

Complex Contagions on Configuration Model Graphs with a Power-Law Degree Distribution

Grant Schoenebeck and Fang-Yi Yu^(✉)

University of Michigan, Ann Arbor, USA
schoeneb@eecs.umich.edu, fayu@umich.edu

Abstract. In this paper we analyze k -complex contagions (sometimes called bootstrap percolation) on configuration model graphs with a power-law distribution. Our main result is that if the power-law exponent $\alpha \in (2, 3)$, then with high probability, the single seed of the highest degree node will infect a constant fraction of the graph within time $O\left(\log^{\frac{\alpha-2}{3-\alpha}}(n)\right)$. This complements the prior work which shows that for $\alpha > 3$ bootstrap percolation does not spread to a constant fraction of the graph unless a constant fraction of nodes are initially infected. This also establishes a threshold at $\alpha = 3$.

The case where $\alpha \in (2, 3)$ is especially interesting because it captures the exponent parameters often observed in social networks (with approximate power-law degree distribution). Thus, such networks will spread complex contagions even lacking any other structures.

We additionally show that our theorem implies that $\omega\left(n^{\frac{\alpha-2}{\alpha-1}}\right)$ random seeds will infect a constant fraction of the graph within time $O\left(\log^{\frac{\alpha-2}{3-\alpha}}(n)\right)$ with high probability. This complements prior work which shows that $o\left(n^{\frac{\alpha-2}{\alpha-1}}\right)$ random seeds will have no effect with high probability, and this also establishes a threshold at $n^{\frac{\alpha-2}{\alpha-1}}$.

1 Introduction

Social behavior is one of the defining characteristics of us as a species. Social acts are influenced by the behavior of others while influencing them at the same time. These interactions have been observed in a wide array of activities including financial practices [8, 14], healthy/unhealthy habits [23], and voting practices [1]. Some of these are beneficial (e.g., adopting a healthy lifestyle) or profitable (e.g., viral marketing), while others are destructive and undesirable (e.g., teenager smoking, drug abuse).

To effectively promote desirable contagions and discourage undesirable ones, the first step is to understand how these contagions spread in networks and what are the important parameters that lead to fast spreading.

G. Schoenebeck—Supported by National Science Foundation Career Award #1452915

F.-Y. Yu—Supported by National Science Foundation Algorithms in the Field Award #1535912.

The high level objective is to understand how these behaviors spread in a social network. Two key factors determine the scope and rate of such diffusion: the model of contagions, i.e., how a node is influenced by its neighbors; and the network topology.

The copying of behaviors leading to a social cascade of behavioral changes is attributed to two effects: the informational benefit (inferring hidden, private information others may know) and direct benefit effects (resulting from coordinated actions or social pressure). In the *threshold model* [19], introduced by Granovetter, each agent has a *threshold* and when an agent's number of infected neighbors reaches her threshold, then she adopts the cascade.

We deal with a simplified version of this model where all agents have the same threshold. This is called *k-complex contagion* [18] or *bootstrap percolation* – the latter is generally used in the physics community where it was originally studied in the context of magnetic disordered systems [2, 13], but has since been applied to describe several complex phenomena including neuronal activity and the dynamics of the Ising model at zero temperature. In the context of social networks, bootstrap percolation provides a model of complex contagions [12] which model for the spread of ideas, beliefs, and behaviors.

A *k-complex contagion* is a deterministic process on a graph G that evolves in rounds. In each round every node has two possible states: it is either infected or uninfected. The network begins with a seed set I of infected nodes. In each subsequent round every uninfected node become infected if it has at least k edges incident on infected neighbors, otherwise it remains uninfected. Once a node has become infected, it remains infected forever.

A key trait of *k-complex contagions* is that they are not “submodular”. This implies that the marginal influence of an additional neighbor is not decreasing. While many cascade models, such as the Independent Cascade model and the Linear Threshold model have the submodularity property [21], many real-world cascades seem not to. Non-submodular contagions are observed by sociologists in the case of the adoption of pricey technology innovations, the change of social behaviors, and the decision to participate in a migration, etc [15, 22], and by data scientists on LiveJournal [6], DBLP [6], Twitter [26], and Facebook [28]. An additional confirmation is crucial, suggesting the model of complex contagion.

Janson et al. [20] show that *k-complex contagions* do not spread on sparse $G(n, p)$ random graphs. Such cascades require $\Omega(n)$ seeds to infect a constant fraction of vertices. [7] extended these results to configuration model graphs with regular degree distributions.

However, many networks do not have regular degree distributions. In a graph with power law degree distribution, the number of nodes having degree d is proportional to $1/d^\alpha$, for a positive constant α . In 1965, Price [25] showed that the number of citations to papers follows a power law distribution. Later, studies of the World Wide Web reported that the network of webpages also has a power law degree distribution [9, 11]. Observations of many different types of social networks also found power law degree distributions, as well as biological, economic and semantic networks [3, 24, 27].

Additional work by [4] studies the configuration model with power-law degree distribution for $\alpha > 3$ and showed and shows theorem which implies (see Sect. 7) that, with high probability, infecting a constant fraction of the nodes requires an initial seed that comprises a constant fraction of the graph.

