

# Contagion on Complex Systems: Structure and Dynamics

by

**Nicholas W. Landry**

B.S., University of New Hampshire, 2014

M.S., University of Colorado Boulder, 2020

A thesis submitted to the  
Faculty of the Graduate School of the  
University of Colorado in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
Department of Applied Mathematics

2022

Committee Members:

Juan G. Restrepo, Chair

James D. Meiss

Zachary P. Kilpatrick

Daniel B. Larremore

Mason A. Porter

Landry, Nicholas W. (Ph.D., Applied Mathematics)

Contagion on Complex Systems: Structure and Dynamics

Thesis directed by Prof. Juan G. Restrepo

Complex systems are important when representing empirical systems in that they can model the underlying structure of interactions. Accounting for this structure can offer important insights for empirical systems such as social networks, biological processes, social phenomena, opinion formation, and many other examples. Pairwise networks are a representation of complex systems comprising a collection of entities (nodes) and pairwise interactions between entities (edges). Hypergraphs are a generalization of pairwise networks where interactions are no longer constrained to be between two nodes, but rather can be of arbitrary size. Modeling dynamics on hypergraphs can uncover rich behavior that one might not see if the dynamics simply occurred on a pairwise network. We focus on the interplay between the structure of a complex system, a particular dynamical process, and the resulting dynamical behavior. In the context of hypergraphs, we explain the effects that degree heterogeneity, assortative mixing, and community structure have on a simple hypergraph contagion model. Likewise, for pairwise networks, we explore both types of structure; structure in the underlying contact network and varying heterogeneity in the infection model. We examine the effect that representing inherently multiplex data (relationships of different types) with uniplex networks (relationships of a single type) has on the resulting dynamical behavior. We present two open source software libraries: (1) XGI, a package for representing complex systems with group interactions and (2) HyperContagion, a package for simulating hypergraph contagion, both of which can be used by the growing community of researchers studying higher-order interactions.

## Dedication

To my wife, Shari.

## Acknowledgements

I want to acknowledge all the people who have supported me throughout this endeavor.

First, I would like to thank my advisor Juan G. Restrepo for his support and guidance. I admire him and the work he does and hope to be like him when I advise students in the future. Thank you to my committee, Jim Meiss, Zack Kilpatrick, Dan Larremore, and Mason Porter for their advice and encouragement.

Thank you to my collaborators, jimi adams, Joel Miller, Leo Torres, Maxime Lucas, Iacopo Iacopini, Giovanni Petri, Sean Shaheen, Jake Perez, Heather Zinn Brooks, Sinan Aksoy, Mirah Shi, and Ilya Amburg. I have appreciated the opportunity to collaborate with such great people who have challenged me to grow and evolve in my perspectives on research.

I would like to thank my peers and cohort at CU Boulder. The community that we created has been deeply meaningful to me. Richie Clancy has been a consistent friend through the ups and downs of the past few years through playing squash, walking on the Boulder Creek path, flying, and hiking. He has been a huge part of getting through graduate school. Amanda Hampton, Subekshya Bidari, Sabina Adhikari, Lyndsey Wong, Shay Gilpin, DeAnna Gilchrist, Seung Kim, and many more have been a support network as we have journeyed together through finding advisors, passing preliminary and comprehensive exams, finishing research, and now onward. Thank you to the Women in Network Science for being a consistent and supportive community when we worked from home. In particular, I would like to thank Alice Schwarze, Mari Kawakatsu, Leonie Neuhauser, Meg Patterson, John Meluso, Maria Ramos, Carolina Mattsson, Jennifer Lawlor. Thank you to the greater network science community, Phil Chodrow, Jean-Gabriel Young, Laurent Hébert-Dufresne,

and many others. To all of you, thank you for your friendship and support.

Thank you to the Applied Math staff and faculty. I would particularly like to thank Anne Dougherty and Silva Chang for helping me to become a more effective teacher. I'd like to thank Jim Curry for supporting and mentoring me throughout my academic career. Ian Grooms, Jem Corcoran, and many other faculty have taken an interest in my career, attended my talks, and encouraged my success. I have grown and learned so much because of them all.

Thank you to my lifelong friends, Rob Lark, Stacy Karthas, Johnson Yang, and Peter Somers, who have always available to chat despite long distance and widely different time zones.

Lastly, I would like to thank my family for their encouragement of my natural curiosity and support throughout my academic journey. I would like to acknowledge my wife, Shari, who has been incredibly supportive. Shari moved out to Colorado to support my dream of getting a PhD and for that, I am overwhelmingly grateful. I couldn't have done it without her.

## Contents

<b>Chapter</b>	
<b>1</b> Introduction	<b>1</b>
<b>2</b> Preliminaries	<b>6</b>
2.1 Hypergraphs . . . . .	6
2.2 Contagion models . . . . .	9
2.3 Notation . . . . .	10
<b>3</b> The effect of heterogeneity on hypergraph contagion models	<b>12</b>
3.1 Mean-Field Analysis . . . . .	13
3.1.1 Hyperedges of sizes 2 and 3 with collective contagion . . . . .	14
3.1.2 Hyperedges of sizes 2 and 3 with individual contagion . . . . .	24
3.1.3 Higher-order healing: hipster effect . . . . .	26
3.1.4 Unfortunate series of events . . . . .	28
3.2 The effect of degree distribution on $\beta_3^c$ . . . . .	29
3.3 Discussion . . . . .	33
<b>4</b> Hypergraph assortativity: a dynamical systems perspective	<b>36</b>
4.1 Motivation . . . . .	37
4.2 Mean-Field Approach . . . . .	39
4.2.1 Uncorrelated m-uniform case . . . . .	39

4.2.2	Derivation of the non-uniform uncorrelated expansion eigenvalue . . . . .	40
4.2.3	Perturbation approach for the correlated case . . . . .	42
4.3	Numerical Results . . . . .	44
4.3.1	Approximating the eigenvalue . . . . .	44
4.3.2	Extinguishing epidemics . . . . .	48
4.4	Discussion . . . . .	50
<b>5</b>	<b>Community structure in hypergraphs and the emergence of polarization</b>	<b>52</b>
5.1	Model . . . . .	54
5.1.1	The stochastic block model for uniform hypergraphs . . . . .	54
5.1.2	Planted partition model for uniform hypergraphs . . . . .	55
5.1.3	The degree-corrected stochastic block model for uniform hypergraphs . . . . .	57
5.1.4	Opinion model . . . . .	58
5.2	Mean-field analysis . . . . .	58
5.2.1	Planted partition model . . . . .	58
5.2.2	Imbalanced communities . . . . .	63
5.3	Discussion . . . . .	67
<b>6</b>	<b>Effect of time-dependent infectiousness on epidemic dynamics</b>	<b>68</b>
6.1	Model . . . . .	69
6.2	Derivation of the population reproductive number . . . . .	71
6.2.1	Fully-mixed population . . . . .	71
6.2.2	Discrete category-mixed population . . . . .	72
6.2.3	The continuum limit . . . . .	74
6.2.4	Examples . . . . .	75
6.2.5	Individual variation in the infectious rate function . . . . .	77
6.2.6	Numerical experiments . . . . .	79
6.3	Discussion . . . . .	82

<b>7</b>	<b>On limitations of uniplex networks for modeling multiplex contagion</b>	<b>84</b>
7.1	Motivation . . . . .	86
7.2	Data . . . . .	87
7.3	Contagion processes . . . . .	89
7.4	Approach . . . . .	89
7.5	Results . . . . .	90
7.6	Discussion . . . . .	94
<b>8</b>	<b>Open-source software</b>	<b>96</b>
8.1	CompleX Group Interactions (XGI) . . . . .	96
8.1.1	Data structure . . . . .	97
8.1.2	Conversion between hypergraph representations . . . . .	97
8.1.3	Algorithms . . . . .	98
8.1.4	Read/write functionality . . . . .	99
8.2	HyperContagion . . . . .	99
8.2.1	Epidemiological models . . . . .	100
8.2.2	Behavioral models . . . . .	101
8.3	CompleX Group Interactions Data (XGI-DATA) . . . . .	103
<b>9</b>	<b>Conclusion</b>	<b>106</b>
	<b>Bibliography</b>	<b>109</b>
	<b>Appendix</b>	
<b>A</b>	<b>Appendix for Chapter 3</b>	<b>121</b>
A.1	Microscopic simulation of the hypergraph SIS model . . . . .	121
A.2	Network models . . . . .	123



A.2.1	The numerical computation of $\beta_3^c$ . . . . .	124
<b>B</b>	Appendix for Chapter 4	<b>126</b>
B.1	More detailed derivation of the perturbed eigenvalue . . . . .	126
B.2	Suppressing epidemics through preferential rewiring . . . . .	128
B.3	Numerical simulations . . . . .	129
<b>C</b>	Appendix for Chapter 5	<b>132</b>
C.1	An efficient algorithm for sampling m-HSBM hypergraphs . . . . .	132
<b>D</b>	Appendix for Chapter 7	<b>135</b>
D.1	Full datasets . . . . .	135
D.2	Larger parameter ranges . . . . .	135
D.3	Different number of seeds . . . . .	136
D.4	Full temporal extent . . . . .	136

## Tables

### Table

2.1	Relevant notation . . . . .	11
4.1	Characteristics of the 3-uniform hypergraph datasets used. . . . .	46
7.1	Statistics of the chosen networks . . . . .	88

## Figures

### Figure

- 2.1 Illustration of a hypergraph. Infected nodes (red) infect a healthy node (grey) via hyperedges of sizes 2 and 3 with rates  $\beta_2$  and  $\beta_3$  respectively. . . . . 7
- 3.1 Schematic illustration of the degree-correlated and uncorrelated cases. In the degree-correlated case (left), nodes with more links are more likely to belong to a triangle. In the uncorrelated case (right), triangles connect nodes with a probability independent of their degree. . . . . 16
- 3.2 Fraction of infected nodes  $U$  versus link infectivity  $\beta_2$  obtained from the mean field equations (3.15)-(3.16) (solid and dashed lines) and from microscopic simulations (connected circles) using  $P(k) \propto k^{-4}$  on  $[67, 1000]$ ,  $\gamma = 2$ , and  $N = 10000$  for  $\beta_3 = 0.0194$  (a), 0.0388 (b), and 0.05482 (c). Refer to the text for an explanation of the discrepancy between the mean field equations and microscopic simulations. . . . 19
- 3.3 Phase diagram for the degree-correlated, collective contagion model. The light pink region labeled “No infection” corresponds to 1 solution of Eq. (3.10), the orange region labeled “Infection, no bistability” to 2 solutions, and the region labeled “Bistability” to 3 solutions. The parameters are  $\gamma = 2$  and  $P(k) \propto k^{-4}$  when  $67 < k < 1000$  and 0 otherwise. . . . . 21

3.4	Bistability index $B$ as a function of $\beta_3$ for (a) $P(k)$ constant for $50 < k < 150$ and 0 otherwise, (b) $P(k) \propto k^{-4}$ for $67 < k < 1000$ and 0 otherwise, and (c) $P(k) \propto k^{-3}$ for $53 < k < 1000$ and 0 otherwise. For each distribution, we considered the uncorrelated case (orange connected circles) and the degree correlated case (blue connected triangles). The dashed lines indicate the value $\beta_3^c$ at which we expect the onset of bistability, obtained from the numerical solution of the mean field equations (3.12) and (3.15)-(3.16). . . . .	22
3.5	Phase diagram for the degree-correlated, individual contagion model with parameters $\gamma = 2$ and $P(k) \propto k^{-4}$ when $67 < k < 1000$ and 0 otherwise. . . . .	25
3.6	Phase diagram for the degree-correlated, higher-order healing with individual contagion with parameters $\gamma = 2$ and $P(k) \propto k^{-4}$ when $67 < k < 1000$ and 0 otherwise. . . . .	28
3.7	$\beta_3^c/\beta_2^c$ as a function of power-law distribution parameters for the degree-correlated case (a) and the uncorrelated case (b). $\beta_3^c$ was calculated numerically from the mean field equations (see Appendix A.2.1), and $\beta_2^c = \gamma\langle k \rangle / \langle k^2 \rangle$ . The parameters are $P(k) \propto k^{-r}$ if $50 \leq k \leq k_{max}$ and $P(k) = 0$ otherwise and $\gamma = 2$ . . . . .	30
3.8	Relative error in the value of $\beta_3^c/\beta_2^c$ obtained from Eq. (3.43) compared with the numerically obtained value shown in Fig. 3.7(b). . . . .	33
4.1	An illustration of disassortative and assortative 3-uniform hypergraphs. The color of the nodes indicates their degree, with low-degree nodes on the left and high-degree nodes on the right. For a given degree sequence, the term $\langle kk_1 \rangle_E$ (the average pairwise product) determines $\rho$ and on average, (a) hyperedges containing nodes with dissimilar degrees decrease this term leading to disassortative hypergraphs and (b) hyperedges containing nodes of similar degree increase this term leading to assortative hypergraphs. . . . .	44

4.2	A comparison of the actual expansion eigenvalue $\lambda$ (connected triangles) to the first-order approximation of the eigenvalue $\lambda^{(0)} + \epsilon\lambda^{(1)}$ (connected circles) for (a) the configuration model, (b) the tags-ask-ubuntu dataset, (c) the congress-bills dataset, and (d) the Eu-Emails dataset. The square marker denotes the original $(\rho, \lambda)$ value of the dataset. Details of the characteristics of these datasets can be found in Table 4.1.	47
4.3	(a) The solid line with markers denotes the fixed value of $\beta_3$ as a fraction of the extinction threshold, $\beta_3/\beta_3^c$ . The dashed line indicates $\beta_3/\beta_3^c = 1$ , below which epidemics are not possible. (b) The epidemic equilibrium (percentage of the population infected) for each hypergraph given the fixed value of $\beta_3$ . The grey bars indicate the standard deviation at each data point.	49
5.1	A phase plot of Eqs. (5.7) - (5.8) with $\tilde{\epsilon}_2 = 0.5$ , $\tilde{\epsilon}_3 = 0.95$ , $\tilde{\beta}_2 = 0.2$ , and $\tilde{\beta}_3 = 4$ .	61
5.2	A plot of the polarization in Eqs. (5.7)-(5.8) for different values of $\epsilon_2$ and $\epsilon_3$ .	62
5.3	A phase plot of Eqs. (5.13) - (5.14) (the imbalanced case where $p_{1,1} = p_{2,2}$ ) with $\rho = 0.48$ , $\langle k \rangle = \langle q \rangle = 20$ , $\epsilon_2 = 10$ , $\epsilon_3 = 20$ , $\beta_2 = 0.2\beta_2^c$ , $\beta_3 = 4\beta_3^c$ , and $\gamma = 1$ .	66
6.1	An illustration of the VL model.	70
6.2	A plot showing how the number of infectious states affects (a) the time at which the infectious peak occurs and (b) the magnitude of the infectious peak for the viral load model in the fully-mixed case. We use two different infectious rate functions described in Section 6.2.6 and show the constant value of the SIR model (1 infectious stage) as a reference. For every data point, $R_0 = 3$ .	75
6.3	Time response of the fraction of infected individuals for different contagion models for (a) the configuration model and (b) the activity model. For both (a) and (b), the dash-dot, dashed, and solid lines indicate the VL model with $\beta_\Gamma(\tau) \propto \tau \exp(-\tau/4)$ , the VL model with $\beta_{const}(\tau) = \langle \beta_\Gamma(\tau) \rangle$ , and the SIR model with a single infection rate of $\beta$ respectively. $\beta = \langle \beta_\Gamma(\tau) \rangle$ and $\beta = 2m\langle a \rangle \langle \beta_\Gamma(\tau) \rangle$ for the configuration and activity models respectively. $R_0 = 3$ for each infection curve.	79

6.4	The number of individuals infected for duration $\tau$ at time $t$ for the configuration model (top) and the activity model (middle). The line plots (bottom) denote the probability distribution of $\tau$ at times 40, 60, and 80, which correspond to normalized vertical cross-sections of $I(t, \tau)$ . . . . .	81
6.5	The epidemic extent plotted as a function of the predicted reproductive number for different contagion models for (a) the configuration model and (b) the activity model. . . . .	82
7.1	A simple multiplex network. The line style corresponds with the relationship type. The node color represents its current status: infected (red), susceptible (white), and unknown (black). . . . .	85
7.2	Comparing the epidemic extent for different choices of network data for the Project 90 data set. . . . .	91
7.3	Comparing the epidemic extent for different choices of network data for the JOAPP data set. . . . .	93
8.1	Comparing the continuous time algorithms to the discrete-time algorithms for the epidemiological models on a Chung-Lu hypergraph of size 1000. . . . .	101
8.2	Examples of hypergraph opinion formation models. . . . .	102
8.3	An example datasheet for the email-Eu dataset. . . . .	105
A.1	Illustration of the bistability index with respect to the solutions to the mean-field equation in the bistable regime . . . . .	124
B.1	Additional plots of preferential rewiring and the corresponding epidemic response. Each subplot corresponds to a particular choice of infectious rate and dataset and follows the same format as Fig. 4.3. . . . .	129
D.1	Considering the full Project 90 and JOAPP data sets. For more details on these plots, see Figs. 7.2b and 7.3b. . . . .	137

D.2	Considering a wider range of contagion parameters. For more details on these plots, see Figs. 7.2b and 7.3b. . . . .	138
D.3	Varying the number of seed nodes for the SI model on the Project 90 data. For more details on these plots, see Figs. 7.2b and 7.3b. . . . .	139
D.4	Varying the number of seed nodes for the threshold model on the JOAPP data. For more details on these plots, see Figs. 7.2b and 7.3b. . . . .	139
D.5	Simulating contagion models until equilibrium is reached for all parameter values. For more details on these plots, see Figs. 7.2b and 7.3b. . . . .	140

# Chapter 1

## Introduction

Complex systems are important in representing real-world systems in that they can model the underlying structure of interactions. Accounting for this structure can offer important insights for real-world systems such as social networks, biological processes, social phenomena, opinion formation, and many other examples. Pairwise networks are a representation of complex systems where we assume that the system can be represented as a collection of entities (nodes) and connections between these entities (edges). When reducing a system to a network, key assumptions are made [1, 2] that simplify the system. One such assumption is that all entities interact only through pairwise interactions. Hypergraphs are a generalization of pairwise networks where interactions are no longer constrained to be between two nodes, but rather, can be of arbitrary size. For this reason, hypergraphs can more accurately model many real-world interaction patterns [3].

Modeling dynamics on hypergraphs can uncover rich behavior that one might not see if the dynamics simply occurred on a pairwise network [4]. It has been shown that higher-order interactions in networks (i.e., interactions involving multiple nodes) can have profound effects on dynamical network processes [5] such as opinion formation [6], biological processes [7], synchronization [8, 9, 10], population dynamics [7], and contagion processes [11, 12, 13]. In the case of contagion processes, it has been shown in Ref. [11] that the addition of higher-order interactions to the susceptible-infected-susceptible (SIS) epidemic model (called the *hypergraph SIS model* in Ref. [12]) results in bistability, where above a critical mass or “tipping point” of individuals, an epidemic will propagate and below which it will die out.



The focus of this dissertation is the interplay between the structure of a complex system, whether a hypergraph or a pairwise network, and the dynamics of a contagion process. The study of contagion processes is a fundamental problem in network science, with applications including epidemics [14, 15, 16, 17, 18, 19, 20], social media [21], opinion formation [22], idea diffusion [23, 24], sudden changes in social convention [25, 26], and many more. Contagion processes can be of many types, ranging from discrete-state models such as the SIS model, to continuous models of opinion formation, to realistic models of disease such as those currently used to model the spread of COVID-19 [27, 28]. Modeling the dynamics of such processes on pairwise interaction networks has been a hallmark of network science, providing many insights into the effect of network structure on the propagation of disease and information. For dynamical processes on hypergraphs, we primarily consider the hypergraph SIS model. In this simple model, with the addition of more complex interaction patterns that include higher-order interactions, rich behavior starts to emerge, including bistability, explosive transitions, hysteresis, and polarization.

One can also consider more complex contagion processes with, for instance, a time-dependent infection rate instead of the simplistic assumption that an agent transmits a contagion at a constant rate [29]. In this case, not only the structure of the underlying interaction network affects the dynamical behavior, but the “structure” of the contagion process as well.

In Chapter 2, we present common terminology and notation that will be used throughout this dissertation.

Chapters 3 to 5 explore the effect of hypergraph structure on simple higher-order contagion processes.

In Chapter 3 we present and analyze a hyperdegree-based mean-field description of the dynamics of the susceptible–infected–susceptible model on hypergraphs, i.e., networks with higher-order interactions, and illustrate its applicability with the example of a hypergraph where contagion is mediated by both links (pairwise interactions) and triangles (three-way interactions). We consider various models for the organization of link and triangle structures and different mechanisms of higher-order contagion and healing. We find that explosive transitions can be suppressed by hetero-

geneity in the link degree distribution when links and triangles are chosen independently or when link and triangle connections are positively correlated. We verify these results with microscopic simulations of the contagion process and with analytic predictions derived from the mean-field model. Our results show that the structure of higher-order interactions can have important effects on contagion processes on hypergraphs.

In Chapter 4 we present the *expansion eigenvalue*, an eigenvalue analogous to the largest eigenvalue of the matrix describing a network’s contact structure, for hypergraph dynamical processes. Using a mean-field approach, we derive an approximation to the expansion eigenvalue in terms of the degree sequence for uncorrelated hypergraphs. We introduce a generative model for hypergraphs that includes degree assortativity, and use a perturbation approach to derive an approximation to the expansion eigenvalue for assortative hypergraphs. We define the *dynamical assortativity*, a dynamically sensible definition of assortativity for uniform hypergraphs, and describe how reducing the dynamical assortativity of hypergraphs through preferential rewiring can extinguish epidemics. We validate our results with both synthetic and empirical datasets.

In Chapter 5 we model the propagation of an ideology with the hypergraph susceptible–infected–susceptible (SIS) model, where the ideology (represented with a binary variable) can spread via pairwise or group interactions through the majority vote process. We construct a hypergraph with nodes divided into two equally-sized communities and approximate our contagion model with a two-variable mean-field model. We find that for sufficiently strong community structure the two communities may hold very different average opinions when higher-order interactions are included. We determine the ranges of link and triangle community strengths for which polarization may occur. We also explore the effect of communities with different relative sizes.

Chapters 6 and 7 are on the importance of data and model choices when considering epidemics on pairwise networks.

In Chapter 6 we compare the average and individual reproductive numbers and epidemic dynamics for a model incorporating time-dependent infectiousness and a standard SIR (susceptible–infected–recovered) model for both fully mixed and category-mixed populations. We find that the

reproductive number depends only on the total infectious exposure and the largest eigenvalue of the mixing matrix and that these two effects are independent of each other. When we compare our time-dependent mean-field model to the SIR model with equivalent rates, the epidemic peak is advanced, and modifying the infection rate function has a strong effect on the time dynamics of the epidemic. We also observe behavior akin to a traveling wave as individuals transition through infectious states.

In Chapter 7 we demonstrate that multiplex spreading processes are not simply the union of spreading processes over their constituent uniplex networks. We use multiplex network data from two different contexts—one representing behavioral networks to represent their potential for infectious disease transmission using a simple epidemiological model, the other from online social network site users to represent their potential for information spread using a threshold-based spreading process. Our results show that spreading on multiplex data is not represented accurately in models developed from the uniplex networks even when they are combined, and that the nature of the differences between the (combined) uniplex and multiplex results depends on the specific spreading process over these networks.

In Chapter 8, we describe software that has been developed to support the algorithms and models developed in Chapters 3 to 5 and to support the larger research community studying complex systems with group interactions.

This dissertation is based on the following publications:

- Chapter 3:

**Nicholas W. Landry**, Juan G. Restrepo, *The effect of heterogeneity on hypergraph contagion models*, Chaos, 2020. DOI: [10.1063/5.0020034](https://doi.org/10.1063/5.0020034)

- Chapter 4:

**Nicholas W. Landry**, Juan G. Restrepo, *Hypergraph assortativity: a dynamical systems perspective*, Preprint, 2021. [arXiv:2109.01099](https://arxiv.org/abs/2109.01099)

- Chapter 5:

**Nicholas W. Landry**, Juan G. Restrepo, *Community structure in hypergraphs and the emergence of polarization*, In Preparation, 2022

- Chapter 6:

**Nicholas W. Landry**, *Effect of time-dependent infectiousness on epidemic dynamics*, Physical Review E, 2021. DOI: [10.1103/PhysRevE.104.064302](https://doi.org/10.1103/PhysRevE.104.064302)

- Chapter 7:

**Nicholas W. Landry**, jimi adams, *On limitations of uniplex networks for modeling multiplex contagion*, In Preparation, 2022

## Chapter 2

### Preliminaries

In this Chapter, we present definitions and background material that we will draw from in subsequent Chapters.

#### 2.1 Hypergraphs

A hypergraph is a mathematical object that can represent group interactions among a set of nodes. We represent it as  $H = (V, E)$ , where  $V$  is the set of nodes and  $E$  is the set of hyperedges, which are subsets of  $V$  and represent interactions of arbitrary size. If a hyperedge has cardinality  $m$ , we call it an  $m$ -hyperedge. When a hypergraph only contains hyperedges of size  $m$ , we call it an  $m$ -uniform hypergraph. It is also useful to consider weighted hypergraphs, where each hyperedge  $e$  has an associated positive weight  $\beta_e$ . We consider a population of  $N$  nodes labeled  $i = 1, 2, \dots, N$  coupled via undirected hyperedges of sizes  $m = 2, 3, \dots, M$ , where a hyperedge of size  $m$  is a set of  $m$  nodes,  $\{i_1, i_2, \dots, i_m\}$ . We define the  $m$ -th order degree of node  $i$ ,  $k_i^{(m)}$ , as the number of hyperedges of size  $m$  to which the node belongs, and its hyperdegree as the vector  $\mathbf{k}_i = [k_i^{(2)}, k_i^{(3)}, \dots, k_i^{(M)}]$ . The 2nd order degree of a node corresponds to the number of pairwise connections of the node, while higher-order degrees measure the node's participation in hyperedges of larger sizes. Figure 2.1 illustrates a hypergraph with hyperedges of sizes 2 and 3, which, for simplicity, we denote as *links* and *triangles* respectively.

For a network or hypergraph, when interactions between nodes are of different types (e.g., Twitter and Facebook connections; train, plane, and bus routes; etc.), we can represent this com-

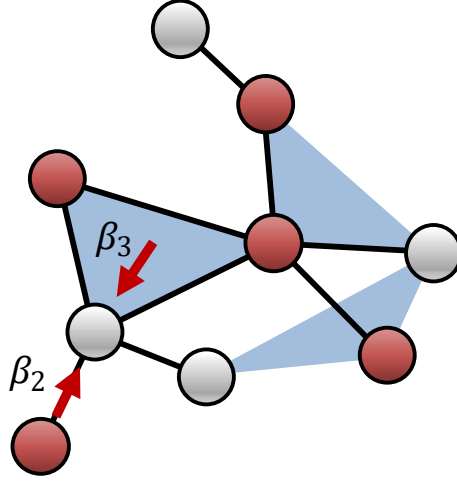


Figure 2.1: Illustration of a hypergraph. Infected nodes (red) infect a healthy node (grey) via hyperedges of sizes 2 and 3 with rates  $\beta_2$  and  $\beta_3$  respectively.

plex system with a *multilayer* network or hypergraph, where each layer corresponds to a different interaction type. When the node set in each layer is identical, it is called a *multiplex* network or hypergraph. For a dynamical process where the nature of the interactions between nodes depends on the cardinality of the hyperedges representing that interaction, it can be sensible to separate the hyperedges by size. Any hypergraph  $H = (V, E)$  can be partitioned into a set of  $m$ -uniform hypergraphs,  $H_m = (V, E_m)$ , where  $E_m = \{\tau \in E \mid |\tau| = m\}$ . We define the *hyperunion* of a finite number of hypergraphs with the same set of nodes  $V$  to be  $\bigcup_m H_m = (V, \bigcup_m E_m)$ . Then  $H = \bigcup_m H_m$  and we can examine each  $H_m$  separately. We also define the  $m$ -th *moment* of a quantity  $\mathbf{q} = [q_1, \dots, q_N]$  associated with each node as  $\langle q^m \rangle = \sum_{i=1}^N q_i^m / N$ .

In this dissertation, we exclusively consider mean-field approaches to the prediction of dynamical behavior. Extending degree-based descriptions of epidemic spreading on networks [30, 31], we will develop a mean-field theory for the propagation of epidemics based on the assumption that nodes with the same hyperdegree have the same statistical properties. We define  $N(\mathbf{k})$  to be the number of nodes with hyperdegree  $\mathbf{k}$  such that  $P(\mathbf{k}) = N(\mathbf{k})/N$  is the degree distribution. For this purpose, we assume that  $N(\mathbf{k})$  is given, and that the probability that nodes with hyperdegrees  $\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_m$  belong to a hyperedge of size  $m$  is given by  $f_m(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_m)$ . This assumes that the

statistical structure of the network is completely described by the hyperdegree distribution  $P(\mathbf{k})$  and the connection probabilities  $f_m(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_m)$ . This is an example of a statistical model for hypergraphs from which one can sample, also known as a *generative model*. Note that, counting the number of hyperedges of size  $m$  in two different ways, the connection probabilities must be normalized such that

$$\frac{1}{m!} \sum_{\mathbf{k}_1, \dots, \mathbf{k}_m} N(\mathbf{k}_1) \dots N(\mathbf{k}_m) f_m(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_m) = \frac{1}{m} \sum_{\mathbf{k}} k^{(m)} N(\mathbf{k}). \quad (2.1)$$

This expression states that the expected number of hyperedges predicted with the connection probability function  $f_m$  must equal the number of hyperedges of size  $m$  specified by the non-normalized degree distribution  $N(\mathbf{k})$ . For example, for the configuration model for networks without higher-order interactions (i.e., only hyperedges of size 2,  $M = 2$ ), the hyperdegree of a node is just the number of links,  $\mathbf{k} = k$ , connecting that node to other nodes and the connection probability is  $f_2(k, k') = kk'/(N\langle k \rangle)$ , where  $\langle k \rangle = \sum_{i=1}^N k_i/N = \sum_k kP(k)$ . For hypergraphs with hyperedges of sizes 2 and 3,  $f_3(k_1, k_2, k_3)$  is the probability that three nodes with degrees  $k_1$ ,  $k_2$ , and  $k_3$  are connected by a hyperedge of size 3. The configuration model for hypergraphs and its associated statistical properties has been studied in Refs. [32, 33] and the associated probability of a size- $m$  hyperedge is  $f_m(k_{i_1}, \dots, k_{i_m}) = f_m^{(0)}(k_{i_1}, \dots, k_{i_m}) = (m-1)!k_{i_1} \dots k_{i_m}/(N\langle k \rangle)^{m-1}$ .

This generative model can produce hypergraphs with heterogeneity in the node hyperdegrees and correlations between hyperdegrees of connected nodes. It can also be easily generalized to account for mixing by additional nodal variables such as community labels or dynamical parameters. Its main limitation is that it does not capture connection patterns that are determined by structures beyond a node's immediate connections (e.g., the model cannot account for hyperedges of size 3 that occur only when there is a clique of 3 nodes connected by links, as one would see in a simplicial complex). Nevertheless, this generative model is a versatile and tractable null model to explore the effect of hypergraph structure on various hypergraph metrics and dynamical behavior. Table 2.1 summarizes the notation and variables presented in this section.

## 2.2 Contagion models

Now we describe the contagion models we will study. In this dissertation, we primarily consider *discrete-state compartment models* in contrast to other models which allow a node's state to be a continuous variable [34, 35]. In these models, we track the number of nodes that have a given opinion or disease state, i.e., infected, susceptible, etc. and describe the rates governing the transitions between these discrete states.

Among the most common discrete-state contagion models are the susceptible – infected (SI), susceptible – infected – susceptible (SIS), and the susceptible – infected – recovered (SIR) models. These models describe contagion that one can adopt and then reject (or contract and recover from) in a cyclic manner and one which an individual adopts only temporarily (or contracts, heals, and becomes immune to) respectively.

For the SIS model, we assume that at any given time  $t \geq 0$ , each node can be in either the susceptible (S) or infected (I) state. Infected nodes heal and become susceptible again at rate  $\gamma$ . Now we specify how hyperedges mediate the contagion process. In contrast to pairwise networks where a susceptible individual can only be exposed to a single infected individual at a time through its interactions, in hypergraphs a susceptible individual can be part of a larger group with many different possibilities of infection statuses. In general, the probability of contagion by a hyperedge could be a function of the number of infected nodes in the hyperedge (e.g., as in Refs. [36, 13]). However, we will only consider the two extreme cases where contagion occurs if all the other members of the hyperedge are infected, or if at least one member of the hyperedge is infected. More precisely, in the *collective contagion* case, a susceptible node that belongs to a hyperedge of size  $m$  gets infected at rate  $\beta_m$  if *all* the other members of the hyperedge are infected; in the *individual contagion* case the node gets infected at rate  $\beta_m$  if *at least* one member is infected. While we will analyze these two cases only, in principle one could treat the case in which at least  $j$  other nodes of the hyperedge need to be infected for contagion to occur using the techniques presented below. This case corresponds to a quorum of size  $j$  and there is evidence for such effects in collective



behavior [37, 38]. For hyperedges of size 2, i.e., links, both cases reduce to the usual contagion via pairwise interactions. The social contagion model of Ref. [11] corresponds to the collective contagion case. The collective contagion process is illustrated in Figure 2.1 for hyperedges of sizes 2 and 3. Applying these contagion models to pairwise networks is a special case of the general framework described above. In this case, we no longer need to consider the fraction of hyperedge neighbors that may transmit that contagion because each edge contains a single neighbor.

The SI model is the same as the SIS model with the key difference that an infected node does not recover. The SIR model is similar to the SIS model in that infected nodes infect susceptible nodes and infected nodes recover, but with the key difference that once a node is infected, it can develop immunity to that contagion and cannot be re-infected.

Empirical studies have shown that *complex contagion*, where a node requires multiple exposures to become infected, can drive behavioral adoption [26]. This type of process has been described in pairwise networks by, for example, the threshold contagion model [24, 22]. This simple model can be described as follows: Consider a pairwise network with a two-state (“0” and “1”) contagion process spreading on top of it. If a node holds the “0” opinion, it will be convinced of the opposite opinion if greater than a critical fraction  $\tau$  of its pairwise neighbors hold the “1” opinion. This process can also be described more generally using hypergraphs where contagion spreads if a critical fraction of members of a hyperedge hold the “1” opinion.

A natural extension of these simple contagion models is to consider time-dependent infectiousness. An individual’s infectiousness varies over the duration of the infection according to their viral load [39, 40]. One can divide the single I compartment into  $k$  stages, each of which has their own associated infectiousness [29, 41] to model the effect that viral load has on epidemic dynamics. This model is commonly known as the SI<sup>k</sup>R model [16].

Table 2.1 summarizes the notation and variables presented in this section.

### 2.3 Notation

This is a brief summary of the notation used in this dissertation.

<b>Variable</b>	<b>Definition</b>
$H$	Hypergraph
$V$	Set of nodes (vertices) of a hypergraph
$E$	Set of hyperedges of a hypergraph
$N$	Number of nodes
$k^{(m)}$	Number of hyperedges of size $m$ to which a node belongs
$\mathbf{k} = [k^{(2)}, \dots, k^{(M)}]$	hyperdegree
$N(\mathbf{k})$	Number of nodes with hyperdegree $\mathbf{k}$
$P(\mathbf{k})$	Probability of a node with hyperdegree $\mathbf{k}$
$\gamma$	Rate of healing
$\beta_m$	Rate of infection by a hyperedge of size $m$
$f_m(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_m)$	Probability that $m$ nodes form a hyperedge of size $m$
$x_{\mathbf{k}}$	Fraction of nodes with hyperdegree $\mathbf{k}$ that are infected
$\tau$	Threshold process adoption threshold

Table 2.1: Relevant notation

## Chapter 3

### The effect of heterogeneity on hypergraph contagion models

It was shown in Ref. [11] that the addition of higher-order interactions to the SIS epidemic model on Erdős-Rényi networks results in bistability, hysteresis, and explosive transitions to an endemic disease state (see also Refs. [42, 43, 36, 44]). The simplicial SIS model has also been extended to scale-free uniform hypergraphs [45]. The fact that the network SIS model with more general higher-order interactions results in bistability has been proven rigorously in Ref. [43]. However, so far there is no general theory explaining how heterogeneity and correlations in the structure of higher-order interactions affect the onset of bistability.

In this Chapter, we present and analyze a degree-based mean-field description of the dynamics of the SIS model in networks with higher-order interactions. To describe higher-order interactions we consider the SIS model on a *hypergraph*, formed by a set of nodes and a set of edges of multiple sizes (so that edges of size larger than two represent higher-order interactions). Our formulation allows us to consider heterogeneous structure in the organization of the edges of a given size, and correlations between the structure of edges of different sizes. Using the illustrative case of networks with edges of sizes 2 and 3, we derive conditions for the appearance of bistability and hysteresis in terms of moments of the degree distribution of the pairwise interaction network. We find that the onset of bistability and hysteresis can be suppressed by heterogeneity in the pairwise interaction network and promoted by positive correlations between the number of pairwise and higher-order interactions a node has. We also consider the effect of healing by higher-order interactions (a “hipster effect”).

The structure of this Chapter is as follows. In Sec. 3.1 we derive a mean-field description of the model and apply it to various illustrative cases. In Sec. 3.2 we study how model parameters affect the onset of bistability. Finally, we discuss our results and present our conclusions in Sec. 3.3.

### 3.1 Mean-Field Analysis

In this section we present a mean-field analysis of the epidemic dynamics on a network specified by the hyperdegree distribution  $P(\mathbf{k}) = N(\mathbf{k})/N$  and the hyperedge connection probabilities  $f_m(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_m)$ . Assuming that all nodes with the same hyperdegree behave similarly, we focus on  $x_{\mathbf{k}}$ , the fraction of nodes with hyperdegree  $\mathbf{k}$  that are infected. The mean-field equation describing the evolution of  $x_{\mathbf{k}}$  is

$$\frac{dx_{\mathbf{k}}}{dt} = -\gamma x_{\mathbf{k}} + (1 - x_{\mathbf{k}}) \sum_{m=2}^M \beta_m \frac{1}{(m-1)!} \sum_{\mathbf{k}_1, \dots, \mathbf{k}_{m-1}} \prod_{l=1}^{m-1} N(\mathbf{k}_l) f_m(\mathbf{k}, \mathbf{k}_1, \dots, \mathbf{k}_{m-1}) G(x_{\mathbf{k}_1}, \dots, x_{\mathbf{k}_{m-1}}) \quad (3.1)$$

$$G(x_{\mathbf{k}_1}, \dots, x_{\mathbf{k}_{m-1}}) = \begin{cases} \prod_{l=1}^{m-1} x_{\mathbf{k}_l}, & \text{collective contagion,} \\ 1 - \prod_{l=1}^{m-1} (1 - x_{\mathbf{k}_l}), & \text{individual contagion.} \end{cases} \quad (3.2)$$

The first term on the right-hand side of Eq. (3.1) corresponds to healing at rate  $\gamma$  and the second term accounts for infection by hyperedges. The number of hyperedges of size  $m$  that can pass an infection to a node with hyperdegree  $\mathbf{k}$  is calculated by considering all the possible hyperdegrees of the other  $m-1$  nodes participating in the hyperedge  $(\mathbf{k}, \mathbf{k}_1, \dots, \mathbf{k}_{m-1})$ , counting how many such combinations there are not counting permutations  $[N(\mathbf{k}_1) \cdots N(\mathbf{k}_{m-1}) / (m-1)!]$ , calculating what fraction of such combinations form a hyperedge with the node in consideration  $[f_m(x_{\mathbf{k}}, x_{\mathbf{k}_1}, \dots, x_{\mathbf{k}_{m-1}})]$ , multiplying by the probability that the hyperedge can transmit the infection  $[G(x_{\mathbf{k}_1} \cdots x_{\mathbf{k}_{m-1}})]$ , and summing over all hyperdegree combinations. The probability that the hyperedge can transmit the infection, given by (3.2), depends on whether the collective contagion or individual contagion model is assumed. Note that the form for  $G$  taken above, and the mean field

treatment in general, assume that the states of nodes are independent. A better approximation that includes correlations between connected nodes has been implemented in Refs. [44, 46] for the case of unstructured hyperedges of sizes 2 and 3, leading to improved quantitative agreement with the results of numerical simulations. Since our interest is in the effects of higher-order structures on qualitative aspects of the epidemic dynamics, we will use the mean field approximation in Eq. (3.1). A similar mean-field equation for a node-based description of the contagion process was recently formulated in Ref. [43]. In the following we will apply the mean-field description to illustrative cases.

