Diffusion Games*

Evan Sadler†

August 20, 2016

Abstract

I study how the structure of a social network affects the diffusion of a new product or technology. The model deals explicitly with the network's discrete structure, in contrast with most extant work that uses a mean-field approximation. My findings highlight important qualitative differences in predicted diffusion patterns: long-run outcomes are stochastic, individuals can remain isolated, and the likelihood of a large cascade is sensitive to early adoption patterns. The analysis requires technical advances that leverage recent mathematical work on random graphs. A key contribution is a set of structural results for a large class of random graph models that can exhibit observed features of real networks—features like homophily and clustering. These results allow us to characterize the extent and rate of diffusion as a function of network structure.

1 Introduction

New technologies often require time to reach everyone who would benefit from them, even if the benefits are clear. Despite supportive evidence that is decades old in some cases, adoption rates of many medical interventions, such as the prescription of beta blockers following a heart attack, remain surprisingly low in certain regions of the United States (Jencks et al., 2003). Economic models going back to Griliches (1957) emphasize the importance of profitability and risk in technology adoption decisions. More recent empirical work shows that education and other measures of social capital are often predictive of adoption (Caselli and Coleman, 2001; Skinner and Staiger, 2005). However, such considerations fail to provide a compelling explanation for the slow rate of adoption in many instances. Physicians undergo two decades of formal education, and there are few economic incentives involved in the decision to prescribe a particular drug. Why do these innovations diffuse so slowly?

*I am grateful to David Pearce for many helpful discussions about this work and to Roy Radner for comments on early versions. I am indebted to Ilan Lobel, Ben Golub, Arun Chandrasekhar, and Jane Zhou for their careful readings and useful feedback. I also thank participants at the 2016 Workshop on Information and Social Economics for comments and suggestions.

†Harvard University – esadler@fas.harvard.edu

This phenomenon is hardly unique to medicine. From agricultural technologies in developing countries to virally marketed products in developed countries, diffusion is often slow, uneven, and difficult to predict.[1] Similar products have vastly different adoption trajectories, and even the same product can experience variance across different countries or regions. Lack of awareness offers one appealing explanation why people lag in their adoption of new technologies—it seems unlikely that many doctors knowingly provide substandard care.

Nevertheless, formal diffusion models typically preclude the possibility of long-term unawareness. This feature is rarely a deliberate choice. Rather, it is an artifact of the available modeling technology. For the sake of tractability, most diffusion models use some version of a mean-field approximation: we assume that the neighbors who influence our choices are always representative of the population. In essence, these models assume that the social network gets completely reshuffled each period, implying that no one can remain isolated from adopters for very long. If we account for the stability of real relationships, different patterns can emerge. Small insular groups may never become aware of a product. People in loosely connected groups might not adopt because those at the boundary are unreceptive. One set of seeds might lead to viral spread while another results in a small niche of users.

I present a framework to study diffusion with strategic interactions in large populations. Players are connected in a network, represented as an undirected graph, and a few random seeds are aware of a new product. Those who are aware choose whether to adopt. Neighbors of adopters become aware and subsequently make decisions themselves. A player's choice may depend on what she expects her neighbors to do because of local complementarities. My main results characterize the extent and rate of adoption as a function of the network.

The diffusion patterns we find are qualitatively different from those in mean-field models. Long-run outcomes are stochastic, rather than deterministic: the same product in the same network will reach more or less of the population depending on who the initial seeds are. The likelihood, extent, and speed of a large cascade depend on details of the network structure beyond density, and there are always pockets of individuals who remain unaware. The early stages of a cascade can tell us a lot about whether it will continue to grow. Greater breadth, rather than depth, is much more likely to produce a large cascade, a finding that agrees with empirical studies (Cheng et al., 2014). In contrast, mean-field models have difficulty generating variance and say nothing about how early patterns correlate with eventual size. These features limit the applicability of mean-field models—a marketer evaluating a viral campaign will care about upside potential and making predictions based on early results.

Studying diffusion in fixed graphs is technically challenging. A central contribution of this paper is a toolkit flexible enough to tackle the diversity of applications that the mean-field literature explores. Two key pieces are the configuration model of random graphs and the use of a branching process to approximate local network structure. The configuration model is a widely-used generative model of random graphs.[2] In its basic version, we start with a

[1] Skinner and Staiger (2005) document large variation in adoption rates of several technologies across geographic regions.

[2] See Bollobás (1980) and Molloy and Reed (1995). There are many other random graph models we could choose, but we can encompass most of them within some version of the configuration model. For instance, the Erdos-Rényi model is asymptotically equivalent to a configuration model with a Poisson degree distribution,

collection of vertices and assign link stubs to each. These link stubs are paired uniformly at random to form links. We can capture realistic network features, like homophily and clustering, if we allow multiple types of vertices or links. Recent mathematical work shows how we can use a branching process to approximate the structure of the basic configuration model (Bollobás and Riordan, 2015). I extend these findings to multi-type graphs, allowing a much wider range of applications.

These tools facilitate new insights on old questions. A more centralized network leads to faster diffusion and makes large cascades more likely, but it may also reduce the long-run extent of adoption. Clustering can inhibit diffusion because each adoption informs fewer new players, but it can also enhance the effects of complementarities, encouraging more adoption. The latter effect is more significant in a dense network. We can decompose the impact of homophily into two distinct effects. Holding behavior fixed, an increase in homophily increases the rate of diffusion because it concentrates connections among those most inclined to adopt. Adoption complementarities create a second effect: homophily exacerbates differences in types' propensities to adopt. Individuals with high (low) valuations become more (less) willing to adopt because neighbors' valuations are positively correlated. With more homophily, large cascades become more common and grow faster, but they are limited to smaller niche communities.

A pair of applications shows how diffusion patterns can impact broader economic outcomes. In a model of Bertrand competition with word-of-mouth communcation among consumers, the network structure affects the level of competitive pressure. A firm might charge above its marginal cost if it expects many customers to purchase before learning about its competitor. However, if the network is large and sparse, the need to spread awareness before the other firm leads to intense competition on price. In a second application, I adapt the model of Lobel et al. (2015) to study the optimal design of referral rewards. The authors abstract away from the dynamics of diffusion to make the analysis tractable. I show that key findings are robust: a monopolist can use referral rewards to price discriminate, and the monopolist prefers to offer lower effective prices to high-degree consumers.

1.1 Related Literature

Models making some version of a mean-field approximation dominate the literature on diffusion processes. The simplest are variants of the SIS model widely used to study the spread of infectious diseases. In one example, Jackson and Rogers (2007) model diffusion using a degree-based mean-field approximation. Individuals become infected and recover at rates that depend on their degrees and the average infection rate among neighbors. The degree distribution in the network uniquely characterizes the steady-state infection rate, and the authors derive comparative statics based on stochastic dominance of degree distributions. The SIS framework is especially tractable, allowing numerous extensions and applications.[3]

and generalized random graphs are asymptotically equivalent to configuration models with appropriately specified degree distributions. See Janson (2010) for a more formal discussion of asymptotic equivalence.

[3]Not all diffusion models using a mean-field assumption fall into the SIS framework. For example, Jackson and Yariv (2007) model diffusion through a version of best response dynamics, leading to steady states that

For instance, López-Pintado (2008) allows infection probabilities following a more general functional form, Jackson and López-Pintado (2013) incorporate multiple types of players, and Galeotti and Rogers (2012) consider how immunization impacts the steady state

An important feature of these models is that adoption decisions are easily reversed: at steady state, many players adopt and an equal number abandon in each period. While rapid turnover makes sense in some contexts, individual use of a product or technology is typically persistent over time. In many cases, adoption is irreversible—at least in the short run—or adoption constitutes a one-time consumption (e.g. watching a video or reading a news story). Making adoption irreversible is a straightforward change, but then steady states become insensitive to the network structure because all players are eventually exposed to neighborhoods comprised entirely of adopters. In this context, Young (2009) motivates different infection functions based on different influence mechanisms—contagion, social utility, social learning—and explores how the mechanisms affect the shape of the adoption curve.

A growing body of empirical research reveals patterns that fail to match the predictions of mean-field models. Studies of recommendation networks (Leskovec et al., 2006) and link sharing (Goel et al., 2012) find widely varying cascade structures, even among similar products, and the size distribution is heavy-tailed. In contrast, cascade sizes in a mean-field model should concentrate around a mean value. No individual in a mean-field model can detectably influence the aggregate outcome, but a large marketing literature documents the outsized role of opinion leaders (e.g Feick and Price, 1987; Tucker, 2008). Recent field experiments use centrality measures to choose more effective seedings (Banerjee et al., 2013, 2016), but these measures have no natural analog in a mean-field setting. A mean-field assumption makes a lot of sense when studying infectious diseases, where exposure can occur inadvertently in a public place, and a person's contacts today may be completely different from those tomorrow. Social relationships, geographic neighbors, and business connections are far less fleeting, making the assumption more questionable in other settings.

Morris (2000) stands out as an early study of diffusion in a fixed network. In each period, players adopt a behavior if a sufficiently high fraction of their neighbors did in the previous period. The central question is when can behavior spread from a finite group to the whole infinite network. The analysis reveals how tightly knit subgroups act as barriers to diffusion, but some degree of clustering—expressed through the "low neighbor growth" condition—facilitates diffusion. Watts (2002) and Campbell (2013) are among the closest precursors to the present paper. Both use the basic configuration model to generate a network and apply older percolation results to study cascade patterns and the extent of diffusion. Watts (2002) assumes each player has a random threshold and will adopt if the fraction of adopting neighbors exceeds this threshold. Due to the tree-like structure of the basic configuration model, cascades depend crucially on players who adopt after a single neighbor does so.

Campbell (2013) goes a step further, studying monopoly pricing when individuals learn about a good through word-of-mouth. In a baseline model, he characterizes demand elasticities and shows that prices should be set lower than the standard monopoly price. Extensions

correspond to Bayes-Nash equilibria in a static network game, and Galeotti and Goyal (2009) study a two-period model in which an agent outside the network chooses seeds to maximize an objective.

show that the qualitative findings are robust to a particular kind of clustering, but homophily might lead to an optimal price higher than the standard monopoly price. The present paper can furnish alternative proofs of these results and make further extensions feasible, like the treatment of Bertrand competition in section 5.

Beyond models of diffusion, I rely on recent studies of network games to endogenize adoption probabilities. Sundararajan (2007) and Galeotti et al. (2010) study one-shot network games in which payoffs can depend on neighbors' choices. I adapt these results to describe individual behavior in the diffusion model, adding a strategic component to the analysis.

1.2 Moving Away from Mean-Field

Before proceeding with technical preliminaries, I present an example to highlight challenges involved in moving away from mean-field models. Suppose a new product is available. Consumer values are independent and uniform on $[0, 1]$, and there is a fixed price p. There are no complementarities, so adoption decisions are very simple: purchase if and only if your value is above p. Purchasing is irreversible, and it may induce awareness among neighbors.

First, consider first a mean-field approach. The population is large, and an initial fraction q_0 is aware of the product. For simplicity, assume everyone has degree d. In each period, a player draws her neighbors uniformly from the population. If at least one of these neighbors has purchased, she becomes aware and makes a decision. If a fraction x of the population has purchased at the beginning of a period, then an unaware player becomes aware with probability $1 - (1 - x)^d$. Regardless of the density d, all players eventually become aware, and the long-run fraction who purchase is $1 - p$.

Now consider the same diffusion process on a fixed graph. Players' degrees no longer provide all the information we need, and small changes to the graph can have a large impact on demand. Figure 1 depicts two graphs with the same degree distribution. Suppose one player is seeded at random. In the left graph, with any initial seed many players will never learn about the product. In the right graph, there is a chance that everyone becomes aware. Nevertheless, if the red player happens to have a low valuation, the two graphs produce similar results. In both graphs, there is a wide distribution of possible outcomes, and this distribution is highly sensitive to individual players' decisions.

To get away from mean-field assumptions, we need to make choices about how to represent the network. It seems too ambitious to characterize outcomes in every possible graph. While we can gain insights from looking at special cases—star networks, ring networks, lattices—these often do not at all resemble the complex networks we observe in the real world. Realistic networks have heavy tailed degree distributions, short path lengths, and a high degree of clustering. As long as we are modeling the network's discrete structure, we should strive to capture realistic features as best we can.

Instead of looking at particular graphs, I focus on a class of random graphs—configuration models—with well-understood statistical properties. Versions of the configuration model generate realistic networks, and the model primitives correspond directly to features we can

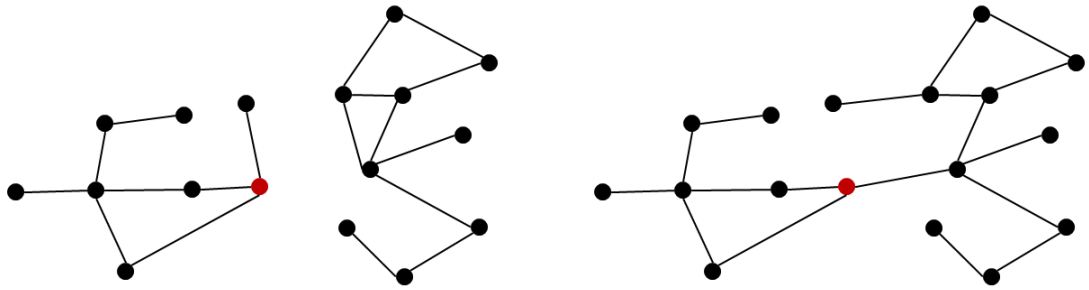

Figure 1: Two graphs with the same degree distribution

measure, so it can be calibrated for applications.[4] Using a random graph model means that any claims we make are about what happens in typical realizations of the graph. We therefore need results that characterize what a typical realization looks like.

2 Preliminaries

This section presents the main technical contribution of the paper. I define several versions of the configuration model, introduce the approximating branching process, and state the key theorem. The proof is given in an appendix. Since the requisite background material may be unfamiliar, appendices also present standard results on branching processes and configuration models.

Any configuration model is generated in two steps. Starting with n vertices, we first assign link stubs to each vertex. Next, we specify a process to randomly join link stubs to form links. Different versions vary the details of these two steps. In the first, we might label vertices or link stubs with types, using these types to restrict which links can form. In the second, we might pair link stubs together, or we might join them in groups of k to add clustering. As long as we have sufficient independence in the joining process, we can apply the same techniques to analyze any version of the configuration model.

Definition 1 (Standard Configuration Model)**.** *Let* $\mathbf{d} = (d_1, d_2, ..., d_n)$ *denote an* n*-vector of positive integers. Given* n *vertices, assign* d_i *link stubs to vertex* i*. Successively pick pairs of stubs uniformly at random, and join the two stubs to form a link. This process produces a distribution over graphs* G *that we call the* ***configuration model with degree sequence*** $\mathbf{d}$*, written* $CM_{n,\mathbf{d}}$*.*

Let D *denote a distribution on* $\mathbb{N}$ *with finite expectation. Given* n *vertices, realize a degree sequence* $\mathbf{d}$ *via* n *independent draws from* D*. Given this sequence, realize a graph according*

[4] Chandrasekhar and Jackson (2015) present methods to estimate the structural parameters of a closely related random graph model.

to $CM_{n,\mathbf{d}}$. This process produces a distribution over graphs G that we call the ***configuration model with degree distribution*** *D.*[5]

The standard configuration model $CM_{n,D}$ assigns a degree to each vertex independently according to D and takes a uniform draw of all possible graphs with the resulting degree sequence. As stated, the definition allows graphs that contain self links or multiple links between the same pair of vertices. When studying diffusion, I condition the configuration model on realizing a *simple graph*—a graph with no self links and at most one link between any two vertices. Results in the appendix show that this conditioning has essentially no impact on the graph's asymptotic structure.

A complication in economic models is that diffusion patterns depend on both the network structure and on individual decisions to adopt. Equilibrium decisions induce a subnetwork of *potential adopters* who will adopt if they become aware of the product. It is the structure of this subnetwork, rather than that of the broader network, that determines diffusion patterns. Consequently, we require the following non-standard result for our analysis.

Suppose we select a subgraph of $CM_{n,D}$ by including each degree d node with independent probability q_d. Write $CM_{n,\mathbf{q},D}$ for the model in which we realize a graph according to $CM_{n,D}$, select nodes according to the probabilities q_d, and retain the subgraph of links between the selected nodes. Let S denote the collection of nodes in a realization of $CM_{n,\mathbf{q},D}$. The subgraph has a realized degree sequence $\mathbf{d}^{(|S|)}$. Conditional on selecting $|S|$ nodes and realizing the sequence $\mathbf{d}^{(|S|)}$, the distribution of $CM_{n,\mathbf{q},D}$ is exactly that of $CM_{|S|,\mathbf{d}^{(|S|)}}$: the subgraph is itself a configuration model.

