



De la structure croissante des réseaux complexes

Approche de la géométrie des réseaux

Mémoire

Charles Murphy

Maîtrise en Physique
Maître ès sciences (M.Sc.)

Québec, Canada

© Charles Murphy, 2018

De la structure croissante des réseaux complexes
Approche de la géométrie des réseaux

Mémoire

Charles Murphy

Sous la direction de:

Louis J. Dubé, directeur de recherche
Antoine Allard, codirecteur de recherche

Résumé

L'internet, le cerveau humain et bien d'autres sont des systèmes complexes ayant un grand nombre d'éléments qui interagissent fortement entre eux selon leur structure. La science des réseaux complexes, qui associe ces éléments et interactions respectivement à des noeuds et liens d'un graphe, permet aujourd'hui de mieux les comprendre grâce aux types d'analyses quantitatives qu'elle rend possible. D'une part, elle permet de définir une variété de propriétés structurelles menant vers une classification de ces systèmes. D'autre part, la compréhension de l'émergence de ces propriétés grâce à certains modèles stochastiques de réseaux devient réalité. Dans les dernières années, un effort important a été déployé pour identifier des mécanismes d'évolution universels pouvant expliquer la structure des réseaux complexes réels.

Ce mémoire est consacré à l'élaboration d'un de ces mécanismes de croissance universels basé sur la théorie de la géométrie des réseaux complexes qui stipule que les réseaux sont des objets abstraits plongés dans des espaces métriques de similarité où la distance entre les noeuds affecte l'existence des liens. Au moyen de méthodes d'analyse avancées, la caractérisation complète de ce mécanisme a été établie et permet le contrôle de plusieurs propriétés structurelles des réseaux ainsi générés. Ce mécanisme général pourrait expliquer, du moins de manière effective, la structure d'un nombre important de systèmes complexes dont la formation est, encore aujourd'hui, mal comprise.

Abstract

The internet and the human brain among others are complex systems composed of a large number of elements strongly interacting according to their specific structure. Nowadays, network science, which construes these elements and interactions respectively as nodes and links of a graph, allows a better understanding of these systems thanks to the quantitative analysis it offers. On the one hand, network science provides the definition of a variety of structural properties permitting their classification. On the other hand, it renders possible the investigation of the emergence of these properties via stochastic network models. In recent years, considerable efforts have been deployed to identify universal evolution mechanisms responsible for the structure of real complex networks.

This memoir is dedicated to one of these universal growth mechanisms based on the network geometry theory which prescribes that real networks are abstract objects embedded in similarity metric spaces where the distance between nodes affect the existence of the links. Thanks to advanced analysis methods, the complete characterization of the mechanism has been achieved and allows the control of structural properties over a wide range. This general mechanism could explain, at least effectively, the structure of a number of complex systems for which the evolution is still poorly understood.

Table des matières

Résumé	iii
Abstract	iv
Table des matières	v
Liste des tableaux	viii
Liste des figures	ix
Liste des abréviations	xii
Liste des symboles et définitions	xiii
Liste des contributions	xvii
Remerciements	xix
Introduction	1
1 La science des réseaux	4
1.1 Concepts de base	4
1.2 Réseaux simples, dirigés et pondérés	5
1.3 Propriétés structurelles des réseaux	5
1.3.1 Degré	5
1.3.2 Corrélacion structurelle	6
1.3.3 Assortativité	7
1.3.4 Agrégation	8
1.4 Propriétés des réseaux réels	10
2 Modèles de réseaux aléatoires	12
2.1 Les modèles fondamentaux	12
2.1.1 Modèle d'Erdős-Rényi	14
2.1.2 Modèle des configurations	15
2.1.3 Modèles de réseaux aléatoires corrélés avec variables cachées	16
2.1.4 Échantillonner les réseaux aléatoires : méthode numérique	20
2.2 Réseaux spatiaux	21
2.3 Réseaux géométriques	22
2.3.1 Espace métrique	23

2.3.2	Espace caché	25
2.3.3	Différentes probabilités de connexion	26
2.3.4	Différentes distributions de noeuds	28
2.3.5	Différents espaces	29
2.4	Modèles de croissance	32
2.4.1	Attachement préférentiel	33
2.4.2	Attachement préférentiel avec densification	36
2.4.3	Règle d'attachement non linéaire	37
2.4.4	Attachement préférentiel spatial	38
3	L'évolution géométrique des réseaux complexes	41
3.1	Avant-propos	42
3.2	Résumé	43
3.3	Abstract	44
3.4	Introduction	45
3.5	Growing geometric networks	46
3.6	Degree Sequence	48
3.6.1	Hidden variables	48
3.6.2	Fixing the degree sequence	49
3.6.3	Scale-free growing geometric networks	50
3.6.4	Finite-size effects	51
3.7	Network history	52
3.7.1	Degree-degree correlations	53
3.7.2	Decreasing order of degrees	54
3.7.3	Increasing order of degrees	54
3.7.4	Random order	55
3.8	Effects of geometry	55
3.8.1	Cold Limit $\tilde{\beta} \rightarrow \infty$	56
3.8.2	Hot limit $\tilde{\beta} \rightarrow \infty$	58
3.8.3	Phase Transition	59
3.9	Conclusion	60
3.10	Complement I : Calculation of $n(\tau)$ on the $(\mathcal{D} + 1)$ -sphere	62
3.10.1	Exact integral form	62
3.10.2	Approximate form	63
3.10.3	Correspondence with existing solutions	64
3.10.4	Linking the exact and approximated solutions	65
3.10.5	Phase transition and dimensionality	66
3.11	Complement II : On the finite size effects	66
3.11.1	Plateau size in the <i>fully-homogeneous</i> limit	67
3.11.2	Plateau size in the <i>fully-heterogeneous</i> limit	68
3.12	Complement III : Evaluation of the parameters	69
3.13	Complement IV : Proof of concept of the geometric evolution inference	71
3.13.1	Maximum likelihood estimation algorithm	71
3.13.2	Network history inference of synthetic networks	72
3.14	Complement V : Solution de la distribution des degrés par équations comparti- mentales	74
3.14.1	Équations compartimentales	75
3.14.2	Solution de l'état à l'équilibre	77

3.14.3 Limites de l'approche par équations compartimentales	78
Conclusions et perspectives	80
Bibliographie	83

Liste des tableaux

1.1	Propriétés structurelles de plusieurs réseaux réels (simples) : le nombre de nœuds N , le nombre de liens M , le degré moyen $\langle k \rangle$ donné par Éq. (1.4), l'exposant de la distribution des degrés si elle est en loi de puissance $P(k) \propto k^{-\gamma}$ ("-" dans le cas contraire), le coefficient d'agrégation moyen C donné par Éq. (1.11), le coefficient d'agrégation local moyen $\langle c \rangle$ donné par Éq. (1.14) et le coefficient d'assortativité par degré r donné par Éq. (1.10). On laisse une case vide lorsque la propriété n'est pas accessible. Ce tableau est tiré de Réf. [86, 91] et a été modifié par soucis d'uniformité.	10
2.1	Propriétés caractéristiques des géométries euclidienne, sphérique et hyperbolique. Ce tableau est tiré de la Réf. [75].	23
2.2	Quelques exemples de probabilités de connexion analysées dans la Réf. [37]. La zone en bleu indique une région où $p(\mathbf{x}, \mathbf{y})$ est non nulle : l'intensité du bleu désigne la valeur de la probabilité.	27

Liste des figures

1.1	Schématisation d'un réseau composé de $N = 10$ nœuds.	5
1.2	Exemple de réseaux (a) assortatifs ($r > 0$) et (b) désassortatifs ($r < 0$). Ces réseaux sont tirés des Refs. [86, 95] et ont été créé à partir de l'algorithme spécifié dans les Refs. [89, 90].	9
2.1	Exemple de réseau géométrique aléatoire : (a) réseau géométrique dont la distribution des nœuds est uniforme et \mathcal{M} est un disque de rayon $R = 1$ identifié par la zone grise, (b) schéma du processus de connexion où la zone en bleu indique l'étendue du voisinage d'un nœud v_i . On a fixé $N = 300$ et $\mu = 0.18$	22
2.2	Exemples de \mathcal{D} -sphère. La 0-sphère correspond à un point, la 1-sphère, à un cercle dans un espace euclidien en 2D et la 2-sphère, à une sphère dans un espace euclidien en 3D.	24
2.3	Schématisation de la géométrie hyperbolique : (a) disque de Poincaré, (b) œuvre de Esher appelée <i>Circle limite III</i> illustrant un espace hyperbolique (les poissons sont tous de la même taille dans cet espace) et (c) exemple de géométrie hyperbolique dans la nature, le chou kale frisé.	25
2.4	RGGs dans des espaces avec rebord : (a) espace euclidien \mathbb{R}_R^2 avec $N = 500$, $R = 1$ et on fixe $\langle k \rangle = 25$ à l'aide de $p \simeq 0.085$ [Éq. (2.53)], (b) espace hyperbolique \mathbb{H}_R^2 avec $N = 500$, $p = 1$ et on fixe $\langle k \rangle = 25$ avec $R \simeq 7.86$ [Éq. (2.55)]. On montre également le degré moyen $\kappa(r)$ d'un nœud selon sa position radiale r . Sur les graphiques, les points bleus sont les données issues du réseau correspondant et les lignes oranges sont les solutions analytiques [Éq. (2.52) pour (a) et Éq. (2.54) pour (b)]. Dans chacun des schémas, on identifie un nœud aléatoirement en rouge ainsi que ses liens. Son volume exclu est également identifié par la région bleue délimitée par la ligne blanche.	31
2.5	Correspondance entre les géodésiques et les parcours les plus courts sur \mathbb{H}_R^2 . On identifie en rouge trois noeuds dont on a mis en évidence les parcours les plus courts en rouge également. Les lignes courbes pointillées en bleu correspondent aux géodésiques entre ces noeuds. Dans ce réseau, on fixe $N = 1000$, $\langle k \rangle = 15$ et $\gamma = 2.2$	32
2.6	Propriétés structurelles dans le MBA : (a) distribution des degrés $P(k)$, (b) corrélation des degrés donnée par $\bar{k}_{nn}(k)$, (c) spectre du coefficient d'agrégation dépendant du degré $\bar{c}(k)$. On indique, par les points bleus, les données tirées de réseaux générés par ce modèle et, par les lignes pleines, les solutions analytiques associées [Éq. (2.71) pour (a) et Éq. (2.72) for (b)]. On fixe $N = 10^4$ et $m = 10$ et on moyenne ces résultats sur 100 simulations.	35

3.1	Illustration of the HA growth mechanism for geometric networks : (a) birth of nodes v_t at time t which connects with three neighbors, including v_s at distance $d_{ts} < \mu(t)$, (b) subsequent birth of node $v_{t'}$, such that $t' > t$, connecting to v_t with $d_{tt'} < \mu(t')$. For illustration purposes, $p(\mathbf{x}_t, \mathbf{x}_s; \mu(t)) = \Theta(\mu(t) - d_{ts})$ such that any node within a circle of radius $\mu(t)$ (gray area) centered on node v_t will become connected to it.	47
3.2	Expected degree $\kappa(\tau)$ as a function of the birth time τ for different parameters. The small dots correspond to data obtained from Monte Carlo simulations averaged over 100 instances where $\tilde{\mu}(\tau)$ is given by Eq. (3.13) with $n(\tau)$ given by Eq. (3.15). The markers (squares, circles and triangles) correspond to the average data. The black solid lines corresponds to Eq. (3.14). The parameters of each dataset are indicated on the plot. The networks have been averaged over 96 instances.	51
3.3	Chemical potential $\tilde{\mu}(\tau)$ vs τ used to generate scale-free geometric networks with different values of α and $\langle k \rangle$ indicated on the plot. The lines and markers correspond to Eq. (3.13) with $n(\tau)$ given by Eq. (3.15). We fixed $N = 10^4$ and $\tilde{\beta} \rightarrow \infty$	52
3.4	Degree-degree correlations for different histories : (a) decreasing degree order (Sec. 3.7.2), (b) increasing degree order (Sec. 3.7.3), (c) random order (Sec. 3.7.4). In each plot, the results from Monte Carlo simulations are indicated by the squares and the analytics by the circles. The degree distributions are also displayed in inset and are power laws with $\gamma = 2.25$ (dashed lines). The average degree assortativity coefficient $\langle r \rangle$ amounts to (a) $\langle r \rangle \simeq 0.68$, for (b) $\langle r \rangle \simeq -0.12$ and for (c) $\langle r \rangle \simeq -0.11$. For each simulation, $\alpha = 0.8$, $\langle k \rangle = 10$, $N = 10^4$ and $\tilde{\beta} \rightarrow \infty$ (Eq. (3.1) as the connection probability). The results have been averaged over 96 instances.	53
3.5	Geometrical representation of Eq. (3.29) with $\mathcal{D} = 1$ in the limit $\tilde{\beta} \rightarrow \infty$. The constraints of the domain are illustrated in shaded parallelograms and the overlapping area corresponds to the resulting value of Eq. (3.29). In this illustrative case, $\tilde{\mu}(\tau) > \tilde{\mu}(\tau')$ and $\tilde{\mu}(\tau'') > \tilde{\mu}(\tau')$	56
3.6	Average clustering coefficient $\langle c \rangle$ as a function of α in the cold regime ($\tilde{\beta} \rightarrow \infty$ with Eq. (3.1) as the connection probability). The solid line corresponds to the numerical integration of Eq. (3.33) and the squares to data from Monte-Carlo simulations. We fixed $N = 10^4$ and the results have been averaged over 96 instances.	57
3.7	Average clustering coefficient $\langle c \rangle$ as a function of the number of nodes N in the hot regime ($\tilde{\beta} \ll 1$). The solid line corresponds to the numerical integration of Eq. (3.38) and the square to data from Monte-Carlo simulations. The scaling of $\langle c \rangle$ is indicated by the black dashed line. We fixed $\alpha = 0$, $\langle k \rangle = 10$ and $\tilde{\beta} = 10^{-3}$. The results have been averaged over 96 instances.	58
3.8	Transition between the hot and the cold regime for different degree sequence heterogeneities : (a) homogeneous degree sequence, $\alpha = 0$, $\langle k \rangle = 10$ and $N = 10^4$, (b) heterogeneous degree sequence, $\alpha = 0.5$, $\langle k \rangle = 10$ and $N = 10^4$. The dashed lines (above) correspond to the cold limit theoretical result while the dotted lines (below) correspond to the hot limit one. The data from Monte-Carlo simulations is denoted by the squares. The results have been averaged over 96 instances.	59
3.9	Average probability $n(\tau)$ as a function of the inverse temperature $\tilde{\beta}$. We fixed $\tilde{\mu}(\tau) = 10^{-2}$. The lines correspond to the exact expression Eq. (3.50) while the dashed lines correspond to the approximated solution Eq. (3.60). The dotted line corresponds to the transition threshold $\tilde{\beta} = \tilde{\beta}_c$ and the shadowed area corresponds to the non-physical case $n(\tau) > 1$	66

3.10	Plateau size τ^* scaling with N in the fully-homogeneous limit : The solid lines correspond to the numerical solution of Eq. (3.69) and the circles correspond to the asymptotic solution Eq. (3.70) in limit $\alpha \rightarrow 0$. The scaling of $N^{-\delta}$ is shown as a reference with the dashed line where $\delta = 1$	67
3.11	Plateau size τ^* scaling with N in the fully-heterogeneous limit : (a) Lambert W-function with the two real branches $k = -1$ and $k = 0$ [96], (b) τ^* as a function of N for different values of $\epsilon \ll 1$. Similarly to Fig. 3.10, the solid lines correspond to the numerical solution of Eq. (3.69) and the triangles correspond of the asymptotic Eq. (3.72) in the limit $\alpha \rightarrow 1$. The scaling with N , as indicated by the dashed line, has an exponent $\delta \approx 0.54 > \frac{1}{2}$	68
3.12	Average degree $\langle k \rangle$ as a function of α : (a) we fixed $\nu = \bar{k}(1 - \alpha)$, (b) we used the numerical procedure elaborated in Sec. 3.12 using $\epsilon = 10^{-6}$. On both (a) and (b), the circles corresponds to the expected average degree calculated from the analytical modified degree sequence and the squares corresponds to the average degree of generated networks with their respective error bars given by the standard deviations. The error bars on (a) are present but too small to appear. The dashed line corresponds to the desired value of $\langle k \rangle = \bar{k}$. For the generated networks, we used $\bar{k} = 10$, $N = 10^3$ and $\tilde{\beta} \rightarrow \infty$. The results are averaged over 96 instances.	70
3.13	Inference of synthetic scale-free geometric networks : (Top) decreasing order of degrees history, (Bottom) random order history. On the left panels, we considered the initialization $H(0)$ in which $\tau_i(0) \sim U(0, 1)$ for all i , where $U(0, 1)$ is a uniform distribution defined between 0 and 1. The right panels show the inference when $H(0)$ is the ground truth. The dashed line is a reference to the target $\tilde{\mu}_{\text{infer}}(\tau) = \tilde{\mu}_{\text{real}}(\tau)$. In the insets is displayed the inferred birth time in H_{infer} as a function of the ground truth birth time in H_{real} . The correlation coefficient for H and $\tilde{\mu}(\tau)$ corresponds to the Pearson coefficient between the inferred and ground truth datasets. The networks are generated with $\alpha = 0.83$, $\gamma = 2.2$, $\langle k \rangle = 35$, $N = 10^2$ and $\tilde{\beta} = 10^5$. The inference data (H_{infer} and $\tilde{\mu}(\tau)_{\text{infer}}$) has been averaged over $M = 200$ histories.	73
3.14	Distribution des degrés $P(k)$ de réseaux synthétiques comparée avec la solution $P_k(t = N)$ donnée par les équations compartimentales. Les points correspondent aux résultats issus de simulations Monte-Carlo moyennés sur 96 instances alors que les lignes pleines sont les solutions des équations compartimentales itérées. On indique également, avec les lignes pointillées noires, le comportement de la distribution des degrés en loi de puissance avec un exposant $\gamma = 1 + \frac{1}{\alpha}$. On considère $N = 10000$ et $\tilde{\beta} \rightarrow \infty$. Le reste des paramètres du modèle sont indiqués sur le graphique.	76

Liste des abréviations

Abbréviations françaises

MER	Modèle d'Erdős-Rényi
MBPS	Modèle de Boguñá-Pastor-Satorras
MBA	Modèle de Barabási-Albert
MKR	Modèle de Krapivsky-Redner

Abbréviations anglaises

ANND	Average Nearest-Neighbor Degree
ERGM	Exponential Random Graph Model
CM	Configuration Model
RGG	Random Geometric Graph
PA	Preferential Attachment
SPA	Spatial Preferential Attachment
HA	Homogeneous Attachment
MLE	Maximum Likelihood Estimation

Liste des symboles et définitions

Dans ce mémoire, nous adoptons la convention que les vecteurs sont écrits en gras avec des **lettres minuscules**. Par exemple, on écrit un vecteur de taille N sous la forme suivante

$$\mathbf{v} = (v_1, v_2, \dots, v_N)^\top .$$

Similairement, les matrices seront écrites en gras avec des **lettres majuscules**, telle que M . Ensuite, on utilisera les accolades $\{ \dots \}$ afin de désigner les ensembles (*set*), et les crochets (\dots) afin de désigner les ensembles ordonnés (*tuple*).

Nous utilisons également la notation *Big O* de Landau

$$f(x) = \mathcal{O}(g(x)) \tag{1}$$

qui spécifie qu'une fonction $f(x)$ varie comme une autre fonction $g(x)$ à une constante mutliplicative près dans la limite où $x \rightarrow \infty$ telle que

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} < \infty . \tag{2}$$

Notation du chapitre 1

Définition d'un réseau

v_i	Noeud indicé i .
a_{ij}	Lien ($v_j \rightarrow v_i$) partant de v_j vers v_i .
\mathcal{V}	Ensemble des noeuds.
\mathcal{E}	Ensemble des liens.
A	Matrice d'adjacence.
N	Nombre de noeuds (correspondant à la cardinalité de \mathcal{V}).
M	Nombre de liens (correspondant à la cardinalité de \mathcal{E}).

$G(\mathcal{V}, \mathcal{E})$	Réseau composé de l'ensemble de noeuds \mathcal{V} et de l'ensemble de liens \mathcal{E} .
ω_{ij}	Poids du lien entre les noeuds v_i et v_j .
Propriétés structurelles	
k_i	Degré du noeud v_i .
\mathbf{k}	Séquence des degrés : vecteur contenant le degré de chaque noeud.
N_k	Nombre de noeuds de degré k .
$P(k)$	Distribution des degrés : probabilité de sélectionner aléatoirement un noeud de degré k .
γ	Exposant d'une distribution des degrés en loi de puissance $P(k) \propto k^{-\gamma}$.
$\langle k \rangle$	Degré moyen d'un réseau.
$P(v_i, v_j)$	Corrélation : probabilité que les noeuds v_i et v_j soient connectés.
$E_{kk'}$	Nombre de liens entre les ensembles des noeuds de degré k et k' .
$P(k, k')$	Corrélation degré-degré : probabilité que deux noeuds de degré k et k' soient connectés.
$\bar{k}_{nn}(k)$	Degré moyen des plus proches voisins des noeuds de degré k .
r	Coefficient d'assortativité par degré.
C	Coefficient d'agrégation moyen.
c_i	Coefficient d'agrégation local du noeud v_i .
$\bar{c}(k)$	Coefficient d'agrégation local moyen des noeuds de degré k .
$\langle c \rangle$	Coefficient d'agrégation local moyen.

Notation du chapitre 2

Ensembles de réseaux aléatoires

\mathcal{G}	Ensemble de réseaux d'un modèle.
$\mathbb{P}(G)$	Probabilité de générer le réseau G à partir d'un modèle donné.
Θ	Ensemble des paramètres d'un modèle.
Z	Fonction de partition.

$H(G)$	Hamiltonien d'un réseau G .
p_{ij}	Nombre de liens moyens en deux noeuds v_i et v_j dans \mathcal{G} .
κ_i	Degré espéré du noeud v_i .
$\kappa_{nn,i}$	Degré espéré des premiers voisins du noeud v_i .

Réseaux géométriques

\mathcal{M}	Espace métrique.
$V_{\mathcal{M}}$	Volume total de l'espace métrique \mathcal{M} .
$d(\mathbf{x}, \mathbf{y})$	Distance entre les points $\mathbf{x}, \mathbf{y} \in \mathcal{M}$.
$\rho(\mathbf{x})$	Distribution spatiale des noeuds.
$p(\mathbf{x}, \mathbf{y})$	Probabilité de connexion entre deux noeuds placés en \mathbf{x}, \mathbf{y} .
μ	Seuil de connexion d'un RGG.
\mathcal{D}	Dimension de \mathcal{M} .
$\mathbb{S}^{\mathcal{D}}$	\mathcal{D} -Sphère.
$\mathbb{H}^{\mathcal{D}}$	\mathcal{D} -Boule de Poincaré.
R	Rayon d'une $(\mathcal{D} + 1)$ -sphère.

Modèle de croissance

t	Temps d'évènement de croissance : compteur des évènements de croissance compris dans l'intervalle $[0, N - 1]$.
q	Probabilité de naissance.
$N_k(t)$	Nombre de noeuds moyen de degré k au temps t .
$p_k(t)$	Distribution moyenne des degrés au temps t .

Notation du chapitre 3

Évolution géométrique des réseaux complexes

$\mu(t)$	Potentiel chimique dépendant du temps : paramètre du modèle qui contraint la portée spatiale des liens créés au temps t .
β	Température inverse du réseau : paramètre du modèle qui contrôle le couplage

entre l'espace métrique et la structure du réseau.

τ	Temps d'évènements de croissance rééchélonné par rapport à N et compris dans l'intervalle $[0, 1)$.
$\tilde{\mu}(\tau)$	Potentiel chimique dépendant du temps rééchélonné par rapport à πR .
$\tilde{\beta}$	Température inverse rééchélonnée par rapport à πR .
\tilde{R}	Rayon rééchélonné d'une $(\mathcal{D} + 1)$ -sphère de rayon R par rapport à πR .
$n(\tau)$	Probabilité de connexion moyenne du noeud τ .
H	Histoire d'un réseau : tuple $(\tau_i), i = 0, \dots, N - 1$ des temps rééchélonnés de chaque noeud dans le réseau.
$\ell(\tau, \tau', \tau'')$	Probabilité de connexion moyenne des noeuds τ, τ' et τ'' (corrélation à trois noeuds).
$N(\tau)$	Nombre de noeuds existants au temps rééchélonné τ .
$\kappa^*(\tau)$	Degré espéré à la naissance du noeud τ : Nombre de liens créés au moment de l'apparition du noeud τ dans le réseau.

Liste des contributions

Articles

- **C. Murphy**, A. Allard, E. Laurence, G. St-Onge L. J. Dubé,
Geometric Evolution of Complex Networks with degree correlations,
Phys. Rev. E **97** (2018), p. 032309.
- G. St-Onge, J.-G. Young, E. Laurence, **C. Murphy** L. J. Dubé,
Phase transition of the susceptible-infected-susceptible dynamics on time-varying configuration model networks,
Phys. Rev. E **97** (2018), p. 022305.
- J.-G. Young, L. Hébert-Dufresne, E. Laurence, **C. Murphy**, G. St-Onge, P. Desrosiers,
Network archaeology : phase transition in the recoverability of network history,
arXiv:1803.09191 (2018).

Conférences

- **C. Murphy**, E. Laurence, G. St-Onge, J.-G. Young, L. J. Dubé,
Time-dependent spatial growth of complex networks,
NetSci (2016), Séoul, Corée du sud.
- **C. Murphy**, A. Allard, E. Laurence, G. St-Onge, L. J. Dubé,
Time-dependent connection threshold in growing random geometric graphs,
NetSci (2017), Indianapolis, États-Unis.
- **C. Murphy**, A. Allard, E. Laurence, G. St-Onge, L. J. Dubé,
Geometric evolution of complex networks,
Macfang (2017), Barcelone, Espagne.

On va toujours trop loin pour les
gens qui [*sic*] vont nulle part.

Pierre Falardeau

Remerciements

Un travail de recherche dans sa forme finale est toujours plus compact, simple et élégant qu'il ne l'a été tout au long de son élaboration. Parfois, on emprunte des chemins compliqués qui peuvent mener à des culs-de-sac. Parfois, on utilise une avenue qui s'avère être infiniment plus simple que la précédente pour exprimer la même idée. En bref, la recherche est une marche aléatoire dans un grand espace compliqué où l'on s'égaré facilement quand on s'y aventure seul. Ce mémoire est donc l'aboutissement de ma propre marche aléatoire supportée, de près ou de loin, par plusieurs personnes avec qui j'ai échangé ou qui m'ont encouragé. Je dédie donc ces quelques lignes à eux pour les remercier.

Bien qu'il est naturel de débiter avec le directeur de recherche, je le ferai, non par convention, mais par son mérite. Ainsi, je remercie Louis J. Dubé pour sa curiosité, son ouverture et son dévouement, mais surtout pour sa confiance et son sens critique. Son enseignement, parfois rude et difficile, s'est toujours avéré juste et est à la base du chercheur que je suis devenu. Le souci du détail, de désir de faire plus et mieux, la confiance en mes capacités, ce sont toutes des qualités qu'il m'a transmises durant ma formation. De plus, le coeur de ce mémoire qui porte ses couleurs n'aurait sans doute jamais vu le jour si ce n'avait été de sa persistance à ce que je poursuive sur cette voie. Enfin, bien que son support fut en grande partie au niveau académique, combien de fois nous sommes-nous assis autour d'un café pour discuter des *romains*, de *Genghis Khan* et de son plus jeune petit fils, *Kublai*, de ses mésaventures de jeunesse ou de tous autres sujets enclins à détendre notre neurone actif. Je suis extrêmement honoré d'avoir été sous sa supervision et lui en suis éternellement reconnaissant.

Ensuite, j'aurais préféré que cette liste de remerciement puisse être parallélisée afin de mettre sur ce même pied d'estale mon excellent codirecteur, Antoine Allard. Je le remercie donc pour sa patience, sa générosité, sa rigueur, son humanité et son sens de l'humour. Je suis également honoré d'avoir été sous sa supervision et d'avoir été son premier étudiant gradué. Les prochains étudiants qu'il supervisera auront un privilège, comme moi, de travailler (et, certainement, de faire le *vermut*) avec un chercheur d'exception.

Je tiens également à remercier messieurs Patrick Desrosiers et Nicolas Doyon pour leur enthousiasme et pour avoir accepté de siéger au comité d'évaluation de ce mémoire.

J'aimerais maintenant tourner mon attention vers amis et collègues de la confrérie *Dynamica* avec

qui j'ai partagé maintes et maintes discussions. D'abord, je remercie le doyen du groupe, Jean-Gabriel, pour sa rigueur scientifique et ses expressions préférées qui me font beaucoup rire : "*absolument pas*", "*c'est convolué tout ça*", "*aight*" et j'en passe. Je remercie ensuite Edward pour son humour cynique (presque malsain) qui déteint un peu sur moi et que j'apprécie beaucoup, Guillaume, mon partenaire avec qui je partage une rivalité saine qui nous pousse à nous dépasser, et Jean-Luc qui parvient à faire survivre et même fructifier la section *lumineuse* du groupe. Je remercie finalement les deux nouveaux, Vincent et Xavier, avec qui j'aurai le plaisir de travailler plus ardemment dans un futur proche.

Bien que le domaine universitaire ait eu un impact important sur mes travaux, celui de ma vie privée a été tout aussi important. C'est pourquoi je souhaite remercier mes amis de longue date, Alexis, Antoine, Charles-David et Edouard, qui, malgré la distance entre nos domaines respectifs, ont toujours trouvé un intérêt dans mes projets ; mes amis du baccalauréat, Félix, Simon, Shanny, Maxime et Jean-Samuel, avec qui j'ai partagé des moments formateurs qui nous ont rapprochés ; et mes amis de l'université, Pierre-Luc et Geneviève, qui m'ont supporté dans les moments les plus difficiles.

Finalement, les seuls qu'il resterait à remercier, qui sont sans doute les plus importants, sont mes parents, Martine et Paul, et ma soeur, Frédérique. Depuis mon plus jeune âge, ils ont toujours cru en moi et ont toujours été derrière moi pour qu'aucune limite ne m'atteigne. Si je suis la personne que je suis aujourd'hui, c'est grâce à eux. Je vous remercie du plus profond de mon coeur.

Introduction

La science du 20^e siècle a vu naître des avancées importantes notamment en physique théorique au niveau des domaines spatio-temporels du microscopique, de la physique quantique et du modèle standard [34], et du macroscopique, de la relativité générale et de la cosmologie [29]. Bien que mathématiquement sophistiquées et complexes, ces disciplines se fondent sur une poignée de lois logiques qui permettent d'élaborer entièrement ces théories. De l'équation de Schrödinger, on peut déduire la dualité onde-particule d'un photon et d'un électron ; de l'équation d'Einstein, on peut déduire l'existence des trous noirs et des ondes gravitationnelles. Ces systèmes sont dits, au sens de la difficulté avec laquelle on les représente, *simples* et *désorganisés*.

À mi-chemin entre ces domaines extrêmes se trouve l'échelle mésoscopique où co-évoluent le règne animal, le cerveau humain et les organisations sociales et technologiques qui s'y rattachent. À cette échelle, bien que tous ces systèmes soient constitués des mêmes particules, leur nature devient absolument fondamentale et trouver un cadre conceptuel universel pour les décrire avec précision devient une tâche d'envergure tout aussi importante. Pourquoi un jaguar est-il différent d'un tigre ? De poser ce genre de questions, vaines aux premiers abords, revient à nous questionner sur l'origine de cette différence et, tel que suggérer par Murray Gell-Mann [55], revient à demander : *pourquoi le jaguar n'est-il pas simplement la somme de ses quarks* ? De façon plus formelle, comment caractérise-t-on les systèmes appartenant au domaine mésoscopique et peut-on prédire leur évolution aussi précisément que dans les domaines microscopique et macroscopique ? Ces systèmes sont conceptuellement différents de ceux discutés en physique habituellement, car ils sont *complexes* et *organisés*, c'est-à-dire que leur organisation ou leur structure a un impact sur leur comportement. Ces systèmes, qui ont notamment été longtemps étudiés de manière qualitative en sciences humaines, en biologie et bien d'autres domaines, s'inscrivent dans la classe des *systèmes complexes*.

Un système complexe, selon Hebert A. Simon [114], correspond à tout système *i.* composé d'un grand nombre (virtuellement infini) d'éléments en interaction, *ii.* ayant une structure et une organisation hiérarchique et *iii.* possédant des propriétés globales et émergentes (qu'elles soient structurelles ou dynamiques) ne pouvant être déduites des seules propriétés de chacun des éléments du système. En mécanique statistique, de travailler sur des systèmes composés d'un grand nombre d'éléments en interaction très faible sans structure, par exemple les gaz de particules, est chose courante. Or, les systèmes complexes ont en général une structure non triviale et très irrégulière qui nous empêche

d'utiliser la mécanique statistique adéquatement. Pour y encapsuler leur structure correctement, les scientifiques du 21^e siècle ont introduit la notion de réseau, dont les éléments sont les noeuds et les liens, leurs interactions, donnant naissance à la *théorie des réseaux complexes*.

Le cadre conceptuel des réseaux complexes [86] unifie la description d'une majorité de systèmes complexes comme des réseaux organisés de noeuds (la *structure*) qui échangent une quelconque information entre eux (la *dynamique*) via leurs liens. La structure, la dynamique et la relation entre les deux permettent ainsi de caractériser et de classer les réseaux indépendamment de leur nature et ont permis d'orienter la recherche dans les avenues structurelles et dynamiques. D'une part, l'étude de la structure s'intéresse, dans un premier temps, à sa caractérisation en définissant une variété de propriétés structurelles et, dans un deuxième temps, à comprendre comment ces propriétés structurelles émergent. D'autre part, l'étude de la dynamique s'intéresse à rattacher les différents comportements dynamiques de ces systèmes à leur structure. À l'heure actuelle, ces deux disciplines abondent de défis conceptuels et théoriques d'envergure étroitement liés à des problèmes d'importance capitale tels que la compréhension du cerveau humain et de ses maladies neurologiques.

Dans ce mémoire, une attention particulière sera portée à la structure des réseaux complexes réels et à ses origines. *Pourquoi observe-t-on certaines propriétés structurelles ?* Cette question directrice a été à l'origine de plusieurs trouvailles intéressantes en théorie des réseaux complexes. Les réseaux complexes sont constamment en évolution : de nouveaux noeuds et liens s'ajoutent et se retirent à chaque moment selon des mécanismes particuliers. Cependant, bien que leur structure soit constamment en changement et que les mécanismes d'évolution puissent être radicalement différents, ils possèdent des propriétés structurelles communes : la distribution des liens entre les noeuds est hétérogène, les noeuds ont une préférence à se connecter avec d'autres noeuds structurellement similaires donnant naissance à des amas, il existe des communautés plus densément connectées de noeuds, etc.

Pour rendre compte de ces propriétés structurelles, des modèles stochastiques de croissance de réseaux ont été élaborés et étudiés. Parmi tous ces modèles, l'attachement préférentiel est sans nul doute le plus connu [11, 113]. Dans sa forme la plus simple, un réseau est créé par attachement préférentiel lorsque de nouveaux noeuds sont créés à chaque temps et se connectent aux noeuds déjà existants sur le réseau avec une probabilité proportionnelle au nombre de liens que possèdent ces noeuds. Ce mécanisme fort simple offre une explication *effective* de l'origine de la distribution hétérogène des liens dans les réseaux réels. Cependant, il ne permet pas de gérer les autres propriétés observées.

Pour y parvenir, une solution élégante a été proposée : les réseaux réels pourraient être "cartographiés" dans des espaces géométriques où la distance entre les noeuds encode directement leur structure [75, 105]. De posséder cette carte beaucoup plus simple à manipuler reviendrait ainsi à connaître entièrement la structure sans pour autant prendre explicitement connaissance. Cette approche s'appelle la *géométrie des réseaux* et explique pourquoi les noeuds similaires possèdent une préférence à se connecter entre eux au moyen d'arguments géométriques.

En combinant l'attachement préférentiel à l'approche géométrique, il a ainsi été possible de créer des modèles de croissance réalistes pouvant être ensuite utilisés pour inférer de manière explicite l'évolution et les cartes géométriques de plusieurs réseaux réels [98, 97, 99]. Bien que ceci soit une prouesse remarquable, ces modèles sont toujours sujets à plusieurs limitations et ne permettent pas de comprendre l'entièreté des propriétés structurelles globales que l'on observe. C'est dans cette veine que le but de ce mémoire se définit : *développer un mécanisme général d'évolution de réseaux complexes qui puisse rendre compte de toutes les propriétés structurelles globales des réseaux complexes.*

Ce mémoire présente nos efforts vers l'accomplissement de cette tâche d'envergure. Au Chapitre 1 sont présentées les notions fondamentales de la science des réseaux complexes. On y introduit notamment quelques définitions et plus particulièrement les propriétés structurelles fondamentales, leur formulation mathématique ainsi que leur correspondance chez une variété de réseaux réels. Le Chapitre 2 offre ensuite une revue des outils de modélisation de ces systèmes en termes d'ensemble de réseaux aléatoires. Les modèles de réseaux aléatoires fondamentaux sont d'abord présentés suivis des modèles géométriques et des modèles de croissance importants. Au Chapitre 3, je présente nos contributions originales pour l'élaboration d'un mécanisme d'évolution nommé *évolution géométrique par attachement homogène*. Une analyse approfondie des propriétés structurelles de ce modèle y est développée sous la forme d'un article en anglais publié chez *Physical Review E*. Finalement, le mémoire est complété d'une conclusion où on y résume les résultats obtenus et y élabore certaines perspectives futures en continuité avec le sujet de ce mémoire.

Chapitre 1

La science des réseaux

Dans ce premier chapitre, on introduit les concepts mathématiques fondamentaux sur lesquels les travaux réalisés sont basés. En particulier, aux Sections 1.1 et 1.2, on introduit la terminologie propre à la caractérisation de la structure d'un réseau complexe. Ensuite, la Sec. 1.3 est consacrée à la définition des propriétés structurelles les plus répandues en science des réseaux. Ces propriétés nous permettront de quantifier plus simplement les réseaux complexes auxquels nous serons confrontés et d'identifier des critères pour orienter notre recherche dans l'élaboration d'un mécanisme de croissance réaliste. On termine ce chapitre avec la Sec. 1.4 offrant une vue d'ensemble des propriétés de plusieurs réseaux complexes réels.

