



HAL
open science

Modéliser l'émergence et détecter les structures de la ségrégation sociospatiale

Antoine Lucquiaud

► **To cite this version:**

Antoine Lucquiaud. Modéliser l'émergence et détecter les structures de la ségrégation sociospatiale. Mathématiques [math]. Université Paris 1 - Panthéon Sorbonne, 2022. Français. NNT: . tel-04085569

HAL Id: tel-04085569

<https://hal.science/tel-04085569>

Submitted on 29 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Modéliser l'émergence de la ségrégation socio spatiale et détecter ses structures

THÈSE

pour l'obtention du grade de
Docteur de l'Université Paris 1 Panthéon Sorbonne

Soutenue le 12 décembre 2022 par

Antoine Lucquiaud

Sous la direction de
Julien Randon-Furling

Spécialité
mathématiques
appliquées

École doctorale n° 386
Sciences
Mathématiques
de Paris Centre

Laboratoire
Statistique,
Analyse et
Modélisation
Multidisciplinaire

EA 4543 – FP2M (CNRS FR2036)

Composition du jury

Laetitia Gauvin, rapporteuse
ISI Foundation, Torino

Niccolò Torri, rapporteur
Université Paris Nanterre

Elsa Arcaute, examinatrice
University College London

Eva Löcherbach, examinatrice
Université Paris 1 Panthéon Sorbonne

Marie Théret, examinatrice
Université Paris Nanterre

Julien Randon-Furling, directeur de thèse
École Normale Supérieure Paris-Saclay

“C’est bien plus beau lorsque c’est inutile !”

E. Rostand, *Cyrano de Bergerac* [147]

Merci

À Julien, pour ta confiance.

À Laetitia Gauvin et Niccolò Torri d'avoir accepté de rapporter cette thèse, ainsi qu'à Elsa Arcaute, Eva Löcherbach et Marie Théret d'avoir accepté de faire partie du jury.

Aux camarades de Tolbiac et de Nanterre, d'avoir su créer et maintenir une atmosphère chaleureuse, légère et désinvolte en toute circonstance : merci Alice, Amanda, Branda, Bruno, Christian, Clara, Clément, Eva, Fabrice, Florent, Florian, Frédérique, Jean-Marc, Kamila, Madalina, Marie, Niccolò, Nicolas, Xavier et tout·e·s les autres.

À Marion, à Claire et à toutes les rencontres de bibliothèque, pour votre présence et votre soutien pendant les longs mois de rédaction. À toutes celles et tous ceux qui ont bien voulu écouter, lire, annoter, critiquer, ajouter, proposer. À celles et ceux que je connais ou non qui luttent contre la casse de l'Université et des services publics en général, et qui s'évertuent à promouvoir un accès libre et gratuit à la recherche scientifique pour tout·e·s. Bref, à vous, qui que vous soyez, qui m'avez aidé à construire ce travail.

À mes parents, pour plein de choses mais surtout d'être venus à la soutenance après avoir lu ce manuscrit pour en traquer les erreurs. À Martin, qui a demandé à le lire puis qui n'est pas venu.

À mes ami·e·s qui se reconnaîtront, de Lorient à Nanterre, de la Kfet au Musée, de la F[ens]foire aux Jackets, rencontres récentes ou amitiés tenaces, qui nous aident mutuellement à avancer.

À Julie, enfin, pour tout. Merci infiniment.

Sommaire

Remerciements	5
Notations	11
Introduction	13
Revue d'une littérature multidisciplinaire	15
1.1 Le modèle originel	17
1.1.1 Premières observations	17
1.1.2 Étude systématique, diagrammes de phase	19
1.1.3 Universalité	21
1.2 Lexique et outils d'analyse	33
1.2.1 Un système à deux échelles	33
1.2.2 Quantités locales et satisfaction	34
1.2.3 Paramètres de contrôle, paramètre d'ordre	36
1.3 La ségrégation : un état d'équilibre?	39
Classification des états de la chaîne de Markov	47
2.1 Démarche et définitions	47
2.1.1 Formalisme markovien	47
2.1.2 Description naïve	49
2.2 Évolution de la densité d'interface	58
2.2.1 Espace d'états	58
2.2.2 Incréments de la densité d'interface	59
2.2.3 Distribution des aimantations	63
2.2.4 Conditionnement de la marche aléatoire	65
2.3 Irréversibilité et invariance	69
2.3.1 Tolérance, échanges autorisés	69
2.3.2 Distribution invariante, comparaison avec les simulations	71
Un système hors équilibre?	79
3.1 Thermodynamique statistique	79
3.2 Un système isolé?	84
3.2.1 Température et réversibilité	84
3.2.2 Entropie microcanonique et température négative	90

3.3	Un système fermé?	95
3.3.1	Grands voisinages unidimensionnels	95
3.3.2	Extension à deux dimensions	99
3.4	Un modèle de nucléation	105
3.4.1	Noyaux homogènes, cœur et interfaces	105
3.4.2	Énergie moyenne du système	107
Détecter les structures de la ségrégation urbaine		117
4.1	Décrire et mesurer la ségrégation urbaine	118
4.1.1	Définir la ségrégation	118
4.1.2	Le problème de la mesure	121
4.2	Détection de structures spatiales	126
4.2.1	Trajectoires et distance focale	126
4.2.2	Inhomogénéités spatiales à Paris	129
4.3	La ville uniforme comme référence	137
4.3.1	Construire une ville uniforme	137
4.3.2	Propriétés des trajectoires	139
4.3.3	Subordination	141
4.4	Discussion	150
Conclusion		153
Annexes		157
A	Stabilité stochastique	157
B	Incréments de la densité d'interface	158
C	Taille de l'espace d'états de la chaîne de Markov	158
D	Présence de sites vacants dans un voisinage	158
E	Entropie microcanonique	159
F	Interfaces et rugosité	160
G	Marche aléatoire bidirectionnelle	161
H	Schelling : un modèle trivial?	163
I	<i>From urban segregation to spatial structure detection</i>	166
Références		183

Notations

σ, Σ	configuration, ensemble des configurations
N	taille du système
κ	taille d'un voisinage
i, j	sites
s_i	contenu du site i
$\partial i, \kappa_i$	voisinage, taille effective du voisinage du site i
m_i, h_i	aimantation et champ effectifs du site i
T	tolérance
d	densité
p	parité
$n_{\bullet\infty}$	nombre de paires dirigées hétérogènes d'agents
n_κ	nombre total de paires dirigées d'agents
$x, \Delta x$	densité d'interface, incrément de la densité d'interface
π	matrice de transition de la chaîne de Markov
\mathcal{C}_x	état de la chaîne de Markov sous-jacente
\mathcal{N}	taille de l'espace des états $\{\mathcal{C}_x\}_x$
S	entropie
β	température inverse
Θ	température
Ω	noyau homogène
Ω	taille d'un noyau homogène
$\partial\Omega$	interface interne d'un noyau homogène
$\partial\Omega_{\text{ext}}$	interface externe d'un noyau homogène
\mathcal{X}	marche aléatoire ($= (X_t)_{t \geq 0}$)
δ_{ij}	symbole de Kronecker
$\mathbb{1}_E$	indicatrice de E

| Introduction

La ségrégation sociospatiale est le constat d'une répartition spatiale inhomogène des groupes sociaux, et d'inégalité d'accès aux infrastructures, aux biens matériels et symboliques [75]. Ce phénomène, observé et documenté dans les milieux urbains, constitue un enjeu de politique publique depuis plusieurs décennies [8, 106, 108, 130, 141]. Plus récemment, son apparition au sein des réseaux sociaux et informationnels (phénomènes de chambres d'écho [5, 18, 68]) confirme la nécessité de l'appréhender scientifiquement, par sa détection, sa mesure et sa quantification, mais aussi en cherchant à comprendre son émergence par la construction et l'étude de modèles. Nous aborderons ces deux aspects dans cette thèse, en commençant par le second.

La ségrégation sociospatiale a été modélisée de diverses manières [150, 151], mais le modèle incontournable est celui de T. Schelling, proposé et développé autour du début des années 1970 [152, 153]. Son modèle est particulièrement intéressant car il met en avant le caractère *complexe*, au sens des systèmes complexes, du phénomène. Le titre de son livre *Micromotives, macrobehaviors* [154] décrit en particulier l'auto-organisation d'un système d'agents et l'émergence en son sein de propriétés collectives surprenantes, inattendues, voire paradoxales au vu des propriétés individuelles. Nous proposons d'éclairer sous un jour nouveau certaines de ces propriétés, avec pour enjeu le calcul analytique de valeurs critiques caractérisant les transitions entre les différents comportements macroscopiques du système.

Après avoir étudié le comportement de ce modèle, nous changerons de point de vue et prendrons en quelque sorte le problème par l'autre bord, celui des données et d'une approche statistique, pour détecter des structures plus fines du phénomène de ségrégation à l'aide de données en accès libre, dans une perspective complémentaire aux travaux des sciences humaines et sociales. En particulier, nous proposerons un *null model* qui pourra servir de référence dans l'optique d'effectuer des mesures comparables entre elles. Nous verrons alors en quoi sa mise en application pose des problèmes non triviaux.

1. | Schelling

Revue d'une littérature multidisciplinaire

LA littérature traitant du modèle de Schelling et de ses variantes est immense. L'accroissement exponentiel [158] du nombre de publications de sciences sociales [28, 36, 37, 94, 104, 157, 178–180] qui y sont liées ne suffit pas à rendre compte de la diversité des approches existantes, en particulier dans d'autres disciplines telles que la théorie des jeux [2, 56], la physique statistique et l'algorithmique théorique [20, 90]. Ces travaux sont majoritairement empiriques ou n'exposent que quelques résultats de simulations numériques. Pour cette raison, ils ne seront que peu présentés dans ce chapitre.

Comparativement, le nombre de publications qui explorent systématiquement l'espace des phases du modèle et de ses variantes est étonnamment réduit. Chaque année depuis plus de vingt ans, la littérature théorique s'est enrichie d'un à cinq articles proposant une solution analytique ou un ensemble rigoureux, à défaut de pouvoir être exhaustif, de simulations d'une variante du modèle de Schelling.

H.P. Young [176] est le premier à décrire l'émergence de la ségrégation dans un système unidimensionnel suivant les règles originelles d'évolution avec le formalisme de la théorie des jeux et des chaînes de Markov. Mais c'est sous la forme d'un système de spins que M. Pollicott et H. Weiss [138] décident d'étudier analytiquement la dynamique de ce modèle, avec les outils de la physique statistique, suivis par L. Gauvin et al. [70, 71] puis N. Goulet-Dominic et al. [74]. Il ressort de ces travaux la présence d'une transition entre deux phases, l'une dont l'état final est *ségrégré* et l'autre à l'état final dit *intégré*.

De plus, l'étude de la phase ségrégré semble indiquer l'existence d'une seconde transition, moins marquée mais correspondant à celle repérée par Schelling, vers un état final intégré pour des agents très tolérants, apportant ainsi un résultat proche de l'intuition. Entre ces deux transitions, des systèmes d'agents tolérants finissent donc dans des configurations ségrégréées. Le "paradoxe" de Schelling peut alors s'énoncer de la façon suivante : *une forte intolérance n'est pas une condition nécessaire à l'émergence de la ségrégation.*

Une autre branche de la littérature [14–17, 26, 88, 123–125, 127] se consacre à la démonstration de l’existence de cette transition ainsi qu’à la détermination analytique de sa valeur critique. Faisant intervenir des réseaux dont le nombre de sites et le degré moyen tendent vers l’infini tout en maintenant une faible densité de liens, ces travaux exploitent certaines propriétés limites des distributions initiales d’agents afin d’étudier la formation de noyaux homogènes de taille conséquente. Ces méthodes seront présentées succinctement dans le chapitre **3**.

Ces multiples travaux conservent une cohérence d’ensemble grâce à la robustesse remarquable du modèle de Schelling aux variantes. Modification de la structure du réseau [9, 12, 22, 77–80, 86, 101, 172], de la répartition initiale des agents [47, 70, 74, 157], jusqu’au détail des règles d’évolution [83, 133, 171, 172], les simulations de ces variations du modèle originel donnent sensiblement les mêmes résultats, ce qui permet à la littérature analytique de s’affranchir de contraintes inutiles ou inversement d’introduire des lois de conservation.

Au regard de cette observation, une exception est à noter. Il s’agit de la catégorie des “jeux de Schelling”, dans la continuité des travaux de Young en théorie des jeux. Récemment, plusieurs articles [2, 34, 55, 56, 95] ont mis en évidence que la convergence de modèles à deux groupes d’agents ou plus vers un état ségrégré ou intégré n’était pas garantie sur tout type de réseau, et qu’il existe une gamme de paramètres faisant émerger des cycles limites.

Ce chapitre vise à familiariser le lecteur avec le modèle originel de Schelling, en présentant ses objets et quantités d’intérêt ainsi que sa robustesse.

Le formalisme dans lequel ces articles seront présentés sera celui dans lequel les résultats de cette thèse seront développés ultérieurement. Ceci devrait permettre, autant que possible, d’avoir une vision d’ensemble des chemins explorés depuis 1971.

1.1 | Le modèle originel

1.1.1 Premières observations

Dans son modèle bidimensionnel originel, T. Schelling [153, 154] dispose aléatoirement des agents sur une grille carrée, de telle sorte qu'un tiers des sites contienne un agent de type \oplus , un tiers un agent de type \ominus et que le tiers restant soit vide.

Schelling identifie alors l'ensemble des agents insatisfaits, c'est-à-dire ceux dont plus de la moitié des huit voisins appartient à l'autre groupe, et les déplace l'un après l'autre (dans un ordre quelconque, par exemple d'en haut à gauche à en bas à droite, ou en partant du centre) sur un des sites vides dont le voisinage est satisfaisant, ou, s'il n'en reste plus, sur un site vide quelconque (Schelling précise que cela arrive rarement).

Tous les agents insatisfaits dans la configuration initiale ayant été déplacés, il suffit de recommencer l'opération jusqu'à satisfaction générale, ou jusqu'à ce qu'aucun mouvement ne soit possible. On atteint alors une configuration stationnaire¹, pour laquelle Schelling observe une ségrégation importante. Ce constat repose en partie sur une simple observation de l'état final, en traçant les contours des noyaux homogènes pour les faire apparaître plus clairement, et en partie sur la mesure du ratio moyen d'agents voisins d'un groupe par rapport à l'autre. Le choix des observables sera un point développé dans la section **1.1.3**, mais dans son article de 1971 déjà, Schelling précise que

*Bien sûr, ces mesures quantitatives n'ont de sens que sur une grille artificielle et il est peu probable qu'elles reflètent une réalité similairement quantifiable.*²

ce qui sera discuté dans le chapitre **4**.

L'intérêt du travail de Schelling est ainsi de montrer que dans un cadre théorique très simple, la ségrégation émerge en dépit d'une tolérance relativement grande d'un groupe d'agents pour l'autre. Son intensité dépend évidemment du seuil de tolérance : Schelling note ainsi que lorsqu'un agent tolère une mixité de $2/3$, la ségrégation observée est modérée, se rapprochant

¹Cette notion et celle d'équilibre seront amplement discutées dans la suite.

²*Quantitative measures, of course, refer exclusively to an artificial checkerboard and are unlikely to have any quantitative analogue in the living world.* [153], p.158 [traduction libre].

de l'intuition qu'un système tolérant devrait être plus intégré, alors qu'elle est frappante lorsque la tolérance descend à $1/2$. Il ne procède toutefois pas à une étude systématique pour toutes les valeurs que peut prendre ce paramètre, il faudra attendre pour cela les travaux de Laurie et Jaggi [101] puis de Gauvin, Vannimetus et Nadal [70, 71]. Qualitativement enfin, il est observé que la baisse de la tolérance tend à accroître le nombre d'agents insatisfaits dans l'état initial, et, plus tard dans la dynamique, à contraindre un nombre plus importants d'agents initialement satisfaits à se déplacer.

Ayant introduit beaucoup de paramètres arbitraires, tant sur la configuration initiale du système que sur ses règles de mises à jour, Schelling revient plus en détail sur certains d'entre eux, dont trois principaux, correspondant aux différentes variantes développées ultérieurement, seront retenus ici :

- la taille du voisinage considéré par un agent qui évalue sa satisfaction en un site donné ;

- la distribution initiale des agents, et notamment ce qu'il se produit lorsque l'une des deux populations est majoritaire ;

- le choix de la condition de satisfaction, et notamment ce qu'il se produit lorsque l'une des deux populations est moins tolérante que l'autre.

Ses constats sont les suivants. Agrandir le voisinage, ou le champ de vision, des agents semble atténuer le phénomène de ségrégation, pour des valeurs de tolérance et de parité modérées. Une population minoritaire aura tendance à se rassembler, formant quelques îlots homogènes et laissant l'autre groupe occuper tout l'espace autour. Enfin, une différence de tolérance entre les populations n'aura pas vraiment d'effet, ce qui illustre le fait que la ségrégation est un phénomène intrinsèquement réciproque. La seule observation notable dans ce dernier cas est une densification de la population la moins tolérante (l'autre restera au contact des sites vides du système).

En conclusion, cette première version du modèle ouvre la voie à une multitude d'études et de variantes. Chacun des choix faits par Schelling peut en effet être remis en question, pour comparer les résultats aux originaux et tester ainsi la robustesse de son modèle. C'est ce qui a été fait depuis plus de vingt ans maintenant et ce qui est présenté dans la suite de ce chapitre, avant une redéfinition de chaque élément du modèle initial pour n'en garder que le squelette et étudier ses propriétés.

1.1.2 Étude systématique, diagrammes de phase

Avant d'explorer variantes et propositions de résolution, voici une présentation des résultats d'analyses plus poussées du modèle, notamment sur l'évolution de l'impact de la tolérance T des agents sur l'état final du système au regard de deux autres paramètres importants : la densité d et la parité p . Formellement définies dans la section **1.2.1**, ces deux quantités mesurent respectivement le ratio de sites occupés dans le système (inversement la densité de sites vides $\rho = 1 - d$ mesure le ratio de sites vides), et le ratio de sites contenant un agent d'un des deux groupes parmi les sites occupés : une parité différente de $1/2$ implique que l'une des deux populations est majoritaire.

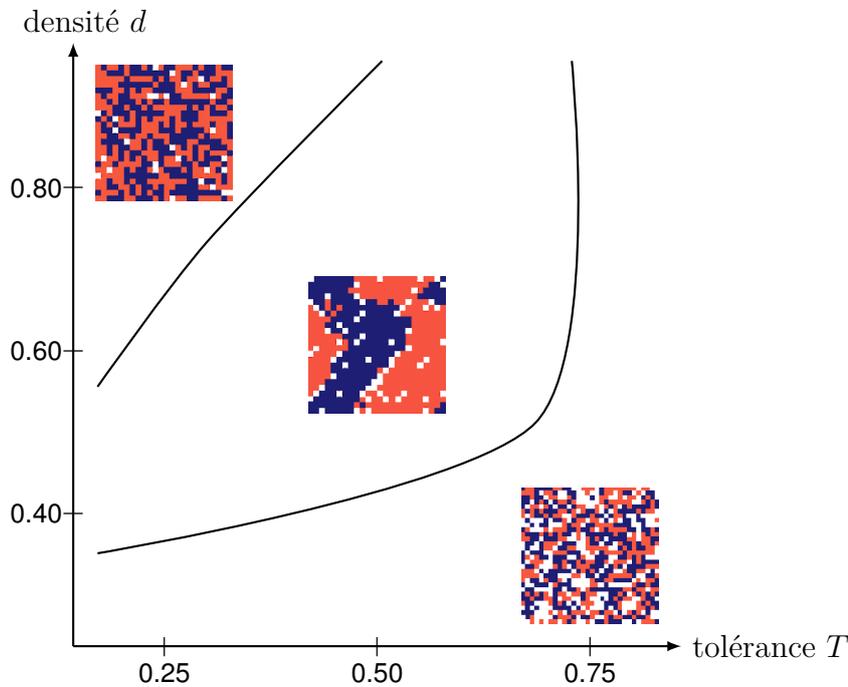


FIGURE 1.1 Représentation schématique du diagramme de phase (d, T) du modèle de Schelling établi par L. Gauvin et al. [70]. Les courbes représentées correspondent aux discontinuités d'observables physiques (énergie, chaleur spécifique, susceptibilité).

Les travaux de L. Gauvin et al. [70] ont permis d'établir le diagramme de phase (d, T) du système, ceux de N. Goulet-Dominic et al. [74] le diagramme de phase (p, T) , bien que celui-ci soit plus schématique du fait du peu de valeurs distinctes que peut prendre la tolérance dans un système sans site vide. On trouve ces diagrammes en figures **1.1** et **1.2** respectivement.

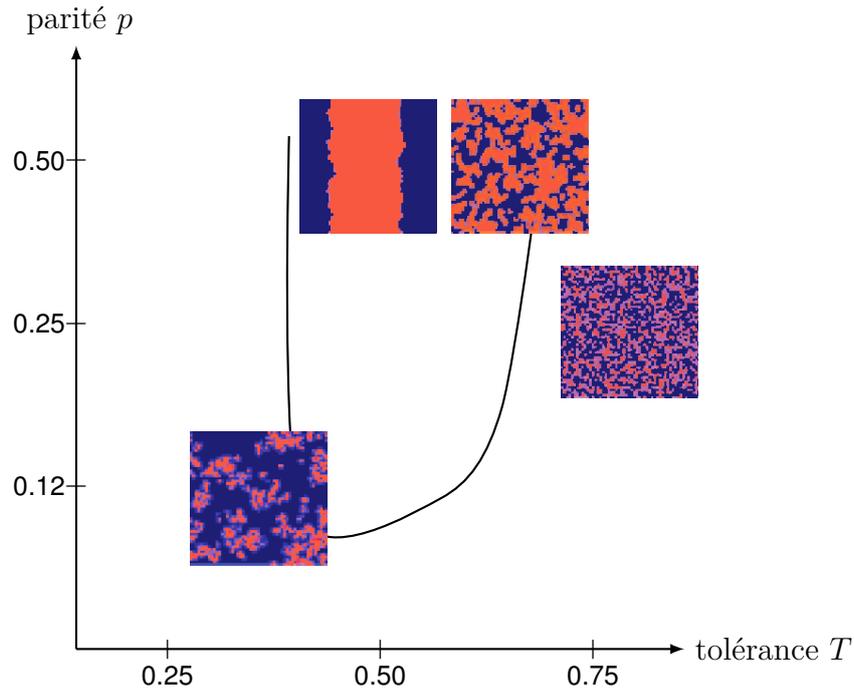


FIGURE 1.2 Représentation schématique du diagramme de phase (p, T) du modèle de Schelling établi par N. Goles Domic et al [74]. Les courbes représentées sont placées plus ou moins arbitrairement pour séparer les phases observées, l'article s'intéressant à l'évolution temporelle du système plus qu'à ses états stationnaires.

Selon [70], le diagramme de phase (d, T) peut être divisé en quatre familles d'états stationnaires : des états ségrégués au centre, des états gelés à forte densité et faible tolérance, des états mixtes lorsque la tolérance et la densité prennent des valeurs "proches" de 1, et une dernière catégorie, celle des états ségrégués "dilués". Dans cette dernière famille, les configurations stationnaires sont "visiblement" ségrégués à l'échelle du système mais l'observable utilisée, la taille des noyaux homogènes, n'est plus une bonne mesure de cette ségrégation car les sites vides sont majoritaires et séparent des petits noyaux homogènes de la même couleur. Il faut donc renormaliser la grille (par une procédure qu'on ne détaillera pas ici consistant à exclure une partie des sites vides entre noyaux de même couleur) pour mesurer la transition entre les phases diluées ségréguée d'une part et mixte de l'autre.

Du point de vue de la modélisation d'une ville, on travaillera principalement avec des densités supérieures à 50%, on peut donc laisser de côté cette dernière distinction pour considérer, à densité d constante, trois états stationnaires : un état gelé pour $T < T_f$, un état ségrégué pour $T_f < T < T_c$ et un état mixte pour $T > T_c$.

Ce même constat est établi dans [74] à l'aide du diagramme (p, T) : dans une population relativement équilibrée ($p \geq 10\%$), on observe également deux seuils de tolérance séparant des états nettement ségrégués d'états mixtes. On ne sait pas ici si ces derniers correspondent à un état gelé dans sa configuration initiale ou à un état en constante évolution à l'échelle locale mais globalement stationnaire.

Cette distinction microscopique dépend, on le verra, du modèle considéré et de ses règles d'évolution (déplacement vers des sites vides, échange ou inversion d'agents...), mais elle ne conditionne pas l'émergence ou non de la ségrégation, qui est un phénomène macroscopique. Cette robustesse aux variantes est une caractéristique notable, déjà mentionnée plus haut, du modèle de Schelling.

1.1.3 Universalité

Dans un article visant à unifier ces variantes dans un formalisme permettant une résolution analytique, T. Rogers et A. McKane [146] identifient quatre choix auxquels toutes les variantes étudiées dans la littérature doivent procéder : celui de la géométrie du système, celui des conditions initiales, celui de la mesure de la satisfaction des agents et celui des probabilités de transition d'une configuration à l'autre.

Afin d'intégrer les travaux réalisés dans la décennie qui sépare cette thèse de leur article de 2011, on explorera la littérature récente pour voir comment ont été traités ces quatre choix, auquel on ajoutera celui de l'observable permettant de quantifier la ségrégation du système à un instant donné. Les auteurs montrent ensuite que la tendance du système à évoluer vers des états ségrégués dans la gamme adéquate de paramètres est robuste malgré les variantes. Ils proposent alors une réécriture stœchiométrique du modèle permettant sa résolution analytique, cohérente avec les résultats des simulations. Cette méthode constitue le point de départ des travaux présentés au chapitre **2** et y sera détaillée.

Géométrie du système Le modèle de Schelling a été étudié sur de nombreux types de réseaux. A. Banos [12] a ainsi étudié l'influence des cliques sur l'émergence de la ségrégation. À cet effet, il a comparé l'état stationnaire atteint sur une grille classique avec ceux qui se dégagent sur des réseaux aléatoires, *scale-free* ou fractals. Définissant un état d'équilibre par la satisfaction totale des agents du réseau, il constate l'existence de trois phases :

un équilibre intégré pour $T > 0.7$, un équilibre ségrégré pour $0.3 < T < 0.7$, et enfin l'apparition de motifs stationnaires hors-équilibre, proches de l'état mixte initial lorsque $T < 0.3$. Ce constat est commun aux quatre géométries, mais la ségrégation est renforcée dans celles dont le coefficient de *clustering* [173] est le plus élevé.

C'est aussi un des premiers articles du corpus qui propose un graphe de la densité d'interface dans l'état stationnaire final, c'est à dire la proportion de paires hétérogènes d'agents voisins, en fonction de la tolérance. Ce graphe est reproduit en figure 1.3.

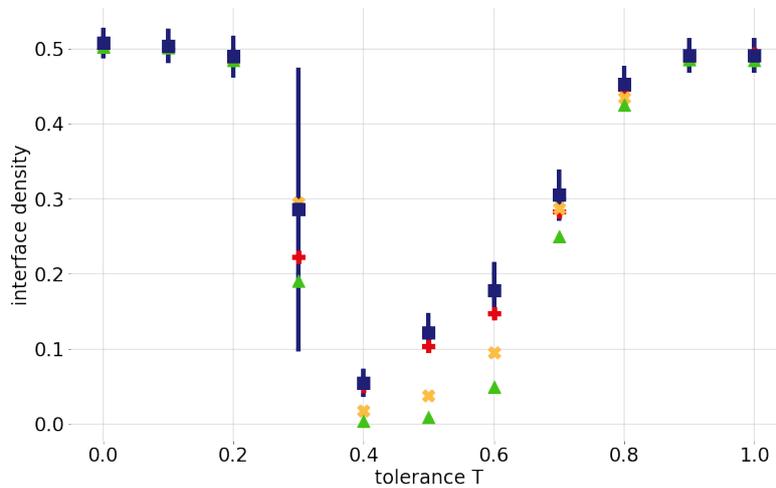


FIGURE 1.3 Densité d'interface $\langle x \rangle$ (moyennée sur 1000 simulations dans l'état stationnaire) en fonction de la tolérance T pour les quatre géométries considérées à $\kappa = 10$ (grille ■, aléatoire ■, *scale-free* ■, fractale ▲)³.

A.D. Henri et al. [86], V. Avetisov et al. [9] ainsi que C. Blex et T. Yasseri [22] s'éloignent encore plus du modèle original puisque les agents ne se déplacent plus : c'est le réseau qui se reconfigure en fonction de la satisfaction des agents présents sur les sites observés à chaque étape. Ainsi l'objectif vise plutôt à étudier l'auto-organisation de communauté que l'émergence de la ségrégation dans un réseau fixe.

Ces quelques travaux permettent d'explorer la dynamique du modèle de Schelling dans des géométries inhabituelles. En comparaison, les suivants ne s'intéressent qu'au cas d'une grille, mais en font varier le voisinage κ .

³Ce graphe est obtenu à partir de la figure 2 de l'article [12]. Dans un souci de clarté, seuls les écarts-types associés à la géométrie en grille sont représentés.

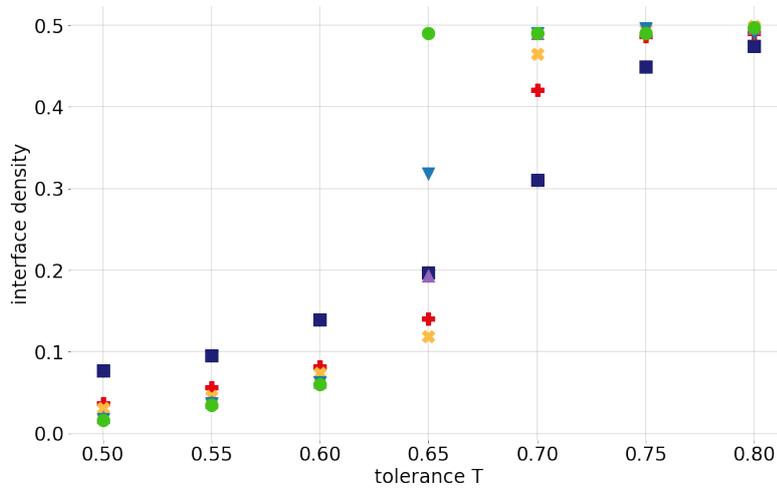


FIGURE 1.4 Densité d’interface $\langle x \rangle$ (moyenne réalisée dans l’état stationnaire) en fonction de la tolérance T pour des voisinages de rayon $r = 2$ (■), 3 (+), 4 (×), 5 (▲), 6 (▼) et 7 (●)⁴.

L’influence de la taille du voisinage est au cœur de l’article de A. Laurie et N. Jaggi [101] qui à cette variation près effectuent les mêmes simulations que dans l’article de L. Gauvin et al. [70]. Un agent sondant son voisinage pour déterminer sa satisfaction, il est naturel de penser qu’étendre son “champ de vision” pourrait jouer un rôle dans la dynamique du système. Les auteurs étudient le cas $T \geq 1/2$, dans l’esprit de Schelling, en considérant que les agents du système tolèrent un voisinage mixte. On n’observe alors que la seconde transition, repérée par la valeur T_c . Cette démarcation en terme de densité d’interface semble se préciser au fur et à mesure de l’accroissement de la taille du voisinage. D’une part, la ségrégation est plus importante lorsque r (et donc κ) augmente : les valeurs de la densité d’interface avant la transition sont plus proches de 0. D’autre part, le passage à la phase mixte est plus abrupt, comme on peut le voir en figures **1.4** et **1.5**. Il semblerait que la valeur de T_c soit légèrement décroissante en fonction de κ , mais cette hypothèse est à considérer avec précaution au vu de la faible quantité de points de données et de l’amplitude des incertitudes.

⁴Ce graphe est obtenu à partir de la figure 7 de l’article [101]. La densité d’interface n’est mesurée que pour des valeurs de T de la forme $n/20$ avec n un entier entre 10 et 16, quelle que soit la taille κ du voisinage, c’est pourquoi les incertitudes sont très grandes. La décroissance de la fonction $\kappa \mapsto T_c(\kappa)$ est donc discutable et pourrait demander une nouvelle étude plus précise, avec T de la forme n/κ par exemple.

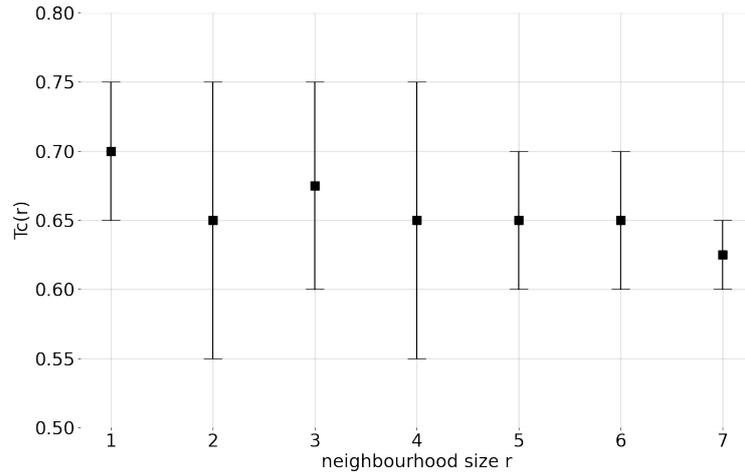


FIGURE 1.5 Évolution de la tolérance critique T_c en fonction de la taille du voisinage $\kappa = 2r(r + 1)$, à partir de la figure 1.4.

Le cas limite du voisinage infini, ou continu, est étudié par D. Vinković et A. Kirman [172], sans toutefois mesurer la densité d’interface ou une observable plus ou moins équivalente, ni explorer toutes les valeurs de T pour repérer d’éventuelles transitions.

Enfin, S. Grauwin et al. [77–80] considèrent des voisinages dits “bornés”. Non-centré sur l’agent considéré, un voisinage borné permet de modéliser une ville découpée en quartiers par des frontières naturelles (rivières...) ou artificielles (routes, voies ferrées, parcs...). Ces frontières sont poreuses en terme de déplacement des agents mais contraignent la mesure de la satisfaction aux individus du quartier, et non aux plus proches voisins. Il n’est pas possible de comparer les résultats de ces travaux avec les précédents pour plusieurs raisons : la tolérance n’y a pas d’équivalent, l’utilité est piquée au lieu d’être en échelon (ce point sera discuté dans la suite d’un des paragraphes suivants), et les échanges sont moins contraints par l’ajout d’une “température” dans le système (ce point est traité en section 3.2). On peut cependant constater que quand il y a ségrégation, celle-ci se fait par bloc, transformant la ville en un ensemble de noyaux homogènes délimités par ces frontières arbitraires.

Conditions initiales Dans le modèle de Schelling original [153], la ville est parsemée de sites vides, permettant le déplacement d'agents insatisfaits par leur voisinage. Comme montré dans la section précédente, les agents étant de deux types, on peut alors construire deux grandeurs, la densité d et la parité p , permettant de rendre compte macroscopiquement de cette configuration initiale. Les diagrammes de phase établis par L. Gauvin et al. [70] et N. Goles Domic et al. [74] représentés en figures **1.1** et **1.2** respectivement rendent compte des variations de comportement du système relativement aux conditions initiales.

L'article de N. Goles Domic et al. [74] s'affranchit totalement des sites vides, en laissant les agents échanger leur position dans le système. C'est principalement ce cas qui sera étudié dans cette thèse. En effet, l'inexistence de sites vides pourrait être considérée comme une meilleure approximation de la réalité des villes modernes, on peut prendre l'exemple de Paris dans laquelle les logements vacants constituent entre 5 et 10 % du parc immobilier, dont à peine 1.3 % sur une durée supérieure à deux ans [7]. De plus, le traitement mathématique est facilité par ce raccourci qui consiste à négliger la durée de vacance du logement.

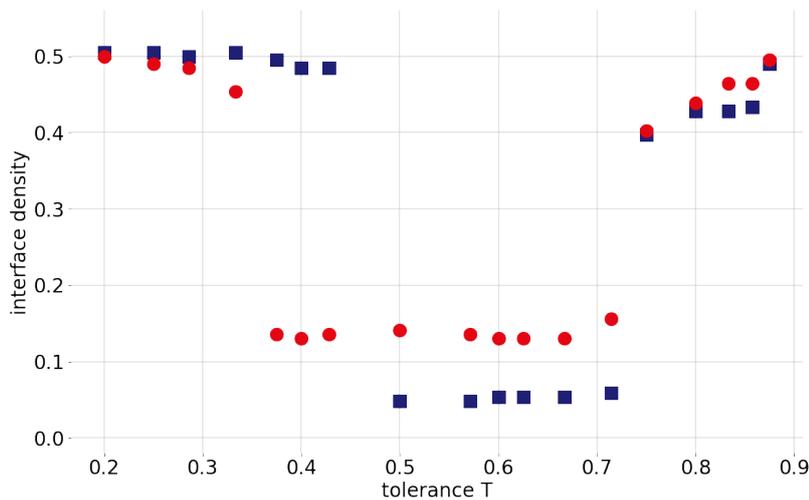


FIGURE 1.6 Densité d'interface $\langle x \rangle$ (moyennée sur 30 000 pas de temps après l'obtention de l'état d'équilibre) en fonction de la tolérance T pour $d = 0.98$ (■) et $d = 0.88$ (●) pour $\kappa = 8$.⁵

⁵Ce graphe est obtenu à partir des données de la figure 7 de l'article [70].

Il convient de noter cependant que la limite $d \rightarrow 1$ de la variante étudiée par L. Gauvin et al. [69–71] ne correspondra pas au cas sans site vide, comme on peut le constater sur la figure **1.6**. En effet, bien que le comportement général de la densité d’interface en fonction de la tolérance soit similaire, les valeurs prises par T_f et T_c sont légèrement différentes de celles prises dans les autres travaux, ce qui sera expliqué en section **2.3.2**. Par ailleurs on peut remarquer que les deux transitions sont bien plus marquées. Il ne s’agit pas d’une caractéristique liée à cette variante à proprement parler, mais d’une conséquence du processus de renormalisation utilisé par les auteur·rice·s.

Utilité : mesurer la satisfaction On emprunte à l’économie la notion d’utilité, fonction quantifiant le “bien-être” ou la satisfaction d’un agent donné en fonction des paramètres du système et de l’influence de ses voisins. Sans unité et restreinte à l’intervalle $[0, 1]$, elle dépendra dans toutes les variantes du modèle de Schelling du nombre ou du ratio de voisins similaires. Ainsi, l’utilité de Schelling est une fonction échelon : un agent est insatisfait tant que le ratio de voisins similaires dans son voisinage est inférieur strictement à $(1 - T)$ (ce qui revient bien à la définition de la tolérance : le ratio maximal toléré d’agents de l’autre groupe est T).

On peut résumer les utilités couramment en usage parmi les variantes présentées ici par le schéma suivant, où la satisfaction est affine par morceaux :

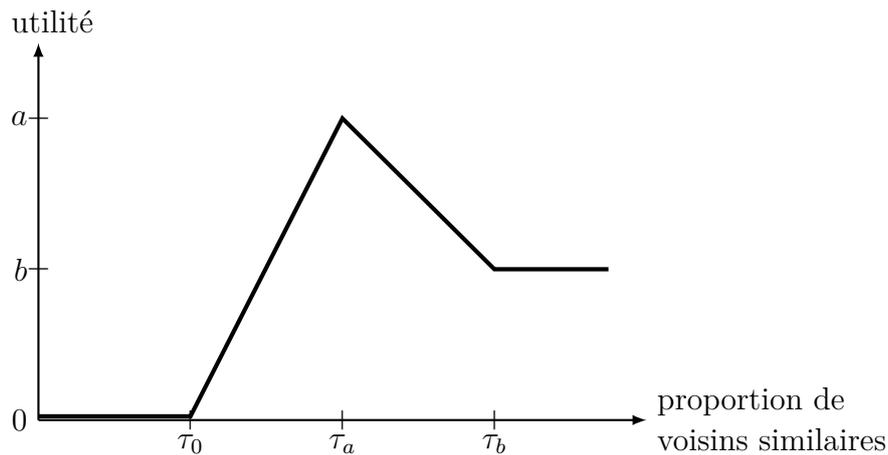


FIGURE 1.7 Schéma de l’utilité “généralisée” utilisée dans le cadre des variantes du modèle de Schelling.

Le modèle de Schelling correspond ainsi à $\tau_0 = \tau_a = \tau_b = 1 - T$ et $a = b = 1$, mais on peut imaginer des situations où la mixité est préférée.

rée localement à un voisinage parfaitement homogène ($a > b$), voire une utilité piquée, où $\tau_0 = \tau_a = \tau_b$ et $a > b = 0$, où la mixité est préférée à toute autre configuration. Enfin, on peut imaginer une utilité strictement croissante [138], où $\tau_0 = 0$ et $\tau_a = \tau_b = 1$.

La robustesse du modèle de Schelling s’illustre ici par le fait que l’utilité, quelle que soit sa forme, permet toujours d’obtenir des états ségrégués pour une certaine gamme de ses paramètres, comme l’ont montré par des simulations R. Pancs et N.J. Vriend [133] ainsi que S. Grauwin et al. [77–80] et J. Zhang [179, 180], du moins lorsque qu’une certaine agitation thermique⁶ est introduite dans le système.

Dans un article récent, A.P. Vieira et al. [171] expérimentent des utilités plus originales pour tester les limites de cette robustesse. Ici, seuls certains types de voisinages sont satisfaisants (par exemple, avoir 1, 3 ou 4 voisins similaires parmi 4). Et en effet, bien que la ségrégation soit constatée dans cet exemple particulier, d’autres comportements émergent, qualifiés d’éponge (2 ou 3 similaires), d’échiquier (0, 1 ou 3 similaires) voire de chaotique (2 similaires uniquement — on retrouve ici le cas piqué avec $\tau = 1/2$).

Comme pour l’introduction de sites vides, on peut se demander quelle réalité cette utilité permet de modéliser, et plus précisément quelle forme d’utilité serait le meilleur compromis entre réalisme sociologique et simplicité mathématique, avec la perspective d’en tirer une solution analytique. La comparaison avec des données réelles recueillies par W. Clark et M. Fossett [37] est menée en introduction du dernier chapitre, en page **113**, mais on peut d’ores et déjà identifier deux formes répondant au critère susmentionné. L’utilité linéaire, pour laquelle une solution analytique existe [138], présente l’inconvénient de supprimer la tolérance comme paramètre pertinent, et empêche donc d’étudier les transitions associées. L’utilité en échelon, introduite par Schelling, et qui semble de fait pouvoir modéliser les préférences réelles des agents : ceux-ci sont aussi satisfaits de vivre dans un voisinage parfaitement homogène que dans un voisinage mixte, jusqu’au seuil repéré par cette valeur de tolérance. L’équivalence entre utilité et “bien-être” ou “bonheur” est donc discutable, puisque comme on le verra, les données empiriques indiquent plutôt une préférence pour la mixité, par rapport à l’isolement de façon marquée, mais aussi par rapport à ce qui serait du regroupement communautaire exclusif.

⁶Se référer à la section **3.2**.

Inertie et non-localité Quelle que soit la forme choisie pour l'utilité, dans l'esprit du modèle de Schelling, un agent ne se déplace qu'à condition que son utilité augmente. Dans le cas d'une utilité en échelon, un agent insatisfait devient alors nécessairement satisfait. Ce n'est pas le cas pour certaines autres formes, qui demandent alors de considérer si la variation d'utilité doit être strictement positive ou si un déplacement peut avoir lieu en conservant l'utilité de l'agent concerné (mais avec l'avantage potentiel d'augmenter celle d'un agent voisin, ou de libérer un site satisfaisant pour un autre agent du système). Cette formulation permet de comprendre pourquoi la première dynamique, qui impose la stricte croissance de l'utilité de l'agent considéré, est parfois appelée "myope". De meilleures configurations sont accessibles, mais elles ne sont pas atteintes faute d'opportunités, qui seraient ouvertes par ces mouvements à utilité constante. On parle alors d'un système à forte *inertie*. L'évolution est lente et le système peut atteindre un état stationnaire, où tout mouvement est impossible mais dans lequel la plupart des agents demeurent insatisfaits. En théorie des jeux, un tel état est appelé *myopic Nash equilibrium* (équilibre de Nash "myope") [133]. En physique, l'évolution d'un système à forte inertie est comparable à celle d'un solide ou d'un verre, tandis qu'un système autorisant des échanges triviaux (à utilité constante) est comparable à un liquide [172].

Une part notable de la littérature [20, 47, 54, 90, 114, 149, 159, 178–180] envisage même d'autoriser une petite proportion de déplacements à utilité décroissante. Ce choix peut paraître anodin voire souhaitable pour étudier un système ayant une propension à atteindre des états stationnaires loin d'un éventuel équilibre "non-myope" : introduire cette *agitation thermique* fluidifierait le système, comme le font déjà les échanges triviaux. Pourtant, comme on le verra suite aux résultats du chapitre **2**, et plus techniquement dans le chapitre **3**, cette variante *perturbative* du modèle de Schelling n'a, à priori, aucune raison de décrire le comportement du modèle "non-perturbé", au sens d'une éventuelle "continuité" lorsque la température tend vers zéro. La nuance est soulignée par H.P. Young [176] qui distingue ainsi les états *absorbants* (au sens des chaînes de Markov) des états *stochastiquement stables*⁷.

Une caractéristique remarquable des systèmes dynamiques stochastiques est que leur comportement aux temps longs (asymptotique)

⁷Se référer à l'annexe **A** pour une définition plus précise.

*totique) peut être extrêmement différent de celui du processus déterministe correspondant, et ce **qu’importe l’amplitude de la perturbation.***⁸

Au cours d’une dynamique non-perturbée, le système peut terminer son évolution dans un état absorbant. Cet état stationnaire ne sera pas nécessairement optimal, au sens d’une mesure, appelons-la énergie, effectuée sur le système. Ainsi, l’évolution d’une copie du système pourrait lui faire atteindre un autre état absorbant de moindre énergie, de même pour un système analogue mais préparé dans des conditions initiales différentes. Les états stochastiquement stables sont en revanche ceux correspondant aux plus basses valeurs du potentiel⁹. Plus physiquement, on peut considérer que les états absorbants du processus non-perturbé sont des équilibres *mé-tastables* de la variante perturbative, des minima locaux de l’énergie, tandis que les états stochastiquement stables représenteraient les équilibres stables de ce même processus, les minima globaux de son énergie. N’étudier que les états d’équilibre de cette variante peut alors conduire à négliger une grande partie des équilibres stationnaires de la dynamique originelle.

L’inertie du système peut également être causée par ce que D. Vinković et A. Kirman [172] appellent son taux de *diffusion* : si la recherche d’un site satisfaisant se fait par proximité, en sélectionnant en priorité les sites voisins, le système est faiblement diffusif, si la recherche a lieu dans le système entier sans considération de distance, alors il est fortement diffusif. Un système fortement diffusif comporte alors de nombreux couplages à longue portée, ce qui entraîne sa *non-localité*.

*La dynamique [de Schelling] est non-locale, au sens où un automate cellulaire est régi par des règles locales. On peut considérer un tel modèle comme un automate cellulaire avec migration.*¹⁰

⁸*A remarkable feature of stochastic dynamical systems is that their long-run (asymptotic) behavior can differ radically from the corresponding deterministic process **no matter how small the noise term is.*** [176], p. 47 [traduction libre].

⁹La définition de Young du potentiel stochastique ne correspond pas à celle de ce même terme en physique : il caractérise la “résistance” des transitions du processus non-perturbé, c’est-à-dire le taux auquel les mêmes transitions dans la variante perturbative convergent vers ces dernières.

¹⁰(...) *The dynamics is not local, in the sense that a cellular automata is determined by local rules. In fact, these models can be thought of as cellular automata with migration.* [138], p.18 [traduction libre].

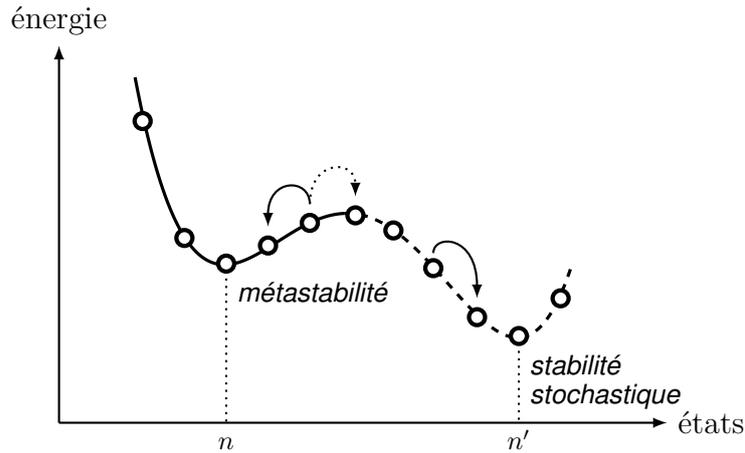


FIGURE 1.8 Schéma du paysage énergétique. Dans le modèle non-perturbé, les transitions autorisées font nécessairement décroître l'énergie (\longrightarrow). Dans ce cas, selon l'état initial, l'état d'équilibre du système sera n ou n' : ces états sont des attracteurs locaux de la chaîne de Markov. Dans la variante perturbative, d'autres transitions sont autorisées (\dashrightarrow). On peut alors distinguer les équilibres précédents : l'état n est dit *métastable* et l'état n' est *stable stochastiquement*.

En plus du lien avec les automates cellulaires [35, 115] établi par M. Pollicott et H. Weiss [138], divers domaines de recherche explorent des systèmes et des dynamiques qui s'apparentent au modèle de Schelling, on peut ainsi citer l'étude de la tension de surface, dans le cadre des phénomènes de nucléation [21, 31] ou des instabilités hydrodynamiques¹¹, ou encore la morphogénèse en biologie [117, 162, 168] et plus généralement les systèmes de réaction-diffusion [170]. Dans la plupart des cas cependant, ces phénomènes font intervenir des interactions à courte portée, et les motifs obtenus résultent de ces interactions locales et/ou des conditions aux limites. Il convient alors d'être attentif à la non-localité de la dynamique de Schelling lors de l'établissement de ces liens, en particulier pour utiliser des résultats de cette littérature.

Pour contourner cette difficulté, plusieurs auteurs [14, 15, 88, 123–125, 127] suppriment simplement le déplacement de leur étude. Au lieu de se rendre sur un site inoccupé, un agent insatisfait change simplement de groupe, ce qui ne le rend pas nécessairement satisfait pour autant. En physique statistique, cette dynamique est dite de Glauber [72], par opposition à une dynamique de Kawasaki [96]. On verra en section **3.3** en quoi les ré-

¹¹Parmi les instabilités hydrodynamiques faisant intervenir la tension superficielle, on peut citer celles de Rayleigh-Taylor [87] et de Bénard-Marangoni [116].

sultats analytiques établis dans ces travaux offrent une solution pertinente à l'étude des transitions entre états ségrégués et états intégrés.

Observables : mesurer la ségrégation La notion d'observable vient de la mécanique quantique, elle a été introduite en 1985 par W. Heisenberg [85] pour désigner toute fonction définie sur l'ensemble des états possibles du système et mesurant une quantité physique.

L'une d'entre elles a déjà été mentionnée pour le modèle de Schelling, il s'agit de la densité d'interface x , dénombrant le ratio de paires hétérogènes non-dirigées de voisins sur toutes les paires non-dirigées de voisins du système¹². C'est elle que l'on retrouve en ordonnées des figures **1.3**, **1.4** et **1.6**, bien que tous les travaux présentés [12, 70, 101], mais également d'autres tels que [40, 47, 74, 83, 138], ne l'aient pas nécessairement utilisée sous cette forme. Elle est en effet liée à l'insatisfaction moyenne des agents du système renormalisée. Cette insatisfaction présente des similarités avec l'énergie d'un modèle de spins d'Ising, elle est donc parfois appelée énergie du système. Il se trouve également qu'elle est reliée, pour des petites valeurs de κ , à la longueur totale de l'interface séparant les noyaux homogènes [138]. Malgré ces "variantes", fonctions affines de la densité d'interface, c'est celle-ci que l'on a choisi pour deux raisons qui seront développées dans la section **1.2.3** : sa normalisation et son rôle de paramètre d'ordre de la transition de phase.

Pour autant, ce n'est pas la seule façon de mesurer la ségrégation. Plusieurs de ses indicateurs seront présentés dans le chapitre **4**. Certains d'entre eux sont utilisés pour le modèle de Schelling, comme les indices de Duncan et de Freeman [41] ou plus simplement la distribution des tailles¹³ des noyaux homogènes, ou clusters, que ce soit par le biais de simulations [70, 133] ou analytiquement [14–17, 26, 88, 123–125, 127].

Comme pour l'utilité, ce choix arbitraire ne semble pas modifier qualitativement les résultats, bien que certaines observables permettent une démarcation plus nette entre les phases du système¹⁴. De nouveau, un compromis est à trouver entre l'aisance mathématique et le coût computationnel ainsi que la qualité des simulations.

D'un point de vue analytique, les analogies qui existent entre certaines

¹²Pour une définition formelle, se référer à la section **1.2.3**.

¹³Ici, la taille d'un noyau ne fait pas référence à une longueur ou une surface mais au nombre d'agents le constituant.

¹⁴Voir à ce sujet la renormalisation proposée par L. Gauvin et al. [70], p. 296.

observables et d'autres grandeurs physiques, comme l'énergie d'un système de spins en physique statistique, permettent de réutiliser techniques et résultats issus des domaines concernés. Il faut toutefois garder à l'esprit que, la parenté qui peut être établie entre le modèle de Schelling et des modèles, comme celui d'Ising, issus de la physique n'est pas une correspondance stricte. Ces notions seront abordées dans le chapitre **3**.

1.2 | Lexique et outils d'analyse

LE SYSTÈME est régi par un nombre fini et restreint de paramètres. À l'instant initial, les agents¹⁵ sont répartis aléatoirement en deux groupes, appelés \oplus et \ominus , sur un réseau de N sites, formant la configuration $\sigma_0 = (s_1, s_2, \dots, s_N)$, où $s_i \in \{-1, 0, 1\}$.

1.2.1 Un système à deux échelles

Une grande partie de la littérature travaillant sur une grille carrée, voici une présentation des voisinages caractéristiques que l'on y retrouve. Lorsque l'on considère les quatre voisins à proximité immédiate, on parle de description de von Neumann [115]. En économie, la description de Moore [113] est plus souvent utilisée [133], où l'on compte également les voisins diagonaux, portant κ à huit. Il s'agit des deux types de voisinages les plus courants pour étudier le modèle de Schelling, ce dernier, économiste, l'ayant d'ailleurs construit avec le formalisme de Moore. On peut généraliser ces définitions : le voisinage de von Neumann étendu du site i est

$$\partial i_{\diamond}(r) = \left\{ j \in G \mid 1 \leq \|\mathbf{x}_j - \mathbf{x}_i\|_1 \leq r \right\} \quad (1.1)$$

faisant intervenir la norme de Manhattan $\|\cdot\|_1$. Il contient $\kappa = 2r(r + 1)$ sites et le voisinage de von Neumann usuel correspond au cas $r = 1$.

Similairement, le voisinage de Moore étendu du site i est

$$\partial i_{\square}(w) = \left\{ j \in G \mid 1 \leq \|\mathbf{x}_j - \mathbf{x}_i\|_{\infty} \leq w \right\} \quad (1.2)$$

faisant intervenir la norme de Tchebychev $\|\cdot\|_{\infty}$. Il contient $\kappa = 4w(w + 1)$ sites et le voisinage de Moore usuel correspond au cas $w = 1$.

Il existe des versions unidimensionnelles¹⁶ [14, 16, 17, 26, 47, 176] du modèle de ségrégation de Schelling. Les définitions des deux voisinages ci-dessus sont alors confondues, et on a alors simplement $\kappa = 2r$.

De manière générale, on notera ∂i l'ensemble des voisins du site i , de cardinal quelconque κ .

La structure en grille du réseau, introduite par Schelling, n'est pas essentielle à l'analyse théorique qui suit. Le modèle de Schelling peut en effet être étudié sur tout graphe G connexe non-pondéré et non-orienté. Lorsque

¹⁵Qui sont plutôt appelés "particules" ou "spins" dans la littérature de physique.

¹⁶La première a d'ailleurs été proposée par Schelling antérieurement à la version étudiée ici.

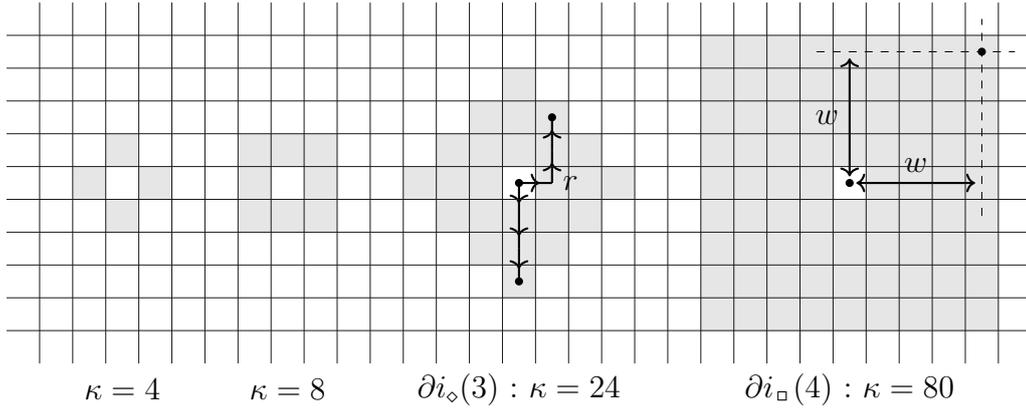


FIGURE 1.9 Voisinages usuels de von Neumann, de Moore et deux exemples de voisinages étendus, utilisés par exemple dans [101] et [15] respectivement.

chaque sommet du graphe a le même degré κ , on appelle G un graphe κ -régulier ; on dit de façon équivalente que chaque agent est muni d'un voisinage de taille κ .