### 3.1.1 Hyperedges of sizes 2 and 3 with collective contagion

Here we focus on the case where the hyperedge sizes are either 2 or 3, i.e.,  $M = 3$ . This corresponds to a network like in Fig. 2.1, with hyperedges of size 2 (links) and 3 (triangles). For simplicity, here we denote the number of links per node as  $k$ , i.e.,  $k = k^{(2)}$ , and the number of triangles a node belongs to by  $q$ , i.e.,  $q = k^{(3)}$ . In addition, we will consider the case where the connection probabilities depend only on the node links, i.e.,  $f_m(\mathbf{k}, \mathbf{k}_1, \dots, \mathbf{k}_{m-1}) = f_m(k, k_1, \dots, k_{m-1})$ . With these assumptions, and using the collective contagion rule in Eq. (3.2), Eq. (3.1) becomes

$$\frac{dx_{k,q}}{dt} = -\gamma x_{k,q} + (1 - x_{k,q})\beta_2 \sum_{k_1, q_1} N(k_1, q_1) f_2(k, k_1) x_{k_1, q_1} \quad (3.3)$$

$$+ (1 - x_{k,q}) \frac{\beta_3}{2} \sum_{k_1, q_1, k_2, q_2} N(k_1, q_1) N(k_2, q_2) f_3(k, k_1, k_2) x_{k_1, q_1} x_{k_2, q_2}, \quad (3.4)$$

where the first term on the right hand side represents healing, the second represents contagion by links, and the third represents contagion by triangles.

Since the connection probabilities do not depend on  $q$ , we can reduce the dynamics to the fraction of nodes with degree  $k$  that are infected,

$$x_k = \frac{\sum_q N(k, q) x_{k,q}}{N(k)}, \quad (3.5)$$

where  $N(k) = \sum_q N(k, q)$  is the number of nodes with degree  $k$ . Multiplying Eq. (3.3) by  $N(k, q)$ ,

summing over  $q$  and dividing by  $N(k)$ , we obtain

$$\begin{aligned} \frac{dx_k}{dt} = & -\gamma x_k + (1 - x_k)\beta_2 \sum_{k_1} N(k_1) f_2(k, k_1) x_{k_1} \\ & + (1 - x_k) \frac{\beta_3}{2} \sum_{k_1, k_2} N(k_1) N(k_2) f_3(k, k_1, k_2) x_{k_1} x_{k_2}. \end{aligned} \quad (3.6)$$

For the link connection probability  $f_2(k, k_1)$ , we will take  $f_2(k, k_1) = k k_1 / (N \langle k \rangle)$ , which corresponds to nodes being connected completely at random according to their degree as in the configuration model. For the triangle connection probability  $f_3$ , we will consider two cases: the *uncorrelated* case and the *degree-correlated case*. In the degree-correlated case, we assume that the connection probability is given by  $f_3(k, k_1, k_2) = 2k k_1 k_2 / (N \langle k \rangle)^2$ , so that nodes that have a higher number of links also belong to more triangles. In the uncorrelated case, we assume instead that  $f_3(k, k_1, k_2) = 2 \langle k \rangle / N^2$ , so that triangles are formed independent of the nodal degrees. The normalization is chosen using Eq. (2.1) so that the mean number of triangles per node,  $\langle q \rangle = \sum_{i=1}^N k_i^{(3)} / N$ , in each case is equal to  $\langle k \rangle$ . We note that the model for triangle formation in Ref. [11] corresponds to the uncorrelated case. We can choose the mean triangle degree independent of the mean network degree by scaling  $f_3(k, k_1, k_2)$  by  $\langle q \rangle / \langle k \rangle$ , but for simplicity, we assume  $\langle q \rangle = \langle k \rangle$ . Figure 3.1 illustrates the difference between the two cases in a small network, where in the degree-correlated case, the triangles cluster around nodes with high pairwise degree, and in the uncorrelated case, the triangles are distributed uniformly at random on the network.

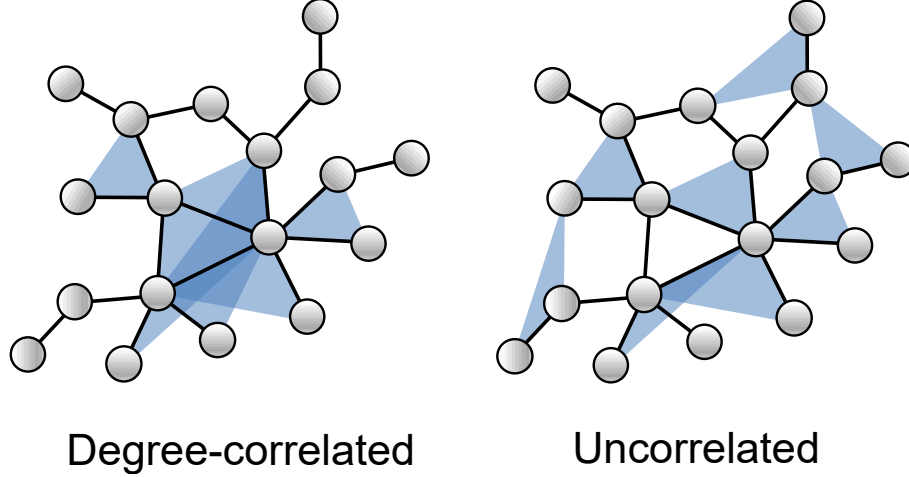


Figure 3.1: Schematic illustration of the degree-correlated and uncorrelated cases. In the degree-correlated case (left), nodes with more links are more likely to belong to a triangle. In the uncorrelated case (right), triangles connect nodes with a probability independent of their degree.

We can also specify the distribution of triangle degrees by defining  $f_2(q, q_1)$  and  $f_3(q, q_1, q_2)$  and then reducing Eq. (3.3) by multiplying by  $N(k, q)$ , dividing by  $N(q)$ , and summing over  $k$  to reduce the dynamics to the fraction of infected nodes with triangle degree  $q$ . For the triangle connection probability, we take  $f_3(q, q_1, q_2) = 2qq_1q_2/(N\langle q \rangle)^2$  which corresponds to three nodes being connected at random according to the configuration model for triangles [32]. For the pairwise links, we define the degree-correlated and uncorrelated cases as before, where in the degree-correlated case,  $f_2(q, q_1) = qq_1/(N\langle q \rangle)$ , and for the uncorrelated case,  $f_2(q, q_1) = \langle q \rangle/N$ . From there, we can use the same formalism as our approach when specifying the pairwise degree.

Now we consider separately the degree-correlated and uncorrelated cases. In the correlated case, where  $f_3(k, k_1, k_2) = kk_1k_2/(N\langle k \rangle)^2$ , Eq. (3.6) can be rewritten in terms of the fraction of infected links

$$V = \sum_k \frac{kN(k)x_k}{N\langle k \rangle} = \sum_k P(k) \frac{kx_k}{\langle k \rangle} \quad (3.7)$$

as

$$\frac{dx_k}{dt} = -\gamma x_k + \beta_2(1-x_k)kV + \beta_3(1-x_k)kV^2. \quad (3.8)$$

In this case, the dynamics of nodes of degree  $k$  is determined by the global variable  $V$ . To

study the qualitative characteristics of the dynamics, we find the steady-state solutions. The fixed point of Eq. (3.8) is

$$x_k = \frac{\beta_2 k V + \beta_3 k V^2}{\gamma + \beta_2 k V + \beta_3 k V^2}. \quad (3.9)$$

Inserting this in (3.7), we obtain a nonlinear equation that determines the fraction of infected links  $V$ :

$$V = \frac{1}{N \langle k \rangle} \sum_k \frac{k N(k) (\beta_2 k V + \beta_3 k V^2)}{\gamma + \beta_2 k V + \beta_3 k V^2}. \quad (3.10)$$

The state with no infection,  $V = 0$ , is a solution to (3.10). However, it is linearly unstable for  $\beta_2 > \beta_2^c = \gamma \langle k \rangle / \langle k^2 \rangle$ , as can be seen by linearizing Eq. (3.8) about  $V = 0$ , multiplying by  $k N(k) / (N \langle k \rangle)$ , and summing over  $k$ , which yields the linearized equation for the evolution of the perturbation  $\delta V$

$$\frac{d\delta V}{dt} = -\gamma \delta V + \beta_2 \frac{\langle k^2 \rangle}{\langle k \rangle} \delta V. \quad (3.11)$$

The nonzero solutions of Eq. (3.10) represent states with a nonzero fraction of infected nodes.

Now we study the uncorrelated case where  $f_3(k, k_1, k_2) = 2 \langle k \rangle / N^2$ . In this case, Eq. (3.6) can be rewritten in terms of the fraction of infected nodes

$$U = \sum_k \frac{N(k) x_k}{N} = \sum_k P(k) x_k, \quad (3.12)$$

and the fraction of infected links  $V$ . In terms of these quantities, Eq. (3.6) reads

$$\frac{dx_k}{dt} = -\gamma x_k + \beta_2 (1 - x_k) k V + \beta_3 (1 - x_k) \langle k \rangle U^2. \quad (3.13)$$

As in the prior case, the equilibrium is

$$x_k = \frac{\beta_2 k V + \beta_3 \langle k \rangle U^2}{\gamma + \beta_2 k V + \beta_3 \langle k \rangle U^2}. \quad (3.14)$$

Evaluating this expression in Eqs. (3.7) and (3.12) we obtain the coupled equations

$$U = \frac{1}{N} \sum_k \frac{N(k) (\beta_2 k V + \beta_3 \langle k \rangle U^2)}{\gamma + \beta_2 k V + \beta_3 \langle k \rangle U^2}, \quad (3.15)$$

$$V = \frac{1}{N \langle k \rangle} \sum_k \frac{k N(k) (\beta_2 k V + \beta_3 \langle k \rangle U^2)}{\gamma + \beta_2 k V + \beta_3 \langle k \rangle U^2}. \quad (3.16)$$



The state with no infection,  $U = 0$ ,  $V = 0$ , is a solution of (3.15)-(3.16). By considering perturbations  $\delta U$ ,  $\delta V$  from this solution, linearizing Eq. (3.13), and evaluating in Eq. (3.7) for the first equation and Eq. (3.12) for the second equation, we obtain the linear system

$$\frac{d\delta V}{dt} = -\gamma\delta V + \beta_2 \frac{\langle k^2 \rangle}{\langle k \rangle} \delta V, \quad (3.17)$$

$$\frac{d\delta U}{dt} = -\gamma\delta U + \beta_2 \langle k \rangle \delta V, \quad (3.18)$$

which shows that the no infection state is linearly unstable for  $\beta_2 > \gamma \langle k \rangle / \langle k^2 \rangle$ , which is the same threshold we obtained for the correlated case.

In summary, nonzero solutions of Eq. (3.10) and Eqs. (3.15)-(3.16) for the degree-correlated and uncorrelated cases, respectively, represent states with a nonzero number of infected nodes. Figure 3.2 shows the fraction of infected nodes  $U$  for the uncorrelated case as a function of the normalized pairwise infectivity  $\beta_2/\beta_2^c$  for three values of the triangle infectivity  $\beta_3$  obtained from numerical solution of Eqs. (3.15)-(3.16) with  $P(k) \propto k^{-4}$  for  $67 < k < 1000$  and 0 otherwise. Different solutions are plotted as solid and dashed lines to indicate stability or instability, respectively. The connected circles are obtained from numerical simulations of the full stochastic microscopic model. In these simulations  $\beta_2$  was slowly increased in small steps up to a maximum value, and subsequently decreased back to its initial value. For each  $\beta_2$ , the average number of infected nodes after transient effects disappeared is shown as a filled circle. For more details about the simulations, see Appendix A.

The behavior of the microscopic simulations is captured qualitatively by the mean field equations. The quantitative disagreement is likely due to the assumptions inherent to the mean-field approximation. In fact, Ref. [44] has shown that, for the particular case of uncorrelated triangles on an Erdős-Rényi network, the disagreement almost disappears when pair correlations are taken into account. Since our interest in this Chapter is on the qualitative dynamics, we use the mean-field theory, but note that the approaches proposed in Refs. [44, 47] could be used to obtain better approximations. The qualitative aspects of interest, captured by the mean field equations and the numerical solution of Eqs. (3.15)-(3.16), are the following.

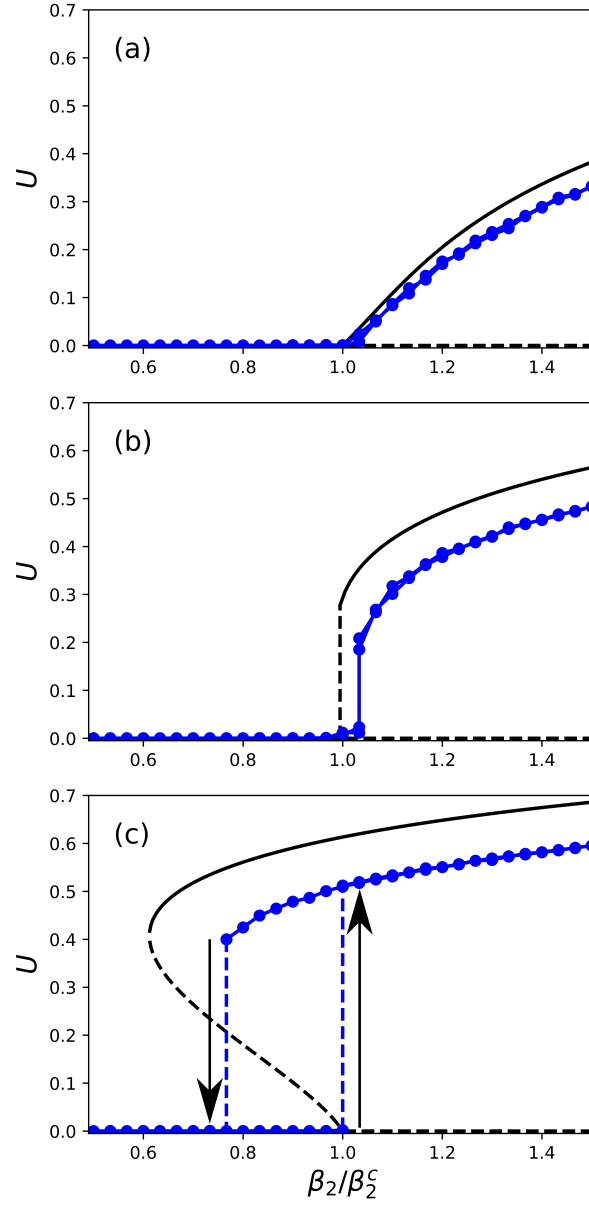


Figure 3.2: Fraction of infected nodes  $U$  versus link infectivity  $\beta_2$  obtained from the mean field equations (3.15)-(3.16) (solid and dashed lines) and from microscopic simulations (connected circles) using  $P(k) \propto k^{-4}$  on  $[67, 1000]$ ,  $\gamma = 2$ , and  $N = 10000$  for  $\beta_3 = 0.0194$  (a), 0.0388 (b), and 0.05482 (c). Refer to the text for an explanation of the discrepancy between the mean field equations and microscopic simulations.

For small values of  $\beta_3$  [Fig. 3.2(a),  $\beta_3 = 0.0194$ ] the bifurcation from the state with no infection ( $U = 0$ ) to the infected state ( $U > 0$ ) is continuous. However, for larger values of  $\beta_3$  [Fig. 3.2(c),  $\beta_3 = 0.0582$ ], the transition is discontinuous: as  $\beta_2$  increases past a critical value  $\beta_2^c$ ,

the fraction of infected links increases explosively towards an epidemic equilibrium (upward arrow). If  $\beta_2$  is subsequently decreased, the fraction of infected links remains high until  $\beta_2$  decreases past the value at which the epidemic equilibrium solution disappears, and then it decreases to zero (downward arrow). For such values of  $\beta_3$ , there is hysteresis, bistability, and explosive transitions. At a critical value  $\beta_3 = \beta_3^c$ , which will be the focus of our interest, there is a transition from the type of bifurcation shown in Fig 3.2(a) to the type of bifurcation shown in Fig. 3.2(c). Fig. 3.2(b) shows  $U$  as a function of  $\beta_2$  for a value  $\beta_3 = 0.0388 \approx \beta_3^c$ . We are interested in exploring how the presence of this bistable regime is affected by the degree distribution  $P(k)$  and other parameters of the model, in particular the triangle infectivity,  $\beta_3$ .

Figure 3.3 shows the phase diagram in the  $(\beta_2, \beta_3)$  plane for the degree-correlated, collective contagion model. The plot was obtained by counting the number of solutions of Eq. (3.10) as a function of  $\beta_2$  and  $\beta_3$  for  $\gamma = 2$ , and  $P(k) \propto k^{-4}$  when  $67 < k < 1000$  and 0 otherwise (all subsequent phase diagram plots are calculated using the same parameters). Light pink indicates that there is only the solution  $V = 0$  corresponding to a stable state with no contagion. Orange indicates two solutions, the unstable  $V = 0$  solution and another stable solution with  $V > 0$ . Finally, dark red indicates a bistable regime with three solutions: the stable  $V = 0$  solution, and a pair of stable and unstable solutions with positive  $V$ . As noted in Refs. [11, 43], this regime is only present for large enough triangle infectivity, i.e., for  $\beta_3 > \beta_3^c$ . The phase space for the uncorrelated case (not shown) is qualitatively similar to the one in Fig. 3.3, but the transition to bistable behavior occurs at a larger value of  $\beta_3$ .

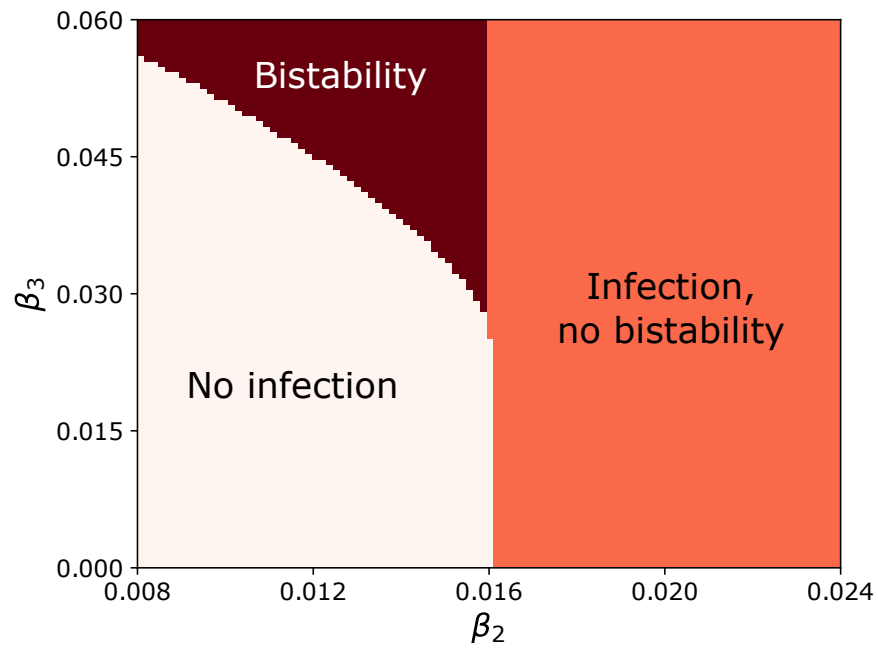


Figure 3.3: Phase diagram for the degree-correlated, collective contagion model. The light pink region labeled “No infection” corresponds to 1 solution of Eq. (3.10), the orange region labeled “Infection, no bistability” to 2 solutions, and the region labeled “Bistability” to 3 solutions. The parameters are  $\gamma = 2$  and  $P(k) \propto k^{-4}$  when  $67 < k < 1000$  and 0 otherwise.

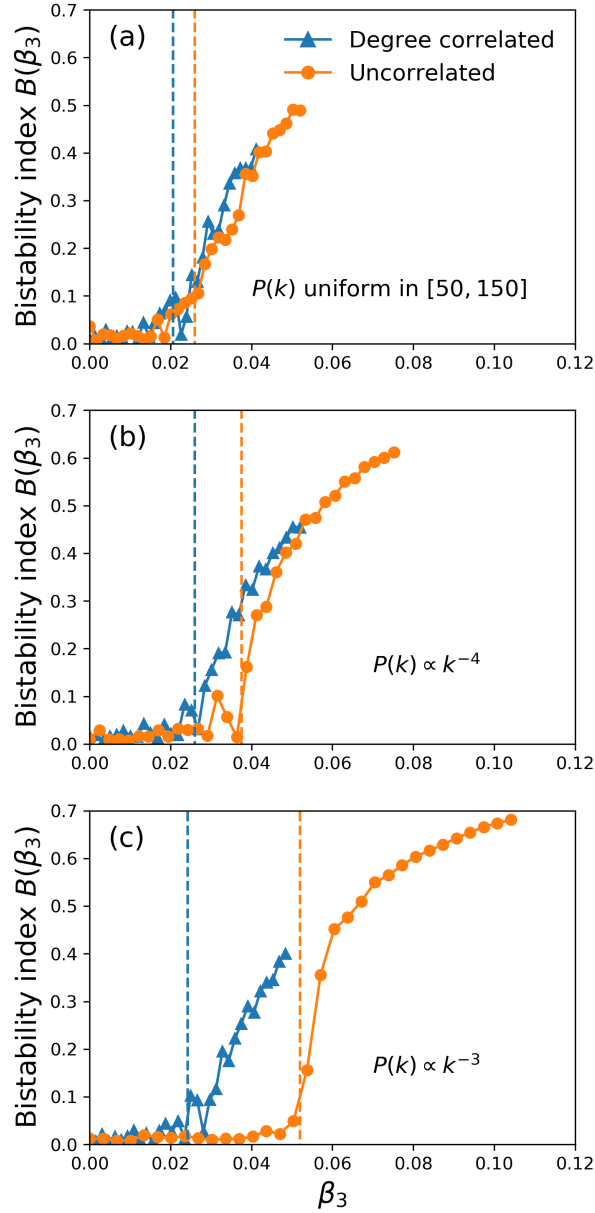


Figure 3.4: Bistability index  $B$  as a function of  $\beta_3$  for (a)  $P(k)$  constant for  $50 < k < 150$  and 0 otherwise, (b)  $P(k) \propto k^{-4}$  for  $67 < k < 1000$  and 0 otherwise, and (c)  $P(k) \propto k^{-3}$  for  $53 < k < 1000$  and 0 otherwise. For each distribution, we considered the uncorrelated case (orange connected circles) and the degree correlated case (blue connected triangles). The dashed lines indicate the value  $\beta_3^c$  at which we expect the onset of bistability, obtained from the numerical solution of the mean field equations (3.12) and (3.15)-(3.16).

To quantify how the onset of bistability depends on the hypergraph parameters, we define the *bistability index*  $B(\beta_3)$  as the maximum separation, over all values of  $\beta_2$ , between the largest

and smallest stable solutions for the fraction of infected nodes  $U$ . The bistability index can be calculated from microscopic simulations of the contagion process such as those used to produce Fig. 3.2, or from numerical solution of Eq. (3.10) for the correlated case and Eqs. (3.15)-(3.16) for the uncorrelated case. In Figure 3.4 we plot the bistability index  $B$  as a function of  $\beta_3$  computed from microscopic simulations for three choices of the link degree distribution  $P(k)$ , all with a mean degree of 100: (a)  $P(k)$  constant for  $50 < k < 150$  and 0 otherwise, (b)  $P(k) \propto k^{-4}$  for  $67 < k < 1000$  and 0 otherwise, and (c)  $P(k) \propto k^{-3}$  for  $53 < k < 1000$  and 0 otherwise. For each distribution, we considered the uncorrelated case (orange connected circles) and the degree correlated case (blue connected triangles). The dashed lines indicate the value  $\beta_3^c$  at which we expect the onset of bistability, obtained from numerical solution of Eqs. (3.10) and (3.15)-(3.16) for the degree correlated and uncorrelated cases, respectively (in Sec. 3.2 we provide analytical expressions for these values). As the degree distribution of the pairwise interaction network  $P(k)$  becomes more heterogeneous from (a) to (c), the value of  $\beta_3$  at which the onset of bistability occurs increases for the uncorrelated case, while it remains almost unchanged for the degree-correlated case. A heuristic interpretation of this phenomenon is the following: in the uncorrelated case, the triangle interactions do not depend on the heterogeneity of the link degree distribution. Therefore, as the link degree distribution  $P(k)$  becomes more heterogeneous, contagion becomes dominated by hubs of the pairwise interaction network, a mechanism which does not result in bistability. Therefore, bistability is suppressed in the uncorrelated case. On the other hand, for the degree correlated case, both triangle and link contagion mechanisms increase their effectiveness in tandem as the heterogeneity of the link degree distribution is increased. It is important to note that the increase in  $\beta_3^c$  with heterogeneity, which is shown here in absolute terms, still occurs if one considers it relative to the value of  $\beta_2^c$  (i.e.,  $\beta_3^c/\beta_2^c$  also increases with heterogeneity), as we will show later.

Another interesting aspect seen in Figure 3.4 is that the transition to bistable behavior seems sharper in the uncorrelated case for the more heterogeneous networks. As we will see in Sec. 3.2, the nature of the bifurcation is indeed different for the uncorrelated case and heterogeneous networks.

Finally, we have to point out that the numerical calculation of the bistability index from

numerical simulations can be challenging. When the unstable solution is small, finite size effects can cause transitions to the nonzero stable solution from the stable zero solution, making the numerical determination of the stable fixed points difficult and the bistability index plots noisy. Nevertheless, the mean field theory predicts well the onset of bistability.

### 3.1.2 Hyperedges of sizes 2 and 3 with individual contagion

Now we consider the case of individual contagion, in which an  $m$ -hyperedge infects a susceptible node with rate  $\beta_m$  when *at least* one member of the hyperedge is infected. For simplicity, we will still consider only links and triangles ( $M = 3$ ) with infection rates of  $\beta_2$  and  $\beta_3$  respectively.

The analogue to Eq. (3.6) for the individual contagion case is

$$\begin{aligned} \frac{dx_k}{dt} = & -\gamma x_k + (1 - x_k)\beta_2 \sum_{k_1} N(k_1) f_2(k, k_1) x_{k_1} + \\ & (1 - x_k) \frac{\beta_3}{2} \sum_{k_1, k_2} N(k_1) N(k_2) f_3(k, k_1, k_2) [1 - (1 - x_{k_1})(1 - x_{k_2})]. \end{aligned} \quad (3.19)$$

For the correlated case,  $f_3(k, k_1, k_2) = 2kk_1k_2/(N\langle k \rangle)^2$ , this can be rewritten as

$$\frac{dx_k}{dt} = -\gamma x_k + (\beta_2 + 2\beta_3)(1 - x_k)kV - \beta_3(1 - x_k)kV^2, \quad (3.20)$$

with fixed point

$$x_k = \frac{(\beta_2 + 2\beta_3)kV - \beta_3kV^2}{\gamma + (\beta_2 + 2\beta_3)kV - \beta_3kV^2}. \quad (3.21)$$

Inserting this into Eq. (3.7) like before, we obtain

$$V = \frac{1}{N\langle k \rangle} \sum_k \frac{kN(k)[(\beta_2 + 2\beta_3)kV - \beta_3kV^2]}{\gamma + (\beta_2 + 2\beta_3)kV - \beta_3kV^2}. \quad (3.22)$$

Linearizing about the  $V = 0$  equilibrium, we find that the epidemic threshold is given by the condition

$$\beta_2 + 2\beta_3 = \gamma \frac{\langle k \rangle}{\langle k^2 \rangle}, \quad (3.23)$$

which defines a linear relationship between  $\beta_2$  and  $\beta_3$  for fixed  $\gamma$ , in contrast to the collective contagion mechanism which does not alter the epidemic threshold  $\beta_2^c = \gamma\langle k \rangle / \langle k^2 \rangle$ . This relationship

can be understood heuristically by noting that, close to the  $V = 0$  solution, the probability that two nodes in a hyperedge are simultaneously infected can be neglected. Under that assumption, infection of a susceptible node by a triangle when at least one other node is infected is equivalent to independent infection by either of the two other nodes in the triangle with rate  $\beta_3$ . Since in the correlated case a node belongs, on average, to the same number of links and triangles, the individual contagion model reduces to the traditional SIS model with contagion rate  $\beta_2^{\text{eff}} = \beta_2 + 2\beta_3$  in the linear regime (we emphasize, however, that the nonlinear behavior can be different).

In Fig. 3.5 we plot the  $(\beta_2, \beta_3)$  phase space for this scenario, with light pink indicating one solution ( $V = 0$ ) to Eq. (3.22) and orange indicating two solutions, the unstable  $V = 0$  solution and a stable  $V > 0$  solution.

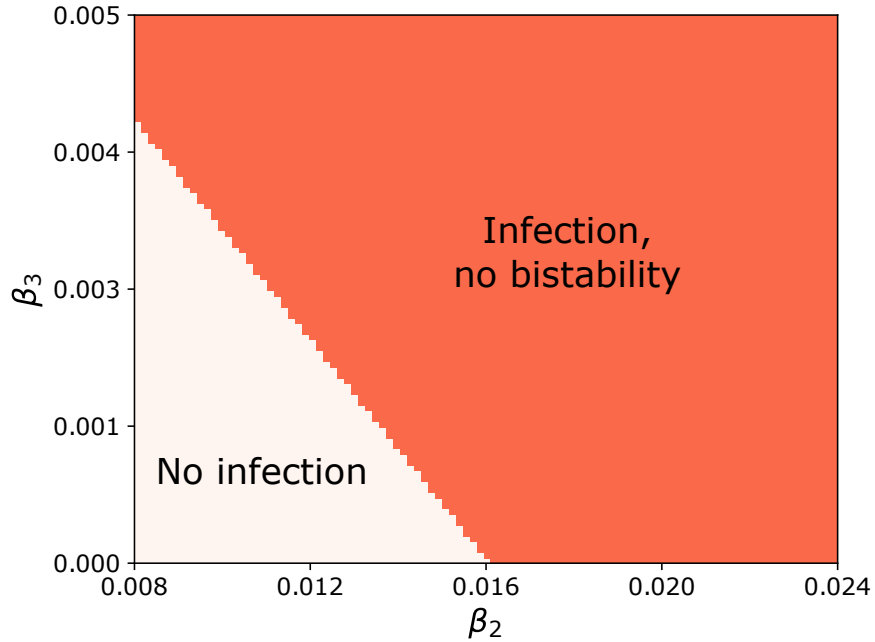


Figure 3.5: Phase diagram for the degree-correlated, individual contagion model with parameters  $\gamma = 2$  and  $P(k) \propto k^{-4}$  when  $67 < k < 1000$  and 0 otherwise.

Considering the uncorrelated case where  $f(k, k_1, k_2) = 2\langle k \rangle / N^2$ , and expressing Eq. (3.19) in



terms of  $U$  and  $V$ , we obtain

$$\begin{aligned} \frac{dx_k}{dt} = & -\gamma x_k + \beta_2(1 - x_k)kV \\ & + 2\beta_3(1 - x_k)\langle k \rangle U - \beta_3(1 - x_k)\langle k \rangle U^2, \end{aligned} \quad (3.24)$$

with equilibrium

$$x_k = \frac{\beta_2 k V + 2\beta_3 \langle k \rangle U - \beta_3 \langle k \rangle U^2}{\gamma + \beta_2 k V + 2\beta_3 \langle k \rangle U - \beta_3 \langle k \rangle U^2}, \quad (3.25)$$

which has different first-order behavior than the degree correlated case. Inserting this expression into Eqs. (3.7) and (3.12), we obtain

$$U = \frac{1}{N} \sum_k \frac{N(k)(\beta_2 k V + 2\beta_3 \langle k \rangle U - \beta_3 \langle k \rangle U^2)}{\gamma + \beta_2 k V + 2\beta_3 \langle k \rangle U - \beta_3 \langle k \rangle U^2}, \quad (3.26)$$

$$V = \frac{1}{N \langle k \rangle} \sum_k \frac{k N(k)(\beta_2 k V + 2\beta_3 \langle k \rangle U - \beta_3 \langle k \rangle U^2)}{\gamma + \beta_2 k V + 2\beta_3 \langle k \rangle U - \beta_3 \langle k \rangle U^2}. \quad (3.27)$$

Linearizing, we obtain the system

$$\delta U = \frac{\langle k \rangle \beta_2}{\gamma} \delta V + \frac{2\langle k \rangle \beta_3}{\gamma} \delta U, \quad (3.28)$$

$$\delta V = \frac{\langle k^2 \rangle \beta_2}{\langle k \rangle \gamma} \delta V + \frac{2\langle k \rangle \beta_3}{\gamma} \delta U. \quad (3.29)$$

Solving this system and canceling the zero solution, we find that the epidemic threshold is defined by a non-linear relationship between the three epidemic parameters

$$\beta_2 = \frac{\langle k \rangle \gamma^2 - 2\langle k \rangle^2 \gamma \beta_3}{\langle k^2 \rangle \gamma - 2(\langle k^2 \rangle - \langle k \rangle^2) \langle k \rangle \beta_3} \quad (3.30)$$

This relationship implies that there is a singularity when  $\beta_3 = \beta_3^* = \gamma \langle k^2 \rangle / [2(\langle k^2 \rangle - \langle k \rangle^2) \langle k \rangle]$ . However, one can check that  $\beta_2$  is negative at  $\beta_3 = \beta_3^*$ , and therefore the singularity is not physically relevant. Note that when  $\langle k^2 \rangle = \langle k \rangle^2$  in the case of a  $k$ -regular network, the threshold reduces to that of the degree-correlated case.

### 3.1.3 Higher-order healing: hipster effect

Here we consider the effect of higher-order healing for both collective and individual contagion. By *higher-order healing* we refer to a situation where infected nodes that belong to a hyperedge of

size  $m > 2$  with other infected nodes heal at rate  $\beta_m$ . This can be thought of as a “hipster effect” where if an idea or trend is popular in groups, then this makes an individual *less* likely to adopt the trend, but the individual can be convinced to adopt the trend by their pairwise connections [48]. For both the collective and individual contagion cases, we comment on the existence of bistability based on numerical phase plots.

When the contagion is collective, the model including higher-order healing can be written as Eq. (3.6) with the sign of the third term changed, and because the triangle healing mechanism is solely higher-order, there is no effect on the epidemic threshold which is obtained by the linearization of the 0 solution. However, we find that explosive transitions do not occur for  $\beta_2, \beta_3 \geq 0$ .

Likewise, for the individual contagion model, higher-order healing can be written as Eq. (3.19) with the third term negative. In this case, the epidemic threshold for both the degree-correlated and uncorrelated case can be obtained by substituting  $-\beta_3$  for  $\beta_3$  in Eq. (3.23) and Eq. (3.30) respectively. Higher-order healing in individual contagion enables explosive transitions to occur for ranges of  $\beta_2, \beta_3 \geq 0$ , as can be seen in Fig. 3.6 which shows the phase space  $(\beta_2, \beta_3)$  for the degree-correlated case. As one might expect, for large enough higher-order healing  $\beta_3$  there is no infection, but there is a narrow band of bistable behavior separating the regions of no infection and monostable infection.

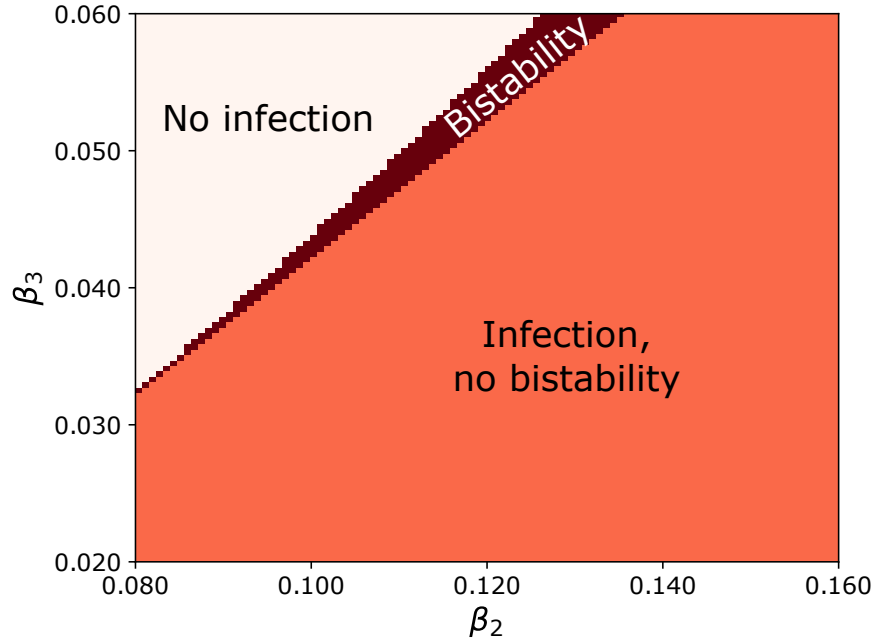


Figure 3.6: Phase diagram for the degree-correlated, higher-order healing with individual contagion with parameters  $\gamma = 2$  and  $P(k) \propto k^{-4}$  when  $67 < k < 1000$  and 0 otherwise.

### 3.1.4 Unfortunate series of events

So far we have considered hypergraphs with hyperedges of sizes 2 and 3 only. We now briefly discuss contagion in networks with hyperedges of all sizes, i.e.,  $M = N$ . In the context of epidemic spreading, hyperedges could be interpreted as participation in social events such as parties, conferences, concerts, and sports events. For simplicity, we will focus on a hypergraph with degree-correlated hyperedges where

$$f_m(\mathbf{k}, \mathbf{k}_1, \dots, \mathbf{k}_{m-1}) = \frac{(m-1)! k k_1 k_2 \dots k_{m-1} \langle k^{(m)} \rangle}{(N \langle k \rangle)^{m-1} \langle k \rangle}, \quad (3.31)$$

such that the average number of hyperedges of size  $m$  a node belongs to is  $\langle k^{(m)} \rangle$ . In this case, by repeating the calculations of Sec. 3.1.2, the fraction of infected nodes of degree  $k$  evolves in terms of the fraction of infected edges  $V$  (3.7) as

$$\frac{dx_k}{dt} = -\gamma x_k + k(1-x_k) \sum_{m=2}^M \frac{\beta_m \langle k^{(m)} \rangle}{\langle k \rangle} [1 - (1-V)^{m-1}], \quad (3.32)$$

for individual contagion and

$$\frac{dx_k}{dt} = -\gamma x_k + k(1 - x_k) \sum_{m=2}^M \frac{\beta_m \langle k^{(m)} \rangle}{\langle k \rangle} V^{m-1}, \quad (3.33)$$

for collective contagion. In the case of collective contagion, larger hyperedges can cause the emergence of new stable fixed points which can lead to richer consensus dynamics [11]. We focus, however, on the case of individual contagion. Linearizing, we find that the solution  $x_k = 0$  becomes unstable when

$$\sum_{m=2}^M \frac{(m-1)\beta_m \langle k^{(m)} \rangle}{\langle k \rangle} > \frac{\gamma \langle k \rangle}{\langle k^2 \rangle}. \quad (3.34)$$

If the sum yields a value larger than  $\gamma \langle k \rangle / \langle k^2 \rangle$  propagating social contagion will result. Social event restrictions implemented as a truncation of the series by prohibiting events larger than a certain size, or practices that reduce contagion in social events and reduce  $\beta_m$  (such as enforcing physical separation) can reduce the value of the sum so that contagion does not propagate [49, 50].

### 3.2 The effect of degree distribution on $\beta_3^c$

In Section 3.1 we expressed the epidemic threshold  $\beta_2^c$  in terms of moments of the degree distribution of the underlying network structure. Similarly, we would like to express the critical value of  $\beta_3$  at which the explosive transitions appear,  $\beta_3^c$ , as a function of hypergraph structure. Explosive transitions and bistability occur when there are two stable steady-state solutions to Eqs. (3.6). For the degree-correlated and uncorrelated cases, this occurs when there are two non-zero solutions to Eq. (3.10) and the coupled system of Eqs. (3.15)-(3.16) respectively. We can compute the critical value of  $\beta_3$  by finding the numerical solution of these mean field equations and determining the value of  $\beta_3^c$  at which bistability appears. This method is much more efficient than using stochastic microscopic simulations of the contagion model to infer the onset of explosive transitions and to map the phase space. Fig. 3.7 shows the predicted value of  $\beta_3^c$  normalized by  $\beta_2^c$  for the correlated (a) and uncorrelated (b) cases as a function of the power-law exponent  $r$  and the maximum degree  $k_{max}$ , where Eqs. (3.10) and (3.15)-(3.16) were solved using  $P(k) \propto k^{-r}$  if  $50 \leq k \leq k_{max}$  and  $P(k) = 0$  otherwise. Larger values of  $r$  and  $k_{max}$  correspond to larger

heterogeneity of the degree distribution. We note that for the most homogeneous network – the  $k$ -regular network –  $\beta_3^c/\beta_2^c$  is 1, and we see in Figs. 3.7(a) and (b) that  $\beta_3^c$  increases relative to  $\beta_2^c$  as  $r$  or  $k_{max}$  increase, except for small values of  $r$  and large values of  $k_{max}$  in the degree-correlated case. Thus, heterogeneity in the degree distribution of the pairwise interaction network appears to suppress explosive transitions. However, this effect is much more pronounced for the uncorrelated case (b) than for the degree-correlated case (a), as we discussed previously. In Appendix A.2.1, we describe in more detail the algorithm employed to find  $\beta_3^c$  from the mean field equations.