Intuitively complex contagions spread well in the presence of additional community structure, and several networks with such structure have been analyzed including the Watts-Strogatz model [18], the Kleinberg Small World graph [17], and the preferential attachment graph [16].

Amini and Fountoulakis [5] also have examined the Chung-Lu model with power-law exponent $2 < \alpha < 3$. They show that there exists a function $a(n) = o(n)$ such that if the number of initial seeds is $\ll a(n)$, the process does not evolve w.h.p.; and if the number of initial seeds is $\gg a(n)$, then a constant fraction of the graph is infected with high probability. However, this function is still super-constant— $n^{\Omega(1)}$.

The question remained open, can non-submodular cascades spread and spread quickly from a constant-sized seed set on sparse graphs with no other structure imposed besides a skewed degree distribution.

1.1 Our Contributions

Our main result is that for a configuration model graph with power-law exponent $\alpha \in (2, 3)$, with high probability, the single seed of the highest degree node will infect a constant fraction of the graph within time $O(\log^{\frac{\alpha-2}{3-\alpha}}(n))$. This complements the prior work which showed that for $\alpha > 3$ boot strap percolation does not spread to a constant fraction of the graph unless a constant fraction of nodes are initially infected. This also establishes a threshold at $\alpha = 3$.

The case where $\alpha \in (2, 3)$ is especially interesting because it captures the exponent parameters often observed in social networks (with approximate power-law degree distribution). Thus, such networks will spread complex contagions even lacking any other structure.

We additionally show that our main theorem implies that $\omega(n^{\frac{\alpha-2}{\alpha-1}})$ random seeds will infect a constant fraction of the graph within time $O(\log^{\frac{\alpha-2}{3-\alpha}}(n))$. This complements the prior work which shows that $o(n^{\frac{\alpha-2}{\alpha-1}})$ random seeds will have no effect with high probability. This also establishes a threshold at $n^{\frac{\alpha-2}{\alpha-1}}$.

To prove these results, we provide new analysis that circumvents previous difficulties. While our results are similar to those of [16] (they study the preferential attachment model, while we study the configuration model), the techniques required are completely different. For example, it is an easy observation that k -complex contagions spread on the configuration model (if k is greater than the minimum degree), but much more difficult to show it spreads quickly.

The previous analyses on the configuration model required that the graph was locally tree-like, an assumption that fails in our case, and then were able to approximate the process using differential equations and obtain rigorous results by applying Wormald's Theorem [30]. However, their analysis fails when the degree distribution is power-law with exponent between 2 and 3.

2 Preliminaries

Let $[m] := \{1, 2, \dots, m\}$. We say an event would be true *with high probability* if its probability of being true is $1 - o(1)$. When we use $\Theta(1)$, the constant may depend on various constant parameters, but should not depend on n .

Definition 1. A k -complex contagion $CC(G, k, I)$ is a contagion that initially infects vertices of $I \subseteq V(G)$ and spreads over the graph G . The contagion proceeds in rounds. At each round, each vertex with at least k edges incident on infected neighbors becomes infected. The vertices of I are called the initial seeds. Let $|CC(G, k, I)|$ denote the random variable of the final size of such a cascade.

We use the *configuration model* introduced by [10] to define a distributions over multigraphs.

Definition 2. Let $\mathbf{d} = (d_1, \dots, d_n)$ be a decreasing degree sequence where the sum of the terms is even. Define $V = [n]$ (Here we use integers $\{1, 2, \dots, n\}$ to denote the vertices, and call nodes with lower indexes “earlier”. Because the degrees decrease, earlier nodes have higher degrees). Let m be such that $2m = \sum_i d_i$. To create the m (multi-)edges, we first assign each node i d_i stubs. Next we choose a perfect matching of the stubs uniformly at random and for each pair of matched stubs construct an edge connecting the corresponding nodes.

We use $CM(\mathbf{d})$ to denote the Configuration Model with respect to the degree distribution \mathbf{d} .

2.1 Power-Law Degree Distributions

For any decreasing degree sequence $\mathbf{d} = (d_1, \dots, d_n)$ where the sum of the terms is even, we define

- the empirical distribution function of the degree distribution $F_{\mathbf{d}}$

$$F_{\mathbf{d}}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}[d_i \leq x] \quad \forall x \in [1, \infty)$$

- the fraction of nodes that have degree less than x .
- Let $N_{\mathbf{d}}(x) = n(1 - F_{\mathbf{d}}(x))$ be the number of nodes with degree at least x .
- Let $S_{\mathbf{d}}(x)$ be the number of stubs from nodes with degree at least x .
- Let $s_{\mathbf{d}}(x)$ be the number of stubs from nodes with index less than x .

We will omit the index \mathbf{d} when there is no ambiguity.