1.1 Concepts de base

Un réseau (ou graphe¹) est un objet mathématique abstrait composé de *nœuds* et de *liens* (*nodes* et *links* en anglais). Introduits pour la première fois par Léonhard Euler dans le problème des *sept ponts de Königsberg* [16], les réseaux sont utilisés pour modéliser et décrire la structure de systèmes composés d'un nombre N d'éléments en interaction dont l'organisation de ces interactions est arbitraire. On associe ces éléments à un ensemble de N nœuds, noté $\mathcal{V} = \{v_i\}$ où $i = 1, 2, \dots, N$, et leurs interactions à un ensemble de M liens, noté $\mathcal{E} = \{(v_j \rightarrow v_i)\}$, où $(v_j \rightarrow v_i)$ dénote un lien partant de v_j vers v_i , pour toutes paires connectées de nœuds v_i et v_j . On définit l'élément $a_{ij} \in \mathbb{N}$ qui quantifie le lien $(v_j \rightarrow v_i)$, où $a_{ij} > 0$ lorsque qu'au moins un lien existe et $a_{ij} = 0$ autrement. Par concision, au lieu d'écrire $(v_j \rightarrow v_i)$, on utilisera simplement a_{ij} comme étant le lien qui relie v_i et v_j . On définit également la *matrice d'adjacence* $\mathbf{A} = (a_{ij}) \in \mathbb{N}^{N \times N}$ qui caractérise la *structure* (ou *topologie*), c'est-à-dire l'organisation et l'arrangement des liens du réseau. Ainsi, un réseau défini par les ensembles \mathcal{V} et \mathcal{E} est noté $G = (\mathcal{V}, \mathcal{E})$.

1. Tout au long de cet ouvrage, aucune distinction ne sera faite entre réseau et graphe. Par contre, le terme « graphe » est généralement utilisé dans un contexte formel et mathématique alors que le terme « réseau » est utilisé dans un contexte de modélisation.

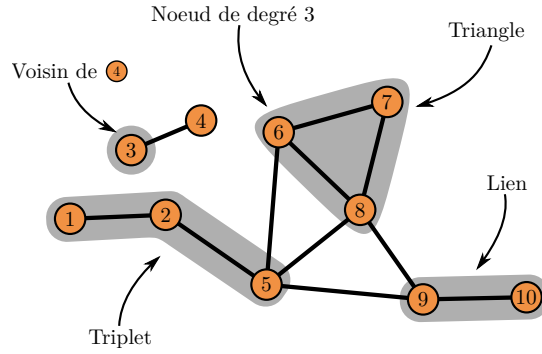


FIGURE 1.1 – Schématisation d'un réseau composé de $N = 10$ nœuds.

1.2 Réseaux simples, dirigés et pondérés

Les réseaux peuvent prendre plusieurs formes [86]. Par exemple, il est possible que les liens soient asymétriques, c'est-à-dire $a_{ij} \neq a_{ji}$, ou qu'ils soient pondérés par une valeur $\omega_{ij} \in \mathbb{R}$ nommée le poids du lien a_{ij} , lorsque les interactions sont inégales, inhibitrices ($\omega_{ij} < 0$) ou excitatrices ($\omega_{ij} > 0$). On appelle ces types de réseaux dirigés et pondérés respectivement. Dans sa forme la plus simple, les liens d'un réseau ne sont ni dirigés, ni pondérés et a_{ij} peut prendre les valeurs 0 ou 1 (aucun multilien) lorsque $i \neq j$ et $a_{ii} = 0$ pour tout i (aucune boucle) autrement. On dit de ces réseaux qu'ils sont *simples*. Bien que leurs structures soient moins sophistiquées que ces variantes dirigée et pondérée, ils demeurent un sujet d'étude actuel puisque leur structure est plus accessibles. Dans ce mémoire, il ne sera question que de réseaux simples.

1.3 Propriétés structurelles des réseaux

Bien que la structure d'un réseau soit entièrement représentée par sa matrice d'adjacence \mathbf{A} , il est souvent plus commode de la décrire en termes de quantités qui sont des fonctions de \mathbf{A} . Du point de la modélisation, ces quantités sont généralement plus intuitives que \mathbf{A} . On dit de ces quantités qu'elles sont les *propriétés structurelles* du réseau [91]. Dans cette sous-section, les propriétés structurelles les plus importantes sont présentées.

1.3.1 Degré

Le *degré* d'un nœud v_i , noté k_i , correspond au nombre de nœuds auxquels il est attaché. Comme le degré est une propriété qui caractérise la structure à un niveau très rudimentaire, c'est-à-dire à l'échelle des nœuds, il est fondamental dans la description de la structure. On observe d'ailleurs que les nœuds de haut degré, les *hubs*, sont généralement les importants du réseau. Par exemple, sur la Fig. 1.1, on voit que le degré du nœud 6 est égal à 3. En termes de la matrice d'adjacence, le degré k_i du nœud v_i s'écrit

$$k_i = \sum_{j=1}^N a_{ij} \quad \Rightarrow \quad \mathbf{k} = \mathbf{A} \mathbf{1} \quad (1.1)$$

où $\mathbf{1} = (1, 1, \dots, 1)^\top$ est un vecteur colonne de 1. Du degré de chaque nœud, on peut calculer le nombre total de liens, M ,

$$M = \sum_{i=1}^N \frac{k_i}{2} = \frac{1}{2} \mathbf{1}^\top \mathbf{A} \mathbf{1} \quad (1.2)$$

où le facteur $\frac{1}{2}$ vient du fait qu'un lien est compté deux fois dans la sommation. Pour un ensemble de nœuds, on définit la *séquence des degrés* comme étant un vecteur $\mathbf{k} = (k_1, k_2, \dots, k_N)^\top$ dont les éléments sont les degrés de chaque nœud. Par contre, il est souvent plus pratique de travailler avec la *distribution des degrés*, notée $P(k)$, qui s'écrit

$$P(k) = \frac{N_k}{N} = \frac{1}{N} \sum_{i=1}^N \delta(k, k_i) \quad (1.3)$$

où N_k est le nombre de nœuds de degré k et $\delta(i, j)$ est la fonction delta de Kronecker. Cette distribution représente la fraction des nœuds dans le réseau possédant un degré égal à k . Cette propriété peut aussi être interprétée comme la probabilité qu'un nœud sélectionné aléatoirement ait un degré égal à k . De la distribution des degrés, on définit le degré moyen $\langle k \rangle$

$$\langle k \rangle = \sum_{k=k_{\min}}^{k_{\max}} k P(k). \quad (1.4)$$

où $k_{\min} \equiv \min(\mathbf{k})$ et $k_{\max} \equiv \max(\mathbf{k})$.

1.3.2 Corrélation structurelle

La *corrélation structurelle* est une information de plus haut niveau que le degré qui caractérise le *voisinage* d'un nœud. Un voisin du nœud v_i correspond à un nœud séparé d'un lien de v_i . On généralise les n -ième voisins d'un nœud comme étant les nœuds séparés par n liens du nœud source. Sur la Fig. 1.1, on y voit que le nœud ③ est premier voisin du nœud ④.

La corrélation structurelle² est définie en fonction de la probabilité jointe $P(v_i, v_j)$ que le nœud v_i soit connecté au nœud v_j . En général, il est difficile d'exprimer $P(v_i, v_j)$ pour un réseau particulier, car cette probabilité ne décrit pas un seul réseau, mais un ensemble de réseaux (pour plus de détails, voir Sec. 2.1). Comme il n'existe qu'une seule instance du lien a_{ij} pour un réseau donné, il est impossible d'établir une statistique rigoureuse de $P(v_i, v_j)$. Il devient ainsi plus pratique de travailler avec la probabilité jointe $P(k, k')$ qu'un nœud de degré k soit connecté à un nœud de degré k' . Cette probabilité s'écrit,

$$P(k, k') = \frac{E_{kk'}}{M} \quad (1.5)$$


2. En probabilité, la corrélation r entre deux variables aléatoires X et Y est explicitement définie par

$$r \propto \langle XY \rangle - \langle X \rangle \langle Y \rangle .$$

Or, dans le cadre de la science des réseaux, la corrélation prend un sens moins direct. Ces deux variables aléatoires correspondent aux demi-liens connectés aux nœuds v_i et v_j formant le lien complet a_{ij} . On dira alors qu'un lien est non corrélé (ou indépendant) si $\langle a_{ij} \rangle \propto \langle k_i \rangle \langle k_j \rangle$.

où $E_{kk'}$ correspond au nombre de liens existants entre des nœuds de degrés k et k' . Ce nombre s'exprime en termes de la matrice d'adjacence de la façon suivante

$$E_{kk'} = \frac{1}{2} \sum_{i,j=1}^N \delta(k, k_i) a_{ij} \delta(k_j, k'), \quad (1.6)$$

où le facteur 2 vient du fait que, pour les réseaux simples, les liens sont comptés deux fois puisqu'ils sont bidirectionnels. Par exemple, le seul lien existant entre des nœuds de degrés $k = 1$ et $k' = 3$ sur la Fig. 1.1 est le lien , donc pour ce réseau $E_{13} = 1$.

Comme la corrélation cible l'existence des liens entre toutes les paires de nœuds, elle nous fournit une information intéressante sur les premiers voisins des nœuds. On définit ainsi le *degré moyen des premiers voisins* $\bar{k}_{nn}(k)$ [102] comme le nombre moyen de premiers voisins que les premiers voisins d'un nœud de degré k possèdent³. Pour écrire correctement $\bar{k}_{nn}(k)$, on définit également la probabilité conditionnelle $P(k'|k)$ qu'un nœud de degré k soit connecté à nœud de degré k' choisi aléatoirement

$$P(k'|k) = \frac{E_{k'k}}{kN_k} = \frac{\langle k \rangle P(k', k)}{kP(k)}. \quad (1.7)$$

Cette probabilité est liée à ce qu'on appelle la *corrélation structurelle des degrés*. Ayant $P(k'|k)$ en main, on calcule directement le degré moyen des premiers voisins,

$$\bar{k}_{nn}(k) = \sum_{k'=k_{\min}}^{k_{\max}} k' P(k'|k). \quad (1.8)$$

1.3.3 Assortativité

L'*assortativité* (ou l'*homophilie*⁴) est un phénomène observé dans les réseaux [89, 90] où les nœuds possédant des propriétés similaires ont tendance à se connecter davantage entre eux. Le phénomène inverse existe également et se nomme *désassortativité*. Cette propriété peut être définie selon deux points de vue : l'assortativité par caractéristiques dénombrables et l'assortativité par caractéristiques scalaires [86].

D'une part, l'appartenance à une *classe* est un exemple de caractéristique dénombrable où il pourrait y avoir de l'assortativité. Si on considère un ensemble de q classes notées $\{\sigma_i\}$ auxquelles les nœuds peuvent appartenir, on peut calculer un *coefficient de corrélation des classes*, noté Q , par

$$Q = \frac{1}{2M} \sum_{i,j=1}^N \left(a_{ij} - \frac{k_i k_j}{2M} \right) \delta(\sigma_i, \sigma_j) = \frac{1}{2M} \mathbf{1}^\top \left[\mathbf{A} - \frac{1}{2M} \mathbf{k} \mathbf{k}^\top \right] \boldsymbol{\Sigma} \mathbf{1} \quad (1.9)$$

où un élément σ_{ij} de la matrice $\boldsymbol{\Sigma} = (\sigma_{ij}) = \{0, 1\}^{N \times N}$ vaut 1 si les nœuds v_i et v_j appartiennent à la même classe et 0 autrement. Le premier terme de la somme correspond à la fraction des liens

3. L'indice *nn* dans $\bar{k}_{nn}(k)$ vient de la traduction anglaise *nearest neighbors*.

4. L'homophilie est, à la base, un terme tiré de la sociologie qui stipule que les gens préfèrent généralement s'associer à d'autres gens ayant des intérêts communs. Dans le cadre des réseaux, l'homophilie caractérise la tendance des nœuds à se connecter avec d'autres ayant des propriétés similaires.

à l'intérieur d'une classe et le second correspond au nombre moyen de liens que l'on s'attendrait à observer s'il n'y avait pas d'organisation en classes. Dans la littérature, cette quantité s'appelle la *modularité* [92, 93] et elle quantifie l'organisation en communautés d'un réseau. Présentement, l'étude des réseaux complexes organisés en communauté est un sujet fort étudié. Toutefois, elle dépasse largement le cadre du présent mémoire et ne sera pas étudiée davantage. Pour plus de détails, consulter les Réfs. [51, 52, 70, 92, 86].

D'autre part, on peut penser au degré des nœuds comme caractéristique scalaire d'assortativité. Dans ce cas, on définit le *coefficient d'assortativité par degré*, noté par r ,

$$r = \frac{1}{2M} \sum_{i,j=1}^N \left(a_{ij} - \frac{k_i k_j}{2M} \right) k_i k_j = \frac{1}{2M} \mathbf{1}^\top \left[\mathbf{A} - \frac{1}{2M} \mathbf{k} \mathbf{k}^\top \right] \mathbf{k} \mathbf{k}^\top \mathbf{1}. \quad (1.10)$$

Ce coefficient est positif lorsque les nœuds de haut degré ont tendance à se connecter fortement entre eux, alors qu'il est négatif si les nœuds de haut degré ont tendance à se connecter majoritairement avec ceux de bas degré. On qualifie ces deux types de réseaux d'*assortatif par degré* et de *désassortatif par degré* respectivement ou simplement assortatif (désassortatif) pour des raisons de concision. À la Fig. 1.2, on retrouve deux exemples de réseaux assortatif et désassortatif. Les réseaux assortatifs par degré sont caractérisés par un cœur dense de nœuds de haut degré et une périphérie de nœuds de bas degré. À l'inverse, les réseaux désassortatifs sont composés majoritairement d'*étoiles*, c'est-à-dire des sous-réseaux composés d'un seul nœud de haut degré auquel est attaché plusieurs nœuds de bas degré.

Il est intéressant de noter que l'effet de l'assortativité par degré peut également être observé avec le degré moyen des premiers voisins $\bar{k}_{nn}(k)$. Si $\bar{k}_{nn}(k)$ est une fonction croissante de k , on observe un régime assortatif $r > 0$ alors que si $\bar{k}_{nn}(k)$ est décroissante, on observe un régime désassortatif $r < 0$. Cependant, le cas $r = 0$ n'implique pas nécessairement que $\bar{k}_{nn}(k)$ est constant en k et que $a_{ij} = \frac{k_i k_j}{2M} \forall i, j$.

1.3.4 Agrégation

L'*agrégation (clustering)* est une information de plus haut niveau que le degré qui caractérise la corrélation entre trois nœuds [86]. Elle désigne la probabilité qu'un *triplet*, c'est-à-dire trois nœuds connectés en chaîne (par exemple le triplet ①-②-⑤ de la Fig. 1.1), soit fermé et qu'il forme un *triangle* (par exemple le triangle ④-⑦-⑤ de la Fig. 1.1). Formellement, l'agrégation d'un réseau est déterminée par le *coefficient d'agrégation moyen (transitivity)* C donné par

$$C = 3 \times \frac{\text{Nombre de triangles}}{\text{Nombre de triplets}} = \frac{\sum_{i,j,k=1}^N a_{ij} a_{jk} a_{ki}}{\sum_{i,j,k=1}^N a_{ij} a_{jk}} = \frac{\mathbf{1}^\top \mathbf{A} \mathbf{A}^\top \mathbf{A} \mathbf{1}}{\mathbf{1}^\top \mathbf{A}^\top \mathbf{A} \mathbf{1}}. \quad (1.11)$$

Le facteur 3 correspond au nombre de fois qu'un triangle participe à une interaction de triplets. Par exemple, dans le réseau de la Fig. 1.1, il y a 3 triangles et 20 triplets, donc $C = \frac{9}{20}$. Il existe une manière de définir localement le coefficient d'agrégation, c'est-à-dire pour chaque nœud individuellement.

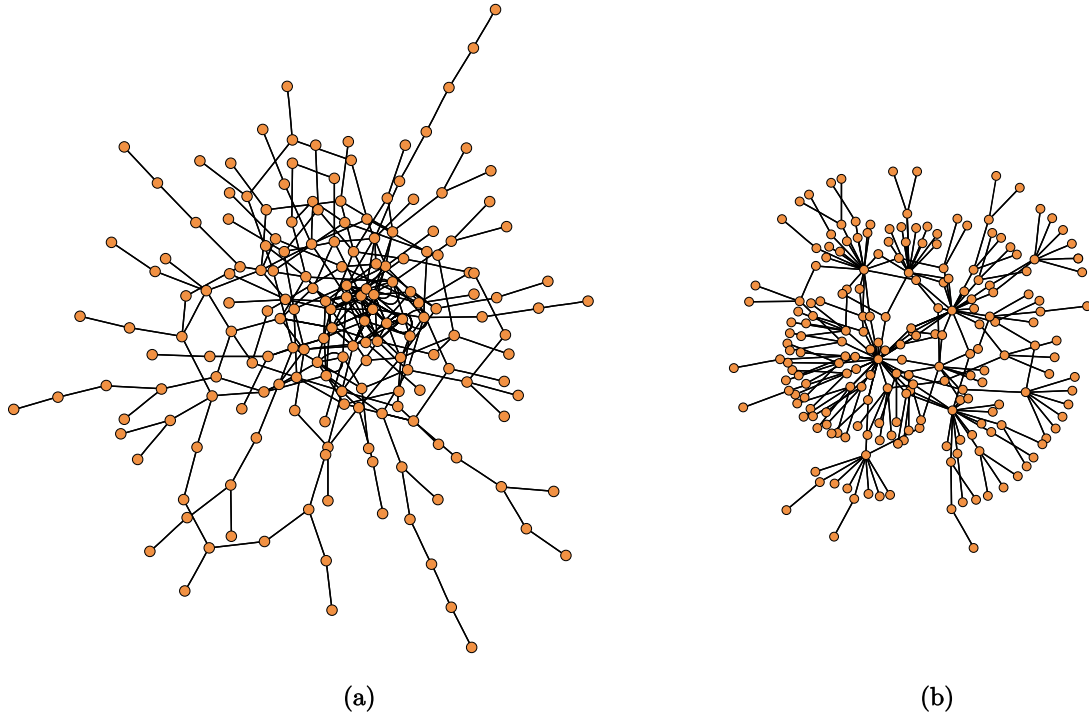


FIGURE 1.2 – Exemple de réseaux (a) assortatifs ($r > 0$) et (b) désassortatifs ($r < 0$). Ces réseaux sont tirés des Refs. [86, 95] et ont été créé à partir de l’algorithme spécifié dans les Refs. [89, 90].

Dans ce cas-ci, au lieu de moyenniser sur le nombre de triangles sur l’ensemble du réseau, on moyenne sur le nombre de triangles auxquels un nœud participe. On définit ainsi le *coefficient d’agrégation local* c_i du nœud v_i comme

$$c_i = \frac{\text{Nombre de paires de voisins de } v_i \text{ connectés}}{\text{Nombre de paires de voisins de } v_i} = 2 \times \frac{\sum_{j,k=1}^N a_{ij} a_{jk} a_{ki}}{k_i(k_i - 1)}. \quad (1.12)$$

De la Fig. 1.1, on observe par exemple que le coefficient d’agrégation du nœud ⑨ est égal à $c_9 = \frac{1}{3}$. Pour quantifier les réseaux, il est souvent plus commode de calculer le *coefficient d’agrégation local moyen dépendant du degré* $\bar{c}(k)$ donné par,

$$\bar{c}(k) = \frac{1}{N_k} \sum_{i=1}^N c_i \delta(k, k_i). \quad (1.13)$$

On peut également définir le *coefficient d’agrégation local moyen* $\langle c \rangle$,

$$\langle c \rangle = \frac{1}{N} \sum_{i=1}^N c_i. \quad (1.14)$$

Les définitions de C et $\langle c \rangle$, bien que comparables, sont fondamentalement différentes et vont donner des résultats différents dans la majorité des cas. Afin d’illustrer ce propos, sur le réseau de la Fig. 1.1, on observe que $C = \frac{9}{20}$ alors que $\langle c \rangle = \frac{17}{60}$.

Type	Réseau	N	M	$\langle k \rangle$	γ	C	$\langle c \rangle$	r	Ref(s).
Social	<i>Telephone call graph</i>	47 000 000	80 000 000	3.16	2.1				[2, 3]
	<i>Biology coauthorship</i>	1 520 251	11 803 064	15.53	-	0.088	0.6	0.127	[87, 88]
	<i>Film actors</i>	449 913	25 516 482	113.43	2.3	0.2	0.78	0.208	[7, 121]
	<i>Math coauthorship</i>	253 339	496 489	3.92	-	0.15	0.34	0.120	[30, 59]
	<i>Physics coauthorship</i>	52 909	245 30	9.27	-	0.45	0.6	0.363	[87, 88]
	<i>Company directors</i>	7 673	55 392	14.44	-	0.59	0.88	0.276	[36, 94]
	<i>Sexual contacts</i>	2 810			3.2				[79, 80]
	<i>Student dating</i>	573	477	1.66	-	0.005	0.001	-0.029	[15]
Information	<i>Word co-occurrence</i>	460 902	16 100 000	66.96	2.7		0.44		[39, 45]
Technologique	<i>Electric circuits</i>	24 097	53 248	4.34	3.0	0.010	0.030	-0.154	[44]
	<i>Internet</i>	10 697	31 992	5.98	2.5	0.035	0.39	-0.189	[31, 43]
	<i>Power grid</i>	4941	6594	2.67	-	0.10	0.08	-0.003	[121]
	<i>Peer-to-peer network</i>	880	1 296	1.47	2.1	0.012	0.011	-0.366	[1, 109]
	<i>Train routes</i>	587	19	66.79	-	0.69		-0.033	[110]
Biologique	<i>Protein interaction</i>	2 115	2 240	2.12	2.4	0.072	0.071	-0.156	[69]
	<i>Metabolic network</i>	765	3 686	9.64	2.2	0.090	0.67	-0.240	[5]

Tableau 1.1 – Propriétés structurelles de plusieurs réseaux réels (simples) : le nombre de nœuds N , le nombre de liens M , le degré moyen $\langle k \rangle$ donné par Éq. (1.4), l'exposant de la distribution des degrés si elle est en loi de puissance $P(k) \propto k^{-\gamma}$ ("- dans le cas contraire), le coefficient d'agrégation moyen C donné par Éq. (1.11), le coefficient d'agrégation local moyen $\langle c \rangle$ donné par Éq. (1.14) et le coefficient d'assortativité par degré r donné par Éq. (1.10). On laisse une case vide lorsque la propriété n'est pas accessible. Ce tableau est tiré de Réf. [86, 91] et a été modifié par soucis d'uniformité.

1.4 Propriétés des réseaux réels

Sur les réseaux réels, l'application de ces mesures permet d'avoir une vue d'ensemble sur leur structure à grande échelle. Le plus intéressant est qu'il existe des tendances dans leurs propriétés [86], indépendantes de leur nature et de leur fonctionnement.

Sur le Tableau 1.1, on présente les propriétés structurelles de 17 réseaux réels tirés de la littérature ayant des provenances différentes. D'une part, on observe que les réseaux d'un même type ont des propriétés structurelles semblables, même si leur nombre de nœuds peut être considérablement différent. Par exemple, dans les réseaux sociaux, on observe généralement que le coefficient d'agrégation moyen (global ou local) est élevé —donnant raison au vieil adage "les amis de mes amis sont mes amis"—, alors qu'il est plus faible pour les réseaux technologiques, pour qui les redondances de liens (les triangles) sont souvent très coûteuses et peu avantageuses. De plus, on observe que le coefficient r d'assortativité par degré des réseaux sociaux est généralement positif, ce qui n'est pas le cas dans les autres types de réseau. Cette homophilie est conforme avec ce qu'on observe dans les sciences sociales [120].

Une propriété qui semble être indépendante de la nature des réseaux est la forme de la distribution des degrés en loi de puissance, c'est-à-dire $P(k) \propto k^{-\gamma}$. Cette particularité a été identifiée [11] et étudiée [32] de façon exhaustive dans les dernières années pour un grand nombre de réseaux réels d'importance. On qualifie ainsi un réseau ayant une telle distribution des degrés comme étant un réseau *invariant d'échelle* (*scale-free*). Contrairement aux autres types de distributions de degrés, la distribution en loi de puissance implique l'existence d'une forte hétérogénéité entre les nœuds : il existe des nœuds de tous les degrés. En d'autres termes, bien qu'une fraction importante de nœuds ait un nombre faible de liens (bas degrés), une fraction non-négligeable possédera de nombreux liens (les hubs de hauts degrés). La présence de cette propriété est déterminante pour plusieurs phénomènes dynamiques importants, tels que la robustesse des réseaux aux attaques ciblées [27] et la chute du seuil d'épidémie dans une population [103, 104]. Il a été démontré dans différents contextes [10, 42, 128] que ce phénomène est lié à la croissance de ces réseaux et à un mécanisme d'attachement appelé l'*attachement préférentiel*. Ce type de modèles sera exploré de façon plus approfondie à la Sec. 2.4.

Il sera important de conserver en tête le Tableau 1.1 et ses évidences empiriques lorsque viendra le temps d'explorer et de développer les différents *modèles* de réseaux. Une correspondance avec la réalité pourra ensuite être considérée.

Chapitre 2

Modèles de réseaux aléatoires

Les réseaux complexes ont généralement un grand nombre de nœuds et de liens. Ceci fait en sorte que, pour une classe de réseaux donnée, même si leur processus de formation sont identiques, il y a de faibles chances qu'ils aient la même structure. En d'autres termes, ils ont une nature *stochastique*. Il devient donc nécessaire des modèles probabilistes de décrivant des ensembles de réseaux afin de pouvoir regrouper et classifier les réseaux réels.

Dans ce chapitre, plusieurs ensembles de réseaux aléatoires, utilisés au coeur du projet de ce mémoire, sont présentés. À la Sec. 2.1, le concept de réseau aléatoire est explicité. Les réseaux aléatoires, formant une classe de modèles mathématiques, permettent de décrire formellement les propriétés structurelles. On présente trois exemples de modèles incontournables : le modèle d'Erdős-Rényi, le modèle des configurations, et le modèle de réseaux aléatoires corrélés avec variables cachées. Puis, à la Sec. 2.2, on aborde le sujet de premier plan des réseaux spatiaux. Les réseaux géométriques aléatoires, formant un modèle simple de réseaux spatiaux, sont discutés à la Sec. 2.3. En particulier, on traite de leur description mathématique, de leurs différentes paramétrisations et des effets que celles-ci entraînent sur les propriétés structurelles. Finalement, à la Sec. 2.4, on discute d'une classe fondamentale des modèles de réseaux aléatoires qui traite explicitement de leur évolution : les modèles de croissance. L'attachement préférentiel, qui constitue une pierre angulaire des modèles de croissance, y est présenté en détails ainsi que plusieurs de ses variantes.

2.1 Les modèles fondamentaux

En général, les réseaux aléatoires sont définis comme des réseaux, dont certaines propriétés structurelles, comme la distribution des degrés $P(k)$ ou le coefficient d'agrégation moyen C , sont fixées alors que le reste de la structure est laissé au hasard. Pour les construire, on doit faire appel à des modèles de réseaux aléatoires. La particularité de ces modèles est que, selon leur formulation, il est possible que la fréquence à laquelle un réseau particulier G apparaît soit différente d'un modèle à l'autre. On définit alors un modèle de réseaux aléatoires par le triplet $(\mathcal{G}, \mathbb{P}, \Theta)$, où \mathcal{G} est l'ensemble des réseaux accessibles par le modèle, $\mathbb{P} : \mathcal{G} \rightarrow [0, 1]$ est la distribution des éléments de l'ensemble et

Θ est l'ensemble des paramètres du modèle. On écrira qu'un réseau G de l'ensemble $\mathcal{G} = \mathcal{G}(\Theta)$ est généré avec probabilité $\mathbb{P}(G) = \mathbb{P}(G|\Theta)$ par le modèle. Cependant, pour des raisons de commodité, la dépendance de \mathcal{G} et \mathbb{P} avec les paramètres Θ sera sous-entendue et omise.

Les modèles les plus fondamentaux de la théorie des réseaux aléatoires ont généralement une distribution $\mathbb{P}(G)$ qui minimise l'information qu'on y injecte [91, 100, 117]. Du point de vue de la théorie de l'information, on dit que la distribution $\mathbb{P}(G)$ de l'ensemble \mathcal{G} maximise l'entropie de Shannon S donnée par

$$S = - \sum_{G \in \mathcal{G}} \mathbb{P}(G) \ln \mathbb{P}(G). \quad (2.1)$$

Avec une approche par multiplicateurs de Lagrange, similaire aux développements de l'ensemble canonique en physique statistique [108], on peut montrer que la distribution qui maximise l'entropie de Shannon, sous certaines contraintes $\langle x_i \rangle = \bar{x}_i$ avec $i = 1, \dots, n$, où le moyennage $\langle \cdot \rangle$ est fait sur $\mathbb{P}(G)$, les x_i sont des propriétés structurelles et les \bar{x}_i sont les valeurs fixées de ces propriétés, a la forme suivante

$$\mathbb{P}(G) = \frac{e^{-H(G)}}{Z} \quad (2.2)$$

où $H(G) = \sum_{i=1}^n \eta_i x_i$ est appelé l'Hamiltonien de G , les η_i sont les multiplicateurs de Lagrange (ou les champs auxiliaires) et $Z = \sum_{G \in \mathcal{G}} e^{-H(G)}$ est la fonction de partition. Cet ensemble particulier est appelé l'ensemble des réseaux aléatoires exponentiels (*exponential random graph model*, ERGM) [100]. À partir de la fonction de partition, on peut ainsi définir une énergie libre $F = -\ln Z$ et fixer la valeur des n multiplicateurs de Lagrange en solutionnant les équations suivantes

$$\frac{\partial F}{\partial \eta_i} = \bar{x}_i. \quad (2.3)$$

Un cas particulier du ERGM est celui où les liens sont décorrélés. Les contraintes fixées correspondent ainsi au nombre moyen de liens que l'on observe entre les nœuds v_i et v_j ,

$$\langle a_{ij} \rangle = p_{ij} \quad (2.4)$$

où $p_{ij} \in [0, 1]$. De là, on calcule explicitement l'Hamiltonien

$$H(G) = \sum_{i < j} \eta_{ij} a_{ij} \quad (2.5)$$

ainsi que la fonction de partition

$$Z = \sum_{\{a_{ij}=0,1\}} \prod_{i < j} e^{-\eta_{ij} a_{ij}} = \prod_{i < j} (1 + e^{-\eta_{ij}}). \quad (2.6)$$

En utilisant l'Éq. (2.3), on relie les champs auxiliaires à p_{ij} de la façon suivante

$$p_{ij} = \frac{1}{1 + e^{\eta_{ij}}} \quad (2.7)$$

qui correspond à une distribution de Fermi-Dirac bien connue en physique statistique¹. On peut finalement calculer la probabilité $\mathbb{P}(G)$ avec l'Éq. (2.2)

$$\mathbb{P}(G) = \prod_{i < j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1 - a_{ij}}. \quad (2.9)$$

Ce modèle très général est associé à plusieurs modèles qui sont des pierres angulaires de la théorie des réseaux aléatoires. Dans cette section, ces modèles fondamentaux sont présentés.

2.1.1 Modèle d'Erdős-Rényi

La forme la plus simple de réseaux aléatoires est donnée par le modèle d'Erdős-Rényi (MER) [41]. Dans ce modèle, un réseau de N noeud est créé en fixant le nombre de liens $M = \bar{M}$. Les liens sont alors distribués uniformément à travers les noeuds. On définit le MER par $(\mathcal{G}_{\text{ER}}, \mathbb{P}_{\text{ER}}, \Theta_{\text{ER}} = \{N, \bar{M}\})$.

Malgré que le MER soit simple, plusieurs propriétés structurelles sont relativement difficiles à calculer. C'est pourquoi il est souvent plus pratique de travailler avec le modèle de Gilbert [56], défini par $(\mathcal{G}_{\text{G}}, \mathbb{P}_{\text{G}}, \Theta_{\text{G}} = \{N, p\})$, où p est la fraction moyenne du nombre de liens existants. Pour faire la correspondance entre le MER et celui de Gilbert, on doit tout simplement poser $p = \frac{M}{M_{\text{max}}}$. Par exemple, la distribution $\mathbb{P}_{\text{G}}(G)$ se calcule beaucoup plus facilement

$$\mathbb{P}_{\text{G}}(G) = p^M (1 - p)^{M_{\text{max}} - M} \quad (2.10)$$

où $M_{\text{max}} = \binom{N}{2} = \frac{N(N-1)}{2}$ est le nombre maximal de liens possibles.

Bien que ces deux modèles soient très semblables, ils demeurent fondamentalement différents puisque, dans le cas du MER, le nombre de liens est fixé exactement, alors que pour le modèle de Gilbert il est fixé en moyenne à $\langle M \rangle = pM_{\text{max}}$. Formellement, ceci implique directement que $\mathcal{G}_{\text{MER}} \subseteq \mathcal{G}_{\text{G}}$. Du point de vue de la physique statistique, le MER correspondrait à un modèle *microcanonique* et le modèle de Gilbert correspondrait à un modèle *canonique*. Ainsi, de la même manière qu'en physique statistique [108], dans la *limite thermodynamique* ($N \rightarrow \infty$), les deux modèles s'équivalent parfaitement. Cette limite particulière est importante en théorie des réseaux aléatoires puisqu'elle facilite un grand nombre de calculs. En fait, dans cette limite, tout ensemble \mathcal{G} ne possède plus qu'un seul élément : un réseau infini auquel tous les éléments de l'ensemble fini tendent. Il devient alors raisonnable de dire qu'une propriété donnée du réseau infini $x = \langle x \rangle$, où $\langle x \rangle$ est la moyenne de x sur l'ensemble fini.

1. La distribution de Fermi-Dirac est associée aux *fermions* en physique statistique. Ce sont des particules qui ne peuvent partager un même état (principe d'exclusion de Pauli). C'est pourquoi, en imposant que le réseau soit simple (sans multilien et boucle), la correspondance entre les liens et les fermions est valide. L'espace entre deux noeuds devient ainsi un état accessible qui peut être (ou non) occupé par un lien (un fermion). Si les multiliens étaient permis, le même développement aurait mené, non pas à la distribution de Fermi-Dirac, mais à la distribution de Bose-Einstein [100],

$$p_{ij} = \frac{1}{1 - e^{\eta_{ij}}} \in [0, \infty). \quad (2.8)$$

Dans ce cas, les liens sont des bosons pour lesquels le principe d'exclusion ne s'applique pas.

Dans le MER, la distribution des degrés peut aussi être calculée par l'intermédiaire du modèle de Gilbert. Un nœud donné de $G \in \mathcal{G}_G$ peut être connecté à chacun de ses $N - 1$ voisins avec une même probabilité p . Donc, la probabilité qu'il ait k voisins est égale au produit $p^k(1 - p)^{(N-1)-k}$, et comme il y a $\binom{N-1}{k}$ manières de choisir k voisins dans le réseau,

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}. \quad (2.11)$$

De cette distribution, le degré moyen se calcule aisément

$$\langle k \rangle = p(N-1). \quad (2.12)$$

Quand on applique la limite thermodynamique, deux régimes se distinguent : le régime *dense* (*dense*) et le régime *creux* (*sparse*). Pour commodité linguistique, on utilisera l'adjectif *sparse* pour le reste du mémoire. Le régime dense est caractérisé par le fait que le degré moyen $\langle k \rangle = \mathcal{O}(N) \rightarrow \infty$ dans la limite thermodynamique. À l'inverse, dans le régime *sparse*, $\langle k \rangle = \mathcal{O}(1)$ pour toutes valeurs de N . Dans la limite thermodynamique, cela implique que $p = \frac{\langle k \rangle}{N-1} = \mathcal{O}(N^{-1}) \rightarrow 0$.

Le régime *sparse* est généralement plus intéressant à étudier puisque la majorité des réseaux réels sont effectivement *sparse*. Évidemment, d'identifier si un réseau est *sparse* ou dense devient difficile pour les cas réels qui sont composés d'un nombre fini de nœuds. On accepte la plupart du temps qu'un réseau soit *sparse* lorsque $\langle k \rangle \ll N$. Dans ce régime, il est possible de développer une autre expression de la distribution des degrés. En effet, à l'aide de l'approximation de Stirling, on peut développer le coefficient binomial tel que $\binom{N-1}{k} \simeq \frac{(N-1)^k}{k!}$ pour $N \rightarrow \infty$. De plus, on calcule $(1-p)^{N-1-k} \simeq e^{-\langle k \rangle}$ avec l'Éq. (2.12) pour obtenir,

$$P(k) \simeq \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!} \quad (2.13)$$

qui est une distribution de Poisson. Comme les liens sont attribués uniformément avec une probabilité p , on peut également se convaincre que $\bar{k}_{nn}(k) = \langle k \rangle$ et que $\langle c \rangle = p = \frac{\langle k \rangle}{N-1}$. Le fait que $\langle c \rangle = p = \mathcal{O}(N^{-1})$ a de fortes implications sur la validité de ce modèle à reproduire la structure des réseaux réels. En effet, comme indiqué au Tableau 1.1, puisque les réseaux complexes ont généralement un coefficient d'agrégation moyen élevé, le MER n'est pas adéquat pour décrire leurs propriétés structurelles.

2.1.2 Modèle des configurations

Le MER est un modèle intéressant du point de vue théorique, mais qui comporte plusieurs lacunes au niveau pratique. Le fait que la distribution des degrés soit binomiale, ce qui est rarement observé dans les réseaux réels, est un problème majeur. En effet, la distribution des degrés étant une propriété fondamentale des réseaux, un modèle se doit de pouvoir la reproduire.

Le modèle des configurations (*Configuration Model*, CM) [94] est un modèle de réseaux aléatoires qui généralise le MER dans ce sens : il génère un réseau aléatoire dont la séquence des degrés $\mathbf{k} = (k_i)$

est fixée. On définit le CM par $(\mathcal{G}_{\text{CM}}, \mathbb{P}_{\text{CM}}, \Theta_{\text{CM}} = \{ \mathbf{k} \})$. Cette formulation du CM considère que les degrés des nœuds sont fixés exactement. On dit alors que les contraintes du CM sur la distribution de degrés sont *rigides* et qu’il constitue un ensemble microcanonique. Il est possible de définir son équivalent canonique simplement en considérant les contraintes *souples* $\langle k_i \rangle = \kappa_i$, où κ_i est le degré espéré² du nœud v_i .

Le CM est un modèle de premier ordre, puisqu’il permet de générer une structure ayant une distribution des degrés arbitraire, sans spécifier les autres propriétés. Ces propriétés sont ainsi entièrement induites par la séquence des degrés. Cette particularité du CM permet de l’utiliser comme *modèle nul*, c’est-à-dire comme un outil de référence de réseaux sans structure. La modularité Q [Éq. (1.9)] et le coefficient d’assortativité par degré [Éq. (1.10)] sont en effet deux exemples de mesure qui comparent une structure donnée à celle d’un réseau fait par le CM. Le terme $\frac{k_i k_j}{2M}$, qui approxime p_{ij} dans le ERGM, est sa signature.

De plus, il est possible de décrire analytiquement les propriétés structurelles du CM à l’aide d’un formalisme de fonctions génératrices³ de la distribution des degrés $P(k)$ [86, 94].

