Social Networks 34 (2012) 451–461

Contents lists available at SciVerse ScienceDirect

Social Networks

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

Critical phenomena in complex contagions

Vladimir Barash, Christopher Cameron, Michael Macy*

Cornell University, Ithaca, NY, United States

ARTICLE INFO

Keywords:
Contagion
Diffusion
Propagation
Social influence
Critical mass
Thresholds
Small worlds
Power laws

ABSTRACT

Why do some contagions "go viral" and others do not? Research on "small world" networks (Watts and Strogatz, 1998) shows how a very small number of long-range ties that bridge between clusters can allow contagions to spread almost as rapidly as on a random network of equal density. Recent research shows how long-range ties that accelerate the spread of information and disease can impede the spread of *complex contagions*—behaviors, beliefs and preferences that diffuse via contact with multiple adopters (Centola and Macy, 2007). In confirming this result analytically and extending the analysis from small world to power law networks, we discovered that complex contagions require a critical mass of infected nodes that corresponds to a phase transition in the ability of the contagion to take advantage of the "shortcuts" created by long-range ties. We demonstrate how this critical mass is related to the dynamics of the contagion process and identify implications for modeling behaviors that spread via social influence, such as viral marketing and social movements.

© 2012 Published by Elsevier B.V.

1. Background

"Critical mass" refers to the point at which a dynamical process becomes self-sustaining, as in nuclear fission. Our interest focuses on critical mass in the spread of complex contagions (Centola and Macy, 2007)—behaviors, beliefs, and preferences that spread via contact with multiple adopters. Examples include the tipping point in the adoption of a virally marketed product, the take-off point at which a new fashion becomes a fad, and the sudden explosion of participation in a new social movement whose appeal increases with its popularity. For complex contagions, critical mass is the level of infection at which the contagion process becomes self-sustaining, such that each additional infected node leads to one or more additional nodes to become infected, until the contagion saturates the target population. We model the dynamics of critical mass in complex contagions and identify the critical value at which the spread of a complex contagion becomes self-sustaining. We show how this value can be derived analytically, and how the value depends on network topology and the network externality of the contagion.

In contrast to simple contagions, complex contagions have a critical value at both the individual and the network levels. At the individual level, network externality refers to the effect of infection of some node in a network on the ability of other infected nodes to infect their neighbors. The spread of many social contagions – which may be risky or costly to adopt – has a positive network externality, so the probability of a node's infection increases with the number of that node's infected neighbors. We focus on a very simple model of positive network externality: the threshold model of contagion, where a node will become infected if and only if a critical fraction of its neighbors have become infected. This property has important implications for the population dynamics: complex contagions may die out early on due to a lack of redundant exposures to multiple infected neighbors. Thus complex contagions can have a critical value at both the individual level and the population level. At the individual level, the critical value is the proportion of neighbors who must be infected in order to infect a given individual, which we define as the "threshold." At the population level, the critical value is the proportion of the population who must be infected in order for the contagion to become self-sustaining, which we define as the "critical mass." Our research investigates the relationship between individual-level thresholds and population-level critical mass in social contagions.

The population-level and individual-level critical values are known to be related (Granovetter, 1978; Schelling, 1971a), but the form of the relationship is not well understood. We show that for two widely studied network structures – small world and power law networks – a population-level critical mass exists that may be far smaller (as a proportion of population size) than is the individual level threshold (as a proportion of neighborhood size, where a neighborhood is defined as a node and all nodes adjacent to it). This critical mass corresponds to a phase transition, from local to global propagation, once the contagion acquires the ability to escape the region of initial infection via long-range ties that bridge across clusters. We demonstrate how this critical mass is related to the dynamics of the contagion. Prior to reaching critical mass, complex

* Corresponding author. Tel.: +1 607 351 8194.
E-mail address: m.macy@cornell.edu (M. Macy).

0378-8733/$ – see front matter © 2012 Published by Elsevier B.V.
doi:10.1016/j.socnet.2012.02.003

contagions are fragile and highly dependent on the idiosyncrasies of local network structure. However, once contagions reach critical mass, they become self-sustaining and highly likely to spread throughout a connected population. In the conclusion, we discuss the theoretical implications of our analysis for explaining how and why contagions "go viral" and the practical implications for viral marketing and social movement mobilization.

Centola and Macy (2007) provide an extended discussion of the micro-mechanisms that underlie a wide range of diffusion processes that are likely to involve complex contagions. These mechanisms include high costs, high risks, low legitimacy, and positive externalities. Although these mechanisms differ widely in their empirical application, they share the theoretically relevant attribute that the spread of the contagion depends on social reinforcement from multiple infected neighbors. Since Centola and Macy (2007) described these mechanisms in considerable detail, we refer readers to their paper rather than rehearsing the discussion here.

The work of Centola and Macy is situated in a context of broad research into contagion models, including stochastic contagion models such as SIR (Anderson and May, 1991), as well as the models developed by Steglich et al. (2010). Stochastic contagion models implement threshold effects as the change in the probability of infection with changes in the number/proportion of an agent's infected neighbors, where this probability is always positive. Thus, stochastic contagion models imply that every contagion will lead eventually to global infection in the limit of infinite time. Deterministic thresholds are a convenient simplification that allows investigation of the contagion dynamics in finite time that are implied by agent's need for social reinforcement. The Centola and Macy model shows how contagion can spread throughout the entire network very quickly, or never spread further than a tiny fraction of the nodes. The contrast between global takeover and "failure" is a useful model for empirical contagions, most of which garner only a few adopters, while a few become immensely popular. Towards the end of this paper, we introduce an extension of our analysis to stochastic models of contagion, and show that stochastic contagions that reach critical mass spread a lot faster than sub-critical stochastic contagions. However, we leave a systematic exploration of the critical mass implications of stochastic models of contagion to future work.

Empirical research into complex contagions is a relatively new field; however, we would like to draw attention to two recent studies. Centola (2010) studies the spread of health behavior in artificially structured online communities, and finds that individual adoption was more likely when participants received social reinforcements from multiple network neighbors and that health behavior spread farther and faster across clustered-lattice networks than across corresponding random networks. Romero et al. (2011) study the diffusion of "hashtags" (which include information, memes, and proxies for certain behaviors like joining a political movement) through the Twitter social network and find that hashtags related to internal Twitter memes are less likely to be adopted after multiple exposures relative to hashtags related to politics, suggesting that political hashtags may more closely resemble complex contagions than memes. Based on these and other studies, we believe that our research into the dynamics of complex contagions has appications to the spread of behaviors and products in the real world.

2. Model

Our research builds on and extends the model of complex contagions developed by Centola and Macy (2007), which is identical to the Watts and Strogatz small-world model (Watts and Strogatz, 1998) except that it allows for individual thresholds greater than one. The model operates on a networked population of N agents, where each agent has one binary state, "infected" or "uninfected," representing whether the individual has adopted a behavior (e.g. acquired a new technology), belief (e.g. an urban legend), or norm (e.g. smoking is uncool). The agents in this population can change their state only from uninfected to infected, in the following deterministic way: if some threshold quantity a, or fraction z/n of an uninfected agent's neighbors are infected, its state changes to infected as well. The quantity a is called the *absolute threshold* of infection for a node, and the fraction z/n is called the *relative threshold* of infection. We first consider the case of absolute thresholds but explore relative threshold models towards the end of our analysis. Following Centola and Macy, we define a "simple contagion" as $a = 1$ for absolute thresholds and $z = 1/n$ for relative thresholds. A contagion is "complex" if $a > 1$ or $z > 1/n$. A small set of nodes A have threshold $a = 0$, corresponding to "innovators" who will buy a product or join a movement without social influence from prior adopters in their neighborhood. Centola and Macy restrict the set A to the neighborhood of a randomly chosen node. Starting with the set A, the contagion is propagated throughout the population until no more agents can be infected (which can happen if all agents are infected or if no uninfected agent has at least a (or z/n) infected neighbors).

