Think Globally, Act Locally: On the Optimal Seeding for Nonsubmodular Influence Maximization

Grant Schoenebeck
University of Michigan, Ann Arbor, USA
http://web.eecs.umich.edu/~schoeneb/
schoeneb@umich.edu

Biaoshuai Tao
University of Michigan, Ann Arbor, USA
http://www-personal.umich.edu/~bstao/
bstao@umich.edu

Fang-Yi Yu
University of Michigan, Ann Arbor, USA
http://www-personal.umich.edu/~fayu/
fayu@umich.edu

Abstract

We study the r-complex contagion influence maximization problem. In the influence maximization problem, one chooses a fixed number of initial seeds in a social network to maximize the spread of their influence. In the r-complex contagion model, each uninfected vertex in the network becomes infected if it has at least r infected neighbors.

In this paper, we focus on a random graph model named the *stochastic hierarchical blockmodel*, which is a special case of the well-studied *stochastic blockmodel*. When the graph is not exceptionally sparse, in particular, when each edge appears with probability $\omega\left(n^{-(1+1/r)}\right)$, under certain mild assumptions, we prove that the optimal seeding strategy is to put all the seeds in a single community. This matches the intuition that in a nonsubmodular cascade model placing seeds near each other creates synergy. However, it sharply contrasts with the intuition for submodular cascade models (e.g., the independent cascade model and the linear threshold model) in which nearby seeds tend to erode each others' effects.

Finally, we show that this observation yields a polynomial time dynamic programming algorithm which outputs optimal seeds if each edge appears with a probability either in $\omega\left(n^{-(1+1/r)}\right)$ or in $o\left(n^{-2}\right)$.

2012 ACM Subject Classification Theory of computation → Social networks; Mathematics of computing → Random graphs

Keywords and phrases Nonsubmodular Influence Maximization, Bootstrap Percolation, Stochastic Blockmodel

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2019.39

Category RANDOM

Funding *Grant Schoenebeck*: National Science Foundation AitF #1535912 and CAREER #1452915
Biaoshuai Tao: National Science Foundation CAREER #1452915
Fang-Yi Yu: National Science Foundation AitF #1535912

1 Introduction

A *cascade*, or a *contagion*[1], is a fundamental process on social networks: starting with some seed agents, the infection then spreads to their neighbors. A natural question known as influence maximization [4, 6, 18, 28] asks how to place a fixed number of initial seeds to maximize the spread of the resulting cascade. For example, which students can most effectively be enrolled in an intervention to decrease student conflict at a school [30]?

Influence maximization is extensively studied when the contagion process is submodular (a node's marginal probability of becoming infected after a new neighbor is infected decreases when the number of previously infected neighbors increases [22]). However, many examples of nonsubmodular contagions have been reported, including pricey technology innovations, the change of social behaviors, the decision to participate in a migration, etc [14, 27, 31, 2, 24]. In this case, a node's marginal influence may increase in the presence of other nodes – creating a kind of synergy.

Network structure and seed placement

We address this lack of understanding for nonsubmodular influence maximization by characterizing the optimal seed placement for certain settings which we will remark on shortly. In these settings, the optimal seeding strategy is to put all the seeds near each other. This is significantly different than in the submodular setting, where the optimal solutions tend to spread out the seeds, lest they erode each others' influence. We demonstrate this in the appendix (Sect. A) by presenting an example of submodular influence maximization where the optimal seeding strategy is to spread out the seeds.

This formally captures the intuition, as presented by Angell and Schoenebeck [1], that it is better to target one market to saturation first (act locally) and then to allow the success in this initial market to drive broader success (think globally) rather than to initially attempt a scattershot approach (act globally). It is also underscores the need to understand the particular nature of a contagion before blindly applying influence maximization tools.

We consider a well-known nonsubmodular cascade model which is also the most extreme one (in terms of nonsubmodularity), the r-complex contagion [19, 7, 8, 16] (a node is infected if and only if at least r of its neighbors are infected, also known as *bootstrap percolation*) when $r \geq 2$.

We consider networks formed by the *stochastic hierarchical blockmodel* [32, 33] which is a special case of the stochastic blockmodel [15, 20, 36] equipped with a hierarchical structure. Vertices are partitioned into m blocks. The blocks are arranged in a hierarchical structure which represents blocks merging to form larger and larger blocks (communities). The probability of an edge's presence between two vertices is based solely on smallest block to which both the vertices belong. This model captures the intuitive hierarchical structure which is also observed in many real-world networks [17, 12]. The stochastic hierarchical blockmodel is rather general and captures other well-studied models (e.g. Erdős-Rényi random graphs, and the planted community model) as special cases.

Result 1. We first prove that, for the influence maximization problem on the stochastic hierarchical blockmodel with r-complex contagion, under certain mild technical assumptions, the optimal seeding strategy is to put all the seeds in a single community, if, for each

[1] As is common in the literature, we use these terms interchangeably.

vertex-pair (u, v), the probability that the edge (u, v) is included satisfies $p_{uv} = \omega(n^{-(1+1/r)})$. Notice that the assumption $p_{uv} = \omega(n^{-(1+1/r)})$ captures many real life social networks. In fact, it is well-known that an Erdős-Rényi graph $\mathcal{G}(n, p)$ with $p = o(1/n)$ is globally disconnected: with probability $1 - o(1)$, the graph consists of a union of tiny connected components, each of which has size $O(\log n)$.

The technical heart of this result is a novel coupling argument in Proposition 16. We simultaneously couple four cascade processes to compare two probabilities: 1) the probability of infection spreading throughout an Erdős-Rényi graph after the $(k+1)$-st seed, conditioned on not already being entirely infected after k seeds; 2) the probability of infection spreading throughout the same graph after the $(k + 2)$-nd seed, conditioned on not already being entirely infected after $k+1$ seeds. This shows that the marginal rate of infection always goes up, revealing the "supermodular" nature of the r-complex contagion. The supermodular property revealed by Proposition 16 is a property for cascade behavior on Erdős-Rényi random graphs in general, so it is also interesting on its own.

Our result is in sharp contrast to Balkanski et al.'s observation. Balkanski et al. [3] studies the stochastic blockmodel with a well-studied submodular cascade model, *the independent cascade model*, and remarks that "when an influential node from a certain community is selected to initiate a cascade, the marginal contribution of adding another node from that same community is small, since the nodes in that community were likely already influenced."

Algorithmic aspects

For influence maximization in submodular cascades, a greedy algorithm efficiently finds a seeding set with influence at least a $(1 - 1/e)$ fraction of the optimal [22], and much of the work following Kempe et al. [22], which proposed the greedy algorithm, has attempted to make greedy approaches efficient and scalable [10, 11, 26, 13, 35, 34].

Greedy approaches, unfortunately, can perform poorly in the nonsubmodular setting [1]. Moreover, in contrast to the submodular case which has efficient constant approximation algorithms, for general nonsubmodular cascades, it is NP-hard even to approximate influence maximization to within an $\Omega(n^{1-\epsilon})$ factor of the optimal [23]. This inapproximability result has been extended to several much more restrictive nonsubmodular models [9, 25, 32, 33]. Intuitively, nonsubmodular influence maximization is hard because the potential synergy of multiple seeds makes it necessary to consider groups of seeds rather than just individual seeds. In contrast, with submodular influence maximization, not much is lost by considering seeds one at a time in a myopic way.

Can the $\Omega(n^{1-\epsilon})$ inapproximability results of Kempe et al. [23] be circumvented if we further assume the stochastic hierarchical blockmodel? On the one hand, the stochastic hierarchical structure seems optimized for a dynamic programming approach: perform dynamic programming from the bottom to the root in the tree-like community structure. On the other hand, Schoenebeck and Tao [32, 33] show that the $\Omega(n^{1-\epsilon})$ inapproximability results extend to the setting where the networks are stochastic hierarchical blockmodels.

Result 2. However, Result 1 (when the network is reasonably dense, putting all the seeds in a single community is optimal) can naturally be extended to a dynamic programming algorithm. We show that this algorithm is optimal if the probability p_{uv} that each edge appears does not fall into a narrow regime. Interestingly, a heuristic based on dynamic programming works fairly well in practice [1]. Our second result theoretically justifies the success of this approach, at least in the setting of r-complex contagions.

