A Framework for Simple and Complex Contagion on Clustered Networks and its Implications

Atul Ganju

Working Paper, December 2, 2024

Due to its increased relevance in economic market research, increased attention has been given towards understanding how behavior diffuses across a social network. This diffusion process is typically modeled as the spread of a contagion across graphs with local structure. In the last decade, independent experimentation has demonstrated that the diffusion of behavior across a social network is a *complex* contagion phenomenon and, as a result, there has been a recent push towards understanding complex contagion in these locally structured networks.

In this paper, we are the first to analyze complex contagion with discrete-time dynamics on graphs with rich local structure. Specifically, we develop a framework for analyzing complex contagion on a family of clustered networks. We then generate several models for complex contagion with this framework and evaluate their accuracy by comparison against large-scale numerical simulation. Using these models, we explore the relationship between network structure and various simple and complex contagion mechanisms. Our results show several interesting phenomena.

1. Simple contagion triggers complex contagion on clustered networks. In previous work, the same phenomenon has been observed to occur on random networks, we extend this result to clustered networks as well.

2. A contagion which diffuses solely through a complex contagion mechanism is more influential and spreads more rapidly on clustered networks. However, a contagion that diffuses through both simple and complex contagion mechanisms performs as well, if not better, on random networks than on clustered networks.

3. Current approximation schemes for modeling complex contagion with discrete dynamics, including our own, prove to have an error which increases with the ratio between a network's average degree and its size.

Finally, it is important to note that the framework developed in this paper can be used to generate complex contagion models with more complicated diffusion mechanisms and on larger graph families than previously possible.

Introduction

The propagation of information, the diffusion of disease, and even the spread of behavior are influenced by the structure of the networks they occur in. The first and second are well studied phenomena for which it was observed that their diffusion is most powerful on random networks, with locally tree-like structure, and that any social reinforcement would inhibit their diffusion process. Therefore, these phenomena are most accurately modeled as a simple contagion — a contagion where one successful exposure is enough to transmit the contagion to an uninfected node []. Alternatively, it has been determined that the spread of behavior is amplified by social reinforcement mechanisms and that its spread is more pronounced on clustered networks [Centola, 2010]. This implies that the spread of behavior is best modeled as a complex contagion

— a contagion where multiple successful exposures are needed to transmit the contagion to an uninfected node. This stresses the importance of understanding complex contagion on clustered networks and provides motivation for our results.

It has proven to be a challenge to analyze complex contagion processes. To do so, we take inspiration from the most elementary framework for analyzing simple contagion, the SI (susceptible-infected) framework. In this framework, each node is either susceptible or infected and once infected remains as such. Since, in complex contagion, nodes may have a larger threshold for adoption, there exist different states for each node depending on its threshold and the number of adopted neighbors it has. We can then define an analogous SI framework for complex contagion, where a node's state is dependent on the number of additional exposures it needs in order to induce its adoption. If this quantity is 0, the node is in the infected state, and if it is greater than 0 it is in the susceptible state.

First attempts to model complex contagion led to the development of an approximation scheme which determined the final fraction of adopted nodes after an SI complex contagion process over a random network. These studies use mean field approximation (MFA) and pair approximation (PA) techniques which approximate the inductive effect of a node's neighbors as an average over all nodes in the network. This approximation scheme has demonstrated its accuracy when compared against numerical simulation [Gleeson and Cahalane, 2007][Min and San Miguel, 2018]. Gleeson and Cahalane use MFA to generate a framework for studying the effect seed size has on the likelihood of a global cascade [Gleeson and Cahalane, 2007]. Extending this framework, Min and Miguel incorporate a mechanism for both arbitrary threshold distributions and probabilistic contagion and use it to investigate when simple contagion can trigger a complex contagion cascade [Min and San Miguel, 2018]. However, in general, since these techniques are limited to locally tree-like networks, they can't accurately model complex contagion on clustered networks.

To address the limitations of MFA methods, O'Sullivan et al. developed the CA (clique-approximation) framework which models complex contagion on clustered networks in continuous time [O'Sullivan et al., 2015]. Their framework makes it possible to approximate the early time rate of diffusion of a complex contagion on a family of clustered networks while also accurately predicting the fraction of adopted nodes at a given time. In this framework, transmission probability of the contagion increases linearly with time and, after some time, eventually reaches 1.

