J. Econ. Theory 226 (2025) 106009

Contents lists available at ScienceDirect

Journal of Economic Theory

journal homepage: www.elsevier.com/locate/jet

Minimal contagious sets: Degree distributional bounds ☆

Itai Arieli [a], Galit Ashkenazi-Golan [b], Ron Peretz [c], Yevgeny Tsodikovich [d],*

[a] Faculty of Data and Decision Sciences, Technion – Israel Institute of Technology, Israel
[b] Department of Mathematics, London School of Economics and Political Science, UK
[c] Department of Economics, Bar Ilan University, Israel
[d] Department of Economics, Ben-Gurion University of the Negev, Israel

ARTICLE INFO

Keywords:
Innovation
Diffusion
Word-of-mouth
Contagious
Attachment

ABSTRACT

Agents in a network adopt an innovation if a certain fraction of their neighbors has already done so. We study the minimal contagious set size required for a successful innovation adoption by the entire population, and provide upper and lower bounds on it. Since detailed information about the network structure is often unavailable, we study bounds that depend only on the degree distribution of the network – a simple statistic of the network topology. Moreover, as our bounds are robust to small changes in the degree distribution, they also apply to large networks for which the degree distribution can only be approximated. Applying our bounds to growing networks shows that the minimal contagious set size is linear in the number of nodes. Consequently, for outside of knife-edge cases (such as the star-shaped network), contagion cannot be achieved without seeding a significant fraction of the population. This finding highlights the resilience of networks and demonstrates a high penetration cost in the corresponding markets.

1. Introduction

The diffusion of innovation is a fundamental process that enables the spread of new technologies, products, and ideas throughout a network. Typically, this process begins with a small group of individuals, known as early adopters or innovators, who are not significantly influenced by the opinions and actions of others. They are often the first to recognize the potential of a new technology and are motivated to adopt it even if it is unproven or untested, creating positive local externalities that encourage other agents to adopt the innovation.

The process of seeding—namely, the task of identifying the group of early adopters within a social network that maximizes the spread of the innovation—has been extensively studied in various domains and contexts, including the dissemination of information, the implementation of microfinance programs, and the adoption of new technologies. Numerous studies have proposed different heuristics to enable the selection of seeds in a manner that will likely improve the outcome of the diffusion process. Many of the proposed seeding strategies rely on detailed topological and structural information about the network; however, since such information is often unavailable (especially in large networks), a more robust approach to assessing the likelihood of contagion is to rely on

☆ For their valuable comments, the authors thank Gabrielle Demange, Tim Hellman, Mathieu Leduc, Dominik Karos, Yakov Babichenko, Tizié Bene, and the participants of the 26th annual Coalition Theory Network conference (Bielefeld, 2023), the workshop on adaptive learning and opinion dynamics in social networks (Ramat Gan, 2023), the Alpine Game Theory Symposium (Grenoble, 2023), and the GSBE-ETBC Seminar (Maastricht, 2023). YT and RP acknowledge the support of the Israel Science Foundation Grant #2566/20. YT acknowledges the support of the Israel Science Foundation Grants #1626/18 & #448/22.

* Corresponding author.
E-mail address: tsodikev@bgu.ac.il (Y. Tsodikovich).

https://doi.org/10.1016/j.jet.2025.106009
Received 8 November 2023; Received in revised form 10 April 2025; Accepted 10 April 2025

Available online 14 April 2025
0022-0531/

coarser information, such as the degree distribution of the network. The bounds on the size of the seed set derived below show that, in general, full contagion is only possible by seeding a significant fraction of the population.

We investigate this problem within the context of strategic interactions in finite social networks. In particular, we consider the well-studied classical model of local strategic interactions [see, e.g., Blume (1995), Ellison (1993), and Morris (2000)], where the strategic interaction is defined by a game in which the players are the network nodes. Each player participates in a coordination game with their neighbors and can take one of two actions: a status-quo action (action 0) or an innovation action (action 1). Each agent is better off playing the innovation action if a proportion of at least $\rho \in [0,1)$ of their neighbors has adopted the innovation and plays the action 1. Similarly to the works mentioned above, we consider the Best-Response Dynamics (BRD), in which players dynamically respond to their neighbors' actions. Importantly, unlike other models of opinion dynamics and consensus formation [such as those presented in Rosenberg et al. (2009) and Bikhchandani et al. (2021)], our model does not assume the existence of either an underlying true state to be learned or a 'correct' action to be taken. The sole driving force in our model is strategic complementarities.

The BRD has two natural resting points, which are equilibria profiles of the static game: all players playing action 0, and all players playing action 1. Our main question is: how many agents must be targeted to divert the population from the status-quo equilibrium to the innovation equilibrium? As the status-quo equilibrium represents the dominant product in a market or a social norm, the answer offers an index that measures resilience to market penetration in social networks.

The answer to the abovementioned question strongly depends on the network structure. For example, in a star-shaped network, a single innovator (the central node) is always sufficient, regardless of the number of agents. While this question can be answered analytically for some specific, well-structured networks [such as grids, as in Chalupa et al. (1979)], finding the minimal contagious set in a general network is an NP-hard problem (Kempe et al., 2003). Hence, similar to Chang and Lyuu (2010), Ackerman et al. (2010), and many others, we resort to obtaining estimations rather than an exact answer. This approach has two merits: it renders the question tractable, and more importantly, our bounds rely mainly on the degree distribution of the network, rather than on its exact structure. Although the degree distribution provides only partial information about the network, we show that it is sufficient for calculating meaningful bounds. This conclusion is crucial because, in practice, the degree distribution may be the only information available on large networks. For example, while an advertiser can easily obtain information about the number of connections (i.e., "friends") a Facebook user, obtaining information about the identity of these connections is difficult or even impossible.

The main results of our study are bounds on the size of the minimal contagious set: two lower bounds (Theorems 1 and 2) and an upper bound (Theorem 5). These bounds show that, except for highly centralized structures, a linear fraction of the population must be initially seeded in order to achieve full contagion. Importantly, the bounds depend only on the degree distribution of the network and not on its detailed topology, making them robust to small changes or approximations in network structure.

We provide two variations of the lower bounds: one that is particularly suited for sparse networks and another that is more appropriate for dense networks. In both cases, the bounds can be expressed in the form αn, where $\alpha \in (0,1)$ depends continuously on the degree distribution under appropriate topologies. This allows us to analyze contagion in large networks where only partial information is available, and to quantify the resilience of such networks to innovation diffusion based on simple, aggregate characteristics. Furthermore, this continuity is useful when considering families of networks whose degree distribution is random but is known to converge to a certain limit; prominent examples include preferential attachment networks (Barabási and Albert, 1999; Albert and Barabási, 2002), wherein the degree distribution converges to a power-law distribution; networks generated from graphons (Erol et al., 2023); and large, real-life networks in which the degree distribution can only be estimated due to their size and the fact that they may gradually change over time. We derive explicit results employing this continuity in Theorems 3 and 4, which address growing sequences of networks with converging degree distributions.

An important insight from our linear lower bounds is that, except in very specific network structures such as the star network, achieving widespread innovation adoption requires seeding a substantial fraction of the population. This insight underscores the inherent resilience of networks to change and highlights a significant penetration cost in relevant markets. Additionally, our results show that such resilience, and the resulting penetration cost, can be understood by examining the degree distribution of the network. In practice, this finding implies that understanding the aggregate network characteristics—rather than the detailed network topology—is a good starting point for analyzing market penetration costs. Our findings have implications for industries and policymakers aiming to accelerate technology adoption and diffusion in real-world social networks, as well as for general marketing campaign planning.

Our upper bound (Theorem 5) only applies to networks of a particular type, which we term attachment networks (Definition 3). In these networks, the agents can be ordered such that, from a certain point m_0, each agent is connected to a fixed number of preceding agents, m; two prominent examples are trees (in which $m_0 = m = 1$) and the preferential attachment model of Barabási and Albert (1999). Similar to our lower bounds, our upper bound is linear in the network size and the linear coefficient depends on the parameters m_0, m, and ρ, as well as on the degree distribution of the network. Our proof of Theorem 5 relies on a method, inspired by Morris (2000, Prop. 2), for constructing contagious sets based on agent ordering. This method can be applied not only to attachment networks but to any type of network; for example, Angel and Kolesnik (2018) used a similar approach on Erdős-Rényi random graphs with a uniform random agent ordering. The important feature of attachment networks is that they have a natural ordering, which prescribes the construction of the contagious set. Such ordering, along with the fact that all agents have the same number of preceding neighbors, m, has enabled us to obtain an explicit upper bound on the size of the contagious set.

