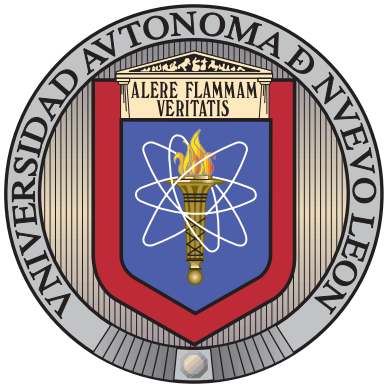
Universidad Autónoma de Nuevo León

Facultad de Ingeniería Mecánica y Eléctrica

Subdirección de Estudios de Posgrado



Simplicial models for higher order interactions in complex networks

por

Gerardo Palafox Castillo

como requisito parcial para obtener el grado de

doctorado en ingeniería de sistemas

Mayo 2024

Universidad Autónoma de Nuevo León

Facultad de Ingeniería Mecánica y Eléctrica

Subdirección de Estudios de Posgrado

Simplicial models for higher order interactions in complex networks

por

Gerardo Palafox Castillo

como requisito parcial para obtener el grado de

doctorado en ingeniería de sistemas

Mayo 2024

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
Facultad de Ingeniería Mecánica y Eléctrica
Posgrado

Los miembros del Comité de Evaluación de Tesis recomendamos que la Tesis "Simplicial models for higher order interactions in complex networks", realizada por el estudiante Gerardo Palafox Castillo, con número de matrícula 1649275, sea aceptada para su defensa como requisito parcial para obtener el grado de Doctorado en Ingeniería de Sistemas.

El Comité de Evaluación de Tesis

José Arturo Berrones Santos
Director

James Melbourne
Co-director

Romeo Sánchez Nigenda
Revisor

Sara Elena Garza Villarreal
Revisor

César Emilio Villarreal Rodríguez
Revisor

Francisco Javier Almaguer Martínez
Revisor

Lilia Alanís López
Revisor

Vo.Bo.

Dr. Simón Martínez Martínez
Subdirector de Estudios de Posgrado

Institución 190001

Programa 557620

Acta Núm. 346

Ciudad Universitaria, a 27 de junio de 2024.

A mis padres, y a Fabiola.

Contents

List of Figures

Acknowledgments

I thank my parents and my wife for the support throughout this and all my previous academic endeavors. I thank my advisors, James and Arturo, for taking an interest in my work and allowing me the academic freedom to explore my ideas with their help and support. I thank CONAHCyT for the funding which made this possible.

Resumen

Gerardo Palafox Castillo.

Candidato para obtener el grado de Doctorado en Ingeniería de Sistemas .

Universidad Autónoma de Nuevo León.

Facultad de Ingeniería Mecánica y Eléctrica.

Título del estudio: Simplicial models for higher order interactions in complex networks.

Número de páginas: 43.

Objetivos y método de estudio: The goal of the present work is to identify interacting systems for which a simplicial complex representation is better suited than the traditional network approach and to prove there is a significant improvement in the results by using higher order models. Computer simulations are used to perform numerical experiments on processes of interest, verifying the results via analytic techniques such as mean field approximations. When the model is too complicated to be studied with analytic methods, statistical tests are performed to give robustness to our simulation results.

Contribuciones y conclusiones: Novel models of processes traditionally defined on networks are studied on simplicial complexes, which show a rich structure to incorporate higher order interactions in complex systems. Results presented show this is the case at least for the known Markovian SIR epidemic model in networks, which when generalized to simplicial complexes exhibits a distinct behavior than its pairwise counterpart.

The following is a list of publications and talks product of the doctoral work surrounding this thesis.

Peer reviewed publications

(Journal, JCR IF 3.3) Palafox-Castillo, G., & Berrones-Santos, A. (2022). Stochastic epidemic model on a simplicial complex. *Physica A: Statistical Mechanics and its Applications, 606,* 128053. DOI: 10.1016/j.physa.2022.128053

(Journal, JCR IF 1.8) Melbourne, J., & Palafox-Castillo, G. (2023). A discrete complement of Lyapunov's inequality and its information theoretic consequences. *The Annals of Applied Probability, 33(6A),* 4322-4340. DOI: 10.1214/22-AAP1919

(Conference) Palafox-Castillo, G. (2021). Simplicial models for higher order interactions in complex networks. *The 10th International Conference on Complex Networks & Their Applications November 30 - December 2, 2021. Book of Abstracts.* ISBN: 978-2-9557050-5-6

Oral presentations

(Poster) "Epidemias estocásticas en grafos" (awarded best poster) at *XIX Escuela de Probabilidad y Estadística,* Centro de investigación en Matemáticas, April 2021.

(Poster) "Simplicial models for higher order interactions in complex networks" at *VII Mexican Workshop on Geometric and Topological Data Analysis*, Centro de investigación en Matemáticas, September 2021.

(Talk) "Simplicial models for higher order interactions in complex networks" at *The 10th International Conference on Complex Networks & Their Applications*, November 2021

(Talk) "Un modelo de contagios en un complejo simplicial" at *Seminario del Posgrado en Ingeniería de Sistemas*, FIME UANL, April 2022

(Poster) "Un revés de la desigualdad de Lyapunov para sucesiones log-cóncavas" (awarded best poster) at *XX Escuela de Probabilidad y Estadística*, Centro de investigación en Matemáticas, April 2022

(Talk) "A stochastic epidemic on a simplicial complex" at *SIAM Workshop on Network Science*, September 2022

(Talk) "Un modelo de contagios en un complejo simplicial" at *Seminario del Posgrado en Ciencias con Orientación en Matemáticas*, FCFM UANL, November 2022

(Talk) "Un modelo de contagios en un complejo simplicial" at *Seminario de Estudiantes*, Centro de Investigación en Matemáticas, August 2023

Firma de los asesores

Dr. José Arturo Berrones Santos
Director

Dr. James C. Melbourne
Co-director

Chapter 1

Introduction

This work is about the behavior of interacting elements in complex systems, and how distinct mathematical objects help us in bettering our understanding of said behavior. By a complex system, one should understand a system with many components. Examples of these abound: crowds, corporations, cities, power grids, financial markets, among others, can all be considered complex systems with many interacting components. For example, if one wishes to study traffic congestion in cities, one has to consider components such as roads with private and public vehicles, traffic lights, public transport outside of roads (e.g. trains), to name a few. Early studies considered elements of the system as particles freely interacting with each other in an environment. For example, when studying the spread of an infectious disease in a population, it was assumed any member of the population could interact (and hence infect) any other member. To loosen this assumption, researchers started modeling systems as *graphs* or *networks*. Informally, a graph or a network consists of points (called *nodes* or *vertices*), some of which are linked by a line (called an edge). This has a natural interpretation in the case of complex systems: elements of a system can be considered vertices, and if two elements interact with each other, one can join the corresponding vertices with an edge. For example, Zachary [33] studied a university karate club, where each member is represented as a vertex and there is an edge between members if they interacted outside the club. This graph can be

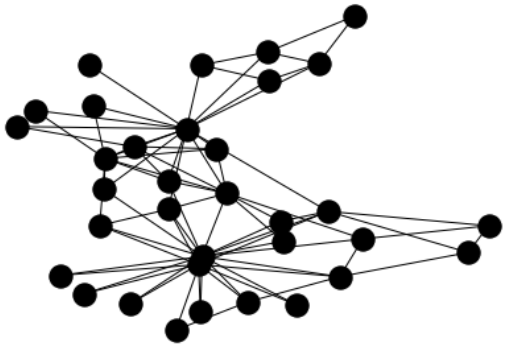


Figure 1.1: Zachary's karate club graph

seen in Figure 1.1. In the context of disease spreading, it is reasonable to assume members of a population (e.g. a city, a school, a workplace) do not interact all with each other, but instead they interact with a subset of the whole population (e.g. neighbors, classmates). With this in mind, a graph can be constructed in a way that each vertex is a member of a population and its joined by an edge with the vertices corresponding to people they interact with. The interaction that gives rise to the existence of a link can be defined differently according to the nature of the infectious disease. For example, for sexually transmitted diseases, two individuals would be joined by an edge if there was a sexual encounter between them. For viruses such as noroviruses, people in the same household would make more sense to link together, while for airborne diseases, any two individuals being face to face should be considered to have interacted. In this context, a disease can only spread from one person (vertex) to another by an edge. While graphs have been studied by mathematicians since the 18th century, it was not until the late 20th century when scientists started to use them in the modelling of complex phenomena.

1.1 Higher order networks

The definition of graphs imposes a limitation on the modeling of interactions: since edges only join two vertices, the interactions captured are only those that occur

pairwise. Consider for example a graph that models academic collaboration, where each vertex corresponds to a researcher and two vertices are joined by an edge if the researchers have published a paper together. If three authors have never co-authored a paper together, but any two of them have co-authored a paper, the corresponding collaboration graph will show a triangle joining these vertices, but one will not be able to discern that there exists no publication where the three of them appear. Another example comes from epidemics. The interaction that occurs between three coworkers does not have the same frequency or intensity as the interaction that occurs between three roommates or members of a family who live together. Thus one would like to have a mathematical object akin to graphs which allow to distinguish between two-types of interactions of more than two elements in a system. A way to remedy this is with *hypergraphs*; if one considers edges in graphs as sets of containing two vertices (the vertices the edge joins), one can then generalize the definition of graph to an hypergraph, which consists of a vertex set and "hyperedges" that can join two or more vertices (instead of only two). An example of an hypergraph can be seen in Figure 1.2. Hypergraphs have been studied in mathematics since the 1980s [4], and have been incorporated to the study of complex systems in more recent times. A particular case of hypergraphs are *simplicial complexes*. Simplicial complexes are just hypergraphs where if e is an hyperedge and a set of vertices e' is contained in e, then e' is also an hyperedge. The example in Figure 1.2 is *not* a simplicial complex since $\{v_1, v_2\}$ is a subset of the hyperedge e_1 but it is not itself an hyperedge. Despite simplicial complexes being a special case of hypergraphs, they are mathematical objects studied in their own right with techniques very different from those used for hypergraphs. Simplicial complexes have been studied since the 1930s [20] by mathematicians, and the results and tools to study these is more robust than for general hypergraphs.