In contrast to the CA framework, ours builds on existing MFA methods and allows us to model complex contagion on a family of clustered networks in discrete time. Using this framework, we generate multiple analytic models to approximate the final fraction of adopted nodes after a different complex contagion process over clique-based networks, which are defined in the following section. Accuracy of our framework is then verified by comparison of the models' outputs against large-scale numerical simulations. With these models, we then explore the relationship between network structure and various simple and complex contagion mechanisms. Finally, we discuss the implications of our results. It is important to note that 1) our framework can be generalized to incorporate additional mechanisms for contagion such as an arbitrary transmission probability and 2) another key difference between the CA framework and ours is that ours can be applied to a larger class of graphs in which every node is not necessarily in an identical environment.

Clique-based Graphs

Complex contagion is more powerful in networks which allow for the influence of social reinforcement, i.e networks with clustering. To analytically model complex contagion on clustered networks, we restrict our attention to a family of networks where we can formally characterize the clustering. We will refer to these as clique-based networks. To understand which networks are in this family, we first define a node's clique environment. Specifically, if a node in the graph is in m cliques of size n, we succinctly denote its clique environment by the tuple (m, n). Then, the graphs in our family are ones where each node's clique environment is (m, n) for some fixed n and m sampled from a distribution $D(m)$.

Any clique-based network has a fixed clique size n, but it can be either non-homogeneous, where an arbitrary node is in m cliques — with m being sampled from $D(m)$ — or (m, n)-homogenous, where all nodes in the network are in exactly m cliques (refer to **Figure 1** for example homogeneous clique-based network topologies).

Although we develop a framework for complex contagion on non-homogeneous clique-based networks,

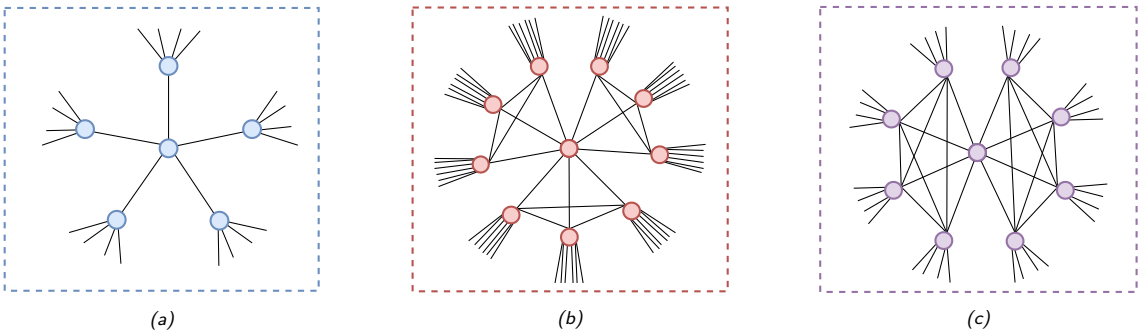

Figure 1: Clique motifs define network topology. In *(a)* we have a 5-regular clique-based network with motif (5,2), in *(b)* we have a 9-regular clique-based network with motif (3,4), and in *(c)* we have an 8-regular clique-based network with motif (2,5).

we only analyze complex contagion on (m, n)-homogeneous clique-based networks for the sake of simplicity. Finally, it is worth noting the following clique-based networks collapse to simpler graphs:

- If $n = 2$ and m is sampled from a distribution $D(m)$, then the network is a tree where the degree distribution of the nodes is given by D.
- If $n = k$ and $m = 1$, where $k \geq 1$, the network is a completely connected component of size k.

While we primarily focus on clique-based networks which contain cycles, we will also examine the model's accuracy on the aforementioned simple clique-based networks as it shows that our model collapses to the one previously studied [Gleeson and Cahalane, 2007][Min and San Miguel, 2018].