1.1. Related literature

Our work is primarily related to the work of Morris (2000) on the contagion of the BRD. Morris considers an infinite graph of bounded degree and investigates the maximal value of ρ such that a finite contagious set exists. As Morris notes, such a finite contagious set can exist only if $\rho < 0.5$.

In contrast to Morris, however, we focus on finite (rather than infinite) graphs and our analog of Morris's results is an asymptotic result that applies to growing sequences of graphs. Unlike Morris, we do not assume a bounded degree, and while Morris only distinguishes between finite and infinite contagious sets, we care about the size of the smallest contagious set. Another difference between our work and that of Morris is that some of our results, namely, Theorems 1 and 3, only apply to cases in which $\rho \geq 0.5$: an assumption that reflects a scenario in which innovation has no an advantage for the players over the status-quo. Therefore, the case in which $\rho \geq 0.5$ corresponds to situations where competing technologies are essentially social norms or products of the same quality. We further demonstrate the differences between $\rho < 0.5$ and $\rho \geq 0.5$ in Section 4.1, below.

In an economic context, Morris's contagion model has been applied to the study of optimal seeding in models related to ours. Erol et al. (2023) and Jackson and Storms (2023) similarly addressed seeding in a strategic diffusion context for a general ρ, but with some key differences: Erol et al. (2023) employed a continuous graphon model to approximate diffusion in large networks, while Jackson and Storms (2023) analyzed equilibrium structures in stochastic block models. Both works examined two optimal-seeding problems: maximizing the proportion of an infected set subject to size restriction on the seed set,[1] and minimizing the size of a seeding set that infects a set of a particular target size. In contrast, we focus on the BRD in *finite* networks with a proportional threshold. In addition, their work considers partial diffusion, whereas our main objective is to derive explicit bounds for complete diffusion.

Rossi et al. (2017) addressed a related issue: they studied the BRD under heterogeneous thresholds on large networks generated by the configuration model[2] focusing on the asymptotic fraction of adopters as a function of the initial seed. By employing a mean-field approach, the authors derived a one-dimensional equation that approximates the evolution of the contagion process. They demonstrate that their reduction provides a good approximation to the total number of adopters in large networks. In contrast, our bounds on the minimal seed set size, based on the degree distribution, are applied for full contagion in finite networks.

Related to the BRD is the model where the threshold for switching to the innovation action is an absolute number, r, rather than a certain proportion of the neighbors. A fixed threshold may better suit word-of-mouth or spontaneous diffusion, whereas a proportional threshold suits strategic environments where adoption decisions are endogenous. Akbarpour et al. (2018) studied the special case of the fixed threshold $r = 1$, applied to random graphs. Using the Susceptible-Infected-Recovered (SIR) diffusion model, the authors analyzed the "value of network information" in seeding problems and demonstrated that a randomly selected seed set could achieve diffusion comparable to that of optimally targeted seeding. This finding contrasts with our work, which employs a proportional threshold model; we show that in the strategic BRD model, targeting agents by their degree outperforms random seeding.

Amini and Fountoulakis (2012) studied bootstrap percolation on power-law networks, showing that a threshold of $r = 3$ ensures a sub-linear contagious set and highlighting that targeted seeding can lead to widespread adoption. Conversely, Amini et al. (2013) demonstrated that, for a random set to achieve contagion on such networks, the seed size must be linear in the size of the network, which underscores the potential inefficiency of random seeding strategies. Furthermore, Freund et al. (2018) provided density conditions on general graphs, which guarantee the existence of a contagious set of a specific size, exploring the interplay between the graph structure and the minimum contagious set size. These results highlight the importance of network structure and seeding strategies in diffusion processes with fixed threshold models, further motivating our analysis of strategic diffusion in finite networks under a proportional threshold rule.

Structure of the paper. We describe the model and the main results in Section 2, followed by the proofs in Section 3. Next, in Section 4.1, we compare majority and minority dynamics and show a discontinuity in the lower bound when the required activation threshold changes from slightly less than 50% to 50%. Then, in Section 4.2, we use regular networks to show that Theorem 1 is tight. The discussion in Section 5 concludes the paper.

2. Model and results

We consider innovation diffusion in a network by using the *Best-Response Dynamics (BRD)* with a threshold $\rho \in [0, 1)$. A social network is a finite undirected simple graph $G = (V, E)$, where V is the set of n agents (vertices) and $E \subseteq \binom{V}{2}$ is the set of edges. The set of neighbors of agent v is denoted N_v and the degree is $d_v = |N_v|$.

An object that plays a fundamental role in our analysis is the cumulative distribution function (CDF) of the degrees $F : \mathbb{R}_+ \to [0, 1]$. Given the graph $G = (V, E)$, the distribution F is defined as $F(x) = \frac{1}{n}|\{v : d_v \leq x\}|$ for every $x \in \mathbb{R}_+$. The average degree of the network is denoted by $\langle F \rangle := \int x \, \mathrm{d}F(x)$. For certain results, it is more convenient to work with continuous distributions; therefore, we denote by $\hat{F} : \mathbb{R}_+ \to [0, 1]$ the piecewise linear interpolation of F from $\mathbb{N}$ to $\mathbb{R}_+$, defined by

$$\hat{F}(x) := (1 - (x - \lfloor x \rfloor))F(\lfloor x \rfloor) + (x - \lfloor x \rfloor)F(\lceil x \rceil). \tag{1}$$

Finally, $\Delta_1(\mathbb{N})$ is defined as the space of all finite-expectation probability measures on $\mathbb{N} = \{0, 1, 2, \dots\}$, therefore, $\mathrm{d}F \in \Delta_1(\mathbb{N})$.

[1] Given an initial seed set, the set obtained after iterative application of the BRD is called the *infected set*.

[2] For definition and discussion on the configuration model, see, e.g., Van Der Hofstad (2024, Chapter 4).

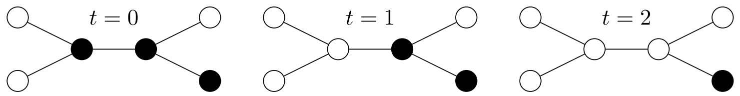

Fig. 1. An example of the dynamics with a seed of three (white) vertices for $\rho = 0.5$. At time $t = 3$, the last black vertex is activated and, therefore, $A_3 = V$ and the seed is contagious.

As in Morris (2000), each agent v plays a two-action coordination game against all his neighbors, where the actions are the status-quo action, 0, and the innovation action, 1. The payoffs of the coordination game define a cutoff threshold $\rho \in [0, 1)$, such that action 1 is the best response of agent v to his neighbors if and only if the fraction of his neighbors who play action 1 is at least ρ. Without loss of generality, indifferent agents choose action 0.

More formally, the dynamic we consider evolves as follows. At the outset, all agents in the network take action 0. A specific innovation is introduced, such that an agent can either adopt it [in which case, he is *activated* or *infected* (Garbe et al., 2018)] or not. At time $t = 0$, a set $A_0 \subset V$ of agents is activated. These agents are called *seeds*, and A_0 is called the *seed set*. While the seeds adhere to the innovation regardless of what others do, the other agents choose between the two actions on each day ($t = 1, 2, \dots$). An inactive agent adopts the innovation if and only if strictly more than a fraction ρ of his neighbors are activated. In this case, we say that the agent sees that a ρ-majority of its neighbors are activated.

Formally, the sets $A_0 \subset A_1 \subset \dots$ of active agents at times $t = 0, 1, \dots$ are defined[3] recursively by

$$A_{t+1} = A_t \cup \{v \in V : |N_v \cap A_t| > \rho d_v\}.$$

Let $A_\infty := \bigcup_{t=0}^{\infty} A_t$. When the graph G is finite, there exists t_0 such that $A_t = A_\infty$ for all $t \geq t_0$. If $A_\infty = V$, then A_0 is said to be *ρ-contagious*. When ρ is clear from the context, we omit it and say that A_0 is *contagious*. Our goal is to determine the size of the smallest contagious seed set, denoted by $h_\rho(G)$, which represents the minimal number of innovators required to shift the population from the status-quo equilibrium to the innovative equilibrium under an optimal seeding strategy, and serves as a lower bound on the contagious seed set size under sub-optimal strategies, such as random seeding. For some of our lower-bound results (Theorems 1 and 3), we discuss the case of $\rho \geq 0.5$, which models scenarios where the innovation has no intrinsic advantage over the status-quo and is adopted solely due to peer pressure. We relax this assumption both for our other lower bounds (Theorems 2 and 4) and for our upper bound (Theorem 5).