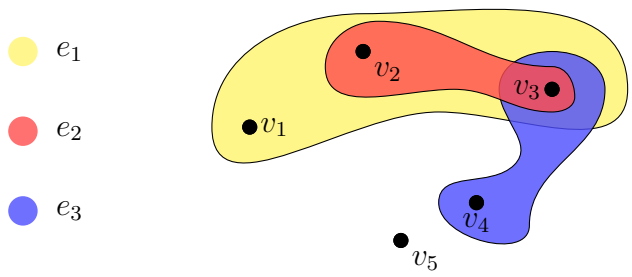


Figure 1.2: Hypergraph with five vertices and three hyperedges.

1.2 Mathematical epidemiology

Of particular interest to us is the study of contagion phenomena in populations. Here, one considers a population where an infectious disease spreads. A typical set up is a *compartmental model*, where members of the population are categorized in mutually exclusive categories (compartments) pertaining their status relative to the disease. For instance, if one only considers the possibility of being susceptible, infectious or recovered (as in diseases with permanent immunity) then one would put every member of the population in one of these compartments. This is called an SIR model, and has different variants which will be discussed in this work. Other popular compartmental models are SI (only susceptible and infectious), SIS (susceptible and infectious, where there is no immunity; infectious individuals turn susceptible immediately after recovering from the disease), SIRS (susceptible, infectious, recovered, susceptible; temporal immunity). Accounting for more compartments tends to make a model more realistic but more complex, and it is not always necessary. For example, consider a disease which gives recovered individuals temporal immunity of about a year. A natural compartmental model to consider in this situation would be the SIRS. However, if one is modelling the transmission of this disease in a smaller time scale (say, one or two months), then the infection will behave as if the immunity was permanent and the simpler SIR model would suffice. The spread of the disease is done via contact of infectious member with susceptible members, so mathematically this processes are akin to the study of population dynamics or

interacting particle systems. Notice that this allows for the study of different phenomena through the same type of models, namely, the spread of information (news, rumors, memes), the spread of computer virus, etc. Commonly, the population is considered to be a finite, fixed number N, and the object of study is to see how the proportions of individuals in each compartment evolves over time; these proportions are usually referred to as densities. It is worth mentioning that while not studied in this present work, models exist where the population is dynamic and not fixed; here, one considers births, deaths and/or migration.

Early studies of the spreading of infectious diseases considered the population to be homogeneously mixed; that is, any infectious member could contact (hence infect) any susceptible members. As mentioned in the preceding section, sometimes it is desirable to model the population in such a way that an infectious individual is limited in who they can infect. This was first achieved by considering epidemics on graphs. To exemplify, a SIR model on a given graph would start with an infectious vertex and the rest susceptible. As time moves forward, an infectious vertex will infect its susceptible *neighbors*, meaning susceptible vertices joined to the infectious vertex by an edge. Infectious vertices will eventually transition to a recovered state, where they will remain for the rest of the process. Eventually there will be no infectious vertices remaining and the epidemic will end by virtue of infectious recovering on their own and there being no re-infections. At the end of the epidemic process, one has a number of recovered individuals, all of whom where infectious at some point. Thus it is common to use this number, called the *final size* of the epidemic, to study the severity of the disease. In the early models with homogeneously mixed population, the only parameters that affected the spread of the disease where the *transmission rate* (how infectious an infected individual is) and the *recovery rate* (how quickly an infectious individual recovers). Studies usually focus on knowing how the available information (parameters) influence the severity (final size) of the disease. With the introduction of graphs to the modelling of disease transmission, there is more information to be taken into account and which play a role in the

dynamics of the disease transmission, for example the average number of neighbors a vertex has. As a consequence, the mathematics of the same compartmental model (e.g. SIR) is quite different if one considers the population homogeneous vs. population as a graph. However, as previously discussed, graphs have limitations. Namely, graphs only capture pairwise interactions and may fail to e.g. distinguish between the interaction of three individuals pairwise or their simultaneous interaction. This fact leads us to study a contagion process on higher-order networks, particularly on simplicial complexes. This generalization will allow for the distinction of "close contacts", modeled by simplices of dimension two or higher. Ball et al. [2] study a similar problem but limited to networks where they incorporate a household structure, but each household had a fixed number of members. The incorporation of these sub-structures within a network (or higher order interactions) is of importance in contagion processes because close contacts can have an effect on the probability of transmitting a disease.

1.3 Hypothesis and objectives

Our hypothesis is that many interaction processes which are currently described and studied by means of networks and pairwise interactions admit a richer, more realistic study via a simplicial complex representation. Current network epidemics may not capture social interactions in the way higher order networks can. The goal of the present work is to present a contact process for which a simplicial complex representation is better suited than the traditional network approach and to prove there is a significant improvement in the results by using higher order models. Some reasons behind using simplicial complexes and not hypergraphs is the possibility of using the tools applied algebraic topology and topological data analysis may offer to better understand the subjacent structure.

1.4 Problem description

The study of epidemics in simplicial complexes has so far been limited to the compartmental model SIS, where individuals are considered either susceptible or infectious, and the dynamics are limited to individuals going from susceptible to infectious and viceversa. Indeed, Iacopini et al. [15] opened the way by introducing a SIS contagion process on a simplicial complex, further studied by Cisneros-Velarde and Bullo [10] and Matamalas et al. [23]. The existent SIR variant in the literature at the time our project started was the work of Ma and Guo [21], who study information transmission on hypergraphs, but used an infection rate independent of the dimension. In this work we develop an SIR model on simplicial complexes, where individuals are either susceptible, infectious or removed (either recovered and immune, or dead) and the dynamics consist of going from susceptible to infectious and from infectious to recovered. Specifically, we generalize the known Markovian stochastic SIR epidemic on a network [8]. These models are useful not only because of the existence of diseases with permanent immunity, but for the modeling early stages of diseases with waning immunity (e.g. influenza, covid-19). Here, we show dynamics of the SIR epidemic on the simplicial complex that are not observed in the network equivalent nor in the SIS epidemic on the simplicial complex.

Chapter 2

Background

To make the present work self contained, this chapter presents the mathematical prerequisites suggested for the reading of the main contributions. While the exposition is original, the material is well known, and can be found in the literature.

2.1 Graph theory

A graph, in the most simple terms, consists of a collection of objects, some of which are paired together. The study of graph theory dates back to Euler, who studied ways of walking through the city of Königsberg, a city separated by a river and connected by bridges, crossing each bridge once. In recent times, graph theory has served as a theoretical foundation for the emerging scientific branch of *network science.* [25]. Network science concerns itself with the study of real-world complex systems which can be modeled by graphs, arising from fields such as biology, engineering, neuroscience, and more. The advent of big data and computers has made this an active area of research, where characterizing the structure, dynamics and statistical properties of large, real-world systems is the main goal. In contrast, graph theory is an area of pure mathematics which aims to prove rigorous results of graphs in greater generality. While results in graph theory can find their way to applications in network science and other areas (e.g. network optimization), this is not the

objective of their practitioners.

A modern treatment of theory of graphs can be found in the book of Bollobás [7]. A book which emphasizes the network scientific approach is the one by Newman [25]. In this section, a brief mathematical account of the concepts needed for our work is given.

A *graph* is a tuple $G = (V, E)$ consisting of a set of vertices V (also called nodes) and a set of edges E joining distinct pairs of vertices. Edges are usually represented as unordered pairs (i.e., two-element sets) $\{u, v\}$ or as a commutative juxtaposition uv, for vertices $u, v \in V$, called the endpoints of the edge. An edge is said to be incident on its endpoints. Graphically, vertices are represented as points, and edges as lines joining these points, e.g. see Figures 2.1 2.2 2.3.

Let $G = (V, E)$ be a graph. Two vertices $u, v \in V$ are said to be *adjacent* or *neighbors* if $uv \in E$. The *neighborhood* of a vertex v, denoted by $N(v)$, is the set of vertices adjacent to v, that is

$$N(v) = \{u \in V : uv \in E\}. \tag{2.1}$$

The degree of a vertex $v \in G$, denoted by $\deg v$, is the number of neighbors v has, i.e. $\deg v := |N(v)|$, where $|S|$ denotes the cardinality (number of elements) of a set S. The degree distribution of G is the discrete probability distribution on $\{0, 1, \ldots, \max_{v \in V} \deg v\}$ such that

$$\mathbb{P}(X = k) = \frac{|\{v \in V : \deg v = k\}|}{|V|}. \tag{2.2}$$

The average degree of a graph $\mathbb{E}[X]$ is commonly denoted by $\langle k \rangle$. A graph $H = (V', E')$ is a subgraph of G if both $V' \subseteq V$ and $E' \subseteq E$. If $S \subseteq V$, the induced subgraph $G[S]$ of G is the graph with vertex set S where the edges are those edges in E which have both endpoints in S. A *path* in G is a sequence of adjacent, distinct vertices $(v_0, v_1, \ldots, v_n)$, where the path is said to connect v_0 and v_n and to be of length n. If any two vertices in a graph are connected by a path, the graph is said to be *connected*. If the graph is not connected, the connected subgraphs that are not

contained in a larger connected subgraph are called *connected componentes*. The *distance* between two vertices u, v in a connected graph, denoted by $d(u, v)$, is the length of the shortest path connecting them. The *diameter* of a connected graph is the largest of said distances, meaning

$$\operatorname{diam} G := \max_{u,v \in E} d(u, v). \tag{2.3}$$

Example 1 (Complete graph). *The complete graph K_n is a graph with n vertices where all of them are adjacent. See Figure 2.1 for a complete graph of six vertices.*

Example 2 (Clique). *A clique of n vertices in a graph G is a complete subgraph K_n of G.*

Example 3 (Erdős-Rényi graph). *An important example of graph is the Erdős-Rényi graph $G(n, p)$, a model of random graphs. The graph $G(n, p)$ has a vertex set $V = \{1, \ldots, n\}$, with each possible edge appearing with a probability p.*

Example 4. *Let $G = (V, E)$ with $V = \{2, \ldots, 10\}$ and $\{a, b\} \in E$ if and only if $a, b, a \neq b$, have common divisors; i.e., if they are distinct and their greatest common divisor is not one. The graph has two connected components. One can see that $N(3) = \{6, 9\}$ so $\deg(3) = 2$; on the other hand, $\deg(7) = 0$. The subgraphs $G[2, 4, 6, 8, 10]$ and $G[3, 6, 9]$ are cliques of 5 and 3 vertices respectively. The distance between 5 and 9 is $d(5, 9) = 3$, obtained from the path $(5, 10, 6, 9)$. One also has $\operatorname{diam} H = 3$, where H is G's largest connected component. See this graph in Figure 2.2a.*

