

MOUTONS ET FORTES TÊTES

Ce chapitre a donné lieu à l'article suivant, présenté à la conférence InfoCom 2010 de San Diego (États-Unis) :

An Interaction-Based Mobility Model for Dynamic Hot Spot Analysis, Frédéric Morlot, Salah-Eddine Elayoubi, François Baccelli, Infocom, Proceedings IEEE (2010).

Par la suite, nous en avons publié une version remaniée et enrichie sur HAL.

Résumé

Les traces laissées par les téléphones GSM dans un réseau de téléphonie pendant les rassemblements urbains révèlent une grande volatilité des comportements. De plus, on assiste fréquemment à la formation de "points chauds" (*hotspots* en anglais) comme on peut le voir sur la figure 3.1, c'est-à-dire des rassemblements significatifs et spontanés. Un phénomène-clé que nous souhaitons étudier est la formation et la disparition de ces rassemblements, qui paraissent aléatoires à la fois dans le temps et dans l'espace. Après avoir prouvé la nécessité d'introduire un nouveau modèle de mobilité pour rendre compte de ces phénomènes, nous présentons le "modèle des moutons et des fortes têtes", fondé sur les interactions entre individus.

3.1 Introduction

3.1.1 Données disponibles

Précisons un peu ce que nous entendons par "aléatoire dans le temps et dans l'espace" lorsque nous parlons des hotspots. On rappelle (voir introduction de la thèse), que les traces utilisées pour notre étude couvraient :

- la fête de la St-Jean à Varsovie et Cracovie (21 juin)
- la Fête de la Musique à Paris (21 juin)
- les derniers tours de l'Euro de Football à Bucarest (13 juin), Madrid et Barcelone (28 juin).

Dans le cas de la St-Jean, les usagers se sont simplement promenés dans la ville sans direction particulière. Le caractère aléatoire des rassemblements est donc bien établi. Seule exception à la règle : un concert sur une scène géante qui a eu lieu dans un parc de Varsovie. Mais même dans ce cas, l'accès était libre si bien que les individus pouvaient vaquer autour et se sont largement déplacés (comme peut en témoigner notre équipe qui était sur place).

Le cas de la Fête de la Musique est assez similaire : à part pour un petit nombre de concerts géants bien identifiés et planifiés, les rassemblements peuvent être considérés comme spontanés. Et même pour lesdits concerts géants, le fait qu'il y en ait plusieurs au même moment à différents endroits incitait les individus à se déplacer en cours de route et réintroduisait une dimension aléatoire.

Enfin, pour ce qui est des matchs de football, la première phase (retransmission sur écran géant) était aléatoire à cause des différentes localisations des écrans dans la ville. Et la deuxième phase (festive après la victoire) était clairement la plus spontanée et la moins planifiée des deux.

Un modèle de mobilité à la fois réaliste et calculable nous permettrait d'améliorer significativement la robustesse des simulateurs en télécommunication et des modèles analytiques, ce qui nous permettrait de mieux comprendre l'impact de la mobilité sur la qualité de service. La prévision des rassemblements permet aussi de mettre au point des algorithmes de gestion dynamique des ressources radio, comme l'allocation dynamique de spectre (voir [Lea04]) ou la radio logicielle (voir [Mit00][Sim07]).



Fig. 3.1: SMS envoyés pendant la Fête de la Musique à Paris, le 21 juin 2008. Il est 23h13. On peut observer d’importants rassemblements dans le centre-ville (Châtelet, Bastille, Saint-Michel), ainsi qu’au Parc des Princes (concert de *Tokio Hotel*) et à l’hippodrome d’Auteuil (concert de France 2).

Dans un premier temps, nous passons en revue les modèles de mobilité existants et nous démontrons la nécessité d’un nouveau modèle pour qui veut modéliser une dynamique des hotspots.

Ensuite, nous introduisons notre modèle markovien. Pour cela, nous divisons le territoire en zones d’attractions potentielles, et nous partons du principe que les individus s’influencent les uns les autres. Dans la partie 3.2, nous étudions un ”modèle à moutons”, pour lequel cette influence est particulièrement grande. Nous commençons par étudier un territoire à deux zones, ce qui nous permet de trouver des formules analytiques pour le temps moyen de formation d’un hotspot ainsi que pour la loi de sa position spatiale. Nous donnons aussi des équivalents asymptotiques de ces quantités, et une borne supérieure de la vitesse de convergence. Enfin, nous généralisons ces résultats à un nombre quelconque de zones. Les hotspots sont bien aléatoires en temps et en espace, mais dès qu’ils se forment ils restent stables. Il nous manque donc encore leur désagrégation.

C'est pourquoi la partie 3.3 introduit un "modèle à moutons et fortes têtes", dont le premier modèle est un cas particulier. Le modèle général est plus réaliste et il reste calculable, ce qui nous permet de modéliser aussi bien la formation de hotspots aléatoires que leur désagrégation.

Dans une dernière partie, nous exploitons les phénomènes d'explosion prédits par notre modèle pour calibrer ce dernier sur les traces de la Fête de la Musique 2008 à Paris.

3.1.2 Pourquoi un nouveau modèle ?

Sur la figure 3.1, on peut voir un instantané des traces laissées par les SMS dans Paris pendant la Fête de la Musique 2008. Tout de suite, on remarque la présence de hotspots. Et sur les animations vidéos de [Urb08], on voit bien que ces hotspots subissent une dynamique. Il ne nous est pas possible de reproduire les vidéos sur papier, mais les clichés de 3.2 donnent une idée de ce qui se passe : la foule commence par quitter petit à petit une zone (nous appellerons cela un "siphonnage") pour en remplir une autre (nous appellerons cela un "remplissage"). La combinaison des deux s'appellera un "transfert". La localisation de tels phénomènes est aléatoire, car lorsque les individus quittent un endroit, nous ne savons pas à l'avance s'ils préfèrent se rendre à tel ou tel concert. Cela dépend de plusieurs paramètres subjectifs, comme leur humeur, l'influence de leur entourage, ou simplement leurs goûts musicaux.

À présent, nous passons en revue les modèles de mobilité existants, et nous étudions s'ils pourraient être adaptés à nos besoins.

Le modèle le plus populaire en télécommunication est probablement le *Random Waypoint Model* (voir [Bou05]). Ce modèle est facile à simuler et de plus il est calculable analytiquement dans un grand nombre de situations, selon les règles de déplacement que l'on donne aux individus et la forme géométrique du territoire. Cependant, il ne s'applique qu'à un seul individu à la fois, ou alors à un groupe d'individus indépendants. En utilisant ce modèle pour faire apparaître de l'attraction, nous devrions choisir une règle de déplacement qui donne plus de poids à certaines zones prédéfinies. Ainsi, la loi de la position d'un individu favoriserait ces zones. Malheureusement, cela empêche toute dynamique au niveau macroscopique. En effet, avec un grand nombre d'individus indépendants superposés, un argument de type loi des grands nombres montre que le résultat empirique constitue un

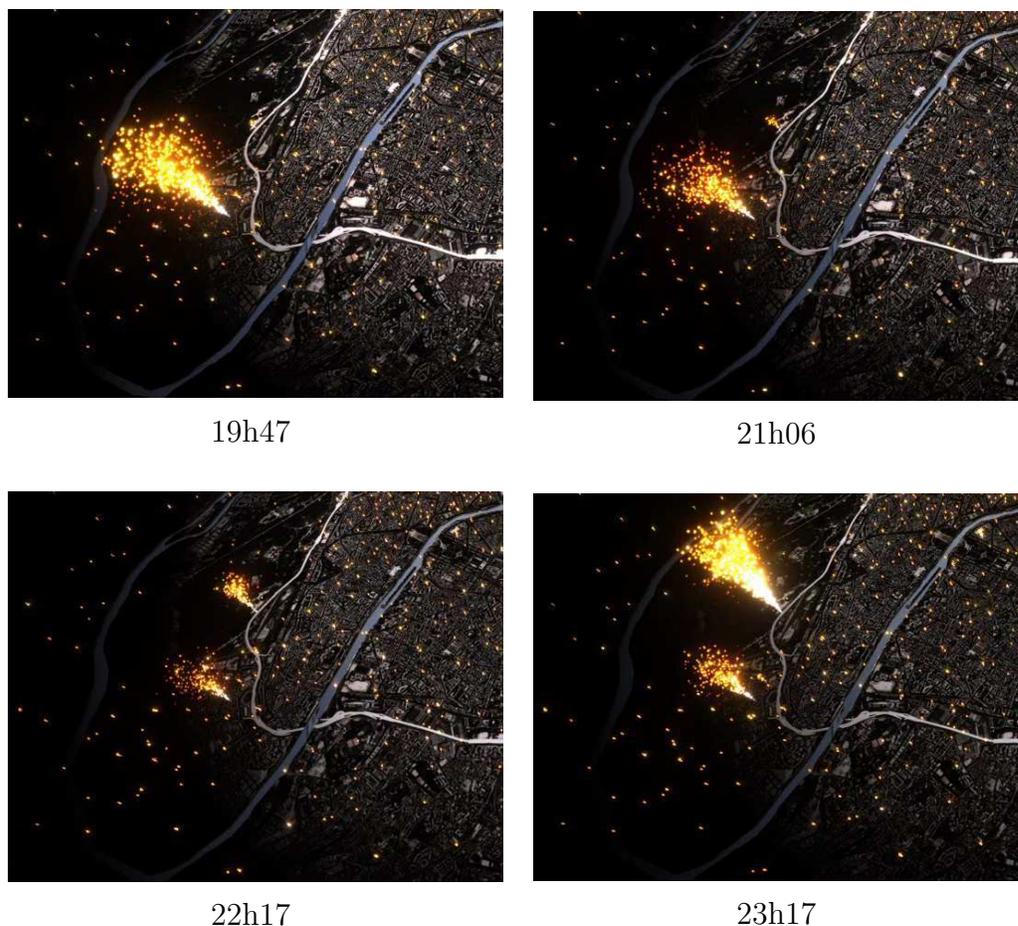


Fig. 3.2: SMS envoyés pendant la Fête de la Musique à Paris, le 21 juin 2008, entre 19h47 et 23h17. On observe un transfert du Parc des Princes à l’hippodrome d’Auteuil.

histogramme quasi-parfait de la loi théorique individuelle.

Plus précisément, cela vient du théorème de Sanov (voir [Dem93], théorème 6.2.10). Soit μ la loi de position individuelle, et X_1, X_2, \dots, X_n les positions de n utilisateurs indépendants qui suivent μ . Soit $\hat{\mu}_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}$ la loi empirique associée à une observation instantanée. Alors, quel que soit l’évènement B , le théorème de Sanov nous indique que pour n assez grand :

$$\mathbb{P}(\hat{\mu}_n \in B) \sim \exp \left[-n \inf_{\nu \in B} H(\nu | \mu) \right],$$

où $H(\nu|\mu)$ désigne la distance de Kullback-Leibler¹ entre μ et ν . En particulier, si B traduit le fait que les utilisateurs ne s'agglutinent pas là où ils devraient d'après μ , $H(\nu|\mu)$ va être grand $\forall \nu \in B$, et donc $\inf_{\nu \in B} H(\nu|\mu)$ va être grand aussi. Alors la probabilité d'une telle observation va être extrêmement faible dès que n est un tant soit peu important. Ainsi, tous les attroupements vont avoir lieu dans les zones prédéfinies, jamais ailleurs, et ils seront toujours présents. On n'aura donc pas de transfert.

D'une manière générale, on aura le même phénomène à chaque fois qu'on partira de la superposition de plusieurs modèles individuels indépendants. Or pendant les grands événements publics, un point-clé est l'interdépendance entre individus, puisque les traces qu'ils laissent dans le réseau correspondent bien souvent à des rendez-vous donnés à tel ou tel endroit. Par exemple pendant la Fête de la Musique, si un concert est de bonne qualité, les spectateurs en informeront leurs amis pour leur proposer de les rejoindre, ce qui renforcera le hotspot.

Le *Nomadic Community Mobility Model* et le *Reference Point Group Mobility Model* (voir [Cam02]) consistent à faire se déplacer globalement de petits groupes d'individus dans une direction donnée, tandis que les utilisateurs eux-mêmes peuvent graviter autour de la tendance globale. Typiquement c'est ce qui se passe avec un groupe de visiteurs qui suit un guide dans un musée. Mais dans notre cas, nous ne souhaitons pas prédéfinir une direction vers laquelle un groupe doit se déplacer. De plus, nous ne souhaitons pas que les mêmes individus restent toujours ensemble, car il n'est pas crédible que les spectateurs d'un même concert se suivent ensuite toute la nuit.

Passons à présent en revue quelques idées qui pourraient nous permettre de traduire l'influence mutuelle des individus.

Comme nous recherchons une grande interdépendance, une tentative naturelle pourrait être d'utiliser des processus déterminantaux ou des processus permanents (voir [Ben09][Sos00]). Les premiers seront étudiés au chapitre 4, ils modélisent bien une certaine répulsion entre les utilisateurs, au contraire des seconds qui introduisent de l'attraction. Cependant, les processus permanents ne sont rien d'autre que des processus de Cox déguisés (voir [Ben09], section 4.9.). Cela signifie que nous devrions choisir une intensité sous-jacente qui varie dans le temps et qui présente une dynamique des zones

¹ c'est-à-dire une distance L^1 entre les distributions de probabilité.

d'attraction, ce qui est aussi difficile que de construire un processus directement avec les mêmes caractéristiques. On est donc renvoyé au problème initial. Quant aux processus déterminantaux, ils ne peuvent rendre compte de la formation des hotspots.

