



Mathematical Institute
University of Oxford

Modelling Higher-Order Interactions with Simplicial Complexes

Contents

1	Introduction	2
2	Simplicial Complexes	2
2.1	Defining Simplicial Complexes	2
3	Representations of Simplicial Complexes	3
3.1	Incidence Matrix	4
3.2	Adjacency Matrix	4
3.2.1	Upper and Lower Adjacency matrices	5
3.2.2	Adjacency Tensor	6
4	Models of Simplicial Complexes	6
4.1	Erdős-Rényi Random Graph	7
4.2	Linial-Meshulam Complex	7
4.3	Random d -complex	8
4.4	Random Clique Complex	9
4.5	Multi-Parameter Complexes	10
4.6	Random Simplicial Complex (RSC) Model	11
5	A Dynamical Application of Simplicial Complexes: Epidemic Modelling	13
5.1	SIR Dynamics on Networks	14
5.2	SIR Dynamics on Simplicial Complexes	15
5.2.1	Governing Equations	15
5.2.2	Mean Field Approach	17
5.2.3	Numerical Simulations	19
6	Concluding Remarks	21
	References	22
A	Simplicial Homology	23

1 Introduction

The applications of classical graphs of dyadic interactions are clearly well established in a multitude of fields, however often the restriction to pairwise interactions can withhold important information about complex systems which exhibit group interactions. Hypergraphs are the natural generalisation of classical graphs that account for higher order interactions, and the literature regarding their use in network science is extensive. We instead consider the option from the field of algebraic topology; simplicial complexes. These complex networks incorporate higher order interactions through implied geometric structures present in an underlying classical graph. The construction of these complexes is not unique; there exists several methods which have different criteria for the implications of these higher dimensional structures. We begin with an introduction of simplicial complexes and their various constructions. We then consider representations and measures of simplicial complexes, namely we generalise the classical incidence and adjacency matrices to the regime of simplicial complexes. As a means to look at an application of simplicial complexes, we consider possible generalisations of the Erdős-Rényi model for classical graphs. Finally, we look at an example of the dynamical application of simplicial complexes by considering simplicial epidemic modelling.

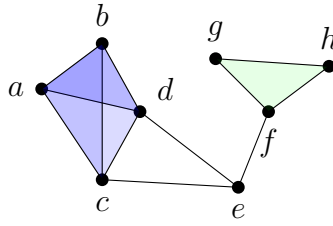
2 Simplicial Complexes

2.1 Defining Simplicial Complexes

We begin by defining these key structures that form the basis of this report. We define a k -dimensional simplex to be the convex hull formed by $k + 1$ affinely independent points. These structures are simply the k -dimensional analogous of cliques formed by $k + 1$ nodes in networks. Following from this definition, 0-simplices are nodes, 1-simplices are edges, 2-simplices are triangles and so on. We define the faces of a simplex to be the simplices formed by subsets of the set of nodes that constitute the simplex. For example, if we denote by $\sigma = [n_0, \dots, n_d]$ an arbitrary d -dimensional simplex formed by the $d + 1$ elements of $\{n_0, \dots, n_d\}$, and similarly by $\tau = [n_{i_0}, \dots, n_{i_k}]$ a k -dimensional simplex, for $k \leq d$. Then τ is a face of σ if $\{n_{i_0}, \dots, n_{i_k}\} \subset \{n_0, \dots, n_d\}$, in this case we say $\tau \subset \sigma$. We can now formally define an abstract simplicial complex. Let \mathcal{X} denote a set of simplices. We say \mathcal{X} is a simplicial complex if both of the following conditions are satisfied:

1. If $\sigma \in \mathcal{X}$ and $\tau \subset \sigma$, then $\tau \in \mathcal{X}$,
2. If $\sigma_1, \sigma_2 \in \mathcal{X}$ and $\sigma_1 \cap \sigma_2 = \tau \neq \emptyset$, then $\tau \subset \sigma_1$ and $\tau \subset \sigma_2$.

Here, the first condition ensures that simplicial complexes are closed under subsets, meaning that all faces of simplices in the simplicial complex are also in the simplicial complex. The second condition ensures that when two simplices in the simplicial complex intersect, the face of intersection is a face of both of the simplices.



$$\mathcal{X} = \{[a, b, c, d], [a, b, c], [a, b, d], [a, c, d], [b, c, d], [f, g, h], [a, b], [a, c], [a, d], [b, c], [b, d], [c, d], [c, e], [d, e], [e, f], [f, g], [f, h], [g, h], [a], [b], [c], [d], [e], [f], [g], [h]\}$$

Figure 1: Example of a 3-dimensional simplicial complex \mathcal{X} containing a tetrahedron and a triangle.

We finally define the k -skeleton of a simplicial complex \mathcal{X} . If we denote by \mathcal{X}^k the set of all simplices of exact dimension k in \mathcal{X} , then we define the k -skeleton $\mathcal{X}^{(k)}$ of \mathcal{X} as

$$\mathcal{X}^{(k)} = \bigcup_{i=0}^k \mathcal{X}^i \quad (1)$$

which is just the set of all simplices of dimension at most k . Often when considering graphs, we are interested in properties such as connectivity. A very powerful generalisation of this property is described by the homology of simplicial complexes. Although this notion is not relevant in our discussions, we include a simple introduction to the concept in appendix A.

3 Representations of Simplicial Complexes

In this section we discuss representations of simplicial complexes analogous to the well-established incidence and adjacency matrices for graphs. The analogue for the incidence matrix follows naturally, however when generalising the adjacency matrix more care needs to be taken as there is some level of ambiguity when dealing with

higher dimensional structures. We consider two options for the generalised adjacency matrix. Throughout this section we use results from [1].

3.1 Incidence Matrix

Recall that the incidence matrix of a graph $G = (V, E)$, where $|V| = n$ and $|E| = m$ is the $n \times m$ matrix I with entries in $\{0, 1\}$, where the (i, j) th element is 1 if node i is included in edge j and 0 otherwise. The natural generalisation of this is to consider the incidence of all k -simplices for $k \geq 1$ on all nodes. We can either separate these matrices by value of k or consider the aggregated matrix of all dimensions[1]. Let i denote a node, and σ some k -simplex, with $k \geq 1$, that belongs to a simplicial complex \mathcal{X} . The aggregated incidence matrix is given by

$$I_{i\sigma} = \begin{cases} 1 & i \in \sigma \\ 0 & \text{else} \end{cases} \quad (2)$$

and the incidence matrix for dimension k is given by

$$I_{i\tau}^{(k)} = \begin{cases} 1 & i \in \tau \\ 0 & \text{else} \end{cases} \quad (3)$$

where τ is a simplex in \mathcal{X} with exact dimension k . Let \mathcal{X} be of dimension D (that is the simplicial complex contains simplices of dimension at most D), and let the number of simplices of dimension exactly k in the complex be denoted by m_k . Then clearly I is of size $n \times \sum_{k=1}^D m_k$ and each $I^{(k)}$ is of size $n \times m_k$. Both representations simplify to the classical incidence matrix when $D = 1$, i.e. when we are in the regime of graphs.

3.2 Adjacency Matrix

The generalisation of the classical adjacency matrix is not as natural. Recall that for the case of graphs, we can construct the adjacency matrix from the incidence matrix I and the $n \times n$ matrix D whose off-diagonal elements are 0, and the i th diagonal element is the degree of node i . Then we define the adjacency matrix by

$$A = II^T - D.$$

This $n \times n$ matrix simply captures where two nodes are adjacent. All entries are 0 except entries (i, j) where nodes i and j are connected by an edge. We explore two

ways to generalise this result. The first is through the notion of upper and lower adjacency, and the second relies on a tensorial representation to account for higher dimensional adjacency.

