

Control and spread of contagion in networks

John Higgins*

Tarun Sabarwal†

john.higgins@ku.edu

sabarwal@ku.edu

May 27, 2021

First draft: November 2020

Abstract

We study proliferation of an action in a network coordination game that is generalized to include a tractable, model-based measure of virality to make it more realistic. We present new algorithms to compute contagion thresholds and equilibrium depth of contagion and prove their theoretical properties. These algorithms apply to arbitrary connected networks and starting sets, both with and without virality. Our algorithms are easy to implement and help to quantify relationships previously inaccessible due to computational intractability. Using these algorithms, we study the spread of contagion in scale-free networks with 1,000 players using millions of Monte Carlo simulations. Our results highlight channels through which contagion may spread in networks. Small starting sets lead to greater depth of contagion in less connected networks. Virality amplifies the effect of a larger starting set and may make full network contagion inevitable in cases where it would not occur otherwise. It also brings contagion dynamics closer to a type of singularity. Our model and analysis can be used to understand potential consequences of policies designed to control or spread contagion in networks.

JEL Numbers: C62, C72

Keywords: Network games, coordination games, contagion, algorithmic computation

*University of Kansas, Lawrence, KS, USA

†University of Kansas, Lawrence, KS, USA

1 Introduction

Proliferation of contagion in connected networks has become an increasingly central topic in society recently. For example, misinformation campaigns in social networks have grown into an alarming problem with consequences that are endangering the very fabric of society. Instances of misinformation include political smear campaigns, anti-vaccination campaigns, climate change skepticism, conspiracy theories, and more [Vosoughi et al., 2018]. Political misinformation circulated during elections exacerbates an already fractious political climate, intensifying divisions across party lines [Grinberg et al., 2019], and allegations of malfeasance and fraud undermine trust in the election process [Frenkel, 2020]. Health misinformation regarding the COVID-19 pandemic has proliferated rapidly and may be causing significant long term damage to public health [Brennen et al., 2020]. Vaccine misinformation has contributed to a decline in vaccination rates, posing another risk to public health [Broniatowski et al., 2018]. This is especially pertinent in the context of COVID-19, where an effective mitigation strategy relies on high vaccination rates. Misinformation has engendered persistent distrust of scientific evidence supporting climate change, preventing the adoption of measures to mitigate further impacts of warming. Crimes against minorities may also have been precipitated by misinformation on social networks. The problems of misinformation are exacerbated by the rapid and widespread diffusion of information in online social networks [Varol et al., 2017] and [Shao et al., 2018]. Collectively, such campaigns are a significant and growing threat to both democratic institutions and society at large.

Even though different misinformation campaigns have different goals, their structural characteristics are similar. These characteristics include the introduction of misinformation by a small group of participants in the network who rely on social network dynamics for proliferation and fan the flames along the way. Proliferation is achieved both by person-to-person spread through individual connectivity in the network and by viral effects based on similar activity by others in different parts of the network. The objective of the misinforming body is typically achieved with partial proliferation of misinformation on the network.

Contagion of an infectious disease in a society of connected individuals follows the same structural pattern. Contagion occurs both due to localized interaction among neighbors (family, friends, teammates, colleagues, and so on) and due to globalized interaction resulting from travel of carriers among different parts of the network or interaction with non-neighbors in public situations.

We study these characteristics of proliferation of misinformation or an infectious disease using a network coordination game. The model and its analysis are applicable to many other situations, including regime change, technology adoption, bank runs, currency crises, run on groceries in a pandemic, marketing new products, segregation and desegregation, success of social platforms, agglomeration in urban economics, and others.

Consider a network in which each person has a preference to coordinate with their friends (or network neighbors) on the choice of actions. Suppose this preference dictates that I find it beneficial to choose an action if a sufficiently large fraction of my friends also choose that action. This may cause me to choose an action based on the choices of some of my friends, which may cause some of my other friends to choose the same action, and in turn, cause additional friends of friends to choose the same action, propagating a chain reaction and causing contagion.

Theoretical properties of contagion arising from individual decisions in coordination games on networks in the manner above have been studied in the existing literature. A central result of interest in this literature is the following. Suppose there are finitely many players in a connected network. Each player can take one of two actions, 0 or 1, and each player is initially playing 0. Suppose an initial subset of players S exogenously switches to play 1 and we are interested in characterizing when best response dynamics starting from S leads to all players in the network playing 1. As a parameter of the network game, suppose that playing 1 is optimal for a player, if, and only if, fraction $q \in [0, 1]$ of the player's neighbors play 1. The parameter q is derived analytically from an underlying tradeoff between benefit of coordination b and cost of miscoordination c to yield $q = \frac{c}{b+c}$. In this sense, q may be viewed as the relative cost of miscoordination as well as a measure of network resilience to contagion. The higher is q , the harder it is for a player to play 1. An adaptation of [Morris, 2000] in [Jackson, 2008] shows that *contagion from S to the entire network occurs if, and only if, the complement of S is uniformly no more than $(1 - q)$ -cohesive*. Recall that a set A of players is $(1 - q)$ -cohesive if each player in A has at least fraction $1 - q$ of their neighbors in A . Set A is uniformly no more than $(1 - q)$ -cohesive if every nonempty subset of A is at most $(1 - q)$ -cohesive.

In order to apply this result, we need to check that every subset of the complement of S is at most $(1 - q)$ -cohesive. When the number of players in a network is large and the starting set is small, this is difficult to implement. For example, with a network size of 1,000 and a starting set size of 100, this would involve checking the cohesiveness of 2^{900} subsets of players. If it took only one nanosecond to check the cohesion of each subset, it would require roughly 2.68×10^{254} years for such an algorithm to complete one run (this is about 2×10^{244} times longer than the age of the universe). If, in addition, we also want to understand how the analysis changes with changes in S , q , and network connectivity, this is impossible with current computing capability. This has been a major obstacle for more than 20 years in applying these results directly to simulation and applications.

In addition to the computational obstacle, the existing result applies only in coordination games without virality, that is, where a player's payoff depends solely on the actions of their neighbors. It does not apply to the case when a player's best response depends on actions of both their neighbors and the rest of the network.

We solve both problems in this paper. We extend the existing theoretical model of a network coordination

game to include a new, flexible, and tractable formulation of virality. Virality is the notion that each player’s decision depends not just on decisions of their neighbors but also on an aggregate based on decisions taken by others in the network. (This is different from the definition sometimes used in the network science literature, which defines an event as viral if the number of times it spreads using person-to-person interaction exceeds a threshold. We use the term depth of contagion for such an incident and include it naturally in our analysis. Our definition allows for a global impact beyond explicit person-to-person spread, capturing a useful feature of the phenomena we study.) We include virality in a manner that generalizes the existing model and includes the case of no virality as a natural special case. Our formulation allows a natural parametric flexibility to study the differential impact of virality on a player’s decision.

Moreover, we provide new and tractable algorithms that can be applied to an arbitrary network coordination game with virality and starting set S to compute equilibrium depth of contagion starting from S and q , and also compute all q for which contagion spreads from S to the entire network in equilibrium. Our algorithms operate at the level of players rather than subsets of players, thereby reducing the computational burden exponentially. In simulations using an ordinary desktop, our algorithms quickly compute contagion thresholds for different networks of 1,000 players, different starting set sizes, and different network parameters. We report results based on millions of runs of our algorithms.

We prove several theoretical properties of our algorithms including a characterization of Nash equilibria in terms of fixed points of Algorithm 1, and starting from an arbitrary set S , the generation of the smallest Nash equilibrium that contains S . Algorithm 1 also has a bootstrapping property that is very useful in studying spread of contagion in an computationally tractable manner. Algorithm 2 delivers a sequence of strictly increasing Nash equilibria and a corresponding strictly decreasing sequence of q to yield the largest q for which contagion occurs from S to the entire network. This gives a new and computationally tractable characterization of the set of all q for which contagion occurs from S to the entire network.

Our algorithms are natural in the sense that they use best response dynamics as their foundation. They are new, general, and transparent in the sense that they apply to all network coordination games with virality and include virality effects explicitly. They are practically useful in the sense that they may be applied using computing power that is easily accessible to most practitioners. Therefore, these algorithms can be used by practitioners to study the effect of different formulations of virality, and also study interaction and tradeoffs among connectivity, virality, network resilience, and other variables of interest. This makes them useful for a potentially wide range of applications. Pseudocode for the algorithms is included to aid implementation.

We apply our theoretical results to study the spread of contagion in the class of Barabási-Albert scale-free networks [Barabási and Albert, 1999] using Monte Carlo methods. Scale-free networks are used widely in modeling social networks and naturally capture the idea of preferential attachment. Given a connectivity parameter m (a positive integer signifying the number of existing nodes to which each new node is connected),

the B-A procedure generates a scale-free network of the desired size.

For each parameter $m \in \{5, 10, 20\}$, we generate 40 scale-free networks of 1,000 players each using the B-A procedure. For each network, we randomly sample 50 starting sets S each of sizes 10 through 990 (at increments of 10). We also consider virality parameters $\alpha \in \{0, 0.5, 1\}$ with $\alpha = 0$ signifying the model without virality, $\alpha = 1$ for the model with virality, and $\alpha = 0.5$ for the case of intermediate impact of virality. For each combination (m, α) of the parameters and for each starting set S , we use our algorithms to compute the contagion threshold and depth of contagion. This yields 1,782,000 complete runs of our algorithms. We conduct robustness checks with additional variations in parameters and conduct several million additional runs of our algorithms.

Several features of contagion on networks emerge in the data. As expected, increasing the starting set size increases both equilibrium depth of contagion and the interval of q for which contagion occurs to the whole network, across the range of parameters.

There is a natural complementarity between starting set size and virality. The larger is starting set size the greater is the impact of virality on depth of contagion. A large starting set directly influences more people to play 1 and increases the impact of virality because more people are playing 1 in the aggregate. This complementarity naturally captures a snowball effect of contagion resulting from increasing returns. Indeed, virality can make this snowball effect irreversible at some point. In low or moderate connectivity networks ($m = 5, 10$) with virality ($\alpha = 1$), if about 40 percent of the network has an incentive to play 1, there is no hope of stopping contagion to the entire network regardless of the relative cost of miscoordination (that is, regardless of the value of q). Without virality, a relative miscoordination cost of $q > 0.584$ is sufficient to stop contagion to the entire network.

Increasing connectivity reduces the depth of contagion and the range of q for which contagion occurs to the entire network. Moreover, in the absence of virality, it is harder to spread contagion at higher q . For $q = 0.75$, there is no proliferation of action 1 beyond the initial set until about 50 percent of the network is already infected for $m = 10$. For $m = 20$, there is no additional proliferation until about 60 percent of the network is affected. With virality, contagion spreads deeper and with smaller starting set sizes. The same computation with virality shows that proliferation starts to occur beyond the initial set for sets as low as 28 percent of the network when $m = 10$ and 30 percent when $m = 20$. For a fixed m , the impact of virality is larger when q is larger.

Virality increases the range of starting set sizes that lead to full contagion and brings contagion dynamics closer to a type of singularity in the sense that there is a narrow interval of starting set sizes below which there is no contagion and above which contagion occurs to the entire network (also termed a *tipping point* in the literature). This interval is about 6-10 percentage points (as a proportion of network size) across the range of parameters.

We can compute an answer to the reverse question as well. Given a depth of contagion, what is the smallest starting set size to achieve this depth? We show that in order to spread contagion to a given depth of the network, smaller starting sets are needed if connectivity is lower (m is low), or relative miscoordination cost is lower (q is low), or virality impact is higher (α is high), or a combination of these. In a network with low connectivity ($m = 5$) and low resilience ($q = 0.25$), to spread contagion to 50 percent of the network requires a starting set size of 6 percent of the network without virality and 5 percent with virality. When resilience is increased to a moderate level ($q = 0.5$), this increases to a starting set size of 16 percent with virality and 27 percent without virality.

Virality has a larger impact on reducing starting set sizes when network resilience is high. In a network with low connectivity ($m = 5$) and high resilience ($q = 0.75$), spreading contagion to the whole network requires a starting set size of 68 percent of the network without virality but only 31 percent with virality, a decrease of 54.4 percent. The larger marginal contribution of virality when q is high can be explained by the degree of complementarity between the two parameters.

Our analysis provides insight into the design of policies that may control or spread contagion on networks. A larger starting set increases the spread of contagion. Therefore, one obvious way to limit spread of contagion is to restrict starting set sizes. For misinformation, this may involve identifying and limiting accounts designed mainly to spread misinformation. For infectious disease, proliferation can be stopped more easily by attempts to contain the spread when the proportion of the infected population is low, perhaps with measures such as quarantining.

Virality amplifies the effect of a larger starting set, and therefore, in networks where virality is a common feature, it is easier to curtail spread of misinformation by nipping it in the bud. Waiting for things to play out will add to the likelihood of a considerably larger spread of contagion. For misinformation, the impact of virality may be curtailed by policies that restrict spread of incorrect or misleading information. For infectious diseases, the effect of virality can be curtailed by restricting cross-network travel and also by limiting non-neighbor interaction in public situations. Taking the reverse actions increases the spread of contagion.

Indeed, virality brings the spread of contagion closer to a type of singularity or tipping point, and therefore, even in more resilient and connected networks, an ability to control virality may be the difference between localized contagion around the starting set or widespread contagion throughout the network. In other words, even in resilient social networks, policies to identify and reduce the sharing of misleading information may be the difference between small-scale spread of misinformation or its contagion to the entire network. Similarly, even in more resilient populations, policies to limit cross-network interaction may be the difference between local containment of an infectious disease or its global spread.

More broadly, our research provides a framework to analyze the effects of different policies in a situation-

dependent manner. The scope of application is enlarged given the many other situations across different fields and disciplines that may be modeled using this framework.

The paper proceeds as follows. Section 2 presents the related literature. Section 3 presents the model of network coordination games with virality, section 4 presents the new algorithms, proves their theoretical properties, and presents an illustrative example. Section 5 is devoted to a Monte Carlo study of scale-free networks using results from section 4. Section 6 provides some policy implications and concludes.

2 Related literature

There is a vast literature that studies contagion using different models and tools. Some of the related literature is as follows.