On calcule par cette approche le degré moyen des premiers voisins

$$\bar{k}_{nn}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle} \quad (2.14)$$

et le coefficient d’agrégation

$$\bar{c}(k) = \frac{[\langle k^2 \rangle - \langle k \rangle]^2}{N \langle k \rangle^3} = \mathcal{O}(N^{-1}) \quad (2.15)$$

où $\langle k^2 \rangle = \sum_{k=0}^{\infty} k^2 P(k)$ est le second moment de $P(k)$. Il est toutefois important de souligner que, selon l’hétérogénéité de la distribution des degrés considérée, il est possible d’observer une différence entre les Éq. (2.14, 2.15) et leur véritable expression. En effet, pour les distributions de degré en loi de puissance $P(k) \sim k^{-\gamma}$ où $\gamma \simeq 2$, on observe que $\bar{k}_{nn}(k)$ et $\bar{c}(k)$ peuvent être des fonctions décroissantes de k . Cette désassortativité présente même dans le CM est entièrement due aux contraintes structurelles imposées par la séquence des degrés [23, 102, 111]. Pour éviter ce genre de corrélations, on doit permettre les multiliens et les boucles ou encore imposer une coupure structurelle dans la séquence des degrés $k_s \sim (\langle k \rangle N)^{1/2}$ telle que $k_i < k_s$ pour tout i [23].

2.1.3 Modèles de réseaux aléatoires corrélés avec variables cachées

Les réseaux complexes sont caractérisés par une variété de propriétés différentes (distribution des degrés en loi de puissance, assortativité, coefficient d’agrégation élevé, etc.) qui ne sont toutefois pas

2. Une distinction subtile entre *moyen* et *espéré* est utilisée dans ce mémoire par rapport à l’ensemble associé sur lequel on prend la moyenne. Pour “moyen”, on utilise l’ensemble des nœuds du réseau alors que pour “espéré”, on utilise l’ensemble de réseaux du modèle. Par contre, pour certaines quantités comme $\bar{k}_{nn}(k)$, les deux définitions peuvent se superposer.

3. La fonction génératrice d’une distribution $P(\lambda)$ de variables aléatoires discrètes λ quelconque s’écrit $g(x) = \sum_{\lambda=0}^{\infty} P(\lambda)x^\lambda$. Même si la forme de $g(x)$ s’apparente aux séries de puissance standards, il n’est toutefois pas nécessaire que $g(x)$ soit une fonction analytique sur tout son domaine de x . Ceci en fait une série formelle.

présentes chez le MER et le CM. Notamment, la corrélation des degrés présente dans les systèmes réels illustre clairement que ces modèles sont inadéquats pour décrire correctement leur structure.

Afin de combler ce vide, M. Boguñá et R. Pastor-Satorras ont développé un modèle très général de réseaux ayant des corrélations arbitraires entre certaines classes de nœuds : un modèle de *réseaux aléatoires corrélés avec variables cachées* [21] (modèle de Boguñá-Pastor-Satorras, MBPS). Une variable cachée h est une variable aléatoire discrète (ou continue⁴) distribuée selon une fonction de poids (ou densité de probabilité) $\rho(h)$ et attribuée à chacun des nœuds du réseau comme un identifiant. Les liens du réseau sont ensuite distribués selon une *probabilité de connexion* $p(h_i, h_j)$, une fonction symétrique qui détermine la probabilité qu'un lien existe entre deux nœuds v_i et v_j de variables cachées respectives h_i et h_j . On comprend que le modèle est entièrement spécifié par les paramètres $\Theta_{\text{BPS}} = \{ \rho(h), p(h, h') \}$. Ainsi, on définit ce modèle par le triplet $(\mathcal{G}_{\text{BPS}}, \mathbb{P}_{\text{BPS}}, \Theta_{\text{BPS}})$. Ce modèle est encore une fois lié au ERGM, où on fixe les contraintes suivantes

$$\langle a_{ij} \rangle = p_{ij} \equiv p(h_i, h_j). \quad (2.16)$$

On calcule donc $\mathbb{P}_{\text{BPS}}(G)$ tel que

$$\mathbb{P}_{\text{BPS}}(G) = \sum_{\{h_i\}} \prod_{i=1}^N \rho(h_i) \left[\prod_{j=i+1}^N p_{ij}^{a_{ij}} (1 - p_{ij})^{1-a_{ij}} \right]. \quad (2.17)$$

où l'on somme sur chaque valeur possible de $h_i \forall i = 1, \dots, N$. En passant par le ERGM, on constate que le MER et le CM sont des cas particuliers du MBPS. En effet, si on considère que $\rho(h)$ est uniforme et que $p(h, h') = p$ est constant, on retrouve le MER. De la même manière, si on considère que $p(h, h') = \frac{hh'}{N\langle h \rangle}$, et qu'on impose $\kappa(h) \simeq h$, alors h correspond à un *degré caché* et, conséquemment, $\rho(k) \simeq P(k)$. Le MBPS émule le CM avec des contraintes souples. Ces deux exemples témoignent de la polyvalence du MBPS.

Il est possible de considérer que la variable cachée soit un vecteur de variables aléatoires, noté $\mathbf{h} = (h_1, \dots, h_{\mathcal{D}})^\top$ de distribution $\rho(h_1, \dots, h_{\mathcal{D}}) = \rho(\mathbf{h})$, dans le cas où plusieurs variables cachées sont impliquées. En général, on considérera que les variables cachées h_i sont statistiquement indépendantes et que $\rho(\mathbf{h}) = \prod_i \rho(h_i)$. Dans ce qui suit, une analyse exhaustive des propriétés structurelles⁵ est présentée pour le cas d'une variable cachée unique. Cependant, tous les calculs sont généralisables pour \mathbf{h} .

Distribution des degrés—On calcule entre autres la distribution des degrés comme suit

$$P(k) = \sum_{\mathbf{h}} g(k|\mathbf{h}) \rho(\mathbf{h}) \quad (2.18)$$

4. Durant les démonstrations qui vont suivre, on utilisera les sommes sur h . Par contre, ce formalisme est uniquement valide pour les variables cachées discrètes. Pour les variables cachées continues, on substitue les sommes par des intégrales.

5. Le développement qui suit est fortement inspiré de la Section III de la Réf. [21]. Toutefois, pour des besoins de clarté et d'uniformité, certaines notions ont été modifiées.

où $g(k|h)$, appelé le *propagateur* en hommage à la mécanique quantique⁶, est la probabilité conditionnelle qu'un nœud de variable cachée h ait un degré k . À partir du propagateur, il est également possible de calculer le degré moyen $\kappa(h)$ des nœuds de variable cachée h

$$\kappa(h) = \sum_k k g(k|h) \quad (2.19)$$

ainsi que le degré moyen

$$\langle k \rangle = \sum_k k P(k) = \sum_h \rho(h) \kappa(h). \quad (2.20)$$

À ce point, le défi réside dans l'obtention d'une forme fermée pour le propagateur. On constate dans un premier temps qu'il peut s'écrire de la façon suivante

$$g(k|h) = \sum_{k_1, \dots, k_n} g_1^{(h)}(k_1|h_1) g_2^{(h)}(k_2|h_2) \cdots g_c^{(h)}(k_n|h_n) \times \delta \left(k, \sum_{i=1}^n k_i \right) \quad (2.21)$$

où n est le nombre de variables cachées possibles et $g_i^{(h)}(k_i|h_i)$ est la probabilité qu'un nœud de variable cachée h se connecte à k_i nœuds de variable cachée h_i . Puisque les liens entre nœuds h et h' sont indépendamment distribués avec la probabilité $p(h, h')$, ce processus est décrit par un processus binomial et la probabilité $g_i^{(h)}(k_i|h_i)$ suit simplement

$$g_i^{(h)}(k_i|h_i) = \binom{N_i}{k_i} p(h, h_i)^{k_i} [1 - p(h, h_i)]^{N_i - k_i} \quad (2.22)$$

où $N_i = N \rho(h_i)$ est le nombre de nœuds h_i . Ensuite, on définit la fonction génératrice du propagateur

$$\hat{g}(z|h) = \sum_k g(k|h) z^k \quad (2.23)$$

et puisque $g(k|h)$ est une convolution des $g_i^{(h)}(k_i|h_i)$, sa fonction génératrice peut s'exprimer comme le produit des fonctions génératrices partielles,

$$\hat{g}_i(z|h) = \left[1 - (1 - z)p(h, h_i) \right]^{N_i}. \quad (2.24)$$

En insérant cette expression dans $\hat{g}(z|h)$ et en prenant le logarithme des deux côtés de l'égalité, on obtient

$$\ln \hat{g}(z|h) = \sum_{i=1}^c N_i \ln \left[1 - (1 - z)p(h, h_i) \right] = N \sum_{h'} \rho(h') \ln \left[1 - (1 - z)p(h, h') \right]. \quad (2.25)$$

6. En mécanique quantique [8], le propagateur $G(x, t; x', t')$ est une fonction qui permet de faire évoluer dans le temps la fonction d'onde $\psi(x, t)$ d'une particule d'un temps t_0 à un temps t tel que

$$\psi(x, t) = \int G(x, t; x', t_0) \psi(x', t_0) dx'$$

où l'intégrale se fait sur le domaine complet de x' .

Ainsi, pour obtenir le propagateur, il suffit de solutionner et inverser cette dernière expression en fonction de $\rho(h)$ et $p(h, h')$. De façon générale, ayant $\hat{g}(z|h)$, il est possible de calculer $\kappa(h)$ simplement en évaluant la dérivée de $\hat{g}(z|h)$ à $z = 1$ étant donné qu'il s'agit de son premier moment. On obtient

$$\kappa(h) = N \sum_{h'} \rho(h') p(h, h') \quad (2.26)$$

et

$$\langle k \rangle = N \sum_{h, h'} \rho(h) p(h, h') \rho(h'). \quad (2.27)$$

Même si une forme fermée générale pour le propagateur n'est pas accessible, il est tout même possible d'en obtenir une pour le cas des réseaux *sparses*, c'est-à-dire avec $\langle k \rangle = \mathcal{O}(1)$ dans la limite thermodynamique. En effet, dans ces conditions, il faut nécessairement que $\sum_{h, h'} \rho(h) p(h, h') \rho(h') = \mathcal{O}(N^{-1})$. Une façon de respecter cette condition survient lorsque $p(h, h') = \mathcal{O}(N^{-1})$. Ainsi, si on définit $p(h, h') = \frac{\tilde{p}(h, h')}{N}$, où $\tilde{p}(h, h')$ est une fonction symétrique indépendante de N , et que l'on prend la limite $N \rightarrow \infty$, on peut utiliser l'approximation $\ln(1 + x) \simeq x$ pour $x \rightarrow 0$ afin d'obtenir

$$\hat{g}(z|h) = \exp \left[(z - 1) \sum_{h'} \rho(h') \tilde{p}(h, h') \right]. \quad (2.28)$$

Cette forme exponentielle est la fonction génératrice d'une distribution de Poisson

$$g(k|h) = \frac{e^{-\kappa(h)} \kappa(h)^k}{k!} \quad (2.29)$$

où $\kappa(h) = \sum_{h'} \rho(h') \tilde{p}(h, h')$. La puissance de ce résultat est remarquable puisqu'il indique l'universalité de la forme du propagateur pour tous réseaux *sparses* indépendamment de $p(h, h')$ ou de $\rho(h)$.

Corrélation des degrés—Pour calculer la corrélation des degrés, on doit obtenir la probabilité conditionnelle $P(k'|k)$. Or, pour la construire, on considère un nœud de degré k qui aurait une variable cachée h avec probabilité $g^*(k|z)$ qui s'obtient avec le théorème de Bayes

$$P(k)g^*(h|k) = \rho(h)g(k|h). \quad (2.30)$$

Ensuite, on définit la probabilité conditionnelle $p(h'|h)$ qu'un nœud de variable cachée h sélectionné soit connecté à un autre nœud h' qui s'écrit

$$p(h'|h) = \frac{\rho(h')p(h, h')}{\sum_{h''} \rho(h'')p(h, h'')} = \frac{N\rho(h')p(h, h')}{\kappa(h)}. \quad (2.31)$$

Ainsi, on écrit

$$P(k'|k) = \sum_{h, h'} g(k' - 1|h)p(h'|h)g^*(h|k) \quad (2.32)$$

où $g(k' - 1|h)$ dénote la probabilité qu'un nœud de variable cachée h ait au final un degré k' . De cette expression, on obtient le degré moyen des premiers voisins $\kappa_{nn}(h)$ d'un nœud de variable cachée h

$$\kappa_{nn}(h) = \sum_{h'} \kappa(h')p(h'|h) = N \sum_{h'} \rho(h')p(h, h') \frac{\kappa(h')}{\kappa(h)} \quad (2.33)$$

et

$$\bar{\kappa}_{nn}(k) = 1 + \frac{1}{P(k)} \sum_h g(k|h)\rho(h)\kappa_{nn}(h). \quad (2.34)$$

Coefficient d'agrégation—Finalement, le coefficient d'agrégation peut également s'obtenir à partir du MBPS. On considère qu'un nœud de variable cachée h peut se connecter à deux autres nœuds h' et h'' avec probabilité $p(h', h''|h)$ et que ces deux nœuds peuvent également être connectés avec probabilité $p(h', h'')$. Ainsi, le coefficient d'agrégation $c(h)$ du nœud de variable cachée h s'obtient avec

$$c(h) = \sum_{h', h''} p(h', h''|h)p(h', h'') = c(h) = \sum_{h', h''} p(h'|h)p(h', h'')p(h''|h) \quad (2.35)$$

où on utilise le fait que $p(h', h''|h) = p(h'|h)p(h''|h)$, puisque les liens sont indépendants entre eux. On obtient alors le coefficient d'agrégation dépendant du degré

$$\bar{c}(k) = \frac{1}{P(k)} \sum_h \rho(h)g(k|h)c(h) \quad \text{pour } k = 2, 3, \dots \quad (2.36)$$

2.1.4 Échantillonner les réseaux aléatoires : méthode numérique

En théorie des réseaux aléatoires, on s'intéressera à déterminer numériquement les propriétés structurelles de l'ensemble complet. Cette approche numérique permettra entre autres de valider nos approches analytiques. Or, ces ensembles sont des espaces discrets généralement de hautes dimensions et il devient impossible de les reconstruire entièrement pour mesurer $\mathbb{P}(G)$ et, par conséquent, la distribution de leurs propriétés structurelles. On utilisera plutôt des méthodes de type *Monte-Carlo* pour les construire.

Une méthode Monte-Carlo est une technique stochastique qui permet d'effectuer des calculs sur des espaces de hautes dimensions à faible coût computationnel. Supposons un modèle de réseaux $(\mathcal{G}, \mathbb{P}, \Theta)$ dont on souhaite calculer la valeur moyenne d'une propriété $x = x(G)$. La valeur moyenne $\langle x \rangle$ sur l'ensemble complet \mathcal{G} distribué selon $\mathbb{P}(G)$ est ainsi donnée par,

$$\langle x \rangle = \sum_{G \in \mathcal{G}} \mathbb{P}(G)x(G). \quad (2.37)$$

La stratégie Monte-Carlo [107] stipule que $\langle x \rangle$ peut être estimée par $\langle x \rangle_{MC}$ correspondant à la valeur moyenne de x sur un échantillon de \mathcal{G} de Λ éléments

$$\langle x \rangle \simeq \langle x \rangle_{MC} = \frac{1}{\Lambda} \sum_{i=1}^{\Lambda} x_i \quad (2.38)$$

où $x_i = x(G_i)$ et les G_i sont Λ variables aléatoires (réseaux) tirées selon la distribution $\mathbb{P}(G)$. Dans la limite où $\Lambda \rightarrow \infty$, on peut évaluer l'incertitude sur I en calculant la variance de l'estimateur

$$\sigma^2 \simeq \frac{1}{\Lambda} \sigma_x^2 = \frac{1}{\Lambda} \left(\langle x^2 \rangle_{MC} - \langle x \rangle_{MC}^2 \right) \quad (2.39)$$

En effet, comme l'estimateur est une somme de $\Lambda \rightarrow \infty$ variables aléatoires x_i dont on suppose les moments bornés supérieurement, on sait que, du théorème limite central, la distribution de l'estimateur sera en effet gaussienne. Ainsi, on constate que, puisque $\sigma = \mathcal{O}(\Lambda^{-1/2})$, la méthode numérique convergera sur la bonne valeur avec la taille de l'échantillon. Cependant, dans la mesure où l'un ou l'autre des deux premiers moments n'est pas borné dans la limite thermodynamique, la méthode Monte-Carlo aura de la difficulté à converger pour certaines propriétés structurelles comme le coefficient d'agrégation [75, 84].

2.2 Réseaux spatiaux

Contrairement aux réseaux aléatoires présentés à la Sec. 2.1, les réseaux réels sont explicitement soumis à des contraintes extérieures de différentes natures qui ont une influence notable sur leurs propriétés structurelles. Notamment, les groupes d'amis dans les réseaux sociaux se forment puisqu'ils ont des similarités : ils vivent dans le même quartier, ils travaillent au même endroit, ils étudient dans le même domaine, etc..

Un type important de contraintes extérieures est lié à la *localisation géographique* des nœuds, c'est-à-dire leur position sur la Terre. Ces contraintes sont dites *spatiales*. Elles apparaissent sous généralement sous la forme d'un coût associé à la longueur des liens, comme pour la formation de synapses dans le cerveau [26], ou d'une proximité nécessaire pour établir une connexion, comme pour les réseaux *ad-hoc* de wifi [77]. Cependant elles sont généralement toutes liées à une minimisation de la distance géographique entre les nœuds. Avec du recul, on peut se convaincre facilement que des contraintes spatiales sont présentes dans un grand nombre de réseaux complexes : les infrastructures urbaines (grilles électriques [4], réseaux routiers [28]), les réseaux de transport (réseaux de lignes aériennes [60, 78]) et les réseaux neuronaux [26] sont tous des exemples typiques de réseaux réels soumis à des contraintes spatiales. Ces réseaux font partie de l'ensemble des *réseaux spatiaux*.

Même si le cadre théorique des réseaux spatiaux fut d'abord développé pour étudier l'effet de l'espace géographique sur la structure, on peut également l'utiliser pour décrire des réseaux sans coordonnées spatiales *a priori*. On parle alors d'*espace de plongement caché*⁷ (*embedding*) que les nœuds occupent et qui affecte la probabilité que deux nœuds soient connectés selon la distance les séparant. Depuis peu, cette représentation permet une description efficace d'une variété de réseaux réels comportant

7. Un plongement réfère à une application injective $f : X \rightarrow Y$ qui envoie un espace topologique X sur un autre Y : par exemple, un réseau sur l'espace géographique. Dans le cadre des réseaux spatiaux, on associe un plongement à l'application qui envoie chacun des nœuds à une position dans l'espace géométrique. Pour en savoir plus, on réfère à la Réf. [85].

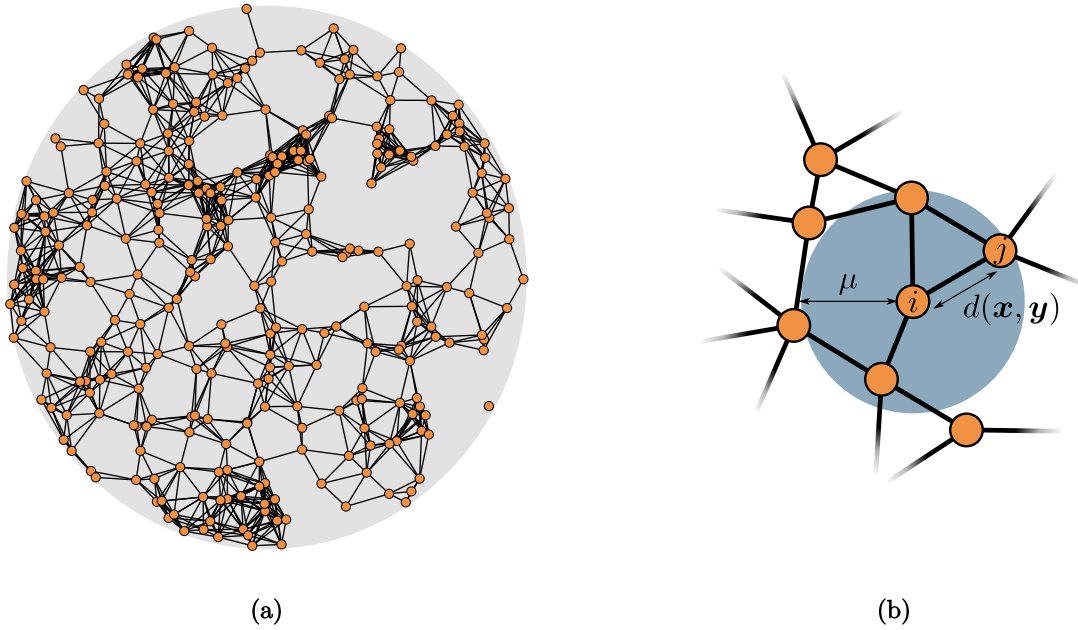


FIGURE 2.1 – Exemple de réseau géométrique aléatoire : (a) réseau géométrique dont la distribution des nœuds est uniforme et \mathcal{M} est un disque de rayon $R = 1$ identifié par la zone grise, (b) schéma du processus de connexion où la zone en bleu indique l'étendue du voisinage d'un nœud v_i . On a fixé $N = 300$ et $\mu = 0.18$.

de l'autosimilarité [112] et un coefficient d'agrégation $\langle c \rangle$ élevé [74]. Ce domaine en effervescence porte aujourd'hui le nom de *géométrie des réseaux* (*network geometry*) et a pour but de fournir une représentation géométriques des réseaux complexes.

2.3 Réseaux géométriques

Le *réseau géométrique* est le cheval de bataille de la géométrie des réseaux. À la base, on le décrit comme un réseau dont chaque nœud possède une position \mathbf{x} dans un *espace métrique* \mathcal{M} de volume $V_{\mathcal{M}}$ et muni d'une distance métrique $d(\mathbf{x}, \mathbf{y})$ entre deux noeuds. Un lien entre deux nœuds placés en \mathbf{x} et \mathbf{y} existe si $d(\mathbf{x}, \mathbf{y}) < \mu$, où $\mu \geq 0$ est appelé le *seuil de connexion* [35, 57, 105] (voir Fig. 2.1). Formellement, on écrit un modèle de réseaux géométriques tel que l'ensemble des paramètres est composé de la position de chaque noeud et la fonction de distance. Bien que fort simple, la dynamique de connexion de ce modèle ressemble beaucoup à celle des réseaux *ad hoc* sans fil [61] où les nœuds sont des appareils qui peuvent communiquer directement entre eux en deçà d'une certaine distance. Ce type de réseaux possède plusieurs applications variées [33] et a souvent été étudié dans le cadre de la théorie de la percolation en milieu continu. Cependant, ceci sort du contexte de ce mémoire et ne sera pas développé davantage. Pour en savoir plus, voir les Réfs. [57, 83].




Propriétés	Euclidien	Sphérique	Hyperbolique
Courbure K	0	> 0	< 0
Nombre de lignes parallèles	1	0	∞
Triangles géodésiques			
Somme des angles d'un triangle	π	$> \pi$	$< \pi$
Périmètre d'un cercle de rayon r	$2\pi r$	$2\pi \sin(\zeta r)$	$2\pi \sinh(\zeta r)$
Aire d'un disque de rayon r	πr^2	$2\pi[1 - \cos(\zeta r)]$	$2\pi[1 - \cosh(\zeta r)]$

Tableau 2.1 – Propriétés caractéristiques des géométries euclidienne, sphérique et hyperbolique. Ce tableau est tiré de la Réf. [75].

2.3.1 Espace métrique

Un espace métrique défini par (\mathcal{M}, d) contient par un ensemble $\mathcal{M} = \{x\}$ et une métrique $d : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}^+$, laquelle est soumise aux quatre conditions suivantes :

1. $d(x, y) \geq 0$ Non-négativité de la métrique,
2. $d(x, y) = 0$ si et seulement si $x = y$ Principe d'identité des indiscernables,
3. $d(x, y) = d(y, x)$ Symétrie de la métrique,
4. $d(x, z) \leq d(x, y) + d(y, z)$ Inégalité du triangle,

où $x, y, z \in \mathcal{M}$. On note que ces quatre conditions ne sont pas nécessaires à la définition de la distance métrique. La condition 1 de non-négativité se déduit des autres et est ainsi redondante. Considérons d'abord l'inégalité du triangle sur le triplet x, y et z tel que $z = x$. On trouve des conditions 3 et 4,

$$d(x, x) \leq d(x, y) + d(y, x) = 2d(x, y). \quad (2.40)$$

Ainsi, de la condition 2, on retrouve la condition 1.

Pour les réseaux géométriques, on écrit ce plongement f formellement de la manière suivante

$$f : \mathcal{V} \mapsto \mathcal{M} \quad (2.41)$$

De cette application, on comprend que les propriétés de l'espace métrique vont transparaître sur la structure du réseau si l'existence des liens est liée à leur longueur sur \mathcal{M} .

Dans les réseaux géométriques, comme les nœuds possèdent des coordonnées sur \mathcal{M} , cela sous-entend que \mathcal{M} est plus spécifiquement une *variété* riemannienne. Une variété est un espace métrique qui ressemble localement à un espace euclidien plat. Cela revient à dire que chacun des points de

\mathcal{M} a un voisinage homéomorphe à l'espace euclidien. En général, on s'intéressera davantage aux variétés isotropes qui sont, dans la hiérarchie des espaces métriques, des variétés riemanniennes dont la courbure gaussienne K est constante. L'intérêt des variétés isotropes est qu'elles ont une géométrie invariante de l'endroit où on la mesure, ce qui facilite grandement le traitement analytique de ces espaces. Il en existe 3 classes particulières : les géométries euclidienne, sphérique et hyperbolique (voir Tableau 2.1). Ces trois types de géométrie sont liés par la constante de courbure $K = \zeta^2$. Pour l'espace euclidien, $K = 0$, alors que pour les espaces sphérique et hyperbolique, la constante est respectivement positive et négative.

Bien que la géométrie euclidienne soit assez connue et très intuitive, les deux autres le sont beaucoup moins. La géométrie sphérique par exemple se replie sur elle-même et peut être plongée directement dans l'espace euclidien. On l'étudie souvent comme la \mathcal{D} -sphère⁸, notée $\mathbb{S}^{\mathcal{D}}$ (voir la Fig. 2.2).

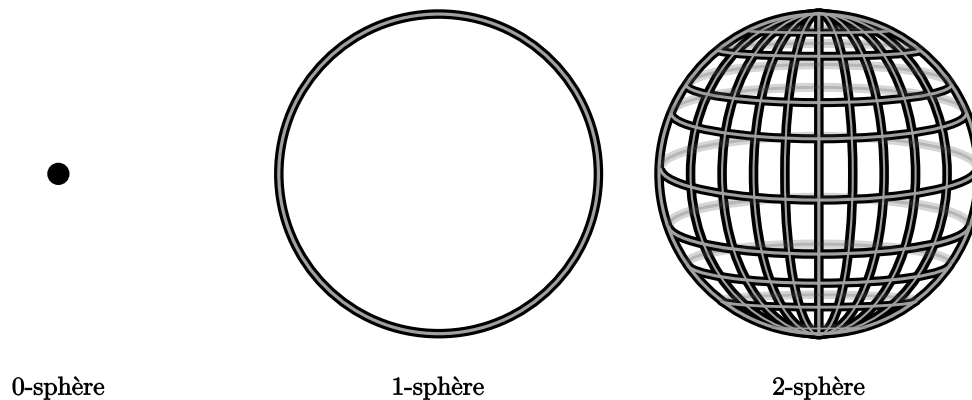


FIGURE 2.2 – Exemples de \mathcal{D} -sphère. La 0-sphère correspond à un point, la 1-sphère, à un cercle dans un espace euclidien en 2D et la 2-sphère, à une sphère dans un espace euclidien en 3D.

La géométrie hyperbolique est encore plus contre-intuitive que la sphérique pour plusieurs raisons. D'une part, sa courbure K étant négative, chaque point de l'espace a la forme d'une selle. D'autre part, il est impossible de plonger un espace hyperbolique dans un espace euclidien, puisqu'il croît exponentiellement en contraste avec l'espace euclidien qui croît de façon polynomiale (voir l'aire des disques sur le Tableau 2.1). Les feuilles d'un chou kale frisé sont des exemples typiques d'objets conventionnels ayant une géométrie hyperbolique. En effet, ces feuilles ne peuvent être étendues sur une table (un espace euclidien bidimensionnel) puisqu'elles occupent "trop" d'espace par rapport à la table, ce qui témoigne de leur nature exponentielle. Une représentation pratique de la géométrie hyperbolique est le *disque de Poincaré* (voir la Fig. 2.3), noté \mathbb{H}^2 , qui correspond à un disque bidimensionnel dont l'intérieur est étiré pour y accommoder *tout* l'espace hyperbolique. Il existe d'autres représentations équivalentes de ce type de géométrie, mais le disque de Poincaré est généralement le plus simple à visualiser. Le disque de Poincaré, comme la \mathcal{D} -sphère, peuvent également être généralisés en \mathcal{D} dimensions. On parle alors de la \mathcal{D} -boule de Poincaré noté $\mathbb{H}^{\mathcal{D}}$.

8. La \mathcal{D} -sphère correspond à la surface d'une $(\mathcal{D} + 1)$ -boule. La distinction entre sphère et boule vient du fait que la sphère est vide, elle ne contient pas les points à l'intérieur de la coquille. Ainsi, la $(\mathcal{D} + 1)$ -boule contient tous les points de la \mathcal{D} -sphère.

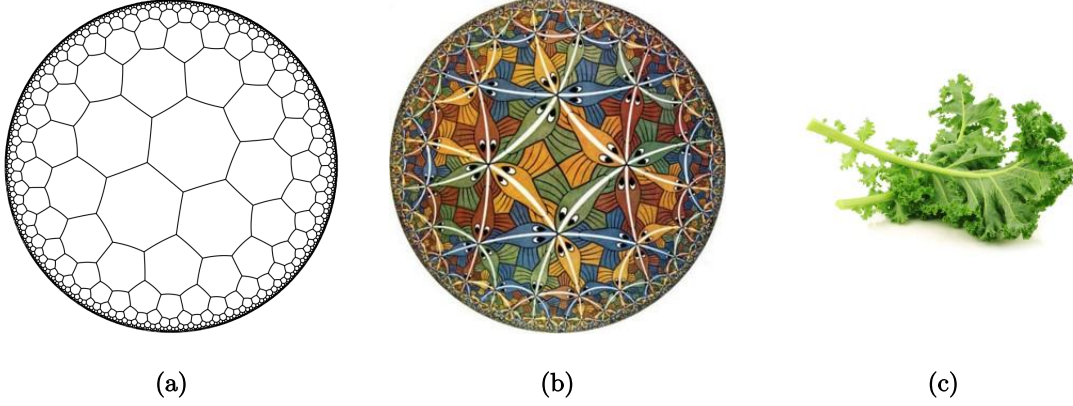


FIGURE 2.3 – Schématisation de la géométrie hyperbolique : (a) disque de Poincaré, (b) œuvre de Escher appelée *Circle limite III* illustrant un espace hyperbolique (les poissons sont tous de la même taille dans cet espace) et (c) exemple de géométrie hyperbolique dans la nature, le chou kale frisé.

2.3.2 Espace caché

On peut définir un modèle de réseaux aléatoires géométriques (RGG⁹) si la position des nœuds et les liens entre eux sont donnés aléatoirement. Formellement, la position \mathbf{x} devient une variable aléatoire dont la distribution est $\rho(\mathbf{x})$ et on détermine l'existence des liens entre deux nœuds en position \mathbf{x} et \mathbf{y} avec une probabilité de connexion $p(\mathbf{x}, \mathbf{y})$. Le modèle des RGGs est ainsi défini par le triplet $(\mathcal{G}_{\text{RGG}}, \mathbb{P}_{\text{RGG}}, \Theta_{\text{RGG}})$, où $\Theta_{\text{RGG}} = \{\rho, (\mathcal{M}, d), p\}$. Dans cette perspective, le MBPS est tout à fait adapté pour représenter cet ensemble de réseaux géométriques aléatoires où les variables cachées sont simplement la position des nœuds dans un *espace caché*.

Afin d'illustrer le propos, considérons le cas le plus simple de RGGs [35]. Soit un RGG dont \mathcal{M} est 1-sphère de rayon R (voir Fig. 2.2), la distribution des nœuds $\rho(\mathbf{x}) = (2\pi R)^{-1}$ est uniforme et deux nœuds se connectent s'ils sont à une distance inférieure ou égale à $\mu > 0$. Selon le MBPS, la position des nœuds x est une variable cachée dont le domaine est $\mathcal{M} = [-\pi R, \pi R]$ et $d(x, y) = \pi R - |\pi R - |x - y||$, où $|\cdot|$ est la fonction valeur absolue. Dans ce contexte, l'existence d'un lien, sachant la position des nœuds, est déterministe. Cependant, la probabilité de connexion peut tout de même s'écrire

$$p(x, y) = \Theta[\mu - d(x, y)] = \begin{cases} 1 & \text{si } d(x, y) < \mu \\ 0 & \text{autrement} \end{cases} \quad (2.42)$$

où $\Theta(x)$ est la fonction Heaviside. On peut alors calculer le degré d'un nœud x

$$\kappa(x) = N \int_{-\pi R}^{\pi R} \rho(y) p(x, y) dy = \frac{N}{2\pi R} \int_{-\pi R}^{\pi R} \Theta \left[\mu - \pi R + |\pi R - |x - y|| \right] dy. \quad (2.43)$$

On constate que, de calculer $\kappa(x)$ revient à déterminer la fraction du volume qu'une boule unidimensionnelle de rayon μ centrée en x occupe dans \mathcal{M} multipliée par le nombre de nœuds. Le volume de

9. L'acronyme RGG est tiré de l'expression anglais *random geometric graphs*.

cette boule est appelé le *volume exclu* [35] et cette nomenclature est encore utilisée dans le contexte des RGGs. En effectuant le changement de variable $x' = |x - y|$, on obtient

$$\kappa(x) = N \frac{\mu}{\pi R}. \quad (2.44)$$

Ainsi, comme $\kappa(x)$ ne dépend pas de x , on a automatiquement $\langle k \rangle = \kappa(x)$ et $\kappa_{nn}(x) = \kappa(x)$. De plus, on obtient la distribution des degrés

$$P(k) = \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}. \quad (2.45)$$

Malgré que ce modèle géométrique et le MER génèrent tous deux des réseaux avec distribution de Poisson, il existe une disparité entre leur coefficient d'agrégation. En effet, pour ces RGGs, on obtient

$$\langle c \rangle = \frac{3}{4} \quad (2.46)$$

qui, on le souligne, est indépendant de N et de μ . Le fait que $\langle c \rangle$ soit constant pour les RGGs est une conséquence directe de l'inégalité du triangle héritée de \mathcal{M} . En effet, l'inégalité du triangle du point de vue d'un réseau favorise la formation de triangles en rapport avec le nombre de triplets : si deux noeuds sont tous les deux près d'un troisième, il est fort probable qu'ils soient également près l'un de l'autre. Le caractère métrique de l'espace devient donc intéressant pour modéliser l'importante formation de triplets connectés de noeuds observée dans les réseaux réels et c'est dans cette optique que la notion d'espace caché prend toute sa force par rapport aux modèles aléatoires classiques.

L'exemple qui vient d'être explicité, où l'espace de plongement est unidimensionnel, est simple. Or, il est possible de le généraliser pour \mathcal{D} dimensions. Dans ce cas, le réseau géométrique est plongé sur $\mathbb{S}^{\mathcal{D}}$. Le calcul du degré moyen se fait de façon similaire avec le volume exclu donné par $V_{\mathcal{D}}(\mu)$,

$$\kappa(x) \simeq N \frac{V_{\mathcal{D}}(\mu)}{V_{\mathcal{M}}} = N \frac{\Gamma(\frac{\mathcal{D}+1}{2})}{2\sqrt{\pi}\Gamma(\frac{\mathcal{D}}{2}+1)} \left(\frac{\mu}{R}\right)^{\mathcal{D}} \quad (2.47)$$

où $V_{\mathcal{M}} = S_{\mathcal{D}}(R)$, la surface d'une boule $(\mathcal{D} + 1)$ -dimensionnel de rayon R . Ce résultat¹⁰ n'est strictement valide que pour $\mu \ll \pi R$. On peut également trouver une solution pour le coefficient d'agrégation moyen [35] qui est approximé pour les grandes valeurs de \mathcal{D} par

$$\langle c \rangle \simeq 3\sqrt{\frac{2}{\pi\mathcal{D}}} \left(\frac{3}{4}\right)^{(\mathcal{D}+1)/2}. \quad (2.48)$$

2.3.3 Différentes probabilités de connexion

La probabilité de connexion est généralement une fonction Heaviside pour les RGGs standards. Cependant, lorsqu'on change $p(\mathbf{x}, \mathbf{y})$ pour une fonction continue, on s'attend à ce que cela ait un effet sur les propriétés structurelles. Dans la Réf. [37], on en fait une analyse rigoureuse. On considère

10. L'approximation utilisée vient du fait que l'on considère que $R = \mathcal{O}(N)$. Ceci revient à considérer que la densité des noeuds dans l'espace $\Delta = N/V_{\mathcal{M}} = \mathcal{O}(1)$ est indépendante de N . On peut ainsi négliger la courbure de l'espace et obtenir ce résultat.

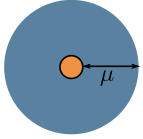
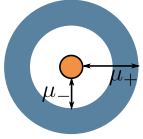
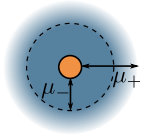
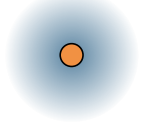
Nom	$p(\mathbf{x}, \mathbf{y})$	Schéma
Disque rigide [35, 57, 105]	$\begin{cases} 1 & d(\mathbf{x}, \mathbf{y}) < \mu \\ 0 & \text{autrement} \end{cases}$	
Anneau rigide [9]	$\begin{cases} 1 & \mu_- < d(\mathbf{x}, \mathbf{y}) < \mu_+ \\ 0 & \text{autrement} \end{cases}$	
Disque quasi-unitaire [77]	$\begin{cases} 1 & d(\mathbf{x}, \mathbf{y}) < \mu_- \\ \frac{\mu_+ - d(\mathbf{x}, \mathbf{y})}{\mu_+ - \mu_-} & \mu_- < d(\mathbf{x}, \mathbf{y}) < \mu_+ \\ 0 & \text{autrement} \end{cases}$	
Waxman [122]	$e^{-\beta d(\mathbf{x}, \mathbf{y})}$	

Tableau 2.2 – Quelques exemples de probabilités de connexion analysées dans la Réf. [37]. La zone en bleu indique une région où $p(\mathbf{x}, \mathbf{y})$ est non nulle : l'intensité du bleu désigne la valeur de la probabilité.

un espace euclidien bidimensionnel et une distribution des nœuds uniforme et on s'intéresse plus particulièrement à l'effet de la forme de $p(\mathbf{x}, \mathbf{y})$ sur la connectivité d'un RGG, c'est-à-dire sa capacité à connecter tous ses nœuds en une même composante¹¹.

