# De la densité spectrale des réseaux extrêmement creux

**Mémoire**

**Olivier Ribordy**

**Maîtrise en physique - avec mémoire**
Maître ès sciences (M. Sc.)

Québec, Canada

© Olivier Ribordy, 2023

# De la densité spectrale des réseaux extrêmement creux

**Mémoire**

**Olivier Ribordy**

Sous la direction de:

Antoine Allard, directeur de recherche

# Résumé

L'étude du spectre des réseaux complexes est un problème riche, ayant une longue histoire et des applications dans de nombreux domaines. La densité spectrale limite d'ensembles de graphes aléatoires, en particulier, fournit une information globale importante pour la compréhension de la topologie et du comportement des modèles de réseaux souvent utilisés comme version jouet des réseaux réels. Si la densité spectrale des réseaux denses et non corrélés est en général bien comprise, celle des réseaux extrêmement creux reste un problème difficile. Bien que plusieurs propriétés de la densité spectrale des réseaux extrêmement creux aient pu être établies, une forme fermée pour celle-ci échappe toujours aux chercheurs et chercheuses. C'est à ce problème que s'attaque ce mémoire. Dans un premier temps, une méthode combinatoire pour le calcul de la densité spectrale et de ses moments est développée. Elle est ensuite appliquée au modèle d'Erdős-Rényi dense avec succès, reproduisant le résultat classique qu'est la loi du demi-cercle de Wigner. Finalement, la méthode est utilisée pour l'étude de la densité spectrale des réseaux extrêmement creux. Sont obtenus ainsi une forme fermée pour la densité spectrale des graphes réguliers aléatoires, la preuve de propriétés importantes de la densité spectrale du modèle d'Erdős-Rényi extrêmement creux, une explication de la présence de pics discontinus dans la densité, une correction pour le coeur de celle-ci, ainsi qu'une forme asymptotique pour ses extrémités conjecturée comme vraie pour tous les modèles de réseaux creux et non corrélés. Malgré tout, une forme fermée est toujours inconnue.

# Abstract

The study of the spectra of complex networks is a rich problem with a long history and varied applications. In particular, the limiting spectral density of random graph ensembles provides important global information on the topology and behavior of the network models often used as toy versions of real networks. Though the spectral density of dense, uncorrelated networks is generally well understood, that of extremely sparse networks remains a difficult problem. Despite the fact that many properties of the spectral density of extremely sparse networks have been established, a closed form still evades researchers. It is that problem which this thesis tackles. First, a combinatorial approach to the calculation of the spectral density and its moments is developed. It is then successfully applied to the dense Erdős-Rényi model, reproducing the classical Wigner semicircle law. Finally, the approach is employed to study the spectral density of extremely sparse networks. Results obtained this way include a closed form for the spectral density of random regular graphs, proofs of important properties of the spectral density of extremely sparse Erdős-Rényi random graphs, an explanation of the presence of discontinuous peaks in the density, a correction for its bulk and an asymptotic form for its extremities, which is conjectured to hold for all models of sparse, uncorrelated networks. However, a closed form remains unknown.

# Table des matières

**Résumé** ii

**Abstract** iii

**Table des matières** iv

**Liste des tableaux** vi

**Liste des figures** vii

**Remerciements** ix

**Introduction** 1

**1 Notions de base** 3
- 1.1 Graphes et propriétés structurelles . . . 3
- 1.2 Matrice d'adjacence . . . 6
- 1.3 Valeurs propres et densité spectrale . . . 9
- 1.4 Graphes aléatoires . . . 11
- 1.5 Séries génératrices . . . 14

**2 Réseaux denses : la loi de Wigner** 16
- 2.1 Nombre attendu de marches fermées . . . 16
- 2.2 Série génératrice des moments . . . 19
- 2.3 Densité spectrale limite . . . 22
- 2.4 Généralisation et limites du résultat . . . 25

**3 Le cas des réseaux extrêmement creux** 29
- 3.1 Observations numériques . . . 29
- 3.2 Formule de récurrence . . . 31
- 3.3 Propriétés de la séquence $I_{n,\ell}$ . . . 34
- 3.4 Conséquences pour la densité spectrale . . . 41
- 3.5 Solution exacte : les graphes réguliers . . . 46
- 3.6 Forme asymptotique pour $x \gg 1$ . . . 49

**Conclusion** 52

**A Formules d'inversion pour la densité spectrale** 54

**B Correction à la loi du demi-cercle pour réseaux très denses** **57**

**Bibliographie** **59**

# Liste des tableaux

3.1 Valeurs propres de divers petits arbres . . . . . 31
3.2 Nombre $I_{n,\ell}$ de marches fermées de demi-longueur $n$ sur des arbres à $\ell$ arêtes. . 34

# Liste des figures


2.1 Ensemble des 6 formes de marches de longueur 6 visitant 2 arêtes distinctes. . . 17
2.2 Densité spectrale du modèle d'Erdős-Rényi dense. . . . . 25
2.3 Densité spectrale corrigée du modèle d'Erdős-Rényi très dense. . . . . 26
2.4 Loi du demi-cercle pour modèle de Chung-Lu et graphes réguliers. . . . . 27
2.5 Loi du demi-cercle pour modèle de Chung-Lu hétérogène et modèle de Watts-Strogatz . . . . . 28

3.1 Densité spectrale de graphes Erdős-Rényi extrêmement creux . . . . . 30
3.2 Valeurs propres de petits arbres . . . . . 31
3.3 Décomposition des marches . . . . . 32
3.4 Construction des sous-marches . . . . . 33
3.5 Les deux configurations possibles du graphe étoile à 5 arêtes enraciné . . . . . 36
3.6 Bornes de la séquence $I_{n,\ell}$ pour $n = 88$ . . . . . 37
3.7 Correction à la loi du demi-cercle . . . . . 45
3.8 Loi de Kesten-McKay . . . . . 48

Vous me dérangez. Je suis en train de cueillir des champignons.

Grigori Perelman

# Remerciements

Merci à Dynamica pour l'accueil dans ce groupe de recherche exceptionnel, malgré ma présence essentiellement à distance. Je serai de retour.

Merci à Antoine Allard pour ton travail de supervision rigoureux, rassurant et presque toujours pertinent. Ta confiance envers mon instinct d'apprenti-chercheur m'a été précieuse.

Merci à Patrick Desrosiers, Nicolas Doyon, Wenjie Fang et Michel Bauer pour votre temps et votre expertise cruciale. Votre lumière a éclairé une grande partie des lignes de ce mémoire.

Merci à ma famille de s'être toujours intéressé à ce que je fais sans en comprendre un traître mot.

Merci à la bande de physiciens éclatés qui me suit depuis maintenant cinq ans d'avoir rendu ce parcours oh combien plus amusant.

Merci à ma partenaire de vie pour ton soutien inconditionnel.

# Introduction

Au tournant du millénaire, les réseaux complexes se sont imposés comme le paradigme dominant pour l'étude des systèmes complexes, favorisés par leur capacité à intégrer les bonds de géants faits par les ordinateurs et la collection d'immenses jeux de données [2]. L'arrivée au devant de la scène des réseaux complexes a stimulé l'intérêt de la communauté scientifique de la complexité pour un sujet plus ancien, les graphes aléatoires, et plus largement la théorie des matrices aléatoires. Si l'étude des graphes aléatoires a ses racines dans les travaux des sociologues de la première moitié du XX$^e$ siècle [31], son avènement comme discipline mathématique peut être attribué aux travaux fondateurs, à la fin des années 1950, de P. Erdős et A. Rényi [15, 16] ainsi que de E. N. Gilbert [19]. Ces derniers établirent les modèles de réseaux aléatoires qui portent maintenant leurs noms, en plus d'en établir en détail les propriétés importantes. Avec l'arrivée des réseaux complexes, la réalisation que la plupart des réseaux réels affichaient des propriétés fort différentes des modèles d'Erdős-Rényi et de Gilbert amena une nouvelle vague de recherche sur les graphes aléatoires, dans le but d'établir de nouveaux modèles reproduisant plus fidèlement les caractéristiques des réseaux réels. Cette vague de recherche mena notamment aux modèles de Watts-Strogatz [44] et de Barabási-Albert [3], intégrant respectivement la propriété *petit monde* et l'*invariance d'échelle* fréquemment observées dans les réseaux réels.

Naturellement, l'arrivée de ces modèles entraîna une vague de recherche sur les propriétés des réseaux générés par ceux-ci. Une des propriétés ayant suscité de l'intérêt est la densité spectrale limite des graphes aléatoires, soit la distribution de leurs valeurs propres dans la limite où la taille du graphe tend vers l'infini. En plus de son intérêt propre comme sujet mathématique, le spectre des graphes et réseaux complexes possède une multitude d'applications autant en chimie qu'en physique et en mathématiques [10, 26], notamment pour la détéction de communautés au sein des réseaux [32]. L'étude des graphes Erdős-Rényi montra que ceux-ci, dans la limite où chaque noeud était connecté à une proportion non nulle de ses voisins – auquel cas les réseaux sont dits denses –, avaient comme densité spectrale limite la loi du demi-cercle de Wigner [21, 46]. Or, il fut vite constaté que les densités spectrales des modèles de Watts-Strogatz et de Barabasi-Albert, ainsi que du modèle d'Erdős-Rényi où le nombre de connexions des noeuds reste constant dans la limite des grands graphes – dit creux –, étaient totalement différentes de la loi du demi-cercle [17]. Vint ainsi une nouvelle vague

d'articles s'attaquant à cette question [4, 11, 25, 33, 39], avec plus ou moins de succès : au moment d'écrire ce mémoire, des formes fermées pour les densités spectrales de ces modèles échappent toujours aux chercheurs et chercheuses.

Ce projet aborde cette question à partir d'une propriété connue de la trace de puissances de matrices permettant de relier les moments de la densité spectrale limite au compte des marches fermées sur le graphe à l'étude. Cette propriété permet de transformer la question de la densité spectrale en un problème combinatoire. Le mémoire est donc divisé en trois chapitres. Le premier met la table pour les chapitres à venir en établissant des notions de bases sur les graphes, les modèles de graphes aléatoires et leurs propriétés, la densité spectrale et les séries génératrices. Le deuxième chapitre fait la démonstration de la méthode combinatoire en établissant le résultat classique de la loi du demi-cercle de Wigner pour les graphes Erdős-Rényi denses. Finalement, le troisième chapitre s'attaque au problème des réseaux extrêmement creux. Si la méthode développée ne permet pas l'obtention d'une forme fermée pour la densité spectrale limite des graphes Erdős-Rényi, elle mène tout de même à des corrections pour la partie centrale de la densité, à une forme fermée pour la densité limite des graphes réguliers aléatoires, ainsi qu'à une forme asymptotique pour les extrémités de la densité spectrale limite des graphes Erdős-Rényi.

# Chapitre 1

# Notions de base

Ce mémoire portant entièrement sur l'étude des graphes ou réseaux [1] aléatoires, de leur matrice d'adjacence et du spectre de celle-ci, il convient d'introduire ces notions comme il se doit.

## 1.1 Graphes et propriétés structurelles

**Définition 1.1.** Un *graphe* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ est une paire de deux ensembles : un ensemble de *sommets* ou *noeuds* $\mathcal{V}$ et un ensemble d'*arêtes* ou *liens* $\mathcal{E}$ connectant des paires de sommets. Il existe une arrête $u \to v$, $u, v \in \mathcal{V}$, si $(u, v) \in \mathcal{E}$.

Il est intéressant de noter que l'écriture d'un graphe spécifique nécéssite d'étiqueter les sommets. Or, un graphe n'a pas à être étiqueté, puisqu'il est simplement défini par les relations de connexion entre ses noeuds. L'ensemble des étiquetages possibles d'un graphe représente une classe d'isomorphismes.

**Définition 1.2.** La *taille* d'un graphe est définie comme la cardinalité de son ensemble de sommets, $|\mathcal{V}| = N$. La notation utilisée pour la cardinalité de l'ensemble d'arêtes est $|\mathcal{E}| = M$.

**Définition 1.3.** Un graphe est dit *connexe* s'il ne contient pas de composante isolée ou, de manière équivalente, s'il existe une suite d'arêtes continue entre chaque paire de sommets du graphe.

**Définition 1.4.** Un *graphe non orienté* est un graphe où les arêtes sont bidirectionnelles, c'est-à-dire que $(u, v) \in \mathcal{E} \iff (v, u) \in \mathcal{E}$. Pour éviter la redondance, si le graphe est non orienté, $(u, v) = (v, u)$, de sorte que le lien n'apparaît qu'une fois dans $\mathcal{E}$. Les sommets $u$ et $v$ sont alors dits *connectés*, $u \sim v$. On dénote l'ensemble des noeuds connectés au noeud $u$ comme le *voisinage* de $u$, et on appelle ces noeuds les *voisins* de $u$.

1. La terminologie "graphe, arêtes, sommets" est parfaitement équivalente à celle "réseau, liens, noeuds". La première est en général préférée par la communauté mathématicienne, alors que la seconde est préférée par la communauté physicienne. Les deux seront utilisées interchangeablement dans ce mémoire.

**Définition 1.5.** Un *graphe simple* est un graphe non orienté et où la connexion entre deux liens est une propriété binaire : le lien $(u, v)$ existe ou non. Les graphes où un poids est associé aux liens sont dénommés *graphes pondérés*. De plus, un graphe simple ne permet pas l'existence d'autoconnexions, soit $(u, u) \notin \mathcal{E}\ \forall u \in \mathcal{V}$.

Ce mémoire se concentre uniquement sur l'étude des graphes simples. Les graphes pondérés et orientés sont fort utiles pour l'étude des réseaux réels, puisque les interactions entre composantes d'un système réel ne sont pas toujours égales ni réciproques. Toutefois, leur étude est très différente de celle des graphes simples et excéderait la portée de la présente recherche. De plus, les graphes simples sont les plus étudiés en analyse spectrale [35]. Ainsi, pour la suite de ce mémoire, tous les résultats présentés concerneront les graphes simples.

**Théorème 1.1.** Pour un graphe simple connexe, le nombre d'arêtes est restreint par $N - 1 \leq M \leq \binom{N}{2}$.

*Preuve.* La borne supérieure correspond au graphe complet, où tous les liens possibles existent. Ces liens sont au nombre de $\binom{N}{2}$ puisque le nombre de liens possible correspond au nombre de paires de noeuds distinctes.

La borne inférieure peut être prouvée par induction : un graphe simple à un seul sommet contient forcément 0 liens, alors que le graphe simple à deux sommets en contient 1. Pour ajouter un noeud et le connecter au reste du graphe, il faut nécessairement ajouter au moins 1 lien à $\mathcal{E}$. Ainsi, le nombre minimal de liens pour un graphe connexe de taille $N$ est $N - 1$. □

La borne inférieure $M = N - 1$ correspond à un type particulier de graphes, nommés *arbres*.

**Définition 1.6.** Un *arbre* est un graphe ne comportant aucun cycle, c'est-à-dire que la suite d'arêtes sans répétition reliant deux sommets est unique.

La propriété des arbres donnée plus haut, $M = N - 1$, peut également être vue comme une définition des arbres, puisqu'elle empêche l'existence de cycles dans le graphe.

**Théorème 1.2.** Un graphe de taille $N$ contenant $N - 1$ arêtes est un arbre.

*Preuve.* Soit un graphe de taille $N$ contenant $N - 1$ arêtes et au moins 1 cycle. Alors, il est possible de retirer un des liens du cycle sans déconnecter le graphe, de sorte que le graphe résultant soit connexe et contienne $N - 2$ arêtes. Or, puisqu'il a été établi que le nombre minimal d'arêtes dans un graphe connexe est $N - 1$, ceci est une contradiction, ce qui complète la preuve. □

Les graphes ont plusieurs propriétés structurelles importantes utilisées pour les caractériser. La définition de trois de celles-ci, qui seront cruciales pour la suite de ce mémoire, est donnée ici.

**Définition 1.7.** Le *degré* d'un noeud $v$, dénoté $k_v$, est le nombre de connexions de ce noeud. La *séquence de degrés* $\boldsymbol{k}$ est le vecteur contenant l'ensemble des degrés du réseau, et la *distribution* des degrés est $P(k) = \frac{N_k}{N}$, où $N_k$ est le nombre de noeuds de degré $k$. Par définition, $\sum_{v \in \mathcal{V}} k_v = 2M$.

La dernière égalité reflète le fait que chaque lien contribue au degré des deux noeuds qu'il connecte. Le degré d'un noeud est une propriété *locale*, puisqu'il est défini uniquement par un noeud et son voisinage immédiat. Toutefois, le degré moyen, $\overline{k} = \frac{1}{N} \sum_{\mathcal{V}} k_v = \frac{2M}{N}$, est une propriété *globale*, puisqu'il encode de l'information sur le réseau dans son entièreté. La prochaine propriété est également une propriété globale.

**Définition 1.8.** La *densité* d'un graphe est définie comme la proportion de liens possibles réalisés dans le graphe. Puisqu'il y a $\binom{N}{2}$ liens possibles, la densité est

$$d = \frac{M}{\binom{N}{2}} = \frac{2M}{N(N-1)} = \frac{\overline{k}}{N-1}. \tag{1.1}$$

**Définition 1.9.** Soit une séquence de matrices de taille N, $\{\mathbf{A}_N\}_{N=1}^{\infty}$. Si $d \to 0$ dans la limite $N \to \infty$, les graphes de la séquence sont dits *creux*[2], alors que si $d \to c$, où $c$ est une constante, les graphes sont dits *denses* [34].

La notion de graphe creux ou graphe dense peut être examinée du point de vue du degré moyen. Puisque $\overline{k} = d(N-1)$, dans une séquence de graphes denses, où $d$ tend vers une constante, le degré moyen croit linéairement avec $N$. Ainsi, un graphe est creux si le degré moyen évolue sous-linéairement avec $N$. Dans le cas particulier où le degré moyen est constant et indépendant de $N$, la densité est proportionnelle à $1/N$, auquel cas le graphe est dit *extrêmement creux*.

La définition des graphes creux et denses est fondamentalement vague : dans bien des cas, notamment dans l'étude des réseaux réels, il est impossible de prendre la limite $N \to \infty$, puisqu'il y a un seul graphe à l'étude et non pas une séquence de graphes. Dans ce cas, la définition donnée ci-dessus n'a pas de sens. La qualification d'un graphe comme creux ou dense revient alors à une classification arbitraire de sa densité comme faible ou élevée. Cependant, dans le cadre de ce mémoire, il sera presque toujours possible de prendre la limite $N \to \infty$, aussi la définition donnée convient-elle.

Il est à noter que les graphes extrêmement creux ne sont pas particulièrement exotiques. Un exemple est le réseau social de la population humaine, où chaque personne serait définie comme un noeud et où deux personnes sont connectées si elles se connaissent. L'accroissement de la population terrestre n'a pas nécessairement amené les personnes à avoir un plus grand réseau de connaissances. Ainsi, le réseau social humain planétaire correspondrait à la définition de réseau extrêmement creux donnée ci-haut.

2. En anglais : *sparse*

Les réseaux réels tendent aussi à exhiber une propriété dite « petit monde » [3], qui correspond à l'adage « les amis de mes amis sont mes amis ». On mesure cette propriété à l'aide du coefficient d'agglomération [4].

**Définition 1.10.** Le *coefficient d'agglomération local* $c_v$ d'un noeud $v$ est défini comme la proportion des paires de voisins de $v$ qui sont également connectées entre elles. Conséquemment, le *coefficient d'agglomération global* $C$ est défini comme la probabilité que deux noeuds soient connectés étant donné qu'ils partagent un voisin commun. Mathématiquement, cela correspond à la proportion de triplets fermés parmi les paires de voisins distincts d'un même noeud dans le graphe, soit

$$C = \frac{3N_\Delta}{N_\wedge},$$

où $N_\Delta$ est le nombre de triplets fermés et $N_\wedge$, le nombre de paires de noeuds ayant un voisin commun [5], et où le facteur 3 au numérateur est attribuable au fait que chaque triplet fermé contient trois paires de voisins distinctes.

Il est à noter que prendre la moyenne du coefficient d'agglomération local sur l'ensemble des noeuds ne mène pas en général au coefficient d'agglomération global, c'est-à-dire que $\frac{1}{N}\sum_{\mathcal{V}} c_v \neq C$. Une explication mathématique en sera donnée à la section suivante. La plupart des graphes réels présentent un coefficient d'agglomération plus grand que la probabilité globale d'observer un lien entre deux noeuds quelconques [22, 44], de sorte qu'il s'agit d'une propriété importante pour que des modèles de réseaux soient considérés comme réalistes.

## 1.2 Matrice d'adjacence

Il existe plusieurs représentations matricielles des graphes, chacune ayant ses avantages et inconvénients. Toutefois, une seule représentation concerne ce mémoire : la matrice d'adjacence.

**Définition 1.11.** La *matrice d'adjacence* $\mathbf{A}$ d'un graphe $\mathcal{G}$ est une matrice dont les colonnes et rangées correspondent aux sommets du graphe, de sorte que l'entrée $a_{ij}$ de la matrice est égale au poids de la connexion $j \to i$.

Puisqu'elle encode chacune des connexions du graphe, la matrice d'adjacence contient toute l'information disponible sur celui-ci. Aucune information n'est donc perdue en faisant appel à cette représentation matricielle.

Conformément à la définition d'un graphe, l'écriture d'une matrice d'adjacence spécifique attribue un étiquetage aux sommets du graphe. Or, cet étiquetage n'étant pas unique, il en

3. En anglais : *small-world*
4. En anglais : *clustering coefficient*
5. Une paire de noeuds peut être comptée plus d'une fois, dans le cas où les neouds de la paire partagent plus d'un voisin.

découle que deux matrices d'adjacence décrivent le même graphe si elles sont égales sous permutation des rangées et des colonnes. Mathématiquement, deux matrices $\mathbf{A}$ et $\mathbf{B}$ représentent le même graphe s'il existe une matrice de permutation $\mathbf{P}$ telle que $\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$.