Dans un graphe régulier, N et κ suffisent à décrire les deux échelles caractéristiques auxquelles sont soumis les agents du système. Pourtant, la complexité de la répartition des populations dans le tissu urbain et sa perception sont plutôt multiscalaires. C'est une des limites du modèle de Schelling et de nombreux indices de mesure, qui sera explorée dans le chapitre 4.

1.2.2 Quantités locales et satisfaction

À l'échelle du voisinage, on définit deux quantités, empruntées à [70], que l'on appellera le voisinage effectif local κ_i et l'aimantation locale m_i . Si le nom de la première renvoie directement à l'étude du réseau, celui de la seconde provient de la littérature des verres de spins et est relié au champ généré par le voisinage de l'agent s_i .

Là où ∂i recense les liens entre les *sites* du réseau, le degré effectif local κ_i compte les liens entre *agents* du système en considérant les sites vides comme contenant des agents de spin $s_j = 0$. Il est donc défini par le nombre de sites voisins occupés :

$$\kappa_i = s_i^2 \sum_{j \in \partial i} s_j^2 \tag{1.3}$$

et prend par construction des valeurs dans $\llbracket 0, \kappa \rrbracket$. Il est à noter que la forme quadratique des spins qui apparaît est parfois remplacée par une valeur absolue ou une fonction signe, ce qui revient au même lorsque $s_i \in \{-1, 0, 1\}$.

L'aimantation locale m_i est analogue à l'interaction ferromagnétique d'un modèle d'Ising [89]. Cela signifie que les interactions entre agents du même groupe sont comptées positivement, et les autres négativement : l'influence du voisinage entier est donc la somme des interactions individuelles avec l'agent i . L'aimantation est une grandeur reliée au champ local h_i , et ces deux quantités sont définies par

$$m_i = s_i \sum_{j \in \partial i} s_j = s_i h_i \quad (1.4)$$

Évoluant dans l'intervalle $[-\kappa_i, \kappa_i]$, l'aimantation sera positive si une majorité d'agents voisins de i appartiennent au même groupe, négative sinon.

On peut définir ces deux quantités m_i et κ_i de façon moins formelle en fonction des nombres d'agents similaires n_s (du même groupe) et différents n_d (de l'autre groupe) dans le voisinage : on a alors simplement

$$\kappa_i = n_{s,i} + n_{d,i} \quad \text{et} \quad m_i = n_{s,i} - n_{d,i} \quad (1.5)$$

ce qui est équivalent à

$$n_{s,i} = \frac{\kappa_i + m_i}{2} \quad \text{et} \quad n_{d,i} = \frac{\kappa_i - m_i}{2} \quad (1.6)$$

Dans la première version du modèle, Schelling mesure la satisfaction d'un agent i à l'aide du ratio d'agents du groupe opposé parmi les κ_i agents présents dans son voisinage. Si ce ratio excède $1/2$, alors l'agent sera insatisfait.

La notion d'utilité introduite dans la section précédente va permettre de formaliser la satisfaction des agents au regard de leur voisinage. Cette fonction, normalisée sur $[0, 1]$ où 0 correspond à l'insatisfaction maximale et 1 à la satisfaction maximale, est déterminée à partir des grandeurs locales définies précédemment par

$$u_i(\boldsymbol{\sigma}) = \mathbb{1}_{[(1-2T)\kappa_i, \kappa_i]}(m_i) \quad (1.7)$$

On peut vérifier sans peine grâce à l'expression (1.6) que comme attendu, l'agent i est satisfait ($u_i(\boldsymbol{\sigma}) = 1$) lorsque $0 \leq n_{d,i}/\kappa_i \leq T$.

Ces quantités locales sont définies pour tout agent sur tout type de graphe et ce sont elles qui vont permettre de déterminer si un agent peut se déplacer ou non dans le cadre de la dynamique de Schelling. Elles sont amenées à évoluer à tout instant. Ce n'est pas le cas, suivant les variantes considérées, des deux quantités globales déjà rencontrées dans les diagrammes

de phase **1.1** et **1.2** : la densité d et la parité p , qui jouent ici le rôle de paramètres de contrôle pour l'étude des éventuelles transitions de phase du système.

1.2.3 Paramètres de contrôle, paramètre d'ordre

À un instant donné, une fraction $d \in [0, 1]$ des sites est occupée par un agent de l'un ou de l'autre type et parmi ces Nd agents, une fraction $p \in [0, 1]$ est de type \oplus . On peut calculer ces deux quantités à partir de σ :

$$d = \frac{1}{N} \sum_{i=1}^N s_i^2 \quad ; \quad p = \frac{1}{2N} \sum_{i=1}^N (s_i + s_i^2) \quad (1.8)$$

Dans un système dit fermé, ces deux quantités sont des constantes : les agents ne font que se déplacer de site en site, sans qu'il y ait de perte ou de création. On pourra alors utiliser les équations (1.8) comme des lois de conservation, avec d et p définis à partir de la configuration initiale σ_0 . On parle dans ce cas de *paramètres de contrôle*.

Dans un système ouvert en revanche, des agents peuvent entrer ou sortir du système à chaque pas de temps, rien ne garantit alors la conservation de la densité d et de la parité p . Une partie de la littérature choisit cependant de se priver d'au moins une de ces lois de conservation pour contourner une caractéristique essentielle du modèle de Schelling : sa *non-localité*¹⁷. Plutôt que d'étudier un système multiscalaire, pour lequel les distances d'échanges entre sites sont comprises dans un immense intervalle allant de l'échelle du voisinage κ et celle du système entier N , dans ces travaux, les agents changent de groupe. On peut envisager cette modification comme un déplacement de l'agent hors de la ville et un remplacement immédiat par un agent extérieur, comme si le système était en interaction avec un "réservoir" d'agents de chacun des groupes. Les deux seules échelles pertinentes à l'étude redeviennent le voisinage d'une part et le système entier d'autre part. Par analogie avec l'étude des systèmes de spin, on parle dans ce cas de variante de Glauber [72], par opposition à des modèles d'échanges d'agents au sein du système, qui se rapprochent des travaux de Kawasaki¹⁸ [96, 97], ou à des déplacement vers des sites vides, qui, même s'ils sont très similaires

¹⁷Se référer au paragraphe *Inertie et non-localité* en section **1.1.3**.

¹⁸À la différence notable que les échanges dans la dynamique de Kawasaki impliquent généralement des sites voisins. On parlera ici de dynamique de Kawasaki même si la portée de l'échange est supérieur à la taille κ du voisinage.

à la variante précédente¹⁹, s'approchent peut-être davantage des modèles de diffusion de lacunes dans un gaz sur réseau [65, 99].

D'un point de vue analytique, cette dernière option (qui correspond pourtant au modèle originel de Schelling) est plus difficile à mettre en œuvre. Déplacer un agent vers un site vide *satisfaisant* impose de pouvoir dresser des listes des sites vides satisfaisants du point de vue de chacun des deux groupes d'agents, et à fortiori de pouvoir évaluer leur taille. Par ailleurs, il n'est pas garanti, surtout pour des densités proches de 1, que ces listes soient non-vides, ce qui aurait pour effet de bloquer la dynamique bien avant d'atteindre un éventuel équilibre.

À partir des quantités locales, il est donc possible de définir des observables qui jouent le rôle de paramètres de contrôle déterminés par l'expérimentateur, en l'occurrence ici par la préparation de la configuration initiale σ_0 . Il est aussi possible de construire des observables qui permettent au contraire de suivre l'évolution du système. Une telle fonction est appelée *paramètre d'ordre* lorsque, par convention, elle s'annule dans la phase "ordonnée" (ici, la phase ségrégée) alors qu'elle prend des valeurs quelconques dans la phase "désordonnée" (ici, l'état intégré).

Dans la dynamique de Schelling, lorsqu'il y a ségrégation, le système se structure en formant des noyaux homogènes, c'est-à-dire contenant des agents d'un seul des deux groupes. Le ratio d'agent du même groupe qu'un agent donné au sein de son voisinage tend alors vers 1, et inversement le taux de contact avec l'autre groupe vers 0. La densité d'interface x , qui est définie comme la proportion de paires dirigées *hétérogènes* d'agents voisins $n_{\bullet\circ}$ parmi toutes les n_κ paires dirigées d'agents voisins présentes dans le système, peut alors être considérée comme un paramètre d'ordre de cette transition de phase. La relation (1.6) permet alors d'établir les identités suivantes :

$$x = \frac{n_{\bullet\circ}}{n_\kappa} = \frac{\sum_i n_{d,i}}{\sum_i \kappa_i} = \frac{1}{2} \left(1 - \frac{\sum_i m_i}{\sum_i \kappa_i} \right) \quad (1.9)$$

Cette quantité n'est pas exactement similaire à la densité de contact proposée par A. Hazan et J. Randon-Furling [83], définie par la moyenne sur

¹⁹Par équivalence entre un déplacement vers un site vide et un échange avec un agent d'un troisième groupe "neutre" qui remplirait les sites vides.

tous les sites du rapport $n_{d,i}/\kappa_i$:

$$x_{\text{HRF}} = \frac{1}{N} \sum_i \frac{n_{d,i}}{\kappa_i} = \frac{1}{2} \left(1 - \frac{1}{N} \sum_i \frac{m_i}{\kappa_i} \right) \quad (1.10)$$

Les deux expressions sont confondues lorsque $d = 1$ car le dénominateur est constant. La première présente l'avantage de pouvoir être reliée à une quantité "physique", l'énergie, qui sera définie dans la section suivante. Une interprétation géométrique de la densité d'interface est esquissée par N. Goles Domic, E. Goles et S. Rica [74]. Pour un voisinage de von Neumann, la densité d'interface est en réalité une mesure du périmètre normalisé de l'interface entre les noyaux homogènes. Pour un voisinage de Moore, le périmètre intervient également mais il faut retrancher les liaisons "en coin". Le schéma suivant permet de mieux appréhender ces correspondances :

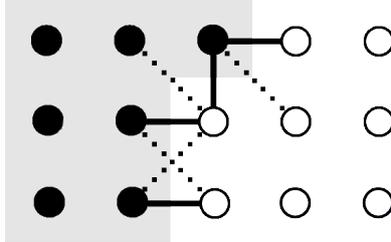


FIGURE 1.10 Schéma de l'interface entre deux noyaux homogènes avec distinction entre les paires hétérogènes d'agents voisins pour le calcul de la densité d'interface en fonction du voisinage.

La densité d'interface correspond, dans ce cas particulier $d = 1$, à

$$x_{\diamond} = \frac{n_{\bullet\text{--}\circ}}{2N} \quad ; \quad x_{\square} = \frac{3n_{\bullet\text{--}\circ} - n_{\bullet\text{--}\circ}}{2N} \quad (1.11)$$

pour les voisinages de von Neumann et Moore respectivement.

1.3 | La ségrégation : un état d'équilibre ?

SCHELLING dans son modèle introduit un système complexe constitué d'un grand nombre d'agents dont les interactions locales ont des conséquences globales : “*d'intentions microscopiques [naissent] des comportements macroscopiques*”²⁰ pouvant parfois sembler paradoxaux. Il entre alors tout naturellement dans le domaine de la physique statistique qui fournit un grand nombre d'outils pour en décrire les propriétés, et notamment via une approche énergétique. L'évolution temporelle des domaines magnétiques du modèle d'Ising et de ses nombreuses variantes se prête en effet à une analogie immédiate avec les caractéristiques du modèle. Ce rapprochement permet d'introduire des quantités mesurables dont la dérivation ne découle pas spécialement des sciences sociales (énergie, susceptibilité...). En revanche, il faut être conscient de ses limites, notamment l'absence d'équivalent, dans les systèmes magnétiques, à des restrictions spécifiques sur le déplacement des agents. Comme mentionné précédemment, une partie de la littérature sur le modèle contourne ce problème en introduisant une part d'aléatoire dans le déplacement des agents, plaçant ainsi leur analyse dans le cadre du formalisme canonique²¹. La conclusion de ces articles étudie alors généralement la limite nulle de cette température extérieure, et en tire des conclusions sur le modèle de Schelling. Cette approche pourrait s'avérer être la bonne, mais en l'état, rien ne justifie que cette limite corresponde au problème initial, du moins pas pour toutes les valeurs de tolérance. C'est pourquoi, pour clore le premier chapitre, ne seront présentés que les travaux proposant une résolution analytique de variantes du modèle de Schelling aux règles déterministes et dont seul le choix des agents subissant l'échange sera aléatoire.

Le cas unidimensionnel Dans son livre *Individual Strategy and Social Structure* publié en 1998, H.P. Young est le premier à en proposer un résultat analytique, pour une variante unidimensionnelle du modèle.

Théorème [176, 177] *Soit deux groupes distincts d'agents \oplus et \ominus répartis sur les N sites d'un anneau unidimensionnel. Chaque agent est muni d'un voisinage de taille $\kappa = 2$ et est considéré insatisfait si et seulement s'il n'est*

²⁰*Micromotives and macrobehavior*, T. Schelling, 1978 [154].

²¹Se référer au chapitre 3, section 3.2.

entouré que d'agents de l'autre groupe :

$$u_i(\sigma) = \mathbb{1}_{[0,2]}(m_i) \tag{1.12}$$

Un tel agent peut alors échanger sa position avec un autre agent, insatisfait ou non²². À l'équilibre, aucun agent n'est insatisfait.

Cette affirmation permet d'identifier les états absorbants de la chaîne de Markov associée à la dynamique de Schelling : dans ces états stationnaires, aucun agent n'est isolé (et inversement, une configuration où un agent est isolé ne peut être un état stationnaire). Sans démontrer l'émergence d'une ségrégation à grande échelle (on peut imaginer la configuration représentée en figure 1.11 pour s'en convaincre), Young prépare tout de même le cadre pour l'étude du modèle de Schelling à l'aide des chaînes de Markov.

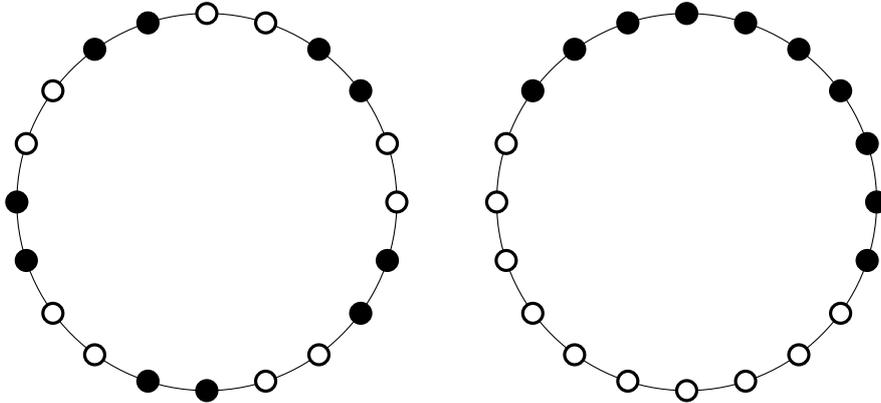


FIGURE 1.11 Un état d'équilibre non-ségrégé ($x = 1/2$) du modèle unidimensionnel non-perturbé (à gauche) et un état d'équilibre stable stochastiquement ségrégé ($x = 1/10$) du modèle unidimensionnel perturbé.

Dans ses autres contributions [176, 177], Young introduit des perturbations aléatoires à cette chaîne pour déterminer l'équilibre *stable stochastiquement*²³. Cet état d'équilibre, au sens thermodynamique, est ségrégé, c'est-à-dire que les agents de chaque groupe forment un noyau homogène continu. Toutefois, rien ne permet d'affirmer a priori qu'il est atteint par un système non-perturbé tel que celui de Schelling. De fait, les résultats démontrés par les auteurs et autrices suivant-e-s portent principalement sur des fonctions caractérisant la dynamique du modèle et non sur l'état d'équilibre atteint.

²²Young imagine une forme de compensation versée à cet autre agent si celui-ci est amené à se déplacer d'un site satisfaisant vers un nouveau site satisfaisant.

²³Se référer au paragraphe *Inertie et non-localité* en section 1.1.3.

Un problème d'optimisation Dans un article publié en 2001, M. Pollicott et H. Weiss développent un modèle simplifiant le modèle de Schelling en introduisant plusieurs variantes. Ils introduisent un indice de ségrégation, l'insatisfaction moyenne des agents, à savoir la moyenne du ratio d'agents voisins de la couleur opposée à celle de l'agent central et du nombre total de voisins, et se placent dans le cadre d'un système sans site vide. La tolérance est elle aussi revisitée : un agent se déplace à condition que sa satisfaction augmente.

Théorème [138] *Soit deux groupes distincts d'agents \oplus et \ominus répartis sur les N sites d'un tore bidimensionnel. Chaque agent est muni du voisinage de von Neumann et sa satisfaction est égale à l'aimantation normalisée m_i/κ .*

$$u_i(\boldsymbol{\sigma}) = m_i \quad (1.13)$$

Deux agents peuvent échanger leurs positions si leur satisfaction augmente suite à cet échange. Alors

$$\mathcal{L} = - \sum_i m_i \quad (1.14)$$

est une fonction de Lyapunov du système et possède une dépendance affine en le périmètre de l'interface.

M. Pollicott et H. Weiss sont les premiers à montrer un résultat analytique dans une variante bidimensionnelle du modèle de Schelling : sans tolérance, avec une utilité linéaire, les agents se déplacent jusqu'à atteindre un minimum local de l'insatisfaction moyenne \mathcal{L}/N . Si le déplacement coûte à l'un des deux agents de l'échange, celui-ci n'a pas lieu. Le système n'atteint donc pas forcément un minimum global de \mathcal{L} , mais à l'aide d'un certain nombre de réalisations, il est possible de s'en approcher, sauf à être absorbé par un état métastable, à partir duquel il est impossible d'explorer d'autres régions du paysage énergétique.

Le fait que l'insatisfaction moyenne soit liée à la longueur de l'interface entre noyaux homogènes ouvre la voie à une analogie entre cette variante de Schelling et la théorie de la nucléation [21, 31]. En effet, il est possible de reformuler ce problème en minimisation d'une surface d'échange (au sens général, ici l'interface est linéique) entre deux phases représentée par les deux groupes d'agents présents. Cette piste sera explorée dans la dernière section du chapitre **3**.

La limite de cette approche est la suppression de la notion de tolérance, cruciale dans le modèle de Schelling, et qui est un des paramètres permettant

notamment la distinction d’avec les modèles de spins de type Ising. Elle permet tout de même d’ouvrir la voie à l’exploration de cette approche énergétique.

Quelques modèles de spins En physique de la matière condensée, les modèles de spins constituent un socle théorique pour l’étude des milieux magnétiques et de leurs transitions de phase puis pour l’étude des transitions de phase de manière générale. Le reste de ce court paragraphe de présentation vise à présenter les modèles inspirés de celui d’Ising [89], qui présente l’intérêt de correspondre à la description qui a été faite du modèle de Schelling (modèle sur réseau, dont le contenu des sites est divisé en deux groupes...). Il en existe cependant une grande diversité, on peut citer le modèle de Potts [139] qui permet d’élargir le précédent à n groupes, le modèle XY qui est la limite continue du modèle de Potts, le modèle classique d’Heisenberg [84] que l’on peut envisager comme une version non-planaire du modèle XY²⁴...

Un des enjeux principaux de ces modèles de spins est de déterminer l’état d’équilibre thermodynamique du système, généralement appelé état fondamental, dans une géométrie donnée. Cet état est défini comme l’état de plus basse énergie²⁵, cette énergie étant issue des interactions entre les spins qui constituent le système ainsi que de contraintes imposées par certains paramètres de contrôle, externes (température Θ , champ magnétique H) ou internes (champ cristallin Δ). On l’exprime à partir du Hamiltonien

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} s_i s_j - K \sum_{\langle i,j \rangle} s_i^2 s_j^2 - H \sum_i s_i - \Delta \sum_i s_i^2 \quad (1.15)$$

où J et K décrivent les interactions d’échange bipolaires et quadrupolaires. Cet Hamiltonien correspond au modèle de Blume-Emery-Griffiths [24], les modèles antérieurs d’Ising [89] et de Blume-Capel [23, 32] correspondant respectivement aux cas $K = \Delta = 0$ et $K = 0$. Ils permettent d’établir des diagrammes de phase de la transition ferromagnétique/paramagnétique mais peuvent également modéliser la transition de phase liquide/superfluide d’un mélange binaire d’isotopes 3 et 4 de l’hélium.

On peut enfin noter que les deux derniers termes du Hamiltonien favorisent respectivement une orientation des spins (un des deux groupes

²⁴Cette présentation par généralisations successives est en grande partie anachronique.

²⁵L’état de plus basse énergie est l’état le plus probable dans le formalisme canonique, se référer au chapitre 3, section 3.2.

d'agents) par rapport à l'autre et une prééminence des sites vides par rapport aux sites occupés. Dans le cas d'un système fermé, ces deux quantités n'ont pas d'équivalent direct dans le modèle de Schelling, on considère donc pour la suite le cas $H = \Delta = 0$.

Premiers diagrammes de phase du modèle de Schelling Introduit par L. Gauvin, J.-P. Nadal et J. Vannimenus [70, 71] dans la littérature du modèle de Schelling, le modèle de Blume-Emery-Griffiths permet de construire une observable physique²⁶, qu'on appellera énergie du système, en remarquant, à partir des règles dynamiques du modèle, que pour des valeurs bien choisies de J et K , le Hamiltonien (1.15) est une fonction décroissante du temps.

Théorème [70, 71] *Soit dN agents séparés en deux groupes distincts \oplus et \ominus répartis sur les N sites d'une grille carrée. Chaque agent est muni du voisinage de Moore et est considéré satisfait lorsque son aimantation m_i est supérieure à $(1 - 2T)\kappa_i$, où κ_i est le nombre de sites occupés dans son voisinage.*

$$u_i(\boldsymbol{\sigma}) = \mathbb{1}_{[(1-2T)\kappa_i, \kappa_i]}(m_i) \quad (1.16)$$

Un agent insatisfait peut se déplacer vers un site vide s'il devient satisfait suite à cet échange. Alors

$$E_S = - \sum_i m_i - (2T - 1) \sum_i \kappa_i \quad (1.17)$$

est une fonction de Lyapunov du système.

Le reste de l'article [70] présente des résultats de simulations numériques et l'établissement du premier diagramme de phase (d, T) en figure 1.1 qui met en évidence les transitions détaillées au début de ce chapitre. Comparaison est également faite avec le diagramme de phase (Δ, Θ) du modèle de Blume-Capel mentionné dans le paragraphe précédent, toutefois il est noté par les auteurs que, d'une part, les quantités considérées ne sont pas analogues d'un modèle à l'autre²⁷ et, d'autre part, que l'ordre des transitions de

²⁶Cette construction physique est complémentaire à celle de la densité d'interface dont l'interprétation géométrique lorsque $\kappa = 4$ devient difficile à généraliser pour des voisinages plus grands.

²⁷Le champ cristallin Δ est une variable conjuguée de la densité d de sites occupés tandis que la température Θ une forme de mesure de la capacité des agents à se déplacer, ces déplacements étant autorisés en fonction de la tolérance T .

phase (c'est-à-dire la continuité ou non de la grandeur mesurée au voisinage de la transition) est inversé.

Dans un article publié un an plus tard, N. Goles-Domic, E. Goles et S. Rica [74] regroupent des configurations stationnaires dans ce qui peut s'apparenter au diagramme de phase (p, T) représenté en figure **1.2** d'une variante du modèle de Schelling pour laquelle tous les sites sont occupés ($d = 1$). La dynamique a donc lieu par échanges d'agents plutôt que par déplacement vers un site inoccupé. Dans ce cadre, le second terme du Hamiltonien (1.17) est constant : l'énergie E_S s'écrit alors exactement comme l'insatisfaction \mathcal{L} de M. Pollicott et H. Weiss.

Théorème [74] *Soit deux groupes distincts d'agents \oplus et \ominus répartis sur les N sites d'un graphe non-orienté quelconque. Chaque agent est muni d'un voisinage de taille κ_i (qui dépend de la géométrie du réseau) et est considéré satisfait lorsque son indice d'homogénéité locale m_i est supérieur à $(1 - 2T)\kappa_i$.*

$$u_i(\boldsymbol{\sigma}) = \mathbb{1}_{[(1-2T)\kappa_i, \kappa_i]}(m_i) \quad (1.18)$$

Un agent insatisfait peut alors échanger sa position avec un autre agent insatisfait. Alors $E = -\sum_i m_i$ est :

- *une fonction de Lyapunov du système si et seulement si $T > 1/2$;*
- *une fonction de Lyapunov p.s. lorsque $T = 1/2$ et $\max(\kappa_i) \ll N$.*

Dans les autres cas, la variation d'énergie causée par un échange est bornée mais peut être positive.

Ces deux résultats permettent d'affirmer que la méthode perturbative n'est pertinente ni dans le modèle original (déplacement vers un site vide satisfaisant) exploré par L. Gauvin et al. ni dans la variante par échanges de N. Goles-Domic et al. lorsque $T > 1/2$. En effet, introduire des perturbations reviendrait à autoriser certaines variations positives de l'énergie, entraînant la possibilité, à partir d'une configuration initiale $\boldsymbol{\sigma}_0$ donnée, d'explorer des états d'équilibre plus stables que les éventuels états absorbants rencontrés au cours de la même dynamique non-perturbée. Ici, rien n'indique en théorie que ces états absorbants sont ségrégués. L'approche perturbative n'est vraisemblablement pas à même de rendre compte des diagrammes de phase obtenus par simulations numériques. Nous nous tournons donc vers le formalisme markovien.

2. | Schelling

Classification des états de la chaîne de Markov

DANS ce chapitre nous présentons une réécriture markovienne du modèle de Schelling, en construisant la matrice de transition du processus, à partir d'un exemple simple que nous généralisons par la suite. Après le calcul des probabilités de transition, qui occupe la deuxième section, nous déterminons une condition sur la tolérance pour que la chaîne de Markov soit irréductible, séparant ainsi les équilibres stationnaires du système en deux catégories : la ségrégation complète et une intégration relative. Les distributions invariantes obtenues sont enfin comparées aux simulations du chapitre précédents.

2.1 | Démarche et définitions

2.1.1 Formalisme markovien

L'ensemble Σ des configurations d'un système de N sites dont une certaine fraction est occupée par des agents séparés en deux groupes est fini et dénombrable. On peut lui adjoindre une tribu \mathcal{A} et une mesure de probabilité \mathbb{P} pour définir un espace de probabilité.

Une suite de variables aléatoires $(X_t)_{t \geq 0}$, définies sur Σ à valeurs dans un ensemble dénombrable E , est une chaîne de Markov de distribution initiale λ_0 si et seulement si, pour tout $t \geq 0$, $n_0, \dots, n_{t-1}, n, n' \in E$

$$\mathbb{P}(X_{t+1} = n' \mid X_0 = n_0, \dots, X_{t-1} = n_{t-1}, X_t = n) = \mathbb{P}(X_{t+1} = n' \mid X_t = n) \quad (2.1)$$

avec la condition initiale $\mathbb{P}(X_0 = n_0) = \lambda_0(n_0)$. Elle est dite homogène si la probabilité de transition

$$\pi_{nn'} = \mathbb{P}(X_{t+1} = n' \mid X_t = n) \quad (2.2)$$

est indépendante de t . $\boldsymbol{\pi} = (\pi_{nn'})_{n, n' \in E}$ est la matrice de transition de cette chaîne de Markov. C'est à partir de ses puissances $\boldsymbol{\pi}^t = (\pi_{nn'}^{(t)})_{n, n' \in E}$ que l'on peut décrire la structure de E . Voici donc quelques définitions permettant de construire deux théorèmes qui seront utiles dans la suite.

On dit que les états n et n' communiquent s'il existe $t, t' \geq 0$ tels que

$$\pi_{nn'}^{(t)} \text{ et } \pi_{n'n}^{(t')} > 0 \quad (2.3)$$

autrement dit, s'il existe un chemin de l'état n vers l'état n' et inversement.

On dit qu'un état n est apériodique s'il existe $T \geq 0$ tel que pour $t \geq T$,

$$\pi_{nn}^{(t)} > 0 \quad (2.4)$$

Un état n est enfin soit récurrent soit transitoire, selon que la série $\sum_{t \geq 0} \pi_{nn}^{(t)}$ diverge ou non. La probabilité que cet état sera visité une infinité de fois au cours de l'évolution du système est alors 1 ou 0, respectivement.

Si tous les états de E communiquent, on dit que la chaîne est irréductible. Sinon, on peut partitionner E en classes communicantes.

L'apériodicité, la récurrence et la transience sont des propriétés de classe, c'est-à-dire que tous les états d'une classe seront apériodiques, récurrents ou transitoires si l'un d'entre eux l'est. On dit alors généralement que la classe elle-même est apériodique, récurrente ou transitoire.

On dit également qu'une classe C est fermée si à partir d'un état $n \in C$ on ne peut accéder qu'à des états également dans C .

Théorème [118] *Toute classe fermée et finie est récurrente.*

Une distribution λ est dite invariante si elle vérifie la relation :

$$\lambda \pi = \lambda \quad (2.5)$$

On l'appelle parfois distribution stationnaire, ou distribution à l'équilibre.

Théorème [118] *Si π est la matrice de transition d'une chaîne de Markov homogène irréductible et apériodique et admet la distribution invariante λ , alors pour tout $n, n' \in E$*

$$\lim_{t \rightarrow \infty} \pi_{nn'}^{(t)} = \lambda_{n'} \quad (2.6)$$

À partir de la loi des probabilités totales, pour tout $n' \in E$,

$$\mathbb{P}(X_{t+1} = n') = \sum_{n \in E} \mathbb{P}(X_{t+1} = n' \mid X_t = n) \mathbb{P}(X_t = n) \quad (2.7)$$

on peut construire l'équation maîtresse suivante :

$$\mathbb{P}(X_{t+1} = n') - \mathbb{P}(X_t = n') = \sum_{n \in E} \left(\mathbb{P}(X_t = n) w_{nn'} - \mathbb{P}(X_t = n') w_{n'n} \right) \quad (2.8)$$

Une condition suffisante pour qu'il existe une distribution d'équilibre possible $\lambda = \left(\mathbb{P}(X_{eq} = n) \right)_{n \in E}$ est celle dite du "bilan détaillé" :

$$\forall n, n' \in E, \quad \mathbb{P}(X_{eq} = n) \pi_{nn'} = \mathbb{P}(X_{eq} = n') \pi_{n'n} \quad (2.9)$$

mais elle impose des contraintes plus fortes que celles nécessaires à la recherche d'une solution stationnaire²⁸. En particulier, si la chaîne de Markov est irréductible ($\pi_{nn'}$ ou $\pi_{n'n} = 0$ pour un ou plusieurs couples $(n, n') \in E^2$), elle devient même inutile.

Par ailleurs, lorsque les conditions du théorème précédent ne sont pas vérifiées, en particulier l'irréductibilité, d'autres formes d'équilibre existent, que l'on peut classifier à l'aide de diverses notions caractérisant leur *stabilité*.

Définition [176] *Un état n est dit stable au sens de Lyapunov si pour tout voisinage ouvert $\mathcal{V}(n)$ de n dans l'espace des états, il existe un autre voisinage ouvert $\mathcal{V}_0(n) \subset \mathcal{V}(n)$ de n tel que*

$$\exists t_0, X_{t_0} \in \mathcal{V}_0(n) \quad \Rightarrow \quad \forall t \geq t_0, X_t \in \mathcal{V}(n)$$

Définition [176] *Un état n est dit stable asymptotiquement s'il est stable au sens de Lyapunov et que*

$$\exists t_0, X_{t_0} \in \mathcal{V}_0(n) \quad \Rightarrow \quad \lim_{t \rightarrow \infty} X_t = n$$

Une autre façon de déterminer la distribution λ à l'équilibre est d'utiliser le théorème de stabilité stochastique [176]. Cette méthode est développée dans l'annexe **A**.

2.1.2 Description naïve

On considère pour l'exemple un système très petit, contenant neuf sites tous remplis à l'exception d'un. Le système est fermé, fini, la densité $d = 8/9$ et la parité $p = 1/2$ sont donc constantes. La seconde partie de ce chapitre traitera de la variante par échange d'agents mais on considère ici le cas, plus simple pour la démonstration, du déplacement d'un agent insatisfait vers l'unique site vacant. La configuration initiale **A** est représentée à gauche sur le schéma **2.1**.

Construire la matrice de transition Lorsque $T = 1/2$, la condition de satisfaction $m_i \geq (1 - 2T)\kappa_i$ devient $m_i \geq 0$: chaque agent tolère un champ local positif ou nul. Les agents insatisfaits sont représentés en noir sur la figure **2.2** (la croix représente alors la position du site vacant). Dans la configuration initiale, c'est alors les agents bleus qui sont insatisfaits et qui

²⁸On reviendra sur ces conditions et leurs conséquences dans la section **3.2**.

sont donc autorisés à se déplacer. En considérant l'invariance du système par rotations de 90° , une seule configuration **B** est accessible à partir de **A**. On a alors $\pi_{AB} = 1$, et $\pi_{An'} = 0$ pour tout $n' \neq \mathbf{B}$.

Dans la configuration **B**, on constate que six agents sont insatisfaits. Cela semble contredire les observations d'une augmentation de la satisfaction au cours du processus de Schelling, puisque le système est apparemment passé de quatre agents insatisfaits à six. Cette mesure binaire est cependant incomplète, et la figure **2.3** résout cette apparente contradiction en montrant que la densité d'interface a bien diminué au cours de cette transition **A** \rightarrow **B**.

Le calcul de π_{BA} , π_{BB} et π_{BC} consiste simplement en un dénombrement des agents dont le déplacement amènerait le système aux configurations accessibles **A**, **B** et **C** respectivement, ramené au nombre total d'agents insatisfaits (ici six). L'invariance par rotation permet ici une transition **B** \rightarrow **B** consistant à déplacer un des trois agents bleus aux coins du système : $\pi_{BB} = 1/2$. Le calcul des $\pi_{nn'}$ suit alors toujours, dans cet exemple simple, la même logique, et le processus se poursuit jusqu'à la configuration **I** qui est stationnaire, au sens où plus aucun agent n'y est insatisfait. On peut alors écrire la matrice de transition du processus entier :

$$\pi_{1/2} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/6 & 1/2 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 0 & 2/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.10)$$

En général, du fait des contraintes de déplacement des agents dans le modèle de Schelling, la matrice de transition π d'un tel système sera creuse. En particulier, l'existence d'une transition d'une configuration de départ σ à une configuration d'arrivée σ' dépend de leur proximité dans l'espace des configurations. On peut ainsi écrire [73, 76, 164]

$$\pi(\sigma, \sigma') = \sum_{i < j} \left(\prod_{k \neq i, j} \delta_{s'_k, s_k} \right) \delta_{s'_j, s_i} \delta_{s'_i, s_j} p_{ij}(\sigma) \quad (2.11)$$

où $p_{ij}(\sigma)$ est la probabilité que les sites i et j de la configuration σ échangent leurs contenus. La configuration d'arrivée, que l'on note σ_{ij} , est identique, à

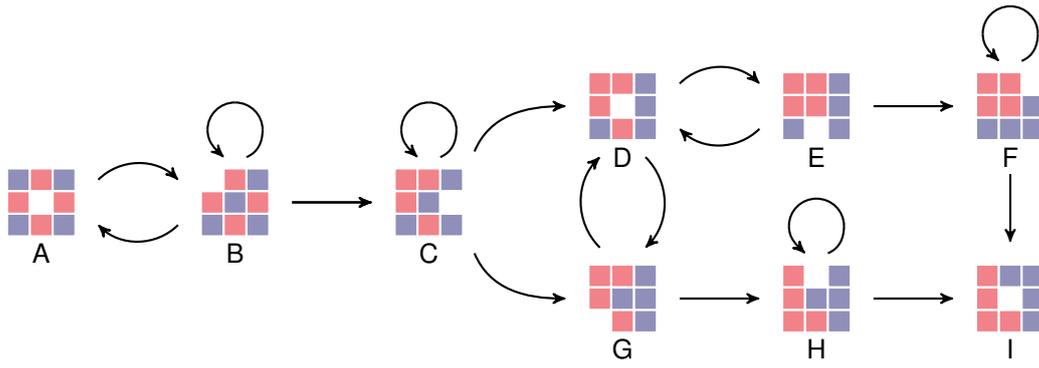


FIGURE 2.1 Représentation schématique de la chaîne de Markov associée à la dynamique de Schelling à partir de la configuration initiale **A** pour une tolérance $T = 1/2$ et un voisinage de Moore.

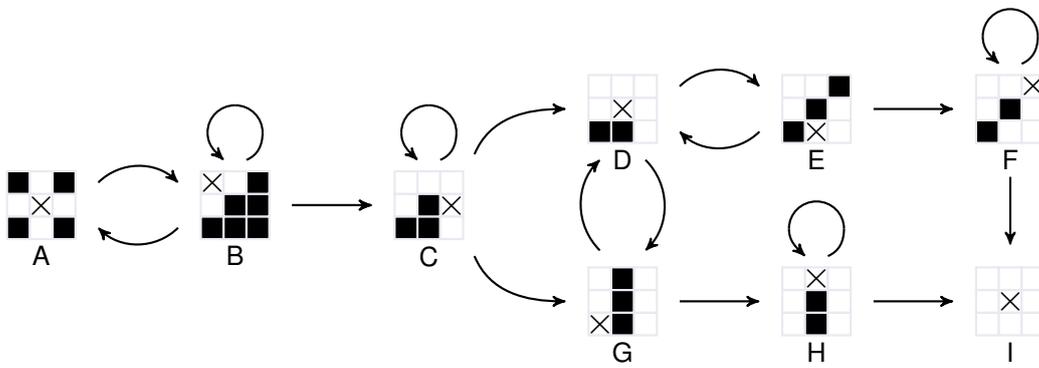
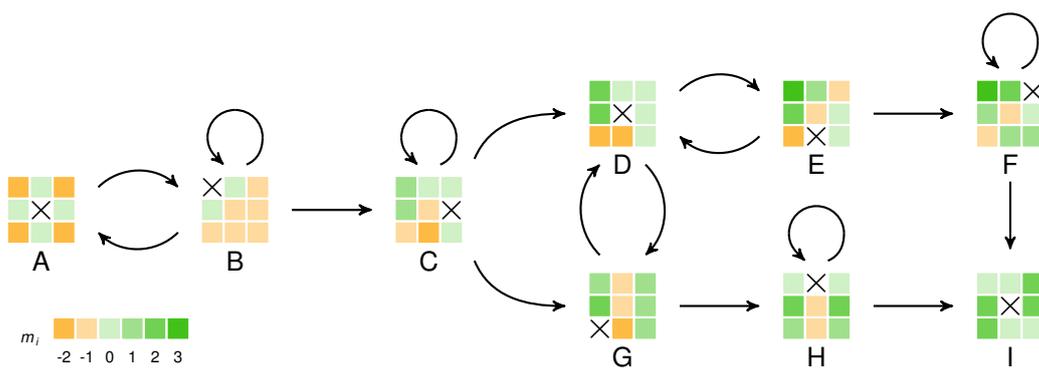


FIGURE 2.2 Recensement des agents insatisfaits (en noir).



Configuration	A	B	C	D	E	F	G	H	I
$\sum_i m_i$	-8	-6	-2	0	2	6	2	4	8
Densité d'interface x	0.6	0.575	0.525	0.5	0.475	0.425	0.475	0.45	0.4

FIGURE 2.3 Représentation des aimantations individuelles.

l'échange près, à la configuration de départ : si $\sigma = \{s_1, \dots, s_i, \dots, s_j, \dots, s_N\}$ alors $\sigma_{ij} = \{s_1, \dots, s_j, \dots, s_i, \dots, s_N\}$. L'espace d'états de cette chaîne de Markov est alors simplement l'ensemble des configurations Σ .

En physique statistique, cette description est celle d'un processus de Kawasaki [96, 97], à ceci près que la somme dans l'équation précédente devrait porter sur des paires de voisins uniquement, afin de modéliser des processus diffusifs, de proche en proche. Ce n'est pas le cas du modèle de Schelling qui permet au contraire la mobilité des agents sans contrainte de distance.

Un autre processus qui lui est souvent associé, ou opposé, dans la littérature des systèmes de spins est le processus de Glauber [72]. Au lieu d'échanger le contenu de deux sites, un unique site occupé est choisi aléatoirement et l'agent qu'il contient change de groupe, ce qui peut modéliser l'influence du voisinage dans un système magnétique, ou bien un agent quittant la ville et étant immédiatement remplacé par un agent entrant de l'autre groupe.

On a alors

$$\pi(\sigma, \sigma') = \sum_i \left(\prod_{j \neq i} \delta_{s'_j, s_j} \right) \delta_{s'_i, -s_i} p_i(\sigma) \quad (2.12)$$

où $p_i(\sigma)$ est la probabilité que l'agent se situant sur le site i change de groupe. La configuration d'arrivée est alors $\sigma_{-i} = \{s_1, \dots, -s_i, \dots, s_N\}$. Choisir l'un ou l'autre de ces processus a des implications thermodynamiques qui seront développées dans le chapitre **3**.

Il est également à noter que ces modèles dynamiques sont construits afin qu'une solution stationnaire soit celle correspondant à l'équilibre thermique, les expressions de $p_{ij}(\sigma)$ ou $p_i(\sigma)$ sont donc conditionnées par cet équilibre. Au contraire, dans le cadre du modèle de Schelling, c'est le choix de $p_{ij}(\sigma)$ ou $p_i(\sigma)$ qui va déterminer l'équilibre, s'il existe, du système. Ces quantités sont simplement liées à la satisfaction des agents impliqués. Dans le formalisme de Kawasaki :

$$p_{ij}(\sigma) = K_\sigma^{-1} (1 - u_i(\sigma)) (1 - u_j(\sigma)) \quad (2.13)$$

où K_σ est une constante de normalisation.

La transition est possible à condition que les deux agents soient insatisfaits dans la configuration de départ.

Dans l'exemple présenté plus haut, aucune contrainte n'était imposée au site vacant : $u_\square(\sigma) = 0$, où \square désigne l'indice du site en question. On

obtient alors

$$p_{i\Box}(\boldsymbol{\sigma}) = \alpha_{\boldsymbol{\sigma}}^{-1}(1 - u_i(\boldsymbol{\sigma})) = K_{\boldsymbol{\sigma}}^{-1} \times \begin{cases} 1 & \text{si } m_i < (1 - 2T)\kappa_i \\ 0 & \text{sinon} \end{cases} \quad (2.14)$$

$$= K_{\boldsymbol{\sigma}}^{-1} \mathbb{1}_{[-\kappa_i, (1-2T)\kappa_i]}(m_i) \quad (2.15)$$

Il est possible de restreindre le déplacement d'un agent au site inoccupé si ce dernier est "satisfaisant" uniquement [70], dans ce cas

$$p_{i\Box}(\boldsymbol{\sigma}) = K_{\boldsymbol{\sigma}}^{-1} \mathbb{1}_{[-\kappa_i, (1-2T)\kappa_i]}(m_i) \mathbb{1}_{[(1-2T)\kappa_{\Box}, \kappa_{\Box}]}(h_{\Box}) \quad (2.16)$$

Dans le cas d'échanges entre agents, qui est celui qui sera développé dans la suite de ce chapitre,

$$p_{ij}(\boldsymbol{\sigma}) = K_{\boldsymbol{\sigma}}^{-1} \mathbb{1}_{[-\kappa_i, (1-2T)\kappa_i]}(m_i) \mathbb{1}_{[-\kappa_j, (1-2T)\kappa_j]}(m_j) \quad (2.17)$$

Enfin, dans une dynamique de type Glauber,

$$p_i(\boldsymbol{\sigma}) = \Gamma_{\boldsymbol{\sigma}}^{-1} \mathbb{1}_{[-\kappa_i, (1-2T)\kappa_i]}(m_i) \quad (2.18)$$

où $\Gamma_{\boldsymbol{\sigma}}$ est une constante de normalisation. Il est ici aussi possible de restreindre cet échange aux cas où l'agent devient satisfait *après* le changement, cette probabilité de transition devient alors

$$\begin{aligned} p_i(\boldsymbol{\sigma}) &= \Gamma_{\boldsymbol{\sigma}}^{-1} \mathbb{1}_{[-\kappa_i, (1-2T)\kappa_i]}(m_i) \times \mathbb{1}_{[(1-2T)\kappa_i, \kappa_i]}(-m_i) \\ &= \Gamma_{\boldsymbol{\sigma}}^{-1} \times \begin{cases} \mathbb{1}_{[-\kappa_i, (1-2T)\kappa_i]}(m_i) & \text{si } T < 1/2 \\ \mathbb{1}_{[-\kappa_i, (2T-1)\kappa_i]}(m_i) & \text{si } T \geq 1/2 \end{cases} \end{aligned} \quad (2.19)$$

On démontre ici, en formalisant l'expression des probabilités de transition, la quasi-symétrie²⁹ autour de $T = 1/2$ constatée empiriquement dans les travaux de G. Barmapalias et al. [15] et de H. Omidvar et M. Franceschetti [124, 125, 127].

Revenons à la matrice $\boldsymbol{\pi}$ définie par (2.10). Sa distribution invariante $\boldsymbol{\lambda}$ est le vecteur unitaire associé à la configuration $\mathbf{l} : \mathbf{e}_1 = (0, 0, 0, 0, 0, 0, 0, 1)$: \mathbf{l} est bien une configuration récurrente de la chaîne de Markov. Cette configuration est ségréguée, visiblement. Sa densité d'interface est de 0.4, ce qui reste proche de la valeur de départ, mais c'est une conséquence de la taille finie et du remplissage du système. Cette observable voit tout de même sa

²⁹La symétrie se rompt exactement en $T = 1/2$, où $m_i = 0$ est une valeur permettant l'échange, mais pas $m_i \rightarrow 0^+$.

valeur diminuer au cours de l'évolution, conformément aux observations de Schelling.

Il est à noter que certaines transitions ($\mathbf{B} \rightarrow \mathbf{A}$, $\mathbf{E} \rightarrow \mathbf{D}$ et $\mathbf{G} \rightarrow \mathbf{D}$) peuvent augmenter la valeur de la densité d'interface. En effet, cette observable n'est pas strictement une fonction de Lyapunov : il a ainsi été démontré [74] dans le cas d'échange entre agents que de tels déplacements entre sites adjacents peuvent occasionnellement augmenter l'insatisfaction générale, il en va de même pour le déplacement vers un site vacant. De fait, il s'agit d'un artefact de la taille finie du système : la probabilité que deux sites adjacents soient choisis pour échanger leur contenu est d'ordre κ/N lorsque ce tirage est uniforme.

Pour une valeur de tolérance de $T = 2/3$, avec les mêmes règles de déplacement et de construction des π_{nm} , la matrice de transition devient simplement (sa chaîne est représentée en figure 2.4) :

$$\boldsymbol{\pi}_{2/3} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \quad (2.20)$$

Sa distribution invariante est le vecteur $\mathbf{e}_{\mathbf{B}} = (0, 1)$, qui correspond à la configuration \mathbf{B} qui est donc récurrente pour cette valeur $T = 2/3$. Plus aucun agent n'étant insatisfait, le processus prend fin, et la densité d'interface ne descend pas sous $1/2$, le système est bien intégré, conformément aux observations de Schelling.

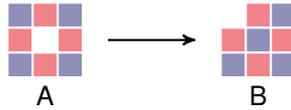


FIGURE 2.4 Représentation schématique de la chaîne de Markov associée à la dynamique de Schelling à partir de la configuration initiale \mathbf{A} pour une tolérance $T = 2/3$ et un voisinage de Moore.

Contraintes géométriques On peut utiliser cet exemple pour interroger plusieurs choix qui ont été faits, en particulier celui du voisinage utilisé et des conditions aux limites.

Ainsi, pour un voisinage de von Neumann, en revenant à une tolérance $T = 1/2$, le processus est beaucoup plus long et donne accès à cinq configurations récurrentes supplémentaires, dénommées \mathbf{F} (qui devient stable lorsque le voisinage est restreint), \mathbf{F}^* (qui est le “négatif” de \mathbf{F} , où

les groupes d'agents ont été inversés), **J**, **K** et **K*** (idem). La distribution invariante est alors une superposition de ces six configurations. La chaîne est représentée de façon simplifiée en figure 2.5 ; le nombre de configurations éclipsées n'est qu'indicatif, il peut s'y trouver quelques redondances, ce qui compte pour la suite est son ordre de grandeur : environ 90 états, soit une matrice de transition de taille 90×90 , pour un système ridiculement petit.

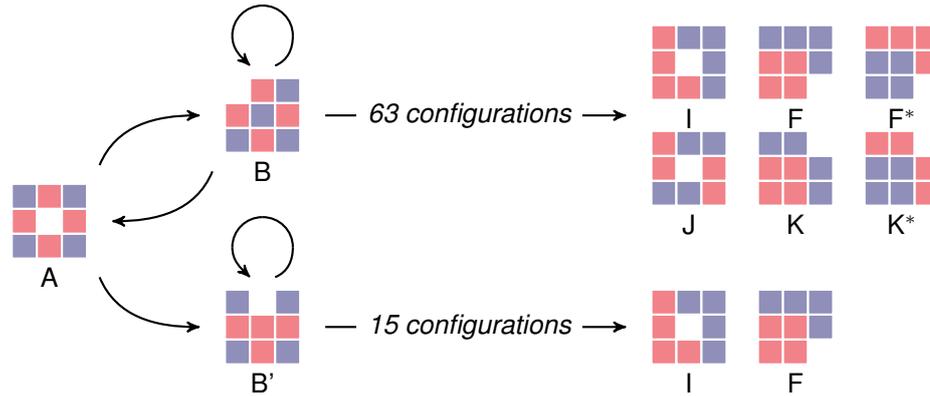


FIGURE 2.5 Représentation schématique de la chaîne de Markov associée à la dynamique de Schelling à partir de la configuration initiale **A** pour une tolérance $T = 1/2$ et un voisinage de von Neumann.

Un autre aspect géométrique particulièrement important dans ce système est l'effet des conditions aux limites. Il est ainsi courant de raisonner sur un tore plan (parfois dénommé “système avec conditions aux limites périodiques”), ce qui a l'avantage de fixer la taille du voisinage κ pour tous les agents (alors qu'ici, lorsque $T \neq 1/2$, la satisfaction des agents doit être déterminée au cas par cas pour les agents au bord du système). Ces inconvénients disparaissent relativement rapidement lorsque le système grandit. Cette dernière variante n'est pas illustrée ici, mais on peut démontrer dans le cas d'un voisinage de Moore, qu'un agent ne pourra jamais changer sa satisfaction : chacun d'entre eux étant virtuellement voisin de tout le système, cette somme reste constante au cours de l'évolution.

Agrégation de la chaîne de Markov Enfin, le dernier point que cet exemple simple va permettre de discuter avant de développer la méthode générale de détermination de $\pi_{nn'}$ et aboutir à une détermination analytique de la matrice de transition et de sa distribution invariante, concerne l'agrégabilité de la chaîne de Markov.

Cette notion, appelée *lumpability* en anglais, ne sera pas développée dans

cette thèse mais son traitement dans les modèles multi-agents a fait l'objet de la thèse de S. Banisch [11] avec application à un modèle plus simple de transmission d'opinion (*voter model*). Elle mérite tout de même d'être abordée au moins succinctement ici, d'autant qu'on l'a déjà utilisée plus haut sans le faire remarquer.

Théorème [11] *Une chaîne de Markov est dite **agrégable** selon une partition $\mathcal{E} = \{\mathcal{E}_1, \dots, \mathcal{E}_r\}$ de E si et seulement si, pour toute paire $\mathcal{E}_i, \mathcal{E}_j$ de sous-ensembles de \mathcal{E} , pour tout $n, n' \in \mathcal{E}_i$*

$$\sum_{m \in \mathcal{E}_j} \pi_{nm} = \sum_{m \in \mathcal{E}_j} \pi_{n'm}$$

Dans la figure **2.4**, un argument d'invariance par rotation est utilisé pour représenter les quatre configurations qui constituent l'état **B** par l'une d'entre elle. Vérifier l'agrégabilité à cette étape consiste à ne plus considérer **A** et **B** comme des configurations mais comme des états regroupant plusieurs configurations : $\mathcal{A} = \{\mathbf{A}\}$ et $\mathcal{B} = \{\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3, \mathbf{B}_4\}$. On calcule alors, pour tout $\mathbf{A} \in \mathcal{A}$,

$$\sum_{\sigma' \in \mathcal{B}} \pi_{\mathbf{A}\sigma'} = \pi_{\mathbf{A}\mathbf{B}_1} + \pi_{\mathbf{A}\mathbf{B}_2} + \pi_{\mathbf{A}\mathbf{B}_3} + \pi_{\mathbf{A}\mathbf{B}_4} = 4 \times 1/4 = 1$$

et pour tout $\mathbf{B}_i \in \mathcal{B}$,

$$\sum_{\sigma \in \mathcal{A}} \pi_{\mathbf{B}_i\sigma} = \pi_{\mathbf{B}_i\mathbf{A}} = 1/6, \quad \forall i$$

Cette chaîne est donc bien agrégable : l'argument d'invariance par rotation permet de regrouper des configurations similaires au sein d'états d'une chaîne de Markov sans avoir à préciser toutes ces configurations. L'agrégation par similitude qui semblait intuitive précédemment est donc bien justifiée.

Cette démonstration s'applique à la chaîne **2.1** en entier : il s'agit d'une chaîne extrêmement simplifiée. Au moment de la brisure de symétrie $\mathcal{B} \rightarrow \mathcal{C}$, il existe en fait plusieurs sous-chaînes parallèles et contenant des configurations semblables à une symétrie ou une rotation près à celles représentées. Il est alors plus juste de considérer que la matrice $\boldsymbol{\pi}$ contient les probabilités de transitions entre les états \mathcal{C} à \mathcal{I} , et non entre les configurations **C** à **I**.

Il peut être plus difficile, voire impossible, de justifier une autre forme d'agrégation dans le modèle de Schelling que l'on va pourtant mettre en œuvre dans la seconde partie de ce chapitre, considérant donc une variante du modèle présenté. Agréger par densité d'interface (ou autre observable pertinente) s'avère en effet discutable. C'est même faux dans l'exemple **2.1**.

Regrouper les configurations par densité d'interface revient à partitionner l'espace d'états selon $\{\{\mathbf{A}\}, \{\mathbf{B}\}, \{\mathbf{C}\}, \{\mathbf{D}\}, \{\mathbf{E}, \mathbf{G}\}, \{\mathbf{F}\}, \{\mathbf{H}\}, \{\mathbf{I}\}\}$. On note alors immédiatement que $\pi_{\mathbf{EF}} \neq \pi_{\mathbf{GF}}$ ou encore que $\pi_{\mathbf{EH}} \neq \pi_{\mathbf{GH}}$, deux égalités qui seraient nécessaires pour établir l'agrégabilité de la chaîne.

Cependant, comme l'ont démontré T. Rogers et A. McKane [146], le traitement analytique d'un modèle proche de celui de Schelling, bien que ne pouvant en être directement tiré sans introduire cette variante, permet d'aboutir à une résolution pertinente à l'explication de l'émergence de la ségrégation dans un tel système. Il n'est donc pas question de balayer la problématique de la potentielle non-agrégabilité de la chaîne de Markov selon une partition définie par la densité d'interface, mais au contraire de poursuivre ce travail en ayant conscience des limites de l'approche employée et de la force de cette hypothèse au moment d'en présenter les résultats : la chaîne de Markov étudiée dans la section suivante n'est pas une chaîne *extraite* du système original. C'est un autre processus discret, dont l'espace d'états est une partition de l'espace des configurations, et que l'on suppose markovien en première approximation.

2.2 | Évolution de la densité d'interface

2.2.1 Espace d'états

La densité d'interface x , ou densité de contact, est une observable du système. Elle compare le nombre $n_{\bullet\circ}$ de paires dirigées de sites voisins hétérogènes, au sens où l'un des sites contient un agent d'un groupe et le second un agent de l'autre groupe, et le nombre total n_κ de paires dirigées de sites voisins du système. Spécifiquement, elle s'écrit

$$x(\sigma) = \frac{n_{\bullet\circ}}{n_\kappa} = \frac{\frac{1}{2} \sum_{\langle i,j \rangle} (s_i^2 s_j^2 - s_i s_j)}{\sum_{\langle i,j \rangle} s_i^2 s_j^2} = \frac{1}{2} \left(1 - \frac{\sum_{i=1}^N m_i}{\sum_{i=1}^N \kappa_i} \right) \quad (2.21)$$

Les quantités κ_i et m_i présentes dans la dernière expression ont déjà été définies, il s'agit respectivement du nombre de sites occupés dans le voisinage d'un agent se trouvant sur le site i (si aucun agent ne s'y trouve alors $s_i = 0$ donc $\kappa_i = 0$, les sites vacants n'interviennent pas dans les sommes), et du champ ou de l'aimantation locale ressentie par ce même agent, qui prend des valeurs entre $-\kappa_i$ et κ_i . Par la suite, on utilisera brièvement les champs locaux effectifs h_i définies par $m_i = s_i h_i$, qui simplifieront quelques expressions bien que n'ayant pas d'analogie physique dans le contexte des verres de spins, contrairement aux deux autres quantités locales.

