

# Sommaire

Abstract . . . . .	iii
Remerciements . . . . .	1
Sommaire . . . . .	2
<b>Introduction</b>	<b>1</b>
<b>1 Sina Weibo et le milieu numérique en Chine</b>	<b>7</b>
1.1 L’Internet chinois : éléments de contexte . . . . .	8
1.2 Médias sociaux en Chine : un paysage morcelé . . . . .	15
1.3 Code, langage et milieu(x) numérique(s) . . . . .	25
<b>2 Les mèmes Internet, objets numériques culturels</b>	<b>45</b>
2.1 Les mèmes : définitions et histoire . . . . .	46
2.2 Mème, mémoire collective et culture . . . . .	53
2.3 Textualité des mèmes et formes d’énonciations numériques . . . . .	61
<b>Problématique et hypothèses</b>	<b>77</b>
<b>3 Réflexions méthodologiques et design expérimental</b>	<b>81</b>
3.1 Sciences sociales du réseau . . . . .	82
3.2 L’analyse de la diffusion sur les réseaux sociaux . . . . .	91
3.3 Méthodes d’identification de mèmes dans un large volume de données . . .	112
<b>4 Résultats et discussions</b>	<b>125</b>
4.1 Un outil de traitement et de visualisation des mèmes . . . . .	127
4.2 Résultats sur un échantillon de mèmes . . . . .	147
4.3 Discussions . . . . .	178
<b>Conclusion</b>	<b>198</b>

<b>Annexes</b>	<b>199</b>
<b>Protocole de recherche</b>	<b>201</b>
<b>Memes : Iconographie</b>	<b>205</b>
<b>Bibliographie</b>	<b>238</b>
<b>Table des Matières</b>	<b>241</b>
<b>Listes</b>	<b>241</b>
Liste des Figures . . . . .	246
Liste des algorithmes . . . . .	247
Liste des tableaux . . . . .	249

# Introduction

PARLER, PARTAGER, COMMENTER, DISCUTER, ÉCRIRE sont les actes quotidiens qui font circuler les informations sur l'Internet. Du silicium des serveurs aux écrans des téléphones, des millions de messages se frayent chaque jour des chemins insoupçonnés à travers les multiples plate-formes du Web. Soutenues par l'appétit d'une industrie florissante, ces nouvelles formes de conversation viennent interroger et renouveler les pratiques médiatiques, politiques, scientifiques et managériales.

Les *mèmes Internet*, petites blagues circulant rapidement sur la Toile, réunissent autour d'eux un grand nombre d'utilisateurs en un temps très court. Satire politique, action commerciale ou simple blague potache, la vitesse et le pouvoir fédérateur de ces simples photos légendées ne cessent de surprendre. Les modèles de leur diffusion, encore largement méconnus, ont jusqu'ici été essentiellement compris par analogie avec ceux du virus biologique. Pourtant, l'observation des tensions entourant l'appropriation de ces puissantes instances médiatiques donne à voir une réalité bien différente.

Ces objets numériques atypiques ont également pénétré les fenêtres des navigateurs Internet en Chine, portés par le sarcasme et l'humour. Le service de microblog *Weibo* lancé par le portail *Sina* a connu un dès son lancement en 2008 un succès fulgurant. La réactivité de ce service de publication instantanée a démultiplié les espaces de conversation en ligne. Rassemblant aujourd'hui plusieurs centaines de millions d'utilisateurs, cette plateforme a amené de nouvelles pratiques de la discussion au sein d'un environnement médiatique chinois traditionnellement très contrôlé (MacKinnon, 2009; Douzet, 2007; Yang, 2008)

La croissance rapide de *Sina Weibo* a été largement soutenue par la politique du gou-

vernement central de Pékin. Le protectionnisme strict appliqué au secteur des industries culturelles et des Technologies de l'Information et de la Communication (TIC) a notamment permis à la firme de se développer dans un environnement non-concurrentiel. Ses homologues américains *Twitter* et *Facebook* ont en effet été bannis du paysage chinois, rendus inaccessibles depuis l'intérieur du territoire (Sullivan, 2012). La valorisation importante sur les marchés d'affaires du titre *Sina*<sup>1</sup> témoigne du succès économique et commercial de l'entreprise. Néanmoins, l'évolution des régulations et utilisations du service lui-même témoigne des tensions constantes entre agenda gouvernemental, désir d'expression des utilisateurs et objectif de rentabilité.

La présente recherche se propose d'examiner les différents régimes d'expression et de discours régissant les usages de *Sina Weibo* grâce à l'étude des contenus à forte circulation - dont les mêmes. Les travaux concernant cette plateforme se polarisent souvent autour des actions de marketing, des stratégies de censure gouvernementale (Ng, 2013b) ou des tactiques de contournement développées par les utilisateurs (Yang, 2014). Nous souhaitons ici dépasser ce clivage en considérant la complexité des relations entre pouvoir politique, industries culturelles et pratiques quotidiennes afin de proposer une lecture plus contrastée (Fernandez et Puel, 2010).

Pour ce faire, nous dresserons un portrait de différents contenus web sur *Sina Weibo* comprenant mêmes absurdistes, scandales politiques, débats de société et campagnes commerciales. Un outil d'analyse et de visualisation de données spécialement conçu nous permettra d'observer dans le détail les interactions entre plate-formes, mots, lieux et utilisateurs lors de leur diffusion.

## Modèles théoriques de la diffusion sur les réseaux sociaux

L'analyse des activités humaines grâce aux informations disponibles en ligne promet de devenir une ressource méthodologique importante dans de multiples champs scientifiques (Schreibman *et al.*, 2007). Stockées en des lieux et formats pas toujours aisément accessibles, ces fragments s'amoncellent pour former d'innombrables bases de données. Parfois désignées par le mot-valise de *Big Data* (Lohr, 2012a), ces traces matérielles pour-

---

1. D'après <http://finance.yahoo.com/echarts?s=SINA+Interactive#symbol=SINA;range=5y>, consulté le 5 Juillet 2014 à 11h10



raient permettre une forme nouvelle d'archéologie des phénomènes humains. L'accès et le traitement de cette mémoire technologique nécessitent un renouvellement de l'écriture scientifique, utilisant à la fois les langages humains et informatiques (Guichard, 2014). La fiabilité des méthodes et la pertinence des découvertes issues de ce vaste champ d'expérimentations restent encore largement à construire (Boyd et Crawford, 2011).

Les données produites par l'usage des services de réseaux sociaux offrent notamment une opportunité nouvelle pour l'étude des pratiques de communication (Zook et Graham, 2007; Nettleton, 2013; Manovich, 2011). La disponibilité en grande quantité de matériels retraçant les échanges quotidiens les plus divers augure de nouvelles formes d'investigation et d'analyse. Depuis la sociométrie de Moreno et Jennings (1938), le modèle des réseaux a progressivement structuré la lecture des phénomènes sociaux (Latour, 1999; Castells, 1989). Guidé par les théories de la complexité (Morin, 1990), cette science des réseaux émergente (Brandes *et al.*, 2013) trouve un terrain de prédilection dans l'analyse des activités sur les réseaux sociaux en ligne. L'observation des dynamiques entourant la diffusion des conversations en ligne fait notamment l'objet de nombreux travaux d'une grande variété méthodologique : modélisation statistique (Steyer *et al.*, 2006), cartographie (Conover *et al.*, 2013; Eisenstein *et al.*, 2012), simulation (Tubaro et Casilli, 2010), etc.

L'analyse des données conversationnelles issues des réseaux sociaux souffre cependant d'assises théoriques encore peu élaborées. La visualisation de conversations sous forme de graphes d'utilisateurs est couramment utilisée dans les travaux sur la communication en ligne. Cette schématisation modélise pourtant les échanges entre individus selon un schéma émetteur-récepteur très basique et largement décrié (Proulx et Breton, 2000). Les réflexions théoriques et méthodologiques sur l'analyse communicationnelle des données provenant de réseaux sociaux émanent le plus souvent de disciplines connexes comme la géomatique (Crampton *et al.*, 2013; Leetaru *et al.*, 2013) ou l'informatique (Bródka *et al.*, 2013; Russel, 2011). Notre travail s'efforce de situer ces pratiques dans une perspective historique. Par conséquent, nous avons été amené à faire dialoguer des approches conceptuelles diverses (espaces, territoires, lieux, rhétorique, discours, énonciation, code...) issues de champs disciplinaires multiples (communication, gestion et systèmes d'information, géographie des technologies, histoire du langage, etc.)

Notre réflexion est conduite autour de l'idée de *milieu numérique*, construit des pro-

tocoles et interfaces technologiques, physiques et institutionnels du web Hui (2012). L’actualisation de ce milieu numérique par la circulation de formes spécifiques de contenus dessine des motifs particuliers que nous nommons *topogrammes*. Ces topogrammes sont connaissables par l’observation des propriétés des structures sémantiques, topologiques et spatio-temporelles représentant le mouvement des informations dans le réseau. Mêmes humoristiques, campagnes publicitaires ou scandales politiques, chaque contenu peut être abordé par la forme particulière de son topogramme, expression de son ou ses milieux numériques.

## Outil d’observation et de visualisation de données

Le terme de *mème*, défini initialement comme une unité de diffusion culturelle, trouve son origine dans une analogie avec le gène biologique (Dawkins, 1989; Blackmore, 2001). La dimension évolutionniste sous-jacente du mot et l’absence de domaines d’application l’ont cantonné pendant longtemps aux marges de la culture scientifique (Jouxte, 2014). Remis au goût du jour avec l’Internet, les mèmes ont depuis été le sujet de plusieurs études, utilisant souvent le virus comme modèle idéal de la diffusion des contenus sur les réseaux sociaux (Leskovec *et al.*, 2005; Adamic *et al.*, 2014). L’idée d’une transmission mécanique par contact représente peu la réalité entrant en jeu dans l’appropriation d’objets technologiques (Orlikowski, 1993) ou informationnels (de Certeau, 1980). De plus, naturaliser des éléments de culture commune en entités autonomes nous semble non seulement réducteur mais aussi dangereux (Elias, 1975).

Aussi nous éloignons nous de la description virale du mème pour observer en détail les topogrammes de différents objets numériques en concevant un dispositif de visualisation de données. À la croisée de l’ingénierie et de la réflexion méthodologique, nous utilisons les technologies de l’analyse des réseaux conversationnels (Weng *et al.*, 2012), du traitement automatique de la langue chinoise (Xue, 2003) et de la visualisation d’information (Cairo, 2012). Nous sélectionnons des contenus de différents types sous la forme de jeux de messages extraits d’un vaste corpus de 200 millions d’interactions retranscrivant l’activité de *Sina Weibo* durant l’année 2012 (Fu *et al.*, 2013). Notre dispositif permet ensuite de montrer différents aspects de leur circulation au sein d’un seul et même espace de représentation.

Cette réflexion s'ancre dans le contexte politico-économique chinois avec le cas particulier de *Sina Weibo*. Nous discuterons du rôle de l'industrie médiatique et des enjeux politiques agissant dans la formation des pratiques en ligne. Nous chercherons également à formuler des recommandations pratiques et méthodologiques concernant la création et l'analyse d'objets digitaux à forte diffusion. Notre dispositif technologique est spécialement conçu avec pour finalité la ré-utilisation sur d'autres terrains, domaines et plate-formes.

## Plan de la thèse

La première partie de cette thèse présentera l'exemple historique du développement de l'Internet en Chine et plus particulièrement du service de microblog *Sina Weibo*. Nous introduirons le concept de *milieu numérique* pour problématiser les rapports existants entre protocoles et usages dans le contexte d'Internet.

La seconde partie discutera la notion de *mème Internet* en proposant une première typologie des différents usages exprimés par ce mot mystérieux. Nous introduirons l'idée de *topogramme* afin de décrire les structures observables lors de la diffusion en ligne.

La partie suivante introduira des réflexions épistémologiques et méthodologiques relatives à l'analyse et au traitement de larges jeux de données issues des réseaux sociaux. Nous discuterons la possibilité d'identifier des mèmes grâce à un protocole expérimental mobilisant différentes approches algorithmiques.

La dernière partie présentera dans le détail les choix retenus lors du développement de notre outil d'analyse et de visualisation. Un échantillon d'une douzaine de mèmes sélectionnés permettra de présenter et discuter les figures obtenues grâce à cet outil. Nous établirons des catégories d'après les modèles de circulation des contenus sur *Sina Weibo*. Enfin, nous nous interrogerons sur la portée et la validité des méthodes d'analyse de données en sciences humaines au regard des résultats de la présente recherche.



# 1

## Sina Weibo et le milieu numérique en Chine

GÉANT DE L'INTERNET, la Chine constitue aujourd'hui une des plus grandes figures de l'ordre médiatique mondial. Avec plus de 538 millions d'internautes recensés<sup>1</sup>, la présence du Web chinois pèse non seulement dans la balance démographique, mais aussi dans la lecture des fondements politiques du réseau. En effet, le discours et les motivations technico-idéologiques qui ont présidé au développement de l'Internet en Chine diffèrent largement de ces précédents américains et européens.

Le regard historique que nous allons porter dans la première partie de ce travail rend compte de la lecture des enjeux industriels et politiques qu'ont porté les cadres du Parti Communiste Chinois (PCC) sur les réseaux d'information. La fermeture de *Facebook*, *Youtube* ou *Twitter* relève en effet autant du protectionnisme économique que de la censure médiatique. L'exemple du service de microblog chinois *Sina Weibo*, qui sera l'objet de

---

1. En Juin 2012 d'après <http://www.internetworldstats.com/asia.htm>, consulté le 12 juin 2014 à 14 :05

notre étude, est sans doute un des exemples les plus parlants des bénéfices apportés par le protectionnisme économique aux industries culturelles chinoises. La réussite commerciale de la firme *Sina*, la surveillance constante exercée par le gouvernement et la possibilité pour les utilisateurs de dialoguer librement sur le service *Weibo* forment un mélange atypique et détonnant que nous allons maintenant décrire.

Tour à tour espace, lieu et territoire, cette plate-forme web articule différents intérêts et usages dans des dynamiques parfois antagonistes. Cette complexité prend corps dans l'existence géographique de ce réseau en Chine. Une relecture historique de l'évolution du concept de milieu nous amène à proposer le concept très simondonien de *milieu numérique* pour décrire l'ensemble des protocoles et pratiques discursives qui préside à la création des objets numériques. Nous discuterons également son actualisation sous des forme spécifiques appelées *topogrammes*.

## 1.1 L'Internet chinois : éléments de contexte

La production de la recherche en sciences sociales s'est longtemps concentrée autour des intérêts politiques et économiques qui entourent les usages et les modes de diffusion de l'Internet en Chine. Les premières études académiques sont issues du champ de l'information-communication et plus spécifiquement du monde anglo-saxon ([Johnson et Liu, 1996](#); [Qiu, 2005](#)). Leurs approches reposent sur le postulat qu'Internet serait un outil favorisant la démocratisation. Ainsi cette littérature peut se résumer en un échange d'arguments autour d'une alternative : l'entrée de la Chine dans la "société de l'information" va de pair avec la modernisation économique et la démocratisation ou à l'inverse avec un contrôle politique accru de l'information et de la société. Les autorités chinoises craignent en effet de voir se développer de nouveaux canaux de circulation et de nouvelles sources d'information échappant à leur autorité traditionnelle sur les médias - et contestant ainsi leur pouvoir. D'un autre côté, le média internet est également perçu par le gouvernement comme un agent indispensable du processus de modernisation politique et économique (dont l'ouverture de marchés commerciaux intérieurs) permettant de faire circuler l'information de façon optimale, en touchant ainsi l'ensemble des foyers. En somme, un Internet "sain" et maîtrisé soutiendrait le développement du système de valeurs gouvernemental,

mais également de l'économie et de la culture. Ici, de nombreuses publications se sont également intéressées au rôle de l'économie en ligne dans la croissance économique de la Chine et aux liens étroits des entreprises web avec la censure (Dann et Haddow, 2008). Ces questions de liberté d'expression et du rapport entre démocratie et médias occupent encore aujourd'hui une large place dans la littérature (MacKinnon, 2009; Douzet, 2007; Yang, 2008).

Les études propres aux usages d'Internet dans le contexte chinois sont toutefois moins nombreuses ou sont généralement le fruit des services marketing des entreprises locales ou internationales en quête de nouveaux marchés (Hwang, 2005; Bergstrom, 2012). Toutefois, on retrouve à la croisée de ces différentes discussions des études plus précises qui cherchent à mettre en relation les dimensions à la fois économique et d'usage du web chinois (Puel, 2009; Fernandez et Puel, 2010). Dans le domaine précis de l'analyse des réseaux sociaux - en anglais *Social Network Analysis* (SNA), les publications internationales en sciences sociales concernant la Chine restent encore peu nombreuses. Le champ de la recherche en informatique propose plusieurs études décrivant les dynamiques des échanges de contenus (Yu *et al.*, 2011) et les dispositifs de censure en place sur *Sina Weibo* (Bamman *et al.*, 2012).

Afin de mieux se situer dans le large paysage de cette littérature, cette première partie rend compte des principaux éléments historiques, politiques et économiques d'infra et d'infrastructure de l'internet chinois.

### 1.1.1 Petite histoire et évolution de l'Internet chinois

L'Internet chinois est aujourd'hui le plus grand réseau national, dépassant depuis 2011 l'Amérique du Nord et l'Europe réunies (CNNIC, 2013)<sup>2</sup>. Alors que l'installation des infrastructures a débuté seulement en 1996, l'expansion en a été fulgurante (Fang et Yen, 2006). 51 millions de nouveaux internautes chinois ont fait leur apparition en 2012, soit une hausse de 10% par rapport à 2011 (CNNIC, 2013).

---

2. 538 millions d'utilisateurs alors même que seuls 40% de ses habitants disposent à ce jour d'une connexion : “ La collection des données se fait grâce à des logiciels, questionnaires en ligne et sondages par téléphones (...) Pour le CNNIC, un internaute est un citoyen chinois âgé de plus de 6 ans qui utilise Internet au moins une heure par semaine ou a utilisé Internet durant les 6 derniers mois ” (d'après le site officiel de CNNIC, agence officielle du gouvernement).

Comme ailleurs dans le monde, les débuts des réseaux informatiques en Chine se déroulent dans le contexte universitaire avec la création en 1994 du CERNET (*China Education and Research Network*) permettant de relier plusieurs grandes universités du pays. Le 17 mai 1994, la Chine effectue sa première connexion au réseau Internet en se reliant au *Stanford Linear Accelerator Center* (SLAC) de l'Université de Stanford aux États-Unis. L'année suivante, les infrastructures *backbones* nécessaires à l'installation de l'Internet à plus large échelle sont déployées (ChinaNet, GBNet, CERNET) et la première licence d'exploitation commerciale est attribuée, effective dès 1996. Le 1er janvier 1997, le *Quotidien du Peuple* lance sa première version en ligne, devenant alors le premier site Internet d'information officielle du pouvoir central Chinois sur Internet. Dans le courant de l'année, le premier FAI privé chinois *China InfoHighway* voit le jour. En novembre, le CNNIC (*China Internet Network Information Center*) est chargé de recueillir et publier des statistiques sur le développement de l'Internet en Chine (Dai, 2007).

Dès son commencement, la mise en place et le contrôle du réseau Internet sont des enjeux importants pour le gouvernement de Pékin. Partie intégrante de leur programme de gouvernance, les dirigeants communistes ont appris depuis longtemps à considérer très sérieusement les nouveaux outils de communication, à la fois comme un risque et une opportunité. Le plus célèbre témoin de cette histoire est sans doute le cinéma soviétique mis en place par Lénine et Trotski dès leur arrivée au pouvoir en 1917. Les *agitki*, courts films de propagande, révéleront de grands cinéastes comme Koulechov, Vertov ou même Eisenstein (Mazuy, 2002). Au début des années 90, la Chine s'est engagée depuis plus de 10 ans dans de vastes réformes (*gaige kaifang*) pour se sortir de l'état de chaos où l'avait laissé la Révolution Culturelle. La politique d'ouverture du pays couplée à un fort protectorat ouvre l'ère du *Made in China* qui concerne l'ensemble des secteurs industriels, y compris ceux des médias et télécommunications dont le potentiel stratégique et économique est énorme. Les dirigeants de Pékin sont également bien conscients que la place forte que la Chine réclame dans le paysage mondial se dessinera notamment par une intégration accrue dans le réseau mondial des TIC. Ainsi, en mars 2000 alors que la population des internautes chinois atteint 16,9 millions d'internautes, le premier ministre Jiang Zemin affirme : “*Internet technology is going to change the international situation, military combat, production, culture, and economic aspects of our daily life significantly.*” (Foster et Goodman, 2000)



Alors qu'Al Gore et l'administration Clinton déploient les *"information superhighways"* aux États-Unis, le programme d'informatisation et de développement des Technologies de l'Information et de la Communication (TIC) *xinxihua* devient une des clés du calendrier politique et économique chinois. Jiang Mianheng, le fils du premier ministre Jiang Zemin, est chargé du déploiement de ce vaste projet. Revenu en 1992 après plusieurs années passées dans les universités américaines puis dans la Silicon Valley chez HP, Jiang Manheng investit largement dans les infrastructures et soutient dès 1999 le lancement du haut-débit en Chine avec la compagnie *Netcom*, aujourd'hui encore deuxième opérateur du pays (Dai, 2007). Fortement influencé par les théories de Alvin Toffler (Tsui, 2007), le gouvernement de Jiang Zemin est bien décidé à ne pas rater la *"troisième vague"* de modernisation de l'industrie : l'informatisation. La mise en place d'une quinzaine de *"golden projects"* informe le mouvement de stratégie globale qui doit dynamiser transversalement tous les secteurs d'activité jusqu'au sein-même de l'administration chinoise. Alors que les *"golden customs"* se chargent des données du commerce extérieur et que la *"golden sea"* devient un outil de communication entre cadres du Parti et administrations locales, les multiples *"golden projects"* sont soutenus par la volonté de mettre en place un *e-government*. Pékin veut faire de l'Internet chinois un espace de dialogue entre citoyens et organes du gouvernement grâce notamment à la mise en place de services administratifs en ligne et d'enquêtes transversales sur la qualité de vie dans les différentes villes de Chine. La gestion proactive de l'Internet national offre également au Parti un vecteur unique pour la diffusion de ses idées politiques. Plusieurs études montrent comment les idées et discours nationalistes sur l'Internet ont bénéficié d'un soutien constant du gouvernement chinois, avec notamment comme objectif l'accès aux communautés chinoises émigrées à l'étranger (Hughes, 2000). Finalement, le média Internet doit servir les intérêts du gouvernement et participer à l'unification territoriale, à l'instar des autres médias dits traditionnels comme la télévision.

### 1.1.2 Censure sur l'Internet chinois

Les modes d'adoption de la technologie Internet par le gouvernement de Pékin illustrent pourtant bien le dilemme constant entre ouverture et protectionnisme qui tiraille la classe politique chinoise depuis la fin de la Révolution Culturelle. L'ouverture au monde et

la politique de réformes du "*nouveau départ*" s'accompagnent pour la Chine de nombreux défis souvent considérés comme extérieurs, que résume avec clarté une phrase célèbre attribuée à Deng Xiaoping : "*Lorsque vous ouvrez une fenêtre pour avoir de l'air frais, vous devez vous attendre à ce que quelques mouches rentrent dans la pièce*". Alors que les promesses d'expansion économique et politique de l'Internet sont bien au rendez-vous, les "mouches" entrées par les fenêtres des navigateurs web commencent à essaimer pour faire de plus en plus de bruit.

Depuis 1998, le Ministère de la Sécurité Publique chinois travaille à la conception d'un projet intitulé *Golden Shield* qui constituerait un fichier global des citoyens chinois utilisable tant pour le contrôle de la démographie que le vol de véhicules ou la sécurité aux frontières (Lyons, 2009). Lors d'un Trade Show intitulé "*Security China 2000*" se déroulant à Pékin en 2000, le projet est présenté publiquement comme une large base de données devant regrouper les informations administratives des citoyens et leurs activités en ligne dans le but de favoriser le travail de la police (Walton, 2001). Titanesque et complexe à réaliser, le projet est peu à peu modifié pour devenir un système de filtrage de contenus et de blocage de sites, basé sur le système du *firewall*. Fang Binxing, un professeur spécialisé dans la sécurité informatique à l'Université de Harbin, est nommé chef-ingénieur du projet *Golden Shield*. Il recrute de nombreux ingénieurs et avec l'aide de l'université de Qinghua et de plusieurs entreprises occidentales (Nortel Networks<sup>3</sup>, Cisco<sup>4</sup>). Ensemble, ils s'attèlent au développement technologique du projet qui devait devenir le système de contrôle de l'Internet chinois aujourd'hui en activité. Baptisé de manière informelle "*The Great Firewall*" (GFW) par analogie avec la "*Grande Muraille de Chine*" (Great Wall), ce système "sociotechnique" hors du commun est aujourd'hui considéré comme une des plus grandes installations d'analyse et de traitement de données en activité. À chaque seconde, GFW traite et scanne des millions de chaînes de caractères issues des requêtes et pages vues par des centaines de millions d'internautes (Winter et Lindskog, 2012). Au-delà de la censure automatique, GFW emploierait aujourd'hui entre

---

3. "Nortel Networks signs contracts valued at over USD 120 Millions during Canadian Prime Minister Jean Chretien's visit to China", PR Newswire, <http://www.prnewswire.co.uk/news-releases/nortel-networks-signs-contracts-valued-at-over-us-dollars-120-million-during-canadian-prime-minister-jean-chretiens-visit-to-china-156751255.html>, consulté le 17 février 2014 à 12 :18

4. "Cisco Leak : 'Great Firewall' of China Was a Chance to Sell More Routers", Sarah Lai Stirland, 05.20.2008, <http://www.wired.com/threatlevel/2008/05/leaked-cisco-do/> consulté le 17 Février 2014 à 12 :39

30 000 et 50 000 personnes<sup>5</sup> : ingénieurs, modérateurs, relecteurs, officiers de police, etc. Phénomène particulier, un groupe de rédacteurs est notamment chargé d'intervenir dans les discussions ou les forums en ligne pour faire valoir le point de vue officiel. L'adage veut que chacun des messages posté soit rémunéré 50 centimes RMB, ce qui a amené les internautes chinois à baptiser non sans humour ces représentants de l'ordre politique en ligne "le Parti à 50 centimes" (*wumao dang*).



**Graphique 1.1.1** – *Jingjing* et *Chacha* sont deux figures créées par les autorités chinoises pour signifier la présence policière en ligne aux internautes. Le nom de ces veilleurs dessinés est une composition issue du mot "police" en chinois (*jingcha*). Source : <http://jswm.newssc.org/system/2008/10/29/011233423.shtml> consultée le 17 Février 2014, à 15 :32

Au-delà de l'activité manuelle de milliers d'employés, GFW opère également plusieurs types de blocages sur les contenus. Techniquement, la plupart des filtrages se déroulent au niveau du fournisseur d'accès avec notamment des adresses particulières qui sont rendues inaccessibles : l'adresse *facebook.com* ou *youtube.com* renvoie une erreur "404 : Le site demandé n'existe pas". Ainsi, de nombreux sites célèbres ne sont pas accessibles (*Twitter*, *Youtube*, *Facebook*, etc.) Les autres blocages effectifs s'effectuent selon les adresses IP ou les serveurs d'attribution de noms de domaine (DNS) menant parfois au blocage de serveurs entiers (*Winter et Lindskog, 2012*). Les URLs des pages de certains sites importants sont également filtrées. Sur Wikipedia notamment, la page "*Tiananmen Square protests of 1989*" est inaccessible depuis la Chine sans que le site Wikipedia soit pour autant intégralement bloqué. Également, des requêtes contenant des mots "interdits" sur les

---

5. "What is internet censorship?". Amnesty International Australia. 28 Mars 2008. Consulté le 17 Février 2014 à 15h12

moteurs de recherche peuvent conduire à des micro-coupures ou un accès restreint au web pendant parfois plusieurs minutes<sup>6</sup>. La liste des sites et mots bloqués n'est pas publiée par le gouvernement et l'ajout sur ses listes s'effectue a priori sur la demande de différentes agences gouvernementales chinoises, sans notification publique. Essentiellement, il s'agit de sites à caractère pornographique (la pornographie est illégale en Chine), de sites liés aux groupes dissidents chinois (Falung Gong, Dalai Lama, Tibet Libre...), de sites du gouvernement taiwanais et d'autres sites revendiquant la liberté d'expression pour la Chine<sup>7</sup>.

Si le GFW est un outil de contrôle politique, il participe également largement au protectorat économique chinois. L'expansion rapide du vaste marché de l'Internet en Chine s'est faite sous un strict contrôle politique et si l'État a largement financé les infrastructures, GFW a été l'un des moteurs de la croissance des grandes sociétés du web chinois. L'absence du géant *YouTube* a notamment permis aux acteurs locaux de la vidéo en ligne de se développer rapidement. *Youku*, son homologue chinois, est aujourd'hui sujet à une valorisation colossale sur les marchés d'affaires. La situation est similaire pour les réseaux sociaux. L'absence de concurrence étrangère due à l'interdiction de *Facebook* en 2008 puis *Twitter* en 2009 a permis aux acteurs chinois de se développer. Leur poids sur le marché intérieur les autorise aujourd'hui à rivaliser avec leurs concurrents américains au niveau mondial tant en nombre d'utilisateurs qu'en revenus directs et indirects générés (CIW, 2012).

Ainsi, GFW a affecté l'économie du pays en profondeur et à ce titre notamment continue d'être une préoccupation première du pouvoir politique. L'évolution technologique de GFW suit de près l'évolution des moyens de contournement des blocages, qui sont nombreux. Le célèbre logiciel *Tor* garantissant l'anonymat sur Internet est aujourd'hui bloqué en Chine (Winter et Lindskog, 2012) ainsi que d'autres technologies communes d'anonymisation (comme le proxy notamment). Pourtant, il reste très facile de "*faire le mur*" (*fanqiang*) et d'accéder aux contenus en contournant les limites du GFW. Les solutions techniques à disposition sont multiples et souvent peu coûteuses à mettre en place ou à utiliser. De nombreux services commerciaux proposent de se connecter depuis

---

6. Testé depuis Shanghai en Septembre 2013.

7. Si ces sites ne sont pas publiés officiellement, une liste est néanmoins maintenue par une entreprise privée sur le site <http://www.greatfirewall.biz>, consulté le 17 février 2014.

d'autres pays à l'aide d'un VPN (*Virtual Private Network*) pour un coût très faible. Le VPN permet d'accéder au Web depuis une machine située dans un autre pays et de bénéficier ainsi de l'accès tel qu'il existe dans le pays où se situe la machine. Le flou juridique qui entoure l'existence de services commerciaux de contournement de GFW témoigne de l'intérêt du gouvernement chinois à protéger largement le marché intérieur en supprimant l'accès aux services majoritaires du web occidental, sans pour autant exercer une traque systématique de chaque personne voulant utiliser *Facebook* ou *Gmail*. De plus, l'absence totale de moyens de contournement interdirait l'accès à des sources précieuses d'informations professionnelles (notamment *Twitter*). Les autorités chinoises ne semblent pas prêt à payer le prix de cette perte d'avantages concurrentiels décisifs pour les entreprises chinoises. Une étude de l'OpenITP parue en 2013 analyse l'usage des outils de contournement de la censure auprès d'un échantillon de 1175 utilisateurs en Chine. Au-delà des solutions technologiques variées, on peut noter que la première raison pour contourner le blocage de l'Internet est l'utilisation des services de *Google* (notamment de *Gmail*)<sup>8</sup>, suivi de la volonté de se rendre sur les sites de réseaux sociaux américains comme *Facebook* et *Twitter*, puis de l'accès aux contenus d'actualité, de vidéo en ligne et de sites à caractère pornographique. Les utilisateurs souhaitant accéder à des contenus à caractère politique ou utilisant l'Internet de manière anonyme pour communiquer de façon plus sécurisée représentent moins de 10% de la population étudiée (OpenITP, 2013). Il est également important de noter que l'immense majorité des internautes chinois n'utilise pas de système de contournement des blocages de l'Internet.

## 1.2 Médias sociaux en Chine : un paysage morcelé

Alors qu'un blocage officiel s'applique sur les plus célèbres sites Internet californiens, de nombreux services se sont développés pour répondre aux besoins et intérêts des internautes chinois. Cette culture particulière des entreprises politiques et économiques du web chinois influe sur l'être-ensemble des usagers et les contenus diffusés sur le réseau.

Profitant de l'absence des grands noms du réseau social en ligne, de nombreux services ont vu le jour sur la Toile chinoise. L'importance du *guanxi* (Yu et King, 2008), élément

---

8. Bloqué en Chine, testé en Juin 2014 à Shanghai et Shenzhen

profond de la culture traditionnelle poussant chaque Chinois à entretenir et exposer avec soin ses relations en société, peut également avoir contribué à créer un terrain idéal pour le développement rapide de ces sites (Yang *et al.*, 2011). Plutôt morcelé, le paysage des SNS en Chine offre une variété de services et d'acteurs qui rassemble les internautes chinois selon leurs centres d'intérêts. *Douban* offre aux jeunes "branchés" de partager lectures, films et musique. *Kaixin001*, plus centré sur les jeux, propose un espace ludique pour les trentenaires au bureau. *Renren* (anciennement *Xiaonei*) est, quant à lui, un véritable clone de *Facebook* et se focalise sur le monde étudiant chinois (Renaud, 2011). Malgré ces nombreux concurrents, le service de messagerie instantanée *QQ* reste le leader incontesté du marché chinois. Aujourd'hui classé 8ème site le plus visité au monde<sup>9</sup>, *QQ* dénombre jusqu'à 100 millions d'utilisateurs connectés simultanément<sup>10</sup>. Quatrième plus grande firme du web mondial, son créateur le géant *Tencent Holdings Limited* a investi depuis le début de l'Internet en Chine dans de nombreux domaines des TIC : jeux, publicité en ligne, e-commerce, etc. Plus qu'une simple messagerie de chat, les services de *QQ* sont multiples : la page de profil de chaque utilisateur (*QZone*) permet de maintenir un blog et d'écouter de la musique (*QQmusic*). Chaque utilisateur peut se créer un avatar en ligne pouvant revêtir de nombreux vêtements et accessoires vendus en ligne (*QQshow*) mais aussi participer à de nombreux jeux multi-joueurs pour tous les âges (*QQ Entertainment*). Le système de monnaie virtuelle (*QCoin*) mis en place pour les achats en ligne a généré dès son lancement en 2005 un nombre important de transactions<sup>11</sup>, poussant même *Tencent* à obtenir une licence bancaire. L'utilisation de la messagerie *QQ*, réseau social avant l'heure, est devenue un véritable phénomène de société porté par la diversification de *Tencent* dans de multiples secteurs sous une marque unique. Au-delà des jeunes et des professionnels de l'Internet, le réseau *QQ* comptait en juillet 2011 plus de 812,3 millions de comptes actifs, faisant de lui le deuxième réseau social mondial après *Facebook*. Du magasin de photocopie de quartier au réseau de prostitution clandestin, *QQ* héberge les discussions quotidiennes et fait pour ainsi dire partie intégrante du paysage des villes modernes. Les chinois échangent plus volontiers leurs numéros de *QQ* que ceux de leurs téléphones portables et ce mode de communication est souvent préféré au mail dans les échanges au

---

9. D'après Alexa.com, consulté le 3 Février 2013.

10. Voir <http://im.qq.com/culture>, consulté le 14 Février 2013.

11. *Central Bank alert on "virtual money"*, People's Daily, 12 Janvier 2007, [http://english.people.com.cn/200701/12/eng20070112\\_340681.html](http://english.people.com.cn/200701/12/eng20070112_340681.html) consulté le 28 Mai 2014

bureau. Convergeant rapidement avec la croissance fulgurante du e-commerce en Chine, les produits dérivés estampillés *QQ* sont devenus une véritable mode en Chine : voitures, téléphones, boissons, etc. Le groupe *Tencent* poursuit son évolution avec le lancement en 2008 de son service de microblog *Tencent Weibo*, qui a connu un véritable succès dès les premiers mois. Aujourd'hui, la firme de Shenzhen continue la conversion de ces utilisateurs *QQ* vers sa plate-forme mobile *WeChat* qui connaît actuellement une très forte croissance, au point de voir les autres services de microblog mis au banc par les utilisateurs. Messagerie écrite et vocale, *WeChat* se diversifie en offrant désormais d'utiliser son compte *QQ* comme moyen de paiement pour de nombreux services du quotidien (taxis, nourritures, etc.)<sup>12</sup>.

### 1.2.1 Microblog en Chine et Sina Weibo

Les sites de *microblogging* (en chinois *weibo*) permettent aux utilisateurs de poster de courts messages composés de photos ou de texte de 140 caractères maximum, puis de les commenter et de les partager avec leurs lecteurs. A l'image de *Twitter*, chaque utilisateur peut souscrire aux fils d'info d'autres utilisateurs afin de recevoir leurs messages et mises à jour.

L'histoire du microblog en Chine débute en 2007 avec plusieurs services se présentant alors comme des clones de *Twitter*. Le service *Fanfou* connaît notamment un succès rapide. De nombreux journalistes l'utilisent pour enquêter et coordonner leurs actions lors de l'arrivée du SRAS ou le tremblement de terre de Wenchuan dans le Sichuan en 2008. *Fanfou* est fermé sur ordre du gouvernement en juillet 2009, suite aux nombreux commentaires suscités par des émeutes s'étant déroulées à Urumqi dans la province du Xinjiang. A peine un mois après cette fermeture, la firme *SINA Corporation* saisit l'opportunité et s'installe en lançant son propre service de microblog intitulé *Sina Weibo*. *Sina Weibo* connaît dès son lancement une croissante soutenue avec plus de 10 millions de nouveaux inscrits par mois et devient en 2012 la plateforme de microblog la plus utilisée en Chine avec 250 millions d'utilisateurs (McKinsey, 2012). Les revenus de *Sina Weibo* ne cessent alors de croître (+19% en 2012<sup>13</sup>) alors que plus de 86 millions de messages sont postés

---

12. *21 million taxi rides have been booked on WeChat in the past month*, Tech in Asia February 12, 2014 <http://www.techinasia.com/wechat-21-million-taxi-rides-booked> consulté le 17 Février 2014 à 15 :36

13. Voir [http://corp.sina.com.cn/chn/Annual\\_Report\\_2011\\_Final.pdf](http://corp.sina.com.cn/chn/Annual_Report_2011_Final.pdf), consulté le 16/02/2013.



chaque jour sur ce service<sup>14</sup>.



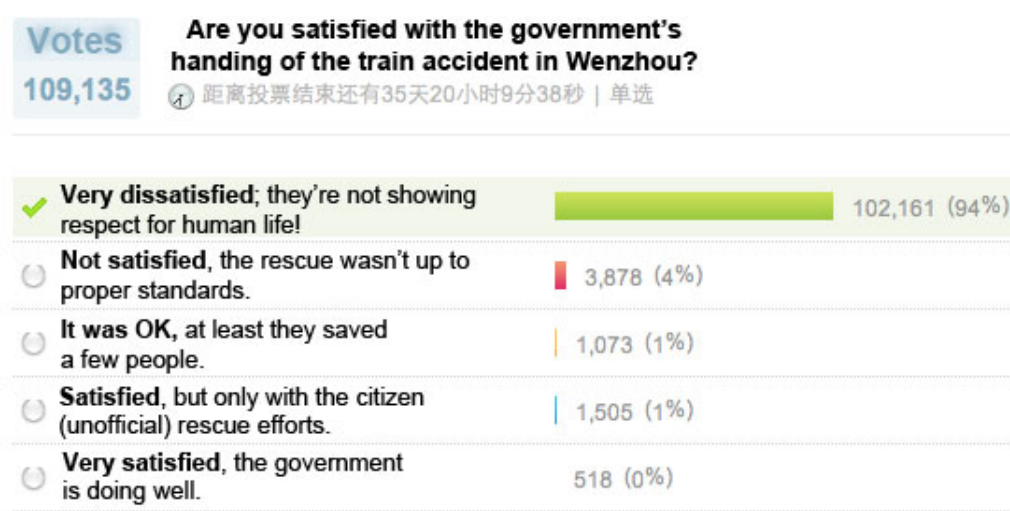
**Graphique 1.2.1** – Capture d'écran de Sina Weibo, réalisé le 9 Avril 2013 à 08 :59

Figure historique de l'Internet chinois, *SINA Corporation* est célèbre pour son portail [sina.net](http://sina.net) et son immense plate-forme de blogs qui en font le fleuron des fournisseurs de contenus en ligne en Chine. Spécialisée dans “l'infotainment” (un mélange très tabloïd d'actualité et de news people), *SINA* est la première compagnie nationale chinoise à avoir été listée au *NASDAQ* dès Avril 2000. Avec son *Weibo*, la firme réussit un coup de force commercial et prouve une fois encore combien la censure gouvernementale est bénéfique à l'industrie du web chinois. Néanmoins, la réussite de *SINA* et de son service de microblog ne se fait pas sans connaître de nombreux ajustements parfois chaotiques. En effet, la stratégie agressive d'acquisition d'audience soutenant la croissance de *Sina Weibo* offre pour garantie aux utilisateurs de pouvoir mieux s'informer et discuter plus facilement en ligne. Dès le début de l'année 2010, les suppressions de comptes utilisateurs et de messages non désirés commencent à se répandre dans le service. Les discussions politiques sont régulièrement effacées et *Sina* se voit contraint de mettre en place un système de censure efficace sous la pression du gouvernement de Pékin. Néanmoins, afin de continuer à garantir la croissance du service, la firme de Pékin laisse une relative liberté aux utilisateurs en étant plutôt souple sur la surveillance des discussions et les actions prises. Des personnalités publiques ou journalistes devenues “weibo-stars” mobilisent régulièrement

14. D'après Sina Corp. Earning Calls - <http://phx.corporate-ir.net/phoenix.zhtml?c=121288&p=irol-EventDetails&EventId=4727394>, consulté le 16/02/2013



l'opinion publique autour de sujets d'actualité, attirant souvent des millions de lecteurs et de commentaires. Plusieurs scandales éclatent en ligne, mettant en cause des officiels et leur famille<sup>15</sup>. Le 23 Juillet 2011, deux trains déraillent sur la ligne reliant Ningbo à Wenzhou inaugurée en fanfare quelques jours auparavant, faisant près de 40 morts et quelques 200 blessés. La colère gronde alors que le gouvernement tarde pendant plusieurs jours à prendre la parole sur ce sujet d'actualité épineux. Sur la toile et *Sina Weibo* en particulier, les discussions vont bon train et les internautes indignés commentent le dernier drame du développement trop rapide de la Chine, où se mêlent détournement de fonds, corruption et sécurité publique.



**Graphique 1.2.2** – Un sondage publié sur *Sina Weibo* (traduction C. Custer, Tech in Asia, 1 Aout 2011), consulté le 24 Février 2014, à 22h12.

*Sina Weibo* désactive alors la fonction de commentaires des messages. Dans les jours suivants, le gouvernement fait enfin une déclaration officielle sur les causes de l'accident de train puis se décide à agir en mettant en garde les internautes trop audacieux de représailles à venir. Les messages controversés sont supprimés, plusieurs comptes d'utilisateurs sont fermés, la police convoque les meneurs des discussions et d'autres mesures d'intimidation sont menées auprès des journalistes et des weibo-stars qui se seraient exprimées un peu trop directement à l'égard du Parti. Dans le même temps, le gouvernement a du mal à se saisir de ce nouvel outil. Alors que les administrations locales, universités et les médias d'État ont plutôt bien réussi le virage de leur stratégie de communication vers

15. Le fils d'un haut-cadre du Parti, arrêté ivre par la police après avoir renversé 5 personnes, annonce : "Mon père s'appelle Li Gang" et se voit immédiatement libéré. Cette impunité provoquera un tollé chez les internautes [http://www.chinadaily.com.cn/china/2011-03/02/content\\_12099500.htm](http://www.chinadaily.com.cn/china/2011-03/02/content_12099500.htm)

le microblog, les membres du gouvernement de Pékin ouvrent des comptes où ils sont parfois raillés, tournés en ridicule et harassés de questions. En Février 2012, les quatre plus grosses sociétés de microblog (dont *Sina Weibo*) annoncent que chaque utilisateur est maintenant contraint de modifier son profil pour mentionner son véritable nom, prénom ainsi que son numéro de carte d'identité. Cette velléité de vérification échoue et est abandonnée quelques semaines plus tard face à la mobilisation des utilisateurs et la difficulté de faire appliquer de telles mesures. Le gouvernement de Pékin édite pourtant une série de règles "*Several Regulations on Microblog Development and Administration Enacted by the Beijing Government*" dont la plus notable sera la possibilité de condamner tous ceux qui auront participé à la diffusion d'informations considérées comme fausses, erronées ou mensongères. Face à la multiplication des actions gouvernementales et à l'apparition d'autres plate-formes, la croissance du nombre d'utilisateurs de *Sina Weibo* est désormais stoppée pour aborder une phase de déclin estimé à près de 10% dans les deux premiers mois de 2014<sup>16</sup>.



**Graphique 1.2.3** – Cours de l'action *SINA* au Nasdaq entre 2009 et Février 2014 - Source : *Google Finance*, consulté le 17 Février 2014 à 15 :28

A la lecture de l'histoire de *Sina Weibo* on constate l'ambivalence des actions officielles du gouvernement chinois dans la réussite économique des entreprises d'Internet. Si la firme *SINA* a bénéficié de prime abord d'un avantage compétitif notoire par l'élimination de la concurrence, elle a par la suite souffert des conséquences du contrôle politique de l'Internet, avec notamment une perte de ses utilisateurs.

---

16. D'après le CNNIC cité dans l'article "*China's Twitter is bleeding users*", 17 Janvier 2014, <http://blogs.marketwatch.com/thetell/2014/01/17/chinas-twitter-is-bleeding-users>, consulté le 17 Février 2014 à 18 :17

### 1.2.2 Sina Weibo, un usage plus ludique que *Twitter*

Dans son article nommé *A Tale of two microblogs*, Jon L. Sullivan (2012) raconte comment l'événement historique de la fermeture de *Twitter* en Chine a vu la communautés des microbloggers chinois se scinder en plusieurs groupes distincts :

- *Twitter* rassemble une communauté avide de libres discussions, souvent très politisées voire radicalement en opposition avec le gouvernement chinois.
- *Tencent Weibo* est utilisé par les utilisateurs de *QQ*, typiquement des personnes aux revenus plus faibles accédant au web depuis leurs mobiles.
- *Sina Weibo* est le favori des travailleurs urbains, souvent plus jeunes ou éduqués, représentant davantage la classe moyenne montante.

Dans la littérature en sciences informatiques, plusieurs articles proposent des comparaisons entre *Twitter* et *Sina Weibo*. Une large analyse quantitative et comparative menée avec des jeux de données des deux services (Gao *et al.*, 2012) nous apprend que le contenu de *Sina Weibo* est davantage corrélé avec des sentiments positifs (analysés automatiquement). Les utilisateurs de *Sina Weibo* parlent davantage de lieux et de personnes alors que les utilisateurs actifs sur *Twitter* s'intéressent plus aux organisations. Également, *Sina Weibo* connaît un pic d'activité le week-end alors que *Twitter* affiche généralement une baisse de régime dans les fins de semaine. Ces différentes indications suggèrent que *Sina Weibo* serait davantage utilisé pour des activités de loisir quand *Twitter* se destinerait à un usage plus professionnel. Une étude s'intéressant aux tendances sur *Sina Weibo* (Yu *et al.*, 2011) indique que la majorité des comptes les plus influents de *Twitter* ont été vérifiés contrairement à *Weibo* où le taux est plus faible chez les grands utilisateurs. La vérification d'un compte se fait par l'authentification auprès du fournisseur de service afin d'attester l'identité de la personne utilisant le compte. C'est un enjeu important pour les figures publiques (marques, stars, hommes politiques, etc.) Cet indicateur nous montre donc l'intérêt professionnel fort entourant *Twitter*, moins pressant dans le cas de *Sina Weibo* où moins de personnes ont ressenti la nécessité de faire officialiser leurs comptes. Sur les deux services de microblog, les utilisateurs inscrits possèdent un réseau de relations identifiables par leur souscription aux fils d'infos d'autres utilisateurs (*follow*). La relation peut être inexistante (*none*), mutuelle (*friend*) ou unidirectionnelle (*follow*, un utilisateur suit un autre mais n'est pas suivi par ce dernier). La comparaison d'échan-

tillons des graphes sociaux issus des deux services (Chen *et al.*, 2012) montre comment les relations sur *Sina Weibo* sont plus dissymétriques et moins réciproques, reflétant une hiérarchie plus forte entre les utilisateurs que *Twitter*.

Un autre facteur important de différenciation entre les deux services est la nature de la diffusion des contenus postés sur *Sina Weibo*. Contrairement à *Twitter* où le texte domine, la majorité des posts de Weibo contiennent des images ou des vidéos (Zhao *et al.*, 2012). Les posts possédant des contenus multimédia (images, vidéos...) sont plus susceptibles d'être diffusés largement et restent en moyenne actifs pour une durée plus longue (Zhao *et al.*, 2012). Également, les contenus sur *Sina Weibo* possèdent une proportion moins élevée de retweets et de commentaires que sur *Twitter* (Zhao *et al.*, 2012; Gao *et al.*, 2012). L'activité de la population de *Twitter* est plus intense, moins tournée vers la diffusion de masse et plus réactive aux influx de nouveaux contenus.

Nous voyons donc que le paysage de *Sina Weibo* se constitue autour de stars et célébrités concentrant l'attention avec des contenus à la diffusion très large. Moins tourné vers l'actualité et la conversation que son homologue *Twitter*, *Sina Weibo* agit comme véhicule de contenus à grande audience, souvent publiés par des personnalités publiques célèbres. Des études quantitatives montrent bien que les contenus les plus échangés et discutés concernent les loisirs et divertissements, la mode, la santé, etc. (Li *et al.*, 2013). Les messages à caractère humoristique (texte, images et vidéos) occupent également une place prépondérante dans les échanges des utilisateurs, contrairement à son homologue américain *Twitter* dominé plutôt par les sujets d'actualité (Yu *et al.*, 2011). *Sina* poursuit ainsi son rôle historique de leader chinois de *l'infotainment*.

Pourtant, la population de jeunes urbains qui soutient la croissance de son service de microblog reflète aussi les transformations en cours dans la société chinoise. Les journalistes et spécialistes de l'information sont les premiers à se saisir de ce nouveau média. Dans un discours à Stanford en 2013, le PDG de *Sina* Charles Chao explique :

*Le plus grand changement apporté par le microblog en Chine concerne d'abord l'industrie des médias elle-même. Aujourd'hui, plus de 30% des actualités ont d'abord été reportées sur Sina Weibo avant d'atteindre les médias traditionnels. Le rôle des médias traditionnels a été déplacé vers un traitement des*

*informations en profondeur (in-depth reporting).*<sup>17</sup> (traduction de l’auteur)

L’omniprésence des supports mobiles (smartphones, tablettes)<sup>18</sup> permet en effet des modes de traitement de l’information jusqu’ici inconnus qui bousculent les hiérarchies très contrôlées des salles de rédaction chinoises. Alors que la population urbaine croît rapidement, le smartphone est “*the first big urban purchase*” (Wallis, 2013) pour les nouveaux arrivants en ville et représente un outil indispensable de participation à la société. En 2008, la Chine était le seul pays en Asie où les moins de 30 ans possèdent plus d’amis en ligne que hors ligne (Hinckley, 2009). Ainsi, les réseaux sociaux jouent un rôle primordial dans la socialisation urbaine et viennent changer les modes d’expression. Le lectorat chinois a perdu toute confiance dans la plupart des médias traditionnels suite à l’absence répétée de courage et à la rétention d’informations cruciales dans les dossiers importants animant le pays.

Le microblog s’installe comme une nouvelle source de confiance pour des millions de citoyens voulant comprendre et prendre part aux changements cruciaux de la société chinoise moderne. Un rapport de l’Institut de Journalisme Reuters à l’Université Oxford paru en 2013 montre comment les usages du microblog ont amené des transformations dans le quotidien des journalistes chinois. Le journalisme d’investigation a notamment connu un essor important grâce au renouvellement des sources et une large diffusion en ligne des sujets. *Sina Weibo* n’a pas amélioré nécessairement la qualité de leurs investigations, mais a par contre permis une plus grande dissémination. Il est à noter que les spécificités de l’écriture chinoise rendent possible l’écriture d’un court texte en 140 caractères, alors qu’une telle longueur autorise seulement une courte phrase dans une écriture utilisant un alphabet latin. La mobilisation des utilisateurs pour la protection des journalistes a également joué un rôle important ainsi que le renforcement de procédés de vérification existant depuis longtemps sur les forums du web chinois. Très populaire dans les années 2000, Le “*moteur de recherche de viande humaine*” (*renrou sousuo*) est une forme de traque d’individus en ligne réalisée par un large nombre d’internautes à partir d’un nom ou d’une

---

17. Charles Chao, PDG de Sina pendant la Stanford Graduate School of Business China 2.0 tenue le 3 Octobre 2013. Disponible en vidéo <http://www.youtube.com/watch?v=tlliivJKHk8>, consultée le 19 Février 2014 à 11 :23

18. Selon l’Universal Telecommunication Union, “*la Chine dépasse 1 milliard d’abonnements mobile, avec 400 millions d’utilisateurs d’Internet mobile dépasse ainsi les États-Unis comme leader du marché des smartphones*”, <http://mobithinking.com/blog/china-top-mobile-market> consulté le 24 Février 2012.

photo. Il s'agit souvent de retrouver quelqu'un désigné comme "coupable" (d'adultère, de corruption, etc) en réunissant un maximum d'informations à travers la Toile afin d'identifier ou de localiser la personne. Devant les dérapages rapides de ce type de procédés, les questions d'éthique sont au cœur des discussions qui entourent le journalisme en ligne. En effet, l'usage des médias sociaux a permis à certains journalistes de faire pression sur les pouvoirs publics, amenant parfois à la censure de leurs travaux, mais a également permis une très large auto-promotion pour de nombreux journalistes devenus des stars de *Weibo*.



**Graphique 1.2.4** – La trace des commentaires supprimés par Sina est encore visible - Page 677 à 708, les commentaires ont été supprimés, soit approximativement 4% messages supprimés (589 messages sur 13452, à raison de 18 à 20 messages par page). Capture d'écran effectuée le 29 Janvier 2013 à 12 :32 :42, <http://www.weibo.com/1701401324/zeoBquVKi>, consulté le 16/02/2013.

Le contrôle des contenus sur *Weibo* est donc une réalité quotidienne et a été depuis son lancement la source de plusieurs études. Les formes les plus courantes sont : la suppression de posts, la suppression de comptes utilisateurs et le blocage de mots-clés. Le blocage de mots-clés s'effectue dans le moteur de recherche interne du site ("pas de résultats" quand vous cherchez un mot bloqué) et plus récemment par l'impossibilité de poster un message contenant des mots ou des adresses web bloqués (Ng, 2013a). La pratique de la suppression de comptes s'est intensifiée en 2013<sup>19</sup> avec notamment la suppression de millions de "zombies" présents sur le site. Les "zombies" sont des comptes utilisateurs créés par des robots dans le but de reposter automatiquement des contenus et d'augmenter le trafic sur le site. Les exigences des annonceurs publicitaires de plus en plus présents

19. "Over 100,000 Sina Weibo Accounts Shut Down or Penalized for Govt Rules Violations" par Gabriela Vatu, 14 November 2013 <http://news.softpedia.com/news/Over-100-000-Sina-Weibo-Accounts-Shut-Down-or-Penalized-for-Govt-Rules-Violations-400289.shtml> consulté le 17 Février à 16 :42

sur le site ont obligé *Sina Weibo* à faire la chasse aux robots sur son site, faisant ainsi diminuer le nombre de comptes actifs de manière significative. La firme Sina est garante auprès du Ministère de la Sécurité Publique chinois des contenus qu'elle diffuse et effectue à ce titre une surveillance constante pour supprimer les messages "non-conformes". Lors de nos recherches, nous avons constaté que l'interface de *Sina Weibo* garde la trace des commentaires supprimés par le système d'administration. Dans les messages supprimés se trouvent à la fois des posts d'utilisateurs "zombies" et les posts jugés incorrects par les administrateurs. Au total, la suppression des messages s'effectue avec un taux estimé à environ 16%, allant jusqu'à plus de 50% dans certaines provinces comme Ningxia ou le Tibet contre seulement 12% à Beijing (Bamman *et al.*, 2012).

## 1.3 Code, langage et milieu(x) numérique(s)

L'espace d'expression offert par l'Internet chinois et ses services de réseaux sociaux héberge donc une variété de pratiques, de questions et de réflexions qui nécessitent d'être appréhendées avec des outils de problématisation et d'analyse spécifiques. Dans la troisième et dernière partie de ce chapitre introductif, nous allons donc nous pencher sur les concepts existants dans la littérature scientifique sur l'existence *in situ* des objets numériques. Afin de mettre en perspective le Web chinois et l'histoire de ces objets, nous introduirons notamment le concept de *milieu numérique*, héritier de l'idée complexe de milieu que nous explorons ci-après.

### 1.3.1 Lieu, espace, territoire et technologies

Géographie, management et diffusion de l'innovation, histoire des technologies, *cultural studies* ou études "nationales", les travaux qui s'intéressent aux relations entre technologies, espace, lieux et territoires sont nombreux et offrent un paysage riche où se croisent de nombreuses disciplines scientifiques. L'influence des réseaux de transports sur l'expérience humaine et le développement des villes a notamment été largement étudiée (Offner, 1993; Doulet, 2001). L'Internet a également fait l'objet de nombreuses études monographiques (par pays) ou comparatives, une étude à l'échelle mondiale présentant en effet des problèmes de données et évidemment d'échelle (Dupuy, 2004). En Chine où l'urbanisation



produit actuellement une vaste migration des ruraux vers la ville, l'utilisation des réseaux sociaux médiatise bien souvent les choix de lieux et les rencontres des nouveaux arrivants. De nombreux groupes de discussions réunissent par exemple les nouveaux acheteurs d'immobilier qui échangent leurs stratégies d'achat et de défense de leurs droits et de leurs biens (Li *et al.*, 2013). L'usage des réseaux sociaux revêt pour les nouveaux arrivants une importance capitale, notamment dans la recherche de groupes similaires et l'échange d'expériences. A travers le navigateur, ils s'approprient la ville étrangère pour s'en construire peu à peu une représentation à leur image.

La carte notamment joue sur le web un rôle important d'abord en tant qu'illustration, puis plus récemment d'interface avec le réel. Loin d'être figée par le territoire, la carte le décrit sous un ou des angles particuliers (Brunet, 1987; Jacob, 1992). Le développement de standards comme le système GPS (Haklay *et al.*, 2008) et de services et outils de cartographie ont contribué à une appropriation de la pratique cartographique par un nombre croissant de personnes (Crampton, 2009). De nouvelles formes de données produites par les utilisateurs parfois appelées “*volunteered geographic information*” (Elwood, 2008) utilisent les services en ligne comme Google Maps pour dessiner un “*miroir du monde offline*” (Graham et Zook, 2011). Le courant dit de la néogéographie fait usage du GIS et des outils en ligne (Google Maps, Flickr, etc.) pour comprendre les pratiques de ces nouvelles formes de “*géographies volontaires*” (Turner, 2006). Ce rôle croissant de la cartographie dans l'usage d'Internet produit un “*Geoweb*” constitué de données et métadonnées spatiales (Crampton, 2009). Le point d'entrée unique qu'offre le marqueur spatial (*geotag*) réunit souvent de vastes quantités de données disparates : Open Data territorial, géo-localisation, GIS, POI<sup>20</sup>, etc. (Torrens, 2010). Cette présence accrue dans le réseau dessine l'enjeu non seulement de cartographier le monde, mais également de cartographier le réseau lui-même, ouvrant ainsi de nouvelles voies pour découvrir la construction sociale des espaces par des pratiques individuelles et de groupe. Dans cette étude, nous avons donc choisi d'interroger les objets numériques afin de comprendre comment se structurent la parole et la conversation dans le contexte unique de l'Internet chinois. Afin d'articuler les multiples dimensions d'analyse qui viennent enrichir notre réflexion, il nous faut donc brosser un portrait en large de l'Internet chinois, en le considérant tour à tour comme un *espace* structurant pour les actions des internautes qui le pratiquent, comme un *territoire*

---

20. GIS : Geographic Information System ; POI : Point-Of-Interest



sujet aux relations de pouvoir actualisées par les groupes et individus et enfin comme un *lieu* habité par ceux qui y construisent chaque jour des significations communes. Nous proposons ici une revue sélective des quelques travaux à même de nous apporter des éclairages pertinents dans la vaste littérature s'intéressant aux dimensions géographiques des TIC.

## Code / space : l'espace transductif des TIC

Dans leurs recherches autour de la géographie des technologies numériques, Dodge & Kitchin ont travaillé à développer le concept de code comme un élément fondateur des espaces modernes dans lesquels nous évoluons. Reflétant l'importance croissante accordée aux TIC dans l'environnement urbain, ils citent un travail sur la production automatique des espaces : “*De plus en plus, les espaces de la vie quotidienne nous parviennent chargés de logiciels (software)*” (Thrift et French, 2002) En effet, si le projet urbain a été guidé pendant le demi-siècle dernier par l'apparition de la technologie automobile dans l'espace des rues, les TIC paraissent prendre le relais avec l'idée d'une ville intelligente et connectée, connue sous le nom de “*smart city*” (Ascher, 2009; Picon, 2014). Dodge & Kitchin ont donc fait du *code* une des pierres d'angles de l'appréhension de l'espace dans leur travail en le définissant comme suit :

*“an instruction or rule that has a single outcome determined by a binary logic (yes/no). The combination of these individuals logic rules produces code (program).”* (Kitchin et Dodge, 2011).

La part croissante des TIC dans nos espaces quotidiens amène les auteurs à envisager l'espace dans son interaction avec le code, symbolisant la suite d'instructions machiniques et électroniques qui permettent à un espace de remplir sa fonction. Dans un article intitulé *Flying through code/space : the real virtuality of air travel*, Dodge & Kitchin analysent la structure des espaces aéroportuaires. De l'achat des tickets jusqu'au vol des avions en passant par la gestion des bagages, le bon fonctionnement d'un aéroport est entièrement régi par les longues successions d'instructions du code. Ici, Dodge et Kitchin proposent le concept de code/space pour décrire ce type d'espace spécifique où lorsque le code échoue (*failure*) alors le code/space tout entier échoue (Dodge et Kitchin, 2004). L'exemple de l'aéroport est parlant : si le système de check-in des bagages ou les machines responsables

du contrôle de sécurité des passagers ne fonctionnent pas, alors l'espace aéroportuaire ne peut exister en tant qu'aéroport. L'analyse de la spatialité ne se situe alors plus dans un domaine sémantique ou narratif, mais plutôt dans les processus et opérations qui s'y déroulent et le code et les technologies y jouent un rôle primordial :

*“Code is employed as the solution to a problem, a particular kind of transduction is occurring.” (Kitchin et Dodge, 2011).*

L'espace n'est pas un donné mais s'explique plutôt comme : *“une forme d'ontogenèse (en perpétuel devenir-au-monde), l'espace est une pratique ; un faire ; un événement (...) qui ne pré-existe pas à son faire (doing)”* (Kitchin et Dodge, 2011). L'espace est considéré non pas comme une production, mais comme une *transduction*. Reprenant le travail de Simondon sur l'individuation par la technologie, Dodge et Kitchin présentent l'espace comme une pratique qui comprend les actes, actions, occurrences, mémoires, perceptions, etc. d'un groupe d'individus s'y trouvant. La fonction de l'espace est structurée par les individus et le code y est considéré comme une entité agissante. Dans le *code/space*, la relation dyadique entre code et espace est bijective : l'un ne peut aller sans l'autre. En terme simondonien, la transduction ne peut être assurée sans code. Si l'exemple de l'aéroport illustre bien cette nécessité du code dans le devenir-espace, Dodge et Kitchin ont également identifié d'autres catégories où cette relation est plus ténue : les *coded spaces*, qui peuvent poursuivre leurs fonctions même lorsque le code échoue ; les *background coded spaces* où les processus de transduction induits par l'espace ne s'appuient pas nécessairement sur le code, mais proposent néanmoins des possibilités de l'activer (machine éteintes ou inactives, etc.)

L'analyse fonctionnelle des rapports entre espace et technologie de Dodge & Kitchin montre comment les TIC peuvent être un facteur *transductif* pour les individus se mouvant dans les espaces de leurs vies quotidiennes. Si nous appuyons pleinement ce constat, il nous semble que le parti-pris des auteurs de considérer le “code” comme une abstraction incluant uniquement les instructions ou “logiques machiniques” ferme la porte à l'immense densité des activités symboliques qui se jouent dans l'usage des technologies. Comment notamment considérer les “contenus” du web dans cette grille de lecture ? Comment resituer dans une perspective historique les logiques de médiation de l'espace par les technologies de l'écriture ? Il nous semble en effet que la faillite fonctionnelle (*failure*)

des code/space précède l'arrivée des technologies et s'opère déjà à un niveau symbolique - la fonction de l'espace du Palais du Louvre après la chute des rois de France se voit radicalement modifiée. La transduction opérée lors de la pratique d'un espace s'effectue dans un jeu d'appropriation symbolique qui passe notamment mais pas seulement par les technologies. Les technologies du langage et de l'information jouent notamment un rôle crucial dans l'affirmation du récit symbolique (*narrative*) qui construit l'espace. L'activité du code dans la structuration des code/space de Dodge & Kitchin existe sous une forme non seulement fonctionnelle mais également sémantique, voire phatique ou même esthétique comme l'a décrite Jakobson dans ces analyses des fonctions du langage ([Jakobson et Halle, 1956](#)). Au-delà de sa dimension machinique, le code possède les caractéristiques d'une *poiesis* dépassant l'idée simple de fonctionnalité pour exister dans la complexité d'une écriture comme traduction du langage humain et machine.

## Codes, discours et territoires des technologies

Le code serait davantage à comprendre comme un mode d'expression humain à travers la technologie, actualisant l'*épistémè* décrit par Foucault dans *Les Mots et Les Choses* comme élément fondamental de la pensée d'une époque et sa considération pour le monde ([Foucault, 1996](#)). État des connaissances scientifiques et littéraires, l'*épistémè* existe comme somme des savoirs d'une époque, présupposée traduite en un regard sur le monde. Le *code* exprime les savoirs d'aujourd'hui dans de nombreux langages écrits. Le code source d'une page Internet, d'un programme informatique ou d'un driver hardware ne s'écrit pas seulement en langage "machine" mais fait appel à plusieurs langages informatiques et humains. A la fois production savante, outil scientifique, vecteur d'expression et interface des savoirs, le code constitue l'expérience narrative du monde par les TIC. Possédant de nombreux mots, aspects et syntaxes issus de multiples langues humaines, les replis de l'écriture informatique laissent transparaître de tout bord leur origine littéraire. L'alphabet se voit augmenté de nombreux caractères qui le rendent compatible avec l'encodage des bases de données - pensons à l'Unicode notamment ([Guichard, 2014](#)). Rappel à l'ordre, l'arrêt soudain de l'ordinateur ou la perte d'un fichier nous laisse frappés d'illettrisme. Seuls une minorité de personnes "lettrées" de l'informatique peuvent déchiffrer et comprendre le "bug". Ainsi, la définition du "code" de Dodge & Kitchin doit être étendue pour recouvrir plus largement les pratiques symboliques liées aux activités de l'écriture

du code dans ces espaces.

Le code ainsi redéfini nous ramène alors à une lecture foucaldienne du discours dans sa relation intime avec le territoire (Foucault, 2004). Dans ces nombreux travaux sur la généalogie, Michel Foucault cherche à comprendre comment les relations de pouvoir créées par les discours portés sur les objets président à la production de territoires et d'interdits comme autant de sujets de ces discours. Nous définissons la *discursivité* comme le processus de construction de ces discours. Historiquement, l'Internet a été très tôt sujet à l'appropriation par le discours de nombreux groupes actifs dans une volonté de territorialisation. La métaphore géographique et spatiale a structuré le vocabulaire de l'Internet dès sa création (Graham, 1998) : site, cyberspace, etc. L'*Electronic Frontier Foundation* se charge de protéger l'u-topie qu'est Internet avec sa fameuse *Déclaration d'Indépendance du Cyberspace* (Barlow, 2001). A l'opposé du spectre, les autoroutes de l'information in-forment le paysage comme autant de géogrammes massifs (Berque, 1999). L'appropriation des protocoles du réseau, notamment par la lutte pour le respect de standards ouverts ou partagés, s'ancre également dans les pratiques du discours. Les mots *free* et *open* cristallisent l'histoire des revendications territoriales de l'Internet (Blondeau, 2000). L'autre grande métaphore constitutive de l'Internet est textuelle avec ses pages, langages et hypertextes (Vandendorpe, 1999). La formule choc "*Code is law*" (Lessig, 2006) résume l'idée que les processus textuels du code mettent en jeu un ensemble de rôles, protocoles et mises en scène qui agissent comme autant d'autorités à travers le discours. La jurisprudence fait loi, comme écrit sur les murs des bureaux de *Facebook* à Palo-Alto : "*Code wins arguments*"<sup>21</sup>. La territorialisation de l'Internet se fait ainsi au travers d'un ensemble de pratiques discursives, méta-grammaire des discours en ligne. La confrontation symbolique au sein des territoires numériques se poursuit dans le discours, réifié dans les pratiques du code.

Sur l'Internet chinois, les pratiques de censure de l'écriture en sont le reflet le plus frappant. Blocage de mots-clés, détournements de langage, suppression et modification de texte sont l'expression de cet affrontement de discursivités parfois antagonistes. La Grande Muraille gouvernementale scanne les masses de texte pour reconnaître et stopper des

---

21. Phrase mentionnée dans la lettre écrite par M. Zuckerberg aux investisseurs lors de l'IPO de *Facebook* [http://www.sec.gov/Archives/edgar/data/1326801/000119312512034517/d287954ds1.htm#toc287954\\_10](http://www.sec.gov/Archives/edgar/data/1326801/000119312512034517/d287954ds1.htm#toc287954_10)

mots tels que “*Printemps arabe*” ou “*événements de Tian-An Men*” (MacKinnon, 2012). Néanmoins, l’état actuel des techniques de *data mining* ne permet pas encore de déceler les phénomènes langagiers comme les jeux de mots ou l’ironie. Bien souvent, les internautes chinois choisissent l’humour pour permettre à leurs idées de se frayer un chemin. Revêtant leurs masques de chat, les internautes chinois sont devenus spécialistes dans la publication de jeux de mots, chansonnettes et petites vidéos d’animaux, comme autant de couperets cinglants pour railler les officiels trop pompeux de Pékin. Dans la guerre de l’information que se livrent sans cesse censeurs et internautes, de simples photos truquées de crabes et de lamas peuvent devenir héroïques. Ces blagues numériques, d’apparence inoffensive, font chaque jour le tour de la Toile chinoise, portant en elles toute la subversion d’internautes aspirant à plus de liberté. Début 2010, alors que pleuvaient les longs discours pieux du Parti sur l’harmonie de la nouvelle société (en chinois *hexie*), on voit apparaître en ligne des essaims de crabes de rivière (se prononçant également *hexie*) couverts de chaînes en or criant : “*Vive l’harmonie*” au volant de leur limousine. Devenus aujourd’hui une image de la corruption des hauts dignitaires du Parti, on croise régulièrement dans les commentaires d’un article officiel un petit crabe de rivière, tel un rapide rappel posté par un lecteur.



