



Sur la représentation géométrique des réseaux orientés

Mémoire

Jérémi Lesage

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

Québec, Canada

Sur la représentation géométrique des réseaux orientés

Mémoire

Jérémi Lesage

Sous la direction de :

Antoine Allard, directeur de recherche

Résumé

Les systèmes complexes sont constitués d'un ensemble d'éléments en interaction, dont les interactions, souvent asymétriques, engendrent des comportements collectifs émergents. Dans le cadre de la modélisation des réseaux complexes, un paradigme géométrique exploite l'espace hyperbolique comme outil pour mieux représenter la structure de ces systèmes et en identifier les propriétés fondamentales. En ce sens, nous présentons diMercator, un algorithme utilisant le modèle \mathcal{S}^1 orienté-réiproque pour réaliser le plongement de réseaux orientés sur le cercle. Adapté de l'algorithme Mercator, son équivalent pour les réseaux non orientés, diMercator infère les paramètres du modèle afin d'optimiser la correspondance avec la topologie du réseau observé. Sans surprise, la prise en compte de l'asymétrie des interactions est essentielle pour modéliser les réseaux orientés, et diMercator répond efficacement à ce défi en réalisant des plongements de réseaux orientés plus spécifiques que ceux obtenus avec Mercator en tenant compte de la directionnalité inhérente des interactions. Nos résultats indiquent que diMercator produit non seulement des plongements de haute fidélité de réseaux orientés empiriques, mais permet d'encoder efficacement plusieurs propriétés observées dans ces réseaux à l'aide d'un petit nombre de coordonnées géométriques. En effet, notre approche permet une réduction dimensionnelle de $N(N - 1)$ à $3N + 2$ paramètres tout en conservant approximativement plusieurs propriétés topologiques essentielles. Parmi les propriétés reproduites, on retrouve notamment la cohérence trophique, la non-normalité, les configurations de triangles, la réciprocité individuelle de chaque noeud, ainsi que le spectre des valeurs singulières et la structure en noeud papillon. Ce qui rend ces résultats particulièrement remarquables, c'est que les propriétés ainsi bien reproduites ne sont pourtant pas explicitement optimisées par le modèle. En ce qui a trait au caractère hyperbolique de notre démarche, la procédure d'inférence développée permet certes un plongement sur le cercle, dont l'équivalence avec l'espace hyperbolique est un résultat connu dans le cas des réseaux non orientés, mais le lien avec l'espace hyperbolique n'est pas formellement démontré dans notre approche. Bien que le lien entre notre modèle et l'espace hyperbolique n'ait pas encore été formellement établi, notre travail constitue une avancée conceptuelle et méthodologique majeure en proposant un modèle probabiliste compact des graphes qui, au-delà de la seule hypothèse d'hyperbolicité, permet une réduction dimensionnelle significative tout en préservant approximativement plusieurs propriétés topologiques essentielles.

Table des matières

Résumé	ii
Table des matières	iii
Liste des figures	iv
Remerciements	vi
Avant-propos	vii
Introduction	1
1 Notions préliminaires	3
1.1 Les systèmes et les réseaux complexes	3
1.2 Point sur la géométrie hyperbolique	10
1.3 Modélisation des réseaux complexes	20
2 Propriétés topologiques et géométrie de réseaux orientés empiriques	42
2.1 Résumé	42
2.2 Abstract	43
2.3 Introduction	43
2.4 Methods	44
2.5 Results	47
2.6 Conclusion	50
2.7 Supplementary Material	51
Conclusion	73
Bibliographie	75
A Résultats mathématiques utiles	82
A.1 Fonction hypergéométrique	82
A.2 Décomposition en fractions partielles	83
B Équations maîtresses approximées générales pour dynamiques binaires sur graphes aléatoires	84
B.1 Travaux connexes	85
B.2 Approche généralisée par types de noeuds	88
B.3 Validation numérique de l'approche généralisée par types de noeuds	91

Liste des figures

1.1	Graphe	4
1.2	Graphe orienté	4
1.3	Types d'interaction dans un graphe orienté	5
1.4	Interaction bidirectionnelle dans un graphe orienté	5
1.5	Exemple de graphe orienté	6
1.6	Configurations de triangles dans un graphe orienté	6
1.7	Structure en noeud papillon	7
1.8	Axiome des parallèles	11
1.9	Modèle de Poincaré	14
1.10	Triangles en géométries euclidienne et non euclidiennes	15
1.11	Modèles analytiques de l'espace hyperbolique	16
1.12	Modèle de l'hyperboloïde	17
1.13	Géodésiques dans différents modèles analytiques de l'espace hyperbolique	19
1.14	Plongement d'un graphe non orienté dans un espace métrique	21
1.15	Modèle de l'hyperboloïde tronqué et représentation abstraite	23
1.16	Modèle \mathcal{H}^2	24
1.17	Modèle \mathcal{S}^1	25
1.18	Plongement d'un graphe orienté dans un espace métrique	29
1.19	Modèle \mathcal{S}^1 orienté	30
1.20	Réciprocité fortuite et réciprocité délibérée	30
2.1	Directed-reciprocal \mathcal{S}^1 model	45
2.2	Embedding on the \mathbb{S}^1 circle with Mercator and diMercator	47
2.3	Procedure for validating the diMercator algorithm on empirical network datasets	48
2.4	Reproducing topological properties of an empirical network dataset	49
2.5	Procedure for comparing the diMercator algorithm to the Mercator algorithm	54
2.6	Comparison between Mercator and diMercator	55
2.7	Trophic incoherence of empirical network datasets	57
2.8	Non-normality of empirical network datasets	58
2.9	Triangle spectrum of empirical network datasets	59
2.10	Individual reciprocity of empirical network datasets	60
2.11	Singular value spectrum of empirical network datasets	62
2.12	Bow-tie structure schematic diagram	63
2.13	Bow-tie structure of empirical network datasets	71
B.1	Transitions compartimentales dans le modèle AME	85
B.2	Transitions compartimentales dans le modèle TAME	88
B.3	Diagramme de phase et évolution temporelle	91

Pour saisir le monde
d'aujourd'hui, nous usons d'un
langage qui fut établi pour le
monde d'hier. Et la vie du passé
nous semble mieux répondre à
notre nature, pour la seule raison
qu'elle répond mieux à notre
langage.

Antoine de Saint-Exupéry,
Terre des hommes (1939)

Remerciements

Je tiens à remercier le groupe Dynamica de m'avoir permis d'entreprendre ce projet de recherche. Le soutien constant d'Antoine a été inestimable et je crois sincèrement avoir eu la chance d'être accompagné par un superviseur d'exception. Antoine et Patrick savent adroite-ment diriger le groupe de recherche ; je salue leur dévouement, leur patience et leur compré-hension. Comme j'aime le dire, Dynamica ne se contente pas du travail convenable, mais vise l'excellence ; et c'est précisément cette excellence qui émane des directeurs du groupe et de ses étudiants. Je remercie Charles d'avoir été un précieux mentor lors de mes balbutiements en recherche. J'aimerais remercier tout particulièrement Simon, qui a été mon soutien moral tout au long de ces deux années. Je remercie également ma famille, en particulier ma mère, qui aime bien me rappeler qu'elle m'a porté dans son ventre. Je suis profondément reconnaissant pour son amour inconditionnel et éternel. Je remercie mon père de m'avoir nourri de réflexions scientifiques parmi les plus fécondes ; il ne reconnaîtra pas, en ce mémoire, uniquement le fruit du hasard et de la nécessité. Merci à Paï et Maï d'avoir généreusement veillé à rassasier ma panse tout au long de ces années d'études. Merci à ma chère Anne-Sophie, douce flamme tranquille et chaleureuse, qui éclaire mon chemin en ces soirées froides de fin d'automne où je rédige ces lignes. Enfin, j'accorde un léger sursis aux choses sérieuses pour remercier Josée di Stasio, qui m'a inspiré le nom de l'algorithme *diMercator*.

Avant-propos

Ce mémoire par article est divisé en deux chapitres. Le premier chapitre introduit les éléments essentiels relatifs au projet de recherche. Le second chapitre contient l'article en préparation *Topological Properties and Geometry of Empirical Directed Networks*. L'auteur principal de cet article est également l'auteur de ce mémoire. Il a développé l'algorithme, écrit le code, produit les résultats et rédigé l'article. Le coauteur, Antoine Allard, a contribué à l'élaboration et au développement du projet de recherche ainsi qu'à la rédaction de l'article. L'annexe B contient la recherche connexe effectuée sous la supervision de l'étudiant au doctorat Charles Murphy lors d'un stage d'été et durant le début de la maîtrise.

Nota bene. – Ce mémoire contient certaines références historiques, littéraires et artistiques. Un lectorat averti saura y reconnaître ma passion pour les sciences, les lettres et les arts. J'ai ainsi choisi d'écrire certaines parties dans un style plus lyrique que strictement scientifique. Comme le disait Louis J. Dubé, fondateur du groupe Dynamica : *Un mémoire sert à se souvenir*. J'y ai donc consigné tout ce dont je souhaitais me souvenir de ces dernières années. C'est ainsi que je conclus mon aventure physicienne, en vous souhaitant la plus agréable des lectures.

Introduction

La pandémie de Covid-19 a mis en lumière l'interconnectivité et la complexité de notre monde moderne. En quelques mois, un virus microscopique a réussi à se propager à l'échelle mondiale, bouleversant les systèmes de santé, les économies et les interactions sociales. Cette propagation rapide et les réponses variées des sociétés à cette crise illustrent parfaitement la dynamique des systèmes complexes. Les interactions entre individus, institutions et nations ont joué un rôle crucial dans l'évolution de la pandémie, démontrant comment des comportements émergents peuvent surgir de réseaux d'interactions apparemment simples. Cette crise sanitaire mondiale nous rappelle l'importance de comprendre les structures sous-jacentes de nos systèmes complexes, et c'est précisément cet enjeu que nous explorons à travers l'étude de la géométrie des réseaux complexes. C'est dans cette optique que s'inscrit le titre de ce mémoire, qui vise à explorer ces structures sous-jacentes.

Que l'on parle de colonies d'insectes, de l'Internet, en passant par le cerveau humain, le système immunitaire ou bien l'économie, ces systèmes apparemment disparates partagent un trait commun : la complexité. Les systèmes dits complexes présentent un comportement collectif émergent, interagissant avec leur environnement et démontrent une capacité d'adaptation qui favorise leur évolution et leur succès continu [1, 2]. Du latin *complexus*, signifiant ce qui est tissé ensemble, un système complexe est constitué d'un ensemble d'éléments individuels en interaction, manifestant un comportement global émergent qui lui permet d'interagir avec son environnement et d'évoluer de manière autonome. Ces systèmes peuvent surgir de diverses sources, qu'elles soient biologiques, technologiques ou sociales, et leurs comportements émergents sont également variés. Par exemple, la stabilité des écosystèmes face à des perturbations externes, le vol coordonné des oiseaux, les fonctions cognitives du cerveau ou encore la propagation des épidémies dans une population illustrent tous la manière dont les systèmes complexes sont fondamentalement plus que la somme de leurs parties [3, 4]. Autrement dit, leur dynamique collective dépasse la somme des comportements individuels. Le cerveau humain est un exemple frappant de ce type de système. Avec ses milliards de neurones interconnectés par des signaux électriques et chimiques via des synapses, le cerveau permet de mémoriser des souvenirs, de formuler des idées, et de contrôler le corps, tout en modifiant ses connexions neuronales en réponse à l'apprentissage et à l'expérience.

Comprendre les propriétés topologiques de ces structures d’interactions est crucial pour expliquer les propriétés émergentes des systèmes complexes et prévoir leur comportement. On les représente mathématiquement à l’aide de graphes dont les sommets sont les éléments et les arêtes les interactions. La théorie des graphes, par l’entremise des réseaux complexes, offre ainsi d’excellents modèles pour décrire de nombreux phénomènes impliquant plusieurs éléments en interaction les uns avec les autres. On peut notamment utiliser les réseaux complexes pour modéliser les chaînes alimentaires d’écosystèmes, pour décrire la topologie du World Wide Web et de l’Internet, pour mieux comprendre les réseaux sociaux ou pour visualiser l’interdépendance entre les protéines d’un organisme vivant. Parmi les nombreuses approches mathématiques pour modéliser ces systèmes [5, 6], celle ayant recours à une géométrie sous-jacente est particulièrement notable [7]. Cette approche postule que les sommets du graphe sont plongés dans un espace métrique où la distance entre deux sommets détermine la probabilité qu’ils soient connectés par une arête. À l’inverse, l’existence d’une arête entre deux sommets du graphe détermine la distance les séparant dans l’espace sous-jacent. Cette classe de modèles permet de reproduire plusieurs propriétés structurelles universelles observées dans les systèmes complexes réels [8–11], notamment en s’appuyant sur l’inégalité du triangle de l’espace métrique sous-jacent. Cette approche permet ainsi de cartographier la structure de systèmes complexes réels en inférant la position de leurs éléments dans l’espace de manière à reproduire la structure d’interaction originale avec haute probabilité [12–15]. Ces cartes révèlent immédiatement la complexité de ces structures et permettent l’utilisation d’outils géométriques pour mieux la caractériser. Il existe cependant une importante classe de systèmes complexes ne pouvant être cartographiés à l’aide de cette approche. En effet, l’approche décrite ci-dessus suppose que les interactions sont bidirectionnelles et donc que les arêtes dans le graphe sont non orientées. Toutefois, de nombreux systèmes complexes présentent des interactions directionnelles comme les synapses entre les neurones du cerveau ou la transmission d’agents infectieux entre individus. Un premier pas a été franchi en adaptant le formalisme mathématique existant pour les graphes orientés [16]. La prochaine étape, qui est l’objectif de ce projet de recherche, consiste à adapter une procédure d’inférence avancée [12] pour cartographier chaque élément de cette classe importante de systèmes complexes.

Ce mémoire se compose de deux chapitres. Le premier chapitre aborde les notions théoriques et mathématiques indispensables à ce projet de recherche. Il présente les concepts fondamentaux des systèmes complexes, la théorie des graphes, et la géométrie hyperbolique. Ce cadre théorique fournit les bases nécessaires à la compréhension de la modélisation des réseaux complexes et leur représentation dans des espaces métriques. Le deuxième chapitre consiste en un article de recherche décrivant en détail l’algorithme développé au cours de ce projet. Il inclut une présentation de l’algorithme, les détails techniques de son implémentation, ainsi que les résultats obtenus à partir de différents ensembles de données empiriques. Une conclusion synthétise les principales contributions de ce projet, analyse les limitations rencontrées et propose des perspectives de futurs développements.

Chapitre 1

Notions préliminaires

1.1 Les systèmes et les réseaux complexes

Un système complexe est composé d'un ensemble d'éléments en interaction, où ces interactions donnent naissance à des comportements collectifs émergents qui ne peuvent être entièrement prévus ni compris en étudiant chaque élément séparément. Ces systèmes sont souvent caractérisés par une grande diversité de leurs éléments, une adaptation dynamique, et une sensibilité aux conditions initiales. Parmi les exemples de systèmes complexes, on trouve notamment les écosystèmes, les marchés financiers, les réseaux sociaux, tout comme de nombreux autres phénomènes naturels et sociaux. La complexité réside ainsi dans le fait que le comportement global du système ne peut être réduit à la somme de ses parties, et qu'il existe souvent des structures auto-organisées qui rendent leur analyse particulièrement exigeante et tout à fait fascinante [17].

1.1.1 Les réseaux complexes et la théorie des graphes

En science de la complexité, les réseaux permettent d'étudier la structure des systèmes complexes en modélisant les relations existant entre leurs éléments. On représente alors les éléments d'un système par des *noeuds* et les interactions entre eux par des *liens*. Dans le cadre de ce travail, on fait la distinction entre un réseau complexe et un graphe tel que le réseau est un graphe auquel on applique une réalité objective, alors que le graphe ne se réfère qu'à l'objet mathématique permettant de décrire un tel réseau. En termes plus techniques, les réseaux sont composés de noeuds et de liens, tandis que les graphes sont constitués de sommets et d'arêtes, ou encore d'arcs.

1.1.2 Le graphe

On définit mathématiquement un graphe G comme un couple (V, E) formé d'un ensemble de sommets V et d'un ensemble d'arêtes $E \subseteq \{\{u, v\} | u, v \in V, u \neq v\}$, lesquelles relient

les sommets entre eux [18]. Pour noter une arête, on utilise les accolades, indiquant que $\{u, v\} = \{v, u\}$, c'est-à-dire que l'ordre des sommets n'a pas d'importance. Le graphe présenté à la figure 1.1 comporte cinq sommets et cinq arêtes. L'objet *graphe* est très utile en science des systèmes et des réseaux complexes puisqu'il permet, entre autres, de caractériser la structure du système. Par exemple, le degré d'un sommet est défini comme son nombre de voisins, c'est-à-dire le nombre de sommets qui lui sont reliés par une arête. Également, on définit l'agrégation comme la mesure du nombre de triangles, où un triangle est constitué de trois arêtes reliant trois sommets distincts. Par exemple, dans le graphe de la figure 1.1, on ne retrouve qu'un seul triangle.



FIGURE 1.1 – Représentation d'un graphe à cinq sommets et cinq arêtes.

1.1.3 Le graphe orienté

La définition précédente d'un graphe étant assez simple, elle ne permet pas de résoudre tous les problèmes posés par la science des systèmes et des réseaux complexes, notamment ceux impliquant des interactions asymétriques. Toutefois, il existe une classe de graphes, les graphes orientés, qui permettent de décrire la nature symétrique ou asymétrique d'une interaction entre deux éléments.



FIGURE 1.2 – Représentation d'un graphe orienté à cinq sommets et sept arcs, dont trois arcs en bleu (représentant les interactions unidirectionnelles) et deux paires d'arcs de sens opposés en orange (représentant les interactions bidirectionnelles).

Dans un graphe orienté, les arêtes deviennent des flèches que l'on nomme arcs et ces arcs possèdent une orientation, représentée par la tête de flèche, conformément au graphe de la figure 1.2. Contrairement au graphe non orienté, où seule l'existence d'une interaction est prise en compte, représentée par une arête entre deux sommets, le graphe orienté, plus spécifique, retient la nature de l'interaction dans l'orientation de l'arc reliant deux sommets. Cette interaction est soit unidirectionnelle, comme les arcs en bleu dans le graphe de la figure 1.2, ou bidirectionnelle (également dite réciproque), comme les paires d'arcs de sens opposés en orange. On définit un graphe orienté G comme un couple (V, E) formé d'un ensemble de sommets V et d'un ensemble d'arcs $E \subseteq \{(u, v) | u, v \in V, u \neq v\}$, lesquels relient les sommets

entre eux [18]. Pour noter un arc, on utilise les parenthèses, indiquant que $(u, v) \neq (v, u)$, c'est-à-dire que l'ordre des sommets revêt de l'importance.

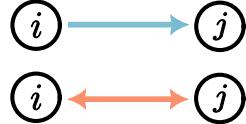


FIGURE 1.3 – Types d’interaction entre deux sommets dans un graphe orienté. Haut : interaction unidirectionnelle représentée par un arc du sommet i au sommet j . Bas : interaction bidirectionnelle représentée par une paire d’arcs de sens opposés (arc du sommet i au sommet j et du sommet j au sommet i).

On retient donc deux types d’interaction dans un graphe orienté : l’interaction unidirectionnelle représentée par un arc et l’interaction bidirectionnelle représentée par une paire d’arcs de sens opposés comme illustré à la figure 1.3. Il est cependant intéressant de noter que l’interaction bidirectionnelle est parfois illustrée comme deux arcs distincts de sens opposés tel que montré à la figure 1.4.

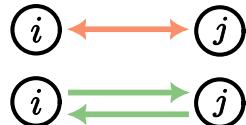


FIGURE 1.4 – Interaction bidirectionnelle entre deux sommets dans un graphe orienté.

Cela nous permet de définir une nouvelle propriété dans le graphe orienté, appelée réciprocité, qui évalue la proportion d’interactions réciproques ou bidirectionnelles dans le graphe. Intuitivement, la réciprocité d’un graphe orienté peut être définie de deux manières. La première consiste à diviser le nombre de paires d’arcs de sens opposés par le nombre total de paires de sommets connectés, autrement dit, à diviser le nombre d’interactions bidirectionnelles par le nombre total d’interactions entre deux sommets. La deuxième approche consiste à compter séparément les deux arcs composant chaque interaction bidirectionnelle, puis à diviser cette quantité par le nombre total d’arcs. [6, 19]. En exprimant le nombre de paires d’arcs de sens opposés par la variable L^{\leftrightarrow} et le nombre total de paires de sommets connectés par L , la réciprocité, notée r , peut être calculée selon les formules suivantes :

$$r_1 = \frac{L^{\leftrightarrow}}{L}, \quad (1.1)$$

$$r_2 = \frac{2L^{\leftrightarrow}}{L^{\leftrightarrow} + L}. \quad (1.2)$$

On exprime la réciprocité r_1 en fonction de la réciprocité r_2 telle que

$$r_1 = \frac{r_2}{2 - r_2}. \quad (1.3)$$

On constate que $r_1 = r_2$ lorsque $L^{\leftrightarrow} = 0$ et $L^{\leftrightarrow} = L$ uniquement, et que $r_1 \leq r_2$ sur le domaine $[0, 1]$ respectif des réciprocités r_1 et r_2 . En guise d'exemple, considérons le graphe illustré à la figure 1.5 : la réciprocité r_1 se rapporte au panneau (a) et se calcule telle que $r_1 = 1/3$ alors que la réciprocité r_2 se rapporte au panneau (b) et se calcule telle que $r_2 = 1/2$. Par un raisonnement similaire, on peut également définir la réciprocité individuelle comme la valeur de réciprocité calculée individuellement pour chaque sommet du graphe selon la nature de ses interactions avec ses voisins.

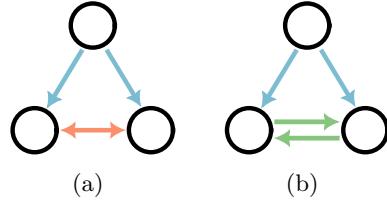


FIGURE 1.5 – Exemple de graphe orienté illustré de deux manières.

Parmi les propriétés importantes des graphes orientés, on retrouve également le degré d'un sommet, qui se décline en un degré entrant (nombre d'arcs dirigés vers le sommet) et un degré sortant (nombre d'arcs dirigés hors du sommet). La propriété d'agrégation, caractérisant le nombre de triangles présents dans le graphe, reste également applicable.

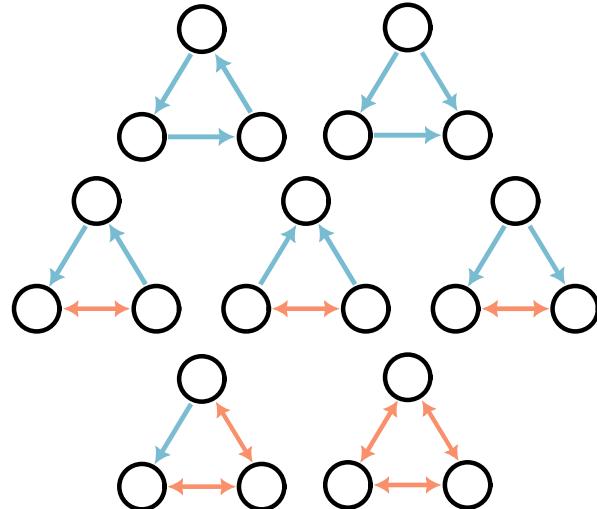


FIGURE 1.6 – Sept configurations de triangles dans un graphe orienté.

Toutefois, puisque nous avons maintenant deux types d'interaction, l'interaction unidirectionnelle et l'interaction bidirectionnelle, il ne s'agit plus d'un seul type de triangle, mais de sept configurations distinctes, conformément aux triangles de la figure 1.6.

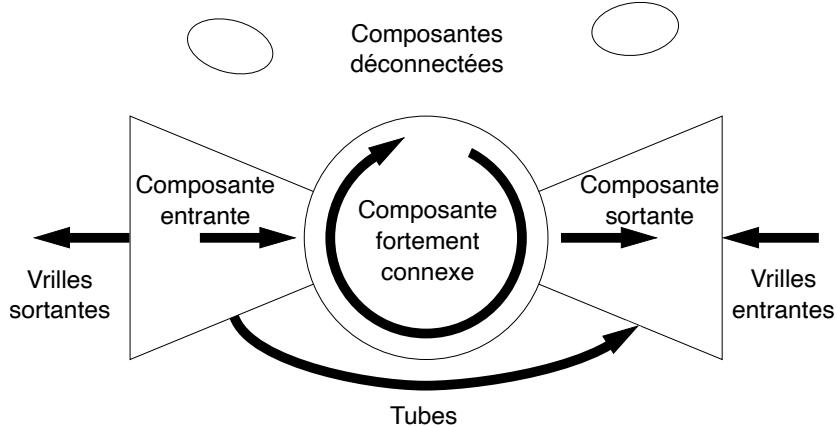


FIGURE 1.7 – Structure en noeud papillon : composantes d'un graphe orienté.

Il est également possible de décomposer un graphe orienté selon ses différentes composantes tel qu'il est montré à la figure 1.7. On nomme cette décomposition la structure en noeud papillon [20]. La composante fortement connexe correspond à l'ensemble des sommets tel que chacun puisse atteindre tous les autres et soit accessible à partir de tous les autres dans l'ensemble suivant un chemin orienté. À chaque composante fortement connexe correspond une composante sortante (l'ensemble de tous les sommets qui peuvent être atteints suivant un chemin orienté à partir de n'importe quel sommet de départ dans la composante fortement connexe) et une composante entrante (l'ensemble des sommets à partir desquels la composante fortement connexe peut être atteinte). Un graphe orienté typique est constitué d'une grande composante fortement connexe et de plusieurs petites composantes, chacune avec une composante entrante et une composante sortante. Par définition, chaque composante entrante ou sortante exclut la composante fortement connexe. Les vrilles sortantes et entrantes contiennent des sommets qui sont accessibles depuis des portions de la composante entrante, ou qui peuvent atteindre des portions de la composante sortante, sans passer par la composante fortement connexe. Il est également possible qu'une vrille sortante soit reliée à une vrille entrante, formant ainsi un tube [6, 20, 21].

1.1.4 Représentation matricielle et propriétés connexes

Il est possible de représenter un graphe non orienté à N sommets par sa matrice d'adjacence : matrice A de taille $N \times N$ dont l'élément $a_{ij} = a_{ji} = 1$ si les sommets i et j sont reliés par une arête et $a_{ij} = a_{ji} = 0$ autrement. Notez que la matrice d'adjacence d'un tel graphe est toujours symétrique. Dans le cas du graphe orienté, on définit la matrice d'adjacence telle que l'élément de matrice $a_{ij} = 1$ s'il existe un arc de i vers j et $a_{ij} = 0$ autrement. Contrairement au cas du graphe non orienté, la matrice d'adjacence d'un graphe orienté n'est pas forcément symétrique.

Décomposition en valeurs singulières

On peut notamment étudier les propriétés de la matrice d'adjacence d'un graphe en effectuant une décomposition en valeurs singulières (aussi appelée SVD) de cette matrice. On considère donc une matrice réelle A de taille $N \times N$ où N correspond au nombre de sommets du graphe. On résout les deux problèmes aux valeurs propres

$$AA^\top \mathbf{u}_i = \lambda_i \mathbf{u}_i \quad \text{et} \quad A^\top A \mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i \in \{1, \dots, N\}, \quad (1.4)$$

où \top dénote la transposée. Notez que les matrices AA^\top et $A^\top A$ sont carrées, symétriques, semi-définies positives et partagent les mêmes valeurs propres non nulles. Le résultat de leur diagonalisation permet d'écrire la décomposition en valeurs singulières de A telle que

$$A = U\Sigma V^\top, \quad (1.5)$$

où U est une matrice carrée orthogonale et ses colonnes correspondent aux vecteurs propres de AA^\top , V est une matrice carrée orthogonale et ses colonnes correspondent aux vecteurs propres de $A^\top A$, et Σ est une matrice diagonale. On dénote les valeurs singulières de A comme les éléments $\sigma_i = \sqrt{\lambda_i}$ pour $i = 1, 2, \dots, N$ de la matrice Σ . Les valeurs singulières d'une matrice sont toujours réelles et non négatives ; on les ordonne habituellement en ordre décroissant de leur valeur tel que $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N$ [22].

La SVD est particulièrement utile en science des systèmes et des réseaux complexes. En effet, la comparaison du spectre en valeurs singulières de deux réseaux peut fournir des informations sur leurs similitudes et leurs différences structurelles. Notamment, le spectre sert de représentation compacte de la matrice d'adjacence et reflète la dimensionnalité du réseau. Plus précisément, il a été montré que la décroissance des valeurs singulières est un indicateur important de la dimension effective d'un réseau et que cette décroissance est typiquement rapide pour les réseaux empiriques [23]. Ainsi, bien que les valeurs singulières ne soient pas des identifiants uniques (puisque différentes matrices peuvent partager le même spectre de valeurs singulières), elles servent de signature interprétable de la complexité de la matrice d'adjacence et du système qu'elle représente.

Cohérence et incohérence trophique

La cohérence trophique (ou incohérence trophique) d'un graphe est liée à la notion de niveaux trophiques en écologie [24–28] et permet de décrire la structure hiérarchique d'un graphe. Toutefois, la définition standard des niveaux trophiques présente certaines limites, notamment la nécessité d'avoir des sommets sources (c'est-à-dire des sommets sans arcs entrants), ce qui restreint son applicabilité. Heureusement, les définitions améliorées du niveau trophique et de l'incohérence proposées par Mackay et al. [28] surmontent ces limitations. Pour chaque sommet $j \in V$ d'un graphe orienté, on définit son degré entrant (nombre d'arcs dirigés vers

le sommet) et son degré sortant (nombre d'arcs dirigés hors du sommet) par

$$k_j^- = \sum_{i \in V} a_{ij} \quad \text{et} \quad k_j^+ = \sum_{i \in V} a_{ji}. \quad (1.6)$$

Nous définissons le degré total du sommet j par

$$u_j = k_j^- + k_j^+, \quad (1.7)$$

et le déséquilibre du sommet j par

$$v_j = k_j^- - k_j^+. \quad (1.8)$$

Nous formons ensuite les vecteurs \mathbf{u} et \mathbf{v} à partir des composantes u_j et v_j respectivement. L'opérateur Laplacien du graphe Λ est défini comme

$$\Lambda = \text{diag}(\mathbf{u}) - A - A^\top, \quad (1.9)$$

et agit sur un vecteur \mathbf{h} selon

$$(\Lambda \mathbf{h})_i = u_i h_i - \sum_{j \in V} (a_{ij} + a_{ji}) h_j. \quad (1.10)$$

Les niveaux trophiques sont alors donnés par la solution \mathbf{h} du système d'équations linéaires

$$\Lambda \mathbf{h} = \mathbf{v}. \quad (1.11)$$

Le système (1.11) admet toujours une solution, mais celle-ci n'est pas unique, car le système est sous-déterminé. Ainsi, pour obtenir une solution unique, on remplace l'équation correspondant à un sommet i_S dans chaque composante fortement connexe $S \subset V$ du graphe par une équation $h_{i_S} = c_S$, pour des constantes arbitraires c_S , par exemple $c_S = 0$. Il existe alors une solution unique pour \mathbf{h} . Cette solution nous permet de définir le paramètre d'incohérence trophique tel que

$$F_0 = \frac{\sum_{ij} a_{ij} (h_j - h_i - 1)^2}{\sum_{ij} a_{ij}}, \quad (1.12)$$

en utilisant les niveaux trophiques \mathbf{h} déterminés ci-dessus. Remarquons que F_0 est strictement compris entre 0 et 1. Un graphe est dit maximalement cohérent s'il satisfait $F_0 = 0$, et maximalement incohérent s'il satisfait $F_0 = 1$. Notez que le choix arbitraire des constantes c_S ne change en rien la valeur de F_0 .

Normalité et non-normalité

Nous qualifions un graphe de normal lorsque sa matrice d'adjacence A est normale [29, 30], c'est-à-dire qu'elle commute avec sa transposée tel que

$$AA^\top = A^\top A. \quad (1.13)$$

En ce qui a trait aux graphes orientés, le caractère orienté des arcs dans le graphe rend la matrice d'adjacence A possiblement asymétrique et non normale. La non-normalité peut ainsi être quantifiée à l'aide d'une mesure spectrale standard issue de la théorie des matrices, telle que l'écart à la normalité d'Henrici [31],

$$d_F(A) = \sqrt{\|A\|_F^2 - \sum_{i=1}^N |\lambda_i|^2}, \quad (1.14)$$

où $\|\cdot\|_F$ dénote la norme de Frobenius et λ_i est déterminée par le problème aux valeurs propres

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i \in \{1, \dots, N\}. \quad (1.15)$$

Une valeur nulle de l'équation (1.14) correspond donc à un réseau normal, tandis que des valeurs plus élevées indiquent une non-normalité plus forte. Dans le cadre de ce travail, on normalise l'équation (1.14) pour que ses valeurs soient comprises dans l'intervalle $[0, 1]$, ce qui donne

$$\tilde{d}_F(A) = \sqrt{1 - \frac{\sum_{i=1}^N |\lambda_i|^2}{\|A\|_F^2}}. \quad (1.16)$$

On référera ainsi à l'équation (1.16) comme la non-normalité.

1.2 Point sur la géométrie hyperbolique

Maintenant que nous avons couvert certaines notions relatives aux systèmes et aux réseaux complexes, nous pouvons nous intéresser au deuxième volet essentiel à ce projet de recherche, à savoir la géométrie hyperbolique.

1.2.1 Une géométrie née du hasard et de la nécessité

Afin de bien comprendre la géométrie hyperbolique, il est d'abord pertinent de se demander ce qui définit une géométrie ; non pas d'un point de vue formel et mathématique, mais plutôt d'un oeil ontologique, d'un oeil humain. Il est de mon avis que le développement de la géométrie puisse se justifier par le désir de l'être humain à rendre compte de l'espace physique qui l'entoure. Il n'y a alors rien de surprenant au fait que la géométrie hyperbolique soit apparue relativement récemment dans notre Histoire alors que les Grecs de l'Antiquité avaient déjà une idée tout à fait formelle et cohérente de la géométrie du monde, la géométrie plane, la géométrie d'Euclide. Dans *La République* [32], Platon s'exprime ainsi quant à la géométrie :

« Qu'elle a pour objet la connaissance de ce qui est toujours et non de ce qui naît et périt. » (Platon, *La République* VII, 527a, vers 380 av. J.-C. [32])

Dans la perspective de célébrer *la connaissance de ce qui est toujours*, ce qui suit se veut une histoire brève, quoique incomplète, et sans aucune prétention d'être adroûtement racontée, de la géométrie (hyperbolique) ainsi qu'une leçon sur comment (ne pas) éviter un trou de lapin.

Les axiomes d'Euclide

Vers 300 av. J.-C., Euclide rédige les *Éléments* [33], un traité mathématique et géométrique comportant une collection de définitions, théorèmes et démonstrations, mais également les axiomes de la géométrie, qui, à ce jour, permettent de définir ce que l'on appelle la géométrie euclidienne, la géométrie qui nous est familière et qui nous est enseignée à l'école. On paraphrase les cinq axiomes d'Euclide comme suit :

1. Une droite peut être tracée entre deux points distincts.
2. Un segment de droite peut être prolongé indéfiniment.
3. Un cercle peut être tracé en prenant une droite comme rayon et son extrémité comme centre.
4. Tous les angles droits sont égaux.

