Diffusion Games[*]

Evan Sadler[†]

March 29, 2018

Abstract

Behaviors and information often spread via person-to-person diffusion. I characterize the extent and rate of diffusion in a large population linked in a fixed network. I argue that standard simplifying assumptions in prior work lead to misguided predictions, such as reversed comparative statics. Our study also uncovers a new strategic effect: when large cascades are possible, exposure to the contagion conveys information about a player's network position, which facilitates coordination. The analysis requires technical advances in the study of random graphs, and a key contribution is a pair of limit theorems for a multi-type configuration model.

1 Introduction

Diffusion processes in disparate contexts share similar features. People are exposed to something new through contact with others. They choose whether to adopt some behavior or pass on information. Preferences are heterogeneous and can depend on neighbors' choices. The ultimate outcome is highly uncertain, occasionally producing "viral" cascades.

As one example, the spread of fake news featured prominently in the 2016 United States elections. False stories targeted to particular groups spread almost exclusively on social media, often reaching more readers than the most widely consumed mainstream stories (Allcott and Gentzkow, 2017). Other important examples include the spread of rumors about India's demonetization policy (Banerjee et al., 2017), the spread of new products and services through referrals (Lobel et al., 2015), and the spread of microfinance in developing countries (Banerjee et al., 2013). Understanding these processes is therefore important for countering misinformation, designing marketing campaigns, and encouraging economic development.

[†]Harvard University – esadler@fas.harvard.edu

The standard approach for studying these processes affords a tractable framework but imperfectly captures key features of person-to-person diffusion. The approach entails two simplifying assumptions:

(a) The effect of social interactions at any time depends only on population averages;

(b) The population is a continuum, so actions by one person are insignificant in aggregate.

I refer to these as the mean-field assumptions. If we take seriously the idea of person-to-person transmission, assumption (a) means that we continually reshuffle links in the network. This neglects the stability of real relationships over time and eliminates any variance in the eventual outcome. Similarly, assumption (b) precludes viral phenomena. Choices within a small group can never trigger widespread diffusion—we need a critical mass of early adopters. Despite these shortcomings, mean-field models may still prove useful to the extent that their predictions align with what a more realistic model could produce.

The present paper makes two main contributions to our understanding of diffusion processes. First, I show that a more realistic model yields substantively different predictions. I illustrate this in an example comparing a mean-field model with an analogous "diffusion game." In the diffusion game, players interact in a fixed network, so the extent of diffusion is sensitive to the initial seeding—some seeds trigger large cascades, while others induce little adoption. In a mean-field model, no one remains unexposed in the long run, while in the diffusion game, there are two distinct reasons people fail to adopt: some make a choice, others are never exposed. This distinction leads to more nuanced comparative statics. Changes to the network structure can have different effects on the speed and extent of diffusion. In some cases, we actually *reverse* the prediction of the mean-field model.

More significantly, the diffusion game exhibits a novel strategic effect. When large cascades are possible, exposure to the contagion conveys information about a player's network position—being exposed is evidence that the player is in a large network component. When adoption entails local complementarities, this information helps players coordinate their choices. Despite players remaining unexposed in the diffusion game, this effect can sometimes sustain higher equilibrium adoption levels.

A second contribution is to render mean-field assumptions unnecessary for tractability: I provide tools to analyze diffusion in a large class of random graphs called multi-type configuration models. Intuitively, a configuration model takes a uniform random draw among all graphs that satisfy a given set of summary statistics. A pair of limit theorems characterizes the component structure of these graphs as the number of players grows—in a large population, essentially all graphs look roughly the same. This makes it meaningful to talk about properties of a "generic" graph, which may prove useful beyond studies of diffusion.

The technical insight is to approximate the local network structure with a branching process. This allows us to study network properties using more standard results for sums of independent random variables. An extensive literature in statistical physics studies related models, but much of this work relies on simulation and heuristic arguments to substantiate its claims (See Durrett, 2007, Ch. 1). Molloy and Reed (1995), van der Hofstad et al.

(2005), and Bollobás and Riordan (2015) provide versions of my limit theorems for the one-type model. To the best of my knowledge, the present paper is the first to supply a formal argument that applies to more general random graphs.

As data on networks becomes increasingly available, a richer modeling approach can help build theories that are more useful to empirical researchers. Recent studies show that the centrality of early adopters predicts uptake of microfinance (Banerjee et al., 2013), that individuals know who is central in their networks (Banerjee et al., 2016), and that people make strategic choices about gathering and sharing information (Bandiera and Rasul, 2006; Banerjee et al., 2012). To fully engage with this work, theory needs to represent diffusion processes in more detail with more realistic features. The present paper includes results that are directly relevant to seeding applications, based on primitives that correspond to the kinds of data that are typically gathered. If one wants to efficiently allocate limited seeds across several communities, then it is important to get network comparative statics right. Section 4.2 illustrates how one can combine demographic data with distributional information about a network to optimally target individual seeds.

After discussing related work, I begin with an example in a one-type model and a careful comparison with the mean-field approach. The following sections develop the general theory of diffusion on networks with heterogeneous players. Section 6 gives an overview of the technical contribution, while formal proofs are in an appendix. Appendices also contain a discussion of several extensions as well as background material on branching processes and configuration models. I conclude with brief remarks.

1.1 Related Work

Mean-field models dominate the theoretical literature on diffusion. Canonical examples assume all prior adopters influence current decisions equally (Bass, 1969; Granovetter, 1978), so we can characterize adoption over time using a logistic curve. More recent work builds on this approach, incorporating heterogeneous agents and interactions. A degree-based mean-field approximation distinguishes individuals according to how many neighbors they have, separately tracking adoption levels among agents with each possible degree (Jackson and Rogers, 2007; López-Pintado, 2006, 2008, 2012). Jackson and López-Pintado (2013) go further, allowing multiple types of agents and studying homophily. In each case, we obtain distinct logistic curves that describe adoption levels for each distinguishable subpopulation, and the degree distribution determines relationships between these curves.

The present paper addresses three limitations of these models. Since steady states are deterministic functions of the degree distribution, long-run outcomes are independent of the initial seeds. This limits our ability to capture the empirically documented role of innovators and opinion leaders. In contrast, early adopters play a crucial role in diffusion on a fixed graph. Relatedly, adoption is typically *reversible* in a mean-field model—without this assumption, steady states become insensitive to the network structure.[1] Sometimes adoption

---

[1]Young (2009) studies irreversible adoption in a mean-field diffusion model, but his focus is on the shape of the adoption curve rather than the long-run steady state.

constitutes a one-time consumption (e.g. watching a video or reading a news story), or learning a piece of information. Even when adoption is reversible in the long-run, switching costs—for instance, the need to invest captial or learn how to use a new technology—can induce short-run irreversibility. Lastly, by rendering individual decisions insignificant for global outcomes, the mean-field assumptions trivialize strategic effects. We find that, even in large networks, a single decision can meaningfully affect the aggregate outcome.

The literature on static network games with incomplete information (e.g. Sundararajan, 2007; Galeotti et al., 2010) is closely related both to this paper and to mean-field diffusion models. In these games, players act simultaneously, and payoffs depend on what a player's network neighbors do. Jackson and Yariv (2007) study a product adoption game in this context, showing that Bayes-Nash equilibria of the network game correspond to the steady states of a mean-field diffusion process. The diffusion game of section 3 features similar decisions and payoff functions, and equilibrium strategies resemble those in the static game. A difference arises because dynamic interactions affect a player's *beliefs about her network position.* Viral belief distortion facilitates coordination, resulting in equilibrium outcomes that are difficult for static models to replicate.

Not all diffusion models rely on mean-field assumptions. In a canonical example, Morris (2000) studies a fixed network in which players adopt a behavior if enough neighbors did so previously. The analysis reveals how cohesive subgroups act as barriers to diffusion and explores conditions that facilitate diffusion. Watts (2002), Campbell (2013), and Akbarpour et al. (2017) are more closely related to this paper, studying diffusion in random graphs.[2] Watts (2002) studies irreversible adoption decisions based on a threshold rule, finding that whether large cascades can occur depends crucially on agents who adopt after a single neighbor does so. Campbell (2013) studies monopoly pricing when consumers learn about a product through word-of-mouth—the information externality results in more elastic demand and lower prices than in a standard model. Akbarpour et al. (2017) explore the value of network information for seeding under a mechanistic transmission process.

I make several innovations on this literature. First, the agents here make strategic choices. Second, by adapting newer mathematical tehcniques, our results permit the study of random networks with heterogeneous agents. This important for applications as individuals often have different preferences and different information, and neighbors' attributes are typically correlated. Moreover, the explicit link to branching processes provides a unifying conceptual framework to study diffusion in random networks. In addition to characterizing the size of large cascades, we can assess the rate of diffusion, the influence of different initial seeds on the long-run outcome, and the microstructure of small components in the network.

The role of social networks in technology adoption is widely studied in the development literature. Bandiera and Rasul (2006) look at social learning effects on the spread of sunflower cultivation in Mozambique. As more of a farmer's friends and family adopt sunflower, the farmer's own propensity to adopt initially increases and later decreases. The authors attribute the latter effect to informational free-riding: a farmer with many adopting friends benefits from others' accumulated knowledge and waits to adopt himself. Other studies ex-

---

[2]See also Chapter 7.2 of Jackson (2008).

plore how different strategies to seed initial adopters influence the diffusion of an innovation (Banerjee et al., 2013; Beaman et al., 2015). This work suggests that seeding based on network centrality measures can significantly outperform alternative approaches.

Collectively, this research demonstrates that information about new technologies—about both the technology's existence and how best to use it—diffuses through social ties,[3] that it matters who the initial seeds are, and that individuals make strategic adoption decisions. The present paper offers a microfounded theory of diffusion that captures all of these elements. Applications may furnish richer sets of predictions to test and help us better understand the mechanisms underlying these empirical phenomena. A parallel literature focuses on other factors that influence technology adoption, like adoption costs (Comin and Hobijn, 2010), education and human capital (Caselli and Coleman, 2001; Skinner and Staiger, 2005), political institutions (Comin and Hobijn, 2004), and geography (Comin et al., 2013). An improved theoretical understanding of diffusion through social ties can help delineate when we should expect social influence to be more or less important relative to these other factors.

Another branch of literature on diffusion lies at the intersection of marketing and computer science. The explosion of internet-based data has allowed researchers to study diffusion phenomena that occur in everyday life. Adoption cascades exhibit a heavy-tailed size distribution across a variety of domains, from product recommendations to communication platforms to games to news stories (Leskovec et al., 2006; Goel et al., 2012). Large scale experiments are able to demonstrate the causal influence of peers on adoption decisions (Bakshy et al., 2012) and document the outsized role of central individuals (Tucker, 2008; Bakshy et al., 2009). Cheng et al. (2014) study whether cascade sizes are predictable, finding that greater breadth, rather than depth, is much more likely to generate a large cascade. These patterns are consistent with the model in this paper, suggesting that the theory I develop can tie together diffusion research in disparate domains.

2   Diffusion versus Steady State Adoption

I first illustrate the paper's main theoretical contribution in the simplest setting with a single type. To highlight the novel features, I present the model together with a mean-field benchmark. For ease of exposition, I frame the benchmark as a simultaneous-move adoption game. Its equilibria correspond to steady states of a mean-field diffusion process (See Jackson and Yariv, 2007).

Two important differences stand out. First, we have a new reason why players fail to adopt: some never get exposed. This small change has significant consequences. In contrast with the mean-field model, whether we get little adoption or a viral cascade depends on a seed's exact location. Random seeds lead to random aggregate outcomes, even in the large network limit. When a viral cascade occurs, some would-be adopters remain unexposed. This proves important for comparative statics. For instance, a mean-preserving spread in

---

[3]A large literature on social learning and technology adoption that does not focus on network structure corroborates this. See for instance Foster and Rosenzweig (1995), Munshi (2004), Conley and Udry (2010), and Dupas (2014).

the degree distribution might lead to more adoption in the mean-field benchmark but less in the diffusion game. Such mistakes are costly in applications. Suppose we want to leverage network information to seed adoption of a new technology in several communities. We might wish to spend more seeds in the communities where viral cascades are less likely. Ignoring the effects of non-exposure may lead to an inefficient allocation.

Second, players in the diffusion game are fully rational and forward looking. As a result, players can learn about the network, and this informs their choices. The main insight here is that viral cascades are self-reinforcing *through the equilibrium strategies*. Getting exposed tells a player that her neighbors are more inclined to adopt. With strategic complementarities, this strengthens the player's own adoption incentives. The diffusion process serves as a coordination device, and this weighs on a number of applications. For instance, understanding these effects may help us better assess the relative merits of different marketing strategies—relying on word-of-mouth not only saves on advertising costs, potential customers may better coordinate their purchases when a product exhibits complementarities.

2.1 The Mean-Field Benchmark

We have a unit mass of players who simultaneously decide whether to adopt some behavior. The players are linked a network with degree distribution D, and each has an i.i.d. private value v drawn uniformly on $[0, 1]$. If a neighobrs adopt, the payoff from adoption is

$$u(v, a) = av - c,$$

where $c > 0$ is the cost of adoption. Non-adoption earns zero.

Each player observes her own value v and number of neighbors (hereafter, degree) d. We consider symmetric equilibria: a strategy profile $\sigma : [0, 1] \times \mathbb{N} \to \{0, 1\}$ gives an adoption decision for each value-degree pair. Each neighbor has a degree drawn independently from the distribution $\tilde{D}$ with

$$\mathbb{P}(\tilde{D} = d) = \frac{d\mathbb{P}(D = d)}{\mathbb{E}[D]}.$$

This represents a standard correction for the friendship paradox: high-degree individuals are overrepresented as neighbors.[4] Therefore, if a player has d neighbors, the number who adopt a follows a binomial distribution with d trials and success probability

$$q_\sigma = \mathbb{E}\left[\sigma(V, \tilde{D})\right]. \tag{1}$$

The strategy profile σ is an equilibrium if

$$\sigma(v, d) = \begin{cases} 1 & \text{if } vdq_\sigma - c > 0 \\ 0 & \text{if } vdq_\sigma - c < 0 \end{cases}$$

for all (v, d). A fraction $p_\sigma = \mathbb{E}[\sigma(V, D)]$ of the population ends up adopting.