Definition 3 (Power-Law Degree Distributions). Adopting the notation of [29], we say a series \mathbf{d} has power-law distribution with exponent α if there exists $0 < C_1 < C_2$ and $0 < x_0$ such that (1) $F_{\mathbf{d}}(x) = 0$ for $x < x_0$; (2) $F_{\mathbf{d}} = 1$ for $x > d_1 = n^{2/(\alpha+1)}$, and (3) for all $x_0 \leq x \leq d_1$ then

$$C_1 x^{-\alpha+1} \leq 1 - F_{\mathbf{d}}(x) \leq C_2 x^{-\alpha+1}$$

Let \mathbf{d} have power-law distribution of *power law with exponent* α then it is easy to check that:

1. $N(x) = \Theta(nx^{-\alpha+1})$
2. $S(x) = \Theta(nx^{-\alpha+2})$.
3. $d(i) = \Theta\left(\left(\frac{n}{i}\right)^{1/(\alpha-1)}\right)$
4. $s(i) = \Theta(n^{1/(\alpha-1)}i^{\frac{\alpha-2}{\alpha-1}})$

3 Main Theorem

In this section, we state and prove our main theorem: in a configuration model graph with the power-law exponent $\alpha \in (2, 3)$, with high probability, the single seed of the highest degree node will infect a constant fraction of the graph within time $O(\log^{\frac{\alpha-2}{3-\alpha}}(n))$.

Theorem 1. *Given a power law distribution $\mathbf{d} = (d_1, \dots, d_n)$ with exponent $\alpha \in (2, 3)$ and $d_1 > n^{\frac{3-\alpha}{\alpha+1}}$, with probability $1 - O\left(\frac{\log^{\frac{\alpha-1}{3-\alpha}} n}{n}\right)$, the k -complex contagion on configuration model $CM(\mathbf{d})$ with constant k and initial infection being the highest degree node $I = \{1\}$, $CC(CM(\mathbf{d}), k, I)$, infects $\Omega(n)$ vertices within time $O(\log^{\frac{\alpha-2}{3-\alpha}} n)$.*

3.1 Proof Setup

We consider a restricted form of contagion where nodes can only be infected by those preceding them in the ordering. Formally, recall the nodes $\{d_i\}$ are ordered in terms of their degree. Node d_i will only be infected if $|\{j : j < i \text{ and } d_j \text{ is infected}\}| \geq k$ neighbors are infected. Hence, the total number of infected nodes in this process will be fewer than the number of infected nodes in original complex contagions, and it is sufficient to prove that a constant fraction of nodes become infected in this restricted contagion with high probability.

Buckets We first partition the nodes $V = [n]$ into buckets. We design the buckets to have at least (and about the same number of) stubs $b = \Theta\left(\frac{n^{\frac{\alpha-2}{3-\alpha}}}{\log^{\frac{\alpha-2}{3-\alpha}} n}\right)$.

We can define N_ℓ as follows

$$N_1 = \frac{n}{\log^{\frac{\alpha-1}{3-\alpha}} n}, \text{ and } N_{\ell+1} = \arg \min_{i > N_\ell} \{s(i) - s(N_\ell) \geq b\}$$

Since $d(N_1) = \Theta(\log^{1/(3-\alpha)} n) = o(b)$ and $\forall i > N_1, d(i) \leq d(N_1)$,

$$b < s(N_{\ell+1}) - s(N_\ell) \leq b + o(b) < 2b.$$

Therefore, we have $\ell b \leq s(N_\ell) \leq 2\ell b$ and $N_\ell = \Theta\left(\frac{n}{\log^{\frac{\alpha-1}{3-\alpha}} n} \ell^{\frac{\alpha-1}{\alpha-2}}\right)$ by (4), and so the total number of buckets is $L \leq \frac{s(n)}{b} = O(\log^{\frac{\alpha-2}{3-\alpha}} n)$.

We define our buckets to be $B_1 = \{1, \dots, N_1\}, B_2 = \{N_1 + 1, \dots, N_2\}, \dots, B_{\ell+1} = \{N_\ell + 1, \dots, N_{\ell+1}\}, \dots, B_L = \{N_{L-1} + 1, \dots, N_L\}$.

Filtration. We now state our filtration.

\mathcal{F}_0 : The node i starts with d_i stubs of edges without revealing any edges.

\mathcal{F}_1 : In the first stage we reveal all edges within the first bucket B_1 ,

$\mathcal{F}_\ell, 1 \leq \ell \leq L$: In the stage $\ell > 1$, we reveal/match all the edges from B_ℓ to early nodes in $B_{<\ell}$.

3.2 Proof Summary

There are two parts of the proof.

1. All of the nodes in the first bucket would be infected with high probability.
2. For some constant $\rho > 0$, in the first $L' = \rho L$ buckets $B_1, \dots, B_{L'}$ a constant fraction ϵ of nodes will be infected. Because $N_{L'} = \Omega(n)$ nodes, the total number of infection also constant fraction.

In the first part of the proof is capture by the following lemma:

Lemma 1 (Base). *Given at \mathcal{F}_0 $d_1 > n^{\frac{3-\alpha}{\alpha+1}}$, at \mathcal{F}_1 all the nodes in B_1 will be infected within $O(\log \log(n))$ steps with probability greater than $1 - O(\frac{1}{n})$.*

To prove this lemma we further decompose the first bucket into $O(\log \log(n))$ finer intervals, which we call bins. We first argue that every node in the first bin will have at least k multi-edges to the first node, and we inductively show the nodes in following bin will have at least k edges to the previous bins. The analysis is by straight-forward probabilistic techniques.