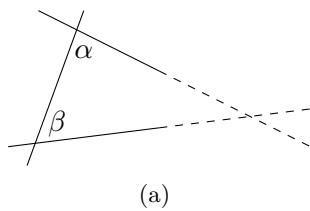
Le cinquième axiome est absent de cette énumération pour la raison qu'il a troublé les mathématiciens pendant plusieurs siècles et mérite donc son propre chapitre dans cette histoire. Avant de poursuivre, prenons un moment pour discuter de ces quatre premiers axiomes et de leur importance en géométrie, mais aussi dans le développement de la pensée scientifique et dans l'Histoire de l'Humanité. La géométrie d'Euclide a été enseignée pendant des siècles comme le paradigme de la pensée logique et rationnelle [34]. Notamment, les axiomes d'Euclide ont profondément influencé les philosophes et les scientifiques comme Descartes dans le développement de la géométrie analytique ou Newton dans les fondations de ses lois de la mécanique [35].

L'axiome des parallèles

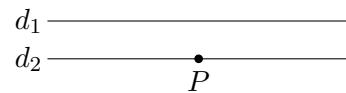
Le cinquième axiome, aussi appelé l'axiome des parallèles, est énoncé comme suit :

« Si une droite tombant sur deux droites fait les angles intérieurs du même côté plus petits que deux angles droits, ces droites, prolongées à l'infini, se rencontreront du côté où les angles sont plus petits que deux angles droits. »

(Euclide, *Éléments*, vers 300 av. J.-C. [33])



(a)



(b)

FIGURE 1.8 – Illustration du cinquième axiome de la géométrie (a) tel que présenté par Euclide et (b) paraphrasé.

La figure 1.8 illustre l’axiome des parallèles. Bien que cet axiome puisse sembler complexe à interpréter au premier abord, on le paraphrase ainsi : par un point P extérieur à une droite d_1 , il passe toujours une seule parallèle d_2 à cette droite.

Ce dernier axiome est sans équivoque le vilain petit canard de la couvée. Difficile à interpréter dans son énonciation initiale, il fascine comme il dérange les mathématiciens et les géomètres qui daignent s’y attarder. C’est pourtant à cette étrange obsession pour l’axiome des parallèles que l’on doit l’émergence de la géométrie hyperbolique.

Euclides ab omni naevo vindicatus

Il faut attendre le Siècle des Lumières pour enfin espérer laver Euclide de toute tache. Durant les siècles précédents, on tente sans grand succès de dériver le cinquième axiome à partir des quatre autres ou tout simplement de raisonner par l’absurde afin d’invalider cet axiome de malheur qui gêne encore et toujours tout grand mathématicien. C’est pourtant durant ce siècle d’effervescence intellectuelle que, sans même se rendre compte de sa découverte, un mathématicien italien du nom de Giovanni Girolamo Saccheri dresse les bases de la géométrie hyperbolique, en faisant fi du cinquième axiome, dans un écrit qui sera oublié jusqu’à sa redécouverte au XIX^e siècle [36]. Saccheri rejette ses propres travaux, estimant que ses résultats contredisent le sens commun et sont incompatibles avec les idées reçues de l’époque. Parmi ces étranges résultats, on retrouve notamment un triangle dont la somme des angles intérieurs est inférieure à π , comble de misère puisque la géométrie d’Euclide stipule que la somme des angles intérieurs d’un triangle est exactement égale à π . Parallèlement, certains mathématiciens des noms de Jean-Henri Lambert et Carl Friedrich Gauss, envers qui l’Histoire sera plus clémence, travaillent aussi sur le cinquième axiome en proposant qu’il existerait peut-être une autre géométrie que celle d’Euclide [37, 38].

Sur une géométrie imaginaire

Entrons dans le XIX^e siècle. En opposition au classicisme empreint de rationalité du siècle précédent, le romantisme de ce siècle des passions se reflète même jusque dans les idées mathématiques. Nous pourrions évidemment discuter des idées du professeur de mathématiques Charles Dodgson, mieux connu sous le nom de plume de Lewis Carroll, et de son pays des merveilles, mais ce n’est pas dans cette direction que notre histoire nous mène. C’est plutôt au mathématicien russe Nikolai Ivanovich Lobatchevski que l’on doit non pas un formalisme géométrique basé sur des axiomes, mais bien sur un ensemble de formules trigonométriques dans lequel le rayon de la sphère est un nombre imaginaire, imaginaire non pas tout droit sorti de l’imagination, mais bien par la nature même de ce nombre dans le plan complexe [39]. La géométrie imaginaire de Lobatchevski n’est alors qu’une curiosité mathématique sans importance pratique puisqu’elle est en totale opposition avec la géométrie réelle de l’espace physique [40]. Les résultats de Lobatchevski sont pourtant confirmés par le mathématicien

hongrois János Bolyai quelques années après leur publication. Au fil du siècle, on obtient graduellement une description de plus en plus étoffée de ces géométries dites *imaginaires*, où plutôt, non euclidiennes [41]. Notamment, le développement de la géométrie différentielle par le mathématicien allemand Bernhard Riemann permet de mieux cerner ces géométries.

Vers des modèles hyperboliques

Comme il a été raconté précédemment dans cette brève histoire tout à fait maladroite, les résultats de Saccheri sont redécouverts à la fin du XIX^e siècle par Eugenio Beltrami, en même temps que l'on poursuit les travaux de Bolyai, Gauss et Riemann sur les géométries non euclidiennes. Entre autres, Beltrami discute de ces nouvelles géométries dans son *Essai d'interprétation de la géométrie non euclidienne* et le mathématicien allemand Felix Klein publie son programme d'Erlangen, programme de recherche sur la géométrie visant à comparer les géométries découvertes dans le dernier siècle et d'en relever les similitudes ainsi que les différences de manière à mieux les décrire individuellement et d'évaluer dans quelle mesure elles se chevauchent [42]. Il semble tranquillement y avoir convergence d'idées en ce qui a trait à la nature de l'axiome des parallèles et de sa relation avec la géométrie ou plutôt *les* géométries admises par les mathématiques. C'est pourtant au philosophe, physicien et mathématicien français Henri Poincaré que l'on doit enfin la preuve de la contingence du cinquième axiome dont son énoncé peut varier selon la géométrie pour laquelle on le postule.

Une nouvelle vision du monde

Au tournant du XX^e siècle, Poincaré dresse un portrait à l'intuition physique de la géométrie hyperbolique qu'il décrit comme suit dans son ouvrage *La Science et l'Hypothèse* :

« Supposons, par exemple, un monde renfermé dans une grande sphère et soumis aux lois suivantes : la température n'y est pas uniforme ; elle est maxima au centre, et elle diminue à mesure qu'on s'en éloigne, pour se réduire au zéro absolu quand on atteint la sphère où ce monde est renfermé. [...] Un objet mobile deviendra alors de plus en plus petit à mesure qu'on se rapprochera de la sphère limite. Observons d'abord que, si ce monde est limité au point de vue de notre géométrie habituelle, il paraîtra infini à ses habitants. Quand ceux-ci, en effet, veulent se rapprocher de la sphère limite, ils se refroidissent et deviennent de plus en plus petits. Les pas qu'ils font sont donc aussi de plus en plus petits, de sorte qu'ils ne peuvent jamais atteindre la sphère limite. »

(Poincaré, *La Science et l'Hypothèse*, 1902 [43])

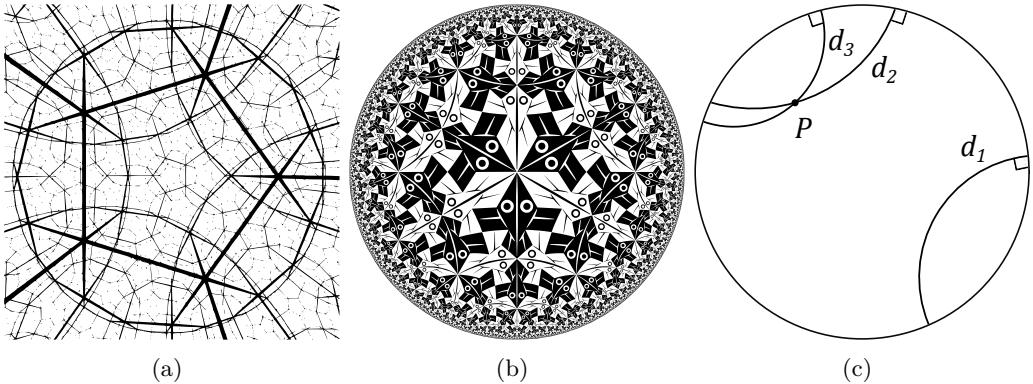


FIGURE 1.9 – (a) Représentation de la sphère de Poincaré à l'aide d'un pavage de l'espace hyperbolique par des dodécaèdres. (b) Représentation du disque de Poincaré (équivalent de la sphère de Poincaré en deux dimensions), *Cercle limite I* par M. C. Escher [44]. (c) Représentation de l'axiome des parallèles en géométrie hyperbolique (sur le disque de Poincaré).

La panneau (a) de la figure 1.9 montre une représentation imaginée de la sphère décrite par Poincaré dans *La Science et l'Hypothèse*. De la même idée, le panneau (b) de la figure 1.9 présente une tranche de cette même sphère pour laquelle on remarque que les illustrations de poissons deviennent de plus en plus petites à mesure que l'on se rapproche du cercle limite du disque. Comme le mentionne Poincaré, les poissons nous semblent devenir de plus en plus petits puisque nous vivons dans un monde à géométrie plane, alors qu'ils sont tous exactement de la même taille pour les habitants de ce monde à géométrie hyperbolique où le disque semble infini puisqu'il grandit exponentiellement. L'artiste néerlandais M. C. Escher, à l'origine de cette illustration du disque de Poincaré réalisée en 1958, s'exprime ainsi :

« Ici, il n'y a plus que des séries en circulation transitoire : tous les poissons de la même série [...] se suivent en nageant, les têtes touchant [les têtes], suivant une trajectoire de bord à bord. Plus ils approchent du centre, plus ils grandissent. [...] Aucun de ces éléments de toutes ces séries ascendantes, telles des fusées s'élevant perpendiculairement à la limite pour s'y perdre de nouveau, n'atteindra jamais celle-ci. Mais au dehors, il y a le néant absolu. »

(Traduction d'Escher, *Le miroir magique de M. C. Escher*, 1994 [45, 46])

Enfin, s'appuyant notamment sur les résultats de Beltrami et de Riemann, l'interprétation de Poincaré permet finalement de résoudre le millénaire mystère du cinquième axiome d'Euclide. La panneau (c) de la figure 1.9 présente une illustration de l'axiome des parallèles revisité pour la géométrie hyperbolique : par un point P extérieur à une droite d_1 , il passe une infinité de parallèles à cette droite ; deux de celles-ci sont montrées.

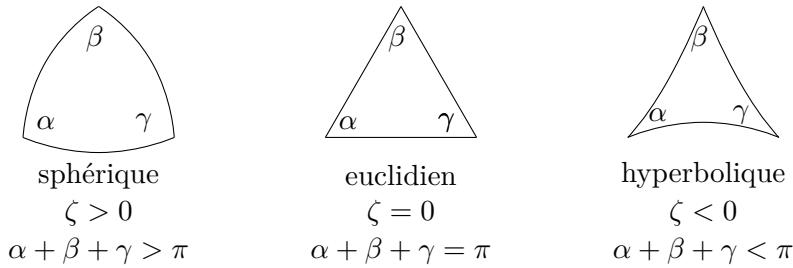


FIGURE 1.10 – Représentation des triangles en géométries euclidienne et non euclidiennes à courbure scalaire ζ constante dans \mathbb{R}^2 .

La figure 1.10 présente les triangles que l'on retrouve dans les trois géométries euclidienne et non euclidiennes à courbure scalaire constante. Ces trois géométries sont ainsi le fruit millénaire du travail acharné de nombreux mathématiciens. On remarque que les triangles sont courbés vers l'extérieur en géométrie sphérique alors qu'ils sont courbés vers l'intérieur en géométrie hyperbolique, ce qui illustre la courbure respective de ces géométries courbes ainsi que les relations trigonométriques qui leur sont propres.

Dans une coquille de noix

Au XIXe siècle, la remise en question du cinquième axiome d'Euclide, l'axiome des parallèles, conduit à l'émergence des géométries non euclidiennes, telles que la géométrie hyperbolique et la géométrie sphérique. Ces nouvelles géométries témoignent que les axiomes d'Euclide ne sont pas les seules bases géométriques admises par les mathématiques, élargissant ainsi la compréhension des structures spatiales et ouvrant la voie à des théories modernes [34, 35]. La découverte des géométries non euclidiennes transforme les mathématiques en introduisant de nouveaux concepts d'espace, et influence également la physique, notamment dans le cadre de la théorie de la relativité générale d'Einstein, qui repose sur une conception non euclidienne de l'espace-temps [47]. En guise de conclusion à cette brève histoire de la géométrie hyperbolique, quoi de mieux qu'une paraphrase d'un des grands dramaturges de notre époque : que celui qui étudie la géométrie hyperbolique, au même titre que le Prince Hamlet de Shakespeare, même enfermé dans une coquille de noix, se sente comme le roi d'un espace infini [48].

1.2.2 Modèles analytiques de l'espace hyperbolique

L'espace hyperbolique est un espace à courbure constante négative, contrairement à l'espace euclidien qui est plat ou encore à l'espace sphérique dont la courbure est constante et positive. On définit l'espace hyperbolique, à isométrie près, comme l'unique variété riemannienne complète, simplement connexe, de dimension n et de courbure sectionnelle constante égale à -1 [49–51]. L'espace de plus petite dimension est le plan hyperbolique \mathbb{H}^2 , une surface lisse qui peut être modélisée comme une nappe d'un hyperboloïde dans l'espace tridimensionnel

de Minkowski, mais aussi en utilisant d'autres modèles équivalents comme le disque de Klein, l'hémisphère ainsi que le disque ou le demi-plan de Poincaré [52]. Ces différents modèles du plan hyperbolique permettent de visualiser l'espace sous différents angles, où chaque modèle admet une intuition propre et des propriétés comme une métrique, qui définit une notion de distance entre les éléments d'un espace, et des géodésiques, qui sont les trajectoires de distance minimale entre deux points de l'espace [37, 53, 54].

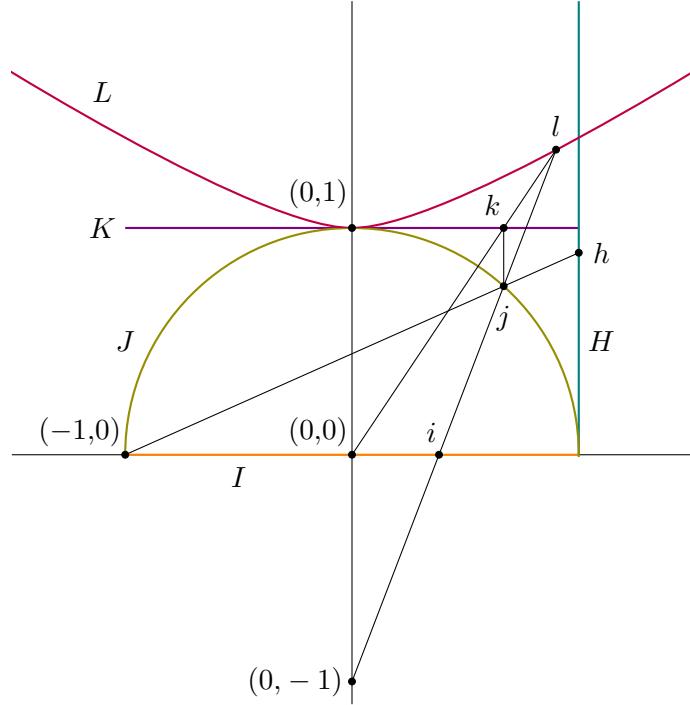


FIGURE 1.11 – Représentation graphique d'une coupe transversale des différents modèles de l'espace hyperbolique en plus haute dimension : l'hyperboloïde L (rouge), le disque de Klein K (violet), l'hémisphère J (vert), le disque de Poincaré I (orange) et le demi-plan de Poincaré H (bleu). Les droites correspondent aux isométries entre les différents modèles avec les points l , k , j , i et h équivalents dans chacun des modèles.

La visualisation de l'espace hyperbolique est tout de même un défi de taille pour quiconque s'intéressant à cette géométrie non euclidienne. De part sa courbure négative, les distances grandissent plus rapidement dans l'espace hyperbolique que dans l'espace euclidien. Par exemple, si l'on trace un disque de rayon arbitraire dans le plan hyperbolique \mathbb{H}^2 ainsi qu'un disque de même rayon dans le plan euclidien \mathbb{R}^2 , on remarque que l'aire du disque hyperbolique est plus grande que l'aire du disque euclidien. L'espace hyperbolique est ainsi *plus grand* que l'espace euclidien. Malgré le caractère fort abstrait de la géométrie hyperbolique, des modèles analytiques de l'espace hyperbolique existent et permettent de décrire adéquatement cette géométrie non euclidienne.

Modèle de l'hyperboloïde

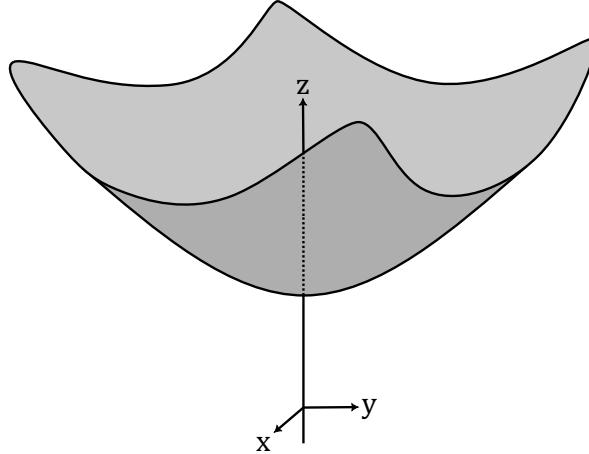


FIGURE 1.12 – Représentation du modèle de l'hyperboloïde.

Parmi les modèles analytiques de l'espace hyperbolique présentés à la figure 1.11, le modèle de l'hyperboloïde est illustré à la figure 1.12 pour $n = 2$ et est défini selon

$$L = \{(x_1, \dots, x_n, x_{n+1}) : x_1^2 + \dots + x_n^2 - x_{n+1}^2 = -1 \text{ et } x_{n+1} > 0\}, \quad (1.17)$$

avec comme métrique

$$ds_L^2 = dx_1^2 + \dots + dx_n^2 - dx_{n+1}^2. \quad (1.18)$$

Ses géodésiques sont des branches d'hyperbole ; une hyperbole étant constituée de deux branches disjointes symétriques l'une de l'autre et possédant deux asymptotes communes. En relativité restreinte, l'espace-temps de Minkowski combine les trois dimensions spatiales et le temps en une seule entité à quatre dimensions. La métrique de Minkowski, donnée par

$$ds^2 = dx^2 + dy^2 + dz^2 - dt^2, \quad (1.19)$$

où x , y , et z représentent les dimensions spatiales et t la dimension temporelle, ne suit pas les règles de la géométrie euclidienne, mais plutôt celles de la géométrie hyperbolique [47]. On constate que la métrique de l'équation (1.18) retombe sur la métrique de Minkowski de l'équation (1.19) lorsque $n = 3$.

Modèle du disque de Klein

Le modèle du disque de Klein est défini selon

$$K = \{(x_1, \dots, x_n, 1) : x_1^2 + \dots + x_n^2 < 1\}, \quad (1.20)$$

avec comme métrique

$$ds_K^2 = \frac{dx_1^2 + \dots + dx_n^2}{1 - x_1^2 - \dots - x_n^2} + \frac{(x_1 dx_1 + \dots + x_n dx_n)^2}{(1 - x_1^2 - \dots - x_n^2)^2}. \quad (1.21)$$

Ce modèle est non conforme, c'est-à-dire qu'il ne préserve pas les angles. En contrepartie, il préserve les lignes droites ; ses géodésiques sont des cordes (ou des lignes droites) sur le disque.

Modèle de l'hémisphère

Le modèle de l'hémisphère est défini selon

$$J = \{(x_1, \dots, x_{n+1}) : x_1^2 + \dots + x_{n+1}^2 = 1 \text{ et } x_{n+1} > 0\}, \quad (1.22)$$

avec comme métrique

$$ds_J^2 = \frac{dx_1^2 + \dots + dx_{n+1}^2}{x_{n+1}^2}. \quad (1.23)$$

Ses géodésiques sont des demi-cercles orthogonaux à la frontière de la demi-sphère.

Modèle du disque de Poincaré

Le modèle du disque de Poincaré est défini selon

$$I = \{(x_1, \dots, x_n, 0) : x_1^2 + \dots + x_n^2 < 1\}, \quad (1.24)$$

avec comme métrique

$$ds_I^2 = \frac{4(dx_1^2 + \dots + dx_n^2)}{(1 - x_1^2 - \dots - x_n^2)^2}. \quad (1.25)$$

Ce modèle est conforme, ce qui signifie qu'il préserve les angles. Ses géodésiques sont des diamètres et des arcs de cercle orthogonaux à la frontière du disque.

Modèle du demi-plan de Poincaré

Le modèle du demi-plan de Poincaré est défini selon

$$H = \{(1, x_2, \dots, x_{n+1}) : x_{n+1} > 0\}, \quad (1.26)$$

avec comme métrique

$$ds_H^2 = \frac{dx_2^2 + \dots + dx_{n+1}^2}{x_{n+1}^2}. \quad (1.27)$$

Ses géodésiques sont des lignes droites semi-infinies verticales et des demi-cercles orthogonaux à la frontière du demi-plan.

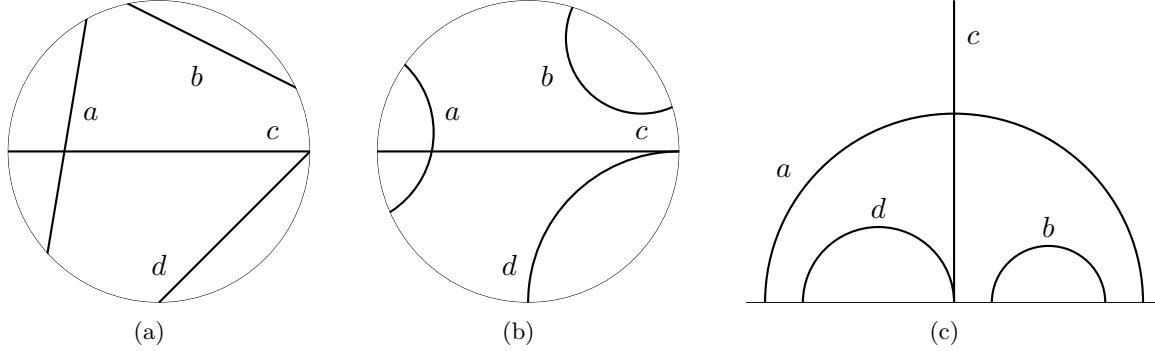


FIGURE 1.13 – Représentation des géodésiques dans différents modèles analytiques de l'espace hyperbolique. (a) Disque de Klein. (b) Disque de Poincaré. (c) Demi-plan de Poincaré.

Isométries entre les modèles analytiques de l'espace hyperbolique

Comme les modèles analytiques de l'espace hyperbolique sont équivalents, il existe des isométries entre eux [53, 54]. Il est donc possible de passer d'un modèle à un autre suivant la transformation associée à une isométrie comme il est montré à la figure 1.13. Ces isométries se déclinent selon les transformations suivantes :

La transformation $\alpha : J \rightarrow H$ est une projection centrale à partir du point $(-1, 0, \dots, 0)$:

$$\alpha : J \rightarrow H, (x_1, \dots, x_{n+1}) \mapsto (1, 2x_2/(x_1 + 1), \dots, 2x_{n+1}/(x_1 + 1)). \quad (1.28)$$

La transformation $\beta : J \rightarrow I$ est une projection centrale à partir du point $(0, \dots, 0, -1)$:

$$\beta : J \rightarrow I, (x_1, \dots, x_{n+1}) \mapsto (x_1/(x_{n+1} + 1), \dots, x_n/(x_{n+1} + 1), 0). \quad (1.29)$$

La transformation $\gamma : K \rightarrow J$ est une projection verticale :

$$\gamma : K \rightarrow J, (x_1, \dots, x_n, 1) \mapsto (x_1, \dots, x_n, (1 - x_1^2 - \dots - x_n^2)^{1/2}). \quad (1.30)$$

La transformation $\delta : L \rightarrow J$ est une projection centrale à partir du point $(0, \dots, 0, -1)$:

$$\delta : L \rightarrow J, (x_1, \dots, x_{n+1}) \mapsto (x_1/x_{n+1}, \dots, x_n/x_{n+1}, 1/x_{n+1}). \quad (1.31)$$

1.2.3 Les aveugles et l'éléphant

Il va de soi que chaque modèle précédemment présenté est une représentation de l'espace hyperbolique comportant ses caractéristiques propres, ses avantages et ses inconvénients. Ces modèles décrivent une seule et même réalité fort abstraite telle qu'elle est restée un mystère pendant de nombreux siècles. C'est du hasard et de la nécessité qu'elle s'est tranquillement dessinée comme un tableau se révélant peu à peu sous le pinceau de l'artiste. Par le hasard, conformément aux plus grandes découvertes de notre Histoire, et par la nécessité d'expliquer le cinquième axiome de la géométrie d'Euclide. Au même titre que la fable des aveugles et de

l’éléphant, où des aveugles doivent examiner un éléphant, mais ne touchant chacun qu’une partie de son corps, les modèles analytiques de l’espace hyperbolique permettent chacun de décrire, sous une lentille propre, cette fascinante géométrie. Longtemps restée une curiosité mathématique, la géométrie hyperbolique a tout de même certaines applications pratiques. Entre autres, l’espace hyperbolique est utilisé en modélisation des réseaux complexes dans l’étude de la structure des systèmes complexes.

1.3 Modélisation des réseaux complexes

Lorsqu’on souhaite étudier les réseaux complexes, plusieurs modèles existent et comportent chacun leurs avantages et leurs limitations. Entre autres, ces modèles peuvent être aléatoires, géométriques, etc. En termes de modèles aléatoires, un des exemples les plus simples est le modèle Erdős–Rényi proposé par les mathématiciens Paul Erdős et Alfréd Rényi en 1959 dans lequel on choisit aléatoirement un graphe parmi l’ensemble des graphes comportant un nombre donné de sommets et d’arêtes [6]. Étant plutôt rudimentaire, ce modèle ne permet pas de décrire adéquatement la structure de systèmes complexes réels ; il offre toutefois une perspective intéressante sur le potentiel de la modélisation des réseaux complexes ainsi que certaines limitations de l’approche. Par ailleurs, les modèles géométriques permettent de reproduire de nombreuses propriétés structurelles fondamentales des réseaux réels obtenus empiriquement tout en restant relativement simples mathématiquement [7]. Cette approche géométrique consiste à positionner chaque sommet du graphe dans une géométrie sous-jacente de manière à obtenir une représentation géométrique dudit graphe où on associe à chaque sommet une coordonnée spatiale. Dans le paradigme de la géométrie des graphes, la distance entre deux sommets détermine la probabilité qu’ils soient connectés par une arête et, à l’inverse, le fait qu’il existe ou non une arête entre deux sommets influence la distance qui les sépare dans la géométrie sous-jacente. Avant de poursuivre, il est nécessaire de bien comprendre ce qu’est un espace métrique et ce que signifie le fait de plonger un graphe dans un tel espace.

1.3.1 Les espaces métriques

Un espace métrique est un couple (M, d) où M est un ensemble non vide et d est une distance sur M définie comme

$$d : M \times M \rightarrow \mathbb{R}. \quad (1.32)$$

La distance $d(x, y)$ entre deux points x et y de l’ensemble M respecte les axiomes suivants.

1. Le principe d’identité des indiscernables : si deux objets x et y possèdent les mêmes propriétés, alors ils sont identiques, c’est donc dire que si l’on considère x et y comme des points de l’espace métrique, alors

$$d(x, y) = 0 \Leftrightarrow x = y. \quad (1.33)$$

2. La non-négativité de la distance :

$$d(x, y) \geq 0. \quad (1.34)$$

3. La symétrie de la distance : la distance de x à y est la même que celle de y à x ,

$$d(x, y) = d(y, x). \quad (1.35)$$

4. L'inégalité du triangle : si x et z sont près de y , alors x et z doivent conséquemment être près l'un de l'autre,

$$d(x, z) \leq d(x, y) + d(y, z). \quad (1.36)$$

C'est donc dire que la mesure d'un côté d'un triangle ne peut être plus grande que la somme des mesures des deux autres côtés. Lorsqu'on parle de théorie des graphes ou de réseaux complexes, où les sommets du triangle sont les éléments du système à l'étude, l'idée se résume bien par l'adage *les amis de mes amis sont mes amis*, où l'on sous-entend que la probabilité de connexion entre deux sommets est une fonction décroissante de la distance entre ces deux sommets.

1.3.2 Plongement dans un espace métrique hyperbolique

Tel que mentionné précédemment, le plongement d'un graphe dans un espace métrique consiste à positionner chaque sommet du graphe dans une géométrie de manière à obtenir une représentation géométrique dudit graphe. Dans la figure 1.14, chaque sommet a une coordonnée spatiale sur la surface sous-jacente. On peut donc mesurer la distance entre les sommets en suivant le plus court chemin admis par l'espace métrique entre les différentes paires de sommets connectés.

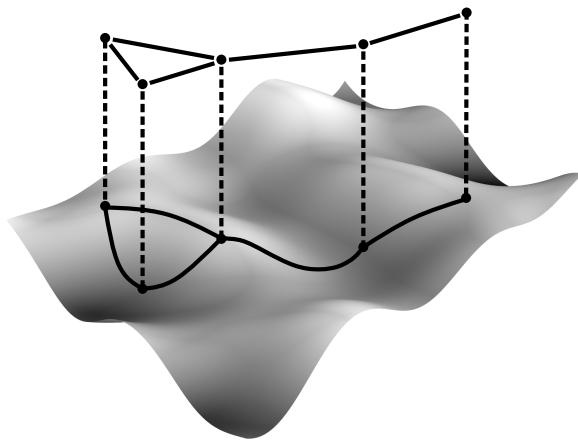


FIGURE 1.14 – Plongement d'un graphe non orienté dans un espace métrique.

L'espace hyperbolique semble être l'espace métrique de choix quand vient le temps d'y plonger un graphe puisque cette géométrie est la seule géométrie connue sur laquelle les graphes aléatoires maximalement entropiques reproduisent simultanément des propriétés fondamentales des réseaux réels observés empiriquement [10, 52]. On dit d'un ensemble de graphes qu'il est maximalement entropique lorsqu'il maximise l'entropie de Shannon sous certaines contraintes données. Il inclut toutes les configurations possibles de graphes compatibles avec ces contraintes (e.g., le nombre de sommets ou d'arêtes, la distribution des degrés attendue, la réciprocité, etc.), et les graphes qui le composent sont aussi aléatoires que possible, c'est-à-dire dépourvus de toute structure additionnelle autre que celle imposée par les contraintes elles-mêmes [6]. Parmi les propriétés reproduites par ces graphes plongés dans l'espace hyperbolique, on retrouve notamment une distribution des degrés en loi de puissance, une présence importante d'agrégation, une structure en communautés ainsi que le caractère petit monde [7, 8, 10]. On dit d'un réseau qu'il a le caractère petit monde si la distance moyenne entre deux noeuds quelconques, en termes de liens, est de l'ordre du logarithme de la taille totale du réseau. Par ailleurs, une distribution des degrés en loi de puissance signifie qu'un petit nombre de noeuds sont très connectés, tandis que la majorité ont peu de connexions. Notamment, cette propriété permet d'étudier la résilience des réseaux aux perturbations correspondant au retrait ou à l'ajout de noeuds dans le réseau. D'autre part, un haut niveau d'agrégation reflète des processus comme la formation de communautés ou la présence d'interactions préférentielles. Cela donne au modèle une meilleure correspondance avec les réseaux observés empiriquement et permet d'étudier des phénomènes comme la diffusion d'information ou la propagation d'épidémies au sein de communautés [6].

Le plongement dans l'espace hyperbolique permet ainsi de visualiser et surtout d'étudier la structure du graphe puisque la position des sommets dans l'espace reflète les relations entre eux. Ces modèles de plongement hyperbolique permettent alors de mieux comprendre le lien entre la structure des réseaux complexes et leur fonctionnement. L'obtention des positions des noeuds dans l'espace hyperbolique offre une perspective géométrique et intuitive sur les interactions complexes du réseau à l'étude ; favorisant non seulement une analyse qualitative et quantitative, mais ouvrant également la voie à la prédiction et la simulation.

Modèle \mathcal{H}^2

Les plongements de graphes non orientés du modèle \mathcal{H}^2 [8, 11, 55] sont générés en distribuant N sommets sur la surface d'un disque de rayon \hat{R} dans le plan hyperbolique \mathbb{H}^2 et en connectant les paires de sommets i et j avec une probabilité $p_{ij} = f(d(i,j) - \hat{R})$, où $d(i,j)$ est la distance hyperbolique entre deux sommets et où $f(\cdot)$ est une fonction monotone décroissante.

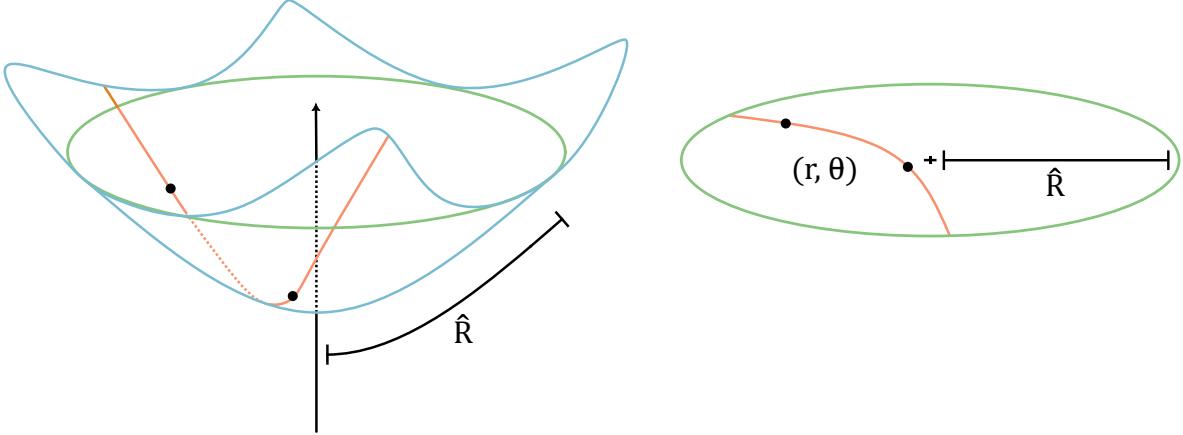


FIGURE 1.15 – Modèle de l’hyperboloïde tronqué à un rayon \hat{R} et représentation abstraite du disque correspondant dans le plan hyperbolique \mathbb{H}^2 paramétré par deux coordonnées (r, θ) .