Fig. 1 depicts the network dynamics for $\rho = 0.5$, where the white vertices are active (A_t) and the black are inactive ($\bar{A}_t := V \setminus A_t$). It is easy to verify, in this example, that $h_{0.5}(G) = 2$.

Our main results are upper and lower bounds on $h_\rho(G)$, in terms of the degree distribution of the agents of G. In contrast to published results [e.g., in Chang and Lyuu (2010)] which depend only on some statistics of that distribution (e.g., the maximal degree), our results rely on the entire distribution of the degrees of the network.

2.1. Lower bounds

We first establish two general lower bounds that apply to any network (Theorems 1 and 2). Next, we present asymptotic results, derived from these general bounds (Theorems 3 and 4), which pertain to sequences of networks whose degree distributions converge in specific modes.

2.1.1. General lower bounds

We present two general lower bounds whose behavior differs with respect to the number of edges in the network. Our first lower bound (Theorem 1) decreases as the set of edges increases; therefore, this theorem is primarily useful for sparse networks, i.e., for networks with relatively few edges. Our second lower bound (Theorem 2) complements the first lower bound because it increases with the edge set under the containment order; therefore, this bound is more applicable to dense networks. Another difference between Theorems 1 and 2 is that the former applies only to $\rho \geq \frac{1}{2}$, while the latter holds for any ρ.

Theorem 1. *For every $\rho \geq \frac{1}{2}$, $n \in \mathbb{N}$, and a network G with n agents and a degree distribution F_G,*

$$h_\rho(G) \geq \alpha(F_G) n,$$

where $\alpha(F_G) = 1 - \hat{F}_G(d_*)$ and d_* is a solution of

$$\hat{F}_G(d_*) = \int_{d_*}^{\infty} \lceil x \rceil \, \mathrm{d}\hat{F}_G(x). \tag{2}$$

[3] In principle, we could allow active agents to become inactive again by postulating $v \in A_{t+1}$ if and only if $v \in A_0$, or $|N_v \cap A_t| > \rho d_v$, or $|N_v \cap A_t| = \rho d_v$ and $v \in A_t$. The two definitions are equivalent since the seeds never become inactive, and by induction on t, any agent that becomes active at time t remains active forever.

The value $\alpha(F_G)$ is obtained by identifying the quantile of the distribution F_G, such that the aggregate degree of the nodes above that quantile equals the quantile value itself. Consequently, $\alpha(F_G)$ increases with F_G and, therefore, it decreases as the edge set grows.

The bound applies when $\rho = \frac{1}{2}$ and, therefore, also when $\rho > \frac{1}{2}$. Interesting directions for future work would be to extend our calculation to identify lower bounds, α_ρ, which depend on ρ, and to obtain bounds that account for agent-dependent thresholds, i.e., distinct $\rho_v \geq \frac{1}{2}$ for each $v \in V$, as in Ackerman et al. (2010) and in Chang and Lyuu (2010).

Despite its appearance, the bound in Theorem 1 is not necessarily linear in n since $\alpha(F_G)$ may depend on n. For example, for the complete network K_n, we obtain $\alpha(F_{K_n}) = \frac{1}{n+1}$, in which case our bound is vacuous. Theorem 3 provides conditions on a family of networks $\{G_n\}_{n=1}^{\infty}$, under which the sequence of coefficients $\alpha(F_{G_n})$ converges to a positive number. Therefore, Theorem 3 ensures that the lower bound on $h_\rho(G_n)$ *is* linear in n.

Our next lower bound generalizes the simple idea that $h_\rho(G)$ must be at least ρ times the minimum degree of G (otherwise, any additional activation cannot occur). The minimal degree may reflect the degrees of outlier agents, rather than the typical ones. Theorem 2 considers this scenario, relying on low percentiles of the degree distribution instead of its exact minimum.

Theorem 2. *For every network G with n agents, and for every $\rho \in (0,1)$,*

$$h_\rho(G) > \max_{k=0,\dots,n} \{k - n \cdot F_G(k\rho^{-1})\}.$$

After establishing the general lower bounds, we turn to their asymptotic implications for sequences of networks of increasing size.

2.1.2. Asymptotic lower bounds

Our asymptotic bounds refer to families of networks whose degree distributions converge in certain modes. The results are stated in terms of the limiting distribution.

In Theorem 3, which is derived from Theorem 1, we assume that the degree distributions converge to the limit in two ways simultaneously: pointwise and in mean. Convergence in mean, in this context, implies that the average degree of the networks converges to the mean of the limiting degree distribution. The preferential attachment model (Albert and Barabási, 2002) exemplifies a process where the degree distribution converges to a specific limit (a power law) as the network grows.

Definition 1. A sequence of networks $(G_n = (V_n, E_n))_{n\in\mathbb{N}}$ of sizes $|V_n| = n$, where F_n is the CDF of the degrees of G_n, is $\mathbb{W}_1$-*converging* if there exists a distribution on $\mathbb{N}$ with a CDF F such that:

1. F is the pointwise limit of F_n, i.e., $F_n(d) \xrightarrow[n\to\infty]{} F(d)$, for every $d \in \mathbb{N}$.
2. The average degree of F_n converges to the average degree of F, i.e., $\langle F_n \rangle \xrightarrow[n\to\infty]{} \langle F \rangle$.

Definition 1 is equivalent to stating that $F_n - F \longrightarrow 0$ in L^1, or that $dF_n \to dF$ in the 1-Wasserstein metric.[4]

Theorem 3 leverages the notion of $\mathbb{W}_1$-convergence to establish an asymptotic lower bound. Specifically, it demonstrates that if a family of networks satisfies $\mathbb{W}_1$-convergence and the limiting degree distribution has a finite mean, then the lower bound on $h_\rho(G_n)$ increases linearly with the size of the network.

Theorem 3. *Let $(G_n)_{n\in\mathbb{N}}$ be a $\mathbb{W}_1$-converging sequence of networks, where each G_n has n agents, F_n is the CDF of the degrees of G_n and F is their limit distribution. If $\langle F \rangle < \infty$, then $h_\rho(G_n)$ is linear in n, i.e., there exists $\alpha > 0$ such that for all $\rho \geq \frac{1}{2}$ and for every n large enough, $h_\rho(G_n) \geq \alpha n$.*

To illustrate the significance of convergence in mean, consider the case where only a pointwise convergence of F_n to F is assumed. For example, in a star-shaped network with n agents, the limit of F_n is $F := \mathbf{1}_{[1,\infty)}$ and $\langle F \rangle = 1$, while $\langle F_n \rangle = \frac{2n-1}{n} \to 2$. Indeed, in this case, $h_\rho(G_n)$ remains constant at 1, failing to grow linearly with n. Hence, pointwise convergence alone does not ensure that $h_\rho(G_n)$ is linear in n. We believe that the finiteness of $\langle F \rangle$ [which generalizes the common assumption of bounded degree distributions, as in Candogan (2022), Manshadi et al. (2020), and Morris (2000)] is essential for the result to hold, but we have been unable to prove this notion.

The finiteness of $\langle F \rangle$ implies that the networks are sparse, i.e., the number of edges is linear in the number of nodes. Therefore, Theorem 3 applies only to families of sparse networks; dense networks are addressed in Theorem 4, which is an asymptotic derivation of Theorem 2. Another difference between Theorems 3 and 4 is the mode of convergence of the degree distributions: while Theorem 3 refers to $\mathbb{W}_1$-convergence, Theorem 4 refers to weak* convergence of the following notion of normalized degree distribution.

Definition 2. The *normalized degree CDF* of a network G with n agents $\mathcal{F}_G : \mathbb{R} \to [0,1]$ is defined by $\mathcal{F}_G(x) = F_G(nx)$.

[4] See Theorem 6.9 on Page 108 in Villani (2009).

The normalized degree CDF captures the degree distribution as a function of the fraction of agents, rather than of their absolute degree. This normalization is crucial when analyzing dense networks, where the degrees of the agents scale linearly with the number of agents. The normalization process ensures that the degree distributions are comparable across networks of different sizes.

Theorem 4. *Let $(G_n)_{n\in\mathbb{N}}$ be a sequence of networks, where each G_n has n agents, and let $\mathcal{F}_n := \mathcal{F}_{G_n}$ be their normalized degree CDFs. Suppose that there exists a CDF $\mathcal{F}$ such that $\mathcal{F}_n(x) \xrightarrow[n\to\infty]{} \mathcal{F}$ at any x at which $\mathcal{F}$ is continuous (i.e., weak* convergence of $\mathrm{d}\mathcal{F}_n$ to $\mathrm{d}\mathcal{F}$). Then $h_\rho(G_n)$ is linear in n, i.e., for every $\alpha < \sup_{x\in[0,1]}\{x - \mathcal{F}(x\rho^{-1})\}$ and for every n large enough, $h_\rho(G_n) \geq \alpha n$.*

To see that Theorem 4 applies to families of dense graphs, note that for the supremum to be positive, there must be a positive proportion $x\rho^{-1}$ of the vertices whose degree is at least xn. Therefore, the theorem is meaningful only when the number of edges is at least $\frac{1}{2}x\rho^{-1}n^2$.