As n grows, the degree distribution of the subgraph approaches a natural limit. Define the distribution $D_{\mathbf{q}}$ in two stages. First, draw $k \in \mathbb{N}$ according to a distribution taking the value i with probability $\frac{q_i p_i}{\sum_{j=0}^{\infty} q_j p_j}$. This is the degree distribution for a random node that gets selected according to $\mathbf{q}$. After drawing k, draw the degree d according to a binomial distribution with k trials and success probability $\frac{\sum_{i=0}^{\infty} i p_i q_i}{\sum_{i=0}^{\infty} i p_i}$, the probability that a random link stub is attached to a node selected according to $\mathbf{q}$. In essence, we draw the degree of a selected node, and we retain each link stub according to the probability that it links to another selected node. Let $p_{\mathbf{q},d}$ denote $\mathbb{P}(D_{\mathbf{q}} = d)$.

Proposition 1. *For any $\epsilon > 0$, there exists $\delta > 0$ such that*

$$\mathbb{P}_{CM_{n,\mathbf{q},D}}\left(\sup_d \left|\frac{n_d(G)}{|S|} - p_{\mathbf{q},d}\right| \geq \epsilon\right) < e^{-\delta n}.$$

Proof. See Appendix. □

Proposition 1 tells us that selecting a subgraph of $CM_{n,D}$ according to $\mathbf{q}$ is asymptotically equivalent to the model $CM_{n,D_{\mathbf{q}}}$. This means that we can apply our structural results for the configuration model to any subgraph generated in this manner—we simply replace the

[5] We need to ensure that $\mathbf{d}$ has an even number of link stubs. There are many ways to make a correction—condition on realizing an even number, add one stub to the last vertex if needed—that are asymptotically equivalent.

degree distribution D with $D_{\mathbf{q}}$.[6] Using an identical argument, we can show that analogs of Proposition 1 hold for more general configuration models that we use through most of the paper.

The main result concerns a version of the configuration model in which multiple types of vertices can have different propensities to link with one another. Let Θ denote a finite set of types. In a multi-type configuration model, we assign to each vertex i a type θ_i and a degree tuple $d^{(i)} = (d_1^{(i)}, d_2^{(i)}, ..., d_{|\Theta|}^{(i)})$. The element $d_\theta^{(i)}$ in the degree tuple indicates the number of link stubs that can connect with type θ vertices.

Definition 2 (Multi-type Configuration Model)**.** *Let $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_n)$ denote an n-vector of types, and let $\mathbf{d}$ denote an n-vector of degree tuples. Given n vertices, assign type θ_i and degree tuple $d^{(i)}$ to vertex i. Successively pick pairs of compatible stubs uniformly at random (e.g. a type θ stub attached to a type θ' vertex and a type θ' stub attached to a type θ vertex) and join the two stubs to form a link. This process produces a distribution over graphs G that we call the* ***multi-type configuration model with type sequence $\boldsymbol{\theta}$ and degree sequence*** $\mathbf{d}$*, written $CM_{n,\boldsymbol{\theta},\mathbf{d}}$.*

Let (T, D) denote a joint distribution on types and degree tuples. Given n vertices, realize a type and degree sequence $(\boldsymbol{\theta}, \mathbf{d})$ via n independent draws from (T, D). Given this sequence, realize a graph according to $CM_{n,\boldsymbol{\theta},\mathbf{d}}$. This process produces a distribution over graphs G that we call the ***multi-type configuration model with distribution*** *(T, D), written $CM_{n,(T,D)}$.*

We can use branching processes to characterize the asymptotic structure of $CM_{n,D}$ and $CM_{n,(T,D)}$. For the standard model $CM_{n,D}$, define the branching process $\mathcal{T}_D$ in two stages. There is a single root that realizes offspring according to D. Each subsequent node in the branching process realizes offspring independently according to the *forward distribution* D', satisfying

$$\mathbb{P}(D' = d) = \frac{(d+1)\mathbb{P}(D = d+1)}{\mathbb{E}[D]}. \tag{1}$$

We write $\nu_D = \mathbb{E}[D']$ for the average forward degree, which captures important properties of $\mathcal{T}_D$. To get an intuition for the forward distribution, consider a breadth first search starting from a random vertex in $CM_{n,D}$. The vertex has neighbors distributed approximately according to D. To get the distribution of neighbors' neighbors, we need to make two adjustments to D. First, higher degree vertices are disproportionately represented as neighbors: a degree d vertex has d chances to connect with a given link stub. We must also avoid counting the link back to the original vertex, so we shift the distribution by one.

We can define the multi-type branching process $\mathcal{T}_{(T,D)}$ analogously. The root realizes a type according to T and a tuple of offspring according to D_θ. The forward distribution depends on the type of the parent and the type of the child, so properly specifying the behavior of $\mathcal{T}_{(T,D)}$ after the root requires $|\Theta|^2$ types, one for each parent and offspring pair.

[6] Combining this Proposition with the main theorem effectively generalizes percolation results to cases with arbitrary heterogeneity in how permeable different links are.

I refer to a type θ vertex with a type θ' parent as a type $\theta\theta'$ vertex. A type $\theta\theta'$ vertex will realize the tuple of offspring $d = (d_1, d_2, ..., d_\Theta)$ with probability

$$\mathbb{P}(D'_{\theta\theta'} = d) = \frac{(d_{\theta'} + 1)\mathbb{P}(D_\theta = d + e_{\theta'})}{\mu_{\theta,\theta'}}. \tag{2}$$

The entry $d_{\tilde{\theta}}$ in d denotes the number of type $\tilde{\theta}\theta$ offspring, the symbol $\mu_{\theta,\theta'}$ is the expected number of type θ' neighbors of a type θ vertex according to (T, D), and $e_{\theta'}$ is a vector of zeros with a one in the entry corresponding to θ'. The number of type $\tilde{\theta}\theta'$ offspring is zero whenever $\theta' \neq \theta$.

To obtain the parameter analogous to ν_D, define $\nu_{\theta_1,\theta_2,\theta_3} = \mathbb{E}\left[(D'_{\theta_2\theta_1})_{\theta_3}\right]$ as the expected number of type $\theta_3\theta_2$ offspring of a type $\theta_2\theta_1$ vertex. We use the collection $\{\nu_{\theta_1,\theta_2,\theta_3}\}$ to populate the entries of the *mean offspring matrix* $M_{(T,D)}$. The matrix $M_{(T,D)}$ is a $|\Theta|^2$ by $|\Theta|^2$ matrix, and we set the entry in row $\theta_3\theta_2$ and column $\theta_2\theta_1$ equal to $\nu_{\theta_1,\theta_2,\theta_3}$. While $M_{(T,D)}$ contains $|\Theta|^4$ entries, note only $|\Theta|^3$ are nonzero.

Given a graph G, write $N_k(G)$ for the number of vertices in components of size k, write $L_1(G)$ for the number of vertices in the largest component, and write $L_2(G)$ for the number of vertices in the second largest component. Moreover, let $H(G)$ be a random variable denoting the length of the shortest path between two vertices chosen uniformly at random. Let $\nu_{(T,D)}$ denote the spectral radius of $M_{(T,D)}$, and let ρ_k denote the probability that $|\mathcal{T}_{(T,D)}| = k$.

Theorem 1. *Suppose $\mathcal{T}_{(T,D)}$ is irreducible and non-singular. For any $\epsilon > 0$, we have*

$$\lim_{n\to\infty} \mathbb{P}_{CM_{n,(T,D)}}\left(\left|\frac{N_k(G)}{n} - \rho_k\right| \geq \epsilon\right) = 0,$$

$$\lim_{n\to\infty} \mathbb{P}_{CM_{n,(T,D)}}\left(\left|\frac{L_1(G)}{n} - \rho_\infty\right| \geq \epsilon\right) = 0, \text{ and}$$

$$\lim_{n\to\infty} \mathbb{P}_{CM_{n,(T,D)}}\left(\frac{L_2(G)}{n} \geq \epsilon\right) = 0.$$

Suppose $\nu_{(T,D)}$ is finite. For any $\epsilon > 0$, we have

$$\lim_{n\to\infty} \mathbb{P}_{CM_{n,(T,D)}}\left(\left|\frac{H(G)}{\log_{\nu_{(T,D)}} n} - 1\right| \geq \epsilon \,\middle|\, H(G) < \infty\right) = 0.$$

Additionally, each of these statements continues to hold if we condition $CM_{n,(T,D)}$ on realizing a simple graph.

Proof. See Appendix. □

The probability that a random vertex in $CM_{n,(T,D)}$ is in a component of size k is approximately the probability that $\mathcal{T}_{(T,D)}$ realizes a tree of size k. As n grows, there is at most

one giant component in $CM_{n,(T,D)}$, and the portion of the network it covers is equal to the survival probability of $\mathcal{T}_{(T,D)}$. These results allow us to characterize the extent of diffusion in latter sections. If a giant component exists, the distance between two random vertices in the component is approximately $\log_{\nu_{(T,D)}} n$. As a result, the parameter $\nu_{(T,D)}$ tells us how fast diffusion spreads through the network.[7]

Real networks exhibit far more clustering than either $CM_{n,D}$ or $CM_{n,(T,D)}$—two individuals with friends in common are more likely to be friends themselves. To add clustering, we can introduce multiple types of link stubs. Modify the standard configuration model by assigning a tuple of link stubs $(d_2^{(i)}, d_3^{(i)}, ..., d_k^{(i)}, ...)$ to each vertex, indicating that vertex i has $d_k^{(i)}$ link stubs of type k. Instead of pairing all stubs, we match link stubs of type k uniformly at random in groups of size k, establishing links between all associated vertices. In the standard model, we have only type 2 stubs. Adding type $k > 2$ creates k-cliques in the graph. Vertex i has total degree $d_i = \sum_{k=2}^{\infty}(k-1)d_k^{(i)}$.

In section 3.3, I use a particular version of the configuration model that allows 3-cliques. Vertices have degree pairs (d_2, d_3), where d_2 is the number of normal stubs and d_3 is the number of triangle stubs. The extent of clustering varies with a single parameter γ.

Definition 3 (Configuration Model with γ-clustering)**.** *Let D denote a degree distribution taking even values with probability one, and fix $\gamma \in [0,1]$. In the model $CM_{n,D,\gamma}$, for each of the n vertices we first realize a degree independently according to D and group the link stubs into pairs. For each pair of link stubs, with independent probability γ merge them into a single triangle stub; otherwise they remain normal stubs. Pair normal stubs uniformly at random into links, and join triangle stubs in groups of three uniformly at random.*

The proof of Theorem 1 applies for the configuration model with γ-clustering, and I use the analogous result in section 3.3.

3 Diffusion in a Random Network

There are n players situated in a network G, which is realized according to the model $CM_{n,(T,D)}$, where D has finite variance.[8] Time is discrete, and initially all players are unaware of some new available product. In period 0, players in a subset S_0 become aware. In period 1, players in S_0 decide whether to adopt. Adoption is irreversible, and any neighbor of an adopter becomes aware. In each period $t \geq 2$, those who are aware and have not yet adopted decide whether to adopt, and any neighbor of an adopter becomes aware.

[7] Bollobás and Riordan (2015) prove an analogous result for component sizes in the standard configuration model. The proof I give closely mirrors their argument. van der Hofstad et al. (2005) prove an analogous result for typical distances in the standard configuration model. My proof is substantially different from theirs, taking advantage of the techniques developed by Bollobás and Riordan (2015). van der Hofstad et al. (2007) give a similar result for the infinite variance case in the standard configuration model, showing that typical distances are much smaller, on the order of $\log\log n$. The proof of Theorem 1 in fact implies a much stronger result than what I state. We can characterize any local property of the configuration model using the branching process $\mathcal{T}_{(T,D)}$.

[8] Finite variance is necessary for the results on the rate of diffusion but not the extent.

Our main questions concern the trajectory of adoption over time—its extent, its rate, and its relationship to individual choices and network structure. In a population with n players, write $X_n(t)$ for the number adopting at time t. Define

$$\alpha_n = \lim_{t\to\infty} \frac{X_n(t)}{n}, \quad \text{and} \quad \tau_n(x) = \min\left\{t \,:\, \frac{X_n(t)}{n} \geq x\right\}.$$

The random variable α_n is the long-run fraction of the population that adopts, and $\tau_n(x)$ is the (random) time it takes for a fraction x of the population to adopt. By convention, we set $\tau_n(x) = \infty$ whenever $x > \alpha_n$. Our results are statements about the limiting distributions of α_n and τ_n as $n \to \infty$.

If a player adopts the product in period t, she earns a payoff that depends on the number of her neighbors A_i who adopt by the end of period $t+1$.[9] Following the model $CM_{n,(T,D)}$, player i has type θ drawn independently according to T. Conditional on this type, she has valuation $v_i \in [0,1]$, drawn independently according to a distribution V_θ, and total degree d_i derived from the configuration model. Player i's payoff from adoption is

$$u_i = u(v_i, d_i, A_i).$$

I assume u is strictly increasing in v and non-decreasing in A—adoption can exhibit complementarities. I normalize the payoff from non-adoption to zero, so player i adopts in period t if

$$\mathbb{E}[u(v_i, d_i, A_i)] > 0.$$

I first present a baseline model in which each player i observes only her valuation v_i and degree d_i. Subsequently, I extend the framework to include information about neighbors. In one example, I suppose players receive signals about neighbors' valuations. In another, players observe whether neighbors have already adopted and may wait for more neighbors to adopt before making a choice.

3.1 Diffusion with Limited Information

Each player i observes her valuation v_i and her degree d_i. Using the distributions (T, D) and $\{V_\theta\}_{\theta\in\Theta}$ as a prior, player i uses the information contained in (v_i, d_i) to update her beliefs over neighbors' valuations and degrees.[10] Upon becoming aware, player i presumes that none of her neighbors have adopted yet, but if she adopts, all will make a decision in the following

[9] This assumption seems innoucuous in situations where players realize payoffs quickly (e.g. reading a news article, downloading a song), make choices quickly (e.g. in response to a limited time offer), or heavily discount the future. If payoffs are realized slowly over time, and these payoffs depend on ongoing interactions with neighbors, the assumption becomes more questionable. However, if we can expect a neighbor who refrains from adopting in period $t+1$ to take a significant amount of time before reconsidering, this can reasonably approximate a player's true discounted payoff.

[10] For finite n, these beliefs deviate from the true distribution, but the differences vanish at an exponential rate as n grows.

period.[11] A symmetric strategy profile is a mapping $\sigma(v,d) : [0,1] \times \mathbb{N} \to \{0,1\}$ giving an adoption decision for each possible valuation and degree. The profile σ is an *equilibrium* if

$$\mathbb{E}_\sigma \left[\sigma(v,d)u(v,d,A)\right] \geq \mathbb{E}_\sigma \left[xu(v,d,A)\right]$$

for all pairs (v,d) and all $x \in \{0,1\}$. Since adoption exhibits complementarities, best responses are monotone in σ, and the existence of minimal and maximal equilibria follows from Tarski's fixed point theorem. Having some examples in mind will help with intuition.

Example 1 (Payoffs linear in number of neighbors)**.** *Generate the network according to the single type model $CM_{n,D}$ with a uniform valuation distribution. Payoffs are*

$$u(v,d,A) = Av - c$$

for some $c > 0$.

Example 2 (Payoffs linear in fraction of neighbors)**.** *Generate the network according to the single type model $CM_{n,D}$ with a uniform valuation distribution. Payoffs are*

$$u(v,d,A) = \frac{Av}{d} - c$$

for some $c > 0$.

In each of these examples, we can characterize equilibria via the probability $P_\sigma = \mathbb{E}[\sigma(V, D'+1)]$ that a random neighbor is a potential adopter. In example 1, we have $\mathbb{E}[u(v,d,A)] = dvP_\sigma - c$, and every equilibrium corresponds to a solution of

$$P_\sigma = \mathbb{P}\left(V > \frac{c}{P_\sigma(D'+1)}\right).$$

Equilibrium strategies are clearly increasing in degree, and the results of Galeotti et al. (2010) give comparative statics: if the degree distribution D_1 first order stochastically dominates (FOSD) D_2, then the set of equilibria under D_1 is higher than that under D_2.

In example 2, we can explicitly solve for the set of equilibria. The expected payoff from adoption is $\mathbb{E}_\sigma[u(v,d,A)] = vP_\sigma - c$, implying

$$P_\sigma = \mathbb{P}\left(V > \frac{c}{P_\sigma}\right) = \max\left(0, 1 - \frac{c}{P_\sigma}\right).$$

[11] This avoids the need to treat initial seeds separately from other players. In the early stages of diffusion, any newly aware player has exactly one adopting neighbor, and adjusting for this has no qualitative impact on the results. If adoption spreads to a positive fraction of the network, many newly aware players will have multiple adopting neighbors. These expectations then understate the value of adoption, and the equilibrium I define gives a lower bound on adoption behavior.

Best replies are always independent of degree, and the set of equilibria is independent of D. The profile in which $\sigma(v,d)=0$ uniformly is always an equilibrium. If $c<\frac{1}{4}$, there are two equilibria with positive adoption, corresponding to

$$P_\sigma = \frac{1\pm\sqrt{1-4c}}{2}.$$

In the maximal equilibrium, a player adopts if $v>\frac{1-\sqrt{1-4c}}{2}$.