Dans cet article, une distinction intéressante est faite entre les types de probabilité de connexion : si $p(\mathbf{x}, \mathbf{y})$ est une fonction qui peut seulement retourner les valeurs 0 ou 1, la fonction Heaviside par exemple, on dit de ces RGGs qu'ils sont *rigides*. À l'inverse, si $p(\mathbf{x}, \mathbf{y})$ est une fonction lisse, c'est-à-dire une fonction dont toutes les dérivées existent en tout point de son domaine, on dit que ces RGGs sont *souples*. Cette terminologie permet de classifier les ensembles de RGGs. Sur le Tableau 2.2, on présente quelques exemples de probabilités de connexion considérés dans la Réf. [37]. Leur analyse confirme que le choix de la probabilité de connexion a un impact sur la connectivité d'un réseau et que, par conséquent, certaines propriétés statistiques, dont $\langle c \rangle$ en est un exemple, vont changer.

11. Une *composante* est un sous-ensemble d'un réseau tel que toute paire de nœuds y appartenant est connectée par une chaîne de liens. On peut donc rejoindre toute paire de nœuds appartenant à une composante.

Cependant, d'autres propriétés, comme la distribution des degrés et la corrélation des degrés, vont rester les mêmes indépendamment de $p(\mathbf{x}, \mathbf{y})$. En effet, on peut se convaincre que le degré du nœud \mathbf{x} ne dépendra jamais de sa position, peu importe $p(\mathbf{x}, \mathbf{y})$, en raison de la nature homogène et isotrope de l'espace euclidien considéré. Par conséquent, la distribution des degrés sera toujours une distribution de Poisson.

Récemment, une forme particulière de probabilité de connexion a été identifiée dans le contexte des RGGs plongé dans l'espace euclidien 1D et permet de générer un ensemble de réseaux aléatoires dont le nombre de triangles, \bar{T} et le nombre de liens, \bar{M} , sont fixés en moyenne [74]. Cette probabilité de connexion prend la forme approximative suivante

$$p(x, y) \simeq \begin{cases} \frac{1}{1+e^{\beta[\varepsilon-\mu]}} & 0 < \varepsilon < 2\mu \\ \frac{1}{1+e^{\beta\mu}} & \text{autrement} \end{cases} \quad (2.49)$$

où $\varepsilon = |x - y|$ et x, y sont des variables aléatoires¹² données à chacun des nœuds et tirés uniformément sur l'intervalle $[0, 1]$. Cette forme Fermi-Dirac s'obtient en maximisant l'entropie de Shannon de cette distribution, sous les contraintes $\langle T \rangle = \bar{T}$ et $\langle M \rangle = \bar{M}$. Ce modèle est donc intimement lié au ERGM. De la forme de l'Éq. (2.49), on constate que cet ensemble de réseaux doit être également lié à celui des RGGs puisque les x peuvent être interprétés comme des positions sur un espace unidimensionnel et ε , comme la distance euclidienne entre deux points. Les champs auxiliaires $\beta > 0$ et $\mu > 0$, liés \bar{T} et \bar{M} , peuvent aussi être respectivement interprétés comme une température inverse et un potentiel chimique. Cette équivalence entre ce modèle et les RGGs unidimensionnels laisse croire que le nombre de triangles dans les réseaux complexes en général est la signature d'un espace métrique sous-jacent.

La validité de l'Éq.(2.49) n'a été démontrée que pour les RGGs unidimensionnels. Cependant, Krioukov prétend dans la Réf. [74] qu'une procédure similaire pourrait être utilisée pour des RGGs dans des espaces courbés et de plus hautes dimensions. La complexité de cette procédure pour tout autre espace de plus haute dimensions ou de courbure non-nulle rend difficile la complétion de la preuve qui demeure, à l'heure actuelle, un problème ouvert. Cependant, une forme similaire à été utilisée dans la Réf. [75] pour générer des RGGs dans des espaces hyperboliques bidimensionnels. Cette probabilité de connexion permettait un contrôle important sur le coefficient d'agrégation. Ceci soutient que le résultat de Krioukov pourrait en effet être une approximation valide indépendante de la nature de l'espace.

2.3.4 Différentes distributions de nœuds

On peut aussi étudier l'effet que la distribution des nœuds vont avoir sur les propriétés des RGGs. Cet effet est étudié plus particulièrement sur la distribution des degrés dans la Réf. [66]. Dans cet article, on considère un type de RGGs rigides avec seuil μ où les nœuds sont distribués selon $\rho(\mathbf{x})$

12. Dans cette représentation de réseaux, on note $p(x, y)$ comme étant un *graphon* [81]. Malgré les apparences, les graphons et les RGGs sont deux modèles différents, mais s'équivalent dans la limite $N \rightarrow \infty$.

sur une variété \mathcal{M} connexe, continue et \mathcal{D} -dimensionnelle. Dans un premier temps, on trouve que, dans la limite thermodynamique $N \rightarrow \infty$ et dans le régime *sparse* $\mu = \mathcal{O}(N^{-\mathcal{D}})$, le degré moyen des nœuds en \mathbf{x} ,

$$\kappa(\mathbf{x}) \simeq \delta \rho(\mathbf{x}) \quad (2.50)$$

où δ est une constante finie et positive. Dans ce régime, comme la distribution des degrés d'un nœud \mathbf{x} est une distribution de Poisson dans le régime *sparse*, il est possible de calculer une expression pour la distribution des degrés,

$$P(k) = \frac{\eta^k}{k!} \int [\rho(\mathbf{x})]^{k+1} e^{-\eta\rho(\mathbf{x})} d\mathbf{x}. \quad (2.51)$$

où $\eta = \frac{\langle k \rangle}{\int_{\mathcal{M}} \rho(\mathbf{x}) d\mathbf{x}}$. L'effet de la distribution des nœuds sur les autres propriétés structurelles, comme la corrélation des degrés, n'est cependant pas clair. On peut se convaincre qu'elles vont certainement changer en fonction de $\rho(\mathbf{x})$. Intuitivement, si on s'attend à ce que les nœuds dans le voisinage de \mathbf{x} ait un degré similaire approximé par $\kappa(\mathbf{x})$, alors on observera de l'assortativité par degré. Ceci pourrait certainement être une avenue de recherche intéressante pour le futur.

La notion de distribution de noeuds dans les RGGs offre une interprétation supplémentaire intéressante. Si on considère une distribution de noeuds non uniforme permettant aux noeuds de se regrouper en amas dans certaines régions plus denses du domaine, on peut généraliser la notion de communauté [54]. En fait, contrairement au concept de structure communautaire que nous avons présentée ultérieurement (Sec. 1.3), où les classes σ_i sont discrètes, on permet naturellement aux communautés d'avoir du chevauchement entre eux. Cette nouvelle définition des communautés en termes géométriques, appelées *communautés souples*, est expliquée en détails dans [54, 127].

2.3.5 Différents espaces

La forme de l'espace, qui fait varier naturellement la métrique $d(\mathbf{x}, \mathbf{y})$ et la densité $\rho(\mathbf{x})$, aura également un impact sur la structure. Cependant, lorsque l'espace est homogène et isotrope, c'est-à-dire dire le cas où la densité des nœuds peut être uniforme, sa courbure ne devrait pas affecter les propriétés structurelles. Ceci s'explique par le fait que, de par l'isotropie de l'espace, toutes les positions sont équivalentes et une translation du point d'origine du système de coordonnées dans n'importe quelle direction ne changera pas les propriétés de l'espace. Les nœuds sont donc toujours statistiquement équivalents, puisque l'on peut toujours placer un nœud donné à l'origine pour étudier ses propriétés.

Espace plat (euclidien)—Il est tout de même possible d'observer des changements lorsqu'on introduit des rebords à l'espace. Ainsi, même si la distribution des nœuds est uniforme, il est clair que les nœuds proches des rebords auront moins de voisins que ceux près du centre. Considérons donc l'exemple d'un espace euclidien en 2D, noté \mathbb{R}_R^2 contraint à un cercle de rayon R . On place ensuite les nœuds uniformément à l'intérieur de ce cercle et on les connecte s'ils sont séparés d'une distance inférieure à R . Les effets de bord sont ainsi maximisés. On peut donc s'attendre à observer qu'un

noeud au centre du disque sera connecté à tous les autres noeuds, et que le degré $\kappa(r)$ d'un noeud à une distance r du centre diminuera avec r . Afin d'empêcher que le réseau ainsi généré ait trop de liens, on choisit de réduire le degré des noeuds en y introduisant un paramètre supplémentaire, p , correspondant à la fraction des liens existants. Le degré moyen d'un noeud à distance r du centre, étant proportionnel à l'intersection du volume exclu d'un noeud (un disque de rayon R centré sur ce noeud) et l'espace disponible (un disque de rayon R également), $V(r)$, se calcule

$$\kappa(r) = \frac{Np}{\pi R^2} V(r) = \frac{2Np}{\pi} \left[\cos^{-1} \left(\frac{r}{2R} \right) - \frac{r}{2R} \sqrt{1 - \frac{r^2}{4R^2}} \right]. \quad (2.52)$$

Ceci permet de calculer le degré moyen $\langle k \rangle$, considérant que la distribution des noeuds sur ce cercle est uniforme, c'est-à-dire $\rho(r) = \frac{2r}{R^2}$,

$$\langle k \rangle = \int_0^R \rho(r) \kappa(r) dr = \frac{Np}{4\pi} (4\pi - 3\sqrt{3}) \Rightarrow p = \frac{4\pi \langle k \rangle}{N(4\pi - 3\sqrt{3})}. \quad (2.53)$$

Sur la Fig 2.4-(a), on montre un exemple du type de réseau obtenu.

Espace de courbure négative (hyperbolique)—En utilisant un processus similaire, on choisit maintenant un *espace courbé* comme un espace hyperbolique (voir Tab 2.1). Considérons comme espace de plongement, par simplicité, un disque fini de Poincaré de rayon R , noté \mathbb{H}_R^2 , dont la courbure $\zeta = 1$. Il est important de mentionner que, contrairement à la précédente définition de \mathbb{H}^2 , \mathbb{H}_R^2 n'inclut pas un espace hyperbolique infini. Dans ce cas, une distribution uniforme implique $\rho(r) = \frac{\sinh(r)}{\cosh(R)-1}$, et on peut montrer que le degré moyen d'un noeud r est approximé, pour $R \gg 1$, par [19, 75],

$$\kappa(r) = \frac{Np}{2\pi(\cosh(R) - 1)} V(r) \simeq \frac{4Np}{\pi} e^{-r/2}. \quad (2.54)$$

Ceci implique

$$\langle k \rangle \simeq \frac{8Np}{\pi} e^{-R/2} \Rightarrow R \simeq 2 \ln \left(\frac{8Np}{\pi \langle k \rangle} \right). \quad (2.55)$$

Ainsi, sans passer par une relaxation de p , on peut fixer $\langle k \rangle$ avec un choix approprié de R .

Il existe plusieurs différences fondamentales entre ces deux modèles. La plus importante est sans doute que le compromis entre $\rho(r)$ et $V(r)$ fait en sorte que la distribution des degrés du modèle hyperbolique est en loi de puissance avec un exposant $\gamma = 3$. La géométrie hyperbolique génère donc naturellement des réseaux *scale-free* ce qui n'est pas le cas des autres géométries. Il est possible de contrôler γ en choisissant une courbure $\zeta \neq 1$ ou en étirant la distribution radiale avec une constante α telle que,

$$\rho(r) = \alpha \frac{\sinh \alpha r}{\cosh \alpha R - 1} \quad (2.56)$$

Dans ce cas, l'exposant est donné par l'expression [75],

$$\gamma = \begin{cases} \frac{2\alpha}{\zeta} + 1 & \text{si } \frac{\alpha}{\zeta} \geq \frac{1}{2} \\ 2 & \text{si } \frac{\alpha}{\zeta} \leq \frac{1}{2} \end{cases} \quad (2.57)$$

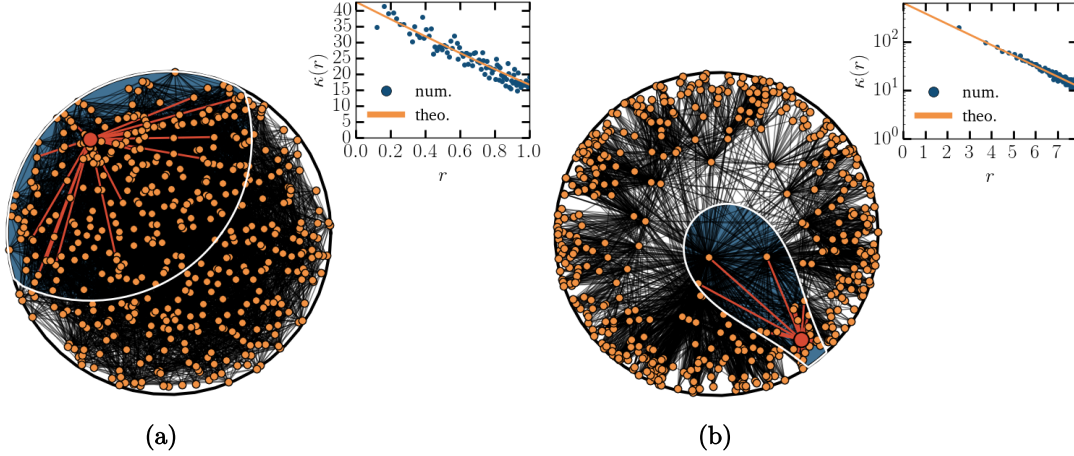


FIGURE 2.4 – RGGs dans des espaces avec rebord : (a) espace euclidien \mathbb{R}_R^2 avec $N = 500$, $R = 1$ et on fixe $\langle k \rangle = 25$ à l’aide de $p \simeq 0.085$ [Éq. (2.53)], (b) espace hyperbolique \mathbb{H}_R^2 avec $N = 500$, $p = 1$ et on fixe $\langle k \rangle = 25$ avec $R \simeq 7.86$ [Éq. (2.55)]. On montre également le degré moyen $\kappa(r)$ d’un nœud selon sa position radiale r . Sur les graphiques, les points bleus sont les données issues du réseau correspondant et les lignes oranges sont les solutions analytiques [Éq. (2.52) pour (a) et Éq. (2.54) pour (b)]. Dans chacun des schémas, on identifie un nœud aléatoirement en rouge ainsi que ses liens. Son volume exclu est également identifié par la région bleue délimitée par la ligne blanche.

Ce modèle hyperbolique a été étudié en détails dans les dernières années [19, 75, 76] et s’est adapté aux différents besoins (changement de distribution $\rho(r)$, de probabilité de connexion $p(\mathbf{x}, \mathbf{y})$, etc.). Dans ces travaux, la géométrie hyperbolique est vue comme étant l’espace de plongement le plus adéquat des réseaux *scale-free*, de par sa structure naturellement hiérarchique, sa correspondance avec le CM souple et avec l’attachement préférentiel (voir Sec. 2.4), d’où l’engouement autour de cette classe de modèles.

Les plus récentes avancées dans ce domaine cherchent à établir un lien entre la structure des réseaux complexes et celle de ces espaces hyperboliques. Plus particulièrement, des algorithmes sont actuellement développés pour inférer les coordonnées des nœuds de ces réseaux dans ce espace [19, 97, 98, 99]. La connaissance des coordonnées devient importante à plusieurs niveaux pour comprendre la relation entre la structure et la fonction d’un réseau. En effet, dans cette représentation, la structure est entièrement encodée dans l’espace. Ce faisant, il n’est plus nécessaire de connaître la matrice d’adjacence (une information qui contient $\mathcal{O}(N^2)$ éléments) pour décrire ces réseaux. Il suffit de connaître les coordonnées des nœuds (une information qui contient $\mathcal{O}(N)$ éléments). Une conséquence importante de ce résultat est que la géodésique entre deux nœuds positionnés dans cette espace approxime, dans la limite thermodynamique, le parcours le plus court les liant (*shortest paths*). Un exemple de cette observation se trouve sur la Fig. 2.5. De ce résultat découle des algorithmes très efficaces pour transmettre des messages et déterminer le parcours le plus court d’un nœud source à un nœud cible. Le plus connu est le *greedy routing*. Dans cet algorithme, les seules informations dont les nœuds disposent sont les coordonnées de leurs voisins et celle du nœud cible. Ainsi, un nœud qui doit transmettre un message à un de ses voisins pour qu’ultimement le message

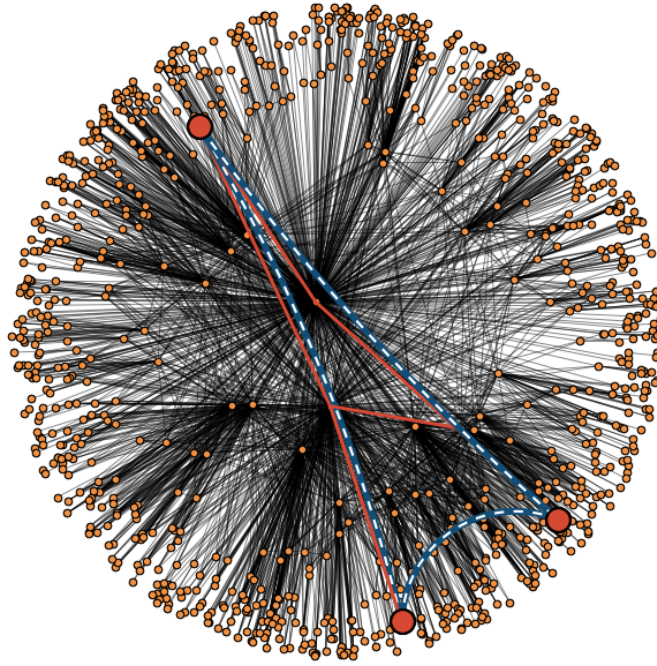


FIGURE 2.5 – Correspondance entre les géodésiques et les parcours les plus courts sur \mathbb{H}_R^2 . On identifie en rouge trois noeuds dont on a mis en évidence les parcours les plus courts en rouge également. Les lignes courbes pointillées en bleu correspondent aux géodésiques entre ces noeuds. Dans ce réseau, on fixe $N = 1000$, $\langle k \rangle = 15$ et $\gamma = 2.2$.

parvienne au noeud cible n'a qu'à choisir son voisin le plus proche, dans l'espace hyperbolique, de la cible.

L'étude de la géométrie hyperbolique des réseaux complexes est un domaine fascinant qui nécessiterait à lui seul un chapitre de ce mémoire. C'est pourquoi, en raison de contraintes d'espace, on envoie le lecteur vers les Réfs. [18, 76, 75, 19, 98, 97, 99, 53, 71, 6] pour en savoir davantage.

2.4 Modèles de croissance

Les modèles vus dans les Sections 2.1–2.2 donnent des outils pratiques pour étudier mathématiquement les propriétés structurelles des réseaux complexes. Toutefois, ils ne nous fournissent aucune explication sur l'évolution elle-même de ces systèmes. *Pourquoi un réseau aurait-il une certaine structure plutôt qu'une autre ?* Malheureusement, cette information est généralement inaccessible puisqu'elle nécessite un enregistrement continu du réseau. Or, il existe des modèles stochastiques, ap-

pelés *modèles génératifs*, permettant de modéliser les mécanismes d'évolution qui créent les noeuds et les liens d'un réseau à partir de processus simples. L'idée justifiant l'utilité de ces modèles est de pouvoir inférer le processus d'évolution d'un réseau complexe réel en comparant statistiquement ses propriétés structurelles avec celles de réseaux synthétiques générés par ce processus. Cette inférence demeure impossible dans la réalité puisque les réseaux complexes sont généralement créés selon des mécanismes également très complexes. On considère plutôt ces processus simples comme étant des *mécanismes effectifs*.

Les modèles génératifs de réseaux forment une grande famille de modèles dont les *modèles de croissance* font partie. Essentiellement, un modèle de croissance fait évoluer la structure en ajoutant à chaque temps t des nouveaux liens et des nouveaux noeuds¹³. Formellement, on définit un modèle de croissance comme un modèle de réseaux aléatoires muni d'une règle d'attachement dépendante du temps, soit une *probabilité de transition*, $p : \mathcal{G} \times \mathcal{G} \rightarrow [0, 1]$. En bref, cette règle évalue la probabilité $p(G(t) \rightarrow G(t+1) | t)$ qu'un réseau $G(t)$ au temps t évolue vers un autre réseau $G(t+1)$ au temps $t+1$. Cette règle d'attachement constitue l'ensemble des paramètres du modèle de réseaux aléatoires, $\Theta = \{p, \Theta_p\}$ où Θ_p est l'ensemble des paramètres de p . Bien que cette classe de modèles appartient à la famille des réseaux temporels [67], pour lesquels les noeuds et les liens peuvent être actifs ou inactifs en fonction du temps, la distinction est que les réseaux en croissance peuvent uniquement gagner de nouveaux noeuds ou de nouveaux liens. En d'autres termes, $|\mathcal{V}(t)| \leq |\mathcal{V}(t+1)|$ et $|\mathcal{E}(t)| \leq |\mathcal{E}(t+1)|$ pour tout t et tous réseaux en croissance.

2.4.1 Attachement préférentiel

Modèle de Simon—Le mécanisme de croissance le plus célèbre est sans aucun doute l'*attachement préférentiel* (PA). D'abord étudié dans le contexte de l'économie [113] par Herbert Simon, c'est un mécanisme qui favorise l'enrichissement des riches : *rich gets richer*. Dans le modèle de Simon, on considère un ensemble d'individus $\{u_i\}$ qui possèdent tous un certain nombre de ressources $k_i = k_i(t)$. Au temps t , on ajoute une ressource au système qui peut être donnée soit à un nouvel individu avec probabilité q ou à un individu existant avec probabilité $1 - q$. Ce dernier est choisi préférentiellement selon son nombre de ressources actuel. En général, la règle d'attachement s'exprime comme suit

$$p_S(k_i \rightarrow k_i + 1 | t) = p_S(k_i \rightarrow k_i + 1) = \begin{cases} q & \text{si } k_i = 0 \text{ (nouvel individu)} \\ (1 - q) \frac{k_i}{Z_S} & \text{si } k_i > 0 \text{ (individu existant)} \end{cases} \quad (2.58)$$

où

$$Z_S = \sum_{j=1}^N k_j = t \quad (2.59)$$

est une constante de normalisation qui correspond au nombre total de ressources dans le système et $\langle N \rangle = qt$. Ce type de systèmes est analogue à ceux que l'on retrouve en physique statistique,

13. Le temps $t \in \mathbb{N}$ ici est un compteur discret d'événements. Ainsi, puisque t compte les événements, la discrétisation du temps n'influence pas la qualité des modèles à représenter la réalité.

comme le modèle d'Ehrenfest [38, 40], qui sont résolubles à l'aide de l'équation maitresse. Dans le présent contexte, on utilise plutôt des équations compartimentales, inspirée de l'équation maitresse, qui décrivent la dynamique des *compartiments* d'individus ayant k ressources au temps t dont la taille est $N_k(t)$. Pour le modèle de Simon, les équations compartimentales ont la forme

$$\begin{aligned} N_k(t+1) &= N_k(t) + p_S(k-1 \rightarrow k)\delta(1, k) + p_S(k-1 \rightarrow k)N_k(t) - p_S(k \rightarrow k+1)N_k(t) \\ &= N_k(t) + q\delta(1, k) + (1-q) \left[\frac{(k-1)N_{k-1}(t)}{t} - \frac{kN_k(t)}{t} \right]. \end{aligned} \quad (2.60)$$

Le premier terme du côté droit correspond au nombre actuel d'individus ayant k ressources, le second correspond à l'ajout d'un nouvel individu possédant $k = 1$ ressource avec probabilité q , le troisième correspond au nombre d'individus qui vont entrer dans le compartiment k et le dernier correspond au nombre d'individus qui vont sortir du compartiment k vers le compartiment $k + 1$. Cette formulation en termes de la taille de compartiments est souvent moins pratique, puisque les $N_k(t)$ sont des quantités qui croissent avec le nombre de ressources total t . On choisit donc de travailler avec $P_k(t) = \frac{N_k(t)}{\langle N \rangle} = \frac{N_k(t)}{qt}$, qui est la proportion d'individus possédant k ressources. Les équations compartimentales se réécrivent ainsi

$$q(t+1)P_k(t+1) = qtP_k(t) + q\delta(1, k) + q(1-q) \left[(k-1)P_{k-1}(t) - kP_k(t) \right]. \quad (2.61)$$

Une particularité intéressante de ce genre de systèmes d'équations est qu'ils sont linéaires. Ils possèdent donc tous une *solution à l'équilibre* vers laquelle toute condition initiale tend. On définit la distribution à l'équilibre indépendante du temps comme $P_k^* \equiv P_k(t+1) = P_k(t)$ et elle se calcule à partir de la relation de récurrence suivante,

$$P_k^* = \frac{(1-q)(k-1)P_{k-1}^* + \delta(1, k)}{1+k(1-q)} \quad \Rightarrow \quad P_k^* = \frac{\prod_{k'=1}^{k-1} k'(1-q)}{\prod_{k'=1}^k [1+k'(1-q)]}, \quad (2.62)$$

qui peut s'écrire en termes d'une fonction Beta d'Euler, $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ où $\Gamma(x) = \int_0^\infty e^{-z} z^{x-1} dz$ est la fonction Gamma, telle que

$$P_k^* = \frac{B(k, \gamma_S)}{1-q} \quad (2.63)$$

avec $\gamma_S = \frac{2-q}{1-q}$. Lorsque $k \gg 1$, on approxime l'Éq. (2.63) à

$$P_k^* \simeq \frac{\Gamma(\gamma_S)}{1-q} k^{-\gamma_S}. \quad (2.64)$$

On voit ainsi que la distribution à l'équilibre des tailles des compartiments suit une loi de puissance d'exposant $\gamma_S = \frac{2-q}{1-q}$. Cette forme particulière de la distribution des ressources est présente dans un grand nombre de systèmes réels, autant en économie qu'en science des réseaux complexes (voir Sec. 1.4), et il est souvent inféré que la forme de cette distribution est due à *i.* la croissance du système et *ii.* un mécanisme effectif PA. En fait, on peut aisément montrer que de retirer l'un ou l'autre de ces deux mécanismes détruit la loi de puissance [86] pour laisser place à une distribution qui diminue exponentiellement avec k .

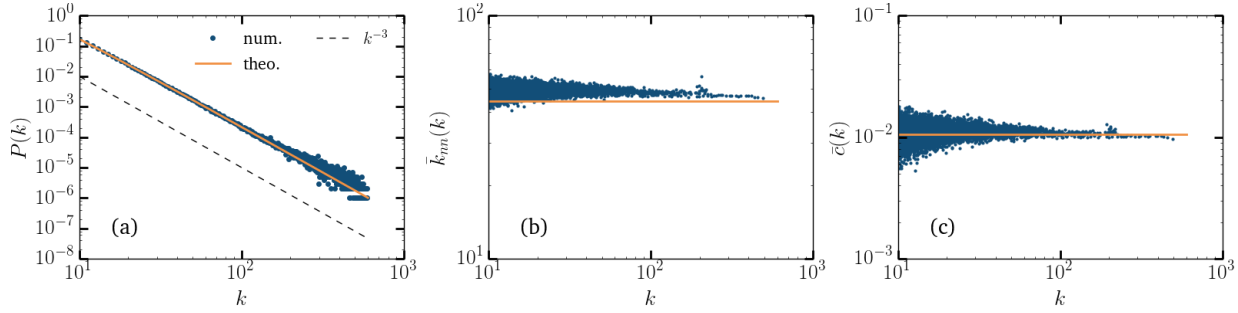


FIGURE 2.6 – Propriétés structurales dans le MBA : (a) distribution des degrés $P(k)$, (b) corrélation des degrés donnée par $\bar{k}_{nn}(k)$, (c) spectre du coefficient d’agrégation dépendant du degré $\bar{c}(k)$. On indique, par les points bleus, les données tirées de réseaux générés par ce modèle et, par les lignes pleines, les solutions analytiques associées [Éq. (2.71) pour (a) et Éq. (2.72) for (b)]. On fixe $N = 10^4$ et $m = 10$ et on moyenne ces résultats sur 100 simulations.

Modèle de Barabási-Albert—Évidemment, le modèle de Simon ne décrit pas l’évolution d’un réseau, mais celle d’individus qui accumulent une certaine ressource. Par contre, il est possible de l’adapter aux réseaux complexes. Dans ce cas-ci, ce sont les nouveaux liens qui sont distribués préférentiellement parmi les noeuds. Une des formes les plus simples de modèles de PA sur réseaux est le modèle de Barabási-Albert (MBA) [11]. Dans ce modèle, un nouveau noeud v_t est créé au temps t et se connecte à m noeuds v_s , où $s < t$, choisis selon une règle d’attachement similaire à celle du modèle de Simon,

$$p_{\text{BA}}\left(k_s \rightarrow k_s + 1 \mid t\right) = p_{\text{BA}}\left(k_s \rightarrow k_s + 1\right) = m \frac{k_s}{Z_{\text{BA}}} \quad (2.65)$$

où $k_s = k_s(t)$ est le degré du noeud v_s au temps t et

$$Z_{\text{BA}} = Z_{\text{BA}}(t) = \sum_{s < t} k_s = 2mt. \quad (2.66)$$

Le degré espéré d’un noeud v_s au temps t , noté $\kappa_s = \kappa_s(t)$, se calcule aisément en solutionnant l’équation différentielle suivante

$$\frac{d\kappa_s}{dt} = p_{\text{BA}}\left(\kappa_s \rightarrow \kappa_s + 1\right) = \frac{\kappa_s}{2t}. \quad (2.67)$$

En utilisant la condition initiale $\kappa_s(s) = m$, on trouve la solution

$$\kappa_s = m \left(\frac{s}{t}\right)^{-1/2}. \quad (2.68)$$

On comprend de cette expression que, puisqu’il s’agit d’une fonction qui décroît avec s , les vieux noeuds auront en moyenne plus de liens que les plus jeunes.

Pour calculer complètement la distribution des degrés, une approche par équations compartimentales similaire à celle du modèle de Simon est mieux adaptée [11]. De façon analogue au modèle de Simon,

on écrit les équations compartimentales suivantes

$$N_k(t+1) = N_k(t) + p_{\text{BA}}(k-1 \rightarrow k)N_{k-1}(t) - p_{\text{BA}}(k \rightarrow k+1)N_k(t) + \delta(m, k) \quad (2.69)$$

$$= N_k(t) + \frac{k-1}{2mt}N_{k-1}(t) - \frac{k}{2mt}N_k(t) + \delta(m, k) \quad (2.70)$$

et, posant $P_k(t) = \frac{N_k(t)}{t}$ ainsi que $P_k(t+1) = P_k(t) = P_k^*$, on obtient l'équation de récurrence pour la distribution à l'équilibre

$$P_k^* = \frac{(k-1)P_{k-1}^* - kP_k^*}{2} + \delta(m, k) \Rightarrow P_k^* = \frac{B(k, 3)}{B(m, 2)} \sim k^{-\gamma_{\text{BA}}} \quad (2.71)$$

où $\gamma_{\text{BA}} = 3$. Il est également possible d'utiliser le même procédé pour calculer la corrélation des degrés [73] et le coefficient d'agrégation [72]. D'une part, on trouve que le coefficient d'assortativité par degré r tend vers 0 dans la limite thermodynamique. Ceci implique nécessairement que la corrélation des degrés tend à être similaire à celle du CM (voir la Fig. 2.6-b) et donc,

$$\bar{k}_{nn}(k) \simeq \frac{\langle k^2 \rangle}{\langle k \rangle}. \quad (2.72)$$

D'autre part, on peut montrer que le coefficient d'agrégation est donné par (voir la Fig. 2.6-c)

$$\bar{c}(k) = \frac{m}{8} \frac{(\ln N)^2}{N} = \langle c \rangle. \quad (2.73)$$

Le coefficient d'agrégation local moyen $\langle c \rangle$ tend ainsi vers 0 dans la limite thermodynamique.

Ce mécanisme est souvent considéré comme à l'origine des réseaux *scale-free* en croissance. Cependant, il possède des lacunes considérables qui le limitent dans sa capacité à modéliser des réseaux complexes réels : *i.* l'exposant de la distribution des degrés est fixé à $\gamma_{\text{BA}} = 3$, *ii.* il ne saisit aucune corrélation des degrés et *iii.* le coefficient d'agrégation local est asymptotiquement indépendant du degré et nul dans la limite thermodynamique. Afin de rendre le modèle plus flexible, plusieurs travaux ont été entrepris et plusieurs variantes différentes du MBA ont été développées.

2.4.2 Attachement préférentiel avec densification

La *densification* peut être utilisée pour modifier la forme de la distribution des degrés du MBA. On nomme densification, l'action de créer un nouveau lien sans créer de nouveau nœud. Un exemple simple serait de considérer qu'à chaque temps t , un nouveau lien est créé dans le réseau et s'attache à un nœud v_s préférentiellement selon son degré. Ensuite, on choisit avec probabilité q de créer un nouveau nœud v_t pour compléter le lien ou, avec probabilité $1 - q$, de choisir un nœud v_s' déjà existant selon la règle d'attachement. En général, la règle d'attachement pour le modèle avec densification s'écrit

$$p_{\text{D}}(k_s \rightarrow k_s+1 | t) = p_{\text{D}}(k_s \rightarrow k_s+1) = \begin{cases} \frac{k_s}{Z_{\text{D}}} & \text{si } v_s \text{ est le premier bout du lien} \\ q & \text{si } v_s \text{ est le second bout et } k_s = 0 \text{ (nouveau nœud)} \\ (1-q) \frac{k_s}{Z_{\text{D}}} & \text{si } v_s \text{ est le second bout et } k_s > 0 \text{ (nœud existant)} \end{cases} \quad (2.74)$$

où

$$Z_D = \sum_{s < t} k_s = 2t. \quad (2.75)$$

Comme pour le modèle de Simon, il est important de mentionner que $N \leq t$ dans ce cas-ci, puisque les évènements correspondent à l'apparition de nouveaux liens, et que $\langle N \rangle = qt$.

L'approche par équations compartimentales nous donne la distribution à l'équilibre suivante,

$$P_k^* = \frac{2}{2-q} B(k, \gamma_D) \quad (2.76)$$

où $\gamma_D = \frac{4-q}{2-q}$. On note également que, lorsque $q = 1$, on retrouve le MBA original et $\gamma_D = \gamma_{BA}$.

Le paramètre q peut être généralisé à une fonction dépendante du temps $q \rightarrow q(t)$. Dans ce contexte, les nœuds vont naître et se densifier à différents taux lors de la croissance. Ce genre de modèles a été récemment étudié dans la Réf. [64]. Dans le cas des réseaux dont la distribution de degrés est en loi de puissance, ils ont d'ailleurs démontré que la forme de $q(t)$ sera contrainte et aura la façon suivante

$$q(t) = a(t + \tau)^{-\alpha} + b \quad (2.77)$$

où $\{\tau, \alpha, \beta\}$ sont des paramètres du modèle et a peut être fixé par la taille de la population.

2.4.3 Règle d'attachement non linéaire

Le MBA ainsi que la majorité de ses variantes considèrent une règle d'attachement linéaire avec le degré. Le modèle de Krapivsky-Redner (MKR) [73] quant à lui considère qu'un nouveau nœud v_t naît au temps t et forme un lien avec un nœud v_s déjà existant choisi selon la règle d'attachement non linéaire suivante

$$p_{\text{KR}}(k_s \rightarrow k_s + 1 | t) = p_{\text{KR}}(k_s \rightarrow k_s + 1) = m \frac{k_s^\nu}{Z_{\text{KR}}} \quad (2.78)$$

où $\nu \geq 0$ est le paramètre qui introduit la non linéarité et

$$Z_{\text{KR}} = \sum_{s < t} k_s^\nu. \quad (2.79)$$

On distingue alors trois régimes : le régime sous-linéaire $0 < \nu < 1$, le régime linéaire $\nu = 1$ et le régime sur-linéaire $\nu > 1$. Dans leur article, une approche par équations compartimentales leur permet de solutionner la distribution des degrés.

Dans le régime sous-linéaire, la distribution des degrés est en loi de puissance caractérisée par un exposant égal à $\gamma_{\text{KR}} \simeq \nu$ et soumis à une décroissance exponentielle en k qui varie en fonction de ν . Lorsque $\nu \rightarrow 1$, la décroissance exponentielle s'étire et fait augmenter l'exposant de la distribution des degrés $\gamma_{\text{KR}} \rightarrow \gamma_{\text{BA}} = 3$. À $\nu = 1$, la règle d'attachement du MKR est linéaire et le modèle est alors complètement équivalent au MBA avec $m = 1$. Dans le régime sur-linéaire, on observe la formation de quelques *étoiles*, c'est-à-dire des nœuds de haut degrés connectés presque uniquement à

des nœuds de degré 1, qui recevront la majorité des nouveaux liens créés à chaque temps. À $\nu = 2$, le nombre d'étoile tombe à un avec haute probabilité et on observe un scénario *winner-takes-all*, c'est-à-dire que le réseau ne sera formé que d'une seule étoile qui reçoit à chaque temps un nouveau voisin avec haute probabilité.

Il est important de souligner que de changer la fonction d'attachement a un impact considérable sur la distribution des degrés, mais ne change pas de façon significative la corrélation des degrés ou le coefficient d'agrégation [73]. L'assortativité observée dans les réseaux réels est visiblement liée à un processus d'évolution différent du PA. En fait, à ce jour, un processus d'évolution permettant de générer ce genre de propriétés structurelles n'a pas encore été découvert et reste une question ouverte. Au Chapitre 3, un processus d'évolution complètement différent de ceux liés au PA sera présenté où les corrélations de degrés sont variables selon les paramètres utilisés. Ce modèle est un premier candidat intéressant pour répondre à cette question ouverte.

2.4.4 Attachement préférentiel spatial

Bien que les corrélations des degrés soient encore difficiles à modéliser à partir d'un processus d'évolution, il se trouve que l'agrégation élevée est adéquatement saisie si on considère qu'un espace géométrique sous-jacent affecte la probabilité de créer les liens. En effectuant une combinaison entre le PA et une fonction de connexion dépendante de la distance séparant deux nœuds, on obtient un processus de croissance permettant de produire des réseaux *scale-free* dont le coefficient d'agrégation moyen demeure élevé dans la limite thermodynamique. Cette classe de processus se nomme PA spatial (SPA).