Une dernière classe de modèles de mobilité que nous pourrions considérer est celle des "processus spatiaux de vie et de mort" (*spatial birth and death processes*, voir [Gar06]). D'abord étudiés par Preston en 1975 (voir [Pre75]), ils peuvent modéliser efficacement certains motifs géométriques, pourvu que l'on choisisse leur distribution stationnaire convenablement. De plus, contrairement au *Random Waypoint Model*, ils présentent une vraie dynamique temporelle. Cependant, on sait qu'ils mènent à des processus de Gibbs, qui ne sont généralement pas calculables analytiquement.

Ce passage en revue nous conforte dans l'idée de trouver un nouveau modèle, qui soit à la fois calculable analytiquement et qui présente une vraie dynamique temporelle des hotspots.

3.2 Le modèle à moutons

Dans cette partie, nous cherchons à modéliser des agrégats d'individus dans des zones prédéfinies. L'idée conductrice est que les gens ont tendance à se diriger là où il y a déjà du monde, ce qui crée des rassemblements auto-alimentés.

3.2.1 Panorama du modèle

Divisons le territoire en K zones géographiques qui ont une identité propre. Ce sont des zones d'influence que l'on peut considérer comme en compétition pour attirer du public. Chaque zone échange donc des individus avec les autres. Notons que nous ne nous intéressons qu'aux mouvements macroscopiques entre zones, pas au détail des déplacements à l'intérieur de chacune d'elles. Dans chaque zone, un individu est supposé rester pendant une durée aléatoire exponentielle de paramètre μ , avant de décider ou non de changer.

Quitte à raffiner notre modèle par la suite, commençons par supposer que les individus sont totalement influencés par leur environnement, c'est-à-dire qu'ils se comportent comme des moutons. Ils décident de se déplacer d'une zone à l'autre proportionnellement au nombre d'individus déjà présents

(voir figure 3.3), y compris dans leur propre zone auquel cas ils décident de ne pas changer. En d'autres termes, si nous notons N le nombre total d'individus, n_1 le nombre d'individus dans la première zone, n_2 le nombre dans la deuxième, etc., un individu de la k -ème zone décide de gagner la k' -ème zone avec la probabilité $n_{k'}/N$. En revanche, il décide de rester avec la probabilité :

$$\frac{n_k}{N} = 1 - \sum_{k' \neq k} \frac{n_{k'}}{N}.$$

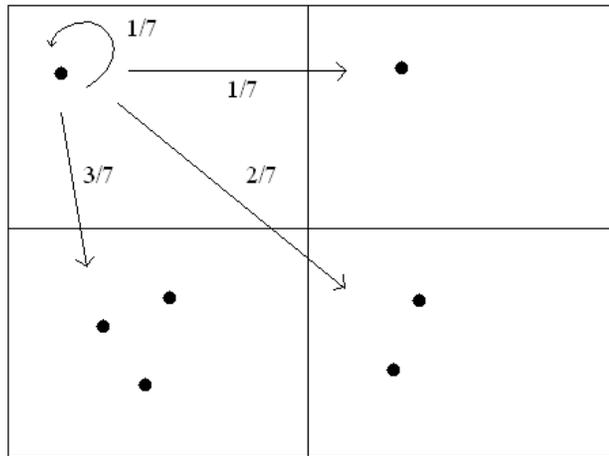


Fig. 3.3: Le modèle à moutons. Ici, $N = 7$ et $K = 4$. L'individu dans le coin en haut à gauche a quatre choix. Les quatre probabilités de décision sont proportionnelles au nombre d'individus.

Nous obtenons alors un processus de Markov à nombre fini d'états codés par (n_1, \dots, n_K) , dont les probabilités de transition sont données ci-dessus. On voit immédiatement qu'il y a K états absorbants, qui correspondent au cas où tous les individus sont regroupés dans une zone (les hotspots stables mentionnés en introduction). Ils sont donnés par :

$$N\mathbf{e}_k = (0, \dots, \underbrace{N}_{k\text{-ème position}}, \dots, 0), \quad 1 \leq k \leq K.$$

Dans un souci de clarté, nous allons d'abord exposer le cas $K = 2$, avant de donner le cas général.

Remarque 3.1. Nous sommes face à ce que la communauté d’algorithmique en calcul distribué appelle un ”problème de consensus” : plusieurs individus peuvent échanger de l’information entre eux et doivent se mettre d’accord sur une valeur commune (ici, le numéro de la zone finale dans laquelle le système est absorbé). En général, il s’agit de savoir si un tel consensus est accessible dans un temps fini, et si oui, de caractériser ce temps (espérance, voire variance). En particulier, dans le cas $K = 2$, nous avons un problème dit de consensus binaire qui a largement été étudié depuis les années 1980. Le résultat dépend notamment des conditions de synchronisation entre les individus et de leur capacité à ne pas faire d’erreur dans leurs décisions. On connaît à ce sujet un célèbre résultat d’impossibilité dû à Fischer, Lynch et Paterson (voir [Fis85]). On trouvera d’autres études intéressantes en [Per09], [Dra10] ou [Mat11].

3.2.2 Le problème à deux zones

Dans cette section nous choisissons $K = 2$. En quelque sorte, notre modèle pourrait être qualifié ”d’anti-Ehrenfest”, d’après le modèle à deux urnes bien connu (voir [Ehr07]).

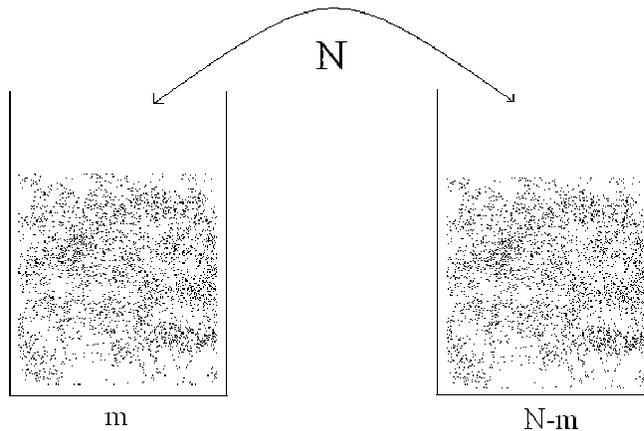


Fig. 3.4: notre modèle à deux zones.

L’état du système à l’instant t est noté :

$$\mathbf{n}(t) = (n_1(t), n_2(t)).$$

Pour plus de simplicité dans les calculs, on se ramène à une seule variable $m(t) = n_1(t)$, l'autre étant donnée par $N - m(t) = n_2(t)$.

Chaque individu reste dans sa zone pendant une durée aléatoire exponentielle de paramètre μ , et tous les temps d'attente sont supposés indépendants. Dans la zone 1, on a donc m horloges exponentielles de paramètre μ en compétition, dont la superposition est une horloge exponentielle de paramètre $m\mu$. Lorsque l'horloge sonne, un individu décide de changer de zone avec la probabilité $(N - m)/N$. Ainsi, la durée jusqu'au prochain changement de zone est une variable exponentielle de paramètre $m\mu(N - m)/N$. La manière la plus rapide de s'en convaincre est d'invoquer la propriété de "*p-thinning*" d'un processus de Poisson (voir [Bac10]). De même, les changements de la zone 2 à la zone 1 sont marqués par une horloge exponentielle de paramètre $(N - m)\mu m/N$. On en déduit les taux de transition de la figure 3.5².

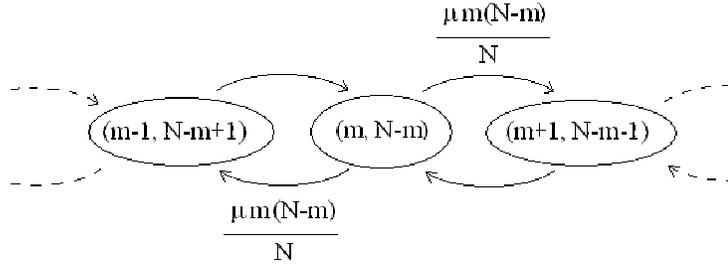


Fig. 3.5: transitions markoviennes du système à deux zones.

Si on indentifie l'état $(m, N - m)$ à l'entier m , le générateur infinitésimal est alors la matrice $(N + 1) \times (N + 1)$ suivante :

² En fait, pour être tout à fait exact, l'intervalle $[T_1, T'_1]$ entre deux transitions $1 \rightarrow 2$ n'est pas une exponentielle de paramètre $m\mu(N - m)/N$, car il peut se produire une transition $2 \rightarrow 1$ entretemps (ce qu'on notera $T_2 \in]T_1, T'_1[$). Dans ce cas, pendant $[T_2, T'_1]$ la probabilité de choisir la zone 2 lorsqu'on est dans la zone 1 passera de $(N - m)/N$ à $(N - m - 1)/N$. Cela dit, nous ne sommes intéressés par le comportement du système que pendant $[T_1, T_2]$, et sur cette durée, le système est équivalent à la concurrence de deux horloges exponentielles indépendantes de paramètre $\mu m(N - m)/N$. Comme T_2 est un temps d'arrêt, les transitions sont bien celles indiquées sur la figure 3.5.

$$Q = \begin{pmatrix} 0 & 0 & & & 0 \\ \ddots & \ddots & \ddots & & \\ & \frac{\mu m(N-m)}{N} & \frac{-2\mu m(N-m)}{N} & \frac{\mu m(N-m)}{N} & \\ & & \ddots & \ddots & \ddots \\ 0 & & & 0 & 0 \end{pmatrix}. \quad (3.1)$$

Pour caractériser la mobilité entre les deux zones, intéressons-nous à τ , le premier instant où l'on atteint un état absorbant :

$$\tau = \inf\{t/\mathbf{n}(t) = (0, N) \text{ ou } (N, 0)\}.$$

L'espérance de τ dépend bien sûr de l'état initial. Si le système démarre en $(m, N - m)$, notons $h_m = \mathbb{E}[\tau | \mathbf{n}(0) = (m, N - m)]$ le temps moyen d'atteinte d'un état absorbant. On a alors la

Proposition 3.1.

$$h_m = \frac{1}{\mu} \left(\sum_{0 \leq i < m} \frac{N - m}{N - i} + \sum_{0 \leq i < N - m} \frac{m}{N - i} \right).$$

Démonstration - On sait (voir [Nor98], théorème 3.3.3. sur les *hitting times*) que le vecteur $(h_m)_m$ est la solution minimale de l'équation :

$$\begin{cases} h_0 = h_N = 0 \\ \forall m \notin \{0, N\}, \quad \sum_{0 < i < N} q_{mi} h_i = -1. \end{cases} \quad (3.2)$$

La deuxième ligne peut donc être réécrite comme :

$$\begin{cases} 2h_1 - h_2 = \frac{N}{\mu(N-1)} \\ 2h_{N-1} - h_{N-2} = \frac{N}{\mu(N-1)} \\ 2h_m - h_{m-1} - h_{m+1} = \frac{N}{\mu i(N-i)} \quad \text{sinon.} \end{cases}$$

On est donc ramené à inverser la matrice $(N-1) \times (N-1)$ suivante :

$$U = \begin{pmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots \\ & & & -1 & 2 & -1 \\ 0 & & & & -1 & 2 \end{pmatrix}.$$

On vérifie facilement que $U^{-1} = \frac{1}{N} \cdot V$, où V est donnée par :

$$v_{ij} = \begin{cases} j(N-i) & \text{si } i \geq j \\ i(N-j) & \text{si } i \leq j. \end{cases}$$

Ainsi :

$$\begin{pmatrix} h_1 \\ \vdots \\ \vdots \\ h_{N-1} \end{pmatrix} = \frac{1}{N} \cdot V \cdot \frac{N}{\mu} \cdot \begin{pmatrix} 1/N-1 \\ 1/2(N-2) \\ \vdots \\ 1/N-1 \end{pmatrix},$$

et :

$$\begin{aligned} h_m &= \frac{1}{\mu} \sum_i \frac{v_{mi}}{i(N-i)} \\ &= \frac{1}{\mu} \sum_{0 \leq i < m} \frac{N-m}{N-i} + \frac{1}{\mu} \sum_{m \leq i < N} \frac{m}{i} \\ &= \frac{1}{\mu} \sum_{0 \leq i < m} \frac{N-m}{N-i} - \frac{1}{\mu} \cdot \frac{N-m}{N} + \frac{1}{\mu} \sum_{0 \leq i \leq N-m} \frac{m}{N-i} \\ &= \frac{1}{\mu} \sum_{0 \leq i < m} \frac{N-m}{N-i} - \frac{1}{\mu} \cdot \frac{N-m}{N} + \frac{1}{\mu} \sum_{0 \leq i < N-m} \frac{m}{N-i} + \frac{1}{\mu} \left(\frac{m}{m} - \frac{m}{N} \right) \\ &= \frac{1}{\mu} \sum_{0 \leq i < m} \frac{N-m}{N-i} + \frac{1}{\mu} \sum_{0 \leq i < N-m} \frac{m}{N-i} \end{aligned}$$

ce qui achève la preuve. ■

Typiquement, N correspond au nombre d'individus dans une ville, et peut donc être très grand. Aussi il n'est pas inutile de trouver un équivalent plus simple de h_m . Si nous notons $x = m/N$, x devient asymptotiquement une variable continue. On a alors la

Proposition 3.2. h_m est bien approché par :

$$h_m \approx -\frac{N}{\mu} \left((1-x) \log(1-x) + x \log(x) \right). \quad (3.3)$$

Pour connaître la marge d'erreur, on sait que la différence :

$$r_m = h_m + \frac{N}{\mu} \left((1-x) \log(1-x) + x \log(x) \right)$$

est uniformément encadrée par :

$$-\frac{2}{\mu} \leq r_m \leq 0.$$

La figure 3.6 compare la courbe approximante avec la courbe réelle.