Given an equilibrium, the results from section 2 let us characterize diffusion outcomes. Define the probabilities $q_{\theta,d,\sigma}=\mathbb{E}[\sigma(V_\theta,d)]$; the model $CM_{n,\mathbf{q}_\sigma,(T,D)}$ describes the subnetwork of potential adopters. The set of people who end up adopting are precisely those in connected components of $CM_{n,\mathbf{q}_\sigma,(T,D)}$ that contain an initial seed. Define $\pi_\sigma=\mathbb{E}[q_{T,D,\sigma}]$ as the probability that a random player is a potential adopter, write $\mathcal{T}_\sigma$ for the branching process that approximates $CM_{n,\mathbf{q}_\sigma,(T,D)}$ (see section 2), write ξ_σ for the extinction probability of $\mathcal{T}_\sigma$, and write ν_σ for the spectral radius of the mean offspring matrix.

Theorem 2. *Suppose S_0 consists of k individuals chosen uniformly at random. As $n\to\infty$, the random variables α_n converge in distribution to α_σ satisfying*

$$\alpha_\sigma = \begin{cases} \pi_\sigma(1-\xi_\sigma) & \text{with probability } 1-\xi_\sigma^k \\ 0 & \text{with probability } \xi_\sigma^k. \end{cases}$$

In particular, if $\nu_\sigma\le 1$, we have $\alpha_\sigma=0$ with probability one, and if $\nu_\sigma>1$, we have $\alpha_\sigma>0$ with positive probability. Furthermore, for any x with $0<x<\max\operatorname{supp}(\alpha)$ and any sufficiently small $\epsilon>0$, we have

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{\tau_n(x)}{\log_{\nu_\sigma} n}-1\right|\ge\epsilon \,\middle|\, \alpha_n>\epsilon\right)=0.$$

Proof. A player is contained in the network of potential adopters with probability π_σ, and is in the giant component of this network with probability $1-\xi_\sigma$. The definition of $\tau_n(x)$ means that a fraction x of players are at distance less than or equal to $\tau_n(x)$ from an initial seed. Both claims now follow from Theorem 1. □

With finitely many seeds, the level of adoption is inherently stochastic. Remarkably, the fraction α_σ can only ever take one positive value. More seeds increases the probability of a large cascade but does not affect its reach: we have a better chance of seeding the giant component, but the size of this component remains the same. Just as the first generations in a branching process tell us a lot about whether it will survive, the first steps in a diffusion tell us a lot about the whether we will get a large cascade. Each adoption among a seed's neighbors is another chance to connect to the giant component. If we see an equal number of adoptions in a long chain, the chances are no larger than in the first step.

Even when there is a large cascade, some individuals never learn about the product. In the event that $\mathbb{E}[\sigma(V_\theta,D_\theta)]<1$ for each type θ, we are guaranteed to have $\xi_\sigma>0$. In this case, the first part of Theorem 1 implies there are many small components in the network

of potential adopters. Each of these small components avoids getting seeded with positive probability. In fact, even if we seed a positive fraction of the network, many players must remain uninformed.

Proposition 2. *Suppose that for each individual i we have $i \in S_0$ with independent probability $1-p$. We then have*

$$\frac{\alpha_\sigma}{\pi_\sigma} = 1 - \xi_\sigma + \sum_{m=1}^{\infty} \mathbb{P}(|\mathcal{T}_\sigma| = m)(1-p^m)$$

with probability one. In particular, if $\overline{m} = \mathbb{E}\left[|\mathcal{T}_\sigma| \,\middle|\, |\mathcal{T}_\sigma| < \infty\right]$, we have

$$\frac{\alpha_\sigma}{\pi_\sigma} \leq 1 - \xi_\sigma + \xi_\sigma(1-p^{\overline{m}}) < 1.$$

Proof. The first part of Theorem 1 tells us that the fraction of potential adopters in connected components of size m converges to the probability that $|\mathcal{T}_\sigma| = m$. A component of size m contains an initial seed with probability $(1-p^m)$, and the first part follows. The second part is an immediate application of Jensen's inequality. □

The close connection to the branching process $\mathcal{T}_\sigma$ facilitates a comparative statics analysis. I decompose this analysis into two parts: the effect of equilibrium shifts and the direct effect of changes in network.[12] The effect of a change in σ is straightforward: higher strategies add players to the network of potential adopters, which increases the extent and rate of diffusion. If we remove some potential adopters and replace them with others who have more neighbors, this also increases the extent and rate of diffusion. Diffusion proceeds further and faster when there are more potential adopters or when potential adopters are more connected.

Corollary 1. *If $\sigma \geq \sigma'$, then α_σ FOSD $\alpha_{\sigma'}$ and $\nu_\sigma \geq \nu_{\sigma'}$.*

If σ and σ' are such that $\pi_\sigma = \pi_{\sigma'}$, and $D_{\mathbf{q}_\sigma}$ FOSD $D_{\mathbf{q}_{\sigma'}}$, then α_σ FOSD $\alpha_{\sigma'}$ and $\nu_\sigma \geq \nu_{\sigma'}$.

Proof. Either condition implies that the offspring distribution for $\mathcal{T}_\sigma$ first order stochastically dominates that of $\mathcal{T}_{\sigma'}$, and the result is immediate. □

The network effect is more nuanced as it depends on the strategies players are using. Fix a strategy profile σ. Given two degree distributions D_1 and D_2, write $\alpha_\sigma^{(1)}$, $\nu_\sigma^{(1)}$, $\alpha_\sigma^{(2)}$, $\nu_\sigma^{(2)}$ for the variables α_σ, ν_σ corresponding to the two respective distributions.

Corollary 2. *Suppose the profile σ is weakly increasing in d. If D_1 FOSD D_2, then $\alpha_\sigma^{(1)}$ FOSD $\alpha_\sigma^{(2)}$ and $\nu_\sigma^{(1)} \geq \nu_\sigma^{(2)}$.*

Suppose the profile σ is weakly convex in d. If D_1 is a mean preserving spread of D_2, then $\nu_\sigma^{(1)} \geq \nu_\sigma^{(2)}$.

[12] A complete analysis would also include how changes in the network shift the equilibrium. The findings of Galeotti et al. (2010) apply directly in the single-type version of this model and could be extended for the multi-type version.

Proof. The assumption that σ is increasing in d implies that an FOSD shift in the degree distribution adds players to the network of potential adopters *and* gives them more links. This implies an FOSD shift in the offspring distribution of the corresponding branching process, and the first part follows. Similarly, if σ is convex in d, a mean preserving spread in the degree distribution produces a FOSD shift in the forward distribution, implying the second part. □

These corollaries apply directly in examples 1 and 2. In the first example, strategies are increasing in degree, so Corollary 2 implies that FOSD shifts in the degree distribution should increase the extent and rate of diffusion. The results of Galeotti et al. (2010) imply this shift also leads to an increase in equilibrium strategies, which exaggerates the change. In the second example, strategies are independent of degree, and FOSD shifts in the degree distribution should again increase the extent and rate of diffusion, but without affecting equilibrum strategies. In this example, Corollary 2 also implies that a mean preserving spread increases the speed of diffusion, but the effect on extent is ambiguous. Information spreads faster with more central hubs, but more peripheral players means more small components.

3.2 Diffusion with Signals

Suppose each player i observes a signal $s_i \in \mathcal{S}$ taking one of finitely many values. The distribution of s_i may depend on v_i and d_i, and it may be correlated with the valuations of player i's neighbors. Again using (T, D) as a prior, player i uses the information in (v_i, d_i, s_i) to update her beliefs over neighbors' valuations and degrees. A symmetric strategy profile is now a mapping $\sigma(v, d, s) : [0, 1] \times \mathbb{N} \times \mathcal{S} \to \{0, 1\}$ giving a decision for each possible triple (v, d, s). The profile σ is an equilibrium if for all (v, d, s) and all $x \in \{0, 1\}$, we have

$$\mathbb{E}_\sigma[\sigma(v, d, s)u(v, d, A)] \geq \mathbb{E}_\sigma[xu(v, d, A)].$$

Just as before, best replies are increasing in σ, so Tarski's fixed point theorem ensures existence of minimal and maximal equilibria.

Signals introduce correlations not captured via the types in Θ. In order to apply Theorem 1, we must augment each player's type with her signal: write $\tilde{\theta} = (\theta, s)$, and write $\tilde{T}$ for the joint distribution of T and S. The configuration model $CM_{n,(\tilde{T},D)}$ is entirely equivalent to $CM_{n,(T,D)}$, except with an extra label attached to each vertex indicating the signal that player receives. A strategy profile determines probabilities $q_{\tilde{\theta},d,\sigma} = \mathbb{E}[\sigma(V_\theta, d, s)]$ with which each augmented type and degree of player will adopt, conditional on becoming aware. The subgraph of potential adopters follows the model $CM_{n,\mathbf{q}_\sigma,(\tilde{T},D)}$. Our problem is now equivalent to that in the previous section, and the earlier results apply.

I give one example to illustrate how signals can impact diffusion.

Example 3. *Generate the network according to the single type model $CM_{n,D}$ with a uniform valuation distribution. Payoffs are*

$$u(v, d, A) = Av - c$$

for some $c > 0$. Player i observes a binary signal $s_{ij}\{0,1\}$ for each neighbor j, with

$$\mathbb{P}\left(s_{ij} = 1 \,|\, v_j > \frac{1}{2}\right) = \mathbb{P}\left(s_{ij} = 0 \,|\, v_j \leq \frac{1}{2}\right) = g,$$

for some $g > \frac{1}{2}$. These signals are conditionally independent given neighbors' valuations. Since neighbors are otherwise indistinguishable, we aggregate the signals $\{s_{ij}\}$ into a single random variable s_i denoting the number of 1 signals player i receives.

We can exploit the structure of this example to reduce the size of the augmented type space. Regardless of the signal precision g, the unconditional signal s_i is a binomial random variable with d_i trials and success probability $\frac{1}{2}$. For this example, I assume a newly aware player observes who informed her and counts this neighbor as a certain adopter in her utility calculation. This implies that, conditional on adopting, player i has no information about her neighbors' signals, and we do not need to augment the type space at all.[13] The forward distribution in the relevant branching process realizes offspring according to D' and retains each with independent probability $P_\sigma = \mathbb{E}\left[\sigma(V, D'+1, B_{D',.5})\right]$, where $B_{n,p}$ denotes a binomal distribution with n trials and success probability p.

Proposition 3. *Assume $c \in (0,1)$ in example 3, and let σ denote the maximal equilibrium. The total population of potential adopters π_σ, the extent of adoption α_σ, and the rate ν_σ are decreasing in the signal precision g.*

Proof. I refer to players with valuations above $\frac{1}{2}$ as high types, and those with valuations below $\frac{1}{2}$ as low types. Write P_σ^- and P_σ^+ for the probabilities that a random low and high type neighbor respectively will adopt. We have

$$P_\sigma^- = \mathbb{E}\left[\sigma(U_{[0,.5]}, D'+1, B_{D',.5})\right], \quad \text{and} \quad P_\sigma^+ = \mathbb{E}\left[\sigma(U_{[.5,1]}, D'+1, B_{D',.5})\right],$$

where $U_{[a,b]}$ denotes a uniform distribution on $[a,b]$. Player i's expected payoff from adoption is

$$\left(1 + s_i(gP_\sigma^+ + (1-g)P_\sigma^-) + (d_i - s_i)(gP_\sigma^- + (1-g)P_\sigma^+)\right) v_i - c.$$

For any realization of s, best responses are increasing in v, which implies that $P_\sigma^+ \geq P_\sigma^-$ in any equilibrium. The last equation then implies equilibrium strategies are increasing in s.

Fixing d and s, there is a threshold valuation

$$\overline{v}_{d,s} = \frac{c}{1 + s_i(gP_\sigma^+ + (1-g)P_\sigma^-) + (d_i - s_i)(gP_\sigma^- + (1-g)P_\sigma^+)}$$

at which a player is indifferent between her two choices. Note the degree and signal pairs (d,s) and $(d, d-s-1)$ are equally likely. Suppose $s \leq \frac{d-1}{2}$, so $\overline{v}_{d,s} \geq \overline{v}_{d,d-s-1}$. Taking the partial derivative of $\overline{v}_{d,s}$ and $\overline{v}_{d,d-s-1}$ with respect to g, we find $\overline{v}_{d,s}$ is increasing in g, and $\overline{v}_{d,d-s-1}$ is decreasing. The magnitude of this change is larger for $\overline{v}_{d,s}$, which means the reduction in P_σ^+ is larger than the increase in P_σ^-. Since $P_\sigma = \frac{P_\sigma^- + P_\sigma^+}{2}$, the claim follows. □

[13] With a different assumption, we would need two types in the augmented space, corresponding to whether the neighbor who informed player i has valuation above or below $\frac{1}{2}$.

Information affects diffusion through the equilibrium, not through the network itself. Intuitively, an increase in g leads to offsetting effects when a player has signal s versus signal $d-s$: those whose neighbors have high valuations become more likely to adopt, and those whose neighbors have low valuations become less likely. With a linear payoff function, the latter effect is more pronounced, suggesting that players who are better informed become more selective and less likely to adopt overall.

3.3 Diffusion with Observable Choices

Suppose each player i observes the number of neighbors $a_i^{(t)}$ who have adopted at the beginning of the current period. A Bayesian analysis in this case is intractable. Rather than assume an ad hoc decision rule, in this section I take as given a symmetric strategy profile $\sigma(v,d,a) : [0,1] \times \mathbb{N}^2 \to \{0,1\}$ that is time invariant. I am deliberately agnostic about how σ arises.[14] Our concern is how the network structure affects diffusion for a given σ.

Due to the tree-like local structure of the configuration model, players with (v,d) such that $\sigma(v,d,1)=1$ play a special role. If the number of initial seeds is small, then players have at most one adopting neighbor until a significant portion of the network adopts. Whether diffusion can spread to a positive fraction of the network depends on the subnetwork of players willing to adopt when only one neighbor has done so. I refer to these players as *potential early adopters*. Define

$$\hat{q}_{\theta,d} = \mathbb{E}[\sigma(V_\theta, d, 1)].$$

The network of potential early adopters follows the model $CM_{n,\hat{\mathbf{q}},(T,D)}$, and we can apply Theorem 1 as before. An important difference from the baseline model is that $CM_{n,\hat{\mathbf{q}},(T,D)}$ fails to capture the full extent of adoption: not all potential adopters are potential early adopters. The size of the giant component in $CM_{n,\hat{\mathbf{q}},(T,D)}$ gives a lower bound on the extent of adoption when there is a large cascade. We can refine this bound by iteratively computing the probability that players have multiple adopting neighbors.

When players wait for multiple neighbors to adopt, clustering can have a positive impact on diffusion. In the baseline model, clustering would only slow diffusion because it reduces the number of newly informed players at each step. Here, clustering can also give us more early adopters, which may more than offset the first effect.

To illustrate, recall the configuration model with γ-clustering $CM_{n,D,\gamma}$. Players realize an even number of link stubs according to D. With independent probability γ, each pair of stubs is collapsed to a single triangle stub, and otherwise the stubs remain normal. Normal stubs are paired uniformly at random, while triangle stubs are joined uniformly at random in triples. Higher values of γ correspond to more triangles and hence more clustering in the network.

In the approximating branching process, the forward degree distribution depends on the type of link to the parent. Conditional on being reached through a normal link, the forward degree distribution is D'. Conditional on being reached through a triangle link, the forward

[14] The network level diffusion results apply regardless of how we specify σ. This is already enough to make predictions about aggregate outcomes if we have data on individual adoption decisions.

degree distribution is $D'-1$—we remove the neighbor that is common to parent and offspring. Define $P_{\sigma,1} = \mathbb{E}[\sigma(V, D, 1)]$, and for $a > 1$ define $P_{\sigma,a} = \mathbb{E}[\sigma(V, D, a)] - P_{\sigma,a-1}$. To simplify what follows, assume that σ does not depend on d.

First, suppose $P_{\sigma,a} = 0$ for all $a > 1$. This corresponds to a profile that could arise in the baseline model: either a player adopts right after becoming aware, or she never adopts. In the network of potential adopters, a player reached through a normal link has in expectation $P_{\sigma,1}(\gamma + (1-\gamma)\nu_D)$ additional neighbors reached through normal links, and $P_{\sigma,1}\gamma(\nu_D - 1)$ neighbors reached through triangle links. A player reached through a triangle link has in expectation $P_{\sigma,1}(1-\gamma)(\nu_D - 1)$ neighbors reached through normal links, and $P_{\sigma,1}\gamma(\nu_D - 1)$ additional neighbors reached through triangle links. The corresponding mean offspring matrix is

$$M_\sigma = P_{\sigma,1}\begin{pmatrix} \gamma + (1-\gamma)\nu_D & \gamma(\nu_D - 1) \\ (1-\gamma)(\nu_D - 1) & \gamma(\nu_D - 1) \end{pmatrix}. \tag{3}$$

We can compute the spectral radius of M_σ as

$$\nu_\sigma = \frac{P_{\sigma,1}}{2}\left(\nu_D + \sqrt{\nu_D^2 - 4\gamma(\nu_D - 1)^2}\right).$$

The effect of an increase in γ, is unambiguous: the likelihood of a large cascade, and the rate of diffusion in a large cascade, goes down. This may change if $P_{\sigma,2} > 0$.