Contagion Framework

We present a framework for complex contagion on clique-based networks over N nodes where dynamics are in discrete time. As discussed before, our clique-based graph is constructed such that each node is in m cliques of size n, where m is sampled independently from some distribution $D(m)$ and n is constant for all nodes. To incorporate the complex contagion mechanism, a node's threshold for adoption is then sampled independently from some distribution $P(\theta)$ where θ represents the number of successful exposures required for a node to transition from the susceptible state to the state of adoption. A node transitions from the susceptible state to the adopted state when the number of adopted neighbors it has is at least its threshold for adoption. At each time step, nodes determine whether they transition from the susceptible state to the adopted state by counting their adopted neighbors and comparing it to their threshold. This process continues until no new nodes become adopted at which time we say the contagion has reached a steady state. Therefore, we are using the generalized SI model for complex contagion defined in the introduction implying nodes that transition to the adopted state remain in the adopted state. To initiate the complex contagion process, some fraction ρ of the nodes, called seeds, are initialized in the adopted state (i.e. they have threshold $\theta = 0$). It is worth noting that by simply taking the threshold distribution to be bimodal over the thresholds 0 and 1 we can model simple contagion on our d-regular clique-based networks.

Using the framework outlined in the next section, we develop self-consistent approximation equations to predict the final fraction of adopted nodes given $\rho, n, D(m)$, and $P(\theta)$. To ensure the validity of our framework, we extensively test it by generating models from it which span the parameter space and compare the predicted output of these models to the output of large-scale numerical simulations (details in our results section). Consequently, we observe a relationship between clustering and various mechanisms of simple contagion and complex contagion and provide further evidence that the diffusion of behavior over a social network is a complex contagion process.

Analytic Approach

We unify ideas from Gleeson and Cahalane's work on approximating the final fraction of adopted nodes on locally tree-like networks and Granovetter's work on threshold models for completely connected networks to overcome the limitations of current approximation methods. Specifically, we define a framework for modeling complex contagion on non-homogeneous clique-based networks and use it to instantiate various complex contagion models. We then analyze these models providing insight into the effect clusters have on the spread of information over a network.

We start by replacing our clique-based networks with one that has an underlying tree-like structure. First, pick a random node from our network to be the root and take it to be at the highest level, T, of our tree. Then, if the root r is in m_r cliques, it is connected to $d_r = m_r(n-1)$ different nodes, which we will take to be in the subsequent level of our tree. These nodes are then the children of our root node. Each child of the root is in a clique containing the root and $n-2$ other children of the root. We refer to these $n-2$ nodes the child's sibling nodes. Now, each of the children, $c_1, \ldots, c_{d_r}$, of the root are in $m_1, \ldots, m_{d_r}$ cliques respectively, meaning that they are connected to $(m_1)(n-1), \ldots, (m_{d_r})(n-1)$ unique nodes. Now, for each child, observe that we have already included the nodes from one of the cliques in the tree, namely the clique containing its siblings and the root. Thus, each of the children is connected to $(m_1-1)(n-1), \ldots, (m_{d_r}-1)(n-1)$ remaining nodes, which we include in the subsequent level of our tree. Continuing this pattern, we can construct a tree structure which underlies the structure of our clique-based network. If we now add edges to this tree between siblings at each level which are in the same clique, we recover a structure that can be used to analyze our clique-based network (see **Figure 2** for a visualization of d-homogeneous clique-based networks as trees).

With this, we can define our framework for complex contagion on clique-based networks. The general approach is to approximate, for an arbitrary susceptible node, the probability one of its neighbors is in the adopted state. We use MFA to accurately model this probability. Specifically, we assume the root is an arbitrary node, its children are in the state they will be in at the end of the complex contagion process, and the effect that each of the root's children can have on it is the average effect over all possible nodes. To approximate the probability that one of the root's neighbors is in the adopted state define a new quantity, q_t, to be the probability a node at level t of the tree has adopted the contagion given its parent has not adopted the contagion and everything up until that level has been updated correctly. Now, if we take the limit of q_t as t goes to infinity, we have an approximation for the probability that one of the root's neighbors is in the adopted state, which we denote as q_∞.

Now, we give a brief overview of the derivation for an expression for q_t. We start by splitting the probability that a node at level t of the tree becomes adopted into two cases: either the node at level t was a seed and started infected or it was not a seed and was infected by its neighbors. As the first case is trivial, let us consider the latter case. The probability that a node at level t becomes adopted is the probability its threshold θ is less than or equal to the sum of the number of adopted children it has, c, and the number of adopted siblings it has, s, in its clique at level t of the tree (i.e. a susceptible node becomes adopted if $c+s \geq \theta$). Now, given that the node has c adopted children and threshold θ, then if we define a quantity $\eta = \max(0, \theta - c)$, we can redefine q_t to be the probability that s is greater than or equal to η (i.e. a susceptible node becomes adopted if $s \geq \eta$). So, if we can solve for the probability distribution over all values of η at level t of the tree, call this quantity $Q_t(\eta)$, and solve for the probability that at least η siblings of a node are infected for any η, call this quantity $Z(\eta)$, then we can solve for q_t.