---

[4]See for instance Jackson and Yariv (2007).

To understand how this seemingly static game represents diffusion, consider a form of myopic best response dynamics. Suppose time is discrete, and σ describes behavior at the end of a period. Next period, a player decides whether to adopt assuming each neighbor does so with independent probability q_σ. This encodes two assumptions. The first is myopia: a player assumes that others follow the same strategy next period. The second is the mean-field assumption: a player assumes her neighbors are a new random draw from the population. The steady states of this process are exactly the static game equilibria. There are typically two equilibria with positive adoption levels: a high stable equilibrium, and the lower "tipping point." If initial adoption is below the tipping point, we converge to no-adoption. To reach the high steady state, we must seed a large fraction of players.

2.2 The Diffusion Game

There are n players in a random network, drawn according to a configuration model with degree distribution D. That is, each player independently realizes a degree d according to D, and we take a uniform random random draw of all graphs with the resulting n-vector of degrees.[5] In the limit as $n \to \infty$, a random player has degree drawn according to D, and a random neighbor has degree drawn according to $\tilde{D}$ as defined above.

Initially, no player is aware of the new behavior. At time zero, a single player chosen uniformly at random adopts the behavior, exposing her neighbors. In each subsequent period, players who are aware make an irreversible choice whether to adopt. If a player adopts, her neighbors become aware. The payoff from adoption is exactly as before. Players have independent private values drawn uniformly on $[0, 1]$, and adopting results in a payoff

$$u(v, a) = av - c,$$

where a is the number of neighbors who *ever* adopt.[6]

On exposure, each player observes her own value v and degree d—importantly, players do not observe the period t.[7] We consider perfect Bayesian equilibria in symmetric strategies $\sigma : [0, 1] \times \mathbb{N} \to \{0, 1\}$. Write $\hat{q}_\sigma$ for the probability, conditional on exposure, that a neighbor is willing to adopt.[8] The profile σ is an equilibrium if for all (v, d), we have

$$\sigma(v, d) = \begin{cases} 1 & \text{if } v(1 + \hat{q}_\sigma(d - 1)) - c > 0 \\ 0 & \text{if } v(1 + \hat{q}_\sigma(d - 1)) < 0, \end{cases}$$

where the leading 1 appears because a newly aware player knows one neighbor has adopted.

---

[5] We must of course condition on realizing a degree sequence for which a graph actually exists. For large n, the only substance of this conditioning is that the sum of all degrees must be even.

[6] All decisions are irreversible. In particular, a player who chooses not to adopt after getting exposed never revists the decision. As a result, when a player adopts and exposes her neighbors, all neighbors make final choices within one additional period. Section 3.2 contains a more complete discussion of the modeling assumptions.

[7] We interpret this as assuming the players do not know how long the diffusion process has been running. Players are not ignorant of the calendar date; they are ignorant of when the seed adopts.

[8] In the large network limit, neighbor adoption decisions are independent.

I give a brief summary analysis based on results proved later—all statements pertain to the limit as $n \to \infty$. Fix a strategy profile σ. I refer to players for whom $\sigma(v, d) = 1$ as *potential adopters*. Let $\sigma_d = \mathbb{E}[\sigma(V, d)]$ denote the probability that a degree d player is a potential adopter, and define the distribution D_σ by

$$\mathbb{P}(D_\sigma = d) = \frac{\mathbb{P}(D = d)\sigma_d}{\sum_{k \in \mathbb{N}} \mathbb{P}(D = k)\sigma_k}.$$

This adjusts for a selection effect, giving the degree distribution for a random potential adopter. Let g and g_σ denote the probability generating functions of D and D_σ respectively. Recall q_σ defined in (1), the *unconditional* probability that a neighbor is a potential adopter, and let ξ_σ denote the minimal solution to

$$g'_\sigma(1)\xi = g'_\sigma(q_\sigma\xi + 1 - q_\sigma). \tag{2}$$

A giant component exists in the subnetwork of potential adopters if and only if $\xi_\sigma < 1$.[9] The value ξ_σ represents the probability that a neighbor does *not* connect to the giant component.

In the limit as $n \to \infty$, the fraction of players who become exposed and adopt is a random variable, taking the value zero with probability $g(q_\sigma\xi_\sigma + 1 - q_\sigma)$ and the value $p_\sigma(1 - g_\sigma(q_\sigma\xi_\sigma + 1 - q_\sigma))$ otherwise. Even when a giant component exists, with positive probability the initial seed is not connected to it, and essentially no adoption takes place. If the initial seed is in the giant component, we get a large cascade of adoption. In this case, we still fail to reach the full population p_σ of potential adopters—a fraction $g_\sigma(q_\sigma\xi_\sigma + 1 - q_\sigma)$ of them are disconnected from the giant component.

Compared with the simultaneous-move game, the diffusion game shares the same choice set, the same payoff structure, and the same essential network structure. The need to spread awareness is an added barrier to adoption—for a player to adopt, the player must both get exposed and have $\sigma(v, d) = 1$. Under many strategy profiles, diffusion reaches only a negligible fraction of players. However, we need not start above a "tipping point" to ensure self-sustaining adoption. A single seed may trigger a cascade that covers a positive fraction of the network.

2.3 A New Effect: Viral Belief Distortion

Forward looking players can learn about their network position. This manifests as a difference between q_σ, the unconditional probability that a neighbor is a potential adopter, and $\hat{q}_\sigma$, the corresponding probability conditional on exposure. If $\xi_\sigma = 1$, there is no difference, and $q_\sigma = \hat{q}_\sigma$. If $\xi_\sigma < 1$, meaning viral cascades are possible, then a player's ex ante likelihood of exposure in a viral cascade dwarfs the corresponding likelihood in a small cascade. As a result, conditional on being exposed, a player is convinced that

(a) A large cascade is in progress, and

---

[9]A giant component is one whose size grows linearly in n as $n \to \infty$.

(b) She is connected to the giant component.

This is the effect I call viral belief distortion.

The degree distribution in the giant component is different from that in the network overall because high-degree players are more likely to be in the giant component. An exposed player learns that her neighbors have degrees drawn according to $\tilde{D}^{(\sigma)}$, with

$$\mathbb{P}(\tilde{D}^{(\sigma)} = d) = \frac{\mathbb{P}(\tilde{D} = d)\left(1 - (q_\sigma \xi_\sigma + 1 - q_\sigma)^d\right)}{\sum_{k \in \mathbb{N}} \mathbb{P}(\tilde{D} = k)\left(1 - (q_\sigma \xi_\sigma + 1 - q_\sigma)^k\right)}. \tag{3}$$

Relative to $\tilde{D}$, the distribution $\tilde{D}^{(\sigma)}$ is skewed towards higher degree players. The conditional probability that a neighbor is a potential adopter is then

$$\hat{q}_\sigma = \mathbb{E}\left[\sigma(V, \tilde{D}^{(\sigma)})\right].$$

Higher degree players are more likely to adopt, so when $\xi_\sigma < 1$, we have $\hat{q}_\sigma > q_\sigma$: learning facilitates coordination, allowing higher equilibrium strategies.

The best reply map illustrates the impact of learning on coordination. Figure 1 depicts $q_{BR(\sigma)}$ as a function of q_σ, where $BR(\sigma)$ is a best reply profile. In the static game, a player with d neighbors adopts if

$$v > \frac{c}{q_\sigma d}.$$

In the diffusion game, this player adopts if

$$v > \frac{c}{1 + \hat{q}_\sigma(d - 1)}.$$

There are two differences in the diffusion game. First, an exposed player knows one of her neighbors has already adopted. Second, when σ crosses a critical threshold, viral belief distortion appears. Looking at (3), we see that this selection effect is most significant when ξ_σ is close to 1. Consequently, as σ increases, the best reply map exhibits a discontinuous jump upwards.

2.4   Reversed Comparative Statics

Consider two degree distributions. The distribution D_1 is degenerate, taking the value 3 for sure, while D_2 is either 1 or 5 with equal probability. The distribution D_2 is a mean preserving spread of D_1. In the mean-field benchmark, adoption is always higher under D_2— if the cost c is positive, adoption is strictly higher. Intuitively, a mean-preserving spread exacerbates the friendship paradox. The neighbor degree distribution becomes more skewed towards high degrees, making neighbors more likely to adopt, yielding stronger adoption incentives. This effect is still present in the diffusion game, but the network also becomes more fragmented, which may result in fewer exposed players. The latter effect can dominate: for c in a neighborhood of zero, equilibrium adoption is higher under D_1.
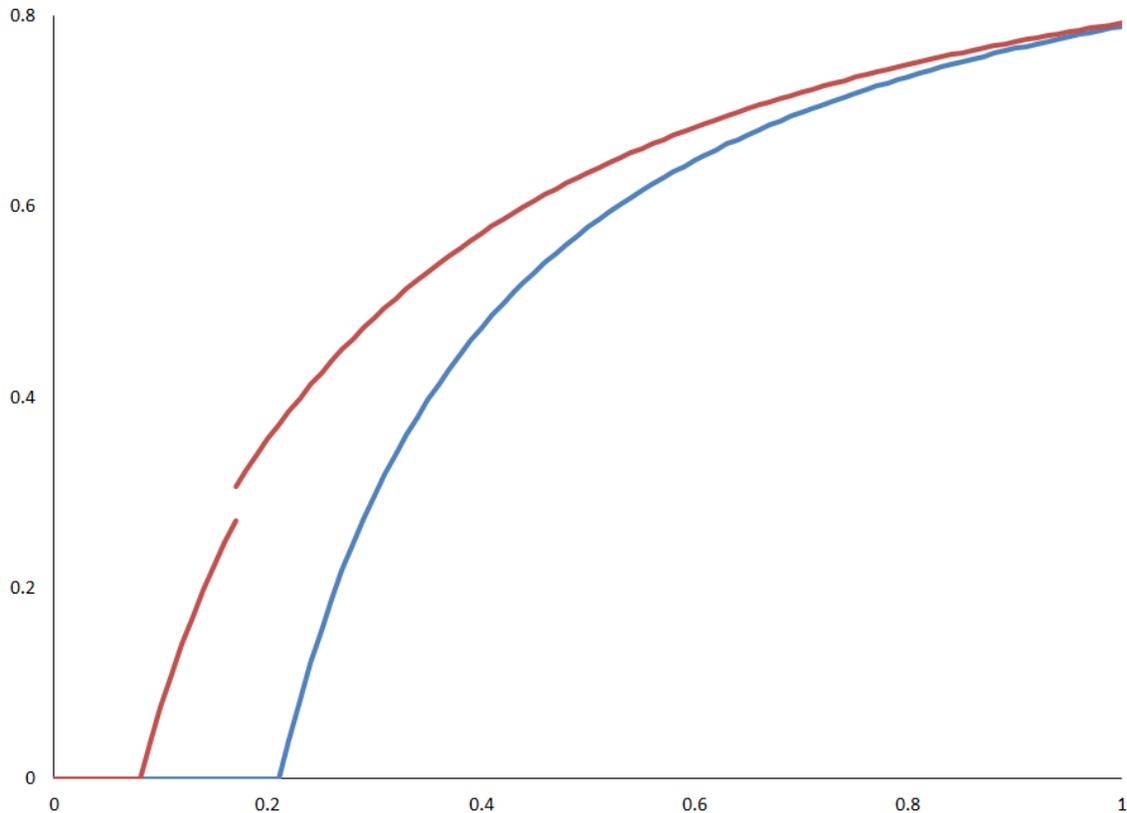
Figure 1: Best reply maps for the simultaneous adoption game and the diffusion game.

If $c = \epsilon \approx 0$, and $D_1 = 3$ with probability one, then nearly all players are potential adopters in equilibrium, and essentially all are connected to the giant component. We have $\xi_\sigma \approx 0$, and nearly the entire population adopts with probabiliy close to one. If we move to the distribution D_2, then essentially all players are still potential adopters, but ξ_σ approximately solves

$$6s = 1 + 5s^4,$$

which gives $\xi_\sigma \approx 0.1673$. The giant component no longer covers the entire network. We get negligible adoption with probability $g(\xi_\sigma) \approx 0.084$, and otherwise we get a fraction

$$1 - g_\sigma(\xi_\sigma) \approx 1 - \frac{1}{2}\left(\xi_\sigma + \xi_\sigma^5\right) \approx 0.916.$$

Under the distribution D_2, any realization of the network contains many small components. The population of potential adopters increases as in the mean-field model, but this fragmentation means that fewer become aware. Later analysis shows that the *speed* of diffusion actually increases when we switch to D_2, highlighting that the speed and extent of diffusion can move in different directions.

2.5 Higher Equilibrium Adoption

The need to spread awareness creates a barrier to adoption. Nevertheless, we can some-times sustain higher equilibrium adoption levels than in the mean-field benchmark. Suppose D takes the values 1 or 5 with equal probability. In the mean-field model, one can verify that an equilibrium with positive adoption exists if and only if $c \leq \frac{25}{24}$. In the diffusion game, I show that viral equilibria exist even with higher costs.

For c above one, we know that players with degree 1 never adopt. I therefore abuse notation letting σ denote the probability that a degree 5 player adopts under strategy profile σ. We can compute $q_\sigma = \frac{5\sigma}{6}$, and the distribution D_σ takes the value 5 with probability one, so $g_\sigma(s) = s^5$. From (2), know that ξ_σ solves

$$\xi = \left(\frac{5\sigma}{6}\xi + 1 - \frac{5\sigma}{6} \right)^4,$$

and one can verify that $\xi_\sigma < 1$ if $\sigma > \frac{1}{5}$. From (3), we have

$$\hat{q}_\sigma = \sigma \mathbb{P}(\tilde{D}^{(\sigma)} = 5) = \sigma \frac{5\left(1 - (q_\sigma\xi_\sigma + 1 - q_\sigma)^5\right)}{5\left(1 - (q_\sigma\xi_\sigma + 1 - q_\sigma)^5\right) + q_\sigma(1 - \xi_\sigma)} > q_\sigma.$$

The profile σ corresponds to an equilibrium if the cutoff value satisfies

$$v = 1 - \sigma = \frac{c}{1 + 4\hat{q}_\sigma},$$

which has a solution $\sigma^* > \frac{1}{5}$ as long as

$$c < \sup_{\sigma \in \left[\frac{1}{5}, 1\right]} (1 - \sigma)(1 + 4\hat{q}_\sigma) \equiv \bar{c}.$$

To estimate $\bar{c}$, note that

$$\bar{c} > \sup_{\sigma \in \left[\frac{1}{5}, 1\right]} (1 - \sigma)(1 + 4q_\sigma)$$

$$= \sup_{\sigma \in \left[\frac{1}{5}, 1\right]} (1 - \sigma)\left(1 + \frac{10\sigma}{3} \right)$$

$$= \left(1 - \frac{7}{20} \right)\left(1 + \frac{10}{3} \cdot \frac{7}{20} \right)$$

$$= \frac{169}{120} > \frac{25}{24}.$$

Even without viral belief distortion, we could sustain adoption in equilibrium with c as high as $\frac{169}{120}$. This is because newly aware players know that one neighbor has already adopted. Viral belief distortion further enhances coordination, allowing us to push c even higher and still obtain viral cascades in equilibrium.