Par construction, la densité de contact x prend un nombre fini \mathcal{N} de valeurs $X_n \in [0, 1]$. La valeur 0 est associée à l'égalité

$$\forall i, \quad m_i = \kappa_i \quad (2.22)$$

qui signifie que toutes les paires d'agents sont homogènes : le système est parfaitement ségrégé et les noyaux homogènes s'y trouvant doivent être séparés, une frontière entre deux noyaux faisant diminuer la valeur de la première somme. Au contraire, la valeur 1 correspond à des configurations vérifiant

$$\forall i, \quad m_i = -\kappa_i \quad (2.23)$$

qui n'est pas possible dans toutes les géométries, mais qu'on peut trouver dans une grille carrée avec voisinage de von Neumann par exemple, sous la forme de configurations en échiquier : chaque agent est entouré de quatre

agents de l'autre groupe qui sont ses seuls voisins. Cet exemple permet de justifier le facteur $1/2$ apparaissant dans l'expression de la densité d'interface. Dans une géométrie de Moore, il est vrai qu'elle ne dépassera que rarement $1/2$ (et lorsque c'est le cas, il s'agit généralement d'effets de bords), ce qui peut sembler non-optimal.

Cette discrétisation de la densité d'interface entraîne l'existence d'une partition $\{\mathcal{C}_n\}_{n \in [0, \mathcal{N}]}$ de l'ensemble Σ des configurations, telle que

$$\mathcal{C}_n = \{\sigma \in \Sigma \mid x(\sigma) = X_n\} \quad (2.24)$$

À des fins de désambiguïsation, un ensemble de configurations de densité d'interface x sera simplement appelée \mathcal{C}_x , la nature discrète de x devenant implicite. La partition $\{\mathcal{C}_x\}$ définit alors l'espace d'états d'une chaîne de Markov $(X_t)_{t \geq 0}$ s'appuyant sur le processus de Schelling sous-jacent.

L'objectif de cette section est de déterminer les probabilités de transition d'un état \mathcal{C}_x à l'autre à partir des propriétés des transitions élémentaires entre configurations σ . Plus spécifiquement, on s'intéressera à la probabilité qu'une telle transition résulte en un incrément Δx de la densité d'interface.

Le nombre \mathcal{N} d'états de cet espace, ou d'éléments dans la partition, n'est pas trivial pour un couple (d, p) donné. Cependant, il sera possible d'en déterminer une borne supérieure dans la suite.

L'espace d'états étant défini, il faut à présent calculer sa matrice de transition π . Pour cela, la méthode détaillée dans la section précédente peut être appliquée.

2.2.2 Incréments de la densité d'interface

Par définition, la densité d'interface $X_t = x$ du système à l'instant t dans une configuration σ appartenant à la classe \mathcal{C}_x vérifie l'équation

$$\sum_{i=1}^N (m_i + (2x - 1)\kappa_i) = 0 \quad (2.25)$$

Après l'échange des contenus des sites a et b , le système se trouve dans la configuration appelée σ' , et sa densité d'interface $X_{t+1} = x'$ vérifie une équation similaire

$$\sum_{i=1}^N (\tilde{m}_i + (2x' - 1)\tilde{\kappa}_i) = 0 \quad (2.26)$$

où \tilde{m}_i et $\tilde{\kappa}_i$ sont respectivement définis comme l'aimantation locale et le degré effectif de chaque agent dans la configuration σ' . Les deux configura-

tions ne diffèrent que du contenu de deux sites, elles sont donc très similaires, c'est-à-dire proches dans l'espace Σ . Ceci implique que les classes auxquelles elles appartiennent sont soit identiques ($\mathcal{C}_{x'} = \mathcal{C}_x$), soit proches dans la partition. Les quantités m_i , \tilde{m}_i et κ_i , $\tilde{\kappa}_i$ partagent donc des valeurs communes pour la majorité des sites. En particulier,

$$\tilde{m}_i = \begin{cases} s_b h_a & i = a \\ s_a h_b & i = b \\ m_i + s_i(s_b - s_a) & i \in \partial a \\ m_i + s_i(s_a - s_b) & i \in \partial b \\ m_i & \text{sinon} \end{cases} \quad \text{et} \quad \tilde{\kappa}_i = \begin{cases} s_b^2 d_a & i = a \\ s_a^2 d_b & i = b \\ \kappa_i + s_i^2(s_b^2 - s_a^2) & i \in \partial a \\ \kappa_i + s_i^2(s_a^2 - s_b^2) & i \in \partial b \\ \kappa_i & \text{sinon} \end{cases} \quad (2.27)$$

avec $m_i = s_i h_i$ et $\kappa_i = s_i^2 d_i$, où d_i est la densité locale $d_i = \sum_{j \in \partial i} s_j^2$. On peut alors réécrire l'équation (2.26) de la manière suivante³⁰ :

$$\sum_{i=1}^N m_i - 2(m_a + m_b) + 2(s_a h_b + s_b h_a) + (2x' - 1) \left(\sum_{i=1}^N \kappa_i - 2(\kappa_a + \kappa_b) + 2(s_a^2 d_b - s_b^2 d_a) \right) = 0 \quad (2.28)$$

et traiter des cas.

Trivialement, si les deux sites tirés sont vacants, $s_a = s_b$, le système reste inchangé.

Si les agents se déplacent dans le système, au sens où seul un site occupé peut transférer son contenu vers un site vacant, alors $s_a = 0$ ou $s_b = s = \pm 1$ (ou inversement, sans perte de généralité). On obtient, à l'aide de (2.25)

$$x' = \frac{x \sum_i \kappa_i - s(h_a - h_b) + s^2(d_a - d_b)}{\sum_i \kappa_i + 2s^2(d_a - d_b)} \quad (2.29)$$

Si deux agents échangent leurs positions, alors $s_a = \pm 1$ et $s_b = \pm 1$. Si $s_a = s_b$, de nouveau, l'échange laisse le système inchangé. En revanche dans le cas où $s_a = -s_b$, on obtient

$$x' = x + \frac{2(m_a + m_b)}{\sum_i \kappa_i} \quad (2.30)$$

Dans la limite $d = 1$, ce dernier cas est le seul pouvant se produire. On peut alors montrer que la somme des deux aimantations est paire à l'aide de la définition (1.5) :

$$m = n_s - n_d = (\kappa - n_d) - n_d = \kappa - 2n_d \quad (2.31)$$

³⁰Se référer à l'annexe **B**.

d'où

$$m_a + m_b = 2(\kappa - n_{d,a} + n_{d,b}) = 2k, \quad \text{avec } k \in \llbracket -\kappa, \kappa \rrbracket \quad (2.32)$$

Cela permet de réduire, lorsque $d = 1$ pour un graphe sous-jacent κ -régulier, les valeurs prises par l'incrément $\Delta x = x' - x$ à l'ensemble

$$\Delta x = x' - x \in \left\{ \frac{4k}{\kappa N} \mid k \in \llbracket -\kappa, \kappa \rrbracket \right\} \quad (2.33)$$

La partition de l'ensemble des configurations peut alors s'écrire, dans ce cas particulier $d = 1$,

$$\{\mathcal{C}_0, \mathcal{C}_{4/\kappa N}, \mathcal{C}_{8/\kappa N}, \dots, \mathcal{C}_1\} \quad (2.34)$$

Elle contient \mathcal{N} classes avec

$$\mathcal{N} = \frac{\kappa N}{4} + 1 \quad (2.35)$$

Dans le cas général, il est plus difficile de déterminer cette quantité, car l'évolution de l'incrément est gouvernée par l'équation (2.29). On peut toutefois en déterminer une borne supérieure en dénombrant les ratios $n_{\bullet\circ}/n_\kappa$ de paires hétérogènes accessibles.

Dans une configuration où tous les sites vacants sont dispersés (*scattered*), le nombre $n_{\text{sc}} = \min(n_\kappa)$ de paires dirigées d'agents est le nombre de paires dirigées du graphe desquelles on retranche les liens rompus par la présence des sites vacants :

$$n_{\text{sc}} = \kappa N - 2\kappa N(1 - d) = \kappa N(2d - 1) \quad (2.36)$$

On peut noter que ces configurations "dispersées" ne sont accessibles que pour $d > 1/2$, faute de quoi cette borne inférieure s'annule. D'autre part, par définition,

$$n_\kappa = \sum_{\langle i,j \rangle} s_i^2 s_j^2 \leq \kappa \sum_{i=1}^N s_i^2 = \kappa N d \quad (2.37)$$

À partir de l'encadrement $n_{\text{sc}} \leq n_\kappa \leq \kappa N d$, il reste à déterminer les valeurs que peut prendre le nombre de paires dirigées hétérogènes $n_{\bullet\circ}$. Par construction, il s'agit d'un multiple de quatre (le nombre de paires de voisins impactées par un échange est pair, et on double cette quantité pour obtenir les paires dirigées) et peut prendre, au plus, lorsque $p = 1/2$, des valeurs dans $\llbracket 0, n_\kappa \rrbracket$. Lorsque $p = 1/2$ et $d = 1$, on peut donc retrouver, puisque $n_\kappa = \kappa N$, que

$$\mathcal{N} = \frac{\kappa N}{4} + 1 \quad (2.38)$$

Dans le cas général, une borne supérieure de \mathcal{N} peut être déterminée³¹ par

$$\mathcal{N} \leq (3d - 1)(1 - d) \frac{(\kappa N)^2}{8} + \frac{7 - 5d}{8} \kappa N + 1 \quad (2.39)$$

Cette inégalité est relativement imprécise à cause des possibles redondances dans les ratios calculés, il est sûrement possible de l'affiner en utilisant l'indicatrice d'Euler par exemple, mais ce n'est pas le propos ici. On peut remarquer que lorsque $d = 1$, on retrouve $\mathcal{N} = \kappa N/4 + 1$

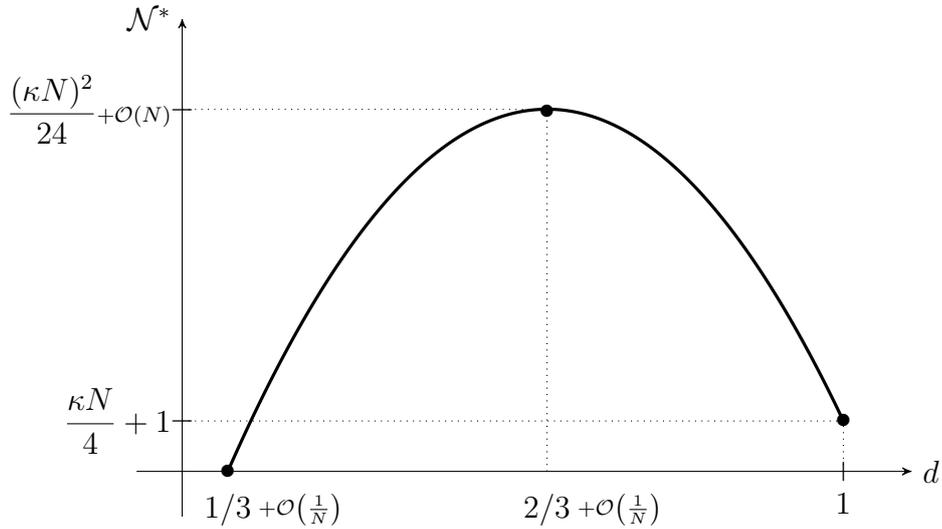


FIGURE 2.6 Un majorant $\mathcal{N}^*(d)$ de la taille de l'espace des états de la chaîne de Markov sous-jacente au processus de Schelling

Utiliser cette chaîne de Markov sous-jacente a donc permis de réduire la taille des états de la chaîne de Markov étudiée de $|\Sigma|$, d'ordre $4^N/\sqrt{N}$, à un nombre d'ordre au plus N^2 , et en particulier lorsque $d = 1$, d'ordre N . Rappelons ici que N est la taille du système, qui modélise une ville. On constate bien là l'intérêt de ne considérer que le processus extrait.

Parmi les \mathcal{N} classes de l'espace d'états de la chaîne de Markov, seule une poignée, caractérisée par l'équation (2.33), est accessible à partir d'une configuration donnée appartenant à \mathcal{C}_x . On s'intéresse à présent à la distribution de probabilité associée à ses classes d'arrivées $\mathcal{C}_{x+\Delta x}$.

³¹Se référer à l'annexe **C**.

2.2.3 Distribution des aimantations

On cherche à présent à déterminer la distribution de la somme $m_a + m_b$ (2.32). Dans un premier temps, on peut établir que, lorsque $s_i = +1$,

$$\mathcal{P}(m_i = m) = \frac{1}{Z} \sum_{n=\max(0,-m)}^{\frac{\kappa-m}{2}} \binom{(1-d)N}{\kappa-m-2n} \binom{dpN}{m+n} \binom{d(1-p)N}{n} \quad (2.40)$$

et lorsque $s_i = -1$

$$\mathcal{P}(m_i = m) = \frac{1}{Z'} \sum_{n=\max(0,-m)}^{\frac{\kappa-m}{2}} \binom{(1-d)N}{\kappa-m-2n} \binom{dpN}{n} \binom{d(1-p)N}{m+n} \quad (2.41)$$

où Z et Z' sont des constantes de normalisation, égales à la somme de ces expressions sur l'intervalle $m \in \llbracket -\kappa, \kappa \rrbracket$.

En toute rigueur, les bornes des sommes (2.40) et (2.41) doivent être corrigées lorsque d , p et N prennent des valeurs extrêmes : on suppose ici que le système que l'on cherche à modéliser n'appartient pas à ces cas de figure, il doit être suffisamment grand ($N \gg \kappa$) et majoritairement occupé par les deux populations ($d \gg \kappa/N$ et $\kappa/N \ll p \ll 1 - \kappa/N$). De plus, on peut remarquer à l'aide de (2.31) que la borne supérieure des deux sommes, bien qu'apparemment fractionnaire, correspond à l'entier n_d .

Dans le cas $d = 1$, les seuls termes non-nuls sont justement ceux d'indice $n = n_d$. Par simplicité d'écriture, on peut alors poser le changement de variable $m = \kappa - 2n$, qui permet de déterminer l'expression

$$\mathcal{P}(m_i = \kappa - 2n) = \frac{p}{Z} \binom{pN}{\kappa-n} \binom{(1-p)N}{n} + \frac{1-p}{Z'} \binom{pN}{n} \binom{(1-p)N}{\kappa-n} \quad (2.42)$$

pour laquelle on peut expliciter les constantes de normalisation à l'aide de l'identité de Vandermonde :

$$Z = Z' = \sum_{n=0}^{\kappa} \binom{pN}{\kappa-n} \binom{(1-p)N}{n} = \binom{N}{\kappa} \quad (2.43)$$

On obtient alors une distribution correspondant à la somme pondérée de deux distributions hypergéométriques symétriques l'une par rapport à l'autre, qui devient elle-même hypergéométrique dans le cas $p = 1/2$. Enfin, la limite thermodynamique $N \gg \kappa$ explicitée plus haut permet d'utiliser la convergence de cette loi vers la loi binomiale. On trouve donc :

$$\mathcal{P}(m_i = \kappa - 2n) = \frac{\binom{N/2}{n} \binom{N/2}{\kappa-n}}{\binom{N}{\kappa}} \underset{N \gg \kappa}{\sim} \frac{1}{2^\kappa} \binom{\kappa}{n} = \frac{1}{2^\kappa} \binom{\kappa}{\frac{\kappa-m}{2}} \quad (2.44)$$

Ces résultats simples étaient attendus : au vu des diverses approximations effectuées, obtenir une valeur donnée $m_i = m$ consiste simplement en un double tirage d'agents de chaque groupe pour peupler le voisinage de taille κ , on retrouve bien une loi hypergéométrique de paramètres $(\kappa, p = 1/2)$.

La convergence vers une distribution binomiale s'explique également par dénombrement : au lieu d'effectuer le tirage des deux groupes d'agents, leur disponibilité devenant illimitée, il suffit pour trouver $m_i = m$ de tirer le bon nombre d'agents du groupe que l'on recherche parmi les κ nécessaires pour peupler le voisinage. On utilisera enfin les propriétés de symétrie du coefficient binomial pour écrire

$$\mathcal{P}(m_i = m) \underset{N \gg \kappa}{\sim} \frac{1}{2^\kappa} \binom{\kappa}{\frac{\kappa+m}{2}} \quad (2.45)$$

Cette dernière approximation permet de conclure quant à la probabilité d'obtenir un incrément donné lors de l'échange du contenu de deux sites : en considérant les variables m_a et m_b indépendantes³² et identiquement distribuées, la loi suivie par la somme de variables de lois binomiales de paramètres r_a et r_b est simplement une loi binomiale de paramètre $r_a + r_b$. On peut donc à présent écrire

$$\mathcal{P}(m_a + m_b = 2k) \underset{N \gg \kappa}{\sim} \frac{1}{4^\kappa} \binom{2\kappa}{\kappa + k} \quad (2.46)$$

et de manière générale

$$\mathcal{P}\left(\sum_{i=1}^{\nu} m_i = y\right) \underset{N \gg \kappa}{\sim} \frac{1}{2^{\nu\kappa}} \binom{\nu\kappa}{\frac{\nu\kappa+y}{2}} \quad (2.47)$$

On utilisera par la suite la notation

$$\mathcal{P}_{(\nu)}(y) = \mathcal{P}\left(\sum_{i=1}^{\nu} m_i = y\right) \quad \text{d'où} \quad \mathcal{P}(m_a + m_b = 2k) = \mathcal{P}_{(2)}(2k) \quad (2.48)$$

L'approximation de $\mathcal{P}(m_a + m_b = 2k)$ semble raisonnable pour les valeurs de κ habituellement choisies dans le cadre de l'étude du modèle de Schelling, on peut en effet montrer lorsque $\kappa = 4$ et 8 que le produit des fonctions génératrices des probabilités des variables indépendantes m_a et m_b vérifie

$$G_{m_a+m_b}(t) = G_{m_a}(t)G_{m_b}(t) = \left(\frac{t+1}{2}\right)^{2\kappa} + \mathcal{O}(1/N) \quad (2.49)$$

³²Cette hypothèse d'indépendance est démontrée formellement en section **2.2.4** et s'appuie sur la probabilité, proche de 1, de ne pas trouver les deux agents a et b dans le voisinage l'un de l'autre.

ce qui correspond, à l'ordre 0, à la fonction génératrice des probabilités d'une variable aléatoire de loi hypergéométrique de paramètre 2κ . Encore une fois, ce résultat est attendu, l'ordre 0 revenant à la fonction génératrice de probabilités d'une variable de loi binomiale.

L'approximation de $\mathcal{P}\left(\sum_{i=1}^{\nu} m_i = y\right)$ est plus grossière que la précédente. Toutefois, si l'on considère la moyenne empirique

$$\bar{m} = \frac{1}{\nu} \sum_{i=1}^{\nu} m_i = \frac{y}{\nu} \quad (2.50)$$

qui en découle, la condition imposée pour pouvoir assimiler la loi hypergéométrique à une loi binomiale, $|\bar{m}| \ll N$, devient $|y| \ll \nu N$. Si cette condition n'est pas respectée, l'indépendance des variables aléatoires $\{m_i\}_i$ doit être remise en cause : de nombreuses paires de sites voisins seront comptabilisées plusieurs fois, créant de fait des corrélations.

2.2.4 Conditionnement de la marche aléatoire

Le calcul de $\mathcal{P}(m_a + m_b = 2k)$ dans le cas où m_a et m_b sont des variables *indépendantes* donne la probabilité d'un incrément $\Delta x = 4k/\kappa N$ de la densité d'interface lors de l'échange du contenu des sites a et b aux voisinages *disjoints*.

Il faut à présent déterminer la probabilité de ce même incrément lors d'une étape quelconque de la marche aléatoire, lorsque le choix des sites est uniforme.

$$P(X_{t+1} = x + \Delta x) = \sum_{a,b} P(m_a + m_b = 2k)P(a, b) \quad (2.51)$$

où $P(a, b)$ est la probabilité de sélectionner les sites a et b parmi les N que compte le système.

Cette somme peut être séparée entre les paires $\{a, b\}$ de sites voisins et les autres, ce qui donne

$$\begin{aligned} &= \mathcal{P}_{(2)}(2k) \sum_{a \notin \partial b} P(a, b) + \sum_{a \in \partial b} P(m_a + m_b = 2k)P(a, b) \\ &\leq \mathcal{P}_{(2)}(2k) \sum_{a \notin \partial b} P(a, b) + \sum_{a \in \partial b} P(a, b) \\ &= \mathcal{P}_{(2)}(2k) \left(1 - \frac{2\kappa}{N-1}\right) + \frac{2\kappa}{N-1} \end{aligned} \quad (2.52)$$

et finalement

$$P(X_{t+1} = x + \Delta x) \underset{N \gg \kappa}{\sim} \mathcal{P}_{(2)}(2k) \underset{N \gg \kappa}{\sim} \frac{1}{4^\kappa} \binom{2\kappa}{\kappa - k} \quad (2.53)$$

Bien que pour la plupart des valeurs de la densité d'interface, la probabilité de ne pas pouvoir sélectionner des sites avec des aimantations données soit négligeable, se rapprocher d'états totalement séparés ou totalement mixtes augmente fortement la probabilité de certaines valeurs de m_i par rapport à d'autres. Intuitivement, dans un système où x peut être arbitrairement proche de 0, la probabilité de choisir un voisinage $m_i = 0$ parfaitement mélangé est nulle. Il faut donc calculer la probabilité que l'on cherche *conditionnellement* à la valeur de la densité d'interface à l'instant précédent. On recherche donc, d'après la définition (2.21),

$$P(X_{t+1} = x + \Delta x \mid X_t = x) = \mathcal{P}(m_a + m_b = 2k \mid \sum_i m_i = \kappa N(1 - 2x)) \quad (2.54)$$

On peut exprimer cette probabilité conditionnelle comme le quotient

$$P(X_{t+1} = x + \Delta x \mid X_t = x) = \frac{\mathcal{P}_{(2)}(2k) \mathcal{P}_{(N-2)}(\kappa N(1 - 2x) - 2k)}{\mathcal{P}_{(N)}(\kappa N(1 - 2x))} \quad (2.55)$$

On peut vérifier que les paramètres utilisés pour écrire les différentes probabilités de la forme $\mathcal{P}_\nu(y)$ appartiennent bien à la gamme dans laquelle les approximations effectuées plus haut sont valides :

$$\begin{aligned} \frac{|\kappa N(1 - 2x) - 2k|}{N - 2} &< \frac{\kappa N|1 - 2x|}{N - 2} + \frac{|2k|}{N - 2} \ll N \\ \frac{\kappa N|1 - 2x|}{N} &= \kappa|1 - 2x| \ll N \end{aligned} \quad (2.56)$$

Enfin,

$$P(X_{t+1} = x + \Delta x \mid X_t = x) = \frac{\binom{2\kappa}{\kappa + k} \binom{\kappa N - 2\kappa}{\kappa N(1 - x) - (\kappa + k)}}{\binom{\kappa N}{\kappa N(1 - x)}} \quad (2.57)$$

que l'on peut réorganiser bien plus commodément en

$$P(X_{t+1} = x + \Delta x \mid X_t = x) = \frac{\binom{\kappa N(1 - x)}{\kappa + k} \binom{\kappa N x}{\kappa - k}}{\binom{\kappa N}{2\kappa}} \quad (2.58)$$

L'interprétation de ce résultat est facilitée en introduisant les quantités $n_{\circ\circ} = \kappa N(1 - x)$ et $n_{\bullet\circ} = \kappa N x$ qui dénombrent respectivement les paires dirigées homogènes et hétérogènes du système.

La conséquence de toutes les hypothèses effectuées est la réduction de cette dynamique d'échanges à un nouveau double tirage sans remise, cette fois de $\kappa + k$ paires homogènes et de $\kappa - k$ paires hétérogènes, dont l'échange induit une variation Δx de la densité d'interface. Le modèle réduit à cette opération perd donc toute géométrie, toute notion d'espace et de proximité ou non des sites avec les autres. Dans la lignée des travaux de T. Rogers et A. McKane [146], le modèle de Schelling est ainsi réduit à un problème de stœchiométrie. On peut d'ailleurs comparer l'équation (2.58) avec l'expression des coefficients de la matrice de transition du modèle présenté dans leur article, bien plus simple :

$$\begin{cases} P(X_{t+1} = x - 4/N \mid X_t = x) &= \alpha(1 - x)^2 \\ P(X_{t+1} = x + 4/N \mid X_t = x) &= \beta x(x - 2/N) \\ P(X_{t+1} = x \mid X_t = x) &= 1 - \alpha(1 - x)^2 - \beta x(x - 2/N) \end{cases} \quad (2.59)$$

avec $0 < \alpha < \beta < 1/2$ des constantes liées à l'utilité des agents³³ et qui jouent ici le rôle de probabilités de transitions entre deux couples de paires d'agents. Dans le cadre présenté dans le premier chapitre, où la satisfaction est nulle en présence d'un voisinage trop différent et égale à 1 pour un agent entouré de voisins du même groupe, ces constantes devraient prendre les valeurs $\alpha = 0$ et $\beta = 1/2$, ce qui amène, dans la limite thermodynamique $N \rightarrow \infty$ à l'équation différentielle ordinaire $x'(t) + 2x^2(t) = 0$, soit

$$x(t) = \frac{x_0}{1 + 2x_0 t} \quad (2.60)$$

On obtient, dans ce cas de figure, une évolution strictement décroissante de la densité d'interface à partir de sa valeur initiale $x(0) = x_0$ jusqu'à prendre des valeurs infinitésimales. L'état "final" d'un tel système est donc toujours la ségrégation.

Par ailleurs, dans le modèle présenté dans ce chapitre, le voisinage d'un agent n'est pas simplement réduit à un autre représentant le groupe localement majoritaire. Il est redéfini aléatoirement à chaque étape, parmi toutes les paires de voisins disponibles. Lorsque $\kappa > 1$, on peut donc étudier l'effet de la tolérance sur cette dynamique d'échanges, contrairement à l'article de T. Rogers et A. McKane, qui ne fait pas état des transitions observées dans les simulations.

³³Ces constantes sont en fait fonction des niveaux de satisfaction liés au fait d'appartenir à une paire hétérogène ou homogène.

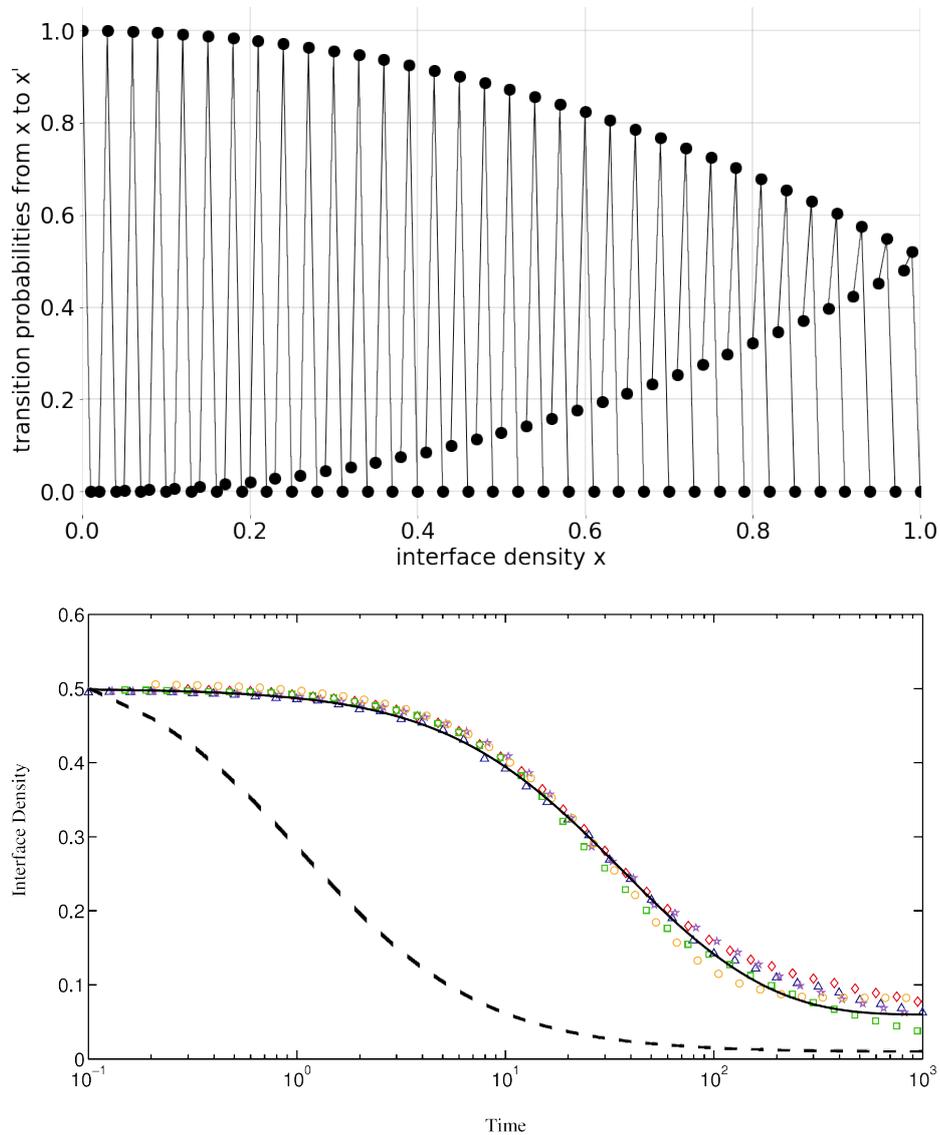


FIGURE 2.7 Probabilités de transition du modèle de Rogers-McKane [146] pour $\alpha = 0$ et $\beta = 1/2$ (en haut). Chacune des courbes juxtaposées représente la probabilité, à partir d'une valeur initiale x , d'effectuer un pas vers la gauche, de ne pas se déplacer (valeur centrale), ou d'effectuer un pas vers la droite. Tous les points de départ ne sont pas représentés dans un souci de clarté. En bas, une reproduction de la figure 3 de l'article présentant ce modèle avec, en trait plein, l'évolution temporelle pour un couple (α, β) de paramètres d'ajustement (non précisés dans l'article) à l'évolution simulée de cinq variantes (symboles). Cette figure est complétée (en trait pointillé) par l'évolution temporelle décrite par l'équation (2.60) correspondant à $(\alpha, \beta) = (0, 1/2)$.

2.3 | Irréversibilité et invariance

2.3.1 Tolérance, échanges autorisés

Pour introduire la tolérance T dans les expressions précédentes, il suffit de considérer $\mathcal{P}_{(2)}(2k) = \mathcal{P}(m_a + m_b = 2k)$ non plus comme la probabilité que l'interversion du contenu des sites a et b entraîne une variation $4k/\kappa N$ de la densité d'interface, mais comme la probabilité qu'un tel échange soit *autorisé*. Par définition de la tolérance, il faut donc que les agents initialement présents dans les sites a et b soient insatisfaits, c'est-à-dire que

$$m_a \text{ et } m_b < (1 - 2T)\kappa \quad (2.61)$$

Les aimantations locales ne pouvant accéder qu'aux valeurs comprises dans l'intervalle $\llbracket -\kappa, \kappa \rrbracket$, on peut alors écrire

$$\begin{aligned} \mathcal{P}_T(2k) &= \mathcal{P}(m_a + m_b = 2k, m_a, m_b < (1 - 2T)\kappa) \\ &= \sum_{\ell} \mathcal{P}(m_a = 2(k - \ell))\mathcal{P}(m_b = 2\ell) \end{aligned} \quad (2.62)$$

où

$$\max(-\kappa, 2k + (2T - 1)\kappa + 1) \leq 2\ell \leq \min((1 - 2T)\kappa - 1, 2k + \kappa) \quad (2.63)$$

On peut enfin obtenir, en utilisant les résultats précédents,

$$P_T \left(X_{t+1} = x + \frac{4k}{\kappa N} \mid X_t = x \right) = \frac{\mathcal{P}_T(2k)}{\mathcal{P}_{(2)}(2k)} \times \frac{\binom{\kappa N(1-x)}{\kappa+k} \binom{\kappa N x}{\kappa-k}}{\binom{\kappa N}{2\kappa}} \quad (2.64)$$

lorsque $k \neq 0$ et

$$P_T \left(X_{t+1} = x \mid X_t = x \right) = 1 - \sum_{k \neq 0} P_T \left(X_{t+1} = x + \frac{4k}{\kappa N} \mid X_t = x \right) \quad (2.65)$$

En effet, le fait de refuser un échange amène le système à demeurer dans son état précédent, donc à conserver sa densité d'interface.

Avant d'étudier l'évolution de ces probabilités en fonction de x et de T , on peut commenter le cas particulier $x = 1/2$, représenté en figure **2.8**.

On observe que selon la tolérance T , certaines densités d'interface x' sont inaccessibles à partir de $x = 1/2$. Plus précisément, on peut remarquer que, l'absence d'incrément ($k = 0$, toujours possible) mise à part,

$$k \leq \kappa_T = (1 - 2T)\kappa - 2 \quad (2.66)$$

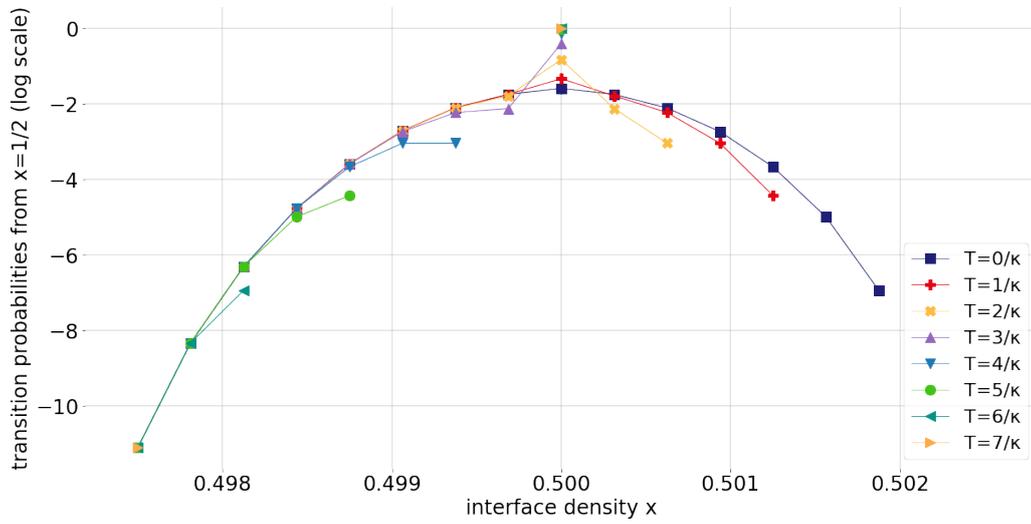


FIGURE 2.8 Évaluation numérique des probabilités de transition $P_T(x' | x = 1/2)$ pour les valeurs de tolérance T pertinentes.

Paramètres : $N = 1600$, $\kappa = 8$, $d = 1$, $p = 1/2$.

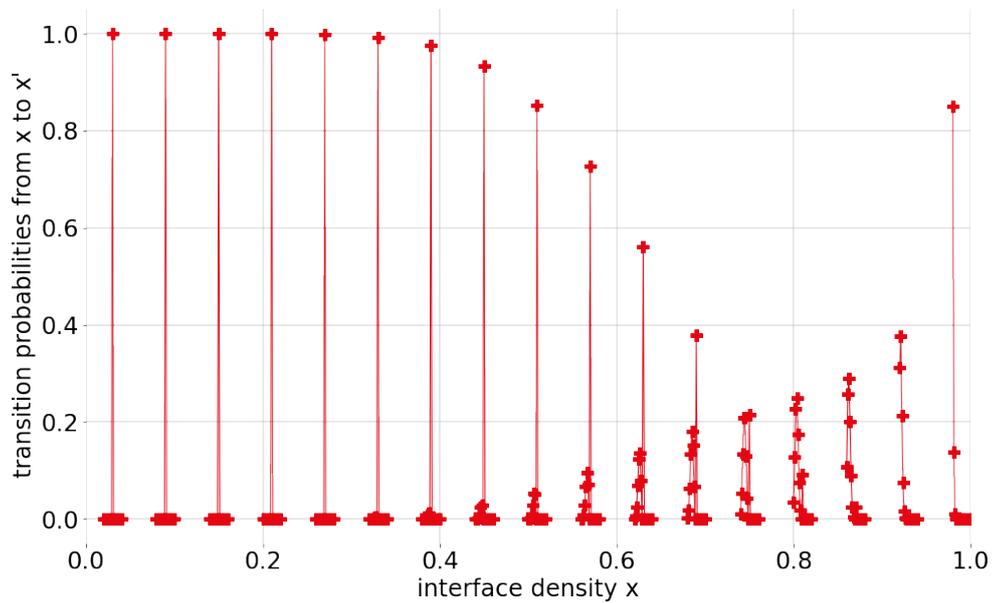


FIGURE 2.9 Probabilités de transition $P_T(x' | x)$ à $T = 1/2$ pour une sélection de valeurs de la densité d'interface x . Les courbes juxtaposées sont équivalentes à celle de la figure ci-dessus pour différentes valeurs de x .

Paramètres : $N = 400$, $\kappa = 8$, $d = 1$, $p = 1/2$.

Cette borne ne découle pas immédiatement des contraintes sur les valeurs de m_a et m_b , qui entraînent que

$$-\kappa \leq k < (1 - 2T)\kappa \quad (2.67)$$

mais plutôt des conditions sur les bornes de la somme (2.62). En effet, lorsque $k = (1 - 2T)\kappa - 1$, on peut vérifier que la condition (2.63) devient

$$2\ell = (1 - 2T)\kappa - 1 \quad (2.68)$$

ce qui n'est satisfait par aucune valeur entière de ℓ .

On peut en déduire que la matrice de transition π_T est une matrice bande. Spécifiquement,

$$\pi_{nn'} = 0 \quad \text{lorsque} \quad n' \notin \llbracket n - \kappa, n + \kappa_T \rrbracket \cup \{n\} \quad (2.69)$$

Pour d'autres densités d'interface, la distribution de probabilité perd sa relative symétrie (voir figure **2.10**), mais les incréments accessibles restent les mêmes. Lorsque x se rapproche de 0 ou 1, les effets réfléchissants des bords du domaine deviennent plus forts et la probabilité des valeurs respectivement négatives ou positives de l'incrément tend vers 0.

Pour une valeur donnée de la tolérance, on peut représenter ces distributions de probabilités $P_T(x' | x)$ bout à bout comme en figure **2.9**. On constate alors que pour $T = 1/2$, quelle que soit la valeur de x , tous les incréments sont négatifs ou nuls. C'est aussi vrai pour $T = 3/4$, mais pas pour $T = 1/4$ (en figure **2.11**).

Pour trouver la valeur critique de la tolérance T séparant ces comportements, il suffit d'inverser la relation (2.66) : on cherche $\kappa_T > 0$, c'est-à-dire

$$T < T_f = \frac{\kappa - 2}{2\kappa} \quad (2.70)$$

2.3.2 Distribution invariante, comparaison avec les simulations

Pour $T < T_f$, chaque état \mathcal{C}_x peut être atteint quelle que soit la valeur initiale x_0 (et donc l'état de départ associé \mathcal{C}_{x_0}). La chaîne est irréductible et son espace d'états, fini, est récurrent. Une distribution invariante existe et reste à déterminer. En revanche, les états à $T \geq T_f$ ne communiquent pas entre eux : les incréments sont uniquement négatifs. Chacun est sa propre classe individuelle. Dans ce cas, l'état \mathcal{C}_0 , correspondant à la valeur $x = 0$, est absorbant. La dynamique est finie (bien que longue au vu des ordres de

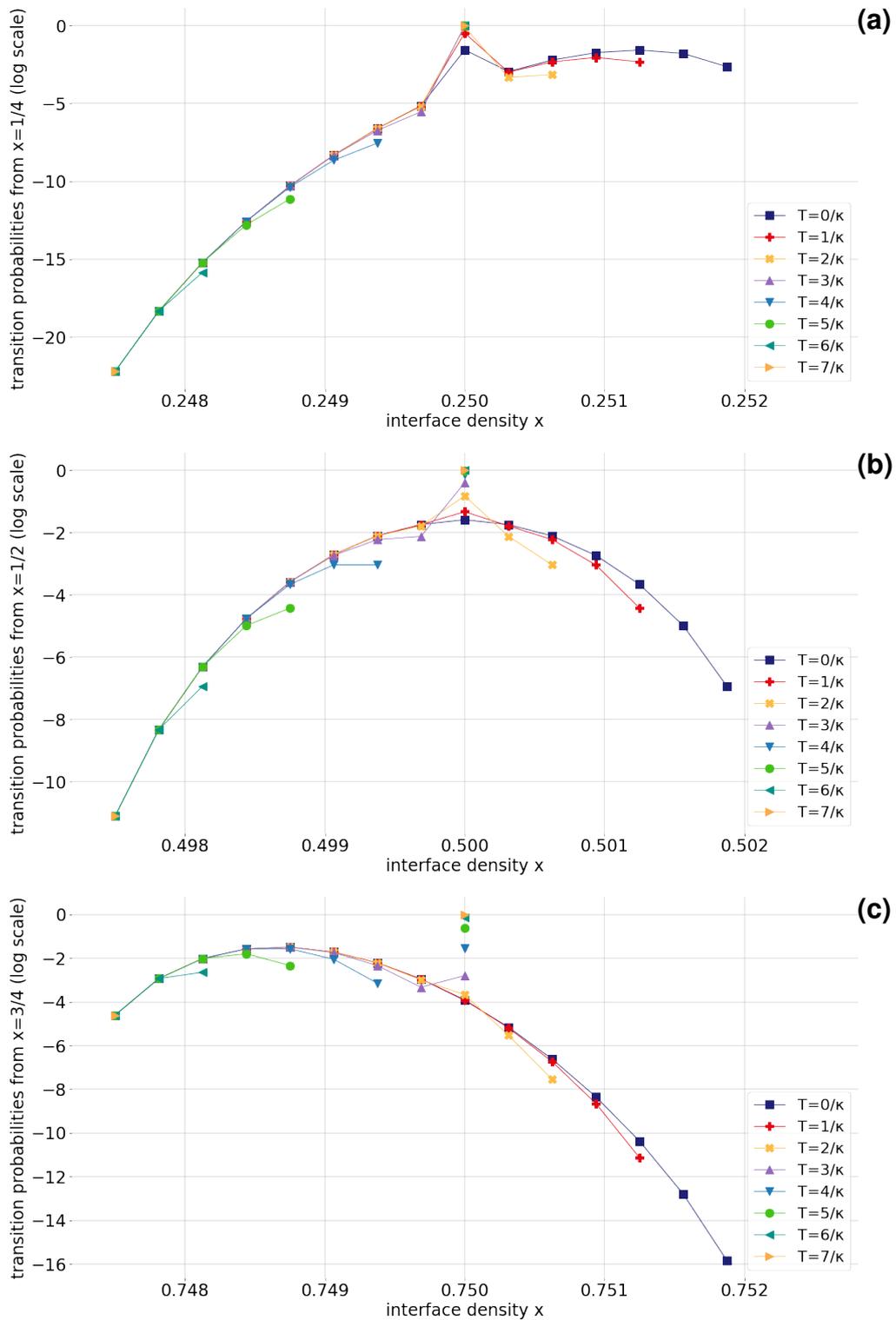


FIGURE 2.10 Évaluation numérique des probabilités de transition $P_T(x' | x)$ pour les valeurs de tolérance T pertinentes avec (a) $x = 1/4$, (b) $x = 1/2$ et (c) $x = 3/4$
Paramètres : $N = 1600$, $\kappa = 8$, $d = 1$, $p = 1/2$.

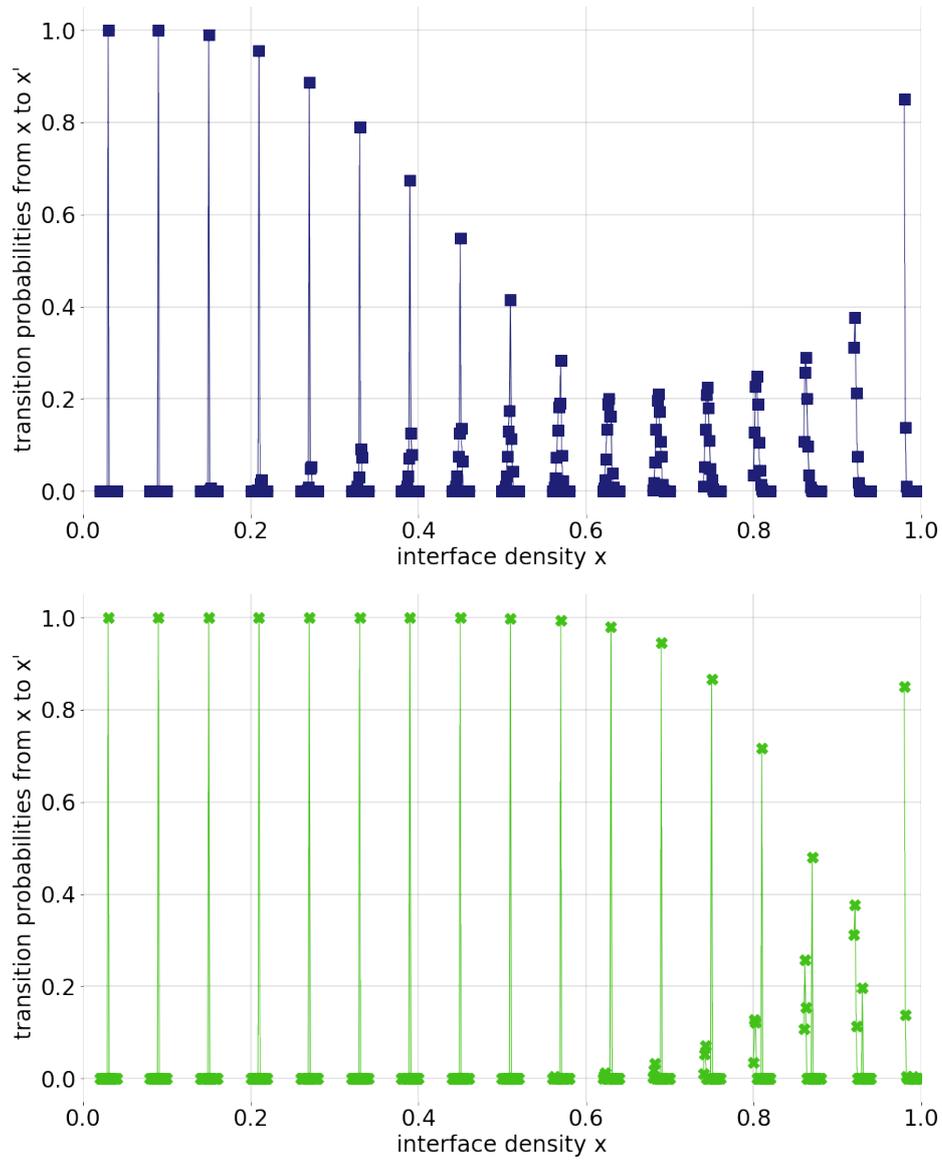


FIGURE 2.11 Évaluation numérique des probabilités de transition $P_T(x' | x)$ à $T = 1/4$ (■), $T = 1/2$ (⊕) et $T = 3/4$ (⊗) pour une sélection de valeurs de la densité d'interface x . Paramètres : $N = 400$, $\kappa = 8$, $d = 1$, $p = 1/2$.

grandeur des probabilités de transition) : l'état final est ségrégé, toutes les paires sont homogènes.

Ces propriétés permettent donc d'envisager l'évolution suivante de la distribution invariante λ en fonction de T . Lorsque $T < T_f$, le système se trouve dans l'état \mathcal{C}_x avec probabilité λ_x . On peut extraire de cette distribution l'espérance

$$x_\infty(T < T_f) = \sum_x x \lambda_x \in (0, 1] \quad (2.71)$$

qui caractérise la densité d'interface "finale" moyenne d'un tel système.

Lorsque $T \geq T_f$, comme expliqué plus haut, l'état \mathcal{C}_0 est absorbant, la densité d'interface "finale" est donc nulle :

$$x_\infty(T \geq T_f) = 0 \quad (2.72)$$

En pratique, on ne peut pas obtenir une expression analytique pour λ en fonction de T , N , κ , ... mais il est possible de déterminer une moyenne empirique $\langle x_\infty \rangle$ à partir de simulations de cette chaîne de Markov. C'est cette quantité que l'on va comparer avec les moyennes empiriques présentées dans le chapitre précédent (paragraphe **1.1.3**). Elle caractérise en effet l'état final de nombreuses simulations du modèle de Schelling et de ses variantes. L'analyse des trois graphes de la figure **2.12** illustre la validité de ce modèle et apporte un éclairage sur certaines variantes.

Dans un premier temps, en comparant simplement les résultats du modèle étudié dans ce chapitre (■) pour différentes valeurs de κ , on remarque que le seuil de tolérance $T_f = (\kappa - 2)/2\kappa$ est parfaitement respecté. Il sépare d'une part les états intégrés à la densité d'interface fluctuant autour de son espérance $x_\infty \in (0, 1]$ et, d'autre part, des états ségrégés pour lesquels $x_\infty = 0$.

On note cependant que la transition est brusque entre des valeurs de densité d'interface proche de $x_\infty = 0.5$ pour $T < T_f$ et $x_\infty \gtrsim 0$, en particulier par comparaison avec les simulations réalisées par A. Banos et al. (graphe **(b)**). Ici, la rupture est moins marquée, l'espérance de la densité d'interface chute prématurément vers des valeurs intermédiaires. Cela ne rentre pas en contradiction avec le modèle puisque x_∞ dépend de la distribution invariante, trop complexe à déterminer analytiquement mais n'étant à priori pas contrainte à demeurer proche de la valeur $1/2$.

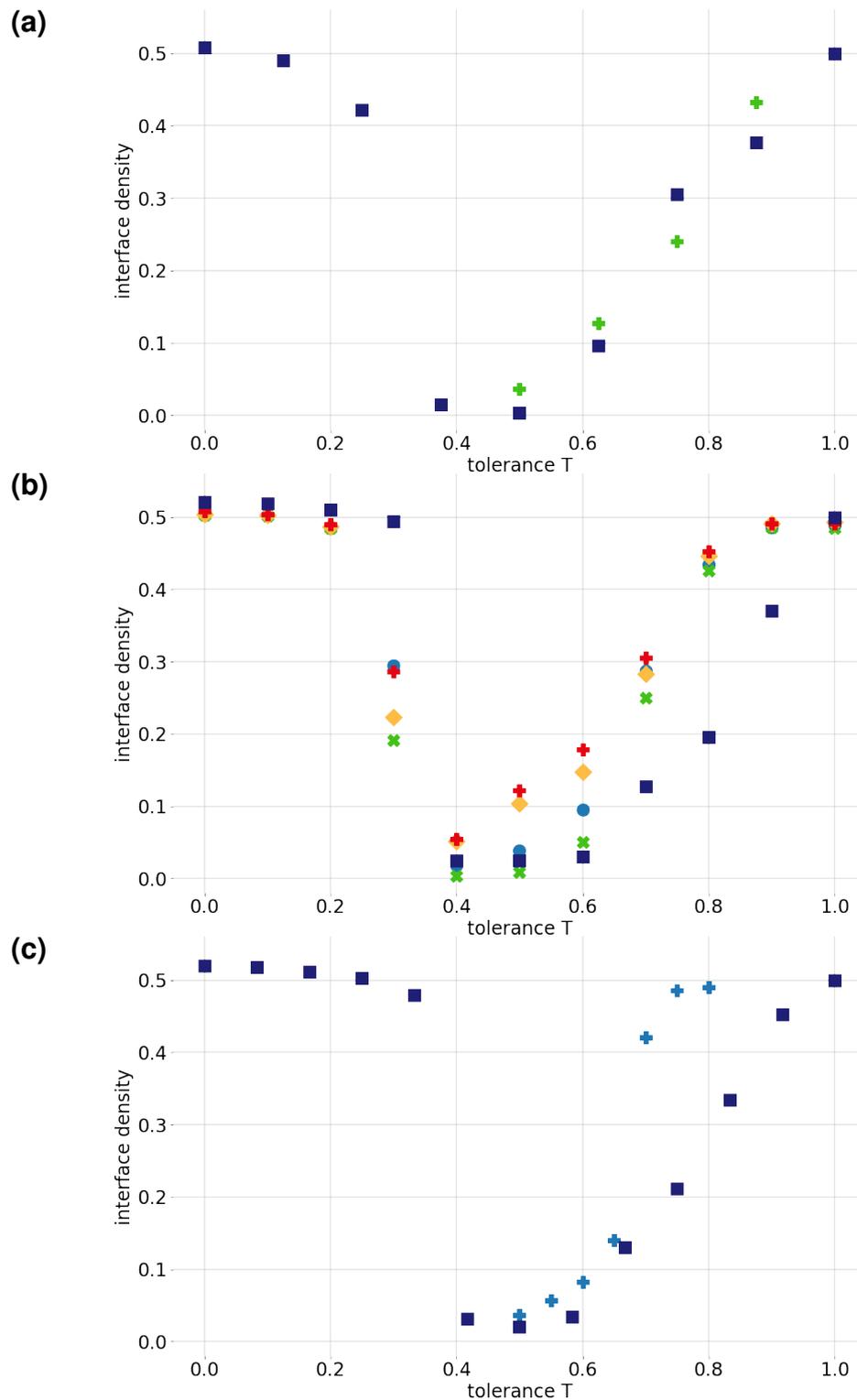


FIGURE 2.12 Comparaison de la moyenne empirique $\langle x_\infty \rangle$ (■) en fonction de T avec les densités d’interface “finales” déterminées dans les articles présentés dans le chapitre précédent :

(a) $\kappa = 8$: V. Cortez et al., 2015 (+) — $T_f = 3/8$

(b) $\kappa = 10$: A. Banos, 2012 (+, ◆, × et ●) — $T_f = 4/10$

(c) $\kappa = 12$: A. Laurie et N. Jaggi, 2003 (+) — $T_f = 5/12$

Avant de discuter de l'autre transition, revenons sur les courbes obtenues par L. Gauvin et al. et leur comparaison avec les valeurs théoriques en figure **2.13**. Lorsque d est proche de 1, on remarque que le seuil de tolérance est bien plus élevé pour ce modèle (où un agent insatisfait se déplace vers un site vacant) que pour la variante simplifiée. Cependant, lorsque $d = 0.88$, on retrouve le seuil prévu pour la dynamique par échanges (de type Kawasaki). Ce qui suit n'est qu'une conjecture, toutefois on peut penser que cette valeur de $d \simeq 7/8$ signifie qu'en moyenne, il existe un site vacant dans le voisinage (de Moore) de chaque site du système³⁴. Ainsi, il est possible pour tout agent a de se déplacer dans le voisinage d'un autre agent b , qui pourra à son tour prendre la place de a , à défaut de pouvoir échanger leur place directement. On retrouverait donc une dynamique suffisamment proche de celle par échanges pour que le résultat théorique soit robuste à cette variante, et ce spécifiquement lorsque $d \simeq 7/8$.

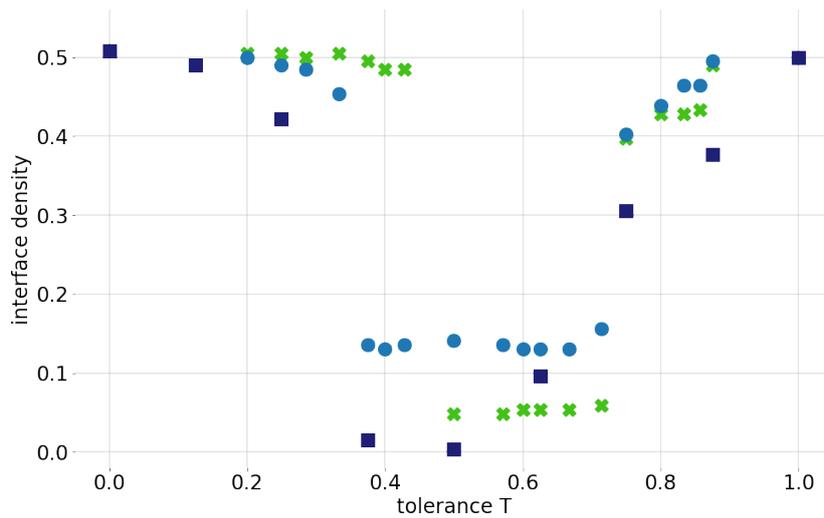


FIGURE 2.13 Comparaison de la moyenne empirique $\langle x_\infty \rangle$ (■) en fonction de T avec les densités d'interface “finales” déterminées dans l'article [70] pour $d = 0.98$ (✕) et $d = 0.88$ (●)

Revenons à la figure **2.12** : suite à cette transition brusque à $T = T_f$, force est de constater que pour les valeurs de $T > T_f$, la convergence vers $x_\infty = 0$ est extrêmement lente, à fortiori en s'approchant de $T = 1$. Ceci s'explique sans mal au vu des ordres de grandeur des probabilités de saut présentées en figures **2.10** et **2.11**. Ce système sans contrainte géométrique peine à converger vers un état de ségrégation totale, et ce de façon sensiblement similaire aux simulations du vrai modèle de Schelling sur une grille

³⁴Se référer à l'annexe **D**.

carrée ($\color{green}{+}, \color{red}{+}, \color{blue}{+}$), où la géométrie contraint le système dans une configuration pour laquelle toute transition vers un état de moindre mixité est interdite. Ce n'est pas nécessairement la seule explication : on pourrait argumenter que lorsque les agents sont très tolérants, l'état initial est globalement satisfaisant, et après quelques échanges marginaux, l'évolution de la densité d'interface devrait s'arrêter à une valeur raisonnablement proche de celle de départ. On peut remarquer que pour les simulations de L. Gauvin et al., en figure **2.13**, cette remontée est plus brutale, mais c'est une conséquence de la procédure de renormalisation appliquée par les auteurs, les variations sont simplement accentuées.