Lets define $D(m)$ to be the distribution which outputs the probability a node in our network is in m cliques. Then, we can derive an expression for $Q_t(\eta)$ in terms of q_{t-1}, an expression for $Z(\eta)$ in terms of Q_t, and in turn a mutually recursive set of equations for q_t:

$$Q_t(\eta) = \sum_{m=1}^{\infty} \frac{m(n-1)D(m)}{z} \sum_{\kappa=0}^{(m-1)(n-1)} \binom{(m-1)(n-1)}{\kappa} q_{t-1}^{\kappa}(1-q_{t-1})^{(m-1)(n-1)-\kappa} \begin{cases} P(\kappa+\eta) & \text{if } \eta > 0, \\ \sum_{\theta=1}^{\kappa} P(\theta) & \text{if } \eta = 0. \end{cases}$$

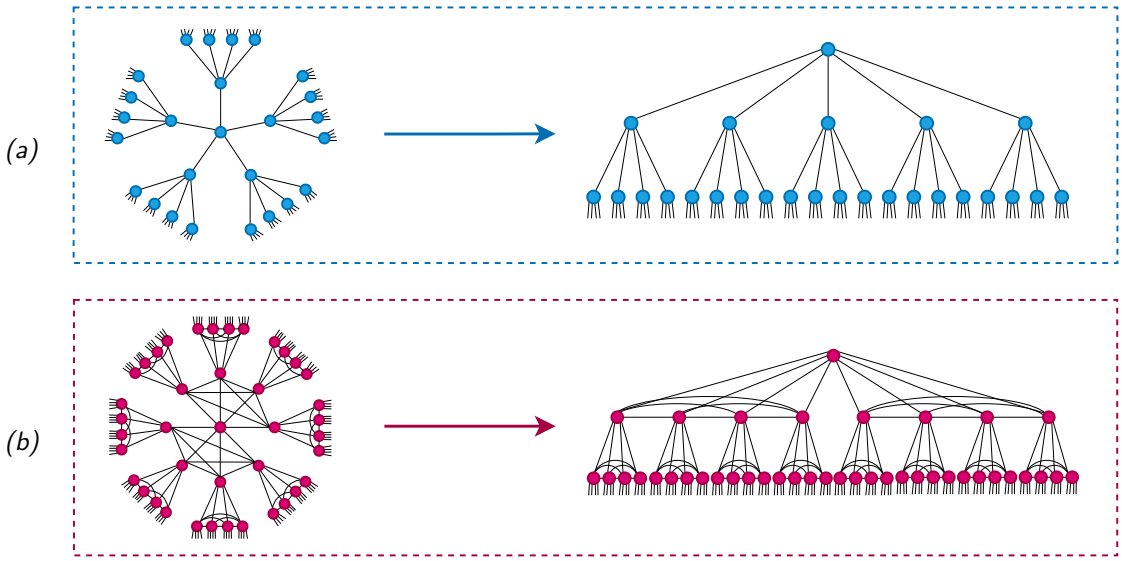

Figure 2: Clique-based networks have an underlying tree-like structure. Above are visualization of two such networks: *(a)* has clique motif (5,2) implying that it is a 5-regular tree and *(b)* has motif (2,5) implying it is a highly clustered 8-regular network.

$$Z(\eta) = \sum_{i=0}^{n-2} \binom{n-2}{i} \rho^i (1-\rho)^{n-2-i} \left[1 - \sum_{j=0}^{\eta-i-1} \binom{n-2-i}{j} \left(1 - \sum_{k=0}^{i+j} Q_t(k) \right)^{n-2-i-j} \prod_{r=i}^{i+j-1} \sum_{s=0}^{r} Q_t(s) \right]$$

$$q_t = \rho + (1-\rho) \sum_{m=1}^{\infty} \frac{m(n-1)D(m)}{z} \sum_{\eta=0}^{n-2} Q_t(\eta) Z(\eta)$$

