

Analysis and control of diffusion processes in networks Kevin Scaman

► To cite this version:

Kevin Scaman. Analysis and control of diffusion processes in networks. General Mathematics [math.GM]. Université Paris Saclay (COmUE), 2016. English. NNT: 2016SACLN039. tel-01531840

HAL Id: tel-01531840 https://theses.hal.science/tel-01531840

Submitted on 2 Jun2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



école———
normale ———
supérieure ———
paris-saclay

NNT: 2016SACLN039

THESE DE DOCTORAT DE L'UNIVERSITE PARIS-SACLAY PREPAREE A L'ECOLE NORMALE SUPERIEURE PARIS-SACLAY

ECOLE DOCTORALE N° 574 Mathématiques Hadamard

Spécialité de doctorat : Mathématiques appliquées

Par

M. Kevin Scaman

Etude et contrôle de phénomènes diffusifs dans un réseau

Thèse présentée et soutenue à Cachan, le 21 octobre 2016 :

Composition du Jury :

M. Massoulié Laurent	Directeur de Recherche, INRIA - Microsoft Research
M. Comets Francis	Professeur, Université Paris Diderot
M. Lelarge Marc	Chargé de Recherche, INRIA - ENS
M. Bordenave Charles	Chargé de Recherche, Université de Toulouse
M. Sturm Karl-Theodor	Professeur, Université de Bonn
M. Vayatis Nicolas	Professeur, ENS Paris-Saclay
M. Kalogeratos Argyris	Chercheur, ENS Paris-Saclay

Président Rapporteur Rapporteur Examinateur Examinateur Directeur de thèse Co-encadrant de thèse

Contents

1	Intr	oductio	on (version française)	17
2	Synt	thèse d	es contributions	19
	2.1	Analys	e des modèles de diffusion	19
		2.1.1	Comportement long-terme de l'influence	20
		2.1.2	Propriétés dynamiques de l'influence	21
	2.2	Contrô	ele de processus diffusifs	22
		2.2.1	Contrôle dynamique d'épidémies SIS	23
		2.2.2	Une analyse détaillée de la planification prioritaire	23
	2.3	Conve	rgence des processus diffusifs	24
		2.3.1	Sur la convergence des réseaux	24
		2.3.2	Continuité de caractéristiques clé et processus	25
3	Intr	oductio	on (english version)	27
4	Sum	mary o	of contributions	29
	4.1	Analys	sis of diffusion processes	29
		4.1.1	Long-term behavior of the influence	30
		4.1.2	Dynamic properties of the influence	31
	4.2	Contro	l strategies for diffusion processes	32
		4.2.1	Dynamic control of SIS epidemics	33
		4.2.2	A detailed analysis of priority planning	33
	4.3	Limitir	ng behavior of diffusion processes	33
		4.3.1	On the convergence of networks	34
		4.3.2	Continuity of key characteristics and processes	34
I	Ana	lysis o	of diffusion processes in networks	37
5	Mod	lels of r	networks and diffusion	39
	5.1	Introdu	action	39
	5.2	Netwo	rk models and characteristics	40
		5.2.1	Basic definitions	40
		5.2.2	Spectral radius	42
		5.2.3	Connected components	42
		5.2.4	Diameter	43
		5.2.5	Degree distribution	43
		5.2.6	Other characteristics	44
	5.3	Rando	m networks	44
		5.3.1	Erdös-Rényi random graphs	44
		5.3.2	Preferential attachment	45
		5.3.3	Configuration model	46

		5.3.4	Geometric random graphs	47		
		5.3.5	Other models of random graphs	48		
	5.4	Diffus	ion processes	49		
		5.4.1	Percolation models	49		
		5.4.2	Epidemic models	50		
		5.4.3	Information cascades	52		
		5.4.4	Multivariate Hawkes processes	53		
6	Lon	g-term	behavior of the influence	55		
	6.1	Introd	luction	56		
		6.1.1	Related works	56		
		6.1.2	Outline	56		
	6.2	Hazar	d matrix, influence and LPC property	56		
		6.2.1	Hazard characteristics of random graphs	57		
		6.2.2	Reachability and influence	58		
		6.2.3	Random graphs with Local Positive Correlation (LPC)	60		
	6.3	Non-a	symptotic upper bounds on the influence	61		
		6.3.1	Deterministic (worst-case) scenario	61		
		6.3.2	Random influencer set with fixed size	62		
		6.3.3	Random influencer set with random size	63		
		6.3.4	Lower bounds	63		
	6.4	Applie	cation to bond percolation	64		
		6.4.1	Size and existence of the giant component	64		
		6.4.2	Number of components of cardinality larger than m	65		
	6.5	Applic	cation to site percolation	66		
	6.6	Applic	cation to epidemiology	66		
		6.6.1	Subcritical behavior in the standard SIR model	67		
		6.6.2	Behavior near the epidemic threshold	68		
		6.6.3	Generic incubation period	69		
	6.7	Applic	cation to Information Cascades	69		
	6.8	8 Discussion				
	6.9	Proofs	S	70		
		6.9.1	Behavior of the Hazard function	70		
		6.9.2	Proofs of the upper bounds on influence	72		
		6.9.3	Proofs of the percolation theorems	80		
7	Dyn	amic p	properties of the influence	85		
	/.1	Introd		86		
		/.1.1		86		
	7.0	7.1.2 Countin	Outline Time Information Constitution	86		
	1.2		Inducus-Hine Information Cascades	8/ 07		
		7.2.1	The Loplace Harris Matrix	ŏ/		
		7.2.2		87		
	7.0	/.2.3 TL	Existence of a critical time of a contagion	88		
	1.3	1 neor	Under the influence of a set of nodes	88		
		7.3.1	Upper bounds on the maximum influence at time 1	88		
		1.3.2	Lower bound on the critical time of a contagion	89		

Applic	ation to particular contagion models	90
7.4.1	Fixed transmission pattern	90
7.4.2	Exponential transmission probabilities	90
7.4.3	SI and SIR epidemic models	91
7.4.4	Discrete-time Information Cascade	92
Experi	mental results	93
Proofs		94
7.6.1	Proof for the definition of critical time	94
7.6.2	Proofs of the upper bounds on influence	95
7.6.3	Proofs of the upper bounds on critical time	97
7.6.4	Proofs for the particular contagion models	97
	Applic 7.4.1 7.4.2 7.4.3 7.4.4 Experi Proofs 7.6.1 7.6.2 7.6.3 7.6.4	Application to particular contagion models7.4.1Fixed transmission pattern7.4.2Exponential transmission probabilities7.4.3SI and SIR epidemic models7.4.4Discrete-time Information CascadeExperimental resultsProofs7.6.1Proof for the definition of critical time7.6.2Proofs of the upper bounds on influence7.6.3Proofs for the particular contagion models

II Control of diffusion processes

8	Dyn	amic co	ontrol of SIS epidemics	103
	8.1	Introdu	uction	. 104
		8.1.1	Related work	. 104
		8.1.2	Outline	. 104
	8.2	Epiden	nic and control model	. 105
		8.2.1	Notations	. 105
		8.2.2	Epidemic model	. 105
		8.2.3	Control model	. 105
	8.3	Quality	y metrics for the control action	. 106
		8.3.1	Extinction time	. 107
		8.3.2	Cumulative costs	. 107
		8.3.3	Area Under the Curve (AUC)	. 107
		8.3.4	Stable infection state	. 108
	8.4	DRA C	Control Strategies	. 108
		8.4.1	Intractability of optimal strategies	. 108
		8.4.2	Score-based strategies	. 108
		8.4.3	LRIE – A greedy approach to DRA	. 110
		8.4.4	MCM – an optimal priority planning strategy	. 112
	8.5	Experi	mental results	. 112
		8.5.1	Quality assessment for DRA strategies	. 112
		8.5.2	Competing strategies	. 113
		8.5.3	Experiments on simulated networks	. 114
		8.5.4	Simulations on real-world networks	. 117
	8.6	Proofs		. 119
9	A de	etailed a	analysis of priority planning	121
	9.1	Introdu	ution	. 122
		9.1.1	Related works	. 122
		9.1.2	Outline	. 122
	9.2	Priorit	v planning	. 122
		9.2.1	A healing plan to gradually remove a contagion	. 122
		9.2.2	Maxcut and cutwidth	. 123

147

167

9.3	Tight I	pounds on the extinction time \ldots \ldots \ldots \ldots \ldots \ldots \ldots 12^4
	9.3.1	Theoretical bounds for the extinction time
	9.3.2	Relationship between critical behavior and maxcut
	9.3.3	Interpretation of results
9.4	The M	<i>axCut Minimization</i> strategy
	9.4.1	Maxcut optimization
9.5	Experi	mental results
	9.5.1	Setup and competitors
	9.5.2	Quality of the theoretical bound
	9.5.3	Empirical evaluation of simulated contagion on real networks 132
9.6	Robust	tness of MCM
	9.6.1	Malicious modification of the network
	9.6.2	Random additive noise
	9.6.3	Uncertainty in the localization of nodes in contact networks 137
9.7	Proofs	
	9.7.1	Main intermediate results and lemmas
	9.7.2	Proofs of theorems
	9.7.3	Proofs of propositions and lemmas 143

III From networks to continuous spaces

10	On t	he con	vergence of networks	1/0
10	10.1	Introdu	action	150
	10.1	10.1.1	Peloted works	150
		10.1.1		150
	10.0	10.1.2		151
	10.2	Probab	Substic mapping for network convergence	152
		10.2.1	Notations	152
		10.2.2	Quantifying near-isometries	153
		10.2.3	The space of spaces	153
		10.2.4	The mapping distance	154
		10.2.5	The Gromov-Wasserstein distance	156
		10.2.6	Examples of distances and mm-spaces	157
		10.2.7	Application to network comparison	158
	10.3	Limits	of several popular graphs	159
		10.3.1	Grid graphs	159
		10.3.2	Totally connected graph	160
		10.3.3	Random sampling from an mm-space	160
		10.3.4	Geometric random networks	161
		10.3.5	Erdös-Rényi random networks and the stochastic block model	161
	10.4	Discus	sion and open problems	162
	10.5	Proofs		163
		10.5.1	Proof of the convergence of product mm-spaces	163
		10.5.2	Proof of the density of undirected graphs	163
		10.5.3	Proofs of the convergence of popular graphs	165

11 Continuity of key characteristics and processes

11.1	Operat	ors of an mm-space
	11.1.1	Mapping distance for operators 168
	11.1.2	Linear operators 169
11.2	Conve	rgence of network characteristics
	11.2.1	Average degree
	11.2.2	Degree distribution
	11.2.3	Spectral radius
	11.2.4	Diameter
11.3	Spectra	al properties of compact self-adjoint operators
	11.3.1	Convergence of spectrum 175
	11.3.2	Convergence of eigenspaces
11.4	Conve	rgence of dynamic processes 176
	11.4.1	Discrete dynamical systems
	11.4.2	Random walks 177
	11.4.3	Differential systems 178
11.5	Proofs	
	11.5.1	Proofs of the convergence of operators
	11.5.2	Proofs of the convergence of spectral properties
	11.5.3	Proofs of the convergence of dynamic processes

Bibliography

187

Résumé

La propagation au sein d'un réseau est un sujet d'étude pour de nombreux domaines scientifiques. Épidémies, marketing viral ou propagation d'information au sein d'un réseau social sont autant de phénomènes réels modélisés par l'évolution d'une caractéristique se propageant à travers un réseau de proche en proche. Ainsi, être capable d'agir sur ces phénomènes de diffusion est un enjeu capital dans de nombreux domaines. Malgré l'abondance de la littérature à ce sujet sur le plan théorique, et notamment la détermination d'un seuil épidémique au dessous duquel la propagation se résorbe, un certain nombre de limitations réduisent l'impact pratique de ces travaux. Dans cette thèse, nous avons travaillé à réduire la distance séparant pratique et théorie, et ce suivant trois axes: la généralisation de résultats théoriques à une classe plus large et réaliste de modèles de propagation, le développement de méthodes de contrôle dynamique efficaces utilisant de manière judicieuse la structure du réseau, et enfin la définition de nouveaux outils mathématiques faisant le lien entre méthodes spatiales et de réseau en épidémiologie. Plus particulièrement, nos travaux permettent l'analyse rigoureuse du comportement des caractéristiques d'un réseau lorsque celui-ci se rapproche, au niveau de sa structure, d'un espace métrique donné, et pourrait permettre l'application de méthodes de contrôle sur réseau à des données spatiales et macroscopiques (notamment à l'aide de données démographiques et de transport) du réseau de contact au sein d'une population.

Abstract

The propagation of a characteristic through a network is the subject of study of many scientific fields. Epidemics, viral marketing or information propagation through a social network are among the many examples of real phenomena modeled by the evolution of a characteristic propagating through the edges of a network. Thus, being capable of acting on these diffusion processes is of capital interest for many fields. Despite the large literature on the theoretical aspects of diffusion processes, and more specifically the discovery of an epidemic threshold under which the propagation is not sustainable, a number of practical limitations prevent the use of these studies in real-life scenarios. In this thesis, we work on reducing the distance separating theory from practice, following three distinct research directions: the generalization of theoretical results to a larger and more realistic class of diffusion models, the development of efficient dynamic control measures utilizing the structure of the network to its advantage, and, finally, the definition of new mathematical tools bridging the gap between *spatial* and *network* approaches in epidemiology. More specifically, our work allows the rigorous analysis of the behavior of a network's characteristics when it converges, in a structural sense, to a given metric space, and could open the way to the application of control strategies that have been developed for networks to scenarios in which only spatial and macroscopic information is available (e.g. transportation or demographic data).

Acknowledgements

First, I would like to thank Nicolas for giving me this unique opportunity to work and learn in such a wonderful environment. These three years have been a collection of great moments and ever renewed enthusiasm for the projects I have had the chance to work on. I would also like to thank Francis Comets and Marc Lelarge for doing me the honor of reviewing this thesis. I am grateful to Charles Bordenave, Laurent Massoulié and Karl-Theodor Sturm, for participating in my thesis committee and taking interest in my work. It is an honor to have you all on my committee.

I would also like to thank all the team at CMLA for all the good moments that we shared, interesting discussions at lunch and well deserved beers: Emile, Julien, Thomas, Charles, Cédric, Nick, Kumar, Ludovic, Juan, Ioannis, and more generally the whole lab, Micheline, Virginie and Véronique, as well as the interns that I had the chance to supervise. My biggest thanks go to Argyris and Rémi who shared with me the thrill of discovery with the same passion as I did, helped me when I needed help, and pushed me when the deadlines were approaching...!

Finally, I would like to thank my parents for giving me the passion for science, Meilin for her day-to-day help, coping with my lack of time and being there for me in the difficult moments, and Fiona, without whom I would never have had the courage to start such an ambitious project.

Notations

General

card(X)	Number of elements of <i>X</i>	p. 40
$\mathcal{P}(X), \mathcal{P}_n(X)$	Set of all subsets of X , and subsets of size n of X	p. 40
[n]	$\{1,, n\}$	p. 40
$\mathbb{1}\{\cdot\}$	Indicator function	p. 40
A a.s.	A holds almost surely, i.e. $\mathbb{P}(A) = 1$	p. 40
A_n a.a.s.	A_n holds asymptotically almost surely, i.e. $\lim_{n\to+\infty} \mathbb{P}(A_n) = 1$	p. 40
$\rho(M)$	Spectral radius (i.e. largest eigenvalue) of the matrix M	p. 42
$f(n) = \theta(g(n))$	$C_1g(n) \le f(n) \le C_2g(n)$ for some C_1 and C_2	p. 43
0, 1	All zeros and all ones vectors	p. 110

Graphs

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$	Graph with node set ${\mathcal V}$ and edge set ${\mathcal E}$	p. 40
A or \mathcal{A}	Adjacency matrix	p. 41
<i>n</i> or N	Number of nodes, i.e. $\operatorname{card}(\mathcal{V})$	p. 41
G(n, A)	Random graph of size n and adjacency matrix A	p. 44
$\mathcal{C}^*(\ell)$	Maximum cut in the ordering ℓ	p. 124

Diffusion processes

β,δ	Epidemic parameters for the SI, SIS and SIR models	p. 50
X(t)	State vector of the diffusion process at time t	p. 49
$\rho(t)$	Resource allocation vector at time t	p. 105
$\sigma_{\mathcal{I}}(t)$	Influence of the set ${\mathcal I}$ at time t in a diffusion process	p. 49

Influence bounds

${\cal H}$	Hazard matrix	p. 57
$\mathcal{L}(s)$	Laplace Hazard matrix	p. 87
ρ_n	Hazard radius	p. 57
$\sigma(\mathcal{I})$	Influence of the set ${\mathcal I}$ in a random graph	p. 58

Convergence of networks

\mathcal{X}	Metric and measureable space (mm-space)	p. 153
$d_{\mathcal{X}}$ and $\mu_{\mathcal{X}}$	Metric and measure of ${\cal X}$	p. 153
$\operatorname{size}_p(\mathcal{X})$	p -size of $\mathcal X$	p. 154
$\Delta_p(\mathcal{X},\mathcal{Y})$	Gromov-Wasserstein distance between ${\mathcal X}$ and ${\mathcal Y}$	p. 156
$\Delta_p(f,g)$	Mapping distance between the functions f and g	p. 154
$\Delta_p(F,G)$	Mapping distance between the operators F and G	p. 168
X_p	Set of all mm-spaces of finite <i>p</i> -size	p. 156
$\operatorname{Opt}_p(\mathcal{X},\mathcal{Y})$	Set of optimal couplings between $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$ w.r.t. Δp	p. 156

Introduction (version française)

"Si nous savions ce que nous faisons, cela ne s'appellerait plus de la recherche, n'est-ce pas ?"

- Albert Einstein

Le concept d'approximation est une notion centrale dans les sciences, et trouver le juste niveau d'abstraction est souvent l'élément clé de la réussite d'une théorie. Dans de nombreux domaines scientifiques, dont l'économie, la sociologie ou encore la biologie, l'infinie complexité d'un individu est souvent réduite à son expression la plus simple: un nœud anonyme et indifférenciable au sein d'un gigantesque réseau. Bien que cette approximation puisse sembler trop forte pour être source d'intuition, les réseaux réels, qu'ils soient informatiques, sociaux ou bancaires, partagent un certain nombre de caractéristiques macroscopiques indépendamment des traits individuels de leurs éléments. En un sens, les caractéristiques macroscopiques de ces systèmes complexes sont formés, non par les propriétés locales de leurs éléments, mais par la structure complexe de leurs interactions. Prédire le comportement des épidémies, diffusion d'information ou crises financières, nécessite donc de comprendre cette structure et comment celle-ci agit sur les aspects dynamiques du réseau. Autrement dit, il s'agit de comprendre comment la propagation d'individu à individu le long des chemins complexes du réseau va influer sur l'évolution temporelle d'une diffusion, et ainsi imposer son comportement macroscopique à l'échelle d'une population.

Les réseaux sociaux offrent aux scientifiques une occasion unique d'observer de tels processus diffusifs en temps réel et avec précision. La propagation d'une information entre utilisateurs, lorsque celle-ci est à caractère publique, est mise à disposition de la communauté scientifique dans d'énormes bases de données. Cette profusion de données a permit le développement rapide du domaine et de ses applications pratiques en marketing. Cependant, l'analyse théorique de ces processus de diffusion n'a pas bénéficié de cet essor, et une partie du travail réalisé au cours de cette thèse a consisté à comprendre quelles caractéristiques structurelles du réseaux déterminent le comportement macroscopique de ces cascades d'information.

En épidémiologie, l'analyse théorique ne constitue qu'une première étape permettant la prédiction précise des épidémies futures, et ne saurait se substituer au développement de mesures préventives. La principale limitation de la littérature à ce sujet consiste à supposer le réseau de contacts entre individus connu. Bien que les algorithmes d'inférence puissent fournir de bonnes approximations d'un réseau lorsque de grandes quantités de données sont disponibles (par exemple dans le cas des réseaux sociaux), le réseau sousjacent est parfois, bien malheureusement, au-delà de notre portée. Afin de contourner cette difficulté, une pratique courante consiste à simplifier le réseau de contacts dans une population par un réseau de villes ou états, sur lesquels un modèle simple d'épidémie est appliqué. Une partie de notre travail dans cette thèse a été consacrée à réduire la distance entre modèles simplifiés et modèles exacts, dans l'optique de créer des modèles plus raffinés à l'intersection de ces deux approches.

2

Synthèse des contributions

Cette thèse est organisé en trois parties. Premièrement, nous analyserons les modèles de cascades de diffusion et présenterons de nouveaux résultats théoriques concernant l'influence, une caractéristique essentielle de ces processus, dans plusieurs scénarios de contagion. Ces résultats génériques nous permettront d'établir de nouvelles bornes supérieures en percolation, épidémiologie et cascades d'information. Ensuite, nous analyserons les propriétés dynamiques des processus de diffusion, et plus particulièrement leur temps critique, à savoir l'instant précis à partir duquel la contagion atteint des dimensions macroscopiques à l'échelle du réseau entier. Deuxièmement, nous développerons de nouvelles stratégies pour le contrôle des processus de diffusion, et des épidémies plus spécifiquement, utilisant des ressources de traitement de manière dynamique. Dans un deuxième temps, nous analyserons l'efficacité d'une large classe de stratégies de contrôle, et montrerons qu'une caractéristique du réseau appelée maxcut, ou coupe maximale, est étroitement liée à l'efficacité de ces stratégies. Enfin, nous développerons de nouveaux outils mathématiques pour l'analyse des réseaux et des processus de diffusion capables de décrire les propriétés limites des réseaux lorsque ceux-ci tendent vers des espaces continus. Ces outils mathématiques pourraient servir de base pour une analyse plus avancée et un contrôle des épidémies sur de très grands réseaux pour lesquels seules les données macroscopiques sont disponibles.

2.1 Analyse des modèles de diffusion

À l'intersection de la physique statistique, de l'épidémiologie et de l'informatique, la littérature relative aux processus de diffusion est vaste et variée. Du fait de leurs contraintes et particularités propres, chaque domaine a défini ses propres modèles de diffusion, et les avancées dans un domaine ne se répercutent pas toujours aux autres. Par exemple, les épidémiologistes connaissent depuis longtemps l'impact du rayon spectral de la matrice d'adjacence (voir Définition 5.5) sur le caractère explosif des épidémies, et leur capacité à se maintenir dans une population (Wang et al., 2003). Des résultats similaires ont été découverts pour les processus de Hawkes, pour lesquels le processus explose en temps fini lorsque le rayon spectral d'une certaine matrice dépasse 1. Cependant, une telle analyse faisait encore défaut pour les *cascades d'information* (voir Modèle 5.12), un modèle plus récent de contagion spécialement conçu pour les réseaux de communication et la diffusion d'information (Gomez-Rodriguez et al., 2011).

Le chapitre 5 résume les principaux concepts et modèles de réseaux et processus de diffusion, et servira d'introduction aux éléments clés de la littérature correspondante. Le reste de la thèse utilisera abondamment les modèles présentés dans cette section. Dans le

chapitre 6, nous développerons de nouvelles bornes supérieures pour l'influence dans un réseau aléatoire, et appliquerons nos résultats à plusieurs processus de diffusion, y compris la percolation, le modèle épidémiologique SIR ainsi que les cascades d'information. Enfin, dans le chapitre 7, nous étudierons les propriétés dynamiques des cascades d'information, notamment à travers la caractérisation d'un *moment critique* à partir duquel la cascade devient *sur-critique* et atteint de vastes parties du réseau.

2.1.1 Comportement long-terme de l'influence

Dans le chapitre 6, nous introduirons la notion de réseaux aléatoires avec corrélation positive locale (CPL) (voir Définition 6.6) et en déduirons des bornes supérieures non asymptotiques pour l'influence dans ce type de réseau. Le concept de réseau aléatoire avec corrélation positive locale généralise, en un sens, la description des phénomènes observés dans les domaines de la percolation, de l'épidémiologie et des cascades d'information. Les bornes supérieures obtenues dépendent du rayon spectral ρ_n d'une matrice particulière construite à partir des probabilités d'occurrence de chaque arête du réseau, appelée la matrice des risques. Nous montrerons que ces bornes révèlent trois régimes: sous-critique, critique et sur-critique, en fonction de la valeur du rayon spectral ρ_n . Pour les réseaux aléatoires de taille *n*, nous montrerons que l'influence est tout au plus un $O(\sqrt{n})$ lorsque $\rho_n < 1$, et en moyenne un O(1). Cependant, quand $\rho_n > 1$, le régime devient sur-critique et l'influence devient potentiellement linéaire en n. Plus précisément, nous montrerons que l'influence est bornée supérieurement par $\gamma_0(\rho_n)n + o(n)$, où $\gamma_0(\rho_n) \in [0,1]$ est une fonction simple (voir Définition 6.3) et que cette borne est satisfaite pour certains réseaux aléatoires. Enfin, dans le régime transitoire où $\rho_n \approx 1$, l'influence est au plus un $O(n^{2/3})$, et en moyenne un $O(\sqrt{n})$. De plus, on obtient également que la taille de ce régime intermédiaire par rapport à ρ_n est proportionnel à $n^{-1/3}$. Tab. 2.1 résume les différents comportements des bornes supérieures pour l'influence dans les réseaux aléatoires avec CPL, dans les régimes sous-critiques, critiques et sur-critiques, obtenus en Sec. 6.3. Dans le scénario aléatoire A, un ensemble de n_0 influenceurs sont tirés au hasard, alors que dans le scénario *aléatoire B* chaque nœud est influenceur avec une probabilité indépendante égale à q.

	Régimes		
Scenario	Sous-critique ($\rho_n < 1$)	Critique ($\rho_n \approx 1$)	Sur-critique ($\rho_n > 1$)
(I) Pire cas	$O(\sqrt{n})$	$O(n^{2/3})$	$\gamma_0(\rho_n)n + O(\sqrt{n})$
(II) Aléatoire A	O(1)	$O(\sqrt{n})$	$\gamma_0(\rho_n)n + O(1)$
(III) Aléatoire B	<i>O</i> (<i>qn</i>)	$O(\sqrt{q}n)$	$\gamma_0(\rho_n)n + O(qn)$

Table 2.1: Résumé des résultats concernant l'influence au sein des réseaux aléatoires à corrélation positive locale.

En corollaire, nous obtenons des bornes supérieures pour la taille de la composante géante en percolation qui améliorent de manière significative les résultats de Bollobás et al. (2010). Plus précisément, nous montrerons que le rayon spectral ρ_n est une quantité clé en percolation, et que la taille de la composante géante $C_1(G)$ est, en moyenne, bornée supérieurement par $O(\sqrt{n})$ quand $\rho_n < 1$, par $O(n^{2/3})$ quand $|\rho_n - 1| = O(n^{-1/3})$, et par $\gamma_0(\rho_n)n + o(n)$ quand $\rho_n > 1$. De plus, nous montrerons qu'une composante géante

ne peut exister que si $\limsup_{n \to +\infty} \rho_n > 1$, puis analyserons la répartition de la taille des composantes connexes en bornant supérieurement le nombre N(m) de composantes connexes de taille supérieure à m en moyenne. Tab. 2.2 résume les différents comportements des bornes supérieures pour la percolation dans les régimes sous-critiques, critiques et sur-critiques.

	Régimes		
Quantité	Sous-critique ($\rho_n < 1$)	Critique ($\rho_n \approx 1$)	Sur-critique ($\rho_n > 1$)
$\mathbb{E}[C_1(G)]$	$O(\sqrt{n})$	$O(n^{2/3})$	$\gamma_0(\rho_n)n + O(\sqrt{n})$
$\mathbb{E}[N(m)]$	$O(nm^{-2})$	$O(nm^{-3/2})$	$\gamma_0(\rho_n)n/m + O(nm^{-3/2})$

Table 2.2: Résumé des résultats en percolation: $C_1(G)$ est la taille de la composante géante, et N(m) est le nombre de composantes connexes de taille supérieure à m.

Enfin, nous appliquerons nos bornes supérieures aux propriétés long-terme du modèle d'épidémies connu sous le nom de Susceptible-Infecté-Retiré (SIR), ainsi que des cascades d'informations en temps continu et en temps discret. Plus précisément, nous améliorerons de manière significative les résultats de Draief et al. (2008) dans le régime sous-critique, et montrerons que, près du seuil épidémiologique, le nombre de nœuds infectés dans le modèle SIR est un $O(n^{2/3})$. De plus, nous étendrons le seuil épidémiologique traditionnel $\beta \rho(\mathcal{A}) = \delta$, où β et δ sont les taux de transmission et de guérison du modèle et \mathcal{A} est la matrice d'adjacence du graphe sous-jacent, à des modèles SIR plus réalistes dans lesquels la période d'incubation peut suivre une distribution non-exponentielle.

2.1.2 Propriétés dynamiques de l'influence

Dans le chapitre 7, nous étendrons la notion de *matrice des risques* afin d'analyser les propriétés dynamiques des cascades d'information en temps continu. Plus précisément, nous définirons la *transformée de Laplace de la matrice des risques* et montrerons que l'influence à l'instant *T* d'un ensemble de nœuds dépend fortement de son rayon spectral. De plus, nous définirons et caractériserons le comportement d'un *temps critique* à partir duquel les processus sur-critiques explosent. Nous montrerons que, avant ce temps, les processus surcritiques se comportent de manière sous-critique et infectent au plus un o(n) nœuds. Nous appliquerons ensuite nos bornes génériques à quatre modèles de cascades particulières dans lesquelles nos limites peuvent être explicitées: les cascades d'information en temps continu avec des probabilités de transmission exponentielles, les cascades d'information en temps discret, les modèles épidémiques Susceptible-Infecté et Susceptible-Infecté-Retiré. Tab. 2.3 résume les bornes sur le temps critique obtenues pour chaque modèle.

En outre, des bornes sur l'influence seront données pour tout temps T dans le cadre générique des cascades d'information en temps continu,

$$\sigma_{\mathcal{I}}(T) \le n_0 + \sqrt{n_0(n - n_0)} \min_{\{s \ge 0 \mid \rho(s) < 1\}} \left(\sqrt{\frac{\rho(s)}{1 - \rho(s)}} e^{sT} \right),$$

où $\sigma_{\mathcal{I}}(T)$ est l'influence du groupe de nœuds \mathcal{I} de taille n_0 à l'instant T (voir Définition 5.10), et $\rho(s)$ est le rayon spectral de la *transformée de Laplace de la matrice des risques*

Modèle	Borne inférieure sur le temps critique
Temps continu	$\ln n/2\rho_n^{-1}(1)$
Temps discret	$\ln n/2\ln \rho_{\alpha}$
Transmissions exponentielles	$\ln n/2\lambda(\rho_{\alpha}-1)$
Susceptible-Infecté-Retiré	$\ln n / \left(2(\delta + \beta)(\ln(1 + \frac{\beta}{\delta})\rho(\mathcal{A}) - 1) \right)$
Susceptible-Infecté	$\frac{\frac{1}{\beta}\sqrt{\frac{\ln n}{2\rho(\mathcal{A}_n)}}(1-e^{-\sqrt{\frac{\ln n}{2\rho(\mathcal{A}_n)}}})$

Table 2.3: Résumé des résultats concernant les bornes inférieures sur le temps critique des processus de diffusion en temps continu. $\rho_n(s)$ désigne le rayon spectral de la *transformée de Laplace de la matrice des risques* (voir Definition 7.1), ρ_{α} le rayon de la matrice des risques, λ le paramètre de la loi exponentielle, et δ et β sont les paramètres de l'épidémie.

(voir Définition 7.1). Bien qu'il ne soit pas possible, en général, d'expliciter ce minimum, nous fournirons ensuite des bornes supérieures explicites dans deux cas particuliers: les cascades d'information en temps discret

$$\sigma_{\mathcal{I}}(T) \leq n_0 + \sqrt{n_0(n-n_0)} \sqrt{\frac{2eT}{T_0}} \rho_{\alpha}^{\frac{T}{T_0}},$$

où ρ_{α} est le rayon spectral de la matrice des risques (Définition 6.2), et les cascades d'information en temps continu avec probabilités de transmission exponentielles

$$\sigma_{\mathcal{I}}(T) \leq n_0 + \sqrt{n_0(n-n_0)}\sqrt{2eT\lambda\rho_{\alpha}}e^{\lambda T(\rho_{\alpha}-1)},$$

où λ est le paramètre des lois exponentielles.

Ces résultats peuvent être utilisés de diverses façons. D'abord, ils fournissent un moyen d'évaluer les algorithmes de maximisation de l'influence sans avoir à tester tous les ensembles possibles d'influenceurs, ce qui est impossible pour de grands réseaux. Deuxièmement, le temps critique permet aux décideurs de savoir combien de temps une contagion restera dans sa phase initiale, avant de prendre des proportions pandémiques et difficilement contrôlables. Enfin, nos résultats fournissent la première formule fermée pour l'estimation, à un temps fixé, de l'influence d'une cascade d'information en temps continu. En effet, nos résultats empiriques indiquent que nos bornes sont optimales pour une grande famille de réseaux au début et à la fin du processus de diffusion.

2.2 Contrôle de processus diffusifs

L'analyse des processus de diffusion et le développement de stratégies de contrôle sont complexes et dépendent fortement de facteurs tels que: i) le type de processus de diffusion, par exemple, chaque nœud peut être sujet à des infections simples ou multiples, ii) la structure du réseau, et iii) le type d'actions de contrôle dont disposent les autorités. Notre objectif est de réaliser *un contrôle dynamique d'une épidémie* à l'aide de *l'allocation de ressources en temps réel*. A chaque instant, un certain budget de ressources est disponible et les autorités doivent décider quels nœuds vont les recevoir en fonction de l'état actuel

2.2. CONTRÔLE DE PROCESSUS DIFFUSIFS

du réseau. Ce cadre de contrôle s'applique aussi aux processus diffusifs indésirables dans un réseau social, comme la propagation de fausses rumeurs, comportements malveillants (par exemple, la violence ou le racisme) ou encore certains comportements à risques ou troubles médicaux tels que l'obésité (dont Christakis and Fowler (2007) ont récemment démontré la diffusivité dans un réseau social) ou le tabagisme.

Dans le chapitre 8, nous étudierons le *contrôle de processus diffusifs*, avec comme cadre de référence l'épidémiologie. Dans le chapitre 9, nous analyserons une classe particulière de stratégies de contrôle appelé *planification prioritaire*, et en tirerons des résultats théoriques sur l'efficacité de ces stratégies.

2.2.1 Contrôle dynamique d'épidémies SIS

Dans le chapitre 8, nous proposerons tout d'abord un modèle pour le contrôle dynamique des processus de diffusion à l'aide de *l'allocation dynamique de ressources*. Deuxièmement, nous étudierons les stratégies d'allocation dynamique et proposerons la stratégie de contrôle nommée *Réduction Maximale des Arêtes Infectieuses* (ou *Largest Reduction in Infectious Edges* en anglais) basée sur la minimisation d'une approximation de second ordre du coût associé à un processus de diffusion. Cette stratégie tire son nom du fait qu'elle minimise à court-terme le nombre *d'arêtes infectieuses* capables de transmettre l'épidémie aux nœuds sains du réseau. De cette manière, elle réduit la diffusion de l'épidémie à travers le réseau et permet un contrôle efficace du processus de diffusion. Troisièmement, notre étude expérimentale sur des réseaux réels ainsi que générés aléatoirement montre que cette stratégie se montre plus efficace que les méthodes basées sur la notion de centralité.

2.2.2 Une analyse détaillée de la planification prioritaire

Dans le chapitre 9, notre contribution majeure se situe dans l'introduction et l'analyse d'une classe particulière de stratégies pour supprimer une épidémie du type Susceptible-Infecté-Susceptible (SIS). L'administrateur du réseau a accès à un ensemble de ressources de traitement dont il peut choisir la distribution tout au long de l'épidémie. Chaque ressource a une action locale et temporelle qui peut influer sur le comportement d'un nœud du réseau. Réagir en temps réel à une diffusion rapide peut être difficile. Afin de simplifier le travail nécessaire au moment de la diffusion, nous considérons une classe simple d'allocation dynamique de ressources reposant sur un ordre de priorité calculé en amont de l'épidémie. En se focalisant sur les premiers nœuds infectés dans l'ordre de priorité, une telle stratégie supprime progressivement la diffusion jusqu'à son extinction totale. Le rôle de la coupe maximale (ou maxcut) d'un ordre de priorité est mis en évidence par des résultats théoriques le reliant étroitement au temps d'extinction de l'épidémie, lorsque le budget est limité et le réseau est totalement infecté au départ de l'épidémie. Plus précisément, nous montrerons que, sous quelques hypothèses techniques, si r est le budget de ressources disponible, β la vitesse de transmission de l'épidémie et $\mathcal{C}^*(\ell)$ la coupe maximale de l'ordre ℓ utilisé par la stratégie de planification prioritaire, alors

$$r = \beta \mathcal{C}^*(\ell)$$

est un seuil critique pour le temps d'extinction de l'épidémie (voir Définition 8.3.1, Théorème 9.4 et Théorème 9.5), séparant les régimes *sous-critique* et *sur-critique* de l'épidémie. Par conséquent, la *largeur de coupe* (ou *cutwidth*, voir Définition 9.5) du réseau détermine le budget de ressources nécessaire pour mettre fin à une épidémie à l'aide de la planification prioritaire, et la stratégie que nous proposons, appelée *Maxcut minimisation* (MCM), apparaît comme une utilisation naturelle et simple de nos résultats théoriques.

2.3 Convergence des processus diffusifs

Cette partie est consacrée à l'analyse de la convergence des réseaux. Plus généralement, notre objectif est de fournir de nouveaux outils mathématiques pour la convergence des matrices et des réseaux, et montrer que ces outils fournissent des résultats nouveaux et intuitifs, tout en étant faciles à utiliser. La convergence des espaces métriques n'est pas un sujet nouveau, et, déjà en 1981, Gromov (1981) définit et analyse une distance entre espaces métriques basée sur la projection des deux espaces métriques dans un espace commun. Toutefois, cette quantité reste relativement complexe et inadaptée aux espaces métriques non bornés. En s'appuyant sur les travaux en transport optimal et leur utilisation judicieuse des *couplages probabilistes*, nous analyserons dans ce chapitre un distance plus récemment introduite par Sturm (2006, 2013) et Mémoli (2011, 2014), à l'aide d'une approche nouvelle mettant l'accent sur la notion de *mapping probabiliste*, une manière élégante et puissante permettant de représenter une "quasi-isométrie" entre deux espaces de cardinalité différente.

Dans le chapitre 10, nous présenterons les principaux concepts utilisés pour la convergence des espaces métriques et mesurables (notées *mm-espaces*), notamment la *distance de Gromov-Wasserstein* entre deux mm-espaces, et donnerons quelques exemples de convergence de réseaux vers des espaces continus. Dans le chapitre 11, nous montrerons que de nombreuses caractéristiques de réseaux et de matrices (et leur généralisation sous forme d'opérateurs dans un mm-espace), tels que leur spectre ou les processus dynamiques qui lui sont associés, sont continues par rapport à la distance de Gromov-Wasserstein.

2.3.1 Sur la convergence des réseaux

Dans le chapitre 10, nous présenterons d'abord une nouvelle distance entre fonctions se trouvant dans deux mm-espaces différents, appelée distance de mapping, basée sur la comparaison des valeurs des deux fonctions via un couplage probabiliste entre les deux espaces. En utilisant cette distance, nous reformulerons la notion de distance de Gromov-Wasserstein comme une distance de mapping entre les mesures des deux mm-espaces. Cette nouvelle représentation de la distance de Gromov-Wasserstein nous permettra, dans le chapitre 11, de généraliser les théorèmes de continuité, non seulement aux opérateurs qui dépendent de la métrique du mm-espace, mais aussi à toute séquence d'opérateurs qui convergent par rapport à une nouvelle distance entre opérateurs, également appelée la distance de mapping en raison de son lien étroit avec la distance de mapping pour les fonctions. Cette distance entre opérateurs peut s'interpréter comme une généralisation de la norme d'opérateur permettant de comparer deux opérateurs se trouvant dans deux espaces différents. Nous montrerons ensuite que les réseaux de type grille, réseaux totalement connectés et réseaux aléatoires géométriques convergent au sens de Gromov-Wasserstein, et l'un des nouveaux résultats de cette section est que les réseaux non dirigés sont denses dans l'espace de tous les mm-espaces. Cela signifie que tout espace continu muni d'une distance et d'une mesure de probabilité est la limite d'une suite de réseaux non dirigés. Ces nouveaux outils mathématiques pourraient fournir une intuition nouvelle et de précieux résultats en théorie des matrices aléatoires, ainsi que pour l'analyse de grands réseaux, notamment les réseaux sociaux.

2.3.2 Continuité de caractéristiques clé et processus

Dans le chapitre 11, nous montrerons que les caractéristiques macroscopiques d'un réseau convergent lorsque le réseau converge au sens de Gromov-Wasserstein. Plus précisément, nous analyserons la convergence du degré moyen, le volume des boules de rayon fixé, la distribution des degrés, le rayon spectral et le diamètre. De plus, nous montrerons la convergence des propriétés spectrales de la matrice d'adjacence, y compris son spectre et ses valeurs propres, et généraliserons ce résultat aux séquences de matrices convergeant par rapport à une nouvelle notion de distance entre opérateurs, appelée distance de mapping. Enfin, nous analyserons la convergence de nombreux processus dynamiques lorsque leurs espaces sous-jacents converge, y compris les systèmes dynamiques discrets, systèmes différentiels, marches aléatoires et processus de diffusion markoviens. Ces résultats justifient l'utilisation de la distance de Gromov-Wasserstein en épidémiologie, notamment lorsque le réseau de contact est très grand, et ouvrent la voie vers des modèles plus raffinés de diffusion utilisant un mm-espace comme approximation du réseau de contact sur lequel l'épidémie se propage. L'un des principaux résultats de cette section est que, lorsqu'une séquence de matrices symétriques converge vers un opérateur d'un espace continu par rapport à la distance de mapping, leur spectre converge, au sens de la distance de Hausdorff, vers le spectre de l'opérateur limite. Ce résultat est assez général et pourra peut-être trouver des applications dans des domaines aussi divers que la théorie des matrices aléatoires ou les statistiques en grande dimension (en fournissant un comportement limite des matrices de covariance pour des données de grande dimension).

Corollaire 11.6 (Spectre des matrices symétriques). Soit F_n une séquence de matrices symétriques, et $F : L_{2,\mu_{\mathcal{X}}} \to L_{2,\mu_{\mathcal{X}}}$ un operateur d'un mm-espace \mathcal{X} . Si $\lim \Delta_2(F_n, F) = 0$,

$$d_H(sp(F_n) \cup \{0\}, sp(F) \cup \{0\}) \le \Delta_2(F_n, F),$$
(2.1)

et $sp(F_n) \cup \{0\}$ converge vers $sp(F) \cup \{0\}$.

Dans ce corollaire, sp(F) est le spectre (ponctuel) de l'opérateur linéaire F, $L_{2,\mu_{\mathcal{X}}}$ est l'espace des fonctions $f : \mathcal{X} \to \mathbb{R}$ de p-norme finie $||f||_{p,\mu_{\mathcal{X}}}$, d_H est la distance de Hausdorff entre sous-ensembles d'un espace, et $\Delta_2(F,G)$ est la distance de mapping entre deux opérateurs F et G de leurs espaces respectifs. Lorsque les opérateurs dépendent des distances de leurs espaces (par exemple les matrices d'adjacence), et que les deux espaces sont proches au sens de Gromov-Wasserstein, alors ce résultat implique que les spectres de ces opérateurs seront également proches.

Introduction (english version)

"If we knew what it was we were doing, it would not be called research, would it?" — Albert Einstein

3

Approximation is a central notion in science, and finding the right level of abstraction is often the key to a theory's success. In many scientific fields, including economy, sociology and biology, the infinite complexity of an individual is often reduced to the most simple of all creatures: a plain and anonymous node in a gigantic network. Although this approximation may seem too strong to provide any valid intuition, real-life networks, whether computer networks, social networks or bank networks, tend to share common characteristics and behave in a similar way, irrespective of the individual traits of their entities. In a certain sense, the macroscopic characteristics of these complex systems are driven, not by the local properties of its elements, but by the *complex structure of their interactions*. Predicting the behavior of epidemics, information diffusion or financial crises thus requires to understand this structure and how it impacts the dynamic properties of the network. In other words, understanding how propagation from one person to the other through the complex paths of a network may influence the temporal evolution of a diffusion, and thus drive its macroscopic behavior.

Social networks give scientists a unique opportunity to observe such diffusion processes in real-time and with a high precision. The propagation of a message among its users, when its content is public, is made available to the research community in very large databases. This wealth of data allowed the fast development of the field and its applications in marketing. However, the theoretical analysis of these diffusion processes did not benefit from it, and part of this thesis consisted in understanding which structural characteristics of networks determine the macroscopic behavior of such information cascades.

In epidemiology, theoretical analysis is only the first step towards effective prediction and preventive measures. The main problem with the corresponding literature is that most works assume that the network of interactions is known. Although inference algorithms may provide good network candidates in situations in which data is available in large quantities (e.g. social networks), sometimes the underlying network is way beyond our reach. In epidemiology, a common practice to circumvent this difficulty is to simplify the contact network in a population by a network of cities or states, upon which the epidemic is assumed to follow a simple epidemic model. Part of our work in this thesis was devoted to bridging the gap between simplified and exact epidemic models, and may allow to create more refined models at the intersection of both approaches.

4

Summary of contributions

This thesis is organized in three parts. First, we analyze models of diffusion cascades and derive new theoretical results concerning the influence -a key characteristic of such processes- in several contagion scenarios. These results allow us to derive new upper bounds in percolation theory, epidemiology, and information cascades. Moreover, we characterize the dynamic properties of diffusion processes by analyzing their *critical time*, i.e. the time when the contagion reaches a non-negligible size. Second, we develop new strategies for the control of diffusion processes, and more specifically epidemics, using treatment resources dynamically. Then, we analyze the quality of a large class of control strategies, and show that a particular network characteristic can effectively describe their efficiency. Third, we develop new mathematical tools for the analysis of networks and diffusion processes capable of describing *limit properties* of networks when these tend to continuous spaces. We consider these mathematical tools as the foundation of more advanced analysis and control of epidemics on extremely large networks for which only macroscopic data is available.

4.1 Analysis of diffusion processes

At the intersection of many scientific fields, including statistical physics, epidemiology, computer science and data mining, the literature of diffusion processes is large and diverse. Unfortunately, due to their specific constraints and limitations, each field defined its own models of diffusion, and the new discoveries in one field are not always applied to the others. For example, the epidemiology literature discovered the impact of the spectral radius of the adjacency matrix (see Definition 5.5) on diffusion processes, and their ability to explode and sustain in a network, more than ten years ago in the work of Wang et al. (2003). A similar analysis was available for Hawkes processes, for which the process explodes in finite time if the spectral radius of a certain matrix reaches 1. However, such an analysis was still lacking for *information cascades* (see Model 5.12), a newer model of contagion specifically designed for communication networks and the spread of information (Gomez-Rodriguez et al., 2011).

Chap. 5 summarizes the main concepts and models of networks and diffusion processes, and will serve as an introduction to the key elements of the literature. The rest of the thesis will use the models presented in this section and heavily depend on them. In Chap. 6, we will develop new upper bounds for the influence in random networks, and apply them to several diffusion processes including percolation, epidemiology models and information cascades. Finally, Chap. 7 will investigate the dynamic properties of information cascades, notably through the characterization of a *critical time* at which the cascade becomes *super-critical* and reaches large areas of the network.

4.1.1 Long-term behavior of the influence

In Chap. 6, we introduce the notion of random graphs with Local Positive Correlation (LPC) (see Definition 6.6) and derive non-asymptotic upper bounds for the influence in this setup. The concept of random graphs with LPC unifies, in some sense, the description of the phenomena observed in the fields of percolation theory, epidemiology and information cascades. The upper bounds obtained depend on the spectral radius ρ_n of a particular matrix built from the edge probabilities, called the *Hazard matrix*. We show that such bounds reveal three regimes: subcritical, critical and supercritical, depending on the value of the spectral radius ρ_n . For random graphs with *n* vertices, we show that the influence is at most a $O(\sqrt{n})$ when $\rho_n < 1$, and in average a O(1). However, when $\rho_n > 1$, the regime becomes supercritical as the influence becomes potentially linear in n. More specifically, we show that the influence is upper bounded by $\gamma_0(\rho_n)n + o(n)$, where $\gamma_0(\rho_n) \in [0,1]$ is a simple function (see Definition 6.3) and that this bound is met for particular random graphs. Finally, in the transitional regime where $\rho_n \approx 1$, the influence is at most a $O(n^{2/3})$, and in average a $O(\sqrt{n})$. Moreover, we also obtain that the size of this intermediate regime w.r.t. ρ_n is proportional to $n^{-1/3}$. Tab. 4.1 summarizes the different behaviors of upper bounds for influence in random graphs with LPC, in the subcritical, critical and supercritical regimes, as provided in Sec. 6.3. In the Random A scenario, a set of n_0 influencers are drawn at random, while in *Random B* each node belongs to the influencer set with independent probability q.

	Regimes		
Scenario	Subcritical ($\rho_n < 1$)	Critical ($\rho_n \approx 1$)	Supercritical ($\rho_n > 1$)
(I) Worst-case	$O(\sqrt{n})$	$O(n^{2/3})$	$\gamma_0(\rho_n)n + O(\sqrt{n})$
(II) Random A	O(1)	$O(\sqrt{n})$	$\gamma_0(\rho_n)n + O(1)$
(III) Random B	O(qn)	$O(\sqrt{q}n)$	$\gamma_0(\rho_n)n + O(qn)$

Table 4.1: Summary of results for influence in random graphs with LPC.

As a corollary, we derive upper bounds for the size of the giant component in bond and site percolation which significantly improve the previous results of Bollobás et al. (2010). More specifically, we show that the spectral radius ρ_n is a key quantity for percolation, and that the size of the giant component $C_1(G)$ is, in expectation, upper bounded by a $O(\sqrt{n})$ when $\rho_n < 1$, by a $O(n^{2/3})$ when $|\rho_n - 1| = O(n^{-1/3})$, and by $\gamma_0(\rho_n)n + o(n)$ when $\rho_n > 1$. Moreover, we prove that a giant component can only exist if $\limsup_{n \to +\infty} \rho_n > 1$. Also, we analyze the distribution of the size of connected components by upper bounding the number N(m) of connected components of size bigger than m in expectation. Tab. 4.2 summarizes the different behaviors of the upper bounds in the subcritical, critical and supercritical regimes, derived for percolation.

Finally, we apply our upper bounds to the late-time properties of the Susceptible-Infected-Removed (SIR) epidemic model, as well as discrete and continuous-time information cascades. More specifically, we significantly improve the results of Draief et al.

	Regimes		
Quantity	Subcritical ($\rho_n < 1$)	Critical ($\rho_n \approx 1$)	Supercritical ($\rho_n > 1$)
$\mathbb{E}[C_1(G)]$	$O(\sqrt{n})$	$O(n^{2/3})$	$\gamma_0(\rho_n)n + O(\sqrt{n})$
$\mathbb{E}[N(m)]$	$O(nm^{-2})$	$O(nm^{-3/2})$	$\gamma_0(\rho_n)n/m + O(nm^{-3/2})$

Table 4.2: Summary of results for bond and site percolation: $C_1(G)$ is the size of the giant component, and N(m) is the number of connected components of size bigger than m.

(2008) in the subcritical regime, and show that, near the epidemic threshold, the number of infected nodes in the SIR model is a $O(n^{2/3})$. Furthermore, we extend the traditional epidemic threshold in $\beta \rho(\mathcal{A}) = \delta$, where β and δ are the transmission and recovery rates and \mathcal{A} is the adjacency matrix of the underlying graph, to more realistic SIR models in which the incubation period may follow a non-exponential distribution.

4.1.2 Dynamic properties of the influence

In Chap. 7, we extend the notion of *Hazard Matrix* in order to analyze the dynamic properties of continuous-time information cascades. More specifically, we define the *Laplace Hazard matrix* and show that the influence at time T of any set of nodes heavily depends on its spectral radius. Moreover, we reveal the existence and characterize the behavior of *critical times* at which super-critical processes explode. We show that, before these times, supercritical processes will behave sub-critically and infect at most o(n) nodes. We then apply our generic bounds to four particular cascade models in which our bounds can be made explicit: continuous-time information cascades with exponential transmission probabilities, discrete-time information cascades, Susceptible-Infected and Susceptible-Infected-Removed epidemic models. Tab. 4.3 summarizes the bounds obtained on the critical time for each model.

Model	Critical time lower bound
Continuous-time	$\ln n/2\rho_n^{-1}(1)$
Discrete-time	$\ln n/2\ln \rho_{\alpha}$
Exponential transmissions	$\ln n/2\lambda(ho_{lpha}-1)$
Susceptible-Infected-Removed	$\ln n / \left(2(\delta + \beta)(\ln(1 + \frac{\beta}{\delta})\rho(\mathcal{A}) - 1) \right)$
Susceptible-Infected	$\frac{1}{\beta}\sqrt{rac{\ln n}{2 ho(\mathcal{A}_n)}}(1-e^{-\sqrt{rac{\ln n}{2 ho(\mathcal{A}_n)}}})$

Table 4.3: Summary of results for lower bounds on the critical time in continuous-time information cascades. $\rho_n(s)$ denotes the spectral radius of the *Laplace Hazard matrix* (see Definition 7.1), ρ_{α} the Hazard radius, λ the parameter of the exponential probabilities, and δ and β are epidemic parameters.

Furthermore, influence bounds are provided for any time T in the generic setting of

continuous-time information cascades,

$$\sigma_{\mathcal{I}}(T) \le n_0 + \sqrt{n_0(n - n_0)} \min_{\{s \ge 0 \mid \rho(s) < 1\}} \left(\sqrt{\frac{\rho(s)}{1 - \rho(s)}} e^{sT} \right),$$

where $\sigma_{\mathcal{I}}(T)$ is the influence of the set \mathcal{I} of n_0 nodes at time T (see Definition 5.10), and $\rho(s)$ is the spectral radius of the *Laplace Hazard matrix* (see Definition 7.1). Although this minimum may not be explicit in general, we then provide explicit upper bounds in two particular cases: the discrete-time information cascades

$$\sigma_{\mathcal{I}}(T) \leq n_0 + \sqrt{n_0(n-n_0)} \sqrt{\frac{2eT}{T_0}} \rho_{\alpha}^{\frac{T}{T_0}},$$

where ρ_{α} is the Hazard radius (Definition 6.2), and the continuous-time information cascade with exponential transmission probabilities

$$\sigma_{\mathcal{I}}(T) \leq n_0 + \sqrt{n_0(n-n_0)}\sqrt{2eT\lambda\rho_{\alpha}}e^{\lambda T(\rho_{\alpha}-1)},$$

where λ is the parameter of the exponential probabilities.

These results can be used in various ways. First, they provide a way to evaluate influence maximization algorithms without having to test all possible set of influencers, which is intractable for large graphs. Secondly, critical times allow decision makers to know how long a contagion will remain in its early phase before becoming a large-scale event, in fields where knowing *when* to act is nearly as important as knowing *where* to act. Finally, they can be seen as the first closed-form formula for anytime influence estimation for continuous-time information cascades. Indeed, we provide empirical evidence that our bounds are tight for a large family of graphs at the beginning and the end of the infection process.

4.2 Control strategies for diffusion processes

The analysis of diffusion processes and the design of control strategies are complex and depend heavily on factors such as: i) the type of the diffusion process, e.g. each node can be prone to single or multiple infections, ii) the network structure, and iii) the type of control actions available to authorities. Our aim is to perform *dynamic epidemic control* using *real-time resource allocation*. At each instant in time, a certain budget of resources is available and the authorities need to decide which nodes should receive them based on the current state of the network. This setting is particularly representative for the control of undesired diffusion processes in a social network, such as the spread of particular interests, malicious behaviors (e.g. violence, racism) or even health related behaviors such as obesity (recently shown to be diffusive through a social network by Christakis and Fowler (2007)) or smoking.

In Chap. 8, we study the *control of diffusion processes* and use the epidemic control as reference. In Chap. 9, we analyze a particular class of control strategies called *priority planning*, and derive theoretical results on the quality of such strategies.

4.2.1 Dynamic control of SIS epidemics

In Chap. 8, we first propose a model formulation for the dynamic control of diffusion processes as a *dynamic resource allocation* (DRA) problem. Second, we investigate DRA strategies and propose the novel *Largest Reduction in Infectious Edges* (LRIE) control strategy based on the minimization of a second-order approximation of the cost associated with a diffusion process. We explain that LRIE greedily minimizes the number of *infectious edges* that can transmit the infection from infected to healthy nodes of the network. This way, it reduces the scattering of the infection across the network and allows for efficient diffusion process control. Third, our experimental study on randomly generated and real-world networks shows that LRIE outperforms well-known centrality-based strategies, whose performance is suboptimal for this particular problem.

4.2.2 A detailed analysis of priority planning

In Chap. 9, our major contribution is the introduction and analysis of a particular class of strategies for suppressing an undesired Susceptible-Infected-Susceptible (SIS) diffusion process. We allow the network administrator to change the distribution of a set of *treat-ment resources* during the diffusion. Each resource represents a *targeted* and *temporal* action that can affect the behavior of an individual node of the network. Since reacting to fast spreading phenomena is difficult to achieve, we consider a simple class of dynamic resource allocation (DRA) strategies that rely on a *priority-order* precomputed offline. By focusing on the first infected nodes in the priority-order, such a strategy gradually suppresses the diffusion and eventually removes the contagion. The role of the *maxcut* of a considered priority-order is highlighted by developing tight bounds for the extinction time of the epidemic, when the budget is bounded and the starting state is the total infection. More specifically, we show that, under technical constraints, if *r* is the available resource budget, β the infection rate of the epidemic and $C^*(\ell)$ the maxcut of the order ℓ used by the priority planning strategy, then

$$r = \beta \mathcal{C}^*(\ell)$$

is a critical threshold for the extinction time of the epidemic (see Definition 8.3.1, Theorem 9.4 and Theorem 9.5), separating the *sub-critical* and *super-critical* behaviors of the epidemic. Hence, the network's *cutwidth* (see Definition 9.5) determines the resource budget required to suppress such epidemic under priority planning, and our proposed DRA strategy, called *maxcut minimization* (MCM), comes as a natural and straightforward utilization of our theoretical findings.

4.3 Limiting behavior of diffusion processes

This part is devoted to the analysis of the convergence of networks. More generally, our aim is to provide new mathematical tools for the convergence of matrices and networks, and show that these tools provide novel and intuitive results, while being easy to use. The convergence of metric spaces is not new and, already in 1981, Gromov (1981) defined and analyzed a distance between metric spaces based on the projection of the two metric spaces in a common space. However, this quantity remains relatively complex, and unsuited to unbounded metric spaces. Relying on the works of optimal transport and

their judicious use of *probabilistic couplings*, we analyze in this chapter a distance more recently introduced by Sturm (2006, 2013) and Mémoli (2011, 2014), using a novel presentation highlighting the notion of *probabilistic mapping*, an elegant and powerful way to represent "*near isometries*" between spaces of different cardinality.

In Chap. 10, we present the main concepts used for the convergence of metric and measurable spaces (denoted as *mm-spaces*), notably the *Gromov-Wasserstein distance* between two mm-spaces, and give a few examples of convergence of networks to continuous spaces. In Chap. 11, we show that many characteristics of networks and matrices (and generic operators of mm-spaces), such as their spectrum or associated dynamical processes, are continuous with respect to the Gromov-Wasserstein distance.

4.3.1 On the convergence of networks

In Chap. 10, we first present a novel distance between functions lying in two different mmspaces, called the *mapping distance*, by comparing the values of the two functions through a probabilistic coupling. Using this distance, we reformulate the notion of Gromov-Wasserstein distance as a mapping distance between the metrics of the two mm-spaces. This new understanding of the distance will allow us, in Chap. 11, to generalize the continuity theorems, not only to operators that depend on the metric of the mm-space, but to any sequence of operators that converge with respect to a new distance between operators, also named the mapping distance for its relation to the mapping distance between function. This distance between operators can be seen as a generalization of the operator norm to operators lying in two different spaces. We then show that grid graphs, totally connected graphs and geometric random graphs converge in the Gromov-Wasserstein sense, and one of the new results of this section is that undirected graphs are dense in the space of all mm-spaces. This means that every continuous space equipped with a metric and a probability measure is the limit of a converging sequence of undirected graphs. These new mathematical tools may provide a new intuition and valuable results for random matrix theory and the analysis of large graphs such as social networks.

4.3.2 Continuity of key characteristics and processes

In Chap. 11, we show that the macroscopic characteristics of a network converge when the network converges in the Gromov-Wasserstein sense. More specifically, we analyze the convergence of the average degree, volume of balls of fixed radius, degree distribution, spectral radius and diameter. Moreover, we show the convergence of the spectral properties of the adjacency matrix, including its spectrum and eigenvalues, and generalize this result to any sequence of matrices that converge with respect to a novel notion of distance between operators, called the mapping distance. Finally, we analyze the convergence of many dynamic processes when their underlying space converges, including discrete dynamical systems, differential systems, random walks and Markovian diffusion processes. These results motivate the use of the Gromov-Wasserstein distance for epidemiology studies on very large networks, and open the way to more refined models of diffusion using mm-spaces as an approximation of the contact graph on which the epidemic is propagating. One of the major results of this section is that, when a sequence of symmetric matrices converges to an operator of a continuous space with respect to the mapping distance, then their spectrum converge in the Hausdorff sense to the spectrum of the limit operator. This result is fairly general and may find applications in fields as diverse as random matrix theory or high-dimensional statistics (by providing a limiting behavior of covariance matrices of high-dimensional data).

Corollary 11.6 (Spectrum of symmetric matrices). Let F_n be a sequence of symmetric matrices, and $F : L_{2,\mu_{\mathcal{X}}} \to L_{2,\mu_{\mathcal{X}}}$ an operator on an mm-space \mathcal{X} . If $\lim_{n \to +\infty} \Delta_2(F_n, F) = 0$, then

$$d_H(sp(F_n) \cup \{0\}, sp(F) \cup \{0\}) \le \Delta_2(F_n, F), \tag{4.1}$$

and $sp(F_n) \cup \{0\}$ converges to $sp(F) \cup \{0\}$.

In this corollary, sp(F) is the (point) spectrum of the linear operator F, $L_{2,\mu_{\mathcal{X}}}$ is the space of functions $f : \mathcal{X} \to \mathbb{R}$ of finite *p*-norm $||f||_{p,\mu_{\mathcal{X}}}$, d_H is the Hausdorff distance between subsets, and $\Delta_2(F,G)$ is the mapping distance between two operators F and G on their respective spaces. When the operators depend on the distances of their spaces (e.g. adjacency matrices), and the two spaces are close with respect to the Gromov-Wasserstein distance, then this result implies that the spectrums of these operators will also be close.
Part I

Analysis of diffusion processes in networks

Models of networks and diffusion

"Thought is an infection. In the case of certain thoughts, it becomes an epidemic." — Wallace Stevens

5

Contents

F 1	Introduction 20		
5.1	muou		39
5.2	Netwo	ork models and characteristics	40
	5.2.1	Basic definitions	40
	5.2.2	Spectral radius	42
	5.2.3	Connected components	42
	5.2.4	Diameter	43
	5.2.5	Degree distribution	43
	5.2.6	Other characteristics	44
5.3	Random networks		
	5.3.1	Erdös-Rényi random graphs	44
	5.3.2	Preferential attachment	45
	5.3.3	Configuration model	46
	5.3.4	Geometric random graphs	47
	5.3.5	Other models of random graphs	48
5.4	Diffusion processes		
	5.4.1	Percolation models	49
	5.4.2	Epidemic models	50
	5.4.3	Information cascades	52
	5.4.4	Multivariate Hawkes processes	53
		*	

5.1 Introduction

Our aim in this chapter is to present the main concepts and definitions of network theory and diffusion processes, as clearly and as precisely as possible. We will then use these definitions in the next chapters for the analysis and control of diffusion processes. This chapter is also the occasion to present a coherent view of the literature, which has the disadvantage of being at the intersection of many different fields including epidemiology, computer science and statistical physics. Thus definitions can sometimes be incoherent between two fields, or too specific (e.g. restricted to undirected networks, suited to epidemic processes only or a definition of random graphs restricted to *constructive* algorithms). We will thus put a particular emphasis on using a unique mathematical framework throughout this entire manuscript, and give proper mathematical definitions of the concepts used in the literature.

