Impact of Community Structure on Cascades

Mehrdad Moharrami, Vijay Subramanian,

Mingyan Liu, Marc Lelarge

June 2, 2016

Abstract

The threshold model is widely used to study the propagation of opinions and technologies in social networks. In this model individuals adopt the new behavior based on how many neighbors have already chosen it. We study cascades under the threshold model on sparse random graphs with community structure to see whether the existence of communities affects the number of individuals who finally adopt the new behavior. Specifically, we consider the permanent adoption model where nodes that have adopted the new behavior cannot change their state. When seeding a small number of agents with the new behavior, the community structure has little effect on the final proportion of people that adopt it, i.e., the contagion threshold is the same as if there were just one community. On the other hand, seeding a fraction of population with the new behavior has a significant impact on the cascade with the optimal seeding strategy depending on how strongly the communities are connected. In particular, when the communities are strongly connected, seeding in one community outperforms the symmetric seeding strategy that seeds equally in all communities.

Keywords: Random Graphs, Galton Watson Multitype Branching Process, Contagion Threshold, Threshold Model, Differential Equation Approximation

---

Authors addresses: M. Moharrami, V. Subramanian, M. Liu, Electrical & Computer Engineering Department, University of Michigan; email: {moharami, vgsubram, mingyan}@umich.edu; M. Lelarge, INRIA-Ecole Normale Superieure, email: marc.lelarge@ens.fr.

1 Introduction

In this paper we investigate a type of cascade problem on graphs that has been used to study the spread of new technology or opinions in social networks, see e.g., [5, 10, 22, 24, 25, 27]. This spread is also referred to as contagions in networks. The underlying model typically consists of a few (selected) initial adopters (nodes in the network) or "seeds" and a particular adoption model that determines the condition under which a node will choose to adopt given the states of its neighbors. A commonly studied model is the threshold model [19, 26], whereby individuals adopt the new technology based on how many neighbors have already chosen it.

Prior work in this area has generally focused on analyzing what happens when the underlying network is given by a single community modeled as a sparse random graph, either heuristically, see e.g., [16, 26], or more rigorously, see e.g., [3, 15]. In this paper we will instead consider graphs with a type of community structure (also known as modular networks), whereby multiple sparse random graphs are weakly interconnected. This could model for instance segments of the population (e.g., different age or ethnic groups), where members of a single segment are more strongly connected (with a relatively high node degree) and cross-segment connections are weak, i.e., fewer members are connected to those from a different segment. This would be a more realistic and interesting model for many practical scenarios and serve as a natural next step to the studies done with a single community. We are particularly interested in whether the existence of communities affects the number of individuals who eventually adopt the new technology. Also of interest is the question whether seeding in all communities is a better strategy in terms of maximizing the number of eventual adopters than exclusively in one community. While earlier works have looked at this problem using heuristic methods, see e.g., [6–9, 16], we set out to present a mathematically rigorous analysis of this problem.

Specifically, we consider the permanent adoption model where nodes that have adopted the new technology cannot change their state. Our analysis presents a differential-equation-based tight approximation to the stochastic process of adoption under the threshold contagion model. While this is a similar approach to the original analysis of contagions in a single community [3], the additional community structure requires significant technical development to establish the validity of this approach in the new scenario. The analysis of the differential equation leads to a correctness proof of a mean-field equation for the contagion in a large network, as well as an algorithm to calculate the properties of the contagion. Using this analysis, we are able to analyze the impact of advertising by means of seeding of the nodes with the new technology or opinion. The differential equation also leads to characterization of the sample-path of the adoption process as well as a sharp characterization of the contagion threshold for the linear threshold model.

Our main contributions can be summarized as follows.

1. We prove the validity of a mean-field analysis of the contagion process over

infinite trees. This analysis yields a fixed point equation whose solution can be used to exactly determine the final fraction of the population that are eventual adopters (the size of the cascade). Furthermore, when the fixed point equation has multiple solutions, we identify the correct solution among these and provide an algorithmic means to calculate it.

2. We provide a tight differential equation approximation to the sample-path of the contagion process. This allows us to track the evolution of the cascade, in particular, how it moves from one community to the other. We also show that the solution of the differential equation can obtained by a much simpler four-dimensional differential equation. This dimension reduction is crucial to developing a comprehensive understanding of the contagion process.

3. For general thresholds we provide a sharp characterization of the contagion threshold in terms of the Perron-Frobenius eigenvalue of an associated matrix.

4. Specializing to Poisson degrees distributions and linear thresholds we prove that the community structure does not matter for global properties like the contagion threshold for the linear threshold model of [19, 26]. In particular, when seeding a small number of agents with the new technology, we find that the community structure has little effect on the final proportion of adopters.

5. We numerically study the impact of the community structure on the viral seeding of nodes. We find that seeding a fraction of population with the new technology has a significant impact on the cascade, with the optimal seeding strategy depending on how strongly the communities are connected.

The remainder of this paper is organized as follows. We present our model in Section 2 and a literature review in Section 3. In Section 4 we present a mean-field approximation of the adoption process, whose validity is then established in Section 5 by constructing a Markov process coupling the evolution of the adoption process with the process generating the random graph. This Markov process is then approximated using a set of ODEs in Section 6, and their analysis is presented in Section 7. We discuss the results on the contagion threshold for general thresholds in Section 8. All the results are specialized to the case of Poisson degree distributions in Section 9. We present numerical results in Section 10 and conclude in Section 11.

This is the condition on the thresholds for which a contagion occurs with a finite set of seed nodes.

3

2 System Model

We consider a set $[n] = \{1, \ldots, n\}$ of agents that are organized into two communities, community 1 $\{1, 2, \ldots, n_1\}$ and community 2 $\{n_1 + 1, \ldots, n\}$ with $n_2 := n - n_1$ individuals. Assume that we are given three sequences of non-negative integers for each of these agents, i.e., $\mathbf{d}_1 = (d_{1,i}^n)_1^{n_1}$, $\mathbf{d}_2 = (d_{2,i}^n)_{n_1+1}^{n}$ and $\mathbf{d}_m = (d_{m,i}^n)_1^n$. We will assume the following conditions on these sequences: 1) $\sum_{i=1}^{n_1} d_{1,i}^n$ is even; 2) $\sum_{i=n_1+1}^{n} d_{2,i}^n$ is even; and 3) $\sum_{i=1}^{n_1} d_{m,i}^n = \sum_{i=n_1+1}^{n} d_{m,i}^n$. The sequence $\mathbf{d}_j$ is the degree sequence of the sub-graph for community j for $j \in \{1, 2\}$ and $\mathbf{d}_m$ the degree sequence of the bi-partite graph connecting the two communities. For ease of explanation we assume the same degree distribution for the cross-community links but we believe that our results hold even if this weren't the case.

We define a two-community random multigraph (allowing for self-loops and multiple links) with given degree sequences $\mathbf{d}_1$, $\mathbf{d}_2$ and $\mathbf{d}_m$ generated by the configuration model [4] as the concatenation of $G^*(n_1, \mathbf{d}_1)$, $G^*(n_2, \mathbf{d}_2)$ (both generated via the configuration model) and a random bi-partite multigraph $G^*(n_1, n_2, \mathbf{d}_m)$: generate half-edges for each node corresponding to the different degree sequences and combine the half-edges into edges by a uniformly random matching of the set of half-edges of each sequence. Conditioned on the random multigraphs being a simple graph, we obtain uniformly distributed random graphs $G(n_1, \mathbf{d}_1)$ and $G(n_2, \mathbf{d}_2)$ with the given degree sequences. If we further condition on the bi-partite random multigraph being simple, we will obtain a simple bi-partite graph $G(n_1, n_2, \mathbf{d}_m)$. The concatenation of these will produce a simple two-community graph $G(n, \mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_m)$ with the desired distributions. In Lemma 5.1 we will assume standard regularity assumptions [18] on our degree sequences, and also include a tail condition needed for our analysis. We will assume that n_1, n_2 are $O(n)$. The stochastic block model [20] is a prototypical example of a two-community graph.

Following Lelarge [15] we will analyze the threshold model of Morris [19] and Watts [26] on the two-community random graph model described above. In this model, we assume that nodes have the choice between two types of opinions/technologies, A and B; we will sometimes also use "inactive" to denote type A and "active" to denote type B. We will assume that all nodes initially start in type A, i.e., are inactive. We will assume that each node has a threshold that is a function of its community and degrees (in the same community and across to the other community); the value of the threshold is fixed and allowed to be any non-negative real number up to the sum of the degrees. If a node finds that the number of its neighbors (across both communities) who have chosen type B is at least its threshold, then it will permanently choose to switch to type B. Again following [15] we will initially seed nodes with type B using a Bernoulli random variable (1 implying that a node gets seeded with type B) that is independently chosen with the mean depending on the node's parameters, namely, community and degree. Note that a degree and/or community-unaware seeding strategy would imply an appropriate uniformity in the means of the seeding random variables. After the seeding process is completed, the remaining

nodes then react to the seed nodes and decide whether to adopt type B. This process continues until a final state of the nodes is reached. A cascade is said to happen if the number of nodes adopting type B is substantially greater than the seed set.

Notation: We will denote random variables by capital letters (sometimes using a bold typeset too); realizations or deterministic quantities are in small letters. We will often denote vectors using a bold typeset and individual components without it. Adhering to game theoretic notation we will often denote a node's community by $j \in \{1, 2\}$ and the other community by $-j = \{1, 2\} \setminus \{j\}$. We will also use *community* and *side* interchangeably in the document.

3 Literature Review

The threshold model [10, 22, 24, 27] is a well accepted model for explaining the adoption of a new technology, opinion or behavior in a population that interacts via a social network. The linear threshold model, where the threshold is function of the degree, was analyzed for the contagion threshold for specific graphs in [19], and using heuristically derived formulae for single community random graphs in [16, 26]. The results on the single community random graphs were rigorously proved using branching processes in [15], where the importance of pivotal players (those whose degree is low enough that one neighbor will make them adopt the new behavior) was identified and studied. Similar results were derived using the differential equation method in [3] with more general threshold functions.

The threshold model has been studied for networks with communities, but using heuristically derived mean-field approximations and approximate differential equations [6–9, 16]. In these studies, it was numerically shown in [6, 8] for the linear threshold model that the community structure leads to a different dynamic in terms of the evolution of the cascade itself. It is important to note that the authors in these works postulate both the mean-field equation and the differential equations in an *ad hoc* manner without a formal proof. This is particularly the case for the multi-community work in [6, 8] where the authors combine the adoption processes in the different communities without a proper mathematical justification.

The problem of maximizing influence propagation in networks, by targeting certain influential nodes that have the potential to influence many others, has been an important follow-up problem [10, 22, 24, 27] once the impact of a social network on behavior adoption was discovered. While this problem is known to be NP hard for many influence models, several approximate methods have been designed, see e.g., [14, 21]. A contrasting strategy to identifying and targeting influential nodes is to use viral marketing [22, 24, 25]. A randomized version of viral marketing, also referred to as seeding or advertising in the paper, was studied in [3, 15] where the resulting cascade was precisely identified. The results in [15] also indicated that targeting higher degree nodes is a better seeding strategy over random seeding. With community structure, [7–9] showed that

5

the seeding strategies could be dramatically different from the one-community optimal strategies. Typically asymmetric seeding strategies, wherein the seeding is principally carried out in one community over another, were shown to perform better than more uniform (over the communities) seeding strategies.

4   Mean-Field Approximation

We start by presenting a mean-field approximation of the process of adoption of type B, i.e., becoming active, in a typical simple graph generated through the configuration model described in Section 2. The graphs that we consider are locally tree-like so that the structure up to any finite depth when viewed from any node of the graph is a tree with high probability. Therefore, simple graphs produced by configuration model converge [1, 2] to a rooted Galton Watson Multi-type Infinite Tree (GWMIT). An example of the limiting rooted GWMIT is shown in Figure 1 where the root node is in community 1. Note that the degree distribution of the root node is $\mathbb{P}_j$ and $\mathbb{P}_m$ if the community of the root is $j \in \{1, 2\}$. The degree distributions of each child is then given by size-biased/sampling-biased distribution for the community of the parent and regular distribution for the other community: if the parent is in community $j \in \{1, 2\}$ and the child node is in community j too, then the degree distribution in community j is the size-biased distribution $\mathbb{P}_j^*$ corresponding to $\mathbb{P}_j$ and the degree distribution in community $-j$ is $\mathbb{P}_m$; on the other hand, if the parent is in community $j \in \{1, 2\}$ and the child node is in community $-j$, then the degree distribution in community $-j$ is $\mathbb{P}_{-j}$ and distribution in community j is the size-biased distribution $\mathbb{P}_m^*$ corresponding to $\mathbb{P}_m$. For a random $D \in \mathbb{Z}_+$ with distribution $\mathbb{P}(\cdot)$ and finite mean $\mathbb{E}[D]$, the size-bias/sampling-biased distribution $\mathbb{P}^*(\cdot)$ is given by $\mathbb{P}^*(d) = d\mathbb{P}(d)/\mathbb{E}[D]$ for all $d \in \mathbb{Z}_+$. We denote a random variable with the size-bias distribution by D^*+1 where D^* takes values in $\mathbb{Z}_+$. For a Poisson random variable with parameter $\lambda > 0$, i.e., $D \sim \text{Poi}(\lambda)$, we have $D^* \sim \text{Poi}(\lambda)$, so that the size-bias/sampling-biased distribution is a shifted Poisson distribution. This is the only distribution with this property.

Assume that we have a rooted GWMIT with root node ψ denoted by T_ψ. For a node $i \neq \psi$ let l be its parent, indicated by $i \longrightarrow l$, and $T_{i \longrightarrow l}$ be the sub-tree rooted at i when the link (l, i) is excised. Then assuming that l is inactive, state of node i only depends on the state of her children in sub-tree $T_{i \longrightarrow l}$. Next we define a few random variables that will aid in describing the mean-field approximation.

$X_\psi^{(j)}$: Bernoulli r.v; $= 1$ if root node ψ of the rooted GWMIT is on side j and inactive.

$Y_l^{(j,j)}$: Bernoulli r.v; $= 1$ if node $l \neq \psi$ on side j is inactive conditioned on having an inactive parent on side j.

$Y_l^{(j,-j)}$: Bernoulli r.v; $= 1$ if node $l \neq \psi$ on side $-j$ is inactive conditioned on having an inactive parent on side j.

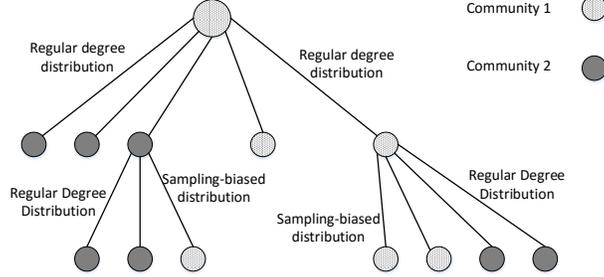$\boldsymbol{\alpha}_l^{(j)}$: Bernoulli r.v; $= 1$ if node l on side j is a seed node.

Figure 1: Illustration of the limiting rooted Galton-Watson multi-type infinite tree.

$K_l^{(j)}$: Threshold of node l on side j that is determined by number of its neighbors in either community, i.e., by the degrees of the node on side j and $-j$.

Then assuming independence we can write down the following equations:

(i) A non-root node $l \neq \psi$ will remain inactive, if it is not seeded initially and the number of her children who are active does not exceed the threshold, i.e.,

$$Y_l^{(j,j)} = \left(1 - \boldsymbol{\alpha}_l^{(j)}\right) \mathbf{1}\Big\{ \sum_{i \longrightarrow l} \left(1 - Y_i^{(j,j)}\right) + \sum_{i \longrightarrow l} \left(1 - Y_i^{(j,-j)}\right) < K_l^{(j)} \Big\}, \quad (1)$$

$$Y_l^{(j,-j)} = \left(1 - \boldsymbol{\alpha}_l^{(-j)}\right) \mathbf{1}\Big\{ \sum_{i \longrightarrow l} \left(1 - Y_i^{(-j,-j)}\right) + \sum_{i \longrightarrow l} \left(1 - Y_i^{(-j,j)}\right) < K_l^{(-j)} \Big\},$$
$$(2)$$

where $\mathbf{1}\{O\}$ is the indicator function of set O. Note that in the equations we are using the fact that the parent node is assumed to be inactive.

(ii) Root node ψ (of community j) will remain inactive if it is not seeded initially and the number of her active children falls below her threshold. This implies that

$$X_\psi^{(j)} = \left(1 - \boldsymbol{\alpha}_\psi^{(j)}\right) \mathbf{1}\Big\{ \sum_{i \longrightarrow \psi} \left(1 - Y_i^{(j,j)}\right) + \sum_{i \longrightarrow l} \left(1 - Y_i^{(j,-j)}\right) < K_\psi^{(j)} \Big\}. \quad (3)$$

For the mean-field approximation it is assumed that the random variables $Y_l^{(1,1)}$, $Y_l^{(1,2)}$, $Y_l^{(2,1)}$, and $Y_l^{(2,2)}$ for $l \neq \psi$ are, respectively, identically distributed too. These random variables are then related via the following Recursive Distributional Equations (RDEs) [1, 2], where equality below should be interpreted in terms of distributions.

$$\tilde{Y}^{(j,j)} \stackrel{d}{=} \left(1 - \bar{\boldsymbol{\alpha}}^{(j)}(D_j^* + 1, D_m)\right) \mathbf{1}\Big\{ \sum_{i=1}^{D_j^*} \left(1 - \tilde{Y}_i^{(j,j)}\right)$$
$$+ \sum_{i=1}^{D_m} \left(1 - \tilde{Y}_i^{(j,-j)}\right) < K^{(j)}(D_j^* + 1, D_m) \Big\}, \quad (4)$$

7

$$\tilde{Y}^{(j,-j)} \stackrel{d}{=} \left(1 - \bar{\boldsymbol{\alpha}}^{(-j)}(D_{-j}, D_m^* + 1)\right) \mathbf{1}\Big\{ \sum_{i=1}^{D_m^*} \left(1 - \tilde{Y}_i^{(-j,j)}\right)$$
$$+ \sum_{i=1}^{D_{-j}} \left(1 - \tilde{Y}_i^{(-j,-j)}\right) < K^{(-j)}(D_{-j}, D_m^* + 1)\Big\}, \tag{5}$$

where for every $j \in \{1,2\}$ $\tilde{Y}^{(j,j)}$ and $\tilde{Y}_i^{(j,j)}$ as well as $\tilde{Y}^{(j,-j)}$ and $\tilde{Y}_i^{(j,-j)}$ are $i.i.d.$ copies (with unknown distribution). We also have a set of independent random variables: D_j is a random variable with the community j degree distribution, $D_j^* + 1$ is a random variable with the size-bias distribution of D_j, D_m has inter-community degree distribution, and D_m^* is a random variable with the size-bias distribution of D_m. We have also assumed, without loss of generality, that the seeding Bernoulli random variables have means that depend on the community and the degrees of the nodes, namely, $\alpha_j(d_j, d_{-j})$ for $j \in \{1,2\}$ and $d_j, d_{-j} \in \mathbb{Z}_+$. We also assume that threshold random variables are deterministic functions of the community and degrees of the nodes, namely, $K_j(d_j, d_{-j})$ for $j \in \{1,2\}$ and $d_j, d_{-j} \in \mathbb{Z}_+$. These are then used to construct the random variables $\bar{\boldsymbol{\alpha}}^{(j)}(D_j^* + 1, D_m)$, $\bar{\boldsymbol{\alpha}}^{(-j)}(D_j, D_m^* + 1)$, $K^{(j)}(D_j^* + 1, D_m)$ and $K^{(-j)}(D_j, D_m^* + 1)$.

