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

Atul Ganju

Working Paper, January 30, 2022

How behavior diffuses in a population is an important sociological question and has received increased attention recently due to its relevance in economics – most prominently in the burgeoning field of market research. The diffusion of a behavior across a population can be modeled as the spread of a contagion across a social network, where a social network is typically a large network with rich local structure. In the last decade, independent experimentation has demonstrated that social reinforcement amplifies the spread of a behavior over a social network. This suggests that social behavior is a complex contagion process implying its spread would be more extensive and rapid across clustered networks versus random networks [Centola, 2010].

To model the spread of behavior on social networks, 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, the relationship between network structure and various simple and complex contagion mechanisms are explored. 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 modeled most accurately as a simple contagion – where one successful exposure is enough to transmit the contagion – which was determined through observation that social reinforcement is an inhibiting mechanism. Therefore, information propagation and disease diffusion are most powerful on random networks which have a locally tree-like structure. Alternatively, it has been determined that the spread of behavior is best modeled as a complex contagion – where multiple successful exposures are needed to induce adoption – which was

determined through observation that social reinforcement is an amplifying mechanism [Centola, 2010]. Therefore, the spread of behavior is most pronounced on clustered networks.

It has proven to be a challenge to analyze complex contagion processes. The most bare-bones model for simple contagion is the SI (susceptible-infected) model, where a node is either susceptible or infected and once infected remains as such. Since, in complex contagion, nodes have a 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 model 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 the spread of a behavior over a social network 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 such a MFA model to study the effect seed size has on the likelihood of a global cascade [Gleeson and Cahalane, 2007]. Extending this model, 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 a clustered network.

To address the limitations of MFA methods, O'Sullivan et al. developed the CA (clique-approximation) scheme which models complex contagion on clustered networks in continuous time [O'Sullivan et al., 2015]. Their scheme 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 model, 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 d-homogeneous 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 scheme and our framework 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. So, in order to analytically model complex contagion on clustered networks, we construct the networks in such a way that we can control the clustering. The way we do this is by defining a family of networks, clique-based networks, that allow us to select the size and number of the cliques each node is a part of. Specifically, a node is in m cliques of size n, where m is node-dependent and sampled from a distribution $D(m)$. For such networks, if a node in the network is in m cliques of size n, then we refer to the tuple (n, m) as the node's clique motif. A clique-based network can be either non-homogeneous, where the network has several clique motifs and a node's environment is then sampled from a distribution over possible values for m, or d-homogeneous, where the network has only one clique motif (refer to **Figure 1** for example homogeneous network topologies)

Although we develop a framework for complex contagion on arbitrary clique-based networks, we use it to generate models for complex contagion on d-homogenous clique-based networks. It is important to note that we choose to analyze models over regular clique-based networks for the sake of simplicity and that the

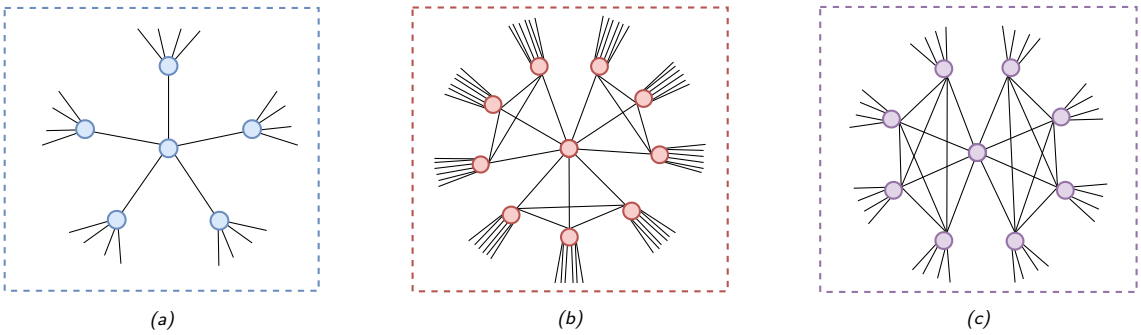

Figure 1: Clique motifs define network topology. In *(a)* we have a 5-regular clique-based network with motif (2,5), in *(b)* we have a 9-regular clique-based network with motif (4,3), and in *(c)* we have an 8-regular clique-based network with motif (5,2).

framework can be used to develop models for complex contagion on non-homogeneous clique-based networks. Finally, it is worth noting the following clique motifs collapse into simpler graphs:

- $n = 2, m = k$ where $k > 2$ is a locally tree-like structure where every node has k children.
- $n = k, m = 1$ where $k \geq 1$ is a completely connected component of size k.