2.1.1 Systems as graphs

Graphs are useful in capturing the pairwise relations in a given set of objects, by considering them vertices joined by an edge whenever an interaction of some sort occurs. For example, in a city, roads can be seen as the edges of a graph, with intersections corresponding to vertices. Similarly, a power grid can be considered as

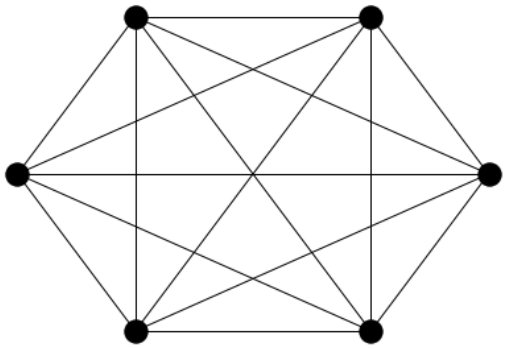
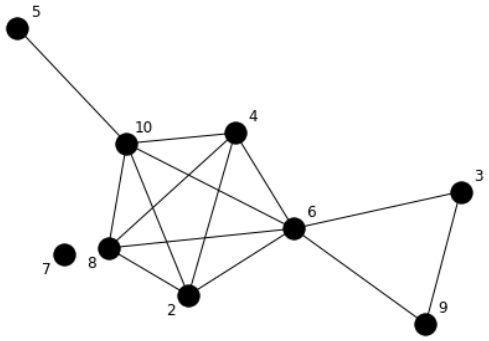
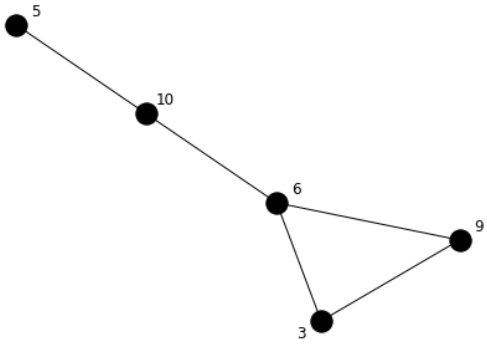


Figure 2.1: Complete graph K_6 of six vertices.



(a) Graph with vertex set $\{2, 3, \ldots, 10\}$ where distinct vertices are joined by an edge if they have common divisors, as described in Example 4.



(b) Subgraph of the graph in Figure 2.2a induced by the vertx subset $\{5, 10, 3, 6, 9\}$.

Figure 2.2: Graph from Example 4 and an induced subgraph.

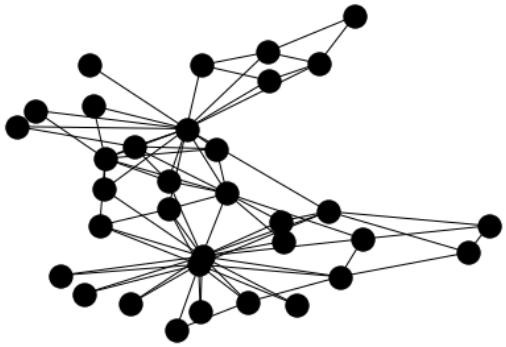


Figure 2.3: Zachary's karate club graph

a graph, with the power generating stations corresponding to vertices and voltage lines being the edges joining these. Distribution systems of natural gas or water can also be represented by graphs. A particular example in the context of spreading processes is that of social networks. Social networks are graphs where each vertex represents a person, and edges correspond to whether there exists an interaction between two given persons. Social networks can be virtual, e.g. Facebook users as vertices and an edge joining Facebook friends, or physical, e.g. a vertex for each person in a school and an edge joining them if they share a class. Similar examples under different rules can be thought for workplaces, sport clubs, etc. For example, Zachary [33] studied a university karate club, where each member is represented as a vertex and there is an edge between members if they interacted outside the club. This graph can be seen in Figure 2.3.

2.2 Higher order models

It is clear from the definition of a graph that they capture only pairwise interactions. Because of this, other combinatorial structures have been used as generalizations of graphs that capture relations of two or more objects. These are known in the literature as *higher order networks*. The two most common generalizations in use are *hypergraphs* and *simplicial complexes*.

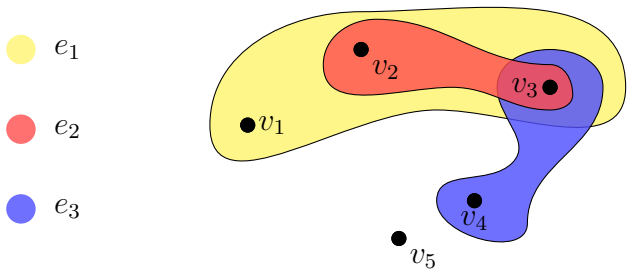


Figure 2.4: Hypergraph with five vertices and three hyperedges.

2.2.1 Hypergraphs

A hypergraph $\mathcal{H}$ is a tuple $\mathcal{H} = (V, E)$ where V is a set of vertices and E is a set of *hyperedges*. Each hyperedge is a subset of V with cardinality greater or equal than two. Notice that if we restrict hyperedges to have cardinality two, we would recover the definition of a graph.

Example 5. *Suppose a university is tracking the collaboration among five members of a department by having a vertex for each researcher and having them be in the same hyperedge if they have coauthored a publication. Write $v_1, \ldots, v_5$ for each of the five researchers, and suppose they have collaborated in three publications: one coauthored by researchers v_1, v_2, v_3; another one coauthored by researchers v_2, v_3; and lastly one coauthored by researchers v_3, v_4. Researcher v_5 has not collaborated in any of these. The resulting collaboration hypergraph would be $V = \{v_1, \ldots, v_5\}$ and $E = \{e_1 = \{v_1, v_2, v_3\}, e_2 = \{v_2, v_3\}, e_3 = \{v_3, v_4\}\}$. See Figure 2.4. Notice that, had this been carried out in a graph, vertices v_1, v_2, v_3 would have formed a clique, but there would be no way to tell that v_1, v_3 have not coauthored a publication on their own without v_2.*

2.2.2 Simplicial complexes

There are different ways of defining a simplicial complex. Some of them are based in geometry, as subsets of Euclidean space, see for example the book by Rotman [26]. For our purposes, it suffices to consider *abstract simplicial complexes*.

Definition 1. *An abstract simplicial complex $\mathcal{K}$ with vertex set V is a collection of subsets of V such that:*

1. *For all $v \in V$, $\{v\} \in \mathcal{K}$;*
2. *if $\sigma \in \mathcal{K}$ and $\tau \subseteq \sigma$ then $\tau \in \mathcal{K}$.*

Each set in the simplicial complex $\mathcal{K}$ is called a simplex. A simplex with $k+1$ elements is called a k-simplex, or it is said to have dimension k. The dimension of the simplicial complex is the maximum dimension of its elements, that is,

$$\dim \mathcal{K} = \max_{\sigma \in \mathcal{K}} (\dim \sigma) = \max_{\sigma \in \mathcal{K}} (|\sigma| - 1) . \tag{2.4}$$

Observe that there is a correspondence between graphs and one-dimensional simplicial complexes. Similarly between simplicial complexes and hypergraphs with hyperedges closed under taking subsets (i.e., if $e' \subseteq e$ for some hyperedge e, then e' is a hyperedge). Thus we have

$$\text{Graphs} \subseteq \text{Simplicial Complexes} \subseteq \text{Hypergraphs}.$$

Observe however that any simplicial complex of dimension higher than one is not a graph, and that not all hypergraphs are simplicial complexes (for example, the hypergraph of Example 5 is not a simplicial complex because $\{v_1, v_3\}$ is not a hyperedge but $\{v_1, v_3\} \subseteq e_1$.

Example 6. *In a given neighborhood, people can be represented by vertices, and households as simplices. Interactions between people of different households can be*

seen as edges (one dimensional simplices). Clearly if a set of people belongs to a household, any subset belong to the same household.

EXAMPLE 7. *Let us revisit Example 4. Given a vertex set $V = \{2, 3, \ldots, 10\}$, consider $\mathcal{K}$ to consist of singletons $\{v\}$ for all $v \in V$, and also, sets of numbers with divisors in common. Clearly if a set of numbers σ share divisors, any subset $\sigma^* \subseteq \sigma$ will consist of numbers that also share divisors. Thus having a simplicial complex of dimension 4. This is a particular example of what is called a clique complex, defined in Example 8.*

EXAMPLE 8 (Clique complex). *Let $G = (V, E)$ be a graph. The clique complex of G is the simplicial complex*

$$\mathcal{K} = \{\sigma \subseteq V : \sigma \text{ is a clique in } G\}.$$

Notice that single vertices and pairs of vertices joined by an edge are cliques of one and two elements, respectively. This is one of the simplest ways of associating a simplicial complex to a graph. Observe that the simplicial complex in Example 7 is the clique complex of the graph of Example 4.

EXAMPLE 9 (d-dimensional clique complex). *Sometimes it is useful to associate a simplicial complex to a graph but limiting its dimension. One way to do so is the following. Let $G = (V, E)$ be a graph. The d-dimensional clique complex of G is the simplicial complex*

$$\mathcal{K} = \bigcup_{k=1}^{d+1} \{\sigma \subseteq V : \sigma \text{ is a clique in } G \text{ with } k \text{ vertices}\}.$$

To any simplicial complex $\mathcal{K}$ one can associate a graph $G = (V, E)$, called its 1-skeleton, where V is the same vertex set from $\mathcal{K}$ and

$$E = \{\sigma \in \mathcal{K} : |\sigma| = 2\},$$

i.e., the edges are the simplices in $\mathcal{K}$ of dimension one. A simplicial complex is connected if its 1-skeleton is connected. Let v be a vertex in a simplicial complex

$\mathcal{K}$. The n-dimensional degree of v is

$$\deg_n v = \big|\{\sigma \in \mathcal{K} : v \in \sigma \text{ and } \dim \sigma = n\}\big|, \tag{2.5}$$

and the average n-dimensional degree is denoted by $\langle k_n \rangle$. Notice that the one-dimensional degree of v coincides with the graph-theoretic definition of vertex degree, since the number of neighbors a vertex has is the same as the number of edges the vertex belongs to, i.e., the number of 1-simplices it belongs to.