3. Analysis

Centola and Macy focus on the spread of complex contagions on a perturbed regular lattice of degree eight, with p ties that are randomly rewired ($0 \leq p \leq 1$) in pairs so as to leave the degree of each node unchanged. If $p = 0$, all ties in the lattice have range 2 since every pair of adjacent nodes has at least one neighbor in common. A small amount of random rewiring transforms the lattice into a "small world" network characterized by a few long-range ties and high average clustering coefficient. That is because random rewiring on a large lattice almost always replaces ties with minimal range (a path length of two steps) with long-range ties that create "shortcuts" for the spread of a contagion over the lattice. These shortcuts allow simple contagions to spread throughout the network far more quickly than would be possible along highly clustered short-range ties. For example, in a small-world network of N nodes, the path length between any two nodes via short-range ties is $O(N^{1/2})$ steps, whereas the path length via long-range shortcuts is $O(log(N))$ steps.

Centola and Macy's contribution was to show that complex contagions spread farther and faster on an unperturbed lattice ($p = 0$) than on a small-world network (e.g. $p = .1$). The spread of complex contagions on a small world network depends on the probability of contact with additional infected neighbors, given that there is contact with one infected neighbor. For nodes whose networks are highly clustered (their neighbors are also neighbors of one another), if a node has one infected neighbor then that node is highly likely to have other infected neighbors, even when the proportion of the population that is infected is still very small. In contrast, for nodes whose neighbors are randomly chosen, when the size of the infected population is small, so too is the probability that a node will have additional infected neighbors, given that the node has one infected neighbor. As the number of infected nodes grows, so too does the probability that random ties will connect an uninfected node with a sufficient number of infected neighbors for that node to also become infected. We identify the critical mass as the point at which the proportion of infected nodes is sufficient for contagions to take advantage of the shortcuts created by long-range ties. For simple contagions, the critical mass is uninteresting, since it is achieved at a single infected node, i.e. the seed node of the

contagion. For complex contagions, the critical mass is always greater than one, and knowing how much greater is important from both a theoretical and practical perspective.

In this section, we use an analytical approach to derive critical mass points for different classes of networks under different conditions (threshold, rewiring probability, degree distribution). The advantage of an analytical approach is that it is deductive and generalizable across parameter settings. However, the main contribution of this paper is not the analysis itself but rather the observation of the critical mass phenomenon via analysis, and the confirmation of this phenomenon via simulation.

3.1. Small world networks

We begin by deriving the critical mass for small world networks, modeled as a perturbed regular lattice. The perturbation is achieved by iteratively taking two edges and, if possible, switching their endpoints: that is, edges (a, b) and (c, d) become (a, d) and (c, b). This model captures two defining properties of empirically observed small world networks – that most ties have minimal range (the modal range is 2 in most empirical networks), while the network also has a relatively small number of long-range ties. The short-range ties correspond to the high level of clustering that is observed in most empirical social networks, while a few long-range ties make possible the surprisingly short mean geodesic (such as the widely observed "six degrees of separation"). Rewired ties on the perturbed lattice are a highly stylized representation of the empirical regularity that nodes in a social network have some mix of highly clustered ties (usually to close friends and family) and long-range ties (usually to acquaintances). Following Centola and Macy, we impose the conservative simplification that infected acquaintances exert as much influence as close friends. Relaxing that assumption is equivalent to increasing a, the threshold number of infected neighbors that are required for a node to become infected. The derivation of the critical mass is based on the probability P_{RW_a} of an uninfected node having a rewired ties to infected nodes, as the number of infected nodes increases. This probability is integral to the critical mass calculations: if the likelihood of having sufficient ties to infected nodes is close to 0, then the contagion is unlikely to spread any further. In contrast, if this probability is close to 1, then we can expect with high probability that at least one more node will become infected.

Theorem 3.1. *Given a randomly rewired lattice of* N *nodes, where every node has probability* p *of having one of its ties rewired, and* I *infected nodes on that network, the probability P_{RW_a} that some uninfected node c has a rewired ties to infected neighbors is given by:*

$$P_{RW_a} \approx 1 - P_{NIa}(c)^{\binom{k}{a}^{(N-I)}} \tag{1}$$

where

$$P_{NIa}(c) = 1 - p^a + p^a \left(\frac{\binom{N-1}{a} - \binom{I}{a}}{\binom{N-1}{a}} \right) \tag{2}$$

Proof. The proof of this theorem is based on an axiom of probability theory that the probability of an event happening for some node is equal to one minus the probability of that event not happening for any node. Accordingly, instead of trying to calculate the positive probability of some uninfected node having a rewired ties to infected nodes, we can calculate the probability that no uninfected node has a rewired ties to infected nodes, which turns out to be a simpler calculation, and then subtract that negative probability from one to arrive at P_{RW_a}:

$$P_{RW_a} = 1 - \prod_{(c) \in V \setminus F} P_{NI}(c) \tag{3}$$

where V is the set of all nodes, F is the set of infected nodes, and $P_{NI}(c)$ is the probability that some uninfected node c does not have rewired ties to a infected nodes. This probability depends on the occurrence of two events: first, c has to have sufficient rewired ties, and second, they have to target a sufficient number of infected nodes. The likelihood of a node having a given number of rewired ties is constant for all nodes, as all nodes on a lattice have the same degree and the rewiring process picks nodes at random. Similarly, the probability of an edge targeting an infected node is uniform across all edges, since rewired edges are targeted at random. The only exception to this case is when $N - I$ is very small so the number of possible targets that are not infected nodes is quickly exhausted. Therefore, given the qualification that $N - I$ is large, we can assume that $P_{NI}(c)$ is uniform across all uninfected nodes c in the network, and rewrite the above as:

$$P_{RW_a} \approx 1 - P_{NI}(c)^{|V \setminus F|} \tag{4}$$

$$P_{RW_a} \approx 1 - P_{NI}(c)^{N-I} \tag{5}$$

$NI(c)$ holds (that is, c does not have rewired ties to a rewired nodes) if we can't pick a of c's ties such that all a are rewired and both point to infected nodes. There are $\binom{k}{a}$ independent ways to pick a of c's ties, where k is c's degree (uniform over all nodes in a lattice). So we can again rewrite:

$$P_{RW_a} \approx 1 - P_{NIa}(c)^{\binom{k}{a}^{(N-I)}} \tag{6}$$

where $P_{NIa}(c)$ is the probability that, having picked some set of c's ties with a elements, at least one of these ties is not rewired and/or does not point to an infected node. The event $NIa(c)$ holds if these conditions are true. Tie rewiring is an independent process (the probability of one tie being rewired does not affect the probability of other ties being rewired), so with probability $1 - p^a$ at least one tie is not rewired and $NIa(c)$ holds. In the opposite case, $NIa(c)$ still holds so long as the targets of all a ties are not in F. Note that the number of possible targets of c's ties is $N - 1$, since c cannot have ties to itself. We can now write $P_{NIa}(c)$ as an interpolation between $1 - p^a$ and p^a:

$$P_{NIa}(c) = 1 - p^a + p^a \left(\frac{\binom{N-1}{a} - \binom{I}{a}}{\binom{N-1}{a}} \right) \tag{7}$$

□

This theorem calculates the most conservative case where a contagion has to spread to some node entirely via long-range ties, even though it is possible for the contagion to spread through any combination of long-range (rewired) and short-range (unrewired) ties. Consequently, the results in this paper slightly understate the point at which a contagion begins to take advantage of shortcuts in the network. We follow the conservative approach for two reasons: greater simplicity of analysis, and (more importantly) the implication of spreading entirely through long-range ties for the growth rate of the contagion, which we discuss towards the end of the paper in Section 4.