General notations

For any set X, we will denote as card(X) its number of elements, $\mathcal{P}_n(X)$ the set of all subsets of X of size n and $X \setminus Y$ the complementary subset of Y in X. We will also use the abbreviation $[|n|] = \{1, ..., n\}$ the set of all integers between 1 and n, and $\mathbb{I}\{\cdot\}$ the indicator function. We will say that a property A holds *almost surely* (abbreviated as a.s.) if $\mathbb{P}(A) = 1$, and that a sequence of properties A_n holds *asymptotically almost surely* (abbreviated as a.a.s.) if $\lim_{n \to +\infty} \mathbb{P}(A_n) = 1$.

5.2 Network models and characteristics

The notion of a network is used in many contexts and scientific fields (biological networks, contact networks, neural networks), technological applications (sensor networks, computer networks, the internet), as well as day-to-day discussions (social networks, transportation networks, business "networking"). All these concepts share a common property, that can be considered as the core definition of a network: they form a *set of interconnected entities*. The simplest way to encode such a notion, are arguably the most widely used an accepted, is through a mathematical concept called a *graph*.

5.2.1 Basic definitions

A graph describes a network by specifying pairs of entities, denoted a *nodes*, that are connected to one another. This connection can be symmetric (e.g. neighborhood or friendship) or asymmetric (e.g. prey vs. predator in a food web, or followee vs. follower in an online social network). We now provide definitions for directed, undirected and weighted graphs.

Definition 5.1 (Directed graph). A directed graph $\mathcal{G} = (\mathcal{V}, \vec{\mathcal{E}})$ is defined via a finite set of nodes \mathcal{V} and a set of edges $\vec{\mathcal{E}} \subset \mathcal{V} \times \mathcal{V}$, i.e. pairs of nodes that are considered neighbors. The size of \mathcal{G} denotes the number of nodes of \mathcal{G} , i.e. $\operatorname{card}(\mathcal{V})$.

We will always assume that a graph has no *self-loops* (i.e. $\forall u \in \mathcal{V}, (u, u) \notin \mathcal{E}$), and no multiple edges on the same pair of nodes. The following definition encodes the notion of a graph whose connections between entities are symmetric.

Definition 5.2 (Undirected graph). An undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a directed graph whose edge set is symmetric, i.e. $\forall (u, v) \in \mathcal{E}, (v, u) \in \mathcal{E}$.



Figure 5.1: Graphical representation of a directed (a), undirected (b) and weighted (c) graph. Directed edges are represented by arrows, and the color intensity represents the weight of the edge.

The outgoing (resp. ingoing) degree of a node v will denote the number of outgoing (resp. ingoing) edges from this node:

$$d_v^{O} = \operatorname{card}(\{(u, w) \in \mathcal{E} : u = v\}),$$
(5.1)

$$d_v^{I} = \operatorname{card}(\{(u, w) \in \mathcal{E} : w = v\}).$$
(5.2)

In the case of undirected graphs, these two quantities are equal, and we will use the term *degree* of node v, noted d_v , for this value. In many applications, the *importance* of a connection between two nodes is variable (e.g. close friends and distant relations will not have the same impact on a person's decisions), and providing each edge with a *weight* is a very natural way to take this imbalance into account.

Definition 5.3 (Weighted graph). A weighted graph $\mathcal{G} = (\mathcal{V}, \vec{\mathcal{E}}, w)$ is a directed graph and a weight vector $w \in \mathbb{R}^{\operatorname{card}(\vec{\mathcal{E}})}_+$ that assigns a non-negative weight for each edge of the network. For each edge $e \in \mathcal{E}$, w_e will denote the weight assigned to edge e.

Unless specified otherwise, we will use the generic term "graph" for a undirected graph, and specify with directed or weighted otherwise. Also, we will usually assume that the nodes of a graph $\mathcal{G} = (\mathcal{V}, \vec{\mathcal{E}})$ of size *n* are indexed from 1 to *n*, i.e. $\mathcal{V} = \{v_1, v_2, ..., v_n\}$, and node *i* will refer to node v_i . A compact and very usefull description of the edge set is called the *adjacency matrix*.

Definition 5.4 (Adjacency matrix). Let $\mathcal{G} = (\mathcal{V}, \vec{\mathcal{E}})$ be a directed graph of size n, and $\mathcal{V} = \{v_1, v_2, ..., v_n\}$ an indexing of the nodes from 1 to n. The adjacency matrix A of \mathcal{G} is an $n \times n$ matrix whose coefficients are: $\forall i, j \in \{1, ..., n\}$,

$$A_{ij} = \begin{cases} 1 & if(v_i, v_j) \in \mathcal{E} \\ 0 & otherwise \end{cases}$$
(5.3)

Many characteristics of graphs can be defined using this matrix, making it a cornerstone of network analysis. For example, the number of edges $E = \text{card}(\mathcal{E}) = ||A||_1 = \sum_{i,j} A_{ij}$, and the outgoing (resp. ingoing) degree of a node is $d_i^O = \sum_j A_{ij}$ (resp. $d_i^I = \sum_j A_{ji}$). Note that the notion of adjacency matrix can also be extended to weighted graphs by associating each edge *e* to its weight w_e instead of 1 in the matrix. Networks of billions of nodes, such as modern social networks, can be extremely highdimensional objects, and very difficult to represent. A large number of characteristics were thus defined in order to provide a better understanding of such graphs, as well as intuition on the *structure* of the network. We now discuss several characteristics of particular interest for diffusion processes.

5.2.2 Spectral radius

One of the most important characteristics for describing the behavior of diffusion processes is the *spectral radius* (see for example Van Mieghem et al. (2009); Prakash et al. (2012); Lemonnier et al. (2014); Scaman et al. (2015b)). This quantity is tightly connected to the explosive behavior of diffusion process, and was found to act as a threshold value between *sub-critical* and *super-critical* epidemics (see Chap. 6 for more details).

Definition 5.5 (Spectral radius). The spectral radius $\rho(G)$ of an undirected graph G is the spectral radius of its adjacency matrix A:

$$\rho(\mathcal{G}) = \max|\lambda_i|,\tag{5.4}$$

where $\{\lambda_1, ..., \lambda_n\}$ are the eigenvalues of *A*.

Note that, for undirected networks, the adjacency matrix A is symmetric and, thus, can be diagonalized. Since epidemiology is mostly interested in undirected networks, the spectral radius was mainly designed for such networks, and a widely adopted generalization to directed networks is still lacking. Our research indicates that the spectral radius of the *symmetrized* adjacency matrix $\frac{A+A^{\top}}{2}$ is a good candidate (or, in a more general setting, the symmetrized Hazard matrix, see Sec. 6.2).

5.2.3 Connected components

For an undirected graph, a connected component is a maximal set of nodes in which each node is connected to any other via a *path*.

Definition 5.6 (Path). Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a (possibly directed) graph, and $u, v \in \mathcal{V}$ two nodes. A path from u to v in \mathcal{G} is a sequence $(v_0, ..., v_K)$ of nodes such that:

- $v_0 = u$ and $v_K = v$,
- $\forall k \in \{0, ..., K-1\}, (v_k, v_{k+1}) \in \mathcal{E}.$

The set of all paths from u to v in G is denoted $Q_{u,v}^{\mathcal{G}}$.

Using paths, we can now give a formal definition of connected components:

Definition 5.7 (Connected component). Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph. A connected component $\mathcal{V}_C \subset \mathcal{V}$ is a set of nodes such that: $\forall u \in \mathcal{V}_C$,

- $\forall v \in \mathcal{V}_C, \mathcal{Q}_{u,v}^{\mathcal{G}} \neq \emptyset$,
- $\forall v \notin \mathcal{V}_{C}, \mathcal{Q}_{uv}^{\mathcal{G}} = \emptyset$.

5.2. NETWORK MODELS AND CHARACTERISTICS

The whole graph is a partition (i.e. disjoint union) of connected components, and the largest connected component is sometimes referred to as *giant component*. The *number of connected components*, and their respective *sizes*, informs us about the *connectivity* of a network, and the size of the largest component is a important quantity for understanding the potential spread of diffusion processes (this is one of the main quantities studied by *percolation theory*, see Sec. 6.4). Finally, when a graph has only one connected component, it is said to be *connected*.

5.2.4 Diameter

The diameter of a graph is the distance between its two most separated nodes:

$$diam(\mathcal{G}) = \max_{u,v\in\mathcal{V}} d(u,v),\tag{5.5}$$

where d(u, v) is the *shortest-path distance* between *u* and *v*, i.e.

$$d(u,v) = \min\{K \ge 0 : \exists (v_0, ..., v_K) \in \mathcal{Q}_{u,v}^{\mathcal{G}}\}.$$
(5.6)

When the graph is not connected, its diameter is infinite, and the diameter of its connected components may provide useful information. In real social networks such as *Twitter* or *Facebook*, the diameter of the giant component was shown to be relatively low, of the order of $\log n$, where *n* is the size of the network (Ugander et al., 2011).

5.2.5 Degree distribution

The degree distribution provides valuable information about the local characteristics of a network: How connected is it? Is the network uniform or presenting denser areas? Are there hubs in the network? Does the network verify some *scale-free* property?

Definition 5.8 (Degree distribution). The degree distribution of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of size *n* is the empirical distribution of its degrees:

$$\pi_D = \frac{1}{n} \sum_{v \in \mathcal{V}} \delta_{d_v} , \qquad (5.7)$$

where d_v is the degree of node v and δ_x is the Dirac probability measure centered on x (i.e. for every borel set A of \mathbb{R} , $\delta_x(A) = \mathbb{1}\{x \in A\}$).

While networks with a small maximum degree will tend to have a relatively uniform structure (e.g. without extremely large hubs that completely dominate the evolution of diffusion processes), a *skewed* degree distribution indicates a network with denser areas (called *center*) and more sparse regions (called *periphery*). Also, the graph is called *scale-free* when the degree distribution exhibits a power-law behavior:

$$\pi_D(]x, +\infty)) = \theta(x^{-\eta}), \tag{5.8}$$

for a fixed parameter $\eta > 0$, that is, $\pi_D(]x, +\infty)x^{\eta}$ is lower and upper bounded for x sufficiently large. Such a power-law degree distribution was observed on many real networks including modern social networks (Ugander et al., 2011) and the internet (Barabási and Albert, 1999), and may indicate a certain preferential attachment procedure for the generation of such graphs (also known as "the rich get richer" effect, see Sec. 5.3.2).

5.2.6 Other characteristics

Due to the large number of network characteristics used in the literature, we omit here many quantities describing the *connectedness*, *clustering*, or *homogeneity* of networks. The interested reader will find information about cliques, clusters, k-cores or centrality measures in the excellent introductory books by Newman (2010) and Kolaczyk (2009).

5.3 Random networks

For any n > 0, let \mathbb{G}_n be the set of all graphs $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of size n, and $\mathbb{G} = \bigcup_{n>0} \mathbb{G}_n$ the set of all possible graphs. Since each set \mathbb{G}_n is finite, \mathbb{G} is countable, and \mathbb{G} can be turned into a probability space using the discrete sigma-algebra $\Sigma = \mathcal{P}(\mathbb{G})$. Note that this sigma-algebra is the only sigma-algebra that contains single elements, i.e. $\forall \mathcal{G} \in \mathbb{G}, \{\mathcal{G}\} \in \Sigma$, and this choice is thus rather natural. A random graph is then a random element in this set of all possible graphs.

Definition 5.9 (Random graph). A random graph G is a random variable in the space of all possible graphs \mathbb{G} equipped with the discrete sigma-algebra $\Sigma = \mathcal{P}(\mathbb{G})$.

We will usually consider random graphs of fixed size, i.e. random graphs $G = (\mathcal{V}, \mathcal{E})$ such that $\operatorname{card}(\mathcal{V}) = n$ a.s. for a certain n > 0. These random graphs are entirely characterized by their *random* adjacency matrices $A \in \{0,1\}^{n^2}$. We will thus use the notation G(n, A) to denote a random graph of size n and random adjacency matrix A, where A_{ij} are Bernoulli random variables indicating the presence or absence of edge (i, j) in the random graph. We will call *undirected* a random graph whose adjacency matrix is *symmetric* almost surely, i.e. $\forall i, j, A_{ij} = A_{ji}$ a.s.. Note that, in general, the edge variables A_{ij} are correlated.

Random graphs are usually created in a *constructive way*, by providing a (randomized) algorithm capable of creating an instance of the random graph. We now provide four standard random graphs of the network literature.

5.3.1 Erdös-Rényi random graphs

The simplest example of undirected random graph is the graph named after Erdös and Rényi (1960). This undirected random graph is usually constructed by connecting each pair of nodes with independent and identical probability $p \in [0,1]$. We now provide a more formal definition of such a random graph.

Model 5.1 (Erdös-Rényi random graph). For n > 0 and $p \in [0,1]$, an Erdös-Rényi random graph $G_{ER}(n,p)$ is an undirected random graph of fixed size n such that its symmetric adjacency matrix A has independent and identically distributed (i.i.d.) coordinates $\{A_{ij} : i < j\}$ and $\mathbb{E}[A_{ij}] = p$.

When p = 1/2, $G_{ER}(n, 1/2)$ is uniformly distributed over the set \mathbb{G}_n of all graphs of size *n*, and is thus the *most random* possible graph. This is, along with its simplicity, one of the main reasons for its popularity, and why much attention was devoted to its analysis. As such, the Erdös-Rényi random graph gives intuition into *average* graphs of size *n* and



Figure 5.2: Erdös-Rényi random graph of 1000 nodes and parameter p = 0.004. The network is unstructured and highly entangled (dark colors indicate high degrees).

the structure of the vast majority of graphs. Note that the original definition of Erdös-Rényi random graphs considered a fixed number of edges m, although this definition is less practical and has very similar properties.

5.3.2 Preferential attachment

Preferential attachment graphs were introduced by Barabási and Albert (1999), in order to model a growing network whose new edges tend to concentrate on nodes with the highest degree. Thus an accumulation effect appears and the degree distribution of such networks tends to be skewed, with a *central* region of the network containing highly connected hubs, and more *peripheral* regions of nodes with only a few connections. More specifically, the degree distribution was shown to exhibit a *power law*, which made this model a particularly well-chosen candidate for social networks generative models (Barabási and Albert, 1999). We now provide the Barabási-Albert model for preferential attachment graphs.

Model 5.2 (Preferential attachment model). Start with a graph $G_{PA}(1, A^1)$ of one node and no edges ($A^1 = 0$). Then, at each iteration, add one node of degree m > 0 to the graph, and connect this node to other nodes of the network with probability proportional to their degree:

$$A_{ij}^{n+1} = \begin{cases} \mathbbm{1}\{j \in \mathcal{N}_{n+1}\} & if i = n+1\\ \mathbbm{1}\{i \in \mathcal{N}_{n+1}\} & if j = n+1\\ A_{ij}^n & otherwise \end{cases}$$
(5.9)

where \mathcal{N}_{n+1} is a set of exactly *m* nodes between 1 and *n* (or $\mathcal{N}_{n+1} = \{1, ..., n\}$ if $m \ge n$) chosen at random, so that

$$\mathbb{P}(i \in \mathcal{N}_{n+1}) \propto d_i. \tag{5.10}$$

The resulting random graph $G_{PA}(n, A^n)$ at iteration n has n nodes and nm - m(m+1)/2 edges almost surely (when $n \ge m$).



Figure 5.3: Preferential attachment graph of 1000 nodes and parameter m = 4. High degree nodes tend to cluster and form a highly connected center (dark colors indicate high degrees).



Figure 5.4: Configuration model of 1000 nodes with 10 nodes of degree 100 and 990 of degree 2 (dark colors indicate high degrees).

5.3.3 Configuration model

Erdös-Rényi and preferential attachment graphs tend to have a fixed degree distribution (a Bernoulli distribution of parameters n and p for Erdös-Rényi, and a power-law for preferential attachment). However, real networks usually exhibit more complex degree distributions, and, in order to improve the accuracy and predictive power of random graph models, Bender and Canfield (1978), and later Molloy and Reed (1995, 1998), introduced the *configuration model*. This random graph models a purely random graph in the set of all random graphs of fixed degree distribution.



Figure 5.5: Geometric random graph of 1000 nodes on the 2D square $[0,1]^2$ and neighborhood radius r = 0.05. The structure of the underlying space is clearly visible (dark colors indicate high degrees).

Model 5.3 (Configuration model). The configuration model G_{CM} of size n and degree distribution $\pi_D = \sum_v \delta_{d_v}$ is a graph drawn uniformly at random among the graphs of size n and degree distribution π_D .

The configuration model is usually generated by considering that each node v has d_v half-edges attached to it, and then pairing half-edges together at random, until no half-edge is left. the resulting graph (except for eventual self-loops that should be removed) is drawn according to the configuration model.

5.3.4 Geometric random graphs

The main modeling issue with the three previous random graphs is that they tend to create *locally tree-like* graphs, i.e. graphs whose neighborhoods around each node *resembles* a tree. For example, these random networks have a very low number of triangles, although triangles are largely present in real social networks (Ugander et al., 2011). We thus present one alternative that maps nodes on a metric space (usually a two-dimensional space) and derives edges based on the distances between pairs of nodes.

Model 5.4 (Geometric random graph). Let (\mathcal{X},d) be a metric space equipped with a probability measure μ , and r > 0 a positive value. The geometric random graph $G_{GR}(\mathcal{X},n,r)$ is an undirected random graph of size n and adjacency matrix

$$A_{ij} = \mathbb{1}\{d(X_i, X_j) \le r\}, \tag{5.11}$$

where $(X_i)_{i \in \{1,...,n\}}$ are *n* i.i.d. random variables drawn according to μ .

These networks are useful for representing sensor or contact networks, that are largely determined by the underlying space in which the agents (sensors or people in a population) interact. A generalization of this model is known as *inhomogeneous random graphs*



Figure 5.6: Inhomogeneous random graph of 1000 nodes drawn randomly on the 2D square $[0,1]^2$ and kernel function $\kappa(x,y) = \exp(-20||x-y||_2)$. Compared to the geometric random graph, long distance edges can appear and perturb the structure of the network (dark colors indicate high degrees).

(Bollobás et al., 2007), in which a generic kernel function κ replaces the metric d, and edges between nodes i and j are drawn independently at random with probability $\kappa(X_i, X_j)$.

Model 5.5 (Inhomogeneous random graphs). Let (\mathcal{X}, μ) be a measurable space, and $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ a symmetric measurable kernel function, and $\mathcal{V} = (x_1, x_2, ...) \subset \mathcal{X}$ a countable node set. An inhomogeneous random graph $G_{IR}^{\mathcal{V}}(n, \kappa)$ is an undirected random graph of size n and adjacency matrix A, where each A_{ij} is an independent Bernoulli random variable of parameter $\mathbb{E}[A_{ij}] = \kappa(x_i, x_j)$.

It is moreover assumed that the empirical distribution of the node set \mathcal{V} converges weakly to the measure μ of \mathcal{X} , i.e.

$$\frac{1}{n} \sum_{i} \delta_{x_i} \stackrel{d}{\to} \mu. \tag{5.12}$$

A detailed analysis of percolation in such networks when $\kappa(x,y) = \theta(1/n)$ is available in the seminal work of Bollobás et al. (2007).

5.3.5 Other models of random graphs

While our aim in this section was to provide an overview of the some of the random graph models used in this document, the corresponding literature also contains a large number of other random graphs. To name a few, the *Stochastic Block Model* was introduced by Holland et al. (1983) and is often used for the theoretical analysis of clustering, *small world* random graphs provide a simplistic representation of communication networks (Watts and Strogatz, 1998), and parametric models such as the *Exponential Random Graph Model* allow the inference of a network using macroscopic information (Robins et al., 2007).

5.4 Diffusion processes

A diffusion process can be described as a sequence of events that *propagates* through a network by means of interconnection. For example, an epidemic spreads through a population by contact between individuals, and this process can be captured by relatively simple mathematical models. In all generality, a diffusion process on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a multivariate stochastic process $\{X_v(t) : v \in \mathcal{V}, t \ge 0\}$ where $X(t) \in \mathcal{S}$ is the state of node v at time t, and \mathcal{S} is a finite set of possible states (e.g. infected or healthy). While each model has its own equations of evolution, the key quantity is usually the number of nodes in a particular state $s \in \mathcal{S}$, called the *influence*. For epidemiology, the goal will be to reduce the number of infected nodes, while for information cascades it will be to increase the reach of an information.

Definition 5.10 (Influence). Let $s \in S$ be a state, $t \ge 0$ and $x \in S^n$ an initial state vector. The influence of x at time t, denoted $\sigma_{x,s}(t)$, is the number of nodes in state s when the process was initialized at X(0) = x:

$$\sigma_{x,s}(t) = \mathbb{E}\left[\sum_{v \in \mathcal{V}} \mathbb{1}\{X_v(t) = s\} \mid X(0) = x\right].$$
(5.13)

Usually, the initial state will be a set of initially infected nodes $\mathcal{I} \subset \mathcal{V}$, and we will use the notation $\sigma_{\mathcal{I},s}(t) = \sigma_{x_{\mathcal{I}},s}(t)$ where $x_{\mathcal{I}}$ is a state vector in which only the nodes of \mathcal{I} are infected. Also, we will usually omit the state *s* (being the *infected* state in most cases) and use the notation $\sigma_{\mathcal{I}}(t)$ hereafter. In Chap. 6, we show that, in many diffusion processes, the long-term influence $\lim_{t\to+\infty} \sigma_{\mathcal{I}}(t)$ is equivalent to the size of connected components of particular random graphs, and then provide upper bounds for the influence in this more general setting. Also, an analysis of the dynamic properties of the influence in the specific setting of information cascades (see Sec. 5.4.3) is available in Chap. 7. We now present several models of diffusion processes used in physics, epidemiology and social network analysis.

5.4.1 Percolation models

Strictly speaking, *percolation theory*, i.e. the study of connected components of random graphs, is not devoted to the analysis of diffusion processes. However, strong connections can be made between percolation and epidemic models or information cascades (see e.g. Newman (2002) or Bollobás et al. (2007)). More specifically, one of the main results of Chap. 6 is the precise mathematical connection between percolation and the *limiting behavior* of several diffusion processes (Susceptible-Infected-Removed model and information cascades, see the following sections) using the notion of *influence in random graphs*. We thus give a precise definition of *bond* and *site* percolation graphs.

Model 5.6 (Bond percolation). A bond percolation graph is an undirected random graph G = G(n, A) of size n with independent edge variables $\{A_{ij} : i < j\}$.

Model 5.7 (Site percolation). A site percolation graph consists in removing the nodes of an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ independently and with probability $1 - p_i$.

Intuitively, a bond (resp. site) percolation graph consists in removing edges (resp. nodes) of an original graph independently at random. Then, percolation theory describes the size (and number) of connected components of the resulting graph. A recent survey of the field can be found in the work of Saberi (2015).

5.4.2 Epidemic models

In epidemiology, several models for the propagation of a disease in a population have been developed (Newman, 2010; Kermack and McKendrick, 1932; Prakash et al., 2012), ranging from simple (e.g. SI, SIS, SIR, see Newman (2010) and Kermack and McKendrick (1932)) to more complex (e.g. SIRS, SEIR, SEIV, see Prakash et al. (2012)) diffusion mechanisms. These diffusion processes are usually modeled using *Markov processes* (Van Mieghem et al., 2009), i.e. *memoryless* stochastic processes entirely defined by their transition matrix. This transition matrix defines the probability for each node to change state during an infinitesimal time window [t, t + dt] (the simultaneous change of more than one node's state is considered improbable). In the following, we will thus use the notation

$$X_i(t): A \to B$$
 at rate $C_i(t)$

to denote the transition rate $C_i(t) \ge 0$ of node $i \in [|n|]$ at time $t \ge 0$ from state $A \in S$ to state $B \in S$. We here focus on three standard models of contagion: Susceptible-Infected (SI), Susceptible-Infected-Removed (SIR) and Susceptible-Infected-Susceptible (SIS) models. For more information on the vast epidemiology literature, we refer the reader to the very recent review by Pastor-Satorras et al. (2015).

Susceptible-Infected model

The Susceptible-Infected (SI) model is the simplest epidemic model, in which nodes can be either *Susceptible* or *Infected* (usually represented by 0 and 1, respectively). Each infected node can transmit the disease to one of its neighboring susceptible node with a rate β , and infected nodes remain infected, and thus contagious.

Model 5.8 (SI model). Let \mathcal{G} be a (possibly weighted) graph of n nodes and adjacency matrix A. The Susceptible-Infected model is a continuous-time Markov process $X(t) \in \{S,I\}^n$ with the following transition rate:

$$X_i(t): S \to I \text{ at rate } \beta \sum_j A_{ji} X_j(t),$$
 (5.14)

where β is the transmission rate of the epidemic.

Since the nodes remain infected, the network is totally infected at the end of the contagion and $\lim_{T\to+\infty} \sigma_A(T) = n$. However, the dynamics of the epidemic are far more complex, and an analysis of this model, and more specifically its explosiveness and diffusion speed, is provided in Sec. 7.4.3.

Susceptible-Infected-Removed model

The Susceptible-Infected-Removed (SIR) model (Kermack and McKendrick, 1932) is a widely used epidemic model designed for epidemic scenarios in which patients present some immunity to the disease after their recovery. In such a case, a recovered person will not transmit the disease anymore, nor will it be subject to a potential infection. An additional state is thus added to the SI model, and each node of the network is either : susceptible (S), infected (I), or removed (R). At t = 0, a subset \mathcal{I} of n_0 nodes is infected. Then, each infected node will transmit the disease to its neighbors at rate β , and recover at rate δ .

Model 5.9 (SIR model). Let \mathcal{G} be a (possibly weighted) graph of n nodes and adjacency matrix A. The Susceptible-Infected-Removed model is a continuous-time Markov process $X(t) \in \{S, I, R\}^n$ with the following transition rate:

$$X_{i}(t): S \to I \text{ at rate } \beta \sum_{j} A_{ji} X_{j}(t)$$

$$X_{i}(t): I \to R \text{ at rate } \delta,$$
(5.15)

where β is the transmission rate of the epidemic.

The graph is usually undirected, and each edge has the same transmission rate. However, more complex scenarios can be modeled using the *inhomogeneous SIR* model, in which each edge has its own transmission rate β_{ij} , and each node its recovery rate δ_i .

This epidemic model is *increasing*, in the sense that, if we order the infection states as follows: S < I < R, then X(t) is increasing w.r.t. the natural partial order on $\{S, I, R\}^n$ (i.e. $X \leq Y \Leftrightarrow \forall i, X_i \leq Y_i$). One of the effects of this property is that *each node will be infected at most once, and removed at most once.* We can thus define, for each node *i*, the time τ_i^I at which a node is infected, and the time τ_i^R at which it is removed. Note that these time may be infinite if these events never happen.

Proposition 5.1. For an SIR epidemic, the infection times τ_i^I of not initially infected nodes verify the following equality:

$$\forall i \notin \mathcal{I}, \tau_i^I = \min_{\{j \in [|n|]: T_{ji} < D_j\}} (\tau_j^I + T_{ji}), \tag{5.16}$$

where T_{ij} and D_i are independent exponential random variables of expected value $1/\beta$ and $1/\delta$, respectively, and $\tau_i^I = +\infty$ if the set $\{j \in [|n|] : T_{ji} < D_j\}$ is empty. Furthermore, the recovery time of each node *i* is

$$\tau_i^R = \tau_i^I + D_i.$$

Upper bounds for the influence of such diffusion processes are provided in Sec. 6.6, while an analysis of its dynamic properties is available in Sec. 7.4.3.

Susceptible-Infected-Susceptible model

Contrary to the SIR model, the Susceptible-Infected-Susceptible (SIS) model is designed for scenarios in which infected people do not develop permanent immunity and remain prone to multiple infections (Van Mieghem et al., 2009). Each node is either *Susceptible* or *Infected*, and the transition rates are similar to that of the SIR model, except that a recovered node is now susceptible again.

Model 5.10 (SIS model). Let \mathcal{G} be a (possibly weighted) graph of n nodes and adjacency matrix A. The Susceptible-Infected-Removed model is a continuous-time Markov process $X(t) \in \{S, I\}^n$ with the following transition rate:

$$X_{i}(t): S \to I \text{ at rate } \beta \sum_{j} A_{ji} X_{j}(t)$$

$$X_{i}(t): I \to S \text{ at rate } \delta,$$
(5.17)

where β is the transmission rate of the epidemic.

Similarly to the SIR model, the *inhomogeneous SIS* model is a generalization of SIS that allows non-uniform transmission and recovery rates. In Chap. 8 and Chap. 9, we will consider control strategies for the containment of an SIS epidemic by adding an additional control action to the model.

5.4.3 Information cascades

In information propagation theory, information cascades have emerged as a relevant model for viral diffusion of ideas and opinions (Kempe et al., 2003; Chen et al., 2009; Rodriguez and Schölkopf, 2012; Gomez-Rodriguez et al., 2011). Instead of modeling the transmission along edges via transmission rates, these models capture the precise temporal dependencies between infection events of neighboring nodes, and are thus better suited to situations in which a large training dataset is available to fit the parameters of the model. Information cascades are of two types:

Model 5.11 (Discrete-Time Information Cascades $DTIC(\mathcal{P})$). At time t = 0, only a set \mathcal{I} of influencers is infected. Given a matrix $\mathcal{P} = (p_{ij})_{ij} \in [0,1]^{n \times n}$, each node i that receives the contagion at time t may transmit it at time t + 1 along its outgoing edge $(i,j) \in \mathcal{E}$ with probability p_{ij} . Node i cannot make any attempt to infect its neighbors in subsequent rounds. The process terminates when no more infections are possible.

Model 5.12 (Continuous-Time Information Cascades $CTIC(\mathcal{F}, T)$). At time t = 0, only a set \mathcal{I} of influencers is infected. Given a matrix $\mathcal{F} = (f_{ij})_{ij}$ of non-negative integrable functions, each node *i* that receives the contagion at time *t* may transmit it at time s > t along its outgoing edge $(i, j) \in \mathcal{E}$ with stochastic rate of occurrence $f_{ij}(s - t)$. The process terminates at a given deterministic time T > 0.

Information cascades are, similarly to the SIR model, increasing stochastic processes, and each *node can be infected only once*. We can thus define, for each node *i*, the time τ_i at which it is infected. Note that this time may be infinite if the node is not infected during the contagion.

Proposition 5.2. For a continuous-time information cascade $CTIC(\mathcal{F},T)$, the infection times τ_i of non-influencer nodes verify the following equality:

$$\forall i \notin \mathcal{I}, \tau_i = \min_{j \in [|n|]} (\tau_j + T_{ji}), \tag{5.18}$$

where $T_{ij} \in \mathbb{R}_+ \cup \{+\infty\}$ are independent random variables of sub-probability density

$$p_{ij}(t) = f_{ij}(t) \exp\left(-\int_0^t f_{ij}(u) du\right).$$

5.4. DIFFUSION PROCESSES

Proof. See for example the work of Du et al. (2013a).

Note that, in general, $p_{ij}(t)$ is *not* a probability density over \mathbb{R}_+ as it does not integrate to one, and $\mathbb{P}(T_{ij} = +\infty) = 1 - \int_0^{+\infty} p_{ij}(t)dt = \exp(-\int_0^{+\infty} f_{ij}(t)dt)$. Upper bounds for the influence of such diffusion processes are provided in Sec. 6.7, while an analysis of its dynamic properties is available in Chap. 7.

5.4.4 Multivariate Hawkes processes

Multivariate Hawkes processes (Oakes, 1975; Liniger, 2009) have emerged in several fields as the gold standard to deal with sequences of correlated events, e.g. earthquake prediction (Vere-Jones, 1978), biology (Reynaud-Bouret et al., 2014), financial (Bauwens and Hautsch, 2009; Alfonsi and Blanc, 2015) and social interactions studies (Crane and Sornette, 2008). The main advantage of this model compared to information cascades is to allow multiple events (e.g. posts, likes or shares in the case of a social network) on a single node. For this reason, more and more attention is devoted to this model in the information cascades community (see for example the recent papers of Farajtabar et al. (2014) and Lemonnier et al. (2016a)).

In a multivariate Hawkes process, an event of type u (e.g. a user posting a new message on a social network) occurring at time t, will increase the conditional rate of occurrence of events of type v at time $s \ge t$ by a rate $g_{uv}(s - t)$.

Model 5.13 (Multivariate Hawkes process). A multivariate Hawkes process $N(t) = \{N_u(t) : u = 1, ..., d, t \ge 0\}$ is a d-dimensional counting process, where $N_u(t)$ represents the number of events along dimension u that occurred during time [0,t]. Each one-dimensional counting process $N_u(t)$ can be influenced by the occurrence of events along the other dimensions. If we denote as $(u_m, t_m)_{m=1}^n$ the event history of the process indicating, for each single event m, its type u_m and time of occurrence t_m , then the non-negative stochastic rate of occurrence of each $N_u(t)$ is defined by:

$$\lambda_u(t) = \mu_u(t) + \sum_{m:t_m < t} g_{u_m u}(t - t_m).$$
(5.19)

In the above, $\mu_u(t) \ge 0$ is the *natural occurrence rate* of events of type u (i.e. along that dimension) at time t, and the *triggering kernel function* evaluation $g_{vu}(s-t) \ge 0$ determines the *increase* in the occurrence rate of events of type u at time s, caused by an event of type v at a past time $t \le s$.

Remark 5.1. A major limitation of multivariate Hawkes processes is that they cannot model super-critical diffusion processes, i.e. contagions that, initialized on a few nodes, can reach a large part of network due solely to its diffusive properties. Examples of such super-critical processes are pandemics in epidemiology, or "buzz" effects in information networks (see Sec. 6.1 for more information on super-critical processes). In such a case, a multivariate Hawkes process would explode in finite time, and a constraint on the parameters of the process is usually added in order to avoid this problem.

 \square

6 Long-term behavior of the influence

"If the facts don't fit the theory, change the facts."

- Albert Einstein

Contents

6.1	Introd	uction	
	6.1.1	Related works 56	
	6.1.2	Outline	
6.2	Hazaro	d matrix, influence and LPC property	
	6.2.1	Hazard characteristics of random graphs	
	6.2.2	Reachability and influence58	
	6.2.3	Random graphs with Local Positive Correlation (LPC)60	
6.3	Non-a	symptotic upper bounds on the influence	
	6.3.1	Deterministic (worst-case) scenario	
	6.3.2	Random influencer set with fixed size62	
	6.3.3	Random influencer set with random size63	
	6.3.4	Lower bounds	
6.4	Application to bond percolation 6		
	6.4.1	Size and existence of the giant component	
	6.4.2	Number of components of cardinality larger than m 65	
6.5	Application to site percolation		
6.6	6 Application to epidemiology		
	6.6.1	Subcritical behavior in the standard SIR model 67	
	6.6.2	Behavior near the epidemic threshold	
	6.6.3	Generic incubation period	
6.7	Applic	cation to Information Cascades	
6.8	Discussion		
6.9	Proofs	5	
	6.9.1	Behavior of the Hazard function	
	6.9.2	Proofs of the upper bounds on influence	
	6.9.3	Proofs of the percolation theorems	

6.1 Introduction

Propagation models over graphs are very popular and particularly well suited to the analysis of epidemics and information cascades. Although different in many technical aspects, the models used in these two fields are similar and can be considered as particular instances of a more generic framework: the analysis of the influence of reachable sets in random networks.

6.1.1 Related works

In epidemiology, the study of diffusion models such as SI, SIS or SIR (Newman, 2010; Kermack and McKendrick, 1932) highlighted the impact of a spectral characteristic on the size of the epidemic: the spectral radius of the underlying network. Moreover, it was shown that this quantity acted as a critical threshold for the size of the epidemic (Van Mieghem et al., 2009; Prakash et al., 2012), and recent work by Draief et al. (2008) provided upper bounds that depend highly on this spectral quantity. Our work can be seen as a generalization of these works, by providing the right spectral quantity to consider in the case of more generic diffusion and percolation models.

In percolation theory, the concept of reachability characterizes the connected components of undirected graphs and the behavior of such components has been the object of several studies. For homogeneous random graphs G(n, p) where removal of edges in the fully connected graph with n vertices occurs independently for every edge with constant probability 1 - p, Erdös and Rényi (1960) showed that a phase transition occurred for $p = \frac{1}{n}$, and their results were later refined by Bollobás (1984) and Łuczak (1990) for the case pn = O(1). For inhomogeneous graphs, we refer to the work by Bollobás et al. (2007) in the special case where the number of edges E is O(n), and Bollobás et al. (2010) when $E = O(n^2)$. These references contain a number of asymptotic results (*i.e.* when $n \to \infty$) including the critical value of the percolation threshold, and upper bounds on the size of connected components.

6.1.2 Outline

This chapter is organized as follows. In Sec. 6.2, we recall the notions of reachable set and influence in random networks, and introduce a generic type of random graphs with *Local Positive Correlation* (LPC). In Sec. 6.3, we derive theoretical bounds for the influence in random graphs with LPC. Finally, in Sec. 6.4, Sec. 6.5, Sec. 6.6 and Sec. 6.7, we show that the previous results apply respectively to the fields of bond percolation, site percolation, epidemiology and information cascades, and improve existing results in these fields. This chapter is joint work with Rémi Lemonnier and Nicolas Vayatis, and is an extended version of the paper (Lemonnier et al., 2016b).

6.2 Hazard matrix, influence and LPC property

In this section, we introduce the main notations and definitions. In particular, we define two novel concepts: the *Hazard matrix*, that will play a key role in the analysis of influence in random graphs, and a generic class of random graphs with *Local Positive Correlation* (LPC).

6.2.1 Hazard characteristics of random graphs

For many diffusion models, the spectral features of the underlying graph were shown to have a drastic impact on the amplitude of the spread (see for example in the work of Van Mieghem et al. (2009) and Prakash et al. (2012) the role of the spectral radius of the adjacency matrix in the epidemiology literature). In order to generalize such results to a broader class of diffusion and percolation phenomena, we introduce two spectral characteristics that extend the spectral radius of the adjacency matrix to the analysis of the influence in random graphs: the *Hazard matrix* and the *Hazard radius*.

Definition 6.1 (Hazard matrix). For a random graph model G(n, A), the Hazard matrix \mathcal{H} is the $n \times n$ matrix whose coefficients \mathcal{H}_{ij} are defined as:

$$\mathcal{H}_{ij} = -\ln(1 - \mathbb{E}[A_{ij}]) . \tag{6.1}$$

The spectral radius of this matrix will play a key role in the quantification of the influence. We recall that for any square matrix M of size n, its spectral radius $\rho(M)$ is defined as the largest of the eigenvalues of M.

Definition 6.2 (Hazard radius). For a random graph model G(n, A) with Hazard matrix \mathcal{H} , we define the Hazard radius as:

$$\rho_n = \rho\left(\frac{\mathcal{H} + \mathcal{H}^\top}{2}\right) \,. \tag{6.2}$$

Remark 6.1. Let $P = (\mathbb{E}[A_{ij}])_{ij}$ be the expected adjacency matrix. When the P_{ij} 's are small, the Hazard matrix is very close to P. This implies that, for small values of P_{ij} , the spectral radius of H will be very close to that of P. More specifically, a simple calculation holds

$$\rho(P) \le \rho(\mathcal{H}) \le \frac{-\ln(1 - \|P\|_{\infty})}{\|P\|_{\infty}} \rho(P), \tag{6.3}$$

where $||P||_{\infty} = \max_{i,j} P_{ij}$. The relatively slow increase of $\frac{-\ln(1-x)}{x}$ for $x \to 1^-$ implies that the behavior of $\rho(P)$ and $\rho(\mathcal{H})$ will be of the same order of magnitude even for large (but lower than 1) values of $||P||_{\infty}$. Moreover, when considering a sequence of random graphs G_n , if $\lim_{n\to+\infty} ||P_n||_{\infty} = 0$, then $\rho(\mathcal{H}_n) \approx \rho(P_n)$ and the subcriticality of the influence is also induced by $\limsup_{n\to+\infty} \rho(\frac{P_n + P_n^{\top}}{2}) \leq 1$ (see Sec. 6.3).

In addition, we introduce here a useful function that will allow the simplification of the upper bounds derived in this chapter.

Definition 6.3 (Hazard function). *let* $\rho \ge 0$ *and* a > 0*. The* Hazard function $\gamma(\rho, a)$ *is defined as the unique solution in* [0,1] *of the following equation:*

$$\gamma - 1 + \exp\left(-\rho\gamma - a\right) = 0. \tag{6.4}$$

We will also use the notation $\gamma_0(\rho) = \lim_{a\to 0^+} \gamma(\rho, a)$ for the limit of the Hazard function *at* 0.

Fig. 6.1 reveals the behavior of $\gamma_0(\rho)$ and $\gamma(\rho, a)$ w.r.t. ρ . For more information on the Hazard function and the bounds used to derive the subcritical, critical and supercritical regimes, we refer to Appendix 6.9.1.



Figure 6.1: Behavior of $\gamma_0(\rho)$ and $\gamma(\rho, a)$ w.r.t. ρ .

6.2.2 Reachability and influence

In this section, we define the *influence* as the size of a *reachable set*. A node *u* is *reachable* from another node *v* if there is a path connecting *u* to *v* in the considered graph (Kolaczyk, 2009). As we will see later, this definition generalizes the notion of long-term influence $\lim_{T\to+\infty} \sigma_{\mathcal{I}}(T)$ for diffusion processes defined in Sec. 5.4, in the case of percolation, information cascades and the SIR epidemic model.

Definition 6.4 (Reachable set). Consider a random graph G(n, A). We call influencers a fixed set $\mathcal{I} \subset [|n|]$ of nodes and we define the reachable set of influencers \mathcal{I} in G(n, A) as the random set of nodes $R(\mathcal{I}, A)$ such that:

$$i \in R(\mathcal{I}, A) \iff i \in \mathcal{I} \text{ or } \prod_{q \in \mathcal{Q}_{\mathcal{I},i}} \left(1 - \prod_{(j,k) \in q} A_{jk} \right) = 0.$$
 (6.5)

where, for any node $i \in [|n|]$, the collection $Q_{\mathcal{I},i} = \{\{(i_0,i_1),(i_1,i_2),...,(i_{k-1},i_k)\} : k \in \mathbb{N}, i_0 \in \mathcal{I}, i_k = i \text{ and all } i_j \text{ are distinct}\}$ is the set of directed paths (removing the loops) from the set \mathcal{I} to node i.

Informally, a node *i* belongs to the reachable set of \mathcal{I} if and only if there is a path from \mathcal{I} to *i* in the random graph. By extension, we will call the *reachable set of node i* the reachable set of $\{i\}$. This definition will be used to characterize the asymptotic behavior of the state vector in a contagion process over a graph, as well as connected components of random undirected networks. As we will see in sections 6.4, 6.6 and 6.7, this setting is general enough to include many reference models used in the fields of percolation theory, epidemiology and information cascades.

We now introduce the notion of *influence* of a set \mathcal{I} of nodes, denoted as $\sigma(\mathcal{I})$, as the expected size of the reachable set of \mathcal{I} with respect to the random graph model G(n, A).

Definition 6.5 (Influence). Given a random graph model G(n, A) and a fixed set of influencers $\mathcal{I} \subset [|n|]$, the influence of \mathcal{I} in G(n, A) is defined as the quantity:

$$\sigma(\mathcal{I}) = \mathbb{E}[\operatorname{card}(R(\mathcal{I}, A))], \qquad (6.6)$$

where $R(\mathcal{I}, A)$ is the reachable set of \mathcal{I} in G(n, A).

Remark 6.2. For information cascades and the SIR model, a node is infected at $T = +\infty$ if and only if its it connected to the initially infected nodes by a path of edges that transmitted the contagion. Hence, this transmission graph is a random graph whose influence is the long-term influence of the respective diffusion processes:

$$\sigma(\mathcal{I}) = \lim_{T \to +\infty} \sigma_{\mathcal{I}}(T).$$

Examples of random graphs and their Hazard radiuses

In order to illustrate the previous concepts, we now relate Hazard radiuses to critical thresholds for influence in four particular random graphs. In the first two, we will show that the critical threshold for influence can be restated as $\rho_n = 1$. The other two examples are cases in which the threshold may differ, sometimes significantly, from $\rho_n = 1$.

Example 6.1 (Erdös-Rényi random graphs). For the Erdös-Rényi random graphs $G(n, \frac{c}{n})$ defined in Sec. 5.3.1, percolation theory (Erdös and Rényi, 1960) states that a threshold phenomenon occurs for c = 1. Moreover, using Definition 6.1, $\mathcal{H}_{ij}^n = -\ln(1 - \frac{c}{n})$ and we have:

$$\rho_n = -n\ln\left(1 - \frac{c}{n}\right) \,. \tag{6.7}$$

Hence, for $G(n, \frac{c}{n})$ *, criticality arises when* $\rho_n \rightarrow 1$ *as n tends to infinity.*

Example 6.2 (Poissonian graph processes). We now consider a particular random graph, called the Poissonian graph process or Norros-Reittu model (Norros and Reittu, 2006) and closely related to random graphs of fixed degree distribution known as the configuration model (see Sec. 5.3.3). More specifically, let $w = (w_i)_{i \in [|n|]}$ be a weight vector, and G(n, w) an undirected random graph of n nodes and adjacency matrix A, where, for $i \leq j$, A_{ij} are independent Bernoulli random variables of parameter $P_{ij} = 1 - \exp\left(-\frac{w_i w_j}{\sum_k w_k}\right)$. Note that self-loops are allowed, but they hardly occur when the weight distribution is close to uniform. Such a random graph has a Hazard radius equal to

$$\rho_n = \frac{\sum_i w_i^2}{\sum_k w_k}.$$
(6.8)

Previous results of Bollobás et al. (2007) showed that, for such graphs, a giant component exists if and only if $\sum_i w_i^2 > \sum_i w_i$, which is equivalent to $\rho_n > 1$.

Example 6.3 (Homogeneous percolation on regular grids). Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a regular cubic grid of n nodes in dimension d, and \mathcal{A} its adjacency matrix. The random graph G(n, A) is the result of homogeneous percolation on \mathcal{G} if $\{A_{ij} : \{i, j\} \in \mathcal{E}\}$ are i.i.d. Bernoulli random variables of fixed parameter $p \in [0, 1]$, and $A_{ij} = 0$ otherwise. The Hazard radius of such a network is $\rho_n = -\rho(\mathcal{A}) \ln(1-p) \rightarrow_{n \to +\infty} -2d \ln(1-p)$. For d > 2, there are no known exact formula for percolation thresholds on cubic grids, although experimental approaches by Grassberger (2003) provided high precision numerical estimates. These estimates seem to coincide rather well with $p = 1 - e^{-1/2d}$ (i.e. the value such that $\lim_{n \to +\infty} \rho_n = 1$) for high-dimensional grids (for d = 13, $p^* = 0.040$ compared to $1 - e^{-1/2d} = 0.038$), while being rather different for lower-dimensional grids (for d = 2, $p^* = 0.5$ compared to $1 - e^{-1/2d} = 0.22$).

Example 6.4 (Star-shaped network). For homogeneous percolation on a star-shaped network centered around 1, the exact influence of $\mathcal{I} = \{1\}$ can be derived explicitly and we have: $\sigma(\{1\}) = 1 + p(n-1)$. As, for i < j, the Hazard matrix coefficients are $\mathcal{H}_{ij} = -\ln(1-p)\mathbf{1}\{i=1\}$, the Hazard radius is given by $\rho_n = -\sqrt{n-1}\ln(1-p)$. When $p = c/\sqrt{n}$, the influence is always sublinear in n, and the threshold value is infinite ($c^* = +\infty$). Hence $\lim_{n \to +\infty} \rho_n = c = 1$ does not bring any particular change in the behavior of the influence.

6.2.3 Random graphs with Local Positive Correlation (LPC)

The analysis developed in this chapter concerns a particular class of random graphs that display correlation at a local scale only.

Definition 6.6 (Random graphs with Local Positive Correlation (LPC)). We consider a random graph G(n, A). For any node pair $(i, j) \in [|n|]^2$, we define A_{-ij} to be the subcollection of edge variables $\{A_{kl} : (k,l) \neq (i,j)\}$, and $\mathcal{N}_{ij} = \{(j,i)\} \cup \{(k,l) : k = i \text{ or } l = j\}$ to be the neighborhood that contains the edge (j,i) plus all edges which share with (i,j) either the same head or the same tail. The random graph G(n, A) is a random graph with Local Positive Correlation (LPC) if the two following conditions hold:

- (H1) $\forall i, j, k, l \text{ such that } (k, l) \notin \mathcal{N}_{ij}, A_{ij} \text{ is independent of } A_{kl}.$
- (H2) $\forall i, j$, the mapping $a \mapsto \mathbb{E}[A_{ij}|A_{-ij}=a]$ is non-decreasing w.r.t. the natural partial order on $\{0,1\}^{n^2-1}$ (i.e. $a \leq a'$ if and only if $\forall i \leq n^2 1, a_i \leq a'_i$).

The first assumption (H1) can be interpreted as a property of long range independence between remote edges, while the assumption (H2) properly states the local positive correlation of neighboring edges. When the random variables A_{ij} are interpreted as indicator variables of the occurrence of transmission events from node *i* to node *j* in a diffusion process, then the long range independence assumption implies pairwise independence for transmission events on nonadjacent edges, and the local positive correlation sets positive correlations at the local level for the transmission events on adjacent edges conditionally to the state of all other edges.

Remark 6.3. Since independence implies positive correlation, all random graphs which assume independence of the edge variables also verify LPC. Hence, most standard models of random networks, including Erdös-Rényi (see Sec. 5.3.1) and the very general class of inhomogeneous random graphs (see Sec. 5.3.4), verify LPC.

The following lemma indicates that the notion of random graphs with LPC covers in particular homogeneous and inhomogeneous percolation.

Lemma 6.1 (Undirected random graph). An undirected random graph G(n, A) is a random graph with LPC if and only if the edge variables $\{A_{ij} : i < j\}$ are independent.

Proof. Since independence implies positive correlation, the LPC property is a direct consequence of $\{A_{ij} : i < j\}$ being independent. Now, if an undirected random graph G(n, A) is a random graph with LPC, then, since $A_{ij} = A_{ji}$, (H1) holds for all $(k, l) \notin \mathcal{N}_{ij} \cap \mathcal{N}_{ji} = \{(i, j), (j, i)\}$, and $\{A_{ij} : i < j\}$ are independent.

We will see later that the conditions of LPC are fulfilled for many popular random graph models used in epidemiology (Sec. 6.6) and in information propagation (Sec. 6.7).

6.3 Non-asymptotic upper bounds on the influence

In this section, we provide tight upper bounds on the influence under three scenarios on the set of influencers:

- I) Deterministic (worst-case) scenario: corresponds to a fixed set $\mathcal I$ of influencers,
- II) Random A with parameter $n_0 < n$: the set \mathcal{I} of influencers is random and drawn according to a uniform distribution $\mathcal{U}([|n|], n_0)$ over the subsets of [|n|] of cardinality n_0 ,
- III) Random B with parameter $q \in [0, 1]$: the set \mathcal{I} of influencers is random and drawn with a distribution $\mathcal{D}([|n|], q)$ such that, for all $i \in [|n|]$, the random variables $B_i = \mathbf{1}\{i \in \mathcal{I}\}$ are independent Bernoulli with parameter q.

In both cases (II) and (III), the influencer set is drawn independently of the particular graph sampled under the random graph model G(n, A).

6.3.1 Deterministic (worst-case) scenario

The first result (Theorem 6.1) applies to any fixed set of influencers \mathcal{I} such that card(\mathcal{I}) = n_0 . Intuitively, this result corresponds to a worst-case scenario since the bound does not depend on \mathcal{I} .

Theorem 6.1. Let G(n, A) be a random graph with LPC. For any fixed set of influencers \mathcal{I} such that $card(\mathcal{I}) = n_0 \leq n$, the influence $\sigma(\mathcal{I})$ is upper bounded by:

$$\sigma(\mathcal{I}) \le n_0 + \gamma_1 \left(\rho_n, \frac{n_0}{n - n_0}\right) (n - n_0) , \qquad (6.9)$$

where $\gamma_1(\rho, a) = \gamma_1$ is the smallest solution in [0,1] of the following equation:

$$\gamma_1 - 1 + \exp\left(-\rho\gamma_1 - \frac{\rho a}{\gamma_1}\right) = 0, \qquad (6.10)$$

and ρ_n is the Hazard radius of the random graph G(n, A).

A refined analysis of the parameter γ_1 in the previous theorem leads to the description of the behavior of the influence with respect to three regimes as shown in the following corollary. **Corollary 6.1.** Under the same conditions as in Theorem 6.1, we have:

$$\sigma(\mathcal{I}) \leq \begin{cases} n_0 + \sqrt{\frac{\rho_n}{1 - \rho_n}} \sqrt{n_0(n - n_0)}, & \text{if } \rho_n < 1 - \delta_n \\\\ n_0 + 2^{4/3} n_0^{1/3} (n - n_0)^{2/3}, & \text{if } |\rho_n - 1| \le \delta_n \\\\ n_0 + (n - n_0) \gamma_0(\rho_n) + c_n \sqrt{n_0(n - n_0)}, & \text{if } \rho_n > 1 + \delta_n \end{cases}$$

where
$$\delta_n = \left(\frac{n_0}{4(n-n_0)}\right)^{1/3}$$
, and $c_n = \sqrt{\frac{(1-\gamma_0(\rho_n))\rho_n}{1-(1-\gamma_0(\rho_n))\rho_n}}$.

Remark 6.4. When considering a fixed set of influencers \mathcal{I} , one can remove the ingoing edges of the influencers in order to get slightly improved results. In such a case, the Hazard matrix is replaced by $\mathcal{H}_{ij}(\mathcal{I}) = \mathbf{1}\{j \notin \mathcal{I}\} \cdot \mathcal{H}_{ij}$.

6.3.2 Random influencer set with fixed size

The second result (Theorem 6.2) applies in the case where the set of influencers is drawn from a uniform distribution over the subpartition of sets of n_0 nodes chosen amongst n. This result corresponds to the average-case scenario in a setting where the initial influencer nodes are not known and drawn independently of the random graph.

Theorem 6.2. Let G(n, A) be a random graph with LPC. Assume the set \mathcal{I} of influencers is drawn from an independent and uniform distribution $\mathcal{U}([|n|], n_0)$ over $\mathcal{P}_{n_0}([|n|])$. Then, we have the following result:

$$\mathbb{E}[\sigma(\mathcal{I})] \le n_0 + \gamma\left(\rho_n, \frac{n_0\rho_n}{n-n_0}\right)(n-n_0), \qquad (6.11)$$

where $\gamma(\rho, a)$ is defined as in Definition 6.3.

Corollary 6.2. Under the same conditions as in Theorem 6.2, we have:

$$\mathbb{E}[\sigma(\mathcal{I})] \leq \begin{cases} \frac{n_0}{1-\rho_n}, & \text{if } \rho_n < 1-\delta'_n \\\\ n_0 + \sqrt{8n_0(n-n_0)}, & \text{if } |\rho_n - 1| \le \delta'_n \\\\ (n-n_0)\gamma_0(\rho_n) + \frac{n_0}{1-\rho_n(1-\gamma_0(\rho_n))}, & \text{if } \rho_n > 1+\delta'_n \end{cases}$$

where $\delta'_n = \sqrt{\frac{n_0}{2(n-n_0)}}.$

6.3.3 Random influencer set with random size

The third result (Theorem 6.3) applies in the case where each node of the network is an influencer independently and with a fixed probability $q \in [0, 1]$. This result corresponds to the randomized scenario in which the initial influencer nodes are not known and drawn independently of the random graph.

Theorem 6.3. Let G(n, A) be a random graph with LPC. Assume each node is an influencer with independent probability $q \in [0,1]$ and denote by $\mathcal{I} \sim \mathcal{D}([|n|],q)$ the random set of influencers that is drawn. Then, we have the following result:

$$\mathbb{E}[\sigma(\mathcal{I})] \le \gamma(\rho_n, -\ln(1-q))n, \tag{6.12}$$

where $\gamma(\rho, a)$ is defined as in Definition 6.3.

Corollary 6.3. Under the same conditions as in Theorem 6.3, we have:

$$\mathbb{E}[\sigma(\mathcal{I})] \leq \begin{cases} \frac{-\ln(1-q)n}{1-\rho_n}, & \text{if } \rho_n < 1-d_q \\\\ n\sqrt{-8\ln(1-q)}, & \text{if } |\rho_n-1| \le d_q \\\\ n\gamma_0(\rho_n) + \frac{-\ln(1-q)(1-\gamma_0(\rho_n))n}{1-\rho_n(1-\gamma_0(\rho_n))}, & \text{if } \rho_n > 1+d_q \end{cases}$$

where $d_q = \sqrt{\frac{-\ln(1-q)}{2}}$.

6.3.4 Lower bounds

The following proposition shows that the upper bounds in Corollary 6.1 are tight, in the sense that, for any Hazard radius, there is a random graph on which the influence has the exact same behavior as the upper bounds in the worst case scenario.

Proposition 6.1. For all $\rho > 0$, there exists a constant $C_{\rho} > 0$ and a sequence of LPC random graphs $(G_n)_{n>0}$ with n vertices and Hazard radius ρ_n , and such that $\lim_{n\to+\infty} \rho_n = \rho$. For n sufficiently large, the influence $\sigma_n(\{1\})$ of node 1 in G_n is lower bounded by:

$$\sigma_{n}(\{1\}) \geq \begin{cases} C_{\rho}\sqrt{n} & if \rho < 1\\ C_{\rho}n^{2/3} & if \rho = 1\\ \gamma_{0}(\rho)n - o(n) & if \rho > 1 \end{cases}$$
(6.13)

This proposition relies on "*random star-networks*", i.e. undirected random graphs such that the $\{A_{ij} : i < j\}$ are independent Bernoulli random variables of parameter $a \in [0,1]$ if i = 1 and b < a otherwise. Intuitively, such a network is the addition of a star network and an Erdös-Rényi random graph. Our theoretical bounds on the influence are particularly tight on this class of random graphs.

6.4 Application to bond percolation

Bounding the influence of reachable sets in random graphs allows to derive non-asymptotic bounds on a celebrated quantity in bond percolation which is the size of the giant component, as well as the distribution of the size of connected components of undirected random graphs. We first recall the inhomogeneous bond percolation model. We recall that, according to Lemma 6.1, the bond percolation graph G = G(n, A) of Model 5.6 is a random graph with LPC. For $k \ge 1$, we denote by $\mathcal{V}_k \subset [|n|]$ the k^{th} -largest connected component of G. We also introduce $C_k(G) = \operatorname{card}(\mathcal{V}_k)$, the size of the connected component with k^{th} greatest cardinality and N(m) the number of connected components of G of cardinality greater than or equal to m.

6.4.1 Size and existence of the giant component

Let a > 0. The key observation is that if each node of *G* is an influencer with independent probability $1 - \exp(-a)$, we can relate the total size of infection and the size of the connected components of *G* in the following way:

$$\mathbb{E}[\operatorname{card}(R(\mathcal{I},A)) \mid A] = \sum_{k} C_{k}(G) \mathbb{P}(\mathcal{V}_{k} \cap \mathcal{I} \neq \emptyset \mid A) = \sum_{k} C_{k}(G) \left(1 - e^{-aC_{k}(G)}\right)$$

Therefore, we have

$$\mathbb{E}\left[\sum_{k} C_{k}(G) \left(1 - e^{-aC_{k}(G)}\right)\right] = \mathbb{E}[\sigma(\mathcal{I})] \leq \gamma(\rho_{n}, a)n$$
(6.14)

The next argument leads to non-asymptotic bounds on the size of the giant component for the inhomogeneous bond percolation model:

Theorem 6.4 (Size of the giant component). Let G(n, A) be an undirected random graph of size n with independent edge variables $\{A_{ij} : i < j\}$. Let a > 0, and ρ_n the Hazard radius of G. The probability distribution of the size of its largest connected component $C_1(G)$ verifies:

$$\mathbb{E}[C_1(G)(1 - e^{-a(C_1(G) - 1)})] \le n\left(1 - e^{-\rho_n\gamma(\rho_{n,a})}\right)$$
(6.15)

where $\gamma(\rho, a)$ is defined as in Definition 6.3.

Corollary 6.4. Under the same conditions as in Theorem 6.4, we have:

$$\mathbb{E}[C_{1}(G)] \leq \begin{cases} \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{n\rho_{n}}{1 - \rho_{n}}}, & \text{if } \rho_{n} < 1 - \kappa n^{-1/3} \\\\ \gamma_{0}(\rho_{n})n + \frac{n^{2/3}}{\sqrt{\kappa}}, & \text{if } |\rho_{n} - 1| \le \kappa n^{-1/3} \\\\ \gamma_{0}(\rho_{n})n + c_{n}\sqrt{n} + 2, & \text{if } \rho_{n} > 1 + \kappa n^{-1/3} \end{cases}$$

where $c_{n} = \frac{2}{\sqrt{e}}\sqrt{\frac{(1 - \gamma_{0}(\rho_{n}))^{2}\rho_{n}}{1 - \rho_{n} + \gamma_{0}(\rho_{n})\rho_{n}}} \text{ and } \kappa = \left(\frac{2e}{27}\right)^{2/3}.$

Remark 6.5 (Homogeneous percolation). Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph and $p \in [0,1]$. If $\mathbb{E}[A_{ij}] = p\mathbf{1}\{(i,j) \in \mathcal{E}\}$, then $\rho_n = -\ln(1-p)\rho(\mathcal{A})$ where \mathcal{A} is the adjacency matrix of the underlying graph \mathcal{G} .

Whereas the latter results hold for any $n \in \mathbb{N}$, classical results in percolation theory study the asymptotic behavior of sequences of graphs when $n \to \infty$. Up to our knowledge, the best result in the inhomogeneous bond percolation model (Bollobás et al., 2007, Corollary 3.2 of section 5) states that: under a given subcriticality condition, $C_1(G_n) = o(n)$ *asymptotically almost surely* (a.a.s.). Combining our previous theorem and Markov's inequality, we are in position of obtaining a significant improvement on the previous result.

Corollary 6.5. Denote by $G_n = G(n, A_n)$ a sequence of undirected random graphs and ρ_n the sequence of spectral radiuses of the corresponding Hazard matrices. Let ω_n be a sequence such that $\lim_{n\to\infty} \omega_n = +\infty$. We have:

$$C_{1}(G_{n}) = \begin{cases} o(n^{1/2}\omega_{n}) \text{ a.a.s.}, & \text{if } \limsup_{n \to \infty} \rho_{n} < 1 \\ \\ o(n^{\max(1-\beta,2/3)}\omega_{n}) \text{ a.a.s.}, & \text{if } \rho_{n} - 1 = O(n^{-\beta}) \end{cases}$$

Moreover, Corollary 6.4 also implies a result of non-existence of the giant component, in the sense that no component has a size proportional to the size of the network.