2.3 Random models

In the absence of a true graph or simplicial complex representation of a system, it is convenient to consider random models of these structures. We repeat here the definition of the Erdős-Rényi random graph model, defined in Example 3, along with other common random graph and random simplicial complex models. Further material can be found in the works of Newman [25] and Kahle [19].

Definition 2 (Erdős-Rényi model). *The Erdős-Rényi graph $G(n, p)$ is a random graph with vertex set $V = \{1, \ldots, n\}$, where each possible edge is included independently with a probability p. A similar, equivalent model denoted by $G(n, m)$ is a random graph with vertex set $V = \{1, \ldots, n\}$ and exactly m edges, placed between randomly chosen pairs of vertices.*

Definition 3 (Configuration model). *The configuration model generates a random graph with a given degree sequence $(k_1, k_2, \ldots, k_N)$, i.e., an ordered list of degrees each vertex will have in the resulting graph. Note that the sum of this list, $\sum_i k_i$, must be an even number. In a first step, one has N vertices with no edges. Then, each vertex i is considered to have k_i "half-edges", also called stubs. Two half-edges are chosen uniformly at random, and are connected to form an edge. Do this repeatedly until no more half-edges remain, resulting in a graph with the pre-defined degree sequence.*

Definition 4 (Barabási-Albert model). *This model generates a random graph with a power-law degree distribution $\mathbb{P}(X = k) \approx k^{-3}$. The model is initialized with a*

graph of m_0 vertices, each of which has at least one edge. At each timestep, a vertex is added and connected to $m \leq m_0$ existing vertices. The probability that a newly added vertex connects to a vertex i is

$$p_i = \frac{\deg(i)}{\sum_j \deg(j)}.$$

After t timesteps, the model generates a graph with $t + m_0$ vertices and $m_0 + mt$ edges.

In the random topology literature, the most common models are described by the general *multi-parameter model*.

DEFINITION 5 (Multiparameter model). *Let $p : \mathbb{N} \longrightarrow [0, 1]$. The multiparameter random simplicial complex $X(n; p_1, p_2, \dots)$ is defined as follows. Starting with n vertices, every edge is added with probability $p(1)$. Conditioned on the presence of all three boundary edges, a 2-simplex is added with probability $p(2)$. Conditioned on the presence of all four boundary triangles, a 3-simplex is added with probability $p(3)$, and so on.*

EXAMPLE 10. *The model resulting from considering $p(1) = p(2) = \cdots = p(d-1) = 1, p = p(d), p(k) = 0$ for $k \geq 0$ in the multiparameter model, is known as the random d-complex $Y_d(n, p)$. The random clique complex $X(n, p)$ is recovered from the multiparameter model by making $p = p_1, p(k) = 0$ for $k \geq 0$. The random clique complex is equivalently obtained by constructing an Erdős-Rény graph and then taking its clique complex, as defined in Example 8.*

In his work on social contagion, Iacopini et al. [15] considers a random simplicial complex model distinct from the ones just defined, and it is as follows: starting with an Erdős-Rényi graph $G(n, p_1)$, every three vertex set is added as a 2-simplex with probability p_2. Note that they do not condition on the existence of the edges between the three vertices; if they were not included as edges when the base Erdős-Rényi graph was constructed, they are added when the two simplex is added. In this work, a variation of this model is used: the construction is the same, but we condition on the connectedness of the subjacent $G(n, p_1)$.

2.4 Poisson processes

The basic theory of probability can be consulted in the book of Casella and Berger [9]. The probability of an event A will be written as $\mathbb{P}(A)$. Similarly, the probability of a random variable X belonging to a set A will be written as $\mathbb{P}(X \in A)$. The expectation of a random variable X and its variance are denoted by $\mathbb{E}[X]$ and $\mathrm{Var}(X)$, respectively. A sequence of random variables $X_1, X_2, \ldots$ is said to converge in probability to a random variable X if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0 \text{ or, equivalently, } \lim_{n\to\infty} \mathbb{P}(|X_n - X| < \varepsilon) = 1.$$

A stochastic process is a collection $\{X(t) : t \in J\}$ of random variables, where the index $t \in J$ is a time variable, making $X(t)$ the state of the stochastic process X at time t. If J is countable (i.e., it has as many elements as there are natural numbers, or less) then the stochastic process is said to be discrete; if J is an interval of the real line, then the stochastic process is said to be continuous. We say a random variable X has exponential distribution with parameter $\lambda > 0$ if

$$\mathbb{P}(X \in (a, b)) = \int_a^b f(x; \lambda)\, dx, \tag{2.6}$$

where

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0; \\ 0 & x < 0. \end{cases} \tag{2.7}$$

If a random variable X has exponential distribution with parameter $\lambda > 0$, one will write $X \sim \mathrm{Exp}(\lambda)$. Let $T_1, T_2, \ldots$ be a sequence of independent random variables, with $T_i \sim \mathrm{Exp}(\lambda)$. The *Poisson process* with parameter λ is the continuous time stochastic process $\{X_t : t \geq 0\}$ defined as

$$X_t = \begin{cases} 0 & T_1 > T; \\ \max\{n \geq 1 : T_1 + \cdots + T_n \leq t\} & \text{else.} \end{cases} \tag{2.8}$$

that is, X_t is the largest $n \in \mathbb{N}$ such that the sum $T_1 + \cdots + T_n$ is less or equal to t. It is this interpretation that allows the use of Poisson processes to model the number

of events occurred up to a certain time t under certain assumptions (time between events occurrences is exponentially distributed and independent). Examples of where Poisson processes are used are to model the arrival of customers to a store, of buses to a bus stop, of claims to an insurance company, among others. Of particular interest for this work is the assumption of some epidemic models that contacts between a person in a social network and its neighbors occur according to a Poisson process.

2.5 Contagion processes

Mathematical models of contagion are relevant in diverse contexts, e.g. for understanding the spreading of viruses (both biological and computer), information (news, rumors), among other phenomena. Assumptions must be made about the contacts between the population where the spreading occurs (e.g., computers, people) and how these occur. A population where contacts can occur between any two of its members its called *homogeneously mixed*. When there is some structure limiting the contacts, it is called *heterogeneously mixed* for example, when the population is considered as a graph and contact occurs only between neighboring vertices. Similarly, the population is homogeneous if members are affected equally by the spreading process or heterogeneous otherwise; for example, a population where a disease spreads more among the elderly, is heterogeneous. The population can be fixed or dynamic, depending whether introducing and removing members is allowed.

2.5.1 Compartmental epidemics

In the case of infectious diseases, a common set up consists of partitioning the population in compartments relating to their status with respect to the disease. Among the most used ones, are the following. The SI model considers all individuals are either susceptible or infectious, where the only transition occurring is from suscep-

tible individuals becoming infectious. The SIS model also partitions individuals in susceptible and infectious, but allows an additional transition of infectious individuals going back to a susceptible state. The SIR model partitions the population in susceptible, infectious or recovered, with individuals transitioning from susceptible to infectious and from infectious to recovered. The SIRS model is similar to the SIR, but recovered individuals may go back to a susceptible state. Despite of the biological language employed in these descriptions, models of compartmental epidemics can be used to understand phenomena beyond infectious diseases. For example, computer viruses can spread through interconnected computers following similar dynamics. Information dissemination can also be explored through the lens of compartmental epidemics, where a piece of knowledge or a rumor is considered akin to a virus, and people who ignore it are thought of as susceptible, while people who transmit it are thought of as infectious.

Of special interest in our investigation is the Markovian SIR epidemic on graphs; its definition is as follows.

Definition 6 (Markovian SIR on graphs [8]). *Consider a graph $G = (V, E)$ where its vertices can be either susceptible, infectious or recovered. At time $t = 0$, all vertices are susceptible, and a randomly chosen one is turned infectious. While a vertex is infectious, it has infectious contacts with each susceptible neighbor in the graph, randomly in time according to independent Poisson processes with rate β. Each infectious individual remains so for a period $I \sim \mathrm{Exp}(\gamma)$ after which it recovers and becomes immune. All infectious periods and contact processes are defined independently. The epidemic goes on until the first time T that there are no infectious individuals and the epidemic stops.*

Notice that the recovered state is absorbent; recovered vertices cannot turn neither susceptible nor infectious. Also, since all infectious individuals remain so for a finite duration of time, eventually the number of infectious vertices will be zero. At the time the infectious number of individuals is zero, the compartments of susceptible and recovered individuals will remain static, and thus the epidemic will have come

to an end. The number of recovered individuals when the epidemic stops is known as the final size of the epidemic. An important question to ask when dealing with an epidemic model is: does there exist a threshold value which determines whether a major outbreak is possible? For the model defined in Definition 6, the answer is yes, as the following results show:

Theorem 1 ([8]). *Let R_0 be the expected number of new infections caused by an infected individual at early stages of an epidemic process. Consider an Erdős-Rényi graph, or a Configuration model graph, having degree distribution D with mean μ_D and variance σ_D^2. The basic reproduction number for the Markovian SIR on graphs (Definition 6) is*

$$R_0 = \frac{\beta}{\beta + \gamma}\left(\mu_D + \frac{\sigma_D^2 - \mu_D}{\mu_D}\right).$$

Additionally, the final fraction getting infected during the entire outbreak, $\tau_n = Z/n$, where Z is the final size of the epidemic and n is the population size, converges to zero in probability if and only if $R_0 \leq 1$.

Chapter 3

Literature review

Higher order network models are popular ways to overcome limitations posed by traditional graphs. There are general works covering the basics of these structures. Two excellent surveys on the study of higher-order network models have been done by Battiston et al. [3], Bick et al. [6], both works covering the basic definitions and typical applications, e.g. applications to neuroscience, social systems, brain networks, contagion models, ecology, among many others. The book by Bianconi [5] provides a superb introduction to the area but focusing on simplicial complexes.

The focus of this work is on contagion occurring on simplicial complexes. In the context of infectious disease modelling on higher order networks, most of the work has been done on SIS models. The work of Iacopini et al. [15] is one of the pioneers in studying contagion in higher-order structures. Their work studies a SIS stochastic model on simplicial complexes. Via numerical simulations and a mean-field approximations, phase transitions of the system are discovered which are not present in traditional, network SIS. Furthermore, the simplicial complexes used in the study are both based from real-world networks, and random models.