This paper belongs to the literature on contagion models in network coordination games. In particular, we study and extend the model of contagion on networks proposed by [Morris, 2000] and its adaptation to finitely many players in [Jackson, 2008] to include a version of virality. [Young, 2011] uses a coordination game on a network to model social innovation and the development of social norms, and provides theoretical results regarding speed of diffusion using perturbed best response dynamics that depend on neighbor actions. [Chwe, 2000] defines the concept of a minimal dependence network and uses it to characterize when coordination is possible in networks. [Jackson and Storms, 2019] investigate behavioral communities where people in some parts of the network adopt a given behavior while people in other parts of the network do not. Comprehensive frameworks for network games are presented in [Galeotti et al., 2010], [Jackson and Zenou, 2015], and [Bramoullé and Kranton, 2016], and book length treatments are available in [Jackson, 2008] and [Goyal, 2009]. Applications to financial networks and contagion are studied in [Elliott et al., 2014], to public goods in [Elliott and Golub, 2019], to innovation in [Dasaratha, 2020], and person-to-person spread in diffusion games is studied in [Sadler, 2020]. Models of learning in social networks are studied in [Golub and Jackson, 2010] and an extensive survey is available in [Golub and Sadler, 2016].

Applications of models of contagion on networks have been stymied by the difficulty of characterizing equilibrium using computationally tractable algorithms. [Jackson and Yariv, 2006], [Jackson and Yariv, 2007] address this by studying diffusion properties using the more tractable mean-field analysis. They also use a Bayesian game formulation where agents are unsure of the larger network in which they reside and use symmetric Bayesian equilibrium as the equilibrium concept. The model in this paper does not require these additional features.

Mean-field models are an example of the more general class of models based on population-level arguments, which have proved to be tractable and yielded valuable insights. These models require same treatment of players with the same degree and cost structure, regardless of their position in the network. (An early

model in the sociology literature is given in [Granovetter, 1978]. More recent examples are in [Watts, 2002] and [Wiedermann et al., 2020]. Applications using experiments include [Centola and Macy, 2007] and [Centola, 2010].) These models rule out situations in which a player closer to a starting set is infected, but a player with the same characteristics further away is not infected. Our model allows for these situations in equilibrium and is based more explicitly on additional micro-foundations, including allowing for heterogeneity in local network topology and starting set location.

Population-based models are related to another literature that is more common in epidemiology based on models of the form Susceptible-Infected-xxx, which includes Susceptible-Infected-Susceptible (SIS), Susceptible-Infected-Removed (SIR) and their many additional variants. These models help to quantify important aspects of disease transmission. They typically do not have an explicit network structure and are based on random matching derived from exogenously specified probabilities of being in particular states. In particular, a common homogeneous mixing assumption implies that each individual may have contact with any other individual, and therefore, even one infected individual eventually leads to infection of any individual in the community. This is the not case in the model studied here.

There is another literature based on bootstrap percolation models from physics ([Adler, 1991]). Those models use node activation based on number or neighbors rather than fraction of neighbors. They don't have a virality feature. Our work may be viewed as providing a foundation for those models in the decentralized and interdependent decision-making framework of game theory and Nash equilibrium, and we extend those models to include virality.

There are additional non-network based game theory models that study contagion. Global games models of contagion include the model proposed by [Carlsson and Van Damme, 1993], and its developments in [Morris and Shin, 2003], [Frankel et al., 2003], [Hoffmann and Sabarwal, 2019], and several others. Applications to signaling in global games, dynamic global games, and coordinating diffusion risk are studied in [Angeletos et al., 2006], [Angeletos et al., 2007], and [Basak and Zhou, 2020]. This literature focuses more on equilibrium selection by embedding a given game in a larger global game using uncertainty in observables and investigating how an equilibrium in dominant strategies in particular states may cause contagion to one of multiple equilibria in other states. These models typically use iterated deletion of strictly dominated strategies as a selection mechanism and do not impose a network structure.

3 Network coordination games with virality

We follow [Morris, 2000] and its adaptation in [Jackson, 2008] to define a network coordination game as follows. Consider a canonical, 2×2 coordination game in which each player chooses action 0 or 1. Action 0 may be viewed as a baseline action and 1 as a new action. The payoff from coordinating on the baseline

action is normalized to 0 for each player. The benefit from coordinating on action 1 is $b \geq 0$ for each player. The cost of deviating to 1 when the other player stays at baseline is $c \geq 0$ for the deviating player. We assume that at least one of b or c is positive, or equivalently, $b + c > 0$. Payoffs are summarized as follows.

	1	0
1	b, b	$-c, 0$
0	$0, -c$	$0, 0$

A *network* is a pair $(\mathcal{I}, \mathbf{g})$, where $\mathcal{I} = \{1, \dots, I\}$ is a finite set of players (or nodes) in the network and \mathbf{g} specifies which players are connected to each other. For each player i in network $(\mathcal{I}, \mathbf{g})$, the neighbors of i are $N_i = \{j \in \mathcal{I} \setminus \{i\} \mid ij \in \mathbf{g}\}$ and the degree of i is the number of its neighbors, denoted $d_i = |N_i|$. A network $(\mathcal{I}, \mathbf{g})$ is *connected*, if for every pair of players i and j , there is a sequence of indices $(i_1 = i, i_2, \dots, i_{k-1}, i_k = j)$ such that $i_\ell i_{\ell+1} \in \mathbf{g}$ for every $\ell = 1, \dots, k-1$.

In a network, incentives for a player to coordinate actions with their friends or neighbors (or more generally, persons to whom they are connected) may be viewed using the 2×2 coordination game above. Suppose action 0 represents a baseline or mainstream belief and action 1 represents a deviation from baseline belief. For example, playing 0 may entail abiding by some societal norm, adhering to a traditional policy, or in the context of misinformation, believing an official narrative. Playing 1 could entail a departure from an established societal norm, adopting a new standard or policy, or choosing to believe a narrative that contradicts an official narrative.

Each person's payoff depends on the choice of their neighbors or friends. If player i plays 1 and a neighbor j of player i also plays 1, player i garners some benefit due to solidarity, conformity effects, or other positive effects of coordination. Or, seeing a neighbor or friend deviate to 1 can provide justification for one's own deviation. This is modeled as benefit $b \geq 0$ that player i receives for each neighbor j with whom i coordinates on playing 1. However, playing 1 is not without its drawbacks. By deviating from the *status quo*, one breaks with existing social norms or accepted policy. For example, suppose a citizen is faced with the choice of expressing belief in 0 (a widely accepted narrative) or 1 (a conspiracy theory). Playing 1 risks conflict and estrangement from those playing 0. Those who accept the mainstream narrative may see those who play 1 as misguided, gullible, or even nefarious. Moreover, disagreement regarding beliefs may create tension, disagreement, and strife in relationships. This is modeled as cost $c \geq 0$ that player i playing 1 incurs for each neighbor j who plays the baseline action 0. For convenience, payoff to a player from playing baseline action 0 is normalized to be 0, so that b and c represent benefits and costs of deviation from baseline.

A **network coordination game** (without virality) is a tuple $\Gamma = (\mathcal{I}, \mathbf{g}, (A_i, u_i)_{i \in \mathcal{I}})$, where $(\mathcal{I}, \mathbf{g})$ is a connected network, the action space of player $i \in \mathcal{I}$ is $A_i = \{0, 1\}$, the joint action space is $A = A_1 \times \dots \times A_I$, and player i payoff, $u_i : A \rightarrow \mathbb{R}$, is given by $u_i(a_i, a_{-i}) = \sum_{j \in N_i} \hat{u}_i(a_i, a_j)$, where $\hat{u}_i(a_i, a_j)$ is player i payoff

from playing the 2×2 coordination game with neighbor j .

In a network coordination game, player i has an incentive to play 1, if, and only if, $u_i(1, a_{-i}) \geq u_i(0, a_{-i})$, which holds exactly when $\frac{1}{d_i} \sum_{j \in N_i} a_j \geq \frac{c}{b+c}$. Let $s_i = \sum_{j \in N_i} a_j$ be the number of neighbors of i who play 1 and let $q = \frac{c}{b+c} \in [0, 1]$ be a reduced form parameter summarizing relative cost of miscoordination. In this case, **player i has incentive to play 1, if, and only if, $\frac{s_i}{d_i} \geq q$** . In other words, 1 is a best response of player i exactly when the fraction of player i 's neighbors who are playing 1 is at least q .

The parameter q may be viewed as an index of network resilience to contagion. When player i plays 1, b indexes intensity of benefit of coordination with neighbors who play 1, c indexes intensity of cost of miscoordination with neighbors who play 0, and $q = \frac{c}{b+c}$ measures the relative cost of miscoordination with neighbors who play 0. When q is low, miscoordination cost is relatively low, and it is easier for player i to switch to 1 (easier for $\frac{s_i}{d_i} \geq q$ to hold). When q is high, miscoordination cost is relatively high and it is harder for player i to switch to 1. In other words, in a network coordination game with low q , the network is less resilient to contagion and in one with high q , the network is more resilient to contagion.

An important feature of spread of contagion on networks is the role of virality. In its simplest form, virality is the notion that player i may take action 1 when a large number of other players take action 1 even if they are not network neighbors of i . In social networks, this may be achieved by publishing trending topics, or the numbers of likes or retweets that are visible to every player. For infectious diseases, this may be achieved when travel to different parts of the network is opened up or restrictions on interactions in common places like markets are lifted. In other words, contagion on networks may be affected by their ability to display or bring ‘‘closer’’ the action of the rest of the network in some aggregated manner.

We extend the network coordination game to include virality as follows. Given player i , and a profile a_{-i} of actions of other players in the network, let $p_i(a_{-i}) = \frac{1}{I-d_i-1} \sum_{j \notin N_i \cup \{i\}} a_j$ be the fraction of the network (not including i or neighbors of i) that plays 1. Player i gets some additional benefit from playing 1 that depends positively on this aggregate proportion. This is formalized by a (weakly) increasing function $\phi_i : [0, 1] \rightarrow \mathbb{R}_+$ such that payoff to player i from playing 1 is given by $u_i(1, a_{-i}) = \phi_i(p_i(a_{-i})) + \sum_{j \in N_i} \hat{u}_i(1, a_j)$. When convenient, we suppress the notation a_{-i} in $p_i(a_{-i})$.

A **network coordination game with virality** is a network coordination game in which payoff of each player i is given by $u_i(1, a_{-i}) = \phi_i(p_i(a_{-i})) + \sum_{j \in N_i} \hat{u}_i(1, a_j)$, and $u_i(0, a_{-i}) = 0$, where $\phi_i : [0, 1] \rightarrow \mathbb{R}_+$ is a (weakly) increasing function. It follows immediately that a network coordination game (without virality) is the special case when each ϕ_i is identically zero.

In a network coordination game with virality, player i has incentive to play 1, if, and only if, $u_i(1, a_{-i}) \geq u_i(0, a_{-i})$. Rearranging terms shows that **player i has incentive to play 1, if, and only if, $\frac{s_i}{d_i} \geq \frac{c}{b+c} - \frac{\phi_i(p_i)}{d_i(b+c)}$** . As ϕ_i is (weakly) increasing, it follows that the threshold $\frac{c}{b+c} - \frac{\phi_i(p_i)}{d_i(b+c)}$ to play 1 decreases in p_i . In other words, in the presence of virality, for any player i to play 1, a smaller fraction of their

neighbors need to play 1 as compared to the case without virality. In this sense, higher virality reduces network resilience to contagion.

To ensure that the threshold fraction $\frac{c}{b+c} - \frac{\phi_i(p_i)}{d_i(b+c)}$ is nonnegative, it is necessary and sufficient that $\phi_i(p_i) \leq cd_i$. We impose this constraint going forward. Moreover, the assumption that $\phi_i(p_i)$ is nonnegative ensures that the threshold fraction is at most 1. There are no other restrictions on ϕ_i in terms of continuity, differentiability, concavity, and so on.

We may rewrite the condition that player i has an incentive to play 1 as $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)} \geq q$, where $q = \frac{c}{b+c}$. When each ϕ_i is identically zero, the model specializes to the case without virality yielding the condition $\frac{s_i}{d_i} \geq q$ in the earlier literature. With virality, there is an additional contribution of b and c in a player's decision, as shown in the term $\frac{\phi_i(p_i)}{d_i(b+c)}$. In this case, considering a reduced form q alone is not sufficient to analyze the model, because there is an indeterminacy among the infinitely many b and c that yield the same q . Without loss of generality, we adopt a normalization to address this by fixing c to be a positive constant throughout the paper, and considering the one-to-one mapping $b \mapsto q = \frac{c}{b+c}$. Intuitively, this normalization fixes miscoordination cost at some baseline value $c > 0$ (for example, $c = 1$) and coordination benefit b is measured relative to baseline cost.

As usual, a profile of actions $a^* = (a_1^*, \dots, a_I^*)$ is a (pure strategy) **Nash equilibrium at q** , if for every player i , a_i^* is a best response to a_{-i}^* .

Following [Morris, 2000], a profile of actions $a = (a_1, \dots, a_I)$ is identified naturally with the subset of players who play 1, that is, with $E = \{i \in \mathcal{I} \mid a_i = 1\}$, and conversely, each subset E of \mathcal{I} is identified with the profile in which player i plays 1, if, and only if, $i \in E$. Thus, if $E \subset \mathcal{I}$ is the set of players playing 1 and $\mathcal{I} \setminus E$ is the set of players playing 0, we may write the fraction of neighbors of i who play 1 as $\frac{s_i}{d_i} = \frac{|N_i \cap E|}{d_i}$ and the fraction of the network excluding i and neighbors of i that plays 1 as $p_i = \frac{|E \setminus (N_i \cup \{i\})|}{I - d_i - 1}$, where, as usual, $|A|$ denotes cardinality of set A . Also following [Morris, 2000], we assume that in case of indifference between 0 and 1, a player plays 1. It follows that in a network contagion game with virality, **set $E \subset \mathcal{I}$ is a Nash equilibrium at q , if, and only if, for every $i \in E$, $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)} \geq q$ and for every $i \in \mathcal{I} \setminus E$, $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)} < q$** . This characterization is helpful in the analysis below.