Since we have RDEs with Bernoulli random variables, we can equivalently obtain the solutions by taking expectations and solving for the means of the underlying random variables. We set $\mathbb{E}\left[X_\psi^{(j)}\right] = \phi_j$, $\mathbb{E}\left[\tilde{Y}^{(j,j)}\right] = \mu^{(j,j)}$ and $\mathbb{E}\left[\tilde{Y}^{(j,-j)}\right] = \mu^{(j,-j)}$, taking expectation in (4)-(5) and then (3) yields

$$\mu^{(j,j)} = \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} \mathbb{P}_j^*(d_j)\mathbb{P}_m(d_{-j})(1 - \alpha_j(d_j, d_{-j}))$$
$$\times Bi(u_j; d_j - 1, 1 - \mu^{(j,j)})Bi(u_{-j}; d_{-j}, 1 - \mu^{(j,-j)}), \tag{6}$$

$$\mu^{(j,-j)} = \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} \mathbb{P}_{-j}(d_{-j})\mathbb{P}_m^*(d_j)(1 - \alpha_{-j}(d_{-j}, d_j))$$
$$\times Bi(u_j; d_j - 1, 1 - \mu^{(-j,j)})Bi(u_{-j}; d_{-j}, 1 - \mu^{(-j,-j)}), \tag{7}$$

$$\phi_j = \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} \mathbb{P}_j(d_j)\mathbb{P}_m(d_{-j})(1 - \alpha_j(d_j, d_{-j}))$$
$$\times Bi(u_j; d_j - 1, 1 - \mu^{(j,j)})Bi(u_{-j}; d_{-j}, 1 - \mu^{(j,-j)}), \tag{8}$$

where $Bi(k; n, p) := \binom{n}{k}p^k(1-p)^{n-k}$ is the PMF of the binomial distribution.

To find the probability of a node in community $j \in \{1,2\}$ remaining inactive, i.e. ϕ_j, one needs to first solve the fixed point equations (6)-(7), and then substitute the result into (8). For ease of understanding we will write the

equations of the mean-field analysis as follows:

$$\boldsymbol{\mu} = F(\boldsymbol{\mu}), \text{ and } \boldsymbol{\phi} = \Phi(\boldsymbol{\mu}), \tag{9}$$

for functions $F(\cdot)$ and $\Phi(\cdot)$ defined component-wise via (6)-(7), and (8), respectively.

A basic question at this point is whether one can rigorously justify (9), particularly given the various independence and uniformity assumptions for the derivation. A few other questions also arise: i) Does a solution to (9) exist; ii) Are there multiple solutions to (9)? Numerically, we observed that there are many cases where (9) has multiple solutions; and iii) Which solution should one pick if there are multiple solutions? Note that for every $\boldsymbol{\mu} \in [0,1]^4$ and $j \in \{1, 2\}$, we have

$$\begin{aligned}
\phi_j = \sum_{d_j, d_{-j}} &\mathbb{P}_j(d_j)\mathbb{P}_m(d_{-j})(1 - \alpha_j(d_j, d_{-j})) \times \\
&\sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} Bi(u_j; d_j - 1, 1 - \mu^{(j,j)})Bi(u_{-j}; d_{-j}, 1 - \mu^{(j,-j)}) \\
\leq \sum_{d_j, d_{-j}} &\mathbb{P}_j(d_j)\mathbb{P}_m(d_{-j})(1 - \alpha_j(d_j, d_{-j})) = \mathbb{P}(\boldsymbol{\alpha}_l^{(j)} = 0)
\end{aligned} \tag{10}$$

so that the seeding distribution gets automatically accounted in any solution of (9), and the final population of active nodes includes at least the seed nodes.

Before proceeding, we should again point out that equations of a similar form were heuristically postulated in the literature [6–9, 16]. An important contribution of our paper is thus to rigorously prove the validity of (9), and to identify the correct solution to choose. As discussed in [15] the existence of multiple solutions and a lack of "monotonicity" makes it extremely challenging to use the techniques developed in [1, 2] to prove the needed results.

5   Markov process of adoption

We next prove the validity of the mean-field equations by constructing a Markov process that couples the evolution of the adoption process with the process of generating the random graph using the configuration model. We will then use techniques developed for the mean-field analysis [3, 17, 28, 29] of the resulting population density-dependent Markov processes to approximate the process by a system of ordinary differential equations (ODEs). We will then show that the resulting ODEs allow us to both track the evolution and identify the final adoption state (using an equilibrium analysis).

The typical way to create a random graph with given degree sequence $(d_i^{(n)})_{i=1}^n$ using the configuration model is to first label nodes of the graph $1, 2, \ldots, n$ such that node i has $d_i^{(n)}$ half-edges sticking out of it. Then we iterate through all the unpaired half-edge so that at each step, we pair half-edges randomly, and declare the final graph as the random graph we desired. In our setting, when

we have two communities, the basic idea of generating the random graph using the configuration model is just the same. Let's assume for nodes in community j, the in-degree (within the same community) sequence is given by $(d_{i,j,j}^{(n)})_{i=1}^n$ and the out-degree (to the other community) sequence is given by $(d_{i,j,-j}^{(n)})_{i=1}^n$. For an agent/node i in community j there are $d_{i,j,j}^{(n)}$ half-edges corresponding to its neighbors in the same community, and $d_{i,j,-j}^{(n)}$ half-edges corresponding to its neighbors in the other community. For obtaining the graph we would then randomly pair half-edges of the same type as described earlier. However, to analyze the adoption process we will work a little differently. We start by realizing the early adopter nodes using the seeding random variables. We set the early adopters to be active and make all their half-edges active. Any other node and its half-edges will initially be counted as inactive. We then run the adoption process and draw the random graph simultaneously by iterating through the active half-edges (if any). At each iteration, we pick an active half-edge, i.e., an half-edge connected to an active node, and connect it to some other half-edge that belongs to the appropriate community. Then we remove both half-edges from the graph. Moreover, if the second half-edge belongs to an inactive node, we reduce its threshold by one. If the threshold of the inactive node becomes zero after this change, we activate this node and also all the half-edges that are still connected to this node. Note that this process stops when all active half-edges have been omitted, and the remainder of the graph (containing only inactive half-edges) is not realized. This process is described in Algorithm 1.

ALGORITHM 1: Process of jointly generating the random graph and running the adoption process.

> **Data**: Realization of degree distribution and early adopters
> **Result**: Sub-graph of the final random graph that contains all Active nodes
> initialization;
> **while** *There is an active half-edge* **do**
> | Randomly choose an active half-edge;
> | Randomly choose another half-edge belongs to proper community ;
> | Omit two selected half-edges from the set of half-edges;
> | Update the state of inactive nodes;
> **end**

We want to keep track of active half-edges, inactive nodes and number of times that the process described in Algorithm 1 picks half-edges from each community. The random variables associated with each of these quantities are given as follows:

$A_j(k)$: Number of active half-edges belonging entirely to community j at time k.

We abuse notation by reusing the same variable but no confusion should arise from this.

10

$A_m^{(j)}(k)$: Number of active half-edges between the two communities at time k that belongs to a node in community j.

$T_j(k)$: Number of times algorithm visits community j up to time k where a visit denotes removing two half-edges within the same community.

$I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k)$: Number of inactive nodes in community j with d_j initially assigned half-edges corresponding to community j where u_j of them have been removed by k, and similarly, d_{-j} initially assigned half-edges corresponding to community $-j$ where u_{-j} of them have been removed by k.

It is easily verified that $\{X^n(k)\}_{k\in\mathbb{Z}_+}$ is a discrete-time Markov chain, where

$$X^n(k) := (A_j(k), A_m^{(j)}(k), T_j(k), I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k), \dots),$$

and $j \in \{1,2\}$, $(d_j, u_j) \in \mathbb{Z}_+^2$ with $u_j \le d_j$, and $(d_{-j}, u_{-j}) \in \mathbb{Z}_+^2$ with $u_{-j} \le d_{-j}$. For ease of explanation we denote the number of edges entirely in community j with n nodes by $m_j(n)$ and the number of edges between the two communities by $m_m(n)$; these can be determined once the degrees have been realized.

The mean-field analysis [3, 17, 28, 29] proceeds by scaling both space and time by n and considering the one-step drift of the scaled process. We will now present the one-step drift analysis of our Markov chain (for the unscaled variables). At each iteration, one of the following events will happen:

1.) Two active half-edges will be omitted. This event results in the half-edges being "wasted", in a manner of speaking. Here two sub-cases are possible:

i.) Both half-edges belongs to community j: This event happens with probability

$$\frac{A_j(k)\,(A_j(k) - 1)}{\left(A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)\right)(2m_j(n) - 2T_j(k) - 1)}.$$

In this case, we should update the corresponding variables as follows:

$$A_j(k + 1) = A_j(k) - 2, T_j(k + 1) = T_j(k) + 1.$$

ii.) Half-edges belongs to different sides: This event happens with probability

$$\frac{2A_m^{(1)}(k)A_m^{(2)}(k)}{\left(A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)\right)(m_m(n) - (k - T_1(k) - T_2(k)))}.$$

In this case, we should update the variables as follows:

$$A_m^{(2)}(k + 1) = A_m^{(2)}(k) - 1, A_m^{(1)}(k + 1) = A_m^{(1)}(k) - 1.$$

2.) One active half-edge and one inactive half-edge will be omitted, while the inactive half-edge belongs to inactive nodes in community j. Four sub-cases arise here:

i.) The inactive node belongs to $I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k)$ and the active half-edges belongs to community j, while $K_j(d_j, d_{-j}) > u_j + u_{-j} + 1$. This event results in

the threshold of an inactive node in community j being lowered by 1 owing a node within its own community. This occurs with probability

$$\frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{(d_j - u_j)\, I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k)}{2m_j(n) - 2T_j(k) - 1}$$

In this case, we should update variables as follows:

$$A_j(k+1) = A_j(k) - 1,\ I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k+1) = I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k) - 1,$$

$$T_j(k+1) = T_j(k) + 1,\ I^{(j)}_{d_j, d_{-j}, u_j+1, u_{-j}}(k+1) = I^{(j)}_{d_j, d_{-j}, u_j+1, u_{-j}}(k) + 1.$$

ii.) The inactive node belongs to $I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k)$ and the active half-edge belongs to community j, while $K_j(d_j, d_{-j}) = u_j + u_{-j} + 1$. Note that an inactive node becomes active during this event and all remaining half-edges also become active. This is an important growth event for our process. This occurs with probability

$$\frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{(d_j - u_j)\, I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k)}{2m_j(n) - 2T_j(k) - 1}$$

Here we should update the variables as follows:

$$A_j(k+1) = A_j(k) - 1 + d_j - u_j - 1,\ T_j(k+1) = T_j(k) + 1,$$

$$A_m^{(j)}(k+1) = A_m^{(j)}(k) + d_{-j} - u_{-j},\ I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k+1) = I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k) - 1.$$

iii.) The inactive node belongs to $I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k)$ and the active half-edge comes from the other community, while $K_j(d_j, d_{-j}) > u_j + u_{-j} + 1$. Here the threshold of an inactive node is being reduced by a node from the other community. This occurs with probability

$$\frac{A_m^{(-j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{(d_{-j} - u_{-j})\, I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k)}{m_m(n) - (k - T_1(k) - T_2(k))}$$

Here we should update the variables as follows:

$$A_m^{(-j)}(k+1) = A_m^{(-j)}(k) - 1,\ I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k+1) = I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k) - 1,$$

$$I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}+1}(k+1) = I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}+1}(k) + 1.$$

iv.) The inactive node belongs to $I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k)$ and the active half-edge comes from the other community, while $K_j(d_j, d_{-j}) = u_j + u_{-j} + 1$. This is another important growth event for our process wherein an inactive node becomes active owing to a node from another community. This occurs with probability

$$\frac{A_m^{(-j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{(d_{-j} - u_{-j})\, I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k)}{m_m(n) - (k - T_1(k) - T_2(k))}$$

Here we should update the variables as follows:

$$A_m^{(-j)}(k+1) = A_m^{(-j)}(k+1) - 1, A_j(k+1) = A_j(k) + d_j - u_j,$$

$$A_m^{(j)}(k+1) = A_m^{(j)}(k+1) + d_j - u_j - 1, \ I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k+1) = I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k) - 1.$$

Finally, note that these random variables satisfy the balance equations given by the realization of degrees. For $j \in \{1,2\}$ we have

$$A_j(k) + \sum_{u_j+u_{-j}<K_j(d_j,d_{-j})} (d_j - u_j)I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k) = 2m_j(n) - 2T_j(k),$$

$$A_m^{(j)}(k) + \sum_{u_j+u_{-j}<K_j(d_j,d_{-j})} (d_{-j} - u_{-j})I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k) = m_m(n) - (k - T_j(k) - T_{-j}(k)),$$

$$(11)$$

where the summations above are understood to be over both the degrees (d_j, d_{-j}) and the used half-edges (u_j, u_{-j}) meeting the constraint listed underneath.

The one step drift for the unscaled random variables is obtained by summing over all possible events, given the history of the Markov process upto time k. The details can be found in Appendix B.

We will conclude this section by stating the regularity conditions of our degree distributions and some consequences of these conditions.

Lemma 5.1. *Let assume the degree distributions over each community and between two communities satisfies the regularity conditions, i.e. for $j, j' \in \{1,2\}$ the corresponding degree sequence is $\mathbf{d} = (d_{i,j,j'}^{(n)})_{i=1}^n$ (degree sequence of nodes in community j connected to the nodes in community j') such that $\sum_i d_{i,j,j'}^{(n)}$ is even and satisfies the following conditions: (i) $|\{i : d_{i,j,j'}^{(n)} = r\}|/n \to p_{j,j',r}$ for every $r \geq 0$ as $n \to \infty$,*
(ii) $\lambda_{j,j'} := \sum_{r \geq 0} r p_{j,j',r} \in (0, \infty)$,
(iii) $\sum_i (d_{i,j,j'}^{(n)})^2 = O(n)$,
where $\mathbf{P}(j,j') = \{p_{j,j',r}\}_0^\infty$ is some probability distribution. Moreover, we also assume that this degree distribution satisfies the following tail condition,
(iv) $|\{i : d_{i,j,j'}^{(n)} > n^{1/3-2\Delta}\}|/n = O(n^{-(1+\Delta)})$ for some $\Delta > 0$, hence, the third moment of degree distribution should be finite. We can rewrite above assumption as follows:
(iv*) $\sum_i \left(d_{i,j,j'}^{(n)} \right)^3 = O(n)$
Then there is a big enough $K \in \mathbb{R}$ such that the corresponding random variables, i.e. $A_j(k)$, $A_m^{(j)}(k)$, $T_j(k)$ and $I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k)$, are bounded by Kn as $n \to \infty$ almost surely.

Proof. See Appendix F.1. □

Some remarks are in order: (1) In order to have a simple random graph, other conditions should also be satisfied. As an example, we should have the

following condition: $\sum_i d_{i,2,1}^{(n)} = \sum_i d_{i,1,2}^{(n)}$, which states that the first moment of $\mathbf{P}(1,2)$ and $\mathbf{P}(2,1)$ should be the same. For the sake of simplicity, we assume that $\mathbf{P}(1,2)$ and $\mathbf{P}(2,1)$ have same distribution which is represented as $P_m(\cdot)$ throughout the paper. (2) Note that assumption (iv) is crucial to bound the tail of the distribution. In most interesting cases such as the Poisson distribution, the tail condition is automatically satisfied given the exponential tail. (3) Under the regularity conditions (i)-(iii), Janson [11] proved that the probability that the random graph generated via the configuration model is simple is strictly positive. Therefore, any results that hold with probability approaching 1 will also hold for simple graphs of this family conditioning on the event that the graph is simple. Under this conditioning and by conditioning on the degree sequences we can study the process on other random graph models [12, 13], e.g., those generated by the Erdos-Renyi model with Poisson degree distributions in the limit of n going to infinity. We will not elaborate further on this.