Deux versions du SPA se distinguent. Dans la première, on considère qu'à chaque temps t , comme dans le MBA, un nouveau nœud v_t est créé et forme m liens avec les nœuds v_s existants. Cependant, ces liens seront distribués proportionnellement aux degrés des nœuds v_s et proportionnellement à une fonction $f(d_{st})$ dépendante de la distance $d_{st} \equiv d(\mathbf{x}_s, \mathbf{x}_t)$ entre deux nœuds v_t et v_s [46, 68]. Dans ce cas, la règle d'attachement est

$$p(k_s \rightarrow k_s + 1 | d_{st} | t) = m \frac{k_s f(d_{st})}{Z_{\text{SPA}}} \quad (2.80)$$

où

$$Z_{\text{SPA}} = \sum_{s < t} k_s f(d_{st}). \quad (2.81)$$

Dans la Réf. [46], il montre que, lorsque \mathcal{M} est homogène et isotrope, la distribution des degrés forme une loi de puissance avec un exposant $\gamma_{\text{SPA}} = \gamma_{\text{BA}} = 3$. Cependant, lorsque \mathcal{M} est courbé et avec rebord ou que la fonction de connexion $f(d)$ est suffisamment rigide, par exemple une fonction Heaviside, la distribution des degrés diffère de plus en plus de celle du MBA.

Dans la deuxième version, le PA émerge d'un mécanisme similaire à celui du MBPS [21] où les variables cachées sont dépendantes du temps. On considère dans ce cas que la variable cachée d'un nœud v_s correspond à l'inverse de sa popularité au temps t donnée par $r_s(t) = \frac{\ln s + \ln t}{2}$ [98]. À

chaque temps t , on introduit un nouveau nœud v_t qui forme m liens avec les nœuds existants v_s , tels que tous $s < t$, avec probabilité de connexion

$$p(s, t) = \frac{1}{1 + e^{r_s(t) + r_t(t) - R(t)}} \quad (2.82)$$

où $R(t) = r_t(t) - \ln\left(\frac{1 - e^{-r_t(t)/2}}{m/2}\right)$ est une constante qui permet de fixer qu'un nouveau nœud créera en moyenne m lien à sa naissance. On note que cette probabilité est choisie de sorte qu'elle correspond à la distribution de liens du ERGM [Éq. (2.7)] lorsque les champs auxiliaires sont donnés par $\eta_{st} = r_s(t) + r_t(t) - R(t)$. Il existe donc une équivalence entre cette version du PA et le ERGM lui permettant de satisfaire le principe d'entropie maximale. De plus, à partir du MBPS, on calcule le degré espéré

$$\kappa_s(t) = m \left(\frac{s}{t}\right)^{-1/2} \quad (2.83)$$

qui correspond exactement à l'Éq. (2.68) du MBA. Cette version du PA est appelée *PA souple* et correspond à une formulation canonique du MBA. En effet, au lieu de créer les liens selon une règle d'attachement proportionnelle au degré des noeuds, les noeuds se connectent avec une probabilité de connexion qui, de manière effective, est une règle d'attachement proportionnelle au degré espéré, d'où l'appellation souple du modèle.

La représentation en termes de variables cachées de ce modèle permet également de démontrer sa correspondance, sous certaines conditions¹⁴, avec le CM lorsqu'on fixe la valeur moyenne de la distribution des degrés à une loi de puissance [128]. Cette correspondance intéressante permet de mettre l'emphase sur l'importance du PA pour la formation des réseaux complexes *scale-free* : en général, si un réseau en croissance est *scale-free*, il a probablement été généré de manière effective par une variante du PA.

On peut ensuite donner à un noeud v_s une seconde variable caché θ_s , soit une position angulaire sur le cercle \mathbb{S}^1 . Cette nouvelle variable cachée permet de définir la *similarité* entre deux noeuds, soit la distance angulaire entre eux sur ce cercle. Après l'introduction de cette seconde variable cachée, la probabilité de connexion se trouve modifiée de la façon suivante

$$p(s, t) = \frac{1}{1 + e^{d_{st} - R(t)}} \quad (2.84)$$

14. Dans la Réf. [128], on discute des différents niveaux de correspondance entre quatre modèles différents. Le premier est l'attachement préférentiel souple, noté SPA (*Soft Preferential attachment*, à ne pas confondre avec notre utilisation de SPA), le second est une transformation de SPA, noté $\widetilde{\text{SPA}}$, qui permet une densification du réseau à l'aide de *liens internes*, le troisième est le CM souple, noté SCM (*Soft Configuration Model*), et le dernier est une version *décalée* du SCM, noté SCM^+ , où on représente le SCM en termes de variables cachées à l'aide du MBPS. Ils introduisent une notion de *forte équivalence* entre deux modèles de réseaux aléatoires, par exemple $(\mathcal{G}_1, \mathbb{P}_1, \Theta_1)$ et $(\mathcal{G}_2, \mathbb{P}_2, \Theta_2)$, lorsque $\mathcal{G}_1 = \mathcal{G}_2$ et $\mathbb{P}_1 = \mathbb{P}_2$ pour une correspondance appropriée des paramètres $\Theta_1 \leftrightarrow \Theta_2$. Les auteurs arrivent à démontrer les relations de correspondance entre les modèles de la manière suivante :

$$\text{SPA} \rightsquigarrow \widetilde{\text{SPA}} \stackrel{\approx}{\approx} \text{SCM}^+ \stackrel{\cong}{\cong} \text{SCM}$$

où $\stackrel{\cong}{\cong}$ dénote une forte équivalence, $\stackrel{\approx}{\approx}$ est une forte équivalence approximée qui devient exacte dans la limite thermodynamique et lorsque $\langle k \rangle \ll N$, et \rightsquigarrow représente la transforme du modèle où les liens internes sont autorisés. Ces différentes conditions permettent d'établir le lien entre leur SPA et le SCM.

où $d_{st} = r_s(t) + r_t(t) + \ln[\theta(s, t)]$ et $\theta(s, t) = \pi - |\pi - |\theta_s - \theta_t||$ est la distance angulaire sur le cercle. Notons que d_{st} est approximativement égale à la distance sur le disque hyperbolique fini $\mathbb{H}_{R(t)}^2$ de rayon $R(t)$. De par la nature des coordonnées (r, θ) précédemment exposées en termes de variables cachées, on interprète $\mathbb{H}_{R(t)}^2$ comme un espace qui encode la *popularité* et la *similarité* des noeuds sur un réseau en croissance.

Chapitre 3

L'évolution géométrique des réseaux complexes

Article original :

Geometric evolution of complex networks with degree correlations

Charles Murphy, Antoine Allard, Edward Laurence, Guillaume St-Onge, Louis J. Dubé

Département de Physique, de Génie Physique et d'Optique,
Université Laval, Québec (Qc), Canada G1V 0A6

Référence : Phys. Rev. E, **97** (2018), p. 032309.

© 2018 The American Physical Society (Sec. 3.4–3.9)

3.1 Avant-propos

Jusqu'à maintenant, nous avons présenté les concepts et outils nécessaires au développement et à l'analyse d'un modèle de croissance de réseaux. Dans ce chapitre, nous présentons donc une application de ces divers concepts, un projet nommé *l'évolution géométrique des réseaux complexes*. Une relation amour-haine particulière entre ce projet et mon parcours s'est développée durant ces deux dernières années de par les renversements que ce projet a pris. Une brève explication de la genèse du projet s'impose.

Le cerveau est un système complexe, peut-être même *le* système complexe par excellence. Dans les dernières années, des avancées significatives en neurosciences, faites en partie grâce à la théorie des réseaux complexes, ont donné naissance à un nouveau domaine transdisciplinaire : *la connectomique* [50, 115]. Cette branche des neurosciences utilise les méthodes d'analyse développées dans le cadre des réseaux complexes pour étudier la structure et le fonctionnement du cerveau. À l'origine, ma tâche était de développer une expertise en connectomique au sein de notre groupe de recherche pour contribuer au développement de ce jeune domaine.

Mon pèlerinage en connectomique m'a amené à apprendre qu'il existe un certain nombre de choses importantes sur le connectome : son organisation en *small-world* [14], la présence du *rich-club* [118, 119] et son organisation hiérarchique [126]. Ces propriétés semblent toutes liées à la nécessité d'un équilibre entre la maximisation de l'efficacité du système à exécuter sa fonction et la minimisation de du coût total de ses liens [26]. Cette fonction de coût, dans le cadre du connectome, augmente en fonction de la longueur des axones et synapses et ainsi contraint fortement leur longueur. Ainsi, on comprend que le cerveau possède nécessairement des propriétés communes aux réseaux géométriques qui sont, eux aussi, contraints dans l'espace. Cependant, pour que le cerveau soit un système complexe efficace et fonctionnel, les *hubs* du *rich-club* violent ce principe en permettant la formation de liens à longue portée pour se connecter entre eux. Ainsi, une question naturelle se pose : pourquoi le connectome permet-il de créer ces liens fort coûteux ? Quel mécanisme est à l'origine de ces anomalies ? C'est avec cette question en tête que nous avons développé un modèle de croissance de réseaux géométriques à la fois simple et élégant.

Ce chapitre contient les résultats principaux de ce projet et constitue le sujet principal de mon mémoire. Il est divisé en deux parties. La première partie (les Sec. 3.4 à 3.9) correspond à un article qui a été accepté dans la revue *Physical Review E* en mars 2018 [84]. Elle est par conséquent rédigée en anglais, à l'exception d'une version française du résumé. La seconde partie (les Sec. 3.10 à 3.14) est composée de quatre compléments à ce travail. Les quatre premiers compléments (les Sec. 3.10 à 3.13) contiennent une analyse de l'impact du nombre de dimensions sur les réseaux générés, un commentaire sur les effets de taille fini, l'élaboration de la méthode numérique qui détermine les paramètres du modèle et une preuve de concept pour l'utilisation de ce modèle pour décrire l'évolution effective des réseaux réels. Enfin, la Sec. 3.13, en lien avec les premiers balbutiements du projet, est destinée à élaborer notre première approche au calcul de la distribution des degrés par équation maîtresse.

3.2 Résumé

Nous présentons une classe générale de mécanismes de croissance de réseaux par attachement homogène dans lesquels les liens créés à un temps donné t sont distribués de manière homogène entre le noeud nouveau-né et les noeuds existants sélectionnés uniformément. On y parvient en créant des nouveaux liens entre des noeuds distribués uniformément dans un espace métrique homogène et continu selon une probabilité de connexion de type Fermi-Dirac avec une température inverse β et un potentiel chimique $\mu(t)$ dépendant du temps et de forme générale. Le potentiel chimique permet de limiter la portée spatiale des liens nouvellement créés. En utilisant un formalisme de variables cachées, nous obtenons une expression analytique de la séquence des degrés et montrons que $\mu(t)$ peut être fixé pour obtenir n'importe quelle distribution des degrés donnée, incluant une distribution des degrés *scale-free*. De plus, nous montrons que, dépendamment de l'ordre d'arrivée des noeuds dans le réseau —son *histoire*—, la corrélation des degrés peut être modifiée pour être assortative ou désassortative. L'effet de la géométrie sur la structure est ensuite étudié à travers le coefficient d'agrégation moyen $\langle c \rangle$. Dans la limite thermodynamique, on identifie une transition de phase entre un régime aléatoire où $\langle c \rangle \rightarrow 0$ lorsque $\beta < \beta_c$ et un régime géométrique où $\langle c \rangle > 0$ lorsque $\beta > \beta_c$.

3.3 Abstract

We present a general class of geometric network growth mechanisms by homogeneous attachment in which the links created at a given time t are distributed homogeneously between a new node and the existing nodes selected uniformly. This is achieved by creating links between nodes uniformly distributed in a homogeneous and continuous metric space according to a Fermi-Dirac connection probability with inverse temperature β and general time-dependent chemical potential $\mu(t)$. The chemical potential limits the spatial extent of newly created links. Using a hidden variable framework, we obtain an analytical expression for the degree sequence and show that $\mu(t)$ can be fixed to yield any given degree distributions, including a scale-free degree distribution. Additionally, we find that depending on the order in which nodes appear in the network—its *history*—the degree-degree correlation can be tuned to be assortative or disassortative. The effect of the geometry on the structure is investigated through the average clustering coefficient $\langle c \rangle$. In the thermodynamic limit, we identify a phase transition between a random regime where $\langle c \rangle \rightarrow 0$ when $\beta < \beta_c$ and a geometric regime where $\langle c \rangle > 0$ when $\beta > \beta_c$.

3.4 Introduction

Random geometric graphs (RGGs) provide a realistic approach to model real complex networks. In this class of models, N nodes are located in a metric space \mathcal{M} and are connected if they are separated by a distance shorter than a given threshold distance μ [35, 105]. While RGGs are naturally associated with spatial networks [13]—such as infrastructure [4], transport [78, 60, 28], neuronal networks [26] and ad-hoc wireless networks [122, 77, 61] —, they can also be used to model real networks with no *a priori* geographical space embedding. In these geometric representations, nodes are positioned in a hidden metric space where the distances between them encode their probability of being connected [98, 97, 99]. This modeling approach allows to reproduce a wide range of topological properties observed in real networks, such as self-similarity [112], high clustering coefficient [74], scale-free degree distribution [76, 75], efficient navigability [18] and distribution of the links' weights [6].

This network geometry approach has been generalized to incorporate network growth mechanisms to further explain the observed structure of real networks under simple principles [48, 49, 46, 98, 97]. Two classes of mechanisms are considered in these approaches. The first one corresponds to a direct generalization of the classical preferential attachment (PA) coupled with a geometric mechanism : spatial or geometric preferential attachment [48, 49, 46] (SPA). In this class of models, nodes are added on a manifold at each time t , similarly to a geometric prescription, but connect with the existing nodes with a probability proportional to their degree and to a distance dependent function $f(d)$. However, the power-law behavior of the degree distribution remains robust to the choice of a specific $f(d)$ and the curvature of space.

The second class involves the interplay between two attractiveness attributes, popularity and similarity, which dictates the connection probability. Contrary to SPA, new nodes connect more likely to nodes of high popularity, denoted by a hidden variable $r(t)$ dependent on the time of birth t of nodes, and of high similarity, denoted by the angular distance between two nodes positioned on a circle. This model is usually referred to as a spatial growing network model with *soft* PA. Its evolution mechanism induces the power-law behavior of the degree distribution, but connects nodes proportionally to their expected degree instead of their real degree. By further tuning the positions of new nodes based on the spatial density of existing ones, this class of models is capable of generating scale-free networks with soft communities using geometry alone. All these network models have an interesting correspondence with static RGGs in hyperbolic geometry [76, 75, 98, 47, 127]. This suggests that the hidden space of real networks might be hyperbolic as well. In fact, several algorithms inferring the positions of nodes of real complex networks in the hyperbolic disk have been proposed based on these realistic models [98, 97, 99].

While the PA mechanism and hyperbolic geometry have been proved to naturally generate networks with power-law degree distribution, they do not capture the whole range of fundamental structural properties characterizing real networks. A good example is the assortative behavior of certain social networks such as scientific collaborations networks [87, 89, 90], film actor networks [7] and Pretty

good privacy web of trust networks [22]. The reason for the lack of assortativity in the PA and hyperbolic models is that they map to the *soft configuration model*: a maximum entropy ensemble in which the degree sequence is fixed with soft constraints such that no degree-degree correlations can be enforced [75, 128]. Growth mechanisms with custom degree-degree correlations are therefore still wanting.

We present a type of growing geometric network model which, in contrast with PA, distributes the links created by newborn nodes *homogeneously* among the existing ones. We call this attachment process *homogeneous attachment* (HA). From a geometric point of view, HA is interpreted as a growing geometric network mechanism where the connection threshold $\mu \equiv \mu(t)$ is a function of the time of birth of the newborn node. While previous works have considered similar scenarios in which the connection threshold is tailored to model social networks [106, 25, 24], to account for scale invariance [123] or to infer hyperbolic maps with external and internal links creation [97, 99], our parameter $\mu(t)$ is general and sufficiently versatile to reproduce a wide range of realistic structural properties. This feature allows the creation of an arbitrary number of links at each time enabling direct specification of the degree distribution *and* the degree-degree correlations.

The paper is organized as follows. In Sec. 3.5, the growing geometric network model is presented in detail. Section 3.6 is devoted to the development of an analytical expression for the degree of each node. This analytical description fixes $\mu(t)$ for any type of degree sequences, and therefore specifies the degree distribution. In Sec. 3.7, we show how the history of a network (the order of appearance of nodes) can be used to tune the degree-degree correlations without altering the degree distribution. In Sec. 3.8, the effects of the underlying geometry on the network topology are studied with special attention given to the average clustering coefficient $\langle c \rangle$. Finally, in Sec. 3.9 we draw some conclusions, limitations of the model and perspectives.

3.5 Growing geometric networks

The HA model, denoted by $(\mathcal{G}_{\text{HA}}, \mathbb{P}_{\text{HA}}, \Theta_{\text{HA}})$, is defined as follows. Let us consider the isotropic, homogeneous and borderless surface of a $(\mathcal{D} + 1)$ -ball of radius R as the metric space \mathcal{M} (dimension \mathcal{D}) in which the growing geometric networks are embedded. This choice simplifies the analytical calculations below, but does not alter the generality of our conclusions. Considering an initially empty network, the growing process goes as follows (see Fig. 3.1):

1. At any time $1 \leq t \leq N$, a new node (noted v_t) is assigned the random position \mathbf{x}_t uniformly distributed on \mathcal{M} .
2. Node v_t connects with the existing nodes v_s , such that any $s < t$, with probability $p(\mathbf{x}_t, \mathbf{x}_s; t)$.
3. Steps 2 and 3 are repeated until a total of N nodes have been reached.

In the model, $p(\mathbf{x}, \mathbf{y}; t)$ is a general function of the birth time t and the positions of the nodes \mathbf{x} and \mathbf{y} that we leave unspecified for the moment. Notice, however, that both the spatial and time dependencies have nontrivial effects on the network topology. On the one hand, the geometry encoded

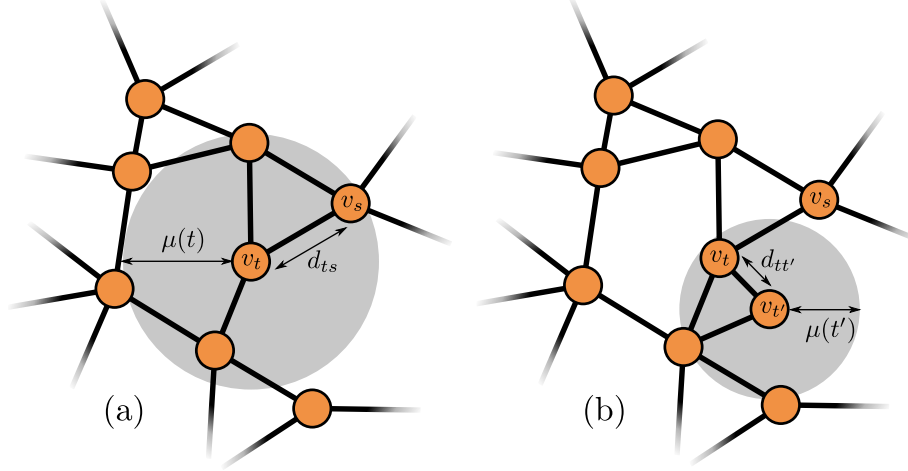


FIGURE 3.1 – Illustration of the HA growth mechanism for geometric networks : (a) birth of nodes v_t at time t which connects with three neighbors, including v_s at distance $d_{ts} < \mu(t)$, (b) subsequent birth of node $v_{t'}$, such that $t' > t$, connecting to v_t with $d_{tt'} < \mu(t')$. For illustration purposes, $p(\mathbf{x}_t, \mathbf{x}_s; \mu(t)) = \Theta(\mu(t) - d_{ts})$ such that any node within a circle of radius $\mu(t)$ (gray area) centered on node v_t will become connected to it.

in $p(\mathbf{x}, \mathbf{y}; t)$ will affect the properties of the networks like the distribution of component sizes [37] and the clustering coefficient [74]. On the other hand, the time dependence will determine *when* distant connections are allowed which, in turn, induce correlations between nodes with different birth times. For instance, if $p(\mathbf{x}, \mathbf{y}; t)$ is a decreasing function of t , older nodes will on average be hubs tightly connected to one another while younger nodes will have lower degrees. The choice of $p(\mathbf{x}, \mathbf{y}; t)$ can therefore induce a hierarchical core-periphery structure, where the hubs are strongly interconnected at the topological center of the structure and where the low degree nodes are mostly connected in chains with the center and occupy the outskirts of the network.

A natural and straightforward generalization of previous works [35, 105] is to add a time dependence to the probability of connection

$$p(\mathbf{x}, \mathbf{y}; \mu(t)) = \Theta(\mu(t) - d(\mathbf{x}, \mathbf{y})) = \begin{cases} 1 & \text{if } d(\mathbf{x}, \mathbf{y}) < \mu(t) \\ 0 & \text{otherwise} \end{cases}, \quad (3.1)$$

where $\Theta(d)$ is the Heaviside step function, $d(\mathbf{x}, \mathbf{y})$ is the metric distance between \mathbf{x} and \mathbf{y} and $\mu(t)$ is the connection threshold. Fixing $\mu(t) = \mu$ reduces to the known *sharp* RGG model which is deterministic in the creation of the links but not always suitable to describe real networks [122, 77, 9].

For more flexibility, we consider a connection probability analog to the Fermi-Dirac distribution

$$p(\mathbf{x}, \mathbf{y}; \mu(t), \beta) = \frac{1}{\exp\left(\beta[d(\mathbf{x}, \mathbf{y}) - \mu(t)]\right) + 1}, \quad (3.2)$$

which fixes the parameters of the HA model to $\Theta_{\text{HA}} = \{\mathcal{D}, \beta, \mu(t)\}$, where $\beta > 0$ is a parameter controlling the clustering coefficient and $\mu(t)$ limits the spatial extent of new links [76, 74]. From a

statistical physics point of view, using this connection probability amounts to consider the links as fermions of energy given by the length of the links, $d(\mathbf{x}, \mathbf{y})$, embedded in an environment maintained at temperature $1/\beta$ with a chemical potential $\mu(t)$.

This connection probability is interesting for two reasons. First, it is very similar to the connection probability of the exponential random graph model. This network ensemble maximizes the Gibbs entropy when the average number of links between any given pair of nodes is fixed [100]. Second, varying β enables us to navigate between the hot regime $\beta \rightarrow 0$ and the cold regime $\beta \rightarrow \infty$. In the limit $N \rightarrow \infty$ and when $\mu(t) = \mu$, the connection probability in the hot regime no longer depends on the position of the nodes, and the corresponding network ensemble is of the Erdős-Rényi type, where $\langle c \rangle = \mathcal{O}(N^{-1}) \rightarrow 0$ [74]. In contrast, in the cold regime, under the same conditions, $\langle c \rangle$ reaches a maximum independent of N [74].

3.6 Degree Sequence

Since $\mu(t)$ limits the spatial extent of potential connections, it has a direct impact on the degree sequence. In this section, we shed light on the relation between $\mu(t)$ and the resulting structure.

3.6.1 Hidden variables

Homogeneous attachment (HA) is a growing mechanism in which new nodes connect to existing nodes with a same probability dependent on their birth time. A convenient way to analyze this mechanism is via the MBPS framework of random graphs with hidden variables [21]. In this class of models, each node $i = 1, \dots, N$ is assigned a hidden variable h_i , sampled from a probability distribution $\rho(h)$, and links are created between nodes with respective hidden variables h and h' with probability $p_{\text{BPS}}(h, h')$. This general model encodes the correlation among nodes via the hidden variables, which can either be random numbers h or vectors of random numbers \mathbf{h} . Although this model is very general and versatile, it is nevertheless amenable to a full mathematical description of the structural properties of the network ensemble such as the degree distribution, the correlations and the clustering.

In our model, there are two hidden variables involved : the time of birth t and the position $\mathbf{x} = (x_1, x_2, \dots, x_{\mathcal{D}+1})$ on the surface \mathcal{M} of the $(\mathcal{D} + 1)$ -ball such that the associate hidden variables are $\mathbf{h} = (t, \mathbf{x})$. Whereas \mathbf{x} is a random variable distributed uniformly on \mathcal{M} by definition, the same cannot be said straightforwardly for t . However, since one node was born at all $t = 1, \dots, N$, randomly choosing a node of birth time t will occur uniformly. This is sufficient to provide a mathematical description of the network ensemble using the hidden variable framework. The hidden variable probability distribution is then

$$\rho(\mathbf{h}) = \frac{1}{NS_{\mathcal{D}}(R)} = \frac{\Gamma\left(\frac{\mathcal{D}+1}{2}\right)}{N2\pi^{(\mathcal{D}+1)/2}R^{\mathcal{D}}} \quad (3.3)$$

where $S_{\mathcal{D}}(R)$ is the surface of the $(\mathcal{D} + 1)$ -ball of radius R and $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is the Gamma function.

For the sake of simplicity, we consider the characteristic time $T \equiv N$ and length $X \equiv \pi R$ to define the normalized hidden variable $\tilde{\mathbf{h}} = (\tau, \boldsymbol{\xi})$, where

$$\tau \equiv \tau(t) = \frac{t}{T} = \frac{t}{N}, \quad (3.4a)$$

$$\boldsymbol{\xi} \equiv \boldsymbol{\xi}(\mathbf{x}) = \frac{\mathbf{x}}{X} = \frac{\mathbf{x}}{\pi R}, \quad (3.4b)$$

and the normalized quantities

$$\tilde{R} \equiv \frac{R}{X} = \frac{1}{\pi}, \quad (3.5a)$$

$$\tilde{\mu}(\tau) \equiv \frac{\mu(\tau T)}{X} = \frac{\mu(\tau N)}{\pi R}, \quad (3.5b)$$

$$\tilde{\beta} \equiv \beta X = \beta \pi R. \quad (3.5c)$$

This choice of normalization implies that

$$\rho \equiv \rho(\tilde{\mathbf{h}}) = \frac{\Gamma\left(\frac{\mathcal{D}+1}{2}\right)}{2\pi^{(1-\mathcal{D})/2}} \quad (3.6)$$

becomes a constant. Note also that as N tends to infinity, the difference between birth times, $\Delta\tau = N^{-1}$, tends to zero which allows to consider the birth times τ as a continuous variable in $[0, 1)$ to facilitate the analytical calculations. Additionally, because \mathcal{M} is homogeneous and isotropic, all analytical calculations will always consider the referenced node to be at $\mathbf{0} \equiv (\tilde{R}, 0, \dots, 0)$ without loss of generality. Finally, considering two nodes with hidden variables $\tilde{\mathbf{h}} = (\tau, \boldsymbol{\xi})$ and $\tilde{\mathbf{h}}' = (\tau', \boldsymbol{\xi}')$, the connection probability becomes

$$\begin{aligned} p_{\text{BPS}}(\tilde{\mathbf{h}}, \tilde{\mathbf{h}}') &= \Theta(\tau - \tau') p(\boldsymbol{\xi}, \boldsymbol{\xi}'; \tilde{\mu}(\tau), \tilde{\beta}) \\ &\quad + \Theta(\tau' - \tau) p(\boldsymbol{\xi}', \boldsymbol{\xi}; \tilde{\mu}(\tau'), \tilde{\beta}) \end{aligned} \quad (3.7)$$

where the step functions select the appropriate probability depending on which nodes appeared first.

3.6.2 Fixing the degree sequence

Since the new links are distributed homogeneously among the existing nodes at any given time, the expected degree of node v_τ with birth time τ *at the end of the process* has the following simple form

$$\kappa(\tau) = N \left[\tau n(\tau) + \int_\tau^1 n(\tau') d\tau' \right], \quad (3.8)$$

where

$$n(\tau) = \int_{\mathcal{M}} p(\mathbf{0}, \boldsymbol{\xi}; \tilde{\mu}(\tau), \tilde{\beta}) \rho d\boldsymbol{\xi} \quad (3.9)$$

which corresponds to the probability that node v_τ will connect to any existing nodes at its birth. This integral can be solved analytically when $\mathcal{D} = 1$ leading to a closed form expression,

$$n(\tau) = \frac{1}{\tilde{\beta}} \ln \left(\frac{\exp(-\tilde{\beta}) + \exp(-\tilde{\beta}\tilde{\mu}(\tau))}{1 + \exp(-\tilde{\beta}\tilde{\mu}(\tau))} \right). \quad (3.10)$$

For $\mathcal{D} \neq 1$, Eq. (3.9) must be solved numerically (see Sec. 3.10). Each term of Eq. (3.8) can be interpreted explicitly : the first one corresponds to the number of links which node τ creates on its arrival while the second term accounts for the links it gains by the creation of the other nodes.

The average degree of node v_τ can be obtained as a function of $\tilde{\mu}(\tau)$ using Eq. (3.9). Indeed, Eq. (3.8) can be inverted such that we obtain $n(\tau)$ as a function of the degree sequence $\kappa(\tau)$. We start by deriving Eq. (3.8) with respect to τ such that we obtain the following differential equation

$$\frac{d\kappa(\tau)}{d\tau} = N\tau \frac{dn(\tau)}{d\tau}. \quad (3.11)$$

Then, solving for $n(\tau)$ by integrating by part from τ to 1, this yields

$$n(\tau) = \frac{1}{N} \left[\frac{\kappa(\tau)}{\tau} - \int_\tau^1 \frac{\kappa(\tau')}{\tau'^2} d\tau' \right] = \mathcal{O}(N^{-1}). \quad (3.12)$$

This represents one of the most interesting assets of the HA model : given an ordered degree sequence in time $\kappa(\tau)$, it is possible to calculate analytically $n(\tau)$ to obtain the appropriate form of $\tilde{\mu}(\tau)$ via Eq. (3.9) which, in the case $\mathcal{D} = 1$, yields

$$\tilde{\mu}(\tau) = \frac{1}{\tilde{\beta}} \ln \left(\frac{1 - \exp(-\tilde{\beta}n(\tau))}{\exp(-\tilde{\beta}n(\tau)) - \exp(-\tilde{\beta})} \right). \quad (3.13)$$

It is therefore possible to obtain the corresponding function $\tilde{\mu}(\tau)$ to reproduce any given degree sequence.

It is important to understand at this point that what is fixed here is the ordered expected degree sequence in time, which is quite different from a configuration model prescription. For example, an increasing degree sequence in time would yield a totally different structure than a decreasing one. We will show that this additional trait of the model allows to tune the level of correlations between the degree of nodes.

3.6.3 Scale-free growing geometric networks

Scale-free networks ($P(k) \propto k^{-\gamma}$) can be generated with our model by assuming that the ordered degree sequence has the following form

$$\kappa(\tau) = \nu\tau^{-\alpha}, \quad (3.14)$$

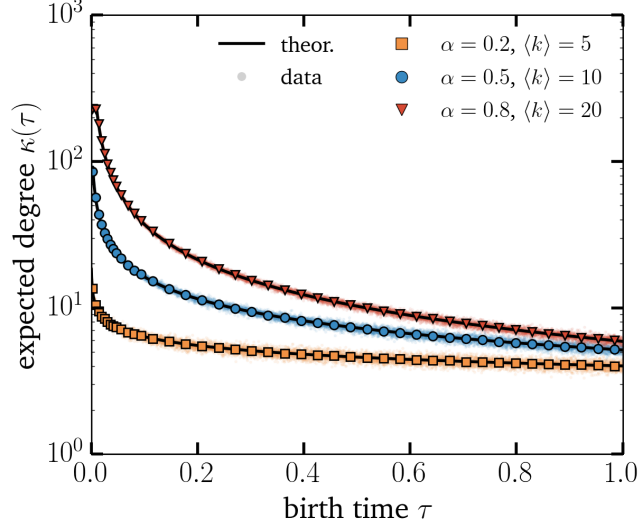


FIGURE 3.2 – Expected degree $\kappa(\tau)$ as a function of the birth time τ for different parameters. The small dots correspond to data obtained from Monte Carlo simulations averaged over 100 instances where $\tilde{\mu}(\tau)$ is given by Eq. (3.13) with $n(\tau)$ given by Eq. (3.15). The markers (squares, circles and triangles) correspond to the average data. The black solid lines corresponds to Eq. (3.14). The parameters of each dataset are indicated on the plot. The networks have been averaged over 96 instances.

where $\nu > 0$ fixes the average degree $\langle k \rangle$, and where $0 \leq \alpha < 1$ amounts for the heterogeneity of the degree sequence and controls the exponent of the degree distribution via $\gamma = 1 + \frac{1}{\alpha}$. It is then possible to calculate explicitly $n(\tau)$ using Eq. (3.12)

$$n(\tau) = \frac{\nu}{N(\alpha + 1)} (\alpha\tau^{-\alpha-1} + 1) . \quad (3.15)$$

Figure 3.2 shows $\kappa(\tau)$ for scale-free geometric networks with different sets of parameters where the chemical potential is given by Eq. (3.13) (see Fig. 3.3). Clearly, Eq. (3.14) correctly describes the behavior of the degree sequence.

3.6.4 Finite-size effects

The agreement between the theoretical predictions and the numerical simulations demonstrates it is indeed possible to reproduce any degree distributions with an appropriate choice of $\tilde{\mu}(\tau)$. There are some limitations however. It is possible for the ordered sequence of expected degrees to be such that Eq. (9) yields $n(\tau) > 1$ meaning that node v_τ would have to create more links than the number of already existing nodes at the moment of its birth (time τ). In the case of scale-free networks with an ordered degree sequence of the form of Eq. (3.14), this happens to any nodes with birth time $\tau < \tau^*$ where

$$\tau^* = \left[\frac{\alpha\nu}{N(\alpha + 1) - \nu} \right]^{\frac{1}{\alpha+1}} , \quad (3.16)$$

which corresponds to the plateau seen on Fig. 3.3, and implies that any nodes with birth time $\tau < \tau^*$ will form a connected clique of hubs. Limiting $n(\tau)$ such that $n(\tau) < 1$ further implies that some

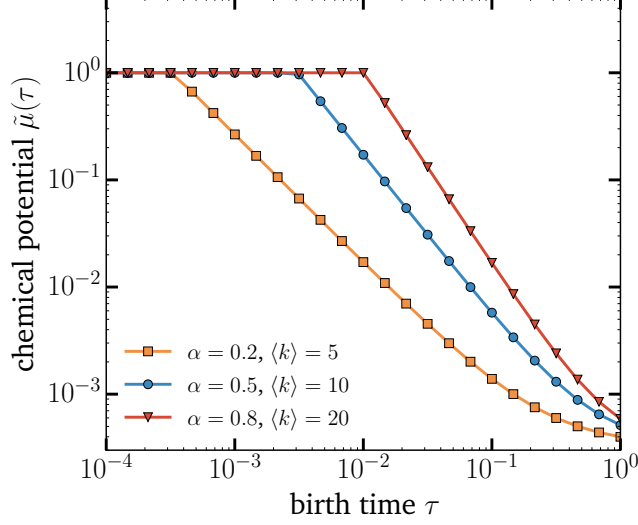


FIGURE 3.3 – Chemical potential $\tilde{\mu}(\tau)$ vs τ used to generate scale-free geometric networks with different values of α and $\langle k \rangle$ indicated on the plot. The lines and markers correspond to Eq. (3.13) with $n(\tau)$ given by Eq. (3.15). We fixed $N = 10^4$ and $\tilde{\beta} \rightarrow \infty$.

links will be missing, and consequently that the degree sequence will differ from the one given by Eq. (3.14). This discrepancy can be investigated through the average degree which, without this correction, is equal to $\langle k \rangle = \frac{\nu}{1-\alpha}$. Considering the clique of connected hubs yields instead

$$\begin{aligned}
 \langle k \rangle &= \int_0^1 \kappa(\tau) d\tau \\
 &= \tau^* \kappa(\tau^*) + \int_{\tau^*}^1 \kappa(\tau) d\tau \\
 &\simeq \frac{\nu}{1-\alpha} \left(1 - \alpha \tau^{*1-\alpha} \right).
 \end{aligned} \tag{3.17}$$

which holds for $\alpha \ll 1$. In general, the size of the plateau scales as $\mathcal{O}(N^{-\delta})$, where $\frac{1}{2} < \delta < 1$, which vanishes in the limit $N \rightarrow \infty$ for all $0 \geq \alpha < 1$ (Sec. 3.11). One possible way to circumvent this effect would be to allow multilinks and self-loops, but this is left as a future improvement of the model.

3.7 Network history

The degree sequence can take many forms : choosing Eq. (3.14) for the degree sequence yields scale-free networks. Yet, this is only one example of degree sequence capable of generating networks with a power-law degree distribution. Another example of such degree sequence would be to choose all the entries of $\kappa(\tau)$ at random from a distribution $P(k) \propto k^{-\gamma}$. However, these two growth processes, despite having the same degree distribution, have different $\tilde{\mu}(\tau)$ which in turn affect the structural organization of the generated networks.

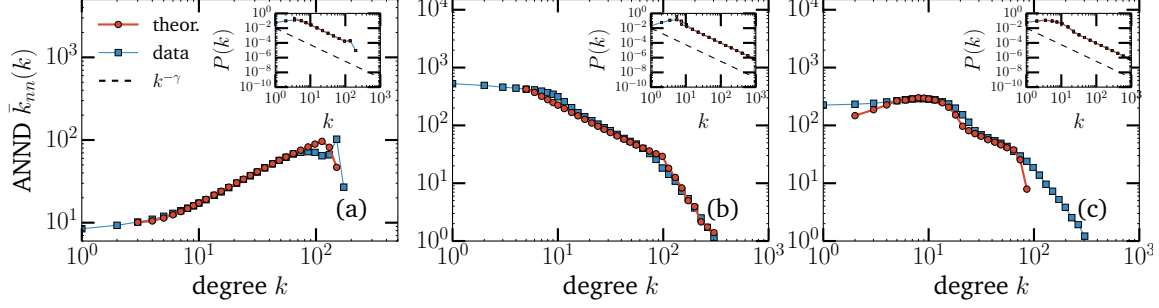


FIGURE 3.4 – Degree-degree correlations for different histories : (a) decreasing degree order (Sec. 3.7.2), (b) increasing degree order (Sec. 3.7.3), (c) random order (Sec. 3.7.4). In each plot, the results from Monte Carlo simulations are indicated by the squares and the analytics by the circles. The degree distributions are also displayed in inset and are power laws with $\gamma = 2.25$ (dashed lines). The average degree assortativity coefficient $\langle r \rangle$ amounts to (a) $\langle r \rangle \simeq 0.68$, for (b) $\langle r \rangle \simeq -0.12$ and for (c) $\langle r \rangle \simeq -0.11$. For each simulation, $\alpha = 0.8$, $\langle k \rangle = 10$, $N = 10^4$ and $\beta \rightarrow \infty$ (Eq. (3.1) as the connection probability). The results have been averaged over 96 instances.

Reordering the degree sequence amounts to changing the *history* of the network. We define a history as an ordered set $H = (\tau_i)$ of the birth time, where τ_i is the birth time of a node v_i ¹ of fixed final degree k_i . For a specific history, the final expected degree of node v_i is simply

$$\kappa(\tau_i) = k_i. \quad (3.18)$$

A network where the degree sequence $\mathbf{k} = (k_i)$ is known can have different histories. Unlike most network growth models, in HA the degree sequence is preserved even if the network history is changed. This particularity makes HA a unique alternative to model growing networks because it induces a specific correlation between nodes born at a different times. In other words, the change of network history affects the correlation between nodes.