Tel que mentionné à la section précédente, il existe de nombreuses représentations du plan hyperbolique \mathbb{H}^2 . Dans le cas du modèle \mathcal{H}^2 , on fait appel au modèle de l’hyperboloïde tronqué à un rayon \hat{R} , dans lequel les points du plan hyperbolique sont décrits par deux coordonnées (r, θ) et où la métrique du plan hyperbolique en coordonnées polaires est

$$ds^2 = dr^2 + \sinh^2 r d\theta^2. \quad (1.37)$$

La représentation de l’espace hyperbolique dans ce modèle est illustrée à la figure 1.15. Notez qu’il ne s’agit pas d’une projection verticale de l’hyperboloïde, mais plutôt d’un ouvert borné par $r < \hat{R}$ dans l’espace des coordonnées (r, θ) . Également, la métrique de l’équation (1.37) est la métrique induite par la métrique de Minkowski sur l’hyperboloïde, de là l’équivalence entre les deux représentations de la figure 1.15. En comparant la métrique de l’équation (1.37) à la métrique du plan euclidien en coordonnées polaires,

$$ds^2 = dr^2 + r d\theta^2, \quad (1.38)$$

on remarque que les périmètres des cercles hyperboliques et conséquemment les aires des disques de rayon r croissent beaucoup plus vite avec r que ceux des cercles euclidiens à cause du terme en sinus hyperbolique dans la métrique de l’équation (1.37) qui est absent de la métrique euclidienne de l’équation (1.38). En particulier, ils croissent de façon exponentielle lorsque $r \gg 1$ [55]. La distance hyperbolique entre deux points de coordonnées radiales r_i et r_j et de séparation angulaire

$$\Delta\theta_{ij} = \min(|\theta_i - \theta_j|, 2\pi - |\theta_i - \theta_j|) \quad (1.39)$$

est calculée à partir de la loi hyperbolique des cosinus

$$\cosh d(i, j) = \cosh r_i \cosh r_j - \sinh r_i \sinh r_j \cos \Delta\theta_{ij}. \quad (1.40)$$

Dans le modèle \mathcal{H}^2 , on obtient un ensemble de graphes maximalement entropique lorsque la probabilité de connexion prend la forme d'une distribution de Fermi-Dirac

$$p_{ij} = \frac{1}{1 + e^{\frac{\beta}{2}(d(i,j) - \hat{R})}}, \quad (1.41)$$

avec β contrôlant l'agrégation. Il est intéressant de noter que l'équation (1.41) suggère que nous pouvons interpréter cet ensemble de graphes, sous la loupe de la physique statistique, comme un ensemble grand canonique, où les arêtes sont considérées comme des fermions dont l'énergie dépend de la distance hyperbolique $d(i,j)$, où β est l'inverse de la température du système, et où \hat{R} est le potentiel chimique contrôlant le nombre attendu de fermions [8]. Le choix de la probabilité définit ainsi les propriétés du graphe [55].

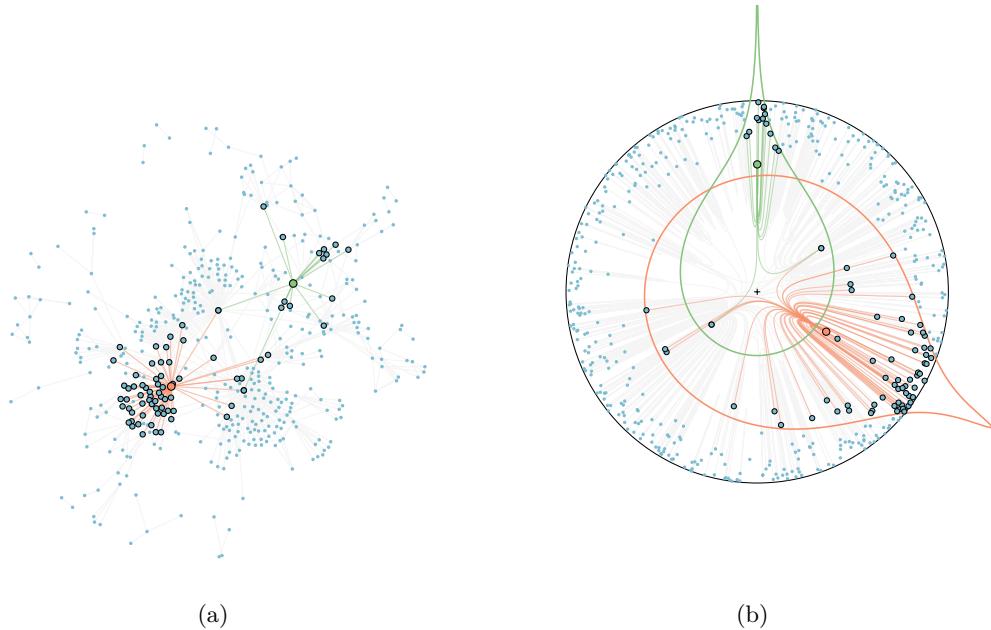


FIGURE 1.16 – Représentation du modèle \mathcal{H}^2 . (a) Exemple de graphe non orienté avec certains sommets (orange et vert) et leurs voisins respectifs en surbrillance. Il est à noter que cette représentation ne tient compte que du patron de connexion entre les sommets ; elle ne contient aucune information géométrique. (b) Modèle \mathcal{H}^2 dans lequel les sommets du graphe sont distribués sur un disque dans le plan hyperbolique et reliés entre eux selon la probabilité de l'équation (1.41) et suivant la géodésique du plus court chemin définie dans le modèle de l'hyperboïde comme étant une branche d'hyperbole. Chaque zone orange ou verte est délimitée par un cercle dont le centre est le sommet orange ou vert. Comme l'aire d'un cercle hyperbolique grandit exponentiellement avec son rayon, dans notre représentation sur le plan, les cercles sont déformés et ont plutôt l'apparence d'une goûte.

Le modèle \mathcal{H}^2 , qui est illustré à la figure 1.16, est donc un modèle purement géométrique dans lequel les propriétés de similarité (mesure de la proximité spatiale des sommets) et de popularité (l'importance de chaque sommet dans le graphe relativement à son degré) émergent

naturellement de l'interprétation géométrique et sont combinées en une valeur de distance dans le plan hyperbolique, de sorte que les sommets les plus proches, en termes de distance hyperbolique, soient plus susceptibles d'être connectés.

Modèle \mathcal{S}^1

Le prochain modèle géométrique proposé, le modèle \mathcal{S}^1 [8, 9, 55], est plus simple que le précédent et tout de même capable de générer des structures réalistes de réseaux non orientés. Il suppose comme espace métrique une sphère unidimensionnelle ou cercle \mathbb{S}^1 . Dans ce modèle, N sommets sont distribués sur le cercle, de sorte que chaque sommet i se voit attribuer une coordonnée angulaire θ_i comprise entre 0 et 2π . Le rayon du cercle R est fixé à $R = N/2\pi$, ce qui rend la densité de sommets unitaire sur le cercle. Chaque sommet i se voit également attribuer une autre variable, un degré caché κ_i . Étant donné que nous nous intéressons aux réseaux empiriques, nous considérons généralement des distributions en loi de puissance avec une borne inférieure κ_0 . Une fois que toutes les variables sont assignées, nous connectons chaque paire de sommets i et j selon la probabilité

$$p_{ij} = \frac{1}{1 + \chi_{ij}^\beta}, \quad \chi_{ij} = \frac{R\Delta\theta_{ij}}{\mu\kappa_i\kappa_j}, \quad (1.42)$$

avec

$$\mu = \frac{\beta \sin \frac{\pi}{\beta}}{2\pi \mathbb{E}[\kappa]}, \quad (1.43)$$

un paramètre contrôlant l'espérance des degrés cachés $\mathbb{E}[\kappa]$ dans le modèle et où $\Delta\theta_{ij}$ est la séparation angulaire minimale entre les deux sommets conformément à l'équation (1.39) et $\beta > 1$ est un paramètre d'agrégation [12, 56]. Comme dans le cas du modèle \mathcal{H}^2 , à la lumière de la physique statistique, cette probabilité de connexion permet d'obtenir un ensemble de graphes maximalement entropique [55].

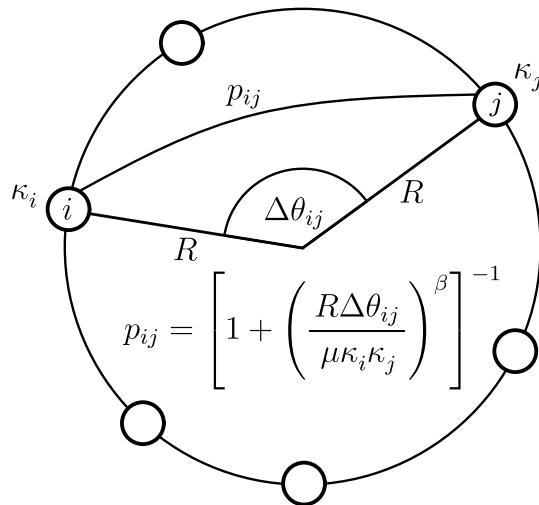


FIGURE 1.17 – Représentation du modèle \mathcal{S}^1 .

La figure 1.17 présente le modèle \mathcal{S}^1 dans lequel les sommets sont distribués sur le cercle et connectés entre eux selon la probabilité de l'équation (1.42). Contrairement au modèle \mathcal{H}^2 , où la similarité et la popularité étaient toutes deux encodées géométriquement, dans le modèle \mathcal{S}^1 , la géométrie représente explicitement la similarité entre les sommets, tandis que la popularité est encodée dans les degrés cachés, qui sont des attributs intrinsèques à chaque sommet, ne contenant aucune information géométrique. Cependant, la dissociation entre popularité et géométrie n'est qu'apparente puisque le modèle \mathcal{S}^1 est approximativement équivalent au modèle \mathcal{H}^2 [56], dans lequel la probabilité de connexion est uniquement fonction de la distance dans l'espace.

Équivalence approximative entre le modèle \mathcal{H}^2 et le modèle \mathcal{S}^1

En géométrie des réseaux, il est possible d'obtenir le modèle \mathcal{S}^1 à partir du modèle \mathcal{H}^2 et vice versa puisqu'il existe une quasi isométrie riemannienne entre ces deux modèles, c'est-à-dire une transformation indéfiniment différentiable et inversible qui préserve approximativement les distances [49, 56]. Cette transformation permet ainsi de passer des coordonnées $\mathbf{r} := (r_1, \dots, r_N)$ et $\boldsymbol{\theta} := (\theta_1, \dots, \theta_N)$ de l'espace \mathbb{H}^2 aux coordonnées $\boldsymbol{\theta} := (\theta_1, \dots, \theta_N)$ du cercle \mathbb{S}^1 et degrés cachés $\boldsymbol{\kappa} := (\kappa_1, \dots, \kappa_N)$ de l'espace $\mathbb{S}^1 \times \mathbb{R}_+$.

On obtient donc l'équivalence entre le modèle \mathcal{S}^1 et le modèle \mathcal{H}^2 en assignant comme valeur de degré caché dans le modèle \mathcal{S}^1

$$\kappa_i = \kappa_0 e^{\frac{1}{2}(\hat{R}-r_i)}, \quad (1.44)$$

où r_i est la coordonnée radiale du sommet i dans le modèle \mathcal{H}^2 ,

$$\hat{R} = 2 \ln \frac{N}{\pi \mu \kappa_0^2} \quad (1.45)$$

est le rayon du disque hyperbolique et κ_0 est la valeur du plus petit degré caché admis par la distribution des degrés cachés. On remplace μ , κ_i et κ_j dans le terme χ_{ij} de la probabilité de connexion de l'équation (1.42) par les équations (1.44) et (1.45) tel que

$$\chi_{ij} = \frac{N \Delta \theta_{ij}}{2\pi} \cdot \frac{\pi \kappa_0^2 e^{\frac{1}{2}\hat{R}}}{N} \cdot \frac{e^{\frac{1}{2}(r_i-\hat{R})}}{\kappa_0} \cdot \frac{e^{\frac{1}{2}(r_j-\hat{R})}}{\kappa_0} = \frac{\Delta \theta_{ij}}{2} e^{\frac{1}{2}(r_i+r_j-\hat{R})}. \quad (1.46)$$

À l'aide des propriétés de l'exponentielle, on regroupe tout dans l'exposant tel que

$$\chi_{ij} = e^{\ln \frac{\Delta \theta_{ij}}{2}} e^{\frac{1}{2}(r_i+r_j-\hat{R})} = e^{\frac{1}{2}(r_i+r_j-\hat{R}+2 \ln(\Delta \theta_{ij}/2))}. \quad (1.47)$$

En posant $\tilde{d}(i, j) = r_i + r_j + 2 \ln(\Delta \theta_{ij}/2)$, on obtient la forme

$$\chi_{ij} = e^{\frac{1}{2}(\tilde{d}(i, j)-\hat{R})}, \quad (1.48)$$

où $\tilde{d}(i, j)$ approxime la distance hyperbolique. Pour le démontrer, on dérive maintenant la distance hyperbolique $d(i, j)$ à partir de la loi hyperbolique des cosinus,

$$\cosh d(i, j) = \cosh r_i \cosh r_j - \sinh r_i \sinh r_j \cos \Delta \theta_{ij}, \quad (1.49)$$

en effectuant une simple mise en évidence sur les cosinus hyperboliques,

$$\cosh d(i, j) = \cosh r_i \cosh r_j \left(1 - \frac{\sinh r_i \sinh r_j}{\cosh r_i \cosh r_j} \cos \Delta\theta_{ij} \right), \quad (1.50)$$

puis en remplaçant le quotient par les tangentes hyperboliques,

$$\cosh d(i, j) = \cosh r_i \cosh r_j (1 - \tanh r_i \tanh r_j \cos \Delta\theta_{ij}). \quad (1.51)$$

On remarque que pour des rayons r_i et r_j suffisamment grands, les fonctions tangentes hyperboliques sont unitaires (i.e., $\lim_{x \rightarrow \infty} \tanh x = 1$). L'approximation est donc meilleure pour les grands graphes puisque la valeur des rayons r_i et r_j augmente avec la taille du graphe N . On exprime alors l'équation comme

$$\cosh d(i, j) \approx \cosh r_i \cosh r_j (1 - \cos \Delta\theta_{ij}). \quad (1.52)$$

En utilisant l'identité trigonométrique $1 - \cos(2x) = 2 \sin^2(x)$, on exprime l'équation comme

$$\cosh d(i, j) \approx \cosh r_i \cosh r_j (2 \sin^2(\Delta\theta_{ij}/2)). \quad (1.53)$$

On remplace dans l'équation par la forme exponentielle de la fonction cosinus hyperbolique,

$$\frac{e^{d(i, j)} + e^{-d(i, j)}}{2} \approx \frac{e^{r_i} + e^{-r_i}}{2} \frac{e^{r_j} + e^{-r_j}}{2} (2 \sin^2(\Delta\theta_{ij}/2)), \quad (1.54)$$

puis, tout en annulant les facteurs 2 de part et d'autre, on effectue une simple mise en évidence du facteur $e^{-d(i, j)}$ dans le membre de gauche tel que

$$\frac{e^{2d(i, j)} + 1}{e^{d(i, j)}} \approx (e^{r_i} + e^{-r_i})(e^{r_j} + e^{-r_j})(\sin^2(\Delta\theta_{ij}/2)). \quad (1.55)$$

On remarque que pour des rayons r_i et r_j suffisamment grands, les termes e^{-r_i} et e^{-r_j} sont nuls (i.e., $\lim_{x \rightarrow \infty} e^{-x} = 0$). On exprime alors l'équation comme

$$\frac{e^{2d(i, j)} + 1}{e^{d(i, j)}} \approx e^{r_i+r_j} \sin^2(\Delta\theta_{ij}/2). \quad (1.56)$$

On obtient une équation quadratique en $e^{d(i, j)}$,

$$e^{2d(i, j)} - e^{d(i, j)} e^{r_i+r_j} \sin^2(\Delta\theta_{ij}/2) + 1 = 0, \quad (1.57)$$

dont la solution est

$$e^{d(i, j)} = \frac{1}{2} \left(e^{r_i+r_j} \sin^2(\Delta\theta_{ij}/2) \pm \sqrt{e^{2(r_i+r_j)} \sin^4(\Delta\theta_{ij}/2) - 4} \right). \quad (1.58)$$

On met en évidence un terme $e^{r_i+r_j} \sin^2(\Delta\theta_{ij}/2)$ de manière à obtenir

$$e^{d(i, j)} = \frac{1}{2} e^{r_i+r_j} \sin^2(\Delta\theta_{ij}/2) \left(1 \pm \sqrt{1 - 4e^{-2(r_i+r_j)} \sin^{-4}(\Delta\theta_{ij}/2)} \right). \quad (1.59)$$

On remarque que pour $4e^{-2(r_i+r_j)} \ll \sin^4(\Delta\theta_{ij}/2)$, le deuxième terme de la racine carrée est nul. Dans ce cas, $\Delta\theta_{ij} > 2 \arcsin\left(\sqrt[4]{4e^{-2(r_i+r_j)}}\right)$. On exprime alors l'équation comme

$$e^{d(i,j)} \approx \frac{1}{2} e^{r_i+r_j} \sin^2(\Delta\theta_{ij}/2) \left(1 \pm \sqrt{1}\right). \quad (1.60)$$

En ne conservant que la branche positive, on obtient

$$e^{d(i,j)} \approx e^{r_i+r_j} \sin^2(\Delta\theta_{ij}/2). \quad (1.61)$$

On isole la distance hyperbolique telle que

$$d(i,j) \approx \ln(e^{r_i+r_j} \sin^2(\Delta\theta_{ij}/2)). \quad (1.62)$$

Par les propriétés du logarithme, on obtient

$$d(i,j) \approx r_i + r_j + 2 \ln \sin(\Delta\theta_{ij}/2). \quad (1.63)$$

On remarque que dans la limite des petits angles, $\sin(x) \approx x$. On peut donc exprimer la distance hyperbolique telle que

$$d(i,j) \approx r_i + r_j + 2 \ln \frac{\Delta\theta_{ij}}{2}. \quad (1.64)$$

On obtient ainsi une approximation adéquate pour la distance hyperbolique $d(i,j)$,

$$\tilde{d}(i,j) = r_i + r_j + 2 \ln \frac{\Delta\theta_{ij}}{2}. \quad (1.65)$$

Q.E.D.

On obtient l'équation (1.42) de la probabilité de connexion du modèle \mathcal{S}^1 en remplaçant par les équations (1.44), (1.45) et (1.64) dans l'équation (1.41) de la probabilité de connexion du modèle \mathcal{H}^2 .

1.3.3 Le cas des graphes orientés

Maintenant que avons vu les modèles de plongement dans l'espace hyperbolique pour les graphes non orientés, nous pouvons nous intéresser au cas des graphes orientés qui pose un défi supplémentaire.

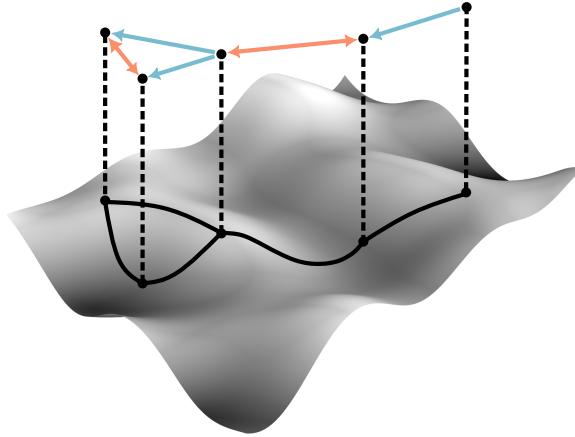


FIGURE 1.18 – Plongement d'un graphe orienté dans un espace métrique.

Contrairement au graphe de la figure 1.14, celui de la figure 1.18 est orienté. Conformément au paradigme de la géométrie des graphes, deux sommets connectés doivent forcément se trouver à une faible distance dans l'espace métrique sous-jacent et vice versa. Or, dans un modèle de graphe orienté, on veut pouvoir favoriser la connexion dans un sens ou dans l'autre, ce qui implique une probabilité de connexion asymétrique. Dans le cas où la connexion entre deux sommets i et j est unidirectionnelle et orientée de i vers j , la probabilité p_{ij} est donc plus grande que la probabilité p_{ji} . Une grande probabilité p_{ij} implique une petite distance entre i et j alors qu'une petite probabilité p_{ji} implique une grande distance entre i et j . La distance ne peut toutefois pas être à la fois petite et grande, car cela ne respecte pas la propriété de symétrie de la distance énoncée à la section 1.3.1. Afin de résoudre ce problème, on utilise un modèle de plongement quasi géométrique, à savoir le modèle \mathcal{S}^1 orienté.

Modèle \mathcal{S}^1 orienté et modèle \mathcal{S}^1 orienté-réiproque

Le modèle \mathcal{S}^1 orienté est adapté du modèle \mathcal{S}^1 aux graphes orientés [16]. Sa principale différence avec son homonyme non orienté est que cette version orientée du modèle admet des degrés cachés entrants et sortants (i.e., κ^- et κ^+) et non seulement des degrés cachés (i.e., κ). Le plongement s'effectue alors dans l'espace $\mathbb{S}^1 \times \mathbb{R}_+ \times \mathbb{R}_+$. Dans ce modèle, la probabilité qu'il y ait un arc du sommet i au sommet j est

$$p_{ij} = \frac{1}{1 + \chi_{ij}^\beta}, \quad \chi_{ij} = \frac{R\Delta\theta_{ij}}{\mu\kappa_i^+\kappa_j^-}, \quad (1.66)$$

où κ_i^+ est le degré caché sortant du sommet i , κ_j^- est le degré caché entrant du sommet j , $\Delta\theta_{ij}$ est la séparation angulaire minimale entre les sommets conformément à l'équation (1.39), $R = N/2\pi$, $\beta > 1$ est un paramètre d'agrégation et μ est un paramètre contrôlant l'espérance

des degrés cachés conformément à l'équation (1.43) où $\mathbb{E}[\kappa] := \mathbb{E}[\kappa^+] = \mathbb{E}[\kappa^-]$ est l'espérance des degrés cachés où l'on fixe le premier moment de la distribution conjointe $\rho(\kappa^+, \kappa^-)$.

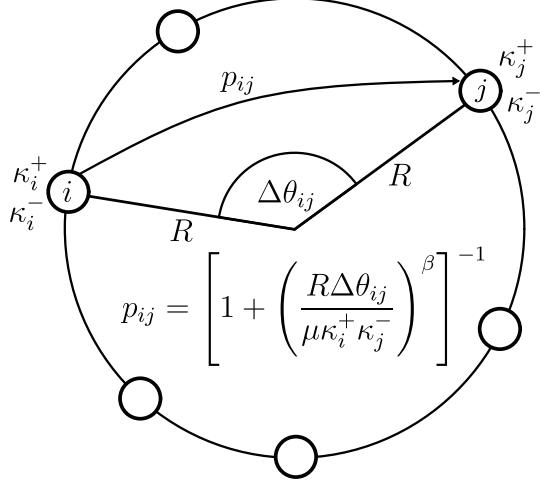


FIGURE 1.19 – Représentation du modèle \mathcal{S}^1 orienté.

La figure 1.19 présente le modèle \mathcal{S}^1 orienté dans lequel les sommets sont distribués sur le cercle et connectés entre eux selon la probabilité de l'équation (1.66).

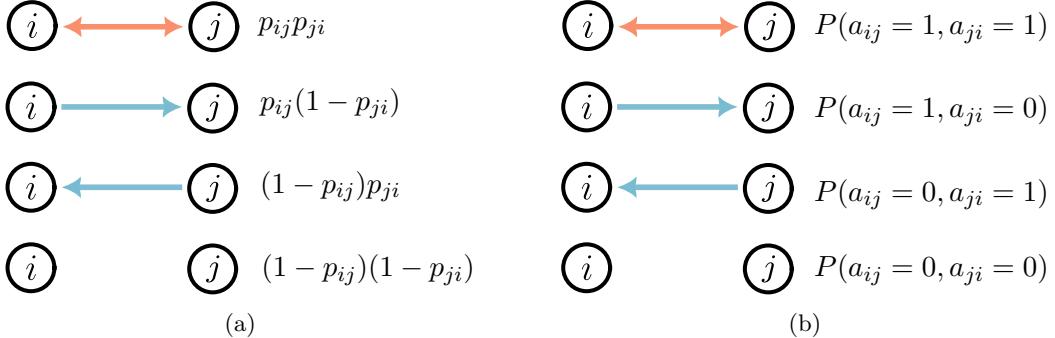


FIGURE 1.20 – (a) Réciprocité fortuite dans le modèle \mathcal{S}^1 orienté. (b) Réciprocité délibérée dans le modèle \mathcal{S}^1 orienté-réiproque.

Dans cette version préliminaire du modèle, la réciprocité caractérisant la nature unidirectionnelle ou bidirectionnelle d'une interaction est fortuite, c'est-à-dire qu'il est du domaine de la probabilité de l'équation (1.66) de connecter de manière unidirectionnelle ou bidirectionnelle deux sommets donnés. Le panneau (a) de la figure 1.20 montre la probabilité d'obtenir chaque type de connexion dans le modèle \mathcal{S}^1 orienté tel qu'il a été défini jusqu'à présent. De manière à exercer un plus grand contrôle sur la réciprocité dans le modèle et de pouvoir ainsi se concentrer indépendamment sur chaque type de connexion existant entre deux sommets tel qu'il est montré dans le panneau (b) de la figure 1.20, il est possible de définir quatre probabilités de

connexion conjointes qui sont conditionnelles à leurs dépendances sur les degrés cachés et la distance angulaire (i.e., $\kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-, \Delta\theta_{ij}$) et qui ont les relations suivantes :

$$\begin{aligned} P(a_{ij} = 1, a_{ji} = 1) + P(a_{ij} = 1, a_{ji} = 0) \\ + P(a_{ij} = 0, a_{ji} = 1) + P(a_{ij} = 0, a_{ji} = 0) = 1, \end{aligned} \quad (1.67a)$$

$$P(a_{ij} = 1, a_{ji} = 1) + P(a_{ij} = 1, a_{ji} = 0) = p_{ij}, \quad (1.67b)$$

$$P(a_{ij} = 1, a_{ji} = 1) + P(a_{ij} = 0, a_{ji} = 1) = p_{ji}. \quad (1.67c)$$

Notez que a_{ij} est l'élément de la matrice d'adjacence du graphe orienté ($a_{ij} = 1$ s'il y a un arc du sommet i au sommet j et $a_{ij} = 0$ autrement). Également, par souci d'alléger la notation, la dépendance sur les degrés cachés et la distance angulaire dans les probabilités des équations (1.67) sera omise sauf lorsque nécessaire. Les équations (1.67) laissent l'une des quatre probabilités $P(a_{ij}, a_{ji})$ non définie, offrant ainsi au modèle un degré de liberté supplémentaire pour ajuster la réciprocité du graphe. Cela peut être fait en considérant le coefficient de corrélation

$$\rho_{ij} = \frac{\mathbb{E}[a_{ij}a_{ji}] - \mathbb{E}[a_{ij}]\mathbb{E}[a_{ji}]}{\sqrt{(\mathbb{E}[a_{ij}^2] - \mathbb{E}[a_{ij}]^2)(\mathbb{E}[a_{ji}^2] - \mathbb{E}[a_{ji}]^2)}} = \frac{P_{ij}(1,1) - p_{ij}p_{ji}}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}}. \quad (1.68)$$

où $\mathbb{E}[\cdot]$ désigne l'espérance sur l'ensemble de graphes défini par les probabilités conjointes [16]. Remarquons que, puisque $P(a_{ij} = 1, a_{ji} = 1) \in [0, 1]$, l'équation (1.68) n'est pas nécessairement bornée entre -1 et 1 . Imposer ces bornes conduit à une expression de $P(a_{ij} = 1, a_{ji} = 1)$ en fonction de p_{ij} , p_{ji} et d'un paramètre $\nu \in [-1, 1]$ contrôlant le niveau de réciprocité entre les sommets i et j ,

$$P(a_{ij} = 1, a_{ji} = 1) = \begin{cases} (1 + \nu)p_{ij}p_{ji} - \nu(p_{ij} + p_{ji} - 1)H(p_{ij} + p_{ji} - 1) & -1 \leq \nu \leq 0 \\ (1 - \nu)p_{ij}p_{ji} + \nu \min\{p_{ij}, p_{ji}\} & 0 \leq \nu \leq 1 \end{cases} \quad (1.69)$$

dans laquelle $H(\cdot)$ est la fonction de Heaviside. La réciprocité est alors délibérée, conformément au panneau (b) de la figure 1.20 qui montre la probabilité d'obtenir chaque type de connexion dans cette version du modèle \mathcal{S}^1 orienté que l'on nomme le modèle \mathcal{S}^1 orienté-réiproque.

Espérance de la distance angulaire dans le modèle \mathcal{S}^1 orienté-réiproque

Le caractère géométrique du modèle \mathcal{S}^1 orienté-réiproque s'inscrit dans la position angulaire de chaque sommet sur le cercle. Il est donc pertinent de s'intéresser à l'espérance de la distance angulaire entre deux sommets connectés selon le type de connexion existant entre ces deux sommets puisque le résultat de ce calcul sera le point de départ pour l'inférence des positions angulaires avec l'algorithme présenté au chapitre 2. La démarche qui suit correspond donc au calcul analytique de l'espérance de la distance angulaire dans le modèle \mathcal{S}^1 orienté-réiproque.

Nous pouvons d'abord exprimer le quotient des produits des degrés cachés tel que

$$\xi_{ij} = \frac{\kappa_i^+ \kappa_j^-}{\kappa_j^+ \kappa_i^-}, \quad (1.70)$$

ce qui facilitera le passage de p_{ij} à p_{ji} tel que

$$\chi_{ji} = \xi_{ij}\chi_{ij}. \quad (1.71)$$

La distribution de la distance angulaire $\Delta\theta_{ij}$ entre deux sommets reliés par une paire d'arcs de sens opposés est obtenue du théorème de Bayes

$$\rho(\Delta\theta_{ij}|a_{ij} = 1, a_{ji} = 1, \kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-) = \frac{P(a_{ij} = 1, a_{ji} = 1|\kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-) \rho(\Delta\theta_{ij})}{P(a_{ij} = 1, a_{ji} = 1|\kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-)}, \quad (1.72)$$

où $P(a_{ij} = 1, a_{ji} = 1|\kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-, \Delta\theta_{ij})$ est obtenue de l'équation (1.69) et où la densité angulaire sur le cercle est $\rho(\Delta\theta_{ij}) = 1/\pi$. On obtient le facteur de normalisation en marginalisant sur $\Delta\theta_{ij}$ tel que

$$P(a_{ij} = 1, a_{ji} = 1|\kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-) = \int_0^\pi P(a_{ij} = 1, a_{ji} = 1|\kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-, \Delta\theta_{ij}) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}. \quad (1.73)$$

L'espérance de la distance angulaire entre les sommets i et j dans le modèle \mathcal{S}^1 orienté-réiproque, sachant que ces sommets sont réciprocement connectés, est ainsi donnée par

$$\mathbb{E} [\Delta\theta_{ij}|a_{ij} = 1, a_{ji} = 1, \kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-] = \int_0^\pi \Delta\theta_{ij} \rho(\Delta\theta_{ij}|a_{ij} = 1, a_{ji} = 1, \kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-) d\Delta\theta_{ij}. \quad (1.74)$$

La démarche suivante explicite les calculs de l'espérance de la distance angulaire pour la probabilité de connexion réciproque $P(a_{ij} = 1, a_{ji} = 1)$ de l'équation (1.69).