(a) (hexie) : Harmonie / Discours du PCC sur l’harmonie



(b) (hexie) : Crabe de rivière / Mème satirique

**Graphique 1.3.1** – 和谐 VS 河蟹 - Mot de la semaine : Crabe de Rivière - China Digital Times du 21 Mars 2012 - Licence Creative Commons <http://chinadigitaltimes.net/2012/03/word-of-the-week-river-crab>, consulté le 15 Février 2013.

On voit bien comment le code décrit un territoire sujet à l’autorité politique sous la forme d’un système de *data mining* cherchant à mettre en forme le discours. La fonction du *code/space* sémantique que forme ici l’espace de l’Internet est réécrit par un jeu de langage. La circulation d’objets digitaux permet de reterritorialiser cet espace en apparence régi par un code strict de censure.

## Les lieux des technologies

Dans son célèbre livre sur les Arts de Faire, **de Certeau (1980)** considère la ville comme un texte dont chaque piéton énonce et révèle (*performe*) des sens nouveaux par son activité de marcheur. Actualisant l'espace urbain par sa marche, l'habitant de la ville s'approprie des lieux restant néanmoins partagés avec d'autres. Au détour des rues, le sens commun des lieux urbains se construit avec les multiples énonciations de ceux qui les habitent et les font vivre. Cette magnifique image de la poésie du texte urbain met en lumière la dualité que nous avons abordée précédemment : l'espace ne peut fonctionner sans les pratiques de ses habitants. Plus encore, l'être-ensemble et le devenir-soi procèdent de la construction de lieux communs, *poiesis* des espaces habités. Les TIC font aujourd'hui souvent partie intégrante des lieux que nous habitons. **Graham (1998)** dans son travail sur l'étude des lieux et de leur rapport à la technologie identifie trois types majeurs d'approches dans la littérature :

1. L'approche "*substitutive*" ou "*transductive*" qui voit dans l'arrivée des TIC la disparition de la valeur des lieux : un idéal de proximité utopique (**McLuhan, 1962**) ou un discours dystopique sur leur proche disparition (**Virilio, 1998; Augé, 1995**). Ces considérations sont les formes traditionnelles du débat accompagnant l'innovation dont sont férus la communication industrielle et la critique des médias (**Ramonet, 2001**), restant souvent purement prospectif et faisant peu de cas des usages.
2. Plus modérée, l'approche qualifiée par Graham de "*co-évolutionniste*" s'interroge sur la façon dont les interactions dynamiques des espaces virtuels (*space of flows*) et réels (*space of places*) (**Castells, 2009**) produisent de nouveaux lieux . Ces études économiques et sociales prennent la forme d'une médiologie de l'espace adaptée aux études stratégiques pour l'urbanisme et l'implantation des télécommunications. Son interprétation par les aménageurs lui donne parfois une dimension déterministe peu utile à la compréhension des phénomènes liés à l'appropriation et l'historicité des technologies (**Offner, 1993**).
3. Une dernière approche plus récente se cristallise autour de l'idée de relations et de réseaux. Afin d'éviter l'écueil des causalités directes et de la notion fataliste d'impact, les lieux sont présentés comme "*des moments articulés dans un réseau de sens et de relation sociales* " (**Massey, 1993**), des assemblages entre objets matériels



“actants” (Latour, 1996), individus et groupes sociaux. Cette approche s’intéresse davantage au lieu comme une géométrie sociale en mouvement, liée au temps et à la situation (May et Thrift, 2001) et refuse l’idée d’une existence “virtuelle” commune à différents objets (Bingham, 1996).

En privilégiant une approche dynamique et relationnelle des lieux comme constructions sociales de sens (Kyle et Chick, 2007), la technologie perd son rôle déterministe de productrice d’espaces et d’usages pour devenir une actualisation d’un espace-temps géographique et historique par des groupes d’individus. Comme le note Cresswell dans son travail sur les lieux : *“places are practiced. People do things in place.”* (Cresswell, 2004). Il propose trois aspects pour décrire l’actualisation des lieux par leurs pratiques : *location* (un point dans l’espace, *“the ‘where’ of place”*), *locale* (les aspects visibles et tangibles du lieu, *“the way a place looks”*) et *sense* (*“the feelings and emotions a place evokes”*). Néanmoins, cette définition ne permet pas d’appréhender l’existence des lieux en ligne (le site web d’un lieu est-il une *locale* ou une partie du *sense*?). Graham et Zook propose le concept de DigiPlace : *“DigiPlace - that is, the use of information ranked and mapped in cyberspace to navigate and understand physical places (...) In other words, DigiPlace represents the simultaneous interaction with software (information) and ‘hard-where’ (place) by an individual.”* (Zook et Graham, 2007). S’inspirant des travaux de Harley sur le pouvoir du cartographe, leur travail sur le rôle de Google Maps dans la présence des lieux sur Internet met à jour l’interaction et l’hybridation entre existence spatiale et existence en ligne constituant les DigiPlace.

Les phénomènes de transduction à l’œuvre dans les pratiques spatiales de l’Internet sont donc le reflet des états du réseau à un moment donné. Les lieux eux-mêmes forment ainsi un réseau basé sur leurs relations et similarités. Brunet dans son *Vocabulaire de la Géographie* propose notamment l’idée de synapses : *“espaces ou lieux par lesquels on passe, par où l’on communique, les isthmes, les détroits, les estuaires, les carrefours, les ports et les ponts, etc. ;”* (Brunet, 1972). Envisagés par leur fonction “synaptique”, les lieux deviennent alors un construit social au rôle clair, évitant l’écueil d’une qualification en-soi. Aéroports, hypermarchés, aires d’autoroutes ou zones industrielles ont été qualifiés de *non-lieux* par Marc Augé, produits d’une hyper-modernité *“qui ne peut se définir ni comme identitaire, ni comme relationnel, ni comme historique.”* (Augé, 1995). Réfutée

plus tard par l’auteur lui-même, cette idée de lieux simplement *produits*, et non construits fait l’impasse sur la tension d’usage, le devenir lieu que modèlent ses utilisateurs assidus ou épisodiques, ceux qui y travaillent voire même y habitent. Le lieu n’est pas nécessairement “patrimonial” comme produit d’une histoire, mais peut “ *être ou ne pas être un non-lieu selon le statut de l’individu envisagé.* ” (Debarbieux, 1993). Les hypermarchés et leurs galeries marchandes sont des haut-lieux de socialisation et de rencontres pour les adolescents vivant en banlieue (Matthews *et al.*, 2000) mais paraissent froids et inhumains aux habitants des rues du centre-ville.

L’étude menée par Puel, Pons et Xiaoting autour des pratiques sociales environnantes les cafés Starbucks de Beijng (Chine) montre également comment la stratégie marketing de la firme s’appuie précisément sur cette “absence patrimoniale” pour faire sa place au travers du vaste territoire chinois. La non-existence d’un “*bon café avec Internet*” dans les villes chinoises offre la possibilité aux usagers de ces cafés récemment apparus de s’y rendre pour utiliser l’Internet ou retrouver leurs amis (Puel *et al.*, 2007). Ainsi, l’absence d’histoire n’interdit pas la constitution de pratiques communes à forte valeur symbolique se comprenant ici dans des réseaux d’appartenances mondiaux (jeune, dynamique, urbain, etc.). Dans son livre *The Great, Good Place*, Ray Oldenburg propose le concept de tiers-lieu pour décrire ces lieux qui, séparés de l’environnement de travail ou de la maison, permettent de se socialiser. Jouant un rôle majeur dans la construction de communautés et la constitution d’une société civile, ces tiers-lieu doivent répondre à certains critères comme : un cout d’accès nul ou modeste, une très bon accessibilité, la présence régulière des “*habitués*”, un lieu accueillant et confortable et enfin un lieu où l’on peut rencontrer facilement de nouvelles personnes (Oldenburg, 1999). Ces tiers-lieux forment donc avant tout un réseau de lieux défini par les usages. L’idée de tiers-lieu virtuels a également été mentionnée pour désigner les chatrooms ou les plate-formes de réseaux sociaux en ligne remplissant des fonctions sociales similaires (Soukup, 2006).

Nous voyons donc qu’en considèrent les lieux de l’Internet, nous mettons à jour un ensemble de pratiques ne s’intéressant plus nécessairement à l’ordre du discours (et aux pratiques de censure notamment) mais plutôt aux phénomènes d’individuation, notamment au travers des multiples actes d’énonciation qui forment les pratiques et usages du web. Dans cette étude nous cherchons donc à comprendre de manière plus profonde comment les internautes habitent leur Internet. Il ne s’agit pas d’analyser le discours en



termes de relations de pouvoir mais plutôt d’essayer de distinguer comment les circulations des objets numériques structurent les pratiques des internautes, comme autant de lieux habités quotidiennement.

### 1.3.2 Le milieu : richesse et désuétude

Afin de problématiser les relations entre protocoles du discours et pratiques locales d’appropriation, nous avons choisi d’introduire le concept de *milieu numérique*. Nous définirons d’abord brièvement ce concept avant de présenter un regard historique sur l’idée de milieu en science. Nous discuterons ensuite de l’acception particulière que nous avons choisie de défendre ici et nous verrons comment cette notion sera utile pour la suite de notre étude sur les réseaux sociaux en Chine.

L’idée de milieu numérique est a priori définie dans les termes suivants :

*“The multiple networks, which are connected by protocols and standards, constitute what I call a digital milieu.” (Hui, 2012)*

L’usage de multiples interfaces et le dédale des réseaux TIC constituent *“un nouveau milieu perceptif”* (Barboza et Weissberg, 2006). Notre milieu physique est aujourd’hui (déc)ouvert par l’existence de notre milieu digital qui nous aiguille. Rencontres, restaurants et voyages sont souvent d’abord médiatisés par l’Internet. Ainsi, nous évoluons dans un milieu numérique agissant comme support de processus de transduction et de connaissance du monde. Le code prend ici pleinement part à la construction de ce milieu digital, à la fois déterminant pour la production des actes de discours et ouvert à l’appropriation des pratiques et usages du quotidien.

Historiquement, le concept de milieu se détache du centre pour resituer et mettre en perspective la relation des êtres à leurs environnements sous des jours parfois contradictoires. Débattue puis écartée mais toujours très usitée, la notion de milieu introduit autant le déterminisme d’un combat pour la survie et l’adaptation, que la liberté créatrice du sujet dans un univers ouvert à sa volonté.

Pour illustrer au mieux la fécondité philosophique de la notion de milieu, nous allons tout d’abord essayer de comprendre la trajectoire de ce mot durant les siècles derniers

(Canguilhem, 1965). Sans remonter à son aube étymologique, nous nous apercevons que le mot milieu décrit un trajet singulier dans le monde des sciences. Employé dès le XVIème siècle par Descartes dans son *Traité de la lumière*, il représente pour Newton une mesure de distance dans l'éther, cette non-matière structurant la gravitation qui fait se mouvoir les objets. Défini plus tard par d'Alembert dans son encyclopédie comme un : "*espace naturel dans lequel un corps est placé, qu'il se meuve ou non*", le mot connaîtra durant tout le XIXème un large développement sémantique en s'étendant de la physique à la biologie. Les naturalistes français de l'époque affirment que le milieu n'est pas seulement *environnant* mais influe sur les êtres vivants. Ainsi Lamarck dès 1809 écrira dans sa *Philosophie Zoologique* : "*le milieu a une grande puissance pour modifier les organes*".

Le XIXème siècle est un moment marquant pour ce concept et structurant pour l'histoire des sciences dans son ensemble (Taylan, 2010). Auguste Comte, inspiré de la biologie, articule le vital au social dans la sociologie naissante et décrit le milieu comme "*l'ensemble des circonstances extérieures (...) nécessaires à l'existence de chaque organisme déterminé*" (Comte, 1838). Comte introduit une première dialectique des rapports avec le milieu comme conditions de possibilité de la vie : "*Tout être vivant (...) modifie sans cesse son milieu.*", écrit-t-il alors. Le concept connaît un succès croissant en France et les penseurs d'Outre-Rhin se le réapproprient en lui donnent un sens différent. S'opposant au francisé *Der Milieu*, le géographe Ratzel introduit dans son *Anthropogéographie* (1899) le terme *Umwelt*. Ce mot se démarque rapidement par sa dimension fortement déterministe. Dans un même mouvement, le physiologiste et biologiste Jakob von Uexküll étudie dans son laboratoire la tique et se rend compte que le milieu de la tique se définit non pas par tout l'environnement qui l'entoure mais seulement par ce qui lui est utile et approprié. Le milieu devenu *Umwelt* s'oppose alors à l'environnement indifférencié et devient l'ensemble des éléments porteurs de significations (*Merkmaltträger*) pour un être. Uexküll propose ainsi une "*biologie subjective*" étudiant les relations de chaque espèce avec son milieu. Dans l'Allemagne du siècle débutant, Uexküll diffuse largement sa théorie qui ne s'adresse pas tant aux animaux qu'à l'humain dont le milieu serai la Patrie (*Heimat*) (Feuerhahn, 2009). Reflétant les débats guerriers entre la *Kultur* allemande et la *Civilisation* française (Elias, 1975), l'idée de Milieu cristallise la tension politique sur les relations entre nature et vivant qui déchirera l'Europe pendant longtemps encore. Foucault dans son cours au Collège de France du 11 Janvier 1978 parle de l'influence de l'idée de milieu

sur la conception du territoire pour les urbanistes du XVIIIème siècle. Sous Louis XIV, les villes sont encore construites dans un espace conçu comme vide (voir *Richelieu* en Indre-et-Loire). A l'inverse, la question de l'urbaniste du XVIIIème est de comprendre la ville dans son évolution future. L'enjeu est devenu l'*adaptation* du milieu existant (à la fois urbain et naturel), la transformation du *donné* compris comme un élément qu'on peut venir modifier. L'idée sous-jacente de milieu ouvre la possibilité de l'appropriation de la nature. En terme foucaldien, cette nouvelle bio-politique se fonde sur la territorialisation du milieu comme nouveau centre des enjeux de pouvoir. L'ère industrielle réalise ce projet d'une adaptation à la fois *au* et *du* milieu vu comme tension nécessaire de l'évolution, passage obligé vers la civilisation. Alors que l'humain est placé dans son rôle central par l'astronomie galiléenne puis l'évolution darwinienne, l'idée de milieu pose comme enjeu majeur du vivant la maîtrise de l'environnement - la lutte pour ne pas être maîtrisé. Pour-suivi par la psychanalyse de Freud qui introduit l'Autre au sein du sujet, l'entreprise de décentrement de l'humain vers son milieu se joue dès l'abord dans les termes de la vie ou de la mort. Plus tard, Lacan identifiera la transition de la petite enfance à l'enfance par le "stade du miroir" comme moment où l'enfant différencie enfin le Milieu (*Umwelt*) du Soi (*InnenWelt*) (Lacan, 2001). Ce "passage au milieu" est donc d'une importance capitale puisque s'y joue la constitution de l'être dans la pensée de l'époque.

L'approche du milieu comme élément unificateur des sciences est un sujet toujours en discussion. L'introduction du concept d'*environnement* a largement recentré le milieu sur le rôle déterminant de l'Homme avec la théorie écologique (Gandolfo, 2008). Les conséquences de ce passage de l'idée de nature à celle de milieu restent profondes, notamment dans le droit civil où nous sommes passés d'un rapport du "droit imposé" de la nature au "droit négocié" du milieu (Papaux, 2008). Méta-réflexion, la discussion sur le "milieu académique" donne lieu à d'intéressants échanges (Stengers *et al.*, 2009) qui interrogent notamment la définition trop abrupte des disciplines scientifiques et leur herméticité. Parfois nommée *mésologie*, l'étude du milieu se donne pour mission de réconcilier des pratiques diverses de la biologie à la sociologie en imaginant une étude par le milieu (Stengers, 2003). Source d'inspiration de la mésologie, les œuvres de Deleuze & Guattari parlent déjà d'une philosophie du milieu : "*Partir au milieu, par le milieu, entrer, sortir, non pas commencer ni finir, [...] renverser l'ontologie, destituer le fondement, annuler fin et commencement.[...] C'est que le milieu n'est pas du tout une moyenne, c'est au*

*contraire l'endroit où les choses prennent de la vitesse.” (Deleuze et Guattari, 1972) La réflexion sur la technique et les technologies s’est notamment saisie à bras le corps de cette notion, avec notamment l’idée de *milieu technique* par Friedman et Leroi-Grouhan (Stiegler, 1998b). Tout geste (du plus banal au plus rare) s’effectuerait dans un milieu technique qui le rend possible. Gilbert Simondon dans son livre *Du mode d’existence des objets techniques* continue cette réflexion avec ce qu’il appelle le milieu associé :*

*“médiateur de la relation entre les éléments techniques fabriqués et les éléments naturels au sein desquels fonctionne l’être technique. (...) C’est ce milieu associé qui est la condition d’existence de l’objet technique inventé.” (Simondon, 1989).*

Simondon problématise le milieu associé comme vecteur de l’*individuation*, où se produit la rencontre entre objets et individus s’actualisant mutuellement. Poursuivant ce travail, Stiegler comprend les technologies de l’information comme un milieu essentiellement social, à la fois autour (environnement) et entre (medium) les individus (Stiegler, 1998a). Dans sa lecture critique des industries culturelles, Stiegler formule l’idée qu’un milieu est *associatif* s’il permet l’individuation. A l’inverse, certains milieux seraient *dissociatifs* car ils ne permettraient pas le devenir individu, à l’image des mass media qui divisent producteurs et consommateurs de symboles. La dynamique industrielle des deux derniers siècles a entraîné une massification des phénomènes culturels, créant un milieu technique considéré comme largement dissociatif car dé-réalisant pour les individualités (Simondon, 1989). Néanmoins, le renouveau technologique porté par l’apparition des technologies numériques ouvre aujourd’hui une page nouvelle pour l’individuation en offrant un milieu extrêmement associatif, fondé pour ainsi dire sur le lien. Voyant un nouvel âge des Lumières (Stiegler, 2012), Stiegler conçoit l’Internet comme un milieu qui ne serait pas structurellement dissociatif et pourrait donc recréer de nouvelles formes plus horizontales d’économie symbolique où existent davantage de symboles partagés.

### 1.3.3 Cyberspace et milieu numérique

Associateur ou dissociateur, les protocoles qui régissent l’accès au milieu sont donc les enjeux politiques du milieu numérique, conditionnant l’existence des objets numériques.

Alors que la notion de milieu disparaît peu à peu pour être remplacée par celle d'environnement (D'Angio, 2001), l'espace des géographes s'est vu aussi augmenté d'une nouvelle réalité à prendre en compte : le cyberspace. Espaces, sites, routes, les nombreuses métaphores géographiques de l'Internet remettent en question des pans entiers de la discipline. Le cyberspace, "*hallucination consensuelle*" décrite par Gibson (1984) déclare bientôt son indépendance (Barlow, 2001) et dessine ainsi une géographie virtuelle (Batty, 1997) qui s'interroge sur les dimensions de ces nouveaux espaces où circule l'information. Les structures spatiales et économiques préexistantes semblent être renforcées par les stratégies territoriales et d'équipements des acteurs. L'importance des Etats-Unis (Zook, 2001; Cukier, 1999) dans la localisation des flux Internet (capital, data centers, noms de domaines...) montre bien comment les évolutions technologiques participent à la fragmentation territoriale à l'échelle mondiale. Néanmoins, les sentiers de l'Internet s'écartent aussi bien souvent des autoroutes de l'information pour venir construire des sens beaucoup plus locaux par les nombreux mécanismes des activités en ligne. "*cyberspace is 'made real' through the language of place*", comme l'écrivent justement Dodge et Kitchin (2007).

Les modèles actuels de l'Internet tendent à structurer les services de réseaux sociaux de façon bien particulière. Dans un Internet géant, le modèle économique des services web est fondé sur la captation et la rétention de l'attention de l'utilisateur et/ou des données qu'il publie. La fonction primordiale du service web est donc l'inclusion c.a.d. la discrimination entre utilisateurs et non-utilisateurs du site. Commercialement, le but actuel des compagnies web est d'acquérir le plus grand nombre d'utilisateurs afin de pouvoir valoriser l'attention auprès des annonceurs puis sur les marchés d'affaires (Ries, 2011). L'usage des réseaux sociaux suppose donc non seulement l'exclusion des non-utilisateurs, mais aussi la conservation des utilisateurs actifs. Ces procédés d'inclusion/exclusion et de rétention, enjeux de la survie économique d'une compagnie web, deviennent alors les fondamentaux du design et du développement de chaque interaction possible sur ce type de plate-forme<sup>22</sup>. Traduit en code, ces impératifs de rentabilité dans l'économie de l'attention structurent le milieu numérique lui-même.

Dans son livre *Rewire : Digital Cosmopolitans in the Age of Connection*, Zuckerman

---

22. Dans la Lettres aux Investisseurs écrite par M. Zuckerberg pour l'IPO de Facebook [http://www.sec.gov/Archives/edgar/data/1326801/000119312512034517/d287954ds1.htm#toc287954\\_10](http://www.sec.gov/Archives/edgar/data/1326801/000119312512034517/d287954ds1.htm#toc287954_10), consulté le 13 Août 2013 à 12 :22

étudie le design des interfaces et algorithmes régissant les relations dans les services de réseaux sociaux les plus utilisés. Le “design social”, produit de l’économie des plate-formes numériques, se fonde sur la segmentation du marché de l’attention, avec notamment les groupes et pages officielles : *“Our challenge is not access to information, it is the challenge of paying attention.”* Zuckerman (2013). Il ne s’agit pas seulement de pouvoir comprendre un message mais également de réussir à prêter un intérêt et une attention suffisante dans une économie de l’attention en ligne ultra-concurrentielle. Rendant hommage au travail de l’urbaniste Jane Jacobs et sa lutte contre les politiques de zonage excessif du plan urbain (Jacobs, 1961), Zuckerman s’attache à comprendre comment s’organisent nos *digital surroundings*. Pour l’utilisateur, l’expérience offerte par les plate-formes en ligne se fonde donc sur une “tribalisation” par petits groupes, nous contraignant à des usages restreints de l’espace d’expression possible. Comme observé par Kumar *et al.* (2006), les services de réseaux sociaux évoluent aujourd’hui vers une structure relationnelle en small-worlds composé de petits groupes très distants. Loin des discours annonçant la fin des frontières avec le village global (Breton, 1997), il semblerait que la “culture web” et plus largement l’usage des technologies de l’Internet soient des facteurs supplémentaires de fragmentation des relations sociales.

### 1.3.4 Topogrammes : association et dissociation dans le milieu numérique

L’étude des relations entre espace et dispositifs socio-techniques ne se comprend donc pas seulement en termes d’infrastructures, mais plus finement dans l’observation et la description d’une géographie du réseau. Les débuts de la géographie au XIXème siècle ont défini cette discipline comme la science des milieux. Vidal de la Blanche cherchait alors à expliquer comment les actions humaines étaient déterminées par des faits “naturels” pré-existants. La sociologie en faisant école a amené les géographes à considérer les œuvres humaines comme partie intégrante du ou des milieux qui les produisent (Demangeot, 1984) en étudiant : *“les relations verticales qui se développent au sein de chaque milieu, et celle des relations horizontales qui mettent en relation les milieux”* (Claval, 1990). Ces réflexions géographiques sont nourries par de vastes controverses sur de nouveaux paradigmes : l’espace, le territoire, le paysage, les lieux qui estompent peu à peu l’idée

déterministe de milieu pour développer un appareil conceptuel plus complexe.

Le développement méthodologique avec notamment la géomatique et les outils d'analyses issus de la statistique permettent de proposer des lectures variées de faits géographiques divers. Brunet définit les *chorèmes* comme des “*structures élémentaires d'organisation de l'espace*” (Brunet, 1980), Berque parle de géogrammes définis comme “*motif éco-techno-symbolique (...) au sein de la relation qu'est l'écoumène*” (Berque, 1999). Inspiré de la philosophie japonaise moderne, l'écoumène de Berque se rapproche de l'idée de milieu et est décrit comme une vaste matrice relationnelle des choses et des êtres - une dimension écologique générale, une “*trajectivité*” (Watsuji et al., 2011). Au sein de cet écoumène, le géogramme offre un modèle des faits géographiques où se rejoignent les aspects techniques, symboliques et sociaux (les relations humaines).

S'opposant aux objets naturels et techniques, les objets numériques sont à comprendre dans leurs relations matérielles et temporelles avec les infrastructures de leur production et archivage dans les mémoires des données du web. Ainsi, le milieu numérique dans lequel chacun évolue se présente sous la forme d'objets numériques actualisés. À l'instar des géogrammes du paysage de Berque ou des chorèmes de l'espace de Brunet, nous pouvons imaginer ici des *topogrammes* permettant de considérer les faits et objets digitaux. Le milieu numérique se constitue à la fois du cyber-espace (le lieu physique où se situent les machines) et des objets digitaux qui lui sont associés. Le topogramme en tant que modèle permet de décrire et considérer sous un jour commun des objets digitaux dissemblables.

Alors que les médias traditionnels ont cherché à définir les territoires du discours, l'enjeu stratégique des médias du web se situe aujourd'hui dans cette définition de fragments d'espaces pour l'énonciation. Marketing, communication politique, journalisme ou activisme social, la fonction du média dans une économie de l'attention devenue hyper-compétitive (Weng et al., 2012) n'est plus communicative (“dire”) mais performative : “faire dire” ou plutôt “faire faire” (cliquer, liker, acheter...). Bâtir l'image d'une marque, d'une entreprise, d'une personne ou d'un fait public nécessite la construction de réseaux sémantiques, conversationnels (sociaux) et territoriaux qui définissent les fondations d'un espace de “participation” où peut se dérouler l'individuation passant nécessairement par une énonciation (Butler, 1993). Ainsi, l'enjeu du média devient le contrôle de cet espace par une gestion stratégique d'un réseau de symboles, de personnes et de lieux, comme

autant de vecteurs de l'énonciation, mémoire partagée en devenir.

Les topogrammes particuliers procèdent donc de constructions existantes sous la forme de relations entre objets des réseaux. Il est possible de caractériser au moins deux types de modèles généraux de topogrammes que nous nommerons associatif et dissociatif (Stiegler, 2008). Reprenant l'idée de milieu associé aux technologies (Simondon, 1989), l'espace de la conversation est dit associatif si il offre une possibilité d'individuation lors de l'énonciation. La conversation suivant un topogramme associatif associe à la conversation, engage et amène à participer. A l'inverse, un topogramme peut être décrit comme dissociatif si il ne propose pas d'être associé à la discussion et offre seulement une énonciation sans transduction, c'est à dire une répétition sans changement.

L'exploration des objets numériques sous la forme de topogrammes apporte un regard sur leur nature dissociative ou associative et nous éclaire sur les modalités de transduction proposées par différents milieux numériques. Cette approche permet également de comprendre les éléments centraux responsables de la transition et des différences entre ces deux modes de structuration. En effet, il ne s'agit pas de caractériser définitivement un milieu mais plutôt d'en considérer l'évolution et les modalités. La médiation, la transition, la transformation ou plus simplement l'interprétation sont autant de phénomènes essentielles qui permettent et autorisent l'accès. Les formes de contenus ou d'énonciation jouent également ce rôle synaptique d'association des espaces dans la structuration du milieu numérique. L'identification précise de caractéristiques propres des différents topogrammes permet de décrire une typologie plus détaillée des lieux du Web.

Afin d'observer comment la notion de topogramme peut participer à décrire le milieu numérique en Chine, nous avons choisi d'étudier empiriquement les dynamiques à l'œuvre autour d'objets numériques particuliers : les mêmes Internet.







# 2

## Les mèmes Internet objets numériques culturels

CONSTRUCTIONS COLLECTIVES ÉPHÉMÈRES, les *mèmes Internet* sont de courts messages comiques se propageant rapidement sur la Toile. Plus que de simples blagues de potache, ils laissent souvent des traces symboliques structurant le milieu numérique et proposent une illustration pertinente des discours multiples prenant place lors des échanges en ligne.

Avant de concerner l'Internet, la notion de *mème* proposée par Dawkins (1989) définit une unité minimale de propagation des cultures. Controversé, ce concept flou teinté d'un évolutionnisme peu convaincant a flotté depuis sa création en marge de la littérature scientifique. Depuis une dizaine d'années, la reprise du terme dans le contexte d'Internet a fait du *mème* un concept populaire. La suite de ce travail resituera l'idée de *mème* dans

la perspective historique des questionnements sur la formation de mémoires collectives.

Notamment, nous interrogerons les pratiques *d'énonciation* et leur performativité pour comprendre comment une blague peut passer de la base de données à la table du déjeuner. Les activités symboliques sous-jacentes aux mèmes seront l'occasion de s'intéresser à leurs formes comme *topoi* ou *lieux communs*. Des exemples nous amèneront à en dresser une première catégorisation.

## 2.1 Les mèmes : définitions et histoire

Le dictionnaire d'Oxford<sup>1</sup> donne deux définitions du mot *mème* :

*Meme* (n.)

1. Un élément d'une culture ou système de comportement passé d'un individu à un autre par imitation ou par d'autres moyens non-génétiques.
2. Une image, vidéo, morceau de texte, etc., la plupart du temps de nature humoristique, qui est copié(e) et propagé(e) rapidement par les utilisateurs d'Internet, souvent après avoir été modifié(e).

Cette définition nous renseigne sur l'usage de ce terme, en le définissant à la fois comme un élément culturel transmissible et une forme particulière de contenu diffusé sur Internet. En nous appuyant sur son évolution dans la littérature, nous allons voir comment les deux versants de ce concept se sont historiquement articulés et à plus forte raison comment cette articulation peut nous servir pour comprendre les phénomènes à l'œuvre sur les réseaux sociaux en Chine.

### 2.1.1 La mémétique : une éthologie culturelle teinté d'évolutionnisme

Le premier usage du mot *mème* est souvent attribué au biologiste Richard Dawkins dans son livre *Le Gène égoïste* (1976).

---

1. Traduction de l'auteur d'après *British & Words English* publié par Oxford University Press en 2014, <http://www.oxforddictionaries.com/definition/english/meme>, consulté le 24 Février 2014 à 21 :50

Dawkins s'inspire des théories évolutionnistes de l'éthologie moderne et propose le concept de *mème* comme un élément moléculaire permettant la transmission de patrimoine culturel, à l'instar du gène des individus biologiques. Considéré comme une « *unité d'information culturelle qui peut être copiée, située dans le cerveau* » (Blackmore, 2001), le mème serait la fondation de pratiques culturelles évoluant selon une sélection naturelle. Représenté comme une « *unité distincte de la pensée* » (Dawkins, 1989), le concept se fonde sur l'analogie entre processus de transmission culturels et génétiques : “*Cultural transmission is analogous to genetic transmission in that, although basically conservative, it can give rise to a form of evolution*” (p.72). Cette approche éthologique de la culture considère donc le *mème* à la fois comme un élément transmissible et un facteur de transmission, doté de la capacité de se reproduire lui-même. Élément actif, le mème serait donc un « *gène égoïste* » agissant de manière isolée et distincte, spécificité d'une “culture”. Plus encore, son but unique serait sa propre pérennisation par sa propagation de cerveau en cerveau (Blackmore, 1997). Le mème agirait donc comme un agent culturel possédant une forme de volonté propre pour se propager. Souvent représenté grâce à l'image du virus, la propagation de la culture sous forme de contamination s'effectue en avant-plan d'une lutte pour la survie et la fécondité des idées. Le mème est un “réplicateur culturel”, une extension à part entière du vivant au-delà du biologique (Bloom, 2000).

Le québécois Fernand Dumont définit la culture comme cette « *maison où l'on habite ensemble* » (Dumont, 1993). Nœud dans une topologie sociale, le mème pourrait donc également se présenter comme un point d'entrée, une porte entrouverte vers ce lieu où d'autres sont déjà passés et se trouvent encore. Le mème devenu particule culturelle définit un seuil, ce lieu de passage si particulier qui « *fonde les espaces* » (Bonnin, 2000) et invite ou interdit d'entrer. Comme on enlève ces chaussures au dojo et qu'on sonne à la porte, les mèmes sont peut-être à envisager comme des rites de franchissement de seuils culturels, pratiques de liaison du vivre-ensemble politique d' Arendt (1995) . Pour le gène comme pour le mème, il ne s'agit pas d'en considérer la fonction mécanique de “réplicateur”, mais d'observer l'altération se déroulant lors de son actualisation pour en comprendre les limites et le rôle. La présence *in potentia* d'une unité culturelle identique ne constitue pas nécessairement une réalité in-formante pour des groupes sociaux ou des individus (Lissack, 2004). Les sciences de la communication ont largement étudié depuis 50 ans les modalités de transmission des informations. Les études sur la réception notamment ont bien montré

qu'il ne suffisait pas qu'un message soit émis pour être décodé et compris (Liebes et Katz, 1990). Déjà avec Shannon et Weaver, l'environnement exprimé par le concept de *bruit* vient altérer largement les phénomènes de transmission tout au long de leur diffusion (Attali, 1978). La mémétique, faute d'étude de cas conséquentes et d'applications théoriques réelles (Jouxte, 2014), a subi de nombreux revers conceptuels en s'appuyant notamment sur l'image peu crédible d'une transmission par réplication. Ignorant la dimension poétique des actes de transmission, cette vision mécaniste issue d'une rationalisation excessive reflète pourtant les écueils non-dits des approches scientifiques modernes. Thierry Bardin dans son livre *Junkware* (2011) effectue une recherche extensive sur les discussions et considérations entourant la partie non-codante de l'ADN appelée "*junk ADN*". Analysant les discussions dans les publications scientifiques, il montre comment plus de 80% des éléments structurant l'ADN ont été très tôt étiquetés comme "bruit" puis "junk", car il était impossible d'identifier leur participation active au codage de protéines. La métaphore de ce "*junk non-codant*" si envahissant et la relative facilité avec laquelle nous nous permettons de l'ignorer démontre la nécessité d'une approche renouvelée des phénomènes complexes du vivant, et notamment de ceux de la transmission culturelle. Déjà clairement identifiées dans les études en communication, les fonctions non-langagières sont indispensables au bon déroulement d'un acte de langage. Ainsi si la suppression du bruit est souvent un préalable méthodologique pour l'étude scientifique, elle peut souvent fausser l'approche expérimentale et les conclusions théoriques en refusant d'admettre sa partialité. L'étude des mêmes est encore largement en quête de reconnaissance scientifique et si la construction théorique permettant d'isoler des éléments culturels pour l'étude paraît intéressante, elle manque d'une réelle prise sur l'observation et l'analyse par l'étude de cas notamment. La fermeture dès 2005 du *Journal of Memetics*, parution de référence de la discipline naissante est annoncé dès 2002 par un article de B. Edmonds (Jouxte, 2014). Intitulé *Three Challenges for the Survival of Memetics*, l'article exhorte les chercheurs intéressés à produire ce que Edmonds juge comme le minimum indispensable pour gagner la reconnaissance des milieux scientifiques : "*a conclusive case-study; a theory for when memetic models are appropriate; and a simulation of the emergence of a memetic process.*" (Edmonds, 2002).

En effet, il paraît impossible d'asseoir scientifiquement la légitimité du concept en se fondant uniquement sur une analogie de phénomènes disparates. Le mème a donc raté sa

cible dans le domaine scientifique. Néanmoins, son acception plus récente sous la forme de contenus Internet a donné au concept une nouvelle vie dans la culture populaire. Dans le même temps, ce sens renouvelé a permis de définir avec l'Internet un domaine d'application idéal.

### 2.1.2 Mèmes Internet : définition, littérature et exemples

Contrairement au concept éthérique de mème présenté dans la partie précédente, la définition des *mèmes Internet* est de prime abord plus pragmatique. Il s'agit de courts messages comiques faits de texte, image, vidéo ou de son gagnant rapidement une forte popularité sur Internet en étant partagés, commentés, réappropriés puis transformés lors de leur diffusion. L'utilisation du terme *mème Internet* pour décrire la diffusion de messages ne recouvre pas nécessairement la dimension évolutive et culturelle du concept initial de Dawkins, mais garde l'idée générale d'une circulation « virale » d'idées parmi des groupes d'individus<sup>2</sup>. Le concept de mème a très fortement gagné en popularité avec cette nouvelle acception. En 2012, il a notamment été sélectionné parmi les 10 mots les plus marquants de l'année par le prestigieux dictionnaire américain Merriam-Webster. Ce choix a été motivé par la très forte popularité sur Internet des images parodiques du politicien Mitt Romney après une bourde commise lors d'une intervention télévisée aux Etats-Unis<sup>3</sup>. Ainsi, le mot *mème* dans une acception que nous prendrons ici soin de nommer *mème Internet* est aujourd'hui entré dans le vocabulaire commun du Web. De nombreux sites spécialisés (knowyourmeme.org, quickmeme.org, memefest.org, etc) ont vu le jour avec comme mission d'archiver et de collecter ces pièces de la culture web. Un des plus anciens mèmes Internet est certainement l'usage des émoticônes ou *smileys*. Ces petites figures servent à exprimer des émotions dans le contexte d'oralité écrite d'Internet. Apparus dans les premiers jours du réseau Internet, les émoticônes répondent à un besoin d'expression non-verbale dans la communication en ligne. Très simples à utiliser ou à modifier, les *smileys* connaissent une popularité rapide et se diversifient partout autour de la toile.

---

2. "The meaning is not that far away from the original. It's anything that goes viral.", Dawkins interviewé par le magazine *Wired* <http://www.wired.co.uk/news/archive/2013-06/20/richard-dawkins-memes>, consulté le 12/08/2013 à 7h53 GMT+8

3. "Words of the year 2012", Merriam-Webster <http://www.merriam-webster.com/info/2012words.htm> consulté le 25 Février 2014 à 19 :01 GMT+1

19-Sep-82 11 :44 Scott E Fahlman :-)  
From : Scott E Fahlman <Fahlman at Cmu-20c>

I propose that the following character sequence for joke markers :

:-)

Read it sideways. Actually, it is probably more economical to mark things that are NOT jokes, given current trends. For this, use

:-(

**Graphique 2.1.1** – 19 Septembre 1982 : la première mention du smiley par Scott Fahlman, que l'on retrouve quelques jours plus tard sur les mailing lists les plus utilisées de l'époque : Arpanet et Usenet, d'après <http://www.cs.cmu.edu/~sef/Orig-Smiley.htm>, archive consultée le 10 Août 2013 à 09 :15 GMT+8

L'usage des émoticônes s'est aujourd'hui largement répandu, notamment chez les adolescents et jeunes adultes (Derks *et al.*, 2007). Une récente étude a même montré que les zones du cerveau stimulées par la vue d'émoticônes étaient similaires à celles stimulées lors de la vue d'un visage, indiquant un ancrage symbolique profond de l'usage de ces signes (Churches *et al.*, 2014). Des émoticônes singuliers se sont également développés dans différentes langues pour exprimer des sentiments particuliers, propres au langage et à ses modes d'expression. Le caractère idéographique des langues chinoise et japonaise se prête particulièrement à ces jeux de dessin langagiers. En chinois, le plus célèbre exemple est sans doute le caractère 冏 (*jiong*) représentant en langue ancienne une fenêtre d'où provient la lumière et signifie « lumineux ». Sa ressemblance avec une figure humaine (un émoticône) a fait renaître ce caractère désuet qui signifie désormais qu'un utilisateur est agacé, embarrassé ou même choqué. La conjonction "冏rz" a même été inventée : 冏 représentant la tête et rz le corps agenouillé d'une personne ; elle signifie l'échec et le désespoir. Cette particularité des langues asiatiques donnent à l'émoticône un rôle central dans la communication en ligne, matérialisée par les interfaces. Les réseaux sociaux chinois proposent tous par défaut de multiples jeux d'émoticônes disponibles pour l'utilisateur qui communiquent ainsi très rapidement en images. L'exemple de l'émoticône illustre la manière dont un élément visuel et langagier vient constituer les pratiques en ligne, en relation proche avec la culture et le lieu qui l'a fait naître.

Il est difficile de définir un même a priori par la nature de son contenu. Néanmoins, la

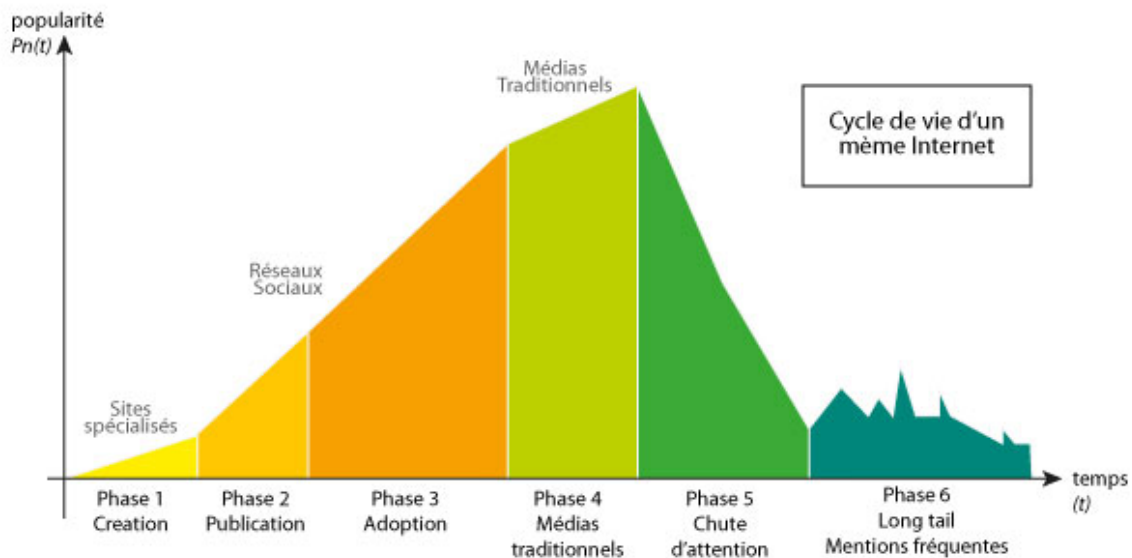




**Graphique 2.1.2** – Un jeu d’émoticônes sur weibo.com – Consulté le 10 août 2013 à 10 :00 GMT+8

structure de la diffusion d’un message peut nous permettre de le décrire comme un mème. Reproduisant en grande partie le cycle de vie classique d’une information comme une rumeur ou une *news*, les mèmes possèdent des modes de diffusion en ligne assez déterminés et prévisibles. La plupart du temps, ils sont mis en circulation sur un petit nombre de sites spécialisés, avant d’être repris dans une première phase par un assez petit nombre d’utilisateurs qui se charge de les publier sur les réseaux sociaux (Bauckhage, 2011). Les réseaux sociaux agissent alors comme une chambre d’écho et déterminent si le “proto-mème” encore en devenir deviendra mème ou restera simple message isolé. Durant cette phase souvent nommée *adoption*, le mème entre en concurrence avec d’autres informations sur les réseaux sociaux où les utilisateurs sont sans cesse sollicités par d’autres informations (Davenport et Beck, 2001). Si l’attention générée par le mème auprès des utilisateurs atteint un pic suffisamment important, il faut environ 2h30 pour que les mèmes rejoignent les pages des médias plus traditionnels en commençant par les blogs, puis les sites d’information (Leskovec et al., 2009). Ensuite, l’attention envers le mème décroît fortement et rapidement. La présence épisodique de citations maintient l’existence du mème apparente dans des groupes définis (Buchel, 2012).

L’évolution du volume de la diffusion permet donc de définir un mème Internet. Néanmoins, il est impossible de donner une estimation du volume minimum pour devenir



**Graphique 2.1.3 – Cycle de vie d'un mème Internet -Clément Renaud - 2013**

“mème” tant ce chiffre dépend de la population étudiée : il existe des mèmes à très forte diffusion comme le smiley ; d'autres mèmes se diffusent seulement au sein de groupes d'individus restreints. Certains mèmes peuvent avoir connu une diffusion très importante dans un groupe, mais rester absolument inconnus du reste de l'Internet. Une étude de 2012 comparant la diffusion de nombreux mèmes sur Twitter montre que les utilisateurs tendent à choisir les mèmes selon la structure de leur réseau social et le moment d'exposition, produisant ainsi une grande hétérogénéité des mèmes dans le réseau (Weng *et al.*, 2012). Aussi est-il hasardeux d'essayer de décrire le concept de mème par son contenu tant les sujets et les discussions varient.

Quelques éléments d'ordre grammaticaux peuvent néanmoins être observés dans la forme que prennent les contenus, appelé parfois “véhicule” du mème. Les mèmes diffusés les plus largement sont composés d'images et de vidéo. L'économie d'attention très limitée de l'Internet et les modes de lecture sur écran dans un contexte d'abondance d'informations font que l'on privilégie souvent les médias visuels sur le texte (Goldhaber, 2006). Un autre élément important est la facilité avec laquelle un message peut être approprié par un utilisateur qui veut le modifier ou tout simplement le diffuser dans le réseau. L'existence des mèmes est en effet largement conditionnée par la possibilité d'une diffusion à moindre coût et effort pour l'utilisateur final, la plupart du temps non-rémunéré. Ici on voit émerger une structure visuelle caractéristique du mème : une image accompagnée

d'une légende écrite en caractères blancs détournés de noir, ou de caractères blancs sur fond noir. L'utilisation de haut contraste de couleur dans les typographies permet de faire apparaître très efficacement des légendes juxtaposées à l'image.



**Graphique 2.1.4** – Exemples de mèmes Internet, d'après <http://knowyourmeme.org>, consulté le 12/08/2013 à 10 :01 GMT+8

L'usage d'images légendées est une des formes les plus communes pour les mèmes Internet, en particulier ceux de nature comique ou absurde. La mise en place de sites générant rapidement ce type d'images légendées (memegenerator.com, mememachine.com, etc.) renforce l'unité formelle des mèmes sur l'Internet, ou plutôt sur l'Internet anglophone et francophone. En effet, on constate que cette forme typique du mème ne se retrouve pas sur l'Internet chinois lequel utilise plus volontiers des montages d'images ou des jeux de mots, avec davantage de diversité dans les formes.

## 2.2 Mème, mémoire collective et culture

### 2.2.1 La mémoire comme trace

Si le concept de mème est souvent considéré comme très récent, on peut néanmoins le resituer dans le vaste paysage des travaux sur la mémoire collective qui ont existé depuis le XIX<sup>ème</sup> siècle (Laurent, 1999). Max Stirner dans son livre *The Ego and Its Own* (1995) énonce déjà l'idée que les individus sont sujets à la circulation de concepts issus

de souvenirs communs ou illusoires, comme notamment le nationalisme et la religion. Le logicien Bertrand Russell reprend par la suite dans son livre *The Analysis of Mind* (1921) les travaux sur la mémoire et l'évolution sociale des physiologistes allemands Semon et Lee (1923). Utilisant le concept central de *mneme* (du grec *mneme*, mémoire), Semon travaille sur l'idée de “*traces mnésiques*” laissées par les diverses expériences au niveau cellulaire comme au niveau de l'organisme tout entier. La psychanalyse a également cherché à saisir cette distance impalpable entre expérience et organisme en interrogeant les marques laissées par les souvenirs. Pour Freud comme pour Semon, “l'appareil psychique” de la mémoire se constitue sous la forme de “traces”, qu'ils se refusent néanmoins à localiser dans des zones spécifiques du cerveau. Lacan après Freud suggèrera que la perception et la mémoire des expériences se structurent dans le langage lui-même, seul outil de connaissance du monde. Depuis les dix dernières années, plusieurs découvertes dans le domaine de la neurologie viennent corroborer cette idée que la mémoire existe sous forme de traces. Les travaux autour de la *plasticité neuronale* montrent notamment l'existence de la mémoire sous la forme de connections, relations ténues ancrées dans notre réseau neuronal global (Magistretti, 2008).

	t=1	t=2	t=3
Freud	expérience	perception	Traces mnémiques et psychiques
Lacan	expérience (signifié)	perception (signifié)	Signifiant (traces structurées dans le langage)
Neurosciences	expérience	Perception	Traces synaptiques, assemblages de neurones

**Tableau 2.2.1** – Etapes constitutives de la mémoire - Convergence entre la trace psychique et synaptique (Ansermet et Magistretti, 2004)

Ces disciplines s'intéressent majoritairement à l'étude de la constitution d'une mémoire individuelle, ne nous livrant que peu de clés pour comprendre les éléments qui fondent une mémoire commune. Le paléontologiste Leroi Gourhan propose dans son livre *L'homme et la Matière* (1971) de considérer que les humains possèdent trois formes de mémoire : une mémoire individuelle sensible, stockée dans les organes du corps ; une mémoire héritée génétiquement, stockée dans l'ADN ; et une troisième forme de mémoire, transmise de générations en générations : la *technologie*. Dans sa lecture de Leroi-Gourhan, Stiegler

(1998b) explique comment *l'objet technologique* porte en lui les traces des expérimentations, réussites et échecs, mémoire cumulative de temps et de sociétés passées, à la fois héritée et commune, transmissible par son usage.

Forme de mémoire	Contenu	Stockage
génétique	Particularités héritées des ancêtres	DNA
épigénétique	Mémoire sensible de l'expérience personnelle	Organes, nerfs, cerveau
technologique	Pratiques de la vie quotidienne en société (usages)	Objets technologiques

**Tableau 2.2.2** – Les trois formes de mémoire d'après Leroi-Gourhan et Stiegler

## 2.2.2 Diffusion de mèmes et structuration d'une mémoire collective

La relation entre mémoire humaine et mémoire technologique est au centre de notre étude. La technologie a depuis bien longtemps été considérée comme une mémoire extérieure. Dans *Phèdre*, Platon raconte l'histoire du roi égyptien Thamous recevant en cadeau du dieu Thot l'écriture, le remède (*pharmakon*) qui devait “soulager la science et la mémoire” (Platon, 274e). Le roi Thamous, effrayé par cette nouvelle technologie qu'est l'écriture se voit saisi de l'angoisse d'une perte de mémoire. La fin de l'oralité va accompagner la disparition de la méthode active des antiques théâtres de la mémoire au profit d'un support inerte et extérieure. Pourrait-elle alors sceller l'avènement d'une nouvelle bêtise ? Aujourd'hui, l'importance grandissante des bases de données et de connaissances soulève encore une fois les mêmes questions, toujours irrésolues. Nicolas Carr constate notamment que “*l'Internet nous rend stupide*” et que son usage répété entraîne une baisse drastique de nos facultés de concentration (Carr, 2010). A l'ère du Big Data et de l'expansion sans fin de notre mémoire numérique, la constitution de nos bases de données interroge notre construction d'une mémoire collective. Les mèmes Internet, d'abord gravés dans les disques durs des serveurs, viennent être actualisés par ceux qui les partagent, les commentent, jouent avec et se les approprient. L'exemple de l'Internet chinois nous

montre la volatilité de cette mémoire numérique, artefact historiographique d'une culture soumise au bon-vouloir des administrateurs du réseau. Le *Manifeste de l'Archiviste* publié par Hui (2014) s'ouvre sur l'angoissante interrogation deleuzienne :

*“Un nouvel archiviste est nommé dans la ville. Mais est-il à proprement parler nommé ? N'est-ce pas sur ses propres instructions qu'il agit ?”* (Deleuze, 1972).

L'existence et l'usage quotidien des bases de données questionnent chaque jour l'assujettissement des symboles de notre mémoire à l'objet technologique, à la fois béquille, prothèse et maquillage postiche de notre détestable devenir bête. Les petits *mêmes Internet*, habitant des profondeurs glacées des *data centers*, nous parviennent en dansant, d'abord sur nos écrans puis dans un coin de notre tête. Avec l'usage répété des technologies de l'écriture numérique, les frontières entre milieu numérique et mémoire collective s'estompent pour laisser entrevoir un enchevêtrement de silicium, d'idées et de chair, constitutifs d'un savoir moderne bien vivant.

Poursuivant l'idée d'une archéologie du présent introduite par Foucault, il s'agit donc de documenter les processus par lesquels ces obscurs habitants des bases de données viennent laisser leurs traces pour constituer des bribes de nos mémoires collectives. Maurice Halbwachs dans son vaste travail aborde les façons dont l'histoire structure l'être-ensemble des groupes humains. En disant que *"l'histoire de notre vie fait partie de l'histoire en général"* (Halbwachs, 1947), il décrit une mémoire autobiographique (personnelle) et une mémoire historique (sociale). Les inquiétudes et considérations autour de la “vie privée” sur Internet illustrent les liens intimes entre ces deux mémoires aux frontières chaque jour plus poreuses. L'acte singulier et autobiographique devient sous l'effet du réseau un fait social, disponible à tout moment dans “l'historique” qui se déroule sous le curseur. L'oubli devient alors un commerce très prisé permettant de garantir la limite entre la mémoire autobiographique de la fin de soirée de samedi dernier et la mémoire socialement acceptable du CV du chercheur d'emploi. À l'inverse, les photos du dernier voyage en Papouasie ou la pose avec une star de la télé témoignent fièrement d'un lien mémoriel entre autobiographie et histoire commune. Les *mêmes* se propagent également d'individus en groupes pour former peu à peu des éléments de mémoire commune. Objets, chansons, histoires, légendes, icônes... se diffusent autour de la toile par des procédés



tenant autant de la copie que de l'appropriation.

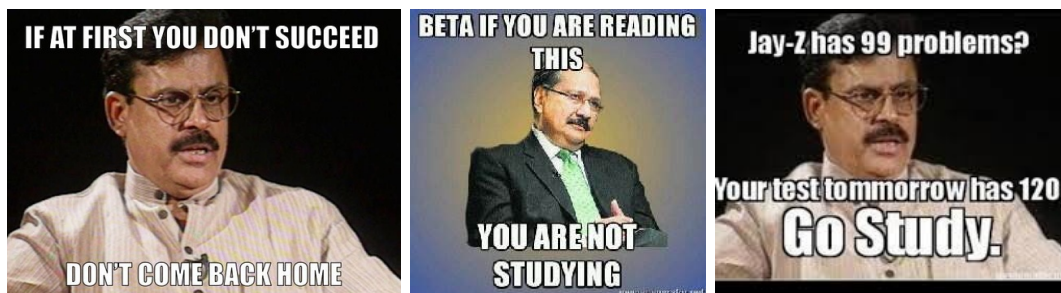
Le site de questions/réponses *Quora.com* offre un regard intéressant sur la question : « *What are some quintessential Indian memes ?* ». Les utilisateurs répondent donc en ajoutant des exemples de mème qui semblent appartenir dans leurs esprits à la “quintessence des mèmes indiens”. L'image la plus citée représente la figure d'un père autoritaire (2.2.2a). Avec plusieurs centaines d'images postées par des utilisateurs majoritairement indiens, nous pouvons constater plusieurs choses :

- **Langue** : à part trois réponses, la totalité des réponses sont en anglais. Cela peut s'expliquer par le fait que l'anglais est une langue de communication majoritaire en Inde, et également par le fait que le site *Quora.com* n'accepte habituellement que des réponses en anglais.
- **Forme** : à l'exception de quatre réponses, les mèmes revêtent tous la forme “classique” : photo retouchée et légendée par un texte en anglais aux lettres blanches sur fond ou détour noir.
- **Humour** : la plupart des réponses sont de nature comique.
- **Récurrence** : certaines images sont très récurrentes et si les légendes diffèrent, le sens reste le même.
- **Thématiques diverses** : de nombreux messages traitent de la vie de famille (parents, mariage), de loisirs (cricket), de la vie quotidienne (achats, école , etc.) et un peu de politique.

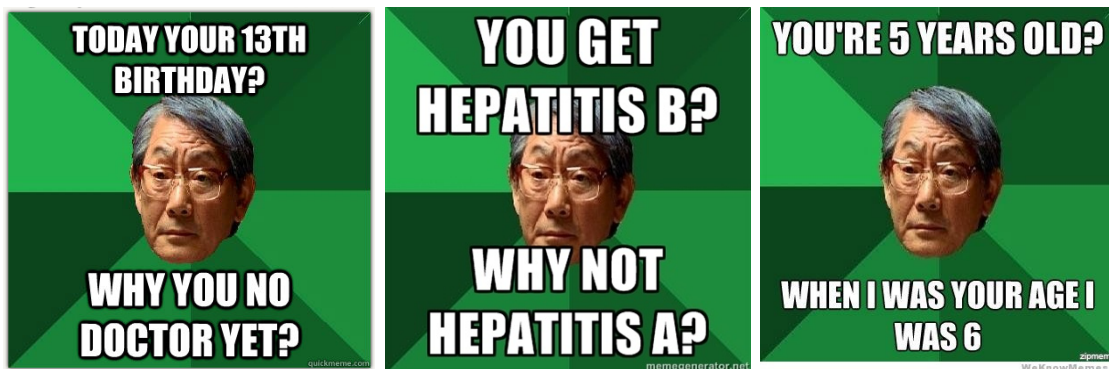


**Graphique 2.2.1** – Exemples de réponses à la question “*What are some quintessential Indian memes ?*” D’après <http://www.quora.com/India/What-are-some-quintessential-Indian-memes>, Consulté le 12/08/2013 à 0 :41 GMT+8

La définition des mèmes faite par Blackmore ne correspond pas nécessairement à ce que nous observons ici. L’unité formelle de l’ensemble de ces mèmes (images avec des caractères blancs cerclés de noirs) propose une définition bien plus restreinte. Néanmoins,



(a) "High Expectations Indian Father"



(b) "High Expectations Asian Father"

**Graphique 2.2.2** – Exemples du mème "High Expectations Asian Father" D'après "What are the funniest High Expectations Asian Father meme images?" sur Quora.com <http://www.quora.com/Memes/What-are-the-funniest-High-Expectations-Asian-Father-meme-images>, Consulté le 12/08/2013 à 10 :56



nous pouvons comprendre qu’il s’agit bien de la manifestation d’une culture particulière. L’image d’une vache assise devant un sigle “*Fuck The Police*” serait en effet un absolu non-sens sans la référence à l’Inde où les vaches sont sacrées et jouissent de droits particuliers que même la police ne peut entraver. Ainsi, il existe bel et bien des pré-requis pour comprendre ou actualiser un mème. Les plus évidents sont :

- **L’accès et l’usage de la bonne technologie** : Il est nécessaire pour un utilisateur de posséder et de savoir utiliser la technologie par laquelle le mème est diffusé.
- **La langue** : Le message possède une légende donc l’utilisateur doit pouvoir la lire et la comprendre.
- **Les implicites** : L’utilisateur doit posséder le socle de références communes et d’implicites qui sont impératifs pour saisir le sens du message.

Nous observons ici de très larges et vagues groupes (« *indiens* », « *asiatiques* »...) sans pouvoir vraiment comprendre dans le détail ce qui peut réellement constituer des éléments communs. Les travaux sur la formation des “communautés” en ligne ont montré comment la circulation des objets digitaux peut posséder une fonction de catharsis pour des groupes plus réduits (Steyer *et al.*, 2006). Néanmoins, on voit bien qu’ici l’appartenance pré-existe puisque le mème nécessite de nombreux pré-requis pour le comprendre. Le langage et son expressivité par l’humour sont notamment des contraintes incompréhensibles pour l’actualisation de ce mème par un individu. Le mème participe à l’affirmation de l’existence du groupe, avec la figure redondante de caractéristiques communes du père, affirmant par la dérision une forme de paternité commune aux membres de ce groupe. D’autres mèmes ne nécessitent pas tant de références, contribuant largement à leurs diffusions. La vidéo du clip musical *Gangnam Style* du chanteur Psy a notamment atteint des records inégalés en termes de diffusion<sup>4</sup> grâce à une très vaste campagne télévisuelle et sur internet. La diminution des implicites et la standardisation de l’écriture utilisée ont sans doute contribué à la diffusion, avec un langage du corps quasi universel, le pas de danse. Formellement, il s’agit d’un vidéo clip très classique dont la structure et le montage sont largement familiers du public. Les attributs des personnages du clip sont également de grands classiques du vidéo clip commercial : voitures, belles filles et bijoux en or. La présence d’un quartier spécifique de Séoul en Corée du Sud agit ici comme un attribut du

---

4. Première vidéo à avoir officiellement dépassée le milliard de vues sur Youtube. 1,733,769,243 vues, consulté le 13/08/2013 à 09 :35 GMT+8 <http://www.youtube.com/watch?v=9bZkp7q19f0>

contenu mais aucune des références ne nécessite de préalable linguistique particulier. De plus, les mystères de la musique et de la scénographie agissent bien évidemment au-delà de toute analyse formelle pour faire de ce hit un des mèmes Internet les plus connus dans le monde.

Les méméticiens disposent classiquement de deux procédés pour analyser la diffusion des mèmes :

— **La contamination**

Le mème se déplace à la manière d'un virus, en contaminant les sujets les plus susceptibles de l'être lors d'une phase d'exposition. L'exemple le plus classique pour ce modèle de contagion est la diffusion des croyances religieuses (Dennett, 2006)

— **La réplication**

Les activités culturelles humaines procèdent de l'imitation, notamment au travers de phases cruciales d'apprentissage. Ainsi, les mèmes existent et se diffusent dans les activités nécessitant une imitation : “ *If we define memes as transmitted by imitation then whatever is passed on by this copying process is a meme.* ” (Blackmore, 2006).

Le modèle épidémique de diffusion vient appuyer la vision éthologique du mème adaptée de la virologie. Les « sujets à risque » seraient plus à même d'être « contaminés » par une « exposition » suffisamment longue à tel ou tel mème (Wang et Wood, 2011). Blackmore définit néanmoins trois phases indispensables pour reconnaître un mème comme tel :

*“Memes fulfill the role of replicator because they exhibit all three of the necessary conditions; that is, heredity (the form and details of the behavior are copied), variation (they are copied with errors, embellishments or other variations), and selection (only some behaviors are successfully copied).”*, d'après Blackmore (2006)

Les trois aspects sont indissociables et forment selon Blackmore un “*véritable processus évolutionniste*” (Blackmore, 2006). La définition quasi tautologique du mème (« *Whatever is passed on* ») montre bien comment le mème en tant que concept est considéré chez

Blackmore non pas comme un phénomène mais plutôt comme un objet en-soi. La définition du mème en tant qu'objet autonome se heurte dès l'abord au risque de devenir un pur artefact de l'observation, n'existant que dans l'esprit de l'observant. Dawkins nous met en garde que son livre *The Selfish Gene* “devrait être lu presque comme s'il était de la science-fiction” (Dawkins, 1989). Les modèles de diffusion des mèmes comme répliqueurs ou virus ne recouvrent que partiellement la réalité observée empiriquement pour les mèmes Internet. Ainsi, l'appartenance des individus à tel ou tel groupe pré-existe au mème. Dans son appropriation se joue davantage l'expression d'un sentiment d'appartenance qu'une contamination qui nierait les termes de sa volonté individuelle pour y substituer l'individu comme sujet du mème.

## 2.3 Textualité des mèmes et formes d'énonciations numériques

### 2.3.1 Le mème comme figure rhétorique de l'écriture intertextuelle

Le mème peut être défini comme une série d'actes *d'énonciation* contribuant à l'existence et la reconnaissance mutuelles d'individus comme groupe. Cette définition s'oppose à l'idée d'une entité de sémantique autonome et constitutive d'une hypothétique culture commune. Les partages, commentaires, réappropriations puis transformations des *mèmes Internet* nous serviront de supports pour comprendre les *actes d'énonciation*. Ces courts messages faits de texte, image ou vidéo sont en quelque sorte les voix, répétitions, annonces et redites d'une foule d'individus et de groupes qui habitent la Toile. Au-delà de l'idée d'un “objet” numérique réifiant les actions en une substance figée, nous nommons *énonciation* les moments d'existence observables où se manifeste un mème. Comme pour les émoticônes, les pratiques actuelles de l'écriture des mèmes Internet sont à envisager comme des formes renouvelées d'oralité. Emails, chat ou réseaux sociaux, ces discussions font partie d'une “oralité seconde” (Ong, 1988) constituante des technologies de l'écrit à l'ère du numérique. Contrairement à l'oralité première des illettrés, cette oralité seconde est structurée par l'usage des technologies et notamment les structures formelles de

l'écriture pour le média :

*“Telephone, radio, television and the various kind of sound tape, electronic technology has brought us into the age of ‘secondary orality’. This new orality has striking resemblance to the old in its participatory mystique, its fostering of a communal sense, its concentration on the present moment, and even its use of formulas. But it is essentially a more deliberate and self-conscious orality, based permanently on the use of writing and print, which are essential for the manufacture and operation of the equipment and for its use as well.” (Ong, 1988)*

L'hypothèse de Ong est ici que l'oralité des écritures numériques est une forme manufacturée de l'expression orale, à laquelle préside la production (industrielle) des technologies. Les services de réseaux sociaux en ligne confirment l'hypothèse première de Ong puisque *Sina Weibo* ou *Twitter* contraignent industriellement l'écriture à une longueur maximum de 140 caractères. Néanmoins, l'oralité du même Internet et plus généralement des écritures numériques ne procèdent pas seulement de cette écriture conditionnée mais également de la mise en relation des textes : l'intertextualité. Là où pour Ong l'industrie médiatique vient contraindre l'écriture pour la réifier en produit industriel simulant l'oral, l'intertextualité vient subvertir ces limites imposées en renvoyant le lecteur à la page suivante.

En effet, le langage des nouveaux médias ne se contente pas de proposer des nouvelles opérations d'interactions et de navigations mais s'inscrit également dans de nouveaux modes de lecture et de narration (Manovich, 2001). Une des grandes difficultés pour l'analyse textuelle et narrative dans le cadre des nouveaux médias est de comprendre où commence et où se termine le texte. La structure éminemment relationnelle du discours narratif en ligne et son intertextualité en font un objet mal défini, que ses auteurs n'ont pas signé d'un point final. L'étude des discours hypertextuels s'apparente donc davantage à l'étude des formes de contes et légendes qu'aux études herméneutiques classiques sur des textes finis (Clément, 1995). Comme le note Lévi-Strauss dans son travail sur Propp, même si les contes possèdent souvent une structure similaire, leur étude en tant qu'élément culturel n'est révélatrice que dans un contexte précis et incarné. Il est inutile de vouloir extraire une supposée intention culturelle du texte car on se doit de le

comprendre lors de son énonciation, sur la place d'un marché au Maroc avec les conteurs de Ben Jelloun ou au pied du lit d'un jeune Européen avec les histoires compilées par les frères Grimm. Héritant à la fois des pratiques culturelles et folkloriques anciennes tout en se renouvelant sous les nouvelles contraintes de la technologie (Barber, 2008), le mème est lui aussi à comprendre dans son contexte d'énonciation. De récents travaux travaillent à comparer les modes de diffusion des mèmes avec ceux des traditions folkloriques (de Seta, 2014). En considérant les mèmes Internet comme un "*folklore numérique*", on comprend mieux la nature presque auto-référentielle de la relation entre le mème et la culture qui le voit naître. La transmission d'éléments particuliers dans la discursivité du mème en fait une figure rhétorique d'énonciation de sa propre origine.

Pourtant, le large pouvoir fédérateur de certains mèmes fascine. On se plaît à imaginer comment ils réunissent en se diffusant des groupes et individus distants, auparavant inconnus et étrangers. Les études sur la formation de communautés en ligne fleurissent, s'intéressant à l'émergence de groupes d'intérêts et d'activités lors de la diffusion d'information. Néanmoins, il nous semble important de se questionner sur la nature des relations créées lors des discussions partagées en ligne. Est-ce bien là le fait d'une réelle rencontre comme le croient les plus enthousiastes ? Ou au contraire est-ce le produit d'une machine médiatique et décérébrante, produisant du lien sans engendrer de rencontres comme le pensent les plus pessimistes ? Cette vaste question est sans doute un des enjeux centraux des questionnements autour de l'Internet, notamment pour le management et les sciences de gestion. Si l'énonciation est une pratique structurante pour un individu ("je suis"), elle peut l'être également pour un groupe ("nous sommes"). Dans la perspective empirique où nous nous situons, il nous faut tout d'abord interroger les pratiques du langage pour mieux comprendre comment les discursivités d'un mème peuvent agir sur les groupes. Le mème considéré comme un acte d'énonciation se caractérise non seulement par sa manifestation langagière, mais également par l'intention qu'il contient. Dans le cadre des réseaux sociaux, nous ne disposons que de très peu d'éléments sur l'intention car les données disponibles sont par définition le résultat d'actions passées (écriture, clics, etc.). Nous devons donc développer un modèle à la fois conceptuel et pratique pour étudier ces phénomènes d'énonciation dans le cas particulier des mèmes Internet. Wittgenstein dans ses *Recherches Philosophiques* défend une analyse pragmatique du langage en écrivant : "*Don't ask for the meaning, ask for the use.*" (Wittgenstein, 2004). Il s'agit de comprendre les jeux lan-

gagiers non pas comme un champ linguistique mais comme un ensemble d'actes qui font sens en contexte et possèdent une intention.

*“ Expressions have meanings even when they are not being used, but it is only in using expressions that a person means something. ” (Bach, 1994)*

Austin dans ses lectures sur William James met au centre du langage sa dimension pragmatique et propose de comprendre comment on « *fait des choses avec les mots* » (Austin, 1975). Ses leçons présentent une manière nouvelle de catégoriser les différents actes d'énonciation et montrent la grande diversité des éléments non-linguistiques présents dans ces actes. Austin introduit le concept de *performativité* pour nommer le processus permettant de construire par le langage une réalité extérieure au langage. Les mots ne nomment pas seulement les choses mais peuvent également les faire changer, voire les fabriquer. Austin s'intéresse donc aux énoncés selon leurs *sens*, leurs intentions (la *force*) ou leurs *effets*. Le concept de *performativité* a depuis continué son chemin tant en linguistique qu'en sciences sociales, et plus récemment dans des champs aussi divers que le management ou l'étude du discours scientifique et des pratiques légales (Denis, 2006). Les économistes ont aussi beaucoup discuté de la performativité des discours économiques sur l'économie réelle (Mackenzie, 2006). Les *cultural studies*, et plus précisément les *gender studies* ont aussi fait un large usage de ce concept pour exprimer l'influence de la matérialité des mots sur les comportements humains (Butler, 1993).

*« Pour qu'ils deviennent de « véritables » performatifs, les faits, les théories ou les formules doivent circuler dans des chaînes de traduction qui consolident l'assemblage des entités qui le composent et leur permet d'acquérir le statut de « matters of fact » (...). C'est lorsqu'ils arrivent à durer, c'est-à-dire à s'inscrire dans le monde (par l'intermédiaire d'objets, de textes, de dispositifs techniques complexes) que leur performativité s'accomplit. » (Denis, 2006)*

Les actes d'énonciation répétés modèlent donc le corps des personnes et de la société, y laissant durablement leurs marques. Les énoncés actualisent les discours des groupes sociaux sur eux-mêmes (Butler, 1993) et ce caractère performatif est constitutif de l'énonciation. Les énoncés collectifs que sont les mêmes Internet possèdent également cette dimension performative, actualisant le discours de certains groupes en réalités tangibles.