2.2. Upper bound

We now determine the upper bound on the minimal contagious set size for the interesting class of attachment networks. These networks are characterized by continuously adding new agents, who connect to existing agents according to specific rules. In many of the leading models of preferential attachment networks, the construction starts with a certain network of size m_0, and then each new agent is connected to a fixed number, $m \leq m_0$, of existing agents. We call networks with the latter property (m_0, m)-*attachment networks*.

Definition 3. An (m_0, m)-*attachment network* is a network of size at least m_0, whose agents can be ordered $v_1, v_2, \dots$ such that, for every $l > m_0$, the degree of v_l in the sub-network induced by $\{v_1, \dots, v_l\}$ is equal to m.

For example, in the model of Barabási and Albert (1999), each new agent is connected to m existing agents, where the probability to connect to a specific agent v is proportional to d_v. The Barabási–Albert Model results in a network where the limiting probability density function (PDF) of the degrees is proportional to d^{-3}, $d \geq m$.

Theorem 5 provides an upper bound on $h_\rho(G)$, which applies only to attachment networks. We show that the two following sets of agents are contagious for attachment networks: the m_0 initial agents and all agents of degree at least m/ρ, and the m_0 initial agents and all agents of degree at most $m/(1-\rho)$. This finding is true for any $\rho \in (0,1)$, dropping the $\rho \geq 0.5$ requirement of Theorems 1 and 3.

Theorem 5. *Fix $\rho \in (0,1)$. For any (m_0, m)-attachment network G with n agents and degree CDF F, we have*

$$h_\rho(G) \leq m_0 + \min\left\{1 - F(\lceil \tfrac{m}{\rho} - 1\rceil), F(\tfrac{m}{1-\rho})\right\} n. \tag{3}$$

The proof relies on the following idea, which is inspired by Proposition 2 in Morris (2000). Consider an ordering of the agents $v_1, \dots, v_n$. The set of all agents for whom less than a fraction ρ of their neighbors have indices smaller than their own is a contagious set (this is proven in Lemma 1 as part of the proof of the theorem). Moreover, any contagious set contains a subset that can be obtained this way with a certain ordering of the agents. Hence, the minimal contagious set can be found among those sets when considering all possible orderings of the agents. While this statement is true for any network, attachment networks are sufficiently structured to produce the closed-form formula in Eq. (3), which is obtained by considering either the ordering that defines attachment networks, or its reversal.

The proof of Theorem 5 provides a simple way to construct sets that are known to be contagious without the need to run the ρ-majority dynamics on the network to verify that the set is indeed contagious. Therefore, it can be beneficial to limit algorithms to searching only sets of this form within the space of all permutations. An exact design of such an algorithm, and the study of the possible improvements compared to existing algorithms, is left for future research.

In the preferential attachment model of Barabási and Albert (1999) with parameters (m_0, m), a family of networks $G_1 \subset G_2 \subset \dots$ is generated such that these networks have a limit degree distribution proportional to d^{-3}, $d \geq m$. Accordingly, Theorem 5 implies that

$$\limsup_{m\to\infty} \limsup_{n\to\infty} h_\rho(G_n)/n \leq \rho^2.$$

In the case of $\rho = \frac{1}{2}$, this upper bound is smaller than the general upper bound $h_{0.5}(G) \leq \lceil n/2 \rceil$, which is attained, e.g., by the complete and the line graphs [see Ackerman et al. (2010)].

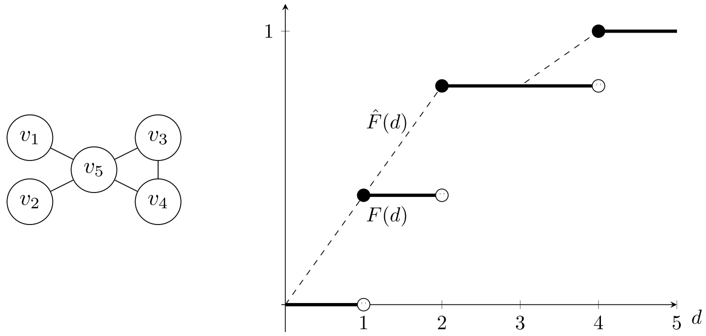

Fig. 2. An illustration of the proof of Eq. (5). The graph has $n = 5$ vertices arranged in a non-decreasing order of degrees $d_{v_1} = d_{v_2} = 1 < d_{v_3} = d_{v_4} = 2 < d_{v_5} = 4$. The graphs of $F(x)$ (solid) and $\hat{F}(x)$ (dashed) are drawn on the right. The points where $\hat{F}(x)$ crosses the heights $0, \frac{1}{5}, \frac{2}{5}, \ldots, 1$ are $d(0) = 0$, $d(1) = \frac{1}{2}$, $d(2) = 1$, $d(3) = 1\frac{1}{2}$, $d(4) = 2$, $d(5) = 4$. The proof of the case $d_{v_m} = d_{v_{m-1}}$ can be exemplified by considering $m = 2$. In the corresponding interval between $d(1) = \frac{1}{2}$ and $d(2) = 1$, the value of $\lceil x \rceil$ is $d_{v_2} = 1$ and, therefore, $\int_{d(1)}^{d(2)} \lceil x \rceil \mathrm{d}\hat{F}(x) = \frac{1}{5} d_{v_2}$. The proof of the case $d_{v_m} > d_{v_{m-1}}$ can be exemplified by considering $m = 5$. The integral is computed in the interval $[d(4), d(5)] = [2, 4]$. The function $\hat{F}$ is flat from 2 to 3, and then it climbs linearly from $\frac{4}{5}$ to 1 in the interval $[3, 4]$.

3. Proofs

3.1. Proof of Theorem 1

Our proof relies on the observation that as the innovation spreads in the network, the number of "disagreements" between neighboring agents decreases. Formally, the *boundary* (also known as *the set of discord*) of a set of agents $A \subset V$ is defined as the set of edges connecting agents in A with agents in $\bar{A}$, denoted $\partial A = \{\{v_1, v_2\} \in E | v_1 \in A, v_2 \in \bar{A}\}$.

At each time $t \geq 0$, each agent in $A_{t+1} \setminus A_t$ has more neighbors in A_t than outside of it, and, in particular, it has more neighbors in A_t than in $\bar{A}_{t+1}$, so $|\partial A_t| > |\partial A_{t+1}|$. More precisely, an agent v is activated if and only if x of its neighbors are active, where $x > \rho d_v$ (and, since x must be an integer, $x \geq \lfloor \rho d_v \rfloor + 1$). The net change in the size of the boundary due to its activation is

$$\Delta_v := |\partial\{v\} \cap \partial A_{t+1}| - |\partial\{v\} \cap \partial A_t| \leq (d_v - x) - x \leq d_v - 2\lfloor \rho d_v \rfloor - 2.$$

It follows that as the activation process unfolds, the size of the set of discord decreases by at least $|\Delta_v|$ with the activation of each new agent v. Hence, $|\partial A_t| - |\partial A_{t+1}| = \sum_{v \in A_{t+1} \setminus A_t} |\Delta_v|$ and, if A is ρ-contagious, then summing over all $t \geq 0$ gives

$$|\partial A| = \sum_{v \notin A} |\Delta_v| \geq \sum_{v \notin A} (2 + 2\lfloor \rho d_v \rfloor - d_v) \geq n - |A|, \tag{4}$$

where the last inequality follows from the fact that $\lfloor \rho d_v \rfloor \geq \frac{1}{2} d_v - \frac{1}{2}$, for $\rho \geq \frac{1}{2}$. Since $h_\rho(G)$ is non-decreasing in ρ, and since we calculate a lower bound, we assume in the remainder of this proof that, w.l.o.g., $\rho = \frac{1}{2}$.