11

3 Diffusion Games with Local Externalities

This section presents the general model of diffusion games—the model with multiple types and arbitrary correlation between neighbors' types. The extension is valuable for two reasons. First, in real networks, neighbors' attributes are correlated. This correlation affects both the mechanics of diffusion and individual decisions in equilibrium. Second, having multiple types creates a clearer link to the empirical diffusion literature. Researchers often gather data on individuals' demographic or other attributes—which could correspond to types—while network data are typically incomplete or noisy. The general framework offers predictions based on the kind of information that is most often available.

Studying diffusion in discrete graphs presents three key challenges. First, tracking a small set of state variables is insufficient—it matters exactly who has adopted and where they are in the network. Second, not everyone adopts if exposed. The subnetwork among those who are inclined to adopt determines the outcome, so we need to understand this subnetwork. Finally, we must account for the inferences of forward-looking players. Beliefs about the network impact behavior, and players can learn about the network during the game.

We get traction from the configuration model's special structure. Locally, the network converges in distribution to a multi-type branching process. Properties of the network— the size of the giant component, typical distances within that component—are then tied to properties of the branching process—its survival probability, its growth rate—which we can assess using standard techniques for studying sums of independent random variables. Taking the limit theorems as given, our analysis reduces to an accounting of selection effects. A random neighbor is different from a random player. A random potential adopter is different from a random neighbor. A potential adopter in the giant component is different from a random potential adopter.

After presenting the model, I divide the analysis among subsequent sections. Section 4 fixes a strategy profile and evaluates the extent and rate of diffusion. I introduce the *characteristic branching process* for the network and define parameters—the *diffusivity* and the *virality*—that characterize outcomes. Section 4.2 highlights the application of the framework to seeding problems, while section 4.3 gives basic comparative statics results. In section 5, I address beliefs and best responses. Finally, section 6 presents the limit theorems on which all of the analysis rests.

3.1 The Model

There are n players connected in an undirected network G. At time $t = 0$, we select $\kappa \in \mathbb{N}$ players uniformly at random to adopt a new behavior. Neighbors of these seeds become aware and make irreversible choices whether to adopt the behavior at time $t = 1$. In each subsequent period, neighbors of those who adopted in the previous period become aware and make their own irreversible choices.

We draw the network G according to a multi-type configuration model. Let Θ denote a finite collection of types, and let $T \in \Delta(\Theta)$ denote a distribution over types. For each type $\theta \in \Theta$, let $D_\theta \in \Delta(\mathbb{N})$ denote a degree distribution, and let $Z_\theta \in \Delta(\Theta)$ denote a distribution

over types. We realize G in four steps:

(a) Assign each player a type drawn independently according to T

(b) For each type θ player, independently draw a degree d according to D_θ and attach d "link stubs" to the player

(c) Assign a type to each link stub, drawn independently according to Z_θ

(d) Pair compatible stubs uniformly at random to form links—for each type θ' link stub attached to a type θ player, match with a type θ stub attached to a type θ' player

Condition this process on realizing matching numbers of compatible stubs and realizing a simple graph.[10] We can interpret this procedure as a uniform random draw over graphs matching the summary statistics T, $\{D_\theta\}_{\theta \in \Theta}$, and $\{Z_\theta\}_{\theta \in \Theta}$. For limits to make sense, we need a consistency condition on the distributions. Define

$$p_\theta = \mathbb{P}(T = \theta), \quad \mu_\theta = \mathbb{E}[D_\theta], \text{ and } q_{\theta'}^{(\theta)} = \mathbb{P}(Z_\theta = \theta').$$

We assume

$$p_\theta \mu_\theta q_{\theta'}^{(\theta)} = p_{\theta'} \mu_{\theta'} q_{\theta}^{(\theta')}$$

to ensure the expected number of type θ' stubs attached to type θ players equals the expected number of type θ stubs attached to type θ' players.

Each type θ player draws a private value $v \in [0, 1]$ independently from the distribution V_θ. A player's payoff depends on her private value v, her degree d, and the number of her neighbors a who adopt—the player earns $u_d(v, a)$ from adopting and 0 otherwise. Assume externalities are positive—the functions $\{u_d\}_{d \in \mathbb{N}}$ are increasing in a—and payoffs are strictly increasing and differentiable in v. Each player observes her type, her private value, and her degree. The network distribution and number of seeds are common knowledge. Importantly, a player *does not* observe the period t in which she makes her choice.

We look at perfect Bayesian equilibria in symmetric strategies. A symmetric strategy profile is a function $\sigma(\theta, v, d) : \Theta \times [0, 1] \times \mathbb{N} \to \{0, 1\}$ giving an adoption decision for each type, value, and degree. Fixing a profile σ, the probabilities

$$\sigma(\theta, d) = \mathbb{E}[\sigma(\theta, V_\theta, d)]$$

determine the distribution over outcomes and hence the players' expectations about whether neighbors will adopt. Write $\mathbb{P}_\sigma^{(\theta, d, n)}$ for the conditional probability measure from the perspective of a type θ player with degree d who gets exposed. The symmetric strategy profile σ is an equilibrium for the n player game if for all (θ, v, d) and each $x \in \{0, 1\}$, we have

$$\mathbb{E}_\sigma^{(\theta, d, n)}[\sigma(\theta, v, d) u_d(v, A)] \geq \mathbb{E}_\sigma^{(\theta, d, n)}[x u_d(v, A)],$$

where A denotes the random number of neighbors who will adopt if they become aware.

---

[10]A simple graph has no self-links and at most one link between any pair of nodes.

Our tools concern limits of the outcome distribution, so we study equilibria under limiting beliefs as $n \to \infty$. Theorem 1 in section 6 implies that the limiting distribution $\mathbb{P}_\sigma^{(\theta,d)}$ over neighbors' actions is well defined. A strategy profile is a **limit equilibrium** if for all (θ, v, d) and each $x \in \{0, 1\}$, we have

$$\mathbb{E}_\sigma^{(\theta,d)} \left[\sigma(\theta, v, d) u_d(v, A) \right] \geq \mathbb{E}_\sigma^{(\theta,d)} \left[x u_d(v, A) \right].$$

Under mild assumptions, equilibria for large n are necessarily close to limit equilibria, so for ease of exposition, all formal statements pertain to limit beliefs and limit equilibria.

Our main questions are: how many players adopt, and how long does the process take? If $X_n(t)$ is the number of adopting players at time t in an n player game, define

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

These random variables describe respectively the long-run fraction of the population that adopts and the time it takes for a fraction x of these players to adopt. We characterize the distributions α_n and τ_n for large n.

3.2   Remarks on the Model

The multi-type configuration model as presented can capture many realistic network features—power-law degree distributions, small-worlds, homophily. One missing feature is clustering: in real networks, linked individuals likely share common neighbors. Clustering is among several extensions I discuss in the appendices. The techniques I develop are robust enough to handle this along with many other adjustments.

There are several ways to interpret types. An econometrician may view them as distinct collections of observable attributes (e.g. demographic or socioeconomic variables), and we could imagine estimating the model's parameters from network data. We can also use types to encode information that players possess—appendix C.1 shows how to incorporate signals about neighbors by expanding the set of types.

I assume choices are irreversible, and players who choose not to adopt cannot reconsider. One's use of a technology or product clearly is reversible in the long run, though switching costs may render this prohibitively expensive in the short run. Since the spread of new technologies is often a slow process (Griliches, 1957; Skinner and Staiger, 2005), the "short run" is economically important. Alternatively, the diffusion of information demands a model of irreversible adoption. One cannot unread a news story or unwatch a video. In natural examples, both the decision to consume information and the decision to share information are important. I explore this in appendix C.2.

Since players do not observe whether neighbors have adopted, formally there is no reason to revisit a decision. We might interpret the model assumption as either forgetfulness or limited attention by the players. However, if players were to observe neighbors' choices, they might wait for several neighbors to adopt. In the absence of clustering, letting players wait changes little because widespread adoption depends on players who adopt after just

one neighbor does so. In a network with clustering, the option to adopt later becomes more significant. I study an example in appendix C.4.

Although timing plays almost no role in the analysis—if a player adopts, all neighbors become aware and immediately make decisions—we should be clear about how players form beliefs about time. In equilibrium, a player knows the number of seeds, the network distribution, and the equilibrium adoption propensities. This induces a distribution over whether and when the player becomes aware. When a player becomes aware, her beliefs about the time period correspond to this distribution conditional on becoming aware. This is a well-defined distribution for any finite n, but in the limit as n goes to infinity, these beliefs do not converge to a well-defined distribution. Nevertheless, the payoff relevant beliefs about whether neighbors will adopt do converge, and these are the beliefs on which we focus.

4   Mapping Strategies to Outcomes

For now, we set aside the choice problem. Fixing a profile σ, the *adoption propensities*

$$\sigma(\theta, d) = \mathbb{E}[\sigma(\theta, V_\theta, d)]$$

fully characterize the outcome distribution. I find it useful to think about three layers in the network. The outermost layer is the social network itself. Strategies determine a subnetwork of *potential adopters*—players who adopt if they get exposed. The propensities $\sigma(\theta, d)$ capture the likelihood of being in this subnetwork. The innermost layer is the network of *actual adopters*. These players are in a connected component of potential adopters linked to a seed. To characterize the extent of adoption α_n, we need to assess component sizes in the potential adopter network. To characterize diffusion times $\tau_n(x)$, we need to assess typical distances within these components.

The subnetwork of potential adopters is itself a configuration model. This is clear after conditioning on who the potential adopters are and which link stubs connect to them: each possible pairing of stubs between potential adopters is equally likely. We first determine the type and degree distributions in the network of potential adopters. We then apply our limit theorems to understand the component structure in this network.

The techniques behind our theorems provide helpful intuition. Figure 2 depicts a network in two ways. On the left is a more standard image, which we reorganize as a tree on the right. A breadth first search defines a branching process—the central node is the root, nodes at distance one are the first generation, nodes at distance two are the second, and so on. For any finite n, the offspring distributions are correlated, and they change as we get further from the root. However, in the limit as $n \to \infty$, the local structure converges in distribution to a standard branching process with independent offspring. I call this limiting distribution the **characteristic branching process**.

We realize the characteristic branching process in two stages. The root realizes a type θ according to T, a number of offspring according to D_θ, and offspring types according to Z_θ.
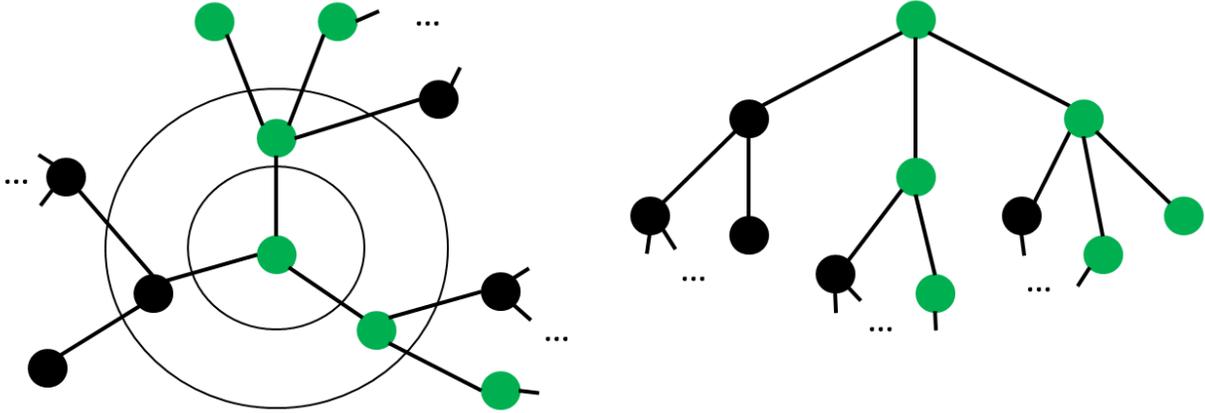
Figure 2: Two representations of the same local network structure

Subsequent nodes realize offspring according to the **forward distributions** D'_θ, defined by

$$\mathbb{P}(D'_\theta = d) = \frac{(d+1)\mathbb{P}(D_\theta = d+1)}{\mu_\theta}. \tag{4}$$

Equation (4) reweights D_θ to correct for the friendship paradox and shifts the distribution by one to avoid double counting the link back to the parent. Understanding diffusion is equivalent to understanding the characteristic branching process. The survival probability of the branching process corresponds to the size of the largest network component. This characterizes the extent of diffusion α_n. The growth rate of the branching process tells us about distances in the network. From this we obtain the diffusion times $\tau_n(x)$.

4.1   Diffusivity and Virality

To characterize the network of potential adopters, we account for selection into this network. In the limit, a random player is a type θ potential adopter with probability

$$\pi_\theta^{(\sigma)} = p_\theta \mathbb{E}\left[\sigma(\theta, D_\theta)\right]. \tag{5}$$

A random type θ potential adopter has degree drawn according to $D_{\sigma,\theta}$, which satisfies

$$\mathbb{P}(D_{\sigma,\theta} = d) = \frac{\mathbb{P}(D_\theta = d)\sigma(\theta, d)}{\sum_{k \in \mathbb{N}} \mathbb{P}(D_\theta = k)\sigma(\theta, k)}. \tag{6}$$

Each type θ link stub leads to a potential adopter with independent probability

$$\sigma_\theta \equiv \mathbb{E}\left[\sigma(\theta, D'_\theta + 1)\right],$$

where D'_θ is the forward distribution defined in (4).