Puisqu'elle contient toute l'information sur un graphe, les diverses propriétés des graphes énoncées à la section 1.1 se traduisent en termes de la matrice d'adjacence. La taille du graphe définit la taille de la matrice d'adjacence, qui est $N \times N$. La matrice d'adjacence est donc une matrice carrée par définition. La notion de graphe orienté correspond à la symétrie de la matrice d'adjacence : pour un graphe non orienté, $a_{ij} = a_{ji}$, de sorte que la matrice d'adjacence est symétrique, ce qui n'est en général pas le cas pour un graphe orienté. Pour un graphe non pondéré, la matrice est binaire, puisque ses entrées sont soit 1 si le lien $i \sim j$ existe et 0 sinon. Finalement, l'absence d'autoconnexions se traduit par une diagonale dont toutes les entrées sont nulles, $a_{ii} = 0 \ \forall i \in \{1, \ldots, N\}$.

En termes de la matrice d'adjacence, le degré du noeud $i$ s'écrit

$$k_i = \sum_{j=1}^{N} a_{ij} = \sum_{j=1}^{N} a_{ji} \tag{1.2}$$

Pour ce qui est du coefficient d'agglomération local, l'expression $a_{ij}a_{jk}a_{ki}$, $i \neq j \neq k$, est égale à 1 si et seulement si le trajet $i \to j \to k \to i$ existe dans le graphe, et à 0 autrement. Ainsi, $\sum_{j=1}^{N}\sum_{k=1}^{N} a_{ij}a_{jk}a_{ki}$ donne deux fois le nombre de triangles dont le noeud $i$ est un sommet, puisque chaque triangle est compté une fois dans chaque direction. Quant au nombre de paires de voisins du noeud $i$, il est donné par $\binom{k_i}{2}$, soit $\frac{k_i(k_i-1)}{2}$. Ainsi,

$$c_i = \frac{\sum_{j=1}^{N}\sum_{k=1}^{N} a_{ij}a_{jk}a_{ki}}{k_i(k_i-1)}, \tag{1.3}$$

avec $c_i = 0$ quand le dénominateur est nul. Pour obtenir le coefficient d'agglomération global, il suffit de sommer le numérateur et le dénominateur de $c_i$ sur l'ensemble des noeuds :

$$C = \frac{\sum_{i,j,k=1}^{N} a_{ij}a_{jk}a_{ki}}{\sum_{i=1}^{N} k_i(k_i-1)}, \tag{1.4}$$

où la notation compacte $\sum_{i,j,k=1}^{N} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}$ est introduite. La différence entre le coefficient d'agglomération global et le coefficient d'agglomération local moyen devient alors apparente : ce dernier est plutôt

$$\overline{c} = \frac{1}{N}\sum_{i=1}^{N} \frac{\sum_{j,k=1}^{N} a_{ij}a_{jk}a_{ki}}{k_i(k_i-1)}. \tag{1.5}$$

Une propriété clé de la matrice d'adjacence est que ses puissances permettent de compter les marches entre les noeuds du graphe.

**Définition 1.12.** Une *marche* est une séquence de noeuds $i \to k \to \ell \to \cdots \to j$ telle que deux noeuds consécutifs sont toujours connectés. Les transitions $i \to j$ sont appelées *pas*. Les noeuds peuvent apparaître à plusieurs reprises dans une marche. La *longueur* d'une marche est définie comme le nombre de pas qui la composent. Le noeud de départ et d'arrivée de la marche sont ses *extrémités*. Une marche dont les deux extrémités sont le même noeud est dite *fermée*.

Pour alléger la notation, les marches sont habituellement écrites sans le symbole $\to$, de sorte que la marche $i \to j \to i \to \ell$ serait simplement écrite $ijil$.

**Théorème 1.3** (Puissances de la matrice d'adjacence)**.** L'élément $[\mathbf{A}^n]_{ij}$ de la $n^e$ puissance de la matrice d'adjacence est égal au nombre de marches de longueur $n$ ayant comme extrémités les noeuds $i$ et $j$.

*Preuve.* Le cas $n = 1$ est trivial. Par définition, $a_{ij} = 1$ si les noeuds $i$ et $j$ sont connectés, de sorte qu'il existe une seule marche de longueur 1 entre les deux, et 0 autrement. Reste à démontrer que le cas $n + 1$ découle du cas $n$. Le nombre de marches de longueur $n + 1$ entre les noeuds $i$ et $j$ est égal à la somme du nombre de marches de longueur $n$ entre le noeud $i$ et les voisins du noeuds $j$, puisque le dernier pas doit mener à $j$. Par définition du produit matriciel,

$$[\mathbf{A}^{n+1}]_{ij} = [\mathbf{A}^n \mathbf{A}]_{ij} = \sum_{k=1}^{N} [\mathbf{A}^n]_{ik} a_{kj} = \sum_{k \sim j} [\mathbf{A}^n]_{ik}.$$

Or, par l'hypothèse d'induction, $[\mathbf{A}^n]_{ik}$ est égal au nombre de marches de longueur $n$ entre les noeuds $i$ et $k$. Puisqu'on somme cette valeur sur le voisinage du noeud $j$, l'expression ci-dessus est égale au nombre de marches de longueur $n + 1$ entre les noeuds $i$ et $j$, ce qui complète la preuve. □

**Corollaire 1.3.1.** Le nombre total de marches fermées de longueur $k$ dans un graphe est donné par $\mathrm{Tr}\big(\mathbf{A}^k\big)$.

*Preuve.* Par le théorème 1.3, $[\mathbf{A}^k]_{ii}$ est le nombre de marches fermées de longueur $k$ ayant le noeud $i$ comme extrémité. Or, par définition,

$$\mathrm{Tr}\Big(\mathbf{A}^k\Big) = \sum_{i=1}^{N} [\mathbf{A}^k]_{ii}.$$

□

**Remarque.** La définition et les résultats ci-dessus prennent pour acquis que le point de départ d'une marche fermée la définit. Ainsi, la marche $i \to j \to k \to i$ est comptée comme différente

de la marche $j \to k \to i \to j$. De même, les marches sont aussi différentiées selon leur direction. Ainsi, par exemple, le nombre total de triangles dans un graphe est donné par

$$N_\Delta = \frac{1}{6}\operatorname{Tr}\big(\mathbf{A}^3\big),$$

puisque chaque triangle est compté six fois : à partir de chaque sommet et dans chaque direction.

Il est possible réécrire les expressions 1.3 et 1.4 à l'aide des puissances de la matrice d'adjacence. En effet,

$$\sum_{j,k=1}^{N} a_{ij}a_{jk}a_{ki} = [\mathbf{A}^3]_{ii},$$

et puisque $\sum_{i=1}^{N} \frac{k_i(k_i-1)}{2}$ compte le nombre de marches ouvertes de longueur 2 dans le graphe,

$$\sum_{i=1}^{N} k_i(k_i-1) = \sum_{i,j=1}^{N} [\mathbf{A}^2]_{ij} - \operatorname{Tr}\big(\mathbf{A}^2\big).$$

Ainsi,

$$c_i = \frac{[\mathbf{A}^3]_{ii}}{k_i(k_i-1)}, \tag{1.6}$$

$$C = \frac{\operatorname{Tr}\big(\mathbf{A}^3\big)}{\sum_{i,j=1}^{N}[\mathbf{A}^2]_{ij} - \operatorname{Tr}\big(\mathbf{A}^2\big)}. \tag{1.7}$$

## 1.3 Valeurs propres et densité spectrale

Bien que la matrice d'adjacence contienne toute l'information sur un réseau, son étude peut être difficile pour des grands systèmes, puisqu'elle contient $N^2$ éléments. Il est donc souvent pratique de se rabattre sur certaines de ses propriétés, comme la séquence de degrés, définie à la section 1.1. Une autre propriété clé des matrices d'adjacence, et la principale à l'étude de ce mémoire, est leur spectre, soit l'ensemble de leurs valeurs propres.

**Définition 1.13.** Les *valeurs propres* $\lambda$ d'une matrice carrée $\mathbf{A}$ sont définies comme les solutions de l'équation

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}.$$

Les vecteurs $\mathbf{v}$ non nuls sont les *vecteur propres* de la matrice $\mathbf{A}$, et chaque valeur propre est associée à un vecteur propre différent. L'ensemble des valeurs propres d'une matrice est appelé son *spectre*, noté $\sigma(\mathbf{A})$.

**Remarque.** Il est à noter que la définition donnée ci-dessus est spécifiquement celle des vecteurs propres *droits*. Les vecteurs propres *gauches* sont définis de manière similaire,

$$\mathbf{u}\mathbf{A} = \lambda \mathbf{u}.$$

Bien que les vecteurs propres droits et gauches soient distincts en général, ils sont associés aux mêmes valeurs propres, de sorte que l'étude de celles-ci ne nécessite pas de considérer les deux types de vecteurs propres. La définition 1.13 sera donc utilisée sans ambiguïté dans l'ensemble de ce travail.

**Théorème 1.4.** Les valeurs propres d'une matrice réelle et symétrique sont réelles.

*Preuve.* Pour une matrice réelle et symétrique, $\mathbf{A} = \mathbf{A}^T = \mathbf{A}^*$, où $\mathbf{A}^T$ dénote la transposée de $\mathbf{A}$ et $\mathbf{A}^*$ sa matrice adjointe, soit la transposée de son conjugué complexe. Ainsi,

$$\begin{aligned}
\mathbf{v}^*\mathbf{A}\mathbf{v} = \mathbf{v}^*\lambda\mathbf{v} &= \lambda\mathbf{v}^*\mathbf{v} \\
\iff \lambda &= \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}} \\
\mathbf{v}^*\mathbf{A}^*\mathbf{v} = (\mathbf{A}\mathbf{v})^* &= (\lambda\mathbf{v})^*\mathbf{v} = \lambda^*\mathbf{v}^*\mathbf{v} \\
\iff \lambda^* &= \frac{\mathbf{v}^*\mathbf{A}^*\mathbf{v}}{\mathbf{v}^*\mathbf{v}} = \frac{\mathbf{v}^*\mathbf{A}\mathbf{v}}{\mathbf{v}^*\mathbf{v}} = \lambda.
\end{aligned}$$

□

Puisque la matrice d'adjacence d'un graphe simple est réelle et symétrique par définition, le spectre sera considéré comme réel, $\sigma(\mathbf{A}) \in \mathbb{R}$, pour la suite de ce mémoire.

**Définition 1.14.** La *densité spectrale empirique* d'une matrice est donnée par la distribution $\mu(z) = \frac{1}{N}\sum_{\lambda\in\sigma(\mathbf{A})}\delta(z-\lambda)$, où $\delta(z)$ est la distribution delta de Dirac. Une telle somme de deltas de Dirac est nommée *peigne* de deltas de Dirac.

**Définition 1.15.** Étant donné une séquence de matrices de taille $N$ $\{\mathbf{A}_N\}_{N=1}^{\infty}$ et la séquence de leurs densités spectrales empiriques $\{\mu_N\}_{N=1}^{\infty}$, leur *densité spectrale limite*, notée $\rho(z)$, est la limite faible de la séquence $\{\mu_N\}$ pour $N \to \infty$, si celle-ci existe, soit

$$\int_{\mathbb{R}} f(x)d\mu_N(x) \to \int_{\mathbb{R}} f(x)d\rho(x)$$

pour toute fonction $f$ continue et bornée sur $\mathbb{R}$.

Si la densité empirique prend la forme d'un peigne de deltas de Dirac, la densité limite peut être – et est souvent – une fonction continue. Puisque le spectre des matrices d'adjacence de graphes simples est réel, les densités spectrales seront traitées comme des fonctions réelles par la suite.

**Théorème 1.5.** Les moments de la densité spectrale empirique sont $\langle x^k \rangle_{\mu_N} := \int_{\mathbb{R}} x^k \mathrm{d}\mu_N(x) = \frac{1}{N} \operatorname{Tr}\left(\mathbf{A}^k\right)$.

*Preuve.* Par la définition de la distribution delta de Dirac,

$$\int_{\mathbb{R}} x^k \mathrm{d}\mu_N(x) = \frac{1}{N} \sum_{\lambda \in \sigma(\mathbf{A})} \lambda^k. \tag{1.8}$$

De plus, il est une propriété connue de la trace que

$$\operatorname{Tr}(\mathbf{A}) = \sum_{\lambda \in \sigma(\mathbf{A})} \lambda.$$

Or, les valeurs propres de $\mathbf{A}^k$ ne sont que les puissances correspondantes des valeurs propres de $\mathbf{A}$. En effet, le cas $k = 1$ étant la propriété connue, le résultat découle par induction mathématique. Avec l'hypothèse $\mathbf{A}^k \boldsymbol{v} = \lambda^k \boldsymbol{v}$,

$$\begin{aligned} \lambda^{k+1} \boldsymbol{v} &= \lambda \lambda^k \boldsymbol{v} \\ &= \lambda \mathbf{A}^k \boldsymbol{v} \\ &= \mathbf{A}^k \lambda \boldsymbol{v} \\ &= \mathbf{A}^k \mathbf{A} \boldsymbol{v} \\ &= \mathbf{A}^{k+1} \boldsymbol{v}. \end{aligned}$$

Ayant établi que les valeurs propres de $\mathbf{A}^k$ sont $\lambda^k \ \forall \ \lambda \in \sigma(\mathbf{A})$,

$$\operatorname{Tr}\left(\mathbf{A}^k\right) = \sum_{\lambda^k \in \sigma(\mathbf{A})} \lambda.$$

La correspondance avec l'équation 1.8 étant directe, ceci complète la preuve. □

Le théorème 1.5 admet une interprétation combinatoire capitale pour la suite du mémoire : par le corollaire 1.3.1, les moments d'ordre $k$ des densités spectrales empirique et limite sont le nombre moyen de marches fermées de longueur $k$ pour un noeud du graphe à l'étude. Cette interprétation change légèrement lorsque le graphe à l'étude est un graphe aléatoire, aussi convient-il de les introduire à l'instant.

## 1.4 Graphes aléatoires

L'étude de réseaux réels s'appuie souvent sur la comparaison de ceux-ci avec des réseaux aléatoires. En effet, ces derniers peuvent servir de modèles nuls permettant de détecter les propriétés distinctives des réseaux réels à l'étude [49] ou comme versions jouets de réseaux réels permettant d'obtenir des résultats analytiques [33]. Ainsi, caractériser adéquatement les différents modèles de réseaux aléatoires est important pour l'application des réseaux à l'étude de systèmes réels.

**Définition 1.16.** Un *graphe aléatoire* est une variable aléatoire dont les valeurs possibles appartiennent à un ensemble de graphes $\mathcal{G}$ auquel est associé une loi de probabilité. L'ensemble et sa loi de probabilité associée définissent un *modèle* de graphes aléatoires. Un élément spécifique de l'ensemble est appelée une *réalisation* du modèle.

La définition 1.16 est plus facilement comprise à l'aide d'exemples.

**Définition 1.17.** Le *modèle d'Erdős–Rényi* est défini par l'ensemble de graphes dont tous les éléments ont $N$ noeuds et la loi de probabilité associée

$$P(G \in \mathcal{G}) = p^M (1-p)^{\binom{N}{2}-M},$$

où $p \in [0,1]$ est une constante représentant la probabilité d'existence de chacun des liens.

Le modèle d'Erdős–Rényi est le modèle le plus simple qui soit : toutes les connexions ont exactement la même probabilité d'existence. Un autre modèle très simple est celui des graphes réguliers aléatoires.

**Définition 1.18.** Les graphes aléatoires *k-réguliers* sont définis par les ensembles de graphes dont tous les éléments ont $N$ noeuds ayant tous un degré exactement $k$, $3 \leq k \leq N-1$, avec tous les éléments de l'ensemble équiprobables.

Bien qu'ils ne soient pas en général un modèle réaliste de réseaux, la simplicité des graphes réguliers aléatoires en fait un outil de choix pour l'étude des propriétés des graphes aléatoires.

Un modèle autrement plus réaliste que les graphes réguliers – et que le modèle d'Erdős–Rényi – est le modèle des configurations.

**Définition 1.19.** Le *modèle des configurations* est défini par l'ensemble de graphes dont les éléments ont $N$ noeuds et une séquence de degrés fixe. Les éléments de l'ensemble sont équiprobables.

Lors de l'analyse de réseaux réels, utiliser comme modèle nul le modèle des configurations, avec la séquence de degré du réseau à l'étude, revient à établir comme hypothèse nulle que la séquence de degrés explique l'entièreté des propriétés du réseau. Ainsi, le modèle des configurations permet de détecter les propriétés du réseau qui découlent plutôt de corrélations entre les noeuds, par exemple le coefficient d'agglomération introduit à la définition 1.10. Le modèle des configurations peut être assoupli pour plutôt obéir à une distribution de degrés attendus plutôt que ferme. Cette variante est appelée *modèle de Chung-Lu*.

Il existe évidemment de nombreux autres modèles de réseaux présentant une variété de propriétés utiles pour l'étude des réseaux. Certains d'entre eux seront introduits plus tard dans ce mémoire lorsqu'ils s'avéreront pertinents pour le problème à l'étude.

Il existe une manière alternative de définir les graphes aléatoires dont la taille est fixée, qui consiste à voir chaque lien possible comme existant avec une probabilité propre. Dans cette vue, le modèle peut être représenté à l'aide d'une matrice d'adjacence aléatoire, dont les entrées hors diagonale sont des variables aléatoires de Bernoulli, soit

$$a_{ij} = a_{ji} \begin{cases} \sim \text{Bernoulli}(p_{ij}) & \text{si } i \neq j \\ 0 & \text{si } i = j, \end{cases} \tag{1.9}$$

où $p_{ij}$ est la probabilité que l'arête $(i, j)$ existe, soit $P((i, j) \in \mathcal{E}(\mathcal{G}))$.

Avec cette définition, le modèle d'Erdős–Rényi, avec $N$ fixe, prend la définition extrêmement simple de $p_{ij} = p \ \forall \ i \neq j$, alors que le modèle des configurations prend la forme $p_{ij} = \frac{k_i k_j}{2M}$, avec $k_i$ et $k_j$ les degrés des noeuds $i$ et $j$ [6], et le modèle de Chung-Lu, $p_{ij} = \frac{w_i w_j}{\sum_{v \in \mathcal{V}} w_v}$, où $w_i$ est le degré attendu du noeud $i$. La nuance distinguant les deux est que pour le modèle des configurations $k_i \in \mathbb{N}$, alors que pour le modèle de Chung-Lu, $w_i \in \mathbb{R}^+$.

Il n'est toutefois pas aussi évident d'appliquer cette définition aux graphes réguliers aléatoires, pour lesquels il n'existe pas de forme fermée pour $p_{ij}$. Cette limitation de la définition est fréquente pour les modèles de graphes où les probabilités de connexion sont corrélées, ou encore lorsque les graphes sont définis de manière essentiellement algorithmique. Néanmoins, c'est la conception de la matrice d'adjacence comme une matrice aléatoire qui est à retenir pour la suite, et celle-ci est toujours possible, même lorsqu'une forme explicite pour $p_{ij}$ n'est pas disponible.

Lorsque les graphes à l'étude sont des graphes aléatoires, les définitions de la section 1.3 doivent être légèrement modifiées. En effet, les valeurs propres d'un graphe aléatoire sont elles-mêmes des variables aléatoires, de sorte que la définition 1.15 change légèrement.

**Définition 1.20.** Étant donné une séquence de matrices aléatoires de taille $N$ $\{\mathbf{A}_N\}_{N=1}^{\infty}$ et la séquence de leurs densités spectrales empiriques $\{\mu_N\}_{N=1}^{\infty}$, leur *densité spectrale limite* est la limite faible, *en probabilité*, de la séquence $\{\mu_N\}$ pour $N \to \infty$, si celle-ci existe.

Or, si la densité spectrale limite existe, les densités spectrales empiriques des réalisations possibles du modèle devraient converger vers celle-ci avec probabilité 1, aussi la définition des moments de la densité spectrale limite n'a-t-elle pas à être modifiée. L'interprétation combinatoire des moments est toutefois légèrement différente. Pour des graphes aléatoires, les moments d'ordre $k$ de la densité limite sont égaux au nombre *attendu* de marches fermées de longueur $k$ pour un noeud du graphe, soit $\frac{1}{N} \sum_{G \in \mathcal{G}} P(G) \operatorname{Tr}\Big(\mathbf{A}(G)^k\Big) := \frac{1}{N} \left\langle \operatorname{Tr}\big(\mathbf{A}^k\big) \right\rangle$, où $\langle \cdot \rangle$ représente l'espérance mathématique. Reste à trouver une manière d'obtenir la densité spectrale à partir

6. Cette forme ne tient en fait que dans la limite $M \gg k_i \ \forall i$, puisque le modèle des configurations permet théoriquement plusieurs liens parallèles entre deux noeuds. Hors de cette limite, cette expression correspond plutôt au nombre attendu de liens entre $i$ et $j$.

de ses moments, ce pour quoi il convient d'introduire un nouvel outil mathématique : les séries génératrices.

## 1.5 Séries génératrices

**Définition 1.21.** Étant donnée une séquence de nombres $\{a_n\}_{n=0}^{\infty}$, sa *série génératrice ordinaire* est la série de puissance dont les coefficients sont les éléments de la séquence, soit

$$g_O(z) = \sum_{n=0}^{\infty} a_n z^n, \quad z \in \mathbb{C},$$

tandis que sa *série génératrice exponentielle* est

$$g_E(z) = \sum_{n=0}^{\infty} \frac{a_n}{n!} z^n.$$

Les séries génératrices sont un outil fondamental de la combinatoire, puisqu'elles permettent d'encapsuler pratiquement toute l'information sur une séquence de nombres en un seul objet analytique. Il est à noter qu'il s'agit de séries *formelles*, dont la convergence n'est pas nécessaire pour leur existence ni leur manipulation. Lorsque la série converge pour un certain rayon de convergence $R$ et qu'une forme fermée existe pour celle-ci, cette forme fermée est nommée *fonction génératrice*. Cette forme n'est une représentation valide de la série génératrice qu'à l'intérieur du domaine de convergence.