6   Convergence to ODE

In this section we present the ODE approximation to the adoption Markov process. We will show that a scaled-version of the Markov process of adoption from Section 5 converges (in probability) to a set of continuous functions obtained from the solution of a set of ODEs. We start by highlighting why the analysis is non-trivial. The first point concerns some of the terms that appear in the one-step drift. Notice that we have many terms like $\frac{A_j(k)}{A_1(k)+A_2(k)+A_m^{(1)}(k)+A_m^{(2)}(k)}$. In terms of the scaled variables, these terms are not Lipschitz unless there is a lower bound on the value of the (scaled) denominator. Owing to this, in our ODE approximation we will have to stop the Markov process of adoption just before the sum of these scaled variables hits zero (corresponding to the denominator above), i.e., before all the active half-edges have been omitted; it is important that this be the sum and not the individual components. The second point is regarding the one-step drift of variables like $A_j(k)$, i.e., the number of active half-edges based on the community structure. The one-step drift can be unbounded as the increase can be as much as the number of nodes (in the appropriate community) minus one. We address this by assumption on the tail of the degree distributions in Lemma 5.1. There is, however, another technical issue with the one-step drift of these quantities as they depend on all terms $I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k)$ with $K_j(d_j,d_{-j}) = u_j + u_{-j} + 1$. For any finite n, we only need to account for a finite number of terms but in the limit we will have a countable number of terms leading to a similar property for the Lipschitz function(s) associated with these variable(s). This precludes the direct application of the results of [28, 29] or even the generalization in [3]. To address the specific scenario outlined above we will also present a generalization of Wormald's theorem [28, 29] that applies to our problem setting: see the statement and the proof of the extension of Wormald Theorem in Appendix E.2.

Assume the following limits hold: $\lim_{n\to\infty} 2m_1(n)/n = \lambda_1$, $\lim_{n\to\infty} m_m(n)/n =$

14

λ_m, and $\lim_{n\to\infty} 2m_2(n)/n = \lambda_2$. Then the ODEs follow by first embedding the discrete-space, discrete-time process in continuous space and time by scaling both:

$$a_j(t) = \lim_{n\to\infty} \frac{A_j(tn)}{n}, \; a_m^{(j)}(t) = \lim_{n\to\infty} \frac{A_m^{(j)}(tn)}{n},$$

$$\tau_j(t) = \lim_{n\to\infty} \frac{T_j(tn)}{n}, \; i_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(t) = \lim_{n\to\infty} \frac{I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(tn)}{n},$$

where all the limits are in probability and sample-path-wise. We can then use the one-step drift from Appendix B to derive the ODE. The details are in Appendix C.

We can prove the Lipschitz property for our one-step drifts and also certain properties of the initial condition; the details are in Appendix D. We then have the following lemma that fulfills the assumptions of the extension of Wormald's Theorem E.2.

Lemma 6.1. *Let assume the degree distributions of the random graphs in each community satisfies the regularity condition from Lemma 5.1. Then the extension of Wormald's Theorem E.2 holds for given random variables, i.e., A_j, $I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}$, T_j and $A_m^{(j)}$ with the following parameters:*

$$\beta = n^{1/3-2\Delta}, \theta = n^{-\Delta}, \gamma = n^{-1-\Delta}, \sigma(n) = O(1), b(n) = 4, a(n) = O(n^4)$$

Where $\Delta > 0$ is defined in Lemma 5.1.

Proof. See Appendix F.3. □

7   Analysis of the ODE

The following lemma characterizes the solution of the differential equations that (with high probability) approximate the adoption process.

Lemma 7.1. *The solution of differential equations* (28)-(31) *with initial condition* (32) *is given by*

$$i_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(t) = \mathbb{P}_j(d_j)\mathbb{P}_m(d_{-j})(1 - \alpha_j(d_j, d_{-j})) \times$$
$$Bi(u_j; d_j, 1 - \mu^{(j,j)}(t))Bi(u_{-j}; d_{-j}, 1 - \mu^{(j,-j)}(t)) \tag{12}$$

$$\tau_j(t) = \frac{\lambda_j}{2}\left(1 - \mu^{(j,j)}(t)^2\right) \tag{13}$$

$$a_j(t) = -\sum_{u_j+u_{-j}<K_j(d_j,d_{-j})} (d_j - u_j)i_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(t) + \lambda_j - 2\tau_j(t) \tag{14}$$

$$a_m^{(j)}(t) = -\sum_{u_j+u_{-j}<K_j(d_j,d_{-j})} (d_{-j} - u_{-j})i_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(t) + \lambda_m - (t - \tau_1(t) - \tau_2(t))$$
$$\tag{15}$$

where $j \in \{1,2\}$ and $\mu^{(j,j)}$ and $\mu^{(j,-j)}$ are the unique solution of following four dimensional differential equation,

$$\frac{-a_j(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)} = \lambda_j \frac{d\mu^{(j,j)}}{dt} \left(\mu^{(j,j)}(t) \right) \tag{16}$$

$$\frac{-a_m^{(-j)}(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)} = \lambda_m \frac{d\mu^{(j,-j)}}{dt} \left(\mu^{(-j,j)}(t) \right) \tag{17}$$

with the initial condition,

$$\left(\mu^{(1,1)}(0), \mu^{(1,2)}(0), \mu^{(2,1)}(0), \mu^{(2,2)}(0) \right) = (1,1,1,1) \tag{18}$$

and $\left(\mu^{(1,1)}(t), \mu^{(1,2)}(t), \mu^{(2,1)}(t), \mu^{(2,2)}(t) \right) \in \mathcal{M}(\epsilon)$ that will be defined later. Functions $a_j(t)$ and $a_m^{(j)}(t)$ are given by equations (14) and (15), respectively.

Proof. See Appendix F.4. $\qquad\square$

The significance of this result is in demonstrating that the set of countable ODEs from Section 6 can be reduced to a set of four dimensional ODEs. Note that this dimension reduction applies to the sample-path of the adoption process and not just the final population of active nodes as suggested by the mean-field approximation of Section 4.

Assume $\mathcal{V}(\epsilon)$ is a subset of $[0,1]^4$ such that $\forall \mu \in \mathcal{V}(\epsilon)$, we have $a_1 + a_2 + a_m^{(1)} + a_m^{(2)} \geq \epsilon$ where definition of a_j and $a_m^{(j)}$ is given by equations (14) and (15). Note that $\mathcal{V}(\epsilon)$ is a closed set. Now set $\mathcal{M}(\epsilon) = \mathcal{V}(\epsilon) \cap [\epsilon, 1]^4$. It is easy to see as long as μ lie inside the set $\mathcal{M}(\epsilon)$, the functions $f_l(\cdot)$ defined in Lemma 7.1 lie inside the set D_ϵ. Note that the dimension reduction proved in Lemma 7.1 is critical in establishing the validity of the mean-field analysis and for any numerical computations.

Tallying all the (scaled) inactive nodes we can determine the total (scaled) number of inactive nodes in community $\phi_j(t)$ for $j \in \{1,2\}$. Note that this tracks the evolution of the cascade and is yet another contribution of our paper. We gather this result in the following easily proved proposition.

Proposition 7.2. *The total (scaled) number of inactive nodes in community j, $\phi_j(t)$ for $j \in \{1,2\}$, is given by*

$$\phi_j(t) = \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} i_{d_j, d_{-j}, u_j, u_{-j}}^{(j)}(t), \tag{19}$$

i.e., $\boldsymbol{\phi}(t) = \Phi(\boldsymbol{\mu}(t))$ where $\Phi(\cdot)$ is defined in (9).

Note that as long as $a_1 + a_2 + a_m^{(1)} + a_m^{(2)} \geq \epsilon$, we can remove the denominator of differential equations for equilibrium analysis and rewrite equations in (14)

and (15) as follows:

$$\frac{d\mu^{(j,j)}}{dt} = F_{(j,j)}(\mu^{(j,j)}, \mu^{(j,-j)}) - \mu^{(j,j)} \tag{20}$$

$$\frac{d\mu^{(j,-j)}}{dt} = F_{(j,-j)}(\mu^{(-j,-j)}, \mu^{(-j,j)}) - \mu^{(j,-j)} \tag{21}$$

for all $j \in \{1, 2\}$ and the initial conditions are given by,

$$\mu^{(j,j)}(0) = \mu^{(j,-j)}(0) = 1$$

such that $\mu(t) \in \mathcal{M}(\epsilon)$ and $F_{(i,j)}(\cdot, \cdot)$ is given as follows:

$$F_{(j,j)}(\mu^{(j,j)}, \mu^{(j,-j)}) = \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} \mathbb{P}_j(d_j)\mathbb{P}_m(d_{-j})\frac{d_j}{\lambda_j} \times$$
$$(1 - \alpha_{d_j, d_{-j}}^{(j)})Bi(u_j; d_j - 1, 1 - \mu^{(j,j)})Bi(u_{-j}; d_{-j}, 1 - \mu^{(j,-j)}) \tag{22}$$

$$F_{(j,-j)}(\mu^{(-j,-j)}, \mu^{(-j,j)}) = \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} \mathbb{P}_{-j}(d_{-j})\mathbb{P}_m(d_j)\frac{d_j}{\lambda_m} \times$$
$$(1 - \alpha_{d_j, d_{-j}}^{(-j)})Bi(u_j; d_j - 1, 1 - \mu^{(-j,j)})Bi(u_{-j}; d_{-j}, 1 - \mu^{(-j,-j)}) \tag{23}$$

Note that the trajectory of differential equations after removing the denominator is same as before; hence, these equations can be used to track the evolution of cascade as well. First, we prove the function $F(\cdot)$ is increasing in each of its components.

Lemma 7.3. *If $\boldsymbol{\mu} \geq \boldsymbol{\mu}'$ component wise with $\boldsymbol{\mu} \neq \boldsymbol{\mu}'$, then $F(\boldsymbol{\mu}) \geq F(\boldsymbol{\mu}')$ component-wise, and at least for one component, $F(\boldsymbol{\mu})$ and $F(\boldsymbol{\mu}')$ are not the same.*

Proof. See Appendix F.5. $\square$

The following lemma states the most important properties of function $F(\cdot)$ which enable us to use the LaSalle Invariance Principle.

Lemma 7.4. *Let's define $U \subseteq [0, 1]^4$ to be the largest connected set containing $(1, 1, 1, 1)$ such that $\forall \boldsymbol{\mu} \in U, \boldsymbol{\mu} \geq F(\boldsymbol{\mu})$. Then we have:*
(i.) $F(U) \subseteq U$.
(ii.) U is closed and compact.
(iii.) $\forall \boldsymbol{u} \in U$, $\lim_{n \to \infty} F^n(\boldsymbol{u})$ converge to some point $\boldsymbol{u}^* \in U$ where, $F(\boldsymbol{u}^*) = \boldsymbol{u}^*$.
(iv.) If $\boldsymbol{\mu} \in U$ is a fixed point of $F(\cdot)$, then for any $\boldsymbol{\mu}' \geq \boldsymbol{\mu}$ s.t. $\boldsymbol{\mu}'_i = \boldsymbol{\mu}_i$ for some $i = 1, 2, 3, 4$, $\boldsymbol{\mu}' \notin U$.

Proof. See Appendix F.6. $\square$

Given all these properties, we have the following theorem on the solution to (9) that corresponds to the final settling point of the adoption process. In

particular, it characterizes the equilibrium point at which the ODE settles starting from our initial point and also provides a means to directly calculate the equilibrium point.

Theorem 7.5. *Consider the following ODE with given initial condition:*

$$\frac{d\boldsymbol{\mu}}{dt} = F(\boldsymbol{\mu}) - \boldsymbol{\mu} \qquad \boldsymbol{\mu}(\boldsymbol{0}) = \boldsymbol{1}, \qquad \boldsymbol{\mu} \in [0,1]^4 \tag{24}$$

The solution to the given ODE converges to the closest equilibrium point to $\boldsymbol{1}$, i.e., the fixed point of $F(\cdot)$ that is equal to $F^\infty(\boldsymbol{1})$, where $F^\infty(\boldsymbol{u}) := \lim_{n \to \infty} F^n(\boldsymbol{u})$ if it exists.

Proof. See Appendix F.7. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Note that the same Liapunov function $V(\boldsymbol{\mu}) = \|\mu - F^\infty(\boldsymbol{1})\|_2^2$ with the same steps can be used to show that the differential equations in (16)-(17) with initial condition (18) also approach a small enough ball around $F^\infty(\boldsymbol{1})$ with ball shrinking to 0 as ϵ approaches 0. Note that $\mathcal{M}(\epsilon) \in \mathcal{N}$. Now we will show that as $\epsilon \to 0$, the random process is well approximated by the ODE up to points arbitrary close to $F^\infty(\boldsymbol{1})$.

Theorem 7.6. *Consider the ODEs given by equations* (16) *and* (17) *with initial condition $\boldsymbol{\mu} = \boldsymbol{1}$ and the domain defined by $\mathcal{M}(\epsilon)$. Then, the trajectory of the solution is same as the trajectory of the solution of* (24)*, as long as it doesn't exit $\mathcal{M}(\epsilon)$. Moreover, as $\epsilon \to 0$, the trajectory hit the boundary of $\mathcal{M}(\epsilon)$ arbitrary close to $F^\infty(\boldsymbol{1})$.*

Proof. See Appendix F.8. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Hence, the process will get arbitrary close to $F^\infty(\boldsymbol{1})$. Now, the question is, whether the process stops at this point or not. The following theorem concludes this section by providing a necessary condition that implies that the process will terminate at $F^\infty(\boldsymbol{1})$.

Theorem 7.7. *Let's define the function $G : [0,1]^4 \to \mathcal{R}$ as $G(\boldsymbol{\mu}) = \boldsymbol{\mu} - F(\boldsymbol{\mu})$. Set $U \subset [0,1]^4$ to be the largest connected set containing $\boldsymbol{1}$ such that $G(\boldsymbol{\mu}) \geq 0$ for all $\boldsymbol{\mu} \in U$. If $U = U \bigcap [F^\infty(\boldsymbol{1}), \boldsymbol{1}]$, then the process terminates at $F^\infty(\boldsymbol{1})$ and therefor, final fraction of inactive nodes is arbitrary close to $\phi = \Phi(F^\infty(\boldsymbol{1}))$. Equivalently, the process terminates at $F^\infty(\boldsymbol{1})$, if,*

$$\nabla_{\boldsymbol{u}} G(F^\infty(\boldsymbol{1})) = \nabla G(F^\infty(\boldsymbol{1}))\boldsymbol{u} \not\geq 0 \qquad \forall \boldsymbol{u} \text{ such that } \|\boldsymbol{u}\|_2 = 1, \boldsymbol{u} \leq 0, \tag{25}$$

where inequalities are understood to hold component-wise, and for a vector $\mathbf{x} \not\geq 0$ implies that no components of $\mathbf{x}$ are positive and at least one component of $\mathbf{x}$ is negative.

Proof. See Appendix F.9. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

This idea can easily be generalized to any number of communities. We conclude this section by generalizing the idea to the case of k communities. The degree conditions may not have been revised for this setting.

Theorem 7.8. *Let's assume the case that we have k communities. Then, the final proportion of adopters is given by $F^\infty(\mathbf{1})$ where $F(\cdot)$ is obtained through recursive analysis of associated Galton-Watson tree with k communities if the condition in Theorem 7.7 holds. Moreover, the evolution of cascade is tractable using the differential equations. The modification needed to apply Wormald extension theorem is as follow over the following parameters: $b(n) = O(k^2), a(n) = O(n^{2k})$. In this case, the number of ODEs that we need to solve is k^2.*

For brevity, we present the proof for the case of 2 communities. The whole approach presented in this paper can be generalized easily to the case of k communities.

8   Contagion Threshold

We have used a Bernoulli random variable to decide if a node is an early adopter. We assumed the mean of the r.v. is a function of the number of her neighbors in each community. Hence, based on the advertisement, node i in community j with $d_{i,j,j}$ neighbors in community j and $d_{i,j,-j}$ neighbors in the other community, is an early adopter with probability $\alpha_j(d_{i,j,j}, d_{i,j,-j})$. If $K_j(d_j, d_{-j}) \equiv \theta(d_1 + d_2)$ for some $\theta \in (0, 1)$, then the largest value of θ that results in a cascade (i.e., $O(n)$ nodes becoming active) when a small number of nodes ($o(n)$, often taken to be a constant number) are initially seeded is called the contagion threshold; denote it by θ^*. Morris [19] showed that $\theta^* \leq 0.5$ and the upper-bound is loose for many graphs. It's argued that the contagion threshold of the graph family can be calculated by choosing $\alpha_j(d_j, d_{-j}) \equiv \alpha$, letting $\alpha \to 0$, and varying θ. In this section we will formalize this intuition and characterize the contagion condition for more general threshold functions.

Let $\boldsymbol{\alpha} = \{\alpha_j(d_j, d_{j'})\}_{d_j, d_{j'}, j, j'}$ represent the seeding strategy. Let's rewrite the function $F(\boldsymbol{\mu})$ as $F(\boldsymbol{\alpha}, \boldsymbol{\mu})$ to emphasis on the dependency of function $F(\cdot)$ over the seeding strategy. The question of interest is the final proportion of adopters, if the seeding effects only a finite number of population, i.e., the proportion of early adopters goes to 0 as $n \to \infty$. The following lemma characterizes the answer to this question.

Theorem 8.1. *Consider an arbitrary sequence $\{\boldsymbol{\alpha}_s\}_{s=1}^{\infty}$ that represent a sequence of non-zero seeding strategies that converges to zero, i.e., $\lim_{s \to \infty} \boldsymbol{\alpha}_s = \mathbf{0}$. Let's define $U \subseteq [0,1]^4$ to be the largest connected set containing $(1,1,1,1)$ such that $\forall \boldsymbol{\mu} \in U, \boldsymbol{\mu} \geq F(\mathbf{0}, \boldsymbol{\mu})$. If U is singleton, i.e., $U = \{\mathbf{1}\}$, then the final proportion of adopters converges to 0 as $\boldsymbol{\alpha}_s \to \mathbf{0}$. Otherwise, the final proportion of adopters is strictly positive. Moreover, we have,*

$$\lim_{s \to \infty} F^\infty(\boldsymbol{\alpha}_s, \mathbf{1}) = F^\infty(\mathbf{0}, \boldsymbol{\mu}) \qquad \forall \mu \in U/\{\mathbf{1}\}$$

Proof. See Appendix F.12. □

A contagion is said to happen if starting from a finite seed set, the final proportion of adopters is nonzero. Theorem 8.1 provides both the necessary

and sufficient condition for this. Note that, if the set U is singleton, then we won't have any contagion. The following corollary provides an easy way to check whether contagion happens or not.