3.2.1 Upper and Lower Adjacency matrices

When considering higher dimensional structures as in simplicial complexes, there are two natural ways in which two k -simplices can be considered adjacent to each other. Let σ_i and σ_j denote two k -simplices in the simplicial complex \mathcal{X} of dimension greater than k . The first type of adjacency follows from the classical notion of adjacency from graphs. Normally we consider if two nodes are adjacent by whether or not there is an edge connecting them, we can generalise this to say two k -simplices are adjacent if they are common faces of a $(k+1)$ -simplex, this form of adjacency in simplicial complexes is known as upper adjacency. The second notion of adjacency is that we consider two k -simplices to be adjacent if they share a common $(k-1)$ -simplex, this is known as lower adjacency. Say $\sigma_i = \{u_0, \dots, u_k\}$ and $\sigma_j = \{v_0, \dots, v_k\}$. We have that σ_i is lower adjacent to σ_j if there exists a $(k-1)$ -simplex $\tau = \{w_0, \dots, w_{k-1}\}$ such that $\tau \subset \sigma_i$ and $\tau \subset \sigma_j$, and σ_i is upper adjacent to σ_j if there exists a $(k+1)$ -simplex $T = \{x_0, \dots, x_{k+1}\}$ such that $\sigma_i \subset T$ and $\sigma_j \subset T$. By this definition, clearly upper adjacency for dimension $k = 0$ describes the classical notion of adjacency. Also, it is important to note that the notion of lower adjacency is only valid for $k \geq 1$ since we cannot have simplices of negative dimension. We denote lower adjacency of σ_i and σ_j by $\sigma_i \sim_L \sigma_j$, and the upper adjacency by $\sigma_i \sim_U \sigma_j$ [3]. We can now define the matrix representations of upper and lower adjacency. For some simplicial complex \mathcal{X} of dimension $D \geq k$, where the number of simplices of dimension k in \mathcal{X} is m_k , we have that the lower adjacency matrix at dimension k is given by the $m_k \times m_k$ matrix

$$(A_L^k)_{ij} = \begin{cases} 1 & \text{if } \sigma_i \sim_L \sigma_j \\ 0 & \text{if } \sigma_i \not\sim_L \sigma_j \text{ or } i = j \end{cases} \quad (4)$$

and the upper adjacency matrix at dimension k is given by the $m_k \times m_k$ matrix

$$(A_U^k)_{ij} = \begin{cases} 1 & \text{if } \sigma_i \sim_U \sigma_j \\ 0 & \text{if } \sigma_i \not\sim_U \sigma_j \text{ or } i = j. \end{cases} \quad (5)$$

It is convenient to note that if $\sigma_i = \{u_0, \dots, u_k\}$ and $\sigma_j = \{v_0, \dots, v_k\}$ are distinct and upper adjacent, then they must also be lower adjacent [3]. These notions of adjacency would prove useful when considering dynamical processes in which a simplex of any dimension could interact with another simplex of any dimension.

3.2.2 Adjacency Tensor

Considering the adjacency of all simplices with all other simplices is not always necessary when considering practical applications of these representations. Another very natural approach to generalising the adjacency matrix without the use of lower and upper adjacency is through a tensor representation. We adapt this representation from [1], changing some notation for convenience. We consider an adjacency tensor for each dimension d where D is the dimension of the simplicial complex \mathcal{X} . For $d = 1$ we have that $A_1 = a_{ij}$, where the entry (i, j) is 1 if $\{i, j\}$ is a 1-simplex in \mathcal{X} and 0 otherwise¹. For $d = 2$ we have

$$\mathcal{A}_2 = a_{ijk} = \begin{cases} 1 & \text{if } \{i, j, k\} \in \mathcal{X} \\ 0 & \text{otherwise .} \end{cases} \quad (6)$$

And we continue in this fashion, if we let σ_{d_j} denote a general index² in a $(d + 1)$ th rank tensor, then

$$\mathcal{A}_d = a_{\sigma_{d_j}} = \begin{cases} 1 & \text{if } \{\sigma_{d_j}\} \in \mathcal{X} \\ 0 & \text{otherwise .} \end{cases} \quad (7)$$

We examine all possible combinations of $d + 1$ 0-simplices in each tensor, hence if we let the number of 0-simplices in \mathcal{X} be N , then \mathcal{A}_d is an $N \times N \times \dots \times N$ tensor, where the tensorial rank is $d + 1$. The d th adjacency tensor captures the adjacency of 0-simplices to d -simplices in \mathcal{X} , this representation is particularly useful when considering dynamical processes where nodes interact with higher order simplices, therefore we will make use of this in our analysis of dynamical systems in section 5. We note that the full connectivity of 0-simplices to higher order simplices in the complex is captured by the set of D tensors $\mathcal{A} = \{\mathcal{A}_1, \dots \mathcal{A}_D\}$.

4 Models of Simplicial Complexes

As is the case for networks, stochastic models can be extremely useful tools for identifying statistically significant features or patterns which are found in empirical data sets. Such models are not intended to exactly, or even closely, match these real data sets, but rather they are meant to be used as an unbiased tool whose statistical properties are convenient for calculations. These are commonly referred to as *null models*.

¹Note this is just the graph adjacency matrix (2nd rank tensor).

²By this we mean that we let $\sigma_{d_j} = i_0, \dots i_d$, where each i_k denotes a 0-simplex in complex and so we consider all possible combinations of $d + 1$ nodes that could form a d -simplex.

In this section we review the foundational stochastic model for networks, namely the Erdős-Rényi (ER) model, and then consider generalisations of this to simplicial complexes. We largely discuss models as stated in the review by Bobrowski and Krioukov [2].

4.1 Erdős-Rényi Random Graph

We denote by $G(N, q)$ the ER random graph with N nodes and with q being the probability that an edge between any two distinct nodes exists, independent of the existence of any other edges. Clearly the maximum number of edges that can exist in this model is given by the binomial coefficient $\binom{N}{2} = N(N-1)/2$, and hence the expected number of edges is simply $qN(N-1)/2$. Since each edge connects 2 nodes and there are N nodes in total, the expected degree is given by $\langle k \rangle = 2/N \cdot qN(N-1)/2 = q(N-1)$. A given node has probability q of being connected to any of the remaining $N-1$ nodes, independently, and probability $1-q$ of not being connected. Hence the probability of the node having degree $k \leq N$ is given by the binomial distribution

$$p(k) = \binom{N-1}{k} q^k (1-q)^{N-1-k}. \quad (8)$$

4.2 Linial-Meshulam Complex

The first generalisation of $G(N, q)$ we consider is the Linial-Meshulam Complex, also known as the random 2-complex. Denoted by $Y(N, q_\Delta)$ this model starts with the complete graph of N nodes, and we proceed to add 2-simplices (triangles) between triplets of nodes independently with probability q_Δ [2].

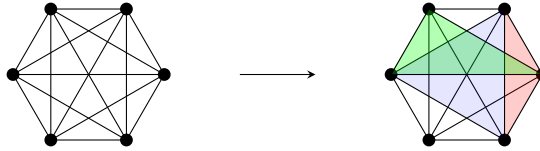


Figure 2: An example of a generated Linial-Meshulam complex $Y(6, 0.25)$. The figure shows the initial complete graph of 6 nodes, and the generated model with 3 triangles formed.