Let $G = (V, E)$ be a network with n agents and degree CDF $F : \mathbb{R}_+ \to [0, 1]$, and let $\hat{F}$ be its linear continuation [see Eq. (1)]. For $m \in \{0, \ldots, n\}$, let $d(m)$ be the lower $\frac{m}{n}$-quantile of $\hat{F}$, namely, the first point where $\hat{F}$ reaches the height m/n, formally defined by

$$d(m) := \min\left\{ x \geq 0 : \hat{F}(x) = \frac{m}{n} \right\}.$$

Rename the agents in a non-decreasing order of degrees, such that $d_{v_1} \leq \cdots \leq d_{v_n}$. We claim that, for every $m \in \{1, \ldots, n\}$,

$$\int_{d(m-1)}^{d(m)} \lceil x \rceil \mathrm{d}\hat{F}(x) = \frac{1}{n} d_{v_m}. \tag{5}$$

The steps of the proof of Eq. (5) can be visualized by considering the example depicted in Fig. 2. For every $m \in \{1, \ldots, n\}$, there are at least m vertices whose degree is at most d_{v_m} and less than m vertices whose degree is less than d_{v_m}; therefore,

$$F(d_{v_m} - 1) < \frac{m}{n} \leq F(d_{v_m}).$$

Since $\hat{F}$ agrees with F on the integers, we have $d_{v_m} - 1 < d(m) \leq d_{v_m}$, namely, $\lceil d(m) \rceil = d_{v_m}$.

Set $d_{v_0} := 0$. If $d_{v_m} = d_{v_{m-1}}$, then $\lceil d(m) \rceil = \lceil d(m-1) \rceil = d_{v_m}$ and Eq. (5) holds. If $d_{v_m} > d_{v_{m-1}}$, then there are exactly $m - 1$ vertices whose degree is at most $d_{v_{m-1}}$ and, therefore, $d(m-1) = d_{v_{m-1}}$. Since F has no jumps between $d_{v_{m-1}}$ and $d_{v_m} - 1$, $\hat{F}(x)$ is constant in that interval. Since $\lceil d(m) \rceil = d_{v_m}$, $\lceil x \rceil \equiv d_{v_m}$ in the interval $x \in (d_{v_m} - 1, d(m)]$. It follows that

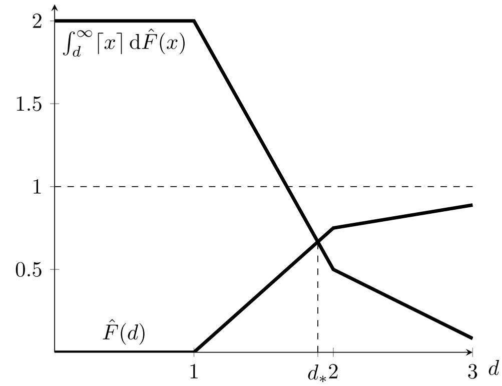

Fig. 3. The intersection of $\hat{F}(d)$ (increasing) and $\int_d^\infty \lceil x\rceil \,\mathrm{d}\hat{F}(x)$ (decreasing) for $F(d) = 1 - \lfloor \max(d,1)\rfloor^{-2}$.

$$\int_{d(m-1)}^{d(m)} \lceil x\rceil \mathrm{d}\hat{F}(x) = \int_{d(m-1)}^{d_{v_m}-1} \lceil x\rceil \mathrm{d}\hat{F}(x) + \int_{d_{v_m}-1}^{d(m)} \lceil x\rceil \mathrm{d}\hat{F}(x) = 0 + (\hat{F}(d(m)) - \hat{F}(d_{v_m}-1))d_{v_m} = (\hat{F}(d(m)) - \hat{F}(d(m-1)))d_{v_m} = \tfrac{1}{n} d_{v_m},$$

which concludes the proof of Eq. (5).

For $A \subset V$, we have by Eq. (5),

$$\tfrac{1}{n} \sum_{i=n-|A|+1}^{n} d_{v_i} = \int_{d(|A|)}^{\infty} \lceil x\rceil \,\mathrm{d}\hat{F}(x).$$

Therefore,

$$\tfrac{1}{n}|\partial A| \le \tfrac{1}{n} \sum_{v\in A} d_v \le \tfrac{1}{n} \sum_{i=n-|A|+1}^{n} d_{v_i} = \int_{n-d(|A|)}^{\infty} \lceil x\rceil \,\mathrm{d}\hat{F}(x).$$

Suppose that A is contagious, then, by Eq. (4) and since $\frac{1}{n}|A| = \hat{F}(d(|A|))$,

$$\tfrac{1}{n}|\partial A| \ge 1 - \hat{F}(d(|A|)),$$

we obtain that

$$\int_{n-d(|A|)}^{\infty} \lceil x\rceil \,\mathrm{d}\hat{F}(x) \ge 1 - \hat{F}(d(|A|)).$$

Recall that $\hat{F}(d)$ is increasing from 0 to 1, whereas $\int_d^\infty \lceil x\rceil \,\mathrm{d}\hat{F}(x)$ is decreasing from a positive number $\langle F\rangle$ to 0. Therefore, by continuity (Fig. 3), there exists a d_* such that

$$\hat{F}(d_*) = \int_{d_*}^{\infty} \lceil x\rceil \,\mathrm{d}\hat{F}(x). \tag{6}$$

Moreover, by monotonicity, $\hat{F}(d_*)$ is uniquely defined and $\hat{F}(d_*) \ge \hat{F}(d(|A|)) = \frac{1}{n}|A|$. Defining $\alpha(F) := 1 - \hat{F}(d_*)$ completes the proof of Theorem 1.

3.2. *Proof of Theorem 2*

Let G be a network with n agents. Suppose that there exists a contagious set A and a number $k \in \{0,\dots,n\}$ such that $|A| \le k - nF_G(k\rho^{-1})$. Note that k must be strictly smaller than n; otherwise, we would have $|A| = 0$.

Let $B \supset A$ be obtained by adding to A a set of $k - |A|$ agents from $\bar{A}$, whose degrees are the smallest. Since $k - |A| \ge nF_G(k\rho^{-1})$, the degrees of all agents outside B are greater than $k\rho^{-1}$. Therefore, these agents need more than k activated neighbors to become active. Since $|B| = k < n$, it cannot activate anyone outside of it, and, therefore, neither B nor $A \subset B$ are contagious.

3.3. Proof of Theorem 3

The proof of Theorem 3 lies in showing that the function $\alpha(F)$ of Theorem 1 is continuous w.r.t. the 1-Wasserstein metric. Let $(G_n)_{n\in\mathbb{N}}$ be a $\mathbb{W}_1$-converging sequence of networks, and let $(F_n)_{n\in\mathbb{N}}$ be their respective degree CDFs with F being the limiting CDF. By Theorem 1, and since $\alpha(F)>0$, it is sufficient to show that $\lim_{n\to\infty}\alpha(F_n)=\alpha(F)$.

Below, integrals of the form $\int_a^b$ should be interpreted as $\int_{(a,b]}$. We first claim that

$$\lim_{n\to\infty}\max_{k\in\mathbb{N}}\left|\int_0^k \lceil x\rceil \mathrm{d}F_n(x)-\int_0^k \lceil x\rceil \mathrm{d}F(x)\right|=0. \tag{7}$$

Let k_n, $n=1,2,\dots$, be maximizers in Eq. (7). If $\{k_n : n=1,2,\dots\}$ is bounded, then the claim holds since $F_n(d)\xrightarrow{n\to\infty}F(d)$ for any integer $d\ge 0$. It remains to show that for any sub-sequence $n'\to\infty$ such that $k_{n'}\to\infty$, we have

$$\limsup_{n'\to\infty}\left|\int_0^{k_{n'}} \lceil x\rceil \mathrm{d}F_{n'}(x)-\int_0^{k_{n'}} \lceil x\rceil \mathrm{d}F(x)\right|=0.$$

Let $\varepsilon>0$ and $m>0$ be large enough so that

$$\int_0^m \lceil x\rceil \mathrm{d}F(x)+\varepsilon>\int_0^\infty \lceil x\rceil \mathrm{d}F(x)=\langle F\rangle.$$

Since $\lim_{n\to\infty}\langle F_n\rangle=\langle F\rangle$ and $\lim_{n\to\infty}\int_0^m\lceil x\rceil \mathrm{d}F_n(x)=\int_0^m\lceil x\rceil \mathrm{d}F(x)$, we have

$$\limsup_{n\to\infty}\int_m^\infty \lceil x\rceil \mathrm{d}F_n(x)<\varepsilon.$$

It follows that

$$\begin{aligned}\limsup_{n'\to\infty}\left|\int_0^{k_{n'}} \lceil x\rceil \mathrm{d}F_{n'}(x)-\int_0^{k_{n'}} \lceil x\rceil \mathrm{d}F(x)\right| &\le \limsup_{n'\to\infty}\left|\int_0^{m} \lceil x\rceil \mathrm{d}F_{n'}(x)-\int_0^{m} \lceil x\rceil \mathrm{d}F(x)\right|\\ &+\limsup_{n'\to\infty}\int_m^\infty \lceil x\rceil \mathrm{d}F_{n'}(x)+\limsup_{n'\to\infty}\int_m^\infty \lceil x\rceil \mathrm{d}F(x)<0+\varepsilon+\varepsilon.\end{aligned}$$

This is true for any $\varepsilon>0$, thus proving Eq. (7).