Dans le cas où $\nu = 1$, nous avons la probabilité

$$P(a_{ij} = 1, a_{ji} = 1|\nu = 1) = \min\{p_{ij}, p_{ji}\}, \quad (1.75)$$

et l'espérance de la distance angulaire

$$\mathbb{E}[\Delta\theta_{ij}|P(a_{ij} = 1, a_{ji} = 1|\nu = 1)] = \frac{\int_0^\pi \Delta\theta_{ij} P(a_{ij} = 1, a_{ji} = 1|\nu = 1) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}}{\int_0^\pi P(a_{ij} = 1, a_{ji} = 1|\nu = 1) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}}. \quad (1.76)$$

De manière à alléger l'écriture du développement mathématique des différentes intégrales, nous allons utiliser les symboles \mathcal{I}_n pour représenter les intégrales. Nous avons d'abord \mathcal{I}_1 , l'intégrale du dénominateur de l'équation (1.76),

$$\begin{aligned} \mathcal{I}_1 &= \int_0^\pi P(a_{ij} = 1, a_{ji} = 1|\nu = 1) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}, \\ \mathcal{I}_1 &= \frac{1}{\pi} \int_0^\pi \min\{p_{ij}, p_{ji}\} d\Delta\theta_{ij}, \\ \mathcal{I}_1 &= \frac{1}{\pi} \int_0^\pi \min\left\{\frac{1}{1 + \chi_{ij}^\beta}, \frac{1}{1 + \chi_{ji}^\beta}\right\} d\Delta\theta_{ij}. \end{aligned}$$

Nous pouvons développer cette intégrale à l'aide de la fonction de Heaviside tel que

$$\begin{aligned}\mathcal{I}_1 &= H(1 - \xi_{ij}) \frac{1}{\pi} \int_0^\pi \frac{1}{1 + \chi_{ij}^\beta} d\Delta\theta_{ij} \\ &\quad + H(\xi_{ij} - 1) \frac{1}{\pi} \int_0^\pi \frac{1}{1 + \chi_{ji}^\beta} d\Delta\theta_{ij}.\end{aligned}$$

Un changement de variable dans l'intégrande nous permet d'exprimer l'intégrale comme

$$\begin{aligned}\mathcal{I}_1 &= H(1 - \xi_{ij}) \frac{2\mu\kappa_i^+ \kappa_j^-}{N} \int_0^{\frac{N}{2\mu\kappa_i^+ \kappa_j^-}} \frac{1}{1 + \chi_{ij}^\beta} d\chi_{ij} \\ &\quad + H(\xi_{ij} - 1) \frac{2\mu\kappa_j^+ \kappa_i^-}{N} \int_0^{\frac{N}{2\mu\kappa_j^+ \kappa_i^-}} \frac{1}{1 + \chi_{ji}^\beta} d\chi_{ij}.\end{aligned}$$

Ces intégrales sont calculées conformément aux identités relatives à la fonction hypergéométrique ${}_2F_1(a, b, c, z)$ énumérées à l'annexe A.1. Nous obtenons alors

$$\begin{aligned}\mathcal{I}_1 &= H(1 - \xi_{ij}) {}_2F_1\left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+ \kappa_j^-}\right)^\beta\right) \\ &\quad + H(\xi_{ij} - 1) {}_2F_1\left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+ \kappa_i^-}\right)^\beta\right).\end{aligned}\tag{1.77}$$

Nous avons \mathcal{I}_2 , l'intégrale du numérateur de l'équation (1.76),

$$\begin{aligned}\mathcal{I}_2 &= \int_0^\pi \Delta\theta_{ij} P(a_{ij} = 1, a_{ji} = 1 | \nu = 1) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}, \\ \mathcal{I}_2 &= \frac{1}{\pi} \int_0^\pi \Delta\theta_{ij} \min\{p_{ij}, p_{ji}\} d\Delta\theta_{ij}, \\ \mathcal{I}_2 &= \frac{1}{\pi} \int_0^\pi \min\left\{\frac{\Delta\theta_{ij}}{1 + \chi_{ij}^\beta}, \frac{\Delta\theta_{ij}}{1 + \chi_{ji}^\beta}\right\} d\Delta\theta_{ij}.\end{aligned}$$

Nous pouvons développer cette intégrale à l'aide de la fonction de Heaviside tel que

$$\mathcal{I}_2 = \frac{1}{\pi} H(1 - \xi_{ij}) \int_0^\pi \frac{\Delta\theta_{ij}}{1 + \chi_{ij}^\beta} d\Delta\theta_{ij} + \frac{1}{\pi} H(\xi_{ij} - 1) \int_0^\pi \frac{\Delta\theta_{ij}}{1 + \chi_{ji}^\beta} d\Delta\theta_{ij}.$$

Un changement de variable dans l'intégrande nous permet d'exprimer l'intégrale comme

$$\begin{aligned}\mathcal{I}_2 &= \frac{1}{\pi} H(1 - \xi_{ij}) \left(\frac{2\pi\mu\kappa_i^+ \kappa_j^-}{N}\right)^2 \int_0^{\frac{N}{2\mu\kappa_i^+ \kappa_j^-}} \frac{\chi_{ij}}{1 + \chi_{ij}^\beta} d\chi_{ij} \\ &\quad + \frac{1}{\pi} H(\xi_{ij} - 1) \left(\frac{2\pi\mu\kappa_i^+ \kappa_j^-}{N}\right)^2 \int_0^{\frac{N}{2\mu\kappa_j^+ \kappa_i^-}} \frac{\chi_{ij}}{1 + (\xi_{ij}\chi_{ij})^\beta} d\chi_{ij}.\end{aligned}$$

Ces intégrales sont calculées conformément aux identités relatives à la fonction hypergéométrique ${}_2F_1(a, b, c, z)$ énumérées à l'annexe A.1. Nous obtenons alors

$$\begin{aligned}\mathcal{I}_2 &= \frac{\pi}{2} H(1 - \xi_{ij}) {}_2F_1\left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, - \left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) \\ &\quad + \frac{\pi}{2} H(\xi_{ij} - 1) {}_2F_1\left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, - \left(\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right)^\beta\right).\end{aligned}\quad (1.78)$$

Une fois les intégrales de l'équation (1.76) calculées, nous obtenons l'espérance de la distance angulaire,

$$\begin{aligned}\mathbb{E}[\Delta\theta_{ij}|P(a_{ij} = 1, a_{ji} = 1|\nu = 1)] \\ = \frac{\pi}{2} \frac{\left[H(1 - \xi_{ij}) {}_2F_1\left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, - \left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) + H(\xi_{ij} - 1) {}_2F_1\left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, - \left(\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right)^\beta\right)\right]}{\left[H(1 - \xi_{ij}) {}_2F_1\left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, - \left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) + H(\xi_{ij} - 1) {}_2F_1\left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, - \left(\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right)^\beta\right)\right]}\end{aligned}\quad (1.79)$$

Dans le cas où $\nu = 0$, nous avons la probabilité

$$P(a_{ij} = 1, a_{ji} = 1|\nu = 0) = p_{ij}p_{ji}, \quad (1.80)$$

et l'espérance de la distance angulaire

$$\mathbb{E}[\Delta\theta_{ij}|P(a_{ij} = 1, a_{ji} = 1|\nu = 0)] = \frac{\int_0^\pi \Delta\theta_{ij} P(a_{ij} = 1, a_{ji} = 1|\nu = 0) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}}{\int_0^\pi P(a_{ij} = 1, a_{ji} = 1|\nu = 0) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}}. \quad (1.81)$$

Nous avons \mathcal{I}_3 , l'intégrale du dénominateur de l'équation (1.81) lorsque $\xi_{ij} = 1$,

$$\begin{aligned}\mathcal{I}_3 &= \int_0^\pi P(a_{ij} = 1, a_{ji} = 1|\nu = 0, \xi_{ij} = 1) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}, \\ \mathcal{I}_3 &= \frac{1}{\pi} \int_0^\pi p_{ij}p_{ji} d\Delta\theta_{ij}, \\ \mathcal{I}_3 &= \frac{1}{\pi} \int_0^\pi \frac{1}{1 + \chi_{ij}^\beta} \frac{1}{1 + \chi_{ji}^\beta} d\Delta\theta_{ij}.\end{aligned}$$

Nous utilisons la relation de l'équation (1.71) afin d'obtenir

$$\mathcal{I}_3 = \frac{1}{\pi} \int_0^\pi \frac{1}{1 + \chi_{ij}^\beta} \frac{1}{1 + \xi_{ij}^\beta \chi_{ij}^\beta} d\Delta\theta_{ij}.$$

Comme $\xi_{ij} = 1$, nous obtenons

$$\mathcal{I}_3 = \frac{1}{\pi} \int_0^\pi \frac{1}{\left(1 + \chi_{ij}^\beta\right)^2} d\Delta\theta_{ij}.$$

Un changement de variable dans l'intégrande nous permet d'exprimer l'intégrale comme

$$\mathcal{I}_3 = \frac{2\mu\kappa_i^+\kappa_j^-}{N} \int_0^{\frac{N}{2\mu\kappa_i^+\kappa_j^-}} \frac{1}{(1 + \chi_{ij}^\beta)^2} d\chi_{ij}.$$

Cette intégrale est calculée conformément aux identités relatives à la fonction hypergéométrique ${}_2F_1(a, b, c, z)$ énumérées à l'annexe A.1. Nous obtenons alors

$$\mathcal{I}_3 = {}_2F_1\left(2, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right). \quad (1.82)$$

Nous avons \mathcal{I}_4 , l'intégrale du numérateur de l'équation (1.81) lorsque $\xi_{ij} = 1$,

$$\begin{aligned} \mathcal{I}_4 &= \int_0^\pi \Delta\theta_{ij} P(a_{ij} = 1, a_{ji} = 1 | \nu = 0, \xi_{ij} = 1) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}, \\ \mathcal{I}_4 &= \frac{1}{\pi} \int_0^\pi \Delta\theta_{ij} p_{ij} p_{ji} d\Delta\theta_{ij}, \\ \mathcal{I}_4 &= \frac{1}{\pi} \int_0^\pi \Delta\theta_{ij} \frac{1}{1 + \chi_{ij}^\beta} \frac{1}{1 + \chi_{ji}^\beta} d\Delta\theta_{ij}. \end{aligned}$$

Nous utilisons la relation de l'équation (1.71) afin d'obtenir

$$\mathcal{I}_4 = \frac{1}{\pi} \int_0^\pi \Delta\theta_{ij} \frac{1}{1 + \chi_{ij}^\beta} \frac{1}{1 + \xi_{ij}^\beta \chi_{ij}^\beta} d\Delta\theta_{ij}.$$

Comme $\xi_{ij} = 1$, nous obtenons

$$\mathcal{I}_4 = \frac{1}{\pi} \int_0^\pi \frac{\Delta\theta_{ij}}{(1 + \chi_{ij}^\beta)^2} d\Delta\theta_{ij}.$$

Un changement de variable dans l'intégrande nous permet d'exprimer l'intégrale comme

$$\mathcal{I}_4 = \frac{1}{\pi} \left(\frac{2\pi\mu\kappa_i^+\kappa_j^-}{N} \right)^2 \int_0^{\frac{N}{2\mu\kappa_i^+\kappa_j^-}} \frac{\chi_{ij}}{(1 + \chi_{ij}^\beta)^2} d\chi_{ij}.$$

Cette intégrale est calculée conformément aux identités relatives à la fonction hypergéométrique ${}_2F_1(a, b, c, z)$ énumérées à l'annexe A.1. Nous obtenons alors

$$\mathcal{I}_4 = \frac{\pi}{2} {}_2F_1\left(2, \frac{2}{\beta}, 1 + \frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right). \quad (1.83)$$

Nous avons finalement l'espérance de la distance angulaire lorsque $\xi_{ij} = 1$,

$$\mathbb{E}[\Delta\theta_{ij} | P(a_{ij} = 1, a_{ji} = 1 | \nu = 0, \xi_{ij} = 1)] = \frac{\pi}{2} \frac{{}_2F_1\left(2, \frac{2}{\beta}, 1 + \frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right)}{{}_2F_1\left(2, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right)}. \quad (1.84)$$

On remarque que dans le cas où $\xi_{ij} = 1$, les produits de degrés cachés $\kappa_i^+ \kappa_j^- = \kappa_j^+ \kappa_i^-$ et donc qu'il est possible d'exprimer l'espérance de la distance angulaire en fonction du produit $\kappa_i^+ \kappa_j^-$ ou du produit $\kappa_j^+ \kappa_i^-$ tout en restant cohérent avec le choix des degrés cachés.

Nous avons également \mathcal{I}_5 , l'intégrale du dénominateur de l'équation (1.81) lorsque $\xi_{ij} \neq 1$,

$$\begin{aligned}\mathcal{I}_5 &= \int_0^\pi P(a_{ij} = 1, a_{ji} = 1 | \nu = 0, \xi_{ij} \neq 1) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}, \\ \mathcal{I}_5 &= \frac{1}{\pi} \int_0^\pi p_{ij} p_{ji} d\Delta\theta_{ij}, \\ \mathcal{I}_5 &= \frac{1}{\pi} \int_0^\pi \frac{1}{1 + \chi_{ij}^\beta} \frac{1}{1 + \chi_{ji}^\beta} d\Delta\theta_{ij}.\end{aligned}$$

Nous utilisons la relation de l'équation (1.71) afin d'obtenir

$$\mathcal{I}_5 = \frac{1}{\pi} \int_0^\pi \frac{1}{1 + \chi_{ij}^\beta} \frac{1}{1 + \xi_{ij}^\beta \chi_{ij}^\beta} d\Delta\theta_{ij}.$$

Un changement de variable dans l'intégrande nous permet d'exprimer l'intégrale comme

$$\mathcal{I}_5 = \frac{2\mu\kappa_i^+ \kappa_j^-}{N} \int_0^{\frac{N}{2\mu\kappa_i^+ \kappa_j^-}} \frac{1}{1 + \chi_{ij}^\beta} \frac{1}{1 + \xi_{ij}^\beta \chi_{ij}^\beta} d\chi_{ij}.$$

Cette intégrale est calculée conformément aux identités relatives à la fonction hypergéométrique ${}_2F_1(a, b, c, z)$ énumérées à l'annexe A.1 et à l'aide de la décomposition en fractions partielles de l'annexe A.2. Nous obtenons alors

$$\begin{aligned}\mathcal{I}_5 &= \frac{1}{1 - \xi_{ij}^\beta} {}_2F_1\left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+ \kappa_j^-}\right)^\beta\right) \\ &\quad - \frac{\xi_{ij}^\beta}{1 - \xi_{ij}^\beta} {}_2F_1\left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+ \kappa_i^-}\right)^\beta\right).\end{aligned}\tag{1.85}$$

Nous avons \mathcal{I}_6 , l'intégrale du numérateur de l'équation (1.81) lorsque $\xi_{ij} \neq 1$,

$$\begin{aligned}\mathcal{I}_6 &= \int_0^\pi \Delta\theta_{ij} P(a_{ij} = 1, a_{ji} = 1 | \nu = 0, \xi_{ij} \neq 1) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}, \\ \mathcal{I}_6 &= \frac{1}{\pi} \int_0^\pi \Delta\theta_{ij} p_{ij} p_{ji} d\Delta\theta_{ij}, \\ \mathcal{I}_6 &= \frac{1}{\pi} \int_0^\pi \Delta\theta_{ij} \frac{1}{1 + \chi_{ij}^\beta} \frac{1}{1 + \chi_{ji}^\beta} d\Delta\theta_{ij}.\end{aligned}$$

Nous utilisons la relation de l'équation (1.71) afin d'obtenir

$$\mathcal{I}_6 = \frac{1}{\pi} \int_0^\pi \Delta\theta_{ij} \frac{1}{1 + \chi_{ij}^\beta} \frac{1}{1 + \xi_{ij}^\beta \chi_{ij}^\beta} d\Delta\theta_{ij}.$$

Un changement de variable dans l'intégrande nous permet d'exprimer l'intégrale comme

$$\mathcal{I}_6 = \frac{1}{\pi} \left(\frac{2\pi\mu\kappa_i^+\kappa_j^-}{N} \right)^2 \int_0^{\frac{N}{2\mu\kappa_i^+\kappa_j^-}} \frac{\chi_{ij}}{(1 + \chi_{ij}^\beta)(1 + \xi_{ij}^\beta\chi_{ij}^\beta)} d\chi_{ij}.$$

Cette intégrale est calculée conformément aux identités relatives à la fonction hypergéométrique ${}_2F_1(a, b, c, z)$ énumérées à l'annexe A.1 et à l'aide de la décomposition en fractions partielles de l'annexe A.2. Nous obtenons alors

$$\begin{aligned} \mathcal{I}_6 &= \frac{\pi}{2} \left[\frac{1}{1 - \xi_{ij}^\beta} {}_2F_1 \left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, - \left(\frac{N}{2\mu\kappa_i^+\kappa_j^-} \right)^\beta \right) \right. \\ &\quad \left. - \frac{\xi_{ij}^\beta}{1 - \xi_{ij}^\beta} {}_2F_1 \left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, - \left(\frac{N}{2\mu\kappa_j^+\kappa_i^-} \right)^\beta \right) \right]. \end{aligned} \quad (1.86)$$

Nous avons finalement l'espérance de la distance angulaire lorsque $\xi_{ij} \neq 1$,

$$\begin{aligned} &\mathbb{E}[\Delta\theta_{ij}|P(a_{ij} = 1, a_{ji} = 1|\nu = 0, \xi_{ij} \neq 1)] \\ &= \frac{\pi}{2} \left[\frac{\frac{1}{1 - \xi_{ij}^\beta} {}_2F_1 \left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, - \left(\frac{N}{2\mu\kappa_i^+\kappa_j^-} \right)^\beta \right) - \frac{\xi_{ij}^\beta}{1 - \xi_{ij}^\beta} {}_2F_1 \left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, - \left(\frac{N}{2\mu\kappa_j^+\kappa_i^-} \right)^\beta \right)}{\frac{1}{1 - \xi_{ij}^\beta} {}_2F_1 \left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, - \left(\frac{N}{2\mu\kappa_i^+\kappa_j^-} \right)^\beta \right) - \frac{\xi_{ij}^\beta}{1 - \xi_{ij}^\beta} {}_2F_1 \left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, - \left(\frac{N}{2\mu\kappa_j^+\kappa_i^-} \right)^\beta \right)} \right]. \end{aligned}$$

Nous pouvons annuler les facteurs $1/(1 - \xi_{ij}^\beta)$ de manière à obtenir

$$\begin{aligned} &\mathbb{E}[\Delta\theta_{ij}|P(a_{ij} = 1, a_{ji} = 1|\nu = 0, \xi_{ij} \neq 1)] \\ &= \frac{\pi}{2} \left[\frac{{}_2F_1 \left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, - \left(\frac{N}{2\mu\kappa_i^+\kappa_j^-} \right)^\beta \right) - \xi_{ij}^\beta {}_2F_1 \left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, - \left(\frac{N}{2\mu\kappa_j^+\kappa_i^-} \right)^\beta \right)}{{}_2F_1 \left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, - \left(\frac{N}{2\mu\kappa_i^+\kappa_j^-} \right)^\beta \right) - \xi_{ij}^\beta {}_2F_1 \left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, - \left(\frac{N}{2\mu\kappa_j^+\kappa_i^-} \right)^\beta \right)} \right]. \end{aligned} \quad (1.87)$$

Dans le cas où $\nu = -1$, nous avons la probabilité

$$P(a_{ij} = 1, a_{ji} = 1|\nu = -1) = (p_{ij} + p_{ji} - 1)H(p_{ij} + p_{ji} - 1), \quad (1.88)$$

et l'espérance de la distance angulaire

$$\mathbb{E}[\Delta\theta_{ij}|P(a_{ij} = 1, a_{ji} = 1|\nu = -1)] = \frac{\int_0^\pi \Delta\theta_{ij} P(a_{ij} = 1, a_{ji} = 1|\nu = -1) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}}{\int_0^\pi P(a_{ij} = 1, a_{ji} = 1|\nu = -1) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}}. \quad (1.89)$$

Nous avons \mathcal{I}_7 , l'intégrale du dénominateur de l'équation (1.89),

$$\begin{aligned} \mathcal{I}_7 &= \int_0^\pi P(a_{ij} = 1, a_{ji} = 1|\nu = -1) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}, \\ \mathcal{I}_7 &= \frac{1}{\pi} \int_0^\pi (p_{ij} + p_{ji} - 1)H(p_{ij} + p_{ji} - 1) d\Delta\theta_{ij}. \end{aligned}$$

Par la définition de la fonction Heaviside,

$$\mathcal{I}_7 = \frac{1}{\pi} \int_0^{\Delta\theta_{ij}^c} (p_{ij} + p_{ji} - 1) d\Delta\theta_{ij}.$$

Nous séparons l'intégrale en trois tel que

$$\mathcal{I}_7 = \frac{1}{\pi} \int_0^{\Delta\theta_{ij}^c} \frac{d\Delta\theta_{ij}}{1 + \chi_{ij}^\beta} + \frac{1}{\pi} \int_0^{\Delta\theta_{ij}^c} \frac{d\Delta\theta_{ij}}{1 + \chi_{ji}^\beta} - \frac{1}{\pi} \int_0^{\Delta\theta_{ij}^c} d\Delta\theta_{ij}.$$

Un changement de variable dans les intégrandes nous permet d'exprimer les intégrales comme

$$\mathcal{I}_7 = \frac{2\mu\kappa_i^+ \kappa_j^-}{N} \int_0^{\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_i^+ \kappa_j^-}} \frac{d\chi_{ij}}{1 + \chi_{ij}^\beta} + \frac{2\mu\kappa_j^+ \kappa_i^-}{N} \int_0^{\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_j^+ \kappa_i^-}} \frac{d\chi_{ji}}{1 + \chi_{ji}^\beta} - \frac{\Delta\theta_{ij}^c}{\pi}.$$

Ces intégrales sont calculées conformément aux identités relatives à la fonction hypergéométrique ${}_2F_1(a, b, c, z)$ énumérées à l'annexe A.1. Nous obtenons alors

$$\begin{aligned} \mathcal{I}_7 = & \frac{\Delta\theta_{ij}^c}{\pi} \left[{}_2F_1 \left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, - \left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_i^+ \kappa_j^-} \right)^\beta \right) \right. \\ & \left. + {}_2F_1 \left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, - \left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_j^+ \kappa_i^-} \right)^\beta \right) - 1 \right]. \end{aligned} \quad (1.90)$$

Nous avons \mathcal{I}_8 , l'intégrale du numérateur de l'équation (1.89),

$$\begin{aligned} \mathcal{I}_8 &= \int_0^\pi P(a_{ij} = 1, a_{ji} = 1 | \nu = -1) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}, \\ \mathcal{I}_8 &= \frac{1}{\pi} \int_0^\pi \Delta\theta_{ij} (p_{ij} + p_{ji} - 1) H(p_{ij} + p_{ji} - 1) d\Delta\theta_{ij}, \\ \mathcal{I}_8 &= \frac{1}{\pi} \int_0^\pi \left(\frac{\Delta\theta_{ij}}{1 + \chi_{ij}^\beta} + \frac{\Delta\theta_{ij}}{1 + \chi_{ji}^\beta} - \Delta\theta_{ij} \right) H \left(\frac{1}{1 + \chi_{ij}^\beta} + \frac{1}{1 + \chi_{ji}^\beta} - 1 \right) d\Delta\theta_{ij}. \end{aligned}$$

Par la définition de la fonction Heaviside,

$$\mathcal{I}_8 = \frac{1}{\pi} \int_0^{\Delta\theta_{ij}^c} \left(\frac{\Delta\theta_{ij}}{1 + \chi_{ij}^\beta} + \frac{\Delta\theta_{ij}}{1 + \chi_{ji}^\beta} - \Delta\theta_{ij} \right) d\Delta\theta_{ij}.$$

Nous séparons l'intégrale en trois tel que

$$\mathcal{I}_8 = \frac{1}{\pi} \left[\int_0^{\Delta\theta_{ij}^c} \frac{\Delta\theta_{ij} d\Delta\theta_{ij}}{1 + \chi_{ij}^\beta} + \int_0^{\Delta\theta_{ij}^c} \frac{\Delta\theta_{ij} d\Delta\theta_{ij}}{1 + \chi_{ji}^\beta} - \int_0^{\Delta\theta_{ij}^c} \Delta\theta_{ij} d\Delta\theta_{ij} \right].$$

Un changement de variable dans les intégrandes nous permet d'exprimer les intégrales comme

$$\mathcal{I}_8 = \frac{\left(\frac{2\pi\mu\kappa_i^+ \kappa_j^-}{N} \right)^2}{\pi} \left[\int_0^{\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_i^+ \kappa_j^-}} \frac{\chi_{ij} d\chi_{ij}}{1 + \chi_{ij}^\beta} + \int_0^{\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_j^+ \kappa_i^-}} \frac{\chi_{ij} d\chi_{ij}}{1 + \chi_{ji}^\beta} \right] - \frac{\Delta\theta_{ij}^c}{2\pi}.$$

Ces intégrales sont calculées conformément aux identités relatives à la fonction hypergéométrique ${}_2F_1(a, b, c, z)$ énumérées à l'annexe A.1. Nous obtenons alors

$$\begin{aligned} \mathcal{I}_8 &= \frac{\Delta\theta_{ij}^c}{2\pi} \left[{}_2F_1\left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, -\left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) \right. \\ &\quad \left. + {}_2F_1\left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, -\left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_j^+\kappa_i^-}\right)^\beta\right) - 1\right]. \end{aligned} \quad (1.91)$$

Nous avons finalement l'espérance de la distance angulaire,

$$\begin{aligned} &\mathbb{E}[\Delta\theta_{ij}|P(a_{ij}=1, a_{ji}=1|\nu=0, \xi_{ij}\neq 1)] \\ &= \frac{\Delta\theta_{ij}^c}{2} \left[\frac{{}_2F_1\left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, -\left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) + {}_2F_1\left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, -\left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_j^+\kappa_i^-}\right)^\beta\right) - 1}{{}_2F_1\left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -\left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) + {}_2F_1\left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -\left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_j^+\kappa_i^-}\right)^\beta\right) - 1} \right]. \end{aligned} \quad (1.92)$$

Il est à noter que la distance angulaire critique $\Delta\theta_{ij}^c$ est solution de l'équation donnée par l'argument de la fonction Heaviside de l'équation (1.69),

$$p_{ij} + p_{ji} - 1 = 0. \quad (1.93)$$

Nous pouvons exprimer l'équation ainsi,

$$1 = \frac{1}{1 + \left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_i^+\kappa_j^-}\right)^\beta} + \frac{1}{1 + \left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_j^+\kappa_i^-}\right)^\beta}. \quad (1.94)$$

De manière à alléger la notation, nous allons effectuer un changement de variable tel que

$$x = \Delta\theta_{ij}^c, \quad (1.95)$$

$$\kappa = \frac{N}{2\pi\mu\kappa_i^+\kappa_j^-}, \quad (1.96)$$

$$\kappa' = \frac{N}{2\pi\mu\kappa_j^+\kappa_i^-}, \quad (1.97)$$

et résoudre l'équation

$$1 = \frac{1}{1 + (\kappa x)^\beta} + \frac{1}{1 + (\kappa' x)^\beta}, \quad (1.98)$$

dont la solution est

$$x = \frac{1}{\sqrt{\kappa\kappa'}}. \quad (1.99)$$

La solution pour $\Delta\theta_{ij}^c$ est donc

$$\Delta\theta_{ij}^c = \frac{2\pi\mu\sqrt{\kappa_i^+\kappa_j^-\kappa_j^+\kappa_i^-}}{N}. \quad (1.100)$$

À partir de l'équation (1.74), il est possible de calculer l'espérance de la distance angulaire pour toutes les probabilités conjointes des équations (1.67) où a_{ij} et a_{ji} prennent les valeurs de 0 ou 1 et en exprimant ces probabilités selon la probabilité de connexion réciproque de l'équation (1.69). L'espérance de la distance angulaire entre deux sommets pour les différentes probabilités de connexion sont données par les équations suivantes :

$$\begin{aligned} \mathbb{E}[\Delta\theta_{ij}|P(a_{ij}=1, a_{ji}=1)] = \\ \left\{ \begin{array}{ll} \frac{\pi}{2} \left[\frac{H(1-\xi_{ij})_2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) + H(\xi_{ij}-1)_2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right)^\beta\right)}{H(1-\xi_{ij})_2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) + H(\xi_{ij}-1)_2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right)^\beta\right)} \right] & \nu = 1 \\ \frac{\pi}{2} \left[\frac{2F_1\left(2, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right)}{2F_1\left(2, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right)} = \frac{\pi}{2} \frac{2F_1\left(2, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right)^\beta\right)}{2F_1\left(2, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right)^\beta\right)} \right] & \xi_{ij} = 1 \\ \frac{\pi}{2} \left[\frac{2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) - \xi_{ij}^\beta 2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right)^\beta\right)}{2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) - \xi_{ij}^\beta 2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right)^\beta\right)} \right] & \xi_{ij} \neq 1 \\ \frac{\Delta\theta_{ij}^c}{2} \left[\frac{2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) + 2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_j^+\kappa_i^-}\right)^\beta\right) - 1}{2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) + 2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_j^+\kappa_i^-}\right)^\beta\right) - 1} \right] & \nu = -1 \end{array} \right. \end{aligned} \quad (1.101a)$$

$$\begin{aligned} & \mathbb{E}[\Delta\theta_{ij}|P(a_{ij}=1, a_{ji}=0)] = \\ & \left\{ \begin{array}{l} \frac{\pi}{2} \left[\frac{2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+ \kappa_j^-}\right)^\beta\right) - 2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+ \kappa_i^-}\right)^\beta\right)}{2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+ \kappa_j^-}\right)^\beta\right) - 2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+ \kappa_i^-}\right)^\beta\right)} \right] \\ \frac{\pi}{2} \left[\frac{2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+ \kappa_j^-}\right)^\beta\right) - 2F_1\left(2, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+ \kappa_j^-}\right)^\beta\right)}{2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+ \kappa_j^-}\right)^\beta\right) - 2F_1\left(2, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+ \kappa_j^-}\right)^\beta\right)} \right] \\ \frac{\pi}{2} \left[\frac{2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+ \kappa_j^-}\right)^\beta\right) - 2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+ \kappa_i^-}\right)^\beta\right)}{2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+ \kappa_j^-}\right)^\beta\right) - 2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+ \kappa_i^-}\right)^\beta\right)} \right] \\ \frac{\pi}{2} \left[\frac{2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+ \kappa_j^-}\right)^\beta\right) - 2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+ \kappa_i^-}\right)^\beta\right)}{2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+ \kappa_j^-}\right)^\beta\right) - 2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+ \kappa_i^-}\right)^\beta\right)} \right] \\ \frac{\Delta\theta_{ij}^c}{2} \left[\frac{\left(\left(\frac{\pi}{\Delta\theta_{ij}^c}\right)^2 - 1\right) 2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+ \kappa_j^-}\right)^\beta\right) - 2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_j^+ \kappa_i^-}\right)^\beta\right) + 1}{\left(\left(\frac{\pi}{\Delta\theta_{ij}^c}\right) - 1\right) 2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+ \kappa_j^-}\right)^\beta\right) - 2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_j^+ \kappa_i^-}\right)^\beta\right) + 1} \right] \end{array} \right\} \end{aligned}$$

$\nu = 1$
 $\xi_{ij} = 1^*$
 $\nu = 0$
 $\xi_{ij} \neq 1$
 $\nu = -1$

$$\begin{aligned} \mathbb{E}[\Delta\theta_{ij}|P(a_{ij}=0, a_{ji}=1)] &= \\ \left\{ \begin{array}{l} \frac{\pi}{2} \left[\frac{2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right)^\beta\right) - 2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right)}{2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right)^\beta\right) - 2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right)} \right] \\ \frac{\pi}{2} \left[\frac{2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) - 2F_1\left(2, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right)}{2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) - 2F_1\left(2, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right)} \right] \\ \frac{\pi}{2} \left[\frac{2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right)^\beta\right) - 2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right)}{2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right)^\beta\right) - 2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^\beta\right)} \right] \\ \frac{\Delta\theta_{ij}^c}{2} \left[\left(\left(\frac{\pi}{\Delta\theta_{ij}^c} \right)^2 - 1 \right) 2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right)^\beta\right) - 2F_1\left(1, \frac{2}{\beta}, 1+\frac{2}{\beta}, -\left(\frac{N\Delta\theta_{ij}^c}{2\pi\mu\kappa_i^+\kappa_j^-}\right)^\beta\right) + 1 \right] \end{array} \right. \\ \nu = 1 & \quad \xi_{ij} = 1^* \\ \nu = 0 & \quad \xi_{ij} \neq 1 \\ \nu = -1 & \quad (1.101c) \end{aligned}$$

* L'équation est aussi valide en remplaçant le produit $\kappa_i^+ \kappa_j^-$ par $\kappa_j^+ \kappa_i^-$ dans le cas où $\xi_{ij} = 1$.

Chapitre 2

Propriétés topologiques et géométrie de réseaux orientés empiriques

Topological Properties and Geometry of Empirical Directed Networks

Jérémie Lesage^{1, 2} and Antoine Allard^{1, 2}

¹ Département de physique, de génie physique et d'optique,
Université Laval, Québec, QC, Canada

² Centre Interdisciplinaire en Modélisation Mathématique de l'Université Laval,
Université Laval, Québec, QC, Canada

2.1 Résumé

Nous introduisons diMercator, un algorithme de maximisation de la vraisemblance conçu pour plonger des réseaux orientés dans l'espace $\mathbb{S}^1 \times \mathbb{R}_+ \times \mathbb{R}_+$ en utilisant le modèle \mathcal{S}^1 orienté-réiproque. L'algorithme infère les paramètres du modèle afin d'optimiser la correspondance avec la structure du réseau observé. Adapté de l'algorithme Mercator, son homologue pour les réseaux non orientés, diMercator réalise des plongements plus spécifiques pour les réseaux orientés en tenant compte de la directionnalité inhérente des interactions. Cette amélioration met en évidence l'importance de modéliser l'asymétrie des interactions dans les graphes orientés. Nos évaluations montrent que diMercator atteint une meilleure fidélité de plongement que Mercator sur une sélection d'ensembles de données de référence. De plus, notre approche permet une réduction dimensionnelle de $N(N - 1)$ à $3N + 2$ paramètres tout en conservant approximativement plusieurs propriétés topologiques essentielles. Parmi ces propriétés clés des réseaux orientés, on retrouve notamment la cohérence trophique, la non-normalité, les confi-

gurations de triangles, la réciprocité individuelle, ainsi que la classification des noeuds par composante du graphe et les informations tirées de la décomposition en valeurs singulières.

2.2 Abstract

We introduce diMercator, a maximum likelihood estimation algorithm designed to embed real-world directed networks into $\mathbb{S}^1 \times \mathbb{R}_+ \times \mathbb{R}_+$ space using the directed-reciprocal \mathcal{S}^1 model. The algorithm infers the model parameters to optimize alignment with the observed network structure. Adapted from the Mercator algorithm, its counterpart for undirected networks, diMercator provides more accurate embeddings of directed networks by accounting for the inherent directionality of the interactions. This improvement underscores the importance of modeling interaction asymmetry in directed graphs. Our evaluations show that diMercator achieves higher embedding fidelity compared to Mercator on benchmark datasets. Moreover, our approach enables a dimensionality reduction from $N(N - 1)$ to $3N + 2$ parameters while approximately preserving several essential topological properties. Our algorithm captures key properties of directed networks, including trophic coherence, non-normality, triangle configurations, individual reciprocities, as well as bow-tie structure and insights derived from singular value decomposition.

2.3 Introduction

Over the last decade, network geometry has emerged as a powerful paradigm for analyzing complex systems, offering geometric representations that enable deeper insights into the structure and behavior of real-world complex networks [7, 55]. This approach assumes that the nodes of the network are embedded in an underlying metric space and that each node has a given position in this space. Thus, connected nodes are expected to be closer in the underlying metric space and, conversely, nodes closer in the underlying metric space are more likely to be connected. Embedding networks in hyperbolic space has notably enabled researchers to optimize network navigation, and model processes such as information diffusion with greater accuracy [57–59].

Traditional approaches to network geometry have primarily focused on undirected networks, where the lack of edge direction simplifies the embedding process and allows for straightforward interpretations of the geometric representation [8, 56]. However, directed networks, which are prevalent in many real-world systems such as the World Wide Web, neural networks, and social networks, introduce additional layers of complexity due to their inherent asymmetry [6, 27]. The directionality of interactions in these networks reflects causal or hierarchical relationships that cannot be captured by undirected models [60]. As a result, embedding directed networks in hyperbolic space poses significant challenges, as it requires the development of new frameworks capable of preserving the directional characteristics while

maintaining the geometric properties of the network structure. For example, quantifying clustering in directed networks is not as straightforward as it is in undirected networks since the directionality of the interaction between two nodes gives rise to seven different types of triangles [61, 62]. Furthermore, quantifying the level of reciprocity, which can be defined as the ratio of the number of reciprocal edges over the total number of edges, gives insight into the global directionality in the network as a whole as well as in the individual reciprocity of each node.

We propose a novel algorithm that extends existing geometric embedding frameworks to accommodate asymmetric interactions, thereby providing a more accurate and meaningful representation of these networks. Our approach not only retains the essential directional information but also leverages the advantages of hyperbolic embeddings while using the embedding space $\mathbb{S}^1 \times \mathbb{R}_+ \times \mathbb{R}_+$. The following work is adapted from Ref. [12] to directed networks and focuses on the inference of the angular positions in the directed-reciprocal \mathcal{S}^1 model from Ref. [16]. We name the resulting algorithm *diMercator*.

2.4 Methods

The directed-reciprocal \mathcal{S}^1 model

In the directed-reciprocal \mathcal{S}^1 model [16], the probability for a directed edge to exist from node i to node j is given by

$$p_{ij} = p(a_{ij} = 1 | \kappa_i^+, \kappa_j^-, \Delta\theta_{ij}) = \left[1 + \left(\frac{R\Delta\theta_{ij}}{\mu\kappa_i^+\kappa_j^-} \right)^\beta \right]^{-1} \quad (2.1)$$

where a_{ij} is the adjacency matrix element (e.g., $a_{ij} = 1$ if there is an edge from node i to node j and $a_{ij} = 0$ otherwise), κ_i^+ is the hidden out-degree of node i , κ_j^- is the hidden in-degree of node j , $\Delta\theta_{ij} = \min(|\theta_i - \theta_j|, 2\pi - |\theta_i - \theta_j|)$ is the angular separation between the nodes, $\beta > 1$ is a clustering parameter, $\mu = \beta \sin(\pi/\beta)/(2\pi\mathbb{E}[\kappa])$, $R = N/2\pi$ and N is the number of nodes. Note that $\mathbb{E}[\kappa] := \mathbb{E}[\kappa^+] = \mathbb{E}[\kappa^-]$ is the expected value of the hidden degrees where we set the first moment of the joint distribution $\rho(\kappa^+, \kappa^-)$.

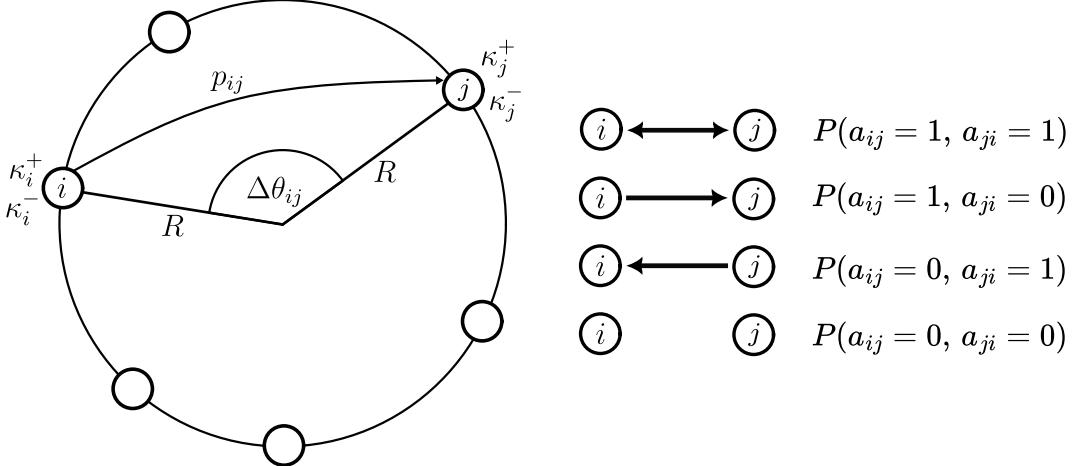


Figure 2.1 – Representation of the directed-reciprocal \mathcal{S}^1 model with the four types of interaction between two nodes.