16

To obtain the degree distributions in the potential adopter network, first realize link stubs according to $D_{\sigma,\theta}$. Each stub then either receives a type or gets deleted. We assign type θ' with probability $q_{\sigma,\theta'}^{(\theta)} = q_{\theta'}^{(\theta)}\sigma_{\theta'}$, and we delete the stub with probability

$$q_{\sigma,0}^{(\theta)} = 1 - \sum_{\theta'=1}^{|\Theta|} q_{\theta'}^{(\theta)}\sigma_{\theta'}.$$

Write $\mathbf{q}_\sigma^{(\theta)}$ for the $|\Theta|+1$-vector with entries $\{q_{\sigma,\theta'}^{(\theta)}\}_{\theta'=0}^{|\Theta|}$. We can think of a network with $|\Theta|+1$ types—the extra type (type 0) encompasses all non-adopters.

The probability generating functions for $D_{\sigma,\theta}$ and D_θ respectively are

$$g_{\sigma,\theta}(s) = \sum_{d=1}^{\infty} \mathbb{P}(D_{\sigma,\theta} = d)s^d \quad \text{and} \quad g_\theta(s) = \sum_{d=1}^{\infty} \mathbb{P}(D_\theta = d)s^d.$$

A simple exercise shows that the generating function for the forward distribution $D'_{\sigma,\theta}$ is $\frac{g'_{\sigma,\theta}(s)}{g'_{\sigma,\theta}(1)}$. Standard results from branching process theory imply that the system

$$g'_{\sigma,\theta}(1)s_\theta = g'_{\sigma,\theta}(\mathbf{q}_\sigma^{(\theta)} \cdot \mathbf{s}), \qquad \theta = 1, 2, ..., |\Theta|, \quad s_0 = 1 \tag{7}$$

has a unique minimal solution $\boldsymbol{\xi}^{(\sigma)} \in [0,1]^{|\Theta|+1}$.[11] This vector helps us define *diffusivities*.

Definition 1. *Fix a strategy profile σ, and let $\boldsymbol{\xi}^{(\sigma)}$ denote the minimal solution of (7). The* **diffusivity of a player with type θ and degree d** *is*

$$\zeta_{\theta,d}^{(\sigma)} = 1 - (\mathbf{q}_\sigma^{(\theta)} \cdot \boldsymbol{\xi}^{(\sigma)})^d. \tag{8}$$

The **diffusivity of a player with type θ** *is*

$$\zeta_\theta^{(\sigma)} = 1 - g_\theta(\mathbf{q}_\sigma^{(\theta)} \cdot \boldsymbol{\xi}^{(\sigma)}) \qquad \theta = 1, 2, ..., |\Theta|. \tag{9}$$

The **diffusivity of a potential adopter with type θ** *is*

$$\phi_\theta^{(\sigma)} = 1 - g_{\sigma,\theta}(\mathbf{q}_\sigma^{(\theta)} \cdot \boldsymbol{\xi}^{(\sigma)}) \qquad \theta = 1, 2, ..., |\Theta|. \tag{10}$$

The **diffusivity of the network** *is*

$$\zeta_\sigma = \boldsymbol{p} \cdot \boldsymbol{\zeta}^{(\sigma)} = \sum_{\theta=1}^{|\Theta|} p_\theta \left(1 - g_\theta(\mathbf{q}_\sigma^{(\theta)} \cdot \boldsymbol{\xi}^{(\sigma)}) \right). \tag{11}$$

The **diffusivity of the potential adopter network** *is*

$$\phi_\sigma = \boldsymbol{\pi} \cdot \boldsymbol{\phi}^{(\sigma)} = \sum_{\theta=1}^{|\Theta|} \pi_\theta \left(1 - g_{\sigma,\theta}(\mathbf{q}_\sigma^{(\theta)} \cdot \boldsymbol{\xi}^{(\sigma)}) \right). \tag{12}$$

---

[11]The appendix provides statements of the relevant results along with references.

The entries of $\boldsymbol{\xi}^{(\sigma)}$ are the **forward extinction probabilities**. In the characteristic branching process, if the root realizes a type θ offspring, that branch of the process dies with probability $\xi_\theta^{(\sigma)}$. Hence, the branch beginning from a random offspring of a type θ player dies with probability $\mathbf{q}_\sigma^{(\theta)} \cdot \boldsymbol{\xi}^{(\sigma)}$. The different diffusivities correspond to different survival probabilities conditional on particular realizations of the root node.

From the limit theorems in section 6, the diffusivities also correspond to probabilities that particular players connect to the giant component of potential adopters. Our first proposition uses ζ_σ and ϕ_σ to characterize the extent of diffusion α_n. The player diffusivities $\zeta_{\theta,d}^{(\sigma)}$, $\zeta_\theta^{(\sigma)}$, and $\phi_\theta^{(\sigma)}$ provide additional information about who adopts in a large cascade, which later allows us to assess players' beliefs. These also function as centrality measures—the next subsection shows how diffusivities can guide optimal seeding.

Proposition 1. *As n grows, the extent of diffusion α_n converges in distribution to a random variable α taking the value 0 with probability $(1 - \zeta_\sigma)^\kappa$ and the value ϕ_σ otherwise.*

Proof. This is immediate from Proposition 16 in the appendix. $\qquad\square$

Proposition 1 tells us how far adoption spreads in the network. Two features stand out. First, even in the large network limit, the extent of adoption is stochastic. For generic parameters, there is a positive probability that diffusion quickly halts, and few individuals ever get exposed. Second, more seeds does not mean more adoption as α only ever takes one positive value. This value depends on the network structure and the adoption propensities, but not the number of seeds. An additional seed increases the likelihood of a large cascade but not its size.

These features are intuitive if we recall what diffusivity represents: ζ_σ is the probability a random player links to the giant component, and the size of this component is ϕ_σ. Each seed links with probability ζ_σ, causing the entire component to adopt. With probability $(1 - \zeta_\sigma)^\kappa$, all seeds miss the giant component, and all other components are negligible in size. We also know how many people get exposed in a large cascade: a fraction ζ_σ of players link to an adopter while a fraction ϕ_σ actually adopt.

To define virality, we first define the mean offspring matrix for the characteristic branching process. For each type θ potential adopter, the average forward degree is

$$\nu_{\sigma,\theta} = \frac{\mathbb{E}\left[D_{\sigma,\theta}(D_{\sigma,\theta} - 1)\right]}{\mathbb{E}[D_{\sigma,\theta}]} = \frac{\mathrm{Var}[D_{\sigma,\theta}]}{\mathbb{E}[D_{\sigma,\theta}]} + \mathbb{E}[D_{\sigma,\theta}] - 1. \tag{13}$$

We then write M_σ for the $|\Theta|$ by $|\Theta|$ matrix with entries

$$m_{ij}^{(\sigma)} = \nu_{\sigma,i} q_{\sigma,j}^{(i)}. \tag{14}$$

After diffusion reaches a type i player, the entry $m_{ij}^{(\sigma)}$ is the average number of type j neighbors who are potential adopters.

Definition 2. *The **virality** ν_σ of the network is the spectral radius of M_σ.*

Proposition 2. *We have $\zeta_\sigma > 0$ if and only if $\nu_\sigma > 1$. In this case, for any $x \in (0,1)$ and any sufficiently small $\epsilon > 0$ we have*

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

Proof. This is immediate from Theorem 2. $\qquad\square$

In a large cascade, the time it takes to reach any fraction of eventual adopters is approximately $\log_{\nu_\sigma} n$. Put differently, most adoption happens in a narrow time window around $\log_{\nu_\sigma} n$. Virality is the growth rate of the characteristic branching process: the number of offspring in each generation grows like ν_σ^t, so it takes pproximately $\log_{\nu_\sigma} n$ generations to get n total progeny. In the early periods, the number of adopters grows as in the branching process. Once a significant fraction of the population adopts, diffusion finishes quickly.

The first part highlights a connection between diffusivity and virality: diffusivity is positive only if virality is sufficiently high. Beyond this, the two may not move in tandem. Virality depends only on the first two moments of the degree distribution, but diffusivity depends on the entire distribution. Virality typically moves in the direction a mean-field model would predict. Ass we saw in the last section, diffusivity may move in the opposite direction.

4.2   A New Approach to Optimal Seeding

I pause here to highlight an application that does not depend on strategic analysis. When attempting to spread a new product or technology, a planner can leverage social influence by targeting key individuals with subsidies or other incentives to adopt. An important question is then how best to select these "seeds." Standard approaches (e.g. Kempe et al. (2003)) face many challenges. They rely on detailed network data and are sensitive to small changes. This is an issue because in most applications we do not have complete data on the relevant network, and the data we do have is noisy. Even with perfect data, finding the optimal seeds is computationally demanding.

The last section suggests a different approach. More often, we have data on demographic or other attributes along with some incomplete information about social ties. We can imagine a player's type θ as a vector of observables. As long as we have data on:

(a) Neighbor adoption propensities σ_θ,

(b) Degree distributions D_θ, and

(c) Neighbor type distributions Z_θ,

then we can compute a vector of diffusivities $\zeta_\theta^{(\sigma)}$ and target individuals based on the observables θ. To simplify further, one might also impose parametric restrictions on the degree distributions or neighbor type distributions. Targeting high-diffusivity types is obviously less effective than what we could achieve through optimal seeding given perfect information

about the network. However, this approach has clear advantages. We can say something potentially useful based on more limited information about the network. Moreover, solving for the diffusivities is far more computationally tractable—it entails numerically solving a system of $|\Theta|$ equations in $|\Theta|$ unknowns.

Consider the following example with three types $\theta \in \{l, m, h\}$. Type l will never adopt ($\sigma_l = 0$), type h will always adopt ($\sigma_h = 1$), and type m adopts with probability $\frac{1}{2}$ ($\sigma_m = \frac{1}{2}$). Suppose all types have 5 neighbors for sure, and we have

$$q_l^{(m)} = q_h^{(m)} = q_m^{(l)} = q_m^{(h)} = \epsilon, \qquad q_l^{(h)} = q_h^{(l)} = 1 - \epsilon, \qquad q_m^{(m)} = 1 - 2\epsilon$$

for some small ϵ. That is, types l and h preferentially link to each other and form occasional links with type m, while type m predominantly forms own-type links.

For small values of ϵ, the giant component is almost entirely comprised of type m players, and these are the ones to target for seeding. While type h players are more likely to adopt, and type l players have neighbors who are more likely to adopt, cascades from these types are sure to end within a few steps of the seed. In contrast, starting from a type m individual, we get in expectation two new adopters at each step. These assumptions are extreme, but they highlights in stark terms that the correlation structure between neighbor types matters, and we can productively use this information to guide seeding.

4.3   Comparative Statics

We now explore how the extent and rate of diffusion vary with the degree distributions $\{D_\theta\}$, the neighbor type distributions $\{Z_\theta\}$, and the adoption propensities $\sigma(\theta, d)$. What matters is how these jointly affect the network of potential adopters. Let $\{D_\theta, Z_\theta, \sigma(\theta, d)\}$ and $\{\hat{D}_\theta, \hat{Z}_\theta, \hat{\sigma}(\theta, d)\}$ denote two distinct sets of parameters—variables with a hat relate to the second set. In this section, we fix the collection of types Θ and the type distribution T.

When the network of potential adopters gets more dense, we clearly get higher diffusivity and virality. However, the degree distribution in the network of potential adopters depends non-trivially on all three fundamental parameters. Proposition 3 parses the effects of different changes. I abuse notation slightly, letting D denote both a degree distribution and a cumulative distribution function. I write D FOSD $\hat{D}$ to mean $D(m) \le \hat{D}(m)$ for all $m \in \mathbb{N}$.

Proposition 3. *Any of the following changes results in higher diffusivity and virality (i.e. $\zeta_\sigma \ge \hat{\zeta}_{\hat{\sigma}}$ and $\nu_\sigma \ge \hat{\nu}_{\hat{\sigma}}$).*

(a) *We increase adoption propensities, holding all else fixed (i.e. $\sigma(\theta, d) \ge \hat{\sigma}(\theta, d)$).*

(b) *Adoption propensities are non-decreasing in d, and we increase network density, holding all else fixed (i.e. D_θ FOSD $\hat{D}_\theta$ for each type θ).*

(c) *We increase network density, and neighbor adoption probabilities satisfy*

$$(1 - \hat{q}_{\sigma,\theta'}^{(\theta)})^{\lceil \hat{D}_\theta^{-1}(D_\theta(m)) \rceil} \ge (1 - q_{\sigma,\theta'}^{(\theta)})^m \tag{15}$$

for each $m \in \mathbb{N}$ and each pair of types θ and θ'.

Proof. See Appendix. $\qquad\square$

Condition (a) is intuitive: adding potential adopters causes adoption to spread further and faster. Condition (b) looks at changes in the total degree distributions $\{D_\theta\}_{\theta\in\Theta}$. As long as adoption propensities weakly increase in degree, a denser network leads to higher ζ_σ and ν_σ. If the propensities decrease in degree, this could offset the additional connectivity. Condition (c) characterizes the tradeoff between adding links and increasing adoption propensities. If D_θ FOSD $\hat{D}_\theta$, then the exponent $\lceil \hat{D}_\theta^{-1}(D_\theta(m)) \rceil$ is (weakly) less than m, providing room to have $\hat{q}_{\sigma,\theta'}^{(\theta)} > q_{\sigma,\theta'}^{(\theta)}$ while still satisfying the condition.

Changes in degree variance have ambiguous effects on the extent of diffusion, but a clear effect on the rate. Equation (13) shows that average forward degrees are increasing in the variance of the degree distributions, so mean preserving spreads tend to increase virality.