Corollary 8.2. *Let's define the set U and function G as in Lemma 7.4 while $\boldsymbol{\alpha}$ is set to be zero. Then if U is singleton, we will not have a contagion. Equivalently, contagion will not happen if and only if,*

$$\nabla_{\boldsymbol{u}} G(\boldsymbol{0}, \boldsymbol{1}) = \boldsymbol{u} - \nabla F(\boldsymbol{0}, \boldsymbol{1})\boldsymbol{u} \ngeq 0 \qquad \forall \boldsymbol{u} \text{ such that } \|\boldsymbol{u}\|_2 = 1, \boldsymbol{u} \leq 0 \qquad (26)$$

Using Corollary **??**, we can check whether contagion happens or not by solving a maximization problem.

Corollary 8.3. *Contagion will happen if and only if,*

$$\max_{\boldsymbol{u} \in [0,1]^4} \quad \boldsymbol{u}^T \nabla F(\boldsymbol{0}, \boldsymbol{1})\boldsymbol{u} > 1 \qquad (27)$$
$$s.t \quad \|\boldsymbol{u}\|_2 = 1, \ \boldsymbol{u} \leq 0.$$

Note from Appendix 7.3 that the gradient of function $F(\cdot)$ is non-negative. The Perron-Frobenius theorem then yields the following equivalent result.

Theorem 8.4. *Contagion will happen if and only if $\rho\big(\nabla F(\boldsymbol{0}, \boldsymbol{1})\big)$ is greater than 1, where $\rho(\cdot)$ is the Perron-Frobenius eigenvalue of a non-negative matrix.*

Note that the discussion on contagion can also be generalized to k communities with exactly the same statement as in Theorem 8.4.

9 Poisson degree distributions

We will now specialize our results to Poisson degree distributions. An Erdos-Renyi random graph is an example of a graph family that asymptotically yields a Poisson degree distribution. The two community stochastic block model is then the appropriate generalization of the Erdos-Renyi random graph that will asymptotically produce Poisson degree distributions within the community and across the community. We will show in the following results that under some symmetry assumptions for the threshold and the advertising strategy, the solution simplifies considerably.

Theorem 9.1. *Let's assume the threshold is a function of number of neighbors, i.e. $K_j(d_j, d_{-j}) = K_j(d_j + d_{-j})$. Moreover, assume that the advertisement is symmetric in each camp over nodes with equal number of neighbors, i.e. $\alpha_j(d_j, d_{-j}) = \alpha_j(d_1 + d_2)$. If the degree distributions are Poisson with parameter λ_j, $j \in \{1, 2, m\}$, then $\mu^{(1,1)}(t) = \mu^{(2,1)}(t)$ and $\mu^{(2,2)}(t) = \mu^{(1,2)}(t)$ so that the dimension of differential equations will reduce to 2. In the case of k communities, the dimension will reduce to k.*

Proof. See Appendix F.10. □

20

Theorem 9.2. *Consider the case that the advertisement and threshold is symmetric, i.e. $\alpha_1(d_1, d_2) = \alpha_2(d_2, d_1)$ and $K_1(d_1, d_2) = K_2(d_2, d_1)$ for all d_1 and d_2. Moreover, also assume that the degree distribution in both communities are the same, i.e. $P_1(\cdot) = P_2(\cdot)$. Then, the number of differential equations can be reduced to 2. In the case of k communities, the dimension will reduce to k.*

Proof. See Appendix F.11. □

Corollary 9.3. *If the assumptions in Theorems 9.1 and 9.2 hold, then number of differential equations reduced to one. In this case, the formulation is just the same as if there was only one community that has a Poisson degree distribution with parameter $\lambda_1 + \lambda_m$. In the case of k communities, the parameter of equivalent one community is given by $\lambda_1 + \sum_{i=1}^{k-1} \lambda_{m,i}$, where $\lambda_{m,i}$ is the mean degree of bi-partite graph between community 1 and community $i + 1$ for $i = 1, \dots, k - 1$.*

Proof. Follows from Theorems 9.1 and 9.2. □

It is interesting to note that in the case of identical Poisson distributions within each community and with additive advertisement strategy, the contagion threshold is the same as if there was only one community. The derivation of contagion threshold match with the ones presented in [3, 15] for case of one community.

10   Results

We present some numerical results using the analysis presented above. The main point is to show how the community structure impacts seeding strategies. Most of our results will be for Poisson degree distributions, owing to analytical simplifications and the fact there is only one parameter to tune. Moreover, for simplicity we assume that the threshold function is given by $k_j(d_j, d_{-j}) = \theta \times (d_j + d_{-j})$ where $\theta = 0.25$. We will evaluate the best action that can be taken by an advertiser of the new behavior based on his information. The nodes that are seeded by the advertisers are early adopters. A few strategies that we will consider are: (1) Random seeding: First, we assume the advertiser does not even know about the existence of two communities. This scenario is named as *global seeding*. Second, we assume the advertiser knows the community structure, and decides to seed just asymmetrically in the two communities. This advertisement strategy is denoted by *local seeding*. (2) Degree-targeted seeding: the advertiser knows the degree distribution of the network and the identity of the nodes that possess a certain degree, but does not know the underlying connectivity structure. For simplicity, here we target a fixed degree.

In Figure 2, we assume both communities have the same degree distribution $Poi(\lambda_{in})$. The inter-community degree distribution is $Poi(\lambda_{out})$. The figure suggests that if the communities are symmetric, and if they are well-connected ($\lambda_{out} = 1$), then the best strategy is to put the whole budget in one community. In Figures 3 & 4, we consider the general case where distributions can be different in the two communities: $Poi(\lambda_{in,1})$ and $Poi(\lambda_{in,2})$, respectively. In
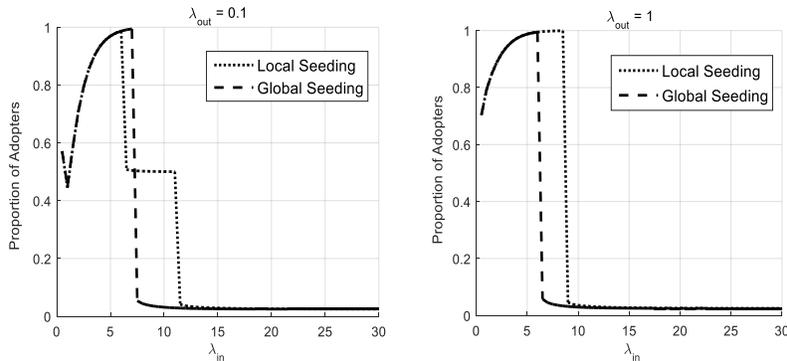
Figure 2: Symmetric communities with random seeding; proportion of early adopters is 5%, x axis denotes the expected number of edges in each community and y the proportion of population that will adopt.

this case, the community structure dramatically changes the cascade potential: there are scenarios where global seeding can cause a cascade while local seeding won't, and *vice-versa*.
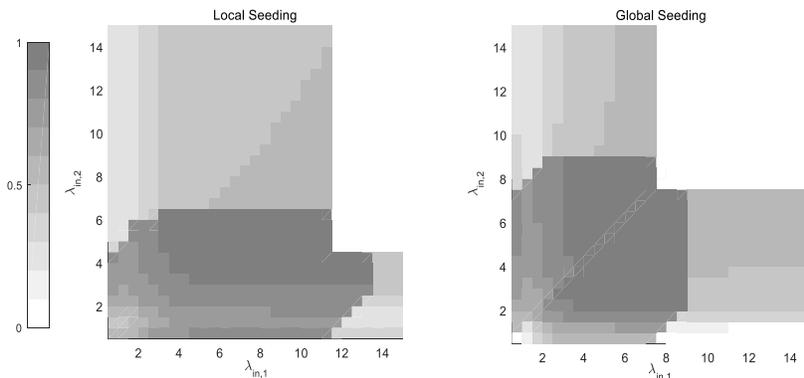

Figure 3: Random Seeding, $\lambda_{out} = 0.1$; proportion of early adopters is 5%, x axis is the expected number of connections in community 1, y the expected number of connections in community 2. Intensity of color denotes the proportion of adopters.

Next we consider degree-targeted seeding in Figures 5 and 6. In general high-degree nodes can potentially stop a cascade if they are not adopters; hence it might make sense to seed these nodes in each community. We will consider the following case: the budget is spread equally in both communities, denoted by $(0.5, 0.5)$; the budget is concentrated in community 1, denoted by $(1, 0)$; and that 25% of budget is in community 1, denoted by $(0.25, 0.75)$. The outer
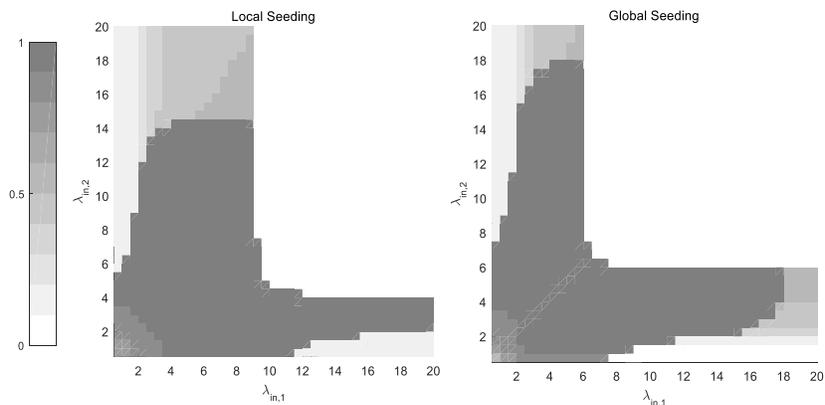
Figure 4: Random Seeding, $\lambda_{out} = 1$, The proportion of early adopters is 5%, x axis the the expected number of connections in community 1 and the y axis is the expected number of connections in community 2. Intensity of color determint the porportion of adopters.

connectivity is given by $\lambda_{out} = 1$ and $\lambda_{out} = 0.1$. The main observation is the dramatic difference in the proportion of final adopters based on how asymmetric the targeting is. Additionally, a higher inter-community connectivity leads to a bigger cascade. Also note that seeding nodes with the highest degree gives better result than random seeding.
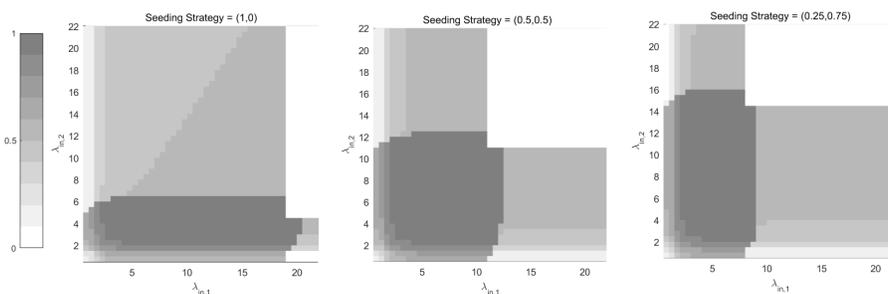

Figure 5: Attacking highest degrees, $\lambda_{out} = 0.1$, The proportion of early adopters is 5%, x axis the the expected number of connections in community 1 and the y axis is the expected number of connections in community 2. Intensity of grayscale indicates the final proportion of adopters.

Next, we discuss the evolution of cascade using Lemma 7.1. Figure 7 represents the evolution of active half-edges and inactive nodes in the second community when $\lambda_{in,1} = 7$, $\lambda_{in,2} = 12$, and $\lambda_{out} = 1$ when the seeding strategy is to put the whole budget in the first community. Figure 4 suggests that global
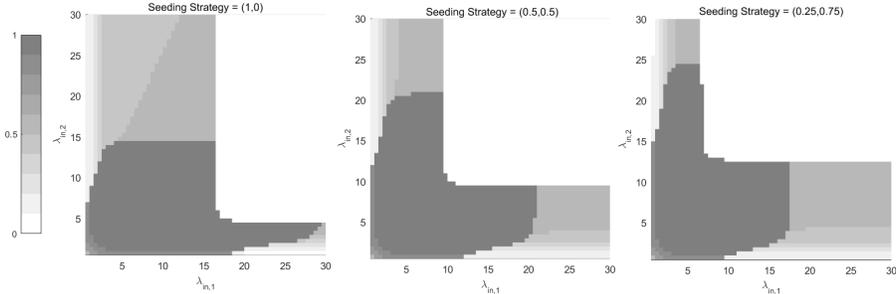
23

Figure 6: Seed highest degrees, $\lambda_{out} = 1$, The proportion of early adopters is 5%, x axis the the expected number of connections in community 1 and the y axis is the expected number of connections in community 2. Intensity of grayscale indicates the final proportion of adopters.

seeding will not result in any cascade. On the other hand, a global cascade emerges following local seeding: it develops in the first community and then moves to the next community; this happens when the inactive nodes in community 2 with $d_1 \geq \theta \times (d_1 + d_2)$ become active, causing a cascade in the second community.

Finally, there are scenarios where neither global nor local seeding causes cascade. Figure 8 represents the evolution of active-half edges, using $(0.25, 0.75)$ seeding strategy. As can be seen from Figure 6, the only seeding strategy (among the ones analyzed) that can cause a global cascade when $\lambda_{in,1} = 17$, $\lambda_{in,2} = 12$ and $\lambda_{out} = 1$ is $(0.25, 0.75)$. Active-half edges in both communities get close to zero, but nevertheless a cascade happens in the second community. This cascade then gets transferred to the first community, and almost all nodes adopt the new technology. This example illustrates the importance of active-half edges $a_m^{(2)}(t)$ in triggering cascade in the first community.

11   Conclusion

We studied cascades under the threshold model and assuming permanent adoption, on sparse random graphs with a community structure to see to what extent cascades are affected by the community structure. When seeding a small number of agents with the new behavior, the community structure has little effect on the final proportion of people that adopt it, i.e., the contagion threshold is the same as if there were just one community. On the other hand, seeding a fraction of population with the new behavior has a significant impact on the cascade with the optimal seeding strategy depending on how strongly the communities are connected.
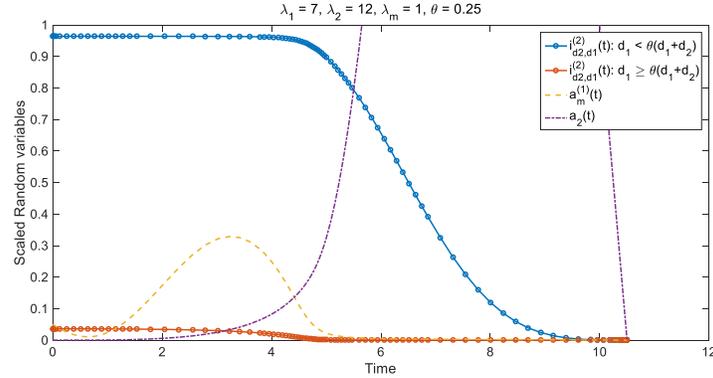
Figure 7: Evolution of cascade, $\lambda_{out} = 1$, $\lambda_{in,1} = 7$ and $\lambda_{in,2} = 12$; proportion of early adopters is 5%, x axis is time and y axis is the quantity of corresponding scaled variables. The seeding is local.

References

[1] Aldous, D. and R. Lyons (2007). Processes on unimodular random networks. *Electron. J. Probab 12*(54), 1454–1508.

[2] Aldous, D. and J. M. Steele (2004). The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, pp. 1–72. Springer.

[3] Amini, H. (2010). Bootstrap percolation and diffusion in random graphs with given vertex degrees. *Electronic Journal of Combinatorics 17*, R25.

[4] Bollobás, B. (2001). Random graphs. 1985. *Academic, London*.

[5] Easley, D. and J. Kleinberg (2010). *Networks, crowds, and markets: Reasoning about a highly connected world.* Cambridge University Press.

[6] Galstyan, A. and P. Cohen (2007). Cascading dynamics in modular networks. *Physical Review E 75*(3), 036109.

[7] Galstyan, A., V. Musoyan, and P. Cohen (2009). Maximizing influence propagation in networks with community structure. *Physical Review E 79*(5), 056102.

[8] Gleeson, J. P. (2008). Cascades on correlated and modular random networks. *Physical Review E 77*(4), 046117.

[9] Gleeson, J. P. and D. J. Cahalane (2007). Seed size strongly affects cascades on random networks. *Physical Review E 75*(5), 056103.
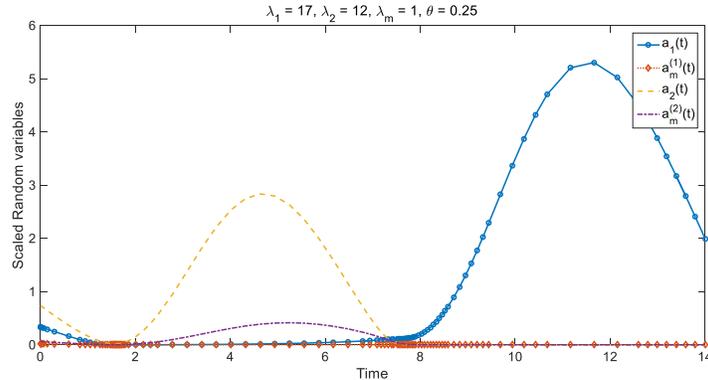
Figure 8: Evolution of cascade, $\lambda_{out} = 1$, $\lambda_{in,1} = 17$ and $\lambda_{in,2} = 12$; proportion of early adopters is 5%, x axis is time and y axis is the quantity of corresponding scaled variables. The seeding is given by $(0.25, 0.75)$.

[10] Granovetter, M. (1978). Threshold models of collective behavior. *American journal of sociology*, 1420–1443.

[11] Janson, S. (2009). The probability that a random multigraph is simple. *Combinatorics, Probability and Computing 18*(1-2), 205–225.

[12] Janson, S. and M. J. Luczak (2007). A simple solution to the k-core problem. *Random Structures & Algorithms 30*(1-2), 50–62.