La circulation et la structuration de ces mêmes vient structurer le milieu numérique et peut ainsi influencer sur la définition de groupes sociaux et des rapports qu'ils entretiennent. Énoncer à son tour une image ou un mot semble un préalable permettant de dénouer l'implicite de l'énoncé et d'accroître la proximité dans l'expérience. Blackmore évoque en termes évolutionnistes le "potentiel transformatif" du même, avec ce qu'elle nomme la *variation* puis la *sélection*. Ici ce sont les actes d'*énonciation* qui forment une praxis du même. Pour nous, le même n'est pas un méta-symbole en évolution mais un réseau de praxis culturelles constitué de multiples d'actes d'énonciation. Ainsi, le même ne peut être simplement « copié » mais a besoin d'être acté pour exister. Dans la définition du même de Blackmore comme dans le cas des réseaux sociaux, nous observons que l'énonciation d'un même procède d'une variation parfois nulle, parfois minimale, parfois importante de sa forme d'origine. Cette déformation due à l'énonciation est le propre de la fonction d'apprentissage, notamment langagier. Le caractère performatif du même devient visible dans l'usage approprié de la variation qui en est fait.

Le succès de l'intention de l'énoncé est visible dans son imitation, avec comme garantie l'erreur ou la variation. Platon dans *La République* puis par la suite Aristote dans sa *Poétique*, s'interrogent sur la notion de *mimesis* définie comme les formes d'imitation permettant soit de reproduire, soit de styliser la nature. Pour Aristote, le but singulier de la mimesis est de mettre à jour la dimension empathique cachée de la nature, de la styliser pour y révéler le continuum de l'expérience propre à tous les êtres. Plus l'artiste s'approche de la nature en se dirigeant vers une imitation « véritable », plus il s'éloigne de la réalité de la nature. L'importance d'une approche rhétorique comme par exemple la stylisation devient le véritable moyen d'accès à la signification profonde des choses. Freud réutilisera le concept de *mimesis* pour décrire l'énonciation du sens d'un événement traumatique passé dans la vie d'un individu au travers d'activités créatives (art, parole, rêves etc.). Dans la continuité d'Aristote, Freud comprend le rêve comme une *mimesis* du passé et du réel, révélant au sujet un objet symbolique enfoui. Peut-être est-il possible de considérer le même Internet comme une mimesis des groupes sociaux et médiatiques qui le produisent. Les processus de symbolisation et de stylisation jouent en effet un rôle déterminant dans sa diffusion. Acte d'énonciation, il prend à chaque fois un sens légèrement différent. Là où le rêve freudien est une manifestation biologique de la mémoire inconsciente lors du sommeil, le même se manifeste sous des formes symboliques visibles et incarnées, *mimesis*

particulière du groupe des individus qui l'énoncent.

### 2.3.2 Typologie des mèmes Internet

Habituellement définie comme « *une forme typique de relation non linguistique entre des éléments discursifs.* »<sup>5</sup>, la figure rhétorique produit de l'intertextualité, des redites et des commentaires pour affirmer l'existence d'une idée.

En nous appuyant sur la littérature concernant les figures rhétoriques, nous pourrions chercher à définir plus précisément une typologie des mèmes fondée sur ses formes dans le discours. En grec ancien, le *topoï* défini à la fois un lieu, un endroit et un ensemble de formes rhétoriques utilisant des motifs particuliers lors de l'argumentaire afin de persuader lors de joutes oratoires. En littérature comme en mathématiques, le terme de *topos* définit également un ensemble de catégories ouvertes mais connues, « *un protocole de description des univers possibles* » (Badiou, 2006). Le mème peut donc être compris comme un *lieu commun*, idée « reçue » utilisant des situations ou des images communes et stéréotypées pour opérer une transformation sémantique en jouant sur la répétition d'éléments (les sèmes du discours). En rhétorique, l'usage du *topos* a pour objectif de contribuer à la persuasion de l'auditeur par la mobilisation subtile d'éléments de culture commune. Tout l'art du rhéteur consiste à trouver un moyen subtile d'actualiser un lieu commun en une situation unique propre au contexte pour convaincre. Dans le discours, il prend bien souvent la forme de l'anecdote que la rumeur se charge de diffuser, sa diffusion étant d'autant plus efficace qu'il possède un caractère amusant ou railleur (Flaubert, 1997). Concevoir le mème comme une forme de pratique rhétorique semblable au lieu commun nous permet de resituer le phénomène des mèmes Internet dans la continuité historique des pratiques de l'écriture et de l'énonciation rhétorique, ainsi que de ses modèles d'analyse socio-textuelle et littéraire (Plantin, 1993). Ainsi, nous pouvons aborder la lecture des mèmes Internet sous le jour de leur existence aussi bien formelle (textuelle) que rhétorique (comme actes d'énonciations et de persuasion). Les mèmes tout comme les lieux communs se donnent à voir d'abord sous la forme de paradoxes, qui deviendront eux-mêmes des lieux communs. L'originalité d'un lieu commun en devenir se définit dans une tension constante entre imitation et nouveauté, subversion et actualisation de formes canoniques : « *dialogue*

---

5. *Les figures de rhétorique*, Laurent Jenny, Université de Genève, 2003



*de l'horizon d'attente et de l'écart esthétique, c'est-à-dire le jeu du classique et du moderne, la tension entre le même et l'autre qui existe, dans tout texte et dans toute lecture, entre le plaisir et la jouissance, pour reprendre les mots de Barthes". (Compagnon, 1997)*

Bauckhage (2011) a identifié des caractéristiques typiques d'après une centaine de mèmes parmi les plus largement diffusés d'Internet :

1. **Humour** : Le mème doit posséder une dimension comique et accrocheuse
2. **Intertextualité** : Le mème mobilise un ou des renvois à d'autres éléments culturels ou textuels, souvent implicites.
3. **Juxtaposition atypique** : Les éléments visuels ou sémantiques mis en jeu dans le mème ne possèdent pas de corrélations apparentes et c'est la mise en relation de plusieurs objets improbables qui en fait un objet intéressant.

L'étude en question nous offre un début de critères mais porte seulement sur un type précis de mèmes à caractère plutôt comique, ignorant les discussions plus sérieuses, d'ordre politique notamment. La forme particulière de juxtaposition atypique observée par les auteurs serait en rhétorique une "*métaphore in praesentia*" décrite comme une « *figure de rapprochement analogique entre deux représentations co-présentes* » (Jenny, 2012). Nous nous trouvons donc en présence d'une forme particulière de mème seulement. Si l'humour sous toutes ces formes (blague, sarcasme, ironie, etc.) est un élément très répandu qui favorise la circulation des contenus en ligne, il nous semble néanmoins un peu réducteur de se limiter à cette définition. De nombreux mèmes Internet existent non pas grâce à l'humour mais grâce au *pathos* qu'ils dégagent. Le mème *Kony2012*, un des plus diffusés de l'histoire d'Internet, présentait le militaire ougandais Joseph Kony dans une vidéo faite d'images de guerre et d'enfants en pleurs<sup>6</sup>. Ainsi, on peut dire que les mèmes Internet utilisent les formes classiques de la rhétorique adaptées au langage médiatique moderne, teintés d'humour ou de sensationnalisme. En se saisissant du monde social et politique, les mèmes Internet produisent d'intéressants discours sur les faits qu'ils relatent - et sur la technologie qui les produit. Les révolutions du Printemps Arabe avaient notamment soulevé l'espoir d'un monde meilleur par l'usage des réseaux sociaux pour instaurer la démocratie (Lotan *et al.*, 2011). Quelques jours plus tard, ces mêmes réseaux sociaux de-

---

6. *KONY2012 : See How Invisible Networks Helped a Campaign Capture the World's Attention, Social Flow*, <http://alturl.com/zniry> consulté le 28 Février 2014 GMT+1

venaient un danger pour la société responsable des flambées de violence durant une série d'émeutes à Londres (Casilli et Tubaro, 2011). Les différentes intentions des discours et actes d'énonciation à l'œuvre dans un même peuvent donc servir de point de départ pour dresser une typologie des mèmes. Cette typologie ne peut être exclusive, tant les glissements sémantiques et symboliques s'opèrent facilement. En nous appuyant sur différents exemples et sur la littérature, nous dressons une première ébauche de typologie des mèmes Internet qui sera ensuite approfondie par l'étude empirique au cœur de cette étude. Pour les exemples, nous essaierons de donner les références et implicites indispensables permettant de comprendre le mème dans son intertextualité.

Objectifs du mème	Exemples célèbres et références
Absurdiste, humour	LOLCats, Tumblr (Bauckhage, 2011)
Actualité, satire, commentaire social	CaoNiMa (Mina, 2012), Cute Cat Theory (Zuckerman, 2008)
Publicité, marketing viral	Gangnam Style (Bolsover <i>et al.</i> , 2013), Memetic Marketing (Flor, 2000)
Marketing politique, soutien, pétition	Obama Ohio Campaign (Walker, 2012), Pétitions en ligne (Adamic <i>et al.</i> , 2014)
Fan clubs, adoration	Fan-fiction, Machinima, etc.
Hoax, spam, canulars	Email spam, "Nigerian scam"

**Tableau 2.3.1** – Les différents type de mèmes Internet observables - tableau réalisé d'après la littérature citée

Cette typologie nous permet de poser un premier regard sur des catégories non-exclusives constituant le paysage des mèmes Internet, inspirées d'exemples et de la littérature existante. Nous pouvons d'ores et déjà noter que l'ensemble de ces différentes pratiques pré-existent à l'Internet et sont souvent la reconduction sous une forme numérique de discours pré-existants (la publicité, la propagande politique ou la rumeur). La forme et l'échelle de la diffusion sont néanmoins des différences importantes ainsi bien sûr que les situations d'énonciation (dans une file d'attente ou devant son ordinateur).

### Absurdiste, humour

Connaissant parfois des succès planétaires, un des exemples les plus caractéristiques des mèmes comiques est celui des "LOLCats", ces vidéos ou images de chats illustrés

de citations et autres accessoires comiques. Omniprésent sur la Toile, l'immense répertoire constitué par les photos et vidéos comiques de chats dans différentes situations (jouant du piano, sautant dans des boîtes en cartons, etc.) est un des plus visionnés de l'Internet (Bauckhage, 2011).



(a) a LOLcat



(b) another LOLcat

**Graphique 2.3.1** – Some LOLcats, d'après le <http://www.lolcats.com>, consulté le 4 Avril 2013 à 15 :32

### Actualité, satire, commentaire social

En regroupant ces trois catégories, nous souhaitons considérer la pratique ancienne des discussions politiques ou polémiques autour de faits divers ou d'actualités que l'Internet vient aujourd'hui renouveler dans la forme. Nous avons vu précédemment avec les "crabes de rivières" (voir 1.3.1) comment la satire politique en Chine se manifestait bien souvent sous la forme de mèmes Internet. Par les voies de l'Internet, de nombreux faits divers sont devenus en Chine des affaires d'état, influençant le jeu politique et offrant un nouveau canal de diffusion pour les pratique subversives. Connue pour son œuvre artistique, l'artiste chinois Ai Weiwei est une des figures célèbres de l'Internet chinois. Il utilise abondamment blog et microblog. Son travail artistique depuis plusieurs années a pris la tournure d'un bras de fer médiatique avec le gouvernement. Son père, poète célèbre du régime communiste sous Mao fut déchu durant la révolution Culturelle, laissant à son fils à la fois un goût pour les arts et une violente amertume pour le pouvoir en place à Pékin. Critiquant ouvertement la corruption des officiels et de la police, Ai Weiwei a vu régulièrement ses comptes Internet de blog et microblog fermés, notamment par le

service Sina Weibo. Sur Twitter néanmoins, il jouit d’une grande popularité avec près de 200.000 “followers” pour son compte officiel @aiww. Au fil de son blog relatant ses nombreux déboires avec le gouvernement, on croise bien souvent une figure mythologique née de la toile chinoise, le *caonima*.



(a) Une image de la vidéo originale représentant le *caonima*. (b) Un caractère chinois composé spécialement par les internautes pour signifier cet animal mythique

**Graphique 2.3.2 – Caonima** : un animal mythique du web chinois, d’après le *Grass-Mud Horse Lexicon Classics* publié par *China Digital Times* (2013)

Luttant contres les crabes de rivière, le *caonima* symbolise la lutte contre la censure pour un Internet libre. Apparue pour la première fois dans une vidéo virale<sup>7</sup>, le *caonima* est un animal de la famille des camélidés (l’alpaga) courant fièrement dans les prés sur une petite musique de dessin animé aux paroles entièrement réécrites pour l’occasion. Littéralement “cheval d’herbe et de boue”, le mot “caonima” (草泥马) cache en fait un double sens puisque son homophone (操你妈) est une grossière interjection à l’intention des génitrices des censeurs de Pékin. Devenu aujourd’hui une véritable icône anti-censure, il n’est pas rare de le croiser sur un tee-shirt ou accroché à un sac dans de nombreux endroits improbables en Chine. Comme le note très justement An Xiao Mina : “les mêmes sont les graffitis du web censuré” (Mina, 2012). Malgré le zèle des pouvoirs publics à les effacer le plus rapidement possible, ils témoignent de l’existence de réalités occultées qui se manifestent souvent de manière dérisoire, grotesque et improbable, mais sont énoncées malgré tout.

7. *Song of the Grass-Mud Horse (Cao Ni Ma)*, Youtube <https://www.youtube.com/watch?v=wKx1aenJK08>, consulté le 28 Février 22 :08

## Publicité et marketing viral

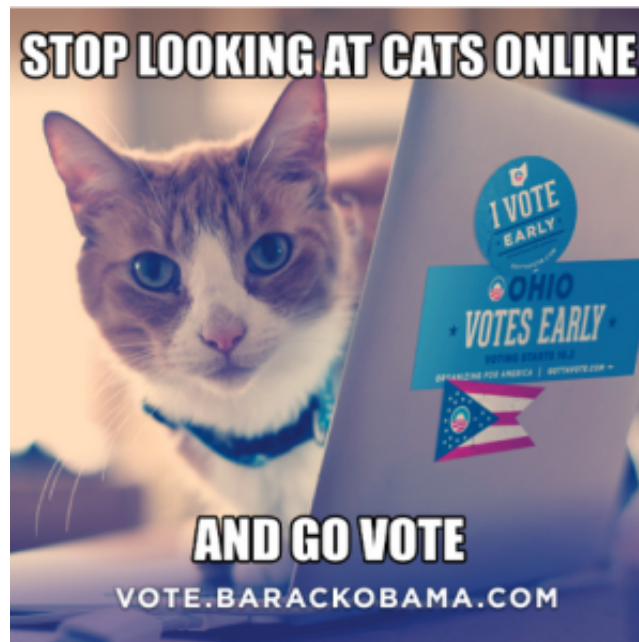
La popularité des services de réseaux sociaux a entraîné de nombreuses marques à se concentrer sur ce nouveau média pour mener leurs campagnes de publicité et de promotion. Les utilisateurs chinois sont très fortement engagés dans la création de nouveaux contenus avec 76% des utilisateurs chinois créant davantage de contenus qu'ils n'en lisent, contre seulement 20% en France (Wang *et al.*, 2013). Ainsi, de nombreuses marques cherchent à concevoir des interactions en ligne utilisant les mêmes comme véhicule de messages commerciaux afin de mettre à contribution les internautes dans la diffusion de leur marque. Une des stratégies marketing les plus répandues consiste à créer un *hashtag* amusant afin d'inciter les internautes à s'en saisir, le partager et créer de nouveaux contenus. La marque de préservatifs *Durex* a notamment connu un large succès avec le hashtag *#BienEtreNocturne* (en Chinois *#夜福利*) sur Sina Weibo. Drôle et un peu osée, cette campagne a généré un fort engagement des internautes, contribuant largement à la propagation du profil utilisateur et du nom de la marque (Shi, 2012). En récupérant les données des utilisateurs impliqués dans la diffusion, la marque peut également procéder à un travail statistique d'analyse pour mieux connaître les utilisateurs intéressés par ses produits. Ce type d'information s'avère précieux dans un marché chinois en mouvement où il existe un fort besoin pour des études de marché précises auprès de segments de populations actifs en ligne (adolescents, jeunes, classe moyenne émergente...) (Bergstrom, 2012). Malgré le paysage morcelé des services de réseaux sociaux chinois, les utilisateurs chinois semblent davantage enclins à suivre et interagir avec les marques que leurs homologues américains notamment <sup>8</sup>.

---

8. D'après Ken Hong, Sina Weibo General Marketing Manager <http://adage.com/article/global-news/questions-sina-weibo-s-ken-hong-china/239508/>, consulté le 19 Février 2014 à 12 :45

## Marketing politique, soutien, pétitions

L'industrie n'est pas seule à s'être emparée des réseaux sociaux. Les grandes entités politiques ont également saisi à bras le corps ce nouveau média. Les plus brillants exemples sont sans doute les campagnes menées par Barack Obama pour la présidence des Etats-Unis. L'importance capitale des réseaux sociaux a placé la stratégie de diffusion virale en haut de la pile des préoccupations pour l'organisation de ces deux campagnes électorales (Miller, 2008) De nombreux mêmes Internet ont été utilisés pour fédérer les votants. Lors de l'étape de campagne auprès des électeurs de l'Ohio, un état décisif de la course présidentielle, l'équipe d'Obama a notamment fait usage d'un des fameux LOCatS pur mobiliser son électorat.



**Graphique 2.3.3** – LOLCat utilisé durant la campagne d'Obama en Ohio - d'après *Obama Campaign Deploys Cat Meme to Get Out the Vote in Ohio* sur Politicker, consulté le 28 février 2008 à 11h41 GMT+1

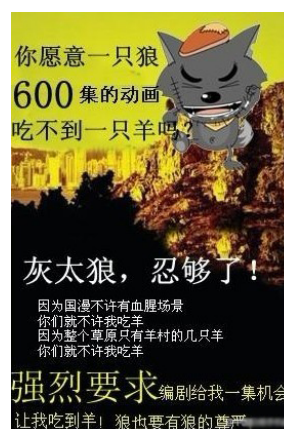
En Chine également, les réseaux sociaux sont très utilisés pour la diffusion d'idées lors de campagnes politiques. Les groupes nationalistes soutenus par le gouvernement se saisissent régulièrement de l'actualité pour rebondir et rassembler les foules (Wu, 2007). Là encore, les mêmes jouent un rôle important dans l'appropriation des discussions au travers de la ré-énonciation de sujets controversés en des termes différents. Les tensions grandissantes entre les habitants du territoire de Hong Kong et ceux venus de la Chine intérieure ont été notamment le sujet d'une discussion intéressante par mêmes interposés. Les visites à Hong Kong sont très



régulées pour les citoyens venus de Chine intérieure mais le flux massif de touristes n'a néanmoins cessé de croître depuis plusieurs années avec l'augmentation des quotas et l'accès aux congés dans les grandes villes de la RPC.



(a) Le tract original



(b) Exemples de tracts réalisés par les internautes en réponse à la campagne

**Graphique 2.3.4** – Détournement de tracts hongkongais anti-chinois, d'après *The Civic Beat*, <http://reader.thecivicbeat.com/2012/03/locusts-and-pandas-and-bears-??-o-mai/> consulté le 1er Mars 2014 à 22h58

De nombreux chinois se rendent donc à Hong Kong pour voyager mais également profiter de la détaxe des produits de consommation et des services publics de bien meilleure qualité. La qualité des hôpitaux de la ville et l'application du droit du sol dans la loi hongkongaise amènent de nombreuses jeunes mères venues de Chine à traverser la frontière pour venir accoucher à HK. Cette pratique très controversée donne à l'enfant le passeport hongkongais et se monnaie à prix d'or, rendant l'accès aux hopitaux de plus en plus cher et attisant la colère des habitants de HK. En septembre 2012, une pétition lancée à Hongkong a cherché à recueillir des votes pour interdire l'accès aux hôpitaux aux chinois venus de RPC. Un tract publié pour cette campagne a largement circulé sur le web chinois ; on pouvait y lire : “*Habitants de Hong Kong, nous avons assez souffert! (...) Ne nous laissons pas envahir par la vermine venue de Chine intérieure.*”, accompagné d'une photo d'insecte (voir fig. 2.3.4)

Les internautes chinois choqués mais armés d'un humour toujours très cinglant ont donc entamé une contre-campagne en proposant des versions modifiées et réécrites du dit tract. Réactions épidermiques, propos nationalistes sommaires, mais aussi

critique du tourisme de masse, dénonciation de la corruption et de la mauvaise qualité des soins hospitaliers en Chine et nombreuses blagues absurdes, les adaptations et réponses singeant le tract d'origine ont montré un panel d'arguments et de réactions qui a permis de crever l'abcès et d'ouvrir un débat sur ce problème épineux entre les habitants de HK et ceux de la RPC.

### Fan clubs, adoration

Comme tout mass media, les réseaux sociaux possèdent une énorme quantité de contenus consacrés aux stars, à leurs vies, leurs coups durs et leurs derniers films et chansons. Développant des stratégies d'envergure sur les médias chinois, de nombreuses stars internationales comme Brad Pitt ou Kobe Bryant ont fait leur arrivée sur *Sina Weibo*. Néanmoins, leur influence reste incomparable à celles des stars originaires de Chine, de Taiwan et de Hong Kong ou bien encore de Corée du Sud, principal producteur de pop culture en Asie (Martel, 2010). Une des stars les plus plébiscitées dans les réseaux sociaux est également une étrangère puisqu'il s'agit de la japonaise Sola Aoi, ex-actrice de film pornographique devenue une des 10 personnes les plus suivies sur Sina Weibo avec près de 15 millions de followers<sup>9</sup>.



**Graphique 2.3.5** – Un message disant “Les peuples de la Chine et du Japon sont des amis” posté par la star du cinéma pornographique japonaise Sola Aoi

9. D’après Sina Weibo <http://www.weibo.com/1739928273/yBJYNt701>, consulté le 1 Mars 2017 à 23h11



Personnalité médiatique publique, la star a depuis plusieurs années utilisé les réseaux sociaux pour nouer contact avec le public chinois qui semble bien la connaître, malgré l'interdiction de la pornographie en Chine. Aujourd'hui retraitée du monde du X à 30 ans, Sola Aoi utilise sa popularité pour défendre une paix durable entre la Chine et le Japon. Alors que les tensions politiques exacerbées entre la Chine et le Japon ont mené à des rixes et des maltraitements envers les ressortissants dans les deux pays, elle a notamment contribué à créer un appel à la non-violence sous forme d'un mème qui a été très diffusé.

Ainsi, les fan clubs et les stars elles-mêmes jouent un rôle important dans la création et la diffusion des mèmes, occupant une large place dans le paysage médiatique et notamment celui des réseaux sociaux.

### **Hoax, spam, canulars**

Un des plus célèbres exemples dans ce domaine sont les emails dits de “fraude 4-1-9” cherchant à extorquer de l'argent au destinataire. Le numéro 419 correspond au numéro de l'article interdisant la pratique de l'escroquerie en ligne dans le code pénal nigérian. En effet, la forme la plus connue de cette arnaque est celle du “prince nigérian” demandant un numéro de compte bancaire pour y transférer rapidement des fonds. Adapté sous de multiples formes, ce type de messages existe également sur les réseaux sociaux sous la forme de *bots*, comptes tenus par des robots postant des messages promotionnels.

I am Stella Amah 19 years of age the only daughter of late Mr Boni Amah whom was killed by the rebels that attacked our country cote d'Ivoire west Africa and took over our town (BOUAKE). I ran to Abidjan the economical capital of cote d'ivoire from were I am contacting you. Before the death of my father he told me that he has a sum of US\$9,000,000(Nine million united states dollars) kept in a private security company here in cote d'ivoire in my name as the next of kin...

**Graphique 2.3.6** – Extrait d'un “Nigerian Scam”, d'après <http://www.hoax-slayer.com/stella-amah-scam.shtml>, consulté le 2 Mars 2014 à 18h50

Au regard des différents exemples données ici, nous voyons qu'il est difficile de classer les mèmes selon des catégories précises et qu'en bien des endroits ces catégories se recoupent : les stars font de la politique alors que les politiques font dans le comique. Ainsi, il ne s'agit pas de dresser des catégories rigides mais de disposer d'une classification flexible pour pouvoir identifier clairement les éléments discursifs présents dans les mèmes

Internet. Structure du réseau social, contenu des mèmes, formes multimédia, éléments de contexte, etc. , ces multiples dimensions des mèmes peuvent être observés empiriquement par l'analyse de données.

# Problématique et hypothèses

## Objet et problématique

L'apparition en Chine d'une industrie dédiée aux contenus des réseaux sociaux s'inscrit dans la continuité historique du développement industriel et politique du Web chinois. Les nombreuses discussions hébergées par le service de microblog *Sina Weibo* recoupent souvent blagues, faits divers, scandales politiques, marketing politique et commercial, etc.

Parmi ces nombreux contenus échangés sur les réseaux sociaux, les *mèmes Internet* offrent une forme typique faisant appel à l'humour pour mobiliser des références culturelles communes. La vitesse et l'échelle de leur circulation en font des objets numériques intéressants pour explorer les nouvelles interactions ayant lieu sur les plate-formes Web.

À mi-chemin entre écriture et oralité, l'appellation de *mèmes Internet* recouvre pourtant des réalités diverses qui semblent procéder de motivations et de régimes d'expression forts différents. La naturalisation implicite du modèle "viral" souvent invoqué nous empêche de considérer l'ensemble hétéroclite de pratiques qui se déroulent lors la diffusion d'informations en ligne. Seule une observation détaillée peut nous permettre d'apporter un éclairage sur les régimes d'expression présidant à la production de ces contenus.

Comment s'opère la diffusion des *mèmes* sur *Sina Weibo* ? Quelles sont les particularités de ce type d'objets par rapport à d'autres formes de contenus plus traditionnels ? Comment pouvons-nous observer ces *mèmes Internet* ? Quelles relations entretiennent-ils avec le milieu numérique qui les produit ? Ont-ils une existence territoriale ?

Afin d'observer les différents types de contenus qui circulent sur la plate-forme de

réseau social *Sina Weibo*, nous allons procéder au développement d'un dispositif utilisant l'analyse et la visualisation de données. La série d'hypothèses infra accompagnera la démarche méthodologique de conception de cet outil.

## Hypothèses

### **La majorité des contenus circulant sur les réseaux sociaux s'apparentent largement à ceux des médias traditionnels**

L'objet numérique communément appelé dans la littérature *même Internet* recouvre différents types de contenus qui ne s'apparentent pas à ce terme. La plupart sont en effet des formes narratives, commerciales ou journalistiques qui pré-existent à l'apparition d'Internet. Qu'il s'agisse de propagande politique ou de campagnes publicitaires, la dimension stratégique de la diffusion médiatique est un enjeu important pour tout système économique et politique. Nous pensons qu'il existe de grandes similarités de diffusion entre les contenus majoritaires des services de réseaux sociaux et ceux des médias plus classiques comme la télévision ou la radio, soumis à des contextes de production industrielle similaires.

### **Les mêmes sont à comprendre comme des actes d'énonciation.**

Si les formes de contenus des médias traditionnels trouvent une continuité sur les réseaux sociaux, la fonctionnalité nouvelle de ce type de service est la possibilité de converser en ligne. Reconduction de phénomènes anciens le même Internet peut se comprendre comme une rumeur, un débat ou une blague à forte circulation. Ces formes archétypales se constituent autour de pratiques essentiellement orales, existantes aujourd'hui sous le régime particulier de l'écriture en ligne. Ainsi, l'étude des mêmes Internet nécessite de mobiliser un cadre conceptuel adéquat considérant ces actes de communication comme formes de l'énonciation.

**L'analyse de la diffusion ne doit pas se contenter de l'analyse conversationnelle, mais doit mobiliser les modèles étudiant l'énonciation.**

Près d'un siècle de recherche en sciences de la communication nous ont permis de développer des modèles d'analyse complexes des actes de communication. Pour les échanges en ligne, nous ne pouvons nous contenter de percevoir uniquement la dimension conversationnelle sans prendre en compte les multiples aspects non-verbaux, para-verbaux ou même sémantiques qui entrent en jeu dans les actes de communication en ligne. Notre première approche cherchera à comprendre non seulement les échanges entre utilisateurs (conversation), mais également les structures lexicales (sémantique), temporelles et géographiques qui émergent de ces échanges. L'analyse de la diffusion des mêmes nous servira de base pour discuter et comprendre comment se structurent ces actes de communication en ligne.

**L'analyse et la visualisation de données permettent d'établir une classification des discussions en ligne selon les structures de leurs diffusions.**

La disponibilité d'immenses volumes de conversations sous la forme de vastes jeux de données issues des réseaux sociaux nous permet aujourd'hui de pratiquer de nouveaux types d'analyses pour comprendre les phénomènes entourant la communication en ligne. Nous visualisons la diffusion d'un ensemble de mèmes issus du réseau social *Sina Weibo* sous formes de graphes nommés *topogrammes*. Cette observation permet de distinguer et classer les mèmes d'autres formes de contenus en ligne pour dresser une typologie des modèles de conversations.

**La circulation des contenus sur les réseaux sociaux accroît la fragmentation du milieu numérique.**

L'observation de ces topogrammes permet de comprendre comment les conversations en ligne actualisent le *milieu numérique*, compris comme l'ensemble d'outils et protocoles technologiques qui les produit (Hui, 2012). Nous pouvons également identifier des formes de communication et d'échanges plus ouvertes à la participation dites *associatives* et d'autres dites *dissociatives* autorisant plus difficilement l'échange. Nous pensons

notamment que la segmentation de l'attention suscitée par la valorisation publicitaire des contenus induit un design largement dissociatif des services de réseaux sociaux, qui conduit à une fragmentation accrue des relations en ligne (une structure en *small worlds*).

Afin d'approcher ces différentes hypothèses, nous avons choisi de nous livrer à la conception d'un outil nous permettant d'analyser de larges jeux de données issus de *Sina Weibo*. La dimension expérimentale de ce travail nous amènera à effectuer de nombreux choix méthodologiques que nous allons maintenant discuter.

# 3

## Réflexions méthodologiques et design expérimental

LA VISUALISATION DE DONNÉES nécessite de se saisir d'outils informatiques, mais également d'allier des savoirs provenant de domaine aussi divers que l'algorithmique, la statistique, la cartographie ou le design graphique.

Notre démarche expérimentale d'observation des mèmes circulant sur *Sina Weibo* implique donc un ensemble de réflexions méthodologiques préalables, notamment sur la validité des méthodes d'analyse de données et leur rôle en sciences sociales. Un protocole expérimental doit également être mis en place pour explorer les données issues des réseaux sociaux. Plus spécifiquement, il s'agit ici de pouvoir détecter, identifier puis visualiser un ensemble de mème dans une large quantité de données comprenant du texte en langue chinoise, des informations géographiques et la trace des interactions entre les utilisateurs.

## 3.1 Sciences sociales du réseau

Pour l'empiriste, l'acte essentiel de la recherche est l'observation. Structure schématique faite de points et de lignes, le réseau n'offre que peu de prise pour une approche empirique. Pourtant, ce concept protéiforme traverse aujourd'hui les disciplines pour se retrouver au centre des débats scientifiques. Couplé à l'autre grand insaisissable de l'étude qu'est *l'information*, le modèle du réseau fait la promesse de nouvelles perspectives en offrant un cadre conceptuel commun pour les sciences et de nouveaux horizons méthodologiques grâce aux outils informatiques. L'interrogation autour des réseaux se trouve donc plus que jamais au cœur du devenir des sciences.

### 3.1.1 Le réseau comme enjeu pour l'étude

Durant la seconde guerre mondiale et jusqu'à la fin des années 1950, les esprits les plus brillants du siècle (Gödel, Von Neumann, Einstein..) se côtoient à *l'Institut des Etudes Avancées* de Princeton (IAS) aux USA. Leurs travaux posent les bases logiques et technologiques d'où émerge la science informatique actuelle. Dans le domaine de la philosophie mathématique tout d'abord, les interrogations amorcées par Russel dans ses *Principia Mathematica* puis leur critique par Gödel dessinent les contours d'une "machine récursive", ancêtre mathématique de la machine de Turing et de l'ordinateur de Von Neumann. Considéré comme le père de l'ordinateur, Von Neuman s'exile d'Allemagne pour rejoindre l'IAS nouvellement fondé dès 1933 où il poursuit des recherches dans le domaine de la physique nucléaire. En 1943, Von Neumann intègre le *Projet Manhattan* dirigé par Oppenheimer où il est chargé de superviser l'immense processus de calcul nécessaire à la construction de la bombe qui s'abattra sur Hiroshima le 6 Août 1945. C'est durant cette période que Von Neumann développe l'architecture encore utilisée aujourd'hui comme base fondamentale du design électronique. Dans ses lectures à l'Université de Yale en 1957 parues sous le titre célèbre de *L'Ordinateur et le Cerveau*, von Neumann (1958) identifie les différentes unités d'un ordinateur : unité d'arithmétique logique, unité de contrôle, mémoire et entrées/sorties. Les travaux teintés d'ombre de ce prestigieux mathématicien font écho à ceux de l'anglais Turing qui travaille également sur la question de l'intelligence des machines depuis le début de la guerre. Leur correspondance témoigne



du respect mutuel qu’entretiennent alors les deux savants, ainsi que de leurs nombreux questionnements sur la possibilité d’une machine intelligente (Istrail et Marcus, 2013). Au cœur de leurs discussions se trouve la quête d’un modèle universel capable d’expliquer le fonctionnement de l’activité cognitive.

Norbert Wiener, un autre prodige des mathématiques prend également part à cette recherche. Sa théorie qu’il nomme *cybernétique* place la notion d’information au cœur de la réflexion sur le fonctionnement des systèmes : “*Information is information, not matter or energy*” (Wiener, 1948, p. 155). Utilisant les concepts de bruit, de messages et de *feedbacks*, il jongle entre mathématiques appliquées et sciences cognitives pour comprendre les phénomènes de transmission de signaux complexes. Dans une lettre du 29 Novembre 1946, Von Neumann écrit à Wiener que l’étude du cerveau “*the most complicated object under the sun, literally*” nécessite de poursuivre une réflexion transverse à de multiples disciplines scientifiques (Masani, 1990). Depuis le début de cette même année 1946, les “conférences cybernétiques” aussi connues sous le nom de “conférences Macy” ont débuté à New York dans le but de définir “*une science générale de l’esprit humain*”<sup>1</sup>. Regroupant physiciens, cognitivistes, biologistes, anthropologues et linguistes, ces conférences sont l’objet de nombreuses publications dont *The Human Use of Human Beings* (Wiener, 1988) qui propose d’étudier la société en considérant les communications entre hommes et machines. Le travail mené lors de ces conférences est souvent considéré comme la pierre fondatrice du champ des études sur la communication (Breton, 1997; Winkin, 1981). Quelques années plus tard, McLuhan grâce à son idée controversée de *village global* (McLuhan, 1962) amène un large auditoire autour des œuvres de Wiener et l’étude des modes de communication. Progressivement se constitue un champ épistémologique pour l’étude de l’information dont la définition reste encore aujourd’hui un enjeu important (Wolton, 1997). En France, les Sciences de l’Information et de la Communication se sont principalement structurées autour de la critique des médias (Mattelart, 1986; Debray, 1991) dans une tradition européenne déjà bien installée (Adorno et Horkheimer, 2005). Aux Etats-Unis, les *cultural studies* (Hall, 2003) s’interrogent davantage sur l’économie politique des symboles et trouvent leurs lettres de noblesse dans les *media studies*, aujourd’hui devenues une discipline universitaire reconnue en Angleterre et aux USA. Toutefois,

---

1. Foundation for Cybernetics, The Macy Conferences <http://www.asc-cybernetics.org/foundations/history/MacySummary.htm> consulté le 10 Mars 2014 à 16h16 GMT+1

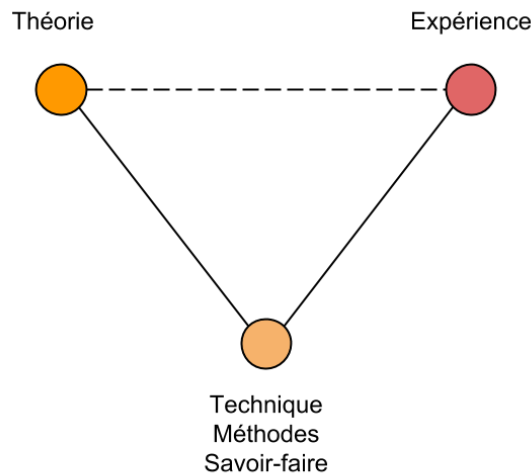
l'émergence d'un "*paradigme communicationnel*" (Bougnoux, 1993) nécessite un dialogue pas toujours évident entre les différentes disciplines constituées autour de l'étude des pratiques de l'information et de la communication : sociologie des médias, études des systèmes d'information, informatique, sciences de l'information et de la communication, etc.

Alors que le programme fixé par la cybernétique a pour ambition d'explorer les structures, possibilités et contraintes des systèmes communicants, l'image fermée et cyclique du système devient rapidement trop étreinte pour une réflexion sur les relations alors en pleine expansion. Deleuze et Guattari proposent dans l'introduction de leur livre *Milles Plateaux* (Deleuze et Guattari, 1972) de s'éloigner de l'arborescence classique du système auto-suffisant pour envisager les liens entre sujets et objets sous la forme ouverte et combinatoire du *rhizome*. Le rhizome se définit par son caractère non-fini et fait de l'étude des causalités un phénomène contextuel aux spécificités pas forcément reproductibles. Imperméable à la catégorisation et aux classifications ordonnées, l'objet d'étude in-formé devient *open-ended*, moment donné à voir et état unique d'un ensemble plus vaste. La vérité ou la validité n'est donc plus à découvrir dans l'analyse, mais plutôt dans l'articulation des contingences entre objets et environnements dont se saisit le champ émergent des études sur la complexité (Morin, 1990). Dans le même temps, les progrès de l'intelligence artificielle et de la robotique viennent également questionner notre définition du vivant face à ces nouveaux êtres (Hofstadter, 1999). Le modèle du *réseau* vient à son tour in-former la pensée scientifique comme un nouvel *episteme* foucaldien offrant une grille de lecture des faits sociaux renouvelée (Castells, 1989; Latour, 1996). Aujourd'hui, la structuration d'un champ de recherche cohérent et multi-disciplinaire autour de l'étude des réseaux reste un enjeu important pour le monde de la recherche (Brandes *et al.*, 2013). Les théories de la complexité offrent notamment des clés de lecture pour une approche scientifique capable de rassembler des disciplines autour d'un même objet. Ce dialogue parfois difficile semble néanmoins nécessaire pour approcher une réconciliation des sciences humaines, des sciences naturelles et des sciences du vivant où cohabiteraient "*l'étude des organismes, des organes et des organisations*" (Morin, 1990)

### 3.1.2 Code, données et les nouveaux outils de l'écriture scientifique

Les objets ouverts et mal définis du réseau posent néanmoins la question d'une continuité avec le projet aristotélicien d'une compréhension du monde par le savoir analytique. La considération d'un objet comme moment du réseau rend caduque l'idée de sa définition *a priori*, en dehors des relations qu'il entretient avec son environnement. A l'inverse, une définition des objets entre origine, fonction et destination ne devrait pas éluder non plus la réflexion ontologique autour de son *être*. Le philosophe américain William James propose une méthode qu'il nomme *empirisme radical* où l'existence des objets doit être considérée par le prisme de l'expérience (James, 1912). L'expérience, même scientifique, existe dans un contexte, un lieu, un temps, un moment, une succession d'étapes. C'est cet ensemble qui nous permet de percevoir les phénomènes étudiés. Si l'acte essentiel de l'empiriste et du philosophe est l'observation, alors elle doit être radicale dans son honnêteté et accepter l'expérience dans son unicité dont de nombreux facteurs sont non-reproductibles. Cette question de l'expérience et de sa reproductibilité est depuis toujours au cœur du débat sur les sciences et se pose comme un des ressorts fondamentaux du savoir scientifique. Les sciences dites "dures" comme la physique, la biologie, l'informatique ou même la géologie valident leurs hypothèses grâce à l'usage extensif d'appareillages multiples devenus une clé de la reproductibilité des expériences. Du simple microscope à l'accélérateur du CERN, la méthodologie d'observation est pour ces disciplines souvent indissociable de la technologie. Le travail de la recherche et de la preuve s'articule ici autour des trois pôles de la théorie, de l'expérience et de la méthode conditionnée à la technique.

Néanmoins, la tension entre la théorie (les concepts), l'expérience (le terrain) et la méthodologie s'articule plus rarement autour d'outillages élaborés. Le langage lui-même est souvent l'outil premier où se construisent des réflexions méthodologiques en sciences sociales. En France, l'intérêt constant porté à l'utilisation des différentes formes de technologie en sciences humaines s'est souvent constitué autour de la critique de méthodologies quantitatives très répandues outre atlantique - la statistique en sociologie, l'étude comportementaliste en psychologie ou le film ethnographique (Becker, 1996). Comme le note justement Latour, la quête de légitimité des Sciences Sociales s'est souvent traduit par une capacité à se procurer et traiter des données : "*Sociology has been obsessed by the*



**Graphique 3.1.1** – Triptyque des pratiques scientifiques d'après [Guichard \(2014\)](#)

*goal of becoming a quantitative science.*” ([Latour, 2003](#)). Pourtant, il serait illusoire de corréler la quantité de données à une quelconque objectivité de la démarche scientifique, tant les démarches et outils nécessaires pour la collecte sont en soi des biais importants.

Tim Berners-Lee, considéré comme l’inventeur du World Wide Web et du Web Sémantique définit la constitution d’un réseau mondial des savoirs comme un projet d’*“ingénierie philosophique”* ([Halpin et Monnin, 2014](#)). Mi-philosophe mi-ingénieur, le chercheur contribuant à ce réseau de savoir doit donc être en capacité de connaître les protocoles et les langages pour y accéder et s’y mouvoir aisément. La méthode scientifique s’écrit notamment avec de nouvelles formulations (fonctions, algorithmes, code...). Les données générées par les usages d’un nombre croissant de machines communicantes et productrices d’information, souvent désignées par le concept « valise » de Big Data ([Lohr, 2012b](#)) offrent des possibilités nouvelles pour l’étude en sciences sociales. L’analyse de ces vastes jeux de données s’accompagne également de nouveaux impératifs et questionnements sur l’observation des phénomènes humains qu’ils représentent. Á la fois barrière et opportunité, une difficulté majeure réside dans le discernement nécessaire entre praxis des outils informatiques, fascination pour ces outils et réflexions pertinentes sur la qualité des méthodes employées. Le traitement quantitatif par ordinateur permet d’extraire de nombreuses connaissances

utiles de jeux de données parfois très importants mais n'assure pas pour autant la qualité des résultats. Une bonne compréhension de la provenance et des méthodes de collection des données est nécessaire afin d'identifier des algorithmes de traitements intéressants, adaptés et efficaces parmi ceux disponibles (Rajaraman et Ullman, 2011).

Les méthodologies d'exploration et de recherche utilisant le "Big Data" comme source nécessitent la mise en œuvre d'une ingénierie complexe soutenue par une connaissance des technologies nécessaires à l'analyse de données. L'algorithmique, la statistique, l'informatique mais également la cartographie et le design graphique doivent se conjuguer pour permettre de produire des résultats à la fois intéressants et fiables. Ce travail d'hypothèses et de vérifications pour l'analyse de données doit donc se prémunir de nombreuses compétences. La définition de la problématique la plus adaptée nécessite une connaissance aigüe du terrain et des outils qui seront articulés au sein d'un système ingénierique parfois très complexe. Le design et la lecture d'algorithme pour le "*data mining*" sont donc une des clés pour le travail du chercheur confronté aux données. Néanmoins, ces algorithmes ne devraient être que la traduction de questions formulées grâce à une connaissance aigüe des problématiques du contexte et des objets étudiés - notamment pour identifier les données manquantes. La "science des données" promet donc d'apporter un véritable renouveau des méthodes et des résultats scientifiques, au prix d'un travail soutenu pour faire face à ces changements d'habitudes et de langages. Les applications statistiques du "big data" permettent aujourd'hui une fiabilité accrue des prédictions par l'augmentation du volume des corpus traités (Breiman, 2001). Le domaine de l'intelligence artificielle (AI) a grandement bénéficié de l'accroissement de la capacité de traitement des données notamment pour la prédiction grâce aux techniques dites de *machine learning*. Néanmoins Peter Norvig, directeur de recherche chez Google, reconnaît lui-même : "*We could draw this curve : as we gain more data, how much better does our system get ? And the answer is, it's still improving—but we are getting to the point where we get less benefit than we did in the past.*" (Somers, 2013). Comme le note Douglas Hofstadter, un des pairs de l'Intelligence Artificielle, à propos du super-ordinateur qui venait de battre Kasparov aux échecs : "*Okay, Deep Blue plays very good chess—so what ? Does that tell you something about how we play chess ? No. Does it tell you about how Kasparov envisions, understands a chessboard ?*" (Somers, 2013).

La question de la réussite des études utilisant les Big Data se corrèle avec la capa-

cité d'une approche méthodologique forgée dans un échange interdisciplinaire grâce à des pratiques renouvelées de l'écriture. Lors du traitement, la complexité des questions abordées rendent nécessaires la discussion entre de multiples connaissances méthodologiques, technologiques et éthiques notamment sur la légitimité, l'utilisation et la provenance des données (Boyd et Crawford, 2011). Aujourd'hui, l'utilisation de l'ordinateur conditionne l'écriture scientifique de l'étude de terrain à la prise de notes en passant par la rédaction ou la publication et se présente ainsi comme un impératif méthodologique pour la réflexion et le travail en sciences humaines et sociales (Wieviorka, 2013). L'appareillage projeté au centre de la pratique quotidienne du chercheur vient modifier le travail de réflexion sur les phénomènes étudiés et s'accompagne de multiples contraintes. La prise en main de ce nouvel "*instrument intellectuel*" (Guichard, 2014) passe par une lente alphabétisation aux langages des machines. L'angoisse latente du "bug" n'est pas apaisée par les rares techniciens présents dans les centres de recherche en sciences humaines, souvent déjà dépassés par la diversité des demandes technologiques. L'écriture, savoir-faire indispensable de la recherche, possède une nouvelle matérialité en disques "durs" produisant une dépendance accrue aux réseaux d'ordinateurs. Cette empreinte digitale laissée par les calculateurs sur les pratiques scientifiques n'est ni anodine ni révolutionnaire. Elle s'inscrit dans la longue tradition des cultures de l'écrit scientifique qui déjà bien avant les moines copistes "*combinent les gestes de la main et les opérations de la pensée.*" (Jacob, 2011)

Ces pratiques méthodologiques doivent trouver leur place dans la continuité historique des disciplines par une exigence de rigueur très scientifique. La portabilité des méthodes et la disponibilité des données sont encore des questions centrales et non-résolues. En sciences sociales notamment, les services de réseaux sociaux en ligne offrent de très larges corpus dont l'utilisation est régie par les exigences commerciales des sociétés privées qui les détiennent. L'analyse des données issues de service de réseaux sociaux en ligne est un champ d'études en rapide expansion (Nettleton, 2013). Pourtant, le débat sur la validité des éclairages apportés par l'analyse des données issues des services de réseaux reste encore largement ouvert et nous entendons dans ce travail de recherche y contribuer.

### 3.1.3 Visualisation et espace perceptif pour l'information

Face à de larges volumes de données, un des grands enjeux est d'en restituer une forme intelligible afin d'identifier des tendances ou des motifs particuliers. La visualisation permet de produire une lecture particulière de parties intéressantes et intelligibles d'un jeu de données (Cairo, 2013). Définie comme « *a process that transforms data, information, and knowledge into a form that relies on the human visual system to perceive its embedded information.* » (Graffieti *et al.*, 2010), la visualisation introduit la question du design visuel au cœur de la problématique d'analyse (Wesolowsky et Tufte, 1992). La visualisation correspond à une série d'actions aboutissant à la production de marqueurs visuels (points, ligne, aires, surface, volume), leurs propriétés rétinienne (couleur, taille, texture, etc.) et leur positionnement dans l'espace visuel (Card et Mackinlay, 1997). Dans son travail, Engelhardt (2007) s'attache à définir les bases d'une grammaire de la visualisation en commençant par en identifier les formes syntaxiques :

1. les objets graphiques montrés (ex. point, flèche, pictogramme, etc.),
2. l'espace graphique donnant sens à l'organisation des objets (ex. systèmes de coordonnées géographiques, timeline, etc.),
3. les propriétés graphiques des objets (couleurs, tailles, etc.),
4. l'organisation des objets en différentes catégories (ex. cadre, liens, légendes, etc.).

Ces choix forment une tâche importante dont l'enjeu n'est pas seulement visuel, mais se situe également dans le champ de la représentation où les objets sont donnés à voir et par là-même à comprendre. Les travaux sur les débuts de la perspective dans la Renaissance italienne ont montré comment l'espace de la représentation visuelle fait écho aux nouvelles questions qui traversent la société européenne de l'époque (Raynaud, 2005). Les œuvres des peintres du Quattrocento témoignent de changements profonds dans la perception : l'espace perceptif se structure désormais autour du sujet et de son "point de vue" qui construit l'ensemble de la représentation (Damisch, 1999). Empreinte de rationalité, la perspective construit comme au théâtre un espace de représentation centré autour du spectateur. En 1639, le mathématicien Desargues modélise les notions intuitives de perspective et d'horizon grâce à la géométrie projective qui permet d'étudier les propriétés inchangées des figures lors de leur projection. Cette géométrie d'un genre nouveau

se structure autour du *plan projectif*, élément topologique qui “*rassemble en une seule surface l’imagination de tous les points de vue possible*” (Petit, 1999). En construisant un plan géométrique fondé sur le point de vue, de nouveaux êtres géométriques aux propriétés étranges voient le jour, dont l’existence logique force notre représentation classique : la bande de Möbius ne possède qu’une seule face et il est impossible de distinguer l’intérieur de l’extérieur d’une bouteille de Klein. Dans ces surfaces dites *unilatères*, le local est traversé en tout point par un tout global. Nous ne pouvons pas traverser d’un bord à l’autre puisque nous sommes toujours sur la même face. Ni extrémité, ni extérieur, ni intérieur, le plan projectif apporte des éléments de réponses conceptuels aux limites de la représentation dans l’espace.

Dans le contexte de systèmes et de réseaux complexes, la visualisation de données structure l’espace perceptif afin de construire une scène à  $n$  dimensions dont l’enjeu est la recherche d’un “point de vue” pour l’étude. Alors que la perspective prend pour parti de matérialiser le sujet au centre de la représentation par des points de fuite, la visualisation de données cherche elle à utiliser les objets comme dimensions du champ de la représentation, en structurant souvent l’espace autour de quantités. Néanmoins, la place du spectateur / utilisateur dans la visualisation reste une des questions majeures encore à explorer. Le nouvel espace de représentation ouvert par la photographie a amorcé ce changement en proposant de nouvelles mises en scènes. En élaborant sa “méthode graphique”, Marey (1885) utilise la chrono-photographie pour créer une nouvelle représentation du mouvement et observer des phénomènes jusqu’ici invisibles. Le psychiatre Charcot cherche dans les images de l’hystérie des témoignages de la folie de ses patients (Didi-Huberman, 2012). Alors que les temporalités de l’écrit ou de la voix sont avant tout linéaires, l’espace visuel permet de manier le réel pour le décomposer en actes logiques (Renaud et Lafont, 2010).

La visualisation scientifique se prévaut donc d’une existence pratique dont le premier objectif serait de « *effectively convey information* » (Kelleher et Wagener, 2011). Son caractère synchrétique et sa capacité à résumer rapidement une large masse d’information en font un des plus importants éléments de la publication en science notamment pour sa diffusion et la facilitation d’accès à une connaissance (Ware, 2004). Dans sa sémiologie graphique, Bertin (1977) distingue deux usages majeurs des graphiques de visualisation : 1) un moyen de communiquer des informations (dans le cas où l’information a déjà été



comprise) 2) un moyen visuel de résoudre des problèmes logiques (quand le graphique est utilisé comme support de lecture et de manipulation d'informations). Ces deux caractéristiques peuvent coexister dans certaines pièces mais la transition entre les deux nécessite souvent un travail important de restructuration visuelle. Dans le traitement et la visualisation des données, *l'interface* joue un rôle primordial (Weissberg et Barboza, 2007). Elle permet d'agir et ainsi d'explorer le sens produit, prolongeant la deuxième catégorie de Bertin. Manovich définit l'interface comme “*the ways to represent ('format') and control the signal.*” (Manovich, 2013). Toujours selon Manovich, les changements induits par l'usage de l'interface dans la pratique de la lecture s'apparentent à de la reconnaissance de *pattern*, symbolisé par l'usage de l'icône et du menu. Si l'interface contraint la lecture, le design de ses formes narratives (les *patterns* de Manovich) sont au cœur de la conception de visualisation d'information. Les signes graphiques doivent refléter les usages par l'utilisateur afin de recréer la coopération textuelle des rôles de lecteur et de designer/auteur nécessaire pour la production du sens (Eco, 1985). La *citizen science* ou encore *night science* a fait de l'interface un paradigme. Utilisant la visualisation, elle amène un grand nombre de collaborateurs à explorer et analyser de vastes jeux de données en distribuant des tâches simples via une interface web (Silvertown, 2009). Le projet *Eyewire* permet à des internautes de contribuer à la classification d'images du cerveau humain en vue de la réalisation d'un modèle 3D (Seung, 2012). Utilisant des scans de tranches de 1mm réalisés par l'institut Max Planck, la modélisation 3D d'un cerveau complet promet une belle contribution pour la découverte des fonctionnements cognitifs. Néanmoins, la tâche est colossale et nécessiterait plusieurs années pour une équipe classique de scientifiques. *Eyewire* propose donc une interface web où un simple jeu de coloriage permet d'identifier les neurones et de contribuer ainsi au dessin du modèle 3D.

## 3.2 L'analyse de la diffusion sur les réseaux sociaux

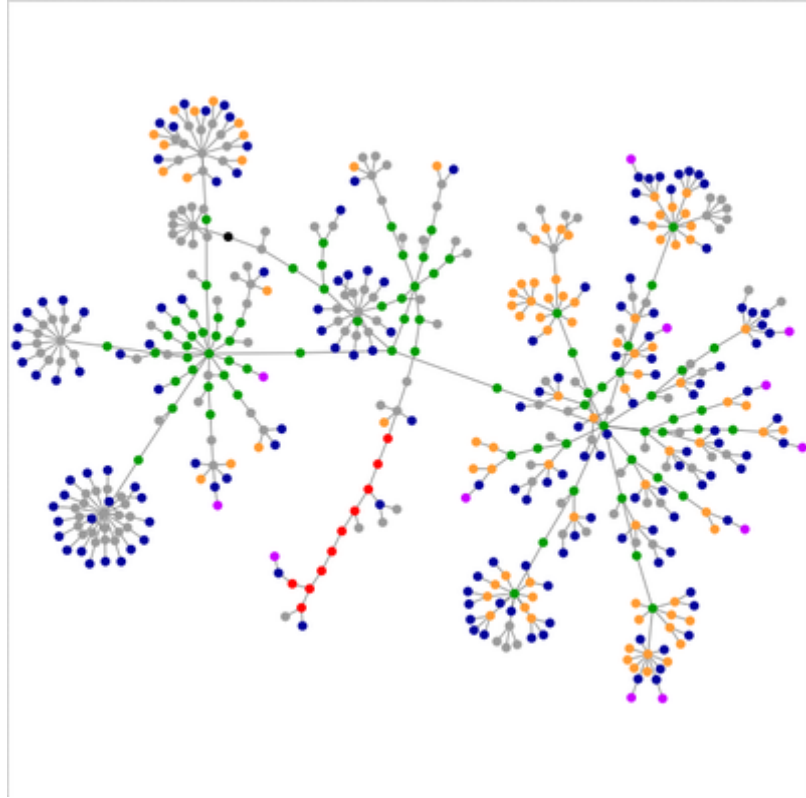
Ces nouvelles pratiques de l'analyse de données trouvent un vaste champ d'applications par l'analyse de réseaux sociaux, en anglais *Social Network Analysis* (ou *SNA*).

### 3.2.1 Anatomie d'un réseau social

La représentation des relations sociales sous forme de graphe trouve son origine dans les travaux des psychologues allemands de la *Gestalt* durant les années 1920-1930 (Scott, 1988). S'inspirant des études sur le cerveau, le psychologue J. L. Moreno s'applique notamment à comprendre les principes organisationnels holistiques des groupes humains et fonde la sociométrie avec comme objectif la qualification et la quantification des relations sociales (Moreno et Jennings, 1938). Moreno veut identifier et isoler des leaders de groupes sociaux définis en étudiant l'asymétrie ou la réciprocité de leurs choix et fréquentations amicales. Cherchant des moyens de représenter les tendances à l'auto-organisation qu'il observe, il cartographie les relations directes et indirectes entre personnes sous forme de *sociogrammes*. Les anthropologues s'emparent rapidement de ce type d'outils pour comprendre les formes tribales (Lundberg, 1975). L'émergence progressive de la topologie comme domaine important des mathématiques vient définir de nouveaux types de relations entre objets disparates, avec notamment la théorie des graphes qui donne à l'étude des réseaux ses modèles logiques (Harary, 1977). Travers et Milgram (1969) voit les relations humaines comme autant de petits mondes (*small-worlds*) connectés entre eux. Granovetter s'intéresse à l'importance des relations ténues et lointaines (*weak ties*) dans l'acquisition d'informations importantes (Granovetter, 1973). L'influence de la théorie des graphes amène notamment les sociologues à expérimenter de nouvelles pratiques venues de la physique ou de la biologie, comme celle de la simulation sociale (Epstein et Axtell, 1996).

La matérialité de l'image du *graphe* structure la représentation du réseau social. Dans la littérature concernant les réseaux, les notions de graphe et de réseau sont interdépendantes et la théorie des graphes sert notamment de système de notation pour la mise en équation des réseaux (Nettleton, 2013). Si cette structure point-ligne semble toutefois revêtir une limite de taille pour la description de phénomènes humains (comment en effet réduire les relations humaines à une simple ligne?), elle semble néanmoins aujourd'hui encore difficile à dépasser.

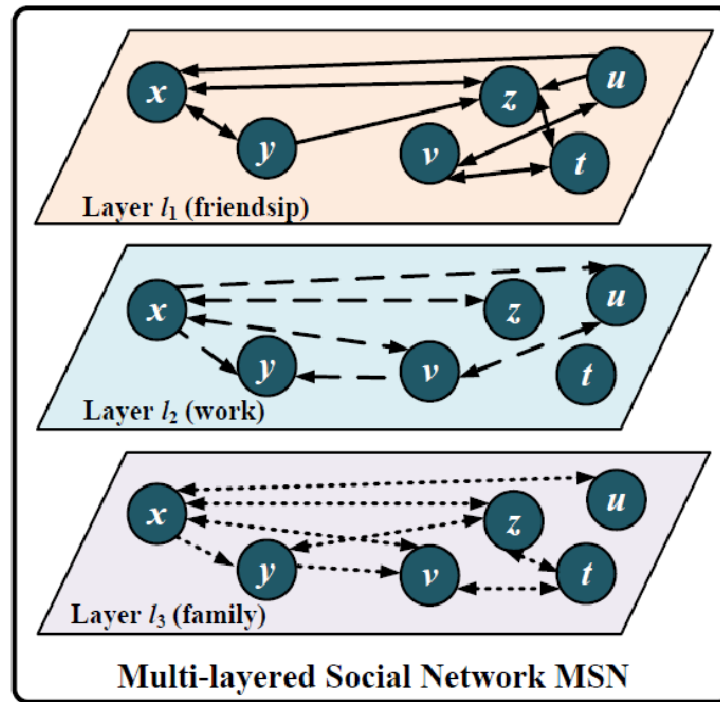
Un réseau considéré comme graphe, noté  $G$ , se compose d'un ensemble de nodes ou vertices (les points) et de liens ou edges (les traits). On représente ainsi un graphe sous la notation  $G(V,E)$  où  $V$  est l'ensemble des nodes du réseau et  $E$  l'ensemble des liens



**Graphique 3.2.1** – Représentation d'un réseau sous forme de graphes.

décrivant leurs relations. Les relations décrites dans  $E$  peuvent être directionnelles (paires de vertices ordonnées) dans le cas d'un graphe *orienté* ou *pondéré* de valeurs particulières. Les relations ainsi exprimées portent sur un aspect unique, quantifiable et isolable. La prise en compte de facteurs multiples, comme notamment l'espace physique, le temps, mais également les multiples réseaux de relations qui peuvent exister entre deux acteurs nous amène à considérer un graphe disposant de multiples couches (*multi-layered*) pour décrire l'ensemble des groupes de relations. Imaginons un graphe de personnes  $G(V, E_n)$  où  $V$  sont les vertices représentant les personnes et  $E_n$  un nombre  $n$  d'ensemble de liens décrivant chaque type de relations spécifiques. La figure 3.2.2 montre un exemple d'un graphe multi-couches où  $E_1$  est l'ensemble des relations amicales,  $E_2$  les relations de travail et  $E_3$  les relations familiales.

Nous pouvons ainsi décrire différents jeux de relations entre un jeu d'acteurs fini, permettant notamment de mieux comprendre les relations entre ses différentes dimensions. Si cette approche résout momentanément la question des  $n$  dimensions à considérer (Bródka *et al.*, 2013), elle augmente également la complexité du graphe et la possibilité d'erreurs de lecture ou de typage des relations.



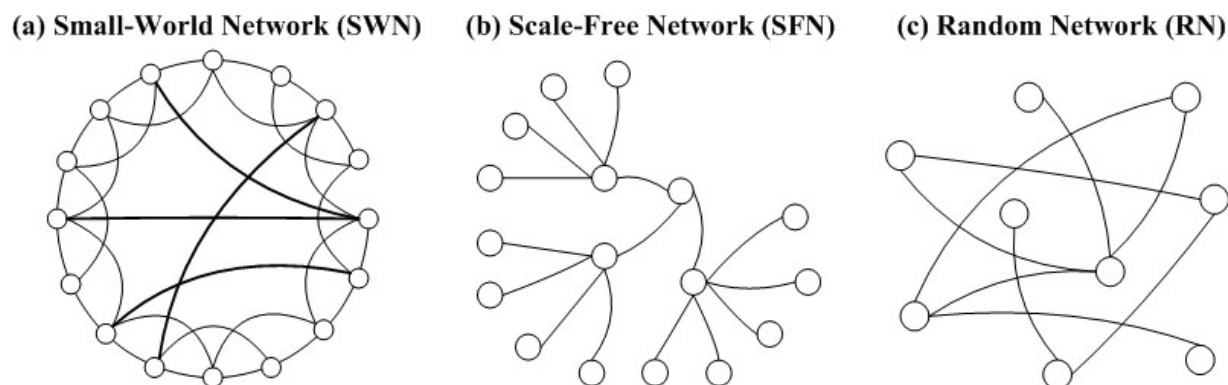
**Graphique 3.2.2** – Un réseau social multi-couches, d'après Bródka *et al.* (2013)

Afin de mieux comprendre l'organisation d'un réseau, nous disposons de plusieurs mesures pour décrire les relations et le rôle des différents acteurs :

- *Degree* (degré ou *valence*) mesure le nombre de connections d'un nœud dans un réseau. Cette valeur indique souvent une possibilité, le potentiel d'un node donné à interagir avec d'autres.
- *Closeness* (proximité) mesure la facilité d'un node à se connecter à un autre. Dans un réseau en ligne, on calcule la proximité en estimant la distance la plus courte entre un node et un autre.
- *Betweenness* (*centralité*) mesure le degré d'importance d'un node dans le réseau en prenant en compte le nombre de nodes dépendant de lui pour établir une connexion entre eux. La centralité représente la capacité à bloquer ou laisser filtrer l'accès à certaines parties du réseau.

En observant ces différentes mesures, nous pouvons définir différentes structures types de réseau. La distribution des degrés dans le graphe permet notamment de comprendre les modèles qui régissent les connexions entre les nodes. L'exemple le plus simple est le *random network*, réseau où les acteurs sont connectés de manière entièrement aléatoire. Un réseau dont le degré de distribution correspond à une loi de puissance est appelé *scale-*

*free network*. Un réseau dont seulement quelques nodes possèdent une centralité élevée et dont la structure d'ensemble est constituée de groupes ou *clusters* interconnectés est appelé un *small-world network*.



**Graphique 3.2.3 – Types communs de réseaux**

Les services de réseaux sociaux en ligne sont structurés en *small-worlds*. Dans une étude analysant de larges corpus issus de différents services (Yahoo!, Flickr, etc.), [Kumar et al. \(2006\)](#) ont montré que ces communautés possèdent toutes une structure et une évolution similaires. Les inscrits de chaque service se répartissent autour de trois grands groupes : *singletons* (membres isolés, inactifs), *giant component* (la majorité des utilisateurs actifs) et *middle region* (des communautés isolées qui interagissent entre elles mais pas avec le reste du réseau). D'après les auteurs, il existe très peu de chances que deux communautés isolées même très similaires se rencontrent dans ce type de réseau, car l'entropie de la structure *small-world* se renforce avec le temps en se fragmentant davantage. Une des raisons principales est que ces communautés isolées suivent souvent un modèle en *étoile*, c'est à dire qu'elles sont construites autour d'un individu central charismatique. Il leur arrive de rejoindre la masse du *giant component* mais elles en deviennent difficilement des acteurs majeurs et restent en périphérie. Dans un réseau social de type small-world comme les OSNS, les acteurs les plus influents sont ceux qui sont capables de 1) renforcer les liens dans un cluster (*closure*) et 2) développer les connections faibles entre des clusters (*brokerage*) ([Burt et al., 2008](#)). Ce "capital social" est inscrit dans la structure du réseau social lui-même ([Lin, 1999](#)) et n'a pas de relations avec le degré du node (son nombre de connections) ([Cha et al., 2010](#)). L'individu le plus puissant d'un réseau est donc celui possédant le plus grand nombre de connections *potentielles*, proches et facilement accessibles et qui bénéficie d'une place privilégiée pour bloquer ou autoriser l'accès aux autres parties du réseau. L'analyse organisationnelle a notamment montré l'importance capitale

des secrétaires pour le maintien d’une bonne circulation des informations dans le réseau de l’entreprise ou la nécessité d’un nombre très faible de connections directes pour les acteurs importants de réseaux terroristes ou mafieux (Russel, 2011).

### 3.2.2 Analyse des réseaux sociaux et de la diffusion d’information en ligne

L’analyse de réseau social permet de comprendre la structure d’un moment du réseau. En effet, les connections entre différents acteurs sont sans cesse en mouvement et l’information en circulant vient modifier l’activité du réseau et bien souvent sa structure même. Notre étude porte sur la diffusion de mèmes au sein du réseau social chinois *Sina Weibo* et s’inscrit ainsi dans la continuité des travaux s’intéressant aux rapports entre diffusion et technologies. Le champ de la diffusion de l’innovation, objet traditionnel de la géographie, s’est largement constitué autour des problématiques technologiques et organisationnelles liées à la situation spatiale du lieu et ses équipements (Crevoisier, 2004). Les études urbaines ont notamment montré comment la réalité économique et géographique présidait à la constitution des savoirs nécessaires à la diffusion de l’innovation technologique (Howells, 2002). Plus largement, les études sur la diffusion sont dominées par l’analogie du virus comme modèle de propagation des messages, comportements et idées. Depuis la fin du XIXe siècle (Le Bon, 1895), ce modèle est prépondérant dans les recherches autour de la diffusion d’information en ligne (Goel *et al.*, 2012). Les membres de groupes dans un réseau seraient *exposés* à un message ou une idée avant d’être *infectés*, devenant alors porteurs puis agents de sa diffusion. Ainsi, en considérant la position d’un individu au sein du réseau de diffusion, il serait possible de définir un “*degré d’infection*” (Cheng *et al.*, 2013) et d’anticiper la diffusion selon une « *probabilité immune* » qui déciderait de la « *qualité infectieuse* » de l’objet diffusé ou de la « *possibilité de rétablissement* » du sujet infecté (Wang et Wood, 2011). Cette analogie du viral propose une vision mécaniste qui fait peu de cas des facteurs contextuels ou psychologiques et ignore les processus de décisions individuelles en jeu (Jackson et Yariv, 2010). Cette vision doit être modérée par une considération plus importante des modalités d’appropriation des technologies comme pratiques propres aux territoires (Fernandez et Puel, 2010). Nous savons notamment que le choix des mots ou la situation des personnes diffusant un message sont des facteurs

décisifs et structurant des processus d'adoption des messages en ligne (Conover *et al.*, 2013). En s'intéressant à la propagation géographique de mots sur les réseaux sociaux, Eisenstein *et al.* (2012) observe que leur diffusion se limite à un domaine géographiquement bien défini, dépendant de facteurs culturels et démographiques. Par exemple, les villes ayant d'importantes communautés afro-américaines ont davantage de chances d'adopter un même mot que d'autres parfois plus proches géographiquement. Cette question est au centre des études de marketing qui s'interrogent notamment sur l'adoption de nouvelles marques ou de slogans. Afin de déterminer statistiquement les possibilités d'adoption de produits et prévoir la pénétration dans un marché précis, la modélisation mathématique des effets de réseaux dans la diffusion est souvent utilisée (Bass *et al.*, 1994). Ici, l'analyse des données des réseaux sociaux devient un grand enjeu pour la prospective économique. L'application de ces modèles mathématiques aux données produites par les utilisateurs permet de simuler des comportements de segments précis de marché. Le marketing politique fait lui aussi grand cas de l'analyse de réseaux sociaux pour comprendre et orienter les discussions. La campagne de réélection d'Obama aux USA en 2012 a fait un usage extensif de l'analyse de données des réseaux sociaux pour identifier, déterminer et cibler des groupes sociaux particuliers grâce au travail d'une vaste équipe d'ingénieurs et de “data scientists”<sup>2</sup>.

Une des grandes interrogations dans ce domaine porte sur les rôles joués par les différents acteurs du processus de diffusion - et la manière de les identifier. Les études concernant les leaders d'opinion, traditionnelles en sciences de la communication (Katz et Lazarsfeld, 1955) trouvent une continuité directe dans l'étude des réseaux sociaux en ligne avec le domaine florissant des recherches sur l'identification des *influenceurs* (Bakshy *et al.*, 2011; Leavitt *et al.*, 2009). L'analyse quantitative permet notamment de mieux comprendre l'influence réelle des acteurs dans le réseau grâce à l'étude de leurs comportements. Le concept d'influence sur les réseaux sociaux revêt en réalité des formes très variables et procède notamment d'une légitimité construite autour de sujets précis par des personnes spécialisées devenues référentes (Cha *et al.*, 2010). D'autres “influenceurs” possèdent une grande capacité d'amplification pouvant par exemple initier le développement

---

2. “Harper Reed, the chief technology officer for the Obama re-election campaign, who heads a team described as “100 data scientists, developers, engineers, analysts, and old-school hackers [that] have been transforming the way politicians acquire data—and what they do with it.”, from The Blaze <http://www.theblaze.com/stories/2012/10/03/very-creepy-details-of-obama-campaigns-voter-data-mining-effort/> consulté le 12 Mars 2014 à 14 :50 GMT+1

d'une "*masse critique*" autour d'une information, définie classiquement comme le seuil d'adoption à partir duquel la diffusion devient pérenne parmi une foule d'acteurs (Oliver et Marwell, 2001). L'image quantitative d'une "masse" uniforme et actionnable dans le réseau est néanmoins remise en question par l'étude de données. L'analyse montre que les relations entre différents acteurs du réseau sont à considérer qualitativement en termes de relations de pouvoirs (Steyer et al., 2006). Les dynamiques d'échanges ne répondent en effet pas tant à une relation pré-existante dans le réseau qu'à un ensemble de situations où les acteurs adoptent des comportements et réactions particulières. La diffusion peut ainsi se comprendre comme une pratique du *bouche-à-oreille* éminemment contextuelle, où certains acteurs sont plus ou moins influents dans telle ou telle situation ou sur tel et tel sujet. Le temps joue néanmoins un rôle déterminant puisqu'en considérant des séries de résultats où des acteurs se côtoient durant plusieurs années, il est souvent difficile d'identifier quel acteur influence l'autre (Aral et al., 2009).