Proposition 4. *Suppose the network follows the model $CM_{n,D,\gamma}$, and fix a strategy profile σ that is independent of d. The parameter ν_σ is increasing in γ if and only if $P_{\sigma,2} > \frac{1}{\nu_D - 1}$.*

Proof. A neighbor j reached through a normal link will adopt if and only if $\sigma(v_j, d_j, 1) = 1$. A neighbor j reached through a triangle link will adopt if either $\sigma(v_j, d_j, 1) = 1$, or if $\sigma(v_j, d_j, 2) = 1$ and the other player reached through that triangle link is a potential early adopter. Carrying out a similar exercise as before, we compute the mean offspring matrix M'_σ for the corresponding branching process

$$M'_\sigma = \begin{pmatrix} P_{\sigma,1}(\gamma + (1-\gamma)\nu_D) & P_{\sigma,1}(1 + P_{\sigma,2})\gamma(\nu_D - 1) \\ P_{\sigma,1}(1-\gamma)(\nu_D - 1) & P_{\sigma,1}(1 + P_{\sigma,2})\gamma(\nu_D - 1) \end{pmatrix} \tag{4}$$

with spectral radius

$$\nu_\sigma = \frac{P_{\sigma,1}}{2}\left((\nu_D + P_{\sigma,2}\gamma(\nu_D - 1)) + \sqrt{(\nu_D + P_{\sigma,2}\gamma(\nu_D - 1))^2 - 4(1 + P_{\sigma,2})\gamma(\nu_D - 1)}\right).$$

Differentiating gives

$$\frac{\partial \nu_\sigma}{\partial \gamma} = \frac{P_{\sigma,1}(\nu_D - 1)}{2}\left(P_{\sigma,2} + \frac{P_{\sigma,2}(\nu_D + P_{\sigma,2}\gamma(\nu_D - 1)) - 2(1 + P_{\sigma,2})}{\sqrt{(\nu_D + P_{\sigma,2}\gamma(\nu_D - 1))^2 - 4(1 + P_{\sigma,2})\gamma(\nu_D - 1)}}\right),$$

which is positive if and only if $P_{\sigma,2} > \frac{1}{\nu_D - 1}$. □

When $P_{\sigma,2} > 0$, an increase in γ has two effects: it decreases the forward degree of a typical neighbor, but it increases the likelihood that a neighbor will adopt. If the average forward degree in the network is too low ($\nu_D \leq 2$) the first effect always dominates. As the ν_D increases, there are more chances to get additional adopting neighbors, and the rate parameter may increase with γ. We could carry out the same analysis for a more general model with clustering—adding k-cliques for any k will have the same qualitative effect.

4 Homophily

Among the most ubiquitous features of social networks is that people disproportionately link to those like themselves. Sociologists call this phenomenon homophily, and many studies document homophily along numerous dimensions like race, religion, political beliefs, and other interests (McPherson et al., 2001). How does varying the extent of homophily affect diffusion? The results of the previous section give us the tools to explore this question.

We start from the model in section 3.1. Suppose there are two types, low and high. Low-type players have uniform valuations on $\left[0, \frac{1}{2}\right]$, and high-type players have uniform valuations on $\left[\frac{1}{2}, 1\right]$. The population is evenly split between the two types, and all have total degree drawn independently according to D. Each link is an own-type link with independent probability $h \in \left[\frac{1}{2}, 1\right]$ and an other-type link otherwise. If $h = \frac{1}{2}$, we get a single type model with uniform valuations and degree distribution D. For $h > \frac{1}{2}$, we have homophily along valuations. Assume $u(v, d, A) = \frac{vA}{d} - c$ for some $c \in [0, 1]$: there is a constant cost to adoption, and a benefit proportional to the fraction of adopting neighbors. As in example 2, this utility ensures that strategies are independent of degree.

A player adopts if her neighbors are potential adopters with probability at least $\frac{c}{v}$. Given σ, let $P_\sigma^+ = \mathbb{E}[\sigma(U_{[.5,1]}, D)]$ denote the probability that a high-type player adopts, and let $P_\sigma^- = \mathbb{E}[\sigma(U_{[0,.5]}, D)]$ denote the probability that a low-type player adopts. For a high-type player, neighbors adopt with probability $hP_\sigma^+ + (1-h)P_\sigma^-$ and for a low type player, with probability $hP_\sigma^- + (1-h)P_\sigma^+$. Expected payoffs are increasing in v and in the neighbor adoption probability. This implies that in the maximal equilibrium, strategies are increasing in v, and we can characterize equilibrium behavior via a single threshold $\overline{v}$.

Proposition 5. *In the maximal equlibrium:*

(a) If $c > \frac{h}{2}$, then $P_\sigma^+ = P_\sigma^- = 0$.

(b) If $\frac{1}{8h} < c \leq \frac{h}{2}$, then $P_\sigma^+ = 1$ and $P_\sigma^- = 0$.

(c) If $c \leq \frac{1}{8h}$, then $P_\sigma^+ = 1$ and $P_\sigma^- = 1 - \frac{1-\sqrt{1-8hc}}{2h}$.

Proof. Suppose there exists an equilibrium with positive adoption and $c > \frac{h}{2}$. Consider the corresponding threshold $\overline{v}$. If $\overline{v} \geq \frac{1}{2}$, this implies that

$$\frac{h}{2\overline{v}} < hP_\sigma^+ + (1-h)P_\sigma^- = h(2 - 2\overline{v}),$$

or that

$$2\overline{v} + \frac{1}{2\overline{v}} < 2,$$

which is impossible. If $\overline{v} < \frac{1}{2}$, it implies

$$\frac{h}{2\overline{v}} < (1-h)P_\sigma^+ + hP_\sigma^- = 1 - h + h(1 - 2\overline{v}) = 1 - 2h\overline{v}.$$

This can only hold if $2\overline{v} + \frac{1}{2\overline{v}} < \frac{1}{h} < 2$, and we conclude that no one adopts in equilibrium.

Now suppose $\frac{1}{8h} < c \leq \frac{h}{2}$. I first show that an equilibrium exists with $P_\sigma^+ = 1$ and $P_\sigma^- = 0$. This corresponds to a threshold $\overline{v} = \frac{1}{2}$, and for this to be a best reply, we must have

$$(1-h)P_\sigma^+ + hP_\sigma^- = 1 - h \leq 2c \leq h = hP_\sigma^+(1-h)P_\sigma^-.$$

Observe that $\frac{1}{4h} \geq 1 - h$ with $h > 0$ is equivalent to $4h^2 - 4h + 1 = (2h-1)^2 \geq 0$, so $P_\sigma^+ = 1$ and $P_\sigma^- = 0$ constitutes an equilibrium when c is in this range.

To finish part (b), we need to show that no equilibrium exists with $P_\sigma^- > 0$ when $c > \frac{1}{8h}$. If such an equilibrium exists, there is $\overline{v} < \frac{1}{2}$ satisfying

$$\frac{1}{8h\overline{v}} < 1 - 2h\overline{v}.$$

This is equivalent to $0 > 1 - 8h\overline{v} + 16(h\overline{v})^2 = (4h\overline{v} - 1)^2$, which is never true.

For part (c), suppose we have an equilibrium with $\overline{v} < \frac{1}{2}$. This implies

$$\frac{c}{\overline{v}} = 1 - 2h\overline{v},$$

which we can solve for

$$\overline{v} = \frac{1 \pm \sqrt{1 - 8hc}}{4h}.$$

For the solution to be defined, we must have $c \leq \frac{1}{8h}$ as we would expect from part (b). The smaller root will correspond to a maximal equilibrium with

$$P_\sigma^- = 1 - 2\overline{v} = 1 - \frac{1 - \sqrt{8hc}}{2h}$$

as desired. □

Homophily exacerbates the difference in preferences because it reinforces complementarities for high types while dampening this effect for low types. As h increases, the interval $\left(\frac{1}{8h}, \frac{h}{2}\right)$ expands on both sides. This increases the cost that high types are willing to pay, and it reduces the cost at which low types start adopting. In the region $c \leq \frac{1}{8h}$, differentiating (c) shows that P_σ^- is strictly decreasing in h. At any fixed c, adoption among high types

is weakly increasing in h, and adoption among low types is weakly decreasing in h. This suggests that homophily tends to limit adoption to niche communities with high valuations.[15]

Changing h changes the network of potential adopters apart from the effects on equilibrium strategies. In this particular example, the forward distribution in the branching process is independent of the parent's type. The forward degree distribution for both types is D', and each link connects to an own-type neighbor with probability h. As a result, we can collapse the four types in $\mathcal{T}_\sigma$ into two types. In the network of potential adopters, the average forward degree tuple for a low-type player is $(\nu_D h P_\sigma^-, \nu_D(1-h)P_\sigma^+)$, and for a high-type player it is $(\nu_D(1-h)P_\sigma^-, \nu_D h P_\sigma^+)$. The mean offspring matrix is

$$M_\sigma = \nu_D \begin{pmatrix} hP_\sigma^- & (1-h)P_\sigma^+ \\ (1-h)P_\sigma^- & hP_\sigma^+ \end{pmatrix}, \tag{5}$$

with spectral radius

$$\nu_\sigma = \frac{\nu_D}{2}\left(h(P_\sigma^- + P_\sigma^+) + \sqrt{h^2(P_\sigma^- + P_\sigma^+)^2 - 4(2h-1)P_\sigma^- P_\sigma^+}\right).$$

From this we can derive comparative statics in homophily.

Proposition 6. *Holding σ fixed, the parameter ν_σ is increasing in h.*

Proof. We compute

$$\frac{\partial \nu_\sigma}{\partial h} = \frac{\nu_D}{2}\left(P_\sigma^- + P_\sigma^+ + \frac{h(P_\sigma^- + P_\sigma^+)^2 - 4P_\sigma^- P_\sigma^+}{\sqrt{h^2(P_\sigma^- + P_\sigma^+)^2 - 4(2h-1)P_\sigma^- P_\sigma^+}}\right)$$

This non-negative if

$$(P_\sigma^- + P_\sigma^+)^2\left(h^2(P_\sigma^- + P_\sigma^+)^2 - 4(2h-1)P_\sigma^- P_\sigma^+\right) \geq \left(4P_\sigma^- P_\sigma^+ - h(P_\sigma^- + P_\sigma^+)^2\right)^2$$

or

$$4P_\sigma^- P_\sigma^+(P_\sigma^- + P_\sigma^+)^2 - 16(P_\sigma^- P_\sigma^+)^2 = 4P_\sigma^- P_\sigma^+\left(P_\sigma^+ - P_\sigma^-\right)^2 \geq 0,$$

implying the result. □

When we increase homophily, the typical potential adopter has more neighbors who are themselves potential adopters. Homophily therefore makes diffusion faster and makes large cascades more likely. This insight is independent of the valuation distribution and the payoff function, so long as $P_\sigma^+ > P_\sigma^-$. The effect is similar to that of a mean-preserving spread in the degree distribution. The network of potential adopters becomes more centralized with denser connections among high types and sparser connections among low types. When we do get a large cascade, the effect on the extent of diffusion is ambiguous.

[15] We find a similar pattern with more general payoff functions. Since $P_\sigma^+ \geq P_\sigma^-$ in the maximal equilibrium, an increase in homophily will tend to increase the neighbor adoption probability for high-type players and decrease it for low-type ones. However, there may be a countervailing effect. If the increase in P_σ^+ is sufficiently great, it may induce more low-type players to adopt; likewise if a decrease in P_σ^- is sufficiently great, it may cause fewer high-type players to adopt. Therefore, it is possible for both P_σ^+ and P_σ^- to increase (decrease) together, but the corresponding shift in the neighbor adoption probability is greater for high-type (low-type) players.

5 Applications

5.1 Bertrand Competition with Word-of-Mouth Communcation

Two firms sell identical products and have the same marginal cost c. Potential consumers are initially unaware of the products, but they learn about a firm's product if a neighbor purchases it. Each consumer has the same value for both products, drawn uniformly on $[0, 1]$. If a firm charges p, a purchasing consumer earns $v - p$. Consumers are myopic: as soon as one learns of a product offering positive utility, she purchases. I assume that both firms choose prices low enough that a giant component of potential adopters exists, and each firm has one random seed in the giant component who is aware of its product. Campbell (2013) studys a monopolist's pricing policy in this setting. The second part of Theorem 1 allows us to proceed a step further and consider the effects of competition.

Proposition 7. *In the equilibrium between the firms, for any $\epsilon > 0$ we have*

$$\lim_{n\to\infty} \mathbb{P}\left(p_1 \geq c + \epsilon\right) = \lim_{n\to\infty} \mathbb{P}\left(p_2 \geq c + \epsilon\right) = 0.$$

Proof. Consider demand for each product when the firms choose prices $p_1 < p_2$. The network of potential adopters for firm 1 contains that for firm 2. If D is the degree distribution, the typical distance between potential adopters of product 1 is $\log_{(1-p_1)\nu_D} n$, and for firm 2 it is $\log_{(1-p_2)\nu_D} n > \log_{(1-p_1)\nu_D} n$. Theorem 1 implies that for large n, essentially all potential adopters will learn of product 1 first, so the demand for firm 2 converges to zero as n grows. The standard argument characterizing equilibrium for a game of Bertrand competition implies the result. □

The proof of Proposition 7 offers more insight than the result. When consumers learn about products through word-of-mouth, the size and structure of the *network among consumers* has a direct effect on the level of *competition between firms*. A firm can earn profits when pricing higher than the competition because some consumers are unaware of the alternative. However, setting a low price helps more consumers learn about a product faster. In a small network, a firm loses relatively less from pricing above its competition because all purchases happen close to an initial seed. Differences in diffusion rates do not have time to create significant differences in demand. In a large network, the high priced firm loses out because the vast majority of the population learns about the lower priced alternative first.

The structure of the network, expressed through the average forward degree ν_D, also influences competitive prices. When $p_1 < p_2$, the difference in typical distances is

$$\log_{(1-p_2)\nu_D} n - \log_{(1-p_1)\nu_D} n = \frac{\ln n}{\ln(1-p_2)\nu_D} - \frac{\ln n}{\ln(1-p_1)\nu_D}.$$

Since the natural logarithm is concave, this difference shrinks with ν_D, indicating that the pressure to lower prices is weaker in a dense network. Intuitively, when density is high, more diffusion takes place close to an initial seed, so again differences in rates matter less.

5.2 Referral Marketing

Consider a monopolist selling a good to a networked population. However, when a consumer makes a purchase, her neighbors do not automatically become aware of the product. Instead, the consumer has the option to send referrals to each of her neighbors, and the firm can offer a reward for each referral that results in a purchase. Assume making referrals is costless, so as long as the firm offers a positive reward, consumers make all possible referrals. Consumer valuations are uniform, and the degree distribution takes one of two values $\underline{d} < \overline{d}$ with equal probability. A single seed is initially aware of the product, and consumers follow the equilibrium from section 3.1.

A referral program allows the monopolist to price discriminate between consumers based on the number of neighbors they can refer. The monopolist chooses a price p and a referral reward r. Since high-degree consumers have more referrals they can send, and they tend to learn about the product earlier, they can extract a larger reward from the referral program. Through an appropriate choice of p and r, the firm can charge any pair of effective prices $(\underline{p}, \overline{p})$ to low and high-degree consumers respectively, as long as $\overline{p} < \underline{p}$. If the firm could perfectly discriminate, it would choose to charge a lower price to high-degree consumers, so the referral program is as good as perfect price discrimination.

Proposition 8. *For any pair of effective prices $(\underline{p}, \overline{p})$ with $\overline{p} \leq \underline{p}$ there exists a price and referral reward pair (p, r) that implements them. The optimal pair of effective prices $(\underline{p}, \overline{p})$ satisfies $\overline{p} \leq \underline{p}$.*

Proof. See Appendix □

Lobel et al. (2015) study a similar referral design problem, allowing more general degree distributions, referral payments, and costs of making referrals. However, to make their analysis tractable, the authors abstract away from the dynamics of the process. Proposition 8 shows that key insights carry over when we fully account for the dynamics: referral programs provide a mechanism for price discrimination, and we generally wish to offer lower effective prices to those who can make more referrals. The latter result is more subtle than it first appears. High-degree consumers make more referrals than low-degree ones, but they are also disproportionately represented among those who become aware of the product. Lowering $\overline{p}$ adds more referrals than a similar decrease in $\underline{p}$, but it is also more costly since the discount applies to a larger fraction of customers. Proposition 8 shows that the first effect is more important for pricing decisions.