3.2. Power law networks

Theorem 3.1 can be extended beyond the rewired lattice. We now present a lemma that derives an approximation to P_{RW_a} for power law networks. In power-law networks, the notion of a "rewired" tie does not apply. Instead, following the Barabasi–Albert model of power-law networks (Albert and Barabási, 2002), we assume that ties are formed according to preferential attachment, with higher-degree nodes more likely to be the targets of ties.

Lemma 3.2. *Given a power-law network of* N *nodes with sum degree* SN *where degree follows a power-law distribution with exponent* α *and ties are formed according to preferential attachment, and* I *infected nodes with sum degree* SI *on that network, the probability* P_{LR_a} *that any uninfected node c has a ties to infected neighbors is approximated by:*

$$P_{LR_a} \approx 1 - \left[1 - \left(\frac{SI}{SN}\right)^a\right]^{k\alpha^{NTa}} \tag{8}$$

where k *is a factor parameter given by:*

$$k = \frac{a^a}{a!} \tag{9}$$

and NT *is a parameter estimated from the degree distribution by:*

$$NT \approx (N-I)\left(\frac{a}{mindeg}\right)^{-\alpha+1} \tag{10}$$

with mindeg *the minimum degree in the network.*

Proof. The proof is very similar to the proof of Theorem 3.1, so here we concentrate only on the differences. First, we have to redefine the probability $P_{NIa}(c)$ that, given some set of c's ties with cardinality a, at least one of those ties is not rewired and/or does not point to an infected node. This probability no longer depends on rewiring. Without preferential attachment, we could write $P_{NIa}(c)$ as simply:

$$P_{NIa}(c) = \frac{\binom{N-1}{a} - \binom{I}{a}}{\binom{N-1}{a}} \tag{11}$$

$$= 1 - \frac{\binom{I}{a}}{\binom{N-1}{a}} \tag{12}$$

$$= 1 - \frac{(I!)/(a!(I-a)!)}{((N-1)!)/(a!(N-1-a)!)} \tag{13}$$

For $N>>I>>a$, we have:

$$\frac{I!}{(I-a)!} \approx I^a \tag{14}$$

$$\frac{(N-1)!}{(N-1-a)!} \approx (N-1)^a \tag{15}$$

so for small a we can approximate:

$$P_{NIa}(c) \approx 1 - \left(\frac{I}{N}\right)^a \tag{16}$$

The only correction we have to make is related to the power-law degree distribution, where each node gets up-weighted by its degree. We can transform these weights into discrete values by counting each node as many times as it has degree. Combinations of nodes from the resulting augmented sets are equivalent to weighted combinations from the original sets. The resulting equation is almost the same, except we substitute SI and SN, the sum degrees of all I infected nodes and all N nodes in the full population:

$$P_{NIa}(c) \approx 1 - \left(\frac{SI}{SN}\right)^a \tag{17}$$

The other difference from the proof of Theorem 3.1 is that the probability $P_{NI}(c)$ that c does not have rewired ties to a infected nodes, is no longer uniform over all nodes, as they do not all have the same degree. This results in the upper exponent $N-I$ being replaced with a sum SE of $N-I$ terms, where each term represents all the ways to choose a nodes from all the neighbors of a particular uninfected node:

$$P_{LR_a} \approx 1 - P_{NI}(c)^{SE}$$
$$SE = \sum_{i \in V \setminus F} \binom{N(i)}{a}$$

Here V is the set of all nodes, F is the set of infected nodes and $N(i)$ is the number of network neighbors of a particular uninfected node i. We now examine the terms in this sum. Each of these terms is a fraction:

$$\binom{N(i)}{a} = \frac{N(i)!}{a!(N(i)-a)!} \tag{18}$$

we can take $1/a\,!$ out of the sum, and transform as follows:

$$SE = \frac{1}{a!}\sum_{i \in V \setminus F} N(i)(N(i)-1)\ldots(N(i)-a+1) \tag{19}$$

For small a and large $N(i)$ (which will dominate the sum), we can approximate as follows:

$$SE \approx \frac{1}{a!}\sum_{i \in V \setminus F} (N(i))^a \tag{20}$$

Note that we can ignore all terms in this sum where i has fewer than a neighbors (since there are no ways to choose a units from a set smaller than a, those terms are 0). Since the degree distribution follows a power-law, the number of uninfected nodes with degree $\geq a$ is given by:

$$NT \approx (N-I)\left(\frac{a}{mindeg}\right)^{-\alpha+1} \tag{21}$$

where *mindeg* is the minimum degree for any node in the network (we can crudely estimate it as 1) and α is the exponent of the power law distribution. So there are NT terms in the sum overall:

$$SE \approx \frac{1}{a!}\sum_{j=1}^{NT} (N(i))^a \tag{22}$$

In a power-law distribution with discrete values, term density thins out at a rate proportional to the exponent, that is, individual degree values will be roughly powers of the exponent α. The smallest of the relevant values lies somewhere between a and αa. Since a will be by far the smallest term in the sum, we can drop it and approximate as follows:

$$SE \approx \frac{1}{a!}\sum_{j=1}^{NT} a^a \alpha^{aj} \tag{23}$$

Finally, we can extract a^a and the sum becomes a geometric series:

$$SE \approx \frac{a^a}{a!}\frac{\alpha^{a(NT+1)}-1}{\alpha^a-1} \approx \frac{a^a}{a!}\alpha^{NTa}$$

Putting everything together, we get:

$$P_{LR_a} \approx 1-\left[1-\left(\frac{SI}{SN}\right)^a\right]^{(a^a/a!)\alpha^{NTa}} \quad (24)$$

□

3.3. Critical behavior

3.3.1. Number of infected nodes

We now examine the behavior of the probabilities of infection via long-range ties, P_{RW_a} and P_{LR_a}, at limiting values of the number of infected nodes I. First, let us consider $I < a$, where a threshold a contagion cannot spread. First, for small world networks, we can rewrite Eq. (2) as follows:

$$P_{NIa}(c) = 1-p^a+p^a\frac{\binom{N-1}{a}-0}{\binom{N-1}{a}} = 1 \quad (25)$$

where a is the threshold, p is the rewiring probability, and N is the number of nodes. For the power-law network we can rewrite Eq. (11) as follows:

$$P_{NIa}(c) = 1-\frac{0}{\binom{N-1}{a}} = 1 \quad (26)$$

with the same variables as above. So $P_{NIa}(c)=0$ which by Eqs. 1 and 8 means $P_{RW_a} = P_{LR_a} = 0$ when $I < a$. So the contagion is indeed guaranteed not to spread through long-range ties when too few nodes are infected, because there are too few infected nodes to which uninfected nodes might be tied. In the opposite case, when $I = N-1$, for the rewired lattice we can rewrite Eq. (1) as:

$$P_{RW_a} = 1-(1-p^a)\binom{k}{a} \quad (27)$$

where p is the rewiring probability, a the threshold and k is node degree which is uniform across all nodes. This equation indicates that the probability of the final node being infected depends only on that node having a rewired ties, as the model suggests. For the power-law case, $P_{NIa}(c)=0$, since it no longer depends on rewiring, hence we can rewrite Eq. (8) as:

$$P_{LR_a} = 1-0^{SE} \quad (28)$$

where SE is the sum across all nodes of different ways of choosing a neighbors of that node. If $SE > 0$ (the last uninfected node has degree a or more), then $P_{RW_a} = 1$, since the node is guaranteed to have a infected neighbors. If $SE = 0$, then $P_{RW_a} = 1-0^0 = 0$, since the node has insufficient ties to become infected.