On semble avoir atteint les limites de l'approche "cinétique" par une chaîne de Markov extraite de la chaîne du modèle complet, qui permet de comprendre la première transition de phase autour de $T_f \simeq 1/2$, mais pas la suivante, observée numériquement autour de $T_c \simeq 3/4$. On verra dans les deuxième et troisième sections du chapitre suivant ce que l'approche thermodynamique, rejetée à priori pour les questions de métastabilité mentionnées dans le chapitre précédent, peut néanmoins apporter à la résolution du problème.

3. | Schelling

Un système hors équilibre ?

DANS ce chapitre nous discutons certaines réécritures thermodynamiques du modèle de Schelling. Après une rapide présentation d'outils et de concepts issus de la littérature de thermodynamique et de physique statistique, nous détaillons deux manières bien distinctes d'introduire une température dans un système régi par une dynamique de Schelling ainsi que leur influence sur les états stationnaires. Nous présentons ensuite des travaux réalisés sur une variante en système ouvert et à grand voisinage du modèle de Schelling, qui permet de déterminer un second seuil de tolérance proche de celui constaté dans les chapitres précédents. Enfin, nous proposons une méthode originale qui permet de déterminer d'une autre manière ce seuil de tolérance en étudiant la stabilité d'un noyau homogène dans la dynamique originelle de Schelling.

3.1 | Thermodynamique statistique

La thermodynamique est la science des systèmes macroscopiques [49], c'est-à-dire constitués d'un grand nombre de particules microscopiques. La physique statistique a pour but d'expliquer le comportement des systèmes macroscopiques *à partir* de leurs caractéristiques microscopiques [48]. Historiquement, la thermodynamique a été développée *avant* la physique statistique, au cours du XVIII^e siècle et au début du XIX^e, sur la base de postulats et de principes justifiés par l'expérience plutôt que par une compréhension des mécanismes sous-jacents. La physique statistique, issue de la théorie cinétique des gaz de Maxwell et Boltzmann dans la seconde moitié du XIX^e siècle et de la modélisation du mouvement brownien par Bachelier et Einstein au début du XX^e, retrouve et généralise ces principes et leurs résultats *ex post facto* et devient par là même un des piliers de la physique moderne.

On considère un système macroscopique. Les particules identiques qui le composent possèdent des caractéristiques mécaniques mesurables individuellement (position, vitesse...) et leur évolution est régie par la mécanique classique (voire la mécanique quantique). Leur nombre est cependant si phénoménal qu'il serait impossible en pratique de réaliser un tel suivi,

sans même se préoccuper des interactions pouvant exister entre elles. On effectue la mesure répétée d'une de ces caractéristiques, appelée X , et on suppose que les quantités obtenues $X_1 \dots X_N$ sont des variables aléatoires indépendantes et identiquement distribuées, en faisant l'hypothèse, raisonnable dans un contexte mécanique, que l'espérance et la variance de cette distribution commune existent. Si on définit la moyenne \bar{X} de X comme la somme pondérée des quantités $X_1 \dots X_N$, le théorème central limite permet de déterminer la variance de \bar{X} , est de la forme

$$\mathbb{V}[\bar{X}] = \mathcal{O}\left(\frac{1}{N}\right) \quad (3.1)$$

Un *système thermodynamique*, ou un système *dans la limite thermodynamique* est un système macroscopique pour lequel on choisit de négliger ces fluctuations statistiques. La variable aléatoire X devient alors une fonction \bar{X} au lieu d'être distribuée selon une loi normale $\mathcal{N}(\bar{X}, \mathbb{V}[\bar{X}])$. À titre d'exemples, l'énergie³⁵ et la pression sont généralement définies par rapport à la vitesse quadratique moyenne des particules du système :

$$E_c = \frac{1}{2}mN\bar{v}^2 \quad ; \quad p = \frac{1}{3} \frac{mN}{V} \bar{v}^2 \quad (3.2)$$

Lorsque ces grandeurs sont définies à partir de la configuration du système, on les appelle *variables internes* ou *observables*. Toutefois, certaines grandeurs peuvent être imposées à un système thermodynamique, on les appelle alors *paramètres extérieurs*³⁶. Elles contraignent le comportement des particules du système, et ainsi l'ensemble des valeurs que peuvent prendre les observables. Dans le modèle de Schelling, le "volume" disponible est donné par la taille de la grille et est une quantité fixe : il s'agit d'un paramètre extérieur.

Le système peut également être soumis à des *lois de conservation*, gouvernant l'évolution des observables en fonction des divers flux entrant et sortant du système, ou, si celui-ci est fermé ou isolé, transformant ces observables en paramètres extérieurs³⁷. Dans le modèle de Schelling, le nombre d'agents n'est jamais amené à changer : il s'agit d'un paramètre extérieur.

³⁵On considère ici l'énergie interne U d'un gaz parfait, c'est-à-dire d'un système de N particules de masse m ponctuelles sans interaction, occupant un volume V . Dans le cas général, il faut ajouter une énergie potentielle à ce terme issu des énergies cinétiques individuelles. Par ailleurs, E et U sont confondues uniquement dans le référentiel dans lequel le système macroscopique est au repos.

³⁶Le volume d'un contenant, le champ magnétique ambiant, la pression dans une enceinte en équilibre mécanique avec l'atmosphère...

³⁷Le nombre de molécules d'eau dans une section de conduite dépend du flux entrant en

Lorsqu'un système est soumis à un ensemble donné de contraintes, il est possible de trouver de nombreuses configurations microscopiques, qu'on appellera *micro-états*, pour lesquelles une ou plusieurs observables présentent des valeurs identiques³⁸. Ces configurations sont alors regroupées dans un ensemble statistique appelé *macro-état*. On considère souvent que l'étude d'un système macroscopique consiste en réalité en l'étude d'une collection de répliques de ce système subissant la même évolution. On peut alors construire une distribution de probabilités $\{P_\ell\}_\ell$ des micro-états, vérifiant les deux contraintes

$$\forall \ell, 0 \leq P_\ell \leq 1 \quad ; \quad \sum_\ell P_\ell = 1 \quad (3.3)$$

La probabilité P_ℓ de trouver le système dans le micro-état ℓ est alors équivalente à la probabilité de sélectionner une copie du système dans l'état ℓ . Cette deuxième approche permet alors d'utiliser les outils classiques en probabilités, en considérant non plus un système unique dans une sorte d'évolution temporelle fluctuante, mais bien un ensemble duquel on peut extraire des propriétés. C'est le *principe ergodique*, qui postule l'équivalence entre moyenne temporelle et moyenne d'ensemble, ou espérance, sur la distribution de probabilités $\{P_\ell\}_\ell$.

Pour calculer ces probabilités, on peut considérer qu'il existe deux approches complémentaires, illustrées en quelque sorte par ce chapitre et le précédent.

L'approche *cinétique* consiste à établir l'équation maîtresse de l'évolution temporelle du système, en explicitant la matrice de transition, dans l'espoir, comme dans le chapitre **2**, de pouvoir déterminer sa distribution invariante. Cette approche demande de pouvoir construire les probabilités de transition à partir des interactions microscopiques ce qui, comme expliqué au début de cette introduction à la thermodynamique et à la physique statistique, est justement impossible pour des systèmes physiques réels, possédant bien trop de degrés de liberté.

amont et sortant en aval, celui dans une cuve hermétique est constant. Dans les deux cas, il y a conservation de la matière : la variation du nombre de molécules entre un instant t et $t + dt$ est directement proportionnelle à la différence des flux entrant et sortant à l'instant t , autrement dit égale à 0 dans le cas fermé.

³⁸Ou, dans le cas d'une variable continue, appartenant à un même intervalle de type $[X_0, X_0 + \delta X]$ avec $\delta X/X_0 \ll 1$.

L'approche *thermodynamique* consiste à déterminer l'équilibre thermodynamique du système, plus ou moins équivalent à la distribution invariante³⁹ directement à partir d'un postulat, appelé le postulat fondamental de la physique statistique et énoncé comme suit⁴⁰ :

À l'équilibre thermodynamique, la distribution de probabilités des micro-états accessibles d'un système est, parmi toutes celles vérifiant les contraintes extérieures imposées à ce système, celle qui maximise l'entropie.

L'entropie est une fonction définie à partir de la distribution de probabilité introduite précédemment :

$$S = -k \sum_{\ell} P_{\ell} \ln P_{\ell}, \quad \lambda > 0 \quad (3.4)$$

Il s'agit d'une des fonctions satisfaisant de nombreux critères⁴¹ (positivité, additivité, continuité...) et on peut montrer sa pertinence pour quantifier le "manque d'information" à l'aide de quelques exemples simples.

Si l'état du système est parfaitement déterminé (il se trouve par exemple dans l'état ℓ' : $\forall \ell P_{\ell} = \delta_{\ell\ell'}$), alors on peut vérifier $S = 0$: le manque d'information est nul.

³⁹La définition de l'*équilibre thermodynamique* met en évidence la capacité de la thermodynamique à construire une théorie aux résultats tout à fait satisfaisants sur une fondation de "bon sens" à la définition imprécise. La définition de l'équilibre dépend en effet de l'observateur et plus précisément, de la *durée d'observation*. Est ainsi considéré comme à l'équilibre un système isolé *invariant* sur une durée largement supérieure à la durée d'observation. Cette notion de durée d'observation permet d'effectuer une classification des états d'équilibre. Ainsi, le diamant n'est pas un état d'équilibre du carbone sous forme solide aux temps longs (très longs), contrairement au graphite.

⁴⁰Ce n'est pas la forme sous laquelle ce postulat est généralement présenté, à savoir :

La distribution de probabilités des micro-états accessibles d'un système isolé à l'équilibre thermodynamique est uniforme.

qui nécessite d'une part de connaître la définition d'un système isolé, que l'on abordera plus loin, et d'autre part, lorsque cette condition n'est pas imposée par les paramètres extérieurs, de construire des systèmes isolés fictifs englobant l'objet de l'étude (se référer à [48], chapitre 3). Cette reformulation est empruntée au précieux cours de physique statistique de B. Diu et al. [48], duquel est par ailleurs tirée une grande partie de l'introduction de ce chapitre. Dans certains cas, il apparaît toutefois indispensable de revenir à la première formulation, plus rigoureuse physiquement.

⁴¹Voir [48], complément I.G, p. 98.

Si à présent on introduit de l'indétermination, en supposant que le système peut se trouver dans un état parmi deux accessibles que l'on suppose équiprobables, l'entropie est $S = k \ln 2$. En théorie de l'information, la constante multiplicative est alors fixée à $k = 1/\ln 2$, ce qui fait de $S = 1$ le nombre de questions à poser sur le système pour en connaître l'état. Dans ce cas binaire, une seule question suffit ("le système est-il dans l'état A?") : le manque d'information est de 1 bit.

Généralisons ce dernier exemple à W états équiprobables. On retrouve ici la formule de Boltzmann (1877)

$$S = k \ln W \quad (3.5)$$

avec $k = k_B$ où k_B est la constante de Boltzmann⁴². Ce choix de k permet la correspondance de cette définition statistique de l'entropie avec celle, thermodynamique, de Clausius (1865). On considèrera par la suite $k = 1$ par simplicité d'écriture, sans que cela change quoi que ce soit aux conclusions tirées.

Le postulat fondamental de la physique statistique ramené à un problème d'optimisation sous contraintes, on peut à présent examiner le cas du modèle de Schelling, de son étude perturbative et des variantes de Glauber. En particulier, on peut définir les trois ensembles thermodynamiques correspondant aux contraintes imposées :

- ensemble *microcanonique* : les échanges de particules et d'énergie avec l'extérieur sont impossibles, le système est *isolé*.
- ensemble *canonique* : les échanges de particules avec l'extérieur sont impossibles, le système est *fermé*.
- ensemble *grand-canonique* : le système est ouvert, au contact d'un thermostat permettant les échanges thermiques et d'un réservoir de particules permettant les échanges de particules.

Dans chacun de ces trois cas, on pourra donc calculer les probabilités P_ℓ associées aux micro-états du système à l'équilibre thermodynamique.

⁴²Dont la valeur a été fixée à $1.380649 \times 10^{-23} \text{ J.K}^{-1}$ lors de la dernière révision du Système international d'unités en 2019.

3.2 | Un système isolé ?

Le choix d'appliquer des principes de la thermodynamique et de la physique statistique au modèle de Schelling est pertinent au vu des définitions qui ont été données de ces disciplines. Schelling voulait présenter l'impact des décisions individuelles des agents sur le comportement collectif du système. De plus, sa sensibilité aux conditions initiales ne fait que renforcer l'attrait pour une approche probabiliste : à défaut de décrire la dynamique d'un système donné, on peut exhiber des caractéristiques communes à l'évolution d'une collection de systèmes macroscopiquement équivalents.

Il l'est moins lorsque l'on considère qu'un système physique, tangible, est soumis à des perturbations extérieures incontrôlables pouvant modifier l'état macroscopique en une durée très courte sans pour autant affecter son énergie totale de manière appréciable. De telles perturbations sont généralement modélisées par un bruit aléatoire tel que l'agitation thermique, qu'il ne faut pas confondre avec les fluctuations statistiques, qui caractérisent la volatilité des grandeurs physiques au sein d'un macro-état donné.

Le modèle de Schelling n'est pas un modèle de particules en interaction, soumises à de l'agitation thermique jusqu'à atteindre son équilibre thermodynamique. Il n'est pas une simulation Metropolis évoluant dans un paysage énergétique donné, de façon globalement décroissante mais s'autorisant ponctuellement des sauts permettant de franchir les barrières de potentiels jusqu'à atteindre l'équilibre le plus stable.

On peut tout de même définir une énergie associée au système étudié qui décroît au cours de son évolution. Mais cette dernière est gouvernée par des mécanismes locaux et ponctuels bien plus proches de ceux que l'on trouve dans les modèles contraints cinétiquement [57, 66, 67] ou KCM (*kinetically constrained models*), sur lesquels on reviendra dans la section **3.2.2**, que par des perturbations aléatoires.

3.2.1 Température et réversibilité

Distribution de Boltzmann Revenons au postulat fondamental de la physique statistique dans le cadre canonique. Le système est considéré comme étant en contact avec un thermostat, c'est-à-dire un réservoir d'énergie avec lequel les échanges sont possibles. À l'équilibre thermodynamique, qui suppose la fin des échanges entre le système et le thermostat (ou à défaut

l'annulation de la résultante des échanges), l'énergie moyenne du système est alors imposée par ce contact.

La maximisation de l'entropie doit donc se faire sous deux contraintes : la condition de normalisation des probabilités et la conservation de l'énergie moyenne

$$\sum_{\ell} P_{\ell} = 1 \quad ; \quad \langle E \rangle = \sum_{\ell} E_{\ell} P_{\ell} = E \quad (3.6)$$

On peut alors écrire le lagrangien suivant en introduisant les multiplicateurs de Lagrange α et β

$$\mathcal{L} = - \sum_{\ell} P_{\ell} \ln P_{\ell} - \alpha \left(\sum_{\ell} P_{\ell} \right) - \beta \left(\sum_{\ell} E_{\ell} P_{\ell} - E \right) \quad (3.7)$$

et dont les états stationnaires sont de la forme

$$\forall \ell, P_{\ell} = \frac{e^{-\beta E_{\ell}}}{Z} \quad \text{avec} \quad Z = \sum_{\ell} e^{-\beta E_{\ell}} \quad (3.8)$$

Ces probabilités, par construction, sont positives, et on peut montrer qu'elles maximisent bien l'entropie. On peut également montrer que les paramètres β et Z , complètement déterminés par les contraintes (3.6), s'ils existent, sont uniques. Cette distribution est appelée la distribution de Boltzmann, il s'agit d'une forme discrète et restreinte à l'intervalle $[\min(E_{\ell}), \max(E_{\ell})]$ ⁴³ d'une distribution exponentielle.

L'entropie peut alors être définie comme une fonction du multiplicateur β :

$$S = - \sum_{\ell} P_{\ell} \ln P_{\ell} = \ln Z + \beta E \quad (3.9)$$

ou encore par rapport à la fonction de partition Z

$$S = \ln Z - \beta \frac{\partial}{\partial \beta} \ln Z \quad (3.10)$$

Reste à définir physiquement cette grandeur β . L'équation (3.9) permet d'établir que S et β^{-1} sont des variables conjuguées pour l'énergie, autrement dit un couple de variables thermodynamiques dont le produit est de la dimension d'une énergie. Pour cela, l'une doit être intensive et l'autre extensive. L'entropie étant une grandeur *extensive*⁴⁴, β^{-1} est donc une grandeur *intensive*⁴⁵.

⁴³Et donc normalisée sur ce même intervalle, contrairement à la distribution exponentielle classique.

⁴⁴L'entropie de l'union de deux systèmes est égale à la somme des entropies de chacun des systèmes : $S(A \cup B) = S(A) + S(B)$. Cette propriété se confond avec l'additivité dans la limite thermodynamique.

⁴⁵C'est-à-dire non-extensive : la propriété ci-dessus n'est pas vérifiée.

On a vu plus haut dans un cas particulier que l'entropie en physique pouvait s'écrire $S = k_B \ln W$, autrement dit, la dimension de S est celle de k_B , ce qui fait de β^{-1} une grandeur en Kelvin, c'est-à-dire une *température*.

Température La température Θ est une notion qui semble intuitive (le "chaud", le "froid") mais qui possède de nombreuses définitions, certaines empiriques, d'autres très techniques, caractérisant toutes une propriété de l'équilibre d'un système. En thermométrie, la loi dite des gaz parfaits est observée

$$pV = Nk_B\Theta \quad (3.11)$$

rendant constant le produit de la pression et du volume d'un gaz isolé (dans une enceinte fermée calorifugée par exemple), lorsque la pression est suffisamment faible. On peut réécrire cette relation pour mettre en évidence la température comme égale, à une constante multiplicative près, au rapport N/pV , où N/V est la densité particulaire du gaz. En physique statistique, à l'aide de l'énergie et de la pression (3.2) d'un système thermodynamique définies en début de chapitre, on peut alors montrer que

$$\frac{3}{2}k_B\Theta = \frac{1}{2}m\overline{v^2} = E_c \quad (3.12)$$

reliant ainsi température et énergie cinétique et justifiant enfin l'appellation "*agitation thermique*" : la température d'un système est directement liée aux fluctuations des vitesses de déplacement des particules qui le composent.

Mais c'est une troisième définition que l'on va utiliser⁴⁶, qui découle de l'identité thermodynamique de Gibbs

$$\frac{1}{\Theta} = \frac{\partial S}{\partial E} \quad (3.13)$$

ce qui permet de vérifier, à l'aide de l'équation (3.9), que $\beta^{-1} = \Theta$. Cette température, appelée température thermodynamique, caractérise l'équilibre. En l'occurrence, dans le cadre canonique, l'équilibre entre le système étudié et le thermostat qui lui impose cette énergie moyenne E . Cette définition reste vraie dans les autres cadres, on l'utilisera pour déterminer la température du modèle de Schelling "non-perturbé" en section **3.2.2**.

⁴⁶Il en existe beaucoup d'autres, se référer à l'excellent article de revue de Casas-Vásquez et Jou [33].

Variantes perturbatives La variante perturbative du modèle de Schelling se place, comme on va le voir, dans l'ensemble canonique. Lorsque sa distribution de probabilités caractéristique (3.8) n'y figure pas directement [20, 47], la règle dynamique couramment employée [77, 78, 80, 90, 111, 159, 178–180] est la suivante⁴⁷

$$w_{nn'}(\boldsymbol{\sigma}) = \frac{1}{1 + e^{-\beta \Delta u_{nn'}}} (1 - \delta_{s_n s_{n'}}) \quad (3.14)$$

où $\Delta u_{nn'}$ est la variation d'utilité des agents impliqués dans l'échange du contenu des sites n et n' à partir de la configuration initiale $\boldsymbol{\sigma}$. Ce choix de probabilité de transition semble impliquer une décision prise à l'échelle du système tout entier, et non à l'échelle du voisinage des agents insatisfaits. D'où vient cette expression et en quoi un système l'utilisant évolue-t-il dans le cadre canonique ?

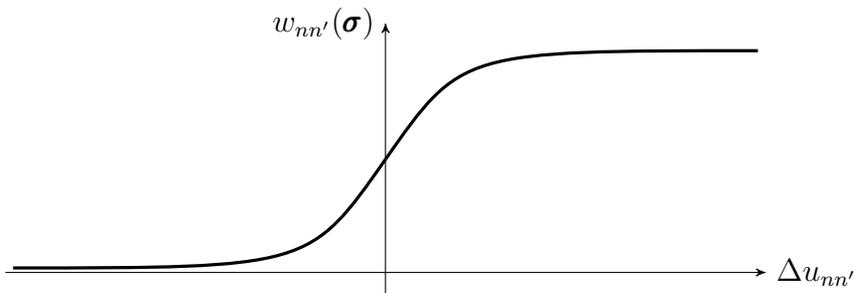


FIGURE 3.1 Règle dynamique (probabilité de transition) couramment utilisée dans la variante perturbative du modèle de Schelling.

Tout d'abord, on peut noter la présence de β , c'est-à-dire d'une température. Mais la dernière définition de cette grandeur est valide dans les autres ensembles thermodynamiques. On peut alors s'interroger sur la forme de cette probabilité de transition et sa signification. La condition entre parenthèses mise à part, il s'agit d'une sigmoïde, ce qui signifie que $w_{nn'}$ est non-nul pour toute valeur de $\Delta u_{nn'}$, il est donc possible d'effectuer tous les échanges imaginables, y compris ceux diminuant l'utilité des agents du système. Dans la limite $\Theta \rightarrow 0$, ou $\beta \rightarrow \infty$, la probabilité d'un échange diminuant l'utilité globale tend vers 0, tandis que la probabilité d'effectuer un échange souhaité tend vers 1, autrement dit

$$\lim_{\beta \rightarrow \infty} w_{nn'}(\boldsymbol{\sigma}) = (1 - \delta_{s_n s_{n'}}) \mathbb{1}_{]0, +\infty[}(\Delta u_{nn'}) \quad (3.15)$$

⁴⁷À une constante de normalisation $z(\boldsymbol{\sigma})$ près. On peut considérer que $w_{nn'} = \tilde{w}_{nn'} z(\boldsymbol{\sigma})$, où $\tilde{w}_{nn'}$ est la probabilité de transition effective.

Si on note $\sigma_{nn'}$ la configuration après échange, alors la variation d'utilité est

$$\Delta u_{nn'} = u(\sigma_{nn'}) - u(\sigma) = \begin{cases} 2 & m_n, m_{n'} < (1 - 2T)\kappa \\ -2 & m_n, m_{n'} \geq (1 - 2T)\kappa \\ 0 & \text{sinon} \end{cases} \quad (3.16)$$

d'où

$$\begin{aligned} \lim_{\beta \rightarrow \infty} w_{nn'}(\sigma) &= (1 - \delta_{s_n s_{n'}}) \mathbb{1}_{]0, +\infty[}(\Delta u_{nn'}) \\ &= (1 - \delta_{s_n s_{n'}}) \mathbb{1}_{[-\kappa, (1-2T)\kappa[}(m_n) \mathbb{1}_{[-\kappa, (1-2T)\kappa[}(m_{n'}) \end{aligned} \quad (3.17)$$

ce qui correspond exactement à l'expression (2.17) utilisée pour décrire le modèle non-perturbé de Schelling avec échanges d'agents.

Pour autant, on peut se demander dans quelle mesure la distribution invariante du processus limite lorsque $\Theta \rightarrow 0$ correspond à celle du modèle originel. Pour cela, il faut comprendre la dérivation de l'expression (3.14). Une façon classique de l'obtenir à partir des grandeurs locales est présentée ici, afin de pouvoir discuter de la validité de certaines de ses étapes. Cette méthode provient de la littérature de physique statistique portant sur les verres de spins [100]. Pour calculer les probabilités de transition, plutôt que de considérer l'utilité des agents, ce sont les énergies d'interaction qui interviennent directement, autrement dit, la somme des aimantations⁴⁸.

$$E_\sigma = - \sum_i m_i \quad (3.18)$$

Après échange, cette énergie devient

$$E_{\sigma_{nn'}} = - \sum_i m_i + 2(m_n + m_{n'}) \quad (3.19)$$

Les probabilités de transition devant favoriser les échanges diminuant l'énergie du système, il est alors conventionnel [100] de choisir

$$w_{nn'}(\sigma) = \alpha (1 - \gamma(m_n + m_{n'})) (1 - \delta_{s_n s_{n'}}) \quad (3.20)$$

Pour déterminer le paramètre γ dans l'ensemble canonique, il suffit de considérer le cas particulier de l'équilibre thermodynamique pour lequel le système peut vérifier le bilan détaillé suivant⁴⁹ :

$$P_\sigma w_{nn'}(\sigma) = P_{\sigma_{nn'}} w_{nn'}(\sigma_{nn'}) \quad (3.21)$$

⁴⁸On considère ici un hamiltonien de type Ising plutôt que Blume-Emery-Griffiths (utilisé dans [70]) car dans un système fermé sans site vide, tous les termes supplémentaires sont constants, on peut alors redéfinir l'origine de l'énergie sans perte de généralité.

⁴⁹Se référer à la section **2.1.1**, en particulier à la dérivation de (2.9).

Or, à l'équilibre thermodynamique, pour une température imposée Θ non-nulle, les configurations (ou micro-états) σ sont distribuées selon

$$P_{\sigma} = \frac{e^{-\beta E_{\sigma}}}{Z} \quad (3.22)$$

Le bilan détaillé permet donc d'écrire

$$\begin{aligned} \frac{w_{nn'}(\sigma)}{w_{nn'}(\sigma_{nn'})} &= e^{-\beta(E_{\sigma_{nn'}} - E_{\sigma})} \\ \Leftrightarrow \frac{1 - \gamma(m_n + m_{n'})}{1 + \gamma(m_n + m_{n'})} &= e^{-2\beta(m_n + m_{n'})} \end{aligned} \quad (3.23)$$

soit

$$\gamma(m_n + m_{n'}) = \tanh\left(\beta(m_n + m_{n'})\right) \quad (3.24)$$

et donc

$$w_{nn'}(\sigma) = \frac{1}{1 + e^{2\beta(m_n + m_{n'})}} \left(1 - \delta_{s_n s_{n'}}\right) \quad (3.25)$$

On voit donc apparaître ici une expression fortement similaire à l'expression recherchée (3.14), dans laquelle on reconnaît un facteur caractérisant non pas la variation d'utilité du système mais bien la variation d'énergie

$$\Delta E_{nn'} = E_{\sigma_{nn'}} - E_{\sigma} = 2(m_n + m_{n'}) \quad (3.26)$$

Qu'a-t-elle de commun avec la variation d'utilité ? Si l'on choisit, comme M. Pollicott et H. Weiss [138], une utilité linéaire par rapport à l'aimantation locale $u(m_n) = m_n$, on obtient bien

$$\Delta u_{nn'} = -2(m_n + m_{n'}) = -\Delta E_{nn'} \quad (3.27)$$

et ainsi l'expression de la probabilité de transition $w_{nn'}$ choisie comme un postulat dans cette variante.

Une majorité des transitions permettant de faire décroître l'énergie du système est acceptée, mais pas toutes, tandis qu'une part de celles faisant augmenter cette énergie est tolérée. Cette chaîne de Markov est intrinsèquement réversible, c'est le principe même permettant d'écrire le bilan détaillé (3.21). Or on a montré dans le chapitre **2** que la chaîne extraite est irréversible pour certaines valeurs de la tolérance. Par ailleurs, lorsqu'elle est effectivement réversible, on pourrait questionner la "continuité" de l'approche perturbative quand on constate l'origine de l'expression (forme de la probabilité de transition choisie par convention avec l'exemple ferromagnétique, changement de définition de l'utilité...).

Dans le cadre de simulations numériques [149], on peut trouver d'autres formes de contraintes, comme celle de Métropolis acceptant toutes les transitions faisant décroître l'énergie du système et certaines la faisant augmenter

$$w_{nn'}(\boldsymbol{\sigma}) = \min \left(1, e^{-\beta \Delta E_{nn'}} \right) \quad (3.28)$$

Cette chaîne reste réversible et s'expose aux mêmes critiques que la précédente.

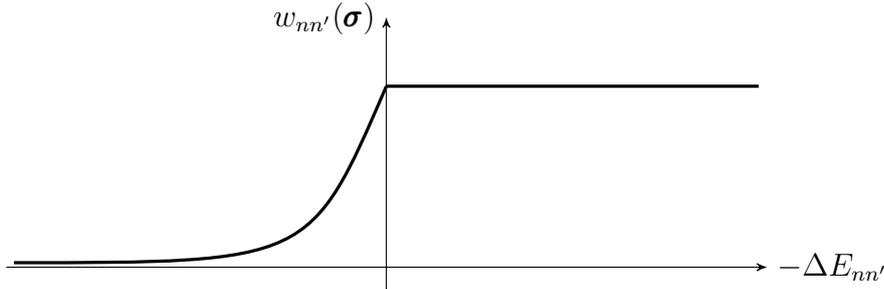


FIGURE 3.2 Règle dynamique (probabilité de transition) de Métropolis.

Enfin, dans certains travaux [114, 119, 159, 160, 163], le modèle de Schelling est traité comme un modèle d'Ising-Glauber pour lequel la tolérance joue le rôle de la température, ce qui correspond en réalité à une toute autre classe de modèles : le rôle de seuil de la tolérance disparaît, et avec lui les valeurs critiques T_f et T_c .

3.2.2 Entropie microcanonique et température négative

Dans le modèle de Schelling, du fait des contraintes sur le déplacement des agents du système, le cadre est nécessairement microcanonique. Il reste possible d'y définir une température, mais elle n'a ici pas le sens d'une mesure de l'agitation aléatoire des agents, simplement celui découlant de sa définition (3.13) à partir de l'entropie.

Dans le formalisme microcanonique, le postulat fondamental de la physique statistique se réduit à un problème d'optimisation de l'entropie sous l'unique contrainte

$$\sum_{\ell} P_{\ell} = 1 \quad (3.29)$$

Comme précédemment, on peut écrire le lagrangien

$$\mathcal{L} = - \sum_{\ell} P_{\ell} \ln P_{\ell} - \alpha \left(\sum_{\ell} P_{\ell} - 1 \right) \quad (3.30)$$

Ses états stationnaires sont de la forme

$$\forall \ell, P_\ell = \frac{1}{W} \quad (3.31)$$

où W est le nombre de micro-états accessibles au système. On retrouve l'autre formulation du postulat : les micro-états d'un système isolé à l'équilibre thermodynamique sont équiprobables. On s'est ramené au système qui permettait de démontrer la formule de Boltzmann

$$S = \ln W \quad (3.32)$$

Dans le formalisme microcanonique, l'entropie est donc simplement une mesure du nombre d'états accessibles au système. Le suivi de l'entropie d'un système subissant la dynamique de Schelling va donc nécessiter de connaître, pour toute énergie, le nombre de configurations accessibles $W(E)$. On en déduit alors $S(E)$, et ainsi, une expression de la température $\Theta(E)$.

L'entropie associée au macro-état d'énergie E fait intervenir le cardinal $W(E)$ de l'ensemble des micro-états le constituant⁵⁰. L'énergie associée au modèle de Schelling s'écrit [70, 71, 74] :

$$E_S(\boldsymbol{\sigma}) = - \sum_i m_i - (2T - 1) \sum_i \kappa_i \quad (3.33)$$

tandis que la densité d'interface est définie par

$$x_S(\boldsymbol{\sigma}) = \frac{1}{2} \left(1 - \frac{\sum_i m_i}{\sum_i \kappa_i} \right) \quad (3.34)$$

Dans un système sans site vide et paritaire, avec $(d, p) = (1, 1/2)$, à tolérance fixée⁵¹, le second terme de l'énergie (et le dénominateur de la densité d'interface) est constant. On peut donc se contenter du premier, et on considère pour la suite

$$E(\boldsymbol{\sigma}) = - \sum_i m_i \quad (3.35)$$

ou, de façon équivalente,

$$E(\boldsymbol{\sigma}) = (2x(\boldsymbol{\sigma}) - 1)\kappa N \quad (3.36)$$

⁵⁰Formellement, cet ensemble est celui des micro-états dont l'énergie est comprise entre E et $E + \delta E$ mais dans ce cadre discret, on pourra simplement considérer δE négligeable devant $\min \{E(\boldsymbol{\sigma}') - E(\boldsymbol{\sigma}) \mid \boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \boldsymbol{\Sigma}\}$.

⁵¹La tolérance est fixée avant l'instant initial. Se référer à [131] pour une variante du modèle de Schelling où la tolérance varie au cours de la dynamique.

Plutôt que de considérer l'entropie et la température comme des fonctions de l'énergie, on prend alors le parti de les exprimer en fonction de l'observable principale de cette thèse, la densité d'interface :

$$\tilde{S}(x) = \ln \tilde{W}(x) = \ln \left(W(E(x)) \right) \quad \text{et} \quad \Theta = \frac{dE}{dx} \left(\frac{d\tilde{S}}{dx} \right)^{-1} \quad (3.37)$$

La probabilité de trouver une configuration dont la densité d'interface est x a été donnée dans le chapitre précédent par l'équation (2.47), mais elle peut aussi s'écrire, par définition,

$$\mathcal{P} \left(\sum_i m_i = (1 - 2x)\kappa N \right) = \frac{\tilde{W}(x)}{W_0} \quad (3.38)$$

où W_0 dénombre l'ensemble de toutes les configurations du système partageant les valeurs des paramètres extérieurs d et p fixées par la configuration initiale σ_0 .

Ainsi, lorsque $d = 1$ et $p = 1/2$, on trouve⁵²,

$$\tilde{W}(x) = \binom{N}{N/2} \times 2^{-\kappa N} \binom{\kappa N}{\kappa N x} \quad (3.39)$$

et l'entropie du macro-état de densité d'interface x peut alors s'écrire

$$\tilde{S}(x) = \phi_{\kappa, N} - \ln \left((\kappa N (1 - x))! \right) - \ln \left((\kappa N x)! \right) \quad (3.40)$$

où $\phi_{\kappa, N}$ est une constante ne dépendant que des paramètres géométriques κ et N .

La température microcanonique est donc donnée par l'expression suivante

$$\Theta(x) = \frac{2}{H_{\kappa N(1-x)} - H_{\kappa N x}} \quad (3.41)$$

où H_n est le n -ième nombre harmonique. Dans la limite thermodynamique, on peut alors montrer que

$$\Theta(x) \sim 2 \left[\ln \left(\frac{1-x}{x} \right) \right]^{-1} \quad \text{lorsque} \quad 1 \ll xN \ll N \quad (3.42)$$

La température Θ présente des propriétés intrigantes à première vue : elle n'est pas définie en $x = 1/2$, est négative sur tout le domaine $]1/2, 1[$... Il s'agit d'un exemple de système semblable au cristal paramagnétique parfait⁵³. Les valeurs négatives prises par la température caractérisent simplement la décroissance de l'entropie en tant que fonction de la densité d'interface, et donc de l'énergie. Sa dérivée, la température inverse, est alors nécessairement négative.

⁵²Se référer à l'annexe **E**.

⁵³Se référer à [48], complément II.B, pp. 180–189.

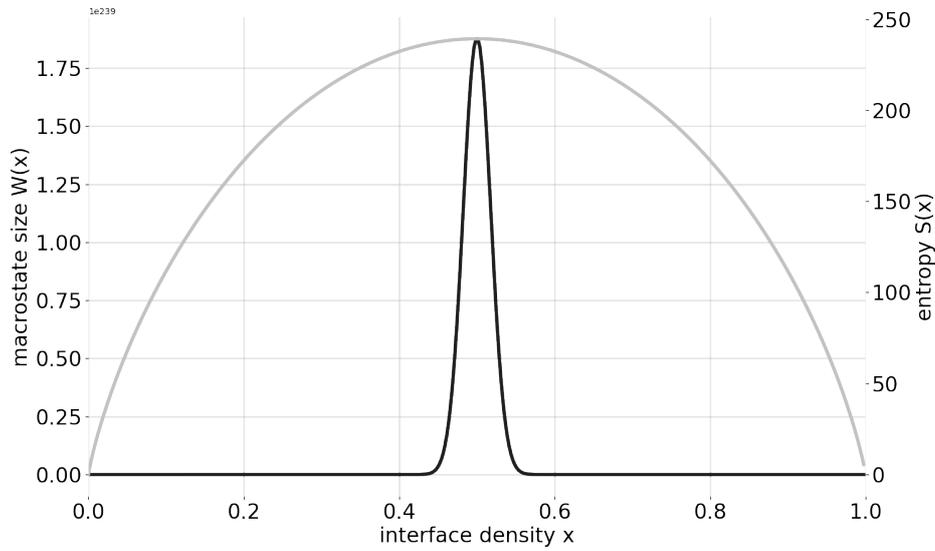


FIGURE 3.3 Nombre de micro-états $\tilde{W}(x)$ (en noir) et entropie $\tilde{S}(x)$ (en gris) en fonction de la densité d'interface x pour $\kappa = 8$, $N = 10^2$. On constate que le maximum de l'entropie correspond aux états de densité d'interface $x = 1/2$. D'après le principe de maximisation de l'entropie, l'équilibre thermodynamique devrait donc correspondre à ce macro-état.

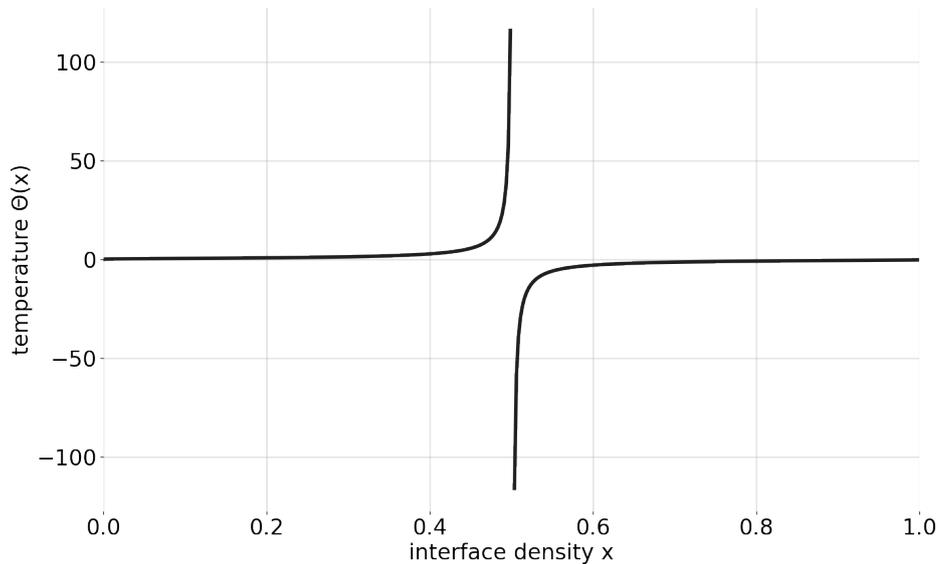


FIGURE 3.4 Température microcanonique $\Theta(x)$ du modèle de Schelling pour $\kappa = 8$, $N = 10^2$. L'entropie est une fonction strictement décroissante de x sur l'intervalle $]1/2, 1]$, ce qui y entraîne $\Theta(x) < 0$.

Cette température étant directement issue des caractéristiques de la dynamique du système, elle ne peut être utilisée comme un équivalent à la température canonique, imposée par l'extérieur. La validité de l'approche canonique est donc hautement discutable, et l'analogie entre le modèle de Schelling et des modèles de spins doit être traitée avec prudence, du moins tant qu'il est question d'étudier la dynamique réelle, contrainte à visiter des états d'équilibres métastables, et non un équilibre théorique inatteignable par le système du fait des contraintes cinétiques qui pèsent sur lui. De ce point de vue, on peut rapprocher ce modèle de ceux décrits dans [57], en particulier celui de G. Fredrickson et H. Andersen [66, 67], qui visent à retrouver la transition vitreuse observée expérimentalement dans des modèles de spins simples comme celui d'Ising à partir de contraintes cinétiques ajoutées à ces modèles. La réductibilité de la chaîne de Markov y est discutée, certains modèles étant robustes à des variantes canoniques mais pas tous. Le modèle de Schelling, où le seuil de la tolérance introduit une discontinuité de l'utilité, semble se trouver dans cette seconde catégorie.

3.3 | Un système fermé ?

Un des enjeux de cette thèse est de calculer de manière analytique les valeurs critiques de la tolérance pour lesquelles on observe une transition de la densité d’interface des états stationnaires du modèle de Schelling. Dans le chapitre **2**, on a ainsi formalisé à l’aide du formalisme markovien une définition de la tolérance T_f séparant des états intégrés et des états ségrégués. Cette méthode, qui s’affranchit de la géométrie du système, n’explique pas la seconde transition observée à T_c .

Dans cette section, on commence par présenter les résultats de travaux de plusieurs groupes de recherche s’intéressant tous à l’étude de systèmes à grand voisinage. La plupart des résultats nécessite en effet que κ soit “grand”, ce terme étant à définir au cas par cas, parfois supérieur à quelques unités, parfois à une fonction de la tolérance, parfois d’ordre $\log(N)$... On s’intéresse dans un premier temps à la variante unidimensionnelle puis à son extension, qui semble nécessiter de se placer dans un nouvel ensemble statistique.

3.3.1 Grands voisinages unidimensionnels

Soit deux groupes distincts d’agents \oplus et \ominus équitablement répartis sur les N sites d’un anneau unidimensionnel. Chaque agent est muni d’un voisinage $\partial i(w)$ de rayon w et contenant donc $\kappa = 2w$ agents. Un agent est considéré satisfait lorsque son indice d’homogénéité locale m_i est supérieur à $(1 - 2T)\kappa$. Son utilité peut donc s’écrire $u_i(\boldsymbol{\sigma}) = \mathbb{1}_{[(1-2T)\kappa, \kappa]}(m_i)$ et dans le cadre d’une dynamique de type Kawasaki, à chaque instant, un agent insatisfait échange, s’il le peut, sa position avec un autre agent insatisfait.

Pour une configuration donnée du système, on peut considérer les paires homogènes de voisins comme reliées par une arête, tandis que les paires hétérogènes sont déconnectées. On obtient alors un sous-graphe séparé en composantes connexes qu’on appellera *noyaux homogènes*, notés Ω et dont la taille Ω peut varier, dans le cadre d’une dynamique par échanges, entre 1 et $N/2$.

On note $\mathbb{P}_T[\Omega = \ell]$ la probabilité que dans l’état final d’une dynamique de Schelling de tolérance T , un agent quelconque, sélectionné uniformément, appartienne à un noyau homogène Ω de longueur ℓ .

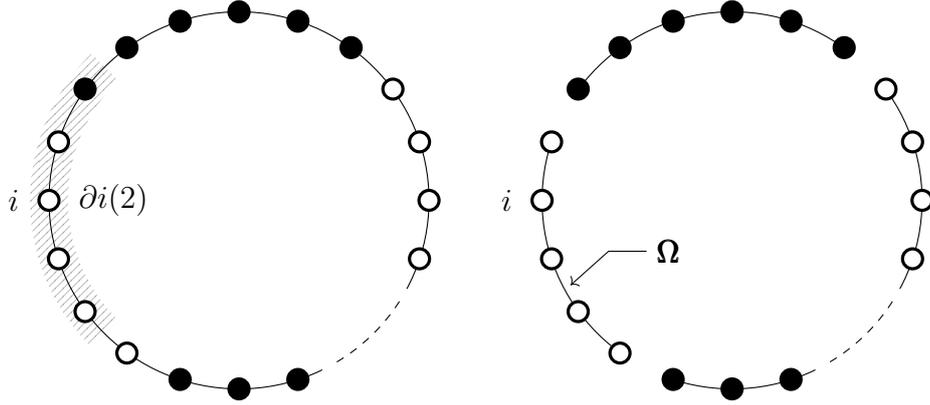


FIGURE 3.5 Représentation du voisinage de l’agent i (à gauche) et du noyau homogène Ω contenant l’agent i (à droite) dans le cas unidimensionnel.

Théorème [26] *Lorsque $T = 1/2$, pour tout w et $N \gg w$, le système atteint un équilibre ne permettant plus aucun échange. Dans cette configuration finale,*

$$\mathbb{E}_{1/2}[\Omega] = \mathcal{O}(w^2) \quad (3.43)$$

De plus, il existe $\eta > 0$ tel que pour tout $\ell > 0$,

$$\mathbb{P}_{1/2}[\Omega \geq \ell] < e^{-\eta\ell/w^2} \quad (3.44)$$

Ce premier théorème indique que lorsque $T = 1/2$, les noyaux à l’équilibre sont quadratiques en la taille du voisinage $w = \kappa/2$, et plus généralement que la distribution des tailles de noyaux est majorée par la distribution exponentielle de paramètre η/w^2 . La ségrégation complète à l’échelle du système n’est pas atteinte, au sens où on n’observe pas deux noyaux bien séparés de taille $N/2$, mais on constate tout de même l’émergence de “bulles” homogènes. Le théorème suivant complète le résultat et introduit la première tolérance critique T_c dont la valeur numérique peut correspondre à la deuxième transition observée dans le diagramme de phase du modèle de Schelling.

Théorème [16] *Soit $T_c \simeq 0.647$ l’unique solution de $2(2T-1)^{2T-1} = (2T)^{2T}$.*

(i) $\mathbf{T} \leq 1/2$. *Pour tout $w > w_T = T/(1-2T)$,*

$$\lim_{N \rightarrow \infty} \mathbb{P}_T[\Omega = N/2] = 1 \quad (3.45)$$

Si $w \leq w_T$, le théorème précédent s’applique.

(ii) $1/2 < \mathbf{T} < \mathbf{T}_c$. *Soit $\varepsilon > 0$. Pour tout w et $N \gg w$, il existe $\eta_\varepsilon > 0$ tel que*

$$\mathbb{P}_T[\Omega = e^{\eta_\varepsilon w}] > 1 - \varepsilon \quad (3.46)$$

(iii) $\mathbf{T} > \mathbf{T}_c$. Soit $\varepsilon > 0$. Pour tout w , il existe $\ell_\varepsilon > 0$ tel que

$$\mathbb{P}_T[\Omega = \ell_\varepsilon] < \varepsilon \quad (3.47)$$

On peut alors représenter les équilibres atteints dans l'espace (κ, T) , en figure 3.6 pour un anneau unidimensionnel. Pour des valeurs croissantes de T à partir de $T = 1/2$, l'équilibre du système est donc un état de ségrégation en noyaux de taille exponentielle puis un état intégré, semblable à la configuration initiale, après la transition $T = T_c$. Selon la valeur de κ , lorsque $T \leq 1/2$, le système peut présenter un état de ségrégation totale ou une agrégation locale en noyaux de taille polynomiale.

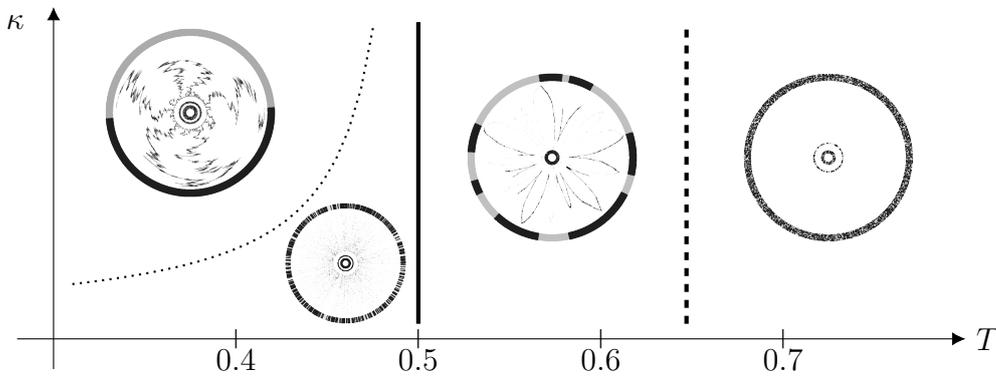


FIGURE 3.6 Diagramme de phase (κ, T) pour un anneau unidimensionnel, où sont représentés les états d'équilibre obtenus par simulation de G. Barmpalias et al. [16]. Le trait plein vertical à $T = 1/2$ sépare les ségrégations polynomiale et exponentielle, le trait pointillé vertical à $T = T_c$ la ségrégation exponentielle et l'intégration, et la courbe pointillée distingue ségrégation polynomiale et ségrégation totale.

Lorsque l'une des deux populations est minoritaire ($p < 1/2$), tolérance et parité influent conjointement sur l'équilibre. D'après le théorème suivant, certains couples (p, T) produisent presque sûrement des équilibres complètement ségrégués (deux noyaux homogènes correspondant aux deux populations). Pour d'autres couples, le système est *quasi-statique* un agent tiré aléatoirement n'a presque sûrement participé à aucun échange : le système est donc toujours dans sa configuration aléatoire initiale.

Théorème [17] *On note Ω_{min} et Ω_{maj} les tailles de noyaux homogènes du groupe minoritaire et majoritaire, respectivement.*

(i) $\mathbf{T} < \mathbf{1}/\mathbf{2}$. Soit $\varepsilon > 0$, pour tout $N \gg w \gg 1/\varepsilon$,

$$\mathbb{P}_T[\Omega_{min} = pN] = \mathbb{P}_T[\Omega_{maj} = (1-p)N] > 1 - \varepsilon \quad (3.48)$$

(ii) $\mathbf{T} > 1/2$. Le système est quasi-statique lorsque $N \rightarrow \infty$ ssi

$$p < \frac{T + T_c}{2(2T_c - 1)} \quad (3.49)$$

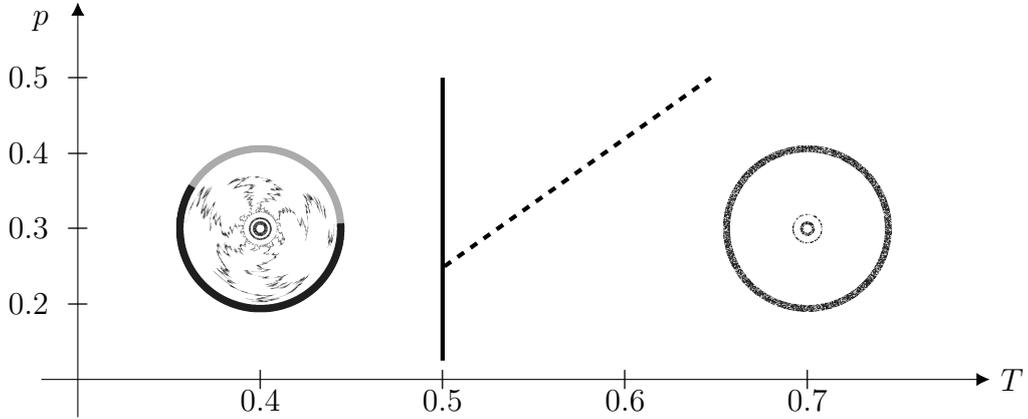


FIGURE 3.7 Diagramme de phase (p, T) pour un anneau unidimensionnel. Le trait plein vertical à $T = 1/2$ marque la limite supérieure de la ségrégation totale, le trait pointillé la limite inférieure de l'intégration. S'ils existent, les équilibres dans la partie intermédiaire sont inconnus.

Enfin, un résultat similaire est donné pour une autre dynamique appelée *dynamique de Glauber*, dans laquelle un agent insatisfait change de groupe à condition que la substitution le satisfasse. Comme on l'a déjà mentionné⁵⁴, pour cette dynamique, le problème est quasi-symétrique par rapport à $T = 1/2$ dans le cas d'un processus sélectif. Tous les résultats suivants sont restreints à $T > 1/2$ qui est le domaine que l'on cherche à étudier.

Théorème [14] *On définit T_c de la même manière que précédemment.*

(i) $1/2 < \mathbf{T} < \mathbf{T}_c$. Soit $\varepsilon > 0$. Pour tout $N \gg w \gg 1/\varepsilon$, la probabilité qu'un agent tiré aléatoirement dans le système à l'état final appartienne à la population initialement majoritaire est

$$\mathbb{P}_T [i \in \Omega_{maj}] > 1 - \varepsilon \quad (3.50)$$

(ii) $\mathbf{T} > \mathbf{T}_c$. Le système est quasi statique lorsque $N \rightarrow \infty$.

⁵⁴Se référer au paragraphe *Construire la matrice de transition* en section 2.1.2.

Cette variante de Glauber appelle un commentaire quant à l'ensemble thermodynamique dans lequel cette étude est menée. En effet, le fait qu'un agent change de groupe est équivalent à sa sortie du système et à son remplacement instantané par un agent de l'autre groupe extérieur au système. Cette reformulation place alors l'étude dans l'ensemble grand-canonique : le système est ouvert. Cette discussion est l'objet de la première partie de la section suivante.

3.3.2 Extension à deux dimensions

Les résultats obtenus par C. Brandt et al. [26] et G. Barmpalias et al. [16, 17] dans le cadre d'une dynamique par échanges (de Kawasaki) dans un système unidimensionnel à grand voisinage n'ont pas d'équivalent bidimensionnel.

En effet, il semblerait que les contraintes exercées par des échanges pouvant avoir lieu à toute échelle, entre proches voisins ou d'un bout à l'autre de l'anneau représentant la ville, se traduisent mal sur une grille plane. Les nombreuses corrélations engendrées, la nécessité de prendre en compte une multitude de "fronts" homogènes a entraîné plusieurs groupes de chercheurs et chercheuses à se concentrer sur la variante de Glauber présentée ci-dessus.

Pourtant, la dynamique de Kawasaki permet d'établir des lois de conservation du système : la parité et la densité sont constantes. Le système est fermé en termes de particules, ce qui est à rapprocher de la discussion sur la température (système isolé ou non). À l'inverse, un système dans une dynamique de Glauber n'est plus fermé en parité, le nombre total de particules est constant mais leur répartition varie, ce qui le place dans un autre ensemble thermodynamique : l'ensemble grand-canonique⁵⁵. En reprenant la conclusion de la section précédente, une telle modification des règles du système pourrait avoir pour conséquence d'en modifier les résultats. Une résolution dans cette variante ne décrirait donc pas nécessairement un système de Schelling "classique".

Pourtant, par comparaison avec l'introduction d'une agitation thermique, on pourrait argumenter en faveur de cette modification pour les raisons suivantes. D'une part, changer la composition du système n'influe pas sur l'inflexibilité du seuil de tolérance. Les échanges ont toujours lieu dans le sens d'une hausse de l'utilité : selon la valeur de T , cela entraînera ou non une di-

⁵⁵Formellement, il faudrait le définir comme un ensemble "pseudo-grand-canonique". En physique, aucun transfert de particule ne peut se faire sans transfert d'énergie, au moins cinétique : un système ouvert ne peut pas être isolé.

mension de la densité d’interface, il y a conservation de la réductibilité (ou de l’irréductibilité, le cas échéant) de la chaîne de Markov. D’autre part, on peut imaginer que le système possède une copie en négatif. Un agent insatisfait échangerait alors sa place avec son équivalent de l’autre groupe dans l’autre système, lui aussi insatisfait par symétrie. Un état de ségrégation devient alors un état où le système est presque entièrement rempli des agents d’un groupe, tandis que sa copie inversée contient tous les autres. La réunion de ces deux versions est alors fermée, on se retrouve dans l’ensemble microcanonique, et on s’évite ainsi l’étude des corrélations internes au système.

Une dernière critique plus générale de ces travaux est celle de la pertinence de considérer des voisinages “infinis”. Du point de vue de la modélisation des mobilités dans la ville, il semble en effet assez peu raisonnable de consulter ses $\log N$ plus proches voisins avant de décider de déménager. La question est heureusement adressée par les auteurs eux-mêmes :

On pourrait objecter que des voisinages de taille arbitrairement grande ne sont pas dans l’esprit des modèles initiaux de ségrégation raciale. Cependant, il ne faut voir ce choix que comme un outil mathématique permettant de déduire le résultat avec une précision arbitrairement élevée. Dans les simulations de cette classe de modèles, ce résultat apparaît dès les premières valeurs de κ ⁵⁶.