Corollary 6.6 (Existence of a giant component). Denote by $G_n = G(n, A_n)$ a sequence of undirected random graphs and ρ_n the sequence of spectral radiuses of the corresponding Hazard matrices. If $\limsup_{n\to\infty} \rho_n \leq 1$, then

$$\mathbb{E}[C_1(G_n)] = o(n), \tag{6.16}$$

and there is no giant component in G_n a.a.s..

Proof. Combining the second equation of Corollary 6.4 (valid for all $\rho_n \ge 0$) and Lemma 6.4, we obtain $\mathbb{E}[C_1(G_n)] = o(n)$. Markov's inequality implies the result in probability. \Box

6.4.2 Number of components of cardinality larger than m

The following result focuses on discovering the expectation of N(m), the number of connected components of *G* having cardinality greater than or equal to *m*. A straightforward observation yields $\mathbb{E}[N(m)] \leq \frac{n}{m}$. For critical and subcritical random graphs, we are able to show that the expectation of N(m) is in fact decreasing much faster with respect to *m*.

Theorem 6.5. Let G(n, A) be an undirected random graph of size n with independent edge variables $\{A_{ij} : i < j\}$. Let m > 0. The expected number of connected components N(m) of cardinality greater than or equal to m is upper bounded by:

$$\mathbb{E}[N(m)] \le \frac{n}{m} \min_{a>0} \left\{ \frac{1 - e^{-\rho_n \gamma(\rho_n, a)}}{1 - e^{a(m-1)}} \right\}$$
(6.17)

where $\gamma(\rho, a)$ is defined as in Definition 6.3.

Corollary 6.7. Under the same conditions as in Theorem 6.5, we have:

$$\mathbb{E}[N(m)] \leq \begin{cases} \frac{n}{m(m-1)} \frac{\rho_n}{1-\rho_n}, & \text{if } \rho_n < 1-\kappa_1 m^{-1/2} \\\\ \frac{n}{m^{3/2}} \frac{1}{\kappa_1}, & \text{if } |\rho_n - 1| \le \kappa_1 m^{-1/2} \\\\ \frac{n}{m} \left(\gamma_0(\rho_n) + \frac{c'_n}{\sqrt{m-1}} + \frac{c_n}{m-1}\right), & \text{if } \rho_n > 1 + \kappa_1 m^{-1/2} \end{cases}$$

where $c_n = \frac{(1-\gamma_0(\rho_n))^2 \rho_n}{1-\rho_n+\gamma_0(\rho_n)\rho_n}$, $c'_n = \sqrt{\gamma_0(\rho_n)c_n}$ and $\kappa_1 = \sqrt{\frac{\eta}{8}} \left(\sqrt{1+\frac{8}{2\eta-1}}-1\right)$ where η is the strictly positive solution of $e^{\eta} = 2\eta + 1$ ($\kappa_1 \approx 0.32$).

6.5 Application to site percolation

In this section, we show that the results of the previous section can further be applied to site percolation as defined in Model 5.7. Although the resulting undirected random graph G_{SP} is not LPC, the size of connected components in G_{SP} can be bounded by the size of reachability sets of a random graph with LPC. Let $A_{ij} = X_j \mathbf{1}\{(i,j) \in \mathcal{E}\}$, where X_i are the Bernoulli random variables indicating the presence or absence of a node i in G_{SP} , and $G'_{SP} = G(n, A)$.

Proposition 6.2. G'_{SP} is a random graph with LPC. Furthermore, if $R(\mathcal{I}, A)$ is the reachable set of $\mathcal{I} \subset [|n|]$ in G'_{SP} , then

$$\operatorname{card}(R(\mathcal{I},A)) \ge \sum_{k} C_{k}(G_{SP}) \mathbb{1}\{\mathcal{V}_{k} \cap \mathcal{I} \neq \emptyset\} \text{ a.s.,}$$
(6.18)

where \mathcal{V}_k is the k^{th} -largest connected component of G_{SP} and $C_k(G_{SP}) = \operatorname{card}(\mathcal{V}_k)$.

Proof. Since the X_i are independent, for all i, j, i', j', A_{ij} and $A_{i'j'}$ are positively correlated if j = j' and independent otherwise, which proves the LPC assumption. In order to prove the inequality, it suffices to see that, if there is an influencer in a connected component \mathcal{V}_k of G_{SP} (i.e. $\mathcal{V}_k \cap \mathcal{I} \neq \emptyset$), then \mathcal{V}_k is in reachable set $R(\mathcal{I}, A)$.

Since the inequality in Proposition 6.2 is the same starting point as the results derived for bond percolation, all the results of Sec. 6.4 also apply to site percolation with the following Hazard radius:

$$\rho_n = \rho\left(-\frac{\ln(1-p_i) + \ln(1-p_j)}{2}\mathbb{1}\{(i,j) \in \mathcal{E}\}\right).$$
(6.19)

6.6 Application to epidemiology

We here focus on the Susceptible-Infected-Removed (SIR) model (see Model 5.9), and show that its long-term behavior is a particular case of LPC random graphs.

6.6.1 Subcritical behavior in the standard SIR model

We show here that Theorem 6.1 (through Corollary 6.1) further improves results on the SIR model in epidemiology. In order to determine the long-term behavior of the epidemic, the following theorem shows that the set R_{∞} of recovered nodes at the end of the epidemic is the reachable set of a random graph with LPC.

Proposition 6.3. Let τ_i^R be the removal times of an SIR epidemic with transmission times T_{ij} and recovery times D_i (see Proposition 5.1). Then, the random graph $G_{SIR} = G(n, A)$ with adjacency matrix $A_{ij} = \mathbf{1}\{\{i, j\} \in \mathcal{E} \text{ and } T_{ij} < D_i\}$ is a random graph with LPC and, if $R(\mathcal{I}, A)$ is the reachable set of \mathcal{I} in G_{SIR} , then the set of recovered nodes at the end of the epidemic $R_{\infty} = \{i \in [|n|] : \tau_i^R < +\infty\}$ is equal to $R(\mathcal{I}, A)$.

Proof. G_{SIR} is a random graph with LPC since only outgoing edges of a node are correlated together, and this correlation is positive due to the fact that $\mathbb{E}[A_{ij}|A_{-ij} = a] = \mathbb{P}(T_{ij} < D_i \mid \max_{k \in \mathcal{N}_{ij}^1(a)} T_{ik} < D_i \leq \min_{k \in \mathcal{N}_{ij}^0(a)} T_{ik})$, where $\mathcal{N}_{ij}^b(a) = \{k \neq j \text{ such that } (i,k) \in \mathcal{E} \text{ and } a_{ik} = b\}$ for $b \in \{0,1\}$, which is non-decreasing w.r.t. a. Finally, $R_{\infty} = \{i \in [|n|] : \tau_i^I < +\infty\}$, and a node i is in R_{∞} if and only if $i \in \mathcal{I}$ or i has an infected neighbor j that transmitted the disease, i.e. such that $j \in R(\mathcal{I}, A), \{j, i\} \in \mathcal{E}$ and $T_{ji} < D_j$, or equivalently $A_{ji} = 1$.

The direct consequence of this proposition is that the number of infected (and then removed) nodes through the epidemic $\lim_{T\to+\infty} \sigma_{\mathcal{I},R}(T)$ is the influence $\sigma(\mathcal{I})$ of \mathcal{I} in the random graph G_{SIR} . The Hazard matrix \mathcal{H} of G_{SIR} is given by $\ln(1 + \frac{\beta}{\delta}) \cdot \mathcal{A}$ and hence $\rho_n = \ln(1 + \frac{\beta}{\delta}) \cdot \rho(\mathcal{A})$. A direct application of Corollary 6.1 leads to the following result.

Corollary 6.8. We consider an SIR epidemic with $\delta > 0$. We denote by \mathcal{A} the symmetric adjacency matrix of \mathcal{G} , by $\rho_n = \ln(1 + \frac{\beta}{\delta}) \cdot \rho(\mathcal{A})$ its Hazard radius, and by \mathcal{I} the initial set of influencers of size n_0 . If $\frac{\beta}{\delta} < \exp(\frac{1}{\rho(\mathcal{A})}) - 1$, then we have

$$\sigma(\mathcal{I}) \le n_0 + \sqrt{\frac{\rho_n}{1 - \rho_n}} \sqrt{n_0(n - n_0)} .$$
(6.20)

It was recently shown by Draief et al. (2008) that, in the case of undirected networks, and if $\beta \rho(A) < \delta$, we have the following bound on the influence for a fixed set of influencer nodes:

$$\sigma(\mathcal{I}) \le \frac{\sqrt{nn_0}}{1 - \frac{\beta}{\delta}\rho(\mathcal{A})} \,. \tag{6.21}$$

As we will show now, Corollary 6.8 improves the result of Draief et al. (2008) and Prakash et al. (2012) in two directions: weaker condition and tighter constants in the upper bound. Indeed, when $\rho(\mathcal{A}) \gg 1$, $\frac{1}{\rho(\mathcal{A})}$ is a good approximation of $\exp(\frac{1}{\rho(\mathcal{A})}) - 1$. However, the two quantities may differ substantially on very sparse networks, for which $\rho(\mathcal{A})$ is close to 1. For example, for an n-cycle graph, we have $\mathcal{A}_{ij} = \mathbf{1}\{j = i \pm 1 \mod n\}$ where mod is the modulo operator, which leads to $\rho(\mathcal{A}) = 2$ and $\exp(\frac{1}{\rho(\mathcal{A})}) - 1 \approx 0.65 > 0.5$. Now, as far as the comparison between the two rates is concerned, we offer the following lemma which assesses the tightness of Corollary 6.8 with respect to the upper bound in Eq. 6.21.

Lemma 6.2. We use the same notations as in Corollary 6.8. If $\beta \rho(\mathcal{A}) < \delta$, then we have:

$$n_0 + \sqrt{\frac{\rho_n}{1 - \rho_n}} \sqrt{n_0(n - n_0)} \le \frac{\sqrt{nn_0}}{1 - \frac{\beta}{\delta}\rho(\mathcal{A})} .$$
(6.22)

Proof. First, $\rho_n = \ln(1 + \frac{\beta}{\delta})\rho(\mathcal{A}) \leq \frac{\beta}{\delta}\rho(\mathcal{A})$. Then, we introduce the function

$$f: r \to n_0 + \sqrt{\frac{r}{1-r}} \sqrt{n_0(n-n_0)} - \frac{\sqrt{nn_0}}{1-r}$$

A simple analysis shows that $\max_{r \in [0,1]} f(r) = n_0 \left(1 - \frac{3}{4}\sqrt{\frac{n}{n_0}} - \frac{1}{4}\sqrt{\frac{n_0}{n}}\right) \le 0$ when $n_0 \le n$, which proves the lemma.

Moreover, these new bounds capture with increased accuracy the behavior of the influence in extreme cases. In the limit $\beta \to 0$, the difference between the two bounds is significant, because Theorem 6.1 yields $\sigma(\mathcal{I}) \to n_0$ whereas Eq. 6.21 only ensures $\sigma(\mathcal{I}) \leq \sqrt{nn_0}$. When $n = n_0$, Theorem 6.1 also ensures that $\sigma(\mathcal{I}) = n_0$ whereas Eq. 6.21 yields $\sigma(\mathcal{I}) \leq \frac{n_0}{1 - \frac{\beta}{\delta}\rho(\mathcal{A})}$. Secondly, Theorem 6.1 also describes the explosive behavior in the SIR model and leads to bounds in the case where $\beta\rho(\mathcal{A}) \geq \delta$, as we will see below.

6.6.2 Behavior near the epidemic threshold

The regime around 1 of $\ln(1 + \frac{\beta}{\delta}) \cdot \rho(\mathcal{A})$ can also be derived from the generic results of Sec. 6.3.

Corollary 6.9 (Critical behavior of SIR). We use the same notations as in Corollary 6.8. If $|\rho_n - 1| \le \left(\frac{n_0}{4(n-n_0)}\right)^{1/3}$, then we have

$$\tau(\mathcal{I}) \le n_0 + 2^{4/3} n_0^{1/3} (n - n_0)^{2/3}.$$
(6.23)

Proof. This is also a direct application of Corollary 6.1 to G_{SIR} .

More specifically, the behavior when $\beta \rho(A) = \delta$ depends on the rate at which the spectral radius of the adjacency matrix diverges w.r.t. *n*.

Corollary 6.10. We use the same notations as in Corollary 6.8. Assume $n_0 = O(1)$ and $\rho(\mathcal{A}) = O(n^{\alpha})$ for $\alpha \ge 0$. If $\beta \rho(\mathcal{A}) = \delta$, then we have

$$\sigma(\mathcal{I}) = O\left(n^{\min\{\frac{1+\alpha}{2},\frac{2}{3}\}}\right).$$
(6.24)

Proof. If the graph is empty, then $\sigma(\mathcal{I}) = 0$. Otherwise, $\rho(\mathcal{A}) \ge 1$ and $\rho_n = \ln(1 + \frac{1}{\rho(\mathcal{A})})\rho(\mathcal{A}) \le 1 - \frac{1 - \ln 2}{\rho(\mathcal{A})}$, the critical bound of Corollary 6.1 implies that $\sigma(\mathcal{I}) = O(n^{2/3})$, while the subcritical bound implies that $\sigma(\mathcal{I}) = O(n^{\frac{1+\alpha}{2}})$.

The behavior in $O(n^{2/3})$ of the size of the epidemic in the critical regime was already shown for the more simple N-intertwined SIR model by Ben-Naim and Krapivsky (2004) (in which the three populations are assumed to be mixed uniformly). However, this result is, up to our knowledge, the first to prove such a behavior in the more general case of epidemics on networks. Finally, note that Proposition 6.1 implies that the behavior in $O(n^{2/3})$ is tight, in the sense that some networks do behave accordingly in the critical regime.

6.7. APPLICATION TO INFORMATION CASCADES

6.6.3 Generic incubation period

Theorem 6.1 applies to more general cases than the classical homogeneous SIR model, and allows infection and recovery rates to vary across individuals. Also, our model allows for incubation times which display a non-exponential behavior, and thus is more adapted to realistic scenarios. Indeed, incubation periods for different individuals generally follow a log-normal distribution (Nelson, 2007), which indicates that SIR with a log-normal recovery rate of removal might be well-suited to model real-world infections.

For each node i, let the incubation time D_i (i.e. the time for an infected node to recover) be a random variable drawn according to a certain probability distribution P_D . In such a case, the Hazard radius is

$$\rho_n = -\rho(\mathcal{A}) \ln\left(\mathbb{E}[e^{-\beta D}]\right),\tag{6.25}$$

and a sufficient condition for subcriticality is $\beta \rho(\mathcal{A}) \mathbb{E}[D] < 1$, where $\mathbb{E}[D] = \int x P_D(x) dx$.

Corollary 6.11 (Generic incubation period). We consider a graph of contaminated nodes obtained after the realization of an SIR contagion process with incubation times drawn according to the probability distribution P_D . We denote by \mathcal{A} its symmetric adjacency matrix, by \mathcal{I} its initial set of influencers of size $n_0 = O(1)$, and $\mathbb{E}[D] = \int x P_D(x) dx$.

If $\beta \rho(\mathcal{A}) \mathbb{E}[D] < 1$, then we have

$$\sigma(\mathcal{I}) = O(\sqrt{n}) . \tag{6.26}$$

Proof. Since, for $G_{SIR} = G(n, A)$, $\mathbb{E}[A_{ij}] = \mathbb{P}(T_{ij} < D_i) = 1 - \mathbb{E}[e^{-\beta D}]$, a direct application of Corollary 6.1 to G_{SIR} returns that the epidemic is subcritical if $\rho_n < 1$, i.e. $-\rho(\mathcal{A})\ln(\mathbb{E}[e^{-\beta D}]) < 1$. Jensen's inequality on $\mathbb{E}[e^{-\beta D}]$ leads to the desired result. \Box

In the log-normal case, Corollary 6.11 gives the following bound on the epidemic threshold:

Corollary 6.12 (Log-normal incubation period). We consider a graph of contaminated nodes obtained after the realization of an SIR contagion process with incubation times drawn according to a log-normal distribution of parameters μ_D and σ_D . We denote by \mathcal{A} its symmetric adjacency matrix, and by \mathcal{I} its initial set of influencers of size $n_0 = O(1)$.

If
$$\mu_D + \frac{\sigma_D^2}{2} < -\ln(\beta\rho(\mathcal{A}))$$
, then we have
 $\sigma(\mathcal{I}) = O(\sqrt{n})$. (6.27)

6.7 Application to Information Cascades

Predicting, and then maximizing, the influence of information cascades (see Model 5.11 and Model 5.12) are among the main goals in information propagation theory, and the subject of many scientific works. In this section, we show that our generic results also apply to these diffusion processes, and allow to derive upper bounds for the long-term influence of a set of influencers. Let $I_t \subset [|n|]$ be the set of infected nodes at time t. In $DTIC(\mathcal{P})$ and $CTIC(\mathcal{F},\infty)$, I_t is non-decreasing w.r.t. t and reaches a limit set $I_{\infty} = \lim_{t \to +\infty} I_t$. Due to the independence of transmission events along the edges of the graph, I_{∞} is the reachable set of \mathcal{I} in a random graph $G_{IC} = G(n, A)$ with independent edge variables A_{ij} . Hence G_{IC} is a random graph with LPC and the results of Sec. 6.3 are applicable.

Proposition 6.4. Let \mathcal{I} be a set of influencers, $\mathcal{P} = (p_{ij})_{ij} \in [0,1]^{n \times n}$ a matrix of transmission probabilities and $\mathcal{F} = (f_{ij})_{ij}$ a matrix of non-negative integrable functions. Then, under $DTIC(\mathcal{P})$ and $CTIC(\mathcal{F},\infty)$, the set of infected nodes at the end of the diffusion process is the reachable set $R(\mathcal{I}, A)$ of \mathcal{I} in a random graph with LPC, and Theorems 6.1, 6.2 and 6.3 are applicable with the following Hazard matrix:

$$\mathcal{H}_{ij} = \begin{cases} -\ln(1-p_{ij}) & \text{for } DTIC(\mathcal{P}) \\ \int_0^\infty f_{ij}(t)dt & \text{for } CTIC(\mathcal{F},\infty) \end{cases}$$
(6.28)

Proof. Since transmission events are independent, we can, prior to the epidemic, draw, respectively, the transmission $X_{ij} \sim \mathcal{B}(p_{ij})$ along each edge $(i,j) \in [|n|]^2$ for $DTIC(\mathcal{P})$, and time to transmit along each edge T_{ij} for $CTIC(\mathcal{F},T)$ (see Proposition 5.2). Then, a node *i* belongs to I_{∞} if and only if there is a path between \mathcal{I} and *i* such that each of its edges transmitted the information. Hence, I_{∞} is the reachable set of \mathcal{I} in the random graph G(n, A) s.t. $A_{ij} = X_{ij}$ for $DTIC(\mathcal{P})$, and $A_{ij} = \mathbf{1}\{T_{ij} < +\infty\}$ for $CTIC(\mathcal{F},T)$. These are independent Bernoulli random variables of parameter p_{ij} for $DTIC(\mathcal{P})$, and $1 - \exp(-\int_0^{\infty} f_{ij}(t)dt)$ for $CTIC(\mathcal{F},T)$, which implies that *G* is a random graph with LPC and the above mentioned Hazard matrices.

6.8 Discussion

In this chapter, we established new bounds on the influence in random graphs, and applied our results to three quantities of major importance in their respective fields: the size of the giant component in percolation, the number of infected nodes in epidemiology and the influence of information cascades. These bounds are a strong indication that the *Hazard radius* plays an important role in the dynamics of diffusion processes in random graphs, and lead to several open questions. First, one may wonder if the LPC property is a necessary condition for the bounds to hold. For example, relaxing the local correlation and allowing positive correlation on larger neighborhoods may still provide random graphs in which criticality is controlled by the Hazard radius. Second, an important class of diffusion models, based on randomized versions of the *Linear Threshold* model, is so far absent of this analysis, and being able to describe such models by a well chosen LPC random graph may lead to new and valuable results. Finally, the Hazard radius may drive the behavior of other diffusion-related quantities in random graphs, such as the volume of neighborhoods of fixed size. Such results would prove critical for understanding the temporal dynamics of diffusion processes in networks.

6.9 **Proofs**

6.9.1 Behavior of the Hazard function

When $\rho \ge 0$ and a > 0, $\gamma - 1 + \exp(-\rho\gamma - a) = 0$ always has a solution in [0,1] and $\gamma(\rho, a)$ is well defined. γ and γ_0 are non-decreasing w.r.t. ρ , $\lim_{\rho \to +\infty} \gamma(\rho, a) = \lim_{\rho \to +\infty} \gamma_0(\rho) = 1$ and, for $\rho \le 1$, $\gamma_0(\rho) = 0$.

Moreover, we have the following upper bounds for $\gamma(\rho, a)$, that we will use to determine the subcritical, critical and supercritical behavior of the influence.

6.9. PROOFS

Lemma 6.3. $\forall \rho \neq 1 \text{ and } a > 0$,

$$\gamma(\rho, a) \le \gamma_0(\rho) + \frac{a(1 - \gamma_0(\rho))}{1 - \rho(1 - \gamma_0(\rho))},$$
(6.29)

and $\forall \rho > 0$ and a > 0,

$$\gamma(\rho, a) \le \gamma_0(\rho) + \sqrt{2a} \min\left\{1, \sqrt{\frac{1}{\rho}}\right\}.$$
(6.30)

Eq. 6.29 is particularly tight, except when $\rho \approx 1$ (i.e. the critical case). In order to derive upper bounds in the critical case, we will thus use the second upper bound.

Proof. By definition of $\gamma(\rho, a)$,

$$\begin{aligned} \gamma(\rho, a) &= 1 - \exp\left(-\rho\gamma_{0}(\rho) - \rho(\gamma(\rho, a) - \gamma_{0}(\rho)) - a\right) \\ &\leq 1 - (1 - \gamma_{0}(\rho))\left(1 - \rho(\gamma(\rho, a) - \gamma_{0}(\rho)) - a\right) \\ &= \gamma_{0}(\rho) + (1 - \gamma_{0}(\rho))\left(\rho(\gamma(\rho, a) - \gamma_{0}(\rho)) + a\right), \end{aligned}$$
(6.31)

hence

$$\gamma(\rho, a) \le \gamma_0(\rho) + \frac{a(1 - \gamma_0(\rho))}{1 - \rho(1 - \gamma_0(\rho))}.$$
(6.32)

For the second inequality, first observe that $\gamma(\rho, a) \geq 1 - \frac{1}{\rho}$ since $\gamma(\rho, a) = 1 - \exp(-\rho\gamma(\rho, a) - a) \geq 1 - \frac{1}{1+\rho\gamma(\rho, a)} = \frac{\rho\gamma(\rho, a)}{1+\rho\gamma(\rho, a)}$ which leads to $\rho\gamma(\rho, a) \geq \rho - 1$. The second inequality follows from an approximation of the derivative of $\gamma(\rho, a)$ w.r.t. a:

$$\frac{\partial \gamma(\rho, a)}{\partial a} = \frac{1 - \gamma(\rho, a)}{1 - \rho(1 - \gamma(\rho, a))} \le \frac{1}{1 - \rho(1 - \gamma(\rho, a))}.$$
(6.33)

Multiplying the two terms by $1-\rho(1-\gamma(\rho,a))>0$ and integrating between 0 and a, we get

$$(1-\rho)(\gamma(\rho,a) - \gamma_0(\rho)) + \frac{\rho}{2}(\gamma(\rho,a)^2 - \gamma_0(\rho)^2) \le a$$
(6.34)

which leads to

$$\gamma(\rho, a) \le 1 - \frac{1}{\rho} + \sqrt{(\gamma_0(\rho) - 1 + \frac{1}{\rho})^2 + \frac{2a}{\rho}} \le \gamma_0(\rho) + \sqrt{\frac{2a}{\rho}}.$$
 (6.35)

using that, $\forall a, b \ge 0$, $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$. Finally, noting that $\gamma(\rho, a)$ is non-decreasing, we get that $\forall \rho \le 1, \gamma(\rho, a) \le \gamma(1, a) \le \sqrt{2a}$, and $\forall \rho \ge 1, \sqrt{\frac{2a}{\rho}} \le \sqrt{2a}$.

We will also use the following bound on $\gamma_0(\rho)$:

Lemma 6.4. $\forall \rho \geq 1, \gamma_0(\rho) \leq 2(\rho - 1).$
Proof. A simple calculation holds that γ_0 is concave on $(1, +\infty)$. Thus, it implies that, $\forall \rho > 1$,

$$\gamma_0(\rho) \le \gamma_0'(1^+)(\rho - 1). \tag{6.36}$$

Finally, $\forall \epsilon > 0$,

$$\gamma_0'(1+\epsilon) = \frac{\gamma_0(1+\epsilon)(1-\gamma_0(1+\epsilon))}{1-(1+\epsilon)(1-\gamma_0(1+\epsilon))} = \frac{\gamma_0'(1^+)}{\gamma_0'(1^+)-1} + o(\epsilon), \tag{6.37}$$

which leads to $\gamma'_0(1^+) \in \{0,2\}$, and $\gamma'_0(1^+) = 0$ is impossible since γ_0 is concave on $(1, +\infty)$ and $\gamma_0(\rho) > 0$ for all $\rho > 1$. Hence, $\gamma'_0(1^+) = 2$ and $\gamma_0(\rho) \le 2(\rho - 1)$ for $\rho \ge 1$.

6.9.2 Proofs of the upper bounds on influence

In this section, we consider G(n, A) a random graph with LPC and $R(\mathcal{I}, A)$ the reachable set of a set of influencers \mathcal{I} in G. We will also define $X_i = \mathbb{1}\{i \in R(\mathcal{I}, A)\}$ the indicators of the reachable set. First, note that all the bounds provided in Sec. 6.3 are infinite when there exists an edge (i, j) such that $\mathbb{E}[A_{ij}] = 1$ (since, in such a case, $\rho_n = +\infty$). Hence, we will assume that $\forall (i, j) \in [|n|]^2$, $\mathbb{E}[A_{ij}] < 1$. In the following paragraphs, we will prove our results for random graphs having a *strictly positive* measure, i.e. such that every graph of *n* nodes has a non-zero probability. When this assumption is not satisfied, the next two lemmas show that we can still derive the desired results by considering a sequence of such graphs converging to *G*.

Definition 6.7 (Perturbed random graph). Let G(n, A) be a random graph and $\varepsilon > 0$. The ε -perturbed version of G, $G^{\varepsilon} = (n, A^{\varepsilon})$, is a random graph such that $A_{ij}^{\varepsilon} = A_{ij}(1 - X_{ij}) + Y_{ij}X_{ij}$ where X_{ij} and Y_{ij} are, respectively, i.i.d. Bernoulli random variables with parameter ε and 1/2, and independent of G.

These *noisy* versions of G have a strictly positive measure, while still verifying the LPC property.

Lemma 6.5. Let $\varepsilon > 0$, G(n, A) a random graph and $G^{\varepsilon} = (n, A^{\varepsilon})$ its ε -perturbed version. Then G^{ε} has a strictly positive measure and, if G is a random graph with LPC, then so does G^{ε} .

Proof. Let X and Y be the random matrices of Definition 6.7. $\forall a \in \{0,1\}^{n^2}$, $\mathbb{P}(A^{\varepsilon} = a) \ge \mathbb{P}(X = \mathbf{1}, Y = a) = (\frac{\varepsilon}{2})^{n^2} > 0$, where **1** is a vector of size *n* filled with ones. Also, since X_{ij} and Y_{ij} are independent, then (H1) and (H2) of the LPC property are still verified for G^{ε} (see Definition 6.6).

Furthermore, the influence and Hazard radius are continuous w.r.t. ε , and thus our results on strictly positive measures can be generalized to any random graph with LPC.

Lemma 6.6. Let \mathcal{I} be a set of influencers, $\varepsilon > 0$, G(n, A) a random graph and $G^{\varepsilon} = G(n, A^{\varepsilon})$ its ε -perturbed version. Let also $\sigma(\mathcal{I})$ be the influence of \mathcal{I} in G, $\sigma^{\varepsilon}(\mathcal{I})$ the influence of \mathcal{I} in G^{ε} , ρ_n the Hazard radius of G and ρ_n^{ε} the Hazard radius of G^{ε} . Then the following results hold:

$$\lim_{\varepsilon \to 0} \sigma^{\varepsilon}(\mathcal{I}) = \sigma(\mathcal{I}), \tag{6.38}$$

6.9. PROOFS

and

$$\lim_{\varepsilon \to 0} \rho_n^{\varepsilon} = \rho_n. \tag{6.39}$$

Proof. Let X and Y be the random matrices of Definition 6.7. $\forall a \in \{0,1\}^{n^2}$, $\mathbb{P}(A^{\varepsilon} = a) = \mathbb{P}(X = \mathbf{0}, A = a) + \mathbb{P}(X \neq \mathbf{0}, A^{\varepsilon} = a)$, where **0** is the vector of size *n* filled with zeros. Hence,

$$|\mathbb{P}(A^{\varepsilon} = a) - \mathbb{P}(A = a)| = |\mathbb{P}(X \neq \mathbf{0}, A^{\varepsilon} = a) - \mathbb{P}(X \neq \mathbf{0})|$$

$$\leq 2\mathbb{P}(X \neq \mathbf{0})$$

$$\leq 2(1 - (1 - \varepsilon)^{n^2})$$

$$\rightarrow_{\varepsilon \to 0} 0.$$
(6.40)

Since $\{0,1\}^{n^2}$ is finite, A^{ε} converges to A in law, and for any function $f: \{0,1\}^{n^2} \to \mathbb{R}$ we have:

$$\lim_{\varepsilon \to 0} \mathbb{E}[f(A^{\varepsilon})] = \mathbb{E}[f(A)].$$
(6.41)

Selecting $f(A) = \operatorname{card}(R(\mathcal{I}, A)) = \operatorname{card}(\mathcal{I}) + \sum_{i \notin \mathcal{I}} \left(1 - \prod_{q \in \mathcal{Q}_{\mathcal{I},i}} (1 - \prod_{(j,l) \in q} A_{jl})\right)$ (see Definition 6.4) implies that $\lim_{\epsilon \to 0} \sigma^{\epsilon}(\mathcal{I}) = \sigma(\mathcal{I})$. The second result comes from the continuity of the spectral radius and that

$$\mathcal{H}_{ij}^{\varepsilon} = -\ln\left(1 - (1 - \varepsilon)\mathbb{E}[A_{ij}] - \frac{\varepsilon}{2}\right) \to_{\varepsilon \to 0} \mathcal{H}_{ij}.$$
(6.42)

Proofs of Theorem 6.1 and Corollary 6.1

We develop here the full proofs for Theorem 6.1 and Corollary 6.1 that apply to any set of influencers. Due to Lemma 6.5 and Lemma 6.6, without loss of generality, we will restrict ourselves to random graphs *G* that have a strictly positive measure. We will first need to prove two useful results: Lemma 6.7, that proves for $j \in [|n|]$ a positive correlation between the events 'node *i* is not reachable from \mathcal{I} through node *j*' and Lemma 6.9, that bounds the probability that a given node is reachable from \mathcal{I} .

Lemma 6.7. $\forall i \notin \mathcal{I}, \{1 - X_i A_{ii}\}_{i \in [|n|]}$ are positively correlated.

Proof. We will make use of a generalization of the FKG inequality due to Holley (1974) (see also otto Georgii et al. (1999) for a more recent presentation of the inequality), that only requires the positive correlation of the edge presence variables A_{ij} (hypothesis (H2) of the LPC property, see Definition 6.6):

Lemma 6.8 (FKG inequality, Theorem 4.11 of otto Georgii et al. (1999) adapted to our notations)). Let \mathcal{L} be finite, S a finite subset of \mathbb{R} , μ a strictly positive probability measure on $S^{\mathcal{L}}$, and $X \in S^{\mathcal{L}}$ a random variable with probability measure μ . If μ is monotone, i.e. $\forall i \in \mathcal{L}$ and $a \in S$, $\xi \mapsto \mathbb{P}_{\mu}(X_i \geq a | X_{\mathcal{L} \setminus \{i\}} = \xi)$ is non-decreasing w.r.t. the natural partial order on $S^{\mathcal{L} \setminus \{i\}}$, then it also has positive correlations: for any bounded non-decreasing functions f and g on $S^{\mathcal{L}}$

$$\mathbb{E}_{\mu}[f(X)g(X)] \ge \mathbb{E}_{\mu}[f(X)]\mathbb{E}_{\mu}[g(X)].$$
(6.43)

In our setting, $\mathcal{L} = [|n|]^2$, $S = \{0,1\}$, and μ is the probability measure of the adjacency matrix A. For a given set of influencers \mathcal{I} , the indicator values of the reachable set $X_i = \mathbf{1}\{i \in R(\mathcal{I}, A)\}$ are deterministic functions of the random variables A_{ij} . Thus, let $f_{ij}(\{A_{i'j'}\}_{(i',j')}) = 1 - X_j A_{ji}$. In order to apply the FKG inequality, we first need to show that each $f_{ij} : \{0,1\}^{n^2} \to \{0,1\}$ is non-increasing with respect to the natural partial order on $\{0,1\}^{n^2}$ (i.e. $X \leq Y$ if $X_i \leq Y_i$ for all *i*). Let $u \in \{0,1\}^{n^2}$ be a given state of the edges of the network. In order to prove the non-increasing behavior of f_{ij} , it is sufficient to show that $f_{ii}(u)$ is non-increasing with respect to every $u_{(i,j)}$.

But from Definition 6.4, it is obvious that $X_i(u) = 1 - \prod_{q \in Q_i} (1 - \prod_{(j,l) \in q} u_{(j,l)})$ is non-decreasing with respect to every $u_{(i,j)}$. This implies that $f_{ij}(u) = 1 - X_j(u)u_{(j,i)}$ is non-increasing with respect to every $u_{(i,j)}$ and that $f_{ij} : \{0,1\}^{n^2} \to \{0,1\}$ is nonincreasing with respect to the natural partial order on $\{0,1\}^{n^2}$.

Finally, since the LPC property implies that the probability measure of A is monotonic, we can apply the FKG inequality to $\{1 - X_j A_{ji}\}_{j \in [|n|]}$, and these random variables are positively correlated.

The next lemma ensures that the variables X_i satisfy an implicit inequality that will be the starting point of the proof of Theorem 6.1.

Lemma 6.9. For any \mathcal{I} such that $card(\mathcal{I}) = n_0 < n$ and for any $i \notin \mathcal{I}$, the probability $\mathbb{E}[X_i]$ that node *i* is reachable from \mathcal{I} in *G* verifies:

$$\mathbb{E}[X_i] \le 1 - \exp\left(-\sum_j \mathcal{H}_{ji}\mathbb{E}[X_j]\right)$$
(6.44)

Proof. We first note that a node $i \notin \mathcal{I}$ is reachable from \mathcal{I} if and only if one of its neighbors is reachable from \mathcal{I} in the graph $G \setminus \{i\}$, and the respective ingoing edge transmitted the contagion. Let X_i^{-i} be a binary value indicating if j is reachable from \mathcal{I} in $G \setminus \{i\}$. Then

$$X_i = 0 \Leftrightarrow \forall j \in [|n|] \setminus \{i\}, X_j^{-i} = 0 \text{ or } A_{ji} = 0,$$
(6.45)

which implies the following alternative expression for X_i :

$$1 - X_i = \prod_{j \neq i} (1 - X_j^{-i} A_{ji}).$$
(6.46)

Moreover, the positive correlation of $\{1 - X_i^{-i}A_{ji}\}_{j \in [|n|] \setminus \{i\}}$ implies that

$$\mathbb{E}\left[\prod_{j\neq i} (1 - X_j^{-i} A_{ji})\right] \ge \prod_{j\neq i} \mathbb{E}\left[1 - X_j^{-i} A_{ji}\right]$$
(6.47)

which leads to

$$\mathbb{E}[X_i] \leq 1 - \prod_{j \neq i} \mathbb{E}[1 - X_j^{-i} A_{ji}] \\
= 1 - \prod_{j \neq i} \left(1 - \mathbb{E}[X_j^{-i}] \mathbb{E}[A_{ji}] \right) \\
\leq 1 - \prod_j \left(1 - \mathbb{E}[X_j] \mathbb{E}[A_{ji}] \right)$$
(6.48)

since X_j^{-i} and A_{ji} are independent, due to hypothesis (H1) of the LPC property (see Definition 6.6) and X_j^{-i} only depends on $(A_{kl})_{(k,l) \notin \mathcal{N}_{ji}}$. The second inequality comes from the fact that $X_i^{-i} \leq X_j$ a.s.. Finally,

$$\mathbb{E}[X_i] \leq 1 - \exp\left(\sum_j \ln(1 - \mathbb{E}[X_j]\mathbb{E}[A_{ji}])\right) \\ \leq 1 - \exp\left(\sum_j \ln(1 - \mathbb{E}[A_{ji}])\mathbb{E}[X_j]\right) \\ = 1 - \exp\left(-\sum_j \mathcal{H}_{ji}\mathbb{E}[X_j]\right)$$
(6.49)

since we have on the one hand, for any $x \in [0,1]$ and a < 1, $\ln(1-ax) \ge \ln(1-a)x$, and on the other hand $\mathbb{E}[A_{ji}] = 1 - \exp(-\mathcal{H}_{ji})$ by definition of \mathcal{H} .

Using Lemma 6.9, we are now ready to start the proof of Theorem 6.1.

Proof of Theorem 6.1. In order to simplify notations, we define $Z_i = (\mathbb{E}[X_i])_i$ that we collect in the vector $Z = (Z_i)_{i \in [1...n]}$. Using Lemma 6.9 and convexity of the exponential function, we have for any $u \in \mathbb{R}^n$ such that $\forall i \in \mathcal{I}, u_i = 0$ and $\forall i \notin \mathcal{I}, u_i \ge 0$,

$$u^{\top}Z \le |u|_1 \left(1 - \sum_{i=1}^{n-1} \frac{u_i}{|u|_1} \exp(-(\mathcal{H}^{\top}Z)_i)\right) \le |u|_1 \left(1 - \exp\left(-\frac{Z^{\top}\mathcal{H}u}{|u|_1}\right)\right)$$
(6.50)

where $|u|_1 = \sum_i |u_i|$ is the L_1 -norm of u.

Now taking $u = (1_{i \notin I} Z_i)_i$ and noting that $\forall i, u_i \leq Z_i$, we have

$$\frac{Z^{\top}Z - n_0}{|Z|_1 - n_0} \le 1 - \exp\left(-\frac{Z^{\top}\mathcal{H}Z}{|Z|_1 - n_0}\right) \le 1 - \exp\left(-\frac{\rho_n(Z^{\top}Z - n_0)}{|Z|_1 - n_0} - \frac{\rho_n n_0}{|Z|_1 - n_0}\right)$$
(6.51)

where $\rho_n = \rho(\frac{\mathcal{H} + \mathcal{H}^{\top}}{2})$. Defining $y = \frac{Z^{\top}Z - n_0}{|Z|_1 - n_0}$ and $z = |Z|_1 - n_0 = \sigma(\mathcal{I}) - n_0$, the aforementioned inequality rewrites

$$y \le 1 - \exp\left(-\rho_n y - \frac{\rho_n n_0}{z}\right) \tag{6.52}$$

But by Cauchy-Schwarz inequality applied to u, $(n - n_0)(Z^T Z - n_0) \ge (|Z|_1 - n_0)^2$, which means that $z \le y(n - n_0)$. We now consider the equation

$$x - 1 + \exp\left(-\rho_n x - \frac{\rho_n n_0}{x(n - n_0)}\right) = 0$$
(6.53)

Because the function $f : x \to x - 1 + \exp\left(-\rho_n x + \frac{\rho_n n_0}{x(n-n_0)}\right)$ is continuous, verifies f(1) > 0 and $\lim_{x\to 0^+} f(x) = -1$, Eq. 6.53 admits a solution γ_1 in]0,1[.

We then prove by contradiction that $z \leq \gamma_1(n-n_0)$. Let us assume $z > \gamma_1(n-n_0)$. Then $y \leq 1 - \exp\left(-\rho_n y - \frac{\rho_n n_0}{\gamma_1(n-n_0)}\right)$. But the function $h: x \to x - 1 + \exp\left(-\rho_n x + \frac{\rho_n n_0}{\gamma_1(n-n_0)}\right)$ is convex and verifies h(0) < 0 and $h(\gamma_1) = 0$. Therefore, for any $y > \gamma_1$, $0 = f(\gamma_1) \leq \frac{\gamma_1}{y} f(y) + (1 - \frac{\gamma_1}{y}) f(0)$, and therefore f(y) > 0. Thus, $y \leq \gamma_1$. But $z \leq y(n-n_0) \leq \gamma_1(n-n_0)$ which yields the contradiction. Proof of Corollary 6.1. Using Lemma 6.3 and observing that:

$$\gamma_1 = \gamma\left(\rho_n, \frac{\rho_n n_0}{\gamma_1(n-n_0)}\right) \leq \gamma\left(\rho_n, \frac{\rho_n n_0}{(\gamma_1 - \gamma_0(\rho_n))(n-n_0)}\right),$$

we obtain the following bounds:

$$\gamma_1 \le \gamma_0(\rho_n) + \frac{\rho_n n_0 (1 - \gamma_0(\rho_n))}{(\gamma_1 - \gamma_0(\rho_n))(n - n_0)(1 - \rho_n (1 - \gamma_0(\rho_n)))},$$
(6.54)

and

$$\gamma_1 \le \gamma_0(\rho_n) + \sqrt{\frac{2n_0}{(\gamma_1 - \gamma_0(\rho_n))(n - n_0)}},$$
(6.55)

which lead to

$$\gamma_1 \le \gamma_0(\rho_n) + \sqrt{\frac{\rho_n(1 - \gamma_0(\rho_n))}{1 - \rho_n(1 - \gamma_0(\rho_n))}} \sqrt{\frac{n_0}{n - n_0}},$$
(6.56)

and

$$\gamma_1 \le \gamma_0(\rho_n) + \left(\frac{2n_0}{n-n_0}\right)^{1/3}.$$
 (6.57)

The subcritical and supercritical regimes are obtained using Eq. 6.56 (recall that $\gamma_0(\rho) = 0$ when $\rho \leq 0$) and the critical regime using Eq. 6.57 and Lemma 6.4.

Proofs of Theorem 6.2 and Corollary 6.2

In this subsection, we develop the proofs for Theorem 6.2 and Corollary 6.2 in the case when the set of influencers \mathcal{I} is drawn from a uniform distribution over $\mathcal{P}_{n_0}([|n|])$.

We start with an important lemma that will play the same role in the proof of Theorem 6.2 than Lemma 6.9 in the proof of Theorem 6.1.

Lemma 6.10. Assume \mathcal{I} is drawn from a uniform distribution over $\mathcal{P}_{n_0}([|n|])$. Then, for any $i \in [|n|]$, the probability $\mathbb{E}[X_i]$ that node i is reachable from \mathcal{I} in G satisfies the following implicit inequality:

$$\mathbb{E}[X_i] \le 1 - \frac{n - n_0}{n} \exp\left(-\frac{n}{n - n_0} \sum_j \mathcal{H}_{ji} \mathbb{E}[X_j]\right)$$
(6.58)

Proof.

$$\mathbb{E}[X_{i}] = \mathbb{E}[1_{\{i \in \mathcal{I}\}}] + \mathbb{E}[1_{\{i \notin \mathcal{I}\}}]\mathbb{E}[\mathbb{E}[X_{i}|\mathcal{I}]|i \notin \mathcal{I}]$$

$$\leq \frac{n_{0}}{n} + \frac{n - n_{0}}{n} \left(1 - \mathbb{E}[\exp\left(-\sum_{j}\mathcal{H}_{ji}\mathbb{E}[X_{j}|\mathcal{I}]\right)|i \notin \mathcal{I}]\right)$$

$$\leq \frac{n_{0}}{n} + \frac{n - n_{0}}{n} \left(1 - \exp\left(-\mathbb{E}[\sum_{j}\mathcal{H}_{ji}\mathbb{E}[X_{j}|\mathcal{I}]|i \notin \mathcal{I}]\right)\right)$$

$$= 1 - \frac{n - n_{0}}{n} \exp\left(-\sum_{j}\mathcal{H}_{ji}\mathbb{E}[X_{j}|i \notin \mathcal{I}]\right)$$

$$\leq 1 - \frac{n - n_{0}}{n} \exp\left(-\frac{n}{n - n_{0}}\sum_{j}\mathcal{H}_{ji}\mathbb{E}[X_{j}]\right)$$
(6.59)

6.9. PROOFS

where the first inequality is Lemma 6.9 and the second one is Jensen inequality for conditional expectations. $\hfill \Box$

Proof of Theorem 6.2. We define $Z_i = (\mathbb{E}[X_i])_i$ that we collect in the vector $Z = (Z_i)_{i \in [1...n]}$. Then, using Lemma 6.10, and convexity of exponential function, we have:

$$\frac{Z^{\top}Z}{|Z|_{1}} \leq 1 - \frac{n - n_{0}}{n} \sum_{i=1}^{n} \frac{Z_{i}}{|Z|_{1}} \exp\left(-\frac{n}{n - n_{0}} (\mathcal{H}^{\top}Z)_{i}\right) \\
\leq 1 - \frac{n - n_{0}}{n} \exp\left(-\frac{n}{n - n_{0}} \frac{Z^{\top}\mathcal{H}Z}{|Z|_{1}}\right) \\
\leq 1 - \frac{n - n_{0}}{n} \exp\left(-\frac{n\rho_{n}}{n - n_{0}} \frac{Z^{\top}Z}{|Z|_{1}}\right),$$
(6.60)

which leads, due to the monotonicity of $x \mapsto 1 - \frac{n-n_0}{n} \exp(-\frac{n\rho_n}{n-n_0}x)$, to

$$\frac{Z^{\top}Z}{|Z|_{1}} \leq \gamma \left(\frac{n\rho_{n}}{n-n_{0}}, -\ln(1-\frac{n_{0}}{n})\right) = (1-\frac{n_{0}}{n})\gamma \left(\rho_{n}, \frac{\rho_{n}n_{0}}{n-n_{0}}\right) + \frac{n_{0}}{n}.$$
 (6.61)

Finally, we have by Cauchy-Schwarz inequality $\sigma_U = |Z|_1 \le n \frac{Z^\top Z}{|Z|_1}$, which proves the proposition.

Proof of Corollary 6.2. Using Lemma 6.3, we obtain the following bounds:

$$\gamma\left(\rho_{n}, \frac{\rho_{n}n_{0}}{n-n_{0}}\right) \leq \gamma_{0}(\rho_{n}) + \frac{\rho_{n}n_{0}(1-\gamma_{0}(\rho_{n}))}{(n-n_{0})(1-\rho_{n}(1-\gamma_{0}(\rho_{n})))},$$
(6.62)

and

$$\gamma(\rho_n, \frac{\rho_n n_0}{n - n_0}) \le \gamma_0(\rho_n) + \sqrt{\frac{2n_0}{n - n_0}},$$
(6.63)

The subcritical and supercritical regimes are obtained using Eq. 6.62 (recall that $\gamma_0(\rho) = 0$ when $\rho \leq 0$) and the critical regime using Eq. 6.63 and Lemma 6.4.

Proofs of Theorem 6.3 and Corollary 6.3

In this subsection, we develop the proofs for Theorem 6.3 and Corollary 6.3 in the case when each node belongs to the set of influencers \mathcal{I} independently at random with probability q.

We start with an important lemma that will play the same role in the proof of Theorem 6.3 than Lemma 6.9 in the proof of Theorem 6.1.

Lemma 6.11. Assume each node is an influencer with independent probability $q \in [0,1]$ and denote by \mathcal{I} the random set of influencers that is drawn. Then, for any $i \in [|n|]$, the probability $\mathbb{E}[X_i]$ that node *i* is reachable from \mathcal{I} in *G* satisfies the following implicit inequality:

$$\mathbb{E}[X_i] \le 1 - (1 - q) \exp\left(-\sum_j \mathcal{H}_{ji} \mathbb{E}[X_j]\right)$$
(6.64)

Proof.

$$\mathbb{E}[X_{i}] = \mathbb{E}[1_{\{i \in \mathcal{I}\}}] + \mathbb{E}[1_{\{i \notin \mathcal{I}\}}]\mathbb{E}[\mathbb{E}[X_{i}|\mathcal{I}]|i \notin \mathcal{I}]$$

$$\leq q + (1-q)\left(1 - \mathbb{E}[\exp\left(-\sum_{j}\mathcal{H}_{ji}\mathbb{E}[X_{j}|\mathcal{I}]\right)|i \notin \mathcal{I}]\right)$$

$$\leq q + (1-q)\left(1 - \exp\left(-\mathbb{E}[\sum_{j}\mathcal{H}_{ji}\mathbb{E}[X_{j}|\mathcal{I}]|i \notin \mathcal{I}]\right)\right)$$

$$= 1 - (1-q)\exp\left(-\sum_{j}\mathcal{H}_{ji}\mathbb{E}[X_{j}|i \notin \mathcal{I}]\right)$$

$$\leq 1 - (1-q)\exp\left(-\sum_{j}\mathcal{H}_{ji}\mathbb{E}[X_{j}]\right)$$
(6.65)

where the first inequality is Lemma 6.9, the second one is Jensen's inequality for conditional expectations, and the third is the positive correlation of X_i and $\mathbf{1}\{i \in \mathcal{I}\}$.

Proof of Theorem 6.3. We define $Z_i = (\mathbb{E}[X_i])_i$ that we collect in the vector $Z = (Z_i)_{i \in [1...n]}$. Then, using Lemma 6.11, and convexity of exponential function, we have:

$$\frac{Z^{\top}Z}{|Z|_{1}} \leq 1 - (1 - q) \sum_{i=1}^{n} \frac{Z_{i}}{|Z|_{1}} \exp\left(-(\mathcal{H}^{\top}Z)_{i}\right) \\
\leq 1 - (1 - q) \exp\left(-\frac{Z^{\top}\mathcal{H}Z}{|Z|_{1}}\right) \\
\leq 1 - (1 - q) \exp\left(-\rho_{n}\frac{Z^{\top}Z}{|Z|_{1}}\right),$$
(6.66)

which leads, due to the monotonicity of $x \mapsto 1 - (1 - q) \exp(-\rho_n x)$, to

$$\frac{Z^{\top}Z}{|Z|_1} \le \gamma(\rho_n, -\ln(1-q)).$$
(6.67)

Finally, we have by Cauchy-Schwarz inequality $\sigma_R = |Z|_1 \le n \frac{Z^\top Z}{|Z|_1}$, which proves the proposition.

Proof of Corollary 6.2. Using Lemma 6.3, we obtain the following bounds:

$$\gamma(\rho_n, -\ln(1-q)) \le \gamma_0(\rho_n) + \frac{-\ln(1-q)(1-\gamma_0(\rho_n))}{1-\rho_n(1-\gamma_0(\rho_n))},$$
(6.68)

and

$$\gamma(\rho_n, -\ln(1-q)) \le \gamma_0(\rho_n) + \sqrt{-2\ln(1-q)},$$
 (6.69)

The subcritical and supercritical regimes are obtained using Eq. 6.68 (recall that $\gamma_0(\rho) = 0$ when $\rho \leq 0$) and the critical regime using Eq. 6.69 and Lemma 6.4.

6.9. PROOFS

Proof of Proposition 6.1

Proof of Proposition 6.1. Let $G_{a,b} = G(n, A)$ be a "random star-network", i.e. an undirected random graph such that $\{A_{ij} : i < j\}$ are independent Bernoulli random variables of parameter $a \in [0,1]$ if i = 1 and b < a otherwise. Then, the next Lemma shows that the influence of node 1 in $G_{a,b}$ is lower bounded by the size of the giant component of an Erdös-Rényi graph.

Lemma 6.12. Let $G_{a,b} = G(n, A)$ be a "random star-network" of parameters a and b, and G(n,p) an Erdös-Rényi graph of size n and parameter p. The influence $\sigma_{a,b}(\{1\})$ of node 1 in $G_{a,b}$ is lower bounded by

$$\sigma_{a,b}(\{1\}) \ge 1 - \frac{1}{ae} + na + (1 - a)\mathbb{E}[C_1(G(n - 1, b))],$$
(6.70)

where $C_1(G)$ denotes the size of the giant component of G.

Proof. Since the edge presence variables A_{ij} are independent, the set of nodes linked to 1 in $G_{a,b}$ is a random set $\mathcal{I}(a)$ such that each node in $\{2,...,n\}$ belongs to it independently with probability a. Also, $\mathcal{I}(a)$ is independent from the subgraph restricted to $\{2,...,n\}$, and since each edge in $\{2,...,n\}$ is drawn independently and has probability b, this subgraph is an Erdös-Rényi graph of size n - 1 and parameter b. Hence, if $\mathbb{E}[\sigma_b(\mathcal{I}(a))]$ is the influence of a random set $\mathcal{I}(a)$ in G(n - 1, b) as defined in Theorem 6.3, then

$$\mathcal{T}_{a,b}(\{1\}) = 1 + \mathbb{E}[\sigma_b(\mathcal{I}(a))]. \tag{6.71}$$

Hence, Eq. 6.14 and the same derivation as in Theorem 6.4 gives that:

$$\sigma_{a,b}(\{1\}) \geq 1 + na + (1-a)\mathbb{E}[C_1(G(n-1,b))(1-(1-a)^{C_1(G(n-1,b))-1})]$$

$$\geq 1 + na + (1-a)\mathbb{E}[C_1(G(n-1,b))] - \frac{1}{-\ln(1-a)e}$$

$$\geq 1 + na + (1-a)\mathbb{E}[C_1(G(n-1,b))] - \frac{1}{ae}.$$

(6.72)

However, a simple calculation holds $\rho_n = \frac{(n-2)b' + \sqrt{(n-2)^2b'^2 + 4(n-1)a'^2}}{2}$, where $a' = -\ln(1-a)$ and $b' = -\ln(1-b)$. We now conclude in the three regimes:

Subcritical regime $(\rho_n < 1)$: In this case, we take b = 0 and $a = \frac{\rho}{\sqrt{n-1}}$, for $\rho \in [0,1)$. Then $\rho_n = \rho + O(\frac{1}{\sqrt{n}})$ and $\sigma_{a,b} = 1 + \rho\sqrt{n-1} \ge \frac{\rho}{2}\sqrt{n}$ for *n* sufficiently large.

Critical and supercritical regime $(\rho_n \ge 1)$: In this case, we take $a = \frac{1}{\sqrt{n \ln n}}$ and $b = \frac{\rho}{n}$. Then $\rho_n = \rho + O(\frac{1}{\ln n})$ and

$$\sigma_{a,b} \ge O(\sqrt{n \ln n}) + \mathbb{E}[C_1(G(n-1,b))].$$
(6.73)

However, classical results in percolation theory (Erdös and Rényi, 1960) state that, for $\eta > 0$ and $\omega(n)$ any function s.t. $\lim_{n \to +\infty} \omega(n) = +\infty$,

$$C_1(G(n, \frac{1}{n})) \ge \frac{n^{2/3}}{\omega(n)}$$
 a.a.s. (6.74)

and

$$\left|\frac{C_1(G(n,\frac{\rho}{n}))}{n} - \gamma_0(c)\right| \le \eta \text{ a.a.s.}$$
(6.75)

Hence, in the first case, Markov's inequality gives $\mathbb{E}[C_1(G(n-1,b))] \ge \frac{n^{2/3}}{\omega(n)}(1-o(1))$, which leads to, for *n* sufficiently large,

$$\mathbb{E}[C_1(G(n-1,b))] \ge C_{\rho} n^{2/3}, \tag{6.76}$$

for some $C_{\rho} > 0$, since assuming $\liminf_{n \to +\infty} \frac{\mathbb{E}[C_1(G(n-1,b))]}{n^{2/3}} = 0$ and taking $\omega(n) = \max_{m \le n} \sqrt{\frac{n^{2/3}}{\mathbb{E}[C_1(G(n-1,b))]}}$ leads to $\liminf_{n \to +\infty} \sqrt{\frac{\mathbb{E}[C_1(G(n-1,b))]}{n^{2/3}}} \ge 1$, which contradicts the assumption. For the second case, Markov's inequality gives

$$\mathbb{E}[C_1(G(n-1,b))] \geq (\gamma_0(\rho)n - \eta n)\mathbb{P}(C_1(G(n-1,b)) \geq \gamma_0(\rho)n - \eta n) \\ \geq \gamma_0(\rho)n - \eta n - o(n).$$
(6.77)

Taking the limit inferior leads to, for all $\eta > 0$:

$$\liminf_{n \to +\infty} \frac{\mathbb{E}[C_1(G(n-1,b))] - \gamma_0(\rho)n}{n} \ge \liminf_{n \to +\infty} \{-\eta - o(1)\} = -\eta , \qquad (6.78)$$

and thus $\liminf_{n \to +\infty} \frac{\mathbb{E}[C_1(G(n-1,b))] - \gamma_0(\rho)n}{n} \ge 0$, which can be rewritten as $\mathbb{E}[C_1(G(n-1,b))] \ge \gamma_0(\rho)n - o(n)$.

6.9.3 **Proofs of the percolation theorems**

The aim of this section is to prove the results obtained in section 6.4 for the bond percolation problem from the general results on reachablity sets of section 6.3. We recall that we consider an undirected random graph G(n, A) of size n with independent edge presence variables $\{A_{ij} : i < j\}$, and denote by $C_k(G)$ the size of its k^{th} -largest connected component, as well as N(m) the number of connected components of G of cardinality greater than or equal to m. We also recall that we are able to relate the distribution of the sizes of connected components of G to the Hazard function through equation 6.14. Let a > 0, then:

$$\mathbb{E}\left[\sum_{k} C_{k}(G) \left(1 - e^{-aC_{k}(G)}\right)\right] \leq \gamma(\rho_{n}, a)n$$

Proofs of Theorem 6.4 and Corollary 6.4

Proof of Theorem 6.4. Theorem 6.4 is simply obtained by combining equation 6.14 and the following observation:

$$\sum_{k} C_{k}(G) \left(1 - e^{-aC_{k}(G)} \right) \le C_{1}(G) \left(1 - e^{-aC_{1}(G)} \right) + (n - C_{1}(G))(1 - e^{-a})$$

Therefore,

$$\mathbb{E}[C_1(G)(1 - e^{-a(C_1(G)-1)})] \le ne^a \left(\gamma(\rho_n, a) - 1 + e^{-a}\right) = n \left(1 - e^{-\rho_n \gamma(\rho_n, a)}\right).$$

6.9. PROOFS

Proof of Corollary 6.4. We first prove the subcritical result. Let $a \ge 0$. When $\rho_n < 1$, we have $\gamma(\rho_n, a) = 0$ and therefore Lemma 6.3 implies $\gamma(\rho_n, a) \le \frac{a}{1-\rho_n}$. By convexity of exponential function, we get:

$$\mathbb{E}[C_1(G)(1 - e^{-a(C_1(G) - 1)})] \le \frac{na\rho_n}{1 - \rho_n}$$

This inequality between two derivable functions of *a* such that $f(a) \le g(a)$ for all $a \ge 0$ and f(0) = g(0) implies that $\frac{\partial f}{\partial a}(0) \le \frac{\partial g}{\partial a}(0)$ which yields:

$$\mathbb{E}[C_1(G)(C_1(G)-1)] \leq \frac{n\rho_n}{1-\rho_n}.$$

The first equation of Corollary 6.4 is then a straightforward resolution of a second-order equation, using the fact that $\mathbb{E}[C_1(G)^2] \ge \mathbb{E}[C_1(G)]^2$.

For the critical and supercritical results, we will make use of the fact that, for all a > 0, $C_1(G)e^{-aC_1(G)} \le \frac{1}{ae}$, which yields:

$$\mathbb{E}[C_1(G)] \le \frac{e^a}{ae} + n\left(1 - e^{-\rho_n \gamma(\rho_{n,a})}\right)$$
(6.79)

which rewrites $\mathbb{E}[C_1(G)] + (n - \mathbb{E}[C_1(G)])(1 - e^{-a}) \leq \frac{1}{ae} + n\gamma(\rho_n, a)$ and therefore implies:

$$\mathbb{E}[C_1(G)] \le \frac{1}{ae} + n\gamma(\rho_n, a) \tag{6.80}$$

From Lemma 6.3, we know that for $\rho_n \neq 0$, $\gamma(\rho_n, a) \leq \gamma_0(\rho_n) + \sqrt{2a}$. We therefore get the critical result using equation 6.80:

$$\mathbb{E}[C_1(G)] \le n\gamma_0(\rho_n) + \min_{a>0} \left\{ \frac{1}{ae} + n\sqrt{2a} \right\} = n\gamma_0(\rho_n) + n^{2/3} \left(\frac{27}{2e}\right)^{1/3}$$

For the supercritical result, we use equation 6.79 and the fact that

$$\begin{split} 1 - e^{-\rho_n \gamma(\rho_n, a)} &= 1 - e^{-\rho_n \gamma_0(\rho_n)} \left(1 - e^{-\rho_n (\gamma(\rho_n, a) - \gamma_0(\rho_n))} \right) \\ &\leq \gamma_0(\rho_n) + \rho_n \left(1 - \gamma_0(\rho_n) \right) \left(\gamma(\rho_n, a) - \gamma_0(\rho_n) \right) \\ &\leq \gamma_0(\rho_n) + \frac{a\rho_n \left(1 - \gamma_0(\rho_n) \right)^2}{1 - \rho_n + \rho_n \gamma_0(\rho_n)}. \end{split}$$

We then choose

$$a = \sqrt{\frac{2(1 - \rho_n + \rho_n \gamma_0(\rho_n))}{2e\rho_n (1 - \gamma_0(\rho_n))^2 n + 1 - \rho_n + \rho_n \gamma_0(\rho_n)}}$$

which gives us

$$\mathbb{E}[C_1(G)] \le n\gamma_0(\rho_n) + \frac{2}{e}\sqrt{\frac{en\rho_n(1-\gamma_0(\rho_n))^2}{1-\rho_n+\rho_n\gamma_0(\rho_n)}} + \frac{1}{2} + \frac{1}{e} + \frac{e^a - 1}{ae}$$

Using the fact that $a < \sqrt{2}$ and $\sqrt{x+y} \le \sqrt{x} + \sqrt{y}$, we finally get:

$$\mathbb{E}[C_1(G)] \le n\gamma_0(\rho_n) + \frac{2}{\sqrt{e}}\sqrt{\frac{n\rho_n(1-\gamma_0(\rho_n))^2}{1-\rho_n+\rho_n\gamma_0(\rho_n)}} + \frac{1+\sqrt{2}+e^{\sqrt{2}}}{\sqrt{2}e}$$

which yields the supercritical result.