Let $\hat F$ (respectively, $\hat F_n$) be the piece-wise linear interpolation from $\mathbb{N}$ to $\mathbb{R}_+$ of F (respectively, F_n), as in Eq. (1). For each n, denote by d_n the solution of Eq. (2) with respect to $\hat F_n$. We next show that

$$\lim_{n\to\infty}\left|\hat F(d_n)-\int_{d_n}^\infty \lceil x\rceil\,\mathrm{d}\hat F(x)\right|=0. \tag{8}$$

We will use the following observations regarding every CDF F supported on $\mathbb{N}$, and for every $k\in\mathbb{N}$:

(O1) $\int_k^{k+1}\lceil x\rceil \mathrm{d}\hat F=\int_k^{k+1}\lceil x\rceil \mathrm{d}F=(k+1)(F(k+1)-F(k))$, and, therefore, $\int_k^\infty\lceil x\rceil \mathrm{d}\hat F=\int_k^\infty\lceil x\rceil \mathrm{d}F$;

(O2) The mapping $d\mapsto\int_d^\infty\lceil x\rceil \mathrm{d}\hat F(x)$ is an affine function over the domain $d\in[k,k+1]$.

For every n,

$$\begin{aligned}\left|\hat F(d_n)-\int_{d_n}^\infty \lceil x\rceil\,\mathrm{d}\hat F(x)\right| &\le \left|\hat F(d_n)-\hat F_n(d_n)\right|+\left|\hat F_n(d_n)-\int_{d_n}^\infty \lceil x\rceil\,\mathrm{d}\hat F_n(x)\right|+\left|\int_{d_n}^\infty \lceil x\rceil\,\mathrm{d}\hat F_n(x)-\int_{d_n}^\infty \lceil x\rceil\,\mathrm{d}\hat F(x)\right|\\ &\le \|\hat F-\hat F_n\|_\infty+0+\max_{k\in\{\lfloor d_n\rfloor,\lceil d_n\rceil\}}\left|\int_k^\infty \lceil x\rceil \mathrm{d}F_n(x)-\int_k^\infty \lceil x\rceil \mathrm{d}F(x)\right|\end{aligned}$$

$$\leq \|\hat{F} - \hat{F}_n\|_\infty + \left|\langle F_n \rangle - \langle F \rangle\right| + \max_{k\in\mathbb{N}} \left| \int_0^k \lceil x \rceil \mathrm{d}F_n(x) - \int_0^k \lceil x \rceil \mathrm{d}F(x) \right| \xrightarrow{n\to\infty} 0.$$

In the second inequality, we first used (O2) to replace d_n by $k \in \{\lfloor d_n \rfloor, \lceil d_n \rceil\}$, and then applied (O1). Equation (8) implies that any partial limit of the sequence $(d_n)_{n\in\mathbb{N}}$ solves Eq. (2) with respect to $\hat{F}$, and, moreover, that d_n is bounded because the set of solutions to this equation is bounded.

Suppose, for the sake of contradiction, that $\alpha(F_n)$ does not converge to $\alpha(F)$. Then there is a subsequence n' in which $\limsup_{n'\to\infty} |\alpha(F_n) - \alpha(F)| > 0$ and a further subsequence n'' in which $d_{n''}$ converges to some d_∞. We have a contradiction by

$$\begin{aligned} |\alpha(F_{n''}) - \alpha(F)| &= |\hat{F}(d_\infty) - \hat{F}_{n''}(d_{n''})| \\ &\leq |\hat{F}(d_\infty) - \hat{F}_{n''}(d_\infty)| + |\hat{F}_{n''}(d_\infty) - \hat{F}_{n''}(d_{n''})| \\ &\leq \|\hat{F} - \hat{F}_{n''}\|_\infty + |d_\infty - d_{n''}| \xrightarrow{n''\to\infty} 0. \end{aligned}$$

This concludes the proof of Theorem 3.

3.4. *Proof of Theorem 4*

Let G_n be a sequence of networks, where n is the number of agents and $\mathcal{F}_n$ is the normalized degree CDF of G_n. Suppose there exists a CDF $\mathcal{F}$ such that $\mathrm{d}\mathcal{F}_n \longrightarrow \mathrm{d}\mathcal{F}$ in the weak* topology. Theorem 2 states that

$$\alpha_n := \max_{k=0,\dots,n} \left\{ \tfrac{k}{n} - \mathcal{F}_n\left(\tfrac{k}{n}\rho^{-1}\right)\right\} < \tfrac{h_\rho(G_n)}{n}. \tag{9}$$

It is sufficient to show that for every $x \in [0,1]$,

$$\alpha(x) := x - \mathcal{F}(x\rho^{-1}) \leq \liminf_{n\to\infty} \alpha_n.$$

Since the continuity points of $\mathcal{F}$ are dense in $[0,1]$ and $\mathcal{F}$ is right continuous at any point, it is sufficient to show that $\liminf_{n\to\infty} \alpha_n \geq \alpha(x)$ for any x at which $\mathcal{F}$ is continuous. Let $x \in [0,1]$ be an arbitrary such point. For $n \in \mathbb{N}$, let $y_n := \lfloor nx \rfloor / n$, which is x rounded down to the nearest integer multiple of $1/n$. Since $\mathcal{F}_n$ is non-decreasing, we have

$$\alpha_n \geq y_n - \mathcal{F}_n(y_n\rho^{-1}) \geq x - \tfrac{1}{n} - \mathcal{F}_n(x\rho^{-1}) \xrightarrow[n\to\infty]{} \alpha(x).$$

This concludes the proof of Theorem 4.

3.5. *Proof of Theorem 5*

An ordering of the agents is an injection $\pi : V \hookrightarrow \mathbb{N}$. We say that $u \in V$ is a left neighbor of $v \in V$ with respect to π if $vu \in E$ and $\pi(u) < \pi(v)$. The left degree of $v \in V$ is defined as the number of left neighbors of v, and it is denoted by $d_v^L(\pi) = |\{u \in N_v : \pi(u) < \pi(v)\}|$. For every $\rho \in (0,1)$, consider

$$\Lambda_\rho(\pi) = \{v \in V : d_v^L(\pi) \leq \rho d_v\},$$

i.e., the set of agents who have at most ρd_v left neighbors with respect to the ordering π. The two following lemmas provide a characterization of minimal contagious sets.

Lemma 1. *For every network $G = (V, E)$, every ordering of the agents $\pi : V \hookrightarrow \mathbb{N}$, and every $\rho \in (0,1)$, the set $\Lambda_\rho(\pi)$ is ρ-contagious.*

Proof. Let $A \subset V$ be the set of agents not activated by $\Lambda_\rho(\pi)$. If $A \neq \emptyset$, let $v \in A$ be the leftmost agent of A, namely, $v = \arg\min\{\pi(u) : u \in A\}$. Since $v \notin \Lambda_\rho(\pi)$, the left degree of v is strictly greater than ρd_v; accordingly, more than ρd_v of his neighbors are active. This implies that v is activated in the next step of the ρ-majority dynamics, which is a contradiction. We conclude that $A = \emptyset$ and $\Lambda_\rho(\pi)$ is contagious. □

Lemma 2. *For every network $G = (V, E)$, every $\rho \in (0,1)$, and every ρ-contagious set $A \subset V$, there is an ordering $\pi : V \hookrightarrow \mathbb{N}$ such that $\Lambda_\rho(\pi) \subset A$.*

Proof. Let $A = A_0 \subset A_1 \subset \dots$ be the sets of activated agents at times $0, 1, \dots$, respectively. Let $\pi : V \hookrightarrow \mathbb{N}$ be an ordering of the agents that agrees with the order of the sets $\{A_t\}$, that is, if $v \in A_t$ and $u \notin A_t$, then $\pi(v) < \pi(u)$. We claim that $\Lambda_\rho(\pi) \subset A$, which will conclude the proof of the lemma.