Proposition 4. *If any of the following conditions hold, we have $\nu_\sigma \geq \hat{\nu}_{\hat{\sigma}}$:*

(a) *We have $m_{ij}^{(\sigma)} \geq \hat{m}_{ij}^{(\hat{\sigma})}$ for all $i, j \in \Theta$*

(b) *We have that $\sigma_{\theta,d} = \hat{\sigma}_{\theta,d}$ is weakly convex in d, that $Z_\theta = \hat{Z}_\theta$ for each type θ, and D_θ is a mean preserving spread of $\hat{D}_\theta$ for each type θ*

Proof. Part (a) is trivial. For part (b), convexity implies that a mean-preserving spread results in (weakly) higher adoption propensities $\sigma_{\theta'}^{(\theta)}$. From the definition, a mean preserving spread also increases ν_θ. The result follows from (14) together with part (a). $\qquad\square$

We can interpret higher variance degrees as higher network centralization: a smaller number of players account for a larger fraction of total links. This reduces path lengths, allowing adoption to spread more quickly. However, more low-degree players may also mean more isolated components. Recalling the example in Section 2.4, Proposition 4 implies we get faster diffusion after a mean preserving spread, but the extent of diffusion may decline.

5   Limit Beliefs and Viral Belief Distortion

I distinguish two types of strategy profiles—**non-viral** and **viral** profiles. In a non-viral profile, we have $\nu_\sigma \leq 1$, which means adoption can never spread beyond a tiny fraction of the population. In a viral profile, we have $\nu_\sigma > 1$, which means that cascades covering a positive fraction of players are possible. The distinction has implications for players' beliefs. In the former case, an exposed player knows exactly one neighbor has adopted, and the remaining neighbors have types and degrees drawn independently from the forward distributions. In the latter case, exposure conveys more information. If adoption fails to "go viral," the asymptotic probability that any given player learns about the product is zero. When adoption does go viral, this probability is strictly positive. Consequently, conditional on exposure, a player knows there is a large cascade, and she is connected to the giant component—I call this adjustment **viral belief distortion**.

Viral belief distortion skews beliefs about neighobrs towards higher degrees and higher adoption propensities—we essentially condition the characteristic branching process on survival. Recall the forward extinction probabilities $\{\xi_\theta^{(\sigma)}\}$ and the adoption-weighted neighbor type probabilities $\{q_{\sigma,\theta'}^{(\theta)}\}$. A type θ player with degree d is connected to the giant component with probability $\zeta_{\theta,d}^{(\sigma)} = 1 - (\mathbf{q}_\sigma^{(\theta)} \cdot \boldsymbol{\xi}^{(\sigma)})^d$, and hence a type θ player is connected to this component with probability $\zeta_\theta^{(\sigma)} = 1 - g_\theta(\mathbf{q}_\sigma^{(\theta)} \cdot \boldsymbol{\xi}^{(\sigma)})$. To obtain players' beliefs, we reweight Z_θ and D_θ to account for the selection effect.

Proposition 5 (Viral Belief Distortion). *If $\nu_\sigma \leq 1$, a type θ player who gets exposed believes that each of her neighbors has type drawn independently according to Z_θ and, conditional on realizing type θ', degree drawn according to $D_{\theta'}' + 1$.*

If $\nu_\sigma > 1$, a type θ player who gets exposed believes that each of her neighbors has type drawn independently according to $\hat{Z}_\theta$ and, conditional on realizing type θ', degree drawn according to $\hat{D}_{\theta'}' + 1$, where

$$\mathbb{P}(\hat{Z}_\theta = \theta') = \frac{q_{\theta'}^{(\theta)} \zeta_{\theta'}^{(\sigma)}}{\sum_{\tilde{\theta} \in \Theta} q_{\tilde{\theta}}^{(\theta)} \zeta_{\tilde{\theta}}^{(\sigma)}}, \quad and \quad \mathbb{P}(\hat{D}_\theta' = d) = \frac{\mathbb{P}(D_\theta' = d)\zeta_{\theta,d+1}^{(\sigma)}}{\sum_{k \in \mathbb{N}} \mathbb{P}(D_\theta' = k)\zeta_{\theta,k+1}^{(\sigma)}}.$$

Proof. The first part is clear as Z_θ and D_θ' correspond to the offspring distribution of the characteristic branching process—we add one to D_θ' because here we count the link back to the parent node. The distributions $\hat{Z}_\theta$ and $\hat{D}_\theta$ reweight the distributions Z_θ and D_θ' according to the probability that a particular type and degree of player is connected to the giant component in the network of potential adopters. $\square$

Some implications of Proposition 5 may appear counterintuitive. Adoption fails to go viral with positive probability, yet in this event, every exposed player has *systematically incorrect beliefs* about her neighbors. This happens because the number of exposed players in a non-viral realization is tiny compared with that in a viral realization. This suggests a reason why users of a niche product might be unrealistically optimistic about the product becoming popular—players reason "if I found out about it, it must be catching on."

The size of this effect on beliefs is largest when diffusivity is small—if few people are in the giant component, then being in this component conveys more information. The ratio

$$\frac{1 - (\mathbf{q}_\sigma^{(\theta)} \cdot \boldsymbol{\xi}^{(\sigma)})^d}{1 - (\mathbf{q}_\sigma^{(\theta)} \cdot \boldsymbol{\xi}^{(\sigma)})^{d'}}$$

captures the relative likelihood that two players of the same type with degrees d and d' are connected to the giant component. When diffusivitiy is small, the value $\mathbf{q}_\sigma^{(\theta)} \cdot \boldsymbol{\xi}^{(\sigma)}$ is close to one. If $d > d'$, the ratio grows with $\mathbf{q}_\sigma^{(\theta)} \cdot \boldsymbol{\xi}^{(\sigma)}$. As the giant component gets smaller, beliefs become more skewed towards neighbors with high degrees. This creates a discontinuity in players' beliefs as we move from non-viral strategy profiles to viral strategy profiles.

6 Random Graphs that Resemble Reality

This section presents the technical contribution: a characterization of component sizes and typical distances in a multi-type configuration model. Real networks, from friendships and coauthorships to webpage links, share several structural regularities. Among these are power-law degree distributions, the "small-worlds" property, clustering, and homophily.[12] The configuration model and its variants can generate all of these features.

While these random graphs are complicated objects—with overlapping neighborhoods, cycles, and multiple disjoint components—a limit theorem ties the network structure to a *characteristic branching process*. This is a far simpler object, amenable to standard tools for studying sums of independent random variables. I present a slightly more general version of the model than what I used earlier—here I do not assume independence of neighbor types. The Law of Large Networks tells us that the local structure of the network converges to the distribution of the characteristic branching process. Moreover, all "large" components are connected in one "giant" component. The Typical Distances theorem tells us that path lengths between random nodes in the giant component concentrate around $\log n$, where the base of the logarithm corresponds to the growth rate of the characteristic branching process. Proofs, along with useful corollaries and extensions, are given in an appendix.

6.1 The Multi-type Configuration Model

To define the model, we need a finite set of types $\Theta = \{1, 2, ..., |\Theta|\}$, a type distribution $T \in \Delta(\Theta)$, and degree distributions $\mathbf{D} = \{D^{(\theta)}\}_{\theta \in \Theta}$ for each type. Here, a degree is not simply a number but a $|\Theta|$-tuple $d = (d_1, d_2, ..., d_{|\Theta|})$ giving the number of links to each type of node. The only assumption we impose on $\mathbf{D}$ is that each component has finite expectation and finite variance.[13] For each of n nodes, we independently draw a type from T. Each type θ node draws a degree tuple independently from $D^{(\theta)}$—imagine assigning link stubs to each vertex, and each stub has a type. We then pair compatible stubs together uniformly at random to form links: for each type θ stub connected to a type θ' node, we draw, from those not yet paired, a type θ' stub connected to a type θ node.

Given n-vectors of types $\boldsymbol{\theta} = (\theta^{(1)}, \theta^{(2)}, ..., \theta^{(n)})$ and degree tuples $\mathbf{d} = (d^{(1)}, d^{(2)}, ..., d^{(n)})$, we say the pair $(\boldsymbol{\theta}, \mathbf{d})$ is **consistent** if

(a) For each pair of types $\theta \neq \theta'$, the total number of type θ stubs attached to type θ' nodes equals the number of type θ' stubs attached to type θ nodes, and

(b) For each type θ, the number of type θ stubs attached to type θ nodes is even.

When we draw $(\boldsymbol{\theta}, \mathbf{d})$ according to $(T, \mathbf{D})$, we condition on realizing a consistent pair. Limit results require a consistency condition on $(T, \mathbf{D})$. Write $\mu_{\theta'}^{(\theta)}$ for the expected number of

---

[12]Barabási and Albert (2002) document the ubiquity of power-law degree distributions, small worlds, and clustering across a variety of networks, including links between webpages, actor collaborations, academic citations, and human sexual contacts. McPherson et al. (2001) review literature on homophily in social networks, and Ugander et al. (2011) find all four features in the Facebook graph.

[13]Finite variance is only necessary for Theorem 2.

type θ' stubs that $D^{(\theta)}$ realizes, and define $p_\theta = \mathbb{P}(T = \theta)$. The pair $(T, \mathbf{D})$ is **consistent in expectation** if for each pair of types θ and θ', we have $p_\theta \mu_{\theta'}^{(\theta)} = p_{\theta'} \mu_\theta^{(\theta')}$. This ensures that the distributions generate the same expected number of stubs on each side of a link. We condition the matching of link stubs on realizing a simple graph—one with at most one link between any pair of nodes and no self links. The following definition summarizes the multi-type configuration model.

Definition 3 (Multi-type Configuration Model). *Given n nodes and a consistent pair* $(\boldsymbol{\theta}, \mathbf{d})$ *of types and degree tuples, we realize the* ***multi-type configuration model with type sequence $\boldsymbol{\theta}$ and degree sequence $\mathbf{d}$***, *written* $CM_{n,\boldsymbol{\theta},\mathbf{d}}$, *as follows. Assign node i type $\theta^{(i)}$ and degree tuple $d^{(i)}$. Successively pick pairs of compatible link stubs uniformly at random, join the two stubs to form a link, and repeat this until all stubs are matched. Condition this process on realizing a simple graph.*

    Given n nodes and distributions $(T, \mathbf{D})$ that are consistent in expectation, we realize the ***multi-type configuration model with distribution $(T, \mathbf{D})$***, *written* $CM_{n,T,\mathbf{D}}$, *as follows. Generate a pair $(\boldsymbol{\theta}, \mathbf{d})$ according to $(T, \mathbf{D})$. Conditional on realizing a consistent pair, generate a graph according to $CM_{n,\boldsymbol{\theta},\mathbf{d}}$.*

Example: The One-Type Configuration Model

    Suppose Θ is a singleton. The collection $\{D^{(\theta)}\}_{\theta \in \Theta}$ reduces to a single distribution $D \in \Delta(\mathbb{N})$ assigning the total number of links to each node. Even in this simple model, different distributions D capture a wide range of network structures. This model is widely applied, and earlier results characterize its structure. Two key questions are:

(a) For large n, is there a giant component?

(b) What is the typical path length between two connected nodes?

If $L_1(G)$ is the largest connected component in G, we could rephrase (a) to ask: is the limit of $\frac{|L_1(G)|}{n}$ positive? The results of Molloy and Reed (1995) tell us that it is if and only if

$$\nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]} > 1.$$

Moreover, we can compute the limit using the probability generating function g of D. Writing ζ for the limiting value of $\frac{|L_1(G)|}{n}$, we have $\zeta = 1 - g(\xi)$, where ξ is the minimal solution to

$$g'(1)\xi = g'(\xi).$$

When a giant component exists, van der Hofstad et al. (2005) show that the distance between two random nodes in this component concentrates around $\log_\nu n$ as n approaches infinity. The theorems in the next subsection generalize both findings to the multi-type model, allowing us to analyze networks of heterogeneous agents with correlated linking propensities.

Example: Two types with Homophily

Suppose Θ contains two types a and b, and T assigns equal probability to each. Fix $h \in [0,1]$. To define $D^{(a)}$, draw a degree from D and label each link type a with independent probability $\frac{1+h}{2}$—label a link type b otherwise. Define $D^{(b)}$ similarly, drawing a degree from D and labeling each link type b with probability $\frac{1+h}{2}$. The parameter h captures homophily. When $h = 0$, the network is structurally equivalent to a one-type configuration model with degree distribution D. When $h > 0$, nodes form more links with others of the same type. When $h = 1$, the two types are completely separated into two disjoint networks.

6.2   Branching Process Approximation

Given $(T, \mathbf{D})$, we define the **characteristic branching process** $\mathcal{T}_{T,\mathbf{D}}$ in two stages. The root node draws a type $\theta \in \Theta$ from T and offspring from $D^{(\theta)}$. Each subsequent node draws offspring independently from a forward distribution $D'^{(\theta,\theta')}$ that depends on both the node's type and its parent's type. A type θ node with a type θ' parent realizes the tuple of offspring $d = (d_1, d_2, ..., d_\Theta)$ with probability

$$\mathbb{P}(D'^{(\theta\theta')} = d) = \frac{(d_{\theta'} + 1)\mathbb{P}(D^{(\theta)} = d + e_{\theta'})}{\mu_{\theta'}^{(\theta)}}, \tag{16}$$

where e_θ is a vector containing a one in entry θ and zeros elsewhere. As before, we correct for the friendship paradox and subtract the link back to the parent.

Definition 4 (Characteristic Branching Process). *Given the distributions $(T, \mathbf{D})$, we realize the **characteristic branching process** $\mathcal{T}_{T,\mathbf{D}}$ as follows. A single root node realizes a type θ according to T and offspring according to $D^{(\theta)}$. Every subsequent type θ node with a type θ' parent realizes offspring independently according to the forward distribution $D'^{(\theta\theta')}$.*

To illustrate, I construct the branching processes for the one-type model and the two-type model with homophily. For the one-type model, the forward distribution D' is

$$\mathbb{P}(D' = d) = \frac{(d + 1)\mathbb{P}(D = d + 1)}{\mathbb{E}[D]}. \tag{17}$$

The root of the characteristic branching process realizes offspring via D, and all subsequent nodes realize offspring independently via D'.

In the example with two types, independent labeling of link stubs implies $D'^{(aa)} = D'^{(ab)} \equiv D'^{(a)}$ and $D'^{(ba)} = D'^{(bb)} \equiv D'^{(b)}$. The forward distributions realize total degree according to D' as in (17), and each link is an own-type link with independent probability $\frac{1+h}{2}$. The root of the characteristic branching process realizes type a or b with equal probability, realizes offspring according to D, and each offspring has the same type with independent probability $\frac{1+h}{2}$. Each subsequent node realizes offspring independently according to the forward distribution, either $D'^{(a)}$ or $D'^{(b)}$, corresponding to its type.

