Diffusion Games*

Evan Sadler†

September 30, 2017

Abstract

Behaviors and information often spread via person-to-person diffusion. I argue that standard simplifying assumptions in diffusion models can lead to misguided predictions. A more detailed model has more realistic qualitative features, and we can reverse standard comparative statics results. We also uncover a new strategic effect: when large cascades are possible, exposure to the contagion conveys information about a player’s network position. This information allows players to better coordinate their actions. 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 often share similar features. People are exposed to something new through contact with others. They make a choice whether to adopt some behavior or pass on information. Preferences are heterogeneous and can depend on neighbors’ choices. Finally, the ultimate outcome of the process is highly uncertain, occasionally producing “viral” cascades.

As one example, the spread of fake news played a prominent role in the 2016 United States elections. False stories targeted to supporters of particular candidates spread almost exclusively on social media, in some cases reaching more readers than the most widely consumed mainstream news 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 referral programs (Lobel et al., 2015), and the spread of microfinance participation in developing countries (Banerjee et al., 2013). Understanding these processes is therefore important for countering misinformation, designing marketing campaigns, and encouraging economic development.

---

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

†Harvard University – esadler@fas.harvard.edu
The standard approach for studying these processes affords us a tractable framework, but it imperfectly captures the 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. It also eliminates any variance in the eventual outcome. Similarly, assumption (b) precludes viral phenomena. The choices of 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 a simple example comparing a mean-field model with an analogous “diffusion game.” In the diffusion game, players interact in a fixed random network, so the extent of diffusion is sensitive to the initial seeding—some seeds trigger large cascades, while others induce little adoption. In the mean-field model, the reshuffling of links means that 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, while 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—shifts in the degree distribution can result in faster spread but a smaller number of eventual adopters. 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 positive local externalities, this information helps players coordinate their adoption decisions. Despite players remaining unexposed in the diffusion game, the dynamic coordination 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 games in a large class of random graph models. A pair of limit theorems characterizes the component structure of a multi-type configuration model as the number of players grows. The technical insight is to approximate the local network structure with a branching process, allowing us to study network properties using 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 single-type model. To the best of my knowledge, the present paper is the first to supply a formal argument that applies to a more general model of 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. This paper contributes tools for the effort. Branching process approximation reduces a complex random network to a much simpler stochastic process. Applying these tools to study diffusion, our findings contrast sharply with standard models. Varied outcomes highlight that it matters who the initial seeds are. We can define notions of centrality that weigh on how far adoption spreads, bringing theory closer to empirical results. Moreover, beliefs and strategies change when we account for diffusion dynamics, revealing new connections between network structure and equilibrium behavior.

After discussing related work, I begin with a simple example 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

Models using mean-field assumptions dominate the theoretical literature on diffusion. Canonical early examples assume all prior adopters influence current adoption 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 some aspects of a network structure. In a degree-based mean-field approximation, we distinguish individuals according to how many neighbors they have, and we separately track 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 all of these models, we obtain distinct but related logistic curves that describe adoption levels for each distinguishable subpopulation. The degree distribution determines the relationships between these curves, and there are detailed results on how steady state adoption varies with this distribution.

The present paper addresses three limitations of mean-field diffusion models. First, steady state adoption in mean-field models is a deterministic function of the network’s degree distribution. Long-run outcomes are therefore independent of who the early adopters are, which limits our ability to capture the empirically documented role of innovators and opinion leaders. In contrast, when we model persistent relationships in a discrete network, early adopters play a crucial role in the diffusion process. A closely related limitation is a reliance on adoption being reversible—without this assumption, steady states become insensitive to
the network structure. Adoption often is reversible in the long-run, but switching costs—for instance, the need to invest capital or learn how to use a new technology—can make it effectively irreversible in the short-run. Sometimes adoption constitutes a one-time consumption (e.g. watching a video or reading a news story), or learning a piece of information, in which case we need to model irreversible choices, as the present paper does. Lastly, by rendering individual decisions insignificant for global outcomes, the mean-field assumptions trivialize strategic effects. This paper finds that even in large networks, the decision of a single individual 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 take actions 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 model with reversible actions. In one view, this correspondence justifies studying static network games—the static game is a meaningful reduced form for a more realistic dynamic process. However, this also means that mean-field dynamics add relatively little to the understanding we have from static games. 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 diffusion dynamics 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. Morris (2000) stands out as an early example, studying a fixed network in which a player adopts a behavior if enough neighbors did so previously. The analysis reveals how cohesive subgroups act as barriers to diffusion and conditions under which clustering facilitates diffusion. Watts (2002) and Campbell (2013) are more closely related to this paper. Both use the basic configuration model to generate a network and study diffusion using older percolation results. 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 a monopolist’s pricing problem when consumers learn about a product through word-of-mouth. Purchasing consumers create an externality by informing others, which results in more elastic demand and lower prices than in a standard model. This work partly addresses the first two limitations of mean-field models, providing insight on which individuals and network features influence adoption when choices are irreversible. However, these models struggle to address heterogeneity, information, and strategic effects.

I make several innovations on this literature. First, the agents in this paper make strategic choices. While earlier models employ a mechanistic rule or myopic best response, I show how we can tractably incorporate beliefs and strategies. Second, by adapting newer mathematical techniques, the results in this paper permit the study of random networks with heterogeneous

---

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.
2 See also Chapter 7.2 of Jackson (2008).
agents. This is an important step for economic applications because individuals often have different preferences and different information, and in real networks, neighbors’ attributes are typically correlated. Moreover, the explicit link to branching processes provides a unifying conceptual framework to study diffusion in random networks. This approach has several advantages over alternatives. 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. Our findings are portable across different settings and allow a more complete analysis of existing models.

The role of social networks in technology adoption is widely studied in the economic 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 explore 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.

\(^3\)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).
2 Diffusion versus Steady State Adoption

This section illustrates how diffusion games differ from mean-field models. I frame the benchmark as a simultaneous-move adoption game, with the understanding that its equilibria correspond to the steady states of a mean-field diffusion process (See Jackson and Yariv, 2007). In both games, a large population of players is linked in a network that we characterize via the degree distribution $D$. In both games, players make binary choices whether to adopt some behavior. In both games, players earn the same utility from adoption

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

where $a$ is the number of neighbors who adopt, the adoption cost is $c$, and players have mutually independent values $v$ drawn uniformly from $[0, 1]$—the default payoff is zero.

2.1 Two Models

I begin with the benchmark game. All players in a continuum simultaneously decide whether to adopt some behavior. A player observes her own value $v$ and number of neighbors $d$. We consider symmetric equilibria, so a strategy profile is a mapping $\sigma : [0, 1] \times \mathbb{N} \to \{0, 1\}$ giving an adoption decision for each value-degree pair. Players know that each neighbor has a degree drawn independently from the distribution $\tilde{D}$ with

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

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

$$q_\sigma = E[\sigma(V, \tilde{D})].$$

(1)

The strategy profile $\sigma$ is an equilibrium if

$$\sigma(v, d) (vdq_\sigma - c) \geq 0$$

for all $(v, d)$. That is, players adopt if doing so yields a positive expected payoff. A fraction $p_\sigma = E[\sigma(V, D)]$ of the population ends up adopting.

To understand how this seemingly static game represents diffusion, consider a particular form of myopic best response dynamics. Suppose $\sigma$ describes adoption behavior at the end of a period. In the next period, a player decides whether to adopt assuming each neighbor does so with independent probability $q_\sigma$. This encodes two assumptions. The first is myopia: players assume that others will follow the same strategy next period. The second is the mean-field assumption: neighbors next period are a random draw from the population. The steady
states of this process are exactly the equilibria of the simultaneous-move adoption game. There are typically two equilibria with positive adoption levels. The higher equilibrium is stable, and the lower one is a “tipping point.” If initial adoption levels are below the tipping point, we converge to the equilibrium with no adoption. To reach the stable outcome with positive adoption levels, we must seed a significant fraction of the population.

I now describe the diffusion game. There are \( n \) players linked in a random network, which we draw according to a configuration model with degree distribution \( D \). That is, each player realizes a number of “link stubs” independently according to \( D \), and the link stubs are paired uniformly at random. We condition on realizing a simple graph, one with no self-links and at most one link between any pair of players.\(^5\) In the limit for large \( n \), a random player has degree drawn according to \( D \), and a random neighbor has degree drawn according to \( \tilde{D} \) as defined above.

Initially, none of the players have been exposed to the behavior and are unaware of it. At time zero, a single player chosen uniformly at random adopts the behavior, and her neighbors become aware. In each subsequent period, players who are aware make an irreversible decision whether to adopt, and if a player adopts, her neighbors are aware in the next period. The structure of the game is common knowledge, and each player observes her own value \( v \) and degree \( d \)—importantly, players do not observe the number of periods that have passed. We again consider symmetric strategy profiles \( \sigma : [0,1] \times \mathbb{N} \rightarrow \{0,1\} \). Write \( \hat{q}_\sigma \) for the probability, conditional on exposure, that a neighbor is willing to adopt.\(^6\) The profile \( \sigma \) is an equilibrium if for all \( (v,d) \), we have

\[
\sigma(v,d) (v (1 + \hat{q}_\sigma(d - 1)) - c) \geq 0,
\]

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

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 use of a discrete fixed graph, instead of reshuffling links each period, is an important difference. The need to spread awareness creates an additional barrier to widespread adoption—for a player to adopt, the player must both get exposed and have \( \sigma(v,d) = 1 \). For many strategy profiles, diffusion can only ever reach a negligible fraction of the population. However, we do not need to cross a “tipping point” to ensure self-sustaining adoption. In some cases, a single seed triggers a cascade covering a positive fraction of the network.

I give a brief summary analysis based on results proved later in the paper—all statements pertain to the limit as \( n \rightarrow \infty \). Given a strategy profile \( \sigma \), 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 \( D_\sigma \) by

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

\(^5\)Later sections describe the configuration model in more detail.