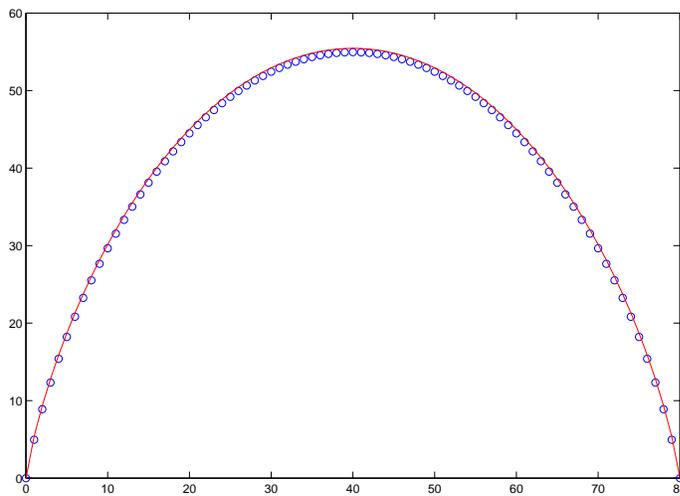


Fig. 3.6: Estimation asymptotique de h_m (en min). $N = 80$, $\mu = 1 \text{ min}^{-1}$. Les cercles représentent h_m , tandis que la ligne continue représente l'estimateur.

Démonstration - Si $m = 0$ ou N , le résultat est encore vrai à condition d'effectuer un prolongement par continuité en posant : $0 \log(0) = 0$.

Sinon, écrivons :

$$\mu h_m = N(1-x) \sum_{0 \leq i < m} \frac{1}{1-i/N} \cdot \frac{1}{N} + Nx \sum_{0 \leq i < N-m} \frac{1}{1-i/N} \cdot \frac{1}{N}$$

Chacune de ces sommes peut être interprétée comme une somme de Riemann. Travaillons sur la première somme par exemple. Par croissance de la fonction $t \mapsto 1/(1-t)$:

$$\int_{-1/N}^{(m-1)/N} \frac{1}{1-t} dt \leq \sum_{0 \leq i < m} \frac{1}{1-i/N} \cdot \frac{1}{N} \leq \int_0^{m/N} \frac{1}{1-t} dt.$$

Comme $m \neq N$, on remarque que $\int_0^{m/N} \frac{1}{1-t} dt$ est bien définie. D'où :

$$\log\left(1 + \frac{1}{N}\right) - \log\left(1 + \frac{1}{N} - x\right) \leq \sum_{0 \leq i < m} \frac{1}{1-i/N} \cdot \frac{1}{N} \leq -\log(1-x).$$

Et *a fortiori* :

$$-\log\left(1 + \frac{1}{N} - x\right) \leq \sum_{0 \leq i < m} \frac{1}{1-i/N} \cdot \frac{1}{N} \leq -\log(1-x).$$

Par convexité de la fonction $y \mapsto -\log(1-x+y)$:

$$-\log(1-x) - \frac{1}{N} \cdot \frac{1}{1-x} \leq \sum_{0 \leq i < m} \frac{1}{1-i/N} \cdot \frac{1}{N} \leq -\log(1-x)$$

En multipliant par le facteur $N(1-x)$ devant la somme, et en effectuant le même travail sur l'autre somme (quitte à changer m en $N-m$), on obtient le résultat demandé. ■

Remarque 3.2. Il s'agit d'une "bonne approximation" au sens où lorsque N tend vers l'infini, elle fournit un équivalent de h_m quel que soit m . En effet, même lorsque h_m est minimal, c'est-à-dire lorsque $m = 1$ ou $N-1$, on a :

$$-\frac{N}{\mu}(x \log(x) - (1-x) \log(1-x)) \sim -\frac{N}{\mu} \left(\frac{1}{N} \log\left(\frac{1}{N}\right) \right) = \frac{\log(N)}{\mu} \rightarrow \infty,$$

alors que r_m reste borné.

Jusqu'ici, nous avons obtenu une formule fermée pour le temps moyen d'absorption ainsi qu'une estimation asymptotique. On sait aussi qu'un des deux états absorbants sera presque sûrement atteint dans un temps fini. Maintenant, nous aimerions connaître la probabilité d'être absorbé en $(N, 0)$ plutôt qu'en $(0, N)$, toujours conditionnellement au point de départ. Soit p_m cette probabilité. On a la

Proposition 3.3. *La probabilité d'être absorbé en $(N, 0)$ vaut :*

$$p_m = \frac{m}{N}.$$

Démonstration - En utilisant l'équation (3.1), on obtient les probabilités de transition de la chaîne de Markov sous-jacente (voir figure 3.7).

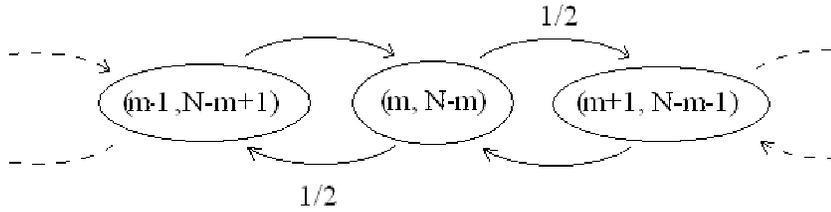


Fig. 3.7: La chaîne de Markov induite par le processus.

En conditionnant par rapport à T_1 , le premier instant de transition (qui est un temps d'arrêt), et en utilisant la propriété de Markov fort, nous obtenons pour $m \notin \{0, N\}$:

$$\begin{aligned} p_m &= \mathbb{P}(n_1(\tau) = 0 | n_1(0) = m) \\ &= \mathbb{P}(n_1(\tau) = 0 | n_1(0) = m, n_1(T_1) = m + 1) \times \mathbb{P}(n_1(T_1) = m + 1) \\ &\quad + \mathbb{P}(n_1(\tau) = 0 | n_1(0) = m, n_1(T_1) = m - 1) \times \mathbb{P}(n_1(T_1) = m - 1) \\ &= \frac{1}{2} p_{m+1} + \frac{1}{2} p_{m-1} \\ p_{m+1} &= 2p_m - p_{m-1}. \end{aligned}$$

Comme de plus $p_0 = 0$, une récurrence immédiate donne le résultat. ■

3.2.3 Le problème à K zones

Tous les résultats de la section précédente sont généralisables, c'est ce que nous allons voir à présent. L'état du système est maintenant décrit par un K -uplet :

$$\mathbf{n} = (n_1, n_2, \dots, n_K),$$

avec $\sum n_k = N$. Ainsi, l'espace des états possibles est une variété de dimension $K - 1$. Une transition de la k -ème zone à la k' -ème zone est donnée par :

$$\mathbf{n} \mapsto \mathbf{n}' = \mathbf{n} + \mathbf{e}_{k'} - \mathbf{e}_k.$$

Il s'agit ni plus ni moins d'un "opérateur de migration" au sens de Kelly (voir [Kel79], paragraphe 2.3.). Au maximum, il y a $K(K - 1)$ transitions possibles, et le taux de transition de $(k \rightarrow k')$ vaut :

$$q_{\mathbf{n}\mathbf{n}'} = \frac{\mu n_k n_{k'}}{N}.$$

À présent, voici une CNS bien commode quoique triviale pour caractériser les transitions :

Lemme 3.4. $\mathbf{n} \mapsto \mathbf{n}'$ a un sens si et seulement si $n_k \neq 0$.

Démonstration - Le sens direct est immédiat. Et réciproquement, si $n_k \neq 0$, alors nécessairement $n_{k'} < N$ car $\sum n_l = N$, et la transition est donc bien définie. Notons que si $n_{k'} = 0$, la transition, bien qu'ayant un sens, est de probabilité nulle. ■

Avec un peu plus de difficulté, on peut généraliser la proposition 3.1, et on obtient le

Théorème 3.5. *Lorsque le système démarre en \mathbf{n} , le temps moyen d'absorption est donné par :*

$$h_{\mathbf{n}} = \frac{1}{\mu} \sum_{1 \leq k \leq K} \sum_{0 \leq i < n_k} \frac{N - n_k}{N - i}. \quad (3.4)$$

Démonstration - Nous cherchons maintenant une solution de l'équation (3.2) en K dimensions. Partons d'abord de l'expression (3.4), et prouvons qu'elle vérifie l'équation.

Première remarque : si l'un des n_k est nul, on peut faire abstraction de la zone k sans modifier la somme dans (3.4), puisque la somme indexée sur $0 \leq i < n_k$ est vide. D'autre part, cela n'altère pas la vérification de (3.2), puisque les transitions vers des zones vides sont interdites. Ainsi on peut se ramener sans perte de généralité au cas où tous les n_k sont non nuls.

Si $K = 1$, on est dans un état absorbant et la formule est vraie. Sinon, introduisons $\epsilon_{\mathbf{n}\mathbf{n}'}$ = $h_{\mathbf{n}'} - h_{\mathbf{n}}$, où \mathbf{n} et \mathbf{n}' sont donnés par la transition de la zone k à la zone k' . On a :

$$\forall \mathbf{n}, \quad \sum_{\mathbf{n}'} q_{\mathbf{n}\mathbf{n}'} h_{\mathbf{n}'} = h_{\mathbf{n}} \underbrace{\sum_{\mathbf{n}'} q_{\mathbf{n}\mathbf{n}'}}_{=0} + \sum_{\mathbf{n}'} q_{\mathbf{n}\mathbf{n}'} \epsilon_{\mathbf{n}\mathbf{n}'} = \sum_{\mathbf{n}' \neq \mathbf{n}} q_{\mathbf{n}\mathbf{n}'} \epsilon_{\mathbf{n}\mathbf{n}'},$$

et :

$$\begin{aligned} \mu h_{\mathbf{n}'} &= \sum_{l \neq k, k'} \sum_{0 \leq i < n_l} \frac{N - n_l}{N - i} + \sum_{0 \leq i < n_k - 1} \frac{N - n_k + 1}{N - i} + \sum_{0 \leq i < n_{k'} + 1} \frac{N - n_{k'} - 1}{N - i} \\ &= \sum_l \sum_{0 \leq i < n_l} \frac{N - n_l}{N - i} - \frac{N - n_k}{N - n_k + 1} + \frac{N - n_{k'}}{N - n_{k'}} \\ &\quad + \sum_{0 \leq i < n_k - 1} \frac{1}{N - i} - \sum_{0 \leq i < n_{k'} + 1} \frac{1}{N - i} \\ &= \mu h_{\mathbf{n}} + \frac{1}{N - n_k + 1} + \sum_{0 \leq i < n_k - 1} \frac{1}{N - i} - \sum_{0 \leq i < n_{k'} + 1} \frac{1}{N - i}, \end{aligned}$$

si bien que :

$$\epsilon_{\mathbf{n}\mathbf{n}'} = \frac{1}{\mu} \left(\sum_{0 \leq i < n_k} \frac{1}{N - i} - \sum_{0 \leq i < n_{k'}} \frac{1}{N - i} \right).$$

Notons que la deuxième somme est bien définie, car comme $n_k \neq 0$ et $K > 1$, alors $n_{k'} < N$ et le dénominateur $N - i$ ne peut pas être nul. Maintenant, nous pouvons prouver que (3.4) vérifie (3.2).

On a :

$$\begin{aligned}
\sum_{\mathbf{n}' \neq \mathbf{n}} q_{\mathbf{nn}'} \epsilon_{\mathbf{nn}'} &= \sum_{(k,k')}^{\neq} \frac{n_k n_{k'}}{N} \left(\sum_{0 \leq i < n_k} \frac{1}{N-i} - \sum_{0 \leq i < n_{k'}} \frac{1}{N-i} \right) \\
&= \underbrace{\sum_{(k,k')}^{\neq} \frac{n_k n_{k'}}{N} \left(\sum_{0 \leq i < n_k} \frac{1}{N-i} - \sum_{0 \leq i < n_{k'}} \frac{1}{N-i} \right)}_{0 \text{ par symétrie}} \\
&\quad - \sum_{(k,k')}^{\neq} \frac{n_k n_{k'}}{N} \cdot \frac{1}{N-n_{k'}} \\
&= - \sum_{k'} \frac{n_{k'}}{N(N-n_{k'})} \cdot (N-n_{k'}) \\
&= -1.
\end{aligned}$$

Il reste à démontrer que dans notre cas, il n'existe qu'une seule solution à (3.2). Autrement dit, nous voulons montrer que la seule solution de :

$$\begin{cases} \forall \mathbf{n} \text{ absorbant, } h_{\mathbf{n}} = 0 \\ \forall \mathbf{n} \text{ non absorbant, } \sum_{\mathbf{n}' \text{ non absorbant}} q_{\mathbf{nn}'} h_{\mathbf{n}'} = 0 \end{cases} \quad (3.5)$$

est le vecteur nul.