Une série génératrice particulière est celle dont les coefficients sont les moments d'une distribution $\rho$, soit

$$G_O(z) = \sum_{n=0}^{\infty} \langle x^n \rangle_\rho z^n$$
$$G_E(z) = \sum_{n=0}^{\infty} \frac{\langle x^n \rangle_\rho}{n!} z^n.$$

Il est possible de relier les séries génératrices des moments à la distribution associée à l'aide de deux relations clés :

$$\rho(x) = -\frac{1}{\pi} \lim_{\epsilon \to 0^+} \operatorname{Im}\left(\frac{1}{x+i\epsilon} G_O\left(\frac{1}{x+i\epsilon}\right)\right) \tag{1.10}$$

$$\rho(x) = \int_{-\infty}^{\infty} e^{2\pi i x s} G_E(-2\pi i s) \mathrm{d}s = \mathcal{F}^{-1}\{G_E(-2\pi i s)\}(x), \tag{1.11}$$

où $\mathcal{F}^{-1}$ dénote la transformée de Fourier inverse. Ces deux relations étant laborieuses à démontrer, et leurs démonstrations étant peu instructives pour la suite de ce mémoire, celles-ci sont présentées à l'annexe A.

Avec ces deux dernières relations, la feuille de route pour calculer la densité spectrale limite de modèles de réseaux aléatoires est établie :

1. Obtenir les moments de la densité spectrale en calculant le nombre attendu de marches fermées pour un noeud du graphe ;
2. Calculer la série génératrice des moments ;
3. Utiliser les relations 1.10 et 1.11 pour calculer la densité spectrale limite.

Évidemment, cette feuille de route est plus facile à établir qu'à exécuter. Néanmoins, il est maintenant possible de s'attaquer au calcul de la densité spectrale pour des modèles simples de réseaux.

# Chapitre 2

# Réseaux denses : la loi de Wigner

Pour réussir l'étape 1 de la feuille de route établie au chapitre 1, il convient de s'attarder d'abord à des modèles simples. C'est pourquoi l'analyse à venir portera en premier sur le modèle d'Erdős-Rényi.

## 2.1 Nombre attendu de marches fermées

Quelques éléments de notations doivent être introduits avant de procéder au calcul

**Définition 2.1.** L'ensemble des marches fermées dans le graphe $\mathcal{G}$ est noté $\mathcal{W}(\mathcal{G})$, où l'argument sera omis lorsque le contexte rend clair le graphe dont il est question. Conséquemment, le sous-ensemble des marches fermées de longueur $n$ est noté $\mathcal{W}_n(\mathcal{G}) \subset \mathcal{W}(\mathcal{G})$, et sa cardinalité, $|\mathcal{W}_n(\mathcal{G})| =: W_n(\mathcal{G})$. L'ensemble $\mathcal{W}(\mathcal{G})$, pour sa part, est toujours infini pour un graphe de taille $N \geq 2$, puisque les passages répétés sur une arrête sont permis.

Le nombre attendu de marches fermées de longueur $n$ pour un graphe aléatoire de matrice d'adjacence $\mathbf{A}$ peut alors s'écrire

$$\langle W_n \rangle = \langle \mathrm{Tr}(\mathbf{A}^n) \rangle = \sum_{w \in \mathcal{W}_n(\mathcal{K})} P(w \in \mathcal{W}_n), \tag{2.1}$$

où $\mathcal{K}$ dénote le graphe complet, soit le graphe où tous les liens sont présents. La matrice d'adjacence de $\mathcal{K}$ est donc $\mathbf{1} - \boldsymbol{I}$, où $\mathbf{1}$ est la matrice dont toutes les entrées sont égales à 1 et $\boldsymbol{I}$ est la matrice identité.

La probabilité qu'une marche quelconque existe, $P(w \in \mathcal{W})$, est simplement égale à la probabilité que chacun des liens qui la composent existe. Si les liens du réseaux sont indépendants,

cette probabilité est simplement

$$P(w \in \mathcal{W}) = \prod_{\epsilon \in \mathcal{E}(w)} P(\epsilon \in \mathcal{E}(\mathcal{G})),$$

où $\mathcal{E}(w)$ dénote l'ensemble des arêtes distinctes visitées par une la marche $w$.

C'est ici que la simplicité du modèle d'Erdős-Rényi intervient : pour ce modèle, $P(w \in \mathcal{W}) = p^\ell$, où $\ell = |\mathcal{E}(w)|$, soit le nombre d'arêtes distinctes visitées par la marche. Ainsi,

$$\langle W_n \rangle = \sum_{\ell=1}^{n} W_{n,\ell}(\mathcal{K}) p^\ell, \tag{2.2}$$

où $W_{n,\ell}$ est le nombre de marches fermées de longueur $n$ visitant $\ell$ arêtes distinctes. La clé est de réussir à calculer cette quantité. Celle-ci peut être décomposée en deux parties : le nombre de *formes* de marches et le nombre de manières d'étiqueter chacune des formes. Le concept de forme est mieux compris à l'aide d'un exemple.

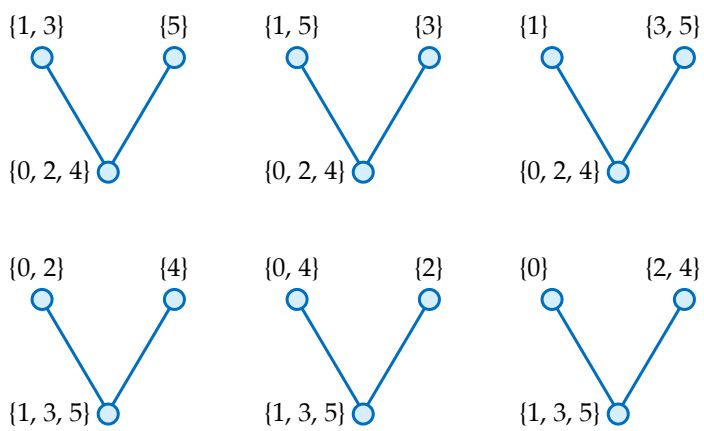



FIGURE 2.1 – Ensemble des 6 formes de marches de longueur 6 visitant 2 arêtes distinctes.

La figure 2.1 présente toutes les formes de marches de longueur 6 visitant 2 arêtes, où la marche est la suite de noeuds 0123450, et où les noeuds eux-mêmes ne sont pas étiquetés. Les formes de marches correspondent donc à toutes les façons distinctes de visiter l'ensemble des arêtes d'un graphe contenant $\ell$ arêtes au cours d'une marche fermée de longueur $n$.

Il est ensuite possible d'étiqueter chacune des formes en assignant un des noeuds du graphe à chacun des noeuds de la forme. Le noeud d'origine de la marche peut être n'importe lequel des $N$ noeuds du graphe, le premier noeud visité n'importe lequel des $N-1$ noeuds restants, et ainsi de suite. Il y a donc $N(N-1)(N-2)\cdots(N-\nu+1) := N^{\underline{\nu}}$ façons d'étiqueter les formes, avec $\nu$ le nombres de noeuds distincts visités par la forme. Puisque $N^{\underline{\nu}} = N^\nu + \mathcal{O}\big(N^{\nu-1}\big)$, dans la limite $N \to \infty$, $N^{\underline{\nu}} = N^\nu$.

Nommant $F_{n,\nu,\ell}(\mathcal{K})$ le nombre de formes de marches fermées de longueur $n$, visitant $\nu$ sommets et $\ell$ arêtes distincts sur le graphe complet, l'équation 2.2 devient donc

$$\langle W_n \rangle = \sum_{\nu=1}^{n} \sum_{\ell=\nu-1}^{n} F_{n,\nu,\ell}(\mathcal{K}) N^{\nu} p^{\ell}. \tag{2.3}$$

Si $p \ll 1$, et dans la limite $N \to \infty$, les termes maximisant $\nu$ et minimisant $\ell$ sont dominants dans l'équation 2.3. Or, par le théorème 1.1, $\max\{\nu - \ell\} = 1$, précisément dans le cas où la forme de la marche prend comme graphe sous-jacent un arbre. Ainsi, fixant $\nu = \ell + 1$,

$$\langle W_n \rangle = N \sum_{\ell=1}^{n} I_{n,\ell} \langle k \rangle^{\ell}, \tag{2.4}$$

où $I_{n,\ell}$ désigne maintenant le nombre de formes de marches fermées sur un arbre de longueur $n$ visitant $\ell$ arêtes distinctes, et avec $\langle k \rangle = Np$, le degré attendu d'un noeud du modèle d'Erdős-Rényi.

Les marches fermées sur un arbre ont une propriété particulière : elles ne peuvent être que de longueur paire. En effet, puisqu'il existe une unique marche entre deux noeuds sur un arbre, fermer une marche implique de repasser sur chacune des arêtes déjà visitées. Ainsi, le nombre de passages sur chaque arête est forcément pair, et par conséquence, la longueur totale de la marche également. C'est donc dire que $\langle W_{2n+1} \rangle = 0 \ \forall n \in \mathbb{N}$[1], soit que tous les moments impairs de la densité spectrale limite sont nuls.

Il est donc déjà possible, sans même poursuivre le calcul, de tirer une conclusion importante sur la densité spectrale limite du modèle d'Erdős-Rényi, soit que cette dernière est symétrique.

**Théorème 2.1.** Si tous les moments impairs d'une fonction de densité existent et sont nuls, alors la densité est symétrique autour de zéro.

*Preuve.* Par définition, les moments impairs d'une fonction de densité $\mu(x)$ sont

$$\left\langle x^{2n+1} \right\rangle_{\mu} = \int_{-\infty}^{\infty} x^{2n+1} \mu(x) \mathrm{d}x.$$

L'intégrale s'effectuant sur un domaine symétrique autour de zéro, $\left\langle x^{2n+1} \right\rangle_{\mu} = 0$ si et seulement si l'intégrande est soit nulle, soit une fonction impaire. Ainsi, puisque $x^{2n+1}$ est forcément non nulle et impaire, si $\exists x \in \mathbb{R}$ tel que $\mu(x) \neq 0$, l'intégrande est une fonction impaire si et seulement si $\mu(x)$ est une fonction paire. □

De plus, le fait que chaque arête visitée le soit au moins deux fois lors d'une marche fermée sur un arbre implique que le nombre maximal d'arêtes distinctes visitées par la marche est $\frac{n}{2}$.

1. Dans ce mémoire, l'ensemble des nombres entiers contient 0, conformément à la norme ISO 80000-2.

Cette borne supérieure correspond au cas où chaque arête est visitée précisément deux fois : un aller et un retour. Ainsi,

$$\langle W_n \rangle = \begin{cases} 0 & n \text{ impair} \\ N \sum_{\ell=1}^{n/2} I_{n,\ell} \langle k \rangle^{\ell} & n \text{ pair.} \end{cases} \tag{2.5}$$

Rappelant la définition 1.8, si le graphe est dense, $d \to c$ dans la limite $N \to \infty$. Or, puisque $d = \frac{\overline{k}}{N-1}$, le réseau est dense si $\overline{k}$ croit linéairement en $N$, ce qui correspond au cas où $p$ est une constante. Ainsi, pour un réseau dense, seul le terme $\ell = \frac{n}{2}$ de l'équation 2.5 est significatif dans la limite $N \to \infty$, de sorte que

$$\langle W_n \rangle \approx \begin{cases} 0 & n \text{ impair} \\ N I_{n,n/2} \langle k \rangle^{n/2} & n \text{ pair.} \end{cases} \tag{2.6}$$

Puisque chaque arête visitée par la marche l'est exactement deux fois, il existe une seule forme de marche de longueur $n$ par arbre contenant $n/2$ arêtes. De plus, la marche spécifie à l'arbre une racine – son noeud de départ et d'arrivée – ainsi qu'un ordre entre les descendants d'un noeud – l'ordre dans lequel ceux-ci sont visités par la marche. Ainsi, $I_{n,n/2}$ compte le nombre d'arbres enracinés et ordonnés contenant $n/2$ arêtes.

Cette quantité est bien connue en combinatoire, et est comptée par une célèbre séquence d'entiers : les nombres de Catalan [18, 24, 41], notés $C_n$ (OEIS A000108).

La première étape de la feuille de route établie au chapitre 1 est alors complétée : les moments de la densité spectrale limite sont

$$\langle x^n \rangle_\rho = \frac{1}{N} \langle W_n \rangle \approx \begin{cases} 0 & n \text{ impair} \\ C_{n/2} \langle k \rangle^{n/2} & n \text{ pair.} \end{cases} \tag{2.7}$$

## 2.2 Série génératrice des moments

La deuxième étape de la feuille de route est de calculer une des séries génératrices des moments de la distribution spectrale limite, soit les sommes

$$G_O(z) = \sum_{n=0}^{\infty} \langle x^n \rangle_\rho z^n,$$
$$G_E(z) = \sum_{n=0}^{\infty} \frac{\langle x^n \rangle_\rho}{n!} z^n.$$

Bien qu'une seule des deux serait nécessaire pour compléter le calcul, les deux seront présentées dans ce chapitre à des fins démonstratives. Avec l'équation 2.7, et le changement d'indice

$n \to 2n$ puisque seuls les moments pairs sont non nuls,

$$G_O(z) = \sum_{n=0}^{\infty} C_n \langle k \rangle^n z^{2n} = \sum_{n=0}^{\infty} C_n (\langle k \rangle z^2)^n,$$
$$G_E(z) = \sum_{n=0}^{\infty} \frac{C_n \langle k \rangle^n}{(2n)!} z^{2n} = \sum_{n=0}^{\infty} \frac{C_n}{(2n)!} (\sqrt{\langle k \rangle} z)^{2n}.$$

La série génératrice ordinaire des moments correspond ainsi simplement à la série génératrice ordinaire des nombres de Catalan, avec comme variable $\langle k \rangle z^2$. Bien la fonction génératrice ordinaire des nombres de Catalan soit un résultat classique, elle est simple à démontrer, aussi la preuve en sera-t-elle faite.

**Théorème 2.2.** Sachant que les nombres de Catalan sont définis par $C_0 \equiv 1$ et la récurrence [24]

$$C_{n+1} = \sum_{i=0}^{n} C_i C_{n-i}, \quad n \geq 1,$$

leur fonction génératrice ordinaire est

$$C(z) = \frac{1 - \sqrt{1 - 4z}}{2z}.$$

*Preuve.* Multipliant des deux côtés l'équation de récurrence par $z^n$, puis sommant sur $n$ de 0 à l'infini,

$$\sum_{n=0}^{\infty} C_{n+1} z^n = \sum_{n=0}^{\infty} \sum_{i=0}^{n} C_i C_{n-i} z^n.$$

Le côté gauche peut être aisément ramené à $C(z)$ :

$$\sum_{n=0}^{\infty} C_{n+1} z^n = \sum_{n=1}^{\infty} C_n z^{n-1} = \frac{1}{z} \sum_{n=1}^{\infty} C_n z^n = \frac{1}{z}(C(z) - C_0) = \frac{1}{z}(C(z) - 1).$$

Quand au côté droit, il correspond au produit de Cauchy de deux séries infinies, soit

$$\sum_{n=0}^{\infty} \sum_{i=0}^{n} C_i C_{n-i} z^n = \left( \sum_{n=0} C_n z^n \right) \left( \sum_{m=0} C_m z^m \right) = C(z)^2.$$

Ainsi,

$$\frac{1}{z}(C(z) - 1) = C(z)^2$$
$$\iff C(z) = \frac{1 \pm \sqrt{1 - 4z}}{2z}.$$

Puisque la définition en série de puissance de la fonction génératrice impose $C(0) = C_0 = 1$, seule la racine négative est une représentation fonctionnelle correcte de la série génératrice. En effet,

$$\lim_{z\to 0} \frac{1-\sqrt{1-4z}}{2z} = 1,$$

alors que

$$\lim_{z\to 0^+} \frac{1+\sqrt{1-4z}}{2z} = \infty, \lim_{z\to 0^-} \frac{1+\sqrt{1-4z}}{2z} = -\infty.$$

Ainsi,

$$C(z) = \frac{1-\sqrt{1-4z}}{2z}.$$

□

La fonction génératrice ordinaire des nombres de Catalan permet d'obtenir une forme fermée pour ceux-ci. En effet, par le théorème du binôme généralisé,

$$\sqrt{1+x} = \sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n} x^n = \sum_{n=0}^{\infty} \frac{(-1)^{n-1}}{4^n(2n-1)} \binom{2n}{n} x^n,$$

de sorte que

$$\sqrt{1-4z} = \sum_{n=0}^{\infty} \frac{-1}{(2n-1)} \binom{2n}{n} z^n = 1 - \sum_{n=1}^{\infty} \frac{1}{(2n-1)} \binom{2n}{n} z^n = 1 - 2\sum_{n=1}^{\infty} \frac{1}{n} \binom{2n-2}{n-1} z^n.$$

Substituant cette forme dans $C(z)$,

$$C(z) = \sum_{n=1}^{\infty} \frac{1}{n} \binom{2n-2}{n-1} z^{n-1} = \sum_{n=0}^{\infty} \frac{1}{n+1} \binom{2n}{n} z^n,$$

ce qui implique

$$C_n = \frac{1}{n+1} \binom{2n}{n}. \tag{2.8}$$

Avec cette formule, il est possible de prouver le théorème suivant :

**Théorème 2.3.** La série génératrice exponentielle paire des nombres de Catalan est

$$C_E(z) := \sum_{n=0}^{\infty} \frac{C_n}{(2n)!} z^{2n} = \frac{I_1(2z)}{z},$$

où $I_1(z)$ est la fonction de Bessel modifiée du premier type, définie en série comme

$$I_1(z) = \sum_{n=0}^{\infty} \frac{1}{n!(n+1)!} \left(\frac{z}{2}\right)^{2n+1}.$$

*Preuve.* Substituant l'équation 2.8 dans $C_E(z)$,

$$\begin{aligned} C_E(z) &= \sum_{n=0}^{\infty} \frac{1}{(n+1)(2n)!}\binom{2n}{n} z^{2n} = \sum_{n=0}^{\infty} \frac{1}{(n+1)(2n)!}\frac{(2n)!}{n!n!} z^{2n} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!(n+1)!} z^{2n} = \frac{1}{z}\sum_{n=0}^{\infty} \frac{1}{n!(n+1)!}\left(\frac{2z}{2}\right)^{2n+1}. \end{aligned}$$

La série infinie dans la dernière équation correspondant à la fonction de Bessel modifiée du premier type,

$$C_E(z) = \frac{I_1(2z)}{z}.$$

□

Par les théorèmes 2.2 et 2.3,

$$G_O(z) = \frac{1-\sqrt{1-4\langle k\rangle z^2}}{2\langle k\rangle z^2} \tag{2.9}$$

$$G_E(z) = \frac{I_1(2\sqrt{\langle k\rangle}z)}{\sqrt{\langle k\rangle}z}, \tag{2.10}$$

ce qui complète la deuxième étape de la feuille de route.

## 2.3 Densité spectrale limite

Reste maintenant à appliquer les relations 1.10 et 1.11 aux fonctions génératrices 2.9 et 2.10, respectivement. Ainsi, pour la fonction génératrice ordinaire,

$$\begin{aligned} \rho(x) &= -\frac{1}{\pi}\lim_{\epsilon\to 0^+} \operatorname{Im}\left(\frac{1}{x+i\epsilon}\frac{1-\sqrt{1-4\langle k\rangle\left(\frac{1}{x+i\epsilon}\right)^2}}{2\langle k\rangle\left(\frac{1}{x+i\epsilon}\right)^2}\right) \\ &= -\frac{1}{\pi}\lim_{\epsilon\to 0^+} \operatorname{Im}\left(\frac{(x+i\epsilon)-\sqrt{(x+i\epsilon)^2-4\langle k\rangle}}{2\langle k\rangle}\right) \\ &= \frac{-1}{2\pi\langle k\rangle}\lim_{\epsilon\to 0^+}\left[\epsilon - \operatorname{Im}\left(\sqrt{(x+i\epsilon)^2-4\langle k\rangle}\right)\right] \\ &= \frac{1}{2\pi\langle k\rangle}\lim_{\epsilon\to 0^+}\operatorname{Im}\left(\sqrt{(x+i\epsilon)^2-4\langle k\rangle}\right). \end{aligned}$$

Pour prendre la racine carrée, il est nécessaire de réécrire son argument sous forme polaire, soit

$$(x+i\epsilon)^2 - 4\langle k\rangle = re^{i\varphi},$$

avec

$$r = \sqrt{(x^2 - \epsilon^2 - 4\langle k\rangle)^2 + 4x^2\epsilon^2}$$
$$\varphi = \arg((x + i\epsilon)^2 - 4\langle k\rangle),$$

où la fonction $\arg(x+iy)$ dénote l'angle entre l'axe des réels positifs et la ligne reliant l'origine à $x + iy$, prenant une valeur dans l'intervalle $(-\pi, \pi]$. Sans ambiguïté, cette valeur est égale à la fonction $\operatorname{atan2}(x + iy)$, définie comme

$$\arg(x+iy) = \operatorname{atan2}(y,x) = \begin{cases} \arctan\left(\frac{y}{x}\right) & \text{si } x > 0, \\ \arctan\left(\frac{y}{x}\right) + \pi & \text{si } x < 0 \text{ et } y \geq 0, \\ \arctan\left(\frac{y}{x}\right) - \pi & \text{si } x < 0 \text{ et } y < 0, \\ +\frac{\pi}{2} & \text{si } x = 0 \text{ et } y > 0, \\ -\frac{\pi}{2} & \text{si } x = 0 \text{ et } y < 0, \\ \text{non défini} & \text{si } x = 0 \text{ et } y = 0. \end{cases}$$

Ainsi,

$$\arg((x+i\epsilon)^2 - 4\langle k\rangle) = \begin{cases} \arctan\left(\frac{2x\epsilon}{x^2-\epsilon^2-4\langle k\rangle}\right) & \text{si } x^2 > 4\langle k\rangle + \epsilon^2, \\ \arctan\left(\frac{2x\epsilon}{x^2-\epsilon^2-4\langle k\rangle}\right) + \pi & \text{si } x^2 < 4\langle k\rangle + \epsilon^2 \text{ et } x \geq 0, \\ \arctan\left(\frac{2x\epsilon}{x^2-\epsilon^2-4\langle k\rangle}\right) - \pi & \text{si } x^2 < 4\langle k\rangle + \epsilon^2 \text{ et } x < 0. \end{cases}$$

La partie imaginaire de la racine carrée est donc

$$\operatorname{Im}\left(\sqrt{(x+i\epsilon)^2 - 4\langle k\rangle}\right) = \sqrt{r}\sin(\varphi/2).$$

Prenant la limite,

$$\lim_{\epsilon\to 0^+} r = \sqrt{(x^2 - 4\langle k\rangle)^2} = \pm(x^2 - 4\langle k\rangle).$$