Similarly, Cisneros-Velarde and Bullo [10] study SIS epidemic model on hypergraphs but focusing on the simplicial complexes case. Authors identify conditions on the parameters under which they can conclude the existence of three regimes:

disease-free, endemic or bistable, this last one being a co-existence of disease-free and endemic equilibrium. Although the results are showed for simplicial complexes of dimension two, the definitions are generalized for higher dimensions and the same techniques can be used to study these.

Jhun et al. [18] study a SIS epidemic on regular hypergraphs, where all hyperedges are the same size. Additionally, their work only considers contagion possible when all but one node in the vertex is infected. For example, a hyperedge consisting of four nodes where two of them are infected will cause no new infections. This does not preclude the possibility of smaller hyperedges being present (as would be the case in a simplicial complex, for example) and resulting in infections.

Ma and Guo [21] study an SIR type of information spreading model on hypergraphs. Their work aims to model the information transmission in informal organizations within an enterprise. The study focuses on a tech enterprise with 62 employees and twenty informal organizations, each considered an hyperedge. Unlike other higher order models, authors use only one infection parameter independent of edges dimension. However, they identify different types of information transmission (one-to-one, one-to-many, etc.).

Outside of the epidemic context, well known stochastic processes have been extended from networks to simplicial complexes. For instance, Schaub et al. [27] study random walks on simplicial complexes. The authors introduced a normalized Hodge Laplacian, which serves as a generalization of the graph Laplacian, and relate these results to the pertinent random walks. Zhao et al. [34] defines a percolation process on simplicial complexes, where removing a node in a 2-simplex causes the rest of the nodes in the simplex to be removed. Their model finds a threshold value for the number of triangles after which simplicial complexes become vulnerable and a phase transition is observed. More generally, Sun and Bianconi [29] study hypergraph robustness via percolation processes. Horstmeyer and Kuehn [13] define a voter model on a simplicial complex, where although the higher-order interaction leaves

the transition to two opinion states intact, it creates significant effects below and above the transition. In the context of dynamical systems, Kuramoto dynamics [24] and consensus models [11] have also been considered on simplicial complexes. Signal processing has been studied in both simplicial complexes and hypergraphs [28]. A summary of the related work can be seen in Table 3.1.

3.1 Opportunity area and expected contributions

From the aforementioned, one readily sees the SIR epidemic model on simplicial complexes is an open problem worth of studying. Our contribution is a suitable model of a stochastic, SIR spreading process on a simplicial complex which is better-suited than networks to represent heterogeneous populations. This model is studied both analytically and through microscopic simulations, the later not commonly seen in the existing literature. Furthermore, we expect our model to be better adjustable to data, and to provide significant analytical and computational characterization of the studied system to provide insights not readily seen in pairwise network models.

Table 3.1: Comparison of related work.

Work	**Model**	**Simplicial complex**	**Hypergraph**
Cisneros-Velarde and Bullo [10]	SIS epidemic	✓	✓
Iacopini et al. [15]	SIS epidemic	✓	×
Jhun et al. [18]	SIS epidemic	×	✓
Wang et al. [31]	Social communication (SIS type)	✓	×
Schaub et al. [27]	Random walk	✓	×
Zhao et al. [34]	Percolation	✓	×
Sun and Bianconi [29]	Percolation	×	✓
Horstmeyer and Kuehn [13]	Voter model	✓	×
DeVille [11]	Consensus model	✓	×
Millán et al. [24]	Kuramoto dynamics	✓	×
Schaub et al. [28]	Signal processing	✓	✓
Ma and Guo [21]	Social communication (SIR type)	×	✓

Chapter 4

Results

The main result of this work is the definition and study of a novel stochastic epidemic model on a simplicial complex, which generalizes the known Markovian SIR on networks. In our work, the simulation of contagion processes on simplicial complexes was done with SageMath [30], a Python-based, open-source mathematics software. The choice to use SageMath was due to the versatility of Python and the rich, topological libraries SageMath includes, for example, it has native support for simplicial complexes and graphs and it automates the construction of clique complexes. The visualization of data obtained from simulations is done with Python's Matplotlib [14].

4.1 Simplicial stochastic SIR

A stochastic SIR epidemic process in a simplicial complex X of dimension d is defined in the following way. Nodes (i.e., 0-simplices) will be either susceptible, infectious or recovered. Susceptible nodes can turn infectious, and infectious nodes eventually recover. Recovered nodes remain that way. At time 0, a randomly selected node is turned infectious, while the rest are susceptible. A k-simplex $\mathcal{K} \in X$ having all its nodes infected will infect a susceptible node x according to a Poisson process with rate β_{k+1} if $\mathcal{K} \cup \{x\}$ is a $k+1$ simplex in the complex. Each infected node remains

infectious for an exponentially distributed period with mean $1/\gamma$. The process stops at the first time T when zero infected nodes remain. The number of recovered nodes at time T is considered the final number of infected throughout the process. Note that the regular node to node transmission observed in networks is preserved in the process definition when $k = 0$. Furthermore, the process defined as above coincides with the Markovian SIR process on networks [8] when the simplicial complex is of dimension one (i.e., it only has nodes and edges). As far as the authors are aware, this process has not been defined in the literature. On simplicial complexes, the SIS and SIRS models have been considered [10, 16, 23, 32]. Ma and Guo [21] study an SIR model of information transmission on a hypernetwork, but with only one transmission parameter and in discrete time. Ball et al. [2] consider an SIR process on a network with household structure, where there is a different infection rate for nodes in the same household. However, their model considers only two rates of transmission, depending on whether nodes are or not in a household. The simplicial model here defined allows for a parameter for each simplicial dimension. Similarly, Fransson and Trapman [12] consider an SIR epidemic on a network with different transmission rates between nodes belonging to a same triangle. However, all triangles show this separate infection rate, in contrast to our model where triangles may not be included as simplices in the complex: i.e., three nodes v_1, v_2, v_3 may be pairwise connected in a network and thus form a triangle, but the simplex $\{v_1, v_2, v_3\}$ can be ommitted from a simplicial complex. Also, since our model has a transmission rate for each dimension, we allow groups of more than three nodes (i.e., simplicial complexes of dimension greater than 2) to have their own transmission rate, although for computational reasons, the experiments in Section 4.2 are limited to two-dimensional simplicial complexes.

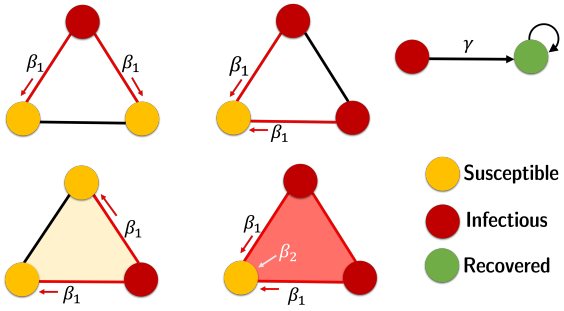


Figure 4.1: Representation of contagion dynamics on a simplicial complex of dimension 2.

4.2 EXPERIMENTS

As a first step to determine whether the simplicial higher order interactions have an effect on the spread of an SIR epidemic, simulations of the process are performed for an Erdős-Rényi graph $G(n = 120, p = 0.1)$ and its two-dimensional clique complex, i.e., all triangles of the graph are included as two-simplices. The resulting network has 120 nodes, 734 edges, 288 cliques of size 3, average degree of $\langle k_1 \rangle = 12.23$, each node belonging to an average of $\langle k_2 \rangle = 7.2$ two-simplices. The distribution of the final fraction of infected nodes in both scenarios will be studied. The parameters used are transmission rates $\beta_1 = 1, \beta_2 = 0$ in the case of the graph (one-dimensional complex) and $\beta_1 = 1, \beta_2 = 15$ in the case of its two-dimensional clique complex. In both cases, the recovery rate is $\gamma = 9$. A thousand runs are done for each process. The top histograms in Figure 4.2 show the distributions of the final fraction of infected nodes, which can be compared for the process with $\beta_2 = 0$ (left, epidemic on a network) and the process with $\beta_2 = 15$ (right, epidemic on the two-dimensional clique complex). The difference in frequency with which the epidemic takes off is noticeable.

The same process is considered in a network of contacts between students in a high school in Marseilles, France [22] and its two-dimensional clique complex. The network has 120 nodes, 348 edges, 272 cliques of size 3, $\langle k_1 \rangle = 5.8$ and $\langle k_2 \rangle = 6.8$.

Simulations with parameters $\beta_1 = 1, \beta_2 = 0, \gamma = 9$ and $\beta_1 = 1, \beta_2 = 15, \gamma = 9$ are done with a thousand runs each. Similarly, the distribution of the final fraction of infected is shown in the bottom histograms of Figure 4.2. It is observed how introducing the simplicial contagion leads to a larger frequency of outbreaks.

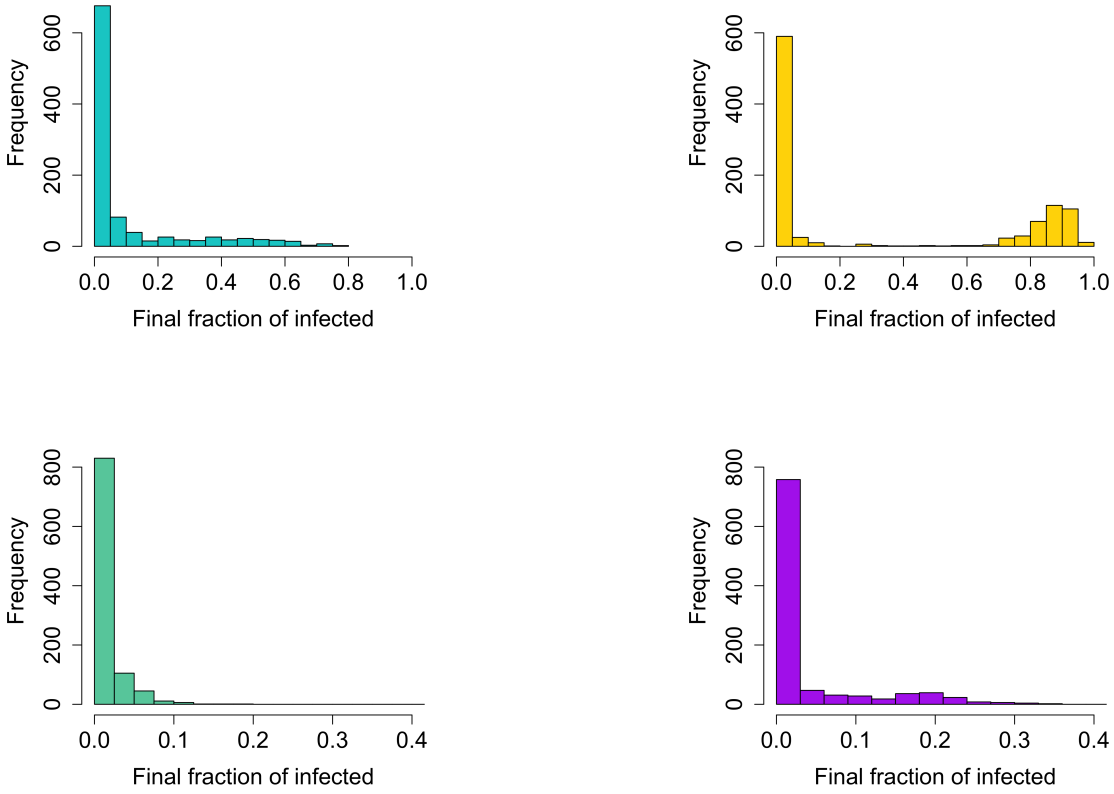


Figure 4.2: Histograms for the final fraction of infected for the simplicial stochastic SIR on graphs (left) and their two-dimensional clique complexes (right). Graphs used were an Erdős-Rényi graph (top) and a high school social network (bottom). Transmission rates were $\beta_1 = 1$ in all cases, $\beta_2 = 0$ for graphs and $\beta_2 = 15$ for the corresponding clique complexes. Recovery rate was $\gamma = 9$ for all cases