Conformément à l'usage, notons $q_{\mathbf{nn}} = -\lambda_{\mathbf{n}}$, et $q_{\mathbf{nn}'} = \lambda_{\mathbf{n}} p_{\mathbf{nn}'}$ si $\mathbf{n} \neq \mathbf{n}'$ (où les $p_{\mathbf{nn}'}$ sont les probabilités de transition). S'il existe une transition réalisable $\mathbf{n} \rightarrow \mathbf{n}'$, alors $p_{\mathbf{nn}'} > 0$, et sinon $p_{\mathbf{nn}'} = 0$.

Comme $h_{\mathbf{n}'} = 0$ si \mathbf{n}' est absorbant, la deuxième ligne de (3.5) peut se réécrire :

$$\begin{aligned}
\sum_{\mathbf{n}'} q_{\mathbf{nn}'} h_{\mathbf{n}'} &= 0 \\
\sum_{\substack{\mathbf{n}' \neq \mathbf{n} \\ \mathbf{n} \rightarrow \mathbf{n}' \text{ existe}}} p_{\mathbf{nn}'} h_{\mathbf{n}'} &= h_{\mathbf{n}}.
\end{aligned}$$

Considérons \mathbf{n} tel que $h_{\mathbf{n}}$ soit maximal. Alors, par convexité, et comme les coefficients $p_{\mathbf{nn}'}$ sont tous strictement positifs, on voit que nécessairement tous les $h_{\mathbf{n}'}$ tels que $\mathbf{n} \rightarrow \mathbf{n}'$ existe avec une probabilité non nulle sont égaux

à $h_{\mathbf{n}}$. De proche en proche, comme on peut atteindre au moins un état absorbant depuis de \mathbf{n} , cela montre que $h_{\mathbf{n}} = 0$.

Enfin, en changeant le vecteur \mathbf{h} en $-\mathbf{h}$ et en appliquant le résultat précédent, on voit que le minimum des $h_{\mathbf{n}}$ est lui aussi nul, ce qui achève la preuve. ■

Sur la figure 3.8, on trouvera une représentation de $h_{\mathbf{n}}$ dans le cas $K = 3$. On peut immédiatement généraliser la proposition 3.2. Si on considère le vecteur $\mathbf{x} = (\frac{n_1}{N}, \dots, \frac{n_K}{N})$, on a le

Théorème 3.6. *Une bonne approximation de $h_{\mathbf{n}}$ est donnée par :*

$$h(\mathbf{x}) \approx -\frac{N}{\mu(K-1)} \sum_k (1-x_k) \log(1-x_k) \quad (3.6)$$

et la différence est uniformément encadrée par :

$$-\frac{K}{\mu} \leq r_{\mathbf{n}} \leq 0.$$

De même, nous pouvons aussi calculer la probabilité d'atteindre tel ou tel état absorbant en partant de \mathbf{n} , disons par exemple $N\mathbf{e}_K$, probabilité que nous noterons $p_{\mathbf{n}}$:

Théorème 3.7. *La probabilité d'atteindre $N\mathbf{e}_K$ en partant de \mathbf{n} vaut :*

$$p_{\mathbf{n}} = \frac{n_K}{N}.$$

Démonstration - On utilise un joli argument d'associativité : on peut considérer un système à deux zones, dont la première zone serait la réunion des $K-1$ premières zones du système initial, et la deuxième zone serait la K -ème zone du système initial. Alors ce nouveau système à deux zones se comporte exactement selon la même règle, et on peut lui appliquer la proposition 3.3, ce qui donne immédiatement le résultat. ■

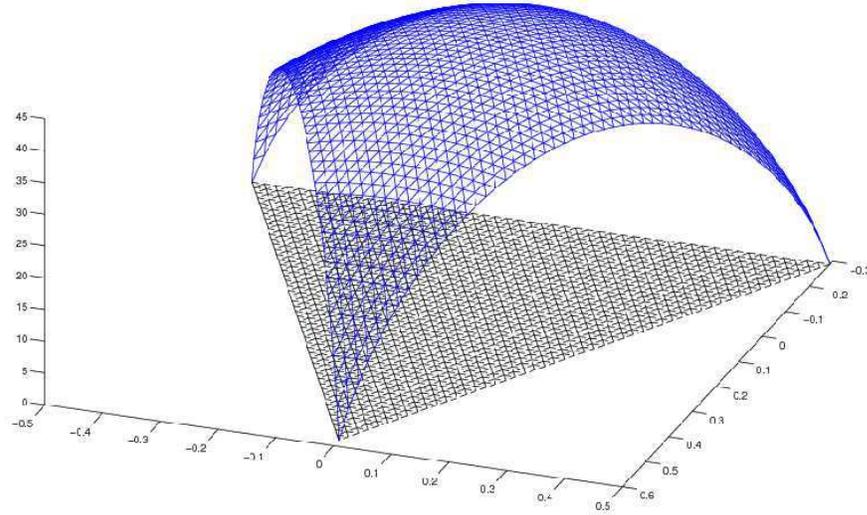


Fig. 3.8: Temps d'absorption moyen (en minutes) dans le cas de 3 zones. $N = 50$, $\mu = 1 \text{ min}^{-1}$. L'axe des z représente le temps, tandis que l'état initial est représenté sur le plan Oxy en coordonnées barycentriques (par rapport à chacun des trois états absorbants).

3.3 Fortes têtes et dynamique des hotspots

Jusqu'ici, nous avons pu modéliser les regroupements d'individus dans une zone donnée, mais ce n'est qu'un aspect de ce que nous observons dans la réalité des réseaux urbains denses. Certes, la zone de regroupement n'est pas définie à l'avance, ce qui est une bonne chose, mais le modèle à moutons possède un inconvénient majeur, qui est la présence d'états absorbants.

À présent, nous proposons un enrichissement du modèle qui nous permet de représenter la désagrégation ultérieure des rassemblements. Pour cela, nous introduisons une petite quantité de "fortes têtes" qui ne se laissent pas influencer par les autres. Lors d'une transition, quand un mouton choisissait sa nouvelle zone proportionnellement aux individus déjà présents, une forte tête choisit sa zone de manière indépendante et uniforme. Plus prosaïquement, nous réintroduisons une petite composante de *Random Waypoint* (voir introduction) dans notre modèle. Grâce aux fortes têtes, il n'y a plus un seul état absorbant, et même lorsque tous les individus se trouvent regroupés dans une zone, tôt ou tard, une forte tête en sortira et on peut espérer qu'elle entraînera des moutons à sa suite.

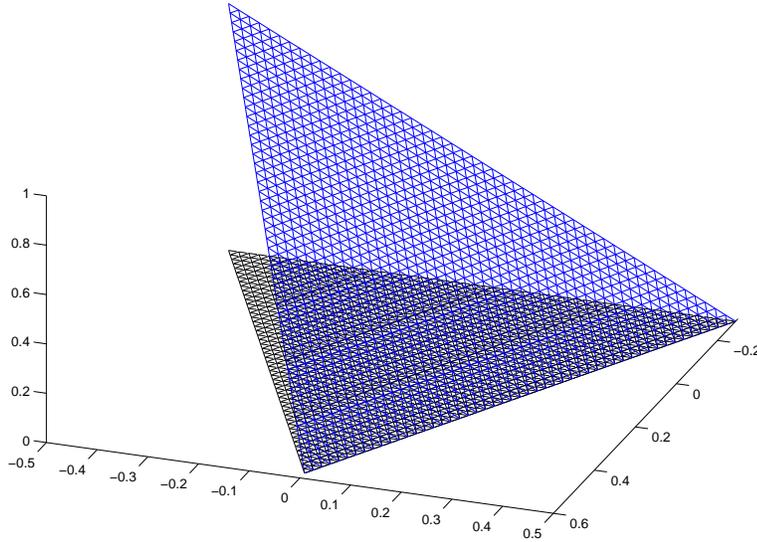


Fig. 3.9: Représentation de p_n pour $N = 50$ et $K = 3$.

Nous choisissons d'introduire les fortes têtes comme suit³ : on introduit un paramètre $\alpha \in]0, 1[$, le taux de fortes têtes, et un individu, au moment de choisir sa zone, se comporte avec la probabilité α comme une forte tête, et avec la probabilité $1 - \alpha$ comme un mouton.

³ on pourrait aussi marquer une fois pour toutes certains individus comme fortes têtes, et ils ne changeraient pas d'état au cours du temps. Mais cette manière de faire comporterait deux inconvénients :

- d'une part, on ne pourrait pas atteindre des proportions de fortes têtes strictement comprises entre 0 et $1/N$. Plus généralement, on ne pourrait atteindre que des proportions de la forme k/N ($k \in \llbracket 0, N \rrbracket$), et le spectre des taux modélisables serait donc discret.
- d'autre part, cela introduirait un biais : il pourrait arriver qu'une zone se vide anormalement rapidement parce qu'elle regroupe à elle seule toutes les fortes têtes. Or une simple considération sociologique montre que les fortes têtes n'ont pas de raison d'être toutes regroupées au même endroit, sauf exception type camp d'internement dans un régime dictatorial - et Paris n'en est pas là, même de l'avis des dissidents les plus enragés. Nous préférons donc faire en sorte qu'à chaque nouveau tirage, les fortes têtes soient uniformément réparties (en loi) sur le territoire.

Sans compter que les calculs s'en trouveraient grandement compliqués, car il faudrait prendre en compte le nombre de fortes têtes dans chaque zone pour le codage des états, et ces derniers deviendraient très nombreux.

Avec les mêmes considérations que précédemment sur les horloges exponentielles en compétition, on montre que le système est encore un processus de Markov, dont les taux de transition sont donnés par :

$$q_{\mathbf{nn}'} = \mu n_k \left(\alpha \cdot \frac{1}{K} + (1 - \alpha) \frac{n_{k'}}{N} \right).$$

3.3.1 Temps moyen d'absorption

À vrai dire, il s'agit d'un léger abus de langage, car les états $N\mathbf{e}_k$ ne sont plus absorbants. On les appellera plutôt "états extrémaux". L'équation (3.2) permet de calculer numériquement $h_{\mathbf{n}}$ par une simple inversion de matrice, et le même raisonnement que pour le théorème 3.5 nous garantit que la solution est unique, et donc que la matrice du système est inversible.

Sur la figure 3.10, nous avons tracé l'évolution de $h_{\mathbf{n}}$ en fonction de α . Lorsque α grandit, on remarque deux choses :

- $h_{\mathbf{n}}$ grandit avec α . En fait, on peut même caractériser cette croissance. Par exemple, si on s'intéresse au $h_{\mathbf{n}}$ maximal (qui correspond au cas où chaque zone contient approximativement N/K individus), la figure 3.11 montre que sa croissance est exponentielle.
- sur la figure 3.10, lorsque α augmente (en d'autres termes, on sature progressivement le système de fortes têtes), la surface prend une forme de plateau (c'est déjà le cas pour $\alpha = 8\%$). Cela signifie qu'à part de toutes petites régions autour des états extrémaux, tous les états sont plus ou moins équivalents (mais pas équiprobables, voir section 3.3.2) car ils fournissent des $h_{\mathbf{n}}$ très proches. Même si on se trouve proche d'un état extrémal, le système fera une longue excursion avant de rejoindre un état extrémal (pas forcément le même). De plus, le fait que l'altitude du plateau grandisse beaucoup montre que les états extrémaux sont de plus en plus difficiles à atteindre (plusieurs dizaines d'heures lorsque pourtant $\mu = 1 \text{ min}^{-1}$): l'effet de saturation se fait sentir.