2 Preliminaries

We study complex contagions on social networks with community structure. This section defines the complex contagion and our model for social networks with community structure.

2.1 r-Complex Contagion

Given a social network modeled as an undirected graph $G = (V, E)$, in a cascade, a subset of nodes $S \subseteq V$ is chosen as the seed set; these seeds, being infected, then spread their influence across the graph according to some specified model.

In this paper, we consider a well-known cascade model named *r-complex contagion*, also known as *bootstrap percolation* and the *fixed threshold model*: a node is infected if and only if at least r of its neighbors are infected. We use $\sigma_{r,G}(S)$ to denote the total number of infected vertices at the end of the cascade, and $\sigma_{r,\mathcal{G}}(S) = \mathbb{E}_{G\sim\mathcal{G}}[\sigma_{r,G}(S)]$ if the graph G is sampled from some distribution $\mathcal{G}$. Notice that the function $\sigma_{r,G}(\cdot)$ is deterministic once the graph G and r are fixed.

Submodularity of a cascade model

Other than the r-complex contagion, most cascade models are stochastic: the total number of infected vertices is not deterministic but rather a *random variable*. $\sigma_G(S)$ usually refers to the *expected* number of infected vertices given the seed set S. A cascade model is *submodular* if, given any graph, subsets of vertices $S \subseteq T \subseteq V$, and any additional vertex $v \in V \setminus T$, we have

$$\sigma_G(S \cup \{v\}) - \sigma_G(S) \geq \sigma_G(T \cup \{v\}) - \sigma_G(T),$$

and it is *nonsubmodular* otherwise. Typical submodular cascade models include *the linear threshold model* and *the independent cascade model* [22], which are studied in an enormous past literature. The r-complex contagion, on the other hand, is a paradigmatic nonsubmodular model.

2.2 Stochastic hierarchical blockmodels

We study the *stochastic hierarchical blockmodel* first introduced in [33]. The stochastic hierarchical blockmodel is a special case of the *stochastic blockmodel* [20]. Intuitively, the stochastic blockmodel is a stochastic graph model generating networks with community structure, and the stochastic hierarchical blockmodel further assumes that the communities form a hierarchical structure. Our definition in this section follows closely to [33].

▶ **Definition 1.** *A* stochastic hierarchical blockmodel *is a distribution $\mathcal{G} = (V, T)$ of unweighted undirected graphs sharing the same vertex set V, where $T = (V_T, E_T, w)$ is a weighted tree called a* hierarchy tree. *The third parameter is the weight function $w : V_T \to [0, 1]$ satisfying $w(t_1) < w(t_2)$ for any $t_1, t_2 \in V_T$ such that t_1 is an ancestor of t_2. Let $L_T \subseteq V_T$ be the set of leaves in T. Each leaf node $t \in L_T$ corresponds to a subset of vertices $V(t) \subseteq V$, where the $V(t)$ sets partition the vertices in V. In general, if $t \notin L_T$, we define $V(t) = \bigcup_{t' \in L_T : t' \text{ is an offspring of } t} V(t')$.*

The graph $G = (V, E)$ is sampled from $\mathcal{G}$ in the following way. The vertex set V is deterministic. For $u, v \in V$, the edge (u, v) appears in G with probability equal to the weight of the least common ancestor of u and v in T. That is $\Pr((u, v) \in E) = \max_{t : u, v \in V(t)} w(t)$.

In the rest of this paper, we use the words *tree node* and *vertex* to refer to the vertices in V_T and V respectively. In Definition 1, the tree node $t \in V_T$ corresponds to community $V(t) \subseteq V$ in the social network. Moreover, if t is not a leaf and $t_1, t_2, \ldots$ are the children of t in V_T, then $V(t_1), V(t_2), \ldots$ partition $V(t)$ into sub-communities. Thus, our assumption that for any $t_1, t_2 \in V_T$ where t_1 is an ancestor of t_2 we have $w(t_1) < w(t_2)$ implies that the relation between two vertices is stronger if they are in a same sub-community in a lower level, which is natural.

To capture the scenario where the advertiser has the information on the high-level community structure but lacks the knowledge of the detailed connections inside the communities, when defining the influence maximization problem as an optimization problem, we would like to include T as a part of input, but not G. Rather than choosing which specific vertices are seeds, the seed-picker decides the number of seeds on each leaf and the graph $G \sim \mathcal{G}(n, T)$ is realized after seeds are chosen. Moreover, we are interested in large social networks with $n \to \infty$, so we would like that a single encoding of T is compatible with varying n. To enable this feature, we consider the following variant of the stochastic hierarchical block model.

▶ **Definition 2.** *A* succinct stochastic hierarchical blockmodel *is a distribution $\mathcal{G}(n, T)$ of unweighted undirected graphs sharing the same vertex set V with $|V| = n$, where n is an integer which is assumed to be extremely large. The hierarchy tree $T = (V_T, E_T, w, v)$ is the same as it is in Definition 1, except for the followings.*

1. *Instead of mapping a tree node t to a weight in $[0, 1]$, the weight function $w : V_T \to \mathcal{F}$ maps each tree node to a function $f \in \mathcal{F} = \{f \mid f : \mathbb{Z}^+ \to [0, 1]\}$ which maps an integer (denoting the number of vertices in the network) to a weight in $[0, 1]$. The weight of t is then defined by $(w(t))(n)$. We assume $\mathcal{F}$ is the space of all functions that can be succinctly encoded.*
2. *The fourth parameter $v : V_T \to (0, 1]$ maps each tree node $t \in V_T$ to a fraction of vertices in $V(t)$. That is: $v(t) = |V(t)|/n$. Naturally, we have $\sum_{t \in L_T} v(t) = 1$ and $\sum_{t' : t' \text{ is a child of } t} v(t') = v(t)$.*

We assume throughout that $\mathcal{G}(n, T)$ has the following properties.

Large communities. For tree node $t \in V_T$, because $v(t)$ does not depend on n, $|V(t)| = v(t)n = \Theta(n)$. In particular, $|V(t)|$ goes to infinity as n does.

Proper separation. $w(t_1) = o(w(t_2))$ for any $t_1, t_2 \in V_T$ such that t_1 is an ancestor of t_2. That is, the connection between sub-community t_2 is asymptotically (with respect to n) denser than its super-community t_1.

Our definitions of w and v are designed so that we can fix a hierarchy tree $T = (V_T, E_T, w, v)$ and naturally define $\mathcal{G}(n, T)$ for any n. As we will see in the next subsection, this allows us to take T as input and then allow $n \to \infty$ when considering InfMax (to be defined soon). This enables us to consider graphs having arbitrarily many vertices.

Finally, we define the *density* of a tree node.

▶ **Definition 3.** *Given a hierarchy tree $T = (V_T, E_T, w, v)$ and a tree node $t \in V_T$, the* density *of the tree node is $\rho(t) = w(t) \cdot (v(t)n)^{1/r}$.*

2.3 The InfMax problem

We study the r-complex contagion on the succinct stochastic hierarchical blockmodel. Roughly speaking, given hierarchy tree T and an integer K, we want to choose K seeds which maximize the expected total number of infected vertices, where the expectation is taken over the graph sampling $G \sim \mathcal{G}(n, T)$ as $n \to \infty$.

▶ **Definition 4.** *The* influence maximization problem INFMAX *is an optimization problem which takes as input an integer r, a hierarchy tree $T = (V_T, E_T, w, v)$ as in Definition 2, and an integer K, and outputs $\boldsymbol{k} \in \mathbb{N}_{\geq 0}^{|L_T|}$ – an allocation of K seeds into the leaves L_T with $\sum_{t \in L_T} k_t = K$ that maximizes*

$$\Sigma_{r,T}(\boldsymbol{k}) := \lim_{n \to \infty} \frac{\mathbb{E}_{G \sim \mathcal{G}(n,T)} \left[\sigma_{r,G}(S_{\boldsymbol{k}})\right]}{n},$$ [2]

the expected fraction of infected vertices in $\mathcal{G}(n,T)$ with the seeding strategy defined by $\boldsymbol{k}$, where $S_{\boldsymbol{k}}$ denotes the seed set in G generated according to $\boldsymbol{k}$.