Let $v \notin A$. We must show that $v \notin \Lambda_\rho(\pi)$. Since A is contagious, there exists $t > 0$ such that $v \in A_t \setminus A_{t-1}$. Since all the neighbors of v who are in A_{t-1} are placed before him by π and form a ρ-majority of its neighbors, we obtain $v \notin \Lambda_\rho(\pi)$. □

Lemmas 1 and 2 immediately imply the following proposition:

Proposition 1. *For every $\rho \in (0,1)$, and for every network G,*

$$h_\rho(G) = \min_{\pi:\, V \hookrightarrow \mathbb{N}} |\Lambda_\rho(\pi)|.$$

We are now ready to prove Theorem 5.

Proof of Theorem 5. Let $G = (V, E)$ be an (m_0, m)-attachment network of size n. By definition, the agents can be indexed $V = \{v_1, \dots, v_n\}$ such that, for every $l > m_0$, there are exactly m edges between v_l and $\{v_1, \dots, v_{l-1}\}$. The induced ordering is defined as $\pi_I :\, v_i \mapsto i$ and the reverse ordering is defined as $\pi_R :\, v_i \mapsto n + 1 - i$.

Note that, if $v_i \in \Lambda_\rho(\pi_I)$, then either $i \le m_0$ or $d_{v_i} \ge m/\rho$. Therefore,

$$|\Lambda_\rho(\pi_I)| \le m_0 + (1 - F(\lceil \tfrac{m}{\rho} - 1 \rceil))n.$$

Similarly, if $v_i \in \Lambda_\rho(\pi_R)$, then either $i \le m_0$ or $d_{v_i} \le \frac{m}{1-\rho}$. Therefore,

$$|\Lambda_\rho(\pi_R)| \le m_0 + F(\tfrac{m}{1-\rho})n.$$

By Lemma 1, both $\Lambda_\rho(\pi_I)$ and $\Lambda_\rho(\pi_R)$ are contagious, which concludes the proof. $\square$

4. Examples

4.1. Discontinuity at $\rho = 0.5$

An interesting aspect of our index $h_\rho(G)$ is the sharp phase transition that it exhibits around $\rho = 0.5$. Theorems 1 and 3 provide linear lower bounds on $h_{0.5}(G)$ (and, therefore, also on $h_\rho(G)$ for $\rho \ge 0.5$). Below, we show that $h_\rho(G)$ may be discontinuous for $\rho = 0.5$, which may explain why these lower bounds do not exist for $\rho < 0.5$.

Let us denote

$$h_{0.5-}(G) := \lim_{\rho \nearrow 0.5} h_\rho(G).$$

While the difference between $h_{0.5}(G)$ and $h_{0.5-}(G)$ is either small or nonexistent in some networks, it is dramatic in others. On one extreme, we find the complete graph K_n, for which $h_{0.5}(K_n) = \lceil n/2 \rceil$ and $h_{0.5-}(K_n) = \lfloor n/2 \rfloor$. Another example is any graph in which all the degrees are odd, for which $h_{0.5}(G) = h_{0.5-}(G)$. On the other extreme, we find the line-shaped graph L_n, for which $h_{0.5}(L_n) = \lfloor n/2 \rfloor$ whereas $h_{0.5-}(L_n) = 1$.

To illustrate the entire range of possible gaps between $h_{0.5}$ and $h_{0.5-}$, we consider the m-dimensional tori with $n = k^m$ vertices, denoted C_k^m. Specifically, consider the m-fold product of a k-cycle graph, $C_k^m = (V, E)$ where $V = \{1, \dots, k\}^m$, and

$$E = \left\{ \{x, y\} \in \binom{V}{2} : \exists i \in [m]\ x_i - y_i = \pm 1 \mod k,\ x_j = y_j\ \forall j \ne i \right\}.$$

Proposition 2, below, states that $h_{0.5-}(C_k^m) = \theta_m \left(n^{1-1/m}\right)$, whereas, by Proposition 3, $h_{0.5}(C_k^m) \ge \frac{n}{m+1}$.

Proposition 2. *For every $k, m \in \mathbb{N}$,*

$$\lfloor k/2 \rfloor^{m-1} \le h_{0.5-}(C_k^m) \le n - (k-1)^m \le m n^{1-1/m}$$

Proof. To prove the upper bound, consider the set $W = V \setminus \{2, \dots, k\}^m$. This set clearly has the required size. It remains to show that it is contagious w.r.t. any $\rho < 0.5$.

Following Morris (2000), a set of vertices $B \subseteq V$ is called *ρ-cohesive* if, for every $v \in B$, a proportion of at least ρ of its neighbors is inside B, namely, $|N_v \cap B| \ge \rho |N_v|$, and a set is ρ-contagious if and only if it intersects all the nonempty $(1-\rho)$-cohesive sets.

Let $\rho < 0.5$ and suppose, for the sake of contradiction, that there exists a nonempty $(1-\rho)$-cohesive set B that does not intersect W, namely, $\emptyset \ne B \subset \{2, \dots, k\}^m$. Let $x = (x_1, \dots, x_m) \in \arg\min_{y \in B} \sum_{i=1}^m y_i$. Since $\deg(x) = 2m$ and B is $(1-\rho)$-cohesive, there must be an i such that $x - e_i \in B$ (as well as $x + e_i \mod k$), where $e_1, \dots, e_m$ is the standard basis of $\mathbb{R}^m$. However, $x - e_i$ contradicts the minimality of x.

To prove the lower bound, it is sufficient to find $\lfloor k/2 \rfloor^{m-1}$ disjoint $(1-\rho)$-cohesive sets. Indeed, for any $i_1, \dots, i_{m-1} \in \{1, \dots, \lfloor k/2 \rfloor\}$, the set

$$\{2i_1 - 1, 2i_1\} \times \cdots \times \{2i_{m-1} - 1, 2i_{m-1}\} \times \{1, \dots, k\}$$

is $(1-\rho)$-cohesive and these sets are disjoint. $\square$

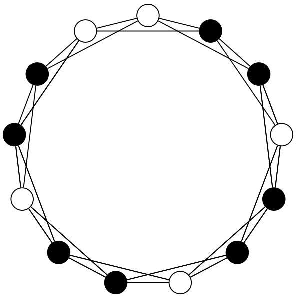

Fig. 4. A Cayley graph with $n = 13$, $d = 4$, and a minimal contagious set (in white).

4.2. Regular networks

Here, we discuss the tightness of Theorem 1 by referring to regular networks of varying degrees. We define a d-regular network as a network in which the degree of all the agents is d, and we first determine the lower bound obtained for such networks. As in the proof of Theorem 1, it is sufficient to consider $\rho = 0.5$. By Eq. (4), we obtain two different expressions for odd and even d.

$$h_{0.5}(G) \geq \begin{cases} \frac{n}{d+1} & d \text{ is odd,} \\ \frac{2n}{d+2} & d \text{ is even.} \end{cases}$$

For the seed to be able to activate any agent, its size must be greater than ρd. We thus have,

$$h_{0.5}(G) \geq \begin{cases} \frac{d+1}{2} & d \text{ is odd,} \\ \frac{d+2}{2} & d \text{ is even.} \end{cases}$$

Combining the two bounds, we obtain the following proposition.

Proposition 3. *For any d-regular network G and any $\rho \geq 0.5$, we have*

$$h_{\rho}(G) \geq \begin{cases} \max\{\frac{n}{d+1}, \frac{d+1}{2}\} & d \text{ is odd,} \\ \max\{\frac{2n}{d+2}, \frac{d+2}{2}\} & d \text{ is even.} \end{cases}$$

Note that Proposition 3 implies that $h_{0.5}(G) = \Omega(\sqrt{n})$ for any regular graph. The following examples prove the tightness of Proposition 3 up to a multiplicative factor of two. For any $d \geq 2$, we construct a growing family of networks along with 0.5-contagious sets whose sizes are obtained by replacing the “maximum” in the expression of Proposition 3 with a “sum.”

For an even d (i.e., Eulerian graphs), consider the Cayley graph of $\mathbb{Z}_n := \mathbb{Z}/n\mathbb{Z}$ with generators $\{\pm 1, \dots, \pm \frac{d}{2}\}$.

Consider the following set of size $\frac{2n}{d+2} + \frac{d}{2}$ (see Fig. 4):

$$A_0 = \{0, \tfrac{d}{2}+1, 2(\tfrac{d}{2}+1), \dots\} \cup \{-1, -2, \dots, -\tfrac{d}{2}\}.$$

This set is contagious since, by induction, $t \in A_t$ for all $t = 0, 1, \dots$.

For an odd d, we construct an example for n divisible by $d+1$: this construction can be slightly modified to fit any size n. Consider the Cayley graph of $\mathbb{Z}_2 \times \mathbb{Z}_{n/2}$ with the generators $\left\{(1, x) : -\frac{d-1}{2} \leq x \leq \frac{d-1}{2}\right\}$.