The time for the first bucket's infection is at most the number of the bins because inclusion of each bin only costs 1 step.

We need some additional notation to state the lemma which will imply the second part. Let X_ℓ be the number of stubs from buckets $B_{<\ell}$ to $B_{\geq\ell}$. Let Y_ℓ be the number of uninfected stubs from $B_{<\ell}$ to $B_{\geq\ell}$ before stage ℓ , of which $Y_\ell^{(1)}$ issue from $B_{<\ell-1}$ and the remaining $Y_\ell^{(2)}$ issue from $B_{\ell-1}$. We use \mathbb{I}_i as the indicator variable that node $i \in B_\ell$ is not infected after stage ℓ . Let $\epsilon > 0$ be some constant we define later. Let $\delta_n = \Theta(\frac{1}{\log^{\frac{\alpha-2}{3-\alpha}} n})$.

Now we can formally define \mathcal{A}_ℓ as the intersection of the following three events:

1. connection: $(1 - \delta_n)\mathbb{E}[X_\ell] \leq X_\ell \leq (1 + \delta_n)\mathbb{E}[X_\ell]$;
2. number of uninfected nodes: $\sum_{i \in B_{\ell-1}} \mathbb{I}_i \leq 2\mu_H$ where $\mu_H = K \frac{|B_\ell|^\frac{3-\alpha}{\alpha-2}}{\log n}$ for some constant K independent of ℓ and n ;
3. number of uninfected stubs: $Y_\ell \leq \epsilon X_\ell$.

Lemma 2 (Induction). *Fix sufficiently small $\epsilon > 0, \rho > 0$. Let $\ell < \rho L$, and suppose $\Pr[\mathcal{A}_\ell] > 0.5$, then we have*

$$\Pr[\mathcal{A}_{\ell+1} | \mathcal{A}_\ell] = 1 - O(1) \frac{(\log n)^{\frac{\alpha-1}{3-\alpha}}}{n\ell^{1/(\alpha-2)}}$$

This lemma will be proved by showing that each of three events happens with high probability conditioned on \mathcal{A}_ℓ . The most technically challenging of these is the second event, where we need to apply Chebychev’s Inequality twice. One challenging is that the edges from $B_{<\ell}$ to B_ℓ are not independent. Another challenge is that if the buckets are too small, we fail to have concentration properties, but if they are too large, then the fraction of infected nodes at each stage will drop too quickly.

3.3 Proof of Theorem 1

Proof. If $\bigcap_{\ell=1}^{L'} \mathcal{A}_\ell$ happens, then the total fraction of infected nodes is $\Omega(n)$.

Using Lemma 1 as the base case and Lemma 2 as the induction steps we see that

$$\Pr \left[\bigcap_{\ell=1}^{L'} \mathcal{A}_\ell \right] \geq 1 - \sum_{\ell=1}^{L'} O(1) \frac{(\log n)^{\frac{\alpha-1}{3-\alpha}}}{n\ell^{1/(\alpha-2)}} - O\left(\frac{1}{n}\right) = 1 - O\left(\frac{\log^{\frac{\alpha-1}{3-\alpha}} n}{n}\right)$$

which is arbitrarily close to 1.

Moreover, the total time spent is the time in first bucket plus the number of buckets (because the infection spreads from bucket to bucket in only 1 step). Therefore the total time spent is

$$O(\log \log n) + O(\log^{\frac{\alpha-2}{3-\alpha}} n) = O(\log^{\frac{\alpha-2}{3-\alpha}} n)$$

which completes our proof.

4 Proof of Lemma 1: Contagion in the First Bucket

In this section, we will show that with high probability, the contagion process infects all nodes within the first bucket. Recall that $N_1 = \frac{n}{\log^{\frac{\alpha-1}{3-\alpha}} n}$ and the number of stubs within the first bucket is $S(N_1) = b$.

We partition the first bucket into finer bins such that $B_1 = \bigcup_{t=1}^T V_t$ and $V_t = \{v_{t-1} + 1, \dots, v_t\}, t = 1, \dots, T$ with ascending order and $v_0 = 1$. The v_t will be specified in Lemma 4. We define the event that every nodes in bin V_t is infected as E_t , then the event that all the nodes in B_1 are infected is equal to $\bigcap_{t=1}^T E_t$.

We recall Lemma 1:

Lemma 1 (Base). *Given at \mathcal{F}_0 $d_1 > n^{\frac{3-\alpha}{\alpha+1}}$, at \mathcal{F}_1 all the nodes in B_1 will be infected within $O(\log \log(n))$ steps with probability greater than $1 - O(\frac{1}{n})$.*

We will use two Lemmas in the proof of Lemma 1, which will be a proof by induction. The first lemma will form the base case of the induction. It states the high degree nodes will all be infected by the first node by showing any high degree node forms k multi-edges to the first node.