Puisque $\operatorname{Im}(z)$ est un nombre réel par définition, $\sqrt{r}$ doit être réel, c'est donc dire que $r$ doit être un nombre non négatif. Ainsi,

$$\lim_{\epsilon\to 0^+} r = \begin{cases} x^2 - 4\langle k\rangle & \text{si } x^2 > 4\langle x\rangle \\ 4\langle k\rangle - x^2 & \text{si } x^2 \leq 4\langle x\rangle. \end{cases}$$

Pour ce qui est de $\varphi$,

$$\lim_{\epsilon\to 0^+} \varphi = \begin{cases} 0 & \text{si } x^2 > 4\langle k\rangle \\ \pi & \text{si } x^2 < 4\langle k\rangle \text{ et } x \geq 0 \\ -\pi & \text{si } x^2 < 4\langle k\rangle \text{ et } x < 0. \end{cases}$$

Ainsi,

$$\rho(x) = \begin{cases} 0 & \text{si } x^2 > 4\langle k\rangle\,, \\ \pm\frac{\sqrt{4\langle k\rangle - x^2}}{2\pi\langle k\rangle} & \text{si } x^2 < 4\langle k\rangle \text{ et } x > 0, \\ \mp\frac{\sqrt{4\langle k\rangle - x^2}}{2\pi\langle k\rangle} & \text{si } x^2 < 4\langle k\rangle \text{ et } x < 0. \end{cases}$$

Puisque $\rho(x)$ est une distribution, par définition, $\rho(x) \geq 0\ \forall x \in \mathbb{R}$. Le signe de $r$ est donc sélectionné pour respecter cette contrainte, de sorte que $\rho(x)$ s'écrit sous forme compacte :

$$\rho(x) = \frac{\sqrt{4\langle k\rangle - x^2}}{2\pi\langle k\rangle}\mathbb{1}_{|x|<2\sqrt{\langle k\rangle}}, \tag{2.11}$$

où $\mathbb{1}_x$ dénote la fonction indicatrice. Puisque l'équation 2.11 est une forme fermée pour la densité spectrale limite, la feuille de route établie au chapitre 1 est complétée. Toutefois, à des fins démonstratives, le calcul à partir de la fonction génératrice exponentielle sera également présenté. Par l'équation 1.11 et le théorème d'inversion de Fourier,

$$G_E(-2\pi i s) = \int_{-\infty}^{\infty} e^{-2\pi i x s}\rho(x)\mathrm{d}x = -\frac{I_1(-4\pi i\sqrt{\langle k\rangle}s)}{2\pi i\sqrt{\langle k\rangle}s}. \tag{2.12}$$

La fonction $I_1$ admet la définition intégrale suivante :

$$I_1(z) = \frac{z}{\pi}\int_{-1}^{1}\sqrt{1-t^2}e^{zt}\mathrm{d}t.$$

Ainsi,

$$\begin{aligned} \frac{I_1(2z)}{z} &= \frac{2}{\pi}\int_{-1}^{1}\sqrt{1-t^2}e^{2zt}\mathrm{d}t \\ \iff -\frac{I_1(-4\pi i\sqrt{\langle k\rangle}s)}{2\pi i\sqrt{\langle k\rangle}s} &= \frac{2}{\pi}\int_{-1}^{1}\sqrt{1-t^2}e^{-4\pi i\sqrt{\langle k\rangle}st}\mathrm{d}t. \end{aligned}$$

Avec la substitution $x = 2\sqrt{\langle k\rangle}t$, $\mathrm{d}t = \frac{1}{2\sqrt{\langle k\rangle}}\mathrm{d}x$ et

$$\begin{aligned} G_E(-2\pi i s) &= \frac{1}{\pi\sqrt{\langle k\rangle}}\int_{-2\sqrt{\langle k\rangle}}^{2\sqrt{\langle k\rangle}}\sqrt{1-\frac{x^2}{4\langle k\rangle}}e^{-2\pi i x s}\mathrm{d}x \\ &= \int_{-\infty}^{\infty}\frac{\sqrt{4\langle k\rangle - x^2}}{2\pi\langle k\rangle}\mathbb{1}_{|x|<2\sqrt{\langle k\rangle}}e^{-2\pi i x s}\mathrm{d}x. \end{aligned}$$

En comparant l'équation ci-dessus avec l'équation 2.12, il est possible de conclure que

$$\rho(x) = \frac{\sqrt{4\langle k\rangle - x^2}}{2\pi\langle k\rangle}\mathbb{1}_{|x|<2\sqrt{\langle k\rangle}},$$

ce qui concorde avec l'équation 2.11.

La densité obtenue à l'équation 2.11 est un résultat classique, et n'est en fait qu'une version modifiée de la loi du demi-cercle de Wigner [21, 32, 46]. La version exacte de cette loi implique

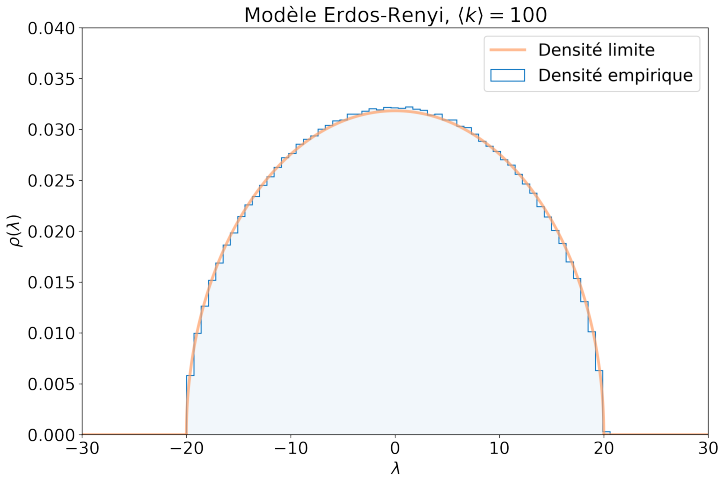



FIGURE 2.2 – Densité spectrale du modèle d'Erdős-Rényi dense. La densité empirique est un histogramme des valeurs propres de 3 réseaux de 10 000 noeuds générés avec le modèle d'Erdős-Rényi et $p = 0.01$. Le trait orange correspond à l'équation 2.11. Tel qu'attendu, la densité spectrale est symétrique et confinée au domaine $x \in [-2\sqrt{\langle k \rangle}, 2\sqrt{\langle k \rangle}]$.

une correction remplaçant $\langle k \rangle$ par $(1-p)\langle k \rangle$ [47], imputable à l'approximation $p \ll 1$ effectuée pour obtenir l'équation 2.4. Dans ce régime, la correction est insignifiante puisque $1 - p \approx 1$, mais pour des grandes valeurs de $p$, elle devient nécessaire, tel qu'illustré à la figure 2.3. Le calcul de cette correction est présenté à l'annexe B.

Puisque la correspondance entre nombre de marches fermées attendu et moments de la densité spectrale reste absolument valide, il devrait être possible de donner une interprétation du facteur de correction $(1-p)$ en termes de marches fermées. Or, en dépit de ses efforts, l'auteur est incapable d'en présenter une dans ces lignes.

## 2.4 Généralisation et limites du résultat

Bien que le calcul présenté jusqu'ici ait été effectué pour le modèle d'Erdős-Rényi, il peut facilement être transposé à d'autres modèles de réseaux. En effet, le modèle n'apparait dans le calcul que lorsque $P(w \in \mathcal{W}) = p^\ell$ est posé pour obtenir l'équation 2.2. Or, cette probabilité d'existence des marches, avec $p$ remplacé par la probabilité de connexion moyenne du modèle en question $\langle p_{ij} \rangle$, est une bonne approximation pour tout modèle respectant les deux conditions suivantes :

1. Les liens du réseau sont indépendants, c'est-à-dire que les entrées de la matrice d'adjacence aléatoire sont des variables aléatoires indépendamment distribuées ;
2. La probabilité de connexion des différents liens est bien approximée par la moyenne, c'est-à-dire que la variance de la distribution des probabilités de connexion $p_{ij}$ est faible.

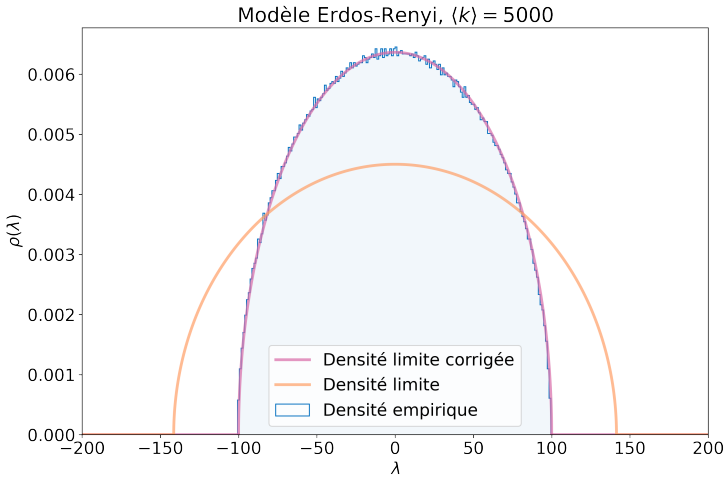



FIGURE 2.3 – Densité spectrale du modèle d'Erdős-Rényi très dense, tel que l'hypothèse $p \ll 1$ ne tient pas. La densité empirique est un histogramme des valeurs propres de 3 réseaux de 10 000 noeuds générés avec le modèle d'Erdős-Rényi et $p = 0.5$. Le trait orange correspond à l'équation 2.11, alors que le trait violet correspond à la version corrigée remplaçant $\langle k \rangle$ par $(1-p)\langle k \rangle$. La correction est significative, ce qui montre que l'hypothèse $p \ll 1$ est essentielle pour que l'équation 2.11 soit valide.

Les réseaux aléatoires respectant la première condition sont dits non corrélés, tandis que ceux respectant la deuxième sont dits homogènes. Dans le cas d'un réseau non corrélé, l'homogénéité du réseau peut également être lue à partir de sa distribution de degrés, auquel cas le réseau est dit homogène si cette dernière a une faible variance. Plus précisément, Metz & Silva ont établi dans [30] qu'avec $\sigma_k^2$, la variance de la distribution de degrés d'un modèle des configurations,

$$\lim_{\langle k \rangle \to \infty} \frac{\sigma_k^2}{\langle k \rangle^2} = 0$$

est une condition nécessaire pour la validité de la loi du demi-cercle. Autrement dit, la distribution de degrés doit devenir concentrée autour du degré moyen pour obtenir la loi du demi-cercle.

Les deux exemples donnés à la figure 2.4 illustrent bien que des modèles de réseaux fondamentalement différents du modèle d'Erdős-Rényi peuvent présenter une densité spectrale bien décrite par le demi-cercle lorsque les réseaux sont suffisamment denses, non corrélés ou homogènes. En effet, le modèle de Chung-Lu n'est pas corrélé par définition, et le choix de la distribution des degrés attendus permet de choisir le degré d'homogénéité du graphe. Quant aux graphes réguliers aléatoires, ils sont par définition homogènes, mais corrélés puisque les degrés ne présentent aucun degré de liberté. Toutefois, lorsque $1 \ll k \ll N$, ces corrélations sont peu significatives. Bauerschmidt, Knowles & Yau [5] ont en effet prouvé que pour $k \geq (\log N)^4$, les graphes réguliers aléatoires obéissent à la loi du demi-cercle.

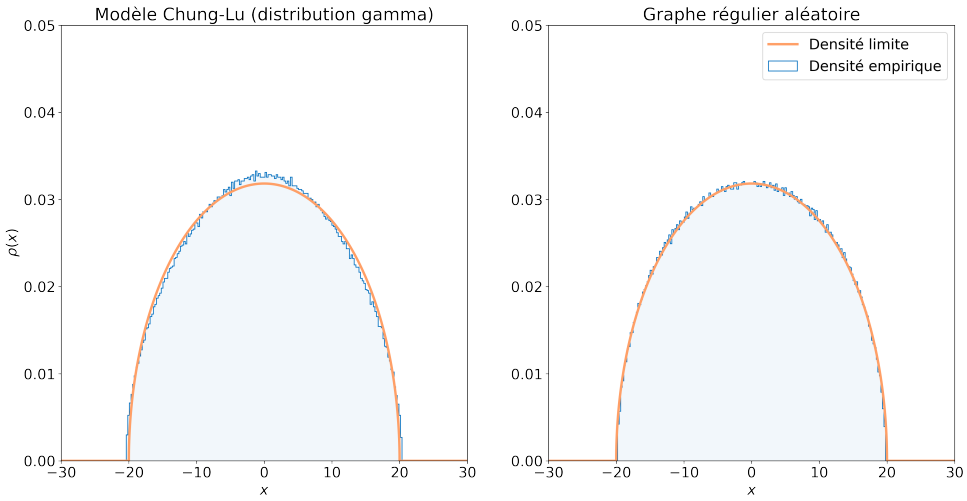



FIGURE 2.4 – Histogramme des densités spectrales empiriques de trois réseaux de 10 000 noeuds avec degré moyen $\langle k \rangle = 100$. Les réseaux de la figure de gauche sont générés avec le modèle de Chung-Lu avec des degrés attendus générés selon une distribution gamma dont les paramètres sont $\alpha = 50$, $\beta = 1/2$. La variance de la distribution de degrés est donc $\mathrm{Var}(k) = \alpha/\beta^2 = 200$. Ceux de la figure de droite sont générés par le modèle de graphes réguliers aléatoires avec $k = 100$. Le trait orange représente la loi du demi-cercle donnée par l'équation 2.11.

La figure 2.5 présente le cas contraire. Le modèle de Chung-Lu y est toujours présenté avec une distribution gamma, mais dont la variance est cinq fois plus grande qu'à la figure 2.4. Les graphes générés par ce modèle ne sont donc plus suffisamment homogène pour que la loi du demi-cercle soit valide. Quant au modèle de Watts-Strogatz, il s'agit d'un modèle où les noeuds sont d'abord placés sur le graphe cycle, puis connectés à leurs $k/2$ voisins les plus proches de chaque côté, avec $k$ le degré moyen désiré. Les liens sont ensuite reconnectés au hasard avec une probabilité $\beta$. Ainsi, le cas $\beta = 0$ correspond à un graphe régulier, alors que le cas $\beta = 1$ correspond au modèle d'Erdős-Rényi. Les valeurs intermédiaires de $\beta$ résultent donc en des réseaux présentant une distribution de degré homogène, mais également des corrélations sous la forme d'un coefficient d'agglomération élevé [44]. La présence de ces corrélations est traduite par l'asymétrie de la distribution spectrale empirique.

Les résultats numériques présentés ci-dessus pointent vers l'existence d'une classe d'universalité de graphes pour lesquels la densité limite correspond à la loi du demi-cercle de Wigner. Il existe une littérature riche sur cette classe d'universalité, remontant aux travaux pionniers de Wigner [45], Dyson [12] et Mehta [29], entre autres. Des travaux plus récents menés notamment par Tao & Vu [42, 43], Erdős et. al [13, 14], Aggarwal [1] et Huang et. al [23] ont permis de raffiner grandement les conditions d'appartenance à cette classe d'universalité. En particulier, il a été montré dans [23] que, pour le modèle d'Erdős-Renyi, la condition $\langle k \rangle \geq N^{\epsilon}$ est suffisante pour

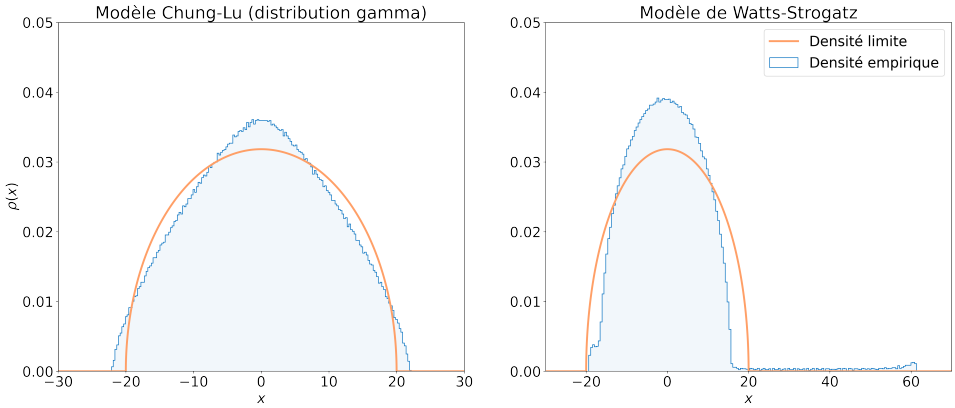



FIGURE 2.5 – Histogramme des densités spectrales empiriques de 3 réseaux de 10 000 noeuds avec degré moyen $\langle k \rangle = 100$. Ceux de la figure de gauche sont générés grâce au modèle de Chung-Lu avec les degrés attendus sont tirés d'une distribution gamma avec paramètres $\alpha = 10$, $\beta = 1/10$. La variance de la distribution de degrés est donc $\mathrm{Var}(k) = \alpha/\beta^2 = 1000$. Les réseaux de la figure de droite sont générés par le modèle de Watts-Strogatz avec $\beta = 0.4$. Le trait orange représente la loi du demi-cercle obtenue à l'équation 2.11.

que la loi du demi-cercle soit valide. Ainsi, même le spectre de graphes pour lesquels $d \to 0$, c'est-à-dire des graphes creux, obéit à la loi du demi-cercle.

Il subsiste toutefois une classe de réseaux qui ne respecte pas la condition de densité pour la loi du demi-cercle. Cette classe, qui est au coeur de ce mémoire, est celle des réseaux extrêmement creux. C'est sur cette classe que se penche le prochain chapitre.

# Chapitre 3

# Le cas des réseaux extrêmement creux

Dans le chapitre 2, il a été montré que le cas des réseaux denses, homogènes et non corrélés est assez simple, les résultats obtenus pointant même vers une classe d'universalité pour laquelle la densité spectrale limite des réseaux est la même malgré leurs différentes propriétés locales. Toutefois, pour les réseaux extrêmement creux, cette classe d'universalité se désintègre : la structure locale des réseaux devient importante pour le calcul de leur densité spectrale limite. Il sera notamment montré dans ce chapitre que les densités limites des graphes réguliers et Erdős-Rényi, qui appartiennent tous deux à la classe d'universalité dans le cas dense, deviennent totalement différentes dans le cas extrêmement creux. Le chapitre se concentre en premier lieu sur le cas du modèle d'Erdős-Rényi.

## 3.1 Observations numériques

La figure 3.1 montre la densité spectrale de réseaux Erdős-Rényi creux. La première observation tirée de ces densités est que, sans surprise, celles-ci semblent tendre vers la loi du demi-cercle lorsque $\langle k \rangle$ augmente. Il devient toutefois apparent que celle-ci ne tient plus pour des réseaux extrêmement creux.

Dans le régime extrêmement creux, des discontinuités apparaissent dans le spectre, qui est alors une combinaison de valeurs discrètes et d'une composante continue. Il est possible d'observer que les discontinuités les plus importantes se trouvent à $x = 0$, $x = \pm 1$ et $x = \pm\sqrt{2}$. Ces valeurs correspondent aux valeurs propres de petits arbres, présentées au tableau 3.1. La figure 3.2 montre bien la correspondance entre celles-ci et les valeurs discrètes de la densité spectrale.

La raison de cette correspondance est simple : il est connu que le modèle d'Erdős-Rényi présente une transition de phase à $\langle k \rangle = 1$, où apparait une composante géante comprenant une fraction finie d'ordre $N^{2/3}$ des noeuds du réseau. Les autres noeuds appartiennent à des composantes finies, dont la plupart sont des arbres. Étant donné un arbre à $n$ sommets $\mathcal{T}_n$, le nombre de composantes finies isomorphes à $\mathcal{T}_n$ est $\frac{N}{A(\mathcal{T}_n)} \langle k \rangle^{n-1} e^{-n\langle k \rangle} + o(N)$, où $A(\mathcal{T}_n)$ est la

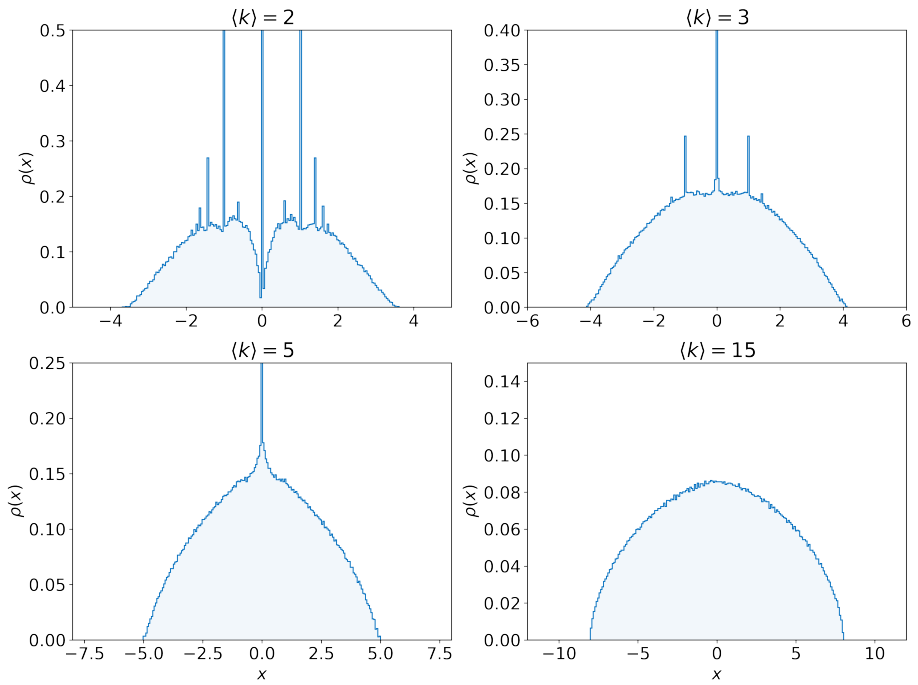



FIGURE 3.1 – Densités spectrales empiriques pour quatre ensembles de trois graphes de 10000 noeuds générés avec le modèle d'Erdős-Rényi et $\langle k \rangle = 2, 3, 5$ et 15.

cardinalité du groupe d'automorphismes de $\mathcal{T}_n$, soit le nombre de permutations possibles de ses sommets qui ne changent pas sa matrice d'adjacence [4, 7, 16]. La décroissance exponentielle en $n$ indique que les arbres avec le moins de sommets sont les plus fréquents, et contribuent donc plus de valeurs propres au spectre du graphe. Ceci explique la prédominance de leurs valeurs propres dans la composante discrète du spectre.

