



La résilience des réseaux complexes

Thèse

Edward Laurence

Doctorat en physique
Philosophiæ doctor (Ph. D.)

Québec, Canada

© Edward Laurence, 2020

La résilience des réseaux complexes

Thèse

Edward Laurence

Sous la direction de:

Louis J. Dubé
Patrick Desrosiers

Résumé

Les systèmes réels subissant des perturbations par l'interaction avec leur environnement sont susceptibles d'être entraînés vers des transitions irréversibles de leur principal état d'activité. Avec la croissance de l'empreinte humaine mondiale sur les écosystèmes, la caractérisation de la résilience de ces systèmes complexes est un enjeu majeur du 21e siècle.

Cette thèse s'intéresse aux systèmes complexes pour lesquels il existe un réseau d'interactions et où les composantes sont des variables dynamiques. L'étude de leur résilience exige la description de leurs états dynamiques qui peuvent avoir jusqu'à plusieurs milliers de dimensions. Cette thèse propose trois nouvelles méthodes permettant de faire des mesures de la dynamique en fonction de la structure du réseau. L'originalité de ce travail vient de la diversité des approches présentées pour traiter la résilience, en débutant avec des outils basés sur des modèles dynamiques définis et en terminant avec d'autres n'exploitant que des données récoltées.

D'abord, une solution exacte à une dynamique de cascade (modèle de feu de forêt) est développée et accompagnée d'un algorithme optimisé. Comme sa portée pratique s'arrête aux petits réseaux, cette méthode signale les limitations d'une approche avec un grand nombre de dimensions. Ensuite, une méthode de réduction dimensionnelle est introduite pour établir les bifurcations dynamiques d'un système. Cette contribution renforce les fondements théoriques et élargit le domaine d'applications de méthodes existantes. Enfin, le problème de retracer l'origine structurelle d'une perturbation est traité au moyen de l'apprentissage automatique. La validité de l'outil est supportée par une analyse numérique sur des dynamiques de propagation, de populations d'espèces et de neurones. Les principaux résultats indiquent que de fines anomalies observées dans la dynamique d'un système peuvent être détectées et suffisent pour retracer la cause de la perturbation. L'analyse témoigne également du rôle que l'apprentissage automatique pourrait jouer dans l'étude de la résilience de systèmes réels.

Abstract

Real complex systems are often driven by external perturbations toward irreversible transitions of their dynamical state. With the rise of the human footprint on ecosystems, these perturbations will likely become more persistent so that characterizing resilience of complex systems has become a major issue of the 21st century.

This thesis focuses on complex systems that exhibit networked interactions where the components present dynamical states. Studying the resilience of these networks demands depicting their dynamical portraits which may feature thousands of dimensions. In this thesis, three contrasting methods are introduced for studying the dynamical properties as a function of the network structure. Apart from the methods themselves, the originality of the thesis lies in the wide vision of resilience analysis, opening with model-based approaches and concluding with data-driven tools.

We begin by developing an exact solution to binary cascades on networks (forest fire type) and follow with an optimized algorithm. Because its practical range is restricted to small networks, this method highlights the limitations of using model-based and highly dimensional tools. We then introduce a dimension reduction method to predict dynamical bifurcations of networked systems. This contribution builds up on theoretical foundations and expands possible applications of existing frameworks. Finally, we examine the task of extracting the structural causes of perturbations using machine learning. The validity of the developed tool is supported by an extended numerical analysis of spreading, population, and neural dynamics. The results indicate that subtle dynamical anomalies may suffice to infer the causes of perturbations. It also shows the leading role that machine learning may have to play in the future of resilience of real complex systems.

Table des matières

Résumé	iii
Abstract	iv
Table des matières	v
Liste des tableaux	viii
Liste des figures	ix
Liste d'abréviations et de notations	xi
Remerciements	xiii
Liste des contributions	xv
Avant-propos	xvi
Introduction	1
1 Théorie de la résilience des systèmes complexes	4
1.1 La structure des réseaux complexes	5
1.1.1 Modèles génératifs	6
1.2 La dynamique sur réseaux	10
1.2.1 Cascade sur réseaux	10
1.2.2 Dynamique SIS	14
1.2.3 Dynamique de Wilson-Cowan	16
1.2.4 Modèle de populations	19
1.3 La résilience des systèmes complexes	21
1.3.1 Détecter et comprendre les perturbations	23
1.4 Méthodes de prédiction	25
1.4.1 Modèle dynamique et structure connus	26
1.4.2 Modèle dynamique et structure partiellement connus	30
1.4.3 Modèle dynamique et structure inconnus	37
1.4.4 Modèle dynamique inconnu et structure connue	40
1.5 Projet de recherche	44
1.5.1 Problématique	44
1.5.2 Objectifs et guide de lecture	44

2	Solution exacte de la dynamique de cascade sur réseaux complexes	47
2.1	Avant-propos	48
2.2	Résumé	48
2.3	Abstract	48
2.4	Introduction	49
2.5	Cascade dynamics on arbitrary graphs	50
2.5.1	Definition of the dynamics	50
2.5.2	Special classes of response functions	51
2.5.3	Exact recursive solution	52
2.6	Enumeration algorithms	54
2.6.1	General recursive solution method	54
2.6.2	Saving time : culling impossible configurations	54
2.6.3	Saving more time : dropping intermediary distributions	56
2.7	Results and applications	57
2.7.1	Calibration : special cases on a directed graph	57
2.7.2	Mixed dynamics	59
2.8	Discussion	61
2.9	Acknowledgements	62
2.10	Supplementary Material	63
2.10.1	Equivalence with bond percolation	63
2.10.2	Explicit example on a tadpole graph	63
2.10.3	Complexity calculation	65
3	Réduction dimensionnelle spectrale des systèmes complexes dynamiques	67
3.1	Avant-propos	68
3.2	Résumé	68
3.3	Abstract	68
3.4	Introduction	69
3.5	Model definition	71
3.5.1	General formalism	71
3.5.2	Examples of possible dynamics	72
3.6	1–dimensional reduction	73
3.6.1	Derivation of the reduction formalism	75
3.6.2	Choice of a universal weight vector	76
3.6.3	Choice of an approximate weight vector	78
3.7	Multidimensional reduction : dominant eigenvectors	81
3.7.1	Cycle reduction	81
3.7.2	Choice of the universal weight vectors	82
3.7.3	Examples : Star and bipartite networks	84
3.8	Multidimensional reduction : including subdominant eigenvectors	86
3.9	Goodness of reduction	88
3.9.1	Impact of the structure	89
3.9.2	Detection of transitions	90
3.9.3	Impact of the dynamics	92
3.10	Conclusion	95
3.11	Acknowledgements	96
3.12	Supplementary Material	97
3.12.1	Derivation of 1–dimensional formalism	97

3.12.2	Derivation of the multidimensional <i>cycle</i> formalism	98
3.12.3	Combination for 2-cycle reduction	99
4	Détection des perturbations structurelles par apprentissage automatique	101
4.1	Avant-propos	102
4.2	Résumé	102
4.3	Abstract	102
4.4	Introduction	103
4.5	Context	105
4.5.1	Graph Neural Networks	105
4.6	Results	106
4.6.1	Spreading dynamics	107
4.6.2	Population dynamics	111
4.6.3	Neural dynamics	112
4.7	Discussion	112
4.8	Materials and Methods	113
4.9	Acknowledgements	118
4.10	Supplementary Material	119
4.10.1	GNN Model	119
4.10.2	Weight decay	122
4.11	Large perturbations and loopy scale-free networks	122
4.12	Challenging edges in scale-free networks	123
4.13	Length of the perturbed time series	124
4.13.1	Likelihood of the SIS model	124
4.14	Likelihood of the SIS model	124
4.15	Ecological networks	126
	Conclusion et perspectives	128
	A Bifurcation du modèle de Wilson-Cowan	131
	B Deux généralisations de la réduction dimensionnelle à plus hautes di-	
	mensions	134
B.1	Contexte	135
B.1.1	Méthode sans contrainte	135
B.1.2	Méthode par approximation d'orthogonalité ou par partition	137
B.2	Système à double coeur	140
B.3	Simulations sur un système réel	142
B.3.1	Dynamique mutualiste	143
	Bibliographie	147

Liste des tableaux

1.1	Exemples de définitions de la résilience selon les domaines scientifiques.	23
4.1	Layer by layer description of the GNN models for each dynamics.	121
4.2	Summary properties of ecological networks	127

Liste des figures

1.1	Exemples de réseaux invariants d'échelle	10
1.2	Solution de la dynamique de cascade.	14
1.3	Transition de phase pour le modèle SIS	17
1.4	Diagramme de stabilité et bifurcation	19
1.5	Temps d'intégration de la dynamique Lotka-Volterra généralisée pour différentes tailles de réseau	25
1.6	Exemples des bifurcations pour différentes stratégies d'attaque sur une dynamique SIS	33
1.7	Exemple de stabilité de la dynamique réduite pour des contextes mutualiste, compétitif et mixte.	35
1.8	Résilience de réseaux réels	37
1.9	Exemple d'estimation de paramètres dynamiques pour une dynamique SIS.	42
2.1	Example of a breath-first exploration of all configurations.	55
2.2	Size distribution and mean size of active component for different dynamics.	56
2.3	Individual response function and activation probabilities.	59
2.4	Complete distribution \mathcal{Q} for the every possible configurations of the network.	60
2.5	Worst-case complexity of a brute-force solution of the exact cascade dynamics.	66
3.1	Observable at equilibrium as a function of the dominant eigenvalue for different dynamics.	74
3.2	Comparison of the observable as a function of the dominant eigenvalues for different Erdős-Rényi networks.	79
3.3	Comparison of different reduction methods on a star network.	80
3.4	Comparison of different reduction methods on a planted partition network.	85
3.5	Spectral density for random and SBM networks.	86
3.6	Network instances produced using the generalized preferential attachment model.	88
3.7	Total error on the prediction of the global state activity at equilibrium for networks obtained from the generalized preferential attachment model	91
3.8	Comparison of different reduction methods on a network instance of the generalized preferential attachment model with $\nu = 1.8$ and $N = 60$	92
4.1	Examples of GNN predictions and time series for epidemic spreading on a scale-free network.	104
4.2	Area under the curve for SIS dynamics on random networks.	108
4.3	Area under the curve of model predictions of perturbation on epidemics on scale-free networks.	110
4.4	Average area under the curve of predictions on 25 real ecological networks with prey-predator dynamics.	111

4.5	Area under the curve of predictions over neural dynamics on the C. Elegans connectome.	112
4.6	Example of inference on the SIS dynamics over a scale-free networks with 5 removed edges.	123
4.7	AUC on various schemes with SIS dynamics.	124
B.1	Comparaisons des méthodes de réduction sur le réseau à double coeur.	141
B.2	Spectre du réseau de Barrett, <i>et al.</i> [16]	142
B.3	Schématisation du réseau de Barrett, <i>et al.</i> [16]	143
B.4	Comparaison du ralentissement critique et la réduction dimensionnelle pour la prédiction des points critiques sur un réseau réel.	144

Liste d'abréviations et de notations

Abréviations

PA	Modèle d'attachement préférentiel (<i>preferential attachment</i>)
SIS	Dynamique de susceptible-infecté-susceptible (<i>Susceptible-infected-Susceptible</i>)
MCMC	Méthode de Monte-Carlo par chaînes de Markov (<i>Monte Carlo Markov Chain</i>)
EWS	Signaux précurseurs (<i>Early-warning signals</i>)
CSD	Ralentissement critique (<i>Critical slowing down</i>)

Symboles

N	Nombre de noeuds.
M	Nombre de liens.
\mathbf{A}	Matrice d'adjacence binaire.
\mathbf{W}	Matrice des poids.
\mathbf{k}	Vecteur des degrés $k_i = \sum_{j=1}^N a_{ij}$.
$\langle k \rangle_{\text{NN}}$	Degré moyen des voisins d'un noeud aléatoire. Pour un réseau donné, alors $\langle k \rangle_{\text{NN}} = \langle k \rangle^2 / \langle k \rangle$. Le double indice N est inspiré de la notation anglais <i>nearest neighbors</i>
\mathbf{k}^{in}	Vecteur des degrés entrants $k_i^{\text{in}} = \sum_{j=1}^N a_{ij}$.
\mathbf{k}^{out}	Vecteur des degrés sortants $k_i^{\text{out}} = \sum_{j=1}^N a_{ji}$.
\mathbf{s}	Vecteur des forces $s_i = \sum_{j=1}^N w_{ij}$.
\mathbf{s}^{in}	Vecteur des forces entrantes $s_i^{\text{in}} = \sum_{j=1}^N w_{ij}$.
\mathbf{s}^{out}	Vecteur des forces sortantes $s_i^{\text{out}} = \sum_{j=1}^N w_{ji}$.
$\text{Pr}(k)$	Distribution des degrés évaluée à k .
$\text{Pr}(k' k)$	Distribution corrélée des degrés. Probabilité qu'un noeud sachant de degré k soit connecté à un noeud de degré k' .

...

Notations générales

\mathbf{X}	Matrice d'éléments x_{ij} . Toute lettre majuscule en gras est une matrice.
\mathbf{x}	Vecteur colonne. Toute lettre minuscule en gras est un vecteur.
\mathbf{x}^T	Vecteur rangée.
$\mathbf{1}$	Vecteur de 1.
\mathbf{I}_N	Matrice identité de taille $N \times N$.
$\mathbf{A} \geq 0$	Tous les éléments de \mathbf{A} sont non négatifs. S'applique aussi aux vecteurs.
$\mathbf{A} < 0$	Tous les éléments de \mathbf{A} sont négatifs. S'applique aussi aux vecteurs.
$\text{eig}[\mathbf{A}]$	Ensemble des valeurs propres de la matrice \mathbf{A} .
$\mathbf{D} = \text{diag}(\mathbf{x})$	Matrice dont les éléments non diagonaux sont nuls et dont les éléments diagonaux sont $[\mathbf{D}]_{ii} = x_i$. Le vecteur \mathbf{x} peut être un vecteur colonne ou rangée.
$\mathbf{D} = \text{diag}(\mathbf{X})$	Matrice dont les éléments non diagonaux sont nuls et dont les éléments diagonaux sont $[\mathbf{D}]_{ii} = [\mathbf{X}]_{ii}$. La matrice \mathbf{X} doit être carrée.
$\mathbf{X}' = \max(c, \mathbf{X})$	Maximum par élément, i.e., $[\mathbf{X}']_{ij} = [\mathbf{X}]_{ij}$ si $[\mathbf{X}]_{ij} > c$ et $[\mathbf{X}']_{ij} = c$ autrement.
δ_{ij}	Delta de Kronecker. $\delta_{ij} = 1$ si $i = j$, et $\delta_{ij} = 0$ autrement.

Remerciements

Je tiens à remercier chaleureusement mon directeur Louis J. Dubé. Votre attitude envers la recherche est inspirante et votre rigueur au travail est exemplaire, bien qu'elle n'ait jamais percolé chez moi autant que l'on aurait voulu. Je garde de joyeux souvenirs de vos partages autour d'un café, que ce soit sur les fonctions Mittag-Leffler, l'exactitude des ratios des masses élémentaires ou les traditions de l'époque romaine. Vous êtes un professeur comme on n'en voit plus.

Je veux remercier mon codirecteur Patrick. Tu as beaucoup contribué à définir mes intérêts scientifiques et à améliorer ma méthode de travail. Avec une grande délicatesse et énormément de patience, tu as toujours clairement communiqué tes exigences tout en me laissant diriger mes recherches et en me guidant dans mes apprentissages. Merci de donner aussi généreusement de ton temps aux étudiants et de t'impliquer dans leurs projets.

Je suis très reconnaissant à mes examinateurs : Antoine Allard, Audrey Durant et Pierre-André Noël, pour vos sages conseils et d'avoir pris le temps de lire cette thèse. Je remercie aussi Nicolas Doyon, un précieux collaborateur avec un enthousiasme contagieux pour la recherche.

Un merci du fond du coeur à tous les membres de Dynamica. Particulièrement à mon senior *préféré* Jean-Gabriel, qui a adroitement joué le rôle de grand-frère scientifique. Ta vision parfois cynique, mais toujours puriste et ton approche audacieuse vis-à-vis de la science sont communicatives, surtout avec l'emploi de l'adjectif [voix rocailleuse] *convolué*. Je n'oublie pas non plus les autres seniors : Antoine, Laurent, Jean-Luc, Denis et Joey, qui sont des exemples inspirants de réussite. À mes acolytes Guillaume, Charles, Vincent et Xavier, cela a été un plaisir d'échanger des idées, d'apprendre à bencher *sniff*, de vous éduquer en design et de partager des séjours à Charlevoix. Je remercie aussi Mohammed et Béatrice pour les agréables discussions auxquelles vous avez pris part.

Mon parcours au doctorat a été marqué par mon travail chez Hectiq.ai. Je remercie sincèrement Martin pour sa confiance et pour l'opportunité qu'il m'a offerte. Tu as toujours facilité la conciliation travail-étude et entretenu un climat de travail stimulant que plusieurs compagnies peuvent envier. Je garde d'excellents souvenirs de ces demi-journées ponctuées de cafés trop corsés et de sessions au tableau blanc. C'est un réel plaisir de travailler avec toi.

Un merci tout spécial à ma famille et mes quelques ami(e)s proches. Bien que vous n'ayez pas eu l'occasion d'assister à la soutenance pour découvrir ce qui m'a occupé ces dernières années, ce sont vos encouragements discrets (et les quelques veaux *parmigiana* entre *Terrible Four*) qui m'ont motivé à décrocher le doctorat et à vous rendre fiers. Je dédie les derniers mots à ma très chère Marie-Ève. C'est dans ton inébranlable confiance que je terminerais ma thèse et les réussites que tu accumules que j'ai pu trouver la motivation de terminer mes études. Tu as fait preuve d'une grande délicatesse dans mes moments d'angoisses et d'une écoute attentive qui t'est naturelle. Merci énormément d'être là et de veiller sur moi.

Liste des contributions

Les contributions suivantes ont été faites lors de mon doctorat.

- [Detecting structural perturbations from time series using deep learning](#)
E. Laurence, C. Murphy, G. St-Onge, X. Roy-Pomerleau et V. Thibeault
En préparation
- [Phase transition in the recoverability of network history](#)
J.-G. Young, G. St-Onge, **E. Laurence**, C. Murphy, L. Hébert-Dufresne et P. Desrosiers
Phys. Rev. X, 9, 041056 (2019)
- [Spectral Dimension Reduction of Complex Dynamical Networks](#)
E. Laurence, N. Doyon, L. J. Dubé et P. Desrosiers
Phys. Rev. X, 9, 011042 (2019)
- [Exact analytical solution of irreversible binary dynamics on networks](#)
E. Laurence, J.-G. Young, S. Melnik et L. J. Dubé
Phys. Rev. E, 97, 032302 (2018)
- [Geometric evolution of complex networks](#)
C. Murphy, A. Allard, **E. Laurence**, G. St-Onge et L. J. Dubé
Phys. Rev. E, 97, 032309 (2018)
- [Phase transition of the susceptible-infected-susceptible dynamics on time-varying configuration model networks](#)
G. St-Onge, J.-G. Young, **E. Laurence**, C. Murphy et L. J. Dubé
Phys. Rev. E, 97, 022305 (2018)
- [Finite size analysis of the detectability limit of the stochastic block model](#)
J.-G. Young, P. Desrosiers, L. Hébert-Dufresne, **E. Laurence** et L. J. Dubé
Phys. Rev. E, 95, 062304 (2016)

Avant-propos

Cette thèse comporte quatre chapitres dont trois sont des contributions publiées ou soumises à des journaux scientifiques. Une seule modification a été apportée au contenu de ces dernières : au Chapitre 3, le terme *Cowan-Wilson* a été remplacé par *Wilson-Cowan*. De plus, le format a été modifié pour satisfaire les normes de la Faculté des études supérieures et postdoctorales de l'Université Laval.

Chapitre 2

— [Exact analytical solution of irreversible binary dynamics on networks](#)

E. Laurence, J.-G. Young, S. Melnik et L. J. Dubé

Phys. Rev. E, 97, 032302

Publié le 2 mars 2018.

Contribution

Tous les auteurs ont élaboré le plan de recherche. J'ai obtenu les principaux résultats analytiques originaux. Jean-Gabriel Young et moi avons programmé les solutions numériques et avons rédigé l'essentiel de l'article. Tous les auteurs ont contribué à la révision de l'article.

Chapitre 3

— [Spectral Dimension Reduction of Complex Dynamical Networks](#)

E. Laurence, N. Doyon, L. J. Dubé et P. Desrosiers

Phys. Rev. X, 9, 011042

Publié le 4 mars 2019.

Contribution

Tous les auteurs ont élaboré le plan de recherche et à l'obtention des résultats analytiques originaux. J'ai programmé les solutions numériques et j'ai rédigé l'essentiel de l'article. Tous les auteurs ont contribué à la révision de l'article.

Chapitre 4

Cette publication est écrite pour un journal acceptant les textes courts. L'ordre de présentation est différent des précédents chapitres puisque les méthodes sont placées à la fin de l'article.

— [Detecting structural perturbations from time series using deep learning](#)

E. Laurence, C. Murphy, G. St-Onge, X. Roy-Pomerleau et V. Thibeault

En préparation

Contribution

Tous les auteurs ont élaboré le plan de recherche et ont contribué à l'obtention des résultats originaux. J'ai programmé les solutions numériques d'intelligence artificielle et j'ai rédigé l'essentiel de l'article. Tous les auteurs ont contribué à la révision de l'article.

Introduction

Bien souvent, certains phénomènes suscitent notre contemplation, car leur complexité nous semble insaisissable. Pensons par exemple à l'apparition soudaine de la congestion routière, la propagation en ligne de fausses rumeurs, ou encore les paniques dans les supermarchés à l'annonce d'une catastrophe. Même la vue d'un prédateur animal peut être suffisante pour nous rappeler l'existence de cet imperceptible et fragile équilibre des écosystèmes. L'explication de ces observations va au-delà de notre connaissance des véhicules, des humains et des animaux. Elle nécessite de prendre en compte les interactions entre les composantes de ces systèmes. Par exemple, la propagation d'un virus dans une communauté ne se prédit normalement pas avec une description médicale détaillée du virus, mais principalement avec une modélisation des contacts à l'intérieur de la population. Dans ce cas-ci, le système est dit *complexe*, car des phénomènes émergent de l'interaction entre plusieurs individus.

Les *réseaux complexes* sont les systèmes dont on peut extraire un graphe des interactions. Dans les dernières années, les réseaux complexes sont aussi devenus un champ d'intérêt scientifique qui a gagné en popularité et a permis une foule d'applications en épidémiologie, neurosciences, écologie, génétique, climatologie, économie et transport [38, 49]. Ces accomplissements sont le fruit de notre capacité croissante à traiter des systèmes de plus en plus compliqués, et ce, de manière plus précise. Cette progression de la science des réseaux est faite de manière incrémentielle : les contributions fondatrices introduisent des modèles très simples et progressivement les nouvelles contributions augmentent la complexité des modèles. Cette thèse n'en fait pas exception.

Plusieurs systèmes réels sont soumis à des perturbations. L'exemple le plus notable est sans doute les écosystèmes qui doivent opérer en présence d'interventions humaines et d'un climat changeant. Ces dérangements mènent parfois à un déséquilibre qui pousse le système vers un nouvel état dynamique. Les manifestations de ces épisodes incluent l'extinction soudaine des truites grises du lac Michigan [105], la transition de la végétation du Sahara à un désert [43, 147] ou encore la dégradation de la barrière de corail des Caraïbes [131]. Bien que les changements soient draconiens, la perturbation appliquée ne l'est pas forcément, un constat bien connu de la théorie du chaos. Comme la dynamique sur un réseau se comporte généralement comme un système non linéaire, l'effet d'une perturbation peut rapidement être amplifié, ou atténué,

par les individus. C'est cet effet couplé qui rend la prédiction des états stables si difficile à établir. Néanmoins, le besoin d'outils pour caractériser la dynamique est capital pour prédire les catastrophes et bâtir des stratégies de gestion durable, notamment en cette époque où les changements climatiques ont des effets sur les écosystèmes.

L'étude de la résilience d'un système dynamique passe avant tout par la prédiction de ses états stables. Que ce soit pour estimer la perturbation maximale qui peut être supportée ou encore de déterminer si une catastrophe est sur le point d'arriver, il faut d'abord décrire les états de la dynamique en fonction des paramètres et du couplage entre les composantes. Toutefois, l'absence de données précises pour valider les conclusions et la haute dimensionnalité des systèmes suggèrent de simplifier le cadre mathématique en ayant recours à des modèles simples. Cela mène en général à des résultats clairs comme démontrer mathématiquement l'existence de phénomènes collectifs ou de tracer le portrait de bifurcation en faibles dimensions. Toutefois, cette direction doit être empruntée avec vigilance. La validité des conclusions ne s'étend pas forcément aux systèmes réels. Plus encore, la précision de ces modèles est inférieure à celle des modèles plus compliqués et moins interprétables. Cet équilibre entre la précision et le pouvoir explicatif des modèles joue un rôle particulièrement important et sera approfondi dans cette thèse.

Le thème d'intérêt de cette thèse est la *résilience des réseaux complexes*. Nous nous intéressons donc à faire la prédiction des états dynamiques dans le cadre de la science des réseaux. Comme mentionné précédemment, le choix entre l'utilisation de modèles simples ou compliqués s'offre à nous. Lorsqu'il s'agit de réseaux complexes, il est surprenant qu'il soit plus difficile de construire des modèles simples que compliqués. En effet, nous ne pouvons facilement ignorer le graphe des interactions qui joue un rôle central. Par conséquent, nous croisons rarement des modèles simples de dynamique qui tiennent correctement en compte une structure effective de faible dimension.

Les objectifs de cette thèse sont de développer des supports mathématiques pour quantifier la résilience des réseaux complexes et identifier des perturbations structurelles. Le premier objectif sera poursuivi en deux volets. Le premier volet est l'obtention d'un modèle de haute dimensionnalité qui devra être en mesure d'accomplir des prédictions quantitatives précises des états dynamiques d'un réseau. Le deuxième volet viendra plutôt réduire la complexité des modèles et devra être d'un faible nombre de dimensions. Les prédictions seront donc plus interprétables et fourniront des indicateurs des bifurcations des dynamiques. Le deuxième objectif complètera le portrait de l'analyse de la résilience en développant un outil capable de détecter les perturbations structurelles qu'un réseau subit. Nous profiterons de cette tâche pour valider l'utilisation de l'apprentissage automatique sur graphes.

Au Chapitre 1, nous présenterons les bases de la science des réseaux. Nous verrons les mécanismes dynamiques ainsi que plusieurs méthodes pour faire des prédictions des états stables.

Nous passerons un certain temps à comparer des approches simples et complexes sur des problèmes équivalents. Dès le Chapitre 2, nous choisirons un problème précis pour développer une approche de hautes dimensions. Plus précisément, nous obtiendrons les probabilités de chaque état stationnaire pour une cascade dynamique sur un réseau quelconque. L'approche qui sera présentée a l'avantage de ne faire aucune approximation sur la structure en graphe, mais montre quelques lacunes en ce qui concerne les prédictions qualitatives.

Au Chapitre 3, nous élargirons le champ d'applications et considérerons une famille de processus dynamiques déterministes sur réseaux. Celle-ci inclut certaines dynamiques de populations, de diffusion et de neurones. L'objectif de ce chapitre est de construire un objet de faibles dimensions qui reproduit les états stationnaires globaux de la dynamique des réseaux. L'outil développé s'applique à une grande gamme de réseaux en plus de surpasser la précision des techniques existantes. Cette contribution permet entre autres de décrire les bifurcations du système et donc de prédire si une perturbation est suffisante pour faire basculer l'état dynamique. En surplus, deux généralisations possibles seront présentées à l'annexe B pour stimuler les travaux futurs.

Enfin, le Chapitre 4 construira sur nos acquis pour développer un moyen d'utiliser les états dynamiques du système afin d'inférer les perturbations que le système subit. Cette tâche sera réalisée en utilisant un modèle d'apprentissage automatique ayant un excellent pouvoir de prédiction et la présentation sera donc principalement de nature numérique. Nous verrons pourquoi l'apprentissage automatique sur graphe est une direction privilégiée pour les travaux demandant une haute précision quantitative.

Chapitre 1

Théorie de la résilience des systèmes complexes

La science des réseaux complexes est naturellement multidisciplinaire. Il est courant de rencontrer certains concepts empruntés à la biologie, à la physique ou bien à la sociologie. Ce chapitre vise à cibler et introduire les concepts nécessaires à l'appréciation des contributions originales de cette thèse.

La première section introduit la structure des réseaux complexes. Celle-ci couvrira d'abord les objets mathématiques de base qui décrivent la structure. Par la suite, nous verrons des modèles structurels permettant de générer des réseaux artificiels dont la structure est contrôlée. À la deuxième section, nous avancerons vers les dynamiques principales sur les réseaux. Avec une bonne description de chaque processus, nous pourrons progresser vers leurs résolutions au moyen de différentes techniques.

À la troisième section, nous aborderons directement le concept central de cette thèse qu'est la résilience des réseaux complexes. Nous détaillerons les problématiques associées. Nous continuerons avec les principales techniques qui s'offrent à nous pour étudier la résilience. Cette section est particulièrement importante pour être en mesure d'apprécier le travail des chapitres suivants.

Finalement, la dernière section est un guide de lecture de cette thèse. Nous ferons le lien entre les concepts déjà présentés et les contributions des prochains chapitres.

...

1.1 La structure des réseaux complexes

Un ensemble de composantes en interaction est un exemple d'un système complexe. Les molécules de l'air qui nous entoure forment un système complexe, et les mots de cette thèse en forment un également. Mais si nous pouvons argumenter que presque tous les systèmes réels sont complexes selon ce point de vue, seulement une fraction d'entre eux profitent d'une représentation par *réseaux complexes*.

La famille des *réseaux complexes* regroupe les systèmes dont les interactions prennent la forme d'un graphe. Par exemple, les neurones du cerveau communiquent via les synapses et forment un réseau de communication. À l'inverse, un gaz profite difficilement de la formulation par graphe puisque les interactions entre les atomes sont imprévisibles et désorganisées. Pour vérifier si un système jouit de la représentation en graphe, on évalue si les propriétés structurelles du graphe permettent de prédire l'émergence de phénomènes fonctionnels. Par exemple, la dureté du diamant est produite par la structure cristalline tétraédrique que forment ses atomes de carbone. Similairement, la structure cristalline en couches du graphite explique sa capacité à s'effriter. La représentation en graphe est donc pertinente pour décrire les cristaux de carbone.

Mathématiquement, un réseau complexe se compose d'un ensemble de N *noeuds* et de M *liens* qui relient les noeuds entre eux [124]. Deux noeuds peuvent interagir seulement s'ils partagent un lien¹. La structure du graphe est décrite par une matrice d'adjacence \mathbf{A} binaire de taille $N \times N$ dont l'élément a_{ij} est 1 s'il existe un lien du noeud j vers le noeud i et 0 autrement. Si le graphe est non dirigé alors \mathbf{A} est symétrique. On introduit également la matrice des poids (ou matrice d'adjacence pondérée) \mathbf{W} dont l'élément $w_{ij} \in \mathbb{R}$ indique le poids du lien du noeud j vers le noeud i . Dans cette thèse, les termes *graphe* et *réseau* sont utilisés comme synonymes.

Degré - Le degré du noeud i est son nombre de liens, soit

$$k_i = \sum_{j=1}^N a_{ij}. \quad (1.1)$$

Pour un graphe dirigé, on distinguera le *degré entrant*

$$k_i^{\text{in}} = \sum_{j=1}^N a_{ij} \quad (1.2)$$

et le *degré sortant*

$$k_i^{\text{out}} = \sum_{j=1}^N a_{ji} \quad (1.3)$$

1. Notons que cette interprétation n'est pas universelle. Par exemple, en apprentissage automatique sur réseaux complexes, il est commun d'utiliser une marche aléatoire sur le graphe pour définir le voisinage généralisé d'un noeud [175].

pour profiter de l'information contenue dans la direction des liens. On note que si le graphe est dirigé, alors le degré total réel du noeud est plutôt $k_i = k_i^{\text{in}} + k_i^{\text{out}}$ contrairement à l'équation (1.1).

Si le graphe est pondérée, on définit la force du noeud i par l'équation

$$s_i = \sum_{j=1}^N w_{ij} \quad (1.4)$$

qui peut être spécifiée entrante $s_i^{\text{in}} = \sum_{j=1}^N w_{ij}$ et sortante $s_i^{\text{out}} = \sum_{j=1}^N w_{ji}$.

À moins d'indications contraires, nous ferons l'hypothèse que les réseaux sont non dirigés et non pondérés.

Le degré des noeuds est souvent la première information qu'on extrait d'un réseau inconnu. D'une part, le degré est un bon indicateur de l'importance de chaque noeud [130], et d'autre part, les statistiques de la distribution des degrés donnent une idée globale de l'architecture du réseau [128]. Même le degré moyen,

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i, \quad (1.5)$$

est indicateur de la connectivité du réseau. Par exemple, si $\langle k \rangle < 2$, alors un noeud ayant un lien entrant n'a en moyenne pas de lien sortant. En suivant un lien au hasard, on ne peut donc pas naviguer sur le réseau : il est probablement non connexe [124].

1.1.1 Modèles génératifs

La carte des connexions d'un cerveau est pratiquement unique à chaque individu². Pourtant, l'humain dispose d'une faculté cognitive qui nous distingue de certains autres animaux. Les cerveaux humains doivent donc partager une structure équivalente, sans pour autant être identique. Cette hypothèse est supportée expérimentalement : nos cerveaux ont des propriétés structurelles communes comme l'existence d'une organisation hiérarchique et d'un rich-club [18, 19, 161].

Règle générale, il existe plusieurs graphes ayant des propriétés émergentes équivalentes. Autrement, le comportement stable des réseaux réels serait pratiquement impossible. Une simple modification d'un lien engagerait le système vers un nouvel état émergent. Plutôt, tout porte à croire qu'une organisation globale, ou une certaine statistique, est suffisante pour l'existence des propriétés émergentes. Par exemple, la capacité d'une personne à propager une rumeur

2. On entend ici que deux réseaux sont différents s'ils ne sont pas isomorphes, c.-à-d. deux graphes sont isomorphes s'il existe une bijection entre les identités des noeuds de deux graphes qui préserve la matrice d'adjacence [124].

dans un réseau social peut être estimée au moyen de son nombre d'amis, sans pour autant connaître l'identité exacte de ses amis.

Par conséquent, il apparaît naturel de regrouper les réseaux dans des familles d'équivalence basées sur certaines statistiques communes. Ces familles sont mieux définies en termes de modèles génératifs qui forment une portion importante de la littérature en science des réseaux. En plus de permettre la classification des systèmes, ils sont utilisés pour valider et inspecter les modèles prédictifs [14, 69, 137, 167] et servent même de distribution a priori à la tâche d'inférence structurelle [137]. Nous présenterons les modèles les plus populaires : le modèle aléatoire d'Erdős-Rényi, le modèle des configurations et d'attachement préférentiel.

Modèle de Erdős-Rényi

Le modèle aléatoire d'Erdős-Rényi, qu'on identifie par $G(N, p)$, produit des graphes non dirigés de N noeuds avec une probabilité de connexion p entre deux noeuds distincts [26, 47]. Le degré moyen attendu est $\langle k \rangle = p(N - 1)$. Comme chaque lien est une réalisation d'un processus de Bernoulli, la distribution binomiale des degrés tend vers une distribution de Poisson lorsque le nombre de noeuds est grand :

$$\Pr(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}, \quad (1.6)$$

où $\lambda = p(N - 1) \approx pN$.

Le modèle d'Erdős-Rényi reproduit une propriété parfois recherchée : le graphe est non assortatif, c.-à-d. les degrés des noeuds composant un lien ne sont pas corrélés. Le degré d'un noeud n'est donc pas prédictif du degré de ses voisins. Mathématiquement, cette propriété se traduit par $\Pr(k'|k) = \Pr(k')$ où $\Pr(k'|k)$ est la probabilité qu'un noeud que l'on sait de degré k soit connecté à un noeud de degré k' . Ceci implique que le degré moyen $\langle k \rangle$ est environ égal au degré moyen des voisins d'un noeud $\langle k \rangle_{\text{NN}}$ ³, soit

$$\langle k \rangle_{\text{NN}} = \sum_{k=1}^N \sum_{k'=1}^N k' \Pr(k) \Pr(k'|k) = \sum_{k=1}^N \sum_{k'=1}^N k' \Pr(k) \frac{k'}{\langle k \rangle} \Pr(k') = \frac{1}{\langle k \rangle} \sum_{k'=1}^N k'^2 \Pr(k') = \frac{\langle k^2 \rangle}{\langle k \rangle} \quad (1.7)$$

Pour une distribution de Poisson, $\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle$. Nous devons conclure que $\langle k \rangle_{\text{NN}} = \langle k \rangle + 1$ pour les réseaux ayant une distribution de degrés en loi de Poisson.

En général, les réseaux du modèle $G(N, p)$ ne forment pas des structures sans cycle, qu'on nomme *structures en arbre*. Pour qu'une structure soit en arbre, la probabilité qu'un noeud i soit connecté à j sachant que i et j sont connectés à k doit être pratiquement nulle :

$$\Pr(a_{ij} = 1 | a_{ik} = 1, a_{jk} = 1) \approx 0. \quad (1.8)$$

3. Notons que le double N en indice de $\langle k \rangle_{\text{NN}}$ est inspiré de la formulation anglaise *nearest neighbors*.

Par contre, cette probabilité est égale à $p > 0$ pour les réseaux aléatoires $G(N, p)$. Comme il est très fréquent d'approximer qu'une structure soit localement en arbre, les réseaux classiques $G(N, p)$ ne sont pas des candidats idéaux pour valider les méthodes approxinant la structure à un arbre. Notons toutefois que dans la limite des grands réseaux $N \rightarrow \infty$ et épars, c.-à-d. que $p \approx O(N^{-1})$, alors la probabilité de former un triangle est très faible.

Une grande faiblesse du modèle $G(N, p)$ est qu'il existe que très peu de réseaux réels possédant une distribution de degrés poissonnienne. On se tourne donc vers le modèle des configurations qui permet de générer aléatoirement des réseaux ayant une distribution de degrés arbitraire.

Modèle des configurations

Le modèle des configurations génère des réseaux pour lesquels on peut choisir la distribution des degrés. On assigne d'abord à chaque noeud un degré $k_i \in K$, où K est une séquence de degrés arbitraires tirés d'une distribution $\Pr(k)$. On connecte les noeuds aléatoirement tout en respectant le degré assigné à chaque noeud. Mentionnons qu'il faut toujours que $\sum_{i=1}^N k_i = 2M$ soit pair pour l'existence d'une structure respectant la séquence des degrés.

La probabilité de connexion entre deux noeuds est proportionnelle à leurs degrés, puisque

$$\Pr(a_{ij} = 1) \approx \frac{k_i k_j}{2M} \propto k_i k_j. \quad (1.9)$$

Deux noeuds de hauts degrés ont une plus grande probabilité d'être connectés que deux noeuds de faibles degrés. Ce constat émerge simplement du fait qu'ils ont plus de connexions.

La densité espérée ρ d'un graphe est la probabilité que deux noeuds aléatoires soient connectés. On obtient celle-ci en moyennant l'équation (1.9) sur toutes les paires (k_i, k_j) :

$$\rho = \frac{1}{2M} \sum_{i=1}^N \sum_{j=1}^N k_i k_j \Pr(k_i) \Pr(k_j) = \frac{\langle k \rangle \langle k \rangle}{2M} = \frac{\langle k \rangle}{N} \quad (1.10)$$

avec $2M = \langle k \rangle N$. Si la moyenne $\langle k \rangle$ existe, alors $\rho \propto N^{-1}$. Pour un grand réseau, la densité tend vers zéro avec la taille du graphe et on accepte alors l'approximation des structures creuses en arbre pour $N \rightarrow \infty$.

Modèle d'attachement préférentiel

Les réseaux générés par les modèles aléatoires $G(N, p)$ et des configurations n'ont pas d'ordre temporel d'arriver des noeuds. Le processus est que tous les noeuds sont créés en même temps et se connecte tous instantanément. Pourtant, les réseaux réels doivent croître selon une certaine mécanique temporelle. Il semble donc manquer un élément important chez les

modèles aléatoires. On complète le tableau en introduisant un mécanisme de croissance : l'attachement préférentiel (PA) [14], aussi connu sous le nom de *rich get richer*.

Le modèle le plus connu d'attachement préférentiel est un processus stochastique dans lequel la probabilité de former de nouvelles connexions est proportionnelle aux degrés des noeuds. Par exemple, un individu ayant déjà 10 relations d'amitié aura 10 fois plus de chance de former une nouvelle amitié qu'un autre individu ayant une seule relation. L'interprétation de cette règle n'est pas que le premier individu soit dix fois plus attachant que l'autre, mais plutôt que son passé est révélateur de son futur.

À l'initialisation du processus, on débute par deux noeuds connectés. À intervalle de temps $\Delta t = 1$, on ajoute un nouveau noeud de degré $k = 1$ et on le connecte à un noeud existant. Ce dernier est choisi avec une probabilité proportionnelle à son degré actuel. On décrit le processus via l'équation maîtresse de l'espérance du nombre de noeuds $N_k(t)$ de degré k au temps t ,

$$N_k(t+1) = N_k(t) + \delta_{k,1} + \frac{(k-1)N_{k-1}(t) - kN_k(t)}{\sum_k kN_k(t)} \quad (1.11)$$

où $\sum_k kN_k(t)$ est le nombre total de liens dans le réseau. Comme un lien est créé à chaque pas de temps et produit deux *stubs*, alors $\sum_k kN_k(t) = 2t$. La distribution normalisée des degrés est $\tilde{N}_k(t) = t^{-1}N_k(t)$. La distribution des degrés à l'équilibre $\tilde{N}_k(t) \approx \tilde{N}_k(t+1) \approx \tilde{N}_k^*$ suit une loi de puissance très caractéristique de ce modèle,

$$\lim_{k,t \rightarrow \infty} \tilde{N}_k^* \propto k^{-\gamma} \quad (1.12)$$

avec $\gamma = 3$. Lorsque la distribution des degrés d'un réseau suit une loi de puissance, on dit que le réseau est *invariant d'échelle*. Ce terme est évidemment inspiré de l'invariance d'échelle d'une loi de puissance, c.-à-d. , pour $f(x) = ax^{-k}$ alors $f(cx) = c^{-k}f(x) \propto f(x)$.

Plusieurs réseaux réels respectent approximativement une distribution de degrés en loi de puissance (Fig.1.1). La validité statistique de l'existence des lois de puissance pour les distributions empiriques est toutefois encore sujette à débat [29, 164].

L'attachement préférentiel s'est rapidement présenté comme un excellent mécanisme pour reproduire la propriété d'invariance d'échelle (34 000 citations en date de 2019 et récolte environ 2 500 citations chaque année) [14]. On ne compte maintenant plus les variations du modèle. Pour n'en nommer quelques-unes, il y a l'utilisation de noyaux d'attachement non linéaires [92, 93], de mécanismes de croissance de noeuds existants [94], de biais naissance/croissance [155] et l'extension aux structures hiérarchiques [68, 69].

...

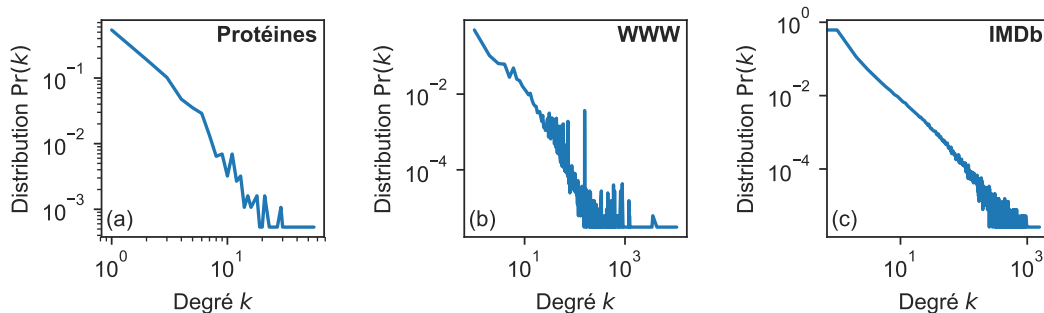


FIGURE 1.1 – Trois exemples de réseaux invariants d’échelle : (a) réseaux des interactions entre les protéines de levure (10^3 protéines), (b) réseaux des liens sortants des sites web nd.edu de l’Université Notre-Dame (10^5 sites), (c) réseau des collaborations d’acteurs (Données de Internet Movie Database, 10^6 noeuds).

1.2 La dynamique sur réseaux

Nous allons maintenant enrichir la description des réseaux complexes par l’ajout de la dynamique sur réseaux. Les noeuds sont porteurs d’un état d’activité, ou une *observable*, qu’on notera $x_i(t) \in \mathbb{R}$ pour l’état du noeud i au temps t et qui sera en général un scalaire.

On distingue deux familles de dynamiques : celles stochastiques et celles déterministes. Les *dynamiques stochastiques* rassemblent tous les processus où l’activité $x_i(t)$ est issue d’un processus stochastique. C’est le cas de plusieurs modèles d’épidémies, où un noeud change d’état selon une probabilité qui dépend de l’état de son voisinage, et des modèles stochastiques de populations d’écosystèmes [112]. Si l’activité n’est pas issue d’un processus stochastique, alors la dynamique est *déterministe*.

Dans cette thèse, nous traiterons des dynamiques reliées à trois domaines : les réseaux de contacts, les écosystèmes et la connectomique.

1.2.1 Cascade sur réseaux

Les cascades sur réseaux décrivent les propagations irréversibles comme la propagation d’un feu de forêt ou d’une rumeur dans un réseau de contacts. Pour notre exposé, nous discuterons des processus discrets de cascades ayant des observables binaires, c.-à-d. que $x_i(t) \in \{0, 1\}$ indique si un noeud est actif ou non [54, 166].

À l’initialisation de la cascade, tous les individus sont dans l’état $x_i(0) = 0$ sauf un ensemble \mathcal{V}_0 de noeuds actifs, soit $x_i(0) = 1$ pour tout $i \in \mathcal{V}_0$, qui représentent une fraction $\rho_0 = N^{-1} \sum_i x_i(0)$ du réseau. À tout moment, un noeud i inactif peut devenir actif si la fraction de voisins actifs est supérieure ou égale à un seuil r_i propre au noeud i . La transition vers

l'état actif est irréversible. La dynamique se termine lorsqu'aucune nouvelle transition n'est possible.

Le mécanisme a tout l'air d'un processus déterministe puisqu'aucun nombre aléatoire n'est tiré durant son exécution. La portion stochastique se cache plutôt dans le choix des seuils puisque r_i est tiré d'une distribution aléatoire prédéfinie $\text{Pr}_s(r)$. À chaque réalisation d'une cascade sur réseau, les seuils r_i doivent être échantillonnés. En choisissant des distributions appropriées $\text{Pr}_s(r)$, il est possible de reproduire un processus équivalent à une dynamique stochastique où un noeud actif a une probabilité p_i d'activer son voisin.

La fraction d'individus infectés ρ_n au pas de temps n est donnée par l'équation

$$\rho_n = N^{-1} \sum_{i=1}^N x_i(n). \quad (1.13)$$

Puisque la dynamique est irréversible, ρ_n est monotone croissante. Par conséquent, la limite

$$\lim_{n \rightarrow \infty} \rho_n = \rho_\infty \quad (1.14)$$

existe. Comme $x_i \in \{0, 1\}$, alors $0 \leq \rho_\infty \leq 1$.

Pour les grands réseaux avec $N \rightarrow \infty$, une cascade réussira à couvrir une fraction macroscopique du réseau si $\rho_\infty > 0$. Si cette condition est respectée, alors le plus grand sous-graphe connexe composé de noeuds actifs sera nommé la *composante géante*. L'observable ρ_∞ est donc d'un grand intérêt, car elle indique que la propagation a réussi ($\rho_\infty > 0$) ou a échoué ($\rho_\infty = 0$). On imagine bien l'importance de ρ_∞ dans le contexte des feux de forêt car il s'agit de la capacité d'un réseau à propager le feu. Or, nous avons tout intérêt à être en mesure de l'estimer avant qu'un feu ne se déclenche, d'où la pertinence d'être en mesure de faire des prédictions de la composante géante.

En première approximation, on estime ρ_∞ à l'aide d'un formalisme en champ moyen. D'abord, considérons que la probabilité qu'un noeud i soit actif $\text{Pr}(x_i(t) = 1)$ ne dépend que de son degré k_i . Cette hypothèse est justifiée pour les réseaux aléatoires car le degré d'un noeud est un excellent indicateur de son voisinage moyen [124]. On considère donc la probabilité $\rho_n^{(k)}$ qu'un noeud de degré k soit actif au temps n . La fraction globale active s'obtient en moyennant sur la distribution des degrés :

$$\rho_n = \sum_{k=1}^{\infty} \text{Pr}(k) \rho_n^{(k)}. \quad (1.15)$$

Deux méthodes d'activation contribuent à l'expression de $\rho_n^{(k)}$: soit le noeud k est un noeud source avec probabilité ρ_0 ou soit il a été activé par ses voisins avec une certaine probabilité qu'on obtient par le raisonnement suivant [57].

La probabilité qu'un noeud de degré k ait m voisins actifs suit une distribution binomiale,

$$\binom{k}{m} (\tilde{\rho}_{n-1})^m (1 - \tilde{\rho}_{n-1})^{k-m}, \quad (1.16)$$

avec $\tilde{\rho}_{n-1}$ la probabilité qu'un voisin soit actif au temps $n - 1$. Cette dernière s'obtient en moyennant la quantité $\rho_{n-1}^{(k)}$ sur les voisins. Comme un noeud de degré k est le voisin de k noeuds, alors la moyenne d'une quantité d'un noeud moyennée sur le voisinage s'obtient en pondérant la quantité par le degré des noeuds,

$$\tilde{\rho}_{n-1} = \sum_{k=1}^{\infty} \frac{k}{\langle k \rangle} \Pr(k) \rho_{n-1}^{(k)}. \quad (1.17)$$

Pour qu'un noeud devienne actif par l'effet de son voisinage, le nombre de voisins actifs m doit être supérieur ou égal à r . On a donc besoin de la probabilité $C(r)$ que le seuil soit inférieur à r , soit

$$C(r) = \int_0^r \Pr_s(r') dr', \quad (1.18)$$

où $\Pr_s(r)$ est la densité de probabilité des seuils évaluée en r . En rassemblant ces équations et en incluant les noeuds initialement actifs, on obtient la probabilité $\rho_n^{(k)}$ qu'un noeud de degré k soit actif au temps n :

$$\rho_n^{(k)} = \rho_0 + (1 - \rho_0) \sum_{m=0}^k \binom{k}{m} (\tilde{\rho}_{n-1})^m (1 - \tilde{\rho}_{n-1})^{k-m} C(m/k). \quad (1.19)$$

En multipliant par ρ_k et en sommant sur les k [Eq. (1.15)] pour $n \rightarrow \infty$, on a que

$$\rho_{\infty} = \rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} \Pr(k) \sum_{m=0}^k \binom{k}{m} (\tilde{\rho}_{\infty})^m (1 - \tilde{\rho}_{\infty})^{k-m} C(m/k). \quad (1.20)$$

Pour évaluer $\tilde{\rho}_{\infty}$, on reprend l'équation (1.19) qu'on multiplie par $\frac{k}{\langle k \rangle} \Pr(k)$ [selon Eq. (1.17)] et on somme sur les k , ce qui permet de déduire que

$$\tilde{\rho}_n = \rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} \frac{k}{\langle k \rangle} \Pr(k) \sum_{m=0}^k \binom{k}{m} (\tilde{\rho}_{n-1})^m (1 - \tilde{\rho}_{n-1})^{k-m} C(m/k) \quad (1.21)$$

On retrouve $\tilde{\rho}_{\infty}$ lorsque $n \rightarrow \infty$.

Numériquement, on estime la fraction espérée des noeuds actifs au terme de la cascade ρ_{∞} via l'équation (1.20) avec $\tilde{\rho}_{\infty}$ solutionnant l'équation (1.21). On note que cette solution est indépendante du nombre de noeuds N . Toutefois, sa validité sera meilleure dans le régime où $N \rightarrow \infty$. En effet, l'équation (1.16) suppose l'indépendance entre les états des voisins d'un noeud. Une supposition qui est d'autant plus vraie lorsque $N \rightarrow \infty$.

À la Fig. 1.2, la solution à l'équation (1.20) est comparée à des résultats de simulations du processus. Le premier constat est que l'équation (1.20) prédit correctement l'apparition d'une composante géante $\rho_{\infty} > 0$, mais elle échoue à estimer le degré moyen maximal d'existence de la composante géante. L'erreur de prédiction s'explique par le manque de direction de la méthode. La cascade est fondamentalement directionnelle, car l'activation croît vers l'extérieur en partant des sources d'activité \mathcal{V}_0 . Pour activer un noeud, il faut donc d'abord activer son

voisinage. Par conséquent, si la structure est organisée en arbre, les noeuds initialement actifs forment le coeur et les noeuds en périphérie ne peuvent être actifs qu'après le temps nécessaire pour activer leur voisinage. Tout cet aspect séquentiel est négligé par le formalisme en champ moyen présenté.

Pour considérer l'aspect séquentiel d'activation et obtenir un nouveau système d'équations, on suppose que la structure est organisée en arbre. On assigne d'abord un niveau à chaque noeud et où le niveau 0 est la racine de l'arbre. Cela nous permet de travailler avec la probabilité q_n qu'un noeud au niveau n soit actif [57]. En utilisant un argument similaire au formalisme précédent, on obtient le système d'équations

$$\rho_\infty = \rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} \text{Pr}(k) \sum_{m=0}^k \binom{k}{m} q_\infty^m (1 - q_\infty)^{k-m} C\left(\frac{m}{k}\right), \quad (1.22)$$

$$q_{n+1} = \rho_0 + (1 - \rho_0) G(q_n), \quad (1.23)$$

$$G(q) = \sum_{k=1}^{\infty} \frac{k}{\langle k \rangle} \text{Pr}(k) \sum_{m=0}^{k-1} \binom{k-1}{m} q^m (1 - q)^{k-1-m} C\left(\frac{m}{k}\right), \quad (1.24)$$

où la démonstration est présentée à la référence [57]. Si on compare avec le système (1.21), on trouve très peu de différences. La plus notable est la présence du $k - 1$ dans la proposition de $G(q)$. L'utilisation d'un degré $k - 1$ plutôt que k est justifiée par une structure en arbre : si la cascade parvient à un noeud par un lien, alors il ne reste que $k - 1$ liens pour se propager. Si la structure n'est pas en arbre, il pourrait y avoir moins de $k - 1$ liens pour propager l'activation car certains noeuds du voisinage pourraient déjà avoir été infectés.

À la Fig. (1.2), la solution du système (1.24) est comparée aux simulations numériques. La solution est maintenant très satisfaisante. Plus précisément, deux transitions sont correctement détectées. La première transition à faible $\langle k \rangle$ survient lorsque le degré moyen est trop faible pour propager la dynamique. La cascade n'a simplement pas suffisamment de chemins pour se propager et s'essouffle rapidement. La deuxième transition, à plus haut degré moyen, survient quand le critère d'activation est trop exigeant pour démarrer la cascade. En effet, le seuil d'activation d'un noeud est donné par la fraction de ses voisins actifs. En augmentant le degré moyen, le nombre de voisins actifs requis pour activer un noeud augmente. La transition survient lorsque l'activation moyenne de la cascade est incapable de satisfaire cette condition.