The characteristic branching process approximates a breadth first search of the network starting from a random node. In a finite network, the offspring distributions in this search process are not independent, and they change as we get further from the root. At some point offspring link to one another. The essence of the first theorem is that these complications are asymptotically insignificant: as n grows, the local structure of the configuration model converges to the that of the characteristic branching process.

Given a graph G, let $N_k(G)$ denote the number of nodes in components of size k, and let $L_i(G)$ denote the ith largest component. Write $\mathbb{P}_{CM_{n,T,\mathbf{D}}}$ for the probability measure over graphs defined by the configuration model $CM_{n,T,\mathbf{D}}$. Define $\rho_k = \mathbb{P}\left(|\mathcal{T}_{T,\mathbf{D}}| = k\right)$, and note that ρ_∞ is the survival probability of $\mathcal{T}_{T,\mathbf{D}}$.

Theorem 1 (Law of Large Networks)**.** *Suppose $\mathcal{T}_{T,\mathbf{D}}$ is irreducible and non-singular.[14] For any $\epsilon > 0$, we have*

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

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

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

Proof. See Appendix. □

Theorem 1 is a central contribution of this paper. The first claim states that the number of nodes in components of size k converges in probability to ρ_k: the distribution of component sizes converges to the law of total progeny for $\mathcal{T}_{T,\mathbf{D}}$. The last two claims state that there is at most one giant component, and it covers a fraction of the network equal to the survival probability of the characteristic branching process. Theorem 1 extends existing results for the one-type configuration model, allowing us to study random graphs with heterogeneous nodes and linking propensities. The proof in fact implies a much stronger result than what is stated. The characteristic branching process captures any "local" property of the configuration model, not just the distribution of component sizes.

If there is a giant component, we can also use the characteristic branching process to assess distances between nodes within it. Recall that $D'^{(\theta\theta')}$ denotes the offspring distribution for a type θ node with a type θ' parent. Write

$$\nu_{\tilde\theta}^{(\theta\theta')} = \mathbb{E}\left[D_{\tilde\theta}^{(\theta\theta')} \right]$$

for the expected number of type $\tilde\theta$ offspring of a type θ node with a type θ' parent. Define the mean offspring matrix $M_{T,\mathbf{D}}$ of $\mathcal{T}_{T,\mathbf{D}}$ as the $|\Theta|^2$ by $|\Theta^2|$ matrix with entries $\nu_{\tilde\theta}^{(\theta\theta')}$— each row and column corresponds to a particular pair of types (θ, θ'). The entry for row

---

[14]Irreducible means that any type of node has all types of offspring at some distance with positive probability. Equivalently, in the configuration model, a path exists between any two types with positive probability. Non-singular means that it is not the case that all types have a single offspring with probability one.

(θ_r, θ'_r) and column (θ_c, θ'_c) is equal to $\nu_{\theta_c}^{(\theta_r \theta'_r)}$ if $\theta'_c = \theta_r$ and 0 otherwise.[15] The entries of the mean offspring matrix record the average number of offspring for each type of node in the characteristic branching process $\mathcal{T}_{T,\mathbf{D}}$.

The spectral radius $\nu_{T,\mathbf{D}}$ of $M_{T,\mathbf{D}}$ characterizes typical distances. Given a graph G, let $H(G)$ denote the length of the shortest path between two vertices chosen uniformly at random from $L_1(G)$. Recall that ρ_∞ is the survival probability of $\mathcal{T}_{T,\mathbf{D}}$.

Theorem 2 (Typical Distances). *Suppose $\mathcal{T}_{T,\mathbf{D}}$ is irreducible and non-singular. We have $\rho_\infty > 0$ if and only if $\nu_{T,\mathbf{D}} > 1$. Moreover, for any $\epsilon > 0$, we have*

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

Proof. See Appendix. $\qquad\square$

Theorem 2 gives a simple condition to check if a giant component exists, and when one does, the distance between two random nodes in this component concentrates near $\log_{\nu_{T,\mathbf{D}}} n$.[16] This illustrates the small-world property: typical distances grow slowly in network size. The result intuitively relates to how a branching process grows over time. Standard results imply that the the number of offspring in successive generations grows exponentially with base $\nu_{T,\mathbf{D}}$. This means the number of generations it takes to get n total progeny is roughly $\log_{\nu_{T,\mathbf{D}}} n$.

Since the forward distributions typically depend on both a node's type and its parent's type, a configuration model with $|\Theta|$ types has a characteristic branching process with $|\Theta|^2$ types. As a practical matter, this presents a problem for computational tractability. This is why in the main body of the paper, I assume the distribution $D^{(\theta)}$ is generated by first realizing a degree according to D_θ and then realizing types independently according to Z_θ. With independent stub types, the parents' type does not influence the forward distribution, and we can reduce the characteristic branching process to one with only $|\Theta|$ types.

7 Final Remarks

Diffusion phenomena are pervasive, and they are relevant in a range of economic applications. Because social ties play an important role in spreading information, the structure of a social network can influence diffusion patterns. Research is making strides in understanding the role of network structure, but existing theory faces several challenges. The tools in this paper allow us to study diffusion with a more faithful representation of the underlying network. We can capture viral phenomena, short-run dynamics, the importance of central individuals, and the strategic effects that emerge when individuals are not myopic.

---

[15]The zeros enter because a type θ_r can only have offspring with type θ_r parents.

[16]van der Hofstad et al. (2005) prove an analogous result for typical distances in the one-type configuration model. The proof in the appendix is substantially different from theirs, taking advantage of the newer techniques used to prove Theorem 1. van der Hofstad et al. (2007) give a similar result for the infinite variance case in the one-type configuration model, showing that typical distances are on the order of $\log \log n$.

Diffusion games give a new account of the relationship between network structure, incentives, and strategic effects. When large cascades are possible, exposure conveys information. This information affects strategies and outcomes, facilitating coordination among players. In addressing key shortcomings of mean-field models, we discover more nuanced adoption patterns. The extent and speed of diffusion can move in different diretcions, and standard comparative statics can reverse. Importantly, the model primitives have a clear link to the kinds of data empirical researchers gather. Our framework offers a direct way to use demographic and network information in tandem to inform seeding strategies or other interventions. Beyond these contributions, the law of large networks and its extensions may prove useful in other contexts. By giving us a meaningful way to characterize "generic" graphs, these tools enable a new conceptual approach for doing network based theory.

The present paper's limitations suggest many avenues for future work. While partly addressed in appendices, we should desire a more complete treatment of networks with clustering, players who can wait to adopt, and the choice to share information separate from the choice to adopt. The same essential techniques can also facilitate systematic studies of optimal seeding and referral program design. We should view the present paper as the first step in a larger research agenda to explore these problems in discrete network models.

References

Akbarpour, Mohammad, Suraj Malladi, and Amin Saberi (2017), "Diffusion, Seeding, and the Value of Network Information." Working Paper.

Allcott, Hunt and Mathew Gentzkow (2017), "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives*, 31, 211–236.

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

Bakshy, Eytan, Brian Karrer, and Lada Adamic (2009), "Social Influence and the Diffusion of User-Created Content." In *Proceedings of the 10th ACM Conference on Electronic Commerce*, 325–334.

Bakshy, Eytan, Itamar Rosenn, Cameron Marlow, and Lada Adamic (2012), "The Role of Social Networks in Information Diffusion." In *Proceedings of the 21st International Conference on World Wide Web*, 519–528.

Bandiera, Oriana and Imran Rasul (2006), "Social Networks and Technology Adoption in Northern Mozambique." *The Economic Journal*, 116, 869–902.

Banerjee, Abhijit, Emily Breza, Arun Chandrasekhar, Esther Duflo, and Matthew Jackson (2012), "Come Play With Me: Experimental Evidence of Information Diffusion about Rival Goods." Working Paper.

Banerjee, Abhijit, Emily Breza, Arun Chandrasekhar, and Ben Golub (2017), "Combating Rumors: Evidence from a Field Experiment During the Indian Demonetization." Working Paper.

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.

Barabási, A. and R. Albert (2002), "Statistical mechanics of complex networks." *Reviews of Modern Physics*, 74, 47–97.

Bass, Frank (1969), "A New Product Growth Model for Consumer Durables." *Management Science*, 15, 215–227.

Beaman, Lori, Ariel BenYishay, Jeremy Magruder, and Ahmed Mushfiq Mobarak (2015), "Can Network Theory-Based Targeting Increase Technology Adoption?" Working Paper.

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.

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.

Comin, Diego, Mikhail Dmitriev, and Esteban Rossi-Hansberg (2013), "The Spatial Diffusion of Technology." NBER Working Paper No. 18534.

Comin, Diego and Bart Hobijn (2004), "Cross-Country Technology Adoption: Making the Theories Face the Facts." *Journal of Monetary Economics*, 51, 39–83.

Comin, Diego and Bart Hobijn (2010), "An Exploration of Technology Diffusion." *American Economic Review*, 100, 2031–2059.

Conley, Timothy and Christopher Udry (2010), "Learning About a New Technology: Pineapple in Ghana." *American Economic Review*, 100, 35–69.

Dupas, Pascaline (2014), "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment." *Econometrica*, 82, 197–228.

Durrett, Rick (2007), *Random Graph Dynamics*. Cambridge University Press.

Foster, Andrew and Mark Rosenzweig (1995), "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *The Journal of Political Economy*, 103, 1176–1209.

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

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.

Granovetter, Mark (1978), "Threshold Models of Collective Behavior." *American Journal of Sociology*, 83, 1420–1443.

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 (2008), *Social and Economic Networks*. Princeton University Press.

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.

Kempe, David, Jon Kleinberg, and Eva Tardos (2003), "Maximizing the Spread of Influence through a Social Network." *KDD Conference Proceedings*.

Klenke, Achim and Lutz Mattner (2010), "Stochastic Ordering of Classical Discrete Distributions." *Advances in Applied Probability*, 42, 392–410.

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 (2006), "Contagion and Coordination in Random Networks." *International Journal of Game Theory*, 34, 371–381.

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

López-Pintado, Dunia (2012), "Influence Networks." *Games and Economic Behavior*, 75, 776–787.

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.

Munshi, Kaivan (2004), "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution." *The Journal of Development Economics*, 73, 185–213.

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.

Ugander, J., B. Karrer, L. Backstrom, and C. Marlow (2011), "The Anatomy of the Facebook Social Graph." Working Paper.

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.

I divide the appendix into three sections. The first appendix presents background material, including standard material on branching processes and the basic configuration model, as well as a few not-so-standard results on the configuration model. The second appendix proves results in the paper, beginning with Theorems 1 and 2 before giving useful corollaries and proofs for the remaining results. The last appendix considers extensions to include local information, choosing to share information, global network externalities, and clustering.

A Background Material

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

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

We can use the generating function to characterize the extinction probability.

Proposition 6. *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 7. *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 require generalizations of these results for multi-type branching processes. Suppose there are r types of individuals in the population. Let $\mathbf{Z}_n$ denote 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 $\mathbf{Z}_0$, 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.

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

Proposition 8. *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 M characterizes the growth rate of the process.

Proposition 9. *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 The Configuration Model

This section focuses mainly on the basic configuration model, meaning the model with a single type. Nodes realize link stubs according to the distribution D, and these are paired uniformly at random. Unlike in the main paper, here *do not* condition on realizing a simple graph—we allow self links and multiple links between the same pair of nodes. This is because proofs of the main theorems start from this version of the configuration model, and we later translate the results to the model conditioned on realizing a simple graph. Write $CM_{n,\mathbf{d}}$ for the basic configuration model with degree sequence $\mathbf{d}$, and write $CM_{n,D}$ for the model where $\mathbf{d}$ is realized via n independent draws from D.

Many results concern limits as n approaches infinity. 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

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

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.[17] For a sequence $\{\mathbf{d}^{(n)}\}_{n \in \mathbb{N}}$, we always assume that (a) and (b) hold, and we

---

[17]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 $n-1$.

write D for the limiting distribution. When degrees are realized independently according to D, the strong law of large numbers implies that (a) and (b) hold almost surely in the limit.

Conditional on realizing a simple graph, the model $CM_{n,\mathbf{d}}$ is equivalent to taking a uniform draw among all simple graphs with the given degree sequence.

Proposition 10. *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. $\qquad\square$

The next result helps translate findings when we condition on a simple graph. In essence, we 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

$$\lim_{n\to\infty} \mathbb{P}_{CM_{n,\mathbf{d}}}\left(|f(G) - \mu| \geq \epsilon\right) = 0, \tag{18}$$

but we need to show that

$$\lim_{n\to\infty} \mathbb{P}_{CM_{n,\mathbf{d}}}\left(|f(G) - \mu| \geq \epsilon \,\big|\, G \text{ is simple}\right) = 0. \tag{19}$$

Since the probability of realizing a simple graph declines subexponentially in n, if we can establish exponential concentration bounds on (18), the bounds translate directly to (19).

Proposition 11. *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). $\qquad\square$

Corollary 1. *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

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

Proof. Choose $\gamma = \frac{\delta}{2}$ in Proposition 11. We have for large n

$$\mathbb{P}_{CM_{n,\mathbf{d}^{(n)}}}\left(|f(G) - \mu| \geq \epsilon \,\big|\, G \text{ is simple}\right) \leq \frac{\mathbb{P}_{CM_{n,\mathbf{d}^{(n)}}}\left(|f(G) - \mu| \geq \epsilon\right)}{\mathbb{P}_{CM_{n,\mathbf{d}^{(n)}}}\left(G \text{ is simple}\right)}$$

$$< \frac{e^{-\delta n}}{e^{-\frac{\delta}{2}n}} = e^{-\frac{\delta}{2}n}.$$

$\qquad\square$

Conditioning on simple graphs does not distort the limiting degree distribution.

Proposition 12. *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*

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

Proof. The corresponding inequality without conditioning on simple G is immediate from the Dvoretzky-Kiefer-Wolfowitz inequality. The result then follows from Corollary 1. □

A complication 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 adopt if they get exposed. The structure of this subnetwork, rather than that of the broader network, determines diffusion patterns. Consequently, we require the following non-standard result.