6 Remarks

Recent mathematical advances provide tools to study diffusion without making mean-field assumptions. In a large class of random graph models—a class that can capture realistic network features—a branching process closely approximates the local structure. This allows us to explore in detail how the network and individual decisions affect diffusion paths. We

can study the randomness in outcomes, understand when a product can "go viral," and relate both the extent and rate of spread to network properties.

The diffusion patterns we found are quite different from what mean-field models predict. Outcomes are inherently stochastic. In generic instances of the model, many individuals never learn about the product. To mimic this, a mean-field model must assume that certain groups never interact, which eliminates the network's role in this phenomenon. In this model, the probability of a large cascade, the extent of adoption, and the set of people who remain uninformed are all tied to the underlying degree distribution. Because we model the discrete network structure, we can make more detailed statements about diffusion paths. Early stages provide information about future growth, and we can replicate the variance in cascade sizes that empirical work documents.

Beyond these differences, the results shed light on how networks interact with economic decisions. A centralized network, or one with more homophily, results in a higher chance for large cascades and a faster rate of spread. As technology makes it easier to sort ourselves into homophilous groups, this suggests an increased incidence of viral phenomena. The Bertrand competition example highlights a mechanism through which the consumer network can affect competition between firms. A large, sparsely connected population of consumers results in the lowest competitive prices.

The key technical contribution should prove useful beyond studies of diffusion. Research on the economics of networks more broadly can benefit from a tractable random graph model that generates realistic networks. These techniques may also aid a closer relationship between theory and empirical work. Explicitly modeling link structure makes it easier to interpret theoretical models. The primitives correspond to things we can measure, and given recent work on methods to estimate random graph models (Chandrasekhar and Jackson, 2015), this raises the possibility of calibration using empirical data.

References

Athreya, Krishna and Peter Ney (1972), *Branching Processes*. Springer-Verlag, New York.

Banerjee, Abhijit, Arun Chandrasekhar, Esther Duflo, and Matthew Jackson (2013), "The Diffusion of Microfinance." *Science*, 341.

Banerjee, Abhijit, Arun Chandrasekhar, Esther Duflo, and Matthew Jackson (2016), "Gossip: Identifying Central Individuals in a Social Network." NBER Working Paper No. 20422.

Bollobás, B. (1980), "A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs." *European Journal of Combinatorics*, 1, 311–316.

Bollobás, Béla and Oliver Riordan (2015), "An Old Approach to the Giant Component Problem." *Journal of Combinatorial Theory, Series B*.

Campbell, Arthur (2013), "Word-of-Mouth Communication and Percolation in Social Networks." *American Economic Review*, 103, 2466–2498.

Caselli, Francesco and Wilbur Coleman (2001), "Cross-Country Technology Diffusion: The Case of Computers." *American Economic Review*, 91, 328–335.

Chandrasekhar, Arun and Matthew Jackson (2015), "A Network Formation Model Based on Subgraphs." Working Paper.

Cheng, Justin, Lada Adamic, P. Alex Dow, Jon Kleinberg, and Jure Leskovec (2014), "Can Cascades be Predicted?" In *Proceedings of the 23rd International World Wide Web Conference*, 925–936.

Feick, Lawrence and Linda Price (1987), "The Market Maven: A Diffuser of Marketplace Information." *Journal of Marketing*, 51, 83–97.

Galeotti, Andrea and Sanjeev Goyal (2009), "A Theory of Strategic Diffusion." *Rand Journal of Economics*, 40, 509–532.

Galeotti, Andrea, Sanjeev Goyal, Matthew Jackson, Fernando Vega-Redondo, and Leeat Yariv (2010), "Network Games." *Review of Economic Studies*, 77, 218–244.

Galeotti, Andrea and Brian Rogers (2012), "Strategic Immunization and Group Structure." *American Economic Journal: Microeconomics*, 5, 1–32.

Goel, Sharad, Duncan Watts, and Daniel Goldstein (2012), "The Structure of Online Diffusion Networks." In *Proceedings of the 13th ACM Conference on Electronic Commerce*, 623–638.

Griliches, Zvi (1957), "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica*, 25, 501–522.

Jackson, M. and L. Yariv (2007), "Diffusion of Behavior and Equilibrium Properties in Network Games." *American Economic Review*, 97, 92–98.

Jackson, Matthew and Dunia López-Pintado (2013), "Diffusion and Contagion in Networks with Heterogeneous Agents and Homophily." *Network Science*, 1, 49–67.

Jackson, Matthew and Brian Rogers (2007), "Relating Network Structure to Diffusion Properties through Stochastic Dominance." *The B.E. Journal of Theoretical Economics*, 7.

Jagers, Peter (1975), *Branching Processes with Biological Applications*. Wiley-Interscience, London.

Janson, Svante (2010), "Asymptotic Equivalence and Contiguity of Some Random Graphs." *Random Structures and Algorithms*, 36, 26–45.

Jencks, Stephen, Edwin Huff, and Timothy Cuerdon (2003), "Change in the Quality of Care Delivered to Medicare Beneficiaries, 1998-99 to 2000-2001." *JAMA*, 289, 305–312.

Leskovec, Jure, Ajit Singh, and Jon Kleinberg (2006), "Patterns of Influence in a Recommendation Network." In *Proceedings of the 10th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, 380–389.

Lobel, Ilan, Evan Sadler, and Lav R. Varshney (2015), "Customer Referral Incentives and Social Media." Forthcoming in Management Science.

López-Pintado, Dunia (2008), "Diffusion in Complex Social Networks." *Games and Economic Behavior*, 62, 573–590.

McPherson, M., L. Smith-Lovin, and J. Cook (2001), "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology*, 27, 415–444.

Molloy, Michael and Bruce Reed (1995), "A Critical Point for Random Graphs with a Given Degree Sequence." *Random Structures and Algorithms*, 6, 161–179.

Morris, Stephen (2000), "Contagion." *Review of Economic Studies*, 67, 57–78.

Skinner, Jonathan and Douglas Staiger (2005), "Technology Adoption from Hybrid Corn to Beta Blockers." NBER Working Paper No. 11251.

Sundararajan, Arun (2007), "Local Network Effects and Complex Network Structure." *The B. E. Journal of Theoretical Economics*, 7.

Tucker, Catherine (2008), "Identifying Formal and Informal Influence in Technology Adoption with Network Externalities." *Management Science*, 55, 2024–2039.

van der Hofstad, Remco, Gerard Hooghiemstra, and Piet Van Mieghem (2005), "Distances in Random Graphs with Finite Variance Degrees." *Random Structures Algorithms*, 27, 76–123.

van der Hofstad, Remco, Gerard Hooghiemstra, and Dmitri Znamenski (2007), "Distances in Random Graphs with Finite Mean and Infinite Variance Degrees." *Electronic Journal of Probability*, 12, 703–766.

Watts, Duncan (2002), "A Simple Model of Global Cascades on Random Networks." *Proceedings of the National Academy of Sciences*, 99, 5766–5771.

Young, Peyton (2009), "Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning." *American Economic Review*, 99, 1899–1924.

A Appendix

A.1 Branching Processes

A Galton-Watson process is a sequence of random variables $\{Z_n\}_{n=0}^{\infty}$, with $Z_0 = 1$ by convention, and

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_{n,i},$$

where $\{X_{n,i}\}_{n,i\in\mathbb{N}}$ is a collection of i.i.d. random variables taking non-negative integer values. Write X for the common distribution. We can interpret the sequence $\{Z_n\}$ as a population growing and shrinking over time, with Z_n the number of individuals in the nth generation. To obtain the next generation, each member of the nth generation has a random number of offspring, generated according to the distribution X.

Standard questions about the Galton-Watson process include: What is the probability of extinction $\eta = \mathbb{P}(\exists n : Z_n = 0)$? What is the distribution of the total population size? How fast does Z_n grow over time? A fundamental tool in the analysis of branching processes is the generating function of X. Write $p_i = \mathbb{P}(X = i)$ for the probability that an individual has exactly i offspring. The generating function is

$$G_X(s) = \mathbb{E}[s^X] = \sum_{i=0}^{\infty} p_i s^i.$$

The following standard result uses the generating function to characterize the extinction probability.

Proposition 9. *The extinction probability η is the smallest solution in $[0,1]$ of*

$$\eta = G_X(\eta).$$

In particular, if $\mathbb{E}[X] < 1$, then $\eta = 1$; if $\mathbb{E}[X] > 1$, then $\eta < 1$. If $\mathbb{E}[X] = 1$, then $\eta = 1$ if $p_1 < 1$ and $\eta = 0$ if $p_1 = 1$.

Proof. See the first chapter in Athreya and Ney (1972). □

Let $\mu = \mathbb{E}[X]$. One can easily verify that $\mathbb{E}[Z_n] = \mu^n$, and the sequence $\frac{Z_n}{\mu^n}$ forms a martingale. This gives us additional information about the size of the branching process, the rate of growth, and implicitly the distribution of the extinction time.

Proposition 10. *Let $T = \sum_{n=0}^{\infty} Z_n$ denote the total progeny of the branching process. If $\mu < 1$, then $\mathbb{E}[T] = \frac{1}{1-\mu}$.*

The sequence $\frac{Z_n}{\mu^n}$ converges almost surely to a non-negative random variable W.

Proof. The first claim follows by summing the geometric sequence of expected values; the second is immediate from the martingale convergence theorem. □

We will require generalizations of these results for multi-type branching processes. Suppose there are r types of individuals in the population. The random variable $\mathbf{Z}_n$ denotes an r-tuple $(Z_{n,1}, Z_{n,2}, ..., Z_{n,r})$, giving the number of individuals of each type in the nth generation. Given an initial population $\mathbf{Z}_0$, we recursively define the sequence $\mathbf{Z}_n$ as

$$\mathbf{Z}_n = \sum_{i=1}^{r} \sum_{j=1}^{Z_{n-1,i}} \mathbf{X}_j^{(i)},$$

where the $\mathbf{X}_j^{(i)}$ are mutually independent, and for each i the $\mathbf{X}_j^{(i)}$ have the same distribution $\mathbf{X}^{(i)}$. Put differently, each type is characterized by its own offspring distribution, where the distribution $\mathbf{X}^{(i)}$ is a distribution over r-tuples of non-negative integers.

We can define an analogous generating function for the multi-type branching process. Let $p^{(i)}(j_1, j_2, ..., j_r)$ denote the probability that $X^{(i)} = (j_1, j_2, ..., j_r)$, and define

$$G_{\mathbf{X}}^{(i)}(\mathbf{s}) = \sum_{(j_1,j_2,...,j_r)\in\mathbb{N}^r} p^{(i)}(j_1, j_2, ..., j_r) s_1^{j_1} s_2^{j_2} ... s_r^{j_r}.$$

The multi-type generating function is the vector $\mathbf{G}_{\mathbf{X}}(\mathbf{s}) = \left(G_{\mathbf{X}}^{(1)}(\mathbf{s}), G_{\mathbf{X}}^{(2)}(\mathbf{s}), ..., G_{\mathbf{X}}^{(r)}(\mathbf{s})\right)$. Let $\eta^{(i)}$ denote the probability of extinction if $\mathbf{Z}_0$ consists of a single individual of type i, and let $\eta = (\eta^{(1)}, \eta^{(2)}, ..., \eta^{(r)})$ denote the vector of extinction probabilities.

In the single type case, we used $\mu = \mathbb{E}[X]$ to characterize whether the branching process goes extinct with probability one. In the multi-type case, we can write an analogous result using the mean offspring matrix. Define $m_{i,j} = \mathbb{E}[\mathbf{X}_j^{(i)}]$ as the expected number of type j children from a type i parent. The mean offspring matrix M is simply the matrix with entries $m_{i,j}$; let ρ denote its spectral radius. We say that a multi-type branching process is *irreducible* if every type of individual has descendants of all other types with positive probability. This is equivalent to M being an irreducible matrix. We say that a multi-type branching process is *non-singular* is there exists a type that does not have a single offspring with probability one.

Proposition 11. *The vector η is the only solution in the unit cube of $\mathbf{G}_{\mathbf{X}}(\eta) = \eta$.*

Suppose the branching process is irreducible and non-singular. If $\rho \leq 1$, then $\eta = \mathbf{1}$. If $\rho > 1$, then $\eta^{(i)} < 1$ for all i.

Proof. See chapter 5 in Athreya and Ney (1972). □

The spectral radius ρ of the mean offspring matrix also characterizes the growth rate of the branching process.

Proposition 12. *There exists a non-negative random vector $\mathbf{W}$ such that $\frac{\mathbf{Z}_n}{\rho^n}$ converges to $\mathbf{W}$ almost surely.*

Proof. Again, see chapter 5 in Athreya and Ney (1972). □

More detailed results on the distribution of T, the distribution of extinction times, and other features are available in the literature, but are beyond what is needed in the present paper. I would direct an interested reader to Athreya and Ney (1972) and Jagers (1975).

A.2 Configuration Models

Many basic results on the configuration model concern limits as n approaches infinity. For $CM_{n,D}$, these limits are well-defined. To make sense of a limit for a sequence $CM_{n,\mathbf{d}^{(n)}}$, the degree vectors $\mathbf{d}^{(n)}$ must converge in an appropriate sense. For an n-vector $\mathbf{d}^{(n)}$ of degrees, let $n_d(\mathbf{d}^{(n)})$ denote the number of entries equal to d, and let $m(\mathbf{d}^{(n)}) = \sum_{i=1}^{n} d_i^{(n)}$ denote the total number of stubs, or twice the number of edges. There are two standard conditions:

(a) There exists $\{p_d\}_{d\in\mathbb{N}}$ such that for each d we have

$$\lim_{n\to\infty} \frac{n_d(\mathbf{d}^{(n)})}{n} = p_d.$$

(b) We have

$$\lim_{n\to\infty} \frac{m(\mathbf{d}^{(n)})}{n} = \sum_{d=0}^{\infty} dp_d < \infty.$$

The sequence $\{p_d\}_{d\in\mathbb{N}}$ describes a limiting degree distribution that takes the value d with probability p_d. These two conditions ensure that $\mathbf{d}^{(n)}$ converges to $\{p_d\}$ in distribution and in expectation.[16] For a sequence $\{\mathbf{d}^{(n)}\}_{n\in\mathbb{N}}$, we always assume that (a) and (b) hold, and we write D for the limiting distribution.

The model $CM_{n,\mathbf{d}}$ may realize a multigraph: it can have self link and multiple links between the same pair of vertices. If there are no self loops, and each pair of nodes has at most one link between them, the graph is *simple*. When we study diffusion, we condition the configuration model on realizing a simple graph. This is equivalent to taking a uniform draw among all graphs with the given degree sequence.

Proposition 13. *Conditional on realizing a simple graph, the model $CM_{n,\mathbf{d}}$ selects a graph uniformly at random from those with n verticies and degree sequence $\mathbf{d}$.*

Proof. This is immediate from the definition as each possible pairing of link stubs that results in a simple graph is equally likely. □

A recurring challenge is translating results for the configuration model to simple graphs. We often wish to show that some function of graphs $f(G)$ concentrates around its mean μ as n grows. Independent link formation makes it relatively easy to show for any $\epsilon > 0$ that

$$\lim_{n\to\infty} \mathbb{P}_{CM_{n,\mathbf{d}}} \left(|f(G) - \mu| \geq \epsilon \right) = 0, \tag{6}$$

[16] At first glance, the second condition might appear redundant, but it is necessary to rule out pathological cases. For instance, suppose $\mathbf{d}^{(n)}$ contains $n-1$ entries equal to 1 and a single entry equal to $n-1$. The sequence converges in distribution to a random variable taking the value 1 with probability 1, which would suggest $\frac{n}{2}$ edges in expectation, but the actual number of edges is always $n-1$.

but we need to show that

$$\lim_{n\to\infty} \mathbb{P}_{CM_{n,\mathbf{d}}} \left(|f(G) - \mu| \geq \epsilon \,\middle|\, G \text{ is simple} \right) = 0. \tag{7}$$

Since the probability of realizing a simple graph declines subexponentially in n, if we can establish exponential concentration bounds on (6), the bounds translate directly to (7).

Proposition 14. *Fix any $\gamma > 0$. For all sufficiently large n we have*

$$\mathbb{P}_{CM_{n,\mathbf{d}^{(n)}}} \left(G \text{ is simple} \right) > e^{-\gamma n}.$$

Proof. This is equivalent to Lemma 21 of Bollobás and Riordan (2015). □

Corollary 3. *Suppose that for any $\epsilon > 0$, there exists δ such that for all sufficiently large n we have*

$$\mathbb{P}_{CM_{n,\mathbf{d}}} \left(|f(G) - \mu| \geq \epsilon \right) < e^{-\delta n}.$$

Then, for any $\epsilon > 0$, there exists δ' such that for all sufficiently large n we have

$$\mathbb{P}_{CM_{n,\mathbf{d}^{(n)}}} \left(|f(G) - \mu| \geq \epsilon \,\middle|\, G \text{ is simple} \right) < e^{-\delta' n}.$$

Proof. Choose $\gamma = \frac{\delta}{2}$ in Proposition 14. We have for large n

$$\begin{aligned} \mathbb{P}_{CM_{n,\mathbf{d}^{(n)}}} \left(|f(G) - \mu| \geq \epsilon \,\middle|\, G \text{ is simple} \right) &\leq \frac{\mathbb{P}_{CM_{n,\mathbf{d}^{(n)}}} \left(|f(G) - \mu| \geq \epsilon \right)}{\mathbb{P}_{CM_{n,\mathbf{d}^{(n)}}} \left(G \text{ is simple} \right)} \\ &< \frac{e^{-\delta n}}{e^{-\frac{\delta}{2} n}} = e^{-\frac{\delta}{2} n}. \end{aligned}$$

□

Corollary 3 allows us to show that conditioning on simple graphs does not distort the limiting degree distribution.