À deux dimensions, la définition du système reste la même. Le système est à présent une grille carrée contenant N sites, de voisinage $\partial i_{\square}(w)$ de rayon w et contenant donc $\kappa = 4w(w + 1)$ agents.

Tout agent tiré aléatoirement dans le système à l’état final appartient à un ensemble fini de *voisinages homogènes* de rayon r , pas nécessairement centrés sur cet agent. On appelle Δ le plus grand d’entre eux, contenant $\Delta = 4R(R + 1)$ éléments, où R est le “rayon” de cette région homogène. La définition ne repose donc plus sur les composantes connexes du sous-graphe extrait. Δ est inclus dans l’équivalent bidimensionnel de Ω . On notera en

⁵⁶*It might be objected, at this point, that neighbourhoods of arbitrarily large size are not in the spirit of models of racial segregation. However, this should be seen as a mathematical device allowing the outcome to be predicted with arbitrarily high precision. In simulations of all models in this family, relatively small values of w are generally sufficient for a clear outcome. [15], p. 1462 [traduction libre].*

gras tous les ensembles correspondant à un voisinage de taille quelconque dans la suite.

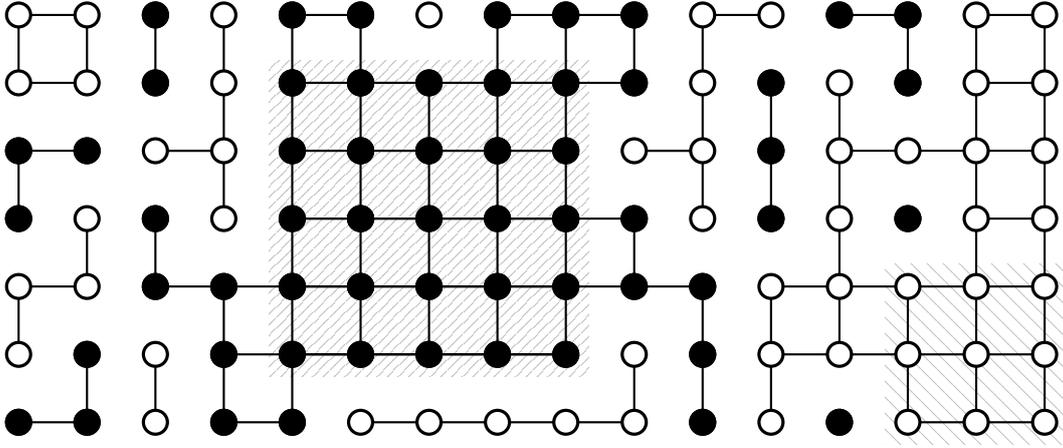


FIGURE 3.8 Représentation des deux plus grands voisinages homogènes Δ de taille supérieure à 1. Tous les autres agents du système sont compris dans un voisinage constitué uniquement d’eux-mêmes. On note que Δ est toujours inclus dans sa composante connexe.

Chronologiquement, les deux premiers résultats donnent un encadrement des valeurs de T dans la phase ségrégée : en 2016, G. Barmpalias et al. [15] démontrent que le système est quasi-statique presque partout lorsque $T > 3/4$, tandis qu’en 2017, N. Immorlica et al. [88] démontrent l’existence de la ségrégation exponentielle lorsque T prend des valeurs proches de $1/2$.

Théorème [88] *On suppose que $T = (1 + \varepsilon)/2$ avec $\varepsilon > 0$. Pour tout $N \gg w \gg 1/\varepsilon$, il existe $k > 0$ tel que*

$$\mathbb{E}_{\frac{1}{2}(1+\varepsilon)}[R] \geq e^{k\varepsilon^2 w^2} \quad (3.51)$$

Il reste à affiner la valeur de tolérance critique T_c . Les paragraphes suivants visent à détailler la méthode par laquelle H. Omidvar et M. Franceschetti [124, 125, 127] parviennent à cet objectif. Ces auteurs s’intéressent plus spécifiquement à la dynamique de constitution d’un voisinage homogène : ils déterminent successivement des intervalles de valeurs de T pour lesquelles une zone favorable existe, pour lesquelles une racine homogène émerge dans cette zone favorable, pour lesquelles la “contagion” par un autre voisinage est impossible, et enfin pour lesquelles la zone homogène créée est macroscopique. Il suffit ensuite de considérer toutes ces conditions réunies pour déterminer la plus grande valeur de T permettant la

ségrégation exponentielle. L'article original est particulièrement technique, la présentation qui en est faite ici permet d'en résumer l'essentiel.

On suppose sans perte de généralité qu'à l'équilibre, le voisinage homogène Δ , s'il existe, est constitué d'agents du groupe $\oplus : \forall i \in \Delta, s_i = 1$.

Un domaine propice à la formation d'un voisinage homogène est un ensemble de sites

$$\mathbf{D}(r) = \left\{ j \in \partial i_{\square}(r) \mid \forall j, h_j \geq (1 - 2T)\kappa \right\} \quad (3.52)$$

Tout agent j tel que $s_j = -1$ initialement présent dans $\mathbf{D}(r)$ sera alors satisfait s'il change de groupe, indépendamment de son état initial de satisfaction. En effet, après un tel changement, $s_j = 1$, donc $m_j = s_j h_j \geq (1 - 2T)\kappa$. D'après les propriétés de la distribution de Bernoulli, utilisée pour construire la configuration initiale, pour $\kappa \gg 1$, on obtient⁵⁷ :

$$\mathbb{P}_T[\mathbf{D}(r = e^{\eta\kappa})] \geq 1 - \frac{5}{\sqrt{\kappa}} \quad \text{pour } T < 1 - \frac{1}{2\kappa} \quad (3.53)$$

pour $\eta = ((1 - s(T)) \ln \sqrt{2})$, où s est l'entropie binaire

$$s(t) = -t \log_2(t) - (1 - t) \log_2(1 - t) \quad (3.54)$$

Pour un voisinage de Moore ($\kappa = 8$, ne respectant donc pas la condition préalable), on trouverait la borne $T < 3/4$.

Au sein de $\mathbf{D}(r)$ peut se former⁵⁸ une "racine" homogène $\mathbf{R}(\varepsilon\kappa)$ avec probabilité

$$\mathbb{P}_T[\mathbf{R}(\varepsilon\kappa)] \geq \kappa^{-2} e^{2\eta(1-\varepsilon^2)\kappa} \quad (3.55)$$

Cette racine s'étend⁵⁹ jusqu'à former un "anneau" homogène $\mathbf{A}(R)$ avec probabilité telle que

$$\lim_{\kappa \rightarrow \infty} \mathbb{P}_T[\mathbf{A}(R = 2\kappa^{-2}e^{\eta\kappa})] = \mathbb{1}_{[1/2, T_c]}(T) \quad (3.56)$$

où $\tau_c = 1 - T_c$ est l'unique solution de

$$3/4 \times [1 - s(3/4\tau)] = [1 - s(\tau)] \quad (3.57)$$

Dans le domaine, s'il existe, cet ensemble d'événements a lieu presque sûrement⁶⁰ avant que la propriété (3.52) ne soit plus vérifiée, ce qui permet de

⁵⁷[126], lemme 21 (annexe A)

⁵⁸*ibid.*, lemme 6

⁵⁹*ibid.*, lemme 8

⁶⁰*ibid.*, lemme 7

considérer cet anneau comme un “pare-feu”, pour reprendre la terminologie de l’article, isolant cette partie du système de rayon R de toute influence extérieure. Ainsi, tout agent qui y est contenu fera indéfiniment partie du groupe \oplus à partir d’un certain point, sans possibilité de contagion par un noyau voisin de l’autre groupe.

Ainsi, on trouve

$$\mathbb{P}_T[R] \geq \mathbb{P}_T[\mathbf{D}(r = e^{\eta\kappa})] P_T[\mathbf{R}(\varepsilon\kappa)] \mathbb{P}_T[\mathbf{A}(R = 2\kappa^{-2}e^{\eta\kappa})] \quad (3.58)$$

et on obtient le théorème suivant.

Théorème [124] *Soit $T_c = 1 - \tau_c$, où τ_c est l’unique solution de*

$$3/4 \times [1 - s(3/4\tau)] = [1 - s(\tau)] \quad (3.59)$$

Lorsque $1/2 < \mathbf{T} < \mathbf{T}_c$ et $\kappa = \mathcal{O}(\ln N)$, il existe $0 < \alpha < \beta$ tels que

$$e^{\alpha\kappa} \leq \mathbb{E}_T[\Delta] \leq e^{\beta\kappa} \quad (3.60)$$

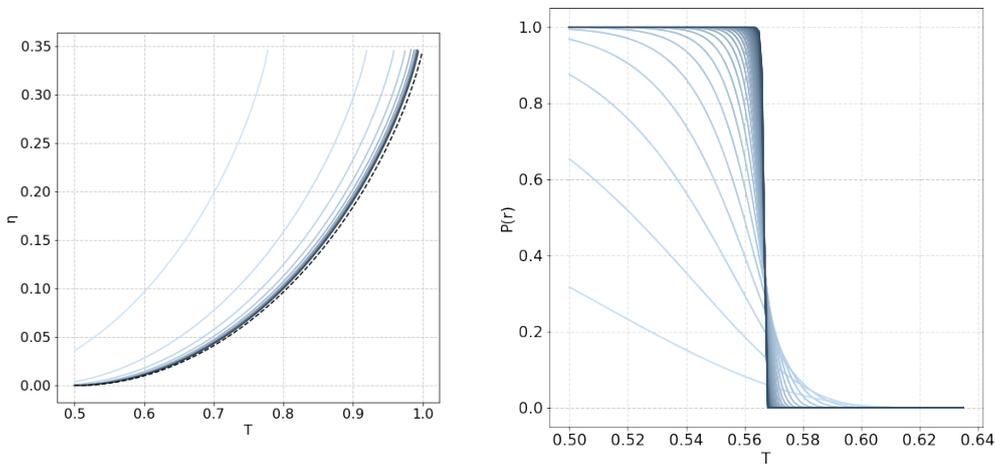


FIGURE 3.9 Détermination du seuil de tolérance au-delà duquel le noyau homogène ne se forme pas (à gauche, échelle caractéristique η propice à la formation d’un noyau homogène en fonction de T) et ne grandit pas (à droite, probabilité $\mathbb{P}_T[\mathbf{A}(R)]$). Simulations effectuées pour $\kappa = 4w(w + 1)$ avec $w \in \{1 \dots 10\}$ (en bleu) jusqu’à $\kappa \rightarrow \infty$ (en noir).

Par la suite [125, 127], les auteurs sont parvenus à démontrer ce résultat pour la taille du domaine Δ elle-même, en s'affranchissant de l'espérance, mais le principe de la preuve reste similaire (bien que plus complexe).

Théorème [125] Soit $T'_c = \frac{18041 - 928\sqrt{30}}{25298} \simeq 0.512$.

Lorsque $1/2 < \mathbf{T} < \mathbf{T}'_c$ et $\kappa = \mathcal{O}(\ln N)$, il existe $0 < \alpha < \beta$ tels que

$$\lim_{\kappa \rightarrow \infty} \mathbb{P}_T [e^{\alpha\kappa} \leq \Delta \leq e^{\beta\kappa}] = 1 \quad (3.61)$$

Un dernier résultat à noter est un encadrement semblable à (3.60) pour des voisinages quasi-homogènes, lui aussi en espérance puis pour Δ . Ces voisinages quasi-homogènes sont définis pour $\varepsilon > 0$ comme des ensembles d'agents adjacents contenant Δ agents d'un groupe et $\delta < \Delta e^{-\kappa^\varepsilon}$ agents de l'autre groupe.

Théorème [124] Soit $T''_c = 21/32 \simeq 0.656$.

Lorsque $\mathbf{T}_c < \mathbf{T} < \mathbf{T}''_c$ et $\kappa = \mathcal{O}(\ln N)$, il existe $0 < \alpha < \beta$ tels que

$$e^{\alpha\kappa} \leq \mathbb{E}_T [\Delta'] \leq e^{\beta\kappa} \quad (3.62)$$

Par cette méthode de croissance stabilisée d'un voisinage homogène, les auteurs parviennent à encadrer de manière de plus en plus précise la valeur critique T_c de la tolérance en deçà de laquelle la ségrégation peut émerger macroscopiquement. Les différentes valeurs déterminées dépendent de la définition de la ségrégation choisie : voisinages homogènes ou quasi-homogènes, de taille exponentielle ou dont l'espérance est exponentielle...

Par ailleurs, les démonstrations sont très techniques, même si leur principe peut être résumé par une construction qui reste assez intuitive. Il serait intéressant de chercher une autre manière plus simple d'approcher cette tolérance critique T_c , c'est ce que nous proposons dans la section suivante.

3.4 | Un modèle de nucléation

3.4.1 Noyaux homogènes, cœur et interfaces

On appelle *noyau homogène* Ω toute composante connexe du sous-graphe obtenu en déconnectant les paires hétérogènes. Un noyau contient Ω^2 sites. Il ne faut pas confondre cette définition du noyau homogène avec celle du *voisinage homogène* [88, 124, 125, 127] défini comme le plus grand voisinage compris dans une composante connexe du sous-graphe obtenu en déconnectant les paires hétérogènes. Ce voisinage est inclus dans l'unique noyau homogène partageant au moins un site en commun avec lui.

L'*interface interne* du noyau est l'ensemble des sites de Ω dont le voisinage (au sens du graphe sous-jacent) contient au moins un site n'appartenant pas à Ω . Un noyau de taille Ω^2 contient $\partial\Omega$ sites interfaciaux. On émet l'hypothèse que l'interface interne est polynomiale en Ω , plus précisément qu'il existe a, b, c , dépendant uniquement de la forme du noyau, tels que

$$\partial\Omega = a\Omega^2 + b\Omega + c \quad (3.63)$$

avec $0 \leq a \leq 1$ et avec la contrainte $\partial\Omega \leq \Omega^2$. Cette hypothèse découle des simulations effectuées sur un modèle de Schelling, dont certains résultats sont représentés en figure **3.10**, et de l'étude des exemples suivants.

Dans le cas d'un noyau *linéique*, couvrant les sites présents le long d'une ligne brisée, alors tous ces sites se trouvent sur l'interface :

$$\partial\Omega \simeq \Omega^2 \quad (3.64)$$

Si ce noyau double d'épaisseur, certains sites se trouvent alors au cœur du noyau, coupés de l'interface. Ce cœur ainsi constitué contient environ la moitié des sites⁶¹ : on trouve alors

$$\partial\Omega \simeq \frac{1}{2}\Omega^2 \quad (3.65)$$

Si on laisse ce cœur grandir jusqu'à occuper presque tout le noyau, il ne reste plus sur l'interface que la différence entre le nombre total de sites Ω^2 et celui dans le cœur, d'ordre $(\Omega - 2\varepsilon)^2$, où ε est un entier caractérisant la largeur de l'interface. On trouve enfin

$$\partial\Omega \simeq 4\varepsilon(\Omega - \varepsilon) \quad (3.66)$$

⁶¹Aux effets de bord près, dépendant du voisinage choisi.

Si on poursuit ce raisonnement, la forme de moindre interface pour une taille Ω^2 devrait être un disque. Dans ce cas on peut s'attendre à trouver, lorsque $\Omega \rightarrow \infty$,

$$\partial\Omega \sim 2\sqrt{\pi}\Omega \quad (3.67)$$

Revenons au noyau carré. Si à partir du cœur se forment des *dendrites*, c'est-à-dire une forme d'arborescence dont chaque "branche" est linéique, et que celles-ci grandissent jusqu'à contenir environ la moitié des agents du noyau, on obtient une interface séparée en un terme linéaire issu du carré original et un terme quadratique provenant de la partie linéique :

$$\partial\Omega \simeq \frac{1}{2}\Omega^2 + 2\sqrt{2}\Omega \quad (3.68)$$

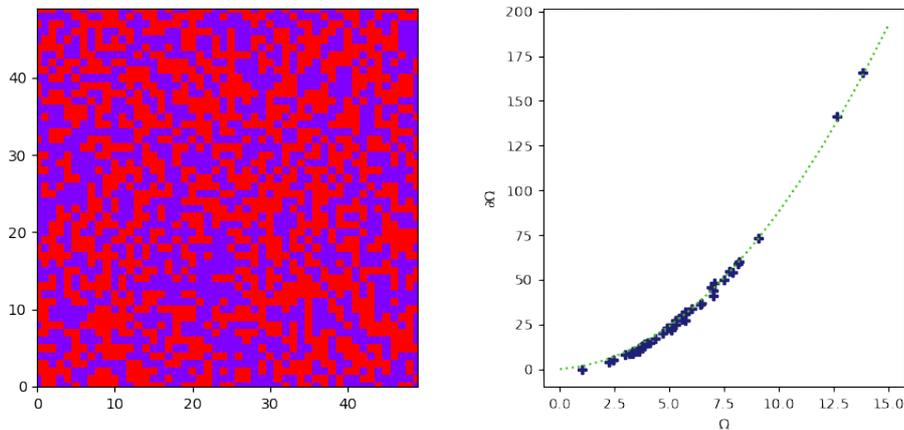


FIGURE 3.10 Dans l'état initial (représenté à gauche), les noyaux sont principalement dendritiques et on trouve $\partial\Omega \simeq a\Omega^2 + b\Omega$. L'ajustement donne $a \simeq b \simeq 0.8$.

On peut en conclure que le choix de (a, b) détermine de façon assez complète la typologie des noyaux rencontrés dans la dynamique de Schelling. Un noyau On peut résumer ces différentes formes en figure **3.11**.

Inversement, l'interface externe est l'ensemble des sites du système n'appartenant pas à Ω et dont le voisinage contient au moins un site appartenant à Ω . Dans le cadre d'un voisinage de Moore, on peut montrer⁶² que le nombre de sites de l'interface extérieure est borné :

$$\partial\Omega + 8 \leq \partial\Omega_{\text{ext}} \leq 8\partial\Omega \quad (3.69)$$

On va donc considérer en général que le nombre de sites de cette interface extérieure est $\partial\Omega_{\text{ext}} = \ell\partial\Omega + k$, où $\ell \geq 1$ est appelé le coefficient de rugosité

⁶²Se référer à l'annexe **F**.

de l'interface, et k est une constante dépendant du voisinage et prenant en compte les effets de bords mentionnés plus haut. Pour un noyau à cœur (carré ou dendritique), $\ell \simeq 1$, en revanche pour un noyau linéique, $\ell \simeq 2$. Il s'agit en première approximation du ratio du nombre de sites de l'interface externe par le nombre de sites de l'interface interne.

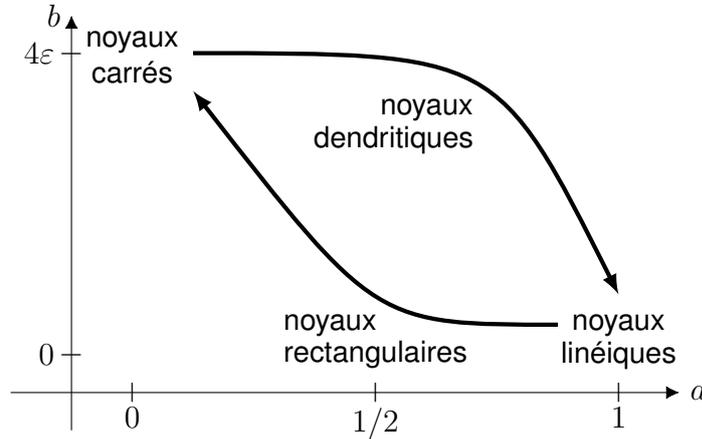


FIGURE 3.11 Représentation schématique des différentes typologies de noyau en fonction de a et b . Sont représentés les deux “chemins” empruntés lors de leur description.

Le principe de la théorie de la nucléation [21, 31] est le suivant : rassembler des agents du même groupe dans le cœur du noyau augmente leur satisfaction, et donc abaisse leur énergie. À l'inverse, créer une interface entre le noyau et le milieu extérieur, supposé aléatoire, est une opération coûteuse. Il peut alors exister une taille caractéristique et une forme “stabilisatrices”. Si le modèle de Schelling avait le même comportement que certains systèmes physiques, on verrait sûrement émerger des “bulles de ségrégation” au sein de la phase aléatoire. On commence par écrire l'énergie moyenne du noyau sous la forme

$$\langle E(\Omega, \ell) \rangle = -e (\Omega^2 - \partial\Omega) + \gamma_{\text{int}} \partial\Omega + \gamma_{\text{ext}} (\ell \partial\Omega + k) \quad (3.70)$$

où les coefficients γ_{int} , γ_{ext} et $-e$ représentent respectivement les énergies moyennes associées à un site de l'interface interne, de l'interface externe et du cœur du noyau.

3.4.2 Énergie moyenne du système

L'énergie moyenne par site dans un système de Schelling est donnée par

$$\langle E_i \rangle = - \left\langle \sum_{j \in \partial i} s_i s_j \right\rangle - (2T - 1) \left\langle \sum_{j \in \partial i} s_i^2 s_j^2 \right\rangle \quad (3.71)$$

où, si l'on considère que tous les sites sont occupés, $\forall i, s_i^2 = 1$.

Les agents présents dans les sites au cœur du noyau contribuent tous identiquement ($s_j = s_i$) ce qui implique

$$-e = \langle E_i \rangle_{\text{cœur}} = -\kappa - (2T - 1)\kappa = -2T\kappa \quad (3.72)$$

Cette contribution est nécessairement négative quelle que soit la valeur de T , elle représente la composante de l'énergie qui stabilise le noyau.

La contribution moyenne des agents hors du noyau ou de son interface est nulle : on note $\langle s_i \rangle_{\infty} = 0$, où la notation $\langle \cdot \rangle_{\infty}$ représente une moyenne calculée au milieu de la phase aléatoire. γ_{ext} étant l'énergie moyenne d'un agent hors du noyau, on a alors

$$\gamma_{\text{ext}} = \langle E_i \rangle_{\text{ext}} = \langle E_i \rangle_{\infty} = -(2T - 1)\kappa \quad (3.73)$$

En revanche, une partie de l'aimantation d'un agent de l'interface interne est strictement positive. Parmi les κ agents de son voisinage, une proportion μ_{κ} appartient à Ω . On a donc

$$\gamma_{\text{int}} = \langle E_i \rangle_{\text{int}} = -\mu_{\kappa}\kappa - (2T - 1)\kappa = -2T\kappa + (1 - \mu_{\kappa})\kappa \quad (3.74)$$

On introduit enfin la quantité

$$\begin{aligned} \gamma &= e + \gamma_{\text{int}} + \ell\gamma_{\text{ext}} \\ &= \kappa(1 - \mu_{\kappa} + \ell - 2\ell T) \end{aligned} \quad (3.75)$$

qui permet, avec (3.63), d'obtenir l'expression suivante pour l'énergie moyenne du noyau :

$$\langle E(\Omega, \ell) \rangle = \Omega^2(a\gamma - e) + b\gamma\Omega + c(\gamma + k\gamma_{\text{ext}}) \quad (3.76)$$

L'énergie moyenne du noyau est donc une fonction affine de ℓ , de pente positive $\partial\Omega$, ce qui laisse envisager qu'un noyau "lisse" sera plus stable qu'un noyau "rugueux". C'est un premier résultat attendu, on sait que la nucléation tend à lisser les interfaces. Elle est toutefois quadratique en la taille caractéristique Ω , ce qui permet d'éventuellement séparer deux régimes, selon les valeurs de a et b .

La fonction $\Omega \mapsto \langle E(\Omega, \ell) \rangle$ atteint son extremum, s'il existe, pour

$$\Omega_*(T, \ell) = \frac{b\gamma}{2(e - a\gamma)} = \frac{b\ell(1 + a\ell)(T_c - T)}{T - \frac{a}{1+a\ell}T_c} \quad (3.77)$$

où la tolérance critique T_c s'exprime en fonction de μ_κ et ℓ sous la forme

$$T_c = \frac{1 + \ell - \mu_\kappa}{2\ell} \quad (3.78)$$

Il s'agit d'un maximum lorsque $a\gamma - e < 0$, c'est-à-dire

$$T > \frac{1 + \ell - \mu_\kappa}{2(1 + \ell)} \quad (3.79)$$

et ce maximum est atteint en $\Omega_* > 0$ lorsque

$$\frac{a}{1 + a\ell} T_c < T < T_c, \quad b > 0 \quad (3.80)$$

Toute cette étude repose sur une décroissance de l'énergie au cours de la dynamique. Cette décroissance a été démontrée lorsque $T > 1/2$ pour différentes variantes [70, 74], et en particulier dans le chapitre **2** pour $T \geq T_f = (\kappa - 2)/2\kappa$. On peut alors vérifier que la condition (3.79) et la minimisation de T imposée par (3.80) sont alors toujours vérifiées.

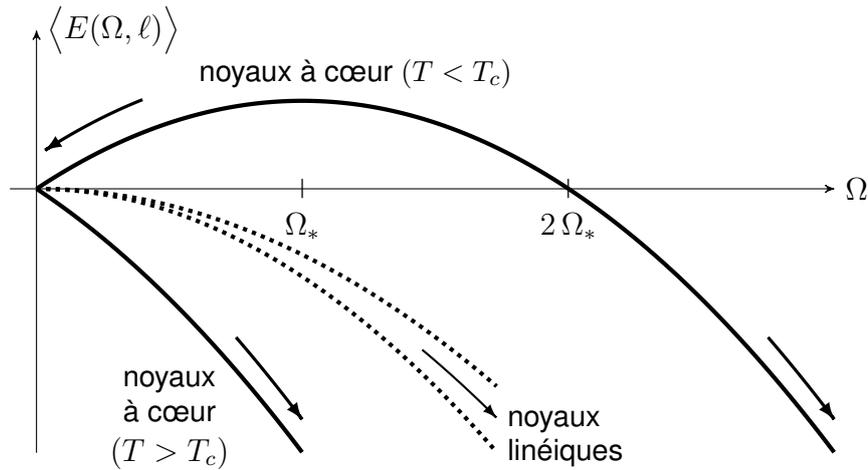


FIGURE 3.12 Représentation schématique des énergies moyennes en fonction de la taille caractéristique Ω pour diverses formes de noyaux homogènes

Lorsque $T > T_c$, les noyaux linéiques, carrés et dendritiques voient tous leur énergie moyenne diminuer si Ω augmente. Tous les noyaux tendraient donc à augmenter de taille au cours de la dynamique : cela entraîne un blocage cinétique car aucun noyau, même contenant des agents insatisfaits, n'a d'intérêt énergétique à réduire sa taille.

En revanche, lorsque $T < T_c$, les noyaux carrés et dendritiques de taille inférieure à Ω_* voient leur énergie décroître avec leur taille. Le blocage est donc levé : le système peut voir des noyaux croître et, à terme, devenir

macroscopiques : la ségrégation peut alors émerger. De plus, les noyaux dendritiques, de rugosité ℓ plus élevée, tendront à la diminuer en devenant plus convexes, donc plus carrés.

On peut alors évaluer cette valeur critique T_c : elle dépend de deux paramètres, la rugosité ℓ et la portion μ_κ de voisins dans le noyau. En supposant l'interface localement lisse (à l'échelle du voisinage κ), on peut montrer⁶³ que

$$\mu_\diamond(r) = \frac{r+2}{2(r+1)} \quad \text{et} \quad \mu_\square(w) = \frac{2w+3}{4(w+1)} \quad (3.81)$$

Ainsi, on peut évaluer pour les voisinages de von Neumann, de Moore et pour des grands voisinages que

$$T_c(\kappa=4) \simeq \frac{5}{8}, \quad T_c(\kappa=8) \simeq \frac{11}{16}, \quad \lim_{\kappa \rightarrow \infty} T_c(\kappa) = \frac{3}{4} \quad (3.82)$$

On obtient alors des valeurs de tolérance critique correspondant aux simulations numériques. La précision de ce modèle est difficile à évaluer au vu de la largeur de la gamme de tolérance correspondant à la transition ségrégation/intégration, comme on peut le voir en figure **2.12** par exemple. Par ailleurs, T_c apparaît croissante en fonction de κ , ce qui n'est pas non plus tout à fait comparable à la figure **1.5**, elle-même imprécise.

Cette méthode pour évaluer la borne supérieure T_c du domaine de stabilité de noyaux homogènes à cœur à partir du constat que le système minimise son énergie lorsque $T \geq T_f$ est plus simple que ce qui existe à ce jour dans la littérature. Par l'étude cinétique du chapitre précédent, couplée à une discussion sur la stabilité des noyaux à l'équilibre, et surtout sans introduire une température qui modifierait grandement la dynamique, nous parvenons à comprendre les mécanismes en compétition dans l'émergence et la stabilité de la ségrégation dans le modèle de Schelling.

⁶³Il suffit de dénombrer les sites appartenant à l'intersection du voisinage d'un agent et de l'interface du noyau homogène. Par construction, on a alors $\lim_{\kappa \rightarrow \infty} \mu_\kappa = 1/2$.

LE modèle de Schelling que nous venons d'étudier en détail sur trois chapitres est fondé sur une hypothèse d'*homophilie* des agents du système, sur leur préférence pour des voisinages mixtes jusqu'à un certain point. Si, comme nous l'avons exploré en section 1.1.3, de nombreuses autres caractéristiques du modèle semblent secondaires à l'émergence de la ségrégation⁶⁴, cette hypothèse semble fondamentale pour saisir la complexité du phénomène.

On peut la comparer aux données extraites de l'étude statistique menée par W. Clark et M. Fossett [37] et représentées en figure 4.0.

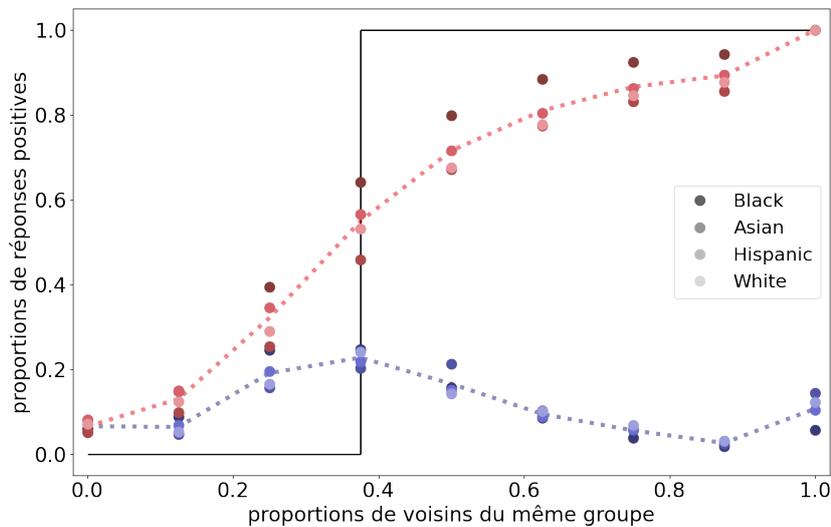


FIGURE 4.0 Préférence pour un voisinage en fonction de la proportion de voisins du même groupe, d'après les données de [37]. En ordonnée, la proportion de réponses (en bleu) et les effectifs cumulés croissants (en rouge). En pointillés, l'évolution de la moyenne pondérée des réponses. En trait plein, l'utilité en échelon proposée par Schelling, pour $T = 5/8$.

En particulier, les auteurs présentent des données recueillies auprès d'habitants de Los Angeles quant à leur voisinage "idéal". Il est demandé aux personnes participant au sondage de renseigner leur groupe ethnique et de choisir la proportion de voisins du même groupe souhaitée dans un voisinage schématique à 8 voisins, très semblable à celui utilisé dans la plupart des variantes du modèle de Schelling. On retrouve les données du tableau 2 [37] (fig. 2, p. 4111) en nuances de bleu. On peut tirer le même constat que l'article, en ajoutant la moyenne de ces préférences, en trait pointillés : la

⁶⁴Toutes les modifications proposées de la forme ou de la taille du voisinage, des modalités de sélections des sites satisfaisants, du déplacement vers ces sites, mènent à des résultats similaires.

plupart des personnes interrogées ont choisi des voisinages dans lesquels la proportion de leur propre groupe est légèrement inférieure à celle additionnée des trois autres groupes. Il y a peu de réponses dénotant une préférence pour être strictement minoritaire (environ 6.5 %) ou strictement majoritaire (environ 10.6 %). La répartition est sensiblement la même pour les quatre groupes interrogés, avec quelques variations marginales.

À partir de ces données, on peut extrapoler la réponse des participants à la question de Schelling : quelle est la proportion maximale tolérée de voisins d'un autre groupe ? Il suffit alors de représenter les proportions cumulées croissantes, en nuances de rouge. Le sondage ne porte pas sur cette question précise, il faudrait bien sûr renouveler l'étude pour un traitement quantitatif, mais on obtient tout de même une tendance, en nuances de rouge, qui permet de justifier qualitativement la forme de l'utilité choisie par Schelling (représentée ici en trait plein pour une tolérance $T = 5/8$).

La mise en relation de ce modèle théorique et des données issues d'une enquête statistique montre la complémentarité des approches. Dans le même esprit, après avoir présenté le modèle de Schelling puis démontré des résultats analytiques dans les chapitres précédents, nous adoptons une approche statistique, en tâchant dans un premier temps d'appréhender les définitions possibles de la ségrégation sociospatiale puis d'en détecter des structures plus complexes⁶⁵ et multiscalaires.

⁶⁵Insistons sur le fait que, malgré sa simplicité, le modèle de Schelling ne conduit pas à une vision simpliste des phénomènes de ségrégation. Tout au contraire, il complexifie la vision que l'on peut en avoir, car il démontre de manière élémentaire que même un bon niveau de tolérance n'est malheureusement pas une condition suffisante pour garantir la mixité sociale. Ce point, qui est le résultat majeur établi par Schelling, n'est pas toujours bien compris. Nous discutons en annexe **H** d'un exemple de critiques qui, à notre sens, passent à côté de ce qu'apporte une telle démarche de modélisation.

4. | Détecter les structures de la ségrégation urbaine

56% de la population mondiale vit en ville en 2020. La proportion d'urbains ne dépassait pas 30% il y a 70 ans, à ce rythme elle pourrait attendre 70% dans 30 ans [53]. Le tissu urbain est non seulement de plus en plus étendu, mais aussi de plus en plus dense.

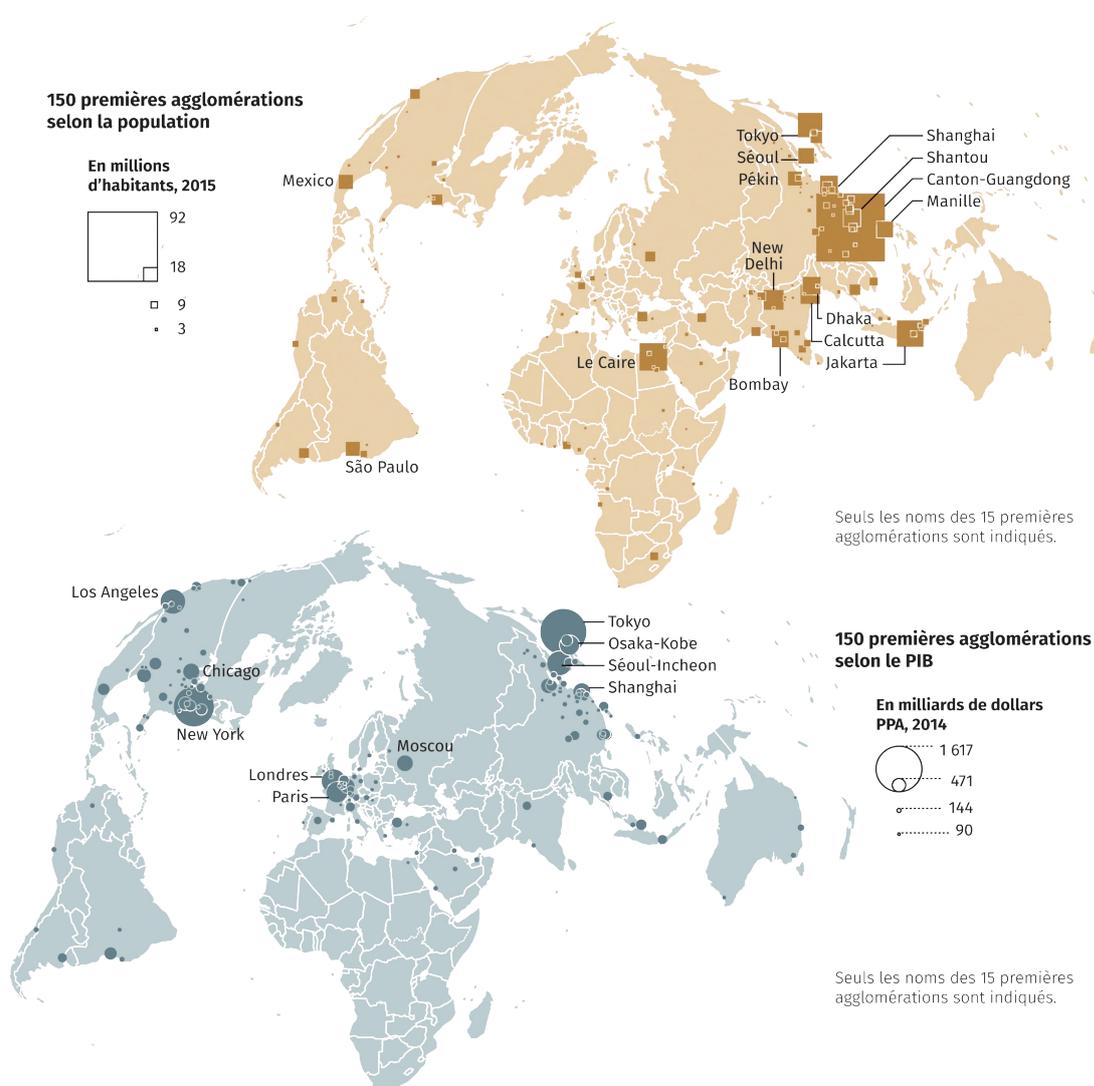


FIGURE 4.1 Les 150 premières agglomérations du monde selon la population de 2014 et le PIB de 2015 [53].

Les villes peuvent être considérées comme un parfait exemple de système complexe : les agents individuels et institutionnels y interagissent à de multiples niveaux au sein de nombreux réseaux, produisant des comportements collectifs et des motifs multiscalaires non triviaux. Parmi tous les sujets de recherche entrant dans le champ de l'étude des systèmes urbains, celui des disparités sociospatiales se distingue par l'intérêt qui lui est porté dans de nombreux domaines scientifiques, renouvelé au cours des dernières décennies.

4.1 | Décrire et mesurer la ségrégation urbaine

4.1.1 Définir la ségrégation

La première représentation qu'évoque le terme "ségrégation" est souvent la ségrégation raciale ou ethnique instaurée à l'échelle d'un État. Deux exemples historiques récents sont particulièrement illustratifs : celui des États-Unis d'Amérique et de l'Afrique du Sud au cours du XX^e siècle.

À partir de 1877, dans certains des États du sud des États-Unis, les lois Jim Crow visent à contrer une partie des droits ouverts par les treizième, quatorzième et quinzième amendements de la Constitution abolissant l'esclavage et garantissant la citoyenneté (par naissance ou naturalisation) et le droit de vote à tous les anciens esclaves. Petit à petit, État par État, des restrictions sont mises en place pour limiter l'accès des Afro-Américains aux infrastructures collectives (transports, écoles...). Puis ce sont les droits au mariage, à la propriété, à la création d'entreprise, et même jusqu'au droit de vote, qui sont entravés légalement. Les mouvements pour les droits civiques des années 1960 parviennent au terme d'une longue lutte à faire abroger la majorité de ces lois ségrégationnistes ou à les rendre inapplicables, via notamment les *Civil Right Acts* de 1964 et 1968.

Rassemblement d'anciennes colonies néerlandaises, françaises et britanniques, l'Union d'Afrique du Sud officialise à partir de 1948 les politiques suprémacistes coloniales dans le système juridique de l'Apartheid, dont le nom même signifie séparation. Distinguant quatre groupes ethniques majoritaires⁶⁶, Bantous (en fait composés d'une dizaine d'ethnies africaines), blancs (descendants de colons européens), métis et Indiens, cet ensemble

⁶⁶Ordonnés ici par proportion décroissante dans l'ensemble de la population : autour de 70 % de Bantous, 20 % de blancs, 9 % de métis et quelques pourcents d'Indiens.

de lois interdit les relations et mariages mixtes, partitionne le territoire en déplaçant ces groupes dans des régions qui leur sont attribuées, et sépare voire supprime certaines listes électorales. Le gouvernement maintient ces lois après la sortie du Commonwealth en 1961 et réprime la contestation, entraînant entre autres l'arrestation de Nelson Mandela en 1964. Après trente ans de recul sur ces politiques racistes sous la pression du Congrès National Africain, de figures politiques réformatrices et de la communauté internationale, la République d'Afrique du Sud abroge les dernières lois de séparation ethnique et spatiale en 1991.

Ces deux illustrations, auxquelles on pourrait ajouter de nombreuses autres, moins référencées dans la culture populaire ou plus contemporaines, ne sont en fait qu'un cas très particulier d'un phénomène bien plus large et général. À partir de la fin des années 1980, des géographes et des sociologues américains commencent en effet à remettre en question la discrimination raciale comme seule explication de la ségrégation urbaine. L'enjeu n'est pas de minimiser le rôle des politiques discriminatoires ni de nier les liens établis [106, 107] entre discrimination raciale, inégalités de revenu et ségrégation résidentielle, mais au contraire de mieux comprendre ce phénomène complexe, sans le réduire à une relation de causalité simpliste, afin de pouvoir le combattre plus efficacement.

Une étude intitulée *Resegregation in American Schools* [130] montre ainsi que la ségrégation raciale revient aux États-Unis sous une forme plus discrète et non officielle. Selon cette étude, le retrait progressif des mesures de "déségrégation" a précédé une tendance observée depuis la fin des années 1980, celle d'un recul de la proportion d'étudiant·e·s non-blanc·he·s dans les écoles à majorité blanche, en particulier au sud des États-Unis. Elle peut en partie être imputée à la relocalisation des familles blanches en banlieue, loin des centres urbains : c'est le *white flight*. L'étude se conclut d'ailleurs par ces phrases, qui illustrent la nécessité de mieux comprendre les mécanismes d'émergence de la ségrégation, et de développer des politiques publiques susceptibles de l'enrayer :

*Nous revenons insidieusement à un système éducatif qui n'a jamais, dans l'Histoire de la nation, permis des écoles égales et prospères. (...) Inverser la tendance d'intensification de la ségrégation et des inégalités sera difficile, mais l'accepter passivement aurait probablement un coût immense.*⁶⁷

⁶⁷ *We are floating back toward an educational pattern that has never in the nation's*

Même si toutes les pratiques foncières ou locatives discriminatoires, illégales depuis 1968, se voient réprimées par un arsenal juridique de plus en plus complet, l'existence d'autres causes à la ségrégation résidentielle la maintiendrait de fait en place.

Les préférences individuelles sortent du cadre des politiques d'équité en matière de logement et les solutions [au problème de la ségrégation] sont nécessairement à chercher en dehors du domaine législatif.⁶⁸

Par ces trois exemples historiques, on commence à appréhender la complexité de l'objet que l'on cherche à définir. Dans un souci de généralisation, on va en proposer des définitions les plus larges et les plus simples possibles, ce qui va déplacer la question de sa complexité au choix des méthodes de mesure.

La ségrégation sociospatiale est le constat d'une répartition spatiale inhomogène des groupes sociaux, et d'inégalité d'accès aux infrastructures urbaines, aux biens matériels et symboliques. D'un point de vue plus mathématique, c'est le constat d'une corrélation entre la distance spatiale dans un espace bien défini et la distance sociale. La ségrégation est également un processus. C'est pour mieux le comprendre que le modèle de Schelling présenté dans les précédents chapitres a été développé, mais l'analyser à partir de données réelles suppose un travail de sociologue ou de géographe qui dépasse largement le cadre de cette thèse. On se concentre donc sur la première définition, descriptive, de la ségrégation sociospatiale d'un espace fixe, à un instant donné.

Une région est ségréguée dans la mesure où les individus des différents groupes vivent dans des quartiers différents de cette région.⁶⁹

history produced equal and successful schools. (...) Reversing the trends of intensifying segregation and inequality will be difficult, but the costs of passively accepting them are likely to be immense. [130], p. 28 [traduction libre].

⁶⁸*Preferences are outside the purview of fair housing law and remedies, if sought, will necessarily be fundamentally different. [37], p. 4110 [traduction libre].*

⁶⁹*A region is segregated to the extent to which individuals of different groups live in different neighborhoods within the region. [143], p. 134 [traduction libre].*

Cet espace peut être matériel (ville, région...), immatériel (réseau social [110]) ou mixte (école, lieu de travail...). On s'intéresse ici exclusivement à la ségrégation sociospatiale urbaine, l'espace de référence sera donc la ville. Une ville peut être divisée à plusieurs échelles, des échelles administratives les plus larges et arbitraires (arrondissements, quartiers administratifs et îlots de recensement pour Paris par exemple) à celle quasi individuelle du lieu de résidence lui-même. Les données à cette dernière échelle sont souvent peu accessibles, quand elles existent. On devra donc composer avec celles, moins fines, dont on dispose.

La distance sociale, peut-être moins évidente à mesurer quantitativement, est pourtant assez intuitive : deux individus sont d'autant plus similaires, ou proches socialement, qu'ils partagent un grand nombre de traits communs, démographiques, sociaux ou économiques. On peut appartenir à la même classe sociale, avoir la même religion, avoir un revenu ou un niveau d'éducation proche, etc. Plusieurs indicateurs de distance sociale sont souvent liés : ségrégation résidentielle, ségrégation scolaire et ségrégation à l'emploi sont ainsi corrélées [25, 108], voire causées l'une par l'autre via les inégalités de revenus et leur influence sur le marché immobilier [103]. Peu mesurée en Europe, la ségrégation par l'origine ethnique peut également être liée aux précédentes [46, 109, 121, 148]. Enfin, il est à noter que cette façon de définir la ségrégation permet d'éviter la distinction entre les situations choisies et les situations subies, entre enclaves, ghettos et citadelles [140], dans une visée quantitative, qui pourra être étoffée par la suite.

4.1.2 Le problème de la mesure

La question d'une quantification pertinente du phénomène complexe qu'est la ségrégation résidentielle demeure ouverte malgré les nombreux travaux de géographes et de sociologues depuis des décennies [27, 39, 63, 91, 144, 155]. Les difficultés sont en effet nombreuses.

De nombreux indices de ségrégation, initialement utilisés par les géographes et qui présentent de fortes corrélations, sont en fait liés à différentes mesures de la "courbe de ségrégation" [52, 175]. On peut tracer celle-ci, en figure 4.2, à partir des données collectées dans la ville : elle représente les proportions cumulées d'un groupe en fonction des proportions cumulées de l'autre. Dans une ville non-ségrégée, cette courbe serait la diagonale d'équité-répartition, reliant l'origine au point (1,1). Tout écart avec la diagonale

représentant l'hypothèse nulle peut donc constituer une mesure pertinente de la situation de la ville réelle.

Depuis ce constat de la redondance d'une partie des indices de la littérature, plusieurs classifications ont été proposées. Celle de D. Massey et N. Denton [105] est la plus couramment utilisée et permet de séparer la définition trop large de la ségrégation sociospatiale en cinq points d'intérêts distincts :

- l'uniformité (*evenness*) de la répartition du groupe minoritaire dans chaque unité géographique de la ville. On peut citer l'indice de dissemblance [52] représenté en figure **4.2** ou l'entropie [167] ;
- l'exposition (*exposure*) ou son pendant, l'isolement, du groupe minoritaire par rapport au groupe majoritaire. Mesurant le degré d'interaction potentielle entre les deux groupes au sein d'une unité, cette catégorie s'approche de la précédente mais fait intervenir les tailles respectives des groupes ;
- la concentration du groupe minoritaire. L'indice Δ [51], par exemple, compare la taille relative du groupe minoritaire à la surface de l'unité dans laquelle cette taille est calculée ;
- la centralité du groupe minoritaire par rapport à la ville entière. Cette catégorie est intéressante dans le cadre de la comparaison de la ségrégation aux États-Unis et en France, les ghettos américains étant plutôt centraux comparés aux banlieues "sensibles" d'Île de France. La même mesure donnera alors lieu à des analyses différentes ;
- l'agrégation (*clustering*) enfin, mesure le nombre d'unités contiguës dans lesquelles le groupe (globalement) minoritaire est (localement) majoritaire.

Ces cinq classes d'indices font néanmoins face à trois limites majeures. La première est donnée par les quatre critères de D. James et K. Taeuber [91] : un indice de ségrégation doit être

- invariant par taille : la population de chaque unité peut être multipliée par un facteur commun en respectant les proportions des deux groupes sans affecter le résultat ;

⁷⁰Ce graphe est obtenu à partir des travaux de O. Duncan et B. Duncan [52].

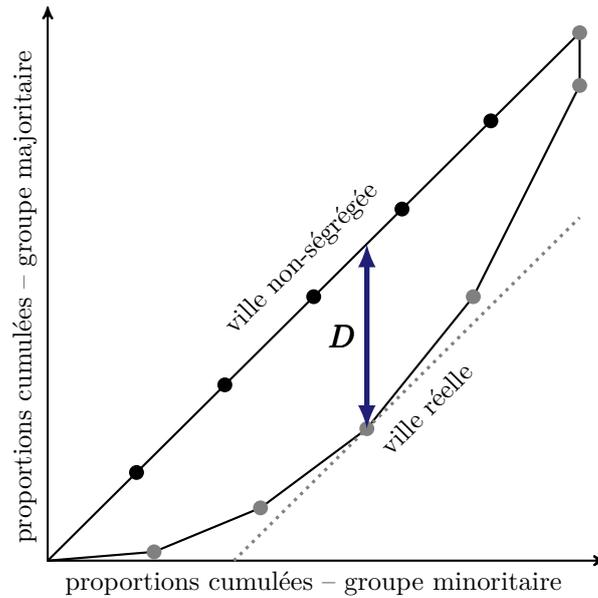


FIGURE 4.2 Construire l'indice de dissemblance à partir de la courbe de ségrégation dans une ville divisée en 6 unités géographiques⁷⁰.

- invariant par composition : le ratio de population des deux groupes peut varier uniformément dans la ville sans affecter le résultat ;
- invariant par agrégation : deux unités de même composition peuvent être agrégées sans affecter le résultat ;
- sensible au transfert d'individus issus de la minorité dans une autre unité, quelle que soit la composition initiale de celle-ci.

Ainsi, une partie des indices cités précédemment ne respecte pas ces critères. L'entropie, non linéaire, n'est pas invariante par composition. La modification proposée par le principe de transfert modifie la courbe de ségrégation, ce qui fait varier de fait une partie des indices pouvant être définis géométriquement à partir d'elle, mais pas nécessairement l'indice de dissemblance D .

La deuxième limite des indices construits à partir de la courbe de ségrégation est qu'ils ne peuvent rendre compte que de distributions binaires, contrairement à l'entropie, par exemple, qui peut être généralisée aux cas où plus de deux groupes cohabitent dans la même ville. De nouveaux indices sont alors inventés pour décrire la ségrégation multiethnique [30, 121, 122, 145], voire la ségrégation selon le niveau de revenu, qui est une variable continue [92].

Enfin, et c'est probablement la limite la plus pertinente pour la suite de ce chapitre, ces indices sont à la fois spatiaux et dépendant de l'échantillonnage choisi, ce qui entraîne deux problèmes "classiques". D'une part, tous les indicateurs, à l'exception de ceux reposant sur l'agrégation d'unités, sont incapables de distinguer une ville en échiquier d'une ville séparée en deux blocs homogènes, c'est le problème dit de l'échiquier. D'autre part, la définition des unités spatiales, forcément arbitraire, influe sur le résultat obtenu, c'est un problème appelé *modifiable areal unit problem* [64, 93, 128, 129]⁷¹. Il s'agit ainsi d'artefacts liés à l'échantillonnage et à la méconnaissance (ou à l'inexistence) des données exactes à l'échelle individuelle.

Similairement, la distinction entre uniformité et agrégation serait un artefact d'une discrétisation trop grossière du problème [143] : selon l'échelle choisie, ces deux mesures peuvent s'avérer équivalentes. Ainsi, le choix d'unités spatiales "trop petites" accole beaucoup de micro-quartiers très similaires, résultant en une valeur élevée d'indice d'agrégation, qui est comparable à une mesure de sous- ou de sur-représentation d'un groupe par rapport aux autres à plus grande échelle (indice d'uniformité). S. Reardon et D. O'Sullivan [144] proposent finalement la distinction selon deux axes, représentée en figure 4.3.

Cette aide visuelle permet entre autres de comprendre que la densité d'interface, utilisée dans les chapitres précédents comme indice de ségrégation associé au modèle de Schelling, suit vraisemblablement un axe oblique. Son nom le laisse suggérer, elle mesure l'interface entre les deux groupes, c'est donc une mesure d'exposition, de contact. Cependant, elle est aussi liée à l'espérance de l'aimantation locale, qui mesure la répartition à l'échelle d'un voisinage. C'est donc aussi une forme de mesure d'uniformité.

Ainsi, une grille séparée en deux noyaux contenant chacun la moitié de la population avec la plus petite interface possible correspondrait à une densité d'interface proche de zéro, et au quadrant en bas à gauche (isolement et agrégation maximales), tandis qu'une configuration en échiquier se trouverait, symétriquement, en haut à droite, avec une densité d'interface à $1/2$, sa valeur maximale.

⁷¹Sa version en sciences politiques est le *gerrymandering*, ou "découpage électoral partisan", à savoir la séparation d'une communauté à l'implantation géographique identifiée dans plusieurs unités administratives afin d'y sous-représenter les suffrages de ses membres.

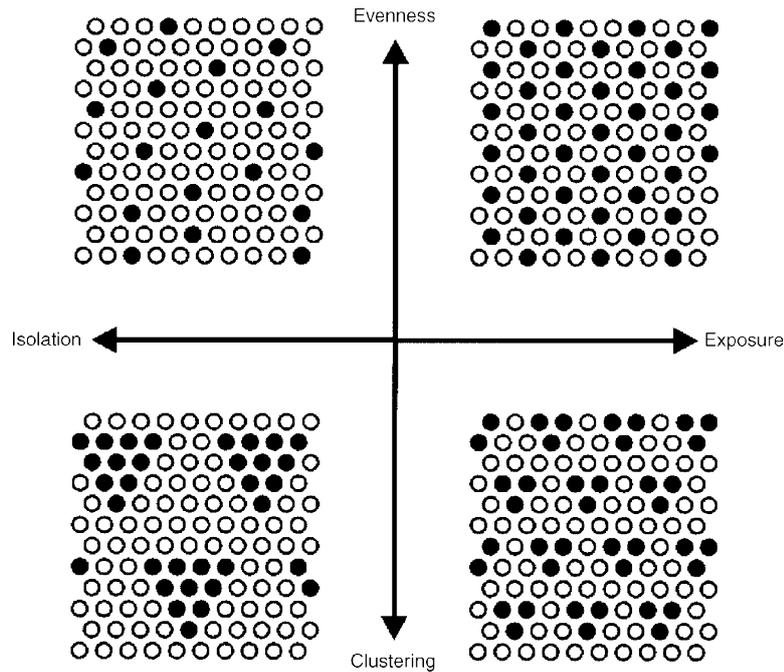


FIGURE 4.3 Représentation schématique des principales caractéristiques mesurées par les indices de ségrégation usuels : uniformité, exposition/isolement et agrégation⁷².

Si la majorité des indices présentés ici permettent de mesurer sous une forme ou une autre l'inhomogénéité de la répartition des groupes dans une ville, ils passent le plus souvent outre la complexité spatiale et multiscalaire du phénomène qu'est la ségrégation. En particulier, la perception qu'un individu peut avoir de la ségrégation de sa ville dépendra fortement de son lieu de résidence. Du fond d'un ghetto ou d'une citadelle, un groupe pourra sembler fortement représenté jusqu'à des échelles très larges par rapport au voisinage individuel, avant qu'on s'aperçoive de l'existence d'autres groupes et qu'on se fasse une image réaliste de leur répartition dans la ville entière. À l'inverse, une personne vivant dans un quartier mixte aura une très bonne image de la ville dans son ensemble et ce à toute échelle spatiale. Dans les deux dernières sections de ce chapitre, on présente une méthode visant à capturer une petite partie de cette complexité, spécifiquement la détection de certaines structures de la ségrégation sociospatiale.

⁷²Figure reproduite d'après [144].

4.2 | Détection de structures spatiales

Dans cette deuxième section, nous présentons une méthode multiscalaire de détection des structures spatiales de la ségrégation à partir de l'analyse de données réelles (taux de logements sociaux, distribution de revenus, distance à des infrastructures de transports ou de services divers...) et son application à Paris. Ce travail a été publié en 2020 [142]. Des améliorations ont été proposées depuis [30, 121, 122] et seront brièvement présentées en conclusion.

La méthode repose sur l'agrégation d'unités spatiales selon un processus bien défini qui permet d'analyser les données, pourtant collectées au niveau individuel, sur une large gamme d'échelles allant jusqu'à la ville entière, tout en évitant, par construction, les problèmes d'agrégation à une échelle arbitraire mentionnés précédemment. Des méthodes similaires existent déjà [3, 6, 38, 42, 82, 102, 132], l'agrégation d'unités spatiales s'avère en effet être un outil précieux pour avoir un meilleur aperçu des inhomogénéités et donc mieux quantifier la ségrégation.