Proofs of Theorem 6.5 and Corollary 6.7

Proof of Theorem 6.5. Let a > 0. In order to prove Theorem 6.5, we start again from equation 6.14 and use the fact that:

$$\sum_{k} C_{k}(G) \left(1 - e^{-aC_{k}(G)} \right) \mathbf{1} \{ C_{k}(G) \ge m \} \le (1 - e^{-am}) \sum_{k} C_{k}(G) \mathbf{1} \{ C_{k}(G) \ge m \} \text{ and}$$
$$\sum_{k} C_{k}(G) \left(1 - e^{-aC_{k}(G)} \right) \mathbf{1} \{ C_{k}(G) < m \} \le (1 - e^{-a}) \left(n - \sum_{k} C_{k}(G) \mathbf{1} \{ C_{k}(G) \ge m \} \right).$$

Therefore, we have:

$$\sum_{k} C_{k}(G) \mathbf{1}\{C_{k}(G) \ge m\} \le \frac{ne^{a} \left(\gamma(\rho_{n}, a) - 1 + e^{-a}\right)}{1 - e^{-a(m-1)}} = \frac{n\left(1 - e^{-\rho_{n}\gamma(\rho_{n}, a)}\right)}{1 - e^{-a(m-1)}}.$$

which proves the theorem, noting that $mN(m) \leq \sum_k C_k(G) \mathbf{1}\{C_k(G) \geq m\}$

Proof of Corollary 6.7. In the subcritical case, we have $\gamma(\rho_n, a) \leq \frac{a}{1-\rho_n}$ which means that, for all a > 0, $\rho_n < 1$:

$$N(m) \leq \frac{n}{m} \frac{a\rho_n}{(1-\rho_n)(1-e^{-a(m-1)})}$$

The right-hand side function of *a* is increasing on the semi-line, and we therefore takes its limit when $a \rightarrow 0$ to get the subcritical result.

For the critical case, we note that Theorem 6.5 implies that, for all a > 0:

$$(1 - e^{-am})mN(m) + (1 - e^{-a})(n - mN(m)) \le n\gamma(\rho_n, a)$$

and therefore

$$N(m) \leq \frac{n}{m} \frac{\gamma(\rho_n, a)}{1 - e^{-am}} \leq \frac{n}{m} \frac{\gamma_0(\rho_n) + \sqrt{2a}}{1 - e^{-am}} = \frac{n}{m^{3/2}} \frac{\gamma_0(\rho_n)\sqrt{m} + \sqrt{2am}}{1 - e^{-am}}.$$

The function $x \mapsto \frac{\sqrt{2x}}{1-e^{-x}}$ admits a unique minimum for x > 0 in η which is the strictly positive solution of $e^{\eta} = 2\eta + 1$. Setting $a = \frac{\eta}{m}$ yields:

$$N(m) \le \frac{n}{m^{3/2}} \frac{\gamma_0(\rho_n)\sqrt{m} + \sqrt{2\eta}}{1 - e^{-\eta}}$$

Hence, if $\rho_n \leq 1 - \nu m^{-1/2}$ for a fixed $\nu > 0$, Lemma 6.4 implies:

$$N(m) \le \frac{n}{m^{3/2}} \frac{2\nu + \sqrt{2\eta}}{1 - e^{-\eta}}.$$

The critical result is given by finding the value ν for which the first orders of the subcritical and critical bounds are equal at the threshold value $\rho = 1 - \nu m^{-1/2}$, i.e. ν is the solution of $\frac{1}{\nu} = \frac{2\nu + \sqrt{2\eta}}{1 - e^{-\eta}}$.

For the supercritical result, we will make use of the fact that

$$1-e^{-\rho_n\gamma(\rho_n,a)}\leq \gamma_0(\rho_n)+\frac{a\rho_n\left(1-\gamma_0(\rho_n)\right)^2}{1-\rho_n+\rho_n\gamma_0(\rho_n)}.$$

6.9. PROOFS

Introducing $B = \frac{\rho_n (1 - \gamma_0(\rho_n))^2}{1 - \rho_n + \rho_n \gamma_0(\rho_n)}$, Theorem 6.5 gives:

$$N(m) \le \frac{n}{m} \left(\frac{\gamma_0(\rho_n) + Ba}{1 - e^{a(m-1)}} \right) \tag{6.81}$$

Deriving the right-hand side with respect to *a* and setting x = a(m-1), we find that the minimizer x^* is given by the unique strictly positive solution of $e^x = 1 + x + \frac{B}{\gamma_0(\rho_n)}(m-1)$. Therefore, we now in particular that $x^* \leq \sqrt{2(e^{x^*} - 1 - x^*)} = \sqrt{2\gamma_0(\rho_n)(m-1)/B}$. The supercritical result is obtained by plugging

$$a = \sqrt{\frac{2\gamma_0(\rho_n)}{B(m-1)}}$$

into equation 6.81.

L		
L		
L	_	

Dynamic properties of the influence

"We are all in the gutter, but some of us are looking at the stars."

- Oscar Wilde

7

Contents

7.1	Introd	luction	86
	7.1.1	Related works	86
	7.1.2	Outline	86
7.2	Conti	nuous-Time Information Cascades	87
	7.2.1	Information propagation and influence in diffusion networks \ldots	87
	7.2.2	The Laplace Hazard Matrix	87
	7.2.3	Existence of a critical time of a contagion	88
7.3	Theor	etical bounds for the influence of a set of nodes	88
	7.3.1	Upper bounds on the maximum influence at time T	88
	7.3.2	Lower bound on the critical time of a contagion	89
7.4	Appli	cation to particular contagion models	90
	7.4.1	Fixed transmission pattern	90
	7.4.2	Exponential transmission probabilities	90
	7.4.3	SI and SIR epidemic models	91
	7.4.4	Discrete-time Information Cascade	92
7.5	Exper	Experimental results	
7.6	Proofs	s	94
	7.6.1	Proof for the definition of critical time	94
	7.6.2	Proofs of the upper bounds on influence	95
	7.6.3	Proofs of the upper bounds on critical time	97
	7.6.4	Proofs for the particular contagion models	97

7.1 Introduction

Diffusion networks capture the underlying mechanism of how events propagate throughout a complex network. In marketing, social graph dynamics have caused large transformations in business models, forcing companies to re-imagine their customers not as a mass of isolated economic agents, but as *customer networks* (Trusov et al., 2009). In epidemiology, a precise understanding of spreading phenomena is heavily needed when trying to break the chain of infection in populations during outbreaks of viral diseases. But whether the subject is a virus spreading across a computer network, an innovative product among early adopters, or a rumor propagating on a network of people, the questions of interest are the same: how many people will it infect? How fast will it spread? And, even more critically for decision makers: how can we modify its course in order to meet specific goals?

7.1.1 Related works

Several papers tackled these issues by studying the influence maximization problem. Given a known diffusion process on a graph, it consists in finding the top-k subset of initial seeds with the highest expected number of infected nodes at a certain time distance T. This problem being NP-hard (Kempe et al., 2003), various heuristics have been proposed in order to obtain scalable suboptimal approximations. While the first algorithms focused on discretetime models and the special case $T = +\infty$ (Chen et al., 2009, 2010), subsequent papers by Gomez-Rodriguez et al. (2011) and Du et al. (2013b) brought empirical evidences of the key role played by temporal behavior. Existing models of continuous-time stochastic processes include multivariate Hawkes processes (Hawkes and Oakes, 1974) where recent progress in inference methods (see for example Zhou et al. (2013) and Lemonnier and Vavatis (2014)) made available the tools for the study of activity shaping (Farajtabar et al., 2014), which is closely related to influence maximization. However, in the most studied case in which each node of the network can only be infected once, the most widely used model remains the Continuous-Time Information Cascade (CTIC) model introduced by Gomez-Rodriguez et al. (2011). Under this framework, successful inference (Gomez-Rodriguez et al., 2011) as well as influence maximization algorithms have been developed (Rodriguez and Schölkopf, 2012; Du et al., 2013a).

However, if recent works by Gomez-Rodriguez et al. (2015) and Pouget-Abadie and Horel (2015) provided theoretical foundations for the inference problem, assessing the quality of influence maximization remains a challenging task, as few theoretical results exist for general graphs. In the infinite-time setting, studies of the SIR diffusion process in epidemiology by Draief et al. (2008) or percolation for specific graphs by Bollobás et al. (2007) provided a more accurate understanding of these processes. In Chap. 6, we showed that the spectral radius of the *Hazard matrix* played a key role in the long-term influence of information cascades, and this chapter investigates the dynamic properties induced by this quantity.

7.1.2 Outline

The rest of this chapter is organized as follows. In Sec. 7.2, we recall the definition of Information Cascades Model and introduce useful notations. In Sec. 7.3, we derive theoretical bounds for the influence. In Sec. 7.4, we illustrate our results by applying them on specific cascade models. In Sec. 7.5, we perform experiments in order to show that our bounds are sharp for a family of graphs and sets of initial nodes. All proof details are provided in Sec. 7.6. This chapter is an extended version of the paper (Scaman et al., 2015b) published at the NIPS 2015 conference in collaboration with Rémi Lemonnier and Nicolas Vayatis.

7.2 Continuous-Time Information Cascades

7.2.1 Information propagation and influence in diffusion networks

Continuous-Time Information Cascades are diffusion processes that were developed for the analysis of communication networks, and more specifically for processes during which each node can be infected only once. The precise definition of this model is given in Sec. 5.4.3, and $p_{ij}(t)$ will denote the sub-probability density of the transmission times T_{ij} of Proposition 5.2. For each node $v \in V$, we will denote as τ_v the (possibly infinite) time at which it is reached by the infection. We also recall that the *influence* of \mathcal{I} at time T, denoted as $\sigma_{\mathcal{I}}(T)$, is defined as the expected number of nodes reached by the contagion at time T originating from \mathcal{I} , i.e.

$$\sigma_{\mathcal{I}}(T) = \mathbb{E}\left[\sum_{v \in \mathcal{V}} \mathbb{1}\{\tau_v \le T\}\right],\tag{7.1}$$

where the expectation is taken over cascades originating from \mathcal{I} (i.e. $\tau_v = 0 \Leftrightarrow v \in \mathcal{I}$). See Definition 5.10 for a generic definition of this quantity. Following the percolation literature, we will differentiate between *sub-critical* cascades whose size is o(n) and *supercritical* cascades whose size is proportional to n, where n denotes the size of the network. In this chapter, we will focus on upper bounding the influence $\sigma_{\mathcal{I}}(T)$ for any given time Tand characterizing the critical times at which phase transitions occur between sub-critical and super-critical behaviors.

7.2.2 The Laplace Hazard Matrix

We extend here the concept of *Hazard matrix* of Chap. 6 which plays a key role for the influence of the information cascade.

Definition 7.1. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph, and p_{ij} be integrable edge transmission probabilities such that $\int_0^{+\infty} p_{ij}(t)dt < 1$. For $s \ge 0$, let $\mathcal{L}(s)$ be the $n \times n$ matrix, denoted as the Laplace Hazard matrix, whose coefficients are

$$\mathcal{L}_{ij}(s) = \begin{cases} -\hat{p}_{ij}(s) \left(\int_0^{+\infty} p_{ij}(t)dt\right)^{-1} \ln\left(1 - \int_0^{+\infty} p_{ij}(t)dt\right) & \text{if } (i,j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}, \quad (7.2)$$

where $\hat{p}_{ij}(s)$ denotes the Laplace transform of p_{ij} defined for every $s \ge 0$ by $\hat{p}_{ij}(s) = \int_0^{+\infty} p_{ij}(t)e^{-st}dt$. Note that the long term behavior of the cascade is retrieved when s = 0 and coincides with the concept of Hazard matrix of Definition 6.1.

7.2.3 Existence of a critical time of a contagion

In the following, we will derive critical times before which the contagion is sub-critical, and above which the contagion is super-critical. We now formalize this notion of critical time via limits of contagions on networks.

Theorem 7.1. Let $(\mathcal{G}_n)_{n \in \mathbb{N}}$ be a sequence of networks of size n, and $(p_{ij}^n)_{n \in \mathbb{N}}$ be transmission probability functions along the edges of \mathcal{G}_n . Let also $\sigma_n(t)$ be the maximum influence in \mathcal{G}_n at time t from a single influencer. Then there exists a critical time $T^c \in \mathbb{R}_+ \cup \{+\infty\}$ such that, for every sequence of times $(T_n)_{n \in \mathbb{N}}$:

- If $\limsup_{n \to +\infty} T_n < T^c$, then $\sigma_n(T_n) = o(n)$,
- If $\liminf_{n \to +\infty} T_n > T^c$, then $\limsup_{n \to +\infty} \sigma_n(T_n)/n > 0$.

Moreover, such a critical time is unique.

In other words, the *critical time* is a time before which the regime is *sub-critical* and after the contagion is *super-critical*. In order to simplify notations, we will omit in the following the dependence in *n* of all the variables whenever stating results holding in the limit $n \rightarrow +\infty$.

7.3 Theoretical bounds for the influence of a set of nodes

We now present our upper bounds on the influence at time T and derive a lower bound on the critical time of a contagion.

7.3.1 Upper bounds on the maximum influence at time T

The next proposition provides an upper bound on the influence at time *T* for any set of influencers \mathcal{I} such that $|\mathcal{I}| = n_0$. This result may be valuable for assessing the quality of influence maximization algorithms in a given network.

Proposition 7.1. Define $\rho(s) = \rho(\frac{\mathcal{L}(s) + \mathcal{L}(s)^{\top}}{2})$. Then, for any \mathcal{I} such that $|\mathcal{I}| = n_0 < n$, denoting by $\sigma_{\mathcal{I}}(T)$ the expected number of nodes reached by the cascade starting from \mathcal{I} at time T:

$$\sigma_{\mathcal{I}}(T) \le n_0 + (n - n_0) \min_{s \ge 0} \gamma_1\left(\rho(s), \frac{n_0}{n - n_0}\right) e^{sT}.$$
(7.3)

where $\gamma_1(\rho, a)$ is as defined in Theorem 6.1.

Corollary 7.1. Under the same assumptions:

$$\sigma_{\mathcal{I}}(T) \le n_0 + \sqrt{n_0(n - n_0)} \min_{\{s \ge 0 \mid \rho(s) < 1\}} \left(\sqrt{\frac{\rho(s)}{1 - \rho(s)}} e^{sT} \right), \tag{7.4}$$

Proof. This is a direct consequence of Corollary 6.1.

7.3. THEORETICAL BOUNDS FOR THE INFLUENCE OF A SET OF NODES

Note that the long-term upper bound in Theorem 6.1 is a corollary of Proposition 7.1 using s = 0. When $\rho(0) < 1$, Corollary 7.1 with s = 0 implies that the regime is subcritical for all $T \ge 0$. When $\rho(0) \ge 1$, the long-term behavior may be super-critical and the influence may reach linear values in n. However, at a cost growing exponentially with T, it is always possible to choose a s such that $\rho(s) < 1$ and retrieve a $O(\sqrt{n})$ behavior. While the exact optimal parameter s is in general not explicit, two choices of s derive relevant results: either simplifying e^{sT} by choosing s = 1/T, or keeping $\gamma(s)$ sub-critical by choosing s s.t. $\rho(s) < 1$. In particular, the following corollary shows that the contagion explodes at most as $\exp(\rho^{-1}(1-\epsilon)T)$ for any $\epsilon \in [0,1]$.

Corollary 7.2. Let $\epsilon \in [0,1]$ and $\rho(0) \ge 1$. Under the same assumptions, the two following bounds hold:

$$\sigma_{\mathcal{I}}(T) \le n_0 + \sqrt{\frac{n_0(n-n_0)}{\epsilon}} \exp\left(\rho^{-1}(1-\epsilon)T\right).$$
(7.5)

and

$$\sigma_{\mathcal{I}}(T) \le n_0 + (2n_0)^{1/3} (n - n_0)^{2/3} \exp\left(\rho^{-1}(1)T\right).$$
(7.6)

Remark. Since this section focuses on bounding $\sigma_{\mathcal{I}}(T)$ for a given $T \ge 0$, all the aforementioned results also hold for $p_{ij}^T(t) = p_{ij}(t)\mathbb{1}\{t \le T\}$. This is equivalent to integrating everything on [0, T] instead of \mathbb{R}_+ , i.e.

$$\mathcal{L}_{ij}(s) = -\ln(1 - \int_0^T p_{ij}(t)dt) (\int_0^T p_{ij}(t)dt)^{-1} \int_0^T p_{ij}(t)e^{-st}dt.$$

This choice of \mathcal{L} is particularly useful when some edges are transmitting the contagion with probability 1, see for instance the SI epidemic model in Sec. 7.4.3.

7.3.2 Lower bound on the critical time of a contagion

The previous section presents results about how explosive a contagion is. These findings suggest that the speed at which a contagion explodes is bounded by a certain quantity, and thus that the process needs a certain amount of time to become super-critical. This intuition is made formal in the following corollary:

Corollary 7.3. Assume that $\forall n \geq 0$, $\rho_n(0) \geq 1$. If the sequence $(T_n)_{n \in \mathbb{N}}$ is such that

$$\limsup_{n \to +\infty} \frac{3\rho_n^{-1}(1)T_n}{\ln n} < 1.$$
(7.7)

Then,

$$\sigma_{\mathcal{I}}(T_n) = o(n). \tag{7.8}$$

In other words, the regime of the contagion is *sub-critical* before $\frac{\ln n}{3q_{r}^{-1}(1)}$ and

$$T^{c} \ge \liminf_{n \to +\infty} \frac{\ln n}{3\rho_{n}^{-1}(1)}.$$
(7.9)

Under an additional technical constraint, this lower bound can be further improved:

Corollary 7.4. With the additional constraint that $\lim_{n\to+\infty} \frac{\rho_n^{-1}(1-\frac{1}{\ln n})}{\rho_n^{-1}(1)} = 1$,

$$T^{c} \ge \liminf_{n \to +\infty} \frac{\ln n}{2\rho_{n}^{-1}(1)}.$$
(7.10)

The technical condition $\lim_{n\to+\infty} \rho_n^{-1}(1-\frac{1}{\ln n})/\rho_n^{-1}(1) = 1$ imposes that, for large n, $\lim_{\epsilon\to 0} \rho_n^{-1}(1-\epsilon)/\rho_n^{-1}(1)$ converges sufficiently fast to 1 so that $\rho_n^{-1}(1-\frac{1}{\ln n})$ has the same behavior as $\rho_n^{-1}(1)$. This condition is not very restrictive, and is met for the different case studies considered in Sec. 7.4. This result may be valuable for decision makers since it provides a safe time region in which the contagion has not reached a macroscopic scale. It thus provides insights into *how long* do decision makers have to prepare control measures. After T^c , the process explodes and immediate action is required.

7.4 Application to particular contagion models

In this section, we provide several examples of cascade models that show that our theoretical bounds are applicable in a wide range of scenarios and provide the first results of this type in many areas, including two widely used epidemic models.

7.4.1 Fixed transmission pattern

When the transmission probabilities are of the form $p_{ij}(t) = \alpha_{ij}p(t)$ s.t. $\int_0^{+\infty} p(t) = 1$ and $\alpha_{ij} < 1$,

$$\mathcal{L}_{ij}(s) = -\ln(1 - \alpha_{ij})\hat{p}(s), \qquad (7.11)$$

and

$$\rho(s) = \rho_{\alpha} \hat{p}(s), \tag{7.12}$$

where $\rho_{\alpha} = \rho(0) = \rho(-\frac{\ln(1-\alpha_{ij})+\ln(1-\alpha_{ji})}{2})$ is the Hazard radius defined in the previous chapter (see Definition 6.2). In these networks, the temporal and structural behaviors are clearly separated. While ρ_{α} summarizes the structure of the network and how connected the nodes are to one another, $\hat{p}(s)$ captures how fast the transmission probabilities are fading through time.

When $\rho_{\alpha} \geq 1$, the long-term behavior is super-critical and the bound on the critical times is given by inverting $\hat{p}(s)$

$$T^{c} \geq \liminf_{n \to +\infty} \frac{\ln n}{2\hat{p}^{-1}(1/\rho_{\alpha})},\tag{7.13}$$

where $\hat{p}^{-1}(1/\rho_{\alpha})$ exists and is unique since $\hat{p}(s)$ is decreasing from 1 to 0. In general, it is not possible to give a more explicit version of the critical time of Corollary 7.4, or of the anytime influence bound of Proposition 7.1. However, we investigate in the rest of this section specific p(t) which lead to explicit results.

7.4.2 Exponential transmission probabilities

A notable example of fixed transmission pattern is the case of exponential probabilities $p_{ij}(t) = \alpha_{ij}\lambda e^{-\lambda t}$ for $\lambda > 0$ and $\alpha_{ij} \in [0, 1[$. Influence maximization algorithms under this

specific choice of transmission functions have been for instance developed by Rodriguez and Schölkopf (2012). In such a case, we can calculate the spectral radii explicitly:

$$\rho(s) = \frac{\lambda}{s+\lambda} \rho_{\alpha},\tag{7.14}$$

where $\rho_{\alpha} = \rho(-\frac{\ln(1-\alpha_{ij})+\ln(1-\alpha_{ji})}{2})$ is again the long-term Hazard matrix. When $\rho_{\alpha} > 1$, this leads to a critical time lower bounded by

$$T^{c} \ge \liminf_{n \to +\infty} \frac{\ln n}{2\lambda(\rho_{\alpha} - 1)}.$$
(7.15)

The influence bound of Corollary 7.1 can also be reformulated in the following way:

Corollary 7.5. Assume $\rho_{\alpha} \ge 1$, or else $\lambda T(1 - \rho_{\alpha}) < \frac{1}{2}$. Then the minimum in Eq. 7.4 is met for $s = \frac{1}{2T} + \lambda(\rho_{\alpha} - 1)$ and Corollary 7.1 rewrites:

$$\sigma_{\mathcal{I}}(T) \le n_0 + \sqrt{n_0(n-n_0)} \sqrt{2eT\lambda\rho_{\alpha}} e^{\lambda T(\rho_{\alpha}-1)}.$$
(7.16)

If $\rho_{\alpha} < 1$ and $\lambda T(1 - \rho_{\alpha}) \ge \frac{1}{2}$, the minimum in Eq. 7.4 is met for s = 0 and Corollary 7.1 rewrites:

$$\sigma_{\mathcal{I}}(T) \le n_0 + \sqrt{n_0(n-n_0)} \sqrt{\frac{\rho_\alpha}{1-\rho_\alpha}}.$$
(7.17)

Note that, in particular, the condition of Corollary 7.5 is always met in the supercritical case where $\rho_{\alpha} > 1$. Moreover, we retrieve the $O(\sqrt{n})$ behavior when $T < \frac{1}{\lambda(\rho_{\alpha}-1)}$. Concerning the behavior in T, the bound matches exactly the infinite-time bound when Tis very large in the sub-critical case. However, for sufficiently small T, we obtain a greatly improved result with a very instructive growth in $O(\sqrt{T})$.

7.4.3 SI and SIR epidemic models

Both epidemic models SI and SIR are particular cases of exponential transmission probabilities, and their respective model definitions are provided in Sec. 5.4.2. When the removing events are not observed, SIR is equivalent to CTIC, except that transmission along outgoing edges of one node are positively correlated. However, our results still hold in case of such a correlation, as shown in the following result.

Proposition 7.2. Assume the propagation follow an SIR model of transmission parameter β and removal parameter δ . Define $p_{ij}(t) = \beta \exp(-(\delta + \beta)t)$ for $(i, j) \in \mathcal{E}$, and let $\mathcal{A} = (\mathbf{1}_{\{(i,j)\in\mathcal{E}\}})_{ij}$ be the adjacency matrix of the underlying undirected network. Then, results of Proposition 7.1 and subsequent corollaries still hold with $\rho(s)$ given by:

$$\rho(s) = \rho\left(\frac{\mathcal{L}(s) + \mathcal{L}(s)^{\top}}{2}\right) = \ln\left(1 + \frac{\beta}{\delta}\right) \frac{\delta + \beta}{s + \delta + \beta}\rho(\mathcal{A})$$
(7.18)

From this proposition, the same analysis than in the independent transmission events case can be derived, and the critical time for the SIR model is

$$T^{c} \geq \liminf_{n \to +\infty} \frac{\ln n}{2(\delta + \beta)(\ln(1 + \frac{\beta}{\delta})\rho(\mathcal{A}) - 1)}.$$
(7.19)

Proposition 7.3. Consider the SIR model with transmission rate β , recovery rate δ and adjacency matrix \mathcal{A}_n . Assume $\liminf_{n \to +\infty} \ln(1 + \frac{\beta}{\delta})\rho(\mathcal{A}_n) > 1$, and the sequence $(T_n)_{n \in \mathbb{N}}$ is such that

$$\limsup_{n \to +\infty} \frac{2(\delta + \beta)(\ln(1 + \frac{\beta}{\delta})\rho(\mathcal{A}_n) - 1)T_n}{\ln n} < 1.$$
(7.20)

Then,

$$\sigma_{\mathcal{I}}(T_n) = o(n). \tag{7.21}$$

This is a direct consequence of Corollary 7.4 and $\rho^{-1}(1) = (\delta + \beta)(\ln(1 + \frac{\beta}{\delta})\rho(\mathcal{A}_n) - 1)$. The SI model is a simpler model in which individuals of the network remain infected and contagious through time (i.e. $\delta = 0$). Thus, the network is totally infected at the end of the contagion and $\lim_{n\to+\infty} \sigma_{\mathcal{I}}(T) = n$. For this reason, the previous critical time for the more general SIR model is of no use here, and a more precise analysis is required. Following the remark of Sec. 7.3.1, we can integrate p_{ij} on [0, T] instead of \mathbb{R}_+ , which leads to the following result:

Proposition 7.4. Consider the SI model with transmission rate β and adjacency matrix A_n . Assume $\liminf_{n\to+\infty} \rho(A_n) > 0$ and the sequence $(T_n)_{n\in\mathbb{N}}$ is such that

$$\limsup_{n \to +\infty} \frac{\beta T_n}{\sqrt{\frac{\ln n}{2\rho(\mathcal{A}_n)}} (1 - e^{-\sqrt{\frac{\ln n}{2\rho(\mathcal{A}_n)}}})} < 1.$$
(7.22)

Then,

$$\sigma_{\mathcal{I}}(T_n) = o(n). \tag{7.23}$$

In other words, the critical time for the SI model is lower bounded by

$$T^{c} \geq \liminf_{n \to +\infty} \frac{1}{\beta} \sqrt{\frac{\ln n}{2\rho(\mathcal{A}_{n})}} (1 - e^{-\sqrt{\frac{\ln n}{2\rho(\mathcal{A}_{n})}}}).$$
(7.24)

If $\rho(\mathcal{A}_n) = o(\ln n)$ (e.g. for sparse networks with a maximum degree in O(1)), the critical time resumes to $T_c \ge \liminf_{n \to +\infty} \frac{1}{\beta} \sqrt{\frac{\ln n}{2\rho(\mathcal{A}_n)}}$. However, when the graph is denser and $\rho(\mathcal{A}_n) / \ln n \to +\infty$, then $T_c \ge \liminf_{n \to +\infty} \frac{\ln n}{2\beta\rho(\mathcal{A}_n)}$.

7.4.4 Discrete-time Information Cascade

A final example is the discrete-time contagion (see Model 5.11) in which a node infected at time *t* makes a unique attempt to infect its neighbors at a time $t + T_0$. This defines the *Information Cascade model*, the discrete-time diffusion model studied by the first works on influence maximization by Kempe et al. (2003), Leskovec et al. (2007) and Chen et al. (2009, 2010). In this setting, $p_{ij}(t) = \alpha_{ij}\delta_{T_0}(t)$ where δ_{T_0} is the Dirac distribution centered at T_0 . The spectral radii are given by

$$\rho(s) = \rho_{\alpha} e^{-sT_0},\tag{7.25}$$

and the influence bound of Corollary 7.1 simplifies to:



Figure 7.1: Empirical maximum influence w.r.t. the spectral radius ρ_{α} defined in Sec. 7.4.2 for various network types. Simulation parameters: n = 1000, $n_0 = 1$ and $\lambda = 1$.

Corollary 7.6. Let $\rho_{\alpha} \geq 1$, or else $T \leq \frac{T_0}{2(1-\rho_{\alpha})}$. If $T < T_0$, then $\sigma_{\mathcal{I}}(T) = n_0$. Otherwise,

$$\sigma_{\mathcal{I}}(T) \le n_0 + \sqrt{n_0(n - n_0)} \sqrt{\frac{2eT}{T_0}} \rho_{\alpha}^{\frac{T}{T_0}}.$$
(7.26)

Moreover, the critical time is lower bounded by

$$T^{c} \ge \liminf_{n \to +\infty} \frac{\ln n}{2 \ln \rho_{\alpha}} T_{0}.$$
(7.27)

A notable difference from the exponential transmission probabilities is that T^c is here inversely proportional to $\ln \rho_{\alpha}$, instead of ρ_{α} in Eq. 7.4.2, which implies that, for the same long-term influence, a discrete-time contagion will explode much slower than one with a constant infection rate. This is probably due to the existence of very small infection times for contagions with exponential transmission probabilities.

7.5 Experimental results

This section provides an experimental validation of our bounds, by comparing them to the empirical influence simulated on several network types. In all our experiments, we simulate a contagion with exponential transmission probabilities (see Sec. 7.4.2) on networks of size n = 1000 and generated random networks of 5 different types (for more information on the respective random generators, see the introductory book written by Newman (2010)): Erdös-Rényi networks, preferential attachment networks, small-world networks, geometric random networks (Penrose, 2003) and totally connected networks with fixed weight $b \in [0,1]$ except for the ingoing and outgoing edges of a single node having, respectively, weight 0 and a > b. The reason for simulating on such totally connected networks is that the influence over these networks tend to match our upper bounds more closely, and plays the role of a best case scenario. More precisely, the transmission probabilities are of the form $p_{ij}(t) = \alpha e^{-t}$ for each edge $(i,j) \in \mathcal{E}$, where $\alpha \in [0,1[$ (and $\lambda = 1$ in the formulas of Sec. 7.4.2).

We first investigate the tightness of the upper bound on the maximum influence given in Proposition 7.1. Fig. 7.1 presents the empirical influence w.r.t. $\rho_{\alpha} = -\ln(1-\alpha)\rho(\mathcal{A})$ (where ${\cal A}$ is the adjacency matrix of the network) for a large set of network types, as well as the upper bound in Proposition 7.1. Each point in the figure corresponds to the maximum influence on one network. The influence was averaged over 100 cascade simulations, and the best influencer (i.e. whose influence was maximal) was found by performing an exhaustive search. Our bounds are tight for all values of $T \in \{0.1, 1, 5, 100\}$ for totally connected networks in the sub-critical regime ($\rho_{\alpha} < 1$). For the super-critical regime ($\rho_{\alpha} > 1$), the behavior in *T* is very instructive. For $T \in \{0.1, 5, 100\}$, we are tight for most network types when ρ_{α} is high. For T = 1 (the average transmission time for the $(\tau_{ij})_{(i,j) \in \mathcal{E}}$), the maximum influence varies a lot across different graphs. This follows the intuition that this is one of the times where, for a given final number of infected node, the local structure of the networks will play the largest role through precise temporal evolution of the infection. Because ρ_{α} explains quite well the final size of the infection, this discrepancy appears on our graphs at ρ_{α} fixed. While our bound does not seem tight for this particular time, the order of magnitude of the explosion time is retrieved and our bounds are close to optimal values as soon as T = 5.

In order to further validate that our bounds give meaningful insights on the critical time of explosion for super-critical graphs, Fig. 7.2 presents the empirical influence with respect to the size of the network n for different network types and values of T, with ρ_{α} fixed to $\rho_{\alpha} = 4$. In this setting, the critical time of Corollary 7.4 is given by $T^{c*} = \frac{\ln n}{2(\rho_{\alpha}-1)\lambda}$. We see that our bounds are tight for totally connected networks for all values of $T \in \{0.2, 2, 5\}$. Moreover, the accuracy of critical time estimation is proved by the drastic change of behavior around $T = T^{c*}$, with phase transitions having occurred for most network types as soon as $T = 5T^{c*}$.

7.6 Proofs

7.6.1 **Proof for the definition of critical time**

Proof of Theorem 7.1. Let $S = \{T \in \mathbb{R}_+ \mid \sigma_n(T) = o(n)\}$. *S* is an interval containing 0 since $\sigma_n(0) = 0$ and, if $T \in S$, then $\forall T' \leq T$, $\sigma_n(T') \leq \sigma_n(T)$ and $T' \in S$. Thus *S* is of the form $[0, T^c]$ or $[0, T^c]$, and let $T^c = \sup S$ (where $T^c \in \mathbb{R} \cup \{+\infty\}$). For all time sequences $(T_n)_{n \in \mathbb{N}}$ such that $\limsup_{n \to +\infty} T_n < T^c$, $\exists T < T^c$ and $n' \geq 0$ s.t., $\forall n \geq n', T_n \leq T$. Hence, by definition of T^c , $\sigma_n(T_n) \leq \sigma_n(T) = o(n)$. Conversely, if $\liminf_{n \to +\infty} T_n > T^c$, then $\exists T > T^c$ and $n' \geq 0$ s.t., $\forall n \geq n', T_n \geq T$. Hence $\limsup_n \sigma_n(T) \geq \limsup_n \sigma_n(T)$



Figure 7.2: Empirical maximum influence w.r.t. the network size for various network types. Simulation parameters: $n_0 = 1$, $\lambda = 1$ and $\rho_{\alpha} = 4$. In such a setting, $T^{c*} = \frac{\ln n}{2(\rho_{\alpha} - 1)\lambda}$. Note the sub-linear (a) versus linear behavior (b and c).

> 0. Now let $T^{c'}$ verify the two constraints of Theorem 7.1. The first constraint implies that $\forall T < T^{c'}$, $T \in S$ and $T \leq T^c$, which leads to $T^{c'} \leq T^c$. Moreover, $\forall T > T^{c'}$, $\limsup_n \sigma_n(T) > 0$, hence $T \notin S$ and $T \geq T^c$ using the second constraint. As a result, $T^{c'} = T^c$ and the critical time is unique.

7.6.2 **Proofs of the upper bounds on influence**

Let $X(t) \in \{0,1\}^n$ be the state vector of a continuous-time information cascade at time *t*. Proposition 5.2 implies the following equation: $\forall t > 0$ and $i \notin \mathcal{I}$,

$$X_{i}(t) = 1 - \prod_{j \in \mathcal{V}} \left(1 - \mathbb{1}\{\tau_{j} + \tau_{ji} < t\} \right),$$
(7.28)

where $\tau_i \in \mathbb{R}_+ \cup \{+\infty\}$ is the infection time of node $i, T_{ij} \in \mathbb{R}_+ \cup \{+\infty\}$ is the transmission time from node i to node j, and \mathcal{I} is a set of influencers. We now develop the proofs for Proposition 7.1 and Corollary 7.1, which rely on upper bounding the Laplace transform of $\sigma_{\mathcal{I}}(T)$.

Lemma 7.1. Define $\rho(s) = \rho(\frac{\mathcal{L}(s) + \mathcal{L}(s)^{\top}}{2})$. Then, for any \mathcal{I} such that $|\mathcal{I}| = n_0 < n$, denoting by $\hat{\sigma}_{\mathcal{I}}(s) = \int_0^{+\infty} \sigma_{\mathcal{I}}(t) e^{-st} dt$ the Laplace transform of the expected number of nodes reached by the cascade starting from \mathcal{I} at time T:

$$s\hat{\sigma}_{\mathcal{I}}(s) \le n_0 + \gamma_1\left(\rho(s), \frac{n_0}{n - n_0}\right)(n - n_0),\tag{7.29}$$

where $\gamma_1(\rho, a)$ is defined as in Theorem 6.1.

This result requires two intermediate lemmas: Lemma 7.2, that proves for $i \in \mathcal{V}$ and t > 0 a positive correlation between the events 'node j did not infect node i before time t' and Lemma 7.4, that bounds the probability that a given node gets infected before t.

Lemma 7.2. $\forall i \notin \mathcal{I} \text{ and } t > 0, \{1 - \mathbb{1}\{\tau_j + \tau_{ji} < t\}\}_{j \in \mathcal{V}} \text{ are positively correlated.}$

Proof. Denoting by $Q_{\mathcal{I},i}$ the collection of directed paths in *G* from the influencers \mathcal{I} to node *i*, we get the following expression for variables $(\tau_i)_{i \in \mathcal{V}}$ (Du et al., 2013a):

$$\tau_i = \min_{q \in \mathcal{Q}_{\mathcal{I},i}} \sum_{(j,l) \in q} \tau_{jl}$$
(7.30)

Therefore, for all $i \notin \mathcal{I}$ and t > 0, the functions $f_{ij}(\tau_{kl})_{(k,l)\in\mathcal{E}} = \{1 - \mathbb{1}\{\tau_j + \tau_{ji} < t\}\}_{j\in\mathcal{V}}$ are increasing with the partial order on $(\tau_{kl})_{(k,l)\in\mathcal{E}}$. We will then make use of the FKG inequality (Fortuin et al., 1971) :

Lemma 7.3. (FKG inequality) Let L be a finite distributive lattice, and μ a nonnegative function on L, such that, for any $(x, y) \in L^2$,

$$\mu(x \lor y)\mu(x \land y) \le \mu(x)\mu(y) \tag{7.31}$$

Then, for any non-decreasing function f and g on L

$$\left(\sum_{x\in L} f(x)g(x)\right)\left(\sum_{x\in L} \mu(x)\right) \ge \left(\sum_{x\in L} f(x)\mu(x)\right)\left(\sum_{x\in L} g(x)\mu(x)\right)$$
(7.32)

Due to the independence of $(\tau_{kl})_{(k,l)\in\mathcal{E}}$, the condition in Lemma 7.3 is met by their joint distribution, which is a product measure on the product space $\mathbb{R}^{\mathcal{E}}$. Lemma 7.2 is then obtained by applying Lemma 7.3 to any couple of functions $(f_{ij}, f_{ik})_{(i,j)\in\mathcal{E},(i,k)\in\mathcal{E}}$. More specifically, in our problem setting, L is the set of all $(\tau_{kl})_{(k,l)\in\mathcal{E}}$, $\mu(x) = \prod_{(k,l)\in\mathcal{E}} \mathbb{P}(\tau_{kl} = t_{kl})$ is the joint probability distribution of the τ_{kl} when $x = (t_{kl})_{(k,l)\in\mathcal{E}}$.

We then show the following lemma that reveals an implicit inequality satisfied by the infection probabilities $\mathbb{E}[X_i(t)]$.

Lemma 7.4. For all $(i,j) \in \mathcal{V}^2$, let p_{ij} be an integrable function such that $\int_0^{+\infty} p_{ij}(t)dt < 1$. For any \mathcal{I} such that $|\mathcal{I}| = n_0 < n$ and for any $i \notin \mathcal{I}$, the probability $\mathbb{E}[X_i(t)]$ that node i will be reached by the contagion originating from \mathcal{I} verifies:

$$\mathbb{E}[X_i(t)] \le 1 - \exp\bigg(-\sum_j (\mathcal{H}_{ji} * \mathbb{E}[X_j])(t)\bigg),\tag{7.33}$$

where $(f * g)(t) = \int f(s)g(t-s)ds$ stands for the convolution of f with g and $\mathcal{H}_{ji}(t) = \frac{-\ln(1-\int_{0}^{+\infty}p_{ji}(s)ds)}{\int_{0}^{+\infty}p_{ji}(s)ds}p_{ji}(t)$.

Proof. Eq. 7.28 and the positive correlation of $\{1 - \mathbb{1}\{\tau_j + \tau_{ji} < t\}\}_{j \in \{1,...,N\}}$ (Lemma 7.2) imply that

$$\mathbb{E}[X_i(t)] = 1 - \mathbb{E}[\prod_j (1 - \mathbb{1}\{\tau_j + \tau_{ji} < t\})] \le 1 - \prod_j \mathbb{E}[1 - \mathbb{1}\{\tau_j + \tau_{ji} < t\}]$$
(7.34)

which leads to

$$\mathbb{E}[X_{i}(t)] \leq 1 - \prod_{j} \left(1 - \mathbb{E}[\mathbb{1}\{\tau_{j} + \tau_{ji} < t\}] \right) \\
= 1 - \prod_{j} \left(1 - \mathbb{E}[\mathbb{E}[X_{j}(t - \tau_{ji})|\tau_{ji}]] \right), \qquad (7.35) \\
= 1 - \prod_{j} \left(1 - \int_{0}^{+\infty} \mathbb{E}[X_{j}(s)]p_{ji}(t - s)ds \right),$$

since $\forall i, j \in \mathcal{V}$, τ_j and τ_{ji} are independent and p_{ji} is the probability density of τ_{ji} . Note that, in our setting, we consider that influencer nodes are infected at time 0, and thus are not infectious before t = 0. We then linearize the product in Eq. 7.35:

$$\mathbb{E}[X_{i}(t)] \leq 1 - \exp\left(\sum_{j} \ln(1 - \int_{0}^{+\infty} \mathbb{E}[X_{j}(s)]p_{ji}(t-s)ds)\right) \\ \leq 1 - \exp\left(\sum_{j} \frac{\ln(1 - \int_{0}^{+\infty} p_{ji}(s)ds)}{\int_{0}^{+\infty} p_{ji}(s)ds} \int_{0}^{+\infty} \mathbb{E}[X_{j}(s)]p_{ji}(t-s)ds\right)$$

$$= 1 - \exp\left(-\sum_{j} (\mathcal{H}_{ji} * \mathbb{E}[X_{j}])(t)\right),$$
(7.36)

since we have on the one hand, for any $x \in [0,1]$ and a < 1, $\ln(1-ax) \ge \ln(1-a)x$ (in Eq. 7.36, we chose $a = \int_0^{+\infty} p_{ji}(s)ds$ and $x = \frac{\int_0^{+\infty} \mathbb{E}[X_j(s)]p_{ji}(t-s)ds}{\int_0^{+\infty} p_{ji}(s)ds}$), and on the other hand $\mathcal{H}_{ji}(t) = \frac{-\ln(1-\int_0^{+\infty} p_{ji}(s)ds)}{\int_0^{+\infty} p_{ji}(s)ds} p_{ji}(t)$ by definition of \mathcal{H} . Note that $\frac{-\ln(1-\int_0^{+\infty} p_{ji}(s)ds)}{\int_0^{+\infty} p_{ji}(s)ds}$ is approximately 1 when $\int_0^{+\infty} p_{ji}(s)ds$ is close to 0.

Proof of Lemma 7.1. Let $f_i(s) = \int_0^{+\infty} \mathbb{E}[X_i(t)]se^{-st}dt$, then, using Jensen's inequality, $\forall i \notin \mathcal{I}$ and $s \ge 0$,

$$f_i(s) \le 1 - \exp\left(-\sum_j \mathcal{L}_{ji}(s)f_j(s)\right),\tag{7.37}$$

where $\mathcal{L}_{ji}(s) = \int_0^{+\infty} \mathcal{H}_{ji}(t) e^{-st} dt$ is the Laplace transform of \mathcal{H}_{ji} . Note also that $\forall i \in \mathcal{I}, f_i(s) = 1$. We are thus is the same scenario as Lemma 6.9 of the previous chapter, and the proof of Theorem 6.1 is directly applicable by substituting $f_i(s)$ to $\mathbb{E}[X_i]$ and $\mathcal{L}_{ji}(s)$ to \mathcal{H}_{ji} . This leads to the desired result since $\sum_i f_i(s) = s\hat{\sigma}_{\mathcal{I}}(s)$.

Using Lemma 7.1, we may now prove Proposition 7.1:

Proof of Proposition 7.1. $\forall s \ge 0, T \ge 0$ and $t \ge 0, e^{-st} \ge e^{-sT} \mathbf{1}_{\{t < T\}}$, and using Lemma 7.1, $s\hat{\sigma}_{\mathcal{I}}(s) = \sum_{i} \mathbb{E}[e^{-s\tau_{i}}] \ge n_{0} + (\sigma_{\mathcal{I}}(T) - n_{0})e^{-sT}$ which leads to the desired inequality. \Box

7.6.3 Proofs of the upper bounds on critical time

Proof of Corollary 7.2. Since e^{-st} is decreasing w.r.t. s, $\mathcal{L}_{ij}(s)$ is decreasing. Thus, the Perron-Frobenius theorem implies that $\rho(s)$ is decreasing. When $\rho(0) \ge 1$, $\rho^{-1}(1-\epsilon)$ exists and is uniquely defined, and using Corollary 7.1 and Proposition 7.1, $\sigma_{\mathcal{I}}(T) \le n_0 + (n-n_0)\gamma(\rho^{-1}(1-\epsilon))e^{\rho^{-1}(1-\epsilon)T} \le n_0 + \sqrt{\frac{n_0(n-n_0)}{\epsilon}}e^{\rho^{-1}(1-\epsilon)T}$. Eq. 7.6 is a direct consequence of the critical bound in Corollary 6.1.

Proof of Corollary 7.4. If $\limsup_{n \to +\infty} \frac{2\rho^{-1}(1)T_n}{\ln n} < 1$, then $\exists \alpha > 0$ and $n' \ge 0$ s.t. $\forall n \ge n'$, $\rho^{-1}(1)T_n \le \frac{(1-\alpha)\ln n}{2}$. Furthermore, $\lim_{n \to +\infty} \frac{\rho^{-1}(1-\frac{1}{\ln n})}{\rho^{-1}(1)} = 1$, thus $\exists n'' \ge n'$ s.t. $\forall n \ge n''$, $n'', \rho^{-1}(1-\frac{1}{\ln n}) \le \frac{1-\alpha/2}{1-\alpha}\rho^{-1}(1)$. Using Corollary 7.2 with $\epsilon = \frac{1}{\ln n}, \sigma_{\mathcal{I}}(T) \le 1 + \sqrt{\ln n(n-1)}e^{\rho^{-1}(1-\frac{1}{\ln n})T} \le 1 + \sqrt{\ln n n^{1-\alpha/4}} = o(n)$.

7.6.4 Proofs for the particular contagion models

Proof of Corollary 7.5. Taking $\rho(s) = \frac{\lambda}{\lambda+s}\rho_{\alpha}$, Corollary 7.1 rewrites

$$\sigma_{\mathcal{I}}(T) \le n_0 + \sqrt{n_0(n-n_0)} \min_{s \ge 0} \left(\sqrt{\frac{\lambda}{s+\lambda(1-\rho_\alpha)}} e^{sT} \right).$$
(7.38)

The function $f(s) = \sqrt{\frac{\lambda}{s+\lambda(1-\rho_{\alpha})}}e^{sT}$ admits a unique minimum in $s_{min} = \frac{1}{2T} + \lambda(\rho_{\alpha} - 1)$. The minimum for $s \ge 0$ is therefore met for $s = s_{min}$ if $\lambda T(1-\rho_{\alpha}) < \frac{1}{2}$ and s = 0 otherwise. The results follow immediately.

Proof of Proposition 7.2. In order to prove Proposition 7.2, it is sufficient to show that Lemma 7.4 still holds for the SIR model, with $p_{ij}(t) = \beta \exp(-(\delta + \beta)t)$ for $(i, j) \in \mathcal{E}$. With the notations of Proposition 5.1 and $\mathcal{Q}_{\mathcal{I},i}^G$ the collection of directed paths in *G* from the influencers \mathcal{I} to node *i*:

$$\tau_{i} = \min_{q \in \mathcal{Q}_{\mathcal{I},i}^{G}} \sum_{(j,k) \in q} T_{jk} \mathbb{1}\{T_{jk} < D_{j}\} + (+\infty) \mathbb{1}\{T_{jk} \ge D_{j}\}$$
(7.39)

Therefore $\forall i \notin \mathcal{I}$ and t > 0, the functions $f_{ij}(T,D) = \{1 - \mathbb{1}\{\tau_j + T_{ji} < t\}\mathbb{1}\{T_{ji} < D_j\}\}_{j \in \mathcal{V}}$ are increasing with respect to the partial order on $\mathbb{R}^{\mathcal{E}} \times \mathbb{R}^{\mathcal{V}}$ defined for any $X^1 = (T^1, D^1), X^2 = (T^2, D^2) \in \mathbb{R}^{\mathcal{E}} \times \mathbb{R}^{\mathcal{V}}$ by:

$$X^{1} \ge X^{2} \iff \begin{cases} T^{1}_{ij} \ge T^{2}_{ij} & \text{for any } (i,j) \in \mathcal{E} \\ D^{1}_{i} \le D^{2}_{i} & \text{for any } i \in \mathcal{V} \end{cases}$$
(7.40)

Variables $(T_{ij})_{(i,j)\in\mathcal{E}}$ and $(D_i)_{i\in\mathcal{V}}$ being independent, we can still apply FKG inequality (Lemma 7.3) and deduce the positive correlation, for any $i \notin \mathcal{I}$ and t > 0, of the random variables $\{1 - \mathbb{1}\{\tau_i + T_{ji} < t\}\mathbb{1}\{T_{ji} < D_j\}_{j\in\mathcal{V}}$. We then introduce, for any $(i,j)\in\mathcal{E}$:

$$\overline{T_{ji}} = \begin{cases} T_{ji} & \text{if } T_{ji} < D_j \\ +\infty & \text{if } T_{ji} \ge D_j \end{cases}$$
(7.41)

It is straightforward that each $\overline{T_{ji}}$ is a random variable over $\mathbb{R}_+ \cup \{+\infty\}$ with probability distribution p_{ij} , and that $\overline{T_{ji}}$ is independent of τ_j . We also have, for any $i \notin \mathcal{I}$, t > 0 and $(i,j) \in \mathcal{E}$:

$$\{1 - \mathbb{1}\{\tau_j + T_{ji} < t\}\mathbb{1}\{T_{ji} < D_j\}\} = \{1 - \mathbb{1}\{\tau_j + \overline{T_{ji}} < t\}\},$$
(7.42)

and Lemma 7.4 for the SIR case (and therefore Proposition 7.2 and its subsequent corollaries) are then proved from following the same steps than in the independent transmission events case, except replacing $(T_{ji})_{(i,j)\in\mathcal{E}}$ by $(\overline{T_{ji}})_{(i,j)\in\mathcal{E}}$.

Proof of Proposition 7.4. $\rho(s) = \frac{\beta T_n}{1 - e^{-\beta T_n}} \frac{\beta}{\beta + s} (1 - e^{-(\beta + s)T_n}) \rho(\mathcal{A}) \leq \frac{\beta^2 T_n \rho(\mathcal{A})}{(1 - e^{-\beta T_n})s}$, which implies $\rho^{-1}(1)T_n \leq \frac{(\beta T_n)^2 \rho(\mathcal{A})}{1 - e^{-\beta T_n}}$. Let $f(x) = \frac{x^2}{1 - e^{-x}}$, f is increasing and $\forall a \geq 0$, $f(x) = a \implies x \geq \sqrt{a}(1 - e^{-\sqrt{a}})$. Hence, if $\limsup_{n \to +\infty} \frac{\beta T_n}{\sqrt{\frac{\ln n}{2\rho(\mathcal{A}_n)}}(1 - e^{-\sqrt{\frac{\ln n}{2\rho(\mathcal{A}_n)}})}} < 1$, then $\exists \alpha > 0$

s.t. $\beta T_n \leq (1-\alpha)\sqrt{\frac{\ln n}{2\rho(\mathcal{A}_n)}}(1-e^{-\sqrt{\frac{\ln n}{2\rho(\mathcal{A}_n)}}})$, and the concavity of $1-e^{-x}$ implies that $\beta T_n \leq \sqrt{\frac{(1-\alpha)\ln n}{2\rho(\mathcal{A}_n)}}(1-e^{-\sqrt{\frac{(1-\alpha)\ln n}{2\rho(\mathcal{A}_n)}}})$. Finally, $f(\beta T_n) \leq \frac{(1-\alpha)\ln n}{2\rho(\mathcal{A}_n)}$ and $\frac{2\rho^{-1}(1)T_n}{\ln n} \leq 1-\alpha$. Applying Corollary 7.4 proves the desired result.

Proof of Corollary 7.6. Taking $\rho(s) = \rho_{\alpha} e^{-sT_0}$, Corollary 7.1 rewrites

$$\sigma_{\mathcal{I}}(T) \le n_0 + \sqrt{n_0(n-n_0)} \min_{s \ge 0} \left(\sqrt{\frac{\rho_{\alpha} e^{-sT_0}}{1-\rho_{\alpha} e^{-sT_0}}} e^{sT} \right).$$
(7.43)

and
$$s = \frac{1}{T_0} \left(\ln \rho_{\alpha} - \ln(1 - \frac{T_0}{2T}) \right)$$
 gives

$$\sigma_{\mathcal{I}}(T) \le n_0 + \sqrt{n_0(n - n_0)} \sqrt{\frac{2T}{T_0} - 1} \left(\frac{\rho_{\alpha}}{1 - \frac{T_0}{2T}} \right)^{\frac{T}{T_0}}.$$
(7.44)

The final result follows by upper bounding $\left(1 - \frac{T_0}{2T}\right)^{\frac{1}{2} - \frac{T}{T_0}}$ by \sqrt{e} due to the monotonic increase of $x \to (x-1)\ln(1-\frac{1}{x})$ on $[1, +\infty[$ and its limit when $x \to +\infty$.

Part II

Control of diffusion processes

B Dynamic control of SIS epidemics

"The work of epidemiology is related to unanswered questions, but also to unquestioned answers."

 $-\operatorname{Patricia}\,\operatorname{Buffler}$

Contents

8.1	Introdu	uction
	8.1.1	Related work
	8.1.2	Outline
8.2	Epiden	nic and control model
	8.2.1	Notations
	8.2.2	Epidemic model
	8.2.3	Control model
8.3	Quality	metrics for the control action
	8.3.1	Extinction time
	8.3.2	Cumulative costs
	8.3.3	Area Under the Curve (AUC) 107
	8.3.4	Stable infection state 108
8.4	DRA C	Control Strategies
	8.4.1	Intractability of optimal strategies 108
	8.4.2	Score-based strategies 108
	8.4.3	LRIE – A greedy approach to DRA 110
	8.4.4	MCM – an optimal priority planning strategy
8.5	Experi	mental results
	8.5.1	Quality assessment for DRA strategies
	8.5.2	Competing strategies
	8.5.3	Experiments on simulated networks 114
	8.5.4	Simulations on real-world networks 117
8.6	Proofs	

8.1 Introduction

This chapter investigates the control of a diffusion process by utilizing real-time information. More specifically, we allow the network administrator to adjust the allocation of control resources, a set of treatments that increase the recovery rate of infected nodes, according to the evolution of the diffusion process.

8.1.1 Related work

The control of diffusion processes has been studied in various fields in the past, including epidemiology and computer networks resilience. The respective literature can generally be divided in three complementary lines of research, the third of which is the line where our work lays:

- a) Static vaccination strategies. Most of the epidemic literature focuses on static control actions such as permanently removing a set of edges or nodes of the network (Cohen et al., 2003; Tong et al., 2012; Wang et al., 2003; Schneider et al., 2011; Preciado et al., 2013a). In this case, the available budget is considered fixed, and the effect of a control action *permanent* (Preciado et al., 2013b; Chung et al., 2009).
- b) Budget optimization. Complementary to resource allocation, the determination of the budget size to be spent at each time step, which aims to fulfill cost and efficiency constraints, is critical for the resulting strategy. Several such studies assume that the network administrator is capable of storing resources for later use (Klepac et al., 2012; Forster and Gilligan, 2007; Khouzani et al., 2011). Also, a usual simplifying assumption is the *uniform mixing*, i.e. the infected nodes are uniformly scattered in the network. Therefore, these studies do not address the problem of how exactly to allocate the resources on the nodes of the network, but rather estimate the budget size that can cause a desired macroscopic result.
- c) *Dynamic resource allocation.* A few studies consider dynamic strategies for allocating resources for dealing with epidemics. One of the most well-known such strategy is *contact-tracing* (Borgs et al., 2010) that consists in healing the neighbors of infected nodes. In practice, this approach was shown inefficient in containing an epidemic, especially when it is beyond a very initial state. In the definition of efficient strategies, among many graph features, the role of the *cutwidth* has already been underlined by Drakopoulos et al. (2014a,b) and Scaman et al. (2014b) independently. In Chap. 9, we further explore the power of this particular concept to analyze and control epidemics in a large variety of contexts from the viewpoint of the diffusion and the resource allocation process.

8.1.2 Outline

This chapter is organized as follows: Sec. 8.2 presents the model of controlled epidemic used in this chapter, while Sec. 8.3 describes several quality metrics for control actions aiming at reducing the spread of an epidemic. In Sec. 8.4, we define a new class of DRA strategies based on scoring the nodes of a network, and then propose the *Largest Reduction in Infectious Edges* (LRIE) control strategy which is based on a greedy minimization of the cost associated to the undesired diffusion, and has the benefits of being efficient and easy

to implement. Finally, Sec. 8.5 provides a comparison between several DRA strategies, and show that the LRIE strategy substantially outperforms its competitors in a wide range of scenarios. This is an extended version of the paper (Scaman et al., 2015a) in collaboration with Argyris Kalogeratos and Nicolas Vayatis.

8.2 Epidemic and control model

We consider the standard Susceptible-Infected-Susceptible (SIS) model (see Model 5.10), and model the control action using *local* and *expensive* treatments. These treatments will be distributed in the network in order to reduce the epidemic under predefined cost constraints. Among other application examples, controlling epidemics using antidotes, limiting rumors via targeted action, or allocating resources geographically to fight against a societal problem, seem valid scenarios for such a diffusion and control model. We now introduce the setup formally and recall the SIS model.

8.2.1 Notations

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a network of $N = |\mathcal{V}|$ nodes with adjacency matrix A, where $A_{ij} = 1$ if $i \neq j$ and edge $(i, j) \in \mathcal{E}$, else $A_{ij} = 0$. Let also **0** and **1** be vectors of size N that are, respectively, all-zeros and all-ones, and $\mathbb{1}_{\{\cdot\}}$ be the indicator function.

8.2.2 Epidemic model

A state vector $X(t) \in \mathbb{R}^N$ represents the state of the diffusion process via the nodes' infection states: for each node $i \in \{1, ..., N\}$, $X_i(t) = 1$ if node i is infected at time t, else $X_i(t) = 0$. We assume no *incubation period*, therefore, a node becomes contagious upon infection. Let the control action be represented as a *resource allocation vector* $\rho(t)$, where $\rho_i(t) > 0$ iff a resource is being given to node i at time t. In such a case, we say that node i is being *healed* by the resource. Following the formalism of Ganesh et al. (2005), we model an epidemic under a control action as a *stochastic process*:

Model 8.1 (SIS epidemic under control action). Consider $\rho(t)$ a vector-valued stochastic process on \mathbb{R}^N_+ . The state vector X(t) of an epidemic under the control action $\rho(t)$ is a stochastic process on $\{0,1\}^N$ evolving with the following transition rates:

$$X_{i}(t): 0 \to 1 \text{ at rate } \beta \sum_{j} A_{ji} X_{j}(t);$$

$$X_{i}(t): 1 \to 0 \text{ at rate } \delta + \rho_{i}(t),$$
(8.1)

where β is the transmission rate over an edge, and δ is the self-recovery rate, both being essential characteristics of the infection.

Dimensionless parameters. We define two dimensionless parameters: $r = \frac{\beta}{\delta}$ the effective spreading rate of the DP, and $e = \frac{\rho}{\delta}$ the treatment efficiency.

8.2.3 Control model

We consider *dynamic resource allocation* (DRA) strategies that take as input the network G, the SIS diffusion characteristics β and δ , the current and past network states $X(t' \le t)$,

and return the distribution of the treatment resources $\rho(t)$ (Eq. 8.1). In other words, a DRA strategy uses the information available up to time *t* in order to determine a control action $\rho(t)$. This constraint is made formal in the following definition.

Definition 8.1 (DRA strategy). A dynamic resource allocation strategy $\rho(t)$ is a stochastic process that is adapted to the natural filtration associated to X(t).

Note that, if $\rho(t)$ only depends on the current state of the network X(t), then X(t) is a Markov process. We will see in Sec. 9.2.1 that *priority planning* is a class of such strategies, that allow us to limit the analysis to the simpler case of continuous-time Markov processes. In addition to this constraint, we also introduce a *resource budget limitation* on the amount of resource available for distribution at each time instance, and a *resource accumulation limitation* regarding the amount of resource that can be allocated on a single node.

Definition 8.2 (Control action under resource budget). Let $r : \mathbb{R}_+ \to \mathbb{R}_+$ be a timedependent resource budget such that r(t) accounts for the available resources budget at time t, and $\rho^* \ge 0$ be a fixed resource threshold limiting the amount of resource that can be allocated on a single node. A control action $\rho(t)$ under limited budget r(t) and resource threshold ρ^* is a DRA strategy such that: $\forall t \ge 0$,

$$\|\rho(t)\|_1 = \sum_{i \in V} \rho_i(t) \le r(t) \text{ and } \|\rho(t)\|_{\infty} = \max_{i \in V} \rho_i(t) \le \rho^*.$$

Example 8.1 (Timewise constant budget). Set r > 0. Here we assume that: r(t) = r, $\forall t$.

Example 8.2 (Finite budget). This case corresponds to a resource budget profile which is a positive Lebesgue integrable function s.t. $\int_0^\infty r(t) dt < \infty$.

Example 8.3 (Fixed number of treatments). With the additional constraints that

$$||\rho(t)||_0 = \sum_i \mathbb{1}_{\{\rho_i(t)>0\}} \le q(t)$$

and that $\forall i, \rho_i(t) \in \{0, r_t\}$, we may consider a setting in which a number q(t) of treatments of resource efficiency r_t are administrated to infected nodes at time t. This control action corresponds to the resource budgets $r(t) = q(t)r_t$ and $\rho^* = r_t$.

Assumption 8.1 (Constant resource budget). Throughout the rest of this chapter, we focus on the constant resource budget setting r(t) = r (Example 8.1). However, the strategies developed hereafter can also be applied to the case of lower bounded varying resource budget $r(t) \ge r$ by simply considering r instead of r(t).

8.3 Quality metrics for the control action

The purpose of a DRA strategy is to minimize a diffusion process using the available resource budget and current information on the epidemic. This minimization is usually performed using a cost function (either due to infected people needing extra medical treatments, the cost of the control action, or the negative impact of the epidemic on a population). Hence, an efficient DRA strategy is a strategy that is able to lower the desired

106

cost function *significantly* and *consistently* (i.e. with high probability). Several quality metrics exist in the literature, which usually capture the characteristics of the influence, i.e. the number of infected nodes during the epidemic. We here mention the major quality metrics and their precise definition.

8.3.1 Extinction time

The extinction time, i.e. the time needed to suppress the epidemic, describes the speed of convergence to a healthy network (see Ganesh et al. (2005)).

Definition 8.3 (Extinction time). For a diffusion process X(t) under control action with initial state X(0) = x, the extinction time of the diffusion process is the random quantity defined as:

$$\tau_x = \min\{t \in \mathbb{R}_+ | X(0) = x, X(t) = 0\}.$$

The extinction time depends on the chosen DRA strategy, and the main quality measure that we will consider for priority planning (see Chap. 9) is its expectation $\mathbb{E}[\tau_x]$. Note that this expectation is never infinite, and may present sub-critical (respectively supercritical) behavior (Ganesh et al., 2005), in the sense that $\mathbb{E}[\tau_x]$ may be upper bounded by a polynomial function (respectively lower bounded by an exponential function) of the network size N. In the sub-critical regime, we say that the DRA strategy removes the epidemic in *reasonable time*, and that, in the super-critical regime, the DRA strategy is not sufficiently efficient to remove the epidemic in *reasonable time*.