La position des transitions critiques s'obtient à partir du système (1.24). D'abord, on développe $G(q)$ sous une forme polynomiale $G(q) = \sum_{l=0}^{\infty} C_l q^l$, avec

$$C_l = \sum_{k=l+1}^{\infty} \sum_{n=0}^l \binom{k-1}{l} \binom{l}{n} (-1)^{l+n} \frac{k}{\langle k \rangle} \text{Pr}(k) C(n/k). \quad (1.25)$$

Pour qu'il y ait une composante géante, il doit y avoir une croissance de la probabilité d'être actif avec le temps, c.-à-d. $q_{n+1}/q_n > 1$. Au premier ordre, $G(q) \approx C_1 q$ et on retrouve donc le

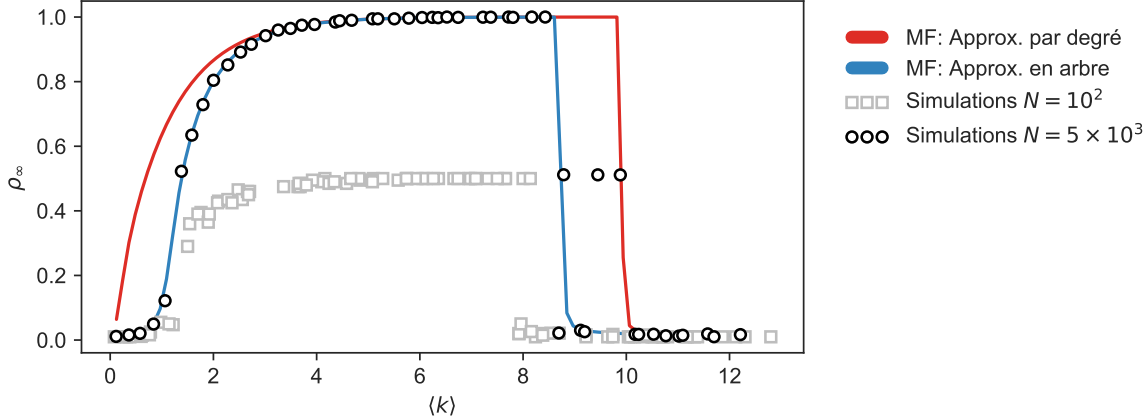


FIGURE 1.2 – Fraction de noeuds actifs ρ_∞ au terme d'un processus de cascade avec un seuil uniforme de $r = 0.18$ et $\rho_0 = 10^{-2}$. Les réseaux sont tirés du modèle de Erdős-Rényi avec $N = 100$ (gris) et $N = 5000$ (noir) avec le paramètre de densité ajusté pour satisfaire les différents degrés moyens. La courbe rouge est solution de Eq. (1.20) et la courbe bleue est solution de (1.24) avec $p_k = \langle k \rangle^k e^{-\langle k \rangle} / k!$. Chaque point est obtenu en moyennant sur 50 simulations pour chaque valeur de $\langle k \rangle$. Pour réduire le bruit, les simulations ayant ρ_∞ entre le 25e et le 75e percentile ont été utilisées pour calculer les moyennes.

critère

$$(1 - \rho_0)C_1 > 1. \quad (1.26)$$

En exprimant C_1 avec l'équation (1.25), on obtient le critère

$$\sum_{k=1}^{\infty} \frac{k(k-1)}{\langle k \rangle} \Pr(k) [C(1/k) - C(0)] > \frac{1}{1 - \rho_0}. \quad (1.27)$$

Notons qu'il est possible de préciser le critère en approximant $G(q)$ jusqu'au terme quadratique [56]. La transition de phase survient lorsqu'il y a égalité à l'inégalité (1.27). Comme $k(k-1)$ est monotone croissant en k , mais que $C(k^{-1})$ est monotone décroissant en k , alors il peut exister au maximum deux solutions en $\langle k \rangle$.

Les approches en champ moyen sont simples à utiliser et donnent des résultats assez précis. Elles sont facilement généralisables aux réseaux modulaires et aux réseaux dont on connaît la distribution corrélée des degrés $\Pr(k'|k)$, c.-à-d. la probabilité qu'un noeud sachant de degré k soit connecté à un noeud de degré k' [54].

1.2.2 Dynamique SIS

Lorsque la transition de l'état actif vers l'état inactif est possible, la dynamique est dite réversible. C'est le cas de la dynamique Susceptible-Infected-Susceptible (SIS) à temps continu qui décrit l'évolution d'une infection sur un réseau. La dynamique SIS est binaire telle que

$x_i \in \{0, 1\}$. Traditionnellement, un noeud actif $x_i = 1$ est dit infecté, tandis qu'un noeud inactif $x_i = 0$ est dit susceptible d'être infecté.

La transition de l'état infecté vers l'état susceptible est faite à taux constant $\alpha \geq 0$ et est indépendante de l'état du voisinage. Un noeud infecté ($x_i = 1$) peut transmettre l'infection à un voisin susceptible ($x_j = 0$) avec un taux $\beta \geq 0$. Sans perte de généralité (en excluant le cas $\alpha = 0$), on normalise les taux tel que $\alpha \rightarrow 1$ et on utilise un taux de transmission $\gamma = \beta/\alpha \geq 0$.

Plusieurs raisons expliquent l'intérêt porté à la dynamique SIS. Premièrement, cette dynamique est polyvalente, car elle modélise plusieurs phénomènes, dont la propagation d'une maladie, d'un vaccin ou d'une rumeur dans un réseau de contacts. Deuxièmement, le mécanisme sous-jacent est simple et facilite la tâche de modélisation. Avec peu de travail et quelques approximations, nous serons en mesure de prédire l'existence d'une transition dans la fraction macroscopique de noeuds infectés, c.-à-d. la moyenne de l'activité au temps t donnée par $\langle x(t) \rangle$. Cette transition est justement le troisième point d'intérêt : la dynamique comporte un changement de phase. Nous verrons que ce changement de phase peut être induit ou bien en modifiant le taux de transmission γ ou bien en perturbant la structure du réseau. Enfin, il est bon de mentionner que la dynamique SIS s'est imposée comme un élément canonique de la dynamique sur réseaux et devient un passage obligé pour valider des modèles prédictifs [15, 91, 124, 135].

Solution par champ moyen

L'analyse de la dynamique débute en utilisant un cadre théorique de champ moyen décorréolé (*quenched mean-field*)⁴ décrivant la probabilité y_i que le noeud i soit infectée,

$$\dot{y}_i = -y_i + \gamma(1 - y_i) \sum_{j=1}^N a_{ij} y_j \quad (1.28)$$

où i est de 1 à N . On s'intéresse aux solutions y_i^* à équilibre qui satisfont $\dot{y}_i = 0$. La solution triviale $y_i^* = 0$ existe toujours, quelque soit les valeurs de γ et a_{ij} . On nomme cette solution la phase *absorbante*. Comme le système (1.28) est quadratique, on peut s'attendre à la présence d'une bifurcation transcritique. On linéarise le système à $y_i \approx 0$ et on obtient la forme suivante :

$$\dot{y}_i = \sum_{k=1}^N (-\delta_{ik} + \gamma a_{ik}) y_k. \quad (1.29)$$

La solution $y_i = 0$ sera instable lorsque la plus grande valeur propre de la matrice d'éléments $-\delta_{ik} + \gamma a_{ik}$ sera positive. On obtient donc le point de bifurcation en

$$\gamma_c = \frac{1}{\lambda_D}, \quad (1.30)$$

4. Notons également que malgré sa haute dimensionnalité, le système (1.28) est valide seulement si les états du voisinage ne sont pas corrélés [136]. Ces termes de corrélation se trouvent dans des termes d'ordre supérieur [110].

où λ_D est la plus grande valeur propre de la matrice d'adjacence \mathbf{A} [110]. À titre d'exemple, la Fig. 1.3 présente une comparaison du seuil prédit et du résultat de simulations de la dynamique.

On simplifie maintenant le système (1.28) en définissant une observable globale ρ qui prendra la valeur de la fraction d'individus infectés,

$$\rho = \frac{1}{N} \sum_{i=1}^N y_i. \quad (1.31)$$

L'observable ρ est la moyenne arithmétique des variables y_i . En sommant le système (1.28) sur l'indice i et en supposant que le réseau est homogène tel que $a_{ij} \sim N^{-1}\langle k \rangle$ et donc $k_i \approx \langle k \rangle$, on décrit l'évolution temporelle de la nouvelle observable, soit

$$\dot{\rho} \approx -\rho + \gamma \langle k \rangle \rho (1 - \rho). \quad (1.32)$$

Le système original à N équations est maintenant exprimé en une seule équation où la structure est fortement réduite à son degré moyen $\langle k \rangle$. On remarque alors beaucoup mieux la forme canonique des bifurcations transcritiques $\dot{x} = rx - x^2$. En solutionnant pour $\dot{\rho} = 0$, on obtient le seuil de transition $\gamma_c = \langle k \rangle^{-1}$ pour lequel $\rho = 0$ pour $\gamma < \gamma_c$ et $\rho \sim \gamma - \gamma_c$ pour $\gamma \geq \gamma_c$ [25, 136]. Ce seuil de transition est cohérent avec le seuil obtenu à l'équation (1.30). En effet, la valeur propre dominante est environ égale au degré moyen pour les réseaux aléatoires Erdős-Rényi.

Il est déjà bon de noter que le choix de la moyenne uniforme pour construire l'observable globale ρ est arbitraire et qu'une moyenne pondérée aurait pu aussi faire l'affaire, c.-à-d. avec $\rho' = \sum_i b_i y_i$ avec $b_i \in \mathbb{R}$. Toutefois, utiliser une moyenne uniforme assure son interprétabilité et simplifie l'analyse. Au Chapitre 3, nous revisiterons l'utilisation de la moyenne uniforme pour construire les observables globales et formaliserons le choix à une moyenne pondérée.

1.2.3 Dynamique de Wilson-Cowan

Parmi les exemples classiques de systèmes dynamiques, il y a celle qui se produit sur le *connectome* qui représente le réseau des connexions entre des neurones ou des groupes de neurones. Ces systèmes sont intrinsèquement dynamiques, car l'activité neuronale est à l'origine des phénomènes émergents comme la motricité et la mémoire. Naturellement, une dynamique neuronale est difficile à modéliser [79, 80]. Nous nous limitons donc à des dynamiques simplifiées. Parmi celles-ci se trouve la dynamique de Wilson-Cowan [171, 172] introduite au début des années 1970 et encore populaire [45].

Considérons qu'un noeud i représente un groupe de neurones. Dans le modèle dynamique de Wilson-Cowan, l'activité y_i du noeud i est donnée par la solution au système suivant :

$$\dot{y}_i = -y_i + \sigma \left(\sum_{j=1}^N w_{ij} y_j \right), \quad (1.33)$$

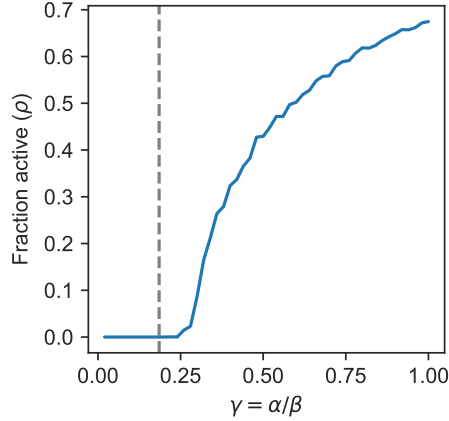


FIGURE 1.3 – Fraction active à l’équilibre en fonction du taux de transmission de la dynamique SIS. La ligne bleue est obtenue par simulations de la dynamique sur un réseau $G(N = 100, p = 0.04)$. Le trait gris est la prédiction de transition par l’équation (1.30).

avec $\sigma(x)$, la fonction d’activation *sigmoïde*,

$$\sigma(x) = \frac{1}{1 + \exp[-\tau(x - \mu)]}, \quad (1.34)$$

où $\mu \geq 0$ est un seuil d’excitation et $\tau > 0$ contrôle la pente de la fonction d’activation. Le terme $-y_i$ de l’équation (1.35) vérifie que l’activité d’un noeud décroît s’il n’est pas entretenu par sa fonction d’activation.

En appliquant le changement de variable $x_i = \sum_{j=1}^N w_{ij}y_j$, on obtient le système suivant :

$$\dot{x}_i = -x_i + \sum_{j=1}^N w_{ij}\sigma(x_j), \quad (1.35)$$

qui sera utilisé dans cette thèse. L’avantage de l’équation (1.35) par rapport à l’équation (1.33) est que la somme sur le voisinage est à l’extérieur de la sigmoïde et obéit donc à une forme plus commune, c.-à-d. $\dot{x}_i = F(x_i) + \sum w_{ij}G(x_i, x_j)$. Dans cette thèse, nous utiliserons l’équation (1.35) lorsqu’il sera mentionné de la dynamique de Wilson-Cowan.

Lorsque les poids des liens sont non négatifs $w_{ij} \geq 0$, on peut montrer que le système converge toujours vers un point fixe stable \mathbf{x}^* qui satisfait $\dot{\mathbf{x}} = 0$ [77].

On s’intéresse à cette dynamique pour plusieurs raisons. Premièrement, le système (1.35) est déterministe. En connaissant les paramètres τ, μ et la structure \mathbf{W} parfaitement, on peut solutionner le système par intégration numérique. Deuxièmement, la forme de la fonction sigmoïde empêche le calcul d’une solution analytique explicite des points fixes \mathbf{x}^* . Une approche numérique ou approximative est obligatoire. Par conséquent, ce système suggère qu’une analyse en faible dimension serait profitable pour en dégager des bifurcations. Troisièmement, nous verrons que la présence de la fonction sigmoïde engendre une hystérèse dans les solutions à

équilibre. Cette hystérèse constitue donc une transition de phase différente de celle observée pour la dynamique SIS.

Pour vérifier l'existence de l'hystérèse à l'équilibre pour l'équation (1.35), on procède à une analyse du diagramme de bifurcation du cas unidimensionnel. On considère $x_i \equiv x$ et $\dot{x} = 0$ dont la solution à l'équilibre satisfait

$$x = c\sigma(x). \quad (1.36)$$

Il existe 1 à 3 intersections entre une droite (terme de gauche) et une sigmoïde (terme de droite). S'il existe trois solutions, alors au moins l'une d'elles doit être instable. Le théorème 1 détaille l'existence des points fixes. Sa démonstration est placée en Annexe A.

Théorème 1. *Considérons l'équation différentielle*

$$\dot{x} = -x + c \frac{1}{1 + \exp[-\tau(x - \mu)]} \quad (1.37)$$

avec $c > 0$, $\tau > 0$ et $\mu \geq 0$ et x défini sur les réels. L'équation (1.37) possède toujours au moins un point fixe. De plus, les paramètres

$$\alpha = c\tau \quad \text{et} \quad \beta = \mu\tau \quad (1.38)$$

régissent le nombre de points fixes tel que

(I). si $\alpha \leq 4$, alors (1.37) possède un seul point fixe.

(II). si $\alpha > 4$, alors (1.37) possède

- i. un seul point fixe si $\beta_- > \beta$ ou $\beta > \beta_+$;
- ii. deux points fixes si $\beta_- = \beta$ ou $\beta = \beta_+$;
- iii. trois points fixes si $\beta_- < \beta < \beta_+$;

avec

$$\beta_{\pm} = \log \left[\frac{1 \pm \gamma}{1 \mp \gamma} \right] + \frac{\alpha}{2}(1 \pm \gamma) \quad \text{et} \quad \gamma = \sqrt{1 - \frac{4}{\alpha}}. \quad (1.39)$$

Selon le théorème 1, l'apparition de deux points fixes survient exactement au point de rebroussement $c\tau = 4$ et $\mu\tau = 2$. Pour un ensemble de paramètres tels que $\mu\tau > 2$, le paramètre c contrôle le nombre de points fixes du système. Si c est très faible ou très grand, le système aura 1 seul point fixe et 2 ou 3 autrement (Fig. 1.4). Cette bifurcation est nommée la *cusp catastrophe* [159]. Le paramètre c joue le rôle de structure effective lorsque comparé au terme de w_{ij} du système (1.35). Ainsi, cette courte analyse suggère que la structure effective peut contrôler l'état dynamique du système et forcer une bifurcation.

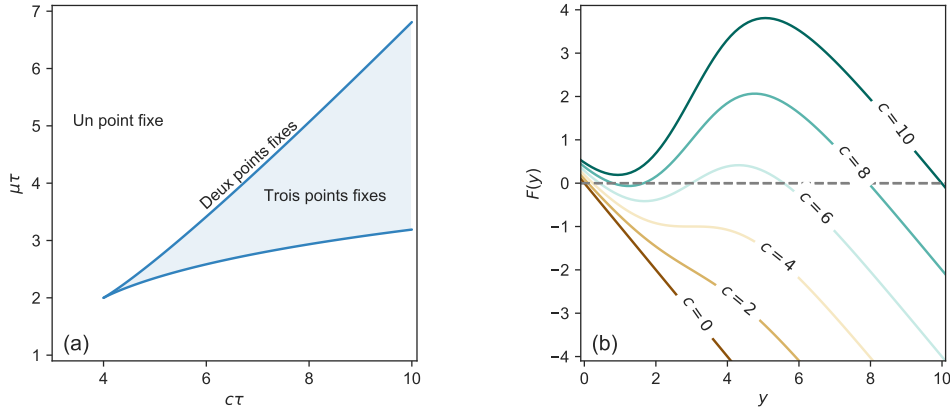


FIGURE 1.4 – (a) Diagramme de stabilité pour le système de Wilson-Cowan (1.35). Le point de rebroussement survient à $c\tau = 4$ et $\mu\tau = 2$. (b) Illustration de $F(y) = y - \sigma(y)$, représentant le terme de droite à l'équilibre de l'équation (1.35) unidimensionnelle, pour $\tau = 1$ et $\mu = 3$ et différentes valeurs de c .

1.2.4 Modèle de populations

Les écosystèmes sont au coeur d'une vaste littérature scientifique [75, 111, 112]. Ils sont étudiés entre autres pour documenter leur stabilité face à l'environnement toujours changeant. Contrairement aux interactions humaines, les interactions entre les espèces d'un écosystème sont généralement de nature prédatrice ou coopérative. Dans les deux cas, elles influencent principalement la population des espèces.

Supposons que les noeuds représentent les différentes espèces et $x_i(t) \in \mathbb{R} \geq 0$ est le nombre d'individus de l'espèce i . Parfois, l'activité $x_i(t)$ est décrite au moyen d'un processus stochastique. Toutefois, lorsque la population est vaste, les variations induites par le processus stochastique sont faibles par rapport à un processus déterministe équivalent [112]. Par conséquent, nous examinerons les processus déterministes des populations.

Nous nous intéressons aux dynamiques de populations pour plusieurs raisons. Premièrement, les retombées scientifiques potentielles de prédictions quantitatives sont énormes. En effet, avec la menace climatique actuelle, être en mesure de prédire le comportement des écosystèmes permettrait de meilleures stratégies de protection des écosystèmes. Deuxièmement, les dynamiques de populations induisent des bifurcations différentes de celles rencontrées avec la dynamique SIS et de Wilson-Cowan. Enfin, l'existence de relations de prédation implique des comportements oscillatoires des populations. Par conséquent, nous pouvons nous attendre à la présence de phénomènes chaotiques.

Il n'existe pas de forme universelle, ni optimale, gouvernant l'évolution temporelle de $x_i(t)$.

La forme la plus commune est sans doute la relation de Lotka-Volterra

$$\dot{x}_i = F(x_i) = c_i x_i + \sum_{j=1}^N x_i w_{ij} x_j \quad (1.40)$$

où $c_i \in \mathbb{R}$. En accordant des valeurs négatives aux poids $w_{ij} \in \mathbb{R}$, alors certaines interactions pourront être compétitives $w_{ij} < 0$ et d'autres coopératives $w_{ij} > 0$.

Ce système constitue un élément pédagogique très populaire en deux dimensions afin d'introduire les cycles limites [159]. Pour trois dimensions et plus, un système aussi simple que (1.40) peut présenter des solutions stables ou instables, de l'oscillation et des trajectoires chaotiques. Nous utiliserons directement cette forme au Chapitre 4 afin d'introduire un bon niveau de complexité.

Le système (1.40) est mutualiste lorsque $w_{ij} \geq 0 \forall (i \neq j)$, c.-à-d. que

$$\frac{\partial \dot{x}_i}{\partial x_j} = x_i w_{ij} \geq 0 \quad (1.41)$$

puisque $x_i \geq 0$. Supposons que le taux de naissance est négatif $c_i < 0$ pour toutes les espèces. Il existe deux solutions stables, soit la solution triviale $\mathbf{x}^* = 0$ et une solution non-négative $\mathbf{x}^* \geq 0$ [112]. Si $c_i < 0$ et $w_{ij} \geq 0$, alors le système diverge vers l'infini.

La pertinence du modèle (1.40) est à toute fin pédagogique. Elle n'a que peu d'intérêt pour des prédictions quantitatives sur des systèmes réels [112]. C'est pourquoi on introduit généralement des mécanismes supplémentaires. Au Chapitre 3, nous considérerons la forme suivante :

$$\dot{x}_i = B_i + x_i \left(1 - \frac{x_i}{K_i}\right) \left(\frac{x_i}{C_i} - 1\right) + \sum_{j=1}^N w_{ij} \frac{x_i x_j}{D_i + E_i x_i + H_i x_j}. \quad (1.42)$$

Le terme de naissance est donné par la forme logistique de Verhulst-Pearl [140],

$$B_i + x_i \left(1 - \frac{x_i}{K_i}\right) \quad (1.43)$$

où $K_i \in \mathbb{R}$ est un paramètre de capacité de l'environnement.

Lorsque la population x_i est plus grande que la capacité K_i , le second terme de (1.43) devient négatif et le nombre d'individus tend à diminuer. Le paramètre $B_i \in \mathbb{R}$ quantifie les naissances par migration. On complète le terme de naissance avec l'effet de Allee [44]

$$\left(\frac{x_i}{C_i} - 1\right) \quad (1.44)$$

où C_i est la population minimale pour observer de la croissance.

Dans le modèle de Lotka-Volterra, les interactions entre les espèces prennent la forme de $x_i w_{ij} x_j$. Ce terme croît avec les populations sans atteindre de saturation. De manière plus

réaliste, on s'attend à observer une saturation si le nombre de prédateurs devenait trop grand. On utilise donc la forme introduite par Holling [74] en 1965,

$$\frac{w_{ij}x_i x_j}{D_i + E_i x_i + H_i x_j} \quad (1.45)$$

où $D_i, E_i, H_i \in \mathbb{R}$. Le dénominateur de cette dernière forme est conçu pour limiter l'interaction entre deux espèces et est paramétrisé par D_i, E_i, H_i .

Plusieurs autres systèmes auraient pu être construits [112]. Nous avons choisi de travailler avec celle-ci pour une raison bien précise. La première étant que cette forme a été introduite dans la publication [52] qui a motivé les travaux du Chapitre 3. Par conséquent, notre étude se compare mieux aux résultats produits par les travaux existants en utilisant la même dynamique. La deuxième raison est que les non-linéarités de la dynamique permettent un réalisme nécessaire pour justifier l'applicabilité des conclusions de nos travaux.

...

1.3 La résilience des systèmes complexes

Les systèmes complexes réels interagissent avec leur environnement. Ils sont entourés de facteurs perturbants qui agissent sur les paramètres structurels, e.g., ajout/retrait de liens ou de noeuds, les paramètres dynamiques, e.g., modification du taux de reproduction des espèces, et l'activité des noeuds, e.g., $\mathbf{X} \leftarrow \mathbf{X} + d\mathbf{X}$. Un système dynamique peut réagir aux perturbations en se dirigeant vers un nouvel état d'équilibre qualitativement différent. La portée des perturbations et tout ce qui entoure le retour à l'équilibre sont rattachés au terme de la *résilience*.

La *résilience* apparaît dans une grande quantité d'ouvrages scientifiques [51] et sa définition diffère d'un exposé à l'autre (voir le tableau 1.1). La grande variété de définitions n'est pas due au hasard. Il serait plutôt naïf de suggérer une définition universelle de la résilience. Notre interprétation de la résilience doit être adaptée au contexte à l'étude : le système est-il oscillatoire, adaptatif, stochastique, etc. ?

Dans les contextes déterministes où l'activité est à l'équilibre, $\dot{\mathbf{x}} = F(\mathbf{x}^*) = 0$, la résilience se pose comme une mesure de la stabilité locale du point d'équilibre \mathbf{x}^* . Si après la perturbation le système revient arbitrairement près de \mathbf{x}^* en un temps fini, souvent par une oscillation amortie ou une transition monotone, alors le point d'équilibre est stable. Plus le système revient vite et plus il semble robuste aux perturbations. Dans de très rares cas (principalement en faibles dimensions), le système peut débiter un mouvement oscillatoire sans amortissement après la perturbation et ne jamais revenir au point d'équilibre.

L'évolution temporelle du retour à l'équilibre pour une faible perturbation est obtenue en linéarisant le système. On suppose le contexte dynamique

$$\dot{\mathbf{x}} = F(\mathbf{x}), \quad (1.46)$$

admettant un point d'équilibre en \mathbf{x}^* si $F(\mathbf{x}^*) = 0$. On déplace l'activité de son point d'équilibre $\mathbf{x} = \mathbf{x}^* + \boldsymbol{\epsilon}$. Si la perturbation est faible, alors Eq. (1.46) est approximativement linéaire et telle que l'évolution temporelle autour du point d'équilibre est

$$\epsilon_i(t) = \sum_{j=1}^N C_{ij} \exp(\lambda_j t) \quad (1.47)$$

où C_{ij} dépend des conditions initiales et λ_j est la j -ième valeur propre de la matrice jacobienne \mathbf{J} de la dynamique, dont l'élément (i, j) est

$$\mathbf{J}_{ij} = \left. \frac{\partial F_i(\mathbf{x})}{\partial x_j} \right|_{\mathbf{x}=\mathbf{x}^*}. \quad (1.48)$$

Puisque le retour à l'équilibre suit une somme d'exponentielles dont les arguments contiennent les valeurs propres λ_i de la matrice jacobienne [Eq. (1.47)], alors la stabilité locale du point fixe se caractérise par le spectre de la matrice jacobienne. Les parties réelles négatives des valeurs propres du spectre seront les plus importantes, car elles sont responsables de l'amortissement de la trajectoire par rapport au point d'équilibre. Le critère de stabilité est alors évident : le système est localement stable si la partie réelle de chacune des valeurs propres sont toutes négatives,

$$\max_{\lambda_i} [\text{Re}(\lambda_i)] < 0. \quad (1.49)$$

S'il y a une ou plusieurs valeurs propres avec des parties réelles positives, alors \mathbf{x}^* sera instable.

Alors que l'équation (1.49) est simple, son applicabilité est loin de l'être. D'abord, si le système est instable ou la perturbation est de trop grande amplitude, la contribution des termes non linéaires ne sera plus négligeable devant celle des termes linéaires. Le critère de stabilité ne sera d'aucune utilité pour prédire le nouvel état, e.g., extinction globale $\mathbf{x} = 0$, nouvel état d'équilibre, cycle limite, etc. Ensuite, la stabilité locale n'est pas un indicateur de la stabilité globale, sauf pour les systèmes parfaitement linéaires. Elle n'est donc pas un bon outil pour traiter les grandes perturbations. Ici, la définition de ce qu'est une trop *grande* perturbation est volontairement imprécise. À quel moment devons-nous accepter la linéarisation comme une bonne approximation ? Il n'y a pas une seule bonne réponse, mais seulement une imprécision qui est soulevée.

On y opposera également que la stabilité locale ne passe pas tout à fait le test de la réalité. Les systèmes réels ne sont pratiquement jamais à un point d'équilibre, mais plutôt dans des états transitoires qui peuvent, ou non, fluctuer autour d'un point d'équilibre. Enfin, les perturbations dites *structurelles*, soit par la modification de la structure du réseau ou des paramètres de la

TABLE 1.1 – Exemples de définitions de la résilience selon les domaines scientifiques. Adapté de Ref. [51].

	Définition
Écologie	Persistance d'un système et capacité à absorber un changement tout en conservant essentiellement la même fonction.
Ingénierie	Capacité d'un système à rétablir ses fonctions d'origine après une perturbation.
Économie	Capacité de répondre à une perturbation pour limiter une certaine fonction de perte.
Psychologie	Capacité d'une personne de maintenir un état d'équilibre malgré les épreuves.

dynamique, sont mal quantifiées par une mesure de stabilité. En modifiant la structure, l'état d'équilibre se retrouve inévitablement modifié. Le système ne retourne donc pas exactement à l'état d'équilibre initial malgré que le nouveau point d'équilibre est souvent qualitativement équivalent.

D'autres mesures de résilience traitent du retour à l'équilibre. Par exemple, Holling propose d'y attacher le volume dans l'espace d'activité \boldsymbol{x} pour lequel le système retournera au point stable \boldsymbol{x}^* [75]. Cette forme est essentiellement la perturbation maximale que peut prendre le système. Toutefois, estimer ce volume est souvent impossible en pratique.

La résilience des systèmes complexes réels ne semble donc pas facilement quantifiable. Notons également que nous n'avons pas discuté que l'analyse de la stabilité se complexifie drastiquement si le système est oscillatoire, e.g., cycle limite.

1.3.1 Détecter et comprendre les perturbations

Dans ce contexte de résilience, notre travail est de contribuer à une meilleure analyse des systèmes complexes perturbés. Dans ce cadre conceptuel, nous dégagons deux tâches d'intérêts :

- **détecter** une perturbation ;
- **estimer** l'effet d'une perturbation.

Éventuellement, nos travaux pourront conclure vers le développement de stratégies d'intervention pour rétablir les fonctions d'origine du système.

Les objets à l'étude sont les réseaux complexes dynamiques tels que décrits dans les sections précédentes. Les perturbations seront toujours structurelles et prendront la forme d'un retrait de liens. Cette forme de perturbation a été choisie, car elle est commune à plusieurs scénarios. Pensons aux écosystèmes où on répand un pesticide sur certaines plantes et qui a pour effet de couper des liens dans le réseau trophique, ou même à une lésion physique dans un système neuronal. Nous aurons l'occasion de découvrir plusieurs autres exemples au cours de cet exposé.

Détecter une perturbation structurelle – Une tâche essentielle en résilience des systèmes, notamment dans le cadre des changements climatiques, est d'être capable de détecter le plus rapidement possible l'existence d'agents perturbants dans un système. Pour s'y prendre, il faut se baser sur des jeux de données décrivant des réseaux complexes. Le défi est de taille considérant parfois les faibles volumes de données disponibles et la complexité des objets.

Une avenue intéressante pour la détection des perturbations sur les réseaux complexes dynamiques est d'étudier les séries temporelles de l'activité des noeuds. De manière similaire à la détection par électroencéphalographie des crises épileptiques, l'activité sur le réseau peut être un indicateur de perturbations.

La détection d'anomalies dans des séries temporelles est bien documentée notamment en ingénierie. Plusieurs méthodes efficaces existent pour accomplir cette tâche [3]. Toutefois, la plupart des méthodes considèrent chaque série de manière indépendante. Mais dans les réseaux complexes, les séries temporelles ne sont pas indépendantes et seront couplées selon une structure en graphe avec beaucoup de composantes, voire des milliers. Ce travail nécessite donc de développer des outils adaptés à des données structurées.

Estimer l'impact de la perturbation – Que ce soit après une éventuelle détection de perturbation ou pour simplement mieux documenter les systèmes réels, nous devons être en mesure d'estimer l'effet d'une perturbation. Est-ce que le réseau peut bifurquer vers un nouvel état dynamique ? Pour être en mesure d'y répondre, il faut avoir une bonne compréhension de la relation structure-dynamique, le *leitmotiv* de la science des réseaux [127]. Il faut également pouvoir accomplir l'estimation dans un délai raisonnable. Ce dernier critère est primordial pour les réseaux avec plusieurs milliers de noeuds où certaines méthodes, dites de *force brute*⁵, ne sont pratiquement pas réalisables (Fig. 1.5).

...

5. Nous verrons que ces méthodes se présentent sous la forme d'intégration numérique de systèmes d'équations à N dimensions de la dynamique.

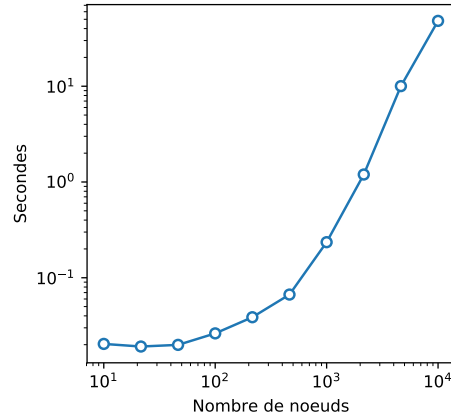


FIGURE 1.5 – Temps d’intégration de la dynamique Lotka-Volterra généralisée pour différentes tailles de réseau. Pour chaque taille de réseau, le temps d’intégration est moyenné sur 10 simulations effectuées avec l’intégrateur *dopri5* avec $dt = 10^{-2}$ et 1000 pas de temps. Les calculs ont été faits sur un ordinateur MacBook Pro 2018 de processeur 2,3 GHz Intel Core i5.

1.4 Méthodes de prédiction

Quelle est la relation entre la structure et la dynamique sur réseaux ? Nous répondons à cette question en étudiant l’activité dynamique sur les réseaux. Plus précisément, il faudra être en mesure de faire des prédictions des états dynamiques pour une structure donnée. Cela nous permettrait à la fois d’estimer l’impact d’une perturbation sur la dynamique, mais également de détecter une perturbation structurelle. Cette direction guidera les travaux de cette thèse.

Prédire l’activité dynamique sur un réseau réel est un défi d’envergure, car nous sommes limités par les données existantes. Parfois, la dynamique nous est pratiquement inconnue et des traces d’activité sont les seuls témoins de la dynamique sur réseau. Dans d’autres cas, nous pouvons raisonnablement modéliser l’activité avec des dynamiques mécanistiques. C’est notamment le cas des dynamiques de contagions qui expliquent certaines données réelles [11, 21].

Dans d’autres contextes, il s’agit plutôt de la structure qui n’est pas parfaitement connue. Par exemple, à défaut de connaître les réseaux réels de contacts sociaux, les travaux en épidémiologie sont faits sur des réseaux de contacts synthétiques (réseaux aléatoires) [12]. Ces structures aléatoires reproduisent certaines propriétés statistiques des systèmes réels, comme la distribution des degrés ou le nombre moyen de contacts. Cela montre qu’il n’est pas forcément nécessaire de connaître parfaitement la structure du système à l’étude et qu’une connaissance partielle de ses propriétés peut être suffisante pour des prédictions de bonne qualité.

Nous distinguons donc quatre contextes :

1. modèle dynamique et structure connus ;
2. modèle dynamique et structure partiellement connus ;

3. modèle dynamique inconnu et structure connue ;
4. modèle dynamique et structure inconnus.

Nous avons omis le contexte où la dynamique est connue et la structure est totalement inconnue, e.g., la matrice d'adjacence, le degré moyen et la distribution des degrés sont inconnus. La raison est simple, ce contexte peut être facilement contourné en composant avec des méthodes issues des quatre contextes exposés. Par exemple, une structure effective peut être inférée à partir d'approches bayésiennes présentées à la sous-section 1.4.4, *modèle dynamique et structure inconnus*, pour ainsi placer le problème sous le paradigme *modèle dynamique et structure connue*. Une autre alternative est d'ignorer l'information de la dynamique et traiter le système directement sous *modèle dynamique et structure inconnus*.

Pour chaque contexte, les outils de prédiction sont très différents. Pour le reste du chapitre, nous présenterons les formalismes existants de chaque contexte.

1.4.1 Modèle dynamique et structure connus

Pour certains systèmes réels, la structure du réseau \mathbf{W} est parfaitement connue : il existe plusieurs bases de données de structures réelles [36]. On considère maintenant que l'activité sur le réseau peut être correctement approximée par un modèle de dynamique. Dans ce contexte, la réponse dynamique à la structure semble facilement quantifiable.

Dans ce premier segment, on considère que la structure du réseau est parfaitement connue via la matrice \mathbf{W} et le modèle de la dynamique également, c.-à-d. les mécanismes dynamiques, exprimés sous la forme d'équations différentielles ou de processus stochastiques, et tous les paramètres du modèle sont connus. Pour décrire la réponse dynamique du système, on doit mettre en lumière les états d'activités pour la structure \mathbf{W} et pour une structure perturbée \mathbf{W}' .

Pour les dynamiques stochastiques (Cascade, SIS, ...), deux pistes de solutions sont possibles. La première est la méthode numérique brute Monte-Carlo. On simule la dynamique à plusieurs reprises et on calcule les distributions qui nous intéressent du comportement dynamique comme la marginale sur le nombre de noeuds actifs. La Sec. 1.4.1 présentera comment réaliser ce travail efficacement. La deuxième méthode consiste à solutionner directement un système d'équations des marginales du système (Sec. 1.4.1). Ce travail se réalise souvent en contexte de passage de messages où la probabilité d'être dans un certain état est propagée jusqu'à atteindre une solution stationnaire.

Pour les dynamiques déterministes (Wilson-Cowan, Lotka-Volterra ...), on peut prendre l'approche de solution par intégrateurs numériques. On travaille alors à l'aveugle, car il faut fixer quelques hyperparamètres (temps d'intégration, type d'intégrateur ...) et il existe une infinité de conditions initiales. Cette direction est hasardeuse et computationnellement coûteuse.

L'autre approche est de suggérer des arguments analytiques pour quantifier la dynamique, par exemple via la stabilité du système [112].

Solution MCMC du modèle SIS

Dans le premier cas d'étude, nous allons présenter une solution par simulations numériques d'une dynamique stochastique sur réseaux. Pour utiliser cette solution, il faut connaître exactement la structure \mathbf{W} et le mécanisme de la dynamique. La grande force des simulations est qu'elles sont très précises et quasi-exactes si correctement effectuées. Elles peuvent être utilisées pour décrire l'impact de la structure sur la dynamique, mais sont également essentielles pour valider des solutions de méthodes approximatives.

En général, il est difficile d'échantillonner *efficacement* les dynamiques stochastiques. Pour le faire rapidement, nous tentons souvent de réduire au maximum le nombre de tirages. Cela se traduit en minimisant le nombre de tirages inutiles (rejets) sans altérer l'exactitude des distributions a posteriori. Nous illustrerons ce propos avec l'exemple de la dynamique SIS.

Algorithme de rejet - Lorsqu'on programme pour la première fois un échantillonneur de la dynamique SIS, on pense d'abord à la méthode de rejet par temps discrets. On définit alors un pas de temps $\Delta t < 1$. À chaque pas, les noeuds actifs ont une probabilité Δt de récupérer et une probabilité $\gamma\Delta t$ de transmettre l'infection. Il faut donc passer à travers tous les évènements possibles et tirer individuellement s'ils doivent être exécutés. Le paramètre Δt doit être suffisamment petit pour qu'il y ait environ un seul évènement exécuté sur tout le réseau à chaque pas de temps.

Le problème de l'algorithme de rejet est que Δt doit être très petit et qu'il peut y avoir plusieurs temps sans évènements. Plus précisément, à chaque pas de temps, $N(1 - \rho) + \langle k \rangle \rho$ évènements possibles d'infection et de récupération sont échantillonnés. Le nombre de tirages est environ égal à $N(1 - \rho) + \langle k \rangle \rho$ et est aussi environ égal au nombre d'évènements rejetés (car 1 évènement réussi par Δt). La méthode est computationnellement très coûteuse.

Algorithme de Doob-Gillepsie - [53] La méthode de Doob-Gillepsie standard échantillonne les évènements directement. L'idée principale est de se déplacer temporellement d'un évènement exécuté à l'autre.

Chaque évènement possible v de transition ou de récupération sur le réseau se fait avec un taux w_v . Le délai entre deux évènements τ suit une distribution exponentielle

$$P(\tau) \propto e^{-W\tau}, \tag{1.50}$$

où $W = \sum_v w_v$ est la somme de tous les taux des évènements. Il nous suffit donc de tirer le délai selon (1.50), et ensuite de choisir l'évènement à exécuter, proportionnellement à w_v .

Après chaque évènement et avant de recommencer le tirage, les taux w_v et la banque des évènements possibles sont mis à jour.

En général, l'algorithme par évènements est suffisamment rapide pour échantillonner des réseaux de taille $N > 10^6$ en des temps raisonnables sur un ordinateur personnel. Mentionnons qu'il existe des algorithmes plus performants, mais plus complexes à programmer [157].

Passage de messages sur réseaux

Certains modèles dynamiques ont une formulation sous la forme d'équations de passage de messages, mieux connu sous le terme passage de messages (MP). Cette technique est pratiquement équivalente à simuler la dynamique car elle propage un message sur le graphe de la même manière que l'activité est propagée. La différence est que le message est une quantité d'intérêt comme une probabilité d'activation. Par rapport aux simulations MCMC, les équations MP peuvent être plus rapides à converger. Toutefois, si comparées à des approches en champ moyen, que nous explorons à la Sec. 1.4.2, les équations MP peuvent s'avérer beaucoup moins efficaces.

La grande force d'un système MP est que la structure est explicitement prise en compte, car les quantités sont propagées à même le graphe. Elles ne nécessitent pas d'imposer la limite des tailles infinies $N \rightarrow \infty$. Pour illustrer le caractère général de l'approche, nous allons développer une solution MP à la dynamique de cascade (Sec. 1.2.1).

Passage de messages sur une dynamique de cascade

Au terme de la cascade, on considère la probabilité $q_{j \rightarrow i}$ que le noeud j soit actif sachant que le noeud i est inactif. Avec une probabilité $1 - \rho_0$, le noeud j ne fait pas partie des noeuds initialement actifs. Dans ce cas et comme le processus est irréversible, le noeud i ne peut avoir activé le noeud j par le passé puisqu'il serait lui-même encore actif. Ainsi, le noeud j ne peut qu'avoir été activé par un membre de son voisinage autre que i . Par conséquent, la probabilité $q_{j \rightarrow i}$ est obtenue en considérant tous les états possibles d'activité du voisinage de j avec i inactif,

$$q_{j \rightarrow i} = \rho_0 + (1 - \rho_0) \sum_{\{x_l\}: l \in \mathcal{N}_j \setminus i} C \left(\frac{\sum_r x_r}{k_j} \right) \prod_{r \in \mathcal{N}_j \setminus i} q_{r \rightarrow j}^{x_r} (1 - q_{r \rightarrow j})^{1 - x_r}, \quad (1.51)$$

où l est un élément de l'ensemble composé des voisins du noeud j excluant le noeud i . En solutionnant l'équation (1.51), on obtient une probabilité $q_{j \rightarrow i}$ pour chaque lien dirigé. On trouve ensuite la probabilité que le noeud i soit actif en considérant tous les états possibles du voisinage du noeud i ,

$$\rho^{(i)} = \rho_0 + (1 - \rho_0) \sum_{\{x_l\}: l \in \mathcal{N}_i} C \left(\frac{\sum_r x_r}{k_i} \right) \prod_{r \in \mathcal{N}_i} q_{r \rightarrow i}^{x_r} (1 - q_{r \rightarrow i})^{1 - x_r}. \quad (1.52)$$

Bien qu'exacte sur les structures parfaitement en arbre, cette probabilité n'est en pratique calculable que pour les petits réseaux avec cycles, c.-à-d. pour $N < 50$. Le nombre de termes dans les équations (1.51) et (1.52) croît exponentiellement $\sim 2^{k_j}$. Notons toutefois qu'en supposant un réseau aléatoire de distribution de degrés $\text{Pr}(k)$ et de taille infinie, les équations (1.51) et (1.52) se réduisent aux équations (1.24) qui s'évaluent numériquement.

Malgré sa complexité, le cadre théorique de MP permet d'estimer la position de la transition de phase de l'activité du réseau (voir Fig. 1.2). À la transition, on estime que $q_{j \rightarrow i} \rightarrow 0$ et que $\rho_0 \rightarrow 0$. On linéarise le système (1.51) en supposant que le noeud j ait exactement un seul voisin actif, c.-à-d. que les termes $(1 - q_{l \rightarrow j})^{1-x_j} \rightarrow 0$, et on déduit que

$$q_{j \rightarrow i} = \sum_{l \in \mathcal{N}_j} C \left(\frac{a}{k_j} \right) B_{i \rightarrow j, j \rightarrow l} q_{l \rightarrow j}, \quad (1.53)$$

où \mathbf{B} est la matrice de non-retour définie sur les liens du graphe et de taille $2M \times 2M$ [65],

$$B_{k \rightarrow l, i \rightarrow j} = \begin{cases} 1, & \text{si } j \neq k \text{ et } i \neq l, \\ 0, & \text{autrement.} \end{cases} \quad (1.54)$$

La matrice de non-retour assure que le seul noeud actif l du voisinage de j n'est pas i . La solution de l'équation (1.53) se calcule par itération comme suit :

$$\mathbf{q}_n = \mathbf{K} \mathbf{B} \mathbf{q}_{n-1} \quad (1.55)$$

où \mathbf{K} est une matrice diagonale, $K_{i \rightarrow j, i \rightarrow j} = C(k_j^{-1})$. Pour qu'il y ait propagation de l'activité, le système (1.55) doit être instable, c.-à-d. que les valeurs propres du système itératif soient plus grandes que 1. On obtient alors le critère pour l'existence d'une composante géante active,

$$\max[\text{eig}(\mathbf{K} \mathbf{B})] > 1, \quad (1.56)$$

qui s'évalue rapidement même pour de grands réseaux. Notons que \mathbf{B} a un spectre différent de la matrice d'adjacence \mathbf{A} , mais qui est borné par la plus grande valeur propre de \mathbf{A} , c.-à-d. que $\max[\text{eig}(\mathbf{B})] \leq \max[\text{eig}(\mathbf{A})]$ [95]. Numériquement, il a été observé que le seuil de l'équation (1.56) est plus près que celui obtenu par le cadre théorique en champ moyen [Sec. 1.2.1].

L'approche par MP donne un critère de propagation d'une cascade sur un réseau donné. Le critère prend en compte la distribution des seuils d'activation, donnée par la matrice \mathbf{K} , et la structure du réseau, donnée par \mathbf{B} .

En conclusion à cette section, nous ouvrons sur d'autres utilisations du MP sur la dynamique SIS. Pour une dynamique réversible, il est naturel d'étendre le cadre théorique par champ

moyen (*quenched mean-field*) [Eq. (1.28)] au MP, à l'aide des équations

$$\dot{y}_i = -y_i + \gamma(1 - y_i) \sum_{j \in N_i} y_{ji}, \quad (1.57)$$

$$\dot{y}_{j \rightarrow i} = -y_{j \rightarrow i} + \gamma(1 - y_j) \sum_{v \in N_j \setminus i} y_{v \rightarrow j}, \quad (1.58)$$

où $y_{j \rightarrow i}$ est la probabilité que le noeud j soit infecté par un ses voisins diffèrent du noeud i , et y_i la probabilité que le noeud i soit infecté. Le prix à payer est l'ajout de $2M$ équations, bien qu'il soit encore possible de solutionner ce système sur des réseaux moyens ($N \approx 10^3$). Toutefois, comme la dynamique est récurrente, le cadre théorique MP n'est pas considérablement plus exact que le champ moyen standard. On a donc peu d'incitatifs à solutionner ces $2M$ équations supplémentaires [154].

Les équations de MP admettent un point critique. En linéarisant autour de $y_{j \rightarrow i} = 0$, on obtient le seuil critique pour que l'activité soit maintenue,

$$\gamma_c^{\text{MP}} = \frac{1}{\max[\text{eig}(\mathbf{B})]}, \quad (1.59)$$

lequel dépend de la matrice de non-retour \mathbf{B} . Comme \mathbf{B} est non négative et irréductible, elle satisfait le théorème de Perron-Frobenius et sa valeur propre dominante est alors positive [30]. On rappelle que le seuil $\gamma_c^{\text{MF}} = 1/\max[\text{eig}(\mathbf{A})]$ obtenu avec le formalisme *quenched* [Eq. (1.30)]. Comme la valeur propre dominante de \mathbf{B} est bornée par celle de la matrice d'adjacence, alors $\gamma_c^{\text{MF}} \leq \gamma_c^{\text{MP}}$.

Le formalisme de MP peut être difficile à solutionner. Le nombre d'équations croît avec le nombre de liens, rendant l'approche impraticable sur de grands réseaux avec cycles. Toutefois, les seuils de transitions sont facilement évaluables et se rapprochent de ce qui est numériquement observé.

...

1.4.2 Modèle dynamique et structure partiellement connu

Supposons que la structure du réseau soit partiellement connue via la distribution des degrés. Les méthodes MCMC et de MP présentées ne s'appliquent plus⁶ et nous devons utiliser des méthodes approximatives. Notons également que même si la structure est parfaitement connue, il est numériquement avantageux d'utiliser des techniques approximatives si les réseaux sont grands. En effet, la capacité de prédiction d'approches approximatives est plus limitée qu'une méthode exacte, mais hautement plus interprétable.

6. Il existe toutefois des variantes. Voir Refs. [87, 88].

Plusieurs méthodes approximatives regroupent les noeuds du réseau en classes d'équivalence. Par exemple, en admettant que tous les noeuds ayant un même degré aient environ le même état dynamique, alors un système à N individus se traduit en un système avec $k_{\max} + 1$ compartiments. Cette approche est fréquemment utilisée pour analyser la dynamique SIS. Les classes de degrés fonctionnent généralement très bien pour les réseaux aléatoires, car le degré d'un noeud est un bon indicateur de son état dynamique. Plus précisément, les noeuds de même degré ont des états dynamiques similaires et partagent des voisinages similaires. Si un réseau possède des corrélations additionnelles, il pourrait être nécessaire de distinguer les noeuds de même degré.

En théorie, les classes d'équivalence peuvent être basées selon n'importe quelle propriété. Mais le travail est plus facile si la propriété choisie est facilement mesurable, discrète et si elle est une bonne indicatrice de la dynamique de chaque groupe. Le degré est donc un candidat idéal. Le degré est discret et possède k_{\max} compartiments. Numériquement, on observe aussi que le degré forme en général un très bon indicateur de l'activité dynamique dans les compartiments.

On ne compte plus le nombre de publications qui emploient un formalisme en champ moyen par degré. Notamment pour les dynamiques SIS où cette avenue a permis une réelle percée dans notre compréhension des phénomènes critiques [110, 158]. Dans cette thèse, nous mettrons de côté cette direction en nous intéressant plutôt à une forme plus compacte. Plutôt que de choisir une propriété structurelle pour regrouper les noeuds, nous explorerons l'idée de créer une classe globale et de trouver les transformations qui permettent de relier les noeuds à cette classe effective.

Réduction universelle

En poussant la théorie des classes d'équivalences vers ses limites, on se rend à construire un seul noeud effectif qui possède certaines propriétés dynamiques (bifurcations, stabilité, etc.) similaires, à une transformation linéaire près, à celles des noeuds du réseau original. Cette approximation est parfois présentée sous la forme d'une réduction de dimensionnalité, car on utilise un système unidimensionnel pour décrire un système N -dimensionnel. En réalité, on peut aussi y voir un formalisme de champ moyen par compartimentage où tous les noeuds ont été placés, sous une certaine transformation, dans la même classe d'équivalence. Pour présenter la démarche, les réseaux seront considérés pondérés et dirigés.

En 2016, un système unidimensionnel de la sorte a été proposé pour les dynamiques déterministes [52]. On considère la dynamique suivante pour l'activité x_i noeud i :

$$\dot{x}_i = F(x_i) + \sum_{j=1}^N w_{ij} G(x_i, x_j), \quad (1.60)$$

où $F(\cdot)$ et $G(\cdot, \cdot)$ sont des fonctions d'activation arbitraires et spécifiques à la dynamique choisie. La dynamique complète se réduit à l'équation effective

$$\dot{R} = F(R) + \alpha G(R, R), \quad (1.61)$$

où R et α sont des observables de l'activité x_i et de la structure \mathbf{W} respectivement. Pour parvenir à ce résultat, il faut faire certaines approximations et hypothèses. La démonstration originale utilise une notation lourde qui obscurcit les approximations et les hypothèses sous-jacentes [52]. En quelque sorte, il est un peu ardu de bien comprendre les implications des approximations qui ont été faites.

On remarque d'abord de (1.60) qu'un noeud ayant k connexions apparaîtra dans $k + 1$ équations, c.-à-d. dans celles régissant l'activité de ses k voisins et la sienne. Ainsi, les noeuds de plus hauts degrés sont plus centraux dans la résolution de la dynamique. On produit donc une observable linéaire pondérée par les degrés des noeuds,

$$R = \frac{1}{c} \sum_{i=1}^N s_i^{\text{out}} x_i, \quad (1.62)$$

où $s_i^{\text{out}} = \sum_{j=1}^N w_{ji}$ est le force sortante de i et $c = \sum_{i=1}^N s_i^{\text{out}}$ est un facteur de normalisation. En multipliant l'équation (1.60) par s_i^{out}/c et en sommant sur i , on déduit l'équation

$$\dot{R} = \frac{1}{c} \sum_{i=1}^N s_i^{\text{out}} \dot{x}_i = \frac{1}{c} \sum_{i=1}^N s_i^{\text{out}} F(x_i) + \frac{1}{c} \sum_{i=1}^N \sum_{j=1}^N s_i^{\text{out}} w_{ij} G(x_i, x_j). \quad (1.63)$$

On estime d'abord que

$$\frac{1}{c} \sum_{i=1}^N s_i^{\text{out}} F(x_i) \approx F\left(\frac{1}{c} \sum_{i=1}^N s_i^{\text{out}} x_i\right) = F(R), \quad (1.64)$$

qui est exacte si $F(x)$ est linéaire⁷. Pour le terme d'interactions, on suppose ensuite que

$$\sum_{i=1}^N \sum_{j=1}^N s_i^{\text{out}} w_{ij} G(x_i, x_j) \approx \alpha G(R, R) \quad (1.65)$$

avec

$$\alpha = \frac{1}{c} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N w_{ji} w_{ik} = \frac{\mathbf{1}^T \mathbf{W} \mathbf{s}^{\text{in}}}{\mathbf{1}^T \mathbf{W} \mathbf{1}}. \quad (1.66)$$

On obtient alors la dynamique unidimensionnelle

$$\dot{R} = F(R) + \alpha G(R, R). \quad (1.67)$$

7. En développant $F(x_i) = F(R) + (x_i - R)F'(R) + O[(x_i - R)^2]$, alors la somme sur les i pondérée par s_i^{out}/c fait disparaître le terme linéaire. En général, les termes quadratiques seront non nuls sauf si $F(x)$ est parfaitement linéaire.

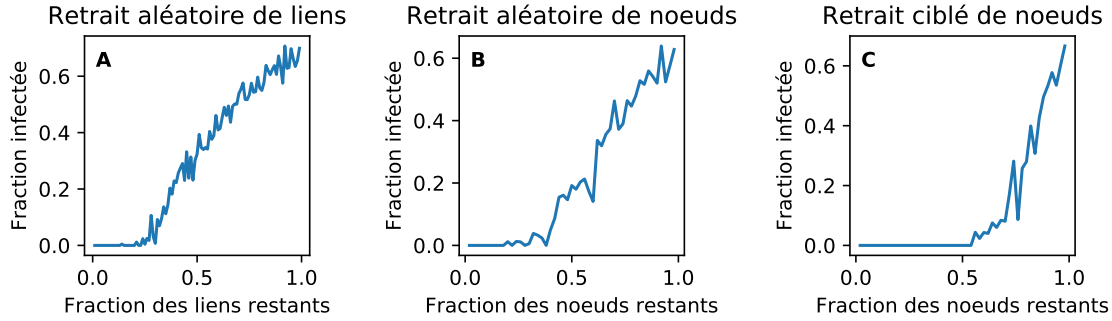


FIGURE 1.6 – Exemples des bifurcations pour différentes stratégies d’attaque sur des réseaux d’attachement préférentiel avec $N = 50$ et $M = 100$, et pour une dynamique SIS de $\alpha = 0.1$ et $\beta = 0.1$. En A, on retire des liens aléatoires du réseau. En B, on retire aléatoirement des noeuds. En C, on retire des noeuds qu’on choisit proportionnellement à leurs degrés. Les figures sont obtenues en simulant la dynamique avec l’algorithme de rejet pour $t = 150$ et $\Delta t = 1$. La fraction infectée est calculée comme la fraction moyenne d’individus infectés entre $t = 145$ et $t = 150$. Chaque point de données est la moyenne de 5 simulations avec les mêmes paramètres d’attaque.