The use of $\beta_2 = 15$ in both cases was rather arbitrary. To further support our findings, the *average* final fraction of infected after 1000 runs of the epidemic process on the contact network were done for fixed $\beta_1 = 1, \gamma = 5$ and varying $\beta_2 = 0, 1, \ldots, 25$, and for fixed $\gamma = 5$ and varying $\beta_1 = 1, 2, \ldots, 10, \beta_2 = 0, 0.5, 1, \ldots, 20$.

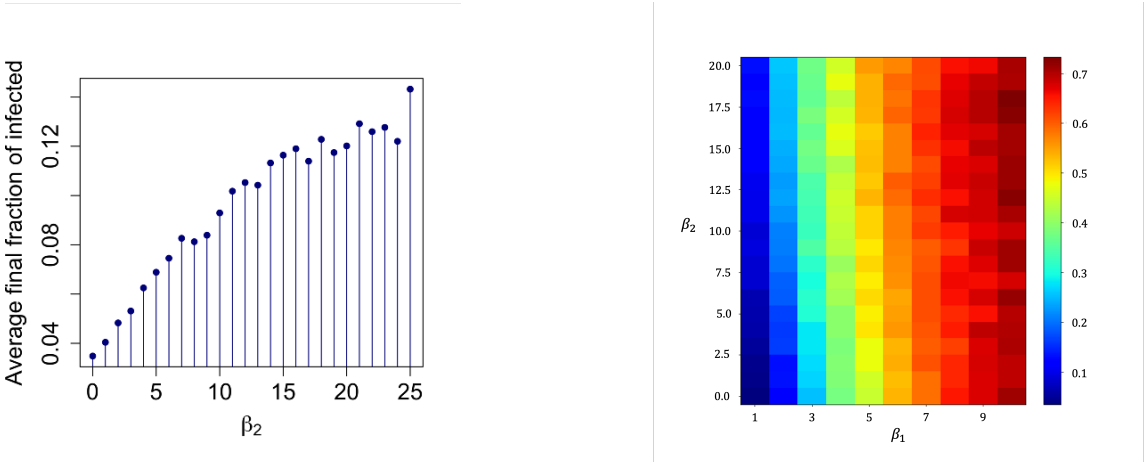


Figure 4.3: Simplicial complex used consisted of 120 nodes, $\langle k_1 \rangle = 5.8, \langle k_2 \rangle = 6.8$. **Left:** Average final fraction of infected over a thousand runs for fixed $\beta_1 = 1, \gamma = 5$ and different values of β_2. **Right:** Average final fraction of infected over a thousand runs for fixed $\gamma = 5$ and different values of β_1, β_2.

The results are displayed in Figure 4.3. In the case of the heat map in Figure 4.3, if the parameter β_2 had no effect, one would expect to see an increment on the average fraction of infected only when β_1 increases, which is the left to right direction in the heatmap. However, a bottom-to-top increment is seen when fixing any of the β_1 values, corresponding to an increment of the parameter β_2. This effect is less noticeable for larger β_1 values.

4.3 MEAN FIELD APPROACH

A mean-field approximation, under an homogeneous degree hypothesis is considered. For a simplicial complex of dimension D, transmission rates $\beta_1, \ldots, \beta_D$, recovery rate γ and average degree $\langle k_d \rangle$ for the d-dimensional simplices, the mean field equations are Equations 4.1 – 4.3:

$$s'(t) = -\sum_{d=1}^{D} \beta_d \langle k_d \rangle i(t)^d s(t), \tag{4.1}$$

$$i'(t) = -\gamma i(t) + \sum_{d=1}^{D} \beta_d \langle k_d \rangle i(t)^d s(t), \tag{4.2}$$

$$r'(t) = \gamma i(t). \tag{4.3}$$

At any given time $s + i + r = 1$. The initial conditions satisfy, $s_o + i_o + r_o = 1$, $r_o = 0$. Due to the dissipative dynamics imposed by the recovered rate equation, the epidemic has an absorbing state, $s_f + i_f + r_f = 1$, $i_f = 0$. Of particular interest is the total fraction of the total population of N nodes which gets infected during the entire epidemic, $I \equiv \int_0^\infty i(t)dt$. Clearly, the total number of infected nodes is NI and $r_f = I$.

Assuming that i, s and r are analytic functions, the mean field equations can be used to study the system close to any given point in the state space (s_a, i_a, r_a) by taking $i(t) = i_a + \phi(t)$, $s(t) = s_a + \sigma(t)$ and $r(t) = r_a + \rho(t)$, where ϕ, σ and ρ are small quantities. Substitution in the susceptible rate equation gives,

$$\begin{aligned} \sigma'(t) &= -[s_a + \sigma(t)] \sum_{d=1}^{D} \beta_d \langle k_d \rangle [i_a + \phi(t)]^d \\ &= -[s_a + \sigma(t)] \sum_{d=1}^{D} \beta_d \langle k_d \rangle \left[\sum_{r=0}^{d} i_a^{d-r} \phi^r(t) \binom{d}{r} \right]. \end{aligned} \tag{4.4}$$

By the corresponding substitution in the infected rate equation, the system is expressed by the following two coupled differential equations,

$$\begin{aligned} &\sigma'(t) + [s_a + \sigma(t)] \left[\sum_{d=1}^{D} \beta_d \langle k_d \rangle \left[i_a^d + i_a^{d-1} \phi(t) \frac{d!}{(d-1)!} \right] \right] \\ &= -[s_a + \sigma(t)] \sum_{d=1}^{D} \beta_d \langle k_d \rangle \left[\sum_{r=2}^{d} i_a^{d-r} \phi^r(t) \binom{d}{k} \right], \end{aligned} \tag{4.5}$$

$$\begin{aligned} &\phi'(t) + \gamma\phi(t) - [s_a + \sigma(t)] \left[\sum_{d=1}^{D} \beta_d \langle k_d \rangle \left[i_a^d + i_a^{d-1} \phi(t) \frac{d!}{(d-1)!} \right] \right] \\ &= [s_a + \sigma(t)] \sum_{d=1}^{D} \beta_d \langle k_d \rangle \left[\sum_{r=2}^{d} i_a^{d-r} \phi^r(t) \binom{d}{k} \right]. \end{aligned}$$

Up to first order in ϕ, Equation 4.5 reads,

$$\sigma'(t) + [s_a + \sigma(t)] \left[\sum_{d=1}^{D} \beta_d \langle k_d \rangle \left[i_a^d + d i_a^{d-1} \phi(t) \right] \right] = 0 \tag{4.6}$$

$$\phi'(t) + \gamma\phi(t) - [s_a + \sigma(t)] \left[\sum_{d=1}^{D} \beta_d \langle k_d \rangle \left[i_a^d + d i_a^{d-1} \phi(t) \right] \right] = 0.$$

The fact that the only possible value for i_f is zero implies that the absorbing state tends to a dynamic system that is independent of the simplices of order greater than $D = 1$. To see this, note that close to $i_f = 0$, only the simplex $D = 1$ survives in Equation 4.6,

$$\begin{aligned} s' &\to -\beta_1 \langle k_1 \rangle s(t)\phi(t) \\ r' &\to \gamma\phi(t), \end{aligned} \tag{4.7}$$

so

$$\frac{ds}{dr} = \frac{-\beta_1 \langle k_1 \rangle s(t)}{\gamma}. \tag{4.8}$$

Therefore,

$$s_f = s_\tau e^{\frac{\beta_1 \langle k_1 \rangle}{\gamma}(1 - s_f)}, \tag{4.9}$$

where τ is a characteristic time scale.

Under the additional assumption that $|\sigma| < |\phi|$, namely that the recovery dynamics is slower than the infection dynamics under small perturbations, which appears to be a plausible assumption for SIR-type epidemics [1, 17], the equation for ϕ,

$$\phi'(t) + \phi(t) \left[\gamma - [s_a + \sigma(t)] \sum_{d=1}^{D} \beta_d \langle k_d \rangle i_a^{d-1} d \right] - [s_a + \sigma(t)] \sum_{d=1}^{D} \beta_d \langle k_d \rangle i_a^d = 0 \tag{4.10}$$

has the solution

$$\phi(t) = \frac{s_a \sum_{d=1}^{D} \beta_d \langle k_d \rangle i_a^d}{\left[\gamma - s_a \sum_{d=1}^{D} \beta_d \langle k_d \rangle i_a^{d-1} d \right]} \left[1 - e^{-\left[\gamma - s_a \sum_{d=1}^{D} \beta_d \langle k_d \rangle i_a^{d-1} d \right] t} \right]. \tag{4.11}$$

From Eq. (4.11) it follows the outbreak threshold for the epidemics, which is affected by the higher order interactions of simplices $D > 1$, is

$$s_o \sum_{d=1}^{D} \beta_d \langle k_d \rangle i_o^{d-1} d > \gamma. \tag{4.12}$$

Clearly, the well known outbreak condition for a SIR epidemics in a graph $D = 1$ is recovered, $s_o > \frac{\gamma}{\beta_1 \langle k_1 \rangle}$. Furthermore, the threshold condition can be verified with the numerical solution of the mean field equations. A process with $\gamma = 10$, $\langle k_1 \rangle \approx 50$, $\langle k_2 \rangle \approx 10$, and one initially infected node out of 300, would have a threshold for values above the line $\beta_2 = 45000/299 - 750\beta_1$ in a (β_1, β_2) plane, which is the straight line which intersects at $\beta_1 = 60/299 \approx 0.2$ and $\beta_2 = 45000/299 \approx 150$, which can be observed in Figure 4.5 as the final fraction of infected in the mean field case (bottom right) is close to zero for this region. The small magnitude of initially infected nodes makes the effect of β_2 small for this threshold condition. If, however, an initial fraction of 15% infected nodes (45/300) is considered, the threshold region is that above the line $\beta_2 = 200/51 - (50/3)\beta_1$ in a (β_1, β_2) plane, which is the straight line which intersects at $\beta_1 = 4/14 \approx 0.235$ and $\beta_2 = 200/51 \approx 3.92$. This can be observed in Figure 4.6. Although an initial fraction of infected of 15% of the population may be unrealistic in some scenarios, it is used here to showcase the correctness of our outbreak threshold.