Before we move on, the following remark is very important throughout the paper.

▶ Remark 5. In Definition 4, n is not part of the inputs to the INFMAX instance. Instead, the tree T is given as an input to the instance, and we take $n \to \infty$ to compute $\Sigma_{r,T}(\boldsymbol{k})$ *after* the seed allocation is determined. Therefore, asymptotically, all the input parameters to the instance, including K, r and the encoding size of T, are *constants* with respect to n. Thus, there are two different asymptotic scopes in this paper: *the asymptotic scope with respect to the input size* and *the asymptotic scope with respect to n*. Naturally, when we are analyzing the running time of an INFMAX algorithm, we should use the asymptotic scope with respect to the input size, not of n. On the other hand, when we are analyzing the number of infected vertices after the cascade, we should use the asymptotic scope with respect to n.

In this paper, we use $O_I(\cdot), \Omega_I(\cdot), \Theta_I(\cdot), o_I(\cdot), \omega_I(\cdot)$ to refer to the asymptotic scope with respect to the input size, and we use $O(\cdot), \Omega(\cdot), \Theta(\cdot), o(\cdot), \omega(\cdot)$ to refer to the asymptotic scope with respect to n. For example, with respect to n we always have $r = \Theta(1)$, $K = \Theta(1)$ and $|V_T| = \Theta(1)$.

Lastly, we have assumed that $r \geq 2$, so that the contagion is nonsubmodular. When $r = 1$, the cascade model becomes a special case of the *independent cascade model* [22], which is a submodular cascade model. As mentioned, for submodular INFMAX, a simple greedy algorithm is known to achieve a $(1 - 1/e)$-approximation to the optimal influence [22, 23, 29].

2.4 r-Complex Contagion on Erdős-Rényi graphs

In this section, we consider the r-complex contagion on the Erdős-Rényi random graph $\mathcal{G}(n,p)$. We review some results from [21] which are used in our paper.

▶ **Definition 6.** *The* Erdős-Rényi random graph $\mathcal{G}(n,p)$ *is a distribution of graphs with the same vertex set V with $|V| = n$. For each pair of vertices u, v, the edge (u,v) in included in E independently with probability p.*

The INFMAX problem in Definition 4 on $\mathcal{G}(n,p)$ is trivial, as there is only one possible allocation of the K seeds: allocate all the seeds to the single leaf node of T, which is the root. Therefore, $\sigma_{r,T}(\cdot)$ in Definition 4 depends only on the *number* of seeds $K = |\boldsymbol{k}|$, not on the seed allocation $\boldsymbol{k}$ itself. In this section, we slightly abuse the notation σ such that it is a function mapping an *integer* to $\mathbb{R}_{\geq 0}$ (rather than mapping *an allocation of K seeds* to $\mathbb{R}_{\geq 0}$ as it is in Definition 4). Let $\sigma_{r,\mathcal{G}(n,p)}(k)$ denote the expected number of infected vertices after the cascade given k seeds. Correspondingly, let $\sigma_{r,G}(k)$ denote the actual number of infected vertices after the graph G is sampled from $\mathcal{G}(n,p)$.

▶ **Theorem 7** (A special case of Theorem 3.1 in [21])**.** *Suppose $r \geq 2$, $p = o(n^{-1/r})$ and $p = \omega(n^{-1})$. We have*

1. *if k is a constant, then $\sigma_{r,\mathcal{G}(n,p)}(k) \leq 2k$ with probability $1 - o(1)$;*
2. *if $k = \omega\left((1/np^r)^{1/(r-1)}\right)$, then $\sigma_{r,\mathcal{G}(n,p)}(k) = n - o(n)$ with probability $1 - o(1)$.*

▶ **Theorem 8** (Theorem 5.8 in [21]). *If $r \geq 2$, $p = \omega(n^{-1/r})$ and $k \geq r$, then we have* $\Pr_{G\sim\mathcal{G}(n,p)}[\sigma_{r,G}(k) = n] = 1 - o(1)$.

When $p = \Theta(n^{-1/r})$, the probability that k seeds infect all the n vertices is positive, but bounded away from 1. We use $\mathrm{Po}(\lambda)$ to denote the Poisson distribution with mean λ.

▶ **Theorem 9** (Theorem 5.6 and Remark 5.7 in [21]). *If $r \geq 2$, $p = cn^{-1/r} + o(n^{-1/r})$ for some constant $c > 0$, and $k \geq r$ is a constant, then*

$$\lim_{n\to\infty} \Pr\left(\sigma_{r,\mathcal{G}(n,p)}(k) = n\right) = \zeta(k, c),$$

for some $\zeta(k, c) \in (0, 1)$. Furthermore, there exist numbers $\zeta(k, c, \ell) > 0$ for $\ell \geq k$ such that

$$\lim_{n\to\infty} \Pr\left(\sigma_{r,\mathcal{G}(n,p)}(k) = \ell\right) = \zeta(k, c, \ell)$$

for each $\ell \geq k$, and $\zeta(k, c) + \sum_{\ell=k}^{\infty} \zeta(k, c, \ell) = 1$.

Moreover, the numbers $\zeta(k, c, \ell)$'s and $\zeta(k, c)$ can be expressed as the hitting probabilities of the following inhomogeneous random walk. Let $\xi_\ell \sim \mathrm{Po}\left(\binom{\ell-1}{r-1}c^r\right)$, $\ell \geq 1$ be independent, and let $\tilde{S}_\ell := \sum_{j=1}^{\ell}(\xi_j - 1)$ and $\tilde{T} := \min\{\ell : k + \tilde{S}_\ell = 0\} \in \mathbb{N} \cup \{\infty\}$. Then

$$\zeta(k, c) = \Pr\left(\tilde{T} = \infty\right) = \Pr\left(k + \tilde{S}_\ell \geq 1 \text{ for all } \ell \geq 1\right) \tag{1}$$

and $\zeta(k, c, \ell) = \Pr(\tilde{T} = \ell)$.

We have the following corollary for Theorem 9, saying that when $p = \Theta(n^{-1/r})$, if not all vertices are infected, then the number of infected vertices is constant. As a consequence, if the cascade spreads to more than constantly many vertices, then all vertices will be infected.

▶ **Corollary 10** (Lemma 11.4 in [21]). *If $r \geq 2$, $p = cn^{-1/r} + o(n^{-1/r})$ for some constant $c > 0$, and $k \geq r$, then*

$$\lim_{n\to\infty} \Pr\left(\phi(n) \leq \sigma_{r,\mathcal{G}(n,p)}(k) < n\right) = 0$$

for any function $\phi : \mathbb{Z}^+ \to \mathbb{R}^+$ such that $\lim_{n\to\infty} \phi(n) = \infty$.

3 Our main result

Our main result is the following theorem, which states that the optimal seeding strategy is to put all the seeds in a community with the highest density, when the root has a weight in $\omega(1/n^{1+1/r})$.

▶ **Theorem 11.** *Consider the* InfMax *problem with $r \geq 2$, $T = (V_T, E_T, w, v)$, $K > 0$ and the weight of the root node satisfying $w(root) = \omega(1/n^{1+1/r})$. Let $t^* \in \operatorname{argmax}_{t\in L_T} \rho(t)$ and $\boldsymbol{k}^*$ be the seeding strategy that puts all the K seeds on t^*. Then $\boldsymbol{k}^* \in \operatorname{argmax}_{\boldsymbol{k}} \Sigma_{r,T}(\boldsymbol{k})$.*

Notice that the assumption $w(\mathrm{root}) = \omega(1/n^{1+1/r})$ captures many real life social networks. In fact, it is well-known that an Erdős-Rényi graph $\mathcal{G}(n, p)$ with $p = o(1/n)$ is globally disconnected: with probability $1 - o(1)$, the graph consists of a union of tiny connected components, each of which has size $O(\log n)$.