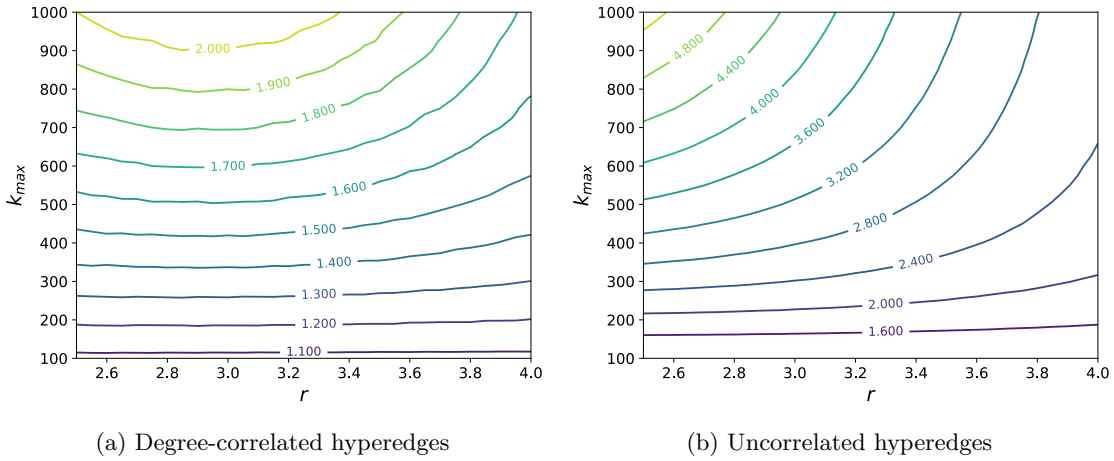


Figure 3.7:  $\beta_3^c/\beta_2^c$  as a function of power-law distribution parameters for the degree-correlated case (a) and the uncorrelated case (b).  $\beta_3^c$  was calculated numerically from the mean field equations (see Appendix A.2.1), and  $\beta_2^c = \gamma\langle k \rangle / \langle k^2 \rangle$ . The parameters are  $P(k) \propto k^{-r}$  if  $50 \leq k \leq k_{max}$  and  $P(k) = 0$  otherwise and  $\gamma = 2$ .

Although this method works well in predicting the value of  $\beta_3^c$ , it does not provide a direct relationship between the network structure and the onset of explosive transitions and is more computationally expensive than an analytical expression. For this reason, we present closed form approximations to  $\beta_3^c$  and describe the parameter regimes over which they are accurate. Starting with the degree-correlated case and canceling the zero solution of Eq. (3.10), we find conditions under which there are at least two solutions to

$$h(V, \beta_2) = \frac{1}{N\langle k \rangle} \sum_k \frac{kN(k)(\beta_2 k + \beta_3 k V)}{\gamma + \beta_2 k V + \beta_3 k V^2} - 1 = 0. \quad (3.35)$$

First, note that  $h(0, \beta_2) = \beta_2/\beta_2^c - 1$  and that  $h(1, \beta_2) < 0$ . Therefore, if  $\frac{\partial h}{\partial V}(0, \beta_2^c) > 0$ , then by continuity, there will be at least two solutions for  $\beta_2$  less than, but sufficiently close to,  $\beta_2^c$ . This condition gives

$$\frac{\beta_3^c}{\gamma} = \frac{\langle k^3 \rangle \langle k \rangle^2}{\langle k^2 \rangle^3}, \quad (3.36)$$

which works well in predicting the onset of bistability for the degree-correlated case. The relative error with respect to the value obtained from directly solving Eq. (3.10) for all distributions tested is less than 2% (not shown).

The analysis for the degree-correlated case was based on the behavior of  $h(V, \beta_2)$  near  $V = 0$ . For the uncorrelated case, however, we find that a saddle-node bifurcation can occur at positive values of  $V$ , and it is necessary to expand Eqs. (3.15)-(3.16) to higher order.

Expanding Eqs. (3.15)-(3.16) to second order, setting  $\beta_2 = \beta_2^c = \gamma \langle k \rangle / \langle k^2 \rangle$ , and subtracting the two equations yields

$$U = \frac{\langle k \rangle^2}{\langle k^2 \rangle} V + \left( \frac{\langle k \rangle \langle k^3 \rangle}{\langle k^2 \rangle^2} - \frac{\langle k \rangle^2}{\langle k^2 \rangle} \right) V^2, \quad (3.37)$$

which, when evaluated in

$$h(V, \beta_2) = \frac{1}{N \langle k \rangle} \sum_k \frac{k N(k) (\beta_2 k V + \beta_3 \langle k \rangle U^2)}{\gamma + \beta_2 k V + \beta_3 \langle k \rangle U^2} - V = 0 \quad (3.38)$$

and expanded to fourth order, again setting  $\beta_2 = \beta_2^c$ , yields

$$h(V, \beta_2^c) = (a_0 + a_1 V + a_2 V^2) V^2, \quad (3.39)$$

where

$$a_0 = -\frac{\langle k \rangle \langle k^3 \rangle}{\langle k^2 \rangle^2} + \frac{\langle k \rangle^5 \beta_3}{\langle k^2 \rangle \gamma}, \quad (3.40)$$

$$a_1 = \frac{\langle k \rangle^2 \langle k^4 \rangle}{\langle k^2 \rangle^3} - 4 \frac{\langle k \rangle^5 \beta_3}{\langle k^2 \rangle^2 \gamma} + 2 \frac{\langle k \rangle^4 \langle k^3 \rangle \beta_3}{\langle k^2 \rangle^3 \gamma}, \quad (3.41)$$

$$a_2 = -\frac{\langle k \rangle^3 \langle k^5 \rangle}{\langle k^2 \rangle^4} + 5 \frac{\langle k \rangle^5 \beta_3}{\langle k^2 \rangle^2 \gamma} + 3 \frac{\langle k \rangle^6 \langle k^3 \rangle \beta_3}{\langle k^2 \rangle^4 \gamma} - 6 \frac{\langle k \rangle^4 \langle k^3 \rangle \beta_3}{\langle k^2 \rangle^3 \gamma} + \frac{\langle k \rangle^3 \langle k^3 \rangle^2 \beta_3}{\langle k^2 \rangle^4 \gamma} - \frac{\langle k \rangle^{10} \beta_3^2}{\langle k^2 \rangle^4 \gamma^2}. \quad (3.42)$$

For continuous transitions to epidemics, there is only one equilibrium for  $V$  at  $\beta_2 = \beta_2^c$ , namely  $V = 0$ . The onset of bistability occurs when a second solution appears, which corresponds

to the first appearance of a root of (3.39) in the interval  $(0, 1)$ . Such a root can appear at  $V = 0$  in a transcritical bifurcation, or at  $V > 0$  as a pair of roots in a saddle-node bifurcation. A pair of roots appears when the discriminant of the quadratic equation  $a_0 + a_1V + a_2V^2 = 0$  is zero. However, this bifurcation is physically meaningless if it occurs for values of  $V$  outside the interval  $[0, 1]$ . Therefore, we impose the constraint that the value of  $\beta_3$  found by solving  $a_1^2 - 4a_0a_2 = 0$  must satisfy the inequality  $0 \leq -a_1/2a_2 \leq 1$ . In addition, we note that because of continuity, the sign of the  $a_2$  term must be negative, because otherwise  $\frac{\partial h}{\partial V}(0, \beta_2^c) > 0$  and the bifurcation has already occurred. The transcritical bifurcation occurs when a root crosses from a negative value to a positive value, which occurs when one root of  $a_0 + a_1V + a_3V^2 = 0$  is  $V = 0$ , implying that  $a_0 = 0$  and  $\beta_3^c = \gamma \langle k^3 \rangle / \langle k \rangle^4$ . Using these conditions, we can construct a piecewise definition of  $\beta_3^c$

$$\beta_3^c = \begin{cases} \text{Solve}(a_1^2 - 4a_0a_2 = 0), & a_2 < 0, 0 \leq -\frac{a_1}{2a_2} \leq 1, \\ \frac{\langle k^3 \rangle}{\langle k \rangle^4} \gamma, & \text{else.} \end{cases} \quad (3.43)$$

The relative error in the value of  $\beta_3^c/\beta_2^c$  obtained from Eq. (3.43) compared with the numerically obtained value shown in Fig. 3.7(b) is shown in Fig. 3.8.

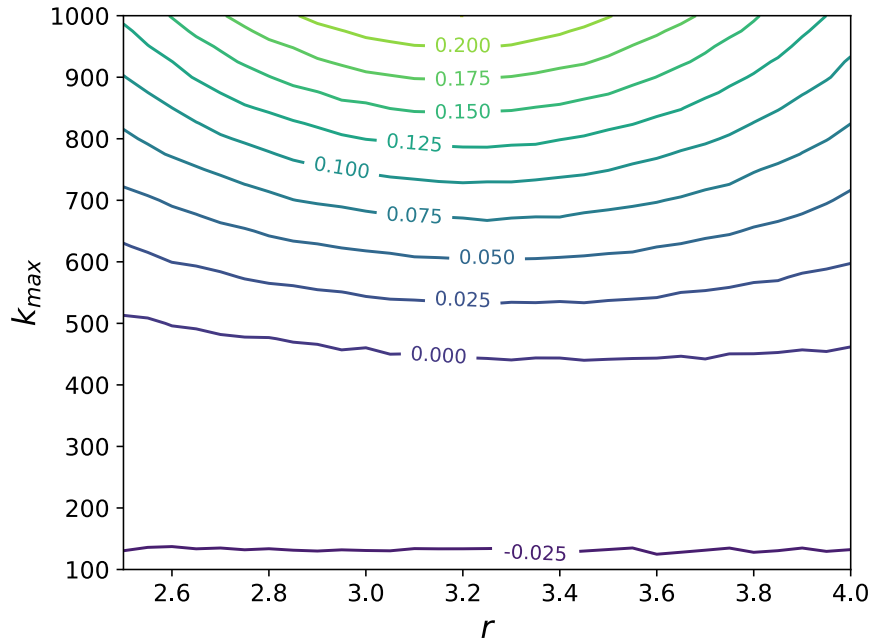


Figure 3.8: Relative error in the value of  $\beta_3^c/\beta_2^c$  obtained from Eq. (3.43) compared with the numerically obtained value shown in Fig. 3.7(b).

In principle, one can expand to higher order to gain accuracy for the most heterogeneous of distributions. However, there is limited utility in increasing the order of the expansion further, because the resulting conditions become extremely complicated.

### 3.3 Discussion

In this Chapter we studied the SIS model of social contagion on hypergraphs with heterogeneous structure. The mean field description in Eq. (3.1) allowed us to explore the effects of hyperedge organization on the epidemic onset and the onset of bistability and explosive transitions. One of our main findings is that with increasing heterogeneity of the pairwise network degree distribution, the onset of explosive transitions is postponed when the pairwise and higher order interactions have independent structure. More generally, when considering a hypergraph contagion model, the group infection and pairwise infection are competing mechanisms by which contagion spreads. Factors that promote contagion via pairwise infection, such as a heterogeneous



degree distribution of the pairwise contact network, suppress discontinuous transitions. Conversely, heterogeneity in the degree distribution of hyperedges of higher order promotes such transitions.

We considered two ways in which the structure of hyperedges of different sizes could be organized: the uncorrelated case, in which they are independent, and the correlated case, in which hyperedges of different sizes connect preferentially to the same nodes. While the organization of hyperedges in real world networks is surely much more complicated, these cases can be considered as null models against which the structure of real-world hypergraphs can be compared.

We studied various forms of higher-order contagion and healing: (i) collective contagion, in which all other members of the hyperedge need to be infected for contagion to occur, (ii) individual contagion, in which at least one member of the hyperdegree needs to be infected, and (iii) higher-order healing, in which pairwise interactions are infectious while higher-order interactions heal. Other forms of higher-order contagion could in principle be studied with the same methodology, but we leave these studies for future research.

Now we mention some of the limitations of our study. First, since we focused on the simplest contagion model, an important question left for future research is whether our results remain valid for more realistic epidemiological models (e.g. such as those used to model COVID-19 [51, 27]). Our model also does not apply to non-Markovian contagion dynamics, which are important when modeling real-world epidemics. From a technical standpoint, another limitation is that we used a mean-field description of the dynamics, and it is known that such a description is not quantitatively accurate for moderate values of the infected population value [52, 44]. Since we were mainly interested in the behavior close to the onset of epidemics, the mean-field approximation was enough for our purposes. However, more precise descriptions could be obtained as in Refs. [44, 53]. Another important limitation of our hypergraph model is that we assume that the probability that two nodes belong to the same hyperedge is a function of their hyperdegrees. While this assumption can be relaxed by considering additional nodal variables, it is possible that such a model might be inadequate to describe some real-world networks. Finally we note that our model relies on knowledge of the functions  $f_m$ , which encode the organization of hyperedges across different

hyperedge sizes. These functions have not yet been estimated from real-world networks, but as progress is made towards understanding the organization of higher-order interactions [54], the determination of these functions could be a natural next step.

While in this Chapter we applied our hyperdegree based mean-field equation to the SIS epidemic model, the same formalism could be applied to other dynamical processes on hypergraphs, such as synchronization, opinion formation, and other types of epidemic models. We believe that this methodology will be useful to study the effect of heterogeneity on these hypergraph dynamical processes.

### **Data Availability**

All code used in this study can be found at <https://github.com/nwlandry/SimplexSIS> [55].

## Chapter 4

### Hypergraph assortativity: a dynamical systems perspective

Complex social systems often exhibit assortative mixing [56, 57], where individuals with similar characteristics connect with each other more often than it would be expected if they were connected at random. Assortativity has been extensively studied in network science [58] and found to have significant effects on synchronization [59], epidemic dynamics [30, 60], stability [61], evolutionary game dynamics [62], and general diffusion processes [63].

The pairwise notion of assortativity has been extended to hypergraphs for categorical node labels [64, 65, 66] and continuous attributes [3]. Assortativity on hypergraphs can provide different insights on the structure of the interactions than assortativity on the pairwise network projection [3] and, as we will show, affect the outcome of hypergraph dynamical processes.

A fundamental problem when studying dynamics on networks is to determine how structural characteristics of the network affect the dynamical behavior. Many dynamical properties such as the onset of epidemic spreading [20], synchronization [67], and percolation [68] are determined by the largest eigenvalue of the network's adjacency matrix (or, in some cases, of the non-backtracking matrix [69]). In turn, this eigenvalue is affected by the network's degree distribution and assortative mixing properties [70] as well as other structural characteristics. In this Chapter we show how the expansion eigenvalue, a suitably generalized eigenvalue for hypergraphs, is similarly modified by the assortative properties of the hypergraph. This eigenvalue has been shown to determine the extinction threshold for the SIS model on hypergraphs [71], and we believe it will also prove useful in relating hypergraph assortative mixing patterns to other dynamical processes.

Our approach is as follows: first, we define and motivate the importance of the expansion eigenvalue on dynamical processes; second, we derive a mean-field approximation of this eigenvalue for hypergraphs without assortativity; third, we present a generative model for assortative hypergraphs; fourth, we employ a perturbation approach to derive the effect of degree-degree mixing on the eigenvalue and define the dynamical assortativity; and lastly, we show how our results can be used to modify hypergraph dynamics through preferential rewiring of hyperedges.

## 4.1 Motivation

We define the expansion eigenvalue and discuss its relevance to dynamical processes on hypergraphs. For a weighted hypergraph, the expansion eigenvalue  $\lambda$  and associated eigenvector  $\mathbf{u}$  are defined by the eigenvalue equation

$$\lambda u_i = \sum_{e=\{i,i_1,\dots,i_{m-1}\}\in E} \beta_e (u_{i_1} + \dots + u_{i_{m-1}}), \quad (4.1)$$

where  $\lambda$  and  $\mathbf{u}$  are the Perron-Frobenius eigenvalue and eigenvector of the nonnegative matrix associated to linear equation (4.1).

Here we present some applications of the expansion eigenvalue. First, just like the Perron-Frobenius eigenvector of a network adjacency matrix represents eigenvector centrality [72], in the unweighted case (i.e.,  $\beta_e = 1$  for every hyperedge  $e$ ), the eigenvector  $\mathbf{u}$  corresponds to the Clique motif Eigenvector Centrality, a generalization of eigenvector centrality for hypergraphs [73]. Second, just as the largest eigenvalue of a network's adjacency matrix is determinant for network dynamics, the expansion eigenvalue plays an important role in dynamical processes on hypergraphs. For example, consider an SIS process on a hypergraph, where a healthy node can get infected via a hyperedge  $e$  to which it belongs at rate  $\beta_e$  if at least one other node in  $e$  is infected (the case referred to as individual contagion in Ref. [12]) and heals spontaneously at rate  $\gamma$ . As discussed in Ref. [71] in Theorem 9.1, the extinction threshold for the exact stochastic process can be bounded above by that for the mean-field dynamics. The mean-field equation for  $x_i$ , the probability that

node  $i$  is infected, is given by

$$\begin{aligned} \frac{dx_i}{dt} &= -\gamma x_i + (1 - x_i) \sum_{e=\{i,i_1,\dots,i_{m-1}\} \in E} \beta_e \\ &\times [1 - (1 - x_{i_1}) \dots (1 - x_{i_{m-1}})]. \end{aligned} \quad (4.2)$$

By inspection,  $x_i = 0$  for all  $i$  is always a fixed point of this equation. We write an ODE for linear perturbations around this equilibrium to derive conditions for the system's stability. To first order, the equation for the perturbations,  $\delta x_i$ , is

$$\frac{d(\delta x_i)}{dt} = -\gamma(\delta x_i) + \sum_{\{i,i_1,\dots,i_{m-1}\} \in E} \beta_e (\delta x_{i_1} + \dots + \delta x_{i_{m-1}}), \quad (4.3)$$

If we assume  $\delta x_i = u_i e^{rt}$ , then

$$(r + \gamma)u_i = \lambda u_i = \sum_{\{i,i_1,\dots,i_{m-1}\} \in E} \beta_e (u_{i_1} + \dots + u_{i_{m-1}}), \quad (4.4)$$

where  $\lambda$  is the expansion eigenvalue and so,  $r = \lambda - \gamma$ . Therefore, a sufficient condition for epidemic extinction is  $\gamma > \lambda$  [71]. For an  $m$ -uniform hypergraph with  $\beta_e = \beta_m$ , the extinction threshold is  $\beta_m/\gamma < 1/\lambda$ , where  $\lambda$  is the expansion eigenvalue of the unweighted hypergraph.

If we rewrite the last term of Eq. (4.3) as a sum over uniform hypergraphs, then

$$\frac{d(\delta x_i)}{dt} = -\gamma(\delta x_i) + \sum_{m=2}^M (m-1) \left( W^{(m)} \delta \mathbf{x} \right)_i, \quad (4.5)$$

where  $W^{(m)}$  is the weighted version of the clique motif matrix defined in Ref. [73] and  $\delta \mathbf{x} = [\delta x_1, \dots, \delta x_N]$ . We can define  $W = \sum_{m=2}^M (m-1) W^{(m)}$  as a linear operator with eigenvalue  $\lambda$  and as before, the extinction threshold is  $\gamma > \lambda$ . In addition, we can determine the relative importance of a node  $i$  with respect to this contagion model (in terms of its probability of infection at the onset of the epidemic) from the  $i$ th entry of the associated eigenvector.

Finally, the importance of the expansion eigenvalue in spreading processes can be understood from the fact that in the unweighted case the number of nodes reachable via hyperedges from a given starting node in  $\ell$  steps grows asymptotically as  $\lambda^\ell$  as demonstrated in Ref. [73]. We discuss limitations of the expansion eigenvalue in Section 4.4.

## 4.2 Mean-Field Approach

### 4.2.1 Uncorrelated $m$ -uniform case

We start by deriving a mean-field approximation for the expansion eigenvalue  $\lambda$  in the case where nodes are connected with hyperedges completely at random (as in the hypergraph configuration model [33, 32, 12, 3]), which we call the *uncorrelated case*, before considering hypergraphs with degree assortativity. In the uncorrelated case, the function  $f_m$  is given by  $f_m(\mathbf{k}_{i_1}, \dots, \mathbf{k}_{i_m}) = f_m^{(0)}(k_{i_1}^{(m)}, \dots, k_{i_m}^{(m)}) = (m-1)! k_{i_1}^{(m)} \dots k_{i_m}^{(m)} / (N \langle k^{(m)} \rangle)^{m-1}$ . For simplicity, from now on we will consider an unweighted  $m$ -uniform hypergraph, and will denote  $k_i^{(m)}$  by  $k_i$  and refer to it as the degree of node  $i$ . Now we assume that all nodes with the same degree are statistically equivalent and that the eigenvector entry of node  $i$  depends only on its degree, i.e.,  $u_i \rightarrow u_{k_i}$ . In Section 4.4, we discuss the limitations of this approach. Henceforth  $\lambda$  will denote the mean-field approximation to the expansion eigenvalue for convenience unless explicitly stated otherwise. Defining  $N(k)$  to be the number of nodes with degree  $k$  such that  $P(k) = N(k)/N$  is the degree distribution, the equation defining the expansion eigenvalue can be written as

$$\begin{aligned} \lambda u_k &= \frac{1}{(m-1)!} \sum_{k_1, \dots, k_{m-1}} N(k_1) \dots N(k_{m-1}) \\ &\times f_m^{(0)}(k, k_1, \dots, k_{m-1})(u_{k_1} + \dots + u_{k_{m-1}}). \end{aligned} \quad (4.6)$$

By symmetry of the function  $f_m^{(0)}$ , we get

$$\lambda u_k = \left[ (m-1) \sum_{k_1} P(k_1) \frac{k_1 u_{k_1}}{\langle k \rangle} \right] k, \quad (4.7)$$

and multiplying both sides by  $k P(k) / \langle k \rangle$  and summing over  $k$ , we obtain for the uncorrelated case

$$\lambda = (m-1) \frac{\langle k^2 \rangle}{\langle k \rangle}, \quad (4.8)$$

and  $u_k \propto k$  from Eq. (4.7).

### 4.2.2 Derivation of the non-uniform uncorrelated expansion eigenvalue

We now relax the assumption of an  $m$ -uniform hypergraph and consider an uncorrelated hypergraph with hyperedges of sizes  $m = 2, \dots, M$  and hyperedge weights of the form  $\beta_e = \beta_{|e|}$ . The expansion eigenvalue equation can be written as

$$\lambda u_i = \sum_{m=2}^M \beta_m \sum_{\{i, i_1, \dots, i_{m-1}\} \in E} (u_{i_1} + \dots + u_{i_{m-1}}). \quad (4.9)$$

The degree-based mean-field eigenvalue equation, where we assume  $u_i = u_{\mathbf{k}_i}$ , can be written as

$$\begin{aligned} \lambda u_{\mathbf{k}} &= \sum_{m=2}^M \beta_m \frac{1}{(m-1)!} \sum_{\mathbf{k}_1, \dots, \mathbf{k}_{m-1}} N(\mathbf{k}_1) \dots N(\mathbf{k}_{m-1}) \\ &\quad \times f_m(\mathbf{k}, \mathbf{k}_1, \dots, \mathbf{k}_{m-1}) (u_{\mathbf{k}_1} + \dots + u_{\mathbf{k}_{m-1}}). \end{aligned} \quad (4.10)$$

Focusing on the uncorrelated case, we assume that

$$\begin{aligned} f_m(\mathbf{k}, \mathbf{k}_1, \dots, \mathbf{k}_{m-1}) &= f_m^{(0)}(k^{(m)}, k_1^{(m)}, \dots, k_{m-1}^{(m)}) \\ &= \frac{(m-1)! k^{(m)} k_1^{(m)} \dots k_{m-1}^{(m)}}{(N\langle k^{(m)} \rangle)^{m-1}}, \end{aligned}$$

so

$$\begin{aligned} \lambda u_{\mathbf{k}} &= \sum_{m=2}^M \beta_m \sum_{\mathbf{k}_1, \dots, \mathbf{k}_{m-1}} N(\mathbf{k}_1) \dots N(\mathbf{k}_{m-1}) \\ &\quad \times \frac{k^{(m)} k_1^{(m)} \dots k_{m-1}^{(m)}}{(N\langle k^{(m)} \rangle)^{m-1}} (u_{\mathbf{k}_1} + \dots + u_{\mathbf{k}_{m-1}}), \end{aligned}$$

and from symmetry,

$$\lambda u_{\mathbf{k}} = \sum_{m=2}^M k^{(m)} \beta_m (m-1) \sum_{\mathbf{k}_1} P(\mathbf{k}_1) \frac{k_1^{(m)} u_{\mathbf{k}_1}}{\langle k^{(m)} \rangle}. \quad (4.11)$$

From Eq. (4.11), we can see that  $u_{\mathbf{k}}$  must be a linear combination of  $k^{(m)}$ . We assume an ansatz of the form

$$u_{\mathbf{k}} = \sum_{m=2}^M v_m k^{(m)} = \mathbf{k}^T \mathbf{v} \quad (4.12)$$

where  $\mathbf{v} = (v_2, \dots, v_M)$  is an unknown vector of nonnegative weights. Renaming the summation indices and evaluating this ansatz in the eigenvalue equation,

$$\begin{aligned} \lambda \sum_{j=2}^M v_j k^{(j)} &= \sum_{i=2}^M k^{(i)} \beta_i (i-1) \\ &\times \sum_{\mathbf{k}_1} P(\mathbf{k}_1) \frac{k_1^{(i)} \sum_{j=2}^M v_j k_1^{(j)}}{\langle k^{(i)} \rangle}. \end{aligned}$$

Changing the order of summation,

$$\begin{aligned} \lambda \mathbf{k}^T \mathbf{v} &= \sum_{i=2}^M \sum_{j=2}^M k^{(i)} \frac{\beta_i (i-1)}{\langle k^{(i)} \rangle} v_j \sum_{\mathbf{k}_1} P(\mathbf{k}_1) k_1^{(i)} k_1^{(j)}, \\ &= \sum_{i=2}^M \sum_{j=2}^M k^{(i)} \frac{\beta_i (i-1) \langle k^{(i)} k^{(j)} \rangle}{\langle k^{(i)} \rangle} v_j, \\ &= \mathbf{k}^T K \mathbf{v}. \end{aligned}$$

We call  $K$  the degree-size correlation matrix, with entries  $K_{ij} = \beta_i (i-1) \langle k^{(i)} k^{(j)} \rangle / \langle k^{(i)} \rangle$  which we call the inter-size correlations. In Ref. [74], the authors derive a similar matrix for higher-order percolation processes. Generically (when  $\mathbf{k}$  is not orthogonal to the range of  $K - \lambda I$ ), this equation has a solution if and only if  $\lambda$  and  $\mathbf{v}$  solve the eigenvalue equation  $\lambda \mathbf{v} = K \mathbf{v}$ . Notice that in the  $m$ -uniform case, we recover the expression we previously derived. Consider the network formed by specifying hyperedge sizes ( $m = 2, \dots, M$ ) to be the nodes and constructing a link between two sizes  $m_1$  and  $m_2$  if at least one node in the original hypergraph is a member of a hyperedge of size  $m_1$  and a hyperedge of size  $m_2$ .  $K$  is irreducible if and only if this network is strongly connected. If this is the case, by the Perron-Frobenius theorem the eigenvalue with largest magnitude is positive and has a corresponding positive eigenvector, and they correspond, respectively, to  $\lambda$  and  $\mathbf{v}$ .



### 4.2.3 Perturbation approach for the correlated case

In contrast with the uncorrelated case, we now assume that nodes are connected with an arbitrary function  $f_m$  determining the connection probability. We define

$$f_m(k_1, \dots, k_m) = f_m^{(0)}(k_1, \dots, k_m) [1 + \epsilon g_m(k_1, \dots, k_m)], \quad (4.13)$$

where  $\epsilon$  is a parameter which will later assume to be small and  $g_m$  an assortativity function for  $m$ -uniform hypergraphs. The assortativity function  $g_m(k_1, \dots, k_m)$  determines how likely it is that nodes with degrees  $k_1, \dots, k_m$  are joined by a  $m$ -hyperedge; if  $\epsilon g_m > 0$  ( $\epsilon g_m < 0$ ) it is more (less) likely than it would be expected if they were connected at random. In order to preserve the expected degree sequence,  $g_m$  must satisfy  $\sum_{k_1, \dots, k_m} f_m^{(0)}(k_1, \dots, k_m) g_m(k_1, \dots, k_m) = 0$ .

We now assume that the parameter  $\epsilon$  is small and develop perturbative approximations to the eigenvalue  $\lambda$  and its eigenvector  $u_k$ . To first order these approximations are

$$\begin{aligned} \lambda &= \lambda^{(0)} + \epsilon \lambda^{(1)}, \\ u_k &= u_k^{(0)} + \epsilon u_k^{(1)}, \end{aligned} \quad (4.14)$$

where  $\lambda^{(0)} = (m-1)\langle k^2 \rangle / \langle k \rangle$  and  $u_k^{(0)} = \alpha k$ , where  $\alpha$  is an arbitrary constant.

Replacing  $f_m^{(0)}$  on the right-hand side of Eq. (4.6) with the  $f_m$  in Eq. (4.13), using Eq. (4.14), assuming symmetry of  $f_m$ , multiplying by  $kP(k)/(N\langle k \rangle)$ , summing over  $k$ , and canceling the zero-order terms, we obtain to first order (see Appendix B.1 for more detailed calculations)

$$\begin{aligned} \lambda^{(1)} &= (m-1) \frac{\langle k \rangle}{\langle k^2 \rangle} \sum_{k, k_1, \dots, k_{m-1}} N(k) N(k_1) \dots N(k_{m-1}) \\ &\quad \times \frac{k^2 k_1^2 k_2 \dots k_{m-1}}{(N\langle k \rangle)^m} g_m(k, k_1, \dots, k_{m-1}). \end{aligned} \quad (4.15)$$

Removing the reference to  $g_m$  using the relation in Eq. (4.13) we find

$$\epsilon \lambda^{(1)} = (m-1) \frac{\langle k \rangle \langle k k_1 \rangle_E}{\langle k^2 \rangle} - \lambda^{(0)}, \quad (4.16)$$

where  $\langle k k_1 \rangle_E$  is the mean pairwise product of degrees over all possible 2-node combinations in each hyperedge in the hypergraph,  $\langle k k_1 \rangle_E = \sum_{e \in E} \sum_{\{i, j\} \subseteq e} k_i k_j / (|E| \binom{m}{2})$ .

Therefore, the expansion eigenvalue can be written, to first order, as

$$\begin{aligned}\lambda &= \lambda^{(0)} + \epsilon\lambda^{(1)} = (m-1)\frac{\langle k \rangle \langle kk_1 \rangle_E}{\langle k^2 \rangle}, \\ &= \lambda^{(0)}(1 + \rho),\end{aligned}\tag{4.17}$$

where we defined

$$\rho = \frac{\langle k \rangle^2 \langle kk_1 \rangle_E}{\langle k^2 \rangle^2} - 1.\tag{4.18}$$

We refer to  $\rho$  as the dynamical assortativity for its relation to hypergraph dynamics. One can verify that the expected value of  $\rho$  for an uncorrelated hypergraph is 0. Interestingly, to first order the expansion eigenvalue does not depend on the particular assortativity function  $g_m$  used, but only on the average of pairwise products of the degrees belonging to the same hyperedge. A schematic of disassortative ( $\rho < 0$ ) and assortative ( $\rho > 0$ ) hypergraphs is shown in Fig. 4.1.

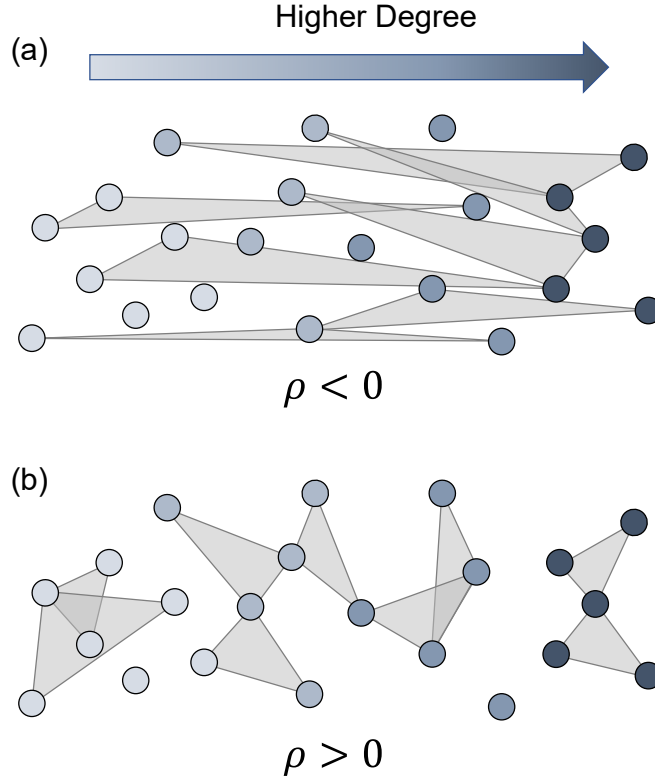


Figure 4.1: An illustration of disassortative and assortative 3-uniform hypergraphs. The color of the nodes indicates their degree, with low-degree nodes on the left and high-degree nodes on the right. For a given degree sequence, the term  $\langle kk_1 \rangle_E$  (the average pairwise product) determines  $\rho$  and on average, (a) hyperedges containing nodes with dissimilar degrees decrease this term leading to disassortative hypergraphs and (b) hyperedges containing nodes of similar degree increase this term leading to assortative hypergraphs.

### 4.3 Numerical Results

#### 4.3.1 Approximating the eigenvalue

We validate our results with numerical simulations on both synthetic and empirical hypergraphs. For both types of data, we modify the dynamical assortativity of the datasets by performing preferential double hyperedge swaps on the hypergraphs.

For each dataset hypergraph  $H$ , we focus on an  $m$ -uniform partition  $H_m$  (i.e., we only consider its hyperedges of size  $m$ ). We set a target dynamical assortativity  $\hat{\rho}$  and swap edges as follows. We choose two hyperedges  $e_1 = \{i_1, i_2, \dots, i_m\}$ ,  $e_2 = \{j_1, j_2, \dots, j_m\}$  and a node from each uniformly at random, say  $i_1$  and  $j_1$ . Then we consider the rewired hypergraph  $H'_m$  obtained by replacing  $e_1$

and  $e_2$  with  $e'_1 = \{j_1, i_2, \dots, i_m\}$  and  $e'_2 = \{i_1, j_2, \dots, j_m\}$  respectively. If the assortativity of  $H'_m$  with this hyperedge swap,  $\rho'$ , reduces the difference between the current assortativity,  $\rho$ , and the desired assortativity,  $\hat{\rho}$ , the swap is accepted and we set  $H_m = H'_m$ . To ensure that the algorithm explores the space of possible hypergraphs, we accept hyperedge swaps which increase the difference between the desired assortativity and the current assortativity with probability  $e^{-[(\hat{\rho}-\rho)^2 - (\hat{\rho}-\rho')^2]/T}$  (we set  $T = 10^{-5}$ ). We terminate the algorithm when  $|\rho - \hat{\rho}|$  is smaller than a prescribed tolerance or when a maximum number of hyperedge swaps have been performed (we used a tolerance of  $10^{-2}$  and  $10^6$  maximum hyperedge swaps).

For the synthetic hypergraph, we constructed a 3-uniform configuration model (CM) hypergraph of size  $N = 10^5$  according to the algorithm described in Ref. [12] with a degree sequence drawn from a truncated power-law distribution,  $P(k) \propto k^{-3}$  on  $[10, 100]$ . We also used the tags-askubuntu (TAU), congress-bills (CB), and Eu-Emails (EE) hypergraph datasets from Refs. [75, 76, 77], filtered to only include hyperedges of size 3. The characteristics of these datasets are described in Table 4.1.

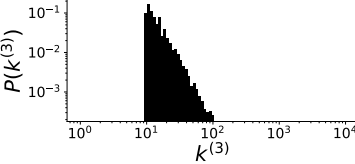
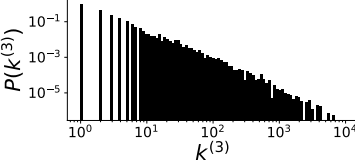
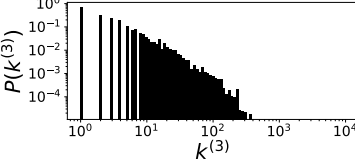
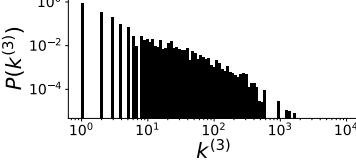
Dataset	$N$	$\langle k \rangle$	$P(k)$
CM	$10^4$	18.2	
TAU	3029	71.2	
CB	1718	20.6	
EE	998	53.7	

Table 4.1: Characteristics of the 3-uniform hypergraph datasets used.

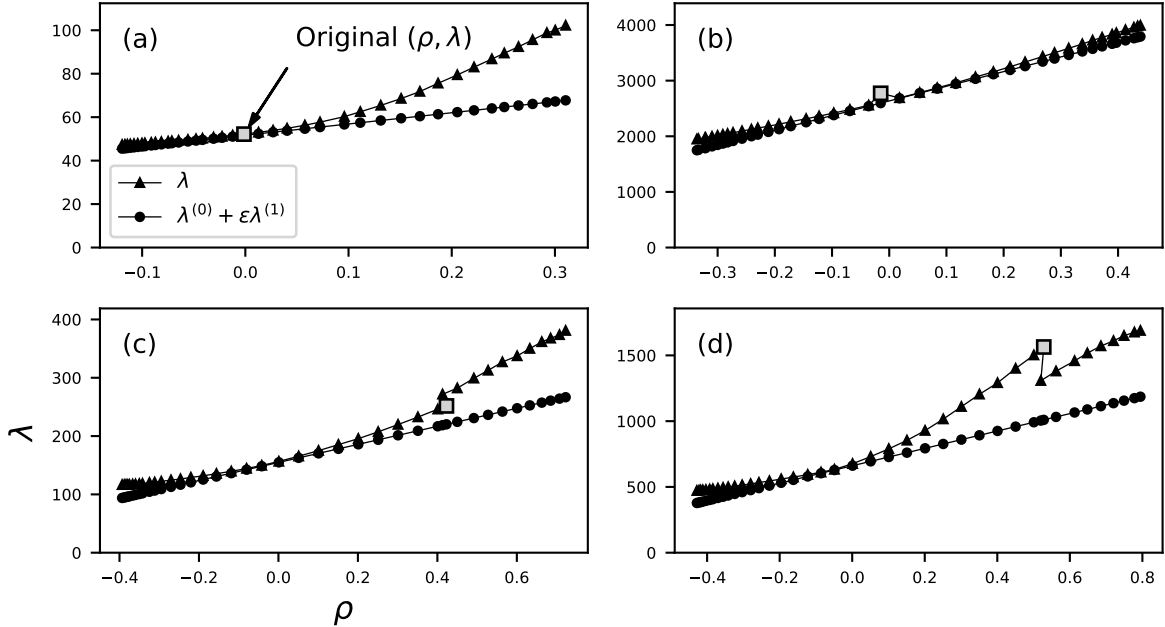


Figure 4.2: A comparison of the actual expansion eigenvalue  $\lambda$  (connected triangles) to the first-order approximation of the eigenvalue  $\lambda^{(0)} + \epsilon\lambda^{(1)}$  (connected circles) for (a) the configuration model, (b) the tags-ask-ubuntu dataset, (c) the congress-bills dataset, and (d) the Eu-E-mails dataset. The square marker denotes the original  $(\rho, \lambda)$  value of the dataset. Details of the characteristics of these datasets can be found in Table 4.1.