8.3.2 Cumulative costs

Cumulative quality metrics represent the overall cost of an epidemic by penalizing high values of the influence through time (see Forster and Gilligan (2007)):

$$\mathcal{C}_{\gamma} = \int_{t=0}^{+\infty} e^{-\gamma t} \sum_{i} X_{i}(t) dt, \qquad (8.2)$$

where $\gamma \ge 0$ is a parameter that reduces the impact of the long-term behavior of the number of infected nodes. These quality metrics describe how many nodes were infected during the whole process, with an emphasis on short-term effects if γ is high. The expectation $\mathbb{E}[C_{\gamma}]$, which can be rewritten as the Laplace transform of the influence (see Definition 5.10), will be the main quality metric considered for the LRIE strategy (see Sec. 8.4.3). Note that, when the behavior of the epidemic is *super-critical*, these quality metrics are extremely large (exponential in the number of nodes), and thus non-informative. In such a case, one might prefer the *stable infection state* metric defined below.

8.3.3 Area Under the Curve (AUC)

This is a special case of *cumulative metrics* with $\gamma = 0$ representing the overall cost of an epidemic. However, AUC uses equal temporal weighting for short- and long-term behavior, thus gives the *total number of infected nodes* during the DP.
8.3.4 Stable infection state

For a particularly aggressive DP, the epidemic becomes *super-critical* and the number of infected nodes $\sum_i X_i(t)$ converges, for a reasonably large time period, to a stable non-zero value. Of course, this is not a true stationary value since, as explained above, $\lim_{t\to+\infty} \sum_i X_i(t) = 0$. However, in the case of a super-critical DP, this convergence happens after an exponentially long time period (Ganesh et al., 2005), while the DP reaches a non-zero stable state in reasonable time. The stable infection state is thus defined as

$$N_{\infty} = \sum_{i} X_{i}(T_{\infty}), \qquad (8.3)$$

for a sufficiently large T_{∞} value (see e.g. Cohen et al. (2003)). In practice, such a quantity is computed by waiting for the influence to reach a stationary value. This can be achieved using a simple statistical test for stationarity.

8.4 DRA Control Strategies

In this section, we present simple strategies for the dynamic resource allocation problem, a greedy optimal strategy called *LRIE*, and discuss the feasibility of optimal strategies in the context of large networks. See Chap. 9 for the more advanced control strategies called *priority planning*.

8.4.1 Intractability of optimal strategies

In theory, optimal strategies can be found using the framework of Markov decision processes (MDP) and optimal control. However, such approaches are computationally intractable due to the very large state space: 2^N states for a network of N nodes. A basic MDP approach would require to store a parameter for each element of the state space, which would be prohibitive in practice for even a network of 50 nodes. For this reason, we investigate other strategies and, specifically, a greedy heuristic for solving the DRA problem presented in Sec. 8.4.3.

8.4.2 Score-based strategies

A wide class of strategies can be described in terms of a scoring function *S* that takes as input the current infection state X(t) and returns a priority order for the nodes of the network. More specifically, we define a *strategy based on score S* as a selection of the $\lceil \frac{r}{\rho^*} \rceil$ top ranked nodes according to S(X(t)):

$$\rho_{i}(t) = \begin{cases}
\rho^{*} & \text{if } S_{i}(X(t)) > \theta(t); \\
r - \rho^{*} \lfloor \frac{r}{\rho^{*}} \rfloor & \text{if } S_{i}(X(t)) = \theta(t); \\
0 & \text{otherwise,}
\end{cases}$$
(8.4)

where $\theta(t)$ is a threshold value, set so that the distributed resources do not exceed the budget, i.e. $\sum_i \rho_i(t) = r$. Note that, while the above formulation is quite general, simple scoring functions that rank the nodes based on their local properties are not well-suited for planning coordinated actions, e.g. taking advantage of the position of other treatments

nodes of the

Algorithm 1 Applying a score-based DRA strategy **Input:** infection state vector X(t), budget size *r*, efficiency limit ρ^* , scoring function S. **Output:** the resource allocation vector $\rho(t)$. if $\sum_i X_i(t) < \lceil \frac{r}{\rho^*} \rceil$ then return X(t)end if Let $\rho(t)$ a zero *N*-dimensional vector Let $V \leftarrow \{S_i(X(t))\}_{i=1}^N$ a vector containing the node scores Sort the elements of V in *descending* order and let I the node indexes of the ranking for i = 1 to $\left\lceil \frac{r}{\rho^*} \right\rceil$ do $\rho_{I(i)}(t) \leftarrow \rho^*$ end for return $\rho(t)$ infected node healthy node a edge between

same state --- infectious edges possible to transmit the disease Figure 8.1: Example network with healthy (*white*) and infected (*red*) nodes. *Dashed* edges

denote infectious edges on which the disease might spread.

when deciding where to allocate the current one. In the rest of the text, we will refer to the strategies that use scoring functions that are independent of the current state X(t) as *static strategies*; they will form a baseline to assess the significance of real-time information for the DRA problem.

Alg. 1 presents a simple pseudocode for applying a score-based strategy. In general, its complexity is $O(E + N \log N)$ due to the sorting of N score values, where E is the number of edges and N the number of nodes of the network. However, when the scoring function depends on *local* properties of the network and DP, the computational cost can be drastically reduced by partially updating the previous calculated node ranking, since only one node can change state at a time and only the scores of its neighboring nodes need to be updated. Any type of node attribute, or measurement associated to each node individually, can potentially be used to define a scoring function *S*. Sec. 8.5.2, along with Tab. 8.1, presents simple scoring functions compared in the experimental section of this chapter.

Fig. 8.1 shows the infection state of a network. Node h is the most connected, d has the highest diffusion rate (three healthy neighbors), e and h are the least and most probable to get reinfected if they recover. Scores that would give emphasis to properties like node *centrality* or *degree*, would tend to assign the highest priority to node h, while a strategy

focusing on the most diffusive nodes would prefer to give a higher priority to node *d*.

8.4.3 LRIE – A greedy approach to DRA

Optimization problem

In order to suppress a diffusion process as quickly and efficiently as possible, we consider the minimization of the cumulative cost (see Sec. 8.3 for a discussion on the quality metrics):

$$\min_{\rho} \mathbb{E}[\mathcal{C}_{\gamma}(X_0, \rho)] = \int_{t=0}^{+\infty} e^{-\gamma t} \sigma_{X_0, \rho}(t) dt, \qquad (8.5)$$

where $\sigma_{X_0,\rho}(t) = \mathbb{E}[\sum_i X_i(t) | X(0) = X_0]$ is the influence under a valid DRA strategy ρ (see Sec. 8.2) and initial state vector X_0 , and $\gamma \ge 0$ is a parameter that reduces the impact of the long-term behavior of the influence. Since the process is Markovian, it is straightforward that such an optimal strategy also minimizes $\mathbb{E}[\mathcal{C}_{\gamma}(X,\rho)]$ for all state vectors $X \in \{0,1\}^N$. In the following, we achieve this optimization by approximating the short-term behavior of $\sigma_{X,\rho}(t)$.

The LRIE solution

Using a second-order approximation of $\sigma_{X,\rho}(t)$, we obtain the following approximation of $\mathbb{E}[\mathcal{C}_{\gamma}(X,\rho)]$:

$$\mathbb{E}[\mathcal{C}_{\gamma}(X,\rho)] = \frac{1}{\gamma} \sum_{i} X_{i} + \frac{1}{\gamma^{2}} \sigma'_{X,\rho}(0) + \frac{1}{\gamma^{3}} \sigma''_{X,\rho}(0) + O(\frac{1}{\gamma^{4}}),$$
(8.6)

The minimization of the first and second order derivatives can be achieved simultaneously, and the resulting strategy, which we name *Largest Reduction in Infectious Edges* (LRIE), selects infected nodes according to the following scoring function (Scaman et al., 2014a, 2015a):

$$S_{\text{LRIE}}(X(t)) = A(\mathbf{1} - X(t)) - A^{\top}X(t) = \left[\sum_{j} [A_{ij}(1 - X_{j}(t)) - A_{ji}X_{j}(t)]\right]_{i=1}^{N}$$
(8.7)

where $\mathbf{1} - X(t)$ is the vector indicating the healthy nodes, and $\mathbf{1}$ is the vector with ones for all coordinates.

This value can be seen as *the difference in the number of infectious edges* (i.e. edges that can transmit the disease from an infected to a healthy node) after healing a specific node. For the situation of Fig. 8.1, five infectious edges would be added if node h was healed. Respectively, one infectious edge would be added if node d was healed, while, if the infection was removed from node e, then the number of such edges would decrease by two. In essence, minimizing the number of infectious edges *reduces the scattering of the infection* across the network. Consequently, a smaller *front* is created separating the healthy region from the infected nodes, and enables better control over the DP. The pseudocode of Alg. 1 can be used for applying the LRIE strategy.

Note that, in theory, the method is also applicable to higher-order approximations of $\mathbb{E}[\mathcal{C}_{\gamma}(X,\rho)]$. However, the mathematical complexity of the derivation of even third-order derivatives makes these improvements less practical, and better left to future investigation.

8.4. DRA CONTROL STRATEGIES

Scalability of LRIE

Eq. 8.7 can be easily computed for the needs of Alg. 1 by updating the score vector of the current state. When a node *i* changes state, only this node and its neighbors need an update of their scores. Then, sorting the scores will only cost $O(d_i \log N)$, where d_i is the degree of node *i*, provided the $(N - d_i)$ other nodes are already sorted.

Technical details

We now provide the derivation of the S_{LRIE} score by computing the first and the second order derivatives of $\sigma_{X,\rho}(t)$. From the formulation of Eq. 8.1, we derive the following formulas:

$$\frac{d}{dt}\mathbb{E}[X_i(t)] = -\delta\mathbb{E}[X_i(t)] - \mathbb{E}[X_i(t)\rho_i(t)] + \beta\sum_j A_{ji}\mathbb{E}[(1 - X_i(t))X_j(t)].$$
(8.8)

And, for two different nodes $i \neq j$:

$$\frac{d}{dt}\mathbb{E}[X_i(t)X_j(t)] = -2\delta\mathbb{E}[X_i(t)X_j(t)]
-\mathbb{E}[X_i(t)X_j(t)(\rho_i(t) + \rho_j(t))]
+\beta\sum_k A_{ki}\mathbb{E}[(1 - X_i(t))X_j(t)X_k(t)]
+\beta\sum_k A_{kj}\mathbb{E}[X_i(t)(1 - X_j(t))X_k(t)].$$
(8.9)

Using Eq. 8.8, we can write the derivative of $\sigma_{X,\rho}(t)$ as:

$$\sigma'_{X,\rho}(t) = -\delta\sigma_{X,\rho}(t) - \mathbb{E}[X(t)^{\top}\rho(t)] + \beta \mathbb{E}[X(t)^{\top}A(\mathbf{1} - X(t))],$$
(8.10)

and

$$\sigma_{X,\rho}'(0) = -\delta \sum_{i} X_i - X^{\top} \rho(0) - \beta X^{\top} A(\mathbf{1} - X).$$
(8.11)

Minimizing this derivative w.r.t. $\rho(0)$ is thus equivalent to only selecting nodes which are infected. In the following, we consider that $X_i(t) = 0 \Rightarrow \rho_i(t) = 0$ and $\sum_i \rho_i(t) = \min(r, \rho^* \sum_i X_i(t))$ (i.e. $\rho(t)$ minimizes $\sigma'_{X(t),\rho}(0)$). Using Eq. 8.9 and 8.10, the second order derivative of $\sigma_{X,\rho}(t)$ can be written as:

$$\sigma_{X,\rho}^{\prime\prime}(t) = -\delta\sigma_{X,\rho}^{\prime}(t) - \frac{d}{dt}\mathbb{E}[X(t)^{\top}\rho(t)] + \beta\frac{d}{dt}\mathbb{E}[X(t)^{\top}A(\mathbf{1} - X(t))] = -\beta\mathbb{E}[\{A(\mathbf{1} - X(t)) - A^{\top}X(t)\}^{\top}\rho(t)] + \mathbb{E}[\Xi(X(t))],$$
(8.12)

where $\Xi : \{0,1\}^n \to \mathbb{R}$ is a function taking a state vector X as input and returning a value (see Sec. 8.6 for the complete derivation). We thus have:

$$\sigma_{X,\rho}(t)''(0) = -\beta \{ A(\mathbf{1} - X) - A^{\top}X \}^{\top} \rho(0) + \Xi(X),$$
(8.13)

and minimizing the second order approximation of $\mathbb{E}[\mathcal{C}_{\gamma}(X,\rho)]$ is equivalent to selecting infected nodes that maximize the following score: $S_{\text{LRIE},i} = \sum_{j} [A_{ij}(1 - X_j(t)) - A_{ji}X_j(t)].$

8.4.4 MCM – an optimal priority planning strategy

The *Maxcut minimization* strategy is a strategy based on a novel analysis of a large class of DRA strategies called *priority planning*. These strategies are similar to score-based strategies except that their score is fixed and precomputed before the epidemic. While this constraint may seem to lead to a substantial drop in efficiency, we show in Chap. 9 that these strategies achieve very good performances and are able to compete, and even outperform, more dynamic strategies that recompute the score at each instant in time. The complete analysis of priority planning strategies is too large to provide here, and we refer the reader to Chap. 9 for more details on these strategies. Also, since the whole purpose of Chap. 9 is the presentation of the MCM strategy, experimental results for this strategy will be provided in Sec. 9.5.

8.5 Experimental results

The DRA strategies were compared using simulations on various random and two realworld networks. To measure the performance of a strategy on a network, 10 to 100 simulations were performed, starting from the same fixed overall infection level of the network (%), but with different random initializations of the nodes' infection state when infection is less than 100%. In all cases we set $\delta = 1$. In order to conduct our simulations we developed a software package in Matlab that we made publicly available for research use at: http://kalogeratos.com/material/lrie-dra/.

8.5.1 Quality assessment for DRA strategies

In literature, many quality metrics are available related to diffusion processes (DP). In our experiments, we used the following metrics: time to extinction, Area Under the Curve (AUC) and stable infection state (see Sec. 8.3 for precise definitions). Regarding the simulation results, they are being illustrated using the following figure types:

- *Line plots* represent, as solid lines, the expected number of infected nodes for each strategy, and their surrounding area is the 95% confidence interval under Gaussian hypothesis¹ (e.g. Fig. 8.2).
- *Heat maps* compare two strategies for a wide range of parameter values for the DP's effective spreading rate r and treatment efficiency e (see Sec. 8.2.2). In these simulations, we consider a total infection at the initial stage. The color of each point (e.g. Fig. 8.3) depicts the ratio $R_{(r,e)}$ of an employed quality metric on the performance of two strategies, for a set of values (r, e). Here AUC is used as quality metric: $\int_{t=0}^{T_{\infty}} \sum_{i} X_{i}(t) dt$, where T_{∞} denotes a sufficiently long time period. Contrary to the general AUC definition of the previous paragraph, here we stop the integration after a relatively long simulation period T_{∞} in order to compare the quality of stable non-zero behaviors as well. In this case, the ratio is equal to the ratio of *stable infection state* metrics.

¹For N_{tests} simulations, this is $2 \frac{\sigma_{N_{tests}}}{\sqrt{N_{tests}}}$, where $\sigma_{N_{tests}}$ is the standard deviation of the measurements.

=

Strategy	Scoring function $S^i(X)$ for node i
RAND	$\sigma(X_i) + R_i$, where R_i is i.i.d. uniform in [0, 1]
MN	$\sigma(X_i) + \sum_j A_{ij}$
PRC	$\sigma(X_i) + P_i$, where P_i is the PageRank score for
	node <i>i</i>
LRSR	$\sigma(X_i) + (\lambda_1 - \lambda_1^{G \setminus i})$, where λ_1 is the largest eigen-
	value of A , and $\lambda_1^{G\setminus i}$ the largest eigenvalue of
	the matrix $A^{G \setminus i}$ for the network without node i
MSN	$\sigma(X_i) + \sum_j A_{ij} (1 - X)_j$
LIN	$\sigma(X_i) - \sum_j A_{ji} X_j$
LRIE	$\sigma(X_i) + \sum_j [A_{ij}(1 - X)_j - A_{ji}X_j]$, sums MSN and LIN

Table 8.1: Various derived DRA scoring functions. In all strategies, $\sigma(1) = 0$ and $\sigma(0) = -\infty$. Also, recall that X(t) is the infection state vector and $(\mathbf{1} - X(t))$ is the vector indicating the healthy nodes (see Sec. 8.4.3).

Technical details. In practice, the duration of the simulation plays a fundamental role in the quality of our results due to the integral definition of AUC. If the behavior is convergent to 0, then the simulation should run until convergence to the absorbent state. Otherwise, $\sum_i X_i(t)$ is stationary for relatively long time, and the simulation can be terminated when stable behavior is reached. We used a statistical test to assess, with a 99% confidence, the non-zero value of the slope of $\sum_i X_i(t)$ under a linear regression assumption. When the slope is sufficiently small, we consider $\sum_i X_i(t)$ to be stationary and we terminate the simulation. In such a case, the AUC is approximately $T_{\infty}N_{\infty}$, where N_{∞} is the stable value of the number of infected nodes.

8.5.2 Competing strategies

The proposed *Largest Reduction in Infectious Edges* (LRIE) method is compared to several other heuristic scoring functions:

- *Random* (RAND): selects nodes uniformly at random, without replacement, among infected nodes.
- *Most Neighbors* (MN): selects the infected nodes with the largest number of neighbors.
- *PageRank Centrality* (PRC): selects the most central infected nodes according to PageRank algorithm (Newman, 2010).
- *Largest Reduction in Spectral Radius* (LRSR): selects the infected nodes which lead to the largest drop in the first eigenvalue of the adjacency matrix of the network.
- *Most Susceptible Neighbors* (MSN): selects the infected nodes with the most non-infected neighbors.
- *Least Infected Neighbors* (LIN): selects the infected nodes with the lowest number of infected neighbors.

Tab. 8.1 provides the expressions to compute the scoring function that are experimentally compared. MN, PRC, and LRSR come from the static vaccination literature, and we will later refer to them as *centrality-based strategies* since they focus on nodes that are central in the network topology.

MSN and LIN are intuitive heuristics based on the assumptions that a node with many susceptible neighbors will spread the virus quickly, while a node with many infected neighbors will get infected with high probability. Notably, MSN and LIN are complementary to each other. Indirectly, the former focuses on *`central'* nodes with large degree, while the latter tends to target nodes at the network *`periphery'*. Thus, MSN and LIN capture different aspects of how critical a node is for the diffusion. The proposed LRIE strategy can also be seen as a combination of MSN and LIN, since it essentially seeks for nodes which are both *diffusive* for many healthy neighbors and, at the same time, *safe* in a mildly infected neighborhood.

8.5.3 Experiments on simulated networks

We used two types of random networks: i) *Erdös-Rényi networks* (see Model 5.1, parameter: the edge probability *p*), ii) *scale-free networks* generated by the Barabási-Albert *preferential attachment* approach (see Model 5.2, parameter: the number of added edges with each node *m*). We generated a different network for each simulation using the same values for model parameters).

Erdös-Rényi random networks

Fig. 8.2 presents simulation results on Erdös-Rényi networks with the same parametrization of the generator, while using different parameter values for r, e, b_{tot} , and two different cases of initial infected population. In all simulations LRIE performs better than the competing strategies. We observe two different behaviors depending on the percentage of initially infected population. If this is low (30% in Fig. 8.2(a)), then centrality-based strategies (MN, PRC, LRSR) perform well and eliminate the DP. However, when this percentage is large (100% in Fig. 8.2(b)) and the budget b_{tot} is low, only LRIE is able to eliminate the DP. More importantly, MN, PRC, and LRSR seem counter-effective, as they present worse results than the random strategy. The reason, in this case, is that central nodes have many infected neighbors which makes them prone to quick reinfection. Fig. 8.2(c) presents a scenario with only moderately effective treatments (e=5). Note that Fig. 8.2 presents worse case scenarios for centrality-based strategies, whereas they do not provide insights whether those situations are usual or rare extreme cases.

In order to allow for more general observations, we present heat maps that compare LRIE to LRSR (the best among the competitors of LRIE) for a range of parameter values. Note that all the presented heat maps consider totally infected networks at the initial stage, and display three characteristic regions: i) the lower-left region in which both strategies are not able to eliminate the DP, ii) the middle blue region where LRIE eliminates the DP but its competitor fails to do the same, and iii) the upper-right region in which both strategies converge to $\sum_i X_i(t) = 0$.

Fig. 8.3(a) considers a range of realistic parameter values (a low treatment efficiency, and moderate budget of resources). We can see that LRIE is always more efficient than LRSR (i.e. ratio < 1) and, in a large region of the space of parameter values, LRIE increases



Figure 8.2: Results for Erdös-Rényi networks: $N=10^4$ nodes, p=0.001.

the relative quality of DP control by 10% or more. Also, there is a thin line where the ratio decreases to 0 indicating that LRIE eliminates the DP while LRSR fails to do so.

Finally, Fig. 8.3(b) compares LRIE to LRSR for a larger range of parameter values. In this setting, the efficiency of LRIE increases and the blue region becomes larger than that in Fig. 8.3(a). The fact that such a region is approximately of the form $a\beta \le \rho \le b\beta$, with $a, b \in \mathbb{R}$, is characteristic of the DRA problem and indicates that, similarly to the epidemic threshold in δ/β in the absence of control (Wang et al., 2003), the diffusion process also displays a sudden switch in its behavior when ρ/β reaches a certain threshold, depending on the strategy.

Scale-free random networks

Scale-free networks are extremely prone to epidemics due to the existence of highly connected nodes. The behavior of the compared strategies are similar to the Erdös-Rényi case (see Fig. 8.4 and Fig. 8.5), except that the DP is more aggressive. In Fig. 8.4(a), some of the



Figure 8.3: Efficiency of LRIE compared to LRSR for an Erdös-Rényi network: N = 1000, p = 0.01. Small and large ranges of values are used for $r = \beta/\delta$, $e = \rho/\delta$.



Figure 8.4: Results for random scale-free networks: $N = 10^4$ nodes, m = 5, r = 2, and $b_{tot} = 10$ resources.

strategies do not manage to converge, despite initiated with a low percentage of infected nodes. As expected, MN is more efficient in this case compared to PRC and LRSR since node degree is more significant attribute in a scale-free network than in a network with uniform random connections. Heat maps in Fig. 8.5 present similar characteristics to those of Fig. 8.3 for Erdös-Rényi networks, except there is a slightly improved performance for LRIE relatively to LRSR.



Figure 8.5: Efficiency of LRIE compared to LRSR for a scale-free network: N = 1000, m = 5. Small and large ranges of values are used for $r = \beta/\delta$ and $e = \rho/\delta$.

8.5.4 Simulations on real-world networks

Specifically:

- *The US air traffic* for the year 2010, containing 2,939 nodes and 30,501 edges. The nodes correspond to the US airports that serviced domestic and international flights, and those non-US airports that serviced flights to US, during that year².
- *A Twitter subgraph* extracted from 1,000 ego-networks of the social network by McAuley and Leskovec (2012). The resulting network contains 81,306 nodes and 1,342,303 edges.

Tab. 8.2 summarizes the simulation results on these networks, where three scenarios were considered:

- *High treatment efficiency:* For the most efficient strategies, the DP reaches zero in reasonable time. In this case, AUC and extinction time are good quality metrics for the comparison of strategies.
- *Moderate treatment efficiency:* Only LRIE is able to eliminate the DP, thus showing the substantial improvement of the method over its competitors.
- *Low treatment efficiency:* In this case, the considered strategies suppress the epidemic but none of them eliminates it. However, LRIE still achieves a far lower stable state infection than its competitors.

In all three regions, LRIE seems robust and substantially outperforms its competitors. Note that for low treatment efficiency, centrality-based strategies become counter-effective

²Source of data: OpenFlights, http://Openflights.org.

Available at T. Opsahl's web post: http://wp.me/poFcY-Vw.

Network	DP scenario				Stratogy	AUC	Τ	$\nabla X_{i}(T)$
	δ	r	е	b_{tot}	Strategy	AOC_{\downarrow}	¹ ext↓	$\sum_{i} X_{i}(1) \downarrow$
Twitter	1	0.2	300	100	RAND	∞	∞	30.6%
subgraph					MN	∞	∞	33.4%
					LRSR	246,476	7.70	0%
					MSN	89,671	2.52	0%
					LRIE	64,425	2.07	0%
	1	0.2	200	100	RAND	∞	∞	37.3%
					MN	∞	∞	42.3%
					LRSR	161,195	5.11	43.2%
					LRIE	87,600	3.03	0%
	1	0.2	50	100	RAND	∞	∞	46.4%
					MN	∞	∞	48.5%
					LRSR	∞	∞	48.9%
					MSN	∞	∞	44.4%
					LRIE	∞	∞	29.2%
US air	1	2	210	50	RAND	∞	∞	26.1%
traffic					MN	∞	∞	73.8%
					LRSR	3,723	1.81	0%
					MSN	3,235	1.65	0%
					LRIE	493	0.43	0%
	1	2	150	50	RAND	∞	∞	38.9%
					MN	∞	∞	76.6%
					LRSR	∞	∞	76.5%
					MSN	∞	∞	76.4%
					LRIE	863	1.08	0%
	1	2	100	50	RAND	∞	∞	49.7%
					MN	∞	∞	79.0%
					LRSR	∞	∞	79.2%
					MSN	∞	∞	77.4%
					LRIE	∞	∞	23.1%

Table 8.2: Results of the simulations on two real networks. An infinite value for AUC and extinction time T_{ext} means that the number of infected nodes reached a non-zero stable infection state $\sum_i X_i(T)$ in our simulations. T = 16 for the Twitter subgraph, and T = 2 for the US air traffic dataset.

with even higher stable infection level than that of RAND. Intuitively, this result is due to the fact that at the beginning of the DP the whole network is infected. Although central nodes have many infected neighbors and are prone to fast reinfections, these strategies will keep their focus on these highly connected nodes and will hence fail to clear the central part of the network, which results in a high stable infection level. Contrary, LRIE indirectly tends to contain the infected nodes in *clusters* and reduces infection's scattering. Even with low treatment efficiency, LRIE will first focus on the periphery and gradually contain the DP to the central part of the network, achieving a significantly lower stable infection percentage.

8.6 **Proofs**

The second order derivative of the number of infected nodes is computed as the sum of three derivatives:

$$\sigma_{X,\rho}^{\prime\prime}(t) = -\delta\sigma_{X,\rho}^{\prime}(t) - \frac{d}{dt}\mathbb{E}[X(t)^{\top}\rho(t)] + \beta\frac{d}{dt}\mathbb{E}[X(t)^{\top}A(\mathbf{1} - X(t))].$$
(8.14)

In the following, we show that only the third derivative $\frac{d}{dt}\mathbb{E}[X(t)^{\top}A(\mathbf{1}-X(t))]$ is not of the form $\mathbb{E}[\Xi(X(t))]$ when $\rho(t)$ already minimizes $\sigma'_{X(t),\rho}(0)$. First, Eq. 8.10 shows that $\sigma'_{X,\rho}(t)$ is of the form $\mathbb{E}[\Xi(X(t))]$ since $\sigma_{X,\rho}(t) = \mathbb{E}[\sum_{i} X_{i}(t)]$ and $X(t)^{\top}\rho(t) =$ $\min(r, \rho^* \sum_{i} X_{i}(t))$. Second, let $H(t) = \min(r, \rho^* \sum_{i} X_{i}(t))$, then $\frac{d}{dt}\mathbb{E}[X(t)^{\top}\rho(t)]$ can be computed as follows:

$$\frac{d}{dt}\mathbb{E}[X(t)^{\top}\rho(t)] = \lim_{\Delta t \to 0} \frac{\mathbb{E}[H(t+\Delta t)] - \mathbb{E}[H(t)]}{\Delta t}.$$
(8.15)

Let Δt be a sufficiently small time interval. Three scenarios are possible:

- either a node is infected during $t' \in [t, t + \Delta t]$, and $\sum_i X_i(t')$ increases by one,
- either a node heals during $t' \in [t, t + \Delta t]$, and $\sum_i X_i(t')$ decreases by one,
- or nothing happens during $[t, t + \Delta t]$ (all other scenarios are negligible when $\Delta t \rightarrow 0$).

Let $\mathbf{1}_{\{c\}} \in \mathbb{R}^{\mathbb{N}}$ be a vector with unit values at dimensions where a certain *condition c* is true, and $\mathbf{1} - X(t)$ be the vector indicating the healthy nodes of the network. We can then write:

$$\mathbb{E}[H(t+\Delta t)|X(t)] = H(t) + \Psi(\sum_{i} X_{i}(t)) \left[\delta \sum_{i} X_{i}(t) - X(t)^{\top} \rho(t)\right] \Delta t - \Psi(\sum_{i} X_{i}(t) + 1) \beta X(t)^{\top} A(\mathbf{1} - X(t)) \Delta t$$
(8.16)
+ $o(\Delta t),$

where $\Psi(x) = \min(r, \rho^* x) - \min(r, \rho^* (x - 1))$. We thus have:

$$\frac{d}{dt}\mathbb{E}[X(t)^{\top}\rho(t)] = \lim_{\Delta t \to 0} \frac{\mathbb{E}[\mathbb{E}[H(t+\Delta t)|X(t)]] - \mathbb{E}[H(t)]}{\Delta t}
= -\mathbb{E}[\Psi(\sum_{i} X_{i}(t))[\delta \sum_{i} X_{i}(t) + H(t)]]
+ \mathbb{E}[\Psi(\sum_{i} X_{i}(t) + 1)\beta X(t)^{\top}A(\mathbf{1} - X(t))],$$
(8.17)

which is of the form $\mathbb{E}[\Xi(X(t))]$.

Finally, $\frac{d}{dt}\mathbb{E}[X(t)^{\top}A(\mathbf{1}-X(t))]$ is derived using Eq. 8.8 and Eq. 8.9:

$$\frac{d}{dt}\mathbb{E}[X(t)^{\top}A(\mathbf{1}-X(t))] = \sum_{i,j}A_{ij}\frac{d}{dt}\mathbb{E}[X_{i}(t)(1-X_{j}(t))]
= \sum_{i,j}A_{ij}\left(\frac{d}{dt}\mathbb{E}[X_{i}(t)] - \frac{d}{dt}\mathbb{E}[X_{i}(t)X_{j}(t)]\right)
= -\delta\sum_{i,j}A_{ij}\mathbb{E}[X_{i}(t)]
-\sum_{i,j}A_{ij}\mathbb{E}[X_{i}(t)\rho_{i}(t)]
+\beta\sum_{i,j,k}A_{ij}A_{ki}\mathbb{E}[(1-X_{i}(t))X_{k}(t)]
+2\delta\sum_{i,j}A_{ij}\mathbb{E}[X_{i}(t)X_{j}(t)]
+\sum_{i,j}A_{ij}\mathbb{E}[X_{i}(t)X_{j}(t)(\rho_{i}(t)+\rho_{j}(t))]
-\beta\sum_{i,j,k}A_{ij}A_{ki}\mathbb{E}[(1-X_{i}(t))X_{j}(t)X_{k}(t)]
-\beta\sum_{i,j,k}A_{ij}A_{kj}\mathbb{E}[X_{i}(t)(1-X_{j}(t))X_{k}(t)].$$
(8.18)

This equation is simplified by the fact that, in order to minimize $\sigma'_{X,\rho}(0)$, resources are only given to infected nodes, which implies $X_i(t)\rho_i(t) = \rho_i(t)$. We can thus rewrite this derivative as:

$$\begin{split} \frac{d}{dt} \mathbb{E}[X(t)^{\top} A(\mathbf{1} - X(t))] &= -\sum_{i,j} A_{ij} \mathbb{E}[\rho_i(t)] + \sum_{i,j} A_{ij} \mathbb{E}[X_j(t)\rho_i(t) + X_i(t)\rho_j(t)] + \mathbb{E}[\Xi(X(t))] \\ &= -\mathbb{E}[\mathbf{1}^{\top} A^{\top} \rho(t)] + \mathbb{E}[X(t)^{\top} A^{\top} \rho(t) + X(t)^{\top} A \rho(t)] + \mathbb{E}[\Xi(X(t))] \\ &= -\mathbb{E}[\{A(\mathbf{1} - X(t)) - A^{\top} X(t)\}^{\top} \rho(t)] + \mathbb{E}[\Xi(X(t))], \end{split}$$

$$(8.19)$$

where $\Xi : \{0,1\}^n \to \mathbb{R}$ is a function taking a state vector X as input and returning a value. This leads to the second order derivative of $\sigma_{X,\rho}(t)$ given in Eq. 8.12.

A detailed analysis of priority planning

"The science of today is the technology of tomorrow." — Edward Teller

Contents

9.1	Introd	ution
	9.1.1	Related works
	9.1.2	Outline
9.2	Priorit	y planning
	9.2.1	A healing plan to gradually remove a contagion
	9.2.2	Maxcut and cutwidth
9.3	Tight	bounds on the extinction time
	9.3.1	Theoretical bounds for the extinction time
	9.3.2	Relationship between critical behavior and maxcut
	9.3.3	Interpretation of results
9.4	The M	<i>laxCut Minimization</i> strategy
	9.4.1	Maxcut optimization
9.5	Experi	imental results
	9.5.1	Setup and competitors
	9.5.2	Quality of the theoretical bound
	9.5.3	Empirical evaluation of simulated contagion on real networks 132
9.6	Robus	tness of MCM
	9.6.1	Malicious modification of the network 136
	9.6.2	Random additive noise
	9.6.3	Uncertainty in the localization of nodes in contact networks 137
9.7	Proofs	
	9.7.1	Main intermediate results and lemmas
	9.7.2	Proofs of theorems
	9.7.3	Proofs of propositions and lemmas 143

9.1 Introdution

In this chapter, we analyze, theoretically and experimentally, a particular class of Dynamic Resource Allocation (DRA) strategies called *priority planning*. These strategies are based on a priority order computed prior to the epidemic, and distribute the resources to the first infected nodes according to the precomputed order. Based on the theoretical analysis of priority planning, we also provide an optimal priority planning strategy called *Maxcut Minimization* (MCM), whose priority order minimizes the *maxcut*, a quantity shown to drive the explosive behavior of the epidemic.

9.1.1 Related works

The most related method to MCM is the recent *CURE policy* by Drakopoulos et al. (2014a) which was developed independently to our work. The differences between our work and that of Drakopoulos et al. (2014a) are summarized in the following points: i) we consider a more general setting that models important additional aspects such as node self-recovery and the allocation of multiple treatment resources, ii) we derive tighter bounds for the expected extinction time, iii) in contrast with the theoretical work by Drakopoulos et al. (2014a), we also provide a test bed for experimental assessment of healing strategies on benchmark real networks and others artificially generated, including a robustness analysis to various noise profiles on the network structure, and iv) we experimentally show that MCM is more efficient than CURE as well as other possible approaches to create a priority-order based on related literature.

9.1.2 Outline

This chapter is organized as follows: Sec. 9.2 describes the class of control actions analyzed in this chapter. Sec. 9.3 provides our derived upper and lower bounds on the extinction time of the diffusion process under a priority-order. These results both give insight into the efficiency of those strategies and motivate an efficient strategy developed in Sec. 9.4. Sec. 9.5 presents experimental results and show that: i) the derived bounds are tight, thus validating the fundamental role of the maxcut in the evaluation of such strategies, ii) the proposed MCM strategy outperforms its competitors in a wide range of scenarios that are of special interest for practical application of epidemic control policies. Finally, Sec. 9.6 provides an analysis of the behavior of the MCM strategy under various noise profiles, and show the robustness of the method with respect to uncertainties in the location of individuals in contact networks. This is an extended version of the papers (Scaman et al., 2014b) and (Kalogeratos et al., 2015) in collaboration with Argyris Kalogeratos and Nicolas Vayatis.

9.2 Priority planning

9.2.1 A healing plan to gradually remove a contagion

According to Definition 8.1, the control action $\rho(t)$ of a dynamic resource allocation (DRA) strategy depends on the history of variations of the contagion process X(t). Among the novelties of our work, we introduce strategies that involve an ordering on the network nodes that accounts for the criticality of each node w.r.t. the overall contagion process. We

name the class of DRA strategies that are based on a priority-order as *priority planning*. These concepts are formally defined below.

Definition 9.1 (Priority-order). A priority-order is a bijective mapping $\ell: \mathcal{V} \to \{1, ..., N\}$ of the N nodes of the network s.t. $\ell(v)$ is the position of node v in the priority-order.

Definition 9.2 (Priority planning). Priority planning is a DRA strategy under limited budget r and resource threshold ρ^* that distributes resources to the top-q infected nodes according to a fixed priority-order ℓ of the network nodes, where q is the number of nodes such that the allocated amount of resources matches the available resource budget r. More specifically, the strategy heals the first $q(t) = \min\left\{ \lceil \frac{r}{\rho^*} \rceil, \sum_i X_i(t) \right\}$ infected nodes according to the respective mapping ℓ , and allocates the resource budget as follows:

$$\rho_{i}(t) = \begin{cases} \frac{r}{\lceil \frac{r}{\rho^{*}} \rceil} & \text{if } X_{i}(t) = 1 \text{ and } \ell(v_{i}) \leq \theta(t); \\ 0 & \text{otherwise,} \end{cases}$$

$$(9.1)$$

where $\theta(t)$ is a threshold adjusted s.t. $\sum_{i} \mathbf{1}_{\{\rho_i(t)>0\}} = q(t)$.

This definition may be regarded as a description of a class of simple planning strategies for the removal of a contagion: a *healing plan*, i.e. a priority-order for healing the nodes, is determined prior to the beginning of the diffusion and is followed no matter how the diffusion process evolves. The plan proceeds from the first to the last node in the priorityorder, hence aims to remove gradually the contagion from the network. In what follows, we refer interchangeably to a priority-order and its corresponding mapping ℓ .

Remark 9.1. Note that changes in the distribution of the resources will only appear when there is a change in the network state (i.e. a new node infection/recovery). Thus, in practice, the distribution of the resources needs to be updated only at those specific times.

9.2.2 Maxcut and cutwidth

The concept of *cutwidth* is well known in graph theory (see for instance Chung and Seymour (1989)). The importance of this concept for containing epidemics on graphs has been pointed out very recently by Drakopoulos et al. (2014a) and Scaman et al. (2014b). We recall here its definition that we adapt with our notations of the priority-order.

Definition 9.3 (Cut of a set of nodes). For an undirected network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with adjacency matrix A, the cut of a set of nodes $\mathcal{I} \subset \mathcal{V}$ is the number of edges between nodes of \mathcal{I} and nodes of its complementary in \mathcal{V} :

$$C(\mathcal{I}) = \sum_{i,j} A_{ij} \mathbb{1}_{\{v_i \in \mathcal{I}, v_j \notin \mathcal{I}\}}.$$
(9.2)

For a priority-order ℓ , we refer to as *cut at position c*, $C_c(\ell)$, the cut of the set of nodes before position *c* in ℓ :

$$C_{c}(\ell) = \sum_{i,j} A_{ij} \mathbb{1}_{\{\ell(v_{i}) < c \le \ell(v_{j})\}}.$$
(9.3)



(b) Priority-order $\ell' : \mathcal{V} \rightarrow \{1, 3, 4, 2, 5\}$

Figure 9.1: Two priority-orders (from left to right) leading to different maxcuts: $C^*(\ell) = 3$ for (a) and $C^*(\ell') = 1$ for (b). The cut (vertical red line) separates the nodes in two sets (white and red). The second priority-order ℓ' is optimal and the network has a cutwidth W = 1.

Definition 9.4 (Maxcut of a priority-order). For a network with N nodes and adjacency matrix A, and for a given priority-order ℓ , the maxcut of ℓ is defined as:

$$\mathcal{C}^*(\ell) = \max_{c=1,\dots,N} C_c(\ell). \tag{9.4}$$

Finally, the minimal value of the maxcut over all possible priority-orders is an inherent property of the network structure known as *cutwidth*.

Definition 9.5 (Cutwidth). *The cutwidth of a network with N nodes and adjacency matrix A is defined as:*

$$\mathcal{W} = \min_{\ell} \mathcal{C}^*(\ell). \tag{9.5}$$

We will see in Sec. 9.3 that this quantity characterizes networks in which priority planning is effective. Fig. 9.1 illustrates two priority-orders for a small network and their respective maxcut (see Definition 9.4). The priority-order of Fig. 9.1(b) is better, indeed optimal, and its maxcut matches the cutwidth W of the network.

9.3 Tight bounds on the extinction time

This section contains a detailed theoretical analysis of priority planning strategies. Since a priority-order is a predefined healing plan that is followed throughout the whole process of suppressing the contagion, the question of whether the contagion will be removed or not depends on the capacity of the control action to accomplish each of the steps of the plan. Intuitively, the most difficult step is at the position(s) where the maxcut of the priority-order lays, where we find the maximum number of infectious edges during the plan. Our results prove the determinant role of the maxcut of the priority-order for the expected extinction time of an epidemic. It is thus justified that this is the right quantity to minimize in order to enforce the removal of a contagion.

9.3.1 Theoretical bounds for the extinction time

Denote by **1** the vector of ones in \mathbb{R}^N . Theorem 9.1 below states an *upper bound* for the expected extinction time $\mathbb{E}[\tau_1]$ under a considered priority-order, when the budget is bounded and the starting state is a total infection $x = \mathbf{1}$. Above a threshold value, it indicates that the diffusion process converges in *reasonable time* to its absorbent state. This threshold depends on the *maxcut*, denoted as $\mathcal{C}^*(\ell)$, of a considered priority-order. Detailed proofs for the theorems are provided in Sec. 9.7.

Theorem 9.1. Let G be a totally infected network of N > 1 nodes, i.e. $X_i(0) = 1, \forall i$, and d is the maximum node degree of the network. Consider a priority planning ℓ under constant resource budget r > 0 as in Assumption 8.1. We set:

• $q = \lceil \frac{r}{\rho^*} \rceil$ the number of treated nodes, and • $\epsilon = \frac{d(3+2\ln N+4q)}{C^*(\ell)}$.

Assume that:

$$r + \delta q > \beta \mathcal{C}^*(\ell) \left(1 + 2\sqrt{\epsilon} + \epsilon \right) \tag{9.6}$$

Then the following upper bound holds for the expected extinction time $\mathbb{E}[\tau_1]$:

$$\mathbb{E}[\tau_{1}] \leq \frac{3N + 6q(1 + \ln q)}{r + \delta q - \beta \mathcal{C}^{*}(\ell) \left(1 + 2\sqrt{\epsilon} + \epsilon\right)}.$$
(9.7)

Eq. 9.7 relates $\mathbb{E}[\tau_1]$ to the number of infectious edges in the worst step of the plan (i.e. its *maxcut*). The next theorem is a simplified version of Theorem 9.1.

Theorem 9.2. Under the hypotheses of Theorem 9.1, the same definitions for ϵ and q values, and the same assumption:

$$r + \delta q > eta \mathcal{C}^*(\ell) \left(1 + 2\sqrt{\epsilon} + \epsilon
ight)$$
 ,

we have:

$$\mathbb{E}[\tau_1] \leq \frac{6N}{\beta} \,. \tag{9.8}$$

Remark 9.2. Due to the increase of $x \mapsto \mathbb{E}[\tau_x]$ (see Lemma 9.1), Theorems 9.1 and 9.2 also hold for any initial infection state $X(0) = x \in \{0,1\}^N$. Thus, when the resource budget is sufficiently high, priority-planning strategies are efficient regardless of the initial infection state.

The next theorem shows that, when there is no self-recovery ($\delta = 0$) and when the resource budget is below a threshold value, the expected extinction time is lower bounded by the exponential of the maxcut.

Theorem 9.3. Let \mathcal{G} be a totally infected network of N > 1 nodes, i.e. $X_i(0) = 1, \forall i$, and d is the maximum node degree of the network. Let also $\delta = 0$, and consider a priority planning ℓ under constant resource budget r > 0 as in Assumption 8.1. We set:

• $q = \left\lceil \frac{r}{\rho^*} \right\rceil$ the number of treated nodes, and

• $\eta \in [0, \frac{1}{2}].$

Assume that $q < \frac{\mathcal{C}^*(\ell)}{d}$ and

$$r < (1 - \eta)\beta \mathcal{C}^*(\ell)(1 - \frac{dq}{\mathcal{C}^*(\ell)})$$
(9.9)

Then the following lower bound holds for the expected extinction time $\mathbb{E}[\tau_1]$:

$$\mathbb{E}[\tau_1] \ge \frac{1}{r} \exp\left(\frac{\eta^2}{12} \left(\frac{\mathcal{C}^*(\ell)}{d} - q\right)\right).$$
(9.10)

We will see in the next section that, under further assumptions on the network type, Theorem 9.3 implies the explosive behavior of the extinction time over the resource threshold $\beta C^*(\ell)$. These results do verify our intuition that completely removing a contagion requires the resource strength to be *as high as needed in order to proceed through the worst step* of the specified plan (see further discussion in Sec. 9.3.3).

9.3.2 Relationship between critical behavior and maxcut

In order to make more apparent the relationship between critical behavior of the expected extinction time and maxcut of the priority-order used for control, we now derive theorems in the restricted setting in which the maximum degree is small w.r.t. the maxcut (see Sec. 9.3.3 for precise examples in which the assumption holds):

Assumption 9.1. Let (\mathcal{G}_N) be a sequence of networks of N nodes and $r_N > 0$ be a sequence of resource budgets. Let also $q_N = \lceil \frac{r_N}{\rho_N^*} \rceil$, d_N be the maximum node degree of \mathcal{G}_N , and assume $\liminf_N r_N > 0$. In the two following theorems, we will assume that:

$$\exists \alpha > 0 \text{ s.t. } \frac{q_N d_N}{\mathcal{C}^*(\ell_N)} = O(N^{-\alpha}) .$$
(9.11)

Under Assumption 9.1, the two next theorems show that $\beta C^*(\ell)$ acts as a threshold between a sub-critical and a super-critical regime for the resource budget r. Below this threshold, the epidemic is removed in reasonable time. Above this value, the epidemic cannot be removed by the considered priority planning.

Theorem 9.4 (Sub-critical behavior). Under the hypotheses of Assumption 9.1, if

$$\liminf_{N} \frac{r_N}{\beta \mathcal{C}^*(\ell_N)} > 1 , \qquad (9.12)$$

then

$$\mathbb{E}[\tau_1] = O(N) . \tag{9.13}$$

Theorem 9.5 (Super-critical behavior). Let $\delta = 0$ (no self-recovery). Under the hypotheses of Assumption 9.1, if

$$\limsup_{N} \frac{r_N}{\beta \mathcal{C}^*(\ell_N)} < 1 , \qquad (9.14)$$

then

$$\mathbb{E}[\tau_1] \ge \exp\left(N^{\alpha/2}\right),\tag{9.15}$$

for N sufficiently large.

126

These results show the existence of a threshold $\beta C^*(\ell)$ similar to the epidemic threshold of the epidemiology literature (Tong et al., 2012), and are fundamental for understanding the behavior of the diffusion process and designing efficient DRA strategies. The simulations in Sec. 9.5 attest that minimizing this resource threshold is an efficient way to dynamically control a diffusion process.

Remark 9.3. While these results consider a fixed priority planning, they also provide a quantitative measure of the quality of priority planning strategies as a whole. More specifically, Theorems 9.4 and 9.5 imply that, under Assumption 9.1, priority planning strategies are unable to suppress an epidemic if the resource budget is below βW (see Definition 9.5). Above this value, a priority-order can be found that achieves this specific goal.

9.3.3 Interpretation of results

Comparison to previous results

A recent work by Drakopoulos et al. (2014a) investigates the use of the cutwidth for designing efficient DRA strategies. Their work introduces a DRA strategy, called the *CURE* policy, using a formalism similar to priority-orders called *crusades*. However, the CURE policy works in a setting with no self-healing ($\delta = 0$) and no limitation on the number of resources that can be allocated to a single node ($\rho = +\infty$). As a result, all the resources are always given to one single node, and the CURE policy is only applicable when the number of treatments to distribute is equal to 1 (i.e. $q = \lceil \frac{r}{\rho^*} \rceil = 1$, see Sec. 9.2). The model considered in the present paper is thus more general and closer to realistic settings. Also, our theoretical results are tighter, as we present a threshold *four times smaller* than that of Drakopoulos et al. (2014a), matching $\beta C^*(\ell)$ in the restricted setting of Assumption 9.1. More specifically, their analysis can be seen as a particular case of ours, as Theorem 9.1 implies the following result, similar to Theorem 1 and Corollary 1b of Drakopoulos et al. (2014a):

Corollary 9.1. Let q=1, $\delta=0$, $\beta=1$ and $N \ge 20$. Consider a priority planning ℓ and assume that:

$$r \ge 4\mathcal{C}^*(\ell) \text{ and } r \ge 16d \log_2 N.$$
 (9.16)

Then the following upper bound holds for the expected extinction time $\mathbb{E}[\tau_1]$:

$$\mathbb{E}[\tau_1] \leq \frac{26N}{r} \,. \tag{9.17}$$

Particular bounds for specific graph models

We present three application examples for our theoretical results.

Example 9.1 (Sparse networks with bounded degrees). The condition $\frac{q_N d_N}{C^*(\ell_N)} = O(N^{-\alpha})$ is verified for sparse networks with bounded node degrees such that the cutwidth \mathcal{W} is a power of N (recall that $\mathcal{C}^*(\ell) \ge \mathcal{W}$, see Definition 9.5), and when the number of treatments q_N is bounded by $\ln N$, i.e. $q_N = O(\ln N)$. The condition on \mathcal{W} is verified for many standard networks (e.g. \sqrt{N} for 2D grids, $\frac{N-1}{2}$ for star networks and $\frac{N^2}{4}$ for complete graphs), with the notable exception of trees (\mathcal{W} is of the order of $\ln N$). See the work of Díaz et al. (2002) for more information on the cutwidth of specific graphs.

Example 9.2 (2D grids). For regular 2D grids of *N* nodes, $\mathcal{W} = \sqrt{N}$ and $d_N = 4$. Hence, if the number of treated nodes is bounded by $q_N = O(N^{\alpha})$ where $\alpha < \frac{1}{2}$, then Assumption 9.1 holds. For example, as in Example 8.3, fixing $\rho_N^* = \frac{r_N}{q}$ where $q \in \mathbb{N}$ is sufficient for Assumption 9.1 to hold. Sec. 9.3.2 implies that, if $r_N < \beta \sqrt{N}$, then no priority-order will manage to remove the epidemic in reasonable time. However, if $r_N > \beta \sqrt{N}$, then at least one priority-order (see Sec. 9.4) can remove the epidemic.

Example 9.3 (Complete networks). Complete networks are trivial settings for which all priority-orders lead to the same efficiency, since all nodes have the same connectivity. We have $\mathcal{W} = \frac{N^2}{4}$ (or $\frac{(N+1)(N-1)}{4}$ if N is odd) and $d_N = N - 1$, hence, if the number of treated nodes is bounded by $q_N = O(N^{\alpha})$ where $\alpha < 1$, then Assumption 9.1 holds. Sec. 9.3.2 implies that, if $r_N < \beta \frac{N^2}{4}$, then no priority-order will remove the epidemic in reasonable time. However, if $r_N > \beta \frac{N^2}{4}$, all priority-orders will.

9.4 The MaxCut Minimization strategy

Based on the analysis of the previous section that uncovered a strong dependency between the critical behavior of the diffusion process and the *maxcut*, we propose a novel DRA algorithm for arbitrary networks. The main idea is to distribute resources to infected nodes in the priority-order that minimizes $C^*(\ell)$, and optimally reaches the cutwidth W of the network. Given a network \mathcal{G} , we compute, prior to the diffusion process, a *priority-order* $\ell_{MCM}(\mathcal{G})$ with minimum maxcut $C^*(\ell)$:

$$\ell_{_{MCM}}(\mathcal{G}) = \operatorname*{argmin}_{\ell} \mathcal{C}^*(\ell), \tag{9.18}$$

using any available optimization algorithm for this problem. Then, during the diffusion, the strategy distributes the resource budget to the infected nodes according to the order $\ell_{MCM}(\mathcal{G})$. Alg. 2 presents the pseudocode of our strategy.

9.4.1 Maxcut optimization

Linear arrangement

Minimizing $C^*(\ell)$ in a network is a standard combinatorial problem which is usually solved under the framework of *linear arrangement* (LA) problems (Díaz et al., 2002; Pantrigo et al., 2012; Pardo et al., 2013). Formally, a linear arrangement is a mapping $\ell: \mathcal{V} \to \{1, ..., N\}$ of the nodes of \mathcal{G} on N discrete positions arranged on a line, by assigning one position (or label) to each node (Fig. 9.1). This is a class of combinatorial optimization problems, which are altogether usually referred to as *graph layout* problems (Díaz et al., 2002), and whose purpose is to minimize some functional ϕ over the space \mathcal{L} of all possible node permutations: $\ell^* = \operatorname{argmin}_{\ell \in \mathcal{L}} \phi(\mathcal{G}, \ell)$. Indicative applications are the graph drawing, VLSI design, and network scheduling (Díaz et al., 2002).

The *minimum maxcut linear arrangement* (MMLA) is an LA problem in which the goal is to minimize the *maxcut*. MMLA is an NP-hard problem, however, approximation heuristics do exist in literature (Pantrigo et al., 2012; Pardo et al., 2013). One of the major difficulties of this problem is that the cost function to optimize is extremely *flat* in the search space, i.e. slight changes in the arrangement will most probably not change $C^*(\ell)$.

Algorithm 2 MCM strategy

▷ Prior to the diffusion process: Compute the priority-order $\ell = \ell_{_{MCM}}(\mathcal{G})$ by minimizing the maxcut $\mathcal{C}^*(\ell)$ Order the nodes of \mathcal{G} according to ℓ , i.e. compute the node list $(v_1, ..., v_N)$ s.t. $\forall i \in \{1, ..., N\}$, $\ell(v_i) = i$ ▷ During the diffusion process: **Input:** network \mathcal{G} , state vector X(t), resource budget r, resource threshold ρ^* **Output:** the resource allocation vector $\rho(t)$ $q \leftarrow \left\lceil \frac{r}{\rho^*} \right\rceil$ if $\sum_i X_i(t) < q$ then return $\frac{r}{a}X(t)$ end if // a zero vector in \mathbb{R}^N $\rho(t) \leftarrow \mathbf{0}$ budget $\leftarrow q$ $i \leftarrow 1$ while budget > 0 do if $X_{v_i}(t) = 1$ then $\rho_{v_i}(t) \leftarrow \frac{r}{a}$ $budget \leftarrow budget - 1$ end if $i \leftarrow i + 1$ end while return $\rho(t)$

Relaxation of the MMLA problem

For the latter reasons, we propose to relax the MMLA problem by optimizing the *sum of the cuts* instead of their *maximum*. This problem is known as the *minimum linear arrangement* problem (MLA) and is part of the larger class of *minimum p-sum linear arrangement* problems (MpLA) (Harper, 1964; Adolphson and Hu, 1973) that minimize the following functional:

MpLA:
$$\phi(\mathcal{G}, \ell) = \left(\sum_{i,j} A_{ij} \left| \ell(v_i) - \ell(v_j) \right|^p \right)^{1/p}$$
. (9.19)

For p = 1, a simple calculation shows that MLA minimizes the average cut in the linear arrangement, instead of its maximum for MMLA (see Definitions 9.3 and 9.4). MLA is easier than MMLA and more suited to gradient descent or simulated annealing methods, and it produces a smoother priority-order w.r.t. the cuts at each position of the ordering.

Practical implementation

MLA is a very challenging problem and, interestingly, most related works conduct experiments on relatively small benchmark networks for which the optimal cost is not known (the same for MMLA). Designing a procedure that can be applied on large social networks with tens of thousands of nodes is by itself a remarkable contribution. We should also note that MCM strategy seeks for an priority-order with as low as possible maxcut, but not necessarily the optimal one. The solver we developed for our simulations follows the



Figure 9.2: Resource threshold w.r.t. the maxcut $C^*(\ell)$ quantity for various random network types of 1,000 nodes, and DRA strategies ($\delta = 1$). The dashed line indicator has slope equal to 1.

steps below and uses a hierarchical approach to take advantage of the group structure of social and contact networks:

- s1) first, we identify dense clusters by applying spectral clustering and we order those clusters (considered as high-level nodes) using spectral sequencing (Juvan and Mohar, 1992),
- **s2)** then, we compute a good ordering of the nodes inside each cluster independently using spectral sequencing followed by an iterative approach which is based on random node swaps (swap heuristics inspired by the work of Rodriguez-Tello et al. (2008)),
- s3) finally, the swap-based approach is reapplied to optimize the overall ordering.

Scalability

The scalability of the MCM strategy is highly dependent on the employed offline algorithm for finding the optimal node order. The whole process described above achieves fairly good results (see for example Tab. 9.1) in reasonable time. Since spectral clustering and spectral sequencing depend on the computation of eigenvectors for the highest eigenvalues of an $N \times N$ sparse matrix with $|\mathcal{E}|$ non-zero entries, the overall complexity of the algorithm is $O(|\mathcal{V}| + |\mathcal{E}|)$ (Arora et al., 2005). Hence, MCM is generally scalable to the size of real social and contact networks. Worth to note that, for networks that are close to planar and can be embedded in the 2D plane without many edge intersections (e.g. contact networks), the clustering step could be skipped since the spectral sequencing method is already a good initial approximation that can be further refined with a subsequent node swapping process.

9.5 Experimental results

9.5.1 Setup and competitors

In the experimental study, we compare MCM against five DRA strategies, which are grouped into three types:

A. Static vaccination

We compare to centrality-based strategies and simple heuristics by considering the output of state-of-the-art static vaccination methods as priority-orders:

- *Most neighbors* (MN): gives priority to high degree nodes, hence aims to first remove the contagion from the network's core before dealing with the periphery.
- *Least neighbors* (LN): gives priority to low degree nodes and works conversely to MN.
- *Largest reduction in spectral radius* (LRSR): gives priority to nodes whose removal will lead to the maximum decrease of the *spectral radius* of the adjacency matrix of the resulting network, and is a state-of-the-art method from the vaccination literature developed by Tong et al. (2012).

B. Uniform mixing

We compare to strategies that assume uniform mixing by considering a random allocation of resources.

• *Random baseline* (RAND): the resource budget is assigned to $\lceil \frac{r}{\rho^*} \rceil$ nodes at random at each time.

One of the primary questions we address in this perspective is whether targeting specific nodes in the network can lead to substantial improvement of the suppression, compared to the treatment of random infected nodes.

C. State-of-the-art direct competitors

• *CURE policy* (CURE): this is a state-of-the-art method developed by Drakopoulos et al. (2014a) in parallel to our work. CURE follows a healing plan with minimal maxcut, called *crusade*. A crusade can be considered to be a priority-order similar to that of MCM, however in the work of Drakopoulos et al. (2014a) this was formulated as a sequence of nested bags which differ by one node each time.

A brief summary of the differences between our work and the work of Drakopoulos et al. (2014a) is given in Sec. 9.1 and a technical discussion regarding the bounds of extinction time can be found in Sec. 9.3.3. Moreover, one of the most significant practical differences between CURE and MCM is that, when many reinfections occur, CURE enters a *waiting phase* in order to return to a previous step of the removal plan. This waiting phase is triggered when the number of infected nodes before the front exceeds $\frac{r}{8d}$. In practice, this threshold value can be very small, and we will see in our experiments that this may lead to substantial delay, or even failure of the healing plan.



Figure 9.3: Simulation of an SIS epidemic in the GermanSpeedway network, under the control of various DRA strategies. N = 1168, $\delta = 0$, $\rho^* = r$ (thus q = 1), $\beta = 1$.

For all our experiments, we consider a constant budget r(t) = r as in Assumption 8.1, and a fixed number of treatments $q \in \{1, ..., 100\}$, as in Example 8.3.

9.5.2 Quality of the theoretical bound

Fig. 9.2 shows the relationship between the *maxcut* $C^*(\ell)$ and the resource threshold r^* under a specific priority planning and budget. The resource threshold is computed by running simulations with a fixed number of treatments q, and finding the resource budget above which the strategy is able to remove the epidemic. Each plotted point is a simulation with fixed network, number of treatments, and epidemic parameters β and δ . To cover a wide range of scenarios, each of the 100 points plotted in each subfigure of Fig. 9.2 involves:

- i) the priority-order of a DRA control strategy, randomly chosen among MCM, RAND, MN, LN, and LRSR,
- ii) a fixed q value, set at random in $\{1, ..., 100\}$, and
- **iii)** a random network of 1,000 nodes, constructed by employing at random a generator for: *Erdös-Rényi*, *preferential attachment*, *small-world*, *geometric random*, and *2D regular grids* (details on these networks are available in Sec. 5.3 and the introductory book written by Newman (2010)).

According to the results illustrated in Fig. 9.2, the resource threshold is always below, but very close, to $\beta C^*(\ell)$ which seems to be a very good approximation of the former. The very stable, nearly linear, behavior holds even for low infectivity where the random self-recoveries of nodes become more significant (Fig. 9.2(b)). Overall, this result justifies the minimization of $C^*(\ell)$ as a proxy for removing a contagion with less resources.

9.5.3 Empirical evaluation of simulated contagion on real networks

In this section, we perform simulations on three real networks matching different use cases of DRA strategies: the GermanSpeedway network (Kaiser and Hilgetag, 2004) for



Figure 9.4: Simulation of an SIS epidemic in the OpenFlights airport network, under the control of various DRA strategies. N = 2939, $\delta = 0$, $\rho^* = r$ (thus q = 1), and $\beta = 1$.

analyzing the growth of an epidemic through the road network, the OpenFlights airport network¹ for epidemics spreading through air routes, and a subgraph of the Twitter network (McAuley and Leskovec, 2012) for the undesired spread of information in a social network (e.g. rumors).

In order to compare to the CURE policy, we consider a simplified setting (matching the limitations set in the work of Drakopoulos et al. (2014a)) for the experiments on the GermanSpeedway network and OpenFlights aiport network: we use only one treatment (q=1) and let no self-healing ($\delta=0$). On the TwitterNet, we use a more realistic scenario with 100 treatments and self-healing to show the robustness of the MCM strategy.

GermanSpeedway network

This is the German Autobahn network from the work of Kaiser and Hilgetag (2004). Due to the spatial properties of road networks, the respective graph is symmetric, with a single connected component, and close to being *planar* (i.e. a graph embedding on the plane would create only very few edge intersections other than the endpoint connections). It contains 1,168 nodes and 1,243 edges, while the degree distribution is particularly flat: 101 nodes are leaves, 971 nodes have degree 2, and 96 nodes have degree 3. Finally, the maximum degree is d = 12.

Two scenarios of SIS epidemics with different resource budget r are shown in Fig. 9.3. In both of them, MN and RAND are the worst strategies. MCM is the best performing strategy, and all the other strategies are strongly affected by the resource budget. LRSR is the second best performing strategy, although the low budget causes an increase in the extinction by a factor of 7. The CURE policy presents a behavior with characteristic ups-and-downs, which is due to its waiting phase. Even for a high resource budget, entering the waiting phase can happen with non negligible probability (see Fig. 9.3(b) for an example of such a scenario) and largely degrade the performances of the CURE policy.

¹Source of data: OpenFlights, http://Openflights.org. Available at T. Opsahl's web post: http://wp.me/poFcY-Vw.

The capacity of the strategies to remove the epidemics is correctly predicted by their maxcuts (except CURE whose behavior depends on whether or not the strategy enters its waiting phase): 650 ± 50 for RAND, 379 for MN and LN, 104 for LRSR, and 29 for CURE and MCM.

OpenFlights airport network

This network represents the US air traffic for the year 2010. The nodes are the US airports, plus those non-US airports connected through flights with the former. We used a symmetric, undirected, and unweighted version of this network containing 2,939 nodes in a single connected component with 30,501 edges. For this network, d = 242.