De plus, M. Bauer et O. Golinelli montrent dans [4] que les arbres reliés par une seule arête à la composante géante contribuent également leurs valeurs propres au spectre discret, accentuant ainsi leur présence dans le spectre. En théorie, il existe toujours ces discontinuités dans le spectre, mais lorsque $\langle k \rangle$ devient grand, leur poids est insuffisant pour apparaître dans l'histogramme de la densité spectrale empirique de réalisations du modèles.

Quant à la composante continue, elle présente notamment des queues beaucoup plus importantes que la distribution du demi-cercle. Il n'est plus aussi évident que la densité soit constrainte à un sous-domaine compact de $\mathbb{R}$. Les calculs analytiques à suivre confirmeront cette impression.

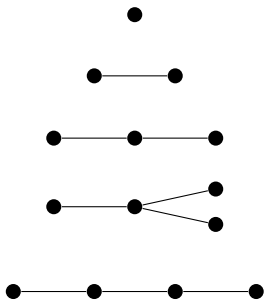
| Arbre | Valeurs propres |
|---|---|
|  | 0 |
| | -1, 1 |
| | $-\sqrt{2}$, 0, $\sqrt{2}$ |
| | $-\sqrt{3}$, 0, 0, $\sqrt{3}$ |
| | $\frac{-1-\sqrt{5}}{2}$, $\frac{1-\sqrt{5}}{2}$, $\frac{-1+\sqrt{5}}{2}$, $\frac{1+\sqrt{5}}{2}$ |

TABLE 3.1 – Valeurs propres de divers petits arbres

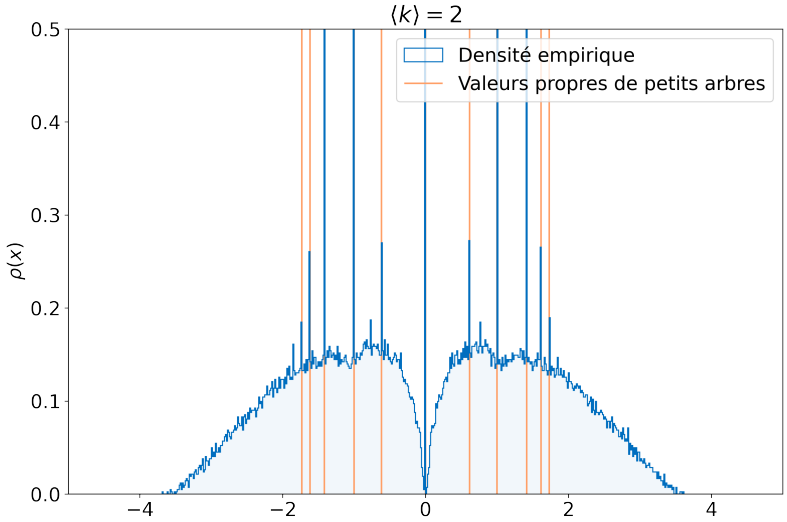



FIGURE 3.2 – Correspondance des valeurs propres de petits arbres calculées au tableau 3.1 et des discontinuités observées dans la densité spectrale empirique de réseaux Erdős-Rényi extrêmement creux. Les discontinuités les plus proéminentes correspondent aux valeurs propres des arbres les plus petits. La densité empirique est celle de trois réseaux de 10000 noeuds générés avec le modèle d'Erdős-Rényi pour $\langle k \rangle = 2$.

## 3.2 Formule de récurrence

Dans le cas du modèle d'Erdős-Rényi extrêmement creux, $p = \mathcal{O}(1/n)$, de sorte que l'hypothèse selon laquelle les termes dominants de l'équation 2.3 sont ceux pour lesquels la marche a un arbre comme graphe sous-jacent devient exacte, et non plus simplement une approximation. Toutefois, $\langle k \rangle$ est maintenant constant dans la limite $N \to \infty$, plutôt que proportionnel à $N$, de sorte que l'équation 2.6 ne tient plus ; il faut plutôt considérer tous les termes de la somme

à l'équation 2.5, et donc obtenir de l'information sur la séquence $\{I_{n,\ell}\}_{n=1,\ell=1}^{\infty,n}$, représentant le nombre de marches de longueur $2n$ sur des arbres et visitant $\ell$ arêtes distinctes. La première façon d'étudier la séquence est d'obtenir une relation de récurrence pour celle-ci.

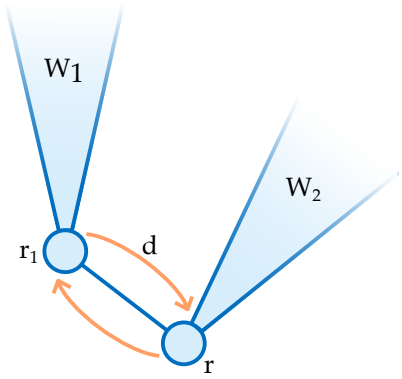



FIGURE 3.3 – Décomposition des marches en une marche $W_1$ émergeant du premier noeud visité $r_1$ et une autre marche, $W_2$, émergeant de la racine $r$, avec $d$ pas effectués sur la première arête traversée $(r, r_1)$.

La démarche présentée à cette section est présentée dans son essence par M. Bauer et O. Golinelli dans [4]. Il est possible d'obtenir une récurrence pour les $I_{n,\ell}$ en décomposant les marches tel que schématisé à la figure 3.3. La marche commence à la racine de l'arbre sur lequel elle prend place, étiquetée $r$. Comme mentionné à la section 2.1, la marche confère un ordre aux descendants de chaque noeud. Graphiquement, cet ordre peut être représenté en plaçant les descendants de gauche à droite dans l'ordre de leurs premières visites. Ainsi, le premier pas visite toujours le descendant le plus à gauche de $r$, étiqueté $r_1$. Or, $r_1$ est elle-même la racine d'un sous-arbre de l'arbre sur lequel se produit la marche. Ainsi, une fois arrivée à $r_1$, la marche réalise une marche fermée – possiblement vide – avant de revenir à $r$. Par le même principe, la marche effectue alors une marche fermée – possiblement vide – sur le sous-arbre de droite dont la racine est $r$, soit l'arbre sous-tendant la marche auquel on a retiré le sous-arbre ayant $r_1$ comme racine et l'arête $(r, r_1)$. La marche retourne alors à $r_1$, effectue une marche fermée, revient à $r$ et ainsi de suite jusqu'à ce qu'elle se termine.

Le résultat est le suivant : une suite de marches fermées sur le sous-arbre dont $r_1$ est la racine et une autre sur le sous-arbre dont $r$ est la racine, entrecoupées par les allers-retours sur la première arête visitée. Or, les différentes marches fermées sur le sous-arbre descendant de $r_1$ peuvent être mises bout à bout pour former une seule marche, tout comme celles sur le sous-arbre descendant de $r$, tel qu'illustré à la figure 3.4. Ce sont ces marches rapiécées qui sont nommées $W_1$ et $W_2$ à la figure 3.3. Ainsi, pour une marche fermée de longueur $2n$, si $W_1$ est de longueur $2n_1$, $W_2$ est de longueur $2(n - n_1 - d)$, avec $d$ le nombre d'allers-retours effectués sur l'arête $(r, r_1)$. De plus, si le sous-arbre couvert par $W_1$ a $\ell_1$ arêtes, celui couvert par $W_2$ a $\ell - \ell_1 - 1$ arêtes, le facteur $-1$ comptant l'arête $(r, r_1)$. Il y a donc $I_{n_1,\ell_1}$ marches $W_1$ possibles

et $I_{n-n_1-d,\ell-\ell_1-1}$ marches $W_2$ possibles.

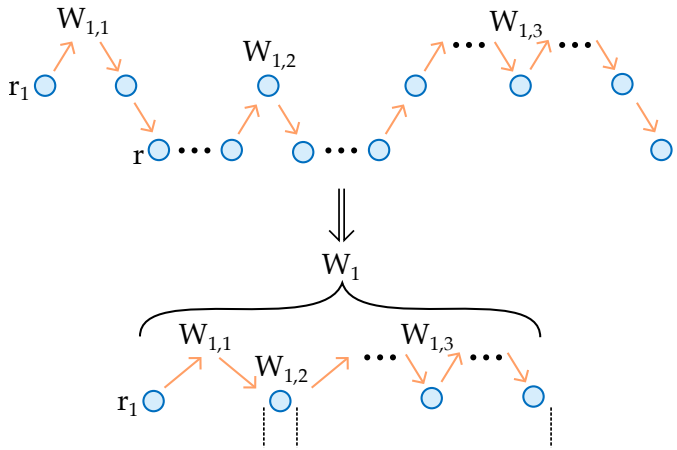



FIGURE 3.4 – Démonstration de la mise bout-à-bout des marches ayant comme racine $r_1$ pour former une marche continue $W_1$. Dans le cas présenté, $d = 3$ et $m_1 = 4$. $W_{1,2}$ est une marche vide et $W_{1,3}$ contient un retour à $r_1$ sans retourner sur l'arête $(r, r_1)$. Supprimer les retours sur $(r, r_1)$ et fusionner les marches $W_{1,i}$ donne de manière unique la marche $W_1$, à laquelle on superpose les retours vers $(r, r_1)$, représentés par les barres pointillées. Le principe est exactement le même pour $W_2$.

À ces comptes, il faut ajouter le nombre de façons d'insérer les $d$ allers-retours sur $(r, r_1)$ dans les marches $W_1$ et $W_2$. Par définition, ces cassures ne peuvent être insérées qu'après un pas revenant à la racine de la marche. Nommant $m$ – respectivement $m_1$ pour $W_1$ et $m_2$ pour $W_2$ – le nombre de ces pas, il y a $\binom{m_1+d-1}{d-1}$ façons d'insérer les $d$ cassures dans $W_1$ et $\binom{m_2+d-1}{d-1}$ de les insérer dans $W_2$. En effet, le problème revient simplement à choisir les $d-1$ emplacements des allers-retours sur $(r, r_1)$ parmi les $m + d - 1$ emplacements possibles, où le facteur $-1$ traduit le fait que le dernier emplacement pour $W_1$ – respectivement le premier pour $W_2$ – est forcément un pas sur $(r, r_1)$. Si $m$ est le nombre de retours à la racine de la marche complète, $m_2 = m - d$, puisque chaque retour à la racine de $W_2$ est un retour à la racine de la marche complète. Mettant tous ces éléments ensemble, la formule de récurrence est :

$$I_{n,\ell,m} = \sum_{n_1=0}^{n-1} \sum_{\ell_1=0}^{\ell-1} \sum_{m_1=0}^{n_1} \sum_{d=1}^{n} \binom{m_1+d-1}{d-1} \binom{m-1}{d-1} I_{n_1,\ell_1,m_1} I_{n-n_1-d,\ell-\ell_1-1,m-d}. \tag{3.1}$$

Les bornes s'expliquent ainsi : la longueur de $W_1$ peut être nulle, mais elle est au plus $n-1$ puisqu'un aller-retour doit être effectué sur l'arête $(r, r_1)$ ; par le même principe, $0 \leq \ell_1 \leq \ell-1$ puisqu'au minimum, l'arête $(r, r_1)$ n'appartient pas à $W_1$ ; $0 \leq m_1 \leq n_1$ puisqu'il est possible que la marche $W_1$ soit vide et ne revienne donc jamais à $r_1$, mais elle peut y revenir au plus $n_1$ fois, dans le cas où tous les pas de la marche font des allers-retours sur des arêtes connectées à $r_1$ ; et finalement, $1 \leq d \leq n$ puisque l'arête $(r, r_1)$ est par définition traversée en premier, et peut contenir tous les pas de la marche. Pour les conditions frontières, $I_{n,\ell,0} = 0$ sauf $I_{0,0,0} = 1$ – la marche vide – et $I_{n,\ell,m} = 0$ si $\ell > n$ ou si $m > n$. Les $I_{n,\ell}$ sont alors obtenus par la

simple formule $I_{n,\ell} = \sum_{m=1}^{n} I_{n,\ell,m}$. À la connaissance de l'auteur, il n'existe pas de formule de récurrence permettant d'éviter un indice supplémentaire relié au nombre de passages sur la racine de la marche.

Grâce à cette récurrence, il est possible de calculer un nombre arbitraire de valeurs pour $I_{n,\ell}$. Avec un peu moins d'une journée de calcul sur un ordinateur personnel, l'auteur a pu obtenir 120 rangées de la séquence [1]. Les neuf premières sont présentées à la table 3.2.

| $n$ \ $\ell$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | | | | | | | | |
| 1 | 0 | 1 | | | | | | | |
| 2 | 0 | 1 | 2 | | | | | | |
| 3 | 0 | 1 | 6 | 5 | | | | | |
| 4 | 0 | 1 | 14 | 28 | 14 | | | | |
| 5 | 0 | 1 | 30 | 110 | 120 | 42 | | | |
| 6 | 0 | 1 | 62 | 375 | 682 | 495 | 132 | | |
| 7 | 0 | 1 | 126 | 1190 | 3248 | 3731 | 2002 | 429 | |
| 8 | 0 | 1 | 254 | 3628 | 14062 | 23020 | 18928 | 8008 | 1430 |

TABLE 3.2 – Nombre $I_{n,\ell}$ de marches fermées de demi-longueur $n$ sur des arbres à $\ell$ arêtes.

## 3.3 Propriétés de la séquence $I_{n,\ell}$

Bien qu'il n'y ait pas de forme fermée connue pour les $I_{n,\ell}$, il est possible d'obtenir une grande quantité d'information sur la séquence. Les résultats de cette section sont essentiellement tirés de la référence [4] et de travaux non publiés de Pr. Nicolas Doyon [2].

### 3.3.1 Valeurs spécifiques

Il est possible d'obtenir une forme fermée pour les premières colonnes et diagonales du tableau 3.2. En effet, par inspection, les premières colonnes sont

$$\begin{aligned} I_{n,0} &= \delta_{n,0} \\ I_{n,1} &= 1 \\ I_{n,2} &= 2^n - 2. \end{aligned}$$

1. La séquence complète peut être transmise au lecteur sur demande. Il s'agit de la séquence OEIS A294439, qui sera mise à jour suite à la publication de ce mémoire.
2. Département de Mathématiques et Statistique, Université Laval. nicolas.doyon@mat.ulaval.ca

Les premières diagonales, cette fois, sont

$$\begin{aligned} I_{n,n} &= C_n = \frac{1}{n}\binom{2n}{n-1} \\ I_{n,n-1} &= \binom{2n}{n-2} \\ I_{n,n-2} &= \frac{n^2+7n-16}{2n+8}\binom{2n}{n-3}. \end{aligned}$$

### 3.3.2 Bornes

Il est possible d'établir une variété de bornes pour la séquence. Une des plus simples et des plus importantes est la suivante :

**Théorème 3.1.** Pour $2 \leq \ell \leq n$,

$$I_{n,\ell} \geq 2\left\{ {n \atop \ell} \right\} \geq 2\ell^{n-\ell},$$

où $\left\{ {n \atop k} \right\}$ sont les nombres de Stirling de seconde espèce.

*Preuve.* Les nombres de Stirling de seconde espèce (OEIS A008277) donnent le nombre de partitions d'un ensemble de $n$ éléments en $k$ sous-ensembles. Pour une marche dont l'arbre sous-jacent est un graphe étoile, c'est-à-dire un graphe dont tous les noeuds sont connectés à un seul noeud central (voir la figure 3.5), tous les pas doivent passer par le noeud central. Ainsi, pour un arbre à $\ell$ arêtes, le compte des marches revient simplement au compte du nombre de façon de distribuer les pas sur les $\ell$ noeuds connectés au noeud central. Or, cette quantité est précisément le nombre de partitions de $n$ éléments en $\ell$ sous-ensemble, c'est-à-dire les nombres de Stirling de seconde espèce. Le facteur 2 vient du fait que la racine de la marche peut être soit le noeud central, soit un des noeuds périphériques. Le cas $\ell = 1$ est exclu puisque le graphe à une arête n'a pas de distinction entre noeud central et noeud périphérique.

Puisque le graphe étoile n'est forcément pas le seul arbre sur lequel les marches comptées par les $I_{n,\ell}$ peuvent se produire, ceci complète la preuve. □

Quant à la deuxième borne, moins précise, il s'agit simplement d'un résultat connu pour les nombres de Stirling [37]. Elle sera nécessaire pour une démonstration ultérieure.

Le théorème 3.1 a une conséquence importante pour la feuille de route établie au chapitre 1. En effet, par ce théorème, pour $\langle k \rangle \geq 1$,

$$\mu_{2n} \geq 2\sum_{\ell=0}^{n} \left\{ {n \atop \ell} \right\} = 2B_n > n!,$$

où les $B_n$ sont les nombres de Bell (OEIS A000110). Or, il est connu que la série génératrice ordinaire de $n!$ ne converge nulle part excepté à $x = 0$, et n'admet ainsi pas de forme fermée.

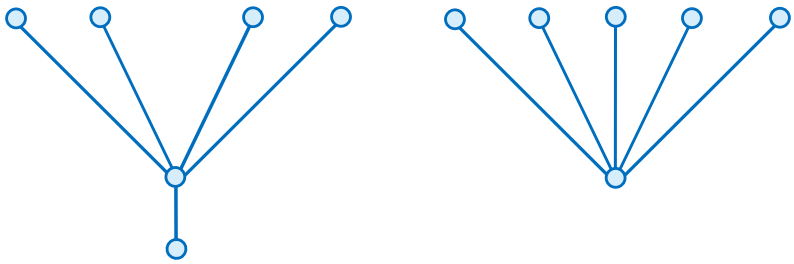


FIGURE 3.5 – Les deux configurations possibles du graphe étoile à 5 arêtes enraciné. À droite, la racine est le noeud central, alors qu'à gauche, la racine est un des noeuds périphériques.

En conséquence, le théorème 3.1 implique que la série génératrice ordinaire des moments de la densité spectrale limite n'admet pas de forme fermée. La série génératrice étant d'abord et avant tout une série formelle, ceci n'est pas un obstacle absolu à la complétion de la feuille de route, mais ce résultat encourage tout de même l'utilisation de la série génératrice exponentielle des moments pour l'étude de la densité spectrale.

Le lemme suivant permet d'obtenir une borne supérieure :

**Lemme 3.2.** Pour tout arbre $T_\ell$ à $\ell$ arêtes, $I_n(T_\ell) \leq \left\{ {n \atop \ell} \right\}$, où $I_n(T_\ell)$ est le nombre de marches fermées de longueur $2n$ couvrant $T_\ell$ et respectant l'ordre des descendants de chaque noeud de l'arbre.

La preuve du lemme 3.2 peut être trouvée en détails à la section 5.5 de [4]. Bien qu'elle soit conceptuellement assez simple, son exécution est longue et compliquée, de sorte qu'elle sera omise pour ce mémoire. Deux cas sont toutefois simples à expliquer : si $n < \ell$, $I_n(T_\ell) = 0$ puisque la marche est trop courte pour couvrir l'arbre, et si $n = \ell$, $I_n(T_\ell) = 1$. Cette dernière condition est équivalente au raisonnement présenté à la section 2.1 voulant qu'il existe une seule marche couvrant l'arbre en ne faisant qu'un aller-retour sur chaque arête, et mène à la conclusion $I_{\ell,\ell} = \sum_{T_\ell} 1 = C_\ell$.

Ce lemme permet d'établir la borne supérieure suivante :

**Théorème 3.3.** Pour $1 \leq \ell \leq n$,

$$I_{n,\ell} \leq C_\ell \left\{ {n \atop \ell} \right\} \leq 4^\ell \left\{ {n \atop \ell} \right\}.$$

*Preuve.* Par la définition de $I_n(T_\ell)$,

$$I_{n,\ell} = \sum_{T_\ell} I_n(T_\ell).$$

Or, par le lemme 3.2,

$$\sum_{T_\ell} I_n(T_\ell) \leq \sum_{T_\ell} \left\{ {n \atop \ell} \right\} = C_\ell \left\{ {n \atop \ell} \right\},$$

où la dernière égalité vient de la relation $\sum_{T_\ell} 1 = C_\ell$ énoncée ci-haut. De plus, $C_n \sim \frac{4^n}{n^{3/2}\sqrt{\pi}}$ [18], de sorte que $C_\ell \leq 4^\ell$. Ceci complète la preuve. □

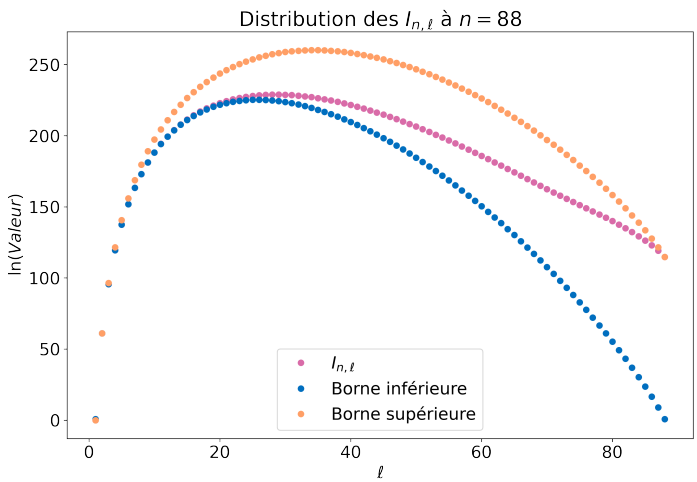



FIGURE 3.6 – Distribution des logarithmes de $I_{n,\ell}$ pour $n = 88$ et des bornes supérieure $C_\ell \left\{ {n \atop \ell} \right\}$ et inférieure $2\left\{ {n \atop \ell} \right\}$.

La figure 3.6 illustre ces bornes comparées à la séquence pour $n = 88$. Il est simple d'y constater que pour la majorité des valeurs de $\ell$, la borne inférieure $2\left\{ {n \atop \ell} \right\}$ est meilleure que la borne supérieure $C_\ell \left\{ {n \atop \ell} \right\}$, à l'exception de $\ell \approx n$. La borne inférieure semble particulièrement bonne autour du mode de la distribution, ce qui sera justifié ultérieurement.