In Fig. 4.2, the expansion eigenvalue  $\lambda$  calculated numerically via the power method from Eq. (4.1) (connected triangles) and the first-order approximation  $\lambda^{(0)} + \epsilon\lambda^{(1)}$  (connected circles) are plotted as a function of  $\rho$  for the four datasets mentioned above. For each dataset, the starting point [i.e., the point  $(\rho, \lambda)$  for the original hypergraph] is shown with a square marker. For the synthetic hypergraph (a), as expected, the first order approximation works well for small values of dynamical assortativity. For the TAU dataset (b) the agreement is even better than for the synthetic dataset for larger values of  $\rho$ . Interestingly, for the CB (c) and EE (d) datasets, and to a much lesser extent for the TAU dataset, the value of  $\lambda$  changes sharply when first increasing (CB dataset and EE datasets), or both increasing and decreasing (TAU dataset) the assortativity. We hypothesize that initial hyperedge swaps might be destroying other structure (such as community structure, clustering, or assortative mixing by unaccounted attributes), causing  $\lambda$  to change abruptly as this structure is destroyed, and then to change slowly as the effects of changing the assortativity

dominate. We note that there appear to be limitations to the extent to which  $\rho$  can be modified. This is similar to the limitations to the values of assortativity that networks and hypergraphs can achieve [78, 79, 80, 81].

In all cases, we see that rewiring the hypergraph to increase the average value of  $\langle kk_1 \rangle_E$  (or, equivalently,  $\rho$ ) has a dramatic effect on the expansion eigenvalue. For example, for the EE dataset  $\lambda$  can be reduced threefold by the rewiring process. Thus, hypergraph rewiring might be a useful theoretical tool to control dynamical processes that depend on the expansion eigenvalue.

### 4.3.2 Extinguishing epidemics

Lastly, we show how modifying the dynamical assortativity by rewiring hypergraphs can extinguish an epidemic. As an example, consider a hypergraph SIS contagion spreading amongst groups of size  $m$  at a *fixed* rate  $\beta_m$ . In Ref. [71], the authors derive a sufficient condition for epidemic extinction for such models. For  $m$ -uniform hypergraphs and  $\beta_e = \beta_m$ , the extinction threshold for the individual contagion model is  $\beta_m < \beta_m^c = \gamma/\lambda$ . By decreasing  $\lambda$  through hyperedge swaps and thus increasing  $\beta_m^c$  so that  $\beta_m^c > \beta_m$ , the epidemic can be extinguished. (Note, however, that this is a sufficient condition;  $\beta_m > \beta_m^c$  may not lead to an epidemic.)

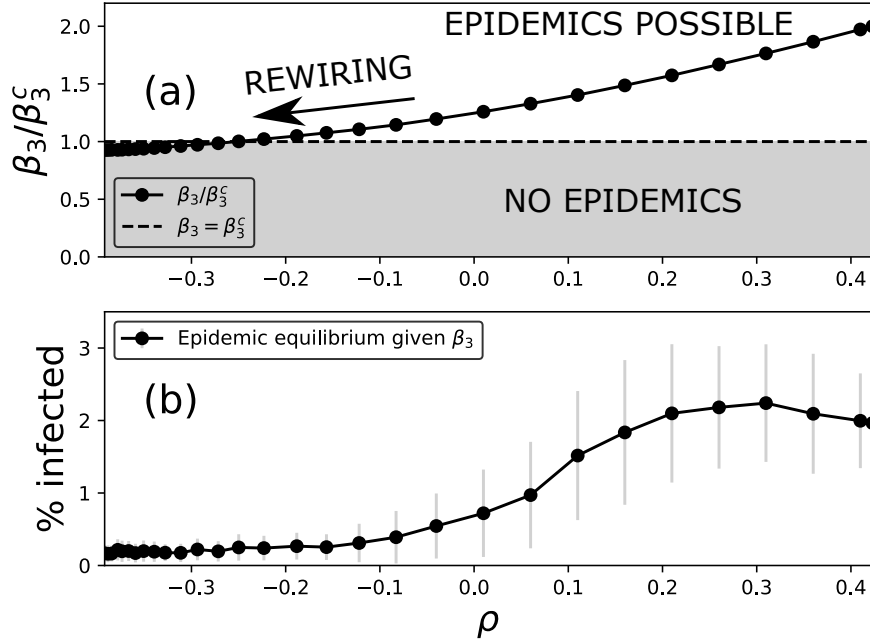


Figure 4.3: (a) The solid line with markers denotes the fixed value of  $\beta_3$  as a fraction of the extinction threshold,  $\beta_3/\beta_3^c$ . The dashed line indicates  $\beta_3/\beta_3^c = 1$ , below which epidemics are not possible. (b) The epidemic equilibrium (percentage of the population infected) for each hypergraph given the fixed value of  $\beta_3$ . The grey bars indicate the standard deviation at each data point.

We present an example based on the CB dataset, and additional cases in Appendix B.2. In this case, we consider  $m = 3$ ,  $\gamma = 1$ , and  $\beta_3 = 7.9 \times 10^{-3}$ . In Fig. 4.3(a), we plot the chosen value of  $\beta_3$  as a fraction of the extinction threshold,  $\beta_3/\beta_3^c$  (solid line with markers), which decreases as  $\beta_3^c$  is increased by hyperedge swaps, and the threshold for extinction (dashed line)  $\beta_3/\beta_3^c = 1$ . Below the dashed line, epidemics are impossible. Above the dashed line, they may be possible. In Fig. 4.3(b), we plot the percentage of the population infected as a function of  $\rho$  (averaged over 100 realizations of the epidemic). For more details about the numerical epidemic simulations see Appendix B.3. For all values of  $\rho$  such that  $\beta_3/\beta_3^c < 1$ , no epidemics occur. For large enough values of  $\rho$ , however, we see that epidemics occur.

We caution, however, that decreasing  $\lambda$  via hyperedge swaps might not necessarily suppress epidemics if  $\beta_3/\beta_3^c$  is not reduced below 1. In principle, epidemics will occur for values of  $\beta_3$  larger than a threshold  $\beta_3^* \geq \beta_3^c$  which depends on the hypergraph structure. If the hyperedge



swaps modify this threshold in such a way that  $\beta_3^c < \beta_3^* < \beta_3$ , when originally  $\beta_3^c < \beta_3 < \beta_3^*$ , epidemics can actually be promoted by the rewiring process (we show examples in Appendix B.2). Therefore, reduction of  $\beta_3/\beta_3^c$  by preferential hyperedge swaps should be attempted only when one can guarantee that  $\beta_3/\beta_3^c$  can be reduced below 1 or when there is already an epidemic.

#### 4.4 Discussion

In this Chapter, we have presented a novel definition of assortativity for hypergraphs, related it to the expansion eigenvalue, and motivated its use in relating assortative structure in hypergraphs to the epidemic behavior. This approach, however, has limitations regarding the application of the expansion eigenvalue to hypergraphs and the calculation of the epidemic threshold.

There are two main limitations of the expansion eigenvalue. The first limitation is that one can think of the matrix associated to the right-hand side of Eq. (4.1) as the weighted adjacency matrix of an effective *pairwise* network, therefore reducing group interactions to multiple pairwise interactions. Such a reduction does not always capture all the complexity of nonlinear dynamical processes [34]. In particular, higher-order dynamical correlations might be missed by this approach. The second (related) limitation is that, since this eigenvalue is, by definition, a quantity related to linear processes, its applicability is restricted in principle only to certain dynamical regimes. However, approaches that reduce a hypergraph to an effective pairwise network have been successful and found application in clustering [82], diffusion and consensus [83], centrality [73], contagion [84], and other areas. In addition, as we showed, the expansion eigenvalue still encapsulates a large amount of information about the hypergraph structure, such as the hyperdegree distribution, correlations between degrees of different order, and assortative mixing. Therefore, the expansion eigenvalue should be considered as a complementary tool to other measures of hypergraph structure.

When deriving approximations to the epidemic threshold for the SIS model in pairwise networks, many approaches may be considered such as using heterogeneous mean-field approaches [30], the largest eigenvalue of the adjacency matrix [20] (the quenched mean-field approach), the largest eigenvalue of the non-backtracking matrix [69], the largest eigenvalue of the branching matrix [69],

message passing approaches [85, 86], and many others. The differences between and advantages of these approaches are discussed at length in Ref. [87]. The same is true for the epidemic threshold in hypergraphs. Although there are more accurate approximations of the epidemic threshold [69, 88], we employ the quenched mean-field approach because of its simple relation to the adjacency tensor of an  $m$ -uniform hypergraph and corresponding explainability. As in the pairwise network case, more sophisticated approaches [44] can yield a better approximation to the epidemic threshold.

Similarly, there are many ways to derive an approximation to the largest eigenvalue for pairwise networks, given an adjacency matrix. The authors in Ref. [89] derive the largest eigenvalue for a Chung-Lu network excluding any correlations. In Ref. [90], the authors extend the approach of Ref. [89] by allowing degree-degree correlations to occur. We use a heterogeneous mean-field approach as in Ref. [89] which requires certain assumptions. First, we enforce that the hypergraph is realizable; that is,  $f_m(k_1^{(m)}, \dots, k_m^{(m)}) \leq 1$  for every combination of  $k_1^{(m)}, \dots, k_m^{(m)}$ , which for the uncorrelated case requires that

$$k_{max}^{(m)} \leq m \sqrt[m]{\left(\sum_{i=1}^N k_i^{(m)}\right)^{m-1}}.$$

For the assortative case, this necessary condition depends on the specific assortativity function used. In addition, the mean-field approximation where we assume that nodes with the same hyperdegree have the same eigenvector entry is valid only when each node has a large number of connections, so that the states of neighbors of nodes with the same hyperdegree are statistically similar.

Despite these limitations, our results provide a way to connect various measures of hypergraph structure with dynamical processes in a systematic way and for a large class of tunable null models. We believe that exploring the role of the expansion eigenvalue in other dynamical processes on hypergraphs will be a fruitful research direction.

## Data Availability

All code used in this study can be found at <https://github.com/nwlandry/hypergraph-assortativity> [91].

## Chapter 5

### Community structure in hypergraphs and the emergence of polarization

Polarization is the extent to which opinions on an issue are opposed [92] and there is evidence of polarization in politics [93], social media [94], and ideology [95]. It is thought that *echo-chambers*, where communities reinforce their own beliefs, may be partially responsible for the polarization seen in politics, media, and society at large [96].

Empirical networks often contain community structure, with dense connections within each community and sparser connections bridging between communities [52]. Community structure in pairwise networks can affect the dynamical behavior in many different ways. Coupled oscillators on networks with community structure can simultaneously exhibit incoherent and synchronized behavior (known as *chimera states*) [97]. In the context of epidemics, community structure has strong effects on disease dynamics [98, 99].

Often, researchers will model polarization explicitly by including mechanisms likely to create polarization, the most common of these being bounded confidence models (BCMs) where the opinions of an agent's neighbors affect the opinion of the agent only if their opinion is close enough. In this model, however, opinion fragmentation or the formation of sects can occur without any community structure whatsoever.

In Ref. [96], the authors present a variant of the Hegselmann-Krause BCM with smooth, non-linear activation functions. The authors run this process on a directed, temporal network specified by the activity model with a power-law degree distribution. The authors also explore this phenomenon by extending a 1-dimensional model to multi-dimensional topic spaces and validating

this model with empirical opinion data [100]. Although this model may produce polarization for populations without community structure, opinions are segregated according to community structure when it does exist.

The Deffuant-Weisbuch BCM model selects an interaction at random for each time step and updates the opinions of the agents participating in that interaction if their opinions are similar enough. The measure of their similarity is called the *confidence bound* and denoted with  $\epsilon$ . For the Deffuant-Weisbuch model, the number of polarized clusters is  $\mathcal{O}(1/\epsilon)$  [101] in the fully-mixed case. In Ref. [35], the authors extend this model to hypergraphs and show that even if community structure exists, if the communities are not completely disconnected, the probability that polarization exists approaches zero as the number of nodes approaches infinity. However, if one limits the intercommunity edge size, polarization is a stable state for any density of inter- or intra-community edges.

In addition to bounded-confidence models, researchers have examined cooperation and partisanship [102], echo chambers on social media [94], news media influence [103], moderates seen as outsiders [104], Congress partisanship [105], and many other topics. In Ref. [106], the authors show that in the case of complex contagion on pairwise networks, particularly the threshold model, community structure leads to a greater number of fixed points which can be thought of as polarized factions.

The dynamical behavior of a complex system is determined by the interplay between the underlying interaction structure and the dynamical system being modeled. In many approaches, however, the focus is on models which, while incorporating sensible mechanisms modeling polarization, will show evidence of polarization in completely homogeneous populations. We examine this problem from a different perspective; rather than focus on a realistic model that will create polarization, we implement a simple hypergraph contagion model and show that with strong enough community structure that polarization can occur. This perspective complements current research on this topic; polarization is likely the result of both structural and dynamical factors. One way that this may occur is an adaptive model where, in addition to opinion dynamics, the structure of

a network may change over time to optimize opinion homophily [107].

In Ref. [108], the authors examined the effect of community structure on the dynamical behavior of hypergraph contagion models. The authors considered a binary state stochastic dynamical model where active nodes spontaneously become inactive according to a Poisson process with a rate  $\delta$  and inactive nodes become active by edge  $e_j$  at a rate  $\lambda_j$  if more than  $\Theta_j$  nodes in edge  $e_j$  are active. Using this model, the authors found the existence of bimodal distributions of active nodes caused by *multistability*, where there are multiple stable fixed points, and *intermittency*, where there is temporal switching between fixed points. We look at a more specific version of this model where  $\lambda_j = \lambda_{|j|}$  and  $\Theta_j = |e_j| - 1$  for edge sizes of 2 and 3, also known as the *hypergraph SIS model* in Ref. [12]. We similarly find that polarization can occur for sufficiently strong community structure in hypergraphs and explore theoretical predictions for the regimes at which this can occur. We also explore the effect of imbalanced communities and offer efficient algorithms for the generation of synthetic hypergraph datasets with community structure.

The structure of this Chapter is as follows: in Section 5.1 we discuss the theoretical generative and contagion models that we use; in Section 5.2, we define and analyze a mean-field model including community structure; and in Section 5.3, we discuss our results.

## 5.1 Model

In this section, we describe the generative hypergraph models and opinion formation model that we use. We define simplistic generative models that incorporate community structure so that we can analyze how the dynamical behavior depends on the strength of the community structure. We refer to Chapter 2 for basic terminology and notation.

### 5.1.1 The stochastic block model for uniform hypergraphs

The simplest random network model incorporating community structure is the *stochastic block model* (SBM), which assumes that the probability that two nodes  $i$  and  $j$  are connected with an edge depends only on their community labels  $g_i$  and  $g_j$ , given by  $p_{g_i, g_j}$ . Extending this notion to

$m$ -uniform hypergraphs gives rise to what is known as the  $m$ -uniform hypergraph stochastic block model ( $m$ -HSBM). Studies have examined this model with two communities of random size [109] and an arbitrary number of communities of equal size [110]. Several studies have looked at this model for general hypergraphs, known as the hypergraph stochastic block model (HSBM) [35, 111].

Given an  $m$ -uniform hypergraph  $H = (V, E)$  where  $N = |V|$  is the number of nodes and  $E$  is the set of hyperedges where  $|e| = m$  for every  $e \in E$ . Suppose that each node  $i$  has a community label,  $g_i$  and that the number of groups is given by  $G$ . Then we define  $P \in [0, 1]^{G \times \dots \times G}$  as the  $m$ -dimensional tensor which, given nodes  $i_1, \dots, i_m$  and community labels  $g_{i_1}, \dots, g_{i_m}$ , specifies the probability of forming a hyperedge as  $P_{g_{i_1}, \dots, g_{i_m}}$ . Because hyperedges are sets and therefore every vertex order is equivalent, we specify that  $P$  is a symmetric tensor, i.e.,  $P_{i_1, \dots, i_m} = P_{\sigma_{i_1}, \dots, \sigma_{i_m}}$ , where  $\sigma$  is any permutation of the indices. When sampling from this random model, naïve algorithms iterate through every edge combination leading to  $\mathcal{O}(N^m)$  complexity. In Appendix C.1, we present more efficient algorithms for sampling from this random model.

### 5.1.2 Planted partition model for uniform hypergraphs

If we add additional constraints to the SBM, namely that there are two equally-sized communities, the mean degree of the network remains fixed, and that a single parameter controls the community structure, we obtain the *planted partition model* for pairwise networks [112]. Consider a pairwise network of size  $N$  with a mean degree of  $\langle k \rangle$ . We define the *imbalance parameter*,  $\epsilon$ , as the difference between the mean degree restricted to links contained within a single group,  $\langle k_{in} \rangle$ , and the mean degree restricted to links connecting two groups,  $\langle k_{out} \rangle$ . The total number of connections that a node has is the sum of its connections within its group and between two groups, i.e.,  $\langle k \rangle = \langle k_{in} \rangle + \langle k_{out} \rangle$ . When  $\epsilon = 0$ , this model specifies an Erdős-Rényi network, and when  $\epsilon = \langle k \rangle$ , we obtain a network with two completely disconnected communities. If nodes  $i$  and  $j$  are within a single community, we connect them with probability  $p_{1,1} = p_{2,2} = p_{in} = p_{ER} + \Delta$  and if nodes  $i$  and  $j$  are in communities 1 and 2 respectively, we connect them with probability  $p_{1,2} = p_{2,1} = p_{in} = p_{ER} - \Delta$ . In both cases, the  $p_{ER}$  term specifies the Erdős-Rényi connec-

tion probability and the second term varies the ratio of the density of links within and between communities. Assuming a simple undirected network,  $N\langle k \rangle = 2\binom{N}{2}p_{ER}$  and we obtain

$$p_{ER} = \frac{\langle k \rangle}{N-1}.$$

Likewise, because  $\langle k_{in} \rangle N/2 = 2\binom{N}{2}p_{in}$ ,

$$\Delta = \frac{2\frac{N}{N-2}\langle k_{in} \rangle - \frac{N}{N-1}\langle k_{in} \rangle - \frac{N}{N-1}\langle k_{out} \rangle}{N}.$$

As  $N \rightarrow \infty$ ,  $p_{ER} = \langle k \rangle/N$  and  $\Delta = \epsilon/N$  asymptotically.

Now we extend this idea to  $m$ -uniform hypergraphs. We assume that intra-community hyperedges are formed with probability  $p_{in}$  and inter-community hyperedges are formed with probability  $p_{out}$  (as in Refs. [110, 109, 64]). We derive the values of  $p_{in}$  and  $p_{out}$  in terms of a specified mean degree  $\langle k \rangle$  and imbalance parameter  $\epsilon_m$ . We consider an  $m$ -uniform hypergraph  $H$  of size  $N$  with mean degree  $\langle k \rangle$  and two communities, both of size  $N/2$ . We say that a hyperedge  $e = \{i_1, \dots, i_m\}$  is *intra-community* if all its members have the same group label and *inter-community* otherwise, in accordance with the ‘‘All-or-Nothing’’ (AoN) definition in Ref. [113]. In principle, one can relax this binary classification to account for the number of communities of which the hyperedge is a part and the proportion of nodes in each community [113], but for simplicity, we do not consider this case. For size  $m$ -hyperedges, the AoN assumption leads to a factor of  $2^{m-1} - 1$  more possible inter-community edges than intra-community edges. To enforce a constant mean degree while varying the imbalance parameter, we scale the intra- and inter-community connection probabilities as

$$p_{in} = p_{ER} + (2^{m-1} - 1)\Delta, \quad (5.1)$$

$$p_{out} = p_{ER} - \Delta. \quad (5.2)$$

Constraining  $N\langle k \rangle = m\binom{N}{m}p_{ER}$  for an  $m$ -uniform Erdős-Rényi uniform hypergraph and employing the asymptotic approximation  $\binom{N}{m} \approx N^m/m!$  for  $N \gg m$ , we obtain

$$p_{ER} = \frac{(m-1)!\langle k \rangle}{N^{m-1}}.$$

Likewise, enforcing that  $\langle k_{in} \rangle N/2 = m \binom{N/2}{m} p_{in}$  and using the same asymptotic approximation,

$$\Delta = \frac{(m-1)! \epsilon_m}{N^{m-1}},$$

where  $\epsilon_m = \langle k_{in} \rangle - \langle k_{out} \rangle / (2^{m-1} - 1)$  is the *size- $m$  imbalance parameter*. When  $m = 2$ , we recover the probabilities derived for pairwise networks. The two communities are completely disconnected when  $\epsilon_m = \langle k \rangle$  and we recover an  $m$ -uniform Erdős-Rényi hypergraph when  $\epsilon_m = 0$ . When  $\epsilon_m < 0$ , nodes in the same groups tend to connect with each less often than nodes in different classes. For pairwise networks, a bipartite network is obtained when  $\epsilon_2 = -\langle k \rangle$ . Any hypergraph can be partitioned into its uniform “layers” [114] and so we can create a hypergraph from its uniform hypergraph constituents, each with their own community structure.

### 5.1.3 The degree-corrected stochastic block model for uniform hypergraphs

When trying to model empirical hypergraph data sets, incorporating the degree distribution into models of community structure can provide additional accuracy [115] and has been used extensively in community detection and inference [116]. The *degree-corrected stochastic block model* (DCSBM) presented in Ref. [115] is defined as follows: consider an undirected, pairwise network of size  $N$ , where each node  $i$  has  $k_i$  neighbors and a community label  $g_i$ . The probability of a link being formed between two nodes  $i$  and  $j$  is defined as

$$p_{ij} = \frac{k_i}{\kappa_{g_i}} \frac{k_j}{\kappa_{g_j}} \omega_{g_i, g_j},$$

where  $\kappa_{g_i}$  is the sum of the degrees of the nodes in group  $g_i$  and  $\omega$  is a symmetric matrix where  $\omega_{g_i, g_j}$  specifies the number of links between communities  $g_i$  and  $g_j$ . The matrix  $\omega$  must satisfy  $\sum_i k_i = \sum_{i,j} \omega_{i,j}$  so that each node’s expected degree and target degree are equal to each other.

Extending this notion to an  $m$ -uniform hypergraph, which we call the  $m$ -uniform hypergraph degree-corrected stochastic block model ( $m$ -HDCSBM), where each node  $i$  belongs to a community  $g_i$  and is a member of  $k_i$   $m$ -hyperedges, we similarly define the connection probability between  $m$  nodes,  $\{i_1, \dots, i_m\}$  to be

$$p_{i_1, \dots, i_m} = \frac{k_{i_1}}{\kappa_{g_{i_1}}} \dots \frac{k_{i_m}}{\kappa_{g_{i_m}}} \omega_{g_{i_1}, \dots, g_{i_m}}, \quad (5.3)$$



Similar to the tensor  $P$  for the  $m$ -HSBM,  $\omega$  is a symmetric tensor of dimension  $m$ . The term  $\kappa_{g_i}$  is the sum of the degrees of the nodes that belong to community  $g_i$ . A similar framework for uniform hypergraphs is presented in Ref. [117] and Ref. [113] presents a model for non-uniform hypergraphs with varying affinity functions controlling the community structure.

#### 5.1.4 Opinion model

Similar to Refs. [12, 11], we study the hypergraph contagion model with the *collective contagion model* described in Ref. [12]. This model allows nodes to hold binary opinions: susceptible (S) and infected (I). An infected node spontaneously transitions to the susceptible state at a rate  $\gamma$  independently of the states of neighboring nodes. A susceptible node may transition to the infected state at a rate  $\beta_m$  if that node is a member of a group of size  $m$  and all of the other members in this group are infected. In this work, we associate the susceptible state with the “0” state and the infected state with the “1” state.

We comment that that this model contains asymmetry in the contagion contagion; in the absence of influence or infection by other groups an individual will almost surely heal (if  $\gamma > 0$ ) given enough time. Likewise, the opinion corresponding to the infected state may only be sustained if the opinion is continually shared with neighbors. This is akin to a peer pressure effect where a phenomenon will likely die out on its own and needs group pressure to sustain it. Both opinions are not interchangeable; each has distinctive dynamical features.

## 5.2 Mean-field analysis

### 5.2.1 Planted partition model

Consider a hypergraph of size  $N$  with hyperedges of sizes 2 (links) and 3 (triangles). We construct a planted partition model of the 2-uniform and 3-uniform subsets of the hypergraph with parameters  $\epsilon_2$  and  $\epsilon_3$  respectively as described in Section 5.1.2. We denote the fraction of infected individuals in each community as  $x_1$  and  $x_2$  and extend the mean-field model described for Erdős-

Rényi hypergraphs in Ref. [11] to include community structure. Using the same notation as in Ref. [12], we define  $\langle k \rangle$  as the mean link degree,  $\langle q \rangle$  as the mean triangle degree,  $\gamma$  as the rate of healing,  $\beta_2$  as the pairwise infectivity, and  $\beta_3$  as the triangle infectivity. Without loss of generality, we derive a rate equation for the fraction of infected nodes in the first community,  $x_1$ . Firstly, for the SIS model, nodes heal independently of their neighbors' states at a rate  $\gamma$  so the healing term is  $-\gamma x_1$ . For the pairwise infection term, the fraction of susceptible nodes is  $(1 - x_1)$  and there are  $N/2$  nodes in each community. The probability that two nodes are connected to each other is  $(\langle k \rangle + \epsilon_2)/N$  if they belong to the same community and  $(\langle k \rangle - \epsilon_2)/N$  if they belong to different communities. We multiply by the probability that a single node is infected which is  $x_1$  and  $x_2$  for communities 1 and 2 respectively. Therefore, the pairwise infection term becomes

$$\frac{\beta_2}{2}(1 - x_1)[(\langle k \rangle + \epsilon_2)x_1 + (\langle k \rangle - \epsilon_2)x_2].$$

For the triangle infection term, the number of unique triangles in each community given a fixed susceptible node is  $\binom{N/2}{2} \approx N^2/8$ . Assuming independence of each node's infection state, the fraction of edges where both neighboring nodes are infected is  $x_i x_j$  depending on the group memberships of nodes  $i$  and  $j$ . The probability that three nodes form a 3-hyperedge is  $2(\langle q \rangle + 3\epsilon)/N^2$  if they are in the same community and  $2(\langle q \rangle - \epsilon)/N^2$  if they are in different communities. Then the triangle infection rate is

$$\frac{\beta_3}{4}(1 - x_1)[(\langle q \rangle + 3\epsilon)x_1^2 + 2(\langle q \rangle - \epsilon)x_1 x_2 + (\langle q \rangle - \epsilon)x_2^2].$$

Putting these terms together we get the following system of equations.

$$\begin{aligned} \frac{dx_1}{dt} = & -\gamma x_1 + \frac{\beta_2}{2}(1 - x_1)[\langle k \rangle(x_1 + x_2) + \epsilon_2(x_1 - x_2)] \\ & + \frac{\beta_3}{4}(1 - x_1)[\langle q \rangle(x_1 + x_2)^2 + \epsilon_3(3x_1^2 - 2x_1 x_2 - x_2^2)], \end{aligned} \quad (5.4)$$

$$\begin{aligned} \frac{dx_2}{dt} = & -\gamma x_2 + \frac{\beta_2}{2}(1 - x_2)[\langle k \rangle(x_1 + x_2) + \epsilon_2(x_2 - x_1)] \\ & + \frac{\beta_3}{4}(1 - x_2)[\langle q \rangle(x_1 + x_2)^2 + \epsilon_3(3x_2^2 - 2x_1 x_2 - x_1^2)]. \end{aligned} \quad (5.5)$$

This system has a fixed point at  $x_1 = x_2 = 0$ . We linearize about this fixed point to obtain

an equation for the perturbations,  $\delta x_1$  and  $\delta x_2$ . The equations governing the perturbations become

$$\begin{pmatrix} \frac{d\delta x_1}{dt} \\ \frac{d\delta x_2}{dt} \end{pmatrix} = \beta_2 \begin{pmatrix} \frac{\langle k \rangle + \epsilon_2}{2} - \frac{\gamma}{\beta_2} & \frac{\langle k \rangle - \epsilon_2}{2} \\ \frac{\langle k \rangle - \epsilon_2}{2} & \frac{\langle k \rangle + \epsilon_2}{2} - \frac{\gamma}{\beta_2} \end{pmatrix} \begin{pmatrix} \delta x_1 \\ \delta x_2 \end{pmatrix}. \quad (5.6)$$

The two eigenvalues of this system are  $\lambda_1 = \beta_2 \langle k \rangle - \gamma$  and  $\lambda_2 = \beta_2 \epsilon_2 - \gamma$ , and because  $\epsilon_2 \leq \langle k \rangle$ ,  $\lambda_1$  is the maximal eigenvalue and the epidemic threshold is  $\beta_2^c / \gamma = 1 / \langle k \rangle$ . Now we assume that  $x_1 = x_2$  and solve for the non-zero fixed point. Substituting  $x_1 = x_2 = x$ , setting the left hand side of Eq. (5.4) equal to zero and canceling the zero solution, we obtain

$$(\beta_2 \langle k \rangle - \gamma) + (\beta_3 \langle q \rangle - \beta_2 \langle k \rangle)x - \beta_3 \langle q \rangle x^2 = 0.$$

Evaluating at  $\beta_2 = \beta_2^c$ , we obtain

$$[(\beta_3 \langle q \rangle - \gamma) - \beta_3 \langle q \rangle x] x = 0.$$

Again, canceling the zero solution, we obtain the fixed point  $x = 1 - \gamma / (\beta_3 \langle q \rangle)$ . For this fixed point to be realizable,  $x > 0$  which predicts that bistability occurs when  $\beta_3 / \gamma > 1 / \langle q \rangle$ . Both the epidemic threshold and the onset of bistability are independent of  $\epsilon_2$  and  $\epsilon_3$  respectively. Therefore, this analysis indicates that balanced community structure has no effect on the epidemic threshold or the onset of bistability when compared to the simpler Erdős-Rényi model. Earlier studies seem to corroborate this result for pairwise networks [118, 119].

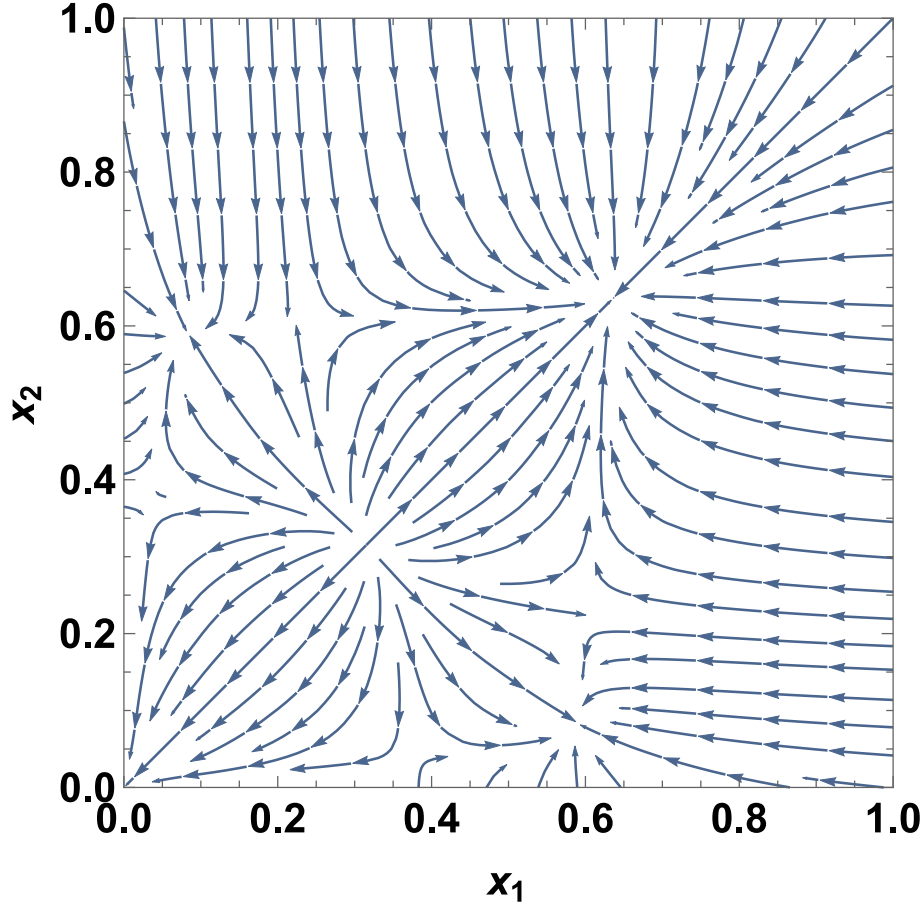


Figure 5.1: A phase plot of Eqs. (5.7) - (5.8) with  $\tilde{\epsilon}_2 = 0.5$ ,  $\tilde{\epsilon}_3 = 0.95$ ,  $\tilde{\beta}_2 = 0.2$ , and  $\tilde{\beta}_3 = 4$ .

We use these results to perform a rescaling. Without loss of generality, assume that  $\gamma = 1$ . Then we define rescaled variables  $\tilde{\beta}_2 = \beta_2/\beta_2^c = \beta_2\langle k \rangle/\gamma$ ,  $\tilde{\beta}_3 = \beta_3/\beta_3^c = \beta_3\langle q \rangle/\gamma$ ,  $\tilde{\epsilon}_2 = \epsilon_2/\langle k \rangle$ , and  $\tilde{\epsilon}_3 = \epsilon_3/\langle q \rangle$ . When  $\epsilon_2$  and  $\epsilon_3$  are zero, this corresponds to the Erdős-Rényi case for 2- and 3-hyperedges respectively, and when they are one, this corresponds to completely disconnected communities by 2- and 3-hyperedges respectively. Then Eqs. (5.4)-(5.5) become

$$\begin{aligned} \frac{dx_1}{dt} = & -x_1 + \frac{\tilde{\beta}_2}{2}(1-x_1)[(x_1+x_2) + \tilde{\epsilon}_2(x_1-x_2)] \\ & + \frac{\tilde{\beta}_3}{4}(1-x_1)[(x_1+x_2)^2 + \tilde{\epsilon}_3(3x_1^2 - 2x_1x_2 - x_2^2)], \end{aligned} \quad (5.7)$$

$$\begin{aligned} \frac{dx_2}{dt} = & -x_2 + \frac{\tilde{\beta}_2}{2}(1-x_2)[(x_1+x_2) + \tilde{\epsilon}_2(x_2-x_1)] \\ & + \frac{\tilde{\beta}_3}{4}(1-x_2)[(x_1+x_2)^2 + \tilde{\epsilon}_3(3x_2^2 - 2x_1x_2 - x_1^2)]. \end{aligned} \quad (5.8)$$

In our analysis, we assumed that  $x_1 = x_2$ , which may not be the case. Indeed, as shown in Fig. 5.1, for certain structural and infectious parameter regimes, we see that there are four stable fixed points: two symmetric ( $x_1 = x_2$ ) fixed points corresponding to bistable behavior and two additional asymmetric ( $x_1 \neq x_2$ ) fixed points where  $x_1 \neq x_2$ . This behavior is not present for the hypergraph contagion model on random null models without community structure. The physical interpretation of this phenomenon is that when two communities are sufficiently disconnected, they are able to, on average, hold different opinions. We refer to this phenomenon as *polarization* and quantify the polarization between two communities  $i$  and  $j$ ,  $\psi_{ij}$ , as the maximum distance between their stable fixed points given a hypergraph and a set of infection parameters, i.e.,  $\psi_{ij} = \max\{|x_i - x_j| \text{ such that } (x_i, x_j) \text{ is a stable fixed point of Eqs. (5.4)-(5.5)}\}$  (or, equivalently, Eqs. (5.7)-(5.8)).

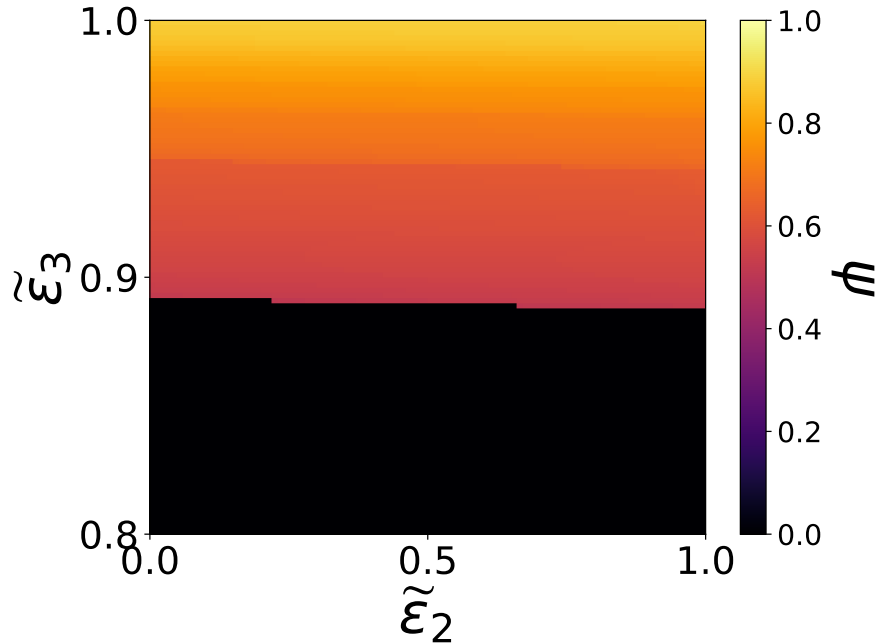


Figure 5.2: A plot of the polarization in Eqs. (5.7)-(5.8) for different values of  $\epsilon_2$  and  $\epsilon_3$ .

When there are only two communities we let  $\psi_{12} = \psi_{21} \equiv \psi$ . When asymmetric fixed points do not exist, the polarization is 0. In the planted partition model, asymmetric fixed points exist

for certain regimes of  $\epsilon_2$  and  $\epsilon_3$ . Using Eqs. (5.4)-(5.5), we can quantify the regimes of  $\epsilon_2$  and  $\epsilon_3$  for which polarization occurs. We do this numerically by doing the following: first, we set the derivatives in Eqs. (5.4)-(5.5) equal to zero; second, we specify that the asymmetric roots should exist in the interval  $[0, 1] \times [0, 1]$ ; third, we specify fixed values of  $\epsilon_2$  and  $\epsilon_3$ ; and fourth, we employ interval root finding methods to find all fixed points of these 2D coupled equations in the specified interval and iterate over values of  $\beta_2 \in [0.01\beta_2^c, 0.95\beta_2^c]$  and  $\beta_3 \in [2\beta_3^c, 10\beta_3^c]$  holding  $\gamma = 1$  fixed. After generating a list of all fixed points for each combination of  $\beta_2, \beta_3$ , we select the stable fixed point with the largest value of  $|x_1 - x_2|$ . The stability of each fixed point is determined by evaluating the spectral abscissa of the Jacobian at that fixed point. We repeat this process for all specified values of  $\epsilon_2$  and  $\epsilon_3$ .

In Fig. 5.2, we see that the model is much more sensitive to the sparsity of 3-hyperedge connections between communities than it is to pairwise interactions between communities; the link structure has very little effect on the polarization.