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 13. *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. 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. $\square$

Proposition 13 tells us that selecting a subgraph of $CM_{n,D}$ according to $\mathbf{q}$ is asymptotically equivalent to the model $CM_{n,D_{\mathbf{q}}}$, so any structural results for the configuration model apply to these subgraphs—we simply replace the degree distribution D with $D_{\mathbf{q}}$.[18]

For the multi-type configuration model, direct analogs to Propositions 10, 11, 12, and 13, and Corollary 1 hold via essentially identical arguments once we impose appropriate conditions on the sequence $(\boldsymbol{\theta}, \mathbf{d})^{(n)}$. Define $p_\theta = \mathbb{P}(T = \theta)$ and $p_d^{(\theta)} = \mathbb{P}(D^{(\theta)} = d)$. Let $\mu_{\theta,\theta'}$ denote the expected number of type θ' neighbors of a type θ vertex. 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 conditions analogous to (a) and (b) above are

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

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

For any pair θ and θ', we must have $p_\theta \mu_{\theta,\theta'} = p_{\theta'} \mu_{\theta',\theta}$. We also impose an irreducibility condition: for any pair θ and θ', there exists a 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.

B   Proofs of Results

B.1   Proofs of Theorems 1 and 2

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

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

where $\bar{d} = \sum_{\theta=1}^{\Theta} d_\theta$ is the total degree associated with d. This measures how much the realized degree sequence deviates from D. It is straightforward to check that $l(\mathbf{d}, D)$ converges

---

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

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*

$$\mathbb{P}\left(\left|N_k(G) - n\mathbb{P}(|\mathcal{T}| = k)\right| \geq \epsilon n\right) \leq e^{-\delta n}. \tag{20}$$

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

$$\mathbb{P}\left(\left|L_1(G) - n\mathbb{P}(|\mathcal{T}| = \infty)\right| \geq \epsilon n\right) \leq e^{-\delta n}, \quad and \quad \mathbb{P}\left(L_2(G) \geq \epsilon n\right) < e^{-\delta n}. \tag{21}$$

I prove Theorem 3 allowing multigraphs—I do not condition on simple graphs. It is easier to work with this version of the configuration model due to full independence in the link stub pairing process. The multi-type analog of Corollary 1 implies the result also holds conditional on realizing a simple graph.

Let G^* to denote a graph realized according to the multigraph configuration model, and let G to denote a generic graph. Let 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 6 and the corresponding tree, viewed as a rooted graph. For a positive integer r, write $G_{v,r}$ for the subgraph of radius r rooted on v (i.e. the subgraph of nodes at distance r or less from v), and similarly write $\mathcal{T}_r$ for the tree $\mathcal{T}$ truncated after the rth generation. I 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 shows that the distribution of component sizes converges to the distribution of tree sizes in 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. Passing to simple graphs 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. $\square$

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 2. *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 3. *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 2 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 14. *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*

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

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 $X_i = \mathbb{E}[f(\pi) \mid \mathcal{F}_i]$ is clearly a martingale. The bijection together with the Lipschitz property implies that

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

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. $\qquad\square$

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. $\qquad\square$

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*

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

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

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 3 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. $\qquad\square$

We can now complete the proof of Proposition 14. Choose Δ sufficiently large so that

$$\left| \mathbb{P}(\mathcal{T} \in \mathcal{P}) - \mathbb{P}(\mathcal{T} \in \mathcal{P} \cap \mathcal{M}_{\Delta,r}) \right| \leq \mathbb{P}(\mathcal{T} \notin \mathcal{M}_{\Delta,r}) \leq \frac{\epsilon}{4}. \tag{22}$$

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

$$\mathbb{P}\left(\left| N_{\mathcal{P}}(G^*) - N_{\mathcal{P} \cap \mathcal{M}_{\Delta,r}}(G^*) \right| \geq \frac{\epsilon n}{2} \right) \leq e^{-\delta_1 n}. \tag{23}$$

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

$$\mathbb{P}\left(\left| N_{\mathcal{P} \cap \mathcal{M}_{\Delta,r}}(G^*) - n\mathbb{P}(\mathcal{T} \in \mathcal{P} \cap \mathcal{M}_{\Delta,r}) \right| \geq \frac{\epsilon n}{4} \right) \leq e^{-\delta_2 n} \tag{24}$$

The inequalities (22), (23), and (24), 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. $\square$

Proposition 14 immediately proves (20) 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 4. *Fix $\epsilon > 0$. For all sufficiently large K, there exists $\delta > 0$ such that if $l(\mathbf{d}, D) < \delta$ we have*

$$\mathbb{P}\left(\left| \sum_{k \geq K} N_k(G^*) - n\mathbb{P}(|\mathcal{T}| = \infty) \right| \geq \epsilon n \right) \leq e^{-\delta n}. \tag{25}$$

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 (20), replacing ϵ with $\frac{\epsilon}{2K}$. $\square$

Corollary 4 implies (21) 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 14 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 14.

Proposition 15. *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. $\qquad\square$

We are now ready to prove (21) 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 4, 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

$$\mathbb{P}\left(L_1 + L_2 \geq n\left(\mathbb{P}(|\mathcal{T}| = \infty) + \frac{\epsilon}{2}\right)\right)$$

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

$$\leq e^{-\delta n}. \tag{26}$$

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

$$\mathbb{P}\left(L_1 \geq n\left(\mathbb{P}(|\mathcal{T}| = \infty) - \frac{\epsilon}{2}\right)\right) \geq 1 - e^{-\delta' n}. \tag{27}$$

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 15. $\qquad\square$

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 4 (or rather, the analog based on Proposition 15), 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

$$\mathbb{P}\left(|S_k| \leq n\left(\mathbb{P}(|\mathcal{T}| = \infty) - \frac{\epsilon}{4}\right)\right) \leq \mathbb{P}\left(|S_k| \leq n\left(\mathbb{P}(|\mathcal{T}'(p)| = \infty) - \frac{\epsilon}{8}\right)\right) \leq e^{-\delta_2 n}. \qquad (28)$$

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 expected 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. $\square$

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

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

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 9 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 $2\tilde{r} = (1 + \epsilon)\log_\nu n$. $\square$

B.2 Practical Implications

In this section I return to the simpler configuration model used throughout most of the paper. We realize the distribution $D^{(\theta)}$ in two steps from a distribution D_θ on $\mathbb{N}$ and a distribution Z_θ on Θ. First, realize the total number of link stubs according to D_θ. For each stub, independently realize a type according to Z_θ. Recall

$$p_\theta = \mathbb{P}(T = \theta), \quad \mu_\theta = \mathbb{E}[D_\theta], \quad q_{\theta'}^{(\theta)} = \mathbb{P}(Z_\theta = \theta'),$$

and write $\mathbf{q}^{(\theta)}$ for the vector of probabilities $\{q_{\theta'}^{(\theta)}\}_{\theta' \in \Theta}$. To ensure $(T, \mathbf{D})$ is consistent in expectation, we need for each pair $\theta, \theta' \in \Theta$ that

$$p_\theta \mu_\theta q_{\theta'}^{(\theta)} = p_{\theta'} \mu_{\theta'} q_\theta^{(\theta')}.$$

Write G for a random graph realized according to this multi-type configuration model, and write $L_1(G)$ for the largest connected component in G.

With this structure, we can rewrite the results in the last subsection using the generating functions of $\{D_\theta\}_{\theta \in \Theta}$. Recall the probability generating function of D_θ is

$$g_\theta(s) = \mathbb{E}\left[s^{D_\theta}\right] = \sum_{k=0}^{\infty} \mathbb{P}(D_\theta = k)s^k.$$

An application of Theorem 1 and standard branching process results gives the following.

Proposition 16. *There exists a unique minimal solution* $\boldsymbol{\xi} \in [0, 1]^{|\Theta|}$ *to the system*

$$g_\theta'(1)s_\theta = g_\theta'(\mathbf{q}^{(\theta)} \cdot \mathbf{s}), \qquad \theta = 1, 2, ..., |\Theta|. \tag{29}$$

xvii

If v_θ is a type θ node chosen uniformly at random, we have

$$\lim_{n\to\infty} \mathbb{P}_{CM_{n,T,\mathbf{D}}} \left(v_\theta \in L_1(G)\right) = 1 - g_\theta \left(\mathbf{q}^{(\theta)} \cdot \boldsymbol{\xi}\right) \equiv \zeta_\theta.$$

If $v_\mathbf{d}$ is a node with degree tuple $\mathbf{d}$ chosen uniformly at random, we have

$$\lim_{n\to\infty} \mathbb{P}_{CM_{n,T,\mathbf{D}}} \left(v_\mathbf{d} \in L_1(G)\right) = 1 - \prod_{\theta=1}^{|\Theta|} \xi_\theta^{d_\theta} \equiv \zeta_\mathbf{d}.$$

Proof. We apply Proposition 8 to compute the survival probability of the characteristic branching process. From the definition of the generating function g_θ for the degree distribution D_θ, the forward distribution D'_θ has the generating function

$$\frac{g'_\theta(s)}{g'_\theta(1)}.$$

A standard property of generating functions is the following. If X is a random variable on $\mathbb{N}$, Z is a random variable on $\mathbb{N}^k$ for some k, and W is the sum of X independent copies of Z, then the generating function g_W of W is $g_X\left(g_Z(\mathbf{s})\right)$. The generating function of Z_θ is $g_{Z_\theta}(\mathbf{s}) = \mathbf{q}^{(\theta)} \cdot \mathbf{s}$. By Proposition 8, the solution $\boldsymbol{\xi}$ to

$$g'_\theta(1)s_\theta = g'_\theta(\mathbf{q}^{(\theta)} \cdot \mathbf{s}), \qquad \theta = 1, 2, ..., |\Theta|$$

gives the vector of extinction probabilities for a multi-type branching process with the off-spring distributions $D'^{(\theta)}$. That is, the element ξ_θ is the probability that a subtree of $\mathcal{T}_{T,\mathbf{D}}$ beginning with a type θ node goes extinct. The claim about nodes with a given degree tuple $\mathbf{d}$ is immediate from this and Theorem 1. The asymptotic probability that a random type θ node is connected to the giant component is then

$$1 - \sum_d \mathbb{P}(D^{(\theta)} = \mathbf{d}) \prod_{\theta=1}^{d_\theta} \xi_\theta^{d_\theta} = 1 - g_{D^{(\theta)}}(\boldsymbol{\xi}) = 1 - g_\theta(\mathbf{q}^{(\theta)} \cdot \boldsymbol{\xi}).$$

$\square$

Proposition 16 restates a key part of Theorem 1 to facilitate computation. While we typically cannot obtain analytic solutions to (29), the system is straightforward to solve using numerical methods. The vector $\boldsymbol{\xi}$ gives the extinction probabilities for a branching process based on the forward distributions: when the characteristic branching process realizes a type θ offspring, that branch of the process dies with probability ξ_θ. We can use this solution to determine how large the giant component is and what nodes are contained in it. An immediate implication is that $\frac{|L_1(G)|}{n}$ converges in probability to a point mass on

$$\zeta = \sum_{\theta=1}^{|\Theta|} p_\theta \zeta_\theta. \tag{30}$$

We can use the last part of the statement to assess the degree distribution within the giant component, which is distinct from the degree distribution of the network as a whole.

We require a few additional definitions to state the result on distances in $L_1(G)$. Define

$$\nu_\theta = \frac{\mathbb{E}\left[D_\theta(D_\theta - 1)\right]}{\mu_\theta} = \frac{\text{Var}\, D_\theta}{\mu_\theta} + \mu_\theta - 1. \tag{31}$$

Moreover, define $m_{ij} = \nu_i q_j^{(i)}$, let M be the $|\Theta|$ by $|\Theta|$ matrix with m_{ij} the entry in row i and column j, and let ν denote the spectral radius of M. Finally, recall that $H(G)$ is a random variable denoting the length of the shortest path between two nodes in $L_1(G)$ chosen uniformly at random.

Corollary 5. *We have $\zeta > 0$ if and only if $\nu > 1$. In this case, we have*

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

Proof. This is immediate from Theorem 2 and our assumptions on the degree distributions. $\square$

Corollary 5 is a specialization of Theorem 2 for the multi-type configuration model with the particular structure of this section. The difference is a reduction in the dimension of M, which makes computing the spectral radius more computationally tractable.

B.3 Other Proofs and Calculations

Proof of Proposition 3

Part (a) is trivial. For part (b), the assumption that adoption propensities are non-decreasing in d means that a FOSD shift in the degree distribution results in more potential adopters. Together with the increase in connectivity, this implies that $D_\sigma^{(\theta)}$ dominates $\hat{D}_\sigma^{(\theta)}$.