3.3.2. Threshold

Next, we consider the behavior of P_{RW_a} at limiting values of threshold a for both the small world (perturbed lattice) and power law networks. For the case $a=0$ for rewired lattices we can rewrite Eq. (2) as:

$$P_{NIa}(c) = 1-p^0+p^0\frac{1-1}{1} = 0 \quad (29)$$

and for the power-law case we can rewrite Eq. (11):

$$P_{NIa}(c) = \frac{1-1}{1} = 0 \quad (30)$$

So, assuming $p > 0$ (for the rewired lattice) and $I < N-1$, $P_{RW_a} = P_{LR_a} = 1$, which shows that a contagion with threshold 0 is guaranteed to spread on all networks.

Now consider the case of simple contagions with threshold $a=1$. For rewired lattices we can rewrite Eq. (1) as:

$$P_{RW_a} = 1-\left(1-\frac{pI}{N-1}\right)^{k*(N-I)} \quad (31)$$

and for the power-law case we can rewrite Eq. (8) as:

$$P_{LR_a} \approx 1-\left[1-\frac{SI}{SN}\right]^{\alpha^{N-I}} \quad (32)$$

This reduction indicates that the spread of simple contagions is a more robust phenomenon than the spread of complex contagions. For rewired lattices, simple contagions will spread across any rewired ties between infected and uninfected nodes, even if there is only a single infected node in the population. For power-law networks, simple contagions will spread across any tie, and will deterministically infect more nodes so long as the infected cluster has non-zero degree, the power law exponent is not zero, and there remain nodes to infect.

3.3.3. Rewiring probability

It is also instructive to consider values of the probability P_{RW_a} of some node having rewired ties to a infected nodes for small world networks at limiting values of rewiring probability p. For $p=0$, Eq. (2) reduces to $P_{NIa}(c)=1$, so $P_{RW_a}=0$ and the contagion cannot take advantage of long-range ties, because there are none. For $p=1$, we can rewrite Eq. (2) as:

$$P_{NIa}(c) = 1-\frac{\binom{I}{a}}{\binom{N-1}{a}} \quad (33)$$

So the spread of the contagion depends entirely on the size of the infected cluster.

3.4. Estimation of function behavior

We analyze other critical points of the probabilities of infection via long-range ties, P_{RW_a} and P_{LR_a}, through numerical estimation rather than by calculating precise solutions. The functional forms for these two probability functions are not readily analyzable, but numeric estimates of the functions show a number of interesting properties.

Fig. 1 shows P_{RW_a}, the probability of infection via long-range ties on a small world network, as a function of the number of infected nodes I for a particular set of parameters: average degree $k=48$, threshold $a=2$, rewiring probability $p=.1$, and number of nodes $N=40{,}000$. There are two important features to note. First, there is an inflection point in P_{RW_a} as it goes from ≈0 to ≈1. This inflection point happens early in the contagion process, with between 10 (or .025%) and 100 (or .25%) nodes infected. Second, there is a rapid drop-off in P_{RW_a} for very high values of I, when almost all nodes are infected. Between the inflection point and the drop-off, the value of P_{RW_a} is very close to 1.

We find that the pattern evident in Fig. 1 persists across the ranges of p, a, and N. Below, we will explore the dynamics of P_{RW_a} for these parameters, but here we focus on the only parameter that changes in the above plot, which is I. This parameter is present in P_{RW_a} in two places: $P_{NIa}(c)$ and the exponent $N-I$ (see Eqs. (1) and

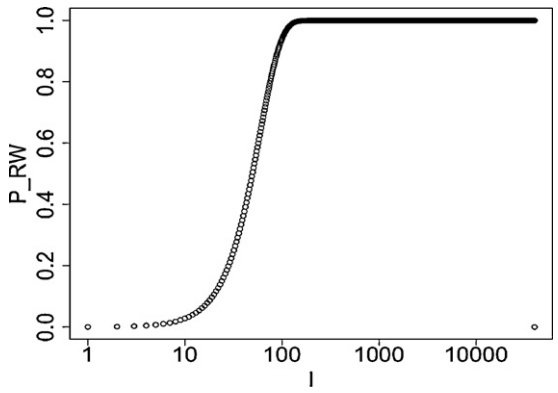

Fig. 1. Inflection and drop-off points in the probability of a rewired ties to an infected node as the level of infection increases on a rewired lattice, $a = 2$, $k = 48$, $p = 0.1$, $N = 40{,}000$.

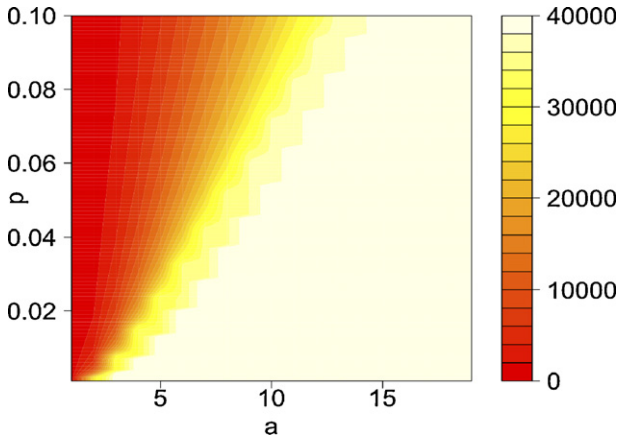

Fig. 2. Critical mass increases with threshold a and decreases with perturbation p on a rewired lattice, $k = 48$, $N = 40{,}000$. Colors indicate critical mass from red ($CM = 1$) to white ($CM = N$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

(2)). As I increases, $P_{NIa}(c)$ decreases, causing an increase in P_{RW_a}, but the exponent $N - I$ decreases as well, causing a decrease in P_{RW_a}. Therefore, at the inflection point the change to $P_{NIa}(c)$ outweighs the change to the exponent, while at the drop-off near $I \approx N$, the reverse happens. In other words, at the inflection point the likelihood of some uninfected node having a ties to infected nodes becomes so high that the smaller pool of uninfected nodes does not bring it down. At the drop-off, the pool of uninfected nodes becomes so small that the very high likelihood of some uninfected node having a ties to infected nodes does not bring the value of P_{RW_a} up.

The results of our numerical estimation suggest that the infection process is self-sustaining between the inflection point and the drop-off. In this region, each additional infected node adds more long-range ties between infected and uninfected nodes, and makes further adoption via long-range ties more likely. The beginning of this region corresponds to a phase transition where the contagion goes from spreading exclusively via short-range ties (because P_{RW_a} is near 0) to spreading via both long- and short-range ties (because P_{RW_a} is near 1). This analysis suggests that I^*, the number of infected nodes at the inflection point in Fig. 1, corresponds to a critical mass in the size of the infected population, above which a complex contagion can leverage long-range ties with a sufficiently high probability for propagation via long-range ties to become self-sustaining (limited only by the declining pool of nodes that remain uninfected).

We now focus on the critical mass phenomenon and explore its values for rewired lattice networks for a parameter space of different thresholds a and rewiring levels p. Fig. 2 below is a heat map that shows critical mass values for a range of values of threshold a (x axis) and rewiring probability p (y axis), holding average degree k constant at 48 and number of nodes N constant at 40,000. Colors of the contour plot correspond to values of the critical mass: red colors indicate low values, yellow values indicate intermediate values, white colors indicate high values.