Lemma 3. *Given $d_1 > n^{\frac{3-\alpha}{\alpha+1}}$ we define node $v_1 = \max\{v : d(v) \geq n^{\frac{3-\alpha}{\alpha+1}}\}$. (Recall nodes are ordered by degree.) Then all the nodes in $V_1 = \{1, \dots, v_1\}$ will be infected in one step with probability*

$$\Pr[E_1] = 1 - n^{\frac{3-\alpha}{\alpha+1}} \exp(-\Theta(1)n^{\frac{3-\alpha}{\alpha+1}}).$$

The second Lemma will form the inductive step in the proof of Lemma 1. It can be proved by induction itself.

Lemma 4. *Let $v \in V_t = \{v_{t-1} + 1, \dots, v_t\}$ and $v_t = \max\{v : d(v) \geq \frac{n^{\frac{\alpha-1}{3-\alpha}}}{\log^{\frac{\alpha-1}{3-\alpha}} n^{(\alpha-2)^t}}\}$, then*

$$\Pr \left[u \text{ is not infected} \mid \bigcap_{s=1}^{t-1} E_s \right] \leq \frac{1}{n^2}$$

Moreover, $T = O(\log \log n)$.

The proofs for Lemmas 3 and 4 are the simple application of a Chernoff bound and a union bound which is in the full version.

Proof (Lemma 1). The proof is by induction. For the base case, Lemma 3 ensures every node in the first bin will be infected. Suppose all nodes before v_{t-1} are infected. We can use a union bound to show every node in V_t will be also infected. Moreover, in each bin the contagion only takes one time step which implies that the infection time for the first bucket is at most $O(\log \log n)$.

For the probability that all these events hold, we apply a union bound.

$$\begin{aligned} & \Pr[\text{all the nodes in } B_1 \text{ are infected}] \\ &= \Pr \left[\bigcap_{t=1}^T E_t \right] \\ &\geq 1 - \Pr[\neg E_1] - \sum_{t=2}^T \Pr \left[\neg E_t \mid \bigcap_{s=1}^{t-1} E_s \right] \quad (\text{union bound}) \\ &\geq 1 - n^{\frac{3-\alpha}{\alpha+1}} \exp(-\Theta(1)n^{\frac{3-\alpha}{\alpha+1}}) - \frac{1}{n^2} |B_1| \quad \text{by Lemmas 3 and 4} \end{aligned}$$

5 Proof of Lemma 2: Contagion from Buckets to Bucket

In this section we prove Lemma 2.

Lemma 2 (Induction). *Fix sufficiently small $\epsilon > 0$, $\rho > 0$. Let $\ell < \rho L$, and suppose $\Pr[\mathcal{A}_\ell] > 0.5$, then we have*

$$\Pr[\mathcal{A}_{\ell+1} | \mathcal{A}_\ell] = 1 - O(1) \frac{(\log n)^{\frac{\alpha-1}{3-\alpha}}}{n \ell^{1/(\alpha-2)}}$$

Recall that \mathcal{A}_ℓ is the intersection of the three events, we will show that at stage ℓ if these three events happen, then the requirements in Lemma 2 will be met, and those events would be proven in Lemmas 5, 6 and 8 respectively.

5.1 First Event: Connection

We first note that the first event holds with high probability. This follows almost immediately from a standard Chernoff bound application, and the proof is in the full version.

Lemma 5. *Let $\delta_n = \Theta\left(\frac{1}{\log^{\frac{\alpha-2}{3-\alpha}} n}\right)$, if $\Pr[\mathcal{A}_\ell] \geq 0.5$*

$$\Pr\left[|X_{\ell+1} - \mathbb{E}[X_{\ell+1}]| \leq \delta_n \mathbb{E}[X_{\ell+1}] \mid \mathcal{A}_\ell\right] \geq 1 - 4 \exp\left(-\Theta\left(\frac{n}{\log^{6 \cdot \frac{\alpha-2}{3-\alpha}} n}\right)\right).$$

Here the constant only depends on the product of δ_n and L .

5.2 Second Event: Number of Infected Nodes

Now we will prove the second events holds with high probability.

Lemma 6 (Number of Uninfected Nodes in a Single Bucket). *For sufficiently small $\epsilon > 0$, conditioned on \mathcal{A}_ℓ*

$$\Pr\left[\sum_{i \in B_\ell} \mathbb{I}_i \geq 2\mu_H \mid \mathcal{A}_\ell\right] \leq O(1) \frac{(\log n)^{\frac{\alpha-1}{3-\alpha}}}{n\ell^{1/(\alpha-2)}}$$

where $\mu_H = K \frac{|B_\ell|\ell^{\frac{3-\alpha}{\alpha-2}}}{\log n}$ and K is independent of ℓ and n .

The proof relies on an application of Chebyshev’s inequality and the following Lemma, which is in turn proved using Chebyshev’s inequality. The full proof is in the full version.