Fig. 9.4 presents two epidemic scenarios similarly to Fig. 9.3. MCM is the best performing strategy and the least affected by the variation in the resource efficiency. LRSR, on the other hand, fails completely with a low resource budget, which is due to the fact that its maxcut is located at the beginning of the considered node ordering (see more details in the experiment in Fig. 9.5). The CURE strategy presents again an unstable behavior as an effect of its waiting phase, which also shows that the conditions under which this policy gets into the waiting phase are not rare at all and can be catastrophic. Indeed, CURE is outperformed even by RAND in these simulations.

Again, the capacity of the strategies to remove the epidemics is correctly predicted by their maxcuts (except CURE): $7,800 \pm 100$ for RAND, 7,504 for MN and LN, 6,223 for LRSR, and 2,231 for CURE and MCM.

TwitterNet social network

This network consists of 1,000 ego-networks extracted from the Twitter social network by McAuley and Leskovec (2012). We use a symmetrized and undirected version of the network which has 81,306 nodes, 1,342,303 edges, and d = 3,383. This network has a single connected component and contains a rich community structure.

Regarding Theorem 9.1, the maximum degree d=3,383 in the network leads to an $\epsilon=20$ for the MCM strategy ($\epsilon=2.1$ for RAND). Tab. 9.1 summarizes the characteristics of the different priority-orders of the compared strategies, where MCM achieves a *five times smaller* $C^*(\ell)$ value than the second best LRSR. This implies that MCM would need a five times smaller resource budget compared to LRSR so as to contain a diffusion process on this network.

The cuts at every position of the LRSR and MCM plans are shown in Fig. 9.5(c) (rotated plot). Fig. 9.5(a)-9.5(b) show two scenarios of full initial infection, varying in the resource efficiency, where MCM performs best in removing the diffusion over time.

Furthermore, Fig. 9.5(d)-9.5(e) provide more insights about the scenario of Fig. 9.5(b). The evolution of each diffusion is illustrated as follows: each line of the figure contains the state of one node of the network throughout the simulation (black: contagious, white: healthy), and the nodes are sorted in the y-axis according to the considered priority-order (aligned y-axis with Fig. 9.5(c)). We can observe that the maxcut acts as a barrier for LRSR: the large cuts at the beginning of the LRSR order prevent the strategy from healing more than the first 5,000 nodes of the priority-order ($\beta C_{5,000}(\ell_{LRSR}) \approx r = 12,000$ for the 5,000-th node). Contrary, MCM gradually reduces the contagion, which is visible by the clear advancement of the front. Note also that some nodes become healthy beyond the front due to self-recovery.

Strategy	Maxcut	Maxcut	Expected resource threshold
		% w.r.t. RAND	$(\delta = 1, \beta = 0.1, q = 100)$
RAND	$670,000 \pm 1000$	100.0 %	67,000
MN	628,571	93.8 %	62,957
LN	628,571	93.8 %	62,957
LRSR	349,440	52.2 %	34,944
МСМ	71,956	10.7 %	7,196

Table 9.1: Maxcut values ($C^*(\ell)$) for different DRA strategies in the TwitterNet dataset (McAuley and Leskovec, 2012). According to Theorem 9.4 and the empirical validation of Fig. 9.2, the expected resource threshold is $\beta C^*(\ell)$.

In agreement with our previous analysis, our results show that: i) the uniform mixing hypothesis leads to a massive drop in efficiency, since MCM substantially outperforms the RAND strategy, ii) despite being efficient in the static vaccination problem, centralitybased priority-orders are suboptimal for the DRA problem, iii) the $\beta C^*(\ell)$ value can be used as a good criterion for assessing the quality of a priority-order, iv) the CURE policy fails to be effective in practical applications due to its waiting phase, and v) MCM outperforms all its competitors in all our experiments.

9.6 Robustness of MCM

Consider that the authorities are prepared to react to an epidemic outbreak using the MCM priority planning approach. They have thus precomputed a priority-order ℓ for the network under threat, which is optimized to have the minimum maxcut, i.e. $C^*(\ell) = W$ and, without loss of generality, let us assume that the maxcut value is unique along ℓ . In essence, the robustness analysis for MCM reduces to a study of how robust the priority-order ℓ is, in the presence of noise or perturbations in the network structure. Specifically, one needs to analyze how the maxcut property is affected by such modifications, since it is the quantity that determines the expected extinction time of the epidemic.

Under this perspective, we briefly discuss about cases where the observed network, utilized by the MCM strategy to compute the priority-order, differs from the real underlying network in which the epidemic is spread. This situation can arise for several reasons, including:

- 1) Malicious modification of the network in order to lessen the efficiency of an employed treatment strategy.
- **2)** Uncertainty in the data used for inferring the network, e.g. the position of people when considering contact networks.
- 3) Misclassification of edges when inferring a network from relational data.
- **4)** Use of outdated data, i.e. a past observation of the network structure which may have changed since then (edges may have appeared or been removed).



Figure 9.5: (a)-(b) Simulation of a diffusion process in the TwitterNet subset of 81,306 nodes (McAuley and Leskovec, 2012). $\delta = 1$, $\beta = 0.1$ and q = 100 (recall $\rho^* = r/q$). MCM clearly outperforms the other heuristics. (c) Cuts for LRSR and MCM. (d)-(e) Visualization of the diffusion of (b) at the node level (infected nodes in black). Nodes are ordered according to LRSR and MCM priority-orders, respectively. The inset figure provides a closer look in (e).

The aforementioned scenarios produce quite different noise profiles, with the malicious intervention (1) as the worst case scenario where the worst modifications are applied to the network intentionally. However, we will see in Sec. 9.6.3 that the MCM exhibits robustness for the very natural scenario of type (2), where there is uncertainty for the position of nodes on the space in which a contact network is formed. The misclassification of edges (3) may, as a first approximation, be modeled by a purely random noise that is adequate in most cases. Nevertheless, correlated noise may appear in practice because an inference algorithm may consistently fail to recover particular structures of edges.

Interestingly, we will see in Sec. 9.6.2 that adding random noise is very similar to malicious modifications w.r.t. the maxcut. Finally, the dynamic evolution of the network can be seen as an intermediate scenario between (1) and (2) since edges can appear anywhere, but, in contact or social networks, they will tend to appear at already strongly connected neighborhoods (i.e. increasing the local edge density by creating triangles).

9.6.1 Malicious modification of the network

Knowing the priority-order ℓ that the authorities use, a malicious agent can make *K* adversarial modifications to the edges of the network. Removing edges may destroy the optimality of ℓ . However, it cannot cause the increase of the maxcut. As for adding new edges, in the worst case, each of them would increase the maxcut by one (by linking one

node before and one node after the maxcut in the ordering). Thus, for the maliciously modified network, we have:

$$\mathcal{C}^*_{\text{MAL}}(\ell) = \mathcal{C}^*(\ell) + K. \tag{9.20}$$

Thus, for networks with large minimum $C^*(\ell)$ (e.g. linear w.r.t. network size, see Sec. 9.3.2 for examples), a noticeable reduction of the effectiveness of the MCM strategy requires the addition of a very large number of edges (e.g. proportional to the size of the network).

9.6.2 Random additive noise

Consider that, for each node pair (i, j), the edge state changes with probability $p \in [0, 1]$. The expected maxcut will then be bounded by:

$$C^{*}(\ell)(1-2p) + pc^{*}(N-c^{*}) \le \mathbb{E}[C^{*}_{\text{RAND}}(\ell)] \le C^{*}(\ell) + \frac{pN^{2}}{2},$$
 (9.21)

where $c^* = \operatorname{argmax}_c C_c(\ell)$ is the position where $\mathcal{C}^*(\ell)$ is located in the priority-order (see Eq. 9.3). When the maxcut does not lay at the beginning or end of the priority-order, $c^* = \alpha N$ where $\alpha \in]0,1[$, which leads to $\mathcal{C}^*(\ell)(1-2p) + pc^*(N-c^*) = \mathcal{C}^*(\ell)(1-2p) + \alpha(1-\alpha)\frac{pN^2}{2}$. Note that, since the number of modified edges is in expectation $K = \frac{pN(N-1)}{2}$, adding random edges is of the same order of magnitude as maliciously adding edges in order to increase the maxcut. This is understandable since, if the maxcut is in the middle of the priority-order, then choosing two nodes at random will give with 50% probability a pair that lays on both sides to the cut position, and hence increase the maxcut by one.

9.6.3 Uncertainty in the localization of nodes in contact networks

For contact networks, a usual scenario consists in acquiring information on the approximate positions of infected, and sometimes healthy, individuals, and then according to it to infer the contact network. Due to the scarcity of the data and the amount of noise, a large scale description of the network is usually preferred, for example by computing the number of cases in large areas, instead of spotting each infected individual. In such a setting, localizing the nodes plays a key role in identifying the *overall* structure of the network, e.g. how cities are connected to one another, while the *local* structure of the contact network is unobservable and highly variable. However, a good property of the MCM strategy is that the optimized priority-order may benefit from the *overall* structure of the network, even though the *local* node connectivity may be inaccurate. In order to investigate this characteristic, we performed simulations on random geometric networks (Penrose, 2003) generated in the following way:

- i) the N nodes are first placed at a position drawn uniformly in the square $[0,1]^2$, and then
- ii) each node is connected to all other nodes that lay within a distance *r* away from them.

These networks are simple models imitating real contact networks, in which people that are closer than a distance r are considered to be in contact.



Figure 9.6: The effect of noise in the node locations of a contact network to the maxcut $C^*(\ell)$: (a) Comparison between the $C^*(\ell)$ of a random priority-order (green), the optimal priority-order that uses the correct node locations (red), and the MCM ordering computed on a noisy version of the network (blue), (b) changes in the network due to the added noise.

Now, we will consider that the position of the nodes used for finding a priority-order ℓ were noisy, and subject to an additive Gaussian noise of standard deviation σ . Hence, the quality of the computed priority-order ℓ becomes poorer as σ grows. The protocol we use is the following. We first generate a random geometric network and compute ℓ by optimizing the maxcut (we use the MLA relaxation, see Sec. 9.4.1). Then we add Gaussian noise to the location of the nodes, recompute the edges of the network, and finally assess the new value of the maxcut along the precomputed priority-order ℓ .

Fig. 9.6(a) shows the increase in the maxcut for the priority-order ℓ as a function of noise intensity. First, we can see that, without any noise, the maxcut of the optimized priority-orders is some orders of magnitude lower than that of a random node ordering (red and blue lines vs. green line). This is expected for networks embedded in 2D spaces (e.g. 2D-grids). Second, as σ increases to become three times bigger than the neighborhood size r, the maxcut of both optimized priority-orders (for the exact and noisy network; red vs blue line) are of the same order of magnitude, and remain low compared to the random ordering.

Finally, using Fig. 9.6(b), one can compare the increase in the maxcut to the number of modifications in edges due to noise. The number of changes increases rapidly w.r.t. σ (note the scale difference between Fig. 9.6(a)-9.6(b)), and for $\sigma = 3r$, most of the edges of the original network are removed. From the perspective of local connectivity and edge to edge comparison, the original network has almost nothing in common with its noisy counterpart. However, the priority-order computed on the noisy contact network is still valid in terms of the large-scale network structure. Therefore, this is a very sound empirical finding for the robustness of the MCM strategy, which testifies that it can deliver good performance in realistic settings where there is high uncertainty for the localization of nodes in a contact network. Note also that this situation is much better than the malicious and random cases, in which the increase in the maxcut is of the same order than the number of the added edges.

9.7. PROOFS

9.7 Proofs

Since the main results hold for any priority-order, we may reorder the nodes of \mathcal{G} according to the priority-order ℓ , and thus consider that $\ell(v_i) = i$ without loss of generality.

Notations. First, let $x \in \{0,1\}^N$ be a state vector of size N (i.e. describing the state of the network during the diffusion process), **0** and **1** vectors of size N that are all-zeros and allones, respectively, and $\bar{x} = \mathbf{1} - x$. For $s \subset \{1, ..., N\}$ a subset of nodes, let also $\mathbf{1}_s = (\mathbb{1}_{\{i \in s\}})_i$ be the indicator vector with ones for nodes in the set s. Then, we define τ_x as the extinction time of the diffusion process starting from the state x, i.e. the time needed for the Markov process to reach its absorbent state $X(t=\tau_x)=\mathbf{0}$ when X(t=0)=x. We also denote the number of infected nodes in network state x as $N_I(x)=\mathbf{1}^\top x$, while $E_{I-S}(x)=x^\top A\bar{x}$ as the number of edges from a contagious to a healthy node, which edges we also refer to as *infectious edges*. We will also make use of the following variables: $q = \lceil \frac{r}{\rho^*} \rceil$, $b(x) = \min\{q, N_I(x)\}$ and $\rho' = \frac{r}{q} + \delta$ is the effective healing rate of a healed node.

We now define infection states that will be useful for proving our main results.

Definition 9.6. For $n \in \{1, ..., N\}$, let $x_n = \mathbf{1}_{\{n,...,N\}}$ be a state vector such that all nodes before n are healthy, and all nodes after n are infected.

Informally, x_n can be seen as the n^{th} step of an ideal removal of the epidemic such that all the nodes after the front remain infected while the nodes are healed one by one following the priority-order. Starting with x_n , adding new infected nodes can either increase or decrease the expected recovery time. In order to bound these changes, we define best and worst additional infections. We will use the following notation for the set of nodes that are possible supports for the increment of infection:

$$S_n^j = \{ s \cup \{n, ..., N\} : s \subset \{1, ..., n-1\}, |s| = j \}$$

Definition 9.7. Let z_n^j be the worst state vector after *j* additional infections from x_n :

$$z_n^j = \underset{s \in S_n^j}{\operatorname{argmax}} \mathbb{E}[\tau_{\mathbf{1}_s}] \,. \tag{9.22}$$

Definition 9.8. Let y_n^j be the best state vector after *j* additional infections from x_n :

$$y_n^j = \underset{s \in S_n^j}{\operatorname{argmin}} \mathbb{E}[\tau_{\mathbf{1}_s}] \,. \tag{9.23}$$

In the following, z_n^j is used for proving upper bounds (Theorems 9.1 and 9.2) and y_n^j lower bounds (Theorem 9.3). Theorems 9.4 and 9.5 of the article are simple corollaries of Theorems 9.1 and 9.3, respectively. The proofs of Theorems 9.1, 9.2, and 9.3 rely on the following lemmas.

9.7.1 Main intermediate results and lemmas

Lemma 9.1. Under priority planning, the function $x \mapsto \mathbb{E}[\tau_x]$ is monotonically increasing with respect to the natural partial order on $\{0,1\}^N$ (i.e. $x \leq y$ if $\forall i, x_i \leq y_i$).

Lemma 9.2. Let $x, y \in \{0, 1\}^N$ be two state vectors such that $x \le y$. Then, denoting $j = \sum_i y_i \bar{x}_i$ as the number of infected nodes in y that are not infected in x, the following inequalities hold:

$$E_{I-S}(x) - jd \le E_{I-S}(y) \le E_{I-S}(x) + jd,$$
 (9.24)

where $d = \max_i \sum_i A_{ij}$ is the highest degree of the network.

Lemma 9.3. Let $\ell(v_i) = i$ for all nodes $v_i \in V$, and x_n be defined as in Definition 9.6. Then, the maxcut of ℓ is equal to:

$$\mathcal{C}^*(\ell) = \max_n E_{I-S}(x_n). \tag{9.25}$$

Lemma 9.4. For every state vector *x*, we have:

$$\mathbb{E}[\tau_x] = \mathbb{E}[T_1 + \tau_{X(T_1)}], \qquad (9.26)$$

where $T_1 = \min\{t \ge 0 : X(t) \neq x\}.$

Proposition 9.1. Set $u_n^j = \mathbb{E}[\tau_{z_n^j}]$ where z_n^j is defined as in Definition 9.7. We have the following recurrence inequality

$$\rho' b(z_n^j)(u_n^j - u_n^{j-1}) \le 1 + \beta E_{I-S}(z_n^j)(u_n^{j+1} - u_n^j), \tag{9.27}$$

Proposition 9.2. Let z_n^j be defined as in Definition 9.7. Then the following bound for the expected extinction time under the ℓ -priority planning and starting from a total infection holds: $\forall K \ge 1, \forall \eta \in]0,1[$ and $r + \delta q > \max\left\{(1 + \frac{1}{\eta})\beta dq, \left[\frac{1}{1-\eta}\sum_{n=1+q}^{N}\prod_{j=0}^{K}\beta E_{I-S}(z_n^{j+q})\right]^{\frac{1}{K+1}}\right\}$

$$\mathbb{E}[\tau_1] \leq \frac{\sum_{k=0}^{K} f(k) + 2q(1+\ln q)}{(r+\delta q)(1-\eta - f(K+1))},$$
(9.28)

where

$$f(k) = \sum_{n=1+q}^{N} \prod_{j=0}^{k-1} \frac{\beta E_{I-S}(z_n^{j+q})}{r+\delta q}.$$
(9.29)

Lemma 9.5. Let $a \ge 0$ and ξ be the (unique) positive solution to $\xi - \ln(1 + \xi) = a$. The following inequality holds:

$$\xi \le a + 2\sqrt{a}.\tag{9.30}$$

9.7.2 **Proofs of theorems**

Proof of Theorem 9.1. Using Lemma 9.2 and Proposition 9.2, we obtain a bound on the extinction time depending on $C^*(\ell) = \max_n E_{I-S}(x_n)$ (using Lemma 9.3),

$$\forall K \ge 1, \forall \eta \in]0,1[\text{ and } \rho' > \max\{(1+\frac{1}{\eta})\beta d, \left[\frac{N}{1-\eta}\prod_{j=0}^{K}\frac{\beta E_{l-S}(z_{n}^{j+q})}{q}\right]^{K+1}\}, \\ \mathbb{E}[\tau_{1}] \le \frac{\sum_{k=0}^{K}f(k) + 2q(1+\ln q)}{\rho'q(1-\eta-f(K+1))},$$

$$(9.31)$$

where

$$f(k) = N \prod_{j=0}^{k-1} \frac{\beta(\mathcal{C}^*(\ell) + (q+j)d)}{\rho' q},$$
(9.32)

9.7. PROOFS

using

$$\sum_{n=1+q}^{N} \prod_{j=0}^{k-1} \frac{\beta(E_{I-S}(x_n) + (q+j)d)}{\rho' q} \le N \prod_{j=0}^{k-1} \frac{\beta(\mathcal{C}^*(\ell) + (q+j)d)}{\rho' q}.$$
 (9.33)

Finally, we need to select proper values for *K* and η and derive the final result. Let $\eta^* = 1 - e^{-1}$, ξ be the unique solution of $\xi - \ln(1 + \xi) = \frac{d(\ln N + 1)}{\hat{C}^*}$ where $\hat{C}^* = C^*(\ell) + qd$, and $K^* = \lfloor \frac{\hat{C}^*}{d} \xi \rfloor$. Using the particular value of K^* ,

$$\begin{split} \sum_{j=0}^{K^*} \ln(1+j\frac{d}{\hat{c}^*}) &\leq \int_0^{K^*+1} \ln(1+x\frac{d}{\hat{c}^*}) dx \\ &= (K^*+1+\frac{\hat{c}^*}{d}) \ln(1+(K^*+1)\frac{d}{\hat{c}^*}) - (K^*+1) \\ &\leq (K^*+1) \ln(1+(K^*+1)\frac{d}{\hat{c}^*}) + \frac{\hat{c}^*}{d} (\ln(1+\xi) - \xi) \\ &= (K^*+1) \ln(1+(K^*+1)\frac{d}{\hat{c}^*}) - \ln(N) - 1, \end{split}$$
(9.34)

where the second inequality is due to $\frac{d}{\hat{C}^*}(K^*+1) \ge \xi$ and the monotonic decrease of $x \mapsto \ln(1+x) - x$ for $x \ge 0$.

From Eq. 9.34, we derive that $f(K^*+1) \leq (1-\eta^*) \left[\frac{\beta}{\rho' q} (\widehat{\mathcal{C}}^* + (K^*+1)d) \right]^{K^*+1}$. We thus have: For $r + \delta q > \beta(\mathcal{C}^*(\ell) + ((1+\frac{1}{\eta^*})q + K^* + 1)d)$,

$$\mathbb{E}[\tau_{1}] \leq \frac{\sum_{k=0}^{K^{*}} f(k) + 2q(1+\ln q)}{\rho' q(1-\eta^{*}-f(K+1))} \\
\leq \frac{N \sum_{k=0}^{K^{*}} \left[\frac{\beta}{\rho' q} (\widehat{C}^{*} + (K^{*}+1)d)\right]^{k} + 2q(1+\ln q)}{\rho' q(1-\eta^{*})(1-f(K^{*}+1))} \\
\leq \frac{3N}{\rho' q - \beta(\widehat{C}^{*} + (K^{*}+1)d)} + \frac{6(1+\ln q)}{\rho' (1-f(K^{*}+1))} \\
\leq \frac{3N + 6q(1+\ln q)}{\rho' q - \beta(\widehat{C}^{*} + (K^{*}+1)d)}.$$
(9.35)

Finally, using Lemma 9.5, $dK^* \leq \widehat{C}^* \xi \leq d(\ln N + 1) + 2\sqrt{\widehat{C}^* d(\ln N + 1)}$, and

$$(1 + \frac{1}{\eta^{*}})q + K^{*} + 1$$

$$\leq 3q + \ln N + 2 + 2\sqrt{\frac{\mathcal{C}^{*}(\ell)}{d}(\ln N + 1)(1 + \frac{qd}{\mathcal{C}^{*}(\ell)})}$$

$$\leq 4q + 2\ln N + 3 + 2\sqrt{\frac{\mathcal{C}^{*}(\ell)}{d}(\ln N + 1)}$$

$$\leq \frac{\mathcal{C}^{*}(\ell)}{d}(\epsilon + 2\sqrt{\epsilon}),$$
(9.36)

using $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ and $2\sqrt{ab} \le a+b$, and where $\epsilon = \frac{d(3+2\ln N+4q)}{C^*(\ell)}$. This final inequality proves the desired bound.

Proof of Theorem 9.2. For $r + \delta q > \beta C^*(\ell) (1 + 2\sqrt{\epsilon} + \epsilon)$, Eq. 9.35 leads to:

$$\mathbb{E}[\tau_{1}] \leq \frac{3N+6q(1+\ln q)}{r+\delta q-\beta(\widehat{C}^{*}+(K^{*}+1)d)} \\
\leq \frac{\eta^{*}(3N+6q(1+\ln N))}{\beta q d} \\
\leq \frac{\eta^{*}(3N+6(1+\ln N))}{\beta} \\
\leq \frac{6N}{\beta},$$
(9.37)

since $q + K^* + 1 \le \frac{\mathcal{C}^*(\ell)}{d} (\epsilon + 2\sqrt{\epsilon}) - \frac{q}{\eta^*}$ according to Eq. 9.36 and $1 + \ln N \le N$.

Proof of Theorem 9.3. Similarly to Proposition 9.2, the expected extinction time of an epidemic starting from a state vector X can be written as the sum of three terms (recall $\delta = 0$ and see proof of Proposition 9.2 for definitions of T_1 and E):

$$\mathbb{E}[\tau_{x}] = \frac{1}{\rho' b(x) + \beta E_{I-S}(x)} + \frac{\rho' b(x)}{\rho' b(x) + \beta E_{I-S}(x)} \mathbb{E}[\tau_{X(T_{1})} | E = 2]$$

$$+ \frac{\beta E_{I-S}(x)}{\rho' b(x) + \beta E_{I-S}(x)} \mathbb{E}[\tau_{X(T_{1})} | E = 3].$$
(9.38)

where $b(x) = \min\{q, N_I(x)\}$ is the number of treatments distributed in the network.

Let $u_n^j = \mathbb{E}[\tau_{y_n^j}]$ where y_n^j is defined as in Definition 9.8. Using Lemma 9.1 and Lemma 9.2, $\forall j \ge q$,

$$\rho' q(u_n^j - u_n^{j-1}) \geq 1 + \beta(E_{I-S}(x_n) - jd)(u_n^{j+1} - u_n^j) \\
\geq \beta(E_{I-S}(x_n) - jd)(u_n^{j+1} - u_n^j),$$
(9.39)

and, $\forall K < n-1$,

$$u_{n}^{q} - u_{n+1}^{q} \geq \prod_{j=q}^{K} \frac{\beta(E_{I-S}(x_{n}) - jd)}{r} (u_{n}^{K+1} - u_{n}^{K}) \\ \geq \frac{1}{r} \prod_{j=q}^{K} \frac{\beta(E_{I-S}(x_{n}) - jd)}{r},$$
(9.40)

since $u_n^{j+1} \ge u_n^j$ using Lemma 9.1 and thus $r(u_n^{K+1} - u_n^K) \ge 1$, and $\rho' q = r$. In particular, let $n = n^*$ be the index of the maxcut, i.e. $C^*(\ell) = E_{I-S}(X_{n^*})$. We thus have:

$$u_{1}^{0} = u_{q+1}^{q}$$

$$= \sum_{n=1}^{N} (u_{n}^{q} - u_{n+1}^{q}) + u_{N+1}^{q}$$

$$\geq u_{n^{*}}^{q} - u_{n^{*}+1}^{q}$$

$$\geq \frac{1}{r} \prod_{j=q}^{K} \frac{\beta(\mathcal{C}^{*}(\ell) - jd)}{r}.$$
(9.41)

The third line is due to the positivity of each $(u_n^q - u_{n+1}^q)$ which is derived from Lemma 9.1.

Let $\eta \in]0, \frac{1}{2}[, \epsilon \in]0, \eta[$, and assume that $q \leq \frac{\mathcal{C}^*(\ell)}{d}$. Let also $K^* = q + \left\lfloor (\eta - \epsilon)(\frac{\mathcal{C}^*(\ell)}{d} - q) \right\rfloor$. Then $K^* \geq q$, $K^* < \frac{\mathcal{C}^*(\ell)}{d} \leq n^* - 1$ due to Lemma 9.2, and the choice of K^* is valid. Eq. 9.41 leads to:

$$u_1^0 \geq \frac{1}{r} \left(\frac{\beta(\mathcal{C}^*(\ell) - K^*d)}{r} \right)^{K^* - q + 1}$$

$$\geq \frac{1}{r} \left((1 - \eta + \epsilon) \frac{\beta(\mathcal{C}^*(\ell) - qd)}{r} \right)^{K^* - q + 1}.$$
(9.42)

If $r < (1 - \eta)\beta(\mathcal{C}^*(\ell) - qd)$, then

$$u_1^0 \geq \frac{1}{r} \left(1 + \frac{\epsilon}{1-\eta} \right)^{K^* - q + 1}$$

$$\geq \frac{1}{r} e^{\ln(1 + \frac{\epsilon}{1-\eta})(\eta - \epsilon)(\frac{\mathcal{C}^*(\ell)}{d} - q)}.$$
(9.43)

 $\begin{array}{l} \mbox{Finally, chosing ϵ^*} = \mbox{argmax}_{\epsilon} \ln(1 + \frac{\epsilon}{1 - \eta})(\eta - \epsilon) \mbox{ leads to } \ln(1 + \frac{\epsilon^*}{1 - \eta})(\eta - \epsilon^*) \geq (\frac{\eta}{2(1 + \ln 2)})^2 \\ \mbox{ and the desired result.} \end{array}$

9.7. PROOFS

Proof of Corollary 9.1. Let q = 1, $\delta = 0$, $\beta = 1$ and $N \ge 20$. If $4C^*(\ell) \ge 16d \ln_2 N$, then, with the notation of Theorem 9.1,

$$\epsilon = \frac{d(7 + 2\ln N)}{\mathcal{C}^*(\ell)} \le \frac{\ln 2}{2} (1 + \frac{7}{2\ln N}), \tag{9.44}$$

and

$$1 + 2\sqrt{\epsilon} + \epsilon \le 4. \tag{9.45}$$

Hence, if $r \ge 4C^*(\ell)$, Theorem 9.1 is applicable and

$$\mathbb{E}[\tau_{\mathbf{1}}] \le \frac{3N+6}{r - \mathcal{C}^*(\ell) \left(1 + 2\sqrt{\epsilon} + \epsilon\right)} \le \frac{26N}{r}.$$
(9.46)

On the contrary, if $4C^*(\ell) < 16d \ln_2 N$, then

$$\mathcal{C}^{*}(\ell)(1+2\sqrt{\epsilon}+\epsilon) = \mathcal{C}^{*}(\ell)+2\sqrt{\mathcal{C}^{*}(\ell)d(7+2\ln N)} + d(7+2\ln N) \\
\leq d\left(7+2(1+\frac{2}{\ln 2})\ln N + \frac{4}{\sqrt{\ln 2}}\sqrt{\ln N(7+2\ln N)}\right) \\
\leq 16d\ln_{2}N,$$
(9.47)

by a simple function analysis. Hence, if $r \ge 16d \ln_2 N$, Theorem 9.1 is applicable and

$$\mathbb{E}[\tau_1] \le \frac{3N+6}{r - \mathcal{C}^*(\ell) \left(1 + 2\sqrt{\epsilon} + \epsilon\right)} \le \frac{26N}{r}.$$
(9.48)

Proof of Lemma 9.1. Let $x, y \in \{0,1\}^N$ be two initial states of the network such that $x \leq y$. If X(t), Y(t) are diffusion processes such that X(0) = x and Y(0) = y, then the strong monotonicity of the Markov process X(t) implies that $\forall t \geq 0$, $\mathbb{P}(X(t) = \mathbf{0}) \geq \mathbb{P}(Y(t) = \mathbf{0})$, which may be rewritten as $\mathbb{P}(\tau_x \leq t) \geq \mathbb{P}(\tau_y \leq t)$. This means that τ_y dominates τ_x and thus $\mathbb{E}[\tau_x] \leq \mathbb{E}[\tau_y]$. The strong monotonicity derives from a standard coupling argument.

Proof of Lemma 9.2. The set of infected nodes at state y consists of the infected nodes at state x and exactly j additional nodes. Since a node can have at most d neighbors, then each of the j additional nodes can add or remove at most d edges to the set of infectious edges of the network. More formally,

$$|E_{I-S}(y) - E_{I-S}(x)| = |\sum_{i,j} A_{ij}(y_i(1-y_j) - x_i(1-x_j))| \\ \le \sum_{i,j} A_{ij}|y_i(1-y_j) - x_i(1-x_j)|.$$
(9.49)

However, since $x \le y$, $|y_i(1-y_j) - x_i(1-x_j)|$ simplifies to $y_i(1-x_i)(1-y_j) + y_j(1-x_j)x_i$. Due to the symmetry of A, we hence have:

$$|E_{I-S}(y) - E_{I-S}(x)| \leq \sum_{i,j} A_{ij} y_i (1 - x_i) (1 - y_j) + \sum_{i,j} A_{ij} y_i (1 - x_i) x_j$$

= $\sum_{i,j} A_{ij} y_i (1 - x_i) (1 - y_j + x_j)$
 $\leq \sum_{i,j} A_{ij} y_i (1 - x_i)$
 $\leq jd.$ (9.50)
Proof of Lemma 9.3. Since $x_n = \mathbf{1}_{\{n,...,N\}}$, we have $E_{I-S}(x_n) = \sum_{ij} A_{ij} x_{ni} (1 - x_{nj}) = \sum_{ij} A_{ij} \mathbf{1}_{\{j < n \le i\}}$ and the maximum over all $n \in \{1, ..., N\}$ matches Definition 9.4 of $\mathcal{C}^*(\ell)$ when $\ell(v_i) = i$ (since A is symmetric).

Proof of Lemma 9.4. The lemma follows from the Markov property of the process X(t) when the control action is a priority planning. The extinction time after T_1 , i.e. the first change of the state vector from X(0) = x to a certain value x', is equal to the extinction time of the process assuming that X(0) = x', hence

$$\mathbb{E}[\tau_x] = \mathbb{E}[\mathbb{E}[\tau_x|T_1, X(T_1)]] = \mathbb{E}[T_1 + \tau_{X(T_1)}].$$
(9.51)

Proof of Proposition 9.1. Three types of events can happen: i) either a node recovers by itself (at a rate δ), ii) a node is healed by a resource (at a rate ρ'), or iii) a node is infected (at a rate β). Let *E* be a random variable representing the type of event that happened at T_1 :

$$E = \begin{cases} 1 & \text{if a node healed by itself at } T_1 \\ 2 & \text{if a node is healed by a resource at } T_1 \\ 3 & \text{if a node is infected at } T_1 \end{cases}$$
(9.52)

Thus:

$$\mathbb{E}[\tau_{x}] = \frac{1}{\delta(N_{I}(x) - b(x)) + \rho' b(x) + \beta E_{I-S}(x)} \left[1 + \delta(N_{I}(x) - b(x)) \mathbb{E}[\tau_{X(T_{1})} | E = 1] + \rho' b(x) \mathbb{E}[\tau_{X(T_{1})} | E = 2] + \beta E_{I-S}(x) \mathbb{E}[\tau_{X(T_{1})} | E = 3] \right].$$
(9.53)

Using Lemma 9.1, we get that $\mathbb{E}[\tau_{X(T_1)}|E=1] \leq \mathbb{E}[\tau_x]$ which leads to:

$$(\rho'b(x) + \beta E_{I-S}(x))\mathbb{E}[\tau_x] \le 1 + \rho'b(x)\mathbb{E}[\tau_{X(T_1)}|E=2] + \beta E_{I-S}(x)\mathbb{E}[\tau_{X(T_1)}|E=3].$$
(9.54)

We finally set $u_n^j = \mathbb{E}[\tau_{z_n^j}]$ to reach the reccurence inequality. Indeed, we have, for all $j \ge q$ (and $j \ge 1$ if n = N + 1), by definition of x_n^{j-1} and x_n^{j+1} , $\mathbb{E}[\tau_{X(T_1)}|E=2] \le u_n^{j-1}$ and $\mathbb{E}[\tau_{X(T_1)}|E=3] \le u_n^{j+1}$. This comes from the fact that, as the order is static, the *j* infected nodes that are among $\{1, ..., n-1\}$ will receive a resource first.

Proof of Proposition 9.2. By iterating Eq. 9.27, we obtain when $n \leq N$:

$$\rho'q(u_n^q - u_{n+1}^q) \\
\leq \sum_{k=0}^{K} \prod_{j=0}^{k-1} \frac{\beta E_{I-S}(z_n^{j+q})}{\rho'q} + \rho'q(u_n^{K+q+1} - u_n^{K+q}) \prod_{j=0}^{K} \frac{\beta E_{I-S}(z_n^{j+q})}{\rho'q} \qquad (9.55) \\
\leq \sum_{k=0}^{K} \prod_{j=0}^{k-1} \frac{\beta E_{I-S}(z_n^{j+q})}{\rho'q} + \rho'qu_1^0 \prod_{j=0}^{K} \frac{\beta E_{I-S}(z_n^{j+q})}{\rho'q},$$

since $u_n^{K+q+1} \le u_1^0$ using Lemma 9.1 and $u_1^0 = \mathbb{E}[\tau_1]$. We can now sum over *n* and use in the following the definition $f(k) = \sum_{n=1+q}^N \prod_{j=0}^{k-1} \frac{\beta E_{I-S}(z_n^{j+q})}{\rho' q}$:

$$\rho' q (1 - f(K+1)) \mathbb{E}[\tau_1] \le \rho' q u_{N+1}^q + \sum_{k=0}^K f(k).$$
(9.56)

The final step consists in upper bounding u_{N+1}^q . Using Eq. 9.27 and Lemma 9.2, $\forall k \leq q$,

$$u_{N+1}^{k} - u_{N+1}^{k-1} \leq \sum_{j=k}^{q} \frac{1}{\rho'_{j}} (\frac{\beta d}{\rho'})^{j-k} + (u_{N+1}^{q+1} - u_{N+1}^{q}) (\frac{\beta d}{\rho'})^{q-k+1} \\ \leq \sum_{j=k}^{q} \frac{1}{\rho'_{j}} (\frac{\beta d}{\rho'})^{j-k} + u_{1}^{0} (\frac{\beta d}{\rho'})^{q-k+1},$$
(9.57)

and, if $\rho' > \beta d$,

$$\begin{aligned}
u_{N+1}^{q} &\leq \sum_{k=1}^{q} \sum_{j=k}^{q} \frac{1}{\rho'_{j}} (\frac{\beta d}{\rho'})^{j-k} + u_{1}^{0} \sum_{k=1}^{q} (\frac{\beta d}{\rho'})^{q-k+1} \\
&\leq \sum_{k=1}^{q} \frac{1}{k} \frac{1}{\rho' - \beta d} + u_{1}^{0} \frac{\beta d}{\rho' - \beta d} \\
&\leq \frac{1 + \ln q}{\rho' - \beta d} + u_{1}^{0} \frac{\beta d}{\rho' - \beta d}.
\end{aligned} \tag{9.58}$$

Replacing u_{N+1}^q with this expression in Eq. 9.56 gives:

$$\mathbb{E}[\tau_{1}] \leq \frac{\sum_{k=0}^{K} f(k) + \frac{q(1+\ln q)\rho'}{\rho' - \beta d}}{\rho' q(1 - \frac{\beta d}{\rho' - \beta d} - f(K+1))},$$
(9.59)

when $1 - \frac{\beta d}{\rho' - \beta d} - f(K+1) > 0$.

Let
$$\eta \in]0,1[$$
. When $\rho' > \max\{(1+\frac{1}{\eta})\beta d, \left[\frac{1}{1-\eta}\sum_{n=1+q}^{N}\prod_{j=0}^{K}\frac{\beta E_{I-S}(z_{n}^{j+q})}{q}\right]^{\frac{1}{K+1}}\},$
$$\frac{\beta d}{\rho'-\beta d} < \eta,$$
(9.60)

$$\frac{\rho' - \beta d}{\rho' - \beta d} < \eta,$$
(9.60)
$$1 - \eta - f(K+1) > 0,$$
(9.61)

and finally

$$\frac{\rho'}{\rho'-\beta d} < 1+\eta < 2, \tag{9.62}$$

which leads to the desired inequality.

Proof of Lemma 9.5. $x - \ln(1+x)$ is convex, thus always above its tangent line: $\forall x_0 > 0$,

$$a = \xi - \ln(1 + \xi) \ge (x_0 - \ln(1 + x_0)) + \frac{x_0}{1 + x_0}(\xi - x_0), \tag{9.63}$$

and thus,

$$\xi \leq \frac{1+x_0}{x_0} (a + \ln(1+x_0)) - 1$$

$$\leq \frac{1+x_0}{x_0} a + x_0.$$
(9.64)

The final result is obtained by setting $x_0 = \sqrt{a}$.

Part III

From networks to continuous spaces

10 On the convergence of networks

"To every complex question there is simple answer... and it is wrong." — H.L. Mencken

Contents

10.1	Introdu	uction
	10.1.1	Related works
	10.1.2	Outline
10.2	Probab	ilistic mapping for network convergence
	10.2.1	Notations
	10.2.2	Quantifying near-isometries
	10.2.3	The space of spaces
	10.2.4	The mapping distance 154
	10.2.5	The Gromov-Wasserstein distance 156
	10.2.6	Examples of distances and mm-spaces 157
	10.2.7	Application to network comparison
10.3	Limits	of several popular graphs
	10.3.1	Grid graphs
	10.3.2	Totally connected graph
	10.3.3	Random sampling from an mm-space
	10.3.4	Geometric random networks 161
	10.3.5	Erdös-Rényi random networks and the stochastic block model 161
10.4	Discus	sion and open problems
10.5	Proofs	
	10.5.1	Proof of the convergence of product mm-spaces
	10.5.2	Proof of the density of undirected graphs 163
	10.5.3	Proofs of the convergence of popular graphs



Figure 10.1: The characteristics of a transportation network are highly dependent on the topology of the space in which it is embedded. NASA Earth Observatory/NOAA NGDC.

10.1 Introduction

In this chapter, our aim is to describe the characteristics of large graphs that display a certain notion of regularity. More specifically, one may argue that a network of millions or billions of nodes will not be affected by adding or removing one node, and that the characteristics of interest have reached a sort of *limiting behavior*. Hence, the description of such a large network may not require the knowledge of each and every node of the network, and its *macroscopic* properties may converge to that of a limit space whose description is simpler and more compact. Such an endeavor is motivated by three reasons:

- 1. First, running complex algorithms (i.e. whose complexity is more than linear in the number of nodes) becomes prohibitive for real-world social networks, and a possible solution to this problem is to find *simpler representations* of such networks.
- Second, the actual network may be impossible to acquire, and working on approximate representations of the network may provide more robust and practical algorithms in real-life applications. A notable example of such a scenario is the control of epidemics using density data.
- 3. Third, the analysis of limits of graphs provides new insights and results for random graphs, which is of particular interest in many practical applications including graph clustering.

We will see in this chapter that a natural way to create such a notion of limit is to define a distance between graphs and continuous spaces, and we will investigate an elegant and powerful method for achieving this using *probabilistic mappings*. A byproduct of our analysis is the definition of convergence for matrices and operators of functional spaces that are easy to use and powerful analytical tools (see Sec. 11.1.1).

10.1.1 Related works

The literature of network convergence is relatively large, and can be divided into four groups: kernel-based distances, graphons and limits of dense graphs, local weak convergence using random rooted graphs, and Gromov-Hausdorff and related distances. A good introduction to network convergence is available in the book by Lovász (2012).

The simplest notions of convergence are based on local characteristics such as the number of triangles, stars, cliques or any particular subgraph of a network. More generally, Borgs et al. (2006, 2008) defined a notion of convergence for dense graphs based on the convergence of the number of subgraphs of any fixed size. However, these notions of convergence usually miss the *macroscopic* structure of the network which drives the large-scale behavior of diffusion processes. One of the most famous notion of limits of graphs is called a *graphon*, and was introduced by Lovász and Szegedy (2006) in order to study the limit of dense graphs in a rigorous way. These objects are very similar to the inhomogeneous random graphs defined in Model 5.5, in which nodes are drawn at random in a certain space ([0,1] for graphons) and then edges are drawn independently at random according to a kernel function. One of the problems of such limits is that they are *locally tree-like*, i.e. the structure of the network is, on a local level, completely random, and triangles are very scarce. However, the contact networks encountered in epidemiology are closer to geometric random networks (see Model 5.4) and have many triangles. As a consequence, graphons are not particularly well-suited to our setup.

A growing body of theoretical works is centered around the *local weak convergence* introduced by Benjamini and Schramm (2001) and later analyzed by Elek (2007, 2012), which describes a graph using random rooted subgraphs whose roots are chosen uniformly at random across the network. While this notion of convergence is relatively strong and a large part of the random matrix literature and theoretical analysis of random networks is based on it, the limiting space is not easy to handle, nor intuitive, and this notion of convergence tends to favor *local* characteristics instead of the macroscopic properties of the network.

In order to impose a more *global* convergence of the *structure* of a network, one may consider the convergence of a graph *as a metric space*. The first works to investigate metric space comparison are probably the works of Gromov (1981, 1999). These theoretical works defined a distance between metric spaces known as the *Gromov-Hausdorff* distance. Several later articles considered the convergence of networks using the Gromov-Hausdorff distance (e.g. Addario-Berry et al. (2012)), in the restricted case of particular random networks.

Finally, the Gromov-Wasserstein distance discussed in this thesis was analyzed in several recent articles, though not for network convergence. In image analysis and object recognition, Mémoli (2011, 2014) used the Gromov-Wasserstein distance to compare 3D objects, and provided interesting and efficient approximations of such distance. In a more theoretical perspective, Sturm (2006, 2013) investigated the properties of the *space of spaces* equipped with the Gromov-Wasserstein distance. More specifically, Strum proved a number of results concerning geodesics, Ricci curvature, tangent space and other geometric characteristics of the space of spaces. More generally, the recent paper by Sturm (2013) is an excellent introduction to this notion of distance, and provides an intuitive discussion and valuable results.

10.1.2 Outline

In Sec. 10.2, we present the notion of *Gromov-Wasserstein distance* between metric and measurable spaces, along with the mathematical tools needed for its analysis. We also define the concept of *mapping distance* between functions lying in two different spaces, which will be necessary for the convergence of diffusion processes when their underlying

space converges. While the notion of Gromov-Wasserstein distance is not new, the presentation given in this section using mapping distances is novel and has the advantages of highlighting the important characteristics of this distance, as well as allowing generalization of such distances based on probabilistic mappings. Sec. 10.3 provides several examples of graphs converging to continuous spaces. Sec. 10.4 provides a discussion on the proposed framework and open problems for future work, while proofs are gathered in Sec. 10.5.

10.2 Probabilistic mapping for network convergence

In this section, we present the main concepts for our analysis of network convergence, as well as a justification of why these specific mathematical concepts were used. More specifically, we consider the problem of network convergence through the perspective of *probabilistic mapping*, that is by comparing two spaces by mapping the first into the second using a coupling of random variables (see Definition 10.1). This approach, in essence very similar to optimal transport theory (see, for example, Villani (2009)), leads to powerful and elegant notions of distance between functions, operators and spaces, that will be used in Chap. 11 to derive a number of continuity theorems for the characteristics of networks.

10.2.1 Notations

Let (\mathcal{X}, μ) be a probability space, $p \ge 1$ a positive number and $f : \mathcal{X} \to \mathbb{R}$ a μ -measurable function. Then, the *p*-norm of *f* is the (possibly infinite) value

$$||f||_{p,\mu} = \mathbb{E}_{\mu}[|f(X)|^{p}]^{1/p} = \left(\int_{\mathcal{X}} |f(x)|^{p} d\mu(x)\right)^{1/p},$$
(10.1)

and let $L_{p,\mu} = \{f : \mathcal{X} \to \mathbb{R} \mid ||f||_{p,\mu} < +\infty\}$ be the space of functions with bounded *p*-norm. Similarly, we define the almost sure maximum value of *f* as

$$||f||_{\infty,\mu} = \inf\{C \in \mathbb{R} \cup \{+\infty\} \mid f(x) \le C \text{ a.s.}\},$$
(10.2)

and $L_{\infty,\mu} = \{f : \mathcal{X} \to \mathbb{R} \mid ||f||_{\infty,\mu} < +\infty\}$ the corresponding functional space. If $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$ are two measures on the spaces \mathcal{X} and \mathcal{Y} , respectively, then $\mu_{\mathcal{X}} \otimes \mu_{\mathcal{Y}}$ is the product measure on the product space $\mathcal{X} \times \mathcal{Y}$. We now define the coupling of two probability measures (see, for example, Villani (2009)):

Definition 10.1 (Coupling). Let $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$ be two probability measures on \mathcal{X} and \mathcal{Y} (respectively). A coupling of $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$ is a probability measure on $\mathcal{X} \times \mathcal{Y}$ such that the marginals over \mathcal{X} and \mathcal{Y} are respectively $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$. The set of all couplings of $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$ is denoted as $\Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$.

Informally, a coupling is a mapping between two probability measures, and is a natural extension of bijective mappings to measure spaces. The most simple example of coupling is the product measure, that creates no correlation between the two measures. For simplicity, when $f : \mathcal{X} \to \mathbb{R}$ is a μ -measurable function on \mathcal{X} and $\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ is a coupling of $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$, we will implicitly extend f to $\mathcal{X} \times \mathcal{Y}$ by $\tilde{f}(x,y) = f(x)$ for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$ and use the notation $||f||_{p,\pi} = ||\tilde{f}||_{p,\pi}$.



Figure 10.2: Isometry problem: can we map one space into another is a way that preserves distances?

10.2.2 Quantifying near-isometries

The intuition behind the *Gromov-Wasserstein distance* is to compare two metric spaces by finding a correspondence between points of \mathcal{X} and \mathcal{Y} that preserves distances, and can be seen as a relaxation of the concept of isometric spaces.

Definition 10.2 (Isometric spaces). Two metric spaces $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ are called isometric *if there exists a mapping* $\phi : \mathcal{X} \to \mathcal{Y}$ such that:

- ϕ is bijective,
- ϕ is an isometry: $\forall (x_1, x_2) \in \mathcal{X}^2$, $d_{\mathcal{Y}}(\phi(x_1), \phi(x_2)) = d_{\mathcal{X}}(x_1, x_2)$.

Isometric spaces are totally equivalent in terms of topology, and would be a natural candidate to compare metric spaces. However, using this notion to compare spaces has several drawbacks:

- 1. Isometries can only identify *perfectly similar* spaces, and do not measure *how far* two spaces are from being isometric.
- The bijection assumption is rather strong, as it imposes the spaces to have the same cardinality. Notably, it is impossible to compare discrete spaces (e.g. graphs) with uncountable spaces (e.g. the real plane ℝ²) since no bijection can exist between a countable and an uncountable set.

The *Gromov-Wasserstein* distance provides a natural solution to these two limitations by relaxing the bijective assumption using a *probabilistic mapping* (i.e. coupling) between the two spaces. The distance then compares the two metrics with a standard L_p -norm and is, in spirit, very similar to the Wasserstein distance between probability measures used in optimal transport theory (Villani, 2009).

10.2.3 The space of spaces

In order to compare two metrics, our strategy is to use a probabilistic coupling between the two spaces. We will thus consider that the metric spaces are equipped with a probability measure. This concept of *metric and measurable space* is formally defined in the work of Sturm (2013), and we use here the same assumptions.

Definition 10.3 (Mm-space). A metric and measurable space (*mm-space for short*) is a triple $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ such that:

- $(\mathcal{X}, d_{\mathcal{X}})$ is a Polish metric space (i.e. complete and separable metric space),
- $\mu_{\mathcal{X}}$ is a Borel probability measure on \mathcal{X} .

Remark 10.1. In the work of Mémoli (2014), $(\mathcal{X}, d_{\mathcal{X}})$ is assumed to be a compact metric space. This assumption can be rather restrictive (e.g. excluding infinite spaces such as $\mathcal{X} = \mathbb{R}$), and is not needed for our purpose.

Remark 10.2. Requiring the metric space to be complete and separable allows the use of advanced probability theory and imposes the measurability of all the sets and functions considered hereafter. This assumption is standard and is also made in optimal transport theory (Villani, 2009).

In the following, we will always denote as $\mu_{\mathcal{X}}$ and $d_{\mathcal{X}}$ the measure and distance of an mm-space \mathcal{X} , and will only refer to \mathcal{X} instead of $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ when there is no ambiguity. The diameter diam (\mathcal{X}) of a metric space \mathcal{X} is the maximum distance between any two points of the space. In order to evaluate *how large* an mm-space is, we define its size as the average distance between two points of the space taken at random, or more formally:

Definition 10.4 (Size). The p-size of an mm-space \mathcal{X} is the quantity

$$size_p(\mathcal{X}) = ||d_{\mathcal{X}}||_{p,\mu_{\mathcal{X}} \otimes \mu_{\mathcal{X}}}$$
(10.3)

where $\mu_{\mathcal{X}} \otimes \mu_{\mathcal{X}}$ is the product measure of $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{X}}$.

Note that, $\forall p \ge 1$, size_{*p*}(\mathcal{X}) \le diam(\mathcal{X}) = $||d_{\mathcal{X}}||_{\infty}$, and hence bounded mm-spaces have a finite *p*-size.

10.2.4 The mapping distance

In order to compare functions lying in different mm-spaces, we now define a distance between functions under a fixed coupling. This distance will later help us define distances between operators and mm-spaces under the probabilistic mapping setting.

Definition 10.5 (Mapping distance for functions). Let \mathcal{X} and \mathcal{Y} be two mm-spaces, and $f \in L_{p,\mu_{\mathcal{X}}}$ and $g \in L_{p,\mu_{\mathcal{Y}}}$ two functions on \mathcal{X} and \mathcal{Y} respectively. Then, the (p,π) -mapping distance between f and g is the p-norm of the difference between the two functions with respect to the coupling π :

$$\mathbf{\Delta}_{p,\pi}(f,g) = \left(\int_{\mathcal{X}\times\mathcal{Y}} |f(x) - g(y)|^p d\pi(x,y)\right)^{1/p},\tag{10.4}$$

where $\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ and $\Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ is the set of all couplings between $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$ (see Definition 10.1). Then, the *p*-mapping distance between *f* and *g* is the minimum of $\Delta_{p,\pi}(f,g)$ over all possible couplings:

$$\Delta_{p}(f,g) = \inf_{\pi \in \Pi(\mu_{\mathcal{X}},\mu_{\mathcal{Y}})} \Delta_{p,\pi}(f,g).$$
(10.5)

Intuitively, this distance compares the values of f and g through a probabilistic mapping (i.e. coupling) between \mathcal{X} and \mathcal{Y} . Note that $\Delta_p(f,g)$ can also be seen as the Wasserstein distance between the probability distributions of f(X) and g(Y), where $X \sim \mu_{\mathcal{X}}$ and $Y \sim \mu_{\mathcal{Y}}$ (see the book of Villani (2009) for more information on this quantity).



Figure 10.3: Probabilistic mapping between a graph and the 2D square $[0,1]^2$. Each node is mapped to a probability distribution (here uniform measures over distinct subsets for simplicity) that add up to the uniform measure on $[0,1]^2$.

Remark 10.3. The Wasserstein distance $W_p(\mu, \nu)$ between two probability measures μ and ν of \mathbb{R} can also be seen as a mapping distance between identity functions of \mathbb{R} equipped with the probability measures μ and ν , respectively:

$$W_p(\mu,\nu) = \Delta_p(Id_{(\mathbb{R},\mu)}, Id_{(\mathbb{R},\nu)}),$$

where $Id_{(\mathbb{R},\mu)}(x) = x$. Hence, the mapping distance is a general framework able to describe the Wasserstein distance, the Gromov-Wasserstein distance, and a generalization of the operator norm for operators of an mm-space (see Sec. 11.1.1).

The mapping distance, as for L_p -norms and their induced distances, is a pseudo-metric for the space of $L_{p,\mu_{\mathcal{X}}}$ functions of all mm-spaces \mathcal{X} , and one can define a space \mathbb{L}_p of equivalence classes for Δ_p in which the mapping distance is indeed a distance.

Theorem 10.1. Δ_p is a pseudometric for the set of $L_{p,\mu_{\mathcal{X}}}$ functions of all mm-spaces \mathcal{X} .

Proof. First, $\Delta_p(f,g)$ is well defined since $|f(x) - g(y)|^p \le 2^{p-1}(f(x)^p + g(y)^p)$ by convexity of $|x|^p$, and thus $\Delta_p(f,g) \le 2^{p-1}(||f||_{p,\mu_{\mathcal{X}}} + ||g||_{p,\mu_{\mathcal{Y}}}) < +\infty$. Second, note that $\Delta_{p,\pi} = ||\tilde{f} - \tilde{g}||_{p,\pi}$ where $\tilde{f}(x,y) = f(x)$ and $\tilde{g}(x,y) = g(y)$. The symmetry, positivity and triangular inequality follow immediately.

Definition 10.6. Let \mathbb{L}_p be the quotient space of the space of all $L_{p,\mu_{\mathcal{X}}}$ functions of all mmspaces \mathcal{X} by the equivalence relation $f \sim_p g \Leftrightarrow \Delta_p(f,g) = 0$:

$$\mathbb{L}_p = \bigcup_{\mathcal{X} \in \mathbb{X}} L_{p, \mu_{\mathcal{X}}} / \sim_{p}, \tag{10.6}$$

where X is the set of all mm-spaces. Then, Δ_p is a distance in \mathbb{L}_p .

Also, two functions f and g are part of the same equivalence class if and only if there exists a coupling such that f(X) = g(Y) a.s. .

Lemma 10.1. Let \mathcal{X} and \mathcal{Y} be two mm-spaces, and $f \in L_{p,\mu_{\mathcal{X}}}$ and $g \in L_{p,\mu_{\mathcal{Y}}}$ two functions on \mathcal{X} and \mathcal{Y} respectively. Then the following assertions are equivalent:

- $\Delta_p(f,g) = 0.$
- There exists X and Y two random variables of marginal probability measure μ_X and μ_Y , respectively, such that f(X) = g(Y) a.s.

10.2.5 The Gromov-Wasserstein distance

In this section, we define the notion of Gromov-Wasserstein distance between mm-spaces, and give a few preliminary results on this quantity. This notion was defined to compare metric and measurable spaces, and is simply the distance between the metrics of the two mm-spaces, up to a probabilistic mapping.

Definition 10.7 (Gromov-Wasserstein distance). Let $p \ge 1$, and \mathcal{X} and \mathcal{Y} two mmspaces of finite *p*-size. The *p*-Gromov-Wasserstein distance between \mathcal{X} and \mathcal{Y} is defined as follows:

$$\Delta_{p}(\mathcal{X},\mathcal{Y}) = \inf_{\pi \in \Pi(\mu_{\mathcal{X}},\mu_{\mathcal{Y}})} \Delta_{p,\pi \otimes \pi}(d_{\mathcal{X}},d_{\mathcal{Y}}),$$
(10.7)

where $\Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ is the set of all couplings of $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$, and $\pi \otimes \pi$ is the product measure.

This distance is also referred to as L_p -distortion distance by Sturm (2013), and can be rewritten as:

$$\Delta_{p}(\mathcal{X},\mathcal{Y}) = \inf_{\pi \in \Pi(\mu_{\mathcal{X}},\mu_{\mathcal{Y}})} ||\tilde{d}_{\mathcal{X}} - \tilde{d}_{\mathcal{Y}}||_{p,\pi \otimes \pi},$$
(10.8)

where $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$ are extended to take values in $(\mathcal{X} \times \mathcal{Y})^2$ as follows: $\tilde{d}_{\mathcal{X}}(x_1, y_1, x_2, y_2) = d_{\mathcal{X}}(x_1, x_2)$ and $\tilde{d}_{\mathcal{Y}}(x_1, y_1, x_2, y_2) = d_{\mathcal{Y}}(y_1, y_2)$. As for the mapping distance, the Gromov-Wasserstein distance is a pseudometric in the space of all mm-spaces, and becomes a distance on the space of equivalence classes w.r.t. Δ_p .

Theorem 10.2. Δ_v is a pseudometric for the set of all mm-spaces of finite p-size.

Proof. First, $\Delta_p(\mathcal{X}, \mathcal{Y})$ is well defined since $|d_{\mathcal{X}}(x_1, x_2) - d_{\mathcal{Y}}(y_1, y_2)|^p \leq 2^{p-1}(d_{\mathcal{X}}(x_1, x_2)^p + d_{\mathcal{Y}}(y_1, y_2)^p)$ by convexity of $|x|^p$, and thus $\Delta_p(\mathcal{X}, \mathcal{Y}) \leq 2^{p-1}(\operatorname{size}_p(\mathcal{X}) + \operatorname{size}_p(\mathcal{Y})) < +\infty$. The symmetry, positivity and triangular inequality follow immediately from Eq. 10.8.

Definition 10.8. Let X_p be the quotient space of the space of all metric and measurable spaces of finite *p*-size by the equivalence relation $\mathcal{X} \sim_p \mathcal{Y} \Leftrightarrow \Delta_p(\mathcal{X}, \mathcal{Y}) = 0$. Then, Δ_p is a distance in X_p .

An important result of the literature is that the infimum in Eq. 10.7 is met by a particular coupling (Sturm, 2013). This result will be useful for our analysis and allow us to remove the infimum in Eq. 10.7 and directly refer to one of the *optimal couplings*.

Lemma 10.2. For each $p \ge 1$ and pair of mm-spaces \mathcal{X} and \mathcal{Y} , the infimum in the definition of $\Delta_p(\mathcal{X}, \mathcal{Y})$ can be reached. That is, there exists a coupling $\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ such that

$$\Delta_{p}(\mathcal{X},\mathcal{Y}) = \Delta_{p,\pi\otimes\pi}(d_{\mathcal{X}},d_{\mathcal{Y}})$$
(10.9)

Proof. See (Sturm, 2013, Lemma 1.7).

Definition 10.9. A coupling is called optimal (for Δ_p) if Eq. 10.9 is satisfied. Moreover, we will denote as $Opt_n(\mathcal{X}, \mathcal{Y})$ the set of all such couplings.



Figure 10.4: Probabilistic mapping of Example 10.2 between a line and a 2D stripe. A point on the former is mapped to a line segment on the latter.

10.2.6 Examples of distances and mm-spaces

We now provide a few simple examples to make the reader familiar with the aforementioned concepts.

Example 10.1. Let $\{0\}$ be the mm-space with only one element. Then, the size of any mm-space \mathcal{X} is equal to its distance to $\{0\}$:

$$\Delta_p(\mathcal{X}, \{0\}) = size_p(\mathcal{X}). \tag{10.10}$$

The mm-space $\{0\}$ plays a similar role to that of the zero value of a vector space, and this intuition is used by Sturm (2013) to define scale-invariant distances between mm-spaces.

Example 10.2. Let $\mathcal{X} = [0,1]$ and $\mathcal{Y} = [0,1] \times [0,a]$ where a > 0. When $a \ll 1$, a very reasonable coupling is to map each point $x \in [0,1]$ uniformly to the horizontal line $\{(y_1, y_2) \in \mathcal{Y} : y_1 = x \text{ and } y_2 \in [0,a]\}$ (see Fig. 10.4). In mathematical terms, this is equivalent to the coupling

$$d\pi(x,y) = \delta_{\{y_1=x\}} dx \frac{dy_2}{\epsilon},\tag{10.11}$$

where $\delta_{\{\cdot\}}$ is the Dirac measure. Then, the Gromov-Wasserstein distance is upper bounded by:

$$\Delta_{p}(\mathcal{X},\mathcal{Y}) \leq \Delta_{p,\pi\otimes\pi}(d_{\mathcal{X}},d_{\mathcal{Y}}) = a\left(\frac{2^{3/2}}{(p+1)(p+2)}\right)^{1/p}.$$
(10.12)

This imposes that $\lim_{a\to 0} [0,1] \times [0,a] = [0,1]$ w.r.t. Δ_p .

More generally, one can show that, if $\lim_{n\to+\infty} \text{size}_p(\mathcal{Y}_n) = 0$, then, under mild technical constraints, $\lim_{n\to+\infty} \mathcal{X} \times \mathcal{Y}_n = \mathcal{X}$ using the following result.

Theorem 10.3. Let \mathcal{X} and \mathcal{Y} be two mm-spaces, and let $\mathcal{X} \times \mathcal{Y}$ be the mm-space equipped with a coupling measure $\mu_{\mathcal{X} \times \mathcal{Y}} \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ and a metric $d_{\mathcal{X} \times \mathcal{Y}}((x, y), (x', y'))$. Furthermore, let us assume that $d_{\mathcal{X} \times \mathcal{Y}}$ verifies the following constraints:

• $\forall (x, x', y) \in \mathcal{X}^2 \times \mathcal{Y}, d_{\mathcal{X} \times \mathcal{Y}}((x, y), (x', y)) = d_{\mathcal{X}}(x, x'),$



Figure 10.5: Mapping $f(\theta) = e^{i\theta}$ of Example 10.3 between the interval $[0, 2\pi]$ and the unit circle. While the blue distance is well preserved, the green one (passing through $\theta = 0$) incurs a large error.