Spécifiquement, l'évolution d'une variable statistique au cours du processus d'agrégation génère une suite de valeurs pour chaque unité de la ville choisie comme point de départ. Ces suites possèdent des caractéristiques communes, notamment leur valeur finale, correspondant à la ville entière. On peut alors s'intéresser à leurs différences, en particulier leur convergence vers cette valeur finale, pour extraire la singularité de chaque unité au sein de la ville toute entière.

4.2.1 Trajectoires et distance focale

On considère une grille comprenant N unités spatiales élémentaires (individus, logements, immeubles, quartiers...). Pour chaque unité i , on définit $g_i(n)$ comme l'ensemble des n unités agrégées récursivement de proche en proche⁷³ autour de i , avec la convention $g_i(1) = \{i\}$. On constate à partir de cette définition que, pour tout unité de départ i , $g_i(N)$ contient l'ensemble des unités de la ville.

On considère à présent la variable statistique ξ définie sur chacun des

⁷³Dans l'exemple présenté plus loin, l'unité j agrégée à l'étape n est celle dont le centroïde est le plus proche du centroïde de $g_i(n)$, mais on pourrait envisager une marche aléatoire contrainte, les unités susceptibles d'être agrégées à l'ensemble précédent étant celles reliées par une arête avec au moins une de celles présentes dans $g_i(n)$.

voisinages concentriques $g_i(n)$ de la ville contenus dans l'ensemble \mathcal{G} :

$$\mathcal{G} = \{g_i(n) \mid 1 \leq i, n \leq N\} \quad (4.1)$$

Pour chaque unité i , on peut alors construire une trajectoire

$$\mathcal{X}_{(i)} = (X_i(n))_{1 \leq n \leq N} = (\xi(g_i(n)))_{1 \leq n \leq N} \quad (4.2)$$

Cette suite représente la trajectoire d'évolution de ξ de sa valeur dans l'unité i à sa valeur sur toute la grille. De façon imagée, on peut se figurer un agent explorant la ville à partir de son lieu de résidence et accumulant ses observations quartier par quartier, jusqu'à pouvoir se représenter une vision globale de la ville. De fait, pour toute unité de départ i , on trouve $X_i(N) = \bar{X}$: toutes les trajectoires convergent vers la moyenne de la ville.

Bien que relativement facile à décrire, le calcul réel des trajectoires présente un certain nombre de difficultés d'un point de vue statistique, en particulier lors de la construction des jeux de données afin qu'ils soient définis sur \mathcal{G} .

La situation standard est la suivante : les données sont disponibles pour chaque unité, sous forme d'un couple (p_i, ξ_i) , où p_i représente la taille de l'unité considérée (nombre d'habitants, de ménages, ...) et ξ_i la valeur locale de la variable ξ , possiblement multidimensionnelle. Il pourra s'agir d'un nombre entier (services accessibles, immeubles de bureaux), d'un taux (logements sociaux), d'une densité (transports en commun), de quantiles d'une distribution voire d'une distribution complète (revenus), etc.

Lors du regroupement de n unités, il faut calculer la valeur de la variable sur ce nouvel agrégat. L'opération est aisée lorsque ξ est simplement un nombre :

$$X_i(n) = \sum_{j \in g_i(n)} \xi_j \quad (4.3)$$

De même, $X_i(n)$ est facilement calculé lorsque ξ est un taux ou une moyenne :

$$X_i(n) = \sum_{j \in g_i(n)} \xi_j p_j \Big/ \sum_{j \in g_i(n)} p_j \quad (4.4)$$

Lorsque ξ est une distribution entièrement connue, on peut construire une distribution empirique en agrégeant simplement les données individuelles du groupe d'unités spatiales considérées.

Un cas plus difficile, hélas très fréquent, est celui où ξ ne contient que certains quantiles d'une distribution. Parmi les exemples présentés par la

suite, on trouvera ainsi les données de revenus de la ville de Paris, disponibles uniquement sous forme de déciles pour chaque unité de recensement. Dans ce cas, on est contraint soit de modéliser la forme de la distribution sous-jacente (dans le cas des revenus, supposer qu'elle est log-normale ou exponentielle), soit de simuler l'ensemble de données avec des hypothèses sur les distributions intradéciles [29, 161].

Toutes les méthodes précédentes reposent sur un recensement discret des données. Il est possible de travailler directement en termes de densité locale, par exemple dans le cas de la disponibilité des transports publics. Ainsi, si $(m_i(\ell), p_i(\ell))$ mesurent respectivement le nombre de stations de métro et la taille de la population présentes dans un disque de rayon ℓ centré sur l'unité i , on peut définir⁷⁴

$$r_i(\ell) = \frac{m_i(\ell)/M}{p_i(\ell)/P} \quad (4.5)$$

Cette quantité est semblable à la *représentation* définie dans l'article de R. Louf et M. Barthélemy [103]. Comparer la représentation à 1 est équivalent à comparer la densité locale de stations de métro par habitant à la densité moyenne de la ville, et permet comme précédemment de construire des trajectoires en faisant varier ℓ jusqu'au rayon complet de la ville.

À partir du faisceau de trajectoires ainsi constitué, deux approches complémentaires sont envisageables.

À l'aide d'algorithmes de classification, il est ainsi possible d'identifier les unités possédant des trajectoires similaires. En étudiant ces trajectoires par échantillonnage [3, 38, 132], il est possible de contourner l'arbitraire des délimitations administratives et l'aplanissement des données qui en résulte, tout en ayant une vision moins limitée que la simple étude des données individuelles aisées à collecter et à représenter sur une carte, mais ne permettant d'établir des comparaisons entre villes, par exemple.

D'autre part, on peut noter que par construction, lorsque n est proche de 1, la valeur $X_i(n)$ dépend fortement de la quantité locale ξ_i . À l'inverse, on a montré que lorsque $n \rightarrow N$, toutes les trajectoires convergent vers la moyenne de la ville \bar{X} . Il existe donc un croisement de ces deux régimes, qui ne peut être détecté que par une étude complète de la trajectoire. Cette deuxième approche consiste à déterminer le nombre d'unités qu'il est nécessaire d'agréger autour de celle de départ pour se rapprocher suffisamment

⁷⁴En notant M le nombre total de stations dans la ville et P la population totale.

de la valeur moyenne \bar{X} de la ville. La *distance focale* est ainsi définie par :

$$f_i(\delta) = \inf_n \left\{ n \leq N \mid \forall n' \geq n, |X_i(n') - \bar{X}| \leq \delta \right\} \quad (4.6)$$

Géographiquement, elle joue le rôle d'un rayon de convergence : c'est une approximation de la zone à explorer localement pour obtenir une perception réaliste de la ville dans son ensemble. Il est possible de décliner cette définition pour différentes expressions de la distance. Ainsi, pour comparer une distribution intermédiaire avec la distribution de la ville entière, on pourra utiliser la divergence de Kulback-Leibler [30, 121] ou toute autre mesure plus adaptée au cas étudié.

4.2.2 Inhomogénéités spatiales à Paris

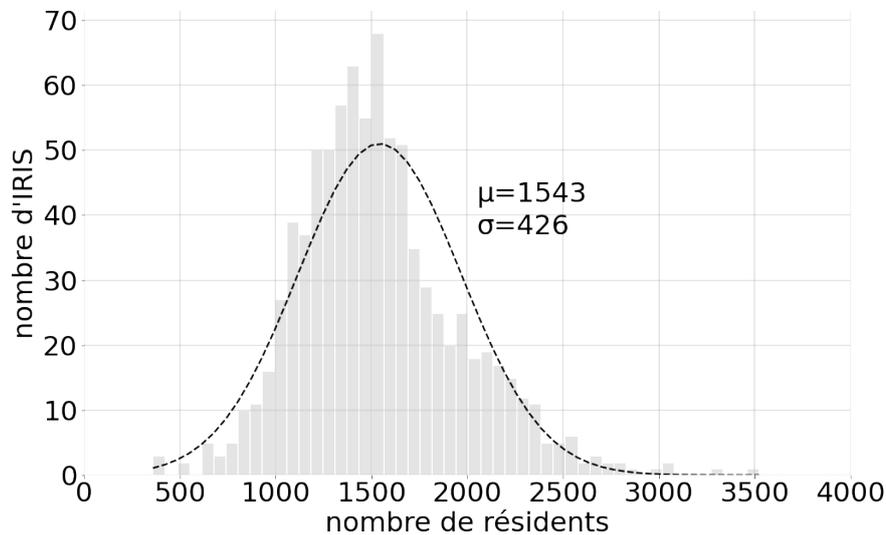


FIGURE 4.4 Histogramme des IRIS parisiens en fonction de leur nombre de résident·e·s. La taille moyenne d'un IRIS est $\mu = 1543$ habitant·e·s, l'écart-type de la distribution est environ $\sigma = 426$.

À titre d'illustration, cette section présente une application de la méthode décrite plus haut à plusieurs types de données INSEE disponibles pour la ville de Paris à l'échelle de l'IRIS : “îlot regroupé pour l'information statistique”⁷⁵. On en dénombre 992 en tout et pour l'année 2016, la dernière disponible au moment de réaliser ces travaux, 936 IRIS sur les 949 d'habitat ou d'activité⁷⁶ dans Paris comptaient au moins un logement, c'est donc sur ceux-là que s'appuie la suite.

⁷⁵<https://www.insee.fr/fr/metadonnees/definition/c1523>, INSEE (2017).

⁷⁶Séparés pour des raisons qui n'interviendront pas dans la suite en fonction du rapport emplois salariés / population résidente : si ce rapport est supérieur à 2 on parle d'IRIS d'activité, sinon d'IRIS d'habitat.

D'autres unités existent pour Paris, comme les 80 quartiers administratifs (échelle utilisée par la préfecture de police, chaque quartier possède un poste de police), les 20 arrondissements (groupements de quatre quartiers), ou encore les 18 circonscriptions législatives. Toutefois, l'IRIS reste la plus petite, tant en superficie qu'en nombre d'habitants ou de ménages, tout en étant suffisamment grande pour éviter les problèmes de sous-échantillonnage mentionnés en section **4.1.2**, un IRIS comptant en moyenne 1543 résident·e·s.

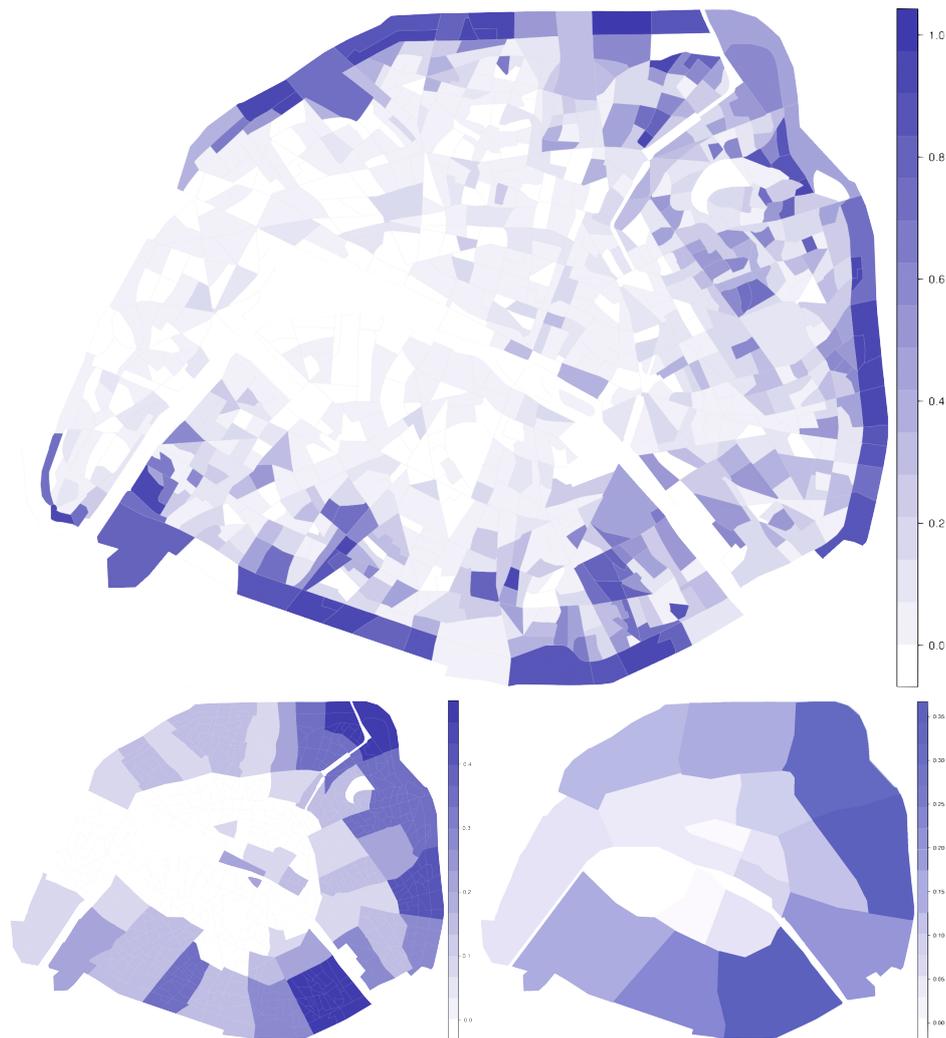


FIGURE 4.5 Taux de logements sociaux (HLM) moyens représentés pour 3 échelles disponibles à Paris (IRIS en haut, quartiers administratifs et arrondissements en bas, à gauche et à droite respectivement)

Pour chaque IRIS, le nombre de logements et parmi ceux-ci, le taux de logements sociaux sont disponibles, c'est le cas le plus simple, correspondant à l'équation (4.4). Il sera alors possible de calculer ce taux pour

tout groupe d'IRIS et donc de construire les trajectoires. En particulier, on peut calculer le taux de logements sociaux par quartier ou par arrondissement (figure 4.5). On procède ainsi à un échantillonnage classique, fonction d'une découpe administrative arbitraire du territoire, consistant à étudier des coupes transversales de ces trajectoires, avant d'évaluer les similarités entre les régions ainsi définies. C'est la première approche présentée plus haut. La seconde consiste à étudier les trajectoires entières.

Logements sociaux et distributions de revenus Représenter les trajectoires obtenues pour une partie des IRIS parisiens (un dixième environ) en figure 4.6 permet de mieux se représenter l'objet de l'étude. On constate, comme anticipé, que toutes les trajectoires convergent vers la moyenne de la ville, environ 17.9%, représentée par une ligne horizontale.

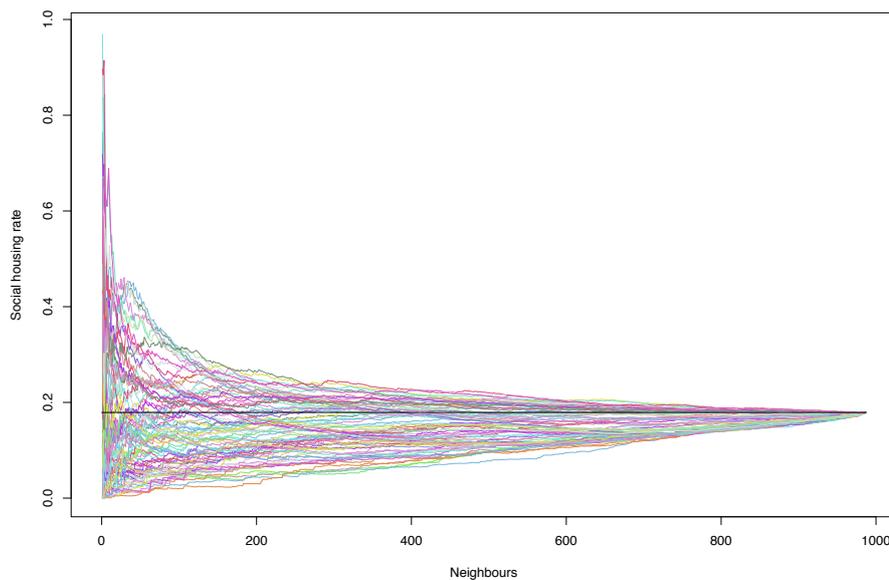


FIGURE 4.6 Trajectoires pour la variable *taux de logements sociaux*, à partir d'une centaine des 936 IRIS parisiens. La ligne continue horizontale correspond à la moyenne de la ville ($\bar{X} = 17.9\%$).

Deux trajectoires sélectionnées au hasard dans ce faisceau peuvent présenter des distances focales très différentes. On constate cependant que cette observable admet une forme de continuité : les voisinages concentriques d'une unité et de sa voisine la plus proche demeurent relativement similaires au cours du processus d'agrégation, le nombre d'unités à agréger pour atteindre l'intervalle $[\bar{X} - \delta, \bar{X} + \delta]$ est donc relativement similaire.

On peut constater cette similarité en figure 4.7, représentant les distances focales $f_i(\delta)$ pour chaque IRIS, où $\delta = 0.05$.



FIGURE 4.7 Distances focales $f_i(\delta)$ des IRIS parisiens, exprimées en nombre d'unités agrégées pour les trajectoires relatives à la variable *taux de logements sociaux* (HLM).

Que peut-on observer ? Tout d'abord, les distances focales des IRIS du centre-ouest de Paris (couvrant les VII, VIII, XVI et XVII^e arrondissements et une partie du I^{er}) sont bien plus élevées que celles de la partie nord-est (XIX et XX^e arrondissements) d'un facteur 2 à 4. Pourtant, ces deux ensembles correspondent en figure 4.5 à des valeurs extrêmes de la variable considérée, avec une concentration de logements sociaux dans la partie nord-est et un taux nettement inférieur à la moyenne dans la partie ouest, et ont des positions géographiquement symétriques. Rien n'indiquait à priori que les distances focales sont à ce point éloignées.

Ensuite, et presque inversement, on observe un arc d'IRIS convergeant très vite vers la moyenne de la ville, indiquant une mixité sociale plus importante. Historiquement, cette zone longe une ancienne enceinte de Paris, le Mur des Fermiers Généraux, érigé à la fin du XVIII^e siècle. Juste à l'extérieur des murs de la ville se trouvaient des faubourgs, qui ont été transformés en boulevards sous le Second Empire. La faible distance focale de cet ensemble d'unités semble ainsi révéler une empreinte durable en terme de diversité sociale, par le prisme de son proxy, le taux de logements sociaux.

On peut appliquer la même méthode aux distributions de revenus. N'ayant accès qu'à certains quantiles de la distribution pour chaque IRIS, les données correspondant au nombre de foyers par voisinage $g_i(n)$ ont dû être simulées à partir de distributions log-normales, la mieux ajustée aux quantiles disponibles. On obtient alors les cartes représentées en figure 4.8.

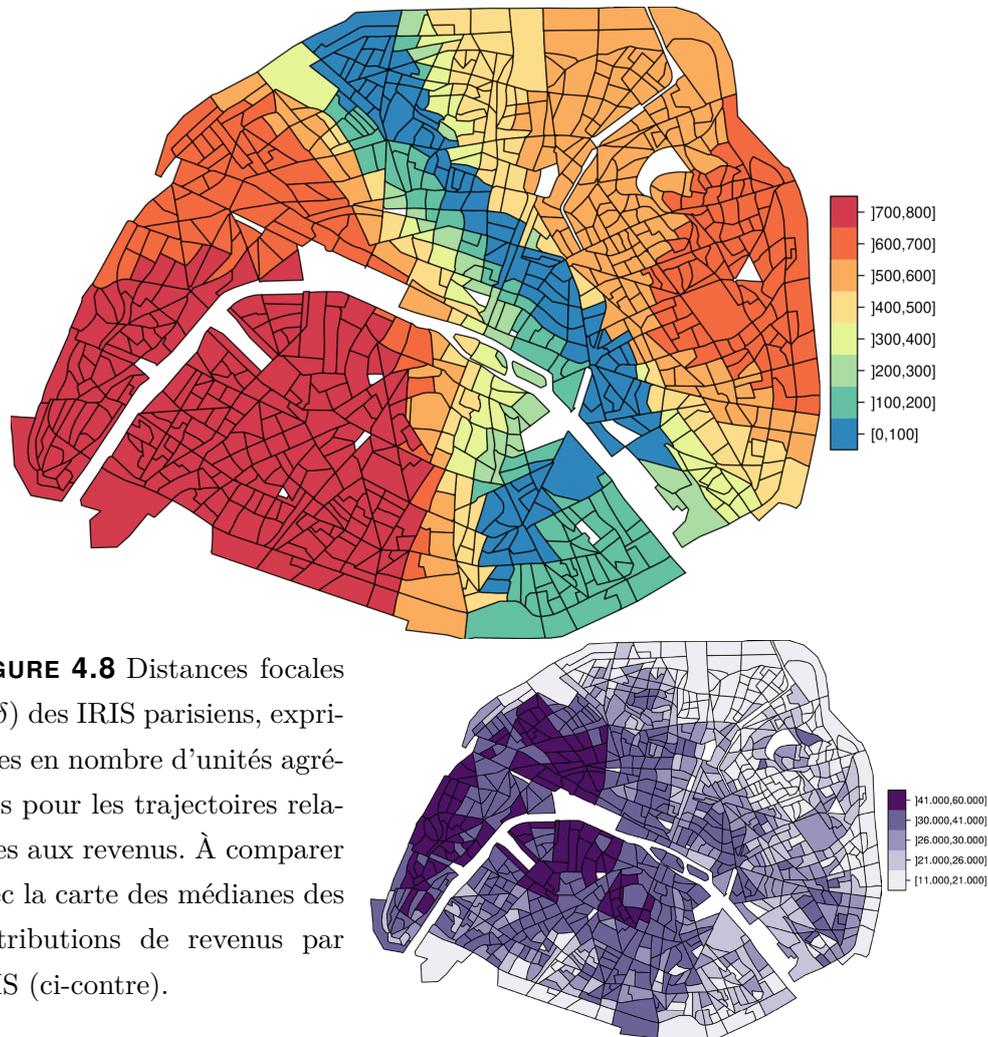


FIGURE 4.8 Distances focales $f_i(\delta)$ des IRIS parisiens, exprimées en nombre d'unités agrégées pour les trajectoires relatives aux revenus. À comparer avec la carte des médianes des distributions de revenus par IRIS (ci-contre).

La carte est bien plus symétrique que pour le taux de logements sociaux. Les régions les plus riches (centre ouest) et les plus pauvres (nord-est) de la ville présentent une convergence relativement lente vers la distribution globale, en comparaison aux quartiers plus mixtes comme les faubourgs. Cependant, une grande partie de la région la plus riche, englobant les VI, VII et XVI^e arrondissements, est de nouveau la dernière à converger. On remarque que les XIV et XV^e arrondissements, pourtant plus inhomogènes, se trouvent également dans cette zone "lente". Plus surprenant, le VIII^e arrondissement ne se distingue pas pour la distribution des revenus. Il est géographiquement beaucoup plus proche des quartiers mixtes que, par exemple, le XVI^e

arrondissement.

Que conclure de ces deux exemples ? Les deux observables utilisées permettent de noter que les grandes distances focales associées aux quartiers les plus favorisés, par comparaison avec les quartiers intermédiaires mais surtout avec les plus défavorisés, indique leur isolement dans la ville. Ce constat est à rapprocher de la mise en évidence par plusieurs chercheurs en sciences sociales de l'existence d'un processus apparemment inverse à celui de la ségrégation, et pouvant se produire lui aussi en banlieue. L'agrégation affinitaire, c'est-à-dire la création de "ghettos du gotha" pour reprendre le titre du livre de M. Pinçon et M. Pinçon-Charlot [136], est le fait des classes aisées et non des classes moins favorisées.

Il est remarquable que ce soit pour le groupe sur lequel les contraintes économiques sont les moins contraignantes que l'on constate ce choix aussi manifeste de vivre ensemble. S'il y a ghetto, c'est donc sur un mode volontaire et maîtrisé.

Plus spécifiquement, E. Maurin [108] constate une fragmentation spatiale reconnue politiquement, en atteste l'action publique dirigée vers les zones identifiées comme défavorisées. Selon son étude, ces aides, trop réduites et trop localisées, sont inefficaces pour contrer cette ségrégation stable depuis une vingtaine d'années. En effet, ces politiques publiques, héritières des théories de l'exclusion des années 1990⁷⁷ ne suffisent pas à compenser une stratification plus complexe, en particulier une tendance collective à éviter les classes sociales immédiatement inférieures sur l'échelle des difficultés. La mobilité résidentielle qui en découle est donc fortement sélective et mène à l'émergence de *citadelles*. Alors que l'attention des pouvoirs publics est mobilisée par la question de l'exclusion sociale, les classes supérieures sont de fait plus concentrées dans les beaux quartiers que ne le sont les plus pauvres dans les quartiers stigmatisés de la périphérie urbaine [8, 134, 141].

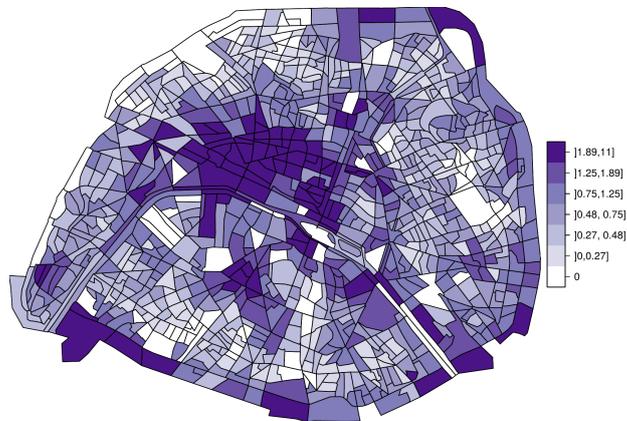
Au regard des cartes, il semble en effet que les classes les plus aisées se regroupent bien plus et sur des échelles bien plus étendues que les classes intermédiaires à pauvres. Ce séparatisme des classes supérieures, qui a été extensivement étudiée depuis [43–45, 134] est donc mis en évidence par notre méthode. Le constat est à nuancer par la délimitation de l'étude à

⁷⁷Se référer notamment aux travaux de A. Touraine : "*Le problème aujourd'hui n'est pas l'exploitation, mais l'exclusion*" [169].

Paris intramuros, alors que les études sociologiques citées précédemment pointent le rôle des banlieues dans ce phénomène. Une étude similaire est alors à envisager à l'échelle de l'Île de France, ou au moins de la métropole du Grand Paris.



FIGURE 4.9 Distances focales $f_i(\delta)$ des IRIS parisiens, exprimées en nombre d'unités agrégées pour les trajectoires relatives à la proximité des transports en commun. À comparer avec la carte des densités de stations de métro et tramway par IRIS (ci-contre).



Mobilités La figure 4.9 illustre un dernier exemple parisien d'utilisation de l'équation (4.5) dans le cas d'une variable continue : la densité de stations de transports en commun (spécifiquement la répartition spatiale des 359 stations de métro et de tramway). On constate que le réseau métropolitain est principalement organisé autour d'un axe Nord-Sud, ce qui n'apparaît pas de façon flagrante sur la carte des densités. Celui-ci met au contraire en avant la présence des deux lignes périphériques de tramway (51 stations réparties le long des frontières sud et est), bien que leur impact en terme de distance focale semble relativement faible.

Pour rester dans le contexte des mobilités, on peut signaler une dernière caractéristique de cette méthode : sa robustesse à un changement du processus d'agrégation. Plutôt que de considérer une marche d'une unité à sa voisine la plus proche, on peut ainsi considérer des séquences définies à partir de trajectoires personnelles (unités où l'on habite, travaille, sort, etc.). Ces trajectoires personnelles transformeraient le réseau global en réseaux personnels. Pour certains, le réseau spatial pourrait être un réseau global, en raison de leur grande mobilité dans les différents quartiers de la ville. Pour d'autres, les trajectoires personnelles resteraient concentrées autour de leur unité de résidence. Une telle étude pourrait être menée à l'aide de sondages de pratiques ou de données de mobilité (par géolocalisation mobile par exemple).

Coefficients de distorsion Une question non-développée dans l'article originel [142] est celle de la définition de la convergence, et spécifiquement de l'arbitraire introduit par le choix de δ . Une technique développée dans des articles plus récents [30, 121, 122] consiste à calculer $f_i(\delta)$ pour toutes les valeurs de δ pertinentes à l'étude (par exemple, pour le taux de logements sociaux, entre 0 et 1) puis à construire pour chaque unité i le *coefficient de distorsion*

$$\mathcal{D}_i = \int_0^\infty f_i(\delta) d\delta \quad (4.7)$$

Il fournit, pour chaque unité i , une mesure de la distorsion de l'image qu'un individu résidant dans l'unité se fait de la ville par rapport à la répartition réelle des groupes dans la ville, et qui lève ainsi le dernier choix arbitraire de la méthode.

Pour construire une version plus interprétable du coefficient de distorsion et pour pouvoir mieux comparer deux villes entre elles, M. Olteanu et al. [122] suggèrent une normalisation de Δ_i par sa valeur obtenue pour la même géographie mais dans la configuration la plus ségréguée possible, donnant ainsi une valeur entre 0 et 1. La comparaison à un modèle de référence est une méthode couramment utilisée en analyse de données, en particulier par rapport à une hypothèse d'absence de l'effet que l'on cherche à mesurer : l'hypothèse nulle. Dans le cas de la ségrégation, il s'agit d'une ville à la répartition uniforme.

4.3 | La ville uniforme comme référence

La généralisation de l'usage des outils statistiques dans de nombreux domaines des sciences sociales, de la géographie urbaine à l'histoire et à l'archéologie [59, 112, 121, 122], conduit au besoin grandissant de construire des modèles s'appuyant sur une hypothèse nulle pour servir de référence à l'étude de ces nouveaux objets. L'hypothèse nulle dans le cas de la ségrégation, définie par une distribution spatiale déviant significativement de celle produite par une distribution aléatoire [103], est donc celle d'une ville uniforme, dont chaque unité présente en moyenne les mêmes caractéristiques que celles de la ville toute entière. Par extension, toute trajectoire définie à partir d'une règle spatiale (agrégation séquentielle d'unités voisines, par exemple) doit constituer une suite de valeurs dans laquelle il ne peut subsister aucune trace de cette règle.

Tout écart par rapport à un tel comportement aléatoire, tout motif dans la distribution spatiale, est le signe qu'une structure sous-jacente conduit à un brassage biaisé des unités.

4.3.1 Construire une ville uniforme

On considère le cas, idéal pour le processus d'agrégation, où les données sont accessibles au niveau individuel. On suppose que dans chaque unité i réside un individu qui appartient au groupe A ou B . On note

$$\xi_i = \begin{cases} 1 & \text{si l'individu de l'unité } i \text{ appartient au groupe } A \\ 0 & \text{si l'individu de l'unité } i \text{ appartient au groupe } B \end{cases} \quad (4.8)$$

On introduit alors

$$S_n = \sum_{i=1}^n \xi_i \quad (4.9)$$

avec $S_0 = 0$. \mathcal{S} est alors la trajectoire correspondant à l'unité $i = 1$. Ce point de départ n'a aucune importance particulière, pas plus que l'ordre dans lequel les unités successives sont choisies (cette ville n'a pas de géométrie prédéfinie). Il existe donc à priori $N!$ copies de cette trajectoire \mathcal{S} , ce qui constitue le faisceau de la ville.

La ville contient un nombre donné d'individus des groupes A et B , toutes les trajectoires partageront donc la même valeur finale $X_N = \rho N$, où ρ est la proportion d'individus du groupe A . Dans le cas où la ville est parfaitement

mixte, la trajectoire est une marche aléatoire d'équation maîtresse

$$P(S_{n+1} = x + 1) = \frac{\rho N - s}{N - n} P(S_n = x) + \left(1 - \frac{\rho N - (x + 1)}{N - n}\right) P(S_n = x + 1) \quad (4.10)$$

Il s'agit du tirage sans remise de n éléments dans une urne contenant ρN éléments du groupe A et $(1 - \rho)N$ éléments du groupe B , on peut donc l'écrire comme la loi hypergéométrique [58] :

$$P(S_n = x \mid S_N = \rho N) = \binom{\rho N}{x} \binom{(1 - \rho)N}{n - x} \binom{N}{n}^{-1} \quad (4.11)$$

D'autres trajectoires peuvent être définies de manière similaire, celle des moyennes pour le groupe A

$$m_n = S_n/n \quad (4.12)$$

celle des moyennes pour le groupe B

$$M_n = (n - S_n)/n \quad (4.13)$$

et celles des différences, où un individu du groupe A ajoute 1 et un individu du groupe B ajoute -1 à la valeur à l'instant précédent

$$D_n = 2S_n - n \quad (4.14)$$

Par construction, ces quatre trajectoires convergent vers des valeurs finales bien définies :

$$S_N = \rho N, \quad m_N = \rho, \quad M_N = 1 - \rho, \quad D_N = (2\rho - 1)N \quad (4.15)$$

D'un point de vue probabiliste, ces quatre marches aléatoires sont entièrement caractérisées par ρ et N . Elles représentent quatre facettes du même modèle de référence : la ville uniforme.

À présent, on peut construire N trajectoires pour la variable ξ . Dans une ville uniforme, l'échantillon de trajectoires ainsi produit présente les propriétés statistiques d'un ensemble de N trajectoires obtenues lors d'un tirage sans remise. Tout écart statistiquement significatif des trajectoires réelles par rapport à ce faisceau signale alors la présence d'inhomogénéités spatiales à plusieurs échelles.

4.3.2 Propriétés des trajectoires

La méthode présentée dans cette partie pour analyser ce faisceau de trajectoires et le comparer aux trajectoires réelles fait intervenir un théorème, appelé théorème du scrutin (*ballot theorem*), ou plutôt un ensemble de théorèmes lié au décompte des voix dans une élection.

Énoncé par J. Bertrand [19], démontré⁷⁸ par D. André [4] et généralisé par E. Barbier [13] en 1887, l'enjeu du théorème du scrutin est d'évaluer la probabilité qu'au cours du décompte des voix d'une élection à deux candidats A et B , sachant que A l'emporte, les voix pour A soient majoritaires à tout instant du dépouillement. À l'aide des marches définies dans le paragraphe précédent, on peut écrire cette probabilité de trois manières équivalentes

$$\begin{aligned} P(S_n > n/2, \forall n < N \mid X_N = \rho N) &= 2\rho - 1 \\ &= P(m_n > 1/2, \forall n < N \mid m_N = \rho) \\ &= P(D_n > 0, \forall n < N \mid D_N = 2\rho - 1) \end{aligned} \quad (4.16)$$

Symétriquement, la probabilité que les voix pour B soient minoritaires à tout instant du dépouillement sont

$$P(M_n < 1/2, \forall n < N \mid M_N = 1 - \rho) = 2\rho - 1 \quad (4.17)$$

En 1962, L. Takács [165] réécrit le théorème sous sa forme la plus couramment employée depuis en considérant la marche \mathbf{X} définie par $\forall n$, $X_n = 2(n - S_n)$:

$$P(X_n < n, \forall n < N \mid X_N = k) = 1 - \frac{k}{N} \quad (4.18)$$

Ce théorème est toujours vrai lorsque les variables ξ_i sont à valeurs dans \mathbb{N} , à condition que $X_N = k \leq N$. Il l'est toujours [166] lorsque \mathbf{X} est un processus stochastique à valeurs réelles bien choisi⁷⁹. On a alors

$$P(X_t \leq t, \forall t \in [0, T] \mid X_T = x) = \left(1 - \frac{x}{T}\right) \mathbb{1}_{[0, T]}(x) \quad (4.19)$$

Quelle interprétation tirer de ce théorème en termes de recensement urbain ? On considère un individu partant d'un point de la ville et considère

⁷⁸La démonstration la plus courante [1, 58] fait généralement intervenir la marche \mathbf{D} et un élégant principe de réflexion.

⁷⁹Séparable, à variations cycliquement échangeables sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et tel que pour presque tout $\omega \in \Omega$, $X_t(\omega)$ soit une fonction croissante, en escalier, et vérifiant $X_0(\omega) = 0$.

le premier logement sur son chemin, puis les deux premiers logements, et ainsi de suite jusqu'à avoir parcouru toute la ville. La probabilité $p(\rho)$ qu'à chaque étape cet individu dénombre une *minorité* de logements sociaux, étant donné que leur proportion dans l'ensemble de la ville est $1 - \rho$, est, d'après l'équation (4.17), $p(\rho) = 2\rho - 1$.

Dans l'exemple de Paris, où $1 - \rho = 17.9\%$, cette probabilité devrait être $p(\rho) = 64.2\%$. Cela signifie que 64.2% des trajectoires \mathcal{M} du faisceau obtenu devraient évoluer strictement sous la droite horizontale $\tau : t \mapsto 1/2$. En réalité, près de 85% d'entre elles sont dans ce cas, comme illustré en figure 4.10. Cet écart au modèle uniforme indique une structuration spatiale des quartiers comptant des logements sociaux, on ne peut pas la qualifier plus précisément avec ce seul résultat, mais il semble probable qu'il s'agisse d'un effet de regroupement de ces logements adressés prioritairement à des populations défavorisées dans certains quartiers plutôt que d'autres.

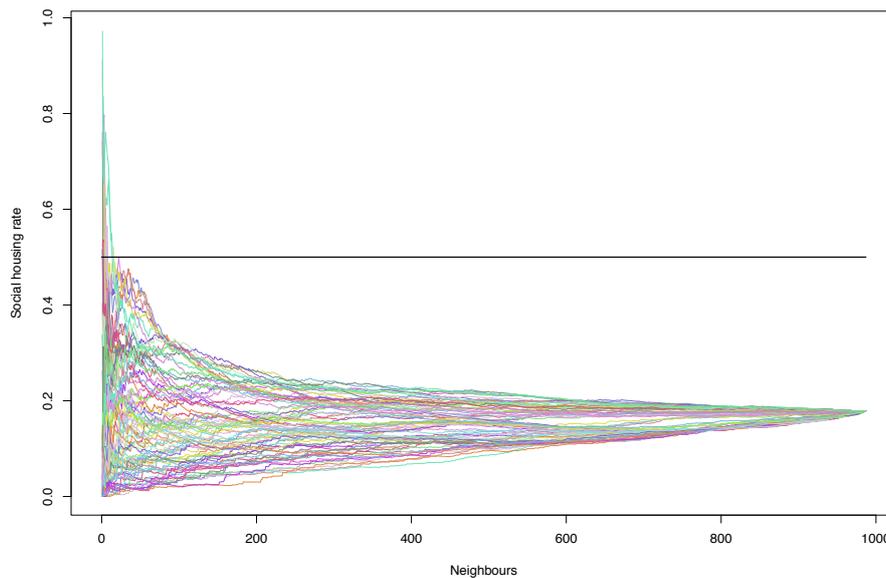


FIGURE 4.10 Les trajectoires \mathcal{M} du taux de logement social, à partir d'une centaine des 936 unités statistiques de Paris. La ligne continue correspond à un taux $\tau = 1/2$. D'après le théorème du scrutin, pour une ville "mixte" en matière de logements sociaux, 64.2% d'entre elles devraient évoluer strictement sous ce seuil. En réalité, près de 85% d'entre elles sont dans ce cas.

La structure sociospatiale de Paris n'est qu'une réalisation possible d'un tel faisceau de trajectoires, le hasard a pu conduire à un écart, observé à l'aide du théorème du scrutin, avec la ville mixte. Il faut alors construire un

test pour déterminer si cet écart est significatif ou un simple effet d'échantillonnage aléatoire. Pour une trajectoire $\mathcal{M}_{(i)}$ partant de l'IRIS i , on définit la variable aléatoire de Bernoulli suivante :

$$\chi_i = \begin{cases} 1 & \text{si } \mathcal{M}_{(i)} \text{ reste sous le seuil } \tau = 1/2 \\ 0 & \text{sinon} \end{cases} \quad (4.20)$$

D'après le théorème du scrutin,

$$\forall i \in \llbracket 1, N \rrbracket, \quad \begin{cases} P(\chi_i = 1) = 2\rho - 1 \\ P(\chi_i = 0) = 2(1 - \rho) \end{cases} \quad (4.21)$$

La somme de ces variables aléatoires pour un tirage de N trajectoires dans une ville uniforme doit suivre une distribution binomiale d'espérance $(2\rho - 1)N$ et de variance $2(1 - \rho)(2\rho - 1)N$. Dans le cas de Paris, ces valeurs deviennent $0.642N$ et $0.230N$ respectivement. Un simple test d'ajustement montre que les trajectoires parisiennes ne satisfont pas au théorème du scrutin, avec une p -value de l'ordre de 10^{-16} . Cette valeur absurde est due au fait que le modèle de la ville uniforme est un cas extrême et irréaliste qui fournit une échelle de comparaison trop extrême, si l'on veut l'utiliser directement comme mesure de ségrégation.

Toutefois, ce modèle est utile si on le considère comme une échelle absolue, une référence au même sens et au même usage que le zéro absolu physique. On peut à présent comparer des villes entre elles à *leur valeur idéale près*.

En pratique cependant, on a rarement accès aux données unitaires. Chaque trajectoire obtenue dans l'exemple parisien est en réalité un échantillonnage d'une trajectoire "réelle" comparable à celle construite par une des marches précédentes. Toute comparaison avec le théorème du scrutin doit en fait être faite à partir de marches aléatoires *subordonnées*, c'est-à-dire observées uniquement pour une séquence d'instantanés, correspondant à la taille d'un IRIS.

4.3.3 Subordination

Dans cette dernière section sont présentées trois méthodes distinctes pour résoudre le problème de l'échantillonnage des données. Pour pouvoir comparer les données réelles, disponibles à l'échelle de l'IRIS, c'est-à-dire environ tous les 1500 logements, à l'hypothèse nulle d'une ville mixte et améliorer le test par le théorème du scrutin présenté précédemment, il faut subordonner la marche aléatoire \mathcal{X} en fonction de la distribution de la taille des IRIS

présentée en figure 4.4. On se place dans le cas simple où les termes de la suite \mathcal{X} sont des nombres.

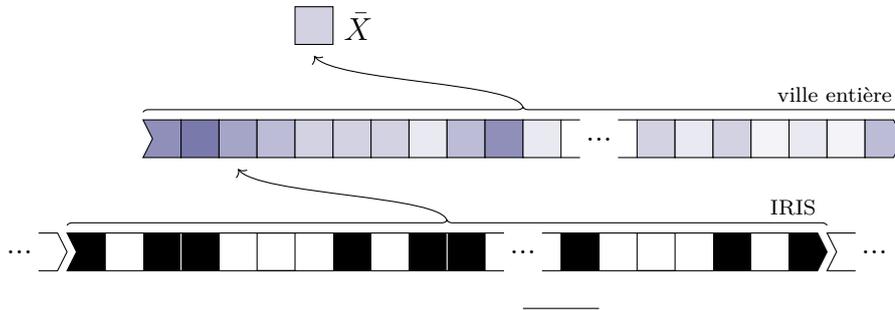


FIGURE 4.11 Représentation schématique du processus de subordination : d’unités individuelles binaires, l’information extraite par recensement est agrégée à l’échelle de l’IRIS, plus petite unité pour laquelle l’information est disponible.

La subordination d’une marche aléatoire est une opération d’échantillonnage : une marche aléatoire subordonnée \mathcal{Y} est obtenue à partir d’une marche aléatoire \mathcal{X} , appelée processus parent, et d’un processus \mathcal{T} à incréments positifs, appelé temps opérationnel, lorsque

$$P(Y_t = x) = \sum_{n=0}^{\infty} P(X_n = x)P(T_t = n) \tag{4.22}$$

Cette équation est présentée dans [58] sous sa forme continue, en faisant intervenir les densités de probabilités.

Dans le cas qui nous intéresse, le processus parent est déterminé par la loi hypergéométrique (4.11) :

$$P(S_n = x) = \binom{\rho N}{x} \binom{(1-\rho)N}{n-x} \binom{N}{n}^{-1} \tag{4.23}$$

Subordination discrète Dans ce paragraphe, on propose une méthode pour déterminer complètement le propagateur⁸⁰ d’une marche aléatoire extraite de la marche principale, que l’on échantillonne comme si on ne l’observait qu’à certains instants, eux aussi déterminés aléatoirement.

Supposons dans un premier temps que tous les IRIS ont la même taille, par exemple la moyenne μ déterminée à partir de l’histogramme 4.4, que

⁸⁰Le propagateur d’une marche aléatoire est la probabilité

$$P_0(X_n = x) = P(X_n = x \mid X_0 = 0)$$

l'on renomme τ_0 . Le propagateur du processus \mathcal{T} est alors

$$P_0(T_t = n) = \delta(n - \tau_0 t) \quad (4.24)$$

Dans ce cas, la marche subordonnée a simplement le même propagateur que le processus parent, avec un temps redimensionné $n = \tau_0 t$

$$P_0(Y_t = x) = \binom{\rho N}{x} \binom{(1-\rho)N}{\tau_0 t - x} \binom{N}{\tau_0 t}^{-1} \quad (4.25)$$

On suppose à présent que les tailles d'IRIS sont contenues dans le vecteur $\boldsymbol{\tau} = (\tau_1, \dots, \tau_M)$, avec $M \leq 936$ pour le cas parisien. On obtient le propagateur suivant pour \mathcal{T} :

$$P_0(T_t = n) = t! \sum_{|\mathbf{K}_t|=t} A(\mathbf{K}_t) \delta(n - \mathbf{K}_t \cdot \boldsymbol{\tau}) \quad (4.26)$$

où $\mathbf{K}_t = (k_1, \dots, k_M)$ est un vecteur d'indices entiers dont la somme vaut t et où le paramètre $A(\mathbf{K}_t) = \prod_{\ell=1}^M a_\ell^{k_\ell} / k_\ell!$ fait intervenir les coefficients a_ℓ dénombrant les IRIS de taille τ_ℓ .

Le nombre de termes présents dans cette somme, qui apparaît aussi dans $P_0(Y_t = x)$, est le nombre de partitions de l'entier t , dont un équivalent [81] lorsque $t \rightarrow \infty$ est :

$$\#\text{termes} \sim \frac{1}{4t\sqrt{3}} \exp\left(\pi\sqrt{\frac{2t}{3}}\right) \quad (4.27)$$

La ville de Paris comptant plus d'un million de logements⁸¹, on peut estimer que $\log(\#\text{termes}) \sim 10^3$, sans qu'il soit ici question de quantifier le reste des opérations à effectuer pour obtenir un résultat exploitable.

Marche bidirectionnelle Le problème du scrutin peut être adapté en une marche aléatoire bidimensionnelle si on remplace les incréments $\{0, 1\}$ par un pas vers la droite ou vers le haut, respectivement. Cette représentation est équivalente, à une transformation linéaire près, à la marche \mathcal{S} de loi hypergéométrique, comme l'illustre le schéma suivant.

L'enjeu pour que cette marche aléatoire soit "victorieuse" dans le cadre du théorème du scrutin est que le chemin de l'origine $O = (0, 0)$ au point final $P = (i, j)$ (avec $i = (1-\rho)N < j = \rho N$) n'entre jamais en contact avec la diagonale (OI) , où $I = (i, i)$. Ce nouveau formalisme permet de réutiliser une partie des résultats de W. Whitworth [174], en particulier :

⁸¹<https://www.insee.fr/fr/statistiques/zones/4177606>, INSEE (2016).

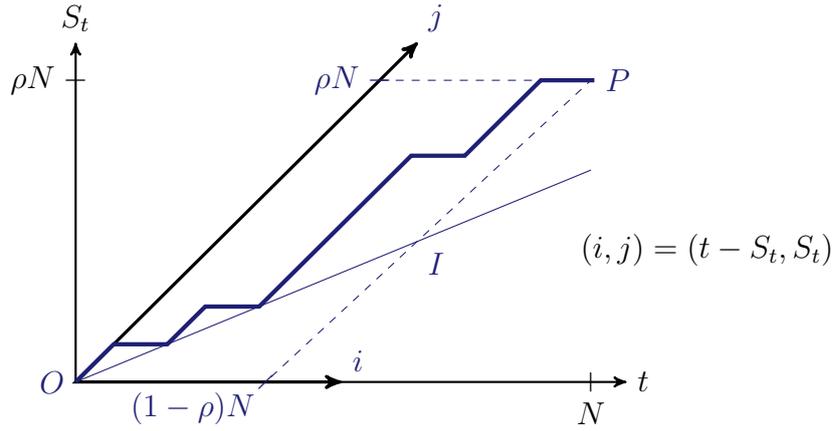


FIGURE 4.12 Représentation schématique de l'équivalence entre la marche hypergéométrique \mathcal{S} et la marche bidirectionnelle de W. Whitworth.

— Le nombre total de chemins OP est

$$c_{i,j} = \frac{(i+j)!}{i!j!} \quad (4.28)$$

— Le nombre de chemins OI ne touchant la diagonale (OI) qu'en leurs extrémités est

$$\phi_{i,i}^* = \frac{2}{i} c_{i-1,i-1} \quad (4.29)$$

— En conséquence⁸², le nombre de chemins OP étant toujours strictement au-dessus de la diagonale (OI) est

$$\phi_{i,j} = \frac{j-i}{i+j} c_{i,j} \quad (4.30)$$

— Le nombre de chemins OP étant toujours strictement au-dessus d'une droite (HH') parallèle à la diagonale (OI) et distante verticalement de celle-ci de h pas *vers le bas* est

$$\psi_{i,i}(h) = \begin{cases} c_{i,j} - c_{i-h+1,j+h-1} & \text{si } i > h - 1 \\ c_{i,j} & \text{sinon} \end{cases} \quad (4.31)$$

La probabilité donnée par le théorème du scrutin est donc bien

$$W_\rho = \frac{\phi_{i,j}}{c_{i,j}} = \frac{j-i}{i+j} = 2\rho - 1 \quad (4.32)$$

On souhaite échantillonner cette marche en la contraignant à passer par les points successifs $P_p = (i_p, j_p)$, $\forall p \in \llbracket 0, n+1 \rrbracket$, avec $i_0 = j_0 = 0$, $i_{n+1} = i$ et

⁸²Se référer à l'annexe **G**

$j_{n+1} = j$. On note $x_p = i_{p+1} - i_p$ et $y_p = j_{p+1} - j_p$ les écarts horizontaux et verticaux entre ces points.

La probabilité d'effectuer le trajet $OP_1P_2\dots P_nP$ en restant strictement au-dessus de la diagonale (OI) est alors

$$W_\rho(P_1, \dots, P_n) = \frac{\phi_{i_1, j_1}}{c_{i_1, j_1}} \prod_{p=1}^n \frac{\psi_{x_p, y_p}(j_p - i_p)}{c_{x_p, y_p}} \quad (4.33)$$

$$= \frac{j_1 - i_1}{i_1 + j_1} \prod_{p \in \mathbb{P}} \left(1 - \frac{x_p! y_p!}{(i_{p+1} - j_p + 1)!(j_{p+1} - i_p - 1)!} \right) \quad (4.34)$$

où l'ensemble \mathbb{P} contient tous les entiers p tels que $i_{p+1} > j_p - 1$.

Cette méthode permet de déterminer la proportion de marches bidimensionnelles restant constamment au-dessus du seuil voulu passant par une série de points choisis par l'observateur.

L'inconvénient de cette méthode est qu'elle nécessite de préciser l'abscisse et l'ordonnée de ces points intermédiaires. Pour obtenir une expression semblable à celle du théorème du scrutin, il faudrait donc sommer ce résultat sur toutes les positions P_p acceptables, et de nouveau on se heurte à leur nombre, d'ordre N^2 .

Ponts browniens Dans la dernière méthode présentée ici, c'est une approche récursive qui est envisagée. La subordination semble coûteuse, en terme de temps de calcul, à mettre en œuvre. En prenant le problème dans l'autre sens, on peut se demander quelle est la probabilité qu'une trajectoire complète soit toujours "victorieuse" sachant que sa trajectoire extraite l'est.

La marche aléatoire étudiée ici n'est plus hypergéométrique (au sens d'une somme contrainte de variables aléatoires i.i.d. de Bernoulli) ni bidimensionnelle mais une marche aléatoire symétrique sur réseau, pour laquelle chaque unité ajoute ± 1 à la somme construite par les unités précédentes. Cette représentation est de nouveau équivalente à l'originale comme l'illustre le schéma suivant.

On considère que la marche \mathcal{Y} est un échantillon de la marche \mathcal{D} aux instants d'observation $\{t_1, \dots, t_n\}$. On note " $\mathcal{Y}_n > 0$ " l'événement

$$\forall p \in \llbracket 1, n \rrbracket, Y_p = D_{t_p} > 0 \quad (4.35)$$

et " $\mathcal{D} > 0$ " l'événement

$$\forall t \in \llbracket 1, N \rrbracket, D_t > 0 \quad (4.36)$$

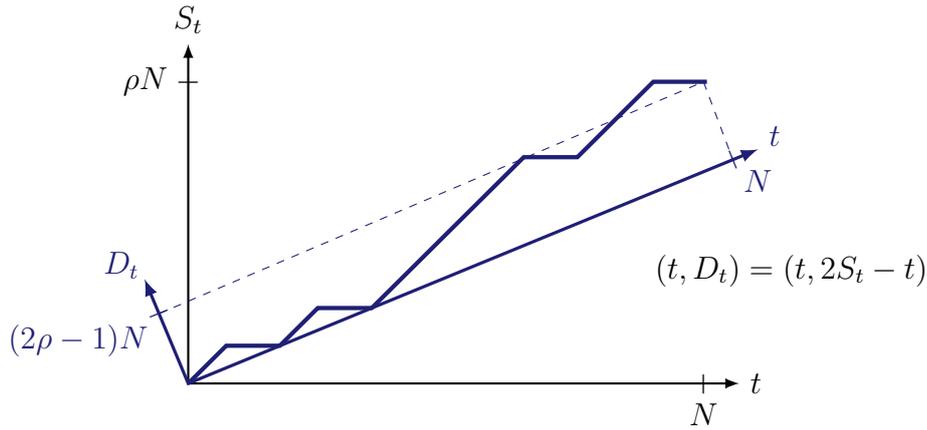


FIGURE 4.13 Représentation schématique de l'équivalence entre la marche hypergéométrique \mathcal{S} et la marche aléatoire \mathcal{D} sur \mathbb{Z} .

Par le théorème du scrutin, on sait que $P(\mathcal{D} > 0) = 2\rho - 1$. De plus, par construction, $P(\mathcal{Y}_n > 0 \mid \mathcal{D} > 0) = 1$. Les propriétés des probabilités conditionnelles donnent donc

$$P(\mathcal{Y}_n > 0) = \frac{2\rho - 1}{P(\mathcal{D} > 0 \mid \mathcal{Y}_n > 0)} \quad (4.37)$$

Déterminer la probabilité au dénominateur nous permettra donc de remonter à la probabilité qu'une marche extraite soit positive. L'événement qu'elle décrit pourrait être formulé de la façon suivante : entre deux points d'observations pour lesquels la trajectoire est positive, la marche aléatoire complète doit demeurer positive. Cette reformulation permet de réduire le problème à une étude entre deux points, c'est-à-dire à étudier des *ponts*.

Pour toute marche aléatoire *symétrique* \mathcal{X} vérifiant $X_0 = 0$, et pour tout entier t , on peut montrer que

$$P_0(X_t > 0) = \frac{1}{2} \quad (4.38)$$

On considère une copie \mathcal{D} de ce processus dont le point final est aussi fixé ($D_N = x > 0$). La probabilité précédente devient

$$P_0(D_t > 0) = P_0(X_t > 0 \mid X_N = x) \quad (4.39)$$

On peut donner plusieurs résultats de dénombrement pour une marche aléatoire *discrète* (sur réseau, avec des incréments ± 1 ici).

On note $t = 2i$ ou $2i + 1$ selon sa parité, $x = 2j$ ou $2j + 1$. Pour tout entier

$0 < t < N$,

$$P_0(D_{N-t} > 0) = \begin{cases} \frac{1}{2^t} \sum_{k=0}^{i+j-1} \binom{t}{k} & x \leq t \text{ et } x, t \text{ sont pairs.} \\ \frac{1}{2^t} \sum_{k=0}^{i+j} \binom{t}{k} & x \leq t, \text{ sinon} \\ 1 & x > t \end{cases} \quad (4.40)$$

Si à présent on cherche à échantillonner cette marche aléatoire à un instant d'observation τ choisi aléatoirement avec une certaine distribution, on obtient

$$P_0(D_\tau > 0) = \sum_{t=1}^{\infty} P(\tau = t) P_0(D_t > 0) \quad (4.41)$$

Pour simplifier les calculs, on peut tenter la même approche en considérant la version continue de cette marche aléatoire symétrique : le mouvement brownien unidimensionnel. Il s'agit d'un processus aléatoire \mathbf{B} satisfaisant l'équation de Langevin

$$\frac{d}{dt}x(t) = \eta(t) \quad (4.42)$$

où η est un bruit blanc, c'est-à-dire un processus stochastique gaussien vérifiant $\langle \eta(t) \rangle = 0$ et $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$.

Le propagateur de \mathbf{B} est alors

$$P(B_t = x \mid B_{t_0} = x_0) = \frac{e^{-x^2/2t^2}}{\sqrt{2\pi t}} \quad (4.43)$$

Un processus brownien dont on fixe le point de départ et le point d'arrivée ($B_0 = 0$, $B_T = x$) est appelé un *pont brownien*. Un processus brownien dont on fixe le point de départ et auquel on impose de rester positif est appelé un *méandre brownien* ($B_0 = 0$, $B_{0 < t < T} > 0$).