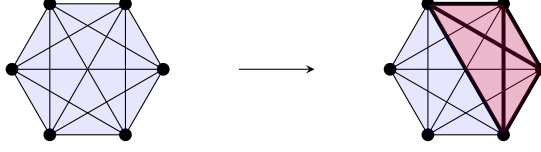


Figure 3: An example of a generated random 3-complex $Y_3(6, 0.1)$. The figure shows the initial complete 2-skeleton (all edges and triangles included), and the generated model with 1 tetrahedron formed.

4.3 Random d -complex

This model is the extension of the Linial-Meshulam complex to d -dimensions. Instead of starting with a complete graph of N nodes and adding 2-simplices independently with probability q_Δ , we start with a complete $(d-1)$ -skeleton on N nodes and add d -simplices independently with probability q_d [2]. Recall that a complete $(d-1)$ -skeleton on N -nodes is the simplicial complex consisting of N 0-simplices, with all possible simplices up to dimension $(d-1)$ included. We denote this model by $Y_d(N, q_d)$. Clearly $Y_1(N, q) = G(N, q)$ and $Y_2(N, q_\Delta) = Y(N, q_\Delta)$. Similar to ER graphs, these models of simplicial complexes have convenient statistical properties so we can easily find some important properties³. Consider the simplicial complex formed by $Y_d(N, q_d)$. If we denote the number of d -simplices in $Y_d(N, q_d)$ by $|Y_d(N, q_d)| = |Y_d|$, we have

$$\max |Y_d| = \binom{N}{d+1} = \frac{N(N-1)\cdots(N-d)}{(d+1)!}$$

since a d -simplex is formed by $d+1$ nodes. Following from this, we have the expectation

$$\langle |Y_d| \rangle = q_d \frac{N(N-1)\cdots(N-d)}{(d+1)!}.$$

We can also derive the d -dimensional degree distribution $p(k_d)$. Given an arbitrarily chosen node, we have $\binom{N-1}{d}$ possible d -simplices that the node can be a member of, hence the probability of the node having degree $k_d \leq \binom{N-1}{d}$ (i.e, it being a part of k_d d -simplices) is given by the binomial distribution

$$p(k_d) = \binom{\binom{N-1}{d}}{k_d} q_d^{k_d} (1 - q_d)^{\binom{N-1}{d} - k_d}. \quad (9)$$

Given that each d -simplex connects $d+1$ nodes, and there are N nodes in total, we have that the average degree is

$$\langle k_d \rangle = \frac{d+1}{N} \langle |Y_d| \rangle = q_d \frac{(N-1)\cdots(N-d)}{d!} \quad (10)$$

³The derivations completed and results regarding the degree distributions of the random d -complex are, as far as the author is aware, original.

Clearly the corresponding properties for the Linial-Meshulam complex are found by setting $d = 2$. We can indeed verify these results numerically, as shown in figure 4.

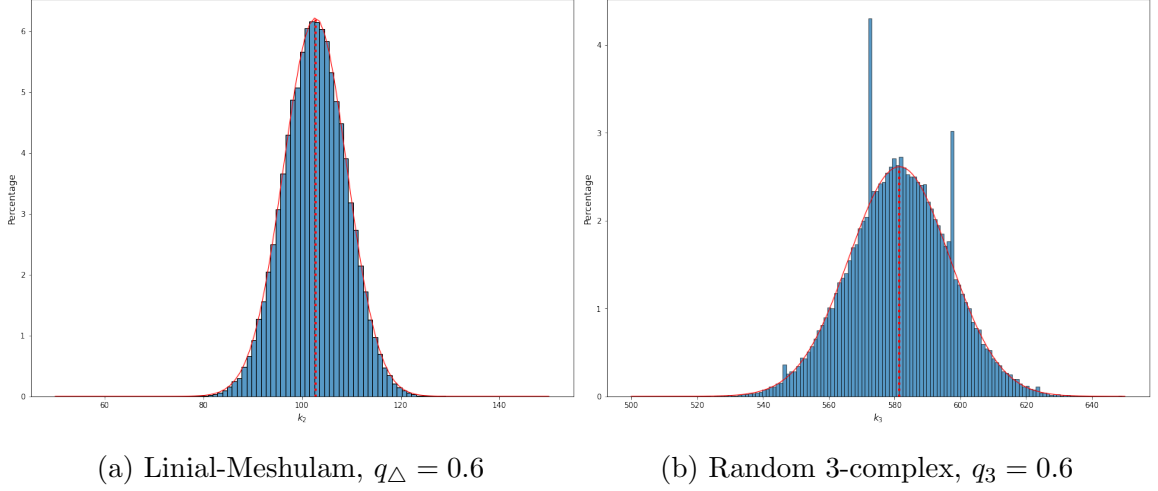


Figure 4: Degree distributions for $N = 100$. We have the histograms formed from 10^6 simulations in each case, along with the degree distributions given by (9) and the average degree given by (10).

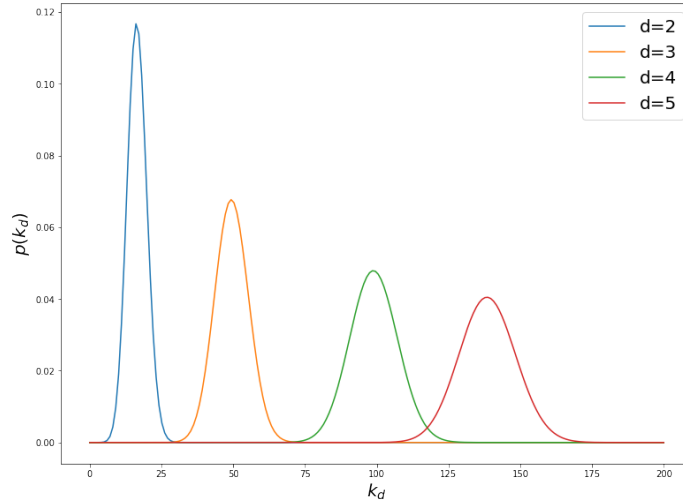


Figure 5: Degree distributions for increasing simplicial complex dimension. Here $q_d = 0.3$ for $d = 2, 3, 4, 5$ and $N = 12$.

4.4 Random Clique Complex

This model is the other natural approach to generalising the ER random graph. We start with the ER random graph $G(N, q)$ and simply ‘fill in’ all possible $(k + 1)$ -cliques

for $k = 1, \dots, N - 1$ [2]. By this we mean that wherever a $(k + 1)$ -clique is formed in the generated ER random graph, we convert it into its corresponding k -simplex. We denote this simplicial complex by $X(N, q)$, where N and q are simply the inputs of the underlying ER random graph. We note that this model is fully defined by its underlying graph.

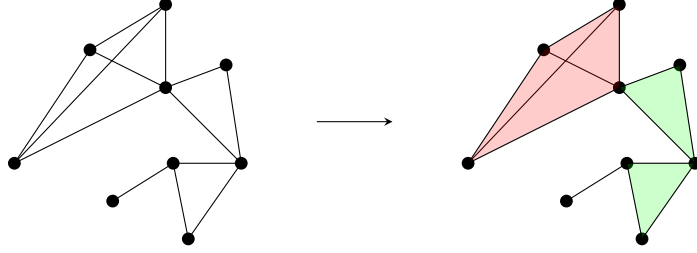


Figure 6: An example of the clique complex of a given graph. All $(k + 1)$ -cliques in the original graph (left) are changed to k -simplices in the simplicial complex (right). In the random clique complex the original graph is an ER random graph.