•
$$\forall (x,y,y') \in \mathcal{X} \times \mathcal{Y}^2, d_{\mathcal{X} \times \mathcal{Y}}((x,y),(x,y')) = d_{\mathcal{Y}}(y,y').$$

Then the Gromov-Wasserstein distance between \mathcal{X} and $\mathcal{X} \times \mathcal{Y}$ is bounded by the size of \mathcal{Y} :

$$\Delta_{p}(\mathcal{X}, \mathcal{X} \times \mathcal{Y}) \leq size_{p}(\mathcal{Y}).$$
(10.13)

Example 10.3. Let $\mathcal{X} = [0, 2\pi[$ be an interval and $\mathcal{Y} = \{x \in \mathbb{C} : |x| = 1\}$ the unit circle. A deterministic coupling exists between these two spaces: $f(\theta) = e^{i\theta}$ for $\theta \in \mathcal{X}$ (see Fig. 10.5). This gives the following bound on the p-Gromov-Wasserstein distance between \mathcal{X} and \mathcal{Y} :

$$\Delta_{p}(\mathcal{X}, \mathcal{Y}) \leq \frac{2\pi}{\left(4(p+1)(p+2)\right)^{1/p}}.$$
(10.14)

10.2.7 Application to network comparison

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a weighted and undirected network of *n* nodes and adjacency matrix *A*. \mathcal{G} has a natural topology (made of the neighborhoods in \mathcal{G}) and uniform probability measure. However, in order to create an mm-space from \mathcal{G} , a metric between elements of \mathcal{V} should be specified. We thus consider three metrics on graphs:

- The shortest path distance $d_{SP}(u, v) = \min_{(c_1, \dots, c_K) \in \mathcal{P}_{uv}} \sum_k A_{c_k c_{k+1}}$ where \mathcal{P}_{uv} is the set of paths in \mathcal{G} from u to v. This is a very natural distance and will be our default metric throughout this document. However, it is very sensitive to noise in the adjacency matrix, and in practical applications we may prefer more robust distances such as the *random walk distance*.
- The random walk distance $d_{RW}(u, v) = \mathbb{E}[\tau_{uv} + \tau_{vu}]$ where τ_{uv} is the time taken by a random walk with initial state u to reach v. In practice, d_{RW} can be computed by taking the pseudo-inverse of the Laplacian matrix K, and return $d_{RW}(v_i, v_j) = \sqrt{K_{ii} + K_{jj} - 2K_{ij}}$.
- Any inferred metric based on the adjacency matrix *A*, e.g. using a maximum likelihood over a certain parametric model of random graphs.

10.3. LIMITS OF SEVERAL POPULAR GRAPHS

In the following, we assume that a metric $d_{\mathcal{G}}$ is chosen, and we denote $(\mathcal{G}, d_{\mathcal{G}}, \mathcal{U}(\mathcal{V}))$ the mm-space of the graph \mathcal{G} , where $\mathcal{U}(\mathcal{V})$ is the uniform distribution over the discrete set \mathcal{V} and, unless specified otherwise, the distance $d_{\mathcal{G}}$ is the shortest-path distance. When comparing two graphs, the Gromov-Wasserstein distance has the following matrix form:

Theorem 10.4. Let \mathcal{G} and \mathcal{G}' be two graphs of size n and n' respectively. Then, the Gromov-Wassertstein distance between \mathcal{G} and \mathcal{G}' can be rewritten as:

$$\boldsymbol{\Delta}_{p}(\mathcal{G},\mathcal{G}') = \inf_{\substack{P \in [0,1]^{n \times n'} \text{ s.t.} \\ P\mathbf{1} = \frac{1}{n}\mathbf{1} \\ P^{\top}\mathbf{1} = \frac{1}{n'}\mathbf{1}}} \left(\sum_{i,j,k,l} |D_{ij} - D'_{kl}|^{p} P_{ik} P_{jl}\right)^{1/p}, \quad (10.15)$$

1 /

where D and D' are the distance matrices of \mathcal{G} and \mathcal{G}' (i.e. $D_{ij} = d_{\mathcal{G}}(i,j)$).

Note that the optimization problem to solve is quadratic in the projection matrix P, and not necessarily convex. Hence, in general, this problem is NP-hard. However, a number of heuristics are available for solving this problem approximately (Mémoli, 2014).

An interesting intuition for generic mm-spaces is that, properly re-normalized, undirected graphs are dense in the space of all mm-spaces X_p . This means that mm-spaces can be seen as limits of graphs, and that *all the characteristics of graphs that are continuous w.r.t.* Δ_p *can also be defined for generic mm-spaces*.

Theorem 10.5. The subset of undirected graphs equipped with a distance metric of the form αd_{SP} for $\alpha > 0$ is dense in the space of mm-spaces X_p with respect to Δ_p .

For sparse networks, rescaling distances is usually required in order for the network to converge to a continuous space, and, for any $\alpha > 0$, we will use the notation $\alpha \mathcal{X}$ for the rescaled mm-space ($\mathcal{X}, \alpha d_{\mathcal{X}}, \mu_{\mathcal{X}}$).

10.3 Limits of several popular graphs

10.3.1 Grid graphs

Grid graphs are among the simplest examples of converging networks, and their convergence to \mathbb{R}^d made them practical tools for approximating continuous processes and partial differential equations for numerical simulations. However, an interesting result of our theory is that grids do not converge to the same metric space, and although *d*-dimensional grids always tend to \mathbb{R}^d , the induced metric on \mathbb{R}^d depends on the number of neighbors of the grid. For example a 2d-regular grid, or *Cartesian grid*, $\mathcal{G}_{CG}(n,d)$ of n^d nodes, in which each node is positioned on the integer coordinates of a d-dimensional space and connected to its left and right neighbor in each dimension, leads to the Manhattan distance $||x - y||_1 = \sum_i |x_i - y_i|$:

$$\Delta_{p}\left(\frac{1}{n}\mathcal{G}_{CG}(n,d), ([0,1]^{d}, ||x-y||_{1}, \mathcal{U}_{[0,1]^{d}})\right) \leq \frac{d}{n},$$
(10.16)

and the *rescaled* grid graph $\frac{1}{n}\mathcal{G}_{CG}(n,d) = (\mathcal{G}_{CG}(n,d), \frac{1}{n}d_{SP}, \mathcal{U}(\mathcal{V}))$, where d_{SP} is the shortest-path distance and $\mathcal{U}(\mathcal{V})$ is the uniform distribution over the nodes of $\mathcal{G}_{CG}(n,d)$, converges to $[0,1]^d$ equipped with the Manhattan distance and uniform distribution.

10.3.2 Totally connected graph

A totally connected graph $\mathcal{T}_n = (\mathcal{V}, \mathcal{E})$ is a graph for which each pair of nodes is connected, i.e. $\mathcal{E} = \mathcal{V}^2$. In this network, the shortest-path distance between each node is exactly 1, and $d_{ij} = \mathbb{1}\{i \neq j\}$. Strictly speaking, totally connected graphs do not converge since, if they did, they would converge to $([0,1],\mathbb{1}\{x \neq y\},\mathcal{U}_{[0,1]})$, which is not a separable metric space. This is actually a simple proof to show that \mathbb{X}_p is not complete, as is can be shown that \mathcal{T}_n is a Cauchy sequence (Sturm, 2013).

However, the notion of Gromov-Wasserstein distance can be extended to *gauged metric spaces* in which \mathcal{T}_n converges (see Sturm (2013)). In our formalism, we can observe that, if $d_{[0,1]}(x,y) = \mathbb{1}\{x \neq y\}$, then

$$\Delta_{p}(d_{\mathcal{T}_{n}}, d_{[0,1]}) = \frac{1}{n^{1/p}},$$
(10.17)

and all the continuity results of the next chapter are applicable. In a *loose* sense, totally connected networks tend to infinite spaces in which each element is at the same distance from any other element. Note also that $d_{\mathcal{T}_n}$ converges to $d_{[0,1]}$ if and only if $p < +\infty$, and this is an example for which the Gromov-Hausdorff distance does not converge (Gromov, 1999), while the Gromov-Wasserstein distance does. This fact was expected since a classical result of the Gromov-Hausdorff distance is that it is a complete metric (Gromov, 1999), and thus the convergence w.r.t. Gromov-Hausdorff would imply the separability of the metric space [0,1] equipped with $d_{[0,1]}$, which yields a contradiction.

10.3.3 Random sampling from an mm-space

When dealing with processes taking place in a continuous space, sampling is a very natural choice, and data processing will usually start by selecting a set of *query points* in the desired space. The assumption is that the query points are sufficiently many and properly spaced so that they represent the continuous space with sufficient accuracy. For example, a water-distribution company may monitor its distribution network by placing several water quality sensors, and the interesting questions that arise are: do these sensors give a truthful view of the water quality in the *whole* network? Is prediction possible using these sensors? Another example is that of a social network in which a company may have gathered information about a subgroup of users (e.g. the age of the few clients that filled an online form), and would like to infer data on its whole customer base.

We will thus consider the convergence of samples of an mm-space, and will focus on the random sampling method. This sampling consists in drawing *n* i.i.d. samples $(X_i)_{i \leq n}$ from an mm-space \mathcal{X} according to its probability measure $\mu_{\mathcal{X}}$. The next theorem shows that this sampled discrete space converges to the original mm-space.

Theorem 10.6. Let $\mathcal{X} \in \mathbb{X}_p$ be an mm-space, and $(X_i)_{i \in \mathbb{N}}$ a sequence of i.i.d. random variables with probability distribution $\mu_{\mathcal{X}}$. Then, if \mathcal{X}_n is a (random) discrete mm-space of n elements, uniform measure and metric $d_{\mathcal{X}_n}(i,j) = d_{\mathcal{X}}(X_i, X_j)$, then:

$$\Delta_p(\mathcal{X}_n, \mathcal{X}) \le 2W_p(\hat{\mu}_n, \mu_{\mathcal{X}}), \tag{10.18}$$

where $W_p(\mu, \nu)$ is the Wasserstein distance between the two measures μ and ν , and $\hat{\mu}_n = \frac{1}{n} \sum_i \delta_{X_i}$ is the empirical distribution of the samples $(X_i)_{i \in \mathbb{N}}$ in \mathcal{X} . As a result,

$$\Delta_p(\mathcal{X}_n,\mathcal{X}) \to 0 \text{ a.s.},$$

10.3. LIMITS OF SEVERAL POPULAR GRAPHS

and \mathcal{X}_n converges to \mathcal{X} almost surely.

Hence, the continuity theorems of the next chapter apply to random sampling. This implies that, intuitively, sampled data have a similar structure as the original sampled space. Note that other sampling methods can be used, for example following a grid structure, and the comparison of the convergence rates of each sampling method would be an interesting work for the future. Moreover, a direct consequence of Theorem 10.6 is that the set of discrete mm-spaces is dense in X_p , and a natural question that arises is whether the convergence rate of random sampling is optimal, i.e. if there exists a sequence of discrete spaces X_n of size n converging to a continuous space X faster than $W_p(\hat{\mu}_n, \mu_X)$.

10.3.4 Geometric random networks

Geometric random graphs are designed to model contact networks, in which edges represent nodes closer than a certain threshold in a certain metric space. A formal definition is given in Model 5.4. These graphs can be seen as the unweighted equivalents of random sampling, and, under technical assumptions, the shortest-path distance in these graphs tends to the distance in the original space (under proper renormalization by the radius r_n). We first define the concept of *geodesic mm-space*.

Definition 10.10 (Geodesic mm-space). A geodesic mm-space \mathcal{X} is an mm-space such that, for each pair of points $(x, y) \in \mathcal{X}^2$, there exists a continuous function $\gamma : [0,1] \to \mathcal{X}$ such that $\gamma(0) = x$, $\gamma(1) = y$, and $d_{\mathcal{X}}(\gamma(a), \gamma(b)) = d_{\mathcal{X}}(x, y)|b-a|$ for all $a, b \in [0,1]$.

Informally, a geodesic mm-space is a *connected* mm-space such that shortest paths exist between any pair of points. In order for geometric random networks to converge, an additional constraint is required: that the volume of balls of fixed radius are lower bounded. More specifically, we will assume that the minimum volume of balls of radius ϵ , $B(\epsilon) = \min_{x \in \mathcal{X}} \mathbb{P}_{\mu_{\mathcal{X}}}(d_{\mathcal{X}}(x, X) \leq \epsilon)$, are positive for any $\epsilon > 0$. This will imply the existence of an inverse, denoted as $B^{-1}(x) = \max\{\epsilon > 0 : B(\epsilon) \leq x\}$. Note that the function $B(\epsilon)$ is closely related to the dimension of the space, and will typically have the form $B(\epsilon) = C\epsilon^d$ for spaces of dimension d, and hence $B^{-1}(x) = C'x^{1/d}$.

Theorem 10.7. Let $\mathcal{X} \in \mathbb{X}_p$ be a geodesic mm-space, and $G_{GR}(\mathcal{X}, n, r_n)$ a sequence of geometric random graphs equipped with the shortest-path distance. If $\lim_{n \to +\infty} r_n = 0$ and $\lim_{n \to +\infty} \frac{r_n}{B^{-1}\left(\frac{4\ln n - \ln B^{-1}(\frac{4\ln n}{n-2})}{n-2}\right)} = +\infty$, then $r_n G_{GR}(\mathcal{X}, n, r_n)$ converges to \mathcal{X} a.s.

Hence, the continuity theorems of the next chapter are, again, applicable. More specifically, this result implies the convergence of epidemics on contact networks to their continuous equivalent on \mathbb{R}^2 equipped with the population's density as probability measure. The technical constraint $\lim_{n\to+\infty} r_n = 0$ requires that the size of neighborhoods tends to zero, while the second constraint imposes that each node is connected to a large number of neighbors (at least $4 \ln n - \ln B^{-1}(\frac{4 \ln n}{n-2})$), in order to ensure that the network is connected with a high probability.

10.3.5 Erdös-Rényi random networks and the stochastic block model

Erdös-Rényi random graphs G(n, p) (see Model 5.1) are dense when p > 0, and of average degree np. Hence, every node is connected to a fraction p of the whole network and,

with high probability, the distance between two different nodes is either 1 if they are neighbors, or 2 otherwise. Unfortunately, Erdös-Rényi random graphs do not converge w.r.t. Δ_p . A lesser statement could be that their adjacency matrices converge w.r.t. the mapping distance defined in Definition 11.3, although this is also false. An analysis of this very interesting example would require a chapter of its own and is left for future work, although a few insights can be provided for it.

As stated in Sec. 11.1, the definition of the mapping distance for operators generalizes the distance induced by the operator norm, and convergence with respect to this norm is among the strongest notion of convergence for operators. A weaker notion of convergence is called the *strong convergence* (as opposed to the *weak convergence*, a yet weaker form of convergence for operators), and is defined as the L_p convergence of the mappings $f \mapsto$ F(f) for each function $f \in L_{p,\mu_X}$. A detailed analysis of dense Erdös-Rényi random graphs shows that their adjacency matrices *converge strongly* to the operator in [0, 1] defined as:

$$\forall x \in \mathcal{X}, \qquad [\mathcal{O}_p(f)](x) = p\mathbb{E}[f(X)], \tag{10.19}$$

where X is a uniform random variable in [0,1]. However, the adjacency matrices do not converge to \mathcal{O}_p in the mapping distance sense, and thus Erdös-Rényi random networks cannot converge w.r.t. Δ_p . However, early results seem to indicate that strong convergence still yields interesting results including the convergence of dynamic processes and a partial result for spectral properties, and the rigorous analysis of the properties of strong convergence would be a notable addition to this work.

10.4 Discussion and open problems

The subject of network convergence is extremely vast, and work is still needed to investigate the theoretical properties and practical advantages of the notions developed herein. We now highlight several open problems and interesting directions for future research:

- First, Erdös-Rényi random graphs do not converge in the Gromov-Wasserstein sense, and relaxing the notion of convergence (by extending the notion of strong convergence of operators instead of the operator norm as in Sec. 11.1.1) may be sufficient to allow the convergence of this class of random graphs while being strong enough to provide continuity theorems similar to that of Chap. 11.
- In addition, proving the convergence of the Stochastic Block Model (Holland et al., 1983) in such a relaxed notion of convergence would be of high practical interest, and may improve our understanding of this random graph model.
- The precise convergence rate of random sampling remains unknown. More generally, one may wonder if random sampling is an optimal approximation strategy. Are there better ways to approximate the structure of a continuous space using a discrete space?
- Efficient algorithms to compute the Gromov-Wasserstein distance are necessary in
 order to use it in practical applications. Also, an experimental analysis would help
 to draw lines between our work and the existing literature of network convergence.

- Approximation is central in computer science due to the fact that any simulation
 performed by a computer is necessarily discrete in nature. Hence, providing good
 inference algorithms for the structure of a space may provide good approximations of
 large relational datasets such as social networks and improve the efficiency and speed
 of computation in these high dimensional objects.
- Finally, our analysis may provide interesting insights for the *clustering* problem, in the sense that the network made by aggregating each cluster into a single node should be a good approximation of the original network in terms of distances.

10.5 Proofs

In all the following proofs, and unless specified otherwise, $(X, Y) \sim \pi$ and $(X', Y') \sim \pi$ are independent pairs of random variables, each of them drawn according to the coupling $\pi \in \Pi(\mu_X, \mu_Y)$ between the two mm-spaces \mathcal{X} and \mathcal{Y} .

10.5.1 Proof of the convergence of product mm-spaces

Proof of Theorem 10.3. For all quadruples $(x, x', y, y') \in \mathcal{X}^2 \times \mathcal{Y}^2$,

$$\begin{aligned} |d_{\mathcal{X}}(x,x') - d_{\mathcal{X}\times\mathcal{Y}}((x,y),(x',y'))| &= |d_{\mathcal{X}\times\mathcal{Y}}((x,y),(x',y)) - d_{\mathcal{X}\times\mathcal{Y}}((x,y),(x',y'))| \\ &\leq d_{\mathcal{X}\times\mathcal{Y}}((x',y),(x',y')) \\ &= d_{\mathcal{Y}}(y,y'). \end{aligned}$$
(10.20)

Hence, using the coupling $d\pi(x',(x,y)) = \delta_{\{x'=x\}} d\mu_{\mathcal{X} \times \mathcal{Y}}(x,y) \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{X} \times \mathcal{Y}})$, we get that

$$\begin{split} \Delta p(\mathcal{X}, \mathcal{X} \times \mathcal{Y}) &\leq ||d_{\mathcal{X}} - d_{\mathcal{X} \times \mathcal{Y}}||_{p, \mu_{\mathcal{X} \times \mathcal{Y}} \otimes \mu_{\mathcal{X} \times \mathcal{Y}}} \\ &\leq ||d_{\mathcal{Y}}||_{p, \mu_{\mathcal{Y}} \otimes \mu_{\mathcal{Y}}} \\ &= \operatorname{size}_{p}(\mathcal{Y}). \end{split}$$
(10.21)

10.5.2 Proof of the density of undirected graphs

Proof of Theorem 10.5. First, note that weighted graphs (with uniform measure and shortest-path distance) are dense due to the convergence of random sampling on any mm-space (see Theorem 10.6). Thus, it is sufficient to show that undirected graphs are dense in the set of weighted graphs.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, w)$ be a weighted graph of n nodes. In order to generate a sequence of undirected graphs converging to \mathcal{G} , we will use two types of structures: cliques and chains. More specifically, for each k > 0, we create an undirected graph $\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k)$ by associating each node in \mathcal{V} to a clique (i.e. a totally connected subgraph) of k^2 nodes, and each edge $(i, j) \in \mathcal{E}$ to a chain of size $d_{ij}^k = \lfloor kd_{SP}(i, j) \rfloor - 1$ (in order to impose the correct distances between the cliques), as presented in Fig. 10.6. Since the cliques are substantially larger than the chains, the uniform measure over the nodes of \mathcal{G}_k will give an almost-zero



Figure 10.6: Example of a weighted graph (left) and its undirected approximation (right).

weight to the chains, thus providing a good approximation of the original graph \mathcal{G} . The exact definitions of \mathcal{V}_k and \mathcal{E}_k are given by:

$$\begin{aligned} \mathcal{V}_{k} &= \bigcup_{i \in \mathcal{V}} \{ v_{i,1}, ..., v_{i,k^{2}} \} \bigcup \cup_{(i,j) \in \mathcal{E}} \{ v_{ij,1}, ..., v_{ij,d_{ij}^{k}} \} \end{aligned}$$
(10.22)
$$\begin{aligned} \mathcal{E}_{k} &= \bigcup_{i \in \mathcal{V}} \{ (v_{i,l}, v_{i,m}) \, : \, 1 \leq l < m \leq k^{2} \} \\ & \bigcup \cup_{(i,j) \in \mathcal{E}} \{ (v_{ij,l}, v_{ij,l+1}) \, : \, 1 \leq l < d_{ij}^{k} \} \\ & \bigcup \cup_{(i,j) \in \mathcal{E}} \{ (v_{ij,l}, v_{i,l}) \, : \, 1 \leq l \leq k^{2} \} \\ & \bigcup \cup_{(i,j) \in \mathcal{E}} \{ (v_{ij,d_{ij}^{k}}, v_{j,l}) \, : \, 1 \leq l \leq k^{2} \}. \end{aligned}$$
(10.23)

By construction, each node in clique *i* is at distance $d_{ij}^k + 1$ from any node in clique *j*. Thus, by coupling each node of a clique $v_{i,l}$ to its original node v_i and each node of a chain to the uniform measure over \mathcal{V} , we obtain the following bound on $\Delta_p(\mathcal{G}, \frac{1}{k}\mathcal{G}_k)$:

$$\begin{split} \mathbf{\Delta}_{p}(\mathcal{G}, \frac{1}{k}\mathcal{G}_{k})^{p} &\leq \sum_{ij} \frac{k^{4}}{\operatorname{card}(\mathcal{V}_{k})^{2}} \left| \frac{d_{ij}^{k} + 1}{k} - d_{SP}(i,j) \right|^{p} \\ &+ \frac{\operatorname{card}(\mathcal{V}_{k})^{2} - n^{2}k^{4}}{\operatorname{card}(\mathcal{V}_{k})^{2}} \left(\operatorname{size}_{\infty}(\mathcal{G}_{k}) + \operatorname{size}_{\infty}(\mathcal{G})\right)^{p} \\ &\leq \frac{n^{2}k^{4}}{\operatorname{card}(\mathcal{V}_{k})^{2}} \frac{1}{k^{p}} + \frac{\operatorname{card}(\mathcal{V}_{k})^{2} - n^{2}k^{4}}{\operatorname{card}(\mathcal{V}_{k})^{2}} \left(3d^{*} + d^{*}\right)^{p} \\ &\leq \frac{1}{k} + \frac{(nk^{2} + \sum_{ij}d_{ij}^{k})^{2} - n^{2}k^{4}}{(nk^{2} + \sum_{ij}d_{ij}^{k})^{2}} \left(4d^{*}\right)^{p} \\ &\leq \frac{1}{k} + \frac{n^{3}k^{2}d^{*} + n^{4}d^{*2}}{n^{2}k^{4}} \left(4d^{*}\right)^{p} \\ &= O\left(\frac{1}{k}\right), \end{split}$$
(10.24)

where $d^* = \max_{ij} d_{SP}(i, j)$ is the diameter of \mathcal{G} , and thus $\lim_{k \to +\infty} \Delta p(\mathcal{G}, \frac{1}{k}\mathcal{G}_k) = 0$ which leads to the desired density result.

10.5.3 Proofs of the convergence of popular graphs

Proof of Theorem 10.6. Let $(I, X) \sim \pi$ and $(I', X') \sim \pi$ be independent couplings between \mathcal{X}_n and \mathcal{X} . Since, for any quadruple $(x, x', y, y') \in \mathcal{X}^4$, $|d_{\mathcal{X}}(x, x') - d_{\mathcal{X}}(y, y')| \leq d_{\mathcal{X}}(x, y) + d_{\mathcal{X}}(x', y')$, we have

$$\begin{split} \mathbf{\Delta}_{p}(\mathcal{X}_{n},\mathcal{X}) &= \inf_{\pi \in \Pi(\mu_{\mathcal{X}_{n}},\mu_{\mathcal{X}})} \mathbb{E}_{\pi \otimes \pi} [|d_{\mathcal{X}}(X_{I},X_{I'}) - d_{\mathcal{X}}(X,X')|^{p}]^{1/p} \\ &\leq \inf_{\pi \in \Pi(\mu_{\mathcal{X}_{n}},\mu_{\mathcal{X}})} \mathbb{E}_{\pi \otimes \pi} [(d_{\mathcal{X}}(X_{I},X) + d_{\mathcal{X}}(X_{I'},X'))^{p}]^{1/p} \\ &\leq 2\inf_{\pi \in \Pi(\mu_{\mathcal{X}_{n}},\mu_{\mathcal{X}})} \mathbb{E}_{\pi} [d_{\mathcal{X}}(X_{I},X)^{p}]^{1/p} \\ &= 2W_{p}(\hat{\mu}_{n},\mu_{\mathcal{X}}), \end{split}$$
(10.25)

where $\hat{\mu}_n = \frac{1}{n} \sum_i \delta_{X_i}$ is the empirical distribution of the samples $(X_i)_{i \in \mathbb{N}}$ in \mathcal{X} (note that the (X_i) are considered fixed in the above equation). Finally, the Glivenko-Cantelli theorem states that the empirical distribution $\mu_{\mathcal{X}_n}$ converges weakly to $\mu_{\mathcal{X}}$ as n tends to infinity, which implies the convergence of the Wasserstein distance to zero (Villani, 2009, Theorem 6.8). For precise convergence rates in \mathbb{R}^d , we refer the reader to Horowitz and Karandikar (1994) and Fournier and Guillin (2015).

Proof of Theorem 10.7. In order to prove the almost-sure convergence of geometric random networks, we will approximate every geodesic by paths in the geometric random graph and show that, with high probability, such approximations exist using the following lemma.

Lemma 10.3. Let $G_{GR}(\mathcal{X}, n, r)$ be a geometric random network on the mm-space $\mathcal{X}, (X_i)_{i \in [|n|]}$ its associated set of points in $\mathcal{X}, d_{SP}(i, j)$ the shortest-path distance in $G_{GR}(\mathcal{X}, n, r)$, and $\epsilon \in (0, r)$. Then, with probability at least $1 - \frac{n^2 \operatorname{size}_1(\mathcal{X})}{r - \epsilon} \exp(-(n - 2)B(\epsilon))$, the random graph $G_{GR}(\mathcal{X}, n, r)$ is connected and, for any pair of nodes $(i, j) \in [|n|]^2$,

$$\left| rd_{SP}(i,j) - d_{\mathcal{X}}(X_i, X_j) \right| \le r + \frac{\epsilon}{r - \epsilon} d_{\mathcal{X}}(X_i, X_j).$$
(10.26)

Proof. For each pair of nodes (i, j), let $\gamma_{ij} : [0, 1] \to \mathcal{X}$ be the geodesic from X_i to X_j . Then, let $K_{ij} = \left\lceil \frac{d_{\mathcal{X}}(X_i, X_j)}{r - \epsilon} \right\rceil$ and, for $k \in \{0, ..., K_{ij}\}$, $\gamma_{ij,k} = \gamma_{ij}(\frac{k}{K_{ij}})$ is the approximation of the geodesic as a chain of K_{ij} points. By definition, we have $d_{\mathcal{X}}(\gamma_{ij,k}, \gamma_{ij,k+1}) = \frac{d_{\mathcal{X}}(X_i, X_j)}{K_{ij}} \leq r - \epsilon$, and, if there exists nodes of the graph that are at distance ϵ from each $\gamma_{ij,k}$, then these nodes will be connected and form a chain between i and j in $G_{GR}(\mathcal{X}, n, r)$. This happens with probability

$$\mathbb{P}(\forall (i,j) \in [|n|]^{2}, \forall k \in \{1, ..., K_{ij} - 1\}, \exists l \in [|n|] \text{ s.t. } d_{\mathcal{X}}(X_{l}, \gamma_{ij,k}) \leq \epsilon) \\
\geq 1 - \sum_{i,j} \mathbb{P}(\exists k \in \{1, ..., K_{ij} - 1\} \text{ s.t. } \forall l \in [|n|], d_{\mathcal{X}}(X_{l}, \gamma_{ij,k}) > \epsilon) \\
\geq 1 - n^{2} \mathbb{E}[\sum_{k=1}^{K_{ij} - 1} \mathbb{P}(\forall l \notin \{i, j\}, d_{\mathcal{X}}(X_{l}, \gamma_{ij,k}) > \epsilon \mid X_{i}, X_{j})] \\
\geq 1 - n^{2} \mathbb{E}[(K_{ij} - 1)(1 - B(\epsilon))^{n-2}] \\
\geq 1 - \frac{n^{2} \text{size}_{1}(\mathcal{X})}{r - \epsilon} (1 - B(\epsilon))^{n-2} \\
\geq 1 - \frac{n^{2} \text{size}_{1}(\mathcal{X})}{r - \epsilon} \exp(-(n - 2)B(\epsilon)).$$
(10.27)

Hence, with probability at least $1 - \frac{n^2 \operatorname{size}_1(\mathcal{X})}{r-\epsilon} \exp\left(-(n-2)B(\epsilon)\right)$, we have

$$d_{SP}(i,j) \le K_{ij} \le \frac{d_{\mathcal{X}}(X_i, X_j)}{r - \epsilon} + 1,$$
(10.28)

and, if $i_0, ..., i_{d_{SP}(i,j)}$ is a shortest-path from *i* to *j*,

$$d_{\mathcal{X}}(X_{i}, X_{j}) \leq \sum_{k=0}^{d_{SP}(i,j)-1} d_{\mathcal{X}}(X_{i_{k}}, X_{i_{k+1}}) \leq rd_{SP}(i,j).$$
(10.29)

Hence, with probability at least $1 - \frac{n^2 \operatorname{size}_1(\mathcal{X})}{r-\epsilon} \exp\left(-(n-2)B(\epsilon)\right)$, we obtain the desired result:

$$\left| rd_{SP}(i,j) - d_{\mathcal{X}}(X_i, X_j) \right| \le r + \frac{\epsilon}{r - \epsilon} d_{\mathcal{X}}(X_i, X_j).$$

$$(10.30)$$

Let
$$\epsilon_n = B^{-1}\left(\frac{4\ln n - \ln B^{-1}(\frac{4\ln n}{n-2})}{n-2}\right)$$
. By assumption, $\epsilon_n = o(r_n)$ and, for *n* sufficiently large, $B^{-1}(\frac{4\ln n}{n-2}) < 1$ and thus $\epsilon_n \ge B^{-1}\left(\frac{4\ln n}{n-2}\right)$. Again, for *n* sufficiently large, $\epsilon_n < r_n/2$ and the probability in Lemma 10.3 is lower bounded by

$$1 - \frac{n^{2}\operatorname{size}_{1}(\mathcal{X})}{r_{n} - \epsilon_{n}} \exp\left(-(n-2)B(\epsilon_{n})\right)$$

$$\geq 1 - \frac{2n^{2}\operatorname{size}_{1}(\mathcal{X})}{\epsilon_{n}} \exp\left(-(n-2)B(\epsilon_{n})\right)$$

$$\geq 1 - \exp\left(\ln(2\operatorname{size}_{1}(\mathcal{X})) + 2\ln n - \ln \epsilon_{n} - (n-2)B(\epsilon_{n})\right)$$

$$\geq 1 - \exp\left(\ln(2\operatorname{size}_{1}(\mathcal{X})) + 2\ln n - \ln \epsilon_{n} - 4\ln n + \ln B^{-1}(\frac{4\ln n}{n-2})\right)$$

$$\geq 1 - \exp\left(\ln(2\operatorname{size}_{1}(\mathcal{X})) - 2\ln n\right)$$

$$= 1 - \frac{2\operatorname{size}_{1}(\mathcal{X})}{n^{2}}.$$
(10.31)

Hence, with probability at least $1 - \frac{2size_1(\mathcal{X})}{n^2}$, we have

$$\begin{split} \mathbf{\Delta}_{p}(\mathcal{X}, r_{n}G_{GR}(\mathcal{X}, n, r_{n})) &\leq \mathbf{\Delta}_{p}(\mathcal{X}, \mathcal{X}_{n}) + \mathbf{\Delta}_{p}(\mathcal{X}_{n}, r_{n}G_{GR}(\mathcal{X}, n, r_{n})) \\ &\leq 2W_{p}(\hat{\mu}_{n}, \mu_{\mathcal{X}}) + \mathbf{\Delta}_{p}(\mathcal{X}_{n}, r_{n}G_{GR}(\mathcal{X}, n, r_{n})) \\ &\leq 2W_{p}(\hat{\mu}_{n}, \mu_{\mathcal{X}}) + r_{n} + \frac{\epsilon_{n}}{r_{n} - \epsilon_{n}} \operatorname{size}_{p}(\mathcal{X}_{n}) \\ &\leq 2W_{p}(\hat{\mu}_{n}, \mu_{\mathcal{X}}) + r_{n} + \frac{2\epsilon_{n}}{r_{n}} \left(\operatorname{size}_{p}(\mathcal{X}) + W_{p}(\hat{\mu}_{n}, \mu_{\mathcal{X}})\right), \end{split}$$

$$(10.32)$$

where \mathcal{X}_n is the sampling space defined in Theorem 10.7 and $\hat{\mu}_n$ is the empirical measure of the $(X_i)_{i \in [|n|]}$. The second inequality is a direct application of Theorem 10.7, and the third inequality is deduced from Lemma 10.3. Since $\frac{2\text{size}_1(\mathcal{X})}{n^2}$ is summable, Borel-Cantelli's lemma implies that, with probability equal to one, only a finite number of values of $\Delta_p(\mathcal{X}, r_n G_{GR}(\mathcal{X}, n, r_n))$ will be above the upper bound of Eq. 10.32, and since this upper bound converges to 0 almost surely (see Theorem 10.7), this implies that

$$\Delta_p(\mathcal{X}, r_n G_{GR}(\mathcal{X}, n, r_n)) \to 0 \text{ a.s.}$$

Continuity of key network characteristics and processes

"Success is not final, failure is not fatal: it is the courage to continue that counts." — Winston Churchill

11

Contents

11.1	Operators of an mm-space 167
	11.1.1 Mapping distance for operators
	11.1.2 Linear operators
11.2	Convergence of network characteristics
	11.2.1 Average degree
	11.2.2 Degree distribution
	11.2.3 Spectral radius
	11.2.4 Diameter
11.3	Spectral properties of compact self-adjoint operators 174
	11.3.1 Convergence of spectrum
	11.3.2 Convergence of eigenspaces
11.4	Convergence of dynamic processes
	11.4.1 Discrete dynamical systems
	11.4.2 Random walks
	11.4.3 Differential systems
11.5	Proofs
	11.5.1 Proofs of the convergence of operators
	11.5.2 Proofs of the convergence of spectral properties
	11.5.3 Proofs of the convergence of dynamic processes

11.1 Operators of an mm-space

Matrix theory is a key ingredient of linear algebra and functional analysis, and matrices are used to define multivariate stochastic processes, structural properties of networks, or

perform signal analysis in graphs. In this chapter, we denote as operator a *function over the space of functions*.

Definition 11.1 (Operator). An operator over the mm-space \mathcal{X} is a function $F : L_{p,\mu_{\mathcal{X}}} \to L_{p,\mu_{\mathcal{X}}}$, i.e. an endomorphism of the space of $L_{p,\mu_{\mathcal{X}}}$ functions on \mathcal{X} .

In the next sections, we will show a number of continuity results for such operators using an extension of the operator norm to situations in which the two operators lie in two different mm-spaces. Also note that proofs are provided in Sec. 11.5.

11.1.1 Mapping distance for operators

The notion of distance presented in Sec. 10.2.4 can be extended to operators $F : L_{p,\mu_{\mathcal{X}}} \to L_{p,\mu_{\mathcal{X}}}$ and $G : L_{p,\mu_{\mathcal{Y}}} \to L_{p,\mu_{\mathcal{Y}}}$. In order to compare two operators F and G, one would like to compare F(f) and G(f), with the same function f. However, this is not possible since F and G do not lie in the same space, and this difficulty is avoided using *projections* of f on \mathcal{X} and \mathcal{Y} , respectively. We now provide a formal definition for such a projection.

Definition 11.2 (Projection). Let $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a function, and $\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ a coupling. The projection of f on \mathcal{X} (resp. \mathcal{Y}), denoted $f|_{\mathcal{X}}^{\pi}$ (resp. $f|_{\mathcal{Y}}^{\pi}$), is defined as:

$$f|_{\mathcal{X}}^{\pi}(x) = \mathbb{E}_{\pi}[f(X,Y) \mid X = x];$$

$$f|_{\mathcal{Y}}^{\pi}(y) = \mathbb{E}_{\pi}[f(X,Y) \mid Y = y].$$
(11.1)

The mapping distance between *F* and *G* is then simply the maximum of the distance between $F(f|_{\mathcal{X}}^{\pi})$ and $G(f|_{\mathcal{Y}}^{\pi})$ over the space of functions on $\mathcal{X} \times \mathcal{Y}$.

Definition 11.3 (Mapping distance between operators). Let \mathcal{X} and \mathcal{Y} be two mmspaces, $F : L_{p,\mu_{\mathcal{X}}} \to L_{p,\mu_{\mathcal{X}}}$ and $G : L_{p,\mu_{\mathcal{Y}}} \to L_{p,\mu_{\mathcal{Y}}}$ two operators on \mathcal{X} and \mathcal{Y} respectively, and $\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ a coupling. Then, the (p, π) -mapping distance between F and G is the maximum over all $L_{p,\pi}$ functions of $\mathcal{X} \times \mathcal{Y}$ of the following quantity:

$$\Delta_{p,\pi}(F,G) = \max_{f \in L_{p,\pi}: ||f||_{p,\pi} \neq 0} \frac{\Delta_{p,\pi}(F(f|_{\mathcal{X}}^{\pi}), G(f|_{\mathcal{Y}}^{\pi}))}{||f||_{p,\pi}},$$
(11.2)

where $f|_{\mathcal{X}}^{\pi}$ and $f|_{\mathcal{Y}}^{\pi}$ are projections of f on \mathcal{X} and \mathcal{Y} , respectively (see Eq. 11.1). Then, the p-mapping distance between F and G is the minimum of $\Delta_{p,\pi}(F,G)$ over all possible couplings:

$$\Delta_{p}(F,G) = \min_{\pi \in \Pi(\mu_{\mathcal{X}},\mu_{\mathcal{Y}})} \Delta_{p,\pi}(F,G).$$
(11.3)

Remark 11.1. The operator norm $|||F||| = \max_{f \in L_2 : ||f||_{2} \neq 0} ||F(f)||_2 / ||f||_2$ can also be rewritten as $|||F||| = \Delta_2(F, \mathbf{0})$, where **0** is the constant operator equal to 0. Also, if F and G are matrices of size n, then $\Delta_{2,I_n}(F,G) = |||F - G|||$ where I_n is a coupling s.t. $dI_n(x,y) = \frac{1}{n} \mathbb{1}\{x = y\}$.

We will see in the following sections that this distance between operators provides new and interesting results for many network related characteristics including the spectrum of an operator and the evolution of a dynamic process. In order to relate $\Delta_p(F,G)$ to the distance between their respective spaces, we define a natural *Lipschitz-like* constraint on operators. **Definition 11.4** (Lipschitz-like inequality). Two operators F and G on X and Y, respectively, are said to satisfy a Lipschitz-like inequality if there exists L > 0 such that:

$$\Delta_p(F,G) \le L\Delta_p(\mathcal{X},\mathcal{Y}). \tag{11.4}$$

When such an inequality is verified, L is called the Lipschitz constant of F and G.

11.1.2 Linear operators

The most simple linear operator is the *identity operator*, denoted Id_X , or simply Id when the underlying space is clear, and defined by

$$Id_{\mathcal{X}}(f) = f.$$

Although $\Delta_p(Id_{\mathcal{X}}, Id_{\mathcal{Y}}) \geq 2$, this simple operator is not, in general, continuous w.r.t. $\Delta_p(\mathcal{X}, \mathcal{Y})$. For example, with $\mathcal{X} = \{0\}$ and $\mathcal{Y} = [0, \epsilon]$, one can show that $\Delta_p(\mathcal{X}, \mathcal{Y}) \leq \epsilon$ for all $p \geq 1$, although $\Delta_p(Id_{\mathcal{X}}, Id_{\mathcal{Y}}) \geq 1/2$ (by taking $f = \mathbb{1}\{y \in [0, \epsilon/2]\}$). However, when the spaces are homeomorphic, and π is a deterministic coupling (i.e. a bijection from \mathcal{X} to \mathcal{Y}), then $\Delta_p(Id_{\mathcal{X}}, Id_{\mathcal{Y}}) = 0$.

Lemma 11.1. If X and Y are two mm-spaces and $\pi \in \Pi(\mu_X, \mu_Y)$ is a deterministic coupling, then

$$\Delta_{p,\pi}(Id_{\mathcal{X}}, Id_{\mathcal{Y}}) = 0.$$
(11.5)

Proof. Let $(X, Y) \sim \pi$ be random variables of coupled distribution π . Since π is deterministic, there exists a bijection $\psi : \mathcal{X} \to \mathcal{Y}$ such that $Y = \psi(X)$. Then $f(X,Y) = f(X,\psi(X)) = f(\psi^{-1}(Y),Y)$ is $\mu_{\mathcal{X}}$ -measurable and $\mu_{\mathcal{Y}}$ -measurable. Hence $f|_{\mathcal{X}}^{\pi}(X) = f|_{\mathcal{Y}}^{\pi}(Y) = f(X,Y)$ and $\Delta_{p,\pi}(Id_{\mathcal{X}}, Id_{\mathcal{Y}}) = \max_{f} \frac{||f|_{\mathcal{X}}^{\pi} - f|_{\mathcal{Y}}^{\pi}||_{p,\pi}}{||f||_{p,\pi}} = 0.$

For the convergence of graphs to continuous spaces, such a bijection is not possible since the spaces do not have the same cardinality. However, the graph \mathcal{G} can be replaced by the continuous space [0,1] equipped with the proper pseudo-distance such that $\Delta_p(\mathcal{G},[0,1]) = 0$. This trick will allow us to provide results on characteristics that are invariant w.r.t. the equivalence relation $\mathcal{X} \sim \mathcal{Y} \Leftrightarrow \Delta_p(\mathcal{X}, \mathcal{Y}) = 0$. As we can see, the quantity $\Delta_p(F,G)$ can be hard to evaluate, even for operators as simple as the identity. However, a large class of linear operators, called integral operators (or Hilbert-Schmidt operators), are more regular, and allow upper bounds on the mapping distance.

Theorem 11.1 (Integral operators). Let $p \ge 1$, $q = \max\{p, \frac{p}{p-1}\}$, \mathcal{X} an mm-space, and $k_{\mathcal{X}} \in L_{q,\mu_{\mathcal{X}} \otimes \mu_{\mathcal{X}}}$ a kernel function. The integral operator $F_{k_{\mathcal{X}}} : L_{p,\mu_{\mathcal{X}}} \to L_{p,\mu_{\mathcal{X}}}$ is the linear operator defined by: $\forall f \in L_{p,\mu_{\mathcal{X}}}, \forall x \in \mathcal{X}$,

$$[F_{k_{\mathcal{X}}}(f)](x) = \int k_{\mathcal{X}}(x, x') f(x') d\mu_{\mathcal{X}}(x').$$
(11.6)

Also, if $k_{\mathcal{X}}$ and $k_{\mathcal{Y}}$ are two kernels on, respectively, \mathcal{X} and \mathcal{Y} , the mapping distance between $F_{k_{\mathcal{X}}}$ and $F_{k_{\mathcal{Y}}}$ is upper bounded by the mapping distance between the kernels:

$$\Delta_{p}(F_{k_{\mathcal{X}}},F_{k_{\mathcal{Y}}}) \leq \inf_{\pi \in \Pi(\mu_{\mathcal{X}},\mu_{\mathcal{Y}})} \Delta_{q,\pi \otimes \pi}(k_{\mathcal{X}},k_{\mathcal{Y}}).$$
(11.7)



Figure 11.1: Value of the parameter *q* w.r.t. *p* in Theorem 11.1. For $p \ge 2$, q = p.

Integral operators generalize the notion of matrices to mm-spaces. Since the adjacency matrix is central to network analysis, we now provide a generalization of the adjacency matrix applicable to any mm-space, and show that, when two spaces are close w.r.t. Δ_p , then the respective adjacency operators are also close.

Theorem 11.2 (Adjacency operators). Let $p \ge 1$, $q = \max\{p, \frac{p}{p-1}\}$, \mathcal{X} an mm-space, and $\phi : \mathbb{R}_+ \to \mathbb{R}$ a real function such that $\phi \circ d_{\mathcal{X}} \in L_{q,\mu_{\mathcal{X}} \otimes \mu_{\mathcal{X}}}$. The adjacency operator $A_{\mathcal{X},\phi} : L_{p,\mu_{\mathcal{X}}} \to L_{p,\mu_{\mathcal{X}}}$ is the linear operator defined by: $\forall f \in L_{p,\mu_{\mathcal{X}}}, \forall x \in \mathcal{X}$,

$$[A_{\mathcal{X},\phi}(f)](x) = \int \phi(d_{\mathcal{X}}(x,x'))f(x')d\mu_{\mathcal{X}}(x').$$
(11.8)

Also, if ϕ is L-Lipschitz and $\mathcal{X}, \mathcal{Y} \in \mathbb{X}_q$, $A_{\mathcal{X}, \phi}$ verifies the following Lipschitz-like inequality:

$$\Delta_{p}(A_{\mathcal{X},\phi}, A_{\mathcal{Y},\phi}) \leq L\Delta_{q}(\mathcal{X}, \mathcal{Y}).$$
(11.9)

Proof. Since $\phi \circ d_{\mathcal{X}}$ is a kernel function, Theorem 11.1 implies that, for $\pi \in \operatorname{Opt}_p(\mathcal{X}, \mathcal{Y})$, $\Delta p(A_{\mathcal{X},\phi}, A_{\mathcal{Y},\phi}) \leq \Delta p_{,\pi \otimes \pi}(\phi \circ d_{\mathcal{X}}, \phi \circ d_{\mathcal{Y}}) \leq L \Delta p_{,\pi \otimes \pi}(d_{\mathcal{X}}, d_{\mathcal{Y}}) = L \Delta p(\mathcal{X}, \mathcal{Y})$.

This property ensures the convergence of characteristics of adjacency operators when the underlying space converges w.r.t. Δ_q , where q depends on p as in Fig. 11.1. When the underlying space is discrete (e.g. a graph), adjacency operators become matrices such that each element in the matrix is of the form $\frac{\phi(D_{ij})}{n}$, where D_{ij} is the distance matrix and nthe size of the space, and a particular choice of ϕ returns the adjacency matrix, hence the name of this class of operators.

Lemma 11.2 (Adjacency matrix). Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected and connected graph of size *n* with shortest-path distance d_{SP} . Then, the adjacency matrix $A_{ij} = \mathbf{1}\{\{i, j\} \in \mathcal{E}\}$ is an adjacency operator of the form $A = A_{\mathcal{G}, \phi}$ where ϕ is *n*-Lipschitz.

Proof. Let $\phi(x) = n(x\mathbf{1}\{x < 1\} + (2 - x)\mathbf{1}\{1 \le x < 2\})$. Then ϕ is *n*-Lipschitz and $A = A_{\mathcal{G},\phi}$.

11.2 Convergence of network characteristics

In this section, we show that a number of characteristics commonly used to describe networks are continuous with respect the Gromov-Wasserstein distance.

11.2.1 Average degree

The average degree of a network is the mean volume of balls of radius one, and can be seen as the norm one of the adjacency matrix. In this section, we thus focus on balls of fixed radius, and show that their volume is continuous w.r.t. Δ_p . More specifically, we extend the notion of volume of balls by integrating any function ϕ of the distances of the considered space.

Definition 11.5 (Volume of a ϕ **-ball).** Let $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ be a positive function, $\mathcal{X} \in \mathbb{X}_1$ an mm-space of finite 1-size, and $x \in \mathcal{X}$. The volume of a ball of radius ϕ centered around x, denoted $B_{x,\phi}$, is defined by

$$B_{x,\phi} = \mathbb{E}[\phi(d_{\mathcal{X}}(x,X))] = A_{\mathcal{X},\phi}(\mathbf{1})(x), \qquad (11.10)$$

where $\mathbf{1}(x) = 1$ is a constant function. The mean volume of balls of radius ϕ is then the integration of $B_{x,\phi}$ over the whole space:

$$B_{\mathcal{X},\phi} = \mathbb{E}[\phi(d_{\mathcal{X}}(X, X'))] = ||A_{\mathcal{X},\phi}(\mathbf{1})||_{1}.$$
(11.11)

The case $\phi(x) = \mathbb{1}\{x \le r\}$ gives the standard volume of balls of fixed radius r, and will be denoted $B_{x,r}$ and $B_{\mathcal{X},r}$. In order to prove the continuity of such characteristics, we now provide a more generic which shows the relation between the p-norm of two operators and their Δ_p distance.

Theorem 11.3. Let F and G be two operators of \mathcal{X} and \mathcal{Y} , respectively, $\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ a coupling, $f \in L_{p,\mu_{\mathcal{X}}}$ a function in \mathcal{X} and $f|_{\mathcal{Y}}^{\pi}$ its projection on \mathcal{Y} through the coupling π as in Definition 11.2. Then, if $q \geq p$,

$$\left| ||F(f)||_{p} - ||G(f|_{\mathcal{Y}}^{\pi})||_{p} \right| \leq \Delta_{q,\pi}(F,G)||f||_{q}.$$
(11.12)

Intuitively, if two operators are close w.r.t. $\Delta_{q,\pi}$, then their values are also close, provided that we map the input function from \mathcal{X} to \mathcal{Y} . Applying this result to adjacency operators with p = 1 and q = 2 leads to the approximation of balls of fixed radius.

Corollary 11.1. The following assertions are true:

- If ϕ is L-Lipschitz, then $|B_{\mathcal{X},\phi} B_{\mathcal{Y},\phi}| \leq L\Delta_2(\mathcal{X},\mathcal{Y}).$
- If ϕ is bounded and $\mathcal{X}_n \xrightarrow{\mathbb{A}_2}_{n \to +\infty} \mathcal{X}$ is a converging sequence of mm-spaces, then $\lim_{n \to +\infty} B_{\mathcal{X}_n,\phi} = B_{\mathcal{X},\phi}$.
- If $\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k)$ is a sequence of (undirected) graphs of n_k nodes and E_k edges converging to \mathcal{X} w.r.t. Δ_2 , then:

$$i) \lim_{k \to +\infty} \frac{1}{n_k^2} \sum_{(i,j) \in \mathcal{V}_k^2} \mathbb{1}\{d_{SP}(i,j) \le r\} = B_{\mathcal{X},r},$$

$$ii) \lim_{k \to +\infty} \frac{2E_k}{n_k^2} = B_{\mathcal{X},1}$$

Proof. Since $B_{\mathcal{X},\phi} = ||A_{\mathcal{X},\phi}(\mathbf{1})||_1$, a direct application of Theorem 11.3 implies that:

$$|B_{\mathcal{X},\phi} - B_{\mathcal{Y},\phi}| \le \Delta_2(A_{\mathcal{X},\phi}, A_{\mathcal{Y},\phi})||\mathbf{1}||_2 \le L\Delta_2(\mathcal{X},\mathcal{Y}).$$
(11.13)

Thus, for any Lipshitz function ϕ , $\mathcal{X} \mapsto B_{\mathcal{X},\phi}$ is continuous w.r.t. Δ_2 . However, since Lipschitz functions are dense in bounded functions and $\phi \mapsto B_{\mathcal{X},\phi}$ is a continuous mapping, this continuity result can be extended to bounded functions ϕ . The last two results follow by considering $\phi(x) = \mathbb{1}\{x \leq r\}$ and $\phi(x) = \mathbb{1}\{x = 1\}$.

From Corollary 11.1, we can see that, if a sequence of graphs \mathcal{G}_k converges w.r.t. Δ_2 , then necessarily:

- either the graphs are dense and the mean degree is linear in the number of nodes, i.e. $E_k = \theta(n_k^2)$,
- or re-normalizing the graph by dividing the distances by a fixed factor is necessary, otherwise the graph can only converge to a totally disconnected space such that, $d(x, x') \ge 2$ a.s.

11.2.2 Degree distribution

Using the formalism of the previous section, the degree distribution can be extended to the distribution of balls of fixed radius. The following result shows that the weak convergence of the distribution of degrees is imposed by the convergence of the spaces.

Theorem 11.4. Let ϕ be a bounded function, $\mathcal{X}_n \xrightarrow{\mathbb{A}_2}_{n \to +\infty} \mathcal{X}$ a converging sequence of mmspaces and $X_n \sim \mu_{\mathcal{X}_n}$ and $X \sim \mu_{\mathcal{X}}$ independent random variables distributed according to the measures of their respective spaces. Then $B_{X_n,\phi}$ converges weakly to $B_{X,\phi}$.

Proof. Let μ_n (resp. μ) be the measure of the random variable $B_{X_n,\phi} = [A_{\mathcal{X}_n,\phi}(\mathbf{1})](X_n)$ (resp. $B_{X,\phi} = [A_{\mathcal{X},\phi}(\mathbf{1})](X)$). Then

$$W_p(\mu_n,\mu) = \Delta_p(A_{\mathcal{X}_n,\phi}(\mathbf{1}), A_{\mathcal{X},\phi}(\mathbf{1})) \leq \Delta_p(A_{\mathcal{X}_n,\phi}, A_{\mathcal{X},\phi}) ||\mathbf{1}||_p \leq L\Delta_p(\mathcal{X}_n, \mathcal{X}),$$

and, since convergence w.r.t. the Wasserstein distance $W_p(\mu_n, \mu)$ implies weak convergence (Villani, 2009, Theorem 6.8), μ_n converges weakly to μ .

This means that the degree distribution, and more generally the distribution of volumes of balls, is preserved through the convergence of the spaces.

11.2.3 Spectral radius

For any matrix M, its operator norm $|||M||| = \max_{x \colon ||x||_2=1} ||Mx||_2$ is equal to its largest singular value. Furthermore, a simple calculation holds $|||M||| = \Delta_p(M,0)$, hence the operator norm, and more generally any *p*-norm, is 1-Lipschitz (and thus continuous) w.r.t. Δ_p .

172

Theorem 11.5. For any $p \ge 1$, the *p*-norm is 1-Lipschitz w.r.t. Δ_p , and $\forall F, G$ operators on \mathcal{X} resp. \mathcal{Y} ,

$$\left| |||F|||_{p} - |||G|||_{p} \right| \leq \Delta_{p}(F,G),$$
(11.14)

where $|||F|||_p = \max_{f : ||f||_p \neq 0} \frac{||F(f)||_p}{||f||_p}$.

Hence, the spectral radius of adjacency operators converges when the spaces converge.

Corollary 11.2. Let \mathcal{X} and \mathcal{Y} be two mm-spaces and ϕ an L-Lipschitz function, then:

$$|\rho(\mathcal{X},\phi) - \rho(\mathcal{Y},\phi)| \le L\Delta_2(\mathcal{X},\mathcal{Y}),\tag{11.15}$$

where $\rho(\mathcal{X}, \phi) = |||A_{\mathcal{X}, \phi}|||$ is the spectral radius of the operator $A_{\mathcal{X}, \phi}$ (see Definition 5.5).

More specifically, when a graph converges w.r.t. Δ_2 , then so does its spectral radius divided by n:

Corollary 11.3. Let \mathcal{G} be an undirected graph and \mathcal{X} an mm-space, then

$$\left|\frac{\rho(\mathcal{G})}{n} - \rho(\mathcal{X}, \phi)\right| \le \Delta_2(\mathcal{G}, \mathcal{X}), \tag{11.16}$$

where $\phi(x) = x\mathbf{1}\{x < 1\} + (2 - x)\mathbf{1}\{1 \le x < 2\}$, and $\rho(\mathcal{G})$ is the spectral radius of \mathcal{G} (see Definition 5.5). Moreover, if \mathcal{G}_n is a sequence of graphs that converges to \mathcal{X} , then

$$\lim_{n \to +\infty} \frac{\rho(\mathcal{G}_n)}{n} = \rho(\mathcal{X}, \mathbb{1}\{x = 1\}) .$$
(11.17)

11.2.4 Diameter

The *p*-size of an mm-space is Lipschitz w.r.t. Δ_p .

Theorem 11.6. The function $\mathcal{X} \mapsto size_p(\mathcal{X})$ is 1-Lipschitz w.r.t. Δ_p , i.e. for all $X, Y \in X_p$,

$$\left|\operatorname{size}_{p}(\mathcal{X}) - \operatorname{size}_{p}(\mathcal{Y})\right| \leq \Delta_{p}(\mathcal{X}, \mathcal{Y}).$$
 (11.18)

Proof. This is a direct consequence of Theorem 11.2 and Theorem 11.3, with $F = A_{\mathcal{X},Id}$, $G = A_{\mathcal{Y},Id}$, q = p, and $f = \mathbf{1}$ the constant function equal to 1.

When $p = +\infty$, the size of an mm-space \mathcal{X} corresponds to its diameter, i.e. the largest distance between two points of \mathcal{X} (in an almost sure sense).

Corollary 11.4. If \mathcal{G}_n is a sequence of graphs converging to the mm-space \mathcal{X} w.r.t. Δ_{∞} , then

$$\lim_{n \to +\infty} diam(\mathcal{G}_n) = size_{\infty}(\mathcal{X}).$$
(11.19)

More usually, social network analysis focuses on the *approximate diameter*, i.e. the distance under which the vast majority of pairewise distances lie. We may define this notion as follows:

$$\operatorname{diam}_{\epsilon}(\mathcal{X}) = \min\{r \ge 0 : \mathbb{P}(d_{\mathcal{X}}(X, X') \le r) \ge 1 - \epsilon\}.$$
(11.20)

This notion of approximate diameter converges also when the graphs (or mm-spaces) only converge w.r.t. Δ_p , where $2 \le p < +\infty$.

Corollary 11.5. If $p \ge 2$, and \mathcal{G}_n is a sequence of graphs converging to the mm-space \mathcal{X} w.r.t. Δ_p , then:

$$\lim_{n \to +\infty} diam_{\epsilon}(\mathcal{G}_n) = diam_{\epsilon}(\mathcal{X}).$$
(11.21)

Proof. The diameter can be rewritten as $diam_{\epsilon}(\mathcal{X}) = min\{r \ge 0 : B_{\mathcal{X},r} \ge 1 - \epsilon\}$, and the convergence of $B_{\chi,r}$ (see Corollary 11.1) implies the desired result.

Spectral properties of compact self-adjoint operators 11.3

In this section, we prove the continuity, and even Lipschitz continuity, of the spectrum of compact self-adjoint operators with respect to Δ_2 . Using this result, we will be able to prove the convergence of the spectrum of symmetric matrices when the sequence of graphs converges w.r.t. Δ_2 . We now give a formal definition of the (point) spectrum, eigenvalues, eigenvectors and eigenspaces.

Definition 11.6 (Spectrum). Let $F : L_{2,\mu_{\mathcal{X}}} \to L_{2,\mu_{\mathcal{X}}}$ be a linear operator. The spectrum of F, noted $sp(F) \subset \mathbb{R}$, is the set of real values λ such that $\exists f \in L_{2,\mu_{\mathcal{X}}}$ s.t. $F(f) = \lambda f$. Such a value λ is called an eigenvalue of *F*, and the corresponding function *f* an eigenvector of *F*.

Note that the definition given here is also called *point spectrum*, and another definition exists for the spectrum of more generic linear operators (see Section 6.3 of the book by Brézis (2011)). However, since we focus on compact self-adjoint operators, the two definitions are equivalent, and for the sake of clarity, we will use the aforementioned definition only (Brézis, 2011, Theorem 6.8).

Definition 11.7 (Eigenspace). Let $F : L_{2,\mu_{\mathcal{X}}} \to L_{2,\mu_{\mathcal{X}}}$ be a linear operator on \mathcal{X} , and $\lambda \in$ sp(F) an eigenvalue of F. Then, the eigenspace $ev(F,\lambda) \subset L_{2,\mu_{\mathcal{X}}}$ is the set of all eigenvectors of eigenvalue λ ,

$$ev(F,\lambda) = \{ f \in L_{2,\mu_{\mathcal{X}}} : F(f) = \lambda f \}.$$
 (11.22)

Since a spectrum is a subset of \mathbb{R} , we will use the Hausdorff distance d_H in order to compare one to another.

Definition 11.8 (Hausdorff distance). Let $X, Y \subset \mathcal{X}$ two subsets of a space \mathcal{X} . The Hausdorff distance between X and Y is the maximum distance between a point of X and its closest match in Y (and respectively for points of Y):

$$d_H(X,Y) = \max\{\sup_{x \in X} \inf_{y \in Y} d(x,y), \sup_{y \in Y} \inf_{x \in X} d(x,y)\}.$$
 (11.23)

Intuitively, the Hausdorff distance describes how far is at least one point of Y of the subset X, or how far is at least one point of X of the subset Y. Hence, upper bounding this quantity $d_H(X,Y) \leq C$ is equivalent to saying that, for each point $x \in X$, there exists a point $y \in Y$ such that $d(x, y) \leq C$ (and similarly for points of Y).

Lemma 11.3. $d_H(X, Y) = 0 \Leftrightarrow X = Y$.

Proof. For all $x \in X$, there exists a $y \in Y$ such that d(x, y) = 0, hence x = y and $x \in Y$. Thus $X \subset Y$, and the symmetry of $d_H(X, Y)$ implies $Y \subset X$, and X = Y.

11.3.1 Convergence of spectrum

The major result of this section is that, for *compact self-adjoint operators*, the spectrum is Lipschitz w.r.t. Δ_2 :

Theorem 11.7. Let F and G be two compact self-adjoint linear operators on \mathcal{X} and \mathcal{Y} , respectively. Then, the Hausdorff distance between their spectrum is bounded by $\Delta_2(F,G)$:

$$d_H(sp(F) \cup \{0\}, sp(G) \cup \{0\}) \le \Delta_2(F, G).$$
(11.24)

Remark 11.2. Note that adding $\{0\}$ to the spectrum is necessary. For example, taking $\mathcal{X}_n = \{0,1\}$ with the uniform measure and $d_{\mathcal{X}_n}(0,1) = 1/n$ leads to $\lim \Delta_2(\mathcal{X}_n, \{0\}) = 0$. However, for $\phi(x) = 1 - x$, $sp(A_{\mathcal{X}_n,\phi}) = \{\frac{1}{2n}, 1 - \frac{1}{2n}\}$ while $sp(A_{\{0\},\phi}) = \{1\}$, and $sp(A_{\mathcal{X}_n,\phi})$ does not converge to $sp(A_{\{0\},\phi})$.

Although compactness may be quite restrictive for linear operators in continuous mmspaces, this assumption is automatically verified for graphs, and hence the result can be generalized to any sequence of converging symmetric matrices.

Corollary 11.6. Let F_n be a sequence of symmetric matrices, and $F : L_{2,\mu_X} \to L_{2,\mu_X}$ an operator on an mm-space \mathcal{X} . If $\lim_{n \to +\infty} \Delta_2(F_n, F) = 0$, then

$$d_H(sp(F_n) \cup \{0\}, sp(F) \cup \{0\}) \le \Delta_2(F_n, F), \tag{11.25}$$

and $sp(F_n) \cup \{0\}$ converges to $sp(F) \cup \{0\}$.

Proof. Each matrix F_n has finite rank, and is thus compact. Let $\pi_n \in \Pi(\mu_{\mathcal{G}_n}, \mu_{\mathcal{X}})$ be a sequence of couplings such that $\lim_{n \to +\infty} \Delta_{2,\pi_n}(F_n, F) = 0$, and $\tilde{F}_n \in L_{2,\mu_{\mathcal{X}}}$ an operator defined by $\tilde{F}_n(f) = F_n(f|_{\mathcal{G}_n}^{\pi_n})|_{\mathcal{X}}^{\pi_n}$. Then \tilde{F}_n has finite rank, and $|||\tilde{F}_n - F||| \leq \Delta_{2,\pi_n}(F_n, F) \rightarrow 0$. Hence, F is the limit of finite rank operators, and is thus compact (Brézis, 2011, Corollary 6.2). Also, F is linear and self-adjoint as a limit of linear and self-adjoint operators, and the conditions for Theorem 11.7 are met.