From Eq. (4.11) the characteristic time scale for the epidemic τ, can be defined as

$$\tau = \frac{1}{\left[\gamma - s_a \sum_{d=1}^{D} \beta_d \langle k_d \rangle i_a^{d-1} d\right]}. \tag{4.13}$$

At time τ, an epidemic outbreak in a susceptible population essentially starts to die out, $\phi \to 0$. It follows that exactly in $t = \tau$,

$$0 = -\gamma i(\tau) + \sum_{d=1}^{D} \beta_d \langle k_d \rangle i(\tau)^d s(\tau). \tag{4.14}$$

By the equation for s' and introducing the definitions $i_\tau \equiv i(\tau)$ and $s_\tau \equiv s(\tau)$,

$$s_\tau = s_o e^{-\tau i_\tau \left[\beta_1 \langle k_1 \rangle + 2\beta_2 \langle k_2 \rangle i_\tau + \ldots + D\beta_D \langle k_D \rangle i_\tau^{D-1}\right]}, \tag{4.15}$$

$$\beta_1 \langle k_1 \rangle + 2\beta_2 \langle k_2 \rangle i_\tau + 3\beta_3 \langle k_3 \rangle i_\tau^2 + \ldots + D\beta_D \langle k_D \rangle i_\tau^{D-1} = \frac{\gamma}{s_\tau}.$$

By the definition of τ, the epidemic peak is bounded from below by i_τ. For $D = 2$,

$$i_\tau = \frac{1}{2\beta_2 \langle k_2 \rangle} \left[\frac{\gamma}{s_\tau} - \beta_1 \langle k_1 \rangle\right] \leq i_*. \tag{4.16}$$

Therefore for $D = 2$ there is a bound for the value s at the infection peak, denoted by s_*,

$$s_* \leq \frac{\gamma}{\beta_1 \langle k_1 \rangle}. \tag{4.17}$$

The mean field approximation is tested numerically by comparing its results with those of simulations in random simplicial complexes. In the Figure 4.3 the final fraction of infected nodes I, given by averaging over 1000 simulations of the microscopic stochastic process is reported. The graph at the right of the Figure 4.3 shows a parametric sweep over the range $\beta_1 \in [1, 10]$, $\beta_2 \in [0, 20]$ with $\gamma = 5$ fixed. The graph in the left is the slice at $\beta_1 = 1$. From Equation 4.16 and taking into account that $s_f \leq s_\tau$, $I \geq i_*$ and $s_f = 1 - I$, mean field predicts the bound given by the relation,

$$I \geq 1 - \frac{\gamma}{\beta_1 \langle k_1 \rangle}. \tag{4.18}$$

This bound is experimentally verified for the parametric sweep considered in the microscopic simulations reported in Figure 4.3 and by the simulations and numerical solutions of the mean field equations reported in the Figures 4.5 and 4.6. For instance, in the conditions at the left graph of Figure 4.3, the resulting bound is $I \geq 0.1379$. The conditions considered in the Figures 4.5 and 4.6 on the other hand, obey the bound $I \geq 0.6799$. Both analytical bounds are in complete accordance with observations. The interplay between fluctuations and high order interactions is however not completely captured by the homogeneous mean field theory, which is exemplified by Figure 4.2, where the histograms of the values for I sampled from the simulations are reported. The uncertainty in I is clearly influenced by the simplicial interactions.

Figure 4.4 shows the evolution of infectious and recovered individuals in time on a simplicial complex with 500 nodes, $\langle k_1 \rangle \approx 25, \langle k_2 \rangle \approx 10$ and an epidemic with parameters $\beta_1 = 1, \beta_2 = 5, \gamma = 5$. Notice that the bound (4.16) is consistent with the observed peak.

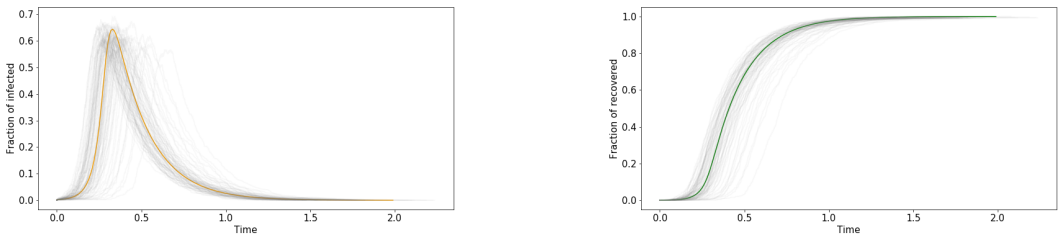


Figure 4.4: Comparing the evolution of infectious (left) and recovered (right) nodes in the realizations of a hundred simulations (gray) and the numerical solution of the mean field equations (color). Simplicial complex used had 500 nodes, $\langle k_1 \rangle \approx 25, \langle k_2 \rangle \approx 10$, and the parameters were $\beta_1 = 1, \beta_2 = 5, \gamma = 5$.

In Figure 4.5 (top), the maximum of infected nodes during the epidemic is compared between mean-field approximations and simulations (in this latter case, via an average of two hundred repetitions). Similarly, the final fraction of infected is compared in Figure 4.5 (bottom). All the comparisons were done on a simplicial complex with 300 nodes, $\langle k_1 \rangle \approx 50$, $\langle k_2 \rangle \approx 10$, and epidemics with parameters $\beta_1 = 0.025, 0.05, \ldots, 0.625$, $\beta_2 = 0, 1, \ldots, 10$, $\gamma = 10$. It is seen that our mean field approximation is consistent with our simulation results, which allows the possibility of future work to be done without expensive simulations.

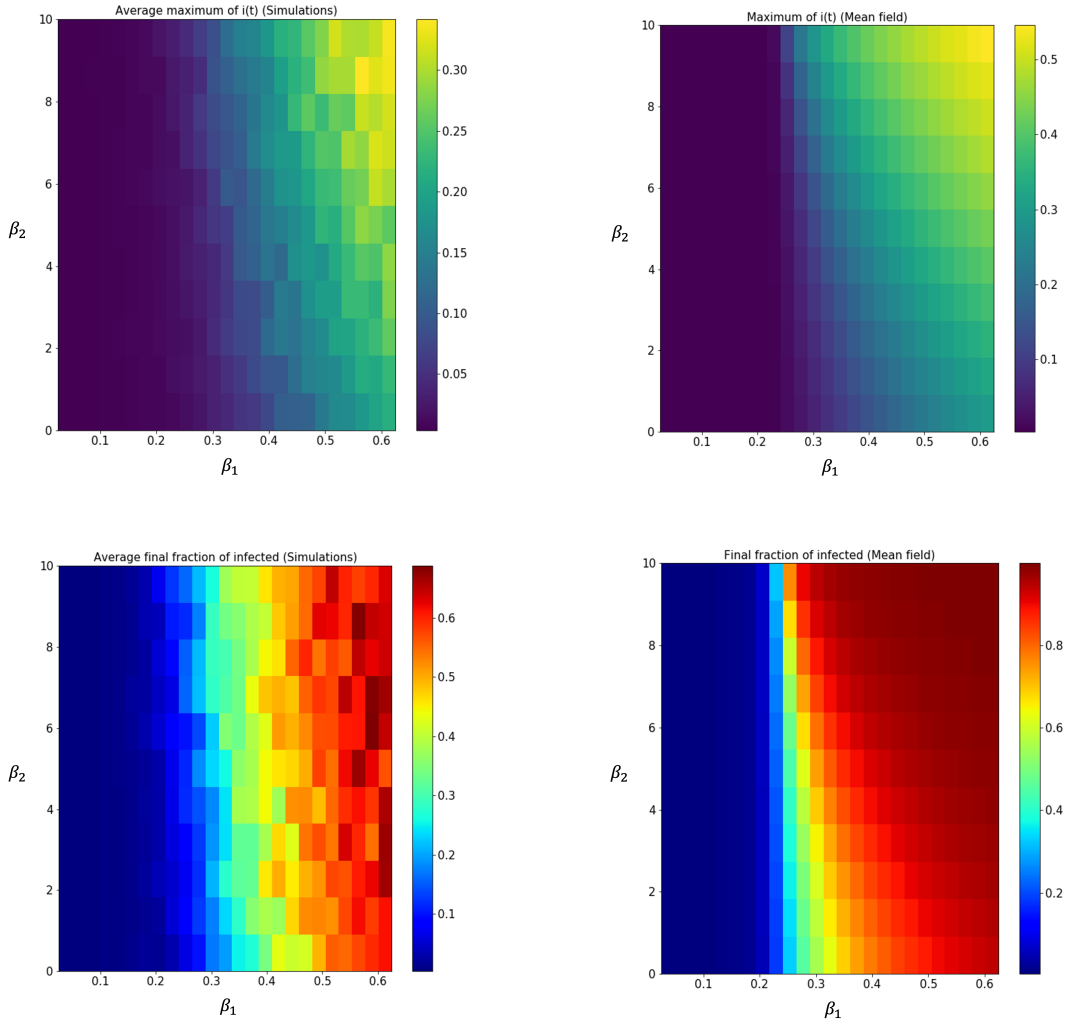


Figure 4.5: Heatmaps comparing average final of infected and maximum of infected obtained by simulation vs. mean field approximation on a random simplicial complex with 300 nodes, $\langle k_1 \rangle \approx 50$, $\langle k_2 \rangle \approx 10$, $\gamma = 10$.