Après moins d'un siècle d'existence, l'analyse de réseaux sociaux sous forme de graphes a donc connu une rapide évolution. Les services de réseaux sociaux en ligne offrent notamment la possibilité d'obtenir des données sur les comportements de groupes sociaux en très vaste quantité. Véritable vivier d'études, ce champ de recherche en pleine expansion trouve ses origines dans des disciplines diverses qui poursuivent souvent des objectifs et des méthodes variées. Le tableau infra présente quelques-unes des méthodes d'analyse de données définies selon leurs domaines d'application. Ces méthodes coexistent souvent lors d'études utilisant le *data mining*; nous présentons ici des exemples qui permettront ensuite de mieux situer nos perspectives de recherche dans ce paysage de pratiques.



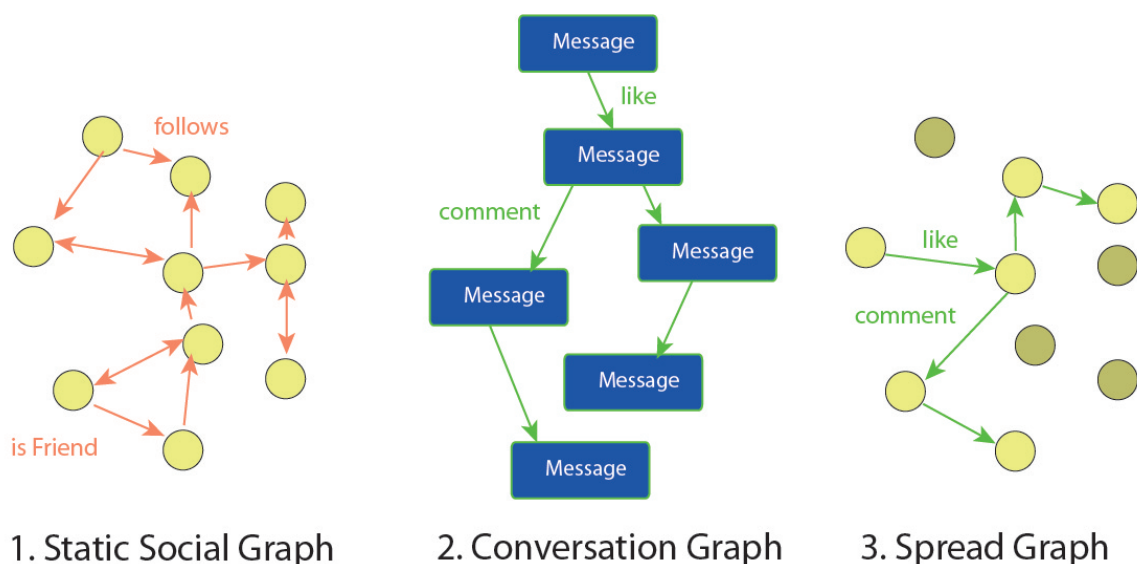
Courant d'analyse	Méthodologie	Exemples d'usages et applications	Auteurs & Publications de référence
1	Graphes sociaux de groupe définis (cartographie de réseau fini)	En partant d'un échantillon fini et déterminé au préalable, on effectue une carte des relations entre les différents acteurs du réseau.	Cette pratique est l'origine de la sociométrie ( <a href="#">Moreno et Jennings, 1938</a> ) et de l'étude des relations sociales sous forme de réseaux.
2	Découverte de groupes par critères (communautés)	Cette méthode permet de réaliser un échantillonnage d'un réseau social à partir d'un ensemble de profils existants appelés <i>seeds</i> . Un logiciel (appelé <i>crawler</i> ) va rechercher et collecter les profils similaires aux <i>seeds</i> selon des critères définis : similarité, différence, profondeur, etc.	Héritage de la tradition de l'échantillonnage " <i>boule de neige</i> " en statistiques ( <a href="#">Rothenberg, 1995</a> ), les algorithmes de <i>crawling</i> sont nombreux et il n'existe pas de véritable consensus sur leur utilisation ( <a href="#">Gjoka et Kurant, 2011</a> )
3	Analyse sémantique des conversations (analyse de contenu)	En utilisant une masse textuelle de posts, un système est chargé d'extraire et de classer les mots et sujets discutés. Ce type de système se fonde sur l'analyse naturelle de langage (NLP) et parfois sur l'analyse structurelle des conversations ( <a href="#">Karandikar, 2010</a> ).	A mi-chemin entre linguistique et informatique ( <a href="#">Russel, 2011</a> ), ce champ est un des grands enjeux actuels du Big Data ( <a href="#">Nettleton, 2013</a> ) avec notamment l'analyse des sentiments ( <a href="#">Pang et Lee, 2008</a> )
4	Analyse de la diffusion (évolution des relations d'après une conversation)	D'après une masse de posts extraites selon des critères précis (souvent un mot-clé ou <i>hashtag</i> ), il s'agit ici de retracer les dynamiques relationnelles qui entourent ou suscitent la conversation en recréant le graphe social de la discussion et son évolution.	Le modèle classique du <i>word-of-mouth</i> ( <a href="#">Steyer et al., 2006</a> ) et l'approche épidémiologique ( <a href="#">Wang et Wood, 2011</a> ) sont souvent utilisés dans l'analyse de la diffusion de contenus en ligne ( <a href="#">Cheng et al., 2013</a> ).

**Graphique 3.2.4** – Tableau récapitulatif de méthodes d'analyse de données de réseaux sociaux en ligne.

Courant d'analyse	Méthodologie	Exemples d'usages et applications	Auteurs & Publications de référence
5 Analyse comportementale et agents de diffusion (classifications et mesures)	Ici on étudie l'activité d'un ou plusieurs agents en analysant leur comportement dans le réseau (volume d'activité, fréquence, etc.) souvent pour définir des critères et mesures qui permettent de classifier par type ou d'anticiper les actions d'un acteur du réseau.	Détection d'influenceurs Mesures de la probabilité de diffusion ( <a href="#">Anagnostopoulos et al., 2012</a> ) Typologie des utilisateurs par comportement Effet psychologique des signaux sur un utilisateur	Présente notamment dans le champ du marketing ( <a href="#">Leskovec et al., 2005</a> ) et de la politique ( <a href="#">Lotan et al., 2011</a> ), ce type d'analyse cherche à comprendre et retracer les processus parfois psychologiques ( <a href="#">Robins, 2013</a> ) de prise de décision individuelle.
6 Analyse contextuelle et géographique	Ce type d'analyse cherche à mettre en perspective des facteurs extérieurs au réseau afin d'en comprendre l'influence et les effets. Entre sciences sociales et informatique, ce type d'étude porte souvent sur l'usage du réseau plus que sur sa structure ou son évolution ( <a href="#">Torrens, 2010</a> ; <a href="#">Leetaru et al., 2013</a> )	Approche sociologique des usages Mesures d'impact géographique ( <i>node locality, geographic clustering coefficient</i> )	L'approche contextuelle dans l'analyse de réseaux reste encore un champ à développer ( <a href="#">Adams et al., 2012</a> ), notamment dans la considération de facteurs géographiques ( <a href="#">Graham, 1998</a> ; <a href="#">Onnela et al., 2011</a> ), culturels ( <a href="#">Gallagher et Savage, 2013</a> ) ou de langage.
7 Simulation sociale	Afin de comprendre les dynamiques en l'absence de données ou pour anticiper une situation à venir, il est possible de modéliser un environnement virtuel où les comportements des acteurs sont simulés ( <a href="#">Macy et Willer, 2002</a> )	Prévision de tendances d'après des données existantes Analyse de faits dont les données sont manquantes ou incomplètes	La découverte de méthodes de modélisation du contexte de l'univers de simulation ( <a href="#">Ronald et al., 2012</a> ) est un des grands enjeux où l'apport de méthodes ethnographiques de terrain peut être crucial ( <a href="#">Tubaro et Casilli, 2010</a> )

### 3.2.3 Modèles pour la visualisation des conversations

Notre recherche s'appuie sur l'étude du graphe social de la diffusion entourant les conversations (cf. tableau 3.2.4-4) et l'analyse sémantique des sujets dominants et sous-jacents aux conversations (cf. tableau 3.2.4-3). Nous souhaitons également amener une plus forte contextualisation des usages (cf. tableau 3.2.4-6) en prenant en compte l'existence spatiale et temporelle des conversations. L'étude des données des SNS par les géographes notamment s'est jusqu'ici largement focalisée sur l'analyse des *geotag* et des cartes en ligne comme principale approche méthodologique (Graham et Zook, 2011; Poorthuis et Zook, 2013), réduisant considérablement le champ de ce type d'études (Crampton *et al.*, 2013).

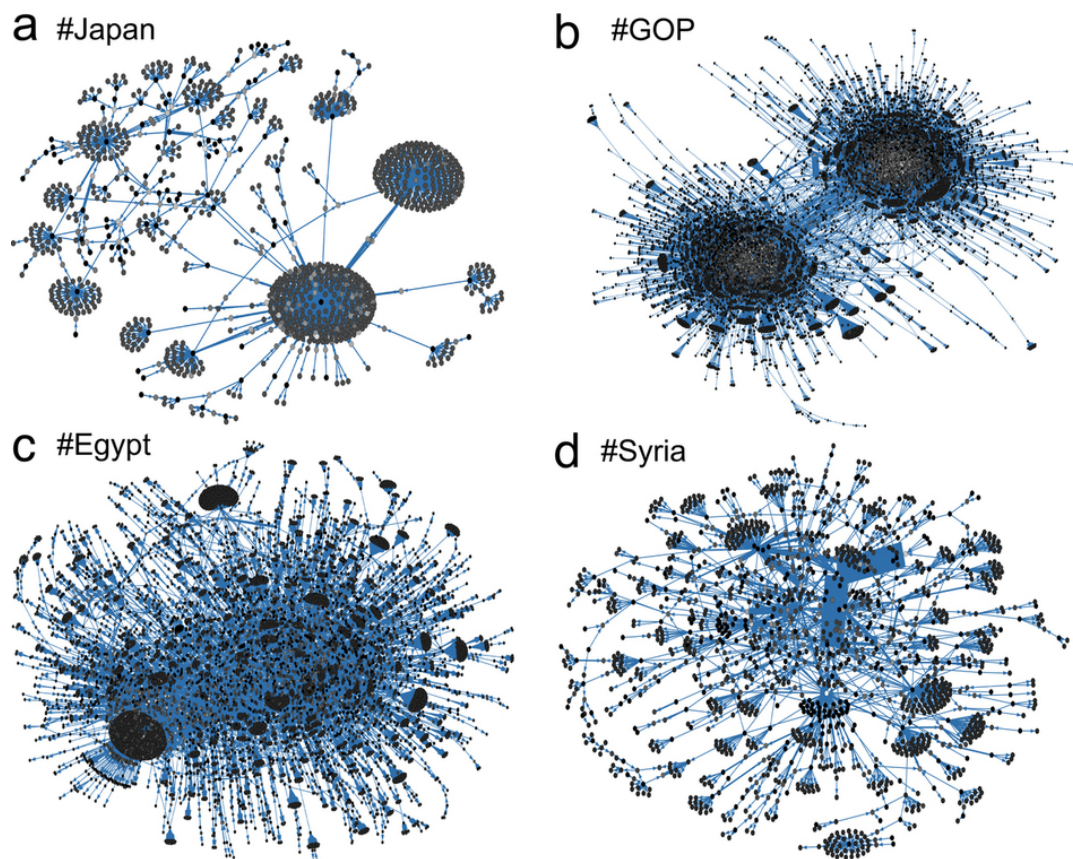


**Graphique 3.2.5** – Les 3 types de graphe classiquement extraits des données de réseaux sociaux

Plusieurs types de graphes représentant différentes relations existantes sur les réseaux sociaux peuvent être extraits :

- **Graphe social statique** : les relations pré-existantes dans la structure du réseau étudié (untel est ami avec untel, untel suit untel, etc.)
- **Graphe conversationnel** : toutes les interactions qui entourent et structurent la diffusion de messages (commentaires, RT, etc.)
- **Graphe de diffusion** : les interactions qui se sont produites entre les acteurs durant la diffusion du message. Ce dernier type de graphe est un recoupement des deux autres.

Dans notre étude, le graphe social statique ne présente pas particulièrement d'intérêt, représentant des relations pas forcément affectées par les discussions. De plus, nous ne disposons dans le jeu de données Weiboscope que de son état final qui ne témoigne pas de l'évolution des relations. Nous voulons obtenir ici les graphes de diffusion sous la forme de conversations structurées (graphe directionnel des réponses et commentaires) de l'ensemble des messages. Dans un article paru dans *Nature* (Weng *et al.*, 2012), des chercheurs de l'Université d'Indiana identifient les caractéristiques des mêmes connaissant le plus de succès (la plus large diffusion). La visualisation des motifs particuliers leur permet de mettre à jour la structure des conversations identifiées par des hashtags sur *Twitter* (voir fig. 3.2.6).



Nodes represent Twitter users, and directed edges represent retweeted posts that carry the meme. The brightness of a node indicates the activity (number of retweets) of a user, and the weight of an edge reflects the number of retweets between two users.

- (a) The *#Japan* meme shows how news about the March 2011 earthquake propagated.
- (b) The *#GOP* tag stands for the US Republican Party and as many political memes, displays a strong polarization between people with opposing views.

Memes related to the Arab Spring and in particular the 2011 uprisings in (c) *#Egypt* and (d) *#Syria* display characteristic hub users and strong connections, respectively.

**Graphique 3.2.6** – Graphe de diffusion de hashtags sur Twitter d'après (Weng *et al.*, 2012)

Notre travail préliminaire regroupe ici plusieurs expérimentations sur la visualisation de graphes conversationnels. Après avoir identifié quelques-uns des hashtags les plus discutés de l'année 2012 sur *Sina Weibo* (voir la section 3.3.3), nous avons procédé à la visualisation des échanges pour chacun d'entre eux. Afin de mettre à jour le graphe conversationnel entourant les hashtags sélectionnés, nous avons choisi d'extraire la séquence d'interactions des messages (mentions, retweets) - voir algorithme 1.

---

**Algorithm 1** Hashtags conversational graphs

---

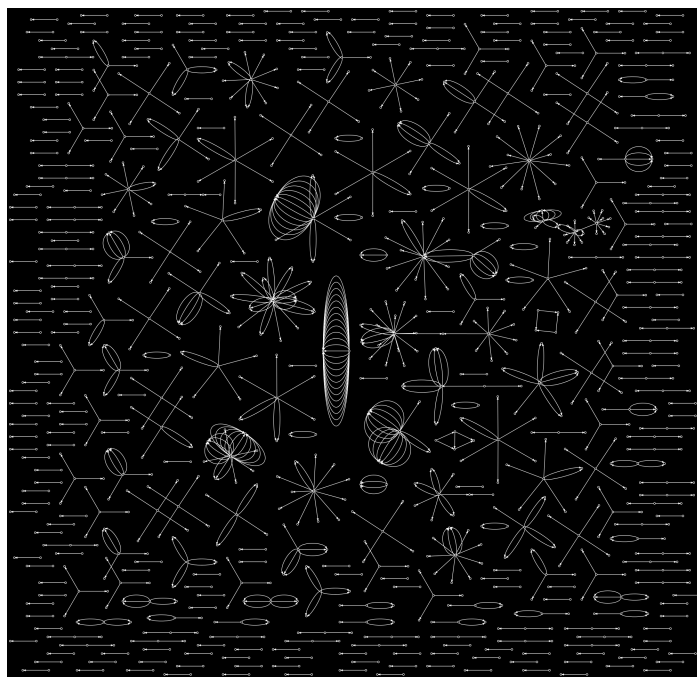
**Require:**  $M$  is a set of microblog messages  
 $H(h, Gh)$  is a set of all hashtags conversational graphs  
**function** HASHTAGSGRAPH( $M$ )  
  **for** message  $m$  in  $M$  **do**  
    **if** hashtag  $h$  in  $m$  **then**  
       $Gh = (Vh, Eh)$  is  $h$  conversational graph Graph(Vertex,Edges)  
      **if** quotes or rt  $e(userA, userB)$  in  $m$  **then**  
         $Eh \leftarrow e(userA, userB)$  and  $Vh \leftarrow userA, userB$   
      **end if**  
       $H \leftarrow (h, Gh)$   
    **end if**  
  **end for**  
**end function**

---

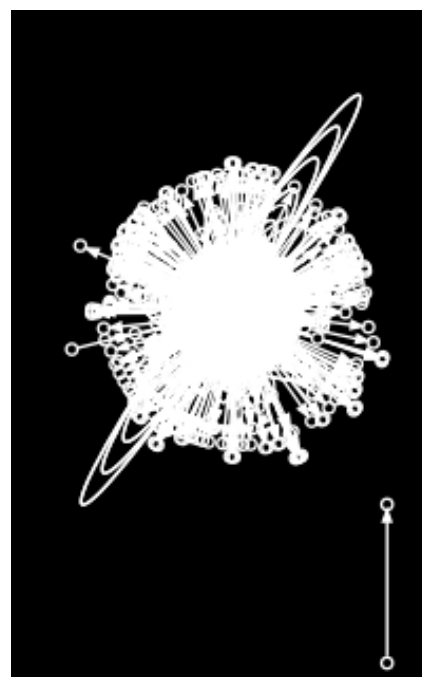
La diffusion de chaque mème peut être représentée sous la forme d'un graphe contenant un node par utilisateur et un edge par échanges visibles dans les textes des messages. Dans un premier temps, le logiciel *Graphviz* nous a permis d'obtenir une représentation basique du graphe conversationnel afin d'avoir un aperçu sur la nature des conversations par l'observation des motifs qui la composent (fig. 3.2.7). Cette première visualisation nous permet de considérer la nature des échanges et l'implication des utilisateurs d'après la structure des motifs conversationnels. Chaque utilisateur est symbolisé par un point et chaque message par un trait reliant deux utilisateurs. Un motif très compact reflète une conversation animée entre des utilisateurs peu nombreux échangeant beaucoup. A l'inverse, un motif disparate reflète des échanges plus brefs et morcelés.

*WeicoPlus* est une application mobile permettant d'utiliser *Sina Weibo*. Le hashtag *#WeicoPlus#* est ajouté automatiquement quand les utilisateurs postent des photos depuis ce service. On remarque que son graphe conversationnel (fig. 3.2.7a) se compose essentiellement de messages simples, mais ne donne pas lieu à une conversation structurée - à l'exception de quelques rapides échanges entre un nombre réduit de personnes. A





(a) Hashtag “WeicoPlus”



(b) Hashtag “Veuve d'enfant unique”

**Graphique 3.2.7** – Visualisation simple des conversations de deux hashtags

l'inverse, le hashtag “Veuve d'enfant unique” #失独母亲# (fig. 3.2.7b) cristallise le débat en une forme très dense qui reflète une surenchère de commentaires et d'actions autour du hashtag, propre d'une conversation animée.

Ces premières visualisations schématiques et peu détaillées montrent comment il est possible d'observer les dynamiques conversationnelles autour d'un hashtag. Nous allons maintenant regarder plus en détail trois exemples de hashtags dont les graphes présentent des modèles de diffusion organisés mais dissemblables. Nous utilisons le logiciel *Gephi* (Bastian *et al.*, 2009) pour examiner de plus près la composition des graphes en manipulant les vertices et les liens pour identifier leurs composantes.

Chaque utilisateur est représenté sous la forme d'un point. La taille des points correspond à l'importance de l'utilisateur dans le réseau total des échanges, caractérisée par une mesure topologique correspondant au nombre de plus courts chemins du graphe passant par cet utilisateur appelée centralité intermédiaire (*betweenness centrality*). Les couleurs représentent la *modularité* du réseau, soit le nombre de communautés engagées dans la conversation définies comme les cliques d'utilisateurs constituant plus de 1% du réseau total d'échange (Blondel *et al.*, 2008). La position des nodes est calculée grâce à l'algorithme

*Force Atlas 2* (Bastian *et al.*, 2009). Grâce à cette modélisation physique, les nodes peu connectés entre eux se repoussent et ceux très connectés s’attirent. Ainsi, la proximité de deux nodes sur le graphe témoigne de l’intensité des échanges dans la conversation (citations, commentaires ou retweets). Pour davantage de visibilité, certaines conversations subalternes représentant moins de 1% du total ont été effacées. Également, les nodes possédant un degré inférieur à 3 ne sont pas représentés (moins d’un échange avec au moins trois autres nodes du graphe).

### Exemple 1 : Pluie torrentielle à Tianjing

Le premier même choisi parle d’une catastrophe naturelle, sous la forme d’une pluie diluvienne qui s’est abattue sur la ville de Tianjin durant la nuit du 21 au 22 Juillet 2012.



**Graphique 3.2.8** – Exemple 1 : Tianjin Baoyu

Sur la figure 3.2.8, on voit nettement 4 groupes constitués autour de gros diffuseurs (les nodes les plus gros) qui composent 85% du graphe. Plusieurs groupes semblent s’emparer de la conversation. L’activité entre les nodes se déroule autour de quelques utilisateurs très centraux. Cela traduit le fait que peu de personnes ont réellement discuté l’information et se sont simplement contentées de la relayer.

La diffusion d’un fait divers très local (il se passe à Tianjing) est entraînée par peu de sources très importantes (les quelques nodes de grande taille), vraisemblablement des journaux et médias locaux qui annoncent la nouvelle (presse, photos-choc d’inondations, etc). La conversation est peu active et très structurée, nous sommes en présence d’un modèle classique de diffusion de masse.



**Graphique 3.2.9** – *Downpour bypasses Beijing, batters neighbor*, in Qinghua News le 2012-07-26 13 :29 :59, [http://news.xinhuanet.com/english/china/2012-07/26/c\\_131740415.htm](http://news.xinhuanet.com/english/china/2012-07/26/c_131740415.htm) consulté le 27 Juin 2014.

## Exemple 2 : Veuve de l’enfant unique

Un autre sujet discuté par une très large quantité de personnes apparaît sous le terme "*shidu muqin*", forme contractée signifiant “*mère qui a perdu son enfant unique*” (#失独母亲#). Ce hashtag désigne un phénomène de société bien connu en Chine où le deuil de la perte d’un enfant se double souvent pour une mère chinoise seule de l’absence de ressources pour vivre. En effet, l’absence de système de retraite fait porter aux enfants la responsabilité de la survie de la famille.

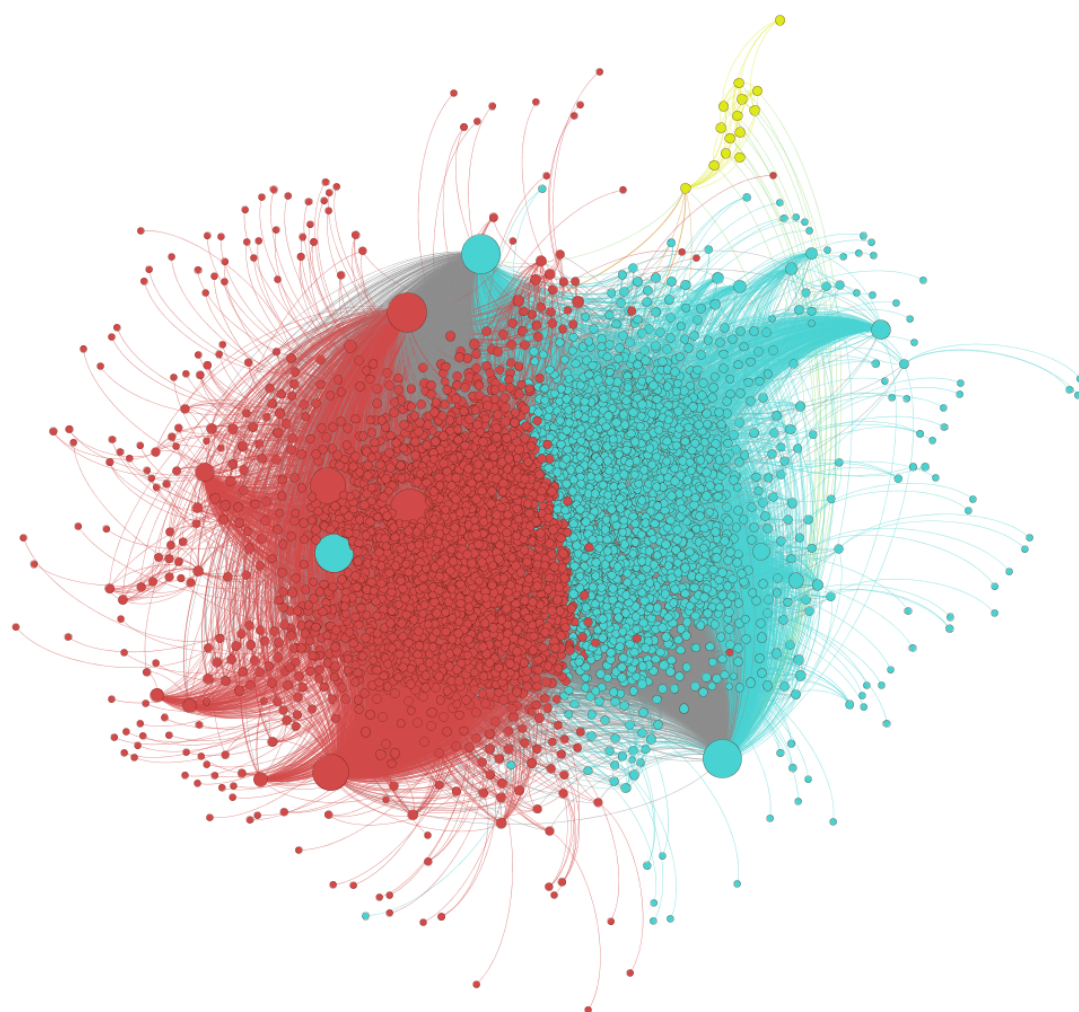




**Graphique 3.2.10** – 一位失独母亲的独语 paru le 17 Juillet 2012 in *Nandu Weekly* le 2012.07.17, <http://www.nbweekly.com/news/special/201207/30571.aspx>, consulté le 27 Juin 2014.

Le graphe 3.2.11 se compose de deux grands groupes regroupant à eux deux près de 95% du graphe total. Les discussions sont très polarisées et menées par peu de participants (les nodes les plus gros sur le graphe). Peu de personnes très influentes concentrent les discussions autour d'eux, accompagnées ou suivies d'une foule de commentateurs.

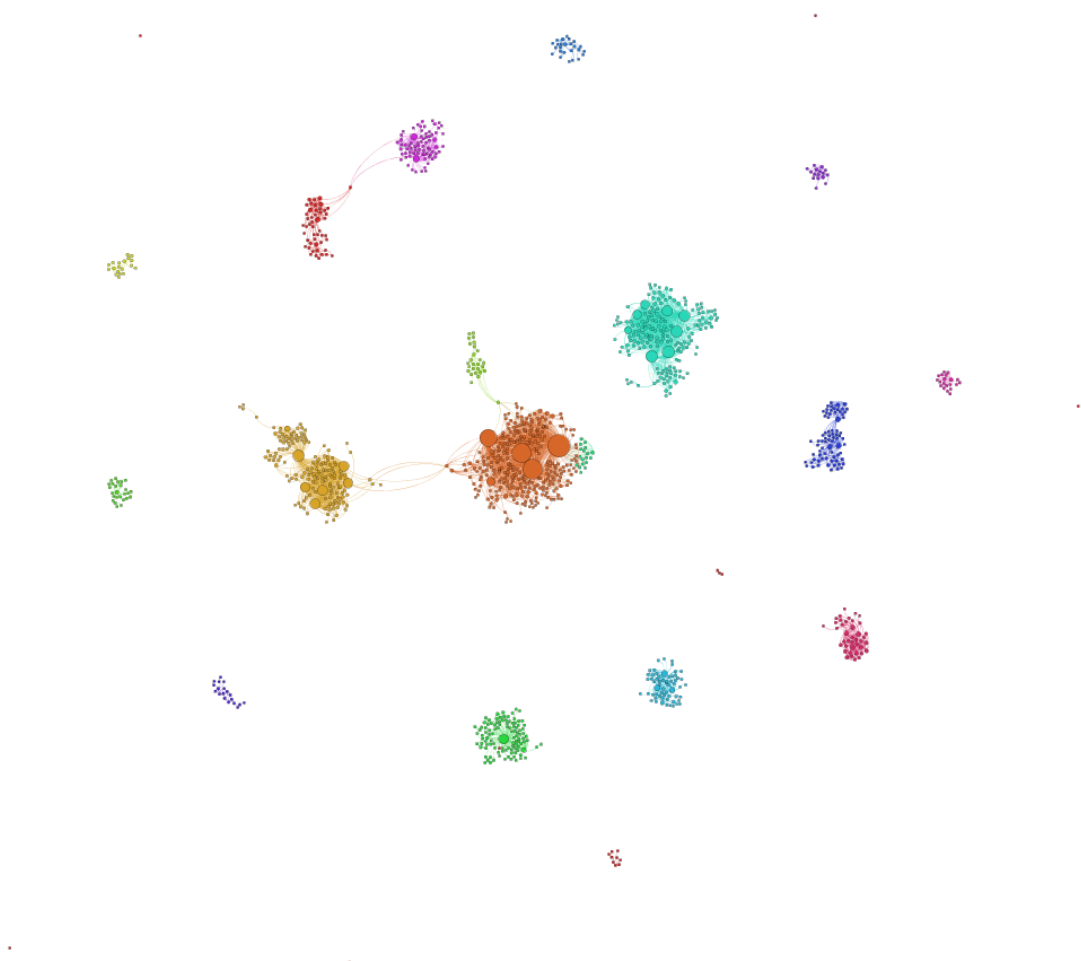
Cet exemple montre des groupes bien définis et très proches où plusieurs acteurs majeurs mènent la discussion. Les dynamiques d'échanges autour d'une question de société (la loi de l'enfant unique en Chine et ses conséquences) s'articulent en groupes distincts sans pour autant amener à des controverses importantes (qui se traduiraient par des discussions longues et houleuses). Ici, les leaders d'opinion font la discussion et la diffusion se fait au travers d'eux.



**Graphique 3.2.11 – “Shidu Muqin”**

### Exemple 3 : Abolition des lois sur la prostitution

Le hashtag “*Abolissons la loi piaowudong nuzui*” est l’expression d’une campagne pour l’abolition d’une législation scandaleuse sur la prostitution en Chine. Depuis les années 80, la loi chinoise interdit la prostitution et prévoit la condamnation des deux parties qui s’adonnent à un échange d’argent. Baptisé “*Piaowudong nuzui*”, cette loi a vu plusieurs cas absurdes impliquant des viols organisés sur mineurs se solder par la condamnation et l’emprisonnement des enfants incriminés. Relayés par les journalistes, les scandales à répétition ont éclaté à plusieurs reprises, impliquant parfois des officiels du Parti souvent blanchis alors que des enfants étaient eux emprisonnés.



**Graphique 3.2.12 – “Piaowudong nuzui”**

Le graphe 3.2.12 des discussions autour de l’abolition de cette loi donne l’impression

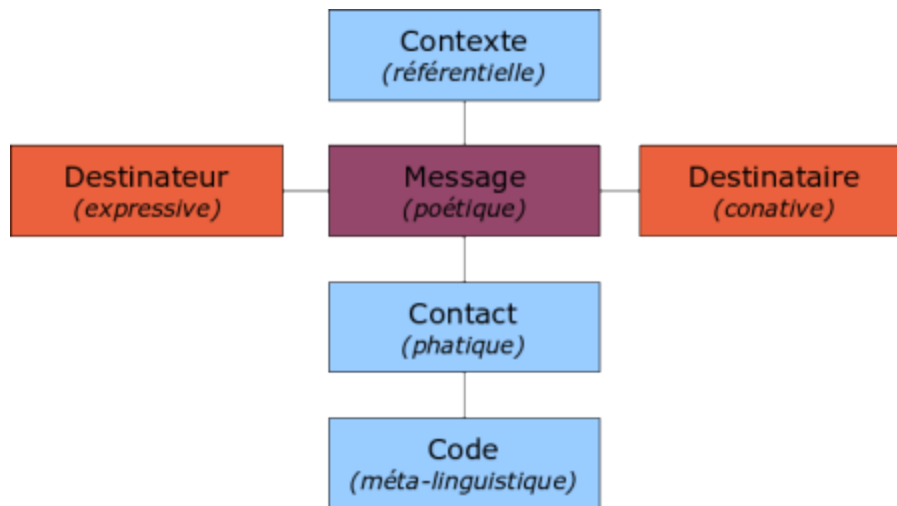
que de nombreux groupes discutent séparément de cette question puisque les premiers 50% du graphe sont déjà constitués de plus d'une quinzaine de clusters. Les groupes sont très éloignés entre eux, n'entretenant que peu de relations et connaissant une activité intense.

Cet exemple présente les caractéristiques d'une conversation très décentralisée dans laquelle de nombreux acteurs différents prennent part. Ce type de discussion fragmentée montre qu'un sujet peut s'étendre sans cohérence interne au média lui-même. Ici un agent extérieur (en l'occurrence un article du journal *Nanfang Zhoumo* sur le sujet) fait naître la conversation sans pour autant l'accaparer ni la centraliser. Plus difficilement détectable et contrôlable, cette dernière configuration est typique du même car elle se développe de façon large et durable entre des groupes à l'origine peu connectés.

Cette première visualisation des graphes conversationnels nous permet d'explorer quelques types précis d'échanges et d'en proposer une première lecture. Le procédé de visualisation reste rudimentaire et soulève plusieurs questions que nous nous donnons pour tâche de continuer à explorer. Premièrement, dans quel espace a lieu cette représentation ? En étalant ainsi ces graphes conversationnels, quelle action réalisons-nous réellement et quelle en est la valeur pour l'analyse ? À plus forte raison, quelle est la relation de cette espace du graphe conversationnel avec les autres formes d'espace, et plus notamment l'espace du réel géographique et l'espace de la représentation par le langage ?

Nous souhaitons ici mettre en perspective ces différentes dimensions afin d'enrichir le modèle d'étude. En effet, l'analyse de la diffusion utilise principalement les graphes des réseaux de diffusion mettant en scène les utilisateurs et leurs interactions en ligne. Loin d'être inintéressant, ce type de schéma est néanmoins très réducteur car il se fonde sur un modèle communicationnel très primaire “émetteur-récepteur” dont on connaît les limites. Les travaux de Jacobson ont notamment permis d'étoffer ce modèle en considérant les différents aspects fonctionnels des actes de communication (fig. 3.2.13).

Ainsi, en nous basant sur l'apport des théories de la communication, nous pouvons peut-être améliorer les modèles méthodologiques d'analyse. Nous proposons ici la notion de *topogramme* comme motif de diffusion des mêmes, considérés comme *topos* ou *lieux communs*. La représentation graphique du topogramme contient les différentes dimensions et dynamiques lisibles dans les données. Ce travail d'observation précis de la diffusion des



**Graphique 3.2.13** – Le modèle de Jakobson présente différents aspects des actes de communication que nous pouvons chercher à adapter dans le contexte des échanges en ligne.

mêmes permettra par la suite une première classification des actes de communication en ligne. Afin de prendre en compte, les différents aspects de la communication, il nous faut donc effectuer une analyse à multi-niveaux, regroupant un ensemble de réseaux à la fois sémantique, conversationnel et géographique.

L'identification de motifs récurrents dans la diffusion des contenus Internet s'impose comme un enjeu de taille notamment pour la compréhension des phénomènes stratégiques de persuasion dans les actes de communication. La capacité à identifier des topogrammes particuliers ouvrent de nouvelles possibilités pour la critique des phénomènes médiatiques et le design de systèmes d'informations à une échelle plus large. L'approche quasi-ontologique des faits de communication par leurs topogrammes permet d'observer la constitution de différents milieux numériques.

### 3.3 Méthodes d'identification de mèmes dans un large volume de données

Choisir un ensemble de mèmes cohérent est une des étapes difficiles de notre recherche. En effet, parmi les millions de messages de notre corpus et plus généralement dans la multitude des échanges quotidiens sur les réseaux sociaux, trouver une prise pour l'étude n'est pas une tâche évidente. Cette section présente les résultats obtenus lors de différentes expérimentations pour constituer des corpus de données représentatifs de mèmes dans un large volume de données.

#### 3.3.1 Constitution et collection d'un corpus de messages

La plupart des services de réseaux sociaux en ligne offrent un large accès à leur données. En effet, il s'agit souvent de la fondation de leurs modèles d'affaires basés sur la valorisation et la revente de ces données pour le marketing ciblé (Ko *et al.*, 2010). Pour entrer en contact avec la base de données, les SNS mettent à disposition une API (*Application Programming Interface*) qui permet à un programme ou une autre application web de se connecter au service pour demander et obtenir des données. L'API est donc la première source d'obtention de données depuis les SNS. Néanmoins, les données des réseaux sociaux sont soumises à d'importants enjeux et contraintes tant commerciales, éthiques que politiques - dans le cas de la Chine notamment. Les conditions d'utilisations techniques et légales (*Terms of Uses*) de ces données sont également soumises à des changements fréquents, étroitement liées à l'évolution commerciale et technologique de compagnies souvent très jeunes. Nous avons établi une liste des limitations et écueils pouvant être rencontrés lors de l'extraction et de l'analyse de données des SNS :

##### Compatibilité

Une solution technologique devient facilement caduque lors de l'évolution d'une API (ex. la version v1.0 de l'API de *Twitter* n'existe plus, ainsi le code doit être réécrit pour la version 1.1).

##### Disponibilité

Chaque API répond à des formats et critères précis et possède ses propres limita-

tions. Pour accéder à l'API du moteur de recherche de *Sina Weibo*, il faut s'identifier auprès de la compagnie grâce à une carte d'identité et des paiements par requête sont exigés.

### Limitations d'usage

Afin de limiter le trafic et conserver le contrôle sur les données distribuées, les SNS mettent en place des limitations d'accès à leurs serveurs, notamment : requêtes par heure, requêtes par machine (basée sur l'adresse IP), utilisateurs connectés. *Twitter* limite à 150 requêtes API par heure pour un compte non-identifié, pouvant augmenter jusqu'à 500 après authentification. Les données datant de plus de 7 jours sont payantes, reflétant la valeur d'un accès en "temps-réel".

### Légalité

Les données sont soumises aux conditions de propriété décrites légalement par la firme qui les publient (Clifton *et al.*, 2006). Ces conditions sont susceptibles de changer. Ainsi, *Twitter* a exigé en 2012 le retrait a posteriori de nombreux jeux de données publiés par des chercheurs depuis plusieurs années parfois (McCreadie *et al.*, 2012). Actuellement, *Twitter* précise notamment dans ses *Terms of Use* : “*You may not resyndicate or share Twitter content, including datasets of Tweet text and follow relationships*”<sup>3</sup>.

### Ethique

Vous pouvez extraire depuis une API des profils d'utilisateurs contenant les informations qu'ils ont auparavant publiées en ligne (Felt et Evans, 2008). En exposant les données personnelles des utilisateurs, le chercheur est responsable des risques qu'il peut faire courir à certains utilisateurs (Rieder, 2005).

Afin de contourner les limitations de l'API, la technique dite du “*scraping*” permet d'obtenir des données. Un robot lit et sauvegarde des parties ou l'intégralité de pages web. Les moteurs de recherche utilisent notamment cette technique pour l'indexation des pages. Le scraping est également soumis à des limitations par les services web (blocage de l'IP source) et se situe à la limite de la légalité, voire est explicitement interdit dans le cas de certains SNS (Petschulat, 2010).

---

3. Terms of use de Twitter, <https://dev.twitter.com/terms/api-terms>, consulté le 12 Mars 2013 à 17h08

Un programme appelé *spider* ou *crawler* est chargé d’obtenir et collecter des informations dans une base de données en effectuant des requêtes régulières à l’API. Plusieurs approches existent dans les techniques d’échantillonnage de réseau social. La première, fondée sur des mots-clés, extrait les posts contenant des mots ou des hashtags particuliers. La seconde utilise l’échantillonnage de graphe, collectant au fil des liens les conversations ou profils des utilisateurs. Classique des études statistiques, cet échantillonnage dit *boule de neige* “élargit l’échantillon en partant d’un *node* original pour s’éloigner vers ses voisins” (Rothenberg, 1995). Ici, deux grandes catégories s’opposent : les techniques transversales où les nodes sont classifiés après avoir été visités et les “walk” aléatoires où l’extension du graphe se fait de manière aléatoire (Gjoka et Kurant, 2011).

Au cours du travail préparatoire de cette recherche, nous avons tout d’abord expérimenté plusieurs algorithmes et outils de collection de données afin d’en comparer les résultats. Une première approche d’extraction par utilisateurs a été infructueuse car la sélection du groupe source (*seeds*) ne permettait pas d’obtenir des résultats cohérents<sup>4</sup>. Par la suite, une autre approche de collecte de données via le développement d’un plug-in pour le navigateur *Google Chrome*<sup>5</sup> nous a permis de tester et d’apercevoir les limites de la collection de données par mots-clés. Cette étape nous a également montré l’intérêt que peut présenter une approche collaborative de la collection de données ou de seeds par un système collaboratif de “curation” pour réduire la masse de données obtenues à des éléments précis de contenus (Ding et al., 2013). Après de multiples tests et comparaisons d’outils et de bibliothèques, plusieurs difficultés majeures limitaient l’obtention d’une quantité de données suffisantes, notamment la nécessité de ressources assez importantes (en terme de développement et de disponibilité des machines), un temps d’acquisition parfois très long et l’exigence d’une veille constante sur les SNS pour identifier un même au bon moment (les tweets de Sina Weibo devenant indisponibles via l’API au-delà de 7 jours). La première limite se situe bien sûr dans la capacité d’une personne seule à mener à bien cette large tâche.

Nous avons donc choisi de considérer les jeux de données déjà collectés et disponibles concernant *Sina Weibo*. Une fois écartés les nombreux jeux tronqués, modifiés ou incomplets, nous avons pu obtenir plusieurs jeux de données provenant de recherches préalables

---

4. Code disponible : <https://github.com/sharismmlab/Pyweibo>, consulté le 14 Mars à 5h32

5. Code disponible : <https://github.com/sharismmlab/battlefield>, consulté le 14 Mars à 5h12



dans le domaine particulier de l'échantillonnage (Ding *et al.*, 2013) ou ayant servi de bases à des études précédentes (Gao *et al.*, 2012). Finalement, nous avons identifié le jeu de données constitué lors du projet *Weiboscope* de l'Université de Hong Kong comme répondant à nos besoins en termes de dimensions (temps, nombre d'utilisateurs observés), taille (nombre de tweets) et contenus (géo-localisation, présence des tweets censurés). Notre travail d'analyse s'appuie donc sur ce jeu de données collecté sur le service de microblog *Sina Weibo* par le *Journalism and Media Studies Centre* (JMSC) de l'Université de Hong Kong lors de son projet *Weiboscope*. Téléchargeable ouvertement, la publication de ce jeu de données a pour objectif de *“enables academic use of the data for better understanding of the social media in China and making the Chinese media system more transparent.”*<sup>6</sup> Il s'agit d'un échantillonnage aléatoire de messages (*random sampling*) effectué quotidiennement durant toute l'année 2012 sur un panel d'environ 350 000 utilisateurs ayant au moins 1000 followers (Fu *et al.*, 2013). La totalité du jeu de données comprend 226,841,122 messages répartis sur 52 semaines, dont des messages ayant été supprimés par les utilisateurs eux-mêmes ou par les administrateurs de Sina Weibo eux-mêmes - parfois sur ordre du gouvernement chinois (*ibid*, 2013).

Ce jeu de données a été mis à disposition sous une forme anonymisée où les identifiants des messages et des utilisateurs ont été remplacés par des pseudo-identifiants. La collection des données a été effectuée sur une série d'utilisateurs (génération aléatoire d'identifiants dont l'existence est ensuite validée) pour donner *“une image représentative des usages et utilisateurs de Sina Weibo (...) dont les études auparavant limitées à des analyses non-aléatoires (...) se cantonnaient aux utilisateurs les plus populaires”* (Fu *et al.*, 2013). Cet échantillon s'attache à refléter les pratiques des utilisateurs “lambda”. Ce jeu de données a déjà été partiellement étudié dans le but de comprendre la démographie des utilisateurs de Sina Weibo, leurs activités et les comportements pouvant permettre de prédire les réactions notamment de censure. La démographie des utilisateurs se compose de 55% d'hommes habitant principalement dans les grandes villes de Chine (Pékin, Canton, Shanghai). Une des découvertes importantes est le très faible taux de création originale de messages malgré une activité importante des utilisateurs, indiquant que l'essentiel de l'activité sur Sina Weibo se constitue de re-posts et de commentaires (Fu *et al.*, 2013). Le

---

6. Le jeu de données Weiboscope est disponible à l'adresse : <http://147.8.142.179/datazip/>, consulté le 14 Mars 2014 à 17h21

Désignation	Description
mid	Unique pseudo message ID
retweeted_status_mid	Pseudo message ID of the original message (Only available if the row of interest is a retweet)
Uid	Pseudo user ID
retweeted_uid	Pseudo user ID of the original poster (Only available if the row of interest is a retweet)
Source	The application name of the client program
Image	With image ? (1= Yes, 0=No)
text	body of the message. Any address handle (@xxxx :) is replaced by either the pseudo user ID or ukn (unknown)
geo	GIS information. Please refer to the Sina Weibo API documentation : <a href="http://goo.gl/Um8SS">http://goo.gl/Um8SS</a>
created_at	Original posting time
deleted_last_seen	The last seen time before this message was missing from the user timeline
permission_denied	'permission denied' status is marked when the message was found missing in the timeline and the API return message was 'permission denied' - See details in (Fu, Chan, Chau 2013)

**Tableau 3.3.1** – Modèle de données du jeu de données *Weiboscope*

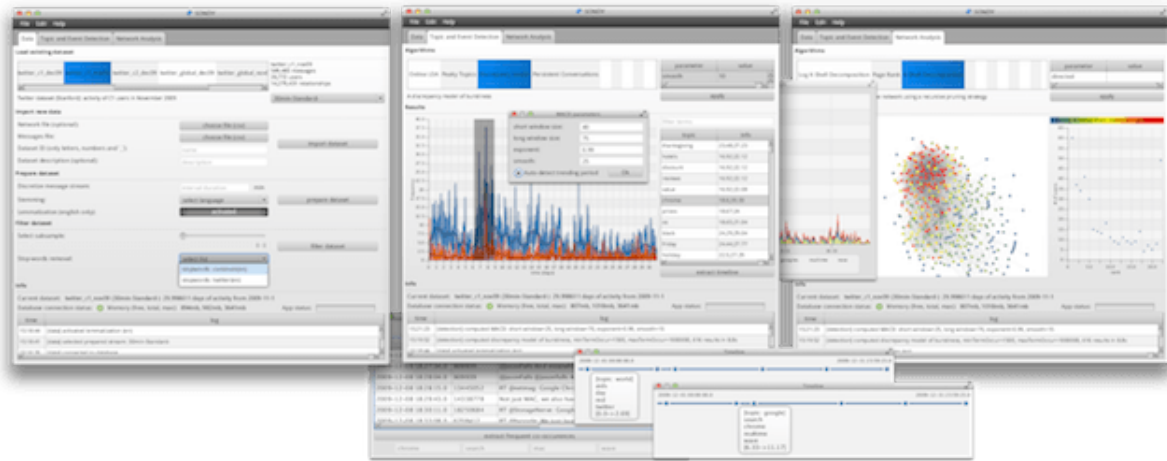
jeu de données est accompagné des informations succinctes de profil des utilisateurs dont le lieu, rempli par eux, sans néanmoins fournir leurs noms d'utilisateurs véritables. Notre travail de recherche s'articule autour d'une nouvelle lecture de ce jeu de données unique.

### 3.3.2 Détection algorithmique de mèmes dans un large corpus de données

Une fois l'acquisition des données effectuée, il s'agit désormais de savoir les analyser correctement pour y déceler les mèmes que nous souhaitons observer. Les travaux dans le domaine de la détection et l'identification de mèmes dans les données de réseaux sociaux restent encore peu nombreux. Une des études pionnières est l'outil *MemeTracker* (devenu *NIFTY*) conçu en 2009 par le *SNA Project* de l'Université de Stanford (Leskovec *et al.*, 2009). Cet outil permet une étude sous forme de graphes de la diffusion de phrases

dans un vaste corpus de texte mais n'est pas adapté à la langue chinoise. La discussion sur la modélisation mathématiques des mèmes ([Ahmad et Teredesai, 2006](#); [Nye, 2011](#)) émane souvent de recherches en informatique cherchant à prendre en considération différents facteurs de diffusion lors de l'analyse machine de données ([Zubiaga et al., 2011](#); [Wang et Wood, 2011](#)), considérant parfois le mème comme un vecteur de modification du réseau lui-même ([Ienco et al., 2010](#)). Plus marginales, de rares études s'intéressent aux dynamiques géographiques des mèmes ([Kamath et al., 2013](#)). Néanmoins, aucune de ces différentes approches ne permet d'apporter une réponse technologique ou algorithmique satisfaisante pour l'identification de mèmes Internet dans un ensemble de données issues des réseaux sociaux. Nous devons a priori déceler les motifs de diffusion particuliers formés par les mèmes dans ce vaste ensemble textuel. La dénomination *machine learning* regroupe un ensemble d'algorithmes qui permettent d'explorer des jeux de données pour en extraire des représentations et y identifier des propriétés (*features*) particulières. Basé sur les sciences statistiques, ces algorithmes font usage de mesures de similarité pour classer les éléments d'un corpus. Les catégories utilisées pour la classification peuvent être définies au préalable par l'utilisateur - c'est le *supervised learning* - ou inférées du jeu de données lui-même - dans le cas du *unsupervised learning* ([Breiman, 2001](#)). La multitude d'algorithmes de *machine learning* disponibles pour la détection de *clusters* au sein d'un corpus textuel ou d'un réseau social ([Nettleton, 2013](#); [Robins, 2013](#)) rend l'identification d'une solution difficile. De plus, les algorithmes utilisés traditionnellement pour l'analyse de documents textuels (*topic modeling*, *LSA*) se heurtent au caractère très disparate des corpus issus des réseaux sociaux (*text sparsity issue*) faisant diminuer drastiquement leur efficacité ([Hong et Davison, 2010](#)). L'outil *Sondy* développé par [Guille et al. \(2013\)](#) nous a notamment permis d'explorer une partie du corpus à l'aide de différents algorithmes (fig. 3.3.1). L'implémentation de l'analyse de langage chinoise et la faible efficacité des algorithmes à détecter des mèmes peu événementiels nous ont néanmoins amené à chercher des solutions plus spécialisées.

[Ferrara et al. \(2013\)](#) propose dans un papier intitulé “*Meme clustering in social media*” un algorithme utilisant la classification automatique non-supervisée pour détecter des mèmes dans un corpus de données de réseaux sociaux. Ce travail récent propose de tenir compte non seulement des textes et hashtags, mais aussi des liens et des modèles de diffusion pour identifier des groupes de messages intéressants et procéder au *clustering*



**Graphique 3.3.1** – Captures d’écran de l’outil d’analyse de microblog *Sondy* d’après <http://mediamining.univ-lyon2.fr/people/guille/software.php> consulté le 15 Mars 2014 à 16 :12

des mêmes. L’algorithme s’articule autour du concept de “protomèmes”, représentant les éléments fondamentaux d’un mème en cours de création (Gabora, 1995). Dans le contexte des médias sociaux, le protomème est défini par les entités (liens, hashtags...), mots-clés et séquences de conversation qui constituent le mème en devenir. En identifiant puis comparant les différents protomèmes présents dans chaque tweet, il est possible d’y détecter des similarités et de deviner les mèmes en formation.



**Graphique 3.3.2** – Algorithme de reconnaissance de mème (clustering) (Ferrara *et al.*, 2013)

Cet algorithme suppose donc d’extraire dans un premier temps les éléments remarquables du corpus de tweets afin d’établir des représentations de ces protomèmes contenant les éléments à comparer : *phrases* (texte brut), *mentions* (@, RT), *hashtags* et *urls*. Pour constituer ces jeux de protomèmes, nous utilisons le pattern *map-reduce* qui permet de chercher et lister des éléments dans un vaste jeu de données (*map*) avant de les regrouper dans une liste (*reduce*). Une fois ces protomèmes constitués, nous procédons à leurs comparaisons selon plusieurs critères :

- *Similarité de texte* : comparant le contenu textuel de chaque protomème

- *Similarité d'utilisateurs* : comparant les utilisateurs contenus dans chaque protomème
- *Similarités de tweet* : recherchant les tweets identiques dans différents protomèmes
- *Similarité de diffusion* : considérant les références aux utilisateurs contenus dans chaque protomème

Ici nous utilisons la *sémantique vectorielle* (Support Vector Machine) afin de comparer les éléments des protomèmes. Cette pratique ancienne de l'algèbre linéaire appliquée à la science informatique (Salton *et al.*, 1975) permet de représenter des objets textuels (mots, identifiants, images...) sous une forme de vecteurs aisément comparable. Pour convertir le texte sous forme vectorielle, l'algorithme classique Tf-idf (*Term Frequency - Inverse Document Frequency*) est utilisé (Soucy et Mineau, 2005). Les autres mesures de similarité sont la *mesure cosinus* (ou *similarité cosinus*) des protomèmes convertis sous forme de vecteurs binaires. Une fois ces différentes valeurs de similarité calculées, nous utilisons les scalaires définis dans le papier de référence pour assigner des poids à chacun des vecteurs et les combiner en une seule valeur (Ferrara *et al.*, 2013). Cette matrice de valeurs de similarité nous permet alors de définir les protomèmes les plus similaires et d'identifier ainsi des *clusters* dans les données correspondant aux mêmes.

Si cette approche offre des résultats probants sur de petits volumes (quelques centaines de tweets), la très grande demande en puissance de calcul et ressources mémoire nécessaires rendent le traitement d'un jeu données plus vaste irréalisable. Les opérations de comparaison et le calcul de similarités sur de vastes volumes de données font croître très rapidement la quantité de calculs à effectuer. Le calcul du coût d'un algorithme se fait au travers des notions dites de domination, avec notamment le "grand O" exprimé  $O(f(n))$  qui fait correspondre à la complexité d'un algorithme une fonction  $f$  de la quantité d'information manipulée  $n$ . Ainsi pour un algorithme courant de complexité  $O(n^2)$ , les ressources de calcul (*computation*) et de mémoire (RAM ou stockage) nécessaires augmentent de manière exponentielle à chaque élément ajouté au corpus  $n$ . L'algorithme de "meme clustering" décrit par Ferrara atteint donc un coût exorbitant devant un large volume de données comme celui du jeu Weibosope. La limite physique de calcul est rapidement atteinte rendant impossible le franchissement du palier expérimental et la vérification des hypothèses de travail à une échelle suffisante. Ainsi, nous allons poursuivre ces expérimentations en nous intéressant désormais aux hashtags.

### 3.3.3 Les hashtags ne sont pas des mèmes

Les *hashtags* (en français “mots-dièses”) sont utilisés dans l’écriture sur les réseaux sociaux et se présentent dans *Sina Weibo* sous la forme d’un mot entouré de deux dièses - ex. *#mot-clé#*. Marqueur particulier, le hashtag permet à un interlocuteur de procéder à une dénotation ou connotation du message original (Romero *et al.*, 2011) ou d’affirmer son caractère événementiel (lors d’un événement sportif, d’une conférence, etc). Facile à identifier dans la masse des données en ligne, il permet de désigner un ensemble de contenus sous un même signe. Ainsi, il est un vecteur important permettant de collecter simplement une large somme d’informations autour d’un mème. La constitution d’un corpus autour d’un “hashtag” présente néanmoins plusieurs limites. Premièrement, le mème est par définition un objet en mutation. Il est difficile de l’identifier une bonne fois pour toute par un ou plusieurs mot-clés. De plus, le mème existe bien souvent sous la forme d’images ou de vidéos qui ne sont pas nécessairement légendées ou référencées et donc peu accessibles à une recherche “plein texte”. Une approche pour la recherche de mèmes ne peut être entièrement textuelle et doit s’intéresser aux autres forme de contenus web (notamment les liens). De plus, l’ajout de hashtags dans les messages est un acte volontaire non systématique. Ainsi, si l’identification de certains mèmes peut se faire grâce à la recherche de hashtags, l’ensemble des messages contenant des hashtags ne recouvre pas systématiquement un mème. Comme nous le verrons, les hashtags sont bien souvent de simples artefacts de campagne marketing en ligne.

Afin de procéder à l’analyse des mèmes, nous avons donc indexé l’ensemble des contenus du corpus Weiboscope contenant des hashtags sur toute l’année 2012 (30 millions de messages sur un total de 200 millions environ). Dans un premier temps, l’ensemble des messages contenant un ou des hashtags a été classifié pour obtenir des jeux de données cohérents par hashtags. L’extraction des hashtags est effectuée grâce à une expression régulière qui scanne le texte pour identifier et préserver uniquement les caractères situés entre les deux signes dièse (#).

Sur un total d’environ 30 millions de tweets contenant des hashtags, nous avons choisi de retenir seulement les hashtags contenus dans plus de 1000 messages et d’ignorer les 1000 hashtags les plus utilisés. En effet, ils ne présentaient pas d’intérêt pour notre étude, étant la plupart du temps des noms de marque ou des mots-clés trop généraux (par exemple :

---

**Algorithm 2** Extract Hashtags from Message

---

**Require:**  $m$  is a microblog text message

```
function EXTRACTENTITIES( $m$ )  
  if begins with # and ends with # then  
     $h \leftarrow \text{hashtag}$   
     $\text{text} \leftarrow u$  from  
  end if  
end function
```

---

**Graphique 3.3.3** – *Algorithme d'extraction d'entités d'après les messages* Les entités extraites des messages sont : les hashtags, les citations et les mots importants.

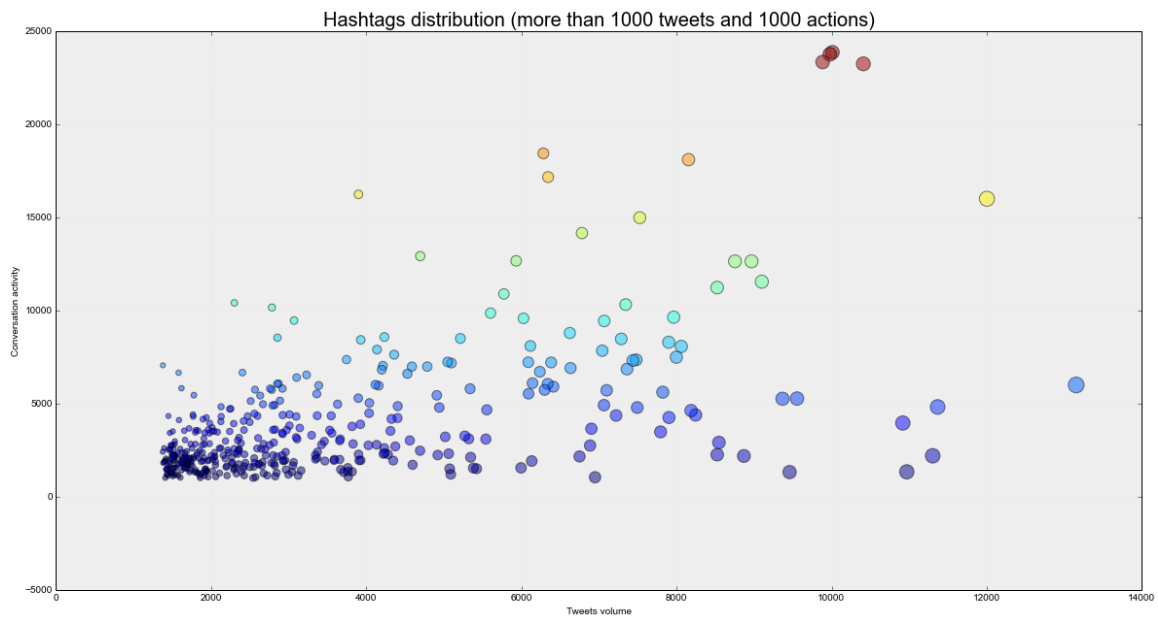
“bonne nuit”, “nouvelles de sports”, “photos de nourriture”, etc).

<i>hashtags</i>	<i>users</i>	<i>actions</i>	<i>tweets</i>
吴奇隆	201	13243	22349
一起到老	182	0	364
春运	92	13	256
轻松一刻	92	11	240
人品值分析	90	490	321
朝阳区	88	49	165
理性态小度	87	0	329
美图GIF	87	101	404
我正在听	86	6	330
微盘签到	84	304	305
2012来了	83	206	309
中级达人	83	0	159
分享	82	87	563
星座	82	5	195

**Graphique 3.3.4** – *Hashtags les plus utilisés pendant la 1ère semaine de 2012* - Les résultats concernant la 1ère semaine de l’année 2012 donnent un aperçu du volume analysé : 5,044,331 tweets, 398 392 utilisateurs uniques cités (dans un total de 2 115 544 mentions), 264 651 urls uniques (pour un total de 426 914) et 44 382 hashtags uniques (pour un total de 244285)

Notre étude vise à analyser les dynamiques conversationnelles et nous devons donc déterminer les plus adéquats parmi des hashtags de nature souvent très différentes. Pour ce faire, nous avons sélectionné pour chaque hashtag deux mesures significatives : premièrement, le volume de messages ; deuxièmement, la quantité d’échanges et d’interactions effectives entre les utilisateurs (commentaires, retweets, etc.). Ces deux mesures nous permettent de nous assurer que 1) nous possédons une quantité suffisante de messages pour mener à bien l’étude et que 2) la discussion a bien eu lieu et qu’il ne s’agit pas de messages

redondants ou non reliés entre eux. Nous avons choisi d’ignorer les échanges dominés à plus de 80% par le même utilisateur pour éviter la pollution de l’étude par l’activité de robots. Le graphe 3.3.5 permet d’observer la distribution de 429 hashtags possédant tous plus de 1000 tweets et 1000 échanges : l’axe vertical représente la quantité d’actions (échanges) et l’axe horizontal le volume des conversations. Tout en bas du graphe se trouvent donc les hashtags ayant été les moins discutés, avec en haut ceux aux conversations les plus intenses. La taille des points illustre le volume de messages et la couleur la quantité de conversations.



**Graphique 3.3.5 – Distribution des 429 hashtags les plus discutés**

En procédant à l’étiquetage des hashtags les plus actifs durant l’année 2012 sur *Sina Weibo* (figure 3.3.5), nous constatons que la plupart sont associés à des activités commerciales, de loisirs ou de divertissement. Ici nous observons que les usages majoritaires du réseau social *Sina Weibo* correspondent pour la plupart à ceux d’autres mass-médias plus traditionnels de par le monde. Le commerce en ligne occupe notamment une place prééminente. La marque de téléphonie mobile chinoise *Xiaomi* est abondamment citée, reflétant son importance croissante dans le marché chinois et surtout sa stratégie commerciale qui cible abondamment les réseaux sociaux avec de nombreux hashtags très discutés (notamment “*Fans de Xiaomi*” #小米粉丝#). Également, de nombreuses campagnes promotionnelles d’ouverture ou d’anniversaire de magasins ont réussi à se hisser dans le jeu de tête des hashtags les plus discutés. Dernier-nés des radio-crochets ou chanteurs reconnus, les stars de la télévision et de la chanson sont aussi présentes dans le peloton de



têtes des discussions sur *Sina Weibo*. Le célèbre chanteur Han Geng notamment compte près d'une dizaine de hashtags le concernant parmi les 500 les plus discutés (*"Han Geng fait une pub pour Nokia"*, *"Han Geng va en Italie"*, *"Han Geng fait une pub Pepsi"*, *"Han Geng refuse une interview"*, etc.) Ici encore, le réseau social agit comme le prolongement des mass média traditionnels, élément-clé des nouvelles stratégies de publicités en ligne, parfois particulièrement agressives comme dans le cas de Han Geng. Les contenus de la télévision sont largement relayés et discutés, notamment les séries télévisuelles. Le cinéma est aussi représenté. Plus grand succès commercial du box-office chinois, le film comique *Lost in Thailand* sorti en Décembre 2012 dépeint les aventures d'un chinois en vacances en Thaïlande. Sa popularité se reflète dans l'importance au sein des discussions en ligne. Les tendances des ventes du livre sont également au rendez-vous avec de nombreux best-seller sur "l'amélioration de soi" ou la "réussite économique". Ce type de hashtags ne se limite pas au support web mais s'origine directement dans d'autres médias plus traditionnels. Le gouvernement lui-même utilise *Sina Weibo* pour faire passer ses messages avec un hashtag *"information officielle"* utilisé notamment pour des démentis publics ou droit de réponse par l'entreprise *Sina*, propriétaire du service. Également outil de conversation, le microblog héberge les discussions de la vie de tous les jours. La situation routière et les bouchons dans chaque ville sont un des grands sujets de discussions. Ce sont dans ces échanges quotidiens que se cristallisent plus particulièrement les enjeux politiques et médiatiques des réseaux sociaux. Nouveau café du commerce, les commentaires sur les faits divers et l'actualité mettent souvent à jour les dysfonctionnements de systèmes politiques, urbains ou légaux.

Il est intéressant néanmoins de noter que parmi les hashtags les plus discutés, les phénomènes de suppression de contenus par les administrateurs (censure) restent très marginaux. Le *China Digital Times* de UC Berkeley maintient une liste des mots interdits sur *Sina Weibo* depuis plusieurs années (Ng, 2013a). En comparant cette liste de mots censurés à celle des hashtags, nous avons pu voir qu'aucun des 3000 hashtags les plus utilisés en 2012 n'a été soumis à une interdiction même temporaire sur *Sina Weibo*. Les hashtags les plus sujets à la censure ne sont pas en lien avec des domaines politiques ou des sujets sensibles, mais plutôt avec des contenus à caractère pornographique (interdits en Chine). Reflétant les usages majoritaires (commerce, loisirs, etc.), les hashtags véhiculent des contenus souvent moins controversés et les "mots censurés" sont plus à

même d'apparaître dans des discussions informelles.

Ces observations nous amènent donc à interroger la pertinence du hashtag comme représentant des mèmes Internet. En effet, la plupart des hashtags semblent être le reflet de campagnes de communication dont la diffusion très planifiée ne laisse que peu de place à des interactions spontanées. L'aspect stratégique du hashtag en fait un artefact de campagnes de diffusion, définition seulement partielle et insatisfaisante du même Internet. Ainsi, nous préférons écarter l'usage de ce marqueur pour observer les mèmes.

Les expérimentations présentées précédemment nous ont permis de réunir, tester et sélectionner un ensemble d'outils et de méthodes parmi ceux disponibles. Ces diverses tentatives ont notamment montré les limites de chacune des approches. À la lumière de ces expérimentations et réflexions, nous allons maintenant présenter les solutions retenues pour la poursuite de cette étude.

# 4

## Résultats et discussions

LE DISPOSITIF TECHNOLOGIQUE que nous avons choisi de développer permet d’observer dans le détail la diffusion de mèmes ou d’autres formes de contenus identifiés sous la forme d’un ensemble de messages. Les expérimentations préliminaires nous ont permis de constater que la détection des mèmes était une tâche peu aisée. Nous allons décrire ici l’approche que nous avons choisi de retenir (choix technologiques, algorithmes et processus de validation). Notre dispositif permet de visualiser différentes dynamiques existantes dans un corpus de messages de microblog constitué d’après une recherche par mots-clés. Nous détaillons ici son fonctionnement et interface. Nous présentons ensuite un ensemble de résultats obtenus grâce à un échantillon d’une douzaine de mèmes de différents types et formes. Nous discutons ensuite des limites, contraintes et écueils possibles d’une approche utilisant l’analyse de données.

## A propos des choix technologiques

Pour mener à bien le développement de cet outil, nous avons été amené à effectuer de nombreux choix technologiques. Chacun des logiciels, programmes ou outils mobilisés sera présentés plus en détail dans les sections infra. Néanmoins, afin d'éclaircir le lecteur sur la nature et les raisons plus générales de ces choix, nous précisons ici les différents aspects qui sont entrés systématiquement en compte lors de chacune de nos décisions :

**Licence : gratuité, open-source et disponibilité** La possibilité de lire, modifier et utiliser les outils et librairies est un des éléments primordiaux dans le cadre de cette recherche. Non seulement, les coûts de développement sont très largement diminués par l'usage de solutions gratuites mais également la disponibilité des outils permet à ceux qui le souhaiteraient de se les réapproprier librement. De plus, le caractère *open source* permet de consulter les rouages du code et s'assurer de sa validité et sa qualité avant de l'invoquer.

**Documentation : communauté et support** Les outils parfois compliqués du développement informatique ne peuvent être utilisés sans un minimum d'explication et de documentation. Si la documentation laisse à désirer, que le code est mal rédigé ou que la communauté des créateurs ou utilisateurs n'est pas présente pour procurer des explications, l'usage de technologies si intéressantes qu'elles soient devient très difficile, voire impossible. Ainsi, nous avons opté pour des solutions possédant une documentation satisfaisante afin de simplifier leur (ré-)utilisation .

**Interopérabilité : déploiement et compatibilité** Le degré de complexité des procédures d'installation et de mise en place est également un facteur décisif dans la sélection d'un outil. En effet, la reproductibilité et la possibilité de réutilisation du code se voient diminuées par un déploiement complexe et laborieux, multipliant souvent les erreurs. Les outils choisis pour ce travail ont été développés pour être compatibles avec le système d'exploitation *Linux Debian*.

**Validation : usage commercial et scientifique** Notre recherche ne se situe pas dans le domaine de la science informatique. Á ce titre, elle fait usage d'outils utilisés tant dans la communauté scientifique que dans l'industrie du Web. Les processus de contrôle de qualité liés aux objectifs commerciaux des acteurs industriels assurent en effet une certaine optimisation du code qui pèse dans nos décisions d'utiliser

telle ou telle technologie.

L'ensemble de ces critères nous a porté à utiliser le langage *Python* comme principal outil de développement de notre outil. Largement utilisé par la communauté scientifique, Python dispose de nombreuses bibliothèques, plugins et outils qui le rendent attractif et efficace pour le prototypage et la recherche. De nombreux outils statistiques pour l'analyse de données ainsi que la plupart des algorithmes récents sont maintenus dans le package *SciPy*. Également, il est simple à déployer et possède de nombreuses ressources pour aider au développement et à la documentation.