L’utilisation d’une dynamique unidimensionnelle change notre vision de la résilience. D’abord, la prédiction est maintenant indépendante du type de perturbation. Par exemple, l’état d’activité est souvent présenté en fonction du nombre de noeuds ou de liens retirés [Voir un exemple à la Fig. 1.6]. Évidemment, le nombre de noeuds et de liens nécessaires pour observer une transition critique de l’activité n’est pas égal puisque le retrait de noeuds est plus dommageable que le retrait de liens. De plus, si l’attaque est ciblée (e.g., en retirant des noeuds de hauts degrés) alors le nombre minimal de noeuds pour atteindre une transition est encore plus faible. Toutes ces considérations rendent l’étude de la résilience spécifique au type de perturbations. En utilisant une approche effective, on élimine la nature des perturbations. Il nous est simplement possible de calculer un α critique. Il y aura bifurcation dès que la perturbation mène le réseau sous ce paramètre critique.

Ensuite, cette méthode est, en théorie, indépendante du type de réseau. Que le réseau soit aléatoire ou issu du modèle d’attachement préférentiel, le paramètre critique est le même.

Enfin, on note aussi qu’une approche unidimensionnelle permet une analyse très simple du diagramme de bifurcations.

Commentaire sur la stabilité de la dynamique réduite – Les dynamiques compétitives et mutualistes diffèrent sur plusieurs aspects. Notamment pour les dynamiques compétitives où du chaos et des cycles limites peuvent apparaître pour deux dimensions et plus, alors que les dynamiques mutualistes convergent en général vers des solutions stables. Il convient donc de préciser la validité d’un système unidimensionnel dans ces deux scénarios.

On considère la dynamique linéaire

$$\dot{x}_i = -x_i + g \sum_{j=1}^N w_{ij} x_j, \quad (1.68)$$

où $g = +1$ lorsque mutualiste et $g = -1$ lorsque compétitif. Sous forme matricielle,

$$\dot{\mathbf{x}} = \mathbf{M}\mathbf{x}, \quad (1.69)$$

où $\mathbf{M} = -\mathbf{I}_N + g\mathbf{W}$ avec \mathbf{I}_N la matrice unité de dimension $N \times N$. Si la plus grande valeur propre de \mathbf{M} est positive, alors la dynamique sera instable et l'activité des noeuds divergera. Ainsi, on supposera que la plus grande valeur propre de \mathbf{W} , qu'on notera λ_D , est inférieure à 1 pour éviter que l'activité ne diverge⁸.

La dynamique réduite représentant l'équation (1.68) est

$$\dot{R} = (-1 + g\alpha)R. \quad (1.70)$$

On retrouve un paramètre de stabilité $\lambda_g = -1 + g\alpha$. Si $\lambda_g > 0$, alors la dynamique réduite sera instable.

Par certains résultats de la théorie spectrale des graphes [30], on a l'inégalité suivante :

$$\langle s \rangle \leq \lambda_D \leq s_{\max}. \quad (1.71)$$

Comme α est donné par la force moyenne du voisinage,

$$\alpha = \frac{\langle s^2 \rangle}{\langle s \rangle} \quad (1.72)$$

et que $\langle s \rangle \leq \langle s^2 \rangle / \langle s \rangle$ ⁹, alors $\alpha \leq \lambda_D$. Comme $\lambda_D \leq 1$, alors $\lambda_g = -1 + g\alpha \leq 0$. Par conséquent, la dynamique réduite de l'équation (1.70) converge toujours vers un point fixe stable pour $|g| \leq 1$.

Lorsque la dynamique est mutualiste $g = +1$, alors la valeur propre de λ_D est une bonne approximation de α et donc $\lambda_+ = -1 + \alpha \approx -1 + \lambda_D$. Ainsi, λ_+ est environ la plus grande valeur propre de la matrice \mathbf{M} . On en conclut donc que λ_+ est un bon indicateur de la stabilité [Fig.1.7(a)]. Plus encore, la dynamique complète est stable seulement si la dynamique réduite l'est. Toutefois, l'observable R est la projection de \mathbf{x} par rapport à la direction la moins stable, c.-à-d. λ_+ est la moins négative, et la plus lente, c.-à-d. λ_+ a le plus petit écart avec 0, de la dynamique.

Lorsque la dynamique est compétitive $g = -1$, la valeur propre dominante de \mathbf{M} est donnée par $-1 - \lambda_S$ où λ_S est la valeur propre la plus négative du spectre de \mathbf{W} . Par conséquent,

8. \mathbf{W} peut toujours être remis à l'échelle pour satisfaire cette condition.

9. Un résultat qui découle de $\langle s^2 \rangle \geq \langle s \rangle^2$ et qui est connu sous le paradoxe de l'amitié. Vos amis ont en moyenne plus d'amis que vous.

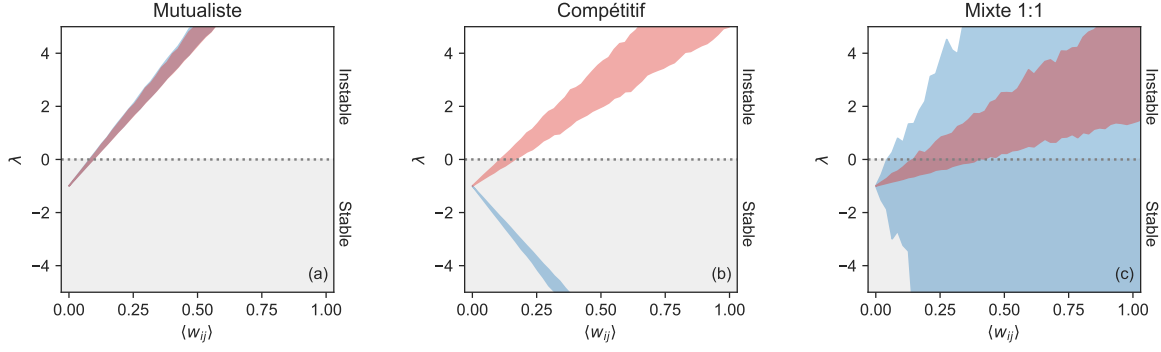


FIGURE 1.7 – Exemple de stabilité de la dynamique réduite pour des contextes mutualiste, compétitif et mixte. Pour la famille des Erdős-Rényi de $N = 10, p = 0.4$, (rouge) l’intervalle du 10-90 percentiles de la valeur propre dominante de $\mathbf{M} = -\mathbf{I}_N + g\mathbf{W}$, où $\mathbf{W} = \mathbf{A}\langle w \rangle$, et (bleu) pour la dynamique réduite $\lambda_g = -1 + g\alpha$ pour différentes valeurs d’interaction moyennes $\langle w_{ij} \rangle$. Les contextes de dynamiques sont mutualistes ($g = +1$), compétitifs ($g = -1$) et mixtes. [Figure inspirée de [10]]

comme $\alpha \approx \lambda_D$, alors $\lambda_- = -1 - \alpha$ est la valeur propre la plus négative du spectre de \mathbf{M} . On en conclut donc que λ_- est la valeur propre la moins indicatrice de la stabilité de la dynamique complète [Fig.1.7(b)].

Pour bien comprendre l’implication du dernier résultat, on considère un réseau prédateur-proie où on assigne aléatoirement les rôles de prédateurs et de proies en utilisant une matrice \mathbf{W} telle que $w_{ij} \in \{-1, 0, 1\}$. Selon la définition de α , le dénominateur $\mathbf{1}^T \mathbf{W} \mathbf{1}$ est environ nul si le nombre de liens de prédateurs et de proies est similaire. Le retrait d’un seul lien peut faire basculer la valeur de α et la solution du système réduit.

Commentaire sur la richesse dynamique du système réduit – En optant pour une réduction unidimensionnelle plutôt qu’une dynamique à N -dimensions, on ne peut généralement pas décrire les oscillations qui peuvent survenir. Pour illustrer le propos, on considère la dynamique de prédation suivante :

$$\dot{x}_i = x_i(1 - x_i) + g \sum_{j=1}^N x_i w_{ij} x_j. \quad (1.73)$$

La dynamique réduite est

$$\dot{R} = R - (1 - g\alpha)R^2. \quad (1.74)$$

À équilibre $\dot{R} = 0$, la dynamique réduite admet la solution triviale $R^* = 0$ et

$$R^* = \frac{1}{1 - g\alpha}. \quad (1.75)$$

Comme la dynamique réduite est unidimensionnelle, elle ne peut admettre aucune forme d’oscillation et n’indique rien à propos du nombre de solutions stables \mathbf{x}^* pour une même valeur de α . Une gamme de complexité est perdue au profit d’une solution effective qui décrit plusieurs solutions de la dynamique complète. Il n’y a aucune manière de savoir si la dynamique complète oscille à partir de la dynamique réduite.

Utilisation du modèle réduit – Le modèle réduit de Gao *et al.* ne se limite pas à faire une prédiction sur l’activité globale du système. Il peut également être utilisé pour classifier la résilience des systèmes en fonction de leur structure, sans doute un des aspects les plus intéressants du formalisme.

On considère le paramètre structurel α [Eq. (1.66)] :

$$\alpha = \frac{\mathbf{1}^T \mathbf{W} \mathbf{s}^{\text{in}}}{\mathbf{1}^T \mathbf{s}^{\text{in}}} = \frac{\langle s^{\text{in}} s^{\text{out}} \rangle}{\langle s \rangle} \quad (1.76)$$

$$= \langle s \rangle + \frac{\langle s^{\text{in}} s^{\text{out}} \rangle}{\langle s \rangle} - \langle s^{\text{in}} \rangle \quad (1.77)$$

en acceptant que $\langle s \rangle = \langle s^{\text{in}} \rangle = \langle s^{\text{out}} \rangle$. On écrit cette expression sous la forme

$$\alpha = \langle s \rangle + SH, \quad (1.78)$$

où

$$\langle s \rangle = \frac{\mathbf{1}^T \mathbf{W} \mathbf{1}}{\mathbf{1}^T \mathbf{1}}, \quad (1.79)$$

$$S = \frac{\langle s^{\text{in}} s^{\text{out}} \rangle - \langle s^{\text{in}} \rangle \langle s^{\text{out}} \rangle}{\sigma^{\text{in}} \sigma^{\text{out}}}, \quad (1.80)$$

$$H = \frac{\sigma^{\text{in}} \sigma^{\text{out}}}{\langle s \rangle}, \quad (1.81)$$

avec σ^{in} et σ^{out} l’écart-type de \mathbf{s}^{in} et \mathbf{s}^{out} respectivement. Le terme $\langle s \rangle \in [0, 1]$ mesure la densité du réseau. La variable $S \in [-1, 1]$ décrit la réciprocité du réseau. Un réseau non dirigé aura $S = 1$. Enfin, le paramètre H mesure l’hétérogénéité du réseau. Plus le paramètre H est grand et plus le réseau se rapproche d’une étoile.

En formulant le paramètre structurel par des paramètres interprétables comme la densité et l’hétérogénéité, on comprend mieux la résilience du réseau. À la Fig. 1.8, des réseaux réels de pollinisateur-plante non dirigés ont été placés dans l’espace $(\langle s \rangle, H)$. Pour une dynamique particulière, on calcule le seuil critique α . Il nous est alors permis de voir les réseaux qui se trouvent sous le seuil critique.

Enfin, l’utilisation d’un système réduit simplifie grandement l’analyse de la résilience. Elle universalise les types de perturbations et les familles de réseaux. Toutefois, elle se limite actuellement aux dynamiques mutualistes et s’applique mieux aux réseaux aléatoires dont le voisinage d’un noeud est le même. Dans le Chapitre 3, nous verrons comment le formalisme peut être amélioré en choisissant une nouvelle observable effective. De plus, nous étendrons les conclusions aux réseaux modulaires et bipartis.

...

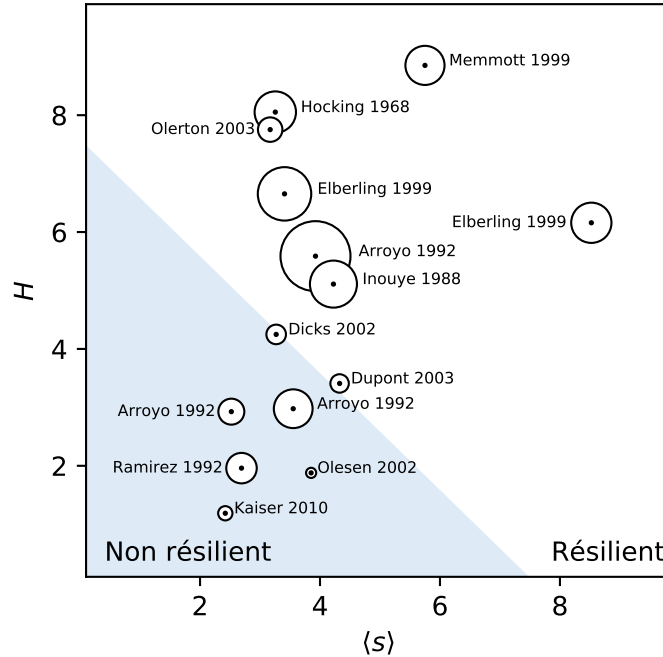


FIGURE 1.8 – Exemple de résilience prédite de réseaux réels pour une dynamique de populations donnée par l'équation (1.42) avec $B_i = 0.1, K_i = 1, C_i = 5, D_i = 5, E_i = 0.9$ et $H_i = 0.1$. La zone de résilience est définie comme l'espace de $(\langle s \rangle, H)$ où la dynamique n'admet qu'une solution stable. La zone de non-résilience est celle où il existe une solution stable non désirée. Les données sont tirées de Ref. [1]. La taille des points est proportionnelle à la taille du réseau.

1.4.3 Modèle dynamique et structure inconnus

Une partie de la complexité des systèmes complexes vient du fait qu'ils ont beaucoup de composantes. Par conséquent, nous n'avons pas toujours une bonne connaissance du réseau sous-jacent. Cela engendre que nous ne pouvons utiliser un modèle dynamique connu pour expliquer l'activité observée. Il s'agit possiblement du scénario le plus difficile à traiter, car tous les outils du paradigme des réseaux complexes semblent alors inutiles. Mais pas tout à fait. L'analyse des dynamiques non linéaires révèle qu'il existe des signes précurseurs quasi universels des catastrophes, détectables directement via la série temporelle. Ces signaux se groupent sous le paradigme des signaux précurseurs [*Early-warning signals* (EWS)].

Les plus connus des EWS sont le *critical slowing down* (CSD). Bien que nous présenterons seulement les CSD, il faut noter qu'il existe bien d'autres mesures [99]. La raison pour laquelle nous nous limitons à présenter les CSD est qu'ils sont les plus utilisés, mais aussi qu'ils s'appliquent aux dynamiques que nous considérons.

Ralentissement critique (Critical Slowing Down)

Le ralentissement critique, mieux connu sous le terme *Critical Slowing Down* (CSD), est le phénomène qui précède les bifurcations de type noeud-selle. Lorsque le système est structurellement perturbé et se dirige vers une transition critique, la réactivité de l'activité diminue d'où le concept de ralentissement critique [40, 41, 99]. Il est indicateur que le système est à l'aube d'une transition et qu'une faible perturbation structurelle est en mesure de changer drastiquement l'état du système.

En pratique, le ralentissement se détecte par l'observation de deux phénomènes :

1. augmentation de l'autocorrélation ;
2. augmentation de la variance.

On observe ces phénomènes en mesurant l'autocorrélation et la variance sur des fenêtres mobiles pour une série temporelle d'activité. L'autocorrélation indique si le temps précédent est un bon prédicteur, via une transformation linéaire, du prochain temps. Naturellement, si le système ralentit, alors les effets non linéaires originaux deviennent négligeables. Similairement, la variance augmente, car les perturbations s'accumulent sans récupération.

Le CSD repose aussi sur de solides fondements théoriques [173]. On considère l'équation dynamique

$$\dot{x}(t) = f(x, \theta), \quad (1.82)$$

où θ est un paramètre structurel. Le point fixe x^* solutionne l'équation

$$f(x^*, \theta) = 0. \quad (1.83)$$

La stabilité locale du point fixe x^* dépend de la pente

$$\lambda = \left. \frac{\partial f(x, \theta)}{\partial x} \right|_{x=x^*}. \quad (1.84)$$

Si $\lambda < 0$, alors x^* est stable.

Pour qu'il y ait une transition critique de l'activité, il doit exister une certaine bifurcation à un paramètre θ_c critique et plusieurs solutions stables. Au minimum, on suppose 2 solutions x_-, x_+ stables et une instable $x_- < x_c < x_+$. Nous sommes donc dans un contexte de bifurcation noeud-selle. La bifurcation survient lorsque $\theta = \theta_c$ et est telle que $x_- = x_c$. Par continuité, on en conclut que $\lambda = 0$ à θ_c .

Sachant que le retour à l'équilibre après une perturbation $x' = x + \epsilon$ décroît exponentiellement,

$$\epsilon(t) = ce^{\lambda t}, \quad (1.85)$$

alors la récupération devient très lente près de la transition puisque $\lambda \rightarrow 0^+$, c'est le CSD.

Pour comprendre pourquoi l'autocorrélation augmente, on considère le modèle régressif à un temps,

$$x_{n+1} - x^* = e^{\lambda\delta t}(x_n - x^*) + \sigma\epsilon_n \quad (1.86)$$

où $\sigma\epsilon_n$ est un bruit blanc. Le coefficient d'autocorrélation est $\alpha = e^{\lambda\delta t}$. L'espérance du modèle sur une série temporelle est

$$E[x_{n+1} - x^*] = \alpha E[x_n - x^*] + E[\sigma\epsilon_n]. \quad (1.87)$$

Pour un bruit blanc, $E[\sigma\epsilon_n] \rightarrow 0$ et donc l'autocorrélation $\alpha \rightarrow 1$ est maximale. Également, la variance du modèle tend vers l'infini :

$$\text{Var}(y_{n+1}) = E[(x_{n+1} - x^*)^2] - E[(x_n - x^*)]^2 = \frac{\sigma^2}{1 - \alpha^2} \rightarrow \infty. \quad (1.88)$$

La présence de CSD n'est pas universelle et se trouve à être fortement limitée aux bifurcations noeud-selle. La revue de Ref. [24] présente bien l'état des choses et mentionne une quantité de bifurcations qui n'admettent pas de CSD. On retrouve maintenant plusieurs études rapportant l'absence de CSD malgré la présence de catastrophes [22, 81, 82]. Par exemple, Wilkat *et al.* indiquent avoir été incapables de montrer l'existence de CSD avant les crises d'épilepsie, en analysant 105 séries temporelles d'activité neuronale pour 28 patients souffrant d'épilepsie [170].

D'un autre côté, les systèmes réversibles multidimensionnels peuvent présenter des signes de CSD malgré l'absence de catastrophes. Un exemple parlant est sans doute la bifurcation Hopf sous-critique (transition d'un point fixe stable à un cycle limite pour les systèmes à au moins deux dimensions). Près de la transition, le système a une valeur propre dominante nulle et ralentit. Par contre, la transition n'est pas catastrophique au sens où la valeur moyenne est continue et il n'y a pas d'hystérésis. Mentionnons également les bifurcations transcritiques où la solution stable non négative change graduellement. Malgré l'absence de catastrophes, les CSD sont bel et bien observées [89].

Dans le contexte multidimensionnel des systèmes complexes, il est possible que les bifurcations soient locales, e.g., extinction de certaines espèces, et donc que l'activité de la plupart des noeuds dans le réseau ne soit pas ralentie [40]. C'est le cas pour les réseaux modulaires où une communauté peut être perturbée sans que l'autre n'en soit affectée. Les CSD peuvent donc donner une mesure difficile à interpréter si seulement une fraction des noeuds sont ralentis. Par ailleurs, il est commun de mesurer la présence de CSD sur une observable agrégée comme l'activité moyenne du réseau. En procédant ainsi, on peut perdre les catastrophes locales ou simplement perdre la signature des CSD due au moyennage.

Un aspect important du CSD est que le système doit s'approcher de la transition critique pour être détectée. Lorsqu'on détecte les CSD, le système est déjà près de la transition. S'il fallait intervenir, la fenêtre temporelle peut être très petite. Les CSD ne sont donc pas une

solution pour suivre l'état global du système, mais peuvent seulement agir comme une alarme à l'approche d'une transition imminente. Sous cette optique, nous développerons au Chapitre 4 un outil, dans le contexte de dynamique inconnue, pour détecter les anomalies dynamiques bien avant d'approcher la transition de phase. Plus encore, nous identifierons la perturbation structurelle défailante.

En conclusion à cette section, le CSD n'est pas universel. Comme toute technique de prédiction, leur présence/absence n'est pas une certitude de la présence/absence d'une catastrophe imminente. Il faut donc agir avec prudence et tenter du mieux possible à se fier à d'autres techniques.

...

1.4.4 Modèle dynamique inconnu et structure connue

Abordons maintenant le scénario où le réseau et des séries temporelles de l'activité sont observés, mais le mécanisme dynamique est inconnu. Ce serait le cas si nous avions à analyser une série temporelle des populations d'un écosystème familier, c.-à-d. où le réseau trophique est connu, mais qu'on ne peut directement associer un modèle de populations réaliste. Nous nous retrouvons donc dans une situation difficile où nous cherchons à documenter la résilience dynamique du réseau sans aucun repère sur les bifurcations possibles de la dynamique sous-jacente. Une option est de choisir un modèle dynamique et d'en inférer les paramètres qui expliquent le mieux la série temporelle observée. Une fois l'inférence faite, on pourra se référer aux techniques présentées à la section *Modèle dynamique et structure connus* (Sec. 1.4.1).

Le problème discuté se formalise comme suit. La matrice d'adjacence \mathbf{W} de la structure est parfaitement connue. Une série temporelle unidimensionnelle pour chaque noeud, regroupée sous une matrice $\mathbf{X} \in \mathbb{R}^{N \times T}$ où T est le nombre d'observations, est également connue. On suppose que le paradigme des réseaux complexes dynamiques s'applique et que la structure a un effet quelconque sur la dynamique,

$$\mathbf{X} = \mathcal{M}(\mathbf{W}, \boldsymbol{\theta}, \mathbf{X}_0), \tag{1.89}$$

où $\boldsymbol{\theta}$ est un ensemble de paramètres cachés, $\mathbf{X}_0 \in \mathbb{R}^{N \times \tau}$ et $\mathcal{M}(\cdot)$ est un mécanisme dynamique. On note ici que \mathbf{X}_0 n'est pas forcément un vecteur et qu'il peut être de longueur τ . Il désigne le plus petit ensemble de conditions initiales requises pour déterminer l'évolution de la dynamique. Par exemple, on aura $\tau = 1$ pour une dynamique markovienne. En posant une forme à $\mathcal{M}(\cdot)$, il sera possible d'inférer les paramètres $\boldsymbol{\theta}$ afin d'utiliser $\mathcal{M}(\mathbf{W}, \boldsymbol{\theta}, \mathbf{X}_0)$ pour décrire la résilience dynamique. Le travail se résume alors à une tâche d'inférence dont on présente les méthodes de résolution.

MCMC

Pour tout problème d'inférence, la première question est : *est-ce que la vraisemblance du modèle est connue ?* Si la réponse est *non*, alors le problème se complexifie. Par contre, si la réponse est *oui*, on emploie une méthode *Monte-Carlo par chaîne de Markov* (MCMC) pour échantillonner l'espace des paramètres et en faire l'inférence dont on fait la description ici.

On suppose une distribution de vraisemblance $P(\mathbf{X}|\boldsymbol{\theta}, \mathbf{W})$ inscrivant la probabilité que \mathbf{X} a été généré sachant le mécanisme dynamique $\mathcal{M}(\cdot)$ et ses paramètres $\mathbf{W}, \boldsymbol{\theta}$.

Pour simplifier la notation, on ne notera plus \mathbf{W} dans les distributions statistiques puisqu'il s'y retrouve partout, e.g., $P(\mathbf{X}|\boldsymbol{\theta}, \mathbf{W}) = P(\mathbf{X}|\boldsymbol{\theta})$. On peut maintenant écrire l'équation de Bayes,

$$P(\boldsymbol{\theta}|\mathbf{X}) = \frac{P(\mathbf{X}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathbf{X})}, \quad (1.90)$$

où $P(\boldsymbol{\theta})$ est la probabilité *a priori* sur les paramètres et $P(\mathbf{X})$ est l'évidence de la série temporelle. On estime les paramètres réels $\boldsymbol{\theta}^*$ soit en maximisant la probabilité a posteriori

$$\hat{\boldsymbol{\theta}}^{\text{MAX}} = \operatorname{argmax}_{\boldsymbol{\theta}} P(\boldsymbol{\theta}|\mathbf{X}) \quad (1.91)$$

ou bien par l'espérance de la distribution a posteriori,

$$\hat{\boldsymbol{\theta}}^{\text{MMSE}} = \int \boldsymbol{\theta} P(\boldsymbol{\theta}|\mathbf{X}) d\boldsymbol{\theta} = \mathbb{E}[\boldsymbol{\theta}|\mathbf{X}]. \quad (1.92)$$

Il est d'ailleurs aisé de montrer que cette forme minimise l'erreur quadratique moyenne¹⁰ (d'où le titre estimateur MMSE - *minimum mean-squared error*),

$$\hat{\boldsymbol{\theta}}^{\text{MMSE}} = \min_{\hat{\boldsymbol{\theta}}} f(\hat{\boldsymbol{\theta}}) = \min_{\hat{\boldsymbol{\theta}}} \int (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^2 P(\boldsymbol{\theta}|\mathbf{X}) d\boldsymbol{\theta}. \quad (1.93)$$

Sans être une loi universelle, il est souvent suggéré de choisir l'estimateur $\hat{\boldsymbol{\theta}}^{\text{MMSE}}$ si le support le permet (si le support est discret, la marginale pourrait ne pas être une solution réaliste).

Approximate Bayesian Computation

Si la fonction de vraisemblance $P(\mathbf{X}|\boldsymbol{\theta})$ est difficile à écrire ou à évaluer, alors on ne peut pas évaluer la formule de Bayes et les estimateurs $\hat{\boldsymbol{\theta}}^{\text{MMSE}}$ et $\hat{\boldsymbol{\theta}}^{\text{MAX}}$ ne sont pas directement mesurables. La méthode *Approximate Bayesian Computation* (ABC) contourne ce problème en travaillant avec des fonctions de coût prédéfinies.

On suppose un mécanisme $\mathbf{X} = \mathcal{M}(\mathbf{W}, \boldsymbol{\theta}, \mathbf{X}_0)$. On échantillonne un ensemble de paramètres $\boldsymbol{\theta}$ selon une distribution *a priori* $P(\boldsymbol{\theta})$. Pour chaque échantillon, on génère une série temporelle

10. Il s'agit d'évaluer $0 = \frac{\partial}{\partial \theta_i} \int |\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}|^2 P(\boldsymbol{\theta}|\mathbf{X}) d\boldsymbol{\theta}$ pour tout i .

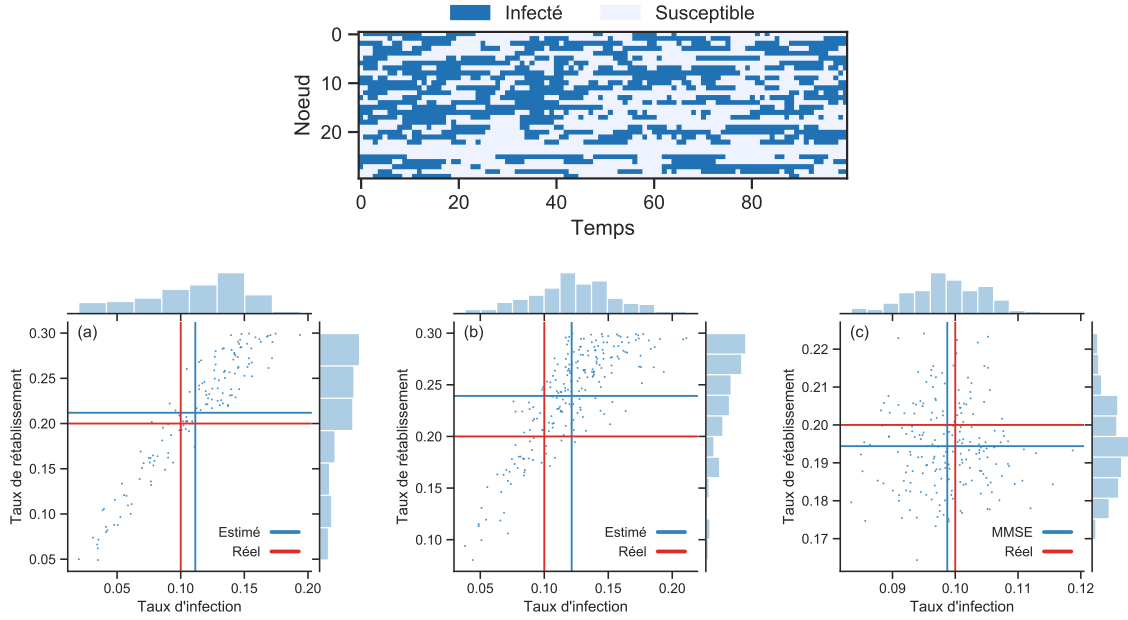


FIGURE 1.9 – En haut, la série temporelle générée de la dynamique SIS avec un taux d’infection $\alpha = 0.1$ et un taux de rétablissement $\beta = 0.2$ sur un réseau Erdős-Rényide $N = 30$ noeuds et $M = 65$ liens. En (a-b), distribution a posteriori $P(\alpha, \beta | \mathbf{X})$ échantillonnée avec la méthode ABC. En (a), la fonction de coût est $L(\mathbf{X}', \mathbf{X}) = |\langle \mathbf{X}' \rangle - \langle \mathbf{X} \rangle|$ alors qu’en (b), $L(\mathbf{X}', \mathbf{X}) = |\langle \mathbf{X}' \rangle - \langle \mathbf{X} \rangle| + |\text{Var}(\mathbf{X}') - \text{Var}(\mathbf{X})|$. En (c), distribution a posteriori estimée par MCMC *simulated annealing*.

$\mathbf{X}' = \mathcal{M}(\mathbf{W}, \boldsymbol{\theta}, \mathbf{X}_0)$ qu’on viendra comparer à la série temporelle originale selon une fonction de coût $L(\mathbf{X}', \mathbf{X})$. On accepte $\boldsymbol{\theta}$ si L est inférieure à un certain seuil ϵ . Si ϵ est suffisamment petit, alors il est attendu que la distribution a posteriori est estimée par

$$P(\boldsymbol{\theta} | \mathbf{X}) \propto P[L(\mathbf{X}' | \mathbf{X}) < \epsilon] P(\boldsymbol{\theta}). \quad (1.94)$$

L’estimateur statistique est donné soit par la maximisation du postérieur soit par la marginale.

Exemple de la dynamique SIS

Pour la dynamique SIS, la vraisemblance est connue en fonction des deux paramètres : le taux de transmission α et de rétablissement β . En effet,

$$P(\mathbf{X} | \boldsymbol{\theta}) = \prod_{t=1}^T \prod_{i=1}^N P(x_i(t) | \mathbf{X}(t-1)) \quad (1.95)$$

avec

$$P(x_i(t) | \mathbf{X}(t-1)) = \left[(1 - (1 - \alpha)^{m_i(t)}) ((1 - \alpha)^{m_i(t)(1-x_i(t))}) \right]^{1-x_i(t-1)} \left[(1 - \beta)^{x_i(t)} \beta^{1-x_i(t)} \right]^{x_i(t-1)} \quad (1.96)$$

où $m_i(t) = \sum_{j=1}^N w_{ij}x_j(t)$ est le nombre de voisins infectés au temps t . Pour échantillonner la distribution a posteriori, on utilise une approche numérique du type Metropolis-Hasting avec Simulated Annealing. Le détail des méthodes est laissé en référence [162]. À la Fig. 1.9, la distribution a posteriori est présentée pour une série temporelle issue d’une dynamique SIS. Si l’on compare ce résultat aux figures 1.9(a-b) obtenues par la méthode ABC, il est évident que la méthode MCMC donne une distribution a posteriori plus uniforme.

Les méthodes ABC souffrent du problème que le choix de la fonction de coût est critique malgré que peu d’indices sont disponibles pour guider notre sélection. Par exemple, en Fig. 1.9(a), le choix est malheureux, car plusieurs couples de paramètres (α, β) satisfont parfaitement la fonction de coût. En effet, le nombre moyen d’individus infectés est gouverné par le ratio α/β . Par conséquent, tous les couples (α, β) tels que $\alpha/\beta = \alpha^*/\beta^*$ seront généralement acceptés par la méthode ABC. Dans l’exemple, l’espace échantillonné est borné par $\alpha < 0.3$ et $\beta < 0.3$ et donc les marginales donnent de manière hasardeuse une bonne estimation des paramètres réels.

La méthode ABC et la formule de Bayes présupposent l’existence d’un mécanisme dynamique. Notre modélisation des séries temporelles est donc limitée par notre capacité à imaginer de nouvelles formes mathématiques pour ces mécanismes. Ainsi, l’expérience du modélisateur à poser un modèle approprié devient critique. Si le modèle est trop différent de la réalité, alors le pouvoir prédictif peut être quasi-nul. Rappelons-nous de l’exemple du virus Zika où un modèle classique SIS donnera un seuil de transmission qualitativement différent d’un modèle épidémiologique avec différents vecteurs de transmission [4].

Suite à ces commentaires, il semble se dégager un besoin évident d’un outil capable à la fois d’inférer un mécanisme effectif et ses paramètres via la structure du réseau et une observation de la dynamique. C’est possiblement ce que le paradigme de l’apprentissage automatique sur graphes est en mesure d’accomplir.

L’apprentissage sur graphes consiste à poser une forme mathématique hautement paramétrée qu’on optimise pour accomplir une tâche particulière [174]. Alors qu’il est commun de les voir résoudre des tâches de détection de communautés, recommandation de produits, classification, et de segmentation pour ne nommer que ceux-ci, ces machines proposent une solution intéressante au problème de modélisation dynamique [64, 163, 165, 179]. Au chapitre 4, nous emprunterons cet outil pour modéliser les séries temporelles structurées et en inférer les perturbations structurelles.

...

1.5 Projet de recherche

1.5.1 Problématique

L'étude de la résilience des réseaux complexes nécessite en général la prédiction des états stables dynamiques. Celle-ci peut être faite par une multitude de techniques. Il n'existe toutefois pas une méthode universelle qui est à la fois simple et précise pour toutes les structures et dynamiques.

Lorsque la dynamique et la structure sont connues, l'approche par passage de messages est à prioriser. Elle mène à une grande précision des résultats. De plus, elle peut être utilisée aussi bien pour les dynamiques récurrentes que non récurrentes. Par contre, lorsque la structure contient plusieurs cycles et s'éloigne de l'approximation en arbre, alors il peut exister des disparités entre les prédictions et les observations. Par conséquent, il semble y avoir un certain manque à combler si le seul objectif est de maximiser la qualité de la prédiction. Considérant que toutes les données (structure et dynamique) sont disponibles, il serait intéressant d'étudier une technique qui repousse cette limite imposée par les structures en arbre.

Pour simplifier l'analyse de la bifurcation des systèmes, nous avons vu l'approche de réduction dimensionnelle proposée par Gao *et al.* [52]. Cette avenue est très intéressante, car elle s'applique à plusieurs systèmes et l'exactitude des prédictions est satisfaisante malgré la simplicité de l'approche. Toutefois, elle n'est pas totalement valide dans le contexte où certaines espèces sont prédatrices [10]. Également, la démonstration mathématique présentée ne paraît pas suffisante pour conclure de sa validité pour les réseaux modulaires et à exposants d'échelle.

Il existe peu de techniques qui traitent des problèmes basés complètement sur les données et qui ne nécessitent pas l'utilisation de modèles explicites. Nous avons vu l'approche bayésienne qui suggère une inférence de paramètres à partir des observations de la dynamique. La grande lacune est la nécessité de poser une forme mathématique au modèle. Dans un contexte réel, la forme posée peut être simplement invalide ou encore trop simple pour la complexité du mécanisme. Il serait intéressant de lever cette inutile difficulté, le tout dans un contexte qui profite à l'analyse de la résilience.

1.5.2 Objectifs et guide de lecture

L'objectif général de cette thèse est de développer des outils mathématiques pour qualifier la résilience des réseaux complexes. À partir des problématiques énoncées, cette direction de recherche peut maintenant être précisée sous quatre sous-objectifs :

1. Développer une méthode très précise de prédiction des états stables pour les structures avec cycles et une dynamique connue ;

2. Valider mathématiquement l'approche de réduction dimensionnelle pour les réseaux modulaires et biparties ;
3. Développer une nouvelle approche de réduction dimensionnelle plus précise ;
4. Valider l'utilisation de l'apprentissage automatique pour les problèmes où la dynamique est inconnue.

Chapitre 2

Le premier objectif sera discuté au Chapitre 2. Pour ce faire, il faut d'abord sélectionner une dynamique précise. Nous utiliserons la cascade sur réseaux, car elle fait partie de celles pouvant être solutionnées par la méthode de passage de messages. Plus encore, la cascade n'est pas récurrente ce qui réduit la difficulté du problème. Les travaux sont inspirés de la contribution de Allard *et al.* [6] où une méthode exacte a été développée pour la dynamique de percolation par liens. Il est attendu que le recours à une méthode exacte implique de lourds calculs. Nous évaluerons donc sa valeur pratique pour différentes tailles de réseaux et tenterons d'optimiser ces calculs. L'objectif 1 se précise ainsi :

1. Obtenir une méthode exacte de prédiction des états stables pour tout réseau et pour une dynamique de cascade ;
2. Établir les limites pratiques du modèle, notamment pour les grands réseaux ;
3. Optimiser le calcul des états stables.

Chapitre 3

Le Chapitre 3 vise à traiter les objectifs 2 et 3. Nous réviserons et étofferons d'abord la démonstration mathématique pour parvenir à la réduction dimensionnelle. À la Sec. 1.4.2, l'observable était obtenue par une moyenne pondérée de l'activité par rapport au poids sortant. Nous construirons une nouvelle technique de réduction où la pondération, notée a_i dans l'équation suivante, sera laissée libre.

$$\text{Gao } et \text{ al. } \rightarrow R \propto \sum_{i=1}^N s_i^{\text{out}} x_i \rightarrow \text{Obtenir } \dot{R} \quad (1.97)$$

$$\text{Nouvelle approche } \rightarrow R \propto \sum_{i=1}^N a_i x_i \rightarrow \text{Obtenir } a_i \text{ et } \dot{R}. \quad (1.98)$$

Notre hypothèse est que la réduction de Gao *et al.* sera validée si la moyenne pondérée par le degré est réobtenue naturellement, c.-à-d. $a_i \propto s_i^{\text{out}}$. Autrement, un nouveau choix d'observables sera introduit. L'objectif 3 sera traité en soumettant le nouveau système à différentes dynamiques et réseaux, dont certains avec des structures modulaires et d'autres biparties. Nous

quantifierons notamment l'erreur sur les prédictions pour différents contextes. Les objectifs 2 et 3 se détaillent ainsi :

1. Valider l'utilisation du poids sortant comme pondération dans la réduction dimensionnelle de Gao *et al* ;
2. Développer un nouveau système réduit en levant la contrainte sur la pondération de l'observable ;
3. Valider le nouveau système réduit pour des dynamiques de populations, de propagation et de neurones, ainsi que pour des réseaux modulaires, aléatoires, biparties et invariant d'échelle.

Chapitre 4

Enfin, l'objectif 4 est abordé au Chapitre 4. Pour valider l'utilisation de l'apprentissage automatique, nous introduirons une nouvelle tâche, soit celle de prédire une perturbation structurelle sur un réseau dont le modèle dynamique est inconnu. L'idée derrière cette tâche est de produire un contexte où l'apprentissage automatique s'impose naturellement par le manque de connaissances *a priori* du système. De plus, cette nouvelle tâche s'inscrit dans le contexte où nous cherchons à documenter les perturbations appliquées sur des réseaux. Pour faire l'inférence, nous admettrons des séries temporelles de l'activité sur le réseau. Cette approche sera comparée à des méthodes de reconstruction fonctionnelles, qui seront introduites en détail dans le Chapitre, et des méthodes où le modèle dynamique est connu. Étant donné que cette contribution est fondée sur des jeux de données, nous ferons la validation de l'approche sur des réseaux réels et des dynamiques variées telles que de populations, de diffusion et de neurones. Le tout se décline sous deux sous-objectifs :

1. Valider l'utilisation de l'apprentissage automatique sur graphes pour détecter une perturbation structurelle et pour une dynamique quelconque ;
2. Documenter l'approche pour des dynamiques de populations, de propagation et de neurones, et sa robustesse au bruit.

Chapitre 2

Solution exacte de la dynamique de cascade sur réseaux complexes

Exact analytical solution of irreversible binary dynamics on networks

Edward Laurence¹, Jean-Gabriel Young¹, Sergey Melnik² et Louis J. Dubé¹,

¹ Département de Physique, de Génie Physique, et d'Optique
Université Laval, Québec, Québec, Canada G1V 0A6

² MACSI, Department of Mathematics & Statistics
University of Limerick, Ireland

Référence : Phys. Rev. E, 97 (2018), p. 032302.

© 2018 The American Physical Society (§2.3–§2.8)

...

2.1 Avant-propos

Pour poursuivre l'objectif de solutionner les états stables d'un réseau dynamique le plus exactement possible, nous devons sélectionner une dynamique particulière. Le choix s'est arrêté sur la dynamique en cascade formulée au moyen de fonctions de transition générales. Cette dynamique peut être solutionnée avec des techniques de passage de messages et d'autres basées sur les fonctions génératrices. Toutefois, il arrive que ces méthodes échouent pour certaines structures, d'où la motivation de ce travail.

...

2.2 Résumé

Au cours du chapitre, nous verrons comment les probabilités de l'état final peuvent être formulées et évaluées exactement. Nous introduirons un ensemble d'équations récursives pour calculer la probabilité que le processus s'achève dans un certain état, sachant l'état initial et les probabilités de transition de chaque noeud. Le temps pour solutionner ces équations croît de manière factorielle avec le nombre de noeuds en utilisant une approche naïve. Par conséquent, nous développerons un algorithme basé sur l'approche de parcours en largeur et optimisé pour solutionner les équations en temps exponentiel.

...

2.3 Abstract

In binary cascade dynamics, the nodes of a graph are in one of two possible states (inactive, active), and nodes in the inactive state make an irreversible transition to the active state, as soon as their precursors satisfy a predetermined condition. We introduce a set of recursive equations to compute the probability of reaching any final state, given an initial state, and a specification of the transition probability function of each node. Because the naive recursive approach for solving these equations takes factorial time in the number of nodes, we also introduce an accelerated algorithm, built around a breath-first search procedure. This algorithm solves the equations as efficiently as possible, in exponential time.

2.4 Introduction

Irreversible binary-state dynamics model rapid information transmission in complex systems. In these dynamics, nodes are in one of two states (inactive, active), and nodes in the inactive state make an irreversible transition to the active state, as soon as their precursors satisfy a predetermined condition. In a popular formulation of the problem [54], transition conditions are expressed by a set of response functions $\{F_i(m)\}_{i=1,\dots,N}$ that give the probabilities that nodes will make a transition from an inactive to an active state, based on the number m of their active precursors. This general formulation, known as *cascade dynamics*, encompasses multiple important dynamics as special cases [54]. Noteworthy examples are site and bond percolation [28, 32], the Watts model of threshold dynamics [166], and Susceptible-Infected (SI) models [126]. As a result, cascade dynamics have relevant applications in fields as diverse as epidemiology [117, 126], economics [63], and neuroscience [84, 86, 178].

Cascade dynamics present a difficult mathematical challenge : Predicting its outcome on arbitrary network topologies is notoriously hard. Only in some special cases do we know of analytical solutions that are both simple and elegant. Perhaps the most famous special case is that of ensembles of tree-like networks [119, 120, 128], for which probability generating functions (PGF) [32, 54, 55, 115, 169], and message passing (MP) [88, 153] methods yield exact analytical predictions of the important observables of the dynamics (e.g., size of the giant component, critical propagation threshold, etc.). In fact, their predictions are so accurate that these methods are routinely applied to real networks, even though the underlying hypotheses are no longer valid ; in many cases, this leads to surprisingly good approximations of the true outcome [102, 114]. However in many important cases, experiment and theory are at odds. It is now understood that these discrepancies can—in part—be traced back to existence of local correlations in real systems, whose effects are overlooked by tree-based theory [114].

There are a few promising methods—developed within the general framework of cascade dynamics—that address this issue by including local correlations using, e.g., a fixed prevalence of cliques on a tree-like backbone [61, 62] or with fixed degree-degree correlations [115]. However, the most versatile proposition to date comes from percolation theory [5, 7, 87] ; it consists in mixing PGF methods with exact solutions on arbitrary motifs. The general idea behind this method is simple : One first decomposes a network in small local subgraphs (motifs), then solves percolation on these graphs, and finally combines the local solutions to obtain a global solution [7].

There appears to be no conceptual obstacles to adapting this method to general cascade dynamics—the clique-based methods of Refs. [61, 62] are in fact special cases of the general

approach. There is, however, a major practical bottleneck : Exactly solving cascade dynamics on small graphs is, at best, tedious [61]. This bottleneck becomes a barrier when motifs are too diverse, or too large. In Refs. [6, 7], a costly but systematic algorithm is introduced to handle enumeration and averaging of the traversal probabilities in the *percolation problem*. The goal of the present paper is to delineate the equivalent enumeration algorithm in the much more general context of *cascade dynamics*.

The paper is organized as follows. We first define cascade dynamics, and obtain recursive equations for the probability of every outcome, on arbitrary network topologies and general cascades (Sec. 2.5). We then discuss the practical aspect of the formalism in Sec. 2.6, i.e., how to compute the solution of the recursive equations as efficiently as possible.¹ In Sec. 2.7, we illustrate the power of the formalism in a case-study of complicated mixed dynamics occurring on small directed networks. We gather our conclusions and perspectives in Sec. 2.8. Three appendices follow. In the first, we prove that our method is equivalent to the analogous method derived in the context of percolation theory (Appendix 2.10.1). In the second, we work out an explicit example (Appendix 2.10.2). In the third and last Appendix, we give a detailed calculations of the worst-case computational complexity of our algorithms (Appendix 2.10.3).

...

2.5 Cascade dynamics on arbitrary graphs

2.5.1 Definition of the dynamics

We consider an irreversible binary-state dynamics occurring on an arbitrary network of N nodes. This dynamic process is well defined on graphs that contain directed edges [54], and we therefore encode the structure of the network in a binary-valued, potentially asymmetric $N \times N$ adjacency matrix \mathbf{A} . We adopt the convention that the element a_{ij} of \mathbf{A} indicates the existence of an edge going from node j to node i . In the directed case, the neighborhood of node i , i.e. the set of nodes that can be reached from i , is hence given by the non-zero entries on the i th column of \mathbf{A} , and the precursors of node i are the set of nodes having a directed edge to node i . The undirected case is obtained by placing a directed edge in both directions.

Following the prescription of Ref. [54], all nodes are initially placed in the inactive state, except for a (potentially disconnected) set of seed nodes, initially active. At each subsequent discrete time steps t , inactive nodes make a transition to the active state if their precursors satisfies

1. We also provide a reference implementation of our solver at github.com/laurencee9/exact_binary_dynamics.

an activation condition. The process ends when no further transitions are possible (i.e., when the state of each node at time t is identical to the state at time $t + 1$).

The transition conditions are encoded in a set of N so-called “influence response functions” $\{F_i(m)\}_{i=1,\dots,N}$, where m is the number of active precursors of node i [54]. We define these functions as the cumulative distribution function (CDF) of the probability $P_i(m)$ that node i has a hidden activation threshold precisely equal to m precursors :

$$P_i(m) = F_i(m) - F_i(m - 1) , \quad (2.1)$$

or in other words, as the CDF of the probability that node i will make a transition to the active state, as soon as m of its precursors reach the active state. For convenience, we also define the complementary probability

$$G_i(m) := 1 - F_i(m) \quad (2.2)$$

that node i has a hidden threshold greater than m active precursors—or equivalently, that it does not make a transition when it has m active precursors. Under this probabilistic description, the process ends when every inactive node at time t fails to make a transition to the active state.

A complete specification of the dynamics thus consists of : the structure of the network, as specified by a $N \times N$ adjacency matrix \mathbf{A} ; a set of seed nodes; and a response function $F_i(m)$ for each node. Because our formalism allows it, we will work with response functions specified on a node-to-node basis. Note, however, that it is not uncommon to group nodes in coarser compartments. In fact, response functions $F(m_i, \theta_i)$ that only depend on m_i and some local parameter θ_i —such as the degree k_i of node i —are widely used [55] : Many well-known dynamics can be recovered as a special case by choosing specific classes of response functions [54].

2.5.2 Special classes of response functions

Let us consider first bond percolation, where a node is part of an active component if at least one of its incoming edges percolates. If one denotes by m_i the number of in-edges, that can percolate, of node i , and by p the percolation probability, then node i remains inactive with probability $G(m_i, p) = (1 - p)^{m_i}$. The response function can therefore be written as the complementary probability

$$F(m_i) = 1 - (1 - p)^{m_i} . \quad (2.3a)$$

For site percolation, where a node percolates with probability p if it has at least a single active precursor, the response function is simply

$$F(m_i) = \begin{cases} 0, & m_i = 0 \\ p, & m_i > 0. \end{cases} \quad (2.3b)$$

In the Watts cascade model [166], a node of degree k_i makes a transition to the active state when a critical fraction r of its k_i precursors reaches the active state, and the thresholds r are drawn from a distribution, whose cumulative distribution is given by $C(r)$. The response function for the resulting dynamics is therefore

$$F(m_i, k_i) = C(m_i/k_i). \quad (2.3c)$$

These are but a few examples, derived in Ref. [54], and reproduced here to showcase the generality of the cascade formalism. Furthermore, there exists many more equivalencies, inherited from its reduction to the Watts threshold model [see Eq. (2.3c)]; the latter can be mapped, with the appropriate choice of thresholds, to site percolation, k -core percolation, diffusion percolation [2], and the generalized epidemic process [118]. Our method therefore provides systematic and exact solutions to a large class of dynamics.

2.5.3 Exact recursive solution

The outcome of a cascade dynamics on a network is a configuration of active and inactive nodes. We encode these configurations in binary vectors $\mathbf{l} = [l_1, \dots, l_N]^\top$ of length N , where $l_i = 1$ if node i is active in the configuration, and 0 otherwise. Thus, the vector $\mathbf{n} := [1, \dots, 1]^\top$ refers to the completely active configuration, and the number of active nodes in a configuration \mathbf{l} is given by the square of its Euclidean norm, e.g., $|\mathbf{n}|^2 = N$.

Cascade dynamics are, by definition, probabilistic processes. So fully solving the cascade amounts to computing the probability of every outcome. Given a fixed initial configuration \mathbf{l}_0 and a fixed structure \mathbf{A} , we define $Q(\mathbf{l}; \mathbf{A}, \mathbf{l}_0)$ as the probability of observing configuration \mathbf{l} when the cascade stops. A solution is therefore a distribution $\mathcal{Q} = \{Q(\mathbf{l}; \mathbf{A}, \mathbf{l}_0)\}_{\mathbf{l}}$ over all $\mathbf{l} \in \{0, 1\}^N$, for a choice of response functions (not explicitly denoted, for the sake of clarity). Even though there are exponentially many possible outcomes ($|\mathcal{Q}| = 2^N$), the calculation of \mathcal{Q} is greatly simplified by the structure underlying the distribution.

Observe that when a cascade process ends, the nodes can be partitioned in two subsets defined by their activity status. In particular, nodes in the inactive subset must have, by definition, a hidden threshold superior to the number of their active precursors. The probability of observing any subset of inactive nodes can thus be written in terms of the complementary probabilities $\{G_i(m)\}$. Specifically, since each activation event is independent, we can write the stopping probability of the cascade in configuration \mathbf{l} as

$$Q(\mathbf{l}; \mathbf{A}_{\mathbf{u}}) = Q(\mathbf{l}; \mathbf{A}_{\mathbf{l}}) \prod_{i=1}^N [G_i(m_i(\mathbf{l}))]^{u_i(1-l_i)}, \quad \mathbf{u} \geq \mathbf{l}, \quad (2.4a)$$

where $m_i(\mathbf{l}) = \sum_j a_{ij} l_j$ is the number of active precursors of node i , and where $\mathbf{u} \geq \mathbf{l}$ is an elementwise inequality $u_i \geq l_i \forall i$. The matrix $\mathbf{A}_{\mathbf{l}}$ denotes the reduced adjacency matrix

$$\mathbf{A}_{\mathbf{l}} = \mathbf{L} \mathbf{A} \mathbf{L},$$

where $\mathbf{L} = \text{diag}(\mathbf{l})$ is a $N \times N$ diagonal matrix, whose entries are given by the vector \mathbf{l} . The reduced adjacency matrix $\mathbf{A}_{\mathbf{l}}$ is almost identical to \mathbf{A} , except that the values on the j^{th} row and column are set to zero if $l_j = 0$. It encodes the structure of the subgraph induced by the set of active nodes in configuration \mathbf{l} . Thus, if $\mathbf{l} = \mathbf{n}$, then $\mathbf{A}_{\mathbf{n}} = \mathbf{A}$.

With this in mind, Eq. (2.4a) is interpreted as follows : It states that the probability that a cascade taking place on $\mathbf{A}_{\mathbf{u}}$ will stop at configuration \mathbf{l} is equal to the probability $Q(\mathbf{l}; \mathbf{A}_{\mathbf{l}})$ of reaching a completely active subgraph induced by \mathbf{l} , times the probability $\prod_{i=1}^N [G_i(m_i)]^{u_i(1-l_i)}$ that the cascade does not spread to the remaining $|\mathbf{u} - \mathbf{l}|^2$ inactive nodes.

This essentially solves the problem. The distribution of outcomes \mathcal{Q} can—in principle—be computed by solving the system of $2^{(N-|\mathbf{l}_0|^2)}$ equations defined in (2.4a), recursively, for configurations with increasing numbers of active nodes. However, it turns out that this system is not complete : Every completely active configuration in a reduced subgraph is associated to a non-informative Eq. (2.4a), because $u_i = l_i = 1 \forall i$ yields the tautology

$$Q(\mathbf{l}; \mathbf{A}_{\mathbf{l}}) = Q(\mathbf{l}; \mathbf{A}_{\mathbf{l}}) .$$

Hence, we are dealing with an underdetermined system of equations. Fortunately, the system can be completed via the normalization condition

$$Q(\mathbf{l}; \mathbf{A}_{\mathbf{l}}) = 1 - \sum_{\mathbf{u} : \mathbf{u} < \mathbf{l}} Q(\mathbf{u}; \mathbf{A}_{\mathbf{l}}) , \quad (2.4b)$$

where $\mathbf{u} : \mathbf{u} < \mathbf{l}$ means that the sum is taken over all \mathbf{u} respecting the elementwise inequality $u_i < l_i \forall i$. Eq.(2.4b) states that the probability of reaching the fully active configuration \mathbf{l} on the reduced adjacency matrix $\mathbf{A}_{\mathbf{l}}$ is given by the complementary probability of reaching *any other state* with fewer active nodes (on $\mathbf{A}_{\mathbf{l}}$). Of course, one also needs to give an initial condition of the form

$$Q(\mathbf{l}_0; \mathbf{A}_{\mathbf{l}_0}) = 1, \quad Q(\mathbf{l}; \mathbf{A}_{\mathbf{l}}) = 0 \quad \forall \mathbf{l} \leq \mathbf{l}_0 , \quad (2.4c)$$

to complete the system.