Much of the literature regarding this construction considers the topological properties analogous to the connectivity of networks. This includes generalising connectivity thresholds⁴ of ER graphs to homological thresholds of these complexes, i.e, conditions on N and q such that the resulting complex has a trivial k th homology group⁵. We omit these results here but refer the interested reader to [5] where Kahle generalises connectivity threshold theorems by Erdős and Rényi.

4.5 Multi-Parameter Complexes

Both random d -complexes (including the Linial-Meshulam complex) and random clique complexes can be seen as special cases of multi-parameter complexes. Denoted $X(N, \mathbf{q})$, these complexes can be seen as the most general case of the ER graph-based complexes. Here \mathbf{q} denotes a $(N - 1)$ -dimensional vector containing probabilities $q_i \in [0, 1]$ for $i = 1, \dots, N - 1$. Here, we start with N (affinely independent) nodes and connect pairs of nodes independently with probability q_1 to construct a graph. Next, we fill in 3-vertex cliques with 2-simplices independently with probability q_2 and continue in this fashion. In general, at step i we fill in $(i + 1)$ -vertex cliques with i -simplices independently with probability q_i [2]. For the Linial-Meshulam complex we start with a complete graph on N nodes, hence $p_1 = 1$. We then include

⁴These are conditions on N and q such that the resultant graph is connected with high probability.

⁵We refer the reader to Appendix A where homology groups are discussed.

2-simplices with probability q_Δ , hence we have

$$Y(N, q_\Delta) = X(N, \{1, q_\Delta, \underbrace{0, 0, \dots, 0}_{N-3}\}).$$

It follows that, in general

$$Y_d(N, q) = X(n, \{\underbrace{1, 1, \dots, 1}_{d-1}, q, \underbrace{0, 0, \dots, 0}_{N-d-1}\}).$$

And for random clique complexes, we start with $G(N, q)$ as the underlying 1-skeleton and fill in all possible d -vertex cliques for $d = 3, \dots, N-1$ with probability 1. Hence we have

$$X(N, q) = X(N, \{p, 1, \dots, 1\}).$$

4.6 Random Simplicial Complex (RSC) Model

An interesting variation of these ER based models is the RSC model as described in [4] which we note cannot be expressed in the form $X(N, \mathbf{q})$. This is because it includes k -simplices depending on the existence of any set of $k+1$ nodes, rather than depending on the existence of an underlying $(k+1)$ -vertex clique in the complex. Hence at each step new edges can be added to form simplices, rather than randomly promoting already existing cliques to simplices. We denote a D -dimensional RSC by $R_d(N, \mathbf{q})$, where N is again the number of initial nodes and \mathbf{q} describes a set of D probability values; $\mathbf{q} = \{q_1, \dots, q_D\}$ where $q_i \in [0, 1]$ for all $i = 1, \dots, D$. Here, similarly to the multi-parameter model, each q_k controls the creation of k -simplices, what we find to differ in this model is the way in which they are created. Clearly by this notion we require $D \leq N-1$. In step 1 of this model, we simply create an ER graph on the N nodes with probability parameter q_1 . Then for step i we independently connect sets of $(i+1)$ nodes with (i) -simplices with probability p_i . It is important to note the key difference between this model and the previously mentioned models: at each step we now consider all combinations of $i+1$ nodes and add simplices accordingly rather than only considering all $(i+1)$ -vertex cliques and filling them in accordingly, as such new edges can be added at each step and therefore each subsequent step increases the expected degree of simplices included in the previous step. To demonstrate this, we elaborate on the derivation [4] of $\langle k_1 \rangle$ and $\langle k_2 \rangle = \langle k_\Delta \rangle$ for the case $D = 2$.

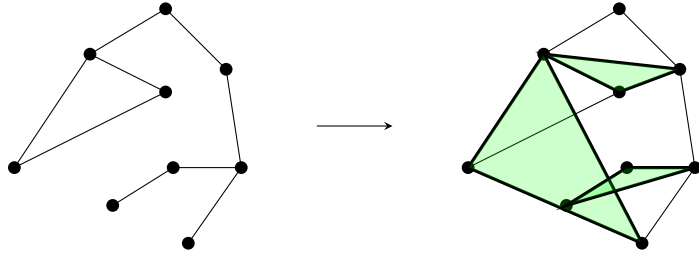


Figure 7: An example of a random simplicial complex formed from an ER random graph. We note the inclusion of new edges in the formation of the simplicial complex as the defining feature of this model.

On the first step of the model, we form the ER graph on N nodes with probability parameter q_1 and so, as stated in our previous discussion of ER graphs, we have

$$\langle k_1^* \rangle = q_1(N - 1)$$

where $*$ denotes that this is not the final value of $\langle k_1 \rangle$. Now on the second step, we add 2-simplices between triplets of nodes independently with probability $q_2 = q_\Delta$. In this case the expected degree of 2-simplices on any nodes is simply given by (10) for $d = 2$, i.e.

$$\langle k_\Delta \rangle = q_\Delta \frac{(N - 1)(N - 2)}{2}.$$

Now implicitly in this step we have also increased $\langle k_1^* \rangle$ since we have added edges to form triangles. To work out by exactly how much we have increased $\langle k_1^* \rangle$ we first consider the possible ways in which a node becomes part of a triangle. For each of these we state how they would increase $\langle k_1^* \rangle$ and the probability in which they occur. From this we can find the expected increase in $\langle k_1^* \rangle$. Taking any node i , there are 3 ways in which it can be included in a triangle, we denote, without loss of generality the two other nodes to be j and k and denote by k_{1_i} the degree of i .

1. (i, j, k) already exists $\mapsto \Delta k_{1_i} = 0$
2. Only one of (i, j) and (i, k) already exist $\mapsto \Delta k_{1_i} = 1$
3. (j, k) already exists $\mapsto \Delta k_{1_i} = 2$

We denote the probability of increasing k_{1_i} by h after step 2 of the model by $\mathbb{P}(\Delta k_{1_i} = h)$. We can note that option 1 does not increase the degree and so we can ignore it in our calculations. Option 2 requires that exactly one of the two other edges are present, hence $\mathbb{P}(\Delta k_{1_i} = 1) = 2 \times q_1 \times (1 - q_1)$ (the factor of two accounting for both

options). Option 3 requires that no edges from i are present, hence we require that (i, j) and (i, j) do not exist, this gives $\mathbb{P}(\Delta k_{1_i} = 2) = (1 - q_1) \times (1 - q_1)$. It follows that the expected increase in degree per included triangle is

$$\begin{aligned}\mathbb{E}(\Delta k_{1_i}) &= 1 \times \mathbb{P}(\Delta k_{1_i} = 1) + 2 \times \mathbb{P}(\Delta k_{1_i} = 2) \\ &= 2(1 - q_1).\end{aligned}$$

And we expect $\langle k_\Delta \rangle$ triangles to be included, so we expect to increase the average degree by $\langle k_\Delta \rangle \times 2(1 - q_1)$. It follows that

$$\begin{aligned}\langle k_1 \rangle &= \langle k_1^* \rangle + 2\langle k_\Delta \rangle(1 - q_1) \\ &= q_1(N - 1) + 2\langle k_\Delta \rangle(1 - q_1).\end{aligned}$$

Though this model is not perhaps as well established as the previously mentioned models, it proves to be convenient for computing purposes for dynamical processes, as shown in the simplicial contagion model, see [4].