[13] Janson, S. and M. J. Luczak (2009). A new approach to the giant component problem. *Random Structures & Algorithms 34*(2), 197–216.

[14] Kempe, D., J. Kleinberg, and E. Tardos (2003). Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, New York, NY, USA, pp. 137–146. ACM.

[15] Lelarge, M. (2012). Diffusion and cascading behavior in random networks. *Games and Economic Behavior 75*(2), 752–775.

[16] López-Pintado, D. (2008). Diffusion in complex social networks. *Games and Economic Behavior 62*(2), 573–590.

[17] Molloy, M. and B. Reed (1998). The size of the giant component of a random graph with a given degree sequence. *Combinatorics, probability and computing 7*(03), 295–305.

[18] Molloy, M. and B. A. Reed (1995). A critical point for random graphs with a given degree sequence. *Random structures and algorithms 6*(2/3), 161–180.

[19] Morris, S. (2000). Contagion. *The Review of Economic Studies 67*(1), 57–78.

[20] Mossel, E., J. Neeman, and A. Sly (2012). Stochastic block models and reconstruction. *arXiv preprint arXiv:1202.1499*.

[21] Mossel, E. and S. Roch (2010). Submodularity of influence in social networks: From local to global. *SIAM Journal on Computing 39*(6), 2176–2188.

[22] Naylor, R. (1990). A social custom model of collective action. *European Journal of Political Economy 6*(2), 201–216.

[23] Samoilenko, A. M. and Y. V. Teplinskii (2003). *Countable systems of differential equations*. Walter de Gruyter.

[24] Schelling, T. C. (1978, October). *Micromotives and Macrobehavior*. W. W. Norton & Company.

[25] Vega-Redondo, F. (2007). *Complex social networks*. Number 44. Cambridge University Press.

[26] Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences 99*(9), 5766–5771.

[27] Watts, D. J. and P. S. Dodds (2007). Influentials, networks, and public opinion formation. *Journal of consumer research 34*(4), 441–458.

[28] Wormald, N. C. (1995). Differential equations for random processes and random graphs. *The annals of applied probability*, 1217–1235.

[29] Wormald, N. C. (1999). The differential equation method for random graph processes and greedy algorithms. *Lectures on approximation and randomized algorithms*, 73–155.

APPENDIX

A   When greedy maximizing is bad

We now give an example showing how the results in [14, 21] break if θ_v is assumed to be fixed. We build a network as follows: start from a $3n \times 3n$ torus, i.e. node (i, j) with $1 \leq i, j \leq 3n$ has four neighbors: $(i+1, j), (i-1, j), (i, j+1), (i, j-1)$ where operation are done modulo $3n$. Now for each $1 \leq j \leq 3n$, and $0 \leq k \leq n-1$, we add a vertex $v(j, k)$ connected to the vertices of the torus $(3k+1, j)$, $(3k+2, j)$ and $(3k+3, j)$. Finally each of these vertices $v(j, k)$ are part of a cycle of size $K \geq 3$ with no other common point with the rest of the graph except through $v(j, k)$. In summary, we have $9n^2$ nodes on the torus, and $3n^2$ disjoint cycles of size K which are connected to the torus only through the vertices $v(j, k)$. There is a total of $9n^2 + 3n^2 K$ vertices. Note that the degree of a node on the torus is 5 (4 neighbors on the torus and 1 on a cycle) as well as for the nodes $v(j, k)$. We take $\theta = 2/5$ so that a node of degree d becomes active as soon as θd of its neighbors are active. In particular a node on the torus or a $v(j, k)$ needs only 2 active neighbors to become active. Moreover, activating a vertex $v(j, k)$ will activate all the K nodes on the cycle. Because of this, it is easy to see that any greedy algorithm with budget $b \leq 3n^2$ will only activate the vertices $v(j, k)$. Note however that by acting the set of nodes on the torus: $(1, 1), (1, 2), \ldots, (1, 3n)$ and $(2, 1)$ will result in a global activation of the network. Hence for any $3n + 1 \leq b \leq 3n^2$, we can find a set activating the $9n^2 + 3n^2 K$ vertices of the networks, whereas the greedy algorithm only activate Kb vertices which is far from the optimum solution.

B   One Step Drift

In this section, the one step drift of each random variable is presented:

- One step drift of $A_j(\cdot)$ for all $j \in \{1, 2\}$ (active half-edges belongs to

community j):

$$\mathbb{E}\left[A_j(k+1) - A_j(k)|H_t\right] =$$
$$- \frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)}$$
$$- \frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{A_j(k) - 1}{2m_j(k) - 2T_j(k) - 1}$$
$$+ \frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times$$
$$\sum_{u_j + u_{-j} + 1 = K_j(d_j, d_{-j})} (d_j - u_j - 1) \times \frac{(d_j - u_j)I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k)}{2m_j(k) - 2T_j(k) - 1}$$
$$+ \frac{A_m^{(-j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times$$
$$\sum_{u_j + u_{-j} + 1 = K_j(d_j, d_{-j})} (d_j - u_j) \times \frac{(d_{-j} - u_{-j})I^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(k)}{m_m(n) - (t - T_1(k) - T_2(k))}$$

- One step drift of $T_j(\cdot)$ for all $j \in \{1, 2\}$ (time the process spent in community j):

$$\mathbb{E}\left[T_j(k+1) - T_j(k)|H_t\right] =$$
$$+ \frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)}$$

- One step drift of $A_m^{(j)}(\cdot)$ for all $j \in \{1, 2\}$ (active half-edges in community

29

j pointing to the other community):

$$\mathbb{E}\left[A_m^{(j)}(k+1) - A_m^{(j)}(k)|H_t\right] =$$

$$-\frac{A_m^{(j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)}$$

$$-\frac{A_m^{(-j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{A_m^{(j)}(k)}{m_m(n) - (t - T_1(k) - T_2(k))}$$

$$+\frac{A_m^{(-j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times$$

$$\sum_{u_j+u_{-j}+1=K_j(d_j,d_{-j})} (d_{-j} - u_{-j} - 1) \times \frac{(d_{-j} - u_{-j})I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k)}{m_m(n) - (t - T_1(k) - T_2(k))}$$

$$+\frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times$$

$$\sum_{u_j+u_{-j}+1=K_j(d_j,d_{-j})} (d_{-j} - u_{-j}) \times \frac{(d_j - u_j)I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k)}{2m_j(n) - 2T_j(k) - 1}$$

- One step drift of $I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(\cdot)$ for all $j \in \{1,2\}$, $(d_j, u_j) \in \mathbb{Z}_+^2$ with $u_j \leq d_j$, and $(d_{-j}, u_{-j}) \in \mathbb{Z}_+^2$ with $u_{-j} \leq d_{-j}$ (Inactive nodes in community j) :

$$\mathbb{E}\left[I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k+1) - I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k)|H_t\right] =$$

$$-\frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{(d_j - u_j)I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k)}{2m_1(k) - 2T_1(k) - 1}$$

$$-\frac{A_m^{(-j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{(d_{-j} - u_{-j})I_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(k)}{m_m(n) - (t - T_1(k) - T_2(k))}$$

$$+\frac{A_j(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k)} \times \frac{(d_j - u_j + 1)I_{d_j,d_{-j},u_j-1,u_{-j}}^{(j)}(k)}{2m_j(k) - 2T_j(k) - 1}$$

$$+\frac{A_m^{(-j)}(k)}{A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(-j)}(k)} \times \frac{(d_{-j} - u_{-j} + 1)I_{d_j,d_{-j},u_j,u_{-j}-1}^{(j)}(k)}{m_m(n) - (t - T_1(k) - T_2(k))}$$

C   Derivation of ODE

The limiting variables are given by the following equations.

$$\frac{da_j}{dt} = f_j(t, a_1, a_2, a_m^{(1)}, a_m^{(2)}, \tau_1, \tau_2, i_{d_1,d_2,u_1,u_2}^{(1)}, i_{d_2,d_1,u_2,u_1}^{(2)})$$

$$:= -\frac{a_j(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)} - \frac{a_j(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)} \times \frac{a_j(t)}{\lambda_1 - 2\tau_1(t)}$$

(28)

$$+ \frac{a_j(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)} \times \sum_{u_j + u_{-j} + 1 = K_j(d_j, d_{-j})} (d_j - u_j - 1) \frac{(d_j - u_j)i_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(t)}{\lambda_j - 2\tau_j(t)}$$

$$+ \frac{a_m^{(-j)}(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)} \times \sum_{u_j + u_{-j} + 1 = K_j(d_j, d_{-j})} (d_j - u_j) \frac{(d_{-j} - u_{-j})i_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(t)}{\lambda_m - (t - \tau_1(t) - \tau_2(t))}$$

$$\frac{da_m^{(j)}}{dt} = f_{j+2}(t, a_1, a_2, a_m^{(1)}, a_m^{(2)}, \tau_1, \tau_2, i_{d_1,d_2,u_1,u_2}^{(1)}, i_{d_2,d_1,u_2,u_1}^{(2)})$$

$$:= -\frac{a_m^{(j)}(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)} - \frac{a_m^{(-j)}(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)} \times \frac{a_m^{(j)}(t)}{\lambda_m - (t - \tau_1(t) - \tau_2(t))}$$

(29)

$$+ \frac{a_m^{(-j)}(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)} \times \sum_{u_j + u_{-j} + 1 = K_j(d_j, d_{-j})} (d_{-j} - u_{-j} - 1) \frac{(d_{-j} - u_{-j})i_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(t)}{\lambda_m - (t - \tau_1(t) - \tau_2(t))}$$

$$+ \frac{a_j(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)} \times \sum_{u_j + u_{-j} + 1 = K_j(d_j, d_{-j})} (d_{-j} - u_{-j}) \frac{(d_j - u_j)i_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(t)}{\lambda_j - 2\tau_j(t)}$$

$$\frac{d\tau_j}{dt} = f_{j+4}(t, a_1, a_2, a_m^{(1)}, a_m^{(2)}, \tau_1, \tau_2, i_{d_1,d_2,u_1,u_2}^{(1)}, i_{d_2,d_1,u_2,u_1}^{(2)}) := \frac{a_j(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)}$$

(30)

$$\frac{di_{d_j,d_{-j},u_j,u_{-j}}^{(j)}}{dt} = f_{j,d_j,d_{-j},u_j,u_{-j}}(t, a_1, a_2, a_m^{(1)}, a_m^{(2)}, \tau_1, \tau_2, i_{d_1,d_2,u_1,u_2}^{(1)}, i_{d_2,d_1,u_2,u_1}^{(2)})$$

$$:= -\frac{a_j(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)} \frac{(d_j - u_j)i_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(t)}{\lambda_j - 2\tau_j(t)}$$

(31)

$$- \frac{a_m^{(-j)}(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)} \frac{(d_{-j} - u_{-j})i_{d_j,d_{-j},u_j,u_{-j}}^{(j)}(t)}{\lambda_m - (t - \tau_1(t) - \tau_2(t))}$$

$$+ \frac{a_j(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(-j)}(t)} \times \frac{(d_j - u_j + 1)i_{d_j,d_{-j},u_j-1,u_{-j}}^{(j)}(t)}{\lambda_j - 2\tau_j(t)}$$

$$+ \frac{a_m^{(-j)}(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(-j)}(t)} \times \frac{(d_{-j} - u_{-j} + 1)i_{d_j,d_{-j},u_j,u_{-j}-1}^{(j)}(t)}{\lambda_m - (t - \tau_1(t) - \tau_2(t))}$$

The initial value is given by the following for $j \in \{1, 2\}$:

$$a_j(0) = \sum_{d_j, d_{-j}} \mathbb{P}_j(d_j)\mathbb{P}_m(d_{-j})d_j\alpha_j(d_j, d_{-j}),$$

$$a_m^{(j)}(0) = \sum_{d_j, d_{-j}} \mathbb{P}_j(d_j)\mathbb{P}_m(d_{-j})d_{-j}\alpha_j(d_j, d_{-j}),$$

$$\tau_j(0) = 0,$$

$$i_{d_j, d_{-j}, u_j, u_{-j}}^{(j)}(0) = \begin{cases} \mathbb{P}_j(d_j)\mathbb{P}_m(d_{-j})\big(1 - \alpha_j(d_j, d_{-j})\big) & \text{if } u_j = u_{-j} = 0 \\ 0 & \text{otherwise} \end{cases}.$$

The balance equations can be written as follows in the scaled variables

$$a_j(t) + \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} (d_j - u_j)i_{d_j, d_{-j}, u_j, u_{-j}}^{(j)}(t) = \lambda_j - 2\tau_j(t)$$

$$a_m^{(j)}(t) + \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} (d_{-j} - u_{-j})i_{d_j, d_{-j}, u_j, u_{-j}}^{(j)}(t) = \lambda_m - (t - \tau_1(t) - \tau_2(t))$$

Note that the balance equations are not related to the differential equations. However, if the solution of ODEs track the random variables closely, then it is expected that balance equations to be satisfied by the solution as well.

D   Technical Conditions

Note that conditions given in Lemma 5.1 are sufficient to use Theorem E.2 to track the evolution of process. We will establish this next. Let define the set $D_\epsilon := \bigcup D_\epsilon(n)$ such that $D_\epsilon(n)$ is an open connected set that contains the closure of:

$$S(n, \epsilon) = \Big\{ \zeta = (t, a_1, a_2, a_m^{(1)}, a_m^{(2)}, \tau_1, \tau_2, i_{d_1, d_2, u_1, u_2}^{(1)}, i_{d_2, d_1, u_2, u_1}^{(2)}, 0, 0, \dots) \mid \zeta \text{ satisfies } \mathcal{A} \Big\},$$

where the conditions in $\mathcal{A}$ are given as follows:
1.) $\zeta(1) = t \geq 0$.
2.) $\zeta(2) + \zeta(3) + \zeta(4) + \zeta(5) = a_1 + a_2 + a_m^{(1)} + a_m^{(2)} > \epsilon$.
3.) $\zeta(1) - \zeta(6) - \zeta(7) = t - \tau_1 - \tau_2 < \lambda_m - \epsilon$.
4.) $2\zeta(6) = 2\tau_1 < \lambda_1 - \epsilon$.
5.) $2\zeta(7) = 2\tau_2 < \lambda_2 - \epsilon$
Moreover, let us assume that all elements of $D_\epsilon(n)$ satisfy a modified version of

balance equations, i.e.

$$\forall \zeta \in D_\epsilon(n):$$

$$a_j + \sum_{u_j+u_{-j}<K_j(d_j,d_{-j})} (d_j - u_j)i^{(j)}_{d_j,d_{-j},u_j,u_{-j}} + 2\tau_j = O(1)$$

$$a_m^{(j)} + \sum_{u_j+u_{-j}<K_j(d_j,d_{-j})} (d_{-j} - u_{-j})i^{(j)}_{d_j,d_{-j},u_j,u_{-j}} + (t - \tau_1 - \tau_2) = O(1)$$

$$\sum_{u_j+u_{-j}<K_j(d_j,d_{-j})} (d_{-j} - u_{-j})^2 i^{(j)}_{d_j,d_{-j},u_j,u_{-j}} = O(1)$$

$$\sum_{u_j+u_{-j}<K_j(d_j,d_{-j})} (d_j - u_j)^2 i^{(j)}_{d_j,d_{-j},u_j,u_{-j}} = O(1)$$

such that for large enough n, all of terms on the right side of the equations are bounded by K provided in Lemma 5.1. Furthermore, we will also assume that all vectors that satisfy the balance equations above are contained in $D_\epsilon(n)$. Finally, note that for any two points $\mathbf{a}, \mathbf{b} \in D_\epsilon(n)$, we have $L(\mathbf{a}, \mathbf{b}) \in D_\epsilon(n)$ where $L(\mathbf{a}, \mathbf{b})$ is the line connecting these two points, i.e. $D_\epsilon(n)$ is convex. It is easy to see that elements of $L(\mathbf{a}, \mathbf{b})$ also satisfy the modified balance equations if $\mathbf{a}$ and $\mathbf{b}$ do so. Now, we will show that for any arbitrary label l, the function $f_l(\cdot)$ is a Lipschitz function within the set D_ϵ.

Lemma D.1. *Let assume D_ϵ is given as above, then it is bounded. Moreover, all functions $f_l(\cdot)$ given in equations (28)-(31) are Lipschitz over D_ϵ with some fixed Lipschitz constant L.*

Proof. See Appendix F.2. □

We need to verify that the initial conditions satisfy

$$\liminf_{n\to\infty} \mathbb{P}\left(\left(\frac{0}{n}, \frac{A_1(0)}{n}, \frac{A_2(0)}{n}, \frac{A_m^{(1)}(0)}{n}, \frac{A_m^{(2)}(0)}{n}, \frac{T_1(0)}{n}, \frac{T_2(0)}{n},\right.\right.$$
$$\left.\left.\frac{I^{(1)}_{d_1,d_2,u_1,u_2}(0)}{n}, \frac{I^{(2)}_{d_2,d_1,u_2,u_1}(0)}{n}, 0, 0, \dots\right) \in D_\epsilon(n)\right) > 0,$$

where n is the total number of nodes. Furthermore, we need all of these random variables to uniformly converge to their means *a.s.* Both of these follow from the regularity conditions and the results of [18]. Proving the Lipschitz hypothesis for functions $f_l(\cdot)$ and satisfying the properties for the initial conditions, the only remaining step to fulfill the assumptions of the extension of Wormald's Theorem E.2 is given by the following lemma. '

E Extension of Wormald's Theorem

We track the approach presented by Wormald closely with some modification to generalize his theorem to our needs. In the proof of generalized version of his theorem, we will use Azuma inequality which is given as follow.

Lemma E.1. *(Azuma inequality) Let $X_0, X_1, \ldots, X_t$ be a supermartingale where $X_0 = 0$ and $|X_i - X_{i-1}| \leq c_i$ for $i \geq 1$ and constants c_i. Then for all $\alpha > 0$,*

$$\mathbb{P}(X_t \geq \alpha) \leq \exp\left(-\frac{\alpha^2}{2\sum c_i^2}\right)$$

The Extension of Wormald's Theorem is as follows.