3.7.1 Degree-degree correlations

To quantify the impact of the history on the resulting network structure, we consider the degree-degree correlation. This measure is fully characterized by the conditional probability that a node of degree k is connected to another node of degree k' denoted by $P(k'|k)$.

We express the degree-degree correlations in terms of the birth times as the conditional probability $p(\tau'|\tau)$ that node v_τ is connected to node $v_{\tau'}$ given by

$$p(\tau'|\tau) = \frac{N}{\kappa(\tau)} \left[\Theta(\tau - \tau')n(\tau) + \Theta(\tau' - \tau)n(\tau') \right]. \quad (3.19)$$

However, it is usually more convenient to calculate the average degree of nearest-neighbors (ANND)

1. Until now, the birth time τ of node v_τ was sufficient to label it. However, in cases where the history is unknown, it is more appropriate to label the nodes according to some integer i unrelated to its birth time.

denoted by $\kappa_{nn}(\tau)$ and defined by

$$\begin{aligned}\kappa_{nn}(\tau) &= \int_0^1 \kappa(\tau') p(\tau'|\tau) d\tau' \\ &= N \left[\int_0^\tau \frac{\kappa(\tau') n(\tau)}{\kappa(\tau)} d\tau' + \int_\tau^1 \frac{\kappa(\tau') n(\tau')}{\kappa(\tau)} d\tau' \right].\end{aligned}\quad (3.20)$$

From this expression, the degree-dependent ANND, denoted $\bar{k}_{nn}(k)$, can be obtained via the hidden variable framework

$$\bar{k}_{nn}(k) = 1 + \frac{1}{P(k)} \int_0^1 g(k|\tau) \kappa_{nn}(\tau) d\tau, \quad (3.21)$$

where $g(k|\tau) = \frac{e^{-\kappa(\tau)} \kappa(\tau)^k}{k!}$ [21]. Having this analytical expression in hand, we can now investigate the effect of different histories on the degree-degree correlations.

3.7.2 Decreasing order of degrees

As we have seen in Sec. 3.6, scale-free degree sequences can be written as $\kappa(\tau) = \nu\tau^{-\alpha}$. This implies a specific type of network history where the degree sequence is a decreasing order of the degrees : hubs are old while low degree nodes are young. The birth time correlations can then be calculated using Eq. (3.14) and (3.15)

$$\kappa_{nn}(\tau) = \frac{\kappa(\tau)}{2(1-\alpha)} [1 + \tau^{2\alpha}], \quad (3.22)$$

which is well approximated, for $\tau \ll 1$, by $\kappa_{nn}(\tau) \simeq \frac{\kappa(\tau)}{2(1-\alpha)}$. Then, the $\bar{k}_{nn}(k)$ can be calculated and is given, for large k , by

$$\bar{k}_{nn}(k) \simeq \frac{k}{2(1-\alpha)}. \quad (3.23)$$

Since $\bar{k}_{nn}(k)$ is essentially a linear function of k , it shows that choosing a history ordered by degree yields assortative networks. Figure 3.4(a) confirms these predictions.

3.7.3 Increasing order of degrees

We now consider an ordering in which the old nodes are assumed to be the low degree ones and the young nodes are the hubs. To generate scale-free networks with an increasing degree order, we use

$$\kappa(\tau) = \nu(1-\tau)^{-\alpha}, \quad (3.24)$$

which yields

$$n(\tau) = \frac{\nu}{N} \left[\frac{(1-\tau)^{-\alpha}}{\tau} - \int_\tau^1 \frac{(1-\tau')^{-\alpha}}{\tau'^2} d\tau' \right]. \quad (3.25)$$

As illustrated on Fig. 3.4(b), an increasing degree history implies a decreasing $\bar{k}_{nn}(k)$ which corresponds to disassortative networks. This can be explained by the fact that, since all the low degree nodes are created early in the history, when the time comes for the hubs to be born, they will connect more frequently to them. Combined with the results of Section 3.7.2, these results show how the degree-degree correlations can be tuned simply by changing the history we consider.

3.7.4 Random order

Our last example is when the set H is random. The network history is then composed of birth times τ_i that are random variables distributed uniformly between $[0, 1)$. From the point of view of the degree sequence, this means that $\kappa(\tau)$, and consequently any function of τ , is also a random variable but drawn, in the case of $\kappa(\tau)$, from the degree distribution. Because of this, $\kappa(\tau)$ is no longer a continuous function and Eqs. (3.12) and (3.20) can no longer be used. Instead, we use the following discrete form

$$n(\tau) = \frac{1}{N} \left[\frac{\kappa(\tau)}{\tau} - \sum_{i=t+1}^N \frac{\kappa(\tau_i)}{\tau_i^2} \Delta\tau_i \right], \quad (3.26)$$

where $t \equiv \tau N$ is the number of nodes when node v_τ is born, τ_i is the birth time of the i^{th} node to arrive in the network and $\Delta\tau_i = \tau_i - \tau_{i-1}$, with $\tau_i = 0$ if $i \leq 0$, is the time step between two birth events. In a similar way, the ANND can be adapted as well

$$\kappa_{nn}(\tau) = N \left[\sum_{i=1}^t \frac{\kappa(\tau_i)n(\tau)}{\kappa(\tau)} \Delta\tau_i + \sum_{j=t+1}^N \frac{\kappa(\tau_j)n(\tau_j)}{\kappa(\tau)} \Delta\tau_j \right]. \quad (3.27)$$

These expressions can be obtained by evaluating the integrals in Eq. (3.12) and (3.20) in the form of Riemann sums. Therefore, in the thermodynamic limit, $\Delta\tau_i \rightarrow 0$ for all i and the continuous and discrete forms are totally equivalent.

As before, to generate scale-free networks under this process, we would have to first determine the degree sequence $\kappa(\tau) \sim P(k) \propto k^{-\gamma}$ and then determine the corresponding $\tilde{\mu}(\tau)$. With that procedure one generates networks with a random history.

As we can see on Fig. 3.4(c), similarly to the increasing degree history, $\bar{k}_{nn}(k)$ is a decreasing function and the networks show disassortativity with $\langle r \rangle = -0.14$. However, the disassortativity observed here is entirely due to structural constraints imposed by the degree sequence: the degree sequence forces the hubs to connect more frequently to the low degree nodes.

3.8 Effects of geometry

The conclusions drawn so far are general, whether the networks are geometric or not. The effect of the geometry becomes manifest at the level of the three-node correlations where the triangle inequality of the underlying metric space implies a non-vanishing clustering coefficient $\langle c \rangle$ in the thermodynamic limit. The choice of $p(\boldsymbol{\xi}, \boldsymbol{\xi}'; \tilde{\mu}(\tau), \tilde{\beta})$ as a Fermi-Dirac distribution [Eq. (3.2)] allows us to adjust the level of clustering by changing the inverse temperature $\tilde{\beta}$.

The average clustering coefficient C is the fraction of triplets—three nodes connected in chains—actually forming a triangle. Adapting this coefficient to each node instead of the whole network yields the local clustering coefficient c , where c corresponds to the fraction of a node's neighbors

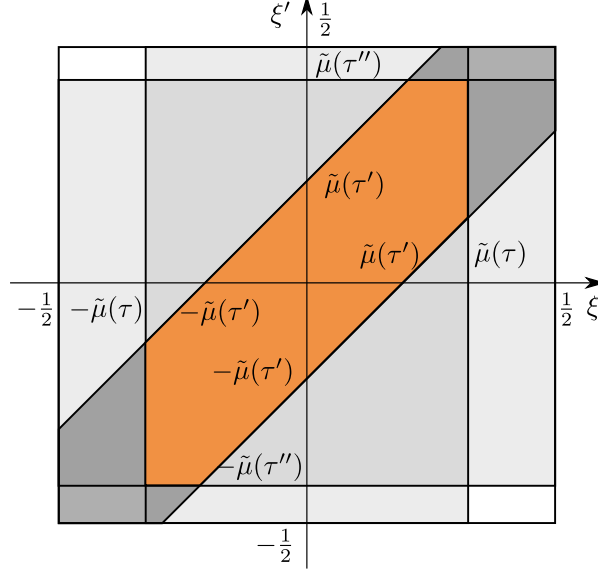


FIGURE 3.5 – Geometrical representation of Eq. (3.29) with $\mathcal{D} = 1$ in the limit $\tilde{\beta} \rightarrow \infty$. The constraints of the domain are illustrated in shaded parallelograms and the overlapping area corresponds to the resulting value of Eq. (3.29). In this illustrative case, $\tilde{\mu}(\tau) > \tilde{\mu}(\tau')$ and $\tilde{\mu}(\tau'') > \tilde{\mu}(\tau')$.

that are connected. For node v_τ born at time τ , this fraction yields

$$c(\tau) = \frac{2N^2\rho^2}{\kappa^2(\tau)} \left[\int_0^\tau \tau' \ell(\tau, \tau', \tau) d\tau' + \tau \int_\tau^1 \ell(\tau, \tau', \tau') d\tau' + \int_\tau^1 \int_\tau^{\tau'} \ell(\tau', \tau', \tau'') d\tau' d\tau'' \right], \quad (3.28)$$

where $\ell(\tau, \tau', \tau'')$ is the probability that three nodes with respective birth times τ , τ' and τ'' form a triangle and is given by

$$\ell(\tau, \tau', \tau'') = \int_{\mathcal{M}} \int_{\mathcal{M}} p(\mathbf{0}, \boldsymbol{\xi}; \tilde{\mu}(\tau), \tilde{\beta}) p(\boldsymbol{\xi}, \boldsymbol{\xi}'; \tilde{\mu}(\tau'), \tilde{\beta}) p(\boldsymbol{\xi}', \mathbf{0}; \tilde{\mu}(\tau''), \tilde{\beta}) d\boldsymbol{\xi} d\boldsymbol{\xi}'. \quad (3.29)$$

Unfortunately, Eq. (3.28) cannot be solved analytically for any $\tilde{\beta}$ and \mathcal{D} . However, the limiting cases consisting of the cold ($\tilde{\beta} \rightarrow \infty$) and hot ($\tilde{\beta} \rightarrow 0$) limit with $\mathcal{D} = 1$ have closed forms for $c(\tau)$.

3.8.1 Cold Limit $\tilde{\beta} \rightarrow \infty$

In this regime, the connection probability $p(\boldsymbol{\xi}, \boldsymbol{\xi}'; \tilde{\mu}(\tau), \tilde{\beta})$ takes the form of a Heaviside step function centered at $\tilde{\mu}(\tau)$, which maximizes the clustering coefficient, $\langle c \rangle$, when $\tilde{\mu}(\tau)$ does not depend on τ [74]. Equation (3.13) then yields

$$\tilde{\mu}(\tau) \simeq n(\tau). \quad (3.30)$$

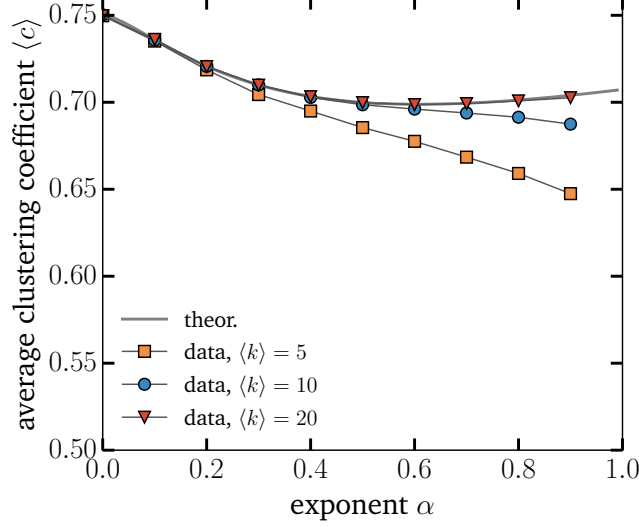


FIGURE 3.6 – Average clustering coefficient $\langle c \rangle$ as a function of α in the cold regime ($\tilde{\beta} \rightarrow \infty$ with Eq. (3.1) as the connection probability). The solid line corresponds to the numerical integration of Eq. (3.33) and the squares to data from Monte-Carlo simulations. We fixed $N = 10^4$ and the results have been averaged over 96 instances.

Additionally, with $\mathcal{D} = 1$, $\ell(\tau, \tau', \tau'')$ can be calculated geometrically from the area of a truncated parallelogram (see Fig. 3.5). It yields

$$\ell(\tau, \tau', \tau'') \simeq \begin{cases} \ell^*(\tau, \tau', \tau''), & \tilde{\mu}_{\max} < \tilde{\mu}_1 + \tilde{\mu}_2 \\ 4\tilde{\mu}_1\tilde{\mu}_2, & \text{otherwise} \end{cases} \quad (3.31)$$

where $\tilde{\mu}_{\max} \equiv \max \{ \tilde{\mu}(\tau), \tilde{\mu}(\tau'), \tilde{\mu}(\tau'') \}$, $\{ \tilde{\mu}_1, \tilde{\mu}_2 \}$ corresponds to the other two values of $\tilde{\mu}(\tau) \neq \tilde{\mu}_{\max}$ and

$$\begin{aligned} \ell^*(\tau, \tau', \tau'') \simeq & 2 \left[\tilde{\mu}(\tau)\tilde{\mu}(\tau') + \tilde{\mu}(\tau')\tilde{\mu}(\tau'') + \tilde{\mu}(\tau'')\tilde{\mu}(\tau) \right] \\ & - \left[\tilde{\mu}^2(\tau) + \tilde{\mu}^2(\tau') + \tilde{\mu}^2(\tau'') \right]. \end{aligned} \quad (3.32)$$

From Eqs. (3.12) and (3.30), we know that $\tilde{\mu}(\tau) = \mathcal{O}(N^{-1})$, which implies that $\ell(\tau, \tau', \tau'') = \mathcal{O}(N^{-2})$ and thus, recalling Eq. (3.28)

$$\langle c \rangle = \int_0^1 c(\tau) d\tau = \mathcal{O}(1). \quad (3.33)$$

Similar scaling arguments will show that $\langle c \rangle$ is also independent of $\langle k \rangle$ in this cold limit as displayed On Fig. 3.6.

Let us examine the effect caused by a specific $\kappa(\tau)$. Consider the case of homogeneous degree sequence $\kappa(\tau) = \langle k \rangle$. Then,

$$\tilde{\mu}(\tau) \simeq \frac{\langle k \rangle}{N}, \quad (3.34)$$

corresponding to the connection threshold for standard RGGs in the 1-dimensional euclidean space and therefore $\langle c \rangle = \frac{3}{4}$ [35, 74]. For a heterogeneous degree sequence such as Eq. (3.14), $\langle c \rangle$ differs from $\frac{3}{4}$ only slightly as seen On Fig. 3.6. Note the remaining discrepancy between the analytical solution and the numerical results for small $\langle k \rangle$ as α is increased. This is due to increasing sampling difficulties in this regime [23] where fluctuations in $\langle k \rangle$ are growing rapidly as $\alpha \rightarrow 1$ ($\gamma = 2$). Figure 3.6 indicates that this difference is indeed reduced considerably with increasing values of $\langle k \rangle$. It can also be avoided altogether with a less constraining degree sequence than the one given by Eq. (3.14).

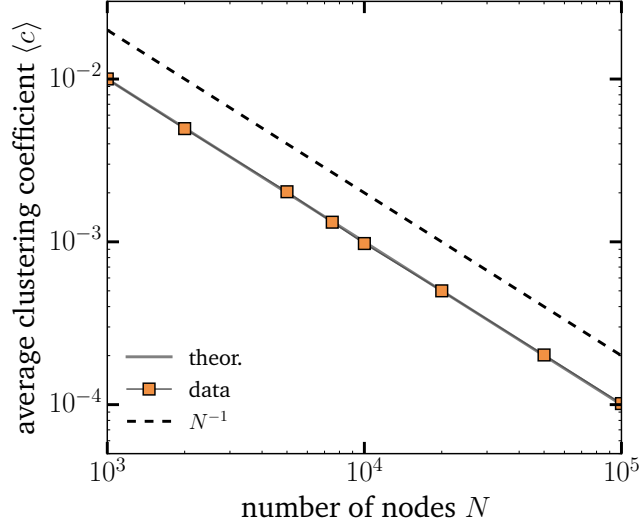


FIGURE 3.7 – Average clustering coefficient $\langle c \rangle$ as a function of the number of nodes N in the hot regime ($\tilde{\beta} \ll 1$). The solid line corresponds to the numerical integration of Eq. (3.38) and the square to data from Monte-Carlo simulations. The scaling of $\langle c \rangle$ is indicated by the black dashed line. We fixed $\alpha = 0$, $\langle k \rangle = 10$ and $\tilde{\beta} = 10^{-3}$. The results have been averaged over 96 instances.

3.8.2 Hot limit $\tilde{\beta} \rightarrow \infty$

We start by evaluating $\tilde{\mu}(\tau)$ $\tilde{\beta} \ll 1$,

$$\tilde{\mu}(\tau) \simeq -\frac{1}{\tilde{\beta}} \ln \left[\frac{1 - n(\tau)}{n(\tau)} \right]. \quad (3.35)$$

which tends to $-\infty$ in the hot limit. Since $\tilde{\mu}(\tau)$ takes on negative values, it cannot be interpreted as a connection threshold anymore. While this may seem counterintuitive, it is a necessary condition to preserve the degree sequence. Now, reinjecting Eq. (3.35) in Eq. (3.2) yields, neglecting $d(\xi, \xi')$ since $d(\xi, \xi') \ll -\tilde{\mu}(\tau)$,

$$p[\xi, \xi'; \tilde{\mu}(\tau), \tilde{\beta}] \simeq \frac{1}{\exp[-\tilde{\beta}\tilde{\mu}(\tau)] + 1} = n(\tau). \quad (3.36)$$

Interestingly, in the hot limit, the connection probability becomes independent of the position ξ and ξ' and is therefore independent of the embedding space \mathcal{M} . Equation (3.29) can then be easily

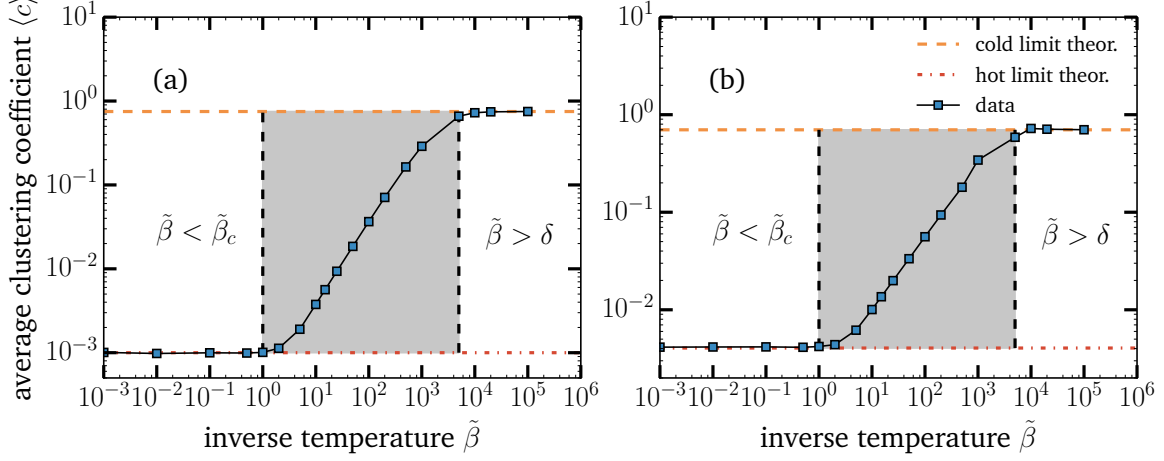


FIGURE 3.8 – Transition between the hot and the cold regime for different degree sequence heterogeneities : (a) homogeneous degree sequence, $\alpha = 0$, $\langle k \rangle = 10$ and $N = 10^4$, (b) heterogeneous degree sequence, $\alpha = 0.5$, $\langle k \rangle = 10$ and $N = 10^4$. The dashed lines (above) correspond to the cold limit theoretical result while the dotted lines (below) correspond to the hot limit one. The data from Monte-Carlo simulations is denoted by the squares. The results have been averaged over 96 instances.

evaluated and we obtain

$$\ell(\tau, \tau', \tau'') \simeq n(\tau)n(\tau')n(\tau''). \quad (3.37)$$

This illustrates that triangle formations are uncorrelated in the hot limit.

Recalling that $n(\tau) = \mathcal{O}(N^{-1})$, Eq. (3.37) informs us that $\ell(\tau, \tau', \tau'') = \mathcal{O}(N^{-3})$, and therefore, together with Eq. (3.28), one is left with

$$\langle c \rangle = \int_0^1 c(\tau) d\tau = \mathcal{O}(N^{-1}). \quad (3.38)$$

This result is validated by Monte-Carlo simulations displayed On Fig. 3.7 for $\kappa(\tau) = \langle k \rangle$.

As the average clustering coefficient vanishes in the thermodynamic limit and the connection probability loses its geometry dependence, the networks have also lost their geometric nature. In fact, this network ensemble is a generalization of Erdős-Renyi random graphs ensemble $\mathcal{G}(N, p)$ where $p \equiv n(\tau)$ is dependent upon the birth time of nodes. For $\kappa(\tau) = \langle k \rangle$, we recover the result $\langle c \rangle = p$ as in the standard $\mathcal{G}(N, p)$.

3.8.3 Phase Transition

Varying $\tilde{\beta}$, we observe a phase transition between the random and geometric phases independently from the heterogeneity of the degree sequence. That is what is shown On Fig. 3.8. Interestingly, the clustering varies between a critical interval $\tilde{\beta} \in [\tilde{\beta}_c, \delta]$, where $\delta \sim \frac{N}{2\pi R} = \frac{N}{2}$ is the density of nodes on the 2-ball, independent of the degree sequence (see shaded region on Fig. 3.8). That $\langle c \rangle$ reaches the cold limit when $\tilde{\beta} > \delta$ is due to the saturation of $\langle c \rangle$.

In the thermodynamic limit, the critical threshold $\tilde{\beta}_c$ is approximately equal to 1, such that $\langle c \rangle = 0$ for $\tilde{\beta} < \tilde{\beta}_c$, whereas $\langle c \rangle$ tends asymptotically to the cold limit as $\delta \rightarrow \infty$.

3.9 Conclusion

In this paper, we have defined a geometric network growth process in which the newly created links attach homogeneously to the existing nodes. A correspondence between our model and a hidden variable framework has allowed us to determine analytical expressions for the most important structural properties, the degree sequence, the degree-degree correlations and the clustering coefficient. Most importantly, we have shown that the parameter $\tilde{\mu}(\tau)$, the network chemical potential characterizing its geometric evolution, can be used as an adjustable function to reproduce just about any degree sequence.

Additionally, we have found that the birth time of nodes in the network, characterized by its network history H , have a strong influence on the form of the degree-degree correlation. This is perhaps one of the more distinctive features of our model, a result that has not been obtained by existing growth processes.

Moreover, we have shown that the other parameter $\tilde{\beta}$, the network inverse temperature, allows to interpolate between a random regime, where the connections are not influenced by geometric constraints, and a geometric regime, where these constraints dominate the connection occurrences. The average clustering coefficient $\langle c \rangle$ varies between two extreme values : the hot limit ($\tilde{\beta} \ll 1$) corresponding to the random phase and the cold limit ($\tilde{\beta} \gg 1$) corresponding to the geometric phase. Notably, the phase transition between the random and geometric phases with a critical threshold $\tilde{\beta}_c \simeq 1$ is similar to the one found in Ref. [75] in hyperbolic geometry.

Some questions remain open, however. First, a problem with certain degree sequences where links are missing when generated by our approach have been identified (see Sec. 3.6). We have discussed that the allowance of multilinks and self-loops would solve the problem, but this implementation is still ongoing.

Second, it is reasonable to ask whether the ensemble of network generated by our model with given degree sequence and equal weights on all histories yields the network ensemble of the configuration model. This correspondence would be important in a number of ways. On the one hand, it would establish an equivalence with the so-called \mathbb{S}^1 model of Ref. [112] and, in turn, with hyperbolic geometry [75] and SPA [98, 68]. On the other hand, the ensemble of networks generated by HA would be a generalization of the configuration model network ensemble where an appropriate distribution on the histories can be chosen to reproduce the degree-degree correlations while preserving the degree sequence.

Third, a further tantalizing question is the possible control of assortativity in the generated networks. While the analysis presented in Sec. 3.7 has shown that the assortativity coefficient is changing

with the type of histories considered, its precise effect on degree-degree correlations has yet to be determined. A better understanding of the role of history on the network structure would allow for a suitable definition of a control parameter of the assortativity coefficient.

In the analysis of the history impact on the structure, one could go even further : inferring an (effective) network history. Although real networks grow or evolve over time according to their own specific dynamics, our model could nevertheless be used to generate random ensembles of surrogates by reconstructing their *effective* growth history. This could be achieved by inferring the model's parameters through, for instance, the maximization of the likelihood that the model has adequately generated the real network structure. This would yield a network ensemble with similar structural properties (degree sequence, correlations, clustering coefficient). Our preliminary work on this aspect has led to promising results, but much remains to be done on this front. We expect to report on this venture in the near future.

3.10 Complement I : Calculation of $n(\tau)$ on the $(\mathcal{D} + 1)$ -sphere

Equation (3.9) cannot be solved exactly for any \mathcal{D} . However, two methods can be used to obtain either an integral form and an approximate closed-form in terms of incomplete Fermi-Dirac integrals for $n(\tau)$. In this section, we investigate both methods to generalize the calculation of $n(\tau)$ to an arbitrary number of dimensions \mathcal{D} .

3.10.1 Exact integral form

The integral of $n(\tau)$ reads, in its explicit form,

$$n(\tau) = \int_{\mathcal{M}} \frac{1}{\exp\left(\tilde{\beta}\left[d(\boldsymbol{\xi}, \boldsymbol{\xi}') - \tilde{\mu}(\tau)\right]\right) + 1} \rho d\boldsymbol{\xi} \quad (3.39)$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_{\mathcal{D}+1})$ and $\boldsymbol{\xi}' = (\xi'_1, \xi'_2, \dots, \xi'_{\mathcal{D}+1})$ are points on the sphere. The symmetry of the sphere helps to solve the problem for two reasons. First, $\boldsymbol{\xi}'$ can be chosen arbitrarily without affecting the final result, and therefore can be fixed to $\boldsymbol{\xi}' = \mathbf{0} = (\tilde{R}, 0, \dots, 0)$ without loss of generality. Second, $(\mathcal{D} + 1)$ -dimensional hyperspherical coordinates centered at the center of the sphere can be chosen to simplify the integral such that $\boldsymbol{\xi} \rightarrow (\tilde{R}, \theta, \varphi_1, \dots, \varphi_{\mathcal{D}-1})$ and $\mathbf{0} \rightarrow (\tilde{R}, 0, \dots, 0)$ where

$$\theta = 2\text{arccot}\left(\frac{\xi_{\mathcal{D}} + \sqrt{\xi_{\mathcal{D}+1} - \xi_{\mathcal{D}}}}{\xi_{\mathcal{D}+1}}\right) \in [0, 2\pi] \quad (3.40)$$

$$\varphi_i = \text{arccot}\left(\frac{\xi_i}{\sqrt{\sum_{j=i+1}^{\mathcal{D}+1} \xi_j^2}}\right) \in [0, \pi], \forall i = 1, 2, \dots, \mathcal{D} - 1 \quad (3.41)$$

and where the element of surface is given by

$$d\boldsymbol{\xi} = \tilde{R}^{\mathcal{D}} d\theta \left[\prod_{i=1}^{\mathcal{D}-1} (\sin \varphi_i)^{\mathcal{D}-i} d\varphi_i \right] \quad (3.42)$$

Additionally, the metric distance between any two points \boldsymbol{x} and \boldsymbol{y} on a sphere is always the length of an arc given by $d(\boldsymbol{x}, \boldsymbol{y}) = \tilde{R}\Delta\sigma$, where $\Delta\sigma$ corresponds to the angle between \boldsymbol{x} and \boldsymbol{y} which can be obtained from the dot product. For $\boldsymbol{\xi}$ and $\mathbf{0}$, we have

$$\begin{aligned} \boldsymbol{\xi} \cdot \mathbf{0} &= \|\boldsymbol{\xi}\| \|\mathbf{0}\| \cos \Delta\sigma = \tilde{R}^2 \cos \Delta\sigma \\ &= \xi_1 \tilde{R} = \tilde{R}^2 \cos \varphi_1 \end{aligned} \quad (3.43)$$

where we recall that, from Eqs. (3.40) and (3.41), $\xi_1 = \tilde{R} \cos \varphi_1$. This simply gives $\Delta\sigma = \varphi_1$ and the metric distance between $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ yields $d(\boldsymbol{\xi}, \mathbf{0}) = \tilde{R}\varphi_1$. Then, the integral yields

$$n(\tau) = \tilde{R}^{\mathcal{D}} \rho \int_0^{2\pi} d\theta \left[\prod_{i=1}^{\mathcal{D}-1} \int_0^{\pi} (\sin \varphi_i)^{\mathcal{D}-i} d\varphi_i \right] \frac{1}{\exp\left\{\tilde{\beta}\left[\tilde{R}\varphi_1 - \tilde{\mu}(\tau)\right]\right\} + 1} \quad (3.44)$$

Since the integrand depends only on φ_1 , there are three types of integrals to solve in order to calculate $n(\tau)$. The first one is trivial

$$\int_0^{2\pi} d\theta = 2\pi \quad (3.45)$$

The second one corresponds to the integrals on φ_i where $i < 1$ and can be computed analytically,

$$\int_0^\pi [\sin(\varphi_i)]^{\mathcal{D}-i} d\varphi_i = \sqrt{\pi} \frac{\Gamma\left(\frac{\mathcal{D}-i+1}{2}\right)}{\Gamma\left(\frac{\mathcal{D}-i+2}{2}\right)} \quad (3.46)$$

Because of its telescoping form, the product is simply given by

$$\prod_{i=2}^{\mathcal{D}-1} \int_0^\pi (\sin \varphi_i)^{\mathcal{D}-i} d\varphi_i = \prod_{i=2}^{\mathcal{D}-1} \sqrt{\pi} \frac{\Gamma\left(\frac{\mathcal{D}-i+1}{2}\right)}{\Gamma\left(\frac{\mathcal{D}-i+2}{2}\right)} = \frac{\pi^{\frac{\mathcal{D}-2}{2}}}{\Gamma\left(\frac{\mathcal{D}}{2}\right)} \quad (3.47)$$

Finally, the last integral is,

$$\int_0^\pi \sin(\varphi)^{\mathcal{D}-1} \frac{d\varphi}{\exp\left\{\tilde{\beta} \left[\tilde{R}\varphi - \tilde{\mu}(\tau)\right]\right\} + 1} = \frac{1}{\tilde{\beta}\tilde{R}} I_{\text{exact}} \quad (3.48)$$

where

$$I_{\text{exact}}(\tau) = \int_0^{\tilde{\beta}} \left[\sin\left(\frac{t}{\tilde{\beta}\tilde{R}}\right) \right]^{\mathcal{D}-1} \frac{dt}{\exp\left[t - \tilde{\beta}\tilde{\mu}(\tau)\right] + 1} \quad (3.49)$$

which cannot be expressed in closed form. The final integral form of $n(\tau)$ is then, replacing $\tilde{R} = \pi^{-1}$ and $\rho = \frac{\Gamma\left(\frac{\mathcal{D}+1}{2}\right)}{2\pi^{-(\mathcal{D}-1)/2}}$

$$n(\tau) = \frac{\pi^{1/2}}{\tilde{\beta}} \frac{\Gamma\left(\frac{\mathcal{D}+1}{2}\right)}{\Gamma\left(\frac{\mathcal{D}}{2}\right)} I_{\text{exact}}(\tau) \quad (3.50)$$

3.10.2 Approximate form

Let us express Eq. (3.9) in terms of a hyperspherical coordinates system centered on node at position ξ that, instead of being in $\mathbb{R}^{\mathcal{D}+1}$, is directly embedded on the surface of $(\mathcal{D} + 1)$ -sphere of radius \tilde{R} . Then, one coordinate of ξ can be written in terms of the other coordinates such that $\xi = \left(\xi_1, \xi_2, \dots, \xi_{\mathcal{D}}, \sqrt{\tilde{R}^2 - \sum_{i=1}^{\mathcal{D}} \xi_i^2}\right)$ and the hyperspherical coordinate system is effectively of dimension \mathcal{D} and expressed, similarly as before,

$$r = \left[\sum_{i=1}^{\mathcal{D}} \xi_i^2 \right] \in [0, 2\tilde{R}] \quad (3.51)$$

$$\theta = 2\text{arccot} \left(\frac{\xi_{\mathcal{D}-1} + \sqrt{\xi_{\mathcal{D}} - \xi_{\mathcal{D}-1}}}{\xi_{\mathcal{D}}} \right) \in [0, 2\pi] \quad (3.52)$$

$$\varphi_i = \text{arccot} \left(\frac{\xi_i}{\sum_{j=i}^{\mathcal{D}} \xi_j^2} \right) \in [0, \pi] \forall i = 1, 2, \dots, \mathcal{D} - 2 \quad (3.53)$$

The element of volume is then given by,

$$d\xi = r^{\mathcal{D}-1} dr d\theta \left[\prod_{i=1}^{\mathcal{D}-2} (\sin \varphi_i)^{\mathcal{D}-1-i} d\varphi_i \right] \quad (3.54)$$

Equation (3.9) can then be rewritten as the following expression

$$n(\tau) = \rho \int_0^{2\pi} d\theta \left\{ \prod_{i=1}^{\mathcal{D}-2} \int_0^\pi [\sin(\varphi_i)]^{\mathcal{D}-1-i} d\varphi_i \right\} \int_0^{\pi\tilde{R}} \frac{r^{\mathcal{D}-1} dr}{\exp\left(\tilde{\beta} \left[d(\xi, \mathbf{0}) - \tilde{\mu}(\tau) \right] \right) + 1} \quad (3.55)$$

In the regime where $\tilde{\mu}(\tau) \ll 1$ and $\tilde{\beta} \gg 1$, we can approximate $r \simeq d(\xi, \mathbf{0})$. This is similar to say that the contributing terms in the integral corresponds to the points on the sphere close to ξ' . This enables us to use the locally Euclidean property of Riemannian manifolds to estimate the distance which, instead of being the length of an arc, to is the distance on the tangent \mathcal{D} -plane of ξ' . Then, each integral can be evaluated separately. The integral on θ simply gives 2π and the integrals on φ_i yields

$$\int_0^\pi [\sin(\varphi_i)]^{\mathcal{D}-1-i} d\varphi_i = \sqrt{\pi} \frac{\Gamma\left(\frac{\mathcal{D}-i}{2}\right)}{\Gamma\left(\frac{\mathcal{D}+1-i}{2}\right)} \quad (3.56)$$

The product can also be evaluated, since it has a telescoping form

$$\prod_{i=1}^{\mathcal{D}-2} \left[\sqrt{\pi} \frac{\Gamma\left(\frac{\mathcal{D}-i}{2}\right)}{\Gamma\left(\frac{\mathcal{D}+1-i}{2}\right)} \right] = \frac{\pi^{(\mathcal{D}-2)/2}}{\Gamma\left(\frac{\mathcal{D}}{2}\right)} \quad (3.57)$$

Finally, the approximated integral on r corresponds to an incomplete form of the Fermi-Dirac integral, with $\pi\tilde{R} = 1$

$$\int_0^1 \frac{r^{\mathcal{D}-1} dr}{\exp\left(\tilde{\beta} \left[r - \tilde{\mu}(\tau) \right] \right) + 1} = \frac{1}{\tilde{\beta}^{\mathcal{D}}} I_{\text{approx}}(\tau) \quad (3.58)$$

where

$$I_{\text{approx}}(\tau) = \int_0^{\tilde{\beta}} \frac{t^{\mathcal{D}-1} dt}{e^{t-\tilde{\beta}\tilde{\mu}(\tau)} + 1} = \Gamma(\mathcal{D}) f_{\mathcal{D}-1}\left(\tilde{\beta}\tilde{\mu}(\tau), \tilde{\beta}\right) \quad (3.59)$$

with $f_j(x, b)$ being the lower incomplete Fermi-Dirac integral. Putting together the results of each integral, we get, using $\rho = \frac{\Gamma\left(\frac{\mathcal{D}+1}{2}\right)}{2\pi^{-(\mathcal{D}-1)/2}}$,

$$n(\tau) \simeq \frac{\pi^{\mathcal{D}-1/2}}{\tilde{\beta}^{\mathcal{D}}} \frac{\Gamma\left(\frac{\mathcal{D}+1}{2}\right)}{\Gamma\left(\frac{\mathcal{D}}{2}\right)} I_{\text{approx}}(\tau) \quad (3.60)$$

3.10.3 Correspondence with existing solutions

We recall that, in the limit $\tilde{\beta} \rightarrow \infty$, the connection probability has the profile of a step function. Therefore, it delineates an area in which the connection occurs with probability equal to one with any node at distance less than $\tilde{\mu}(\tau)$ from any source node. The approximated value of $n(\tau)$ will therefore be proportional to the volume of a \mathcal{D} -sphere of radius $\tilde{\mu}(\tau)$ corresponding to the excluded

volume [35] (see Sec. 2.3 for additional details). In order to validate this previously obtained result, we investigate different values of \mathcal{D} and examine this correspondence. In general, when $\tilde{\beta} \rightarrow \infty$, the approximated integral $I_{\text{approx}}(\tau)$ simply yields

$$\lim_{\tilde{\beta} \rightarrow \infty} I_{\text{exact}}(\tau) = \tilde{\beta}^{\mathcal{D}} \frac{\tilde{\mu}(\tau)^{\mathcal{D}}}{\mathcal{D}} \quad (3.61)$$

and the fraction of created links gives

$$n(\tau) \simeq \frac{\pi^{\mathcal{D}-1/2} \Gamma\left(\frac{\mathcal{D}+1}{2}\right)}{\mathcal{D} \Gamma\left(\frac{\mathcal{D}}{2}\right)} \tilde{\mu}^{\mathcal{D}}(\tau) \quad (3.62)$$

For $\mathcal{D} = 1$, we find the result given by Eq. (3.30) which is $n(\tau) \simeq \tilde{\mu}(\tau)$ which was expected. Also, given that the surface of this 2-sphere is $S_1(\tilde{R}) = 2\pi\tilde{R} = 2$, we find

$$n(\tau) \simeq \frac{2\tilde{\mu}(\tau)}{2} = \frac{V_1(\tilde{\mu}(\tau))}{S_1(\tilde{R})} \propto V_1(\tilde{\mu}(\tau)) \quad (3.63)$$

where $V_1(\tilde{\mu}(\tau))$ is the volume of a 1-sphere. In general, we can write from Eq. (3.62)

$$n(\tau) = \frac{V_{\mathcal{D}}(\tilde{\mu}(\tau))}{S_{\mathcal{D}}(\tilde{R})} \quad (3.64)$$

where

$$V_{\mathcal{D}}(\tilde{\mu}(\tau)) = \frac{\pi^{\mathcal{D}/2}}{\Gamma(\mathcal{D}/2 + 1)} \tilde{\mu}(\tau)^{\mathcal{D}}, \quad S_{\mathcal{D}}(\tilde{R}) = \frac{2\pi^{(\mathcal{D}+1)/2}}{\Gamma((\mathcal{D}+1)/2)} \tilde{R}^{\mathcal{D}} \quad (3.65)$$

3.10.4 Linking the exact and approximated solutions

The two expressions for $n(\tau)$, Eq. (3.50)-(3.60), have been obtained through two conceptually different, yet quite technically similar, paths. To complete the portrait of this analysis, it is therefore necessary to determine if these two expressions are linked through the assumptions made, i.e. $\tilde{\mu}(\tau) \ll 1$ and $\tilde{\beta} \gg 1$.