For a computationally tractable formulation that we use in Monte Carlo simulations, suppose $\phi_i(p_i) = \alpha cd_i p_i$, where $\alpha \in [0, 1]$ is a parameter capturing the intensity of virality on payoffs. This captures the different impacts succinctly. The function ϕ_i is (weakly) increasing in p_i (and strictly increasing when $\alpha > 0$). The term cd_i captures the scale effect of virality to keep the threshold nonnegative. The term α captures intensity of impact of virality; when $\alpha = 0$, virality has no impact on payoff and $\alpha = 1$ indexes the full impact of virality. The formulation is computationally tractable in the sense that *player i has incentive to play 1, if, and only if, $\frac{s_i}{d_i} \geq q(1 - \alpha p_i)$* . Both α and p_i serve to lower the switching threshold in an intuitive and analytically tractable manner. There is a complementarity between α and p_i ; the higher is one, the

greater is the impact of the other to lower the threshold to play 1.

4 Contagion algorithms with virality

A central problem in the study of contagion on networks is to understand the proliferation of an action in the network. Suppose there is an initial set S of players who have an incentive to play 1 (either based on the configuration of play in the network or due to an exogenous switch). Suppose player $i \notin S$ is initially playing 0. If a sufficiently large proportion of i 's neighbors start to play 1, player i finds it optimal to play 1 as well. This may further induce some of player i 's other neighbors who were playing 0 to play 1, spreading action 1 further along. This is the underlying dynamic for person-to-person spread. Moreover, as more players in the network play 1, this knowledge may induce a given player to play 1 even if fewer of the player's neighbors are playing 1. This is the virality impact. Our model allows for both effects simultaneously to determine best response to a given configuration of play in the network.

Several questions are of interest. Starting from an arbitrary set of players S playing 1 at q , how far will action 1 proliferate in the network? What characteristics does this limit set have? Can we compute this limit? How does this limit depend on S and q ? Under what conditions is the limit set the entire network?

For network coordination games without virality, [Morris, 2000] characterizes equilibrium behavior in terms of cohesiveness of subsets of players in a network, as follows. For $r \in [0, 1]$, subset S of \mathcal{I} is *r-cohesive*, if every $i \in S$ has at least proportion r of its neighbors in S . Equivalently, S is *r-cohesive*, if, and only if, $\min_{i \in S} \frac{s_i}{d_i} \geq r$, where $s_i = |N_i \cap S|$ is the number of neighbors of i in S . The cohesiveness of S is the largest r such that S is *r-cohesive*. Set S is *uniformly no more than r-cohesive*, if every nonempty subset of S is no more than *r-cohesive*. [Jackson, 2008] adapts [Morris, 2000] to characterize when contagion occurs from S to \mathcal{I} at q in terms of uniform cohesion. Both results are as follows.

Theorem 1. *Let Γ be a network coordination game (without virality) and $q \in [0, 1]$.*

1. *[Morris, 2000] A nonempty set $E \subset \mathcal{I}$ is a Nash equilibrium at q , if, and only if, E is q -cohesive and $\mathcal{I} \setminus E$ is $(1 - q)$ -cohesive.*
2. *[Jackson, 2008] Contagion occurs from S to \mathcal{I} at q , if, and only if, $\mathcal{I} \setminus S$ is uniformly no more than $(1 - q)$ -cohesive.*

Proof. Confer Propositions 9.7 and 9.8 in [Jackson, 2008]. □

Both statements provide helpful insights about the role of mutual connectivity of sets of neighbors, but are hard to implement computationally. Statement (1) is an equilibrium condition. In order to find such an equilibrium set E , we have to consider each nonempty subset S of \mathcal{I} and check whether it is q -cohesive

and its complement is $(1 - q)$ -cohesive. This involves potentially checking $2^{|I|}$ sets, which is computationally infeasible even for moderately sized networks. Similarly, statement (2) involves checking cohesiveness of each subset of the complement of S . This involves checking $2^{|I \setminus S|}$ sets, which is also computationally infeasible as network size increases. The problem is exacerbated if we also want to understand how the analysis changes with changes in S , q , virality, and network connectivity.

Both statements in Theorem 1 apply to network coordination games without virality. Neither statement is necessarily true in network coordination games with virality, because in that case, actions of non-neighbors of i also influence the choice of player i , whereas cohesion conditions are local conditions about mutual connectivity.

We consider network coordination games with virality and present algorithms to compute the extent of proliferation of action 1 starting from an arbitrary subset S of players playing 1, an arbitrary threshold $q = \frac{c}{b+c} \in [0, 1]$, and using best response dynamics with virality. We prove several theoretical properties of our algorithms including a characterization of Nash equilibria in terms of fixed points of Algorithm 1, and the generation of Nash equilibria starting from an arbitrary set S . Algorithm 2 delivers the largest q for which contagion occurs from S to the entire network and characterizes the set of all q for which contagion occurs from S to the entire network.

Our algorithms are natural in the sense that they use best response dynamics as their foundation. They are new, general, and transparent in the sense that they apply to all network coordination games with virality and include virality effects explicitly. They are practically useful in the sense that in many problems they can be applied using computing power that is easily accessible to most practitioners. (We use millions of iterations in Monte Carlo simulations of networks with 1,000 players using a regular desktop computer.) Therefore, these algorithms can be used by practitioners to explore the effect of different formulations of virality, and also interaction and tradeoffs among connectivity, virality, network resilience, and other variables of interest. This makes them useful for a potentially wide range of applications.

Let S be a nonempty set of players with an incentive to play 1 and $q = \frac{c}{b+c} \in [0, 1]$. That is, for every $i \in S$, $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)} \geq q$. Set S may include players who have an incentive to play 1 based on the configuration of play in the network. It can also include players who have been infected exogenously to play action 1 or have an interest in proliferating action 1 in the network by choosing 1 themselves. These latter type of players want to play 1 regardless of decisions of others. This can be formalized by positing that for each such player i , $\phi_i(p_i)$ is constant at cd_i , and therefore, $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)} = \frac{s_i}{d_i} + \frac{c}{b+c} \geq q$. This makes action 1 (weakly) dominant for such players. Alternatively, we can specify $u_i(1, a_{-i}) = 1$ for such players. This makes action 1 (strictly) dominant for these players. Starting at S and q , Algorithm 1 computes how far action 1 proliferates in the network.

Algorithm 1 (Depth of contagion starting from (S, q))

1. Let S be a nonempty set of players with incentive to play 1 and $q = \frac{c}{b+c} \in [0, 1]$. Define $C_0 = S$ and $B_0 = \mathcal{I} \setminus S$.
2. For $t \geq 0$, recursively define
$$F_{t+1} = \left\{ i \in B_t \mid \frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)} \geq q \right\}, C_{t+1} = C_t \cup F_{t+1}, \text{ and } B_{t+1} = \mathcal{I} \setminus C_{t+1}.$$
3. Let T be the first t such that $C_{t+1} = C_t$. Let $C(S, q) = C_T$.

Algorithm 1 starts with an initial, nonempty set $C_0 = S$ of players who have an incentive to play 1. At each step t , the algorithm identifies players not in the contagion set C_t who have an incentive to play 1 and includes them in C_{t+1} . Notice that computation of p_i in F_{t+1} depends on C_t , which is the set of players who play 1 in the previous step. This notation is suppressed for convenience. A new player has an incentive to play 1 for two reasons. First, as more people play 1, their neighbors may find it beneficial to play 1 as well, due to the neighborhood coordination effect. Second, as more people play 1, every player in the network has a growing incentive to play 1 due to the virality effect. Both effects are incorporated in the model-based threshold computation $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)}$. The algorithm uses this to compute F_{t+1} , identifying players who have an incentive to play 1 at each stage and including these players in the contagion set C_{t+1} . With this new information on who is playing 1, the algorithm considers the remaining players $B_{t+1} = \mathcal{I} \setminus C_{t+1}$ and loops back to start the process again. The algorithm stops when there are no more players left to flip, or the first t such that $F_{t+1} = \emptyset$. As the number of players is finite, the algorithm stops in finitely many iterations (there are never more than the number of players in $\mathcal{I} \setminus S$). The ending contagion set is denoted $C(S, q)$. Moreover, when convenient, we denote stage t dependence of the algorithm on the starting parameters (S, q) by $F_t = F_t(S, q)$, $C_t = C_t(S, q)$, and $B_t = B_t(S, q)$. As indicated in Algorithm 1, its application assumes that the initial set S is nonempty.

Theorem 2 (Properties of Algorithm 1). *Let Γ be a network coordination game with virality.*

1. *A nonempty subset E of \mathcal{I} is a Nash equilibrium at q , if, and only if, $C(E, q) = E$.*
2. *For every (S, q) , $C(S, q)$ is a Nash equilibrium at q .*
3. *For every q , if $S \subseteq S'$, then $\forall t$, $C_t(S, q) \subseteq C_t(S', q)$, and therefore, $C(S, q) \subseteq C(S', q)$.*
4. *For every S , if $q' \leq q$, then $\forall t$, $C_t(S, q) \subseteq C_t(S, q')$, and therefore, $C(S, q) \subseteq C(S, q')$.*
5. *For every (S, q) , $C(S, q)$ is the smallest Nash equilibrium at q that contains S .*
6. *Consider (S, q) and let $A = C(S, q)$. If $q' \leq q$, then $C(A, q') = C(S, q')$.*

Proof. Statement (1) is proved as follows. If a nonempty E is a Nash equilibrium at q , then for every $i \in E$, $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)} \geq q$, and for every $i \notin E$, $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)} < q$, and therefore, Algorithm 1 terminates in one step ($T = 0$) to yield $C(E, q) = E$. In the other direction, suppose for a nonempty starting set E at q , $C(E, q) = E$. Then as E is a starting set, $i \in E$ implies $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)} \geq q$, and by the algorithm, $i \notin C(E, q)$ implies $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)} < q$, whence E is a Nash equilibrium.

Statement (2) is proved as follows. If the algorithm starts at (S, q) and stops at $E = C(S, q)$, then restarting it at E with no change in q will terminate it in one step yielding $C(E, q) = E$. Statement (1) now implies that $C(S, q)$ is a Nash equilibrium at q .

Statement (3) is proved as follows. Suppose $S \subseteq S'$. We use induction on t . For $t = 0$, $C_0(S, q) = S \subseteq S' = C_0(S', q)$. Suppose $C_t(S, q) \subseteq C_t(S', q)$. Let $A = C_t(S, q)$, $A' = C_t(S', q)$, and $E = A' \setminus A$ so $A \cup E = A'$. Consider (A, q) and (A', q) . We note that $C_0(A, q) = A$ and $B_0(A, q) = A^c = B_0(A', q) \cup E$. Furthermore, $C_0(A', q) = A'$ and $B_0(A', q) = (A')^c$. Consider $F_1(A, q)$ and $C_1(A, q) = C_0(A, q) \cup F_1(A, q)$. Also consider $F_1(A', q)$ and $C_1(A', q) = C_0(A', q) \cup F_1(A', q)$. Notice that $F_1(A, q) \subseteq F_1(A', q) \cup E$ and therefore, $C_1(A, q) = C_0(A, q) \cup F_1(A, q) \subseteq C_0(A, q) \cup E \cup F_1(A', q) = C_1(A', q)$. Consequently, $C_{t+1}(S, q) = C_1(A, q) \subseteq C_1(A', q) = C_{t+1}(S', q)$.

Statement (4) is proved as follows. Suppose $q' \leq q$. We use induction on t . For $t = 0$, $C_0(S, q) = S = C_0(S, q')$. Suppose $C_t(S, q) \subseteq C_t(S, q')$. Let $A = C_t(S, q) \subseteq C_t(S, q')$. Let $A = C_t(S, q)$ and $A' = C_t(S, q')$. Then $A \subseteq A'$. Consider (A, q) and (A, q') . Then $F_1(A, q) \subseteq F_1(A, q') \implies C_1(A, q) \subseteq C_1(A, q')$. Therefore, $C_{t+1}(S, q) = C_1(A, q) \subseteq C_1(A, q') \subseteq C_1(A', q') = C_{t+1}(S, q')$, where the second inclusion follows from statement (3).

Statement (5) is proved as follows. If E is a Nash equilibrium at q that contains S , then $S \subseteq E$ implies $C(S, q) \subseteq C(E, q) = E$, where the inclusion follows from statement (3) and equality from statement (1).

Statement (6) is proved as follows. Note that $S \subseteq A \implies C(S, q') \subseteq C(A, q')$ by statement (3). Let $A' = C(S, q')$. Statement (4) implies that $A = C(S, q) \subseteq C(S, q') = A'$. Applying statement (3) we have that $C(A, q') \subseteq C(A', q') = C(S, q')$, where the last equality follows from statement (2). \square

For a fixed q , Algorithm 1 may be viewed as a mapping $S \mapsto C(S, q)$, from an initial set S of infected players to a (weakly) larger set $C(S, q)$ of all players who are infected by players in S using both person-to-person spread and virality impact. Fixed points of Algorithm 1 are Nash equilibria in the network coordination game with virality. Indeed, for a fixed q , statement (1) shows that *the set of all Nash equilibria in which at least one player has an incentive to play 1 is equal to the set of fixed points of Algorithm 1*. We may define $C(\emptyset, q) = \emptyset$ in Algorithm 1 to signify the case that when E is the empty set, everyone plays 0, and as that is the initial Nash equilibrium in the game, there is trivially no spread of action 1. This strengthens statement (1) to apply to every subset E , and therefore, *the set of Nash equilibria at q is equal to the set of*

fixed points of Algorithm 1.

Statement (2) shows that regardless of the starting configuration of players with an incentive to play 1, Algorithm 1 expands this set to always deliver a Nash equilibrium. Statement (5) makes statement (2) stronger by proving that Algorithm 1 delivers the smallest Nash equilibrium that contains a given starting configuration of players with an incentive to play 1. We use this to define (equilibrium) **depth of contagion from (S, q) as a proportion of the network** by the fraction $\delta(S, q) = \frac{|C(S, q)|}{|\mathcal{I}|}$.

Statements (3) and (4) highlight comparative statics of equilibrium outcomes generated by Algorithm 1. 1. If starting set increases, the number of players who play 1 in the resulting Nash equilibrium goes up (weakly). If network resilience parameter q goes down, the number of players who end up playing 1 in the resulting Nash equilibrium goes up (weakly). These two statements generalize the corresponding statements in [Morris, 2000] to include virality. An expansion of the Nash equilibrium set increases depth of contagion as well.