## 4.1 Un outil de traitement et de visualisation des mèmes

Dans cette section, nous présenterons les développements spécifiques que nous avons menés pour construire cet outil. Les comment et pourquoi de nos différents choix technologiques seront explicités dans une présentation pas à pas du fonctionnement et de l'utilisation de ce système d'analyse.

### 4.1.1 Constitution de corpus par mots-clés pour chaque mème

Comme nous l'avons vu précédemment, ni les méthodes de détection algorithmiques (section 3.3.2) ni l'usage de l'objet hashtag (section 3.3.3) n'ont permis de saisir de manière satisfaisante l'objet mème dans notre vaste jeu de données (section 3.3.1). Ainsi, nous avons choisi de considérer une troisième approche qui consiste à décrire le mème sous la forme de mots-clés. Les travaux en sciences humaines sur les mèmes ([Bauckhage, 2011](#); [Coscia, 2013](#); [Knobel et Lankshear, 2007](#)) et les sites spécialisés ([Buchel, 2012](#); [Bernstein et al., 2011](#)) présentent bien souvent un mème sous la forme d'un titre (mot ou groupe de mots) accompagné d'une collection d'images ou/et de vidéos, ainsi qu'un texte explicatif (mise en contexte) et des indications sur les dates de parution et son évolution dans le temps. Ici, un ou des auteurs se saisissent des matériaux bruts du web et les rassemblent pour reconstituer une vision particulière du mème, aussi représentative que possible. Cette approche nécessite peu de connaissances technologiques (un copier/coller suffit souvent),

mais exige néanmoins une forte connaissance du terrain.

Pour chaque mème, nous pouvons donc procéder à l'extraction d'un jeu de données contenant l'ensemble des messages correspondant à une requête définie composée de différents mots. La faiblesse intrinsèque de cette approche est qu'elle nécessite de savoir ce qu'on veut chercher, à l'inverse d'une détection des mèmes qui permettrait de les identifier sans connaître leur existence a priori. Cette méthode prend également peu en compte les éléments audio et visuels (images, vidéos) qui constituent pourtant une partie importante des mèmes Internet.

Néanmoins, elle permet une approche intuitive par l'usage de mots-clés et une vérification itérative de la qualité des résultats obtenus, à l'inverse de la détection algorithmique notamment. Les biais de la recherche plein-texte peuvent être rétablis : effacement des éléments exogènes, reconstitution des éléments manquants du corpus...

## Indexation du corpus Weiboscope pour la recherche plein-texte

La fonction de recherche plein-texte est un des secteurs des technologies numériques qui a connu la plus forte expansion depuis les 10 dernières années de l'Internet, comme en témoigne l'histoire de *Google*. L'accroissement du volume des données et la nécessité de naviguer en leur sein ont fait des moteurs de recherche un élément incontournable du paysage de nos navigateurs. Chaque site ou presque possède aujourd'hui un champ "*search*". Suivant cet important développement, des solutions de plus en plus fiables et performantes ont été mises à disposition sous des licences ouvertes, permettant une appropriation facile.

Dans cette étude, nous avons choisi d'utiliser le moteur de recherche *ElasticSearch*<sup>1</sup> pour accéder au contenu du corpus. Utilisant la technologie robuste du moteur d'indexation *Apache Lucene*<sup>2</sup>, il permet d'indexer de larges corpus dans une base de données non-relationnelles, mettant à disposition une API efficace et structurée suivant la norme *REST* (Massé, 2012). Largement utilisé et documenté, *ElasticSearch* permet l'indexation des caractères chinois grâce au plugin *Lucene SmartChinese Analyzer*, utilisant un algorithme dit de *Markov caché*. Appliqué à un vaste *training corpus* de texte Chinois et de

---

1. Voir le site <http://www.elasticsearch.org>, consulté le 22 Avril 2014 à 12 :23

2. <http://lucene.apache.org/> consulté le 20 Juin 2014 à 11 :22

dictionnaires<sup>3</sup>, il permet de segmenter le texte chinois en mots. Les mots sont ensuite indexés dans la base de données de *Lucene*. Chaque requête dans le moteur de recherche permet alors d'établir la similarité entre les instructions (mots-clés) et les documents textuels contenus dans la base de données en attribuant un poids spécifique à chaque document. Enfin, le moteur de recherche renvoie les résultats sous la forme d'un objet au format standard *JSON*.

Afin de procéder à la recherche de mêmes dans le texte, nous avons donc indexé dans *ElasticSearch* les parties du corpus *Weiboscope* à l'aide d'un script<sup>4</sup>. Chacun des 52 fichiers au format *zip* contenant des données au format *csv* (texte des microblog en chinois et méta-données) déposé dans un dossier sélectionné est indexé. Cette manipulation est également reproductible pour tout type de fichiers en langue chinoise ou anglaise (d'autres langues nécessitent simplement l'ajout d'un analyseur de texte *Apache Lucene*).

### Sélection qualitative de mots-clés

Une fois le texte indexé, il nous faut maintenant définir les requêtes appropriées qui nous permettront d'extraire un corpus intéressant pour l'étude d'un ou plusieurs mêmes. La définition des mots-clés est un exercice compliqué car il suppose que les mêmes ne recoupent pas ou peu des mots utilisés couramment usités. Or, nous avons vu que le propre du même est justement l'intertextualité et à ce titre joue littéralement sur les mots. Les résultats d'une recherche pour les mots *crabe* et *rivière* (voir figure 1.3.1) dans le moteur de recherche n'auront que peu de liens avec le même *hexie* somme toute très marginal dans la masse des contenus. Ici, la syntaxe propre du moteur de recherche peut nous venir en aide, permettant d'inclure (*AND*) ou d'exclure (*NO*) certains mots précis ou d'exiger du moteur de considérer des groupes de mots entiers ("*\""). Une autre possibilité est de limiter la recherche dans le temps en se concentrant sur la période où le même connaît sa période de plus fort intérêt.*

La qualité de la requête et des résultats renvoyés par le moteur de recherche pour chaque même peut être évaluée selon trois aspects importants :

---

3. Voir <http://www.ictclas.org/> consulté le 7 Juillet 2014 à 11 :32

4. consultable ici [https://github.com/clemsos/mitras/blob/master/es\\_build\\_index.py](https://github.com/clemsos/mitras/blob/master/es_build_index.py) consulté le 7 Juillet 2014 à 11 :27

**Volume** Taille significative et volume suffisant de données pour constituer un corpus représentatif du mème.

**Qualité** Possibilité de vérifier manuellement la qualité du contenu par la lecture d'un échantillon parmi les messages obtenus.

**Dimensions** Les différents paramètres retournés (dates, lieu, utilisateur précis, contenu censuré...) correspondent à ceux choisis dans la requête

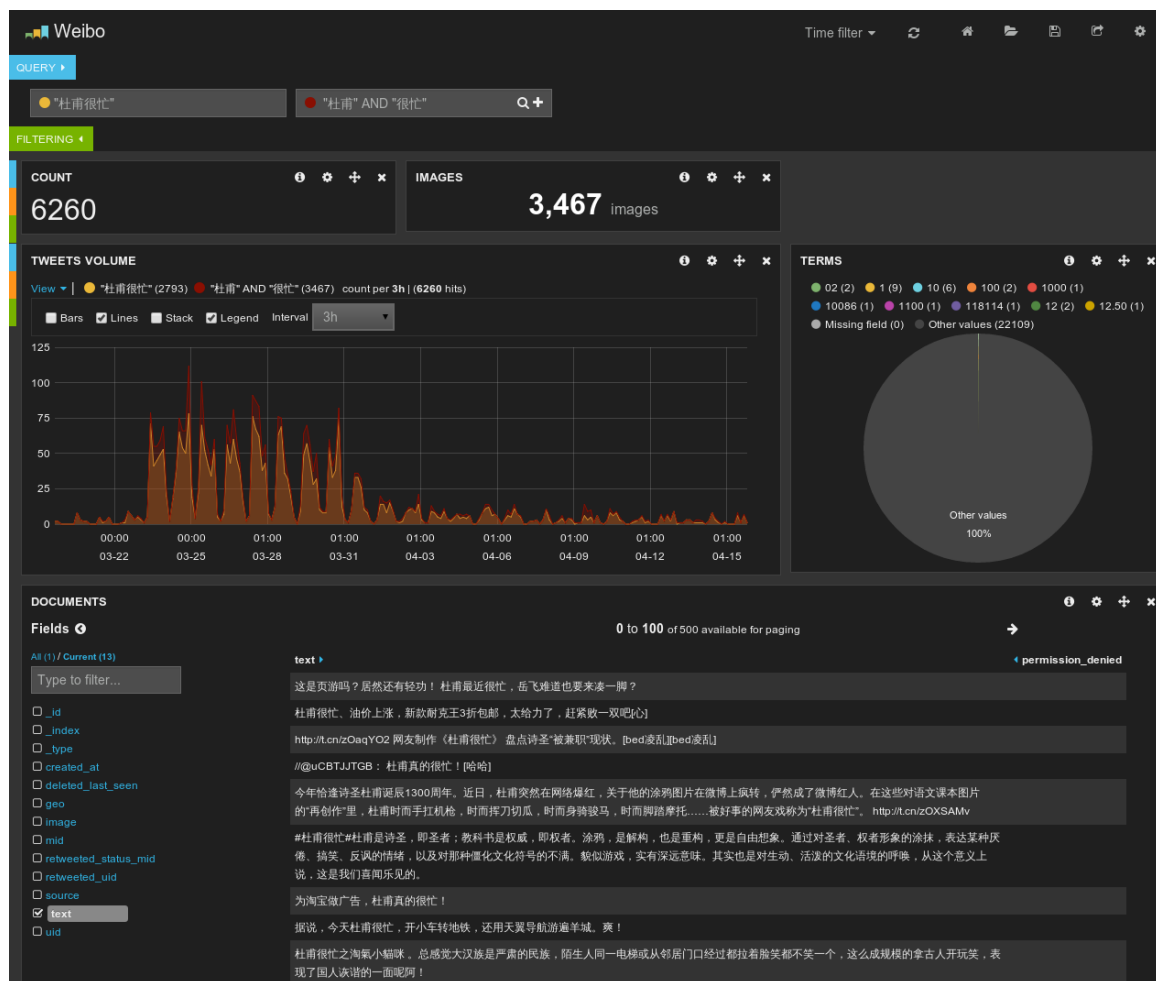
La définition de la requête répondant au mieux aux besoins du chercheur ne peut donc se faire a priori sans un certain nombre d'essais et tentatives. L'obtention de résultats satisfaisants passe par une phase itérative qui permet de définir et cerner une définition du mème sous forme d'une requête dans le moteur de recherche. Cette démarche implique non seulement la possibilité de pouvoir s'adonner à de multiples essais de formulations de la requête mais également la nécessité de lire les résultats sous une forme compréhensible. Si le format JSON reçu ou renvoyé par le moteur de recherche est un standard de l'échange de données, il n'est néanmoins pas adapté à la lecture et l'écriture par des humains. Ainsi, afin de pouvoir mener ce travail de façon intuitive et pratique, nous avons mis en place un tableau de bord à l'aide du logiciel *Kibana* qui nous permet de contrôler la pertinence des résultats en comparant plusieurs requêtes différentes (fig. 4.1.1).

## Constitution d'un corpus pour chaque mème

En contrôlant la qualité de ses différents critères nous pouvons donc identifier une requête adéquate pour décrire le mème. Une fois cette étape effectuée, nous procédons à l'extraction des messages qui vont venir constituer le corpus d'études de notre mème. Pour s'assurer de la fiabilité des résultats, nous disposons de l'index de poids du moteur de recherche qui représente la similarité entre notre requête et les messages obtenus. Pour limiter le nombre de résultats, nous pouvons fixer un indice de similarité minimum et ainsi limiter le bruit en ignorant les messages peu en lien avec notre requête.

Une fois extrait, ce jeu de données est stocké dans un fichier de type *csv* respectant le format initial et l'encodage des données de Weiboscope. Afin de parfaire sa complétude, les messages mentionnés et les retweets ne contenant pas le mot-clé (commentaires, réponses, etc.) sont rassemblés pour donner finalement un ensemble de messages représentatifs du mème qui va pouvoir être analysé.





**Graphique 4.1.1** – Ce tableau de bord permet de comparer la qualité de différentes requêtes dans le corpus. Capture d'écran réalisée le 23 Mars 2014 à 16h18

### 4.1.2 Analyse des graphes et traitement des données

Il s'agit pour nous d'observer plusieurs aspects importants de la diffusion des mèmes : champ sémantique, dynamique des discussions et dimension spatio-temporelle de ces échanges. Principalement, nous choisissons de représenter différents réseaux sémantiques, conversationnels et géographiques extraits du corpus de message. Les croisements de ces différents réseaux de diffusion peuvent mettre à jour des phénomènes intéressants, en ramenant de la géographie dans le langage et les mots des conversations.

Les choix méthodologiques de cette étude cherchent notamment à démontrer que les actes de communication ne peuvent simplement se comprendre comme des échanges “sociaux” mais doivent être appréhendés plus largement comme des actes d'énonciation complexes possédant de multiples dimensions sémantiques, temporelles, conversationnelles toutes localisées.

Nous définissons les trois dimensions de l'analyse comme suit :

**Langagier** Le champ sémantique d'un mème est constitué des mots qui sont prononcés lors de sa diffusion. L'association de mots - souvent sous la forme du jeu de mots - est un des propres du mème et définit ainsi une part importante de son existence. Ainsi, le mème produit à proprement parler des réseaux de mots en dessinant des liens entre des signifiants souvent improbables qui en font le succès (Bauckhage, 2011). Observer les étapes de la construction du réseau de mots peut donc apporter un éclairage nouveau.

**Conversationnel** Un mème se constitue sous la forme d'un échange. Les mots s'articulent dans une conversation où les différents acteurs discutent, commentent et se saisissent des actions disponibles sur la plateforme web (like, retweets, etc.) pour converser. Comme nous l'avons vu précédemment, nous pouvons identifier et considérer un graphe conversationnel créé par le mème en se diffusant pour identifier des structures de diffusion particulières.

**Réel** Au-delà des échanges en ligne, ces discussions possèdent une existence physique, premièrement sous la forme de l'activité électrique des machines qui sont utilisées lors de ce processus. Également, dans l'approche d'une géographie humaine des échanges numériques, nous considérons l'existence physique des mèmes par celle

des utilisateurs - de leurs corps.

Nous allons donc procéder à la collection de connaissances sur chacun de ces aspects d'après le corpus de données du ou des mêmes sélectionnés. Chaque corpus est traité selon une série de procédures que nous allons maintenant définir.

## Traitement naturel de la langue chinoise

Le texte de chaque message est d'abord analysé de façon à ne conserver que les éléments significatifs et utiles à l'analyse : mots importants, mentions d'utilisateurs, urls et hashtags. Les éléments propres à *Sina Weibo* peuvent aisément être identifiés à l'aide d'expressions régulières car ils suivent des patterns stricts :

- le hashtag débute et termine par le symbole # (ex. #hello#)
- les urls suivent une forme classique (ex. `http://t.cn/SVWAfP`)
- les mentions d'utilisateurs débute normalement par @ suivi du noms de l'utilisateur mentionné. Le corpus Weiboscope ayant été anonymisé, le nom des utilisateurs a été remplacé par un identifiant commençant par la lettre *u* suivi de 8 caractères (ex. *uY02ZFLAN*)

Le texte rédigé présent dans les messages est par contre un élément beaucoup plus complexe à traiter. De plus, la langue écrite chinoise, contrairement aux langues de l'alphabet latin, n'utilise pas d'espace pour séparer les mots d'une phrase. Ainsi, alors qu'il est plutôt simple de repérer les différents mots d'une phrase en langue française ou allemande en recherchant les espaces dans le texte, la phrase chinoise nécessite d'être segmentée en mots avant de pouvoir être traitée. Le nom commun chinois prend le plus souvent la forme d'un, deux ou trois caractères, voire plus pour de nombreux mots techniques, noms propres ou expressions plus complexes. La segmentation de la phrase chinoise est donc une première contrainte qu'il nous faut prendre en compte dans le cadre de cette étude, le chinois mandarin constituant le langage majoritaire de notre corpus.

Le domaine du *Natural Language Processing* (NLP) en chinois est un sujet de recherche encore très discuté (Qiu *et al.*, 2013). De nombreuses solutions, librairies et logiciels sont disponibles pour la segmentation des phrases chinoises et l'extraction de mots-clés notamment. L'identification de la meilleure option parmi la multitude des outils a constitué un

des aspects préliminaires importants de notre travail d'analyse de données. Nous avons bénéficié ici de l'aide de Yuan Mingli, directeur du centre de Recherche et Développement de la compagnie *Guokr*<sup>5</sup>. Situé à Beijing, *Guokr* est un site ressource pour la culture scientifique en Chine qui propose notamment des cours en ligne (MOOC) et des listes de discussions sur de nombreux sujets en lien avec les sciences et les technologies. Une partie de l'équipe de recherche dédie son travail au développement et à la maintenance de solutions pour le Web chinois, notamment pour l'analyse naturelle de la langue chinoise. Ayant travaillé auparavant sur des dispositifs d'analyse de données avec Yuan Mingli dans le cadre du projet *Sharism Lab*<sup>6</sup>, nous nous sommes rendus à Pékin lors d'un séjour de recherche en Septembre 2013 pour identifier les solutions les plus intéressantes concernant l'analyse de la langue chinoise.

Nous avons ainsi réalisé une comparaison des différents algorithmes de segmentation existants afin de définir le plus approprié à ce travail. Nous avons testé plusieurs solutions, dont plusieurs développées par l'équipe de *Guokr* elle-même :

**gkSeg** Cette solution utilise un algorithme appelé *Conditional Random Fields* qui s'intéresse à chacun des caractères. Il est conçu pour pouvoir être utilisé aussi bien sur des textes en chinois ancien<sup>7</sup>.

**Stanford CoreNLP Chinese Segmenter** Le célèbre outil de NLP développé par Stanford dans sa version 3.4 permet de découper la phrase chinoise. La librairie *stan-cn-seg*<sup>8</sup> permet de l'utiliser simplement sous la forme d'un serveur.

**neuseg** Ce protocole expérimental utilise les réseaux neuronaux et le modèle vectoriel de segmentation de la phrase chinoise<sup>9</sup>.

**jieba** *Jieba*<sup>10</sup> est un segmenteur open-source très répandu et utilisé commercialement dans de nombreux projets. Plus généreux en termes de fonctionnalités, il permet de meilleures performances sur de larges corpus. Il utilise des *tries* pour organiser

---

5. Voir le site officiel <http://www.guokr.com/> consulté le 5 Juillet à 15h45

6. Site officiel <http://sharismmlab.com>, consulté le 6 Juillet 2014 à 13h46

7. gkSeg : “Yet another Chinese word segmentation package based on character-based tagging heuristics and CRF algorithm” <https://github.com/guokr/gkseg>, consulté le 14 Mars à 18 :28

8. <https://github.com/guokr/stan-cn-seg> consulté le 7 Juillet à 16 :01

9. <https://github.com/guokr/neuseg> consulté

10. <https://github.com/fxsjy/jieba>, consulté le 14 Mars à 18 :30

des *graphes orientés acycliques* recréant l'ordonnancement des mots de la phrase. Les erreurs dûes aux mots inconnus sont détectées et corrigées grâce à l'algorithme *Viterbi*.

Après plusieurs tentatives, nous avons préféré le segmenteur *Jieba*, plus adapté aux dimensions de notre large corpus. *Neuseg* n'était pas assez mature et ne permettait pas d'obtenir des résultats satisfaisants, connaissant beaucoup d'erreurs et manquant de stabilité. *Stanford NLP* nous contraignait à l'usage d'un serveur externe rendant le déploiement plus complexe et l'ensemble plus compliqué à maintenir. *Gkseg* ne présentait pas de performances suffisantes en terme de temps de traitement, devenant handicapant pour notre étude. De plus, *Jieba* se présente sous la forme d'une librairie du langage *Python*.

---

**Algorithm 3** Extract Words from Message

---

**Require:**  $m$  is a microblog text message

**Require:** *stopwords* is a list of common words that should be excluded

**Require:** JIEBA NLP(*text*) is a Chinese language word segmenter

$words \leftarrow$  JIEBA NLP( $m$ ) segment the sentence

remove *stopwords* from *words*

---

La phrase est donc segmentée pour obtenir un ensemble de mots du même (voir algo 3). Néanmoins, de nombreux mots dans une phrase sont redondants. Afin de s'intéresser seulement aux mots les plus importants, il est donc important de supprimer les mots les plus communs, appelés *stopwords*<sup>11</sup>. La première forme de stopwords est celle propre à la langue chinoise. De nombreux mots très usités ne possèdent que peu d'information pour notre étude et peuvent donc être supprimés (articles, pronoms, etc.). Nous utilisons pour cela la fonctionnalité de *Jieba* qui propose un dictionnaire comprenant caractères traditionnels et simplifiés et permet d'enlever les mots les plus courants. Une seconde liste de stopwords spécifiques à *Weibo* a été ajoutée pour parfaire le tri des informations. Cette liste a été constituée lors du projet commercial *Echidna* réalisé en Avril 2013 à Shanghai avec l'équipe constituée par Ricky Ng-Adam, ex-ingénieur chez *Google* à Mountain View (USA) pour le compte de la société *Transist* (创思). *Echidna* était une solution permettant de connaître les tendances de mots-clés grâce à un tableau de bord et une infrastructure d'analyse temps-réel de données de *Tencent Weibo*. Lors de ce projet, de nombreux outils et designs ont été développés. Une partie de la logique utilisée pour la communication

---

11. La liste des mots ignorés (*stopwords*) dans cette recherche est disponible en ligne à <https://github.com/clemsos/mitras/tree/master/lib/stopwords>

entre les différentes parties du système ainsi qu’une liste de stopwords constituée sur une période de plusieurs mois ont été réutilisés dans le présent travail.

## Graphe lexical

Après avoir extrait l’ensemble des entités pour chaque message, nous allons maintenant constituer le graphe sémantique représentant les relations entre les mots-clés du mème. Un graphe pondéré et non-dirigé représente les co-occurrence des 500 mots les plus importants dans le corpus de messages.

---

### Algorithm 4 Word Graph extraction from Meme messages corpus

---

**Require:**  $M$  is a set of messages representing a meme  
 $topWords$  is the 500 most used words in  $M$   
 $G_w(N_w, E_w)$  is a weighted directed graph of words co-occurrence

```

function CREATEWORDGRAPH( $M$ )
  for message  $m$  in  $M$  do
    EXTRACTENTITIES( $m$ )  $\leftarrow words$ 
    if  $word$  in  $topWords$  then
      for  $word_A, word_B$  in  $words$  do
        if  $\exists G_w(word_A, word_B)$  then
           $G_u(user_A, user_B)$  weight +1
        else
           $G_w \leftarrow (word_A, word_B, 1)$ 
        end if
      end for
    end if
  end for
end function

```

---

Chaque mot est un nœud dans le réseau. La co-occurrence de deux mots dans une même phrase définit une relation entre ces mots. L’ensemble des relations constitue le réseau sémantique. Nous limitons également le graphe aux 500 mots les plus utilisés (dont les occurrences sont les plus nombreuses) afin de préserver la lisibilité lors de l’étape suivante de visualisation.

## Graphe conversationnel

Les actions effectuées par les utilisateurs lors des conversations (mentions, citations et retweets) nous permettent de reconstituer le graphe conversationnel des échanges. Chaque

utilisateur  $y$  est un node et chaque interaction un edge. Le graphe est dirigé car les interactions vont depuis un utilisateur à un autre et pondéré afin de pouvoir retranscrire l'intensité des échanges. La taille du graphe est également limitée aux utilisateurs ayant effectué au moins 2 échanges - en ignorant les *edges* du graphe ayant un poids inférieur à 2. Seuls les 500 utilisateurs les plus actifs sont utilisés pour l'analyse afin de réduire le temps de traitement et la complexité du réseau pour la visualisation. Pour procéder à cette limitation, nous constituons une *whitelist* des 500 utilisateurs les plus actifs qui autorise ou empêche l'ajout d'utilisateurs au réseau conversationnel.

---

**Algorithm 5** Extract User Graph from Meme Corpus

---

**Require:**  $M$  is a set of messages representing a meme  
 $topUsers$  is the 500 most active users in  $M$   
 $G_u(N_u, E_u)$  is a weighted directed graph of user conversations

```

function CREATEUSERGRAPH( $M$ )
  for message  $m$  in  $M$  do
     $E$  edges in the message
    EXTRACTENTITIES( $m$ )  $\leftarrow$   $mentions, RTuser, author$ 
    if  $author, RT, mentions$  in  $topUser$  then
      if  $RTuser$  in  $m$  then
         $E \leftarrow e(RTuser, author)$ 
      end if
      for  $user$  in  $mentions$  do
         $E \leftarrow e(author, @user)$ 
      end for
      for  $edges$  in  $E$  do
         $G_u \leftarrow (e, weight_e + 1)$ 
      end for
    end if
  end for
end function

```

---

Une fois l'ensemble du graphe constitué, nous utilisons l'algorithme dit de Louvain et son implémentation par Blondel *et al.* (2008)<sup>12</sup> pour identifier les communautés dans notre réseau. Cet algorithme relativement simple attribue à chaque utilisateur un groupe unique. De plus, il est plus rapide que d'autres algorithmes répandus pour cet usage comme *k-means* par exemple.

---

12. Pour la version Python voir : <http://perso.crans.org/aynaud/communities/index.html> consulté le 22 Avril 2014 à 14 :24

## Localisation des utilisateurs

Les données de géo-localisation contenues dans le corpus Weiboscope sont seulement partielles et ne permettent pas d'utiliser le géo-tag attaché à chaque message comme support pour l'analyse. En effet, très peu de messages sont géo-localisés. Néanmoins, une annexe donne accès aux localités mentionnées par les utilisateurs dans leurs profils. Cette information doit être utilisée avec précaution : les utilisateurs peuvent indiquer ce que bon leur semble et mettent rarement à jour leurs informations de profil lors de leurs déplacements, voire leurs déménagements. Néanmoins, la présence d'une information de vérification du compte dans les données permet de savoir si le compte a été vérifié par la firme *Sina*. Il s'agit bien souvent de personnalités publiques, de stars ou de compagnies privées qui possèdent des comptes officiels. Pour ce type d'utilisateur, l'information de localisation revêt plus d'intérêt car elle a été précédemment acceptée comme valide. Le nombre des localités proposées par *Sina Weibo* aux utilisateurs est restreinte par l'interface elle-même : l'ensemble des provinces de Chine continentale, Hong Kong, Macau, Taiwan, “à l'étranger” (*haiwai*) et “autres” (*qita*)<sup>13</sup>.

Afin d'observer les aspects géographiques des discussions, nous assignons à chaque utilisateur l'information géographique mentionnée dans son profil. Afin de pouvoir accéder facilement à cette information, une base de données utilisant *MongoDB* regroupe toutes les informations de localisation des utilisateurs disponibles, accessibles grâce à leur nom d'utilisateurs<sup>14</sup>.

### 4.1.3 Serveur et moteur de visualisation interactive

Une fois l'ensemble des connaissances rassemblées, une interface de visualisation nous permet d'explorer et d'observer les différents aspects des processus de diffusion.

---

13. Voir la liste des choix disponibles sur <https://github.com/clemsos/mitras/blob/master/lib/cities/provinces.csv>, consulté le 7 Juillet 2014

14. Voir le code <https://github.com/clemsos/mitras/blob/master/lib/users.py>, consulté le 8 Juillet 2014 à 17 :22



## Enjeux de la visualisation

Plusieurs niveaux de lecture interdépendants nous permettent de considérer différents aspects de chaque mème. Les informations de graphe, peu claires et difficilement exploitables sous forme de données, doivent faire l'objet d'une représentation graphique. La visualisation permet d'explorer les différentes dimensions du mème, articulées à quelques mesures indicatives sur la structure des réseaux.

- **Sémantique** mots-clés et relations entre les mots
- **Conversationnel** échanges et communautés d'utilisateurs
- **Socio-sémantique** relations entre les mots et les utilisateurs
- **Socio-géographique** relations entre les utilisateurs et les provinces
- **Géo-sémantique** relations entre les mots et les provinces

Aucun outil existant ne permet de réaliser cette cartographie particulière. Nous avons donc choisi de développer une solution adaptée pour considérer sous différents angles l'ensemble des graphes obtenus. Les choix conceptuels et technologiques présidant au design de cet outil de visualisation sont détaillés ci-après.

Dès lors que se pose la question de structurer un espace de perception, nous allons au devant de plusieurs difficultés formelles majeures. Afin de pouvoir considérer les relations entre mots, communautés et territoires dans l'espace graphique disponible, il est nécessaire de résoudre plusieurs problèmes :

**Unité** Donner à voir chaque type d'entités de façon reconnaissable (provinces chinoises, communautés, mots, temps)

**Cohésion** Donner à voir les relations entre les différentes entités (graphe multiples)

**Cohérence** A l'image du milieu numérique, l'existence même de ce plan réconciliant global, local, réel, symbolique et imaginaire pose question. Quelle doit être sa couleur ? Quelle peut en être la justification logique dans l'espace de la visualisation ? Et que penser de l'écran ? De quoi est-t-il le bord ?

**Complexité** Pallier l'augmentation de la complexité visuelle et préserver une lisibilité

**Publication** L'exigence de rigueur et la rigidité des formats de publications scientifiques rajoutent à la complexité de cette entreprise. Il est indispensable de pouvoir exporter les résultats sous forme de figures de façon propre et ordonnée.

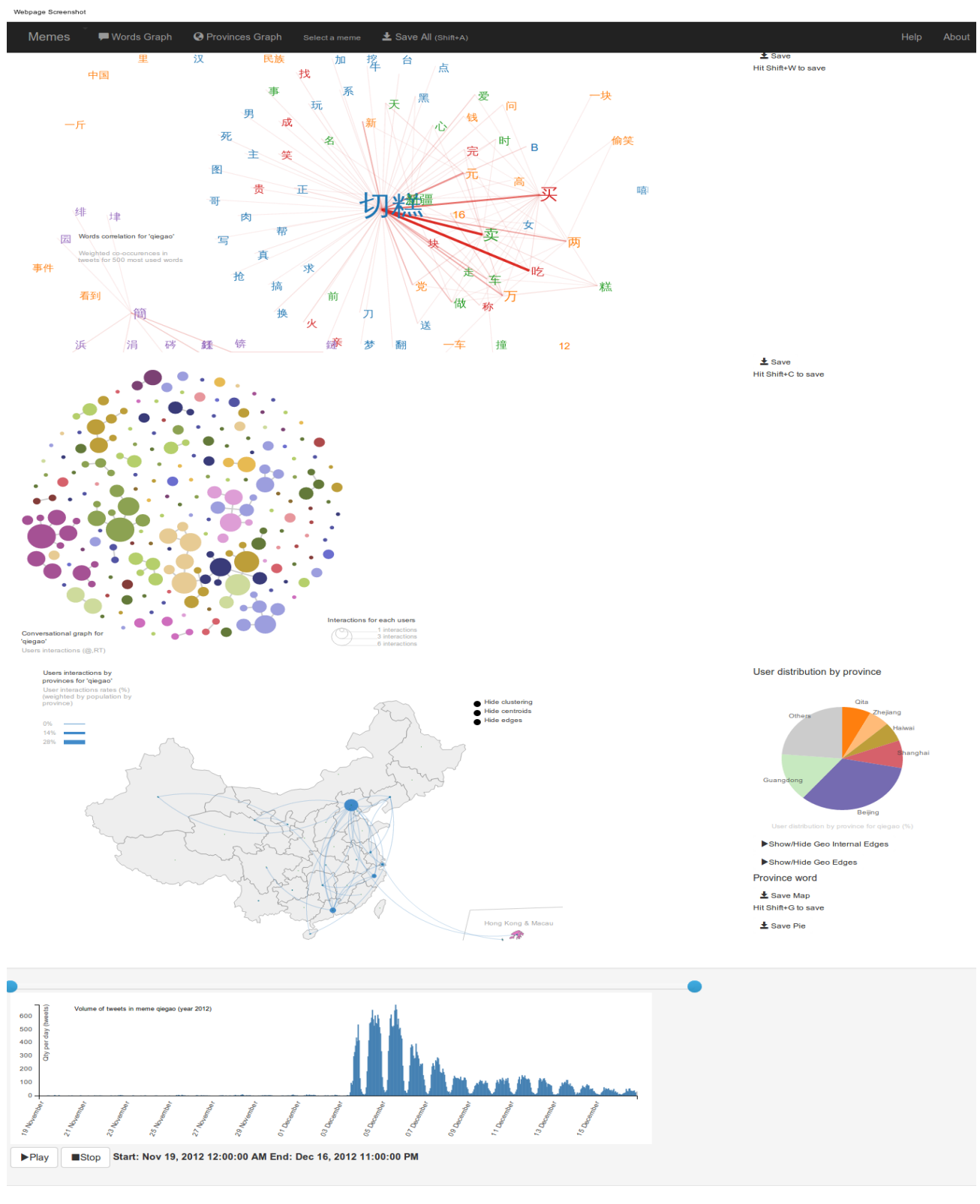
## Organisation de l'espace visuel et perceptif

Nous avons donc choisi de construire un espace sur plusieurs niveaux .

Devant la disparité des données en présence, l'espace perceptif doit être hiérarchisé pour considérer des entités dissemblables (utilisateurs, lieux, mots et leurs relations mutuelles). La tâche consiste à réconcilier dans le plan de l'espace graphique de la visualisation (ici celui de l'écran) les différents graphes. Ici, nous devons prendre en compte la nature des données affichées. La première décision est d'utiliser un défilement vertical qui permet de naviguer entre les différents graphes. Nous structurons l'espace en plusieurs parties clairement identifiées.

Le centre de la visualisation est occupé par le réseau représentant les utilisateurs. En structurant l'espace graphique autour des utilisateurs, nous souhaitons montrer le rôle central des processus humains dans cette étude. Dans notre représentation, chaque utilisateur est un point et chaque interaction correspond à un trait. La couleur des points montre l'appartenance de l'utilisateur à une même communauté de conversation calculée grâce à un algorithme de Louvain (Blondel *et al.*, 2008). La taille des points exprime le nombre total de connexions entrantes et sortantes d'un utilisateur. La distance entre les groupes d'utilisateurs est définie en utilisant l'algorithme *force* de *d3.js* (Bostock *et al.*, 2011) pour calculer leur répulsion : plus les utilisateurs sont proches dans la conversation, plus les points les représentant sont proches dans le graphe. La forme circulaire du graphe est due aux propriétés physiques d'attraction utilisées lors du calcul qui permettent aux différents nodes du graphe de rester agrégés ensemble sans se disséminer. Pour des questions de lisibilité, seuls les 500 utilisateurs les plus actifs ont été représentés. Les utilisateurs possédant moins de deux interactions ont été supprimés pour rendre le graphe lisible.

Le graphe de mots occupe lui la position haute. Éthérique et peu enracinée, la représentation de ce nuage sémantique trouve sa place dans le haut de l'espace visuel permettant de saisir d'un coup d'œil la teneur générale des débats pour s'attarder sur les nœuds mystérieux créés dans le langage lors des conversations. Les graphes de mots présentés sont construits grâce aux co-occurrences de mots dans les messages. Si deux mots sont utilisés dans un même message, un *edge* est ajouté entre eux, recréant ainsi un graphe pondéré représentant une structure sémantique relationnelle d'ensemble entre les mots. La taille



**Graphique 4.1.2** – Interface d'exploration et de visualisation des données. Capture d'écran réalisée le 23 Mai 2014 à 16h18

Menu	
Words Semantic graph	Options
Users Conversational graph	
Geo Map of China	
Time Timeline	

**Graphique 4.1.3** – Structure de l'interface d'exploration et de visualisation des données

des mots représente le nombre de fois où il sont cités dans l'ensemble du corpus. Les couleurs définissent des communautés de mots qui ont été calculées grâce à l'algorithme de Louvain (Blondel *et al.*, 2008). L'épaisseur des traits représente l'intensité des relations (leur nombre de co-occurrences dans le corpus). La disposition des mots utilise également l'algorithme *force-based* de d3js (Bostock *et al.*, 2011) pour calculer la proximité des mots sur la base de l'intensité de leurs relations. Afin de limiter la taille du graphe, nous avons sélectionné uniquement les 500 mots les plus utilisés.

Plancher des vaches, la carte géographique représentant la Chine est positionnée en-dessous du graphe des utilisateurs (voir section suivante 4.1.3). Enfin, l'axe temporel se situe tout en bas de la page, montrant le volume des conversations dans le temps. Sur la droite, un espace est laissé disponible pour permettre l'ajout de différentes options et l'affichage d'informations secondaires.

## Cartographie et visualisation interactive

La représentation des échanges entre les utilisateurs sur une carte géographique a présenté plusieurs difficultés. La première étape a été de reconstituer une carte de la Chine par provinces comprenant également Taiwan, Hong Kong et Macau. Chacun de ces territoires possède une influence médiatique certaine en Chine et mérite à ce titre d'être représenté. De plus, il était important de faire correspondre la carte aux informations géographiques disponibles dans le jeu de données (voir chapitre 4.1.2). Néanmoins, le statut



**Graphique 4.1.4** – Carte interactive de la Chine (exemple contenant des données générées aléatoirement)

politique particulier de chacun de ces territoires nous a obligé à reconstituer une carte où ils figuraient tous. Nous avons pour ce faire utilisé les fichiers *Shapefile* disponibles via *Natural Earth* (Administrative 1 :10m)<sup>15</sup> pour constituer un seul et même fichier compatible au format *geoJson* utilisant les noms standard *ISO 3166-1 alpha-3* pour nommer les différents lieux.<sup>16</sup> Nous avons choisi d’agrandir l’échelle de Hong Kong et Macau en les mettant en exergue sur la droite de la carte afin qu’ils soient visibles à l’échelle choisie pour le reste du territoire chinois. Également, nous avons choisi d’ignorer les données concernant les utilisateurs ayant choisi “autres” ou “reste du monde” comme leurs lieux de résidence dans leurs informations de profil.

La représentation des dynamiques entre province se base sur les échanges entre utilisateurs. Nous avons précédemment extrait pour chaque même le réseau conversationnel des interactions entre utilisateurs. En associant chaque utilisateur à sa province d’origine, nous pouvons reconstituer un réseau des interactions entre provinces.

L’agencement spatial des discussions a d’abord été représenté par la quantité des utilisateurs par province. Néanmoins, cette carte montre avant tout les grandes villes et contient donc très peu d’informations. Pour faire face à ce biais, nous avons d’abord choisi

15. <http://www.naturalearthdata.com/>, consulté le 9 Juillet 2014 à 10 :42

16. Le détail des procédures utilisées ainsi que le code pour générer la carte sont disponibles à <https://github.com/clemsos/d3-china-map>, consulté le 9 Juillet 2014 à 10 :41

de représenter les relations entre provinces sous formes d'une liste de noms reliés par des traits. La figure obtenue, complexe et peu lisible, a donc été remplacé par la projection du réseau pondéré et dirigé des interactions sur une carte. Pour des raisons de lisibilité, les liens entre les provinces sont dirigés non pas vers la capitale de la province mais vers le *centroid* (barycentre) de la forme géométrique qui la représente. Afin de rétablir un équilibre et de faire apparaître des dynamiques plus particulières, les résultats ont été pondérés par le pourcentage de chaque province dans la population totale.

Des interactions permettant de focaliser sur des zones ou des caractéristiques précises des graphes ont été ajoutées pour explorer les différentes dimensions des graphes. Il est également possible d'exporter un ou plusieurs graphes sous forme d'images au format *PNG Portable Network Graphic* pouvant être ensuite utilisées comme figures (voir section 4.2). Des raccourcis claviers offrent la possibilité de capturer des instantanés obtenus par la manipulation des graphes.

## Moteur de visualisation

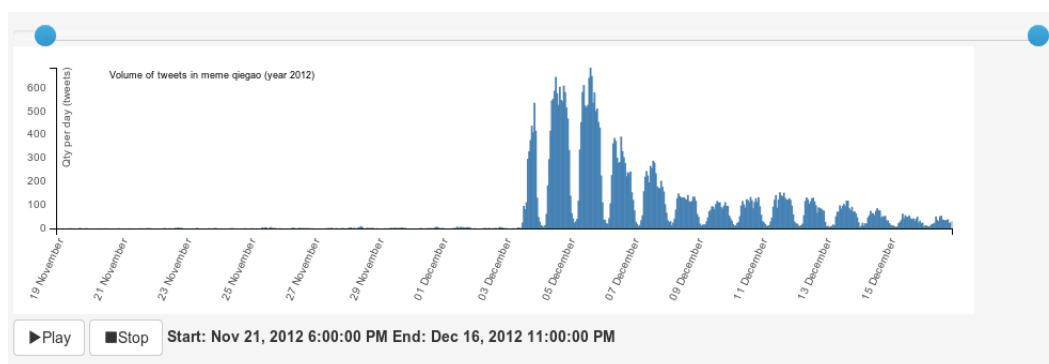
Les protocoles de communication entre base de données, serveur et affichage des graphes sont une des clés qui rendent possible la visualisation interactive. En effet, afin de pouvoir se concentrer sur des résultats ou des éléments spécifiques, il est nécessaire de pouvoir isoler un ensemble de données intéressant. L'utilisateur doit d'abord sélectionner un mème parmi ceux existant dans la base de données. Il dispose d'un menu décrivant succinctement le mème par un titre et les mots clés utilisés lors de la recherche qui lui permettent de le reconnaître. En cliquant sur les boutons du menu (fig. 4.1.5), l'utilisateur accède à une page spécifique présentant le mème (fig. 4.1.2).



**Graphique 4.1.5** – Menu de sélection des mèmes

Une fois le mème affiché, l'utilisateur peut sélectionner différentes parties qui l'inté-

ressent. Une des fonctionnalités importantes de ce logiciel est la sélection d'une période de temps à afficher. Déplacer le curseur situé au dessus du graphe temporel permet de choisir un temps de début et de fin. L'affichage est automatiquement mis à jour pour montrer uniquement la période sélectionnée. Un bouton *play* permet également d'animer automatiquement les graphes, montrant leurs évolutions sur la période de temps choisie (figure 4.1.6).



**Graphique 4.1.6** – Cette timeline interactive permet de sélectionner une période de temps

Cette fluidité dans l'interface est obtenue par la synchronisation de nombreux composants permettant : 1) la communication avec la base de données, 2) l'obtention des données suivant la limite de temps et 3) la mise à jour de l'affichage. Pour ce faire, nous utilisons des technologies basées sur le langage *Javascript*. Le serveur chargé de communiquer avec la base de données est une application écrite en *Node*<sup>17</sup> qui est un framework généraliste de Javascript pour les serveurs. L'ensemble de l'outil de visualisation dans le navigateur se fonde sur HTML5, CSS3 et Javascript. La gestion des données dans le navigateur et la communication avec le serveur se fait grâce à la librairie *AngularJS*<sup>18</sup>. Conçu par *Google*, *AngularJS* nous permet de standardiser et d'optimiser les fonctionnalités de traitement des données côté client. La visualisation des données utilise la librairie "*Data-Driven Documents*" (*Bostock et al.*, 2011)<sup>19</sup>. Les menus et boutons sont construits à l'aide du framework *Bootstrap*<sup>20</sup>. L'ensemble du schéma de communication entre le serveur et le moteur de visualisation est disponible dans les annexes de ce document (figure .0.2).

17. "Node.js is a platform built on Chrome's JavaScript runtime for easily building fast, scalable network applications." d'après <http://nodejs.org/>, consulté le 8 Juillet à 15 :05

18. "HTML enhanced for web apps" d'après <https://angularjs.org/>, consulté le 8 Juillet à 15 :08

19. D3js at <http://d3js.org/>, consulté le 24 Avril 2014 à 14 :58

20. "Bootstrap is the most popular HTML, CSS, and JS framework for developing responsive, mobile first projects on the web." d'après <http://getbootstrap.com/>, consulté le 8 Juillet 2014 à 15 :16

Ces technologies sont utilisées en production par des sites possédant un grand nombre d'utilisateurs (comme *Twitter* pour *Bootstrap* ou *Youtube* pour *Angular*). Cette grande fiabilité accompagnée d'une documentation très complète a permis d'obtenir des résultats rapides lors des processus itératifs du prototypage. Enfin, l'usage de bibliothèques déjà reconnues rend cet outil plus lisible pour les utilisateurs ou ré-utilisateurs qui voudraient s'en saisir.



## 4.2 Résultats sur un échantillon de mèmes

Nous présentons maintenant les résultats obtenus grâce à notre dispositif d'analyse sur un échantillon d'une dizaine de mèmes. La typologie de mèmes identifiée précédemment (voir fig. 2.3.1) a servi de point de départ pour cette sélection. Une recherche mêlant littérature scientifique et secondaire (journaux, blogs, encyclopédies en ligne, sites ressources) ainsi que notre propre expérience du web chinois nous a permis de définir un panel de mèmes observables dans par notre corpus (voir annexe ).

Nom	Description
<i>Sextape</i>	Un scandale d'adultère concernant un homme politique dans la ville de Chongqing dans le centre de la Chine.
<i>The Voice</i>	La première saison d'une émission de télé-crochet musical en Chine.
<i>Qiegao</i>	Un débat de société sur la condition du peuple Uyghur suivant un fait divers autour d'une rixe lors d'une vente de gâteau.
<i>Dufu</i>	Un mème comique mettant en scène un poète chinois dans des situations burlesques.
<i>Biaoge</i>	Un scandale entourant la collection de montres de marques prestigieuses d'un homme politique de la province du Shaanxi
<i>Moyan</i>	L'attribution du Prix Nobel de littérature à l'écrivain chinois Mo Yan
<i>Yuanfang</i>	La reprise d'une citation d'une série télévisée sous forme humoristique
<i>Ccp</i>	Les débats entourant le 18ème Congrès du Parti Communiste Chinois

**Tableau 4.2.1** – Dénomination et description des mèmes étudiés

Dans un premier temps, nous avons listé différents mèmes et événements web de l'année 2012 sur *Sina Weibo* (voir annexe .0.1). Ensuite, nous avons cherché à constituer un ensemble cohérent montrant la diversité des contenus pour l'étude (voir 4.2.1). Les différentes sources consultées proposent des définitions très différentes de l'objet mème. De nombreux sites Web effectuent a posteriori une revue des faits et contenus qu'ils considèrent comme les plus marquants de l'Internet chinois en 2012. Micro-historiographie du web, ce type de classification affirmant l'importance supposée des événements web reflète néanmoins plutôt l'orientation éditoriale de l'auteur ou du média qu'une réalité historique. Sur les sites anglophones, les contenus marquant du web social chinois en 2012 ont tous un caractère politique et polémique. Les listes les plus largement reprises ont été constituées

par des sites généralistes comme *Global Voices*<sup>21</sup> ou le *Wall Street Journal*<sup>22</sup>. Largement repris par des professionnels, journalistes ou blogueurs, ces supports influents proposent une sélection de mèmes tournés vers la satire politique. Plusieurs sont directement liés à des campagnes de soutien pro-démocratie dont la diffusion dans le paysage médiatique contrôlée de la Chine reste extrêmement marginale, voire souterraine. D'autres actualités fortes ont donné lieu à des débats d'ampleur nationale. Les faits divers et scandales éblouissant des hommes politiques chinois impliquent très largement l'audience des réseaux sociaux, qui commentent abondamment sur la situation politique alarmante et la corruption des gouvernements locaux.

L'affaire Yang Dacai, baptisé "Brother Watch" (*biaoge*, 表哥), a notamment fait couler beaucoup d'encre numérique sur les écrans chinois. En Août 2012, Yang Dacai, directeur du Bureau de Supervision de la Sécurité Routière de la province du Shanxi, se rend sur les lieux d'un accident de bus ayant fait 36 morts. Une photo diffusée sur Internet le montre discutant avec un policier, arborant un grand sourire. Agacés par cet excès d'incivilité, les internautes commencent à faire des recherches à son sujet et remarquent que l'officiel chinois arbore à chaque apparition une montre différente, toutes de marques prestigieuses. Il est rapidement surnommé *le frère aux montres* (*biaoge*). Les collections de photos de ses différentes montres de luxe circulent comme autant de preuves flagrantes de sa corruption. Quelques jours plus tard, Yang Dacai est démis de ces fonctions. Il sera par la suite jugé et condamné pour corruption, assorti d'une peine de 12 ans de prison (voir les détails dans l'annexe ).

Parmi les autres affaires remarquables en 2012 se trouve celle de la *sextape* montrant Lei Zhengfu, le secrétaire du Parti dans le district de Beibei à Chongqing dans le plus simple appareil en compagnie d'une de ses maîtresses. A peine 63h après la publication de cette vidéo avec le titre "*Lei, the secretary who accepts sex bribes*", Lei fut mis à pied de ses fonctions politiques. La rapidité de cette action est alors présentée publiquement comme le résultat du nouveau programme de *surveillance du réseau par l'opinion publique*<sup>23</sup>. Cette vidéo qui avait été tournée 4 ans auparavant était utilisée depuis plusieurs années par

---

21. Global Voices, <http://globalvoicesonline.org/2012/12/07/top-10-chinese-internet-memes-of-2012/>, consulté le 22 Avril 2014 à 12 :10

22. Wall Street Journal <http://blogs.wsj.com/chinarealtime/2012/12/19/the-top-10-chinese-internet-memes-of-2012/>, consulté le 22 Avril 2014 à 12 :10

23. <http://www.chinanews.com/gn/2012/11-24/4354915.shtml>, consulté le 9 Juillet à 12 :41



(a) La photo montrant Yang Dacai “frère montre” souriant sur la scène du crash.



(b) Un an plus tard, Yang Dacai plaide coupable pour corruption lors de son jugement

**Graphique 4.2.1** – Illustrations du même *biaoge* d’après le Guardian <http://www.theguardian.com/world/2013/sep/05/china-brother-wristwatch-yang-dacai-sentenced>, consulté le 7 Juillet 2014 à 12 :11. et la BBC <http://www.bbc.com/news/world-asia-china-23956170>, consulté le 7 Juillet 2014 à 12 :23.

un promoteur immobilier de la ville de Chongqing pour faire chanter Lei Zhengfu (voir annexe ).

Un dernier fait divers n’impliquant cette fois pas des hommes politiques mais des citoyens chinois de l’ethnie ouïghour a également retenu notre attention. Originaire d’Asie centrale et orientale, le peuple ouïghour vit dans la région du Xinjiang, en Chine occidentale. Turcophones et majoritairement musulmans, les us et coutumes de ce peuple entrent bien souvent en contradiction avec ceux des chinois orientaux. La forte présence militaire chinoise dans la région et le ressenti qui anime les relations entre les peuples chinois et ouïghour conduisent régulièrement à des affrontements. Cette forte tension est également palpable dans les grandes villes de la côte comme Canton, Shanghai ou Pékin où les habitants du Xinjiang ont immigré en masse. La plupart travaillent dans des petits commerces de bouche. Une des spécialités de la région du Xinjiang est un délicieux gâteau à base de riz gluant, de noix et de jujube nommé *qiegao* (切糕). Vendu le plus souvent dans la rue, il se coupe difficilement. Son prix est plutôt élevé (jusqu’à 10 euros le kilo). Les vendeurs sont connus pour jouer de cette découpe compliquée comme excuse pour servir de généreuses portions aux clients, parfois plus larges que celles demandées. Une part un peu trop grosse découpée par un des vendeurs a mené à une rixe avec un client chinois fort mécontent. Escaladant en bagarre générale, cette altercation s’est terminée par l’arrivée

de la police qui a séparé rapidement les belligérants. Cette histoire fort banale ce serait arrêtée là si les internautes n'avaient pas eu vent du prix de 200 000 RMB (25 000 euros) réclamé par les vendeurs en dédommagement de leur marchandise et des dommages causés (blessures, motos brisées, etc). Parcourant rapidement la toile, cette nouvelle soulève de nombreuses questions des internautes, raillant sur le coût faramineux du gâteau : *“A piece of Xinjiang nut cake about 1.6 square meters in size cost RMB 160,000, which means about RMB 100,000 per square meter. Every 1 square meter of Xinjiang nut cake can buy about 3 square meters of apartment in Beijing.”*<sup>24</sup>. Entre commentaires racistes, discussions houleuses et blagues de tous bords, ce fait divers du *qiegao* amène à un débat national de plusieurs jours sur la condition des minorités ouïgoures dans les villes chinoises impliquant médias, universitaires et citoyens.



(a) Une image de l'affrontement avec la police posté par un utilisateur de Sina Weibo

(b) Création d'un internaute montrant combien le *qiegao* est cher

#### Graphique 4.2.2 – Quelques images du même *qiegao*

Note étude sur les hashtags a montré que les réseaux sociaux en Chine sont également le lieu de vastes campagnes commerciales. Dans les plus grands événements du Web en Chine, le site spécialisé *Danwei*<sup>25</sup> note que de nombreuses émissions de télévision ont généré une participation considérable sur Internet également. Ainsi, nous avons choisi pour représenter cette catégorie la version chinoise de l'émission internationalement connue de

24. d'après Off Beat China <http://offbeatchina.com/an-unbelievably-expensive-piece-of-xinjiang-nut-cake-and-what-it-tells-about-the-ethnic-policy-in-china>, consulté le 7 Juillet à 13 :02

25. Danwei <http://www.danwei.com/chinas-hottest-styles-of-2012/> consulté le 22 Avril 2014 à 12 :12

radio-crochet “The Voice” (中国好声音, *zhongguo hao shengyin*). Ayant battu en 2012 tous les records d’audience, *The Voice of China* offre un exemple parlant des grandes campagnes trans-médiatiques qui touche *Sina Weibo*.

Si la représentation des réseaux sociaux chinois comme un outil d’expression politique et de domination médiatique est souvent portée par les médias étrangers, les médias chinois dont *Baidu* publient une sélection des mots les plus marquants à chaque fin d’année<sup>26</sup>. Le mot *mème* ne possède pas de traduction chinoise à proprement parler (Renaud, 2014). Ainsi, la sélection de fin d’année se fait la plupart du temps sous la forme d’une liste de “mots à la mode” (流行语, *liuxingci*). Aucune des campagnes pro-démocratie ni aucun des “événements web” considérés comme marquants par les médias généralistes occidentaux n’est présent dans cette liste. On y retrouve par contre un humour noir et sarcastique, un des traits caractéristiques de l’Internet chinois d’aujourd’hui. Ces expressions humoristiques en disent long sur la société chinoise moderne. *Yalishan da* (压力山大) est un homophone du nom chinois d’Alexandre le Grand mais signifie littéralement “La montagne du stress est grande.”. Ce jeu de mot invoquant la figure mythologique du conquérant témoigne de l’immense pression pour la survie et la réussite qui s’exerce au quotidien sur les jeunes urbains chinois. Ici, le rire caustique qui fait de ce simple jeu de mots un succès sur l’Internet semble agir comme un exutoire dans la compétition sans relâche que réclame la société chinoise moderne.

Les interrogations sur la réalité quotidienne se poursuivent avec en seconde place du classement des mots de l’Internet en 2012 la question “*Es-tu heureux ?*” (你幸福吗). Avant de devenir une des citations les plus commentées de l’année 2012, cette question fut posée lors d’une interview sur la première chaîne officielle chinoise *CCTV1* au prix Nobel de littérature chinois Mo Yan. Sa réponse incluant “*avoir une belle maison*” et “*une bonne éducation*” déclencha de longues discussions et débats en ligne sur l’infinie question de la définition du bonheur dans une Chine en plein développement (“2012年，开发商，你们幸福吗?”).

Également, de nouveaux mots ont fait leur apparition comme le très notable *diaosi* (屌丝) aujourd’hui largement usité dans le langage commun<sup>27</sup>. Depuis 2010, ce mot désigne

---

26. Lire *Les 10 mots de l’Internet en 2012* sur *Baidupedia* <http://baike.baidu.com/view/9710412.htm>, consulté le 9 Juillet 2014 à 10 :32

27. “ a buzzword, referring to a young man born in a humble family, with mediocre appearance and



un jeune homme d'origine modeste, célibataire, sans argent ni bien, qui travaille dans un métier sans intérêt ni avenir, souvent derrière un ordinateur. Proche de l'idée de *No Life* en anglais ou de *otaku* en japonais, *diaosi* porte également en lui le mal-être des jeunes générations d'urbains désargentés, face à l'impossibilité de trouver un métier ou une vie familiale plaisante.

Ces quelques exemples montrent bien les nombreuses discussions sur la condition moderne et la société actuelle en Chine qui animent les internautes chinois. La noirceur de l'humour et le réalisme des débats donnent une image plutôt sérieuse voir obséquieuse des réseaux sociaux chinois. Nous devons nous empresser de nuancer ce propos : les photos d'animaux domestiques ou de nourritures y restent avant tout les contenus les plus courants. L'humour peut également être beaucoup plus léger avec des mèmes aux contenus aussi absurdes que comiques. Une série policière très prisée présentant un couple de détectives enquêtant dans la Chine médiévale procura notamment de bonnes doses de rire aux spectateurs. Dialogue récurrent de la série, la phrase “*Yuanfang, qu'en penses-tu ?*” (元芳, 你怎么看?) s'est hissée en peu de temps à une célébrité semblable au “*élémentaire*” de Sherlock Holmes au Dr. Watson. Le poète chinois Du Fu de la dynastie Tang a également connu un retour fulgurant sous la forme de peintures détournées le mettant en scène dans des situations improbables : “Du Fu est vraiment très occupé” (杜甫很忙).



(a) Dufu is very busy



(b) Yuanfang, qu'en penses-tu ?

### Graphique 4.2.3

L'échantillon que nous avons choisi d'étudier reflète la diversité de ces contenus (voir

---

working in a dead-end job” sur <https://en.wikipedia.org/wiki/Diaosi>, consulté le 7 Juillet 2014 à 11 :12

la liste 4.2.1). Ainsi, l'échantillon contient autant d'actualités nationales importantes (comme le 18ème Congrès du Parti Communiste Chinois ou l'attribution du prix Nobel de littérature à l'écrivain chinois Mo Yan) que des mèmes comiques et absurdes ("Yuanfang, qu'en penses-tu?" et "Du Fu est vraiment très occupé"). Également, nous avons retenu plusieurs faits divers (la sextape de Lei Zhengfu, les montres de *biaoge* ou le gâteau *qie-gao* du Xinjiang) et une émission de télévision *The Voice*. Les différents contenus choisis vont nous permettre de mieux comprendre les modèles de diffusion en établissant des comparaisons.

Une présentation plus détaillée de chacun des mèmes choisis est disponible en annexe de cette thèse. Pour une meilleure commodité de lecture, un nom a été attribué à chacun d'eux afin de pouvoir les évoquer facilement dans la suite de l'étude. A l'aide de l'outil que nous avons développé, nous allons maintenant observer dans le détail la diffusion de chacun de ces mèmes.

### 4.2.1 Structures temporelles

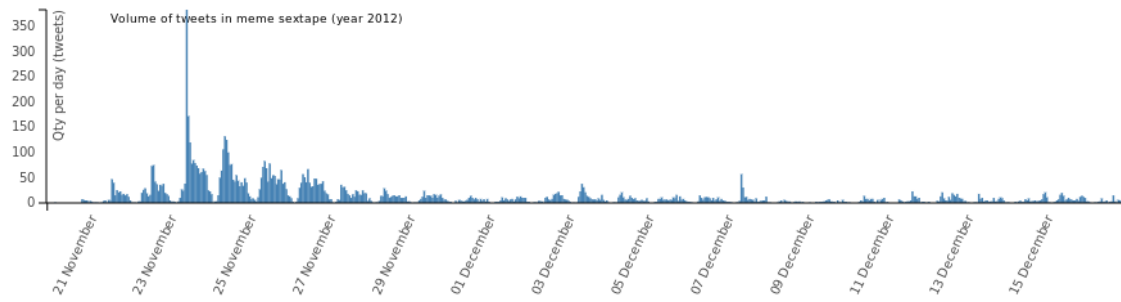
Dans un premier temps, nous avons choisi de considérer les différentes structures temporelles qui émergent de la diffusion de ces mèmes. La première série de figures ci-dessous (fig. 4.2.4) représente chronologiquement le volume de messages échangés par heure dans chacun des jeux de données. La période totale couverte par les échantillons est de 3 semaines. La première observation concerne la diminution quotidienne du volume de messages dessinant des pics réguliers. Cette forme spécifique correspond tout simplement à la baisse de l'activité des utilisateurs de *Sina Weibo* pendant la nuit.

Nous voyons également émerger différentes formes représentatives pour chacune des catégories choisies : le fait divers (“breaking news”) représenté par le mème *qiegao* débute très nettement avec une augmentation abrupte de la quantité de message postés en très peu de temps (Fig. 4.2.4c). Atteignant un niveau très élevé quasiment instantanément, ce pic est suivi d'une retombée rapide de l'attention quelques jours plus tard.

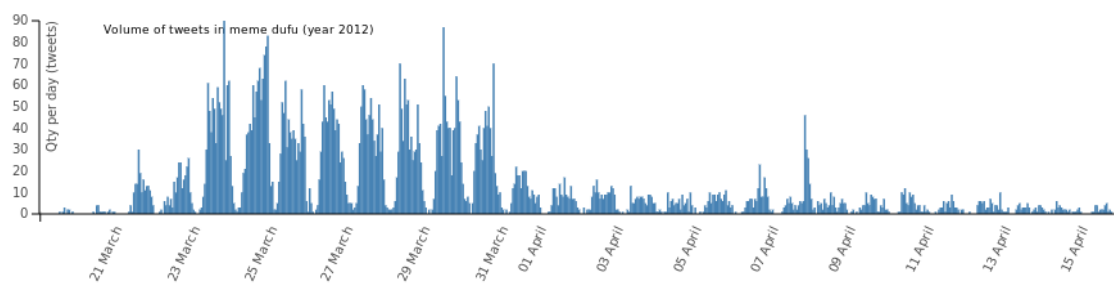
A l'inverse, la discussion plus informative autour du procès de l'homme politique de Chongqing et sa “sextape” (Fig. 4.2.4a) croît doucement pour connaître un pic très fort de discussion. L'attention retombe rapidement, consécutivement à la prise en main rapide de cette affaire par le gouvernement et les médias locaux.

L'émission de télévision *The Voice* est, elle, absolument événementielle. Son graphe montre nettement le rythme clairsemé des émissions, accompagné de très peu de discussions seulement entre chacune d'elles (Fig. 4.2.4d). Néanmoins, le volume de messages échangés est nettement plus important que dans les autres cas, avec un pic de plus de 2000 tweets par heure (dans notre corpus) le soir de la première émission, dépassant très largement le volume des trois autres mèmes présentés ici.

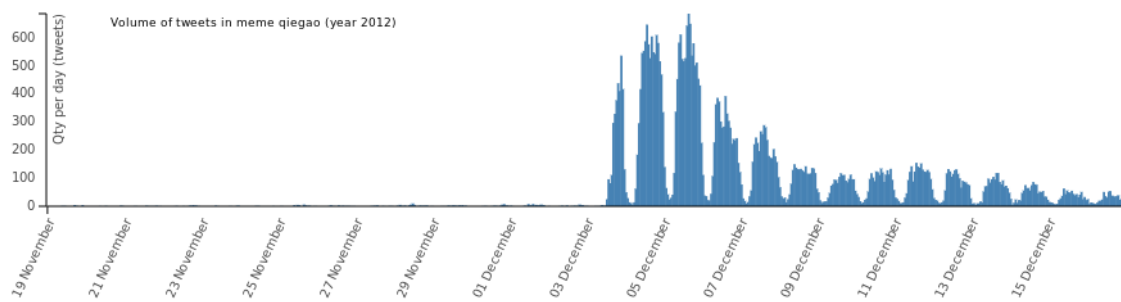




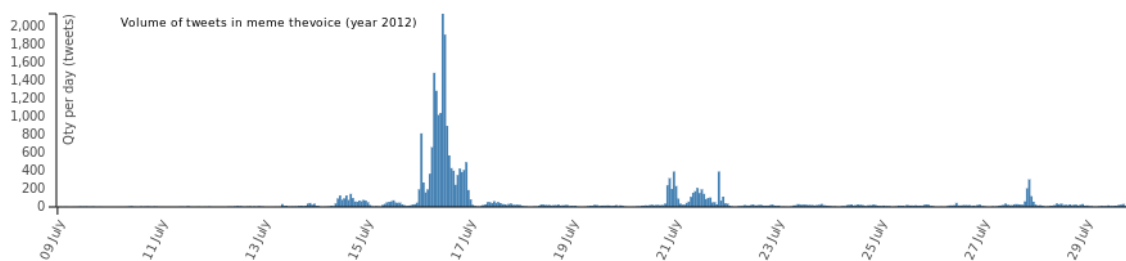
(a) sextape



(b) dufu



(c) qiegao

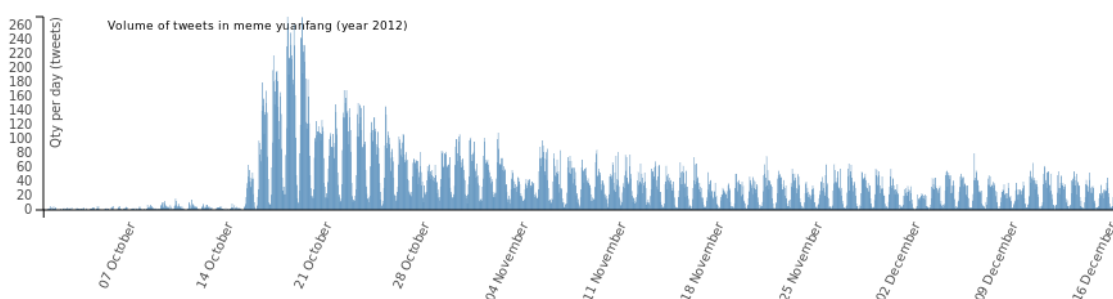


(d) The Voice

**Graphique 4.2.4** – Graphe temporel représentant le volume de messages échangés autour de 4 mèmes sur une durée de 3 semaines

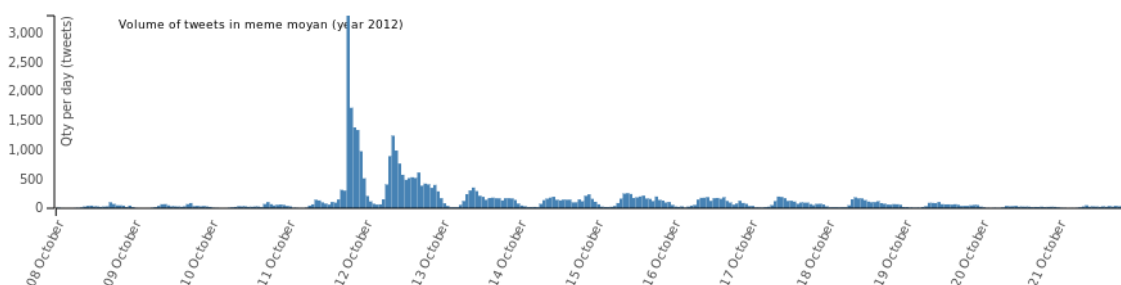
Le graphe du même absurdiste *dufu* connaît quant à lui une lente montée d'attention puis un plateau de plusieurs jours (fig. 4.2.4b). La blague semble durer et même connaître un regain d'intérêt une semaine plus tard. Elle est ensuite régulièrement mentionnée durant les jours suivants, montrant davantage de persistance que les autres types de contenus.

Proportionnellement, nous constatons que le même absurdiste est donc celui qui retient le plus l'attention, le plus longtemps (Fig. 4.2.4b). A l'inverse la campagne en ligne de *The Voice* n'est qu'un artefact de l'émission (Fig. 4.2.4d) mais draine une audience très importante pendant un temps restreint. Le cycle des actualités semble être plus court, arrivant rapidement pour disparaître tout aussi vite.



**Graphique 4.2.5** – Graphe temporel représentant le volume de messages échangés autour du même *yuanfang* entre le 7 Octobre et le 16 Décembre 2012.

En observant un second contenu comique (*yuanfang*) sur une période beaucoup plus longue de plusieurs mois (ici du 7 Octobre au 16 Décembre 2012) (fig. 4.2.5), nous pouvons vérifier la tendance des même humoristiques à durer dans le temps. *Yuanfang* continue d'être mentionné très régulièrement même des mois après sa première apparition.



**Graphique 4.2.6** – Graphe temporel représentant le volume de messages échangés autour du même *moyan* entre le 8 et le 21 Octobre 2012.

A l'inverse, les actualités du type “news” ont une durée de vie très courte, même lorsqu'elles sont très débattues. Dans le cas de *moyan*, on voit un très fort volume d'activité

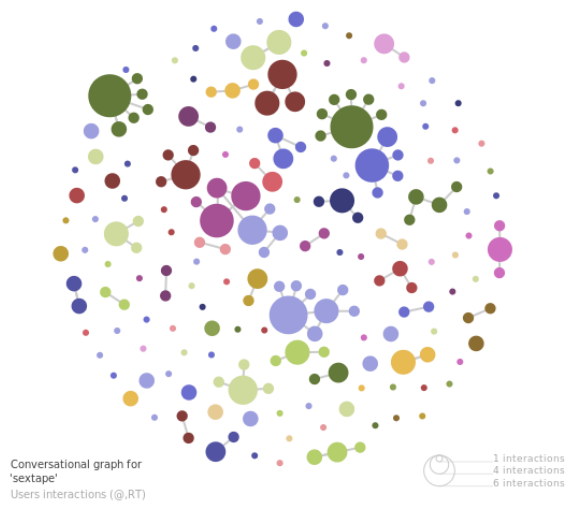
(3000 messages par heure au plus haut) mais une disparition quasi-totale des débats au bout en moins d’une dizaine de jours (Fig. 4.2.6).

Les cycles de vie des actualités ou des scandales en ligne sont déjà largement connus et documentés. Cette approche montre néanmoins une nouveauté : la dimension humoristique et absurdiste de certains mêmes leur permet une diffusion à moyen terme.

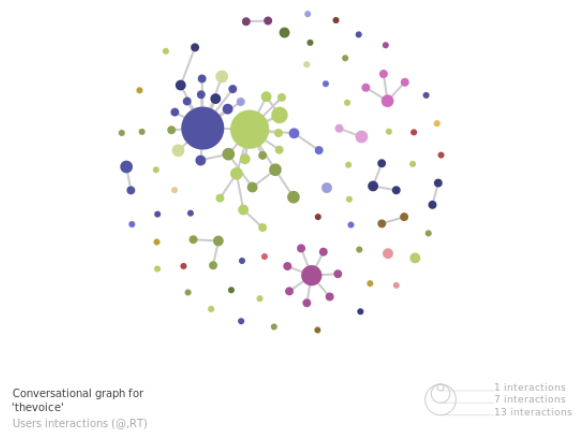
### 4.2.2 Structures conversationnelles

Les conversations sur Sina Weibo sont faites d’échanges entre les utilisateurs sous la forme de commentaires, de promotion d’un message (*zhuanfa* ou 转发 équivalent du “retweet” de Twitter) ou de réponse au message. Mises bout à bout, ces différentes interactions forment une large conversation où les utilisateurs interagissent, se parlent, se répondent, discutent en utilisant les différentes fonctionnalités de la plate-forme. En simplifiant ces échanges sous la forme de relations dirigées entre deux utilisateurs (*“A interagit avec B”*), nous pouvons reconstituer le réseau dirigé des interactions que nous visualisons ensuite sous la forme d’un graphe. Dans ce réseau conversationnel, les individus sont symbolisés par des points au sein des communautés notées par différentes couleurs sélectionnées pour leur cohérence et leur visibilité (Lin *et al.*, 2013). La taille des cercles indique leur degré dans le réseau global (total des connections entrantes et sortantes).