Ce qui est intéressant, c'est que cela se produit même pour des valeurs de α relativement faibles, en tout cas bien inférieures à 50% (pour plus de détail, voir la section 3.3.2). Même si cela ne constitue pas encore une preuve, on possède déjà l'intuition que les phénomènes de

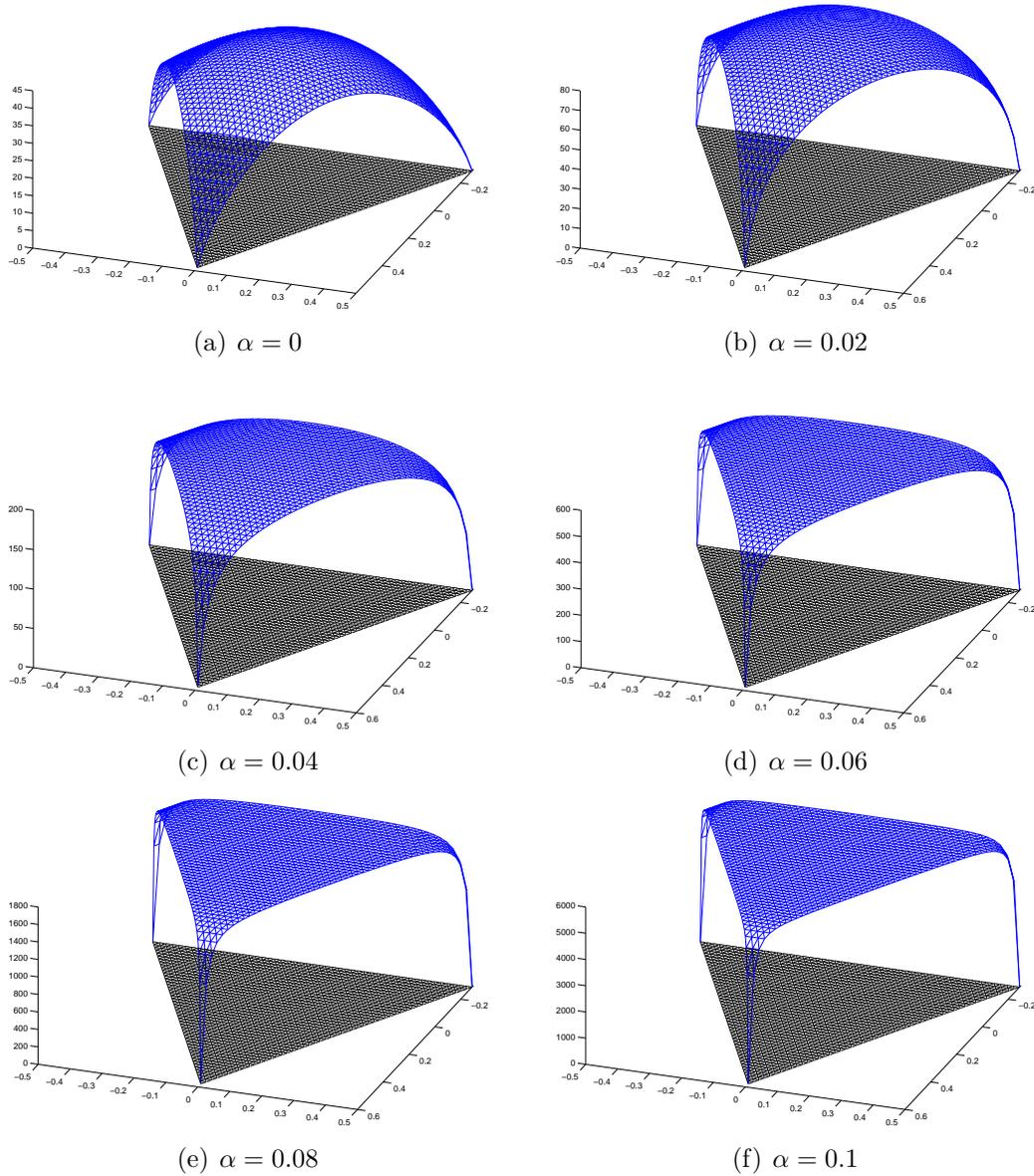


Fig. 3.10: Temps moyen d'absorption (en min) pour $N = 50$, $K = 3$ et $\mu = 1 \text{ min}^{-1}$. Le taux de fortes têtes α varie de 0 à 0.1. Bien noter que l'échelle varie sur l'axe des z .

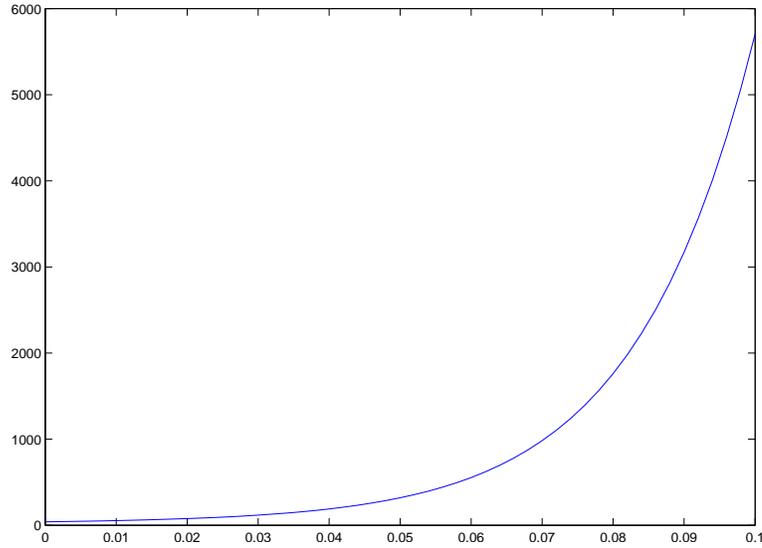


Fig. 3.11: Temps moyen d'absorption maximal (en minutes) en fonction de α . $N = 50$, $K = 3$ et $\mu = 1 \text{ min}^{-1}$.

désagrégation pourront facilement se produire (pour une étude systématique, voir section 3.3.3).

3.3.2 Transition de phase

On souhaite caractériser plus précisément cet effet de saturation, et savoir s'il existe un seuil critique à partir duquel il se fait sentir. Pour cela, on calcule la mesure invariante π de notre processus de Markov, et on montre qu'il existe une valeur de α telle que tous les états soient équiprobables.

π est donné par l'équation $\pi Q = 0$, où on rappelle que Q est la matrice de transition. On montre alors le

Théorème 3.8. *La mesure invariante est donnée par :*

$$\pi(\mathbf{n}) = A \prod_{k=1}^K \frac{1}{n_k!} \prod_{i=0}^{n_k-1} (N\beta + i),$$

où $\beta = \frac{\alpha}{K(1-\alpha)}$, et A est une constante de normalisation qui ne dépend pas de \mathbf{n} .

Démonstration - Nous allons expliquer progressivement comment arriver à l'équation. Pour commencer, supposons que $K = 2$. Comme on est face à un processus de naissance et de mort, qui est donc réversible, on sait qu'on peut se restreindre aux équations de balance locale :

$$\forall m \in \llbracket 0, N-1 \rrbracket, q_{m,m+1}\pi(m) = q_{m+1,m}\pi(m+1).$$

On en déduit immédiatement :

$$\pi(m) = \pi(0) \prod_{i=1}^m \frac{(N-i+1)(N\beta+i-1)}{i(N\beta+N-i)}.$$

Afin de généraliser cette formule à l'ordre K , on aimerait la réécrire d'une manière symétrique en n_1 et n_2 . De fait, on a :

$$\pi(m) = \pi(0) \frac{\prod_{i=1}^{n_1} \frac{N\beta+i-1}{i} \prod_{i=1}^{n_2} \frac{N\beta+i-1}{i}}{\prod_{i=1}^N \frac{N\beta+i-1}{i}},$$

comme on peut s'en rendre compte en effectuant le changement de variable $i \mapsto N+1-i$ dans le produit de droite et dans celui du bas. Ainsi le théorème est prouvé dans le cas $K = 2$.

À partir de là, on peut raisonnablement espérer que la formule fonctionne aux ordres supérieurs. En fait, c'est assez immédiat. En effet, grâce à la forme produit de notre formule, nous savons que dans le cas général, les équations de balance locale continuent à être satisfaites, ce qui achève la preuve. ■

Nous donnons maintenant une estimation asymptotique de $\pi(\mathbf{n})$:

Théorème 3.9. $\pi(\mathbf{n})$ est donné par :

$$\pi(\mathbf{n}) = (1+r) \frac{AN^{-K/2}}{\beta^{K(N\beta-1/2)}} \prod_{k=1}^K \frac{f(\beta, x_k)}{f\left(\frac{1}{N}, x_k\right)},$$

où

$$f(\beta, x_k) = (\beta + x_k)^{N(\beta+x_k)-1/2}$$

et l'erreur relative r est dominée par :

$$e^{-K(1-\alpha)} - 1 \leq r \leq e^{\frac{K(1-\alpha)}{N\beta}} - 1.$$

Remarque 3.3. En anticipant sur la suite, on va voir que $N\beta$ est souvent de l'ordre de 1. Aussi, l'approximation ci-dessus ne sera réellement utile que pour des valeurs de α très proches de 1, c'est-à-dire dans un système complètement saturé. Sinon, on préférera utiliser la formule exacte.

Démonstration - Pour commencer, écrivons :

$$\log \left(\prod_{i=0}^{n_k-1} (N\beta + i) \right) = n_k \log(N) + \sum_{i=0}^{n_k-1} \log \left(\beta + \frac{i}{N} \right),$$

et approximations le deuxième terme par une intégrale :

$$\begin{aligned} & N \int_0^{x_k} \log(\beta + t) dt - \sum_{i=0}^{n_k-1} \log \left(\beta + \frac{i}{N} \right) \\ &= N \sum_{i=0}^{n_k-1} \int_{\frac{i}{N}}^{\frac{i+1}{N}} \left(\log(\beta + t) - \log \left(\beta + \frac{i}{N} \right) \right) dt. \end{aligned}$$

Or d'après l'inégalité des accroissements finis :

$$\left(t - \frac{i}{N} \right) \frac{1}{\beta + \frac{i+1}{N}} \leq \log(\beta + t) - \log \left(\beta + \frac{i}{N} \right) \leq \left(t - \frac{i}{N} \right) \frac{1}{\beta + \frac{i}{N}}.$$

Ainsi, en intégrant sur $[\frac{i}{N}, \frac{i+1}{N}]$ et en sommant les intégrales :

$$\begin{aligned} \frac{1}{2} \sum_{i=1}^{n_k} \frac{1}{N\beta + i} &\leq N \int_0^{x_k} \log(\beta + t) dt - \sum_{i=0}^{n_k-1} \log \left(\beta + \frac{i}{N} \right) \\ &\leq \frac{1}{2} \sum_{i=0}^{n_k-1} \frac{1}{N\beta + i}. \end{aligned}$$

Approchons chacune des deux bornes par une intégrale. On peut écrire :

$$\begin{aligned} \frac{i}{N} \leq t \leq \frac{i+1}{N} &\Rightarrow \frac{1}{\beta + \frac{i}{N}} \geq \frac{1}{\beta + t} \geq \frac{1}{\beta + \frac{i+1}{N}} \\ &\Rightarrow \frac{1}{N\beta + i} \geq \int_{i/N}^{(i+1)/N} \frac{1}{\beta + t} dt \geq \frac{1}{N\beta + i + 1}. \end{aligned}$$

Pour $i \geq 1$, on a donc :

$$\int_{i/N}^{(i+1)/N} \frac{1}{\beta + t} dt \leq \frac{1}{N\beta + i} \leq \int_{(i-1)/N}^{i/N} \frac{1}{\beta + t} dt.$$

En particulier :

$$\frac{1}{2} \sum_{i=1}^{n_k} \frac{1}{N\beta + i} \geq \frac{1}{2} \int_{1/N}^{x_k+1/N} \frac{dt}{\beta + t} = \frac{1}{2} \log \left(1 + \frac{x_k}{\beta + \frac{1}{N}} \right)$$

d'où :

$$\sum_{i=0}^{n_k-1} \log \left(\beta + \frac{i}{N} \right) \leq N \int_0^{x_k} \log(\beta + t) dt - \frac{1}{2} \log \left(1 + \frac{x_k}{\beta + \frac{1}{N}} \right).$$

Par ailleurs :

$$\frac{1}{2} \sum_{i=0}^{n_k-1} \frac{1}{N\beta + i} \leq \frac{1}{N\beta} + \int_0^{x_k} \frac{dt}{\beta + t}$$

d'où :

$$\sum_{i=0}^{n_k-1} \log \left(\beta + \frac{i}{N} \right) \geq N \int_0^{x_k} \log(\beta + t) dt - \frac{1}{2} \log \left(1 + \frac{x_k}{\beta} \right).$$

Pour plus de simplicité, nous cherchons à comparer $\log \left(1 + \frac{x_k}{\beta+1/N} \right)$ à $\log \left(1 + \frac{x_k}{\beta} \right)$. En appliquant l'inégalité des accroissements finis à la fonction $u \mapsto \log \left(1 + \frac{x_k}{u} \right)$ sur l'intervalle $[\beta, \beta + \frac{1}{N}]$, on obtient :

$$\left| \log \left(1 + \frac{x_k}{\beta + 1/N} \right) - \log \left(1 + \frac{x_k}{\beta} \right) \right| \leq \frac{1}{N} \cdot \frac{x_k}{\beta(\beta + x_k)}.$$

On peut donc conclure :

$$\begin{aligned} N \int_0^{x_k} \log(\beta + t) dt - \frac{1}{2} \log \left(1 + \frac{x_k}{\beta} \right) &\leq \sum_{i=0}^{n_k-1} \log \left(\beta + \frac{i}{N} \right) \\ &\leq N \int_0^{x_k} \log(\beta + t) dt - \frac{1}{2} \log \left(1 + \frac{x_k}{\beta} \right) + \frac{x_k}{N\beta(\beta + x_k)}. \end{aligned}$$