Statement (6) presents an important feature of Algorithm 1. If we start at (S, q) and end at $C(S, q) \neq \mathcal{I}$, then statement (4) says that we can reduce q to q' and end up with a (weakly) larger contagion set $C(S, q')$. Statement (6) says that instead of running the entire analysis again by starting at (S, q') , we can start the algorithm at $C(S, q)$ and q' and end up with the same result as starting at (S, q') . This bootstrapping from the end of one run of the algorithm to a continuation of the next to reach the same final outcome is an important step in providing a greatly improved algorithm to compute the contagion threshold using Algorithm 2. It is useful to compute the equilibrium depth of contagion for every q as well.

Properties of Algorithm 1 in Theorem 2 can be used to prove new characterizations of earlier results for network coordination games without virality.

Corollary 3. *Let Γ be a network coordination game (without virality) and $q \in [0, 1]$.*

1. *A nonempty set $E \subset \mathcal{I}$ is q -cohesive and $\mathcal{I} \setminus E$ is $(1 - q)$ -cohesive, if, and only if, E is a fixed point of Algorithm 1.*
2. *For every nonempty set S of players with incentive to play 1, $C(S, q)$ is the smallest set E containing S such that E is q -cohesive and $\mathcal{I} \setminus E$ is $(1 - q)$ -cohesive.*

Proof. Statement (1) follows from statement (1) of Theorem 1 and statement (1) of Theorem 2. Statement (2) follows from statement (1) of Theorem 1 and statement (5) of Theorem 2. \square

Statement (2) in Corollary 3 shows that Algorithm 1 can be used to find the smallest q -cohesive set with a $(1 - q)$ -cohesive complement starting from an arbitrary starting set S of players with an incentive to play 1, in any network coordination game without virality. When $C(S, q) \neq \mathcal{I}$, Algorithm 1 provides a simple way to generate an equilibrium in which a subset of the network plays 1 and the remainder plays 0, or as termed in [Morris, 2000], an equilibrium with coexistence of conventions.

Algorithm 1 shows that if we start from an arbitrary set S of players playing 1 and an arbitrary threshold q , then action 1 proliferates to the set $C(S, q)$ of players in \mathcal{I} . It is also of interest to know when action 1 proliferates to the entire network. Algorithm 2 answers this question by computing q^* such that $C(S, q^*) = \mathcal{I}$.

Algorithm 2 (Full network contagion starting from S)

1. In a network coordination game with virality, let S be a nonempty set of players with an incentive to play 1. Define $A_0 = S$ and $q_0 = 1$ (and therefore, $b_0 = 0$).
2. For $n \geq 0$, define (A_{n+1}, q_{n+1}) recursively as follows:
 - (a) Apply Algorithm 1 with (A_n, q_n) to determine $A_{n+1} = C(A_n, q_n)$.
 - (b) If $A_{n+1} \neq \mathcal{I}$, let $q_{n+1} = \max_{i \notin C(A_n, q_n)} \frac{cs_i}{cd_i - \phi_i(p_i)}$. Let $b_{n+1} = \frac{c(1-q_{n+1})}{q_{n+1}}$.
3. Let N be the first n such that $A_{n+1} = \mathcal{I}$. Let $q^* = q_N$.

Algorithm 2 starts with an initial, nonempty set S of players who have an incentive to play 1 and recursively searches for the next highest q that allows the contagion set resulting from Algorithm 1 to keep expanding. To motivate statement 2(b), notice that $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)} = \frac{s_i}{d_i} + \frac{\phi_i(p_i)}{cd_i}q$, and therefore, the threshold to switch $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)} = q$ can be written as $\frac{cs_i}{cd_i - \phi_i(p_i)} = q$. Moreover, $i \notin C(A_n, q_n)$ implies $\phi_i(p_i) < cd_i$ so that the denominator is positive.

The algorithm starts at the top, with $q_0 = 1$, and applies Algorithm 1 to determine $C(S, 1)$. If contagion has not spread to the entire network with $C(S, 1)$, the algorithm looks at the threshold computations of all the players who play 0 (using $C(S, 1)$ as the set of players who play 1) and selects one with the highest threshold and sets q_1 to be this threshold. This q_1 is the threshold closest to 1 that is *just enough* to give this “marginal” player an incentive to play 1. Algorithm 1 is run again to see the full ramification of lowering the threshold to q_1 . If contagion has not yet spread to the entire network, we follow the same process to lower q just enough to switch the next marginal player and apply Algorithm 1 again. Algorithm 2 stops at the first time when contagion spreads to the entire network. As the number of players is finite, the algorithm stops in finitely many iterations (there are never more than the number of players in $\mathcal{I} \setminus S$, and there may be considerably fewer depending on spread of contagion in Algorithm 1 at each step). The resulting q^* is the contagion threshold for the network with starting set S . Pseudocode for Algorithm 2 (with nested Algorithm 1) is presented in Algorithm 3. Properties of Algorithm 2 are as follows.

Theorem 4 (Properties of Algorithm 2). *Let Γ be a network coordination game with virality and S be as in Algorithm 2.*

1. $(A_n)_{n=1}^{N+1}$ is a strictly increasing sequence of Nash equilibria, starting at $C(S, 1)$ and ending at \mathcal{I} .
2. $(q_n)_{n=0}^N$ is a strictly decreasing sequence in $[0, 1]$, starting at 1 and ending at q^* .

Algorithm 3 Pseudocode for depth of contagion and full contagion (Algorithms 1 and 2 combined)

```

1: Generate network  $\mathcal{I}$ 
2: Define starting set  $A$ 
3: Create dictionary Depth
4: function PROP( $i, S$ )
5:    $H \leftarrow \mathcal{I} \setminus (N_i \cup \{i\})$ 
6:    $p \leftarrow 0$ 
7:    $n \leftarrow |H|$ 
8:   for  $j \in H$  do:
9:     if  $j \in S$  then:
10:       $p \leftarrow p + 1$ 
11:     end if
12:   end for
13:   return  $\frac{p}{n}$ 
14: end function
15:  $q \leftarrow 1$ 
16:  $b \leftarrow 0$ 
17:  $c \leftarrow 1$ 
18: while  $|A| < |\mathcal{I}|$  do:
19:    $F \leftarrow \emptyset$ 
20:    $B \leftarrow \mathcal{I} \setminus A$ 
21:    $B' \leftarrow \emptyset$ 
22:   while  $B \neq B'$  do:
23:     for  $i$  in  $B$  do:
24:        $p_i \leftarrow \text{PROP}(i, A)$ 
25:       if  $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)} \geq q$  then
26:          $F \leftarrow F \cup \{i\}$ 
27:       end if
28:     end for
29:      $A \leftarrow A \cup F$ 
30:      $B' \leftarrow B$ 
31:      $B \leftarrow \mathcal{I} \setminus A$ 
32:   end while
33:   if  $F = \emptyset$  then
34:     if  $F = \mathcal{I}$  then
35:       Depth( $q$ )  $\leftarrow 1$ 
36:       return  $q$ 
37:     else
38:       Depth( $q$ )  $\leftarrow \frac{|S|}{|\mathcal{I}|}$ 
39:       for  $i \in B$  do:
40:          $p_i \leftarrow \text{PROP}(i, A)$ 
41:       end for
42:        $q \leftarrow \max_{i \in B} \frac{cs_i}{cd_i - \phi_i(p_i)}$ 
43:        $b \leftarrow \frac{c(1-q)}{q}$ 
44:     end if
45:   end if
46: end while

```

3. For every $n \in \{0, \dots, N\}$, $A_{n+1} = C(S, q_n)$, and therefore, $C(S, q^*) = \mathcal{I}$.

4. For every $q \in [0, 1]$, $C(S, q) = \mathcal{I}$, if, and only if, $q \leq q^*$.

Proof. For statement (1), $(A_n)_{n=1}^{N+1}$ is a sequence of Nash equilibria follows from statement (2) of Theorem 2. It is increasing follows from the fact that each term in the sequence is the output of Algorithm 1 applied to the immediately preceding term in the sequence. To see that it is strictly increasing, fix $n \in \{1, \dots, N\}$ and notice that $A_n \neq \mathcal{I}$. Let $i_0 \notin A_n$ be such that $q_n = \frac{cs_{i_0}}{cd_{i_0} - \phi_{i_0}(p_{i_0})} = \max_{i \notin A_n} \frac{cs_i}{cd_i - \phi_i(p_i)}$. Notice that $i \notin A_n$ implies $\phi_i(p_i) < cd_i$ so that the denominator is positive. Rearranging terms yields $\frac{s_{i_0}}{d_{i_0}} + \frac{\phi_{i_0}(p_{i_0})}{d_{i_0}(b_n + c)} = q_n$, and therefore, $i_0 \in A_{n+1}$.

To prove statement (2), notice that for each $n \in \{0, \dots, N-1\}$, $C(A_n, q_n) = A_{n+1} \neq \mathcal{I}$. Therefore, for every $i \notin C(A_n, q_n)$, it must be that $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b_n + c)} < q_n$, or equivalently, $\frac{cs_i}{cd_i - \phi_i(p_i)} < q_n$, whence $q_{n+1} = \max_{i \notin C(A_n, q_n)} \frac{cs_i}{cd_i - \phi_i(p_i)} < q_n$.

Statement (3) is proved using induction on n . We know that $A_1 = C(S, q_0)$. Suppose $A_n = C(S, q_{n-1})$. Using statement (6) of Theorem 2, it follows that $A_{n+1} = C(A_n, q_n) = C(S, q_n)$.

To prove statement (4), in one direction, suppose $q \leq q^*$. Then $\mathcal{I} = C(S, q^*) \subseteq C(S, q) \subseteq \mathcal{I}$, where the equality follows from statement (3) and the first inclusion follows from statement (4) of Theorem 2. In the other direction, to show the contrapositive, suppose $q^* < q$. Then $\exists n \in \{1, \dots, N\}$ such that $q \in (q_n, q_{n-1}]$. In this case, $q \leq q_{n-1}$ implies $C(S, q_{n-1}) \subseteq C(S, q)$, using statement (4) of Theorem 2. Suppose $i_0 \in C(S, q)$ and $i_0 \notin C(S, q_{n-1})$. Then $q \leq \frac{cs_{i_0}}{cd_{i_0} - \phi_{i_0}(p_{i_0})} \leq \max_{i \notin C(S, q_{n-1})} \frac{cs_i}{cd_i - \phi_i(p_i)} = q_n$, where the first inequality follows from $i_0 \in C(S, q)$, and the second inequality follows from $i_0 \notin C(S, q_{n-1})$. This contradicts $q_n < q$. Therefore, $C(S, q) = C(S, q_{n-1})$. As $C(S, q_{n-1}) \neq \mathcal{I}$, it follows that $C(S, q) \neq \mathcal{I}$. \square

Statements (1) and (2) of Theorem 4 show that Algorithm 2 outputs a strictly increasing sequence of Nash equilibria that start from $C(S, 1)$ and expand to the entire network. At each step, the expansion at the end of that step is the smallest Nash equilibrium extension of the beginning set for that step. Moreover, at the end of each step, Algorithm 2 determines the smallest decrease in q that is needed to continue further proliferation of action 1 in the network. In this sense, for an arbitrary starting set S , Algorithm 2 provides an expanding equilibrium-by-equilibrium proliferation of action 1 to the entire network, using a minimal decrease in q to expand contagion from one equilibrium to the next.

Statement (3) equates the equilibrium-by-equilibrium expansion up to step $n+1$ with the one step contagion from initial S using threshold q_n , using the bootstrapping feature of Algorithm 1 listed in statement (6) of Theorem 2. This immediately yields one direction of statement (4), that is, for every threshold $q \leq q^*$, contagion occurs from S to \mathcal{I} at q . The design of the algorithm shows the other direction in statement (4).

Statement (4) proves that a special characteristic of the q^* computed by Algorithm 2 is that it is the largest q for which $C(S, q) = \mathcal{I}$. Following [Morris, 2000], we term this the contagion threshold. Let Γ be a network coordination game with virality and consider a configuration S of players with incentive to play 1. The **contagion threshold for S** is $q^* = \sup\{q \in [0, 1] \mid C(S, q) = \mathcal{I}\}$. The definition applies to

network coordination games with virality. As 0 is trivially in the set over which the supremum is taken, the supremum is well-defined. Algorithm 2 shows that the supremum is in the set itself. These are formalized in the following corollary.

Corollary 5. *Let Γ be a network coordination game with virality and S a nonempty set of players with incentive to play 1.*

1. *The value q^* from Algorithm 2 is the contagion threshold for S .*
2. *The set of all q for which $C(S, q) = \mathcal{I}$ is the interval $[0, q^*]$.*

Proof. Both statements follow from statement (4) of Theorem 4. □

The interval $[0, q^*]$ gives a full characterization of when contagion occurs from S to the entire network. In other words, for a given starting set S , the most resilient network that can be fully infected is the one with relative cost of miscoordination given by q^* . Networks with resilience less than or equal to q^* will be fully infected while those with higher resilience will not be fully infected.

In addition to determining the threshold q^* guaranteeing contagion to the entire network, Algorithm 2 can be used to compute how far contagion progresses if it does not proliferate to the entire network, that is, for every $q > q^*$. This is helpful to understand proliferation of contagion to a fraction of the network. As mentioned in the introduction, a misinformation campaign or health epidemic may wreak havoc with proliferation of an action to a fraction of the network. For example, misinformation related to the COVID-19 pandemic did not convince the entire US population yet still caused significant societal disruption. Misinformation regarding vaccines has instilled doubt about the safety of vaccines and led to a decrease in vaccination rates. It is therefore important to understand not only full contagion, but also partial contagion to a fraction of the network.