\(^6\)In the large network limit, neighbor adoption decisions are independent.
This adjusts for selection effects, giving the degree distribution for a random potential adopter. Let \( g \) and \( g_\sigma \) denote the probability generating functions of \( D \) and \( D_\sigma \) respectively. Recall \( q_\sigma \) defined in (1), the unconditional probability that a neighbor is a potential adopter, and let \( \xi_\sigma \) denote the minimal solution to

\[
 g'_\sigma(1)\xi = g'_\sigma(q_\sigma\xi + 1 - q_\sigma).
\]  

(2)

A giant component exists in the subnetwork of potential adopters if and only if \( \xi_\sigma < 1 \). The value \( \xi_\sigma \) 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 \( p_\sigma \left( 1 - g_\sigma(q_\sigma\xi_\sigma + 1 - q_\sigma) \right) \) 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. Even then, however, we fail to reach the full population \( p_\sigma \) of potential adopters—a fraction \( g_\sigma(q_\sigma\xi_\sigma + 1 - q_\sigma) \) of them are not connected to the giant component.

2.2 A New Effect: Viral Belief Distortion

In the diffusion game, forward looking players can learn about their network position. This manifests itself as a difference between \( q_\sigma \), 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 for any given player, the likelihood of being exposed in a viral cascade dwarfs the likelihood of being exposed in a small cascade. As a result, conditional on being exposed, a player is convinced that

(a) A large cascade is in progress, and
(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 the degree distribution in the network overall because high-degree players are more likely to be in it. An exposed player learns that her neighbors have degrees drawn according to \( \tilde{D}^{(\sigma)} \), with

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

(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

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

\( ^7 \)A giant component is one whose size grows linearly in \( n \) as \( n \to \infty \).
With our assumed payoff structure, higher degree players are more likely to adopt, so when \( \xi_\sigma < 1 \), have \( \hat{q}_\sigma > q_\sigma \): learning facilitates coordination, allowing higher equilibrium strategies.

The impact of the diffusion process on coordination shows up clearly in the best reply map. Figure 1 depicts \( q_{BR(\sigma)} \) as a function of \( q_\sigma \), where \( BR(\sigma) \) is a best reply profile. In the simultaneous move 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. First, in the diffusion game an exposed player knows one of her neighbors has already adopted. Second, when \( \sigma \) crosses a critical threshold, we have the new effect of viral belief distortion. Looking at (3), we see that this selection effect is most significant when \( \xi_\sigma \) is close to 1. Consequently, as \( \sigma \) increases, the best reply map exhibits a discontinuous jump upwards.

Figure 1: Best reply maps for the simultaneous adoption game and the diffusion game.
2.3 Reversed Comparative Statics

Consider the degree distribution $D_1$ taking the value 3 with probability 1, and the degree distribution $D_2$ taking the values 1 and 5 with equal probability. The distribution $D_2$ is a mean preserving spread of $D_1$. The results of Jackson and Yariv (2007) imply that adoption is always higher under $D_2$ than under $D_1$—if the cost $c$ is positive, equilibrium adoption is strictly higher. For $c$ in a neighborhood of zero, the opposite holds for the diffusion game: equilibrium adoption is higher under $D_1$.

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 probability close to one. If we move to the distribution $D_2$, then essentially all players are still potential adopters, but $\xi_\sigma$ 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 a fraction

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

adopts.

Having a fixed graph is the key difference behind this finding. If we do not reshuffle links, then low-degree players cause the network to fragment. When all players have degree 3, essentially everyone is connected. 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.

2.4 Higher Equilibrium Adoption

The need to spread awareness creates a barrier to adoption in the diffusion game. Nevertheless, we can sometimes sustain higher equilibrium adoption levels than in a comparable mean-field model. 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 will never adopt. We therefore abuse notation by letting $\sigma$ denote the probability that a degree 5 player adopts under strategy profile $\sigma$. We can compute $q_\sigma = \frac{5\sigma}{6}$, and the distribution $D_\sigma$ takes the value 5 with probability one, so $g_\sigma(s) = s^5$. From (2), know that $\xi_\sigma$ solves

$$\xi = \left(\frac{5\sigma}{6} \xi + 1 - \frac{5\sigma}{6}\right)^4,$$
and one can verify that $\xi_\sigma < 1$ if $\sigma > 1/5$. From (3), we have
\[
\hat{q}_\sigma = \sigma \mathbb{P}(\tilde{D}^{(\sigma)} = 5) = \sigma \frac{5(1 - (q_\sigma \xi_\sigma + 1 - q_\sigma)^5)}{5(1 - (q_\sigma \xi_\sigma + 1 - q_\sigma)^5) + q_\sigma (1 - \xi_\sigma)} > q_\sigma.
\]
The profile $\sigma$ corresponds to an equilibrium if the cutoff value satisfies
\[
v = 1 - \sigma = \frac{c}{1 + 4\hat{q}_\sigma},
\]
which has a solution $\sigma^* > 1/5$ as long as
\[
c < \sup_{\sigma \in [1/5, 1]} (1 - \sigma)(1 + 4\hat{q}_\sigma) \equiv \overline{c}.
\]
To estimate $\overline{c}$, note that
\[
\overline{c} > \sup_{\sigma \in [1/5, 1]} (1 - \sigma)(1 + 4\hat{q}_\sigma) = \sup_{\sigma \in [1/5, 1]} (1 - \sigma) \left(1 + \frac{10\sigma}{3}\right) = \left(1 - \frac{7}{20}\right) \left(1 + \frac{10}{3} \cdot \frac{7}{20}\right) = \frac{169}{120} > \frac{25}{24}.
\]
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 adopted. Viral belief distortion further enhances coordination, allowing us to push $c$ even higher and still obtain viral cascades in equilibrium.

### 2.5 Summary

Relative to a mean-field model, the diffusion game better captures the salient features of person-to-person diffusion. Viral phenomena occur naturally in this framework, and the location of the initial seed matters for the aggregate outcome. Changes to the network structure have two distinct affects on the extent of diffusion: equilibrium strategies can change, and the set of players who get exposed can change. As a result, we sometimes reverse comparative statics from mean-field models. When large cascades are possible, we uncover a new effect: players who become exposed learn about their own network positions. When there are positive externalities, this reinforces adoption incentives, helping players better coordinate their actions. Consequently, we sometimes get more adoption in a diffusion game than in a comparable mean-field model.

The remainder of the paper develops a general framework to study diffusion games with heterogeneous players. I extend the analysis of this section to multi-type random networks in which different types of players can have different propensities to link with one another. The techniques I develop allow us to study diffusion games in networks with realistic features like homophily and, as I show in the appendix, clustering.
3 Diffusion Games with Local Externalities

Analyzing diffusion in discrete graphs presents three major challenges. First, unlike in a mean-field model, we cannot keep track of diffusion using a small number of state variables. It matters exactly who has adopted and where they are in the fixed network. Second, it’s not just the network structure that matters. Not everyone will adopt if exposed, and the subnetwork among those who are willing to adopt is what ultimately determines the outcome. We need to assess the structure of this subnetwork. Finally, we need to account for the inferences of forward-looking players. The information players have about the network impacts their behavior, and players can learn about the network during the game.

We can get traction on each of these challenges because of the configuration model’s special structure. Due to the uniform random matching of link stubs, the local structure of the configuration model converges in distribution to that of a multi-type branching process. Properties of the network—the size of the giant component, typical distances within that component—are 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 these limit theorems as given, most of the analysis becomes an exercise in keeping track of selection effects. The subnetwork among players who are willing to adopt if exposed—the potential adopters—also follows a configuration model, but with different asymptotic degree distribution. We need to account for the fact that a random potential adopter is systematically different from a random player, and not every neighbor is a potential adopter. Likewise, assessing players’ beliefs in a viral cascade means accounting for selection into the giant component.

The rest of this section presents the general diffusion game in a multi-type network with random seeding. Section 4 takes a strategy profile as given 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 feed into the characterization. This analysis deals mostly with selection into the network of potential adopters. Section 4.2 provides basic comparative statics results, holding the strategy profile fixed. In section 5, I address equilibrium beliefs, which means accounting for selection into the giant component. Finally, section 6 presents the limit theorems on which all of the analysis rests.

3.1 The Model

A population consists of $n$ players connected in a network $G$, which we represent as an undirected graph. 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 decisions whether to adopt this 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 $\Theta$ 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. First, assign each player a type drawn independently
according to $T$. Second, for each type $\theta$ player, independently draw a degree $d$ according to $D_\theta$ and attach $d$ “link stubs” to the player. Next, assign a type to each link stub, drawn independently according to $Z_\theta$. Finally, pair compatible link stubs uniformly at random to form links—that is, for each type $\theta'$ link stub attached to a type $\theta$ player, match it with a random type $\theta$ link stub attached to a type $\theta'$ player.

We condition this process on realizing matching numbers of compatible link stubs and realizing a simple graph. Limit results for large $n$ require a consistency condition on the distributions $T$ and $\{D_\theta, Z_\theta\}_{\theta \in \Theta}$. Define $p_\theta = \mathbb{P}(T = \theta)$, $\mu_\theta = \mathbb{E}[D_\theta]$, 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 that the expected number of type $\theta'$ stubs attached to type $\theta$ players equals the expected number of type $\theta$ stubs attached to type $\theta'$ players.

Each type $\theta$ player draws a private value $v \in [0, 1]$ independently from the distribution $V_\theta$. 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 equilibria in symmetric strategies. A symmetric strategy profile is a function $\sigma(\theta, v, d) : \Theta \times [0, 1] \times \mathbb{N} \rightarrow \{0, 1\}$ giving an adoption decision for each type, value, and degree. Fixing a profile $\sigma$, 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 $\theta$ player with degree $d$ who becomes informed. The symmetric strategy profile $\sigma$ is an equilibrium for the $n$ player game if for all $(\theta, v, d)$ and each $x \in \{0, 1\}$, we have

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

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

Equilibrium analysis rests on understanding the conditional distributions $\mathbb{P}_{\sigma}^{(\theta, d, n)}$. Our tools concern limits of the outcome distribution, so we shall study equilibria under limiting beliefs as $n \rightarrow \infty$. Theorem 1 in section 6 implies that the limiting distribution $\mathbb{P}_{\sigma}^{(\theta, d)}$ over neighbors’ actions is well defined. We say that a strategy profile is a limit equilibrium if for all $(\theta, v, d)$ and each $x \in \{0, 1\}$, we have

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

A simple graph has no self-links and at most one link between any pair of nodes.
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. Results characterize the distributions $\alpha_n$ and $\tau_n$ for large $n$.