Ainsi :

$$\prod_{i=0}^{n_k-1} (N\beta + i) = N^{n_k} \left(\frac{(\beta + x_k)^{(\beta+x_k)}}{\beta^\beta e^{x_k}} \right)^N \sqrt{\frac{\beta}{\beta + x_k}} \cdot (1 + r_k),$$

avec

$$0 \leq r_k \leq \exp \left(\frac{x_k}{N\beta(\beta + x_k)} \right) - 1.$$

Or ce calcul est parfaitement valable si l'on pose $\beta = \frac{1}{N}$. Dans ce cas, on a :

$$\begin{aligned} n_k! &= \prod_{i=0}^{n_k-1} (N\beta + i) \\ &= N^{n_k} \left(\frac{(1/N + x_k)^{(1/N+x_k)}}{(\frac{1}{N})^{1/N} e^{x_k}} \right)^N \sqrt{\frac{1/N}{1/N + x_k}} \cdot (1 + r'_k) \\ &= N^{n_k+1} (1/N + x_k)^{n_k+1} e^{-n_k} \cdot \sqrt{\frac{1}{1 + n_k}} \cdot (1 + r'_k) \\ &= (n_k + 1)^{n_k+1/2} e^{-n_k} (1 + r'_k), \end{aligned}$$

avec

$$0 \leq r'_k \leq \exp \left(\frac{n_k}{n_k + 1} \right) - 1.$$

Après avoir multiplié terme à terme les K approximations, on obtient l'équivalent annoncé. Reste à dominer l'erreur :

$$1 \leq \prod_k (1 + r_k) \leq \exp \left(\sum_k \frac{x_k}{N\beta(\beta + x_k)} \right).$$

Or :

$$\frac{1}{N\beta} \sum_k \frac{x_k}{\beta + x_k} = \frac{1}{N\beta} \left(K - \sum_k \frac{\beta}{\beta + x_k} \right),$$

et d'après l'inégalité entre moyenne arithmétique et moyenne harmonique :

$$\frac{\sum_k \frac{\beta}{\beta + x_k}}{K} \geq \frac{K}{\sum_k \frac{\beta + x_k}{\beta}} = \frac{K}{K + \frac{1}{\beta}} = \frac{K}{K + \frac{(1-\alpha)K}{\alpha}} = \alpha.$$

Ainsi :

$$\frac{1}{N\beta} \sum_k \frac{x_k}{\beta + x_k} \leq \frac{1}{N\beta} (K - K\alpha),$$

et on obtient :

$$1 \leq \prod_k (1 + r_k) \leq \exp \left(\frac{K(1 - \alpha)}{N\beta} \right).$$

De même, on a :

$$1 \leq \prod_k (1 + r'_k) \leq \exp (K(1 - \alpha)).$$

Le résultat s'en déduit. ■

Sur la figure 3.12, on a représenté la mesure invariante en fonction de α dans le cas $K = 2$. Comme nous l'avons déjà remarqué sur la figure 3.10, lorsque α grandit l'effet "fortes têtes" se fait de plus en plus sentir. À la limite, pour les grandes valeurs de α , la masse de la mesure invariante est concentrée sur les états centraux, c'est-à-dire ceux où il y a sensiblement le même nombre d'individus dans les deux zones.

Apparemment, il existe une valeur critique α_c où le profil de la courbe s'inverse. Les deux questions que nous nous posons sont les suivantes :

- au moment où la courbe s'inverse, passe-t-elle par un stade où elle est complètement plate ?
- si oui, ce résultat est-il généralisable à K zones ?

En fait, le théorème 3.8 montre immédiatement que la réponse est à chaque fois oui : il suffit de choisir $\beta = \frac{1}{N}$ pour obtenir une mesure uniforme. Ainsi on a le

Corollaire 3.10. *Pour la valeur critique de α suivante :*

$$\alpha_c = \frac{K}{N + K},$$

la mesure invariante est uniforme sur tous les états.

Ainsi :

- pour $\alpha < \alpha_c$, l'influence des moutons est encore notable. C'est le cas le plus intéressant, puisque on assiste à la formation de hotspots.

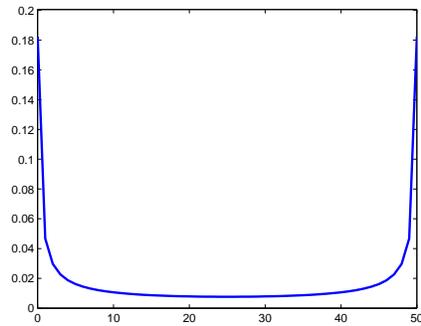
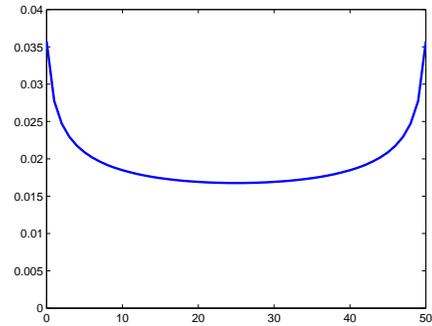
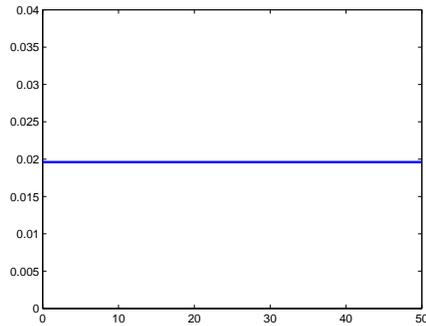
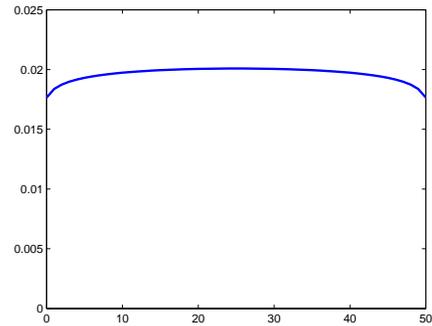
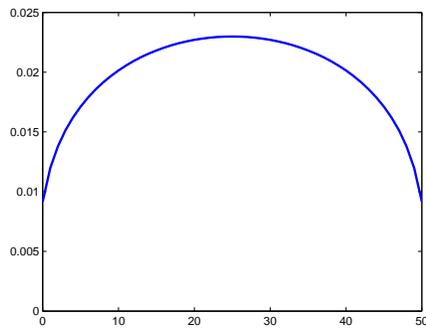
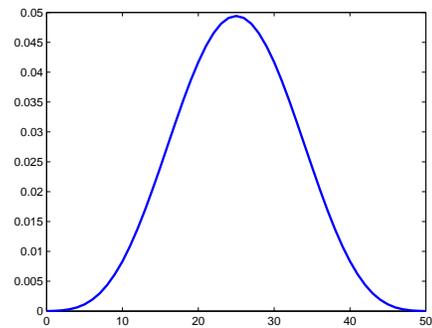
(a) $\alpha = 0.01$ (b) $\alpha = 0.03$ (c) $\alpha = 0.038$ (d) $\alpha = 0.04$ (e) $\alpha = 0.05$ (f) $\alpha = 0.2$

Fig. 3.12: Représentation de la mesure invariante pour $N = 50$ et $K = 2$. Lorsque α dépasse 0.038, le profil de la courbe s'inverse.

- pour $\alpha = \alpha_c$, tous les états sont équiprobables. Lorsque N est grand, comme $\alpha_c \sim \frac{K}{N}$, on obtient une conséquence assez surprenante : il suffit d'une seule forte tête par zone pour perturber tout le système.
- pour $\alpha > \alpha_c$, les moutons perdent toute influence. C'est dans ce cas que le théorème 3.9 se révèle le plus utile, mais ce n'est pas l'objet de notre étude ici.

Comme nous cherchons à modéliser la formation de hotspots, nous supposerons dans tout le chapitre que $\alpha < \alpha_c$.

3.3.3 Macro-états et dynamique des hotspots

Notre système est entièrement caractérisé par quatre paramètres, qui sont N , K , μ et α . Les trois premiers paramètres peuvent être directement mesurés, mais α non, car son sens est plus abstrait. Pourtant, il est de première importance de le déterminer. Une idée pourrait être de trouver une grandeur caractéristique du système, facilement mesurable et étroitement liée à α . C'est ce que nous proposons ici.

Dans le cas $\alpha < \alpha_c$, on peut observer aussi bien la formation de hotspots que leurs désagrégation. Définir un hotspot comme la totalité des individus dans une zone serait trop extrême, car un tel événement est extrêmement rare et il n'est pas réaliste de vouloir le mesurer. En revanche, on peut demander qu'une proportion significative de la population se retrouve dans une zone donnée. C'est ce que nous appellerons un "macro-état", à savoir la réunion de plusieurs états proches d'un état extrémal.

Introduisons le macro-état A_k défini par :

Définition 3.1. $A_k = \{\mathbf{n} : n_k > N_{\text{sup}}\}$, avec $N_{\text{sup}} < N$,

où N_{sup} est un nombre d'individu qualifié de "significatif". De même définissons :

Définition 3.2. $B_k = \{\mathbf{n} : n_k < N_{\text{inf}}\}$, avec $N_{\text{inf}} > 0$.

Alors une désagrégation sera simplement une transition de A_k à B_k . on dira aussi que la zone k a été "siphonnée". Plus précisément, pour k fixé, nous définissons le "temps de siphonnage" S_k comme la durée entre le premier

instant où on se trouve dans A_k et le premier instant ultérieur où on se trouve dans B_k ⁴:

Définition 3.3. Le temps de siphonnage de la zone k est donné par :

$$\begin{aligned} S_k &= \tau(A_k \rightarrow B_k) \\ &= \inf\{t/\mathbf{n}(t) \in B_k \text{ et } \exists u < t, \mathbf{n}(u) \in A_k\} - \inf\{u/\mathbf{n}(u) \in A_k\}. \end{aligned}$$

Le premier temps d'atteinte de A_k est un temps d'arrêt, donc par application de la propriété de Markov fort le futur du système est indépendant de tout ce qui s'est passé auparavant. Aussi, si on s'intéresse à la loi de S_k , on peut supposer sans perte de généralité que le système démarre déjà en A_k . On a alors le

Théorème 3.11. *Sachant que le système démarre en A_k , le temps de siphonnage moyen de la zone k est donné par :*

$$\mathbb{E}[S_k | \mathbf{n}(0) = \mathbf{n} \in A_k] = \frac{N}{\mu(1-\alpha)} \sum_{l=N_{\text{inf}}}^{n_k} \sum_{j=l}^N \frac{\prod_{l \leq i < j} \delta(i)}{j(N(K-1)\beta + N - j)},$$

où

$$\delta(i) = \frac{(N-i)(N\beta+i)}{i(N(K-1)\beta + N - i)}.$$

Démonstration - Comme d'habitude, commençons par le cas $K = 2$ et $k = 1$. Nous sommes dans l'état initial $(m, N - m)$, et nous cherchons le temps d'atteinte h_m de l'ensemble $\{\mathbf{n}/n_1 < N_{\text{inf}}\}$. Pour l'instant, on ne suppose pas forcément que $m > N_{\text{sup}}$, et on cherche une formule analytique de h_m (disons pour $N_{\text{inf}} \leq m \leq N$, puisque sinon $h_m = 0$).

Si nous introduisons $u_m = h_m - h_{m-1}$, une équation du même type que 3.2 nous fournit la récurrence suivante :

$$\begin{aligned} -1 &= q_{m,m+1}h_{m+1} + q_{m,m}h_m + q_{m,m-1}h_{m-1} \\ &= q_{m,m+1}h_{m+1} - (q_{m,m-1} + q_{m,m+1})h_m + q_{m,m-1}h_{m-1} \\ u_{m+1} &= \frac{q_{m,m-1}}{q_{m,m+1}}u_m - \frac{1}{q_{m,m+1}} \\ &= \frac{m(N\beta + N - m)}{(N - m)(N\beta + m)}u_m - \frac{N}{\mu(1-\alpha)(N - m)(N\beta + m)}. \end{aligned}$$

⁴ pourquoi ne pas plutôt mesurer la durée à partir du dernier instant où on quitte A_k avant de rejoindre B_k ? Parce que ce n'est pas un temps d'arrêt, ce qui rendrait plus compliqué l'usage de la théorie des processus de Markov.