Fig. 2 shows how, holding p constant, low thresholds yield a smaller critical mass, and so a faster rate of spread, than high thresholds. For any $p > 0$, as a increases, the probability for an uninfected node to have a rewired ties to infected nodes necessarily decreases. For simple contagions ($a = 1$), the critical mass has its minimal value (1 infected node) regardless of p. Conversely, for contagions with very high threshold ($a \geq 15$), the critical mass has its maximum value N regardless of p. For intermediate thresholds, the critical mass decreases in p. Intuitively, the more ties that are rewired, the higher the probability that a node will have a rewired ties to infected nodes, hence a given expected value requires fewer infected nodes.

We conclude this section by replicating our analysis of P_{RW_a} on the probability of infection via long-range ties for power-law networks, P_{LR_a}. As power-law networks have no rewiring, we add a new parameter r to represent the ratio SI/SN, which models the degree of the infected nodes relative to that of the rest of the population. Formally:

$$r = \frac{SI/SN}{I/N} \tag{34}$$

where SI is the sum degree of infected nodes, SN is the sum degree of all nodes, I the number of infected nodes and N the total number of nodes in the network. Fig. 3 shows the results for a particular combination of parameter settings, $a = 2$, $N = 40{,}000$, $\alpha = 5$ and $r = 1$ (infected nodes have the same average degree as all nodes). We find the same overall pattern for P_{LR_a} as for P_{RW_a} with an even sharper transition (which appears as a step function due to floating-point precision limitations).

In Fig. 4, we explore values of the critical mass (defined as the number of infected nodes at the inflection point in P_{LR_a}) for a range of values of threshold a (x axis) and degree ratio r (y axis), keeping the number of nodes N constant at 40,000 and the power law exponent α constant at 2. Fig. 4 shows that, for a given threshold, complex contagions on power law networks have a much smaller critical mass than complex contagions on rewired lattices. That is because the higher variance in degree in a power law network, relative to a small-world network, even with the same mean degree

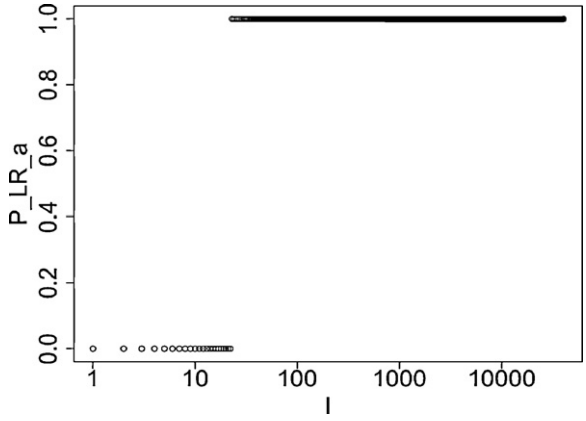

Fig. 3. Inflection and drop-off points in the probability of a rewired ties to an infected node as the level of infection increases on a power law network, $a = 2$, $r = 1$, $N = 40{,}000$, $\alpha = 2$.

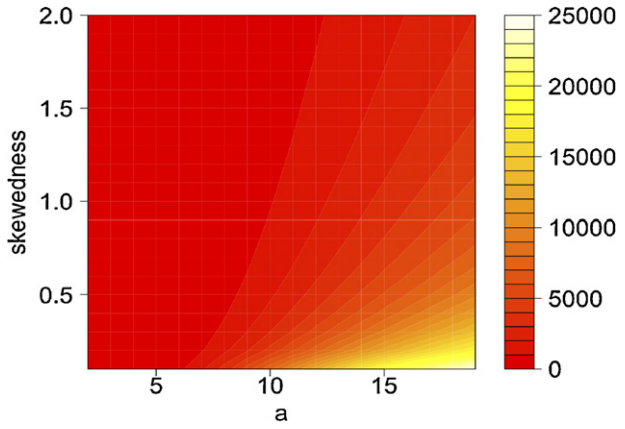

Fig. 4. Critical mass increases with threshold a and decreases with degree ratio r on a power law network, $N = 40{,}000$, $\alpha = 2$. Colors indicate critical mass from red ($CM = 1$) to white ($CM = N$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

($k = 8$) makes it more likely that an uninfected node will have a ties to a few hubs that have become infected.

Finally, it is important to consider the distribution of nodes in the infected cluster throughout the network with respect to critical mass. Given that the calculation of critical mass depends on long-range ties from uninfected nodes to infected nodes, the distribution of nodes in the infected cluster is relevant only with respect to the distribution of ties between infected and uninfected nodes.

In the case of small-world networks, long-range ties are distributed randomly throughout the network. Therefore, any distribution of infected nodes which is independent of the distribution of long-range ties is equally likely to reach critical mass. In the cases of a contiguous or random distribution, this condition holds, as neither a contiguous distribution of nodes nor a random distribution of nodes is dependent on the distribution of long-range ties. The infected cluster would have to be specifically distributed to avoid (or include) long-range ties to have an effect on the critical mass point. As an extreme example of an infected cluster that is distributed to avoid long-range ties, consider a contagion model where propagation is dependent on the number of rewired ties in nodes. Let this model be equivalent to the complex contagion model described in Section 2, but nodes change their state from uninfected to infected if and only if they have sufficient infected neighbors and none of their ties are rewired. In that case, the contagion would never reach critical mass because the infected cluster would never contain nodes with rewired ties.

In the case of power-law networks, long-range ties, like all ties, are targeted disproportionately at high-degree nodes. This means that the distribution of infected nodes throughout the network with respect to critical mass depends mainly on the degree of these nodes. An infected cluster that consists of hubs will reach critical mass earlier than an infected cluster that consists of low-degree nodes. Hubs also tend to be more central in power-law networks, so the infected clusters in the core of the network (short average path length to other nodes) will reach critical mass faster than infected clusters in the periphery (long average path length to other nodes).

3.5. Absolute and relative thresholds

The preceding analysis assumed absolute thresholds. In this section, we briefly consider relative thresholds z/n. In the relative threshold model, a fraction z/n of a node's neighbors must become infected for the node to switch its state to infected. For degree-regular networks such as the rewired lattice, there is no difference between absolute thresholds a and relative thresholds z/n with respect to analysis of contagion dynamics. For networks with a non-uniform degree distribution, such as the power-law network, using a relative threshold makes the analysis more complex, but it is still possible to make broad observations about the dynamics of the contagion. First, we have to rewrite the probability of infection via long-range ties P_{LR_a} to incorporate the relative threshold z.

Lemma 3.3. *Given a power-law network of N nodes with sum degree SN where degree follows a power-law distribution and ties are formed according to preferential attachment, and I infected nodes with sum degree SI on that network, the probability that for some uninfected node c, a fraction z of c's n ties is to infected neighbors, is approximated by:*

$$P_{LR_z} \approx 1 - \prod_{n \in deg(N)} \left[1 - \left(\frac{SI}{SN}\right)^{\lceil zn \rceil}\right] \binom{n}{\lceil zn \rceil}^{n^{-\alpha}} \tag{35}$$

where deg(N) is the set of all distinct degree values in the network and α the power law exponent.