### 3.2 Remarks on the Model

The multi-type configuration model can represent many realistic network features—we can model arbitrary degree distributions, heterogeneity, and homophily. One feature this model fails to capture is clustering. In real networks, those linked to each other are likely to share many 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 player types. An econometrician may view them as distinct collections of observable attributes (e.g. demographic and socioeconomic variables). We could then imagine estimating the configuration model’s parameters from actual network data. We could also use types to encode information that players possess. In the model as described, players observe how many neighbors they have, but they observe no other information about their neighbors. Another extension (See C.1) shows how to incorporate signals about neighbors by expanding the set of types.

I assume choices are irreversible, and individuals who refrain from adopting cannot revisit this decision. One’s use of a technology or product clearly is reversible in the long run, but switching costs may render this prohibitively expensive in the short run. The spread of new technologies is often a slow process (Griliches, 1957; Skinner and Staiger, 2005), so the “short run” is economically important. Alternatively, the diffusion of information necessitates a model with irreversible adoption. A person cannot unread a news story or unwatch a video. In the most natural examples, both the decision to consume information and the decision to share information are important. Another extension allows players to choose whether to inform their neighbors.

Since players do not observe whether neighbors have adopted, formally there is no reason to revisit a decision. However, if players were to observe their neighbors’ choices, they might wait until several neighbors do so before adopting themselves. In the model as described, this would change little because of the network’s tree-like local structure—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, and I analyze an example in the appendix, but the qualitative features of the diffusion game do not change.

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 $\sigma$, the adoption propensities

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

capture everything relevant to the outcome distribution. We can imagine three layers in the network. The outermost layer is the social network itself. The strategy profile determines a subnetwork of potential adopters—the players who would choose to adopt upon being exposed. Each player with type $\theta$ and degree $d$ is a potential adopter with independent probability $\sigma(\theta, d)$. The innermost layer is the network of actual adopters. The players who actually adopt are in a connected component of potential adopters that contains a seed. To characterize the extent of diffusion $\alpha_n$, we need to assess the size of components in the network of potential adopters. To characterize diffusion times $\tau_n(x)$, we need to assess typical distances within these connected components.

The subnetwork of potential adopters is itself a configuration model. This is clear once we condition on who the potential adopters are and which link stubs connect to them: after conditioning, each possible pairing of link stubs between potential adopters is equally likely. Our task in this section therefore has two steps. First, we determine the type and degree distributions in the network of potential adopters. Second, we apply limit theorems for the multi-type configuration model to understand the component structure of this network.

The techniques underlying our limit theorems provide helpful intuition. Figure 2 depicts the same 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 of offspring, 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 and exhaust more nodes in the graph. However, in the limit for large $n$, the local structure converges in distribution to a standard multi-type 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 $\theta$ according to $T$, a number of offspring according to $D_\theta$, and offspring types according to $Z_\theta$. Subsequent nodes realize offspring according to the forward distributions $D'_{\theta}$, defined by

$$\mathbb{P}(D'_{\theta} = d) = \frac{(d + 1)\mathbb{P}(D_{\theta} = d + 1)}{\mu_\theta}. \quad (4)$$
The forward distributions incorporate two corrections. First, reweighting the distribution according to players’ degrees accounts for the disproportionate representation of high-degree players as neighbors. The second correction is to subtract one—this is necessary to avoid double counting the link back to the parent.

Understanding diffusion in a large random graph is equivalent to understanding the characteristic branching process. The survival probability of the branching process corresponds to the size of the largest component in the network. From this we can characterize the extent of diffusion $\alpha_n$. The growth rate of the branching process tells us about distances in the configuration model. From this we obtain the diffusion times $\tau_n(x)$. These two parameters are exactly the network’s **diffusivity** and **virality** defined later in this section.

### 4.1 Diffusivity and Virality

To obtain the type and degree distributions in the network of potential adopters, we must account for selection into this network via the strategy profile $\sigma$. In the large network limit, a random player is a type $\theta$ potential adopter with probability

$$\pi^{(\sigma)}_\theta = p_\theta \mathbb{E}[\sigma(\theta, D_\theta)].$$

(5)

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

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

(6)

Each type $\theta$ link stub leads to a potential adopter with independent probability

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

where $D'_\theta$ is the forward distribution defined in (4).
We generate the degree distribution for a type $\theta$ player in the network of potential adopters in three steps. First, realize a total number of link stubs according to $D_{\sigma,\theta}$. Next, assign types independently for each stub, giving type $\theta'$ with probability $q_{\sigma,\theta'}^{(\theta)} = q_{\sigma,\theta'}^{(\theta)}\sigma_{\theta'}$. Finally, delete stubs left without a type, which occurs with probability

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

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

To define diffusivity, we need the probability generating functions for the degree distributions $\{D_{\sigma,\theta}\}_{\theta \in \Theta}$ and $\{D_{\theta}\}_{\theta \in \Theta}$. The probability generating function of $D_{\sigma,\theta}$ is

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

Likewise, the probability generating function of $D_{\theta}$ is

$$g_{\theta}(s) = \sum_{d=1}^{\infty} \mathbb{P}(D_{\theta} = d)s^d.$$ 

A simple exercise shows that the generating function for the forward distribution $D'_{\sigma,\theta}$ is

$$g'_{\sigma,\theta}(s) = g_{\sigma,\theta}(q_{\sigma}^{(\theta)} \cdot s), \quad \theta = 1, 2, \ldots, |\Theta|, \quad s_0 = 1.$$ 

has a unique minimal solution $\xi^{(\sigma)} \in [0, 1]^{|\Theta|+1}$.

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

$$\zeta^{(\sigma)}_{\theta,d} = 1 - (q_{\sigma}^{(\theta)} \cdot \xi^{(\sigma)})^d.$$ 

The **diffusivity of a player with type $\theta$** is

$$\zeta^{(\sigma)}_{\theta} = 1 - g_{\theta}(q_{\sigma}^{(\theta)} \cdot \xi^{(\sigma)}) \quad \theta = 1, 2, \ldots, |\Theta|.$$ 

The **diffusivity of a potential adopter with type $\theta$** is

$$\phi^{(\sigma)}_{\theta} = 1 - g_{\sigma,0}(q_{\sigma}^{(\theta)} \cdot \xi^{(\sigma)}) \quad \theta = 1, 2, \ldots, |\Theta|.$$ 

\(^9\)The appendix provides statements of the relevant results along with references.
The diffusivity of the network is
\[
\zeta_\sigma = p \cdot \zeta^{(\sigma)} = \sum_{\theta=1}^{\Theta} p_{\theta} \left( 1 - g_{\theta}(q_{\theta}^{(\theta)} \cdot \xi^{(\sigma)}) \right).
\]

The diffusivity of the potential adopter network is
\[
\phi_\sigma = \pi \cdot \phi^{(\sigma)} = \sum_{\theta=1}^{\Theta} \pi_{\theta} \left( 1 - g_{\theta,\sigma}(q_{\theta}^{(\theta)} \cdot \xi^{(\sigma)}) \right).
\]

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

The limit results in section 6 imply that the diffusivities also correspond to the asymptotic probability that particular players are connect to the giant component of potential adopters. Our first proposition below uses the diffusivities \( \zeta_\sigma \) and \( \phi_\sigma \) to characterize the extent of diffusion \( \alpha_n \). The diffusivities of players \( \zeta^{(\sigma)}_{\theta,d}, \zeta^{(\sigma)}_{\theta}, \) and \( \phi^{(\sigma)}_{\theta} \) provide additional information about who adopts in a large cascade. This proves important in section 5 when we assess players’ beliefs. Player diffusivities also function as centrality measures for this model, giving probabilities that certain players can trigger a large cascade. Unlike standard centrality measures, diffusivities depend on both the network structure and the adoption propensities.

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

**Proof.** This is immediate from Proposition 16 in the appendix. \( \square \)

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 very few individuals ever get exposed. Second, more seeds does not mean more adoption: the random variable \( \alpha \) 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—stochastic outcomes and extent that is independent of \( \kappa \)—are intuitive if we recall what diffusivity represents. The diffusivity \( \zeta_\sigma \) is the probability that a random player has a link to the largest component of potential adopters. The value \( \phi_\sigma \) is the size of this component. Each seed connects to it with independent probability \( \zeta_\sigma \), 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, yielding insignificant adoption. We can also tell how many people become exposed in a large cascade: a fraction \( \zeta_\sigma \) of players link to an adopter while a fraction \( \phi_\sigma \) actually adopt.
To define virality, we first define the mean offspring matrix for the characteristic branching process. For each type $\theta$ potential adopter, the average forward degree is

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

We then write $M_\sigma$ for the $|\Theta| \times |\Theta|$ matrix with entries

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

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* $\nu_\sigma$ of the network is the spectral radius of $M_\sigma$.

**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

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

**Proof.** This is immediate from Theorem 2. \qed

When we get a large cascade, the time it takes to reach any positive fraction $x$ of eventual adopters is approximately $\log_{\nu_\sigma} n$. Put differently, most adoption takes place in a narrow time window around $\log_{\nu_\sigma} n$. The virality corresponds to the growth rate of the characteristic branching process: the number of offspring in each generation grows roughly like $\nu_\sigma t$, so the number of generations it takes to obtain $n$ total progeny is approximately $\log_{\nu_\sigma} n$. In the early stages of diffusion, the number of adopters grows as in the branching process. Once a significant fraction of the population adopts, the process 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 do not necessarily move in tandem with changes to the network. Virality depends only on the first two moments of the degree distribution, but diffusivity depends in subtle ways on the entire distribution. As we change the degree distribution, virality moves in the direction a mean-field model would predict, but diffusivity can move in the opposite direction, reversing the corresponding comparative static for the extent of adoption.

### 4.2 Comparative Statics

We now explore how the extent and rate of diffusion vary with key parameters—the degree distributions $\{D_\theta\}$, the neighbor type distributions $\{Z_\theta\}$, and the adoption propensities $\sigma(\theta, d)$. What ultimately matters is how these jointly influence the degree distribution in the network of potential adopters. When this network gets denser, measured through first order stochastic dominance, adoption spreads further and faster. When it gets more centralized, measured through second order stochastic dominance, adoption spreads faster, but
it may not spread as far. 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—the hat notation will indicate variables that correspond to the second set. Throughout this section, we fix the collection of types $\Theta$ and the type distribution $T$.

Our first result concerns first order stochastic dominant shifts in the degree distribution. When the network of potential adopters is denser, we clearly get increased diffusivity and virality. However, the degree distribution in the network of potential adopters depends in non-trivial ways on all three of our fundamental parameters. Proposition 3 helps parse the effect of different changes. I abuse notation slightly, letting $D$ denote both a degree distribution and a cumulative distribution function. I write $D \text{ FOSD } \hat{D}$ to mean $D(m) \leq \hat{D}(m)$ for all $m \in \mathbb{N}$.