En théorie, la récurrence est valable pour $N_{\text{inf}} + 1 \leq m \leq N - 1$, mais comme $N_{\text{inf}} > 0$ et $h_{N_{\text{inf}}-1} = 0$, elle est encore valable pour $m = N_{\text{inf}}$. Ensuite l'idée est d'introduire :

$$\delta(i) = \frac{(N-i)(N\beta+i)}{i(N\beta+N-i)} \quad \text{et} \quad v_m = \frac{u_m}{\prod_{i=m}^{N-1} \delta(i)}.$$

On obtient :

$$\begin{aligned} v_{m+1} &= \frac{u_{m+1}}{\prod_{i=m+1}^{N-1} \delta(i)} \\ &= \frac{u_m}{\delta(m) \prod_{i=m+1}^{N-1} \delta(i)} - \frac{N}{\mu(1-\alpha)(N-m)(N\beta+m) \prod_{i=m+1}^{N-1} \delta(i)} \\ &= v_m - \frac{N}{\mu(1-\alpha)} \cdot \frac{1}{m(N\beta+N-m) \prod_{i=m}^{N-1} \delta(i)}, \end{aligned}$$

si bien que pour $N_{\text{inf}} \leq m \leq N$:

$$v_m = v_N + \frac{N}{\mu(1-\alpha)} \sum_{j=m}^{N-1} \frac{1}{j(N\beta+N-j) \prod_{i=j}^{N-1} \delta(i)}.$$

Comme $q_{N,N}h_N + q_{N,N-1}h_{N-1} = -1$, on a aussi :

$$u_N = -\frac{1}{q_{N,N}} = \frac{2}{\mu\alpha N} = v_N.$$

Finalement :

$$\begin{aligned} v_m &= \frac{2}{\mu\alpha N} + \frac{N}{\mu(1-\alpha)} \sum_{j=m}^{N-1} \frac{1}{j(N\beta+N-j) \prod_{i=j}^{N-1} \delta(i)} \\ &= \frac{N}{\mu(1-\alpha)} \sum_{j=m}^N \frac{1}{j(N\beta+N-j) \prod_{i=j}^{N-1} \delta(i)}, \end{aligned}$$

et :

$$u_m = \frac{N}{\mu(1-\alpha)} \sum_{j=m}^N \frac{\prod_{m \leq i < j} \delta(i)}{j(N\beta+N-j)}.$$

Pour déterminer les h_m , il ne reste plus qu'à remarquer que $u_{N_{\text{inf}}} = h_{N_{\text{inf}}}$:

$$h_{N_{\text{inf}}} = \frac{N}{\mu(1-\alpha)} \sum_{j=N_{\text{inf}}}^N \frac{\prod_{N_{\text{inf}} \leq i < j} \delta(i)}{j(N\beta+N-j)},$$

et plus généralement :

$$h_m = \frac{N}{\mu(1-\alpha)} \sum_{l=N_{\text{inf}}}^m \sum_{j=l}^N \frac{\prod_{l \leq i < j} \delta(i)}{j(N\beta + N - j)}.$$

Intéressons-nous maintenant au cas général, et invoquons le même argument d'associativité que précédemment : il est parfaitement licite de considérer l'union des $K - 1$ zones différentes de la zone k comme une seule grande zone. Un mouton restera dedans avec la probabilité $\frac{\sum_{l \neq k} m_l}{N}$, alors qu'une forte tête y restera avec la probabilité $\frac{K-1}{K}$. C'est cette dernière probabilité qui constitue la seule adaptation que nous ayons à faire. En l'occurrence, nous obtenons la nouvelle récurrence suivante :

$$u_{m+1} = \frac{m(N\beta' + N - m)}{(N - m)(N\beta + m)} u_m - \frac{N}{\mu(1-\alpha)(N - m)(N\beta + m)},$$

avec $\beta' = (K - 1)\beta$. Le reste du calcul est exactement le même. \blacksquare

Dans le cas où on ne démarre pas en A_k , on sait par continuité qu'au premier instant où on entrera dans A_k , on aura presque sûrement $n_k = N_{\text{sup}} + 1$. On a donc le

Corollaire 3.12. *Sachant que le système ne démarre pas en A_k , le temps de siphonnage moyen de la zone k est donné par :*

$$\mathbb{E}[S_k | \mathbf{n}(0) = \mathbf{n} \notin A_k] = \frac{N}{\mu(1-\alpha)} \sum_{l=N_{\text{inf}}}^{N_{\text{sup}}+1} \sum_{j=l}^N \frac{\prod_{l \leq i < j} \delta(i)}{j(N(K-1)\beta + N - j)}.$$

Comme on pouvait s'y attendre, ce nombre est indépendant de \mathbf{n} et k . On le note $\mathbb{E}[S]$.

La figure 3.13 représente l'évolution de $\mathbb{E}[S]$ en fonction de α , et ce pour plusieurs valeurs de N . Sur cette figure on a aussi représenté α_{min} , la valeur de α à laquelle $\mathbb{E}[S]$ est minimal.

De la même manière, on peut calculer le "temps de remplissage moyen" R d'une zone :

Définition 3.4. Le temps de remplissage R_k de la zone k est la durée entre le premier instant où on se trouve dans B_k et le premier instant où on se trouve dans A_k .

On a alors le

Théorème 3.13. *Sachant que le système démarre en B_k , le temps de remplissage moyen de la zone k est donné par :*

$$\mathbb{E}[R_k | \mathbf{n}(0) = \mathbf{n} \in B_k] = \frac{N}{\mu(1-\alpha)} \sum_{l=N-N_{\text{sup}}}^{N-n_k} \sum_{j=l}^N \frac{\prod_{l \leq i < j} \epsilon(i)}{j(N\beta + N - j)},$$

où

$$\epsilon(i) = \frac{1}{\delta(N-i)} = \frac{(N-i)(N(K-1)\beta + i)}{i(N\beta + N - i)}.$$

Démonstration - Le début de la récurrence est le même que pour le théorème 3.11. Seule la plage de valeurs change : pour $1 \leq m \leq N_{\text{sup}}$ on a toujours :

$$\begin{aligned} v_{m+1} &= v_m - \frac{N}{\mu(1-\alpha)(N-m)(N\beta + m) \prod_{i=m+1}^{N-1} \delta(i)} \\ &= v_m - \frac{N}{\mu(1-\alpha)} \cdot \frac{\prod_{i=1}^{N-m-1} \epsilon(i)}{(N-m)(N\beta + m)}, \end{aligned}$$

soit, pour $2 \leq m \leq N_{\text{sup}} + 1$:

$$\begin{aligned} v_m &= v_{m-1} - \frac{N}{\mu(1-\alpha)} \cdot \frac{\prod_{i=1}^{N-m} \epsilon(i)}{(N-m+1)(N\beta + m - 1)} \\ &= v_1 - \frac{N}{\mu(1-\alpha)} \sum_{j=2}^m \frac{\prod_{i=1}^{N-j} \epsilon(i)}{(N-j+1)(N\beta + j - 1)}. \end{aligned}$$

Puis, de manière symétrique par rapport au cas précédent, on a :

$$\begin{aligned} -1 &= q_{0,0}h_0 + q_{0,1}h_1 \\ u_1 &= -\frac{1}{q_{0,1}} \\ &= -\frac{2}{\mu\alpha N} \\ v_1 &= -\frac{2}{\mu\alpha N \prod_{i=1}^{N-1} \delta(i)} \\ &= -\frac{2 \prod_{i=1}^{N-1} \epsilon(i)}{\mu\alpha N}, \end{aligned}$$

d'où :

$$v_m = -\frac{N}{\mu(1-\alpha)} \sum_{j=1}^m \frac{\prod_{i=1}^{N-j} \epsilon(i)}{(N-j+1)(N\beta+j-1)}$$

et :

$$u_m = \frac{v_m}{\prod_{i=1}^{N-m} \epsilon(i)} = -\frac{N}{\mu(1-\alpha)} \sum_{j=1}^m \frac{\prod_{N-m < i \leq N-j} \epsilon(i)}{(N-j+1)(N\beta+j-1)}.$$

Il ne reste plus qu'à remarquer que $h_{N_{\text{sup}}} = -u_{N_{\text{sup}}+1}$ pour conclure :

$$h_{N_{\text{sup}}} = -\frac{N}{\mu(1-\alpha)} \sum_{j=1}^{N_{\text{sup}}+1} \frac{\prod_{N-N_{\text{sup}} \leq i \leq N-j} \epsilon(i)}{(N-j+1)(N\beta+j-1)},$$

et plus généralement :

$$\begin{aligned} h_m &= \frac{N}{\mu(1-\alpha)} \sum_{l=m+1}^{N_{\text{sup}}+1} \sum_{j=1}^l \frac{\prod_{N-l < i \leq N-j} \epsilon(i)}{(N-j+1)(N\beta+j-1)} \\ &= \frac{N}{\mu(1-\alpha)} \sum_{l=N-N_{\text{sup}}}^{N-m} \sum_{j=l}^N \frac{\prod_{l \leq i < j} \epsilon(i)}{j(N\beta+N-j)}. \end{aligned}$$

La fin de la preuve ne change pas. ■

De même, on a le

Corollaire 3.14. *Sachant que le système ne démarre pas en B_k , le temps de remplissage moyen de la zone k est donné par :*

$$\mathbb{E}[R_k | \mathbf{n}(0) = \mathbf{n} \notin B_k] = \frac{N}{\mu(1-\alpha)} \sum_{l=N-N_{\text{sup}}}^{N-N_{\text{inf}}+1} \sum_{j=l}^N \frac{\prod_{l \leq i < j} \epsilon(i)}{j(N\beta+N-j)},$$

Comme on pouvait s'y attendre, ce nombre est indépendant de \mathbf{n} et k . On le note $\mathbb{E}[R]$.

Remarque 3.4. Il est facile de trouver des formules de récurrence pour $\mathbb{E}[S]$ et $\mathbb{E}[R]$, ce qui permet de ramener la complexité des calculs à un $O(N)$.

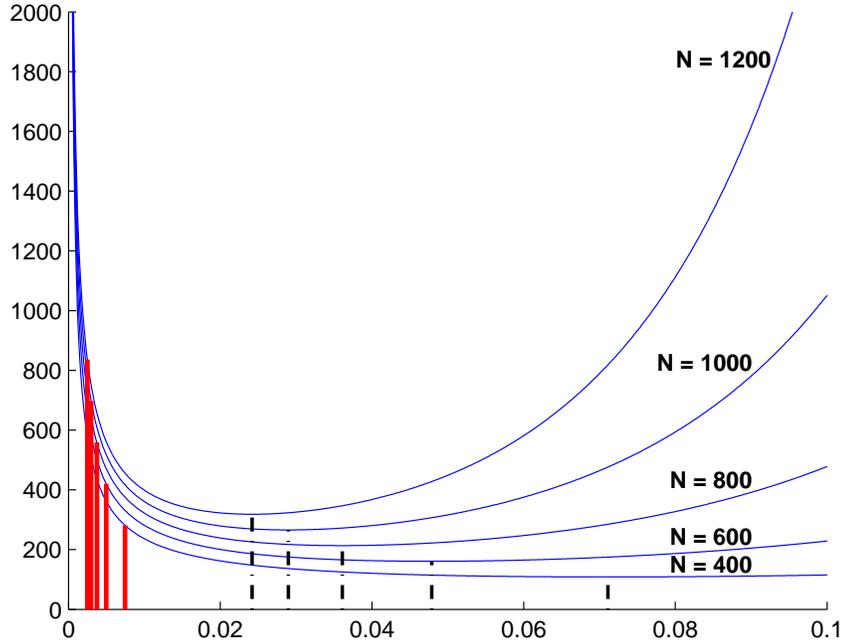


Fig. 3.13: Temps de siphonnage moyen (en minutes) en fonction de α . $N_{\text{sup}}/N = 80\%$, $N_{\text{inf}}/N = 20\%$, $K = 3$ et $\mu = 1 \text{ min}^{-1}$. Pour chaque courbe, le trait plein vertical correspond à la valeur critique α_c , tandis que la ligne pointillée correspond à α_{min} .

3.4 Calibration à l'aide des traces

Dans cette section, nous regardons si les mouvements de foule observés à l'aide de nos traces mobiles peuvent raisonnablement être décrits par le modèle des moutons et des fortes têtes. Nous proposons d'utiliser $\mathbb{E}[S]$ et/ou $\mathbb{E}[R]$ comme grandeur de calibration. En les mesurant suffisamment précisément, nous pourrions en déduire α en inversant les formules développées ci-dessus. Si la fonction $\mathbb{E}[S] = f(\alpha)$ par exemple est injective, au moins pour $\alpha < \alpha_c$, alors le problème inverse pourra être résolu sans ambiguïté. La figure 3.13 suggère que la fonction $f(\alpha)$ est toujours convexe, et donc injective sur l'intervalle $]0, \alpha_{\text{min}}]$, si bien que si $\alpha_c < \alpha_{\text{min}}$ alors notre problème aura au plus une solution. Malheureusement, c'est faux en général. Un contre-exemple est donné par $K = 2$, $N = 100$ et $N_{\text{sup}}/N = 0.99\%$.

En pratique, α_{min} paraît difficile à calculer analytiquement, à tel point

que nous n'avons pas pu trouver ne serait-ce qu'une condition suffisante simple pour garantir $\alpha_c < \alpha_{\min}$. Et quand bien même ce serait le cas, rien ne nous garantit que la fonction $f(\alpha)$ soit systématiquement convexe, même si intuitivement, on peut s'attendre à ce que ce soit toujours le cas : l'effet mouton est en lutte contre l'effet forte tête, et selon la valeur de α c'est l'un ou l'autre qui l'emporte. Dans les valeurs extrêmes de α , les temps de siphonnage sont extrêmement longs : soit il n'y a pratiquement que des moutons, auquel cas les zones mettent évidemment beaucoup de temps à se vider, soit il n'y a pratiquement que des fortes têtes, auquel cas en permanence toutes les zones sont plus ou moins équilibrées en effectifs.

L'expérience montre que f est toujours convexe, et que les contre-exemples à $\alpha_c < \alpha_{\min}$ sont extrêmement rares, surtout si $K \geq 3$. Cependant, s'il apparaissait qu'il est impossible de choisir N_{\inf} et N_{\sup} correctement, alors nous recommandons de croiser les mesures de $\mathbb{E}[S]$ avec des mesures de $\mathbb{E}[R]$ pour lever autant d'ambiguïtés que possible.