5 A Dynamical Application of Simplicial Complexes: Epidemic Modelling

Epidemic compartmental models are amongst the most commonly studied dynamical systems on classical networks. Such models were initially created as a way to model the spread of infectious diseases on a population and have been very successful in doing so. More recently higher-dimensional spreading has become a topic of interest due to its potential in increasing the accuracy of such models. Higher order compartmental models also play a key role in other social dynamics such as the spread of information throughout a population. For example, if we were considering the spreading of a rumour, perhaps the rate of spread would be further increased if we considered neighbouring nodes to be dependent on each other - the rumour may spread faster if it's coming from a group. Classical networks only account for pairwise interactions and thus miss out on these potentially important higher-order interactions. In this section, we first review the SIR compartmental model on standard networks, and then we move onto consider SIR dynamics on some of the stochastic models of simplicial complexes presented in Section 4.

5.1 SIR Dynamics on Networks

In SIR models, each node belongs to one of three states at any given time - susceptible (S), infected (I) or recovered/removed(R). The distinction between recovered and removed depends on the contagion being modelled, for brevity we will generally refer to this compartment as recovered. Nodes belonging to the susceptible compartment are susceptible to becoming infected, and this may happen if they are adjacent to at least one infectious node. A susceptible node adjacent to a single infected node will become infected at a rate β , it follows that if the susceptible node is adjacent to m infected nodes this rate becomes $m\beta$. We can rephrase the infection rate β into a probability of infection in some small time step Δt , for a sufficiently small time step the probability that the susceptible node becomes infected is $\beta\Delta t$. In the case the node becomes infected, it will be removed from the susceptible compartment and will be added to the infected compartment. If a node belongs to the infected compartment, then as already stated it can infect adjacent susceptible nodes, but it can also become recovered from the infection. A given infected node will recover at some rate μ , similarly to the infection rate this can be rephrased as a probability in a sufficiently small time step; $\mu\Delta t$. If an infected node recovers then it is removed from the infected compartment and added to the recovered compartment. Nodes in the recovered compartment will remain there. It follows that given any infection rate β , so long as $\mu > 0$, eventually there will be no infected nodes⁶.

If we consider a network consisting of N total nodes, and denote by S , I and R the proportion of nodes that are susceptible, infected and recovered respectively, we have the following governing system:

$$\begin{aligned}\frac{dS(t)}{dt} &= -\beta S(t)I(t), \\ \frac{dI(t)}{dt} &= \beta S(t)I(t) - \mu I(t), \\ \frac{dR(t)}{dt} &= \mu I(t).\end{aligned}\tag{11}$$

Where we specify initial conditions $I(0) = I_0$, $S(0) = 1 - I_0$, $R(0) = 0$. Typically, we choose a small proportion of initially infected $I_0 \ll 1$. Clearly 0 nodes should initially be recovered and so to keep a constant population we require the stated

⁶We could rephrase this as all nodes will eventually recover, however this is assuming that the network is connected, otherwise there may be susceptible nodes that never become infected and so will never recover.

initial condition on S .

It is important to note the assumptions we have made when constructing this model, as these also apply for higher order spreading. Firstly, we have neglected births and deaths in the population. We could justify this assumption by either claiming that birth rate \approx death rate, or by claiming that we are considering dynamics on an overall timescale which is much less than that required for births and deaths to have a significant effect. An implicit consequence of this assumption is that we have a constant population size N , and so $S(t) + I(t) + R(t) = 1$ for all times⁷. Finally, we have assumed that the population is well-mixed. This means that the underlying network is homogeneous and any two distinct nodes are adjacent with equal probability. Such assumptions allow for the deterministic system (11), which though an idealised model, can provide very useful insight into predicting the spread of contagions on real populations.

5.2 SIR Dynamics on Simplicial Complexes

We now consider SIR dynamics on stochastic models of simplicial complexes. These models follow the same rules as on networks, however now we generalise the infection process to beyond the pairwise nature of classical networks. On a given simplicial complex we will still consider the nodes (0-simplices) as agents that belong to each compartment, but the process by which a susceptible node becomes infectious is extended. Given an arbitrary susceptible node, we now have distinct infection rates for each dimension of simplex it can be a part of. Namely in a d -dimensional simplicial complex \mathcal{X} , we have d infection rates $\{\beta_1, \dots, \beta_d\}$. Each β_i represents the rate of infection that an i -simplex has on an adjacent 0-simplex, given that the 0-simplex is susceptible and the remaining i 0-simplices that constitute the i -simplex are infected.

5.2.1 Governing Equations

We adapt derivations completed in [7] to our model. In doing so we shall make use of the adjacency tensor representation (7). We consider the probability that a given node n_i is in each of the possible compartments at a given time t . We denote these as P_i^S , P_i^I and P_i^R for susceptible, infected and recovered respectively. In doing this we need to check which simplices n_i belongs to - at each intermediate dimension $d \leq D$ we have $\binom{N}{d+1}$ possible combinations of d -simplices that n_i can be a member of.

⁷This is evident from the system of ODEs as they sum to 0, hence $S(t) + I(t) + R(t)$ is constant. The constant is found to be 1 from the prescribed initial conditions.

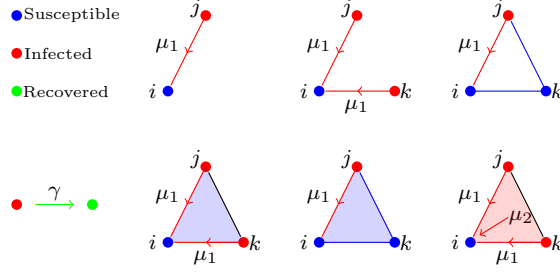


Figure 8: Illustration of possible interactions for $D = 2$. We note that only interactions that act to change node i are considered (in the actual processes, infected node j being edge adjacent to susceptible node k would also give a probability of transition), as well as the recovery process of nodes in general. This is an adaptation of the figure used for the SIS model considered in [4].

We first consider the case for $D = 2$ (i.e, up to triangles) to gain the intuition behind the derivation, and then provide a generalisation of this for arbitrary $D \geq 1$. For the case $D = 2$ we note the possible ways in which a susceptible node becomes infected. Of course, it can be infected if it is adjacent to a single infected node via an edge, however now we also have the case where it can become infected due to it being adjacent to a 2-simplex in which both other nodes of the 2-simplex are infected. If n_k and n_l are two arbitrary nodes, then n_i is adjacent to the 2-simplex $\{n_i, n_k, n_l\}$ if it exists in the complex, i.e, if $(\mathcal{A}_2)_{ikl} = a_{i,k,l} = 1$. Similarly, we only consider possible infections via edges if $(\mathcal{A}_1)_{ij} = a_{i,j} = 1$. We have that n_i is adjacent to

$$\sum_{\{i,k,l\} \in \mathcal{X}} a_{i,k,l}$$

triangles and

$$\sum_{\{i,j\} \in \mathcal{X}} a_{i,j}$$

edges. We note that in the above summations i is held constant since we are considering effects on a single node n_i , hence we sum over the i th face of the $N \times N \times N$ tensor \mathcal{A}_2 , and over the i th row of the $N \times N$ matrix \mathcal{A}_1 . In our model we have that infections can occur if a susceptible node is adjacent to an infected node via an edge with probability μ_1 in a small timestep Δt . Similarly, if the node is adjacent to a triangle where both other nodes are infected, then on top of the infection probabilities from the adjacent edges from each node, we have an extra contribution from the triangle. The triangle infects the node with probability μ_2 in the timestep. It follows

that the governing equation for the rate of change of the state probability P_i^S is