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

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

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

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

$$
(1 - q_{\sigma,\theta}^{(\theta)}[D_{\theta}^{-1}(D_\theta(m))]) \geq (1 - q_{\sigma,\theta'}^{(\theta)})^m
$$

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

**Proof.** See Appendix.

Condition (a) is intuitive: if we add agents to the network of potential adopters, then diffusion spreads further and faster. Condition (b) looks at changes in the total degree distributions, holding the type distributions $\{Z_\theta\}$ and adoption propensities fixed. As long as adoption propensities are weakly increasing in degree, a denser network leads to higher $\zeta_\sigma$ and $\nu_\sigma$. If adoption propensities decrease in degree, then the lower propensity to adopt could offset the additional connectivity. Condition (c) characterizes the tradeoff between adding links and increasing adoption propensities. If $D_\theta \text{ FOSD } \hat{D}_\theta$, then the exponent $[\hat{D}_\theta^{-1}(D_\theta(m))]$ 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 $\nu_\sigma \geq \hat{\nu}_\sigma$:

(a) We have $m_{ij}^{(\sigma)} \geq \hat{m}_{ij}^{(\sigma)}$ 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 $\theta$, and $D_\theta$ is a mean preserving spread of $\hat{D}_\theta$ for each type $\theta$
Proof. Part (a) is trivial. For part (b), convexity implies that a mean-preserving spread results in (weakly) higher adoption propensities $\sigma^{(\theta)}$. From the definition, a mean preserving spread also increases $\nu_\theta$. The result follows from (14) together with part (a).

We can interpret higher variance as higher centralization in the network: more links connect to a smaller number of central individuals. This change reduces path lengths, allowing adoption to spread more quickly. For related reasons, in mean-field diffusion models, mean preserving spreads typically result in higher steady state adoption. However, in our setting this can mean more isolated components. Recalling the example in Section 2.3, Proposition 4 implies that we get faster diffusion after a mean preserving spread, but we already saw that the extent of diffusion declined. While adoption spreads faster in a more centralized network, it may not spread to as many people.

5 Limit Beliefs and Viral Belief Distortion

We can distinguish two types of strategy profiles—non-viral and viral profiles. In a non-viral profile, the parameter $\nu_\sigma$ is less than or equal to one, and from Proposition 2, this means adoption can never spread beyond a tiny fraction of the population. A viral profile has $\nu_\sigma > 1$, meaning that large cascades covering a positive fraction of the network are possible. The distinction has important implications for players’ beliefs.

In a non-viral profile, a player who gets exposed knows that exactly one neighbor has adopted, and the remaining neighbors have types and degrees drawn independently from the forward distributions. In a viral profile, determining players’ beliefs is a more involved exercise. 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 becoming aware, a player is certain there is a large cascade. Therefore, in a viral strategy profile, each newly aware player conditions her beliefs about her neighbors on the event that they are connected to the giant component—I call this adjustment viral belief distortion.

With viral belief distortion, players put more probability on neighbors having high degrees and higher adoption propensities. To derive limit beliefs, we have to condition the characteristic branching process on survival. Recall the forward extinction probabilities $\{\xi^{(\sigma)}_\theta\}$ and the adoption-weighted neighbor type probabilities $\{q_{\sigma,\theta'}^{(\theta)}\}$ defined in the previous section. A type $\theta$ player with degree $d$ is connected to the giant component with probability $\zeta^{(\sigma)}_{\theta,d} = 1 - (q^{(\theta)}_{\sigma'} \cdot \xi^{(\sigma)})^d$, and hence a type $\theta$ player is connected to this component with probability $\zeta^{(\sigma)}_{\theta} = 1 - g^{(\theta)}_{\theta'} (q^{(\theta)}_{\sigma'} \cdot \xi^{(\sigma)})$. To obtain the distribution of types and degrees for players connected to the giant component, we need to reweight $Z_\theta$ and $D_\theta$ to account for the selection effect.

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

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

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

Some implications of Proposition 5 may appear counterintuitive. Adoption fails to go viral with positive probability, and in this event, every player who becomes aware has systematically incorrect beliefs about her neighbors. This is possible because the number of individuals who become aware in a non-viral realization is insignificant compared with that in a viral realization. Hence, the result suggests a reason why users of a niche product might be unrealistically optimistic about the product becoming popular—we could summarize their reasoning as “if I found out about it, it must be catching on.”

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

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

captures the relative likelihood that two players of the same type with degrees $d$ and $d'$ are connected to the giant component. When diffusivity is small, the value $q^{(\theta)}_\sigma \cdot \xi^{(\sigma)}_\sigma$ is close to one. If $d > d'$, the ratio grows with $q^{(\theta)}_\sigma \cdot \xi^{(\sigma)}_\sigma$. 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, and this jump in beliefs can make it surprisingly easy to sustain a viral equilibrium.

### 6 Random Graphs that Resemble Reality

This section presents the paper’s technical contribution: a characterization of component sizes and typical distances in a class of random graph models that can represent realistic network features. 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.\(^{10}\) Degree distributions typically follow

\(^{10}\)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.
a power-law, meaning that a few nodes with many links account for a significant fraction of all links. The small-worlds property says roughly that distances between nodes are small relative to the network’s size. Clustering means that neighboring nodes often have neighbors in common. Homophily is the tendency to link preferentially to those with similar attributes. The configuration model and its variants can generate all of these features.

While the random graph is a complicated object—with overlapping neighborhoods, cycles, and multiple disjoint components—a limit theorem ties its structure to a characteristic branching process. The branching process is a far simpler object, amenable to analysis via standard tools for studying sums of independent random variables. I first describe the multi-type configuration model. This presentation is slightly more general than that given earlier in the paper as I do not assume independence of neighbor types. I subsequently define the characteristic branching process and state the main theorems. 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 some useful corollaries and extensions, are given in an appendix.

6.1 The Multi-type Configuration Model

We start from a finite collection of types $\Theta = \{1, 2, \ldots, |\Theta|\}$, a distribution over types $T$, and a collection of degree distributions $D = \{D(\theta)\}_{\theta \in \Theta}$, one for each type. With multiple types of nodes, a degree is not simply a number but a $|\Theta|$-tuple $d = (d_1, d_2, \ldots, d_{|\Theta|})$: a node’s degree records the number of links to nodes of each possible type. The only assumption I impose on the distributions in $D$ is that each component has finite expectation and finite variance. We assign a type to each of $n$ nodes, drawn independently from the distribution $T$. Each node of type $\theta$ then draws a degree tuple independently according to $D(\theta)$. Think of this step as assigning link stubs to each vertex, where each stub is labeled with the type of node to which it can connect. After realizing the link stubs, we pair compatible stubs together uniformly at random to form links: for each type $\theta$ stub connected to a type $\theta'$ node, we draw, from those not yet paired, a type $\theta'$ stub connected to a type $\theta$ node.

For this procedure to make sense, we need matching numbers of compatible stubs. Given an $n$-vector of types $\theta = (\theta(1), \theta(2), \ldots, \theta(n))$ and an $n$-vector of degree tuples $d = (d^{(1)}, d^{(2)}, \ldots, d^{(n)})$, we say the pair $(\theta, d)$ is consistent if

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

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

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

\[11\] Finite variance is only necessary for Theorem 2.
θ' stubs that $D^{(θ)}$ realizes, and define $p_θ = \mathbb{P}(T = θ)$. We say the pair $(T, D)$ is **consistent in expectation** if for each pair of types $θ$ and $θ'$, we have $p_θ \mu_θ^{(θ)} = p_θ \mu_θ^{(θ')}$. This ensures that the distributions generate the same expected number of stubs on each side of a link.

The pairing of stubs may yield multiple links between the same two nodes, or even self links. A graph without such features is called a **simple graph**. We further condition the process on realizing a simple graph. The following definition summarizes the procedure to generate the multi-type configuration model.

**Definition 3 (Multi-type Configuration Model).** Given $n$ nodes and a consistent pair $(θ, d)$ of types and degree tuples, we realize the **multi-type configuration model with type sequence $θ$ and degree sequence $d$**, written $CM_{n,θ,d}$, as follows. Assign node $i$ type $θ(i)$ 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 $(T, D)$ that are consistent in expectation, we realize the **multi-type configuration model with distribution $(T, D)$**, written $CM_{n,T,D}$, as follows. Generate a pair $(θ, d)$ according to $(T, D)$. Conditional on realizing a consistent pair, generate a graph according to $CM_{n,θ,d}$.

I conclude this subsection with a pair of examples to illustrate the model.

**Example: The Basic Configuration Model**

Suppose $Θ$ contains just a single type. There is then only one type of link, and the collection $\{D^{(θ)}\}_{θ ∈ Θ}$ reduces to a single degree distribution $D$ assigning the total number of links to each node. By specifying different distributions $D$, even the basic model can capture a large range of different network structures. This version of the configuration model is widely applied, and existing results provide many tools to analyze 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

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

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 $ζ = 1 - g(ξ)$, where $ξ$ is the minimal solution to

$$g'(1)ξ = g'(ξ).$$
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 $\Theta$ consists of two types, type $a$ and type $b$, and $T$ assigns equal probability to each type. Let $h$ denote a number in $[0, 1]$. To obtain the distribution $D^{(a)}$, we first realize the total number of links according to $D$. We then label each link type $a$ with independent probability $\frac{1+h}{2}$ and type $b$ otherwise. We obtain $D^{(b)}$ similarly, realizing the total number of links according to $D$ and then labeling each type $b$ with independently probability $\frac{1+h}{2}$ and type $a$ otherwise. The parameter $h$ varies the extent of homophily in the network. When $h = 0$, the network is structurally equivalent to a basic 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. The theorems in the next subsection allow us to derive analogous results on the component structure and typical distances for the two type model with homophily.