Algorithm 2 yields a decreasing sequence of thresholds q_n and a corresponding increasing sequence of contagion sets $C(S, q_n)$ that shows how far action 1 can spread at q_n . Combined with the fact that for each $q \in (q_n, q_{n-1}]$, $C(S, q) = C(S, q_{n-1})$ (as shown in proof of statement (4) of Theorem 4), we define the **contagion depth function from S** as follows:

$$\delta(S, q) = \begin{cases} 1 & \text{if } q \leq q^*, \text{ and} \\ \frac{|C(S, q_n)|}{|\mathcal{I}|} & \text{if } q_{n+1} < q \leq q_n, \text{ for } n \in \{0, \dots, N-1\}. \end{cases}$$

For a fixed starting set S , δ measures the proportion of the network that is infected as a function of network resilience parameter $q \in [0, 1]$. For a fixed S , it is a step function that starts at 1 for $q \leq q^*$ (low network resilience leads to full contagion) and strictly decreases in a step-wise manner to $|C(S, q_n)|/|\mathcal{I}|$ over each interval $(q_{n+1}, q_n]$ as network resilience increases. The contagion depth function is an equilibrium notion.

For each S and q , it provides the fraction of the network infected by the smallest Nash equilibrium that contains S .

The contagion depth function allows us to analyze scenarios in which an action may spread to a significant proportion of a network but not the entire network. These scenarios are important in light of the aforementioned cases of partial contagion causing significant societal disruption. As it is important to understand both full contagion and partial contagion to some proportion of a network, we will show in the next section how to invert this relation to understand what starting set size is needed to reach a given depth of contagion for a given q .

Special cases of properties of Algorithm 2 in Theorem 4 can be used to prove new characterizations of earlier results for network coordination games without virality.

Corollary 6. *Let Γ be a network coordination game (without virality), S a nonempty set of players with incentive to play 1, and q^* the contagion threshold from Algorithm 2.*

For every $q \in [0, 1]$, the complement of S is uniformly no more than $(1 - q)$ -cohesive, if, and only if, $q \leq q^$.*

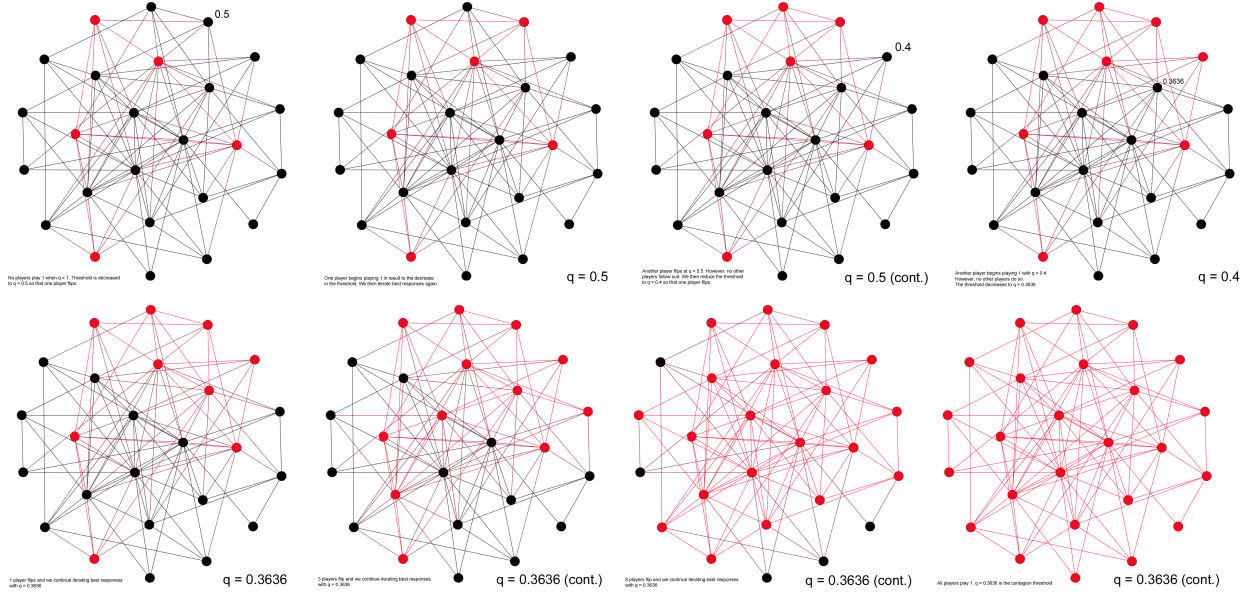
Proof. Follows immediately from statement (2) of Theorem 1 and statement (4) of Theorem 4. \square

For network coordination games without virality, this corollary shows how Algorithm 2 helps to compute that the complement of S is uniformly no more than $(1 - q)$ -cohesive even for significantly large networks which were not amenable to such analysis earlier.

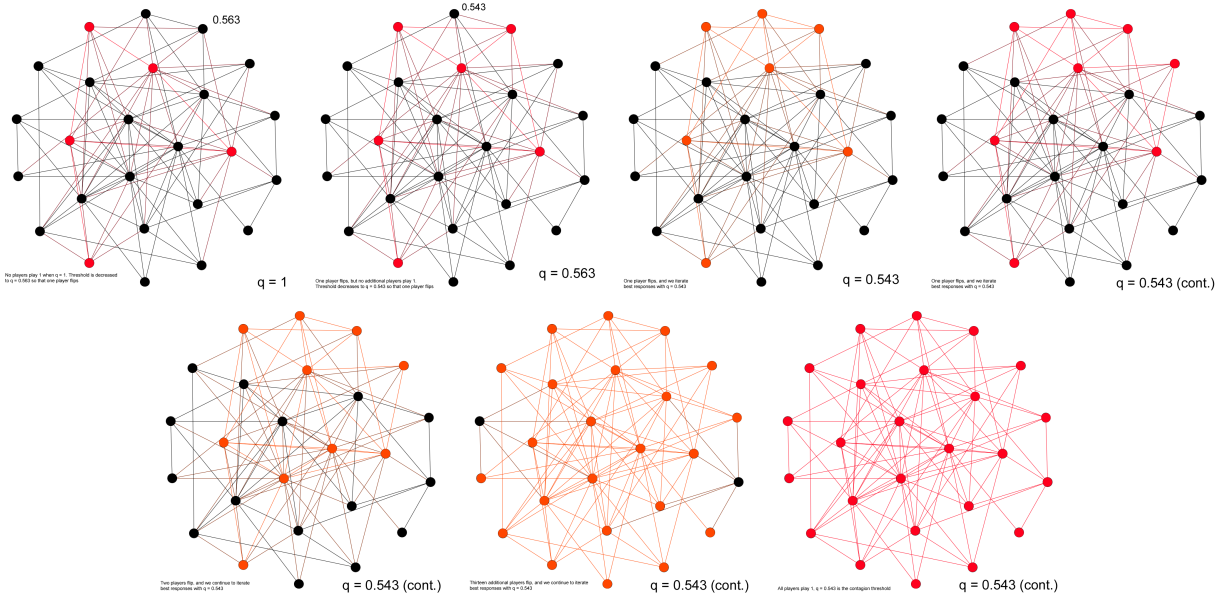
As an explicit illustration of algorithms 1 and 2, we use the Barabási-Albert procedure to generate a scale-free network of 25 players with attachment parameter $m = 5$. The generated network is presented in Figure 1 (each panel shows the same network). Virality impact is modeled by $\phi_i(p_i) = \alpha c d_i p_i$ where cost of miscoordination is normalized to $c = 1$. When $\alpha = 0$ there is no impact of virality, when $\alpha = 1$ there is a virality impact, and intermediate values of α correspond to differential impact of virality on payoffs. For convenience, we focus on the cases when $\alpha = 0$ and $\alpha = 1$.

Suppose there are 5 players who have exogenously decided to play 1 or who have been exogenously infected with 1. (This can be formulated by exogenously setting their payoff from playing 1 to 1, regardless of the actions of their neighbors.) These comprise the initial set S , identified by the red nodes in the first panel of the first row in both Figure 1a and Figure 1b.

First consider the case of no virality, $\alpha = 0$, presented in Figure 1a. The red nodes in the first panel in the first row are the initial set of 5 players with an incentive to play 1. We start Algorithm 2 with $q_0 = 1$ and notice that no other player becomes infected at step 1 ($A_1 = A_0 = S$). Checking incentives for the remaining players to play 1 shows the next step down is $q_1 = 0.5$ (labeled next to the player who will switch at this threshold). Running Algorithm 1 with (A_1, q_1) shows that only 1 player switches at step 1 of Algorithm 1 (shown in the second panel in top row of Figure 1a), and 1 more player flips at step two of Algorithm 1



(a) Contagion threshold without virality is $q^* = 0.3636$ ($I = 25$, $m = 5$, $\alpha = 0$)



(b) Contagion threshold with virality is $q^* = 0.543$ ($I = 25$, $m = 5$, $\alpha = 1$)

Figure 1: Illustration of network contagion algorithm

(third panel in top row), and then Algorithm 1 stops. This set of 7 players is how far contagion will spread from S when $q = 0.5$. It is the smallest Nash equilibrium of players playing 1 if we start from (S, q_1) .

Algorithm 2 denotes this set A_2 , checks the remaining players, and computes the next level for q to be $q_2 = 0.4$ (labeled next to the player who will switch at this threshold). Running Algorithm 1 with (A_2, q_2) flips 1 player only and stops after one iteration (last panel in the first row). The resulting set of 8 players

is A_3 , the next larger Nash equilibrium of players playing 1. The next threshold is $q_3 = 0.3636$. Using (A_3, q_3) as input to Algorithm 1 shows that 1 player flips in iteration one (first panel of row 2 in Figure 1a), 3 more players flip in iteration two (second panel), 8 more players flip in iteration three (third panel), and the remaining 5 players flip in iteration four (last panel). The contagion threshold for this network and starting set S is $q^* = 0.3636$. Algorithm 2 also yields depth of contagion for different values of q . Depth (in terms of proportion of the network infected) is 1 when $q \leq 0.3636$, is $\frac{7}{25}$ when $q \in (0.3636, 0.4]$, is $\frac{6}{25}$ when $q \in (0.4, 0.5]$, and is $\frac{5}{25}$ when $q \in (0.5, 1]$.

The entire process uses a total of 11 iterations (3 for Algorithm 2 and 8 for Algorithm 1) on shrinking sets of players not playing 1. This is negligible as compared to the number of subsets of players not in S , which is $2^{20} = 1,048,576$.

Now consider the case with virality, $\alpha = 1$, presented in Figure 1b. We start with the same initial set of 5 players in S (first panel in first row of Figure 1b) and $q_0 = 1$ and find that no other player gets infected. The next level is $q_1 = 0.563$ at which the next marginal player will flip. (Notice that this is the same one as in the earlier case, but the threshold at which this player flips is higher due to the virality impact which contributes 0.063 to the earlier threshold of 0.5 without virality.) Only this player flips and the next threshold is $q_2 = 0.543$. In this case, 1 player flips in the first iteration of Algorithm 1, 1 more in the second iteration, 2 in the third, thirteen in the fourth, and the remaining two players in the fifth iteration. The contagion threshold with virality for starting set S is $q^* = 0.543$. Depth (in terms of proportion of the network infected) is 1 when $q \leq 0.543$, is $\frac{6}{25}$ when $q \in (0.543, 0.563]$, and is $\frac{5}{25}$ when $q \in (0.563, 1]$. The entire process uses a total of 8 iterations (2 for Algorithm 2 and 6 for Algorithm 1).

The contagion threshold with virality is considerably higher than the one without virality. In other words, for every $q \in (0.3636, 0.543]$ there is full contagion with virality but not without it. More generally, for every $q \in [0, 1]$, depth of contagion at q is (weakly) higher with virality than without.

5 Monte Carlo Simulation

Algorithm 3 provides a computational tool to study various properties of contagion in networks. We use it in Monte Carlo simulation of scale-free networks following the Barabási-Albert algorithm with different attachment parameters. Scale-free networks are used widely to study social networks and they naturally capture the idea of preferential attachment. Our algorithm delivers results for a large number of combinations of network configurations, attachment parameters, relative miscoordination costs q , starting set sizes S , and virality. The data generating process is as follows.

1. Fix an attachment parameter $m \in \{5, 10, 20\}$ and randomly generate 40 scale-free networks of size 1,000 with attachment parameter m using Barabási-Albert algorithm in NetworkX Python.

2. Fix a generated network and a virality parameter $\alpha \in \{0, 0.5, 1\}$. For each starting set size $n \in \{10\lambda \mid \lambda = 1, \dots, 99\}$, randomly select 50 sets S of nodes in the generated network, each set of size $|S| = n$. For each S , use algorithm to determine full contagion threshold and depth of contagion starting from S .
3. Repeat the process for each virality parameter α and each attachment parameter m .

For each m and α , this process generates 198,000 ($40 \times 50 \times 99$) complete runs of Algorithm 3. Varying $m \in \{5, 10, 20\}$ and $\alpha \in \{0, 0.5, 1\}$ yields 1,782,000 complete runs of Algorithm 3 (9 scenarios, each with 198,000 runs). We report results based on these simulations.

The data generating process simulates networks with 1,000 players. The parameter values facilitate analysis for three levels, approximately corresponding to low, medium, and high parameter values. For the attachment parameter m , the value $m = 5$ captures the idea of low attachment (5 connections for each new player in a network with 1,000 players), $m = 10$ is medium attachment (10 connections for each new player) and $m = 20$ is high attachment (20 connections for each new player). Similarly, $q = 0.25$ is low (relative) cost of miscoordination, $q = 0.5$ is medium (relative) cost of miscoordination, and $q = 0.75$ is high (relative) cost of miscoordination. For virality, we use the functional form presented in the model section, $\phi(p_i) = \alpha c d_i p_i$. The value $\alpha = 0$ corresponds to the model without virality, $\alpha = 1$ presents the model with virality, and $\alpha = 0.5$ is an intermediate level of impact of virality. A large number of robustness checks conducted with additional variation in the values of m , q , and α yield similar results.

5.1 Contagion Thresholds

Figure 2 shows the distribution of contagion thresholds for each of the nine scenarios. In each panel, starting set size (as proportion of the network) is on the x -axis and network contagion threshold on the y -axis. Each dot is one run of the algorithm to determine contagion threshold q^* for a given starting set size. Each panel includes 198,000 outcomes. The average contagion threshold for the 2,000 outcomes for each starting set size is overlaid as a bold line. A summary of selected data is included in Table 1.

Several features of contagion on networks emerge in Figure 2. In all 9 cases, the average contagion threshold is increasing in starting set size, as expected. In each case, for a given starting set size, the variation in the 2,000 thresholds is due to different network configurations and different nodes in the starting set. In each case, the contagion thresholds are distributed tightly around the average, for the most part.