Lemma 7 (Infection of a single node). *If $\mathcal{F}_\ell \subseteq \mathcal{A}_\ell$ for some constant $0 < \epsilon < 1/2$ and $\delta_n = \Theta\left(\frac{1}{\log^{\frac{\alpha-2}{3-\alpha}} n}\right) < 1/2$, then the probability any node $i \in B_\ell$ is not infected is*

$$\Pr[\mathbb{I}_i \mid \mathcal{A}_\ell] \leq O(1) \frac{\ell^{\frac{3-\alpha}{\alpha-2}}}{\log n}$$

where the constant $O(1)$ only depends on α, k, ρ if δ_n, ϵ is small enough, and $\rho \leq 0.3 \frac{\alpha-1}{\alpha-2} k^{\frac{\alpha-2}{3-\alpha}}$.

The main proof idea of Lemma 7 is that because the events that a infected stub from $B_{<\ell}$ to a node i in B_ℓ are negative dependent, the variance of the number of infected stubs from $B_{<\ell}$ to node i is small, and we can use Chebyshev’s inequality to show each node has a high chance of being infected when fraction of uninfected stubs from $B_{<\ell}$, ϵ , is small. The full proof is in the full version.

5.3 Third Event: Number of Uninfected Stubs

Lemma 8. *Suppose A_ℓ , the first event, $(1 - \delta_n)\mathbb{E}[X_{\ell+1}] \leq X_{\ell+1} \leq (1 + \delta_n)\mathbb{E}[X_{\ell+1}]$ and the second event, $\sum_{i \in B_\ell} \mathbb{I}_i \leq 2\mu_H$ is true (this is the conclusion of Lemma 6), then*

$$\Pr \left[Y_{\ell+1} \leq \epsilon X_{\ell+1} \mid |X_{\ell+1} - \mathbb{E}[X_{\ell+1}]| \leq \delta_n \mathbb{E}[X_{\ell+1}] \wedge \sum_{i \in B_\ell} \mathbb{I}_i \leq 2\mu_H \wedge \mathcal{A}_\ell \right]$$

is greater than $1 - \exp\left(-\Theta\left(\frac{n}{\log^{5 \cdot \frac{\alpha-2}{3-\alpha}}}\right)\right)$ when $\rho > 0$ is small enough and $\delta_n > 0$ is smaller than some constant.

For the third event, in Lemma 8 we want to argue the fraction of uninfected stubs is smaller than ϵ after stage ℓ . That requires both that $X_{\ell+1}$ is large and that $Y_{\ell+1}$ —which is the summation of $Y_{\ell+1}^{(1)}$ and $Y_{\ell+1}^{(2)}$ —is small. Upper bounds on $Y_{\ell+1}^{(1)}$ and $Y_{\ell+1}^{(2)}$ will be proven by Lemmas 9 and 10 respectively. The full proof for Lemma 8 is in the full version.

Lemma 9. *Let $Y_\ell^{(1)}$ be the number of free uninfected stubs from $B_{<\ell}$ to $B_{>\ell}$ over the probability space $\mathcal{F}_{\ell+1} | \mathcal{F}_\ell$, then*

$$\Pr \left[Y_{\ell+1}^{(1)} \geq (1 + \delta_n)\epsilon X_\ell | \mathcal{A}_\ell \right] \leq \exp\left(-\Theta\left(\frac{n}{\log^{5 \cdot \frac{\alpha-2}{3-\alpha}}}\right)\right)$$

Here the constant only depends on $\delta_n \cdot L$, ϵ and ρ .

Lemma 10. *Suppose A_ℓ and the $\sum_{i \in B_\ell} \mathbb{I}_i \leq 2\mu_H$ is true (this is the conclusion of Lemma 6), then $Y_{\ell+1}^{(2)}$, the total number of uninfected stubs from B_ℓ to $B_{>\ell}$ is*

$$Y_{\ell+1}^{(2)} = O(1) \frac{(\log n)^{\frac{\alpha}{3-\alpha}}}{n \ell^{2/(\alpha-2)}}$$

The full proofs for Lemmas 9 and 10 are in the full version.

5.4 Proof of Lemma 2

Proof. Recall the the event $\mathcal{A}_{\ell+1}$ is the intersection of the three events, so

$$\Pr[\mathcal{A}_{\ell+1} | \mathcal{A}_\ell] \geq 1 - \Pr[\neg(|X_\ell - \mathbb{E}[X_\ell]| \leq \delta_n \mathbb{E}[X_\ell]) | \mathcal{A}_\ell] \tag{1}$$

$$- \Pr \left[\sum_{i \in B_{\ell-1}} \mathbb{I}_i \geq 2\mu_H | \mathcal{A}_\ell \right] \tag{2}$$

$$- \Pr \left[Y_\ell \leq \epsilon X_\ell (|X_\ell - \mathbb{E}[X_\ell]| \leq \delta_n \mathbb{E}[X_\ell]) \wedge \sum_{i \in B_{\ell-1}} \mathbb{I}_i \leq 2\mu_H \wedge \mathcal{A}_\ell \right] \tag{3}$$

Applying Lemma 5 to Eq. 1, Lemma 6 to Eq. 2, and Lemma 8 to Eq. 3, and we have

$$\begin{aligned} \Pr[\mathcal{A}_{\ell+1}|\mathcal{A}_\ell] &\geq 1 - 4 \exp\left(-\Theta\left(\frac{n}{\log^{6 \cdot \frac{\alpha-2}{3-\alpha}} n}\right)\right) \\ &\quad - O(1) \frac{(\log n)^{\frac{\alpha-1}{3-\alpha}}}{n\ell^{1/(\alpha-2)}} \\ &\quad - \exp\left(-\Theta\left(\frac{n}{\log^{5 \cdot \frac{\alpha-2}{3-\alpha}}}\right)\right) \end{aligned}$$