Before considering the practical aspect of this formalism, we note that an analogous derivation leads to almost identical equations in the case of bond percolation [6]. In fact, we show in Appendix 2.10.1, that upon substitution of the response function (2.3a) in the equations, one recovers the formalism derived in Ref. [6] for bond percolation.

...

2.6 Enumeration algorithms

2.6.1 General recursive solution method

In practice, and as stated above, Eqs. (2.4) must be solved recursively. One must feed the results of Eq. (2.4a) into Eq. (2.4b), and back into Eq. (2.4a), using the initial condition Eq. (2.4c) as a starting point. To make things more concrete, let us follow through with the first few steps of the computation.

First, we compute the probabilities $Q(\mathbf{l}_0; \mathbf{A})$ using Eq. (2.4a) and the initial condition Eq. (2.4c). Then, we proceed to configurations with one more active node, i.e., configurations \mathbf{l}_i such that $|\mathbf{l}_i|^2 = |\mathbf{l}_0|^2 + 1$. These probabilities, of the form $Q(\mathbf{l}_i; \mathbf{A})$, can be computed from Eq. (2.4a). But in so doing, we will need to resort to the normalization condition [Eq. (2.4b)]. Once the stopping probabilities of all configurations with one extra active node are computed, it is only a matter of repeating the process for all configurations with one more active node (i.e., \mathbf{l}_i such that $|\mathbf{l}_i|^2 = |\mathbf{l}_0|^2 + 2$), progressing toward larger and larger configurations, all the way up to the complete configuration of $|\mathbf{n}|^2 = N$ nodes. This scheme obviously constructs the complete distribution \mathcal{Q} (as well as a complete set of distributions \mathcal{Q}' on every possible subgraph of \mathbf{A}). We work out an explicit example for a small tadpole graph in Appendix 2.10.2.

2.6.2 Saving time : culling impossible configurations

The method just discussed is less than optimal, precisely because it goes through every single configuration of active nodes and every single induced subgraphs. This is due to the fact that many response functions and many graphs are associated with large sets of *impossible* configurations, i.e., sets of configurations \mathbf{l} with null probabilities $Q(\mathbf{l}; \mathbf{A}_u) = 0$. It is easy to see that a configuration \mathbf{l} is *necessarily* impossible if it contains at least one node i not connected to a seed (or a spontaneously activated node) via a path in $\mathbf{A}_\mathbf{l}$: There is then simply no possible path for the cascade to have propagated to node i . Such configurations abound.

If we can find and avoid these impossible configurations more efficiently than through brute-force enumeration of $\mathbf{l} \in \{0, 1\}^N$, then we will have accelerated the calculation of \mathcal{Q} by completely ignoring large portions of its support. This can be done in most practical cases by using a breath-first search (BFS) *over the configurations*, that accounts for the presence of spontaneously activated nodes [i.e. $F_i(0) > 0$] and disconnected seeds, see Fig. 2.1. In this modified BFS, nodes can be in one of two states : either discovered or undiscovered. We seed the search with the initial condition \mathbf{l}_0 , i.e., by labeling the $|\mathbf{l}_0|^2$ seed nodes as discovered. We then enumerate all *neighboring* configurations containing $|\mathbf{l}_0|^2 + 1$ discovered nodes, i.e.,

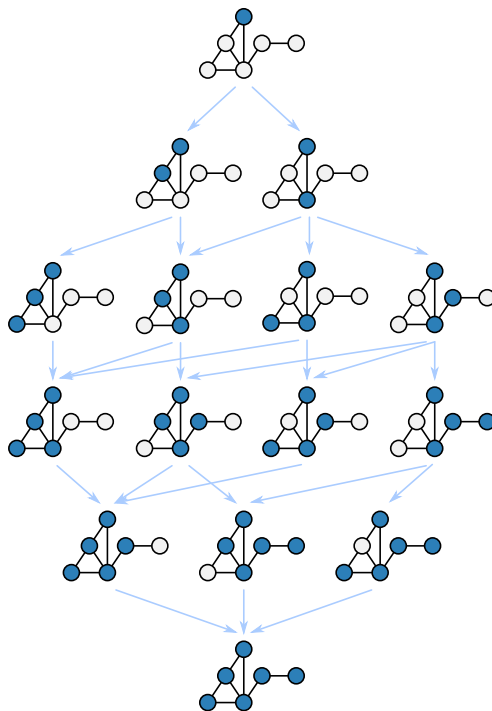


FIGURE 2.1 – Example of a breath-first exploration of all configurations. The configurations are explored by rows, i.e., by the number of discovered (active) nodes. Nodes are labeled as discovered (blue) or undiscovered (white). Blue arrows indicate possible transitions.

all configurations with $|\mathbf{l}_0|^2 + 1$ discovered nodes that are either neighbor of an initial seed node or a spontaneous active node. We then repeat the process, expanding outwards. Importantly, every new configuration is constructed by taking one of the previous configurations and converting one undiscovered node, adjacent to either an already discovered or a spontaneously activating nodes.

By carrying this process recursively, we mimic the cascade and only generate possible configurations. When there are no undiscovered nodes adjacent to a discovered node and no undiscovered spontaneous nodes, we therefore have enumerated all possible final configurations, and the algorithm terminates.

The advantage of using this search is that we can intertwine the enumeration procedure and the recursion of Eqs. (2.4). As soon as the number of active nodes increases by one in the BFS, say from m to $m + 1$, we know that we have encountered all possible active configurations with $m' \leq m$ nodes. We can thus set the probability of all the missing configurations of less than $m + 1$ nodes to 0. At best, we will have saved vast amount of unnecessary calculation. At worst, the complexity of the whole calculation will remain more or less the same—the BFS will only have ordered the enumeration of configurations.

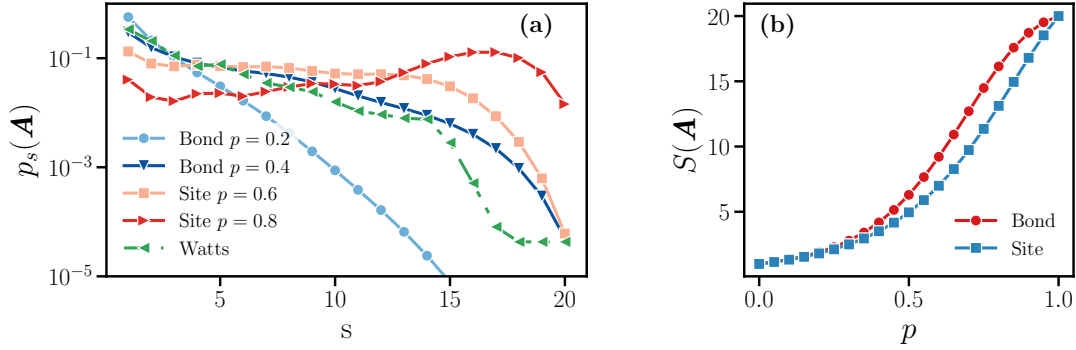


FIGURE 2.2 – (a) Size distribution of the active components under different dynamic processes : bond percolation, site percolation and the Watts threshold model. Symbols are the results of 10^7 Monte Carlo simulations, and curves show the exact probability given by our approach Eq. (2.6). (b) Exact mean size, Eq. (2.7), of the final active component for bond and site percolations as a function of the occupation probability p . The graph has $N = 20$ and 54 directed edges.

2.6.3 Saving more time : dropping intermediary distributions

For most graphs, carrying out the full recursive solution of Eqs. (2.4) is impossible, because too much information must be held either in memory, or because the calculation is simply too long—even with the help of the BFS strategy.

Besides the obvious exponential size of the solution, this slowdown can be traced to another important bottleneck : The evaluation of Eq. (2.4b) requires a complete knowledge of the distribution for all \mathbf{u} with $\mathbf{u} < \mathbf{l}$, i.e., all $\{\mathcal{Q}'\}$ over induced subgraphs. In fact, one can show that $|\{\mathcal{Q}'\}| \sim N!$ (see Appendix. 2.10.3). If we could do without the information contained in $\{\mathcal{Q}'\}$, then the limiting factor would become the (exponential) size of the solution \mathcal{Q} rather the (factorial) complexity of the algorithm—an appreciable speedup. This would of course come at the price of dropping the distributions $\{\mathcal{Q}'\}$ on smaller subgraphs. One should therefore stick to the BFS enhanced recursion method of Sec. 2.6.2 if this information is needed.

However, provided that $\{\mathcal{Q}'\}$ can be dropped, it is possible to compute \mathcal{Q} directly. The strategy consists in combining the recursive equations, and using BFS to enumerate the elements of the distribution \mathcal{Q} in an orderly manner. We build upon the observation that two configurations that share a common core of active nodes have related probabilities. Specifically : the probability $Q(\mathbf{u}; \mathbf{A}_\mathbf{l})$ can be computed from the probability $Q(\mathbf{u}; \mathbf{A})$ by considering the contribution of the inactive nodes absent from the subgraph $\mathbf{A}_\mathbf{l}$:

$$Q(\mathbf{u}; \mathbf{A}_\mathbf{l}) = Q(\mathbf{u}; \mathbf{A}) \left[\prod_{i=1}^N G_i(m_i(\mathbf{u}))^{(1-l_i)(1-u_i)} \right]^{-1}. \quad (2.5a)$$

[we merely “factor out” complementary probabilities hidden in $Q(\mathbf{u}; \mathbf{A})$.] Inserting Eq. (2.5a)

into Eq. (2.4), we can then write the elements of \mathcal{Q} directly as

$$Q(\mathbf{l}; \mathbf{A}) = [1 - Z(\mathbf{l}, \mathbf{A})] \prod_{i=1}^N [G_i(m_i(\mathbf{l}))]^{(1-l_i)}, \quad (2.5b)$$

where

$$Z(\mathbf{l}, \mathbf{A}) := \sum_{\mathbf{u}: \mathbf{u} < \mathbf{l}} \frac{Q(\mathbf{u}; \mathbf{A})}{\prod_{i=1}^N G_i(m_i(\mathbf{u}))^{(1-l_i)(1-u_i)}}. \quad (2.5c)$$

This transformation is useful because Eqs. (2.5b)–(2.5c) explicitly include the normalization condition and can be solved without recursion. The procedure is simple. We start by listing and ordering all possible configurations of the complete graph using BFS. It is important to keep the discovery order. Then, we solve Eq. (2.5b) from the smallest configuration to the largest configuration (i.e. the discovery order). At each evaluation, every $Q(\mathbf{u}; \mathbf{A})$ appearing in Eq. (2.5c) is already calculated and memorized—as imposed by the discovery order. Under this systematic method, the previously calculated information is reinjected in the set of equations. All intermediate distributions are thus simply dropped, which considerably reduces the computational complexity of the algorithm (see Appendix 2.10.3).

...

2.7 Results and applications

Our main motivation in deriving recursive equations is to elaborate methods that combines exactly solved motifs on a tree-like backbone, in the spirit of Refs. [7, 87]. However, for small graphs, the method can also be useful in itself. For instance, one can marginalize \mathcal{Q} to obtain the individual activation probabilities of each node, and these probabilities can then be used for diagnosis purposes, e.g., the importance of a node with regard to spreading [76]. With this in mind, we give two practical examples in the next section, to illustrate the power of exact solutions on small graphs.

2.7.1 Calibration : special cases on a directed graph

To calibrate the method and verify its validity, we first use the algorithm of Sec. 2.6.3 to obtain cascade results on a known network : The directed network of 20 nodes and 54 directed edges appearing in Ref. [6]. We compute the solutions for three special cases of cascade dynamics, previously introduced in Sec. 2.5.2 : bond percolation [see Eq. (2.3a)], site percolation [see Eq. (2.3b)], and the Watts cascade model [see Eq. (2.3c)]. Furthermore, we compare the predictions of the formalism with Monte Carlo simulations.

For site (bond) percolation, the Monte Carlo simulations are done by randomly selecting a seed node and then occupying the neighbors (adjacent edges) with a probability p . This step is then repeated for the new neighborhood (adjacent edges), until the process stops or until the network is exhausted. For the Watts threshold model, the simulation follows the standard description of a cascade process [166]. We first assign a threshold C_i between 0 and 1 to each node i , drawn from a uniform distribution. A randomly chosen node—the seed—is then activated. We then enter the propagation phase : If the threshold C_i of node i is lower or equal to the fraction of its precursors that are active m_i/k_i , then the node makes a transition to the active state. This process is repeated until no transitions are possible. For all these Monte Carlo simulations, the estimator $\hat{Q}(\mathbf{l}; \mathbf{A}, \mathbf{l}_0)$ of the outcome probability is computed as the frequency of the final configuration \mathbf{l} ; standard results tell us that the variance of \hat{Q} will decrease as the square root of the number of trials.

Note that because we seed Monte Carlo simulations at random, we compare against results computed using exact probabilities \mathcal{Q} that are marginalized over every initial configuration $|\mathbf{l}_0|^2 = 1$, i.e.,

$$Q(\mathbf{l}; \mathbf{A}) = \sum_{\mathbf{l}_0: |\mathbf{l}_0|^2=1} Q(\mathbf{l}; \mathbf{A}, \mathbf{l}_0) .$$

This is consistent with the typical way in which Monte Carlo simulations are carried out.

The graph contains 20 nodes, meaning that there are roughly 10^6 different configurations—we cannot possibly visualize the complete distribution \mathcal{Q} . Thus, we will focus on summary statistics, the size distribution [Fig. 2.2 (a)] and the mean size [Fig. 2.2 (b)]. Using the distribution \mathcal{Q} , we calculate the probability p_s that a cascade will reach s nodes as

$$p_s(\mathbf{A}) = \sum_{\mathbf{l} \in \{0,1\}^N} Q(\mathbf{l}; \mathbf{A}) \mathbb{I}(|\mathbf{l}|^2 = s) , \quad (2.6)$$

where $\mathbb{I}(\cdot)$ is an indicator function, equal to 1 when its argument is true, and 0 otherwise. We also compute the mean size of the active component

$$S(\mathbf{A}) = \sum_{s=1}^N s p_s(\mathbf{A}) = \sum_{\{\mathbf{l}\}} Q(\mathbf{l}; \mathbf{A}) |\mathbf{l}|^2 . \quad (2.7)$$

Figure 2.2 (a) shows the size distribution for different dynamics and occupation probabilities. The perfect fit of the exact results and the Monte Carlo simulations confirms the validity the equations and the algorithm. In the case of the selected graph (see Ref. [6]), roughly half of the $2^{20} = 1\,048\,576$ possible configurations go into the calculations of the size distributions. Notably, computing these distributions is trivial once the recursive procedure of Sec. 2.6 has been carried out : The equations yield a large polynomial in $\{G_i(x)\}$, which can be evaluated easily and many times, by specifying the response functions and parameters. Thus, generating a family of distributions is not significantly more costly than generating a single one. Another

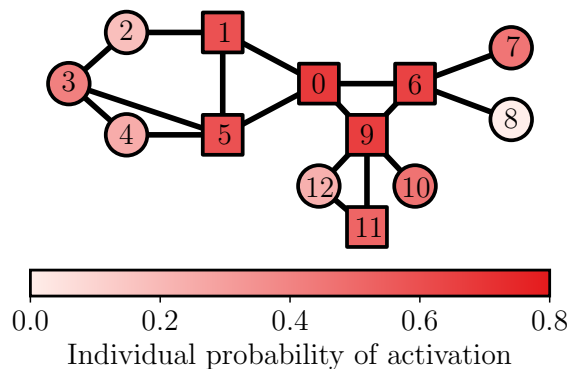


FIGURE 2.3 – Individual response function and activation probabilities. The colors of the nodes are used to show their activation probability, summed over all possible outcomes. Square nodes can make a spontaneous transition [with response function $F_s(p, q)$, see Eq. (2.9)], while round nodes need active precursors to make a transition [they have the response function $F_w(p, \tau)$, see Eq. (2.8)]. The parameters of the spontaneously activating nodes are : $p = 0.4$ and $q = 0.3$ for nodes 1, 5 and 11 ; and $p = 0.6$ and $q = 0.1$ for nodes 0, 6 and 9. The parameters of the threshold nodes are : $p = 0.6$ and $\tau = 2$ for node 2, 4, 8 and 12 ; and $p = 0.7$ and $\tau = 1$, for nodes 3, 7 and 10.

important aspect, also raised in Ref. [6] is the irregularities of the size distributions (in the case of bond percolation), which indicates that the solution is non-trivial and depends on the intricacies of the structure of the graph. This tells us that any close form solution will necessarily be just as complex. Furthermore, we see in Fig. 2.2 (a) that bond percolation behaves, in fact, quite simply in comparison with the other dynamics (site percolation, Watts model); exact solutions on small graphs therefore appear even more useful in the case of general cascade dynamics than in the special case of bond percolation.

Figure 2.2 (b) shows the theoretical mean size component for two dynamics. Again, we benefit from the fact that the corresponding polynomial can be evaluated at will once the recursive procedure has been completed : Our approach allows us to zoom-in on any point of the curve, and to investigate specific regions of the parameter space at little extra computational costs.

2.7.2 Mixed dynamics

In the previous case study, the network had a global response function, independent of the node identity. In real situations, it is fair to assume that nodes of different types do not respond identically to stimuli. Thus, it is natural and perfectly general to consider a mixed dynamics that stems from diverse response functions. As highlighted in Sec. 2.5, our formalism can handle this generalization without any additional complexity.

We consider a mixed dynamics where nodes are associated with one of two types of parame-

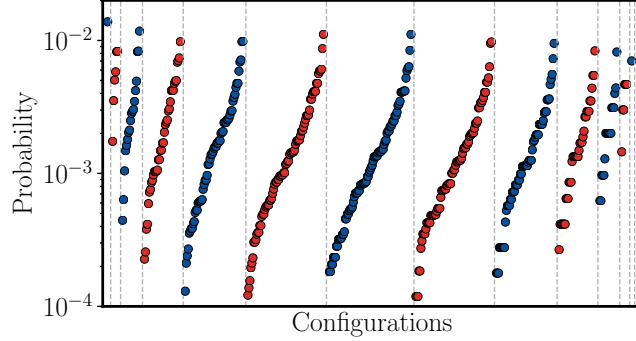


FIGURE 2.4 – Complete distribution \mathcal{Q} for the every possible configurations of the network, using the dynamics described in Fig. 2.3. Configurations are grouped together according to the number of active nodes (from 0 to 12), and then ordered with ascending probabilities. Since node 8 is always inactive, the completely active configuration (13 active nodes) has a zero probability and, consequently, is not shown. Gray lines are placed where the number of active nodes increases.

terizable response function,

$$F_w(p, \tau) = \begin{cases} 0 & m \leq \tau - 1, \\ p & m \geq \tau, \end{cases} \quad (2.8)$$

$$F_s(p, q) = \begin{cases} p & m = 0, \\ 1 - (1 - p)q^m & m > 0. \end{cases} \quad (2.9)$$

The function $F_w(p, \tau)$ describes a threshold dynamics, where τ is the activation threshold and p is the activation probability once the threshold is exceeded. The function $F_s(p, q)$ describes a node that can either activate spontaneously, or through contagion; p is the probability that the node will spontaneously activate (at time $t = 0$), whereas q controls its sensitivity to active precursors. We assign these response functions to the $N = 13$ nodes of a handmade small network, shown in Fig. 2.3. The network is constructed to showcase heterogeneous patterns of activation.

Because we have introduced random seeds, we will no longer sum the probabilities over all seeds; instead, we always begin the process in the inactive configuration $\mathbf{l}_0 = [0, 0, \dots, 0]_{\mathbb{T}}$ and let spontaneous activations determine the outcome. In Fig. 2.4, the probability $Q(\mathbf{l}; \mathbf{A})$ is displayed for the complete set of possible outcomes \mathbf{l} . Since the network is relatively sparse and small ($N = 13$), these results can be computed extremely quickly (less than a minute on a modern personal computer).

In Fig. 2.3, we have colored nodes with their probability of being active by summing over all possible outcomes. Combining the information given by both of these graphs, we see that the fully active configuration is impossible, because nodes 8 has a degree smaller than its threshold. Similar and much more in depth analyses could obviously be carried out on real small networks, e.g., on a power grid.

2.8 Discussion

In this paper, we have derived a new set of recursive equations that solve cascade dynamics on arbitrary networks, and we have introduced two practical strategies to manage the complexity of solving such a large set of equations. These developments are inspired by analogous methods, introduced within the framework of percolation theory [6].

Due to the generality of the cascade dynamics formulation [54, 118], our method leads to exact solutions for a wide range of dynamics, including well-known examples such as site and bond percolation and the Watts threshold model. But its power goes beyond such simple cases ; it also generates exact solution for many exotic dynamics and connectivity patterns, e.g., directed, self-referencing edges, weighted graphs, disconnected active components, spontaneous activity and disconnected initial seeds activation. Furthermore, the exact solution is valid in the much general context of individual activation functions (i.e., each node can have a different activation function). To the best of our knowledge, our framework is the only one able to do so.

We cannot understate the fact that our formalism solves a problem whose solution grows exponentially with N . As such, the method is, by necessity, intractable². On the other hand, the usual methods for solving numerically cascade dynamics, the tree-based theory and the message-passing framework [57, 88, 102], are surprisingly accurate for sparse networks and are commonly used on real large networks [114]. Our contribution is therefore not designed for computing ensemble statistics in the large N limit, where Monte Carlo simulations or approximations schemes would be more appropriate, if the specifics of the configurations do not matter. Instead, our method is designed to yield exact solutions for small graphs, where the exponential dependency is still manageable. This is useful in at least three scenarios : (i) on small graphs where accurate *configuration* probabilities are needed (c.f. Ref. [76]), (ii) in formalisms where motif distributions are specified and traversal probabilities must preferably be computed in closed form (c.f. Refs. [5, 6, 7]), and (iii) on graphs with mixed dynamics. In all cases, our formalism involves calculations that are no more complex than Monte Carlo simulations, with the added advantage of producing exact probabilities.

In summary, despite the unwieldiness of the calculations involved, our results open the way to new theoretical predictions, since it solves cascades on small motifs, both exactly and systematically.

2. In the worst case scenario of a complete graph, our implementation is able to handle roughly 20 nodes on a modern single-CPU computer.

...

2.9 Acknowledgements

We are thankful to Antoine Allard and Patrick Desrosiers for useful comments and suggestions. We are grateful to Andrey Lokhov for pointing out Ref. [102]. This work was funded by the Fonds de recherche du Québec-Nature et technologies (J.-G.Y and E.L.), the Conseil de recherches en sciences naturelles et en génie du Canada (L.J.D.), the Irish Research Council (New Foundations grant to S.M.), Science Foundation Ireland (Grant No. 11/PI/1026, S.M.) and Sentinel North (E.L., J.-G.Y).

E.L. and J.-G.Y. contributed equally to this work.

...

2.10 Supplementary Material

2.10.1 Equivalence with bond percolation

An exact solution to bond percolation is given in Ref. [6], in the form of a set of recursive equations. We show that our equations generalize these equations. As stated in Sec. 2.5.2, bond percolation can be mapped to a cascade dynamic process by setting $F(m_i) = 1 - (1 - p)^{m_i}$ and $G(m_i) = (1 - p)^{m_i}$ for all nodes. We insert these relations in Eqs. (2.4a) and obtain

$$Q(\mathbf{l}; \mathbf{A}_\mathbf{u}) = Q(\mathbf{l}; \mathbf{A}_\mathbf{l}) \prod_{i=1}^N [(1 - p)^{m_i}]^{u_i - l_i} \quad \mathbf{u} \geq \mathbf{l} \quad (2.10)$$

Next, writing out the number of active precursors of i as $m_i(\mathbf{l}) = \sum_k a_{ik} l_k$, we get

$$\begin{aligned} \prod_{i=1}^N (1 - p)^{\sum_k a_{ik} l_k} &= (1 - p)^{\sum_{k,i} a_{ik} l_k} , \\ &= (1 - p)^{\mathbf{l}^T \mathbf{A} \bar{\mathbf{l}}} , \end{aligned}$$

where $\bar{l}_i := u_i - l_i$. Substituting back into Eq. (2.10), we find, for $\mathbf{u} = \mathbf{n}$,

$$Q(\mathbf{l}; \mathbf{A}) = Q(\mathbf{l}; \mathbf{A}_\mathbf{l}) (1 - p)^{\mathbf{l}^T \mathbf{A} \bar{\mathbf{l}}} , \quad (2.11)$$

i.e., Eq. (3) of Ref. [6]. The normalization condition is simply generalized since it does not depend on the specifics of the cascade dynamics. This completes the proof of equivalence.

2.10.2 Explicit example on a tadpole graph

In this Appendix, we work out several of the first steps of the full recursive procedure, on a small tadpole graph of 4 nodes, i.e., the graph whose adjacency matrix is given by

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} . \quad (2.12)$$

We choose an initial configuration where every node is inactive, $\mathbf{l}_0 = [0, 0, 0, 0]^T$. To keep the response function as general as possible, we will not further specify the dynamics. Therefore, spontaneous activation [$F_i(m) > 0$] is assumed to occur, otherwise the only possible outcome is \mathbf{l}_0 . To simplify the demonstration, we suppose identical response functions for each node, i.e., $F_i(m) = F(m_i) \forall i$.

There exists 2^4 different configurations of the cascade dynamics. We start with the initial configuration and Eq. (2.4a), to yield

$$Q(\mathbf{l}_0; \mathbf{A}) = Q(\mathbf{l}_0; \mathbf{A}_{\mathbf{l}_0})G(0)^4 = G(0)^4,$$

since $Q(\mathbf{l}_0; \mathbf{A}_{\mathbf{l}_0}) = 1$ by definition [see the initial condition in Eq. (2.4c)]. Moving to configurations with one more active node, we begin with $\mathbf{l}_1 = [1, 0, 0, 0]^\top$ and follow the same procedure

$$\begin{aligned} Q(\mathbf{l}_1; \mathbf{A}) &= Q(\mathbf{l}_1; \mathbf{A}_{\mathbf{l}_1})G(1)^2G(0) \\ &= [1 - Q(\mathbf{l}_0; \mathbf{A}_{\mathbf{l}_1})]G(1)^2G(0) \\ &= [1 - Q(\mathbf{l}_0; \mathbf{A}_{\mathbf{l}_0})G(0)]G(1)^2G(0) \\ &= F(0)G(1)^2G(0), \end{aligned} \tag{2.13}$$

where we have used the definition $F(m) = 1 - G(m)$ at the last step. Notice how the normalization (2.4c) intervenes and how $Q(\mathbf{l}_0; \mathbf{A}_{\mathbf{l}_1})$ and $Q(\mathbf{l}_1; \mathbf{A}_{\mathbf{l}_1})$ are computed as a by-product of this step, producing the complete distribution of outcomes \mathcal{Q}' for a cascade taking place on $\mathbf{A}_{\mathbf{l}_1}$. Further steps will generate similar distributions.

Completing our calculations for the other 1 node configurations, we observe that the symmetry of the graph leads to $Q(\mathbf{l}_2; \mathbf{A}) = Q(\mathbf{l}_1; \mathbf{A}) = F(0)G(1)^2G(0)$ with $\mathbf{l}_2 = [0, 1, 0, 0]^\top$, while

$$\begin{aligned} Q(\mathbf{l}_3 = [0, 0, 1, 0]^\top; \mathbf{A}) &= F(0)G(1)^3, \\ Q(\mathbf{l}_4 = [0, 0, 0, 1]^\top; \mathbf{A}) &= F(0)G(1)G(0)^2, \end{aligned}$$

following the procedure of Eq. (2.13).

For configurations with two active nodes such as $\mathbf{l}_5 = [1, 1, 0, 0]^\top$, two steps of recursions are required. First, we use Eq. (2.4a)

$$Q(\mathbf{l}_5; \mathbf{A}) = Q(\mathbf{l}_5; \mathbf{A}_{\mathbf{l}_5})G(2)G(0),$$

and apply Eq. (2.4b)

$$Q(\mathbf{l}_5; \mathbf{A}_{\mathbf{l}_5}) = [1 - Q(\mathbf{l}_0; \mathbf{A}_{\mathbf{l}_5}) - Q(\mathbf{l}_1; \mathbf{A}_{\mathbf{l}_5}) - Q(\mathbf{l}_2; \mathbf{A}_{\mathbf{l}_5})].$$

None of these terms are a priori known. We must use Eq. (2.4a) again for each of them

$$\begin{aligned} Q(\mathbf{l}_0; \mathbf{A}_{\mathbf{l}_5}) &= Q(\mathbf{l}_0; \mathbf{A}_{\mathbf{l}_0})G(0)^2 = G(0)^2, \\ Q(\mathbf{l}_1; \mathbf{A}_{\mathbf{l}_5}) &= Q(\mathbf{l}_1; \mathbf{A}_{\mathbf{l}_1})G(1) = F(1)G(1), \\ Q(\mathbf{l}_2; \mathbf{A}_{\mathbf{l}_5}) &= Q(\mathbf{l}_2; \mathbf{A}_{\mathbf{l}_2})G(1) = F(1)G(1). \end{aligned}$$

This leads to

$$Q(\mathbf{l}_5; \mathbf{A}) = [1 - G(0)^2 - 2F(1)G(1)]G(2)G(0).$$

This process can obviously be carried out systematically for the 5 other configurations with 2 active nodes. We then proceed to larger configurations until we reach $\mathbf{l} = \mathbf{n}$, which requires a special treatment. Rather than using Eq. (2.4a), we directly use the normalization (2.4b) to find $Q(\mathbf{n}; \mathbf{A})$. This completes the calculation of \mathcal{Q} on \mathbf{A} , and all other distributions $\mathcal{Q}' < \mathcal{Q}$ have been computed in the process.

2.10.3 Complexity calculation

In the main text, we mention that the exact recursive solution of Eqs. (2.4) is much more computationally complex than the accelerated solution, found using Eqs. (2.5b), since the latter skips the computation of \mathcal{Q}' on the induced subgraphs. This Appendix clarifies and quantifies this statement. We consider the worst case : An undirected complete graph of N nodes, with an undefined dynamics and no initial activation, i.e., $|\mathbf{l}_0|^2 = 0$. In this setup, the number of possible configurations is 2^N and the BFS is unnecessary. We assume that the complexity of the process is well represented by the number of $Q(\mathbf{l}; \mathbf{A}_{\mathbf{u}})$ required to obtain a solution.

We begin with the analysis of the complexity of the algorithm of Sec. 2.6.3 (the accelerated solution). For a configuration \mathbf{l} , we count the number of terms $\mathcal{N}_{|\mathbf{l}|^2}$ involved in the evaluation of $Q(\mathbf{l}; \mathbf{A})$ from Eq. (2.5b).

$$\mathcal{N}_{|\mathbf{l}|^2} = O\left(\sum_{|\mathbf{u}| < |\mathbf{l}|} Q(\mathbf{u}; \mathbf{A}_{\mathbf{n}})\right). \quad (2.14)$$

For each $|\mathbf{u}|^2$ of the sum, there exists $\binom{|\mathbf{l}|^2}{|\mathbf{u}|^2}$ different configurations, meaning that

$$\mathcal{N}_{|\mathbf{l}|^2} = \sum_{|\mathbf{u}|^2=0}^{|\mathbf{l}|^2-1} \binom{|\mathbf{l}|^2}{|\mathbf{u}|^2} \sim 2^{|\mathbf{l}|^2} \quad (2.15)$$

For a configuration \mathbf{l} , roughly $2^{|\mathbf{l}|^2}$ terms are needed to obtain a solution to Eq. (2.5b). To obtain the complexity of the complete distribution $\{Q(\mathbf{l}; \mathbf{A})\}_{\mathbf{l}}$, we sum up the complexity of all 2^N equations,

$$\mathcal{N}_{\text{total}} = \sum_{\mathbf{l}} \mathcal{N}_{|\mathbf{l}|^2} \quad (2.16)$$

$$= \sum_{|\mathbf{l}|^2=0}^N \binom{N}{|\mathbf{l}|^2} 2^{|\mathbf{l}|^2} = 3^N. \quad (2.17)$$

We conclude that the total number of operations scales approximately as 3^N for a complete set of configurations for an arbitrary dynamics on a complete graph, using the algorithm of Sec. 2.6.3.

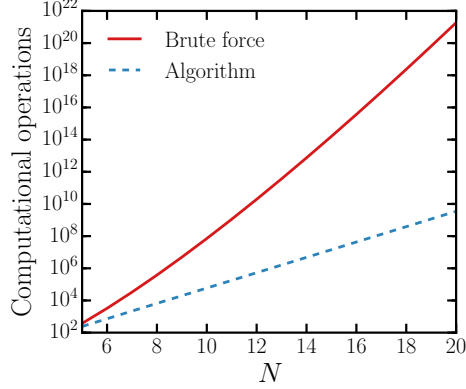


FIGURE 2.5 – Worst-case complexity of a brute-force solution of the Eqs. (2.4), and of the accelerated solution Eq. (2.5b). The complexity is assumed to be proportional to the number of $Q(\mathbf{l}; \mathbf{A}_{\mathbf{u}})$ required to obtain a solution. The worst-case corresponds to an arbitrary dynamics with spontaneous activation on a complete graph of N nodes.

Next, we analyze the complexity of the naive recursion algorithm (see Sec. 2.6.1) in the same manner. We call this method the brute-force method, since it does not use any shortcuts. Again, the complexity is represented by the number of $Q(\mathbf{l}; \mathbf{A}_{\mathbf{u}})$ required to obtain a solution. The dominant contribution to the complexity comes from Eq. (2.4b), i.e., the normalization which needs to be evaluated for every Q' on all the induced subgraphs. We denote by $\mathcal{M}_{|\mathbf{l}|^2}$ the complexity of the brute force calculation of $Q(\mathbf{l}; \mathbf{A}_{\mathbf{u}})$. The complexity can be written as a recursive equation :

$$\mathcal{M}_{|\mathbf{l}|^2} = O \left(\sum_{|\mathbf{u}| < |\mathbf{l}|} Q(\mathbf{u}; \mathbf{A}_{\mathbf{l}}) \right), \quad (2.18)$$

$$\mathcal{M}_{|\mathbf{l}|^2} = \sum_{|\mathbf{u}|^2=0}^{|\mathbf{l}|^2-1} \mathcal{M}_{|\mathbf{u}|^2} \binom{|\mathbf{l}|^2}{|\mathbf{u}|^2}, \quad (2.19)$$

since the normalization must be invoked all the way down to the initial condition. Finally, summing $\mathcal{M}_{|\mathbf{l}|^2}$ over each configuration to obtain the total complexity for a complete set of outcomes, we find

$$\mathcal{M}_{\text{total}} = \sum_{|\mathbf{l}|^2=0}^N \binom{N}{|\mathbf{l}|^2} \mathcal{M}_{|\mathbf{l}|^2}. \quad (2.20)$$

For large $|\mathbf{l}|^2$, the last element of the series, i.e., $|\mathbf{u}|^2 = |\mathbf{l}|^2 - 1$, is dominant and $\mathcal{M}_{|\mathbf{l}|^2} \sim |\mathbf{l}|^2!$ (see Eq. (2.19)). Thus, $\mathcal{M}_{\text{total}} \sim N!$ for large N .

Figure 2.5 compares Eqs. (2.17) and (2.20) for different graph sizes. It is clear that the brute-force method becomes impractically *much faster* than the accelerated method.

Chapitre 3

Réduction dimensionnelle spectrale des systèmes complexes dynamiques

Spectral dimension reduction of complex dynamical networks

Edward Laurence^{1,2}, Nicolas Doyon^{2,3,4}, Louis J. Dubé^{1,2} et Patrick Desorisiers^{1,2,4}

¹ Département de Physique, de Génie Physique, et d'Optique
Université Laval, Québec, Québec, Canada G1V 0A6

² Centre Interdisciplinaire en Modélisation Mathématique de l'Université Laval (CIMMUL)
Université Laval, Québec, Québec, Canada G1V 0A6

³ Département de mathématiques et de statistique
Université Laval, Québec, Québec, Canada G1V 0A6

⁴ Centre de recherche CERVO
Université Laval, Québec, Québec, Canada G1V 0A6

Référence : Phys. Rev. X, 9 (2019), p. 011042.

© 2019 The American Physical Society

...

3.1 Avant-propos

Un des enjeux importants des systèmes réels est la présence de transitions abruptes. Comme les dynamiques sont de hautes dimensions, il n'est pas évident de tracer le diagramme des bifurcations afin de prévoir ces transitions. Dans ce chapitre, nous travaillerons ce problème en introduisant une nouvelle approche de réduction dimensionnelle de réseaux dynamiques. Cette technique vise à produire un système à 1 ou 2 dimensions qui reproduit une observable globale du système réel comme l'activité moyenne. Ce passage en faible dimension nous permettra de documenter simplement les bifurcations de la réduction.

...

3.2 Résumé

Les réseaux dynamiques sont des outils puissants pour modéliser une vaste gamme de systèmes réels, dont les marchés financiers, le cerveau et les écosystèmes. Ils expriment comment les éléments interagissent tous ensemble et évoluent. Malgré les récents progrès, plusieurs inconnus persistent à savoir pourquoi de légers changements à la structure peuvent mener à un changement important de la dynamique. Nous traitons cet enjeu en introduisant une méthode pour réduire la dimensionnalité des réseaux complexes.

L'originalité de cette méthode est l'utilisation des vecteurs propres de la matrice d'adjacence pour réduire le graphe des interactions. Nous vérifierons sa validité sur des dynamiques de propagation, de populations et de neurones et des structures modulaires, aléatoires et biparties. Les résultats suggèrent que cette approche pourrait être d'un intérêt à la fois fondamental et pratique pour la détection des transitions critiques.

...

3.3 Abstract

Dynamical networks are powerful tools for modeling a broad range of complex systems, including financial markets, brains, and ecosystems. They encode how the basic elements (nodes) of these systems interact altogether (via links) and evolve (nodes' dynamics). Despite substantial progress, little is known about why some subtle changes in the network structure, at the

so-called critical points, can provoke drastic shifts in its dynamics. We tackle this challenging problem by introducing a method that reduces any network to a simplified low-dimensional version. It can then be used to describe the collective dynamics of the original system. This dimension reduction method relies on spectral graph theory and, more specifically, on the dominant eigenvalues and eigenvectors of the network adjacency matrix. Contrary to previous approaches, our method is able to predict the multiple activation of modular networks as well as the critical points of random networks with arbitrary degree distributions. Our results are of both fundamental and practical interest, as they offer a novel framework to relate the structure of networks to their dynamics and to study the resilience of complex systems.

...

3.4 Introduction

Critical breakdowns generally arise unexpectedly in complex dynamical systems [144]. Noteworthy examples are financial crises [60, 149], epileptic seizures [142], and species extinctions [156]. These breakdowns are typically identified by using global-scale indicators that collapse at the critical point, such as stock market indices, neural synchronization, and species biomass. While much effort has been devoted to forecast breakdowns [23], no simple and universal method has yet been found. This is mostly due to the inherent complexity of the problem : real systems are composed of multiple units that participate to the global state in highly complicated patterns of interactions.

Network Science addresses this problem and offers a unifying framework where a complex system with N fundamental units is described as a network of N components (nodes). The state of each unit is encoded into an activity variable and the evolution of the states in the whole system is governed by N coupled dynamical equations that depend on both the activity variables and a set of weighted interactions (links). Therefore, the dynamical properties of the system strongly depend on the underlying network structure. Although recent development has clarified how small targeted perturbations in the network structure can provoke drastic changes in the structure itself [103, 121], much less is known about the dynamical effects of these perturbations. One promising approach is to use dimension reduction to transform the original N -dimensional representation into a simplified version with $n \ll N$ effective dimensions.

Recently, Gao *et al.* have presented a dimension reduction formalism that collapses any N -dimensional network into a 1-dimensional effective version and used it to predict the global activity of the original network [52]. The authors have proposed to measure the global activity as the degree-weighted average activity in which the nodes with high degree, i.e., high number

of links, contribute more to the average than those with low degree. The rationale behind this choice is that the highly connected nodes have a higher impact on the dynamics. Moreover, they have shown that the degree-weighted connectivity is sufficient to explain the global level of activity. Their formalism can be applied to a wide variety of complex systems, thus suggesting that the degree-weighted averages are in fact *universal* predictors. For instance, they accurately predict the minimum level of interaction between species to prevent biomass extinction, the so-called *critical point*, of real ecosystems. Yet, this spectacular outcome is not totally satisfactory since no fundamental reason is provided that would explain why degree is the key property to any network structure, particularly those with degree correlations.

In an attempt to determine the critical points of 59 bipartite mutualistic ecosystems [83], Jiang *et al.* have proposed a 2-dimensional reduction that divide each original ecosystem into two populations for which they obtain the average interaction strength. From numerical explorations, they conclude that the degree may not always be the key predictive property of a network. Their results also suggest that 2-dimensional reductions can lead to better predictions than the 1-dimensional formalism of Gao *et al.*

It remains unclear whether the dimension reduction procedure of Gao *et al.* can lead to accurate predictions for arbitrary network structures and why some dynamical networks should require 2-dimensional reductions. Besides, strong theoretical foundations are clearly lacking to answer those questions. The goal of this paper is to address these issues. We rely on a simple and strong hypothesis : one can predict the evolution of a small number of variables describing the global activities of a network. These variables are constructed a priori unknown weighted-averages of the individual node activities. By enforcing this hypothesis, we provide theoretical justifications for the required number of effective dimensions and quantify the contribution of each component to the universal global activities. Beyond the mere improvement in precision over existing approaches, our method allows the detection of dynamical breakdowns that would be missed altogether with previous reductions.

This study provides a reliable tool for researchers who want to study critical breakdowns of complex systems. Using our approach, once a system is framed as a network, one can first determine the number of effective dimensions required to adequately reduce the system at hand, and then find the variables of interest that describe the global state. Moreover, our method also identifies the units that, if perturbed, can induce large reactions in the system. Our findings thus lead to a deeper understanding of how critical breakdowns occur and how to prevent them.

The paper is structured as follows. We first present the general framework for complex dynamical networks (Sec. 3.5). We then describe a general method to obtain a 1-dimensional reduction for mutualistic networks (Sec. 3.6). We show that the reduction scheme of Gao *et al.* emerges as an approximation of our general approach when specifically considering random

networks. In Sec. 3.7 we develop the *cycle reduction*, a multidimensional approach useful to reduce heterogeneous and bipartite networks. Next, in Sec. 3.8, we complete the method by including subdominant contributions of the structure. We finally assess the goodness of these reductions, as a function of the structure, and the nature of the dynamics (Sec. 3.9).

...

3.5 Model definition

The diverse nature of complex systems requires to establish a common ground. In Sec. 3.5.1, we regroup dynamical complex networks under a general model that encodes the structure and the dynamics. Then, in Sec. 3.5.2, we provide examples of contrasting models of dynamics satisfying the formalism used afterward in the paper to illustrate the dimension reduction methods.

3.5.1 General formalism

We consider a complex network of N units, called *nodes*, for which the interactions are encoded in the weighted and directed adjacency matrix \mathbf{W} . The element $w_{ij} \in \mathbb{R}$ of \mathbf{W} is interpreted as the strength of the directed interaction from node j to node i .

Each node has an activity $x_i \in \mathbb{R}$ whose evolution is governed by the general equation

$$\dot{x}_i = F(x_i) + \sum_{j=1}^N w_{ij} G(x_i, x_j), \quad (3.1)$$

where $F(x_i)$ and $G(x_i, x_j)$ are real-valued functions. For technical reasons that will become clear in the next section, both $F(x_i)$ and $G(x_i, x_j)$ are required to have continuous derivatives of second order. The product $w_{ij}G(x_i, x_j)$ specifies the type of interactions.

If $w_{ij}\partial G(x_i, x_j)/\partial x_j \leq 0$, the interaction is competitive and the increase of activity of node j tends to decrease the activity of node i . If $w_{ij}\partial G(x_i, x_j)/\partial x_j \geq 0$, the interaction is mutualistic and therefore node j activity benefits node i . For mixed dynamics of mutualistic and competitive interactions, it is common to fix $\partial G(x_i, x_j)/\partial x_j \geq 0$ and use negative weights $w_{ij} < 0$ for competitive interactions. Unless specified, we will only consider mutualistic dynamics with $w_{ij}\partial G(x_i, x_j)/\partial x_j \geq 0$. Furthermore, we have concentrated our studies of the possible dynamical models (Sec. 3.5.2) in parameter ranges for which $\partial G(x_i, x_j)/\partial x_j \geq 0$ such that w_{ij} are all non-negative as well.

To describe the evolution of the whole system at the macro- and the mesoscopic scales, it is convenient to focus on *observables*. We define an observable as a *smooth* function mapping the

activities x_1, \dots, x_N to a real number. Among all observables, the linear observables, functions of the form $L(\mathbf{x}) = \sum_i a_i x_i \in \mathbb{R}$, are of particular interest for different reasons.

The first reason to use a linear observable is the numerical evidence given by Ref. [52] that suggests that a linear combination of the activity could be a good indicator of the global state of the network. We have followed this direction, which has led us to a more general linear dimension reduction formalism. The second reason is that a linear combination is far more intuitive than a general non-linear observable of the form $\Omega(\mathbf{x}) = \sum_{n,j} a_{nj} x_j^n$. It is not excluded that such non-linear observables could provide adequate dimension reductions, but their interpretability would most surely be limited. Finally, there is a more formal and practical reason to select linear observables. The set of all observables, which are smooth functions from \mathbb{R}^N to \mathbb{R} , forms a real vector space \mathcal{V} of infinite dimension. Dimension reduction can thus be seen as a search for $n < N$ observables, among infinitely many possible choices, whose evolution can be well approximated by n coupled differential equations. At first sight, there is no obvious way to determine the best observables to choose. However, \mathcal{V} contains a subspace \mathcal{L} formed by all linear observables. The dimension of \mathcal{L} is precisely N and one can show that any basis of \mathcal{L} has a dynamics that is exactly described by N differential equations very similar to the original ones. This simplifies the search immensely : rather than looking in the whole space \mathcal{V} , the spectral properties of the adjacency matrix to find n elements of interest in \mathcal{L} offer themselves as a natural choice.

3.5.2 Examples of possible dynamics

A number of dynamical systems satisfy the form of Eq. (3.1). For instance, in computational neuroscience, the *Wilson-Cowan model* [171] describes the firing-rate activity of a population of neurons as

$$\dot{x}_i = -x_i + \sum_{j=1}^N w_{ij} \frac{1}{1 + \exp[-\tau(x_j - \mu)]}, \quad (3.2a)$$

where τ and μ are parameters controlling the steepness of the activation function and the firing-rate threshold, respectively¹.

In biology, the *generalized Lotka-Volterra dynamics* describes the evolution of the population of species in an ecosystem as

$$\dot{x}_i = \omega x_i + x_i \sum_{j=1}^N w_{ij} x_j, \quad (3.2b)$$

where ω is the intrinsic growth rate [59], and x_i is the population of individuals of species i . To prevent unbounded growth and account for species migration and the *Allee effect*, a more

1. This is actually a modified version of the original Wilson-Cowan model, which describes output activities of the specific form of Eq. (3.1). In Eq. (3.2a), x_i must be interpreted as the input activity to neuron i .

complex model of ecological networks has been proposed [52, 72] :

$$\begin{aligned} \dot{x}_i = & B_i + x_i \left(1 - \frac{x_i}{K_i} \right) \left(\frac{x_i}{C_i} - 1 \right) \\ & + \sum_{j=1}^N w_{ij} \frac{x_i x_j}{D_i + E_i x_i + H_i x_j}, \end{aligned} \quad (3.2c)$$

where all parameters are real-valued, B_i accounts for the migration rate, $K_i > 0$ for the ecosystem capacity, and $C_i > 0$ for the minimum abundance for species growth. The parameters D_i, E_i, H_i control the strengths of the interactions between the species.

The *Michaelis-Menten equation* is yet another example [8]. It applies to the gene regulatory networks and governs the concentration of substrates as

$$\dot{x}_i = -c x_i^a + \sum_{j=1}^N w_{ij} \frac{x_j^b}{x_j^b + 1}, \quad (3.2d)$$

where $a, b, c \in \mathbb{R}$ are parameters.

In social networks, the spreading of a virus or rumors can be described using the *Susceptible-Infected-Susceptible model* (SIS) [136]. In this context, the activity $x_i \in [0, 1]$ is interpreted as the probability of being infected and evolves according to

$$\dot{x}_i = -x_i + \gamma(1 - x_i) \sum_{j=1}^N w_{ij} x_j, \quad (3.2e)$$

with $\gamma \geq 0$ as the normalized infection rate.

...

3.6 1–dimensional reduction

The systems described by Eqs. (3.2) are N –dimensional, their dynamics governed by N coupled differential equations. As the number of nodes N grows, the computational cost of solving N coupled equations increases which raises a number of issues [46, 97]. Moreover, the state of the original system given by the N –dimensional vector \mathbf{x} becomes less intelligible and less insightful, and does not provide much into the general properties of the solutions.

Hence, we must rely on measures, or observables, to reduce N –dimensional systems to more practical and accessible objects. For instance, the unweighted average activity could be a measure on how dissimilar the system state is compared to a specifically chosen state. We also want to make predictions on those measures to anticipate dynamical breakdowns and locate the global state of the system on a standardized bifurcation diagram. However, when solely

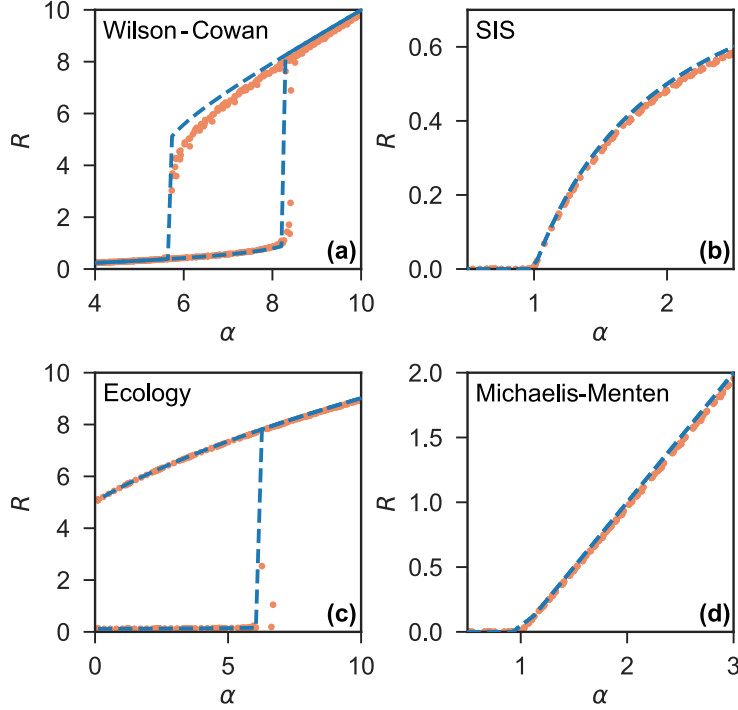


FIGURE 3.1 – Observable $R^* = \mathbf{a}^T \mathbf{x}^*$ at equilibrium as a function of the dominant eigenvalue α of \mathbf{W}^T , for different dynamics on Erdős-Rényi networks of $N = 100$ nodes and connection probability $p = 0.1$. (a) Wilson-Cowan dynamics Eq. (3.2a) with $\tau = 1, \mu = 3$, (b) SIS dynamics Eq. (3.2e) with $\gamma = 1$, (c) Mutualistic ecological dynamics Eq. (3.2c) with $B_i = 0.1, C_i = 1, K_i = 5, D_i = 6, E_i = 0.9, H_i = 0.1$ [14], (d) Michaelis-Menten dynamics Eq. (3.2d) with $a = 1, b = 1, c = 1$. Dashed lines are theoretical predictions obtained from Eq. (3.9) while dots are equilibrium states resulting from the evolution of the whole N -dimensional system. For each dynamics and network ensemble, 100 networks are generated. For each network, we scale the edge weights by a constant random factor as $w_{ij} \mapsto cw_{ij}$, so that the dominant eigenvalue of the adjacency matrix is located in the region of interest. Then, the dynamics are integrated to equilibrium and an orange dot is placed at the corresponding point (α, R^*) . Next, the network is perturbed by removing an edge and the dynamics is brought back to equilibrium, and a new dot at (α', R'^*) is placed. The perturbation step is repeated 50 times for each network.

based on the unweighted average activity, the predictions are often non-representative of the original system [83].

Alternatively, a *weighted* activity seems more reliable as we inject additional information on the importance of nodes and has already been proven to be a promising avenue of breakdown predictions [52, 83]. In the next subsections, we introduce a general procedure to select a weighted activity and to predict its evolution.

3.6.1 Derivation of the reduction formalism

Let us consider a real linear observable R of the activity :

$$R = \sum_{i=1}^N a_i x_i = \mathbf{a}^T \mathbf{x}, \quad (3.3)$$

where $a_i \in \mathbb{R}$, the i -th component of the column vector \mathbf{a} , is a normalized weight so that

$$\sum_{i=1}^N a_i = \mathbf{1}^T \mathbf{a} = 1. \quad (3.4)$$

In general, some components a_i can be positive or negative, and R represents a weighted activity. In many instances however, a_i will be non-negative, i.e. the normalized vector \mathbf{a} will be a probability vector, and our observable R could then be called justifiably a weighted *average* activity, where a_i is the relative contribution, or *centrality*, of node i to the observable.

The linear observable R is a function that takes the instantaneous activity of each node and returns a real number that describes the global state of the network. For instance, for $a_i = 1/N$, R describes the unweighted average activity. Although the average activity is attractive because of its simplicity, it may not be easy to predict its value using only the structure of the network and the nature of the dynamics. Thus, we hypothesize that \mathbf{a} should be specific to the structure.

Let us explain how the weight vector \mathbf{a} is constrained by the adjacency matrix \mathbf{W} . By taking the time derivative of Eq. (3.3) and using Eq. (3.1) (refer to Appendix 3.12.1 for complete derivation), we obtain that the dynamics of R – truncated up to second-order terms $O[(x_k - R)^2]$ – is given by the 1–dimensional equation

$$\dot{R} \approx F(R) + \alpha G(\beta R, R), \quad (3.5)$$

where β is a structural parameter given by

$$\beta = \frac{1}{\alpha} \frac{\mathbf{a}^T \mathbf{K} \mathbf{a}}{\mathbf{a}^T \mathbf{a}} \quad (3.6)$$

and \mathbf{K} is a $N \times N$ diagonal matrix of diagonal elements $K_{ii} = k_i^{\text{in}} = \sum_{j=1}^N w_{ij}$, the in-degree of node i . The parameter α can be measured directly on the network as the weighted in-degree,

$$\alpha = \sum_{i=1}^N a_i k_i^{\text{in}} = \mathbf{a}^T \mathbf{k}^{\text{in}}. \quad (3.7)$$

Interestingly, we show, in Appendix 3.12.1, that the closed form of Eq. (3.5) is satisfied only if \mathbf{a} is a normalized eigenvector of the transposed adjacency matrix \mathbf{W}^T with eigenvalue α ,

$$\mathbf{W}^T \mathbf{a} = \alpha \mathbf{a}. \quad (3.8)$$

We now have obtained a single equation [Eq. (3.5)] that governs the evolution of the weighted activity of a complex network, and constrained the weight vector \mathbf{a} to be adapted to the structure under study.