Figure 2.1 illustrates the directed-reciprocal \mathcal{S}^1 model. Since our approach focuses on each pair of directed edges between two nodes, we define four joint connection probabilities which are conditioned to their dependencies over hidden degrees and angular distance (i.e., κ_i^+ , κ_i^- , κ_j^+ , κ_j^- , $\Delta\theta_{ij}$) and have the following relations:

$$\begin{aligned} &P(a_{ij} = 0, a_{ji} = 0) + P(a_{ij} = 1, a_{ji} = 0) \\ &\quad + P(a_{ij} = 0, a_{ji} = 1) + P(a_{ij} = 1, a_{ji} = 1) = 1, \end{aligned} \quad (2.2a)$$

$$P(a_{ij} = 1, a_{ji} = 0) + P(a_{ij} = 1, a_{ji} = 1) = p_{ij}, \quad (2.2b)$$

$$P(a_{ij} = 0, a_{ji} = 1) + P(a_{ij} = 1, a_{ji} = 1) = p_{ji}. \quad (2.2c)$$

Equations (2.2) leave one of the four probabilities $P(a_{ij}, a_{ji})$ undefined, giving the model an extra degree of freedom to fix the reciprocity. This can be done by considering the correlation coefficient

$$\rho_{ij} = \frac{\mathbb{E}[a_{ij}a_{ji}] - \mathbb{E}[a_{ij}]\mathbb{E}[a_{ji}]}{\sqrt{(\mathbb{E}[a_{ij}^2] - \mathbb{E}[a_{ij}]^2)(\mathbb{E}[a_{ji}^2] - \mathbb{E}[a_{ji}]^2)}} = \frac{P_{ij}(1,1) - p_{ij}p_{ji}}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}}. \quad (2.3)$$

where $\mathbb{E}[\cdot]$ corresponds to an average over the graph ensemble defined by the joint probabilities [16]. Note that, because $P(a_{ij} = 1, a_{ji} = 1) \in [0,1]$, Eq. (2.3) is not guaranteed to be bounded between -1 and 1 . Enforcing these bounds yields an expression for $P(a_{ij} = 1, a_{ji} = 1)$ in terms of p_{ij} , p_{ji} and a parameter $\nu \in [-1,1]$ controlling the level of reciprocity between nodes i and j ,

$$P(a_{ij} = 1, a_{ji} = 1) = \begin{cases} (1+\nu)p_{ij}p_{ji} - \nu(p_{ij} + p_{ji} - 1)H(p_{ij} + p_{ji} - 1) & -1 \leq \nu \leq 0 \\ (1-\nu)p_{ij}p_{ji} + \nu \min\{p_{ij}, p_{ji}\} & 0 \leq \nu \leq 1 \end{cases}, \quad (2.4)$$

where $H(\cdot)$ is the Heaviside step function. The distribution of angular distance $\Delta\theta_{ij}$ between two reciprocally connected nodes is obtained from Bayes' rule as

$$\rho(\Delta\theta_{ij}|a_{ij} = 1, a_{ji} = 1, \kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-) = \frac{P(a_{ij} = 1, a_{ji} = 1 | \kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-, \Delta\theta_{ij}) \rho(\Delta\theta_{ij})}{P(a_{ij} = 1, a_{ji} = 1 | \kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-)}, \quad (2.5)$$

where $P(a_{ij} = 1, a_{ji} = 1 | \kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-, \Delta\theta_{ij})$ is obtained from Eq. (2.4) and the angular distance density on the circle is $\rho(\Delta\theta_{ij}) = 1/\pi$. We obtain the normalization factor by integrating the numerator over $\Delta\theta_{ij}$ as

$$P(a_{ij} = 1, a_{ji} = 1 | \kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-) = \int_0^\pi P(a_{ij} = 1, a_{ji} = 1 | \kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-, \Delta\theta_{ij}) \rho(\Delta\theta_{ij}) d\Delta\theta_{ij}. \quad (2.6)$$

Since the geometric aspect of the model is solely encoded in the angular position of each node and that our framework focuses on modeling the type of interaction existing between each pair of nodes, we can derive the expected angular distance between each pair of connected nodes. The expected angular distance between nodes i and j in the directed-reciprocal \mathcal{S}^1 model given that they are connected is given by

$$\mathbb{E} [\Delta\theta_{ij} | a_{ij} = 1, a_{ji} = 1, \kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-] = \int_0^\pi \Delta\theta_{ij} \rho(\Delta\theta_{ij} | a_{ij} = 1, a_{ji} = 1, \kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-) d\Delta\theta_{ij}, \quad (2.7)$$

which integral we can compute numerically. From Eqs. (2.2) and Eq. (2.7), we can easily obtain the expected distance from any of the four joint connection probabilities.

In the limit where the clustering parameter $\beta \rightarrow 0$, the directed-reciprocal \mathcal{S}^1 model recovers its non-geometric counterpart, the directed-reciprocal soft configuration model, but unlike the latter model, it has a nonvanishing clustering in the limit $N \rightarrow \infty$ due to the triangle inequality of its embedding space [16]. The directed-reciprocal soft configuration model is thus defined as the unique ensemble of unbiased sparse random graphs whose entropy is maximized across all graphs with a given expected joint in- and out-degree distribution [16]. It consists of N nodes, each of which is assigned a pair of hidden degrees κ^- and κ^+ according to the distribution $\rho(\kappa^-, \kappa^+)$. In this model, the probability for a directed edge to exist from node i to node j is given by

$$p_{ij} = p(a_{ij} = 1 | \kappa_i^+, \kappa_j^-) = \left[1 + \frac{N\mathbb{E}[\kappa]}{\kappa_i^+ \kappa_j^-} \right]^{-1}, \quad (2.8)$$

where $\mathbb{E}[\kappa]$ is the expected value of the hidden degrees. Note that the model has the same marginal probabilities as the ones from Eqs. (2.2). The directed-reciprocal \mathcal{S}^1 model can therefore be seen as the geometric extension of the directed-reciprocal soft configuration model.

The diMercator inference algorithm

We propose an inference algorithm adapted from the Mercator algorithm from Ref. [12]. The algorithm, named diMercator, embeds directed graphs in $\mathbb{S}^1 \times \mathbb{R}_+ \times \mathbb{R}_+$ space by inferring

the values of the $3N + 2$ parameters of the directed-reciprocal \mathcal{S}^1 model. Of these $3N + 2$ parameters, β , ν and $\kappa := (\kappa_1^+, \kappa_1^-, \dots, \kappa_N^+, \kappa_N^-)$ are inferred with the algorithm proposed by Ref. [16]. The N missing parameters, the angular coordinates $\theta := (\theta_1, \dots, \theta_N)$, are then inferred by adapting the approach of Ref. [12] to the formalism introduced in Ref. [16]. The inference of the angular coordinates is performed in two steps: we first initialize the angular coordinates using the Laplacian Eigenmaps (LE) algorithm with the expected angular distance from Eqs. (2.2) and Eq. (2.7) as input and then we maximize the likelihood that the model generated the original graph one node at the time with a sampling procedure. Subsequently, we can also update the hidden degrees based on the inferred positions. The detailed procedure for the diMercator algorithm is described in section 2.7.1 of the supplementary material. Notably, the computational complexity of the diMercator algorithm scales quadratically as $\mathcal{O}(N^2)$ for networks with N nodes.

2.5 Results

The diMercator algorithm addresses the specific challenges of modeling directed networks, and the following results serve as a proof of concept to demonstrate its effectiveness and accuracy at reproducing the structure of real-world directed networks. We tested diMercator on various empirical network datasets listed in Sec. 2.7.4 of the supplementary material.

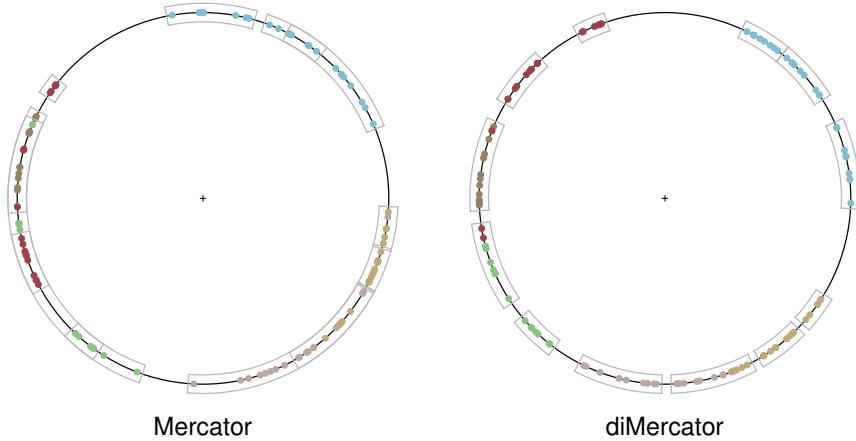


Figure 2.2 – Embedding on the \mathbb{S}^1 circle with Mercator and diMercator of the contacts between high school students (dataset: `sp_high_school_diaries` [63]). In the network, nodes represent students and edges represent their friendships. In the embeddings, nodes are colored according to each student’s classroom group. The boxes correspond to the geometric communities on the circle that maximize the modularity in the original network.

We first compare diMercator to Mercator, its equivalent for undirected networks. To that end, we embed a directed network with diMercator and use the undirected network corre-

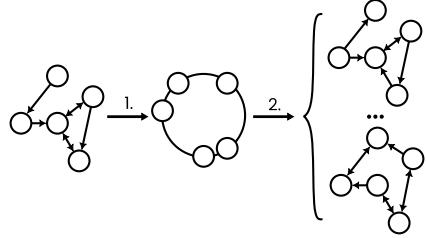


Figure 2.3 – Procedure for validating the diMercator algorithm on empirical network datasets. Step 1: we perform an embedding of an empirical directed network with diMercator. Step 2: we generate a specified number of synthetic directed networks from this embedding (step 1). The properties of these synthetic networks are then compared to those of the original empirical network.

sponding to this directed network for Mercator. Focusing solely on the embedding process and excluding other network properties, diMercator demonstrates improvements over Mercator. As shown in Fig. 2.2, same color nodes (or students belonging to the same classroom group) are less scattered all over the circle in the diMercator embedding compared to the Mercator embedding. For instance, it is challenging to distinguish between the groups dominated by green, red and grey nodes in the Mercator embedding. In contrast, diMercator more clearly separates these communities into distinct green, red and grey groups. Additionally, diMercator reduces overlap between classroom communities, minimizing misassociation between a node (student and their actual classroom group) and its geometric community on the circle (predicted classroom group). Further analysis of the comparison between diMercator and Mercator is provided in Sec. 2.7.2 of the supplementary material.

We also tested diMercator on various real-world directed networks. First, we embedded each empirical network using the algorithm. Then, we generated a specified number of synthetic directed networks using the inferred parameters with Eqs. (2.1)–(2.4). This procedure is illustrated in Fig. 2.3. Figure 2.4 shows how the diMercator algorithm reproduces topological properties observed in real-world networks. Regarding trophic incoherence and normality, the values computed from the synthetic networks are centered on the empirical value. Sec. 2.7.3 of the supplementary material provides a thorough derivation of these two metrics. Regarding the triangle configurations, the synthetic networks consistently encompass the observed value in the original empirical network. For individual reciprocities, while the relative frequencies differ, the two distributions exhibit similar trends.

Furthermore, regarding the singular value spectrum resulting from the singular value decomposition (SVD) of the networks' adjacency matrices, the diMercator singular value spectrum closely matches the spectrum of the empirical network, suggesting that the adjacency matrices of the synthetic networks closely resemble those of the empirical network and that diMercator captures the empirical network's structure with great accuracy. Indeed, comparing the singular value spectrum of two networks can provide insights into their structural similarities and

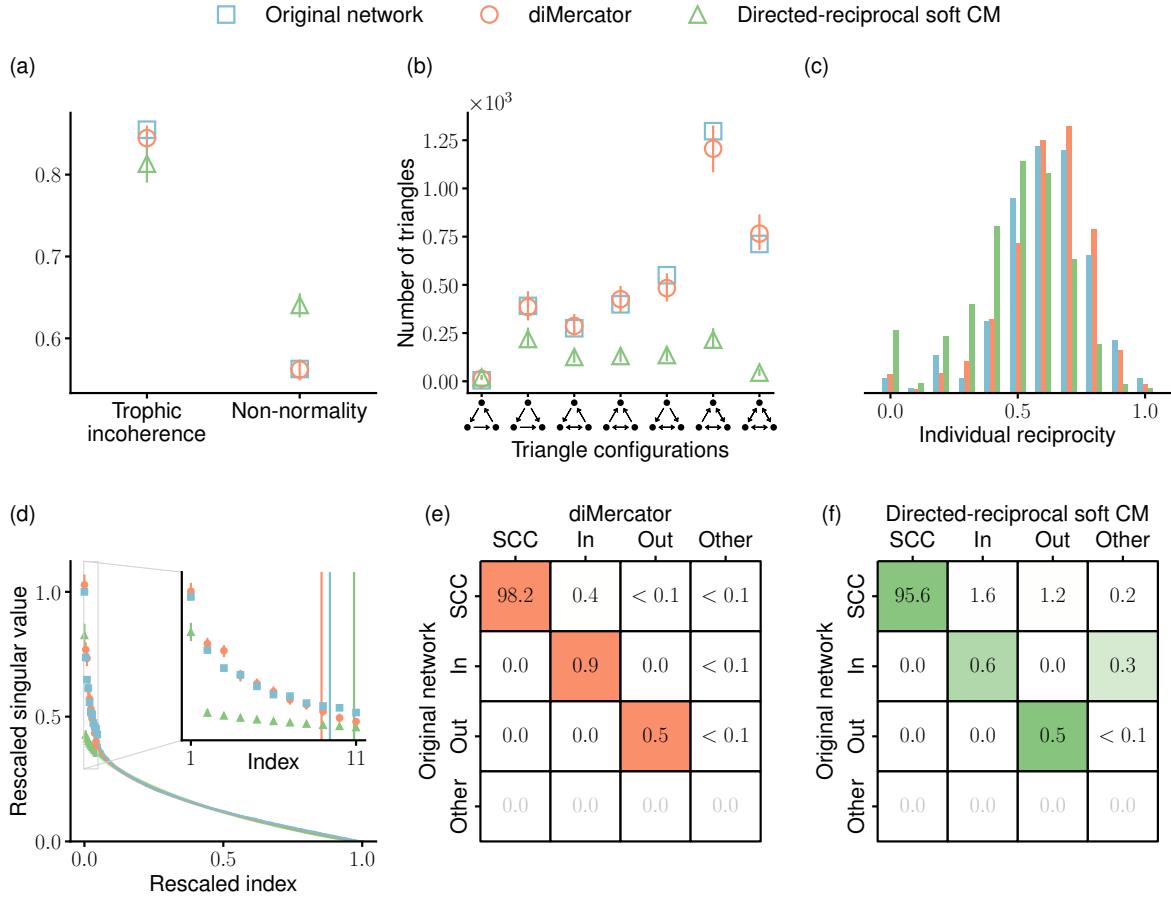


Figure 2.4 – Reproducing topological properties such as (a) trophic incoherence and non-normality, (b) triangle spectrum, (c) individual reciprocity (i.e., each node’s reciprocity), (d) singular value spectrum (with vertical bars denoting the stable rank) and (e)-(f) bow-tie structure as confusion matrices (with numerical values representing total percentage of nodes and color grading showing row normalization) of the friendships among students living in a residence hall (dataset: `residence_hall` [64]). A total of 1000 synthetic networks were generated. Error bars in panels (a), (b) and (d) correspond to a 95% confidence interval.

differences. The spectrum serves as a compact representation of the adjacency matrix and reflects the dimensionality of the network. It has been shown that the decrease of singular values is an important indicator of the effective dimension of a network and that this decrease is typically rapid for empirical networks [23]. As for the stable rank, the value computed from the synthetic networks is close to the empirical value, further assessing the above conclusions. The definition of the stable rank and its relation to the SVD can be found in Sec. 2.7.3 of the supplementary material.

Finally, diMercator replicates very accurately the bow tie structure of the empirical network. On this basis, nodes are classified according to their membership to a component of the directed graph (i.e., strongly connected component (scc), in-component, out-component or other components). The bow-tie structure is thoroughly described in Sec. 2.7.3 of the supplementary material. We use a confusion matrix [65] to evaluate the performance of our algorithm in correctly associating nodes with the appropriate components of the bow-tie structure. In our framework, each row of the confusion matrix corresponds to an actual class (i.e., components of the original network), while each column corresponds to a predicted class (i.e., components of the synthetic networks). Furthermore, we clearly see the increased accuracy of diMercator over the directed-reciprocal soft configuration model at reproducing topological properties since diMercator’s results are always closer to the ground truth than those of its non-geometric counterpart, highlighting the importance of geometry in empirical directed networks.

Additional analyses of thirty other real-world directed networks are included in Sec. 2.7.3 of the supplementary material, further supporting the accuracy of diMercator.

2.6 Conclusion

We have adapted the paradigm of network geometry to directed graphs by developing an embedding algorithm that preserves the asymmetry of interactions using the directed-reciprocal \mathcal{S}^1 model. The relevance of diMercator for studying directed networks has been demonstrated through comparisons to Mercator, its counterpart for undirected networks. Our results show that diMercator captures key properties observed in real-world networks, such as trophic coherence, non-normality, triangle configurations, individual reciprocity, singular value decomposition and bow tie structure. Such properties are not directly fitted by the model, yet they are well reproduced. We have expanded our analysis by validating diMercator on several datasets, including social networks, connectomes, trophic networks, and transcription networks. The network geometry toolbox has proven its usefulness time and time again in the past, but it had a significant blind spot that we have now begun to address with diMercator. Even though the connection between the directed-reciprocal \mathcal{S}^1 model and hyperbolic space has not been formally established, our work constitutes a significant conceptual and method-

ological advancement by proposing a compact probabilistic graph model that, beyond the sole assumption of hyperbolicity, enables substantial dimensionality reduction, from $N(N - 1)$ to $3N + 2$ parameters, while approximately preserving several essential topological properties. This contribution stands as a central achievement in its own right. Looking ahead, establishing a formally proven connection between the model and the underlying hyperbolic space, rather than relying on intuitive analogies, would represent a valuable contribution to the field. Furthermore, it would be interesting to further adapt the network geometry paradigm to directed graphs by developing a directed \mathcal{S}^D model [66] capable of producing multidimensional embeddings of real networks on the D -sphere \mathbb{S}^D or even by developing a directed version of the \mathcal{H}^2 model [8], which would close the loop on hyperbolic embedding models adapted to directed networks. While this is a challenging task, it holds significant promise for advancing the study of such complex systems.

2.7 Supplementary Material

2.7.1 The diMercator inference algorithm

The diMercator algorithm embeds directed graphs on the circle by inferring the values of the $3N + 2$ parameters of the directed-reciprocal \mathcal{S}^1 model. Of these $3N + 2$ parameters, β , ν and $\boldsymbol{\kappa} := (\kappa_1^+, \kappa_1^-, \dots, \kappa_N^+, \kappa_N^-)$ are inferred with the algorithm proposed by Ref. [16]. The N missing parameters, the angular coordinates $\boldsymbol{\theta} := (\theta_1, \dots, \theta_N)$, are then inferred by adapting the approach of Ref. [12] to the formalism introduced in Ref. [16].

Inferring the initial positions and ordering

We use a modified version of the Laplacian Eigenmaps (LE) algorithm from Ref. [67] which generates a weighted graph from a given set of points by assigning a distance between two points if they are connected and mapping them to \mathbb{R}^2 . This embedding in \mathbb{R}^2 can be radially normalized so that all points lay in \mathbb{S}^1 space. Note that the algorithm was originally designed for undirected graphs but that we can adapt it to directed graphs by symmetrizing the weight matrix. The procedure is as follows:

1. Since geometric information comes from the triangles, which are the topological manifestation of the triangle inequality in the metric space, we remove degree-one nodes and work with the subgraph of nodes with undirected degree (i.e., degree from the corresponding undirected graph) greater than one. We construct the weight matrix W for which the elements w_{ij} are assigned some value

$$w_{ij} = e^{-d(i,j)^2/t}, \quad (2.9)$$

where $d(i,j)$ is the distance between the points that we replace with the expected dis-

tance from Eq. (2.7)

$$d(i,j) = 2 \cdot \sin \left(\frac{1}{2} \cdot \mathbb{E} \left[\Delta\theta_{ij} | a_{ij}, a_{ji}, \kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^- \right] \right), \quad (2.10)$$

and t is a scaling factor fixed as the average of the squares of all distances.

2. Since the graph is directed, the weight matrix is not symmetric and we symmetrize it according to

$$W^{\text{sym}} = \frac{W + W^\top}{2}. \quad (2.11)$$

3. We compute the eigenvalues and eigenvectors for the generalized eigenvalue problem

$$L\mathbf{v} = \lambda D\mathbf{v}, \quad (2.12)$$

where D is the diagonal weight matrix which entries $d_{ii} = \sum_j w_{ij}^{\text{sym}}$ are the sums of each row of W^{sym} and $L = D - W^{\text{sym}}$ is the Laplacian matrix of the symmetrized graph.

4. We assign to every point i the coordinates $\mathbf{y}_i = (v_1^i, v_2^i)$, where v_j^i is the i th component of the j th eigenvector with non-null eigenvalue (the eigenvectors are ordered according to their eigenvalues). It can be shown that these coordinates minimize the weighed squared distances between connected pairs $\epsilon = \sum_{i,j} |\mathbf{v}_i - \mathbf{v}_j|^2 w_{ij}$.
5. The LE algorithm yields a sequence of angular coordinates $S = (\theta_1, \dots, \theta_N)$ for the nodes in the subgraph, where the angles in S are ordered in increasing order and computed as

$$\theta_i = \text{atan2}(v_2^i, v_1^i), \quad (2.13)$$

where v_1^i and v_2^i are the x and y coordinates of node i found by LE.

6. Degree-one nodes are reintroduced in the sequence S by replacing every node i with l undirected degree-one neighbors by the sequence $(n_1^i, \dots, n_{\lfloor l/2 \rfloor}^i, i, n_{\lfloor l/2 \rfloor + 1}^i, \dots, n_l^i)$, where n_j^i is the j th undirected degree-one neighbor of node i in arbitrary order.

Having obtained an initial ordering of the nodes, we set every node's coordinate such that the angular separation between two consecutive nodes in S is proportional to its expected value as derived from Eq. (2.7). The procedure is as follows:

1. Let g_i be the angular separation between two consecutive nodes i and $i + 1$ in S . The density of g_i can be obtained using Bayes' rule as

$$\rho(g_i | a_{i,i+1}, a_{i+1,i}) = \frac{P(a_{i,i+1}, a_{i+1,i} | g_i) \rho(g_i)}{\int_0^\pi P(a_{i,i+1}, a_{i+1,i} | g_i) \rho(g_i) dg_i}, \quad (2.14)$$

where the joint connection probability $P(a_{i,i+1}, a_{i+1,i} | g_i)$ is obtained from Eqs. (2.2) and Eq. (2.4) and where $\rho(g_i) = N/2\pi \cdot \exp\{-N/2\pi \cdot g_i\}$ is the exponential distribution of parameter $N/2\pi$. Note that we omitted the hidden degrees dependency (i.e., $\kappa_i^+, \kappa_i^-, \kappa_{i+1}^+, \kappa_{i+1}^-$) to simplify the notation. The expected angular separation $\mathbb{E}[g_i]$ is then computed as

$$\mathbb{E}[g_i] = \frac{\int_0^\pi g_i P(a_{i,i+1}, a_{i+1,i} | g_i) \rho(g_i) dg_i}{\int_0^\pi P(a_{i,i+1}, a_{i+1,i} | g_i) \rho(g_i) dg_i}, \quad (2.15)$$

and we define

$$g_i = \max(\mathbb{E}[g_i], g_i^{\text{LE}}), \quad (2.16)$$

where $g_i^{\text{LE}} = \min(|\theta_{i+1} - \theta_i|, 2\pi - |\theta_{i+1} - \theta_i|)$ is computed from the sequence S .

2. By applying the last step to every pair of consecutive nodes, we obtain a sequence of N expected angular separations which need to sum up to 2π , so we normalize each g_i according to

$$\tilde{g}_i = 2\pi \frac{g_i}{\sum_{i=1}^N g_i}. \quad (2.17)$$

Finally, we assign every node's coordinate sequentially starting with $\theta_1 = 0$ and following as $\theta_i = \theta_{i-1} + \tilde{g}_{i-1}$, $i = 2, \dots, N$.

Maximizing the likelihood

This step of the embedding algorithm adjusts the angular coordinates that maximize the likelihood for the observed network to be generated by the model. Henceforth, we visit every node once and propose several new angular coordinates for it, keeping the one with higher log-likelihood. We visit the nodes in the order defined by the corresponding undirected network's onion decomposition [68] and break ties randomly. For every node i , we select the optimal coordinate among candidate positions generated in the vicinity of the average angular coordinate of its neighbors (both in-going and out-going). The procedure is as follows:

1. Let $\mathcal{N}(i)$ be the neighbors (both in-going and out-going) of node i and $k_i = |\mathcal{N}(i)|$ be the undirected degree of node i . Since nodes lay on a circle, the average angular coordinate $\bar{\theta}_i$ is given by

$$\bar{\theta}_i = \text{atan2} \left(\sum_{j \in \mathcal{N}(i)} \frac{\sin \theta_j}{(\kappa_j^+ + \kappa_j^-)^2}, \sum_{j \in \mathcal{N}(i)} \frac{\cos \theta_j}{(\kappa_j^+ + \kappa_j^-)^2} \right), \quad (2.18)$$

where the hidden degrees' arithmetic mean in the above expression weights the contribution of every neighbor's position. Note that the average angular coordinate $\bar{\theta}_i$ is rescaled in the interval $[0, 2\pi]$.

2. We draw $\lfloor 100 \cdot \max(\ln N, 1) \rfloor$ candidate angular coordinates from a normal distribution of mean $\bar{\theta}_i$ and standard deviation

$$\sigma_i = \max \left(\frac{\pi}{N}, \frac{\Delta\theta_i^{\max}}{2} \right), \quad (2.19)$$

where N is the number of nodes and $\Delta\theta_i^{\max}$ is the angular distance between $\bar{\theta}_i$ and the most distant neighbor of node i ,

$$\Delta\theta_i^{\max} = \max_{j \in \mathcal{N}(i)} \{ \min(|\theta_j - \bar{\theta}_i|, 2\pi - |\theta_j - \bar{\theta}_i|) \}. \quad (2.20)$$

3. We compute the local log-likelihood of each candidate position and use the one which maximizes it according to

$$\ln \mathcal{L}_i = \sum_{j \neq i} \ln P(a_{ij}, a_{ji} | \kappa_i^+, \kappa_i^-, \kappa_j^+, \kappa_j^-, \Delta\theta_{ij}). \quad (2.21)$$

Adjusting the hidden degrees

We now optimize the hidden degrees such that the expected degrees of the directed-reciprocal \mathcal{S}^1 model's graphs obtained from the embedding match the degrees of the original graph according to $\mathbb{E}[k_i^+ | \boldsymbol{\kappa}, \boldsymbol{\theta}] = k_i^+$ and $\mathbb{E}[k_i^- | \boldsymbol{\kappa}, \boldsymbol{\theta}] = k_i^-$.

$$\epsilon^{\max} = \max_{i \in \{1, \dots, N\}} \left\{ \max \left(|\mathbb{E}[k_i^+ | \boldsymbol{\kappa}, \boldsymbol{\theta}] - k_i^+|, |\mathbb{E}[k_i^- | \boldsymbol{\kappa}, \boldsymbol{\theta}] - k_i^-| \right) \right\}. \quad (2.22)$$

While ϵ^{\max} is greater than some threshold ϵ , the hidden degrees of each node are updated according to

$$\kappa_i^+ \leftarrow \kappa_i^+ + [k_i^+ - \mathbb{E}[k_i^+ | \boldsymbol{\kappa}, \boldsymbol{\theta}]] \cdot u^+, \quad (2.23a)$$

$$\kappa_i^- \leftarrow \kappa_i^- + [k_i^- - \mathbb{E}[k_i^- | \boldsymbol{\kappa}, \boldsymbol{\theta}]] \cdot u^-, \quad (2.23b)$$

where u^+ and u^- are random variables drawn from a uniform distribution $U(0,1)$. The expected degrees are computed according to

$$\mathbb{E}[k_i^+ | \boldsymbol{\kappa}, \boldsymbol{\theta}] = \sum_{\substack{j=1 \\ j \neq i}}^N \frac{1}{1 + \left(\frac{R\Delta\theta_{ij}}{\mu\kappa_i^+ \kappa_j^-} \right)^\beta}, \quad (2.24a)$$

$$\mathbb{E}[k_i^- | \boldsymbol{\kappa}, \boldsymbol{\theta}] = \sum_{\substack{j=1 \\ j \neq i}}^N \frac{1}{1 + \left(\frac{R\Delta\theta_{ij}}{\mu\kappa_j^+ \kappa_i^-} \right)^\beta}, \quad (2.24b)$$

where $\boldsymbol{\kappa} := (\kappa_1^+, \kappa_1^-, \dots, \kappa_N^+, \kappa_N^-)$ and $\boldsymbol{\theta} := (\theta_1, \dots, \theta_N)$.

2.7.2 Validation of the diMercator inference algorithm

We compare the diMercator algorithm to its counterpart for undirected networks, the Mercator algorithm [12], by giving them both an input of a reciprocal network, which means that all edges are reciprocated as in an undirected network. The procedure for comparing both algorithms is illustrated in Fig. 2.5.

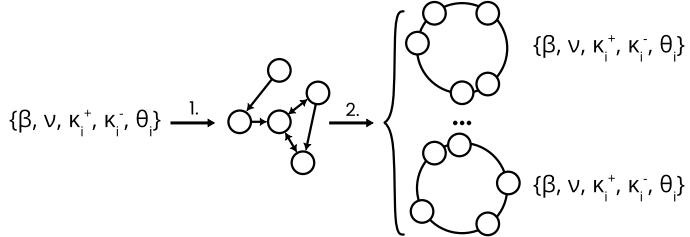


Figure 2.5 – Procedure for comparing the diMercator algorithm to the Mercator algorithm. Step 1: we generate a reciprocal synthetic network from an arbitrary set of embedding parameters. Step 2: we perform a specified number of embeddings of this network with both Mercator and diMercator. Then, we compare the output sets of parameters to the original one.

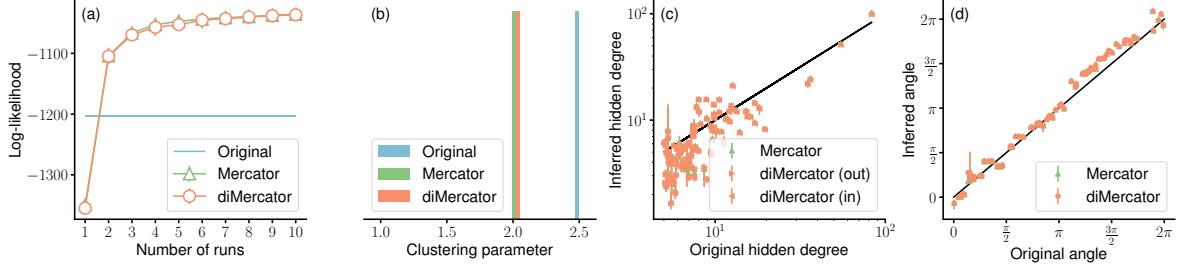


Figure 2.6 – Comparison between the Mercator and the diMercator algorithms on synthetic networks with maximum reciprocity (i.e., all reciprocal edges) which implies $\nu = 1$. (a) Log-likelihood maximization. (b) Clustering parameter β . (c) Hidden degrees. (d) Angular coordinates. Initial clustering parameters are chosen arbitrarily while initial hidden degrees are sampled from a truncated Pareto distribution between 5 and 100 with exponent 2.5 and initial angular coordinates are sampled from a uniform distribution in the interval $[0, 2\pi]$. A total of 100 embeddings were done. Error bars correspond to a 95% confidence interval.

Figure 2.6 compares the Mercator algorithm to the diMercator algorithm on a reciprocal network. Panel (a) shows the evolution of the log-likelihood with respect to the number of refining runs of the algorithm (i.e., the number of times we optimized the values of the embedding parameters). As expected, both Mercator and diMercator yield the same results. The log-likelihood seems to converge to a value higher than the value of the original set of parameters because the original set might not have been the most probable set of parameters for generating the resulting synthetic networks. As for the parameters themselves, panels (b), (c) and (d) show that diMercator yields the same output parameters as Mercator, which is expected if we give the same network to both algorithms as input. For the clustering parameter, the values obtained from diMercator and Mercator appear to diverge from the original value. This discrepancy stems from the same reason discussed above in relation to the log-likelihood. Nonetheless, this analysis confirms that the diMercator algorithm behaves as expected.

2.7.3 Reproducing topological properties of empirical directed networks

In the following subsections, we compute various metrics on the synthetic networks generated by diMercator and other baseline algorithms. For each case, we compare the average metrics’ value across the synthetic networks to the corresponding value measured on the original network, which serves as our ground truth. Additionally, we systematically compare the results obtained using diMercator to those from two alternative models: a version of diMercator in which angular coordinates are assigned uniformly at random in the interval $[0, 2\pi]$, and the directed-reciprocal soft configuration model. Note that all network datasets were downloaded from the Netzschleuder network catalogue and repository (<https://networks.skewed.de>) and are listed in Sec. 2.7.4.

Trophic coherence

The trophic coherence (or incoherence) of a network is related to the notion of trophic levels in ecology [24–28]. However, the standard definition of trophic levels suffers from drawbacks such as requiring basal nodes (i.e., nodes with no incoming edges), which limit its applicability. Fortunately, Ref. [28] presents improved definitions of trophic level and incoherence that overcome these limitations. Hence, we use the latter definitions in our framework. In Ref. [28], we consider directed networks with set V of vertices or nodes and set E of directed edges. We denote the adjacency matrix of a directed network as A with entries $a_{ij} = 1$ if there is an edge from node i to node j and $a_{ij} = 0$ otherwise. For each node j , we define its in-degree and out-degree by

$$k_j^- = \sum_{i \in V} a_{ij} \quad \text{and} \quad k_j^+ = \sum_{i \in V} a_{ji}. \quad (2.25)$$

We define the total degree of the node j by

$$u_j = k_j^- + k_j^+, \quad (2.26)$$

and the imbalance for node j by

$$v_j = k_j^- - k_j^+, \quad (2.27)$$

the latter representing the difference between the flow into and out of the node. We make vectors \mathbf{u} and \mathbf{v} from the u_j and v_j . The graph-Laplacian operator Λ is defined by

$$\Lambda = \text{diag}(\mathbf{u}) - A - A^\top, \quad (2.28)$$

where superscript \top denotes the transpose. When applied to a vector \mathbf{h} , the operator acts as

$$(\Lambda \mathbf{h})_i = u_i h_i - \sum_{j \in V} (a_{ij} + a_{ji}) h_j. \quad (2.29)$$

Then, the trophic levels are given by the solution \mathbf{h} of the linear system of equations

$$\Lambda \mathbf{h} = \mathbf{v}. \quad (2.30)$$

Equations (2.30) always have a solution but it is non-unique because the linear system is undetermined. Thus, to find a solution, one can replace the equation for one node i_S in each weakly connected component $S \subset V$ of the network by an equation $h_{i_S} = c_S$, for arbitrary constants c_S . Then, there is a unique solution for \mathbf{h} . The trophic incoherence parameter is then given by

$$F_0 = \frac{\sum_{ij} a_{ij} (h_j - h_i - 1)^2}{\sum_{ij} a_{ij}}, \quad (2.31)$$

using the trophic levels \mathbf{h} determined above. Note that F_0 is strictly between 0 and 1. We say a network is maximally coherent if it has $F_0 = 0$ and maximally incoherent if it has $F_0 = 1$. The trophic coherence is then defined as $1 - F_0$.