### 5.2.2 Imbalanced communities

Up until this point, we have assumed that the two communities are the same size. We now relax this assumption and specify that a fraction,  $\rho$ , of nodes belongs to the first community and the remaining fraction of nodes,  $1 - \rho$ , belongs to the second community. As before, we seek to preserve the expected number of hyperedges regardless of the value of  $\epsilon_m$ . We denote the probability of forming a hyperedge solely containing members of the first community and solely containing members of the second community as  $p_{1,1}$  and  $p_{2,2}$  respectively. We define  $p_{1,2}$  as the probability of forming a hyperedge containing nodes from both communities. The average number of hyperedges solely containing members of the first community and solely containing members of the second community to which a node is connected is denoted  $\langle k_{1,1} \rangle$  and  $\langle k_{2,2} \rangle$  respectively. The average number of hyperedges containing members from both communities to which a node is connected is denoted  $\langle k_{1,2} \rangle$ . Because the number of unique hyperedges in the first and second communities is  $\binom{\rho N}{m}$  and  $\binom{N - \rho N}{m}$  respectively and following the same steps as in Section 5.1.2, we

derive that

$$p_{1,1} = \frac{(m-1)! \langle k_{1,1} \rangle}{(\rho N)^{m-1}} \quad (5.9)$$

and

$$p_{2,2} = \frac{(m-1)! \langle k_{1,1} \rangle}{([1-\rho]N)^{m-1}}. \quad (5.10)$$

Because these communities are of different sizes, we can make one of two assumptions: (1) that  $\langle k_{1,1} \rangle = \langle k_{2,2} \rangle$  or (2) that  $p_{1,1} = p_{2,2}$ . If we enforce that  $p_{1,1} = p_{2,2}$ , then  $\langle k_{1,1} \rangle$  and  $\langle k_{2,2} \rangle$  must differ, and vice-versa. For both cases, the number of unique connections between the two communities is the number of leftover connections,  $\binom{N}{m} - \binom{\rho N}{m} - \binom{N-\rho N}{m} \approx (1 - \rho^m - [1-\rho]^m)N^m/m!$  and so

$$p_{1,2} = \frac{(m-1)! \langle k_{1,2} \rangle}{(1 - \rho^m - [1-\rho]^m)N^{m-1}}. \quad (5.11)$$

Examining the first case where  $\langle k_{1,1} \rangle = \langle k_{2,2} \rangle = \langle k_{in} \rangle$ . Then Eq. (5.11) can be simplified to

$$p_{out} = p_{ER} - \frac{(m-1)! \epsilon_{\rho,m}}{N^{m-1}},$$

where  $\epsilon_{\rho,m} = \langle k_{in} \rangle - \langle k_{out} \rangle / r_{\rho,m}$  and  $r_{\rho,m} = (\rho^m + [1-\rho]^m)^{-1} - 1$ . When  $\rho = 1/2$ , we recover the expression for the size- $m$  imbalance parameter that we derived previously. Looking at both communities together,  $\langle k_{in} \rangle N = m \left[ \binom{\rho N}{m} - \binom{N-\rho N}{m} \right] p_{in}$ , and so

$$\begin{aligned} p_{in} &= \frac{(m-1)! \langle k_{in} \rangle}{(\rho^m + [1-\rho]^m)N^{m-1}} \\ &= p_{ER} + r_{\rho,m} \frac{(m-1)! \epsilon_{\rho,m}}{N^{m-1}}. \end{aligned} \quad (5.12)$$

Because the mean degrees are identical for each community, we can calculate the probabilities for each community:  $p_{1,1} = (\rho^m + [1-\rho]^m)/\rho^{m-1} p_{in}$  and  $p_{2,2} = (\rho^m + [1-\rho]^m)/[1-\rho]^{m-1} p_{in}$ . For case (2) where  $p_{1,1} = p_{2,2} = p_{in}$  we use Eq. (5.12) along with Eqs. (5.9) and (5.10) to derive  $\langle k_{1,1} \rangle = \rho^m/(\rho^m + [1-\rho]^m) \langle k_{in} \rangle$  and  $\langle k_{1,1} \rangle = [1-\rho]^m/(\rho^m + [1-\rho]^m) \langle k_{in} \rangle$ .

For the second case, performing the same calculations as in Section 5.2.1, we obtain

$$\begin{aligned} \frac{dx_1}{dt} = & -\gamma x_1 + \beta_2(1-x_1)[\rho(\langle k \rangle + r_{\rho,2}\epsilon_{\rho,2})x_1 + (1-\rho)(\langle k \rangle - \epsilon_{\rho,2})x_2] \\ & + \beta_3(1-x_1)[\rho^2(\langle q \rangle + r_{\rho,3}\epsilon_{\rho,3})x_1^2 + (1-\rho)(\langle q \rangle - \epsilon_{\rho,3})(2x_1x_2 + x_2^2)], \end{aligned} \quad (5.13)$$

$$\begin{aligned} \frac{dx_2}{dt} = & -\gamma x_2 + \beta_2(1-x_2)[(1-\rho)(\langle k \rangle + r_{\rho,2}\epsilon_{\rho,2})x_2 + \rho(\langle k \rangle - \epsilon_{\rho,2})x_1] \\ & + \beta_3(1-x_2)[(1-\rho)^2(\langle q \rangle + r_{\rho,3}\epsilon_{\rho,3})x_2^2 + \rho(\langle q \rangle - \epsilon_{\rho,3})(2x_2x_1 + x_1^2)], \end{aligned} \quad (5.14)$$

and to obtain the mean-field equation corresponding to the first case, we can simply multiply the first terms in the square brackets on each line of Eq. (5.14) by the multipliers relating  $p_{in}$  to  $p_{1,1}$  and  $p_{2,2}$  respectively, which restores symmetry to the terms relating to neighbors in the same community. Unlike the balanced planted partition model, the derived epidemic threshold now has a dependence on  $\epsilon_2$  for both cases.



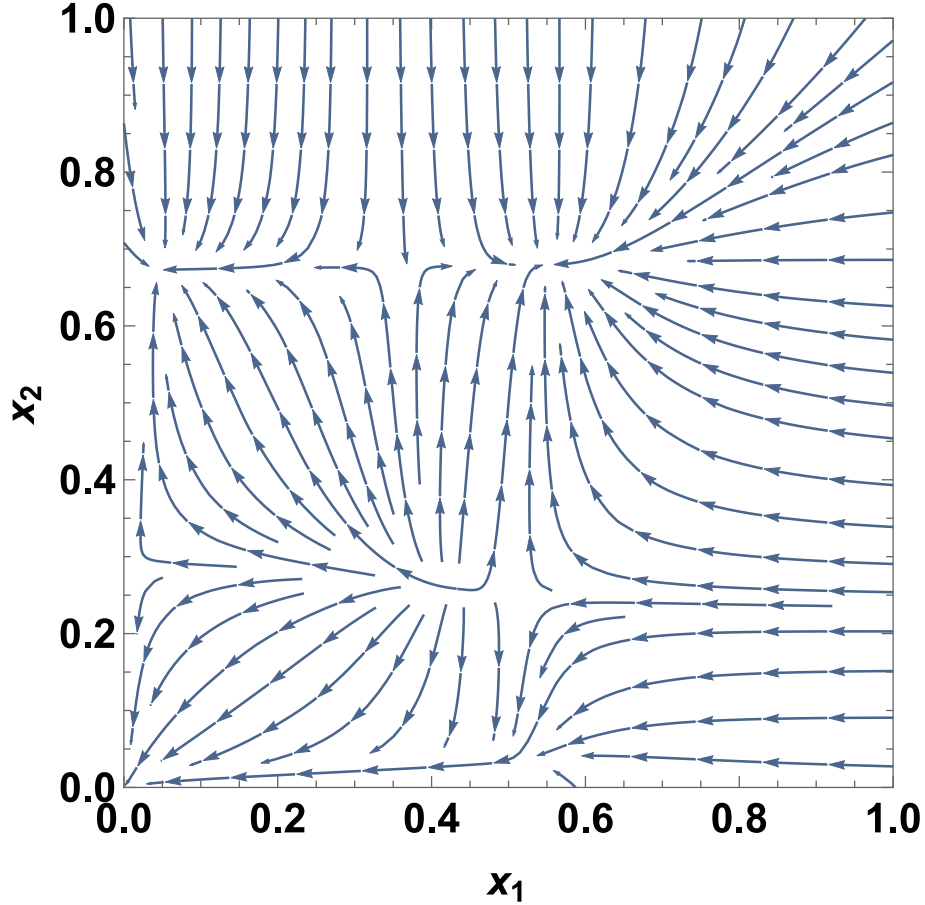


Figure 5.3: A phase plot of Eqs. (5.13) - (5.14) (the imbalanced case where  $p_{1,1} = p_{2,2}$ ) with  $\rho = 0.48$ ,  $\langle k \rangle = \langle q \rangle = 20$ ,  $\epsilon_2 = 10$ ,  $\epsilon_3 = 20$ ,  $\beta_2 = 0.2\beta_2^c$ ,  $\beta_3 = 4\beta_3^c$ , and  $\gamma = 1$ .

In Fig. 5.3, we can see that changing the relative sizes of the communities even slightly destroys the symmetry of the phase plot such that the only symmetric fixed point is  $x_1 = x_2 = 0$ . In addition, the asymmetric fixed point where  $x_1 \gg x_2$  is no longer stable and the asymmetric fixed point where  $x_1 \ll x_2$  develops a larger basin of attraction. For small enough  $\rho$ , the only remaining stable fixed point is the one where  $x_1 \ll x_2$ . This makes physical sense because as  $\rho$  decreases, the size of the first community decreases and there are fewer neighbors to sustain the contagion in the first community.

### 5.3 Discussion

We presented a mean-field model of hypergraph contagion incorporating community structure and used this model to examine the structural characteristics of hypergraphs for which polarization occurs. We also modified this framework to explore the effect of changing the relative sizes of communities.

In this study, we solely considered the hypergraph SIS model, which is a simple model of social contagion. In the future, it may be fruitful to consider more realistic models of opinion formation. In addition, we assumed that the dynamics of these opinion formation processes do not affect the underlying structure of the hypergraph. In empirical case studies, the formation of echo chambers may be the result of intrinsic opinion formation dynamics and the hypergraph adapting to dissolve interactions with heterogeneous opinions.

Empirical systems often contain degree heterogeneity and deriving additional results for the  $m$ -HDCSBM will help bridge the gap between theory and empirical datasets. Numerical experiments through stochastic simulations on the  $m$ -HSBM and the  $m$ -HDCSBM as well as empirical datasets will help to validate these results. In addition, using statistical inference to fit empirical datasets the  $m$ -HDCSBM will improve our numerical results.

Nonetheless, our work offers a helpful framework for predicting the existence of polarization and shows that simple models of contagion can drive polarization in systems with community structure and group interactions.

## Chapter 6

### Effect of time-dependent infectiousness on epidemic dynamics

In traditional literature, the SIR model is a canonical example of modeling the spread of disease with total immunity. This model has common extensions such as the SEIR (Susceptible – Exposed – Infected – Recovered) when one wants to incorporate a latent period which captures delays between transmission and infectiousness. With most of these models, however, a key assumption is that an individual’s infectivity is constant. However, we know that an individual’s infectiousness varies over the duration of the infection, according to their viral load [39, 40]. We define a framework to extend the SIR model by dividing the single infectious compartment into  $n$  stages as has been considered by Ref. [120], known as the  $SI^KR$  model in Ref. [16], and assigning each stage a different infection rate as in Refs. [121, 41]. Other approaches have been considered, such as the message-passing approach [85, 122], mapping an individual’s viral load to an infection probability [123], and looking at an infection density function [16, 124]. We use this approach to examine fully-mixed populations and theoretical networks constructed from category-based mixing, both static and temporal.

The structure of this Chapter is as follows. In Section 6.1 we describe a framework for modeling time-dependent infectiousness. In Section 6.2 we use this model to create theoretical predictions for the reproductive number, apply these predictions to several common cases, and validate our theory with numerical simulations. Lastly, in Section 6.3 we discuss the implications of our theory.

## 6.1 Model

We propose a general mean-field model to describe the spread of an epidemic including time-dependent infectiousness. In the following, we will refer to this model as the viral load (VL) model.

We consider a population of  $N$  nodes. We assume that a node  $i$ 's intrinsic infectiousness is solely determined by the amount of time it has been infected,  $\tau$ , and its corresponding viral load at that time, denoted  $v_i(\tau)$ , although other factors may be involved as well [125]. Several studies have examined the correspondence between an individual's viral load and their infectiousness [39, 40] but for this study, we simply define  $\beta_i(\tau)$ , the infectious rate function, as the rate at which node  $i$  transmits infection having been infected for a duration of time  $\tau$ . Note that in the case where an infectious threshold exists [126, 123], we can express the function as  $\beta_i(\tau)I_{\tau \in \delta}$ , where  $\delta = \{\tau \mid \beta_i(\tau) \geq \eta\}$  and  $\eta$  is the infectious threshold. This infectious rate function can vary in response to many factors such as asymptomatic versus symptomatic infection or severity of symptoms and be considered as being drawn according to some distribution. For much of this study, however, we assume that while  $\beta_i(\tau)$  is heterogeneous in time, that every member of the population has the same infectious rate function, i.e.,  $\beta_i(\tau) = \beta(\tau)$ ,  $i = 1 \dots N$ , although we relax this assumption later. We assume that nodes start in the susceptible compartment ( $S$ ) and that an infected individual infected for time  $\tau$  infects a susceptible node with rate  $\beta(\tau)$ . We approximate  $\beta(\tau)$  by evaluating it at  $n$  discrete times  $\tau_j = j\Delta\tau$ , where  $\Delta\tau$  is fixed and  $n\Delta\tau = \tau_R$ , the recovery time. Then we divide the infectious compartment,  $I$  into  $n$  stages,  $I_j$ ,  $j = 1 \dots n$ , each with an associated infection rate  $\beta_j$ , in a similar manner to Refs. [120, 41]. Lastly, nodes that transition through all infection states accumulate in the recovered ( $R$ ) compartment.

We assume that the flow of infected individual between subsequent infectious compartments is deterministic and that upon entering the first infectious stage, an individual passes through all the subsequent stages as shown in Fig. 6.1, meaning that  $\gamma_i = 1/\Delta\tau$  where  $\Delta\tau = \tau_R/n$ .

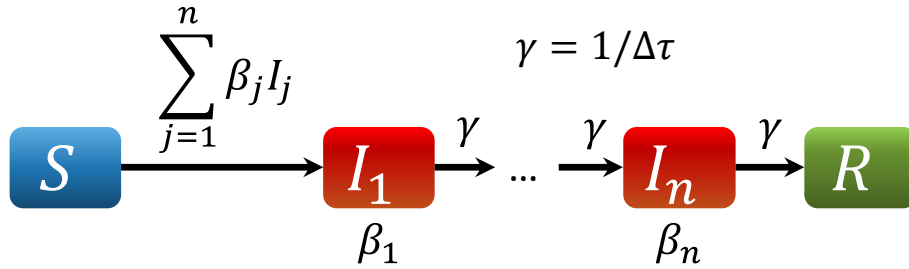


Figure 6.1: An illustration of the VL model.

In the following, we define the *m*-th moment of a quantity  $q$  as  $\langle q^m \rangle = \sum_{i=1}^N q_i^m / N$  when  $q$  is a discrete quantity and as  $\langle q^m(\tau) \rangle = \int_0^{\tau_R} [q(\tau)]^m d\tau / \tau_R$  when  $q$  is a continuous function of  $\tau$ .

There are many studies exploring the effect of more realistic infectious behavior. In Ref. [120], the authors use the  $n$ -stage  $SI^KR$  model with constant infectiousness on a fully-mixed network so that the infectious waiting time is gamma-distributed. In Ref. [41], the authors explore the  $SI^KR$  model with variable infectiousness for fully-mixed networks. For both of these models, the authors allow healing and recovery to occur at every infectious stage. In Ref. [127], the authors explore the  $SI^KR$  link-closure model with a constant infection rate and solely consider static networks. They simulate their model numerically on homogeneous and Erdős-Rényi networks. In Ref. [85], the authors consider a message-passing approach to model time-dependent infectiousness and simulate their results on a static network. In Ref. [122], the authors present a non-Markovian edge-based compartment model, prove its equivalence to the message-passing model, and describe how other models compare to the message-passing approach. In Refs. [120, 41] the authors solely consider the fully-mixed case and in Refs. [127, 122, 85] the authors solely consider static networks. In contrast, our approach encompasses fully-mixed, static, and temporal networks. In Refs. [120, 127], although the authors consider an  $SI^KR$  model, they specify that the infectious rate is constant in contrast to our model where we allow the rate to vary over time. In addition, Refs. [120, 41, 127, 127] assume Markovian transitions between infectious states in contrast to our approach which enforces deterministic transitions between infectious states (as in Ref. [123]).

## 6.2 Derivation of the population reproductive number

We derive the reproductive number for the viral load model described above that has been cast as a system of mean-field ODEs. First, we derive the reproductive number for a fully-mixed model and second, we derive the reproductive number for an arbitrary category-mixed population. We comment on the continuum limit for both cases and derive specific closed-form solutions for the reproductive number for a configuration model static network, and an activity model temporal network.

### 6.2.1 Fully-mixed population

Consider a fully-mixed population of  $N$  individuals and an infectious rate function,  $\beta(\tau)$ . In our formalism, we denote the fraction of the population in the susceptible,  $j$ th infectious stage, and the recovered stage as  $S$ ,  $I_j$ ,  $j = 1 \dots n$ , and  $R$  respectively and note that  $S + \sum_{j=1}^n I_j + R = 1$  by conservation. Assuming that an individual's infection status is independent of the infection status of its neighbors, as done in Ref. [41], we can write the following system of mean-field equations as

$$\frac{dS}{dt} = -S \sum_{j=1}^n \beta_j I_j, \quad (6.1a)$$

$$\frac{dI_1}{dt} = -\frac{I_1}{\Delta\tau} + S \sum_{j=1}^n \beta_j I_j, \quad (6.1b)$$

$$\frac{dI_j}{dt} = \frac{I_{j-1} - I_j}{\Delta\tau}, \quad j = 2 \dots n, \quad (6.1c)$$

$$\frac{dR}{dt} = \frac{I_n}{\Delta\tau}. \quad (6.1d)$$

By construction, an infected node will always transition through all the infectious states until it reaches the recovered state. However, we are not interested in whether infected nodes transition through all the states, but rather whether susceptible nodes become infected. In Ref. [128], the authors introduce the notion of a *next generation matrix* (NGM) which decomposes the linearized system into infectious transmissions,  $T$ , and non-infectious transitions,  $\Sigma$ , where transmissions move

susceptible nodes to infected compartments and transitions move infected nodes to other infectious states. As done in Ref. [128], we exclude the susceptible and recovered states. The linearized system can be written as

$$\mathbf{I}' = \frac{1}{\Delta\tau} \begin{pmatrix} -1 + \beta_1\Delta\tau & \beta_2\Delta\tau & \dots & \dots & \beta_n\Delta\tau \\ 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & -1 \end{pmatrix} \mathbf{I},$$

where  $\mathbf{I} = (I_1, \dots, I_n)^T$ . Splitting the matrix into transmissions and transitions according to the procedure outlined in Ref. [128], we find that the reproductive number  $R_0$  is given by  $\rho(-T\Sigma^{-1})$ , where  $\rho$  is the spectral radius, which for the fully mixed case evaluates to

$$R_0 = \sum_{i=1}^n \beta_i \Delta\tau, \quad (6.2)$$

which matches the value found in Ref. [41].

This result indicates that any infectious rate function that has the same total infectiousness or *exposure* yields the same reproductive number, regardless of the particular function. This, however, does not hold for the time scale on which the epidemic spreads as we will see later.

### 6.2.2 Discrete category-mixed population

Now we consider a population with  $N$  individuals, each of which belongs to a category  $c_i, i = 1 \dots n_c$ . These mixing categories can encode many different characteristics such as degree-based mixing [31], age-mixing [129], spatial meta-population mixing [51], mixing due to travel, and many other types of mixing.

We denote the probability that sub-populations  $c_i$  and  $c_j$  interact with each other as  $p(c_i, c_j)$  and the probability that a node belongs to category  $i$  as  $p(c_i)$ . We discretize the infectious states not only by the progression of the infection, but by the category to which that individual belongs

as well. This model has  $(n+2)n_c$  states:  $n_c$  susceptible states,  $S^{c_1}, \dots, S^{c_{n_c}}$ ;  $nn_c$  susceptible states,  $I_1^{c_1}, \dots, I_1^{c_{n_c}}, \dots, I_n^{c_1}, \dots, I_n^{c_{n_c}}$ ; and  $n_c$  recovered states,  $R^1, \dots, R^{n_c}$ . Then the mean-field model becomes for each category  $c$

$$\frac{dS^c}{dt} = -S^c \sum_{i=1}^{n_c} \sum_{j=1}^n p(c, c_i) p(c_i) \beta_j I_j^{c_i}, \quad (6.3a)$$

$$\frac{dI_1^c}{dt} = -\frac{I_1^c}{\Delta\tau} + S^c \sum_{i=1}^{n_c} \sum_{j=1}^n p(c, c_i) p(c_i) \beta_j I_j^{c_i}, \quad (6.3b)$$

$$\frac{dI_j^c}{dt} = \frac{I_{j-1}^c - I_j^c}{\Delta\tau}, \quad j = 2 \dots n, \quad (6.3c)$$

$$\frac{dR^c}{dt} = \frac{I_n^c}{\Delta\tau}. \quad (6.3d)$$

The linearized ODE is the following block-matrix system of equations:

$$\mathbf{I}' = \frac{1}{\Delta\tau} \begin{pmatrix} -I + \beta_1 \Delta\tau P & \beta_2 \Delta\tau P & \dots & \dots & \beta_n \Delta\tau P \\ I & -I & 0 & \dots & 0 \\ 0 & I & -I & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & I & -I\tau \end{pmatrix} \mathbf{I},$$

where

$$P = \begin{pmatrix} p(c_1, c_1) p(c_1) & \dots & p(c_1, c_{n_c}) p(c_{n_c}) \\ \vdots & \ddots & \vdots \\ p(c_{n_c}, c_1) p(c_1) & \dots & \beta_i p(c_{n_c}, c_{n_c}) p(c_{n_c}) \end{pmatrix},$$

$\mathbf{I} = (I_1^{c_1}, \dots, I_1^{c_{n_c}}, \dots, I_n^{c_1}, \dots, I_n^{c_{n_c}})^T$ , and  $I$  is the identity matrix.

Splitting the matrix into transmissions and transitions, the next-generation matrix (see [128] for details) is

$$-T\Sigma^{-1} = \begin{pmatrix} P \sum_{i=1}^n \beta_i \Delta\tau & P \sum_{i=2}^n \beta_i \Delta\tau & \dots & P \beta_n \Delta\tau \\ 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 \end{pmatrix}. \quad (6.4)$$



Then, the reproductive number evaluates to

$$R_0 = \rho(P) \sum_{i=1}^n \beta_i \Delta\tau, \quad (6.5)$$

which indicates that the epidemic threshold depends both on the infectious exposure and the matrix of mixing probabilities and that these two quantities are independent.

### 6.2.3 The continuum limit

For each case described prior, it is natural to take the limit as the number of infectious compartments approaches infinity and  $\Delta\tau \rightarrow 0$ . For the fully-mixed case, the reproductive number becomes

$$R_0 = \int_0^{\tau_R} \beta(\tau) d\tau, \quad (6.6)$$

and similarly, for category-based mixing, it is

$$R_0 = \rho(P) \int_0^{\tau_R} \beta(\tau) d\tau. \quad (6.7)$$

Alternatively, we can treat  $\tau$  as a continuous quantity and track the infectiousness,  $I(t, \tau)$ , as a function of the overall time and how long an individual has been infected. When  $\tau$  is continuous,  $\Delta\tau \rightarrow 0$  and the finite difference  $(I_{j-1} - I_j)/\Delta\tau$  in Eqns. (6.1c) and (6.3c) becomes a derivative with respect to  $\tau$ . With these assumptions, our ODE model can be expressed as the transport equation with boundary conditions handling the infection and recovery. For the fully-mixed case, this is

$$\frac{\partial I(t, \tau)}{\partial t} = -\frac{\partial I(t, \tau)}{\partial \tau}, \quad (6.8a)$$

$$I(t, 0) = S \int_0^{\tau_R} \beta(\tau) I(t, \tau) d\tau, \quad (6.8b)$$

$$S = 1 - \int_0^{\tau_R} I(t, \tau) d\tau - \int_0^t \frac{\partial I(t, \tau)}{\partial \tau} \Big|_{\tau=\tau_R} dt, \quad (6.8c)$$

$$I(t, \tau_R) = 0. \quad (6.8d)$$

The transport equation admits traveling wave solutions and this perspective lends physical interpretation to our model; an infected individual is transported through the infectious stages and

the boundaries merely introduce new individuals into this transport process and remove recovered individuals at the other boundary. We can see this behavior in Fig. 6.4 for both static and temporal networks.

Because our approach approximates the infectious rate function with discrete infectious compartments, we perform numerical experiments to analyze the number of states at which we can expect the mean-field ODE model to reasonably approximate the continuous rate function. For a small number of states, the discretized values of the infectious rate function fluctuate, leading to non-monotone and non-smooth trends, so we only look at the viral load model with greater than 4 infectious states. As the number of infectious states is increased, the epidemic dynamics converge to that of the continuous VL model with a continuous infectious rate function. From Fig. 6.2, approximately 100 infectious states are necessary to capture key features of the epidemic response.

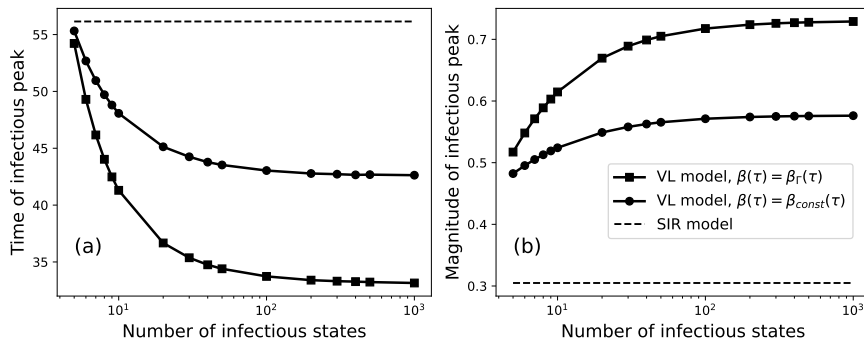


Figure 6.2: A plot showing how the number of infectious states affects (a) the time at which the infectious peak occurs and (b) the magnitude of the infectious peak for the viral load model in the fully-mixed case. We use two different infectious rate functions described in Section 6.2.6 and show the constant value of the SIR model (1 infectious stage) as a reference. For every data point,  $R_0 = 3$ .

## 6.2.4 Examples

In the following, we apply our category-mixing framework to two cases, a static degree-based configuration model and a temporal activity-based model.

### 6.2.4.1 Configuration model

Consider a network of size  $N$  with a degree sequence  $\mathbf{k} = (k_1, \dots, k_N)^T$  and nodes connected by links at random, which specifies the configuration model (described more in Ref. [130]). Networks generated with the configuration model may have a non-negligible number of self-loops and multi-edges in the infinite size limit [131], leading to correlated simple networks. In this study, however, we consider a bounded degree distribution and so we can assume the configuration model to be uncorrelated for large enough  $N$ . For the standard SIR model on a configuration model network, the reproductive number is  $R_0 = \beta \langle k^2 \rangle / (\gamma \langle k \rangle)$  [30]. We assume that a node's degree completely specifies its dynamic behavior, which ignores effects from a node's other characteristics. From the degree sequence  $\mathbf{k}$ , we can compute the discrete probability distribution  $p(k) = N(k)/N$ , where  $N(k)$  is the number of nodes in the degree sequence that have degree  $k$ , and the list of unique degrees in the degree sequence,  $\mathbf{k}_u$ . From our general formalism in Section 6.2.2, the degree mixing matrix is

$$P = \frac{1}{\langle k \rangle} (\mathbf{k}_u \mathbf{p})^T \mathbf{k}_u. \quad (6.9)$$

where  $\mathbf{k}_u \mathbf{p} = (k_1 p(1), \dots, k_{max} p(k_{max}))^T$  and  $\mathbf{k}_u = (k_1, \dots, k_{max})^T$ . The largest eigenvalue of this matrix is  $\langle k^2 \rangle / \langle k \rangle$  and so the reproductive number is

$$R_0 = \frac{\langle k^2 \rangle}{\langle k \rangle} \int_0^{\tau_R} \beta(\tau) d\tau. \quad (6.10)$$

Setting  $\gamma = 1/\tau_R$  and  $\beta = \langle \beta(\tau) \rangle = \int_0^{\tau_R} \beta(\tau) d\tau / \tau_R$  for the SIR model yields the reproductive numbers derived in Ref. [30].

### 6.2.4.2 Activity model

Our category-based framework applies not only to static contact structures, but to temporal networks as well. We consider the *activity model* first presented in Ref. [132]. Given a temporal network of size  $N$ , suppose that each node  $i$  has an *activity rate*  $a_i$ , which denotes the probability per unit time that the node is active. At each discrete time, each node is either active or idle,

and each active node forms  $m$  connections with other nodes, active or inactive. Unlike degrees which are discrete for an unweighted network, these activity rates are continuous, and to use our category-based mixing framework, we assume that we can bin these rates into discrete categories,  $a_i, i = 1 \dots n_a$  and later take the continuum limit as before. We denote the probability that a node has an activity rate  $a_i$  as  $p(a_i)$ . Then the probability that nodes with activity rates  $a_i$  and  $a_j$  are connected at any given time is  $(a_i + a_j) \frac{m}{N}$  and the time-averaged mixing matrix is

$$P_{ij} = \frac{m(a_i + a_j)}{N} p(a_j),$$

which can be written  $P = \mathbf{1b}^T + \mathbf{c}\mathbf{p}^T$  where

$$\mathbf{b} = (m a_1 p(a_1), \dots, m a_{n_a} p(a_{n_a}))^T,$$

$$\mathbf{c} = (m a_1, \dots, m a_{n_a})^T,$$

$$\mathbf{p} = (p(a_1), \dots, p(a_{n_a}))^T.$$

Observing that this is a rank-2 matrix, the analytical solution for the Perron-Frobenius eigenvalue is  $(m\langle a \rangle + m\sqrt{\langle a^2 \rangle})$  and

$$R_0 = (m\langle a \rangle + m\sqrt{\langle a^2 \rangle}) \int_0^{\tau_R} \beta(\tau) d\tau. \quad (6.11)$$

In Ref. [132], they derive the epidemic threshold for the activity model as  $\beta/\gamma = 2\langle a \rangle / (\langle a \rangle + \sqrt{\langle a^2 \rangle})$ . As before, setting  $\gamma = 1/\tau_R$  and  $\beta = \langle k \rangle \langle \beta(\tau) \rangle = 2m\langle a \rangle \langle \beta(\tau) \rangle$  yields the same result.

### 6.2.5 Individual variation in the infectious rate function

In Ref. [133], the authors consider heterogeneous susceptibility and recovery rate for the SIR model. Similarly, we now relax the assumption that the infectious rate function is the same for every individual. We extend our results in Section 6.2.2 for a distribution of infectious rate functions over the population. In our analysis, we assume that the particular infectious rate function is distributed independently of any other nodal characteristic such as its degree. We denote  $p_b(b)$  as the fraction of the population with an infectious rate function of  $\beta_b(\tau)$  and an associated recovery time of  $\tau_{R_b}$ , where the number of unique infectious rate functions is  $n_b$ . We enforce that the number of infectious

states regardless of recovery time is  $n$  so the time between infectious compartments is  $n\Delta\tau_b = \tau_{R_b}$ .

We define the discretized values  $\beta_i(\tau_j) = \beta_i(j\Delta\tau_i)$  as  $\beta_j^i$  and denote the  $j$ th infectious stage with infectious rate function  $\beta_b(\tau)$  and category  $c$  as  $I_j^{b,c}$ . Then the mean-field equations become

$$\frac{dS^{b,c}}{dt} = -S^{b,c} \sum_{i=1}^{n_b} \sum_{j=1}^{n_c} \sum_{k=1}^n p_b(b_i) \beta_k^{b_i} p(c, c_j) p(c_j) I_k^{b_i, c_j}, \quad (6.12a)$$

$$\begin{aligned} \frac{dI_1^{b,c}}{dt} &= -\frac{I_1^{b,c}}{\Delta\tau} \\ &+ S^{b,c} \sum_{i=1}^{n_b} \sum_{j=1}^{n_c} \sum_{k=1}^n p_b(b_i) \beta_k^{b_i} p(c, c_j) p(c_j) I_k^{b_i, c_j}, \end{aligned} \quad (6.12b)$$

$$\frac{dI_j^{b,c}}{dt} = \frac{I_{j-1}^{b,c} - I_j^{b,c}}{\Delta\tau}, \quad j = 2 \dots n, \quad (6.12c)$$

$$\frac{dR^{b,c}}{dt} = \frac{I_n^{b,c}}{\Delta\tau}. \quad (6.12d)$$

Linearizing these equations, we obtain  $\mathbf{I}' = \mathbf{A}\mathbf{I}$ , where  $A = \Sigma + T$ .  $\Sigma$  and  $T$  are each  $n \times n$  block matrices of size  $n_c n_b \times n_c n_b$  with blocks of size  $n_c \times n_c$ .

$$\Sigma_{i,j} = \begin{cases} \text{diag}(I/\Delta\tau_1, \dots, I\Delta\tau_{n_b}), & i = j \\ \text{diag}(-I/\Delta\tau_1, \dots, -I\Delta\tau_{n_b}), & i = j + 1 \end{cases}$$

and

$$T_{i,j} = \begin{cases} \begin{pmatrix} p_b(b_1) \beta_j^1 P & \dots & p(b_{n_c}) \beta_j^{n_b} P \\ \vdots & \ddots & \vdots \\ p_b(b_1) \beta_j^1 P & \dots & p(b_{n_b}) \beta_j^{n_b} P \end{pmatrix}, & i = 1 \\ \mathbf{0}, & \text{else.} \end{cases}$$

Then, the reproductive number (the maximal eigenvalue of  $-T\Sigma^{-1}$ ) is

$$R_0 = \rho(P) \sum_b p_b(b) \sum_{j=1}^n \beta_j^b \Delta\tau_b.$$

As  $n \rightarrow \infty$ ,  $\Delta\tau_b \rightarrow 0$  for every  $b$  and we obtain

$$R_0 = \rho(P) \sum_b p_b(b) \int_0^{\tau_{R_b}} \beta_b(\tau) d\tau, \quad (6.13)$$

which is the value obtained for the category-mixed case with the key difference that the *exposure* is now the average exposure with respect to the distribution of infectious rate functions.

### 6.2.6 Numerical experiments

We compare the time dynamics of the SIR model with that of the VL model with different infectious rate functions. For the following figures, we fixed  $N = 10^4$ ,  $R_0 = 3$ ,  $\tau_R = 21$  days, and  $\arg \max_{\tau} \beta(\tau) = 4$  days, unless otherwise noted. We considered the configuration model with a power-law degree distribution  $p(k) \propto k^{-3}$  on  $[10, 1000]$  and the activity model with activity rates  $p(a) \propto k^{-3}$  on  $[0.01, 1]$ ,  $m = 10$ , and  $\Delta t = 1$ . We used the following contagion models: the VL model with  $\beta_{\Gamma}(\tau) \propto \tau \exp(-\tau/4)$  as in Ref. [40], the VL model with a constant-valued infectious rate function,  $\beta_{const}(\tau) = \langle \beta_{\Gamma}(\tau) \rangle$ , and the SIR model with a single infectious rate of  $\beta = \langle \beta_{\Gamma}(\tau) \rangle$  for the configuration model and  $\beta = 2m\langle a \rangle \langle \beta_{\Gamma}(\tau) \rangle$  for the activity model. These relations were chosen such that the reproductive numbers are the same for each infection model.

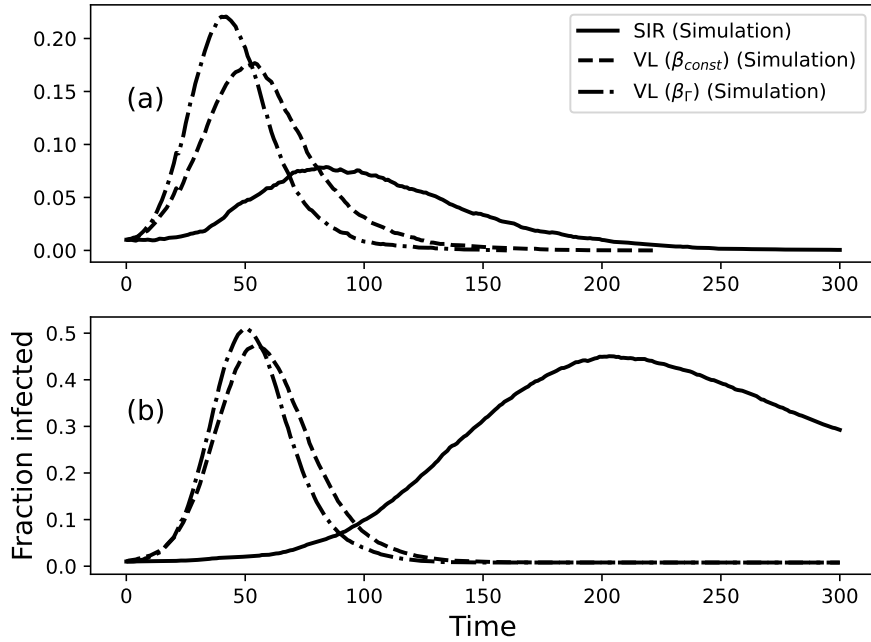


Figure 6.3: Time response of the fraction of infected individuals for different contagion models for (a) the configuration model and (b) the activity model. For both (a) and (b), the dash-dot, dashed, and solid lines indicate the VL model with  $\beta_{\Gamma}(\tau) \propto \tau \exp(-\tau/4)$ , the VL model with  $\beta_{const}(\tau) = \langle \beta_{\Gamma}(\tau) \rangle$ , and the SIR model with a single infection rate of  $\beta$  respectively.  $\beta = \langle \beta_{\Gamma}(\tau) \rangle$  and  $\beta = 2m\langle a \rangle \langle \beta_{\Gamma}(\tau) \rangle$  for the configuration and activity models respectively.  $R_0 = 3$  for each infection curve.

We simulated all the contagion models described in discrete time with  $\Delta t = \Delta \tau = 1$ . We simulated the SIR model as a discrete time Markov process using the parameters  $\gamma = 1/\tau_R$  and  $\beta = \langle \beta_\Gamma(\tau) \rangle$  and  $\beta = 2m\langle a \rangle \langle \beta_\Gamma(\tau) \rangle$  for the configuration and activity models respectively. For the viral load model, we store the time at which node  $i$  has been infected as  $t_i^*$  and at time  $t$ , the rate of infection of that node is  $\beta(t - t_i^*)$ , for example, and when  $t - t_i^* \geq \tau_R$ , the node recovers. When simulating on temporal networks, we store the temporal network as an array, where each entry is a static network corresponding to a particular snapshot in time.

From Fig. 6.3 we see that the peak of the SIR model is delayed relative to both viral load models. For the static case the epidemic peak is significantly less pronounced. We comment that the viral load model fundamentally changes the time scale of the epidemic when compared to the SIR model. Not all infectious compartments are created equal, however; someone at their peak infectiousness contributes much more to the spread of an epidemic than someone who has just gotten infected or almost recovered. For this reason, we now plot the number of individuals in each infectious stage over time. We now relax the assumption that  $\beta_\Gamma(\tau)$  and  $\tau_R$  are identical for each member of the population. We assume that  $\arg \max_\tau \beta(\tau) \sim \text{Uniform}(2, 6)$  and that  $\tau_R \sim \text{Uniform}(16, 26)$  similar to Ref. [123]. At given times  $t$ , we plot the number of individuals as a function of the infectious duration  $\tau_i = t - t_i^*$  and  $t$ .

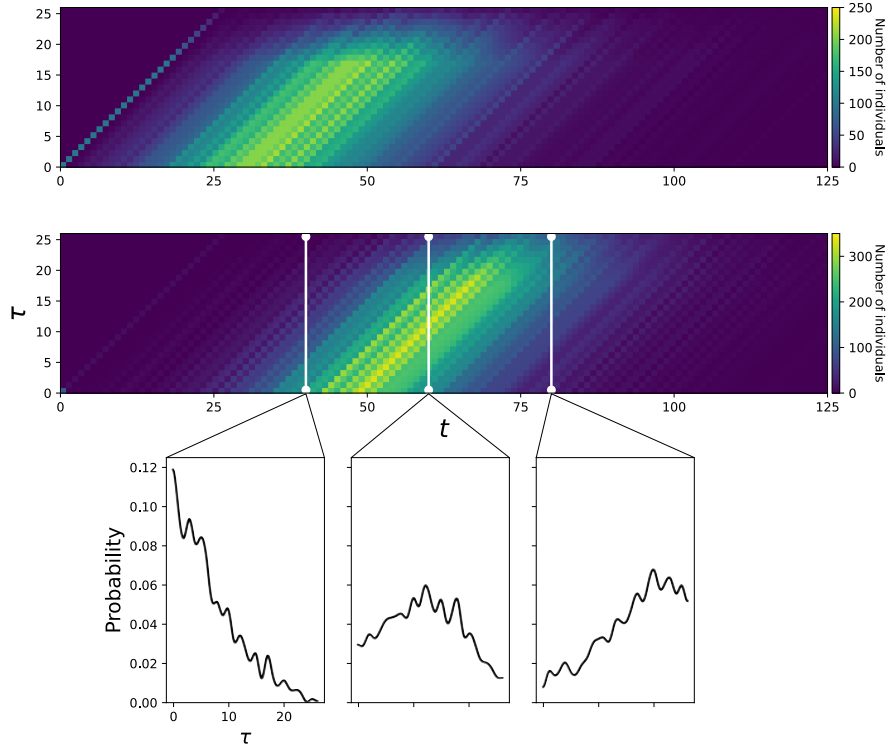


Figure 6.4: The number of individuals infected for duration  $\tau$  at time  $t$  for the configuration model (top) and the activity model (middle). The line plots (bottom) denote the probability distribution of  $\tau$  at times 40, 60, and 80, which correspond to normalized vertical cross-sections of  $I(t, \tau)$ .