Increasing virality increases the contagion threshold, making the network more susceptible to contagion. Moreover, the same increase in starting set size causes a larger increase in the contagion threshold for higher virality levels, as shown in the higher slope of the average contagion threshold line. Both can be seen moving downward in a column through the panels.

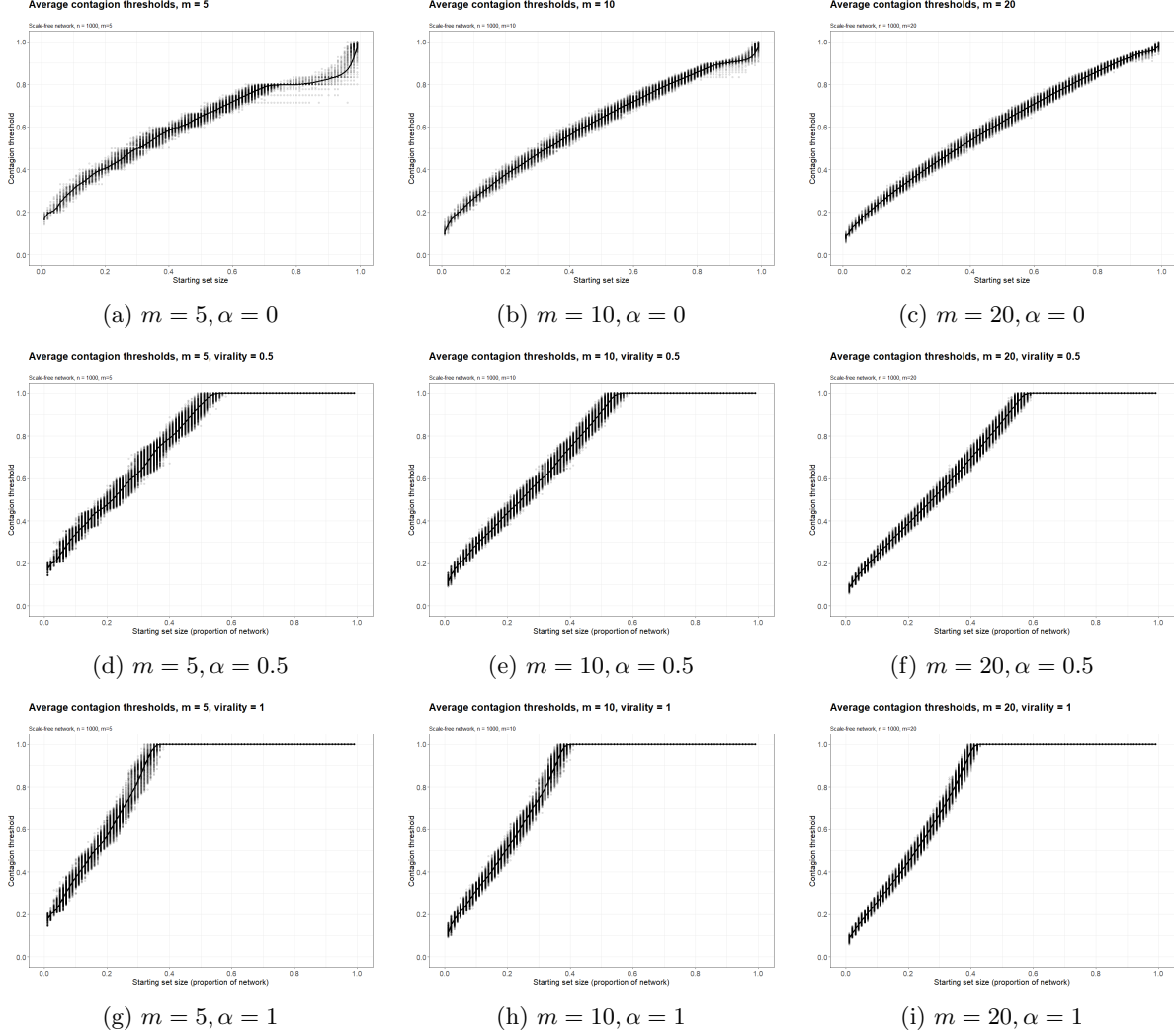


Figure 2: Contagion thresholds for scale-free networks

This shows a complementarity in the effect of starting set size and virality on contagion thresholds. A large starting set implies that on average, a higher proportion of a given player's neighbors are playing 1 and also that more players are affected by the higher level of aggregate action. This, in turn, increases the size of the set of players playing 1, which makes the impact of virality stronger leading to contagion in networks with higher resilience (higher q). For example, in a low connectivity network ($m = 5$), without virality ($\alpha = 0$), when starting set size increases from 5 percent (50 players) to 10 percent, the contagion threshold increases by 36.3 percent (from 0.226 to 0.308), but with virality ($\alpha = 1$), the comparable increase is 48.6 percent (from 0.253 to 0.376). An increase of starting set size from 5 percent to 20 percent leads to a 79.2 percent increase in the contagion threshold (from 0.226 to 0.405) without virality, but a 125.7 percent (from 0.253 to 0.571) with virality, on average.

This complementarity between virality and starting set size captures a snowball effect of contagion re-

Size	$m = 5$			$m = 10$			$m = 20$		
	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$
0.01	0.171	0.174	0.177	0.109	0.110	0.111	0.085	0.086	0.087
0.02	0.194	0.198	0.201	0.143	0.146	0.149	0.105	0.107	0.108
0.03	0.201	0.206	0.211	0.166	0.170	0.175	0.126	0.129	0.132
0.04	0.209	0.217	0.227	0.182	0.189	0.196	0.142	0.147	0.151
0.05	0.226	0.239	0.253	0.195	0.203	0.212	0.157	0.162	0.168
0.06	0.246	0.262	0.279	0.210	0.217	0.228	0.173	0.180	0.188
0.07	0.263	0.284	0.306	0.222	0.235	0.249	0.187	0.196	0.206
0.08	0.280	0.304	0.331	0.238	0.254	0.272	0.199	0.211	0.223
0.09	0.294	0.322	0.354	0.253	0.271	0.293	0.212	0.226	0.241
0.1	0.308	0.340	0.376	0.265	0.288	0.313	0.225	0.242	0.261
0.2	0.405	0.477	0.571	0.377	0.437	0.512	0.340	0.386	0.449
0.3	0.498	0.627	0.831	0.474	0.587	0.746	0.442	0.538	0.676
0.4	0.584	0.790	1	0.562	0.748	1	0.536	0.698	0.966
0.5	0.650	0.946	1	0.643	0.916	1	0.624	0.870	1
0.6	0.720	1	1	0.720	1	1	0.709	1	1
0.7	0.788	1	1	0.794	1	1	0.788	1	1
0.8	0.800	1	1	0.858	1	1	0.862	1	1
0.9	0.825	1	1	0.903	1	1	0.928	1	1

Table 1: Average thresholds for full network contagion for different starting set sizes (proportion of network)

sulting from increasing returns. A marginal increase in the starting set size causes a more than proportional increase in the contagion threshold in the presence of virality.

Indeed, virality can make the snowball effect of contagion irreversible at some point. In low or moderate connectivity networks ($m = 5, 10$) with virality ($\alpha = 1$), the contagion threshold is 1 with a starting set size of 40 percent. In other words, in low or moderate connectivity networks with virality, if about 40 percent of the network is infected (has an incentive to play 1), there is no hope of stopping contagion to the entire network, regardless of how much a player dislikes miscoordination with neighbors who play 0 (that is, even with unbounded miscoordination cost). But if we can stop virality ($\alpha = 0$), contagion from 40 percent of the network will not spread to the whole network whenever relative miscoordination cost is above 0.584, on average.

Increasing connectivity makes the average contagion threshold curve less steep implying that for a given starting set size, increasing connectivity lowers the contagion threshold, making it harder to propagate an action across the network. This can be seen by moving across in a row through the panels. For a fixed q , higher connectivity implies a higher degree, which lowers the incentive for a player to play 1, as shown by the threshold calculation $\frac{s_i}{d_i} + \frac{\phi_i(p_i)}{d_i(b+c)} \geq q$. For example, suppose the relative cost of miscoordination is $q = 0.25$ and consider a network with virality ($\alpha = 1$). For a starting set size of 5 percent, full contagion occurs with low attachment parameter ($m = 5$) but does not occur with moderate or high attachment

parameter ($m = 10, 20$). In other words, contagion occurs for a larger range of q in less connected networks, and therefore, it is easier to propagate an action in less connected networks. Intuitively, in a less connected network, flipping one neighbor is more likely to flip a given player.

5.2 Depth of contagion

As mentioned earlier, there are several cases where it is more fruitful to understand proliferation of contagion to a fraction of the network. For each q , α , and starting set S , we use the contagion depth function to compute the proportion of the network to which action 1 spreads starting from S . This is reported in Figure 3 (for scale-free networks with attachment parameter $m = 5$), in Figure 4 (for $m = 10$), and in Figure 5 (for $m = 20$). In each panel of each figure, the starting set size (as proportion of the network) is on the horizontal axis and depth of contagion is on the vertical axis. In each panel, for each starting set size, there are 2,000 computations of depth of contagion (40 networks of type m and 50 starting sets of the given size). Each dot represents one such computation. As the computation of depth of contagion uses equilibrium sets of the form $C(S, q)$, each dot is a Nash equilibrium in the corresponding network for starting set S and q . The average of these depths is overlaid as a bold line across different starting set sizes.

In each panel, an increase in starting set size increases depth of contagion (both due to greater person-to-person spread and due to greater virality impact). From Theorem 2, we know that $S \subseteq S' \implies C(S, q) \subseteq C(S', q)$, and therefore, the depth of contagion is a (weakly) increasing function of starting set size. Moreover, as the set to which contagion spreads from a given starting set contains the starting set ($S \subseteq C(S, q)$), every observation lies (weakly) above the diagonal. The difference between average depth of contagion and the diagonal can be viewed as a measure of how far contagion spreads from a given starting set. Points on the diagonal show that there is no further contagion beyond the corresponding starting set size ($S = C(S, q)$).

In the absence of virality, it is harder to spread contagion at higher q . This can be seen by going across any row in each figure. For example, consider Figure 3 ($m = 5$) and consider the first row ($\alpha = 0$). When q is low ($q = 0.25$), depth of contagion rises steeply in a narrow range of small starting set sizes (about 5-10 percent of the network) and spreads quickly to the entire network beyond that. When q increases to 0.5, contagion spreads more slowly for smaller starting set sizes, speeds up if about 25-30 percent of the network has an incentive to play 1, and spreads quickly to the entire network shortly thereafter. When q is high ($q = 0.75$), it is hard to spread contagion beyond the starting set till about 35 percent of the network, it spreads more in the range of 35-65 percent of the network before spreading to the entire network.

With virality, contagion spreads deeper and using smaller starting set sizes. This can be seen by going down in a column in each figure. For example, consider Figure 3 ($m = 5$) and consider the third column ($q = 0.75$). Without virality ($\alpha = 0$), starting set sizes need to go beyond about 35 percent of the network for contagion to start spreading and it takes about 68 percent of the network to be affected before contagion

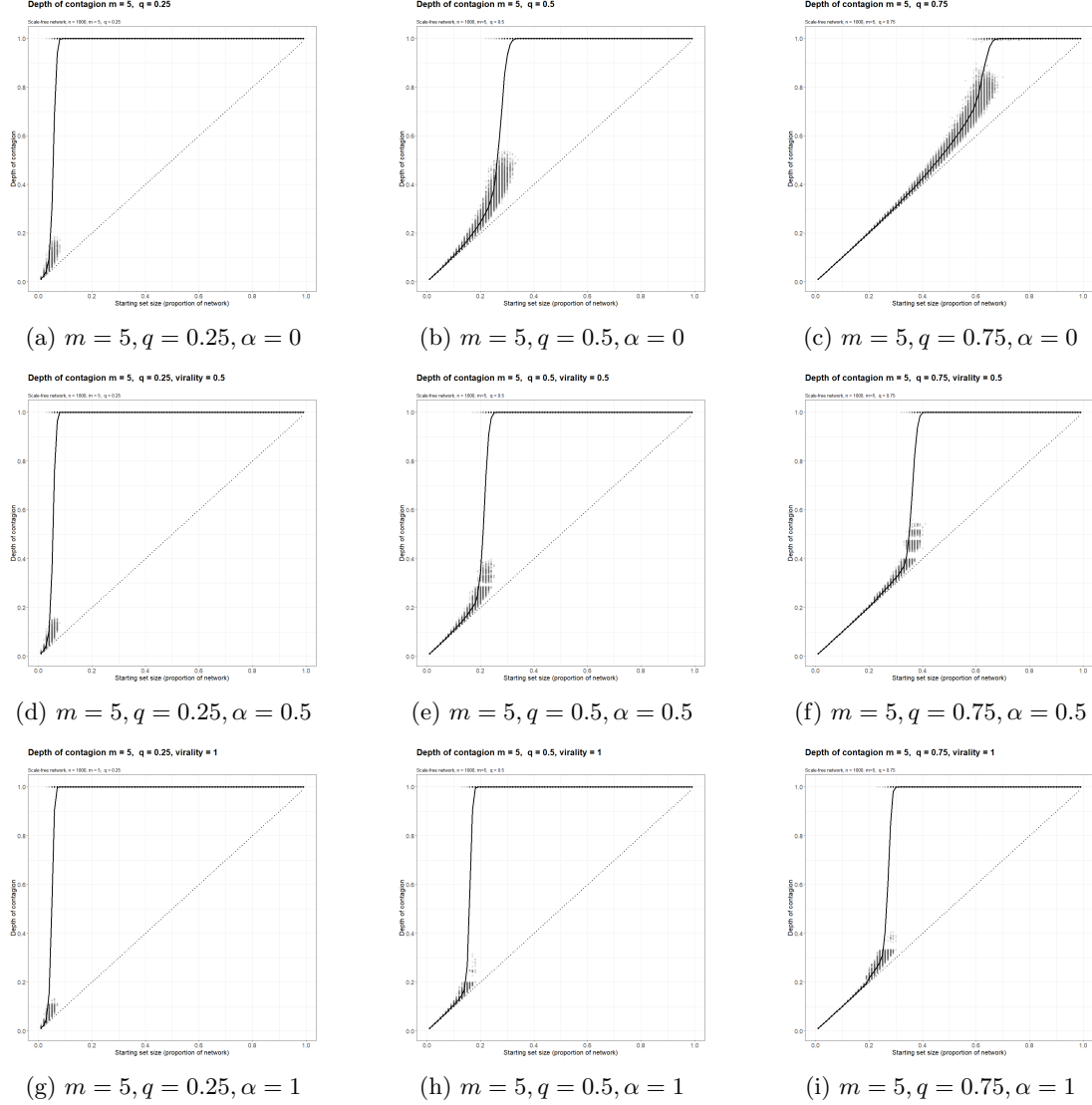


Figure 3: Depth of contagion for scale-free networks with $m = 5$

spreads to the entire network. With $\alpha = 0.5$, contagion starts to spread earlier, with starting set size around 25 percent, and spreads to the entire network once about 40 percent of the network is affected. With $\alpha = 1$, contagion starts to spread even earlier, at around 20 percent of the network and spreads to the entire network once about 30 percent of the network is infected. Without virality, q would have to be below 0.5 for a 30 percent starting set size to cause full contagion. In other words, virality (or susceptibility to global behavior) can rollback the effect of a higher q (higher resilience to local behavior).