where $q_0 = \rho$ and z is the mean degree of the network. The expected final fraction of adopted nodes R is then given by:

$$R = \rho + (1-\rho) \sum_{m=1}^{\infty} D(m) \sum_{\kappa=1h}^{m(n-1)} \binom{m(n-1)}{\kappa} q_\infty^\kappa (1-q_\infty)^{m(n-1)-\kappa} \sum_{\theta=1}^{\kappa} P(\theta)$$

where q_∞ is the steady state probability that the neighbor of an unadopted node is adopted. One can solve for the steady state probability numerically by calculating the fixed points of the recursive relation for q_t between 0 and 1. A complete derivation of the self consistent equations is given in the appendix. From this framework, we generate several models for complex contagion on d-regular clique-based networks. In the next section, we discuss the results of our experiments.

Experiments and Results

In this section, we analyze the accuracy of several models generated from our framework. We generate three models from our framework – specifically models for d-homogeneous clique-based networks with a bi-modal, uniform, and Poisson threshold distributions – and compare the outputs of these models to numeric simulations.

Model Accuracy

We start by validating the accuracy of models generated from our framework of SI complex contagion on d-homogeneous clique-based networks with several threshold distributions. Below is a table containing our experimental results.

The complex contagion framework above, along with others which approximate the final fraction of adopted nodes, exhibits error when cluster size is comparable to network size. Intuitively this result is a product of the underlying structure of the graph no longer being locally-treelike. As cluster size increases this creates a small world effect in the network which contains short, unpredictable loops. We believe the accuracy of models generated from this framework can be increased by accounting for the average length of a loop in the family of clique-based networks with N nodes.

References

[Centola, 2010] Centola, D. (2010). The spread of behavior in an online social network experiment. *Science (New York, N.Y.)*, 329:1194–7.

[Gleeson and Cahalane, 2007] Gleeson, J. and Cahalane, D. (2007). Seed size strongly affects cascades on random networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 75:056103.

[Min and San Miguel, 2018] Min, B. and San Miguel, M. (2018). Competing contagion processes: Complex contagion triggered by simple contagion. *Scientific Reports*, 8(1).

[O'Sullivan et al., 2015] O'Sullivan, D. J., O'Keeffe, G., Fennell, P., and Gleeson, J. (2015). Mathematical modeling of complex contagion on clustered networks. *Frontiers in Physics*, 3.

Appendix

A Derivation of the Self Consistent Equations

We proceed in a top-down manner, starting with the derivation of R, which is the final fraction of adopted nodes. To start, we partition into two cases. First, we note that ρ fraction of the nodes are seeds. The remaining $1-\rho$ fraction are not initially adopted but could become adopted as the contagion spreads throughout the network. Thus, we can say that

$$R = \rho + (1-\rho)\cdot B$$

, where B represents the fraction of non-seed nodes that become adopted. We can equivalently view B as the probability that a node becomes adopted provided that it is not a seed. To reason about this latter interpretation, we observe that the randomness in our network is entirely captures by the number of cliques a node is in. In other words, we can view B as an expectation that is taken over the number of cliques that a node is contained in. A node is a part of m cliques with probability $D(m)$ and thus has $m(n-1)$ neighbors. Now, let q_∞ denote the steady state probability of an unadopted node having an adopted neighbor. It follows that the probability of our selected node having κ adopted neighbors is

$$\binom{m(n-1)}{\kappa} q_\infty^\kappa (1-q_\infty)^{m(n-1)-\kappa}.$$

Now, given that the node has κ adopted neighbors, it becomes adopted provided that its threshold θ, sampled from $P(\theta)$ is at most κ. That is, the node becomes adopted with probability $\sum_{\theta=1}^{\kappa} P(\theta)$. Putting all of this together, we obtain that

$$R = \rho + (1-\rho)\sum_{m=1}^{\infty} D(m) \sum_{\kappa=1h}^{m(n-1)} \binom{m(n-1)}{\kappa} q_\infty^\kappa (1-q_\infty)^{m(n-1)-\kappa} \sum_{\theta=1}^{\kappa} P(\theta)$$