We see traveling wave behavior as described in Section 6.2.3 for both static and temporal networks. The amplitude of this wave varies in response to the introduction of new infected individuals, but the distribution shows a clear transition to the latter infectious stages as the epidemic progresses. This behavior is corroborated by the three normalized vertical cross-sections, showing the probability distribution at selected times. We notice that, despite identical values of  $\beta_i(\tau)$  and  $\tau_{R_i}$  for every node, the temporal behavior is different for static and temporal networks. For the temporal network, it seems evident that individuals with the longest infection duration seem to be driving the epidemic based on the minimal decrease in individuals for large  $\tau$  in comparison to the static network case.

We also plot the epidemic extent for different values of  $R_0$  in Fig. 6.5 to validate our predictions of the reproductive number. For each data point, we averaged over 100 simulations but use



the same network realization for all simulations for both the configuration and activity models. We ran each simulation until there were no longer any infected individuals.

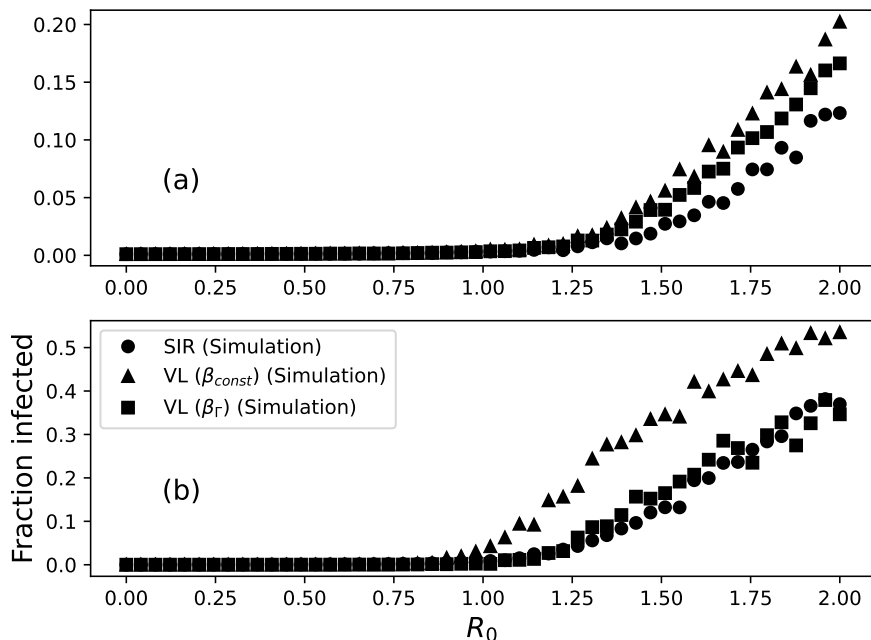


Figure 6.5: The epidemic extent plotted as a function of the predicted reproductive number for different contagion models for (a) the configuration model and (b) the activity model.

We see that for both static and temporal networks, the predictions from our theory do as well as the predictions for the SIR model in Refs. [30] and [132]. The gradual transition is due to the heterogeneity of the networks and agrees with prior results on power-law networks [133].

### 6.3 Discussion

In our analysis, we theoretically derived and numerically validated predictions of the population reproductive number for static and temporal networks for a contagion model accounting for time-dependent infectiousness. We see that time-dependent infectiousness causes a fundamental change in the time dynamics compared to the dynamics of the SIR model, despite an epidemic threshold matching classical theory.

Although time-dependent infectiousness does not affect predictions on whether an epidemic will initially grow or die out, it has strong implications on how the epidemic progresses in time. In the continuum limit, the viral load model can be written as the transport equation PDE with an infectious boundary condition, which indicates that the distribution of infectious durations changes in time similar to the propagation of a traveling wave. This prediction is validated with numerical simulations.

In this study, we have only considered the population reproductive number, although it is well known that merely studying the population reproductive number without examining the heterogeneity in the number of secondary infections leaves out key information [125]. Superspreading events are the result of this stochasticity and can often be responsible for the transmission of an epidemic. The VL framework could be used to model the distribution of secondary infections resulting from a combination of contact-based and infectiousness-based heterogeneity.

## Data Availability

All code used in this study can be found at <https://github.com/nwlandry/time-dependent-infectiousness> [134].

## Chapter 7

### On limitations of uniplex networks for modeling multiplex contagion

Studies of contagion on networks regularly rely on data that only represent one type of relationship at a time. These sorts of data can be helpful for understanding contagion dynamics on relationships of that type. However, when the underlying contagion processes are multiplex, they are less well suited for understanding the ultimate extent over a population and the timescale over which these processes occur.

As an example, consider research aiming to model an epidemic of a particular sexually transmitted infection such as HIV. HIV is transmissible via sexual contact, shared needles, and other exposures to blood or other bodily fluids. In network terms, this means that any spread of HIV through the population requires an unbroken chain of susceptible cases exposed to those who are currently infectious [135]. That chain of exposure could be contained within relationships of a particular type (e.g., could be completely composed of sexual contacts), or could combine relationships of multiple types (i.e., sexual and needle-sharing contacts).

In many populations, members have active contacts of both types, and they are not directly overlapping — for example, not all needle-sharing contacts are with the same partners involving sexual contacts [136]. In that case, studying only one type of relationship at a time could substantially distort any predictions of epidemic extent within the population. As a simple example, suppose person 1 in Figure 7.1 is currently infected, and we want to know the likelihood of person 4 becoming infected. If we examine only sexual relationships (denoted by solid lines), both 1 and 4 are sexually active with a single partner (2 and 3, respectively) and those direct relationships would

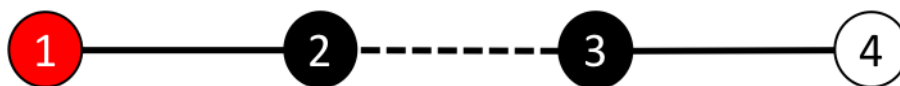


Figure 7.1: A simple multiplex network. The line style corresponds with the relationship type. The node color represents its current status: infected (red), susceptible (white), and unknown (black).

be accurately represented with data on the sexual network along with any models using those data. However, the potential infection-carrying relationship between 2 and 3 providing an indirect route of transmission between 1 and 4 would not be observed, because it stems from a needle-sharing relationship (dashed line). With only data on sexual relationships, we would be underestimating 1's likelihood of *indirectly* infecting 4, and even misunderstanding how these nodes' sexual partnerships contribute to that possible route of transmission. We see that even in this simple example, only considering one tie type can lead to very different predictions of the likelihood of infection and the resulting epidemic extent.

Similarly, there has been a recent proliferation of studies examining propagation of memes, ideas, even voting behavior across social media sites like Facebook or Twitter [137, 138, 139]. However, while people mix their social media usage and have non-overlapping connectivity patterns across platforms [140], these studies overwhelmingly focus on modeling contagion over a single platform at a time. It is reasonable to expect this approach could potentially give a limited picture of the complete contagion potential across the population, as in the sex/needle example for STIs above.<sup>1</sup> For example, people may primarily use one platform with their friends and family (ties between  $\{1, 2\}$  and  $\{3, 4\}$  in Figure 7.1), while using another with their professional contacts (tie  $\{2, 3\}$  in Figure 7.1). In that case, studies that rely on detailed data from a single platform at a time would miss important cross-platform linkages that are vital to understanding the contagion dynamics across the population of interest.

In this Chapter, we demonstrate the potential limitations of inference about population contagion potential and timescale from networks of a single relationship type (uniplex networks), when

<sup>1</sup> Moreover, simply increasing the coverage of the sample used will not overcome these potential sampling biases [141].

contagion processes realistically involve spreading over relationships of multiple types (multiplex networks).

We proceed as follows: first, we motivate the importance of these limitations on contagion modeling; second, we discuss the data sets and contagion processes considered; third, we discuss our approach in quantifying discrepancies in the epidemic response for different choices of data; fourth, we present numerical results validating our premise; and lastly, we discuss limitations and interpretations of our model.

## 7.1 Motivation

Our central aim is to show that modeling capacity in contagion research is strongly constrained by choices in the data collection and representation. Many studies consider a single risk behavior or platform for potential idea spread. This choice can result from simple limitations of data availability, or may draw on the recognition that domain expertise often improves our capacity to collect high quality data on particular topics (e.g., drug sharing networks have considerable barriers to data elicitation [142]). It is important to understand the limitations introduced by these choices when predicting the extent to which behaviors or diseases will spread. We draw on two simple case studies to explore the discrepancies between the contagion extent resulting from spreading over a single relationship type compared to a model with two different relationship types.

Many studies have presented theoretical approaches for describing contagion spread on multiplex networks [143, 144, 145, 146]. These studies have been crucial in understanding of dynamics on multiplex networks, but take a fundamentally different approach to ours. This study focuses not on predicting the behavior of a multiplex network, but rather exploring the effect of the underlying data representation on the limitations that arise when modeling contagion processes from data representing a single relationship type. This is similar to the approach in Ref. [147] where the authors examine how well networks constructed from different data types effectively capture transmission potential. We show that one cannot generate accurate expectations of epidemic extent from uniplex data, even if combining results across layers to reconstruct the multiplex reality, as

an estimate of their potential combined effects.

## 7.2 Data

We use two data sets to represent multiplex networks. For each, we examine (1) the entire multiplex network and (2) the multiplex network constructed with nodes that are members of the largest connected component of the network when considering both relationship types.

The first data set comes from “Project 90” in Colorado Springs, which studied a network of commercial sex workers, people who inject drugs, and their sex and needle-sharing partners between 1988-1992 [148]. Here we use the complete network of 7,677 individuals representing all respondents and the ‘risk behavior’ partners they nominated as well as the largest connected component of 4,385 individuals. Although these data were collected over five different waves, we aggregated them into a single network for our purposes. The other data set represents a multiplex network of on-line social network site interactions among a sample of 1,672 Twitter and Foursquare users [149] hereafter, ‘JOAAP’ as well as the largest connected component of 1,564 individuals. <sup>2</sup>

In the Project 90 data, we considered sexual ties as one type of tie and drug or needle-sharing ties as a second tie type. When we refer to the *multiplexed data*, we aggregate these uniplex layers by forming a link between two individuals if they have a sexual relationship or a drug or needle-sharing relationship, whereas the uniplex networks are formed from each tie type separately.

The JOAPP data, which are much higher density, representing Foursquare co-check-ins (frequenting the same locales) as one type of tie and “follower” links on Twitter as a second type of tie. We symmetrized the Twitter data for consistent analysis, even though these data are inherently directional.

Analyses begin with the observed multiplex Project 90 and JOAAP data sets [151, 148]. We then decompose each of these multiplex networks into their respective uniplex networks with each tie-type representing a unique network. For each data set, this produces three networks. For the

---

<sup>2</sup> These data sets are available for download from <https://doi.org/10.3886/ICPSR22140.v1> [150] and <https://doi.org/10.6084/m9.figshare.4585270.v1> [151], respectively.

Data set	N	average degree
<b>Full data set</b>		
1. Project 90	7677	1.70
2a. Project 90 - Sex	7677	0.85
2b. Project 90 - Drugs	7677	1.17
3. JOAPP	1672	45.7
4a. JOAPP - Foursquare	1672	35.22
4b. JOAPP - Twitter	1672	16.85
<b>Largest component of multiplexed data</b>		
1. Project 90	4375	2.46
2a. Project 90 - Sex	4375	1.22
2b. Project 90 - Drugs	4375	1.71
3. JOAPP	1564	36.21
4a. JOAPP - Foursquare	1564	24.67
4b. JOAPP - Twitter	1564	18.02

Table 7.1: Statistics of the chosen networks

Project 90 data set, this produces (1) the multiplex composite risk behavior network, (2a) the sexual contact uniplex network, and (2b) the needle-sharing uniplex network. For the JOAAP data set, this produces (3) the multiplex composite online network, (4a) the Foursquare uniplex network, and (4b) the Twitter uniplex network. All analyses described below are then conducted on each of these networks separately, and the results compared between the multiplex networks (1 and 3) and the union of the uniplex layers (2a and 2b, and 4a and 4b, respectively).

For each data set, we extract the list of node labels by tabulating all the nodes that are connected to either link type. For each type of network (uniplex or multiplex), we start with this node list and then add the links corresponding to the desired tie types. Although in principle one could weight the links according to the frequency of their occurrence, for simplicity, we consider unweighted uniplex and multiplex networks, even if two ties of different types connect the same node pairs, accounting for why the average degree of the two uniplex networks do not sum to the average degree of the multiplexed network.

In our study, we use the Project 90 data as a model of a behavioral network over which epidemics can spread and the JOAPP data as a model of online networks for behavior adoption. Table 7.1 presents descriptive statistics for each data set.

### 7.3 Contagion processes

In this study, we considered two models of contagion: the susceptible – infected (SI) model representing an epidemiological simple contagion model and the threshold model representing a complex contagion process, which represent two common model forms of contagion processes.

For the SI model, we define two states: susceptible (S) and infected (I). A susceptible node can be infected by one of its infected neighbors with probability  $\beta\Delta t$ , where  $\beta$  is the infection rate and  $\Delta t$  is the interval at which the node states are updated [152]. In this study,  $\Delta t = 1$  day and we update the nodes at the next time synchronously. Once a susceptible node is infected, it will remain infected. Thus, for a connected network, given enough time, every node will eventually become infected.

The threshold model (also known as “complex contagion”) is a common model for behavior adoption on networks [22, 24]. For the threshold model, we define, as above, two states: non-adopting (S) and adopting (I). We fix an adoption threshold  $\tau$  between 0 and 1. A non-adopting node adopts the opinion if the fraction of its neighbors who have already adopted is larger than  $\tau$ . Once a node has adopted it will not change its state to become non-adopting. We update the opinions of all the nodes at the next state synchronously; the threshold model is deterministic once the initial state is specified. Unlike the SI model, a contagion spreading via the threshold process may not reach the entirety of a network, even if it is connected [153].

For both models, we select seed nodes uniformly at random to infect and this defines the initial state of our system. We run these contagion processes for a sufficient duration to ensure the epidemic extents have reached equilibrium.

### 7.4 Approach

We demonstrate how network data representation choices alter the epidemic extent and rate of contagion spread. The data sets described above are inherently multiplex in nature and serve as case studies in answering our central question.



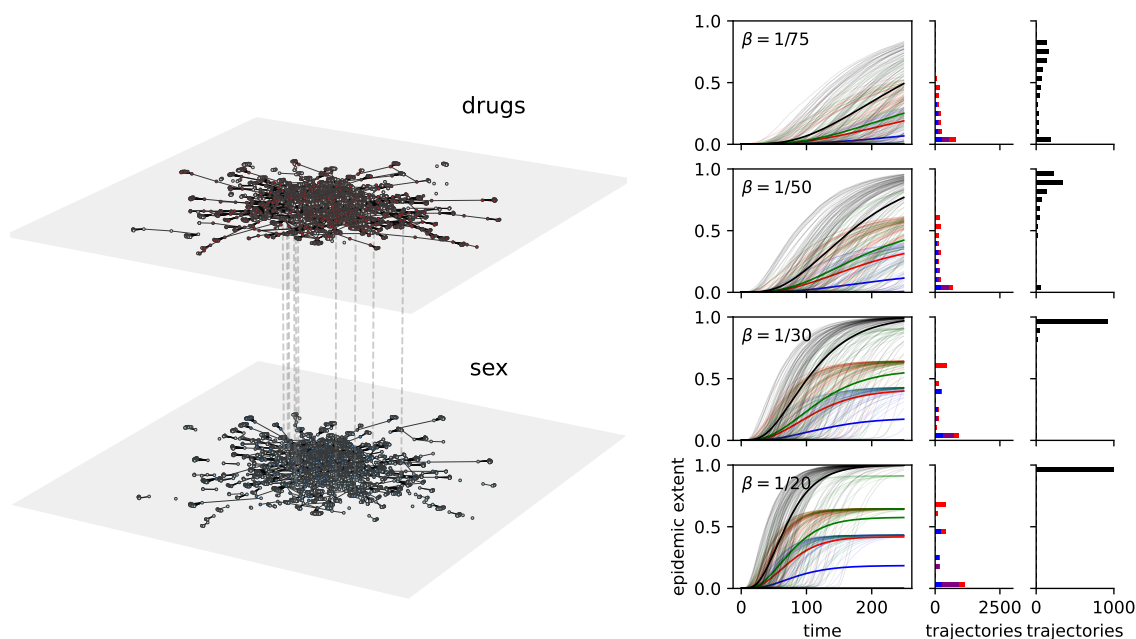
We start by constructing networks from the data sets as described above. For each data set, we produce the observed multiplex network containing both tie types and two constituent uniplex networks from each dataset.

For each simulation run, we fix a single seed node for each of the four settings of interest. We begin by running these respective contagion processes (the SI model on the Project 90 dataset and the threshold model on the JOAPP dataset) for the following: (1) the respective multiplex network data, (2) and (3) each of the constituent uniplex layers decomposed from those multiplex networks (as described above), and (4) the union of the infected nodes in each layer of (2) and (3) for each time step in the simulated contagion processes. We present results for the proportion of infected network members at each time step and final epidemic extent. For each setting combination, we generate many realizations of these simulations to form an ensemble of time series.

## 7.5 Results

For all contagion processes and each parameter value, we set the time step to a week, i.e.,  $\Delta t = 1$  and infected a single node at random initially (although we present additional results in the Supplement using different numbers of seed nodes). We ran 1000 simulations to form an ensemble for robustness.

For the SI process, we used the following infection rates (which, because  $\Delta t = 1$ , are also infection probabilities):  $\beta = 1/75, 1/50, 1/30, 1/20$  (simulations for wider parameter ranges in the Supplement provide a robustness check). We simulated up until a time of  $t_{max} = 250$  (i.e., 5 years; although we present longer simulations in the Supplement as a robustness check). For the threshold process, we used the following adoption thresholds:  $\tau = 1/8, 1/10, 1/12, 1/15$  (simulations for wider parameter ranges in the Supplement provide a robustness check). We simulated until  $t_{max} = 30$  (although we present longer simulations in the Supplement as a robustness check).



(a) Network visualization of the Project 90 data set. The dashed lines are not exhaustive, but illustrate that the nodes in each layer are identical. The lightly colored nodes in each layer are disconnected from the largest connected component in that layer. One can see that each layer in this multiplex network has different contact structure.

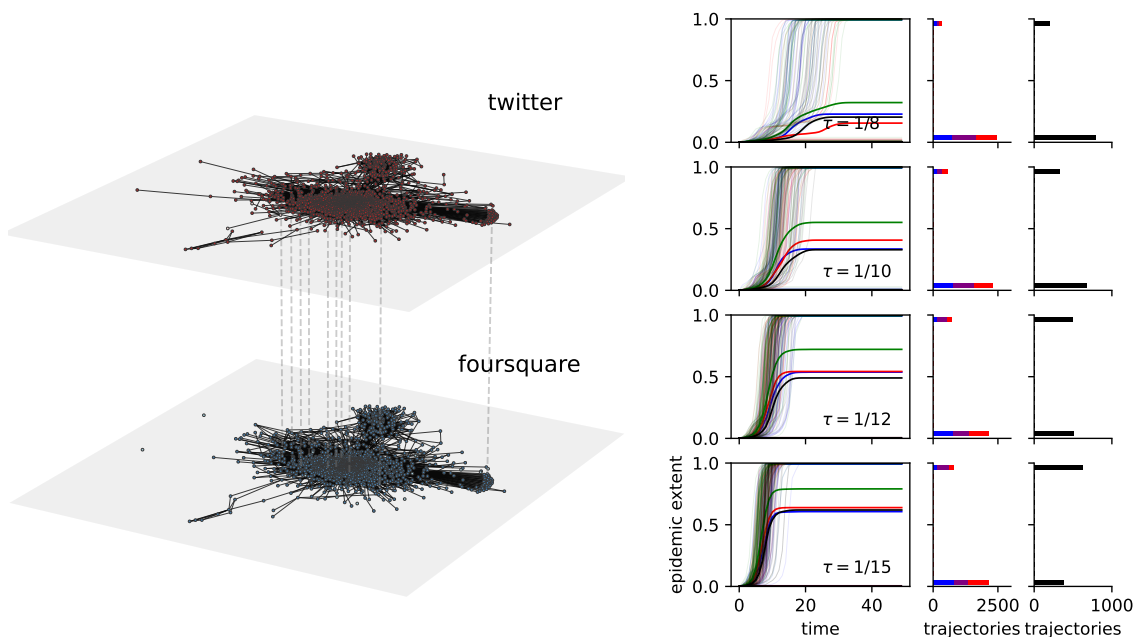
(b) The SI model simulated on the Project 90 data. The left column displays a plot of the epidemic extent with respect to time for the uniplex sex network (blue), uniplex drug and needle network (red), the union of uniplex networks (green), and the multiplexed data (black). The thick lines indicate the average epidemic extent and the thin lines are individual realizations of this model. The center column illustrates the histogram of epidemic extents for the number of nodes solely accessed from the sex network (blue), the drug network (red), and both (magenta). The right column illustrates the distribution of epidemic extents for the multiplexed data.

Figure 7.2: Comparing the epidemic extent for different choices of network data for the Project 90 data set.

To examine our central question, we draw the reader’s attention to comparisons between the black and green curves in Figs. 7.2b and 7.3b. For the SI model, Fig. 7.2b shows that the multiplex data leads to a larger epidemic extent than when we consider the union of the separate uniplex processes. This result demonstrates that to capture the “true” potential of the contagion process on the multiplex network, it is not enough to simply combine the results from their constituent uniplex

layers. As shown in Fig. 7.1, simply considering relationships of one type ignores how contagion can rely on the complementarities across relationships of different types. These uniplex networks, when combined into the multiplex nature of reality, unlock connection patterns that reach more of the population than can be accessed by the combination of the two independently. This indicates, as illustrated in Fig. 7.1, that there is “leap-frog” behavior occurring, where a contagion must pass through connections of alternating types to reach certain nodes.

Looking at the histogram of final extents, we see that for the multiplexed data, by construction, the contagion will spread to the entire population given enough time. The uniplex data sets do not reach the entirety of the population indicating that the uniplex networks have components that are smaller than those of the full multiplex networks.



(a) Network visualization of the JOAPP data set. The dashed lines are not exhaustive, but illustrate that the nodes in each layer are identical. The light colored nodes in each layer are disconnected from the largest connected component in that layer.

(b) The threshold model on the JOAPP data. As in Fig. 7.3b the left column displays a plot of the epidemic extent with respect to time for the uniplex Twitter network (blue), uniplex Foursquare network (red), the union of uniplex networks (green), and the multiplexed data (black). The thick lines indicate the average epidemic extent and the thin lines are individual realizations of this model. The center column illustrates the histogram of epidemic extents for the number of nodes solely accessed from the Twitter network (blue), the Foursquare network (red), and both (magenta). The right column illustrates the distribution of epidemic extents for the multiplexed data.

Figure 7.3: Comparing the epidemic extent for different choices of network data for the JOAPP data set.

For the threshold model, the situation is a bit less straightforward. In particular, here the results are more bimodal in that the ultimate contagion extent typically either (a) stalls before taking off at all or (b) reaches the vast majority of the population. This result reflects that both the connectedness and the density of the network determine the epidemic extent (see for example, Ref. [153]). This result again demonstrates that understanding contagion processes on

uniplex data representation, even when combining them, does not straightforwardly translate to account for contagion on the related multiplex network. Indeed, in Fig. 7.3b, we see that for each parameter value the epidemic extent for the multiplexed data may be lower than that of both the uniplex network ( $\tau = 1/10, 1/12$ ), larger than that of the Twitter uniplex network but smaller than that of the Foursquare uniplex network ( $\tau = 1/15$ ), and larger than that of the Foursquare uniplex network but smaller than that of the Twitter uniplex network ( $\tau = 1/15$ ). This is likely a complicated interplay between the connectedness and density of these networks. On the one hand, the uniplex networks are more disconnected than the multiplexed network. On the other hand, the multiplexed network is denser than each of the uniplex networks, which discourages the spread of contagion for a given threshold value. If we set  $\tau = 0$ , this removes the first of these two factors and the relations between the epidemic extents becomes the same as for the SI model. By definition, the epidemic extent for the union of these is larger and we see that considering both types of relationships separately may be a *worse* estimate than simply considering a single uniplex network (for example  $\tau = 1/10, 1/12$  in Fig. 7.3b).

## 7.6 Discussion

In this Chapter, we have drawn on two cases to show how modeling inherently multiplex contagion processes with uniplex network representations can misrepresent the predicted fraction of the population that a contagion process will reach. We illustrate this limitation separately for models of epidemiological and behavioral contagion to highlight the effect of the data representation on the resulting epidemic behavior. For simplicity, within each model we used the same transmission parameters for each link type. Contagion on empirical networks may have different rates of spread depending on the type of link through which contagion is transmitted (e.g., needle sharing transmits most STIs more efficiently than sexual contacts). In addition, contagion spread can be modeled by different processes spreading on each layer. For example, one might model the spread of awareness as a behavioral process on one network layer and the spread of disease on another network layer [154].

In the case of the SI model, the multiplex epidemic extent was consistently larger than that of the union of the two constituent uniplex layers. The magnitude of this discrepancy depends on the way that the connected components in each uniplex layer overlap. The case where each uniplex network has the same connected components will lead to the same epidemic extent. This is a trivial case where one link type may be a proxy for another link type, which may limit the usefulness of a multiplex representation, although is likely to be uncommon empirically. For the threshold process, gaps remain, but differ in their nature. Whether the union of the constituent uniplex data underestimates or overestimates the true contagion extent depends upon the threshold levels, and these results depend on combinations of network structural characteristics in ways that are important to examine further.

It may be fruitful to estimate how the magnitude of the under or overestimates seen here differ when examining datasets with different structural features. Along the same lines, reproducing these results on synthetic datasets and predicting them may offer additional insight to the behavior that we observe.

Our results offer evidence supporting our premise that uniplex data is inadequate for modeling inherently multiplex processes. We offer this study as a cautionary tale for researchers modeling the spread of contagion: the choice of network data is an important assumption baked into contagion models and should be carefully considered. As a more positive recommendation, a primary takeaway of our results is that future data collection efforts should prioritize faithfully capturing the multiplex realities of the underlying processes intended to be examined as in Ref. [147], rather than relying solely on data of a single type.

## Chapter 8

### Open-source software

In this Chapter, we discuss software that has been developed to support the projects in Chapters 3 to 5.

#### 8.1 CompleX Group Interactions (XGI)

The accessibility and inclusivity of a scientific community rests, in part, on its commitment to share resources. For the pairwise network community, packages such as NetworkX [155], graph-tool [156], and igraph have provided a common language and tools allowing cross-disciplinary researchers to more easily collaborate and contribute to the field of network science.

The sub-field of higher-order network science has grown rapidly over the past few years, garnering wide attention for its ability to offer additional insights for machine learning, the structure of complex systems, the dynamics of complex systems, and many other applications. This topic is relevant to many different fields such as computer science, infectious diseases, dynamical systems, behavioral science, and many others. Open-source code is available for hypergraph dynamics [11, 13, 12], inference [54, 113, 157], generative models [3, 12, 158], algorithms [159, 73], and data [75] written in languages such as Python, Julia, C++, and Matlab, but these resources are relatively piecemeal. We have developed the CompleX Group Interactions (XGI) Python library to provide an open-source resource for this community [160].

This library alleviates many difficulties in hypergraph research. It does so by providing an efficient data structure for hypergraphs, standard algorithms, conversions between data represen-

tations, and functions to read and write from different file formats.

### 8.1.1 Data structure

The data structure employed by XGI leverages the bipartite representation of a hypergraph [4] which allows users to efficiently query the hyperedges of which a specified node is a part and the nodes that are members of a specified hyperedge. Each hyperedge is assigned an ID which allows for multi-hyperedges. In addition, the members of a hyperedge and the hyperedges to which a node belongs are stored as multisets which allows for loopy hyperedges.

In addition, the core data structure provided by XGI offers methods for easily getting the following:

- The number of nodes
- The number of hyperedges
- The degree sequence
- The hyperedge size sequence
- The singleton edges
- Subhypergraphs
- Attributes of nodes and hyperedges (Such as weights, names, etc.)

### 8.1.2 Conversion between hypergraph representations

A hypergraph may be represented in a variety of ways including an incidence matrix, a bipartite network, a list of hyperedges, and many others [4]. Different applications require different hypergraph representations for efficient computation. For example, when modeling contagion, a node's infection status may change dependent on the statuses of its neighbors, indicating that a representation allowing efficient access to the node's edge neighbors is desirable. Likewise, when



one is interested in computing properties of a hypergraph that is averaged over the hyperedges such as assortativity or modularity, it may be most efficient to represent a hypergraph by a list of its hyperedges.

XGI provides methods for users to convert between a hypergraph and the following formats:

- A list of hyperedges
- An adjacency list
- An incidence matrix
- A bipartite network

### 8.1.3 Algorithms

There are two types of algorithms that XGI supports: generative random models and methods for analyzing the connectedness of a hypergraph. Generative models are important for the generation of synthetic datasets, which are useful as null models to compare empirical datasets, create datasets with similar characteristics as another dataset, or control structural properties such as assortative mixing or degree heterogeneity. XGI currently implements several different generative models:

- The Chung-Lu model for hypergraphs
- The Degree-Corrected Stochastic Block Model (DCSBM) model for hypergraphs
- The Erdős-Rényi model for hypergraphs
- The configuration model for uniform hypergraphs

Measuring connectedness in hypergraphs is important as a data exploration tool because many theoretical frameworks assume that a hypergraph is connected and this may not be the case. XGI currently provides the following functionality for users:

- Determining whether or not a hypergraph is connected

- Finding the connected components of a hypergraph
- Finding the component of which a specified node is a part

#### 8.1.4 Read/write functionality

There are many sources of hypergraph data [75, 161, 162], often in very different formats which is a significant overhead cost for researchers. XGI alleviates this cost in two ways: first, by implementing methods for importing and writing hypergraphs from several different formats and second, by implementing a standard for hypergraph data in JSON format, described in more detail in Section 8.3.

The XGI library offers 4 main types of file input and output:

- A list of hyperedges, where each line of the file is a hyperedge
- A list of bipartite edges, where each line is a (node, edge) entry
- An incidence matrix, where each column corresponds to a particular hyperedge, and the non-zero entries correspond to the nodes that participate in that hyperedge.
- A JSON file, the structure of which is described in more detail in Section 8.3.

## 8.2 HyperContagion

Just as there are limited open-source libraries for the representation and analysis of higher-order interaction networks, similarly, there are no standardized libraries for the simulation of contagion processes on higher-order interaction networks. There are specialized repositories for the hypergraph SIS model both discrete [12, 11] and continuous [163], hypergraph bounded confidence models [35], and others. All of these models, however, rely on different data structures and implementation.

HyperContagion is a library for the efficient and standardized simulation of hypergraph contagion models. Currently the package implements two basic classes of models: epidemiological

models and behavioral models. The epidemiological models are discrete and continuous time compartment models (such as the SIS or SIR models). The behavioral models are based on a node's *ideology*, whether a continuous or discrete variable. Although these different classes of models were developed with different applications in mind, epidemiological models can inform ideological processes [164] and behavioral models contribute to our understanding of how diseases spread [154].

### 8.2.1 Epidemiological models

There are many choices of epidemiological models in the literature, but for simplicity's sake, we solely implement the SIS and SIR models. We implement 4 different contagion models in this module: the discrete SIR model, the discrete SIS model, the Gillespie SIR model, and the Gillespie SIS model. For more details on the discrete-time algorithm, see Appendix A.1 and for more details on the Gillespie algorithm, see Appendix B.3.

Unlike epidemic spreading on pairwise networks, one must make rules for when contagion spreads on hyperedges with greater than two members. These rules are expressed as contagion functions  $G$  (as presented in Eqs. (3.1)-(3.2)) and may take a variety of forms. In the exact process, these functions act as indicator functions and several examples are the following:

- Collective contagion:

$$G(x, x_1, \dots, x_{m-1}) = \prod_{i=1}^{m-1} I_{x_i=I'}$$

- Individual contagion:

$$G(x, x_1, \dots, x_{m-1}) = \max\{I_{x_i=I'} \mid i = 1, \dots, m-1\}$$

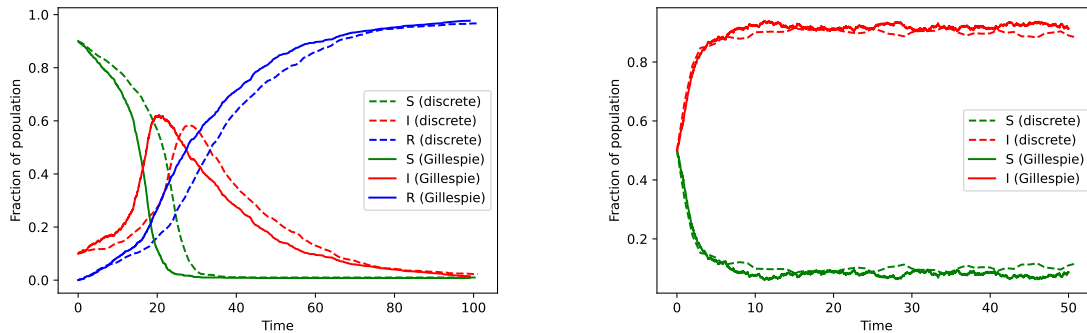
- Threshold contagion:

$$G(x, x_1, \dots, x_{m-1}, \tau) = I_{|\{I_{x_i=I'} \mid i=1, \dots, m-1\}|/m \geq \tau}$$

- Majority vote:

$$G(x, x_1, \dots, x_{m-1}) = I_{|\{I_{x_i=I'} \mid i=1, \dots, m-1\}|/m \geq 1/2}$$

The choice of contagion function can have strong effects on the dynamical behavior (for example, individual vs. collective contagion as presented in Ref. [12]). When the contagion function is non-linear (like all of the examples presented above), the resulting dynamics cannot be represented as a simple pairwise contagion process as shown in Ref. [34]. When considering a pairwise network, all these contagion functions reduce to what one would expect from classical network epidemic models. In HyperContagion, the above functions are implemented as standard functions, with the option for users to define custom contagion functions.



(a) SIR model (discrete model with  $\Delta t = 1$ , both models with  $\gamma = 0.05$  and  $\beta_m = 0.1$ ) (b) SIS model (discrete model with  $\Delta t = 1$ , both models with  $\gamma = 0.05$  and  $\beta_m = 0.1$ )

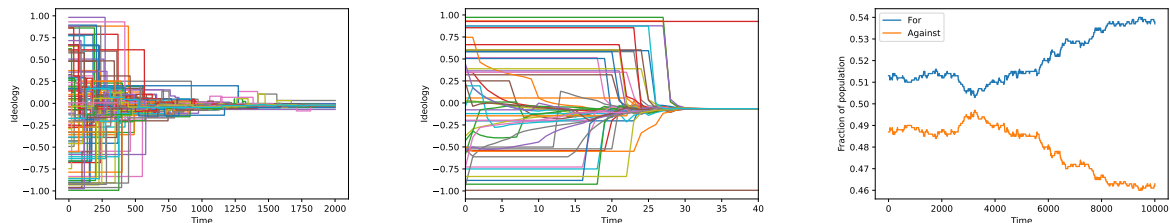
Figure 8.1: Comparing the continuous time algorithms to the discrete-time algorithms for the epidemiological models on a Chung-Lu hypergraph of size 1000.

In the limit that  $\Delta t \rightarrow 0$  the expected dynamics of the discrete models approach that of their continuous counterparts.

### 8.2.2 Behavioral models

The space of behavioral models is vast, each model seeking to incorporate a specific reality-informed mechanisms for behavior adoption. Researchers have looked at voter models [6], bounded confidence models [35, 103], threshold models [24, 22], and many others. It would be a vast undertaking to consider all possible models, so instead we implement four general classes of models: continuous opinions with random hyperedge updates, continuous opinions with synchronous up-

dates to node states, discrete opinions with hyperedge updates, discrete opinions with synchronous updates to node states.



(a) A simulation of the Deffuant-Weisbuch model on a Chung-Lu hypergraph of size 1000. This model is characterized by large jumps in opinions [35]. (b) A simulation of the Hegselmann-Krause model on a Chung-Lu hypergraph of size 1000. (c) A simulation of the voter model on a Chung-Lu hypergraph of size 1000.

Figure 8.2: Examples of hypergraph opinion formation models.

Opinion formation models with continuous opinions and random hyperedge updates describe the formation of consensus as a process driven by random group interactions, where one, some, or all of the nodes in that group update their opinion based on the opinions present in that group. A common example of this model class is the Deffuant-Weisbuch model, a type of bounded confidence model where all of the nodes in the hyperedge update their state if their opinions are close enough together [35, 165].

Opinion formation models with continuous opinions and synchronous node updates assume perfect information amongst a node's neighborhood; a node instantaneously is aware of all of its neighbors' states through the interactions it participates in. A common example of this class of opinion model is the Hegselmann-Krause model [166] which adopts the average opinion of its neighbors if their opinions are close enough.

Opinion formation models with discrete opinions and random hyperedge updates, like the continuous opinion case, assume that the formation of consensus is driven by random group interactions, where one, some, or all of the nodes in that group update their opinion based on the opinions present in that group. Unlike a continuous opinion spectrum, this class of model assumes

that there are only two meaningful positions to adopt: for or against. A common example of this class of opinion model is the voter model (for example, see Ref. [6] for the voter model on simplicial complexes).

Lastly, opinion formation models with discrete states and synchronous node updates, like the continuous opinion spectrum case, assume perfect information between a given node and its neighbors. In the pairwise network case, a common example is the threshold model [24], where nodes hold one of two opinions and all the groups to which a node belongs affect whether that node updates its opinion.

### 8.3 CompleX Group Interactions Data (XGI-DATA)

With larger datasets becoming more widely available, it is important to close the gap between dataset creators and consumers [167] because using data without an underlying knowledge of its creation and limitations is an incomplete picture. Although there are many excellent collections of hypergraph datasets [75, 162, 161], the format of each dataset and the information about how and why it was collected varies widely. Creating a *datasheet* for each dataset detailing characteristics, limitations, and surrounding factors that informed the collection of that data is a responsible practice in the curation of datasets [167]. There is significant work to be done so that researchers can easily access a summary of the dataset’s statistics, a description of the collection process, limitations of the dataset, and other relevant information. In addition, a standard format for storing hypergraph datasets with an accompanying standard will reduce the overhead time and increase accuracy for researchers working on cross-disciplinary datasets.

The XGI-DATA repository [168] is a collection of openly available hypergraph datasets in JSON format with documentation more extensively describing the datasets. There is also a rudimentary inspection script for checking that datasets are in the proper format.

An overview of the JSON data structure for all datasets:

- “hypergraph-data”: This tag accesses the attributes of the entire hypergraph dataset such

as the authors or dataset name.

- “node-data”: This tag accesses the nodes of the hypergraph and their associated properties as a dictionary where the keys are node IDs and the corresponding values are dictionaries. If a node doesn’t have any properties, the associated dictionary is empty.
  - \* “name”: This tag accesses the node’s name if there is one that is different from the ID specified in the hyperedges.
  - \* Other tags are user-specified based on the particular attributes provided by the dataset.
- “edge-data”: This tag accesses the hyperedges of the hypergraph and their associated attributes.
  - \* “name”: This tag accesses the edge’s name if one is provided.
  - \* “timestamp”: This is the tag specifying the time associated with the hyperedge if it is given. All times are stored according to the ISO8601 standard.
  - \* Other tags are user-specified based on the particular attributes provided by the dataset.
- “edge-dict”: This tag associates edge IDs with the nodes that participate in that hyperedge.

## email-Eu

### Overview

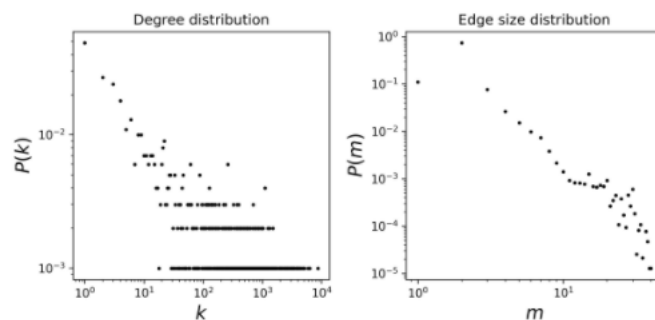
This hypergraph dataset was generated using email data from a large European research institution for a period from October 2003 to May 2005 (18 months). Information about all incoming and outgoing email between members of the research institution has been anonymized. The e-mails only represent communication between institution members (the core), and the dataset does not contain incoming messages from or outgoing messages to the rest of the world.