Clearly, Eq. (3.5) shares similarities with Eq. (3.1). We can interpret the former as a reduced system of one dynamical node that interacts with itself. The nature of its dynamics is identical to the one from the original system, i.e. specified by $F(x_i)$ and $G(x_i, x_j)$, and the coupling is parametrized by α, β . The activity R of the single node describes the weighted activity of the original network, and is obtained by solving Eq. (3.5) which solutions are solely controlled by the nature of the dynamics and the structural parameters α, β .

3.6.2 Choice of a universal weight vector

We have seen that the weight vector \mathbf{a} must be a normalized eigenvector of \mathbf{W}^T so that the observable R satisfies Eq. (3.5). In principle, any eigenvectors of \mathbf{W}^T , except those that satisfy $\mathbf{1}^T \mathbf{a} = 0$, could be used for the dynamical reduction. However, the larger the modulus of α is, the stronger is the influence of the structure on the weighted activity. An eigenvector whose eigenvalue has a low modulus leads to a linear observable R that does not properly take into account the network structure, which in turn leads to correction terms $O[(x_k - R)^2]$ greater than those produced by eigenvectors with a higher modulus. The choice of \mathbf{a} as the eigenvector with the largest eigenvalue modulus seems to impose itself : \mathbf{a} is the dominant eigenvector.

For an arbitrary weighted adjacency matrix, the dominant eigenvalue and the components of the dominant eigenvector can be complex. In this case, the observable R as well as the structural parameters α and β are complex. The 1-dimensional dynamical system of Eq. (3.5) becomes complex too and can be interpreted as a 2-dimensional real dynamical system.

There is however a large class of networks for which the dominant eigenvalue and the components of the dominant eigenvector are all real. For instance, strongly connected (in practice, sufficiently connected) undirected and directed networks with non-negative weights w_{ij} and fall into this class. In fact, the Perron-Frobenius theorem guarantees that if the network is strongly connected, i.e. a path exists between each pair of nodes, and that all edge weights satisfy $w_{ij} \geq 0$, then the dominant eigenvalue λ_D of \mathbf{W}^T is non-negative $\lambda_D \geq 0$, and the dominant eigenvector is elementwise positive [78]. Moreover, in practice, the dominant eigenvector can be efficiently computed using the power method.

The procedure to apply this 1-dimensional dimension reduction is straightforward. First, we compute the dominant eigenvalue α and the corresponding eigenvector \mathbf{v}_D of \mathbf{W}^T . Second, we define the normalized eigenvector $\mathbf{a} = \mathbf{v}_D / (\mathbf{1}^T \mathbf{v}_D)$, and obtain β according to Eq. (3.6). In most cases, we want to determine the weighted activity at equilibrium R^* , determined by

solving

$$0 = F(R^*) + \alpha G(\beta R^*, R^*). \quad (3.9)$$

This is a universal equation in the sense that α and β are independent of the dynamics, controlled by $F(x_i)$ and $G(x_i, x_j)$; α and β only depend upon the network structure, encoded in \mathbf{W} . The 1-dimensional reduction process has been applied to different dynamics for small random uncorrelated networks and led to surprisingly accurate predictions (Fig. 3.1). For larger networks, we expect the formalism to maintain a similar level of accuracy. However, since the number of nodes does not intervene explicitly in the formalism, we are generally unable to analytically describe how the quality of the reduction varies with the network size. We must rely on a numerical investigation.

Our numerical experiments indicate that the network size *by itself* has no significant impact on the quality of the reduction. Rather, the accuracy strongly depends on the network structure and, in particular, on the degree variance. In a nutshell, our findings can be summarized as follows :

- a. For a given N , the larger the average *number of edges per node* $\langle \mathbf{k}_B^{in} \rangle^2$, the better the dimensional reduction will be ;
- b. For a fixed value of $\langle \mathbf{k}_B^{in} \rangle$, the residual error of the reduction tends rapidly to a finite limit as N is increased ;
- c. For large enough $\langle \mathbf{k}_B^{in} \rangle$, the reduction error is small and insensitive to the precise value of N (Corollary of a. and b.).

In Fig. 3.2, we examine these conclusions by comparing different network ensembles. Although displayed for undirected Erdős-Rényi (ER) networks, $G(N, p)$, the results are representative of a larger set of calculations, and synthesize nicely our general conclusions on this issue. In Fig. 3.2(a), dynamics on networks of different $\langle \mathbf{k}_B^{in} \rangle = p(N - 1)$ but equal number of nodes $N = 200$, are differently reproduced by the 1-dimensional reduction. Denser networks (large p , large $\langle \mathbf{k}_B^{in} \rangle$) are better represented by the reduction than sparser (small p , small $\langle \mathbf{k}_B^{in} \rangle$) networks. In Fig. 3.2(b), we compare networks of different sizes but equal $\langle \mathbf{k}_B^{in} \rangle$. For a fixed $\langle \mathbf{k}_B^{in} \rangle$ (here $\langle \mathbf{k}_B^{in} \rangle = 10$), the connection probability p is adjusted to compensate for the growing number of nodes. As the number of nodes is increased, the residual error on the prediction rapidly tends to a finite limit, and no further deterioration of the quality of the dimensional reduction is observed. In other words, the goodness of the reduction is practically invariant of the network size, if large enough, and is mostly governed by the average number of edges per node.

2. The average number of edges per node $\langle \mathbf{k}_B^{in} \rangle = N^{-1} \langle \mathbf{1}^T \mathbf{A} \mathbf{1} \rangle$ is computed from the binary adjacency matrix \mathbf{A} and must be distinguished from the average degree $\langle \mathbf{k}^{in} \rangle = N^{-1} \langle \mathbf{1}^T \mathbf{W} \mathbf{1} \rangle$ computed from the weighted adjacency matrix \mathbf{W} . In the latter equations, $\langle \mathbf{1}^T \mathbf{A} \mathbf{1} \rangle$ either denotes the expected value of $\mathbf{1}^T \mathbf{A} \mathbf{1} = \sum_{i,j} A_{ij}$ over a random network ensemble (in the case of random networks such as ER networks) or simply $\sum_{i,j} A_{ij}$ (in the case of a particular network), and similarly for $\langle \mathbf{1}^T \mathbf{W} \mathbf{1} \rangle$.

These observations extend to other types of networks and confirm that the quality of the reduction is more affected by the network connectivity than by the network size. A dynamical explanation goes as follows. In sparse networks (e.g. $N = 200, p = 0.02$), one must use large scaling factors $c > 1$ ($\mathbf{W} \mapsto c\mathbf{W}$ which implies $\mathbf{W}\mathbf{1} = \mathbf{k}^{in} \mapsto c\mathbf{k}^{in}$, $\alpha = \mathbf{a}^T \mathbf{k}^{in} \mapsto c\alpha$) to reach the desired range of the dominant eigenvalue $\alpha \in [4, 10]$ (see caption of Fig. 3.2 : the scaling does not alter the topology, only the strength of what is injected in the dynamical equations through the parameter α). This accentuates the inequalities between the degrees of the nodes, $\text{Var}(\mathbf{k}^{in}) \mapsto c^2 \text{Var}(\mathbf{k}^{in})$, and eventually splits the populations into active and inactive nodes. Hence, the reduction is unable to describe the two populations with a single variable. In the opposite extreme of dense networks (e.g. $N = 200, p = 0.25$), the edge weights must be scaled down, $c < 1$, to reach the same dominant eigenvalue. Therefore, the nodes follow a more global activation scheme that can be described with high accuracy by the 1-dimensional reduction. This explanation matches the observed finite limit for the quality of the reduction [Fig. 3.2(b)]. For a fixed value of $\langle \mathbf{k}_B^{in} \rangle = p(N-1)$, when N grows larger, p tends to zero so that the variance $\text{Var}(\mathbf{k}_B^{in}) \approx (N-1)p(1-p)$ tends to a limit $\text{Var}(\mathbf{k}_B^{in}) \rightarrow \langle \mathbf{k}_B^{in} \rangle$ and no further deterioration of the quality is observed.

The quantitative impact of the degree variance is best explained by a simple example. Consider an ER network, ($N = 200, p = 0.1$) with $\langle \mathbf{k}_B^{in} \rangle \approx 20$ and $c = 1$ so that the variance of the edge weights is $\text{Var}(\mathbf{k}^{in}) \approx 18$ and the dominant eigenvalue is $\alpha \approx 20$. For a denser network ($N = 200, p = 0.5$) with $\langle \mathbf{k}_B^{in} \rangle \approx 100$,

one must use a small scaling factor, $\mathbf{W} \mapsto \mathbf{W}/5$, to have the same dominant eigenvalue $\alpha \approx 20$, and the scaled variance of the edge weights is now $\text{Var}(\mathbf{k}^{in}) \approx 2$. Therefore, the nodes in the denser networks are more uniformly activated than in the sparser networks and are better amenable to the 1-dimensional reduction.

3.6.3 Choice of an approximate weight vector

Recently, Gao *et al.* [52] have introduced a different 1-dimensional reduction for dynamics of the form of Eq. (3.1). In this section, we show how their reduction is a special case of our 1-dimensional reduction when applied to uncorrelated random networks.

Uncorrelated random networks are a family of networks for which the degree distribution can be arbitrary but the probability of connection between two nodes is independent of the presence or absence of any other edge [128]. We generate our random networks using the configuration model [124]. We first sample the nodes in- and out- expected degrees $\kappa^{in}, \kappa^{out}$ from an arbitrary degree distribution. Then, we connect node j to node i with probability

$$P_{ij} = \frac{\kappa_i^{in} \kappa_j^{out}}{m}, \quad (3.10)$$

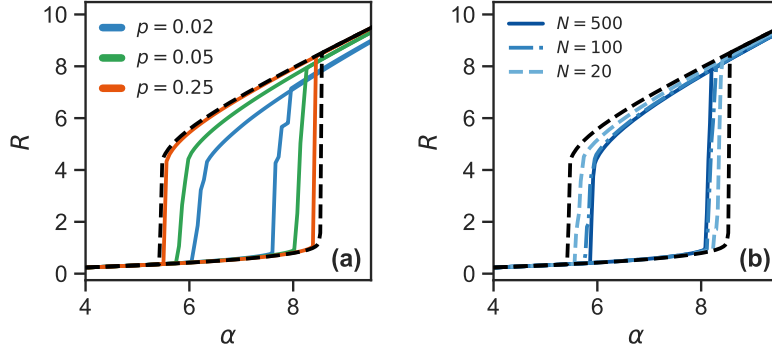


FIGURE 3.2 – (a) Comparison of the observable $R^* = \mathbf{a}^T \mathbf{x}^*$ at equilibrium as a function of the dominant eigenvalue α of \mathbf{W}^T for different connection probabilities p of undirected Erdős-Rényi networks $G(N, p)$ and the Wilson-Cowan dynamics Eq. (3.2a) with $\tau = 1, \mu = 3$. For each value of p , 10 networks of 200 nodes are generated and their dynamics are solved for $\alpha \in [4, 10]$. Lines are computed as a binned average over same connection probability networks. (b) Comparison of Erdős-Rényi networks $G(N, p)$ of different sizes $N \in \{20, 100, 500\}$ but equal average number of edges per node $\langle k_B^{in} \rangle = 10$. For each N , we adjust the connection probabilities $p \in \{0.50, 0.10, 0.02\}$ to match the expected value of $\langle k_B^{in} \rangle$ and solve the dynamics on 10 network realizations. Lines are computed as a binned average over same size networks. Dashed lines are theoretical predictions obtained from Eq. (3.9). To obtain a specific $\alpha = \mathbf{a}^T \mathbf{k}^{in}$, we multiply each edge weight by a constant scaling factor $w_{ij} \mapsto cw_{ij}$ so that the dominant eigenvalue $\alpha \mapsto c\alpha$ falls in the range $[4, 10]$ using the dominant eigenvector \mathbf{a} as weight vector. We then solve the dynamics at equilibrium and measure the observable $R^* = \mathbf{a}^T \mathbf{x}^*$.

where $m = \sum_i \kappa_i^{\text{out}}$ is the expected total number of edges. If the resulting network is strongly connected, the Perron-Frobenius theorem guarantees that the dominant eigenvector \mathbf{v}_D of \mathbf{W}^T will have only non-negative elements. We may then use this dominant eigenvector to construct the observable

$$R = \mathbf{a}^T \mathbf{x} = \frac{\mathbf{v}_D^T \mathbf{x}}{\mathbf{1}^T \mathbf{v}_D}. \quad (3.11)$$

For networks that satisfy Eq. (3.10), spectral graph theory [35] informs us that the elements of the dominant eigenvector \mathbf{v}_D of \mathbf{W}^T (i.e. the weights of the reduced system) can be approximated by the vector of out-degrees \mathbf{k}^{out} as

$$a_i = [\mathbf{v}_D]_i \approx \frac{k_i^{\text{out}}}{\sum_{j=1}^N k_j^{\text{out}}}. \quad (3.12)$$

if the rather mild condition

$$\frac{\langle (k^{\text{out}})^2 \rangle}{\langle k^{\text{out}} \rangle} > \sqrt{\max[k_i^{\text{out}}] \ln(N)}, \quad (3.13)$$

is satisfied. It then results from Eq. (3.7) that α measures the average neighbor in-degree, that is

$$\alpha \approx \frac{\sum_{i=1}^N k_i^{\text{out}} k_i^{\text{in}}}{\sum_{i=1}^N k_i^{\text{out}}}, \quad (3.14a)$$

and Eq. (3.6) reduces to $\beta = 1$. Therefore, R is simply the average neighbor activity :

$$R \approx \frac{\sum_{i=1}^N k_i^{\text{out}} x_i}{\sum_{i=1}^N k_i^{\text{out}}}. \quad (3.14b)$$

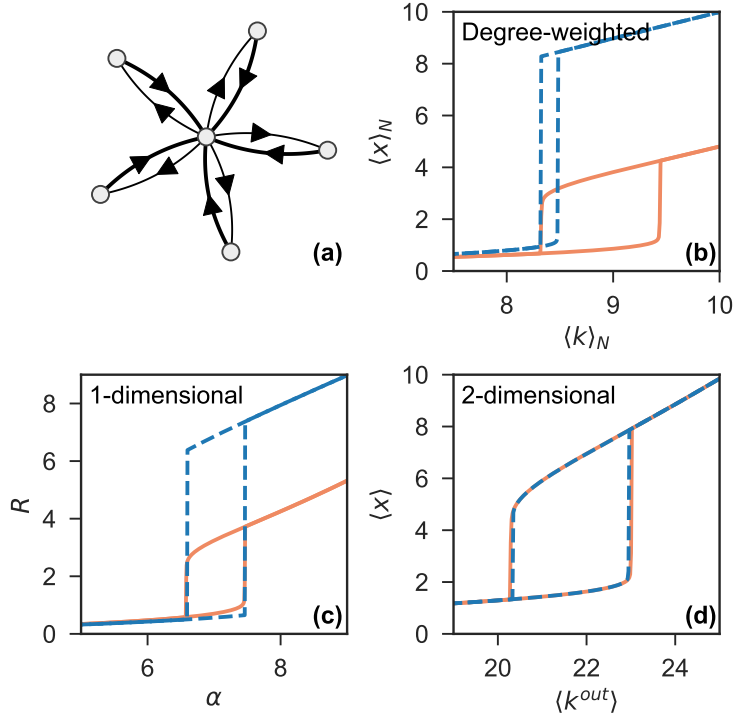


FIGURE 3.3 – (a) Schematisation of the star network of $N = 6$ nodes where the edge weight toward the core is twice the weight of an edge toward the periphery. (b) Average neighbor activity at equilibrium as a function of its structural parameter using the degree-weighted reduction of Gao *et al.*[52]. (c) Dominant eigenvector weighted activity at equilibrium $R^* = \mathbf{v}_D^T \mathbf{x}^*$ as a function of the dominant eigenvalue α for the 1-dimensional reduction [Eq. (3.5)]. (d) Average activity at equilibrium obtained by a combination of two observables $\langle x \rangle^* = N^{-1}[R_1^* + (N - 1)R_2^*]$, as a function of the average out-degree $\langle k^{out} \rangle = N^{-1}[\alpha_1 + (N - 1)\alpha_2]$ computed using the 2-dimensional reduction formalism [Eq. (3.28b)]. Full lines are results from simulations and dashed lines are theoretical predictions. The network dynamics is the Wilson-Cowan model with $\tau = 1, \mu = 3$ [Eq. (3.2a)].

It turns out that this special case is exactly the formalism proposed by Gao *et al.*[52] with $R = x_{\text{eff}}$ and $\alpha = \beta_{\text{eff}}$, in their notation. It also means that the formalism of Gao *et al.* is mostly appropriate for random networks [Eq. (3.10)] respecting Eq. (3.13). Moreover, a recent work [33] has introduced a corrected eigenvalue approximation for random networks with power-law degree distribution $p(k) \sim k^{-\gamma}$ with $\gamma > 5/2$. This may further limit the accuracy of Gao *et al.* approach with respect to our 1-dimensional reduction scheme.

...

3.7 Multidimensional reduction : dominant eigenvectors

In the 1–dimensional reduction, it has been supposed that the dynamical global state of a network is dominated by the information contained in a single dominant eigenvalue and corresponding eigenvector. Therefore, it also presupposes that other eigenvalues can be safely neglected and do not provide relevant information about the dynamics on the network. But, if the network admits many eigenvalues of similarly large modulus, it is plausible to expect that all these eigenvalues are important as well, and should be included in a n –dimensional reduction. In this section, we address this problem by introducing a n –dimensional approach to predict the evolution of n coupled observables. We argue that if the spectrum $\{\lambda_1, \lambda_2, \dots, \lambda_N\}$ of \mathbf{W}^T satisfies

$$|\lambda_1| \approx |\lambda_2| \approx \dots \approx |\lambda_n| \gg |\lambda_{n+1}| \geq \dots \geq |\lambda_N|,$$

then n observables should be considered, leading to a n –dimensional reduced dynamical system.

3.7.1 Cycle reduction

Let us consider n observables R_j , $1 \leq j \leq n$, each being a different linear combination of the activity

$$R_j = \sum_{i=1}^N [\mathbf{a}_j]_i x_i, \quad (3.15)$$

where \mathbf{a}_j is a real-value weight vector associated with the observable R_j and normalized $\sum_i [\mathbf{a}_j]_i = 1$. As in the 1–dimensional reduction, \mathbf{a}_j are yet undetermined. While there are several choices for \mathbf{a}_j that are a priori plausible, we discuss the *cycle* reduction, which is natural when several dominant eigenvalues are approximately of equal modulus.

Using a similar approach as the 1–dimensional reduction (See Appendix 3.12.2), one finds that the evolution of the observables is given by

$$\dot{R}_j \approx \begin{cases} F(R_j) + \alpha_j G(\beta_j R_j, R_{j+1}), & j < n \\ F(R_j) + \alpha_j G(\beta_j R_j, R_1), & j = n \end{cases} \quad (3.16)$$

where

$$\beta_j = \frac{1}{\alpha_j} \frac{\mathbf{a}_j^T \mathbf{K} \mathbf{a}_j}{\mathbf{a}_j^T \mathbf{a}_j} \quad (3.17)$$

and α_j is an observable of the weighted neighbor in-degree,

$$\alpha_j = \sum_{i=1}^N [\mathbf{a}_j]_i k_i^{\text{in}}. \quad (3.18)$$

To satisfy Eqs. (3.16), the weight vectors are constrained by the structure and must transform according to

$$\mathbf{a}_j = \frac{\mathbf{W}^T \mathbf{a}_{j-1}}{\alpha_{j-1}}, \quad (3.19)$$

preserves the positiveness and the required normalization. Moreover, Eq. (3.19) needs to be a periodic application, i.e. $\mathbf{a}_{j+n} = \mathbf{a}_j$, in order to close the system to n observables. The initial choice of \mathbf{a}_1 is then highly constrained to satisfy this condition. In the following section, we will explain how the weight vectors can be computed using the dominant eigenvectors of \mathbf{W}^T .

In contrast to Eq. (3.5) where a single observable is used, we now have developed a closed n -dimensional system of observables that are coupled by a set of structural parameters $\{\alpha_j, \beta_j\}_{j=1, \dots, n}$.

3.7.2 Choice of the universal weight vectors

It is yet unclear if one should use a n -dimensional reduction or a 1-dimensional reduction for a certain network structure. By answering this question, we also address how to set the weight vectors of the cycle reduction.

Recall that the Perron-Frobenius theorem guarantees that \mathbf{W}^T has a non-negative dominant eigenvector \mathbf{v}_D only if \mathbf{W} is a connected graph. Therefore, we can rule out that it is always possible to construct a 1-dimensional reduction relying on the dominant eigenvector.

But the same reasoning also implies that we can always construct a n -dimensional system using the dominant eigenvector \mathbf{v}_D of \mathbf{W}^T . One could use $\mathbf{a}_1 = \mathbf{v}_D$ and apply iteratively Eq. (3.19) to obtain the set of weight vectors $\{\mathbf{a}_j\}_{1, \dots, n}$, as prescribed. In doing so, the resulting weight vectors would all be identical $\mathbf{a}_1 = \mathbf{a}_2 = \dots = \mathbf{a}_n = \mathbf{v}_D$, as it obviously satisfies both Eq. (3.19) and the periodicity condition $\mathbf{a}_{j+n} = \mathbf{a}_j$. Hence, we find n identical observables $R_1 = R_2 = \dots = R_n$, and the constructed n -dimensional system is no better than the 1-dimensional system. For this reason, a n -dimensional cycle reduction is only advantageous if we can construct a set of *distinct* weight vectors $\mathbf{a}_1 \neq \mathbf{a}_2 \neq \dots \neq \mathbf{a}_n$ from the dominant eigenvectors.

The maximum number of significant and distinct observables that we can construct is determined by the periodicity of the transposed adjacency matrix \mathbf{W}^T . The periodicity n is the number of eigenvalues λ_m of modulus equal to the spectral radius $r \geq 0$, i.e. $|\lambda_m| = r$. From the Perron-Frobenius theorem, they must be uniformly distributed on a circle, centered at the origin, in the complex plane. Thus, the m th dominant eigenvalue can be written as

$$\lambda_m = r e^{2\pi i m/n},$$

for a given periodicity n . Since λ_m is an eigenvalue of \mathbf{W}^T , it must have an eigenvector \mathbf{v}_m that satisfies

$$\mathbf{W}^T \mathbf{v}_m = r e^{2\pi i m/n} \mathbf{v}_m.$$

By multiplying both sides by $(\mathbf{W}^T)^{(n-1)}$, we find

$$(\mathbf{W}^T)^n \mathbf{v}_m = r^n \mathbf{v}_m.$$

Therefore, r^n is a real-value positive n times degenerated eigenvalue of $(\mathbf{W}^T)^n$. Since each eigenvector of \mathbf{W}^T is also eigenvector of $(\mathbf{W}^T)^n$, we can combine those eigenvectors to construct new distinct eigenvectors of $(\mathbf{W}^T)^n$ with eigenvalue r^n and use them as weight vectors. We construct the first weight vector as

$$\mathbf{a}_1 = \frac{\sum_{m=1}^n c_m \mathbf{v}_m}{\sum_{m=1}^n c_m \mathbf{1}^T \mathbf{v}_m}, \quad (3.20)$$

where $c_m \in \mathbb{C}$ are arbitrary coefficients. From Eq. (3.19), we iteratively compute \mathbf{a}_j from \mathbf{a}_{j-1} . By doing so, we both satisfy the periodic condition $\mathbf{a}_{j+n} = \mathbf{a}_j$ and construct distinct weight vectors.

The reduction only requires to arbitrarily choose $\mathbf{c} = (c_1, \dots, c_n)$ to construct \mathbf{a}_1 . We propose to select \mathbf{c} by minimizing the scalar product of the first two weight vectors

$$\mathbf{c} = \underset{\mathbf{c}}{\operatorname{argmin}} |\mathbf{a}_1^T \mathbf{a}_2| \quad (3.21)$$

where $\mathbf{a}_2 = \alpha_1^{-1} \sum_{m=1}^n c_m \lambda_m \mathbf{v}_m$ from Eq. (3.19). In Appendix 3.12.3, we give a general and exact solution of \mathbf{c} for $n = 2$.

In summary, the cycle reduction method goes as follows. First, compute a set of n eigenvectors $\{\mathbf{v}_m\}_{m=1, \dots, n}$ of \mathbf{W}^T whose eigenvalues have a modulus equal to the spectral radius r . Second, obtain $\{c_i\}_{i=1, \dots, n}$ by solving Eq. (3.21). Third, iteratively construct \mathbf{a}_i from Eq. (3.19). Finally, compute α_i, β_i and solve R_i at equilibrium from Eqs. (3.16).

An interesting aspect of this method is that it allows to combine the information of each observable to construct a global observable

$$R_{\text{global}} = \sum_{j=1}^n \phi_j R_j, \quad (3.22)$$

where $\phi_j \in \mathbb{R}$. Since $R_j = \sum_i [\mathbf{a}_i]_j x_j$, the contribution p_j of node j to the global observable is

$$p_j = \sum_i \phi_i [\mathbf{a}_i]_j. \quad (3.23)$$

We can then tune ϕ_i to reach the desired node contributions. For instance, to access the unweighted average activity $R_{\text{global}} = \langle x \rangle$, one solves $\mathbf{A} \boldsymbol{\phi} = N^{-1} \mathbf{1}_N$, where $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_n]$. The solution is $\boldsymbol{\phi} = N^{-1} \mathbf{A}^+ \mathbf{1}_N$, where \mathbf{A}^+ is the Moore-Penrose pseudo-inverse of matrix \mathbf{A} [78].

3.7.3 Examples : Star and bipartite networks

We give an example of the cycle reduction for a highly heterogeneous family of networks : star networks. We construct a star network of N nodes where the strength of a directed edge to a periphery nodes is s_{pc} and s_{cp} for edges directed toward the central node. Hence, the adjacency matrix is

$$\mathbf{W} = \begin{bmatrix} 0 & s_{cp} & s_{cp} & \dots & s_{cp} \\ s_{pc} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{pc} & 0 & 0 & \dots & 0 \end{bmatrix}.$$

One finds that \mathbf{W}^T has two eigenvalues of modulus equal to the spectral radius, $\lambda_+ = \sqrt{s_{pc}s_{cp}(N-1)}$ and $\lambda_- = -\sqrt{s_{pc}s_{cp}(N-1)}$. From the previous analysis, this signals that we can construct a 2-dimensional reduction. The associated eigenvectors are

$$\mathbf{v}_+^T = \begin{bmatrix} \frac{\lambda_+}{s_{cp}} & 1 & 1 & \dots & 1 \end{bmatrix}, \quad (3.24a)$$

$$\mathbf{v}_-^T = \begin{bmatrix} \frac{\lambda_-}{s_{cp}} & 1 & 1 & \dots & 1 \end{bmatrix}. \quad (3.24b)$$

We construct the 2-dimensional reduction by combining the eigenvectors \mathbf{v}_+ , \mathbf{v}_- to minimize $|\mathbf{a}_1^T \mathbf{a}_2|$ and to satisfy the normalization $\mathbf{1}^T \mathbf{a}_1 = 1$. The following linear combinations fulfill the requirements :

$$\mathbf{a}_1 = \frac{\mathbf{v}_+ - \mathbf{v}_-}{\mathbf{1}^T \mathbf{v}_+ - \mathbf{1}^T \mathbf{v}_-}, \quad (3.25)$$

$$\mathbf{a}_2 = \frac{1}{(N-1)s_{cp}} \frac{(\lambda_+ \mathbf{v}_+ - \lambda_- \mathbf{v}_-)}{\mathbf{1}^T \mathbf{v}_+ - \mathbf{1}^T \mathbf{v}_-}, \quad (3.26)$$

where the second vector has been obtained from Eq. (3.19). Note that the overlap $\mathbf{a}_1^T \mathbf{a}_2 = 0$ exactly. Explicitly, in component form, we have

$$[\mathbf{a}_1]_i = \delta_{i,1}, \quad (3.27a)$$

$$[\mathbf{a}_2]_i = \frac{1}{N-1} (1 - \delta_{i,1}). \quad (3.27b)$$

One also finds that $\beta_1 = 1, \beta_2 = 1, \alpha_1 = s_{cp}(N-1), \alpha_2 = s_{pc}$. Hence, the 2-dimensional reduction reads

$$\dot{R}_1 = F(R_1) + s_{cp}(N-1)G(R_1, R_2), \quad (3.28a)$$

$$\dot{R}_2 = F(R_2) + s_{pc}G(R_2, R_1). \quad (3.28b)$$

One notes that R_1 is exactly equal to the activity of the central node and R_2 is the activity of a periphery node. Thus, the 2-dimensional formalism is an exact reduction in this example. This is confirmed in Fig. 3.3 where simulations and predictions are compared for the Gao

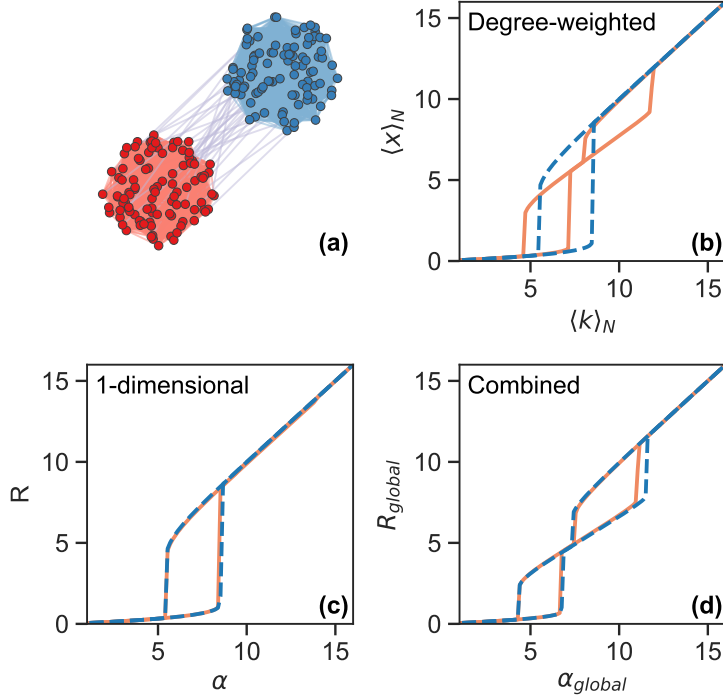


FIGURE 3.4 – (a) Schematisation of the undirected planted partition network of two communities of 100 nodes each with in-densities $p_{in} = 0.4$ and $p_{in} = 0.7$, and out-density $p_{out} = 3 \times 10^{-3}$. (b) Average neighbor activity at equilibrium as a function of its structural parameter using the degree-weighted reduction of Gao *et al.* [52]. (c) Dominant eigenvector-weighted activity $R = \mathbf{v}_D^T \mathbf{x}$ at equilibrium as a function of the associated eigenvalue α for the 1-dimensional reduction [Eq. (3.9)]. (d) Combination of the uncoupled observables $R_{global} = (R_D + R_{SD})/2$ at equilibrium as a function of the structural parameter α_{global} [See Eq. (3.30b)]. Full lines result from simulations and dashed lines are theoretical predictions. The dynamics is the Wilson-Cowan model with $\tau = 1, \mu = 3$ [Eq. 3.2a].

et al., the 1-dimensional reduction and the 2-dimensional reduction for the Wilson-Cowan dynamics.

Star networks are not the only systems conforming to the 2-dimensional reduction; all bipartite networks also do. Bipartite networks have nodes that can be separated into two groups such that connections only exist between nodes of different groups [124]. This network architecture is common in many real systems such as plant-pollinator interactions [20], scientific collaborations [125], and actor-film networks [167].

Bipartite networks exhibit a remarkable and useful spectral property. Each eigenvalue is paired, i.e.

$$\lambda_j = -\lambda_{N-j+1}, \quad (3.29)$$

for all $j = 1, 2, \dots, N$, assuming that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ [39]. Thus, a bipartite graph contains two eigenvalues λ_1, λ_N of modulus equal to the spectral radius. This suggests that a 2-

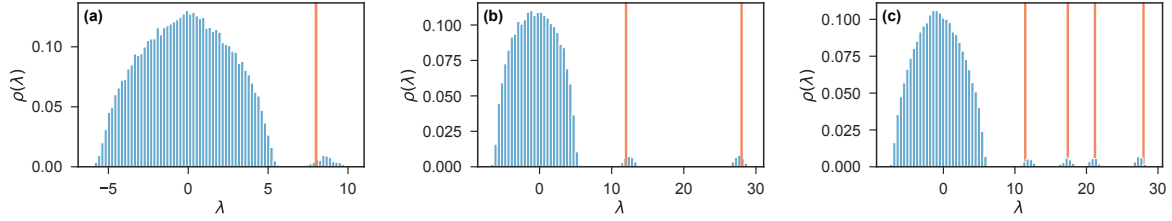


FIGURE 3.5 – Spectral density $\rho(\lambda)$ for (a) random networks of 80 nodes with connection probability $p = 0.1$, (b) of SBM with two communities of 40 nodes with in-connection probabilities $p_{11} = 0.3$ and $p_{22} = 0.7$ and out-connection probabilities $p_{12} = p_{21} = 0.01$, and (c) of SBM with four communities of 40 nodes with in-connection probabilities $p_{11} = 0.4$, $p_{22} = 0.6$, $p_{33} = 0.7$, and $p_{44} = 0.9$ and out-connection probability $p_{rs} = 0.05 \forall r \neq s$. The orange lines highlight the position of the dominant and subdominant eigenvalues in the expected network. Each spectral density is produced by collecting spectra from 500 network instances.

dimensional representation could always be constructed out of these two eigenvalues, following the prescription of Sec. 3.7.1.

We therefore gain a clear understanding of the results of Jiang *et al.*[83], where they present numerical evidences on real bipartite networks, suggesting that 2-dimensional reductions are better predictors of critical points for bipartite networks than the 1-dimensional approaches. Moreover, it solves the problem of selecting the right weighted combination : The eigenvector-weighted combination should always be favored over others.

...

3.8 Multidimensional reduction : including subdominant eigenvectors

Until now, we have developed a direct method to construct n -dimensional reduced systems. Using only the network structure, we can first identify the number of dimensions, i.e. the number of eigenvalues of modulus equal to the spectral radius, and then construct the weight vectors \mathbf{a}_j to predict the observables R_j .

Yet, our method is compelling only if the observables R_j are good indicators of the global states of the network, which requires that each region of the network contributes significantly to at least one observable R_j . Since we use the dominant eigenvectors, we do not control the contribution of each node. Thus, if the dominant eigenvector assigns negligible weights to some nodes, it may result in an incomplete description of the network. Modular networks fall into this category. Let us introduce the stochastic block model (SBM) to understand the underlying problem of misrepresentation.

The SBM is a generative model of modular networks [73]. Nodes are first assigned to modules. Then, we connect a node from module s to a node from module r with probability p_{rs} . This simple method generates accurately modular random networks.

The spectrum of a SBM network is rather different than the spectrum of a random network (Fig. 3.5). We first note that we only have a single eigenvalue of modulus equal to the spectral radius, indicating that we should use a 1-dimensional reduction. However, the eigenvalues are distributed in a multimodal distribution with as many dominant and subdominant eigenvalues as they are modules.

For instance, let us consider a network with two communities of equal size $N/2$. Using the spectral theory of random matrices, we estimate the two dominant eigenvalues

$$\lambda_1 = \frac{(p_{11} + p_{22}) + [(p_{11} - p_{22})^2 + 4p_{12}p_{21}]^{1/2}}{2},$$

$$\lambda_2 = \frac{(p_{11} + p_{22}) - [(p_{11} - p_{22})^2 + 4p_{12}p_{21}]^{1/2}}{2},$$

and their corresponding eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} \frac{p_{12}}{p_{12} + \lambda_1 - p_{11}} \mathbf{1}_{N/2} \\ \frac{\lambda_1 - p_{11}}{p_{12} + \lambda_1 - p_{11}} \mathbf{1}_{N/2} \end{bmatrix},$$

$$\mathbf{v}_2 = \begin{bmatrix} \frac{\lambda_2 - p_{22}}{\lambda_2 - p_{22} + p_{21}} \mathbf{1}_{N/2}, \\ \frac{p_{21}}{\lambda_2 - p_{22} + p_{21}} \mathbf{1}_{N/2} \end{bmatrix}.$$

If $p_{rs}/p_{rr} \approx 0 \quad \forall r \neq s$, the dominant eigenvector \mathbf{v}_1 assigns a negligible weight to the nodes in the community $r = 1$ ³. Thus, if we solely use a 1-dimensional reduction with $\mathbf{a} = \mathbf{v}_1$, the observable R will not take into account the activity of half the network. Fortunately, the second-dominant eigenvector accounts for the remaining nodes. Thus, we must apply the 1-dimensional reduction of Sec. 3.6.1 twice and construct two uncoupled observables, one with the dominant eigenvector $R_D = \mathbf{v}_1^T \mathbf{x}$, and one for the subdominant eigenvector $R_{SD} = \mathbf{v}_2^T \mathbf{x}$, for which the dynamics follow,

$$\dot{R}_D = F(R_D) + \alpha_D G(\beta_D R_D, R_D), \quad (3.30a)$$

$$\dot{R}_{SD} = F(R_{SD}) + \alpha_{SD} G(\beta_{SD} R_{SD}, R_{SD}). \quad (3.30b)$$

3. Note that if $(p_{11} - p_{22})^2 \ll 4p_{12}p_{21}$, then the problem of misrepresentation is absent and it reduces to a single 1-dimensional representation.

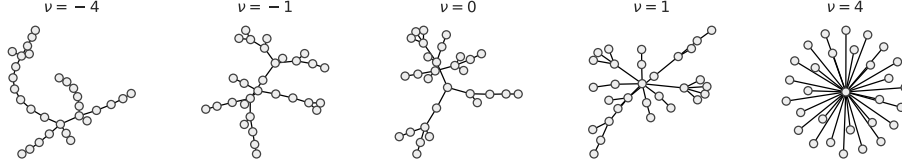


FIGURE 3.6 – Network instances produced using the generalized preferential attachment model. The model tends to generate chain-like networks for negative values of ν and star networks for positive values. Figure inspired from [176].

To make a global prediction, we simply combine the observables

$$R_{\text{global}} = \frac{R_D + R_{SD}}{2}. \quad (3.31)$$

It follows that the global structural parameter is also a linear composition

$$\alpha_{\text{global}} = \frac{\alpha_D + \alpha_{SD}}{2}. \quad (3.32)$$

In general, we can construct as many uncoupled observables as the number of modules in the network, by using the eigenvectors associated with the eigenvalues detached from the bulk of the spectrum.

We show a numerical example of this method for a network of two communities (Fig 3.4). The Gao *et al.* formalism predicts a single bifurcation, which almost coincides with the transition of the densest community. In Fig 3.4(b), the 1–dimensional reduction, using the first dominant eigenvector, predicts the activity accurately. However, the first dominant eigenvector omits half the network and we only see a single bifurcation. Thus, even if we are highly accurate, we miss characterizing the distinctive multistep bifurcation, a prominent feature of interacting networks [106]. Finally, combining observables as in Eq. (3.31), we recover the bifurcations of both modules [Fig 3.4(d)].

...

3.9 Goodness of reduction

The goodness of the reduction method, i.e. how accurate are the predictions of the low-dimensional representation compared with the observations on the original network, depends on the nature of the dynamics and on the network structure. For instance, some complex patterns of interactions may be less amenable to a low-dimensional formalism, resulting in disparities between the predicted and exact values of the observables. In this section, we explore the impacts of the structure and the dynamics on the goodness of the reduction methods.

3.9.1 Impact of the structure

To measure the impact of the structure, we introduce a generative model of networks called *generalized preferential attachment model* [93]. Parameters of this model can be continuously tuned to obtain networks ranging from chain-like networks to star networks, with scale-free networks as an intermediate state. Scale-free systems are an important family of networks, recognizable by their power-law degree distribution $p(k) \sim k^{-\gamma}$ [14]. Due to their lack of well-defined characteristic scale, it is a priori unclear whether these systems can be efficiently reduced.

The growth process of the generalized preferential attachment goes as follows. We initialize the network with two connected nodes. Then, at each time step t , we add a new node to the network. It is connected to an existing node chosen with probability

$$w_i(\nu, t) = \frac{s_i^\nu(t)}{\sum_{j=1}^{N(t)} s_j^\nu(t)} \quad (3.33)$$

where $\nu \in \mathbb{R}$ is the exponent of the attachment kernel, $s_j(t)$ is the number of connections of node j at time t , and $N(t)$ is the number of nodes at time t .

The generative model is solely tuned by the kernel parameter $\nu \in \mathbb{R}$. It controls the inequalities of the attachment probability. On the one hand, if $\nu \gg 1$, the generated networks are star-like as we always attach new nodes to the richest node. On the other hand, if $\nu \ll 0$, the networks are more chain-like as we always connect to the least connected node [93]. The classic preferential attachment model is found for $\nu = 1$. Therefore, for $0 < \nu < 1$, we observe a continuum of network organizations which gradually become more scale-free as the parameter ν is increased. Examples of networks generated from this model are illustrated in Fig. 3.6.

From now on, we will distinguish the predicted observable from the reduced system, denoted $\tilde{R}(\alpha)$, and the measured observable $R(\alpha) = \mathbf{a}^T \mathbf{x}$ from the original network.

We have applied the degree-weighted, the 1-dimensional and the 2-cycle reductions to networks generated with the generalized preferential attachment model for $\nu \in [-1, 2]$. For each network, we have computed the total error Δ_R between the measured activity $R^*(\alpha) = \mathbf{a}^T \mathbf{x}^*$ at equilibrium on the original network and the predicted activity $\tilde{R}^*(\alpha)$ by the reduction system,

$$\Delta_R = \int_0^\infty |\tilde{R}^*(\alpha) - R^*(\alpha)| d\alpha. \quad (3.34)$$

We have found a transition in the dimension reduction accuracy for all methods at $\nu = 1$, corresponding to the preferential attachment model [Fig 3.7]. As we enter the star-like region $\nu > 1$, the average error reaches a plateau to specific values for the 1-dimensional reductions while the 2-dimension reduction remains highly effective.

We argue that this transition is not dynamics-specific but mostly due to the network architecture. The nature of the error can be interpreted by a careful examination of the generative model. First, negative values of ν tend to homogenize the degree of the nodes. Thus, the more uniform the network is, the easier it is to capture its behavior in a 1–dimension reduction. For positive values of ν , the reduced model tends to favor degree inequalities which are best achieved when the networks are star-like.

For $0 < \nu < 1$, the degree distribution resembles a power law with exponential cutoffs [93]. However, a pure power-law distribution is only achieved precisely at $\nu = 1$. Thus, this transition in the degree distribution forces the reduction of the degree-weighted reduction to predict inaccurate observables. Finally, the region $\nu > 1$ is dominated by star-like networks, which has been previously shown to be better represented by the 2–dimensional reduction than any 1–dimensional reduction (See Subsec. 3.7.3).

3.9.2 Detection of transitions

Most dynamical systems exhibit activity bifurcations when a certain structural threshold is reached [168]. Therefore, the goodness of the reduction should display at least qualitative changes as the structural threshold is crossed. We investigate this kind of prediction using the SIS model.

The SIS model has been thoroughly studied in the last decade [135]. In the SIS model, nodes reversibly switch from susceptible to infected states with a certain probability that depends on their neighborhood. We can formulate this dynamics using a mean-field approach,

$$\dot{x}_i = -x_i + \gamma(1 - x_i) \sum_{j=1}^N w_{ij} x_j, \quad (3.35)$$

where x_i is the probability that node i is infected and $\gamma \geq 0$ is the normalized infection rate. In this model, the average fraction of infected node $\langle x \rangle = N^{-1} \sum_i x_i$ undergoes a critical transition at a certain threshold γ_C . The classical problem in the study of the SIS model consists in estimating the value of γ_C above which a significant fraction of the whole system is infected [158].

We will however study a related problem : the parameter γ is fixed and the structure is evolving. Using the dimension reduction procedure, we investigate the critical structural parameter α_{global} , or an equivalent parameter depending on the reduction approach, for which the global state at equilibrium R^* undergoes a critical transition characterized by

$$\frac{d^2 R^*(\alpha_C)}{d\alpha^2} = 0. \quad (3.36)$$

In Fig. 3.8, we investigate the errors on the position of the critical transition for the degree weighted 1–dimension approach, the eigenvector-weighted 1–dimensional approach, and the 2–

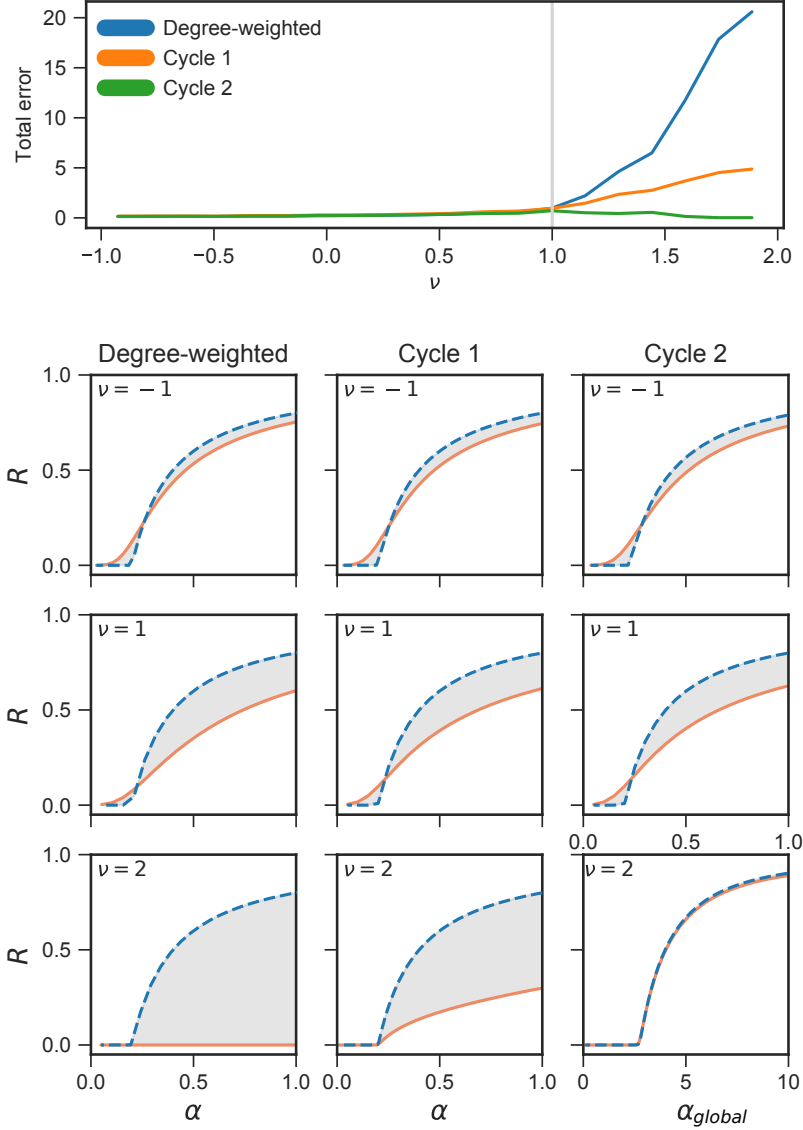


FIGURE 3.7 – (Top) Total error on the prediction of the global state activity at equilibrium for networks obtained from the generalized preferential attachment model [See Eq. (3.34)]. A gray line indicates the classical preferential attachment model at $\nu = 1$ where a critical transition in the reduced descriptions is found. The total error is averaged from 300 networks of $N = 200$ generated uniformly on the domain $\nu \in [-1, 2]$. The activity on the network is the SIS model with $\gamma = 1$ [See Eq. (3.2e)]. (Bottom) Instances of bifurcation diagrams for the three reduction schemes (columns) and for $\nu = \{-1, 1, 2\}$ (rows). For the 2-dimensional cycle reduction, the x-axis is the average of the structural parameters : $\alpha_{global} = (\alpha_1 + \alpha_2)/2$. The blue dashed lines are predictions from reduced systems and the orange lines are the measured activities on the original networks. The gray regions indicate the absolute errors [See Eq. (3.34)].

dimensional *cycle* reduction. We use a network of $N = 60$ nodes generated from the generalized preferential attachment model with $\nu = 1.8$. We observe that the two proposed approaches based on dominant eigenvectors are able to accurately predict the critical transition while the degree-weighted approach does not. This behavior is typical for reductions of networks

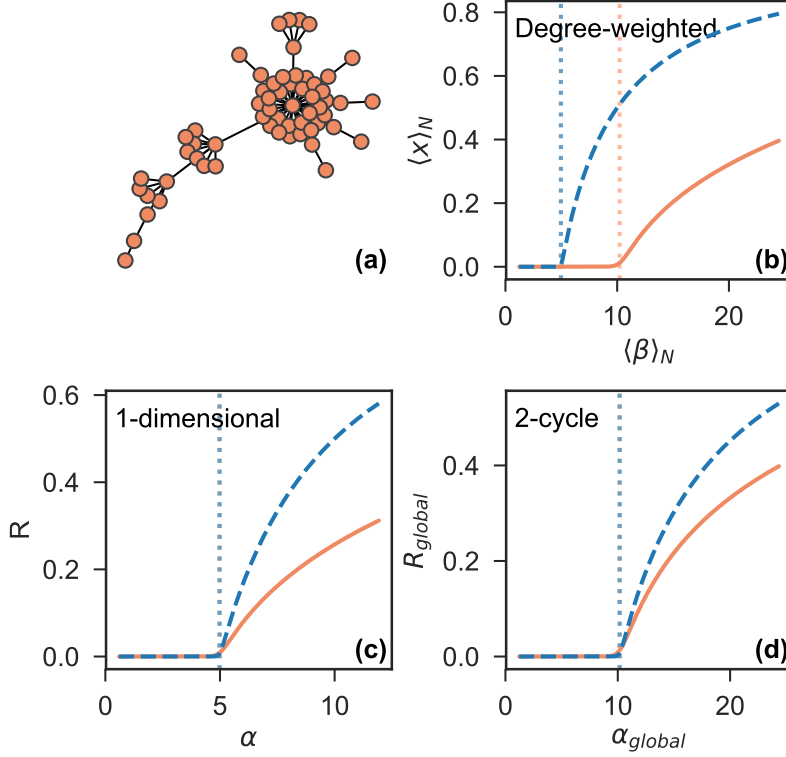


FIGURE 3.8 – (a) Schematisation of the undirected network generated using the generalized preferential attachment model with $\nu = 1.8$ and $N = 60$. (b) Average neighbor activity at equilibrium as a function of its structural parameter using the degree-weighted reduction [52]. (c) Dominant eigenvector-weighted activity at equilibrium $R^* = \mathbf{v}_D^T \mathbf{x}^*$ as a function of the associated eigenvalue α for the 1-dimensional reduction [Eq. (3.5)]. (d) Solution at equilibrium of the 2-dimensional system $R_{global}^* = (R_1^* + R_2^*)/2$ as a function of the structural parameter $\alpha_{global} = (\alpha_1 + \alpha_2)/2$. Full lines are results from simulations and dashed lines are theoretical predictions. Dotted lines indicate the position of the transition α_{global} . The dynamics is the SIS model with $\gamma = 0.2$ [Eq. (3.2e)].

generated from $\nu \in [1, \infty)$. Our results indicate that even if the 1-dimensional observable fails to predict the true level of activity R , it still predicts with high accuracy the onset of the epidemic. We conclude that the largest eigenvalue is a reliable indicator of the onset for correlated networks. Perhaps, this conclusion is not surprising as it has been previously discovered under a different approach [25]. Nonetheless, it supports the proposed reductions as valuable candidates for predicting the onset of critical transitions.

3.9.3 Impact of the dynamics

As previously discussed, the goodness of the reduction is highly dependent on the network structure. The other element that impacts the goodness of the reduction is the nature of the dynamics. For instance, linear dynamics such as $F(x_i) = x_i$, $G(x_i, x_j) = x_j$ lead to exact 1-dimensional reduction. However, typical dynamics are nonlinear and may add significant

contributions to the quadratic terms that have been neglected (See Appendix 3.12.1). We investigate this aspect by looking at two contrasting dynamics : Wilson-Cowan model and Lotka-Volterra model.

First, let us introduce the relative error Δ_α on the structural parameter that predicts the original network activity. We compare the measured structural parameter $\alpha = \mathbf{a}^T \mathbf{k}^{\text{in}}$ on the original network with the structural parameter $\tilde{\alpha}$ that matches, from the 1-dimensional reduction, the measured activity $R^* = \mathbf{a}^T \mathbf{x}^*$ on the original network. The relative error can be written as

$$\Delta_\alpha = \frac{\alpha - \tilde{\alpha}(R^*)}{\alpha}.$$

Notice that this is different from Eq. (3.34) : Δ_α is the horizontal error in the space (α, R) while Δ_R measures the vertical error. Both errors convey distinct information and are complementary. However, we will use Δ_α since it can be written as a function that depends explicitly on the nature of the dynamics for the 1-dimensional reduction.

From Eq. (3.5), we obtain $\tilde{\alpha}$ at the dynamical equilibrium $\dot{R} = 0$. This leads to

$$\Delta_\alpha = 1 + \frac{F(R^*)}{G(R^*)\alpha}. \quad (3.37)$$

One can then evaluate the error for specific dynamics as in [160]. In the following paragraphs, we give two examples of dynamics and compare the errors for the Gao *et al.* formalism and our 1-dimensional reduction.

Error on the Wilson-Cowan dynamics

The Wilson-Cowan dynamics describes the firing-rate activity of populations of neurons. The evolution of a node activity is given by Eq. (3.2a), and repeated here as

$$\dot{x}_i = -x_i + \sum_{j=1}^N w_{ij} \frac{1}{1 + \exp[-\tau(x_j - \mu)]}, \quad (3.38)$$

where $\tau > 0$. The equilibrium solution \mathbf{x}^* cannot be found analytically, so it must be evaluated numerically, even for $N = 1$. However, \mathbf{x}^* is well approximated for extreme values of activities. We derive error estimates for extreme regimes $x_j^* \gg \mu$ and $x_j^* \ll \mu$.

In general, from Eq. (3.37), the error can be written as

$$\Delta_\alpha = 1 - \frac{[1 + \exp[-\tau(R^* - \mu)]]R^*}{\alpha}. \quad (3.39)$$

In the limit of high levels of activity $x_j^* \gg \mu$, the exponential vanishes and one finds that $\mathbf{x}^* \approx \mathbf{W}\mathbf{1} = \mathbf{k}^{\text{in}}$. Therefore, the error is approximately

$$\Delta_\alpha \approx 1 - \frac{R^*}{\alpha} \approx 1 - \frac{\mathbf{a}^T \mathbf{W}\mathbf{1}}{\alpha}.$$

For the 1–dimensional reduction, \mathbf{a} is an eigenvector of \mathbf{W}^T such that $\mathbf{a}^T \mathbf{W} = \alpha \mathbf{a}^T$. Since \mathbf{a} is normalized, $\mathbf{a}^T \mathbf{1} = 1$, it follows that $\mathbf{a}^T \mathbf{W} \mathbf{1} = \alpha$. Thus, the error Δ_α vanishes in the limit of large activity.