Tree Representation of the Network

In order to better reason about q_∞, we reorient our clique-based network into an underlying tree-based structure. First, pick a random node from our network to be the root and take it to be at the highest level, T, of our tree. Then, if the root r is in m_r cliques, it is connected to $d_r = m_r(n-1)$ different nodes,

which we will take to be in the subsequent level of our tree. These nodes are then the children of our root node. Each child of the root is in a clique containing the root and $n-2$ other children of the root. We refer to these $n-2$ nodes the child's sibling nodes. Now, each of the children, $c_1, \ldots, c_{d_r}$, of the root are in $m_1, \ldots, m_{d_r}$ cliques respectively, meaning that they are connected to $(m_1)(n-1), \ldots, (m_{d_r})(n-1)$ unique nodes. Now, for each child, observe that we have already included the nodes from one of the cliques in the tree, namely the clique containing its siblings and the root. Thus, each of the children is connected to $(m_1-1)(n-1), \ldots, (m_{d_r}-1)(n-1)$ remaining nodes, which we include in the subsequent level of our tree. Continuing this pattern, we can construct a tree structure which underlies the structure of our clique-based network. If we now add edges to this tree between siblings at each level which are in the same clique, we recover a structure that can be used to analyze our clique-based network (see **Figure 2** for a visualization of (m, n)-homogeneous clique-based networks as trees).

Now, given our tree representation of the network, we again start with some fraction ρ of the nodes being infected but then imagine the contagion propagating up the tree through the edges connecting the seed nodes with its siblings and parents. This will allow us to think of the propagation in a more recursive manner. In other words, we can imagine the contagion spreading by starting at seed nodes and following edges to nodes in the level above, and so on. The key idea to note here is that the degree distribution of the node we reach by following a randomly chosen edge is different from the degree distribution of the overall network, which is $D(m)$. To see this, observe that a node that is in more cliques has a proportionally larger degree, thereby making it proportionally more likely to reach the node after following a randomly chosen edge. It follows from *insert citation here* that the degree distribution of this node is given by

$$\frac{m(n-1)D(m)}{z}$$

Derivation of q_∞

Now, that we have motivated the tree representation of our network, we derive q_∞ in a recursive manner. Formally, we define q_t to be the probability of a node at level t being infected given the state of the nodes at lower levels, and q_∞ as the limit of this quantity as t goes to infinity. Now, a node at level t can be infected in two ways. it is initially infected as part of the set of seed nodes (with probability ρ) or it is not a seed node (with probability $1-\rho$) but becomes infected later via the spread of the contagion. We now want to determine the probability with which the node becomes infected through the spread of the contagion. We begin by imagining the contagion starting at a nodes in level t or $t-1$ and following a randomly chosen edge to a node at level t. We then enumerate over all possibilities of the degree of this node, recognizing from the previous section that the degree distribution is not the same as that of the overall network. Thus, summing over the possible degrees of the node, we get

$$\sum_{m=1}^{\infty} \frac{m(n-1)D(m)}{z}$$

The last thing we need to determine is the likelihood of a particular node being infected. Now, observe that by definition of q_t, we are given the number of infected children of a node at level t (since we know the states of nodes at lower levels) but not the number of infected siblings. This suggests that we think about infected siblings and infected children separately. In other words, if we know that a certain number of a node's children are infected, we also know that if at least a certain number of its siblings are infected, then the node is infected. Concretely, if there are c infected children of a node, we define $\eta = \max(0, \theta - c)$, where *theta* is the threshold of the node and drawn from the distribution $P(\theta)$. Then, if we develop a notion of the likelihood that we need η siblings to be infected after seeing how many children are infected as well as the likelihood that at least η siblings are infected, then we can characterize the probability of the node being infected. Formally, let $Q_t(\eta)$ be a distribution that outputs the probability that η siblings of a node at level t need to be infected in order for it to be infected, and let $Z(\eta)$ denote the probability that at least η siblings of a node become infected. Note that $Q_t(\eta)$ is indexed by the level of the tree whereas $Z(\eta)$ is not. This is due to the fact that $Z(\eta)$ is inherently a distribution that is dependent on only sibling nodes, and each node other than the root has $n-2$ siblings, regardless of what level of the tree it is on. Now, putting this together, the probability that a node becomes infected is $Q_t(\eta)Z(\eta)$. Observing that η lies between 0 and

$n-2$ and summing over these values gives our equation for q_t:

$$q_t = \rho + (1-\rho)\sum_{m=1}^{\infty} \frac{m(n-1)D(m)}{z} \sum_{\eta=0}^{n-2} Q_t(\eta) Z(\eta)$$