This is a temporal hypergraph dataset, which here means a sequence of timestamped hyperedges where each hyperedge is a set of nodes. Timestamps are in ISO8601 format. In email communication, messages can be sent to multiple recipients. In this dataset, nodes are email addresses at a European research institution. The original data source only contains directed temporal edge tuples (sender, receiver, timestamp), where timestamps are recorded at 1-second resolution. The hyperedges are undirected and consist of a sender and all receivers grouped such that the email between the sender and each receiver has the same timestamp.

### Statistics

Some basic statistics of this dataset are:

- number of nodes: 1,005
- number of timestamped hyperedges: 235,263



Hypergraph degree and edge size distributions

### Source of original data

Source: [email-Eu dataset](#)

### References

If you use this dataset, please cite these references:

- [Simplicial closure and higher-order link prediction](#), Austin R. Benson, Rediet Abebe, Michael T. Schaub, Ali Jadbabaie, and Jon Kleinberg. Proceedings of the National Academy of Sciences (PNAS), 2018.
- [Local Higher-order Graph Clustering](#), Hao Yin, Austin R. Benson, Jure Leskovec, and David F. Gleich. Proceedings of KDD, 2017.
- [Graph Evolution: Densification and Shrinking Diameters](#), Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. ACM Transactions on Knowledge Discovery from Data, 2007.

Figure 8.3: An example datasheet for the email-Eu dataset.

In Fig. 8.3, we see an example of a simple datasheet that accompanies a dataset. From this example, we see that this information sheet provides easy access to the size of the dataset, the distribution of degrees and edge sizes, a description of the data being represented, and references that one must cite to use the dataset.



## Chapter 9

### Conclusion

In this dissertation we have shown that structure in complex systems, particularly those that include group interactions, can have strong effects, both qualitatively and quantitatively, on contagion dynamics. This work is important as we seek to understand how empirical systems behave using simple models incorporating common types of structure.

In the context of hypergraphs, we have shown that the heterogeneity in the contact structure of complex systems strongly affects when tipping point behavior appears. We have also shown that degree-degree mixing in hypergraphs can be described in terms of a dynamical system and that changing the assortativity of a hypergraph may modify when epidemics appear. Lastly, we show that when community structure is present, in addition to tipping point behavior, stable polarized states emerge that one would not see in pairwise networks for this contagion model. Empirical systems often contain group interactions whether that be through social interactions, political committees, email networks, biological systems and others. Community structure, assortativity, and heterogeneity are common types of structure that have been examined for pairwise networks and it is important to understand how these notions of structure affect the dynamical behavior of a hypergraph.

In the context of pairwise networks, we have explored two important limitations that must be considered when modeling epidemics. Firstly, one must be careful when using simplified epidemic models, when there is intrinsic heterogeneity in the transmission rates of an individual over the course of their infection. Depending on the model to which one compares, the time scale

on which the epidemic spreads may be vastly different when considering time-dependent infectiousness. Secondly, empirical networks are often inherently multiplex due to different types of relationships. Choosing to use a dataset constructed from a single relationship type compared with one constructed from several relationship types is an important modeling assumption and strongly affects predictions of the epidemic extent. These considerations are important when seeking to more accurately model contagion spread on networks.

This work, of course, is limited in scope and primarily looks at contagion from a theoretical perspective. In Chapters 3 to 5, random models of hypergraphs were almost exclusively used to validate mean-field predictions. There is a gap in the literature between theoretical research where one creates simple null models in order to make analytical predictions and data science oriented research which quantifies structural properties of empirical datasets. Empirical data is far messier but hopefully these studies help close the gap between structure and dynamical behavior in hypergraphs. One could incorporate more accurate hypergraph null models and statistical inference techniques to help close this gap. In addition, in these Chapters, we presented a simple model of contagion which was not validated against empirical time series. We comment that there may be models of contagion that more accurately model behavioral adoption in groups. One may also use statistical inference to determine the best dynamical model to predict the behavior of a particular system. This may be particularly useful in epidemiology, where many models of contagion are pairwise or fully-mixed in nature. A study like this could show that incorporating group interactions may indeed provide additional information about the spread of disease. Another limitation in the work presented is that almost every project utilized a mean-field model which is well-known to have limited accuracy. In the future, one may improve our predictions by considering other approaches.

Despite these limitations, this dissertation represents a strong contribution to the fields of complexity science, higher-order interactions, and dynamical systems. It has presented a novel hyperdegree-based mean-field approach for analyzing hypergraphs and applied this framework to study the effect of structure observed in empirical datasets. It has also presented important limitations of modern contagion modeling. Lastly, it has provided several open-source resources in the

form of software and data which will we hope benefit the field of higher-order interactions for many years to come.

## Bibliography

- [1] Leo Torres, Ann S. Blevins, Danielle S. Bassett, and Tina Eliassi-Rad. The why, how, and when of representations for complex systems. [arXiv:2006.02870 \[cs, q-bio\]](https://arxiv.org/abs/2006.02870), June 2020.
- [2] Carter T. Butts. Revisiting the Foundations of Network Analysis. *Science*, 325(5939):414–416, July 2009.
- [3] Philip S Chodrow. Configuration models of random hypergraphs. *Journal of Complex Networks*, 8(cnaa018), June 2020.
- [4] Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Pataña, Jean-Gabriel Young, and Giovanni Petri. Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 874:1–92, August 2020.
- [5] Mason A. Porter. Nonlinearity + Networks: A 2020 Vision. In Panayotis G. Kevrekidis, Jesús Cuevas-Maraver, and Avadh Saxena, editors, *Emerging Frontiers in Nonlinear Science, Nonlinear Systems and Complexity*, pages 131–159. Springer International Publishing, Cham, 2020.
- [6] Leonhard Horstmeyer and Christian Kuehn. Adaptive voter model on simplicial complexes. *Physical Review E*, 101(2):022305, February 2020.
- [7] Jacopo Grilli, György Barabás, Matthew J. Michalska-Smith, and Stefano Allesina. Higher-order interactions stabilize dynamics in competitive network models. *Nature*, 548(7666):210–213, August 2017.
- [8] Per Sebastian Skardal and Alex Arenas. Abrupt Desynchronization and Extensive Multistability in Globally Coupled Oscillator Simplexes. *Physical Review Letters*, 122(24):248301, June 2019.
- [9] Can Xu, Xuebin Wang, and Per Sebastian Skardal. Bifurcation analysis and structural stability of simplicial oscillator populations. *Physical Review Research*, 2(2):023281, June 2020.
- [10] Ana P. Millán, Joaquín J. Torres, and Ginestra Bianconi. Explosive Higher-Order Kuramoto Dynamics on Simplicial Complexes. *Physical Review Letters*, 124(21):218301, May 2020.
- [11] Iacopo Iacopini, Giovanni Petri, Alain Barrat, and Vito Latora. Simplicial models of social contagion. *Nature Communications*, 10(1):2485, June 2019.

- [12] Nicholas W. Landry and Juan G. Restrepo. The effect of heterogeneity on hypergraph contagion models. Chaos: An Interdisciplinary Journal of Nonlinear Science, 30(10):103117, October 2020.
- [13] Guillaume St-Onge, Vincent Thibeault, Antoine Allard, Louis J. Dubé, and Laurent Hébert-Dufresne. Social confinement and mesoscopic localization of epidemics on networks. Physical Review Letters, 126(9):098301, March 2021.
- [14] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. Reviews of Modern Physics, 87(3):925–979, August 2015.
- [15] Pieter Trapman. On analytical approaches to epidemics on networks. Theoretical Population Biology, 71(2):160–173, March 2007.
- [16] István Z. Kiss, Joel Miller, and Péter L. Simon. Mathematics of Epidemics on Networks: From Exact to Approximate Models. Interdisciplinary Applied Mathematics. Springer International Publishing, 2017.
- [17] Thomas House. Modelling epidemics on networks. Contemporary Physics, 53(3):213–225, May 2012.
- [18] M. E. J. Newman. Spread of epidemic disease on networks. Physical Review E, 66(1):016128, July 2002.
- [19] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics in finite size scale-free networks. Physical Review E, 65(3):035108(R), March 2002.
- [20] Yang Wang, D. Chakrabarti, Chenxi Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In 22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings., pages 25–34, October 2003.
- [21] Fei Xiong and Yun Liu. Opinion formation on social media: An empirical approach. Chaos: An Interdisciplinary Journal of Nonlinear Science, 24(1):013130, March 2014.
- [22] Duncan J. Watts and Peter Sheridan Dodds. Influentials, Networks, and Public Opinion Formation. Journal of Consumer Research, 34(4):441–458, December 2007.
- [23] Robin Cowan and Nicolas Jonard. Network structure and the diffusion of knowledge. Journal of Economic Dynamics and Control, 28(8):1557–1575, June 2004.
- [24] Thomas W. Valente. Network Models of the Diffusion of Innovations. Hampton Press, 1995.
- [25] Malcolm Gladwell. The Tipping Point: How Little Things Can Make a Big Difference. Back Bay Books, Boston, January 2002.
- [26] Damon Centola, Joshua Becker, Devon Brackbill, and Andrea Baronchelli. Experimental evidence for tipping points in social convention. Science, 360(6393):1116–1119, June 2018.
- [27] Alex Arenas, Wesley Cota, Jesús Gómez-Gardeñes, Sergio Gómez, Clara Granell, Joan T. Matamalas, David Soriano, and Benjamin Steinegger. A mathematical model for the spatiotemporal epidemic spreading of COVID19. medRxiv, page 2020.03.21.20040022, March 2020.

- [28] Buddhananda Banerjee, Pradumn Kumar Pandey, and Bibhas Adhikari. A model for the spread of an epidemic from local to global: A case study of COVID-19 in India. [arXiv:2006.06404 \[physics, q-bio\]](#), June 2020.
- [29] Nicholas Landry. The effect of time-dependent infectiousness on epidemic dynamics. [arXiv:2106.10384 \[physics, q-bio\]](#), June 2021.
- [30] Marián Boguñá and Romualdo Pastor-Satorras. Epidemic spreading in correlated complex networks. *Physical Review E*, 66(4):047104, October 2002.
- [31] Joel C. Miller, Anja C. Slim, and Erik M. Volz. Edge-based compartmental modelling for infectious disease spread. *Journal of The Royal Society Interface*, 9(70):890–906, May 2012.
- [32] Owen T. Courtney and Ginestra Bianconi. Generalized network structures: The configuration model and the canonical ensemble of simplicial complexes. *Physical Review E*, 93(6):062311, June 2016.
- [33] Jean-Gabriel Young, Giovanni Petri, Francesco Vaccarino, and Alice Patania. Construction of an efficient sampling from the simplicial configuration model. *Physical Review E*, 96(3):032312, September 2017.
- [34] Leonie Neuhäuser, Andrew Mellor, and Renaud Lambiotte. Multibody interactions and non-linear consensus dynamics on networked systems. *Physical Review E*, 101(3):032310, March 2020.
- [35] Abigail Hickok, Yacoub Kureh, Heather Z. Brooks, Michelle Feng, and Mason A. Porter. A Bounded-Confidence Model of Opinion Dynamics on Hypergraphs. *SIAM Journal on Applied Dynamical Systems*, pages 1–32, January 2022.
- [36] Guilherme Ferraz de Arruda, Giovanni Petri, and Yamir Moreno. Social contagion models on hypergraphs. *Physical Review Research*, 2(2):023032, April 2020.
- [37] Ashley J. W. Ward, David J. T. Sumpter, Iain D. Couzin, Paul J. B. Hart, and Jens Krause. Quorum decision-making facilitates information transfer in fish shoals. *Proceedings of the National Academy of Sciences*, 105(19):6948–6953, May 2008.
- [38] Stephen C. Pratt, Eamonn B. Mallon, David J. Sumpter, and Nigel R. Franks. Quorum sensing, recruitment, and collective decision-making during colony emigration by the ant *Leptothorax albipennis*. *Behavioral Ecology and Sociobiology*, 52(2):117–127, July 2002.
- [39] Michael Marks, Pere Millat-Martinez, Dan Ouchi, Chrissy h Roberts, Andrea Alemany, Marc Corbacho-Monné, Maria Ubals, Aurelio Tobias, Cristian Tebé, Ester Ballana, Quique Bassat, Bàrbara Baro, Martí Vall-Mayans, Camila G-Beiras, Nuria Prat, Jordi Ara, Bonaventura Clotet, and Oriol Mitjà. Transmission of COVID-19 in 282 clusters in Catalonia, Spain: A cohort study. *The Lancet Infectious Diseases*, 21(5):629–636, May 2021.
- [40] Xi He, Eric H. Y. Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y. Wong, Yujuan Guan, Xinghua Tan, Xiaoneng Mo, Yanqing Chen, Baolin Liao, Weilie Chen, Fengyu Hu, Qing Zhang, Mingqiu Zhong, Yanrong Wu, Lingzhai Zhao, Fuchun Zhang, Benjamin J. Cowling, Fang Li, and Gabriel M. Leung. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, 26(5):672–675, May 2020.

- [41] James M. Hyman, Jia Li, and E. Ann Stanley. The differential infectivity and staged progression models for the transmission of HIV. Mathematical Biosciences, 155(2):77–109, February 1999.
- [42] Guilherme Ferraz de Arruda, Michele Tizzani, and Yamir Moreno. Phase transitions and stability of dynamical processes on hypergraphs. arXiv:2005.10891 [cond-mat, physics:physics], May 2020.
- [43] Pedro Cisneros-Velarde and Francesco Bullo. Multi-group SIS Epidemics with Simplicial and Higher-Order Interactions. arXiv:2005.11404 [physics], February 2021.
- [44] Joan T. Matamalas, Sergio Gómez, and Alex Arenas. Abrupt phase transition of epidemic spreading in simplicial complexes. Physical Review Research, 2(1):012049(R), February 2020.
- [45] Bukyoung Jhun, Minjae Jo, and B. Kahng. Simplicial SIS model in scale-free uniform hypergraph. Journal of Statistical Mechanics: Theory and Experiment, 2019(12):123207, December 2019.
- [46] Giulio Burgio, Alex Arenas, Sergio Gómez, and Joan T. Matamalas. Network clique cover approximation to analyze complex contagions through group interactions. Communications Physics, 4(1):1–10, May 2021.
- [47] Beniamino Guerra and Jesús Gómez-Gardeñes. Annealed and mean-field formulations of disease dynamics on static and adaptive networks. Physical Review E, 82(3):035101, September 2010.
- [48] Jonathan D. Touboul. The hipster effect: When anti-conformists all look the same. Discrete & Continuous Dynamical Systems - B, 24(8):4379, 2019.
- [49] Guillaume St-Onge, Vincent Thibeault, Antoine Allard, Louis J Dubé, and Laurent Hébert-Dufresne. School closures, event cancellations, and the mesoscopic localization of epidemics in networks with higher-order structure. arXiv preprint arXiv:2003.05924, 2020.
- [50] Benjamin M Althouse, Edward A Wenger, Joel C Miller, Samuel V Scarpino, Antoine Allard, Laurent Hébert-Dufresne, and Hao Hu. Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2. arXiv preprint arXiv:2005.13689, 2020.
- [51] Duygu Balcan, Bruno Gonçalves, Hao Hu, José J. Ramasco, Vittoria Colizza, and Alessandro Vespignani. Modeling the spatial spread of infectious diseases: The GLocal Epidemic and Mobility computational model. Journal of Computational Science, 1(3):132–145, August 2010.
- [52] Mark Newman. Networks. Oxford University Press, Oxford, United Kingdom ; New York, NY, United States of America, 2nd edition edition, September 2018.
- [53] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. Reviews of Modern Physics, 74(1):47–97, January 2002.
- [54] Austin R. Benson, Rediet Abebe, Michael T. Schaub, Ali Jadbabaie, and Jon Kleinberg. Simplicial closure and higher-order link prediction. Proceedings of the National Academy of Sciences, 115(48):E11221–E11230, November 2018.

- [55] Nicholas Landry. [nwlandry/simplexsis](#): Initial release, September 2020.
- [56] M. E. J. Newman. Assortative Mixing in Networks. Physical Review Letters, 89(20):208701, October 2002.
- [57] M. E. J. Newman. Mixing patterns in networks. Physical Review E, 67(2):026126, February 2003.
- [58] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. Complex networks: Structure and dynamics. Physics Reports, 424(4):175–308, February 2006.
- [59] Juan G. Restrepo and Edward Ott. Mean-field theory of assortative networks of phase oscillators. EPL (Europhysics Letters), 107(6):60006, September 2014.
- [60] Yamir Moreno, Javier B. Gómez, and Amalio F. Pacheco. Epidemic incidence in correlated complex networks. Physical Review E, 68(3):035103(R), September 2003.
- [61] Markus Brede and Sitabhra Sinha. Assortative mixing by degree makes a network more unstable. [arXiv:cond-mat/0507710](#), July 2005.
- [62] Zhihai Rong, Xiang Li, and Xiaofan Wang. Roles of mixing patterns in cooperation on a scale-free networked game. Physical Review E, 76(2):027101, August 2007.
- [63] Gregorio D’Agostino, Antonio Scala, Vinko Zlatić, and Guido Caldarelli. Robustness and assortativity for diffusion-like processes in scale-free networks. EPL (Europhysics Letters), 97(6):68006, March 2012.
- [64] Bogumił Kamiński, Valérie Poulin, Paweł Prałat, Przemysław Szufel, and François Thériège. Clustering via hypergraph modularity. PLOS ONE, 14(11):e0224307, November 2019.
- [65] Ilya Amburg, Nate Veldt, and Austin Benson. Clustering in graphs and hypergraphs with categorical edge labels. In Proceedings of The Web Conference 2020, WWW ’20, pages 706–717, New York, NY, USA, April 2020. Association for Computing Machinery.
- [66] Philip Chodrow and Andrew Mellor. Annotated hypergraphs: Models and applications. Applied Network Science, 5(1):1–25, December 2020.
- [67] Juan G. Restrepo, Edward Ott, and Brian R. Hunt. Onset of synchronization in large networks of coupled oscillators. Physical Review E, 71(3):036151, March 2005.
- [68] Juan G. Restrepo, Edward Ott, and Brian R. Hunt. Weighted Percolation on Directed Networks. Physical Review Letters, 100(5):058701, February 2008.
- [69] Brian Karrer, M. E. J. Newman, and Lenka Zdeborová. Percolation on Sparse Networks. Physical Review Letters, 113(20):208702, November 2014.
- [70] Juan G. Restrepo, Edward Ott, and Brian R. Hunt. Approximating the largest eigenvalue of network adjacency matrices. Physical Review E, 76(5):056119, November 2007.
- [71] Desmond J. Higham and Henry-Louis de Kergorlay. Epidemics on hypergraphs: Spectral thresholds for extinction. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 477(2252):20210232, August 2021.



- [72] Phillip Bonacich. Power and Centrality: A Family of Measures. American Journal of Sociology, 92(5):1170–1182, March 1987.
- [73] Austin R. Benson. Three Hypergraph Eigenvector Centralities. SIAM Journal on Mathematics of Data Science, 1(2):293–312, January 2019.
- [74] Hanlin Sun and Ginestra Bianconi. Higher-order percolation processes on multiplex hypergraphs. Physical Review E, 104(3):034306, September 2021.
- [75] Austin Benson. Data! <https://www.cs.cornell.edu/~arb/data/>, 2021.
- [76] James H. Fowler. Connecting the Congress: A Study of Cosponsorship Networks. Political Analysis, 14(4):456–487, 2006.
- [77] James H. Fowler. Legislative cosponsorship networks in the US House and Senate. Social Networks, 28(4):454–465, October 2006.
- [78] Nate Veldt, Austin R. Benson, and Jon Kleinberg. Higher-order Homophily is Combinatorially Impossible. arXiv:2103.11818 [cs], March 2021.
- [79] Nelly Litvak and Remco van der Hofstad. Uncovering disassortativity in large scale-free networks. Physical Review E, 87(2):022801, February 2013.
- [80] Remco van der Hofstad and Nelly Litvak. Degree-Degree Dependencies in Random Graphs with Heavy-Tailed Degrees. Internet Mathematics, 10(3):1560, July 2014.
- [81] Matteo Cinelli, Leto Peel, Antonio Iovanella, and Jean-Charles Delvenne. Network constraints on the mixing patterns of binary node metadata. Physical Review E, 102(6):062310, December 2020.
- [82] Koby Hayashi, Sinan G. Aksoy, Cheong Hee Park, and Haesun Park. Hypergraph Random Walks, Laplacians, and Clustering. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM '20, pages 495–504, New York, NY, USA, October 2020. Association for Computing Machinery.
- [83] Jürgen Jost and Raffaella Mulas. Hypergraph Laplace operators for chemical reaction networks. Advances in Mathematics, 351:870–896, July 2019.
- [84] Ágnes Bodó, Gyula Y. Katona, and Péter L. Simon. SIS Epidemic Propagation on Hypergraphs. Bulletin of Mathematical Biology, 78(4):713–735, April 2016.
- [85] Brian Karrer and M. E. J. Newman. Message passing approach for general epidemic models. Physical Review E, 82(1):016101, July 2010.
- [86] Alec Kirkley, George T. Cantwell, and M. E. J. Newman. Belief propagation for networks with loops. Science Advances, 7(17):eabf1211, April 2021.
- [87] Wei Wang, Ming Tang, H. Eugene Stanley, and Lidia A. Braunstein. Unification of theoretical approaches for epidemic spreading on complex networks. Reports on Progress in Physics, 80(3):036603, February 2017.
- [88] A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes. Percolation on correlated networks. Physical Review E, 78(5):051105, November 2008.

- [89] Fan Chung, Linyuan Lu, and Van Vu. Spectra of random graphs with given expected degrees. Proceedings of the National Academy of Sciences, 100(11):6313–6318, May 2003.
- [90] Claudio Castellano and Romualdo Pastor-Satorras. Relating Topological Determinants of Complex Networks to Their Spectral Properties: Structural and Dynamical Effects. Physical Review X, 7(4):041024, October 2017.
- [91] Nicholas Landry. `nwlandry/hypergraph-assortativity: v0.6`, January 2022.
- [92] Paul DiMaggio, John Evans, and Bethany Bryson. Have American’s Social Attitudes Become More Polarized? American Journal of Sociology, 102(3):690–755, November 1996.
- [93] Clio Andris, David Lee, Marcus J. Hamilton, Mauro Martino, Christian E. Gunning, and John Armistead Selden. The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives. PLOS ONE, 10(4):e0123507, April 2015.
- [94] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. The echo chamber effect on social media. Proceedings of the National Academy of Sciences, 118(9), March 2021.
- [95] Political Polarization in the American Public, June 2014.
- [96] Fabian Baumann, Philipp Lorenz-Spreen, Igor M. Sokolov, and Michele Starnini. Modeling Echo Chambers and Polarization Dynamics in Social Networks. Physical Review Letters, 124(4):048301, January 2020.
- [97] Erik A. Martens, Christian Bick, and Mark J. Panaggio. Chimera states in two populations with heterogeneous phase-lag. Chaos: An Interdisciplinary Journal of Nonlinear Science, 26(9):094819, September 2016.
- [98] Zonghua Liu and Bambi Hu. Epidemic spreading in community networks. Europhysics Letters, 72(2):315, September 2005.
- [99] Marcel Salathé and James H. Jones. Dynamics and Control of Diseases in Networks with Community Structure. PLOS Computational Biology, 6(4):e1000736, April 2010.
- [100] Fabian Baumann, Philipp Lorenz-Spreen, Igor M. Sokolov, and Michele Starnini. Emergence of Polarized Ideological Opinions in Multidimensional Topic Spaces. Physical Review X, 11(1):011012, January 2021.
- [101] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. Reviews of Modern Physics, 81(2):591–646, May 2009.
- [102] Mari Kawakatsu, Yphtach Lelkes, Simon A. Levin, and Corina E. Tarnita. Interindividual cooperation mediated by partisanship complicates Madison’s cure for “mischiefs of faction”. Proceedings of the National Academy of Sciences, 118(50), December 2021.
- [103] Heather Z. Brooks and Mason A. Porter. A model for the influence of media on the ideology of content in online social networks. Physical Review Research, 2(2):023041, April 2020.
- [104] Vicky Chuqiao Yang, Tamara van der Does, and Henrik Olsson. Falling through the cracks: Modeling the formation of social category boundaries. PLOS ONE, 16(3):e0247562, March 2021.

- [105] Vicky Chuqiao Yang, Daniel M. Abrams, Georgia Kernell, and Adilson E. Motter. Why Are U.S. Parties So Polarized? A “Satisficing” Dynamical Model. *SIAM Review*, 62(3):646–657, January 2020.
- [106] Rui-Sheng Wang and Réka Albert. Effects of community structure on the dynamics of random threshold networks. *Physical Review E*, 87(1):012810, January 2013.
- [107] Unchitta Kan, Michelle Feng, and Mason A. Porter. An Adaptive Bounded-Confidence Model of Opinion Dynamics on Networks. [arXiv:2112.05856](https://arxiv.org/abs/2112.05856) [nlin, physics:physics], December 2021.
- [108] Guilherme Ferraz de Arruda, Giovanni Petri, Pablo Martín Rodríguez, and Yamir Moreno. Multistability, intermittency and hybrid transitions in social contagion models on hypergraphs. [arXiv:2112.04273](https://arxiv.org/abs/2112.04273) [cond-mat, physics:physics], December 2021.
- [109] Chiheon Kim, Afonso S. Bandeira, and Michel X. Goemans. Stochastic Block Model for Hypergraphs: Statistical limits and a semidefinite programming approach. [arXiv:1807.02884](https://arxiv.org/abs/1807.02884) [cs, math], July 2018.
- [110] Sam Cole and Yizhe Zhu. Exact recovery in the hypergraph stochastic block model: A spectral algorithm. *Linear Algebra and its Applications*, 593:45–73, May 2020.
- [111] Debarghya Ghoshdastidar and Ambedkar Dukkipati. Consistency of spectral hypergraph partitioning under planted partition model. *The Annals of Statistics*, 45(1):289–315, February 2017.
- [112] Mark Jerrum and Gregory B. Sorkin. The Metropolis algorithm for graph bisection. *Discrete Applied Mathematics*, 82(1):155–175, March 1998.
- [113] Philip S. Chodrow, Nate Veldt, and Austin R. Benson. Generative hypergraph clustering: From blockmodels to modularity. *Science Advances*, July 2021.
- [114] Despina Stasi, Kayvan Sadeghi, Alessandro Rinaldo, Sonja Petrović, and Stephen E. Fienberg.  $\beta$  models for random hypergraphs with a given degree sequence. [arXiv:1407.1004](https://arxiv.org/abs/1407.1004) [cs, math, stat], July 2014.
- [115] Brian Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, January 2011.
- [116] Tiago P. Peixoto. Nonparametric Bayesian inference of the microcanonical stochastic block model. *Physical Review E*, 95(1):012317, January 2017.
- [117] Zheng Tracy Ke, Feng Shi, and Dong Xia. Community Detection for Hypergraph Networks via Regularized Tensor Power Iteration. [arXiv:1909.06503](https://arxiv.org/abs/1909.06503) [math, stat], January 2020.
- [118] Erik M. Volz, Joel C. Miller, Alison Galvani, and Lauren Ancel Meyers. Effects of Heterogeneous and Clustered Contact Patterns on Infectious Disease Dynamics. *PLOS Computational Biology*, 7(6):e1002042, June 2011.
- [119] Clara Stegehuis and Thomas Peron. Network processes on clique-networks with high average degree: The limited effect of higher-order structure. [arXiv:2104.14776](https://arxiv.org/abs/2104.14776) [physics], April 2021.

- [120] Alun L. Lloyd. Realistic Distributions of Infectious Periods in Epidemic Models: Changing Patterns of Persistence and Dynamics. Theoretical Population Biology, 60(1):59–71, August 2001.
- [121] Junling Ma and David J. D. Earn. Generality of the Final Size Formula for an Epidemic of a Newly Invading Infectious Disease. Bulletin of Mathematical Biology, 68(3):679–702, April 2006.
- [122] Neil Sherborne, Joel C. Miller, Konstantin B. Blyuss, and Istvan Z. Kiss. Mean-field models for non-Markovian epidemics on networks. Journal of Mathematical Biology, 76(3):755–778, February 2018.
- [123] Daniel B. Larremore, Bryan Wilder, Evan Lester, Soraya Shehata, James M. Burke, James A. Hay, Milind Tambe, Michael J. Mina, and Roy Parker. Test sensitivity is secondary to frequency and turnaround time for COVID-19 surveillance. medRxiv, page 2020.06.22.20136309, September 2020.
- [124] G. Röst, Z. Vizi, and I. Z. Kiss. Pairwise approximation for SIR-type network epidemics with non-Markovian recovery. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 474(2210):20170695, February 2018.
- [125] Benjamin M. Althouse, Edward A. Wenger, Joel C. Miller, Samuel V. Scarpino, Antoine Allard, Laurent Hébert-Dufresne, and Hao Hu. Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and control. PLOS Biology, 18(11):e3000897, November 2020.
- [126] Michael J. Mina, Roy Parker, and Daniel B. Larremore. Rethinking Covid-19 Test Sensitivity — A Strategy for Containment. New England Journal of Medicine, 383(22):e120, November 2020.
- [127] N. Sherborne, K. B. Blyuss, and I. Z. Kiss. Dynamics of Multi-stage Infections on Networks. Bulletin of Mathematical Biology, 77(10):1909–1933, October 2015.
- [128] O. Diekmann, J. a. P. Heesterbeek, and M. G. Roberts. The construction of next-generation matrices for compartmental epidemic models. Journal of The Royal Society Interface, 7(47):873–885, June 2010.
- [129] Dina Mistry, Maria Litvinova, Ana Pastore y Piontti, Matteo Chinazzi, Laura Fumanelli, Marcelo F. C. Gomes, Syed A. Haque, Quan-Hui Liu, Kunpeng Mu, Xinyue Xiong, M. Elizabeth Halloran, Ira M. Longini, Stefano Merler, Marco Ajelli, and Alessandro Vespignani. Inferring high-resolution human mixing patterns for disease modeling. Nature Communications, 12(1):323, January 2021.
- [130] Bailey K. Fosdick, Daniel B. Larremore, Joel Nishimura, and Johan Ugander. Configuring Random Graph Models with Fixed Degree Sequences. SIAM Review, 60(2):315–355, January 2018.
- [131] Michele Catanzaro, Marián Boguñá, and Romualdo Pastor-Satorras. Generation of uncorrelated random scale-free networks. Physical Review E, 71(2):027103, February 2005.
- [132] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani. Activity driven modeling of time varying networks. Scientific Reports, 2(1):469, June 2012.

- [133] Wei Gou and Zhen Jin. How heterogeneous susceptibility and recovery rates affect the spread of epidemics on networks. Infectious Disease Modelling, 2(3):353–367, August 2017.
- [134] Nicholas Landry and Karen (Ren) Stengel. Time-dependent-infectiousness. Zenodo, October 2021.
- [135] James Moody, Jimi Adams, and Martina Morris. Epidemic potential by sexual activity distributions. Network Science, 5(4):461–475, December 2017.
- [136] jimi adams, James Moody, and Martina Morris. Sex, Drugs, and Race: How Behaviors Differentially Contribute to the Sexually Transmitted Infection Risk Network Structure. American Journal of Public Health, 103(2):322–329, February 2013.
- [137] Sinan Aral, Lev Muchnik, and Arun Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proceedings of the National Academy of Sciences, 106(51):21544–21549, December 2009.
- [138] Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. A 61-million-person experiment in social influence and political mobilization. Nature, 489(7415):295–298, September 2012.
- [139] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. The science of fake news. Science, March 2018.
- [140] Leticia Bode and Emily K. Vraga. Studying Politics Across Media. Political Communication, 35(1):1–7, January 2018.
- [141] Eszter Hargittai. Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites. The ANNALS of the American Academy of Political and Social Science, 659(1):63–76, May 2015.
- [142] J Potterat, Donald E Woodhouse, Stephen Q Muth, Richard B Rothenburg, William W Darrow, Alden S Klovdahl, and John B Muth. Network dynamism: History and lessons of the colorado springs study. In Network Epidemiology: A Handbook for Survey Design and Data Collection, pages 87–114. Oxford University Press, New York, 2004.
- [143] Manlio De Domenico, Clara Granell, Mason A. Porter, and Alex Arenas. The physics of spreading processes in multilayer networks. Nature Physics, 12(10):901–906, October 2016.
- [144] S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Pérez-Vicente, Y. Moreno, and A. Arenas. Diffusion Dynamics on Multiplex Networks. Physical Review Letters, 110(2):028701, January 2013.
- [145] Filippo Radicchi and Alex Arenas. Abrupt transition in the structural formation of interconnected networks. Nature Physics, 9(11):717–720, November 2013.
- [146] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin. The structure and dynamics of multilayer networks. Physics Reports, 544(1):1–122, November 2014.

- [147] Kayla Kauffman, Courtney S. Werner, Georgia Titcomb, Michelle Pender, Jean Yves Rabezara, James P. Herrera, Julie Teresa Shapiro, Alma Solis, Voahangy Soarimalala, Pablo Tortosa, Randall Kramer, James Moody, Peter J. Mucha, and Charles Nunn. Comparing transmission potential networks based on social network surveys, close contacts and environmental overlap in rural Madagascar. Journal of The Royal Society Interface, 19(186):20210690, January 2022.
- [148] A. S. Klovdahl, J. J. Potterat, D. E. Woodhouse, J. B. Muth, S. Q. Muth, and W. W. Darrow. Social networks and infectious disease: The Colorado Springs study. Social Science & Medicine, 38(1):79–88, January 1994.
- [149] Mahdi Jalili, Yasin Orouskhani, Milad Asgari, Nazanin Alipourfard, and Matjaž Perc. Link prediction in multiplex online social networks. Royal Society Open Science, 4(2):160863, February 2017.
- [150] Martina Morris and Richard Rothenberg. HIV transmission network metastudy project: An archive of data from eight network studies, 1988–2001, 2011.
- [151] Mahdi Jalili, Yasin Orouskhani, Milad Asgari, Nazanin Alipourfard, and Matjaž Perc. JOAPP from Link Prediction in Multiplex Online Social Networks, 2017.
- [152] Robert M. May and Roy M. Anderson. Transmission dynamics of HIV infection. Nature, 326(6109):137–142, March 1987.
- [153] Douglas Guilbeault and Damon Centola. Topological measures for identifying and predicting the spread of complex contagions. Nature Communications, 12(1):4430, July 2021.
- [154] Clara Granell, Sergio Gómez, and Alex Arenas. Dynamical Interplay between Awareness and Epidemic Spreading in Multiplex Networks. Physical Review Letters, 111(12):128701, September 2013.
- [155] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using NetworkX. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, Proceedings of the 7th Python in Science Conference, pages 11–15, Pasadena, CA USA, 2008.
- [156] Tiago P. Peixoto. The graph-tool python library. figshare, 2014.
- [157] Jean-Gabriel Young, George T Cantwell, and M E J Newman. Bayesian inference of network structure from unreliable data. Journal of Complex Networks, 8(cnaa046), March 2021.
- [158] Alessia Antelmi, Gennaro Cordasco, Bogumił Kamiński, Paweł Prałat, Vittorio Scarano, Carmine Spagnuolo, and Przemysław Szufel. Analyzing, Exploring, and Visualizing Complex Networks via Hypergraphs using SimpleHypergraphs.jl. Internet Mathematics, April 2020.
- [159] Brenda Praggastis, Dustin Arendt, Cliff Joslyn, Emilie Purvine, Sinan Aksoy, and Kyle Monsen. PNNL/HyperNetX.
- [160] Nicholas Landry, Leo Torres, Iacopo Iacopini, Maxime Lucas, Giovanni Petri, and Alice Patania. XGI.

- [161] Aaron Clauset, Ellen Tucker, and Matthias Sainz. *The Colorado Index of Complex Networks*, 2016.
- [162] Tiago P. Peixoto. *Netzschleuder*, 2021.
- [163] Nicholas W. Landry and Juan G. Restrepo. Hypergraph assortativity: A dynamical systems perspective. [arXiv:2109.01099 \[physics\]](https://arxiv.org/abs/2109.01099), January 2022.
- [164] Alexandria Volkening, Daniel F. Linder, Mason A. Porter, and Grzegorz A. Rempala. Forecasting Elections Using Compartmental Models of Infection. *SIAM Review*, 62(4):837–865, January 2020.
- [165] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 03(01n04):87–98, January 2000.
- [166] Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, 5:1–24, 2002.
- [167] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, November 2021.
- [168] Nicholas Landry, Leo Torres, Maxime Lucas, Iacopo Iacopini, Giovanni Petri, and Alice Patania. *xgi-data*.
- [169] Joel C. Miller and Aric Hagberg. Efficient Generation of Networks with Given Expected Degrees. In Alan Frieze, Paul Horn, and Paweł Prałat, editors, *Algorithms and Models for the Web Graph*, Lecture Notes in Computer Science, pages 115–126, Berlin, Heidelberg, 2011. Springer.

## Appendix A

### Appendix for Chapter 3

#### A.1 Microscopic simulation of the hypergraph SIS model

We simulated the stochastic SIS model on a given hypergraph as a discrete-time Markov process on the nodes with transitions to infected and healthy states through the modalities described in Section 2.2, a variant of which was simulated in Ref. [84]. We denote the binary states of the nodes at a time step  $t$  by a vector  $\mathbf{X}^t = (X_1^t, X_2^t, \dots, X_N^t)$  where  $X_i^t = 0$  if node  $i$  is healthy and  $X_i^t = 1$  if it is infected. In this model, we assume that the events that a hyperedge infects node  $i$  and that a pairwise connection infects node  $i$  are independent. Likewise, we assume that an infected neighbor, whether through a pairwise or group connection, infects a node independently of any other neighboring node. The probability that a single infected node infects its pairwise neighbor in the time interval  $[t, t + \Delta t]$  is  $\beta_2 \Delta t$ , and so the probability that no neighboring node infects a given node is

$$(1 - \beta_2 \Delta t)^{(A\mathbf{X})_i}$$

where  $A$  is the adjacency matrix with entries  $A_{ij} = 1$  if nodes  $i$  and  $j$  are connected by a link and 0 otherwise.

In the collective contagion model, the probability that a triangle infects a node in the time interval  $[t, t + \Delta t]$  is  $\beta_3 \Delta t$  provided the other two nodes are infected. Therefore, the probability of no triangles infecting node  $i$  can be written as

$$(1 - \beta_3 \Delta t)^{\sum_{\{i_1, i_2, i\}} X_{i_1}^t X_{i_2}^t},$$



where the sum is over all triangles  $\{i_1, i_2, i\}$  with node  $i$  as a member. Lastly, the rate of healing is constant and independent of the infection status of any neighboring nodes so the probability that an infected node heals in a time interval  $[t, t + \Delta t]$  is  $\gamma\Delta t$ .

The Markov process can then be described as

$$P(X_i^{t+1} = 1 \mid X_i^t = 0) = 1 - (1 - \beta_2\Delta t)^{(A\mathbf{X}^t)_i} (1 - \beta_3\Delta t)^{\sum_{\{i_1, i_2, i\}} X_{i_1}^t X_{i_2}^t}, \quad (\text{A.1})$$

$$P(X_i^{t+1} = 0 \mid X_i^t = 1) = \gamma\Delta t. \quad (\text{A.2})$$

In our simulations, we updated the status of the nodes synchronously at times

$$t = 0, \Delta t, 2\Delta t, \dots, n\Delta t$$

where  $\Delta t = 0.1$ .

Our specific implementation is described in what follows. We note that for all mechanisms of infection and healing described next,  $u_i \sim \text{Uniform}(0, 1)$  and this variable is drawn independently for each modality and each node  $i$ . At each time step, we iterate through every node and follow the following conditional logic. If a node  $i$  is already infected, it is healed if  $u_i < \gamma\Delta t$  and remains infected otherwise. Next, if the node  $i$  is currently healthy, it is infected by its pairwise neighbors if  $u_i < 1 - (1 - \beta_2\Delta t)^{(A\mathbf{X})_i}$  and remains healthy otherwise. If node  $i$  still remains healthy after being subjected to pairwise infection, the node is infected by its triangle neighbors if

$$u_i < 1 - (1 - \beta_3\Delta t)^{\sum_{\{i_1, i_2, i\}} X_{i_1}^t X_{i_2}^t}$$

and remains healthy otherwise. Note that each infection mechanism is only dependent on the prior time step so the order of these steps does not matter.