### 3.3.3 Mode de la séquence à n fixe

L'investigation numérique de la séquence $I_{n,\ell}$ indique que pour $n$ fixe, la distribution est unimodale, tel qu'apparent à la figure 3.6. Pour identifier la position du mode, il est utile de définir la fonction $f_n(\ell)$, $\ell \in \mathbb{R}^+$ tel que $\ell \leq n$,

$$f_n(\ell) = (n - \ell) \ln \ell.$$

Par les théorèmes 3.1 et 3.3,

$$\exp\{f_n(\ell)\} \leq I_{n,\ell} \leq \exp\big(f_n(\ell) + \ln\big(4e^2\big)\ell\big) \equiv \exp(g_n(\ell)).$$

En effet, $\left\{ {n \atop \ell} \right\} \sim \frac{\ell^n}{\ell!}$, et par l'approximation de Stirling, $\ell! \sim \frac{\ell^\ell}{e^{\ell(1+o(1))}}$, ce qui permet d'obtenir l'inégalité de droite.

**Théorème 3.4.** La fonction $f_n(\ell)$ a un maximum unique, situé à la valeur

$$\ell^* = \frac{n(\ln n + \ln\ln n)}{(\ln n)^2} + \mathcal{O}\left(\frac{n}{(\ln(n))^2}\right).$$

*Preuve.* La dérivée première de $f_n(\ell)$ est

$$f_n'(\ell) = \frac{n}{\ell} - 1 - \ln \ell,$$

alors que sa dérivée seconde est

$$f_n''(\ell) = -\frac{n}{\ell^2} - \frac{1}{\ell}.$$

Puisque $\lim_{\ell\to 0^+} f_n'(\ell) = \infty$ et $\lim_{\ell\to\infty} f_n'(\ell) = -\infty$, le maximum existe. De plus, puisque $f_n''(\ell)$ est toujours négative, le maximum est unique. Par définition, le maximum est la valeur qui satisfait l'équation

$$f_n'(\ell^*) = \frac{n}{\ell^*} - 1 - \ln \ell^* = 0 \tag{3.2}$$

Cette équation admet une solution exacte en termes de la fonction W de Lambert. Toutefois, aux fins de la démonstration du prochain théorème, il est utile d'en donner une forme en termes de fonctions élémentaires. Soit l'approximation de départ $\ell_1^* = \frac{n}{\ln n}$. Alors,

$$f_n'(\ell_1^*) = \ln n - 1 - \ln n + \ln\ln n = \ln\ln n - 1$$

et

$$f_n''(\ell_1^*) = -\frac{(\ln n)^2 - \ln n}{n}.$$

Ainsi, par la méthode d'approximation de Newton-Raphson, l'approximation peut être améliorée par

$$\begin{aligned}
\ell_2^* &= \ell_1^* - \frac{f'(\ell_1^*)}{f_n''(\ell_1^*)} \\
&= \frac{n}{\ln n} + \frac{n\ln\ln n - n}{(\ln n)^2 - \ln n} \\
&= \frac{n}{\ln n} + \frac{n\ln\ln n}{(\ln n)^2} + \mathcal{O}\left(\frac{n}{(\ln n)^2}\right).
\end{aligned}$$

Puisque

$$\begin{aligned}
f_n'(\ell_2^*) &= \ln n\left(1 + \frac{\ln\ln n}{\ln n} + \mathcal{O}\left(\frac{1}{\ln n}\right)\right)^{-1} - 1 - \ln n + \ln\ln n + \mathcal{O}\left(\frac{\ln\ln n}{\ln n}\right) \\
&= \ln n\left(1 - \frac{\ln\ln n}{\ln n} + \mathcal{O}\left(\frac{(\ln\ln n)^2}{(\ln n)^2}\right)\right) - 1 - \ln n + \ln\ln n + \mathcal{O}\left(\frac{\ln\ln n}{\ln n}\right) \\
&= \mathcal{O}(1),
\end{aligned}$$

la prochaine correction apportée par la méthode de Netwon-Raphson serait $\mathcal{O}(1)$. Ainsi,

$$\begin{aligned}\ell^* &= \frac{n}{\ln n} + \frac{n \ln\ln n}{(\ln n)^2} + \mathcal{O}\left(\frac{n}{(\ln n)^2}\right) \\ &= \frac{n(\ln n + \ln\ln n)}{(\ln n)^2} + \mathcal{O}\left(\frac{n}{(\ln n)^2}\right). \quad \square\end{aligned}$$

De ce résultat découle le théorème suivant :

**Théorème 3.5.** Pour $n$ fixe, la valeur de $\ell$ qui maximise $I_{n,\ell}$ est

$$\ell_{\max} = \frac{n}{\ln n} + \mathcal{O}\left(\frac{n}{(\ln n)^{3/2}}\right) = (1 + o(1))\frac{n}{\ln n}.$$

*Preuve.* Soit $\tilde{\ell}$, la valeur de $\ell$ qui maximise $g(\ell) \equiv f(\ell) + \ell \ln\big(4e^2\big)$. Par la même démarche qu'à la preuve du théorème 3.4, il est possible de montrer que

$$\tilde{\ell} = \frac{n(\ln n + \ln\ln n)}{(\ln n)^2} + \mathcal{O}\left(\frac{n}{(\ln n)^2}\right).$$

De plus,

$$\begin{aligned}f_n''(\tilde{\ell}) &= -\frac{n}{\tilde{\ell}^2} - \frac{1}{\tilde{\ell}} \\ &= -\frac{(\ln n)^2}{n}(1 + o(1)).\end{aligned}$$

Ainsi, puisque $f(\ell) \leq g(\ell)$, et en développant $g(\ell)$ en série de Taylor autour de $\tilde{\ell}$,

$$f(\ell) \leq f(\tilde{\ell}) + \tilde{\ell} \ln\big(4e^2\big) - \frac{(\ln n)^2}{2n}(\ell - \tilde{\ell})^2,$$

où $g'(\tilde{\ell}) = 0$ par définition. En particulier, $f(\ell) \leq g(\ell_{max})$, de sorte que

$$\begin{aligned}& f(\ell) \leq f(\tilde{\ell}) + \tilde{\ell} \ln\big(4e^2\big) - \frac{(\ln n)^2}{2n}(\ell_{max} - \tilde{\ell})^2 \\ \iff & \frac{(\ln n)^2}{2n}(\ell_{max} - \tilde{\ell})^2 \leq (f(\tilde{\ell}) - f(\ell)) + \tilde{\ell} \ln\big(4e^2\big).\end{aligned}$$

Puisque l'inégalité ci-dessus est vraie pour tout $\ell$, en particulier pour $\ell = \tilde{\ell}$,

$$\begin{aligned}\frac{(\ln n)^2}{2n}(\ell_{max} - \tilde{\ell})^2 \leq \tilde{\ell} \ln\big(4e^2\big) &= \frac{n \ln\big(4e^2\big)(\ln n + \ln\ln n)}{(\ln n)^2} + \mathcal{O}\left(\frac{n}{(\ln n)^2}\right) \\ \iff (\ell_{max} - \tilde{\ell})^2 &\leq \frac{2n^2 \ln\big(4e^2\big)(\ln n + \ln\ln n)}{(\ln n)^4} + \mathcal{O}\left(\frac{n^2}{(\ln n)^4}\right) \\ \iff \left|\ell_{max} - \tilde{\ell}\right| &\leq \frac{\sqrt{2\ln(4e^2)}\, n\sqrt{\ln n + \ln\ln n}}{(\ln n)^2} + \mathcal{O}\left(\frac{n}{(\ln n)^2}\right).\end{aligned}$$

Ainsi,

$$\left|\ell_{max} - \tilde{\ell}\right| \leq \mathcal{O}\left(\frac{n}{(\ln n)^{3/2}}\right),$$

de sorte que

$$\begin{aligned}
\ell_{max} &= \tilde{\ell} + \mathcal{O}\left(\frac{n}{(\ln n)^{3/2}}\right) \\
&= \frac{n(\ln n + \ln\ln n)}{(\ln n)^2} + \mathcal{O}\left(\frac{n}{(\ln n)^{3/2}}\right) \\
&= \frac{n}{\ln n} + \mathcal{O}\left(\frac{n}{(\ln n)^{3/2}}\right) = (1 + o(1))\frac{n}{\ln n},
\end{aligned}$$

ce qui complète la preuve. □

### 3.3.4 Comportement asymptotique

Pour $\ell \geq 2$ et $\ell$ fixe, donc $\ell \ll n$,

$$I_{n,\ell} \sim 2\left\{{n \atop \ell}\right\}. \tag{3.3}$$

Une preuve rigoureuse de cet énoncé peut être trouvée dans [4]. Le raisonnement présenté ici-bas est plutôt un argument heuristique permettant de saisir l'idée de la preuve. Celui-ci découle de bornes conditionnelles sur les $I_{n,\ell}$.

**Théorème 3.6.** Soit $I_{n,\ell,u}$, le nombre de marches de demi-longueur $n$ sur un arbre à $\ell$ sommets et contenant $u$ *séquences ascendantes*. Alors

$$I_{n,\ell,u} \leq \binom{2n-2}{2u-1}\binom{n}{\ell}\ell^u.$$

*Preuve.* Une séquence ascendante est une séquence de pas dans la marche telle que chaque pas de la séquence visite un noeud descendant du précédent noeud visité. En d'autres mots, il s'agit d'une séquence ne contenant aucun retour sur ses pas. Une marche peut alors toujours être divisée en ses séquences ascendantes et descendantes, qui alternent forcément. Ainsi, cette suite de séquences ascendantes et descendantes peut être décrite simplement en marquant le début et la fin de chaque séquence ascendante.

Pour une marche de longueur $2n$, comme le premier pas est forcément le début d'une séquence ascendante et que le dernier pas ne peut être la fin d'une de ces séquences, il y a $2n-2$ endroits possibles pour placer les $s$ marqueurs de début et fin des séquences. Placer $s$ marqueurs revient à définir $\frac{s+1}{2} = u$ séquences ascendantes. Il y a donc

$$\binom{2n-2}{s} = \binom{2n-2}{2u-1}$$

manières de placer les marqueurs des séquences dans la marche.

De plus, puisqu'il n'existe qu'un seul chemin entre deux noeuds dans un arbre, le trajet d'une séquence ascendante est entièrement déterminé par ses extrémités. Ainsi, étant donné un point de départ, le nombre de manières de choisir ce trajet est au plus $\ell$, le nombre de destinations possibles. Enfin, il faut garantir que la marche visite toutes les arêtes au moins une fois. Puisqu'une arête ne peut être visitée la première fois qu'au cours d'une séquence ascendante, le nombre de manières de choisir les positions des premières visites des arêtes est au plus $\binom{n}{\ell}$. □

Le théorème 3.6 permet d'établir que le nombres de marches avec un grand nombres de séquences ascendantes est borné moins sévèrement que pour un petit nombre de séquences ascendantes, puisque la borne supérieure évolue comme $\ell^u$. Ainsi, il est raisonnable de croire que pour $\ell \to \infty$, les marches maximisant $u$ deviennent dominantes dans le compte de $I_{n,\ell}$. Or, $\max u = n$, précisément dans le cas du graphe étoile. Or, il a déjà été établi dans la preuve du théorème 3.1 que les marches sur le graphe étoile sont au nombre de $2\left\{{n \atop \ell}\right\}$.

Cet argument est imparfait puisque les marches sur le graphe étoile commençant sur un noeud périphérique ont plutôt $u \geq n/2$, et sont donc équivalentes aux marches sur n'importe quel arbre à deux niveaux dont un des niveaux contient une seule arête. L'argument peut être renforcé à l'aide d'une autre borne conditionnelle.

**Théorème 3.7.** Soit $I_{n,\ell,d}$, le nombre de marches fermées de demi-longueur $n$ sur un arbre à $\ell$ arêtes et dont le degré maximal est $d$. Alors

$$I_{n,\ell,d} \leq (d-1)^n.$$

*Preuve.* Chaque pas ascendant de la marche a au plus $d-1$ choix de noeuds à visiter. Puisqu'il y a $n$ pas ascendants dans la marche, la borne est établie. □

Le théorème 3.7 établit de manière similaire au théorème 3.6 que les marches dominant le compte de $I_{n,\ell}$ pour $\ell \to \infty$ ont probablement un noeud de degré très élevé. Or, $\max d = n$, encore une fois dans le cas du graphe étoile. Par le même calcul que précédemment, ceci indique $I_{n,\ell} \sim 2\left\{{n \atop \ell}\right\}$. Les vérifications numériques effectuées confirment également l'exactitude asymptotique de la borne donnée au théorème 3.1.

## 3.4 Conséquences pour la densité spectrale

Si l'absence de forme fermée pour les $I_{n,\ell}$ ou pour leur série génératrice ne permet pas d'obtenir une forme exacte pour la densité spectrale du modèle d'Erdős-Rényi extrêmement creux, les résultats assemblés jusqu'ici donnent tout de même de l'information sur celle-ci.

### 3.4.1 Propriétés de la densité spectrale

La première propriété importante de $\rho(x)$ que les résultats de la section 3.3 permettent d'établir est qu'elle est entièrement déterminée par ses moments, c'est-à-dire qu'il existe une seule distribution sur $\mathbb{R}$ dont les moments sont ceux calculés avec l'équation 2.5. Avant d'énoncer ce résultat, il convient d'introduire un théorème classique qui sera nécessaire pour le démontrer.

**Théorème 3.8.** Condition de Carleman [40]. Soit $m$, une mesure sur $\mathbb{R}$ telle que ses moments

$$\mu_n = \int_{-\infty}^{\infty} x^n \mathrm{d}m(x)$$

sont finis pour tout $n$. Alors $m$ est la seule mesure sur $\mathbb{R}$ dont $\{\mu_n\}_{n=0}^{\infty}$ est la séquence de moments si

$$\sum_{n=1}^{\infty} \mu_{2n}^{-1/2n} = \infty.$$

Il est maintenant possible de démontrer l'énoncé suivant :

**Théorème 3.9.** La densité spectrale limite du modèle d'Erdős-Rényi extrêmement creux, $\rho(x)$, définie sur $\mathbb{R}$, est entièrement déterminée par ses moments.

*Preuve.* Puisque la condition de Carleman ne demande qu'une connaissance des moments pairs de la mesure pour être testée, il est suffisant de savoir que les moments de $\rho(x)$ sont

$$\mu_{2n} = \sum_{\ell=0}^{n} I_{n,\ell} \langle k \rangle^{\ell}.$$

Par le théorème 3.3,

$$\begin{aligned}
\mu_{2n} &\leq \sum_{\ell=0}^{n} C_\ell \left\{ {n \atop \ell} \right\} \langle k \rangle^{\ell} \\
&\leq \sum_{\ell=0}^{n} C_n \left\{ {n \atop \ell} \right\} \langle k \rangle^{n} \\
&= C_n B_n \langle k \rangle^{n},
\end{aligned}$$

où $B_n$ sont les nombres de Bell, qui comptent le nombre de partitions de $n$ objets. La deuxième inégalité vient du fait que $C_\ell \leq C_n \ \forall \ell \leq n$ et que $\langle k \rangle^{\ell} \leq \langle k \rangle^{n}$. La dernière égalité résulte de $\sum_{\ell=0}^{n} \left\{ {n \atop \ell} \right\} = B_n$, par définition. Le comportement asymptotique des nombres de Catalan est

$$C_n \sim \frac{4^n}{n^{3/2}\sqrt{\pi}} < 4^n.$$

Quant aux nombres de Bell, la borne suivante a été établie par D. Berend et T. Tassa dans [6] :

$$B_n < \left(\frac{0.792n}{\ln(n+1)}\right)^n < n.$$

Ainsi, il existe un $n^*$ tel que

$$\begin{aligned} \mu_{2n} &\leq (4\langle k\rangle n)^n \quad \forall n > n^* \\ \implies \mu_{2n}^{-1/2n} &\geq (4\langle k\rangle n)^{-1/2} \quad \forall n > n^*. \end{aligned}$$

Or, il est connu que la série $\sum_{n=1}^{\infty} n^{-1/2}$ diverge. Ainsi, la série $\sum_{n=0}^{\infty} \mu_{2n}^{-1/2n}$ diverge également, et la condition de Carleman est respectée. □

La deuxième propriété de $\rho(x)$ qu'il est possible d'établir est la suivante :

**Théorème 3.10.** La densité spectrale limite du modèle d'Erdős-Rényi extrêmement creux est non bornée, c'est-à-dire que son support est $\mathbb{R}$.

*Preuve.* Soit $f(x)$ la densité dont les moments impairs sont nuls et les moments pairs sont deux fois les polynômes de Touchard,

$$S_n(\langle k\rangle) = \sum_{\ell=0}^{n} \left\{ {n \atop \ell} \right\} \langle k\rangle^{\ell}.$$

Comme les polynômes de Touchard sont les moments d'une distribution de Poisson, il est possible de les réécrire comme

$$S_n(\langle k\rangle) = e^{-\langle k\rangle} \sum_{\ell=0}^{\infty} \frac{\langle k\rangle^{\ell}\, \ell^n}{\ell!}.$$

Par l'équation 1.11,

$$\begin{aligned} f(x) &= \int_{-\infty}^{\infty} e^{2\pi ixs} \sum_{n=0}^{\infty} \frac{2S_n(\langle k\rangle)(-2\pi is)^{2n}}{(2n)!} \mathrm{d}s \\ &= 2e^{-\langle k\rangle} \int_{-\infty}^{\infty} e^{-2\pi ixs} \sum_{n=0}^{\infty} \sum_{\ell=0}^{\infty} \frac{\langle k\rangle^{\ell}\, (2\pi is\sqrt{\ell})^{2n}}{\ell!(2n)!} \mathrm{d}s. \end{aligned}$$

Inversant l'ordre de sommation, et avec $\cosh(x) = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!}$,

$$\begin{aligned} f(x) &= 2e^{-\langle k\rangle} \int_{-\infty}^{\infty} e^{-2\pi ixs} \sum_{\ell=0}^{\infty} \frac{\langle k\rangle^{\ell}}{\ell!} \cosh\Big(2\pi is\sqrt{\ell}\Big) \mathrm{d}s \\ &= 2e^{-\langle k\rangle} \int_{-\infty}^{\infty} e^{-2\pi ixs} \sum_{\ell=0}^{\infty} \frac{\langle k\rangle^{\ell}}{\ell!} \cos\Big(2\pi s\sqrt{\ell}\Big) \mathrm{d}s \\ &= 2e^{-\langle k\rangle} \sum_{\ell=0}^{\infty} \frac{\langle k\rangle^{\ell}}{\ell!} \int_{-\infty}^{\infty} e^{-2\pi ixs} \cos\Big(2\pi s\sqrt{\ell}\Big) \mathrm{d}s \\ &= e^{-\langle k\rangle} \sum_{\ell=0}^{\infty} \frac{\langle k\rangle^{\ell}}{\ell!} \Big[\delta(x-\sqrt{\ell}) + \delta(x+\sqrt{\ell})\Big], \end{aligned}$$

où la dernière égalité est un résultat classique de la transformée de Fourier. Puisque $f(x)$ est un peigne de deltas de Dirac se situant à toutes les racines carrées d'entiers, il est clair que $f(x)$ est non bornée, au sens où il n'existe pas de constante $M$ telle que $f(x) = 0\ \forall |x| \geq M$. Or, par le théorème 3.1, les moments de $\rho(x)$ sont tous supérieurs ou égaux à ceux de $f(x)$. Ainsi, $\rho(x)$ est également non bornée. □

Le théorème 3.10 établit une différence qualitative importante avec le cas dense. En effet, la distribution du demi-cercle, pour sa part, est bornée : elle est nulle pour $|x| \geq 2\sqrt{\langle k \rangle}$.

### 3.4.2 Corrections à la loi du demi-cercle

Les valeurs spécifiques énoncées à la sous-section 3.3.1 permettent d'introduire des corrections à la distribution du demi-cercle pour la rapprocher de la densité spectrale limite du modèle d'Erdős-Rényi extrêmement creux. En effet, pour $\langle k \rangle = c \gg 1$ avec $c$ une constante, il est logique de ne considérer que les quelques premiers termes de la série à l'équation 2.5. La première de ces contributions sous-dominantes est celle apportée par la série $\ell = n - 1$, soit

$$\left\langle x^{2n} \right\rangle_{\rho,sd} = I_{n,n-1} \langle k \rangle^{n-1}$$

Étant donné la linéarité de la partie imaginaire dans l'équation 1.10, la contribution de cette série peut être traitée séparément et ajoutée comme correction au demi-cercle.

La séquence $I_{n,n-1} = \binom{2n}{n-2}$ est la séquence OEIS A002694. Sa fonction génératrice exponentielle paire est aisément accessible :

$$\begin{aligned}
G_E(x) &= \sum_{n=0}^{\infty} \binom{2n}{n-2} \frac{x^{2n}}{(2n)!} \\
&= \sum_{n=0}^{\infty} \frac{1}{(n-2)!(n+2)!} x^{2n} \\
&= \sum_{n=0}^{\infty} \frac{1}{n!(n+4)!} x^{2n+4} \\
&= I_4(2x)
\end{aligned}$$

$$\begin{aligned}
\implies G_{E,sd}(x) &= \sum_{n=0}^{\infty} \binom{2n}{n-2} \langle k \rangle^{n-1} \frac{x^{2n}}{(2n)!} \\
&= \frac{1}{\langle k \rangle} \sum_{n=0}^{\infty} \binom{2n}{n-2} \frac{(\sqrt{\langle k \rangle} x)^{2n}}{(2n)!} = \frac{I_4(2\sqrt{\langle k \rangle} x)}{\langle k \rangle}.
\end{aligned}$$

Ainsi, la correction apportée par cette série à la densité spectrale peut être calculée de manière similaire à la section 2.3 :

$$G_{E,sd}(-2\pi i s) = \int_{-\infty}^{\infty} e^{-2\pi i x s} \rho_{sd}(x) \mathrm{d}x = \frac{I_4(-4\pi i \sqrt{\langle k \rangle} s)}{\langle k \rangle}. \tag{3.4}$$

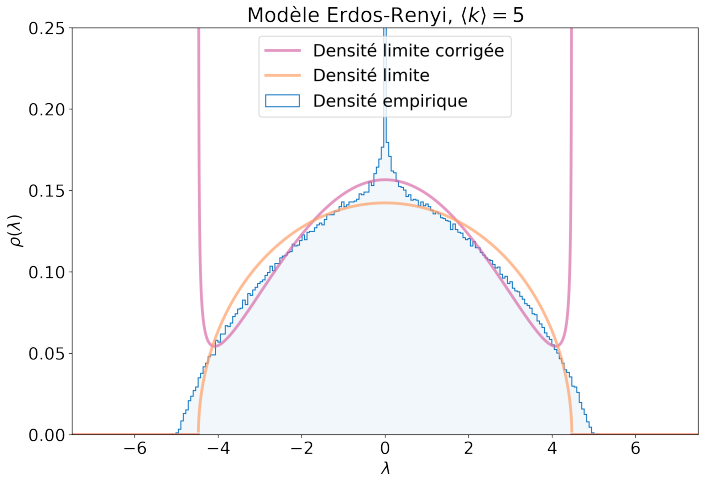



FIGURE 3.7 – Comparaison entre la densité spectrale empirique et les densités spectrales limites données par la loi du demi-cercle et par le demi-cercle corrigé avec l'équation 3.5.