Theorem E.2. *For $1 \leq l \leq a(n)$, where $a(n)$ is an increasing function of n, let assume $Y_l^n(t)$ is a real-valued random variable such that $Y_l^n(t) \leq C_0 n$ for some constant C_0. Let define $Y_l^n(t) = 0$ for all $l > a(n)$. Assume the following conditions hold, where $D(n)$ is some bounded connected open set containing the closure of:*

$$\{(0, z_1, z_2, \ldots) : \mathbb{P}(Y_l^n(0) = z_l n, \quad \forall l) > 0\}$$

and $D := \bigcup_k D(k)$

i. *(Boundedness hypothesis.) For some functions $\beta = \beta(n) \geq 1$, $\gamma = \gamma(n)$ and $b(n)$ we have,*

$$\mathbb{P}(\max_{l \leq b(n)} |Y_l^n(t+1) - Y_l^n(t)| \leq \beta \mid H_t) \geq 1 - \gamma \tag{33}$$

$$\mathbb{P}(\sup_{l > b(n)} |Y_l^n(t+1) - Y_l^n(t)| \leq \beta \mid H_t) = 0 \tag{34}$$

for all t, where H_t is history upto time t

ii. *(Trend hypothesis.) For some function $\theta_1 = \theta_1(n) = o(1)$ and for all $l \leq a(n)$,*

$$|\mathbb{E}(Y_l^n(t+1) - Y_l^n(t) \mid H_t) - f_l(t/n, Y_1^n(t)/n, Y_2^n(t)/n, \ldots)| \leq \theta_1$$

condition on boundedness hypothesis.

iii. *(Lipschitz hypothesis.) The functions f_l satisfy the Cauchy-Lipschitz condition with uniformly bounded Lipschitz constant L on the set $D \bigcap \{(t, z_1, z_2, \ldots) : t \geq 0\}$,*

$$|f_l(x_1, x_2, x_3, \ldots) - f_l(x_1', x_2', x_3', \ldots)| \leq L \sup_i \|x_i - x_i'\|,$$

In particular, they are continuous in their first component, i.e.,

$$f_l(x + \Delta x, y_1, y_2, \ldots) - f_l(x, y_1, y_2, \ldots) \to 0 \qquad as \qquad \Delta x \to 0$$

iv. *The following inequalities hold for some fix and bounded values $y_1, y_2, y_3, \ldots$:*

$$|f_l(x, y_1, y_2, y_3, \ldots)| \leq F(x)$$

for some continuous function $F(\cdot)$. Note that f_ls are continuously differentiable according to x_1.

Then the following are true.

(a) *For $(0, \hat{z}_1, \hat{z}_2, \dots) \in D$ the system of differential equations*

$$\frac{dz_l}{dx} = f_l(x, z_1, z_2, \dots),$$

has a unique solution in D for $z_l : \mathbb{R} \to \mathbb{R}$ passing through

$$z_l(0) = \hat{z}_l,$$

and which extends to points arbitrary close to the boundary of D.

(b) *Let $\theta > \theta_1 + C_0 n\gamma$ with $\theta = o(1)$. For a sufficiently large constant C, with probability $1 - O(b(n)\sigma(n)n\gamma + \frac{a(n)\beta}{\theta}exp(-\frac{n\theta^3}{\beta^3}))$,*

$$Y_l^n(t) = nz_l(t/n) + O(\theta n)$$

uniformly (convergence implicit in the $o(1)$ terms) for $0 \le t \le \sigma n$ and for each l, where $z_l(x)$ is solution in (a) with $\lim_{n \to \infty} max_{1 \le l \le a(n)} |Y_l^n(0) - \hat{z}_l n| = 0$, and $\sigma = \sigma(n)$ is the supremum of those x to which the solution can be extended before reaching within l^∞-distance $C\theta$ of the boundary of D. Note that $\sigma(n)$ is $O(1)$.

Proof. Authors in [23, Ch. 1, Sec. 1, p. 7] showed that the given system of differential equation in (a) has a unique solution in the domain H:

$$x \in (-\infty, +\infty) \quad \sup\{|z_1|, |z_2|, \dots\} < \infty,$$

which contains D; since for any k, $D(k)$ lay inside the set $\mathbb{R} \times [0, C_0]^{\mathbb{N}}$. Note that the assumptions in the statement of the theorem are not as same as the ones presented in [23, Ch. 1, Sec. 1, p. 7]. The changes are almost trivial. One can even prove that given condition *iii.* there is no need to impose condition *iv.* on functions $f_l(\cdot)$.
Take $\theta > \theta_1$ as in (b), define

$$w = \left\lceil \frac{n\theta}{\beta} \right\rceil$$

and let $t \ge 0$. If $\beta/\theta > n^{1/3}$ then $\beta/\theta \times exp(-n\theta^3/\beta^3) \ge O(n^{1/3})$ and the bound in (b) is trivially true and not useful. Hence, let's consider the case that $w \ge n^{2/3}$. Moreover, based on the assumption $\theta = o(1)$, we can pick n large enough such that $\theta < 1$. We will show that expected trend in rate of change of $Y_l^n(t)$ is followed almost surely by demonstrating concentration of

$$Y_l^n(t + w) - Y_l^n(t)$$

Let assume for some large positive constant C, $(t/n, Y_1^n(t)/n, Y_2^n(t)/n, \dots)$ in l^∞-distance at least $C\theta$ from the boundary of D almost surely where choice of C is based on Lipschitz constants in the Lipschitz hypothesis. This will guarantee

that in the following arguments, we are always inside D and Lipschitz hypothesis is valid. For $0 \leq k \leq w$, note that $k\beta/n = O(\theta)$. By trend hypothesis we have

$$\mathbb{E}(Y_l^n(t+k+1)-Y_l^n(t+k) \mid H_{t+k}) = f_l((t+k)/n, Y_1^n(t+k)/n, \dots)+O(\theta_1+C_0\gamma n) \tag{35}$$

Note that, if the boundedness hypothesis fails(and hence, the trend hypothesis will fail), then $|Y_l^n(t+k+1)-Y_l^n(t+k)| < 2C_0 n$ with probability γ. In this case, the expected change in equation (35) is bounded by $O(C_0\gamma n)$ which denote the added term to equation (35). Base on Lipschitz hypothesis, condition on the boundedness hypothesis (which holds with probability $1-\gamma$),

$$
\begin{aligned}
f_l((t+k)/n, Y_1^n(t+k)/n, \dots) &\leq f_l((t+k-1)/n, Y_1^n(t+k-1)/n, \dots) + \\
&\quad L \max_{1 \leq l \leq a(n)} |Y_l^n(t+k)/n - Y_l^n(t+k-1)/n| \\
&\leq f_l((t+k-1)/n, Y_1^n(t+k-1)/n, \dots) + L\beta/n
\end{aligned}
$$

Conditioning on boundedness hypothesis for each step,

$$
\begin{aligned}
&f_l((t+k)/n, Y_1^n(t+k)/n, \dots) + O(\theta_1 C_0\gamma n) = \\
&\qquad f_l(t/n, Y_1^n(t)/n, Y_2^n(t)/n, \dots) + O(\theta_1 + Lk\beta/n) \quad \forall l \leq a(n)
\end{aligned}
$$

Hence, there is a function $g(n) = O(\theta_1 + Lw\beta/n) = O(\theta)$ such that conditioned on H_t and boundedness hypothesis,

$$Y_l^n(t+k) - Y_l^n(t) - kf_l(t/n, Y_1^n(t)/n, Y_2^n(t)/n, \dots) - kg(n)$$

$(k = 0, 1, \dots, w)$ is a supermartingle in k with respect to the sequence of σ-fields generated by $H_t, \dots, H_{t+w}$ for all $l \leq a(n)$. The differences in this supermartingale is bounded by:

$$\beta + |f_l(t/n, Y_1^n(t)/n, Y_2^n(t)/n, \dots)| + |g(n)| \leq \beta + O(1) + O(\theta) \leq \kappa\beta,$$

for some fix constant $\kappa > 0$ that is independent of l, since based on Lipschitz hypothesis,

$$
\begin{aligned}
f_l(t/n, Y_1^n(t)/n, Y_2^n(t)/n, \dots) &\leq \tag{36} \\
\max_{1 \leq l \leq a(n)} |f_l(0, Y_1^n(0)/n, Y_2^n(0)/n, \dots)| &+ L \max(\beta/n, t/n) = O(1)
\end{aligned}
$$

Note that D is bounded and Lipschitz hypothesis applies since variables do not leave D. Now based on Lemma E.1), since the differences in this suppermartingale is bounded by $\kappa\beta$ we have:

$$\mathbb{P}\left(Y_l^n(t+w) - Y_l^n(t) - wf_l(t/n, Y_1^n(t)/n, Y_2^n(t)/n, \dots) - wg(n) \geq \kappa\beta\sqrt{2w\alpha} \mid H_t\right)$$

$$\leq \exp\left(-\frac{2w\alpha\kappa^2\beta^2}{2w\kappa^2\beta^2}\right) = e^{-\alpha}$$

for all α. Let fix α as follows:

$$\alpha = \frac{n\theta^3}{\beta^3}$$

Note that lower tail of the $Y_l^n(t+w) - Y_l^n(t) - wf_l(t/n, Y_1^n(t)/n, Y_2^n(t)/n, \dots)$ can be bounded by exactly the same method using submartingale. Hence (after adjusting κ) we have

$$\mathbb{P}\left(|Y_l^n(t+w) - Y_l^n(t) - wf_l(t/n, Y_1^n(t)/n, Y_2^n(t)/n, \dots)| \geq wg(n) + \kappa\beta\sqrt{w\alpha} \mid H_t\right) \leq 2e^{-\alpha} \tag{37}$$

Now, let's define $k_i = iw$, $i = 0, 1, 2, \dots, i_0$ where $i_0 = \lfloor \sigma n/w \rfloor$. We will show that

$$\mathbb{P}\left(|Y_l^n(k_j) - z_l(k_j/n)n| \geq B_j \text{ for some } j \leq i \text{ and } l \leq a(n)\right) = O(2^j a(n) i e^{-\alpha}) \tag{38}$$

where

$$B_j = B_0\left(1 + \frac{Bw}{n}\right)^j + Bw\left(\theta + \frac{w}{n}\right)\left(\left(1 + \frac{Bw}{n}\right)^j - 1\right)\frac{n}{Bw}$$

for some constant B and B_0. B_0 is an arbitrary fix number. The value of B will be determined. The basis of the induction is proven as follows,

$$\lim_{n\to\infty} \max_{1\leq l\leq a(n)} |Y_l^n(0)/n - \hat{z}_l| = 0$$

then there exits N such that,

$$\mathbb{P}\left(|Y_l^n(0) - z_l(0)n| \geq B_0 \text{ for all } l\right) = 0 \quad \forall n > N$$

Let assume the statement is correct upto i. Note that,

$$|Y_l^n(k_{i+1}) - z_l(k_{i+1}/n)n| = |A_1 + A_2 + A_3 + A_4|$$

where

$$
\begin{aligned}
A_1 &= Y_l^n(k_i) - z_l(k_i/n)n \\
A_2 &= Y_l^n(k_{i+1}) - Y_l^n(k_i) - wf_l(k_i/n, Y_1^n(k_i)/n, Y_2^n(k_i)/n, \dots) \\
A_3 &= wz_l'(k_i/n) + z_l(k_i/n)n - z_l(k_{i+1}/n)n \\
A_4 &= wf_l(k_i/n, Y_1^n(k_i)/n, Y_2^n(k_i)/n, \dots) - wz_l'(k_i/n)
\end{aligned}
$$

Note that in the all following parts, we will take the advantage of following lemma:

Lemma E.3. *Let assume* $A_1, A_2, A_3, \dots$ *is a sequence of events. Then we have:*

$$P\left(\bigcap_i A_i\right) = 1 - P\left(\bigcup_i A_i^c\right) \geq 1 - \sum_i P(A_i^c)$$

By inductive hypothesis
$$|A_1| < B_i,$$

together with similar inequalities for $j < i$ and for all l, with probability $1 - O(2^j a(n)ie^{-\alpha})$. From inequality (37) and recalling that $w = n\theta/\beta =$, $g(n) = O(\theta)$ and $\alpha = n\theta^3/\beta^3$, condition on boundedness hypothesis,

$$|A_2| < wg(n) + \kappa\beta\sqrt{2w\alpha} < B'w\theta,$$

for some universal constant B' for all l with probability $1 - O(2^{j+1}a(n)e^{-\alpha})$ which is driven by Lemma E.3. Since z_l is the solution given in (a) and because of Lipschitz hypothesis over f which guarantee $f \in C^1$,

$$
\begin{aligned}
|A_3|/n &= |z_l(k_i/n) - z_l(k_{i+1}/n) - w/nz_l'(k_i/n)| \\
&= w^2/nz_l''(\varsigma) \\
&< B''w^2/n
\end{aligned}
$$

for a suitable B'' and $l \leq a(n)$. Finally, by Lipshitz hypothesis and noting that z_l is the solution of (a),

$$
\begin{aligned}
A_4 &= wf_l(k_i/n, Y_1^n(k_i)/n, Y_2^n(k_i)/n, \dots) - wf_l(k_i/n, z_1(k_i), z_2(k_i), \dots) \\
&\leq Lw \times \max\left(\max_{1\leq l\leq a(n)} |Y_l^n(k_i)/n - z_l(k_i)|, \sup_{l>a(n)} z_l(k_i)\right) \\
&\leq Lw \times \max(B_i, C_0) \\
&\leq \frac{B'''wB_i}{n}
\end{aligned}
$$

with probability $1 - O(2^j a(n)ie^{-\alpha})$. Now, let set $B = \max\{B', B'', B'''\}$. Summing the bounds gives,

$$B_{i+1} = B_i + B\frac{w^2}{n} + B\frac{wB_i}{n} + Bw\theta$$

hence,

$$
\begin{aligned}
B_{i+1} &= B_i\left(1 + \frac{Bw}{n}\right) + B\frac{w^2}{n} + Bw\theta \\
&= B_0\left(1 + \frac{Bw}{n}\right)^{i+1} + Bw\left(\theta + \frac{w}{n}\right)\left(\left(1 + \frac{Bw}{n}\right)^{i+1} - 1\right)\frac{n}{Bw}
\end{aligned}
$$

Condition on boundedness hypothesis and based on Lemma E.3, with probability greater than

$$1 - O(2^{j+1}a(n)e^{-\alpha}) - O(2^j a(n)ie^{-\alpha}) - O(2^j a(n)ie^{-\alpha}) = 1 - O(2^{j+1}a(n)(i+1)e^{-\alpha})$$

we have,

$$|Y_l^n(k_{i+1}) - z_l(k_{i+1}/n)n| < B_{i+1}$$

Note that Bw/n is $o(1)$ since β is bounded below and θ is $o(1)$. Hence, $B_0(1 + Bw/n)^i$ is $O(1)$. By same argument, we have:

$$B_i = O(n\theta + w) = O(n\theta)$$

38

since β is bounded below. Also for any $t \leq \sigma n$, put $i = \lfloor t/w \rfloor$, then from time k_i to t the change in Y_l for each l is at most $w\beta = O(n\theta)$ condition on boundedness hypothesis and the change in z_l is bounded by $O(w\beta) = O(n\theta)$ using first order approximation over z_l by noting that z_l is the solution of (a) and boundedness of f_l which we discussed in equation (36). Hence

$$|Y_l^n(t) - z_l(t/n)n| = O(n\theta)$$

with probability greater than $1 - O(a(n)i_0 \exp\left(-\frac{n\theta^3}{\beta^3}\right))$ conditioned on boundedness hypothesis. By removing this condition and replace it with probability that the boundedness hypothesis holds at every single step which is equal to $1 - O(b(n)\gamma\sigma(n)n)$ and noting that $i_0 = O(n/w) = O(\theta/\beta)$,

$$|Y_l^n(t) - z_l(t/n)n| = O(n\theta)$$

with probability greater than $1 - O(b(n)\gamma n + a(n)\frac{\theta}{\beta} \exp\left(-\frac{n\theta^3}{\beta^3}\right))$ $\qquad\square$

F   Proofs of Theorems and Lemmas

F.1   Proof of Lemma 5.1

Proof. The proof is straight forward. Note that given balance equations, the random variables are almost surely bounded by $2m_j(n)$ which is bounded as follow:

$$2m_j(n) = \sum_i d_{i,j}^{(n)} \leq O(n)$$

Choosing big enough constant K, the result follows. $\qquad\square$

F.2   Proof of Lemma D.1

Proof. Note that boundedness is a direct consequence of modified regularity conditions we imposed over the elements of $D_\epsilon(n)$. For showing Lipschits property, pick some function f_l from equations (28)-(31). Note that f_l is $\mathcal{C}^1$ and in D_ϵ. Now using mean value theorem, we have:

$$\mid f_l(\mathbf{a}) - f_l(\mathbf{b}) \mid \leq \mid \mathbf{a} - \mathbf{b} \mid_\infty \sup_{x \in L(\mathbf{a},\mathbf{b})} \|f_l'(x)\|_2,$$

for any arbitrary label l such that $L(\mathbf{a}, \mathbf{b})$ is the line connecting $\mathbf{a}$ and $\mathbf{b}$. Using this property, and nothing that $\|f_l'(\cdot)\|_2$ is a continuous function, we have:

$$\mid f_l(\mathbf{a}) - f_l(\mathbf{b}) \mid_\infty \leq \mid \mathbf{a} - \mathbf{b} \mid_\infty \sup_{x \in D^\epsilon} \|f_l'(x)\|_2,$$

it might be unclear that why $\sup_{x \in D^\epsilon} \|f_l'(x)\|_2 < \infty$. We highlight the modified regularity condition to see why this is correct. Note that the only probable

unbounded term in $f_l(\cdot)$ which appear in $\|f_l'(\cdot)\|_2$ as well is the followings:

$$\sum_{u_j + u_{-j} + 1 = K_j(d_j, d_{-j})} (d_j - u_j - 1) \times (d_j - u_j) i^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(t)$$

which is an $O(1)$ term. Finally, note that these bound work for all functions if it works for first 6 equations. Hence, one can pick a fix Lipschitz constant $L(\epsilon)$ that works for all functions. $\qquad\square$