The same applies for Gao *et al.* formalism for high activity. By using $\mathbf{a} = \mathbf{k}^{\text{in}}$, one finds that $\alpha = (\mathbf{k}^{\text{in}})^T \mathbf{k}^{\text{out}}$ and $R^* = (\mathbf{k}^{\text{out}})^T \mathbf{k}^{\text{in}}$ so that $\Delta_\alpha \rightarrow 0$.

Using a similar procedure for $x_j \ll \mu$, one can also show that $\Delta_\alpha \approx 0$ for both methods.

We conclude that in the extreme regimes of high and low activities, both reduction methods provide a practically exact solution. However, we are more often interested in the hysteresis region where the activity collapses rapidly. Unfortunately, analytic error estimates are lacking in this regime. Still, numerical results and theoretical insights suggest that the proposed 1–dimensional reduction should always be favored over the degree-weighted reduction. This is confirmed below for a more tractable dynamics.

Error on the Lotka-Volterra dynamics

Let us consider the Lotka-Volterra dynamics governing the evolution of species populations. The N –dimensional system goes as

$$\dot{\mathbf{x}} = \omega \mathbf{x} + \mathbf{x} \circ \mathbf{W} \mathbf{x}, \quad (3.40)$$

where \circ denotes an elementwise multiplication. At equilibrium, \mathbf{x}^* satisfies

$$-\omega \mathbf{1} = \mathbf{W} \mathbf{x}^*. \quad (3.41)$$

With Eq. (3.37), we write the expected error as

$$\Delta_\alpha = 1 + \frac{\omega}{\beta \alpha R^*} = 1 + \frac{\omega}{\beta \alpha \mathbf{a}^T \mathbf{x}^*}.$$

Using the 1–dimensional reduction,

$$\mathbf{W}^T \mathbf{a} = \alpha \mathbf{a},$$

or $\mathbf{a}^T \mathbf{W} = \alpha \mathbf{a}^T$. Furthermore,

$$\mathbf{a}^T \mathbf{W} \mathbf{x}^* = -\omega \mathbf{a}^T \mathbf{1} = \alpha \mathbf{a}^T \mathbf{x}^*. \quad (3.42)$$

Thus, the error depends only on β :

$$\Delta_\alpha = 1 - \frac{1}{\beta}. \quad (3.43)$$

Therefore, the difference between the exact value for $\mathbf{a}^T \mathbf{x}^*$ and the approximate value derived from the 1–dimensional reduced system is only $1 - 1/\beta$. Given the expression for β [Eq. (3.48)], it should be close to $\beta \approx 1$, so the error goes to zero $\Delta_\alpha \approx 0$. This contrasts with results derived by Tu *et al.* [160] for the method of Gao *et al.*, where they reported non-vanishing error averages and variances.

...

3.10 Conclusion

We have built systematic methods of dimension reduction adapted to different families of networks (random, star-like, bipartite, SBM). The activity of the reduced systems is used as an indicator of the global activity of large networks. Without further restriction than imposing a linear form of the global activity, we have found that the dominant eigenvectors of the adjacency matrix are central to the global states' evolution. Moreover, when considering the *cycle* reduction, the dimension of the reduced systems corresponds to the periodicity of the adjacency matrix.

We have further shown that the proposed reduction of Gao *et al.* is a special case of the general scheme when applied to uncorrelated random networks. Moreover, the range of applicability of our method extends to modular, heterogeneous and bipartite networks.

Our results suggest, both numerically and theoretically, that the eigenvector-weighted reduction should be preferred over the degree-weighted reduction. Originally, the degree-weighted reduction has been used to approximate the state of a network by the state of the average neighbor. But, the degree of a node is only a local centrality measure since it does not provide information to whom a node is connected with. In contrast, in the eigenvector-weighted reduction, the dominant eigenvector yields a more global node centrality since it contains the information on how each node is connected with the rest of the network [124]. Therefore, the eigenvector-based reduction brings a new light on the influence of each node on the global states of a network.

On a more practical side, our general method is able to predict the correct number of bifurcation points. The expected number of predicted bifurcation points depends on the dimension of the reduction. Intuitively, and confirmed by our investigations, a single linear observable provides a good reduction if the network is homogeneous or if, for instance, the degree variance is small. When this is not the case, however, different parts of the network behave differently and a single observable is no longer sufficient to capture the characteristics of the global dynamics, which effectively becomes multidimensional. In the SBM case, the 2-dimensional reduction reveals additional bifurcation points that are missed altogether by all 1-dimensional reductions.

As a closing remark, although our reduction method has been designed to access large dynamical networks through low-dimensional formalisms, it was not clear from the outset how the dimensional reductions would fare with respect to the size of the networks. Our findings on the matter have been comforting since size by itself has a secondary effect on the quality

of the reduction procedure, leaving precedence to connectivity and dynamics. Hence, beyond the addition to the theoretical arsenal, our systematic and versatile approach can now be used to address concrete problems of real-world systems. To name a few, it could be used to describe with high accuracy the bifurcation patterns, to identify dynamical vulnerabilities, to suggest intervention strategies to prevent dynamical breakdowns, or to classify networks on a standardized diagram.

...

3.11 Acknowledgements

We are thankful to Charles Murphy, Guillaume St-Onge, Vincent Thibeault, and Jean-Gabriel Young for useful comments and suggestions. This work was funded by the Fonds de recherche du Québec-Nature et technologies (EL, PD), the Natural Sciences and Engineering Research Council of Canada (LJD), and Sentinel North, financed by the Canada First Research Excellence Fund (EL, PD, LJD, ND).

...

3.12 Supplementary Material

3.12.1 Derivation of 1-dimensional formalism

In this section, we detail the analytical derivation of the evolution of R for the 1-dimensional reduction. We consider the observable

$$R = \sum_{i=1}^N a_i x_i$$

with $a_i \in \mathbb{R}$ and $\sum_i a_i = \mathbf{1}^T \mathbf{a} = 1$. We first take the time derivative and insert Eq. (3.1), which leads to

$$\begin{aligned} \dot{R} &= \sum_{i=1}^N a_i \dot{x}_i \\ &= \sum_{i=1}^N a_i \left[F(x_i) + \sum_{j=1}^N w_{ij} G(x_i, x_j) \right]. \end{aligned}$$

We wish to show that if \mathbf{a} is chosen correctly, then the right-hand side can be written, up to second order corrections, in terms of R only. To do so, we develop each function around the observable :

$$F(x_i) = F(R) + (x_i - R)F'(R) + \mathcal{O} [(x_i - R)^2]. \quad (3.44)$$

Thus,

$$\begin{aligned} \sum_{i=1}^N a_i F(x_i) &= F(R) + F'(R) \sum_{i=1}^N a_i (x_i - R) + \mathcal{O} [(x_i - R)^2] \\ &= F(R) + \mathcal{O} [(x_i - R)^2]. \end{aligned}$$

It means that $F(x_i)$ does not impose any constraint on \mathbf{a} . Now, for the function $G(x_i, x_j)$, we develop around $x_i = \beta R$ and $x_j = \gamma R$:

$$\begin{aligned} G(x_i, x_j) &\approx G(\beta R, \gamma R) + (x_i - \beta R)G_1(\beta R, \gamma R) \\ &\quad + (x_j - \gamma R)G_2(\beta R, \gamma R) \end{aligned}$$

where second order terms have been neglected. Letting $\alpha = \sum_{i,j} a_i w_{ij}$, we find that $\sum_{i,j} a_i w_{ij} G(x_i, x_j)$ is given by

$$\begin{aligned} \sum_{i,j} a_i w_{ij} G(x_i, x_j) &\approx \alpha G(\beta R, \gamma R) \\ &+ G_1(\beta R, \gamma R) \sum_{i,j} a_i w_{ij} (x_i - \beta R) \\ &+ G_2(\beta R, \gamma R) \sum_{i,j} a_i w_{ij} (x_j - \gamma R). \end{aligned}$$

The left-hand side is a function of R only if the linear terms cancel out exactly, which is possible if and only if

$$\alpha \beta R = \sum_{ij} a_i w_{ij} x_i = \mathbf{x}^T \mathbf{K} \mathbf{a}, \quad (3.45a)$$

$$\alpha \gamma R = \sum_{ij} a_i w_{ij} x_j = \mathbf{x}^T \mathbf{W}^T \mathbf{a}. \quad (3.45b)$$

Since $R = \mathbf{x}^T \mathbf{a}$, we conclude that the last two equations are satisfied for all $\mathbf{x} \in \mathbb{R}^N$ only if \mathbf{a} is an eigenvector of both matrices \mathbf{K} and \mathbf{W}^T , with corresponding eigenvalues $\beta \alpha$ and $\gamma \alpha$. Although we cannot solve these two equations simultaneously in general, we can enforce that at least one equation is satisfied exactly. Choosing \mathbf{a} as an eigenvector of \mathbf{W}^T , we can prove that $\gamma = 1$ if the vector \mathbf{a} is normalized $\mathbf{1}^T \mathbf{a} = 1$: If $\mathbf{W}^T \mathbf{a} = \lambda \mathbf{a}$ and $\lambda = \alpha \gamma$, then $\lambda = \mathbf{1}^T \mathbf{W}^T \mathbf{a} = \mathbf{a}^T \mathbf{W} \mathbf{1} = \mathbf{a}^T \mathbf{k}^{\text{in}} = \alpha$, so $\gamma = 1$.

We then choose β to best satisfy Eq. (3.45a) by minimizing the *mean square error* (MSE) :

$$\beta^* = \underset{\beta}{\operatorname{argmin}} \|\mathbf{K} \mathbf{a} - \beta \alpha \mathbf{a}\|^2, \quad (3.46)$$

where the symbol $\|\cdot\|$ denotes the standard euclidean norm. Basic calculus leads to

$$\beta^* = \frac{1}{\alpha} \frac{\mathbf{a}^T \mathbf{K} \mathbf{a}}{\mathbf{a}^T \mathbf{a}} = \frac{1}{\alpha} \frac{\sum_i a_i^2 k_i^{\text{in}}}{\sum_i a_i^2}. \quad (3.47)$$

Note that β^* is a ratio of weighted averages

$$\beta^* = \frac{\mathbf{b}^T \mathbf{k}^{\text{in}}}{\mathbf{a}^T \mathbf{k}^{\text{in}}}, \quad (3.48)$$

where \mathbf{b} is normalized $\mathbf{1}^T \mathbf{b} = 1$ and has for elements $b_i = a_i^2 / \sum_{i=1}^N a_i^2$. From the construction of \mathbf{b} , we deduce that \mathbf{b} must be similar to \mathbf{a} and β^* close to 1, which has been confirmed throughout most of the simulations.

3.12.2 Derivation of the multidimensional *cycle* formalism

For the *cycle* reduction, we construct n observables

$$R_k = \sum_{i=1}^N [\mathbf{a}_k]_i x_i \quad (3.49)$$

with normalized weights $\mathbf{1}^T \mathbf{a}_j = 1$. Using Eq. (3.1), we find that the dynamics of R_k is equal to

$$\dot{R}_k = \sum_i [\mathbf{a}_k]_i F(x_i) + \sum_{ij} [\mathbf{a}_k]_i w_{ij} G(x_i, x_j). \quad (3.50)$$

As for the 1-dimensional reduction, one finds that

$$\sum_i [\mathbf{a}_k]_i F(x_i) \approx F(R_k) \quad (3.51)$$

up to the second order of corrections. We then develop $G(x_i, x_j)$ around $x_i = \beta_k R_k$ and $x_j = \gamma_k R_{k+1}$, which yields

$$\begin{aligned} G(x_i, x_j) &\approx G(\beta_k R_k, \gamma_k R_{k+1}) \\ &\quad + (x_i - \beta_k R_k) G_1(\beta_k R_k, \gamma_k R_{k+1}) \\ &\quad + (x_j - \gamma_k R_{k+1}) G_2(\beta_k R_k, \gamma_k R_{k+1}). \end{aligned}$$

Using the same arguments as in the 1-dimensional reduction, one can prove that $\sum_{i,j} [\mathbf{a}_k]_i w_{ij} G(x_i, x_j) \approx \alpha_k G(\beta_k R_k, \gamma_k R_{k+1})$, with $\alpha_k = \mathbf{a}^T \mathbf{k}^{\text{in}}$, only if the following equations are satisfied simultaneously

$$\mathbf{x}^T \mathbf{K} \mathbf{a}_k = \beta_k R_k \alpha_k, \quad (3.52)$$

$$\mathbf{x}^T \mathbf{W}^T \mathbf{a}_k = \gamma_k R_{k+1} \alpha_k. \quad (3.53)$$

The second equation is satisfied if :

$$\mathbf{W}^T \mathbf{a}_k = \alpha_k \mathbf{a}_{k+1}, \quad (3.54)$$

with $\gamma_k = 1$. After n applications of Eq. (3.54), we close the system with $a_{n+1} = a_1$, which is the respected if \mathbf{a}_1 is eigenvector of $(\mathbf{W}^T)^n$.

As for the parameter β_k , we minimize the MSE

$$\beta_k^* = \underset{\beta_k}{\text{argmin}} \|\mathbf{K} \mathbf{a}_k - \beta_k \alpha_k \mathbf{a}_k\|^2 \quad (3.55)$$

and find

$$\beta_k = \frac{1}{\alpha_k} \frac{\mathbf{a}_k^T \mathbf{K} \mathbf{a}_k}{\mathbf{a}_k^T \mathbf{a}_k}. \quad (3.56)$$

3.12.3 Combination for 2-cycle reduction

In this Appendix, we show that, for an adjacency matrix \mathbf{W}^T , non-negative and of periodicity 2, the first weight vector \mathbf{a}_1 of the reduction is an equipartition of the two dominant eigenvectors.

For such an adjacency matrix, the Perron-Frobenius theorem states that \mathbf{W}^T admits two eigenvalues of modulus equal to the spectral radius. The eigenvectors satisfy

$$\mathbf{W}^T \mathbf{v}_1 = r \mathbf{v}_1 \quad ; \quad \mathbf{W}^T \mathbf{v}_2 = -r \mathbf{v}_2, \quad (3.57)$$

where r is the spectral radius and \mathbf{v}_i are the eigenvectors normalized as $\mathbf{v}_i^T \mathbf{v}_i = 1$.

Now, let us consider the first weight vector as a linear combination of the dominant eigenvectors :

$$\mathbf{a}_1 = \frac{c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2}{c_1 \mathbf{1}^T \mathbf{v}_1 + c_2 \mathbf{1}^T \mathbf{v}_2}. \quad (3.58)$$

From transformation (3.19), we get the equation

$$\mathbf{a}_2 = \frac{1}{\alpha_1} \frac{c_1 r \mathbf{v}_1 - c_2 r \mathbf{v}_2}{c_1 \mathbf{1}^T \mathbf{v}_1 + c_2 \mathbf{1}^T \mathbf{v}_2}. \quad (3.59)$$

We want to find c_1, c_2 such that $c_1 + c_2 = 1$ and $S = |\mathbf{a}_1^T \mathbf{a}_2|$ is minimized. The former condition is chosen for definiteness, while the latter condition favors the weight vectors that represent almost exclusive groups of nodes. The scalar product is then simply :

$$S \propto |c_1^2 - c_2^2|.$$

and minimized with $c_1 = c_2 = 1/2$.

Chapitre 4

Détection des perturbations structurelles par apprentissage automatique

Detecting structural perturbations from time series using deep learning

Edward Laurence^{1,2}, Charles Murphy^{1,2}, Guillaume St-Onge^{1,2}, Xavier Roy-Pomerleau^{1,2} et Vincent Thibeault^{1,2},

¹ Département de Physique, de Génie Physique, et d'Optique
Université Laval, Québec, Québec, Canada G1V 0A6

² Centre Interdisciplinaire en Modélisation Mathématique de l'Université Laval (CIMMUL)
Université Laval, Québec, Québec, Canada G1V 0A6

...

4.1 Avant-propos

Un des enjeux majeurs que pose la présence d’agents perturbants dans un système complexe est d’être en mesure de détecter rapidement les perturbations et d’intervenir dans un délai raisonnable. Les méthodes d’*early-warning signals* de la Section 1.4.3 comblent en partie ce travail fastidieux de détection, sans toutefois proposer une solution universelle et très précise.

...

4.2 Résumé

Dans ce chapitre, nous introduirons une méthode d’apprentissage automatique pour la détection de perturbations dans le contexte de réseaux complexes dynamiques. L’approche ne requiert aucune connaissance quant à la nature de la dynamique ou de la perturbation. Nous validerons la technique sur des dynamiques de propagation (SIS), de populations animales dans un écosystème et neuronales se déroulant sur des réseaux aléatoires, invariants d’échelle et réels. Les performances indiquent que cette approche surpasse les méthodes de reconstruction fonctionnelle et possède la qualité des prédictions d’une approche bayésienne. Cette méthode offre une alternative pratique aux méthodes de *early-warning signals* et comble un aspect important de l’étude de la résilience des systèmes complexes.

...

4.3 Abstract

Small disturbances can trigger functional breakdowns in complex systems. A challenging task is to infer the structural cause of a disturbance in a networked system, soon enough to prevent a catastrophe. We present a graph neural network approach, borrowed from the deep learning paradigm, to infer structural perturbations from functional time series. We show our data-driven approach outperforms typical reconstruction methods while meeting the accuracy of Bayesian inference. We validate the versatility and performance of our approach with epidemic spreading, population dynamics, and neural dynamics, on various network structures : random networks, scale-free networks, 25 real food-web systems, and the *C. Elegans* connectome. Moreover, we report that our approach is robust to data corruption. This work uncovers a practical avenue to study the resilience of real-world complex systems.

4.4 Introduction

Complex systems can shift abruptly and irreversibly into pathological states following a disturbance. Mass extinction, stock market crash, lake eutrophication are concrete examples of such catastrophes [71, 113, 148]. For systems with a clear underlying network structure, stresses may take the form of structural defects such as removing nodes or edges. These perturbations bring the system closer to a tipping-point until an obvious dynamical shift is observed [145]. To prevent breakdowns to occur, early detection is essential but is notoriously difficult [99]. Indeed, the effect of the disturbance can be negligible and comparable to the system's noise [146]. Even more difficult than early detection of catastrophes is the challenge of *inferring* the removed edges/nodes.

Given observations of a network dynamics, the issue of identifying structural defects has remained largely unexplored. Most closely related are numerous studies that tackle the problem of detecting early-warning signals of functional transitions [37, 42, 146, 173]. Mostly grouped under the critical slowing down paradigm, these methods are far from universal as many have reported erroneous detections or failed to forecast an eventual catastrophe [22, 81, 89, 170]. Moreover, few early-warning signals take into account the structure of interactions, even though data may be available, so the structural cause of the disturbance cannot be inferred. Other approaches coming from the study of dynamical complex networks investigate the functional effect of removing edges [52, 83, 96]. However, they are restricted to specific dynamical models which are generally inadequate to describe the rich behavior of real complex systems.

Reconstruction methods have been frequently used, especially in neuroscience, to infer a functional structure from complex time series [17, 27, 34, 143, 150, 152]. Most of these methods shine by their simplicity as they usually do not require any parameter fitting and can readily be applied to almost any type of dynamics. While being originally designed to reconstruct whole networks, they nevertheless seem a natural fit to identify structural defects. Yet, the precision of the reconstructed structure is often limited when the methods are applied on real-world datasets. They tend to detect a large variety of functional (dense) relationships between the nodes, which are only indirectly related to the (sparse) structure [50, 134]. Other reconstruction approaches rely on assumptions of dynamical [138, 151] or structural models [123, 133, 177]. Their main assets are their high accuracy on toy models with ground-truth data and their quantification of the confidence interval of the reconstructed structure. However, these approaches are limited to specific instances where the model assumptions are reasonable [31], which limit their scope for real-world applications.

In this paper, we directly tackle the issue of inferring structural perturbations by introducing

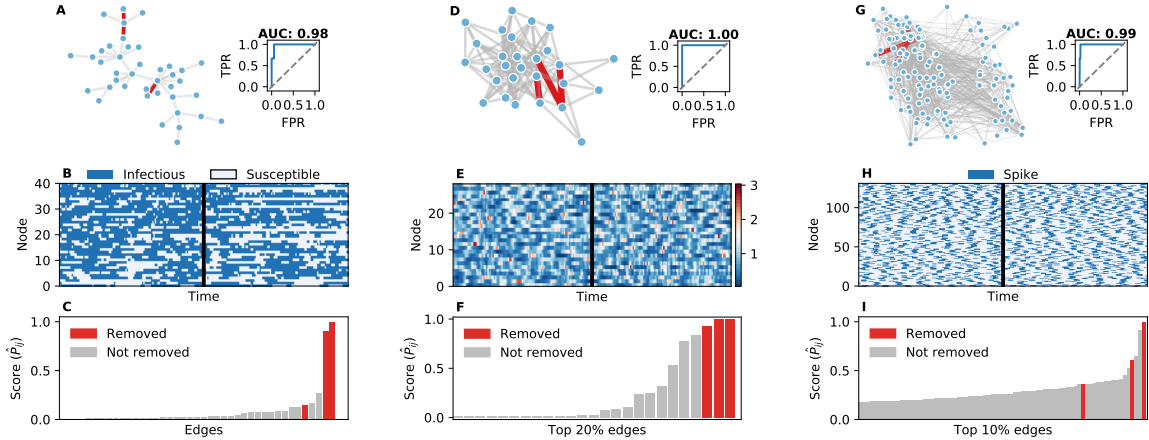


FIGURE 4.1 – Examples of GNN predictions and time series for epidemic spreading on a scale-free network (A-B-C), population dynamics on a real ecological network [9] (D-E-F), and neural dynamics on a C.Elegans connectome (G-H-I). Halfway through the time series (black lines in B-E-H), three random edges, shown by the thick red edges in (A-D-G), are removed from the original networks displayed in (A-D-G). The bars of (C-F-I) show the predicted scores by the GNN model on individual edges with true removed edges colored in red. The ROC curves of the GNN predictions are shown with the corresponding AUC.

a method based on graph neural networks oriented toward real-world applications. We achieve this by training a deep learning model to forecast complex dynamics after being trained on time series of activity with a given original network. Our approach is self-supervised and predicts structural defects without prior knowledge of the nature of the perturbations, nor of the dynamical process that generated the time series. It can thus be applied to various dynamics and networks with minimal modifications. We benchmark its performance on epidemiological, population, and neural dynamics over various synthetic and real networks. We show that it outperforms standard reconstruction methods while achieving a high level of precision, and proves to be greatly robust to noisy datasets.

Our work provides a novel method to study and monitor stressed complex networks. Beyond paving the way of bringing deep learning frameworks into the study of dynamical networks, our versatile approach can be used to address concrete problems of real-world systems. To name a few, it includes monitoring ecological systems [141], documenting temporal evolution of gene regulatory networks [70], detecting leaks in water flow networks [139], and evaluating pathological structures of brains disorders [104].

...

4.5 Context

Consider a graph G composed of N nodes and a set \mathcal{E} of $M = |\mathcal{E}|$ edges. We denote the adjacency matrix by \mathbf{A} whose element $a_{ij} = 1$ if there exists an interaction from node j to node i and $a_{ij} = 0$ otherwise. We consider T discrete observations of the nodes activity represented by a time series $\mathbf{X} = \{\mathbf{x}(t) \in \mathbb{R}^N\}_{t=1\dots T}$, where the element $x_i(t) = [\mathbf{x}(t)]_i$ is the activity of node i at time t . The time series is generated from a hidden and potentially stochastic dynamical mechanism,

$$\mathbf{X} = \mathcal{M}[\mathbf{A}, \mathbf{X}_0, \boldsymbol{\theta}], \tag{4.1}$$

where \mathbf{X}_0 is the initial state of the system and $\boldsymbol{\theta}$ are unknown parameters. The dynamics can take any form respecting the condition that the time evolution of a node activity depends on itself and the activity of its neighbors only. It ensures that the adjacency matrix truly governs the interactions.

We consider the following scenario : For $t < \tau$, the dynamics governed by (4.1) is taking place on the original graph whose edges are \mathcal{E} . At $t = \tau > 0$, the graph is perturbed by removing a set $d\mathcal{E} \subset \mathcal{E}$ of edges. The effect of the perturbation is a shift in the adjacency matrix $\mathbf{A}' = \mathbf{A} - \mathbf{P}$ where \mathbf{P} is called *the perturbation matrix*. For $t \geq \tau$, the dynamics takes place on the perturbed graph of adjacency matrix \mathbf{A}' . Note that the change point τ is only introduced to simplify the notation and is assumed to be known. Our analysis could have been done equivalently with the more general scheme of having two distinct and disconnected time series, the dynamics occurring on the original graph and the dynamics on the perturbed graph.

We now address the task of inferring the perturbation matrix \mathbf{P} relying on the initial adjacency matrix \mathbf{A} , time series \mathbf{X} , and the moment of perturbation τ . Three examples illustrating the task are displayed in Fig. 4.1. Reasonably, this problem can be solved by inferring the perturbed structure, $\mathbf{A}' \approx \hat{\mathbf{A}}' = f(\mathbf{X})$ from the observation of the nodes activity. The estimated perturbation matrix $\hat{\mathbf{P}}$ is composed of the original edges missing from the inferred structure, i.e., $\hat{\mathbf{P}} = \mathbf{A} - \hat{\mathbf{A}}'$. However, reconstructing networks from time series is a notoriously hard problem and, yet, no universal method exists [108]. Assuming that the structure \mathbf{A} and nodes' activity before the perturbation $t < \tau$ are known, properly incorporating these features into the reconstruction process becomes a critical facet of the inference methods, i.e., $\hat{\mathbf{A}}' = f(\mathbf{A}, \mathbf{X})$.

4.5.1 Graph Neural Networks

We introduce a Graph Neural Networks (GNN) to leverage the given structural information. In recent years, GNN have been developed to be used on structural datasets [174] with graph-based tasks [64, 90]. The developed model is trained on the self-supervised task of forecasting

the nodes activity while optimizing on numerous internal parameters. More precisely, our GNN model forecasts a node activity from previous dynamical states of the neighbors,

$$\hat{\mathbf{x}}(t+1) = \begin{cases} \text{GNN}(\mathbf{A}, \mathbf{X}^{(a)}(t), \mathbf{\Lambda}), & \text{if } t < \tau, \\ \text{GNN}(\mathbf{A} - \sigma(\hat{\mathbf{h}}_{\text{GNN}}), \mathbf{X}^{(a)}(t), \mathbf{\Lambda}), & \text{if } t \geq \tau \end{cases} \quad (4.2)$$

where $\mathbf{X}^{(a)}(t) = \{\mathbf{x}(t')\}_{t'=t-a, \dots, t}$ is the a past steps of activity, $\hat{\mathbf{x}}(t+1)$ is the forecasted activities at time $t+1$, $\mathbf{\Lambda}$ and $\hat{\mathbf{h}}_{\text{GNN}} \in \mathbb{R}^{N \times N}$ are trainable parameters, and $\sigma(\cdot)$ is the sigmoid function. In (4.2), the graph of interactions is given by the structure $\mathbf{A} - \sigma(\hat{\mathbf{h}}_{\text{GNN}})$ if the forecast is next to the perturbation $t > \tau$, and simply \mathbf{A} otherwise. Hence, the perturbation and the dynamical mechanisms are parametrized separately using $\hat{\mathbf{h}}_{\text{GNN}}$ and $\mathbf{\Lambda}$ respectively. The inferred perturbation matrix is given by $\hat{\mathbf{P}} = \sigma(\hat{\mathbf{h}}_{\text{GNN}})$. Note that the sigmoid function $\sigma(\cdot)$ is used to limit the perturbation amplitude between 0 and 1.

During training, examples of previous and future activities are presented to the model and the error on a loss function $L[\hat{\mathbf{x}}, \mathbf{x}]$ between the forecasted activity $\hat{\mathbf{x}}$ and the observed activity \mathbf{x} is backpropagated through the model for parameters optimization [See the Materials and Methods for details]. Eventually, the model reaches a stable minimum of the loss function and the perturbation matrix can be estimated as $\hat{\mathbf{P}} = \sigma(\hat{\mathbf{h}}_{\text{GNN}})$. Model and training details are provided in the Materials and Methods section.

By design, the GNN is inductive, which means that its predictions are independent of the target node identity and only depend on the states of the target node and its neighbors. The learned dynamical mechanism is shared among all nodes, although the neighborhood activity may differ. It also means that forecasting all N node states is equivalent to operate N independent forecasts, one for each node of the graph. Hence, the number of parameters of $\mathbf{\Lambda}$ does not have to scale up with the number of nodes in the graph and the GNN model can be lightweight memory-wise and computationally efficient.

...

4.6 Results

We introduce two types of inference methods to compare with the GNN model. The first are functional reconstruction algorithms. The adjacency matrix after the perturbation \mathbf{A}' is approximated by ad-hoc metrics such as the correlation matrix [17] or the Granger Causality [27, 34, 58, 143, 150], and the perturbation matrix is estimated by finding the original edges missing from the inferred structure $\hat{\mathbf{P}} = \mathbf{A} - \hat{\mathbf{A}}'$.

For the second type, we assume that the dynamical mechanisms are perfectly known. It cuts out the challenge of learning the dynamics to focus on detecting the perturbation. In the context

of stochastic dynamics, we develop a Bayesian inference method, inspired by Ref. [138], that estimates the perturbation by sampling the posterior distribution $\Pr(\mathbf{P}|\mathbf{A}, \mathbf{X}, \boldsymbol{\theta})$. This method is deeply advantaged compared to the GNN model as the dynamical mechanism $\mathcal{M}[\cdot, \cdot, \boldsymbol{\theta}]$ is assumed to be known and is used explicitly to compute the posterior distribution. Therefore, it serves as an idealized reference point to compare with our GNN model.

We measure the performance of the algorithms using the area under the curve (AUC) of the ROC curve. The AUC can be interpreted as the probability that the model distinguishes whether an edge is present or absent from the perturbation set $d\mathcal{E}$, independently of the threshold applied on the estimated perturbation matrix $\hat{\mathbf{P}}$. Note that an AUC equals to 0.5 is achieved with a uniform and non-informative baseline, whereas an AUC equals to 1 indicates that all the edges from the ground-truth perturbation have the top scores on the inferred perturbation matrix $\hat{\mathbf{P}}$.

4.6.1 Spreading dynamics

We evaluate the GNN model over an epidemic spreading dynamics called the susceptible-infectious-susceptible model (SIS) [135]. Nodes can be either infected ($x_i = 1$) or susceptible ($x_i = 0$). At each time step, infected nodes can infect their susceptible neighbors with a probability α . Infected individuals, on the other hand, become susceptible again with probability β . An example of a time series and prediction of the perturbations is given in Fig. 4.1(A,B,C).

Random networks

In Fig. 4.2, we report the AUC for spreading dynamics over the Erdős-Rényi random networks [124] for various parameters. It is striking that the GNN performs as well as the Bayesian model, a result that is consistent over all experiments. This may come as a surprise as the GNN has been given a more difficult task : Learning both the dynamics and the perturbation. Thus, it suggests that the GNN has simultaneously achieved the task of learning the dynamical mechanism and struggles as much as the Bayesian method on detecting the perturbation.

Another important result is that the GNN model outperforms functional reconstruction methods over all experiments. It demonstrates the importance of explicitly considering the dynamical mechanism and the prior network structure, instead of using a purely functional-based definition of the connectivity.

All methods perform best on low connectivity networks [Fig. 4.2(A)]. Perturbations on denser networks are harder to infer, independently of the model used. At least two phenomena justify this. First, the number of edges involved in the perturbation is fixed while the total number of edges grows with the average connectivity of the network. Hence, finding the right removed

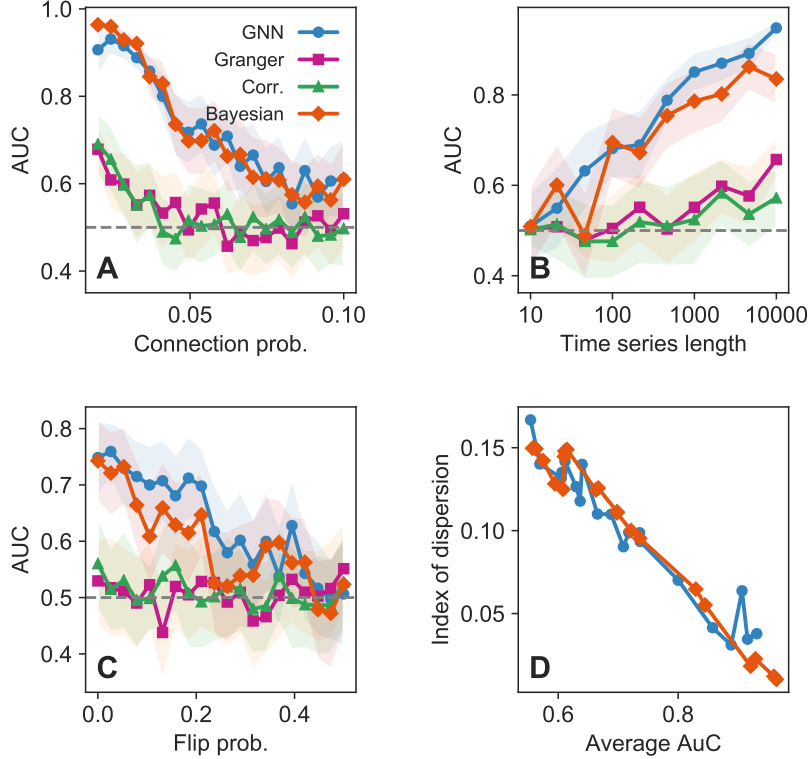


FIGURE 4.2 – Area under the curve for SIS dynamics on Erdős-Rényi networks. (A) With different network connection probability p and $N = 100$, $T = 300$, $\alpha = \beta = 0.1$. (B) With different time series length and $N = 100$, $p = 0.05$, and transition probabilities $\alpha = 0.2$, $\beta = 0.1$. (C) AUC under various flip probabilities with $N = 100$, $p = 0.05$, $T = 300$, $\alpha = \beta = 0.1$. Each symbol is the average over 100 independent simulations with similar configurations and standard deviation is displayed by a shaded region. The gray dashed lines are the non-informative baseline $AUC = 0.5$. (D) Index of dispersion $D = \sigma^2/\mu$ of AUC for data from (A).

edges in an increasing number of candidates naturally lead to a more difficult task, irrespectively of the method employed. Second, the average number of infected nodes monotonically increases with the connectivity of the network [See Supplementary Information]. Therefore, the probability of being infected remains almost the same after the removal of a single edge in dense networks. Consequently, perturbations are barely detectable.

Figure 4.2(B) shows that increasing the length of the time series leads to better performance, especially for the GNN and Bayesian approaches. Larger time series are equivalent to having a larger dataset, hence a more robust inference. This is beneficial for the GNN in two ways since both the inference of the dynamics and of the perturbations are enhanced. Building on results of Fig. 4.2(A), we conclude that perturbations on dense random networks can be inferred with high confidence ($AUC > 0.9$) if using sufficiently large time series, e.g., $T \sim 10^4$ entries.

The consistency of the inference can be expressed using the index of dispersion $D = \sigma^2/\mu$ of the AUC over experiments sharing similar settings. We observe that the dispersion decreases with

the average AUC [Fig. 4.2(D)]. It means that the GNN model becomes even more consistent on similar experiments with high AUC outcomes. Moreover, because the dispersion coefficient is highly similar between the GNN model and the Bayesian model, it suggests that the error source is similar. We hypothesize that large AUC variance is produced by inherent fluctuations in the generated dataset rather than the sensitivity of the algorithms.

Scale-free networks

Networks featuring hubs that dominate the connectivity are common in nature [13]. We benchmark the models on tree-like scale-free networks generated from the Barabási–Albert model [14](Fig. 4.3). In the Supplementary Information, we provide a validation of the GNN model with non-tree scale-free networks as well.

In general, the GNN model performance is much higher on scale-free networks than on random networks. This may be due to the tree-like structure of the generated networks. Perturbations tend to break up networks into disconnected components. Smaller and peripheral components tend to quickly deactivate. The removed edges may then be easily identified as the bridges between the dynamically distinct components. This could further be supported by their notable contributions in the dynamical likelihood [See Supplementary Information for details].

Fig. 4.3(A) shows a decrease in AUC for the functional reconstruction methods as the infection probability increases. Based on the previous discussion, increasing the infection probability raises the contrast between the inactive and active components, which would normally help infer the perturbation. To explain this behavior, note that the correlation (4.14) between an active node $\mathbf{x}_i \approx \mathbf{1}_T$ and an inactive node $\mathbf{x}_j \approx \mathbf{0}_T$ is roughly equal to zero, but so is the correlation for pairs of active nodes. Similar arguments can be made for the Granger causality. Therefore, for high infection probability, structural information becomes hidden from functional reconstruction methods, explaining their poor performance. They perform best when the infection is sparse. It highlights the importance of using techniques that take the nature of the dynamics into consideration.

In Fig. 4.3(B), the models are tested against increasing network size. We observe that the GNN model performance is roughly constant up to 10^3 nodes. This may come as a surprise when recalling that the perturbation is composed of a single edge and that the model has only 300 time steps to learn both the dynamics and the perturbation. Note that we observe a slight decrease in the performance of the Bayesian model over large networks. However, this trend is due to an insufficient number of sampling steps compared to the network size.

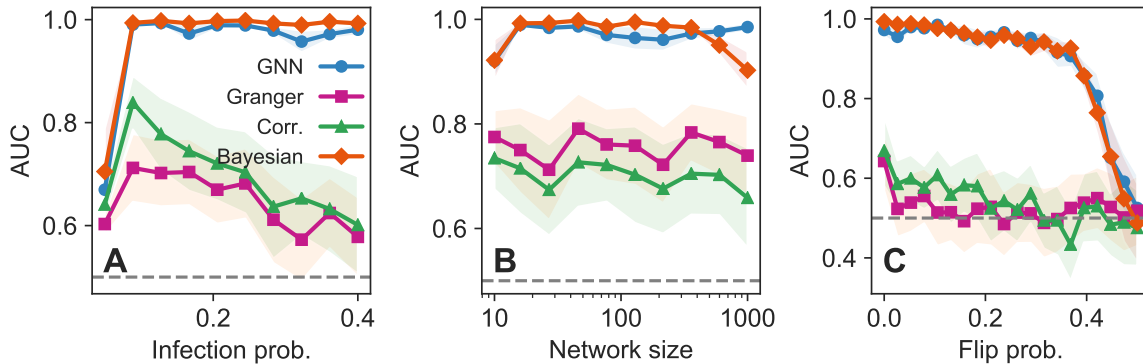


FIGURE 4.3 – Area under the curve for SIS dynamics on Barabási–Albert scale-free networks. (A) Prediction over various infection probability α on networks of $N = 100$ nodes. (B) AUC over increasing network size with $\alpha = 0.2$. (C) AUC over increasing flip probability with $\alpha = 0.2$. The likelihood of the Bayesian model of (C) is corrected to integrate the flip mechanism. Dots are averages over 100 simulations with similar configurations, and standard deviations are indicated by shaded regions. Epidemics spreading are simulated over $T = 300$ time steps and $\beta = 0.1$. Perturbations occur at $\tau = 150$ and consists in removing a single random edge.

Noisy time series

Typical real datasets are noisy [107]. Time series come as a mixture of some hidden dynamical mechanism masked by a layer of noise. We investigate the robustness to noise of the GNN model. To simulate noise, entries of the time series are flipped with probability p_{flip} , i.e., 0 goes to 1 and 1 to 0. For $p_{\text{flip}} = 0.5$, we obtain uncorrelated time series (independent Bernoulli processes).

For the next experiments, we use a corrected Bayesian model with a likelihood function that explicitly considers the flip mechanism, including the ground-truth flip probability [See Materials and Methods]. This modification to the Bayesian model is essential, as new transitions originally forbidden become possible with the introduction of noise. Precisely, in the original Bayesian version, null values of likelihood appear and sampling of the posterior distribution becomes impossible. While the Bayesian model must be revisited, our GNN model remains unchanged, incorporating automatically the effect of noise in its predictions during training.

We compare the performance over various flip probabilities on scale-free networks [Fig. 4.3(C)] and random networks [Fig. 4.2(C)]. We report that the GNN model is as precise as the corrected Bayesian model. On scale-free networks, it maintains high AUC over large flip probability. This performance is quite remarkable as the GNN model has to learn the dynamical mechanism from a noisy time series while inferring the missing edge. For random networks, the results show high AUC variances. It indicates that the noise process leads to larger fluctuations in precision, which is shared among the GNN and the Bayesian models. We again report that functional reconstruction methods are outperformed by the GNN approach.

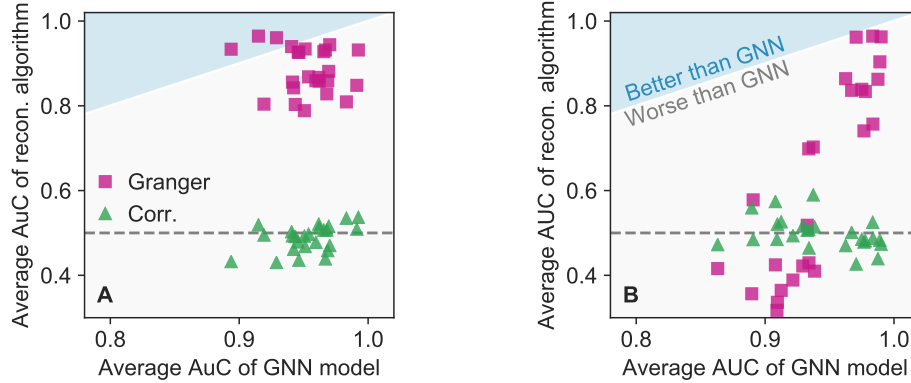


FIGURE 4.4 – Average area under the curve of predictions on 25 real ecological networks with predator-prey dynamics. Each marker is associated to a specific real network. The dynamics are simulated on $T = 10^3$ steps with (A) $\Delta t = 1/5$ and (B) $\Delta t = 1$. The functional reconstruction algorithms perform better when the dots are in the upper blue region of the line $x = y$. Random baselines are indicated with dashed gray lines. Averages are computed over 100 runs on each network. The perturbation is the removal of a single random edge at $\tau = 500$.

4.6.2 Population dynamics

Population dynamics are valuable tools to study ecological systems. Species population varies according to their predator-prey relationships.

In these systems, each node represents a different species and its population $x_i(t)$ varies in time according to a predator-prey relationship. We benchmark our models on a simple population dynamics given by

$$\dot{x}_i = c_i x_i + x_i \sum_{j=1}^N (a_{ij} - a_{ji}) x_j, \quad (4.3)$$

where $c_i = c \sum_j (a_{ji} - a_{ij}) \in \mathbb{R}$ is the fraction of the number of interactions that a species has [See Materials and Methods]. Note that introducing a species-dependent parameter c_i raises the difficulty for dynamical learning as the dynamical parameters are now specific to each node. Now that the dynamics is deterministic, the Bayesian approach previously developed no longer applies. Hence, the GNN model is compared with the correlation and the Granger causality approaches on 26 real directed ecological networks [Example in Fig. 4.1(D,E,F)]. We also introduce the sampling parameter Δt that indicates the time interval between two steps in the time series. Large Δt implies a coarser sampling of time series.

Our results show that the GNN (Granger) approach reaches an average AUC of 0.95 (0.88) over all simulations for $\Delta t = 1/5$ and 0.94 (0.64) for $\Delta t = 1$ [Fig. 4.4]. The GNN performance suggests that it is a valuable candidate for deterministic context, and even when the dynamical parameters of each node are dependent of their local structure.

Interestingly, the GNN model maintains a high AUC for all 26 real networks and large sampling

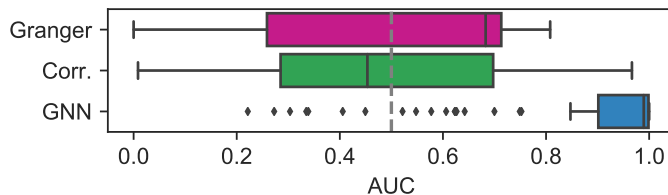


FIGURE 4.5 – Area under the curve of predictions over neural dynamics on the C. Elegans connectome. The perturbation is composed of three random edges and dynamics spans over $T = 500$ time steps with $\tau = 250$. Boxplots are computed over 100 simulations. Outliers (outside of 75th percentile plus 1.5 interquartile range) are indicated with markers.

steps Δt . The GNN performance is mostly unaffected by the characteristic of the network structure.

4.6.3 Neural dynamics

We challenge the versatility of our GNN model on a neural dynamics simulated on the C.Elegans connectome. First, we simulate the continuous noisy theta model of (4.4) on the directed network structure [48]. Then, we threshold the time series to only record the spikes of activity. Doing so cuts off valuable information for learning the dynamics which greatly increase the difficulty of the inference task [Example in Fig. 4.1(G,H,I)]. Also, the simulated theta model includes a random noise injection during integration.

Results indicate that the GNN model is highly capable of handling this spiking neural dynamics [Fig. 4.5]. While 17% of the runs show AUC below 0.75, most of them are above the expected values from functional reconstruction algorithms. On average, the AUC is 0.91 ± 0.19 with a median value at 0.99. It demonstrates the GNN’s high accuracy in the deterministic context, even in the presence of partial information.

...

4.7 Discussion

Detecting perturbations of complex systems is a challenge of paramount importance, especially in modern times with the impacts of climate changes on ecosystems and our more than ever entangled societies. Yet, the lack of reliable predictive tools to infer the presence and the structural cause of disturbances is of serious concern. Without taking full advantage of the graph structure, functional effect of perturbations can be so subtle that it could remain undetected at the local and global levels.

In this paper, we have introduced a method relying on recent deep learning advances to infer structural perturbations from time series of activity. The core and original idea is the self-supervised training scheme that neither requires prior information on the perturbation strategy nor on the dynamical mechanism. Minimal modifications of the model are required to apply the method on datasets of different natures, such as continuous, noisy, or discrete time series.

We have tested the method in three contrasting contexts : spreading, ecological, and neural dynamics. We show that our approach outperforms standard functional reconstruction methods while being comparable to Bayesian models with a priori knowledge of the dynamical mechanisms. Our results suggest that GNN are promising candidates to predict structural disturbances in practical applications.

Apart from its effectiveness, there are multiple advantages of using the GNN model. First, the optimization procedure scales, both in speed and accuracy, to large networks, and is fairly robust to the hyperparameters choice. Second, GNN models are part of an active field of research so that future improvements are expected, e.g., enhanced neural network architecture and improved gradient descent optimizer [101]. Third, GNN are versatile to different dynamics while being able to cope with large levels of noises. Finally, they are well adapted to support other types of information collected on the graph that could help the predictions.

Our study provides a new avenue for monitoring complex systems. It can be used for detecting disturbances in various classes of complex networks, ranging from ecology to neuroscience, and paves the way for the development of targeted intervention strategies for sustainable management.

...

4.8 Materials and Methods

Empirical Networks

The ecological networks are food webs obtained from the Web of Life database (www.web-of-life.es). Supplementary information gives an overview of their structural properties. The C.Elegans connectome is obtained from Ref. [85].

Generated networks

Random networks are generated from the Erdős-Rényi model : N nodes are randomly connected with probability p [124]. The scale-free networks are generated from the Barabási-Albert model [14]. We first start with a connected pair of nodes. Then, at each step, we add a new node to the network and connect it to $m = 1$ existing node chosen with probability proportional to their degrees.

Simulated dynamics

For all experiments, we first initialize the nodes activity and run the dynamics for 100 burning steps which are then discarded. Then, we simulate the dynamics for τ steps. At $t = \tau$, we remove random edges from the original network. Then, the dynamics is continued until $t = T$. The time series $\mathbf{X} = \{\mathbf{x}(t) \in \mathbb{R}^N\}_{t=1\dots T}$ is then constructed from the vectors of nodes activity at each time step.

Susceptible-Infected-Susceptible dynamics (SIS)

Nodes activity is binary $\mathbf{x}(t) \in \{0, 1\}^N$ where $x_i(t) = 1$ indicates that node i is infected at time step t and $x_i(t) = 0$ if susceptible. At each time step, a susceptible node i is infected with probability $1 - (1 - \alpha)^{l_i}$, where $l_i = \sum_j a_{ij}x_j$ is its number of infected neighbors and α is the transmission probability. Infected nodes recover with constant probability β . The nodes activity is initialized with equiprobable binary values.

Lotka-Volterra dynamics

We have simulated mixed ecosystems from a generalized Lotka-Volterra dynamics (4.3). The inherent growth of species is given by $c_i x_i$. We chose $c_i = 0.25 \sum_j (a_{ji} - a_{ij})$. The diagonal of the adjacency matrix is set to zero, $a_{ii} = 0$, to prevent intraspecies competition. The initial population is uniformly distributed between 0 and 1. The integration is done using an eighth order Runge-Kutta method.

Noisy neural dynamics

Neural dynamics on the C. Elegans are simulated using the noisy theta model. First, we integrate the ODEs for the the phase of oscillators,

$$\begin{aligned} \dot{y}_i = & [1 - \cos(y_i)] \\ & + [1 + \cos(y_i)] \left[I_i + \theta N^{-1} \sum_{j=1}^N a_{ij} [1 - \cos(y_j)] \right], \end{aligned} \quad (4.4)$$

where $y_i(0)$ is initialized uniformly between 0 and 2π . The parameters I_i control the noise level and are sampled from a uniform distribution $U(0.05, 0.15)$ and $\theta = 5$. We integrate the system using an Euler integrator with $dt = 0.1$. The external current I_i is sampled on each time step. A spike is said to be produced when $\cos(y_i) > 0.95$, i.e., when y_i is close to π . The continuous variable y_i is transformed into a discrete state variable $x_i = H[\cos(y_i) - 0.95]$ where $H(y) = 1$ if $y > 0$ and $H(y) = 0$ otherwise. The time series \mathbf{y} is then discarded and only \mathbf{x} is used for inference.

GNN model architecture

We have trained graph neural networks to forecast the activity. For a single forecast of the activity of node i , the GNN takes as input an adjacency matrix \mathbf{A} and a window $\mathbf{X}^{(a)} = \{\mathbf{x}_i^{(a)}\}_{i=1\dots N} \in \mathbb{R}^{a \times N}$ of length a of consecutive time steps of activity of all N nodes, and it outputs the forecasted activity $\hat{\mathbf{x}} = \{\hat{x}_i\}_{i=1\dots N}$ of all nodes,

$$\hat{\mathbf{x}} = \text{GNN}(\mathbf{A}, \mathbf{X}^{(a)}, \mathbf{\Lambda}). \quad (4.5)$$

The GNN is formed of two main operations. For an individual node i ,

$$\mathbf{y}_i = h_{\mathbf{\Lambda}} \left(\mathbf{x}_i^{(a)} + \sum_{j=1}^N a_{ij} \mathbf{x}_j^{(a)} \right) \quad (4.6a)$$

$$\hat{x}_i = F(\mathbf{y}_i), \quad (4.6b)$$

where (4.6a) is a parametrized and non-linear operation called a GINConv layer [174], and $F(\cdot)$ is a trainable function. The operation $h_{\mathbf{\Lambda}}$ is a sequence of linear transformations and ReLU activations with the exact form detailed in the Supplementary information.

For directed networks, (4.5) is replaced with a linear transformation of two distinct neural networks

$$\hat{\mathbf{x}} = \mathbf{W}[\text{GNN}(\mathbf{A}, \mathbf{X}^{(a)}, \mathbf{\Lambda}_1) \parallel \text{GNN}(\mathbf{A}^T, \mathbf{X}^{(a)}, \mathbf{\Lambda}_2)]. \quad (4.7)$$

All GNN model parameters $\mathbf{\Lambda}$ are randomly initialized using Kaiming’s method [66].

GNN training

We first initialize $\hat{\mathbf{h}}_{\text{GNN}}$ so that $\sigma(\hat{\mathbf{h}}_{\text{GNN}}) = 10^{-2}$ for existing edges of \mathbf{A} and 0 otherwise.

A training step goes as follows. We randomly slice a window at randomly chosen time of length $a + 1$ in \mathbf{X} . The first a values, $\mathbf{X}^{(a)}$, are given to the model while the last values are used as the forecasting target \mathbf{x} . We used $a = 1$ for the SIS and predator-prey dynamics, and $a = 20$ for the neural dynamics. The neural dynamics gets a longer activity history because the actual state of the system is hidden due to the binary spike transformation. Therefore, the information from the observed state is only partial and the time series are non-Markovian. For these reasons, the task is inherently more difficult and requires a longer activity history.

We apply a forward step following the prescription of (4.2). We optimize the target forecast over a cross-entropy loss for binary dynamics (epidemics, neural) and the absolute error loss for continuous dynamics (ecological) with the RAdam optimizer [101].

We use a learning rate of 10^{-3} for the parameters $\mathbf{\Lambda}$, with a decay of 0.1 after 1000 steps, and a learning rate of 10^{-2} for the perturbation matrix $\hat{\mathbf{h}}_{\text{GNN}}$. The model is optimized on 10^5 steps. It roughly takes 30 seconds per training on a personal computer to achieve an inference over a network of $N = 100$ nodes with time series of length $T = 300$.

Bayesian model

We estimate the perturbed structure from the expected value of the posterior distribution as

$$\hat{\mathbf{A}}' = \sum_{\mathbf{A}^*} \mathbf{A}^* \Pr(\mathbf{A}^* | \mathbf{X}, \mathbf{A}, \boldsymbol{\theta}), \quad (4.8)$$

and the perturbation matrix from $\hat{\mathbf{P}} = \mathbf{A} - \hat{\mathbf{A}}'$. The posterior can be rewritten using Bayes theorem

$$\Pr(\mathbf{A}^* | \mathbf{X}, \mathbf{A}, \boldsymbol{\theta}) \propto \Pr(\mathbf{X} | \mathbf{A}, \mathbf{A}^*, \boldsymbol{\theta}) \Pr(\mathbf{A}^* | \mathbf{A}), \quad (4.9)$$

where $\Pr(\mathbf{X} | \mathbf{A}, \mathbf{A}^*, \boldsymbol{\theta})$ is the likelihood and $\Pr(\mathbf{A}^* | \mathbf{A})$ is the prior distribution, that we consider uniform. Moreover, since only the portion $t > \tau$ is needed to infer the perturbed matrix, we can drop the dependence on \mathbf{A} in the likelihood, i.e., $\Pr(\mathbf{X} | \mathbf{A}^*, \boldsymbol{\theta})$.

The posterior of (4.8) is sampled using a Metropolis-Hasting scheme [116]. We denote the adjacency matrix at sampling step ν by \mathbf{A}_ν that we have initialized with the unperturbed graph $\mathbf{A}_0 = \mathbf{A}$.

At each sampling step, we randomly select an edge $(i, j) \in \mathcal{E}$. If the edge (i, j) is present ($[\mathbf{A}_\nu]_{ij} = 1$), we remove it, $[\mathbf{A}_{\nu+1}^*]_{ij} = 0$, otherwise we add it back, $[\mathbf{A}_{\nu+1}^*]_{ij} = 1$. Since graphs are undirected for the SIS dynamics, each graph proposition is symmetrically applied, e.g., $[\mathbf{A}_{\nu+1}^*]_{ij} = [\mathbf{A}_{\nu+1}^*]_{ji}$.