Proposition 15. *Let $n_d(G)$ denote the number of degree d vertices in the graph G and let $p_d = \mathbb{P}(D = d)$. For any $\epsilon > 0$, there exists $\delta > 0$ such that*

$$\mathbb{P}_{CM_{n,D}} \left(\sup_d \left| \frac{n_d(G)}{n} - p_d \right| \geq \epsilon \,\middle|\, G \text{ is simple} \right) < e^{-\delta n}.$$

Proof. The corresponding inequality without conditioning on simple G is immediate from the Dvoretzky-Kiefer-Wolfowitz inequality. The result then follows from Corollary 3. □

Moving to the multi-type configuration model, we require similar conditions on the sequence $(\boldsymbol{\theta}, \mathbf{d})^{(n)}$ to state limit results. Define $r_\theta = \mathbb{P}(T = \theta)$ and $p_{\theta,d} = \mathbb{P}(D = d \,|\, T = \theta)$. Let $\mu_{\theta,\theta'}$ denote the expected number of type θ' neighbors of a type θ vertex according to (T, D). Given a sequence $(\boldsymbol{\theta}, \mathbf{d})^{(n)}$, let $n_{\theta,d}\left((\boldsymbol{\theta}, \mathbf{d})^{(n)}\right)$ denote the number of type θ entries with corresponding degree tuple d, and let $m_{\theta,\theta'}\left((\boldsymbol{\theta}, \mathbf{d})^{(n)}\right)$ denote the number of edges between type θ and θ' nodes. The analogous conditions are

(a)

$$\lim_{n\to\infty} \frac{n_{\theta,d}\left((\boldsymbol{\theta},\mathbf{d})^{(n)}\right)}{n} = r_\theta p_{\theta,d}.$$

(b)

$$\lim_{n\to\infty} \frac{m_{\theta,\theta'}\left((\boldsymbol{\theta},\mathbf{d})^{(n)}\right)}{n} = r_\theta \mu_{\theta,\theta'} < \infty.$$

Observe for any pair of types θ and θ', we must have $r_\theta \mu_{\theta,\theta'} = r_{\theta'}\mu_{\theta',\theta}$. We often also impose an irreducibility condition: for any pair of types θ and θ', there exists some sequence of types $\theta_0, \theta_1, ..., \theta_l$ such that

- $\theta_0 = \theta$ and $\theta_l = \theta'$
- $\prod_{i=1}^{l} \mu_{\theta_{i-1},\theta_i} > 0$.

This ensures that a path can exist between any two types with positive probability. Direct analogs to Propositions 1, 13, 14, and 15 , and Corollary 3 hold via identical arguments.

A.3 Proofs of Technical Results

Proof of Proposition 1

From the construction of $D_{\mathbf{q}}$, we can compute the probabilities $p_{\mathbf{q},d}$ as

$$p_{\mathbf{q},d} = \sum_{k\geq d} \frac{q_k p_k}{\sum_{i=0}^{\infty} q_i p_i} \binom{k}{d} \left(\frac{\sum_{i=0}^{\infty} i q_i p_i}{\sum_{i=0}^{\infty} i p_i}\right)^d \left(\frac{\sum_{i=0}^{\infty} i(1-q_i) p_i}{\sum_{i=0}^{\infty} i p_i}\right)^{k-d}.$$

I first show that $\mathbb{E}_{CM_{n,\mathbf{q},D}}\left[\frac{n_d(G)}{|S|}\right]$ converges to $p_{\mathbf{q},d}$. Let $n_{d,\mathbf{q}}(G)$ denote the number of degree d nodes in $CM_{n,D}$ that get selected according to $\mathbf{q}$, let $m_{\mathbf{q}}(G)$ denote the number of link stubs attached to nodes selected according to $\mathbf{q}$, and let $m(G)$ denote the total number of link stubs. The Dvoretzky-Kiefer-Wolfowitz inequality implies that for any $\epsilon > 0$ there exists $\delta > 0$ such that

$$\mathbb{P}_{CM_{n,D}}\left(\sup_d \left|\frac{n_{d,\mathbf{q}}(G)}{|S|} - \frac{q_d p_d}{\sum_{i=0}^{\infty} q_i p_i}\right| \geq \epsilon\right) < \frac{1}{3} e^{-\delta n}.$$

An application of Hoeffding's inequality ensures we can choose this δ so that also

$$\mathbb{P}_{CM_{n,D}}\left(\left|\frac{m_{\mathbf{q}}(G)}{n} - \sum_{i=0}^{\infty} i q_i p_i\right| \geq \frac{\epsilon}{2}\right) < \frac{1}{3} e^{-\delta n}, \text{ and } \mathbb{P}_{CM_{n,D}}\left(\left|\frac{m(G)}{n} - \sum_{i=0}^{\infty} i p_i\right| \geq \frac{\epsilon}{2}\right) < \frac{1}{3} e^{-\delta n}.$$

Write A for the event that none of the three inequalities hold, which happens with probability at least $1 - e^{-\delta n}$.

The expected fraction of nodes in $CM_{n,\mathbf{q},D}$ with degree d is the probability that a random vertex in S has degree d. Write D_v for the degree of a vertex in $CM_{n,\mathbf{q},D}$ chosen uniformly

at random, and D_v^+ for the degree of this vertex in the graph $CM_{n,D}$ before removing the vertices that were not selected. Also, write $\pi_{d,k}$ for the set of vectors in $\{0,1\}^k$ with exactly d entries equal to 1. We have

$$\mathbb{P}(D_v = d) = \sum_{k \geq d} \mathbb{P}(D_v^+ = k) \mathbb{E}\left[\sum_{\pi \in \pi_{d,k}} \frac{\prod_{i=0}^{k-1} (m_{\mathbf{q}}(G) - i)^{\pi_i} (m(G) - m_{\mathbf{q}}(G) - i)^{1-\pi_i}}{\prod_{i=0}^{k-1} (m(G) - i)} \right].$$

In event A, with n sufficiently large, we have

$$\left| \mathbb{P}(D_v^+ = k) - \frac{q_k p_k}{\sum_{i=0}^{\infty} q_i p_i} \right| < \epsilon,$$

$$\left| \frac{\prod_{i=0}^{k-1} (m_{\mathbf{q}}(G) - i)^{\pi_i} (m(G) - m_{\mathbf{q}}(G) - i)^{1-\pi_i}}{n^k} - \left(\sum_{i=0}^{\infty} i q_i p_i \right)^d \left(\sum_{i=0}^{\infty} i (1 - q_i) p_i \right)^{k-d} \right|$$
$$< (1+\epsilon)^k - 1, \text{ and}$$

$$\left| \frac{\prod_{i=0}^{k-1} (m(G) - i)}{n^k} - \left(\sum_{i=0}^{\infty} i p_i \right)^k \right| < (1+\epsilon)^k - 1.$$

Choosing ϵ sufficiently small, and letting n tend to infinity, we see that $\mathbb{P}(D_v = d)$ must converge to $p_{\mathbf{q},d}$.

To complete the argument, we need to show that $\frac{n_d(G)}{|S|}$ concentrates around its mean. To see this, we define a sequence of random variables that slowly reveal the realized value of $\frac{n_d(G)}{|S|}$. Given a realization of $CM_{n,D}$, let X_0 denote the expected value of $\frac{n_d(G)}{|S|}$. Label the vertices 1 through n, and one vertex at a time, reveal whether that vertex is selected according to q; let X_i for $i = 1, 2, ..., n$ denote the conditional expectation of $\frac{n_d(G)}{|S|}$ after the ith revelation. By definition, $X_n = \frac{n_d(G)}{|S|}$, and the sequence X_i is a martingale with bounded increments. The Azuma-Hoeffding inequality finishes the proof. □

Proof of Theorem 1

For the first part of Theorem 1, I prove a stronger result which immediately implies it. To state the result, I must define a few terms and notation. I use $\mathbf{d} = \{(d_1, d_2, ..., d_\Theta)_{i,\theta_i}\}_{i=1}^N$ to denote the sequence of realized types and degree tuples. Fixing $\mathbf{d}$, I write $m_{\theta,\theta'}(\mathbf{d})$ for the number of edges connecting type θ nodes and type θ' nodes, and I write $n_{\theta,d}(\mathbf{d})$ for the number of type θ nodes with degree tuple d. Finally, I define the *configuration distance* as

$$l(\mathbf{d}, D) = \max \left\{ \frac{1}{N}, \sum_{\theta=1}^{\Theta} \sum_{d} \left| \overline{d} \frac{n_{\theta,d}(\mathbf{d})}{N} - \overline{d} p_\theta \mathbb{P}(D_\theta = d) \right| \right\},$$

where $\overline{d} = \sum_{\theta=1}^{\Theta} d_\theta$ is the total degree associated with d. This is a measure of how much the realized degree sequence deviates from that prescribed by the distribution D. It is straightforward to check that $l(\mathbf{d}, D)$ converges to zero as N grows if and only if the realized degree sequence converges to D both in distribution and in expectation. Hence, the law of large numbers implies that the configuration distance converges to zero almost surely as N goes to infinity.

Theorem 3. *For any $\epsilon > 0$ and any $k \geq 1$, there exists $\delta > 0$ such that if $l(\mathbf{d}, D) < \delta$ we have*

$$\mathbb{P}\left(\left|N_k(G) - N\mathbb{P}(|\mathcal{T}| = k)\right| \geq \epsilon N\right) \leq e^{-\delta N}. \tag{8}$$

If additionally the degree distribution D assigns positive probability to some type having 3 or more neighbors, then there exists $\delta > 0$ such that if $l(\mathbf{d}, D) < \delta$ we have

$$\mathbb{P}\left(\left|L_1(G) - N\mathbb{P}(|\mathcal{T}| = \infty)\right| \geq \epsilon N\right) \leq e^{-\delta N}, \quad \textit{and} \quad \mathbb{P}\left(L_2(G) \geq \epsilon N\right) < e^{-\delta N}. \tag{9}$$

I prove Theorem 3 for a configuration model that allows multigraphs. That is, we generate the network as described, but we do not condition on realizing a simple graph: we permit self-links and multiple links between the same pair of nodes. This version of the configuration model is easier to work with because we can allow the stubs to be paired uniformly at random without any further conditioning. The multi-type analog of Corollary 3 implies the result also holds conditional on realizing a simple graph.

I will use G^* to denote a graph realized according to the multigrah configuration model, and I use G to denote a generic graph. I use v to denote a generic node in G, and G_v to denote the graph G rooted on v. I use $\mathcal{T}$ to refer both to the branching process defined in section 2 and the corresponding tree, viewed as a rooted graph. For a positive integer r, I write $G_{v,r}$ for the subgraph of radius r rooted on v (i.e. the graph comprising only nodes at distance r or less from v), and similarly I write $\mathcal{T}_r$ for the tree $\mathcal{T}$ truncated after the rth generation. I will occasionally abuse notation, writing $G_{v,r} = \mathcal{T}_r$ to indicate that $\mathcal{T}_r$, viewed as a rooted graph, is isomorphic to $G_{v,r}$.

There are two major steps in the proof. The first step is to show that the distribution of component sizes converges to the distribution of tree sizes generated by the branching process $\mathcal{T}$. We can prove this for the multigraph case using a straightforward coupling argument, matching the branching process with a breadth first search process starting at a random node. However, passing to simple graphs later requires a more powerful concentration result, giving exponential bounds on the rate of convergence. We obtain these bounds by applying the Azuma-Hoeffding inequality to a martingale that arises through a process that explores possible link stub pairings.

Once we establish the correspondence between component sizes and tree sizes, we show that essentially all "large" components are connected in one "giant" component. This relies on a coloring and sprinkling argument in which we first retain links independently with some probability $p \in (0, 1)$, and then sprinkle the remaining links back in, taking advantage of conditional independence between the retained links and the sprinkled links. Large components that exist in the thinned graph are likely to be connected by the additional links. The

assumption that at least one type has three or more neighbors with positive probability is necessary for this step. It ensures that the survival probability of the thinned tree converges to that of $\mathcal{T}$ as p approaches 1. An argument showing that the results carry over if we condition on realizing a simple graph completes the proof.

The Branching Process Approximation

The first part of the proof establishes a coupling between rooted graphs of finite size $G^*_{v,r}$ and truncated trees $\mathcal{T}_r$. This in turn implies that any property of the rooted graph G^*_v, which depends only on those vertices within distance r of v, is asymptotically characterized by the branching process $\mathcal{T}$. This is the sense in which $\mathcal{T}$ captures the "local" properties of G^*. The bulk of this section is devoted to proving bounds on the probability of deviations.

Lemma 1. *Let v be a vertex of G^* chosen uniformly at random, and suppose $\{\mathbf{d}_N\}_{N\in\mathbb{N}}$ is a sequence for which $l(\mathbf{d}_N, D)$ converges to zero. For any finite r, we can couple the random graphs $G^*_{v,r}$ and $\mathcal{T}_r$ so that they are isomorphic with probability approaching 1 as N approaches infinity.*

Proof. Begin with a realized sequence $\mathbf{d}$, and suppose that $l(\mathbf{d}, D) < \epsilon$ for some $\epsilon > 0$. We will reveal the rooted graph $G^*_{v,r}$ one node at a time, following a breadth first search procedure, coupling it with $\mathcal{T}_r$ at each step and bounding the probability that the coupling fails. Given our assumption on the configuration distance, we can couple the degree d of the root v with the offspring distribution of the root of $\mathcal{T}_r$ with probability at least $1-\epsilon$.

At each subsequent step, we start with a node of some type θ and reveal a partner for a link of another type θ'. At the jth step, the probability that this is an unvisited node with degree tuple d is precisely

$$\frac{d_\theta \left(n_{\theta',d}(\mathbf{d}) - u_{\theta',d,j}\right)}{m_{\theta,\theta'}(\mathbf{d}) - u_{\theta,\theta',j}},$$

where $u_{\theta',d,j}$ is the number of type θ' nodes with degree vector $\mathbf{d}$ that have been visited before the jth step, and $u_{\theta,\theta',j}$ is the number of completed edges between type θ and type θ' nodes before the jth step. Note that $u_{\theta',d,j} \leq j$ and $u_{\theta,\theta',j} \leq j$. This implies that for any fixed j, the difference between this quantity and $\mathbb{P}(D'_{\theta,\theta'} = d)$ is no more than $\epsilon + o(1)$, so the coupling succeeds with probability $1 - \epsilon - o(1)$.

To complete the proof, note that for any $\epsilon > 0$, there is a constant M such that $|\mathcal{T}_r| \leq M$ with probability at least $1-\epsilon$, and for sufficiently large N, we have $l(\mathbf{d}, D) < \epsilon$ with probability at least $1-\epsilon$. For N larger than this, the probability that the coupling fails is no more than $2\epsilon + M(\epsilon + o(1))$, and the conclusion follows. □

One immediate consequence of this lemma is that the rooted graphs $G^*_{v,r}$ are trees with probability approaching 1. More generally, the branching process $\mathcal{T}$ characterizes any "local" property of the graph G^*. Let $\mathcal{P}$ be a property of rooted graphs, meaning a set of rooted graphs that is closed under isomorphisms. We can also think of $\mathcal{P}$ as a property of vertices, taking the root of the graph as the relevant vertex. We write $G_v \in \mathcal{P}$ to say that the graph G rooted on v has the property $\mathcal{P}$ and we write $N_{\mathcal{P}}(G)$ for the number of vertices with property

$\mathcal{P}$. For any positive integer r, we say that $\mathcal{P}$ is *r-local* if whether $G_v \in \mathcal{P}$ depends only on $G_{v,r}$. The following corollary is immediate from Lemma 1

Corollary 4. *Let $\mathcal{P}$ be a r-local property of rooted graphs, let v be a vertex of G^* chosen uniformly at random, and suppose $\{\mathbf{d}_N\}_{N\in\mathbb{N}}$ is a sequence for which $l(\mathbf{d}_N, D)$ converges to zero. Then,*

$$\lim_{N\to\infty} \mathbb{P}\left(G_v^* \in \mathcal{P}\right) = \mathbb{P}\left(\mathcal{T} \in \mathcal{P}\right).$$

Equivalently, for any $\epsilon > 0$, there exists N_ϵ such that if $N \geq N_\epsilon$ we have

$$\left|\mathbb{E}[N_{\mathcal{P}}(G^*)] - N\mathbb{P}(\mathcal{T} \in \mathcal{P})\right| \leq \epsilon N.$$

We require a slightly modified version of this result, which follows from the previous corollary.