F.3 Proof of Lemma 6.1

Proof. Note that the regularity condition assure that number of edges is roughly $O(n)$, hence, the algorithm should terminate in $O(n)$ steps which shows that $\sigma(n) = O(1)$. Moreover, Lipschitz hypothesis is proved by Lemma D.1. Trend hypothesis is almost trivial. Finally, note that boundedness hypothesis is a direct consequence of boundedness condition we impose over degree distribution in Lemma 5.1. $\qquad\square$

F.4 Proof of Lemma 7.1

Proof. Note that initial conditions (32) match the initial condition given by equation (18) and the form of solution (12)-15. Before plug the solutions in differential equations, we discuss some properties of four dimensional differential equation. Note that equation (13) implies,

$$\lambda_j - 2\tau_j(t) = \lambda_j \mu^{(j,j)}(t)^2 \tag{39}$$

hence,

$$\frac{d\mu^{(j,j)}}{dt}\left(\mu^{(j,j)}(t)\right)^{-1} = \lambda_j \frac{d\mu^{(j,j)}}{dt}\left(\mu^{(j,j)}(t)\right) \times \left(\lambda_j \mu^{(j,j)}(t)^2\right)^{-1} \tag{40}$$

$$= \frac{-a_j(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)}(\lambda_j - 2\tau_j(t))^{-1}$$

Another useful relation that we will use throughout the proof is the following. Based on equation (17),

$$\lambda_m \frac{d\mu^{(1,2)}}{dt}\left(\mu^{(2,1)}(t)\right) + \lambda_m \frac{d\mu^{(2,1)}}{dt}\left(\mu^{(1,2)}(t)\right) = \frac{-a_m^{(1)}(t) - a_m^{(2)}(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)} \tag{41}$$

$$= -1 + \frac{a_1(t) + a_2(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)}$$

$$= -1 + \frac{d\tau_1}{dt} + \frac{d\tau_2}{dt}$$

Now, integrating both sides of the equation (39),

$$\lambda_m\left(\mu^{(2,1)}(t)\right)\left(\mu^{(1,2)}(t)\right) = \lambda_m - (t - \tau_1 - \tau_2) \tag{42}$$

Note that, the constant of integration is determined by equation (18). Using equation (42), we get

$$\frac{d\mu^{(j,-j)}}{dt}\left(\mu^{(j,-j)}(t)\right)^{-1} = \lambda_m \frac{d\mu^{(j,-j)}}{dt}\mu^{(-j,j)}(t)\left(\lambda_m\mu^{(-j,j)}(t)\mu^{(j,-j)}(t)\right)^{-1}$$

(43)

$$= \frac{-a_m^{(-j)}(t)}{a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)}\left(\lambda_m - (t - \tau_1 - \tau_2)\right)^{-1}$$

Using these properties, we can check the validity of the solutions by plugging the form of solution and calculating the derivatives. First, we will check the form of $i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}$ by taking the derivative of equation (12),

$$\frac{di^{(j)}_{d_j,d_{-j},u_j,u_{-j}}}{dt} = \mathbb{P}_j(d_j)\mathbb{P}_m(d_{-j})(1 - \alpha_j(d_j,d_{-j}))\times$$

$$\left\{(d_j - u_j)\frac{d\mu^{(j,j)}}{dt}\binom{d_j}{u_j}\left(1 - \mu^{(j,j)}(t)\right)^{u_j}\times\right.$$

$$\left(\mu^{(j,j)}(t)\right)^{d_j-u_j-1}Bi\left(u_{-j};d_{-j},1 - \mu^{(j,-j)}(t)\right)$$

$$+ (d_{-j} - u_{-j})\frac{d\mu^{(j,-j)}}{dt}\binom{d_{-j}}{u_{-j}}\left(1 - \mu^{(j,-j)}(t)\right)^{u_{-j}}\times$$

$$\left(\mu^{(j,-j)}(t)\right)^{d_{-j}-u_{-j}-1}Bi\left(u_j;d_j,1 - \mu^{(j,j)}(t)\right)$$

$$- u_j\frac{d\mu^{(j,j)}}{dt}\binom{d_j}{u_j}\left(1 - \mu^{(j,j)}(t)\right)^{u_j-1}\times$$

$$\left(\mu^{(j,j)}(t)\right)^{d_j-u_j}Bi\left(u_{-j};d_{-j},1 - \mu^{(j,-j)}(t)\right)$$

$$+ u_{-j}\frac{d\mu^{(j,-j)}}{dt}\binom{d_{-j}}{u_{-j}}\left(1 - \mu^{(j,-j)}(t)\right)^{u_{-j}-1}\times$$

$$\left.\left(\mu^{(j,-j)}(t)\right)^{d_{-j}-u_{-j}}Bi\left(u_j;d_j,1 - \mu^{(j,j)}(t)\right)\right\}$$

substituting each term with corresponding function from solution set yields,

$$\frac{di^{(j)}_{d_j,d_{-j},u_j,u_{-j}}}{dt} = (d_j - u_j)\times\frac{d\mu^{(j,j)}}{dt}\left(\mu^{(j,j)}(t)\right)^{-1}\times i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t) \quad (44)$$

$$+ (d_{-j} - u_{-j})\times\frac{d\mu^{(j,-j)}}{dt}\left(\mu^{(j,-j)}(t)\right)^{-1}\times i^{(j)}_{d_j,d_{-j},u_j,u_{-j}}(t)$$

$$- (d_j - u_j + 1)\times\frac{d\mu^{(j,j)}}{dt}\left(\mu^{(j,j)}(t)\right)^{-1}\times i^{(j)}_{d_j,d_{-j},u_j-1,u_{-j}}(t)$$

$$- (d_{-j} - u_{-j} + 1)\times\frac{d\mu^{(j,-j)}}{dt}\left(\mu^{(j,-j)}(t)\right)^{-1}\times i^{(j)}_{d_j,d_{-j},u_j,u_{-j}-1}(t)$$

Now equation (31) will be obtained by substituting proper values from equation (43) and (41) to equation (44) and equation (30) can be verified by plugging equation (16) to following equation,

$$\frac{d\tau_j}{dt} = -\lambda_j \mu^{(j,j)} \frac{d\mu^{(j,j)}}{dt}$$

which is obtained by taking the derivative of both sides of equation (13). Next, we will verify the solution of $a_j(t)$. Note that, as we expected, the form of solution given in Lemma 7.1 satisfies the balance equations.

$$\frac{da_j}{dt} = 2\lambda_j \frac{d\mu^{(j,j)}}{dt} \mu^{(j,j)}(t) - \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} (d_j - u_j) \frac{di^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}}{dt} \tag{45}$$

$$= 2\lambda_j \frac{d\mu^{(j,j)}}{dt} \left(\mu^{(j,j)}(t) \right) \tag{46}$$

$$- \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} (d_j - u_j) \times (d_j - u_j) \times \frac{d\mu^{(j,j)}}{dt} \left(\mu^{(j,j)}(t) \right)^{-1} \times i^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(t)$$

$$- \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} (d_j - u_j) \times (d_{-j} - u_{-j}) \times \frac{d\mu^{(j,-j)}}{dt} \left(\mu^{(j,-j)}(t) \right)^{-1} \times i^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(t)$$

$$+ \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} (d_j - u_j) \times (d_j - u_j + 1) \times \frac{d\mu^{(j,j)}}{dt} \left(\mu^{(j,j)}(t) \right)^{-1} \times i^{(j)}_{d_j, d_{-j}, u_j - 1, u_{-j}}(t)$$

$$+ \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} (d_j - u_j) \times (d_{-j} - u_{-j} + 1) \times \frac{d\mu^{(j,-j)}}{dt} \left(\mu^{(j,-j)}(t) \right)^{-1} \times i^{(j)}_{d_j, d_{-j}, u_j, u_{-j} - 1}(t)$$

Note that equation (46) is obtained by plugging equation (44) to equation (45). Simplifying it further, we have,

$$\frac{da_j}{dt} = \lambda_j \frac{d\mu^{(j,j)}}{dt} \mu^{(j,j)}(t) \tag{47}$$

$$\lambda_j \frac{d\mu^{(j,j)}}{dt} \mu^{(j,j)}(t) \times \left(1 - \left(\lambda_j \mu^{(j,j)}(t)^2 \right)^{-1} \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} (d_j - u_j) i^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(t) \right)$$

$$- \frac{d\mu^{(j,j)}}{dt} \left(\mu^{(j,j)}(t) \right)^{-1} \sum_{u_j + u_{-j} + 1 = K_j(d_j, d_{-j})} (d_j - u_j - 1) \times (d_j - u_j) i^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(t)$$

$$- \frac{d\mu^{(j,-j)}}{dt} \left(\mu^{(j,-j)}(t) \right)^{-1} \sum_{u_j + u_{-j} + 1 = K_j(d_j, d_{-j})} (d_j - u_j) \times (d_{-j} - u_{-j}) i^{(j)}_{d_j, d_{-j}, u_j, u_{-j}}(t)$$

Substituting proper values from equations (39), (40) and (43) to equation (47) will justify the solution of differential equation (28). Finally, the procedure to

verify equation (15) satisfies equation (29) is just as same as previous cases.

$$\frac{da_m^{(j)}}{dt} = -1 + \frac{d\tau_1}{dt} + \frac{d\tau_2}{dt} - \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} (d_{-j} - u_{-j}) \frac{di_{d_j, d_{-j}, u_j, u_{-j}}^{(j)}}{dt} \tag{48}$$

$$= -1 + \frac{d\tau_1}{dt} + \frac{d\tau_2}{dt} \tag{49}$$

$$- \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} (d_{-j} - u_{-j}) \times (d_j - u_j) \times \frac{d\mu^{(j,j)}}{dt} \left(\mu^{(j,j)}(t)\right)^{-1} \times i_{d_j, d_{-j}, u_j, u_{-j}}^{(j)}(t)$$

$$- \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} (d_{-j} - u_{-j}) \times (d_{-j} - u_{-j}) \times \frac{d\mu^{(j,-j)}}{dt} \left(\mu^{(j,-j)}(t)\right)^{-1} \times i_{d_j, d_{-j}, u_j, u_{-j}}^{(j)}(t)$$

$$+ \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} (d_{-j} - u_{-j}) \times (d_j - u_j + 1) \times \frac{d\mu^{(j,j)}}{dt} \left(\mu^{(j,j)}(t)\right)^{-1} \times i_{d_j, d_{-j}, u_j - 1, u_{-j}}^{(j)}(t)$$

$$+ \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} (d_{-j} - u_{-j}) \times (d_{-j} - u_{-j} + 1) \times \frac{d\mu^{(j,-j)}}{dt} \left(\mu^{(j,-j)}(t)\right)^{-1} \times i_{d_j, d_{-j}, u_j, u_{-j} - 1}^{(j)}(t)$$

After some algebra, the equation (49) can be written as follows. Note that equation (41) is used to simplify the answer,

$$\frac{da_m^{(j)}}{dt} = \lambda_m \frac{d\mu^{(-j,j)}}{dt} \left(\mu^{(j,-j)}(t)\right) \tag{50}$$

$$+ \lambda_m \frac{d\mu^{(j,-j)}}{dt} \mu^{(-j,j)}(t) \times$$

$$\left(1 - \left(\lambda_m \mu^{(j,-j)}(t)\mu^{(-j,j)}(t)\right)^{-1} \sum_{u_j + u_{-j} < K_j(d_j, d_{-j})} (d_{-j} - u_{-j}) i_{d_j, d_{-j}, u_j, u_{-j}}^{(j)}(t)\right)$$

$$- \frac{d\mu^{(j,-j)}}{dt} \left(\mu^{(j,-j)}(t)\right)^{-1} \sum_{u_j + u_{-j} + 1 = K_j(d_j, d_{-j})} (d_{-j} - u_{-j} - 1) \times (d_{-j} - u_{-j}) i_{d_j, d_{-j}, u_j, u_{-j}}^{(j)}(t)$$

$$- \frac{d\mu^{(j,j)}}{dt} \left(\mu^{(j,j)}(t)\right)^{-1} \sum_{u_j + u_{-j} + 1 = K_j(d_j, d_{-j})} (d_{-j} - u_{-j}) \times (d_j - u_j) i_{d_j, d_{-j}, u_j, u_{-j}}^{(j)}(t)$$

Using equations (39), (40) and (43), we get the equation (29). $\qquad\square$

F.5 Proof of Lemma 7.3

Proof. Let's define: $Y(x) = Bi(u, d, 1 - x)$. Then we have:

$$\frac{dY}{dx} = \binom{d}{u} \left((d - u) \times x^{d-u-1}(1 - x)^u - u \times x^{d-u}(1 - x)^{u-1}\right)$$

$$= d \times (Bi(u, d - 1, 1 - x) - Bi(u - 1, d - 1, 1 - x))$$

43

Without loss of generality, we will show that first component of $F(\cdot)$ is an increasing function of it's arguments.

$$
\begin{aligned}
\frac{\partial F_{(1,1)}}{\partial \mu^{(1,1)}} &= \sum_{u1+u2<K_1(d1,d2)} \mathbb{P}_1(d1)\mathbb{P}_m(d2)\frac{d1}{\lambda_1}(1-\alpha^{(1)}_{d1,d2})Bi(u2;d2,1-\mu^{(1,2)}) \times (d1-1) \\
&\qquad\qquad \left(Bi(u1;d1-2,1-\mu^{(1,1)}) - Bi(u1-1;d1-2,1-\mu^{(1,1)}) \right) \\
&= \sum_{u2<K_1(d1,d2)} \mathbb{P}_1(d1)\mathbb{P}_m(d2)\frac{d1}{\lambda_1}(1-\alpha^{(1)}_{d1,d2})(d1-1) \times \\
&\qquad\qquad Bi(u2;d2,1-\mu^{(1,2)})Bi(K_1(d1,d2)-u2-1;d1-2,1-\mu^{(1,1)})
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial F_{(1,1)}}{\partial \mu^{(1,2)}} &= \sum_{u1+u2<K_1(d1,d2)} \mathbb{P}_1(d1)\mathbb{P}_m(d2)\frac{d1}{\lambda_1}(1-\alpha^{(1)}_{d1,d2})Bi(u1;d1-1,1-\mu^{(1,1)}) \times d2 \\
&\qquad\qquad \left(Bi(u2;d2-1,1-\mu^{(1,2)}) - Bi(u2-1;d2-1,1-\mu^{(1,2)}) \right) \\
&= \sum_{u1<K_1(d1,d2)} \mathbb{P}_1(d1)\mathbb{P}_m(d2)\frac{d1}{\lambda_1}(1-\alpha^{(1)}_{d1,d2}) \times \\
&\qquad\qquad d2 \times Bi(u1;d1-1,1-\mu^{(1,1)})Bi(K(d1,d2)-u1-1;d2-1,1-\mu^{(1,2)})
\end{aligned}
$$

Hence, each component of F is an increasing function of its arguments . $\qquad\square$

F.6   Proof of Lemma 7.4

Proof. *i.* Fix $\boldsymbol{\mu} \in U$. Consider the following connected set:

$$
\mathcal{S} = \{\boldsymbol{x} : F(\boldsymbol{\mu}) \le \boldsymbol{x} \le \boldsymbol{\mu}\}
$$

Note that the line connecting $\boldsymbol{\mu}$ to $F(\boldsymbol{\mu})$, $\mathcal{L} = \{\alpha\boldsymbol{\mu} + (1-\alpha)F(\boldsymbol{\mu}) : \alpha \in [0,1]\}$, is a subset of $\mathcal{S}$. Finally, it is easy to see that $\mathcal{S} \subset U$:

$$
\forall \boldsymbol{s} \in \mathcal{S}, \ \boldsymbol{s} \le \mu \longrightarrow F(\boldsymbol{s}) \le F(\boldsymbol{\mu}) \quad \text{by Lemma 7.3}
$$

ii. It is sufficient to show that U is closed. Consider an arbitrary sequence of elements of U, i.e. $\{\boldsymbol{u}_i\}_{i=1}^{\infty}$ that converges to $\boldsymbol{u}^*$. We will show that $\boldsymbol{u}^* \in U$. Let define $G(\boldsymbol{x}) = \boldsymbol{x} - F(\boldsymbol{x})$. If $G(\boldsymbol{u}^*) \ge 0$ then we are done. Let assume $G(\boldsymbol{u}^*) = -\epsilon < 0$, since $G(\cdot)$ is continuous, there is a δ such that

$$
\forall \boldsymbol{x}, \quad \|\boldsymbol{u}^* - \boldsymbol{x}\|_2 < \delta \Longrightarrow \|G(\boldsymbol{u}^*) - G(\boldsymbol{x})\|_2 < \frac{\epsilon}{2}
$$

Moreover, since $\lim_{n\to\infty} \boldsymbol{u}_n = \boldsymbol{u}^*$, there is an N such that:

$$
\forall k > N, \quad \|\boldsymbol{u}^* - \boldsymbol{u}_k\|_2 < \delta
$$

which contradict with the assumption that $\boldsymbol{u}_i \in U$

iii. Let consider the sequence $\left\{F^k(\boldsymbol{u})\right\}_{k=1}^\infty$, where $\boldsymbol{u} \in U$ is chosen arbitrary. Since U is compact, there is a subsequence $\boldsymbol{u}_i$ of $\left\{F^k(\boldsymbol{u})\right\}_{k=1}^\infty$ that converges to some point $\boldsymbol{u}^* \in U$. Now we will show that the whole sequence converge to $\boldsymbol{u}^*$:

$$u_i \to u^* \implies \forall \epsilon, \exists N \ s.t. \ \forall k > N : \|u_k - u^*\|_2 < \epsilon \tag{51}$$

$$\left\{F^k(u)\right\}_{k=1}^\infty \ is \ increasing \implies \forall i, \exists j : u_j \leq F^i(u) \leq u_{j+1} \tag{52}$$

hence, we have $\lim_{k\to\infty} F^k(\boldsymbol{u}) = \boldsymbol{u}^*$. Now we will show that $\boldsymbol{u}^*$ is a fixed point of $F(\cdot)$. Let assume the converse is true, i.e. $\|F(\boldsymbol{u}^*) - \boldsymbol{u}^*\|_2 = \epsilon$.

$$F(\cdot) \ is \ continuous \implies \exists \delta : \|\boldsymbol{u} - \boldsymbol{u}^*\|_2 < \delta : \|F(\boldsymbol{u}) - F(\boldsymbol{u}^*)\|_2 < \frac{\epsilon}{3} \tag{53}$$

$$\lim_{k\to\infty} F^k(\boldsymbol{u}) = \boldsymbol{u}^* \implies \exists N, \forall k > N : \left\|F^k(\boldsymbol{u}) - \boldsymbol{u}^*\right\|_2 < \min(\delta, \frac{\epsilon}{3}) \tag{54}$$

Let fix $k > N$. Given (53)) and (54)), we have $\left\|F^k(\boldsymbol{u}) - F(\boldsymbol{u}^*)\right\|_2 < \epsilon/3$ and $\left\|F^k(\boldsymbol{u}) - \boldsymbol{u}^*\right\|_2 < \epsilon/3$. Using triangular inequality, $\|\boldsymbol{u}^* - F(\boldsymbol{u}^*)\|_2 < 2\epsilon/3$ which is a contradiction.

iiii. There are four possible cases which first two are as follow:

1. $\boldsymbol{\mu}'_1 = \boldsymbol{\mu}_1$: Note that in this case if $\boldsymbol{\mu}'_2 > \boldsymbol{\mu}_2$ then $F_1(\boldsymbol{\mu}') > \boldsymbol{\mu}'_1$ which means $\boldsymbol{\mu}' \notin U$. Let consider $\boldsymbol{\mu}'_2 = \boldsymbol{\mu}_2$. Then either $\boldsymbol{\mu}'_3 > \boldsymbol{\mu}_3$ or $\boldsymbol{\mu}'_4 > \boldsymbol{\mu}_4$ we have $F_2(\boldsymbol{\mu}') > \boldsymbol{\mu}'_2$.