### 6.2 Branching Process Approximation

Given a type distribution $T$ and a collection of degree distributions $D = \{ D^{(\theta)} \}_{\theta \in \Theta}$, we define the characteristic branching process $T_{T,D}$. The process $T_{T,D}$ is realized in two stages. We begin with a root node, which draws a type $\theta \in \Theta$ according to $T$ and realizes offspring according to $D^{(\theta)}$. Each subsequent node realizes offspring independently according to a distribution that depends on both the node’s type and its parent’s type. A type $\theta$ node with a type $\theta'$ parent realizes the tuple of offspring $d = (d_1, d_2, \ldots, d_{\Theta})$ with probability

$$
P(D^{(\theta\theta')} = d) = \frac{(d_{\theta'} + 1)P(D^{(\theta)} = d + e_{\theta'})}{\mu^{(\theta)}_{\theta'}},$$  

where $e_{\theta}$ is a vector containing a one in the entry corresponding to type $\theta$ and zeros elsewhere. The distribution $D^{(\theta\theta')}$ differs from $D^{(\theta)}$ for two reasons. First, we need to account for the “friendship paradox:” high-degree nodes are overrepresented as neighbors. Second, we need to subtract the link back to the parent. Intuitively, we obtain $D^{(\theta\theta')}$ from $D^{(\theta)}$ by reweighting realizations proportionally to the number of type $\theta'$ links and then subtracting one of these links. I refer to the distributions $D^{(\theta\theta')}$ as the forward distributions.

**Definition 4 (Characteristic Branching Process).** Given the distributions $(T, D)$, we realize the characteristic branching process $T_{T,D}$ as follows. A single root node realizes a type $\theta$ according to $T$ and offspring according to $D^{(\theta)}$. Every subsequent type $\theta$ node with a type $\theta'$ parent realizes offspring independently according to the forward distribution $D^{(\theta\theta')}$. Returning to our two examples—the basic configuration model and the two type model with homophily—we can construct the characteristic branching processes. Recall that the
basic configuration model has a single type and a single degree distribution \( D \). The forward distribution in this case is \( D' \) defined by
\[
P(D' = d) = \frac{(d + 1)P(D = d + 1)}{E[D]}.
\]
The forward distribution \( D' \) corresponds to a standard degree-weighting of the distribution and then subtracting one. This is the degree distribution for a random neighbor rather than a random node. The characteristic branching process begins with a root realizing offspring according to \( D \), and all subsequent nodes realize offspring independently according to \( D' \).

In the example with two types, nodes realize total degrees according to \( D \), and each link is an own-type link with independent probability \( \frac{1 + h}{2} \). Due to the independent labeling of link stubs, the forward distributions are independent of the parent node's type: we have \( D'(aa) = D'(ab) \equiv D'(a) \) and \( D'(ba) = D'(bb) \equiv D'(b) \). Taking \( D' \) as in (17), the forward distributions \( D'(a) \) and \( D'(b) \) realize total degree according to \( D' \), and each link is an own-type link with independent probability \( \frac{1 + h}{2} \) and an other-type link otherwise. The characteristic branching process begins with a root node that is either type \( a \) or \( b \) with equal probability, this node realizes offspring according to \( D \), and each offspring has the same type as the root 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 configuration model 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, and as we exhaust more nodes in the graph, the distribution of new offspring gets sparser. 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 \( i \)th largest component. Write \( P_{CM_n,T,D} \) for the probability measure over graphs defined by the configuration model \( CM_n,T,D \). Define \( \rho_k = P(|T_{T,D}| = k) \), and note that \( \rho_\infty \) is the survival probability of \( T_{T,D} \).

**Theorem 1** (Law of Large Networks). Suppose \( T_{T,D} \) is irreducible and non-singular.\(^{13}\) For any \( \epsilon > 0 \), we have
\[
\lim_{n \to \infty} P_{CM_n,T,D} \left( \left| \frac{N_k(G)}{n} - \rho_k \right| \geq \epsilon \right) = 0,
\]
\[
\lim_{n \to \infty} P_{CM_n,T,D} \left( \left| \frac{|L_1(G)|}{n} - \rho_\infty \right| \geq \epsilon \right) = 0, \text{ and}
\]
\[
\lim_{n \to \infty} P_{CM_n,T,D} \left( \left| \frac{|L_2(G)|}{n} \right| \geq \epsilon \right) = 0.
\]

\(^{12}\)See for instance Jackson and Yariv (2007).

\(^{13}\)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.
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 \( \rho_k \): the distribution of component sizes converges to the law of total progeny for \( \mathcal{T}_{T,D} \). 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. Tying network structure to a characteristic branching process significantly simplifies random graph analysis because standard tools for sums of independent random variables permit a detailed understanding of \( \mathcal{T}_{T,D} \). Theorem 1 extends findings for the basic 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 this component. To do this, we define the mean offspring matrix \( M_{T,D} \) of \( \mathcal{T}_{T,D} \). Recall that \( D_{\theta}(\theta') \) denotes the offspring distribution for a type \( \theta \) node with a type \( \theta' \) parent. Write

\[
\nu_{\bar{\theta}}^{(\theta\theta')} = \mathbb{E}[D_{\bar{\theta}}^{(\theta\theta')}] 
\]

for the expected number of type \( \bar{\theta} \) offspring of a type \( \theta \) node with a type \( \theta' \) parent. The mean offspring matrix \( M_{T,D} \) is a \(|\Theta|^2 \times |\Theta|^2 \) matrix in which each row and column coordinate corresponds to a particular pair of types \((\theta, \theta')\). The entry for row \((\theta_r, \theta'_r)\) and column \((\theta_c, \theta'_c)\) is equal to \( \nu_{\bar{\theta}}^{(\theta_r, \theta'_r)} \) if \( \theta'_c = \theta_r \) and 0 otherwise.\(^{14}\) The entries of the mean offspring matrix record the average number of offspring for each type of node in the characteristic branching process \( \mathcal{T}_{T,D} \): the entry in column \((\theta_c, \theta'_c)\) and row \((\theta_r, \theta'_r)\) is the average number of type \( \theta_c \) nodes with type \( \theta'_c \) parents that are realized as offspring of a type \( \theta_r \) node with a type \( \theta'_r \) parent.

The spectral radius \( \nu_{T,D} \) of \( M_{T,D} \) is the key parameter in our characterization of typical distances. Given a graph \( G \), let \( H(G) \) be a random variable denoting the length of the shortest path between two vertices chosen uniformly at random from \( L_1(G) \). Recall that \( \rho_\infty \) is the survival probability of the characteristic branching process \( \mathcal{T}_{T,D} \).

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

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

Proof. See Appendix.

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

\(^{14}\)The zeros enter because a type \( \theta_r \) can only have offspring with type \( \theta_r \) parents.
log_{\nu T,D} n.\) This illustrates how the configuration model exhibits the small-world property: typical distances grow logarithmically in network size. This is intuitive if we think about how a branching process grows over time. Standard results imply that the the number of offspring in successive generations grows exponentially with base \(\nu T,D\). Conversely, this means the number of generations it takes to get \(n\) total progeny is on the order of \(\log_{\nu T,D} n\).

The giant component in the associated configuration model inherits this same property: essentially all nodes are within a distance \(\log_{\nu T,D} n\) of one another.

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 \(D_\theta\) and then realizing types independently according to \(Z_\theta\). 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 theoretical work based on mean-field assumptions faces several limitations. The tools in this paper allow us to study diffusion with a more faithful representation of the underlying network. By modeling the network as a discrete random graph, we can capture viral phenomena, short-run dynamics, the importance of central individuals, and the strategic effects that emerge when individuals are not myopic.

Viral belief distortion is a central phenomenon that appears in diffusion games. When large cascades are possible, the fact that an individual gets exposed provides information about her network position. This information changes equilibrium strategies and outcomes. In a game with positive local externalities, this creates a dynamic coordination effect. Viral belief distortion makes players more optimistic about others adopting, leading to higher equilibrium strategies. In some cases, this effect allows us to sustain high adoption levels when no one would adopt otherwise.

A richer view of network structure yields more subtle adoption patterns. As we change the network structure, the extent and speed of diffusion can move in different directions, and we can reverse some standard comparative statics results. Unlike a mean-field model—in which we implicitly assume that links are continually reshuffled—the persistence of links means we can define meaningful notions of centrality for individual players. This brings theory closer

\(^{15}\)van der Hofstad et al. (2005) prove an analogous result for typical distances in the basic 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 basic configuration model, showing that typical distances are much smaller, on the order of \(\log \log n\).
to a recent body of empirical research that shows network centrality can predict adoption.

With new tools come new avenues for future work. One obvious direction to pursue is a systematic study of optimal seeding in various settings. The distinction between non-viral and viral equilibria can provide guidance on whether a small number of seeds will do or if we need a more intensive campaign. In both cases, we can ask who to target for early adoption based on observable attributes. Another direction to explore is to what extent we can decentralize the use of this information. Referral programs are one way to incentivize people to spread information about an innovation, and a well-designed program may be able to leverage the knowledge people have about their networks. The law of large networks may prove useful beyond studies of diffusion. Other applications in network economics may benefit from an approach based on random graphs that can capture a network’s discrete structure.

References


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 \( n \)th generation. To obtain the next generation, each member of the \( n \)th 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 \( \eta \) 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 $Z_n$ denote an $r$-tuple $(Z_{n,1}, Z_{n,2}, \ldots, Z_{n,r})$, giving the number of individuals of each type in the $n$th generation. Given an initial population $Z_0$, recursively define the sequence $Z_n$ as

$$Z_n = \sum_{i=1}^{r} \sum_{j=1}^{r} X_{j}^{(i)} Z_{n-1,i},$$

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

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

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

The multi-type generating function is the vector $G_X(s) = (G^{(1)}(s), G^{(2)}(s), \ldots, G^{(r)}(s))$. Let $\eta^{(i)}$ denote the probability of extinction if $Z_0$ consists of a single individual of type $i$, and let $\eta = (\eta^{(1)}, \eta^{(2)}, \ldots, \eta^{(r)})$ 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 $\rho$ 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 if there exists a type that does not have a single offspring with probability one.

Proposition 8. The vector $\eta$ is the only solution in the unit cube of $G_X(\eta) = \eta$.

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

Proof. See chapter 5 in Athreya and Ney (1972).
The spectral radius $\rho$ of $M$ characterizes the growth rate of the process.