$$\frac{dP_i^S(t)}{dt} = -(\underbrace{\mu_1 P_i^S(t) \sum_{\{i,j\} \in \mathcal{X}} a_{i,j} P_j^I(t)}_{\text{1-simplex contribution}} + \underbrace{\mu_2 P_i^S(t) \sum_{\{i,k,l\} \in \mathcal{X}} a_{i,k,l} P_k^I(t) P_l^I(t)}_{\text{2-simplex contribution}}) \quad (12)$$

since for an infection to occur we require n_i to be susceptible, which occurs with probability P_i^S , and any edge-adjacent node n_j to be infected which happens with probability P_j^I . For triangles we require both other nodes n_k and n_l to be infectious, which happens with probability $P_k^I P_l^I$. Noting that recovery is only a node dependent process, and letting γ be the probability that an infected node becomes recovered in the timestep, the remaining equations follow naturally:

$$\frac{dP_i^I(t)}{dt} = -\gamma P_i^I(t) + \mu_1 P_i^S(t) \sum_{\{i,j\} \in \mathcal{X}} a_{i,j} P_j^I(t) + \mu_2 P_i^S(t) \sum_{\{i,k,l\} \in \mathcal{X}} a_{i,k,l} P_k^I(t) P_l^I(t) \quad (13)$$

and

$$\frac{dP_i^R(t)}{dt} = \gamma P_i^I(t). \quad (14)$$

It follows that for arbitrary dimension $d \leq D$, these equations generalise to the dynamical system

$$\begin{aligned} \frac{dP_i^S(t)}{dt} &= -P_i^S(t) \sum_{d=1}^D \mu_d \sum_{\{i,j_1,\dots,j_d\} \in \mathcal{X}} a_{i,j_1,\dots,j_d} \prod_{k=1}^d P_{j_k}^I \\ \frac{dP_i^I(t)}{dt} &= -\gamma P_i^I(t) + P_i^S(t) \sum_{d=1}^D \mu_d \sum_{\{i,j_1,\dots,j_d\} \in \mathcal{X}} a_{i,j_1,\dots,j_d} \prod_{k=1}^d P_{j_k}^I \\ \frac{dP_i^R(t)}{dt} &= \gamma P_i^I(t) \end{aligned} \quad (15)$$

Since for each higher dimension we have an additional node that is required to be in the infected state, hence the necessity for the product in (15).

5.2.2 Mean Field Approach

Mean field theories simplify the model to allow for a form of deterministic solution that can be used as at least a sanity check for numerical simulations on stochastic models. We approximate the system (15) using such a mean field approach, where we consider the dynamics on a fully connected simplicial complex, and with all nodes having equal probabilities of being in a given state at any given time, i.e, $P_i^{S,I,R}(t) = P_j^{S,I,R}(t) = P^{S,I,R}(t)$ for any distinct nodes n_i and n_j . Since all nodes are equally

likely to belong to any state, we can take the exact probabilities at some time t simply as the proportion of nodes in each state at this time. Namely we let $P^S = S$, $P^I = I$ and $P^R = R$, where S , I and R are the time dependent proportions of nodes in the susceptible, infected and recovered compartments respectively. We have the simplified dynamical system

$$\begin{aligned}\frac{dS(t)}{dt} &= - \sum_{d=1}^D \mu_d \langle k_d \rangle S(t) I^d(t) \\ \frac{dI(t)}{dt} &= -\gamma I + \sum_{d=1}^D \mu_d \langle k_d \rangle S(t) I^d(t) \\ \frac{dR(t)}{dt} &= \gamma I\end{aligned}\tag{16}$$

where $\langle k_d \rangle$ is the average degree of d -simplices. Along with this system we prescribe initial conditions $S(0) = 1 - I_0/N$, $I(0) = I_0/N < 1$ and $R(0) = 0$, where N is the total number of nodes. We note that from now on we strictly consider the case $D = 2$. Some important properties of the mean field system (16) are derived by Palafox-Castillo and Arturo Berrones-Santos in their 2022 paper [6] which we explain and slightly change to allow for an arbitrary number of initially infected nodes. If we enforce that $dI/dt > 0$, simply rearranging the second of the equations gives the condition $S > \gamma/(\mu_1 \langle k_1 \rangle + \mu_2 \langle k_\Delta \rangle I)$ at any given time. In particular, if we want this to hold at $t = 0$, when $S = 1 - I_0/N$ and $I = I_0/N$, we end up with the condition that

$$\gamma < \mu_1 \langle k_1 \rangle + \frac{I_0}{N} (\mu_2 \langle k_\Delta \rangle - \mu_1 \langle k_1 \rangle) - \frac{I_0^2}{N^2} \mu_2$$

which, for large N ⁸, gives a rough condition for the epidemic to grow initially⁹, namely

$$\gamma < \mu_1 \langle k_1 \rangle.\tag{17}$$

We denote by γ^* the critical value $\gamma^* = \mu_1 \langle k_1 \rangle$. Finally, setting $dI(t^*)/dt = 0$ implies that either $I = 0$, which indicates the long-time steady state, or

$$I(t^*) = \frac{\gamma - \mu_1 \langle k_1 \rangle S(t^*)}{\mu_2 \langle k_\Delta \rangle S(t^*)}\tag{18}$$

which indicates the peak of the epidemic at time t^* [6].

⁸We note that we have prescribed initial conditions which explicitly include the total number of nodes N as opposed to the governing equations of standard SIR models, where this value is implicit in I_0 . We could alternatively make N implicit in the initial conditions and consider small I_0 in this step.

⁹This equation implies initial epidemic growth as it is the requirement for $\frac{dI(0)}{dt} > 0$.

5.2.3 Numerical Simulations

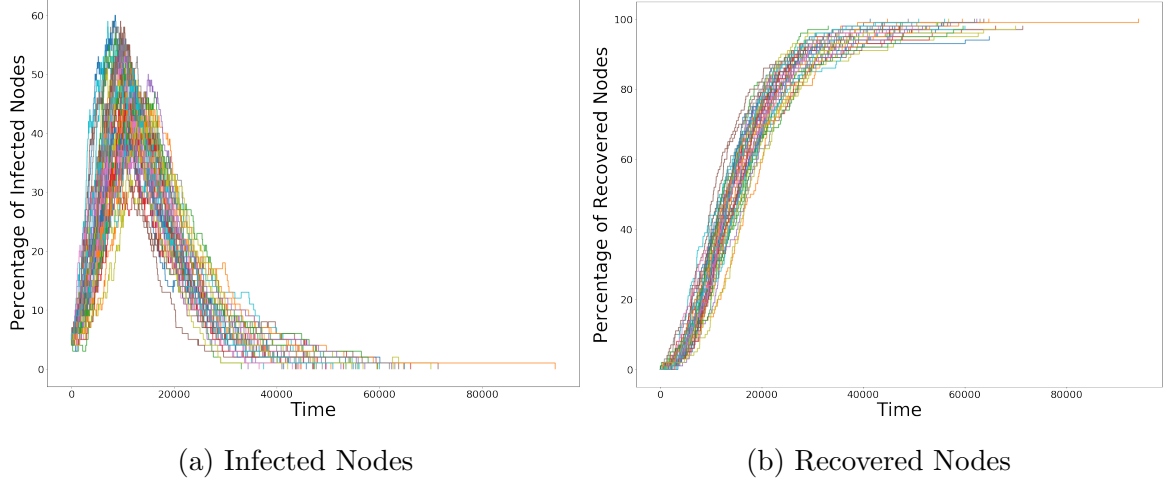


Figure 9: SIR dynamics on the Linial-Meshulam simplicial complex model $Y(100, 0.001)$, the resulting graph has $\langle k_1 \rangle = 99$ and $\langle k_\Delta \rangle = 4.89$. Other parameters: $N = 100$, $\mu_1 = 5 \times 10^{-6}$, $\mu_2 = 10^{-5}$, $\gamma = 0.25 \times \mu_1 \langle k_1 \rangle = 1.24 \times 10^{-4}$. We plot results for 50 simulations on the same simplicial complex which randomly selects 5% of nodes to infect initially .