Figure 2.7 shows the trophic incoherence for all datasets.

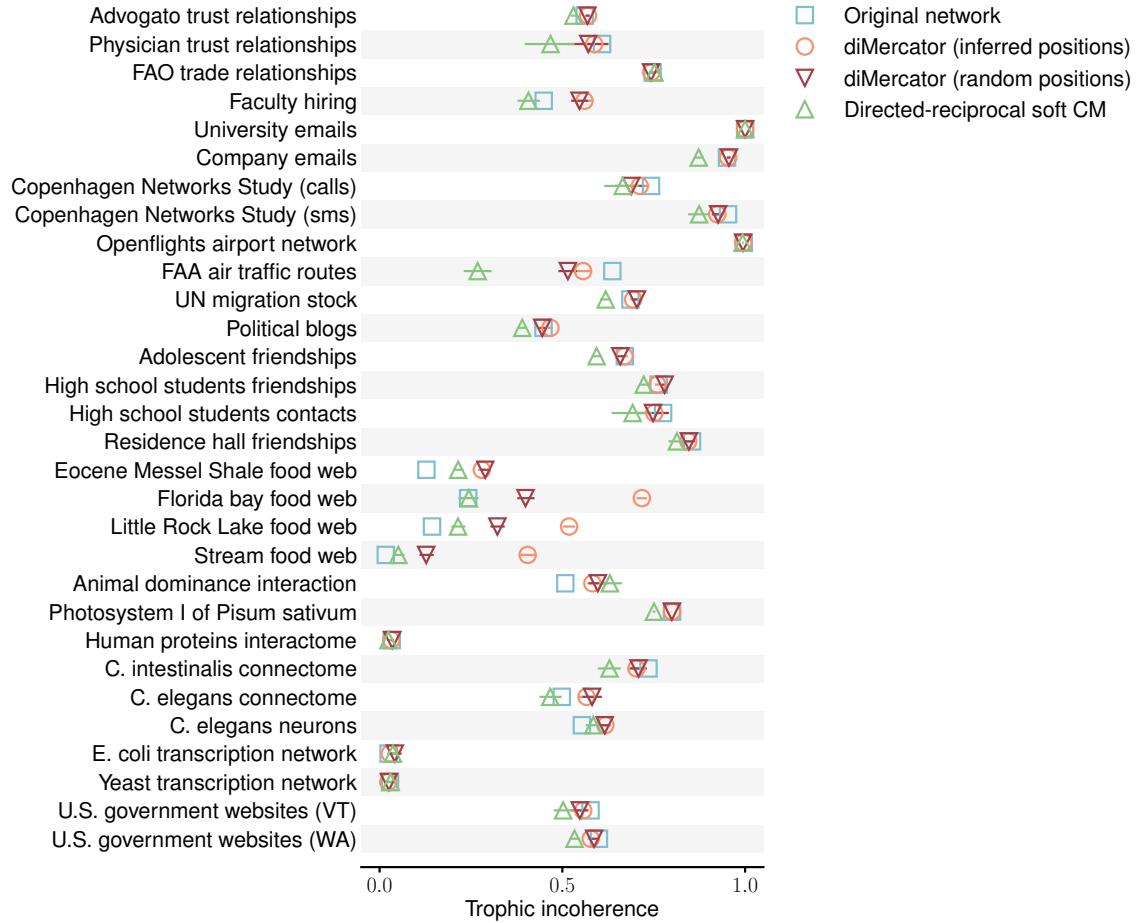


Figure 2.7 – Trophic incoherence. For each dataset, a total of 1000 synthetic networks were generated. Error bars correspond to a 95% confidence interval. See Sec. 2.7.4 for details on the datasets.

Non-normality

We call non-normal a network whose adjacency matrix A is non-normal [29, 30]. A non-normal matrix is a square matrix that does not commute with its conjugate transpose. Hence, in the case of a network's adjacency matrix which is a real matrix, the matrix A is non-normal if

$$AA^\top \neq A^\top A, \quad (2.32)$$

where A^\top is the transpose of A . On the other hand, a matrix that does commute with its conjugate transpose is called a normal matrix. It is thus clear that A needs to be asymmetric to be non-normal, or, equivalently, the network needs to be directed to be non-normal. We consider directed networks with N nodes. Non-normality can hence be quantified using a

standard spectral measure borrowed from matrix theory, such as Henrici's departure from normality [31],

$$d_F(A) = \sqrt{\|A\|_F^2 - \sum_{i=1}^N |\lambda_i|^2}, \quad (2.33)$$

where $\|\cdot\|_F$ is the Frobenius norm and λ_i are the eigenvalues of the adjacency matrix. A zero value is associated with a normal network, while the larger the values, the stronger the non-normality. In our framework, we aim to rescale Eq. (2.33) in the interval $[0, 1]$, hence yielding

$$\tilde{d}_F(A) = \sqrt{1 - \frac{\sum_{i=1}^N |\lambda_i|^2}{\|A\|_F^2}}. \quad (2.34)$$

We will therefore reference Eq. (2.34) as non-normality.

Figure 2.8 shows the non-normality for all datasets.

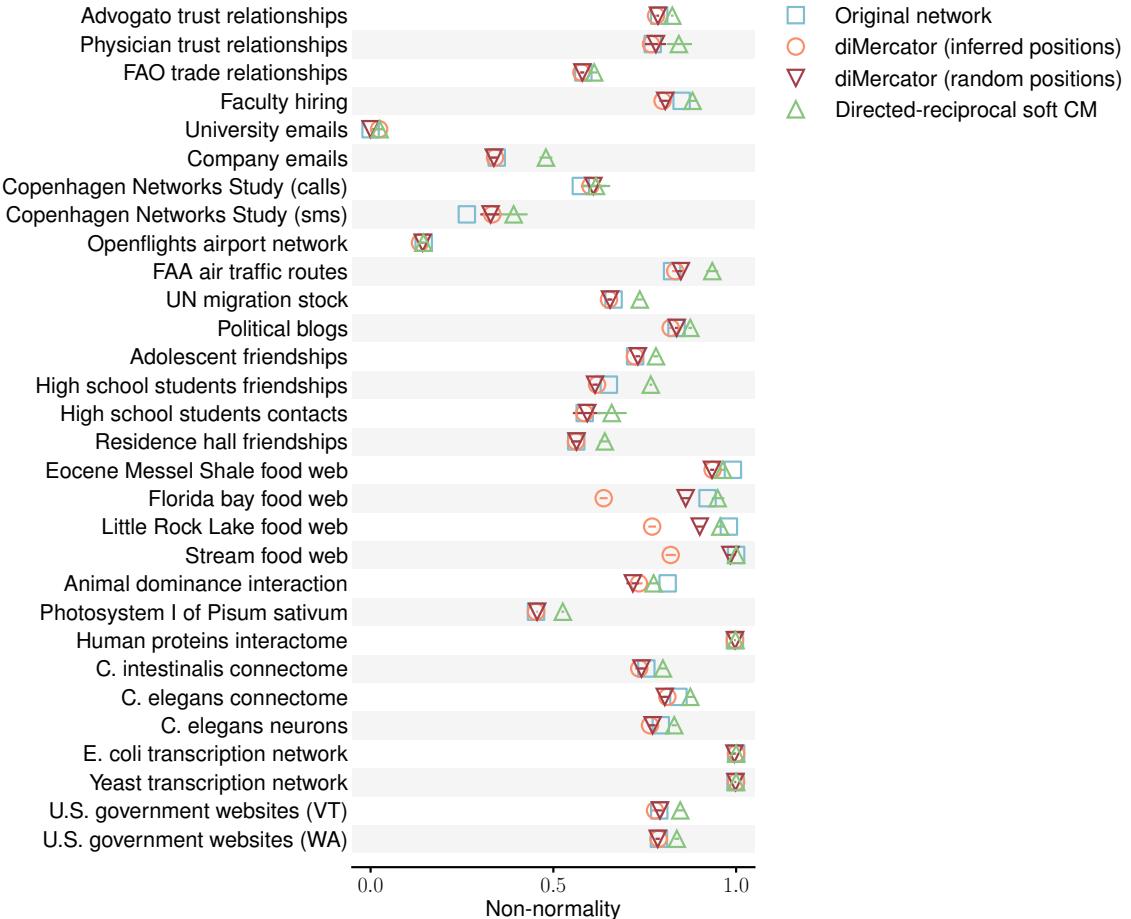


Figure 2.8 – Non-normality. For each dataset, a total of 1000 synthetic networks were generated. Error bars correspond to a 95% confidence interval. See Sec. 2.7.4 for details on the datasets.

Triangle spectrum

Figure 2.9 shows the triangle spectrum for all datasets.

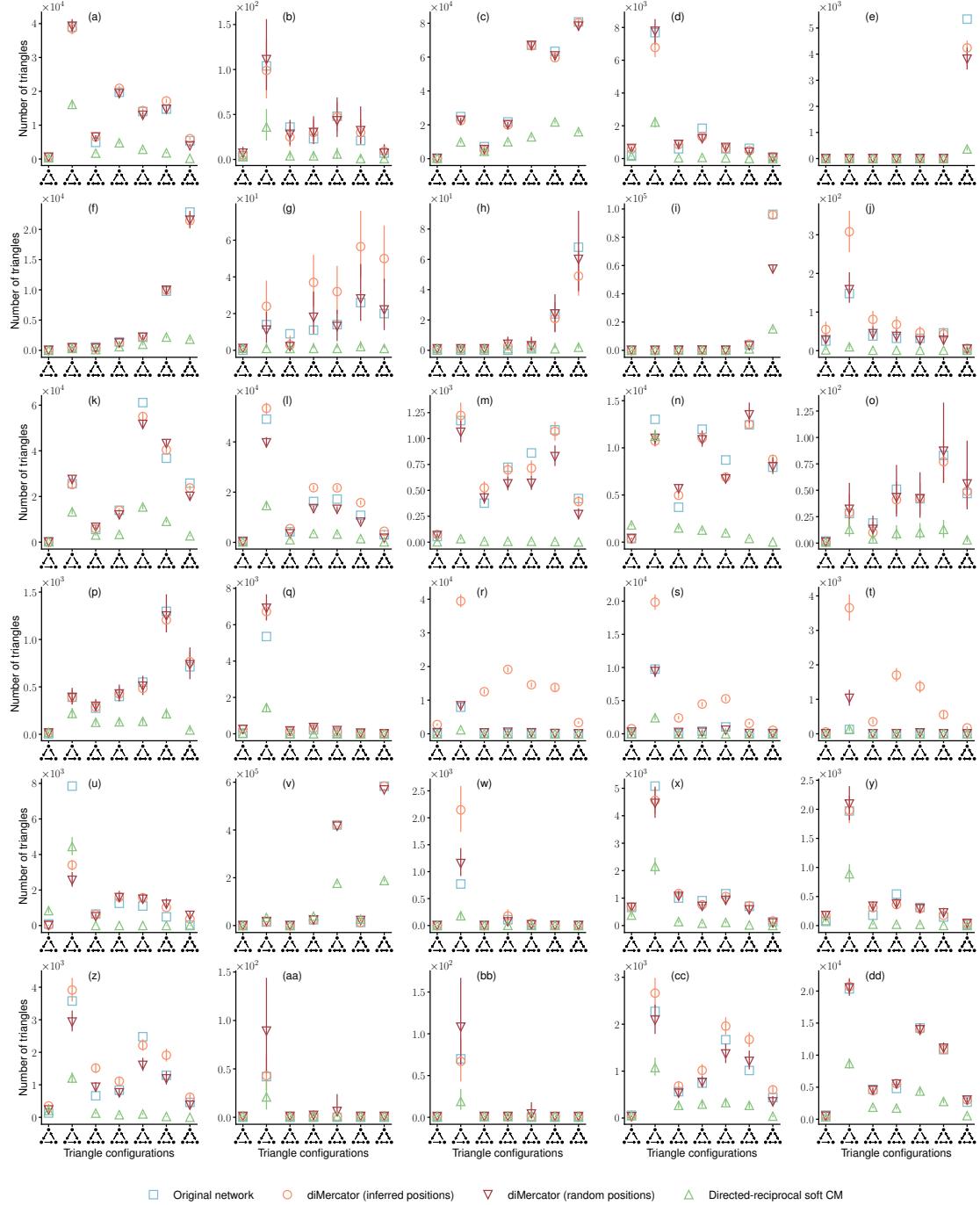


Figure 2.9 – Triangle spectrum. For each dataset, a total of 1000 synthetic networks were generated. Error bars correspond to a 95% confidence interval. See Sec. 2.7.4 for details on the datasets.

Individual reciprocity

Figure 2.10 shows the individual reciprocity (i.e., each node's reciprocity) for all datasets.



Figure 2.10 – Individual reciprocity (each node's reciprocity) distributions. For each dataset, a total of 1000 synthetic networks were generated. See Sec. 2.7.4 for details on the datasets.

Singular value spectrum

One can study the properties of a network's adjacency matrix A by performing a singular value decomposition (also known as SVD) of this matrix. We thus consider a real matrix A of size $N \times N$, where N corresponds to the size of the network. We solve the two eigenvalue problems

$$AA^\top \mathbf{u}_i = \lambda_i \mathbf{u}_i \quad \text{and} \quad A^\top A \mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i \in \{1, \dots, N\}, \quad (2.35)$$

where \top denotes the transpose. Note that the matrices AA^\top and $A^\top A$ are square, symmetric, positive semi-definite, and share the same non-zero eigenvalues. The result of their diagonalization allows us to write the singular value decomposition of A as

$$A = U\Sigma V^\top, \quad (2.36)$$

where U is an orthogonal square matrix whose columns correspond to the eigenvectors of AA^\top , V is an orthogonal square matrix whose columns correspond to the eigenvectors of $A^\top A$, and Σ is a diagonal matrix. The singular values of A are denoted as the elements $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, 2, \dots, N$ of the matrix Σ . The singular values of a matrix are always real and non-negative; they are typically ordered in decreasing magnitude as $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N$ [22].

The rank of A , denoted by $\text{rank}(A)$, is the dimension of the vector space spanned by its rows (row rank) or its columns (column rank),

$$\text{rank}(A) = \dim(\text{Col}(A)) = \dim(\text{Row}(A)). \quad (2.37)$$

It is a fundamental theorem of linear algebra that the row rank and the column rank of any matrix are always equal [69]. In terms of SVD, the rank of A defined by Eq. (2.36) corresponds to the number of strictly positive singular values of A . The stable rank of a matrix A , denoted $\text{srank}(A)$, is defined as

$$\text{srank}(A) = \sum_{i=1}^N \sigma_i^2 / \sigma_1^2, \quad (2.38)$$

where σ_1 is the largest singular value of A . The stable rank cannot exceed the usual rank,

$$\text{srank}(A) \leq \text{rank}(A). \quad (2.39)$$

Figure 2.11 shows the singular value spectrum as well as the stable rank for all datasets.

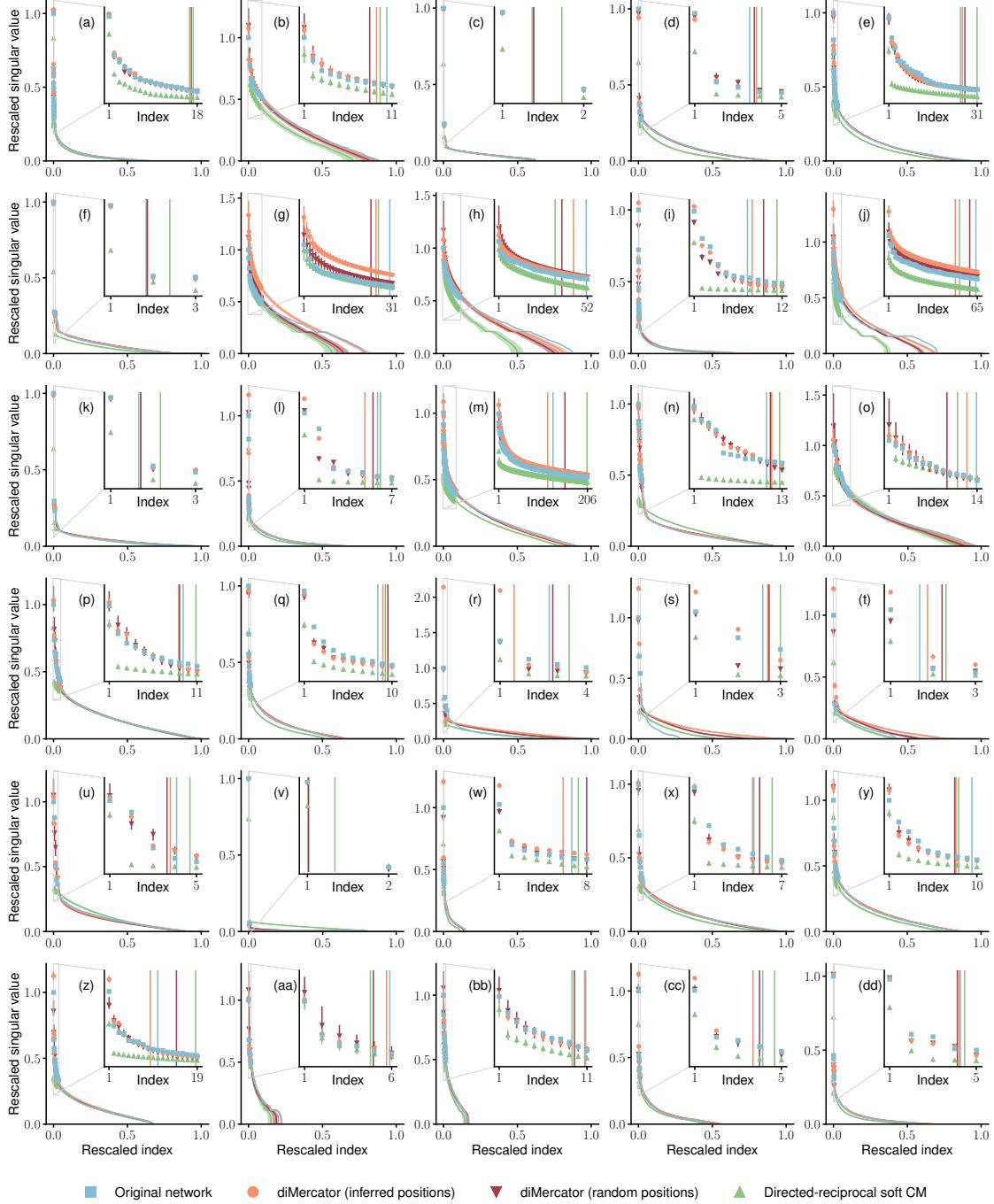


Figure 2.11 – Singular value spectrum from singular value decomposition (SVD). To improve visual clarity, only the first values are shown as individual points; the later values are plotted as a continuous line. In the inset, the vertical bars denote the stable rank. For each dataset, a total of 1000 synthetic networks were generated. Error bars correspond to a 95% confidence interval. See Sec. 2.7.4 for details on the datasets.

Bow-tie structure

It is possible to decompose a directed network into its different components, as shown in Figure 2.12.

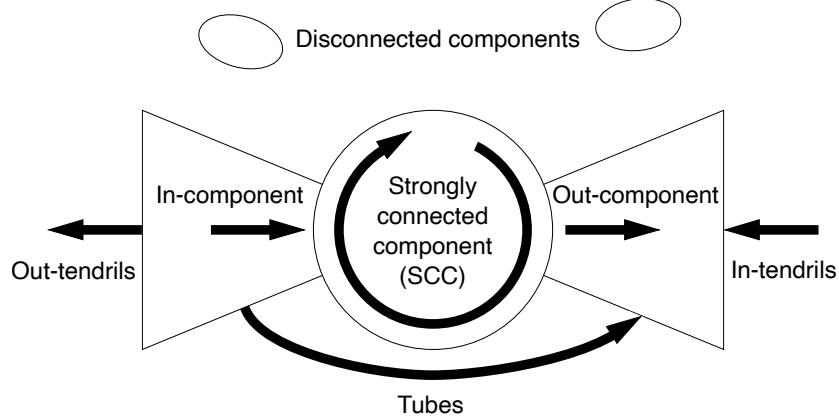


Figure 2.12 – Bow-tie structure of a directed network.

This decomposition is called the bow-tie structure [20, 21]. The strongly connected component (SCC) corresponds to the set of nodes such that each can reach all others and be reached from all others within the set via a directed path. Each strongly connected component is associated with an out-component (the set of all nodes that can be reached via a directed path from any node in the strongly connected component) and an in-component (the set of all nodes from which the strongly connected component can be reached). A typical directed network consists of a large strongly connected component and several smaller components, each with its own in-component and out-component. Note that each in- or out-component excludes the strongly connected component. Out-tendrils and in-tendrils contain nodes that are reachable from portions of the in-component, or that can reach portions of the out-component, without passage through the SCC. It is possible for an out-tendril to be hooked into an in-tendril, forming a tube.

We use a confusion matrix [65] to evaluate the performance of our algorithm in correctly associating nodes with the appropriate components of the bow-tie structure. In our framework, each row of the confusion matrix corresponds to an actual class (i.e., components of the original network), while each column corresponds to a predicted class (i.e., components of the synthetic networks).

Figure 2.13 shows the bow-tie structure as confusion matrices for all datasets.

(a) Advogato trust relationships

		diMercator (inferred positions)							
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	In	SCC	54.0	3.2	3.9	< 0.1	0.2	< 0.1	0.9
		In	0.4	10.7	< 0.1	< 0.1	0.4	< 0.1	1.2
Original network	Out	SCC	0.7	< 0.1	19.6	< 0.1	< 0.1	0.3	1.8
		Out	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
Original network	In-T.	SCC	< 0.1	0.8	< 0.1	< 0.1	0.6	< 0.1	0.2
		In-T.	< 0.1	0.3	< 0.1	< 0.1	0.2	< 0.1	0.1
Original network	Out-T.	SCC	0.0	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1
		Out-T.	< 0.1	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1
Original network	Disc.	SCC	0.0	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1
		Disc.	< 0.1	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1

diMercator (random positions)

		diMercator (random positions)							
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	In	SCC	53.8	3.1	3.8	< 0.1	0.2	0.1	1.2
		In	0.4	9.6	< 0.1	< 0.1	0.3	< 0.1	2.4
Original network	Out	SCC	0.8	< 0.1	16.3	< 0.1	< 0.1	0.4	4.9
		Out	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
Original network	In-T.	SCC	< 0.1	1.0	< 0.1	< 0.1	< 0.1	< 0.1	0.5
		In-T.	< 0.1	0.4	< 0.1	< 0.1	< 0.1	< 0.1	0.2
Original network	Out-T.	SCC	0.0	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1
		Out-T.	< 0.1	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1
Original network	Disc.	SCC	0.0	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1
		Disc.	< 0.1	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1

Directed-reciprocal soft CM

		Directed-reciprocal soft CM							
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	In	SCC	47.1	5.3	6.1	< 0.1	0.4	0.2	3.2
		In	0.2	7.9	< 0.1	< 0.1	0.4	< 0.1	4.1
Original network	Out	SCC	0.5	0.1	13.1	< 0.1	< 0.1	0.5	8.3
		Out	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
Original network	In-T.	SCC	< 0.1	0.7	< 0.1	< 0.1	< 0.1	< 0.1	0.8
		In-T.	< 0.1	0.3	< 0.1	< 0.1	< 0.1	< 0.1	0.3
Original network	Out-T.	SCC	0.0	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1
		Out-T.	< 0.1	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1
Original network	Disc.	SCC	0.0	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1
		Disc.	< 0.1	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1

(b) Physician trust relationships

		diMercator (inferred positions)							
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	In	SCC	73.8	6.3	0.9	< 0.1	< 0.1	< 0.1	< 0.1
		In	1.1	9.7	< 0.1	< 0.1	0.2	< 0.1	< 0.1
Original network	Out	SCC	0.0	0.0	7.4	0.0	0.0	0.2	< 0.1
		Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	SCC	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	SCC	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	SCC	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Disc.	< 0.1	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1

		diMercator (random positions)							
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	In	SCC	69.9	9.5	1.5	< 0.1	< 0.1	< 0.1	0.2
		In	1.1	9.5	< 0.1	< 0.1	< 0.1	< 0.1	0.3
Original network	Out	SCC	0.0	0.0	6.7	0.0	0.0	0.4	0.6
		Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	SCC	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	SCC	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	SCC	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Disc.	< 0.1	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1

		Directed-reciprocal soft CM							
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	In	SCC	61.8	< 0.1	0.4	0.0	0.0	< 0.1	< 0.1
		In	0.0	34.2	0.0	0.0	< 0.1	0.0	3.7
Original network	Out	SCC	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	SCC	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	SCC	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	SCC	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Disc.	< 0.1	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1

(d) Faculty hiring

		diMercator (inferred positions)							
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	In	SCC	76.9	< 0.1	4.1	0.0	< 0.1	< 0.1	< 0.1
		In	0.3	0.2	0.0	0.0	0.0	0.0	0.0
Original network	Out	SCC	9.7	< 0.1	8.7	< 0.1	< 0.1	< 0.1	< 0.1
		Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	SCC	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	SCC	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	SCC	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Disc.	< 0.1	< 0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1

		diMercator (random positions)							
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	In	SCC	76.0	< 0.1	5.0	0.0	0.0	< 0.1	< 0.1
		In	0.3	0.1	0.0	0.0	0.0	0.0	0.0
Original network	Out	SCC	8.7	< 0.1	9.7	< 0.1	< 0.1	< 0.1	< 0.1
		Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network									

(e) University emails

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	95.7	< 0.1	< 0.1	0.0	0.0	< 0.1	4.3
	In	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

diMercator (random positions)

		diMercator (random positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	93.0	< 0.1	< 0.1	0.0	0.0	0.0	7.0
	In	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Directed-reciprocal soft CM

		Directed-reciprocal soft CM						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	81.8	< 0.1	< 0.1	0.0	< 0.1	0.0	18.2
	In	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(f) Company emails

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	74.8	< 0.1	0.7	0.0	0.0	0.0	< 0.1
	In	0.4	14.3	0.2	0.0	< 0.1	< 0.1	1.2
Original network	Out	6.6	< 0.1	1.1	0.0	0.0	0.0	< 0.1
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.3	0.0	0.0	< 0.1	0.0	0.3
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

		diMercator (random positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	75.0	0.0	0.5	0.0	0.0	0.0	< 0.1
	In	0.4	10.5	0.2	0.0	< 0.1	0.0	5.1
Original network	Out	7.1	< 0.1	0.6	0.0	0.0	< 0.1	< 0.1
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.4	0.0	0.0	0.0	0.0	0.2
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

		Directed-reciprocal soft CM						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	72.9	< 0.1	2.0	0.0	< 0.1	0.0	0.5
	In	< 0.1	2.1	0.2	0.0	< 0.1	0.0	13.8
Original network	Out	2.0	0.1	3.8	0.0	< 0.1	0.0	1.9
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	< 0.1	0.0	0.0	< 0.1	0.0	0.5
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(g) Copenhagen Networks Study (calls)

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	41.2	1.6	1.4	< 0.1	0.3	< 0.1	0.9
	In	8.6	4.4	0.7	< 0.1	0.3	0.2	0.8
Original network	Out	8.5	0.6	10.7	< 0.1	0.5	0.6	0.9
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	1.5	1.6	< 0.1	< 0.1	0.4	< 0.1	0.1
	Out-T.	3.2	0.3	2.0	< 0.1	0.1	0.3	0.4
Original network	Disc.	3.0	1.7	1.5	< 0.1	0.5	0.4	0.4
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

		diMercator (random positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	69.2	2.2	2.6	< 0.1	0.3	0.2	13.8
	In	2.3	1.3	< 0.1	< 0.1	< 0.1	< 0.1	1.5
Original network	Out	2.3	< 0.1	1.1	0.0	< 0.1	< 0.1	1.3
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	1.1	< 0.1	0.2	< 0.1	< 0.1	< 0.1	0.2
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

		Directed-reciprocal soft CM						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	42.4	4.1	4.7	< 0.1	0.6	0.6	36.1
	In	1.1	0.9	< 0.1	< 0.1	< 0.1	< 0.1	3.1
Original network	Out	1.0	< 0.1	0.8	< 0.1	< 0.1	< 0.1	2.8
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.7	< 0.1	0.1	< 0.1	< 0.1	< 0.1	0.6
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(i) Openflights airport network

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	90.3	0.5	0.6	< 0.1	< 0.1	< 0.1	7.3
In	In	< 0.1	0.3	< 0.1	0.0	< 0.1	< 0.1	0.3
Original network	In	< 0.1	< 0.1	0.3	0.0	< 0.1	< 0.1	0.2
Original network	Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

diMercator (random positions)

		diMercator (random positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	85.9	0.5	0.5	0.0	< 0.1	< 0.1	11.8
In	In	< 0.1	0.4	< 0.1	0.0	< 0.1	< 0.1	0.3
Original network	In	< 0.1	< 0.1	0.3	0.0	0.0	< 0.1	0.2
Original network	Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Directed-reciprocal soft CM

		Directed-reciprocal soft CM						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	76.2	0.6	0.8	0.0	< 0.1	< 0.1	21.1
In	In	< 0.1	0.2	< 0.1	0.0	< 0.1	0.0	0.4
Original network	In	< 0.1	< 0.1	0.2	0.0	< 0.1	< 0.1	0.3
Original network	Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(j) FAA air traffic routes

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	42.8	7.5	9.3	0.2	1.1	0.6	3.1
In	In	1.1	1.4	0.7	< 0.1	0.2	0.1	0.6
Original network	In	8.2	3.2	14.4	0.2	1.0	0.8	2.8
Original network	Out	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
Original network	Disc.	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	0.2
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

diMercator (random positions)

		diMercator (random positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	36.1	9.1	11.5	0.2	1.6	0.7	5.4
In	In	1.2	1.3	0.7	< 0.1	0.2	< 0.1	0.7
Original network	In	7.2	3.3	13.1	< 0.1	0.7	0.8	5.3
Original network	Out	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
Original network	Disc.	< 0.1	0.2	< 0.1	< 0.1	< 0.1	< 0.1	0.1
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Directed-reciprocal soft CM

		Directed-reciprocal soft CM						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	10.7	9.9	13.0	0.7	4.3	2.8	23.2
In	In	0.2	0.7	0.5	< 0.1	0.4	0.1	2.3
Original network	In	0.9	2.3	7.6	0.2	1.3	2.0	16.3
Original network	Out	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
Original network	Disc.	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	0.3
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(k) UN migration stock

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	98.8	0.7	< 0.1	0.0	0.0	0.0	< 0.1
In	In	0.0	0.4	0.0	0.0	0.0	0.0	0.0
Original network	In	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

diMercator (random positions)

		diMercator (random positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	99.2	0.4	< 0.1	0.0	0.0	0.0	< 0.1
In	In	0.0	0.4	0.0	0.0	0.0	0.0	< 0.1
Original network	In	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Directed-reciprocal soft CM

		Directed-reciprocal soft CM						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	95.9	3.2	0.4	0.0	0.0	< 0.1	0.1
In	In	0.0	0.4	0.0	0.0	0.0	0.0	< 0.1
Original network	In	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(l) Political blogs

|
<th colspan="
| |

(m) Adolescent friendships

		diMercator (inferred positions)						diMercator (random positions)						Directed-reciprocal soft CM								
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	76.1	5.1	2.7	< 0.1	0.2	0.2	0.6	76.0	5.3	2.7	< 0.1	0.1	0.1	0.6	66.2	10.4	5.5	< 0.1	0.4	0.3	1.9
	In	1.0	4.1	< 0.1	< 0.1	0.1	< 0.1	0.3	1.1	3.9	< 0.1	< 0.1	< 0.1	< 0.1	0.4	0.8	3.8	< 0.1	< 0.1	0.1	< 0.1	0.8
Original network	Out	0.2	< 0.1	8.1	< 0.1	< 0.1	0.3	0.4	0.2	< 0.1	7.8	< 0.1	< 0.1	0.1	0.9	0.2	< 0.1	6.9	< 0.1	< 0.1	0.3	1.7
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Original network	In-T.	0.0	< 0.1	0.0	0.0	< 0.1	0.0	< 0.1	0.0	0.0	0.1	0.0	0.0	< 0.1	0.1	0.0	< 0.1	0.0	< 0.1	0.0	< 0.1	0.2
	Out-T.	< 0.1	< 0.1	0.2	< 0.1	< 0.1	0.1	< 0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	< 0.1	< 0.1	0.2	< 0.1	< 0.1	0.1	< 0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(n) High school students friendships

		diMercator (inferred positions)						diMercator (random positions)						Directed-reciprocal soft CM								
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	90.8	0.0	0.2	0.0	0.0	0.0	0.0	90.7	0.0	0.3	0.0	0.0	0.0	0.0	90.6	< 0.1	0.4	0.0	0.0	0.0	0.0
	In	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Original network	Out	0.1	0.0	8.9	0.0	0.0	0.0	0.0	0.1	0.0	8.9	0.0	0.0	0.0	0.0	0.0	0.2	0.0	8.8	0.0	0.0	0.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	< 0.1	< 0.1	0.2	0.0	< 0.1	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(o) High school students contacts

		diMercator (inferred positions)						diMercator (random positions)						Directed-reciprocal soft CM								
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	64.8	4.2	3.6	< 0.1	0.1	0.2	0.5	59.4	5.0	5.8	0.1	0.5	0.5	2.0	59.7	4.9	5.4	< 0.1	0.2	0.2	3.0
	In	18.3	4.2	1.3	< 0.1	0.1	0.1	0.1	18.3	2.8	2.0	< 0.1	0.2	0.1	0.6	19.2	2.4	1.3	< 0.1	< 0.1	< 0.1	1.1
Original network	Out	< 0.1	< 0.1	1.2	0.0	< 0.1	0.2	0.2	0.9	< 0.1	0.5	< 0.1	< 0.1	< 0.1	0.2	0.7	< 0.1	0.6	< 0.1	< 0.1	< 0.1	0.3
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	Out-T.	0.0	0.0	0.7	0.0	0.0	< 0.1	< 0.1	0.0	0.0	0.0	0.0	0.0	< 0.1	< 0.1	0.0	0.0	0.0	0.0	0.0	0.0	
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	Out-T.	< 0.1	< 0.1	0.2	0.0	< 0.1	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

(p) Residence hall friendships

		diMercator (inferred positions)						diMercator (random positions)						Directed-reciprocal soft CM								
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	98.2	0.4	< 0.1	0.0	0.0	0.0	< 0.1	97.6	0.8	0.2	0.0	< 0.1	< 0.1	< 0.1	95.6	1.6	1.2	< 0.1	< 0.1	< 0.1	0.2
	In	0.0	0.9	0.0	0.0	0.0	0.0	< 0.1	0.0	0.9	0.0	0.0	0.0	0.0	< 0.1	0.0	0.6	0.0	< 0.1	0.0	0.3	
Original network	Out	0.0	0.0	0.5	0.0	0.0	< 0.1	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	Out-T.	< 0.1	< 0.1	0.2	0.0	< 0.1	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