The probability of accepting the proposition $\mathbf{A}_\nu \rightarrow \mathbf{A}_{\nu+1}^*$ at step $\nu + 1$ is given by

$$\rho = \min \left(1, \frac{\Pr(\mathbf{X}|\mathbf{A}_{\nu+1}^*, \boldsymbol{\theta})}{\Pr(\mathbf{X}|\mathbf{A}_\nu, \boldsymbol{\theta})} \right). \quad (4.10)$$

The likelihood $\Pr(\mathbf{X}|\mathbf{A}^*, \boldsymbol{\theta})$ of the SIS model can be approximated by the product over individual transition probabilities for each node

$$\Pr(\mathbf{X}|\mathbf{A}^*, \boldsymbol{\theta}) = \prod_{i=1}^N \prod_{t>\tau}^T \Pr[x_i(t)|x_i(t-1); \mathbf{A}^*, \mathbf{X}(t-1), \boldsymbol{\theta}]. \quad (4.11)$$

which is exact if the process is noiseless, $p_{\text{flip}} \equiv q = 0$. For a node i of degree k , with $l \equiv \sum_j a_{ij} x_j(t-1)$ infected neighbors, $x_i(t-1) \equiv v$ and $x_i(t) \equiv w$, the transition probabilities are

$$\begin{aligned} \Pr(w|v; l, k, \boldsymbol{\theta}) = & R(v, q)R[w, R(\beta, q)] , \\ & + [1 - R(v, q)]R[w, R(F_{kl}, q)] , \end{aligned} \quad (4.12)$$

where $R(y, z) \equiv y(1 - z) + (1 - y)z$ and

$$F_{kl} = [q + (1 - q)(1 - \alpha)]^l [1 - q + q(1 - \alpha)]^{k-l}. \quad (4.13)$$

Recall that α and β are given SIS parameters, and q is the flip probability. Complete derivation is covered in the Supplementary information. We sample the posterior distribution over $\Gamma = 2000$ propositions. We estimate the posterior distribution of (4.8) by counting the relative number of steps held by a certain graph, i.e., $\Pr(\mathbf{A}^*|\mathbf{X}, \mathbf{A}, \boldsymbol{\theta}) = \Gamma^{-1} \sum_\nu \mathbb{I}(\mathbf{A}_\nu - \mathbf{A}^*)$ where $\mathbb{I}(\cdot) = 1$ only if the argument is element-wise zero and $\mathbb{I}(\cdot) = 0$ otherwise.

Functional reconstruction methods

The correlation matrix has elements

$$\hat{a}_{ij} = \left| \frac{(\mathbf{x}_i - \mathbf{1}\mu_i)^T (\mathbf{x}_j - \mathbf{1}\mu_j)}{\sigma_i \sigma_j} \right|, \quad (4.14)$$

where $\mu_i = \langle \mathbf{x}_i \rangle$ is the average activity of node i and σ_i is the standard deviation of \mathbf{x}_i . In the Granger causality method [58], the element (i, j) of the adjacency matrix is estimated as the ratio of the standard deviation of the forecasting error of linear models $F(\cdot)$ trained with solely the activity of node i and the other, $G(\cdot, \cdot)$, trained with the activity of nodes i and j ,

$$\hat{a}_{ij} = \frac{\text{std}[\mathbf{x}_i - F(\mathbf{x}_i)]}{\text{std}[\mathbf{x}_i - G(\mathbf{x}_i, \mathbf{x}_j)]}. \quad (4.15)$$

For both methods, the estimated adjacency matrix $\hat{\mathbf{A}}'$ is computed using only the nodes activity after the perturbation. To suppress initially absent edges, we also perform an element-wise multiplication between $\hat{\mathbf{A}}'$ and the original adjacency matrix \mathbf{A} . If the network is undirected, we average the score over in and out edges, i.e., $\hat{\mathbf{A}}' \leftarrow \frac{1}{2}(\hat{\mathbf{A}}' + \hat{\mathbf{A}}'^T)$. We then normalize the estimated adjacency matrix and evaluate the perturbation matrix by $\hat{\mathbf{P}} = \mathbf{A} - \hat{\mathbf{A}}'$.

Evaluation metrics

Predicted removed edges are compared to ground-truth edges under a Receiver Operating Characteristic (ROC) curve analysis. For each inference, models output a weighted matrix of perturbation $\hat{\mathbf{P}}$. Note that only existing edges of \mathbf{A} are assigned to a non-zero score because adding edges is prohibited. Perturbation scores are then rescaled between 0 and 1 where a score of 1 is the most likely of being removed. Then, we compute the True/False Positive Rates (TPR/FPR) under a varying threshold r from 0 to 1 :

$$\text{TPR}(r) = \frac{N_{\text{TP}}(r)}{N_{\text{TP}}(r) + N_{\text{FN}}(r)} \quad (4.16)$$

$$\text{FPR}(r) = \frac{N_{\text{FP}}(r)}{N_{\text{FP}}(r) + N_{\text{TN}}(r)}, \quad (4.17)$$

where the positive edges are those with a higher score than the threshold, i.e., the edge (i, j) with $\hat{P}_{ij} > r$. The AUC is computed as the area under the ROC curve and is independent of the threshold r applied. The AUC can be interpreted as the probability that the model distinguishes whether an edge is present or absent from the perturbation set $d\mathcal{E}$.

...

4.9 Acknowledgements

This work was funded by the Fonds de recherche du Québec-Nature et technologies, the Natural Sciences and Engineering Research Council of Canada, and Sentinel North, financed by the Canada First Research Excellence Fund. The authors want to thank Antoine Allard, Patrick Desrosiers, and Louis J. Dubé for supports and helpful comments, Mohamed Babine for fruitful discussions and code sharing, and Martin Laprise for useful comments and for allocating computational resources. We acknowledge Calcul Québec for using their computing facilities.

...

4.10 Supplementary Material

4.10.1 GNN Model

Layers

The input of the GNN model for the target node i is its current activity $\mathbf{x}_i^{(a)}$, and the activity of its neighbors $\mathbf{x}_j^{(a)}$ for all j with $a_{ij} = 1$. Here, $\mathbf{x}_j^{(a)}$ denotes a slice of a consecutive time steps of node j activity.

The length a is a parameter that must be set beforehand. In practice, it is best if set to the expected memory of the dynamical mechanism. For instance, the memory of a markovian process is $a = 1$. If a is too small, then the input activity will not be sufficient to forecast the future activity. On the contrary, if a is much larger than the memory of the dynamical system, the model should learn to vanish the weights allocated for the oldest activity steps. A large a also exposes the model to overfitting by finding superfluous patterns in previous states of activity.

For a single forward pass, the GNN model applies the following operations :

$$\mathbf{y}_i = h_{\Lambda} \left(\mathbf{x}_i^{(a)} + \sum_{j=1}^N a_{ij} \mathbf{x}_j^{(a)} \right) \quad (4.18a)$$

$$\hat{x}_i = F(\mathbf{y}_i) \quad (4.18b)$$

where $h_{\Lambda}(\cdot)$ is a trainable layer, and $F(\cdot)$ is a function that may have trainable parameters. The function $h_{\Lambda}(\cdot)$ is called a GINConv [174]. It aggregates the neighborhood of node i by summing its neighbors' activity, and then by applying a non-linear operation described by $h_{\Lambda}(\cdot)$. Thanks to the sum operation on j , the layer is independent of the number of neighbors of node i . It makes the GNN model inductive as neither the size of the graph, nor the number of neighbors, are hard-coded in the layers.

During training, a loss function is computed between the model output and the ground-truth future activity,

$$L(\hat{x}_i, x_i). \quad (4.19)$$

The spreading and the neural dynamics use the binary cross entropy,

$$L(\hat{x}_i, x_i) = x_i \log(\hat{x}_i) + (1 - x_i) \log(1 - \hat{x}_i), \quad (4.20)$$

while the predator-prey dynamics uses the L1 loss,

$$L(\hat{x}_i, x_i) = |x_i - \hat{x}_i|. \tag{4.21}$$

The GINConv layer is the core operation of the GNN model. By explicitly using the adjacency matrix \mathbf{A} in Eq. (4.18a), the model is able to backpropagate the error into the adjacency matrix \mathbf{A} which will be crucial for inferring the perturbation. Note that the backpropagation requires the mathematical operations applied by the layers to be differentiable, which obviously is satisfied when using the sum aggregation.

The GINConv could have been equivalently written as a sparse operation on the node’s neighborhood \mathcal{N}_i ,

$$\mathbf{y}_i = h_{\Lambda} \left(\mathbf{x}_i + \sum_{j \in \mathcal{N}_i} \mathbf{x}_j \right), \tag{4.22}$$

where the independence with the number of nodes N in the graph still holds. Yet, slicing beforehand the nodes activity directly on the neighborhood of the node i would disable the differentiability of the adjacency matrix \mathbf{A} . Hence, the essential backpropagation of \mathbf{A} would fail.

Table 4.1 summarizes the mathematical operations of the model for the different dynamics. There is no guarantee that this selection of hyperparameters, e.g., number of linear transformations or the number of hidden dimensions, is the best possible configuration for our task neither that it will fit to a different dynamics. From our experiments, similar results were achieved with other sets of hyperparameters. Worse performance resulted from a bad choice of learning rates rather than an inappropriate model design. An ablation study could be made in the future to validate our observations.

It is worth noting that other GNN layers could have been used instead of the GINConv, such as the Graph Attention Layer (GAT) [163] and the SAGEConv [64].

We have used the GINConv mostly by convenience. Most deep learning frameworks of GNN models use a sparse representation of the adjacency matrix, which makes the adjacency matrix not differentiable. Only few layers have been coded by the PyTorch community for using a dense adjacency matrix.

The function $F(\cdot)$ is an activation layer used to control the shape of the output and its bounds. It receives as input \mathbf{y}_i , i.e., the output of the GINConv. We list the functions used for the

TABLE 4.1 – Layer by layer description of the GNN models for each dynamics. For each sequence, the operations are applied in order from top to bottom. A linear operation $\text{Linear}(n, m)$ is defined as $f(\mathbf{x}) = \mathbf{W}(\mathbf{x}) + \mathbf{c}$ with $\mathbf{W} \in \mathbb{R}^{m \times n}$ and $\mathbf{c} \in \mathbb{R}^m$. The ReLU operation is the element-wise maximum $\text{ReLU}(\mathbf{x}) = \max(0, \mathbf{x})$.

Dynamics	$h_{\Lambda}(\cdot)$	$F(\cdot)$	Number of parameters
SIS	Linear(1, 16)	Sigmoid	1 185
	ReLU		
Predator-prey	Linear(16, 64)	Linear(128, 1)*	37 569
	ReLU		
	Linear(64, 1)		
	Linear(1, 64)		
Neural	ReLU	Linear(64, 1)*	17 441
	Linear(64, 128)		
	ReLU		
	Linear(128, 64)	Sigmoid	
	Linear(20, 64)		
ReLU			
Linear(64, 64)			
ReLU			
Linear(64, 32)			

*The linear function is twice the size of the $h_{\Lambda}(\cdot)$ because the networks are directed and the outputs of two GINConvs are concatenated.

selected dynamics,

$$\text{SIS} : F(\mathbf{y}) = \frac{1}{1 + e^{-\mathbf{y}}}; \quad (4.23)$$

$$\text{Predator-prey} : F(\mathbf{y}) = \mathbf{W}\mathbf{y} + \mathbf{c}; \quad (4.24)$$

$$\text{Neural dynamics} : F(\mathbf{y}) = \frac{1}{1 + e^{-\mathbf{W}\mathbf{y} + \mathbf{c}}}, \quad (4.25)$$

with detailed dimensions in Table 4.1.

For the binary dynamics (SIS and neural dynamics), we use the sigmoid function to bound the output between 0 and 1. In doing so, the output can be interpreted as a probability that the next activity is the active state. For the predator-prey dynamics, the output is not required to be bounded. Therefore, we simply apply a linear transformation to match the desired output size. Note that \mathbf{W} and \mathbf{c} are trainable parameters.

For directed networks, we use two GINConvs so that the GNN model can distinguish between

in- and out- edges,

$$\mathbf{y}_i^{(1)} = h_{\Lambda}^{(1)} \left(\mathbf{x}_i + \sum_{j=1}^N a_{ij} \mathbf{x}_j \right) \quad (4.26a)$$

$$\mathbf{y}_i^{(2)} = h_{\Lambda}^{(2)} \left(\mathbf{x}_i + \sum_{j=1}^N a_{ji} \mathbf{x}_j \right) \quad (4.26b)$$

$$\hat{\mathbf{x}}_i = F(\mathbf{y}_i^{(1)} \parallel \mathbf{y}_i^{(2)}) \quad (4.26c)$$

where \parallel is the concatenation operator. This technique is essential to correctly infer on directed networks.

4.10.2 Weight decay

L2 weight decay is often used to prevent overfitting. In our task, we found that using weight decay impairs the learning of the perturbation. When applying a weight decay on the trainable perturbation matrix $\hat{\mathbf{h}}_{\text{GNN}}$, we penalize the model on the norm of $|\hat{\mathbf{h}}_{\text{GNN}}|^2$. Therefore, it pushes the elements to zero, $[\hat{\mathbf{h}}_{\text{GNN}}]_{ij} \rightarrow 0$. Recall that the forecast is done using

$$\hat{\mathbf{x}} = \text{GNN}(\mathbf{A} - \sigma(\hat{\mathbf{h}}_{\text{GNN}}), \mathbf{x}, \Lambda). \quad (4.27)$$

Hence, the direction $[\hat{\mathbf{h}}_{\text{GNN}}]_{ij} \rightarrow 0$ contributes to false positive detection $[\sigma(\hat{\mathbf{h}}_{\text{GNN}})]_{ij} \rightarrow 0.5$. We would rather look for the direction $[\sigma(\hat{\mathbf{h}}_{\text{GNN}})]_{ij} \rightarrow 0$ to push most edges toward a vanishing contributions except for truly removed edges. For these reasons, we did not use a weight decay in our simulations.

Note that the weight decay could be applied only on the inner parameters Λ without harming the perturbation. From experiments, using L2 weight decay did not lead to any noticeable improvements.

...

4.11 Large perturbations and loopy scale-free networks

We explore how the model is affected by removing a large number of edges. From Fig. 4.7(C,D), we conclude that the GNN model performs almost as well as the Bayesian model in multiple edges perturbations. A slight deviation between the GNN and Bayesian model is noticeable for scale-free networks. This issue is detailed in the next section. Moreover, the performance measured with the AUC is quite stable up to 15 removed edges. In the considered scale-free networks of Fig. 4.7(D), the perturbation represents 15% of all edges. In Fig. 4.7(B), we present

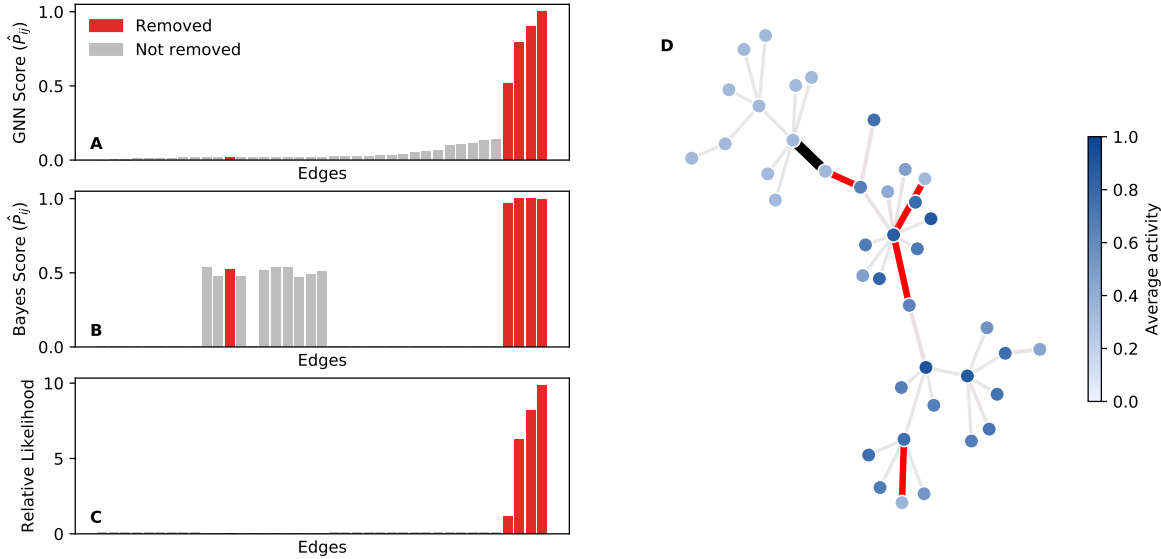


FIGURE 4.6 – Example of inference on the SIS dynamics with $\alpha = 0.12$ and $\beta = 0.1$ over a scale-free networks with 5 removed edges. In (A-B), the scores of the inferred perturbation matrix for all edges ordered by the GNN predicted scores. In (C), the relative variation of the likelihood $1 - \Pr(\mathbf{X}|\mathbf{A}^*, \boldsymbol{\theta})/\Pr(\mathbf{X}|\mathbf{A}, \boldsymbol{\theta})$ when removing the edge ordered similarly as (A-B). In D, the network used for the example where red edges and the black edge are the perturbation. The black edge is a false negative according to the model. Only the red edges have been correctly predicted by the GNN model with a high score. The nodes are colored according to their average activity after the perturbation.

the AUC for loopy scale-free graphs. We conclude again that the GNN model performs as well as the Bayesian approach.

...

4.12 Challenging edges in scale-free networks

Interestingly, for scale-free networks, we observe that some removed edges can be more challenging to infer than usual. This emergent aspect is caused by removing core edges that disconnect the network into small inactive components. Therefore, if certain edges have been removed deeper down into the smaller components of the disconnected tree structure, they would be practically impossible to infer because of the absence of activity. The only time frame to infer them would be just after the perturbation, while the activity is rapidly decaying in the periphery. We show an example in Fig. 4.6 where an edge from the perturbation does not induce a large likelihood drop. It is clearly shown that this edge connects two nodes with vanishing level of activity after the perturbation, which explains how difficult it can be to infer it.

...

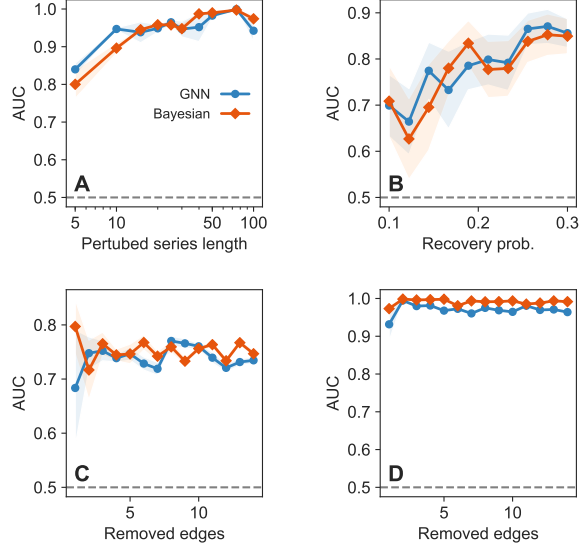


FIGURE 4.7 – AUC on various schemes with SIS dynamics. In (A), we have varied when the change point τ occurs. The length $T - \tau$ of the perturbed time series is displayed on the x-axis (Scale-free graphs $N = 100, m = 1$ with $T = 200$ and $\alpha = 0.2, \beta = 0.1$). (B) AUC on a scale-free graph with loops $N = 100, m = 3$ and $T = 200, \alpha = 0.1$. (C-D) AUC as a function of the number of removed edges (C : random graphs with $N = 100, p = 0.1, \alpha = 0.1, \beta = 0.1$) (D : scale-free graphs $N = 100, m = 1, \alpha = 0.2, \beta = 0.1$)

4.13 Length of the perturbed time series

We investigate the effects of decreasing the size of the perturbed time series. More precisely, the moment of perturbation τ is moved closer to the end of the time series. In Fig. 4.7(A), we have simulated the dynamics over a fixed $T = 200$ time steps. We execute the perturbation at various $t = \tau \in [0, 200]$. We observe that the maximum AUC is achieved with only 10 perturbed time steps.

4.13.1 Likelihood of the SIS model

4.14 Likelihood of the SIS model

We now detail the likelihood of the SIS model. Consider the transition probability $\Pr[x_i(t)|x_i(t-1); \mathbf{A}, \mathbf{X}(t-1), \boldsymbol{\theta}]$ for the noiseless SIS model $p_{\text{flip}} = 0$.

If the node i is infected at time $t - 1$, i.e., $x_i(t - 1) = 1$, then it may have recovered with probability β , i.e., $x_i(t) = 0$, or stayed infected with probability $1 - \beta$, i.e., $x_i(t) = 1$. Its contribution to the transition probability is

$$x_i(t - 1)[x_i(t)[1 - \beta] + [1 - x_i(t)]\beta]. \quad (4.28)$$

If the node is susceptible at time $t - 1$, i.e., $x_i(t - 1) = 0$, it may get infected with probability $1 - (1 - \alpha)^{l_i}$, where $l_i = \sum_j a_{ij} x_j(t - 1)$ is the number of infected neighbors, or stay susceptible with probability $(1 - \alpha)^{l_i}$. Its contribution to the transition probability is

$$[1 - x_i(t - 1)][x_i(t)[1 - (1 - \alpha)^{l_i}] + [1 - x_i(t)](1 - \alpha)^{l_i}]. \quad (4.29)$$

Combining (4.28) and (4.29), the transition probabilities are

$$\Pr[w|v; \mathbf{A}, \mathbf{X}(t - 1), \alpha, \beta] = vR(w, \beta) + [1 - v]R(w, g_l) \quad (4.30a)$$

where $w \equiv x_i(t)$, $v \equiv x_i(t - 1)$, $g_l = (1 - \alpha)^{l_i}$, and

$$R(y, z) = y(1 - z) + (1 - y)z. \quad (4.30b)$$

Because this is a Markov process, the likelihood of a time series is the product of the transition probabilities,

$$\Pr(\mathbf{X}|\mathbf{A}, \boldsymbol{\theta}) = \prod_{i=1}^N \prod_{t>\tau}^T \Pr[x_i(t)|x_i(t - 1); \mathbf{A}, \mathbf{X}(t - 1), \boldsymbol{\theta}]. \quad (4.31)$$

We now consider the SIS model with a flip probability p_{flip} . We introduce the observed variables \mathbf{X} and the noiseless activity \mathbf{Y} , i.e., equivalent to the true activity before executing the flips. Note that since we do not observe the whole state of the system (\mathbf{X} and \mathbf{Y}), correlation remains between $\mathbf{X}(t)$ and $\mathbf{X}(t - 2)$, even after conditioning on $\mathbf{X}(t - 1)$. Formally speaking, the process on the variables \mathbf{X} alone is non-Markovian. However, we expect the correlation to decrease quickly with time, so we use a Markovian approximation to write the likelihood. In other words, we use only $\mathbf{X}(t - 1)$ to describe $\mathbf{X}(t)$, as before.

First, consider that the node i is truly infected at time $t - 1$, i.e., $y_i(t - 1) = 1$. Assuming that we observe $x_i(t - 1)$, we can write the conditional probability by considering that the flip succeeded with probability p_{flip} and failed with probability $1 - p_{\text{flip}}$,

$$\Pr(y_i(t - 1) = 1|v) = v[1 - p_{\text{flip}}] + [1 - v]p_{\text{flip}} \equiv R[v, p_{\text{flip}}], \quad (4.32)$$

where $v = x_i(t - 1)$. For a truly infected node, two processes take place. First, the node can recover with probability β or stay infected with probability $1 - \beta$. Second, the flip process succeeds with probability p_{flip} or fails with probability $1 - p_{\text{flip}}$. This composition is enclosed in $R[x_i(t), R(\beta, p_{\text{flip}})]$. Therefore, the first contribution to the transition probability is

$$R[x_i(t - 1), p_{\text{flip}}]R[x_i(t), R(\beta, p_{\text{flip}})]. \quad (4.33)$$

The second contribution to the transition probability adopt that the node i is truly susceptible at time $t - 1$, i.e., $y_i(t - 1) = 0$. This occurs with probability $1 - R[x_i(t - 1), p_{\text{flip}}]$. Again, two

processes form the step : The possible infection of node i and the flip process with probability p_{flip} . Therefore, the second contribution is

$$(1 - R[x_i(t-1), p_{\text{flip}}])R[x_i(t), R(F_i, p_{\text{flip}})]. \quad (4.34)$$

where F_i is the probability that the node i *does not get infected by its neighbors*, i.e., $\Pr(y_i(t) = 0 | y_i(t-1) = 0, \mathbf{A}, \alpha, \beta)$. To obtain F_i , we need to consider the complete neighborhood of the node i . If a neighbor j is observed infected $x_j(t-1) = 1$, then two events could have occurred :

1. Node j is infected at time $t-1$, i.e., $y_j(t-1) = 1$, in which case the flip failed with probability $(1 - p_{\text{flip}})$. It fails to infect node i with probability $1 - \alpha$;
2. Node j is susceptible at time $t-1$, i.e., $y_j(t-1) = 0$, in which case the flip noise succeed with probability p_{flip} , and it failed to infect node i with probability 1.

As we observe l_i infected nodes, the contribution of observed infected neighbors is

$$[p_{\text{flip}} + (1 - p_{\text{flip}})(1 - \alpha)]^{l_i}. \quad (4.35)$$

For a node i with k_i neighbors, a similar argument can be made for the $k_i - l_i$ susceptible neighbors of node i , i.e., $x_j(t-1) = 0$ for all j that $a_{ij} = 1$. The total contribution is

$$[(1 - p_{\text{flip}}) + p_{\text{flip}}(1 - \alpha)]^{k_i - l_i}. \quad (4.36)$$

Combining (4.35) with (4.36), we obtain

$$F_i = [p_{\text{flip}} + (1 - p_{\text{flip}})(1 - \alpha)]^{l_i} [1 - p_{\text{flip}} + p_{\text{flip}}(1 - \alpha)]^{k_i - l_i}. \quad (4.37)$$

From (4.33) and (4.34), the transition probabilities of the SIS model with flips are

$$\Pr[w|v; \mathbf{A}, \mathbf{X}(t-1), \alpha, \beta] = R[v, p_{\text{flip}}]R[w, R(\beta, p_{\text{flip}})] + (1 - R[v, p_{\text{flip}}])R[w, R(F_i, p_{\text{flip}})] \quad (4.38)$$

where $w \equiv x_i(t)$, $v \equiv x_i(t-1)$, and F_i given by (4.37). As a reminder, correlations remain between the observed states of node separated by more than 1 time step. Therefore, the likelihood is only approximately equal to the product of the transition probabilities [Eq. (4.31)].

...

4.15 Ecological networks

The ecological networks are food webs downloaded from the Web of Life database (www.web-of-life.es). A summary of structural properties is presented in Table 4.2. Nodes are species while the edges are observed interactions of predation.

TABLE 4.2 – Summary properties of ecological networks. Average, maximum, and minimum values of structural properties of the 25 food webs we have used. The average local clustering is the average of the fraction of existing over the possible triangles through each node [124].

Network property	Mean	Min	Max
Nodes	76.2	14	249
Edges	487.4	38	3313
Density	0.097	0.0293	0.263
Average local clustering	0.10	0.016	0.358

Conclusion et perspectives

Les changements climatiques et la hausse de l’empreinte humaine nécessitent d’établir les tensions limites que peuvent supporter les systèmes complexes réels, notamment les écosystèmes naturels. D’un point de vue de la physique, cette tâche exige la prédiction des états dynamiques du système couplé en fonction de ses paramètres et de sa structure. Ce travail se révèle toutefois ardu, car les dynamiques se produisent sur des graphes complexes d’interactions, potentiellement de très grande taille. C’est pourquoi il n’existe pas une solution universelle et qu’il faut employer un éventail de méthodes. Certaines devront être exactes, d’autres devront viser des mesures approximatives et globales des bifurcations, et d’autres seront plus pratiques et exigeront d’être basées sur les données observées.

Dès le Chapitre 1, nous avons introduit une sélection d’approches *qualifiant* et *quantifiant* la dynamique d’un réseau : simulations MCMC, passage de messages, formalisme en champ moyen, réduction dimensionnelle, *critical slowing down* et inférence bayésienne. Malgré leurs différences, ces outils peuvent s’utiliser pour étudier les mêmes systèmes. La dynamique SIS se prête à en faire l’exercice. Ce qui favorise l’utilisation d’une approche par rapport à l’autre est le degré d’exactitude et d’explicabilité de leurs conclusions. Motivés par cette dualité et après avoir examiné les limitations des méthodes, nous avons énoncé les objectifs de cette thèse :

1. Développer une méthode très précise de prédiction des états stables pour les structures avec cycles et une dynamique connue ;
2. Valider mathématiquement l’approche de réduction dimensionnelle pour les réseaux modulaires et bipartis ;
3. Développer une nouvelle approche de réduction dimensionnelle plus précise ;
4. Valider l’utilisation de l’apprentissage automatique pour les problèmes où la dynamique est inconnue.

Tous les objectifs ont été atteints.

Au Chapitre 2, nous avons complété le premier objectif en formulant une solution exacte de la dynamique en cascade sur réseau. Le formalisme est composé d’un système fermé de $2^N - 1$ équations et d’un algorithme optimisé pour solutionner ce système. Cet outil bénéficie d’une connaissance exacte de la matrice d’adjacence d’un réseau ainsi que les fonctions d’activation

de chacun des noeuds. Il se compare donc aux méthodes de simulations MCMC et au passage de messages. Notre méthode se distingue en étant la seule qui assure une solution exacte pour tous les réseaux. Toutefois, son potentiel est miné par son impraticabilité à des réseaux de grandes tailles.

En ce qui concerne la résilience, notre approche exacte permet de calculer la probabilité qu'un certain noeud soit actif, ou encore la fraction active de noeuds, au terme d'un processus de cascade en fonction des perturbations structurelles qu'on applique au réseau. Toutefois, on constate que cette démarche est plutôt pédestre, car elle nécessite de résoudre le système exact pour chaque perturbation à documenter. Son utilisation se révèle plus pertinente dans d'autres contextes. Par exemple, au temps de la pandémie de la Covid-19 où plusieurs initiatives cherchent à évaluer les probabilités qu'un individu soit infectieux, cet outil pourrait être utilisé pour mesurer cette probabilité dans de petits groupes d'individus et en fonction des stratégies d'immunisation.

Au Chapitre 3, nous avons suggéré une nouvelle méthode de réduction dimensionnelle faisant usage du spectre de la matrice d'adjacence, en lien direct avec les objectifs 2 et 3. Cette avenue permet entre autres de mieux décrire l'activité moyenne des réseaux bipartis et modulaires que l'approche comparable de Gao *et al.* [52]. Comme son développement mathématique est accessible et instinctif, la porte est ouverte aux améliorations significatives, comme celles de l'Annexe B et des Refs. [109, 132], et aux conclusions aux dynamiques compétitives.

Cette nouvelle réduction peut être exploitée pour l'analyse de la résilience d'un système. D'abord, elle expose les bifurcations dynamiques en faibles dimensions. Ce faisant, elle fournit une mesure de la perturbation maximale qu'un système peut soutenir sans la présence de changements irréversibles. Ensuite, la réduction procure certaines prédictions quantitatives globales, dont les points critiques de réseaux réels [83].

Tout approche qui repose sur la réduction dimensionnelle pourrait être critiquée par son manque de prédictions quantitatives précises sur des systèmes réels. En effet, la méthode n'a pas été développée avec cette tâche en tête. La méthode requiert des modèles dynamiques et des structures connues. Toutefois, son rôle n'en est pas moins essentiel pour une analyse complète de la résilience d'un système.

En effet, les méthodes purement quantitatives nous exposent au risque d'ignorer les effets néfastes d'une petite altération d'un paramètre d'un système complexe. L'approche du Chapitre 3 nous assure d'identifier la nature et la position des bifurcations du système. Nous pouvons nous fier à ses conclusions pour documenter les perturbations maximales et éventuellement en établir des stratégies d'intervention.

Finalement, l'objectif 4 a été étudié au Chapitre 4. Nous avons introduit un processus d'apprentissage automatique pour détecter des perturbations structurelles et dont l'architecture du

modèle repose sur les plus récentes avancées du domaine. Le fonctionnement est simple, mais efficace. En apprenant des séries temporelles de l'activité des noeuds, le modèle détecte les anomalies dynamiques et les associe à des perturbations structurelles. Les principaux résultats montrent que la méthode performe très bien pour une variété de dynamiques et sur différents réseaux.

Notre méthode se compare au *critical slowing down* et à l'inférence bayésienne en ce qui concerne la détection de perturbations basée sur des séries de données. Notre approche se distingue par au moins deux aspects : (i) sa haute performance malgré l'absence d'information injectée sur la nature de la dynamique et (ii) sa grande capacité à détecter de faibles perturbations. Plus encore, notre méthode est en mesure d'inférer dans un régime où les CSD sont absents, soit lorsque le système a peu de risque de faire une transition critique.

Les bons résultats du Chapitre 4 avec l'apprentissage automatique suggèrent de poursuivre la recherche dans cette direction. Par exemple, comment peut-on traiter des dynamiques hétérogènes où chaque noeud a une fonction d'activation différente ? Cette avenue pourrait même s'étendre à identifier des variations paramétriques, telle qu'une hausse de la probabilité d'infection ou des fonctions d'activation individuelles. Une autre perspective consiste à combiner les acquis des chapitres 3 et 4 pour extraire la représentation réduite d'un système à partir des données et de l'apprentissage automatique. Cette approche qu'on nomme *auto-encodage* est courante en apprentissage automatique, mais n'a jamais encore été exploitée à cette fin.

Les travaux de cette thèse contribuent aux aspects théorique et pratique de la résilience. Les deux sont essentiels pour comprendre les rôles des variables d'un système et les traduire en prédictions quantitatives. Le Chapitre 4 témoigne de la grande utilité de l'apprentissage automatique pour des tâches fondées sur des données et il est tentant de redoubler les efforts dans cette direction. Elle risque de se voir attribuer un rôle important dans le développement d'outils numériques du 21e siècle. Néanmoins, les travaux du Chapitre 3 réitèrent la pertinence d'un contexte théorique solide reposant sur des dizaines d'années de recherche en dynamique non linéaire. Ces analyses théoriques nous fournissent une connaissance fondamentale des systèmes qui sera requise lorsqu'il sera temps de développer des stratégies d'intervention. Attendons maintenant de voir la place que l'avenir laissera aux travaux fondamentaux et à l'apprentissage automatique sur l'étude de la résilience des systèmes complexes.

Annexe A

Bifurcation du modèle de Wilson-Cowan

Théorème 1. *Considérons l'équation différentielle*

$$\dot{x} = -x + c \frac{1}{1 + \exp[-\tau(x - \mu)]} \quad (\text{A.1})$$

avec $c > 0$, $\tau > 0$ et $\mu \geq 0$ et x défini sur les réels. L'équation (A.1) possède toujours au moins un point fixe. De plus, les paramètres

$$\alpha = c\tau \quad \text{et} \quad \beta = \mu\tau \quad (\text{A.2})$$

régissent le nombre de points fixes tel que

(I). si $\alpha \leq 4$, alors (A.1) possède un seul point fixe.

(II). si $\alpha > 4$, alors (A.1) possède

- i. un seul point fixe si $\beta_- > \beta$ ou $\beta > \beta_+$;
- ii. deux points fixes si $\beta_- = \beta$ ou $\beta = \beta_+$;
- iii. trois points fixes si $\beta_- < \beta < \beta_+$;

avec

$$\beta_{\pm} = \log \left[\frac{1 \pm \gamma}{1 \mp \gamma} \right] + \frac{\alpha}{2}(1 \pm \gamma) \quad \text{et} \quad \gamma = \sqrt{1 - \frac{4}{\alpha}}. \quad (\text{A.3})$$

Démonstration. Considérons le changement de variables $y = \tau(x - \mu)$, $c\tau = \alpha$ et $\mu\tau = \beta$ qui a pour effet de simplifier l'écriture de l'équation (A.1),

$$\dot{y} = -y - \beta + \alpha \frac{1}{1 + \exp[-y]}, \quad (\text{A.4})$$

$$= -y - \beta + \alpha\sigma(y) \equiv F(y). \quad (\text{A.5})$$

où $\sigma(y)$ est la fonction sigmoïde. Notons que le domaine de $F(y)$ définie en (A.5) est l'ensemble des réels.

À l'équilibre $F(y) = 0$, l'équation (A.5) est

$$y + \beta = \alpha\sigma(y). \quad (\text{A.6})$$

Comme $\alpha\sigma(y)$ est une fonction monotone croissante sur $y \in \mathbb{R}$ et est bornée entre $-\alpha$ et α , et que $y + \beta$ est croissante sur $y \in \mathbb{R}$, alors il existe au moins une solution à l'équation (A.6). Par conséquent, l'équation (A.5) a toujours au moins un point fixe.

Les points fixes de (A.4) sont les zéros de (A.5), i.e., y^* est un point fixe si et seulement si $F(y^*) = 0$. Pour déterminer le nombre de zéros, nous étudierons la dérivée de la fonction $F(y)$ qui est égale à

$$F'(y) = \frac{dF(y)}{dy} = -1 + \alpha\sigma(y)[1 - \sigma(y)]. \quad (\text{A.7})$$

(I) Supposons $\alpha \leq 4$. Par la propriété de la fonction sigmoïde, $\sigma(y)[1 - \sigma(y)] \leq 1/4$ est vraie pour tout y et l'égalité est obtenue à $y = 0$. Par conséquent, $F'(y)$ telle que définie par (A.7) est négative sur tous les réels pour $\alpha \leq 4$. La fonction $F(y)$ est alors décroissante et elle possède exactement un seul zéro. Nous devons donc conclure qu'il y a un seul point fixe à (A.1) pour $c\tau \leq 4$.

(II) Supposons $\alpha > 4$. Nous établirons que $F'(y)$ a exactement 2 zéros et que $F(y)$ a au minimum 1 et au maximum 3 zéros. D'abord, remarquons que (A.7) est une équation quadratique en $\sigma(y)$. Il existe donc au maximum deux zéros y^{**} à l'équation (A.7) et donc un maximum de 3 zéros à (A.5). Les zéros de (A.7) solutionnent

$$F(y^{**}) = 0 = -1 + \alpha\sigma^{**}[1 - \sigma_{\pm}^{**}]; \quad (\text{A.8})$$

$$\sigma_{\pm}^{**} = \frac{1}{2} \left[1 \pm \left(1 - \frac{4}{\alpha} \right)^{1/2} \right]. \quad (\text{A.9})$$

où σ_{\pm}^{**} est une notation abrégée de $\sigma(y_{\pm}^{**})$. Notons que (A.9) admet toujours deux solutions car $\alpha > 4$. Nous introduisons la variable

$$\gamma = \left(1 - \frac{4}{\alpha} \right)^{1/2} \quad (\text{A.10})$$

et noterons les deux solutions de (A.8) par y_{-}^{**} et y_{+}^{**} . Leurs valeurs sont obtenues en inversant l'équation (A.9),

$$\frac{1}{1 + e^{-y_{\pm}^{**}}} = \frac{1}{2} [1 \pm \gamma] \quad (\text{A.11})$$

$$y_{\pm}^{**} = -\log \left[\frac{1 \pm \gamma}{1 \mp \gamma} \right] \quad (\text{A.12})$$

Pour déterminer le nombre de zéros de (A.5), nous devons inspecter les signes de $F(y_{\pm}^{**})$.

Supposons que y_-^{**} ou y_+^{**} soient des zéros de (A.5) et satisfassent

$$F(y_{\pm}^{**}) = -y_{\pm}^{**} - \beta + \alpha\sigma_{\pm}^{**} = 0. \quad (\text{A.13})$$

Alors $\beta = -y_{\pm}^{**} + \alpha\sigma_{\pm}^{**} \equiv \beta_{\pm}$ où

$$\beta_{\pm} = \log \left[\frac{1 \pm \gamma}{1 \mp \gamma} \right] + \frac{\alpha}{2}(1 \pm \gamma). \quad (\text{A.14})$$

Notons que $\beta_- < \beta_+$ est toujours vérifié, car la fonction logarithme et le second terme de (A.14) sont monotones croissants.

Si $\beta = \beta_{\pm}$ est respectée, alors un des zéros de $F'(y)$ coïncide avec un zéro de $F(y)$ et il existe exactement 2 zéros à $F(y)$ pour $\alpha > 4$. Notons que β ne peut jamais être simultanément égal à β_- et β_+ , car cela requièrerait que $\beta_+ = \beta_-$, qui est impossible, car $\beta_- < \beta_+$ est toujours vérifié.

L'équation $F(y)$ admettra trois points fixes lorsque $F(y_-^{**})$ et $F(y_+^{**})$ auront des signes inversés. Cette condition survient si et seulement si

$$\beta_- < \beta < \beta_+. \quad (\text{A.15})$$

L'équation $F(y)$ admettra un seul point fixe lorsque $F(y_-^{**})$ et $F(y_+^{**})$ auront le même signe et seront différents de zéros. Comme $\beta_- < \beta_+$, alors il ne peut qu'y avoir les cas où

$$\beta < \beta_- \quad \text{et} \quad \beta > \beta_+. \quad (\text{A.16})$$

□

Annexe B

Deux généralisations de la réduction dimensionnelle à plus hautes dimensions

Au Chapitre 3, nous avons introduit un modèle de réduction dimensionnelle basée sur les vecteurs propres dominants de la matrice des poids \mathbf{W} . Cette analyse a mené à deux types de systèmes réduits : le premier est unidimensionnel et le second est bidimensionnel, car il s'applique aux réseaux bipartis. Pour modéliser les réseaux modulaires, nous avons suggéré d'utiliser autant d'équations *indépendantes* qu'il a de modules dominants. Toutefois, cette approche présente plusieurs imperfections. Premièrement, elle favorise l'utilisation de vecteurs propres sous-dominants, lesquels contiennent des éléments négatifs. Le problème pouvant alors survenir est présenté dans le commentaire de la Ref. [10], il est essentiellement reproduit à la sous-section 1.4.2. La stabilité du système réduit s'éloigne de la stabilité du système original, rendant les prédictions moins fiables. Plus encore, l'interprétation des vecteurs de poids comme des vecteurs de probabilité devient invalide pour les éléments négatifs. Deuxièmement, il semble plus naturel qu'un système de n modules couplés ensemble soit décrit par n équations *couplées*, plutôt que n équations indépendantes. Tout ceci suggère un modèle à plus hautes dimensions. Dans les prochaines sections, nous introduirons deux de ces modèles.

...

B.1 Contexte

Reprenons le contexte introduit au Chapitre 3 où un réseau à N noeuds est décrit par une matrice d'adjacence pondérée \mathbf{W} . La dynamique est donnée par un système de N équations,

$$\dot{x}_i = F(x_i) + \sum_{j=1}^N w_{ij} G(x_i, x_j). \quad (\text{B.1})$$

Dans les prochaines sections, nous présenterons deux réductions dimensionnelles de $n \geq 1$ dimensions (B.1).

B.1.1 Méthode sans contrainte

On suppose un ensemble de n vecteurs de poids qu'on place dans la matrice $\mathbf{V} \in \mathbb{R}^{n \times N}$. L'élément v_{ki} est le poids du noeud $i \in \{1, \dots, N\}$ dans le groupe $k \in \{1, \dots, n\}$. Sans perte de généralité, on supposera que chaque vecteur de poids est normalisé selon la règle suivante :

$$\sum_{i=1}^N v_{ki} = 1 \quad \forall k \in \{1, \dots, n\}. \quad (\text{B.2})$$

Les vecteurs de poids permettent de mesurer les observables dynamiques,

$$y_k = \sum_{i=1}^N v_{ki} x_i \quad (\mathbf{y} = \mathbf{V} \mathbf{x}). \quad (\text{B.3})$$

Ce sont des moyennes pondérées de l'activité sur le réseau. On cherche maintenant à établir les équations dynamiques de y_k ainsi qu'une forme contrainte des vecteurs de poids.

En dérivant l'équation (B.3) par rapport au temps, on obtient que la dynamique de y_k est donnée par les équations

$$\dot{y}_k \approx F(y_k) + \sum_{i=1}^N \sum_{j=1}^N v_{ki} w_{ij} G(x_i, x_j). \quad (\text{B.4})$$

En effet, il a déjà été démontré au Chapitre 3 que $\sum_i v_{ki} F(x_i) \approx F(y_k)$. Pour le deuxième terme de droite, on décompose $G(x_i, x_j)$ autour de (u_k, t_k) :

$$G(x_i, x_j) \approx G(u_k, t_k) + (x_i - u_k) G_i + (x_j - t_k) G_j. \quad (\text{B.5})$$

où G_i et G_j sont respectivement les dérivées partielles de $G(x_i, x_j)$ par rapport à x_i et x_j , évaluées en (u_k, t_k) . En incluant la somme de droite de l'équation (B.4), on élimine les termes linéaires en (x_i, x_j) de l'équation (B.5) si, d'une part,

$$\sum_{i=1}^N \sum_{j=1}^N v_{ki} w_{ij} x_i = \sum_{i=1}^N \sum_{j=1}^N v_{ki} w_{ij} u_k \quad (\text{B.6a})$$

et, d'autre part,

$$\sum_{i=1}^N \sum_{j=1}^N v_{ki} w_{ij} x_j = \sum_{i=1}^N \sum_{j=1}^N v_{ki} w_{ij} t_k. \quad (\text{B.6b})$$

Sous forme matricielle, ces équations sont

$$\mathbf{V} \mathbf{S} \mathbf{x} = (\mathbf{V} \mathbf{W} \mathbf{1}) \circ \mathbf{u}, \quad (\text{B.7a})$$

$$\mathbf{V} \mathbf{W} \mathbf{x} = (\mathbf{V} \mathbf{W} \mathbf{1}) \circ \mathbf{t}, \quad (\text{B.7b})$$

avec $\mathbf{S} = \text{diag}(\mathbf{W} \mathbf{1})$ une matrice diagonale dont les éléments sur la diagonale sont $S_{ii} = [\mathbf{W} \mathbf{1}]_{ii}$. Ces dernières équations sont équivalentes à la forme

$$\mathbf{D} \mathbf{V} \mathbf{S} \mathbf{x} = \mathbf{u}, \quad (\text{B.8a})$$

$$\mathbf{D} \mathbf{V} \mathbf{W} \mathbf{x} = \mathbf{t}, \quad (\text{B.8b})$$

où $\mathbf{D} = \text{diag}([1/\alpha_1, 1/\alpha_2, \dots, 1/\alpha_n])$ avec

$$\boldsymbol{\alpha} = \mathbf{V} \mathbf{W} \mathbf{1}. \quad (\text{B.9})$$

Si les conditions des équations (B.8) sont respectées, alors l'équation (B.4) est

$$\dot{\mathbf{y}} \approx F(\mathbf{y}) + \boldsymbol{\alpha} \circ G(\mathbf{u}, \mathbf{t}) \quad (\text{B.10})$$

où $F(\mathbf{y})$ et $G(\mathbf{u}, \mathbf{t})$ désignent des vecteurs colonnes de composantes $[F(\mathbf{y})]_k = F(y_k)$ et $[G(\mathbf{u}, \mathbf{t})]_k = G(u_k, t_k)$.

On vérifie comment les conditions des Eqs. (B.8) peuvent être respectées. On pose que u_k et t_k sont des combinaisons linéaires des variables réduites :

$$\mathbf{u} = \boldsymbol{\beta} \mathbf{y} = \boldsymbol{\beta} \mathbf{V} \mathbf{x}, \quad (\text{B.11})$$

$$\mathbf{t} = \boldsymbol{\theta} \mathbf{y} = \boldsymbol{\theta} \mathbf{V} \mathbf{x}. \quad (\text{B.12})$$

En plaçant ces expressions dans le système (B.8), on obtient que les matrices $\boldsymbol{\beta}$ et $\boldsymbol{\theta}$ doivent être

$$\boldsymbol{\beta} = \mathbf{D} \mathbf{V} \mathbf{S} \mathbf{V}^+, \quad (\text{B.13})$$

$$\boldsymbol{\theta} = \mathbf{D} \mathbf{V} \mathbf{W} \mathbf{V}^+, \quad (\text{B.14})$$

où \mathbf{V}^+ est le pseudo-inverse de la matrice \mathbf{V} et $\mathbf{D} = \text{diag}([1/\alpha_1, 1/\alpha_2, \dots, 1/\alpha_n])$.

Ces expressions concluent la réduction pour un ensemble de poids \mathbf{V} . De manière explicite, on a le système à n dimensions

$$\dot{\mathbf{y}} \approx F(\mathbf{y}) + (\mathbf{V} \mathbf{W} \mathbf{1}) \circ G(\mathbf{D} \mathbf{V} \mathbf{S} \mathbf{V}^+ \mathbf{y}, \mathbf{D} \mathbf{V} \mathbf{W} \mathbf{V}^+ \mathbf{y}). \quad (\text{B.15})$$

En supposant que la matrice de poids \mathbf{V} est déterminée, on construit le système (B.15) avec pour n observables les variables dynamiques données à l'équation (B.3).

Contrairement au formalisme du Chapitre 3, cette réduction n'impose aucune contrainte quant à la forme de la matrice de poids \mathbf{V} . En théorie, les poids peuvent être réels et chaque vecteur de poids doit être normalisé. Par contre, en pratique, il arrive souvent qu'un mauvais choix des vecteurs de poids entraîne une divergence des observables. La stabilité du système (B.15) est donc grandement affectée par le choix des vecteurs de poids.

Pour la k -ième observable, le système (B.15) admet la forme

$$\dot{y}_k = F(y_k) + c_k G(s_k, t_k) \quad (\text{B.16})$$

où c_k, s_k, t_k sont utilisés pour simplifier la notation. La caractéristique principale de cette forme est que la fonction G n'apparaît qu'une seule fois. Par conséquent, l'observable y_k est de manière effective couplée à un noeud unique d'activité t_k . Est-ce vraiment la forme du couplage auquel nous nous attendions? Prenons par exemple un réseau modulaire à n modules. Si chaque observable y_k représente un module distinct et dont le couplage attendu avec une observable y_j devrait être proportionnel au couplage observé entre les noeuds de ces deux modules (k, j) . Nous aurions alors un système qui satisfait la forme suivante :

$$\dot{y}_k = F(y_k) + \sum_{j=1}^n c_{kj} G(y_k, y_j). \quad (\text{B.17})$$

B.1.2 Méthode par approximation d'orthogonalité ou par partition

Dans le formalisme multidimensionnel sans contrainte, la structure du couplage entre les variables dynamiques impose une forme très particulière. Dans cette section, nous proposons un nouveau formalisme plus contraint, mais qui permet un couplage complet entre les variables dynamiques.

On considère une matrice de poids \mathbf{V} qui construit un ensemble de n variables réduites $\mathbf{y} = \mathbf{V}\mathbf{x}$. La dynamique de y_k obéit aux équations

$$\dot{y}_k \approx F(y_k) + \sum_{i=1}^N \sum_{j=1}^N v_{ki} w_{ij} G(x_i, x_j). \quad (\text{B.18})$$

On impose une contrainte d'exclusivité sur les vecteurs de poids, soit

$$\forall i \text{ appartenant au groupe } k : v_{ri} \approx 0 \quad \forall r \neq k. \quad (\text{B.19})$$

Cette contrainte se traduit par une partition disjointe des noeuds, donc chaque noeud appartient à un et un seul groupe. La contrainte d'exclusivité assure que deux vecteurs de poids distincts ne peuvent placer simultanément un poids non nul sur un même noeud i . On construit une matrice indicatrice $\tilde{\mathbf{V}}$ dont les éléments sont $\tilde{v}_{lj} = v_{lj}/\mu_j$ avec

$$\boldsymbol{\mu} = \mathbf{1}^T \mathbf{V}. \quad (\text{B.20})$$

En pratique, μ_j ne sera égale qu'à v_{lj} lorsque le noeud j appartiendra au groupe l . Si la contrainte d'exclusivité est respectée, alors $\tilde{v}_{lj} = 1$ seulement si le poids du noeud j est placé dans le groupe l et 0 autrement. La matrice $\tilde{\mathbf{V}}$ est une forme de matrice indicatrice de l'adhésion des noeuds aux communautés décrites par les n vecteurs de poids.

On note également que, par construction, la matrice $\tilde{\mathbf{V}}$ respecte

$$\mathbf{1}^T \tilde{\mathbf{V}} = \mathbf{1}. \quad (\text{B.21})$$

On procède aux mêmes opérations que précédemment pour obtenir la dynamique de \mathbf{y} ,

$$\dot{y}_k \approx F(y_k) + \sum_{i=1}^N \sum_{j=1}^N v_{ki} w_{ij} G(x_i, x_j). \quad (\text{B.22})$$

Cette fois-ci, on développe $v_{ki}G(x_i, x_j)$ autour de $x_i = u_k$ et $x_j = t_l$. Toutefois, on somme sur tous les $l \in \{1, \dots, n\}$ et on pondère par la fonction indicatrice

$$v_{ki}G(x_i, x_j) \approx v_{ki} \sum_{l=1}^n \tilde{v}_{lj} [G(u_k, t_l) + (x_i - u_k)G_i + (x_j - t_l)G_j]. \quad (\text{B.23})$$

où G_i et G_j sont respectivement les dérivées partielles de $G(x_i, x_j)$ par rapport à x_i et x_j et évaluées à (u_k, t_l) . Malgré la somme sur l'indice l , la décomposition de $G(x_i, x_j)$ n'est en réalité qu'un seul terme non nul, celui où l est tel que $\tilde{v}_{lj} = 1$.

Les termes linéaires de l'équation (B.23) sont nuls si

$$\mathbf{V} \mathbf{S} \mathbf{x} = (\mathbf{V} \mathbf{W} \mathbf{1}) \circ \mathbf{u} \quad (\text{B.24})$$

$$\mathbf{V} \mathbf{W} \mathbf{x} = \mathbf{V} \mathbf{W} \tilde{\mathbf{V}}^T \mathbf{t} \quad (\text{B.25})$$

où la relation (B.21) est utilisée. Si ces conditions sont respectées, alors la dynamique de y_k est

$$\dot{y}_k \approx F(y_k) + \sum_{l=1}^n \gamma_{kl} G(u_k, t_l) \quad (\text{B.26})$$

avec la matrice de couplage

$$\boldsymbol{\gamma} = \mathbf{V} \mathbf{W} \tilde{\mathbf{V}}^T. \quad (\text{B.27})$$

On pose maintenant que \mathbf{u} et \mathbf{t} sont des combinaisons linéaires des variables réduites

$$\mathbf{u} = \beta \mathbf{y} = \beta \mathbf{V} \mathbf{x}, \quad (\text{B.28})$$

$$\mathbf{t} = \theta \mathbf{y} = \theta \mathbf{V} \mathbf{x}. \quad (\text{B.29})$$

On solutionne donc Eq. (B.25) par

$$\beta = \mathbf{D} \mathbf{V} \mathbf{S} \mathbf{V}^+ \quad (\text{B.30})$$

$$\theta = (\mathbf{V} \mathbf{W} \tilde{\mathbf{V}}^T)^{-1} \mathbf{V} \mathbf{W} \mathbf{V}^+ \quad (\text{B.31})$$

où $\mathbf{D} = \text{diag}([1/\alpha_1, 1/\alpha_2, \dots, 1/\alpha_n])$ avec

$$\boldsymbol{\alpha} = \mathbf{V} \mathbf{W} \mathbf{1}. \quad (\text{B.32})$$

Nous avons toutes les pièces en main pour construire un système réduit à partir de la matrice \mathbf{W} et \mathbf{V} .

La matrice γ a une importance particulière, car elle est la source du couplage entre toutes les observables. Elle représente un réseau réduit de matrice d'adjacence $\gamma = \mathbf{V} \mathbf{W} \tilde{\mathbf{V}}^T$. De par la forme de cette construction, un esprit averti identifiera la matrice caractéristique \mathbf{V} et la matrice des quotients γ , des termes empruntés de la théorie spectrale des graphes [30, 129]. Ces termes sont propres au problème de partitionnement spectral des graphes. En effet, il faut choisir une représentation \mathbf{V} qui divise, de manière pondérée, les noeuds du réseau dans des groupes exclusifs pour ainsi respecter notre contrainte d'exclusivité. Comme il est connu que plusieurs algorithmes de partitionnement sont basés sur des arguments spectraux [122], il semble probable que \mathbf{V} puisse être construite par combinaisons des vecteurs propres dominants de la matrice \mathbf{W} .