This result is quite generic and may provide interesting and novel results in random matrix theory. The result also holds for symmetric integral operators, and more specifically for adjacency operators.

Corollary 11.7. Let \mathcal{X} and \mathcal{Y} be two mm-spaces, and $k_{\mathcal{X}} \in L_{2,\mu_{\mathcal{X}} \otimes \mu_{\mathcal{X}}}$ and $k_{\mathcal{Y}} \in L_{2,\mu_{\mathcal{Y}} \otimes \mu_{\mathcal{Y}}}$ two symmetric kernels. Then

$$d_H(sp(F_{k_{\mathcal{X}}})\cup\{0\}, sp(F_{k_{\mathcal{Y}}})\cup\{0\}) \le \min_{\pi\in\Pi(\mu_{\mathcal{X}},\mu_{\mathcal{Y}})}\Delta_{2,\pi\otimes\pi}(k_{\mathcal{X}},k_{\mathcal{Y}}).$$
(11.26)

Proof. Integral operators are compact (Brézis, 2011, Theorem 6.12), and since the kernels are symmetric, then $F_{k_{\chi}}$ and $F_{k_{\chi}}$ are self-adjoint and compact. Hence, the conditions of Theorem 11.7 are met, and Theorem 11.1 gives the final bound in $\Delta_{2,\pi\otimes\pi}(k_{\chi},k_{\chi})$.

Corollary 11.8. If ϕ is L-Lipschitz, then $\mathcal{X} \mapsto sp(A_{\mathcal{X},\phi}) \cup \{0\}$ is Lipschitz w.r.t. Δ_2 and the Hausdorff distance:

$$d_H(sp(A_{\mathcal{X},\phi}) \cup \{0\}, sp(A_{\mathcal{Y},\phi}) \cup \{0\}) \le L\Delta\!\!\!\!\Delta_2(\mathcal{X},\mathcal{Y}). \tag{11.27}$$

Proof. The adjacency operator $A_{\mathcal{X},\phi}$ is an integral operator with a symmetric kernel $\phi \circ d_{\mathcal{X}}$, and Theorem 11.2 gives the final bound in $\Delta_2(\mathcal{X},\mathcal{Y})$.

11.3.2 Convergence of eigenspaces

When the operators converge, the eigenspaces also converge in the Hausdorff sense.

Theorem 11.8. Let F (resp. G) be a compact self-adjoint linear operator in \mathcal{X} (resp. \mathcal{Y}), and $\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ a coupling. Let also $\lambda \in sp(F) \setminus \{0\}$ and $\lambda' \in sp(G) \setminus \{0\}$ be non-zero eigenvalues of F and G, respectively. Then, $\forall f \in ev(F, \lambda)$, $\exists g \in ev(G, \lambda')$ such that

$$\Delta_{2,\pi}(f,g) \le \frac{|\lambda - \lambda'| + \Delta_{2,\pi}(F,G)}{\min\{|\lambda'|, e_g(G,\lambda')\}} ||f||_2, \qquad (11.28)$$

where $eg(G, \lambda') = \inf_{\lambda'' \in sp(G) : \lambda'' \neq \lambda'} |\lambda'' - \lambda'|$ is the eigengap of G for the eigenvalue λ' .

Hence, the convergence of F_n to F w.r.t. Δ_2 and $\lambda_n \in sp(F_n)$ to $\lambda \in sp(F)$ implies the convergence of the corresponding eigenspaces in the Hausdorff sense.

Corollary 11.9. Let F be a compact self-adjoint linear operator on \mathcal{X} , and $(F_n)_{n \in \mathbb{N}}$ a sequence of compact self-adjoint linear operators on \mathcal{X}_n converging to F w.r.t. Δ_2 . Let also $\lambda \in sp(F) \setminus \{0\}$ an eigenvalue of F and $\lambda_n \in sp(F_n) \setminus \{0\}$ a sequence of eigenvalues of F_n converging to λ . Then,

$$\lim_{n \to +\infty} d_H(ev_1(F_n, \lambda_n), ev_1(F, \lambda)) = 0,$$
(11.29)

where $ev_1(F,\lambda) = \{f \in ev(F,\lambda) : ||f||_2 = 1\}$ is the set of eigenvectors of norm 1.

This result may be useful to provide approximation and convergence results for Fourier transforms and convolution for graph signals, and hence justify their use as tools performing operations similar to that of functional analysis in \mathbb{R}^d .

11.4 Convergence of dynamic processes

Simulating a continuous process on a computer requires to discretize the underlying space (usually using a uniform grid or a triangulation) and a large research community is devoted to the study of *when* and *how* is such an approximation valid. In this section, we will use the Gromov-Wasserstein distance in order to present new and intuitive results for such an endeavor and show that, in a certain sense, convergence of a process only requires the convergence of the space on which it lies w.r.t. Δ_p . While the approximation of continuous processes using discrete spaces is not new, a second application of such theoretical results is to find limits of processes on graphs, e.g. epidemic processes. In order to achieve this, we will need to compare functions on different spaces, and optimal couplings are natural mappings between the two spaces that will help us compare functions.

11.4.1 Discrete dynamical systems

Let \mathcal{X} be an mm-space, $F: L_{p,\mu_{\mathcal{X}}} \to L_{p,\mu_{\mathcal{X}}}$ an operator on \mathcal{X} , and $f \in L_{p,\mu_{\mathcal{X}}}$ a function on \mathcal{X} . We say that a sequence of functions $f_n \in L_{p,\mu_{\mathcal{X}}}$ is an (F, f)-discrete dynamical system if $f_0 = f$ and f_n is subject to the following equation of evolution: $\forall x \in \mathcal{X}, \forall n \in \mathbb{N}$,

$$f_{n+1}(x) = [F(f_n)](x), \tag{11.30}$$

When the operator F only depends on the distances between points, for example when F is an adjacency operator, then two processes of similar initial value will tend to stay close if the spaces are similar. In order to show this behavior, we consider two spaces \mathcal{X} and \mathcal{Y} , and relate the distance between two dynamical processes f_n and g_n to the distance between their respective operators. Since we would like their initial value to be similar, we use $f_0 = f |_{\mathcal{X}}^{\pi}$ and $g_0 = f |_{\mathcal{Y}}^{\pi}$, where $f \in L_{p,\pi}$ is a function in $\mathcal{X} \times \mathcal{Y}$.

Theorem 11.9. Let $L \ge 1$, F (resp. G) an L-Lipschitz operator in \mathcal{X} (resp. \mathcal{Y}), $\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ a coupling, $f \in L_{p,\pi}$ a function on $\mathcal{X} \times \mathcal{Y}$, and f_n (resp. g_n) an $(F, f|_{\mathcal{X}}^{\pi})$ -dynamical system (resp. $(G, f|_{\mathcal{Y}}^{\pi})$ -dynamical system). Then, the distance between the two processes is bounded by the distance between F and G:

$$\forall n > 0, \ \Delta_{p,\pi}(f_n, g_n) \le nL^{n-1}\left(||f||_{p,\pi} + n||G(\mathbf{0})||_p\right) \Delta_{p,\pi}(F, G).$$
(11.31)

This theorem implies the convergence of dynamical systems when the corresponding operators converge and their Lipschitz constants are upper bounded by a fixed value.

Corollary 11.10. Let L > 0, $f \in L_{p,\mu_{\mathcal{X}}}$ a function on \mathcal{X} , F an L-Lipschitz operator, and F_k a sequence of L-Lipschitz operators converging to F w.r.t. Δ_p . Let also f_n^k (resp. f_n) an $(F_k, f|_{\mathcal{X}_k}^{\pi})$ -dynamical system (resp. (F, f)-dynamical system). Then, f_n^k converges to f_n for all positive n:

$$\forall n > 0, \lim_{k \to +\infty} \Delta_p(f_n^k, f_n) = 0.$$
(11.32)

Proof. If L < 1, then F and F_k are also 1-Lipschitz, and applying Theorem 11.9 leads to the desired result.

11.4.2 Random walks

One notable example of the previous result is the convergence in probability of random walks on \mathcal{X} . More specifically, we allow the random walker to jump to neighboring positions with a probability that depends on the distance.

Definition 11.9 (Random walk on \mathcal{X}). Let \mathcal{X} be an mm-space and $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ a non-negative function. A ϕ -random walk on \mathcal{X} is a sequence of random variables X_n in \mathcal{X} such that, for each n > 0 and conditionally on $X_n = y$, X_{n+1} is distributed according to a probability density function proportional to $\phi(d_{\mathcal{X}}(x,y))$:

$$p_{X_{n+1}|X_n=y}(x) \propto \phi(d_{\mathcal{X}}(x,y)).$$

In the case of graphs (or discrete spaces equipped with a uniform probability measure), a random walk is simply a Markov chain such that $\mathbb{P}(X_{n+1} = x | X_n = y) \propto \phi(d_{\mathcal{X}}(x, y))$. More generally, a ϕ -random walk on \mathcal{X} is a Markov chain of transition operator equal to the integral operator with kernel

$$k_{\mathcal{X},\phi}^{RW}(x,x') = \frac{\phi(d_{\mathcal{X}}(x,x'))}{\int \phi(d_{\mathcal{X}}(x',x'')d\mu_{\mathcal{X}}(x'')},$$

and thus applying our previous result to a $(F_{k_{X,\phi}^{RW}}, f)$ -dynamical system, where f is the initial distribution of the random walk, leads to the convergence of the random walk X_n in probability.

Theorem 11.10. Let ϕ be a non-negative L-Lipschitz function, \mathcal{X} and \mathcal{Y} two mm-spaces, and X_n (resp. Y_n) a ϕ -random walk on \mathcal{X} (resp. \mathcal{Y}) with initial value X_0 (resp. Y_0) distributed according to the density $p_{X_0} = p|_{\mathcal{X}}^{\pi}$ (resp. $p_{Y_0} = p|_{\mathcal{Y}}^{\pi}$), where $\pi \in Opt_1(\mathcal{X}, \mathcal{Y})$ is an optimal coupling between \mathcal{X}_k and \mathcal{X} , and $p \in L_{1,\pi}$ is a density function in $\mathcal{X} \times \mathcal{Y}$. Then, if p_{X_n} (resp. p_{Y_n}) is the density of X_n (resp. Y_n),

$$\Delta_{1}(p_{X_{n}}, p_{Y_{n}}) \leq nL\Delta_{\infty}(\mathcal{X}, \mathcal{Y}) \frac{size_{\infty}(\mathcal{X}) + size_{\infty}(\mathcal{Y})}{M_{\mathcal{X}, \phi}M_{\mathcal{Y}, \phi}}, \qquad (11.33)$$

where $M_{\mathcal{X},\phi} = \min_{x \in \mathcal{X}} \mathbb{E}[\phi(d_{\mathcal{X}}(X,x))]$ and $M_{\mathcal{Y},\phi} = \min_{y \in \mathcal{Y}} \mathbb{E}[\phi(d_{\mathcal{Y}}(Y,y))].$

As a result, random walks converge when the underlying spaces are converging w.r.t. Δ_{∞} . Also, the bound gives intuition into the quality of the approximation with respect to the convergence rate. For example, for renormalized 2D grid graphs $\frac{1}{m}\mathcal{G}_{CG}(m,2)$ of m^2 nodes, the distance to the limit space is inferior or equal to $\frac{2}{m}$ (see Sec. 10.3.1). Hence, the approximation error is proportional to $\frac{n}{m}$, which provides a natural way of selecting the size of the grid *m* for a certain time window *n*.

11.4.3 Differential systems

For all times $t \ge 0$, let $f_t \in L_{p,\mu_{\mathcal{X}}}$ be a function subject to the following equation of evolution: $\forall x \in \mathcal{X}, f_0(x) = f(x)$, and $\forall t \ge 0$,

$$\frac{\partial}{\partial t}f_t(x) = [F(f_t)](x), \tag{11.34}$$

where $F: L_{p,\mu_{\mathcal{X}}} \to L_{p,\mu_{\mathcal{X}}}$ is a Lipschitz operator on \mathcal{X} , and $f \in L_{p,\mu_{\mathcal{X}}}$ is a function on \mathcal{X} . We will denote as (F, f)-differential system the solution $f_t(x)$ of Eq. 11.34 and, under proper regularity assumptions on F, we will see that the associated differential system converges in the sense of Δ_{v} . Note that the existence (and uniqueness) of a solution for all $t \ge 0$ si due to the global Cauchy-Lipschitz theorem and the fact that F is Lispchitz.

Theorem 11.11. Let *F* (resp. *G*) be an *L*-Lipschitz operator on \mathcal{X} (resp. \mathcal{Y}), $\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ a coupling, $f \in L_{p,\mu_{\mathcal{X}}}$ a function on \mathcal{X} , and f_t (resp. g_t) the solution of an (F, f)-differential system (resp. $(G, f|_{\mathcal{V}}^{\pi})$ -differential system). Then, the distance between the original process f_t and its approximation $g_t|_{\mathcal{X}}^{\pi}$ is bounded by the distance between F and G:

$$\forall t \ge 0, \ ||f_t - g_t|_{\mathcal{X}}^{\pi}||_p \le te^{Lt} \left(||f||_p + \frac{||G(\mathbf{0})||_p}{L} \right) \Delta_{p,\pi}(F,G) \ . \tag{11.35}$$

Intuitively, this theorem means that it is possible to perform the evolution of the differential system with another operator (for example on a discretized space), and then project back to the original space, as long as the two operators are sufficiently close to one another. Thus, this theorem implies the convergence of differential systems when the corresponding operators converge and their Lipschitz constants are upper bounded by a fixed value.

Corollary 11.11. Let $f \in L_{p,\mu_{\mathcal{X}}}$ a function on \mathcal{X} , F an L-Lipschitz operator, F_k a sequence of L-Lipschitz operators converging to F w.r.t. \mathbf{A}_p , and $\pi_k \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{X}_k})$ a sequence of couplings such that $\lim_{k\to+\infty} \Delta_{p,\pi_k}(F,G) = 0$. Then, if f_t^k (resp. f_t) is an $(F_k, f|_{\mathcal{X}_k}^{\pi_k})$ -differential system (resp. (F, f)-differential system), $f_t^k |_{\mathcal{X}}^{\pi_k}$ converges to f_t for all positive times:

$$\forall t > 0, \ \lim_{k \to +\infty} ||f_t^k|_{\mathcal{X}}^{\pi_k} - f_t||_p = 0.$$
(11.36)

Proof. Using Theorem 11.3, $||F_k(\mathbf{0})||_p = ||F(\mathbf{0})||_p$, and applying Theorem 11.11 leads to the desired result.

11.5 Proofs

In all the following proofs, and unless specified otherwise, $(X, Y) \sim \pi$ and $(X', Y') \sim \pi$ are independent pairs of random variables, each of them drawn according to the coupling $\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ between the two mm-spaces \mathcal{X} and \mathcal{Y} .

11.5.1 Proofs of the convergence of operators

Proof of Theorem 11.1. Let us first prove that $F_k(f)$ is well defined and in L_{p,μ_X} for all $f \in L_{p,\mu_X}$. Indeed, $x' \mapsto k(x,x')f(x')$ is almost surely integrable w.r.t. μ_X : let $g(x) = \mathbb{E}[|k(x,X)f(X)|] \in \mathbb{R} \cup \{+\infty\}$,

$$g(x) \le ||k(x, \cdot)||_{r, \mu_{\mathcal{X}}} ||f||_{p, \mu_{\mathcal{X}}}, \tag{11.37}$$

where $r = \frac{p}{p-1}$ using Hölder inequality. Although $||k(x, \cdot)||_{r,\mu_{\mathcal{X}}}$ may not be finite for every x, its p-norm is, and thus g(x) is infinite at most on a negligible subset of \mathcal{X} .

$$||g||_{p,\mu_{\mathcal{X}}} \leq ||f||_{p,\mu_{\mathcal{X}}} \mathbb{E}[|\mathbb{E}[|k(X',X)|^{r} | X']|^{p/r}]^{1/p} \\ \leq ||f||_{p,\mu_{\mathcal{X}}} ||k||_{q,\mu_{\mathcal{X}} \otimes \mu_{\mathcal{X}}}$$

$$< +\infty,$$
(11.38)

where $q = \max\{p, r\}$ and the second inequality is Jensen inequality on $|x|^{p/r}$ concave (resp. convex) when p < 2 (resp. $p \ge 2$). Hence $F_k(f)$ is well defined and

$$||F_{k}(f)||_{p,\mu_{\mathcal{X}}} \le ||f||_{p,\mu_{\mathcal{X}}} ||k||_{q,\mu_{\mathcal{X}} \otimes \mu_{\mathcal{X}}} < +\infty.$$
(11.39)

Now, let $\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ a coupling, $f \in L_{p,\pi}$ a function on $\mathcal{X} \times \mathcal{Y}$ and $x \in \mathcal{X}$. By definition,

$$F_{k_{\mathcal{X}}}(f|_{\mathcal{X}}^{\pi})(x) = \mathbb{E}[k_{\mathcal{X}}(x,X)f|_{\mathcal{X}}^{\pi}(X)]$$

$$= \mathbb{E}[k_{\mathcal{X}}(x,X)\mathbb{E}_{\pi}[f(X,Y)|X]]$$

$$= \mathbb{E}_{\pi}[k_{\mathcal{X}}(x,X)f(X,Y)]$$

$$= F_{\bar{k}_{\mathcal{X}}}(f),$$

(11.40)

where $\tilde{k}_{\mathcal{X}}(x, y, x', y') = k_{\mathcal{X}}(x, x')$ is a function on $\mathcal{X} \times \mathcal{Y}$. Hence,

$$\begin{split} \mathbf{\Delta}_{p,\pi}(F_{k_{\mathcal{X}}}(f|_{\mathcal{X}}^{\pi}),F_{k_{\mathcal{Y}}}(f|_{\mathcal{Y}}^{\pi})) &= ||F_{\tilde{k}_{\mathcal{X}}}(f) - F_{\tilde{k}_{\mathcal{Y}}}(f)||_{p,\pi} \\ &= ||F_{\tilde{k}_{\mathcal{X}}-\tilde{k}_{\mathcal{Y}}}(f)||_{p,\pi} \\ &\leq ||f||_{p,\pi} ||\tilde{k}_{\mathcal{X}}-\tilde{k}_{\mathcal{Y}}||_{q,\pi\otimes\pi} \\ &= ||f||_{p,\pi} \mathbf{\Delta}_{q,\pi\otimes\pi}(k_{\mathcal{X}},k_{\mathcal{Y}}), \end{split}$$
(11.41)
using Eq. 11.39. Since f is any function in $L_{p,\pi}$, we obtain an upper bound for $\Delta_{p,\pi}(F_{k_{\chi}}, F_{k_{\chi}})$:

$$\Delta_{p,\pi}(F_{k_{\mathcal{X}}},F_{k_{\mathcal{Y}}}) = \max_{f \in L_{p,\pi} : ||f||_{p,\pi \neq 0}} \frac{\Delta_{p,\pi}(F_{k_{\mathcal{X}}}(f|_{\mathcal{X}}^{\pi}),F_{k_{\mathcal{Y}}}(f|_{\mathcal{Y}}^{\pi}))}{||f||_{p,\pi}}$$

$$\leq \Delta_{q,\pi \otimes \pi}(k_{\mathcal{X}},k_{\mathcal{Y}}).$$
(11.42)

and taking the minimum over $\Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ returns the desired result.

Proof of Theorem 11.3. Let $\tilde{f}(x,y) = f(x)$ and $q \ge p$. Then

$$\begin{aligned} |||F(f)||_{p} - ||G(f|_{\mathcal{Y}}^{\pi})||_{p}| &= |||F(f)||_{p,\pi} - ||G(f|_{\mathcal{Y}}^{\pi})||_{p,\pi} \\ &\leq ||F(f) - G(f|_{\mathcal{Y}}^{\pi})||_{p,\pi} \\ &= \Delta_{p,\pi}(F(f), G(\tilde{f}|_{\mathcal{Y}}^{\pi})) \\ &\leq \Delta_{q,\pi}(F(f), G(\tilde{f}|_{\mathcal{Y}}^{\pi})) \\ &\leq \Delta_{q,\pi}(F,G)||\tilde{f}||_{q,\pi} \\ &= \Delta_{q,\pi}(F,G)||f||_{q} \,. \end{aligned}$$
(11.43)

11.5.2 Proofs of the convergence of spectral properties

Proof of Theorem 11.7. Let \mathcal{X} and \mathcal{Y} be two mm-spaces and $\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ be a coupling of $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$. The proof relies on three steps:

- First, we show that the spectra of *F* and *G* are equal to the spectra of two operators on $\mathcal{X} \times \mathcal{Y}$ (except for zero).
- Second, we show that an eigenvector for one operator is nearly an eigenvector for the second operator.
- Finally, we show that this implies the existence of an eigenvalue of the second operator close to that of the first one.

Let \tilde{F} and \tilde{G} be linear operators on $\mathcal{X} \times \mathcal{Y}$ defined by: $\forall \phi \in L_2(\mathcal{X} \times \mathcal{Y}), \forall (x, y) \in \mathcal{X} \times \mathcal{Y},$

$$\begin{split} &[\tilde{F}(f)](x,y) = F(f|_{\mathcal{X}}^{\pi})(x), \\ &[\tilde{G}(f)](x,y) = G(f|_{\mathcal{Y}}^{\pi})(y). \end{split}$$
(11.44)

where $f|_{\mathcal{X}}^{\pi}$ and $f|_{\mathcal{Y}}^{\pi}$ are defined as in Eq. 11.1.

Lemma 11.4. *For* $H \in \{F, G\}$ *,*

$$sp(H) \subset sp(\tilde{H}) \subset sp(H) \cup \{0\}.$$
 (11.45)

Proof. First, note that, if f(x,y) = g(x) (i.e. f only depends on the variable x), then $f|_{\mathcal{X}}^{\pi}(x) = \mathbb{E}_{\pi}[g(X) \mid X = x] = g(x)$. Now, let $\lambda \in sp(F)$ and $f \in L_2(\mathcal{X})$ s.t. $F(f) = \lambda f$, and $\tilde{f} \in L_2(\mathcal{X} \times \mathcal{Y})$ s.t. $\tilde{f}(x,y) = f(x)$. Then $\forall (x,y) \in \mathcal{X} \times \mathcal{Y}$,

$$\begin{split} [\tilde{F}(\tilde{f})](x,y) &= F(\tilde{f}|_{\mathcal{X}}^{\pi})(x) \\ &= F(f)(x) \\ &= \lambda f(x) \\ &= \lambda \tilde{f}(x,y), \end{split}$$
(11.46)

and $\lambda \in sp(\tilde{F})$, hence $sp(F) \subset sp(\tilde{F})$.

Conversely, let $\lambda \in sp(\tilde{F})$ and $\tilde{f} \in L_2(\mathcal{X} \times \mathcal{Y})$ s.t. $\tilde{F}(\tilde{f}) = \lambda \tilde{f}$. Then $\forall (x,y) \in \mathcal{X} \times \mathcal{Y}$,

$$\lambda f(x,y) = [\tilde{F}(f)](x,y)$$

= [F($\tilde{f}|_{\mathcal{X}}^{\pi}$)](x), (11.47)

and $\lambda \tilde{f}(x,y)$ does not depend on y. So either $\lambda = 0$, or $\tilde{f}(x,y)$ does not depend on y. In the second case, let $f(x) = \tilde{f}(x,y)$ for any y. Then $F(f)(x) = F(\tilde{f}|_{\mathcal{X}}^{\pi})(x) = \tilde{F}(\tilde{f})(x,y) = \lambda \tilde{f}(x,y) = \lambda f(x)$ and $\lambda \in sp(F)$, hence $sp(\tilde{F}) \subset sp(F) \cup \{0\}$. The proof for \mathcal{Y} is identical.

Let $\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}), \lambda \in sp(\tilde{G}) \text{ and } f \in L_2(\mathcal{X} \times \mathcal{Y}) \text{ s.t. } \tilde{G}(f) = \lambda f$. Then,

$$||\tilde{F}(f) - \lambda f||_{2,\pi} = ||\tilde{F}(f) - \tilde{G}(f)||_{2,\pi}$$

= $\Delta_{2,\pi}(F(f|_{\mathcal{X}}^{\pi}), G(f|_{\mathcal{Y}}^{\pi}))$
 $\leq \Delta_{2,\pi}(F,G)||f||_{2,\pi}$, (11.48)

by definition of $\Delta_{2,\pi}(F,G)$. Finally, Eq. 11.48 implies that there is an eigenvalue of \tilde{F} , denoted λ' , such that $|\lambda' - \lambda| \leq \Delta_{2,\pi}(F,G)$ due to the following lemma:

Lemma 11.5. Let *F* be a self-adjoint and compact operator on a Hilbert space $\mathcal{H}, \lambda \in \mathbb{R}$ and $C \ge 0$. If $\exists x \in \mathcal{H} \text{ s.t. } ||Fx - \lambda x||_2 \le C||x||_2$, then

$$\inf_{\lambda' \in sp(F)} |\lambda' - \lambda| \le C.$$
(11.49)

Proof. Due to the spectral theorem (Brézis, 2011, Theorem 6.11), a self-adjoint and compact operator on a Hilbert space can be diagonalized on an orthonormal basis $(e_i)_{i \in I}$ where I is a countable index set. So $Fx - \lambda x = \sum_{i \in I} (\lambda_i - \lambda) < e_i, x > e_i$ where $< \cdot, \cdot >$ is the scalar product associated to \mathcal{H} . Now $||Fx - \lambda x||_2^2 = \sum_{i \in I} (\lambda_i - \lambda)^2 < e_i, x >^2 \geq ||x||_2^2 \inf_{i \in I} (\lambda_i - \lambda)^2$, which leads to the desired result.

Since \tilde{F} is compact and using Lemma 11.5, $\inf_{\lambda' \in sp(F)} |\lambda' - \lambda| \leq \Delta_{2,\pi}(F,G)$. Since the same result holds for eigenvalues of \tilde{G} , we get that:

$$d_H(sp(\tilde{F}), sp(\tilde{G})) \le \Delta_{2,\pi}(F, G), \tag{11.50}$$

which leads to the desired result using Lemma 11.4 and observing that the result holds for any coupling $\pi \in \Pi(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$.

Proof of Theorem 11.8. Let \tilde{F} and \tilde{G} be the extensions of F and G to $\mathcal{X} \times \mathcal{Y}$ defined by $[\tilde{F}(f)](x,y) = F(f|_{\mathcal{X}}^{\pi})(x)$ and $[\tilde{G}(f)](x,y) = G(f|_{\mathcal{Y}}^{\pi})(y)$. Let also $\lambda \in sp(F) \setminus \{0\}$, $\lambda' \in sp(G) \setminus \{0\}$ and $f \in ev(F,\lambda)$. Then, $\tilde{f}(x,y) = f(x)$ is an eigenvector of \tilde{F} of eigenvalue λ , and let $\tilde{g} \in ev(\tilde{G},\lambda')$ be the orthogonal projection of f on $ev(\tilde{G},\lambda')$. Since \tilde{G} is a compact self-adjoint operator, there exists an orthonormal basis of eigenvectors of \tilde{G} , denoted $(e_i)_{i\in\mathbb{N}}$, with corresponding eigenvalues $(\lambda_i)_{i\in\mathbb{N}}$, and

$$\begin{aligned} ||\tilde{G}(\tilde{f}-\tilde{g})-\lambda'(\tilde{f}-\tilde{g})||_{2,\pi}^2 &= \sum_i (\lambda_i-\lambda')^2 < e_i, \tilde{f}-\tilde{g} >^2\\ &\geq \inf_{\lambda'' \in sp(\tilde{G}) \ : \ \lambda'' \neq \lambda'} (\lambda''-\lambda')^2 ||\tilde{f}-\tilde{g}||_{2,\pi'}^2 \end{aligned}$$
(11.51)

Also, the quantity $||\tilde{G}(\tilde{f}-\tilde{g})-\lambda'(\tilde{f}-\tilde{g})||_{2,\pi}$ is upper bounded by

$$\begin{aligned} ||\tilde{G}(\tilde{f} - \tilde{g}) - \lambda'(\tilde{f} - \tilde{g})||_{2,\pi} &= ||\tilde{G}\tilde{f} - \lambda'\tilde{f}||_{2,\pi} \\ &\leq |\lambda - \lambda'|||\tilde{f}||_{2,\pi} + ||\tilde{G}\tilde{f} - \tilde{F}\tilde{f}||_{2,\pi} \\ &\leq (|\lambda - \lambda'| + \Delta_{2,\pi}(F,G)) ||f||_{2}. \end{aligned}$$
(11.52)

Combining Eq. 11.51 and Eq. 11.52 leads to

$$||\tilde{f} - \tilde{g}||_{2,\pi} \leq \frac{|\lambda - \lambda'| + \Delta_{2,\pi}(F,G)}{\inf_{\lambda'' \in sp(G) \cup \{0\} : \ \lambda'' \neq \lambda'} |\lambda'' - \lambda'|} ||f||_2,$$

and noting that $\tilde{g}(x,y)$ only depends on y (since $\lambda \tilde{g}(x,y) = G(g|_{\mathcal{Y}}^{\pi})(y)$) implies that $g|_{\mathcal{Y}}^{\pi}$ is an eigenvector of G of eigenvalue λ' and $||\tilde{f} - \tilde{g}||_{2,\pi} = \Delta_{2,\pi}(f, \tilde{g}|_{\mathcal{Y}}^{\pi})$ which leads to the desired result.

Proof of Corollary 11.9. First of all, the spectral theorem (Brézis, 2011, Theorem 6.8) implies that the spectrum of F, $sp(F) = (\lambda_i)_{i \in \mathbb{N}}$ is either discrete or $\lim_{i \to +\infty} \lambda_i = 0$, and thus, since $\lambda \neq 0$, $eg(F, \lambda) > 0$. For every eigenvector f_n in $ev_1(F_n, \lambda_n)$, Theorem 11.8 implies that there is an eigenvector $f \in ev(F, \lambda)$ such that

$$\begin{split} \mathbf{\Delta}_{2,\pi}(f_{n}, \frac{f}{||f||_{2}}) &\leq \mathbf{\Delta}_{2,\pi}(f_{n}, f) + ||f - \frac{f}{||f||_{2}}||_{2} \\ &= \mathbf{\Delta}_{2,\pi}(f_{n}, f) + ||f||_{2} - 1 | \\ &= \mathbf{\Delta}_{2,\pi}(f_{n}, f) + ||f||_{2} - ||f_{n}||_{2} | \\ &\leq 2\mathbf{\Delta}_{2,\pi}(f_{n}, f) \\ &\leq 2\mathbf{\Delta}_{2,\pi}(f_{n}, f) \\ &\leq 2\frac{|\lambda_{n} - \lambda| + \mathbf{\Delta}_{2,\pi}(F_{n}, F)}{\min\{|\lambda|, eg(F, \lambda)\}}. \end{split}$$
(11.53)

Similarly, for every eigenvector f in $ev_1(F,\lambda)$, there is an eigenvector $f_n \in ev(F_n,\lambda_n)$ such that

$$\begin{split}
\boldsymbol{\Delta}_{2,\pi}(f, \frac{f_n}{||f_n||_2}) &\leq 2\boldsymbol{\Delta}_{2,\pi}(f, f_n) \\
&\leq 2\frac{|\lambda_n - \lambda| + \boldsymbol{\Delta}_{2,\pi}(F_n, F)}{\min\{|\lambda_n|, eg(F_n, \lambda_n)\}} \\
&\leq 2\frac{|\lambda_n - \lambda| + \boldsymbol{\Delta}_{2,\pi}(F_n, F)}{\min\{|\lambda|, eg(F, \lambda)\} - |\lambda_n - \lambda| - \boldsymbol{\Delta}_{2,\pi}(F_n, F)'}
\end{split}$$
(11.54)

11.5. PROOFS

when min{ $|\lambda|, eg(F, \lambda)$ } - $|\lambda_n - \lambda| - \Delta_{2,\pi}(F_n, F) > 0$ (which happens when *n* is sufficiently large since $\lim_{n\to+\infty} |\lambda_n - \lambda| = 0$ and $\lim_{n\to+\infty} \Delta_{2,\pi}(F_n, F) = 0$). Thus,

$$d_{H}(ev_{1}(F_{n},\lambda_{n}),ev_{1}(F,\lambda)) \leq 2\frac{|\lambda_{n}-\lambda| + \Delta_{2}(F_{n},F)}{\min\{|\lambda|,eg(F,\lambda)\} - |\lambda_{n}-\lambda| - \Delta_{2}(F_{n},F)}, \quad (11.55)$$

which tends to zero as n tends to infinity.

Proof of Theorem 11.9. Let $[\tilde{F}(f)](x,y) = [F(f|_{\mathcal{X}}^{\pi})](x)$ and $[\tilde{G}(f)](x,y) = [G(f|_{\mathcal{Y}}^{\pi})](x)$ be the extensions of F and G to the space $\mathcal{X} \times \mathcal{Y}$, respectively. Then, for all n > 0,

$$f_n(x) = \tilde{f}_n(x,y)$$
 and $g_n(x) = \tilde{g}_n(x,y)$,

where \tilde{f}_n and \tilde{g}_n are (\tilde{F}, f) - and (\tilde{G}, f) -discrete dynamical systems, respectively. Thus,

$$\begin{split} \mathbf{\Delta}_{p,\pi}(f_{n},g_{n}) &= ||\tilde{f}_{n} - \tilde{g}_{n}||_{p,\pi} \\ &= ||\tilde{F}(\tilde{f}_{n-1}) - \tilde{G}(\tilde{g}_{n-1})||_{p,\pi} \\ &\leq ||\tilde{F}(\tilde{f}_{n-1}) - \tilde{F}(\tilde{g}_{n-1})||_{p,\pi} + ||\tilde{F}(\tilde{g}_{n-1}) - \tilde{G}(\tilde{g}_{n-1})||_{p,\pi} \\ &\leq L||\tilde{f}_{n-1} - \tilde{g}_{n-1}||_{p,\pi} + \mathbf{\Delta}_{p,\pi}(F(\tilde{g}_{n-1}|_{\mathcal{X}}^{\pi}), G(\tilde{g}_{n-1}|_{\mathcal{Y}}^{\pi})) \\ &\leq L\mathbf{\Delta}_{p,\pi}(f_{n-1},g_{n-1}) + \mathbf{\Delta}_{p,\pi}(F,G)||\tilde{g}_{n-1}||_{p,\pi}. \end{split}$$
(11.56)

A simple recursion leads to

$$\begin{split} \mathbf{\Delta}_{p,\pi}(f_n, g_n) &\leq L^n || \tilde{f}_0 - \tilde{g}_0 ||_{p,\pi} + \mathbf{\Delta}_{p,\pi}(F, G) \sum_{k=0}^{n-1} || \tilde{g}_k ||_{p,\pi} L^{n-k-1} \\ &\leq \mathbf{\Delta}_{p,\pi}(F, G) \sum_{k=0}^{n-1} || \tilde{g}_k ||_{p,\pi} L^{n-k-1}, \end{split}$$
(11.57)

since $\tilde{f}_0 = \tilde{g}_0 = f$. However, for $k \le n$,

$$||\tilde{g}_{k}||_{p,\pi} = ||\tilde{G}(\tilde{g}_{k-1})||_{p,\pi}$$

$$\leq L||\tilde{g}_{k-1}||_{p,\pi} + ||G(\mathbf{0})||_{p,\pi}$$

$$\leq L^{k}||f||_{p,\pi} + \sum_{i=0}^{k-1} L^{i}||G(\mathbf{0})||_{p,\pi}$$

$$\leq L^{k} (||f||_{p,\pi} + n||G(\mathbf{0})||_{p,\pi}).$$
(11.58)

since $L \ge 1$. Using Eq. 11.58, Eq. 11.57 becomes

$$\Delta_{p,\pi}(f_n, g_n) \leq \Delta_{p,\pi}(F, G) \sum_{k=0}^{n-1} L^{n-1} \left(||f||_{p,\pi} + n ||G(\mathbf{0})||_{p,\pi} \right)$$

$$\leq \Delta_{p,\pi}(F, G) n L^{n-1} \left(||f||_{p,\pi} + n ||G(\mathbf{0})||_{p,\pi} \right).$$
(11.59)

Proof of Theorem 11.11. Similarly to the proof of Theorem 11.9, let $[\tilde{F}(f)](x,y) = [F(f|_{\mathcal{X}}^{\pi})](x)$ and $[\tilde{G}(f)](x,y) = [G(f|_{\mathcal{Y}}^{\pi})](x)$ be the extensions of F and G to the space $\mathcal{X} \times \mathcal{Y}$, respectively. Also, let $\tilde{f}(x,y) = f(x)$, $\tilde{g}(x,y) = g(x)$, and $h(t) = ||f_t - g_t|_{\mathcal{X}}^{\pi}||_p$

be the distance between the two processes at time t. The function h(t) is differentiable and

$$\frac{d}{dt}h(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left(||\tilde{f}_{t+\Delta t} - \tilde{g}_{t+\Delta t}|^{\pi}_{\mathcal{X}}||_{p,\pi} - ||\tilde{f}_{t} - \tilde{g}_{t}|^{\pi}_{\mathcal{X}}||_{p,\pi} \right)
\leq \lim_{\Delta t \to 0} \frac{1}{\Delta t} ||\tilde{f}_{t+\Delta t} - \tilde{g}_{t+\Delta t}|^{\pi}_{\mathcal{X}} - \tilde{f}_{t} + \tilde{g}_{t}|^{\pi}_{\mathcal{X}}||_{p,\pi}
= \lim_{\Delta t \to 0} \left\| \left(\frac{\tilde{f}_{t+\Delta t} - \tilde{f}_{t}}{\Delta t} - \frac{\tilde{g}_{t+\Delta t} - \tilde{g}_{t}}{\Delta t} \right) \right|^{\pi}_{\mathcal{X}} \right\|_{p,\pi}
\leq \lim_{\Delta t \to 0} \left\| \frac{\tilde{f}_{t+\Delta t} - \tilde{f}_{t}}{\Delta t} - \frac{\tilde{g}_{t+\Delta t} - \tilde{g}_{t}}{\Delta t} \right\|_{p,\pi}
= ||\tilde{F}(\tilde{f}_{t}) - \tilde{G}(\tilde{g}_{t})||_{p,\pi}
\leq ||\tilde{F}(\tilde{f}_{t}) - \tilde{F}(\tilde{g}_{t})||_{p,\pi} + ||\tilde{F}(\tilde{g}_{t}) - \tilde{G}(\tilde{g}_{t})||_{p,\pi}$$

$$\leq Lh(t) + \Delta_{p,\pi}(F,G)||g_{t}||_{p}.$$
(11.60)

Now, $||\tilde{g}_t||_p$ can also be bounded using

$$\frac{d}{dt}||g_t||_p \leq ||G(g_t)||_p \\
\leq ||G(\mathbf{0})||_p + L||g_t||_p,$$
(11.61)

and hence

$$||g_t||_p \le -\frac{||G(\mathbf{0})||_p}{L} + \left(||f||_p + \frac{||G(\mathbf{0})||_p}{L}\right)e^{Lt}.$$
(11.62)

Using this bound in Eq. 11.60, we obtain

$$\frac{d}{dt}h(t) \le Lh(t) + \Delta_p(F,G) \left[-\frac{||G(\mathbf{0})||_p}{L} + \left(||f||_p + \frac{||G(\mathbf{0})||_p}{L} \right) e^{Lt} \right], \quad (11.63)$$

which leads to the desired result by considering $h(t) = a(t)e^{Lt}$ and integrating the bound on $\frac{d}{dt}a(t)$.

Proof of Theorem 11.10. First, let us verify that $p_{X_{n+1}} = F_{k_{\mathcal{X},\phi}^{RW}}(p_{X_n})$. For all functions $f \in L_1$,

$$\begin{split} \mathbb{E}_{\mu_{\mathcal{X}}}[f(X)p_{X_{n+1}}(X)] &= \mathbb{E}[f(X_{n+1})] \\ &= \mathbb{E}[\mathbb{E}[f(X_{n+1})|X_n]] \\ &= \mathbb{E}[\mathbb{E}[f(X)\frac{\phi(d_{\mathcal{X}}(X_n,X)|X_n]}{\mathbb{E}[\phi(d_{\mathcal{X}}(X_n,X)|X_n]}|X_n]] \\ &= \mathbb{E}[f(X)\frac{\phi(d_{\mathcal{X}}(X_n,X)|X_n]}{\mathbb{E}[\phi(d_{\mathcal{X}}(X',X)|X'_n]}] \\ &= \mathbb{E}[f(X)\frac{\phi(d_{\mathcal{X}}(X',X)}{\mathbb{E}[\phi(d_{\mathcal{X}}(X',X)|X'_n]}p_{X_n}(X')] \\ &= \mathbb{E}[f(X)\mathbb{E}[\frac{\phi(d_{\mathcal{X}}(X',X)|X'_n]}{\mathbb{E}[\phi(d_{\mathcal{X}}(X',X)|X'_n]}p_{X_n}(X')|X]] \\ &= \mathbb{E}[f(X)F_{k_{\mathcal{X},\phi}^{RW}}(p_{X_n})(X)], \end{split}$$
(11.64)

where X and X' are independent random variables drawn according to $\mu_{\mathcal{X}}$ (and independent of X_n). Since this equality holds for all functions $f \in L_1$, this implies that $p_{X_{n+1}} =$

184

 $F_{k_{\mathcal{X},\phi}^{RW}}(p_{X_n})$. Also, $F_{k_{\mathcal{X},\phi}^{RW}}$ is 1-Lipschitz since it is linear and, for all $f \in L_1$,

$$\begin{aligned} ||F_{k_{\mathcal{X},\phi}^{RW}}(f)||_{1} &= \mathbb{E}\left[\left|\mathbb{E}\left[\frac{\phi(d_{\mathcal{X}}(X',X)}{\mathbb{E}[\phi(d_{\mathcal{X}}(X',X)|X']}f(X')|X]}\right|\right] \\ &\leq \mathbb{E}\left[\frac{\phi(d_{\mathcal{X}}(X',X)}{\mathbb{E}[\phi(d_{\mathcal{X}}(X',X)|X']}|f(X')|]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\frac{\phi(d_{\mathcal{X}}(X',X)}{\mathbb{E}[\phi(d_{\mathcal{X}}(X',X)|X']}|f(X')||X']\right] \\ &= \mathbb{E}[|f(X')|] \\ &= ||f||_{1}. \end{aligned}$$
(11.65)

Thus, the conditions for Theorem 11.9 are met, and

$$\forall n > 0, \ \Delta _{1,\pi}(p_{X_n}, p_{Y_n}) \le n \Delta _{1,\pi}(F_{k^{RW}_{\mathcal{X},\phi}}, F_{k^{RW}_{\mathcal{Y},\phi}}).$$

$$(11.66)$$

In order to derive the desired result, we now need to upper bound the distance between $F_{k_{\mathcal{X},\phi}^{RW}}$ and $F_{k_{\mathcal{Y},\phi}^{RW}}$. Since these operators are integral operators, Theorem 11.1 implies that

$$\begin{split} \mathbf{\Delta}_{1,\pi}(F_{k_{\mathcal{X},\phi}^{RW}},F_{k_{\mathcal{Y},\phi}^{RW}}) &\leq \mathbf{\Delta}_{\infty,\pi}(k_{\mathcal{X},\phi}^{RW},k_{\mathcal{Y},\phi}^{RW}) \\ &\leq \left| \left| \frac{\phi \circ d_{\mathcal{X}}}{\mathbb{E}[\phi(d_{\mathcal{X}}(\cdot,X)]} - \frac{\phi \circ d_{\mathcal{Y}}}{\mathbb{E}[\phi(d_{\mathcal{Y}}(\cdot,Y)]} \right| \right|_{\infty,\pi} \\ &\leq \frac{\left| \left| \mathbb{E}[\phi(d_{\mathcal{Y}}(\cdot,Y)]\phi \circ d_{\mathcal{X}} - \mathbb{E}[\phi(d_{\mathcal{X}}(\cdot,X)]\phi \circ d_{\mathcal{Y}}] \right|_{\infty,\pi}}{M_{\mathcal{X},\phi}M_{\mathcal{Y},\phi}} \right| \\ &\leq \frac{L\mathbf{\Delta}_{\infty,\pi}(\mathcal{X},\mathcal{Y}) \left(\text{size}_{\infty}(\mathcal{X}) + \text{size}_{\infty}(\mathcal{Y}) \right)}{M_{\mathcal{X},\phi}M_{\mathcal{Y},\phi}}. \end{split}$$
(11.67)

Bibliography

Addario-Berry, L., N. Broutin, and C. Goldschmidt 2012. The continuum limit of critical random graphs. <i>Probability Theory and Related</i> <i>Fields</i> , 152(3):367–406.	
Adolphson, D. and T. C. Hu 1973. Optimal linear ordering. <i>SIAM Journal on Applied Mathematics</i> , 25(3):403–423.	
Alfonsi, A. and P. Blanc 2015. Dynamic optimal execution in a mixed-market-impact Hawkes price model. <i>Fi-</i> <i>nance and Stochastics</i> , 20(1):183–218.	
Arora, S., E. Hazan, and S. Kale 2005. Fast algorithms for approximate semidefinite programming using the multiplica- tive weights update method. In <i>Proceedings of the 46th IEEE Symposium on Foundations</i> <i>of Computer Science</i> , pp. 339–348.	
Barabási, AL. and R. Albert 1999. Emergence of scaling in random networks. <i>Science</i> , 286(5439):509–512.	
Bauwens, L. and N. Hautsch 2009. Modelling financial high frequency data using point processes. Springer.	
Ben-Naim, E. and P. L. Krapivsky 2004. Size of outbreaks near the epidemic threshold. <i>Physical Review E</i> , 69:050901.	
Bender, E. A. and E. Canfield 1978. The asymptotic number of labeled graphs with given degree sequences. <i>Journal</i> of Combinatorial Theory, Series A, 24(3):296–307.	
Benjamini, I. and O. Schramm 2001. Recurrence of distributional limits of finite planar graphs. <i>Electronic Journal of</i> <i>Probability</i> , 6(23):1–13.	
Bollobás, B. 1984. The evolution of random graphs. <i>Transactions of the American Mathematical</i> <i>Society</i> , 286(1):257–274.	
Bollobás, B., C. Borgs, J. Chayes, and O. Riordan 2010. Percolation on dense graph sequences. <i>The Annals of Probability</i> , 38(1):150–183.	
Bollobás, B., S. Janson, and O. Riordan 2007. The phase transition in inhomogeneous random graphs. <i>Random Structures &</i> <i>Algorithms</i> , 31(1):3–122.	
Borgs, C., J. Chayes, A. Ganesh, and A. Saberi 2010. How to distribute antidote to control epidemics. <i>Random Structures & Algorithms</i> ,	

37(2):204-222.

Borgs, C., J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi 2006. Topics in Discrete Mathematics: Dedicated to Jarik Nešetřil on the Occasion of his 60th Birthday, chapter Counting Graph Homomorphisms, pp. 315–371. Springer Berlin Heidelberg.

Borgs, C., J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi 2008. Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6):1801–1851.

Brézis, H.

2011. Functional analysis, Sobolev spaces and partial differential equations, Universitext. New York, Heidelberg, London: Springer.

Chen, W., C. Wang, and Y. Wang

2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1029–1038. ACM.

Chen, W., Y. Wang, and S. Yang

2009. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 199–208. ACM.

Christakis, N. A. A. and J. H. H. Fowler 2007. The spread of obesity in a large social network over 32 years. *New England Journal* of Medicine, 357(4):370–379.

Chung, F., P. Horn, and A. Tsiatas 2009. Distributing antidote using PageRank vectors. *Internet Mathematics*, 6(2):237–254.

Chung, F. R. and P. D. Seymour 1989. Graphs with small bandwidth and cutwidth. *Discrete Mathematics*, 75(1-3):113– 119.

Cohen, R., S. Havlin, and D. Ben-Avraham 2003. Efficient immunization strategies for computer networks and populations. *Physi*cal Review Letters, 91(24):247901.

Crane, R. and D. Sornette 2008. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653.

Díaz, J., J. Petit, and M. Serna 2002. A survey of graph layout problems. *ACM Computing Surveys*, 34(3):313–356.

Draief, M., A. Ganesh, and L. Massoulié 2008. Thresholds for virus spread on networks. *Annals of Applied Probability*, 18(2):359– 378.

Drakopoulos, K., A. Ozdaglar, and J. N. Tsitsiklis 2014a. An efficient curing policy for epidemics on graphs. *IEEE Transactions on Network Science and Engineering*, 1(2):67–75.

- Drakopoulos, K., A. Ozdaglar, and J. N. Tsitsiklis 2014b. An efficient curing policy for epidemics on graphs. *ArXiv e-prints*, 1407.2241v1.
- Du, N., L. Song, M. Gomez-Rodriguez, and H. Zha 2013a. Scalable influence estimation in continuous-time diffusion networks. In Advances in Neural Information Processing Systems, pp. 3147–3155.

Du, N., L. Song, H. Woo, and H. Zha

2013b. Uncover topic-sensitive information diffusion networks. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics*, pp. 229–237.

Elek, G.

2007. Note on limits of finite graphs. Combinatorica, 27(4):503-507.

Elek, G.

2012. Finite graphs and amenability. Journal of Functional Analysis, 263(9):2593 - 2614.

Erdös, P. and A. Rényi

1960. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61.

Farajtabar, M., N. Du, M. Gomez-Rodriguez, I. Valera, H. Zha, and L. Song 2014. Shaping social activity by incentivizing users. In Advances in Neural Information Processing Systems, pp. 2474–2482.

Forster, G. A. and C. A. Gilligan 2007. Optimizing the control of disease infestations at the landscape scale. *Proceedings* of the National Academy of Sciences, 104(12):4984–4989.

Fortuin, C. M., P. W. Kasteleyn, and J. Ginibre 1971. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22(2):89–103.

Fournier, N. and A. Guillin

2015. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3):707–738.

Ganesh, A., L. Massoulié, and D. Towsley 2005. The effect of network topology on the spread of epidemics. In Proceedings of the 24th Conference of the IEEE Computer and Communications Societies, volume 2 of INFOCOM, pp. 1455–1466. IEEE.

- Gomez-Rodriguez, M., D. Balduzzi, and B. Schölkopf 2011. Uncovering the temporal dynamics of diffusion networks. In *Proceedings of the* 28th International Conference on Machine Learning, pp. 561–568.
- Gomez-Rodriguez, M., L. Song, H. Daneshmand, and B. Schoelkopf 2015. Estimating diffusion networks: Recovery conditions, sample complexity & softthresholding algorithm. *Journal of Machine Learning Research*.

Grassberger, P.

2003. Critical percolation in high dimensions. Physical Review E, 67(3):036101.

Gromov, M.

1981. *Structures métriques pour les variétés riemanniennes*, Textes mathématiques. CEDIC/Fernand Nathan.

Gromov, M.

1999. Metric structures for Riemannian and non-Riemannian spaces. Birkhäuser Boston.

Harper, L. H.

1964. Optimal assignments of numbers to vertices. *Journal of the Society for Industrial and Applied Mathematics*, 12(1):131–135.

Hawkes, A. G. and D. Oakes

1974. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, pp. 493–503.

Holland, P. W., K. B. Laskey, and S. Leinhardt 1983. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137.

Holley, R.

1974. Remarks on the FKG inequalities. *Communications in Mathematical Physics*, 36(3):227–231.

Horowitz, J. and R. L. Karandikar

1994. Mean rates of convergence of empirical measures in the Wasserstein metric. *Journal of Computational and Applied Mathematics*, 55(3):261–273.

Juvan, M. and B. Mohar

1992. Optimal linear labelings and eigenvalues of graphs. *Discrete Applied Mathematics*, 36(2):153–168.

Kaiser, M. and C. C. Hilgetag 2004. Spatial growth of real-world networks. *Physical Review E*, 69:036103.

Kalogeratos, A., K. Scaman, and N. Vayatis 2015. Learning to suppress SIS epidemics in networks. In Networks in the Social and Information Sciences workshop.

Kempe, D., J. Kleinberg, and E. Tardos

2003. Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137–146. ACM.

Kermack, W. O. and A. G. McKendrick

1932. Contributions to the mathematical theory of epidemics. II. the problem of endemicity. *Proceedings of the Royal society of London. Series A*, 138(834):55–83.

Khouzani, M., S. Sarkar, and E. Altman

2011. Optimal control of epidemic evolution. In *Proceedings of the 30th Conference of the IEEE Computer and Communications Societies*, pp. 1683–1691. IEEE.

Klepac, P., O. N. Bjørnstad, C. J. E. Metcalf, and B. T. Grenfell 2012. Optimizing reactive responses to outbreaks of immunizing infections: Balancing case management and vaccination. *PLoS ONE*, 7(8):e41428.

Kolaczyk, E. D.

2009. *Statistical analysis of network data : methods and models*, Springer series in statistics. New York, London: Springer.

Lemonnier, R., K. Scaman, and A. Kalogeratos

2016a. Multivariate Hawkes processes for large-scale inference. *ArXiv e-prints*, 1602.08418.

Lemonnier, R., K. Scaman, and N. Vayatis

2014. Tight bounds for influence in diffusion networks and application to bond percolation and epidemiology. In *Advances in Neural Information Processing Systems*, pp. 846–854.

Lemonnier, R., K. Scaman, and N. Vayatis

2016b. Spectral bounds in random graphs applied to spreading phenomena and percolation. *ArXiv e-prints*, 1603.07970.

Lemonnier, R. and N. Vayatis

2014. Nonparametric Markovian learning of triggering kernels for mutually exciting and mutually inhibiting multivariate Hawkes processes. In *Machine Learning and Knowledge Discovery in Databases*, pp. 161–176. Springer.

Leskovec, J., A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance 2007. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 420– 429. ACM.

Liniger, T. J.

2009. *Multivariate Hawkes processes*. PhD thesis, Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 18403.

Lovász, L.

2012. Large networks and graph limits, volume 60. American Mathematical Society.

Lovász, L. and B. Szegedy

2006. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957.

Łuczak, T.

1990. Component behavior near the critical point of the random graph process. *Random Structures & Algorithms*, 1(3):287–310.

McAuley, J. J. and J. Leskovec

2012. Learning to discover social circles in ego networks. In Advances in Neural Information Processing Systems, pp. 548–556.

Mémoli, F.

2011. Gromov–Wasserstein distances and the metric approach to object matching. *Foun- dations of Computational Mathematics*, 11(4):417–487.

Mémoli, F.

2014. The Gromov-Wasserstein distance: A brief overview. Axioms, 3(3):335-341.

Molloy, M. and B. Reed

1995. A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, 6(2-3):161–180.

Molloy, M. and B. Reed

1998. The size of the giant component of a random graph with a given degree sequence. *Combinatorics probability and computing*, 7(3):295–305.

Nelson, K. E.

2007. Epidemiology of infectious disease: general principles. *Infectious Disease Epidemiology Theory and Practice. Gaithersburg, MD: Aspen Publishers,* pp. 17–48.

Newman, M.

2010. Networks: An Introduction. New York, NY, USA: Oxford University Press.

Newman, M. E.

2002. Spread of epidemic disease on networks. Physical review E, 66(1):016128.

Norros, I. and H. Reittu

2006. On a conditionally Poissonian graph process. *Advances in Applied Probability*, 38(1):59–75.

Oakes, D.

1975. The Markovian self-exciting process. Journal of Applied Probability, pp. 69-77.

otto Georgii, H., O. Häggström, and C. Maes

1999. The random geometry of equilibrium phases. *Phase Transitions and Critical Phenomena*, 18:1–142.

Pantrigo, J. J., R. Martí, A. Duarte, and E. G. Pardo 2012. Scatter search for the cutwidth minimization problem. *Annals of Operations Research*, 199(1):285-304.

Pardo, E. G., N. Mladenović, J. J. Pantrigo, and A. Duarte 2013. Variable formulation search for the cutwidth minimization problem. *Applied Soft Computing*, 13(5):2242–2252.

Pastor-Satorras, R., C. Castellano, P. Van Mieghem, and A. Vespignani 2015. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87:925–979.

Penrose, M.

2003. Random geometric graphs, volume 5. Oxford University Press.

Pouget-Abadie, J. and T. Horel

2015. Inferring graphs from cascades: A sparse recovery framework. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 977–986.

- Prakash, B. A., D. Chakrabarti, N. C. Valler, M. Faloutsos, and C. Faloutsos 2012. Threshold conditions for arbitrary cascade models on arbitrary networks. *Knowledge and Information Systems*, 33(3):549–575.
- Preciado, V. M., M. Zargham, C. Enyioha, A. Jadbabaie, and G. Pappas 2013a. Optimal vaccine allocation to control epidemic outbreaks in arbitrary networks. *ArXiv e-prints*, 1303.3984.
- Preciado, V. M., M. Zargham, C. Enyioha, A. Jadbabaie, and G. J. Pappas 2013b. Optimal resource allocation for network protection: A geometric programming approach. *ArXiv e-prints*, 1309.6270.
- Reynaud-Bouret, P., V. Rivoirard, F. Grammont, and C. Tuleau-Malot 2014. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *Journal of Mathematical Neurosciences*, 4(1):1–41.
- Robins, G., P. Pattison, Y. Kalish, and D. Lusher 2007. An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2):173–191.

Rodriguez, M. G. and B. Schölkopf 2012. Influence maximization in continuous time diffusion networks. In *Proceedings of* the 29th International Conference on Machine Learning, pp. 313–320.

Rodriguez-Tello, E., J.-K. Hao, and J. Torres-Jimenez 2008. An effective two-stage simulated annealing algorithm for the minimum linear arrangement problem. *Computers & Operations Research*, 35(10):3331–3346.

Saberi, A. A.

2015. Recent advances in percolation theory and its applications. *Physics Reports*, 578:1–32.

Scaman, K., A. Kalogeratos, and N. Vayatis 2014a. Dynamic treatment allocation for epidemic control in arbitrary networks. In *Diffusion Networks and Cascade Analytics workshop*.

Scaman, K., A. Kalogeratos, and N. Vayatis 2014b. What makes a good plan? An efficient planning approach to control diffusion processes in networks. *ArXiv e-prints*, 1407.4760.

Scaman, K., A. Kalogeratos, and N. Vayatis

2015a. A greedy approach for dynamic control of diffusion processes in networks. In *Proceedings of the 27th IEEE International Conference on Tools with Artificial Intelligence*, pp. 652–659. IEEE.

- Scaman, K., R. Lemonnier, and N. Vayatis 2015b. Anytime influence bounds and the explosive behavior of continuous-time diffusion networks. In *Advances in Neural Information Processing Systems*, pp. 2026–2034.
- Schneider, C. M., T. Mihaljev, S. Havlin, and H. J. Herrmann 2011. Suppressing epidemics with a limited amount of immunization units. *Physical Review E*, 84(6):061911.

Sturm, K.-T.

2006. On the geometry of metric measure spaces. Acta Mathematica, 196(1):65-131.

Sturm, K.-T.

2013. The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces. *ArXiv e-prints*, 1208.0434.

Tong, H., B. A. Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos 2012. Gelling, and melting, large graphs by edge manipulation. In *Proceedings of the 21st* ACM International Conference on Information and Knowledge Management, pp. 245–254. ACM.

Trusov, M., R. E. Bucklin, and K. Pauwels

2009. Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of marketing*, 73(5):90–102.

Ugander, J., B. Karrer, L. Backstrom, and C. Marlow 2011. The anatomy of the Facebook social graph. *ArXiv e-prints*, 1111.4503.

Van Mieghem, P., J. Omic, and R. Kooij 2009. Virus spread in networks. *IEEE/ACM Transactions on Networking*, 17(1):1–14.

Vere-Jones, D.

1978. Earthquake prediction – statistician's view. Journal of Physics of the Earth, 26(2):129–146.

Villani, C.

2009. *Optimal transport: old and new*, Grundlehren der mathematischen Wissenschaften. Berlin: Springer.

Wang, Y., D. Chakrabarti, C. Wang, and C. Faloutsos
2003. Epidemic spreading in real networks: An eigenvalue viewpoint. In *Proceedings of the 22nd International Symposium on Reliable Distributed Systems*, pp. 25–34. IEEE.

Watts, D. J. and S. H. Strogatz

1998. Collective dynamics of 'small-world' networks. Nature, 393(6684):440-442.

Zhou, K., H. Zha, and L. Song

2013. Learning triggering kernels for multi-dimensional Hawkes processes. In *Proceedings of the 30th International Conference on Machine Learning*, pp. 1301–1309.



Titre : Etude et contrôle de phénomènes diffusifs dans un réseau

Mots clés : Réseaux, Diffusion, Processus Stochastiques, Epidémies, Contrôle

Résumé : La propagation au sein d'un réseau est un sujet d'étude pour de nombreux domaines scientifiques. Épidémies, marketing viral ou propagation d'information au sein d'un réseau social sont autant de phénomènes réels modélisés par l'évolution d'une caractéristique se propageant à travers un réseau de proche en proche. Ainsi, être capable d'agir sur ces phénomènes de diffusion est un enjeu capital dans de nombreux domaines. Malgré l'abondance de la littérature à ce sujet sur le plan théorique, et notamment la détermination d'un seuil épidémique au dessous duquel la propagation se résorbe, un certain nombre de limitations réduisent l'impact pratique de ces travaux. Dans cette thèse, nous avons travaillé à réduire la distance séparant pratique et théorie, et ce suivant trois axes :

la généralisation de résultats théoriques à une classe plus large et réaliste de modèles de propagation, le développement de méthodes de contrôle dynamique efficaces utilisant de manière judicieuse la structure du réseau, et enfin la définition de nouveaux outils mathématiques faisant le lien entre méthodes spatiales et de réseau en épidémiologie. Plus particulièrement, nos travaux permettent l'analyse rigoureuse du comportement des caractéristiques d'un réseau lorsque celui-ci se rapproche, au niveau de sa structure, d'un espace métrique donné, et pourrait permettre l'application de méthodes de contrôle sur réseau à des données spatiales et macroscopiques à l'aide données (notamment de démographiques et de transport) du réseau de contact au sein d'une population.

Title : Study and control of diffusion processes in networks

Keywords : Networks, Diffusion, Stochastic processes, epidemics, control

Abstract : The propagation of a characteristic through a network is the subject of study of many scientific fields. Epidemics, viral marketing or information propagation through a social network are among the many examples of real phenomena modeled by the evolution of a characteristic propagating through the edges of a network. Thus, being capable of acting on these diffusion processes is of capital interest for many fields. Despite the large literature on the theoretical aspects of diffusion processes, and more specifically the discovery of an epidemic threshold under which the propagation is not sustainable, a number of practical limitations prevent the use of these studies in real-life scenarios. In this thesis, we work on reducing the distance separating theory from practice,

following three distinct research directions: the generalization of theoretical results to a larger and more realistic class of diffusion models, the development of efficient dynamic control measures utilizing the structure of the network to its advantage, and, finally, the definition of new mathematical tools bridging the gap between spatial and network approaches in epidemiology. More specifically, our work allows the rigorous analysis of the behavior of a network's characteristics when it converges, in a structural sense, to a given metric space, and could open the way to the application of control strategies that have been developed for networks to scenarios in which only spatial and macroscopic information is available (e.g. transportation or demographic data).