(q) Eocene Messel Shale food web

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	10.9	0.8	0.2	< 0.1	< 0.1	< 0.1	< 0.1
	In	3.1	5.3	< 0.1	< 0.1	< 0.1	0.0	< 0.1
Original network	Out	35.0	1.2	40.3	< 0.1	0.1	0.2	0.2
	Tubes	0.7	0.3	0.3	< 0.1	< 0.1	< 0.1	< 0.1
Original network	In-T.	< 0.1	1.1	0.0	0.0	< 0.1	0.0	0.0
	Out-T.	0.0	0.0	< 0.1	0.0	0.0	< 0.1	< 0.1
Original network	Disc.	0.0	0.0	< 0.1	0.0	0.0	< 0.1	< 0.1
	Tube	0.0	0.0	< 0.1	0.0	0.0	< 0.1	< 0.1

(r) Florida bay food web

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	79.8	0.3	0.4	0.0	0.0	0.0	0.0
	In	14.2	3.1	0.5	< 0.1	< 0.1	< 0.1	< 0.1
Original network	Out	1.5	0.0	< 0.1	0.0	0.0	0.0	0.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Tube	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(s) Little Rock Lake food web

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	12.0	0.0	< 0.1	0.0	0.0	0.0	0.0
	In	82.8	4.4	0.2	0.0	0.0	< 0.1	< 0.1
Original network	Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	< 0.1	< 0.1	0.3	0.0	0.0	< 0.1	0.1
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Tube	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(t) Stream food web

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	0.5	0.0	0.4	0.0	0.0	0.0	0.0
	In	22.0	9.7	0.1	< 0.1	0.3	< 0.1	< 0.1
Original network	Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	34.5	2.4	4.3	< 0.1	< 0.1	0.5	< 0.1
Original network	Disc.	7.6	12.7	1.8	0.2	1.2	0.2	1.4
	Tube	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(u) Animal dominance interaction

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	87.6	0.9	0.9	0.0	0.0	0.0	0.0
	In	0.0	1.9	0.0	0.0	0.0	0.0	0.1
Original network	Out	0.0	0.0	8.4	0.0	0.0	0.0	0.2
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

diMercator (random positions)

		diMercator (random positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	84.6	1.4	3.1	< 0.1	< 0.1	< 0.1	0.3
	In	0.0	1.7	0.0	0.0	< 0.1	0.0	0.2
Original network	Out	0.0	0.0	7.6	0.0	0.0	< 0.1	0.9
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Directed-reciprocal soft CM

		Directed-reciprocal soft CM						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	82.9	1.4	4.7	0.0	< 0.1	< 0.1	0.4
	In	0.0	1.7	0.0	0.0	< 0.1	0.0	0.3
Original network	Out	0.0	0.0	7.6	0.0	0.0	< 0.1	1.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(v) Photosystem I of Pisum sativum

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	80.2	0.0	0.0	0.0	0.0	0.0	0.0
	In	0.0	18.8	0.0	0.0	0.0	0.0	0.0
Original network	Out	0.0	0.0	1.0	0.0	0.0	0.0	0.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

		diMercator (random positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	80.2	0.0	0.0	0.0	0.0	0.0	0.0
	In	0.0	18.8	0.0	0.0	0.0	0.0	0.0
Original network	Out	0.0	0.0	1.0	0.0	0.0	0.0	0.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

		Directed-reciprocal soft CM						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	80.2	0.0	< 0.1	0.0	0.0	0.0	0.0
	In	0.0	18.8	0.0	0.0	0.0	0.0	0.0
Original network	Out	0.0	0.0	1.0	0.0	0.0	0.0	0.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(w) Human proteins interactome

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	0.3	< 0.1	< 0.1	< 0.1	< 0.1	0.0	0.0
	In	< 0.1	1.0	< 0.1	< 0.1	0.5	< 0.1	< 0.1
Original network	Out	0.1	< 0.1	5.7	0.2	< 0.1	5.9	0.3
	Tubes	0.1	0.3	0.3	0.7	0.3	0.3	0.1
Original network	In-T.	< 0.1	1.4	< 0.1	0.3	3.9	< 0.1	1.4
	Out-T.	< 0.1	< 0.1	9.1	0.2	< 0.1	41.0	5.2
Original network	Disc.	< 0.1	0.1	1.6	< 0.1	0.9	11.6	6.3
	Out-T.	< 0.1	0.1	1.6	< 0.1	0.9	11.6	6.3
Original network	Disc.	< 0.1	0.1	1.6	< 0.1	0.9	11.6	6.3

		diMercator (random positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	92.9	2.8	2.2	0.0	< 0.1	< 0.1	0.1
	In	0.0	0.5	0.0	0.0	0.0	0.0	0.0
Original network	Out	< 0.1	< 0.1	1.2	0.0	0.0	< 0.1	0.2
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

		Directed-reciprocal soft CM						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	79.7	9.2	7.2	< 0.1	< 0.1	0.2	1.7
	In	0.0	0.5	0.0	0.0	< 0.1	0.0	< 0.1
Original network	Out	0.0	0.0	0.8	0.0	0.0	< 0.1	0.6
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Original network	Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(x) C. intestinalis connectome

|--|

(y) *C. elegans* connectome

		diMercator (inferred positions)						
	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	SCC	75.9	2.7	1.7	< 0.1	< 0.1	< 0.1	< 0.1
	In	0.3	4.9	< 0.1	< 0.1	< 0.1	< 0.1	0.1
	Out	4.1	0.1	4.6	< 0.1	0.2	< 0.1	< 0.1
	Tubes	< 0.1	0.3	< 0.1	< 0.1	< 0.1	< 0.1	0.0
	In-T.	0.0	0.2	0.0	0.0	4.2	0.0	0.3
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

		diMercator (random positions)						
	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	SCC	74.2	3.9	2.2	< 0.1	< 0.1	< 0.1	0.1
	In	0.6	4.5	< 0.1	0.0	< 0.1	0.0	0.2
	Out	4.6	0.2	4.1	< 0.1	< 0.1	< 0.1	< 0.1
	Tubes	0.2	0.1	< 0.1	0.0	< 0.1	< 0.1	< 0.1
	In-T.	0.0	2.6	0.0	0.0	0.4	0.0	1.7
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

		Directed-reciprocal soft CM						
	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	SCC	60.2	10.2	8.7	< 0.1	0.4	0.1	0.7
	In	0.2	4.2	< 0.1	< 0.1	0.3	< 0.1	0.6
	Out	2.6	0.3	5.9	< 0.1	< 0.1	< 0.1	0.3
	Tubes	< 0.1	0.2	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
	In-T.	0.0	0.3	0.0	0.0	0.1	0.0	4.3
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

(z) *C. elegans* neurons

		diMercator (inferred positions)						
	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	SCC	58.7	0.5	1.3	< 0.1	< 0.1	< 0.1	< 0.1
	In	1.3	27.7	< 0.1	0.0	< 0.1	< 0.1	0.3
	Out	0.0	0.0	1.1	0.0	0.0	< 0.1	< 0.1
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	0.1	0.0	0.0	0.4	0.0	< 0.1
	Out-T.	3.2	0.2	0.2	< 0.1	0.0	0.1	< 0.1
Disc.	0.0	4.0	0.0	0.0	0.5	0.0	0.2	

		diMercator (random positions)						
	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	SCC	58.5	0.5	1.4	< 0.1	< 0.1	< 0.1	0.1
	In	1.4	26.3	< 0.1	< 0.1	< 0.1	< 0.1	1.5
	Out	0.0	0.0	1.1	0.0	0.0	< 0.1	< 0.1
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	0.3	0.0	0.0	< 0.1	0.0	0.2
	Out-T.	3.5	< 0.1	0.1	< 0.1	0.0	< 0.1	< 0.1
Disc.	0.0	3.8	0.0	0.0	< 0.1	0.0	0.8	

		Directed-reciprocal soft CM						
	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	SCC	56.7	1.1	2.6	< 0.1	< 0.1	< 0.1	0.2
	In	1.2	25.6	< 0.1	< 0.1	< 0.1	< 0.1	2.4
	Out	0.0	0.0	1.0	0.0	0.0	< 0.1	< 0.1
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	0.3	0.0	0.0	< 0.1	0.0	0.3
	Out-T.	3.3	0.1	0.2	0.0	< 0.1	< 0.1	< 0.1
Disc.	0.0	3.4	0.0	0.0	< 0.1	0.0	1.2	

(aa) *E. coli* transcription network

		diMercator (inferred positions)						
	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	SCC	0.2	0.0	0.0	0.0	0.0	< 0.1	0.1
	In	0.0	0.4	0.0	0.0	< 0.1	0.0	0.2
	Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Out-T.	0.0	0.0	< 0.1	0.0	0.0	0.2	< 0.1
Disc.	0.2	2.2	0.2	< 0.1	2.5	0.8	92.8	

		diMercator (random positions)						
	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	SCC	< 0.1	0.0	< 0.1	0.0	0.0	< 0.1	0.3
	In	0.0	< 0.1	0.0	0.0	< 0.1	0.0	0.5
	Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	0.0	< 0.1	0.0	0.0	< 0.1	0.2
	Out-T.	0.0	0.0	< 0.1	0.0	0.0	< 0.1	0.2
Disc.	0.7	3.3	0.7	0.2	12.0	1.4	80.6	

		Directed-reciprocal soft CM						
	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	SCC	0.3	0.0	0.0	0.0	0.0	< 0.1	< 0.1
	In	0.0	< 0.1	0.0	0.0	< 0.1	0.0	0.6
	Out	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	0.0	< 0.1	0.0	0.0	< 0.1	0.3
	Out-T.	< 0.1	1.0	0.2	< 0.1	2.3	1.3	93.8

(bb) Yeast transcription network

		diMercator (inferred positions)						
	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	SCC	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	0.3
	In	0.0	0.3	0.0	0.0	0.4	0.0	2.1
	Out	< 0.1	0.0	< 0.1	0.0	0.0	< 0.1	< 0.1
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	< 0.1	0.0	0.0	0.2	0.0	1.0
	Out-T.	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	0.2	0.6
Disc.	0.2	1.5	0.2	< 0.1	3.9	2.1	86.5	

		diMercator (random positions)						
	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	SCC	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	0.3
	In	0.0	0.2	0.0	0.0	0.4	0.0	2.6
	Out	< 0.1	0.0	< 0.1	0.0	0.0	< 0.1	< 0.1
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	< 0.1	0.0	0.0	0.1	0.0	1.1
	Out-T.	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	0.2	0.6
Disc.	0.3	3.0	0.4	0.1	7.8	3.1	79.8	

		Directed-reciprocal soft CM						
	SCC	In	Out	Tubes	In-T.	Out-T.	Disc.	
Original network	SCC	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	0.3
	In	0.0	0.2	0.0	0.0	< 0.1	0.0	2.6
	Out	< 0.1	0.0	< 0.1	0.0	0.0	< 0.1	< 0.1
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	< 0.1	0.0	0.0	< 0.1	0.0	1.2
	Out-T.	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	0.2	0.7
Disc.	0.2	2.7	< 0.1	< 0.1	1.1	4.1	86.2	

(cc) U.S. government websites (VT)

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	49.7	4.4	3.2	< 0.1	0.2	0.2	0.8
	In	0.1	13.6	< 0.1	0.0	0.3	< 0.1	0.3
	Out	1.7	< 0.1	24.4	< 0.1	< 0.1	0.3	0.1
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	0.1	0.0	0.0	< 0.1	0.0	< 0.1
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

		diMercator (random positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	47.5	5.0	3.8	< 0.1	0.1	0.4	1.6
	In	0.1	12.6	< 0.1	< 0.1	0.1	< 0.1	1.6
	Out	1.7	0.1	17.5	< 0.1	< 0.1	1.3	6.2
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	0.1	0.0	0.0	< 0.1	0.0	< 0.1
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

		Directed-reciprocal soft CM						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	39.7	8.0	6.3	< 0.1	0.2	0.6	3.7
	In	< 0.1	11.2	< 0.1	< 0.1	0.2	< 0.1	2.9
	Out	1.5	0.2	13.6	< 0.1	< 0.1	1.3	10.2
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	< 0.1	0.0	0.0	< 0.1	0.0	0.1
	Out-T.	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

(dd) U.S. government websites (WA)

		diMercator (inferred positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	67.0	5.0	4.4	< 0.1	0.1	0.2	0.5
	In	0.6	14.1	< 0.1	0.0	< 0.1	< 0.1	0.3
	Out	< 0.1	< 0.1	6.9	< 0.1	< 0.1	0.2	0.3
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	< 0.1	0.0	0.0	< 0.1	0.0	< 0.1
	Out-T.	0.0	0.0	< 0.1	0.0	0.0	< 0.1	< 0.1
Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

		diMercator (random positions)						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	66.6	5.6	3.9	< 0.1	< 0.1	0.2	1.0
	In	0.5	12.9	< 0.1	0.0	< 0.1	< 0.1	1.5
	Out	< 0.1	< 0.1	5.9	0.0	< 0.1	0.2	1.2
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	< 0.1	0.0	0.0	< 0.1	0.0	< 0.1
	Out-T.	0.0	0.0	< 0.1	0.0	0.0	< 0.1	< 0.1
Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

		Directed-reciprocal soft CM						
		SCC	In	Out	Tubes	In-T.	Out-T.	Disc.
Original network	SCC	50.2	12.5	8.0	< 0.1	0.1	0.9	5.6
	In	0.3	10.0	< 0.1	< 0.1	0.1	< 0.1	4.5
	Out	< 0.1	< 0.1	4.0	0.0	< 0.1	0.4	3.0
	Tubes	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	In-T.	0.0	< 0.1	0.0	0.0	< 0.1	0.0	< 0.1
	Out-T.	0.0	0.0	< 0.1	0.0	0.0	< 0.1	< 0.1
Disc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Figure 2.13 – Bow-tie structure (strongly connected component (SCC), in-component (In), out-component (Out), tubes (Tubes), in-tendrils (In-T.), out-tendrils (Out-T.) and disconnected nodes (Disc.)) as confusion matrices. Numerical values represent total percentage of nodes while color grading shows row normalization. For each dataset, a total of 1000 synthetic networks were generated. See Sec. 2.7.4 for details on the datasets.

2.7.4 Network datasets

The network datasets used in the article have been made publicly available by the original authors and were downloaded from The Netzschleuder network catalogue and repository (<https://networks.skewed.de>). The letter identifying each dataset, its name and its reference are listed below:

- (a) Advogato trust relationships (dataset: `advogato_trust` [70])
- (b) Physician trust relationships (dataset: `physician_trust` [71])
- (c) FAO trade relationships (dataset: `fao_trade` [72])
- (d) Faculty hiring (dataset: `faculty_hiring_computer_science` [73])
- (e) University emails (dataset: `uni_email` [74])
- (f) Company emails (dataset: `email_company` [75])
- (g) Copenhagen Networks Study (calls) (dataset: `copenhagen_calls` [76])
- (h) Copenhagen Networks Study (sms) (dataset: `copenhagen_sms` [76])
- (i) Openflights airport network (dataset: `openflights` [77])
- (j) FAA air traffic routes (dataset: `faa_routes` [78])
- (k) UN migration stock (dataset: `un_migration` [79])
- (l) Political blogs (dataset: `polblogs` [80])
- (m) Adolescent friendships (dataset: `add_health_commm50` [81])
- (n) High school students friendships (dataset: `spanish_highschools_6` [82])
- (o) High school students contacts (dataset: `sp_high_school_diaries` [63])
- (p) Residence hall friendships (dataset: `residence_hall` [64])
- (q) Eocene Messel Shale food web (dataset: `messel_shale` [83])
- (r) Florida bay food web (dataset: `foodweb_baywet` [84])
- (s) Little Rock Lake food web (dataset: `foodweb_little_rock` [85])
- (t) Stream food web (dataset: `fresh_webs_stony` [86])
- (u) Animal dominance interaction (dataset: `dom_Strauss_2019d` [87])
- (v) Photosystem I of Pisum sativum (dataset: `psi` [88])
- (w) Human proteins interactome (dataset: `interactome_figeys` [89])
- (x) C. intestinalis connectome (dataset: `cintestinalis` [90])
- (y) C. elegans connectome (dataset: `celegansneural` [91])
- (z) C. elegans neurons (dataset: `celegans_2019_male_chemical` [92])
- (aa) E. coli transcription network (dataset: `ecoli_transcription` [93])
- (bb) Yeast transcription network (dataset: `yeast_transcription` [94])
- (cc) U.S. government websites (VT) (dataset: `us_agencies_vermont` [95])
- (dd) U.S. government websites (VT) (dataset: `us_agencies_washington` [95])

Conclusion

Vous l'aurez ainsi compris, les réseaux orientés ne constituent pas une exception, mais bien la norme parmi les systèmes complexes que l'on retrouve dans la nature et dans le monde qui nous entoure. L'asymétrie des interactions au sein de certains systèmes en est une qualité essentielle et souvent indispensable à leur bon fonctionnement. Ce projet de maîtrise visait à adapter le paradigme de la géométrie des réseaux au cas des graphes orientés, en développant un algorithme de plongement hyperbolique capable de préserver l'asymétrie des interactions tout comme les propriétés topologiques telles que les configurations de triangles ou encore la réciprocité individuelle de chaque noeud.

C'est ainsi que nous avons conçu l'algorithme diMercator et démontré sa pertinence dans l'étude des réseaux orientés. Son développement s'inscrit dans une démarche visant à enrichir l'arsenal d'outils disponibles pour l'analyse des réseaux complexes, aux côtés de Mercator, son homologue pour les réseaux non orientés. L'algorithme repose sur le modèle \mathcal{S}^1 orienté-réiproque, modèle géométrique de graphe orienté simple et robuste, permettant d'effectuer des plongements sur le cercle. Nos résultats montrent que diMercator, en plus de réaliser des plongements de haute fidélité, reproduit des propriétés clés observées dans les réseaux empiriques. Ces propriétés incluent notamment la cohérence trophique, la non-normalité, les configurations de triangles, la réciprocité individuelle, la structure en noeud papillon ainsi que la décomposition en valeurs singulières. Ce qui rend ces résultats particulièrement remarquables, c'est que les propriétés ainsi bien reproduites ne sont pourtant pas explicitement optimisées par le modèle. L'évaluation de l'algorithme sur des ensembles de données variés, incluant des réseaux sociaux, des connectomes, des réseaux trophiques et des réseaux de transcription, confirme son efficacité et souligne la pertinence de diMercator dans le cadre du paradigme géométrique des réseaux.

L'algorithme et le modèle ne sont pourtant pas sans défauts. En ce qui a trait au plongement hyperbolique comme tel, l'objectif n'est pas complètement atteint. La procédure d'inférence développée permet un plongement sur le cercle, mais le lien avec l'espace hyperbolique n'a pas encore été formellement démontré. Qui plus est, la complexité algorithmique quadratique de diMercator contribue à diminuer son efficacité sur de grands réseaux, augmentant considérablement le temps de calcul. Or, les approximations faites dans le modèle, notamment

en ce qui a trait à la distance hyperbolique, ont une meilleure validité plus on augmente la taille du réseau. Également, bien que le modèle soit défini dans un espace unidimensionnel, son extension à des dimensions supérieures pour reproduire des caractéristiques sous-jacentes supplémentaires pourrait augmenter de manière considérable les exigences de calcul. De plus, le modèle ne tient pas compte des métadonnées (attributs des noeuds) qui sont souvent importantes dans les réseaux réels (e.g., le groupe classe auquel appartient chaque élève d'un réseau d'amitié dans une école). Le modèle n'est pas non plus adapté aux réseaux pondérés ; l'étendre pour tenir compte des poids des liens (e.g., la force du lien d'amitié entre deux élèves d'une école) introduirait un niveau de complexité supplémentaire.

Pour l'avenir, il serait intéressant d'approfondir davantage l'approche de la géométrie des réseaux orientés en démontrant formellement le lien unifiant les graphes orientés à un espace hyperbolique sous-jacent tout en développant une version orientée du modèle \mathcal{H}^2 , permettant de plonger des graphes orientés directement sur le disque hyperbolique sans la nécessité d'avoir recours à des approximations. L'espace topologique engendré par ce modèle hypothétique n'est pourtant pas trivial, puisqu'il devrait encapsuler à la fois les propriétés des graphes orientés ainsi qu'une carte locale d'un espace hyperbolique. Puisque le modèle \mathcal{S}^1 orienté-réiproque réalise un plongement dans l'espace en trois dimensions $\mathbb{S}^1 \times \mathbb{R}_+ \times \mathbb{R}_+$ des coordonnées $\boldsymbol{\theta} := (\theta_1, \dots, \theta_N)$ sur le cercle \mathbb{S}^1 et des degrés cachés $\boldsymbol{\kappa} := (\kappa_1^+, \kappa_1^-, \dots, \kappa_N^+, \kappa_N^-)$, il semblerait naturel d'envisager un espace hyperbolique en trois dimensions comme \mathbb{H}^3 . Or, cet espace est paramétré par une coordonnée radiale et deux coordonnées angulaires alors que l'espace $\mathbb{S}^1 \times \mathbb{R}_+ \times \mathbb{R}_+$ suggère plutôt un modèle où il y aurait deux coordonnées radiales et une coordonnée angulaire. Il serait alors nécessaire d'effectuer un changement de variables bijectif et différentiable entre les deux espaces. Une autre hypothèse serait de réaliser deux plongements dans l'espace \mathbb{H}^2 : un premier plongement en ne considérant que le degré entrant de chaque sommet et un second en ne considérant que le degré sortant, à l'instar de ce que l'on retrouve dans la littérature [96, 97]. Ces deux plongements seraient ensuite combinés en un seul où chaque sommet aurait alors une coordonnée angulaire et deux coordonnées radiales étant respectivement liées aux degrés entrant et sortant du graphe initial.

Une telle avancée, bien que techniquement complexe, ouvrirait la voie à une modélisation encore plus riche des systèmes complexes aux interactions asymétriques, complétant ainsi le cadre des plongements hyperboliques pour les graphes orientés. Cela représenterait un outil analytique particulièrement puissant et enrichirait un arsenal méthodologique toujours plus sophistiqué. Bien que le lien entre les réseaux orientés et un éventuel espace hyperbolique sous-jacent reste à éclaircir pleinement, ce mémoire pose les fondations nécessaires au développement de ce formalisme et présente des résultats susceptibles de nourrir l'intérêt croissant pour cette classe de modèles. En contribuant à étendre le cadre géométrique aux graphes orientés, ce travail ouvre de nouvelles perspectives pour approfondir la compréhension des systèmes et des réseaux complexes.

Bibliographie

- [1] Sayama, H. *Introduction to the Modeling and Analysis of Complex Systems*. SUNY Geneseo (2015).
- [2] Parisi, G. *Complex systems : a physicist's viewpoint*. Physica A **263**, 557–564 (1999).
- [3] Anderson, P. W. *More Is Different*. Science **177**, 393–396 (1972).
- [4] De Domenico, M., Brockmann, D., Camargo, C., Gershenson, C. et al.. *Complexity Explained*. University of Exeter Repository (2019).
- [5] Latora, V., Nicosia, V. et Russo, G. *Complex Networks : Principles, Methods and Applications*. Cambridge University Press (2017).
- [6] Newman, M. E. J. *Networks*. Oxford University Press (2018).
- [7] Boguñá, M., Bonamassa, I., De Domenico, M., Havlin, S. et al.. *Network Geometry*. Nature Reviews Physics **3**, 114–135 (2021).
- [8] Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A. et Boguñá, M. *Hyperbolic Geometry of Complex Networks*. Physical Review E **82** (2010).
- [9] Serrano, M. Á., Krioukov, D. et Boguñá, M. *Self-Similarity of Complex Networks and Hidden Metric Spaces*. Physical Review Letters **100**, 078701 (2008).
- [10] Boguñá, M., Krioukov, D., Almagro, P. et Serrano, M. Á. *Small worlds and clustering in spatial networks*. Physical Review Research **2**, 023040 (2020).
- [11] Krioukov, D., Papadopoulos, F., Vahdat, A. et Boguñá, M. *Curvature and temperature of complex networks*. Physical Review E **80**, 035101 (2009).
- [12] García-Pérez, G., Allard, A., Serrano, M. Á. et Boguñá, M. *Mercator : uncovering faithful hyperbolic embeddings of complex networks*. New Journal of Physics **21**, 123033 (2019).
- [13] Alanis-Lobato, G., Mier, P. et Andrade-Navarro, M. A. *Efficient embedding of complex networks to hyperbolic space via their Laplacian*. Scientific Reports **6**, 30108 (2016).

- [14] Muscoloni, A., Thomas, J. M., Ciucci, S., Bianconi, G. et Cannistraci, C. V. *Machine learning meets complex networks via coalescent embedding in the hyperbolic space*. Nature Communications **8**, 1615 (2017).
- [15] Boguñá, M., Papadopoulos, F. et Krioukov, D. *Sustaining the Internet with hyperbolic mapping*. Nature Communications **1**, 62 (2010).
- [16] Allard, A., Serrano, M. Á. et Boguñá, M. *Geometric description of clustering in directed networks*. Nature Physics **20**, 150–156 (2024).
- [17] Mitchell, M. *Complexity : A Guided Tour*. Oxford University Press (2009).
- [18] Bondy, A. et Murty, U. S. R. *Graph Theory*. Springer Science and Business Media LLC (2008).
- [19] Garlaschelli, D. et Loffredo, M. I. *Patterns of Link Reciprocity in Directed Networks*. Physical Review Letters **93**, 268701 (2004).
- [20] Broder, A., Kumar, R., Maghoul, F., Raghavan, P. et al.. *Graph structure in the web*. Computer Networks **33**, 309–320 (2000).
- [21] Newman, M. E. J., Strogatz, S. H. et Watts, D. J. *Random graphs with arbitrary degree distributions and their applications*. Physical Review E **64**(2) (2001).
- [22] Riley, K. F., Hobson, M. P. et Bence, S. J. *Mathematical methods for physics and engineering*. Cambridge University Press, 3e édition (2006).
- [23] Thibeault, V., Allard, A. et Desrosiers, P. *The low-rank hypothesis of complex systems*. Nature Physics **20**, 294–302 (2024).
- [24] Levine, S. *Several measures of trophic structure applicable to complex food webs*. Journal of Theoretical Biology **83**(2), 195–207 (1980).
- [25] Antràs, P., Chor, D., Fally, T. et Hillberry, R. *Measuring the Upstreamness of Production and Trade Flows*. American Economic Review **102**(3), 412–416 (2012).
- [26] Johnson, S., Domínguez-García, V., Donetti, L. et Muñoz, M. A. *Trophic coherence determines food-web stability*. Proceedings of the National Academy of Sciences **111**(50), 17923–17928 (2014).
- [27] Johnson, S. *Digraphs are different : why directionality matters in complex systems*. Journal of Physics : Complexity **1**, 015003 (2020).
- [28] MacKay, R. S., Johnson, S. et Sansom, B. *How directed is a directed network ?* Royal Society Open Science **7**(9) (2020).

- [29] Asllani, M. et Carletti, T. *Topological resilience in non-normal networked systems*. Physical Review E **97**(4) (2018).
- [30] Asllani, M., Lambiotte, R. et Carletti, T. *Structure and dynamical behavior of non-normal networks*. Science Advances **4**(12) (2018).
- [31] Trefethen, L. N. et Embree, M. *Spectra and Pseudospectra : The behavior of nonnormal matrices and operators*. Princeton University Press (2020).
- [32] Baccou, R. *Platon. Œuvres complètes. La République. Traduction nouvelle avec introduction et notes*. Librairie Garnier Frères, Paris (1936).
- [33] Peyrard, F. *Les éléments de géométrie d'Euclide, traduits littéralement et suivis d'un traité du cercle, du cylindre, du cône et de la sphère, de la mesure des surfaces et des solides, avec des notes*. École Polytechnique, Paris (1804). Bibliothèque nationale de France.
- [34] Kuhn, T. S. *La structure des révolutions scientifiques*. Flammarion (2018).
- [35] Gingras, Y., Keating, P. et Limoges, C. *Du scribe au savant*. Boréal (2012).
- [36] Nicotra, L. *La verità in matematica : da assoluta a relativa*. ArteScienza **III**, 71–146 (2016).
- [37] Needham, T. *Visual Differential Geometry and Forms : A Mathematical Drama in Five Acts*. Princeton University Press, Princeton (2021).
- [38] Stillwell, J. *Mathematics and Its History*. Springer Science and Business Media LLC (2010).
- [39] Lobatchevski, N. *Géométrie imaginaire et pangéométrie*. Nabu Press (2011).
- [40] Battaglini, G. *Sur la géométrie imaginaire de Lobatcheffsky*. Rencondiconto della R. Accademia delle Scienze fisiche e matematiche di Napoli, Naples (1868).
- [41] Vassilief, A. *Éloge historique de Nicolas-J. Lobatchevski*. Librairie scientifique A. Hermann, Paris (1896).
- [42] Beltrami. *Essai d'interprétation de la géométrie non euclidienne*. Université de Bologne, Bologne (1868).
- [43] Poincaré, H. *La Science et l'hypothèse*. Flammarion, Paris (1902).
- [44] Escher, M. C. *Cercle limite 1* (1958).
- [45] Ernst, B. *Le miroir magique de M. C. Escher*. Taschen (1994).

- [46] Saint-Aubin, Y. *Pavages hyperboliques*. Accromath **5**, 4–9 (2010).
- [47] Carroll, S. M. *Spacetime and Geometry : An Introduction to General Relativity*. Cambridge University Press (2019).
- [48] Shakespeare, W. *La tragique histoire d'Hamlet, prince de Danemark* (1603).
- [49] Ratcliffe, J. G. *Foundations of Hyperbolic Manifolds*. Springer Science and Business Media LLC, 3e édition (2019).
- [50] Désy, B. *Effets de la dimension des réseaux hyperboliques sur la modélisation de la structure communautaire*. Mémoire de maîtrise, Université Laval, Québec, Canada (2022).
- [51] Stillwell, J. *Sources of Hyperbolic Geometry*. American Mathematical Society and London Mathematical Society (1996).
- [52] Désy, B., Desrosiers, P. et Allard, A. *Dimension matters when modeling network communities in hyperbolic spaces*. PNAS Nexus **2** (2023).
- [53] Cannon, J. W., Smillie, J., Ball, K. et Bollobas, B. *Flavors of Geometry*, volume 31. Cambridge University Press (1997).
- [54] Walkden, C. *Lecture notes on Hyperbolic Geometry*. University of Manchester (2019).
- [55] Serrano, M. Á. et Boguñá, M. *The Shortest Path to Network Geometry : A Practical Guide to Basic Models and Applications*. Elements in the Structure and Dynamics of Complex Networks. Cambridge University Press (2022).
- [56] García-Pérez, G. *A geometric approach to the structure of complex networks*. Thèse de doctorat, Universitat de Barcelona, Barcelona, Spain (2018).
- [57] Kitsak, M., Ganin, A., Elmokashfi, A., Cui, H. et al.. *Finding shortest and nearly shortest path nodes in large substantially incomplete networks by hyperbolic mapping*. Nature Communications **14** (2023).
- [58] Qu, B., Li, Q., Havlin, S., Stanley, H. E. et Wang, H. *Nonconsensus opinion model on directed networks*. Physical Review E **90**, 052811 (2014).
- [59] Klaise, J. et Johnson, S. *From neurons to epidemics : How trophic coherence affects spreading processes*. Chaos **26** (2016).
- [60] Peralta-Martinez, K. et Méndez-Bermúdez, J. A. *Directed random geometric graphs : Structural and spectral properties*. Journal of Physics : Complexity **4**, 015002 (2022).
- [61] Holland, P. et Leinhardt, S. *Local Structure in Social Networks*. Sociological Methodology **7**, 1 (1976).

- [62] Ahnert, S. E. et Fink, T. M. A. *Clustering signatures classify directed networks*. Physical Review E **78**, 036112 (2008).
- [63] Mastrandrea, R., Fournet, J. et Barrat, A. *Contact Patterns in a High School : A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys*. PLOS ONE **10**, e0136497 (2015).
- [64] Freeman, L. C., Webster, C. M. et Kirke, D. M. *Exploring social structure using dynamic three-dimensional color images*. Social Networks **20**, 109–118 (1998).
- [65] Stehman, S. V. *Selecting and interpreting measures of thematic classification accuracy*. Remote Sensing of Environment **62**(1), 77–89 (1997).
- [66] Jankowski, R., Allard, A., Boguñá, M. et Serrano, M. Á. *The D-Mercator method for the multidimensional hyperbolic embedding of real networks*. Nature Communications **14** (2023).
- [67] Belkin, M. et Niyogi, P. *Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering*. Dans *Advances in Neural Information Processing Systems*, volume 14. MIT Press (2001).
- [68] Hébert-Dufresne, L., Grochow, J. A. et Allard, A. *Multi-scale structure and topological anomaly detection via a new network statistic : The onion decomposition*. Scientific Reports **6**, 31708 (2016).
- [69] Horn, R. A. et Johnson, C. R. *Matrix Analysis*. Cambridge University Press (1985).
- [70] Massa, P., Salvetti, M. et Tomasoni, D. *Bowling Alone and Trust Decline in Social Network Sites*. Dans *2009 eighth IEEE international conference on dependable, autonomic and secure computing*, p. 658–663. IEEE (2009).
- [71] Coleman, J., Katz, E. et Menzel, H. *The Diffusion of an Innovation Among Physicians*. Sociometry **20**, 253 (1957).
- [72] De Domenico, M., Nicosia, V., Arenas, A. et Latora, V. *Structural reducibility of multilayer networks*. Nature Communications **6** (2015).
- [73] Clauset, A., Arbesman, S. et Larremore, D. B. *Systematic inequality and hierarchy in faculty hiring networks*. Science Advances **1**(1) (2015).
- [74] Guimerà, R., Danon, L., Díaz-Guilera, A., Giralt, F. et Arenas, A. *Self-similar community structure in a network of human interactions*. Physical Review E **68**(6) (2003).
- [75] Michalski, R., Palus, S. et Kazienko, P. *Matching Organizational Structure and Social Network Extracted from Email Communication*. Dans *Business information systems*, pp. 197–206. Springer Science and Business Media LLC, Berlin, Heidelberg (2011).

- [76] Sapiezynski, P., Stopczynski, A., Lassen, D. D. et Lehmann, S. *Interaction data from the Copenhagen Networks Study*. Scientific Data **6**(1) (2019).
- [77] *OpenFlights.org : Flight logging, mapping, stats and sharing — openflights.org*.
- [78] *National Airspace System — nasstatus.faa.gov*.
- [79] *Welcome to the United Nations — un.org*.
- [80] Adamic, L. A. et Glance, N. *The Political Blogosphere and the 2004 U.S. Election : Divided They Blog*. Dans *Proceedings of the 3rd international workshop on Link discovery*, KDD05. ACM (2005).
- [81] Moody, J. *Peer influence groups : identifying dense clusters in large networks*. Social Networks **23**(4), 261–283 (2001).
- [82] Ruiz-García, M., Ozaita, J., Pereda, M., Alfonso, A. et al.. *Triadic influence as a proxy for compatibility in social relationships*. Proceedings of the National Academy of Sciences **120**(13) (2023).
- [83] Dunne, J. A., Labandeira, C. C. et Williams, R. J. *Highly resolved early Eocene food webs show development of modern trophic structure after the end-Cretaceous extinction*. Proceedings of the Royal Society B : Biological Sciences **281**(1782), 20133280 (2014).
- [84] Ulanowicz, R., Bondavalli, C. et Egnatovich, M. *Network Analysis of Trophic Dynamics in South Florida Ecosystem, FY 97 : The Florida Bay Ecosystem*. Annual Report to the United States Geological Service Biological Resources Division. (1998).
- [85] Martinez, N. D. *Artifacts or Attributes ? Effects of Resolution on the Little Rock Lake Food Web*. Ecological Monographs **61**, 367–392 (1991).
- [86] Thompson, R. M. et Townsend, C. R. *Impacts on stream food webs of native and exotic forest : An intercontinental comparison*. Ecology **84**(1), 145–161 (2003).
- [87] Strauss, E. D., DeCasien, A. R., Galindo, G., Hobson, E. A. et al.. *DomArchive : a century of published dominance data*. Philosophical Transactions of the Royal Society B : Biological Sciences **377**(1845) (2022).
- [88] Montepietra, D., Bellingeri, M., Ross, A. M., Scotognella, F. et Cassi, D. *Modelling photosystem I as a complex interacting network*. Journal of The Royal Society Interface **17**(172), 20200813 (2020).
- [89] Ewing, R. M., Chu, P., Elisma, F., Li, H. et al.. *Large-scale mapping of human protein-protein interactions by mass spectrometry*. Molecular Systems Biology **3**(1) (2007).