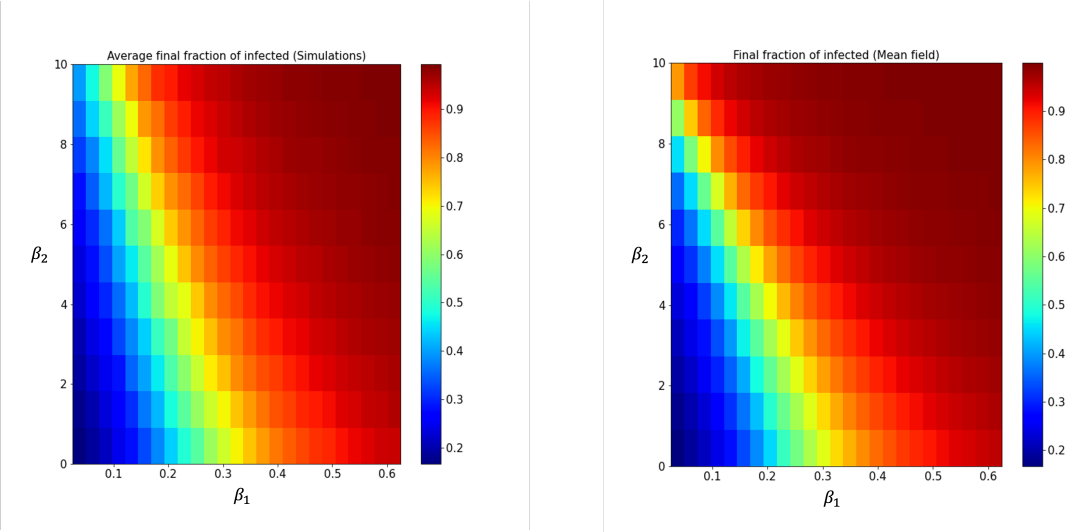


Figure 4.6: Heatmaps comparing average final of infected obtained by simulation vs. mean field approximation on a random simplicial complex with 300 nodes, $\langle k_1 \rangle \approx 50$, $\langle k_2 \rangle \approx 10$, $\gamma = 10$ but considering 15% of initially infected nodes.

Chapter 5

Conclusions

The dissipative dynamics of the SIR contagion process permits a very complete mean field characterization by which it is even possible to obtain an explicit solution for the changes in the fraction of infected nodes close to any given state space point for any simplex dimension. Other similar contagion processes like SIS [10, 16, 23] or SIRS [32] have been recently studied on simplicial complexes, but our model offers a framework in which the mean field theory, besides its good agreement with the underlying microscopic process (as shown in Figures 4.4, 4.5 and 4.6), it is also capable of describing the effects of arbitrarily higher order interactions. This makes our model an ideal framework for the study of variations beyond contagion processes on simplicial networks. From the results in Sections 4.2 and 4.3 it is clear that the presence of higher order simplicials has an impact in all of the prominent features of the epidemic. Take for instance the epidemic's peak, which is bounded according to Eq. (4.16) at interactions up to $D = 2$. It turns out that an increase in the strength of the interactions within the simplex has the not very intuitive effect of diminishing the lower bound for the infection's peak. The following intriguing remarks are however worth noting at this matters. It is plausible to consider the simplicial interactions to model other types of relationships among nodes besides infection, for example risk perception or social pressure-induced prophylaxis. In these situations the interaction strength parameters very well can be of positive

or negative sign. In such a case, the mathematical structure of the mean field linearized solutions does not forbid the existence of stationary states different from the absorbing $i = 0$ state. A second remark is that the microscopic simulations shown in Figure 4.2 indicate that although the average final number of infected nodes is almost independent of the simplicial interactions, the fluctuations around that average are not and this can be indicative of some interplay between the simplicial interactions and the underlying stochasticity. This aspect cannot be analyzed by the developed homogeneous mean field description, so further study via more extensive simulations or more advanced mean field approximations is worth to pursue for future works on the model.

5.1 Future work

Future work may consist using theory of branching processes, to analytically study the contagion model here defined. The choice of focusing on simplicial complexes and not other generalizations, such as hypergraphs, was in part because of the rich mathematical theory in algebraic topology about this combinatorial structure. As such, a line of research is to study the structural properties of simplicial complexes and see how these affect a contagion process spreading on it. Additional work may consist of incorporating the multi-layered paradigm to higher order network structures, by studying a suitable definition of multi-layered simplicial complexes, their properties and processes therein. More realistic versions of the models here described can be defined, losing perhaps the ability to analytically study them, but still well suited for computational studies. For example, in the context of contagion processes, be it rumour spreading or disease spreading, nodes can be considered agents in a multiagent system, with properties such as memory or behavioral traits.

Bibliography

[1] Augeraud-Véron, E. and Sari, N. (2014). Seasonal dynamics in an SIR epidemic system. *Journal of mathematical biology*, 68(3):701–725.

[2] Ball, F., Sirl, D., and Trapman, P. (2009). Threshold behaviour and final outcome of an epidemic on a random network with household structure. *Advances in Applied Probability*, 41(3):32.

[3] Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.-G., and Petri, G. (2020). Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 874:1–92.

[4] Berge, C. (1984). *Hypergraphs: combinatorics of finite sets*, volume 45. Elsevier.

[5] Bianconi, G. (2021). *HIGHER ORDER NETWORKS: An Introduction to Simplicial Complexes*. Cambridge University Press.

[6] Bick, C., Gross, E., Harrington, H. A., and Schaub, M. T. (2021). What are higher-order networks?

[7] Bollobás, B. (1998). *Modern graph theory*, volume 184. Springer Science & Business Media.

[8] Britton, T. (2020). Epidemic models on social networks – with inference. *Statistica Neerlandica*, 74(3):222–241.

[9] Casella, G. and Berger, R. (2001). *Statistical Inference*. Duxbury Press, Florence, AL, 2nd edition.

[10] Cisneros-Velarde, P. and Bullo, F. (2020). Multi-group SIS epidemics with simplicial and higher-order interactions.

[11] DeVille, L. (2021). Consensus on simplicial complexes: Results on stability and synchronization. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(2):023137.

[12] Fransson, C. and Trapman, P. (2019). Sir epidemics and vaccination on random graphs with clustering. *Journal of Mathematical Biology*, 78(7).

[13] Horstmeyer, L. and Kuehn, C. (2020). Adaptive voter model on simplicial complexes. *Physical Review E*, 101(2):022305.

[14] Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

[15] Iacopini, I., Petri, G., Barrat, A., and Latora, V. (2019a). Simplicial models of social contagion. *Nature Communications*, 10(1).

[16] Iacopini, I., Petri, G., Barrat, A., and Latora, V. (2019b). Simplicial models of social contagion. *Nature Communications*, 10(1).

[17] Jardón-Kojakhmetov, H., Kuehn, C., Pugliese, A., and Sensi, M. (2021). A geometric analysis of the SIR, SIRS and SIRWS epidemiological models. *Nonlinear Analysis: Real World Applications*, 58(103220).

[18] Jhun, B., Jo, M., and Kahng, B. (2019). Simplicial SIS model in scale-free uniform hypergraph. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12).

[19] Kahle, M. (2017). *Random Simplicial Complexes*. CRC Press, 3rd edition.

[20] Lefschetz, S. (1930). *Topology*, volume 12. American Mathematical Soc.

[21] Ma, T. and Guo, J. (2018). Study on information transmission model of enterprise informal organizations based on the hypernetwork. *Chinese journal of physics*, 56(5):2424–2438.

[22] Mastrandrea, R., Fournet, J., and Barrat, A. (2015). Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLOS ONE*, 10(9).

[23] Matamalas, J. T., Gómez, S., and Arenas, A. (2020). Abrupt phase transition of epidemic spreading in simplicial complexes. *Physical Review Research*, 2(1):012049.

[24] Millán, A. P., Torres, J. J., and Bianconi, G. (2020). Explosive higher-order kuramoto dynamics on simplicial complexes. *Physical Review Letters*, 124(21).

[25] Newman, M. (2018). *Networks*, volume 1. Oxford University Press.

[26] Rotman, J. J. (2013). *An introduction to algebraic topology*, volume 119. Springer Science & Business Media.

[27] Schaub, M. T., Benson, A. R., Horn, P., Lippner, G., and Jadbabaie, A. (2020). Random walks on simplicial complexes and the normalized hodge 1-laplacian. *SIAM Review*, 62(2):353–391.

[28] Schaub, M. T., Zhu, Y., Seby, J.-B., Roddenberry, T. M., and Segarra, S. (2021). Signal processing on higher-order networks: Livin'on the edge... and beyond. *Signal Processing*, 187.

[29] Sun, H. and Bianconi, G. (2021). Higher-order percolation processes on multiplex hypergraphs. *Physical Review E*, 104(3):034306.

[30] The Sage Developers (2011). Sagemath, the Sage Mathematics Software System (Version 9.0). `https://www.sagemath.org`.

[31] Wang, D., Zhao, Y., Leng, H., and Small, M. (2020). A social communication model based on simplicial complexes. *Physics Letters A*, 384(35).

[32] Wang, D., Zhao, Y., Luo, J., and Leng, H. (2021). Simplicial sirs epidemic models with nonlinear incidence rates. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(5):053112.

[33] Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of anthropological research*, 33(4):452–473.

[34] Zhao, D., Li, R., Peng, H., Zhong, M., and Wang, W. (2022). Higher-order percolation in simplicial complexes. *Chaos, Solitons & Fractals*, 155.

Resumen autobiográfico

Gerardo Palafox Castillo

Candidato para obtener el grado de
Doctorado en Ingeniería de Sistemas

Universidad Autónoma de Nuevo León
Facultad de Ingeniería Mecánica y Eléctrica

Tesis:
Simplicial models for higher order interactions in complex networks

Nacido en Tampico, Tamaulipas, el 16 de Marzo de 1994, hijo de Gerardo Palafox Sánchez y Patricia Elena Castillo Torres. Licenciado en matemáticas por la Facultad de Ciencias Físico Matemáticas de la Universidad Autónoma de Nuevo León, y maestro en ciencias con orientación en matemáticas por la University of Western Ontario.