La fonction de Bessel modifiée $I_4$ admet la définition intégrale suivante :

$$I_4(z) = \frac{1}{\pi}\int_0^{\pi} e^{z\cos(\theta)}\cos(4\theta)\mathrm{d}\theta.$$

Ainsi,

$$I_4(-4\pi i\sqrt{\langle k\rangle}s) = \frac{1}{\pi}\int_0^{\pi} e^{-4\pi i\sqrt{k}s\cos(\theta)}\cos(4\theta)\mathrm{d}\theta.$$

Pour obtenir la forme désirée, il convient d'effectuer la substitution

$$\begin{aligned} x &= 2\sqrt{\langle k\rangle}\cos(\theta), \\ \theta &= \arccos\left(\frac{x}{2\sqrt{\langle k\rangle}}\right) \\ \mathrm{d}\theta &= \frac{-1}{\sqrt{4\langle k\rangle - x^2}}\mathrm{d}x. \end{aligned}$$

Pour évaluer la quantité $\cos(4\theta)$ avec la substitution, il faut invoquer les polynômes de Tchebyshev, définis comme

$$T_n(\cos(\theta)) = \cos(n\theta).$$

Ainsi,

$$\cos(n\arccos(z)) = T_n(z).$$

Or,

$$T_4(z) = 8x^4 - 8x^2 + 1,$$

de sorte que

$$\cos\left(4\arccos\left(\frac{x}{2\sqrt{\langle k\rangle}}\right)\right) = \frac{x^4}{2\langle k\rangle^2} - \frac{2x^2}{\langle k\rangle} + 1$$
$$= \frac{x^4 - 4\langle k\rangle x^2 + 2\langle k\rangle^2}{2\langle k\rangle^2}$$

et

$$\frac{I_4(-4\pi i\sqrt{\langle k\rangle}s)}{\langle k\rangle} = \frac{1}{\pi}\int_{-2\sqrt{\langle k\rangle}}^{2\sqrt{\langle k\rangle}} e^{-2\pi ixs}\left(\frac{x^4 - 4\langle k\rangle x^2 + 2\langle k\rangle^2}{2\pi\langle k\rangle^3\sqrt{4\langle k\rangle - x^2}}\right)\mathrm{d}x$$
$$= \frac{1}{\pi}\int_{-\infty}^{\infty} e^{-2\pi ixs}\left(\frac{x^4 - 4\langle k\rangle x^2 + 2\langle k\rangle^2}{2\pi\langle k\rangle^3\sqrt{4\langle k\rangle - x^2}}\mathbb{1}_{|x|\leq 2\sqrt{\langle k\rangle}}\right)\mathrm{d}x$$

En comparant avec l'équation 3.4, il appert que

$$\rho_{sd}(x) = \frac{x^4 - 4\langle k\rangle x^2 + 2\langle k\rangle^2}{2\pi\langle k\rangle^3\sqrt{4\langle k\rangle - x^2}}\mathbb{1}_{|x|\leq 2\sqrt{\langle k\rangle}}.$$

La densité spectrale limite corrigée est donc

$$\rho(x) = \left(\frac{\sqrt{4\langle k\rangle - x^2}}{2\pi\langle k\rangle} + \frac{x^4 - 4\langle k\rangle x^2 + 2\langle k\rangle^2}{2\pi\langle k\rangle^3\sqrt{4\langle k\rangle - x^2}}\right)\mathbb{1}_{|x|\leq 2\sqrt{\langle k\rangle}}. \tag{3.5}$$

Sa forme est présentée à la figure 3.7. S'il est clair que la correction améliore le résultat, le défaut fondamental de cette approche apparait : les corrections apportées ainsi présentent des discontinuités aux bords du domaine du demi-cercle. De plus, $\rho(x)$ est encore nulle hors du domaine du demi-cercle, alors qu'il a été prouvé à la section précédente que ce n'est pas le cas pour les réseaux extrêmement épars. Cette approche ne peut donc pas mener à une description satisfaisante de la densité spectrale pour $|x| > 2\sqrt{\langle k\rangle}$.

## 3.5 Solution exacte : les graphes réguliers

Il est possible d'obtenir une forme exacte pour la densité spectrale limite de graphes extrêmement épars dans le cas d'un modèle particulier : les graphes réguliers aléatoires. En effet, la certitude quant au degré de chaque noeud permet d'obtenir une forme fermée pour la série génératrice des marches fermées sur ces graphes.

Pour un graphe $k$-régulier aléatoire, $\langle x^{2n}\rangle_\rho$ est le nombre de marches fermées sur un arbre $k$-régulier commençant à la racine. Dans un tel arbre, la racine a $k$ descendants, et tous les autres noeuds ont $k-1$ descendants et un ancêtre. Si une marche revient à la racine pour la première fois après $2(i+1)$ pas, le premier pas descend forcément d'un niveau, avec $k$ choix possibles pour le noeud visité. Puis, les $2i$ pas subséquents forment une marche fermée restant sous la racine, où chaque pas est soit un pas descendant, pour lequel il y a $k-1$ choix, ou un

pas ascendant, pour lequel il n'y a qu'une seule possibilité. Le nombre de manières de choisir l'ordre des pas descendants et ascendants est équivalent au nombre de manières d'ouvrir et de fermer $2i$ parenthèses de sorte que toutes les parenthèses ouvertes soient fermées, mais qu'à aucun moment il n'y ait plus de parenthèses fermées que d'ouvertes. Or, ce nombre est précisément donné par le nombre de Catalan $C_i$. La marche revient ensuite à la racine au pas $2(i+1)$, laissant $\left\langle x^{2(n-i-1)}\right\rangle_\rho$ façons de continuer. Ainsi,

$$\left\langle x^{2n}\right\rangle_\rho = k\sum_{i=0}^{n-1} C_i(k-1)^i \left\langle x^{2(n-i-1)}\right\rangle_\rho .$$

Sommant sur $n$,

$$\begin{aligned} G_O(x) = \sum_{n=0}^{\infty}\left\langle x^{2n}\right\rangle_\rho x^{2n} &= 1 + k\sum_{n=1}^{\infty}\sum_{i=0}^{n-1} C_i(k-1)^i \left\langle x^{2(n-i-1)}\right\rangle_\rho x^{2n} \\ &= 1 + kx^2\sum_{n=0}^{\infty}\sum_{i=0}^{n} C_i(k-1)^i \left\langle x^{2(n-i)}\right\rangle_\rho x^{2n} \\ &= 1 + kx^2\left(\sum_{i=0}^{\infty} C_i(k-1)^i x^{2i}\right)\left(\sum_{n=0}^{\infty}\left\langle x^{2n}\right\rangle_\rho x^{2n}\right). \end{aligned}$$

Avec la série génératrice ordinaire des nombres de Catalan, donnée au théorème 2.2, l'équation ci-dessus devient

$$G_O(x) = 1 + k\frac{1-\sqrt{1-4(k-1)x^2}}{2(k-1)}G_O(x),$$

de sorte que

$$G_O(x) = \left(1 - k\frac{1-\sqrt{1-4(k-1)x^2}}{2(k-1)}\right)^{-1} = \frac{2(k-1)}{k-2+k\sqrt{1-4(k-1)x^2}}. \tag{3.6}$$

Ainsi,

$$\begin{aligned} \rho(x) &= -\frac{1}{\pi}\lim_{\epsilon\to 0^+}\operatorname{Im}\left(\frac{1}{x+i\epsilon}\frac{2(k-1)}{k-2+k\sqrt{1-\frac{4(k-1)}{(x+i\epsilon)^2}}}\right) \\ &= -\frac{1}{\pi}\lim_{\epsilon\to 0^+}\operatorname{Im}\left(\frac{2(k-1)}{(k-2)(x+i\epsilon)+k\sqrt{(x+i\epsilon)^2-4(k-1)}}\right) \end{aligned}$$

De manière similaire à la section 2.3,

$$(x+i\epsilon)^2 - 4(k-1) = re^{i\varphi}$$

avec

$$\begin{aligned} r &= \sqrt{(x^2-\epsilon^2-4(k-1))^2+4x^2\epsilon^2} \\ \varphi &= \begin{cases} \arctan\left(\frac{2x\epsilon}{x^2-\epsilon^2-4(k-1)}\right) & \text{si } x^2 > 4(k-1)+\epsilon^2, \\ \arctan\left(\frac{2x\epsilon}{x^2-\epsilon^2-4(k-1)}\right)+\pi & \text{si } x^2 < 4(k-1)+\epsilon^2 \text{ et } x \geq 0, \\ \arctan\left(\frac{2x\epsilon}{x^2-\epsilon^2-4(k-1)}\right)-\pi & \text{si } x^2 < 4(k-1)+\epsilon^2 \text{ et } x < 0, \end{cases} \end{aligned}$$

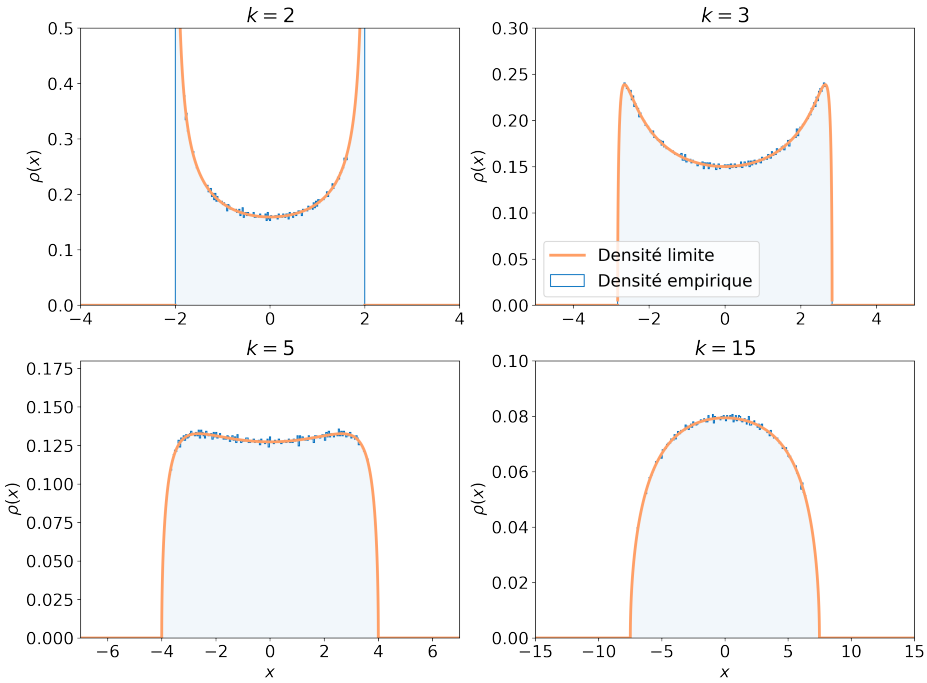



FIGURE 3.8 – Densité spectrale empirique et loi de Kesten-McKay pour quatre ensembles de trois graphes réguliers aléatoires de 10000 noeuds générés avec $k = 2, 3, 5$ et 15.

de sorte que

$$\begin{aligned}
\rho(x) &= -\frac{1}{\pi} \lim_{\epsilon \to 0^+} \operatorname{Im}\left( \frac{2(k-1)}{(k-2)x + k\sqrt{r}\cos(\varphi/2) + i((k-2)\epsilon + k\sqrt{r}\sin(\varphi/2))} \right) \\
&= \frac{2(k-1)}{\pi} \lim_{\epsilon \to 0^+} \frac{(k-2)\epsilon + k\sqrt{r}\sin(\varphi/2)}{(k-2)^2(x^2+\epsilon^2) + 2k(k-2)\sqrt{r}(x\cos(\varphi/2) + \epsilon\sin(\varphi/2)) + k^2 r} \\
&= \frac{2(k-1)}{\pi} \lim_{\epsilon \to 0^+} \frac{k\sqrt{r}\sin(\varphi/2)}{(k-2)^2 x^2 + 2k(k-2)\sqrt{r}\cos(\varphi/2)x + k^2 r} \\
&= \begin{cases} 0 & \text{si } x^2 > 4(k-1) \\ \frac{2(k-1)}{\pi} \frac{k\sqrt{4(k-1)-x^2}}{(k-2)^2x^2 + 4k^2(k-1) - k^2x^2} & \text{si } x^2 < 4(k-1). \end{cases}
\end{aligned}$$

Ainsi, finalement,

$$\rho(x) = \frac{k\sqrt{4(k-1)-x^2}}{2\pi(k^2-x^2)} \mathbb{1}_{|x| \leq 2\sqrt{k-1}}. \tag{3.7}$$

Cette densité est un résultat classique connu sous le nom de *loi de Kesten-McKay* [28, 48]. Sa forme est présentée à la figure 3.8. Il est évident que pour des petites valeurs de $k$, cette densité est totalement différente de celle du modèle d'Erdős-Rényi pour des degrés moyens égaux (voir figure 3.1), ce qui confirme l'idée que les propriétés locales des graphes gouvernent

maintenant leur densité spectrale, plutôt que la propriété globale qu'est leur degré moyen. Il a également été prouvé que, dans la limite $\langle k \rangle \to \infty$, cette loi converge bel et bien vers la loi du demi-cercle [5].

Il est à noter que l'équation 3.7 montre que la densité spectrale des graphes réguliers aléatoires reste bornée pour toute valeur de $k$, contrairement à celle des graphes Erdős-Rényi. Cette différence indique que le taux de croissance des moments de la densité spectrale des graphes réguliers est largement inférieur à celui des moments de celle des graphes Erdős-Rényi. Ce résultat n'est pas particulièrement surprenant : en effet, en permettant une variation importante du degré des noeuds visités, le modèle d'Erdős-Rényi admet un bien plus grand nombre de formes de marches fermées que les graphes réguliers. Il n'est donc pas surprenant que le nombre de marches fermées attendu croisse plus rapidement pour le modèle d'Erdős-Rényi.

L'idée permettant d'obtenir la loi de Kesten-McKay peut être généralisée à d'autres modèles plus spécifiques pour lesquels la certitude sur les degrés des noeuds où d'autres propriétés des graphes permet d'obtenir une forme exacte pour la série génératrice des marches fermées. Cette avenue a été explorée en détail par C. D. Godsil et B. Mohar dans [20]. Par exemple, pour les graphes biréguliers – soit les graphes où seuls deux degrés sont permis et les liens n'existent qu'entre des noeuds de différent degré – avec degrés $k$ et $l$, ces derniers trouvent la densité limite

$$\rho(x) = \frac{kl\sqrt{-\left(x^2 - kl + (p-1)^2\right)\left(x^2 - kl + (p+1)^2\right)}}{\pi(k+l)\left(kl - x^2\right)|x|}\mathbb{1}_{|\sqrt{k-1}-\sqrt{l-1}|\leq|x|\leq\sqrt{k-1}+\sqrt{l-1}},$$

où $p = \sqrt{(k-1)(l-1)}$. Il est facile de vérifier que lorsque $k = l$, cette densité est égale à la loi de Kesten-McKay.

## 3.6 Forme asymptotique pour $x \gg 1$

Ensemble, le théorème 3.5 et l'équation 3.3 suggèrent que la distribution dont les moments impairs sont nuls et les moments pairs sont deux fois les polynômes de Touchard devrait être une bonne approximation de la densité spectrale limite. En effet, l'équation 3.3 est une approximation particulièrement bonne pour $\ell \ll n$. Or, le mode de $I_{n,\ell}$ à $n$ fixe se trouvant à $\frac{n}{\ln n}(1 + o(1)) \ll n$, il est logique de poser que $\mu_{2n} \approx 2S_n(\langle k \rangle)$ est une bonne approximation à petit $\langle k \rangle$. Plus exactement, puisque l'équation 3.3 est asymptotique, cette approximation est meilleure pour les moments d'ordre supérieur. Ainsi, la distribution en résultant devrait être une approximation particulièrement bonne de $\rho(x)$ pour $x \gg 1$. Il est donc pertinent de calculer cette distribution.

Ce calcul a été réalisé avec la série génératrice exponentielle à la preuve du théorème 3.10, mais la distribution résultante prenait la forme d'un peigne de deltas de Dirac à chaque racines d'entiers. Or, comme la distance entre les racines d'entiers est $\sqrt{k} - \sqrt{k-1} \sim \frac{1}{2\sqrt{k}} \to 0$, il

est possible de croire qu'une forme continue de cette distribution puisse être obtenue, à tout le moins pour $x \gg 1$. L'autre possibilité est donc de passer par la série génératrice ordinaire. Soit $\left\langle x^{2n} \right\rangle_\rho \approx 2S_n(\langle k \rangle) = 2\left\langle x^n \right\rangle_{\text{Poisson}}$. Alors

$$\begin{aligned} G_{O,\rho}(z) &:= \sum_{n=0}^{\infty} \left\langle x^{2n} \right\rangle_\rho z^{2n} \\ &\approx 2\sum_{n=0}^{\infty} \left\langle x^n \right\rangle_{\text{Poisson}} z^{2n} \\ &=: 2G_{O,\text{Poisson}}(z^2). \end{aligned}$$

Ainsi,

$$\begin{aligned} \rho(x) &:= \frac{1}{\pi} \lim_{\epsilon \to 0^+} \operatorname{Im}\left( \frac{1}{x+i\epsilon} G_{O,\rho}\left( \frac{1}{x+i\epsilon} \right) \right) \\ &\approx \frac{2}{\pi} \lim_{\epsilon \to 0^+} \operatorname{Im}\left( \frac{1}{x+i\epsilon} G_{O,\text{Poisson}}\left( \frac{1}{(x+i\epsilon)^2} \right) \right) \\ &= \frac{2}{\pi} \lim_{\epsilon \to 0^+} \operatorname{Im}\left( (x+i\epsilon) \frac{1}{(x+i\epsilon)^2} G_{O,\text{Poisson}}\left( \frac{1}{(x+i\epsilon)^2} \right) \right). \end{aligned}$$

Or,

$$\operatorname{Im}\{\operatorname{Re}\{z_1\} + i\operatorname{Im}\{z_1\})((\operatorname{Re}\{z_2\} + i\operatorname{Im}\{z_2\})\} = \operatorname{Re}\{z_1\}\operatorname{Im}\{z_2\} + \operatorname{Im}\{z_1\}\operatorname{Re}\{z_2\},$$

de sorte que

$$\begin{aligned} \operatorname{Im}\left( (x+i\epsilon) \frac{1}{(x+i\epsilon)^2} G_{O,\text{Poisson}}\left( \frac{1}{(x+i\epsilon)^2} \right) \right) &= x \operatorname{Im}\left( \frac{1}{(x+i\epsilon)^2} G_{O,\text{Poisson}}\left( \frac{1}{(x+i\epsilon)^2} \right) \right) \\ &\quad + \epsilon \operatorname{Re}\left( \frac{1}{(x+i\epsilon)^2} G_{O,\text{Poisson}}\left( \frac{1}{(x+i\epsilon)^2} \right) \right). \end{aligned}$$

Puisque typiquement, $\operatorname{Re}\left( \frac{1}{(x+i\epsilon)^2} G_{O,\text{Poisson}}\left( \frac{1}{(x+i\epsilon)^2} \right) \right)$ ne diverge pas lorsque $\epsilon \to 0^+$, dans cette limite, le deuxième terme disparaît et seul le premier doit être considéré. Ainsi,

$$\rho(x) \approx \frac{2}{\pi} x \lim_{\epsilon \to 0^+} \operatorname{Im}\left( \frac{1}{(x+i\epsilon)^2} G_{O,\text{Poisson}}\left( \frac{1}{(x+i\epsilon)^2} \right) \right).$$

Avec le changement de variable $(x+i\epsilon)^2 = x^2 - \epsilon^2 + i2\epsilon = x' + i\epsilon'$, et notant que $\lim_{\epsilon \to 0^+} = \lim_{\epsilon' \to 0^+}$ puisque $\epsilon' = 2\epsilon$,

$$\begin{aligned} \rho(x) &\approx \frac{2}{\pi} x \lim_{\epsilon' \to 0^+} \operatorname{Im}\left( \frac{1}{x'+i\epsilon'} G_{O,\text{Poisson}}\left( \frac{1}{x'+i\epsilon'} \right) \right) \\ &= 2|x| \operatorname{Poisson}(x^2), \end{aligned} \tag{3.8}$$

où $x \to |x|$ a été posé par contrainte de symétrie de la densité spectrale limite. Ainsi,

$$\begin{aligned} \rho(x) &\approx 2|x| \operatorname{Poisson}(x^2) \\ &\sim 2|x| \frac{e^{-\langle k \rangle} \langle k \rangle^{x^2}}{\Gamma(1+x^2)}. \end{aligned}$$

Avec l'approximation de Stirling, pour $x$ suffisamment grand, $\Gamma(1+x^2) \approx |x|\sqrt{2\pi}\left(\frac{x^2}{e}\right)^{x^2}$, de sorte que

$$\rho(x) \sim \sqrt{\frac{2}{\pi}} e^{-\langle k \rangle} \left(\frac{e\,\langle k \rangle}{x^2}\right)^{x^2}. \tag{3.9}$$

La dérivation ci-dessus admet quelques entorses à la rigueur mathématique, notamment en considérant la distribution de Poisson comme une fonction continue. Malgré tout, cette expression a déjà été obtenue par G. J. Rodgers et A. J. Bray dans [38] ainsi que par G. Semerjian et L. F. Cugliandolo dans [39]. Rodgers et Bray l'obtiennent suite à un traitement perturbatif complexe d'une solution pour la densité spectrale en termes d'une équation intégrale non linéaire. Quant à Semerjian et Cugliandolo, leur résultat provient d'une analyse par méthode des répliques et par méthode du point col de la fonction $\frac{1}{z}G_O\left(\frac{1}{z}\right)$. Toutefois, à la connaissance de l'auteur, la dérivation présentée dans ce mémoire est originale.