Corollary 5. *Let $\mathcal{P}$ be a r-local property of rooted graphs, and let v be a vertex of G^* chosen uniformly at random. For any $\epsilon > 0$, there exists $\delta > 0$ such that if $l(\mathbf{d}, D) < \delta$, then conditional on the degree sequence $\mathbf{d}$ we have*

$$\left|\mathbb{E}[N_{\mathcal{P}}(G^*)] - N\mathbb{P}(\mathcal{T} \in \mathcal{P})\right| \leq \epsilon N.$$

We focus on the k-local property $\mathcal{P}_k$ that a vertex is in a graph component with exactly k nodes, meaning

$$N_{\mathcal{P}_k}(G) = N_k(G), \quad \text{and} \quad \mathbb{P}(\mathcal{T} \in \mathcal{P}_k) = \mathbb{P}(|\mathcal{T}| = k).$$

Corollary 4 of course implies convergence of $\frac{N_k(G^*)}{N}$ to $\mathbb{P}(|\mathcal{T}| = k)$, but we require a stronger bound on the rate of convergence. We make repeated use of the following concentration result.

Proposition 16. *Let $\mathcal{P}$ be a r-local property of rooted graphs. For any $\epsilon > 0$, there exists $\delta > 0$ such that if $l(\mathbf{d}, D) < \delta$ then*

$$\mathbb{P}\left(\left|N_{\mathcal{P}}(G^*) - N\mathbb{P}(\mathcal{T} \in \mathcal{P})\right| \geq \epsilon N\right) \leq e^{-\delta N}.$$

The first step to obtain this bound is a lemma using the Azuma-Hoeffding inequality. Fixing a degree sequence $\mathbf{d}$, we can consider different pairings of stubs. We say that two pairings π_1 and π_2 are *related by a switching* if we can obtain π_2 from π_1 by deleting two pairs of the same type $\{a, b\}$ and $\{c, d\}$ and replacing them with the pairs $\{a, d\}$ and $\{c, b\}$. Let f be a real-valued function defined on pairings of $\mathbf{d}$. We say that f is *C-Lipschitz* if for any π_1 and π_2 related by a switching, we have $|f(\pi_1) - f(\pi_2)| \leq C$.

Lemma 2. *Let f be a C-Lipschitz function of pairings of some degree sequence $\mathbf{d}$, let M denote the total number of pairs. If π is chosen uniformly at random from all pairings of $\mathbf{d}$, then for any $r \geq 0$ we have*

$$\mathbb{P}\left(\left|f(\pi) - \mathbb{E}[f(\pi)]\right| \geq r\right) \leq 2e^{-\frac{r^2}{2C^2M}}.$$

Proof. Let $S_\theta^{\theta'} = \{s_1, s_2, ..., s_m\}$ denote the set of stubs leading from type θ nodes to type θ' nodes, with $S_{\theta'}^{\theta} = \{s'_1, s'_2, ..., s'_m\}$ the set of potential partners. We consider a random process in which we sequentially reveal the pairing. Conditional on the partners of $s_1, ..., s_i$, let Ω denote the set of pairings between $S_\theta^{\theta'}$ and $S_{\theta'}^{\theta}$ that are consistent with the information revealed so far. For any possible partner b of s_{i+1}, let Ω_b denote the subset of Ω containing all possible pairings in which s_{i+1} is matched to b. For any two potential partners b and c, there is a bijection between Ω_b and Ω_c in which each $\pi_1 \in \Omega_b$ is related by a switching to its image $\pi_2 \in \Omega_c$: just switch the pairs $\{s_{i+1}, b\}$ and $\{s_j, c\}$ to $\{s_{i+1}, c\}$ and $\{s_j, b\}$.

Iterate the revelation process over each type of link, and let $\mathcal{F}_i$ be the sigma-field generated by the sequential revelation process up to s_i. The process $X_i = \mathbb{E}[f(\pi) \,|\, \mathcal{F}_i]$ is clearly a martingale. The bijection together with the Lipschitz property implies that

$$\left|\mathbb{E}[f(\pi) \,|\, \mathcal{F}_i] - \mathbb{E}[f(\pi) \,|\, \mathcal{F}_{i+1}]\right| \leq C.$$

The sequence $\{X_i\}_{i=0}^{M}$, with $X_0 = \mathbb{E}[f(\pi)]$ and $X_M = f(\pi)$, is a martingale with differences bounded by C, and the result follows from the Azuma-Hoeffding inequality. □

This lemma is sufficient to prove our concentration result for a local property $\mathcal{P}$ if $N_{\mathcal{P}}(G)$ is C-Lipschitz for some C, but this is not universally true for all local properties. However, if we modify the property to avoid high-degree vertices, we can obtain a C-Lipschitz function of the graph and use it to prove the concentration bounds. For $\Delta \geq 2$ and $r \geq 0$, let $\mathcal{M}_{\Delta,r}$ be the property of rooted graphs that every node within distance r of the root has degree at most Δ. This is a $r+1$-local property.

Lemma 3. *Let $\mathcal{P}$ be a r-local property, and let $\mathcal{Q} = \mathcal{P} \cap \mathcal{M}_{\Delta,r}$. The number of vertices $N_{\mathcal{Q}}(G)$ with property $\mathcal{Q}$ is $16\Delta^r$-Lipschitz.*

Proof. Suppose v is a vertex of G such that exactly one of G_v and $(G+e)_v$ has property $\mathcal{Q}$, for some edge e. This implies that G_v has property $M_{\Delta,r}$ since removing an edge can only reduce the degree of a vertex. Suppose x and y are the endpoints of e. Since only one of G_v and $(G+e)_v$ has property $\mathcal{Q}$, one of x and y is connected through a path of length at most r to v in G, and each vertex along this path has degree at most Δ. For each endpoint of e, there can be at most $(1 + \Delta + ... + \Delta^r) \leq 2\Delta^r$ such paths, so adding or removing an edge can change the number of vertices with property $\mathcal{Q}$ by at most $4\Delta^r$. Since a switching corresponds to removing two edges and adding two edges, the result follows. □

The next lemma formalizes the idea that we can safely ignore high-degree vertices.

Lemma 4. *For any $r \geq 0$ and $\epsilon > 0$, there exist $\delta > 0$ and an integer Δ such that whenever $l(\mathbf{d}, D) < \delta$ we have*

$$\mathbb{P}(\mathcal{T} \in \mathcal{M}_{\Delta,r}) \geq 1 - \frac{\epsilon}{4}, \text{ and}$$

$$\mathbb{P}\left(N_{\mathcal{M}_{\Delta,r}}(G^*) \leq N(1-\epsilon)\right) \leq e^{-\delta N}.$$

Proof. The first part is immediate since the total number of offspring in $\mathcal{T}$ within distance r of the root is finite with probability one. For δ sufficiently small, Corollary 5 then implies that $\mathbb{E}\left[N_{\mathcal{M}_{\Delta,r}}(G^*)\right] \geq N\left(1-\frac{\epsilon}{2}\right)$. Apply Lemma 3 to the trivial r-local property (i.e. the property that always holds), which shows that $N_{\mathcal{M}_{\Delta,r}}(G^*)$ is a C-Lipschitz function for some C. The second part now follows from Lemma 2. □

We can now complete the proof of Proposition 16. Choose Δ sufficiently large so that

$$\left|\mathbb{P}(\mathcal{T} \in \mathcal{P}) - \mathbb{P}(\mathcal{T} \in \mathcal{P} \cap \mathcal{M}_{\Delta,r})\right| \leq \mathbb{P}(\mathcal{T} \notin \mathcal{M}_{\Delta,r}) \leq \frac{\epsilon}{4}. \tag{10}$$

Let $B = N - N_{\mathcal{M}_{\Delta,r}}(G^*)$ denote the number of high-degree vertices in the graph G^*. Since $\left|N_{\mathcal{P}}(G^*) - N_{\mathcal{P}\cap\mathcal{M}_{\Delta,r}}(G^*)\right| \leq B$, Lemma 4 implies that for some $\delta_1 > 0$, whenever $l(\mathbf{d}, D) < \delta_1$ we have

$$\mathbb{P}\left(\left|N_{\mathcal{P}}(G^*) - N_{\mathcal{P}\cap\mathcal{M}_{\Delta,r}}(G^*)\right| \geq \frac{\epsilon N}{2}\right) \leq e^{-\delta_1 N}. \tag{11}$$

Lemma 3 implies that $N_{\mathcal{P}\cap\mathcal{M}_{\Delta,r}}(G^*)$ is C-Lipschitz for some C, so Corollary 5 and Lemma 2 together imply that for another $\delta_2 > 0$, whenever $l(\mathbf{d}, D) < \delta_2$ we have

$$\mathbb{P}\left(\left|N_{\mathcal{P}\cap\mathcal{M}_{\Delta,r}}(G^*) - N\mathbb{P}(\mathcal{T} \in \mathcal{P} \cap \mathcal{M}_{\Delta,r})\right| \geq \frac{\epsilon N}{4}\right) \leq e^{-\delta_2 N} \tag{12}$$

The inequalities (10), (11), and (12), with an application of the triangle inequality, now imply that for some $\delta < \min(\delta_1, \delta_2)$, whenever $l(\mathbf{d}, D) < \delta$ we have

$$\mathbb{P}\left(\left|N_{\mathcal{P}}(G^*) - N\mathbb{P}(\mathcal{T} \in \mathcal{P})\right| \leq \epsilon N\right) \geq 1 - e^{-\delta N},$$

proving the result. □

Proposition 16 immediately proves (8) for the multigraph configuration model. Summing over component sizes above some lower bound, we also find that the number of vertices in "large" components concentrates around $N\mathbb{P}(|\mathcal{T}| = \infty)$.

Corollary 6. *Fix $\epsilon > 0$. For all sufficiently large K, there exists $\delta > 0$ such that if $l(\mathbf{d}, D) < \delta$ we have*

$$\mathbb{P}\left(\left|\sum_{k\geq K} N_k(G^*) - N\mathbb{P}(|\mathcal{T}| = \infty)\right| \geq \epsilon N\right) \leq e^{-\delta N}. \tag{13}$$

Proof. For sufficiently large K, we have $\sum_{k=1}^{K} \mathbb{P}(|\mathcal{T}| = k) \geq 1 - \frac{\epsilon}{2} - \mathbb{P}(|\mathcal{T}| = \infty)$. The result follows from (8), replacing ϵ with $\frac{\epsilon}{2K}$. □

Corollary 6 implies (9) if $\mathbb{P}(|\mathcal{T}| = \infty) = 0$, and it will also play a key role in the next section as we address the case in which $\mathbb{P}(|\mathcal{T}| = \infty) > 0$.

Coloring and Sprinkling

Having established branching process approximation results for component sizes, we now show that essentially all "large" components are connected. I assume throughout this section that $\mathbb{P}(|\mathcal{T}| = \infty) > 0$. The basic idea of the argument is to thin the graph G^* by retaining edges with some probability p. For p close to 1, the component structure of the thinned graph is similar to that of G^*. When we "sprinkle" back in the remaining edges, any large components are very likely joined together.

I choose only one type of edge to thin. By assumption there exists a type θ_1 which has three or more neighbors with positive probability. Since the graph is irreducible and $\mathcal{T}$ survives with positive probability, there exists a type θ_2 that connects to type θ_1 nodes with positive probability and has 2 or more neighbors with positive probability. These conditions ensure that in the branching process $\mathcal{T}$, with positive probability we will encounter both type θ_1 parents with type θ_2 offspring and type θ_2 parents with type θ_1 offspring. Let G' denote the subgraph of G^* that we obtain by deleting edges between type θ_1 and type θ_2 nodes independently with some probability $p \in (0, 1)$, and let G'' denote the subgraph formed by the deleted edges. We can also view G^* as a colored graph, in which the edges of G' are red and those of G'' are blue. I will sometimes write $G^*(p)$ to emphasize that I am talking about the colored version of G^*. Let $\mathbf{d}'$ denote the degree sequence of G', and let $\mathbf{d}''$ denote the degree sequence of G''. The sprinkling argument relies on the following lemma.

Lemma 5. *For any* $\mathbf{d}$ *and any* $0 < p < 1$*, the random graphs* G' *and* G'' *are conditionally independent given* $\mathbf{d}'$*.*

Proof. This follows from the definition of the configuration model. The graph G^* is a uniform random pairing of the stubs determined by $\mathbf{d}$. Color each pair red, except color edges between type θ_1 and type θ_2 nodes blue with independent probability $1-p$. Given the set of stubs in red pairs, which determines $\mathbf{d}'$ and $\mathbf{d}''$, the pairing of these stubs is uniformly random, and similarly the blue stubs are paired uniformly at random. □

The method used to prove Proposition 16 allows us to state similar concentration results for the colored subgraphs. Let $\mathcal{T}(p)$ denote the branching process $\mathcal{T}$ in which we color edges between type θ_1 and type θ_2 nodes blue with independent probability $1-p$. Let $\mathcal{T}'(p)$ denote the red subtree containing the root, and let D_p denote the thinned degree distribution. Note that D_p is the asymptotic degree distribution of G', and $\mathcal{T}'(p)$ is the corresponding branching process that approximates rooted graphs in G'. I omit the proof of the following result as it is essentially identical to that of Proposition 16.

Proposition 17. *Let* $\mathcal{P}$ *be a* r*-local property of colored rooted graphs, and fix* $\epsilon > 0$ *and* $p \in (0, 1)$*. There exists* $\delta > 0$ *such that if* $l(\mathbf{d}, D) < \delta$ *then*

$$\mathbb{P}\left(\left|N_{\mathcal{P}}\left(G^*(p)\right) - N\mathbb{P}(\mathcal{T}(p) \in \mathcal{P})\right| \geq \epsilon N\right) \leq e^{-\delta N}.$$

We also require a simple lemma bounding the probability that no links are formed between sets of stubs. Recall that $m_{\theta_1,\theta_2}(\mathbf{d})$ is the number of edges connecting type θ_1 and type θ_2 nodes, given the degree sequence $\mathbf{d}$.

Lemma 6. *Let $\{A_i\}_{i=1}^2$ and $\{B_i\}_{i=1}^2$ be disjoint sets of stubs, with A_1 and B_1 containing stubs attached to type θ_1 nodes leading to type θ_2 nodes, and vice verse for A_2 and B_2. The probability that no stubs in $A_1 \cup A_2$ are paired to stubs in $B_1 \cup B_2$ is no more then*

$$e^{-\frac{|A_1||B_1|+|A_2||B_2|}{2m_{\theta_1,\theta_2}(\mathbf{d})}}.$$

Proof. Without loss of generality, assume $|A_1| \leq |B_1|$, and conduct the following exercise. One at a time, select a random unpaired stub in A_1 and reveal its partner. Conditional on having no matches in B_1 yet, the probability of finding a partner in B_1 is at least $\frac{|B_1|}{m_{\theta_1,\theta_2}(\mathbf{d})}$. Hence, the probability that we have no matches in B_1 is at most

$$\left(1 - \frac{|B_1|}{m_{\theta_1,\theta_2}(\mathbf{d})}\right)^{|A_1|} \leq e^{-\frac{|A_1||B_1|}{2m_{\theta_1,\theta_2}(\mathbf{d})}}.$$

Repeat the argument for A_2 and B_2, and the result follows. □

We are now ready to prove (9) for the multigraph configuration model. Let $L_i = L_i(G^*)$ denote the number of vertices in the ith largest component of G^*, and fix $\epsilon > 0$. By Corollary 6, there are constants K and $\delta > 0$ such that if $l(\mathbf{d}, D) < \delta$, then

$$\mathbb{P}\left(\sum_{k \geq K} N_k(G^*) \geq N\left(\mathbb{P}(|\mathcal{T}| = \infty) + \frac{\epsilon}{4}\right)\right) \leq e^{-\delta N}.$$

Trivially, we know $L_1 + L_2 \leq 2K + \sum_{k \geq K} N_k(G^*)$. For sufficiently large N, we have $K \leq \frac{\epsilon N}{8}$, implying

$$\begin{aligned}
&\mathbb{P}\left(L_1 + L_2 \geq N\left(\mathbb{P}(|\mathcal{T}| = \infty) + \frac{\epsilon}{2}\right)\right) \\
&\quad \leq \mathbb{P}\left(\sum_{k \geq K} N_k(G^*) \geq N\left(\mathbb{P}(|\mathcal{T}| = \infty) + \frac{\epsilon}{4}\right)\right) \\
&\quad \leq e^{-\delta N}. \qquad (14)
\end{aligned}$$

To complete the proof for the multigraph G^*, it suffices to show for some δ', with $0 < \delta' \leq \delta$, that if $l(\mathbf{d}, D) < \delta'$ we have

$$\mathbb{P}\left(L_1 \geq N\left(\mathbb{P}(|\mathcal{T}| = \infty) - \frac{\epsilon}{2}\right)\right) \geq 1 - e^{-\delta' N}. \qquad (15)$$

As p approaches 1, the distribution D_p converges to the distribution D. Here we make use of the assumption that at least one type has three or more neighbors with positive probability. This means that in the forward distribution D'' for the branching process $\mathcal{T}$, there is a positive probability of having two or more offspring. This rules out the case in which a node in $\mathcal{T}$ always has one child, and one can check that under this assumption, the

survival probability $\mathbb{P}(|\mathcal{T}'(p)| = \infty)$ converges to $\mathbb{P}(|\mathcal{T}| = \infty)$. For the rest of the proof, fix a p such that $\mathbb{P}(|\mathcal{T}'(p)| = \infty) \geq \mathbb{P}(|\mathcal{T}| = \infty) - \frac{\epsilon}{8}$.