On cherche dans notre cas à conditionner un processus positif par l'événement $\{B_0 = 0, B_T = x\}$. On parle dans ce cas de *méandre contraint*⁸³.

$$P_0(W_t > 0) = \int_0^{\infty} P_0(B_t = y \mid B_T = x) dy \quad (4.44)$$

À l'aide du propagateur (4.43) et de la formule de Bayes, on peut établir que

$$P_0(B_t = y \mid B_T = x) = \frac{e^{x^2/2T} \sqrt{T}}{\sqrt{2\pi t(T-t)}} \exp\left(-\frac{(x-y)^2}{2(T-t)}\right) \exp\left(-\frac{y^2}{2t}\right) \quad (4.45)$$

⁸³Avec ces définitions empruntées à J. Pitman [137], le pont brownien standard correspond au cas $x = 0$ et l'excursion brownienne standard est un méandre contraint à revenir en $x = 0$.

et ainsi,

$$P_0(W_t > 0) = \frac{1}{2} + \operatorname{sgn}(x) \operatorname{erf} \left(|x| \sqrt{\frac{t}{2T(T-t)}} \right) \quad (4.46)$$

où on note erf la fonction erreur

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (4.47)$$

De là, l'équation (4.41) pour un instant d'observation choisi aléatoirement devient

$$P(W_\tau > 0) = \int_0^T f(\tau = t) P_0(W_t > 0) dt \quad (4.48)$$

Si l'on suppose que cet instant est tiré uniformément sur l'intervalle $]0, T]$, et que $x \geq 0$,

$$P(W_\tau > 0) = \frac{1}{2} + \frac{1}{T} \int_0^T \operatorname{erf} \left(x \sqrt{\frac{t}{2T(T-t)}} \right) dt \quad (4.49)$$

$$= \frac{1}{2} + \frac{\sqrt{\pi}}{2} \frac{x}{\sqrt{2T}} e^{x^2/2T} \left(1 - \operatorname{erf} \left(\frac{x}{\sqrt{2T}} \right) \right) \quad (4.50)$$

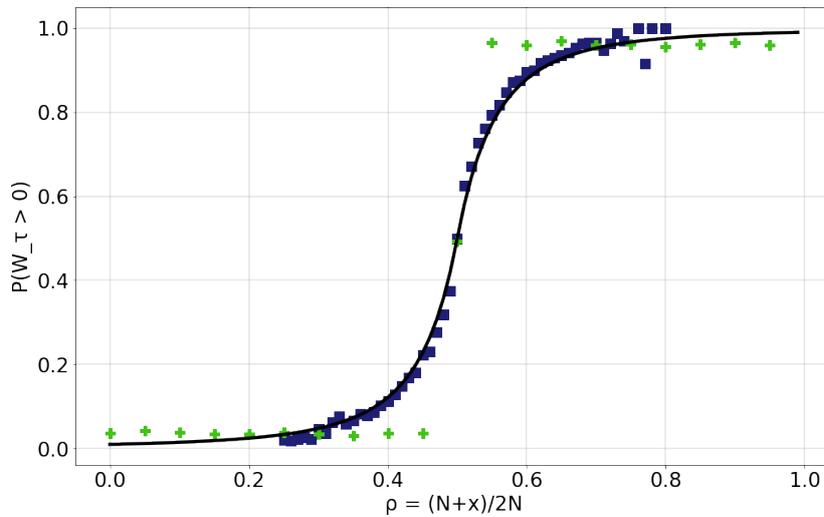


FIGURE 4.14 Simulations de $P(W_\tau > 0)$ (■) pour x dans l'intervalle $[0, N]$, comparaison avec $P(L_\tau > 0)$ (+) où \mathcal{L} est un processus de Cauchy de paramètres $(0, 10^{-3})$, et comparaison avec l'ajustement brownien (en trait plein).

Ajouter un point d'observation supplémentaire entraîne des difficultés calculatoires. Cette méthode ne permet donc toujours pas de résoudre le problème de pré-agrégation des données mises à disposition par les institutions chargées du recensement. La même approche pour un processus de Cauchy ne permet a priori même pas d'obtenir une formule pour un seul point d'observation. Toutefois, elle est facile à simuler, se référer à

la figure **4.14**. On peut remarquer sur cette même figure la symétrie par rapport à $\rho = 1/2$. Cette symétrie est brisée lorsque le nombre de points d'observation tend vers N , en effet le théorème du scrutin prévoit

$$P(D_n > 0, \forall n < N \mid D_N = 2\rho - 1) = \begin{cases} 0 & \rho \leq 1/2 \\ 2\rho - 1 & \rho > 1/2 \end{cases} \quad (4.51)$$

Il y a donc une brisure de symétrie au cours du processus d'échantillonnage, entre 1 et N points d'observations, qu'il serait intéressant d'aller explorer.

4.4 | Discussion

Au cours de tout ce chapitre, nous avons utilisé l'agrégation d'unités statistiques comme une opportunité d'analyser les disparités spatiales, en construisant ces trajectoires et ces distances focales qui mettent en lumière des structures de ségrégation qu'une simple observation des données ne suffit pas à appréhender. Le calcul des distances focales n'est qu'une des informations que l'on peut tirer de ces trajectoires individuelles, une façon d'estimer la transition entre spécificité locale et homogénéisation globale. Dans l'exemple de Paris, les structures mises en évidence par leur calcul et leur représentation graphique correspondent aux citadelles aisées recensées par les sociologues et les géographes. Elle peut donc tout à fait constituer un outil de mesure complémentaire aux enquêtes statistiques.

Précisons ici que les applications des distances focales sont plus générales que la seule étude de la ségrégation, il s'agit d'une manière de visualiser les disparités s'appuyant sur une construction simple à réaliser et riche d'enseignements. La part d'arbitraire qui y subsiste, introduite par le choix d'une définition pour la convergence des trajectoires, a été discutée dans des travaux récents [30, 121, 122]. Nous espérons que notre contribution en la matière fera l'objet d'autres développements, tant en sciences des données pour les questions théoriques et techniques soulevées par sa mise en œuvre qu'en sciences humaines et sociales, comme un outil d'analyse parmi d'autres.

Dans la perspective d'en faire un outil de mesure et de comparaison, nous avons proposé la construction d'une référence commune reposant sur la ville uniforme, c'est-à-dire un système dans lequel l'effet que l'on cherche à mesurer, ici la ségrégation, n'est pas présent. La détermination d'un système témoin, dépendant de la structure géographique de la ville, permettrait alors de ne plus comparer simplement les unités statistiques d'une ville donnée mais plusieurs villes entre elles, chacune associée à sa référence. Ce calcul repose sur une collecte fine des données, à l'échelle individuelle, ce qui est envisageable en Europe du Nord ou aux États-Unis d'Amérique mais qui pose des difficultés pratiques dans le cas de Paris par exemples, où les données sont pré-agrégées.

Conclusion

AU cours de cette thèse, nous nous sommes attachés à étudier la ségrégation sociospatiale selon deux axes de recherche complémentaires : la modélisation de son émergence et l'étude statistique de ses structures géographiques.

L'émergence de la ségrégation dans le modèle de Schelling a fait l'objet de dizaines de milliers de travaux, dont nous avons proposé une ample lecture analytique [12, 70, 71, 74, 101, 133, 138, 171, 172, 176] qui nous est rapidement apparue indispensable. En effet, la dispersion entre plusieurs disciplines, conjuguée à l'absence de schéma commun, rendait même difficile de savoir ne serait-ce quels résultats exactement avaient été établis. Ce travail est en grande partie présenté dans le chapitre **1**. Nous nous sommes efforcés d'uniformiser les énoncés pour démontrer la robustesse du modèle, des états stationnaires qui le caractérisent et des transitions séparant ces états stationnaires. En particulier, nous avons passé en revue ceux démontrant l'existence d'une phase ségrégée.

Dans la lignée de l'article de T. Rogers et A. McKane [146], le chapitre **2** introduit le formalisme markovien, dont les limites sont discutées, et qui permet de préciser cette observation. En déterminant des propriétés de l'espace d'états du système et en calculant les probabilités de transition d'une chaîne de Markov plus proche du modèle original que celle étudiée par T. Rogers et A. McKane, nous avons pu déterminer la valeur critique T_f caractérisant la transition d'un état stationnaire "gelé" dans une configuration proche de l'état initial à un état stationnaire complètement ségrégé.

Cette méthode permet également de déterminer numériquement la distribution invariante de la matrice de transition et de la comparer avec succès aux simulations présentées dans le premier chapitre. La seconde transition est également remarquée, mais ne peut être expliquée analytiquement. Au contraire, elle permet de confirmer notre intuition que son existence est sûrement due à la métastabilité de configurations proche de l'état initial. Cette piste est étudiée dans le chapitre **3**.

Auparavant, grâce à l'approche markovienne, nous affirmons le caractère approximatif des analogies trop rapides entre le modèle de Schelling et

les systèmes de spins, spécifiquement la possibilité d'introduire de l'agitation thermique pour fluidifier la dynamique. Cette variante fait basculer le système dans un nouvel ensemble statistique ; l'équilibre déterminé dans ce formalisme et l'équilibre "réel" peuvent alors être très proches comme très éloignés dans l'espace d'états. Le calcul de l'entropie et de la température du modèle de Schelling permet enfin de le rapprocher des systèmes de spins à contraintes cinétiques. Cette analogie plus rigoureuse que nous établissons pourrait ouvrir l'accès à de nouveaux résultats pour caractériser le modèle.

À la suite de ces considérations thermodynamiques, nous poursuivons la discussion sur la stabilité de structures caractérisant une ségrégation macroscopique en présentant des travaux [15, 88, 124, 125, 127] faisant intervenir des grands voisinages et sans échange d'agents pour déterminer la tolérance T_c associée à la seconde transition ségrégation/mixité. Nous introduisons alors une nouvelle analogie avec la nucléation homogène pour dériver une autre expression de cette même tolérance critique à partir de considérations énergétiques, ceci d'une manière plus simple, peut-être aussi plus intuitive, et réutilisant certains résultats démontrés au chapitre précédent.

Nous pouvons alors conclure que la ségrégation apparaît dans notre variante par échanges du modèle de Schelling lorsque

$$T_f \leq T < T_c \tag{1}$$

où T_f et T_c sont respectivement définies par les expressions (2.70) et (3.78).

Ce résultat établi pour la ségrégation dans le modèle de Schelling, nous nous sommes tournés vers les données réelles dans le chapitre **4**. Après être revenus sur les définitions et les méthodes de mesure de la ségrégation urbaine, nous avons proposé une nouvelle méthode de mesure des disparités sociospatiales reposant sur un processus d'agrégation d'unités statistiques. Ce processus fournit un faisceau de trajectoires duquel on peut extraire tant des propriétés de la ville entière que des traces de la singularité de chaque unité individuelle. Une de ces caractéristiques individuelles est la distance focale, qui illustre visuellement dans l'exemple de Paris des constats établis par les sciences sociales : agrégation affinitaire des populations aisées, mixité historique des grands boulevards. La méthode présentée dans ce chapitre s'inscrit dans une démarche de développement d'outils mathématiques pour les sciences sociales capables de capter la complexité des systèmes étu-

diés. Dans le même esprit, nous construisons un “modèle de référence”, correspondant à une ville “bien mélangée”, afin de fournir une hypothèse nulle pour les mesures statistiques et permettre de comparer différents systèmes entre eux. Sa mise en application dans un contexte où les données sont pré-agrégées pose un problème non trivial d'échantillonnage de processus stochastiques pour lequel nous proposons des pistes de résolution.

Suite à ce travail, les perspectives de recherche sont nombreuses, voire déjà engagées [30, 121, 122] : explorer d'autres facettes du faisceau de trajectoires, consolider le modèle de référence... Le dialogue avec des chercheurs et chercheuses en sciences sociales semble essentiel pour avancer sur ces questionnements.

Quant au modèle de Schelling, on peut compter en dizaines, si ce n'est en centaines, le nombre d'articles publiés ou prépubliés pendant la simple rédaction de ce manuscrit, traitant de diverses variantes et altérations au modèle original. S'il nous a semblé important de comprendre rigoureusement le comportement d'un modèle au plus proche de celui proposé par T. Schelling en 1971, il va de soi qu'ajouts et améliorations sont toujours envisageables. Les liens entre modèle de Schelling, nucléation et verres de spins nous semblent une de ces perspectives, une branche à étudier dans la littérature des systèmes complexes.

Annexes

Annexe A | Stabilité stochastique

Cette annexe reprend les notions introduites par Young [176] avec les notations de cette thèse. Il peut être utile d'avoir lu la section **2.1.1**.

Soit une chaîne de Markov de matrice de transition $W = (w_{nn'})_{n,n' \in E}$. On peut perturber ce processus en altérant cette matrice W pour construire la matrice $W(\beta)$. Le processus en résultant est appelé chaîne de Markov perturbative si et seulement si

- il est irréductible pour tout β dans un voisinage de $+\infty$;
- $\forall n, n', \lim_{\beta \rightarrow \infty} w_{nn'}(\beta) = w_{nn'}$;
- $\forall n, n', \exists r_{nn'} \in \bar{\mathbb{R}}, 0 < \lim_{\beta \rightarrow \infty} (\beta^{-r_{nn'}} w_{nn'}(\beta)) < \infty$.

Ces taux $r_{nn'}$ sont les résistances des transitions $n \rightarrow n'$, une forme de taux de convergence des probabilités avec perturbation vers les probabilités non-perturbées. Lorsque $w_{nn'} \neq 0$, la résistance associée est nulle.

La chaîne perturbative étant finie et irréductible pour β au voisinage de $+\infty$, elle possède une distribution invariante $\lambda(\beta)$. Alors pour tout état n ,

$$\lim_{\beta \rightarrow \infty} \lambda_n(\beta) = \lambda_n^* \tag{A.1}$$

existe, et $\lambda^* = (\lambda_n^*)_{n \in E}$ est une distribution invariante de la chaîne de Markov non-perturbée.

Définition [176] *Un état n est dit stable stochastiquement lorsque $\lambda_n^* > 0$.*

Il est utile de noter que ce théorème et la définition qui en découle ne font nullement état de l'*unicité* de cette distribution invariante λ^* . La chaîne de Markov initiale peut tout à fait ne pas être irréductible, la distribution λ^* qualifie alors un équilibre possible, qu'il peut être tentant de rapprocher, dans le cas de systèmes pour lesquels la comparaison est appropriée et lorsque la perturbation s'apparente à l'effet d'une agitation thermique, de l'équilibre *thermodynamique*.

Annexe B | Incrément de la densité d'interface

En écrivant \sum_i pour condenser $\sum_{i=1}^N$ et $\sum_{i \neq a,b}$ pour signaler que la somme porte sur tous les sites hors des voisinages de a et de b (a et b inclus),

$$\begin{aligned} \sum_i \tilde{m}_i &= \sum_{i \neq a,b} m_i + s_b M_a + s_a M_b + \sum_{i \in \partial a} m_i + s_i (s_b - s_a) + \sum_{i \in \partial b} m_i + s_i (s_a - s_b) \\ &= \sum_i m_i - m_a - m_b + s_b M_a + s_a M_b + (s_b - s_a) \sum_{i \in \partial a} s_i + (s_a - s_b) \sum_{i \in \partial b} s_i \\ &= \sum_i m_i - 2(m_a + m_b) + 2(s_a M_b + s_b M_a) \end{aligned} \quad (\text{B.1})$$

$$\begin{aligned} \sum_i \tilde{\kappa}_i &= \sum_{i \neq a,b} \kappa_i + s_b^2 K_a + s_a^2 K_b + \sum_{i \in \partial a} \kappa_i + s_i^2 (s_b^2 - s_a^2) + \sum_{i \in \partial b} \kappa_i + s_i^2 (s_a^2 - s_b^2) \\ &= \sum_i \kappa_i - \kappa_a - \kappa_b + s_b^2 K_a + s_a^2 K_b + (s_b^2 - s_a^2) \sum_{i \in \partial a} s_i^2 + (s_a^2 - s_b^2) \sum_{i \in \partial b} s_i^2 \\ &= \sum_i \kappa_i - 2(\kappa_a + \kappa_b) + 2(s_a^2 K_b - s_b^2 K_a) \end{aligned} \quad (\text{B.2})$$

On obtient alors l'équation (2.28).

Annexe C | Taille de l'espace d'états de la chaîne de Markov

$$\begin{aligned} \mathcal{N} &\leq \sum_{n=n_{sc}}^{\kappa N d} \left(\left\lfloor \frac{n}{4} \right\rfloor + 1 \right) \leq \sum_{n=\kappa N(2d-1)}^{\kappa N d} \left(\frac{n}{4} + 1 \right) \\ &= \frac{1}{8} \left(\kappa N d (\kappa N d + 1) - \kappa N d (2d - 1) (\kappa N d (2d - 1) - 1) \right) \end{aligned} \quad (\text{C.1})$$

$$+ \kappa N d - \kappa N (2d - 1) + 1 \quad (\text{C.2})$$

$$= (3d - 1)(1 - d) \frac{(\kappa N)^2}{8} + \frac{7 - 5d}{8} \kappa N + 1 \quad (\text{C.3})$$

On obtient alors l'équation (2.39).

Annexe D | Présence de sites vacants dans un voisinage

Dans le modèle de L. Gauvin et al. [70], la probabilité de trouver un voisinage de Moore contenant k sites vacants dans un système de densité d est

$$\mathcal{P}(k \text{ sites vacants}) = \frac{\binom{dN}{8-k} \binom{(1-d)N}{k}}{\binom{N}{8}} \quad (\text{D.1})$$

C'est une loi hypergéométrique de paramètres $(8, 1 - d, N)$ et son espérance est alors

$$\mathbb{E}(k) = 8(1 - d) \quad (\text{D.2})$$

soit $\mathbb{E}(k) = 1$ lorsque $d = 7/8 \simeq 0.88$, ce qui correspond à une des courbes de la figure **2.13**.

Annexe E | Entropie microcanonique

À l'initialisation du système, $W = 3^N$ configurations sont à priori accessibles. Construire la configuration initiale σ_0 est équivalent à un tirage aléatoire successif avec remise dans $\{-1, 0, 1\}^N$.

Au cours de la dynamique de Kawasaki, le mécanisme d'évolution est basé sur l'échange à intervalles réguliers des contenus de deux sites choisis aléatoirement. Cette évolution possède deux invariants indépendants, la densité $d \in [0, 1]$ et la parité $p \in [0, 1]$. Une fois l'état initial fixé tous les échanges effectués dans la dynamique de Kawasaki laissent ces deux quantités invariantes. À partir d'une configuration σ_0 donnée, il existe alors W_0 configurations accessibles

$$W_0 = \binom{N}{dN} \binom{dN}{dpN} \quad (\text{E.1})$$

Dans un système paritaire et sans site vide ($d = 1$ et $p = 1/2$), on a alors

$$W_0 = \binom{N}{N/2} \quad (\text{E.2})$$

et donc, avec l'expression (2.47)

$$\begin{aligned} \tilde{W}(x) &= W_0 \times \mathcal{P}\left(\sum_i m_i = (1 - 2x)\kappa N\right) \\ &= \binom{N}{N/2} \times 2^{-\kappa N} \binom{\kappa N}{\kappa N(1 - x)} \end{aligned} \quad (\text{E.3})$$

que l'on peut réécrire, en utilisant l'identité $\binom{m}{n} = \binom{m}{m-n}$, sous la forme de l'équation (3.39).

L'entropie est ensuite simplement égale au logarithme de $\tilde{W}(x)$, et sa dérivée par rapport à x est

$$\frac{d\tilde{S}}{dx}(x) = \kappa N \left(\psi(\kappa N(1 - x) + 1) - \psi(\kappa N x + 1) \right) \quad (\text{E.4})$$

où ψ est la fonction digamma définie sur \mathbb{C} par $\psi(z) = \Gamma'(z)/\Gamma(z)$. Les propriétés de récurrence de ces fonctions restreintes à l'ensemble des entiers naturels donnent

$$\psi(n + 1) = -\gamma + H_n \quad (\text{E.5})$$

ce qui permet de trouver

$$\frac{d\tilde{S}}{dx}(x) = \kappa N (H_{\kappa N(1-x)} - H_{\kappa N x}) \quad (\text{E.6})$$

où H_n est le n -ième nombre harmonique :

$$H_n = \sum_{k=1}^n \frac{1}{k} \quad (\text{E.7})$$

et donc, par composition,

$$\Theta = 2\kappa N \times \frac{1}{\kappa N (H_{\kappa N(1-x)} - H_{\kappa N x})} \quad (\text{E.8})$$

On obtient alors l'équation (3.41).

Annexe F | Interfaces et rugosité

On se place dans le cadre du voisinage de Moore. Pour déterminer la rugosité d'un noyau homogène, on utilise le théorème de Pick [135] sur le graphe dual. L'aire du polygone exactement compris sous l'interface interne $\partial\Omega$ est alors

$$A = \Omega^2 - \partial\Omega + \frac{1}{2}\partial\Omega - 1 \quad (\text{F.1})$$

En effet, le nombre de sites strictement inclus dans ce polygone est le nombre de sites au cœur du noyau, tandis que ses sommets sont exactement les sites interfaciaux. Cette aire peut être déterminée d'une autre façon, en dénombrant les sites interfaciaux "rectilignes", "concaves" et "convexes". On peut montrer que

$$A = (\Omega^2 - \partial\Omega) + \frac{1}{2}r + \frac{1}{4}cv + \frac{3}{4}cc \quad (\text{F.2})$$

Alors

$$\partial\Omega = r + \frac{1}{2}cv + \frac{3}{2}cc + 2 \quad (\text{F.3})$$

L'intérêt de cette méthode est la facilité avec laquelle on peut calculer $\partial\Omega_{\text{ext}}$ à partir des grandeurs utilisées précédemment, il suffit de considérer le polygone englobant entièrement le noyau, on obtient alors de la même façon

$$A' = \Omega^2 + \frac{1}{2}\partial\Omega_{\text{ext}} - 1 \quad (\text{F.4})$$

$$= \Omega^2 + \frac{1}{2}r + \frac{5}{4}cv - \frac{1}{4}cc \quad (\text{F.5})$$

D'où

$$\partial\Omega_{\text{ext}} = r + \frac{5}{2}cv - \frac{1}{2}cc + 2 \quad (\text{F.6})$$

Un simple argument géométrique suffit alors à déterminer une relation entre ces deux grandeurs. Pour un noyau de taille minimum 2×2 , il existe au minimum 4 coins convexes. Plus précisément s'il existe $cc = n$ coins concaves, alors $cv = 4 + n$. On a donc

$$\partial\Omega_{\text{ext}} = \partial\Omega + 8 \quad (\text{F.7})$$

En recherchant d'autres exemples, on s'aperçoit qu'il ne reste alors que les noyaux linéaires d'épaisseur 1, pour lesquels on trouve directement

$$\partial\Omega_{\text{ext}} = 2\partial\Omega + 6 \quad (\text{F.8})$$

Enfin, le cas le plus extrême est celui du noyau de taille 1, pour lequel $\partial\Omega = \Omega^2 = 1$ et $\partial\Omega_{\text{ext}} = \kappa$, ainsi

$$\partial\Omega_{\text{ext}} = 8\partial\Omega \quad (\text{F.9})$$

On obtient l'équation (3.69) d'où on postule

$$\partial\Omega_{\text{ext}} = \ell\partial\Omega + k \quad (\text{F.10})$$

Annexe G | Marche aléatoire bidirectionnelle

Les coordonnées des points utiles à l'étude sont $O = (0, 0)$, $P = (i, j)$ et $I = (i, i)$. Il existe $c_{i,j} = (i+j)!/(i!j!)$ chemins reliant O et P qui se partagent en $\phi_{i,j}$ chemins se trouvant toujours strictement au-dessus de la diagonale et $x_{i,j}$ chemins passant par au moins un point K de la diagonale (OI). On considère que ce point K est celui de premier contact (éventuellement le seul). Alors

$$x_{i,j} = c_{i,j} - \phi_{i,j} = \sum_{k=1}^i \phi_{k,k}^* c_{i-k,j-k} \quad (\text{G.1})$$

Dans le cas particulier $i = j$ (on note dans ce cas que les chemins ne touchant pas la diagonale peuvent être soit strictement au-dessus soit strictement en-dessous de la diagonale, d'où la notation différente : $\phi_{i,i}^* = 2\phi_{i,i}$),

$$c_{i,i} - \phi_{i,i}^* = \sum_{k=1}^i \phi_{k,k}^* c_{i-k,i-k} \quad (\text{G.2})$$

et, par récurrence

$$\phi_{i,i}^* = \frac{2}{i} c_{i-1,i-1} \quad (\text{G.3})$$

Notons que le résultat est différent de [174], car $x_{i,j}$ y compte les chemins *couplant* la diagonale, pas ceux qui restent au-dessus sauf en un point.

On peut alors combiner les équations (G.1) et (G.3) :

$$\phi_{i,j} = c_{i,j} - 2 \sum_{k=1}^i \frac{c_{k-1,k-1} c_{i-k,j-k}}{k} \quad (\text{G.4})$$

et après un peu de dénombrement (preuve dans l'article 7 [174]),

$$\phi_{i,j} = c_{i,j} - 2c_{i-1,j} = \frac{j-i}{i+j} c_{i,j} \quad (\text{G.5})$$

On retrouve bien l'équation (4.30).

Annexe H | Schelling : un modèle trivial ?

Depuis un premier preprint de 2006, sobrement intitulé “La ségrégation spatiale selon Schelling : la perversité est ailleurs” et jusqu’au numéro spécial Schelling de la revue *Émulations* en 2019, M. Forsé et M. Parodi [60–62], deux sociologues français, cherchent à démontrer la trivialité du modèle de Schelling et surtout son inutilité pour comprendre le phénomène de la ségrégation, les contraintes appliquées étant selon eux trop fortes et “irréalistes”.

La ségrégation urbaine a de nombreuses causes, qu’il est plus urgent d’étudier que la fiction d’un monde ségrégué “malgré nous”.

M. Forsé et M. Parodi, 2006 (conclusion de l’abstract)

A. Kirman a déjà commenté avec amusement [98] les arguments avancés par les deux sociologues, nous reviendrons sur une partie d’entre eux et compléterons sa lecture à l’aide de la revue établie au premier chapitre et de notions de physique abordées dans le chapitre précédent.

Tout d’abord, la (trop grande) simplicité du modèle. Les auteurs reprochent à T. Schelling de négliger bien trop de contraintes structurelles et insistent sur l’impossibilité de modéliser un phénomène complexe tel que la ségrégation à partir d’un unique facteur (l’attachement des agents à être majoritairement entourés de voisins du même groupe qu’eux). “La ségrégation urbaine”, avancent-ils, “a de nombreuses causes politiques, économiques et sociales” et le modèle de Schelling ne prend “pas au sérieux la question de [son] application”⁸⁴. Pourtant, T. Schelling n’a jamais prétendu expliquer la ségrégation comme un simple effet pervers d’une société tolérante. Il se borne à se demander si, toute autre cause possible écartée, une société peut tendre vers la ségrégation et trouve que oui, et ce pour une large gamme de tolérance : une forte intolérance n’est pas une condition nécessaire à l’émergence de la ségrégation. Il est certes possible d’améliorer le modèle, tout en gardant un certain niveau d’abstraction nécessaire à la description schématique d’un aspect commun à des situations très diverses. On peut ainsi agrémenter la dynamique de Schelling par un marché de l’immobilier [178], en y intégrant des lieux de rencontre [156] (centres commerciaux, lieux de travail, cafés, restaurants...) ou encore en introduisant un biais (similaire à un champ magnétique pour les systèmes de spins et pouvant modéliser une forme de discrimination systémique) favorisant un groupe plutôt qu’un

⁸⁴[62], p. 101.

autre [131]. Il ne faut simplement pas perdre de vue l'objectif simple du modèle originel.

Ensuite, une impossibilité physique. À partir de deux postulats, *la ségrégation est un état ordonné et tout système évolue vers son état de plus grand désordre*, les auteurs “démontrent” l'impossibilité pour le modèle de Schelling de converger vers un état ségrégué. Toute expérience montrant le contraire serait un ersatz de certaines conditions dans lesquelles s'effectuent la simulation, et M. Forsé et M. Parodi en relèvent deux.

Premièrement, la présence de sites vacants fausse la tolérance des agents. En faisant diminuer le nombre effectif moyen de voisins par agent, ils augmentent artificiellement le seuil de voisins similaires requis pour être satisfait. Ce biais existe effectivement et il est quantifiable, néanmoins le modèle de Schelling est robuste à des variantes sans sites vacants, on peut citer les variantes par échanges [40, 74, 138], et à fortiori celles à grand voisinage [16, 17, 26]⁸⁵, voire à voisinage continu [172]. Le biais n'influe donc pas de façon dramatique sur les résultats des simulations.

Deuxièmement, la règle de non-déplacement des agents satisfaits serait l'unique raison de la stabilité, voire de la formation, des noyaux homogènes. En introduisant cette règle, Schelling contraindrait le système dans un équilibre métastable factice, un équilibre de ségrégation ne correspondant pas à l'équilibre réel, prévu par le principe de maximisation de l'entropie : la mixité. Encore une fois, ce postulat est invalidé par les travaux utilisant d'autres règles pour l'utilité, notamment ceux dans lesquels tous les agents sont autorisés à se déplacer pour maximiser leur utilité[138].

Ce que la “démonstration” initiale et ce dernier argument montrent, c'est l'incompréhension par les auteurs des notions d'entropie, d'ordre et d'équilibre présentées dans le chapitre précédent. Tout d'abord, la ségrégation n'est pas le seul état “ordonné” que peut prendre le système. Une configuration en échiquier, par exemple, est tout aussi ordonnée, pourtant c'est la configuration la plus intégrée possible⁸⁶. Le calcul rigoureux, en section **3.2**, de l'entropie en fonction de la densité d'interface montre d'ailleurs la symétrie des situations d'intégration maximale et de ségrégation maximale. D'autre part, le principe de maximisation de l'entropie est aussi mal com-

⁸⁵L'échantillonnage des valeurs de la tolérance contraint les estimations des seuils T_c et T_f à des intervalles de taille $1/\kappa$, ces valeurs sont donc d'autant plus précises que κ est grand.

⁸⁶Pour un voisinage de von Neumann la densité d'interface est même de 1.

pris : l'équilibre qu'il fournit est un équilibre *thermodynamique*. Or, comme on l'a expliqué plus haut, le système subit une dynamique *cinétique*, qui peut le contraindre à demeurer dans un équilibre métastable, loin du précédent. Les auteurs ne semblent pas se rendre compte que leur argumentation contredit des dizaines de travaux et de résultats bien établis, notamment sur l'auto-organisation des systèmes complexes [10, 50, 120].

Enfin, et cette correction provient de la théorie des jeux, les auteurs semblent confondre l'optimisation par les agents de leur utilité individuelle et l'obtention d'un état final ségrégé qui serait donc désiré par ces mêmes agents. Dans la construction du modèle, les agents n'ont aucune préférence *globale*, ni pour la ségrégation ni pour l'intégration : toute la dynamique ne résulte que des préférences *locales*. On est dans une situation qui pourrait alors rappeler le dilemme du prisonnier, c'est particulièrement flagrant pour des utilités piquées⁸⁷ : des agents qui souhaitent tous avoir un voisinage mixte se retrouvent dans une situation ségrégée, certes plus favorable qu'une configuration en échiquier, où chaque voisinage est entièrement constitué d'agents de l'autre groupe, mais tout de même éloignée de leur utilité maximale. On retrouve ici la description de l'état final comme un "équilibre de Nash myope".

En conclusion, s'il est certain que le développement de l'étude des systèmes complexes doit être transdisciplinaire, et qu'à ce titre il est heureux de voir des chercheurs et chercheuses d'une discipline se saisir de modèles jusque là étudiés par d'autres, il serait sans doute plus profitable scientifiquement parlant de le faire de manière moins péremptoire. Il est facile en venant de la physique statistique de reprendre et de corriger une partie des arguments de cet exemple en particulier, mais le problème existe évidemment aussi dans l'autre sens, nombre de physiciens pensant pouvoir expliquer *in extenso* des phénomènes socio-économiques complexes à partir de modèles simplistes sans se donner la peine de chercher à comprendre les avancées des sciences sociales sur ces phénomènes.

⁸⁷Se référer au paragraphe *Utilité : mesurer la satisfaction* en section **1.1.3**.

From urban segregation to spatial structure detection

Environment and Planning B: Urban Analytics
and City Science
0(0) 1–17

© The Author(s) 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/2399808318797129

journals.sagepub.com/home/epb



Julien Randon-Furling , **Madalina Olteanu**
and **Antoine Lucquiaud**

Université Paris I Panthéon-Sorbonne, France

Abstract

We develop a ‘multifocal’ approach to reveal spatial dissimilarities in cities, from the most local scale to the metropolitan one. Think, for instance, of a statistical variable that may be measured at different scales, e.g. ethnic group proportions, social housing rate, income distribution, or public transportation network density. Then, to any point in the city there corresponds a sequence of values for the variable, as one zooms out around the starting point, all the way up to the whole city – as if with a varifocal camera lens. The sequences thus produced encode spatial dissimilarities in a precise manner: how much they differ from perfectly random sequences is indeed a signature of the underlying spatial structure. We introduce here a mathematical framework that allows to analyse this signature, and we provide a number of illustrative examples.

Keywords

Segregation, dissimilarities, pattern detection, random sequences, Ballot Theorem

Introduction

Cities may be regarded as an epitome of complex systems (Batty, 2007): individual and institutional agents interact on multiple levels of numerous networks (both physical and virtual), leading to nontrivial collective behaviour and nontrivial patterns at many scales. Among a number of intricate questions that have emerged in the study of urban systems, that of sociospatial dissimilarities stands out as one that has aroused interest for decades – and across a wide range of fields thanks to, especially, the pioneering multi-agent models of segregation developed by Schelling (1971) and Sakoda (1971). Urban segregation phenomena will serve here as archetypal examples to illustrate the mathematical methods and tools we are presenting.

Corresponding author:

Julien Randon-Furling, Université Paris I Panthéon-Sorbonne, 90 rue de Tolbiac, Paris 75013, France.

Email: j.randon-furling@cantab.net

Segregation is often simply perceived as spatial separation of two or more groups, and therefore measured in terms of the relative proportions of each group in the different neighbourhoods of a city. However, segregation is essentially a spatial and multiscalar phenomenon, as pointed out, for instance, by Leckie et al. (2012), Östh et al. (2015), and Louf and Barthelemy (2016). An individual perceives segregation all the more acutely as she has to go *a longer way* from her home to discover what the city in its entirety might look like. Imagine the extreme case of a city where two groups *A* and *B* live in total separation, thus forming two ghettos. An individual living at the heart of one of the ghettos would have to explore the whole city to find out that it actually comprises equal proportions of both groups. On the contrary, starting from some parts of the city that are *better mixed*, for instance on the boundary between the two ghettos, there would be no need to cover so large an area to come to the same realization. This basic observation has led us to imagine a mathematical framework that allows to capture and measure spatial dissimilarities as a multiscalar phenomenon across the city.

Consider ever larger neighbourhoods around a starting point or areal unit. As detailed in the first section below, enlarging the area all the way up to the whole city produces a trajectory for each point in the city. That is, the sequence of values taken by the variable under consideration from its value at the most local level around the starting point to its average value at the metropolitan level. For a given starting point, how ‘long’ it takes for the sequence to reach the city’s average thus indicates how ‘distant’ the point is from the city’s global distribution.

Further, the same mapping of points into trajectories allows to easily characterize a null model. In a perfectly mixed city (the ‘uniform’ city), the aggregation process used to build the trajectories would be equivalent to adding units drawn at random (as when drawing balls in an urn without replacement). We introduce in the second section below various ways of quantifying deviations from such random sequences. In both sections, we illustrate our ideas and methods with public data available for the city of Paris.

From bespoke neighbourhoods to trajectories

Data are often available at a given spatial level – for example census blocks in the USA or *Îlots regroupés pour l’information statistique* (IRIS) in France, see US Census Bureau (1994) and INSEE (2017). Aggregating such blocks poses a number of statistical questions, the most famous of which is the so-called modifiable areal unit problem, as stated in Openshaw and Taylor (1979) and Openshaw (1984). Also, in most instances, elementary spatial units will exhibit dissimilarities. Indeed, if the data give the number of medical practitioners, or the quantiles of the income distribution within the unit, then units will generally differ from one another (Glaeser et al., 2001). Such dissimilarities may or may not present spatial patterns. If they do, the geographical system may be said to exhibit spatial segregation, especially if the variables considered correspond to the relative proportions of different population groups (Brown and Chung, 2006; Caridi et al., 2017; Cortese et al., 1976; Cottrell et al., 2017; Cowgill and Cowgill, 1951; Duncan and Duncan, 1955; Farley and Taeuber, 1968; Feitosa et al., 2007, 2012; Fossett, 2017; James and Taeuber, 1985; Massey and Denton, 1988; Reardon and Firebaugh, 2002; Schwartz and Winship, 1980; White, 1983).

Modifiable areal units may actually prove to be a valuable tool to quantify segregation and spatial patterns of dissimilarities, see Leckie et al. (2012), Clark et al. (2015), Jones et al. (2015), Manley et al. (2015), Leckie and Goldstein (2015), Harris and Owen (2017), and Adrestani et al. (2017). Instead of only comparing individual units, one may gain greater

insight into relative spatial differences by taking into account the broader picture: that is, by considering ever larger neighbourhoods around a unit, as if through a zoom lens – see Figure 7 in the Supplemental Material. Evaluating a given statistical variable at each step of the aggregation procedure produces a sequence of values that reflects not only the initial point but also the singularity of its position within the city. The sequence of bespoke neighbourhoods around a starting point acts as a sequence of filters: unveiling step by step the city’s ‘face’. Indeed the final value of all trajectories, whatever the starting point, will be the city’s average for the variable considered. But details of a given trajectory, such as the number of aggregating steps needed to converge to the city’s average, characterize how different and how singular the starting point is from the whole city.

Although relatively easy to describe, the actual computing of trajectories presents a number of difficulties from a statistical point of view. We recall some of them in the following paragraphs.

Defining and computing trajectories

Consider a grid comprising N fundamental spatial units (typically, individuals or the smallest statistical units available). For a given unit, define $G_n(i)$, the cluster formed by itself and its $n - 1$ nearest neighbouring units. Thus, $G_1(i)$ is just the i th unit on its own, while $G_N(i)$ is the whole grid (whichever i from which one starts).

Now suppose one has a statistical variable ξ defined on each element of

$$\mathcal{G} = \{G_n(i), 1 \leq i, n \leq N\}$$

Note that \mathcal{G} is simply the set of all bespoke neighbourhoods, with all possible starting points in the city. Then, define for each starting point i a function f_i such that $f_i(n)$ gives the value of ξ computed on $G_n(i)$.

Interpreting n as an index, one has for each starting unit i a sequence representing the trajectory that takes ξ from its value on unit i to its value on the whole grid. Formally, $(f_i(n))_{1 \leq n \leq N}$ is what we call the trajectory of i for the variable ξ .

Once trajectories have been built from the data available in a given city, a statistical analysis may be carried out in order to:

- identify units that exhibit similar trajectories – this may be done, for instance, using clustering algorithms;
- detect the crossing over between two regimes: the local one and the global one – in the former, ξ may take values significantly distinct from the one on the whole grid; in the latter, it takes values very similar to the city’s average.

Let us emphasize links between the first point above and multiscalar approaches used in recent papers (e.g. Andersson and Malmberg, 2015; Clark et al., 2015; Östh et al., 2015). In particular, Andersson and Malmberg (2015) use vertical slices or cross-sections of the trajectories defined here, i.e. values at certain points only. But looking at full trajectories allows us to examine the second point listed above: the number of aggregated neighbouring blocks needed around a starting block to get close enough to the city’s average value for the variable under consideration. We shall call this the ‘*radius of convergence*’. Geographically this plays the role of a local urban radius: it is a proxy for the area one needs to explore locally to obtain a reasonably good perception of the city as a whole. We give examples of such analysis further in this section.

Constructing trajectories from actual data

One of the statistical subtleties arises upon building datasets so that they are defined on \mathcal{G} , the full set of all possible bespoke neighbourhoods in the city.

Consider indeed the following standard situation: data are available in the form of a size factor s_i (e.g. a number of inhabitants) for each spatial unit, as well as the value of a possibly multi- or even infinite dimensional variable ξ . This could be the number of offices or services of such and such type available in the unit, the social housing rate, an average income, a local density of public transportation, quantiles of a distribution, a full distribution, etc.

When grouping n units, one needs to compute the value of the variable on the new aggregate. This is done easily in the case when ξ is simply a number

$$\xi(G_n(i)) = \sum_{j \in G_n(i)} \xi(j) \quad (1)$$

Similarly, $\xi(G_n(i))$ is readily computed when $\xi(i)$ is a rate or an average

$$\xi(G_n(i)) = \frac{\sum_{j \in G_n(i)} s_j \xi(j)}{\sum_{j \in G_n(i)} s_j} \quad (2)$$

with s_i the population of unit i (or another relevant size factor).

When ξ is a distribution that is known entirely (e.g. one knows the income of every single household in any of the units), one simply aggregates the individual datasets to obtain the dataset for a group of spatial units, and the corresponding empirical distribution is obtained readily.

A more difficult case, alas very frequent, is when ξ retains only certain percentiles of a distribution. For instance

$$\xi(i) = \{2354; 4684; 6546; 8138; 10542; 13622; 17058; 22202; 30862\} \quad (3)$$

are the deciles of the income distribution (in euros per year) for a given IRIS (census block) in the northern part of Paris.

In this case, one is led to either (or both) rely on an Ansatz for the shape of the underlying distribution (e.g. assume that it is log-normal, or exponential or other) or simulate the full dataset (with assumptions on the intradecile distributions), see Stoker (1984), Genest and Zidek (1986), Fotheringham and Wong (1991), Genest (1992), Jelinski and Wu (1996), and Buseti (2014). An example is given in the next section, as we are faced with this problem when we consider income data for the city of Paris, available only in the form of quantiles for each census block.

Lastly, one does not have to be working in the discrete framework of statistical blocks or population count data. For certain variables, such as availability of public transport networks, one may work directly in terms of a local density. For instance, let M be the total number of metro stations in Paris. Let $m_i(l)$ be the number of stations in a disk of radius l centred on IRIS i , $p_i(l)$ the population living in the same disk, and P the total population in Paris. Then define

$$r_i(l) = \frac{m_i(l)/M}{p_i(l)/P} = \frac{m_i(l)/p_i(l)}{M/P} \quad (4)$$

This is similar to the ‘representation’ defined for social classes in Louf and Barthelemy (2016). It quantifies whether the local density of metro stations per inhabitant is smaller ($r_i(l) < 1$), larger ($r_i(l) > 1$), or equal ($r_i(l) = 1$) to the city’s average density. And it allows to build trajectories, for each IRIS, letting l vary from some small value ε to the full radius of the city.

Example: Spatial dissimilarities in Paris

As an illustration, let us work with two types of data available for the city of Paris from France’s census bureau, called INSEE (*‘Institut national de la statistique et des études économiques’*).

Social housing rate. For each IRIS, the number of housing units and the number, among these, of social housing ones are available. One is then in a situation where computing the social housing rate for any group of IRIS is easy and the trajectories can be computed for all IRIS in Paris. Some of these trajectories are shown in Figure 1.

One observes groups of trajectories that tend to start higher or lower than the metropolitan average, and converge to it more or less quickly, obviously with a strong spatial dependency as far as the initial units are concerned.

Let us define the radius of convergence, that is the point where each path enters (and remains) into a given interval (here ± 0.05) around the city’s average social housing rate. Looking at these radii reveals different scales of convergence to the global mean, from one district to another and, inside each district, from one IRIS to another, as can be seen in Figure 2: IRIS blocks from Paris’s 11th district converge much ‘faster’ than those in the eighth district, for instance.

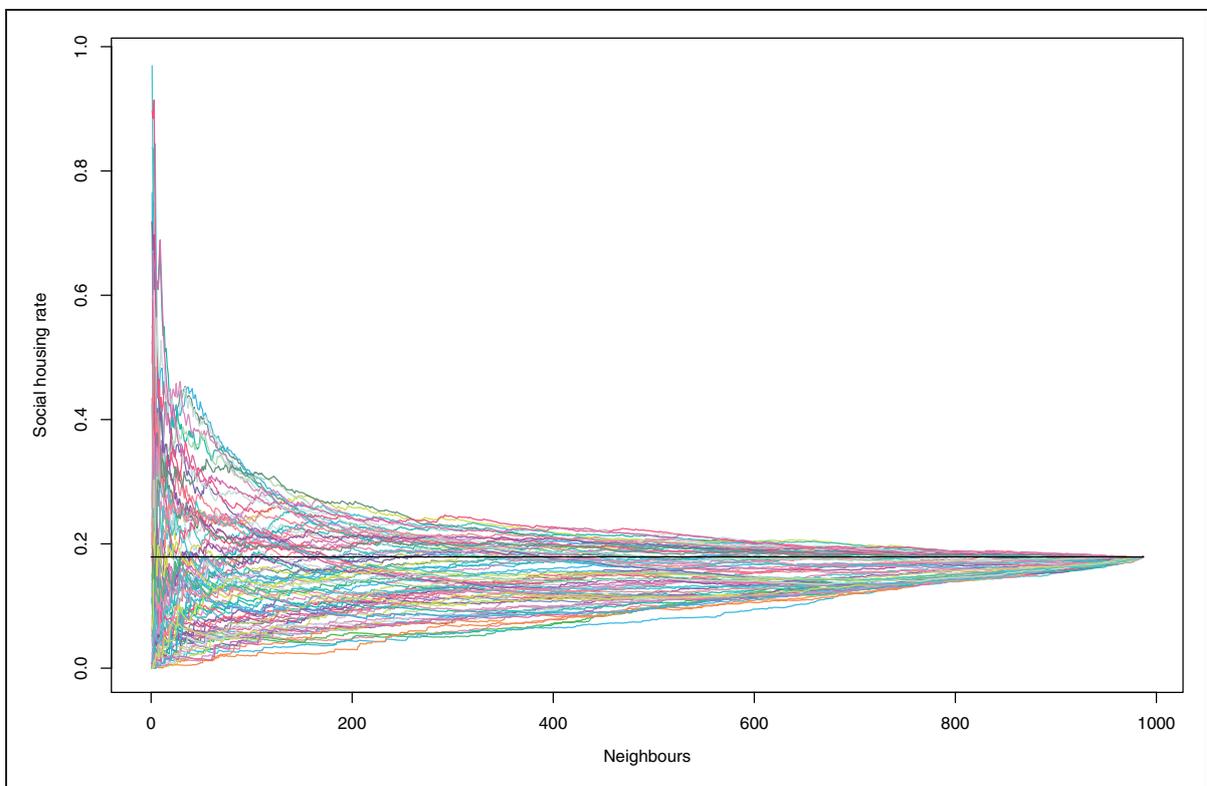


Figure 1. Trajectories for the social housing rate, starting from some (10%) of the 936 statistical units (so-called IRIS) in Paris. The solid line corresponds to the city’s average (17.9%).

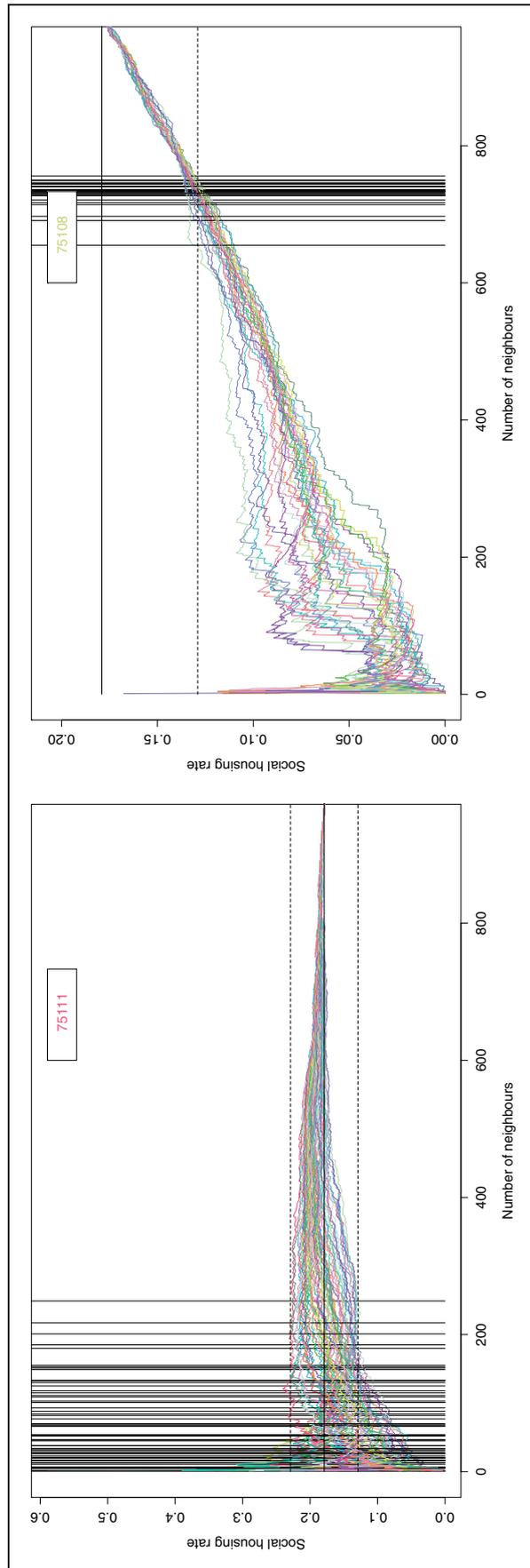


Figure 2. Trajectories for the social housing rate, starting from each IRIS (census block) in Paris's 11th (left) and eighth (right) districts. Colours correspond to each different IRIS taken as starting point. The solid flat line gives the city's average social housing rate (17.9%) and the dashed lines correspond to ± 0.05 around this average. Solid vertical lines correspond to radii of convergence: the radius of convergence of a trajectory is defined as the point where the path last enters the ± 0.05 interval (and therefore remains inside the interval afterwards). One observes that IRIS in the 11th district converges much faster to the city's average than IRIS in the eighth district.

One may then classify IRIS blocks according to their radii of convergence to the city's mean and represent them accordingly on a geographical map of Paris (Figure 3). First, note that boundary effects are clearly not predominant as peripheral western (W) and north-eastern (NE) parts of the city, for instance, do not exhibit the same radii of convergence to the city's average. In fact, boundaries, by forcing the aggregation of IRIS blocks closer to the city centre and beyond (rather than neighbouring IRIS blocks just outside the city) tend to smooth patterns rather than exacerbate them. A particularly interesting feature revealed by our method is that the W part of the city is 'further away' from the whole city than the NE part: trajectories for IRIS in the W part need 2–4 times as many aggregation steps to converge to the city's average than IRIS in the NE part. Both parts correspond to extreme points for the variable in question, with a concentration of social housing in the NE part and a substantially lower than average rate in the W part (local social housing rates are also visible in Figure 3). Starting from these extremal points above and below the city's average, with symmetrical geographical position in the city, the W and NE parts could have had similar radii of convergence to the city's mean. However, this is clearly not the case, thus revealing a higher level of singularity, as far as social housing is concerned, in the W part of the city than in the NE part.

Conversely, one observes a quasi ring of IRIS blocks with relatively short scales of convergence to the city's mean. These are the historically socially mixed areas along the *boulevards*, the former *faubourgs* that used to be just outside the city's walls before these were transformed into *boulevards*. Our method thus reveals a lasting imprint visible in terms of distance to the city's average for a variable, the social housing rate, that may be taken as a proxy to social diversity.

Income distribution. We now consider a second example and build trajectories for the income distribution in Paris.

As explained at the beginning of this section, when working with distributions available only through their quantiles, computing points for each group of blocks is a slightly more involved task. In this case, we chose to estimate the parameters of the best-fitting distribution (which happened to be log-normal) given a block's quantiles. Then, from this distribution, we simulated data corresponding to the number of households in the block. Thus, we obtained a full ensemble of simulated households for every possible group of neighbouring blocks, from one block only to the whole city.

Let us look at trajectories for IRIS blocks in three Parisian districts of comparable sizes and population numbers: the 13th, 16th, and 20th districts (see Figure 4). The first one corresponds to the south-western peripheral part of the city, the second one to the south-eastern peripheral part of the city, and the third one to the NE peripheral part.

The 16th tends to be further away from the full city's picture, as it takes generally longer for its blocks to converge to the city's distribution than for blocks in the 13th or in the 20th districts. The 13th district exhibits an interesting behaviour, with some trajectories that first come close to the city's mean but then bounce up again further from it. This can be understood as follows: some IRIS blocks in the 13th district belong to relatively well-mixed neighbourhoods, so initially their trajectory approaches the city's average distribution. But the aggregation process around them leads to incorporate much less wealthy blocks on the border of the city, at a stage where wealthier blocks towards the centre of the city and beyond have not yet been aggregated in large enough numbers to counterbalance the former. This sends trajectories away from the city's distribution again. A similar effect may (and does) take place for neighbourhoods relatively close to blocks that are much wealthier than the city as a whole (e.g. in the north-western 17th district, not shown here).

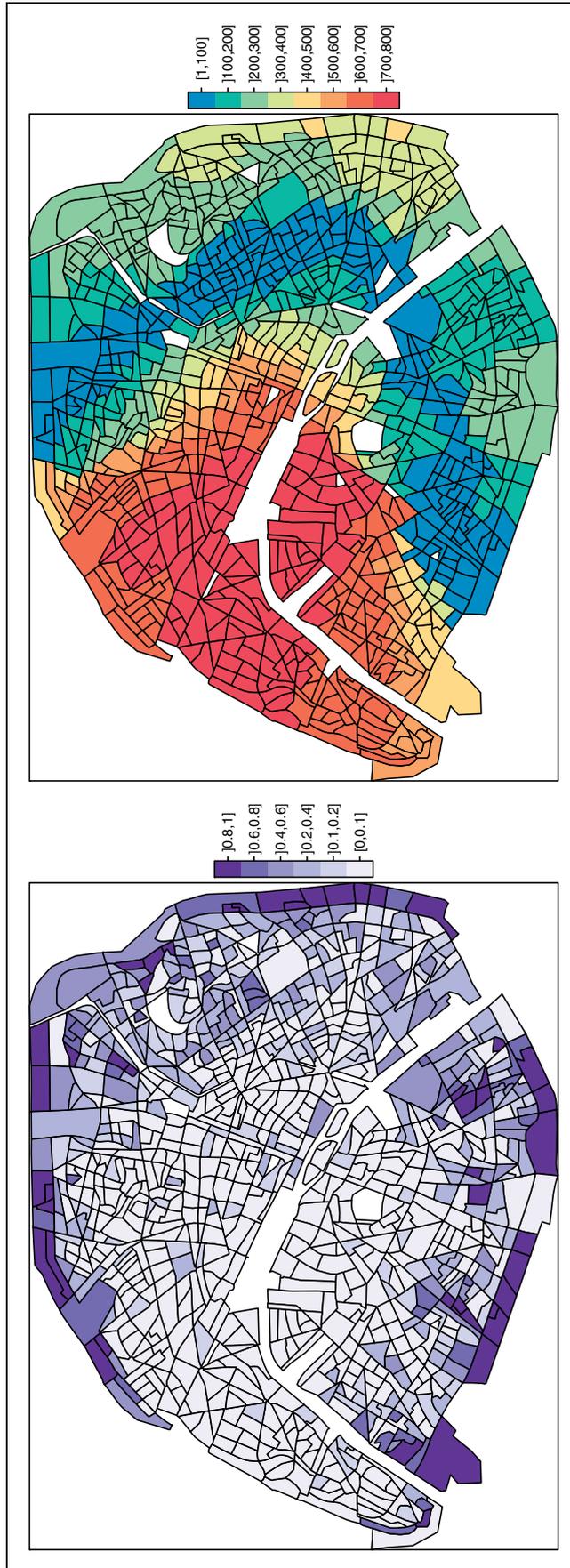


Figure 3. IRIS blocks coloured according to their radius of convergence to the global average social housing rate (right). Radii are expressed in terms of the number of aggregation steps. For comparison, the map on the left-hand side shows IRIS areas coloured according to the local rate of social housing. (White areas correspond to parks, riverbanks, and other IRIS where data are not available.) Observe that IRIS in the peripheral W part and the peripheral eastern and NE part of Paris presents similar geographical positions and corresponds to extreme values for the social housing rate (very high in the NE part, very low in the W part). Nonetheless, IRIS in the W part converges much slower to the city's mean, revealing a higher level of seclusion.