Therefore

$$\Pr[\mathcal{A}_{\ell+1}|\mathcal{A}_\ell] \geq 1 - O(1) \frac{(\log n)^{\frac{\alpha-1}{3-\alpha}}}{n\ell^{1/(\alpha-2)}}$$

6 Infection with Random Seeds

Theorem 1 together with prior results in Ebrahimi et al. [16] immediately implies the following corollary:

Corollary 1. *For a configuration model graph with power-law exponent α , if $\Omega(n^{\frac{\alpha-2}{\alpha-1}})$ initially random seeds are chosen, then with probability $1 - o(1)$ k -complex contagion infects a constant fraction of nodes.*

We first restate two results from [16].

Proposition 2 [16]. *For any graph, let u be a node with degree d . If $\Omega(d/n)$ initial random seeds are chosen, then with probability $1 - o(1)$ u is infected after one round.*

Proof. The initial node has $\Theta(n^{\frac{1}{\alpha-1}})$ neighbors. If there are $\Omega(n^{\frac{\alpha-2}{\alpha-1}})$ initial seeds then by Proposition 2 the first seed is infected with probability $1 - o(1)$. However, then by Theorem 1 a constant fraction of the remaining nodes are infected in $\log^{O(\alpha)}(1)$ rounds.

This is tight as in Ebrahimi et al. [16] the following was proven:

Proposition 3 [16]. *For any graph, with power law distribution α , if $o(n^{\frac{\alpha-2}{\alpha-1}})$ initially random seeds are chosen, then with probability $1 - o(1)$, no additional nodes are infected.*

7 $\alpha > 3$

For the case of power-law degree distribution with $\alpha > 3$, Amini [4] shows how to analyze k -complex contagions using a differential equation method [30]. This approach heavily depends on the variance of the degree distribution and fails when $\alpha < 3$. For the case where the seed set contains all nodes with degree greater than $\rho > 0$ we can state their theorem as follows:

Theorem 4 [4]. *Given a power law distribution \mathbf{d} with exponent $\alpha > 3$ and $d_1 < n^{1/\alpha-1}$, the k -complex contagion on configuration $CM(\mathbf{d})$ with constant k and seed set $I_\rho = \{i | d_i \geq \rho\}$ where $0 \leq \rho \leq n$. Then with high probability*

$$|CC(CM(\mathbf{d}), k, I_\rho)| = n \left(1 - \sum_{\substack{1 \leq d < \rho, \\ 0 \leq j < k}} p_{\mathbf{d}}(d) \binom{d}{j} (y^*)^{d-j} (1 - y^*)^j + o(1) \right) \quad (4)$$

where $p_{\mathbf{d}}(d) = (F_{\mathbf{d}}(d + 1) - F_{\mathbf{d}}(d))$ and $0 < y^* \leq 1$ is the largest root such that $f(y) = 0$ and

$$f(y) = y^2 \left(\sum_{1 \leq d} d p_{\mathbf{d}}(d) \right) - y \left(\sum_{\substack{1 \leq d < \rho, \\ 0 \leq j < k}} d p_{\mathbf{d}}(d) \binom{d-1}{j} y^{d-1-j} (1 - y)^j \right) \quad (5)$$

Before stating our corollary, we wish to give a brief idea of the proof of Theorem 4. They consider a Markov chain which results in the same number infected nodes as a k -complex contagion, but proceeds using the randomness of the configuration model. The Markov chain starts with the initially infected nodes and at each step the process reveals one of the unmatched edges from the set of infected nodes. This process needs only track: the number of unmatched edges, and the number of d -degree uninfected nodes with j infected neighbors, for each $j < k$. The Markov chain stops when all the agent are infected, or there are no unmatched edges from already infected nodes. It turns out, that if $\alpha > 3$, the process is smooth and we can use the corresponding differential equations to approximate this Markov chain and derive the fraction of infections.

With their results we can prove that to infect a constant fraction of nodes requires the initial seed need to also be a constant fraction of nodes. Note that if our initial seed set infects the highest degree nodes, but does not infect a constant fraction of the nodes, then the greatest degree node not in the initially infected set has degree $\omega(1)$.

Corollary 2. *Given a power law distribution \mathbf{d} with exponent $\alpha > 3$ and $d_1 < n^{1/\alpha-1}$, the k -complex contagion on configuration $CM(\mathbf{d})$ with constant k and seed set $I_\rho = \{i | d_i \geq \rho\}$ where $\rho = \omega(1)$, the $|CC(CM(\mathbf{d}), k, I_\rho)| = o(n)$ with high probability.*

The proof of the corollary requires some delicate calculations and is in the full version.