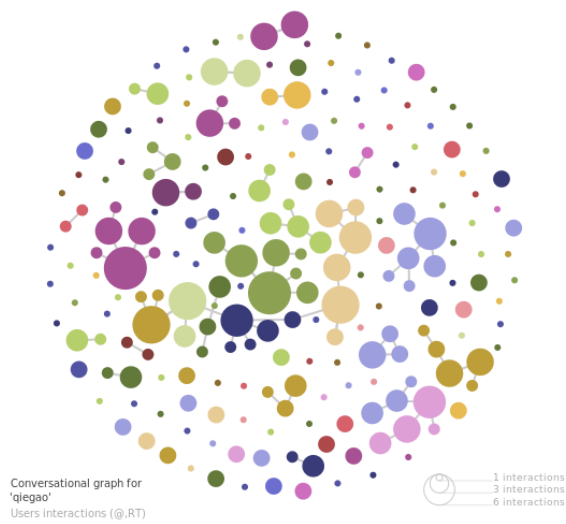
Les graphes de conversation permettent d’identifier des spécificités et de lire plusieurs indicateurs. Ils rendent notamment bien compte de la taille des conversations, très variables, ou plutôt de la dimension des interactions. On voit que les discussions autour d’aspects politiques et sociétaux suscitent beaucoup plus d’échanges, alors que les deux autres semblent moins importantes. Dans le cas de *The Voice* (Fig. 4.2.7b), la conversation est très structurée autour de peu d’acteurs qui centralisent les débats. Le même absurdiste *dufu* (fig. 4.2.7d) est lui composé d’utilisateurs très distants qui ne communiquent entre eux que faiblement ou par petits groupes. La conversation *sextape* autour du fait divers politique se constitue en groupes structurés mais plus distants. Repoussés vers les bords du graphe, l’existence de nombreux foyers de discussions montre bien que les débats ont été canalisée autour de groupes spécifiques. A l’inverse, la conversation société *qiegao* autour des peuples du Xinjiang (fig. 4.2.7c) semble se constituer en un vaste foyer de discussions où viennent s’agréger de nombreux utilisateurs très actifs.



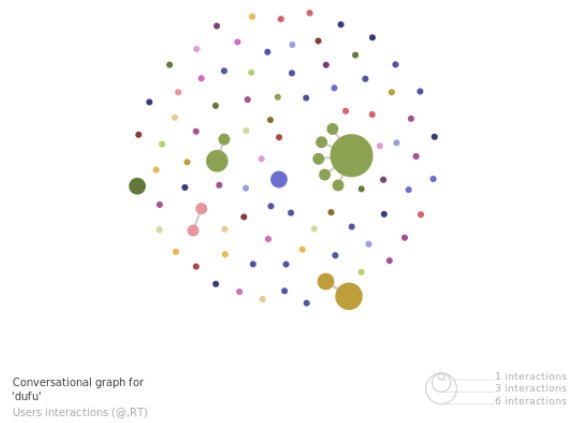
(a) Sextape



(b) The Voice



(c) Qiegao



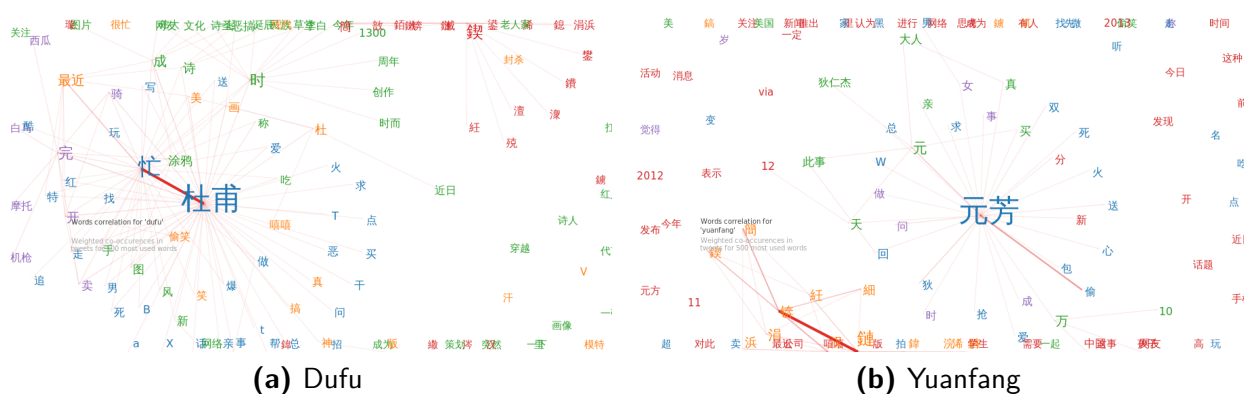
(d) Dufu

**Graphique 4.2.7 – Graphe conversationnel pour quatre mèmes (sur une durée de 3 semaines)**

### 4.2.3 Structures sémantiques

La dimension lexicale des contenus web présente une autre approche intéressante à explorer pour mieux comprendre comment se structure l'activité lors de la diffusion des mèmes. La représentation classique du nuage de mots s'intéresse principalement à la quantité de mots en laissant toutefois de côté la mise en relations des mots, pourtant fondamentale dans l'activité symbolique et expressive. Ainsi, nous avons plutôt choisi de donner à voir les réseaux de mots pour comprendre la construction de sens autour de signes particuliers.

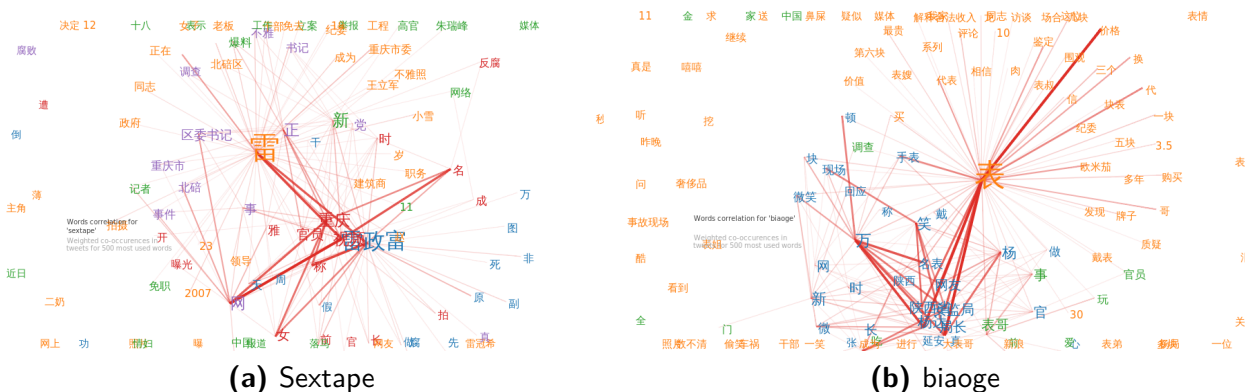
Les relations entre les mots représentent leurs co-occurrences dans un même message. La proximité des mots lors des discussions est matérialisée par un graphe disposant les mots par groupe grâce à un algorithme de force. La taille des mots indique le nombre de fois qu'ils ont été cités dans les conversations. Les couleurs représentent les communautés détectées par l'algorithme de Louvain (Blondel *et al.*, 2008) (voir section 4.1.3 pour plus de détails). Les corpus analysés ici ont été constitués grâce à une recherche par mots-clés dans un moteur d'indexation. Ainsi, les mots les plus proéminents sont bien souvent à l'origine de la constitution du corpus (les mots-clés de la requête utilisée). Ce biais important est à prendre en compte dans la lecture de ces graphes.



**Graphique 4.2.8** – Graphe sémantique de co-occurrences des mots pour les mèmes dufu et yuanfang

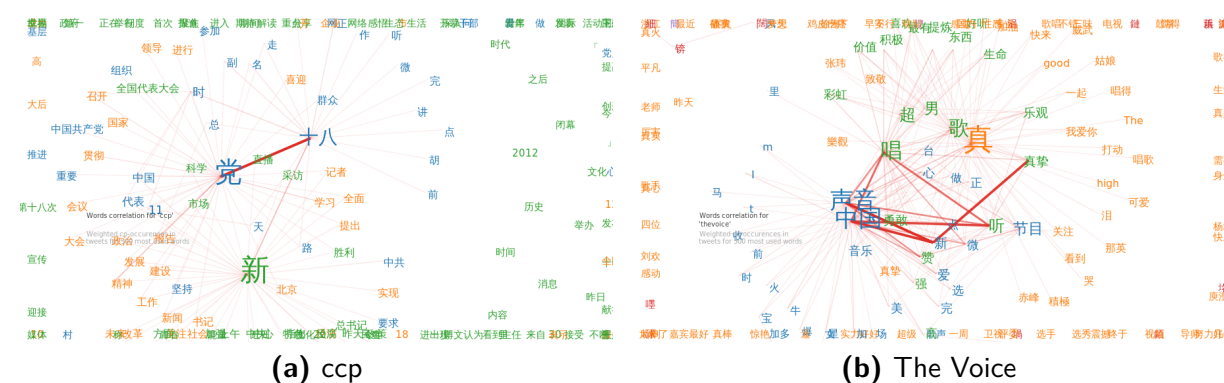
En regardant ces figures, nous constatons tout d'abord que les graphes des mèmes absurdistes (fig. 4.2.8a & 4.2.8b) se composent de peu de mots centraux très fortement connectés et entourés d'une myriade de mots de faible densité. Ces mots de faible importance correspondent aux multiples déclinaisons qui entrent en jeu dans l'appropriation du

même par jeux de mots.



**Graphique 4.2.9** – Graphe sémantique de co-occurrences des mots pour les mêmes Sextape et Biaoge

Les réseaux sémantiques de la discussion concernant les faits divers politiques (fig. 4.2.9a & 4.2.9b) sont plus structurés. Ils sont organisés en communautés de mots et de sens plutôt bien définies : une communauté décrit les détails de l'affaire (noms de lieux et de personnes, verbes d'actions), l'autre se compose d'adjectifs discutant le caractère houleux de ces histoires, la dernière est faite de mots plus anecdotiques.

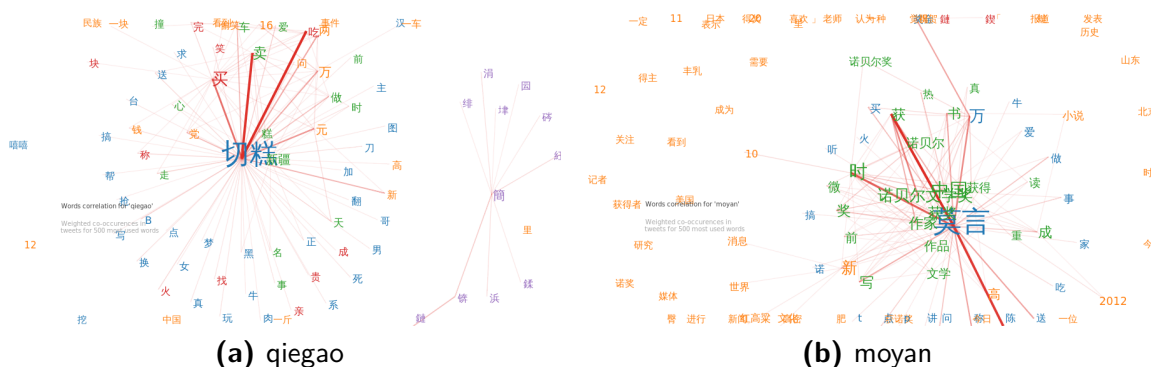


**Graphique 4.2.10** – Graphe sémantique de co-occurrences des mots pour les mêmes CCP et The Voice

La structure lexicale qui entoure *The Voice* et *CCP* semble très architecturée, avec des associations plus convenues. Pour *The Voice* (fig. 4.2.10b), nous sommes sans surprise en présence du champ lexical du chant (voix-chant, chant-chanson...) entouré par de nombreuses émotions et sentiments (arc-en-ciel-amour, trac-courage, etc.). Nous voyons également que de très nombreux mots ont été propulsés sur les bords du graphe car peu reliés entre eux ou avec le reste du réseau. Reconnus algorithmiquement comme une seule

communauté, on y trouve notamment le vocabulaire propre de l'émission qui ne s'intègre pas au reste des discussions.

Dans le cas du congrès du parti communiste chinois (fig. 4.2.10a), on retrouve également cette organisation sémantique très hiérarchique autour du mot central “parti”, avec un ensemble de signifiants issus du programme politique (construction, développement, loi, science, etc.). Un autre élément prédominant s'articule autour du caractère “nouveau” (révolution, changement, futur...). Comme dans le cas de *The Voice*, nous sommes en présence d'un langage stéréotypé qui donne lieu à un graphe très structuré selon des axes de communication définis. Proportionnellement, une grande partie des discussions est située aux frontières du graphe, très peu reliée avec le reste de l'activité.



**Graphique 4.2.11** – Graphe sémantique de co-occurrences des mots pour les mêmes Qiegao et Moyan

*Qiegao* (fig. 4.2.11a), quant à lui, possède une structure très centralisée articulée autour d'un mot-clé précis qui relie plusieurs groupes sémantiques (communautés de mots par couleurs) autour de lui. On retrouve cette caractéristique d'ensemble dans le graphe de *moyan* (fig. 4.2.11b) même si la structuration des communautés sémantiques est plus accentuée. La hiérarchie émergente du graphe montre des mots-clés qui orientent la discussion. Ici, les associations sont plus librement définies mais suivent des champs sémantiques distincts.

#### 4.2.4 Dynamiques géographiques et discussions entre provinces

Nous avons proposé dans cette recherche le terme de milieu numérique afin notamment de problématiser les rapports existants entre les pratiques de la discussion en ligne sur *Sina Weibo* et l'espace de la Chine urbaine moderne. Afin de poursuivre notre interrogation,

nous allons désormais observer comment les topologies de diffusion en ligne peuvent se corréler avec la réalité géographique. La diffusion d'un même suit-elle la hiérarchie urbaine classique ou bien la géographie urbaine de l'Internet ? Existe-t-il des modèles de diffusion selon les types de contenus ? Quels sont les *patterns* géographiques décrits par la circulation des contenus ?

Pour tenter de répondre à ces différentes questions, nous disposons de données concernant les provinces d'origine des utilisateurs grâce à leurs informations de profil de Sina Weibo. Une première analyse de la diffusion géographique des mèmes se fera par la projection sur des cartes des échanges entre les utilisateurs. La province d'origine est assignée à chacun des membres des réseaux conversationnels précédemment identifié afin de représenter la dynamique des échanges. Les dynamiques mineures de contenus apparaissent car le volume d'information pour chaque province est pondéré par sa présence dans l'échantillon de la population totale. L'épaisseur des traits exprime le volume des échanges en pourcentage du total sur la période observée.



**Graphique 4.2.12 – The Voice : Interaction des utilisateurs par province entre le 9 et le 29 Juillet 2012**



**Graphique 4.2.13 – The Voice : Interaction des utilisateurs par province entre le 16 et le 17 Juillet 2012**

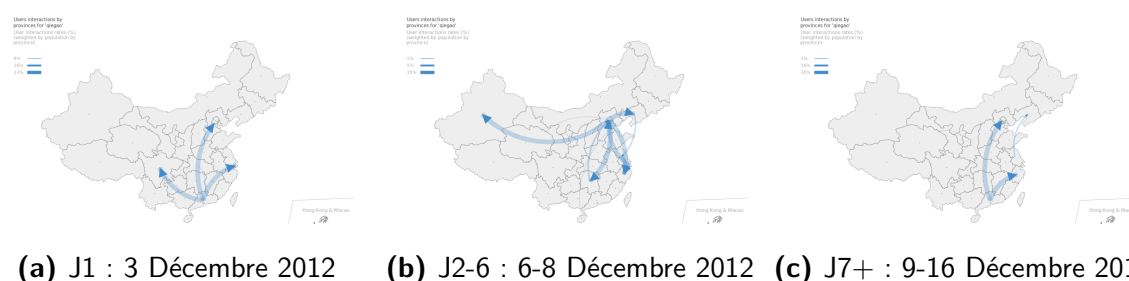
Dans le cas de *The Voice* (fig. 4.2.12), nous pouvons voir que l'interaction est structurée autour d'un axe fort entre la province de Zhejiang et Pékin. Cela s'explique assez simplement par le fait que l'émission est diffusée par *Zhejiang TV* dont le compte officiel joue un rôle important dans la diffusion et la promotion de la discussion. En isolant uniquement le soir de la première diffusion (fig. 4.2.13), on voit que cette structure s'accroît encore davantage. Les interactions se structurent alors entre le Zhejiang, Pékin et Shanghai ainsi que quelques axes plus minoritaires.





**Graphique 4.2.14** – Qiegao : Interaction des utilisateurs par province entre le 19 Novembre et le 15 Décembre 2012

A l'inverse, nous voyons que pour *Qiegao* (fig. 4.2.14) les patterns semblent beaucoup plus diversifiés. Guangzhou joue un rôle de diffuseur important et beaucoup d'informations convergent vers Shanghai et Pékin. De nombreuses conversations se déroulent également dans l'ouest de la Chine. En effet, la discussion concerne ici les peuples Uyghur et on constate que les habitants du Xinjiang participent davantage à la conversation que dans d'autres mèmes.



**Graphique 4.2.15** – Qiegao : Interaction des utilisateurs par province

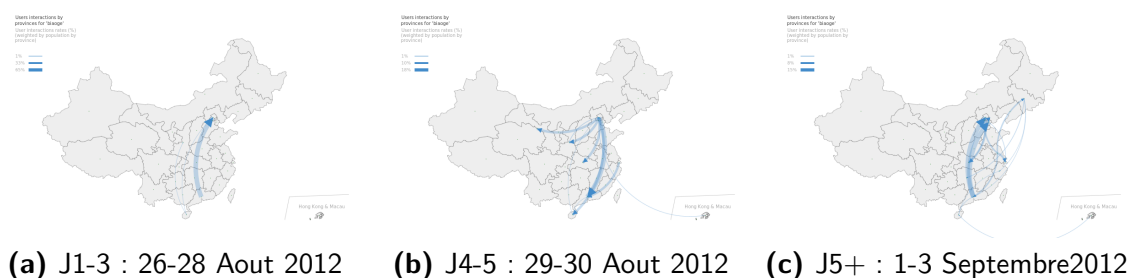
En considérant l'évolution de *qiegao* dans le temps, nous voyons qu'il émane d'abord de Canton (fig. 4.2.15a), puis se diffuse à Pékin et Shanghai qui semble jouer un rôle d'amplificateur en le diffusant plus largement (fig. 4.2.15c), notamment vers l'Ouest (fig. 4.2.15b).

On retrouve également ce pattern pour d'autres faits de société comme *sextape* ou *biaoge*. Cela illustre bien le rôle particulier des médias du sud de la Chine et plus spécialement celui de Guangzhou (fig. 4.2.16a & 4.2.17a).

Avec une plus grande liberté de ton et une plus grande latitude dans leur propos, les

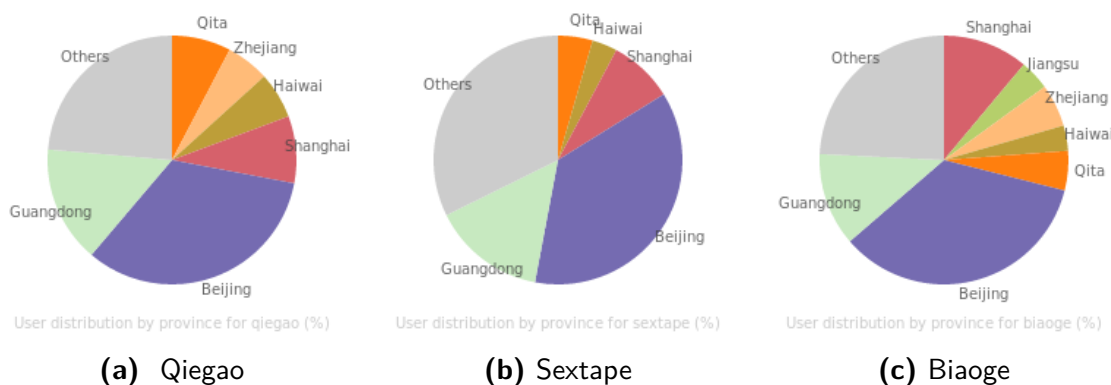


**Graphique 4.2.16 – Sextape : Interaction des utilisateurs par province**



**Graphique 4.2.17 – Biaoge : Interaction des utilisateurs par province**

médias cantonnais sont bien souvent instigateurs d'affaires importantes alors que les médias pékinois, plus conservateurs agissent plutôt comme des diffuseurs.



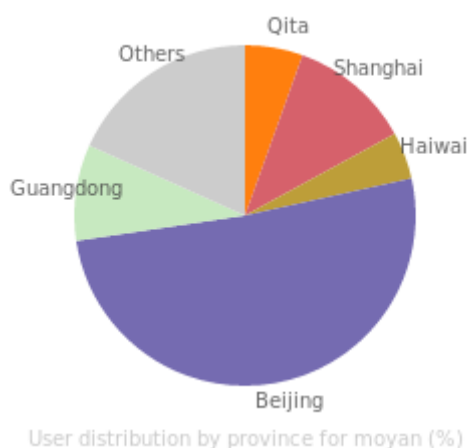
**Graphique 4.2.18 – Répartition des utilisateurs par province (en % du total)**

En observant la répartition des utilisateurs par province pour chacun de ces trois mêmes (fig. 4.2.18), on remarque que Beijing, Canton, Shanghai ont une place importante, ainsi que les personnes situées à l'étranger (Hawaii) et le Zhejiang. Néanmoins, le reste des provinces (représentés ici par "Others" en gris) constitue une part importante des échanges.



**Graphique 4.2.19** – Moyan : Interaction des utilisateurs par province

Dans le cas de nouvelles plus officielles comme pour *Moyan* notamment, nous pouvons voir que l'information s'origine généralement de Pékin (fig. 4.2.19a) et se diffuse vers le sud et le centre (fig. 4.2.19b).



**Graphique 4.2.20** – Répartition des utilisateurs par province pour Moyan (en % du total)

La composition des utilisateurs (fig. 4.2.20) montre bien que les débats autour de Moyan sont très largement dominés par Pékin avec plus de 50% de l'activité impliquant au moins un utilisateur identifié à Pékin.

Les mêmes absurdistes et comiques ne semblent pas se construire selon les patterns classiques que nous avons pu observer auparavant. On voit notamment que le même *dufu* (fig. 4.2.21a) débute dans la région du Hubei alors que le Liaoning joue un rôle clé dans la diffusion de *yuanfang* (fig. 4.2.21b).

En considérant les graphes de temps, on remarque également que si Pékin est bien présent, les échanges des premiers jours se font principalement avec le Hubei (fig. 4.2.22a) et le Yunnan (fig. 4.2.22c), deux provinces typiquement peu actives. Dans un second temps, l'activité se concentre sur les régions cotières (fig. 4.2.22d).



(a) dufu



(b) yuanfang

**Graphique 4.2.21 – Interaction des utilisateurs par province**



(a) (Dufu) J1-2 : 23 au 25 Mars 2012



(b) (Dufu) J3+ : 25 Mars au 15 Avril



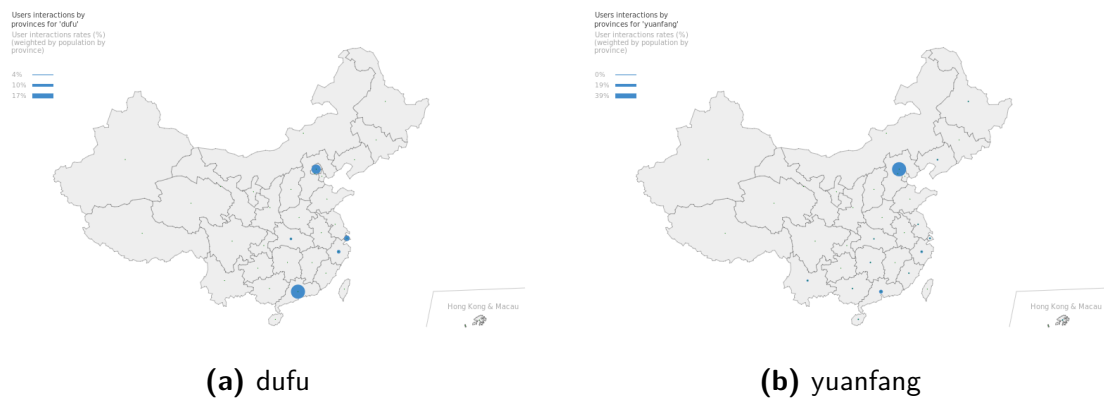
(c) (Yuanfang) 1-2 : 15 au 17 Octobre 2012



(d) (Yuanfang) J3+ : 18 Oct au 16 Décembre 2012

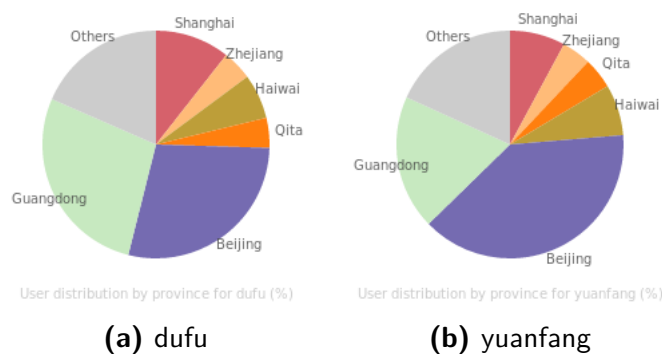
**Graphique 4.2.22 – Evolution des interactions des utilisateurs par province pour Dufu et Yuanfang**

En nous intéressant aux échanges qui se déroulent au sein de chaque province (d'un utilisateur situé dans une province vers un autre utilisateur situé dans la même province) (fig. 4.2.23b & 4.2.23a), nous constatons que les échanges internes sont également importants au sein des villes principales (Canton, Pékin et Shanghai).



**Graphique 4.2.23** – Interaction entre utilisateurs de la même province

Néanmoins, de nombreuses villes prennent aussi part à la discussion (fig. 4.2.24b & 4.2.24a).



**Graphique 4.2.24** – Répartition des utilisateurs par province (en % du total)

La plupart des contenus paraissent suivre des modèles connus, à l'exception des mêmes absurdistes qui présentent des caractéristiques différentes.

## 4.2.5 Communautés, clusters et groupes de provinces

Un récent article du département d'urbanisme de l'Université de Nanjing (*Zhen et al., 2013*) analyse les similarités et différences entre le réseau urbain et le réseau des relations friends/followers entre les utilisateurs sur *Sina Weibo*. Il y est notamment montré que le

réseau des relations entre utilisateurs vient renforcer la hiérarchie classique du système urbain. Zhen explique comment Internet vient accélérer l’agglomération selon des modèles spatiaux pré-existants comme celui des “*Three majors and four smalls*”, accentuant ainsi la différence très marquée entre l’Ouest et l’Est de la Chine.

### Three Majors

1. Pearl River Delta (Guangzhou, Shenzhen)
2. Beijing-Tianjin-Hebei region (Beijing)
3. the Yangtze River Delta (Shanghai, Hangzhou, Nanjing)

### Four Smalls

1. Chengdu-Chongqing region (Chengdu, Chongqing)
2. Hercynian region (Fuzhou, Xiamen)
3. Wuhan (central) region (Wuhan, Changsha)
4. Northeast China (Shenyang, Harbin, Changchun)

**Graphique 4.2.25** – Le modèle urbain chinois du Three majors and four smalls d’après *Zhen et al. (2013)*

Les résultats obtenus par l’analyse du réseau des followers sur *Sina Weibo* peuvent être comparés avec les phénomènes observés lors de la diffusion de contenus. En effet, si on peut affirmer que le réseau “d’amis” structure les canaux de diffusion de *Sina Weibo*, ces canaux ne sont pas nécessairement sujets à une utilisation fréquente et donc à l’actualisation de ces structures. Il peut donc être intéressant d’observer les façons dont ces canaux sont activés lors des conversations en ligne.

Reprenant le réseau des interactions entre provinces constitué par les échanges des utilisateurs, nous souhaitons désormais observer les communautés de discussion se formant entre des provinces spécifiques lors de la diffusion des contenus sur *Sina Weibo*. Nous soumettons donc le réseau d’interactions entre provinces à un algorithme de Louvain (*Blondel et al., 2008*) qui nous permet d’identifier les provinces ayant le plus interagi ensemble. Sur les cartes, les couleurs représentent les différentes communautés calculées depuis le réseau d’interactions.

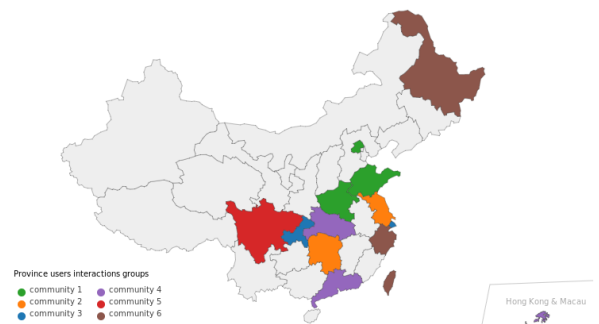
À la première lecture de ces cartes, nous constatons en effet que les provinces de l’Ouest sont nettement moins engagées que celles de la moitié Est de la Chine. Ce fait reflète la concentration de la population des utilisateurs de Weibo dans les grandes villes en développement de la côte et du centre (*Fu et al., 2013*).

Users interactions by  
provinces for 'sextape'



(a) sextape

Users interactions by  
provinces for 'thevoice'



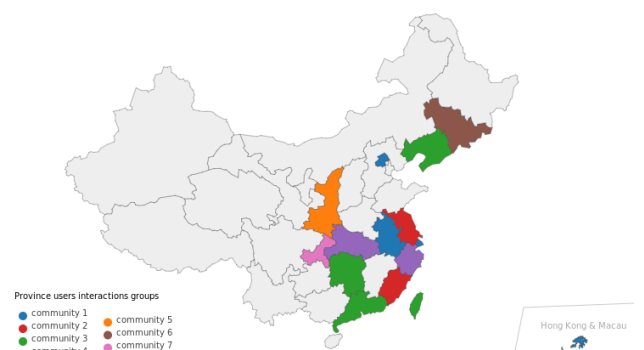
(b) TheVoice

Users interactions by  
provinces for 'qiegao'



(c) Qiegao

Users interactions by  
provinces for 'dufu'



(d) Dufu

**Graphique 4.2.26** – Communauté de provinces dessinées par les échanges entre utilisateurs au-tour de chaque mème (Algorithme de Louvain).

Le même absurdiste *dufu* possède une diffusion plus concentrée (fig. 4.2.26d) alors que les discussions politiques semblent regrouper des utilisateurs d'origines plus diverses (fig. 4.2.26a & 4.2.26c). Nous remarquons également que Taiwan est absente des discussions plus en lien avec la politique et la société de Chine continentale, alors qu'elle est bien présente dans le cas des mêmes absurdistes et de l'émission de loisir (fig. 4.2.26b). Parmi la diversité de formes visibles sur les cartes, peu de pattern sont néanmoins décelables a priori. La circulation des contenus ne semble pas suivre de patterns définis, mais plutôt posséder des modèles différents pour chaque contenu. Il nous est donc impossible de valider ou réfuter le modèle proposé par l'analyse des relations friends / followers par *Zhen et al.* (2013).

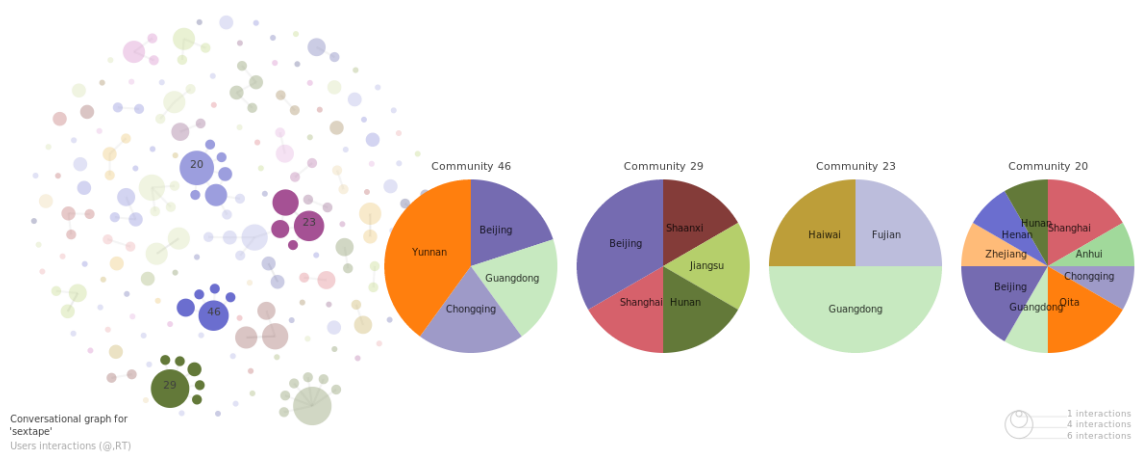
## 4.2.6 Dimensions géographiques des graphes socio-sémantiques

Ces premières observations nous ont montré que s'il est possible de mieux comprendre les logiques et dynamiques de diffusion des contenus, il est néanmoins plus périlleux de tirer des conclusions définitives sur les dynamiques géographiques de la diffusion sur l'ensemble d'un territoire. De plus, nous ne disposons que de très peu d'aspects "géo-localisés" dans ces données (pas de géotag sur chaque message à proprement parler). Nous savons également que la province d'origine du profil ne peut être considérée comme une source absolument fiable. Les utilisateurs sont en effet libres de la remplir selon leur bon vouloir. Face à ces diverses réserves, il est important de recentrer l'étude sur les dimensions plus facilement observables de la diffusion, notamment socio-sémantiques. Le croisement et les corrélations entre mots, utilisateurs et provinces peuvent nous permettre d'explorer des dynamiques entourant les échanges en ligne.

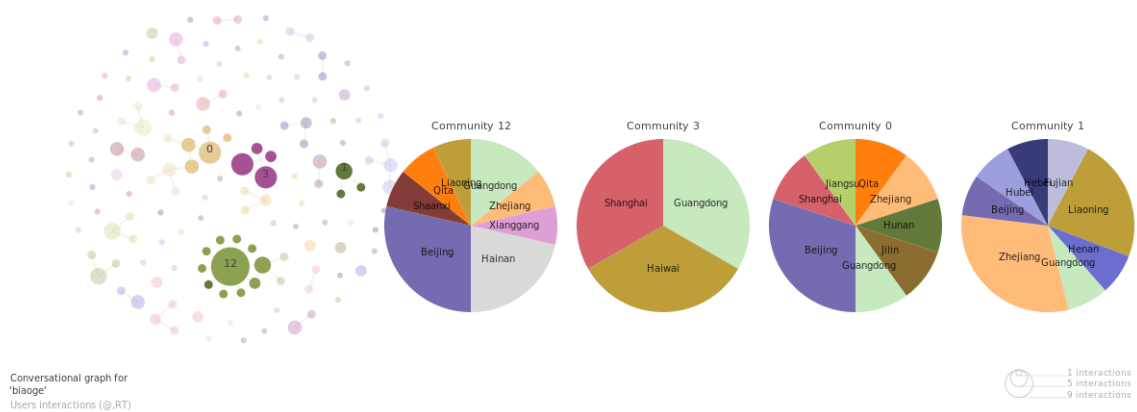
Tout d'abord, nous nous proposons d'observer la distribution des provinces d'origine des communautés d'utilisateurs les plus importantes. En sélectionnant les communautés les plus centrales pour chaque graphe, nous pouvons mieux comprendre comment se répartissent les "influenceurs" sur le territoire pour chacun des mèmes choisis ici.

Les communautés formées lors de la discussion autour des mèmes à teneur plus politique présentent une grande diversité de provenances. On voit dans le cas de *biaoge* (fig. 4.2.28) et *sextape* (fig. 4.2.27) que de nombreuses provinces sont représentées.





**Graphique 4.2.27** – Communauté de provinces dessinées par les échanges entre utilisateurs autour de Sextape

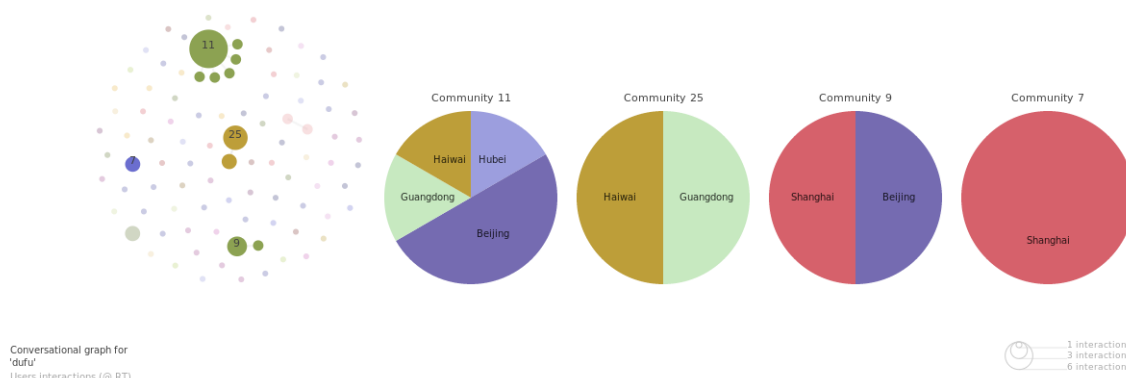


**Graphique 4.2.28** – Communauté de provinces dessinées par les échanges entre utilisateurs autour de Biaoge



**Graphique 4.2.29** – Communauté de provinces dessinées par les échanges entre utilisateurs autour de The Voice

À l'inverse, les communautés d'utilisateurs les plus impliquées dans *The Voice* (fig. 4.2.29) ou *Moyan* (fig. 4.2.30) sont pour la plupart localisées à Beijing. Par rapport à la cartographie précédente qui montrait les relations fortes avec le Zhejiang pour *The Voice* (fig. 4.2.29), nous voyons ici que la province est très faiblement représentée dans les communautés majoritaires.

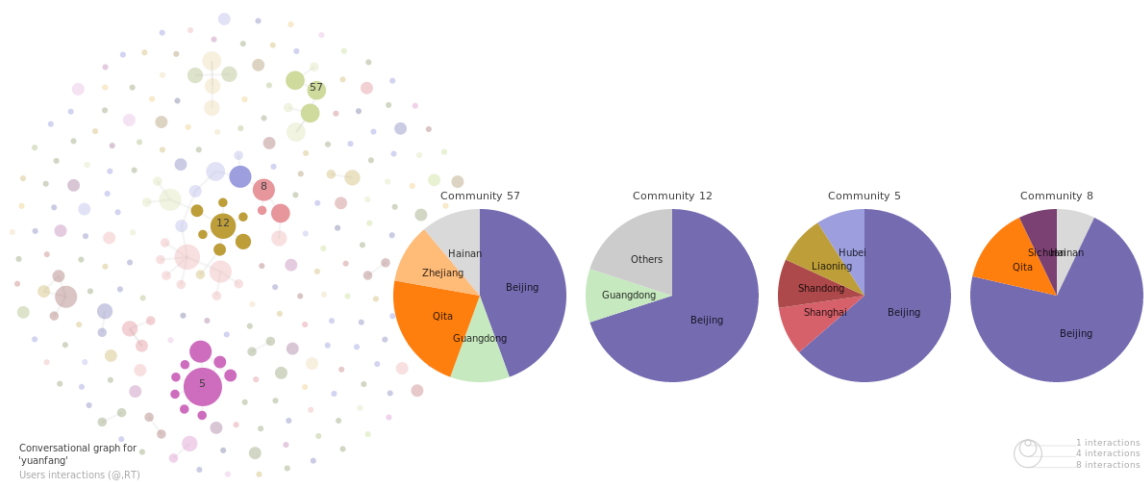


**Graphique 4.2.30** – Communauté de provinces dessinées par les échanges entre utilisateurs autour de Moyan

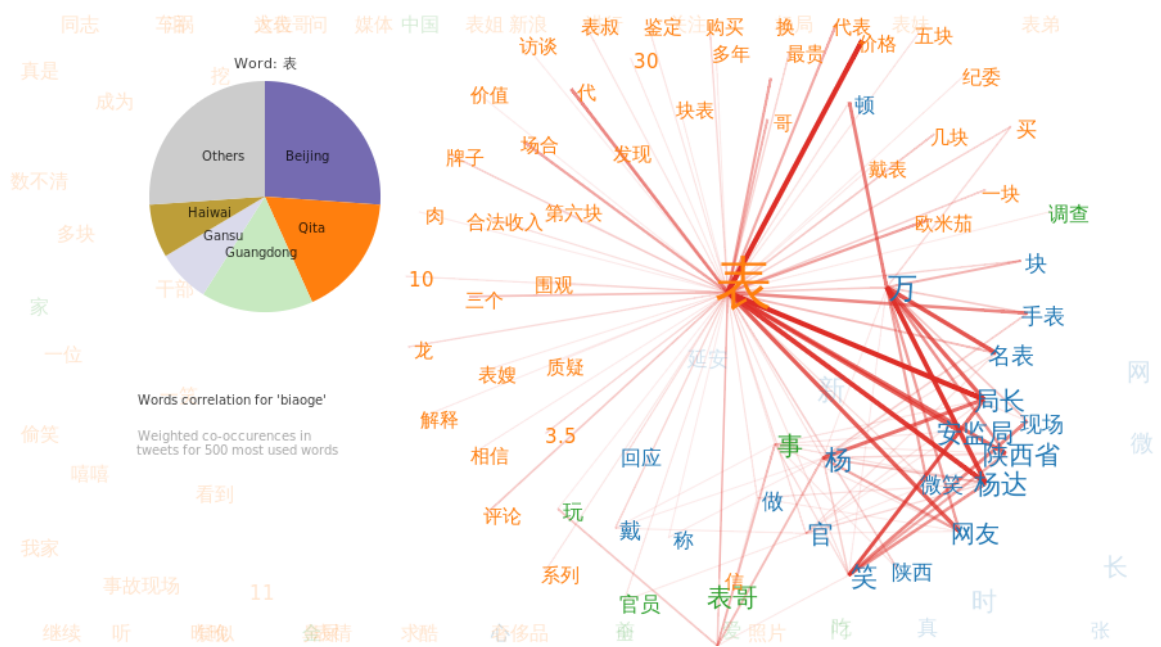
Le même *dufu* (fig. 4.2.31) montre quant à lui un pattern différent où les communautés sont organisées plus régionalement, avec moins de diversité d'origine des utilisateurs engagés dans les discussions. Des groupes plutôt perméables se forment autour de différentes régions.

Une dimension importante du même absurdiste semble être l'organisation des conversations en communautés plutôt locales (utilisateurs de même province). Une première analyse des communautés les plus centrales dans le graphe de *yuangfang* ne nous permet pas de retrouver ce pattern. Pourtant, en nous intéressant aux communautés de seconde et troisième importance, nous voyons que la plupart se constituent autour de 1 ou 2 provinces seulement.

Pour continuer notre exploration de l'existence géographique des relations socio-sémantiques dans nos mêmes, nous allons maintenant nous intéresser à une seconde dimension qui est celle de la distribution des mots par province. Nous sélectionnons pour chaque même les mots les plus centraux du graphe sémantique et décomposons leur usage afin de comprendre la diversité ou l'unicité de l'origine des utilisateurs.

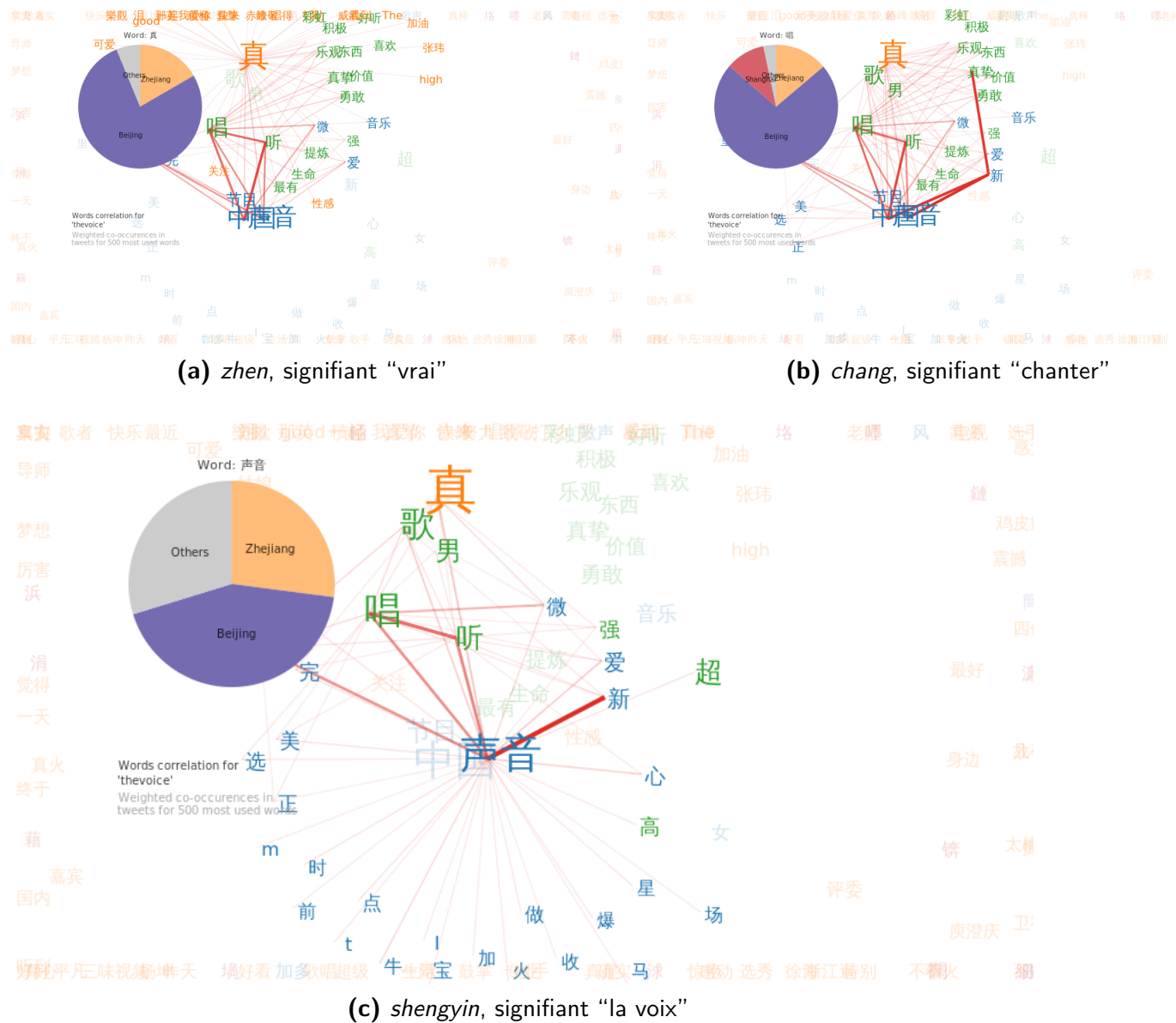


**Graphique 4.2.31** – Communauté de provinces dessinées par les échanges entre utilisateurs au tour de Dufu



**Graphique 4.2.32** – biaoqe : distribution des citations par provinces pour le mot biao (montre)

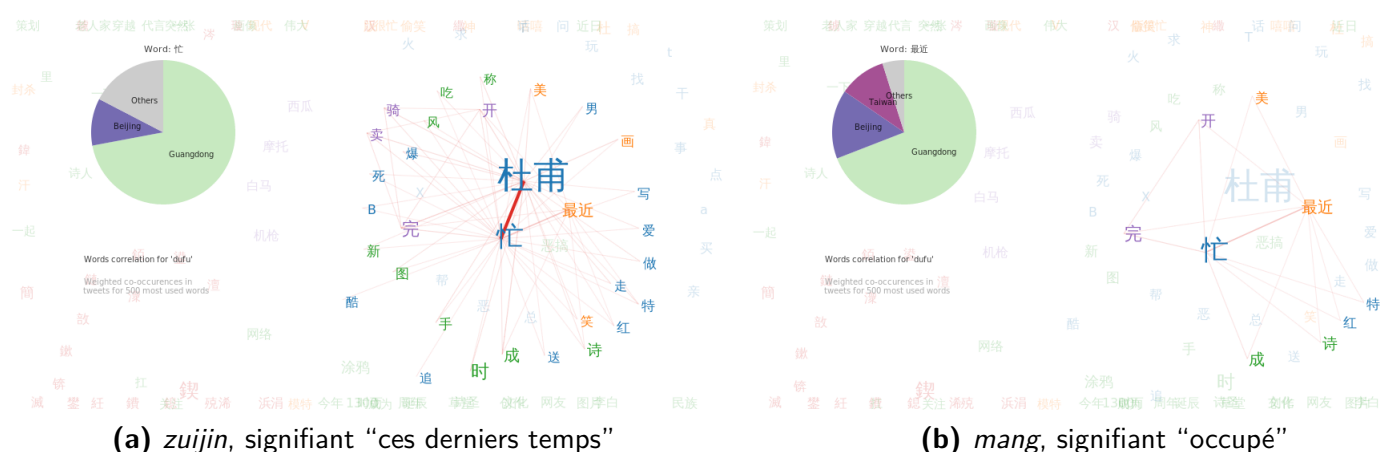
Dans le cas du même *biaoge* (fig. 4.2.32 & 4.2.32), nous voyons que les mots principaux portent une forte diversité et confirment l'hypothèse d'une diffusion géographique plus large de la discussion.



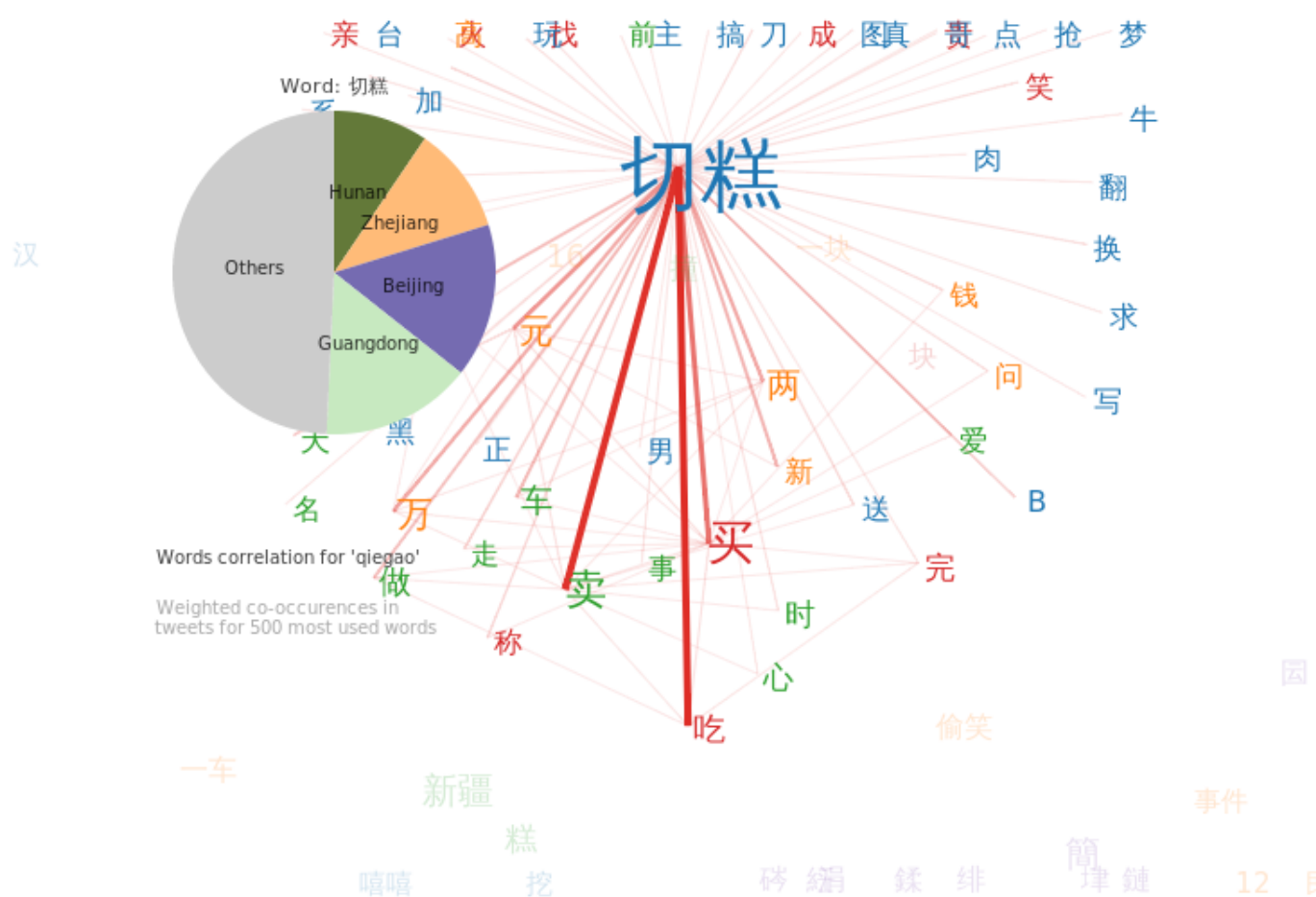
**Graphique 4.2.33 – The Voice : distribution des citations de mots par provinces**

À l'inverse, les mots-clés de *The Voice* (fig. 4.2.33) sont nettement dominés par des communautés d'utilisateurs de Pékin ou du Zhejiang qui semblent réellement fixer les termes de la discussion.

Dans le cas de dufu, les trois mots les plus importants *dufu*, *zuijin* et *mang* qui forment la phrase-mème “*dufu est très occupé récemment*” sont tous très fortement liés à la région de Canton (fig. 4.2.34). Cette information, pas forcément proéminente dans les analyses précédentes, montre que le développement de ce mème est largement le fait d'utilisateurs de la région de Canton.



**Graphique 4.2.34 – Dufu : distribution des citations de mots par provinces**



**Graphique 4.2.35 – Qiegao : origine géographique des utilisateurs**

## 4.2.7 Résultats

Les différentes observations faites grâce aux figures ci-dessus sont résumées dans le tableau synthétique ci-dessous (tableau 4.2.2).

Catégorie	Comique, Absurdiste	Actualité et société	Campagne médiatique	Faits Divers
Exemples	<i>“Du Fu is very busy”, “Yuanfang, qu’en penses-tu ?”</i>	<i>“Mo Yan prix Nobel”, “Qiegao”</i>	<i>“The Voice of China”, “Congrès du PCC”</i>	<i>“La sextape”, “Biaoge et le scandale des montres”</i>
Temps	Moins de volume, mais présent plus longtemps dans les conversations	Intense discussion de courte durée	Peu de discussions en-dehors des actions planifiées	Irruption abrupte d’un sujet qui disparaît rapidement
Conversation	Utilisateurs distants qui communiquent faiblement ou par petits groupes	Groupes très actifs mais distants entre eux	Peu d’acteurs centralisent les débats	Un foyer de discussion agrège les utilisateurs actifs
Lexique	Peu de mots centraux entourés de mots de faible importance	Champ lexical ouvert mais clairement organisé	Champ lexical de mots et d’associations convenus	Champ sémantique hiérarchisé et limité
Géographie	Importante diffusion interne aux villes, lieux atypiques	Large diffusion souvent centrée autour de Pékin	Articulée autour de la localisation des émetteurs principaux	Large portée des messages.
Influenceurs	Regroupés localement par régions	Pékin diffuse principalement, repris localement	Majoritaire-ment dans la ville d’origine du média	Grande diversité de provenance
Localisation	Mots-clés fixés par la région d’origine, puis multitude de régions	Pékin fixe les termes du débat	Faible engagement des autres provinces	Mots peu localisés et diffusés très largement

**Tableau 4.2.2 – Résumé des résultats**

## 4.3 Discussions

Un outil d'analyse et de visualisation nous a permis de constituer les topogrammes d'une dizaine de mèmes ayant circulé sur le service de réseau social chinois *Sina Weibo* : comique, commercial, publicitaire, actualité, faits divers et scandale politique. En comparant les dimensions sémantiques, conversationnelles, temporelles et géographiques de leurs diffusions, nous avons dressé une première liste de spécificités pour chacun de ces types de contenus.

En rappelant chacune des hypothèses formulées au début de ce travail (voir section 2.3.2), nous allons maintenant proposer une lecture critique des résultats obtenus et des démarches mises en œuvre lors de la réalisation de ce travail.

### Les industries culturelles et le modèle chinois

*Hypothèse : La majorité des contenus circulant sur les réseaux sociaux s'apparente largement à ceux des médias traditionnels*

L'analyse des hashtags les plus mentionnés montre que les contenus majoritairement discutés sur *Sina Weibo* sont largement similaires à ceux des médias traditionnels : divertissements, produits culturels, trafic routier, sports, etc. Également, les hashtags sont la plupart du temps un artefact de campagnes de marketing promotionnel politique, médiatique ou commercial. Les contenus qui leurs sont associés sont largement similaires à ceux des médias traditionnels.

En Chine, l'intégration très rapide des réseaux sociaux à la chaîne des médias traditionnels a été rendue possible par un ensemble de conditions. La première est le protectionnisme économique qui a été mis en place sur tout le secteur des médias, lié à la volonté de contrôle politique et le développement industriel. Comme nous l'avons montré (chap. 1.1), les TIC ont été dès le début conçues par Pékin comme un des piliers chargés de soutenir la croissance économique du pays. Ce protectionnisme sur tout le secteur des industries culturelles a permis l'apparition rapide de nombreuses firmes qui ont bénéficié d'un marché captif et non soumis à la concurrence internationale. Les opportunités au sein d'un marché contrôlé ont donc été largement favorables à la verticalisation des activités médiatiques.



La nécessité d'un contrôle de "qualité" des informations publiées par ces services a également accéléré cette intégration entre producteurs et distributeurs de contenus. Les fournisseurs et diffuseurs de contenus chinois sont soumis à des règles strictes de contrôle politique. Ils doivent en effet s'assurer que les contenus qu'ils publient sont "en harmonie" avec les règles et tendances définies par le gouvernement de Pékin. La gestion de ce risque a rendu indispensable la mise en place d'un contrôle continu des informations publiées. De ce fait, l'intégration verticale entre producteurs et distributeurs de contenu s'impose comme la solution la plus logique - la moins coûteuse. Les processus de gestion des "risques politiques" liés à la publication d'information ont donc favorisé l'intégration des réseaux sociaux au cycle de production des médias traditionnels.

La perspective de gérer la conformité des contenus aux législations et à l'agenda politique d'un gouvernement peut paraître au premier abord une spécificité de l'Internet chinois. Les termes du débat et le cadre légal qui régissent les relations avec les gouvernements sont bien différents dans le cas des firmes américaines comme *Facebook* ou *Twitter*. Néanmoins, la gestion des aspects légaux liés à la possession et l'usage des données utilisateurs deviennent une compétence de plus en plus centrale pour les services de réseaux sociaux. Culminant autour de la récente "affaire Snowden" (Greenwald *et al.*, 2013), les tensions entre les questions de "vie privée" (*privacy*) et de surveillance via Internet viennent notamment faire pression pour davantage de responsabilité légale de la part des services web. Également, l'inégalité de traitement des flux de données est au centre des débats législatifs et managériaux avec la question de la neutralité des réseaux (Schafer et Le Crosnier, 2011). La pression grandissante pour légiférer et faire appliquer les décisions légales auprès des firmes Internet va nécessiter un contrôle accru des différents maillons de la chaîne de production d'information. Ainsi, on peut gager que l'intégration verticale sera également une des premières solutions pour rentabiliser les coûteux dispositifs de contrôle à mettre en place. Comme dans le cas de *Sina Weibo* et des firmes de l'industrie médiatique en Chine, la pression légale et politique pourrait amener à une intégration des services de production et de diffusion de contenus.

Publicité, radio, télévision, le secteur des industries culturelles est historiquement sujet à une très forte concentration (Martel, 2010). Les grands groupes médiatiques fonctionnent selon ce principe d'intégration verticale, élargissant même largement leurs activités à d'autres secteurs. Comme nous l'avons vu avec *Sina Weibo*, les informations

publiées par médias *online* et *offline* sont de plus en plus homogènes et sont quotidiennement reprises. La dichotomie entre “web” et “traditionnels” s’estompe rapidement, avec le partage quotidien d’articles de journaux ou d’émissions de télévision sur la Toile. Les canaux de distribution eux-mêmes se recentrent autour du format mobile. Télévision web, web documentaire ou data journalisme convergent et les services de réseaux sociaux y jouent souvent le rôle de diffuseur principal. Dans cette dynamique, le rapprochement entre *pure players* du web comme *Twitter* ou *Facebook* et les firmes médiatiques traditionnelles ne paraît pas impensable. Cette tendance se voit notamment très clairement dans la nécessité de rentabiliser au maximum le placement d’information pour l’obtention de revenus publicitaires sur des écrans toujours plus petits. Les dispositifs de captation de l’attention que sont les services de réseaux sociaux deviennent la clé de voûte de l’industrie médiatique.

Ainsi, l’histoire récente des médias chinois et notamment celle de *Sina Weibo* ne doit pas être lue seulement dans les termes d’une lutte politique entre les utilisateurs et le gouvernement de Pékin. La dimension stratégique du développement industriel des médias en Chine apporte aujourd’hui des clés de lecture éclairantes non seulement pour le pays, mais pour le monde entier. L’équilibre entre contrôle politique, réussite commerciale et développement industriel de l’Internet chinois crée un précédent dont les conséquences pour le reste du réseau ne doivent pas être minimisées. L’intégration forte entre entités gouvernementales, fournisseurs de contenus et canaux de distribution de l’information propose également un modèle inédit. Les études sur l’Internet chinois doivent s’intéresser davantage aux détails des mécanismes et relations entre ces différents acteurs. Cette relation d’apparence fusionnelle des différents pouvoirs dans le champ médiatique pourrait même être considérée comme une forme de maturité de l’industrie, acquise lors d’un développement hyper-rapide permis par un environnement protectionniste et autoritaire. En effet, les discussions entourant le cadre législatif et les lois cherchant à réguler l’usage et la gouvernance d’Internet (*Hadopi* en France, *SOPA* aux États-Unis, *Data Protection Directive* de l’Union Européenne) tendent à une plus forte intégration des pouvoirs dans la sphère médiatique pour une gestion plus importante des risques politiques. L’expérience de l’Internet chinois établit dans le réseau mondial un modèle radicalement différent de celui imaginé initialement par les fondateurs du réseau ARPANET. S’il est possible de l’écarter comme un archaïsme face au discours *new-age* des géants californiens, son exis-

tence a néanmoins établi un mode de développement et de gouvernance qui s’est jusqu’ici montré d’une efficacité et d’une rentabilité sans faille et pourrait bien dans le futur séduire davantage d’états et de gouvernements.

## **Plate-formes web et lieux d’énonciation**

Cette dimension infra-structurelle de l’Internet n’est bien sûr qu’une seule des facettes qui entoure la production de l’espace web. L’immense diversité des usages et les nombreuses actions quotidiennes des utilisateurs sont l’autre pilier de ce large développement industriel. Les services de réseaux sociaux peuvent être analysés selon deux activités majeures qui s’entrelacent constamment : la lecture ou consultation de ressources médiatiques (textes, images, vidéos, photos, etc.) et l’écriture (liée à la discussion et la conversation). La rencontre de ces deux usages au quotidien donne lieu à de multiples phénomènes, portant notamment la pratique du commentaire vers des sphères nouvelles.

Les analyses machines préliminaires de notre corpus de données pour l’identification de mêmes ont montré la difficulté d’isoler chacune de ces pratiques de façon satisfaisante. Le même Internet se situe en effet à l’exacte croisée de ces deux usages, utilisant la déclinaison d’un texte médiatique comme fondement de la conversation. Ce statut particulier du même rend son approche complexe mais en fait également un objet passionnant.

Hypothèse : *Les mêmes sont à comprendre comme des actes d’énonciation.*

Les analyses de notre corpus nous ont permis de montrer que les hashtags ne recouvraient pas le statut particulier que nous attribuons ici au même (voir section 3.3.3). Objet hybride, le hashtag reste avant tout un artefact exogène à la conversation, plutôt qu’un support discursif *per se*. L’irruption du caractère dièse (#) dans le langage témoigne de ce caractère peu naturel. Même s’il tend à se généraliser, son usage nécessite une appropriation, souvent par jeu de connotation ou de dénotation, et de ce fait participe d’une sous-culture des plate-formes web. A l’opposé, le même ne se construit pas comme artefact remplissant une fonction annexe dans la conversation (identifier un événement, un lieu, une marque) mais bien comme une discussion en soi. Comme nous avons pu le constater de manière expérimentale (voir chapitre 3.3), identifier un même dans un ensemble de conversations est une tâche difficile. Les approches algorithmiques ne peuvent s’appuyer sur des constructions en-dehors du langage et sont donc inefficaces (voir 3.3.2). Finale-

ment, c'est l'entrée par le mot lui-même qui permet d'accéder à la représentation la plus fidèle du même (voir 4.1.1). Le régime particulier de l'énonciation que nous avons choisi pour appréhender l'existence des mêmes nous permet de considérer différents aspects de la conversation. Ce faisant, nous mettons également à jour la limite de notre méthodologie fondée sur l'analyse de données, à jamais incapable de voir l'action ou l'intention qui a motivé un acte de parole en ligne.

Nos résultats semblent avant tout indiquer que plusieurs régimes d'expression et de conversation co-existent sur les réseaux sociaux. Les mêmes humoristiques semblent mieux correspondre au régime de l'énonciation. Proches de l'humour, ils définissent des lieux intimes au sein du vaste espace d'expression qu'offre les plate-formes Web. A l'inverse, les campagnes de marketing ou de communication procèdent plus du régime du discours, cherchant à territorialiser des domaines précis notamment par la création de signes nouveaux (les hashtags par exemple, équivalent des marques du Web). Ces deux régimes co-existent largement au quotidien et ne sont pas nécessairement exclusifs. C'est dans la rencontre de ces différents modes d'expression que se nichent sûrement les approches les plus intéressantes de l'analyse des échanges en ligne.

Hypothèse : *L'analyse de la diffusion doit mobiliser les modèles étudiant l'énonciation*

Les modèles à construire de l'analyse de données doivent donc s'inspirer du dialogue mais aussi prendre en compte les parties manquantes ou incomplètes lors de l'analyse. Les actes de communication observables sur les réseaux sociaux existent dans une continuité physique, sont le produit de corps qui parlent ou écrivent. Le statut de trace que constitue le texte écrit sur les réseaux sociaux donne à la conversation une nouvelle matérialité. Considérer de façon autonome l'existence textuelle des messages échangés peut être intéressant dans une étude linguistique. Néanmoins, dans le cadre d'une étude sur la communication, il paraît frauduleux de réifier la conversation en un simple échange textuel. De même, considérer uniquement les conversations comme des "flux" entre utilisateurs échoue à prendre en compte la pré-existence de leurs relations et du contexte sur l'acte de communication.

Le système de visualisation que nous avons choisi de développer voulait mettre en perspective les différentes dimensions de la conversation dans un même espace visuel (voir section 4.1.3). L'analyse d'un échantillon de contenus issus de *Sina Weibo* nous

a permis de vérifier partiellement notre hypothèse méthodologique. En donnant à voir différents aspects des échanges et conversations, nous avons cherché à observer les phénomènes de diffusion en prenant en compte les quoi, qui, comment et où de ce qui est dit. Le développement d'un dispositif technologique sur-mesure nous a permis d'obtenir des visualisations répondant à ce besoin. Néanmoins, la disparité des données n'offre pas une granularité suffisante pour observer de façon détaillée chacun de ces aspects. Nous avons au demeurant pu montrer une image générale et observer que chacun d'eux possède des particularités. Les différentes figures obtenues que nous nommons *topogrammes* montrent plusieurs spécificités pour chaque type de contenus qui encouragent à poursuivre vers une plus grande diversité d'approche dans l'analyse des échanges en ligne.

*Hypothèse : L'analyse et la visualisation de données permette d'établir une classification des discussions en ligne selon les structures de leurs diffusions.*

La lecture des différents topogrammes nous permet d'ébaucher une première typologie des contenus en ligne selon les modalités de leurs diffusions (voir fig. 4.2.2). Néanmoins, notre étude à dimension expérimentale s'intéresse à un nombre restreint de mêmes. Cette première validation notre hypothèse méthodologique nécessiterait une systématisation à des corpus plus diversifiés afin de confirmer qu'il est réellement possible de constituer une typologie générale et cohérente des *topogrammes* des différents contenus disponibles en ligne. Nos premières observations permettent de tirer des conclusions provisoires sur les spécificités des différents types de contenus.

La diffusion des faits d'actualité est entourée de grands volumes de discussion, qui lui permettent de se propager rapidement. Dans le cas du fait divers, son démarrage brutal puis sa transformation par le commentaire en débat de société parfois brûlants se caractérise par un grand nombre de communautés très actives dans la discussion pendant un laps de temps plutôt court. La conversation s'oriente généralement autour de termes simples et structurés selon des chemins assez typiques. Les conversations autour de faits d'actualité sont au début composées de communautés diverses. L'enjeu pour le contrôle de la discussion est de s'approprier le cœur de la diffusion, souvent en centralisant et fédérant l'ensemble des conversations. Cette stratégie qui cherche à couper court à une diffusion non-programmée de l'information semble particulièrement présente en Chine. Une des stratégies est notamment de relayer et de diffuser le maximum d'information afin

de recentrer la discussion autour de mots et de groupes d'individus bien définis.

Les discussions entourant une actualité politique créent des communautés importantes qui restent néanmoins relativement distantes. Discutant sans vraiment se rencontrer, les différentes communautés se livrent à des échanges très polarisées avec peu de dialogues. La projection cartographique montre que Canton joue un rôle de précurseur dans les faits d'actualité et paraît "sortir" les affaires. Pékin agit plutôt comme diffuseur en disséminant largement l'information. Ce résultat renforce l'hypothèse d'une similarité entre réseaux sociaux et médias traditionnels, puisque Canton possède traditionnellement une presse plus portée sur l'investigation que celle plus officielle de Pékin. Les disparités entre Est et Ouest sont également flagrantes lors de la diffusion, avec une domination des grandes zones urbaines de l'Est chinois qui, même pondérées, restent omniprésentes sur les cartes de la diffusion. Les zones de l'Ouest souvent moins urbanisées semblent être moins associées aux discussions, sauf dans le cas où celles-ci les concernent directement (comme le Xinjiang)

L'émission de télévision *The Voice* propose en quelque sorte un contre-exemple du même. Le réseau de conversations entourant cette émission s'organise autour de très peu de diffuseurs dialoguant peu mais entourés de très nombreux fans. Les mots de moindre importance sont exclus des conversations, contraints à la marginalité dans le réseau sémantique. La diffusion de *The Voice* ne possède pas à proprement parler les spécificités qu'on attendrait d'une diffusion sur les réseaux sociaux avec peu de conversations. Son topogramme se caractérise par une très forte agrégation et une très faible modularité (peu de groupes fortement dominés par des diffuseurs importants). Il ne se pose pas du tout comme un lieu commun du Web, ouvert et associatif, mais plutôt comme une zone d'échanges clairement identifiée et territorialisée.