We need a lower bound on the number of stubs in G'' that are attached to large components of G'. Given $\Delta \geq 2$ and $r \geq 0$, for a vertex v, we define the r-local property $\mathcal{H}_{\Delta,r}$, which is satisfied if two conditions hold. First, no vertex within distance r of v has more than Δ neighbors in G'. Second, at least one of the following statements is true:

(a) In the component of G' containing v, no vertex lies at distance r or greater from v

(b) Within distance r of v in G', there exists both a type θ_1 node with a stub in G'' and a type θ_2 node with a stub in G''.

Lemma 7. *Fix $\epsilon > 0$. We can choose Δ and r for which there exists $\delta_1 > 0$ such that if $l(\mathbf{d}, D) < \delta_1$ we have*

$$\mathbb{P}\left(N_{\mathcal{H}_{\Delta,r}}\left(G^*(p)\right) \leq N\left(1 - \frac{\epsilon}{8}\right)\right) \leq e^{-\delta_1 N}.$$

Proof. Choosing r sufficiently large ensures that, conditional on surviving until the rth generation, the red subtree $\mathcal{T}'_r(p)$ has blue stubs of both types with probability at least $1 - \frac{\epsilon}{24}$. By Lemma 4 we can find Δ so that $\mathbb{P}\left(\mathcal{T}'(p) \in \mathcal{M}_{\Delta,r}\right) \geq 1 - \frac{\epsilon}{24}$. Consequently, we have

$$\mathbb{P}\left(\mathcal{T}(p) \in \mathcal{H}_{\Delta,r}\right) \geq 1 - \frac{\epsilon}{12}.$$

The result follows from Proposition 17. □

Fix the Δ and r obtained in Lemma 7, let S_k denote the set of vertices in components of G' with at least k vertices, and let $M = Np_{\theta_1}\mathbb{E}[D_{\theta_1} \cdot \mathbf{e}_{\theta_2}]$ be the expected number of links between type θ_1 and type θ_2 nodes. By Corollary 6 (or rather, the analog based on Proposition 17), there exists $k \geq \max\left(K, \frac{M}{N}\frac{16\Delta^{2r}}{\epsilon^2}\right)$ and $\delta_2 > 0$ such that whenever $l(\mathbf{d}, D) < \delta_2$ we have

$$\mathbb{P}\left(|S_k| \leq N\left(\mathbb{P}(|\mathcal{T}| = \infty) - \frac{\epsilon}{4}\right)\right) \leq \mathbb{P}\left(|S_k| \leq N\left(\mathbb{P}(|\mathcal{T}'(p)| = \infty) - \frac{\epsilon}{8}\right)\right) \leq e^{-\delta_2 N}. \quad (16)$$

Call a partition (X, Y) of S_k a *potentially bad cut* if both $|X| \geq \frac{\epsilon N}{4}$ and $|Y| \geq \frac{\epsilon N}{4}$, and there are no edges of G' connecting X and Y. The partition is a *bad cut* if additionally no edge in G'' connects X and Y. Each component of G' in S_k must lie entirely in X or in Y, so in any realization there are at most

$$2^{\frac{|S_k|}{k}} \leq 2^{\frac{N}{k}} \leq e^{\frac{N}{k}}$$

potentially bad cuts.

Fix a realization of $\mathbf{d}'$ and G' such that

$$N_{\mathcal{H}_{\Delta,r}}\left(G^*(p)\right) \geq N\left(1 - \frac{\epsilon}{8}\right).$$

Suppose that (X, Y) is a potentially bad cut. Both X and Y contain at least $\frac{\epsilon N}{8}$ vertices with property $\mathcal{H}_{\Delta,r}$. Since $k \geq \Delta^r$, and no vertex in $\mathcal{H}_{\Delta,r}$ can reach more than Δ^r other

vertices within r links in G', we know that each of these vertices satisfies condition (b). For any particular stub in G'', there are no more than $2\Delta^r$ paths of length r connecting it to a vertex in $\mathcal{H}_{\Delta,r}$. Therefore, both X and Y contain at least $\alpha N = \frac{\epsilon N}{16\Delta^r}$ stubs of each type in G''.

For small enough δ, the graph G'' contains no more than M edges. By Lemma 6, the probability that no edges in G'' connect X and Y is no more than

$$e^{-\frac{\alpha^2 N}{M}} \leq e^{-\frac{2N}{k}}.$$

This implies that the exepected number of bad cuts, given $\mathbf{d}'$ and G', is at most $e^{-\frac{N}{k}}$, and the probability of having any bad cuts is at most $e^{-\frac{N}{k}}$. If there are no bad cuts, then

$$L_1 \geq |S_k| - \frac{\epsilon N}{4} \geq N\left(\mathbb{P}(|\mathcal{T}| = \infty) - \frac{\epsilon}{2}\right).$$

Taking $\delta' < \min\left(\delta, \delta_1, \delta_2, \frac{1}{k}\right)$ completes the proof for the multigraph configuration model. □

Typical Distances

The results above allow a simple proof of the typical distance claim. We do not require the path counting arguments of van der Hofstad et al. (2005). I drop the subscript from $\nu_{(T,D)}$ in what follows for notational convenience. In essence, we show that the neighborhood $G_{v,r}$ is well approximated by $\mathcal{T}_r$ for r of order $\frac{1}{2}\log_\nu n$, implying that for a random vertex in the giant component, we have $|G_{v,r}|$ of order $\sqrt{n}$. Lemma 6 then implies that two such neighborhoods are connected with high probability, giving typical distances of order $\log_\nu n$.

First, we establish the lower bound:

$$\lim_{n\to\infty} \mathbb{P}\left(H(G) \leq (1-\epsilon)\log_\nu n\right) = 0.$$

Let v denote a randomly chosen vertex. From the branching process approximation we have

$$\mathbb{E}[|G_{v,r}| \leq \mathbb{E}[|\mathcal{T}_r|] = 1 + \sum_{i=1}^{r} \mathbb{E}[|D|]\nu^{i-1} = 1 + \mathbb{E}[|D|]\frac{\nu^r - 1}{\nu - 1}.$$

We can bound our probability as

$$\begin{aligned}\mathbb{P}\left(H(G) \leq (1-\epsilon)\log_\nu n\right) &\leq \frac{\mathbb{E}\left[|G_{v,(1-\epsilon)\log_\nu n}|\right]}{n} \\ &\leq \frac{1}{n} + \frac{\mathbb{E}[|D|]}{n(\nu-1)}(n^{1-\epsilon} - 1),\end{aligned}$$

which converges to zero as $n \to \infty$.

For the upper bound, we need a more precise estimate of $|G_{v,r}|$. Write Z_r for the rth generation of $\mathcal{T}$, and write $N_{v,r}$ for the set of vertices at distance exactly r from v. An implication of Proposition 12 is that, given any $\epsilon > 0$ and conditional on survival, there

exists $0 < c_\epsilon < C_\epsilon < \infty$ such that $\mathbb{P}(c_\epsilon \nu^k < |Z_k| < C_\epsilon \nu^k, \forall\, k) > 1 - \epsilon$. Consequently for $\tilde{r} = \frac{1+\epsilon}{2} \log_\nu n$ we have with probability at least $1 - \epsilon$ that

$$c_\epsilon \frac{\nu^{\tilde{r}} - 1}{\nu - 1} \leq |\mathcal{T}_{\tilde{r}-1}| \leq C_\epsilon \frac{\nu^{\tilde{r}} - 1}{\nu - 1}$$

or

$$\frac{c_\epsilon}{\nu - 1} \left(n^{\frac{1+\epsilon}{2}} - 1 \right) \leq |\mathcal{T}_{\tilde{r}-1}| \leq \frac{C_\epsilon}{\nu - 1} \left(n^{\frac{1+\epsilon}{2}} - 1 \right).$$

The upper bound applies to $|G_{v,\tilde{r}}|$ as well since vertices in $N_{v,r}$ might link to each other, or link to the same new vertex in the next extended neighborhood. However, as long as $|G_{v,r}|$ is small relative to n, the distribution of neighborhoods $N_{v,r}$ will closely track that of Z_r. In particular, as long as $|G_{v,r}| < cn^{\frac{1+\epsilon}{2}}$, the probability that a link stub from a vertex in $N_{v,r}$ connects to a redundant vertex is at most $c' n^{-\frac{1-\epsilon}{2}}$ for some constant c'. Hence, conditional on $N_{v,\tilde{r}} \neq \emptyset$, with probability at least $1 - \epsilon$, we have

$$|N_{v,r}| \geq c_\epsilon \nu^r \left(1 - c' n^{-\frac{1-\epsilon}{2}}\right)^r \geq c_\epsilon \nu^r \left(1 - c' n^{-\frac{1-\epsilon}{2}}\right)^{n^{\frac{1-\epsilon}{2}}} \geq \frac{c_\epsilon}{e^{c'}} \nu^r,$$

for $r \leq \tilde{r}$ and n sufficiently large.

Take two random vertices v and w, and consider $G_{v,\tilde{r}}$ and $G_{w,\tilde{r}}$. Conditional on $N_{v,\tilde{r}} \neq \emptyset$ and $N_{w,\tilde{r}} \neq \emptyset$, with probability at least $1 - \epsilon$ we have $|N_{v,\tilde{r}}| \geq Cn^{\frac{1+\epsilon}{2}}$ and $N_{w,\tilde{r}} \geq Cn^{\frac{1+\epsilon}{2}}$. Moreover, the law of large numbers implies that we can choose two types θ and θ' and a constant C' such that $|N_{v,\tilde{r}}|$ has at least $C' n^{\frac{1+\epsilon}{2}}$ type $\theta\theta'$ link stubs leading away, and $|N_{w,\tilde{r}}|$ has at least $C' n^{\frac{1+\epsilon}{2}}$ type $\theta'\theta$ link stubs leading away, with probability at least $1 - \epsilon$. Lemma 6 implies that the probability that there is no link between $N_{v,\tilde{r}}$ and $N_{w,\tilde{r}}$ is at most e^{-cn^ϵ} for another constant c. Taking n large, we see that with arbitrarily high probability, the distance between v and w is at most $2\tilde{r} = (1 + \epsilon) \log_\nu n$. $\square$

Proof of Proposition 8

Given a pair of effective prices $(\underline{p}, \overline{p})$, the giant component in the network of potential adopters contains some number $\underline{q}$ of low-degree consumers and $\overline{q}$ of high-degree ones. A random consumer in the giant component on average will make one successful referral, with high-degree (low-degree) consumers making $\overline{r}$ ($\underline{r}$) successful referrals in expectation, with $\overline{r} > \underline{r}$. If ρ is the referral reward, a low-degree consumer expects to earn $\rho\underline{r}$ from making referrals, and a high-degree consumer expects $\rho\overline{r}$. If $\overline{p} < \underline{p}$, the firm can implement the effective prices $(\underline{p}, \overline{p})$ by choosing

$$\rho = \frac{\underline{p} - \overline{p}}{\overline{r} - \underline{r}}, \quad p = \underline{p} + \rho\underline{r}.$$

I now show that an optimal pair of effective prices satisfies $\overline{p} \leq \underline{p}$. Define $\tilde{p} = \frac{\underline{d}\underline{p} + \overline{d}\overline{p}}{\underline{d} + \overline{d}}$ as the expected price a random neighbor faces. Let X denote the degree distribution in the

network of potential adopters, and let $B_{n,p}$ denote a binomial distribution with n trials and success probability p. The branching process $\mathcal{T}_X$ has extinction probability ξ that solves $\mathbb{E}[X]\xi = G'_X(\xi)$. A low-degree potential adopter is connected to the giant component with probability

$$1 - \mathbb{E}[\xi^{B_{\underline{d},1-\tilde{p}}}] = 1 - (\tilde{p} + (1-\tilde{p})\xi)^{\underline{d}},$$

and a high-degree potential adopter is connected to the giant component with probability

$$1 - \mathbb{E}[\xi^{B_{\overline{d},1-\tilde{p}}}] = 1 - (\tilde{p} + (1-\tilde{p})\xi)^{\overline{d}}.$$

The profit the firm expects to earn from offering effective prices $(\underline{p}, \overline{p})$ is then

$$\pi(\underline{p}, \overline{p}) = \frac{1}{2}\left[\underline{p}(1-\underline{p})\left(1 - (\tilde{p} + (1-\tilde{p})\xi)^{\underline{d}}\right) + \overline{p}(1-\overline{p})\left(1 - (\tilde{p} + (1-\tilde{p})\xi)^{\overline{d}}\right)\right].$$

Define $x = \tilde{p} + (1-\tilde{p})\xi$. First order conditions for optimal pricing imply that

$$\frac{\partial \pi}{\partial \underline{p}} = (1-2\underline{p})(1-x^{\underline{d}}) - \left[\underline{p}(1-\underline{p})\underline{d}x^{\underline{d}-1} + \overline{p}(1-\overline{p})\overline{d}x^{\overline{d}-1}\right]\frac{\partial x}{\partial \underline{p}} = 0, \text{ and}$$

$$\frac{\partial \pi}{\partial \overline{p}} = (1-2\overline{p})(1-x^{\overline{d}}) - \left[\underline{p}(1-\underline{p})\underline{d}x^{\underline{d}-1} + \overline{p}(1-\overline{p})\overline{d}x^{\overline{d}-1}\right]\frac{\partial x}{\partial \overline{p}} = 0.$$

This implies at an optimum we have

$$\frac{\frac{\partial x}{\partial \underline{p}}}{\frac{\partial x}{\partial \overline{p}}} = \frac{(1-2\underline{p})(1-x^{\underline{d}})}{(1-2\overline{p})(1-x^{\overline{d}})}.$$

To finish the proof, we show that the ratio $\frac{\frac{\partial x}{\partial \underline{p}}}{\frac{\partial x}{\partial \overline{p}}}$ is always less than $\frac{\underline{d}}{\overline{d}}$. Since $\frac{1-x^{\underline{d}}}{1-x^{\overline{d}}} \geq \frac{\underline{d}}{\overline{d}}$, this implies that $\overline{p} < \underline{p}$ at an optimum.

To prove the claim, we can compute from the definition

$$\frac{\partial x}{\partial \underline{p}} = \frac{\underline{d}}{\underline{d}+\overline{d}}(1-\xi) + (1-\tilde{p})\frac{\partial \xi}{\partial \underline{p}}, \quad \frac{\partial x}{\partial \overline{p}} = \frac{\overline{d}}{\underline{d}+\overline{d}}(1-\xi) + (1-\tilde{p})\frac{\partial \xi}{\partial \overline{p}}.$$

From the equation $\mathbb{E}[X]\xi = G'_X(\xi)$, we compute

$$(1-\tilde{p})\frac{(1-\underline{p})\underline{d} + (1-\overline{p})\overline{d}}{2-\underline{p}-\overline{p}}\xi = \frac{(1-\underline{p})\underline{d}}{2-\underline{p}-\overline{p}}\xi^{\underline{d}-1} + \frac{(1-\overline{p})\overline{d}}{2-\underline{p}-\overline{p}}\xi^{\overline{d}-1},$$

or equivalently

$$(1-\tilde{p})\left((1-\underline{p})\underline{d} + (1-\overline{p})\overline{d}\right) = (1-\underline{p})\underline{d}\xi^{\underline{d}-2} + (1-\overline{p})\overline{d}\xi^{\overline{d}-2}.$$

Differentiating, along with a little algebra, yields

$$\frac{\frac{\partial \xi}{\partial \overline{p}}}{\frac{\partial \xi}{\partial \underline{p}}} = \frac{\frac{\overline{d}}{\underline{d}+\overline{d}}\left((1-\underline{p})\underline{d}+(1-\overline{p})\overline{d}\right)+\overline{d}(1-\tilde{p})-\overline{d}\xi^{\overline{d}-2}}{\frac{\underline{d}}{\underline{d}+\overline{d}}\left((1-\underline{p})\underline{d}+(1-\overline{p})\overline{d}\right)+\underline{d}(1-\tilde{p})-\underline{d}\xi^{\underline{d}-2}} > \frac{\overline{d}}{\underline{d}},$$

which in turn implies

$$\frac{\frac{\partial x}{\partial \overline{p}}}{\frac{\partial x}{\partial \underline{p}}} > \frac{\overline{d}}{\underline{d}}$$

as desired. □