The remaining part of this section is dedicated to proving Theorem 11. We assume $w(\mathrm{root}) = \omega(1/n^{1+1/r})$ in this section from now on. It is worth noting that, in many parts of this proof, and also in the proof of Theorem 23, we have used the fact that an infection of $o(n)$ vertices contributes 0 to the objective $\Sigma_{r,T}(\boldsymbol{k})$, as we have taken the limit $n \to \infty$ and divided the expected number of infections by n in Definition 4.

▶ **Definition 12.** *Given $T = (V_T, E_T, w, v)$, a tree node $t \in V_T$ is* supercritical *if $w(t) = \omega(1/n^{1/r})$, is* critical *if $w(t) = \Theta(1/n^{1/r})$, and is* subcritical *if $w(t) = o(1/n^{1/r})$.*

From the results in Sect. 2.4, if we allocate $k \geq r$ seeds on a supercritical leaf $t \in L_T$, then with probability $1 - o(1)$ all vertices in $V(t)$ will be infected; if we allocate k seeds on a subcritical leaf $t \in L_T$, at most a negligible number of vertices, $2k = \Theta(1)$, will be infected; if we allocate $k \geq r$ seeds on a critical leaf $t \in L_T$, the number of infected vertices in $V(t)$ follows Theorem 9.

We say a tree node $t \in V_T$ is *activated* in a cascade process if the number of infected vertices in $V(t)$ is $v(t)n - o(n)$, i.e., almost all vertices in $V(t)$ are infected. Given a seeding strategy $\boldsymbol{k}$, let $P_{\boldsymbol{k}}$ be the probability that at least one tree node is activated when $n \to \infty$. Notice that this is equivalent to at least one leaf being activated. The proof of Theorem 11 consists of two parts. We will first show that, $P_{\boldsymbol{k}}$ completely determines $\Sigma_{r,T}(\boldsymbol{k})$ (Lemma 13). Secondly, we show that placing all the seeds on a single leaf with the maximum density will maximize $P_{\boldsymbol{k}}$ (Lemma 14).

▶ **Lemma 13.** *Given any two seeding strategies $\boldsymbol{k}_1, \boldsymbol{k}_2$, if $P_{\boldsymbol{k}_1} \leq P_{\boldsymbol{k}_2}$, then $\Sigma_{r,T}(\boldsymbol{k}_1) \leq \Sigma_{r,T}(\boldsymbol{k}_2)$.*

▶ **Lemma 14.** *Let $\boldsymbol{k}$ be the seeding strategy that allocates all the K seeds on a leaf $t^* \in \operatorname{argmax}_{t \in L_T}(\rho(t))$. Then $\boldsymbol{k}$ maximizes $P_{\boldsymbol{k}}$.*

Lemma 13 and Lemma 14 imply Theorem 11. The proof of Lemma 13 is available in the full version. We prove Lemma 14 in the next section.

3.1 Proof of Lemma 14

We first handle some corner cases. If $K < r$, then the cascade will not even start, and any seeding strategy is considered optimal. If T contains a supercritical leaf, the leaf with the highest density is also supercritical. Putting all the $K \geq r$ seeds in this leaf, by Theorem 8, will activate the leaf with probability $1 - o(1)$. Therefore, this strategy makes $P_{\boldsymbol{k}} = 1$, which is clearly optimal. In the remaining part of this subsection, we shall only consider the case $K \geq r$ and all the leaves are either critical or subcritical. Notice that, by the proper separation assumption, all internal tree nodes of T are subcritical.

We split the cascade process into two stages. In Stage I, we restrict the cascade within the leaf blocks ($V(t)$ where $t \in L_T$), and temporarily assume there are no edges between two different leaf blocks (similar to if $w(t) = 0$ for all $t \notin L_T$). After Stage I, Stage II consists of the remaining cascade process.

Proposition 15 shows that maximizing $P_{\boldsymbol{k}}$ is equivalent to maximizing the probability that a leaf is activated in Stage I. Therefore, we can treat T such that all the leaves, each of which corresponds to a $\mathcal{G}(n, p)$ random graph, are isolated.

▶ **Proposition 15.** *If no leaf is activated after Stage I, then with probability $1 - o(1)$ no vertex will be infected in Stage II, i.e., the cascade will end after Stage I.*

We defer the proof of Proposition 15 to Appendix C. Notice that Proposition 15 is the only part where we have used the proper separation assumption.

Since Theorem 7 suggests that any constant number of seeds will not activate a subcritical leaf, we should only consider putting seeds in critical leaves. In Proposition 16, we show that in a critical leaf t, the probability that the $(i+1)$-th seed will activate t conditioning on the first i seeds failing to do so is increasing as i increases. Intuitively, Proposition 16 reveals a

super-modular nature of the r-complex contagion on a critical leaf, making it beneficial to put all seeds together so that the synergy is maximized, which intuitively implies Lemma 14. The proof of Proposition 16 is the most technical result of this paper, we will present it in Sect. 4.

▶ **Proposition 16** (log-concavity of $\lim_{n\to\infty}\Pr(E_k^n)$)**.** *Consider an Erdős-Rényi random graph $\mathcal{G}(n,p)$ with $p = cn^{-1/r} + o(n^{-1/r})$, and assume an arbitrary order on the n vertices. Let E_k^n be the event that seeding the first k vertices does not make all the n vertices infected. We have $\lim_{n\to\infty}\Pr(E_{k+2}^n \mid E_{k+1}^n) < \lim_{n\to\infty}\Pr(E_{k+1}^n \mid E_k^n)$ for any $k \geq r-1$.*

Equipped with Proposition 16, to show Lemma 14, we show that the seeding strategy that allocates $K_1 > 0$ seeds on a critical leaf t_1 and $K_2 > 0$ seeds on a critical leaf t_2 cannot be optimal. Firstly, it is obvious that both K_1 and K_2 should be at least r, for otherwise those K_1 (K_2) seeds on t_1 (t_2) are simply wasted.

Let E_k^n be the event that the first k seeds on t_1 fail to activate t_1 and F_k^n be the event that the first k seeds on t_2 fail to activate t_2. By Proposition 16, we have $\lim_{n\to\infty}\Pr(E_{K_1+1}^n \mid E_{K_1}^n) < \lim_{n\to\infty}\Pr(E_{K_1}^n \mid E_{K_1-1}^n)$ and $\lim_{n\to\infty}\Pr(F_{K_2+1}^n \mid F_{K_2}^n) < \lim_{n\to\infty}\Pr(F_{K_2}^n \mid F_{K_2-1}^n)$, which implies

$$\begin{aligned}&\lim_{n\to\infty}\frac{\Pr(E_{K_1+1}^n)\Pr(F_{K_2-1}^n)}{\Pr(E_{K_1}^n)\Pr(F_{K_2}^n)}\cdot\frac{\Pr(E_{K_1-1}^n)\Pr(F_{K_2+1}^n)}{\Pr(E_{K_1}^n)\Pr(F_{K_2}^n)}\\ =&\lim_{n\to\infty}\frac{\Pr(E_{K_1+1}^n \mid E_{K_1}^n)\Pr(F_{K_2+1}^n \mid F_{K_2}^n)}{\Pr(E_{K_1}^n \mid E_{K_1-1}^n)\Pr(F_{K_2}^n \mid F_{K_2-1}^n)} < 1.\end{aligned}$$

Therefore, we have either $\lim_{n\to\infty}\frac{\Pr(E_{K_1+1}^n)\Pr(F_{K_2-1}^n)}{\Pr(E_{K_1}^n)\Pr(F_{K_2}^n)}$ or $\lim_{n\to\infty}\frac{\Pr(E_{K_1-1}^n)\Pr(F_{K_2+1}^n)}{\Pr(E_{K_1}^n)\Pr(F_{K_2}^n)}$ is less than 1. This means either the strategy putting K_1+1 seeds on t_1 and K_2-1 seeds on t_2, or the strategy putting K_1-1 seeds on t_1 and K_2+1 seeds on t_2 makes it more likely that at least one of t_1 and t_2 is activated. Therefore, the strategy putting K_1 and K_2 seeds on t_1 and t_2 respectively cannot be optimal. This implies an optimal strategy should not allocate seeds on more than one leaf.