At  $t = 0$ , the network is randomly and uniformly seeded with a small fraction ( $p = 0.001$ ) of infected nodes and at each subsequent step, the current state is iterated as described above and the population average,  $x^t = \sum_{i=1}^N X_i^t / N$  is stored. To avoid the absorbing state  $\mathbf{X}^t = \mathbf{0}$ , we infect a single randomly chosen node if the population becomes completely healthy. To mitigate the effect of variability in the stochastic simulation, we average the time response of  $x^t$  over a sufficient time

window (determined from the average infected response curves) after it reached the steady-state. In this study, we ran the simulation for a fixed set of parameters  $\{\gamma, \beta_2, \beta_3\}$  for 1000 time steps and averaged over the last 300 time steps.

To find the bistability index, we initialize the simulation with a small fraction of infected agents for a fixed  $\beta_3$  value and incrementally increase  $\beta_2$  from a sufficiently small value (typically  $\beta_2^c/2$ ) to a value above the critical value of  $\beta_2$  (typically  $3\beta_2^c/2$ ), and then incrementally decrease the value of  $\beta_2$  down to its original value. As described previously, if the equilibrium value while increasing the value of  $\beta_2$  is distinct from the equilibrium value while decreasing the value of  $\beta_2$  for the same  $\beta_2$  values, this indicates the presence of bistability. We simulated several equilibrium curves corresponding to different  $\beta_3$  values to observe the value of  $\beta_3$  at which the response curve starts to show bistability, and thus infer the value of  $\beta_3^c$ .

## A.2 Network models

We exclusively considered networks generated using the configuration model in order to isolate the effect of the degree distribution. Although the configuration model has the potential to contain both self-loops and multi-edges, in practice, the fraction of these types of edges is small[52] and in our numerical experiments, the number of self-loops was approximately 1% of the total number of nodes.

We used networks of size  $N = 10^4$  in the simulation of the hypergraph SIS model because this was sufficiently large enough to reduce the finite-size effects. Because the network realization was relatively large, we did not average over an ensemble of these random graphs as in Ref. [11]. We have described in Section 3.1.1 the particular distributions examined.

We generated the triangles in two different ways corresponding to the two separate cases; degree-correlated and uncorrelated. For the first case, we used the same degree sequence as used to generate the network using the configuration model and extended the configuration model to triangles as has been done in prior work[33, 32]. Because this is analogous to the construction of the network configuration model, there is also the possibility for self-loops and multi-edges, but

this probability is low. For the independently-distributed triangles, we drew with replacement a fixed number of triples (enforcing the mean triangle degree) containing node indices and assigned these nodes to a triangle. Again, as with the standard configuration model, there is the possibility for self-loops and multi-edges, but the probability of either occurring is small.

### A.2.1 The numerical computation of $\beta_3^c$

In Section 3.2, we plotted the numerical solution of  $\beta_3^c$  for truncated power law distributions as a function of the maximum degree and power-law exponent. In this section, we discuss the specific methodology in generating these results.

First, we describe the process for finding the bistability index accurately from the mean-field equations (3.10) and (3.15)-(3.16) for the correlated and uncorrelated cases respectively.

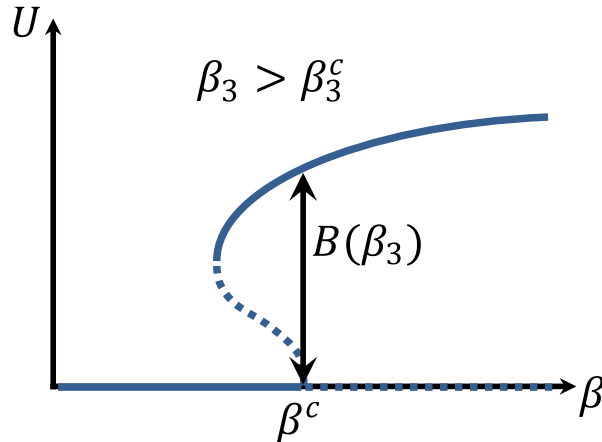


Figure A.1: Illustration of the bistability index with respect to the solutions to the mean-field equation in the bistable regime

Since the  $V = 0$  solution becomes linearly unstable at  $\beta_2 = \beta_2^c$  and the stable  $V > 0$  solution is monotonically increasing with  $\beta_2$ , the bistability index  $B(\beta_3)$  coincides with the value of the largest root of Eq. (3.10) for the correlated case [or Eqs. (3.15)-(3.16) for the uncorrelated case] at  $\beta_2 = \beta_2^c$ , as shown schematically in Fig. A.1. Therefore, using our analytical knowledge of  $\beta_2^c$ , we set  $B(\beta_3) \approx V_\epsilon^*$ , where  $V_\epsilon^*$  is the largest root at  $\beta_2 = \beta_2^c - \epsilon$  with  $\epsilon = 10^{-5}$  a small number added

for numerical robustness. (We verified that this method gives numerically accurate results when compared with other methods which do not require knowledge of  $\beta_2^c$  but are more computationally intensive.)

Being able to compute  $B(\beta_3)$ , we find  $\beta_3^c = \sup\{\beta_3 \mid B(\beta_3) = 0\}$  by bisection: starting with an interval  $[\beta_3^{\min,0}, \beta_3^{\max,0}]$  such that  $B(\beta_3^{\min,0}) = 0$ ,  $B(\beta_3^{\max,0}) > 0$ , we recursively define the interval  $[\beta_3^{\min,i+1}, \beta_3^{\max,i+1}]$  as  $[\beta_3^{\min,i}, \tilde{\beta}^i]$  if  $B(\tilde{\beta}^i) > 0$  and  $[\tilde{\beta}^i, \beta_3^{\max,i}]$  if  $B(\tilde{\beta}^i) = 0$ , where  $\tilde{\beta}^i = (\beta_3^{\min,i} + \beta_3^{\max,i})/2$ . When the length of the interval  $[\beta_3^{\min,i}, \beta_3^{\max,i}]$  is less than the tolerance  $10^{-4}$ , we set  $\beta_3^c = \beta_3^{\min,i}$ .

## Appendix B

### Appendix for Chapter 4

#### B.1 More detailed derivation of the perturbed eigenvalue

We start with the expansion of Eq. (2) in the main text to first order (we recall that we are considering an  $m$ -uniform hypergraph), which is

$$\begin{aligned}
\alpha\lambda^{(0)}k + \epsilon\lambda^{(0)}u_k^{(1)} + \alpha\epsilon\lambda^{(1)}k &= \alpha \sum_{k_1, \dots, k_{m-1}} N(k_1) \dots N(k_{m-1}) \frac{k k_1 \dots k_{m-1}}{(N\langle k \rangle)^{m-1}} (k_1 + \dots + k_{m-1}) \\
&+ \epsilon \sum_{k_1, \dots, k_{m-1}} N(k_1) \dots N(k_{m-1}) \frac{k k_1 \dots k_{m-1}}{(N\langle k \rangle)^{m-1}} (u_{k_1}^{(1)} + \dots + u_{k_{m-1}}^{(1)}) \\
&+ \alpha\epsilon \sum_{k_1, \dots, k_{m-1}} N(k_1) \dots N(k_{m-1}) \frac{k k_1 \dots k_{m-1}}{(N\langle k \rangle)^{m-1}} \\
&\times g_m(k, k_1, \dots, k_{m-1})(k_1 + \dots + k_{m-1}).
\end{aligned} \tag{B.1}$$

From the 0th order approximation, the first terms on both sides of the equation are equal and we can cancel them. Secondly, assuming symmetry of  $f_m$  and  $g_m$ , we can simplify the right-hand side as

$$\begin{aligned}
\epsilon\lambda^{(0)}u_k^{(1)} + \alpha\epsilon\lambda^{(1)}k &= \epsilon(m-1)k \sum_{k_1} P(k_1) \frac{k_1 u_{k_1}^{(1)}}{\langle k \rangle} \\
&+ \alpha\epsilon(m-1) \sum_{k_1, \dots, k_{m-1}} N(k_1) \dots N(k_{m-1}) \frac{k k_1^2 k_2 \dots k_{m-1}}{(N\langle k \rangle)^{m-1}} g_m(k, k_1, \dots, k_{m-1}).
\end{aligned}$$

We multiply both sides by  $k P(k)/\langle k \rangle$  and sum over  $k$  which yields

$$\begin{aligned} \epsilon \lambda^{(0)} \sum_k P(k) \frac{k u_k^{(1)}}{\langle k \rangle} + \alpha \epsilon \lambda^{(1)} \sum_k P(k) \frac{k^2}{\langle k \rangle} &= \epsilon(m-1) \sum_k P(k) \frac{k^2}{\langle k \rangle} \sum_{k_1} P(k_1) \frac{k_1 u_{k_1}^{(1)}}{\langle k \rangle} \\ &+ \alpha \epsilon(m-1) \sum_{k, k_1, \dots, k_{m-1}} N(k) N(k_1) \dots N(k_{m-1}) \\ &\times \frac{k^2 k_1^2 k_2 \dots k_{m-1}}{(N\langle k \rangle)^m} g_m(k, k_1, \dots, k_{m-1}). \end{aligned}$$

Because  $\lambda^{(0)} = (m-1)\langle k^2 \rangle / \langle k \rangle$ , the first terms on both sides are equal and we cancel them, yielding

$$\epsilon \lambda^{(1)} = \epsilon(m-1) \frac{\langle k \rangle}{\langle k^2 \rangle} \sum_{k, k_1, \dots, k_{m-1}} N(k) N(k_1) \dots N(k_{m-1}) \frac{k^2 k_1^2 \dots k_{m-1}}{(N\langle k \rangle)^m} g_m(k, k_1, \dots, k_{m-1}). \quad (\text{B.2})$$

We can use the relation that

$$f_m(k_1, \dots, k_m) = (m-1)! k_1 \dots k_m / (N\langle k \rangle)^{m-1} [1 + \epsilon g_m(k_1, \dots, k_m)]$$

to remove the reference to  $g_m$ , obtaining

$$\begin{aligned} \epsilon \lambda^{(1)} &= \frac{(m-1)}{(m-1)!} \frac{\langle k \rangle}{\langle k^2 \rangle} \sum_{k, k_1, \dots, k_{m-1}} N(k) N(k_1) \dots N(k_{m-1}) \frac{k k_1}{N\langle k \rangle} f_m(k, k_1, \dots, k_{m-1}) \\ &- (m-1) \frac{\langle k \rangle}{\langle k^2 \rangle} \sum_{k, k_1} P(k) P(k_1) \frac{k^2 k_1^2}{\langle k \rangle^2}. \end{aligned}$$

The term

$$\frac{1}{2!(m-2)!} \sum_{k, k_1, \dots, k_{m-1}} N(k) N(k_1) \dots N(k_{m-1}) k k_1 f_m(k, k_1, \dots, k_{m-1})$$

represents the expected sum of all products of degrees for pairs of nodes belonging to the same hyperedge (where the factors  $2!$  and  $(m-2)!$  correct for overcounting permutations of  $k, k_1$  and  $k_2, k_3, \dots, k_{m-1}$  respectively). Since the number of possible pairwise products in an  $m$ -uniform hypergraph is given by

$$\begin{aligned} \sum_{e \in E} \sum_{k, k' \in e, k \neq k'} 1 &= \binom{m}{2} |E| = \binom{N\langle k \rangle}{m} \binom{m(m-1)}{2} \\ &= \frac{(m-1)N\langle k \rangle}{2}, \end{aligned} \quad (\text{B.3})$$

letting  $|E|$  be the number of edges, we can express  $\lambda^{(1)}$  in terms of

$$\langle kk_1 \rangle_E = \frac{1}{(m-1)!} \sum_{k, k_1, \dots, k_{m-1}} N(k)N(k_1) \dots N(k_{m-1}) \frac{k k_1}{N \langle k \rangle} f_m(k, k_1, \dots, k_{m-1}),$$

the average of pairwise degree products over pairs of connected nodes, as

$$\epsilon \lambda^{(1)} = (m-1) \frac{\langle k \rangle \langle kk_1 \rangle_E}{\langle k^2 \rangle} - \lambda^{(0)}.$$

Therefore,

$$\lambda = \lambda^{(0)} + \epsilon \lambda^{(1)} = (m-1) \frac{\langle k \rangle \langle kk_1 \rangle_E}{\langle k^2 \rangle} = \lambda^{(0)}(1 + \rho), \quad (\text{B.4})$$

where

$$\rho = \frac{\langle k \rangle^2 \langle kk_1 \rangle_E}{\langle k^2 \rangle^2} - 1. \quad (\text{B.5})$$

## B.2 Suppressing epidemics through preferential rewiring

In this Section, we include additional plots of the effect of disassortative rewiring on the epidemic extent. We consider the CM and EE datasets described in the main text. The following plots have the same structure as that in the main text so we omit the legend for simplicity.

In Fig. B.1a, we see the same behavior as that of the CB dataset. We comment that, as we expect, the epidemic threshold is fairly close to the predicted extinction threshold. In Figs. B.1b and B.1d, we see behavior that differs from that of the CB dataset, but is consistent with our theoretical approach. In Fig. B.1b, we see that the epidemic extent is roughly less than 0.25% for all values of  $\rho$ . This does not contradict the bounds we derived because there is no epidemic below the extinction threshold. The behavior in Fig. B.1d indicates that additional structure is present in the original hypergraph that seems to be suppressing the epidemic as well and warrants further study.

As discussed in the text, it is possible that if hyperedge swaps do not bring  $\beta_3/\beta_3^c$  below 1 as in Fig. B.1d, the process results in an epidemic. While we only see this for the EE dataset, one should be cautious of rewiring the hypergraph unless one can guarantee that  $\beta_3/\beta_3^c < 1$  can be achieved.

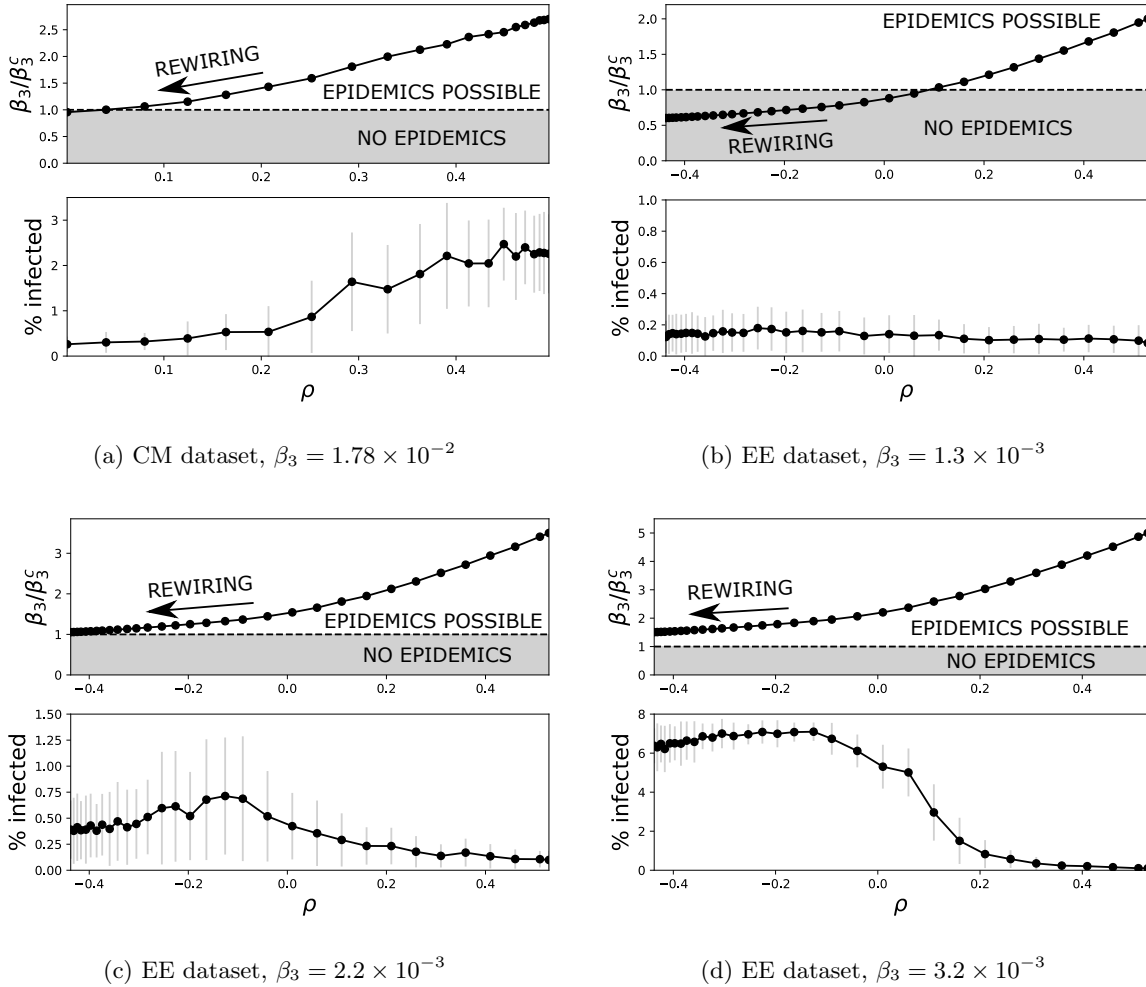


Figure B.1: Additional plots of preferential rewiring and the corresponding epidemic response. Each subplot corresponds to a particular choice of infectious rate and dataset and follows the same format as Fig. 4.3.

### B.3 Numerical simulations

We model the hypergraph SIS contagion process as a continuous-time discrete-state (CTDS) Markov process, in contrast to Refs. [11, 12] which assume a discrete-time (DTDS) process. In Ref. [46], the authors find that discrete-time processes inaccurately model contagion processes due to higher-order correlations. We note that as the time step in a DTDS process approaches zero, we recover the dynamics of the continuous time process.



As described in the main text, we consider a 3-uniform hypergraph of size  $N$ . We specify a spontaneous healing rate  $\gamma$  and an infection rate  $\beta_3$  at which an infected 3-hyperedge infects a susceptible node. The total rate at which infected nodes recover is given by the number of infected nodes  $N_I$  multiplied by the healing rate  $\gamma$ . The rate at which each susceptible node  $i$  is infected is given by the number of infected hyperedges (hyperedges with at least one infected neighbor) of which it is a member,  $N_i^E$ , multiplied by the infection rate  $\beta_3$ , and the total infection rate is  $\beta_3 \sum_{i=1}^N N_i^E$ . The total rate at which these disjoint events occur is their sum, i.e.,  $R = \gamma N_I + \beta_3 \sum_{i=1}^N N_i^E$ .

For this CTDS process, the time between events,  $\tau$ , is drawn from the exponential distribution with rate  $R$ . Once this time has been determined, we must determine which type of event occurred. The probability that a node recovers is  $p = \gamma N_I / R$  and the probability that a node becomes infected is  $1 - p$  and so we can draw the event from a Bernoulli distribution with parameter  $p$ . Next we must determine the node for which this event occurred. If an infected node has recovered, we select this node uniformly at random from the list of infected nodes. If a susceptible node has become infected, we select a node from the list of infected nodes according to the probabilities  $p_i = N_i^E / \left( \sum_{i=1}^N N_i^E \right)$  for each node  $i$ .

Once the time increment, event type, and affected node have been determined, we first increment the time  $t_i$  by  $\tau$ ; second, we increment the number of infected individuals by one and decrease the number of susceptible individuals by one for an infection event (vice-versa for a recovery event); and lastly, we update the list of susceptible and infected nodes as well as the rates of each mechanism. We repeat these steps until either  $t$  exceeds a maximum specified time or the number of infected nodes is zero. We refer to this termination time as  $T$  and the corresponding number of discrete data points as  $N_T$ . Modeling the SIS contagion process as a CTDS process can be more efficient than a DTDS process when  $R$  is small because the exponential distribution allows the simulation to take large steps in time when  $R$  is small.

To recover the equilibrium from these simulations, we average over the last 10% of the simulation time, i.e., the interval  $[T_0, T]$ , where  $T_0 = 0.9T$ . We calculate an average of the number

of infected nodes at each time step,  $i = 1, \dots, N_T$ , weighted by the interevent time, i.e.,

$$\langle I \rangle = \left[ I_1(T - T_0)/N_T + \sum_{i=2}^{N_T} I_i(t_i - t_{i-1}) \right] / T.$$

## Appendix C

### Appendix for Chapter 5

#### C.1 An efficient algorithm for sampling $m$ -HSBM hypergraphs

Simply sampling naïvely from the list of all possible hyperedges of size  $m$  and accepting them with probability  $p \ll 1$  is  $\mathcal{O}(N^m)$  which becomes increasingly inefficient as  $m$  increases. We modify the algorithm presented in Ref. [169] to  $m$ -uniform hypergraphs to sample from the  $m$ -HSBM with time complexity  $\mathcal{O}(m(N + |E|))$ .

Although in Section 5.1.1 we specify that no multiedges or self-loops may occur, given an index in the list of all possible unique hyperedges, it is expensive to recover the hyperedge corresponding to that index, particularly for large hyperedges. If instead we allow these artifacts to occur, then it is  $\mathcal{O}(m)$  to recover the hyperedge of interest in contrast to iterating through all combinations which is  $\mathcal{O}(mN)$ . Because of this modification, in practice, we divide every probability derived in Section 5.1.2 by  $m!$  to account for the increase in possible hyperedges because of multiedges and hyperedges containing self-loops.

Consider an  $m$ -uniform hypergraph with  $N$  nodes and  $G$  groups. As in Section 5.1.1,  $P_{g_{i_1}, \dots, g_{i_m}}$  specifies the probability that nodes  $i_1, \dots, i_m$  with community labels  $g_{i_1}, \dots, g_{i_m}$ , form a hyperedge. The function  $\mathcal{G}$  that returns the members in a group  $g$  is defined as

$$\mathcal{G} : g \mapsto (i \mid g_i = g, i = 1, \dots, N)^T,$$

where  $|\mathcal{G}(g)|$  is the number of nodes with community label  $g$ . We iterate through each entry  $g_1, \dots, g_m$  of the tensor  $P$  and for fixed group assignments for the nodes in a hyperedge,  $\mathbf{g} =$

$g_{i_1}, \dots, g_{i_m}$ , the probability of generating a hyperedge is constant and we generate all the hyperedges associated with these ordered community assignments. The hyperedges are elements of the set formed by the Cartesian product of the indices in each partition and the maximum index is given by the product  $\prod_{g \in \mathbf{g}} |\mathcal{G}(g)|$ .

Our algorithm is an extension of the algorithm in Ref. [169] and the main idea is this: instead of iterating through all possible edges and accepting an edge with probability  $P_{g_{i_1}, \dots, g_{i_m}} = p$ , which is expensive when  $p \ll 1$ , we simply “skip” the edges that would be rejected by sampling from a geometric distribution. While the current index is less than the maximum index, we increment the index with steps  $s \sim \text{Geometric}(p)$ . For a given index, we convert to a list of  $m$  node labels with Algorithm 1. Because we simulate the community connection probability tensor patch-by-patch, for a given patch, we specify the community to which each node belongs as an ordered list.

---

**Algorithm 1:** Get an  $m$ -hyperedge from a specified index, given community sizes (Index-ToEdge)

---

**Input:**  $i, \mathbf{g}, \mathcal{G}$

**Output:**  $e$

$e = \emptyset$

$r = |\mathbf{g}|$

**while**  $r > 0$  **do**

$$\left[ \begin{array}{l} w = \left\lfloor i / \left( \prod_{p=r+1}^{|\mathbf{g}|} |\mathcal{G}(\mathbf{g}_p)| \right) \right\rfloor \bmod |\mathcal{G}(\mathbf{g}_r)| \\ v = \mathcal{G}(\mathbf{g}_r)_w \\ e \leftarrow e \cup v \\ r \leftarrow r - 1 \end{array} \right.$$

**return**  $e$

---

The algorithm for sampling from the  $m$ -HSBM is given in Algorithm 2.

---

**Algorithm 2:** Generating the  $m$ -uniform stochastic block model for hypergraphs ( $m$ -HSBM)

---

**Input:**  $N, m, P, G, \mathcal{G}$

**Output:**  $E$

$B = 1, \dots, G$

$E = \emptyset$

**for**  $\mathbf{g} \in B \times \dots \times B$  **do**

$\mathcal{N} = (|\mathcal{G}(\mathbf{g}_1)|, \dots, |\mathcal{G}(\mathbf{g}_m)|)$

$M_{max} = \prod_{g \in \mathbf{g}} |\mathcal{G}(g)|$

$p = P_{\mathbf{g}_1, \dots, \mathbf{g}_m}$

$i \sim \text{Geometric}_1(p)$

**while**  $i < M_{max}$  **do**

$e = \text{IndexToEdge}(i, \mathbf{g}, \mathcal{G})$

$E \leftarrow E \cup e$

$i \leftarrow i + \text{Geometric}_1(p)$

**return**  $E$

---

## Appendix D

### Appendix for Chapter 7

#### D.1 Full datasets

In this section, we show that our results hold for the full datasets, not just the largest component of the multiplexed data.

In contrast to Figures 2b and 3b in Chapter 7, the multiplexed data set is no longer fully connected, leading to some epidemic trajectories that reach very few nodes, resulting in the bimodal distribution of epidemic extents. For the Project 90 dataset, we see in Fig. D.1a that the relative epidemic extents are preserved when compared with the epidemic extents of the largest connected component of the multiplexed data. This is not the case with the JOAPP data set in Fig. D.1b, but as discussed in the Chapter 7, this is to be expected due to the two competing factors that determine the epidemic extent for the threshold contagion process.

#### D.2 Larger parameter ranges

In this section, we simulate each contagion process for a wider range of parameters to verify that our results are not dependent on the choice of parameters.

For the SI model on the Project 90 data set in Fig. D.2a, we notice the same differences in epidemic extents, albeit on different time scales. This should be anticipated; if we plot the epidemic extent with respect to  $\beta t$ , the average time series will be roughly identical. For the threshold process on the JOAPP data set in Fig. D.2b, there are parameter values where the epidemic extents are trivially the same. First, if we choose a threshold greater than the maximum possible fraction of

infected neighbors that the network structure and number of seed nodes allow, then contagion will never occur. Second, if the threshold is low enough and the multiplex and uniplex representations are all fully connected, then the entire population will be infected no matter the data representation.

### D.3 Different number of seeds

In this section, we show that our results hold for 2, 5, and 10 seeds as well as a single seed node as presented in Chapter 7.

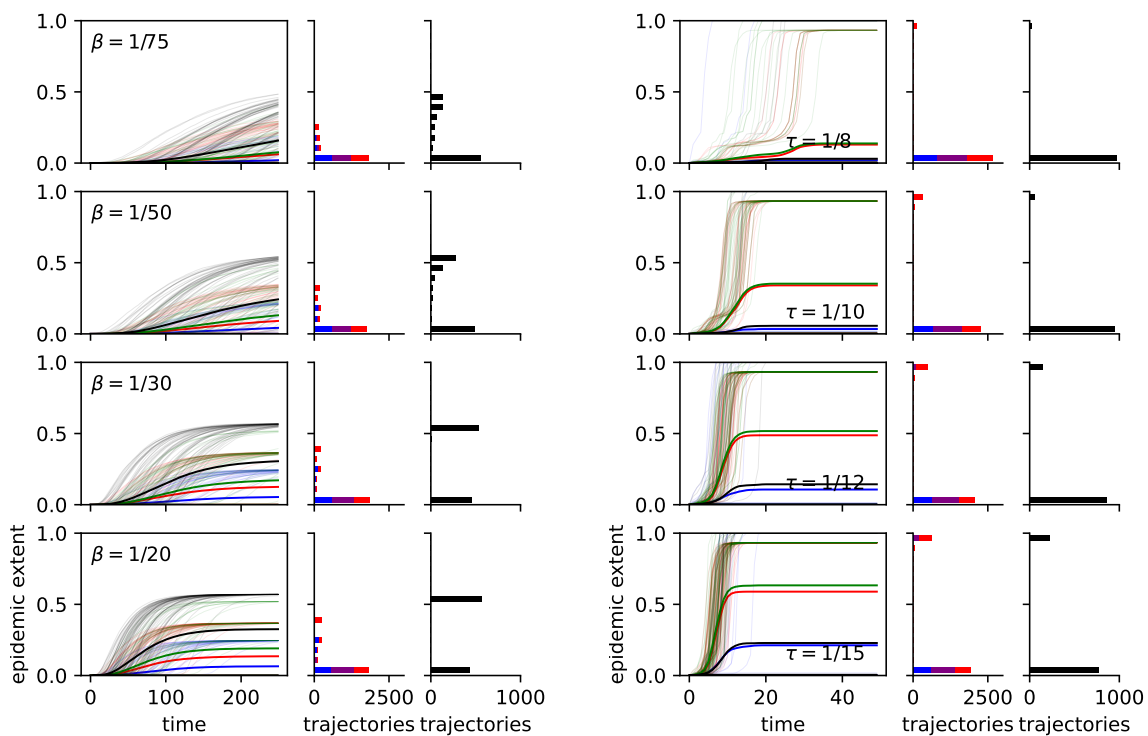
In Fig. D.3, we see that the relative differences in the epidemic extents are preserved. The largest differences are than with a larger number of seed nodes, the likelihood that there is an epidemic trajectory that spreads to very few nodes is much smaller as can be seen in the figure.

In Fig. D.4, we see that, like in Fig. D.3, there is a smaller chance of trajectories dying out. In addition, we see that increasing the number of seed nodes effectively raises the maximum threshold for which the contagion will spread to the entire network (for example,  $\tau = 1/10$  in Fig. D.4) and lead to trivial results as discussed prior.

### D.4 Full temporal extent

This section, we run the simulation for a long enough time to remove any temporal censoring for larger values of  $\tau$  for the threshold model and smaller values of  $\beta$  for the SI model.

In Fig. D.5, we see that the epidemic extents are consistent with our results in Chapter 7. For the SI model, this should be expected as explained prior; rescaling time by the infection probability should yield very similar epidemic responses in expectation.

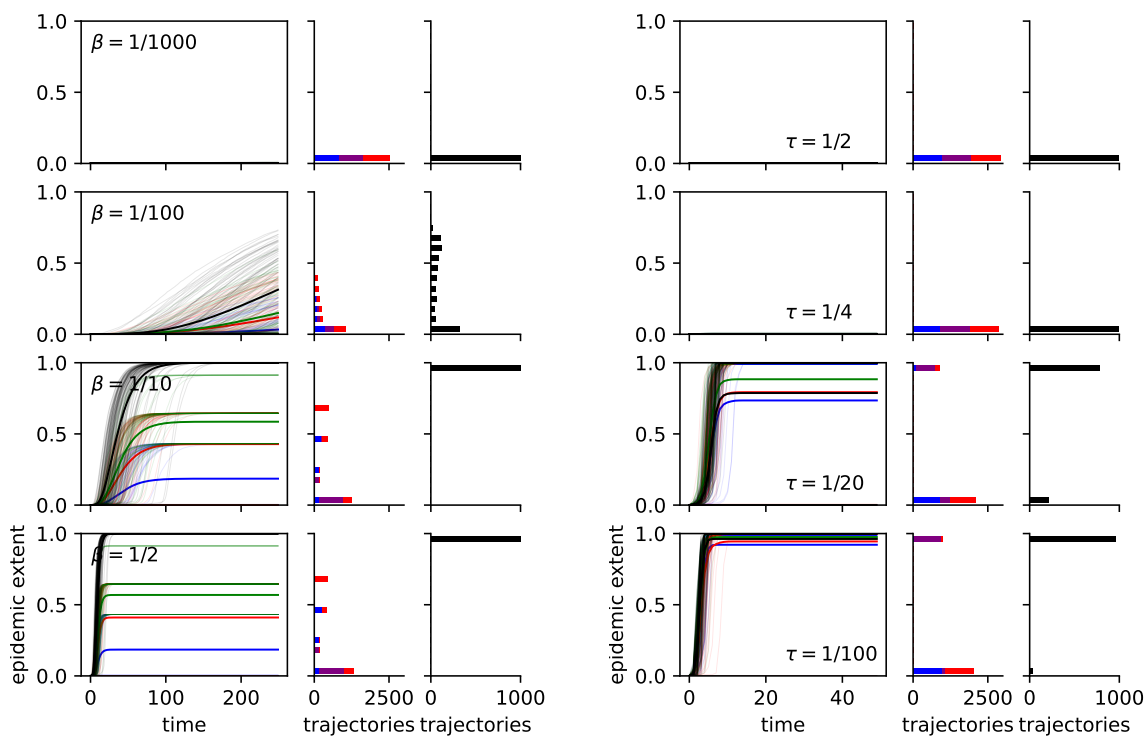


(a) Full Project 90 dataset.

(b) Full JOAPP dataset.

Figure D.1: Considering the full Project 90 and JOAPP data sets. For more details on these plots, see Figs. 7.2b and 7.3b.





(a) Project 90 data set with a wider range of parameters. (b) JOAPP data set with a wider range of parameters.

Figure D.2: Considering a wider range of contagion parameters. For more details on these plots, see Figs. 7.2b and 7.3b.

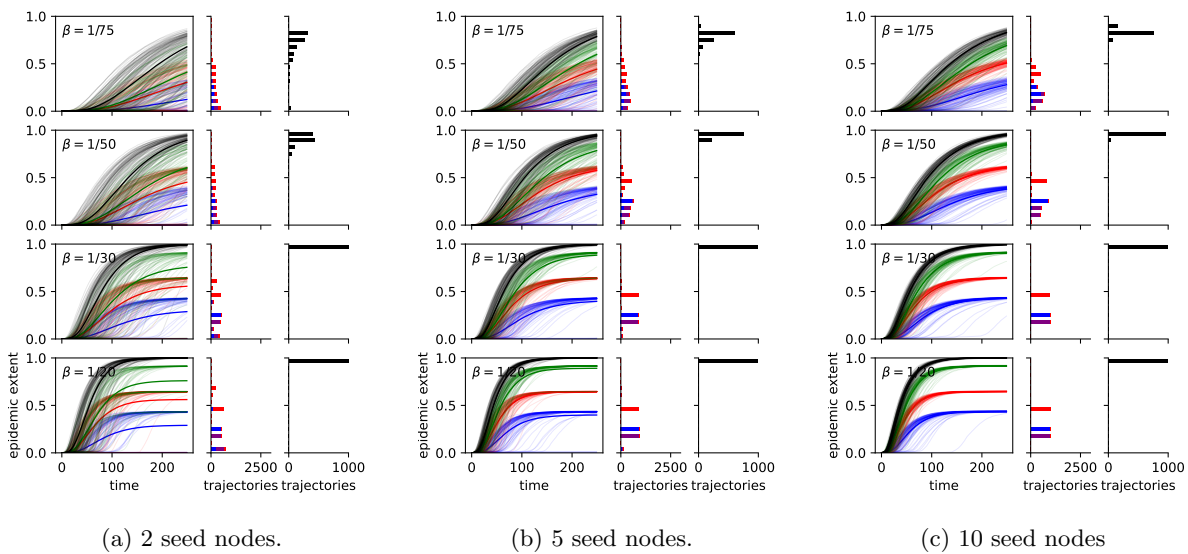


Figure D.3: Varying the number of seed nodes for the SI model on the Project 90 data. For more details on these plots, see Figs. 7.2b and 7.3b.

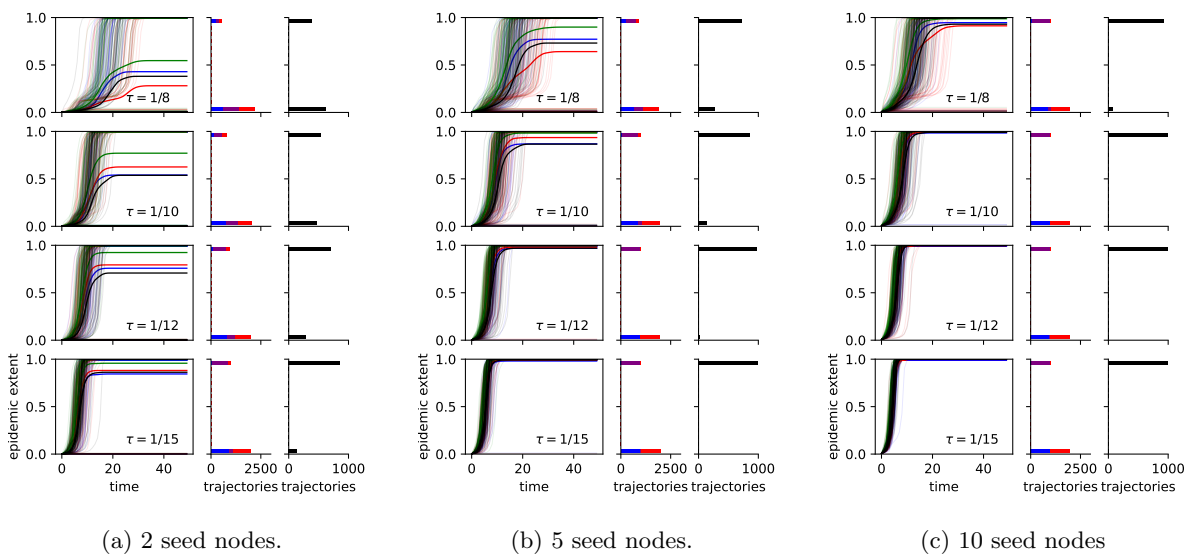
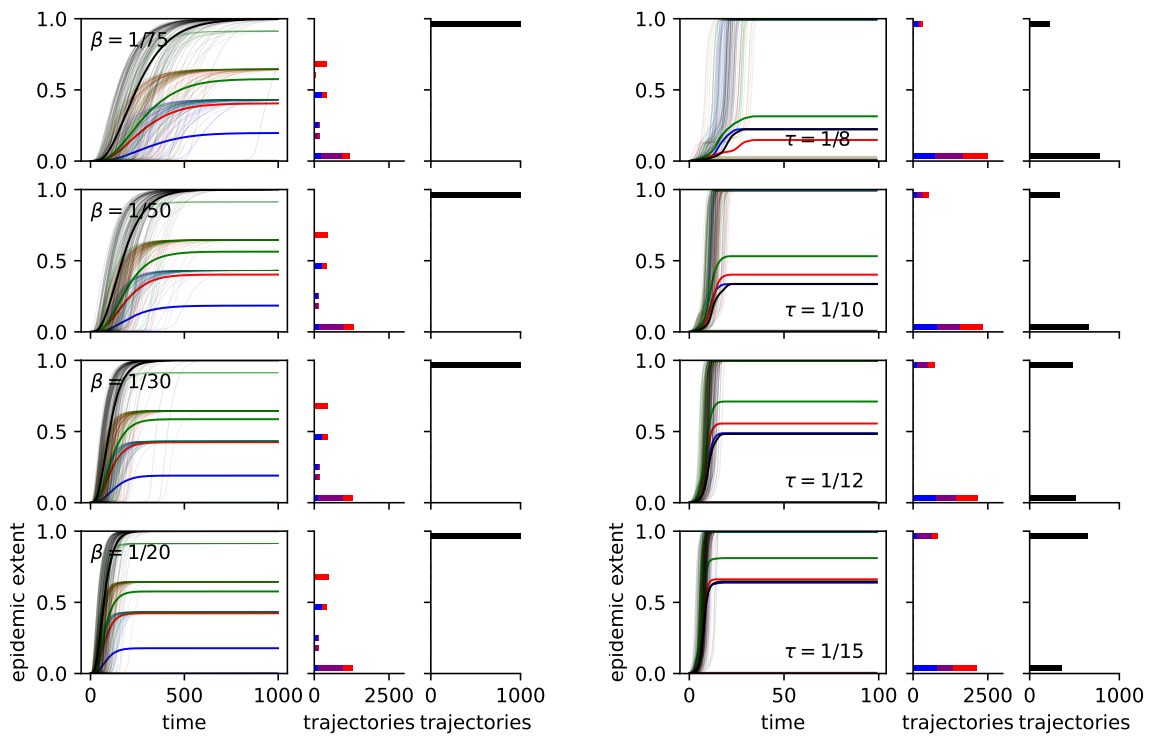


Figure D.4: Varying the number of seed nodes for the threshold model on the JOAPP data. For more details on these plots, see Figs. 7.2b and 7.3b.



(a) Project 90 dataset.

(b) JOAPP dataset.

Figure D.5: Simulating contagion models until equilibrium is reached for all parameter values. For more details on these plots, see Figs. 7.2b and 7.3b.