**Proposition 9.** There exists a non-negative random vector $W$ such that $Z_n^{\rho n}$ converges to $W$ 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 $CM_{n,d}$ for the basic configuration model with degree sequence $d$, and write $CM_{n,D}$ for the model where $d$ 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 $CM_{n,d(n)}$, the degree vectors $d^{(n)}$ must converge in an appropriate sense. For an $n$-vector $d^{(n)}$ of degrees, let $n_d(d^{(n)})$ denote the number of entries equal to $d$, and let $m(d^{(n)}) = \sum_{i=1}^{n} d_i^{(n)}$ 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

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

(b) We have

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

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

---

$^{16}$At first glance, the second condition might appear redundant, but it is necessary to rule out pathological cases. For instance, suppose $d^{(n)}$ 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 $n-1$. 

iii
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 $CM_{n,d}$ 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 $CM_{n,d}$ selects a graph uniformly at random from those with $n$ vertices and degree sequence $d$.

*Proof.* This is immediate from the definition as each possible pairing of link stubs that results in a simple graph is equally likely. \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 $\mu$ as $n$ grows. Independent link formation makes it relatively easy to show for any $\epsilon > 0$ that

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

but we need to show that

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

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

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

*Proof.* This is equivalent to Lemma 21 of Bollobás and Riordan (2015). \square

**Corollary 1.** Suppose that for any $\epsilon > 0$, there exists $\delta$ such that for all sufficiently large $n$ we have

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

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

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

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

$$
\mathbb{P}_{CM_{n,d(n)}} (|f(G) - \mu| \geq \epsilon \bigm| G \text{ is simple}) \leq \frac{\mathbb{P}_{CM_{n,d(n)}} (|f(G) - \mu| \geq \epsilon)}{\mathbb{P}_{CM_{n,d(n)}} (G \text{ is simple})} < \frac{e^{-\delta n}}{e^{-\frac{\delta}{2} n}} = e^{-\frac{\delta}{2} n}.
$$

\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

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

**Proof.** The corresponding inequality without conditioning on simple \( G \) is immediate from the Dvoretzky-Kiefer-Wolfowitz inequality. The result then follows from Corollary 1. \( \square \)

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 \( CM_{n,q,D} \) 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 \( CM_{n,q,D} \). The subgraph has a realized degree sequence \( d(\vert S \vert) \). Conditional on selecting \( \vert S \vert \) nodes and realizing the sequence \( d(\vert S \vert) \), the distribution of \( CM_{n,q,D} \) is exactly that of \( CM_{\vert S \vert, d(\vert S \vert)} \): the subgraph is itself a configuration model.

As \( n \) grows, the degree distribution of the subgraph approaches a natural limit. Define the distribution \( D_q \) 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 \( q \). After drawing \( k \), draw the degree \( d \) according to a binomial distribution with \( k \) trials and success probability \( \frac{\sum_{i=0}^{\infty} i q_i p_i}{\sum_{i=0}^{\infty} i p_i} \), the probability that a random link stub is attached to a node selected according to \( q \). 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 \( p_{q,d} \) denote \( \mathbb{P}(D_q = d) \).

**Proposition 13.** For any \( \epsilon > 0 \), there exists \( \delta > 0 \) such that

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

**Proof.** From the construction of \( D_q \), we can compute the probabilities \( p_{q,d} \) as

\[
p_{q,d} = \sum_{k \geq d} \frac{q_k p_k}{\sum_{i=0}^{\infty} q_i p_i} \left( \begin{array}{c} k \\ d \end{array} \right) \left( \frac{\sum_{i=0}^{\infty} i q_i p_i}{\sum_{i=0}^{\infty} i p_i} \right)^d \left( \frac{\sum_{i=0}^{\infty} (1 - q_i) p_i}{\sum_{i=0}^{\infty} i p_i} \right)^{k-d}.
\]

I first show that \( \mathbb{E}_{CM_{n,q,D}} \left[ \frac{n_d(G)}{\vert S \vert} \right] \) converges to \( p_{q,d} \). Let \( n_{d,q}(G) \) denote the number of degree \( d \) nodes in \( CM_{n,D} \) that get selected according to \( q \), let \( m_q(G) \) denote the number of link...
stubs attached to nodes selected according to \( q \), 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

\[
P_{CM_{n,D}} \left( \sup_d \frac{n_d q(G)}{|S|} - \frac{g_d p_d}{\sum_{i=0}^\infty q_i p_i} \geq \epsilon \right) < \frac{1}{3} e^{-\delta n}.
\]

An application of Hoeffding’s inequality ensures we can choose this \( \delta \) so that also

\[
P_{CM_{n,D}} \left( \left| \frac{m(G)}{n} - \sum_{i=0}^\infty i q_i p_i \right| \geq \epsilon \right) < \frac{1}{3} e^{-\delta n}, \quad \text{and} \quad P_{CM_{n,D}} \left( \left| \frac{m(G)}{n} - \sum_{i=0}^\infty i p_i \right| \geq \epsilon \right) < \frac{1}{3} e^{-\delta n}.
\]

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 \( CM_{n,q,D} \) with degree \( d \) is the probability that a random vertex in \( S \) has degree \( d \). Write \( D_n \) for the degree of a vertex in \( CM_{n,q,D} \) 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

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

In event \( A \), with \( n \) sufficiently large, we have

\[
\left| \mathbb{P}(D_v^+ = k) - \frac{q_k p_k}{\sum_{i=0}^\infty q_i p_i} \right| < \epsilon, \quad \text{and} \quad \left| \prod_{i=0}^{k-1} \frac{m_q(G) - i}{n}^{\pi_i} \frac{m(G) - m_q(G) - i}{n}^{1-\pi_i} - \left( \sum_{i=0}^\infty i q_i p_i \right)^d \left( \sum_{i=0}^\infty i q_i p_i \right)^{k-d} \right| < (1 + \epsilon)^k - 1, \quad \text{and} \quad \left| \prod_{i=0}^{k-1} \frac{m(G) - i}{n}^{k} - \left( \sum_{i=0}^\infty i p_i \right)^k \right| < (1 + \epsilon)^k - 1.
\]

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

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 \( i \)th 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 $q$ is asymptotically equivalent to the model $CM_{n,D,q}$, so any structural results for the configuration model apply to these subgraphs—we simply replace the degree distribution $D$ with $D_q$.

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 $(\theta, d)$. 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 $\theta'$ neighbors of a type $\theta$ vertex. Given a sequence $(\theta, d)^{(n)}$, let $n_{\theta,d} ((\theta, d)^{(n)})$ denote the number of type $\theta$ entries with corresponding degree tuple $d$, and let $m_{\theta, \theta'} ((\theta, d)^{(n)})$ denote the number of edges between type $\theta$ and $\theta'$ nodes. The conditions analogous to (a) and (b) above are

(a) \[ \lim_{n \to \infty} \frac{n_{\theta,d} ((\theta, d)^{(n)})}{n} = p_{\theta} p_d^{(\theta)}. \]

(b) \[ \lim_{n \to \infty} \frac{m_{\theta, \theta'} ((\theta, d)^{(n)})}{n} = p_{\theta} \mu_{\theta, \theta'} < \infty. \]

For any pair $\theta$ and $\theta'$, we must have $p_{\theta} \mu_{\theta, \theta'} = p_{\theta'} \mu_{\theta', \theta}$. We also impose an irreducibility condition: for any pair $\theta$ and $\theta'$, 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 $d = \{(d_1, d_2, ..., d_\Theta)_{i, \theta}\}_{i=1}^{n}$ to denote the sequence of types and degree tuples. Fixing $d$, write $m_{\theta, \theta'} (d)$ for the number of edges connecting type $\theta$ and type $\theta'$ nodes, and write $n_{\theta,d}(d)$ for the number of type $\theta$ nodes with degree tuple $d$. Finally, define the configuration distance as

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

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 $l(d, D)$ converges

\[ 17 \text{Combining this Proposition with the main theorem effectively generalizes percolation results to cases with arbitrary heterogeneity in how permeable different links are.} \]
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 \( l(d, D) < \delta \) we have

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

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 \( l(d, D) < \delta \) we have

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

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 \( G^* \) denote a graph realized according to the multigraph configuration model, and let \( G \) denote a generic graph. Let \( v \) denote a generic node in \( G \), and \( G_v \) to denote the graph \( G \) rooted on \( v \). I use \( T \) 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 \( T_r \) for the tree \( T \) truncated after the \( r \)th generation. I abuse notation, writing \( G_{v,r} = T_r \) to indicate that \( T_r \), 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 \( T \). 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 \( T \) 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 \( T_r \). 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 \( T \). This is the sense in which \( T \) 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 \( \{d_n\}_{n \in \mathbb{N}} \) is a sequence for which \( l(d_n, D) \) converges to zero. For any finite \( r \), we can couple the random graphs \( G^*_{v,r} \) and \( T_r \) so that they are isomorphic with probability approaching 1 as \( n \) approaches infinity.

**Proof.** Begin with a realized sequence \( d \), and suppose that \( l(d, D) < \epsilon \) 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 \( T_r \) 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 \( T_r \) with probability at least \( 1 - \epsilon \).

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

\[
\frac{d_{\theta}(n_{\theta,d}(d) - u_{\theta,d,j})}{m_{\theta,d}(d) - u_{\theta,j}},
\]

where \( u_{\theta,d,j} \) is the number of type \( \theta' \) nodes with degree vector \( d \) that have been visited before the \( j \)th step, and \( u_{\theta,j} \) is the number of completed edges between type \( \theta \) and type \( \theta' \) nodes before the \( j \)th step. Note that \( u_{\theta,d,j} \leq j \) and \( u_{\theta,j} \leq j \). This implies that for any fixed \( j \), the difference between this quantity and \( \Pr(D_{\theta} = d) \) 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 \( |T_r| \leq M \) with probability at least \( 1 - \epsilon \), and for sufficiently large \( n \), we have \( l(d, D) < \epsilon \) with probability at least \( 1 - \epsilon \). For \( N \) larger than this, the probability that the coupling fails is no more than \( 2\epsilon + M(\epsilon + o(1)) \), and the conclusion follows.