L'observation des spécificités de diffusion des faits d'actualité ou des campagnes commerciales nous a permis de vérifier des choses déjà connues : l'actualité se propage selon des cycles courts, les médias de Canton sont plus avant-gardistes que ceux de Pékin, la communication télévisuelle est soigneusement préparée, etc. Comme souvent dans les études utilisant l'analyse de données, ces quelques observations viennent infirmer des connaissances préalables sans vraiment apporter de nouveautés.

Les modèles peu connus des mèmes absurdistes sont une des dimensions intéressantes mises à jour dans cette étude. En effet, les topogrammes des mèmes comiques possèdent

des structures atypiques. Contrairement aux topogrammes qui présentent des graphes conversationnels bien organisés, ceux des mêmes *dufu* et *yuanfang* sont fragmentés en nombreux petits groupes de discussions très intégrés qui communiquent très peu entre eux. Tout se passe comme si nous n’assistions pas à une vaste conversation mais plutôt à de nombreuses petites conversations. Les origines géographiques diverses des utilisateurs montrent que la blague fonctionne localement. Le caractère anecdotique et humoristique semble être un fort vecteur de diffusion avec une présence beaucoup plus durable dans le temps que dans le cas des actualités. Leurs graphes sémantiques se composent autour de peu de mots-clés qui sont réutilisés de manière non-définitive, permettant ainsi une grande variation et une grande appropriation par des utilisateurs même isolés. Cette structuration des conversations en groupes lexicaux ouverts aux associations peu communes semblent favoriser la diffusion des mêmes et leur permettre de durer dans le temps. Les utilisateurs de Taiwan sont présents dans la diffusion des mêmes absurdistes, alors qu’ils sont absents des discussions politiques. Ainsi, les mêmes possèdent des topogrammes associatifs qui encouragent une participation durable.

Parmi les éléments observés dans notre échantillon, seuls *dufu* et *yuanfang* semblent réellement correspondre à la définition de “*mème*”. Les débats de société et les faits d’actualité s’apparentent davantage à des “événements web”, reconduction de phénomènes pré-existants à l’Internet. Les faits divers ont toujours été débattus, même si l’Internet en change radicalement l’échelle. Les modèles issus des médias traditionnels ont également été reconduits comme le montre l’exemple de *The Voice*. Ce sont bien les mêmes comiques qui présentent le modèle le plus atypique de cet ensemble.

L’humour, s’il pré-existe heureusement à l’Internet, procède d’un régime éminemment conversationnel et prend parfois la forme particulière des mêmes sur Internet. Le jeu de mots, le détournement d’images ou de slogans sont des pratiques communes qui se voient reconduites sous des formes nouvelles en ligne. L’étude de ces formes humoristiques permet d’accéder à des conversations pas nécessairement planifiées et possédant un caractère plus informel. Ce type d’échanges presque spontané semble pourtant suivre des modèles identifiables qu’il est possible d’étudier plus en détails grâce à l’usage des topogrammes.

Hypothèse : *La circulation des contenus sur les réseaux sociaux accroît la fragmentation du milieu numérique.*

Les différents topogrammes montrent la structure atomique et rhizomique des échanges en ligne. Nous avons pu voir comment les communautés de discussions se constituent. Ce modèle fondé sur l’inclusion ou l’exclusion de groupes dans les échanges est un facteur de fragmentation (voir 1.3.3). Les mêmes humoristiques présentent notamment la particularité intéressante de générer beaucoup de conversations pendant une longue durée, mais seulement au sein de petits groupes et non entre ses groupes. L’effet “*private joke*”, composante importante des mêmes Internet, fait de la blague partagée un signe de reconnaissance mutuelle (voir chap. 2). Néanmoins, dans cette circulation ne se joue pas nécessairement une “socialisation” conçue comme le propre des réseaux mais plutôt une fragmentation en petits groupes à l’identité forte. Cette différenciation entre utilisateurs au sein d’une même plate-forme constitue lors de la circulation des contenus des communautés d’appartenance.

La lecture des topogrammes permet d’inférer la nature plus ou moins dissociative ou associative de certains contenus. Nous avons vu comment des diffusions plus ou moins centralisées et contrôlées invitent ou non à la conversation. Nous avons également montré que les modèles dissociatifs et associatifs coexistent largement sur le réseau social *Sina Weibo*, même si l’activité historique de *Sina* comme fournisseur de contenus semble produire un espace de discussion où les contenus plus dissociatifs dominent. La vivacité des débats entourant les développements économiques, sociaux et politiques de la Chine moderne donne lieu à de nombreuses tentatives d’associer la population chinoise par les médias. Les mêmes quant à eux remettent en question cette dichotomie entre nature associative et dissociative d’un événement ou d’un média. Les vastes conversations des mêmes humoristiques impliquent activement de nombreux utilisateurs, sans pour autant les associer réellement au sein d’un dialogue. En définissant des groupes, les mêmes renforcent la fragmentation du réseau mais n’associent pas nécessairement les acteurs entre eux.

La faible taille de notre échantillon ne nous permet pas d’observer de manière empirique les structures récurrentes à un échelle suffisante. Néanmoins, les topogrammes montrent que la description des actes de communication ne peut suivre une simple dichotomie. La multitude des situations rend difficile la qualification de façon définitive d’un espace de conversation en ligne. La tâche devient réellement périlleuse lorsqu’il s’agit de qualifier un milieu numérique. L’approche “par le milieu” que nous avons présentée en introduction connaît donc des limites importantes. La plasticité théorique du terme mi-



lieu en fait également un outil faiblement opérationnel du point de vue méthodologique. L'histoire du mot teintée de déterminisme et plus généralement l'ambition globalisante du concept nous dirigent vers des généralisations tentantes mais néanmoins hasardeuses. L'analyse de données comme la recherche de terrain connaissent bien des difficultés à observer empiriquement ce "milieu" dont l'existence et les limites restent finalement non résolues par l'étude. Les données contenaient seulement une indication de lieu vague donnée par les utilisateurs. L'absence de géolocalisation en tant que telle n'a pas permis une étude extensive des modèles géographiques mis en œuvre lors de la diffusion des mèmes. Les observations faites dans les résultats viennent tout au plus valider des éléments déjà connus, en disant plus sur l'outil et la méthode que sur les phénomènes réellement observés.

La tâche plus large de compréhension et de description du rôle des échanges en ligne dépasse largement les seuls réseaux sociaux. Loin d'être une exception chinoise, le contrôle du caractère dissociatif et associatif des nouvelles formes langagières du Web est une problématique entourant les technologies de la parole dans le monde entier. Les premiers apports méthodologiques de cette étude demandent désormais à être poursuivis, testés et discutés sur différents terrains d'études. La description des invariants et variations au sein d'un corpus plus large de mèmes ou d'autres formes textuelles (e-mail, commentaires, etc.) pourrait permettre de définir les canons des nouvelles formes d'énonciation en ligne, comme autant d'archétypes de la façon dont s'écrivent les lieux communs du Web. Une étude plus large concernant un ensemble de corpus de différents médias pourrait donc à terme permettre de classer les topogrammes et de dresser ainsi une typologie des actes de communication publiques ou privés.

Une fois identifiés, les éclairages apportés par les topogrammes sur les faits médiatiques peuvent supporter l'analyse de la diffusion mais également produire un support plus stratégique de conception des actes de communication. En identifiant des topogrammes précis, on peut offrir une lecture qualitative des caractéristiques à différents moments (diffusion de masse, faits divers, artefacts de communication, etc.). Ce procédé de vérification peut s'avérer utile dans la vérification des sources journalistiques sur les réseaux sociaux ou la communication en situation de crise notamment. Les notions d'impact et d'influence peuvent également être réajustées à l'aune de facteurs paramétriques et contextuels propres à des topogrammes particuliers, notamment en terme de moment et de

centralité dans les réseaux sémantiques, géographiques ou conversationnels. De même, une réévaluation du concept de milieu numérique pour une prise en compte plus importante des phénomènes physiques (appartenances territoriales, réseaux de lieux, etc.) offrirait une meilleure assise théorique à cette méthodologie d'analyse, permettant notamment à l'analyse de données de s'articuler plus facilement à un travail de terrain.

Cette limite conceptuelle inhérente s'exprime ici notamment dans la visualisation des conversations sous forme de graphe. L'espace de la représentation où le graphe est projeté possède des propriétés topologiques non définies qui devraient pourtant être celles du milieu numérique, s'il eût été observable.

Comment pourtant définir un milieu s'il est propre à un usage voire à une personne ? La description des technologies mises en œuvre dans les actes de communication est-elle suffisante ? Comment le milieu et ses protocoles existent-t-il lors de la transduction et l'individuation ? Cette existence est-elle alors observable ? Au regard de ses limites conceptuelles fortes, l'idée de milieu numérique gagnerait sûrement à être problématisée autour d'une définition plus précise des relations directes existantes entre le cyber-espace dans sa forme physique (les serveurs, machines, usagers, etc.) et sa manifestation "en ligne". L'histoire des sciences et plus particulièrement de la géographie ont montré comment le concept plutôt déterministe de milieu peut être repensé en termes d'espace, de lieu et de territoire, voire de paysages offrant des prises théoriques et des approches méthodologiques nouvelles.

Peut-être la vision trop généralisatrice voire éthérique induite par l'usage du concept de milieu pourrait être précisée par une approche incarnée dans des lieux, des espaces et des territoires, support structurant physiquement à la fois pour l'étude mais également pour la représentation. Le choix et le statut de l'espace de la représentation des phénomènes d'échanges en ligne est donc une des questions centrales que soulèvent à la fois l'usage de la visualisation et le concept de milieu numérique. Ainsi, l'ambition de cette recherche a été d'apporter une pierre à cette réflexion en réfléchissant à l'articulation possible de différents espaces de projection. Nous avons notamment essayé de mobiliser des apports théoriques de différentes disciplines afin de problématiser notre lecture dans un contexte élargi, donnant à voir une existence à la fois langagière et géographique des actes de communication.

Le cas de la Chine nous a fourni un terrain idéal pour s'interroger sur la production et la matérialité de cet objet physique devenu espace d'expression qu'est l'Internet. Un vaste travail doit encore être mené sur les usages de l'Internet pour repenser la séparation artificielle entre “bits” et “atoms” et resituer la magie du “*cloud*” dans le domaine de l'actuel. La distinction chaque jour plus ténue entre *réseaux*, *software* et *hardware* rend notamment le travail sur la réalité physique et situationnelle des actes de communication de plus en plus pressant. Pour rendre lisible les processus d'individuation en jeu dans les usages des technologies du réseau, nous avons besoin d'un cadre de lecture qui s'intéresse aux particuliers de l'Internet et à leurs relations avec l'environnement, jusqu'aux mines dont proviennent les ordinateurs.

La linéarité de la *timeline* constituant normalement l'interface de lecture des réseaux sociaux ne doit pas non plus éclipser la diversité des formes d'expressions qui s'y déroulent. L'analyse de différents types de contenus nécessite le recours à des approches particulières, héritées des réflexions sur les différents régimes du discours. Les méthodes mobilisées pour l'étude des actes de communication en ligne doivent être capables de recourir aux savoirs qui les précèdent, sur le langage notamment. Nous avons discuté dans cette étude des travaux sur l'énonciation, des lieux communs et de l'analyse rhétorique ou encore du modèle *word-of-mouth* de la diffusion en ligne. Ces théories considérées conjointement forment un ensemble en apparence hétéroclite. Néanmoins, chacune de ces approches peut faire sens si elle est invoquée pour l'analyse d'un type de contenu adéquat. Identifier le régime d'expression qui accompagne un objet numérique, voire son existence et son évolution entre discours et énonciation peut permettre une approche plus juste des échanges sur Internet. Au-delà d'une grammatologie des contenus, l'étude des usages doit chercher à prendre en compte les acteurs multiples prenant part aux phénomènes observables sur les plate-formes du Web.

L'interrogation sur l'existence ou la conception d'objets numériques doit donc s'accompagner d'une recherche dans la cohérence des intentions de ces différents acteurs. Notamment, la recherche d'utilisateurs “représentatifs” semble vaine tant les intentions qui président à la production d'un contenu peuvent différer. Sur une plate-forme comme *Sina Weibo* ou *Facebook*, un même utilisateur peut dans la même heure partager une blague avec l'un de ses amis, promouvoir une action collective ou sa propre marque puis commenter un article de journal. La recherche d'une cohérence intrinsèque dans l'ensemble de ces

pratiques doit considérer la diversité de ces différents modes d'expression. Ainsi, l'étude de l'activité ayant cours sur les réseaux sociaux et à plus forte raison sur Internet doit prendre en compte les moments et situations qui amènent à la création des objets numériques. D'une part, l'actualisation de notre milieu numérique sous une forme particulière (même, news, etc.) dépend de la structure technologique économique et politique de l'espace d'expression défini par la plate-forme Web - ici *Sina Weibo*. D'autre part, l'intention dans l'acte de communication doit être interrogée comme une des clés de compréhension qui permettront une lecture pertinente des usages. La complexité de cette tâche nécessite une attention soutenue aux contenus ainsi qu'une approche méthodologique renouvelée dont nous avons tenté de montrer ici un exemple.

## Validité et limites des méthodologies Big Data

La validation des hypothèses par un travail d'analyse de données est une tâche délicate. Nous avons pu voir comment l'accroissement du volume de données amène à l'augmentation de la complexité des processus de traitement de l'information. Les algorithmes pouvant donner des résultats utiles sont souvent sophistiqués, obligeant à un difficile travail d'optimisation des calculs. Ces couches successives de développements ingénieriques doivent la plupart du temps être effectuées spécifiquement pour une tâche ou un jeu de données précis. La cohérence et la pertinence des résultats est alors conditionnée à un travail d'écriture du code.

Le concept de *validité interne* décrit la pertinence ou la justesse de l'explication fournie par le chercheur pour expliquer les résultats qu'il a obtenus dans son expérience (Yin, 2009). Dans le cas de l'analyse de données, la reproductibilité est conditionnée à la disponibilité des données et la disponibilité du code et des outils qui ont servi à mettre en œuvre l'analyse. Ainsi, nous avons choisi de publier le code et d'utiliser un jeu de données largement téléchargeable. Le travail d'écriture associé à la production de code est à la fois pratique et réflexif. La possibilité de consulter les lignes ayant été écrites pour réaliser une analyse spécifique permet de s'approprier dans le détail la méthode utilisée. En effet, contrairement aux domaines de l'algorithmique ou des mathématiques appliquées qui se fondent sur des développements logiques, les méthodologies de l'analyse de données se construisent sur l'éventualité d'obtenir un résultat. Bien souvent, l'analyse statistique

ne nécessite pas de créations algorithmiques complexes, mais plutôt l'utilisation d'outils existants. Un des travaux importants menés par la communauté scientifique en informatique est la publication de librairies pour l'usage scientifique aux qualités d'écriture et performances rigoureuses. De plus, ces librairies sont souvent conçues pour être utilisées par d'autres personnes, étape de validation supplémentaire du travail.

Le code écrit avec comme but l'analyse de données n'est pas soumis aux mêmes contraintes. Tout d'abord, il est très peu publié ou réutilisé et à ce titre ne nécessite pas de rigueur particulière dans l'écriture (tests unitaires, commentaires, lisibilité, disponibilité, etc.). Sa formulation sous forme algorithmique ne représente pas non plus nécessairement fidèlement son action. En effet, le premier travail de l'analyste de données est de repérer parmi l'abondance des librairies disponibles les outils les plus adaptés. Ces choix sont difficilement communicables puisqu'ils dépendent d'un ensemble de critères souvent définis par l'analyste lui-même. Le travail d'écriture et de référencement des librairies utilisées se fait directement au contact de son jeu de données. Alors que la science algorithmique est un développement ayant pour base la logique, l'analyse de données s'apparente davantage à une discussion avec un jeu de données. Le travail préliminaire pour l'identification de modèles (voir section 3.3) se présente sous la forme d'un ensemble d'expérimentations *nécessaires* qui ont permis à l'analyse des données d'obtenir des résultats. La tâche dans cette partie du travail était de pouvoir décrire un modèle de même compréhensible par un ordinateur. La définition complexe et débattue du modèle ne s'écrit donc pas une fois pour toute, mais procède d'une itération qui nécessite de nombreuses expérimentations. Ces séries d'essais permettent de construire une représentation de l'objet recherché. Les informaticiens disent d'ailleurs souvent que l'un des problèmes les plus complexes de la programmation est de nommer les objets<sup>28</sup>. Comme dans le travail de terrain, les modèles conceptuels doivent passer cet "épreuve du réel". Ce dialogue avec les données s'écrit dans un ou des langages informatiques. Comme pour les entretiens de l'anthropologue, il semble nécessaire de documenter le code écrit de façon rigoureuse car il constitue finalement un des aspects fondamentaux de la méthodologie et des résultats de ce type d'étude. La formulation sous forme d'algorithmes ne saurait remplacer le texte original ayant servi à construire l'analyse.

---

28. "There are only two hard problems in Computer Science : cache invalidation and naming things.", citation attribuée au co-fondateur de Netscape Phil Karlton

Le passage à une validité dite externe, c'est à dire à une généralisation à d'autres terrains ou populations, dépend notamment de la nature de ces préalables. Quelles sont les questions exactes qui ont été posées et comment ont-elles été formulées ? La longue réflexion des sciences ethnographiques et anthropologiques sur le rôle du regard et la présence de l'observateur doit se poursuivre dans les méthodologies liées à l'analyse de données en sciences sociales. Des conclusions faites sur la base de questions mal formulées sont nuisibles. Le texte de ces questions est écrit sous la forme de scripts, référençant et mobilisant des travaux d'auteurs précédents. Ces travaux se trouvent pour la plupart sous la forme de librairies, patiemment testées et sélectionnées. Comme celui de l'étude de terrain, ce savoir-faire doit être sujet à une réflexion critique sur la base de références et d'exigences communes.

Les données sont les traces d'activités passées et doivent être comprises en tant que telles. Une vaste reconstruction du contexte et des intentions est nécessaire. Il s'agit même bien souvent de l'enjeu central de ce type de méthodologie : définir un manque et le discuter. Le présent travail s'est écrit en forme d'interrogation sur les limites et les enjeux de l'usage des données issues de réseaux sociaux dans les études en sciences humaines. À ce stade, il est difficile voire impossible de généraliser les résultats de notre étude. L'absence d'informations de géolocalisation a notamment rendu difficile l'approche cartographique. Caché, mal cerné, le "terrain" est quasi-absent et exige donc d'être reconstruit ou donné à voir par l'intermédiaire d'angles et de questionnements précis.

Alors, le dialogue des disciplines scientifiques s'impose comme la seule solution pour pallier ces limites intrinsèques de l'analyse de données. La faible connaissance des langages informatiques dans les équipes en sciences sociales à l'heure actuelle (Wieviorka, 2013) rend difficile la définition d'un effort commun dans ce domaine. Pourtant, la compréhension d'un jeu de données spécifique offre une opportunité intéressante de construire une approche pluri-disciplinaire autour d'un même objet. Les échecs et réussites montrent l'importance des choix méthodologiques justes dans la mise en perspective des nombreux savoirs disponibles.

Finalement, comme tout travail scientifique, les biais des méthodes et la fragilité des résultats existent comme autant d'opportunités invitant à la discussion et aux débats pour en repousser les limites.

# Conclusion

LE DISPOSITIF D'ANALYSE que nous avons présenté ici nous a permis d'observer la diversité des contenus circulant sur le réseau social chinois *Sina Weibo*. Nous avons pu aborder les différents aspects sémantiques, conversationnels et spatio-temporels entrant en jeu dans la construction des objets numériques. Le regard porté sur ces échanges quotidiens nous a mené vers une première typologie des faits de communication d'après leurs *topogrammes*. Cette modélisation schématique a également permis de mettre à jour l'articulation entre énonciation et discours dans la production de différentes formes médiatiques.

## Topogrammes et régimes d'expression en ligne

Le même *Internet*, proche de la blague en ligne, procède d'une énonciation spontanée. Son topogramme montre des petits groupes discutant très activement sans pour autant converser entre eux. Les mêmes absurdistes sont avant tout des jeux de mots. Leur champ lexical se constitue d'une grande variété de déclinaisons autour d'un ou deux mots-clés situés au centre de l'intérêt. Suivant des parcours géographiques atypiques, ils se constituent autour des foyers d'adoption très localisés et offre finalement une diffusion peu prévisible. Parmi les différents contenus observés, les mêmes comiques semblent être les plus tenaces, apparaissant régulièrement même longtemps après leur première apparition.

A l'inverse, certaines constructions comme les campagnes par *hashtags* semblent vouloir territorialiser des espaces d'expression entiers par des stratégies d'actions planifiées. L'enjeu de la communication stratégique en ligne et performatif, il s'agit de "faire dire" plutôt que de simplement exprimer. Le champ sémantique entourant une émission de di-

vertissements télévisée se caractérise par une architecture convenue autour de mots-clés attendus. La conversation se centralise autour de peu de diffuseurs qui posent les questions, suggèrent les émotions et crée ainsi une activité très intense pendant un temps très bref. Une analyse portant sur les hashtags les plus utilisés dans le corpus a notamment montré les grandes similarités existant entre les contenus majoritaires des réseaux sociaux et ceux d'autres médias de masse comme la télévision (divertissements, loisirs, produits culturels, etc.)

Les faits divers et les débats sur l'actualité regroupent quant à eux un grand nombre d'utilisateurs qui discutent de façon plutôt ordonnée en petits groupes indépendants. Leurs diffusions semblent également suivre les modèles géographiques observés traditionnellement en Chine. Les actualités plus controversées sont portées par des médias plus libéraux de Canton alors que Pékin joue un rôle de diffuseur central dans les annonces officielles. Certains médias et télévisions locales rayonnent largement dans des domaines particuliers - comme le divertissement - sans pour autant occuper une place prédominante ailleurs dans la vie médiatique chinoise.

De nombreux modèles de diffusion coexistent donc sur les réseaux sociaux chinois. Chacun d'entre eux peut être considéré selon un régime d'expression particulier. La diffusion fortement organisée et planifiée des campagnes publicitaires semble satisfaire une lecture foucauldienne des relations de pouvoir établies par le discours. Les discussions entourant les scandales et les faits d'actualité peuvent être envisagées comme un espace de négociation de l'ordre du discours par l'énonciation. Les mêmes, quant à eux, ne procèdent pas d'un régime discursif établi et bénéficieraient d'être envisagés comme des actes d'énonciation.

Nous constatons que les phénomènes observables sur les réseaux sociaux sont très hétérogènes et difficilement envisageables dans leur ensemble. Leur cohérence peut être abordée par l'étude des plate-formes et des outils, souvent riches en enseignements. Une étude des contenus et des usages doit prendre en compte la diversité des pratiques et des régimes d'expression en présence. Sur Internet comme dans la vie, on ne parle pas de la même façon à un inconnu, un ami, un collègue ou un représentant d'une marque de vêtement. Ces différences ne sont pas seulement des subtilités de l'analyse, mais bien des composantes structurelles de l'existence de ces conversations, comme nous l'avons vu à partir des différents topogrammes.



Ainsi, il paraît hasardeux de vouloir qualifier a priori les objets numériques. La notion de *milieu numérique* nous a permis de problématiser les relations entre usagers, protocoles et producteurs de contenus. Néanmoins, il semble difficile de définir précisément son champ d'application. Les topogrammes portent peu de traces de l'actualisation d'un milieu constitué de protocoles et témoignent plutôt de la diversité des usages possibles. Des études plus spécifiques et localisées seraient plus à même de vérifier l'opérabilité du concept de milieu numérique.

## **Les médias sociaux de masse**

La forte contextualisation de cette étude autour du cas *Sina Weibo* nous a permis de considérer le rôle du contexte économique et politique dans la production et la circulation des contenus. Nous avons vu comment les faits divers et les scandales politiques faisaient couler beaucoup d'encre en mobilisaient particulièrement l'attention dans les grandes villes. Ces échanges sont un reflet de la vivacité des débats entourant les développements économiques, sociaux et politiques de la Chine moderne. Les multiples contradictions et questions amenées par l'urbanisation croissante du pays trouvent des expressions quotidiennes, notamment par la satire et la dérision. Les mêmes comiques sont souvent teintés d'un humour noir, exutoire et passe-nerfs des angoisses actuelles des internautes chinois.

Cette nécessité de s'exprimer et de communiquer se corrèle de façon directe à une stratégie économique et politique de développement industriel du secteur médiatique. Le protectionnisme et les investissements d'État soutiennent activement la croissance de services Web commercialement très profitables. En contrepartie, *Sina Weibo* doit bien souvent jouer le rôle d'intermédiaire dans les échanges parfois houleux entre gouvernement et utilisateurs. En s'orientant vers le divertissement, les industries de contenus des magazines ou la télévision trouvent des débouchés directs dans le microblog. L'intégration du diffuseur au reste de la chaîne de production de l'information est à la fois une tradition des grands groupes médiatiques, mais également une nécessité pour réduire les coûts engendrés par les régulations toujours plus drastiques de la publication d'information. En effet, l'instabilité de la qualité du service due aux décisions politiques amène des pertes non-négligeables chez les entreprises Web. Ce manque-à-gagner peut être compensé par une diversification, permettant également un contrôle accru de la production

d'information en amont.

Ces phénomènes de contrôle des contenus, souvent regroupés un peu rapidement sous le nom de “censure”, ont des équivalents directs pour la plupart des services Web dans le monde. La tendance législative en Europe tend actuellement à établir un cadre réglementaire pour les fournisseurs et distributeurs publiant des informations en ligne. A ce titre, *Sina Weibo* en fait pas seulement figure d'exemple exotique, mais établit dans ce domaine un précédent tant par sa réussite économique que son modèle de coopération gouvernemental.

## Limites et opportunités de l'analyse de données

L'analyse d'un vaste ensemble de données nous a permis de construire une typologie enrichie des diverses formes typiques de communication en ligne. Un outil de visualisation interactif montrant la circulation des mots et phrases entre les utilisateurs sur le territoire chinois nous a permis d'observer les topogrammes de plusieurs contenus choisis. Le choix des exemples étudiés et des outils utilisés a nécessité un processus expérimental qui constitue le cœur de ce travail. De nombreuses expérimentations et tentatives parfois infructueuses nous ont mené vers le développement d'un outil finalement utilisable.

La difficulté d'identifier des contenus pertinents dans une large masse de données a notamment été un des problèmes importants. L'approche algorithmique de détection par des méthodes dites de *machine learning* n'a pas abouti, nécessitant trop de ressources de calcul pour un résultat peu sûr. La seconde approche utilisant comme support les hashtags n'a pas permis de découvrir de mêmes intéressants. Nous avons pu mettre à jour d'autres aspects complémentaires contenus dans notre jeu de données. Enfin, l'indexation et la recherche plein-texte ont offert une solution satisfaisante, accompagnée d'une recherche parallèle dans la littérature académique et secondaire.

L'ensemble de ces tentatives plus ou moins réussies montre que le travail d'analyse de données s'effectue de manière itérative et non-définitive. Tâtonnements, tentatives et hésitations ont jalonné ce travail. L'exploration progressive du jeu de données grâce à des travaux préliminaires ont façonné les réflexions méthodologiques et théoriques finalement produites ici. Contrairement à la recherche en informatique et algorithmique qui s'intéresse aux possibilités et capacités d'analyse, la pratique de l'analyse de données est

conditionnée par un résultat extérieur au processus lui-même. L’ajustement des étapes du traitement et de la visualisation doit donc refléter sincèrement l’approche première de l’objet étudié. Ainsi, une connaissance aigüe du terrain doit se corrélérer à une maîtrise des outils et langages nécessaires à la réalisation de l’étude. Cette rencontre se produit dans la conjugaison de langages informatiques et humains afin de formuler et d’écrire un travail juste. La disponibilité des données et du code en donne le texte intégral et permet une ré-utilisation de tous les éléments produits ici <sup>29</sup>.

Le présent travail s’est donné pour tâche d’établir les fondements méthodologiques et conceptuels qui permettent de mener à bien une étude plus systématique des formes de circulation des contenus sur Internet. Les relations entre régimes d’expression et topogrammes définies dans ce premier travail nécessitent d’être affinées et approfondies. Nous avons notamment fait attention à ne pas généraliser trop rapidement des analyses portant sur le terrain bien particulier de *Sina Weibo*. Un projet de recherche post-doctorale financé par l’Agence National de la Recherche (ANR) autour des modèles d’innovation ouvertes en Chine <sup>30</sup> permettra de poursuivre les réflexions amorcées dans ce travail sur un autre terrain. Nous pourrons confronter les résultats observés par l’analyse de données sous forme de topogrammes aux expériences transcrites lors d’interviews ou de méthodes d’exploration plus qualitatives.

Ce travail continuera de questionner les objectifs, limites et objets dont peut se saisir le vaste champ d’études pluri-disciplinaire ouvert par l’analyse de données et l’usage des langages informatiques en contexte scientifique.

---

29. Le code et les données sont disponibles depuis <http://github.com/topogram>, consulté le 4 Juillet 2014 à 16h15

30. Appel à projet *Sociétés innovantes* (INOV) en 2013 : *OPIMPUC - OPen Innovation : Models and Places in Urban China*. Voir l’annonce officielle [http://www.agence-nationale-recherche.fr/en/anr-funded-project/?tx\\_lwmsuivibilan\\_pi2%5BCODE%5D=ANR-13-SOIN-0006](http://www.agence-nationale-recherche.fr/en/anr-funded-project/?tx_lwmsuivibilan_pi2%5BCODE%5D=ANR-13-SOIN-0006), consulté le 5 Juillet à 17h12



# Annexes



# Protocole de recherche

## Calendrier et étapes de la recherche

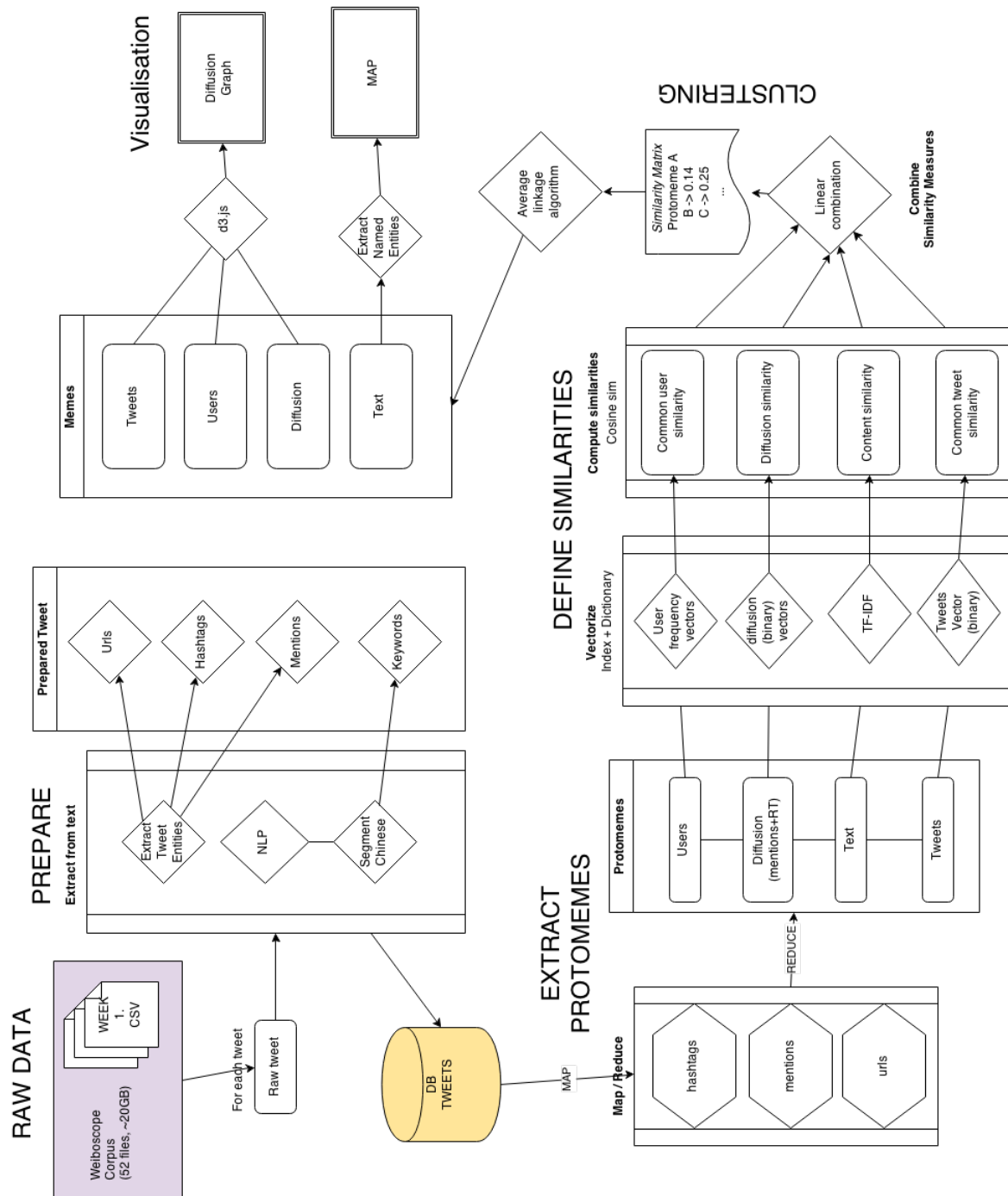
Le tableau suivant retrace les grandes étapes du déroulement de cette recherche.

Tâches	Date	Rendu	Encadrement
<b>2012</b>			
21 février	Inscription en doctorat à ParisTech Telecom	Paris	Mme Valérie FERNANDEZ & M. GILLES PUEL
12 Avril	Première présentation publique	Paris	ParisTech Telecom
28 Mai	CaiYuan Pei 2012	Toulouse-Nanjing	M. ZHEN FENG
Juin-Octobre	Social Brain Framework (SNS mining)	Online	Sharism Lab & M. Isaac MAO
Octobre-Décembre	Séjour de recherche 1	Nanjing / Shanghai	M. ZHEN FENG
10-12 Novembre	Meeting CaiYuan Pei	Université de Nanjing	Equipe Cai Yuanpei
8 Décembre	TEDx Wuxi	Wuxi	TEDx
<b>2013</b>			
Février	Echidna project	Shanghai	Ricky Ng-Adam (Xinchejian)
22-24 Mai	AIM2013 Big Data	Lyon	Colloque
6 Juin	Journées i3	Paris	Institut Mines-Telecom
Août-Octobre	Séjour de recherche 3	Nanjing / Shanghai	M. ZHEN FENG
Septembre	Analyse langue chinoise	Beijing	Yuan Mingli & <i>Guokr</i>
20 Septembre	Talk at Anti-Tectonics	Beijing	Beijing Design Week
10-30 Septembre	Smart City Exhibition	Beijing	Beijing Design Week
Novembre-Décembre	Collectes de données	Lyon	Atelier des Médias
<b>2014</b>			
Janvier	Algorithmique et détection	Lyon	Adrien GUILLE ET FABIEN VECLIN (LYON 2)
Février-Avril	Développement de l'outil	Lyon	Atelier des Médias
Janvier-Avril	Séminaire <i>Atelier Internet Lyonnais</i>	Lyon	Éric Guichard (ENSSIB)
Mai-Juin	Ecriture thèse	Lyon	Atelier des médias
Mai	AIM2014	Aix-en-Provence	Colloque AIM
Juin	Chinese Internet Research Conference	Hong Kong	Polytechnic HK
Septembre	Soutenance thèse	Paris	EDITE / ParisTech Telecom



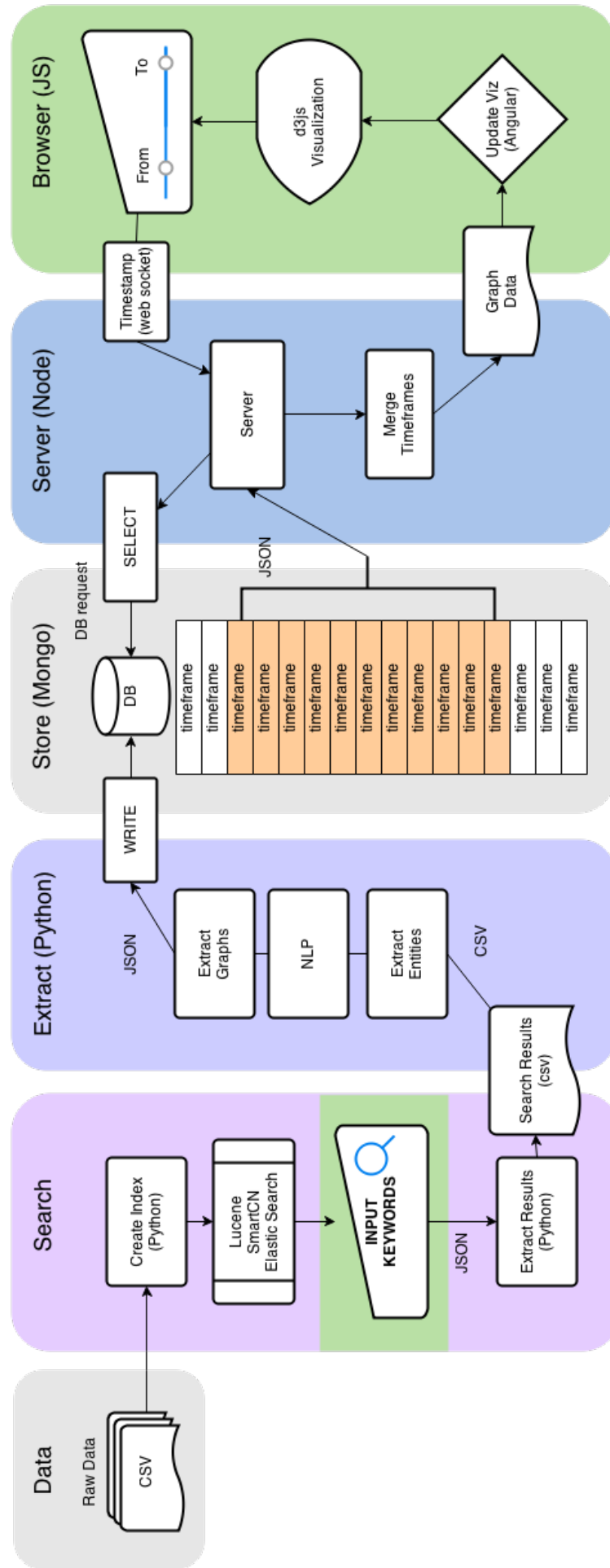
[

Schémas des algorithmes]Schémas des algorithmes



Graphique .0.1 – Protomemes algorithm workflow

**Graphique .0.2 – Data mining and visualisation workflow**



# Mememes : Iconographie

Cette annexe regroupe une sélection de documents iconographiques et textuels collectés pour décrire les mèmes et contenus abordés dans ce travail.

- Une liste des mèmes permet de considérer les classifications et mots-clés utilisés pour réaliser la typologie et extraire les données de chacun des mèmes.
- Une seconde partie détaille un par un les mèmes abordés en proposant une rapide description ainsi que quelques images ou messages qui lui sont associés.

Début	Nom	Type	Mots-Clés
March 21, 2012	Détournement calligraphie : “Du Fu is very busy”	Absurdiste, humour	杜甫很忙, 李白不服气了
July 13, 2012	The Voice : lancement de l’émission TV	Fan clubs, adoration	中国好声音, 吴莫愁
August 26, 2012	Yang Dacai, l’officiel “souriant” et le scandale des montres	Actualité, satire, commentaire social	表叔, 表哥, 微笑局长, 杨达才
October 1, 2012	Blague : “Yuan Fang, qu’est-ce que tu en penses ?”	Absurdiste, humour	远芳, 你怎么看?, 我觉得此事有蹊跷, 神探狄仁杰
October 11, 2012	Mo Yan reçoit le Prix Nobel de littérature	Fan clubs, adoration	莫言, 管谟业, 2012诺贝尔文学奖, 诺贝尔
November 8, 2012	18ème Congrès du Parti Communiste	Marketing politique, soutien, pétition	第十八次全国代表大会, 中共十八大, 18大, 十八大
November 20, 2012	La sex tape de Lei Zhengfu publiée par un journaliste	Actualité, satire, commentaire social	雷政富
December 3, 2012	Qiegao : échauffourées autour d’un gâteau du Xinjiang	Actualité, satire, commentaire social	切糕

**Tableau .0.1** – Liste des mèmes identifiés

## Dufu is very busy

*Du Fu* (712-770) (en chinois 杜甫 est un poète chinois de la dynastie Tang, célèbre pour sa prose teintée de réalisme où se croise scènettes de la vie quotidienne et descriptions des arcanes du pouvoir. Le mème “Du Fu est vraiment très occupé” 杜甫很忙 détourne une calligraphie bien connue pour mettre en scène le poète dans des situations improbables :

Voir la page sur *Know Your Meme* : <http://knowyourmeme.com/memes/du-fu-is-busy> et sur l'encyclopédie Baidu <http://wenku.baidu.com/view/941e25c805087632311212aa.html>



(a)



(b)

(c)



(d)



(e)



(f)

Graphique .0.3 – Dufu is very busy

## The Voice of China

*The Voice of China* (en chinois 中国好声音) est une émission de télé-crochet musical, diffusée en Chine depuis le 13 juin 2012 sur *Zhejiang Television*. *The Voice of China* est une adaptation de l'émission *The Voice of Holland*, dont les droits ont été vendues dans le monde entier. Un jury composé de membres de l'industrie musicale chinoise prend part à l'émission afin de juger l'interprétation de chansons populaires par des candidats choisis sur casting. Diffusés depuis Juillet 2012, les épisodes sont visionnés en moyenne par 7 millions de téléspectateurs et plus de 70 millions d'internautes (d'après Laurie Burkitt, "Why The Voice Is China's No. 1 TV Show" in *The Wall Street Journal* du 19/09/2013)

Voir la page officielle : <http://www.zjstv.com/voice/>



(a) Logo Officiel de The Voice of China

(b) Les membres du Jury de l'édition 2014



(c) "The Voice devient le sujet le plus discuté du mois de Juillet", d'après Hi News, <http://www.hinews.cn/news/system/2012/08/25/014863154.shtml>, consulté le 7 Juillet à 14 :32

(d) Un dessin posté par un utilisateur de Sina Weibo

Graphique .0.4 – The Voice of China : Illustrations



## Biaoge

En Août 2012, Yang Dacai, directeur du Bureau de Supervision de la Sécurité Routière de la province du Shanxi, se rend sur les lieux d'un accident de bus ayant fait 36 morts. Une photo diffusée sur Internet le montre discutant avec un policier, arborant un grand sourire. Agacé par cette excès d'incivilité, les internautes commencent à faire des recherches à son sujet et se rendent compte que l'officiel chinois arborent à chaque apparition une montre différente, toutes de marques prestigieuses. Les internautes le renomment rapidement *le frère aux montres* (*biaoge*) et collectent les photos de ses différentes montres de luxe comme autant de preuves manifestes de sa corruption. Quelques jours plus tard, Yang Dacai est démis de ces fonctions. Il sera par la suite jugé et condamné pour corruption, assorti d'une peine de 12 ans de prison.

Voir aussi : <http://www.scmp.com/topics/yang-dacai> ou <http://www.chinasmack.com/2012/pictures/chinese-government-official-smiling-at-tragic-accident-scene.html>



(a)



(b)



(c)



(d)

**Graphique .0.5 – Images du même entourant les montres de Yang Dacai**

## Yuan Fang, qu'en penses tu ?

Un dialogue issu d'une série policière très prisée présente un couple de détectives enquêtant dans la Chine médiévale. La phrase "Yuanfang, qu'en penses-tu" (元芳, 你怎么看?) est devenue aussi célèbre que le "élémentaire" de Sherlock Holmes au Dr. Watson.



(a)

(b)



(c)



(d)



(e)



(f)

Graphique .0.6 – Exemples d'images du même Yuanfang



## Moyan

Mo Yan (1955-) (en chinois 莫言) littéralement : « celui qui ne parle pas ») est un écrivain chinois célèbre. Le 11 octobre 2012, il a reçu le prix Nobel de littérature. Lors d'une interview sur une grande chaîne de télévision, il déclare ne pas pouvoir être heureux, trop stressé par sa récente réussite et par l'inquiétude que lui procure les changements de la Chine. Cette déclaration provoque rires, jalousies et moqueries sur la Toile.

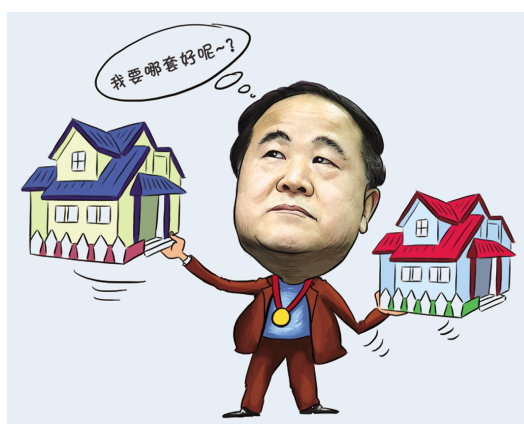
Voir sa page sur Wikipedia [http://fr.wikipedia.org/wiki/Mo\\_Yan](http://fr.wikipedia.org/wiki/Mo_Yan)



(a) La cérémonie du Prix Nobel de Littérature



(b) Moyan répondant "Je ne sais pas" à la question "Êtes-vous heureux?" par lors d'une interview sur la chaîne CCTV13



(c) Dessin d'un internaute moquant Mo Yan "Je ne sais pas laquelle choisir", d'après IT Times, <http://www.ittime.com.cn/index.php?m=content&c=index&a=show&catid=23&id=2662>, consulté le 7 Juillet à 14 :32



(d) Un dessin posté par un utilisateur de Sina Weibo

**Graphique .0.7 – Mo Yan Prix Nobel : Illustrations**

## 18ème Congrès du Parti Communiste

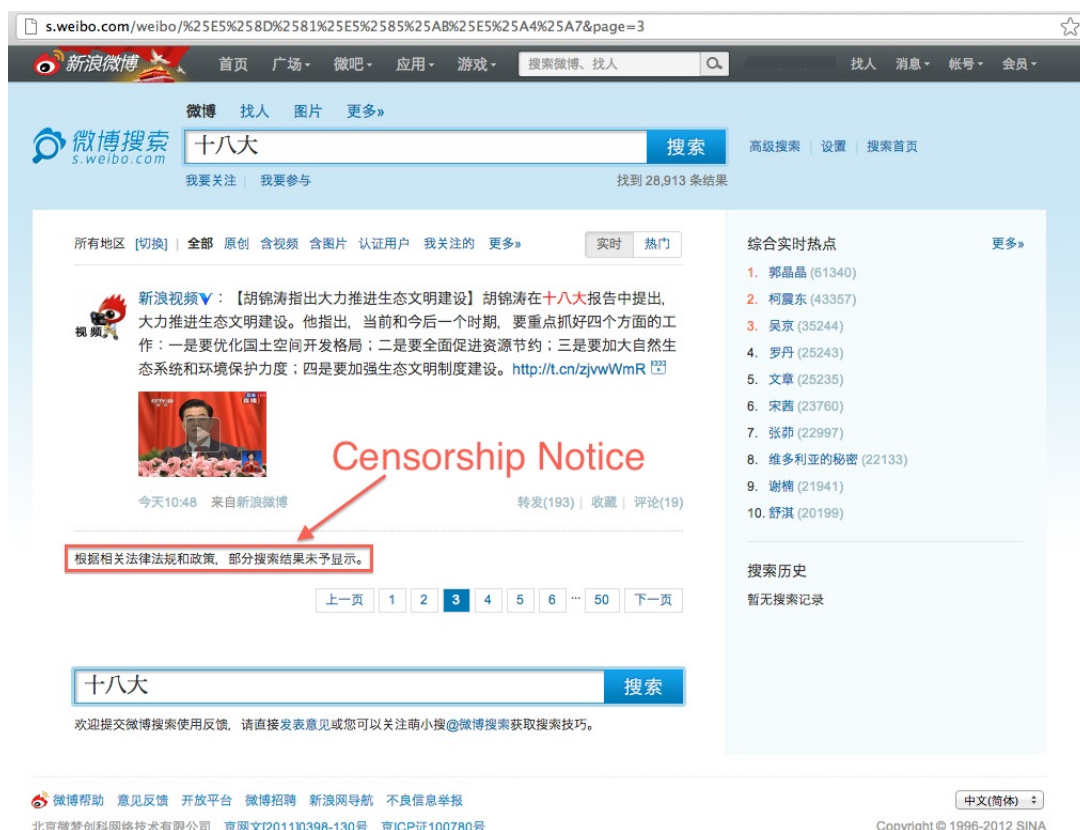
Le XVIII<sup>e</sup> congrès national du Parti communiste chinois (en chinois 中国共产党第十八次全国代表大会) s'est tenu en novembre 2012. Réunissant plus de 2270 délégués, ce congrès a élu Xi Jinping, l'actuel président chinois. La couverture médiatique de cet évènement majeur de la vie politique chinois en a fait un des temps forts de l'année 2012 dans les médias.



(a) Le siège du 18ème Congrès du Parti Communiste à Pékin



(b) L'ancien président chinois Hu Jintao (à gauche) avec son successeur Xi Jinping (à droite)



(c) Post bloqués lors de la recherche "18ème Congrès" sur Sina Weibo, d'après Feichang Dao, <http://blog.feichangdao.com/2012/11/in-week-before-party-congress-sina.html>, consulté le 27 Juillet à 16 :53

**Graphique .0.8 – 18ème Congrès du Parti Communiste : Illustrations**

## Sex Tape

Une vidéo montrant Lei Zhengfu, le secrétaire du Parti dans le district de Beibei à Chongqing, dans le plus simple appareil en compagnie d'une de ses maîtresses est publiée en ligne. A peine 63h après la publication de cette vidéo avec le titre "Lei, the secretary who accepts sex bribes", Lei est mis à pied de ses fonctions politiques. Cette vidéo qui avait été tournée 4 ans auparavant été utilisée depuis plusieurs années par un promoteur immobilier de la ville de Chongqing pour faire chanter Lei Zhengfu.

Voir aussi : [http://en.wikipedia.org/wiki/Lei\\_Zhengfu](http://en.wikipedia.org/wiki/Lei_Zhengfu) et <http://www.viddler.com/v/41bf67b2>, consulté le 7 Juillet à 12 :32



(a) Image tirée de la vidéo



(b) Photo de la femme



(c) Détournement d'un internaute



(d) Détournement d'un internaute



**Graphique .0.9** – Exemples d'image du même concernant la sextape de Lei Zhengfu. Images d'après l'article "Sex tape official stands trial in Chongqing" dans le *People's Daily*, <http://english.peopledaily.com.cn/90882/8290628.html> et <http://www.dailymail.co.uk/news/article-2239300/Zhao-Hongxia-Teenage-honeytrap-brought-Chinese-Communist-Party-official-sex-tape-pictures-leaked-online.html> consulté le 7 Juillet à 12 :32



## Qiegao

Qiegao (切糕) est une gâteau au sucre et aux noix, spécialité de la région du Xinjiang souvent vendus par des vendeurs itinérants. Le 3 décembre 2012, une dispute entre un client villageois du Hunan et un vendeur originaire de l'extrême-ouest de la Chine a tourné en altercation, avant de dégénérer en bagarre générale. Selon la police, le montant total des gâteaux abîmés a été estimé par la police à 160 000 yuans (environ 19 500€). Les internautes se sont emparés de cette affaire pour en faire une plaisanterie, comparant la valeur du Qiegao à celle de l'or.

Voir également <http://world.time.com/2012/12/05/dont-let-them-eat-cake-how-ethnic-tensions-in-china-explode-on-the-streets/> ou <http://chine.blogs.rfi.fr/article/2012/12/05/hans-ouighours-c-est-pas-du-gateau>



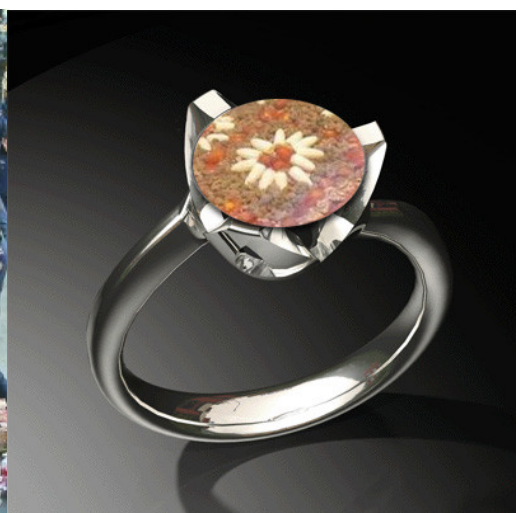
(a) Un vendeur de qiegao



(b) Le gâteau



(c) Une image de l'affrontement avec la police posté par un utilisateur de Sina Weibo



(d) Création d'un internaute montrant combien le qiegao est cher

Graphique .0.10 – Qiegao

# Bibliographie

- LA ADAMIC, TM LENTO, Eytan ADAR et PC NG : Information Evolution in Social Networks. *arXiv preprint arXiv :1402.6792*, pages 1–19, 2014. URL <http://arxiv.org/abs/1402.6792>.
- Jimi ADAMS, Katherine FAUST et Gina S. LOVASI : Capturing context : Integrating spatial and social network analyses. *Social Networks*, 34(1):1–5, janvier 2012. ISSN 03788733. URL <http://linkinghub.elsevier.com/retrieve/pii/S0378873311000608>.
- Theodor ADORNO et Max HORKHEIMER : The Culture Industry : Enlightenment as Mass Deception. *Culture*, 12(1944):1–24, 2005. URL <http://www.marxists.org/reference/archive/adorno/1944/culture-industry.htm>.
- Muhammad A AHMAD et Ankur TEREDESAI : Modeling Spread of Ideas in Online Social Networks. *Reproduction*, 61:185–190, 2006. URL <http://dl.acm.org/citation.cfm?id=1273833>.
- Aris ANAGNOSTOPOULOS, Luca BECCHETTI, Carlos CASTILLO, Aristides GIONIS et Stefano LEONARDI : Online team formation in social networks. *Proceedings of the 21st international conference on World Wide Web - WWW '12*, page 839, 2012. URL <http://dl.acm.org/citation.cfm?doid=2187836.2187950>.
- François ANSERMET et Pierre MAGISTRETTI : *À chacun son cerveau : Plasticité neuronale et inconscient*. Odile Jacob, 2004. ISBN 2738115322. URL <http://books.google.com/books?id=jY7HXE7inQIC&pgis=1>.
- Sinan ARAL, Lev MUCHNIK et Arun SUNDARARAJAN : Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51):21544–21549, 2009. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2799846&tool=pmcentrez&rendertype=abstract>.
- Hannah ARENDT : *Qu'est-ce que la politique ?* Seuil, Paris, points ess édition, 1995. ISBN 2020481901. URL <http://www.amazon.fr/Qu'est-ce-que-politique-Hannah-Arendt/dp/2020481901>.

- François ASCHER : *L'âge des métapoles*. Éd. de l'Aube, 2009. ISBN 2815900173. URL <http://www.amazon.com/Lâge-métapoles-François-Ascher/dp/2815900173>.
- J ATTALI : *Bruits*. PUF, Paris, {FR}, 1978.
- Marc AUGÉ : *Non-lieux*. 1995. ISBN 1859840515. URL [http://books.google.fr/books/about/Non\\_lieux.html?id=LMr8\\_pXJgdwC&pgis=1](http://books.google.fr/books/about/Non_lieux.html?id=LMr8_pXJgdwC&pgis=1).
- J L AUSTIN : *How to do things with words*, volume 3 de *Syntax and Semantics*. Harvard University Press, 1975. ISBN 019824553X.
- Kent BACH : Meaning , Speech Acts , and Communication. In Robert M. HARNISH, éditeur : *Basic Topics in the Philosophy of Language*, pages 1–23. Prentice-Harll, 1994.
- Alain BADIOU : *L'Etre et l'événement : Tome 2, Logiques des mondes*. Seuil, 2006. ISBN 2020843242. URL <http://www.amazon.fr/LEtre-lévénement-Tome-Logiques-mondes/dp/2020843242>.
- Eytan BAKSHY, J. HOFMAN, W. MASON et D. WATTS : Identifying ‘Influencers’ on Twitter. In *Fourth ACM International Conference on Web Search and Data Mining (WSDM)*, Hong Kong, 2011. URL <http://kdpaine.blogs.com/files/twitterinfluencershofmanetal.pdf>.
- David BAMMAN, Brendan O’CONNOR et Noah A SMITH : Censorship and deletion practices in Chinese social media. *First Monday*, 17(3-5):1–19, 2012. URL <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3943/3169><http://brenocon.com/censorship.bamman+oconnor+smith.fm2012.pdf>.
- K. BARBER : Orality, the media and new popular cultures in Africa. In Mena LAFKIOUI et Daniela MEROLLA, éditeurs : *Oralité et nouvelles dimensions de l’oralité Intersections théoriques et comparaisons des matériaux*, volume 00, pages 16–28, Paris, France, 2008. INALCO. ISBN 9782858311750.
- Pierre BARBOZA et Jean-Louis WEISSBERG : *L’image actée : Scénarisations numériques, parcours du séminaire - L’action sur l’image*. Editions L’Harmattan, 2006. ISBN 2296153666. URL <http://books.google.com/books?id=ax07uqf8oWIC&pgis=1>.
- Thierry BARDINI : *Junkware*. U of Minnesota Press, 2011. ISBN 0816667500. URL <http://books.google.com/books?id=7hZD11G700YC&pgis=1>.
- Jp BARLOW : *Déclaration d’indépendance du Cyberspace*. L’Eclat, 2001.
- F M BASS, T V KRISHNAN et D C JAIN : Why the Bass model fits without decision variables. *Marketing Science*, 13(3):203–223, 1994. ISSN 07322399. URL <http://www.jstor.org/stable/10.2307/183674>.

- Mathieu BASTIAN, Sebastien HEYMANN et Mathieu JACOMY : Gephi : An Open Source Software for Exploring and Manipulating Networks. *Artificial Intelligence*, 2(2):361–362, 2009. ISSN 14753898. URL <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/download/154/1009>.
- Michael BATTY : Virtual geography. *Futures*, 1997. URL <http://www.sciencedirect.com/science/article/pii/S0016328797000189>.
- Christian BAUCKHAGE : Insights into internet memes. In *Proc. ICWSM2011*, pages 42–49, 2011. URL <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/2757/3304>.
- H S BECKER : The epistemology of qualitative research. *Ethnography and human development : Context and meaning in social inquiry*, pages 53–71, 1996. ISSN 19398441.
- Mary BERGSTROM : *All Eyes East : Lessons from the Front Lines of Marketing to China's Youth*. Palgrave Macmillan, 2012. ISBN 0230120628. URL <http://www.amazon.com/All-Eyes-East-Lessons-Marketing/dp/0230120628>.
- Michael S BERNSTEIN, Drew HARRY, Paul ANDR, Katrina PANOVICH et Greg VARGAS : An Analysis of Anonymity and Ephemerality in a Large Online Community. (Coleman), 2011.
- Augustin BERQUE : Géogrammes, pour une ontologie des faits géographiques. *Espace géographique*, 28(4):320–326, 1999. ISSN 0046-2497. URL [http://www.persee.fr/web/revues/home/prescript/article/spgeo\\_0046-2497\\_1999\\_num\\_28\\_4\\_1275](http://www.persee.fr/web/revues/home/prescript/article/spgeo_0046-2497_1999_num_28_4_1275).
- Jacques BERTIN : *Sémiologie graphique : Les diagrammes - Les réseaux - Les cartes*. Editions de l'Ecole des Hautes Etudes en Sciences, 1977. ISBN 2713212774.
- Nick BINGHAM : Object-ions : from technological determinism towards geographies of relations, 1996. ISSN 0263-7758.
- Susan BLACKMORE : The power of the meme meme. *Skeptic*, (2):43, 1997. ISSN 10639330. URL <http://search.ebscohost.com/login.aspx?direct=true&db=afh&AN=9711050421&site=ehost-live>.
- Susan BLACKMORE : Evolution and Memes : the Human Brain As a Selective Imitation Device. *Cybernetics and Systems*, 32(1-2):225–255, 2001. ISSN 01969722. URL <http://www.informaworld.com/openurl?genre=article&doi=10.1080/019697201300001867&magic=crossref>.
- Susan BLACKMORE : Why we need memetics. *Behavioral and Brain Sciences*, 29(04):349–350, 2006. ISSN 0140525X. URL <http://www.lib.ncsu.edu/cgi-bin/proxy.pl?server=http://search.ebscohost.com/login.aspx?direct=true&db=ufh&AN=23334578&site=ehost-live&scope=site>.
- Olivier BLONDEAU : *Libres enfants du savoir numérique : Une anthologie du Libre*. L'éclat édition, 2000.

- Vincent D BLONDEL, Jean-Loup GUILLAUME, Renaud LAMBIOTTE et Etienne LEFEBVRE : Fast unfolding of communities in large networks. *Journal of Statistical Mechanics : Theory and Experiment*, 2008(10):P10008, octobre 2008. ISSN 1742-5468. URL <http://arxiv.org/abs/0803.0476>.
- Howard BLOOM : *Global Brain : The Evolution of Mass Mind from the Big Bang to the 21st Century*. Wiley, 2000. ISBN 0471295841. URL [http://books.google.com/books?id=V\\_IgAQAIAAJ&pgis=1](http://books.google.com/books?id=V_IgAQAIAAJ&pgis=1).
- Gillian BOLSOVER, William H. DUTTON, Ginette LAW et Soumitra DUTTA : Social Foundations of the Internet in China and the New Internet World : A Cross-National Comparative Perspective. *SSRN Electronic Journal*, juin 2013. ISSN 1556-5068. URL <http://papers.ssrn.com/abstract=2276482>.
- Philippe BONNIN : Dispositifs et rituels du seuil. *Communications*, 70(1):65–92, 2000. ISSN 0588-8018. URL [http://www.persee.fr/web/revues/home/prescript/article/comm\\_0588-8018\\_2000\\_num\\_70\\_1\\_2064](http://www.persee.fr/web/revues/home/prescript/article/comm_0588-8018_2000_num_70_1_2064).
- Michael BOSTOCK, Vadim OGIEVETSKY et Jeffrey HEER : D<sup>3</sup> : Data-Driven Documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–9, décembre 2011. ISSN 1941-0506. URL <http://www.ncbi.nlm.nih.gov/pubmed/22034350>.
- Daniel BOUGNOUX : *Sciences de l'information et de la communication*. Larousse, 1993. ISBN 978-2037410106.
- Danah BOYD et Kate CRAWFORD : Six Provocations for Big Data. In *SSRN Electronic Journal*, septembre 2011. URL <http://papers.ssrn.com/abstract=1926431>.
- Ulrik BRANDES, Garry ROBINS, Ann MCCRANIE et Stanley WASSERMAN : What is network science? *Network Science*, 1(01):1–15, avril 2013. ISSN 2050-1242. URL [http://www.journals.cambridge.org/abstract\\_S2050124213000027](http://www.journals.cambridge.org/abstract_S2050124213000027).
- Leo BREIMAN : Random Forests. *Machine Learning*, 45(1):5–32, 2001. ISSN 14787954. URL <http://www.springerlink.com/index/U0P06167N6173512.pdf>.
- Philippe BRETON : *L'Utopie de la Communication*. Poche, Paris, {FR}, essais la édition, 1997.
- Piotr BRÓDKA, Przemysław KAZIENKO et Multi-layered Social NETWORKS : Multi-layered Social Networks Key Points. In *Encyclopedia of Social Network Analysis*. Springer, 2013.
- Roger BRUNET : Organisation de l'espace et cartographie de modèles. *Espace géographique*, 1(1): 43–48, 1972. ISSN 0046-2497. URL [http://www.persee.fr/web/revues/home/prescript/article/spgeo\\_0046-2497\\_1972\\_num\\_1\\_1\\_1298](http://www.persee.fr/web/revues/home/prescript/article/spgeo_0046-2497_1972_num_1_1_1298).
- Roger BRUNET : La composition des modèles dans l'analyse spatiale. *Espace géographique*, pages 253–265, 1980. URL [http://www.persee.fr/web/revues/home/prescript/article/spgeo\\_0046-2497\\_1980\\_num\\_9\\_4\\_3572](http://www.persee.fr/web/revues/home/prescript/article/spgeo_0046-2497_1980_num_9_4_3572).



- Roger BRUNET : *La carte : mode d'emploi*. Fayard, 1987. ISBN 2213018480. URL <http://www.amazon.fr/La-carte-demploi-Roger-Brunet/dp/2213018480>.
- Branislav BUCHEL : *Internet Memes as Means of Communication*. Master thesis, MASARYK UNIVERSITY, 2012.
- Ron BURT, Damon CENTOLA et Steve KAHL : A Shortcut to Efficiency? Implications of the Small but Stratified World. *mittedu*, pages 1–52, 2008. URL [http://web.mit.edu/ewzucker/www/A\\_ShortCut\\_to\\_Efficiency.pdf](http://web.mit.edu/ewzucker/www/A_ShortCut_to_Efficiency.pdf).
- Judith BUTLER : *Bodies that matter : on the discursive limits of "sex"*. Numéro Book, Whole. Routledge, 1993. ISBN 0415903661. URL <http://www.amazon.com/dp/0415903661>.
- Alberto CAIRO : Infographics and Visualization and exploration. In *The Functional Art*, pages 15–25. New Riders, 2012. ISBN 978-0321834737. URL <http://www.thefunctionalart.com/>.
- Alberto CAIRO : Forms and Functions : Visualization as a Technology An Information. In *The Functional Art*, chapitre 2, pages 25–44. New Riders, 2013. ISBN 978-0321834737. URL <http://www.thefunctionalart.com/>.
- Georges CANGUILHEM : *La Connaissance De La Vie*. Vrin, 1965. ISBN 2711611329. URL [http://books.google.fr/books/about/La\\_Connaissance\\_De\\_La\\_Vie.html?id=A7EUiqvUHZsC&pgis=1](http://books.google.fr/books/about/La_Connaissance_De_La_Vie.html?id=A7EUiqvUHZsC&pgis=1).
- SK CARD et Jock MACKINLAY : The structure of the information visualization design space. In *Information Visualization, 1997*, 1997. URL [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=636792](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=636792).
- Nicholas CARR : How the internet makes us stupid. *Sydney Morning Herald 10 September 2010*, pages 13–15, 2010.
- AA CASILLI et Paola TUBARO : Why net censorship in times of political unrest results in more violent uprisings : A social simulation experiment on the UK riots. *SSRN eLibrary*, 2011. URL [http://www.sfu.ca/cmns/courses/2012/801/1-Readings/Examples-EmpricalResearchwithClearDescriptionsofMethodology\(onMediaUse\)/CasilliTubanoNetcensorshipandviolenceUKriots.pdf](http://www.sfu.ca/cmns/courses/2012/801/1-Readings/Examples-EmpricalResearchwithClearDescriptionsofMethodology(onMediaUse)/CasilliTubanoNetcensorshipandviolenceUKriots.pdf).
- Manuel CASTELLS : *The Informational City : Information Technology, Economic Restructuring, and the Urban Regional Process*. Oxford, {UK}, Cambridge, {MA}, 1989.
- Manuel CASTELLS : The Networked City : Réseaux, espace, société. *Textuel*, 2009.
- Meeyoung CHA, Hamed HADDADI, F. BENEVENUTO et K.P. GUMMADI : Measuring user influence in Twitter : The million follower fallacy. In *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 10–17, Washington, 2010. AAAI. URL <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/download/1538/1826>.

- Zhaoqun CHEN, Pengfei LIU, Xiaohan WANG et Yuantao GU : Follow Whom? Chinese Users Have Different Choice. page 9, décembre 2012. URL <http://arxiv.org/abs/1212.0167>.
- Jun-Jun CHENG, Yun LIU, Bo SHEN et Wei-Guo YUAN : An epidemic model of rumor diffusion in online social networks. *The European Physical Journal B*, 86(1):29, janvier 2013. ISSN 1434-6028. URL <http://link.springer.com/10.1140/epjb/e2012-30483-5>.
- Owen CHURCHES, Mike NICHOLLS, Myra THIESSEN, Mark KOHLER et Hannah KEAGE : Emoticons in mind : An event-related potential study Emoticons in mind : An event-related potential study. *Social Neuroscience*, 9(2):196–202, 2014. ISSN 1747-0919. URL <http://dx.doi.org/10.1080/17470919.2013.873737>.
- CIW : China Social Media Whitepaper 2012. Rapport technique, 2012. URL <http://www.chinainternetwatch.com/whitepaper>.
- Paul CLAVAL : L'évolution de quelques concepts de base de la géographie. In JF STASZAK, éditeur : *Les discours du géographe*, pages 89–118. L'Harmattan, Paris, France, histoire e édition, 1990.
- Jean CLÉMENT : Du texte à l'hypertexte : Vers une épistémologie de la discursivité hypertextuelle. *Hypertextes et hypermédias : Réalisations, Outils, Méthodes.*, 1995.
- Christopher CLIFTON, Deirdre MULLIGAN et Raghu RAMAKRISHNAN : Data Mining and Privacy : An Overview. *Privacy and Technologies of identity*, (2):191–208, 2006. URL [http://dx.doi.org/10.1007/0-387-28222-X\\_11](http://dx.doi.org/10.1007/0-387-28222-X_11).
- CNNIC : Statistical Report on Internet Development in China China Internet Network Information Center. Rapport technique January, Beijing, 2013. URL [http://www1.cnnic.cn/IDR/ReportDownloads/201209/t20120928\\_36586.htm](http://www1.cnnic.cn/IDR/ReportDownloads/201209/t20120928_36586.htm).
- Antoine COMPAGNON : Théorie du lieu commun. *Cahiers de l'Association internationale des études francaises*, 49(1):23–37, 1997. ISSN 0571-5865. URL [http://www.persee.fr/web/revues/home/prescript/article/caief\\_0571-5865\\_1997\\_num\\_49\\_1\\_1269](http://www.persee.fr/web/revues/home/prescript/article/caief_0571-5865_1997_num_49_1_1269).
- Auguste COMTE : *Cours de philosophie positive : La philosophie chimique et la philosophie biologique, Volume 3*. Bachelier, 1838. URL <http://books.google.com/books?id=X3YrAAAAYAAJ&pgis=1>.
- Michael D CONOVER, Clayton DAVIS, Emilio FERRARA, Karissa MCKELVEY, Filippo MENCZER et Alessandro FLAMMINI : The Geospatial Characteristics of a Social Movement Communication Network. *PLoS ONE*, 8(3):e55957+, 2013. URL <http://dx.doi.org/10.1371/journal.pone.0055957>.
- Michele COSCIA : Competition and Success in the Meme Pool : a Case Study on Quickmeme.com. avril 2013. URL <http://arxiv.org/abs/1304.1712>.