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

Contagion Model

We present a model 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 from a distribution $D(m)$ and n is constant. 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 nothing 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 $n, m, \rho,$ and $P(\theta)$. To ensure the accuracy of our model, and in turn our framework, we compare its output to large-scale numerical simulations over a large parameter space, detailed 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

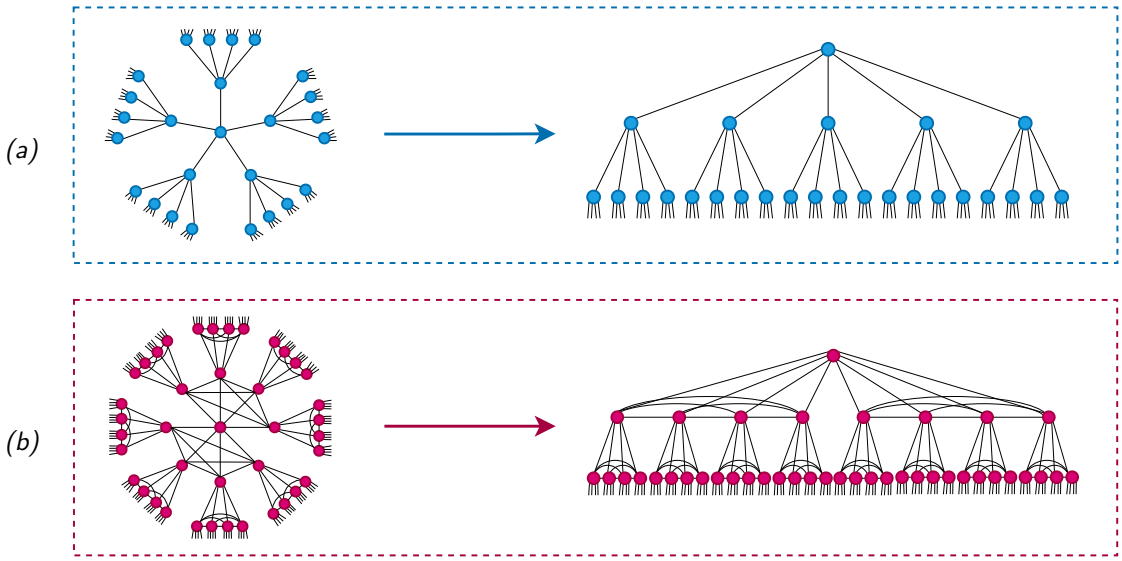

Figure 2: Clique-based networks have an underlying tree-like structure. Above are visualization of two such networks: *(a)* has clique motif (2,5) implying that it is a 5-regular tree and *(b)* has motif (5,2) implying it is a highly clustered 8-regular network.

overcome the limitations of current approximation methods. Specifically, we define a framework for modeling complex contagion on clique-based networks and use it to develop models for complex contagion on clique-based networks. We then analyze these models over several parameter sets providing insight into the effect clusters have on the spread of information over a network.

We start by demonstrating that clique-based networks have 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 we take the root to be in m_r cliques, the root 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. Ignoring this clique, if each of the children, $c_1, \ldots, c_{d_r}$, of the root are in $m_1, \ldots, m_{d_r}$ cliques respectively, then they are connected to $(m_1 - 1), \ldots, (m_{d_r} - 1)$ unique nodes which we can imagine to be in the next lower level of our tree. Continuing this pattern, we can construct a tree 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 have our clique-based network (see **Figure 2** for a visualization of d-homogeneous clique-based networks as trees). In the limit where the number of nodes in our network $N \to \infty$ we have that our tree gets infinitely large and therefore $T \to \infty$. This visualization of a clique-based network then inspires a set of mutually recursive, self-consistent equations to approximate the final fraction of adopted nodes on any locally clique-based network in the limit where $N \to \infty$.

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 susceptible 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 an average 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 derive 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 thanmor 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 it $Q_t(\eta)$, 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}, and in turn a mutually recursive set of equations for q_t:

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

$$Z(n, \eta, i) = \begin{cases} 1 - \sum_{j=0}^{\eta-i-1} \binom{n-2-i}{j} \left(1 - \sum_{k=0}^{i+j} Q_t(k)\right)^{n-2-i-j} \prod_{r=i}^{i+j-1} \sum_{s=0}^{r} Q_t(s) & \text{when } i < \eta \\ 1 & \text{when } i \geq \eta \end{cases}$$

$$q_t = \rho + (1 - \rho) \sum_{m=1}^{\infty} \frac{m(n-1)D(m)}{z} \sum_{\eta=0}^{n-2} Q_t(\eta) \left[\sum_{i=0}^{n-2} \binom{n-2}{i} \rho^i (1 - \rho)^{n-2-i} Z(n, \eta, i) \right]$$

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:

$$R = \rho + (1 - \rho) \sum_{m=1}^{\infty} D(m) \sum_{\kappa=0}^{m(n-1)} \binom{m(n-1)}{\kappa} q_{\infty}^{\kappa} (1 - q_{\infty})^{m(n-1)-\kappa} \sum_{\theta=1}^{\kappa} P(\theta)$$

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.

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.