2. $\boldsymbol{\mu}'_2 = \boldsymbol{\mu}_2$: Note that in this case if $\boldsymbol{\mu}'_3 > \boldsymbol{\mu}_3$ or $\boldsymbol{\mu}'_4 > \boldsymbol{\mu}_4$ then by same argument as above, $\boldsymbol{\mu}' \notin U$. Assuming equality for $\boldsymbol{\mu}'_3$ and $\boldsymbol{\mu}'_4$, same should holds for $\boldsymbol{\mu}'_1$.

The proof of other cases, i.e. $\boldsymbol{\mu}'_3 = \boldsymbol{\mu}_3$ and $\boldsymbol{\mu}'_4 = \boldsymbol{\mu}_4$, are just the same. $\qquad \square$

F.7 Proof of Theorem 7.5

Proof. We will prove this theorem by the following steps:

Step 1. By Lemma 7.4, we know that $F^\infty(\boldsymbol{1}) = \lim_{n\to\infty} F^n(\boldsymbol{1})$ exist and belongs to U; moreover, it is a fixed point of given differential equation.

Step 2. Let set $\mathcal{N} = U \bigcap [F^\infty(\boldsymbol{1}), \boldsymbol{1}]$. Note that for any arbitrary $u \in U$, $F(u) - u$ is directing toward the set $\mathcal{S}_u = \{x : F(u) \leq x \leq u\}$, i.e.

$$\forall \delta < 1 : u + \delta(F(u) - u) \in \mathcal{S}_u$$

Note that, $\mathcal{S}_u$ do not contain any point on the border of $[F^\infty(\boldsymbol{1}), \boldsymbol{1}]$ other than $F^\infty(\boldsymbol{1})$ over the planes contain $F^\infty(\boldsymbol{1})$, base on Lemma 7.4 part iv. Moreover, base on the proof of part i of same lemma, $\mathcal{S}_u$ belongs to U. Hence, $\mathcal{S}_u \in \mathcal{N}$ and $\mathcal{N}$ is a positive invariant set. Furthermore, it is close and compact (Lemma 7.4

part ii) and has only one fixed point which is $F^\infty(\mathbf{1})$ (Lemma 7.4 part iv and definition of $\mathcal{N}$). Let set $V(\mu) = \|\mu - F^\infty(\mathbf{1})\|_2^2$:

$$
\begin{aligned}
\frac{\dot{V}(\mu)}{2} &= \frac{\nabla V(\mu)^T \dot{\mu}}{2} \\
&= \left(\mu - F^\infty(\mathbf{1})\right)^T (F(\mu) - \mu) \\
&= \left(\mu - F(\mu) + F(\mu) - F^\infty(\mathbf{1})\right)^T (F(\mu) - \mu) \\
&= -\|\mu - F(\mu)\|_2^2 + \left(F(\mu) - F^\infty(\mathbf{1})\right)^T (F(\mu) - \mu)
\end{aligned}
$$

Note that for any point $\upsilon \in \mathcal{N}$, base on definition of $\mathcal{N}$, we have $\upsilon \geq F^\infty(\mathbf{1})$. Moreover, since $F(\upsilon) \in \mathcal{N}$, we have $F(\upsilon) \geq F^\infty(\mathbf{1})$. Hence,

$$
\frac{\dot{V}(\mu)}{2} = -\|\mu - F(\mu)\|_2^2 - \left(F(\mu) - F^\infty(\mathbf{1})\right)^T (\mu - F(\mu)) \leq 0
$$

which is equal to zero iff μ is a fixed point of F. Hence, using LaSalle Invariance Principle, noting that E consist of only one element which is $F^\infty(\mathbf{1})$ complete the proof. Specifically, all trajectories started in $\mathcal{N}$ will end up at $F^\infty(\mathbf{1})$. Note that Lemma 7.4 part iv and the fact that $\mathcal{N}$ has only one fixed point guarantee that $F^\infty(\mathbf{1})$ is the closest fixed point to $\mu(0) = \mathbf{1}$.

□

F.8   Proof of Theorem 7.6

Proof. First off, note that as long as the trajectory of ODEs (16)-(17) lie inside $M(\epsilon)$ removing the term $a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t)$ from denominator, will note change the trajectory (since we are multiplying all equations with same positive function, hence, the field generated by ODEs are the same); hence, the trajectory of ODEs given by (16)-(17) and the one given by (24) are the same inside the set $M(\epsilon)$ while their speed might be different. Moreover, note that if $\epsilon < \epsilon'$, then $M(\epsilon') \subset M(\epsilon)$. Let consider the sequence $(\epsilon_i)_{i=0}^\infty$ such that $\epsilon_m = 1/m$. Let call the first point that the trajectory of ODE (24) hit the boundary of $M(\epsilon_m)$ to be $\boldsymbol{p}_m$. First off note that for all $n < m$, $\boldsymbol{p}_n \in M(\epsilon_m)$. Moreover, since the trajectory of ODE (24) is non-increasing, we also have $\boldsymbol{p}_n \leq \boldsymbol{p}_m$. Hence, the sequence $(\boldsymbol{p}_i)_{i=1}^\infty$ is a non-decreasing sequence in a compact set, hence it will converge to some point $\boldsymbol{p}^*$. Let define the set U as in Lemma 7.4, to be the largest connected set containing $\mathbf{1}$ such that $\forall \boldsymbol{\mu} \in U, \boldsymbol{\mu} \geq F(\boldsymbol{\mu})$. Let define $\mathcal{V} = U \bigcap [0, F^\infty(\mathbf{1})]^4$. Note that for any arbitrary ϵ, we have $M(\epsilon) \subset \mathcal{V}$. Now, we will show p^* is $F^\infty(\mathbf{1})$ by following steps:

  i. Note that $\mathcal{V}$ is close and compact. Moreover, the sequence $(\boldsymbol{p}_i)_{i=1}^\infty$ lie inside the set $\mathcal{V}$; hence, $\boldsymbol{p}^* \in \mathcal{V}$.

ii. Note that for every ϵ_i, one of conditions over the ODEs should be violated at the point $\boldsymbol{p}_i$. Hence, there is two possible choices for point $\boldsymbol{p}^*$:

    a) $\boldsymbol{p}^*$ is the point such that $\boldsymbol{p}^*_m = 0$ for some m(one of its components is equal to zero), but $\boldsymbol{p}^*$ is not equal to $F^\infty(\boldsymbol{1})$. Note that $\mathcal{V}$ is an invariant set and the trajectory of ODE (24), i.e. $\boldsymbol{\mu}(t)$, is non-increasing and passes through the point $\boldsymbol{p}^*$; hence, for some t', $\boldsymbol{\mu}_m(t') = 0$. This contradict with Lemma 7.4 part iv, since $\boldsymbol{p}^* \geq F^\infty(\boldsymbol{1})$.

    b) $\boldsymbol{p}^*$ is the point such that $a_1(t) + a_2(t) + a_m^{(1)}(t) + a_m^{(2)}(t) = 0$. Hence, the only possible candidate for the point $\boldsymbol{p}^*$ is $F^\infty(\boldsymbol{1})$.

$\square$

F.9   Proof of Theorem 7.7

Proof. We will prove this theorem by contradiction. Let assume the process continues even after hitting $F^\infty(\boldsymbol{1})$, i.e., proportion of active half-edges will become positive again. Hence, there is some $\delta > 0$ such that $A_1(k) + A_2(k) + A_m^{(1)}(k) + A_m^{(2)}(k) > n\delta$. Note that for any arbitrary positive value of $\varepsilon < \delta$, corresponding scaled variables are in ε-neighborhood of $F^\infty(\boldsymbol{1})$ which is a contradiction, since, there is no feasible point less that $F^\infty(\boldsymbol{1})$ in ε-neighborhood of it. $\square$

F.10   Proof of Theorem 9.1

Proof. Using equations (22)-(23) we have,

$$
\begin{aligned}
F_{(j,j)}(\mu^{(j,j)}, \mu^{(j,-j)}) &= \sum_{u_j + u_{-j} < K_j(d_j + d_{-j})} \mathbb{P}_j(d_j - 1)\mathbb{P}_m(d_{-j})(1 - \alpha_j(d_j + d_{-j})) \times \\
&\qquad Bi(u_j; d_j - 1, 1 - \mu^{(j,j)})Bi(u_{-j}; d_{-j}, 1 - \mu^{(j,-j)}) \\
&= \sum_{u_j + u_{-j} < K_j(d_j + d_{-j} + 1)} \mathbb{P}_j(d_j)\mathbb{P}_m(d_{-j})(1 - \alpha_j(d_j + d_{-j} + 1)) \times \\
&\qquad Bi(u_j; d_j, 1 - \mu^{(j,j)})Bi(u_{-j}; d_{-j}, 1 - \mu^{(j,-j)}) \\
F_{(-j,j)}(\mu^{(j,j)}, \mu^{(j,-j)}) &= \sum_{u_j + u_{-j} < K_j(d_j + d_{-j})} \mathbb{P}_j(d_j)\mathbb{P}_m(d_{-j} - 1)(1 - \alpha_j(d_j + d_{-j})) \times \\
&\qquad Bi(u_{-j}; d_{-j} - 1, 1 - \mu^{(j,-j)})Bi(u_j; d_j, 1 - \mu^{(j,j)}) \\
&= \sum_{u_j + u_{-j} < K_j(d_j + d_{-j} + 1)} \mathbb{P}_j(d_j)\mathbb{P}_m(d_{-j})(1 - \alpha_j(d_j + d_{-j} + 1)) \times \\
&\qquad Bi(u_{-j}; d_{-j}, 1 - \mu^{(j,-j)})Bi(u_j; d_j, 1 - \mu^{(j,j)})
\end{aligned}
$$

Hence, if $\mu^{(1,1)} = \mu^{(2,1)}$ and $\mu^{(2,2)} = \mu^{(1,2)}$, then $F_{(1,1)}(\boldsymbol{\mu}) = F_{(2,1)}(\boldsymbol{\mu})$ and $F_{(2,2)}(\boldsymbol{\mu}) = F_{(1,2)}(\boldsymbol{\mu})$. Finally, note that since these equalities hold at the initial

point of ODE (24), it will preserved over the trajectory, i.e. $\mu^{(1,1)}(t) = \mu^{(2,1)}(t)$ and $\mu^{(2,2)}(t) = \mu^{(1,2)}(t)$ for all t. □

F.11 Proof of Theorem 9.2

Proof. If $\mu^{(1,1)} = \mu^{(2,2)}$ and $\mu^{(2,1)} = \mu^{(1,2)}$ then we will have $F_{(1,1)}(\boldsymbol{\mu}) = F_{(2,2)}(\boldsymbol{\mu})$ and $F_{(2,1)}(\boldsymbol{\mu}) = F_{(1,2)}(\boldsymbol{\mu})$. Hence, by same type of argument as in proof of Theorem 9.2, we have $\mu^{(1,1)}(t) = \mu^{(2,2)}(t)$ and $\mu^{(2,1)}(t) = \mu^{(1,2)}(t)$ for all t. □

F.12 Proof of Theorem 8.1

Note that the function $F(\boldsymbol{\alpha}_s, \cdot)$ converges point-wise to $F(\mathbf{0}, \cdot)$ as $s \to \infty$.

Lemma F.1. *If the sequence $\{\boldsymbol{\alpha}_s\}_{s=1}^{\infty}$ converges to $\mathbf{0}$, then $\lim_{s \to \infty} F(\boldsymbol{\alpha}_s, \boldsymbol{\mu}) = F(\mathbf{0}, \boldsymbol{\mu})$*

Proof. Sine $\boldsymbol{\alpha}_s \to \mathbf{0}$, for every ϵ, there is large enough S such that for all $s > S$, each element of $\boldsymbol{\alpha}_s$ is smaller than ϵ. Hence, $F(\boldsymbol{\alpha}_s, \boldsymbol{\mu}) > F(\epsilon \mathbf{1}, \boldsymbol{\mu})$ component-wise for all $\boldsymbol{\mu} \subset [0,1]^4$. Using this property,

$$\|\mathbf{F}(\mathbf{0}, \boldsymbol{\mu}) - \mathbf{F}(\boldsymbol{\alpha}_s, \boldsymbol{\mu})\|_2 \quad \leq \quad \|\mathbf{F}(\mathbf{0}, \boldsymbol{\mu}) - \mathbf{F}(\epsilon \mathbf{1}, \boldsymbol{\mu})\|_2 \qquad (55)$$
$$= \quad \epsilon \|\mathbf{F}(\mathbf{0}, \boldsymbol{\mu})\|_2 \qquad (56)$$

The equality (56), is coming from the fact that $\mathbf{F}(\epsilon \mathbf{1}, \boldsymbol{\mu}) = (1 - \epsilon) \mathbf{F}(\mathbf{0}, \boldsymbol{\mu})$. Arbitrary choice of ϵ completes the proof. □

Let's define the set $U(\boldsymbol{\alpha})$ to be largest connected set containing $(1,1,1,1)$ such that for every $\boldsymbol{\mu} \in U(\boldsymbol{\alpha})$, we have $\boldsymbol{\mu} \geq F(\boldsymbol{\alpha}, \boldsymbol{\mu})$. It is easy to see that for any $\boldsymbol{\alpha}_s \neq \mathbf{0}$, we have $F(\boldsymbol{\alpha}_s, \boldsymbol{\mu}) < F(\mathbf{0}, \boldsymbol{\mu})$. Hence, we have $U(\mathbf{0}) \subset U(\boldsymbol{\alpha}_s)$. Similarly, since $\boldsymbol{\alpha}_s \to \mathbf{0}$, for every ϵ, there is a large enough S such that for all $s > S$, each element of $\boldsymbol{\alpha}_s$ is smaller than ϵ. Hence, $F(\boldsymbol{\alpha}_s, \boldsymbol{\mu}) > F(\epsilon \mathbf{1}, \boldsymbol{\mu})$ component-wise and $U(\boldsymbol{\alpha}_s) \subset U(\epsilon \mathbf{1})$. Now, consider the sequence $\{1/k\}_{k=0}^{\infty}$. Then we have,

$$\forall k, \ \exists S, \ s.t. \ \forall s > S, \qquad U(\mathbf{0}) \subset U(\boldsymbol{\alpha}_s) \subset U(1/k \times \mathbf{1})$$

Using Lemma F.1, we can see that $U(1 \times \mathbf{1}) \supset U(1/2 \times \mathbf{1}) \supset \ldots$ and $\bigcap U(1/k \times \mathbf{1}) = U(\mathbf{0})$.

Finally, note that $F^{\infty}(\boldsymbol{\alpha}_s, 1)$ is the furthermost point in $U(\boldsymbol{\alpha}_s)$ from $(1,1,1,1)$. Using properties of this point given by Lemma 7.4, it is easy to see that the sequence $\{F^{\infty}(\boldsymbol{\alpha}_s, 1)\}_{s=0}^{\infty}$ is sandwiched between the sequence $\{F^{\infty}(1/k \times \mathbf{1}, 1)\}_{s=0}^{\infty}$ and $\boldsymbol{\xi}^*$ where $\boldsymbol{\xi}^*$ is equal to $\mathbf{1}$ if $U(\mathbf{0})$ is singleton, and is equal to $F^{\infty}(\mathbf{0}, \boldsymbol{\mu})$ for some $\boldsymbol{\mu} \in U(\mathbf{0})$ not equal to $(1,1,1,1)$ other wise. Hence, if $U(\mathbf{0})$ is singleton, then the final proportion of adopters converges to 0 as $\boldsymbol{\alpha}_s \to \mathbf{0}$. Otherwise, the final proportion of adopters is strictly positive. Moreover, we have,

$$\lim_{s \to \infty} F^{\infty}(\boldsymbol{\alpha}_s, \mathbf{1}) = F^{\infty}(\mathbf{0}, \boldsymbol{\mu}) \qquad \forall \mu \in U/\{\mathbf{1}\}$$

Note that, this result is independent of choice of the sequence $\{\boldsymbol{\alpha}_s\}_{s=0}^{\infty}$.