Proof. The proof builds upon the derivation of P_{LR_a}. As noted in Section 3.2, the quantity $P_{NIa}(c)$ (Eq. (16)) is uniform across all nodes with the same degree. For nodes with a particular degree n, we can rewrite Eq. (16) as:

$$P_{NIz}(c) \approx 1 - \left(\frac{SI}{SN}\right)^{\lceil zn \rceil} \tag{36}$$

For each value of n, the number of nodes with degree n in a power-law network is $n^{-\alpha}$. Finally, for each of those nodes, the number of ways to choose $\lceil zn \rceil$ nodes out of n nodes is:

$$\binom{n}{\lceil zn \rceil}$$

Plugging these terms into the overall form for P_{RW_a} (see Eq. (3)) we get:

$$P_{LR_z} \approx 1 - \prod_{n \in deg(N)} \left[1 - \left(\frac{SI}{SN}\right)^{\lceil zn \rceil}\right] \binom{n}{\lceil zn \rceil}^{n^{-\alpha}} \tag{37}$$

□

To analyze P_{LR_z}, consider the relationship between $P_{NIz}(n)$ and n. This probability has three terms, which govern its dynamics. The first is $\binom{n}{\lceil zn \rceil}$, which follows the inequality:

$$z^{\lceil zn \rceil} \leq \binom{n}{\lceil zn \rceil} \leq (ez)^{\lceil zn \rceil} \tag{38}$$

By the inequality, the first term is exponentially increasing in node degree n. The second term is $(n^{-\alpha}$, which is polynomially decreasing in n. The third term is $1 - \left(\frac{SI}{SN}\right)^{\lceil zn \rceil}$, which approaches 1 at an exponential rate in n. On balance, these terms indicate that $P_{NIz}(n)$ will approach 1 at a polynomial rate in n. Furthermore, as we discussed in Section 3.2, individual degree values will be roughly powers of the exponent α. These two factors in combination suggest that elements of $P_{NIz}(n)$ for large n will be very close to 1. That is, the very small number of nodes with high degree and the very large number of neighbors required to infect them will ensure that hubs of the network are nearly immune to infection.

It follows that the success or failure of contagions with relative thresholds depends decisively on the infection of low-degree nodes. There are two factors at play: the choice of seed cluster (hub vs. non-hub) and the interconnectivity between low-degree nodes. If the contagion seed cluster does not include a hub, then a very high

level of interconnectivity among the low-degree nodes is needed to ensure successful propagation. This observation parallels a well-known result by Morris (2000) – that behaviors with a high relative threshold will spread best through local neighborhoods with a high degree of clustering. Even if one of the seed nodes is a hub, and the network contains multiple hubs, the contagion can only reach these uninfected hubs by spreading through low degree nodes that are sufficiently clustered to propagate a complex contagion [1]. The fewer hubs in the seed cluster, the greater the clustering that is needed among the low-degree nodes. A single infected hub may put its low-degree neighbors over the threshold, but unless those neighbors can infect further nodes, the contagion will not spread. We leave a more detailed analysis of relative-threshold contagions to future investigation.

4. Thresholds and contagion dynamics

We turn now from the identification of critical mass to the consequences for the propagation dynamics of complex contagions, focusing on the contagion growth rate, that is, the proportional increase in the number of infected nodes as the contagion spreads throughout the network. More precisely, we define the perimeter of an infected region as the number of nodes about to be infected given I nodes already infected with the contagion, the area as the total number of nodes already infected by the contagion, and the growth rate as the size of the perimeter relative to the area.

Definition 1. Given a contagion with threshold a (or z/n) contagion and I infected nodes, the *perimeter* $x(I)$ of that infected set is the number of uninfected nodes that have a or more (or z/n or more) infected neighbors.

The growth rate of a contagion over time is the ratio of the number of nodes about to be infected (the perimeter) to the number of nodes already infected, expressed as a function of I.

Definition 2. The growth rate of a contagion $\lambda(I)$ is given by $x(I)/I$ as a function of I.

Consider the growth rates of a complex contagion on a perturbed lattice before and after it reaches critical mass, starting with a single infected neighborhood A. Since the lattice has uniform degree, the analysis applies equally to absolute and relative thresholds. Before critical mass, the contagion is unlikely to spread through long-range ties since it is unlikely that an uninfected node will have a random ties to infected nodes. So, the contagion will spread by leveraging the local neighbors of A. Spatially, the neighborhood A resembles a square (for a two-dimensional lattice; this analysis extends into lattices of dimension d). The local neighbors of A (i.e. those adjacent to A) form a perimeter around this square. As these neighbors in turn become infected, the size of the infected square increases, as does the perimeter around it. In general, until the contagion reaches critical mass, the infected set will always form a square with area I^2 and its perimeter will always form a square perimeter of size $O(I)$. then the growth rate prior to critical mass, $\lambda_{pre}(I)$ is given by

$$\lambda_{pre}(I) \approx \frac{I}{I^2} \approx I^{-1} \tag{39}$$

where I is the number of infected nodes. Thus the growth rate prior to critical mass drops quickly as I increases, for the simple reason that the perimeter of a square becomes smaller relative to its area as the area increases.

After reaching critical mass, the picture is dramatically different. The perimeter of the contagion is not limited by the lattice structure, but also includes the expected number of nodes infected via rewired ties. Spatially, the infected set now consists of the infected square containing A plus the set of randomly distributed nodes with a random ties to infected nodes. As the size of the infected square continues to grow, one or more of these randomly infected nodes may eventually be able to help infect one of its local neighbors, and a second infected square will emerge which grows locally, as does the perimeter around it. And then a third square, and so on, each happening more quickly than the last, given the increasing overall number of infected nodes and thus the increasing probability that an uninfected node will have random ties to a infected neighbors. More formally, consider the quantity $1 - P_{NI}(c)$, the probability of some particular node c on a rewired lattice having a or more infected neighbors:

$$1 - P_{NI}(c) \approx 1 - \left[1 - \left(\frac{pI}{N}\right)^a\right] \binom{k}{a} \tag{40}$$

where p is the rewiring probability, I the number of infected nodes, N the total number of nodes, a the threshold, and k the degree uniform across all nodes. We see that $1 - P_{NI}(c)$ is a monotonically increasing polynomial function of I. Now consider the quantity NIc, the expected number of nodes that have a or more infected neighbors. By an argument similar to that stated in Theorem 3.1, $P_{NI}(c)$ is not exactly uniform over all c as it tends to 1 as the set of available targets (infected neighbors) is exhausted. The expected number of nodes with a or more infected neighbors will, then, be a sum of terms dominated by $1 - P_{NI}(c)$. The total number of such terms will be $N - I$, the number of nodes remaining uninfected. We can now write the post-critical perimeter as follows:

$$x(I) = O(f(I^a)(N - I)) \tag{41}$$

where I is the number of infected nodes and $f(I^a)$ is some monotonically increasing polynomial function of I (i.e. $1 - P_{NI}(c)$). Then, the growth rate is given by:

$$\lambda(I) = O(f(I^{a-1})(N - I)) \tag{42}$$

This is a product of two functions, one monotonically increasing in I, the other monotonically decreasing in I. At first, $\lambda(I)$ is dominated by the first term (many uninfected nodes available to infect through rewired ties), and grows in I. As the number of uninfected nodes declines, the second terms begins to dominate and $\lambda(I)$ falls in I. In summary, the growth rate of a complex contagion on a rewired lattice at first drops quickly in I but then, if the contagion reaches critical mass, the growth rate suddenly takes off and increases in I and then once again drops in I as the contagion runs out of nodes to infect.

The analysis in this section helps shed light on the conservative nature of the P_{RW_a} estimator derived in Eq. (1). That estimator is based on the assumption that nodes are infected entirely via long-range ties, whereas in simulation nodes can be infected via a combination of short-range and long-range ties. The extent to which P_{RW_a} is a conservative estimator can be approximated by the growth rate of the infected set in simulation, λ_s, and the estimated perimeter of the infected set, which considers only long-range ties, λ_e. The first quantity is governed by the dynamics of $\lambda_{pre}(I)$ (Eq. (39)) and $\lambda(I)$ (Eq. (42)). The second quantity is governed solely by the dynamics of $\lambda(I)$. At first, $\lambda_s > \lambda_e$ as $\lambda(I)$ is close to 0. This difference is small, however, as $\lambda_{pre}(I)$ is rapidly shrinking to 0. Next, the increased number of infected nodes from short-range ties means that the simulated contagion reaches critical mass earlier, so once again $\lambda_s > \lambda_e$. However, at this point the difference between the simulated and estimated contagion growth rates stays more or less

[1] To the extent that hubs are themselves more likely to have ties to other hubs (degree homophily), the critical number of hubs that needs to be activated is reduced.