The impact of virality is higher when q is higher. When q is low (consider the first column, for $q = 0.25$), contagion spreads easily without virality ($\alpha = 0$) and adding virality ($\alpha = 0.5, 1$) contributes less to further increase the spread of contagion. More generally, for each q and S , increasing virality increases depth of contagion and steepens the curve.

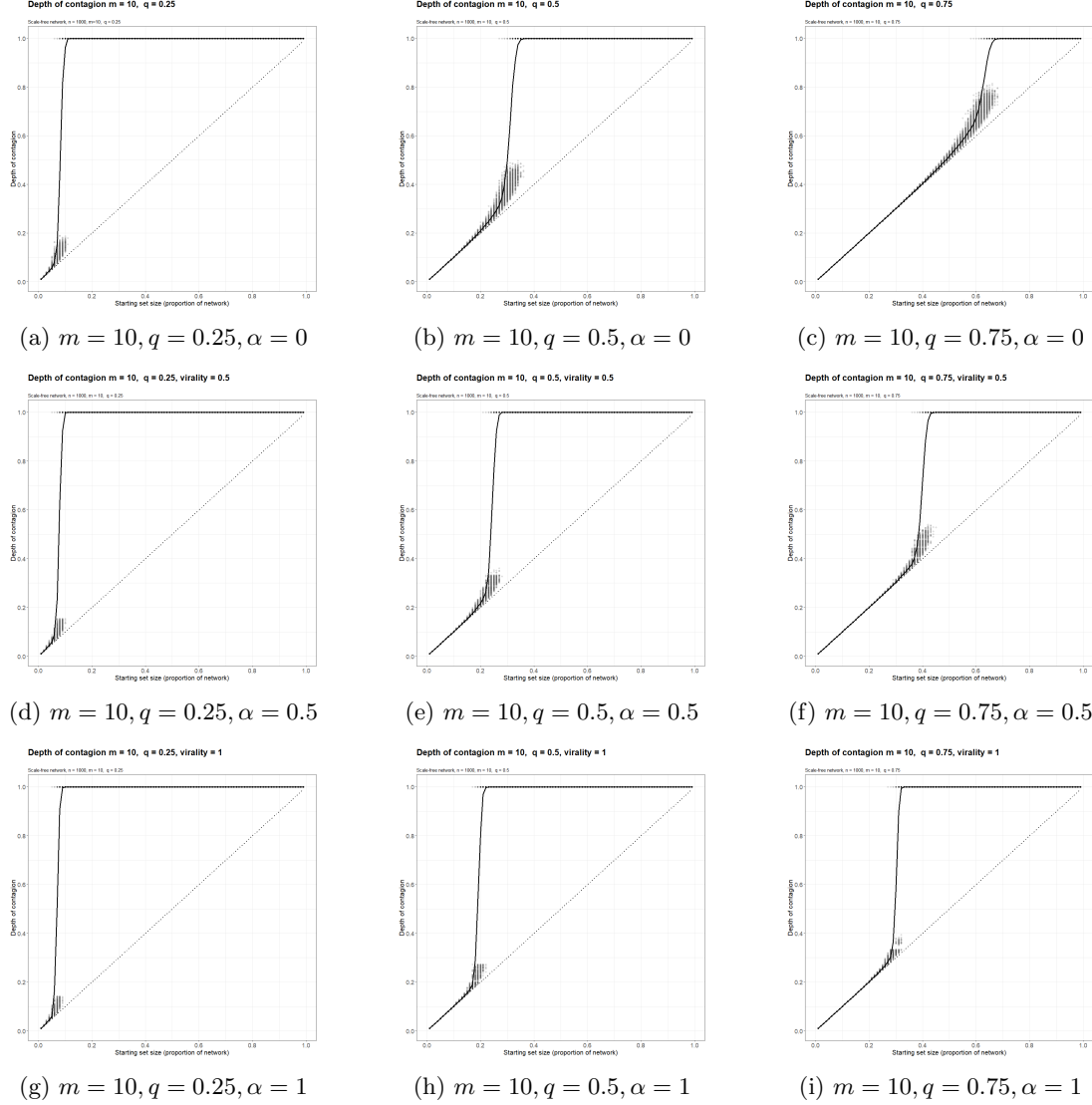


Figure 4: Depth of contagion for scale-free networks with $m = 10$

Virality increases the range of starting set sizes that lead to full contagion and brings contagion dynamics closer to a type of singularity (a narrow interval of starting set sizes below which there is no contagion and above which there is full contagion).

When $q = 0.25$, there is a narrow interval of initial starting set sizes below which contagion to any additional depth rarely occurs and beyond which contagion occurs to the entire network. For $m = 5$, this interval is about $[4,9]$ percent of the network, for $m = 10$, it is about $[6,11]$ percent of the network, and for $m = 20$, it is about $[9,15]$ percent of the network. Above this range, contagion shoots up quickly to the entire network.

When $q = 0.5$, this interval increases and virality plays a more visible role in shrinking this interval. For $m = 5$, this interval is about $[13,35]$ percent of the network without virality, but $[12,19]$ percent with virality.

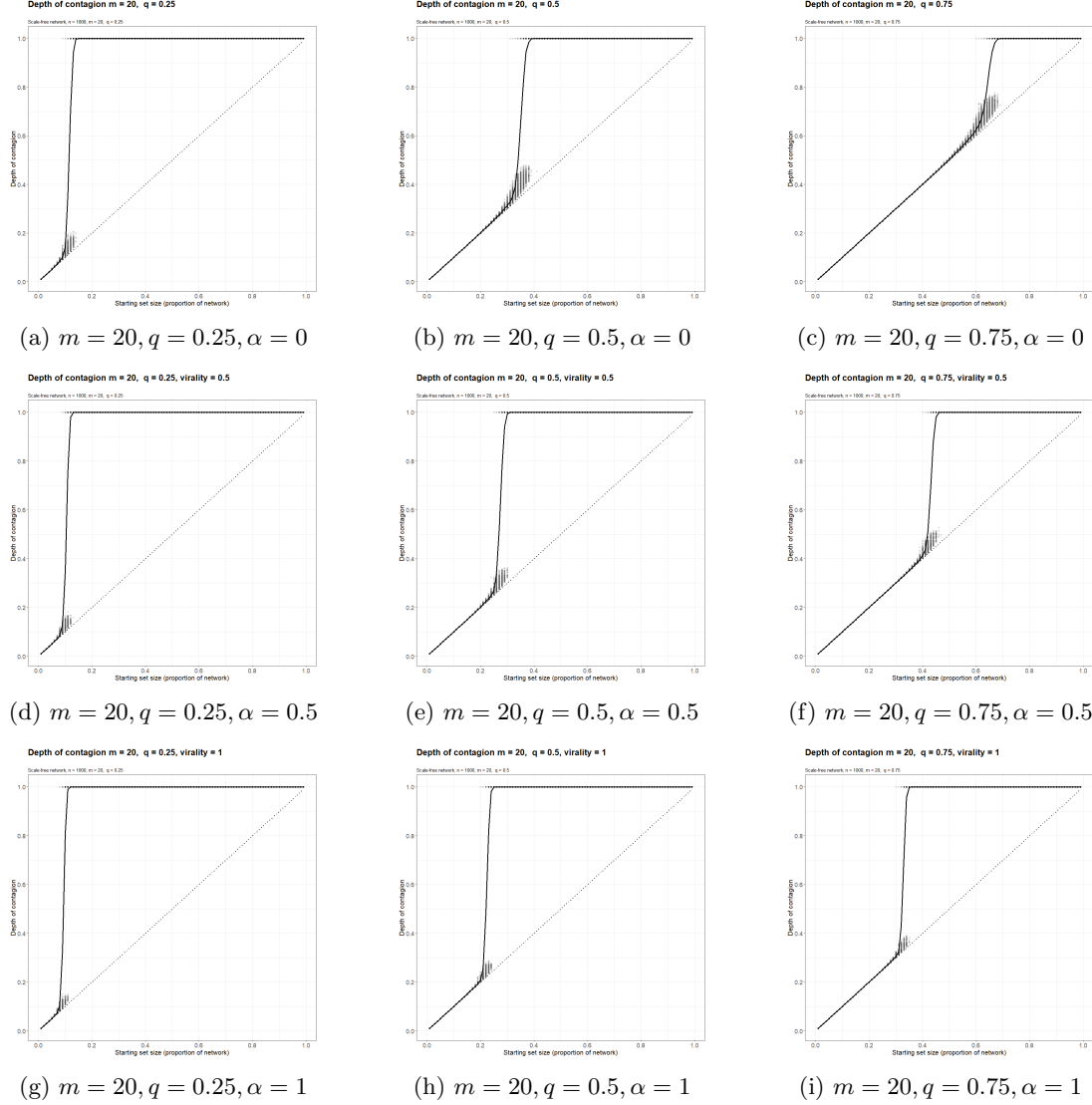


Figure 5: Depth of contagion for scale-free networks with $m = 20$

For $m = 10$, the interval is smaller without virality, about $[20,37]$ percent, and shrinks more with virality, about $[17,23]$ percent with virality. For $m = 20$, it is about $[29,40]$ percent of the network without virality, but $[20,26]$ percent with virality.

When $q = 0.75$, this interval increases more without virality and virality plays an even stronger role in shrinking this range. For $m = 5$, this interval is about $[30,68]$ percent of the network without virality, but $[20,30]$ percent with virality. For $m = 10$, it is about $[47,69]$ percent without virality, but about $[26,33]$ percent with virality. For $m = 20$, it is about $[56,69]$ percent of the network without virality, but $[30,36]$ percent with virality. Even with high network resilience and connectivity, virality reduces the range of the interval to about 6 percentage point difference.

In all three cases, in the presence of virality, there is a narrow range of about 6-10 percentage point

difference in starting set sizes below which there is no contagion and beyond which contagion spreads to the entire network. In this sense, with virality, the variability of outcomes is reduced and once contagion starts, its spread to the entire network becomes more inevitable.

Going across the figures shows the effect of network connectivity (m) on depth of contagion. Overall, increasing connectivity makes it harder to spread contagion, because for a fixed q and α , greater degree means that more players need to flip before causing a given player to play 1. The shape of the curves is qualitatively similar across networks with different attachment parameters.

5.3 Inverse depth of contagion

Our tools can be used to answer the reverse question as well. That is, given a depth of contagion, what is the minimum starting set size to achieve this depth, on average? As average depth of contagion is a (weakly) increasing function of starting set size, we can invert it to derive the (smallest) starting set size needed to achieve a desired depth of contagion. These numbers are shown in Table 2 for depths of contagion from 0.1 (10 percent of network) to 1 (100 percent of network) at steps of 0.1.

In order to spread contagion to a given depth of the network (on average), smaller starting set sizes are needed if connectivity is lower (m is low), or if relative miscoordination cost is lower (q is low), or if virality impact is higher (α is high), or a combination of these. Different combinations of these parameters have differing impacts on the starting set size needed to achieve a given depth of contagion.

When $m = 5$ and $q = 0.25$, in order to spread contagion to one-half of the network (depth = 0.5), a starting set size of 6 percent (60 players) of the network is required without virality and with virality when $\alpha = 0.5$, and 5 percent is required when $\alpha = 1$. When q is low, person-to-person spread is high and virality has a lower additional impact. This can be explained by examining the incentive to play 1, given by $\frac{s_i}{d_i} \geq q(1 - \alpha p_i)$. The partial derivative of the right-hand side with respect to α is $-qp_i$, which is small in absolute value when q is low. When q is low, an increase in α has a smaller contribution to lowering the threshold to play 1.

When $m = 5$ and $q = 0.5$, in order to spread contagion to one-half of the network on average, a starting set size of 27 percent of the network is required if $\alpha = 0$ and 16 percent if $\alpha = 1$. That is, instead of needing 270 players in the initial starting set, virality enables us to achieve the same expected depth with 160 players. This is a 40.7 percent reduction in required starting set size. For full contagion, we need 35 percent starting set size when $\alpha = 0$ and 19 percent when $\alpha = 1$, implying that virality yields a 45.7 percent reduction in starting set size.

When $m = 5$ and $q = 0.75$, full contagion requires 68 percent of the network without virality and only 31 percent when $\alpha = 1$. This is a 54.4 percent reduction in required starting set size. We go from needing more than $\frac{2}{3}$ of the network in the starting set when $\alpha = 0$ to requiring less than $\frac{1}{3}$ when $\alpha = 1$.