One immediate consequence of this lemma is that the rooted graphs \( G_{v,r}^* \) are trees with probability approaching 1. More generally, the branching process \( T \) 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 \( \{d_n\}_{n \in \mathbb{N}} \) is a sequence for which \( l(d_n, D) \) converges to zero. Then,

\[
\lim_{n \to \infty} \Pr(G_v^* \in \mathcal{P}) = \Pr(T \in \mathcal{P}).
\]
Equivalently, for any \( \epsilon > 0 \), there exists \( N_\epsilon \) such that if \( n \geq N_\epsilon \) we have

\[
|E[N_P(G^*)] - n\mathbb{P}(T \in \mathcal{P})| \leq \epsilon n.
\]

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 \( l(d, D) < \delta \), then conditional on the degree sequence \( d \) we have

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

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

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

Corollary 2 of course implies convergence of \( \frac{N_k(G^*)}{n} \) to \( \mathbb{P}(|T| = k) \), 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 \( l(d, D) < \delta \) then

\[
\mathbb{P}\left(\left|\frac{N_{\mathcal{P}}(G^*)}{n} - \mathbb{P}(T \in \mathcal{P})\right| \geq \epsilon n\right) \leq e^{-\frac{\delta^2 n}{2}}.
\]

The first step to obtain this bound is a lemma using the Azuma-Hoeffding inequality. Fixing a degree sequence \( d \), we can consider different pairings of stubs. We say that two pairings \( \pi_1 \) and \( \pi_2 \) are related by a switching if we can obtain \( \pi_2 \) from \( \pi_1 \) 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 \( d \). We say that \( f \) is \( C \)-Lipschitz if for any \( \pi_1 \) and \( \pi_2 \) 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 \( d \), let \( M \) denote the total number of pairs. If \( \pi \) is chosen uniformly at random from all pairings of \( d \), then for any \( r \geq 0 \) we have

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

**Proof.** Let \( S_\theta' = \{s_1, s_2, ..., s_m\} \) denote the set of stubs leading from type \( \theta \) nodes to type \( \theta' \) nodes, with \( S'_\theta = \{s'_1, s'_{2}, ..., s'_m\} \) the set of potential partners. We consider a random process in which we sequentially reveal the pairing. Conditional on the partners of \( s_{i+1} \), let \( \Omega \) denote the set of pairings between \( S'_\theta \) and \( S'_\theta \) that are consistent with the information revealed so far. For any possible partner \( b \) of \( s_{i+1} \), let \( \Omega_b \) denote the subset of \( \Omega \) 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 $\Omega_b$ and $\Omega_c$ in which each $\pi_1 \in \Omega_b$ is related by a switching to its image $\pi_2 \in \Omega_c$: just switch the pairs $\{s_{i+1}, b\}$ and $\{s_j, c\}$ to $\{s_{i+1}, c\}$ and $\{s_j, b\}$.

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

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

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

This lemma is sufficient to prove our concentration result for a local property $\mathcal{P}$ if $N_{\mathcal{P}}(G)$ is $C$-Lipschitz for some $C$, but this is not universally true for all local properties. However, if we modify the property to avoid high-degree vertices, we can obtain a $C$-Lipschitz function of the graph and use it to prove the concentration bounds. For $\Delta \geq 2$ and $r \geq 0$, let $\mathcal{M}_{\Delta, r}$ be the property of rooted graphs that every node within distance $r$ of the root has degree at most $\Delta$. 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 $\mathcal{M}_{\Delta, r}$ 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 $\Delta$. For each endpoint of $e$, there can be at most $(1 + \Delta + \ldots + \Delta^r) \leq 2\Delta^r$ such paths, so adding or removing an edge can change the number of vertices with property $\mathcal{Q}$ by at most $4\Delta^r$. Since a switching corresponds to removing two edges and adding two edges, the result follows.

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

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

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

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

**Proof.** The first part is immediate since the total number of offspring in $\mathcal{T}$ within distance $r$ of the root is finite with probability one. For $\delta$ sufficiently small, Corollary 3 then implies that $\mathbb{E}\left[N_{\mathcal{M}_{\Delta, r}}(G^*)\right] \geq n \left(1 - \frac{\epsilon}{3}\right)$. Apply Lemma 3 to the trivial $r$-local property (i.e. the property that always holds), which shows that $N_{\mathcal{M}_{\Delta, r}}(G^*)$ is a $C$-Lipschitz function for some $C$. The second part now follows from Lemma 2.
We can now complete the proof of Proposition 14. Choose $\Delta$ sufficiently large so that
\[ |\mathbb{P}(T \in \mathcal{P}) - \mathbb{P}(T \in \mathcal{P} \cap \mathcal{M}_{\Delta,r})| \leq \mathbb{P}(T \notin \mathcal{M}_{\Delta,r}) \leq \frac{\epsilon}{4}. \] (22)
Let $B = n - N_{\mathcal{M}_{\Delta,r}}(G^*)$ denote the number of high-degree vertices in the graph $G^*$. Since $|N_{\mathcal{P}}(G^*) - N_{\mathcal{P} \cap \mathcal{M}_{\Delta,r}}(G^*)| \leq B$, Lemma 4 implies that for some $\delta_1 > 0$, whenever $l(d, D) < \delta_1$ we have
\[ \mathbb{P}\left(\left| N_{\mathcal{P}}(G^*) - N_{\mathcal{P} \cap \mathcal{M}_{\Delta,r}}(G^*) \right| \geq \frac{\epsilon n}{2}\right) \leq e^{-\delta_1 n}. \] (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 $l(d, D) < \delta_2$ we have
\[ \mathbb{P}\left(\left| N_{\mathcal{P} \cap \mathcal{M}_{\Delta,r}}(G^*) - n\mathbb{P}(T \in \mathcal{P} \cap \mathcal{M}_{\Delta,r}) \right| \geq \frac{\epsilon n}{4}\right) \leq e^{-\delta_2 n} \] (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 $l(d, D) < \delta$ we have
\[ \mathbb{P}\left(\left| N_{\mathcal{P}}(G^*) - n\mathbb{P}(T \in \mathcal{P}) \right| \leq \epsilon n\right) \geq 1 - e^{-\delta n}, \]
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 $n\mathbb{P}(|T| = \infty)$.

**Corollary 4.** Fix $\epsilon > 0$. For all sufficiently large $K$, there exists $\delta > 0$ such that if $l(d, D) < \delta$ we have
\[ \mathbb{P}\left(\left| \sum_{k \geq K} N_k(G^*) - n\mathbb{P}(|T| = \infty) \right| \leq \epsilon n\right) \geq 1 - e^{-\delta n}. \] (25)
**Proof.** For sufficiently large $K$, we have $\sum_{k=1}^{K} \mathbb{P}(|T| = k) \geq 1 - \frac{\epsilon}{2} - \mathbb{P}(|T| = \infty)$. The result follows from (20), replacing $\epsilon$ with $\frac{\epsilon}{2K}$. \(\square\)

Corollary 4 implies (21) if $\mathbb{P}(|T| = \infty) = 0$, and it will also play a key role in the next section as we address the case in which $\mathbb{P}(|T| = \infty) > 0$.

**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 $\mathbb{P}(|T| = \infty) > 0$. 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 \( \theta_1 \) 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 \( \theta_2 \) that connects to type \( \theta_1 \) 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 \( \theta_1 \) parents with type \( \theta_2 \) offspring and type \( \theta_2 \) parents with type \( \theta_1 \) offspring. Let \( G' \) denote the subgraph of \( G^* \) that we obtain by deleting edges between type \( \theta_1 \) and type \( \theta_2 \) 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 \( d' \) denote the degree sequence of \( G' \), and let \( d'' \) denote the degree sequence of \( G'' \). The sprinkling argument relies on the following lemma.

**Lemma 5.** For any \( d \) and any \( 0 < p < 1 \), the random graphs \( G' \) and \( G'' \) are conditionally independent given \( d' \).

**Proof.** This follows from the definition of the configuration model. The graph \( G^* \) is a uniform random pairing of the stubs determined by \( d \). Color each pair red, except color edges between type \( \theta_1 \) and type \( \theta_2 \) nodes blue with independent probability \( 1 - p \). Given the set of stubs in red pairs, which determines \( d' \) and \( d'' \), 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 \( \theta_1 \) and type \( \theta_2 \) 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 \( l(d, D) < \delta \) then

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

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

**Lemma 6.** Let \( \{A_1\}_{i=1}^2 \) and \( \{B_1\}_{i=1}^2 \) be disjoint sets of stubs, with \( A_1 \) and \( B_1 \) containing stubs attached to type \( \theta_1 \) nodes leading to type \( \theta_2 \) nodes, and vice versa 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 than

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

Hence, the probability that we have no matches in $B_1$ is at most

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

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

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 $i$th largest component of $G^*$, and fix $\epsilon > 0$. By Corollary 4, there are constants $K$ and $\delta > 0$ such that if $l(d, D) < \delta$, then

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

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

$$P\left(L_1 + L_2 \geq n \left(P(|T| = \infty) + \frac{\epsilon}{2}\right)\right) \leq P\left(\sum_{k \geq K} N_k(G^*) \geq n \left(P(|T| = \infty) + \frac{\epsilon}{4}\right)\right) \leq e^{-\delta n}.$$ (26)

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

$$P\left(L_1 \geq n \left(P(|T^*| = \infty) - \frac{\epsilon}{2}\right)\right) \geq 1 - e^{-\delta' n}.$$ (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 $T$, there is a positive probability of having two or more offspring. This rules out the case in which a node in $T$ always has one child, and one can check that under this assumption, the survival probability $P(|T(p)| = \infty)$ converges to $P(|T| = \infty)$. For the rest of the proof, fix a $p$ such that $P(|T(p)| = \infty) \geq P(|T| = \infty) - \frac{\epsilon}{8}$.

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 $H_{\Delta, r}$, which is satisfied if two conditions hold. First, no vertex within distance $r$ of $v$ has more than $\Delta$ 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 $\theta_1$ node with a stub in $G''$ and a type $\theta_2$ node with a stub in $G''$.