- J W CRAMPTON : Cartography : maps 2.0. *Progress in Human Geography*, 33(1):91–100, 2009. ISSN 03091325. URL <http://phg.sagepub.com/cgi/doi/10.1177/0309132508094074>.
- Jeremy W. CRAMPTON, Mark GRAHAM, Ate POORTHUIS, Taylor SHELTON, Monica STEPHENS, Matthew W. WILSON et Matthew ZOOK : Beyond the geotag : situating ‘big data’ and leveraging the potential of the geoweb. *Cartography and Geographic Information Science*, 40(2):130–139, mars 2013. ISSN 1523-0406. URL <http://dx.doi.org/10.1080/15230406.2013.777137>.
- Tim CRESSWELL : *Place : A Short Introduction*, volume 58 de *Short introductions to geography*. Wiley-Blackwell, 2004. ISBN 1405106727. URL <http://www.amazon.co.uk/dp/1405106727>.
- Olivier CREVOISIER : The Innovative Milieus Approach : Toward a Territorialized Understanding of the Economy? *Economic Geography*, 80(4):367–379, 2004. ISSN 00130095. URL <http://onlinelibrary.wiley.com/doi/10.1111/j.1944-8287.2004.tb00243.x/abstract>.
- Kenneth Neil CUKIER : Bandwidth Colonialism ? The Implications of Internet Infrastructure on International e-commerce. In *INET '99*, San Jose, CA, 1999.
- Candice Tran DAI : L’Internet en République populaire de Chine : vers un outil de démocratisation ou de renforcement de l’État-parti? Rapport technique, Fondation pour la Recherche Stratégique, 2007.
- Hubert DAMISCH : *L’origine de la perspective*. Flammarion, champs art édition, 1999. ISBN 2080816055. URL <http://www.amazon.fr/gp/product/2080816055?ie=UTF8&tag=leproidi-21&linkCode=as2&camp=1642&creative=6746&creativeASIN=2080816055>.
- Richard D’ANGIO : Enquête sur le milieu. *L’information géographique*, 65(4):372–382, 2001. ISSN 0020-0093.
- G Elijah DANN et Neil HADDOW : Just Doing Business or Doing Just Business : Google, Microsoft, Yahoo! and the Business of Censoring China’s Internet. *Journal of Business Ethics*, 79:219–234, 2008.
- Thomas H DAVENPORT et John C BECK : *The Attention Economy*, volume 2. Harvard Business School Press, 2001. ISBN 157851441X. URL <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/519/440>.
- Richard DAWKINS : *The selfish gene*. Oxford University Press, 1989. ISBN 0192860925. URL <http://books.google.com/books?id=WkH09HI7koEC&pgis=1>.
- Michel de CERTEAU : *L’invention du quotidien. Vol. 1, Arts de faire*, volume 1. 1980.
- Gabriele de SETA : “Meng? It Just Means Cute” : A Chinese Online Vernacular Term in Context, mars 2014. ISSN 14412616. URL <http://journal.media-culture.org.au/index.php/mcjournal/article/view/789>.

- M Bernard DEBARBIEUX : Non lieux. *Espace géographique*, 22(1):90–91, 1993.
- Régis DEBRAY : *Cours de médiologie générale*. Gallimard, Paris, France, 1991. ISBN 978-2070722921.
- Gilles DELEUZE : *Un Nouvel archiviste*. 1972. URL [http://books.google.fr/books/about/Un\\_Nouvel\\_archiviste.html?id=tRMOAAAAAAAJ&pgis=1](http://books.google.fr/books/about/Un_Nouvel_archiviste.html?id=tRMOAAAAAAAJ&pgis=1).
- Gilles DELEUZE et Felix GUATTARI : Rhizome, introduction. 1972.
- Jean DEMANGEOT : *Les milieux "naturels" du globe*. Masson, 1984. ISBN 220026450X. URL <http://www.amazon.fr/Les-milieux-naturels-du-globe/dp/220026450X>.
- Jérôme DENIS : Préface : Les nouveaux visages de la performativité. *Études de communication*, (29):8–24, décembre 2006. ISSN 1270-6841. URL <http://edc.revues.org/344>.
- Daniel C DENNETT : *Breaking the Spell : Religion As a Natural Phenomenon*. Viking, 2006. ISBN 067003472X. URL <http://books.google.com/books?id=yWtwDDqR61QC&pgis=1>.
- Daantje DERKS, Arjan E R BOS et Jasper Von GRUMBKOW : Emoticons and social interaction on the Internet : the importance of social context. *Computers in Human Behavior*, 23(1): 842–849, 2007. ISSN 07475632. URL <http://linkinghub.elsevier.com/retrieve/pii/S0747563204002079>.
- Georges DIDI-HUBERMAN : *Invention de l'hystérie : Charcot et l'iconographie photographique de la Salpêtrière*. Macula, Paris, France, 2012. ISBN 286589004X. URL <http://www.amazon.fr/Invention-lhystérie-liconographie-photographique-Salpêtrière/dp/286589004X>.
- Cong DING, Yang CHEN et Xiaoming FU : Crowd Crawling : Towards Collaborative Data Collection for Large-scale Online Social Networks. pages 183–188, 2013.
- Martin DODGE et Rob KITCHIN : Code , space and everyday life. 2004.
- Martin DODGE et Rob KITCHIN : Virtual places, 2007.
- Jean-François DOULET : Les nouveaux enjeux de la mobilité urbaine dans les villes chinoises - Rapport d ' une mission en Chine. Rapport technique, Institut pour la ville en mouvement (PSA Peugeot-Citroën), Paris, France, 2001.
- F DOUZET : Les frontières chinoises de l'Internet. *Hérodote*, 125:127–142, 2007.
- Fernand DUMONT : Le sort de la culture. 1993.
- Gabriel DUPUY : Internet : une approche géographique à l'échelle mondiale. *Flux*, pages 5–19, 2004. URL <http://www.cairn.info/revue-flux-2004-4-page-5.htm>.
- Umberto ECO : *Lector in fabula ou La Coopération interprétative dans les textes narratifs*. Grasset, 1985. ISBN 2246342619. URL <http://www.amazon.fr/Lector-fabula-Coopération-interprétative-narratifs/dp/2246342619>.

- Bruce EDMONDS : Three Challenges for the Survival of Memetics. *Journal of Memetics Evolutionary Models of Information Transmission*, (6), 2002.
- Jacob EISENSTEIN, Brendan O'CONNOR, Noah A. SMITH et Eric P. XING : Mapping the geographical diffusion of new words. In *NIPS 2012 Workshop on Social Network and Social Media Analysis*, octobre 2012. URL <http://arxiv.org/abs/1210.5268>.
- Norbert ELIAS : *La Dynamique de l'Occident*. Poche Pocket, Paris, {FR}, 1975.
- Sarah ELWOOD : Volunteered geographic information : future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal*, 72(3-4):173–183, 2008. ISSN 03432521. URL <http://www.springerlink.com/index/10.1007/s10708-008-9186-0>.
- Y ENGELHARDT : Syntactic structures in graphics. *Computational Visualistics and Picture ...*, pages 23–36, 2007. URL [http://www.gib.uni-tuebingen.de/own/journal/pdf/buch\\_image5b.pdf#page=25](http://www.gib.uni-tuebingen.de/own/journal/pdf/buch_image5b.pdf#page=25).
- J M EPSTEIN et R AXTELL : *Growing artificial societies*. 1996. ISBN 0262550253.
- Xiang FANG et David C YEN : Demographics and behavior of Internet users in China. *Technology in Society*, 28:363–387, 2006.
- Adrienne FELT et David EVANS : Privacy Protection for Social Networking Platforms. *Social Networks*, 2(May):1–8, 2008. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.143.5642&rep=rep1&type=pdf>.
- Valérie FERNANDEZ et Gilles PUEL : *La gouvernance territoriale des TIC*. Hermes Science/Lavoisier, Paris, mondialisa édition, 2010. URL <http://editions.lavoisier.fr/notice.asp?ouvrage=2139905>.
- Emilio FERRARA, Mohsen JAFARIASBAGH, Omur VAROL, Vahed QAZVINIAN, Filippo MENCZER et Alessandro FLAMMINI : Clustering Memes in Social Media. octobre 2013. URL <http://arxiv.org/abs/1310.2665>.
- Wolf FEUERHAHN : Du milieu à l'Umwelt : enjeux d'un changement terminologique. *Revue philosophique de la France et de l'étranger*, 134(4):419, 2009. ISSN 0035-3833. URL <http://www.cairn.info/revue-philosophique-2009-4-page-419.htm>.
- Gustave FLAUBERT : *Le Dictionnaire des idées reçues*. Le Livre de Poche, 1997. ISBN 2253098361. URL <http://www.amazon.fr/Dictionnaire-idees-receues-Gustave-Flaubert/dp/2253098361>.
- Nick V. FLOR : Memetic Marketing, 2000. URL <http://www.informit.com/articles/article.aspx?p=19996>.
- William FOSTER et Seymour E GOODMAN : *The Diffusion of the Internet in China*. University of Arizona, 2000.

- Michel FOUCAULT : *Les Mots et les Choses : Une Archéologie du Savoir*. Gallimard, Paris, {FR}, 1996.
- Michel FOUCAULT : *Sécurité, territoire, population. Cours au Collège de France (1977-78)*. Gallimard/Seuil (Collection « Hautes Études »), Paris, 2004.
- King-wa FU, Chung-hong CHAN et Michael CHAU : Assessing Censorship on Microblogs in China : Discriminatory Keyword Analysis and the Real-Name Registration Policy. *IEEE Internet Computing*, 17(3):42–50, mai 2013. ISSN 1089-7801. URL <http://www.computer.org/csdl/mags/ic/2013/03/mic2013030042.html>.
- L GABORA : Meme and Variations : A Computational Model of Cultural Evolution. In L NADEL et D L STEIN, éditeurs : *D Stein Ed*, numéro Figure 1, pages 471–485. Addison Wesley, 1995. URL <http://cogprints.org/531/1/mav.htm>.
- Silvia Elena GALLAGHER et Timothy SAVAGE : Cross-cultural analysis in online community research : A literature review. *Computers in Human Behavior*, 29(3):1028–1038, mai 2013. ISSN 07475632. URL <http://linkinghub.elsevier.com/retrieve/pii/S0747563212002683>.
- Gabriel GANDOLFO : Le concept de milieu dans les sciences du vivant. *Noesis*, 14(2):237–247, 2008.
- Qi GAO, Fabian ABEL, GJ HOUBEN et Yong YU : A Comparative Study of Users' Microblogging Behavior on Sina Weibo and Twitter. *User Modeling, Adaptation, and ...*, 2012. URL [http://link.springer.com/chapter/10.1007/978-3-642-31454-4\\_8](http://link.springer.com/chapter/10.1007/978-3-642-31454-4_8).
- William GIBSON : *Le Neuromancien*. J'ai Lu / live de Poche, 1984.
- Minas GJOKA et Maciej KURANT : Practical recommendations on crawling on-line social networks. *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS on Measurement of Internet Topologies*, 29(9):1872–1892, 2011. URL [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6027868](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6027868)[http://www.minasgjoka.com/papers/jsac11\\_crawling\\_osns.pdf](http://www.minasgjoka.com/papers/jsac11_crawling_osns.pdf).
- Sharad GOEL, DJ WATTS et DG GOLDSTEIN : The structure of online diffusion networks. In *EC '12 Proceedings of the 13th ACM Conference on Electronic Commerce*, volume 1, 2012. ISBN 9781450314152. URL <http://dl.acm.org/citation.cfm?id=2229058>.
- Michael GOLDHABER : The value of openness in an attention economy. *First Monday*, 11(6):3, 2006. URL <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1334/1254>.
- Michele GRAFFIETI, Gaia SCAGNETTI, Donato RICCI, Luca MASUD et Mario PORPORA : Tell them anything but the truth : They will find their own. How we visualized the map of the future with respect to the audience of our story. *Leonardo*, 44(3):250–251, 2010.

- Mark GRAHAM et Matthew ZOOK : Visualizing Global Cyberscapes : Mapping User-Generated Placemarks. *Journal of Urban Technology*, 18(1):115–132, janvier 2011. ISSN 1063-0732. URL <http://www.tandfonline.com/doi/abs/10.1080/10630732.2011.578412>.
- S GRAHAM : The end of geography or the explosion of place? Conceptualizing space, place and information technology. *Progress in Human Geography*, 22(2):165–185, 1998. ISSN 03091325. URL <http://phg.sagepub.com/cgi/doi/10.1191/030913298671334137>.
- Mark S GRANOVETTER : The Strength of Weak Ties. *American Journal of Sociology*, 78(6):1360–1380, 1973. ISSN 00029602. URL <http://www.jstor.org/stable/2776392>.
- Glenn GREENWALD, Ewen MACASKILL et The GUARDIAN : NSA Prism program taps in to user data of Apple , Google and others, 2013.
- Éric GUICHARD : L'internet et les épistémologies des sciences humaines et sociales. *Revue Sciences/Lettres*, (2), octobre 2014. ISSN 2271-6246. URL <http://rsl.revues.org/389>.
- Adrien GUILLE, Cécile FAVRE, Hakim HACID et D.A. ZIGHED : SONDY : An Open Source Platform for Social Dynamics Mining and Analysis. In *2013 ACM SIGMOD International Conference on Management of Data (SIGMOD 2013)*, pages 6–9, 2013. ISBN 9781450320375.
- Muki HAKLAY, Alex SINGLETON et Chris PARKER : Web Mapping 2.0 : The Neogeography of the GeoWeb. *Geography Compass*, 2(6):2011–2039, novembre 2008. ISSN 17498198. URL <http://doi.wiley.com/10.1111/j.1749-8198.2008.00167.x>.
- Maurice HALBWACHS : *La mémoire collective et le temps*, volume 1945. UQAC, Les classiques des sciences sociales, 1947. URL [http://www.psychanalyse.com/pdf/memoire\\_collective.pdf](http://www.psychanalyse.com/pdf/memoire_collective.pdf).
- Stuart HALL : Cultural identity and diaspora. *Theorising Diaspora : A Reader*, 2:222–237, 2003. ISSN 1475925X. URL [http://www.unipa.it/~michele.cometa/hall\\_cultural\\_identity.pdf](http://www.unipa.it/~michele.cometa/hall_cultural_identity.pdf).
- Harry HALPIN et Alexandre MONNIN, éditeurs. *Philosophical Engineering : Toward a Philosophy of the Web (Metaphilosophy)*. Wiley-Blackwell, 2014. ISBN 111870018X. URL <http://www.amazon.com/Philosophical-Engineering-Toward-Philosophy-Metaphilosophy/dp/111870018X>.
- Frank HARARY : Graph theory 1736–1936, 1977. ISSN 03150860.
- June HINCKLEY : MTV Music Matters. Rapport technique 5, MTV / Synovate, mars 2009. URL <http://mej.sagepub.com/cgi/doi/10.2307/3399627>.
- Douglas R HOFSTADTER : *Gödel, Escher, Bach*. Klett-Cotta, 1999. ISBN 360893037x.
- Liangjie HONG et Brian D. DAVISON : Empirical study of topic modeling in Twitter. *Proceedings of the First Workshop on Social Media Analytics - SOMA '10*, pages 80–88, 2010. URL <http://portal.acm.org/citation.cfm?doid=1964858.1964870>.

- Jeremy R L HOWELLS : Tacit Knowledge , Innovation and Economic Geography. *Urban Studies*, 39(5-6):871–884, 2002. ISSN 00420980. URL <https://webvpn.ethz.ch/http/0/www.informaworld.com/smpp/1814185133-11623592/content~content=a770985056>.
- Christopher R HUGHES : Nationalism in Chinese cyberspace. *Cambridge review of international*, 13(2):195–209, 2000.
- Yuk HUI : What is a digital object ? *Metaphilosophy*, 43(4):380–395, 2012.
- Yuk HUI : Pour un nouvel archiviste (manifeste). 2014.
- J HWANG : Coffee Goes to China : An Examination of Starbucks' Market Entry Strategy. *Journal of Undergraduate Research*, 6(8), 2005.
- Dino IENCO, Francesco BONCHI et Carlos CASTILLO : The Meme Ranking Problem : Maximizing Microblogging Virality. In *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, pages 328–355, Sydney, NSW, 2010.
- Sorin ISTRAIL et Solomon MARCUS : Alan Turing and John von Neumann - Their Brains and Their Computers. In Erzsébet CSUHAJ-VARJÚ, Marian GHEORGHE, Grzegorz ROZENBERG, Arto SALOMAA et György VASZIL, éditeurs : *13th International Conference on Membrane Computing CMC 2012*, volume 7762 de *Lecture Notes in Computer Science*, pages 26–35, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. ISBN 978-3-642-36750-2. URL <http://link.springer.com/10.1007/978-3-642-36751-9>.
- MO JACKSON et Leeat YARIV : Diffusion, strategic interaction, and social structure. in *Handbook of Social Economics*, pages 1–50, 2010. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.207.4193&rep=rep1&type=pdf>.
- Christian JACOB : *L'Empire des cartes : Approche théorique de la cartographie à travers l'histoire*. Albin Michel, 1992. ISBN 2226060839.
- Christian JACOB : *Lieux de savoir : Tome 2, Les mains de l'intellect*. Editions Albin Michel, 2011. ISBN 978-2226187291. URL <http://www.amazon.fr/Lieux-savoir-Tome-mains-lintellect/dp/2226187294><http://lieuxdesavoir.hypotheses.org/lieux-de-savoir-2-les-mains-de-lintellect-2>.
- Jane JACOBS : *The Death and Life of Great American Cities*, volume 71 de *Vintage Series*. Random House, 1961. ISBN 067974195X. URL <http://www.amazon.com/dp/B001PGMCXG>.
- Roman JAKOBSON et Morris HALLE : Phonemic patterning. *Fundamentals of language*, pages 37–44, 1956.
- William JAMES : *Essays in radical Empiricism*. 1912.
- Laurent JENNY : *Le Style en acte*. METISPRESES, voltiges édition, 2012. ISBN 2940406375. URL [http://www.amazon.fr/Style-en-acte-Collectif/dp/2940406375/ref=la\\_B004N7XYE0\\_1\\_4?s=books&ie=UTF8&qid=1401958702&sr=1-4](http://www.amazon.fr/Style-en-acte-Collectif/dp/2940406375/ref=la_B004N7XYE0_1_4?s=books&ie=UTF8&qid=1401958702&sr=1-4).



- D. Lamont JOHNSON et Leping LIU : China and the Internet :, 1996. ISSN 0738-0569.
- Pascal JOUXTEL : La mémétique, une science à l'état sauvage. novembre 2014. ISSN 0767-9513.  
URL <http://documents.irevues.inist.fr/handle/2042/51885>.
- KY KAMATH, James CAVERLEE, Kyumin LEE et Zhiyuan CHENG : Spatio-Temporal Dynamics of Online Memes : A Study of Geo-Tagged Tweets. In *WWW 2013*, Rio de Janeiro, Brazil, 2013. IW3C2. URL <http://students.cse.tamu.edu/kykamath/papers/www2013/fp113-kamath.pdf>.
- Anand KARANDIKAR : Clustering short status messages : A topic model based approach. *Work*, page 55, 2010. URL <http://ebiquity.umbc.edu/get/a/publication/518.pdf>.
- E KATZ et P F LAZARFELD : *Personal influence : The part played by people in the flow of mass communications*, volume 21 de *Foundations of communications research*. Free Press, 1955. ISBN 9781412805070. URL <http://www.ncbi.nlm.nih.gov/pubmed/13409025>.
- Christa KELLEHER et Thorsten WAGENER : Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling Software*, 26(6):822–827, 2011. ISSN 13648152. URL <http://linkinghub.elsevier.com/retrieve/pii/S1364815210003270>.
- Rob. KITCHIN et Martin DODGE : *Code, Space : Software and Everyday Life*. MIT Press, 2011. ISBN 0262042487. URL <http://books.google.com/books?hl=en&lr=&id=ZHez2BXgIeQC&pgis=1http://mitpress.mit.edu/books/codespace>.
- Michele KNOBEL et Colin LANKSHEAR : *A new literacies sampler*. Peter Lang Publishing, Inc., New York City, USA, 2007. ISBN 9780820495231. URL <http://books.google.com/books?hl=en&lr=&id=4Gjs8uT6dxIC&oi=fnd&pg=PR9&dq=A+New+Literacies+Sampler&ots=F14cYpvwjs&sig=4gyFVE88QJSJvbC-jSnIUESPf-V8>.
- Moo Nam KO, Gorrell P CHEEK, Mohamed SHEHAB et North CAROLINA : CONNECT SERVICES. *Computer*, 43(August):37–43, 2010. ISSN 00189162. URL [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5551044](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5551044).
- Ravi KUMAR, Jasmine NOVAK et Andrew TOMKINS : Structure and evolution of online social networks. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining KDD 06*, 106(2):611, 2006. URL <http://portal.acm.org/citation.cfm?doid=1150402.1150476>.
- Haewoon KWAK, Changhyun LEE, Hosung PARK et Sue MOON : What is Twitter , a Social Network or a News Media? Categories and Subject Descriptors. In *WWW 2010 : Proceedings of the 19th international conference on World Wide Web*, Raleigh, North Carolina, USA., 2010. ISBN 9781605587998.
- Gerard KYLE et Garry CHICK : The Social Construction of a Sense of Place. *Leisure Sciences*, 29(3):209–225, 2007. ISSN 01490400. URL <http://www.tandfonline.com/doi/abs/10.1080/01490400701257922>.

- Jacques LACAN : *Autres écrits*. Seuil, Paris, le champ f édition, 2001.
- Bruno LATOUR : On actor-network theory. *Soziale Welt*, 47(4):369–381, 1996. URL <http://www.cours.fse.ulaval.ca/edc-65804/latour-clarifications.pdf>.
- Bruno LATOUR : On recalling ANT. *Network*, 46(3):15–25, 1999. ISSN 1467954X. URL <http://books.google.com/books?hl=en&lr=&id=0sNv1Mq38-IC&oi=fnd&pg=PA15&dq=On+Recalling+ANT&ots=1ZhmFhhDco&sig=KnP0wF18jQvRu3dU09eDq-u0ZQ8>.
- Bruno LATOUR : Tarde’s idea of quantification. In *The Social After Gabriel Tarde : Debates and Assessments*, pages 145–162. ed M. Candea, Routledge, London, UK, 2003.
- John LAURENT : A Note on the Origin of Memes & Mnemes. *Journal of Memetics Evolutionary Models of Information Transmission*, 3(1), 1999.
- Gustave LE BON : *Psychologie des foules*. Presses Universitaires de France - PUF, 1895. ISBN 2130542972. URL <http://www.amazon.fr/Psychologie-foules-Gustave-Le-Bon/dp/2130542972>.
- Alex LEAVITT, Evan BURCHARD, David FISHER et Sam GILBERT : The Influentials : New Approaches for Analyzing Influence on Twitter, 2009. URL <http://www.webecologyproject.org/2009/09/analyzing-influence-on-twitter/>.
- Kalev LEETARU, Shaowen WANG, Guofeng CAO, Anand PADMANABHAN et Eric SHOOK : Mapping the global Twitter heartbeat : The geography of Twitter, avril 2013. ISSN 13960466. URL <http://firstmonday.org/ojs/index.php/fm/article/view/4366/3654>.
- André LEROI-GOURHAN : *Évolution et techniques : L’homme et la matière*. Albin Michel, 1971. ISBN 978-2226062130.
- Jure LESKOVEC, Lada A ADAMIC et Bernardo A HUBERMAN : The Dynamics of Viral Marketing. *ACM Transactions on the Web*, 1(1):5–es, 2005. URL <http://arxiv.org/abs/physics/0509039>.
- Jure LESKOVEC, Lars BACKSTROM et Jon KLEINBERG : Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’09*, page 497, New York, New York, USA, juin 2009. ACM Press. ISBN 9781605584959. URL <http://dl.acm.org/citation.cfm?id=1557019.1557077>.
- Lawrence LESSIG : *Code : Version 2.0*. Numéro September. Basic Books, 2006. ISBN 9780465039142. URL <http://pdf.codev2.cc/Lessig-Codev2.pdf>.
- Yuan LI, Haoyu GAO, Mingmin YANG, Wanqiu GUAN et H MA : What are Chinese Talking about in Hot Weibos. *CoRR*, 2(abs), 2013. URL <http://arxiv.org/abs/1304.4682>.

- Tamar LIEBES et Elihu KATZ : The Export of Meaning : Cross-Cultural Readings of Dallas. 1990.
- Nan LIN : Building a Network Theory of Social Capital. *Connections*, 22(1):28–51, 1999. ISSN 14691930. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.3792&rep=rep1&type=pdf>.
- Sharon LIN, Julie FORTUNA, Chinmay KULKARNI, Maureen STONE et Jeffrey HEER : Selecting Semantically-Resonant Colors for Data Visualization. *graphics.stanford.edu*, 32(3), 2013. URL <http://graphics.stanford.edu/~sharonl/papers/semanticColors.pdf>.
- Michael R LISSACK : The Redefinition of Memes. *Journal of Memetics Evolutionary Models of Information Transmission*, (1):1–14, 2004. ISSN 13664786. URL <http://search.ebscohost.com/login.aspx?direct=true&db=afh&AN=15171172&site=ehost-live>.
- S LOHR : The age of big data, février 2012a. URL [http://rkbusinessschool.com/uploads/2\\_The\\_New\\_York\\_Times\\_on\\_The\\_Age\\_of\\_Big\\_Data.pdf](http://rkbusinessschool.com/uploads/2_The_New_York_Times_on_The_Age_of_Big_Data.pdf).
- Steve LOHR : Big Data’s Impact in the World, février 2012b. URL [http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?\\_r=1&pagewanted=1&ref=technology](http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?_r=1&pagewanted=1&ref=technology).
- Gilad LOTAN, Erhardt GRAEFF, Mike ANANNY, Devin GAFFNEY, Ian PEARCE et Danah BOYD : The Arab Spring| The Revolutions Were Tweeted : Information Flows during the 2011 Tunisian and Egyptian Revolutions. *International Journal of Communication*, 5, 2011. URL <http://ijoc.org/ojs/index.php/ijoc/article/view/1246>.
- CC LUNDBERG : Patterns of acquaintanceship in society and complex organization : A comparative study of the small world problem. *Pacific Sociological Review*, 18:206 – 222, 1975.
- Dave LYONS : China’s Golden Shield Project. In *CIRC09*, 2009. URL <http://www.scribd.com/doc/15919071/Dave-Lyons-Chinas-Golden-Shield-Project>.
- Donald MACKENZIE : Is Economics Performative? Option Theory and the Construction of Derivatives Markets, 2006. ISSN 1053-8372.
- Rebecca MACKINNON : China’s Censorship 2.0 : How companies censor bloggers. *First Monday*, 14(2-2), 2009. URL <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewarticle/2378/2089>.
- Rebecca MACKINNON : The Not-So-Great Firewall of China, 2012. URL [http://www.foreignpolicy.com/articles/2012/04/17/the\\_not\\_so\\_great\\_firewall\\_of\\_china](http://www.foreignpolicy.com/articles/2012/04/17/the_not_so_great_firewall_of_china).
- Michael W MACY et Robert WILLER : From Factors to Actors : Computational Sociology and Agent-Based Modeling. *Annual Review of Sociology*, 28(1):143–166, 2002. ISSN 03600572. URL <http://www.annualreviews.org/doi/abs/10.1146/annurev.soc.28.110601.141117>.

- Pierre MAGISTRETTI : Neurosciences et psychanalyse : une rencontre autour de l'émergence de la singularité. 2008. URL <http://www.cerimes.fr/le-catalogue/colloque-neurosciences-et-psychanalyse.html>.
- Lev MANOVICH : *Language of new media*. {MIT} pres édition, 2001.
- Lev MANOVICH : Trending : The Promises and the Challenges of Big Social Data. *In Debates in the Digital Humanities*, pages 1–17. ed M.K.Gold. The University of Minnesota Press, Minneapolis, MN, 2011.
- Lev MANOVICH : Media After Software. *Journal of Visual Culture*, 12(1):30–37, 2013.
- Étienne Jules MAREY : *La méthode graphique dans les sciences expérimentales : et principalement en physiologie et en médecine*. G. Masson, 1885. URL <http://books.google.com/books?id=JnsAAAAAMAAJ&pgis=1>.
- Frédéric MARTEL : *Mainstream*. Flammarion, Paris, 2010. ISBN 2081236176. URL <http://www.amazon.fr/Mainstream-Frédéric-Martel/dp/2081236176>.
- P. R. MASANI : The Cybernetical Movement and von Neumann's Letter, 1946. *In Norbert Wiener 1894–1964*, pages 239–250. Birkhäuser Basel, Basel, 1990. ISBN 978-3-0348-9963-5. URL <http://www.springerlink.com/index/10.1007/978-3-0348-9252-0>.
- Mark MASSÉ : *REST Api Design Rulebook*. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, 2012. ISBN 9781449310509.
- Doreen MASSEY : Power-geometry and a progressive sense of place. *In J BIRD, B CURTIS, T PUTNAM, G ROBERTSON et L TICKNER, éditeurs : Mapping the futures Local cultures global change*, numéro 1984, chapitre 4, pages 59–69. Routledge, 1993. ISBN 041507018X. URL [http://books.google.com/books?id=307\\_ioVON0gC](http://books.google.com/books?id=307_ioVON0gC).
- Armand MATTELART : *Penser les médias*. La Découverte, Paris, France, 1986. ISBN 978-2707116192.
- Hugh MATTHEWS, Marilyn TAYLOR, Barry PERCY-SMITH et Melanie LIMB : The unacceptable flaneur : The shopping mall as a teenage hangout. *Childhood*, 7(3):279–294, 2000. ISSN 09075682. URL <http://chd.sagepub.com/cgi/doi/10.1177/0907568200007003003>.
- Jon MAY et N J THRIFT : *TimeSpace : geographies of temporality*. Routledge, 2001. ISBN 041518083X\0415180848 (pbk.).
- Rachel MAZUY : *Croire plutôt que voir ? : Voyage en Russie soviétique*. Odile Jacob, 2002. ISBN 273811153X.
- Richard MCCREADIE, Ian SOBOROFF, Jimmy LIN, Craig MACDONALD, Iadh OUNIS et Dean MCCULLOUGH : On building a reusable Twitter corpus. *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '12*, page 1113, 2012. URL <http://dl.acm.org/citation.cfm?doid=2348283.2348495>.

- McKINSEY : The social economy : Unlocking value and productivity through social technologies. Rapport technique July, McKinsey, 2012.
- Marshall McLuhan : *The Gutenberg Galaxy*. University édition, 1962.
- Claire Cain MILLER : How Obama's Internet Campaign Changed Politics, 2008. URL <http://bits.blogs.nytimes.com/2008/11/07/how-obamas-internet-campaign-changed-politics/>.
- An Xiao MINA : A Curated History of the Grass Mud Horse Song, 2012. URL <http://www.88-bar.com/2012/02/a-curated-history-of-the-grass-mud-horse-song/>.
- Jacob Levi MORENO et Helen JENNINGS : Statistics of Social Configurations. *Sociometry*, 1(3/4): 342–374, 1938. ISSN 00380431. URL <http://www.jstor.org/stable/10.2307/2785588>.
- Edgar MORIN : *Introduction à la pensée complexe*. Le seuil édition, 1990.
- David F. NETTLETON : Data mining of social networks represented as graphs. *Computer Science Review*, 7:1–34, février 2013. ISSN 15740137. URL <http://linkinghub.elsevier.com/retrieve/pii/S1574013712000445>.
- Jason Q. NG : *Blocked on Weibo : What Gets Suppressed on China's Version of Twitter (And Why)*. New Press, The, 2013a. ISBN 159558871X.
- Jason Q NG : *Chinese Social Media as Laboratory : What we can learn about China from research into Sina Weibo*. Thèse de doctorat, University of Pittsburgh, 2013b.
- BD NYE : *Modeling memes : A memetic view of affordance learning*. Thèse de doctorat, 2011. URL <http://repository.upenn.edu/edissertations/336/>.
- Jean-Marc OFFNER : Les « effets structurants » du transport : mythe politique, mystification scientifique. *Espace géographique*, 22(3):233–242, 1993. ISSN 0046-2497. URL [http://www.persee.fr/web/revues/home/prescript/article/spgeo\\_0046-2497\\_1993\\_num\\_22\\_3\\_3209](http://www.persee.fr/web/revues/home/prescript/article/spgeo_0046-2497_1993_num_22_3_3209).
- Ray OLDENBURG : *The Great Good Place*. Marlowe & Company, 1999. ISBN 1569246815. URL <http://books.google.com/books?id=0a0jHGdSKLMC>.
- Pamela E OLIVER et Gerald MARWELL : Whatever Happened to Critical Mass Theory? A Retrospective and Assessment. *Sociological Theory*, 19(3):292–311, 2001. ISSN 07352751. URL <http://doi.wiley.com/10.1111/0735-2751.00142>.
- Walter J ONG : *Orality and Literacy : The Technologizing of the Word*. 1988.
- Jukka-Pekka ONNELA, Samuel ARBESMAN, Marta C GONZÁLEZ, Albert-László BARABÁSI et Nicholas A CHRISTAKIS : Geographic Constraints on Social Network Groups. *PLoS ONE*, 6(4):7, 2011. URL <http://arxiv.org/abs/1011.4859>.

- OPENITP : A Snapshot of Chinese Internet Users Circumventing Censorship. Rapport technique April, Open Internet Tools Project, 2013.
- Wanda ORLIKOWSKI : Learning from notes : Organizational issues in groupware implementation, 1993. ISSN 0197-2243.
- Bo PANG et Lillian LEE : Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2:1–135, 2008. ISSN 1554-0669.
- Alain PAPAUX : De la nature au « milieu » : l’homme plongé dans l’environnement. *Revue interdisciplinaire d’études juridiques*, 60(1):29–57, 2008.
- Jean-Pierre PETIT : *Les Aventures d’Anselme Lanturlu : Le Topologicon*. Belin, 1999. ISBN 978-2701104669.
- Stephen PETSCHULAT : Other people’s data. *Communications of the ACM*, 53(1):53, 2010. ISSN 00010782. URL <http://portal.acm.org/citation.cfm?doid=1629175.1629196>.
- Antoine PICON : *La ville des réseaux - Un imaginaire politique*. Manucius, modélisati édition, 2014. ISBN 2845784147.
- Christian PLANTIN : *Lieux communs, topoi, stéréotypes, clichés*. Editions Kimé, 1993. ISBN 2908212765. URL [http://books.google.fr/books/about/Lieux\\_communs\\_topoi\\_stéréotypes\\_clich.html?hl=fr&id=718UAAAAIAAJ&pgis=1](http://books.google.fr/books/about/Lieux_communs_topoi_stéréotypes_clich.html?hl=fr&id=718UAAAAIAAJ&pgis=1).
- Ate POORTHUIS et Matthew ZOOK : Spaces of Volunteered Geographic Information. *SSRN Electronic Journal*, mai 2013. ISSN 1556-5068. URL <http://papers.ssrn.com/abstract=2259845>.
- Serge PROULX et Philippe BRETON : *L’explosion de la communication à l’aube du XXIème siècle*. Boréal édition, 2000.
- G PUEL : Géographie des lieux d’accès à Internet. Les conditions de l’accès public et les modèles d’usages dans les grandes villes de Chine. *Espace géographique*, 38:17–29, 2009.
- Gilles PUEL, Blandine PONS et Jin XIAOTING : New Forms of Mobility and Social Practices : The Starbucks Cafés in Beijing. *China Perspectives*, (April), 2007. URL <http://hal.archives-ouvertes.fr/hal-00125873/>.
- Jack Linchuan QIU : The Internet in China : technologies of freedom in a statist society. In Manuel CASTELLS, éditeur : *The Network Society : A Cross-Cultural Perspective*, pages 99–125. Edward Elgar Pub, 2005. ISBN 1845424352.
- Xipeng QIU, Qi ZHANG et Xuanjing HUANG : FudanNLP : A toolkit for Chinese natural language processing. *Proceedings of ACL*, pages 49–54, 2013. URL <https://www.aclweb.org/anthology/P/P13/P13-4009.pdf>.

- Anand RAJARAMAN et Jeffrey David ULLMAN : *Mining of Massive Datasets*. Cambridge University Press, Cambridge, 2011. ISBN 9781139058452. URL <http://ebooks.cambridge.org/ref/id/CB09781139058452>.
- Ignacio RAMONET : *La tyrannie de la communication*. Folio essa édition, 2001.
- Dominique RAYNAUD : L ' émergence de l ' espace perspectif : Effets de croyance et de connaissance. In Odile JACOB, éditeur : *Les Espaces de l'Homme, Symposium annuel du Collège de France*, 2005.
- Clément RENAUD : Facebook face aux réseaux sociaux chinois, 2011. URL [http://www.lemonde.fr/technologies/article/2011/04/27/facebook-face-aux-reseaux-sociaux-chinois\\_1513168\\_651865.html](http://www.lemonde.fr/technologies/article/2011/04/27/facebook-face-aux-reseaux-sociaux-chinois_1513168_651865.html).
- Clément RENAUD : Meme observation and classification on a large corpus of tweets from Sina Weibo Memes on Sina Weibo. In *CIRC14*, pages 1–9, HK, 2014.
- Jean-Pierre RENAUD et Jeanne LAFONT : *La surface de boy et l'aventure graphique de la topologie*. Efedition édition, 2010. ISBN 9782913786189.
- Bernhard RIEDER : Networked control : Search engines and the symmetry of confidence. *International Review of Information Ethics*, 3:26–32, 2005. URL [http://www.i-r-i-e.net/inhalt/003/003\\_full.pdf#page=28](http://www.i-r-i-e.net/inhalt/003/003_full.pdf#page=28).
- Eric RIES : The Lean Startup. *Working Paper*, pages 1–28, 2011. URL <http://lean.st/>.
- Garry ROBINS : A tutorial on methods for the modeling and analysis of social network data. *Journal of Mathematical Psychology*, pages 1–14, mars 2013. ISSN 00222496. URL <http://linkinghub.elsevier.com/retrieve/pii/S0022249613000126>.
- Daniel M. ROMERO, Brendan MEEDER et Jon KLEINBERG : Differences in the Mechanics of Information Diffusion Across Topics : Idioms, Political Hashtags, and Complex Contagion on Twitter. In *Proceedings of the 20th ACM International World Wide Web Conference*, Hyderabad, 2011. ACM. URL <http://www2011india.com/proceeding/proceedings/p695.pdf>.
- Nicole RONALD, Virginia DIGNUM, Catholijn JONKER, Theo ARENTZE et Harry TIMMERMANS : On the engineering of agent-based simulations of social activities with social networks. *Information and Software Technology*, 54(6):625–638, 2012. ISSN 09505849. URL <http://linkinghub.elsevier.com/retrieve/pii/S0950584911002485>.
- Richard B ROTHENBERG : Commentary : Sampling in Social Networks. *Connections*, 18(1):104–110, 1995.
- Matthew A RUSSEL : *Mining the Social Web*. O'Reilly Media, 2011.
- Bertrand RUSSELL : *The Analysis of Mind*. 1921. URL [http://books.google.fr/books/about/The\\_Analysis\\_of\\_Mind.html?id=4dYLAAAAIAAJ&pgis=1](http://books.google.fr/books/about/The_Analysis_of_Mind.html?id=4dYLAAAAIAAJ&pgis=1).

- G. SALTON, A. WONG et C. S. YANG : A vector space model for automatic indexing, 1975. ISSN 00010782.
- Valérie SCHAFER et Hervé LE CROSNIER : *La neutralité de l ' Internet : un enjeu de communication*. Coll. « Les essentiels d'Hermès », 2011.
- Susan SCHREIBMAN, Ray SIEMENS et John UNSWORTH : *A Companion to Digital Humanities*. John Wiley and Sons, 2007.
- John SCOTT : Social Network Analysis. *Sociology*, 22(1):109–127, 1988. ISSN 0038-0385, 1469-8684. URL [http://soc.sagepub.com/content/22/1/109\\$\\delimiter"026E30F\\$nhhttp://soc.sagepub.com/content/22/1/109.full.pdf\\$\\delimiter"026E30F\\$nhhttp://soc.sagepub.com/content/22/1/109.short](http://soc.sagepub.com/content/22/1/109$\\delimiter).
- Richard Wolfgang SEMON et Vernon LEE : *Mnemic psychology*. G. Allen & Unwin, ltd., 1923. URL <http://books.google.com/books?id=LC4XAAAAYAAJ&pgis=1>.
- Sebastian SEUNG : *Connectome : How the Brain's Wiring Makes Us Who We Are*. Houghton Mifflin Harcourt, 2012. ISBN 0547508174. URL <http://books.google.com/books?id=GXwEu0Yl3wQC&pgis=1>.
- Meng SHI : *Brand Management via China's Sina Weibo : Case Studies on Durex and VANCL*. Thèse de doctorat, Faculty of the American University, 2012.
- Jonathan SILVERTOWN : A new dawn for citizen science. *Trends in Ecology & Evolution*, 24(9):467–471, 2009. URL <http://www.ncbi.nlm.nih.gov/pubmed/19586682>.
- Gilbert SIMONDON : *Du mode d'existence des objets techniques*. Aubier philosophie, ISSN 1248-8658. Aubier, 1989. ISBN 2700734149. URL <http://www.sudoc.abes.fr/DB=2.1/SRCH?IKT=12&TRM=056398972>.
- James SOMERS : The Man Who Would Teach Machines to Think. *The Atlantic*, 2013. URL <http://www.theatlantic.com/magazine/archive/2013/11/the-man-who-would-teach-machines-to-think/309529/>.
- Pascal SOUCY et Guy W MINEAU : Beyond TFIDF Weighting for Text Categorization in the Vector Space Model. *Corpus*, 19(1):1130–1135, 2005. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.105.4957&rep=rep1&type=pdf>.
- C SOUKUP : Computer-mediated communication as a virtual third place : building Oldenburg's great good places on the world wide web. *New Media & Society*, 8(3):421–440, 2006. ISSN 14614448. URL <http://nms.sagepub.com/cgi/doi/10.1177/1461444806061953>.
- Isabelle STENGERS : Penser les sciences par leur milieu. *Rue Descartes*, 41(3):41, août 2003. ISSN 1144-0821. URL [http://www.cairn.info/article.php?ID\\_ARTICLE=RDES\\_041\\_0041](http://www.cairn.info/article.php?ID_ARTICLE=RDES_041_0041).
- Isabelle STENGERS, Brian MASSUMI et Erin MANNING : Histoire du milieu : entre macro et mésopolitique. *Inflexions*, (3):1–17, 2009.



- Alexandre STEYER, Renaud GARCIA-BARDIDIA et Pascale QUESTER : Online Discussion Groups as Social Networks : An Empirical Investigation of Word-of-Mouth on the Internet. *Journal of Interactive Advertising*, 6(2):51–60, 2006. ISSN 15252019. URL <http://www.jiad.org/article80>.
- Bernard STIEGLER : Leroi- Gourhan : l'inorganique organisé. *Cahier de Médiologie*, (6):187–194, 1998a. URL <http://mediologie.org/cahiers-de-mediologie/>.
- Bernard STIEGLER : Temps et individuation technique, psychique, et collective dans l'oeuvre de Simondon. *Intellectica*, 1-2(26-27):241–256, 1998b.
- Bernard STIEGLER : *Réenchâter le monde : La valeur esprit contre le populisme industriel*. Flammarion, Paris, France, champs ess édition, 2008. ISBN 2081217848.
- Bernard STIEGLER : The Aufklärung in the Age of Philosophical Engineering. In *WWW12 Conference Proceedings*, 2012.
- Max STIRNER : *Stirner : The Ego and Its Own*. Cambridge University Press, 1995. ISBN 0521456479. URL <http://books.google.com/books?id=Ap2VLD38XYAC&pgis=1>.
- J. SULLIVAN : A tale of two microblogs in China. *Media, Culture & Society*, 34(6):773–783, août 2012. ISSN 0163-4437. URL <http://mcs.sagepub.com/content/34/6/773.extract>.
- Ferhat TAYLAN : Centralité et modifiabilité. *Le Portique*, 25(25), août 2010. ISSN 1283-8594. URL <http://leportique.revues.org/index2475.html>.
- Nigel THRIFT et Shaun FRENCH : The automatic production of space. *Transactions of the Institute of British Geographers*, 27(3):309–335, septembre 2002. ISSN 0020-2754. URL <http://www.blackwell-synergy.com/links/doi/10.1111/1475-5661.00057>.
- Paul M TORRENS : Geography and computational social science. *GeoJournal*, 75(2):133–148, 2010. ISSN 03432521. URL <http://www.springerlink.com/index/10.1007/s10708-010-9361-y>.
- Jeffrey TRAVERS et Stanley MILGRAM : An Experimental Study of the Small World Problem. *Sociometry*, 32:425–443, 1969. URL <http://www.jstor.org/stable/2786545>.
- Lokman TSUI : An Inadequate Metaphor : The Great Firewall and Chinese Internet Censorship. *Global Dialogue*, 9(1-2), avril 2007. URL <http://www.worlddialogue.org/content.php?id=400>.
- Paola TUBARO et Antonio A CASILLI : An ethnographic seduction” : how qualitative research and Agent-based models can benefit each other. *Bulletin de Méthodologie Sociologique*, 106(1): 59–74, 2010. URL <http://dx.doi.org/10.1177/0759106309360111>[http://hal.archives-ouvertes.fr/docs/00/66/14/25/PDF/EthnoSeductionBMSvol106n1pp116\\_draft.pdf](http://hal.archives-ouvertes.fr/docs/00/66/14/25/PDF/EthnoSeductionBMSvol106n1pp116_draft.pdf).
- Andrew J TURNER : *Introduction to Neogeography*. O'Reilly, short cuts édition, 2006. ISBN 9780596529956.

- Christian VANDENDORPE : *Du Papyrus à l'hypertexte. Essais sur les mutations du texte et de la lecture*. Editions l'édition, 1999.
- Paul VIRILIO : *La bombe informatique*. Galilée, 1998. ISBN 2718605073. URL <http://www.amazon.fr/La-bombe-informatique-Virilio-P/dp/2718605073>.
- John von NEUMANN : *The Computer and the Brain : Second Edition (Mrs. Hepsa Ely Silliman Memorial Lectures)*. Yale University Press, 1958. ISBN 0300084730. URL <http://www.amazon.com/The-Computer-Brain-Silliman-Memorial/dp/0300084730>.
- Hunter WALKER : Obama Campaign Deploys Cat Meme to Get Out the Vote in Ohio, 2012. URL <http://politicker.com/2012/11/obama-campaign-deploys-cat-meme-to-get-out-the-vote-in-ohio/>.
- Cara WALLIS : *Technomobility in China : Young Migrant Women and Mobile Phones (Critical Cultural Communication)*. NYU Press, 2013. ISBN 0814795269. URL <http://www.amazon.com/Technomobility-China-Critical-Cultural-Communication/dp/0814795269>.
- Greg WALTON : China ' s Golden Shield : Corporations and the Development of Surveillance Technology in China. Rapport technique, International Centre for Human Rights and Democratic Development, 2001.
- Bryan WANG, Dane ANDERSON, Andrew BARTELS, Michael BARNES et Rita XIA : China Tech Market Outlook : 2012 To 2013. Rapport technique, Forrester, 2013. URL <http://www.forrester.com/China+Tech+Market+Outlook+2012+To+2013/fulltext/-/E-RES81022>.
- Lin WANG et Brendan C. WOOD : An epidemiological approach to model the viral propagation of memes. *Applied Mathematical Modelling*, 35(11):5442–5447, novembre 2011. ISSN 0307904X. URL <http://linkinghub.elsevier.com/retrieve/pii/S0307904X11002824>.
- Colin WARE : *Information Visualization : Perception for Design*, volume 22 de *The Morgan Kaufmann series in interactive technologies / ed. Stuart Card*. - San Francisco : Morgan Kaufmann. Morgan Kaufmann, 2004. ISBN 1558608192. URL [http://books.google.com/books?id=ZmG\\_FiqqygC&pgis=1](http://books.google.com/books?id=ZmG_FiqqygC&pgis=1).
- Tetsurô WATSUJI, Augustin BERQUE, Pauline COUTEAU et Kurodamm AKINOBU : *Fûdo : Le milieu humain*. CNRS, Paris, 2011. ISBN 2271071372.
- J L WEISSBERG et P BARBOZA : *L'image actée, Scénarisations numériques, parcours du séminaire*. L'harmatta édition, 2007.
- L. WENG, A. FLAMMINI, A. VESPIGNANI et F. MENCZER : Competition among memes in a world with limited attention. *Scientific reports*, 2:335, janvier 2012. ISSN 2045-2322. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3315179&tool=pmcentrez&rendertype=abstract>.

- Al B. WESOLOWSKY et Edward R. TUFTE : The Visual Display of Quantitative Information. *Journal of Field Archaeology*, 19(1):121, 1992. ISSN 00934690. URL <http://www.jstor.org/stable/530384?origin=crossref>.
- Norbert WIENER : *Cybernetics : Or the Control and Communication in the Animal and the Machine*. {MIT} pres édition, 1948.
- Norbert WIENER : *The Human Use of Human Beings : Cybernetics and Society*. The Da Capo Series in Science. Da Capo Press, 1988. ISBN 0306803208. URL <http://www.amazon.co.uk/dp/0306803208>.
- Michel WIEVIORKA : *L'impératif numérique ou La nouvelle ère des sciences humaines et sociales ?* CNRS, 2013. ISBN 978-2-271-07981-7. URL <http://lectures.revues.org/12677>.
- Yves WINKIN : La nouvelle communication. In *La nouvelle communication*, pages 115–145. Seuil, 1981. ISBN 2020427842.
- Philipp WINTER et Stefan LINDSKOG : How China Is Blocking Tor. *Arxiv preprint arXiv:1204.0447*, (March), 2012. URL <http://arxiv.org/abs/1204.0447>.
- Ludwig WITTGENSTEIN : *Recherches philosophiques*. 2004.
- Dominique WOLTON : *Penser la communication*. Flammarion, Paris, France, champs ess édition, 1997. ISBN 2081212994.
- Xu WU : Chinese Cyber Nationalism : Evolution, Characteristics and Implications. 2007.
- Nianwen XUE : Chinese Word Segmentation as Character Tagging. *Computational Linguistics*, 8:29–48, 2003.
- G. YANG : Political contestation in Chinese digital spaces : Deepening the critical inquiry. *China Information*, 28(2):135–144, juillet 2014. ISSN 0920-203X. URL <http://cin.sagepub.com/cgi/doi/10.1177/0920203X14539910>.
- Guobin YANG : Activism beyond Virtual borders : Chinese Internet-mediated Networks and information politics in China. *Command Lines : The Emergence of Governance in Global Cyberspace*, 7, 2008.
- Jiang YANG, Mark S ACKERMAN et Lada A ADAMIC : Virtual Gifts and Guanxi : Supporting Social Exchange in a Chinese Online Community. *CSCW*, (March 19-23):45–54, 2011. URL <http://portal.acm.org/citation.cfm?id=1958832>.
- Robert K YIN : *Case Study Research : Design and Methods*, volume 5. 2009. ISBN 9781412960991. URL <http://books.google.com/books?id=FzawIAdilHkC&pgis=1>.
- Louis YU, Sitaram ASUR et Bernardo A. HUBERMAN : What Trends in Chinese Social Media. In *The 5th SNA-KDD Workshop'11 (SNA-KDD'11)*, San Diego CA USA, juillet 2011. ACM.

ISBN 9781450302258. URL <http://www.ssrn.com/abstract=1888779><http://arxiv.org/abs/1107.3522>.

Louis YU et Valerie KING : Guanxi in the Chinese Web-A Study of Mutual Linking. *In WWW2008*, volume 30, pages 35–41, Beijing, China, 2008. A postdoctoral researcher in the Social Computing Lab at HP Labs., ACM. URL <http://portal.acm.org/citation.cfm?doid=1367497.1367706>.

Xun ZHAO, Feida ZHU, Weining QIAN et Aoying ZHOU : Impact of Multimedia in Sina Weibo : Popularity and Life Span. *In The Joint Conference of the Sixth Chinese Semantic Web Symposium and the First Chinese Web Science Conference (CSWS & CWSC '12)*, pages 1–11, 2012. URL <http://www.mysmu.edu/faculty/fdzhu/paper/CWSC'12.pdf>.

Feng ZHEN, Bo WANG, Guangliang XI et Yinxue CHEN : Research on Chinese Urban Network Based on the Relations between Micro-blog Users. *In* Pietro ELISEI MANFRED SCHRENK, VASILY V. POPOVICH, PETER ZEILE, éditeur : *REAL CORP 2013*, 2013. URL <http://www.corp.at>.

M. a. ZOOK : Old Hierarchies or New Networks of Centrality? : The Global Geography of the Internet Content Market. *American Behavioral Scientist*, 44(10):1679–1696, juin 2001. ISSN 0002-7642. URL <http://abs.sagepub.com/cgi/doi/10.1177/00027640121958113>.

MA ZOOK et Mark GRAHAM : From cyberspace to DigiPlace : Visibility in an age of information and mobility. *Societies and cities in the age of instant access*, pages 241–254, 2007. URL [http://link.springer.com/content/pdf/10.1007/1-4020-5427-0\\_16.pdf](http://link.springer.com/content/pdf/10.1007/1-4020-5427-0_16.pdf).

Arkaitz ZUBIAGA, Damiano SPINA et Raquel MARTÍNEZ : Classifying Trending Topics : A Typology of Conversation Triggers on Twitter. *Evaluation*, pages 8–11, 2011. URL <http://nlp.uned.es/~damiano/pdf/zubiaga2011trendingtopics.pdf>.

Ethan ZUCKERMAN : The Cute Cat Theory Talk at ETech. *In O'Reilly E-tech conference2*. O'Reilly Conferences, 2008. URL <http://www.ethanzuckerman.com/blog/2008/03/08/the-cute-cat-theory-talk-at-etech/http://en.oreilly.com/et2008/public/schedule/detail/1597>.

Ethan ZUCKERMAN : *Rewire : Digital Cosmopolitans in the Age of Connection*. W. W. Norton & Company, 2013. ISBN 0393240622. URL <http://books.google.com/books?id=fpbTUSM6Kb8C&pgis=1>.

# Table des matières

Abstract . . . . .	iii
Remerciements . . . . .	1
Sommaire . . . . .	2
<b>Introduction</b>	<b>1</b>
<b>1 Sina Weibo et le milieu numérique en Chine</b>	<b>7</b>
1.1 L’Internet chinois : éléments de contexte . . . . .	8
1.1.1 Petite histoire et évolution de l’Internet chinois . . . . .	9
1.1.2 Censure sur l’Internet chinois . . . . .	11
1.2 Médias sociaux en Chine : un paysage morcelé . . . . .	15
1.2.1 Microblog en Chine et Sina Weibo . . . . .	17
1.2.2 Sina Weibo, un usage plus ludique que <i>Twitter</i> . . . . .	21
1.3 Code, langage et milieu(x) numérique(s) . . . . .	25
1.3.1 Lieu, espace, territoire et technologies . . . . .	25
1.3.2 Le milieu : richesse et désuétude . . . . .	35
1.3.3 Cyberspace et milieu numérique . . . . .	38
1.3.4 Topogrammes : association et dissociation dans le milieu numérique . . . .	40
<b>2 Les mèmes Internet, objets numériques culturels</b>	<b>45</b>
2.1 Les mèmes : définitions et histoire . . . . .	46
2.1.1 La mémétique : une éthologie culturelle teinté d’évolutionnisme . . . . .	46
2.1.2 Mèmes Internet : définition, littérature et exemples . . . . .	49
2.2 Mème, mémoire collective et culture . . . . .	53

2.2.1	La mémoire comme trace . . . . .	53
2.2.2	Diffusion de mèmes et structuration d’une mémoire collective . . . . .	55
2.3	Textualité des mèmes et formes d’énonciations numériques . . . . .	61
2.3.1	Le mème comme figure rhétorique de l’écriture intertextuelle . . . . .	61
2.3.2	Typologie des mèmes Internet . . . . .	66
<b>Problématique et hypothèses</b>		<b>77</b>
<b>3 Réflexions méthodologiques et design expérimental</b>		<b>81</b>
3.1	Sciences sociales du réseau . . . . .	82
3.1.1	Le réseau comme enjeu pour l’étude . . . . .	82
3.1.2	Code, données et les nouveaux outils de l’écriture scientifique . . . . .	85
3.1.3	Visualisation et espace perceptif pour l’information . . . . .	89
3.2	L’analyse de la diffusion sur les réseaux sociaux . . . . .	91
3.2.1	Anatomie d’un réseau social . . . . .	92
3.2.2	Analyse de réseaux sociaux et de la diffusion d’information en ligne . . .	96
3.2.3	Modèles pour la visualisation des conversations . . . . .	101
3.3	Méthodes d’identification de mèmes dans un large volume de données . . . . .	112
3.3.1	Constitution et collection d’un corpus de messages . . . . .	112
3.3.2	Détection algorithmique de mèmes . . . . .	116
3.3.3	Les hashtags ne sont pas des mèmes . . . . .	120
<b>4 Résultats et discussions</b>		<b>125</b>
4.1	Un outil de traitement et de visualisation des mèmes . . . . .	127
4.1.1	Constitution de corpus par mots-clés pour chaque mème . . . . .	127
4.1.2	Analyse des graphes et traitement des données . . . . .	132
4.1.3	Serveur et moteur de visualisation interactive . . . . .	138
4.2	Résultats sur un échantillon de mèmes . . . . .	147
4.2.1	Structures temporelles . . . . .	154
4.2.2	Structures conversationnelles . . . . .	157
4.2.3	Structures sémantiques . . . . .	159

4.2.4	Dynamiques géographiques et discussions entre provinces . . . . .	161
4.2.5	Communautés, clusters et groupes de provinces . . . . .	167
4.2.6	Dimensions géographiques des graphes socio-sémantiques . . . . .	170
4.2.7	Résumé des résultats . . . . .	177
4.3	Discussions . . . . .	178
<b>Conclusion</b>		<b>198</b>
<b>Annexes</b>		<b>199</b>
<b>Protocole de recherche</b>		<b>201</b>
<b>Memes : Iconographie</b>		<b>205</b>
<b>Bibliographie</b>		<b>238</b>
<b>Table des Matières</b>		<b>241</b>
<b>Listes</b>		<b>241</b>
	Liste des Figures . . . . .	246
	Liste des algorithmes . . . . .	247
	Liste des tableaux . . . . .	249





# Table des figures

1.1.1	Jingjing et Chacha, les policiers de l'Internet chinois . . . . .	13
1.2.1	Capture d'écran de Sina Weibo . . . . .	18
1.2.2	Sondage Weibo concernant l'accident de train de Wenzhou . . . . .	19
1.2.3	Cours de l'action SINA au Nasdaq entre 2009 et Février 2014 . . . . .	20
1.2.4	Commentaires supprimés par Sina . . . . .	24
1.3.1	Hexie, les crabes de rivières . . . . .	31
2.1.1	la première mention du smiley par Scott Fahlman . . . . .	50
2.1.2	Un jeu d'émoticônes de Weibo . . . . .	51
2.1.3	Cycle de vie d'un mème Internet . . . . .	52
2.1.4	Exemples de mèmes internet . . . . .	53
2.2.1	Réponses à la question à propos des mèmes indiens sur Quora . . . . .	57
2.2.2	Mème "High Expectations Asian Father" d'après Quora.com . . . . .	58
2.3.1	Some LOLcats . . . . .	69
2.3.2	<i>Caonima</i> : un animal mythique du web chinois . . . . .	70
2.3.3	Lolcat utilisé lors la campagne d'Obama . . . . .	72
2.3.4	Détournement de tracts hongkongais anti-chinois . . . . .	73
2.3.5	La star du X japonaise Sola Aoi sur Sina Weibo . . . . .	74
2.3.6	Extrait d'un spam du type <i>Nigerian Scam</i> . . . . .	75
3.1.1	Triptyque des pratiques scientifiques . . . . .	86
3.2.1	Représentation d'un réseau sous forme de graphes. . . . .	93
3.2.2	réseau social multi-couches . . . . .	94

3.2.3	Types communs de réseaux . . . . .	95
3.2.4	Tableau récapitulatif de méthodes d'analyse de données de réseaux sociaux en ligne. . . . .	100
3.2.5	3 modèles de réseau . . . . .	101
3.2.6	Graphe de diffusion de hashtags sur Twitter d'après (Weng <i>et al.</i> , 2012) . . .	102
3.2.7	Visualisation simple des conversations de deux hashtags . . . . .	104
3.2.8	Exemple 1 : Tianjin Baoyu . . . . .	105
3.2.9	Photo de Tianjin durant la pluie torentielle en Juillet 2012 . . . . .	106
3.2.10	Photo illustrative de Shidu Muqin . . . . .	107
3.2.11	“Shidu Muqin” . . . . .	108
3.2.12	“Piaowudong nuzui” . . . . .	109
3.2.13	Modèle de Jakobson . . . . .	111
3.3.1	Captures d'écran de l'outil d'analyse de microblog Soudy . . . . .	118
3.3.2	Algorithme de reconnaissance de mème (clustering) (Ferrara <i>et al.</i> , 2013) . . .	118
3.3.3	Algorithme d'extraction d'entités d'après les messages . . . . .	121
3.3.4	Hashtags les plus utilisés pendant la 1ère semaine de 2012 . . . . .	121
3.3.5	Distribution des 429 hashtags les plus discutés . . . . .	122
4.1.1	Tableau de bords requêtes par mots-clés . . . . .	131
4.1.2	Interface d'exploration et de visualisation des données. Capture d'écran réalisée le 23 Mai 2014 à 16h18 . . . . .	141
4.1.3	Structure de l'interface d'exploration et de visualisation des données . . . . .	142
4.1.4	Carte interactive de la Chine (exemple contenant des données générées aléatoirement) . . . . .	143
4.1.5	Menu de sélection des mèmes . . . . .	144
4.1.6	Cette timeline interactive permet de sélectionner une période de temps . . . .	145
4.2.1	Illustrations du mème <i>biaoge</i> d'après le Guardian <a href="http://www.theguardian.com/world/2013/sep/05/china-brother-wristwatch-yang-dacai-sentenced">http://www.theguardian.com/world/2013/sep/05/china-brother-wristwatch-yang-dacai-sentenced</a> , consulté le 7 Juillet 2014 à 12 :11. et la BBC <a href="http://www.bbc.com/news/world-asia-china-23956170">http://www.bbc.com/news/world-asia-china-23956170</a> , consulté le 7 Juillet 2014 à 12 :23. . . . .	149
4.2.2	Quelques images du mème <i>qiegao</i> . . . . .	150
4.2.3	. . . . .	152

4.2.4	Graphe temporel représentant le volume de messages échangés autour de 4 mèmes sur une durée de 3 semaines . . . . .	155
4.2.5	Graphe temporel représentant le volume de messages échangés autour du mème <i>yuanfang</i> entre le 7 Octobre et le 16 Décembre 2012. . . . .	156
4.2.6	Graphe temporel représentant le volume de messages échangés autour du mème <i>moyan</i> entre le 8 et le 21 Octobre 2012. . . . .	156
4.2.7	Graphe conversationnel pour quatre mèmes (sur une durée de 3 semaines) . .	158
4.2.8	Graphe sémantique de co-occurrences des mots pour les mèmes dufu et yuanfang	159
4.2.9	Graphe sémantique de co-occurrences des mots pour les mèmes Sextape et Biaoge . . . . .	160
4.2.10	Graphe sémantique de co-occurrences des mots pour les mèmes CCP et The Voice . . . . .	160
4.2.11	Graphe sémantique de co-occurrences des mots pour les mèmes Qiegao et Moyan	161
4.2.12	The Voice : Interaction des utilisateurs par province entre le 9 et le 29 Juillet 2012 . . . . .	162
4.2.13	The Voice : Interaction des utilisateurs par province entre le 16 et le 17 Juillet 2012 . . . . .	162
4.2.14	Qiegao : Interaction des utilisateurs par province entre le 19 Novembre et le 15 Décembre 2012 . . . . .	163
4.2.15	Qiegao : Interaction des utilisateurs par province . . . . .	163
4.2.16	Sextape : Interaction des utilisateurs par province . . . . .	164
4.2.17	Biaoge : Interaction des utilisateurs par province . . . . .	164
4.2.18	Répartition des utilisateurs par province (en % du total) . . . . .	164
4.2.19	Moyan : Interaction des utilisateurs par province . . . . .	165
4.2.20	Répartition des utilisateurs par province pour Moyan (en % du total) . . . . .	165
4.2.21	Interaction des utilisateurs par province . . . . .	166
4.2.22	Evolution des interactions des utilisateurs par province pour Dufu et Yuanfang	166
4.2.23	Interaction entre utilisateurs de la même province . . . . .	167
4.2.24	Répartition des utilisateurs par province (en % du total) . . . . .	167
4.2.25	Le modèle urbain chinois du Three majors and four smalls d'après Zhen <i>et al.</i> (2013) . . . . .	168

4.2.26	Communauté de provinces dessinées par les échanges entre utilisateurs autour de chaque même (Algorithme de Louvain). . . . .	169
4.2.27	Communauté de provinces dessinées par les échanges entre utilisateurs autour de Sextape . . . . .	171
4.2.28	Communauté de provinces dessinées par les échanges entre utilisateurs autour de Biaoge . . . . .	171
4.2.29	Communauté de provinces dessinées par les échanges entre utilisateurs autour de The Voice . . . . .	171
4.2.30	Communauté de provinces dessinées par les échanges entre utilisateurs autour de Moyan . . . . .	172
4.2.31	Communauté de provinces dessinées par les échanges entre utilisateurs autour de Dufu . . . . .	173
4.2.32	biaoge : distribution des citations par provinces pour le mot biao (montre) . .	173
4.2.33	The Voice : distribution des citations de mots par provinces . . . . .	174
4.2.34	Dufu : distribution des citations de mots par provinces . . . . .	175
4.2.35	Qiegao : origine géographique des utilisateurs . . . . .	176
.0.1	Protomemes algortithm workflow . . . . .	203
.0.2	Data mining and visualisation workflow . . . . .	204
.0.3	Dufu is very busy . . . . .	207
.0.4	The Voice of China : Illustrations . . . . .	208
.0.5	Images du même entourant les montres de Yang Dacai . . . . .	209
.0.6	Exemples d'images du même <i>Yuanfang</i> . . . . .	210
.0.7	Mo Yan Prix Nobel : Illustrations . . . . .	211
.0.8	18ème Congrès du Parti Communiste : Illustrations . . . . .	212
.0.9	Exemples d'image du même concernant la sextape de Lei Zhengfu. Images d'après l'article "Sex tape official stands trial in Chongqing" dans le <i>People's Daily</i> , <a href="http://english.peopledaily.com.cn/90882/8290628.html">http://english.peopledaily.com.cn/90882/8290628.html</a> et <a href="http://www.dailymail.co.uk/news/article-2239300/Zhao-Hongxia-Teenage-honeytrap-brought-Chinese-Communist-Party-official-sex-tape-pictures-leaked-online.html">http://www.dailymail.co.uk/news/article-2239300/Zhao-Hongxia-Teenage-honeytrap-brought-Chinese-Communist-Party-official-sex-tape-pictures-leaked-online.html</a> consulté le 7 Juillet à 12 :32 . . . . .	213
.0.10	Qiegao . . . . .	214

# Liste des algorithmes

1	Hashtags conversational graphs . . . . .	103
2	Extract Hashtags from Message . . . . .	121
3	Extract Words from Message . . . . .	135
4	Word Graph extraction from Meme messages corpus . . . . .	136
5	Extract User Graph from Meme Corpus . . . . .	137



# Liste des tableaux

2.2.1 Etapes constitutantes de la mémoire - Convergence entre la trace psychique et synaptique (Ansermet et Magistretti, 2004) . . . . .	54
2.2.2 Les trois formes de mémoire d'après Leroi-Gourhan et Stiegler . . . . .	55
2.3.1 Typologie des mèmes . . . . .	68
3.3.1 Modèle de données du données Weiboscope . . . . .	116
4.2.1 Dénomination et description des mèmes étudiés . . . . .	147
4.2.2 Résumé des résultats . . . . .	177
.0.1 Liste des mèmes identifiés . . . . .	206

# Conception d'un outil d'analyse et de visualisation des mèmes Internet - Le cas du réseau social chinois Sina Weibo

Clément RENAUD

**RESUME :** Nous concevons et développons un outil permettant d'analyser la diffusion d'information sur les services de réseaux sociaux en ligne sous la forme de motifs appelés *topogrammes*. Fruit d'une réflexion méthodologique sur la visualisation de données, ce dispositif permet d'observer les relations entre les dimensions conversationnelles, sémantiques, temporelles et géographiques des actes de communication en ligne. En mobilisant des méthodes issues de l'analyse des réseaux et du traitement automatisé de la langue chinoise, nous procédons à l'analyse d'un ensemble de contenus d'après un vaste corpus de 200 millions de messages représentant l'activité sur le réseau social *Sina Weibo* durant l'année 2012. Courts messages se propageant rapidement sur la Toile, les mèmes Internet articulent discussions personnelles, débats sociétaux et vastes campagnes médiatiques selon des modèles encore mal connus. Une visualisation interactive nous permet de comprendre les réseaux de mots et d'échanges entre utilisateurs en jeu dans la circulation d'un ensemble de mèmes Internet, ainsi que leurs projections sur des cartes géographiques. Une première analyse portant sur les hashtags indique que les usages majoritaires de *Sina Weibo* sont similaires à ceux des médias traditionnels (publicité, divertissement, loisirs...). L'approche retenue étudie une dizaine de mèmes sélectionnés pour leurs intentions diverses : humour, actualité, scandale politique, faits divers et marketing promotionnel. Les figures obtenues permettent d'ébaucher une typologie de la diffusion de ces contenus d'après les topogrammes décrits lors de la circulation de ces contenus en ligne. Nous observons que la tension entre énonciation et discours sur les plate-formes Web se manifeste par la cohabitation de différents régimes d'expression. Nous concluons en formulant des recommandations pour l'analyse et la conception de stratégies de communication d'organismes tant privés que publics.

**MOTS-CLEFS :** Internet chinois, topogrammes, Sina Weibo, mèmes, visualisation, analyse de réseaux sociaux

**ABSTRACT :** We develop a data mining and visualisation toolkit to study how the information is shared on online social network services. This software allows to observe relationships between conversational, semantical, temporal and geographical dimensions of online communication acts. Internet memes are short messages that spread quickly through the Web. Following models that remain largely unknown, they articulate personal discussions, societal debates and large communication campaign. We analyse a set of Internet memes by using methods from social network analysis and Chinese natural language processing on a large corpus of 200 million tweets which represents/reflects the overall activity on the Chinese social network Sina Weibo in 2012. An interactive visualisation interface showing networks of words, user exchanges and their projections on geographical maps provides a detailed understanding of actual and textual aspects of each meme spread. An analysis of hashtags in the corpus shows that the main content from Sina Weibo is largely similar to the ones in traditional media (advertisement, entertainment, etc.) Therefore, we decided to not consider hashtags as memes representatives, being mostly byproducts of well-planned strategic or marketing campaigns. Our final approach studies a dozen of memes selected for the diversity of their topic : humor, political scandal, breaking news and marketing.

**KEYWORDS :** Chinese Internet, topograms, Sina Weibo, memes, visualization, social network analysis