Attention, si on décide de calibrer α uniquement sur $\mathbb{E}[R]$, il faut être plus prudent, car l'expérience montre que l'injectivité de $g(\alpha) = \mathbb{E}[R]$ sur $]0, \alpha_c[$ est moins souvent remplie.

Revenons à notre calibration sur les données de la Fête de la Musique. Il n'est pas question de proposer ici une analyse statistique extensive, d'autant plus que le jeu de données dont nous disposons est trop réduit pour cela, mais plutôt d'illustrer sur un exemple simple la marche à suivre. Considérons par exemple le transfert mis en évidence sur la figure 3.2, entre le Parc des Princes et l'hippodrome d'Auteuil. En observant précisément nos données, nous nous sommes aperçus qu'il y avait trois zones d'attraction :

- Auteuil lui-même
- le Parc des Princes, zone Nord
- le Parc des Princes, zone Sud

Ainsi, nous avons $K = 3$. Sur la figure 3.14, on peut voir l'évolution du nombre d'individus dans chaque zone au cours du temps. Le nombre total de SMS reste plus ou moins constant au cours du temps, si bien que nous choisissons la valeur moyenne $N = 4830$. Cela fournit $\alpha_c = 6.2 \times 10^{-4}$. Dans nos données, un signal SMS dure environ 1 seconde : à chaque fois qu'un SMS est reçu ou envoyé, un nouvel individu est comptabilisé, mais il est systématiquement effacé après une seconde. Ainsi, même si ce n'est pas

vrai dans la réalité mais afin de rester cohérents (puisque seule son influence compte), nous faisons comme si chaque utilisateur restait 1 seconde dans sa zone avant de disparaître, ce qui nous donne $\mu = 1 \text{ sec}^{-1}$. Ainsi, nous avons $K = 3$. Sur la figure 3.14, on peut voir l'évolution du nombre d'individus dans chaque zone au cours du temps. Le nombre total de SMS reste plus ou moins constant au cours du temps, si bien que nous choisissons la valeur moyenne $N = 4830$. Cela fournit $\alpha_c = 6.2 \times 10^{-4}$. Dans nos données, un signal SMS dure environ 1 seconde : à chaque fois qu'un SMS est reçu ou envoyé, un nouvel individu est comptabilisé, mais il est systématiquement effacé après une seconde. Ainsi, même si ce n'est pas vrai dans la réalité mais afin de rester cohérents (puisque seule son influence compte), nous faisons comme si chaque utilisateur restait 1 seconde dans sa zone avant de disparaître, ce qui nous donne $\mu = 1 \text{ sec}^{-1}$.

En choisissant de faire notre calibration sur le remplissage d'Auteuil par exemple, nous prenons $N_{\text{sup}}/N = 55\%$ et $N_{\text{inf}}/N = 45\%$. Après avoir vérifié que $g(\alpha)$ était injective sur $]0, \alpha_c[$, nous obtenons $\alpha = 0.6 \times 10^{-4}$, c'est-à-dire $\alpha \ll \alpha_c$. Afin de vérifier cette valeur de α , nous mesurons également le siphonnage du Parc des Princes en zone Sud, ce qui nous donne $\alpha = 0.5 \times 10^{-4}$. Encore une fois, cela ne saurait constituer une vérification certaine de α , et si l'on a à sa disposition un plus grand jeu de données, il est bon d'effectuer une moyenne sur un échantillon aussi important que possible.

Maintenant, si notre modèle est consistant, il doit être capable de prédire d'autres siphonnages ou remplissages à travers Paris. Considérons le triangle Châtelet-Saint-Michel-Bastille par exemple, dans le cœur de Paris. Nous observons que les trois zones subissent chacune des remplissages de durée à peu près égale (voir figure 3.15). Nous considérons aussi la zone complémentaire, qui connaît un siphonnage. Ainsi $K = 4$.

D'après le corollaire 3.14, le remplissage d'une zone de 5% à 28% par exemple devrait durer environ 10800 secondes, selon la valeur de α que nous avons calculée précédemment. Or le remplissage de Bastille dure environ 11300 secondes, celui de Châtelet également, et celui de Saint-Michel dure 11000 secondes, ce qui est très proche de nos prédictions.

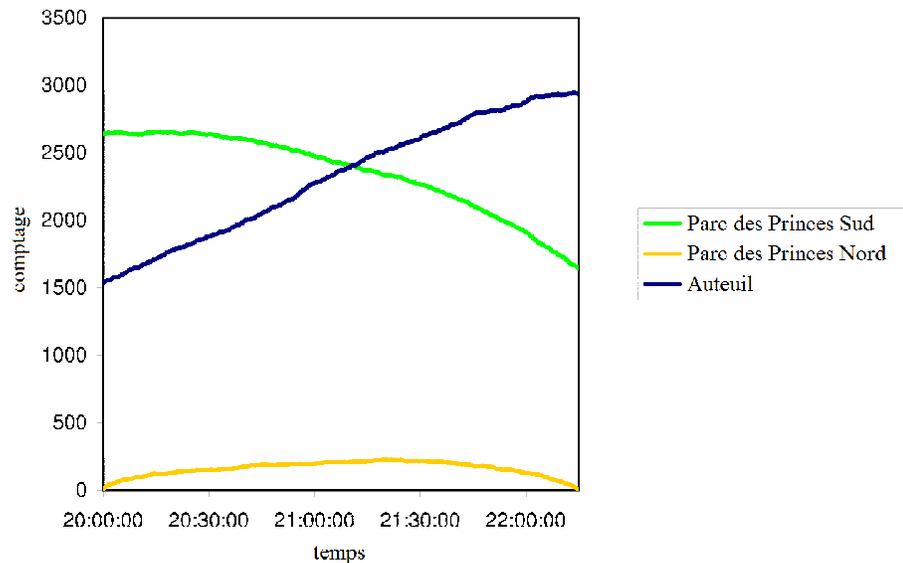


Fig. 3.14: SMS envoyés pendant la Fête de la Musique entre les trois zones d'attraction.

3.5 Conclusion et perspectives

Nous avons proposé un nouveau modèle de mobilité pour représenter la formation de hotspots dans un mouvement de foule, ainsi que leur dynamique aléatoire. Nous avons prouvé que ce modèle était calculable analytiquement en proposant des formules fermées ainsi que des estimations asymptotiques. En utilisant les traces des utilisateurs d'un réseau GSM, nous avons montré comment calibrer ce modèle pour qu'il suive au plus près la réalité. Enfin, ce modèle a montré qu'il pouvait prédire la formation de hotspots dans différentes situations spatio-temporelles.

Notons que d'autres modèles de décision peuvent être imaginés pour les individus, comme d'avoir des fortes têtes qui privilégient une zone en particulier. À ce sujet, on pourra consulter [Bin11].

Avant de passer à quelques applications de ce modèle, voici des réponses à deux objections qui pourraient venir à l'esprit du lecteur :

- pourquoi y a-t-il un nombre constant d'utilisateurs ?

À cela nous répondons d'abord que l'on travaille sur des grands effectifs de population (plusieurs dizaines de milliers), si bien que statistique-

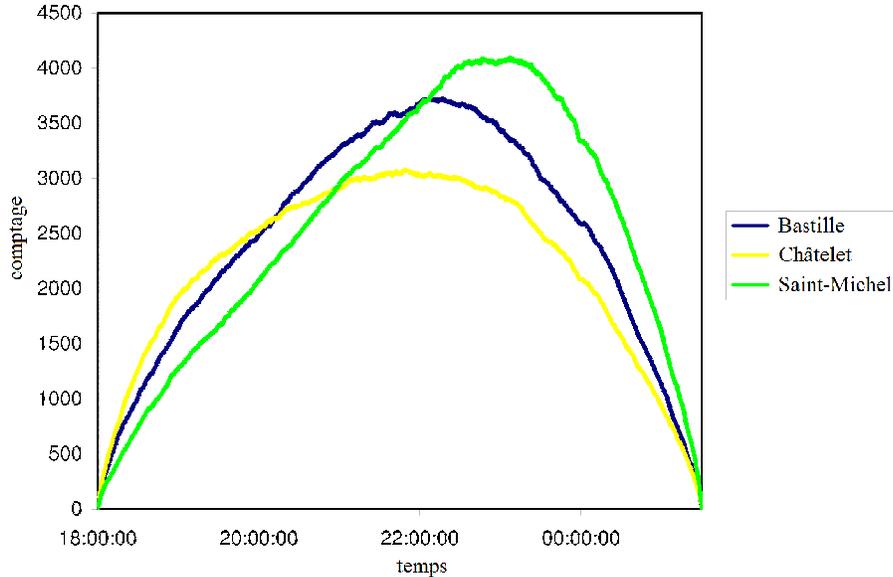


Fig. 3.15: SMS envoyés pendant la Fête de la Musique dans le triangle Châtelet-Saint-Michel-Bastille.

ment, l'évolution du nombre total d'utilisateurs est lente. Sur de petite périodes, on peut considérer que ce nombre est à peu près constant, et qui plus est nous disposons d'équivalents asymptotiques pour la plupart de nos formules. Le fait qu'il existe de tels équivalents donne un sens précis à la notion d'approximation.

- est-ce que les transitions sont instantanées ?

À vrai dire, il n'y a aucune raison *a priori* de supposer les transitions instantanées comme nous l'avons fait, surtout si deux zones sont éloignées géographiquement. Cependant, nous avons considéré qu'à partir du moment où un individu avait décidé de se déplacer dans une nouvelle zone, alors sa décision était prise dès l'instant où il quittait l'ancienne zone. Et c'est cela qui compte, puisque dans un événement comme la Fête de la Musique par exemple, les messages qu'il échange avec ses proches lui servent bien souvent à leur donner rendez-vous à tel ou tel endroit (selon l'âge de l'utilisateur, messages du type: "venez dans la cour du musée Carnavalet, il y a un concert formidable" ou "ramène-toi à Pigalle, y a trop une ambiance de ouf"). Aussi à partir

du moment où il se met en route pour une zone donnée, on peut considérer qu'il y est virtuellement déjà, puisqu'il est désormais capable d'influencer d'autres individus pour venir à sa suite.

Au-delà d'une des applications les plus immédiates, qui est de fournir un modèle de mobilité facilement implémentable pour les simulateurs de réseau, voici quelques autres applications qui reposent sur la capacité à prévoir les siphonnages et les remplissages :

- mettre de la donnée en mémoire cache dans des réseaux ad-hoc denses

Dans un réseau ad-hoc dense, il peut être utile d'ajouter quelques serveurs fixes qui contiennent les données les plus souvent demandées. Plusieurs algorithmes existent déjà dans ce sens (voir par exemple [Har01][Har03][Yin06]). La question est : où et quand doit-on mettre de la donnée en cache ? L'idée est de sélectionner quelques endroits stratégiques où les individus ont l'habitude de se rassembler. Ce sont nos K zones. Supposons que de la même manière que dans la section 3.4, nous ayons préalablement évalué α , éventuellement à un autre endroit de la ville avec d'autres individus. Alors, en mesurant en temps réel le nombre N d'individus et en calculant le temps de siphonnage correspondant, dès qu'un regroupement se produit on sait combien de temps en moyenne il va durer. En d'autres mots, on sait si cela est utile de placer de la donnée en cache à cet endroit. Réciproquement, si une zone est presque vide, en calculant son temps de remplissage on peut savoir s'il existe un risque d'attroupement à cet endroit qui obligerait à placer de la donnée en cache.

- allocation dynamique de spectre

L'allocation dynamique de spectre (*Dynamic Spectrum Allocation* dans la littérature anglo-saxonne, ou DSA), consiste à allouer des ressources variables en bande passante selon les zones et les densités de population. Elle constitue aujourd'hui un champ de recherche très actif. Comme dans le point précédent, si nous savons prévoir à l'avance le remplissage ou le siphonnage d'une zone sur le long terme, nous pouvons décider de la quantité de ressource à allouer.

- améliorer la robustesse de certains optimisateurs

Plus généralement, certains algorithmes d'optimisation sont réputés dépendre d'un réglage fin des paramètres. Par exemple, [Kau07] propose un algorithme pour optimiser les ressources radio dans un réseau 802.11, basé sur un échantillonneur de Gibbs (*Gibbs sampler*). Le "paramètre de température" de l'échantillonneur doit être réglé très soigneusement, afin que l'optimisateur puisse réagir plus rapidement que le temps d'évolution typique du système lui-même. Plus concrètement, dans notre cas, supposons que l'on veuille prendre en compte la mobilité des utilisateurs dans un réseau 802.11; si l'on peut anticiper une désagrégation alors on pourra adapter l'optimisateur d'une manière convenable.

Dans le cas d'un algorithme génétique également, il peut être intéressant d'adapter les taux de mutation et de cross-over de l'optimisateur, afin de savoir réagir à une évolution soudaine du système. En effet, plus ces taux sont hauts, plus la diversité de la population s'accroît. Dans notre cas, un éclatement de population est susceptible de modifier totalement le système à optimiser, aussi il est bon de pouvoir s'y préparer.