Derivation of $Q_t(\eta)$

We first enumerate over all possible degrees of a node at level t. Now, by definition of $Q_t(\eta)$, we are only concerned with the children of a node. Recall that if a node has $n-1$ siblings in the tree, so if it is a part of m cliques, it has $(m-1)(n-1)$ children. Of these children, which are on level $t-1$ of the tree, we want to think about the probability with which κ of them are infected. Observe that we know q_{t-1} which gives us the probability that a node at level $t-1$ is infected given everything up to that level has been updated. Then, enumerating over all κ and using the binomial distribution with success probability q_{t-1}, we get the probability of κ children of our node being infected.

$$\sum_{m=1}^{\infty} \frac{m(n-1)D(m)}{z} \sum_{\kappa=0}^{(m-1)(n-1)} \binom{(m-1)(n-1)}{\kappa} q_{t-1}^{\kappa}(1-q_{t-1})^{(m-1)(n-1)-\kappa}$$

The last step of the derivation is to compute the probability of at least η siblings needing to be infected once κ children are infected. Observe that if $\eta = 0$, the threshold of the node, θ, must be less than or equal to κ. Thus probability of this is

$$\sum_{\theta=1}^{\kappa} P(\theta)$$

Now if $\eta > 0$, then the only way that we need η siblings to be infected once κ children are infected is if the node's threshold is $\kappa + \eta$, which occurs with probability $P(\kappa+\eta)$. Putting this together, we obtain the equation for $Q_t(\eta)$.

$$Q_t(\eta) = \sum_{m=1}^{\infty} \frac{m(n-1)D(m)}{z} \sum_{\kappa=0}^{(m-1)(n-1)} \binom{(m-1)(n-1)}{\kappa} q_{t-1}^{\kappa}(1-q_{t-1})^{(m-1)(n-1)-\kappa} \begin{cases} P(\kappa+\eta) & \text{if } \eta > 0, \\ \sum_{\theta=1}^{\kappa} P(\theta) & \text{if } \eta = 0. \end{cases}$$

Derivation of $Z(\eta)$

We conclude by deriving $Z(\eta)$, which denotes the probability that at least η siblings of a node are infected. Now, we begin by considering the probability with with i of the $n-2$ siblings are seeds. This is given by

$$\binom{n-2}{i} \rho^i (1-\rho)^{n-2-i}$$

Now, if we let K denote the probability that less that $\eta - i$ of the non-seed siblings become infected and enumerate over all i, we obtain that

$$Z(\eta) = \sum_{i=0}^{n-2} \binom{n-2}{i} \rho^i (1-\rho)^{n-2-i}(1-K)$$

Now, to reason about K, we want to compute the probability of exactly j non-seed siblings being infected and enumerate up until $\eta - i - 1$. For this to happen, we need one of the siblings to need at most i siblings to be infected, which happens with probability

$$\sum_{s=0}^{i} Q_t(s).$$

Since we have i siblings who are seeds, this node will be infected. Subsequently, we need a node who needs at most $i+1$ siblings to be infected, which happens with probability

$$\sum_{s=0}^{i+1} Q_t(s).$$

We iteratively apply this logic, ending with requiring a sibling who needs at most $i+j-1$ siblings to be infected. This gives us the probability of at least j additional siblings being infected

$$\prod_{r=i}^{i+j-1} \sum_{s=0}^{r} Q_t(s)$$

. Now, recall that we don't what *at least* j additional siblings to be infected but rather *exactly* that amount. Thus, we need to compute the probability that the remaining $n-2-i-j$ siblings do not become infected, which is equivalent to the probability that each of them need greater than $i+j$ siblings to be infected, which is

$$\left(1 - \sum_{k=0}^{i+j} Q_t(k)\right)^{n-2-i-j}$$

. Putting all of this together, we obtain the final result,

$$Z(\eta) = \sum_{i=0}^{n-2} \binom{n-2}{i} \rho^i (1-\rho)^{n-2-i} \left[1 - \sum_{j=0}^{\eta-i-1} \binom{n-2-i}{j} \left(1 - \sum_{k=0}^{i+j} Q_t(k)\right)^{n-2-i-j} \prod_{r=i}^{i+j-1} \sum_{s=0}^{r} Q_t(s)\right]$$