References

1. Adamic, L.A., Glance, N.: The political blogosphere, the: divided they blog. In: Proceedings of the 3rd International Workshop on Link Discovery, pp. 36–43. ACM (2005)

2. Adler, J.: Bootstrap percolation. *Phys. A: Stat. Theor. Phys.* **171**(3), 453–470 (1991)
3. Albert, R., Barabási, A.-L.: Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002). doi:[10.1103/RevModPhys.74.47](https://doi.org/10.1103/RevModPhys.74.47). <http://link.aps.org/doi/10.1103/RevModPhys.74.47>
4. Amini, H.: Bootstrap percolation and diffusion in random graphs with given vertex degrees. *Electron. J. Comb.* **17**(1), 1–20 (2010)
5. Amini, H., Fountoulakis, N.: What I tell you three times is true: bootstrap percolation in small worlds. In: Goldberg, P.W. (ed.) *WINE 2012*. LNCS, vol. 7695, pp. 462–474. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-35311-6_34](https://doi.org/10.1007/978-3-642-35311-6_34)
6. Backstrom, L., Huttenlocher, D., Kleinberg, J., Lan, X.: 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*, pp. 44–54 (2006)
7. Balogh, J., Pittel, B.: Bootstrap percolation on the random regular graph. *Random Struct. Algorithms* **30**, 257–286 (2007)
8. Banerjee, A., Chandrasekhar, A.G., Duflo, E., Jackson, M.O.: The diffusion of microfinance. *Science* **341**(6144), 1236498 (2013)
9. Barabási, A., Albert, R.: Emergence of scaling in random networks. *Science* **286**, 509–512 (1999)
10. Bollobás, B., McKay, B.D.: The number of matchings in random regular graphs and bipartite graphs. *J. Comb. Theory, Series B* **41**(1), 80–91 (1986)
11. Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., Wiener, J.: Graph structure in the web. In: *Proceedings of the 9th International World Wide Web Conference on Computer Networks*, pp. 309–320 (2000)
12. Centola, D., Macy, M.: Complex contagions and the weakness of long ties¹. *Am. J. Sociol.* **113**(3), 702–734 (2007)
13. Chalupa, J., Leath, P.L., Reich, G.R.: Bootstrap percolation on a bethe lattice. *J. Phys. C: Solid State Phys.* **12**(1), L31 (1979)
14. Coleman, J., Katz, E., Menzel, H.: The diffusion of an innovation among physicians. *Sociometry* **20**(4), 253–270 (1957)
15. Coleman, J.S., Katz, E., Menzel, H.: *Medical Innovation: A Diffusion Study*. Bobbs-Merrill Co., Indianapolis (1966)
16. Ebrahimi, R., Gao, J., Ghasemiefteh, G., Schoenebeck, G.: How complex contagions spread quickly in the preferential attachment model, other time-evolving networks. arXiv preprint [arXiv:1404.2668](https://arxiv.org/abs/1404.2668) (2014)
17. Ebrahimi, R., Gao, J., Ghasemiefteh, G., Schoenebeck, G.: Complex contagions in Kleinberg’s small world model. In: *Proceedings of the Conference on Innovations in Theoretical Computer Science*, pp. 63–72. ACM (2015)
18. Ghasemiefteh, G., Ebrahimi, R., Gao, J.: Complex contagion, the weakness of long ties in social networks: revisited. In: *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, pp. 507–524, June 2013
19. Granovetter, M.: Threshold models of collective behavior. *Am. J. Sociol.* **83**(6), 1420–1443 (1978)
20. Janson, S., Luczak, T., Turova, T., Vallier, T.: Bootstrap percolation on the random graph $G_{n,p}$. *Ann. Appl. Prob.* **22**(5), 1989–2047 (2012)
21. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137–146 (2003)

22. Macdonald, J.S., Macdonald, L.D.: Chain migration, ethnic neighborhood formation and social networks. *Milbank Meml. Fund Q.* **42**(1), 82–97 (1964)
23. Mermelstein, R., Cohen, S., Lichtenstein, E., Baer, J.S., Kamarck, T.: Social support and smoking cessation and maintenance. *J. Consult. Clin. Psychol.* **54**(4), 447 (1986)
24. Newman, M.E.J.: The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003)
25. De Solla Price, D.: Networks of scientific papers. *Science* **149**(3683), 510–515 (1965). doi:[10.1126/science.149.3683.510](https://doi.org/10.1126/science.149.3683.510)
26. Romero, D.M., Meeder, B., Kleinberg, J.: Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In: *Proceedings of the 20th International Conference on World Wide Web*, pp. 695–704 (2011)
27. Steyvers, M., Tenenbaum, J.B.: The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cogn. Sci.* **29**, 41–78 (2005)
28. Ugander, J., Backstrom, L., Marlow, C., Kleinberg, J.: Structural diversity in social contagion. *Proc. Natl. Acad. Sci.* **109**(16), 5962–5966 (2012)
29. Van Der Hofstad, R.: Random graphs and complex networks, p. 11 (2009). <http://www.win.tue.nl/rhofstad/NotesRGCN.pdf>
30. Wormald, N.C.: Differential equations for random processes and random graphs. *Ann. Appl. Prob.* **5**(4), 1217–1235 (1995)