Let us examine the behavior of $I_{\text{exact}}(\tau)$ in this regime. First, for $\tilde{\beta}$ relatively large, the majority of the contributions of the integral are located around $t = 0$. Second, the term in the sine function will necessarily tend to zero as $\tilde{\beta}$ is large. Therefore, we can develop the sine function around $t = 0$ and keep only the first term

$$I_{\text{exact}}(\tau) \simeq \int_0^{\tilde{\beta}} \left(\frac{\pi t}{\tilde{\beta}}\right)^{\mathcal{D}-1} \frac{dt}{\exp\left[t - \tilde{\beta}\tilde{\mu}(\tau)\right] + 1} = \left(\frac{\pi}{\tilde{\beta}}\right)^{\mathcal{D}-1} I_{\text{approx}}(\tau) \quad (3.66)$$

Replacing the approximated expression of $I_{\text{exact}}(\tau)$ in Eq. (3.50), we obtain the same expression as Eq. (3.60).

3.10.5 Phase transition and dimensionality

An observation that one might make is that in the random phase changing \mathcal{D} should not affect any structural properties.

In the random phase, the structural properties of the networks should not depend on \mathcal{D} since the space dependence on the network structure is negligible. Therefore, we expect $n(\tau)$ to be invariant of \mathcal{D} when $\tilde{\beta} < \tilde{\beta}_c$. This is confirmed on Fig. 3.9.

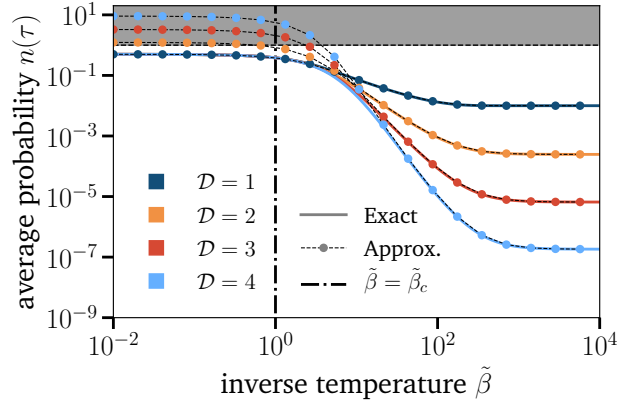


FIGURE 3.9 – Average probability $n(\tau)$ as a function of the inverse temperature $\tilde{\beta}$. We fixed $\tilde{\mu}(\tau) = 10^{-2}$. The lines correspond to the exact expression Eq. (3.50) while the dashed lines correspond to the approximated solution Eq. (3.60). The dotted line corresponds to the transition threshold $\tilde{\beta} = \tilde{\beta}_c$ and the shadowed area corresponds to the non-physical case $n(\tau) > 1$.

This observation suggests that the phase transition could be investigated through the dependence of $n(\tau)$ with \mathcal{D} at $\tilde{\beta} \rightarrow \tilde{\beta}_c$. However, because $n(\tau)$ is independent of N for fixed $\tilde{\mu}(\tau)$, it might not be an appropriate order parameter to characterize this phase transition. Indeed, as $N \rightarrow \infty$ and $\tilde{\beta} \rightarrow \tilde{\beta}_c$, the discrepancy between each curves on Fig. 3.9 does not tend to 0 as it should for standard phase transitions.

3.11 Complement II : On the finite size effects

In Sec. 3.6, it was established that, for a degree sequence such as Eq. (3.14), all nodes born at any time $\tau < \tau^*$, where

$$\tau^* = \left[\frac{\alpha\nu}{N(\alpha+1) - \nu} \right]^{\frac{1}{\alpha+1}}, \quad (3.67)$$

the average connexion probability $n(\tau) = 1$. First, this implies that these older nodes will form a τ^* -core. Second, because we impose that $n(\tau) \leq 1$ for all τ , the degree of these nodes will be less than we expected in the first place. Since all nodes belonging to the τ^* -core are all statistically equivalent, the structural properties as functions of τ will always be constant for $\tau < \tau^*$ forming a plateau. Hence, we will refer to τ^* as the *plateau size*. In Sec. 3.6.4, these alterations to the degree

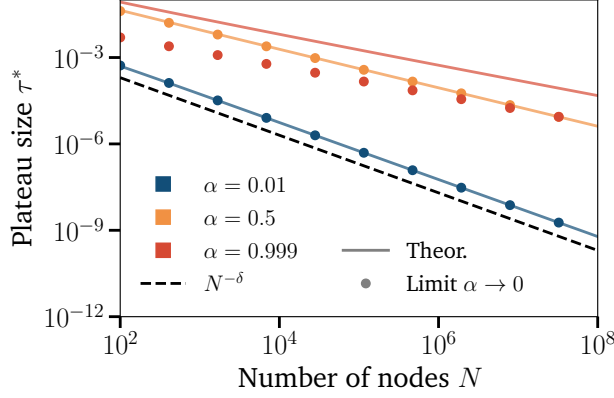


FIGURE 3.10 – Plateau size τ^* scaling with N in the fully-homogeneous limit : The solid lines correspond to the numerical solution of Eq. (3.69) and the circles correspond to the asymptotic solution Eq. (3.70) in limit $\alpha \rightarrow 0$. The scaling of $N^{-\delta}$ is shown as a reference with the dashed line where $\delta = 1$.

sequence were declared to be *finite size effects*, but no explicit demonstration of this was shown. In this Complement, we adress this topic.

First, let us determine the equation to calculate τ^* as a function of α , $\langle k \rangle$ and N . We begin by calculating ν from Eq. (3.17)

$$\nu = \frac{\langle k \rangle (1 - \alpha)}{1 - \alpha \tau^{*1-\alpha}}. \quad (3.68)$$

Introducing this expression into Eq. (3.67), we obtain the implicit expression

$$\tau^* = \left(\frac{\alpha \lambda}{1 + \lambda - \alpha \tau^{*1-\alpha}} \right)^{\frac{1}{1+\alpha}} \Rightarrow \alpha \tau^{*2} - (1 + \lambda) \tau^{*1+\alpha} + \alpha \lambda = 0 \quad (3.69)$$

where $\lambda \equiv \frac{\langle k \rangle (1-\alpha)}{N(1+\alpha)}$. In order to obtain τ^* , one must in general solve Eq. (3.69) numerically for arbitrary α , $\langle k \rangle$ and N . However, to determine if the plateau is a finite-size effect, a solution of Eq. (3.69) can be determined for τ^* as a function of N , in the limits $\alpha \rightarrow 0$, the fully-homogenous limit, and $\alpha \rightarrow 1$, the fully-heterogeneous limit.

3.11.1 Plateau size in the *fully-homogeneous* limit

Letting $\alpha = 0$ in Eq. (3.69) yields the solution $\tau^* = 0$. However, we are interested in an asymptotic solution in function of N . Thus, we consider instead Eq. (3.67) assuming that, as $\alpha \rightarrow 0$, $\nu \simeq \langle k \rangle (1 - \alpha)$, $\tau^* \ll 1$ and $N \gg \nu$

$$\tau^* \simeq \left(\alpha \frac{(1 - \alpha) \langle k \rangle}{(1 + \alpha) N} \right)^{\frac{1}{1+\alpha}} = \mathcal{O}(N^{-\delta}). \quad (3.70)$$

The plateau will therefore shrink as N increases, as shown on Fig. 3.10, demonstrating that, in the fully-homogeneous limit $\alpha \rightarrow 0$, this modification to the degree sequence is truly a finite size effect with scaling exponent $\delta \rightarrow 1$.

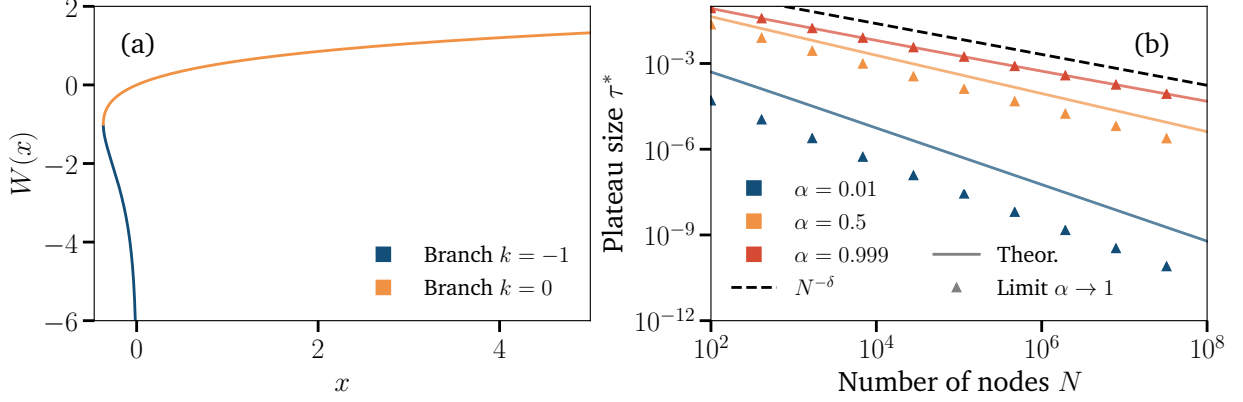


FIGURE 3.11 – Plateau size τ^* scaling with N in the fully-heterogeneous limit : (a) Lambert W-function with the two real branches $k = -1$ and $k = 0$ [96], (b) τ^* as a function of N for different values of $\epsilon \ll 1$. Similarly to Fig. 3.10, the solid lines correspond to the numerical solution of Eq. (3.69) and the triangles correspond to the asymptotic Eq. (3.72) in the limit $\alpha \rightarrow 1$. The scaling with N , as indicated by the dashed line, has an exponent $\delta \approx 0.54 > \frac{1}{2}$.

3.11.2 Plateau size in the *fully-heterogeneous* limit

Contrary to the fully-homogeneous case, letting $\alpha = 1$ in Eq. (3.69) does not lead to any solution. This is in line with the fact that the model is not well-behaved in this limit. Instead, we have to consider that $\alpha \rightarrow 1$. Developing Eq. (3.68) around $\alpha = 1$ and keeping only the zero-th order term yields

$$\nu \simeq \frac{\langle k \rangle}{1 - \ln \tau^*}. \quad (3.71)$$

With this approximate expression of ν , Eq. (3.69) can be solved using Lambert W-function $W(x)$, defined as the inverse function of $x = W(x)e^{W(x)}$ [96]. Starting with Eq. (3.67) with ν expressed in terms of Eq. (3.71) and letting $\varphi \equiv \frac{\langle k \rangle}{N}$, we obtain

$$\begin{aligned} \tau^* &\simeq \left[\frac{\alpha \langle k \rangle}{N(1 + \alpha)(1 - \ln \tau^*) - \langle k \rangle} \right]^{\frac{1}{1 + \alpha}} \\ \tau^{*1 + \alpha} \left[(1 + \alpha)(1 - \ln \tau^*) - \varphi \right] &\simeq \alpha \varphi \\ (1 + \alpha)(1 - \ln \tau^*) - \varphi &\simeq \alpha \varphi \tau^{*-1 - \alpha} \\ \ln \left(\tau^{*-1 - \alpha} \right) + [\alpha - \varphi + 1] &\simeq \alpha \varphi \tau^{*-1 - \alpha} \\ \ln \left(\tau^{*-1 - \alpha} \right) - \alpha \varphi \tau^{*-1 - \alpha} &\simeq -[\alpha - \varphi + 1]. \end{aligned}$$

Then, by adding $\ln(-\alpha\varphi)$ on both sides leads to

$$\ln \left(-\alpha \varphi \tau^{*-1 - \alpha} \right) - \alpha \varphi \tau^{*-1 - \alpha} \simeq -[\alpha - \varphi + 1] + \ln(-\alpha\varphi)$$

which, letting $\Omega = -\alpha\varphi\tau^{*-1 - \alpha}$ and applying the exponential on both sides, yields

$$\Omega e^{\Omega} = -\alpha\varphi e^{-[\alpha - \varphi + 1]} \Rightarrow \Omega = -\alpha\varphi\tau^{*-1 - \alpha} = W \left(-\alpha\varphi e^{-[\alpha - \varphi + 1]} \right)$$

This expression has a closed form in terms of $W(x)$,

$$\tau^* \simeq \left[\frac{-\alpha\varphi}{W(-\alpha\varphi e^{-[\alpha-\varphi+1]})} \right]^{\frac{1}{1+\alpha}} = \mathcal{O}(N^{-\delta}) \quad (3.72)$$

where one must choose the branch $k = -1$ in the W-function (see Fig. 3.11(b)) in order for τ^* to converge on a real value. We find that, according to numerical observations, the scaling exponent is well approximated by $\delta \approx 0.54 > \frac{1}{2}$ as shown on Fig. 3.11(a). Although this is surely not the end of the asymptotic analysis, we can however safely say that the plateau is also a finite size effect in the fully-heterogeneous limit.

3.12 Complement III : Evaluation of the parameters

In order to generate scale-free geometric networks with the HA model, we have access to the set of parameters $\Theta_{\text{HA}} = \{ \alpha, \nu, N, \tilde{\beta}, H \}$, where α is the exponent of the degree sequence, ν is a constant fixing the average degree $\langle k \rangle$, N is the number of nodes, $\tilde{\beta}$ is controlling the clustering coefficient and H is the history. Fixing these parameters appropriately can generate scale-free networks with given degree distribution exponent γ , average degree $\langle k \rangle$, and, to a certain extent, with given average clustering coefficient $\langle c \rangle$ and level of assortativity. While some of these parameters are easily fixed, such as $\alpha = \frac{1}{\gamma-1}$ and N , some others, such as $\tilde{\beta}$, H and ν , are more delicate to evaluate.

First, $\langle c \rangle$ can be fixed using an appropriate $\tilde{\beta}$, but it implies finding the solution of a $(2\mathcal{D}+3)$ -dimensional integral² in Eq. (3.28). The issue is that this $(2\mathcal{D}+3)$ -dimensional integral does not have a closed-form and can only be evaluated numerically, for instance, using Monte-Carlo integration [107]. Second, while it is true that H can be fixed to generate a given degree assortativity coefficient r , it remains, at least for now, a difficult combinatorial optimization problem to solve (See. Sec. 3.13). The easiest parameter to evaluate in this subset of difficult parameters is ν . One can use it to fix $\langle k \rangle$, provided that α , N and H are given, with an appropriate numerical procedure. For these reasons, we focus only on this problem of finding ν in this Complement.

The naive approach to evaluate ν would be to calculate $\langle k \rangle$ analytically as a function of ν and to solve this expression for ν letting $\langle k \rangle = \bar{k}$, where \bar{k} is an additional parameter corresponding to the desired value of average degree. In particular, we have

$$\langle k \rangle = \bar{k} = \frac{\nu}{1-\alpha} \quad \Rightarrow \quad \nu = \bar{k}(1-\alpha) \quad (3.73)$$

However, as we have determined in Sec. 3.6.4, some finite size effects are involved and alters the degree sequence and, consequently, $\langle k \rangle$. It is true that, for a decreasing order of degrees history, one could take explicitly these finite size effects into account in the calculation of $\langle k \rangle$. Unfortunately, this procedure cannot be used for general H . The reason is that, when we consider a scale-free degree sequence with a history H , we begin by generating a degree sequence given by Eq. (3.14) and then

2. In total, there are 3 integrals for τ , τ' and τ'' and $2\mathcal{D}$ integrals for ξ and ξ' to solve for $\langle c \rangle$.

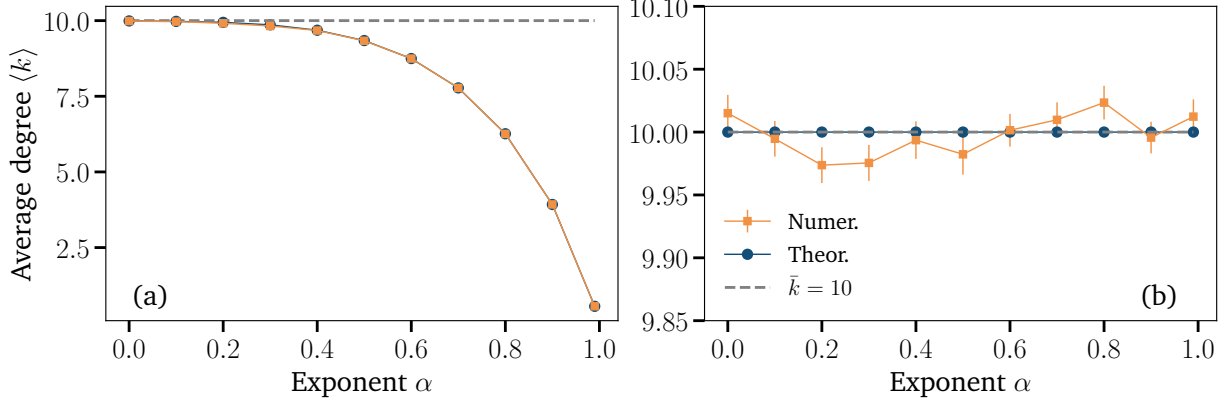


FIGURE 3.12 – Average degree $\langle k \rangle$ as a function of α : (a) we fixed $\nu = \bar{k}(1 - \alpha)$, (b) we used the numerical procedure elaborated in Sec. 3.12 using $\epsilon = 10^{-6}$. On both (a) and (b), the circles corresponds to the expected average degree calculated from the analytical modified degree sequence and the squares corresponds to the average degree of generated networks with their respective error bars given by the standard deviations. The error bars on (a) are present but too small to appear. The dashed line corresponds to the desired value of $\langle k \rangle = \bar{k}$. For the generated networks, we used $\bar{k} = 10$, $N = 10^3$ and $\tilde{\beta} \rightarrow \infty$. The results are averaged over 96 instances.

we rearrange it to match H . Then, calculating the appropriate $\tilde{\mu}(\tau)$ using Eq. (3.12) to obtain $n(\tau)$ and then Eq. (3.13) to get $\tilde{\mu}(\tau)$, we have to impose that $0 \leq n(\tau) \leq 1$. This operation is highly non linear and is therefore untractable for general H . We therefore need to evaluate ν in order to be able to fix $\langle k \rangle = \bar{k}$, without an explicit knowledge of these non linear operations.

We define an algorithm in which ν is evaluated iteratively. The algorithm goes as follows :

1. At iteration $i = 0$, we generate a scale-free degree sequence $\kappa_0(\tau) = \nu_0 \tau^{-\alpha}$, where the initial $\nu_0 = \bar{k}(1 - \alpha)$. The initial degree sequence is then rearranged to the appropriate history H .
2. At iteration $i > 0$, the degree sequence $\kappa_i(\tau)$ is injected in Eq. (3.12) to obtain $n_i(\tau)$.
3. A non linear operation is applied, namely

$$n_{i+1}(\tau) = \begin{cases} 0 & \text{if } n_i(\tau) < 0 \\ 1 & \text{if } n_i(\tau) > 1 \\ n_i(\tau) & \text{otherwise} \end{cases} \quad (3.74)$$

for all τ .

4. A modified degree sequence $\tilde{\kappa}_i(\tau)$ is calculated with Eq. (3.8) using $n_{i+1}(\tau)$.
5. The parameter ν_{i+1} is updated using

$$\nu_{i+1} = \frac{\bar{k}}{\langle k \rangle_i} \nu_i \quad (3.75)$$

where

$$\langle k \rangle_i = \int_0^1 \tilde{\kappa}_i(\tau) d\tau \quad (3.76)$$

6. A new degree sequence $\kappa_{i+1}(\tau)$ is calculated with

$$\kappa_{i+1}(\tau) = \frac{\nu_{i+1}}{\nu_i} \kappa_i(\tau) \quad (3.77)$$

7. Steps 2-6 are repeated until $|\langle k \rangle - \langle k \rangle_i| \leq \epsilon$, where ϵ is some error tolerance.

The idea is to generate a mapping $f(\nu_i) = \nu_{i+1}$ such that the appropriate value of ν matching $\langle k \rangle$ is its only attractive fixed point. Interestingly, this procedure can also be applied to any form of degree sequence, because it is always possible to define a multiplicative parameter ν in control of the scale of $\kappa(\tau)$ independently of its form. On Fig. 3.12, we show the results for $\langle k \rangle$ as a function of α using the naive method and the iterative procedure. The naive method fails completely to obtain the desired $\langle k \rangle$ as α tends to 1, whereas the iterative procedure is able to evaluate ν with a high level of accuracy.

3.13 Complement IV : Proof of concept of the geometric evolution inference

Although real networks evolve over time according to their own specific dynamics, the HA model can be used to generate random ensembles of surrogates by reconstructing their *effective* growth history. This is achieved by inferring the model's parameters by maximizing the likelihood that the model generated the real network structure, which then yields a network ensemble with similar structural properties like the degree sequence, the correlations and the clustering coefficient.

Given a network structure characterized by the adjacency matrix \mathbf{A} , the ensemble of surrogates is defined by the function $\tilde{\mu}(\tau)$ obtained from the history H and the parameter $\tilde{\beta}$ both inferred from a maximum likelihood estimation algorithm. In principle, to obtain H correctly, we would have to infer the model parameters entirely : that is, inferring $H = (\tau_i)$ and $X = \{\xi_i\}$, where X is the set of all the nodes position. However, nodes are distributed homogeneously and therefore the inference of H and X can be decoupled. Since the inference of X has already been investigated in Refs. [19, 98, 97, 99], we will focus on the inference of H .

3.13.1 Maximum likelihood estimation algorithm

The likelihood $\mathbb{P}_{\text{HA}}(G|H)$ of generating a network G with history H is

$$\mathbb{P}_{\text{HA}}(G|H) = \prod_{i=1}^N n(\tau_i)^{2[\kappa^*(\tau_i)]} [1 - n(\tau_i)]^{2[N(\tau_i) - \kappa^*(\tau_i)]}, \quad (3.78)$$

where $\kappa^*(\tau)$ is the expected degree at birth of node v_τ born at time τ and $N(\tau)$ is the number of nodes existing at time τ . Note that the fraction $n(\tau)$ of links created by node v_τ can be rewritten as

$$n(\tau) = \frac{\kappa^*(\tau)}{N(\tau)}. \quad (3.79)$$

To infer H , we use the following Metropolis-Hasting algorithm :

1. We consider an initial history $H(0) = (\tau_i(0))$, where $\tau_i(m) \in (0, 1]$ is the birth time of node v_i at step m .
2. At step m , node v_i is randomly selected and is reassigned a new birth time such that $\tau_i(m - 1) \rightarrow \tau_i(m)$, where $\tau_i(m)$ is distributed uniformly between 0 and 1.
3. The step is accepted with probability

$$q = \min \left\{ \frac{\mathbb{P}_{\text{HA}}(G|H(m))}{\mathbb{P}_{\text{HA}}(G|H(m-1))}, 1 \right\} \quad (3.80)$$

and rejected with probability $1 - q$.

4. Steps 2 and 3 are repeated until $m = M$.

5. H_{infer} and $\tilde{\mu}_{\text{infer}}(\tau)$ are calculated from the average over all the set of samples $\{H(m)\}$ [125]. Averaging over all $H(m)$ guarantees that the inferred $\tilde{\mu}_{\text{infer}}(\tau)$ will not be overfitted. Additionally, since the $H(m)$ is chosen according to their likelihood, we expect $\tilde{\mu}_{\text{infer}}(\tau)$ to be a good estimator of the maximum likelihood configuration in the limit $M \rightarrow \infty$.

3.13.2 Network history inference of synthetic networks

To calibrate the algorithm and examine its validity, we consider the case of scale-free networks. While networks with a homogeneous degree sequence could also be inferred, the choice of a scale-free degree sequence is more suited for a proof of concept of the method. First, in a homogeneous degree sequence, each degree is well represented by the average degree and therefore the birth-time should not be as meaningful. Second, it is harder for homogeneous networks to have assortative (disassortative) behavior because of the lack of hubs.

Degree ordered history

The first case we investigate is that of degree ordered history $\kappa(\tau) = \nu\tau^{-\alpha}$. In these configurations, the traces left by the growth process are significant and well represented by a strong assortativity. To have a better understanding of the inference of this type of network, we considered two different initial conditions for the MLE algorithm : (a) $H(0)$ is randomly chosen and (b) $H(0)$ corresponds to the answer we are looking for, denoted the *ground truth*. On the one hand, considering a random $H(0)$ is similar to saying that we know nothing of the ground truth we seek. It should then be an effective test of the MLE algorithm if it shows that, from anywhere in the history space, the inferred history is correlated with the ground truth. On the other hand, considering that $H(0)$ corresponds to the ground truth, it will show how well the ground truth is representing the actual structure of the network. Because the network properties are fixed on average, we expect it to be slightly off the ground truth. The discrepancy between the inferred history and the ground truth will illustrate how well the MLE algorithm can actually do.

On Fig. 3.13 (a)-(b), we show the results of the algorithm for synthetic networks. First, we can see that the two initial conditions yield approximately the same correlation between H and $\tilde{\mu}(\tau)$. Second,

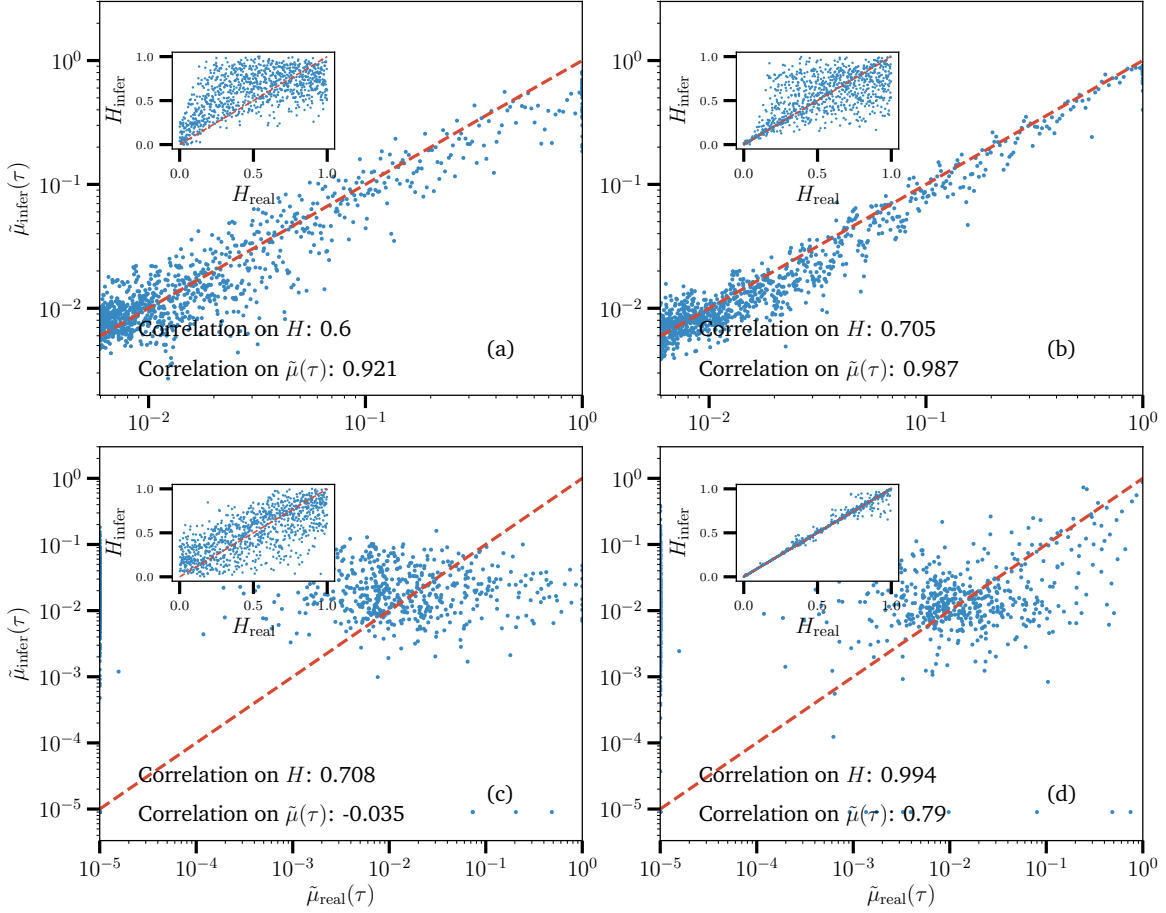


FIGURE 3.13 – Inference of synthetic scale-free geometric networks : (Top) decreasing order of degrees history, (Bottom) random order history. On the left panels, we considered the initialization $H(0)$ in which $\tau_i(0) \sim U(0, 1)$ for all i , where $U(0, 1)$ is a uniform distribution defined between 0 and 1. The right panels show the inference when $H(0)$ is the ground truth. The dashed line is a reference to the target $\tilde{\mu}_{\text{infer}}(\tau) = \tilde{\mu}_{\text{real}}(\tau)$. In the insets is displayed the inferred birth time in H_{infer} as a function of the ground truth birth time in H_{real} . The correlation coefficient for H and $\tilde{\mu}(\tau)$ corresponds to the Pearson coefficient between the inferred and ground truth datasets. The networks are generated with $\alpha = 0.83$, $\gamma = 2.2$, $\langle k \rangle = 35$, $N = 10^2$ and $\beta = 10^5$. The inference data (H_{infer} and $\tilde{\mu}(\tau)_{\text{infer}}$) has been averaged over $M = 200$ histories.

notice how, while the correlation between H_{infer} and H_{real} is poorer, $\tilde{\mu}_{\text{infer}}(\tau)$ and $\tilde{\mu}_{\text{real}}(\tau)$ are highly correlated. This is due to the fact that the nodes for which the birth-time has been falsely inferred are actually the low degree ones which barely affect the likelihood. The important part is that the birth-time of the hubs is well captured, because they have the most contribution in the likelihood. The high correlation between $\tilde{\mu}_{\text{infer}}(\tau)$ and $\tilde{\mu}_{\text{real}}(\tau)$ demonstrates the validity of the algorithm for this type of network.

Random history

Random history is the second case we examine. As we have seen in Sec. 3.7, in these configurations, $\kappa(\tau) \sim P(k)$ where $P(k) \propto k^{-\gamma}$. On Fig. 3.13, the results of the inference for these synthetic networks are shown.

Interestingly, in comparison with the decreasing degree history inference, we obtain different correlations in the inference. First, the knowledge of the ground truth seems to have a significant impact on the inference as H_{infer} and H_{real} have an almost perfect match. The likelihood for the optimal history must be extremely peaked and difficult to maximize with the algorithm. Additionally, even if H_{infer} and H_{real} are almost perfectly correlated, the inference of $\tilde{\mu}(\tau)$ yields a poor result. This is due to the fact that the variance on the degree sequence is high for random history geometric network.

3.14 Complement V : Solution de la distribution des degrés par équations compartimentales

Le modèle d'évolution géométrique présenté au Chapitre 3 permet d'utiliser le formalisme de réseaux corrélés par variables cachées [21] pour calculer analytiquement certaines propriétés structurales, comme la distribution des degrés, la corrélation des degrés ou le coefficient d'agrégation. Bien que fort pratique, cette méthode ne peut toutefois pas être utilisée pour tous les modèles de réseaux aléatoires. Un exemple de modèle n'ayant pas de représentation en termes de variables cachées est le modèle de BA dans sa forme conventionnelle³.

À la Section 2.4, nous avons vu une méthode différente pour calculer certaines propriétés structurales dont la distribution des degrés : l'approche par équations compartimentales. Cette approche est généralement utilisée dans les modèles de croissance de type PA [11, 46, 62, 63, 65, 73, 124] ainsi que pour l'étude des dynamiques de propagation d'épidémie [58, 82, 103] où l'on étudie l'évolution de la santé d'une population à l'aide de différents *compartiments*⁴. Dans le cas qui nous intéresse,

3. Tel que mentionné à la Section 2.4, il est possible de modifier le modèle de BA de sorte que les nouveaux noeuds ne se connectent pas aux plus vieux proportionnellement à leur degré, mais à leur degré espéré. De cette manière, on peut définir une variable cachée r liée au degré espéré afin d'obtenir une telle représentation du MBA (voir la Sec. 2.4). On nomme ce modèle le modèle souple d'attachement préférentiel (*soft preferential attachment, SPA*).

4. Un exemple en épidémiologie de compartimentalisation serait de définir un compartiment pour les individus en santé mais *susceptibles*, noté S , un autre pour ceux *infectés* par la maladie, noté I , et un dernier ceux qui se sont *rétablis* de cette maladie, noté R . On par alors d'une dynamique épidémiologique SIR [12].

un compartiment correspond à une classe de noeuds possédant une propriété commune, le degré k d'un noeud, et les équations compartimentales décrivent l'évolution de la taille moyenne de chaque compartiment.

Étant une approche particulièrement standard pour l'étude des modèles de croissance, il est donc naturel de vouloir l'appliquer à au modèle d'évolution géométrique pour en calculer la distribution des degrés. Il est d'ailleurs digne de mentionner que cette approche fut employée avant d'utiliser la représentation en termes de variables cachées du modèle HA à l'aide du MBPS. Cependant, il s'est avéré qu'elle était plus limitée que l'approche par le MBPS. Les résultats de cette analyse ont toutefois aidé à mieux comprendre ce mécanisme et c'est pourquoi, dans cette annexe, nous présentons la méthode par équations compartimentales appliquée au modèle HA d'évolution géométrique.

3.14.1 Équations compartimentales

Dans le modèle d'évolution géométrique, les noeuds arrivent un à un à chaque temps. De plus, comme montré au Chapitre 3, il est possible de définir, pour un potentiel chimique donné $\tilde{\mu}(\tau)$, une probabilité de connexion moyenne $n(\tau)$ indépendante de la géométrie selon l'Éq. (3.9). Il est important de mentionner que la notation en termes du temps normalisé τ ne simplifie par le développement des équations compartimentales. C'est pourquoi nous réutiliserons désormais le temps d'évènements t et on considèrera, pour alléger la notation, $n(\tau) = n\left(\frac{t}{N}\right) = \bar{n}(t)$. Cette probabilité de connexion moyenne est en fait directement liée à la règle d'attachement du modèle HA qui s'exprime comme

$$p\left(k_i \rightarrow k_i + k \mid t\right) = \begin{cases} \bar{n}(t) & \text{si } k_i > 0 \text{ et } k = 1 \text{ (noeuds existants)} \\ \binom{t}{k} \bar{n}(t)^k [1 - \bar{n}(t)]^{t-k} & \text{si } k_i = 0 \text{ (nouveau noeud)} \end{cases} \quad (3.81)$$

Cette probabilité peut ainsi être interprétée comme une probabilité de transition d'un noeud entre deux compartiments k et $k + 1$. On définit le système d'équations compartimentales de la manière suivante,

$$\begin{aligned} N_k(t+1) &= N_k(t) + p\left(k-1 \rightarrow k \mid t\right) N_{k-1}(t) - p\left(k \rightarrow k+1 \mid t\right) N_k(t) + p\left(0 \rightarrow k \mid t\right) \\ &= N_k(t) + \bar{n}(t) [N_{k-1}(t) - N_k(t)] + \binom{t}{k} \bar{n}(t)^k [1 - \bar{n}(t)]^{t-k} \end{aligned} \quad (3.82)$$

où $N_k(t)$ correspond au nombre moyen de noeuds ayant un degré k au temps t . Comme à la Sec. 2.4, chacun des termes de cette équation a une signification particulière : le premier correspond au nombre de noeuds ayant un degré k au temps t avant que le nouveau noeud soit attaché au réseau, le second correspond au nombre de noeuds de degré $k - 1$ entrant dans le compartiment k au temps $t + 1$ avec probabilité $\bar{n}(t)$ (il y a gain d'un lien pour ces noeuds) et, similairement, le troisième correspond au nombre de noeuds de degré k quittant ce dernier compartiment. Le dernier terme quant à lui est donné par la probabilité de transition que le nouveau noeud né au temps $t + 1$ gagne k liens à sa création. On applique également les conditions aux frontières $N_k(t) = 0$ pour tout $k < 0$, $k \geq N$ et $t = 1..N$. Nous avons donc $N + 1$ équations couplées à solutionner.

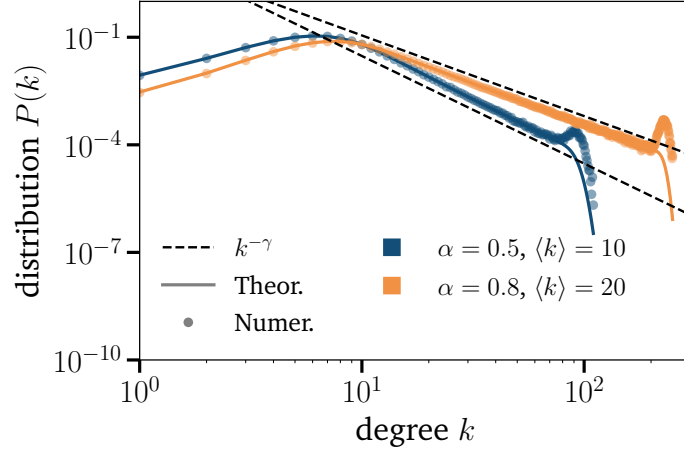


FIGURE 3.14 – Distribution des degrés $P(k)$ de réseaux synthétiques comparée avec la solution $P_k(t = N)$ donnée par les équations compartimentales. Les points correspondent aux résultats issus de simulations Monte-Carlo moyennés sur 96 instances alors que les lignes pleines sont les solutions des équations compartimentales itérées. On indique également, avec les lignes pointillées noires, le comportement de la distribution des degrés en loi de puissance avec un exposant $\gamma = 1 + \frac{1}{\alpha}$. On considère $N = 10000$ et $\tilde{\beta} \rightarrow \infty$. Le reste des paramètres du modèle sont indiqués sur le graphique.

Le nombre de noeuds $N_k(t)$ est une quantité extensive, c'est-à-dire qui croît avec la taille du réseau. C'est pourquoi il est souvent plus pratique de travailler avec $P_k(t) = \frac{N_k(t)}{t}$, la fraction des noeuds ayant un degré k au temps t . Dans cette perspective, le système devient

$$N(t+1)P_k(t+1) = tP_k(t) + t\bar{n}(t) \left[P_{k-1}(t) - P_k(t) \right] + \binom{t}{k} \bar{n}(t)^k \left[1 - \bar{n}(t) \right]^{t-k} \quad (3.83)$$

Lorsque $t = N$, $P_k(t)$ correspond à la distribution des degrés moyenne après le processus de croissance avec $\tilde{\mu}(\tau) = \tilde{\mu}(t/N)$. Les équations compartimentales peuvent être solutionnées simplement en itérant chaque équation du système successivement jusqu'au temps $t = N$ avec la condition initiale $P_k(0) = 0 \forall k \geq 0$.