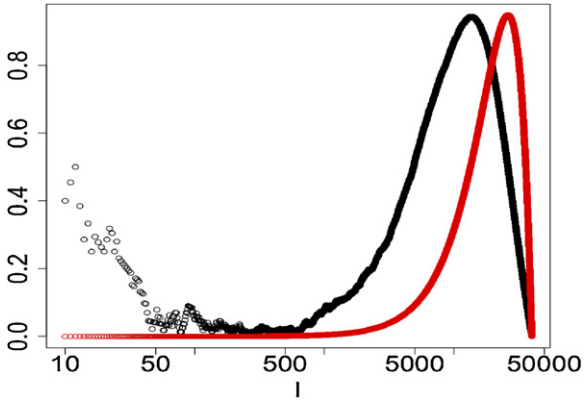

Fig. 5. Contagion growth rate (black) and probability that a random node will be uninfected and have a random ties to infected nodes (red), for a regular lattice with $k=8$, $a=2$, $N=40{,}000$, $p=.1$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

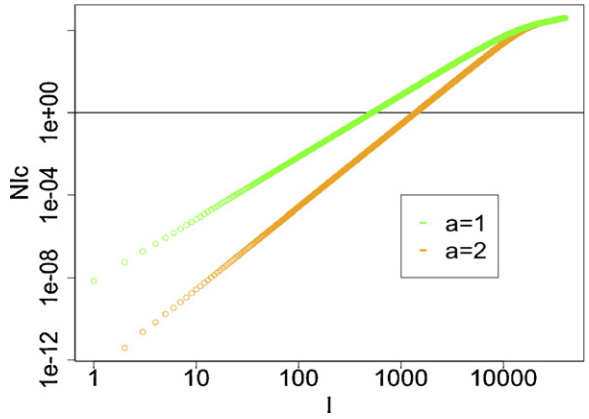

Fig. 6. *NIc* (logged) as a function of the log of the area on a perturbed lattice, $k=48$, $p=.1$, $N=40{,}000$, and $a=2$ (green) and $a=3$ (brown). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

constant, as the growth rate from short-range ties is close to zero. Finally, the earlier onset of critical mass for the simulated contagion means an earlier onset of the point when the contagion runs out of nodes to infect. At this point, the difference between λ_s and λ_e rapidly shrinks and the estimated growth rate may at some point exceed the simulated contagion growth rate.

We demonstrate this pattern empirically by plotting the observed growth rate (renormalized as $\frac{\hat{f}(I)}{I}$ so the minimal growth rate is 0) against I for a simulated contagion with degree $k=8$, threshold $a=2$, rewiring probability $p=.1$, and number of nodes $N=40{,}000$ (Fig. 5). The growth rate, shown in black in Fig. 5, goes through three phases – first a rapid drop, then a sharp rise, followed by another drop. Fig. 5 also shows (in red) the corresponding quantity $(1-P_{NI}(c))(N-I)$ (based on Eq. (41)), which shows the change in the probability of infection through long-range ties. Note that, as described in the previous paragraph, the red line first lags behind the black, and reaches critical mass later (between 1000 and 2000 nodes infected instead of between 500 and 1000); however, as the contagion begins to run out of nodes to infect, the red line overtakes the black and at the very end of the simulation the estimator (which is derived from P_{RW_a}) is actually overly optimistic about the growth rate of the contagion.

5. Beyond the threshold model

Following Centola and Macy, the preceding analyses assume discrete and deterministic thresholds of adoption. However, empirical contagions may be more plausibly modeled as continuous stochastic decisions rather than thresholds. Previous studies (Backstrom et al., 2006; Leskovec et al., 2006; Valente, 2005) show that having multiple adopter friends does increase the likelihood of adoption, sometimes in a non-linear way. We can formalize this relationship as follows:

$$P(\text{adopt}|a\ \text{adopterfriends}) = f(a) \tag{43}$$

where f is a monotonically increasing function. For small a (region around $a=2$ in Backstrom et al. (2006)), f is convex. For large a, f is concave, with each additional adopter friend contributing a diminishing marginal likelihood of adoption.

We can use this formalism to adapt the analytical results above to contagions that do not have a deterministic threshold, but do have a positive relationship between likelihood of adoption and number of adopter friends (that is, the threshold is stochastic rather than deterministic). As in the previous section, we focus on the quantity *NIc*, the expected number of uninfected nodes that have a or more infected neighbors. We can simulate the behavior of *NIc* over values of number of infected nodes I and threshold a. Fig. 6 shows the log of *NIc* (y axis) as a function of the log of I (x axis) on a rewired lattice with degree $k=48$, rewiring probability $p=.1$ and number of nodes $N=40{,}000$ and thresholds $a=2$ (green) and $a=3$ (brown).

Fig. 6 shows that for values I below 10,000 (1/4 of the nodes infected), *NIc* is an exponential function that appears linear on a logged axis. We can also see that the difference between the number of nodes that have 2 or more infected neighbors, and 3 or more infected neighbors, diminishes exponentially and disappears for $I>10{,}000$. This general pattern applies across the range of N, k and p, but for smaller values of k (holding all other parameters constant), the lines for $a=2$ and $a=3$ do not converge before the pool of uninfected nodes is exhausted.

The dynamics of *NIc* as shown in Fig. 6 suggest an important property of the diffusion of stochastic threshold contagions across regular networks: as a contagion leverages long-range ties, the number of nodes that are exposed to a and not $a+1$ infected neighbors shrinks exponentially. Given the relationship between a and likelihood of infection given by f above, this means that for small values of a, the likelihood of infection increases rapidly in the early stages of contagion diffusion. Larger values of a yield a smaller marginal likelihood, and do not need to be considered as closely. However, predictive models of contagion adoption that take into account f should outperform models that ignore the marginal likelihood.

As an illustrative example, we consider the same network as above (rewired lattice, $p=.1$, $k=48$, $N=40{,}000$), and calculate the expected number of adopters at the next step for a given number of current adopters based on two different functions f. The first function, $f1$, will be linear in a, the second function, $f2$, will be non-linear in a. For simplicity, we consider only thresholds up to $a=3$. The specific forms of $f1$ and $f2$ are as follows:

$$f1 = ((1, .02), (2, .03), (3, .04))$$
$$f2 = ((1, .02), (2, .05), (3, .07))$$

Then we can calculate the expected number of adopters by determining the number of uninfected nodes that have a neighbors and multiplying by the appropriate value of $f(a)$. We also include a "baseline" expected number of adopters that is based solely on $f(1)$, ignoring the marginal likelihood of adoption due to multiple exposures. The results are summarized in Fig. 7.

For $I<1000$ and $I\approx N$, the expected number of adopters is the same whether using $f1(1)$, all of $f1$, or $f2$. In the intermediate range

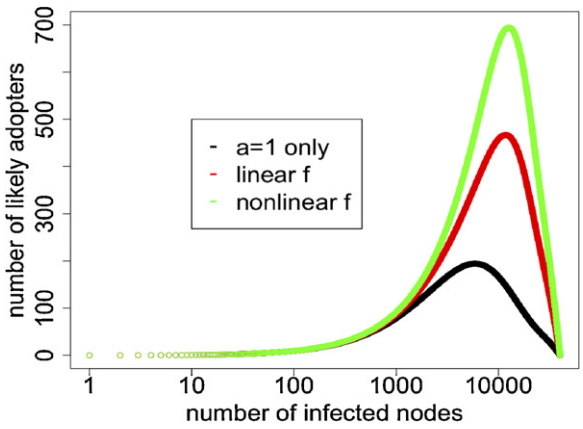

Fig. 7. Expected number of adopters as a fiunction of I using only $f1(1)$ (black), all of $f1$ (red), and $f2$ (green), $N = 40{,}000$, $k = 48$, $p = .1$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

$I \in [1000, N)$ (2.5% and 25% of all nodes infected), there is a significant difference in new adopter expectation. At the peak, around $I = 10{,}000$ nodes, the "baseline" expectation is half as much as the expectation that uses a linear f, and a third as much as the expectation that uses a non-linear f.