**Lemma 7.** Fix $\epsilon > 0$. We can choose $\Delta$ and $r$ for which there exists $\delta_1 > 0$ such that if $l(d, D) < \delta_1$ we have
\[
\mathbb{P} \left( N_{H_{\Delta,r}}(G^*(p)) \leq n \left( 1 - \frac{\epsilon}{8} \right) \right) \leq e^{-\delta_1 n}.
\]

**Proof.** Choosing $r$ sufficiently large ensures that, conditional on surviving until the $r$th generation, the red subtree $T_r^r(p)$ has blue stubs of both types with probability at least $1 - \frac{\epsilon}{24}$. By Lemma 4 we can find $\Delta$ so that $\mathbb{P} \left( T^r_r(p) \in M_{\Delta,r} \right) \geq 1 - \frac{\epsilon}{24}$. Consequently, we have
\[
\mathbb{P} \left( T(p) \in H_{\Delta,r} \right) \geq 1 - \frac{\epsilon}{12}.
\]
The result follows from Proposition 15. \qed

Fix the $\Delta$ 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 $M = np_{\theta_1}\mathbb{E}[D_{\theta_1} \cdot e_{\theta_2}]$ be the expected number of links between type $\theta_1$ and type $\theta_2$ nodes. By Corollary 4 (or rather, the analog based on Proposition 15), there exists $k \geq \max \left( K, \frac{M 16 \Delta r}{n^2} \right)$ and $\delta_2 > 0$ such that whenever $l(d, D) < \delta_2$ we have
\[
\mathbb{P} \left( |S_k| \leq n \left( \mathbb{P}(|T| = \infty) - \frac{\epsilon}{4} \right) \right) \leq \mathbb{P} \left( |S_k| \leq n \left( \mathbb{P}(|T'(p)| = \infty) - \frac{\epsilon}{8} \right) \right) \leq e^{-\delta_2 n}.
\]

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
\[
2^{\frac{|S_k|}{2}} \leq 2^{\frac{n}{2}} \leq e^\frac{n}{2}
\]
potentially bad cuts.

Fix a realization of $d'$ and $G'$ such that
\[
N_{H_{\Delta,r}}(G^*(p)) \geq n \left( 1 - \frac{\epsilon}{8} \right).
\]

Suppose that $(X, Y)$ is a potentially bad cut. Both $X$ and $Y$ contain at least $\frac{\epsilon n}{8}$ vertices with property $H_{\Delta,r}$. Since $k \geq \Delta r$, and no vertex in $H_{\Delta,r}$ can reach more than $\Delta r$ 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 $H_{\Delta,r}$. 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 $\delta$, 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
\[
e^{-\frac{\alpha^2 n}{M}} \leq e^{-\frac{\alpha n}{2}}.
\]
This implies that the expected number of bad cuts, given $d'$ and $G'$, is at most $e^{-\frac{n}{4}}$, and the probability of having any bad cuts is at most $e^{-\frac{n}{4}}$. If there are no bad cuts, then

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

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

**Typical Distances**

The results above allow a simple proof of the typical distance claim. We do not require the path counting arguments of van der Hofstad et al. (2005). I drop the subscript from $\nu(T,D)$ in what follows for notational convenience. In essence, we show that the neighborhood $G_{v,r}$ is well approximated by $T_r$ 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:

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

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

$$\mathbb{E}(|G_{v,r}|) = 1 + \sum_{i=1}^{r} \mathbb{E}(|D|)\nu^{i-1} = 1 + \mathbb{E}(|D|)\nu^r - 1 \nu - 1.$$

We can bound our probability as

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

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 $r$th generation of $T$, 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

$$c_\epsilon \nu^\tilde{r} - 1 \nu - 1 \leq |T_{\tilde{r} - 1}| \leq C_\epsilon \nu^\tilde{r} - 1 \nu - 1$$

or

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

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

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

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 $|N_{v,\tilde{r}}| \geq C n^{\frac{1+\epsilon}{2}}$ and $N_{w,\tilde{r}} \geq C n^{\frac{1+\epsilon}{2}}$. Moreover, the law of large numbers implies that we can choose two types $\theta_i$ and $\theta_2$ and a constant $C'$ such that $|N_{v,\tilde{r}}|$ has at least $C' n^{\frac{1+\epsilon}{2}}$ type $\theta_1\theta_2$ link stubs leading away, and $|N_{w,\tilde{r}}|$ has at least $C' n^{\frac{1+\epsilon}{2}}$ type $\theta_2\theta_1$ 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 $e^{-c n^{\epsilon}}$ for another constant $c$. Taking $n$ large, we see that with arbitrarily high probability, the distance between $v$ and $w$ is at most $2\tilde{r} = (1 + \epsilon) \log_\nu n$. □

### 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 $D_\theta$ on $\mathbb{N}$ and a distribution $Z_\theta$ on $\Theta$. First, realize the total number of link stubs according to $D_\theta$. For each stub, independently realize a type according to $Z_\theta$. Recall

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

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

$$p_\theta \mu_\theta q_\theta^{(\theta')} = p_{\theta'} \mu_{\theta'} q_{\theta}^{(\theta')}.\quad (29)$$

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 $D_\theta$ is

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

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

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

$$g_\theta'(1)s_\theta = g_\theta'(q^{(\theta)} \cdot s), \quad \theta = 1, 2, \ldots, |\Theta|.$$  

---

The above text is a mathematical section from a document discussing properties of random graphs and their configurations. It delves into the distribution of neighborhoods and the implications for the structure of these graphs, particularly focusing on the probability of link formation and the distribution of types in a random graph setting. The text is dense with mathematical notation and concepts, typical of a research paper in graph theory or probability theory.
If \( v_\theta \) is a type \( \theta \) node chosen uniformly at random, we have
\[
\lim_{n \to \infty} \mathbb{P}_{CM,n,T,D}(v_\theta \in L_1(G)) = 1 - g_\theta \left( q^{(\theta)} \cdot \xi \right) \equiv \zeta_\theta.
\]

If \( v_d \) is a node with degree tuple \( d \) chosen uniformly at random, we have
\[
\lim_{n \to \infty} \mathbb{P}_{CM,n,T,D}(v_d \in L_1(G)) = 1 - \prod_{\theta=1}^{\lvert \Theta \rvert} \xi_{d_\theta} \equiv \zeta_d.
\]

**Proof.** We apply Proposition 8 to compute the survival probability of the characteristic branching process. From the definition of the generating function \( g_\theta \) for the degree distribution \( D_\theta \), the forward distribution \( D'_\theta \) has the generating function
\[
g'_\theta(s) = g'_\theta(1).
\]

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 \( g_X(g_Z(s)) \). The generating function of \( Z_\theta \) is
\[
g_{Z_\theta}(s) = q^{(\theta)} \cdot s.
\]

By Proposition 8, the solution \( \xi \) to
\[
g'_\theta(1)s_\theta = \left( g'_\theta(1) \cdot s \right) \cdot \xi, \quad \theta = 1, 2, ..., \lvert \Theta \rvert
\]
gives the vector of extinction probabilities for a multi-type branching process with the offspring distributions \( D'(\theta) \). That is, the element \( \xi_\theta \) is the probability that a subtree of \( T_{T,D} \) beginning with a type \( \theta \) node goes extinct. The claim about nodes with a given degree tuple \( d \) is immediate from this and Theorem 1. The asymptotic probability that a random type \( \theta \) node is connected to the giant component is then
\[
1 - \sum_d \mathbb{P}(D^{(\theta)} = d) \prod_{\theta=1}^{\lvert \Theta \rvert} \xi_{d_\theta} = 1 - g_{D^{(\theta)}}(\xi) = 1 - g_\theta(q^{(\theta)} \cdot \xi).
\]

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 \( \xi \) gives the extinction probabilities for a branching process based on the forward distributions: when the characteristic branching process realizes a type \( \theta \) offspring, that branch of the process dies with probability \( \xi_\theta \). 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
\[
\zeta = \sum_{\theta=1}^{\lvert \Theta \rvert} p_\theta \zeta_\theta.
\]
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

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

Moreover, define $m_{ij} = \nu_i q^{(i)}_j$, let $M$ be the $|\Theta|$ by $|\Theta|$ matrix with $m_{ij}$ the entry in row $i$ and column $j$, and let $\nu$ 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

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

**Proof.** This is immediate from Theorem 2 and our assumptions on the degree distributions.

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 $D^{(\theta)}$ dominates $\hat{D}^{(\theta)}$.

For part (c), note that conditional on realizing degree $m$, a type $\theta$ agent has type $\theta'$ neighbors (who are potential adopters) distributed according to a binomial distribution with $m$ trials and success probability $q^{(\theta')}_{\sigma^{(\theta)}}$. 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 $D_\theta$ to realizations of $\hat{D}_\theta$ such that $\hat{D}_\theta$ realizes a value no higher than $[\hat{D}_\theta^{-1}(D_\theta)]$, and if the above inequality holds for each pair of binomial distributions that arise in this coupling, then $D^{(\theta)}$ dominates $\hat{D}^{(\theta)}$.

**Calculation for Section 2**
Fixing $\sigma$, the forward extinction probability for a high-degree player is the minimal solution to

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

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

The virality $\nu_{\sigma}$ is

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

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

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

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

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

The expected payoff from adoption is then

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

Hence, the best reply map for the diffusion game is

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

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 $\sigma(\theta, d)$ analogously to those in section 4, where $d \in \mathbb{N}^{[\Theta]}$ is a degree tuple—the entry $d_\theta$ corresponds to the number of type $\theta$ neighbors. To map strategies to diffusion

xx
outcomes, we follow the exact same analysis as in section 4, replacing the diffusivities $\zeta^{(\sigma)}_{\psi,d}$ with diffusivities $\zeta^{(\sigma)}_{\theta,d}$ 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 $\sigma_a$ for the probability that a player is willing to adopt, and write $\sigma_i$ for the probability that a player informs her neighbors. Note that the extent and rate of diffusion depends only on $\sigma_i$, while equilibrium decisions will also depend on $\sigma_a$.

The forward extinction probability is the minimal solution to

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

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

$$v (1 + 2(1 - \xi)\sigma_a) - c,$$
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 $\sigma_a$, and the probability she informs her neighbors $\sigma_i$. 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 $\xi$) 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 $\kappa$ 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 $\theta$ has a value $v \in [0, 1]$ drawn independently according to the distribution $V_\theta$. 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

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

Each player observes her type $\theta$, 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 \( \sigma : \Theta \times [0, 1] \times \mathbb{N} \rightarrow \{0, 1\}^2 \), 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 \( \sigma \) and \( \sigma_i(\theta, v, d) \) to denote the sharing decision—to economize on notation, going forward I omit the dependence on \( \theta, v, \) and \( d \) when there is little risk of confusion. The payoff from following the strategy \( \sigma \) is

\[
U(v, X, \sigma_a, \sigma_i).
\]

The value \( \kappa \), 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.

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

xxiii
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.

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 \( CM_{n,T,D} \)—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 \((d_2, d_3, \ldots, d_k)\), where \( k \geq 2 \) 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, ..., \( k \). We successively select link stubs of type \( j \) uniformly at random in groups of size \( j \), and we establish links between all associated nodes.

xxiv
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 \( j = 2, 3, ..., k \) 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 \( d_i = \sum_{j=2}^{k}(k-1)d_j^{(i)} \).

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

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

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.