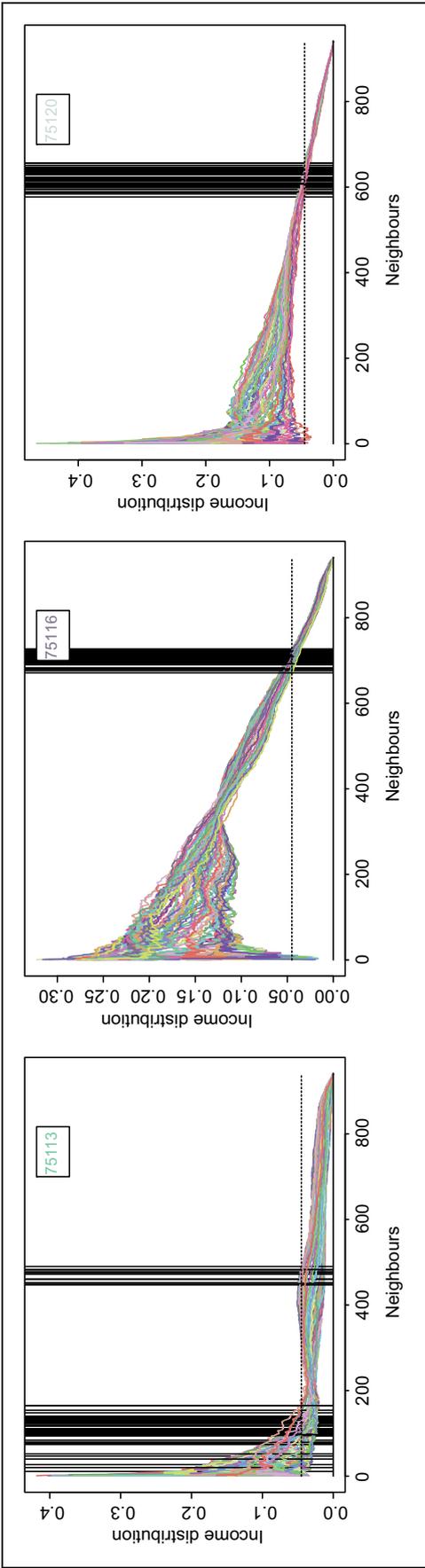


Figure 4. Trajectories for the income distribution, starting from each IRIS (census block) in Paris's 13th (left), 16th (middle), and 20th (right) districts. Ordinates correspond to the Kolmogorov–Smirnov distance (KS) between a group of blocks' income distribution and the whole city's income distribution. Solid vertical lines correspond to radii of convergence: the radius of convergence of a trajectory is defined as the point where the path last enters within 0.05 of the city's average in terms of KS distance.

Let us now look at a map of Paris with IRIS blocks coloured according to their radius of convergence for the income distribution (Figure 5). The picture here is much more symmetric than for the social housing rate. Indeed, wealthier (south-western) and poorer (NE) parts of the city both exhibit similar, slower convergence to the city's global distribution, compared to well-mixed areas such as the *faubourgs*. However, the wealthier, south-western part is again the last one to converge and constitutes a larger, more secluded area, including all of the seventh, 16th, and 17th districts, and most of the sixth and (more surprisingly) the 14th district. Remarkably, the eighth district (around the *Champs Élysées*) does not stand out as a secluded area for the income distribution: this is because, geographically, the eighth is much closer to well-mixed neighbourhoods than, for instance, the 16th district.

The uniform city as a null model: Random walks to the mean

The importance of null models has perhaps become more palpable in the last few decades, as new mathematical and statistical tools have come into play in a wider variety of fields, from urban geography to history and archaeology (Filet, 2017; Mills, 2017).

The null model against which segregation should be defined is the 'unsegregated' city, as pointed out by Louf and Barthelemy (2016). In the unsegregated city, distributions are uniform across space. Therefore, for instance, if the rate of social housing in the whole city is ρ , then any housing unit in the city, irrespective of its spatial position, has probability ρ of being a social housing unit: there is no spatial structure in the distribution of social housing units. This means that when considering a sequence of units defined solely according to some spatial rule (e.g. aggregating sequentially nearest neighbours), no sign of this spatial rule will appear in the sequence: everything will be as if one were drawing units at random from a well-mixed urn.

Any deviation from such random behaviour will be a sign that an underlying structure leads to a biased shuffling of the units. Note that the underlying structure may be a spatial one, as in the viewpoint we are taking here, but it may also be of another nature depending on the type of data one is examining.

A null model

Let us examine here the somewhat ideal case when one is dealing with count data, i.e. data at the individual level. Suppose each individual belongs to one of two groups A and B , e.g. people belonging to one of two social groups, or housing units being of one of two types. Start from a given point in the city, and increase the population, one by one, from the individual located there to the whole population in the city. Write $X_i = 1$ if the i th individual is of type A , $X_i = 0$ otherwise. Then setting $S_0 = 0$ and

$$S_n = \sum_{i=1}^n X_i \quad (5)$$

produces a trajectory, that is a sequence of points (n, S_n) , with n varying from 0 to the total number N of individual units in the city. S is simply the count data for group A .

In the case when the city is perfectly mixed, the trajectory thus produced will be a random walk S that, given its current value S_{n-1} , either moves up with probability

$$\rho_n = \frac{\rho N - S_{n-1}}{N - n + 1} \quad (6)$$

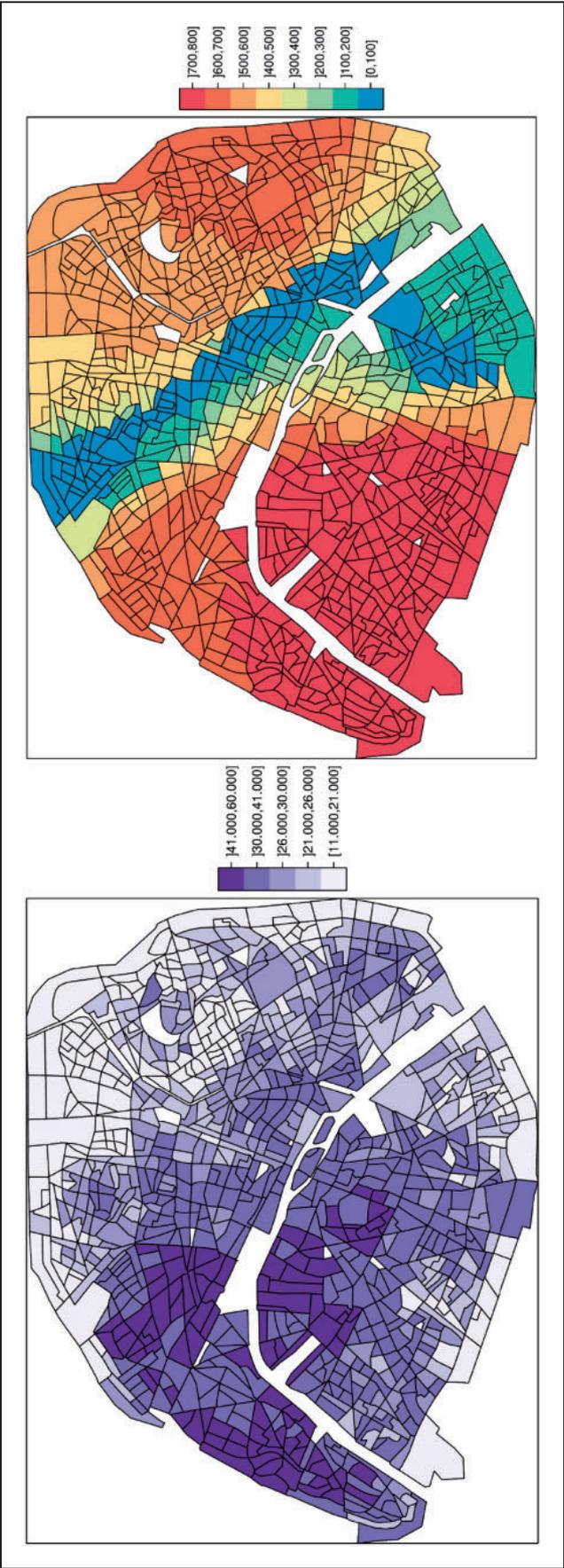


Figure 5. IRIS blocks coloured according to their radius of convergence to the city's income distribution (right). For comparison, the map on the left-hand side shows IRIS areas coloured according to the median of the (local) income distribution. Distance between distributions is measured with the KS distance. (White areas correspond to parks, riverbanks, and other IRIS where data are not available.)

or stays put with probability $1 - \rho_n$, where ρ represents the fraction of group A in the total population (N). For the more mathematically inclined, such a random walk may also be viewed as a standard problem of drawing without replacement from an urn containing N balls, the ρN of which being of one type and the rest being of another type. S_n would thus follow a well-known hypergeometric distribution – see Feller (1968).

Other sequences of interest that may be defined similarly are

- the sequence of averages

$$M_n = S_n/n$$

- the sequence of differences D_n , when one counts +1 for an individual of group A and –1 otherwise

$$D_n = 2S_n - n$$

Both sequences, M and D , are random walks that move up with probability ρ_n or down with probability $1 - \rho_n$, ρ_n being as defined above in equation (6).

All three types of trajectories converge to well-identified final values: S converges to $N\rho$, M converges to ρ , and D converges to $N(2\rho - 1)$. From a probabilistic point of view, any of these three sequences is fully characterized by ρ and N . Here they represent three facets of the same null model: a perfectly mixed city.

Now, if one has count data for a given variable in a city, one may build N sequences (one for each starting point). One thus produces a sample of trajectories that would, if the city were well mixed, exhibit the statistical properties of a set of N random sequences obtained by drawing balls without replacement from a well-mixed urn. Any statistically significant deviation from such a set of random sequences signals the presence of spatial dissimilarities at multiple scales. We look at such deviations in the next subsection.

Deviation from random sequences

Count data sequences are similar to sequences that one encounters when counting votes in an election. Imagine an election with two candidates, A and B . Counting the votes cast for candidate A produces the S -sequence defined in the previous subsection, with $N\rho$ the final number of votes for A . A famous result, or rather a famous set of results, linked to count data in elections, is the so-called Ballot Theorem (Addario-Berry and Reed, 2008; Barton and Mallows, 1965; Bertrand, 1887; Feller, 1968; Takács, 1962; Whitworth, 1886). In its simplest form, it gives the probability that A be always in the lead, all through the counting process, given that A wins the election with a final share ρ of the votes. This is $2\rho - 1$ (note that if A wins the election, necessarily $\rho > 1/2$).

A very natural interpretation of the Ballot Theorem in terms of urban count data is the following. If I start from some point in the city and look at the first two housing units around me, and then the first three, and so on, what is the probability that I will always see a minority of social housing units, given that the fraction of social housing in the whole city is, say, 20%? The Ballot Theorem tells us this probability is 60%. Therefore, even if social housing is in a strong minority and if the city is perfectly mixed, when zooming out from a point chosen at random there is only a 60% probability that one will never come across a bespoke neighbourhood where social housing is in the majority.

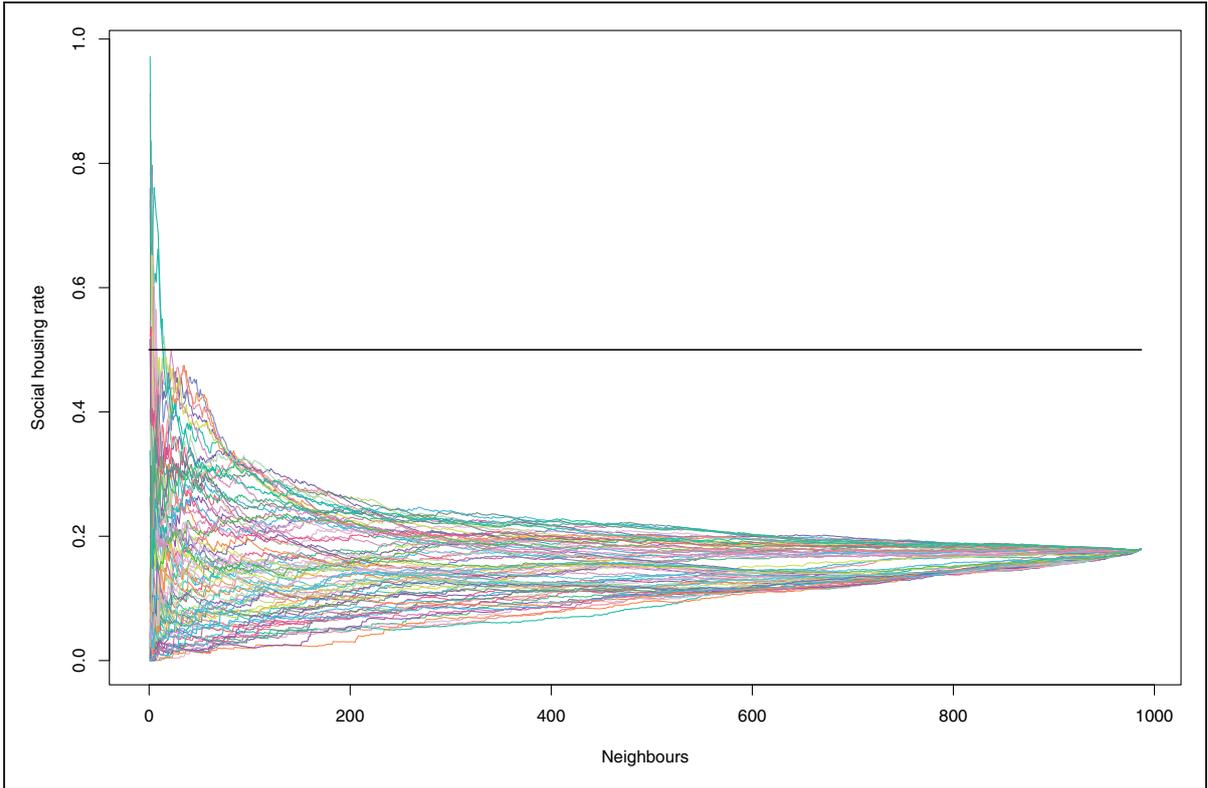


Figure 6. Trajectories for the social housing rate, starting from some (10%) of the 936 statistical units (so-called IRIS) in Paris. The solid line corresponds to a rate of 0.5. According to the Ballot Theorem, were the city ‘well mixed’ in terms of social housing, just above 64% of the trajectories would always remain below the 0.5 mark. In Paris, almost 85% of them do so.

On count data, a first test for the presence of an underlying spatial structure therefore simply consists in examining significant deviation from the Ballot Theorem. In Paris, for instance, the total fraction of social housing in the city is about 17.9%. Therefore, 64.2% of the trajectories should always stay below the horizontal line $y = 0.5$. In actuality, almost 85% of them always stay below 0.5 (Figure 6).

Obviously, one should account for the fact that we only have a finite sample of sequences – this is typically done through statistical hypothesis testing. In this specific context, we need a test to decide whether the observed deviation from the Ballot Theorem is significant or could be a simple random sampling effect. To this end, define the following sequence of Bernoulli random variables

$$Y_i = \begin{cases} 1 & \text{if trajectory for the } i\text{-th IRIS is always below } 0.5 \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

According to the Ballot Theorem

$$\text{Prob}(Y_i = 1) = 0.642, \quad \forall 1 \leq i \leq N \quad (8)$$

If all N trajectories were ‘drawn’ independently from a well-mixed city, the sum $\sum_{i=1}^N Y_i$ would follow a binomial distribution with mean $0.642N$ and variance $0.23N$. Here a simple

goodness-of-fit test shows that the Parisian trajectories for the social housing rate do not obey the Ballot Theorem, with a p -value of the order 10^{-16} .¹

Further, deviations from the null model of an unsegregated city may be characterized by a more complete statistical test. Indeed, with N starting points in a city, one builds N sequences, which should be compared to N full drawings of a well-mixed urn. By full drawing, we mean drawing all balls, one by one, without replacement, until the urn is therefore left empty. In partial drawings, where one draws n balls from an urn containing N balls, standard statistical tests include the hypergeometric one, similar to Fisher's exact test (Chvátal, 1979; Fisher, 1925, 1945). In fact, one could use such a test at every single point of any of the trajectories. However, this would not be using the available information to the full. In future work, we will make use of a multiple test for the full set of trajectories' points $S_n^{(j)}$, with $1 \leq n \leq N$ and $1 \leq j \leq N$ – see, for instance, Simes (1986). We will also present ways of characterizing deviations from the null model for data types other than count data. This will lead to new ways of measuring segregation levels. A thorough, systematic investigation of these new indices, including rigorous comparisons with commonly used indices (see, e.g. Fossett, 2017; Massey and Denton, 1988), will be carried out.

Conclusion

Aggregation of spatial data units has been so far generally considered as difficult to circumvent rather than an opportunity to analyse spatial dissimilarities. The new method introduced in this paper uses aggregation as a means to extract information on the relative singularity of each spatial unit within the city as a whole. Aggregation procedures may indeed reveal features that are not so easily seized from other perspectives. Another example is the recent use by Arcaute et al. (2015) of aggregation to explore the behaviour (and more specifically the scaling laws) of various statistical variables across a phase space corresponding to almost all possible city boundaries' definitions in England and Wales. Aggregation is akin to an exploration of space from a given starting point, and it is this observation that forms the basis of the method we have presented here.

A question we have not enlarged upon in this paper is that of the definition of convergence. For instance, we chose a ± 0.05 interval for convergence to the city's average social housing rate. In fact, one may tune this threshold so as to study in greater detail areas that converge more slowly or more quickly. We will investigate this in future work. Other extensions that we will seek to develop include devising a visual representation of multi-group trajectories (other than a synthetic one such as a Kullback–Leibler distribution distance) and identifying relevant features on which to apply machine learning algorithms in order to obtain unsupervised clustering of trajectories. It will also be interesting to explore models able to reproduce stylized aspects of observed trajectories: hidden Markov models, multi-agent models, intermittent diffusion models.

We will also address a difficulty that arises in practice: one rarely has access to individual unit data. The data are often already aggregated at some basic statistical unit level. This means trajectories are observed only at certain points determined by the sequence of sizes of the statistical units instead of at every individual point. So that comparisons with Ballot Theorem results need to be made through what is called subordinated random walks which are precisely sequences observed at a number of different points rather than at every single step.

An interesting feature of our method is that there is no need to restrict oneself to the aggregation of nearest neighbour units. One may also consider sequences defined from personal trajectories across the city (and thus aggregating units according to a personal

‘distance’: units where one lives, works, goes out, etc.). These personal trajectories transform the spatial network of blocks into personal networks reflecting personal paths and lives in the city. For some, the initial spatial network may turn into a small-world network, thanks to an ease of mobility across the different parts of the city. For others, personal trajectories may well be extremely concentrated around their local block. This is another aspect of segregation that we will seek to explore, using data from surveys and geo-mobility (e.g. mobile phone location data), and comparisons with random walks across the spatial network.

Authors’ note

Madalina Olteanu is also affiliated with INRA – Université Paris-Saclay

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Julien Randon-Furling  <http://orcid.org/0000-0001-9497-2297>

Note

1. The minuteness of this p -value is due to the fact that the perfectly mixed city is an extreme, unrealistic case. It provides a scale that is too extreme, if one wishes to use it directly as a measure of segregation. However, it is an absolute scale, much in the same sense as the absolute temperature in physics. As such, it may be used reliably for comparisons.

References

- Addario-Berry L and Reed B (2008) Ballot theorems, old and new. *Horizons of Combinatorics* 17: 9–35.
- Adrestani BM, O’Sullivan D and Davis P (2017) A multi-scaled agent-based model of residential segregation applied to a real metropolitan area. *Computers, Environment and Urban Systems* 69: 1–16.
- Andersson EK and Malmberg B (2015) Contextual effects on educational attainment in individualised, scalable neighbourhoods: Differences across gender and social class. *Urban Studies* 52(12): 2117–2133.
- Arcaute E, Hatna E, Ferguson P, et al. (2015) Constructing cities, deconstructing scaling laws. *Journal of the Royal Society Interface* 12(102): 20140745.
- Barton DE and Mallows CL (1965) Some aspects of the random sequence. *The Annals of Mathematical Statistics* 36(1): 236–260.
- Batty M (2007) *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. Cambridge (MA): MIT Press.
- Bertrand J (1887) Solution d’un probleme. *Comptes rendus de l’Académie des Sciences* 105: 369.
- Brown LA and Chung SY (2006) Spatial segregation, segregation indices and the geographical perspective. *Population, Space and Place* 12(2): 125–143.

- Busetti F (2014) Quantile aggregation of density forecasts. *Oxford Bulletin of Economics and Statistics* 79(4): 495–512.
- Caridi I, Pinasco J, Saintier N, et al. (2017) Characterizing segregation in the Schelling Voter model. *Physica A: Statistical Mechanics and its Applications* 487: 125–142.
- Chvátal V (1979) The tail of the hypergeometric distribution. *Discrete Mathematics* 25(3): 285–287.
- Clark WAV, Andersson E, Östh J, et al. (2015) A multiscale analysis of neighborhood composition in Los Angeles, 2000–2010: A location-based approach to segregation and diversity. *Annals of the Association of American Geographers* 105(6): 1260–1284.
- Cortese CF, Falk RF and Cohen JK (1976) Further considerations on the methodological analysis of segregation indices. *American Sociological Review* 41(4): 630–637.
- Cottrell M, Hazan A, Olteanu M, et al. (2017) Multidimensional urban segregation: An exploratory case study. In: *Twelfth international Workshop on Self-Organizing Maps and learning vector quantization (WSOM+)*. IEEE eXplore.
- Cowgill DO and Cowgill MS (1951) An index of segregation based on block statistics. *American Sociological Review* 16(6): 825–831.
- Duncan OD and Duncan B (1955) A methodological analysis of segregation indexes. *American Sociological Review* 20(2): 210–217.
- Farley R and Taeuber KE (1968) Population trends and residential segregation since 1960. *Science* 159(3818): 953–956.
- Feitosa FF, Camara G, Monteiro AMV, et al. (2007) Global and local spatial indices of urban segregation. *International Journal of Geographical Information Science* 21(3): 299–323.
- Feitosa FF, Le QB, Vlek PL, et al. (2012) Countering urban segregation in Brazilian cities: Policy-oriented explorations using agent-based simulation. *Environment and Planning B: Planning and Design* 39(6): 1131–1150.
- Feller W (1968) *An Introduction to Probability Theory and Its Applications*. Vol.1. New York: John Wiley & Sons.
- Filet C (2017) An attempt to estimate the impact of the spread of economic flows on Latenian urbanization. *Frontiers in Digital Humanities* 3: 10.
- Fisher RA (1925) *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd.
- Fisher RA (1945) A new test for 2×2 tables. *Nature* 156(388): 10–1038.
- Fossett M (2017) *New Methods for Measuring and Analyzing Segregation*. Vol.42. New York: Springer.
- Fotheringham AS and Wong DW (1991) The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23(7): 1025–1044.
- Genest C (1992) Vincentization revisited. *The Annals of Statistics* 20(2): 1137–1142.
- Genest C and Zidek JV (1986) Combining probability distributions: A critique and an annotated bibliography. *Statistical Science* 1(1): 114–135.
- Glaeser EL, Kolko J and Saiz A (2001) Consumer city. *Journal of Economic Geography* 1(1): 27–50.
- Harris R and Owen D (2017) Implementing a multilevel index of dissimilarity in R with a case study of the changing scales of residential ethnic segregation in England and Wales. *Environment and Planning B: Urban Analytics and City Science*. Epub ahead of print, 18 December 2017. DOI: 10.1177/2399808317748328.
- INSEE (2017) IRIS – Definition. Available at: <https://www.insee.fr/en/metadonnees/definition/c1523> (accessed 5 November 2017).
- James DR and Taeuber KE (1985) Measures of segregation. *Sociological Methodology* 15: 1–32.
- Jelinski DE and Wu J (1996) The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology* 11(3): 129–140.
- Jones K, Johnston R, Manley D, et al. (2015) Ethnic residential segregation: A multilevel, multigroup, multiscale approach exemplified by London in 2011. *Demography* 52(6): 1995–2019.
- Leckie G and Goldstein H (2015) A multilevel modelling approach to measuring changing patterns of ethnic composition and segregation among London secondary schools, 2001–2010. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178(2): 405–424.
- Leckie G, Pillinger R, Jones K, et al. (2012) Multilevel modeling of social segregation. *Journal of Educational and Behavioral Statistics* 37(1): 3–30.

- Louf R and Barthelemy M (2016) Patterns of residential segregation. *PLoS One* 11(6): e0157476.
- Manley D, Johnston R, Jones K, et al. (2015) Macro-, meso-and microscale segregation: Modeling changing ethnic residential patterns in Auckland, New Zealand, 2001–2013. *Annals of the Association of American Geographers* 105(5): 951–967.
- Massey DS and Denton NA (1988) The dimensions of residential segregation. *Social Forces* 67(2): 281–315.
- Mills BJ (2017) Social network analysis in archaeology. *Annual Review of Anthropology* 46(1): 379–397.
- Openshaw S (1984) *The Modifiable Areal Unit Problem*. Norwich: Geo Abstracts University of East Anglia.
- Openshaw S and Taylor PJ (1979) A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. *Statistical Applications in the Spatial Sciences* 21: 127–144.
- Östh J, Clark WAV and Malmberg B (2015) Measuring the scale of segregation using k-nearest neighbor aggregates. *Geographical Analysis* 47(1): 34–49.
- Reardon SF and Firebaugh G (2002) Measures of multigroup segregation. *Sociological Methodology* 32(1): 33–67.
- Sakoda JM (1971) The checkerboard model of social interaction. *Journal of Mathematical Sociology* 1(1): 119–132.
- Schelling TC (1971) Dynamic model of segregation. *Journal of Mathematical Sociology* 1: 143–186.
- Schwartz J and Winship C (1980) The welfare approach to measuring inequality. *Sociological Methodology* 11: 1–36.
- Simes RJ (1986) An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73(3): 751–754.
- Stoker TM (1984) Completeness, distribution restrictions, and the form of aggregate functions. *Econometrica: Journal of the Econometric Society* 887–907.
- Takács L (1962) Ballot problems. *Probability Theory and Related Fields* 1(2): 154–158.
- US Census Bureau (1994) *Geographic Areas Reference Manual*. Washington DC: US Dpt of Commerce.
- White MJ (1983) The measurement of spatial segregation. *American Journal of Sociology* 88(5): 1008–1018.
- Whitworth WA (1886) *Choice and Chance*. 4th ed. Cambridge: Deighton Bell.

Julien Randon-Furling is a Senior Lecturer in Mathematical Sciences at Université Paris 1 Panthéon-Sorbonne.

Madalina Olteanu is a Senior Lecturer in Statistics at Université Paris 1 Panthéon-Sorbonne, currently on sabbatical leave at INRA, Université Paris-Saclay.

Antoine Lucquiaud is a doctoral researcher at the Centre for Statistics, Analysis, and Multidisciplinary Modelling, Université Paris 1 Panthéon-Sorbonne.

Références

- [1] Addario-Berry L, Reed B. Ballot theorems, old and new. *Horizons combinatorics* **17**, 9–35 (2008).
- [2] Agarwal A, Elkind E, Gan J, Voudouris AA. Swap stability in Schelling games on graphs in *Proc. AAAI Conf. on Artificial Intelligence*. **34**(2), 1758–1765 (2019).
- [3] Andersson EK, Malmberg B. Contextual effects on educational attainment in individualised, scalable neighbourhoods: Differences across gender and social class. *Urban Studies*. **52**(12), 2117–2133 (2015).
- [4] André D. Solution directe du problème résolu par M. Bertrand. *Comptes-rendus de l'Académie des Sciences* **105**, 436–437 (1887).
- [5] Aral S, Muchnik L, Sundararajan A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *PNAS* **106**(51), 21544–21549 (2009).
- [6] Arcaute E, Hatna E, Ferguson P, Youn H, Johansson A, Batty M. Constructing cities, deconstructing scaling laws. *J. The Royal Soc. Interface* **12**(102), 20140745 (2015).
- [7] Arènes JF, Tanvé G. *18 600 logements durablement vacants à Paris*. Atelier Parisien d'Urbanisme, note 206 (2021).
- [8] Avenel C. La question des quartiers dits “sensibles” à l’épreuve du ghetto. *Revue économique* **67**(3), 415–441 (2016).
- [9] Avetisov V, Gorsky A, Maslov S, Nechaev S, Valba O. Phase transitions in social networks inspired by the Schelling model. *Phys. Rev. E* **98**(3), 032308 (2018).
- [10] Bak P, Tang C, Wiesenfeld K. Self-organized criticality. *Phys. Rev. A* **38**(1), 364–374 (1988).
- [11] Banisch S. *Markov Chain Aggregation for Agent-Based Models*. Ph.D Thesis, Universität Bielefeld (2013).
- [12] Banos A. Network effects in Schelling’s model of segregation: new evidence from agent-based simulation. *EPB: Planning and Design* **39**(2), 393–405 (2012).
- [13] Barbier E. Généralisation du problème résolu par M. J. Bertrand. *Comptes-rendus de l'Académie des Sciences* **105**, 407 (1887).
- [14] Barmpalias G, Elwes R, Lewis-Pye A. Tipping points in 1-dimensional Schelling models with switching agents. *J. Stat. Phys.* **158**(4), 806–852 (2015).

- [15] Barmpalias G, Elwes R, Lewis-Pye A. Unperturbed Schelling segregation in two or three dimensions. *J. Stat. Phys.* **164**(6), 1460–1487 (2016).
- [16] Barmpalias G, Elwes R, Lewis-Pye A. Digital morphogenesis via Schelling segregation. *Nonlinearity* **31**(4), 1593 (2018).
- [17] Barmpalias G, Elwes R, Lewis-Pye A. Minority population in the one-dimensional Schelling model of segregation. *J. Stat. Phys.* **173**(5), 1408–1458 (2018).
- [18] Baumann F, Lorenz-Spreen P, Sokolov IM, Starnini M. Modeling echo chambers and polarization dynamics in social networks. *Phys. Rev. Lett.* **124**(4), 048301 (2020).
- [19] Bertrand J. Solution d’un problème. *Comptes-rendus de l’Académie des Sciences* **105**, 369 (1887).
- [20] Bhakta P, Miracle S, Randall D. Clustering and Mixing Times for Segregation Models on \mathbb{Z}^2 in *Proc. 25th ACM-SIAM Symp. on Discrete Algorithms*. 327–340 (2014).
- [21] Binder K, Virnau P. Overview: Understanding nucleation phenomena from simulations of lattice gas models. *The J. Chem. Phys.* **145**(21), 211701 (2016).
- [22] Blex C, Yasseri T. Positive algorithmic bias cannot stop fragmentation in homophilic social networks. *J. Math. Sociol.* **46**(1), 80–97 (2020).
- [23] Blume M. Theory of the first-order magnetic phase change in UO_2 . *Phys. Rev.* **141**(2), 517 (1966).
- [24] Blume M, Emery VJ, Griffiths RB. Ising model for the λ transition and phase separation in He^3 – He^4 mixtures. *Phys. Rev. A* **4**, 1071–1077 (1971).
- [25] Bouzouina L, Mignot D. Disparités de revenus à différentes échelles spatiales en France de 1985 à 2001 in *Proc. 55e colloque de l’ASRDLF*. (2005).
- [26] Brandt C, Immorlica N, Kamath G, Kleinberg R. An analysis of one-dimensional Schelling segregation in *Proc. 44th ACM Symposium on Theor. Comput.* 789–804 (2012).
- [27] Brown LA, Chung SY. Spatial segregation, segregation indices and the geographical perspective. *Population, space and place* **12**(2), 125–143 (2006).
- [28] Bruch EE, Mare RD. Neighborhood choice and neighborhood change. *Am. J. Sociol.* **112**(3), 667–709 (2006).
- [29] Busetti F. Quantile aggregation of density forecasts. *Oxford Bulletin Econ. Stat.* **79**(4) (2014).

- [30] Bézenac C de, Clark WAV, Olteanu M, Randon-Furling J. Measuring and visualizing patterns of ethnic concentration: The role of distortion coefficients. *Geogr. Analysis* **54**(1), 173–196 (2022).
- [31] Cammarota C, Cavagna A. A novel method for evaluating the critical nucleus and the surface tension in systems with first order phase transition. *The J. Chem. Phys.* **127**(21), 214703 (2007).
- [32] Capel H. On the possibility of first-order phase transitions in Ising systems of triplet ions with zero-field splitting. *Physica* **32**(5), 966–988 (1966).
- [33] Casas-Vázquez J, Jou D. Temperature in non-equilibrium states: a review of open problems and current proposals. *Reports on Prog. Phys.* **66**(11), 1937–2023 (2003).
- [34] Chauhan A, Lenzner P, Molitor L. Schelling segregation with strategic agents in *Proc. Int. Symp. Algorithmic Game Theory*. 137–149 (2018).
- [35] Chopard B, Droz M. *Cellular automata modeling of physical systems*. Cambridge University Press (1998).
- [36] Clark WAV. Residential preferences and neighborhood racial segregation: a test of the Schelling segregation model. *Demography* **28**(1), 1–18 (1991).
- [37] Clark WAV, Fossett M. Understanding the social context of the Schelling segregation model. *PNAS* **105**(11), 4109–4114 (2008).
- [38] Clark WAV, Andersson E, Östh J, Malmberg B. A multiscalar analysis of neighborhood composition in Los Angeles, 2000–2010: A location-based approach to segregation and diversity. *Annals Assoc. Am. Geogr.* **105**(6), 1260–1284 (2015).
- [39] Cortese CF, Falk RF, Cohen JK. Further considerations on the methodological analysis of segregation indices. *Am. Sociol. Rev.* **41**(4), 630–637 (1976).
- [40] Cortez V, Medina P, Goles E, Zarama R, Rica S. Attractors, statistics and fluctuations of the dynamics of the Schelling’s model for social segregation. *Eur. Phys. J. B* **88**, 25 (2015).
- [41] Cortez V, Rica S. Dynamics of the Schelling social segregation model in networks. *Procedia Comput. Sci.* **61**, 60–65 (2015).
- [42] Cottrell M, Hazan A, Olteanu M, Randon-Furling J. Multidimensional urban segregation: an exploratory case study in *Proc. 12th Int. Workshop on Self-Organizing Maps and Learning Vector Quantization*. (2017).
- [43] Cousin B. Quartiers bourgeois in *Dictionnaire des inégalités*. Armand Colin, 322–370 (2014).

- [44] Cousin B. Ségrégation résidentielle in *Dictionnaire des inégalités*. Armand Colin, 322–370 (2014).
- [45] Cousin B. Ségrégation résidentielle (mesure) in *Dictionnaire des inégalités*. Armand Colin, 322–370 (2014).
- [46] Cutler DM, Glaeser EL. Are ghettos good or bad? *The Q. J. Econ.* **112**(3), 827–872 (1997).
- [47] Dall’Asta L, Castellano C, Marsili M. Statistical physics of the Schelling model of segregation. *J. Stat. Mech.* **07**, L07002 (2008).
- [48] Diu B, Guthmann C, Lederer D, Roulet B. *Éléments de physique statistique*. Hermann (1989).
- [49] Diu B, Guthmann C, Lederer D, Roulet B. *Thermodynamique*. Hermann (2007).
- [50] Drossel B, Schwabl F. Self-organized critical forest fires. *Phys. Rev. Lett.* **69**(11), 1629 (1992).
- [51] Duncan OD, Ray PC, Beverly D. *Statistical Geography: problems in analyzing areal data*. Glencoe (1961).
- [52] Duncan OD, Duncan B. A methodological analysis of segregation indexes. *Am. Sociol. Rev.* **20**(2), 210–217 (1955).
- [53] Durand MF. *Espace mondial : l’Atlas*. Les Presses de Sciences Po (2018).
- [54] Durrett R, Zhang Y. Exact solution for a metapopulation version of Schelling’s model. *PNAS* **111**(39), 14036–14041 (2014).
- [55] Echzell H, Friedrich T, Lenzner P, Molitor L, Pappik M, Schöne F, Sommer F, Stangl D. Convergence and hardness of strategic Schelling segregation in *Proc. Int. Conf. Web and Internet Econ.* 156–170 (2019).
- [56] Elkind E, Gan J, Igarashi A, Suksompong W, Voudouris AA. Schelling games on graphs in *Proc. 28th Int. Joint Conf. on Artificial Intelligence*. 266–272 (2019).
- [57] F R, P S. Glassy dynamics of kinetically constrained models. *Adv. physics* **52**(4), 219–342 (2003).
- [58] Feller W. *An introduction to probability theory and its applications 1 & 2*. John Wiley & Sons (1968).
- [59] Filet C. An attempt to estimate the impact of the spread of economic flows on Latenian urbanization. *Frontiers in Digit. Human.* **3**(10) (2017).
- [60] Forsé M, Parodi M. La ségrégation spatiale selon Schelling : la perversité est ailleurs. *OFCE, Sci. Po* **05** (2006).
- [61] Forsé M, Parodi M. Low levels of ethnic intolerance do not create large

- ghettos: a discussion about an interpretation of Schelling's model. *L'Année sociologique* **60**(2), 445–473 (2010).
- [62] Forsé M, Parodi M. Retour critique sur le modèle de ségrégation urbaine de Schelling. *Émulations* **31**, 91–104 (2019).
- [63] Fossett M. *New methods for measuring and analyzing segregation*. Springer (2017).
- [64] Fotheringham AS, Wong DW. The modifiable areal unit problem in multivariate statistical analysis. *Envir. and Plan. A: Econ. and Space* **23**(7), 1025–1044 (1991).
- [65] Franz S, Mulet R, Parisi G. Kob-Andersen model: A nonstandard mechanism for the glassy transition. *Phys. Rev. E* **65**(2), 021506 (2002).
- [66] Fredrickson GH, Andersen HC. Kinetic ising model of the glass transition. *Phys. Rev. Lett.* **53**(13), 1244 (1984).
- [67] Fredrickson GH, Andersen HC. Facilitated kinetic ising models and the glass transition. *The J. chemical physics* **83**(11), 5822–5831 (1985).
- [68] Garrett RK. Echo chambers online? politically motivated selective exposure among internet news users. *J. computer-mediated communication* **14**(2), 265–285 (2009).
- [69] Gauvin L. *Modélisation de systèmes socio-économiques à l'aide des outils de physique statistique*. Ph.D Thesis, Université Pierre et Marie Curie (2010).
- [70] Gauvin L, Vannimenus J, Nadal JP. Phase diagram of a Schelling segregation model. *Eur. Phys. J. B* **70**, 293–304 (2009).
- [71] Gauvin L, Nadal JP, Vannimenus J. Schelling segregation in an open city: a kinetically constrained Blume-Emery-Griffiths spin-1 system. *Phys. Rev. E* **81**(6), 066120 (2010).
- [72] Glauber RJ. Time-dependent statistics of the Ising model. *J. Math. Phys.* **4**(2), 294–307 (1963).
- [73] Godoy M, Figueiredo W. Critical behavior of the mixed-spin Ising model with two competing dynamics. *Phys. Rev. E* **65**(2), 026111 (2002).
- [74] Goles Domic N, Goles E, Rica S. Dynamics and complexity of the Schelling segregation model. *Phys. Rev. E* **83**(5), 056111 (2011).
- [75] Grafmeyer Y. *Sociologie Urbaine*. Armand Colin (1994).
- [76] Grandi BCS, Figueiredo W. Critical exponents of the Ising model with competing Glauber and Kawasaki dynamics. *Phys. Rev. E* **53**(5), 5484 (1996).

- [77] Grauwin S. *Exploring Social Phenomena with Complex Systems Tools*. Ph.D Thesis, École normale supérieure de Lyon (2011).
- [78] Grauwin S, Bertin E, Lemoy R, Jensen P. Competition between collective and individual dynamics. *PNAS* **106**(49), 20622–20626 (2009).
- [79] Grauwin S, Goffette-Nagot F, Jensen P. Dynamic models of residential segregation: an analytical solution. *J. Publ. Econ.* **96**, 124–141 (2012).
- [80] Grauwin S, Hunt D, Bertin E, Jensen P. Effective free energy for individual dynamics. *Adv. Complex Syst.* **14**(04), 529–536 (2011).
- [81] Hardy GH, Ramanujan S. Asymptotic formulæ in combinatory analysis. *Proc. Lond. Math. Soc.* **2**(1), 75–115 (1918).
- [82] Harris R, Owen D. Implementing a multilevel index of dissimilarity in R with a case study of the changing scales of residential ethnic segregation in England and Wales. *EPB: Urban Analytics and City Science* **45**(6), 1003–1021 (2017).
- [83] Hazan A, Randon-Furling J. A Schelling model with switching agents: decreasing segregation via random allocation and social mobility. *Eur. Phys. J. B* **86**(10), 1–9 (2013).
- [84] Heisenberg W. Zur Theorie des Ferromagnetismus. *Z. Phys.* **49**, 619–636 (1928).
- [85] Heisenberg W. Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. *Z. Phys.* **33**, 879–893 (1955).
- [86] Henry AD, Prałat P, Zhang CQ. Emergence of segregation in evolving social networks. *PNAS* **108**(21), 8605–8610 (2011).
- [87] Homsy GM. Viscous fingering in porous media. *Annu. Rev. Fluid Mech.* **19**, 271–311 (1987).
- [88] Immorlica N, Kleinberg R, Lucier B, Zadomighaddam M. Exponential segregation in a two-dimensional Schelling model with tolerant individuals in *Proc. 28th ACM-SIAM Symp. on Discrete Algorithms*. 984–993 (2017).
- [89] Ising E. Beitrag zur Theorie des Ferromagnetismus. *Z. Phys.* **31**(1), 253–258 (1925).
- [90] Istrate G. Stochastic stability in Schelling’s segregation model with markovian asynchronous update in *Proc. Int. Conf. on Cellular Automata*. 416–427 (2018).
- [91] James DR, Taeuber KE. Measures of segregation. *Sociol. Methodol.* **15**, 1–32 (1985).
- [92] Jargowsky PA. Take the money and run: Economic segregation in US metropolitan areas. *Am. Sociol. Rev.* **61**(6), 984–998 (1996).
- [93] Jelinski DE, Wu J. The modifiable areal unit problem and implications

- for landscape ecology. *Landscape Ecology* **11**(3), 129–140 (1996).
- [94] Jones FL. Simulation models of group segregation. *Aust. N.Z. J. Sociol.* **21**(3), 431–444 (1985).
- [95] Kanellopoulos P, Kyropoulou M, Voudouris AA. Modified Schelling games. *Theor. Comput. Sci.* **880**, 1–19 (2021).
- [96] Kawasaki K. Diffusion constants near the critical point for time-dependent Ising models I. *Phys. Rev.* **145**(1), 224 (1966).
- [97] Kawasaki K. Kinetics of Ising models in *Phase transitions and critical phenomena*, ed. Domb, Lebowitz. **2**, 224–501 (1972).
- [98] Kirman A. A comment on “Low levels of ethnic intolerance do not create large ghettos” by Michel Forsé and Maxime Parodi. *L’Année sociologique* **60**(2), 475–480 (2010).
- [99] Kob W, Andersen HC. Kinetic lattice-gas model of cage effects in high-density liquids and a test of mode-coupling theory of the ideal-glass transition. *Phys. Rev. E* **48**(6), 4364 (1993).
- [100] Krapivsky PL, Redner S, Ben-Naim E. *A kinetic view of statistical physics*. Cambridge University Press (2010).
- [101] Laurie AJ, Jaggi NK. Role of “vision” in neighbourhood racial segregation: a variant of the Schelling segregation model. *Urban Studies.* **40**(13), 2687–2704 (2003).
- [102] Leckie G, Pillinger R, Jones K, Goldstein H. Multilevel modeling of social segregation. *J. Educ. Behav. Stat.* **37**(1), 3–30 (2012).
- [103] Louf R, Barthelemy M. Patterns of residential segregation. *PloS One* **11**(6), e0157476 (2016).
- [104] Macy MW, Willer R. From factors to factors: computational sociology and agent-based modeling. *Annu. Rev. Sociol.* **28**(1), 143–166 (2002).
- [105] Massey DS, Denton NA. The dimensions of residential segregation. *Soc. forces* **67**(2), 281–315 (1988).
- [106] Massey DS, Denton NA. *American apartheid: Segregation and the making of the underclass*. Harvard University Press (1993).
- [107] Massey DS, Eggers ML. The ecology of inequality: Minorities and the concentration of poverty, 1970-1980. *Am. J. Sociol.* **95**(5), 1153–1188 (1990).
- [108] Maurin E. *Le ghetto français*. Seuil (2004).
- [109] Mayer SE. How did the increase in economic inequality between 1970 and 1990 affect children’s educational attainment? *Am. J. Sociol.* **107**(1), 1–32 (2001).
- [110] Mele A. A structural model of dense network formation. *Econometrica* **85**(3), 825–850 (2017).

- [111] Meyer-Ortmanns H. Immigration, integration and ghetto formation. *Int. J. Mod. Phys. C* **14**, 311 (2003).
- [112] Mills BJ. Social network analysis in archaeology. *Annu. Rev. Anthropol.* **46**(1) (2017).
- [113] Moore EF. Machine models of self-reproduction in *Proc. Symp. in Appl. Math.* **14**(5), 17–33 (1962).
- [114] Müller K, Schulze C, Stauffer D. Inhomogeneous and self-organised temperature in Schelling-Ising model. *Int. J. Mod. Phys. C* **19**, 385 (2008).
- [115] Neumann J von, Burks AW. *Theory of Self-Reproducing Automata*. University of Illinois Press (1966).
- [116] Nield DA. Surface tension and buoyancy effects in cellular convection. *J. Fluid Mech.* **19**(3), 341–352 (1964).
- [117] Nielsen AV, Gade AL, Juul J, Strandkvist C. Schelling model of cell segregation based only on local information. *Phys. Rev. E* **92**(5), 052705 (2015).
- [118] Norris JR. *Markov Chains*. Cambridge University Press (1998).
- [119] Ódor G. Self-organizing, two-temperature Ising model describing human segregation. *Int. J. Mod. Phys. C* **19**(03), 393–398 (2008).
- [120] Olami Z, Feder HJS, Christensen K. Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes. *Phys. Rev. Lett.* **68**(8), 1244 (1992).
- [121] Olteanu M, de Bézenac C, Clark WAV, Randon-Furling J. Revealing multiscale segregation effects from fine-scale data: A case study of two communities in Paris. *Spatial Demogr.* **8**(3), 255–267 (2020).
- [122] Olteanu M, Randon-Furling J, Clark WAV. Segregation through the multiscalar lens. *PNAS* **116**(25), 12250–12254 (2019).
- [123] Omidvar H. *Self-Organization of a Long-Range Schelling-Type Interacting Particle System*. Ph.D Thesis, University of California San Diego (2021).
- [124] Omidvar H, Franceschetti M. Self-organized segregation on the grid. *J. Stat. Phys.* **170**(4), 748–783 (2018).
- [125] Omidvar H, Franceschetti M. Shape of diffusion and size of monochromatic region of a two-dimensional spin system in *Proc. 50th ACM SIGACT Symp. on Theory of Comput.* 100–113 (2018).
- [126] Omidvar H, Franceschetti M. Self-organized segregation on the grid (version complète). *arXiv preprint:1705.08586* (2018).
- [127] Omidvar H, Franceschetti M. Improved intolerance intervals and size bounds for a Schelling-type spin system. *J. Stat. Mech.* **7**, 073302

- (2021).
- [128] Openshaw S. *The modifiable areal unit problem*. University of East Anglia (1984).
 - [129] Openshaw S, Taylor PJ. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. *Stat. Appl. Spatial Sci.* **21**, 127–144 (1979).
 - [130] Orfield G, Yun JT. *Resegregation in American schools*. The Civil Rights Project, Harvard University (1999).
 - [131] Ortega D, Rodríguez-Laguna J, Korutcheva E. A Schelling model with a variable threshold in a closed city segregation model. *Physica A: Stat. Mech. Appl.* **574**, 126010 (2021).
 - [132] Östh J, Clark WAV, Malmberg B. Measuring the scale of segregation using k-nearest neighbor aggregates. *Geogr. Analysis* **47**(1), 34–49 (2015).
 - [133] Pans R, Vriend NJ. Schelling’s spatial proximity model of segregation revisited. *J. Public Econ.* **91**(1), 1–24 (2007).
 - [134] Paugam S, Cousin B, Giorgetti C, Naudet J. *Ce que les riches pensent des pauvres*. Seuil (2017).
 - [135] Pick G. Geometrisches zur Zahlenlehre. *Lotos : Z. Naturwissenschaften* **19**, 311–319 (1899).
 - [136] Pinçon-Charlot M, Pinçon M. *Les ghettos du gotha*. Seuil (2007).
 - [137] Pitman J. Brownian motion, bridge, excursion, and meander characterized by sampling at independent uniform times. *Electron. J. Probab.* **4**, 1–33 (1999).
 - [138] Pollicott M, Weiss H. The dynamics of Schelling-type segregation models and a nonlinear graph laplacian variational problem. *Adv. Appl. Math.* **27**(1), 17–40 (2001).
 - [139] Potts RB. Some generalized order-disorder transformations. *Math. Proc. Cambridge Philos. Soc.* **48**(1), 106–109 (1952).
 - [140] Poulser M, Forrest J, Johnston R. From modern to post-modern? Contemporary ethnic residential segregation in four US metropolitan areas. *Cities* **19**(3), 161–172 (2002).
 - [141] Préteceille E. La ségrégation sociale a-t-elle augmenté? la métropole parisienne entre polarisation et mixité. *Sociétés contemporaines* **62**(2), 69–93 (2006).
 - [142] Randon-Furling J, Olteanu M, Lucquiaud A. From urban segregation to spatial structure detection. *EPB: Urban Analytics and City Science* **47**(4), 645–661 (2020).
 - [143] Reardon SF. A conceptual framework for measuring segregation and

- its association with population outcomes in *Methods in social epidemiology*. John Wiley & Sons, 169–192 (2017).
- [144] Reardon SF, O’Sullivan D. Measures of spatial segregation. *Sociol. Methodol.* **34**(1), 121–162 (2004).
- [145] Reardon SF, Firebaugh G. Measures of Multigroup Segregation. *Sociol. Methodol.* **32**(1), 33–67 (2002).
- [146] Rogers T, McKane A. A unified framework for Schelling’s model of segregation. *J. Stat. Mech.* **7**, P07006 (2011).
- [147] Rostand E. *Cyrano de Bergerac*. Acte V, Scène VI (1897).
- [148] Sahasranaman A, Jensen HJ. Ethnicity and wealth: The dynamics of dual segregation. *PloS One* **13**(10), e0204307 (2018).
- [149] Sahasranaman A, Jensen HJ. Dynamics of transformation from segregation to mixed wealth cities. *PloS One* **11**(11), e0166960 (2016).
- [150] Sakoda JM. *Minidoka: an analysis of changing patterns of social interaction*. Ph.D Thesis, University of California, Berkeley (1949).
- [151] Sakoda JM. The checkerboard model of social interaction. *J. Math. Sociol.* **1**(1), 119–132 (1971).
- [152] Schelling TC. Models of segregation. *The Am. Econ. Rev.* **59**(2), 488–493 (1969).
- [153] Schelling TC. Dynamic model of segregation. *J. Math. Sociol.* **1**, 143–186 (1971).
- [154] Schelling TC. *Micromotives and Macrobehavior*. WW Norton (1978).
- [155] Schwartz J, Winship C. The welfare approach to measuring inequality. *Sociol. Methodol.* **11**, 1–36 (1980).
- [156] Silver D, Byrne U, Alder P. Venues and segregation: a revised Schelling model. *PLoS One* **16**(1), e0242611 (2021).
- [157] Singh A, Vainchtein D, Weiss H. Schelling’s segregation model: parameters, scaling, and aggregation. *Demogr. Res.* **21**, 341 (2009).
- [158] Skvoretz J. Introduction to special edition on ethnic preferences, social distance dynamics, and residential segregation: Theoretical explanations using simulation analysis by Mark Fossett in *J. Math. Sociol.* **30**, 181–184 (2006).
- [159] Stauffer D, Solomon S. Ising, Schelling and self-organising segregation. *Eur. Phys. J. B* **57**, 473–479 (2007).
- [160] Stauffer D. A biased review of sociophysics. *J. Stat. Phys.* **151**(1-2), 9–20 (2013).
- [161] Stoker TM. Completeness, distribution restrictions, and the form of aggregate functions. *Econometrica* **52**(4), 887–907 (1984).
- [162] Strandkvist C. *Physical Mechanisms for the Emergence of Order in*

- Biological Systems*. Ph.D Thesis, University College London (2016).
- [163] Sumour MA, El-Astal AH, Radwan MA, Shabat MM. Urban segregation with cheap and expensive residences. *Int. J. Mod Phys. C* **19**(4), 637–645 (2008).
- [164] Szolnoki A. Phase transitions in the kinetic Ising model with competing dynamics. *Phys. Rev. E* **62**(5), 7466 (2000).
- [165] Takács L. Ballot problems. *Probab. Theory Relat. Fields* **1**(2), 154–158 (1962).
- [166] Takács L. A combinatorial theorem for stochastic processes. *Bull. Am. Math. Soc.* **71**, 649–650 (1965).
- [167] Theil H, Finizza AJ. A note on the measurement of racial integration of schools by means of informational concepts. *J. Math. Sociol.* **1**, 187–194 (1971).
- [168] Tirico M. *Morphogenesis of complex networks : an application in urban growth*. Ph.D Thesis, Université Le Havre Normandie (2020).
- [169] Touraine A. Face à l’exclusion. *Esprit* **169**(2), 7–13 (1991).
- [170] Turing AM. The chemical basis of morphogenesis. *Philos. Transactions Royal Soc. Lond. Ser. B* **237**(641), 37–72 (1952).
- [171] Vieira AP, Goles E, Herrmann HJ. Dynamics of extended Schelling models. *J. Stat. Mech.* **1**, 013212 (2020).
- [172] Vinković D, Kirman A. A physical analogue of the Schelling model. *PNAS* **103**(51), 19261–19265 (2006).
- [173] Watts DJ, Strogatz SH. Collective dynamics of ‘small-world’ networks. *Nature* **393**(6684), 440–442 (1998).
- [174] Whitworth WA. Arrangements of m things of one sort and n things of another sort, under certain conditions of priority. *Messenger Math.* **8**, 105–114 (1878).
- [175] Winship C. A revaluation of indexes of residential segregation. *Soc. Forces* **55**(4), 1058–1066 (1977).
- [176] Young HP. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press (1998).
- [177] Young HP. Social Dynamics: Theory and Applications in *Handbook of Computational Economics*. Elsevier, (2006).
- [178] Zhang J. A dynamic model of residential segregation. *J. Math. Sociol.* **28**(3), 147–170 (2004).
- [179] Zhang J. Residential segregation in an all-integrationist world. *J. Econ. Behav. & Org.* **54**(4), 533–550 (2005).
- [180] Zhang J. Tipping and residential segregation: a unified Schelling model. *J. Reg. Sci.* **51**(1), 167–193 (2011).

Résumé

Cette thèse a pour but d'étudier certains aspects de la ségrégation sociospatiale selon deux axes de recherche complémentaires : la modélisation de son émergence et l'étude statistique de ses structures géographiques.

L'émergence de ségrégation sociospatiale à partir des comportements individuels est étudiée dans le cadre du modèle multi-agents de Schelling. Un vaste panorama des résultats analytiques relatifs à ce modèle et à ses transitions est dressé. Il permet d'identifier deux valeurs remarquables de son paramètre principal, la tolérance des agents, encadrant une phase macroscopiquement ségrégée. L'enjeu des deux chapitres centraux est de déterminer des méthodes de calcul de ces valeurs particulières. Ecrire la chaîne de Markov associée à la densité d'interface du système permet d'en extraire des propriétés menant à la détermination de la première valeur de transition. La seconde est obtenue par l'étude de la stabilité mésoscopique du système, par analogie avec la nucléation homogène en physique statistique. Cette approche thermodynamique permet en outre de caractériser l'entropie du système au cours de son évolution.

Enfin, une méthode multiscalair de détection des hétérogénéités dans une ville réelle est présentée et discutée. La construction d'un faisceau de trajectoires à partir de l'évaluation d'une variable statistique du voisinage local jusqu'à la ville entière pour chaque unité de départ possible révèle la structure sous-jacente de la ville. L'exemple de Paris est alors utilisé pour illustrer la méthode. Pour finir, un modèle probabiliste de ville homogène est développé pour comparer les trajectoires qui en sont issues à celles d'une ville réelle.

Mots-clés modèle de ségrégation de Schelling – ségrégation sociospatiale – milieux hétérogènes – chaînes de Markov – nucléation homogène – systèmes complexes – transitions de phase

Abstract

This thesis aims at studying certain aspects of sociospatial segregation along two complementary research axes: a model of its emergence and the statistical study of its geographical structures.

The emergence of sociospatial segregation from individual behaviours is studied within the framework of the multi-agent Schelling model. A broad spectrum of analytical results related to this model and its transitions is reviewed, showcasing two bounds for the macroscopically segregated phase in terms of the main parameter, the agents' tolerance. Chapters two and three deal with methods to compute these two transition values. The properties of the Markov chain in terms of the interface density are displayed and lead to the first one. The second value is obtained through the study of the system's mesoscopic stability, similarly to homogeneous nucleation theory. This thermodynamical approach also yields the variations of the system's entropy.

Then, a multiscale method to detect spatial dissimilarities in real cities is presented and discussed. Any statistical variable can be measured from the most local scale to the metropolitan one, starting from any unit block, yielding a beam of sequences revealing the city's underlying spatial structure. We study the case of Paris to exemplify this method. Finally, a probabilistic model, the uniform city, is developed to compare its beam of sequences with the one we obtain from real data.

Keywords Schelling segregation model – sociospatial segregation – heterogeneous media – Markov chains – homogeneous nucleation theory – complex systems – phase transitions