- [90] Ryan, K., Lu, Z. et Meinertzhagen, I. A. *The CNS connectome of a tadpole larva of Ciona intestinalis (L.) highlights sidedness in the brain of a chordate sibling.* eLife **5**, e16962 (2016).
- [91] Watts, D. J. et Strogatz, S. H. *Collective dynamics of ‘small-world’ networks.* Nature **393**(6684), 440–442 (1998).
- [92] Cook, S. J., Jarrell, T. A., Brittin, C. A., Wang, Y. et al.. *Whole-animal connectomes of both Caenorhabditis elegans sexes.* Nature **571**(7763), 63–71 (2019).
- [93] Shen-Orr, S. S., Milo, R., Mangan, S. et Alon, U. *Network motifs in the transcriptional regulation network of escherichia coli.* Nature Genetics **31**, 64–68 (2002).
- [94] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N. et al.. *Network Motifs : Simple Building Blocks of Complex Networks.* Science **298**, 824–827 (2002).
- [95] Kosack, S., Coscia, M., Smith, E., Albrecht, K. et al.. *Functional structures of US state governments.* Proceedings of the National Academy of Sciences **115**, 11748–11753 (2018).
- [96] Kovács, B. et Palla, G. *Model-independent embedding of directed networks into Euclidean and hyperbolic spaces.* Communications Physics **6**(1) (2023).
- [97] Wu, Z., Di, Z. et Fan, Y. *An Asymmetric Popularity-Similarity Optimization Method for Embedding Directed Networks into Hyperbolic Space.* Complexity **2020**, 1–16 (2020).
- [98] Gleeson, J. P. *High-Accuracy Approximation of Binary-State Dynamics on Networks.* Physical Review Letters **107**, 068701 (2011).
- [99] Gleeson, J. P. *Binary-State Dynamics on Complex Networks : Pair Approximation and Beyond.* Physical Review X **3**, 021004 (2013).
- [100] St-Onge, G., Thibeault, V., Allard, A., Dubé, L. J. et Hébert-Dufresne, L. *Social Confinement and Mesoscopic Localization of Epidemics on Networks.* Physical Review Letters **126**, 098301 (2021).
- [101] Nematzadeh, A., Ferrara, E., Flammini, A. et Ahn, Y.-Y. *Optimal Network Modularity for Information Diffusion.* Physical Review Letters **113**, 088701 (2014).
- [102] Rodriguez, N., Izquierdo, E. et Ahn, Y.-Y. *Optimal modularity and memory capacity of neural reservoirs.* Network Neuroscience **3**, 551–566 (2019).
- [103] Allard, A., Hébert-Dufresne, L., Young, J. et Dubé, L. J. *General and exact approach to percolation on random graphs.* Physical Review E **92**, 062807 (2015).

Annexe A

Résultats mathématiques utiles

A.1 Fonction hypergéométrique

Les intégrales suivantes ont trait à la fonction hypergéométrique ${}_2F_1(a, b, c, z)$ et sont valides sur le domaine $x > 0$, $\beta > 1$, $\kappa > 0$ et $C \in \mathbb{R}$ [16]. Les identités sont les suivantes :

$$\int \frac{1}{1+x^\beta} dx = x \cdot {}_2F_1\left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -x^\beta\right) + C \quad (\text{A.1})$$

$$\int \frac{x}{1+x^\beta} dx = \frac{1}{2}x^2 \cdot {}_2F_1\left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, -x^\beta\right) + C \quad (\text{A.2})$$

$$\int \frac{1}{(1+x^\beta)^2} dx = x \cdot {}_2F_1\left(2, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -x^\beta\right) + C \quad (\text{A.3})$$

$$\int \frac{x}{(1+x^\beta)^2} dx = \frac{1}{2}x^2 \cdot {}_2F_1\left(2, \frac{2}{\beta}, 1 + \frac{2}{\beta}, -x^\beta\right) + C \quad (\text{A.4})$$

$$\int \frac{1}{1+x^\beta} \frac{1}{1+(\kappa x)^\beta} dx = \begin{cases} x \cdot {}_2F_1\left(2, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -x^\beta\right) + C & \kappa = 1 \\ \frac{x}{1-\kappa^\beta} {}_2F_1\left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -x^\beta\right) \\ -\frac{\kappa^\beta x}{1-\kappa^\beta} {}_2F_1\left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -(\kappa x)^\beta\right) + C & \kappa \neq 1 \end{cases} \quad (\text{A.5})$$

$$\int x \cdot \frac{1}{1+x^\beta} \frac{1}{1+(\kappa x)^\beta} dx = \begin{cases} \frac{1}{2}x^2 \cdot {}_2F_1\left(2, \frac{2}{\beta}, 1 + \frac{2}{\beta}, -x^\beta\right) + C & \kappa = 1 \\ \frac{x^2}{2(1-\kappa^\beta)} {}_2F_1\left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, -x^\beta\right) \\ -\frac{\kappa^\beta x^2}{2(1-\kappa^\beta)} {}_2F_1\left(1, \frac{2}{\beta}, 1 + \frac{2}{\beta}, -(\kappa x)^\beta\right) + C & \kappa \neq 1 \end{cases} \quad (\text{A.6})$$

A.2 Décomposition en fractions partielles

Le développement explicite de l'intégrale

$$\int \frac{1}{1+x^\beta} \frac{1}{1+(\kappa x)^\beta} dx = \frac{1}{1-\kappa^\beta} \int \frac{1}{1+x^\beta} dx - \frac{\kappa^\beta}{1-\kappa^\beta} \int \frac{1}{1+(\kappa x)^\beta} dx \quad (\text{A.7})$$

est réalisé suivant la décomposition en fractions partielles de l'intégrande

$$\begin{aligned} \frac{1}{1+x^\beta} \frac{1}{1+(\kappa x)^\beta} &= \frac{1}{(1+x^\beta)(1+(\kappa x)^\beta)} \\ &= \frac{1-\kappa^\beta}{(1-\kappa^\beta)(1+x^\beta)(1+(\kappa x)^\beta)} \\ &= \frac{1-\kappa^\beta + (\kappa x)^\beta - (\kappa x)^\beta}{(1-\kappa^\beta)(1+x^\beta)(1+(\kappa x)^\beta)} \\ &= \frac{(1-(\kappa x)^\beta) - \kappa^\beta(1+x^\beta)}{(1-\kappa^\beta)(1+x^\beta)(1+(\kappa x)^\beta)} \\ &= \frac{1}{(1-\kappa^\beta)(1+x^\beta)} - \frac{\kappa^\beta}{(1-\kappa^\beta)(1+(\kappa x)^\beta)} \\ &= \frac{1}{1-\kappa^\beta} \frac{1}{1+x^\beta} - \frac{\kappa^\beta}{1-\kappa^\beta} \frac{1}{1+(\kappa x)^\beta}. \end{aligned} \quad (\text{A.8})$$

Le développement explicite de l'intégrale

$$\int x \cdot \frac{1}{1+x^\beta} \frac{1}{1+(\kappa x)^\beta} dx = \frac{1}{1-\kappa^\beta} \int \frac{x}{1+x^\beta} dx - \frac{\kappa^\beta}{1-\kappa^\beta} \int \frac{x}{1+(\kappa x)^\beta} dx \quad (\text{A.9})$$

est réalisé suivant la décomposition en fractions partielles de l'intégrande

$$\begin{aligned} x \cdot \frac{1}{1+x^\beta} \frac{1}{1+(\kappa x)^\beta} &= \frac{x}{(1+x^\beta)(1+(\kappa x)^\beta)} \\ &= \frac{(1-\kappa^\beta)x}{(1-\kappa^\beta)(1+x^\beta)(1+(\kappa x)^\beta)} \\ &= \frac{(1-\kappa^\beta + (\kappa x)^\beta - (\kappa x)^\beta)x}{(1-\kappa^\beta)(1+x^\beta)(1+(\kappa x)^\beta)} \\ &= \frac{(1-(\kappa x)^\beta)x - \kappa^\beta(1+x^\beta)x}{(1-\kappa^\beta)(1+x^\beta)(1+(\kappa x)^\beta)} \\ &= \frac{x}{(1-\kappa^\beta)(1+x^\beta)} - \frac{\kappa^\beta x}{(1-\kappa^\beta)(1+(\kappa x)^\beta)} \\ &= \frac{1}{1-\kappa^\beta} \frac{x}{1+x^\beta} - \frac{\kappa^\beta}{1-\kappa^\beta} \frac{x}{1+(\kappa x)^\beta}. \end{aligned} \quad (\text{A.10})$$

Annexe B

Équations maîtresses approximées générales pour dynamiques binaires sur graphes aléatoires

Le projet *CoDyn* a été entamé lors d'un stage à la session d'été 2021 sous la supervision de l'étudiant au doctorat Charles Murphy, puis continué au début de la maîtrise. Un article scientifique est en cours de rédaction en date du dépôt de ce mémoire. Ce document consolide le travail effectué dans le projet et présente le modèle développé ainsi que les résultats préliminaires obtenus.

Dans ce projet, les équations maîtresses d'une dynamique binaire sur réseau selon le modèle épidémiologique SIS (susceptible-infecté-susceptible) ainsi que leurs approximations sont dérivées selon différentes approches. Dans le modèle SIS, nous avons la probabilité qu'un noeud soit susceptible $s_{k,m}(t)$ ou infecté $i_{k,m}(t)$ avec k , le degré du noeud, et m , le nombre de voisins infectés, puis les taux d'activation (taux d'infection) $\alpha_{k,m} = m\alpha$ et de désactivation (taux de rétablissement) $\beta_{k,m} = \beta$. Les paramètres θ_s , θ_i , ϕ_s et ϕ_i correspondent respectivement au taux auquel un noeud S connecté à un noeud S (lien S – S) ou un noeud I (lien S – I) de degré k est infecté et au taux auquel un noeud I connecté à un noeud S (lien I – S) ou un noeud I (lien I – I) de degré k redevient susceptible.

B.1 Travaux connexes

La dérivation mathématique suivante est conforme à celle de Gleeson [98, 99].

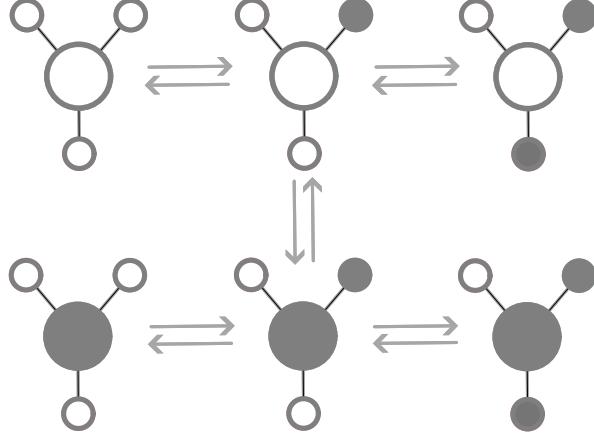


FIGURE B.1 – Transitions compartimentales dans le modèle AME pour un noeud de degré $k = 3$ et de degré actif $m = 1$. Les noeuds foncées sont actifs, tandis que les noeuds clairs sont inactifs.

B.1.1 Équations maîtresses approximées (AME)

Les équations maîtresses sont

$$\begin{aligned} \frac{d}{dt} s_{m|k} &= -\alpha_{k,m} s_{m|k} + \beta_{k,m} i_{m|k} \\ &\quad - \theta_s(k-m) s_{m|k} + \theta_s(k-m+1) s_{m-1|k} \\ &\quad - \phi_s m s_{m|k} + \phi_s(m+1) s_{m+1|k}, \end{aligned} \tag{B.1a}$$

$$\begin{aligned} \frac{d}{dt} i_{m|k} &= -\beta_{k,m} i_{m|k} + \alpha_{k,m} s_{m|k} \\ &\quad - \theta_i(k-m) i_{m|k} + \theta_i(k-m+1) i_{m-1|k} \\ &\quad - \phi_i m i_{m|k} + \phi_i(m+1) i_{m+1|k}, \end{aligned} \tag{B.1b}$$

avec les taux de transition d'état

$$\theta_s = \frac{\langle \sum_{m=0}^k (k-m) \alpha_{k,m} s_{m|k} \rangle_k}{\langle \sum_{m=0}^k (k-m) s_{m|k} \rangle_k}, \tag{B.2a}$$

$$\phi_s = \frac{\langle \sum_{m=0}^k (k-m) \beta_{k,m} i_{m|k} \rangle_k}{\langle \sum_{m=0}^k (k-m) i_{m|k} \rangle_k}, \tag{B.2b}$$

$$\theta_i = \frac{\langle \sum_{m=0}^k m \alpha_{k,m} s_{m|k} \rangle_k}{\langle \sum_{m=0}^k m s_{m|k} \rangle_k}, \tag{B.2c}$$

$$\phi_i = \frac{\langle \sum_{m=0}^k m \beta_{k,m} i_{m|k} \rangle_k}{\langle \sum_{m=0}^k m i_{m|k} \rangle_k}. \tag{B.2d}$$

Dans les équations (B.1a) et (B.1b), les deux premiers termes correspondent à la probabilité qu'un noeud de degré k à m voisins infectés devienne infecté (équation (B.1a)) ou susceptible (équation (B.1b)). Les deux termes suivant correspondent à la probabilité qu'un voisin susceptible devienne infecté alors que les deux derniers termes correspondent à la probabilité qu'un voisin infecté redevienne susceptible.

B.1.2 Approximation par pairs (PA)

L'approximation par pairs permet de définir de nouvelles probabilités d'infection et de rétablissement proportionnelles à une distribution binomiale selon

$$s_{m|k} = s_k B_{m|k}(p_k), \quad (\text{B.3a})$$

$$i_{m|k} = i_k B_{m|k}(q_k), \quad (\text{B.3b})$$

où la loi de probabilité binomiale est

$$B_{m|k}(x) = \binom{k}{m} x^m (1-x)^{k-m}.$$

Nous avons maintenant la relation

$$i_k = 1 - s_k,$$

qui permet de simplifier le calcul des nouvelles probabilités correspondant aux probabilités qu'un voisin d'un noeud susceptible ou infecté devienne infecté

$$p_k = \frac{\sum_{m=0}^k m s_{m|k}}{k s_k}, \quad (\text{B.4a})$$

$$q_k = \frac{\sum_{m=0}^k m i_{m|k}}{k i_k}. \quad (\text{B.4b})$$

Les équations d'évolution sont

$$\frac{d}{dt} s_k = s_k \sum_{m=0}^k \eta_{k,m}, \quad (\text{B.5a})$$

$$\frac{d}{dt} p_k = \sum_{m=0}^k \left(\frac{m}{k} - p_k \right) \eta_{k,m} + \theta_s (1 - p_k) - \phi_s p_k, \quad (\text{B.5b})$$

$$\frac{d}{dt} q_k = \sum_{m=0}^k \left(\frac{m}{k} - q_k \right) \xi_{k,m} + \theta_i (1 - q_k) - \phi_i q_k, \quad (\text{B.5c})$$

avec de nouveaux paramètres

$$\eta_{k,m} = -\alpha_{k,m} B_{m|k}(p_k) + \left(\frac{1 - s_k}{s_k} \right) \beta_{k,m} B_{m|k}(q_k), \quad (\text{B.6a})$$

$$\xi_{k,m} = -\beta_{k,m} B_{m|k}(q_k) + \left(\frac{s_k}{1 - s_k} \right) \alpha_{k,m} B_{m|k}(p_k), \quad (\text{B.6b})$$

et les taux de transition d'état

$$\theta_s = \frac{\langle \sum_{m=0}^k (k-m) \alpha_{k,m} s_k B_{m|k}(p_k) \rangle_k}{\langle k s_k (1-p_k) \rangle_k}, \quad (\text{B.7a})$$

$$\phi_s = \frac{\langle \sum_{m=0}^k (k-m) \beta_{k,m} (1-s_k) B_{m|k}(p_k) \rangle_k}{\langle k (1-s_k) (1-q_k) \rangle_k}, \quad (\text{B.7b})$$

$$\theta_i = \frac{\langle \sum_{m=0}^k m \alpha_{k,m} s_k B_{m|k}(q_k) \rangle_k}{\langle k s_k p_k \rangle_k}, \quad (\text{B.7c})$$

$$\phi_i = \frac{\langle \sum_{m=0}^k m \beta_{k,m} (1-s_k) B_{m|k}(q_k) \rangle_k}{\langle k (1-s_k) q_k \rangle_k}. \quad (\text{B.7d})$$

B.1.3 Approximation par champ-moyen (MF)

L'approximation par champ-moyen permet la simplification des probabilités selon

$$p_k = q_k = \omega, \quad (\text{B.8})$$

où la probabilité qu'un lien aléatoire soit connecté à au moins un noeud infecté est

$$\omega = \frac{\langle k(1-s_k) \rangle_k}{\langle k \rangle_k}. \quad (\text{B.9})$$

Les équations d'évolution deviennent

$$\frac{d}{dt} s_k = \sum_{m=0}^k \eta_{k,m} s_k, \quad (\text{B.10})$$

avec le paramètre

$$\eta_{k,m} = -\alpha_{k,m} B_{m|k}(\omega) + \left(\frac{1-s_k}{s_k} \right) \beta_{k,m} B_{m|k}(\omega). \quad (\text{B.11})$$

B.2 Approche généralisée par types de noeuds

Supposons maintenant que nous avons un réseau dont les noeuds sont classés selon différents types d'un espace des types \mathcal{B} avec une distribution de types ρ_μ correspondant à la probabilité de trouver un noeud de type μ et la fonction de probabilité conditionnelle type-degré $\rho_{\mathbf{k}|\mu,k}$ qui donne la probabilité qu'un noeud de degré k et de type μ ait un vecteur type-degré \mathbf{k} où la composante k^ν est le degré du noeud par rapport au type ν . Sachant que $\rho_{k|\mu}$ existe, la fonction de probabilité conjointe type-degré est $\rho_{\mathbf{k},\mu} = \rho_\mu \rho_{k|\mu} \rho_{\mathbf{k}|\mu,k}$.

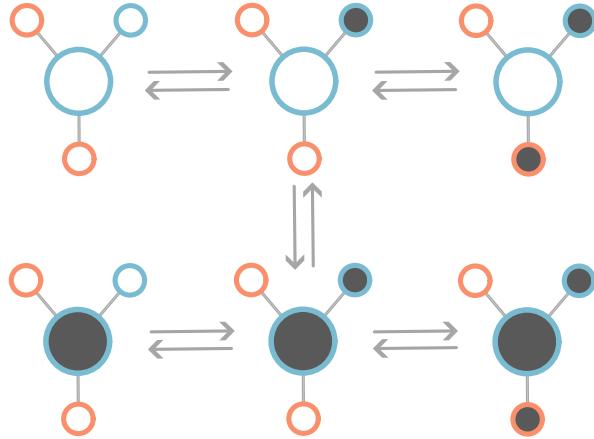


FIGURE B.2 – Transitions compartimentales dans le modèle TAME pour un noeud bleu de degré généralisé $\mathbf{k} = (2,1)$ et de degré actif généralisé $\mathbf{m} = (0,1)$ (les premières entrées correspondent aux voisins oranges, et les secondes entrées, aux voisins bleus). Le contour des noeuds code leur appartenance à un type : orange ou bleu. Les noeuds foncées sont actifs, tandis que les clairs sont inactifs.

B.2.1 Équations maîtresses approximées typées (TAME)

Afin d'approximer les équations maîtresses, nous devons d'abord définir de nouvelles probabilités de trouver un noeud susceptible ou infecté selon la classification par types. Ces probabilités sont $s_{\mathbf{m}|\mathbf{k}}^\mu(t)$ et $i_{\mathbf{m}|\mathbf{k}}^\mu(t)$ où un élément m^ν du vecteur \mathbf{m} est le nombre de voisins infectés de type ν , un élément k^μ du vecteur \mathbf{k} est le degré du noeud de type ν , le degré $k = \sum_{\nu \in \mathcal{B}} k^\nu$ est la somme des éléments du vecteur \mathbf{k} et le nombre total de voisins infectés $m = \sum_{\nu \in \mathcal{B}} m^\nu$ est la somme des éléments du vecteur \mathbf{m} . Les équations maîtresses deviennent

$$\begin{aligned} \frac{d}{dt} s_{\mathbf{m}|\mathbf{k}}^\mu &= -\alpha_{k,m} s_{\mathbf{m}|\mathbf{k}}^\mu + \beta_{k,m} i_{\mathbf{m}|\mathbf{k}}^\mu \\ &+ \sum_{\nu \in \mathcal{B}} \left\{ \theta_s^{\mu,\nu} \left[-(k^\nu - m^\nu) s_{\mathbf{m}|\mathbf{k}}^\mu + (k^\nu - m^\nu + 1) s_{\mathbf{m}-\mathbf{e}_\nu|\mathbf{k}}^\mu \right] \right. \\ &\quad \left. + \phi_s^{\mu,\nu} \left[-m^\nu s_{\mathbf{m}|\mathbf{k}}^\mu + (m^\nu + 1) s_{\mathbf{m}+\mathbf{e}_\nu|\mathbf{k}}^\mu \right] \right\}, \end{aligned} \quad (\text{B.12a})$$

$$\begin{aligned} \frac{d}{dt} i_{\mathbf{m}|\mathbf{k}}^{\mu} &= -\beta_{k,m} i_{\mathbf{m}|\mathbf{k}}^{\mu} + \alpha_{k,m} s_{\mathbf{m}|\mathbf{k}}^{\mu} \\ &+ \sum_{\nu \in \mathcal{B}} \left\{ \theta_i^{\mu,\nu} \left[-(k^{\nu} - m^{\nu}) i_{\mathbf{m}|\mathbf{k}}^{\mu} + (k^{\nu} - m^{\nu} + 1) i_{\mathbf{m}-\mathbf{e}_{\nu}|\mathbf{k}}^{\mu} \right] \right. \\ &\left. + \phi_i^{\mu,\nu} \left[-m^{\nu} i_{\mathbf{m}|\mathbf{k}}^{\mu} + (m^{\nu} + 1) i_{\mathbf{m}+\mathbf{e}_{\nu}|\mathbf{k}}^{\mu} \right] \right\}. \end{aligned} \quad (\text{B.12b})$$

avec les taux de transition d'état

$$\theta_s^{\mu,\nu} = \frac{\langle \sum_{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}} (k^{\mu} - m^{\mu}) \alpha_{k,m} s_{\mathbf{m}|\mathbf{k}}^{\nu} \rangle_{\mathbf{k}|\nu}}{\langle \sum_{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}} (k^{\mu} - m^{\mu}) s_{\mathbf{m}|\mathbf{k}}^{\nu} \rangle_{\mathbf{k}|\nu}}, \quad (\text{B.13a})$$

$$\phi_s^{\mu,\nu} = \frac{\langle \sum_{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}} (k^{\mu} - m^{\mu}) \beta_{k,m} i_{\mathbf{m}|\mathbf{k}}^{\nu} \rangle_{\mathbf{k}|\nu}}{\langle \sum_{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}} (k^{\mu} - m^{\mu}) i_{\mathbf{m}|\mathbf{k}}^{\nu} \rangle_{\mathbf{k}|\nu}}, \quad (\text{B.13b})$$

$$\theta_i^{\mu,\nu} = \frac{\langle \sum_{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}} m^{\mu} \alpha_{k,m} s_{\mathbf{m}|\mathbf{k}}^{\nu} \rangle_{\mathbf{k}|\nu}}{\langle \sum_{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}} m^{\mu} s_{\mathbf{m}|\mathbf{k}}^{\nu} \rangle_{\mathbf{k}|\nu}}, \quad (\text{B.13c})$$

$$\phi_i^{\mu,\nu} = \frac{\langle \sum_{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}} m^{\mu} \beta_{k,m} i_{\mathbf{m}|\mathbf{k}}^{\nu} \rangle_{\mathbf{k}|\nu}}{\langle \sum_{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}} m^{\mu} i_{\mathbf{m}|\mathbf{k}}^{\nu} \rangle_{\mathbf{k}|\nu}}. \quad (\text{B.13d})$$

B.2.2 Approximation par pairs typée (TPA)

L'approximation par pairs permet de définir de nouvelles équations d'évolution tenant compte de la structure typée selon

$$\frac{d}{dt} s_{\mathbf{k}}^{\mu} = \sum_{\{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}\}} \eta_{\mathbf{k},\mathbf{m}}^{\mu} s_{\mathbf{k}}^{\mu}, \quad (\text{B.14a})$$

$$\frac{d}{dt} p_{\mathbf{k}}^{\mu,\nu} = \sum_{\{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}\}} \left(\frac{m^{\nu}}{k^{\nu}} - p_{\mathbf{k}}^{\mu,\nu} \right) \eta_{\mathbf{k},\mathbf{m}}^{\mu} s_{\mathbf{k}}^{\mu} + \theta_s^{\mu,\nu} (1 - p_{\mathbf{k}}^{\mu,\nu}) - \phi_s^{\mu,\nu} p_{\mathbf{k}}^{\mu,\nu}, \quad (\text{B.14b})$$

$$\frac{d}{dt} q_{\mathbf{k}}^{\mu,\nu} = \sum_{\{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}\}} \left(\frac{m^{\nu}}{k^{\nu}} - q_{\mathbf{k}}^{\mu,\nu} \right) \xi_{\mathbf{k},\mathbf{m}}^{\mu} (1 - s_{\mathbf{k}}^{\mu}) + \theta_i^{\mu,\nu} (1 - q_{\mathbf{k}}^{\mu,\nu}) - \phi_i^{\mu,\nu} q_{\mathbf{k}}^{\mu,\nu}, \quad (\text{B.14c})$$

avec les paramètres

$$\eta_{\mathbf{k},\mathbf{m}}^{\mu} = -\alpha_{k,m} \prod_{\nu \in \mathcal{B}} B_{m^{\nu}|k^{\nu}}(p_{\mathbf{k}}^{\mu,\nu}) + \left(\frac{1 - s_{\mathbf{k}}^{\mu}}{s_{\mathbf{k}}^{\mu}} \right) \beta_{k,m} \prod_{\nu \in \mathcal{B}} B_{m^{\nu}|k^{\nu}}(q_{\mathbf{k}}^{\mu,\nu}), \quad (\text{B.15a})$$

$$\xi_{\mathbf{k},\mathbf{m}}^{\mu} = -\beta_{k,m} \prod_{\nu \in \mathcal{B}} B_{m^{\nu}|k^{\nu}}(q_{\mathbf{k}}^{\mu,\nu}) + \left(\frac{s_{\mathbf{k}}^{\mu}}{1 - s_{\mathbf{k}}^{\mu}} \right) \alpha_{k,m} \prod_{\nu \in \mathcal{B}} B_{m^{\nu}|k^{\nu}}(p_{\mathbf{k}}^{\mu,\nu}), \quad (\text{B.15b})$$

et les taux de transition d'état

$$\theta_s^{\mu,\nu} = \frac{\langle \sum_{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}} (k^{\mu} - m^{\mu}) \alpha_{k,m} s_{\mathbf{k}}^{\nu} \prod_{\nu'} B_{m^{\nu'}|k^{\nu'}}(p_{\mathbf{k}}^{\nu,\nu'}) \rangle_{\mathbf{k}|\nu}}{\langle k^{\mu} s_{\mathbf{k}}^{\nu} (1 - p_{\mathbf{k}}^{\mu,\nu}) \rangle_{\mathbf{k}|\nu}}, \quad (\text{B.16a})$$

$$\phi_s^{\mu,\nu} = \frac{\langle \sum_{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}} (k^{\mu} - m^{\mu}) \beta_{k,m} (1 - s_{\mathbf{k}}^{\nu}) \prod_{\nu'} B_{m^{\nu'}|k^{\nu'}}(q_{\mathbf{k}}^{\nu,\nu'}) \rangle_{\mathbf{k}|\nu}}{\langle k^{\mu} (1 - s_{\mathbf{k}}^{\nu}) (1 - q_{\mathbf{k}}^{\mu,\nu}) \rangle_{\mathbf{k}|\nu}}, \quad (\text{B.16b})$$

$$\theta_i^{\mu,\nu} = \frac{\langle \sum_{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}} m^\mu \alpha_{k,m} s_{\mathbf{k}}^\nu \prod_{\nu'} B_{m^{\nu'}|k^{\nu'}}(p_{\mathbf{k}}^{\nu',\nu'}) \rangle_{\mathbf{k}|\nu}}{\langle k^\mu s_{\mathbf{k}}^\nu p_{\mathbf{k}}^{\mu,\nu} \rangle_{\mathbf{k}|\nu}}, \quad (\text{B.16c})$$

$$\phi_i^{\mu,\nu} = \frac{\langle \sum_{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}} m^\mu \beta_{k,m} (1 - s_{\mathbf{k}}^\nu) \prod_{\nu'} B_{m^{\nu'}|k^{\nu'}}(q_{\mathbf{k}}^{\nu',\nu'}) \rangle_{\mathbf{k}|\nu}}{\langle k^\mu (1 - s_{\mathbf{k}}^\nu) q_{\mathbf{k}}^{\mu,\nu} \rangle_{\mathbf{k}|\nu}}. \quad (\text{B.16d})$$

B.2.3 Approximation par champ-moyen typée (TMF)

L'approximation par champ-moyen permet la simplification des probabilités telles que

$$p_{\mathbf{k}}^{\mu,\nu} = q_{\mathbf{k}}^{\mu,\nu} = \omega^{\mu,\nu}, \quad (\text{B.17})$$

où la probabilité qu'un lien aléatoire soit connecté à au moins un noeud infecté est

$$\omega^{\mu,\nu} = \frac{\langle k^\mu (1 - s_{\mathbf{k}}^\nu) \rangle_{\mathbf{k}|\nu}}{\langle k^\mu \rangle_{\mathbf{k}|\nu}}. \quad (\text{B.18})$$

Les équations d'évolution deviennent

$$\frac{d}{dt} s_{\mathbf{k}}^\mu = \sum_{\{\mathbf{0} \leq \mathbf{m} \leq \mathbf{k}\}} \eta_{\mathbf{k},\mathbf{m}}^\mu s_{\mathbf{k}}^\mu, \quad (\text{B.19})$$

avec le paramètre

$$\eta_{\mathbf{k},\mathbf{m}}^\mu = -\alpha_{k,m} \prod_{\nu \in \mathcal{B}} B_{m^{\nu}|k^{\nu}}(\omega^{\mu,\nu}) + \left(\frac{1 - s_{\mathbf{k}}^\mu}{s_{\mathbf{k}}^\mu} \right) \beta_{k,m} \prod_{\nu \in \mathcal{B}} B_{m^{\nu}|k^{\nu}}(\omega^{\mu,\nu}). \quad (\text{B.20})$$

B.3 Validation numérique de l'approche généralisée par types de noeuds

La structure mésoscopique, qui encapsule les modèles de connectivité émergeant à l'échelle des groupes, joue un rôle fondamental dans la diffusion de l'information sur les réseaux complexes. Par exemple, il a été démontré qu'elle régissait en partie le phénomène de localisation dans les processus de propagation, une prise de conscience à l'origine de la conception de meilleures mesures de confinement [100]. La capacité de mémoire du cerveau a également été liée à la modularité optimale de l'organisation à méso-échelle des connectomes [101, 102]. Cependant, les modèles mathématiques derrière ces résultats sont généralement spécifiquement adaptés au problème à résoudre, ce qui limite leur portée et nécessite des formalismes plus généralisés.

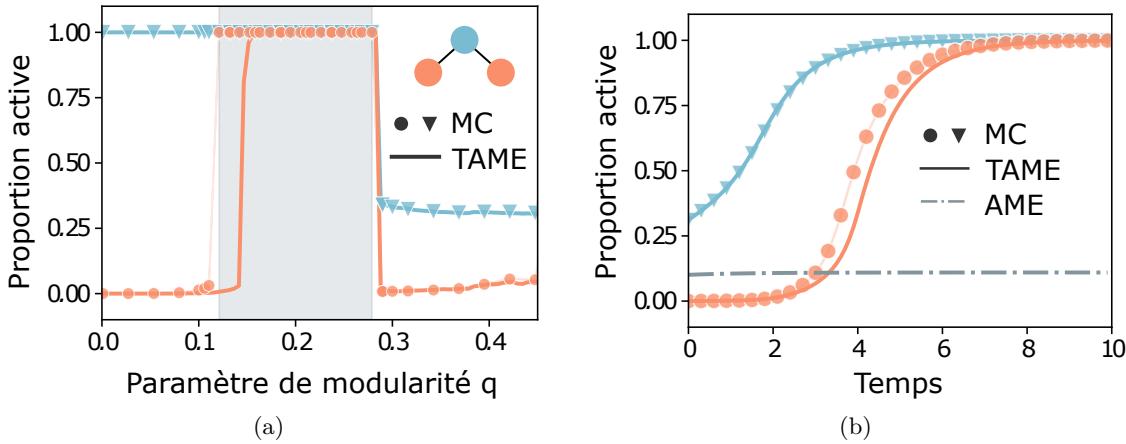


FIGURE B.3 – (a) Diagramme de phase de la dynamique de seuil avec $\theta = 0,3$, sur des graphes modulaires réguliers de taille $N = 10^5$ avec $M = 10^6$ arêtes et trois groupes de deux types (orange et bleu), en moyenne sur 50 réalisations. Les communautés sont connectées selon le graphe de communauté au-dessus de la légende – dans une chaîne sans arête reliant les communautés orange –, dans lequel les arêtes internes existent proportionnellement à $1 - q$ et les arêtes externes à q . Le paramètre $q \in [0; 0,5]$ contrôle la modularité et le degré dans chaque groupe est contraint par q et M . Initialement, seuls les noeuds du groupe bleu peuvent être actifs de telle sorte qu'une fraction ρ des noeuds du graphe soit actif. La zone ombragée désigne une région de modularité optimale semblable à celle de Nematzadeh et al. [101]. (b) Évolution temporelle de la dynamique de seuil prédictive par TAME. Les détails techniques sont les mêmes que ceux du panneau (a), sauf que le paramètre de modularité est fixé à $q = 0,2$. De plus, nous montrons la prédiction du modèle AME de Gleeson [98, 99].

Ce projet de recherche est une généralisation directe du cadre d'équations maîtresses approximées (AME) de Gleeson [98, 99], mais où la structure du réseau sous-jacent est spécifiée par un modèle des configurations avec des types de noeuds [103]. Le formalisme développé est utilisé pour étudier l'optimalité de la diffusion de l'information pour la dynamique de seuil sur des graphes modulaires avec différentes structures mésoscopiques. Nous montrons que

TAME reproduit avec précision la région de modularité optimale de Nematzadeh et al. [101], comme le montre la figure B.3 (a), et l'évolution temporelle de la fraction de noeuds actifs dans chaque groupe, comme le montre la figure B.3 (b). Enfin, nous montrons comment la généralité de notre approche étend naturellement le concept de modularité optimale à des structures mésoscopiques plus riches, ainsi qu'à d'autres dynamiques binaires, y compris la contagion complexe et les modèles de consensus.