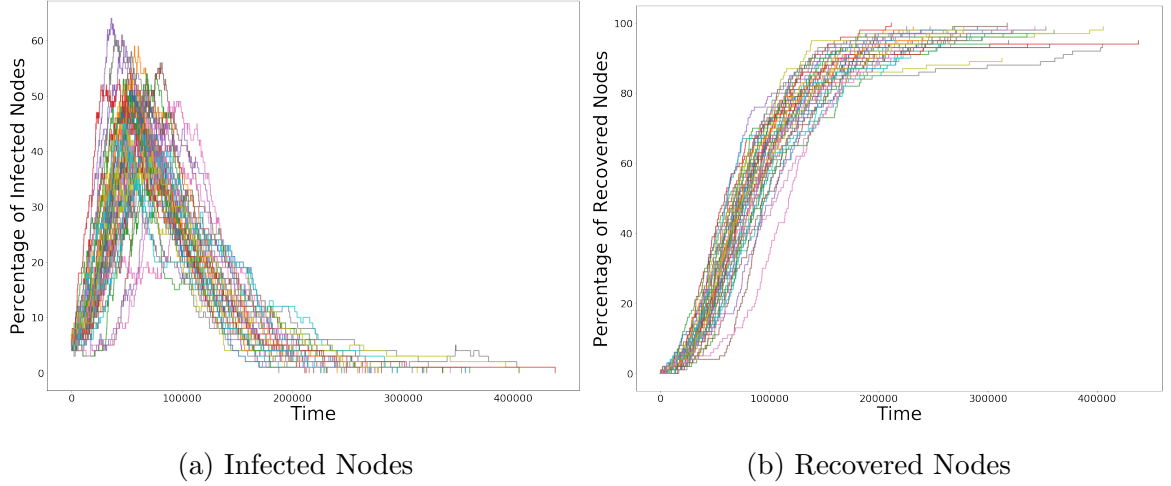


Figure 10: SIR dynamics on the Random Simplicial Complex model $R_2(100, 0.1, 0.001)$, the resulting graph has $\langle k_1 \rangle = 18.18$ and $\langle k_\Delta \rangle = 4.74$. Other parameters: $N = 100$, $\mu_1 = 5 \times 10^{-6}$, $\mu_2 = 10^{-5}$, $\gamma = 0.25 \times \mu_1 \langle k_1 \rangle = 2.27 \times 10^{-5}$. We plot results for 50 simulations on the same simplicial complex which randomly selects 5% of nodes to infect initially .

We simulate the spread of a contagion on the RSC and Linial-Meshulam models¹⁰. From 9 and 10 we see the SIR dynamics on two of our models. We note that in the simulations for the Linial-Meshulam model we see a faster peak in infected nodes at $t \approx 1.2 \times 10^4$ as compared to the RSC model, from which we see a peak at $t \approx 8 \times 10^4$. This is indicative of the lower edge degree for the RSC model, which is perhaps a shortcoming of the Linial-Meshulam model for such dynamics, since it will always have maximal degree $\langle k_1 \rangle = N - 1$. The relative importance of μ_1 and μ_2 are perhaps overlooked by these simulations. Instead, to see how changing these infectivity parameters effects the overall epidemic we plot heat maps of different values of μ_1 and μ_2 against peak infections in the mean field approximation given by (16).

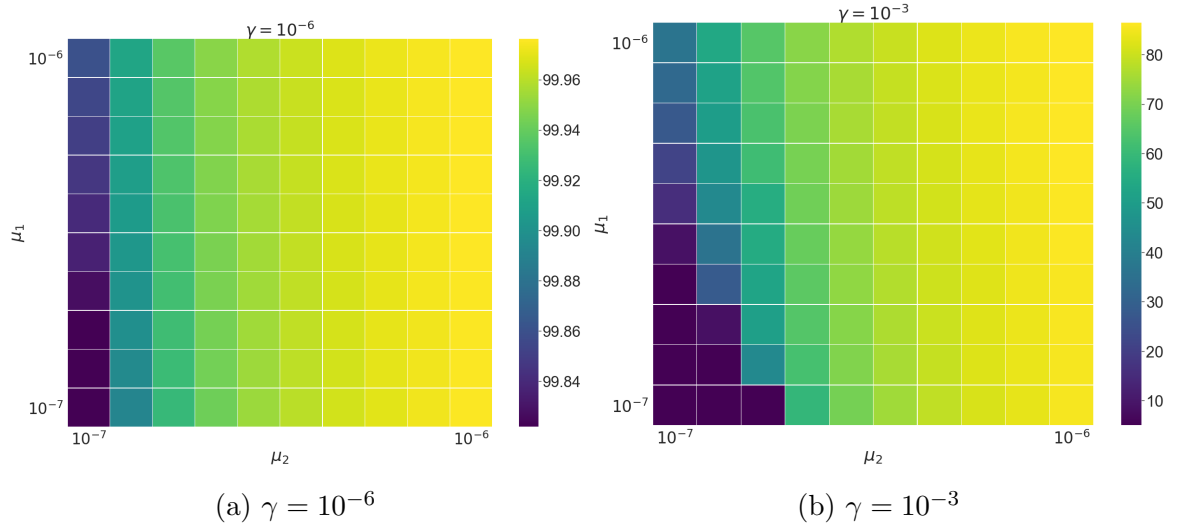


Figure 11: Heat maps showing the percentage of the peak infected for varying μ_1 and μ_2 in the mean field approximation. Parameters: $k_1 = 20$, $k_\Delta = 6$, $S(0) = 0.95$, $I(0) = 0.05$, $R(0) = 0$.

We see from figure 11 the relative contributions of μ_1 and μ_2 on the peak infections caused by 5% initially infected nodes. Given the parameters we have chosen, we can use the critical condition for epidemic growth from (17) to see that when $\mu_1 = 5 \times 10^{-7}$ (the average of the values we are sampling from) we have $\gamma^* = 20 \times 5 \times 10^{-7} = 10^{-5}$. This justifies the choices of $\gamma = 10^{-6}$ and $\gamma = 10^{-3}$ to investigate the relative importance of μ_1 and μ_2 with regard to the prevalence of an epidemic outbreak.

¹⁰Simulations are completed largely using the NetworkX package for Python, see <https://networkx.org/>. We also adapt some of the code used in [4] for the construction of the RSC model, however we note that SIR dynamics on the Linial-Meshulam complex have not been considered in the literature, as far as the author is aware.

We notice that when $\gamma = 10^{-6}$, we see in all cases that the peak infected reaches nearly 100% of the population. This is a consequence of the very high infection probability in comparison to the recovery probability. It is interesting to note that even at this resolution, we see a clear favouring of infections with higher μ_2 relative to similar changes in μ_1 , indicating the importance of higher order interactions in such models. For the case $\gamma = 10^{-3}$ we indeed see a better spread of peak infected, as is expected when we are slightly above the epidemic threshold. In this case we see similar contributions to peak infectivity from both μ_1 and μ_2 since the gradient roughly increases from bottom left to top right. An interesting point to note here is the slight favouring in the importance of μ_2 , as the gradient increases more towards the lower axis. This result further motivates the importance of higher order interactions. An interesting extension would be to consider the implications of higher dimensional simplicial complexes but due to scaling of the binomial coefficient $\binom{N}{d+1}$, assigning simplices in the complex becomes much more demanding for higher dimensions.