Afin d'étudier la validité de l'approche compartimentale, on considère le cas de réseaux géométriques *scale-free*, c'est-à-dire avec $\tilde{\beta} \rightarrow \infty$ et

$$\tilde{\mu} \left(\frac{t}{N} \right) = \frac{\nu}{N(\alpha+1)} \left[\alpha \left(\frac{t}{N} \right)^{-1-\alpha} + 1 \right] \quad (3.84)$$

où la distribution des degrés est en loi de puissance $P(k) \sim k^{-\gamma}$ où l'exposant s'écrit $\gamma = 1 + \frac{1}{\alpha}$ (voir Chapitre 3). Puisque $\tilde{\beta} \rightarrow \infty$, on rappelle que, dans ce régime, $\bar{n}(t) = \tilde{\mu}(t/N)$ et donc on peut générer des réseaux géométriques à partir de l'Éq. (3.84) et solutionner le système d'équation avec les valeurs de $\bar{n}(t)$ correspondantes.

3.14.2 Solution de l'état à l'équilibre

Comme nous l'avons vu à la Section 2.4, il est possible de solutionner dans la limite $t = N \rightarrow \infty$ les équations compartimentales en calculant la solution de l'état à l'équilibre. En général, dans le cas où les probabilités de transition sont indépendantes du temps [38], la solution de l'état à l'équilibre s'obtient en posant que la variation de la taille d'un compartiment, dans la limite $t \rightarrow \infty$ est nulle. Ceci se traduit, dans le cas qui nous concerne, par $P_k(t+1) = P_k(t) \equiv P_k^*$, où P_k^* est défini comme la solution stationnaire. Cette solution, correspondant à l'état à l'équilibre, est indépendante du temps. Dans la limite où $t \rightarrow \infty$, on peut réduire les équations maîtresses à la forme récursive suivante

$$P_k^* = t\bar{n}(t) \left[P_{k-1}^* - P_k^* \right] + \binom{t}{k} \bar{n}(t)^k \left[1 - \bar{n}(t) \right]^{t-k} \quad (3.85)$$

$$= \left(\frac{t\bar{n}(t)}{t\bar{n}(t) + 1} \right) P_{k-1}^* + \frac{1}{t\bar{n}(t) + 1} \binom{t}{k} \bar{n}(t)^k \left[1 - \bar{n}(t) \right]^{t-k} \quad (3.86)$$

Cette dernière équation peut être solutionnée de manière récursive en remplaçant P_{k-1}^* par la même expression avec $k \rightarrow k - 1$. On trouve la solution finale pour l'état à l'équilibre

$$P_k^* = \left(\frac{\phi(t)}{\phi(t) + 1} \right)^k P_0^* + \frac{1}{\phi(t) + 1} \sum_{i=0}^{k-1} \binom{t}{k-i} \bar{n}(t)^{k-i} \left(\frac{\phi(t)}{\phi(t) + 1} \right)^i \left[1 - \bar{n}(t) \right]^{t-k+i} \quad (3.87)$$

où $\phi(t) \equiv t\bar{n}(t)$ peut être interprété comme le nombre moyen de liens créés au temps t et, en considérant que $\bar{n}(t) = \mathcal{O}(N^{-1})$, on calcule

$$P_0^* = \frac{\left[1 - \bar{n}(t) \right]^t}{\phi(t) + 1} \simeq \frac{e^{-\phi(t)}}{\phi(t) + 1} \quad (3.88)$$

Ainsi, on peut approximer la solution à l'équilibre à

$$P_k^* \simeq \frac{e^{-\phi(t)}}{\phi(t) + 1} \left(\frac{\phi(t)}{\phi(t) + 1} \right)^k \sum_{i=0}^k \binom{t}{j} \left[\left(\frac{\bar{n}(t)}{1 - \bar{n}(t)} \right) \left(\frac{1 + \phi(t)}{\phi(t)} \right) \right]^j \quad (3.89)$$

En utilisant l'approximation de Stirling, on approxime, pour $n \gg k$, $\binom{n}{k} \simeq \frac{n^k}{k!}$ et on obtient

$$P_k^* \simeq \frac{e^{-\phi(t)}}{\phi(t) + 1} \left(\frac{\phi(t)}{\phi(t) + 1} \right)^k \sum_{i=0}^k \frac{1}{j!} \left[\frac{1 + \phi(t)}{1 - \bar{n}(t)} \right]^j \quad (3.90)$$

$$\simeq \frac{e}{\phi(t) + 1} \left(\frac{\phi(t)}{\phi(t) + 1} \right)^k \frac{\Gamma(k + 1, 1 + \phi(t))}{\Gamma(k + 1)} \quad (3.91)$$

où on approxime $\bar{n}(t) \ll 1$ et on évalue la somme de l'Éq. (3.90) par

$$\sum_{i=0}^k \frac{x^i}{i!} = \frac{e^x \Gamma(k + 1, x)}{\Gamma(k + 1)} \quad (3.92)$$

L'Éq. (3.91) ne correspond pas encore tout à fait à la solution stationnaire, puisqu'il existe toujours une dépendance en t . On considère alors

$$\phi = \lim_{t \rightarrow \infty} \phi(t) = \lim_{t \rightarrow \infty} t\bar{n}(t) \quad (3.93)$$

On distingue trois régimes. Le premier cas $\phi \rightarrow 0$ implique que $\bar{n}(t)$ tend vers 0 plus vite que $\frac{1}{t}$. On le nomme le régime *intensif* puisque le nombre moyen de liens créé à chaque temps tend vers 0. Ceci fait en sorte que le réseau ne croîtra plus après un certain temps et le nombre moyen total de liens sera fini dans la limite thermodynamique. Pour le régime intensif, la solution de la distribution des degrés dans l'état d'état est simple. Elle tend exponentiellement avec $k \ln \phi$ vers

$$P_k^{int} \simeq \delta(k, 0) \quad (3.94)$$

où $\delta(k, k')$ est la fonction Delta de Kronecker. Ce résultat particulièrement inintéressant démontre qu'à l'état à l'équilibre, une fraction importante des noeuds n'ont aucun voisin. Évidemment, toute l'information sur la structure du sous-réseau connecté, où tous les noeuds de degrés zéro sont retirés, est perdue dans ce formalisme. Dans le second régime, que l'on nomme le régime *extensif*, $\phi \rightarrow \infty$ qui implique que $\bar{n}(t)$ tend vers 0 moins vite que $\frac{1}{t}$. Dans ce cas-ci, le nombre moyen de liens créés à chaque temps augmente avec le nombre de noeuds et on peut approximer la solution stationnaire de la manière suivante, avec $\Gamma(s, x) \simeq x^{s-1}e^{-x}$ pour $x \rightarrow \infty$,

$$P_k^{ext} = \frac{(\phi + 1)^{k-1} e^{-\phi}}{\Gamma(k + 1)} \simeq \frac{\phi^k e^{-\phi}}{\Gamma(k + 1)} \quad (3.95)$$

Cette distribution s'approxime lorsque $k \gg 1$ par une distribution de Poisson dont le premier moment est donné par ϕ . Finalement, on appelle le régime *balancé* celui où ϕ est une constante non nulle et la distribution des degrés est donnée par

$$P_k^{bal} = \frac{e}{\phi + 1} \left(\frac{\phi}{\phi + 1} \right)^k \frac{\Gamma(k + 1, 1 + \phi)}{\Gamma(k + 1)}. \quad (3.96)$$

3.14.3 Limites de l'approche par équations compartimentales

L'approche par équations compartimentales s'est révélée être utile pour calculer numériquement avec beaucoup de précision la distribution des degrés du modèle d'évolution géométrique et pour éclaircir l'évolution de la distribution des degrés avec l'état à l'équilibre. Cependant, cette approche préliminaire possède des limitations majeures justifiant l'utilisation d'une méthode plus adaptée comme l'approche par variables cachées [21].

Solutionner le systèmes d'équations compartimentales—En dehors de l'état à l'équilibre, les équations compartimentales n'ont pas de solution fermée et ne peuvent être solutionnées qu'itérativement. Ceci est un problème puisque, dans le cas où la distribution des degrés est en loi de puissance, il devient impossible d'aller calculer l'exposant à partir seulement des équations compartimentales.

Ainsi, on ne bénéficie d'aucune intuition mathématique de la forme générale de la distribution des degrés complète. De plus, comme on doit itérer explicitement les équations compartimentales à chaque temps, lorsque $N \rightarrow \infty$, la solution prend un temps d'intégration considérable à calculer. Des solutions à ce problème existent, notamment transformer le système d'équation compartimentales en un système d'équations différentielles. Toutefois, même pour ces méthodes, l'intégration demeure plus coûteuse que l'approche par variables cachées.

Choisir la distribution des degrés—Comme il est difficile d'avoir une intuition mathématique sur la forme générale de la solution de les équations compartimentales, il devient également difficile de choisir la forme de la distribution des degrés. Au Chapitre 3, nous avons montré que l'expression pour le degré espéré $\kappa(\tau)$ d'un noeud né au temps τ peut être inversé de telle sorte que l'on puisse calculer $\tilde{\mu}(\tau)$ pour reproduire en moyenne une séquence de degrés $\mathbf{k} = (\kappa(\tau))$. Ainsi, selon l'approche par variables cachées, il devient possible de choisir la distribution des degrés simplement en choisissant la forme de $\tilde{\mu}(\tau)$ appropriée. Ce résultat, qui constitue l'un des points forts de notre contribution, n'aurait pas pu être identifié seulement avec l'approche par les équations compartimentales.

Calculer d'autres propriétés structurelles—L'un des nets avantages de l'approche par variables cachées est celui de pouvoir calculer la distribution des degrés ainsi que d'autres propriétés structurelles telles que la corrélation de degrés et le coefficient d'agrégation. Bien qu'il serait possible d'écrire un système d'équations similaire à celui décrit par l'Éq. (3.82) pour calculer ces deux propriétés, il est clair qu'elle serait au moins aussi difficile à manipuler que l'Éq. (3.82).

Conclusions et perspectives

Le jaguar est-il seulement la somme de ses quarks ? L'étude des systèmes complexes nous enseigne que les quarks d'un jaguar sont organisés selon une structure mésoscopique bien spécifique qui fait en sorte qu'on le distingue d'un tigre, et ce, même s'ils sont composés des mêmes particules élémentaires. C'est dans la structure que réside la complexité. La science des réseaux complexes permet de modéliser cette structure et de la quantifier sous le concept unificateur de réseau, lequel se compose de noeuds et de liens [86]. C'est grâce à cette avancée conceptuelle qu'il a été possible de classer un grand nombre de systèmes complexes de différentes natures, allant du cerveau humain à l'internet, selon leurs propriétés structurelles. Plus important encore, il est maintenant possible d'identifier des propriétés structurelles communes et invariantes d'échelle liant un grand nombre de ces systèmes.

Ces systèmes ne sont toutefois pas statiques : ils évoluent dans le temps, croissent et se détruisent en fonction de contraintes auxquelles ils sont soumis. Modéliser leur évolution devient alors d'une importance capitale pour comprendre l'émergence de ces propriétés communes. Les modèles de croissance stochastiques comme l'attachement préférentiel permettent d'expliquer l'émergence d'un sous-ensemble important de ces propriétés structurelles, la distribution des degrés en loi de puissance notamment [11]. Cependant, on doit faire appel à des mécanismes additionnels, tels que l'existence d'un espace métrique caché, pour justifier la présence d'autres propriétés structurelles. La combinaison de ces deux mécanismes mène à des modèles de croissance hautement réalistes et permet, d'une part, de comprendre l'évolution temporelle d'un grand nombre de réseaux réels et, d'autre part, d'inférer des cartes bidimensionnelles très simples de ces réseaux dans des espaces hyperboliques [19, 97, 99].

Ces modèles ne sont cependant réalistes que dans une certaine mesure, car plusieurs propriétés structurelles demeurent encore hors de leur portée. L'une d'entre elles est l'assortativité liée à la corrélation des degrés des noeuds dans le réseau. Cette propriété, difficilement saisie par la majorité des modèles de croissance proposés jusqu'à présent, quantifie la force du coeur des noeuds de haut degré dans le réseau et constitue ainsi une propriété globale de premier plan. L'objectif des travaux présentés dans ce mémoire s'orientait autour de cette problématique : développer un modèle de croissance de réseaux qui puisse reproduire le plus grand nombre de propriétés structurelles, dont l'assortativité, à partir d'un mécanisme d'évolution simple et général.

Pour placer cette problématique en contexte, nous avons d'abord présenté au Chapitre 1 les différentes propriétés structurelles et leurs distributions à travers plusieurs réseaux réels. Nous avons ainsi été en mesure d'identifier les propriétés structurelles d'intérêt pour l'élaboration du modèle de croissance que nous convoitions : le caractère *sparse*, la distribution en degré en loi de puissance (*scale-free*), l'assortativité (désassortativité) par degré et le coefficient d'agrégation élevé. Par la suite, au Chapitre 2, nous avons présenté les modèles fondamentaux de réseaux aléatoires permettant de modéliser ces propriétés. Les modèles importants ont d'abord été présentés : le modèle d'Erdős-Rényi [41], le modèle des configurations [94] et le modèle de Boguñá-Pastor-Satorras de variables cachées [21], ajoutant ainsi toujours un degré de complexité additionnel aux modèles. Ces analyses nous ont ensuite orientés dans deux directions différentes : les réseaux géométriques [35, 75] et les modèles de croissance [11, 73, 98], le coeur des modèles d'évolution de réseaux complexes les plus élaborés à ce jour.

Ce travail de recherche s'est concrétisé au Chapitre 3, où nous avons présenté un modèle original de réseaux géométriques en croissance. La contribution majeure de ce travail est conceptuelle. En effet, au lieu de faire appel à l'attachement préférentiel pour définir notre mécanisme, comme la majorité des modèles de croissance, nous faisons appel à un mécanisme que nous appelons attachement homogène, où les nouveaux noeuds se connectent aux plus vieux sans préférence selon une probabilité uniforme constante. Toutefois, pour obtenir les propriétés structurelles voulues, cette probabilité varie en fonction du temps de naissance. En termes de réseaux géométriques, cela revient à dire que le seuil de connexion, qui s'apparente à un potentiel chimique en physique statistique, est une fonction dépendante du temps. Une analyse en termes de variables cachées nous a, par la suite, permis de décrire mathématiquement la structure des réseaux de cet ensemble avec grande précision, et ce pour une dépendance temporelle arbitraire du seuil.

Trois résultats découlant de cette analyse se sont démarqués. Le premier est qu'il est possible de fixer la séquence des degrés attendus de manière exacte en choisissant une forme temporelle appropriée pour le seuil. De cette manière, on peut saisir la propriété invariante d'échelle de la distribution des degrés. Comme second résultat, nous avons montré que l'ordre dans lequel les noeuds apparaissent, l'histoire du réseau, considérant que leur degré final est fixé, a un impact sur la corrélation des degrés. Ceci permet notamment de naviguer d'une structure assortative à une structure désassortative simplement en inversant l'ordre d'apparition des noeuds (le n -ième plus vieux devient le plus n -ième jeune et *vice-versa* pour tout $n = 0, \dots, N - 1$). Finalement, le troisième résultat est l'identification d'une transition de phase du coefficient d'agrégation moyen du réseau. Nous avons montré qu'en changeant le paramètre de température du réseau, qui correspond au couplage entre l'espace et la structure, le réseau se retrouve soit dans une phase aléatoire, où le coefficient d'agrégation moyen tend vers zéro dans la limite thermodynamique, soit dans une phase géométrique, où il tend vers une constante non-nulle dans cette limite.

Ces travaux ne sont cependant qu'une première itération dans l'élaboration des classes de modèles de croissance par attachement homogène. En effet, plusieurs généralisations du modèle peuvent être

implémentées dans le futur pour améliorer notre compréhension du modèle et pour le rendre plus flexible. La première serait de permettre l'existence de multiliens, de boucles et de directionnalité des liens dans le réseau. Cette modification mineure permettrait de régler des problèmes liés à la génération de certaines séquences de degrés hétérogènes dans un ordre particulier.

La seconde généralisation serait d'améliorer le formalisme de variables cachées pour admettre des ensembles d'histoires dans les calculs. Cette généralisation permettrait d'étudier de manière formelle un cas d'étude que nous avons présenté, les histoires aléatoires. Nous pensons, entre autres, que l'ensemble des réseaux générés avec une histoire aléatoire ne favorise ni un comportement assortatif, ni un comportement disassortatif. Il devient ainsi naturel de poser la question si cet ensemble de réseaux pourrait être lié à celui du CM et, par conséquent, à celui des réseaux géométriques hyperboliques. Un tel formalisme admettant des ensembles d'histoires pourrait ainsi s'attaquer à cette question.

La dernière contribution future, qui fut présentée comme preuve de concept en complément du Chapitre 3, est l'inférence de ces plongements couplés à la croissance. En effet, nous avons montré qu'il est possible, seulement à partir d'une capture de la structure à un certain temps, de réinférer le processus d'évolution. Ceci est effectué en choisissant l'histoire (et donc le processus) qui maximise la vraisemblance d'un réseau ayant une structure donnée. L'objectif serait ainsi d'étendre cette méthode pour inférer, d'une part, l'âge des noeuds et, d'autre part, leur position dans l'espace. Cette méthode pourrait ensuite être utilisée sur des réseaux complexes réels pour en inférer les positions dans l'espace.

Le discours principal de ce mémoire était surtout orienté autour de la structure des réseaux complexes et de leur évolution. Par contre, je crois que ces contributions vont au-delà du modèle de croissance ultimement proposé. En effet, pour parvenir à développer ce modèle, nous avons fait appel à une expertise nouvelle et un ensemble de méthodes novatrices liés à la géométrie des réseaux complexes. Comme nous l'avons mentionné lors de l'introduction, la science des réseaux se divise en deux parties ; la première étant la structure et la seconde, la dynamique. Dans ce mémoire, nous avons exploré la structure d'un point de vue géométrique. Il serait ainsi naturel de vouloir explorer la dynamique dans cette même direction.

Modéliser la propagation de maladies, de nouvelles, de rumeurs ou de toute autre forme d'information sur un réseau réel [12, 17, 101] a toujours été chose difficile. Les plus récentes avancées dans ce domaine doivent limiter leur analyse à des structures relativement simples [20, 101, 116] pour éviter un cauchemar mathématique. Nous pensons qu'une perspective géométrique offre une approche plus réaliste qui pourrait avoir impact considérable sur l'avancement de ce domaine de recherche.

Bibliographie

- [1] Adamic, L. A., Lukose, R. M., Puniyani, A. R., and Huberman, B. A. (2001). Search in power-law networks. *Phys. Rev. E*, 64(4) :046135.
- [2] Aiello, W., Chung, F., and Lu, L. (2000). A random graph model for massive graphs. In *Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing*, pages 171–180. ACM.
- [3] Aiello, W., Chung, F., and Lu, L. (2002). Random evolution in massive graphs. In *Handbook of massive data sets*, pages 97–122. Springer.
- [4] Albert, R., Albert, I., and Nakarado, G. L. (2004). Structural vulnerability of the north american power grid. *Phys. Rev. E*, 69(2) :025103.
- [5] Albert, R., Oltvai, Z., Barabasi, A., Jeong, H., and Tombor, B. (2000). The large-scale organization of metabolic networks. *Nature*, 407(6804) :651–654.
- [6] Allard, A., Serrano, M. Á., García-Pérez, G., and Boguñá, M. (2017). The geometric nature of weights in real complex networks. *Nat. Commun.*, 8 :14103.
- [7] Amaral, L. A. N., Scala, A., Barthélemy, M., and Stanley, H. E. (2000). Classes of small-world networks. *Proc. Natl. Acad. Sci.*, 97(21) :11149–11152.
- [8] Auletta, G., Fortunato, M., and Parisi, G. (2009). *Quantum Mechanics*. Cambridge University Press.
- [9] Balister, P., Bollobás, B., and Walters, M. (2004). Continuum percolation with steps in an annulus. *Ann. Appl. Probab.*, pages 1869–1879.
- [10] Barabási, A.-L. (2016). *Network Science*. Cambridge University Press.
- [11] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439) :509–512.
- [12] Barrat, A., Barthélemy, M., and Vespignani, A. (2008). *Dynamical Processes on Complex Networks*. Cambridge University Press.
- [13] Barthélemy, M. (2011). Spatial networks. *Phys. Rep.*, 499(1) :1–101.

- [14] Bassett, D. S. and Bullmore, E. (2006). Small-world brain networks. *Neurosci.*, 12(6) :512–523.
- [15] Bearman, P. S., Moody, J., and Stovel, K. (2004). Chains of affection : The structure of adolescent romantic and sexual networks. *Am. J. Sociol.*, 110(1) :44–91.
- [16] Biggs, N., Lloyd, E. K., and Wilson, R. J. (1976). *Graph Theory, 1736-1936*. Oxford University Press.
- [17] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. (2006). Complex networks : Structure and dynamics. *Phys. Rep.*, 424(4) :175–308.
- [18] Boguna, M., Krioukov, D., and Claffy, K. C. (2009). Navigability of complex networks. *Nat. Phys.*, 5(1) :74–80.
- [19] Boguná, M., Papadopoulos, F., and Krioukov, D. (2010). Sustaining the internet with hyperbolic mapping. *Nat. Commun.*, 1 :62.
- [20] Boguná, M. and Pastor-Satorras, R. (2002). Epidemic spreading in correlated complex networks. *Phys. Rev. E*, 66(4) :047104.
- [21] Boguná, M. and Pastor-Satorras, R. (2003). Class of correlated random networks with hidden variables. *Phys. Rev. E*, 68(3) :036112.
- [22] Boguñá, M., Pastor-Satorras, R., Díaz-Guilera, A., and Arenas, A. (2004). Models of social networks based on social distance attachment. *Phys. Rev. E*, 70(5) :056122.
- [23] Boguná, M., Pastor-Satorras, R., and Vespignani, A. (2004). Cut-offs and finite size effects in scale-free networks. *Eur. Phys. J.*, 38(2) :205–209.
- [24] Bonato, A., Gleich, D. F., Kim, M., Mitsche, D., Prałat, P., Tian, Y., and Young, S. J. (2014). Dimensionality of social networks using motifs and eigenvalues. *PloS One*, 9(9) :e106052.
- [25] Bonato, A., Janssen, J., and Prałat, P. (2012). Geometric protean graphs. *Internet Math.*, 8(1-2) :2–28.
- [26] Bullmore, E. and Sporns, O. (2012). The economy of brain network organization. *Nat. Rev. Neurosci.*, 13(5) :336–349.
- [27] Callaway, D. S., Newman, M. E., Strogatz, S. H., and Watts, D. J. (2000). Network robustness and fragility : Percolation on random graphs. *Phys. Rev. Lett.*, 85(25) :5468.
- [28] Cardillo, A., Scellato, S., Latora, V., and Porta, S. (2006). Structural properties of planar graphs of urban street patterns. *Phys. Rev. E*, 73(6) :066107.
- [29] Carroll, S. M. (2004). *Spacetime and Geometry. An Introduction to General Relativity*. Number 1. Pearson.

- [30] Castro, R. and Grossman, J. W. (1999). Famous Trails to Paul Erdős. *Math. intell.*, 3(21) :51–53.
- [31] Chen, Q., Chang, H., Govindan, R., and Jamin, S. (2002). The origin of power laws in internet topologies revisited. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 608–617. IEEE.
- [32] Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Rev.*, 51(4) :661–703.
- [33] Cordeiro, C. D. M. and Agrawal, D. P. (2011). *Ad Hoc and Sensor Networks : Theory and Applications*. World Scientific.
- [34] Cottingham, W. N. and Greenwood, D. A. (2007). *An Introduction to the Standard Model of Particle Physics*. Cambridge University Press.
- [35] Dall, J. and Christensen, M. (2002). Random geometric graphs. *Phys. Rev. E*, 66(1) :016121.
- [36] Davis, G. F., Yoo, M., and Baker, W. E. (2003). The small world of the american corporate elite, 1982-2001. *Strateg. Organ.*, 1(3) :301–326.
- [37] Dettmann, C. P. and Georgiou, O. (2016). Random geometric graphs with general connection functions. *Phys. Rev. E*, 93(3) :032313.
- [38] Diu, B., Roulet, B., Guthmann, C., and Lederer, D. (1989). *Eléments de Physique Statistique*. Hermann.
- [39] Dorogovtsev, S. N. and Mendes, J. F. F. (2001). Language as an evolving word web. *Proc. Royal Soc. Lond. B : Biol. Sci.*, 268(1485) :2603–2606.
- [40] Ehrenfest, P. and Ehrenfest, T. (2002). *The Conceptual Foundations of the Statistical Approach in Mechanics*. Courier Corporation.
- [41] Erdős, P. and Rényi, A. (1959). On random graphs. *Publ. Math. Debrecen*, 6 :290–297.
- [42] Eriksen, K. A. and Hörnquist, M. (2001). Scale-free growing networks imply linear preferential attachment. *Phys. Rev. E*, 65(1) :017102.
- [43] Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On power-law relationships of the internet topology. In *ACM SIGCOMM Computer Communication Review*, volume 29, pages 251–262. ACM.
- [44] Ferrer i Cancho, R., Janssen, C., and Solé, R. V. (2001). Topology of technology graphs : Small world patterns in electronic circuits. *Phys. Rev. E*, 64(4) :046119.
- [45] Ferrer i Cancho, R. and Solé, R. V. (2001). The small world of human language. *Proc. Royal Soc. Lond. B : Biol. Sci.*, 268(1482) :2261–2265.

- [46] Ferretti, L. and Cortelezzi, M. (2011). Preferential attachment in growing spatial networks. *Phys. Rev. E*, 84(1) :016103.
- [47] Ferretti, L., Cortelezzi, M., and Mamino, M. (2014). Duality between preferential attachment and static networks on hyperbolic spaces. *Eur. Phys. Lett.*, 105(3) :38001.
- [48] Flaxman, A. D., Frieze, A. M., and Vera, J. (2006). A geometric preferential attachment model of networks. *Internet Math.*, 3(2) :187–205.
- [49] Flaxman, A. D., Frieze, A. M., and Vera, J. (2007). A geometric preferential attachment model of networks ii. *Internet Math.*, 4(1) :87–111.
- [50] Fornito, A. and Bullmore, E. T. (2015). Connectomics : a new paradigm for understanding brain disease. *Eur. Neuropsychopharmacol.*, 25(5) :733–748.
- [51] Fortunato, S. (2010). Community detection in graphs. *Phys. Rep.*, 486(3) :75–174.
- [52] Fortunato, S. and Barthélemy, M. (2007). Resolution limit in community detection. *Proc. Natl. Acad. Sci.*, 104(1) :36–41.
- [53] García-Pérez, G., Boguñá, M., Allard, A., and Serrano, M. Á. (2016). The hidden hyperbolic geometry of international trade : World Trade Atlas 1870–2013. *Sci. Rep.*, 6 :33441.
- [54] García-Pérez, G., Serrano, M., and Boguñá, M. (2017). Soft communities in similarity space. *arXiv preprint arXiv :1707.09610*.
- [55] Gell-Mann, M. (1995). *The Quark and the Jaguar : Adventures in the Simple and the Complex*. W. H. Freeman & Co., New York, NY, USA.
- [56] Gilbert, E. N. (1959). Random graphs. *Ann. Math. Stat.*, 30(4) :1141–1144.
- [57] Gilbert, E. N. (1961). Random plane networks. *J. Soc. Ind. Appl. Math.*, 9(4) :533–543.
- [58] Gleeson, J. P. (2011). High-accuracy approximation of binary-state dynamics on networks. *Phys. Rev. Lett.*, 107(6) :068701.
- [59] Grossman, J. W. and Ion, P. D. (1995). On a portion of the well-known collaboration graph. *Congr. Numerantium*, pages 129–132.
- [60] Guimera, R. and Amaral, L. A. N. (2004). Modeling the world-wide airport network. *Eur. Phys. J. B*, 38(2) :381–385.
- [61] Haenggi, M., Andrews, J. G., Baccelli, F., Dousse, O., and Franceschetti, M. (2009). Stochastic geometry and random graphs for the analysis and design of wireless networks. *IEEE J. Sel. Areas Commun.*, 27(7).

- [62] Hébert-Dufresne, L., Allard, A., Marceau, V., Noël, P.-A., and Dubé, L. J. (2011). Structural preferential attachment : Network organization beyond the link. *Phys. Rev. Lett.*, 107(15) :158702.
- [63] Hébert-Dufresne, L., Allard, A., Marceau, V., Noël, P.-A., and Dubé, L. J. (2012). Structural preferential attachment : Stochastic process for the growth of scale-free, modular, and self-similar systems. *Phys. Rev. E*, 85(2) :026108.
- [64] Hébert-Dufresne, L., Allard, A., Young, J.-G., and Dubé, L. J. (2016). Constrained growth of complex scale-independent systems. *Phys. Rev. E*, 93(3) :032304.
- [65] Hébert-Dufresne, L., Laurence, E., Allard, A., Young, J.-G., and Dubé, L. J. (2015). Complex networks as an emerging property of hierarchical preferential attachment. *Phys. Rev. E*, 92(6) :062809.
- [66] Herrmann, C., Barthélemy, M., and Provero, P. (2003). Connectivity distribution of spatial networks. *Phys. Rev. E*, 68(2) :026128.
- [67] Holme, P. and Saramäki, J. (2012). Temporal networks. *Phys. rep.*, 519(3) :97–125.
- [68] Jacob, E. and Mörters, P. (2015). Spatial preferential attachment networks : Power laws and clustering coefficients. *Ann. Appl. Probab.*, 25(2) :632–662.
- [69] Jeong, H., Mason, S. P., Barabasi, A.-L., and Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *arXiv preprint cond-mat/0105306*.
- [70] Karrer, B. and Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E*, 83(1) :016107.
- [71] Kitsak, M., Papadopoulos, F., and Krioukov, D. (2017). Latent geometry of bipartite networks. *Phys. Rev. E*, 95(3) :032309.
- [72] Klemm, K. and Eguiluz, V. M. (2002). Growing scale-free networks with small-world behavior. *Phys. Rev. E*, 65(5) :057102.
- [73] Krapivsky, P. L. and Redner, S. (2001). Organization of growing random networks. *Phys. Rev. E*, 63(6) :066123.
- [74] Krioukov, D. (2016). Clustering implies geometry in networks. *Phys. Rev. Lett.*, 116(20) :208302.
- [75] Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A., and Boguná, M. (2010). Hyperbolic geometry of complex networks. *Phys. Rev. E*, 82(3) :036106.
- [76] Krioukov, D., Papadopoulos, F., Vahdat, A., and Boguná, M. (2009). Curvature and temperature of complex networks. *Phys. Rev. E*, 80(3) :035101.
- [77] Kuhn, F., Wattenhofer, R., and Zollinger, A. (2003). Ad-hoc networks beyond unit disk graphs. In *Proceedings of the 2003 Joint Workshop on Foundations of Mobile Computing*, pages 69–78. ACM.

- [78] Li, W. and Cai, X. (2004). Statistical analysis of airport network of china. *Phys. Rev. E*, 69(4) :046106.
- [79] Liljeros, F., Edling, C. R., and Amaral, L. A. N. (2003). Sexual networks : implications for the transmission of sexually transmitted infections. *Microbes Infect.*, 5(2) :189–196.
- [80] Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., and Aberg, Y. (2001). The web of human sexual contacts. *arXiv preprint cond-mat/0106507*.
- [81] Lovász, L. (2012). *Large Networks and Graph Limits*, volume 60. American Mathematical Society Providence.
- [82] Marceau, V., Noël, P.-A., Hébert-Dufresne, L., Allard, A., and Dubé, L. J. (2010). Adaptive networks : Coevolution of disease and topology. *Phys. Rev. E*, 82(3) :036116.
- [83] Meester, R. and Roy, R. (1996). *Continuum Percolation*, volume 119. Cambridge University Press.
- [84] Murphy, C., Allard, A., Laurence, E., St-Onge, G., and Dubé, L. J. (2018). Geometric evolution of complex networks with degree correlations. *Phys. Rev. E*, 97(3) :032309.
- [85] Nakahara, M. (2003). *Geometry, Topology and Physics*. CRC Press.
- [86] Newman, M. (2010). *Networks : An Introduction*. Oxford University Press.
- [87] Newman, M. E. (2001a). Scientific collaboration networks. i. network construction and fundamental results. *Phys. Rev. E*, 64(1) :016131.
- [88] Newman, M. E. (2001b). The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci.*, 98(2) :404–409.
- [89] Newman, M. E. (2002). Assortative mixing in networks. *Phys. Rev. Lett.*, 89(20) :208701.
- [90] Newman, M. E. (2003a). Mixing patterns in networks. *Phys. Rev. E*, 67(2) :026126.
- [91] Newman, M. E. (2003b). The structure and function of complex networks. *SIAM Rev.*, 45(2) :167–256.
- [92] Newman, M. E. (2006). Modularity and community structure in networks. *Proc. Natl. Acad. Sci.*, 103(23) :8577–8582.
- [93] Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2) :026113.
- [94] Newman, M. E., Strogatz, S. H., and Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64(2) :026118.
- [95] Newman, M. E. J. and Girvan, M. (2003). *Mixing Patterns and Community Structure in Networks*, pages 66–87. Springer, Berlin, Heidelberg.

- [96] Olver, F. W. J., Olde Daalhuis, A. B., Lozier, D. W., Schneider, B. I., R. F., B., Clark, C. W., Miller, B., and Saunders, B. V. (2010–2018). Nist digital library of mathematical functions. <http://dlmf.nist.gov/>, Release 1.0.18 of 2018-03-27.
- [97] Papadopoulos, F., Aldecoa, R., and Krioukov, D. (2015a). Network geometry inference using common neighbors. *Phys. Rev. E*, 92(2) :022807.
- [98] Papadopoulos, F., Kitsak, M., Serrano, M., Boguná, M., and Krioukov, D. (2011). Popularity versus similarity in growing networks. *arXiv preprint arXiv :1106.0286*.
- [99] Papadopoulos, F., Psomas, C., and Krioukov, D. (2015b). Network mapping by reur. phys. lett.aying hyperbolic growth. *IEEE/ACM Trans. Netw.*, 23(1) :198–211.
- [100] Park, J. and Newman, M. E. (2004). Statistical mechanics of networks. *Phys. Rev. E*, 70(6) :066117.
- [101] Pastor-Satorras, R., Castellano, C., Van Mieghem, P., and Vespignani, A. (2015). Epidemic processes in complex networks. *Rev. Mod. Phys.*, 87(3) :925.
- [102] Pastor-Satorras, R., Vázquez, A., and Vespignani, A. (2001). Dynamical and correlation properties of the internet. *Phys. Rev. Lett.*, 87(25) :258701.
- [103] Pastor-Satorras, R. and Vespignani, A. (2001). Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86(14) :3200.
- [104] Pastor-Satorras, R. and Vespignani, A. (2002). Epidemic dynamics in finite size scale-free networks. *Phys Rev. E*, 65(3) :035108.
- [105] Penrose, M. (2003). *Random Geometric Graphs*. Number 5. Oxford University Press.
- [106] Prałat, P. and Wormald, N. (2007). Growing protean graphs. *Internet Math.*, 4(1) :1–16.
- [107] Press, W. H. (2007). *Numerical Recipes 3rd Edition : The Art of Scientific Computing*. Cambridge University Press.
- [108] Reif, F. (2009). *Fundamentals of Statistical and Thermal Physics*. Waveland Press.
- [109] Ripeanu, M. and Foster, I. (2002). Mapping the gnutella network : Macroscopic properties of large-scale peer-to-peer systems. In *International Workshop on Peer-to-Peer Systems*, pages 85–93. Springer.
- [110] Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P., Mukherjee, G., and Manna, S. (2003). Small-world properties of the indian railway network. *Phys. Rev. E*, 67(3) :036106.
- [111] Serrano, M. Á., Boguñá, M., and Pastor-Satorras, R. (2006). Correlations in weighted networks. *Phys. Rev. E*, 74(5) :055101.

- [112] Serrano, M. A., Krioukov, D., and Boguná, M. (2008). Self-similarity of complex networks and hidden metric spaces. *Phys. Rev. Lett.*, 100(7) :078701.
- [113] Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3/4) :425–440.
- [114] Simon, H. A. (1965). The architecture of complexity. *General Systems*, 10(1965) :63–76.
- [115] Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome : a structural description of the human brain. *PLoS Comput. Biol.*, 1(4) :e42.
- [116] St-Onge, G., Young, J.-G., Laurence, E., Murphy, C., and Dubé, L. J. (2018). Phase transition of the susceptible-infected-susceptible dynamics on time-varying configuration model networks. *Phys. Rev. E*, 97 :022305.
- [117] Strauss, D. (1986). On a general class of models for interaction. *SIAM Rev.*, 28(4) :513–527.
- [118] Towilson, E. K., Vértes, P. E., Ahnert, S. E., Schafer, W. R., and Bullmore, E. T. (2013). The rich club of the c. elegans neuronal connectome. *J. Neurosci.*, 33(15) :6380–6387.
- [119] Van Den Heuvel, M. P. and Sporns, O. (2011). Rich-club organization of the human connectome. *J. Neurosci.*, 31(44) :15775–15786.
- [120] Wasserman, S. and Faust, K. (1994). *Social network analysis : Methods and applications*, volume 8. Cambridge University Press.
- [121] Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, 393(6684) :440.
- [122] Waxman, B. M. (1988). Routing of multipoint connections. *IEEE J. Sel. Areas Commun.*, 6(9) :1617–1622.
- [123] Xie, Z. and Rogers, T. (2016). Scale-invariant geometric random graphs. *Phys. Rev. E*, 93(3) :032310.
- [124] Young, J.-G., Hébert-Dufresne, L., Allard, A., and Dubé, L. J. (2016). Growing networks of overlapping communities with internal structure. *Phys. Rev. E*, 94(2) :022317.
- [125] Young, J.-G., Hébert-Dufresne, L., Laurence, E., Murphy, C., St-Onge, G., and Desrosiers, P. (2018). Network archaeology : phase transition in the recoverability of network history. *arXiv preprint arXiv :1803.09191*.
- [126] Zhou, C., Zemanová, L., Zamora, G., Hilgetag, C. C., and Kurths, J. (2006). Hierarchical organization unveiled by functional connectivity in complex brain networks. *Phys. Rev. Lett.*, 97(23) :238103.
- [127] Zuev, K., Boguná, M., Bianconi, G., and Krioukov, D. (2015). Emergence of soft communities from geometric preferential attachment. *Sci. Rep.*, 5 :9421.

- [128] Zuev, K., Papadopoulos, F., and Krioukov, D. (2016). Hamiltonian dynamics of preferential attachment. *J. Phys. A*, 49(10):105001.