Finally, a critical leaf t with $v(t)n$ vertices and weight $w(t)$ can be viewed as an Erdős-Rényi random graph $\mathcal{G}(m,p)$ with $m = v(t)n$ and $p = w(t) = \rho(t)\cdot(v(t)n)^{-1/r} = \rho(t)m^{-1/r}$, where $\rho(t) = \Theta(1)$ when t is critical. Taking $c = \rho(t)$ in Theorem 9, we can see that ξ_ℓ has a larger Poisson mean if c is larger, making it more likely that the $\mathcal{G}(m,p)$ is fully infected (to see this more naturally, larger c means larger p if we fix m). Thus, given that we should put all the K seeds in a single leaf, we should put them on a leaf with the highest density. This concludes Lemma 14.

4 Proof for Proposition 16

Since the event E_{k+1}^n implies E_k^n, we have $\Pr(E_{k+1}^n|E_k^n) = \Pr(E_{k+1}^n)/\Pr(E_k^n)$. Therefore, the inequality we are proving is equivalent to $\lim_{n\to\infty}\Pr(E_{k+2}^n)/\Pr(E_{k+1}^n) < \lim_{n\to\infty}\Pr(E_{k+1}^n)/\Pr(E_k^n)$, and it suffices to show that

$$\lim_{n\to\infty}\Pr(E_{k+2}^n)\lim_{n\to\infty}\Pr(E_k^n) < \lim_{n\to\infty}\Pr(E_{k+1}^n)\lim_{n\to\infty}\Pr(E_{k+1}^n). \tag{2}$$

Proposition 16 shows that the failure probability, $\lim_{n\to\infty}\Pr(E_k^n)$, is logarithmically concave.

The remaining part of the proof is split into four parts: In Sect. 4.1, we begin by translating Eqn (2) in the language of inhomogeneous random walks. In Sect. 4.2, we present a coupling of two inhomogeneous random walks to prove Eqn. (2). In Sect. 4.3, we prove the validity of the coupling. in Sect. 4.4, we finally show the coupling implies Eqn. (2).

4.1 Inhomogeneous random walk interpretation

We adopt the inhomogeneous random walk interpretation from Theorem 9, and view the event E_k^n as the following: the random walk starts at $x = k$; in the i-th iteration, x moves to the left by 1 unit, and moves to the right by $\alpha(i) \sim \mathrm{Po}\left(\binom{i-1}{r-1}c^r\right)$ units; Let $\mathcal{E}_k$ be the event that the random walk reaches $x = 0$. By Theorem 9, $\Pr(\mathcal{E}_k) = \lim_{n\to\infty} \Pr(E_k^n)$. Thus, $\lim_{n\to\infty} \Pr(E_{k+2}^n) \lim_{n\to\infty} \Pr(E_k^n) = \Pr(\mathcal{E}_{k+2}) \Pr(\mathcal{E}_k)$. In this proof, we let $\lambda(i) = \binom{i-1}{r-1}c^r$, and in particular, $\lambda(0) = \lambda(1) = \cdots = \lambda(r-1) = 0$. Note that as i increases, the expected movement of the walk increases, and make it harder to reach 0. This observation is important for our proof.

To compute $\Pr(\mathcal{E}_{k+2}) \Pr(\mathcal{E}_k)$, we consider the following process. A random walk in $\mathbb{Z}^2$ starts at $(k+2, k)$. In each iteration i, the random walk moves from (x, y) to $(x - 1 + \alpha(i), y - 1 + \beta(i))$ where $\alpha(i)$ and $\beta(i)$ are sampled from $\mathrm{Po}(\lambda(i))$ independently. If the random walk hits the axis $y = 0$ after a certain iteration $\mathcal{T}$, then it is stuck to the axis, i.e., for any $i > \mathcal{T}$, the update in the i-th iteration is from $(x, 0)$ to $(x - 1 + \alpha(i), 0)$; similarly, after reaching the axis $x = 0$, the random walk is stuck to the axis $x = 0$ and updates to $(0, y - 1 + \beta(i))$. Then, $\Pr(\mathcal{E}_{k+2}) \Pr(\mathcal{E}_k)$ is the probability that the random walk starting from $(k+2, k)$ reaches $(0, 0)$.

To prove (2), we consider two random walks in $\mathbb{Z}^2$ defined above. Let A be the random walk starting from $(k+2, k)$, and let B be the random walk starting from $(k+1, k+1)$. Let H_A and H_B be the event that A and B reaches $(0, 0)$ respectively. To prove (2), it is sufficient to show:

$$\Pr(H_A) < \Pr(H_B).$$

To formalize this idea, we define a coupling between A and B such that: 1) whenever A reaches $(0, 0)$, B also reaches $(0, 0)$, and 2) with a positive probability, B reaches $(0, 0)$ but A never does.

In defining the coupling, we use the idea of splitting and merging of Poisson processes [5]. We reinterpret the random walk by breaking down each *iteration* i into $J(i)$ *steps* such that it is symmetric in the x- and y-directions (with respect to the line $y = x$) and the movement in each step is "small".

If at the beginning of iteration i the process is at (x, y) with $x > 0$ and $y > 0$:

- At step 0 of iteration i, we sample $J(i) \sim \mathrm{Po}(2\lambda(i))$, set $(\alpha(i, 0), \beta(i, 0)) = (-1, -1)$, and update $(x, y) \mapsto (x + \alpha(i, 0), y + \beta(i, 0))$;
- At each step j for $j = 1, \ldots, J(i)$, $(\alpha(i, j), \beta(i, j)) = (1, 0)$ with probability 0.5, and $(\alpha(i, j), \beta(i, j)) = (0, 1)$ otherwise. Update $(x, y) \mapsto (x + \alpha(i, j), y + \beta(i, j))$;[3]

On the other hand, if $x = 0$ (or $y = 0$) at the beginning of iteration:

- At step 0 of iteration i, we sample $J(i) \sim \mathrm{Po}(2\lambda(i))$, set $\big(\alpha(i, 0), \beta(i, 0)\big) = (0, -1)$ (or $(-1, 0)$ if $y = 0$), and update $(x, y) \mapsto \big(x + \alpha(i, 0), y + \beta(i, 0)\big)$;
- At each step j for $j = 1, \ldots, J(i)$, with probability 0.5 $\big(\alpha(i, j), \beta(i, j)\big) = (1, 0)$, (or $\big(\alpha(i, j), \beta(i, j)\big) = (0, 1)$) and $(\alpha(i, j), \beta(i, j)) = (0, 0)$, otherwise. Update $(x, y) \mapsto \big(x + \alpha(i, j), y + \beta(i, j)\big)$;

If at the end of iteration i, $(x, y) = (0, 0)$ we stop the process.

[3] Standard results from Poisson process indicate that, $\sum_{j=1}^{J(i)} \alpha(i, j) \sim \mathrm{Po}(\lambda(i))$, and $\sum_{j=1}^{J(i)} \beta(i, j) \sim \mathrm{Po}(\lambda(i))$ which are two independent Poisson random variables.

Notice that we only switch from one type of iteration to the other if $x = 0$ (or $y = 0$) at the *end* of an iteration i. Here way say the random walk is stuck to the axis $x = 0$ (or the axis $y = 0$). If this happens, it will be stuck to this axis forever. Also, notice that in each step we have at most 1 unit movement. Also, in steps $j = 1, \ldots, J(i)$ the walk can only move further away from both axes $y = 0$ and $x = 0$.

Let $\big(x(i,j), y(i,j)\big)$ be the position of the random walk after iteration i step j, and $\big(x(i), y(i)\big)$ be its position at the end of iteration i. Moreover, let $\alpha(i) = \sum_{j=1}^{J(i)} \alpha(i,j)$ be the net movement in x direction during iteration i excluding the movement in Step 0, and let $\bar{\alpha}(i) = \alpha(i) + \alpha(i,0)$ be the net movement including movement at step 0. Similarly define y-directional movements $\beta(i) = \sum_{j=1}^{J(i)} \beta(i,j)$ and $\bar{\beta}(i)$.

4.2 The coupling

We want to show that the probability of A reaching the origin is less that of B. To this end, we create a coupling between the two walks, which we outline here. Fig. 1 and Fig. 2 illustrate most aspects of this coupling. In the description of the coupling, we will let B move "freely", and define how A is "coupled with" B.