6 Concluding Remarks

We have reviewed how we can extend classical graphs to simplicial complexes which encompass higher order interactions and have shown how representations of graphs can be generalised to simplicial complexes, discussing where appropriate how these different representations capture the connectivity of a simplicial complex. Through a detailed review of the literature, we have shown several ways in which the ER random graph can be extended. In particular we have derived the general degree distributions for the random d -complexes and have verified the results numerically. We finally considered a dynamical application of simplicial complexes through epidemic modelling, through both a computational and mean-field approach. We identified the prevalence of higher order interactions in our analysis and considered simulations of spreading on the Linial-Meshulam and RSC models. It would be an interesting extension to consider the other stochastic models mentioned, and also to compare the dynamics of the system with regard to the dimension of the model used. It should be noted that the use of simplicial complexes goes far beyond their abilities to capture many body interactions; we could instead consider how the shape of data changes in a temporal simplicial complex. Such notions are fundamental in the rapidly growing field of topological data analysis.

References

- [1] Federico Battiston et al. “Networks beyond pairwise interactions: Structure and dynamics”. In: *Physics Reports* 874 (2020). Networks beyond pairwise interactions: Structure and dynamics, pp. 1–92. ISSN: 0370-1573. DOI: <https://doi.org/10.1016/j.physrep.2020.05.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0370157320302489>.
- [2] Omer Bobrowski and Dmitri Krioukov. “Random simplicial complexes: Models and phenomena”. In: *arXiv preprint arXiv:2105.12914* (2021).
- [3] Timothy E Goldberg. “Combinatorial Laplacians of simplicial complexes”. In: *Senior Thesis, Bard College* (2002).
- [4] Iacopo Iacopini et al. “Simplicial models of social contagion”. In: *Nature Communications* 10 (June 2019), p. 2485. DOI: 10.1038/s41467-019-10431-6.
- [5] Matthew Kahle. “Topology of random clique complexes”. In: (2006). DOI: 10.48550/ARXIV.MATH/0605536. URL: <https://arxiv.org/abs/math/0605536>.
- [6] Gerardo Palafox-Castillo and Arturo Berrones-Santos. “Stochastic epidemic model on a simplicial complex”. In: *arXiv preprint arXiv:2203.06241* (2022).
- [7] Dong Wang et al. “Simplicial SIRS epidemic models with nonlinear incidence rates”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 31.5 (2021), p. 053112. DOI: 10.1063/5.0040518. eprint: <https://doi.org/10.1063/5.0040518>. URL: <https://doi.org/10.1063/5.0040518>.

A Simplicial Homology

When considering classical networks, so often we are concerned about connectedness and in particular the notion of connected components. We can generalise this notion to these higher order representations by finding homology groups of simplicial complexes. Informally, simplicial homology gives a procedural method to finding the number of holes or voids in a simplicial complex. Without going into the technical details of algebraic topology, we build up to a definition of homology groups for simplicial complexes.

Let \mathcal{X} be a simplicial complex. We define a k -chain c_k of \mathcal{X} to be

$$c_k = \sum_i a_i \sigma_i \quad (19)$$

where $\sigma_i \in \mathcal{X}^k$ is an oriented simplex, and $a_i \in \mathbb{F}$, where \mathbb{F} is some field. Firstly, by oriented we simply mean that summing the oriented simplex with the same simplex with opposite orientation will give the empty set. Secondly, though the choice of \mathbb{F} is arbitrary since this construction can be very general, we pick $\mathbb{F} = \mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$ to simplify the derivation. This choice is justified because binary coefficients will signify if a simplex belongs to the k -chain or not; if $a_i = 1$ then σ_i is part of the k -chain, and if $a_i = 0$ then σ_i is not part of the k -chain. We denote by C_k the vector space describing the set of all possible k -chains of \mathcal{X} .

We next consider the boundary operator. This abstract operator can be seen as a tool to return the $(k-1)$ -simplices that constitute the boundary of some k -simplex. We first give an informal definition in the context of linear algebra. Since we are expecting to return some $(k-1)$ -simplices given some k -simplex, it follows that the boundary operator is defined by some linear map¹¹ $\partial_k : C_k \mapsto C_{k-1}$. Moreover, if we prescribe that \mathcal{X} has exactly N k -simplices and M $(k-1)$ -simplices ($|\mathcal{X}^k| = N$ and $|\mathcal{X}^{k-1}| = M$), then it follows that the boundary operator can be described as the linear map $\partial_k : \mathbb{Z}_2^N \mapsto \mathbb{Z}_2^M$. As such the boundary operator can be fully described as an $M \times N$ matrix with entries in \mathbb{Z}_2 . In general, we have that the boundary operator acting on some k -simplex $\sigma = [n_0, \dots, n_k]$ is given by the formal sum of its $k-1$ -dimensional faces, in particular we have

$$\partial_k(\sigma) = \sum_{i=0}^k (-1)^i [n_0, \dots, n'_i, \dots, n_k] \quad (20)$$

¹¹This map is actually a linear homomorphism as it retains the structure of the input simplex.

where n'_i denotes that this node is to be removed. Typically, we let the orientation of the simplex be described by the ordering of the nodes in a counter-clockwise direction. To close the definition, we enforce that the boundary of a node is trivial, $\partial_0 \equiv 0$. We also note that the boundary operator only returns the boundary of simplices of exact dimension $k-1$, this implies that the boundary of a boundary is trivial: $\partial_{k-1}\partial_k(\sigma) = 0$. The final step before defining simplicial homology groups is to consider the notion of k -dimensional cycles and boundaries. We define the space of k -cycles as

$$Z_k = \ker(\partial_k) \quad (21)$$

and the space of k -boundaries as

$$B_k = \text{im}(\partial_{k+1}). \quad (22)$$

We can think of $Z_k \subset C_k$ as the space of chains that do not have a boundary, and B_k as the space of k -simplices that are boundaries of higher dimensional objects. Following from the idea that the boundary of a boundary is trivial, it is clear that the simplices that are boundaries must also be simplices that do not have boundaries¹², i.e. $B_k \subset Z_k$. We can finally define the k -th homology group of a simplicial complex \mathcal{X} of dimension at least k . Denoted H_k , the k th homology group is defined as the quotient group

$$H_k = Z_k / B_k. \quad (23)$$

We also define the k th Betti β_k number as the dimension of this:

$$\beta_k = \dim(H_k). \quad (24)$$

Notice that the k th homology group is nothing but the space of all cycles of dimension k modulo those that are boundaries of higher dimensional objects. What we are left with is k -dimensional voids, and the Betti number simply counts how many voids we have. These abstract notions have significant application in topological data analysis. Essentially they can be used to distinguish geometric features of data, which can provide unique insight. Though we do not go into significant details regarding their usage, we could consider a sequence of simplicial complexes parameterised by some distance parameter δ . We can define persistent homology through the temporal analogue of the Betti number, namely barcodes. These are commonly used in topological data analysis to distinguish features of data sets from noise. Longer bars represent

¹²Evidently not having a boundary does not imply that the object is a boundary, so the converse does not necessarily hold

higher persistence given the distance parameter and therefore are more likely to be features, and shorter bars are more likely to represent noise. The features in question here are simply holes/voids in the data set.