For part (c), note that conditional on realizing degree m, a type θ agent has type θ' neighbors (who are potential adopters) distributed according to a binomial distribution with m trials and success probability $q_{\theta'}^{(\theta)} \sigma_{\theta'}^{(\theta)}$. Write $B(n, p)$ for a binomial distribution with n trials and success probability p. Results on the stochastic ordering of binomial distributions (Klenke and Mattner, 2010) imply that $B(n_1, p_1)$ FOSD $B(n_2, p_2)$ if

$$(1 - p_2)^{n_2} \geq (1 - p_1)^{n_1}.$$

We can couple realizations of D_θ to realizations of $\hat{D}_\theta$ such that $\hat{D}_\theta$ realizes a value no higher than $\lceil \hat{D}_\theta^{-1}(D_\theta) \rceil$, and if the above inequality holds for each pair of binomial distributions that arise in this coupling, then $D_\sigma^{(\theta)}$ dominates $\hat{D}_\sigma^{(\theta)}$. $\square$

Calculation for Section 2

Fixing σ, the forward extinction probability for a high-degree player is the minimal solution to

$$\xi_\sigma^{\frac{1}{\overline{d}-1}} = \frac{\underline{d}(1-p)}{\underline{d}(1-p)+\overline{d}p} + \frac{\overline{d}p}{\underline{d}(1-p)+\overline{d}p}(1-\sigma) + \frac{\overline{d}p}{\underline{d}(1-p)+\overline{d}p}\sigma\xi_\sigma$$

$$= 1 - \frac{\overline{d}p}{\underline{d}(1-p)+\overline{d}p}\sigma(1-\xi_\sigma).$$

The virality ν_σ is

$$\nu_\sigma = (\overline{d}-1)\sigma\frac{\overline{d}p}{\underline{d}(1-p)+\overline{d}p}.$$

If $\sigma \leq \frac{\underline{d}(1-p)+\overline{d}p}{\overline{d}p(\overline{d}-1)} \equiv \underline{\sigma}$, then $\nu_\sigma \leq 1$, and there is no belief distortion. A high-degree player's expected payoff from adoption is

$$v\left(1 + (\overline{d}-1)\sigma\frac{\overline{d}p}{\underline{d}(1-p)+\overline{d}p}\right) - c.$$

If $\sigma > \underline{\sigma}$, a player believes each neighbor has degree $\overline{d}$ with probability

$$\frac{\overline{d}p(1-\xi_\sigma^{\overline{d}})}{\underline{d}(1-p)(1-\xi_\sigma^{\underline{d}})+\overline{d}p(1-\xi_\sigma^{\overline{d}})}.$$

The expected payoff from adoption is then

$$v\left(1 + (\overline{d}-1)\sigma\frac{\overline{d}p(1-\xi_\sigma^{\overline{d}})}{\underline{d}(1-p)(1-\xi_\sigma^{\underline{d}})+\overline{d}p(1-\xi_\sigma^{\overline{d}})}\right) - c.$$

Hence, the best reply map for the diffusion game is

$$BR_d(\sigma) = \begin{cases} 1 - c\left(1 + (\overline{d}-1)\sigma\frac{\overline{d}p}{\underline{d}(1-p)+\overline{d}p}\right)^{-1} & \text{if } \sigma \leq \underline{\sigma} \\ 1 - c\left(1 + (\overline{d}-1)\sigma\frac{\overline{d}p(1-\xi_\sigma^{\overline{d}})}{\underline{d}(1-p)(1-\xi_\sigma^{\underline{d}})+\overline{d}p(1-\xi_\sigma^{\overline{d}})}\right)^{-1} & \text{if } \sigma > \underline{\sigma}. \end{cases}$$

C Extensions

C.1 Local Information

In realistic settings, individuals have some information about their close neighbors. We can capture this in our framework if we suppose that players observe the types of their neighbors in addition to how many neighbors they have—we can choose the types in our model to reflect the information that people have about each other. Define the adoption propensities $\sigma(\theta, \mathbf{d})$ analogously to those in section 4, where $\mathbf{d} \in \mathbb{N}^{|\Theta|}$ is a degree tuple—the entry d_θ corresponds to the number of type θ neighbors. To map strategies to diffusion

outcomes, we follow the exact same analysis as in section 4, replacing the diffusivities $\zeta_{\theta,d}^{(\sigma)}$ with diffusivities $\zeta_{\theta,\mathbf{d}}^{(\sigma)}$ that depend on the entire degree tuple. This entails more notation and book keeping, but solving the problem is conceptually the same as before.

In the equilibrium analysis, two complications arise. First, since neighbors are partly distinguishable, it matters whether a player observes who informed her or not. I think the most natural assumption is that a player observes who informs her, but then we need to specify what happens if multiple neighbors do so simultaneously. One possibility is to have the player attribute becoming informed to a single such neighbor, chosen uniformly at random. A second complication is that we must assess players' beliefs about neighbors of neighbors. If I adopt and my neighbors become aware, what is the distribution of their neighbors' types? In a non-viral equilibrium, such beliefs come directly from the distributions that underlie the multi-type configuration model, but in a viral equilibrium, we need to correct for the selection effect. Viral belief distortion no longer affects my beliefs about my neighbors' types—because I observe them—but it does affect my beliefs about the types of my neighbors' neighbors.

C.2 Choosing to Share Information

If a person adopts a behavior, her neighbors may not automatically become aware of it. For instance, one can choose to read a news story and subsequently choose whether to share that story on social media. In the model, we can augment a player's decision problem by adding a choice whether to broadcast information to neighbors. When a player becomes aware, she chooses whether to adopt. Conditional on adopting, she chooses whether to inform her neighbors, and her neighbors only become aware if she informs them. The decision to inform neighbors may affect the player's payoff (e.g. a communication cost, or a referral benefit).

Consider a simple example. There is one type, all players have degree 3, and values are uniform on $[0, 1]$. The payoff from adoption is

$$u(v, a) = va - c$$

for some positive cost c. Moreover, there is a cost $\delta > 0$ to inform neighbors. Given a strategy profile, write σ_a for the probability that a player is willing to adopt, and write σ_i for the probability that a player informs her neighbors. Note that the extent and rate of diffusion depends only on σ_i, while equilibrium decisions will also depend on σ_a.

The forward extinction probability is the minimal solution to

$$\xi = \left(1 - \sigma_i + \sigma_i \xi\right)^2,$$

giving $\xi = 1$ if $\sigma_i \leq \frac{1}{2}$, and $\xi = \left(\frac{1-\sigma_i}{\sigma_i}\right)^2$ otherwise. For a player with value v, the payoff from adopting, but not informing neighbors, is

$$v \left(1 + 2(1 - \xi)\sigma_a\right) - c,$$

where the term $2(1 - \xi)\sigma_a$ appears because a neighbor may get informed by someone else. The payoff from adopting and informing neighbors is

$$v(1 + 2\sigma_a) - c - \delta.$$

As long as costs are not too high, in equilibrium we have

$$1 - \sigma_a = \min\left(\frac{c}{1 + 2(1 - \xi)\sigma_a}, \frac{c + \delta}{1 + 2\sigma_a}\right), \quad 1 - \sigma_i = \frac{\delta}{2\xi\sigma_a}.$$

Notice that for low values of c, there is necessarily a gap between the probability a player adopts σ_a, and the probability she informs her neighbors σ_i. In the extreme case of $c = 0$, we have $\sigma_a = 1$ but $\sigma_i = 1 - \frac{\delta}{2\xi\sigma_a}$. When informing neighbors is costly, widespread diffusion (i.e. low ξ) undermines incentives to share information.

C.3   Diffusion Games with Global Externalities

In some instances, adoption externalities are less direct. When more people use iPhones, there is a larger market for applications, and users benefit from more varied and higher quality apps. When more people hear about a new hit musical, tickets are harder to obtain. These externalities often depend on the total population of adopters, rather than a person's local neighborhood. Building from the last extension, we can fruitfully apply the diffusion model to these settings as well.

As before, there are n players connected in a random network, which we generate according to a multi-type configuration model. At time zero, we choose κ players uniformly at random to adopt a behavior and inform all of their neighbors. In each subsequent period t, players who are aware make an irreversible decision whether to adopt, and conditional on adoption, they choose whether to inform their neighbors.

A player's payoff is a function of whether she adopts, her private value, the total number of players who end up adopting, and whether she informs her neighbors. Each player of type θ has a value $v \in [0, 1]$ drawn independently according to the distribution V_θ. Write $X(t)$ for the number of players who have adopted at time t, and $X = \lim_{t \to \infty} X(t)$ for the number who eventually adopt. A player earns the payoff $u(v, X)$ from adoption, and we normalize the payoff from non-adoption to zero. I assume $u(v, X)$ is differentiable in v with $\frac{\partial u}{\partial v} > 0$ for each X. I also assume that the limit

$$\lim_{X \to \infty} u(v, X)$$

exists and is finite for each v. Additionally, if a player informs her neighbors, she earns the payoff c, which may be positive or negative. The action $(a, i) \in \{0, 1\}^2$, where $a = 1$ indicates adoption and $i = 1$ indicates informing neighbors, yields the total payoff

$$U(v, X, a, i) = a(u(v, X) + ic).$$

Each player observes her type θ, her value v, and how many neighbors d she has in the network. Players do not observe the period t. We can express a symmetric strategy

profile as a function $\sigma : \Theta \times [0,1] \times \mathbb{N} \to \{0,1\}^2$, giving an adoption and sharing decision for each possible type, value, and degree. I use $\sigma_a(\theta, v, d)$ to denote the adoption decision associated with σ and $\sigma_i(\theta, v, d)$ to denote the sharing decision—to economize on notation, going forward I omit the dependence on θ, v, and d when there is little risk of confusion. The payoff from following the strategy σ is

$$U(v, X, \sigma_a, \sigma_i).$$

The value κ, the configuration model, the value distributions $\{V_\theta\}_{\theta \in \Theta}$, and the payoffs are common knowledge.

We can define limit equilibria of the diffusion game with global externalities in a way analogous to the earlier model. Viral belief distortion manifests in a particularly simple way for this model: under a viral strategy profile, limit beliefs assign probability one to the event $X = \infty$, regardless of a player's action. This leads to an extreme version of the free-rider effect when externalities are positive and an extreme version of the tragedy of the commons when externalities are negative.

C.3.1 Positive Externalities

Platform services, like social networking sites or online marketplaces, exhibit positive network externalities, as do many software applications and entertainment products. In this section, I assume u is strictly increasing in X, and I focus on the case where $c < 0$. The latter assumption means it is costly to inform neighbors about the product. We can interpret c as an attention cost or possibly a social cost (e.g. being ostracized for sending friends too much spam). When informing neighbors is costly, players can only justify choosing $i = 1$ if the corresponding increase in X justifies the cost; otherwise, each adopter trivially informs her neighbors in equilibrium.

Viral belief distortion creates a discontinuity in the set of limit equilibria at $c = 0$. Conditional on a viral equilibrium, a player believes infinitely many others will adopt regardless of what she does, so her choice has no effect on the value of externalities. If $c < 0$, this means she never informs her neighbors, and therefore a viral equilibrium cannot exist.

Proposition 17. *Suppose $u(v, X)$ is increasing in X for each v. If $c < 0$, there is no viral limit equilibrium. If $c \geq 0$, every adopting player informs her neighbors in any equilibrium of the n player game.*

Proof. This is immediate from the argument in the previous paragraph. $\square$

The impossibility of viral equilibria with positive costs represents an extreme version of the free-rider effect. Widespread adoption depends on players sharing information with their neighbors. When doing so is costly, players only wish to share if they expect to have a meaningful impact on the extent of diffusion, but in a large cascade this is never true. Even when externalities are substantial, a small cost of informing neighbors results in extreme inefficiency. This highlights the important role of viral product features—features that automate sharing, or greatly reduce its cost—and referral payments in spreading new

products with positive externalities. The severity of the free-rider problem means that modest interventions can have an enormous impact on aggregate adoption.

C.3.2   Negative Externalities

In some instances, the spread of information creates negative externalities. If information is rival—like valuable knowledge about a company's stock, which becomes less valuable if more people trade on it—or if a service is subject to congestion problems, then diffusion reduces the value. Suppose $u(v, X)$ is strictly decreasing in X. Mirroring the section the analysis of positive externalities, I assume $c > 0$—without some benefit to sharing, players never inform neighbors in equilibrium. Such a benefit may arise if players have altruistic preferences towards their neighbors, if they expect reciprocal favors, or if there is some explicit incentive to share information.

Unlike games with positive externalities, where viral belief distortion eliminated incentives to share information, here viral belief distortion eliminates the cost of sharing. With $c > 0$, viral equilibria are self-reinforcing because players perceive no meaningful effect of sharing information.

Proposition 18. *Suppose $u(v, X)$ is decreasing in X for each v, and $c > 0$. In any viral equilibrium, all adopting players inform their neighbors. If the network is sufficiently dense and $u(v, \infty)$ is positive for some v, then a viral equilibrium exists.*

Proof. If the network is dense enough, viral spread is possible when all players who can earn a positive payoff adopt and share information. If this is the case, we can sustain a viral equilibrium because players perceive no cost of sharing. $\square$

In network games with negative externalities, viral belief distortion creates an extreme version of the tragedy of the commons. Players fail to internalize any cost of sharing information, and as a result, we can sustain viral equilibria in any sufficiently dense network. With local incentives to share, it is hard to keep a secret.

C.4   Clustering

Real networks exhibit far more clustering than the model $CM_{n,T,\mathbf{D}}$—two individuals with friends in common are more likely to be friends themselves. This last subsection discusses how to introduce clustering in the configuration model while retaining the ability to analyze structure through a branching process approximation. We do this by adding k-cliques to the network, groups of k nodes that are all linked to one another, for arbitrary k. For ease of exposition, we assume a single type of node throughout this subsection.

We characterize a configuration model with k-cliques via a degree distribution D on tuples $(d_2, d_3, ..., d_{\overline{k}})$, where $\overline{k} \geq 2$ is an arbitrary positive integer. Starting with n nodes, we take independent random draws from D to generate degree tuples for each. The associated link stubs are labeled with the types 2,3,...,$\overline{k}$. We successively select link stubs of type j uniformly at random in groups of size j, and we establish links between all associated nodes.

For instance, if j is three, then we select 3 type 3 link stubs, and we create a 3-clique with the nodes attached to those stubs. We repeat this for each $j = 2, 3, ..., \overline{k}$ until all link stubs are used. In the basic model, we only have type 2 stubs. Adding type $k > 2$ creates k-cliques in the graph. Node i has total degree $d_i = \sum_{j=2}^{\overline{k}} (k-1) d_j^{(i)}$.

As before, we require a consistency condition. If $\mathbf{d} = \{d^{(i)}\}_{i=1}^n$ is the n vector of degree tuples, then it is **consistent** if for all $j = 2, 3, ..., \overline{k}$ we have that

$$\frac{1}{j} \sum_{i=1}^n d_j^{(i)}$$

is an integer. Given n nodes and the distribution D, we condition the network formation procedure on realizing a consistent $\mathbf{d}$ and on realizing a simple graph.

As we did in section 6.2, we can define a characteristic branching process for this version of the configuration model. We begin with a root node that draws offspring according to D. A single type j link stub corresponds to $j-1$ offspring, which we label as type j. For each subsequent node of type j, we realize offspring independently according to the forward distribution D_j', defined by

$$\mathbb{P}\left(D_j' = d\right) = \frac{(d_j + 1)\mathbb{P}(D = d + e_j)}{\mathbb{E}[D_j]},$$

where e_j is a tuple of zeros with a 1 in the entry corresponding to type j. Again, a type j link stub in the forward distribution corresponds to $j-1$ offspring. The proofs of Theorems 1 and 2 apply for configuration model with k-cliques with only minor modifications.