Recall that A starts at $(k+2, k)$ and B starts at $(k+1, k+1)$. At the beginning, we set A's movement to be identical to B's. Before one of them hits the origin, either of the following two events must happen: A and B become symmetric to the line $x = y$ at some step, $\mathcal{E}_{\mathsf{symm}}$, or A reaches the axis $y = 0$ at the end of some iteration, $\mathcal{E}_{\mathsf{skew}}$. This is called Phase I and is further discussed in Sect. 4.2.1.

In the first case $\mathcal{E}_{\mathsf{symm}}$, the positions of A and B are symmetric. We set A's movement to mirror B's movement. Therefore, in this case, A and B will both hit the origin, or neither of them will. This is called Phase II Symm and is further discussed in Sect. 4.2.2.

For the latter case $\mathcal{E}_{\mathsf{skew}}$, A reaches the axis $y = 0$ at iteration $\mathcal{T}_{\mathsf{skew}}$. We call the process is in Phase II Skew and further discussed in Sect. 4.2.3. Because B starts one unit above A and one unit to the left of A, at iteration $\mathcal{T}_{\mathsf{skew}}$, B is at the axis $y = 1$ and one unit to the left of A. Next we couple A's movement in the x-direction to be identical to B's, so that B is always one unit to the left of A. This coupling continues unless B hits the axis $x = 0$. Denote this iteration $\mathcal{T}^*$. At time $\mathcal{T}^*$, A is one unit to the right of the axis $x = 0$. Recall that at iteration $\mathcal{T}_{\mathsf{skew}}$ when $\mathcal{E}_{\mathsf{skew}}$ happens, B is one unit above the axis so that $y = 1$. Therefore, we can couple the movement of A in the x-direction after iteration $\mathcal{T}^*$ with B's movement in the y-direction after iteration $\mathcal{T}_{\mathsf{skew}}$. Because $\lambda(i)$ increases with i, we can couple the walks in such a way as to ensure that A moves toward the origin at a strictly slower rate than B does. Therefore, A only reaches the y-axis $x = 0$ if B reaches the x-axis $y = 0$, and we have shown that A is less likely to reach the origin than B does.

Let $\big(x^A(i,j), y^A(i,j)\big)$, and $\big(x^B(i,j), y^B(i,j)\big)$ be the coordinates for A and B respectively after iteration i step j. Similarly, let $J^A(i)$ and $J^B(i)$ be the number of steps for A and B in iteration i. Let $\alpha^A(i,j)$ and $\alpha^B(i,j)$ be the x-direction movements of both walks in iteration i step j, and $\beta^A(i,j)$ and $\beta^B(i,j)$ be the corresponding y-direction movements.

4.2.1 Phase I

Starting with $\big(x^A(0), y^A(0)\big) = (k+2, k)$ and $\big(x^B(0), y^B(0)\big) = (k+1, k+1)$, A moves in exactly the same way as B, i.e., $J^A(i) = J^B(i)$, $\alpha^A(i,j) = \alpha^B(i,j)$ and $\beta^A(i,j) = \beta^B(i,j)$, until one of the following two events happens.

Event $\mathcal{E}_{\mathsf{symm}}$. The current position of A and B are symmetric with respect to the line $y = x$, i.e., $x^A(i,j) - x^B(i,j) = y^B(i,j) - y^A(i,j)$ and $x^A(i,j) + x^B(i,j) = y^A(i,j) + y^B(i,j)$. Notice that $\mathcal{E}_{\mathsf{symm}}$ may happen in some middle step j of an iteration i. When $\mathcal{E}_{\mathsf{symm}}$ happens, we move on to Phase II Symm.

Event $\mathcal{E}_{\mathsf{skew}}$. A hits the axis $y = 0$ *at the end of an iteration*. Notice that this means A is then stuck to the axis $y = 0$ forever. When $\mathcal{E}_{\mathsf{skew}}$ happens, we move on to Phase II Skew. Note that B is one unit away from the axis $y = 0$, $y^B = 1$. We remark that the in the third part we show, if event $\mathcal{E}_{\mathsf{skew}}$ happens, B has a higher chance to reach $(0,0)$ than A.

The following three claims will be useful.

▷ Claim 17. A is always below the line $y = x$ before $\mathcal{E}_{\mathsf{symm}}$ happens, so A will never hit the axis $x = 0$ in Phase I.

Proof. To see this, A can only have four types of movements in each step: lower-left $(x,y) \mapsto (x-1, y-1)$, up $(x,y) \mapsto (x, y+1)$, and right $(x,y) \mapsto (x+1, y)$. It is easy to see that, 1) A will never step across the line $y = x$ in one step, and 2) if A ever reaches the line $y = x$ at (w,w) for some w, then A must be at $(w, w-1)$ in the previous step. However, when A is at $(w, w-1)$, B should be at $(w-1, w)$ according to the relative position of A, B. In this case event $\mathcal{E}_{\mathsf{symm}}$ already happens. ◁

▷ Claim 18. $\mathcal{E}_{\mathsf{symm}}$ and $\mathcal{E}_{\mathsf{skew}}$ cannot happen simultaneously.

Proof. Suppose $\mathcal{E}_{\mathsf{symm}}$ and $\mathcal{E}_{\mathsf{skew}}$ happen at the same time, then it must be that A is at $(1,0)$ and B is at $(0,1)$, as the relative position of A and B is unchanged in Phase I, and this must be at the end of a certain *iteration*. In the previous iteration, A must be at $(2,1)$, since $\mathcal{E}_{\mathsf{skew}}$ did not happen yet and A is below the line $y = x$. However, B is at $(1,2)$ when A is at $(2,1)$, implying that case $\mathcal{E}_{\mathsf{symm}}$ has already happened in the previous iteration, which is a contradiction. ◁

▷ Claim 19. B cannot reach the axis $x = 0$ before either $\mathcal{E}_{\mathsf{symm}}$ or $\mathcal{E}_{\mathsf{skew}}$ happen.

Proof. If $\mathcal{E}_{\mathsf{symm}}$ happens before $\mathcal{E}_{\mathsf{skew}}$, B cannot reach the axis $x = 0$ before $\mathcal{E}_{\mathsf{symm}}$ as A is always below the line $y = x$ and B is always on the upper-left diagonal of A. If $\mathcal{E}_{\mathsf{skew}}$ happens before $\mathcal{E}_{\mathsf{symm}}$, B cannot reach the axis $x = 0$ before $\mathcal{E}_{\mathsf{skew}}$, or even by the time $\mathcal{E}_{\mathsf{skew}}$ happens: by the time $\mathcal{E}_{\mathsf{skew}}$ happens, A can only at one of $(2,0), (3,0), (4,0), \ldots$ (A cannot be at $(1,0)$, for otherwise $\mathcal{E}_{\mathsf{symm}}$ and $\mathcal{E}_{\mathsf{skew}}$ happen simultaneously, which is impossible as shown just now), in which case B will not be at the axis $x = 0$. ◁

4.2.2 Phase II Symm

Let A move in a way that is symmetric to B with respect to the line $y = x$: $J^A(i) = J^B(j)$, $\alpha^A(i,j) = \beta^B(i,j)$ and $\beta^A(i,j) = \alpha^B(i,j)$. Notice that, in Phase II Symm, A may cross the line $y = x$, after which A is above the line $y = x$ while B is below.