Depth	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$m = 5, q = 0.25$										
$\alpha = 0$	0.05	0.05	0.05	0.06	0.06	0.06	0.07	0.07	0.07	0.09
$\alpha = 0.5$	0.04	0.05	0.05	0.06	0.06	0.06	0.06	0.07	0.07	0.08
$\alpha = 1$	0.04	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.08
$m = 5, q = 0.5$										
$\alpha = 0$	0.10	0.18	0.23	0.26	0.27	0.28	0.28	0.29	0.30	0.35
$\alpha = 0.5$	0.10	0.17	0.20	0.21	0.21	0.22	0.22	0.23	0.23	0.26
$\alpha = 1$	0.10	0.15	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.19
$m = 5, q = 0.75$										
$\alpha = 0$	0.10	0.20	0.30	0.38	0.46	0.53	0.59	0.62	0.64	0.68
$\alpha = 0.5$	0.10	0.20	0.29	0.34	0.35	0.36	0.37	0.37	0.38	0.41
$\alpha = 1$	0.10	0.20	0.25	0.26	0.27	0.27	0.28	0.28	0.29	0.31
$m = 10, q = 0.25$										
$\alpha = 0$	0.07	0.08	0.08	0.08	0.09	0.09	0.09	0.09	0.10	0.11
$\alpha = 0.5$	0.07	0.07	0.08	0.08	0.08	0.08	0.09	0.09	0.09	0.11
$\alpha = 1$	0.06	0.07	0.07	0.07	0.07	0.08	0.08	0.08	0.08	0.10
$m = 10, q = 0.5$										
$\alpha = 0$	0.10	0.20	0.27	0.29	0.31	0.31	0.32	0.32	0.33	0.37
$\alpha = 0.5$	0.10	0.19	0.23	0.24	0.24	0.25	0.25	0.26	0.26	0.28
$\alpha = 1$	0.10	0.18	0.19	0.19	0.19	0.20	0.20	0.20	0.21	0.23
$m = 10, q = 0.75$										
$\alpha = 0$	0.10	0.20	0.30	0.40	0.49	0.57	0.61	0.63	0.64	0.69
$\alpha = 0.5$	0.10	0.20	0.30	0.37	0.39	0.40	0.40	0.41	0.42	0.45
$\alpha = 1$	0.10	0.20	0.28	0.30	0.30	0.31	0.31	0.31	0.31	0.33
$m = 20, q = 0.25$										
$\alpha = 0$	0.09	0.11	0.11	0.12	0.12	0.12	0.12	0.13	0.13	0.15
$\alpha = 0.5$	0.09	0.10	0.10	0.11	0.11	0.11	0.11	0.12	0.12	0.13
$\alpha = 1$	0.08	0.09	0.09	0.10	0.10	0.10	0.10	0.10	0.11	0.12
$m = 20, q = 0.5$										
$\alpha = 0$	0.10	0.20	0.29	0.34	0.35	0.35	0.36	0.36	0.37	0.40
$\alpha = 0.5$	0.10	0.20	0.26	0.27	0.27	0.28	0.28	0.29	0.29	0.31
$\alpha = 1$	0.10	0.20	0.22	0.22	0.23	0.23	0.23	0.23	0.24	0.25
$m = 20, q = 0.75$										
$\alpha = 0$	0.10	0.20	0.30	0.40	0.50	0.59	0.63	0.65	0.66	0.69
$\alpha = 0.5$	0.10	0.20	0.30	0.40	0.42	0.43	0.44	0.44	0.45	0.47
$\alpha = 1$	0.10	0.20	0.30	0.32	0.33	0.33	0.33	0.34	0.34	0.36

Table 2: Starting set size needed to achieve a given depth of contagion

The differential impact of virality is especially pronounced when q increases from 0.5 to 0.75. Without virality ($\alpha = 0$), the required starting set increases from 35 percent of the network to 68 percent (a 94.3 percent increase), whereas with virality ($\alpha = 1$), the required starting set size increases from 19 percent to 31 percent (a 63.2 percent increase). Virality reduces not only the size of the starting set required, but also the percent by which the required starting set must increase in response to a higher q . As above, this can also be explained by examining the threshold to play 1, given by $\frac{s_i}{d_i} \geq q(1 - \alpha p_i)$. The partial derivative of the right-hand side with respect to α is $-qp_i$, which is large in absolute value when q is high. When q is high, an increase in α has a larger contribution to lowering the threshold to play 1.

A large starting set implies a higher level of aggregate action, meaning that players observe a higher p_i and therefore have more incentive to play 1. This, in turn, increases the size of the set of players playing 1 (which itself makes virality stronger). In other words, the additional player in the starting set both increases the best responses of all their neighbors and also increase the aggregate level of action. These two channels of influence act together to make the rest of the network more willing to play 1. This feedback loop enabled by virality shows how starting set size exhibits increasing returns to scale (for contagion). Hence, increases in the virality parameter (which controls the sensitivity of players to aggregate action) will magnify the marginal contribution of an additional player in the starting set through this feedback loop, thus increasing the marginal benefit of starting set size.

The implication is that to achieve a given depth of contagion, virality has the largest impact in scenarios where larger starting sets are required in the absence of virality and when the network is especially resilient to contagion (when q is high). In these cases, virality can have an outsized impact in reducing the required starting set size.

6 Policy implications

Our results show that in several classes of situations, contagion can occur from a relatively small starting set to a relatively large set. These include situations in which network resilience (q) is low, or connectivity (m) is low, or virality (α) is high. Combinations of these may amplify or dampen the spread of contagion. This has numerous implications for social interactions affecting societal cohesion and democratic institutions, and for the spread of infectious diseases.

If a group of people seeks to sow discord within a country by spreading misinformation, they only need to convince a relatively small minority of citizens to believe an incendiary piece of information, and this may be sufficient to influence a large proportion of society. The larger the initial set of people convinced, the greater the potential spread of misinformation. Similarly, for infectious diseases, the larger is the initial proportion of people infected, the greater is the spread of the disease. Therefore, one obvious way to limit

spread of contagion is to restrict starting set sizes. For misinformation, this may involve identifying and limiting accounts designed mainly to spread misinformation. For infectious disease, proliferation can be stopped more easily by attempts to contain the spread when the proportion of the infected population is low, perhaps with measures such as quarantining.

Our results show that virality amplifies the effect of a larger starting set. As shown in Table 1 and Figure 2, without virality ($\alpha = 0$), even with very large starting sets, contagion does not occur to the entire network. With virality ($\alpha = 1$), once about 40 percent of the network has an incentive to play 1, the spread of contagion to the entire network is inevitable regardless of network resilience (q), at least for low and moderate connectivity networks ($m = 5, 10$). Therefore, in networks where virality is a common feature, there is a greater chance to curtail spread of misinformation by nipping it in the bud. Waiting for things to play out will add to the likelihood of a considerably larger spread of contagion in the presence of virality. For misinformation, the impact of virality may be curtailed by network wide policies that restrict spread of incorrect or misleading information. For infectious diseases, the effect of virality can be curtailed by restricting cross-network travel and also by limiting non-neighbor interaction in public situations. Taking the reverse actions increases the spread of contagion.

We show that higher network resilience and higher connectivity reduce the spread of contagion but an increase in virality can rollback these effects. As shown in Figures 3, 4, and 5, for a given starting set size, reducing network resilience (lowering q) or reducing connectivity (lowering m) causes deeper contagion, and increasing virality (higher α) also causes deeper contagion. Indeed, virality brings the spread of contagion closer to a type of singularity in the sense that there is a small interval of starting set sizes below which contagion does not spread beyond initial conditions and above which contagion spreads to the entire network. This shows that even in more resilient and connected networks, an ability to control virality may be the difference between localized contagion around the starting set or widespread contagion throughout the network. In other words, even in resilient social networks, policies to identify and reduce the sharing of misleading information may be the difference between localized spread of misinformation or its network-wide spread. Similarly, even in more resilient populations, policies to limit cross-network interaction may be the difference between local containment of an infectious disease or its global spread.

Our results show that it is not difficult to achieve large scale contagion with a small starting set, especially in cases of lower network connectivity and resilience. As shown in Table 2, a starting set size of about 6 percent of the network is sufficient to infect about 80 percent of the network when $m = 5$, $q = 0.25$, and $\alpha = 1$. Even in highly resilient and connected networks ($m = 20$, $q = 0.75$, $\alpha = 1$), a 36 percent starting set leads to global contagion. Therefore, whether it is spread of misinformation or an infectious disease, identifying the phenomenon early increases the chance of implementing measures to prevent its spread and prevent greater damage. Conversely, delaying action is more likely to result in larger spread of misinformation

or more infection and to lead toward potentially large and destabilizing systemic impact.

More broadly, the analysis here shows how different policies have different effects depending on the situation and how combinations of different interventions can be used to design more effective policies to achieve the desired effects in a situation-dependent manner. This may help to control better the spread of contagion in connected networks.

To conclude, we model the spread of contagion using a coordination game on a network and extend the base model to include a flexible formulation of aggregative virality. We provide tractable new algorithms to compute depth of contagion and the threshold for network contagion for an arbitrary, finite, connected network and starting from an arbitrary set, solving a long-standing problem in the related literature. Our results are founded in the game theory framework of interdependent decision-making and Nash equilibrium. Our algorithms allow for more extensive simulation, characterization, and study of networks than was previously accessible in the network games literature. The research here may help other researchers study additional applications using these methods.

References

- [Adler, 1991] Adler, J. (1991). Bootstrap percolation. *Physica A: Statistical Mechanics and its Applications*, 171(3):453–470.
- [Angeletos et al., 2006] Angeletos, G.-M., Hellwig, C., and Pavan, A. (2006). Signaling in a global game: Coordination and policy traps. *Journal of Political Economy*, 114(3):452–484.
- [Angeletos et al., 2007] Angeletos, G.-M., Hellwig, C., and Pavan, A. (2007). Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks. *Econometrica*, 75(3):711–756.
- [Barabási and Albert, 1999] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.
- [Basak and Zhou, 2020] Basak, D. and Zhou, Z. (2020). Diffusing coordination risk. *American Economic Review*, 110(1):271–97.
- [Bramoullé and Kranton, 2016] Bramoullé, Y. and Kranton, R. (2016). Games played on networks. In Bramoullé, Y., Galeotti, A., and Rogers, B., editors, *The Oxford Handbook of the Economics of Networks*. Oxford University Press.
- [Brennen et al., 2020] Brennen, J. S., Simon, F. M., Howard, P. N., and Nielsen, R. K. (2020). Types, sources, and claims of COVID-19 misinformation. *Reuters Institute*.
- [Broniatowski et al., 2018] Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., Quinn, S. C., and Dredze, M. (2018). Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. *American Journal of Public Health*, 108(10):1378–1384.
- [Carlsson and Van Damme, 1993] Carlsson, H. and Van Damme, E. (1993). Global games and equilibrium selection. *Econometrica*, pages 989–1018.
- [Centola, 2010] Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197.
- [Centola and Macy, 2007] Centola, D. and Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113(3):702–734.
- [Chwe, 2000] Chwe, M. S.-Y. (2000). Communication and coordination in social networks. *The Review of Economic Studies*, 67(1):1–16.
- [Dasaratha, 2020] Dasaratha, K. (2020). Innovation and strategic network formation. Working paper.
- [Elliott and Golub, 2019] Elliott, M. and Golub, B. (2019). A network approach to public goods. *Journal of Political Economy*, 127(2):730–776.
- [Elliott et al., 2014] Elliott, M., Golub, B., and Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, 104(10):3115–53.
- [Frankel et al., 2003] Frankel, D. M., Morris, S., and Pauzner, A. (2003). Equilibrium selection in global games with strategic complementarities. *Journal of Economic Theory*, 108(1):1–44.
- [Frenkel, 2020] Frenkel, S. (2020). Meet the top election misinformation ‘superspreaders.’ Nov 23, 2020. *New York Times*.
- [Galeotti et al., 2010] Galeotti, A., Goyal, S., Jackson, M. O., Vega-Redondo, F., and Yariv, L. (2010). Network games. *The Review of Economic Studies*, 77(1):218–244.
- [Golub and Jackson, 2010] Golub, B. and Jackson, M. O. (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–49.

- [Golub and Sadler, 2016] Golub, B. and Sadler, E. (2016). Learning in social networks. In *The Oxford Handbook of the Economics of Networks*. Oxford University Press.
- [Goyal, 2009] Goyal, S. (2009). *Connections: An introduction to the economics of networks*. Princeton University Press.
- [Granovetter, 1978] Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443.
- [Grinberg et al., 2019] Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., and Lazer, D. (2019). Fake news on Twitter during the 2016 US Presidential Election. *Science*, 363(6425):374–378.
- [Hoffmann and Sabarwal, 2019] Hoffmann, E. J. and Sabarwal, T. (2019). Global games with strategic complements and substitutes. *Games and Economic Behavior*, 118:72–93.
- [Jackson, 2008] Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press.
- [Jackson and Storms, 2019] Jackson, M. O. and Storms, E. C. (2019). Behavioral communities and the atomic structure of networks. Working Paper.
- [Jackson and Yariv, 2006] Jackson, M. O. and Yariv, L. (2006). Diffusion on social networks. *Economie publique/Public economics*, (16).
- [Jackson and Yariv, 2007] Jackson, M. O. and Yariv, L. (2007). Diffusion of behavior and equilibrium properties in network games. *American Economic Review*, 97(2):92–98.
- [Jackson and Zenou, 2015] Jackson, M. O. and Zenou, Y. (2015). Games on networks. In *Handbook of Game Theory with Economic Applications*, volume 4, pages 95–163. Elsevier.
- [Morris, 2000] Morris, S. (2000). Contagion. *The Review of Economic Studies*, 67(1):57–78.
- [Morris and Shin, 2003] Morris, S. and Shin, H. S. (2003). Global games: Theory and applications. In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Volume 1*, pages 56–114. Cambridge University Press.
- [Sadler, 2020] Sadler, E. (2020). Diffusion games. *American Economic Review*, 110(1):225–70.
- [Shao et al., 2018] Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., and Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, 9(1):4787.
- [Varol et al., 2017] Varol, O., Ferrara, E., Davis, C. A., Menczer, F., and Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. In *Eleventh international AAAI conference on web and social media*.
- [Vosoughi et al., 2018] Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.
- [Watts, 2002] Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771.
- [Wiedermann et al., 2020] Wiedermann, M., Smith, E. K., Heitzig, J., and Donges, J. F. (2020). A network-based microfoundation of Granovetter’s threshold model for social tipping. *Scientific Reports*, 10(1):1–10.
- [Young, 2011] Young, H. P. (2011). The dynamics of social innovation. *Proceedings of the National Academy of Sciences*, 108(Supplement 4):21285–21291.