On associera également le problème de choisir les vecteurs de poids à la *factorisation matricielle* en apprentissage automatique [98]. Dans sa forme originale, il s'agit de trouver une décomposition à la matrice $\mathbf{W} \in \mathbb{R}^{N \times N}$,

$$\mathbf{W} \approx \mathbf{U} \mathbf{H} \quad (\text{B.33})$$

où $\mathbf{U} \in \mathbb{R}^{N \times n}$ et $\mathbf{H} \in \mathbb{R}^{n \times N}$. Les rangées de la matrice \mathbf{U} sont souvent utilisées comme représentation vectorielle des noeuds. Notons que l'égalité à l'équation (B.33) peut être obtenue si le rang de la matrice \mathbf{W} est supérieur ou égal à n . Pour obtenir \mathbf{U} et \mathbf{H} , il a été proposé d'optimiser une simple fonction de coût de la reconstruction de la matrice \mathbf{W} [67]. Une autre méthode profite de la décomposition en valeurs singulières de la matrice \mathbf{W} pour construire les matrices \mathbf{U} et \mathbf{H} . La factorisation matricielle a rapidement été popularisée, notamment grâce à son utilisation pour la victoire au concours de recommandation de Netflix en 2005¹.

1. Notons que maintenant l'utilisation des réseaux neuronaux sur graphes, comme ceux utilisés au Chapitre 4, est priorisée en intelligence artificielle pour construire une représentation des noeuds [175].

Dans le cas qui nous intéresse, la matrice \mathbf{W} doit être factorisée sous la forme

$$\mathbf{W} \approx \mathbf{V}^T \mathbf{H} \quad (\text{B.34})$$

où \mathbf{V} est une matrice non négative des poids semi-orthogonale, i.e., $\mathbf{V}\mathbf{V}^T = \mathbf{I}$. Une revue des algorithmes permettant de satisfaire approximativement ces contraintes est présentée à la Ref. [100].

À l'état actuel du projet, la forme exacte de la construction de \mathbf{V} est encore inconnue. Cette direction de recherche est laissée ouverte.

Il est intéressant de constater que $\boldsymbol{\theta}$ et $\boldsymbol{\beta}$ sont tous les deux invariants sous une transformation d'échelle de \mathbf{W} . En effet, une transformation $\mathbf{W} \mapsto c\mathbf{W}$ n'aura aucun impact sur la valeur de leurs éléments. Par ailleurs, selon le système (B.26), le rôle de $\boldsymbol{\theta}$ et $\boldsymbol{\beta}$ est de corriger les interactions entre les observables directement via la fonction $G(\cdot, \cdot)$.

Par construction, $\boldsymbol{\beta}$ est une matrice diagonale, mais $\boldsymbol{\theta}$ ne l'est pas forcément. Il semble même que les éléments non diagonaux de $\boldsymbol{\theta}$ témoignent de son rôle correctif. Ainsi, la recherche de la matrice \mathbf{V} pourrait s'effectuer par la minimisation de la distance entre $\boldsymbol{\theta}$ et la matrice identité.

...

B.2 Système à double coeur

Nous avons comparé les deux méthodes de réduction sur un réseau formé de deux étoiles connectées (Fig. B.1). Ce réseau a quatre valeurs propres non nulles. Comme il est biparti, les valeurs propres sont organisées en couple de valeurs opposées.

Un ensemble de vecteurs de poids est construit en combinant les quatre vecteurs propres de manière à ce que les deux groupes des noeuds en périphéries et les deux coeurs forment quatre groupes distincts. En considérant que les noeuds 1 et $N_1 + 2$ sont les coeurs du réseau, alors les vecteurs de poids ont la forme

$$\mathbf{v}_1^T = [1, 0, \dots, 0], \quad (\text{B.35a})$$

$$\mathbf{v}_2^T = [0, \dots, 0, 1, 0, 0, \dots, 0], \quad (\text{B.35b})$$

$$\mathbf{v}_3^T = [0, N_1^{-1}, N_1^{-1}, \dots, N_1^{-1}, 0, 0, \dots, 0], \quad (\text{B.35c})$$

$$\mathbf{v}_4^T = [0, \dots, 0, N_2^{-1}, N_2^{-1}, \dots, N_2^{-1}], \quad (\text{B.35d})$$

avec N_1 et N_2 désignant le nombre de noeuds dans chaque périphérie des deux coeurs.

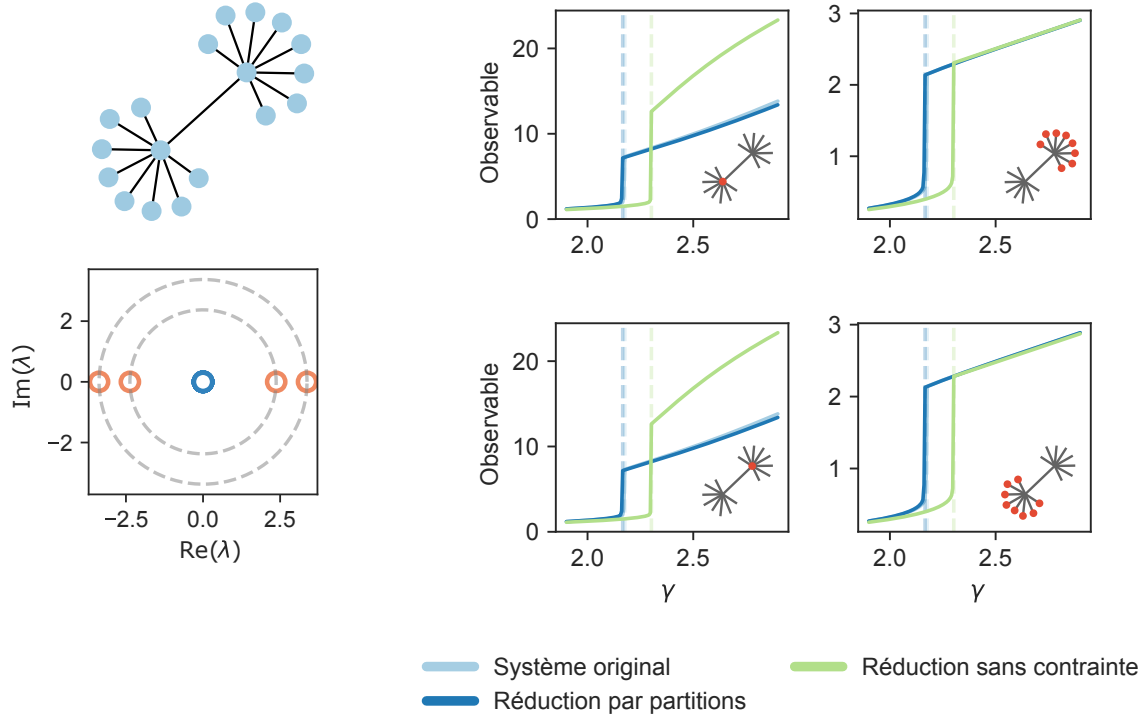


FIGURE B.1 – En haut à gauche, schématisation du réseau à double coeur [$N = 18$, $N_1 = N_2 = 8$]. En bas à gauche, spectre du réseau à double coeur. On observe quatre valeurs propres dominantes et 14 valeurs propres nulles. À droite, activités moyennes y_k^* pour les quatre vecteurs de poids utilisés [Eq. (B.35)]. Pour chaque vecteur de poids, une schématisation de la partition est donnée où la taille des noeuds est proportionnelle à leur pondération au vecteur de poids. La courbe en bleu pâle se trouve sous la courbe en bleu foncé. Les lignes pointillées suggèrent l’emplacement de la transition critique prédit par les différentes approches. Le terme γ est le facteur d’échelle par lequel la matrice binaire du réseau est multipliée, c.-à-d. $\mathbf{W}' = \gamma \mathbf{A}$, permettant d’échantillonner un même réseau avec un poids moyen variable. La dynamique utilisée est celle du modèle de Wilson-Cowan.

Les vecteurs \mathbf{v}_1^T et \mathbf{v}_2^T décrivent les coeurs du réseau alors que les vecteurs de poids \mathbf{v}_3^T et \mathbf{v}_4^T décrivent les deux périphéries. Les éléments de la matrice de poids sont $v_{ki} = [\mathbf{v}_k^T]_i$ avec le vecteur \mathbf{v}_k^T donné par (B.35).

En suivant le formalisme avec contrainte d’orthogonalité, on obtient le système réduit suivant :

$$\dot{y}_1 = F(y_1) + N_1 G(y_1, y_3) + G(y_1, y_3), \quad (\text{B.36a})$$

$$\dot{y}_2 = F(y_2) + N_2 G(y_2, y_4) + G(y_2, y_1), \quad (\text{B.36b})$$

$$\dot{y}_3 = F(y_3) + G(y_3, y_1), \quad (\text{B.36c})$$

$$\dot{y}_4 = F(y_4) + G(y_4, y_1), \quad (\text{B.36d})$$

avec $y_k = \sum_i v_{ki} x_i$. Le système réduit résout *exactement* la dynamique au premier ordre. Nous avons donc une partition idéale du réseau.

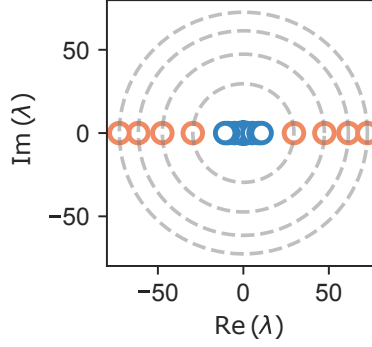


FIGURE B.2 – Spectre du réseau de Barrett, *et al.* [16]. Les valeurs propres dominantes sont coloriées en orange.

En utilisant les mêmes poids \mathbf{V} , la méthode sans contrainte donne une dynamique réduite de la forme

$$\dot{y}_1 = F(y_1) + \alpha_1 G(s_1, t_1), \quad (\text{B.37a})$$

$$\dot{y}_2 = F(y_2) + \alpha_2 G(s_2, t_2), \quad (\text{B.37b})$$

$$\dot{y}_3 = F(y_3) + \alpha_3 G(s_3, t_3), \quad (\text{B.37c})$$

$$\dot{y}_4 = F(y_4) + \alpha_4 G(s_4, t_4). \quad (\text{B.37d})$$

La différence est notable dans la forme du couplage entre les variables dynamiques.

Numériquement, pour un petit système à $N = 18$ noeuds, le système avec contrainte arrive à prédire nettement mieux les observables que la dynamique sans contrainte (Fig. B.1).

...

B.3 Simulations sur un système réel

Nous avons exploré la pertinence de la méthode avec contrainte sur un réseau réel. Parmi les 70 réseaux réels de la banque de données de `www.web-of-life.es` [1], nous avons sélectionné un réseau réel ($N = 112$ et $M = 166$) de plantes et de pollinisateurs dont le spectre présente des valeurs propres sous-dominantes [16] [Fig. B.3]. Ces valeurs propres indiquent la présence de plusieurs modules, d'où l'intérêt d'utiliser une méthode à plusieurs dimensions [Fig. B.2].

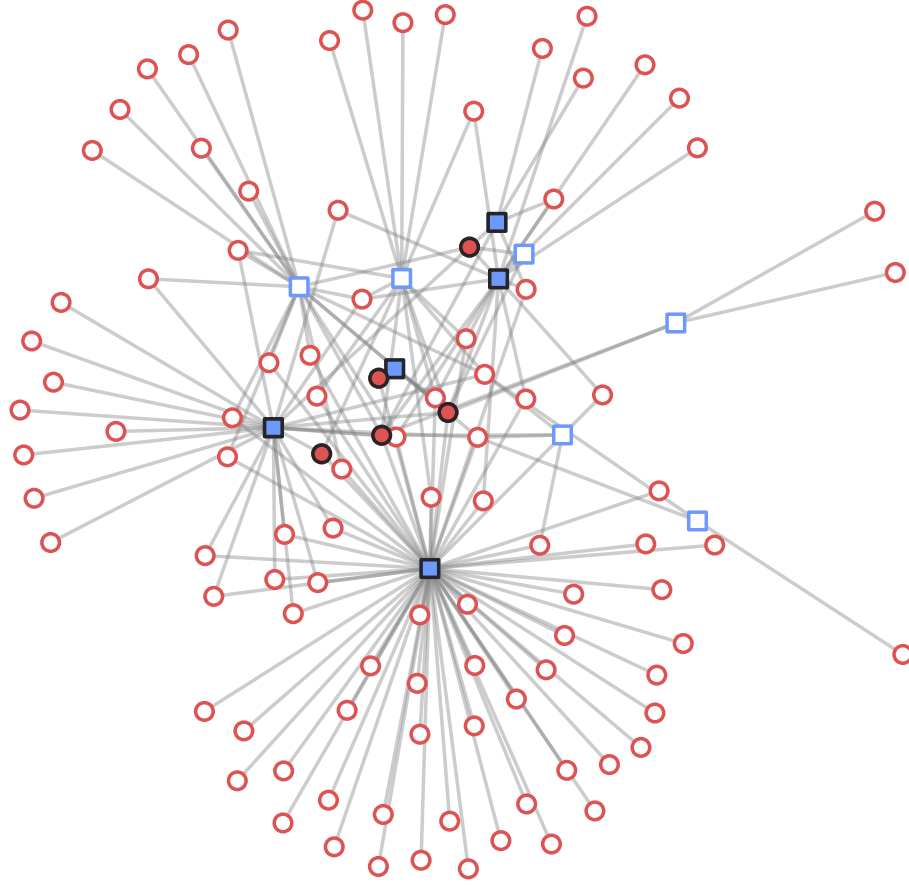


FIGURE B.3 – Schématisation du réseau de Barrett, *et al.* [16] avec $N = 112$ et $M = 166$. Les carrés sont réservés aux espèces des plantes et les cercles aux pollinisateurs. Les noeuds avec les contours foncés sont les noeuds ayant des contributions supérieures à 0.1 pour la réalisation de la figure B.4(e).

B.3.1 Dynamique mutualiste

La dynamique mutualiste suivante est simulée :

$$\dot{x}_i = a_i x_i - \sum_{j=1}^N x_i b_{ij} x_j + \sum_{j=1}^N x_i \frac{w_{ij} x_j}{1 + c x_i + d x_j} + e, \quad (\text{B.38})$$

qui est inspirée des Refs.[40, 52, 96]. Les éléments de la matrice \mathbf{B} représentent la compétition entre les noeuds d'une même espèce (plante-plante et pollinisateur-pollinisateur), alors que les éléments de la matrice \mathbf{W} mesurent les interactions pollinisateur-plante. Une description de la dynamique est spécifiée à la section 1.2.4. On fixe les paramètres $c = 0.5$ et $d = 0.1$. Le vecteur \mathbf{a} assure la décroissance naturelle d'une population et est un échantillon d'une distribution uniforme entre -1 et -5. Enfin, le paramètre $e = 10^{-5}$ fournit une croissance constante des

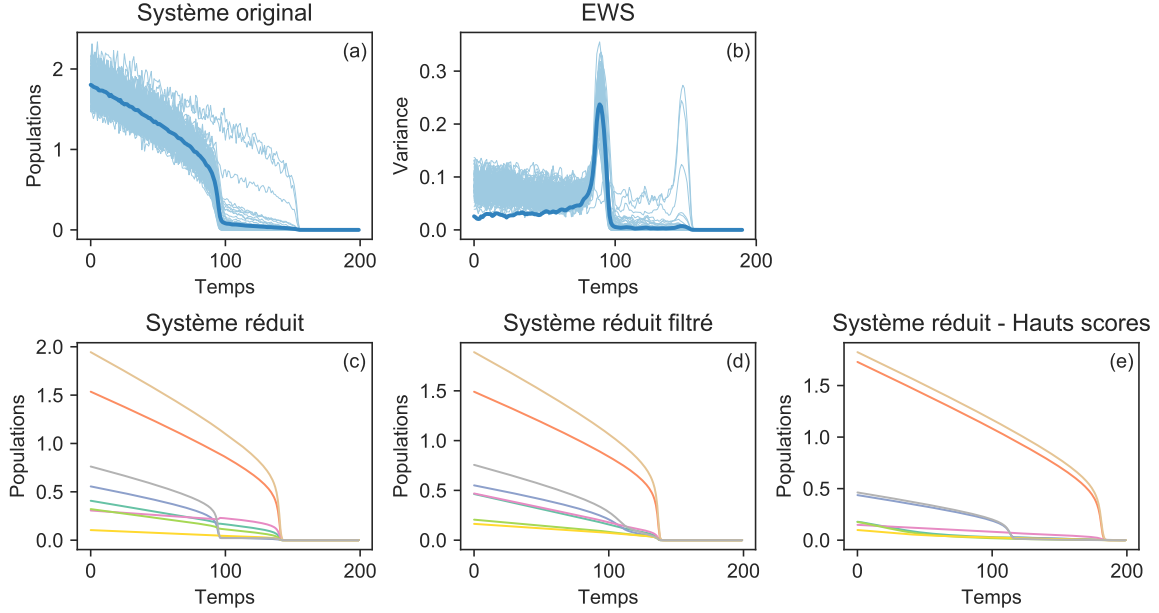


FIGURE B.4 – Simulation d’une dynamique mutualiste sur le réseau de Barrett, *et al.* [16]. (a) Population de chaque espèce (pâle) et population moyenne (foncé) en fonction du temps. (b) Variance d’une fenêtre mobile de taille $\tau = 10$ pour chaque espèce (pâle) et de la population moyenne (foncé) en fonction du temps. (c) Solution du système réduit de 8 dimensions en fonction du temps pour chaque variable réduite. Chaque variable effective est représentée par une courbe. (d) Même système que (c), mais les éléments négatifs des vecteurs de poids sont mis à zéro, c.-à-d. que $\mathbf{V} = \max(0, \mathbf{V})$. (e) Tous les éléments des vecteurs de poids qui sont inférieurs à 0.1 sont ramenés à zéro. Seuls 10 noeuds ont des poids non nuls (voir la Fig. B.3 pour une identification de ces noeuds). À chaque temps, la transformation $\mathbf{W}(t) = \gamma(t)\mathbf{A}$ est appliquée avec $\gamma(t) = \gamma_0 - t\Delta\gamma$, où $\Delta\gamma = 1/40$ et $\gamma_0 = 12$.

populations.

La simulation de la dynamique suit la procédure de la Ref. [40]. On initialise d’abord les paramètres \mathbf{a}, \mathbf{b} , et $\mathbf{W} = \gamma_0\mathbf{A}$, où \mathbf{A} est donnée par la structure non pondérée initiale du réseau. On amène les populations à l’équilibre dynamique. Le choix initial γ_0 vérifie que l’activité moyenne sur le réseau soit grande. Ensuite et sans quelconque réinitialisation, on varie l’échelle via $\gamma(t) = \gamma_0 - t\Delta\gamma$ par 200 étapes avec $\Delta\gamma = (\gamma_0 - \gamma_1)/200$. Dans le cas en question, on utilise $\gamma_0 = 12$, $\gamma_1 = 8$. À chaque modification de $\Delta\gamma$, on laisse la dynamique évoluée pendant $\Delta t = 1$, et on ajoute ensuite au vecteur de population un bruit stochastique $d\mathbf{x} \in \mathbb{R}^N$ de moyenne 0 et de variance σ .

Au terme de la simulation, une série temporelle $\mathbf{X} \in \mathbb{R}^{N \times T}$ bruitée est obtenue sur un réseau dont le poids moyen de la structure diminue temporellement. L’objectif est de vérifier si plusieurs transitions de l’activité sont visibles ainsi que si elles peuvent être décrites via le formalisme de faible dimension.

Un exemple de simulations est présenté à la Fig. B.4. On construit un système dynamique

réduit représentatif du réseau en question. Pour ce faire, on utilise les 8 vecteurs propres ayant les valeurs propres de plus grand module pour obtenir 8 vecteurs de poids par combinaisons linéaires (voir Fig. B.2 pour les valeurs propres utilisées). Grâce à une astuce simple, on construit deux ensembles de 4 vecteurs dont chaque ensemble n'a de poids que sur les plantes ou les pollinisateurs (voir 3.12). Ensuite, le système réduit avec contrainte (B.26) est résolu en fonction du temps pour $\mathbf{W}(t) = \gamma(t)\mathbf{A}$. La méthode sans contrainte n'est pas présentée, car le système réduit résultant est fortement instable, c.-à-d. que les observables divergent rapidement et l'intégration numérique échoue.

On compare la méthode par réduction dimensionnelle à la détection de *critical slowing down* (EWS). Pour cette dernière, on mesure les variances des populations dans une fenêtre mobile de taille τ . Près des points critiques, la variance doit atteindre un maximum.

La réduction parvient à détecter les transitions multiples. Comme le modèle réduit est de 8 dimensions, le réseau n'a pas un point critique unique. En effet, certaines espèces se dirigent plus rapidement à l'extinction que d'autres. Le système réduit expose grossièrement ces transitions multiples. L'exemple montre que la variance est également un bon prédicteur d'une transition critique. Une complication de cette approche est que l'augmentation significative de la variance survient à quelques pas $\Delta t \approx 10$ de la transition critique. Ainsi, la détection ne laisse qu'une très courte fenêtre de temps pour intervenir et rétablir un équilibre dynamique. Également, la détection des multiples transitions ne peut se faire qu'en mesurant la variance de l'activité de chaque noeud. En effet, la variance de l'activité moyenne est un faible indicateur de la double transition.

La présence de plusieurs trajectoires suggère des transitions locales où certains noeuds préservent plus longtemps un bon niveau d'activité. Par conséquent, certains noeuds sont plus importants pour prédire les transitions locales que d'autres. En effet, en filtrant les poids de la réduction pour ne laisser que les contributions dominantes [Fig. B.4(d-e)], on obtient que les transitions sont encore bien décrites. À la Fig. B.4(e), le filtrage est d'une telle importance que les vecteurs de poids ont des contributions non nulles pour une dizaine de noeuds, c.-à-d. que chaque groupe est représenté par un ou deux noeuds. Évidemment, on perd une certaine précision dans les trajectoires des observables et de la position des transitions critiques. Malgré tout, le système prédit encore 2 transitions distinctes.

Ces observations montrent que les vecteurs de poids utilisés placent pratiquement tout le poids de l'observation sur quelques noeuds. La majorité des noeuds ne contribuent pratiquement pas au système réduit. On en conclut donc que la dynamique est dominée par ces noeuds et que la réduction offre une avenue intéressante pour les identifier.

Cette preuve de concept sur un réseau réel est un exemple de la pertinence d'une approche de réduction dimensionnelle. Elle permet de discriminer les noeuds les plus importants à la dynamique et fournit une approximation des points critiques d'un système. Des analyses

quantitatives seront toutefois nécessaires pour approfondir la précision de cette méthode.

Bibliographie

- [1] *The Web of Life*. <http://www.web-of-life.es>. Visité le 2019-03-26.
- [2] ADLER J. ET AHARONY A., *Diffusion percolation. I. Infinite time limit and bootstrap percolation*, J. Phys. A : Math. Gen., 21 (1988), p. 1387.
- [3] AGHABOZORGI S., SHIRKHORSHIDI A. S. ET WAH T. Y., *Time-series clustering—A decade review*, Inf. Syst., 53 (2015), p. 16–38.
- [4] ALLARD A., ALTHOUSE B. M., HÉBERT-DUFRESNE L. ET SCARPINO S. V., *The risk of sustained sexual transmission of Zika is underestimated*, PLOS Pathog., 13 (2017), p. 1006633.
- [5] ALLARD A., HÉBERT-DUFRESNE L., NOËL P.-A., MARCEAU V. ET DUBÉ L. J., *Bond percolation on a class of correlated and clustered random graphs*, J. Phys. A, 45 (2012), p. 405005.
- [6] ———, *Exact solution of bond percolation on small arbitrary graphs*, Europhys. Lett., 98 (2012), p. 16001.
- [7] ALLARD A., HÉBERT-DUFRESNE L., YOUNG J.-G. ET DUBÉ L. J., *A general and exact approach to percolation on random graphs*, Phys. Rev. E., 92 (2015), p. 062807.
- [8] ALON U., *An Introduction to Systems Biology : Design Principles of Biological Circuits*, CRC press, 2006.
- [9] ANGELINI R. ET VAZ-VELHO F., *Ecosystem structure and trophic analysis of Angolan fishery landings*, Sci. Mar., 75 (2011), p. 309–319.
- [10] ARNOLDI J.-F., HAEGEMAN B., REVILLA T. ET LOREAU M., *Particularity of “universal resilience patterns in complex networks”*, bioRxiv :056218, (2016).
- [11] AZAELE S., MARITAN A., BERTUZZO E., RODRIGUEZ-ITURBE I. ET RINALDO A., *Stochastic dynamics of cholera epidemics*, Phys. Rev. E, 81 (2010), p. 051901.

- [12] BANSAL S., GRENFELL B. T. ET MEYERS L. A., *When individual behaviour matters : homogeneous and network models in epidemiology*, J. R. Soc. Interface, 4 (2007), p. 879–891.
- [13] BARABÁSI A.-L., *Scale-free networks : A decade and beyond*, Science, 325 (2009), p. 412–413.
- [14] BARABÁSI A.-L. ET ALBERT R., *Emergence of scaling in random networks*, Science, 286 (1999), p. 509–512.
- [15] BARRAT A., BARTHELEMY M. ET VESPIGNANI A., *Dynamical Processes on Complex Networks*, Cambridge University Press, 2008.
- [16] BARRETT S. C. ET HELENURM K., *The reproductive biology of boreal forest herbs. I. Breeding systems and pollination*, Can. J. Bot., 65 (1987), p. 2036–2046.
- [17] BARZEL B. ET BARABÁSI A.-L., *Network link prediction by global silencing of indirect correlations*, Nat. Biotechnol., 31 (2013), p. 720–725.
- [18] BASSETT D. S., BULLMORE E., VERCHINSKI B. A., MATTAY V. S., WEINBERGER D. R. ET MEYER-LINDENBERG A., *Hierarchical organization of human cortical networks in health and schizophrenia*, J. Neurosci., 28 (2008), p. 9239–9248.
- [19] BASSETT D. S., PORTER M. A., WYMBBS N. F., GRAFTON S. T., CARLSON J. M. ET MUCHA P. J., *Robust detection of dynamic community structure in networks*, Chaos, 23 (2013), p. 013142.
- [20] BASTOLLA U., FORTUNA M. A., PASCUAL-GARCÍA A., FERRERA A., LUQUE B. ET BASCOMPTE J., *The architecture of mutualistic networks minimizes competition and increases biodiversity*, Nature, 458 (2009), p. 1018.
- [21] BERTUZZO E., FINGER F., MARI L., GATTO M. ET RINALDO A., *On the probability of extinction of the Haiti cholera epidemic*, Stoch. Env. Res. Risk. A, 30 (2016), p. 2043–2055.
- [22] BOERLIJST M. C., OUDMAN T. ET ROOS A. M.D. , *Catastrophic collapse can occur without early warning : Examples of silent catastrophes in structured ecological models*, PLOS one, 8 (2013), p. e62033.
- [23] BOETTIGER C. ET HASTINGS A., *Tipping points : From patterns to predictions*, Nature, 493 (2013), p. 157.
- [24] BOETTIGER C., ROSS N. ET HASTINGS A., *Early warning signals : the charted and uncharted territories*, Theor. Ecol., 6 (2013), p. 255–264.

- [25] BOGUNÁ M. ET PASTOR-SATORRAS R., *Epidemic spreading in correlated complex networks*, Phys. Rev. E, 66 (2002), p. 047104.
- [26] BOLLOBÁS B. ET BÉLA B., *Random graphs*, no 73, Cambridge University Press, 2001.
- [27] BRESSLER S. L. ET SETH A. K., *Wiener–Granger causality : a well established methodology*, NeuroImage, 58 (2011), p. 323–329.
- [28] BROADBENT S. R. ET HAMMERSLEY J. M., *Percolation processes*, dans Math. Proc. Camb. Philos. Soc., vol. 53, Cambridge University Press, 1957, p. 629–641.
- [29] BROIDO A. D. ET CLAUSET A., *Scale-free networks are rare*, Nat. Commun., 10 (2019), p. 1017.
- [30] BROUWER A. E. ET HAEMERS W. H., *Spectra of Graphs*, Springer Science & Business Media, 2011.
- [31] BRUGERE I., GALLAGHER B. ET BERGER-WOLF T. Y., *Network structure inference, a survey : Motivations, methods, and applications*, ACM Computing Surveys (CSUR), 51 (2018), p. 1–39.
- [32] CALLAWAY D. S., NEWMAN M. E. J., STROGATZ S. H. ET WATTS D. J., *Network robustness and fragility : Percolation on random graphs*, Phys. Rev. Lett., 85 (2000), p. 5468.
- [33] CASTELLANO C. ET PASTOR-SATORRAS R., *Relating topological determinants of complex networks to their spectral properties : Structural and dynamical effects*, Phys. Rev. X, 7 (2017), p. 041024.
- [34] CHEN Y., BRESSLER S. L. ET DING M., *Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data*, J. Neurosci. Methods, 150 (2006), p. 228–237.
- [35] CHUNG F., LU L. ET VU V., *Spectra of random graphs with given expected degrees*, Proc. Natl. Acad. Sci., 100 (2003), p. 6313–6318.
- [36] CLAUSET A., TUCKER E. ET SAINZ M., *The Colorado Index of Complex Networks*. <https://icon.colorado.edu>. Visité le 2020-03-26.
- [37] CLEMENTS C. F., MCCARTHY M. A. ET BLANCHARD J. L., *Early warning signals of recovery in complex systems*, Nat. Commun., 10 (2019), p. 1681.
- [38] COSTA L. D. F., OLIVEIRA JR O. N., TRAVIESO G., RODRIGUES F. A., VILLAS BOAS P. R., ANTIQUEIRA L., VIANA M. P. ET CORREA ROCHA L. E., *Analyzing and modeling real-world phenomena with complex networks : A survey of applications*, Adv. Phys., 60 (2011), p. 329–412.

- [39] CVETKOVIĆ D. M., DOOB M. ET SACHS H., *Spectra of Graphs : Theory and Application*, vol. 87, Academic Press, 1980.
- [40] DAKOS V. ET BASCOMPTE J., *Critical slowing down as early warning for the onset of collapse in mutualistic communities*, Proc. Natl. Acad. Sci., 111 (2014), p. 17546–17551.
- [41] DAKOS V., CARPENTER S. R., BROCK W. A., ELLISON A. M., GUTTAL V., IVES A. R., KEFI S., LIVINA V., SEEKELL D. A., NES E. H.v. *et al.*, *Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data*, PLOS One, 7 (2012), p. 41010.
- [42] DAKOS V., CARPENTER S. R., NES E. H.v. ET SCHEFFER M., *Resilience indicators : prospects and limitations for early warnings of regime shifts*, Philos. Trans. R. Soc. B, 370 (2015), p. 20130263.
- [43] DEMENOCAL P., ORTIZ J., GUILDERTON T., ADKINS J., SARNTHEIN M., BAKER L. ET YARUSINSKY M., *Abrupt onset and termination of the African Humid Period : Rapid climate responses to gradual insolation forcing*, Quat. Sci. Rev., 19 (2000), p. 347–361.
- [44] DENNIS B., *Allee effects : population growth, critical density, and the chance of extinction*, Nat. Resour. Model., 3 (1989), p. 481–538.
- [45] DESTEXHE A. ET SEJNOWSKI T. J., *The Wilson–Cowan model, 36 years later*, Biol. Cybern., 101 (2009), p. 1–2.
- [46] ELIASMITH C., STEWART T. C., CHOO X., BEKOLAY T., DEWOLF T., TANG Y. ET RASMUSSEN D., *A large-scale model of the functioning brain*, Science, 338 (2012), p. 1202–1205.
- [47] ERDŐS P. ET RÉNYI A., *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci., 5 (1960), p. 17–60.
- [48] ERMENTROUT G. B. ET KOPELL N., *Parabolic bursting in an excitable system coupled with a slow oscillation*, SIAM Journal on Applied Mathematics, 46 (1986), p. 233–253.
- [49] ESTRADA E., *The Structure of Complex networks : Theory and Applications*, Oxford University Press, 2012.
- [50] FEIZI S., MARBACH D., MÉDARD M. ET KELLIS M., *Network deconvolution as a general method to distinguish direct dependencies in networks*, Nat. Biotechnol., 31 (2013), p. 726.
- [51] FRACCASCIA L., GIANNOCCARO I. ET ALBINO V., *Resilience of complex systems : State of the art and directions for future research*, Complexity, 2018 (2018).

- [52] GAO J., BARZEL B. ET BARABÁSI A.-L., *Universal resilience patterns in complex networks*, Nature, 530 (2016), p. 307.
- [53] GILLESPIE D. T., *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J. Comput. Phys., 22 (1976), p. 403–434.
- [54] GLEESON J. P., *Cascades on correlated and modular random networks*, Phys. Rev. E, 77 (2008), p. 046117.
- [55] GLEESON J. P., *Binary-state dynamics on complex networks : pair approximation and beyond*, Phys. Rev. X, 3 (2013), p. 021004.
- [56] GLEESON J. P. ET CAHALANE D. J., *Seed size strongly affects cascades on random networks*, Phys. Rev. E, 75 (2007), p. 056103.
- [57] GLEESON J. P. ET PORTER M. A., *Message-passing methods for complex contagions*, dans Complex Spreading Phenomena in Social Systems, Springer, 2018, p. 81–95.
- [58] GRANGER C. W., *Investigating causal relations by econometric models and cross-spectral methods*, Econometrica, (1969), p. 424–438.
- [59] GRILLI J., ADORISIO M., SUWEIS S., BARABÁS G., BANAVAR J. R., ALLESINA S. ET MARITAN A., *Feasibility and coexistence of large ecological communities*, Nat. Commun., 8 (2017), p. 0.
- [60] GUALDI S., TARZIA M., ZAMPONI F. ET BOUCHAUD J.-P., *Tipping points in macroeconomic agent-based models*, J. Econ. Dyn. Control, 50 (2015), p. 29–61.
- [61] HACKETT A. ET GLEESON J. P., *Cascades on clique-based graphs*, Phys. Rev. E, 87 (2013), p. 062801.
- [62] HACKETT A., MELNIK S. ET GLEESON J. P., *Cascades on a class of clustered random networks*, Phys. Rev. E, 83 (2011), p. 056107.
- [63] HALDANE A. G. ET MAY R. M., *Systemic risk in banking ecosystems*, Nature, 469 (2011), p. 351–355.
- [64] HAMILTON W., YING Z. ET LESKOVEC J., *Inductive representation learning on large graphs*, dans NeurIPS, 2017, p. 1024–1034.
- [65] HASHIMOTO K.-I., *Zeta functions of finite graphs and representations of p -adic groups*, dans Automorphic Forms and Geometry of Arithmetic Varieties, Elsevier, 1989, p. 211–280.
- [66] HE K., ZHANG X., REN S. ET SUN J., *Delving deep into rectifiers : Surpassing human-level performance on Imagenet classification*, dans Proceedings of the ICCV, 2015, p. 1026–1034.

- [67] HE X., ZHANG H., KAN M.-Y. ET CHUA T.-S., *Fast matrix factorization for online recommendation with implicit feedback*, dans SIGIR, 2016, p. 549–558.
- [68] HÉBERT-DUFRESNE L., ALLARD A., MARCEAU V., NOËL P.-A. ET DUBÉ L. J., *Structural preferential attachment : Network organization beyond the link*, Phys. Rev. Lett., 107 (2011), p. 158702.
- [69] HÉBERT-DUFRESNE L., LAURENCE E., ALLARD A., YOUNG J.-G. ET DUBÉ L. J., *Complex networks as an emerging property of hierarchical preferential attachment*, Phys. Rev. E, 92 (2015), p. 062809.
- [70] HECKER M., LAMBECK S., TOEPFER S., VAN SOMEREN E. ET GUTHKE R., *Gene regulatory network inference : data integration in dynamic models—A review*, BioSystems, 96 (2009), p. 86–103.
- [71] HOEGH-GULDBERG O., MUMBY P. J., HOOTEN A. J., STENECK R. S., GREENFIELD P., GOMEZ E., HARVELL C. D., SALE P. F., EDWARDS A. J., CALDEIRA K. *et al.*, *Coral reefs under rapid climate change and ocean acidification*, Science, 318 (2007), p. 1737–1742.
- [72] HOLLAND J. N., DEANGELIS D. L. ET BRONSTEIN J. L., *Population dynamics and mutualism : Functional responses of benefits and costs*, Am. Nat., 159 (2002), p. 231–244.
- [73] HOLLAND P. W., LASKEY K. B. ET LEINHARDT S., *Stochastic blockmodels : First steps*, Soc. Networks, 5 (1983), p. 109–137.
- [74] HOLLING C. S., *The functional response of predators to prey density and its role in mimicry and population regulation*, Mem. Entomol. Soc. Can., 97 (1965), p. 1–60.
- [75] HOLLING C. S., *Resilience and stability of ecological systems*, Annu. Rev. Ecol. Evol. Syst., 4 (1973), p. 1–23.
- [76] HOLME P., *Three faces of node importance in network epidemiology : Exact results for small graphs*, Phys. Rev. E, 96 (2017), p. 062305.
- [77] HOPFIELD J. J., *Neurons with graded response have collective computational properties like those of two-state neurons*, Proc. Natl. Acad. Sci., 81 (1984), p. 3088–3092.
- [78] HORN R. A. ET JOHNSON C. R., *Matrix Analysis*, Cambridge University Press, 1990.
- [79] IZHIKEVICH E. M., *Simple model of spiking neurons*, IEEE Trans. Neural Netw., 14 (2003), p. 1569–1572.
- [80] ———, *Dynamical Systems in Neuroscience*, MIT press, 2007.

- [81] JÄGER G. ET FÜLLSACK M., *Systematically false positives in early warning signal analysis*, PLOS one, 14 (2019), p. e0211072.
- [82] JÄGER G., HOFER C., KAPELLER M. ET FÜLLSACK M., *Hidden early-warning signals in scale-free networks*, PLOS One, 12 (2017), p. 0189853.
- [83] JIANG J., HUANG Z.-G., SEAGER T. P., LIN W., GREBOGI C., HASTINGS A. ET LAI Y.-C., *Predicting tipping points in mutualistic networks through dimension reduction*, Proc. Natl. Acad. Sci. U.S.A., 115 (2018), p. E639–E647.
- [84] KAISER M., GÖRNER M. ET HILGETAG C. C., *Criticality of spreading dynamics in hierarchical cluster networks without inhibition*, New J. Phys., 9 (2007), p. 110.
- [85] KAISER M. ET HILGETAG C. C., *Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems*, PLOS Comput. Biol., 2 (2006).
- [86] KAISER M. ET HILGETAG C. C., *Optimal hierarchical modular topologies for producing limited sustained activation of neural networks*, Front. Neuroinform., 4 (2010), p. 1662–5196.
- [87] KARRER B. ET NEWMAN M. E. J., *Random graphs containing arbitrary distributions of subgraphs*, Phys. Rev. E., 82 (2010), p. 066118.
- [88] KARRER B., NEWMAN M. E. J. ET ZDEBOROVÁ L., *Percolation on sparse networks*, Phys. Rev. Lett., 113 (2014), p. 208702.
- [89] KÉFI S., DAKOS V., SCHEFFER M., VAN NES E. H. ET RIETKERK M., *Early warning signals also precede non-catastrophic transitions*, Oikos, 122 (2013), p. 641–648.
- [90] KIPF T. N. ET WELLING M., *Semi-supervised classification with graph convolutional networks*, arXiv preprint arXiv :1609.02907, (2016).
- [91] KISS I. Z., MILLER J. C., SIMON P. L. *et al.*, *Mathematics of Epidemics on Networks*, Springer, 2017.
- [92] KRAPIVSKY P. L. ET REDNER S., *Organization of growing random networks*, Phys. Rev. E, 63 (2001), p. 066123.
- [93] KRAPIVSKY P. L., REDNER S. ET LEYVRAZ F., *Connectivity of growing random networks*, Phys. Rev. Lett., 85 (2000), p. 4629.
- [94] KRAPIVSKY P. L., RODGERS G. J. ET REDNER S., *Degree distributions of growing networks*, Phys. Rev. Lett., 86 (2001), p. 5401.
- [95] KRZAKALA F., MOORE C., MOSSEL E., NEEMAN J., SLY A., ZDEBOROVÁ L. ET ZHANG P., *Spectral redemption in clustering sparse networks*, Proc. Natl. Acad. Sci., 110 (2013), p. 20935–20940.

- [96] LAURENCE E., DOYON N., DUBÉ L. J. ET DESROSIERS P., *Spectral dimension reduction of complex dynamical networks*, Phys. Rev. X, 9 (2019), p. 011042.
- [97] LAURENCE E., YOUNG J.-G., MELNIK S. ET DUBÉ L. J., *Exact analytical solution of irreversible binary dynamics on networks*, Phys. Rev. E, 97 (2018), p. 032302.
- [98] LEE D. D. ET SEUNG H. S., *Learning the parts of objects by non-negative matrix factorization*, Nature, 401 (1999), p. 788–791.
- [99] LENTON T. M., *Early warning of climate tipping points*, Nat. Clim. Change, 1 (2011), p. 201–209.
- [100] LI T. ET DING C., *The relationships among various nonnegative matrix factorization methods for clustering*, dans ICDM, IEEE, 2006, p. 362–371.
- [101] LIU L., JIANG H., HE P., CHEN W., LIU X., GAO J. ET HAN J., *On the variance of the adaptive learning rate and beyond*, arXiv preprint arXiv :1908.03265, (2019).
- [102] LOKHOV A. Y., MÉZARD M. ET ZDEBOROVÁ L., *Dynamic message-passing equations for models with unidirectional dynamics*, Phys. Rev. E, 91 (2015), p. 012811.
- [103] LÜ L., CHEN D., REN X.-L., ZHANG Q.-M., ZHANG Y.-C. ET ZHOU T., *Vital nodes identification in complex networks*, Phys. Rep., 650 (2016), p. 1–63.
- [104] LYNALL M.-E., BASSETT D. S., KERWIN R., MCKENNA P. J., KITZBICHLER M., MULLER U. ET BULLMORE E., *Functional connectivity and brain networks in schizophrenia*, J. Neurosci., 30 (2010), p. 9477–9487.
- [105] MADENJIAN C. P., BREIDERT B., BOYARSKI D., BRONTE C. R., DICKINSON B., DONNER K., EBENER M. P., GORDON R., HANSON D., HOLEY M. *et al.*, *2016 Lake Michigan lake trout working group report*, dans Lake Michigan Committee Meeting, Great Lakes Fishery Commission, 2017.
- [106] MAJDANDZIC A., BRAUNSTEIN L. A., CURME C., VODENSKA I., LEVY-CARCIENTE S., STANLEY H. E. ET HAVLIN S., *Multiple tipping points and optimal repairing in interacting networks*, Nat. Commun., 7 (2016), p. 10850.
- [107] MANN M. E. ET LEES J. M., *Robust estimation of background noise and signal detection in climatic time series*, Clim. Change, 33 (1996), p. 409–445.
- [108] MARBACH D., COSTELLO J. C., KÜFFNER R., VEGA N. M., PRILL R. J., CAMACHO D. M., ALLISON K. R., ADERHOLD A., BONNEAU R., CHEN Y. *et al.*, *Wisdom of crowds for robust gene network inference*, Nat. Methods, 9 (2012), p. 796.
- [109] MASUDA N. ET HOLME P., *Detecting sequences of system states in temporal networks*, Sci. Rep., 9 (2019), p. 1–11.

- [110] MATA A. S. ET FERREIRA S. C., *Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks*, EPL, 103 (2013), p. 48003.
- [111] MAY R. M., *Will a large complex system be stable ?*, Nature, 238 (1972), p. 413.
- [112] MAY R. M., *Stability and complexity in model ecosystems*, vol. 6, Princeton University Press, 2001.
- [113] MAY R. M., LEVIN S. A. ET SUGIHARA G., *Ecology for bankers*, Nature, 451 (2008), p. 893–894.
- [114] MELNIK S., HACKETT A., PORTER M. A., MUCHA P. J. ET GLEESON J. P., *The unreasonable effectiveness of tree-based theory for networks with clustering*, Phys. Rev. E, 83 (2011), p. 036112.
- [115] MELNIK S., PORTER M. A., MUCHA P. J. ET GLEESON J. P., *Dynamics on modular networks with heterogeneous correlations*, Chaos, 24 (2014), p. 023106.
- [116] METROPOLIS N., ROSENBLUTH A. W., ROSENBLUTH M. N., TELLER A. H. ET TELLER E., *Equation of state calculations by fast computing machines*, J. Chem. Phys., 21 (1953), p. 1087–1092.
- [117] MEYERS L., *Contact network epidemiology : Bond percolation applied to infectious disease prediction and control*, Bull. Amer. Math. Soc., 44 (2007), p. 63–86.
- [118] MILLER J. C., *Equivalence of several generalized percolation models on networks*, Phys. Rev. E, 94 (2016), p. 032313.
- [119] MOLLOY M. ET REED B., *A critical point for random graphs with a given degree sequence*, Random Struc. Algor., 6 (1995), p. 161–180.
- [120] ———, *The size of the giant component of a random graph with a given degree sequence*, Comb. Probab. Comput., 7 (1998), p. 295–305.
- [121] MORONE F. ET MAKSE H. A., *Influence maximization in complex networks through optimal percolation*, Nature, 524 (2015), p. 65.
- [122] NADAKUDITI R. R. ET NEWMAN M. E. J., *Graph spectra and the detectability of community structure in networks*, Phys. Rev. Lett., 108 (2012), p. 188701.
- [123] NEWMAN M. E., *Estimating network structure from unreliable measurements*, Phys. Rev. E, 98 (2018), p. 062321.
- [124] ———, *Networks*, Oxford university press, 2018.
- [125] NEWMAN M. E. J., *Scientific collaboration networks. I. Network construction and fundamental results*, Phys. Rev. E, 64 (2001), p. 016131.

- [126] NEWMAN M. E. J., *Spread of epidemic disease on networks*, Phys. Rev. E, 66 (2002), p. 016128.
- [127] NEWMAN M. E. J., *The structure and function of complex networks*, SIAM review, 45 (2003), p. 167–256.
- [128] NEWMAN M. E. J., STROGATZ S. H. ET WATTS D. J., *Random graphs with arbitrary degree distributions and their applications*, Phys. Rev. E, 64 (2001), p. 026118.
- [129] NICA B., *A Brief Introduction to Spectral Graph Theory*, European Mathematical Society, 2018.
- [130] NOH J. D. ET RIEGER H., *Random walks on complex networks*, Phys. Rev. Lett., 92 (2004), p. 118701.
- [131] NYSTRÖM M., FOLKE C. ET MOBERG F., *Coral reef disturbance and resilience in a human-dominated environment*, Trends Ecol. Evol., 15 (2000), p. 413–417.
- [132] PAN L., YANG D., WANG W., CAI S., ZHOU T. ET LAI Y.-C., *Complete phase diagrams of the coevolving spreading dynamics in complex networks*, arXiv preprint arXiv :2001.09565, (2020).
- [133] PAN L., ZHOU T., LÜ L. ET HU C.-K., *Predicting missing links and identifying spurious links via likelihood analysis*, Sci. Rep., 6 (2016), p. 1–10.
- [134] PARK H.-J. ET FRISTON K., *Structural and functional brain networks : From connections to cognition*, Science, 342 (2013), p. 1238411.
- [135] PASTOR-SATORRAS R., CASTELLANO C., VAN MIEGHEM P. ET VESPIGNANI A., *Epidemic processes in complex networks*, Rev. Mod. Phys., 87 (2015), p. 925.
- [136] PASTOR-SATORRAS R. ET VESPIGNANI A., *Epidemic dynamics and endemic states in complex networks*, Phys. Rev. E, 63 (2001), p. 066117.
- [137] PEIXOTO T. P., *Hierarchical block structures and high-resolution model selection in large networks*, Phys. Rev. X, 4 (2014), p. 011047.
- [138] PEIXOTO T. P., *Network reconstruction and community detection from dynamics*, arXiv preprint arXiv :1903.10833, (2019).
- [139] PÉREZ R., PUIG V., PASCUAL J., PERALTA A., LANDEROS E. ET JORDANAS L., *Pressure sensor distribution for leak detection in Barcelona water distribution network*, Water Sci. Tech.-W. Sup., 9 (2009), p. 715–721.
- [140] PIELOU E. C. *et al.*, *An Introduction to Mathematical Ecology*, New York, USA, Wiley-Inter-science, 1969.

- [141] PIRES M. M., *Rewilding ecological communities and rewiring ecological networks*, *Perspect. Ecol. Conser.*, 15 (2017), p. 257–265.
- [142] RICHARDSON M. P., *Large scale brain models of epilepsy : Dynamics meets connectomics*, *J. Neurol. Neurosurg. Psychiatry*, (2012), p. jnnp–2011.
- [143] ROEBROECK A., FORMISANO E. ET GOEBEL R., *Mapping directed influence over the brain using Granger causality and fMRI*, *NeuroImage*, 25 (2005), p. 230–242.
- [144] SCHEFFER M., *Critical Transitions in Nature and Society*, Princeton University Press, 2009.
- [145] ———, *Foreseeing tipping points*, *Nature*, 467 (2010), p. 411–412.
- [146] SCHEFFER M., BASCOMPTE J., BROCK W. A., BROVKIN V., CARPENTER S. R., DAKOS V., HELD H., VAN NES E. H., RIETKERK M. ET SUGIHARA G., *Early-warning signals for critical transitions*, *Nature*, 461 (2009), p. 53.
- [147] SCHEFFER M. ET CARPENTER S. R., *Catastrophic regime shifts in ecosystems : Linking theory to observation*, *Trends Ecol. Evol.*, 18 (2003), p. 648–656.
- [148] SCHINDLER D. W., *Recent advances in the understanding and management of eutrophication*, *Limnol. Oceanogr.*, 51 (2006), p. 356–363.
- [149] SCHWEITZER F., FAGIOLO G., SORNETTE D., VEGA-REDONDO F., VESPIGNANI A. ET WHITE D. R., *Economic networks : The new challenges*, *Science*, 325 (2009), p. 422–425.
- [150] SETH A. K., BARRETT A. B. ET BARNETT L., *Granger causality analysis in neuroscience and neuroimaging*, *J. Neurosci.*, 35 (2015), p. 3293–3297.
- [151] SHANDILYA S. G. ET TIMME M., *Inferring network topology from complex dynamics*, *New J. Phys.*, 13 (2011), p. 013004.
- [152] SHEIKHATTAR A., MIRAN S., LIU J., FRITZ J. B., SHAMMA S. A., KANOLD P. O. ET BABADI B., *Extracting neuronal functional network dynamics via adaptive Granger causality analysis*, *Proc. Natl. Acad. Sci. U.S.A.*, 115 (2018), p. E3869–E3878.
- [153] SHRESTHA M. ET MOORE C., *Message-passing approach for threshold models of behavior in networks*, *Phys. Rev. E*, 89 (2014), p. 022805.
- [154] SHRESTHA M., SCARPINO S. V. ET MOORE C., *Message-passing approach for recurrent-state epidemic models on networks*, *Phys. Rev. E*, 92 (2015), p. 022821.
- [155] SIMON H. A., *Models of Man ; Social and Rational*, Wiley, 1957.

- [156] SOLE R. V. ET MONTOYA M., *Complexity and fragility in ecological networks*, Proc. Royal Soc. B, 268 (2001), p. 2039–2045.
- [157] ST-ONGE G., YOUNG J.-G., HÉBERT-DUFRESNE L. ET DUBÉ L. J., *Efficient sampling of spreading processes on complex networks using a composition and rejection algorithm*, Comput. Phys. Commun., 240 (2019), p. 30–37.
- [158] ST-ONGE G., YOUNG J.-G., LAURENCE E., MURPHY C. ET DUBÉ L. J., *Phase transition of the susceptible-infected-susceptible dynamics on time-varying configuration model networks*, Phys. Rev. E, 97 (2018), p. 022305.
- [159] STROGATZ S. H., *Nonlinear dynamics and chaos : With applications to physics, biology, chemistry, and engineering*, CRC Press, 2018.
- [160] TU C., GRILLI J., SCHUESSLER F. ET SUWEIS S., *Collapse of resilience patterns in generalized Lotka-Volterra dynamics and beyond*, Phys. Rev. E, 95 (2017), p. 062307.
- [161] VAN DEN HEUVEL M. P. ET SPORNS O., *Rich-club organization of the human connectome*, J. Neurosci., 31 (2011), p. 15775–15786.
- [162] VAN LAARHOVEN P. J. ET AARTS E. H., *Simulated annealing*, dans *Simulated Annealing : Theory and Applications*, Springer, 1987, p. 7–15.
- [163] VELIČKOVIĆ P., CUCURULL G., CASANOVA A., ROMERO A., LIÒ P. ET BENGIO Y., *Graph attention networks*, arXiv :1710.10903, (2017).
- [164] VOITALOV I., HOORN P.V. D. , HOFSTAD R.V. D. ET KRIOUKOV D., *Scale-free networks well done*, Phys. Rev. Res., 1 (2019), p. 033034.
- [165] WANG Y., SUN Y., LIU Z., SARMA S. E., BRONSTEIN M. M. ET SOLOMON J. M., *Dynamic graph CNN for learning on point clouds*, ACM Trans. Graph, 38 (2019), p. 1–12.
- [166] WATTS D. J., *A simple model of global cascades on random networks*, Proc. Natl. Acad. Sci., 99 (2002), p. 5766–5771.
- [167] WATTS D. J. ET STROGATZ S. H., *Collective dynamics of ‘small-world’ networks*, Nature, 393 (1998), p. 440.
- [168] WEISBUCH G., *Complex Systems Dynamics*, CRC Press, 2018.
- [169] WILF H., *Generatingfunctionology*, Academic Press, New York, 1990.
- [170] WILKAT T., RINGS T. ET LEHNERTZ K., *No evidence for critical slowing down prior to human epileptic seizures*, Chaos, 29 (2019), p. 091104.

- [171] WILSON H. R. ET COWAN J. D., *Excitatory and inhibitory interactions in localized populations of model neurons*, Biophys. J., 12 (1972), p. 1–24.
- [172] ———, *A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue*, Kybernetik, 13 (1973), p. 55–80.
- [173] WISSEL C., *A universal law of the characteristic return time near thresholds*, Oecologia, 65 (1984), p. 101–107.
- [174] XU K., HU W., LESKOVEC J. ET JEGELKA S., *How powerful are graph neural networks ?*, arXiv preprint arXiv :1810.00826, (2018).
- [175] YING R., HE R., CHEN K., EKSOMBATCHAI P., HAMILTON W. L. ET LESKOVEC J., *Graph convolutional neural networks for web-scale recommender systems*, arXiv :1806.01973, (2018).
- [176] YOUNG J.-G., ST-ONGE G., LAURENCE E., MURPHY C., HÉBERT-DUFRESNE L. ET DESROSIERS P., *Phase transition in the recoverability of network history*, Phys. Rev. X, 9 (2019), p. 041056.
- [177] YOUNG J.-G., VALDOVINOS F. S. ET NEWMAN M. E., *Reconstruction of plant–pollinator networks from observational data*, bioRxiv, (2019), p. 754077.
- [178] ZHOU D. W., MOWREY D. D., TANG P. ET XU Y., *Percolation model of sensory transmission and loss of consciousness under general anesthesia*, Phys. Rev. Lett., 115 (2015), p. 108103.
- [179] ZHOU J., CUI G., ZHANG Z., YANG C., LIU Z., WANG L., LI C. ET SUN M., *Graph neural networks : A review of methods and applications*, arXiv :1812.08434, (2018).