The set $A_0 = \{0\} \times \left(\left\{-1, \dots, -\frac{d-1}{2}\right\} \cup \frac{d+1}{2}\mathbb{Z}_{n/2}\right)$ of size $\frac{n}{d+1} + \frac{d-1}{2}$ is shown below to be contagious, and the construction is depicted in Fig. 5. Since $A_0 \supset \{0\} \times \left(\left\{0, \dots, -\frac{d-1}{2}\right\} \cup \left\{\frac{d+1}{2}\right\}\right)$,

$$A_1 \supset \{1\} \times \left\{1, \dots, -\tfrac{d-1}{2}\right\},$$

and, therefore,

$$A_2 \supset \{0\} \times \left(\left\{1, \dots, -\tfrac{d-1}{2}\right\} \cup \left\{\tfrac{d+1}{2}\right\}\right).$$

It follows, by induction on t, that

$$A_{2t-1} \supset \{1\} \times \left\{t, \dots, -\tfrac{d-1}{2}\right\}$$

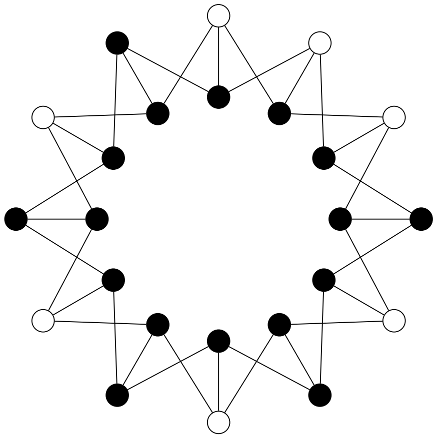

Fig. 5. A bipartite graph with $n = 24$ and $d = 3$, corresponding to $\mathbb{Z}_2 \times \mathbb{Z}_{12}$, and a minimal contagious set (in white).

and

$$A_{2t} \supset \{0\} \times \left(\left\{ t, \dots, -\frac{d-1}{2} \right\} \cup \left\{ \left\lfloor \frac{2t}{d+1} + 1 \right\rfloor \frac{d+1}{2} \right\} \right),$$

for every $t \geq 1$.

5. Discussion

We study innovation diffusion in the classical product adoption model and provide a lower bound on the seed size, which depends on the size of the network and the degree distribution (Theorems 1 and 2). We show that for a wide range of networks of increasing size, the bound is linear with respect to the network size, provided that their degree distribution converges (Theorems 3 and 4). This linear dependence suggests that, except in specific network structures such as star networks, achieving widespread adoption necessitates seeding a substantial fraction of the population—a finding that highlights the inherent resilience of networks to change and the significant cost of achieving high penetration. Moreover, we show that this network resilience and the high penetration cost can be understood by examining the degree distribution of the network. Finally, for the class of attachment networks, we provide an upper bound for the minimal contagious set (Theorem 5), which can be easily extended for other well-structured networks.

Our study sheds light on the dynamics of innovation diffusion, offering practical insights for industries and policymakers striving to promote technology adoption and diffusion. By elucidating the minimal contagious set size required for successful innovation adoption, we can develop targeted seeding strategies to leverage early adopters and accelerate the diffusion process. In practice, this implies that understanding aggregate network characteristics, rather than detailed network topology, serves as a good starting point for analyzing market penetration costs.

For our lower bound, we focus on asymptotic results where the network size increases, and we do not assume that the degree distribution is bounded. This approach contrasts with previous studies which either considered networks of fixed sizes or sequences of networks with growing sizes (including infinite networks) but under the assumption that the degree is uniformly bounded (Candogan, 2022; Manshadi et al., 2020; Morris, 2000). In addition, we study best-response dynamics on attachment networks, which induce a proportional activation threshold.

Our work regarding growing networks focuses on one aspect of this type of network: the converging degree distribution. We do not consider other properties of the network, such as the inner connectivity of high-degree vertices. Taking this and other properties into account can help close some of the gap between our upper and lower bounds. Future research directions may also include elucidating the path of the contagion process inside the network, its convergence time, and the optimal seeding strategy.

Notably, unlike other models [e.g., Amini et al. (2013)], our model is scalable such that, to activate a linear fraction of the network, the seed set must itself be a linear fraction of the network; for example, activating half of the network requires at least (an order of) $0.5h_\rho(G)$ seeds.

For our upper bound, we do not impose any restrictions on the network or on the activation threshold. Therefore, our upper bound can be used for any network with either majority or minority dynamics. Since we assume that the network is an attachment network only to obtain the formula in the last step of the proof of Theorem 5, both the theorem and its proof can be applied to non-attachment networks that are sufficiently structured, resulting in a different formula. Moreover, the proof provides a simple way to construct contagious sets, which can be used to design an improved algorithm to find an approximation to the seed set; this endeavor, however, is left to future research.

CRediT authorship contribution statement

Itai Arieli: Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Galit Ashkenazi-Golan:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Ron Peretz:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Yevgeny Tsodikovich:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

Data availability

No data was used for the research described in the article.

References

Ackerman, E., Ben-Zwi, O., Wolfovitz, G., 2010. Combinatorial model and bounds for target set selection. Theor. Comput. Sci. 411 (44–46), 4017–4022.

Akbarpour, M., Malladi, S., Saberi, A., 2018. Diffusion, seeding, and the value of network information. In: Proceedings of the 2018 ACM Conference on Economics and Computation, p. 641.

Albert, R., Barabási, A.-L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74 (1), 47.

Amini, H., Fountoulakis, N., 2012. What I tell you three times is true: bootstrap percolation in small worlds. In: Internet and Network Economics: 8th International Workshop, WINE 2012, Proceedings 8. Liverpool, UK, December 10–12, 2012. Springer, pp. 462–474.

Amini, H., Fountoulakis, N., Panagiotou, K., 2013. Discontinuous bootstrap percolation in power-law random graphs. In: The Seventh European Conference on Combinatorics, Graph Theory and Applications: EuroComb 2013. Springer, pp. 431–436.

Angel, O., Kolesnik, B., 2018. Sharp thresholds for contagious sets in random graphs. Ann. Appl. Probab. 28 (2), 1052–1098.

Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512.

Bikhchandani, S., Hirshleifer, D., Tamuz, O., Welch, I., 2021. Information Cascades and Social Learning.

Blume, L.E., 1995. The statistical mechanics of best-response strategy revision. Games Econ. Behav. 11 (2), 111–145.

Candogan, O., 2022. Persuasion in networks: public signals and cores. Oper. Res. 70 (4), 2264–2298.

Chalupa, J., Leath, P.L., Reich, G.R., 1979. Bootstrap percolation on a Bethe lattice. J. Phys. C, Solid State Phys. 12 (1), L31.

Chang, C.-L., Lyuu, Y.-D., 2010. Bounding the number of tolerable faults in majority-based systems. In: Algorithms and Complexity: 7th International Conference, CIAC 2010, Proceedings 7. Rome, Italy, May 26–28, 2010. Springer, pp. 109–119.

Ellison, G., 1993. Learning, local interaction, and coordination. Econometrica, 1047–1071.

Erol, S., Parise, F., Teytelboym, A., 2023. Contagion in graphons. J. Econ. Theory 211, 105673.

Freund, D., Poloczek, M., Reichman, D., 2018. Contagious sets in dense graphs. Eur. J. Comb. 68, 66–78.

Garbe, F., Mycroft, R., McDowell, A., 2018. Contagious sets in a degree-proportional bootstrap percolation process. Random Struct. Algorithms 53 (4), 638–651.

Jackson, M.O., Storms, E., 2023. Behavioral communities and the atomic structure of networks. ArXiv preprint arXiv:1710.04656.

Kempe, D., Kleinberg, J., Tardos, É., 2003. Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137–146.

Manshadi, V., Misra, S., Rodilitz, S., 2020. Diffusion in random networks: impact of degree distribution. Oper. Res. 68 (6), 1722–1741.

Morris, S., 2000. Contagion. Rev. Econ. Stud. 67 (1), 57–78.

Rosenberg, D., Solan, E., Vieille, N., 2009. Informational externalities and emergence of consensus. Games Econ. Behav. 66 (2), 979–994.

Rossi, W.S., Como, G., Fagnani, F., 2017. Threshold models of cascades in large-scale networks. IEEE Trans. Netw. Sci. Eng. 6 (2), 158–172.

Van Der Hofstad, R., 2024. Random Graphs and Complex Networks. Cambridge University Press.

Villani, C., 2009. Optimal Transport: Old and New, vol. 338. Springer.