4.2.3 Phase II Skew

If event $\mathcal{E}_{\mathsf{skew}}$ happens, we need a more complicated coupling. Suppose Phase II Skew starts after iteration $\mathcal{T}_{\mathsf{skew}}$. Here we use $\mathcal{T}_S^A$ (and $\mathcal{T}_S^B$) to denote the hitting time of A (and B) to a set of states S which is the first iteration of the process into the set S. For example $i = \mathcal{T}_{y=1}^B$ is the hitting time of B such that $y^B(i) = 1$. Here we list six relevant hitting times and their relationship.

$$\mathcal{T}_{\mathsf{skew}} = \mathcal{T}_{y=1}^B = \mathcal{T}_{y=0}^A < \mathcal{T}_{y=0}^B, \text{ and } \mathcal{T}_{\mathsf{skew}} < \mathcal{T}_{x=0}^B = \mathcal{T}_{x=1}^A < \mathcal{T}_{x=0}^A.$$

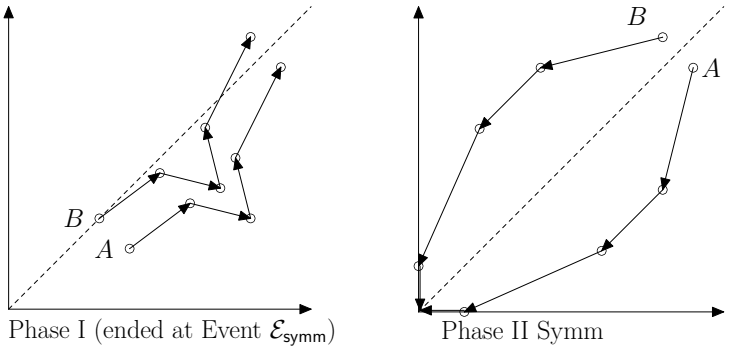

Figure 1 The coupling with Phase I ended at Event $\mathcal{E}_{\mathsf{symm}}$.

Back to the coupling, we first let the x-direction movement of A be the same with that of B. To be specific, in each iteration $\mathcal{T}_{\mathsf{skew}} < i \le \mathcal{T}^B_{x=0}$, set $J^A(i) = J^B(i)$. At step j, we set $\alpha^A(i,j) = \alpha^B(i,j)$ and $\beta^A(i,j) = 0$ ($\beta^A(i,j)$ is always 0 now, as A is stuck to the axis $y = 0$). Till now, the relative position of A and B in x-coordinate is preserved $x^A(i,j) = x^B(i,j) + 1$. Let $\mathcal{E}^*$ be the event that B reaches the axis $x = 0$, and let $\mathcal{E}^*$ happens at the end of iteration $\mathcal{T}^* = \mathcal{T}^B_{x=0}$. We further define $\Delta = \mathcal{T}^* - \mathcal{T}_{\mathsf{skew}}$ to be the additional time before $x^B = 0$ (if both stopping times exist), and $L = \mathcal{T}^B_{y=0} - \mathcal{T}_{\mathsf{skew}}$ to be the additional time before $y^B = 0$ (if both stopping times exist).

At the end of iteration $\mathcal{T}^*$, the positions for A is one unit to the right of the origin. That is $x^A(\mathcal{T}^*) = 1$ while $y^A(\mathcal{T}^*)\big) = 0$. Informally, we want to couple the movement of A from $(1,0)$ at $\mathcal{T}^*$ to the movement of B in the y-direction at $\mathcal{T}_{\mathsf{skew}}$ which is one unit above the axis at $y = 1$. Formally, starting at $(1,0)$, A is a 1-dimensional random walk on the axis $y = 0$, and we couple it to B in the following way.

- For each $t = 1, \dots, L$, we couple A's movement in the x direction at iteration $\mathcal{T}^* + t$ with B's movement Δ steps earlier in the y direction at iteration $\mathcal{T}^* + t - \Delta = \mathcal{T}_{\mathsf{skew}} + t$ such that $\alpha^A(\mathcal{T}^* + t) \sim \mathrm{Po}(\lambda(\mathcal{T}^* + t))$ and $\alpha^A(\mathcal{T}^* + t) \ge \beta^B(\mathcal{T}_{\mathsf{skew}} + t)$. [4]
- We do not couple A to B for future iterations after $\mathcal{T}^* + L$.

A key property of this coupling is that the x-coordinate of A at $\mathcal{T}^* + t$ is always greater or equal to the y-coordinate of B at iteration $\mathcal{T}_{\mathsf{skew}} + t$.

▷ Claim 20. For all $t = 1, \dots, L$, $x^A(\mathcal{T}^* + t) \ge y^B(\mathcal{T}_{\mathsf{skew}} + t)$.

Proof. We use induction. For the base case, we have $1 = x^A(\mathcal{T}^*) = y^B(\mathcal{T}_{\mathsf{skew}})$ from the definitions of $\mathcal{T}_{\mathsf{skew}}$ and $\mathcal{T}^*$. For the inductive case, $\alpha^A(\mathcal{T}^* + t) \ge \beta^B(\mathcal{T}_{\mathsf{skew}} + t)$ due to our coupling. ◁

[4] Here is an example of such a coupling. Consider iteration $i = \mathcal{T}^* + t$ for A, and we want to couple it with B's movement at iteration $\iota = \mathcal{T}_{\mathsf{skew}} + t$. Let $J^B(\iota)$ be the number of steps of B in the iteration ι which is not necessary equal to the number of steps of A after iteration $\mathcal{T}^*$. At step 0, we sample a non-negative integer $d(i) \sim \mathrm{Po}(2(\lambda(i) - \lambda_\iota))$ independent to $J^B(\iota)$, and set the number of steps of A to be $J^A(i) = J^B(\iota) + d(i)$. Then set $\alpha^A(i,0) = -1$ and $\beta(i,0)^A = 0$. At each step $j = 1, \dots, J^B(\iota)$, we set $(\alpha^A(i,j), \beta^A(i,j)) = (\beta^B_{\iota j}, 0)$. At the later steps $j = J^B(\iota) + 1, \dots, J^A(i)$, we set $(\alpha^A(i,j), \beta^A(i,j)) = (1,0)$ with probability 0.5, or $(0,0)$ otherwise.

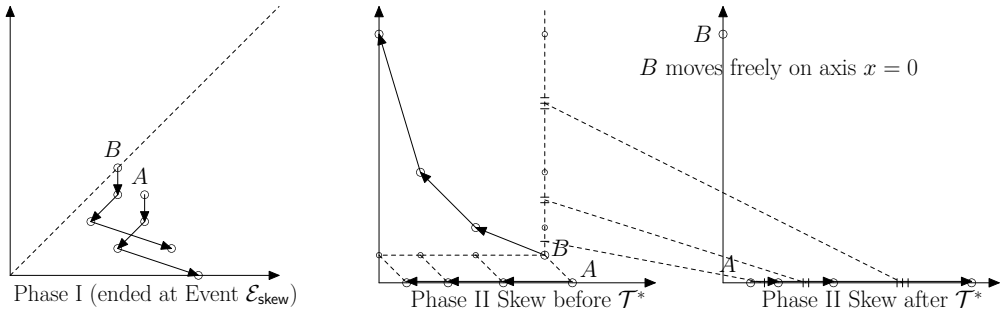

Figure 2 The coupling with Phase I ended at Event $\mathcal{E}_{\mathsf{skew}}$, if $\mathcal{E}^*$ happens.

4.3 Validity of the coupling

The coupling induces the correct marginal random walk process for B, as we have defined the coupling in a way that B is moving "freely" and A is being "coupled" with B. The only non-trivial part is to show that the coupling induces the correct marginal random walk process for A. It is straightforward to check that the marginal probabilities are correct during Phase I, before the event $\mathcal{E}^*$ occurs, or if the event $\mathcal{E}^*$ does not occur. If the process enters Phase II Skew and B reaches the axis $x = 0$, the movement of A in the x direction is coupled with B's movement in y direction $\Delta = \mathcal{T}^* - \mathcal{T}_{\mathsf{skew}}$ iterations ago. We note that B's movements in the x direction and the y direction are independent and A does not contain two iterations that are coupled to a same iteration of B. Therefore, the movements of A in x direction after $\mathcal{T}^*$ are independent to its previous movement, so the marginal distribution is correct. Fig. 3 illustrates the coupling time line.

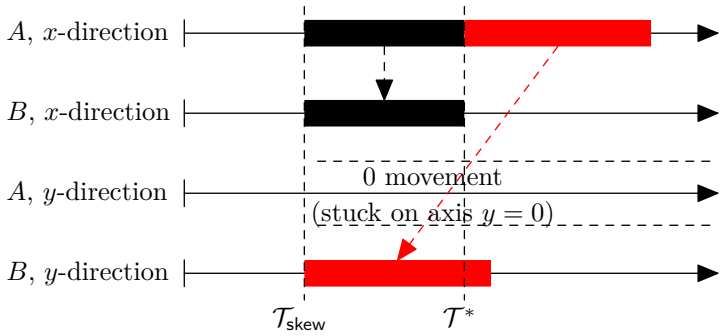

Figure 3 The time line for the coupling after event $\mathcal{E}_{\mathsf{skew}}$ happens.