6. Discussion and conclusion

The key contribution of this paper is the demonstration of a bifurcation point in the spread of complex contagions—the critical mass. For simple contagions like information and disease, this bifurcation does not exist. Such contagions can leverage long-range ties even with only one infected node, hence a single seed is sufficient to create a critical mass.

For complex contagions, in contrast, the growth process will have two phases separated by a very sharp transition. Initially, the contagion can only spread locally, that is, via short-range ties. Once every node that is reachable via short-range ties is infected, propagation terminates if the level of infection remains sub-critical.

However, if the region reachable via local propagation is sufficiently large, the contagion will reach critical mass and the contagion will "go viral." On this side of the bifurcation point, the contagion can now spread via long-range ties that allow the contagion to "jump" to a fresh area. This area of infection then rapidly expands through short-range ties. The increase in the size of the infected population in turn increases the probability that the contagion can spread to yet another fresh region of the network via long-range ties. Simply put, once a complex contagion reaches critical mass, it begins to spread in the same way as a simple contagion—taking advantage of shortcuts to distant regions and eventually reaching every node in a connected network

The existence of a bifurcation point points to the non-linearity of the complex contagions model. Bifurcation occurs when a "small smooth change made to the parameter values (the bifurcation parameters) of a system causes a sudden 'qualitative' or topological change in its behaviour" (Blanchard et al., 2006). This sudden change is not a sign of degeneracy or model simplicity: both the analysis in Centola and Macy (2007) and this work show that the complex contagion model is capable of generating highly nonlinear dynamics of contagion diffusion that lead to widespread contagion in some cases, and failure of the contagion to spread beyond its seed in others.

Our analysis also has an important theoretical implication for understanding why some contagions "go viral" and others do not. For a simple contagion to escape the region of initial infection, it need only reach a node with a long-range tie or a hub that can broadcast the contagion more widely. On a small world or undirected power law network, that is guaranteed to eventually occur, so long as the contagion remains capable of passing from one node to another (e.g. there is no decline in infectiousness such as might happen in a news cycle). For a complex contagion to go viral, it must infect sufficient nodes that a long-range tie or susceptible hub can make a difference. There has not been, to our knowledge, any empirical analysis of the Centola–Macy complex contagions model to explain the success or failure of virally marketed products or other contagions. We encourage future work in this area to apply the model in an empirical setting.

The existence of a bifurcation point in the propagation of complex contagions has a potentially valuable practical implication for the ability to predict the eventual outcome at the early stages of a viral marketing campaign. Prior to critical mass, the growth rate decays, but as soon as the contagion is able to spread via long-range ties, the growth rate reverses and rapidly accelerates. This qualitative change in the rate of growth from negative to positive is a statistical signature of critical mass.

Another early indicator that the contagion has gone viral is if the contagion is spotted in a fresh region of the network, not contiguous with the seed area. Because propagation via long-range ties is self-sustaining, the first occurrence can be a useful indicator that the contagion has reached critical mass and will therefore continue to spread via every available tie, regardless of its range. The caveat is that a false positive can also be caused by the homophilous clustering of nodes with low thresholds ("early adopters"), but even in the case where a jump simply indicates an expansion to a set of "early adopters," the consequent increase in the number of infected nodes in turn increases the likelihood of a true positive.

This analysis also has practical implications for marketing strategy. The greater the need for social reinforcement to persuade individuals to adopt an innovation, the larger the size of the initial region of local propagation required for the contagion to go viral. Thus, in deciding where to launch an innovation, the proportion of highly susceptible nodes in the initial region is less important than the overall size, so long as the nodes are sufficiently susceptible that the contagion can spread through short-range ties.

The need for a critical mass also carries implications for initial pricing. For simple contagions, it may be optimal to set prices initially high, in order to maximize profits from the most interested customers (early adopters). In contrast, for complex contagions, it is better to set prices initially low to improve the chances that the contagion will reach critical mass.

We note these implications of the existence of a critical mass not as policy recommendations but as suggested directions for theoretical and empirical research. Our analysis assumes highly stylized topologies composed of nodes with homogenous attributes. Much more research is needed before we can have confidence in the predicted existence of a bifurcation point in the propagation of complex contagions. Residential neighborhoods, college dormitories, and soccer stadia may loosely resemble a regular lattice, and other empirical social networks have been shown to have degree distributions that approximate a power law. Yet other social networks have degree distributions that are more irregular than a lattice and less skewed than a power law. In addition, the nodes in empirical networks have heterogenous attributes, thresholds that vary between innovators, early adopters, and laggards (Berry and Keller, 2003; Katz and Lazarsfeld, 1955), and influence that varies between influentials, opinion leaders, and followers (Rogers, 2003). Moreover, these attributes may be homophilously clustered. These complications preclude the ability to use a set of formal results to confidently predict the critical mass in natural settings, and systematic empirical research is needed to see if the predicted existence of a bifurcation point is observed in empirical social networks. The important contribution of the present study is not that we have

settled the question but quite the opposite—the predicted existence of a critical mass opens up an important direction for both theoretical and empirical research on why some contagions "go viral" and others die.

References

Albert, R., Barabási, A.-L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74 (1), 47–97.

Anderson, R.M., May, R.M., 1991. Infectious Diseases of Humans. Oxford University Press, Oxford.

Backstrom, L., Huttenlocher, D., Kleinberg, J., Lan, X., 2006. Group formation in large social networks: membership, growth, and evolution. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Berry, J., Keller, Ed., 2003. The Influentials: One American in Ten Tells the Other Nine How to Vote, Where to Eat, and What to Buy. Simon and Schuster, NY.

Blanchard, P., Devaney, R., Hall, G.R., 2006. Differential Equations. Thompson, London.

Centola, D., 2010. The spread of behavior in an online social network experiment. Science 329, 1194–1197.

Centola, D., Macy, M., 2007. Complex contagions and the weakness of long ties. Am. J. Sociol. 113, 702–734.

Granovetter, M., 1978. Threshold models of collective behavior. Am. J. Sociol. 83 (6), 1420–1443.

Katz, E., Lazarsfeld, P.F, 1995. Personal influence: the part played by people in the flow of mass communications. ISBN 1412805074 (new edition), p. 309ff.

Leskovec, J., Adamic, L., Huberman, B., 2006. The dynamics of viral marketing. In: Proceedings of the 7th ACM Conference on Electronic Commerce.

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

Rogers, E.M., 2003. Diffusion of Innovations, 5th ed. Free Press, New York.

Romero, D.M., Meeder, B., Kleinberg, J., 2011. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter, in Intl. WWW Conference, (Hyderabad, India, 2011), ACM, 695–704.

Schelling, T., 1971a. Dynamic models of segregation. J. Math. Sociol. 1 (1), 143–186.

Steglich, C., Snijders, T.A.B., Pearson, M., 2010. Dynamic networks and behavior: separating selection and influence. Sociol. Methodol. 40, 329–393.

Valente, T.W., 2005. Network Models and Methods for Studying the Diffusion of Innovations. Cambridge University Press.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440–442.