L'équation 3.8 présente un résultat somme toute étonnant, soit que la densité spectrale limite du modèle d'Erdős-Rényi est directement reliée à la distribution de Poisson. Or, la distribution de Poisson est intimement reliée au modèle d'Erdős-Rényi, puisqu'elle en est la distribution des degrés. En effet, pour le modèle d'Erdős-Rényi, la distribution des degrés est binomiale,

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

Or, dans la limite $n \to \infty$ avec $np = \langle k \rangle$ constant, il est connu que la distribution binomiale converge vers la distribution de Poisson

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

Il appert improbable que l'apparition de la distribution des degrés du modèle dans l'équation 3.8 soit une coïncidence, aussi est-il tentant de conjecturer que, pour tout modèle de réseau non corrélé avec $\langle k \rangle$ fini dans la limite $n \to \infty$,

$$\rho(x) \cong 2|x|P(x^2), \tag{3.10}$$

avec $P(k)$ la distribution des degrés du modèle en question. Cette conjecture est corroborée par les réseaux invariants d'échelle[3], dont la distribution de degrés est en loi de puissance, $P(k) \approx P_0 k^{-\gamma}$. Pour ces réseaux, F. Chung, L. Lu et V. Vu ont montré dans [9] que les plus grandes valeurs propres sont distribuées selon $P(k) \sim k^{1-2\gamma}$. Or, $2|x|P(x^2) = 2P_0|x|^{1-2\gamma}$.

La preuve de la conjecture 3.10 excède largement la portée de ce mémoire, mais a été donnée par S. N. Dorogovtsev et al. dans [11].

3. En anglais, *scale-free networks*

# Conclusion

L'objectif de ce projet de recherche était de développer une méthode utilisant les outils de la combinatoire pour étudier la densité spectrale limite de graphes aléatoires non corrélés. Une telle procédure a été établie avec succès au chapitre 1, reliant les séries génératrices du nombre de marches fermées attendu sur le graphe à l'étude à la densité spectrale limite de celui-ci. Il a donc été établi qu'il est suffisant de connaître le nombre attendu de marches fermées de longueur $n$ pour un modèle de graphe aléatoire, ou même simplement sa série génératrice, pour pouvoir calculer sa densité spectrale. La procédure a été utilisée avec succès au chapitre 2, obtenant la loi du demi-cercle de Wigner comme forme fermée pour la densité spectrale limite du modèle d'Erdős-Rényi. En plus de confirmer ce résultat classique, la procédure développée dans ces pages permet de donner une interprétation combinatoire des moments de la loi du demi-cercle. Les résultats analytiques et numériques obtenus suggèrent également que cette loi peut être étendue à la densité spectrale de plusieurs autres modèles de réseaux.

Au chapitre 3, une analyse numérique a montré que le spectre de graphes Erdős-Rényi extrêmement creux diffère fortement de la loi du demi-cercle, et présente des discontinuités à des valeurs précises. Il a été montré par calcul direct que ces valeurs correspondent aux valeurs propres de petits arbres. Le problème de la densité spectrale de réseaux extrêmement creux a ensuite été étudié à l'aide de la méthode combinatoire. La méthode a permis de reproduire un résultat connu quant à la forme fermée de la densité spectrale de graphes réguliers aléatoires, mais n'a pas permis d'en obtenir une pour le modèle d'Erdős-Rényi extrêmement creux. Malgré tout, il a été possible d'étudier en détail les moments de la densité spectrale limite du modèle d'Erdős-Rényi. Cette étude a permis de déterminer une correction pour le coeur de la densité, en plus de déterminer une forme asymptotique pour les bords de celle-ci. Finalement, il a été possible de poser une conjecture sur la forme asymptotique des bords de la densité spectrale de tous les modèles de réseaux extrêmement creux et non corrélés.

Le problème de la densité spectrale limite du modèle d'Erdős-Rényi extrêmement creux est un problème coriace, résistant de manière remarquable aux méthodes présentées dans ce mémoire ainsi que dans la littérature scientifique. Il est toutefois pertinent de noter que l'essentiel de cette littérature se concentre, directement ou indirectement, sur la série génératrice ordinaire des moments de la densité spectrale – voir par exemple [11, 39]. Or, il a été noté à la section

3.3 que la croissance rapide des moments empêche l'existence d'une forme fermée pour la série génératrice ordinaire. De plus, les marches fermées sont naturellement des objets étiquetés, puisque chaque pas est associé à sa position dans la marche. Or, il est connu que les séries génératrices exponentielles sont en général plus appropriées pour l'étude des objets étiquetés que les séries génératrices ordinaires. Ainsi, il est de l'avis de l'auteur que l'utilisation de la série génératrice exponentielle pour l'étude de la densité spectrale a plus de potentiel pour obtenir une forme fermée pour la densité spectrale limite du modèle d'Erdős-Rényi extrêmement creux, et que la recherche à venir sur le sujet devrait se concentrer sur celle-ci.

# Annexe A

# Formules d'inversion pour la densité spectrale

Cette annexe contient les preuves des formules 1.10 et 1.11. La preuve formelle de la formule 1.10 demande des notions plus avancées d'analyse complexe et de convergences de mesures et d'intégrales, et excède donc la portée de ce mémoire. La démonstration présentée ici prend donc quelques libertés avec ces notions, mais expose l'idée de la preuve formelle, s'inspirant des présentations faites dans [27] et [47]. La formule et sa démonstration formelle sont toutefois un résultat classique, pouvant être consulté par exemple dans la référence [8]. Dans cette référence – au changement de notation près –, elle est présentée comme

$$\rho(x) = -\frac{1}{\pi} \lim_{\epsilon \to 0^+} \mathrm{Im}\,(m(x + i\epsilon)), \tag{A.1}$$

avec $m(z)$ appelée la *transformée de Stieltjes*, *transformée de Stieltjes-Perron*, *résolvente* ou encore *fonction de Green*, et définie par

$$m(z) = \int_{-\infty}^{\infty} \frac{\rho(x')}{z - x'} \mathrm{d}x'.$$

Il est évident par sa définition que $m(z)$ est définie pour $z$ strictement complexe, soit $z \in \mathbb{C}/\mathbb{R}$, puisque le dénominateur est nul pour $z$ réel. Or, $m(z)$ peut être exprimée en termes de la fonction génératrice ordinaire des moments de $\rho(x)$. En effet,

$$\begin{aligned}\frac{1}{z} G_O\left(\frac{1}{z}\right) &:= \frac{1}{z} \sum_{k=0}^{\infty} \frac{\left\langle x'^k \right\rangle_\rho}{z^k} \\ &= \frac{1}{z} \sum_{k=0}^{\infty} \int_{-\infty}^{\infty} \rho(x') \left(\frac{x'}{z}\right)^k \mathrm{d}x'.\end{aligned}$$

Inversant la somme et l'intégrale,

$$\begin{aligned}\frac{1}{z}G_O\left(\frac{1}{z}\right) &= \frac{1}{z}\int_{-\infty}^{\infty}\frac{\rho(x')}{1-\frac{x'}{z}}\mathrm{d}x' \\ &= \int_{-\infty}^{\infty}\frac{\rho(x')}{z-x'}\mathrm{d}x' := m(z).\end{aligned}$$

Posant $z = x + i\epsilon$,

$$\begin{aligned}\mathrm{Im}\left(m(x+i\epsilon)\right) &= \mathrm{Im}\left(\int_{-\infty}^{\infty}\frac{\rho(x')}{x-x'+i\epsilon}\mathrm{d}x'\right) \\ &= -\int_{-\infty}^{\infty}\frac{\epsilon\rho(x')}{(x-x')^2+\epsilon^2}\mathrm{d}x'.\end{aligned}$$

La limite $\epsilon \to 0^+$ peut toujours être prise tout en restant dans le domaine de $m(z)$, puisque $z$ reste strictement complexe. Or, il est connu que

$$\lim_{\epsilon\to 0^+}\frac{1}{\pi}\frac{\epsilon}{x^2+\epsilon^2} = \delta(x),$$

où $\delta(x)$ est la distribution delta de Dirac. Ainsi,

$$\begin{aligned}\lim_{\epsilon\to 0^+}\mathrm{Im}\left(m(x+i\epsilon)\right) &= -\pi\int_{-\infty}^{\infty}\rho(x')\delta(x-x')\mathrm{d}x' \\ &= -\pi\rho(x).\end{aligned}$$

Conséquemment,

$$\rho(x) = -\frac{1}{\pi}\lim_{\epsilon\to 0^+}\mathrm{Im}\left(m(x+i\epsilon)\right).$$

□

Pour ce qui est de la formule 1.11, la démonstration reproduite ici est celle faite par S. Rabbani dans [36]. Soit la fonction génératrice exponentielle des moments,

$$G_E(z) = \sum_{k=0}^{\infty}\left\langle x^k\right\rangle_\rho\frac{z^k}{k!}.$$

En posant $z = -2\pi i s$,

$$\begin{aligned}G_E(-2\pi i s) &= \sum_{k=0}^{\infty}\left\langle x^k\right\rangle_\rho\frac{(-2\pi i s)^k}{k!} \\ &= \sum_{k=0}^{\infty}\int_{-\infty}^{\infty}x^k\rho(x)\frac{(-2\pi i s)^k}{k!}\mathrm{d}x.\end{aligned}$$

Inversant la somme et l'intégrale,

$$\begin{aligned}G_E(-2\pi i s) &= \int_{-\infty}^{\infty}\rho(x)\sum_{k=0}^{\infty}\frac{(-2\pi i x s)^k}{k!}\mathrm{d}x \\ &= \int_{-\infty}^{\infty}\rho(x)e^{-2\pi i x s}\mathrm{d}x.\end{aligned}$$

Or, il s'agit précisément de la définition en fréquence de la transformée de Fourier, de sorte que

$$G_E(-2\pi is) = \mathcal{F}\{\rho(x)\}(s).$$

Par le théorème d'inversion de Fourier,

$$\rho(x) = \mathcal{F}^{-1}\{G_E(-2\pi is)\}(x) = \int_{-\infty}^{\infty} e^{2\pi ixs} G_E(-2\pi is)\mathrm{d}s.$$

$\square$

# Annexe B

# Correction à la loi du demi-cercle pour réseaux très denses

La version exacte de la loi de Wigner pour les réseaux denses peut être calculée à partir de la matrice d'adjacence centrée $\hat{\mathbf{A}}$, définie comme

$$\hat{\mathbf{A}} := \mathbf{A} - \langle \mathbf{A} \rangle \,.$$

Pour le modèle d'Erdős-Rényi, $\langle \mathbf{A} \rangle = p(\mathbf{1} - \mathbf{I})$, avec $\mathbf{1}$ la matrice dont toutes les entrées sont 1 et $\mathbf{I}$ la matrice identité. Ainsi, $\langle \mathbf{A} \rangle$ est une matrice de rang 1, de sorte que les valeurs propres de $\hat{\mathbf{A}}$ et de $\mathbf{A}$ sont entrelacées [21] :

$$\lambda_n^{\hat{\mathbf{A}}} \leq \lambda_n^{\mathbf{A}} \leq \lambda_{n-1}^{\hat{\mathbf{A}}} \ldots \leq \lambda_1^{\hat{\mathbf{A}}} \leq \lambda_1^{\mathbf{A}}.$$

C'est donc dire que le coeur des deux densités est le même dans la limite $N \to \infty$, avec seulement la plus grande valeur propre différant. Celle-ci est bien étudiée [21, 32], mais ne sera pas au menu de cette annexe. Les moments de la densité spectrale de $\hat{\mathbf{A}}$ sont, tel qu'établi à la section 1.4,

$$\begin{aligned} \langle x^n \rangle_\rho &= \frac{1}{N} \left\langle \mathrm{Tr}\left\{ \hat{\mathbf{A}} \right\} \right\rangle \\ &= \frac{1}{N} \sum_{\mathcal{W}_n(\mathcal{K})} \langle \hat{a}_{i_1 i_2} \hat{a}_{i_2 i_3} \cdots \hat{a}_{i_n i_1} \rangle \,, \end{aligned}$$

où les indices $i_k$ sont ceux des noeuds visités par la marche $w \in \mathcal{W}_n(\mathcal{K})$ correspondant au terme choisi. Sous hypothèse d'indépendance des entrées de $\hat{\mathbf{A}}$, l'espérance peut être distribuée sur chacun des liens visités par la marche. Ainsi, de manière analogue au chapitre 2, la somme ci-dessus peut être découpée en trois contributions :

— $N^{\underline{\nu}} \approx N^\nu$ façons d'attribuer des noeuds aux indices $i_k$ de la marche, où $\nu$ est le nombre de noeuds distincts visités par la marche ;

— Un facteur $F_{n,\nu,\ell}$ comptant le nombre de formes de marches possibles ;

— Le produit des contributions $\left\langle \hat{a}_{ij}^{\mu} \right\rangle$ de chacune des arêtes visitées à la marche, où $\mu$ est le nombre de passages sur chaque arête.

Or, par définition, $\langle \hat{a}_{ij} \rangle = 0$, de sorte que seules les marches passant au moins deux fois sur chaque arête contribuent à $\langle x^n \rangle_\rho$. De plus, dans la limite $N \to \infty$, seules les marches marches maximisant $\nu$ sont significatives. Or, avec la contrainte que chaque arête soit visitée au moins deux fois, la valeur maximale de $\nu$ est $\frac{n}{2}+1$, précisément dans le cas (à nouveau !) où le graphe sous-jacent est un arbre et où chaque arête est traversée exactement deux fois. Ainsi, par le même raisonnement qu'au chapitre 2,

$$\langle x^n \rangle_\rho = \frac{1}{N} \left\langle \mathrm{Tr}\left(\hat{\mathbf{A}}\right) \right\rangle = \begin{cases} 0 & n \text{ impair} \\ C_{n/2} \left( N \left\langle \hat{a}_{ij}^2 \right\rangle \right)^{n/2} & n \text{ pair.} \end{cases} \tag{B.1}$$

Or, $\hat{a}_{ij}$ a deux valeurs possibles : $1-p$, avec probabilité $p$, et $-p$, avec probabilité $1-p$. Ainsi,

$$\begin{aligned} \left\langle \hat{a}_{ij}^2 \right\rangle &= p(1-p)^2 + (1-p)p^2 \\ &= p(1-p)(1-p+p) \\ &= p(1-p). \end{aligned}$$

Avec $Np = \langle k \rangle$, il en découle que

$$\left\langle x^{2n} \right\rangle_\rho = C_n(\langle k \rangle (1-p))^n.$$

Le reste du calcul étant identique au chapitre 2, la correction est introduite.

# Bibliographie


[1] Amol Aggarwal : Bulk universality for generalized Wigner matrices with few moments. *Probability Theory and Related Fields*, 173(1-2), février 2019.

[2] Albert-László Barabási : The network takeover. *Nature Physics*, 8(1):14–16, 2012.

[3] Albert-László Barabási et Réka Albert : Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999.

[4] M. Bauer et O. Golinelli : Random incidence matrices : moments of the spectral density. *Journal of Statistical Physics*, 103(1/2):301–337, 2001.

[5] Roland Bauerschmidt, Antti Knowles et Horng-Tzer Yau : Local semicircle law for random regular graphs. *Communications on Pure and Applied Mathematics*, 70(10):1898–1960, 2017.

[6] Daniel Berend et Tamir Tassa : Improved Bounds on Bell Numbers and on Moments of Sums of Random Variables. *Probability and Mathematical Statistics*, 30:185–205, 2010.

[7] Béla Bollobás : *Modern Graph Theory*. Springer, 1998.

[8] Sang-Il Choi : The inversion formula of the Stieltjes transform of spectral distribution. *Journal of the Chungcheong Mathematical Society*, 22(3):6, 2009.

[9] Fan Chung, Linyuan Lu et Van Vu : Eigenvalues of Random Power law Graphs. *Annals of Combinatorics*, 7(1):21–33, 2003.

[10] Dragos M Cvetkovic et Ivan Gutman : *Applications of graph spectra*. Matematički institut SANU, 2009.

[11] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes et A. N. Samukhin : Spectra of complex networks. *Physical Review E*, 68(4):046109, 2003.

[12] Freeman J. Dyson : A Brownian-Motion Model for the Eigenvalues of a Random Matrix. *Journal of Mathematical Physics*, 3(6):1191–1198, novembre 1962.

[13] László Erdős, Antti Knowles, Horng-Tzer Yau et Jun Yin : Spectral Statistics of Erdős—Rényi Graphs I : Local Semicircle Law. *The Annals of Probability*, 41(3), 2013.

[14] László Erdős, Horng-Tzer Yau et Jun Yin : Bulk universality for generalized Wigner matrices. *Probability Theory and Related Fields*, 154(1-2), octobre 2012.

[15] P. Erdős et A. Rényi : On random graphs i. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[16] P. Erdős et A Rényi : On the evolution of random graphs. *In Publication Of The Mathematical Institute Of The Hungarian Academy Of Sciences*, pages 17–61, 1960.

[17] Illés J. Farkas, Imre Derényi, Albert-László Barabási et Tamás Vicsek : Spectra of "real-world" graphs : Beyond the semicircle law. *Physical Review E*, 64(2):026704, 2001.

[18] Philippe Flajolet et Robert Sedgewick : *Analytic combinatorics*. Cambridge University Press, 2009.

[19] E. N. Gilbert : Random Graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[20] C.D. Godsil et B. Mohar : Walk generating functions and spectral measures of infinite graphs. *Linear Algebra and its Applications*, 107:191–206, 1988.

[21] Alice Guionnet : Bernoulli Random Matrices. *arXiv :2112.05506 [math]*, 2021.

[22] Paul W. Holland et Samuel Leinhardt : Transitivity in Structural Models of Small Groups. *Comparative Group Studies*, 2(2):107–124, 1971.

[23] Jiaoyang Huang, Benjamin Landon et Horng-Tzer Yau : Bulk universality of sparse random matrices. *Journal of Mathematical Physics*, 56(12), décembre 2015.

[24] Thomas Koshy : Catalan Numbers. *In Catalan Numbers with Applications*. Oxford University Press, 2008.

[25] Reimer Kuehn : Spectra of Sparse Random Matrices. *Journal of Physics A : Mathematical and Theoretical*, 41(29):295002, 2008.

[26] Edward Laurence, Nicolas Doyon, Louis J. Dubé et Patrick Desrosiers : Spectral Dimension Reduction of Complex Dynamical Networks. *Physical Review X*, 9(1):011042, 2019.

[27] Giacomo Livan, Marcel Novaes et Pierpaolo Vivo : *Introduction to Random Matrices - Theory and Practice*, volume 26. Springer, 2018.

[28] Brendan D. McKay : The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its Applications*, 40:203–216, 1981.

[29] Madan Lal Mehta : *Random Matrices*. Elsevier, octobre 2004.

[30] Fernando L. Metz et Jeferson D. Silva : Spectral density of dense random networks and the breakdown of the Wigner semicircle law. *Physical Review Research*, 2(4), octobre 2020.

[31] J. L. Moreno et H. H. Jennings : Statistics of Social Configurations. *Sociometry*, 1(3/4):342–374, 1938.

[32] Raj Rao Nadakuditi et M. E. J. Newman : Spectra of random graphs with arbitrary expected degrees. *Phys. Rev. E*, 87:012803, 2013.

[33] M. E. J. Newman : *Random graphs as models of networks*, chapitre 2, pages 35–68. John Wiley & Sons, 2002.

[34] M. E. J. Newman : *Networks*. Oxford University Press, 2018.

[35] Bogdan Nica : *A Brief Introduction to Spectral Graph Theory*. EMS, mai 2018.

[36] Sahand Rabbani : Probability Density Function in Terms of Moments. https://srabbani.com/moments2.pdf, 2008. Consulté le 2022-04-26.

[37] B.C. Rennie et A.J. Dobson : On Stirling numbers of the second kind. *Journal of Combinatorial Theory*, 7(2):116–121, 1969.

[38] G. J. Rodgers et A. J. Bray : Density of states of a sparse random matrix. *Physical Review B*, 37(7), 1988.

[39] Guilhem Semerjian et Leticia F. Cugliandolo : Sparse random matrices : the eigenvalue spectrum revisited. *Journal of Physics A : Mathematical and General*, 35(23):4837–4851, 2002.

[40] James Alexander Shohat et Jacob David Tamarkin : *The Problem of Moments*. American Mathematical Society, 1950.

[41] Richard P. Stanley : *Enumerative Combinatorics*, volume 1 de *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2$^e$ édition, 2011.

[42] Terence Tao et Van Vu : Random Matrices : Universality of Local Eigenvalue Statistics up to the Edge. *Communications in Mathematical Physics*, 298(2):549–572, septembre 2010.

[43] Terence Tao et Van Vu : Random matrices : Universality of local eigenvalue statistics. *Acta Mathematica*, 206, mars 2011.

[44] Duncan J. Watts et Steven H. Strogatz : Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[45] Eugene P. Wigner : Characteristic Vectors of Bordered Matrices With Infinite Dimensions. *Annals of Mathematics*, 62(3):548–564, 1955. Publisher : Annals of Mathematics.

[46] Eugene P. Wigner : On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, 67(2):325–327, 1958.

[47] Jean-Gabriel Young : *De la détection de la structure communautaire des réseaux complexes*. Thèse de Maîtrise, Université Laval, 2014.

[48] Yufei Zhao : Spectral Distributions of Random Graphs. https://web.mit.edu/18.338/www/2012s/projects/yz_report.pdf. Consulté le 2022-04-26.

[49] Katharina A. Zweig : Random graphs as null models. *In Lecture Notes in Social Networks*, pages 183–214. Springer, 2016.