▶ Remark 21. The coupling of the two random walks A and B in $\mathbb{Z}^2$ in the proof above can be alternatively viewed as a coupling of four independent random walks in $\mathbb{Z}$ (this is why we have said that "we simultaneously couple four cascade processes" in the introduction), as the x-directional and y-directional movements for both A and B correspond to the four terms in inequality (2), which are intrinsically independent.

4.4 Proof of Inequality (2)

It suffices to show that in our coupling $H_A \subseteq H_B$ and $H_B \setminus H_A$ is not empty, because this implies inequality (2): $\Pr(H_A) = \Pr(H_B \cap H_A) < \Pr(H_B \cap H_A) + \Pr(H_B \setminus H_A) = \Pr(H_B)$. We aim to show that:

1. if the coupling never moves to Phase II, neither A nor B reaches $(0,0)$;
2. if the coupling moves to Phase II Symm, A reaches $(0,0)$ if and only if B reaches $(0,0)$;
3. if the coupling moves to Phase II Skew, A reaches $(0,0)$ implies that B also reaches $(0,0)$;
4. with a positive probability, there is an event such that B reaches $(0,0)$ but A does not.

The first, second, and third show $H_A \subseteq H_B$. The last one shows $H_B \setminus H_A$ has a positive probability.

1 is trivial. 2 follows from symmetry.

To see 3, first notice that in Phase II Skew, $\mathcal{E}^*$ must happens if A ever reaches $(0,0)$: because A can move to the left by at most 1 unit in each iteration, A must first reach $(1,0)$, but at this point $x^B = 0$ and event $\mathcal{E}^*$ happens. Now consider the case that B never reaches the origin after event $\mathcal{E}^*$. Then the x movement of A remains coupled to the y-movement of B in such a way that $\bar{\alpha}^A(\mathcal{T}^* + t) \geq \bar{\beta}^B(\mathcal{T}_{\mathsf{skew}} + t)$. Walk A starts at $x^A = 1$, and walk B starts at $y^B = 1$. Therefore, A cannot reach the origin if B does not. In the case walk B meets the origin, the statement is vacuously true.

For 4, to show $\Pr(H_B \setminus H_A) > 0$, we define the following event which consists of four parts. i) For all $i = 1, \ldots, k$, it happens that $\alpha^A(i) = \beta^A(i) = 0$, in which case the event $\mathcal{E}_{\mathsf{skew}}$ happens at $\mathcal{T}_{\mathsf{skew}} = k$ and A reaches $(2,0)$. ii) For $i = k+1$, it happens that $\alpha^A(i) = 0$ and $\beta^B(i) = 1$, in which case A reaches $(1,0)$ and B reaches $(0,1)$, and the process B reaches the axis $x = 0$ at iteration $\mathcal{T}^* = k+1$. iii) In iteration $i = \mathcal{T}^* + 1$, it happens that $\beta^B(i) = 0$, so B reaches $(0,0)$. On the other hand, by the coupling $\alpha^A(\mathcal{T}^* + 1) \geq \beta^B(\mathcal{T}_{\mathsf{skew}} + 1) = 1$, so A does not reach $(0,0)$ at iteration $\mathcal{T}^* + 1 = k+2$. iv) Finally, it happens that $\alpha^A(i) \geq 1$ for all $i > k+2$. It is straightforward the i), ii), and iii) happen with positive probabilities. By direct computations, iv) happens with a positive probability as well.[5] Since the above event consisted of i), ii), iii) and iv) belongs to $H_B \setminus H_A$ and each of the four sub-events happens with a positive probability, 4 is implied.

From 2, 3, and 4, we learn that the probability that B reaches $(0,0)$ is strictly larger than that of A, which implies inequality (2) and concludes the proof.

References

1 Rico Angell and Grant Schoenebeck. Don't be greedy: leveraging community structure to find high quality seed sets for influence maximization. In *International Conference on Web and Internet Economics*, pages 16–29. Springer, 2017.

2 Lars Backstrom, Daniel P. Huttenlocher, Jon M. Kleinberg, and Xiangyang Lan. Group formation in large social networks: membership, growth, and evolution. In *ACM SIGKDD*, 2006.

3 Eric Balkanski, Nicole Immorlica, and Yaron Singer. The Importance of Communities for Learning to Influence. In *Advances in Neural Information Processing Systems*, pages 5862–5871, 2017.

4 Frank M Bass. A new product growth for model consumer durables. *Management science*, 15(5):215–227, 1969.

5 Dimitri P Bertsekas and John N Tsitsiklis. *Introduction to probability*, volume 1. Athena Scientific Belmont, MA, 2002.

6 Jacqueline Johnson Brown and Peter H Reingen. Social ties and word-of-mouth referral behavior. *Journal of Consumer research*, 14(3):350–362, 1987.

[5] The event that $\alpha^A(i) \geq 1$ for all $i > k+2$ happens with probability $\prod_{i>k+2} \Pr(\mathrm{Po}(\lambda(i)) \geq 1) = \prod_{i>k+2}(1 - \exp(-\lambda(i))) \geq \prod_{i \geq r+1}(1 - \exp(-\binom{i-1}{r-1}c^r))$ which is a positive constant depending on r and c.

7 Damon Centola and Michael Macy. Complex contagions and the weakness of long ties. *American journal of Sociology*, 113(3):702–734, 2007.

8 John Chalupa, Paul L Leath, and Gary R Reich. Bootstrap percolation on a Bethe lattice. *Journal of Physics C: Solid State Physics*, 12(1):L31, 1979.

9 Wei Chen, Tian Lin, Zihan Tan, Mingfei Zhao, and Xuren Zhou. Robust influence maximization. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 795–804. ACM, 2016.

10 Wei Chen, Yajun Wang, and Siyu Yang. Efficient influence maximization in social networks. In *ACM SIGKDD*, pages 199–208. ACM, 2009.

11 Wei Chen, Yifei Yuan, and Li Zhang. Scalable influence maximization in social networks under the linear threshold model. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 88–97. IEEE, 2010.

12 Aaron Clauset, Cristopher Moore, and Mark EJ Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98, 2008.

13 Edith Cohen, Daniel Delling, Thomas Pajor, and Renato F Werneck. Sketch-based influence maximization and computation: Scaling up with guarantees. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 629–638. ACM, 2014.

14 James Samuel Coleman, Elihu Katz, and Herbert Menzel. *Medical innovation: A diffusion study*. Bobbs-Merrill Co, 1966.

15 Paul DiMaggio. Structural analysis of organizational fields: A blockmodel approach. *Research in organizational behavior*, 1986.

16 John W Essam. Percolation theory. *Reports on Progress in Physics*, 43(7):833, 1980.

17 Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

18 Jacob Goldenberg, Barak Libai, and Eitan Muller. Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review*, 9(3):1–18, 2001.

19 Mark Granovetter. Threshold Models of Collective Behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978. URL: `http://www.journals.uchicago.edu/doi/abs/10.1086/226707`.

20 Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

21 Svante Janson, Tomasz Łuczak, Tatyana Turova, and Thomas Vallier. Bootstrap percolation on the random graph $G_{N,P}$. *The Annals of Applied Probability*, 22(5):1989–2047, 2012.

22 David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

23 David Kempe, Jon Kleinberg, and Éva Tardos. Influential nodes in a diffusion model for social networks. In *International Colloquium on Automata, Languages, and Programming*, pages 1127–1138. Springer, 2005.

24 Jure Leskovec, Lada A. Adamic, and Bernardo A. Huberman. The dynamics of viral marketing. In *EC*, pages 228–237, 2006. `doi:10.1145/1134707.1134732`.

25 Qiang Li, Wei Chen, Xiaoming Sun, and Jialin Zhang. Influence Maximization with ε-Almost Submodular Threshold Functions. In *NIPS*, pages 3804–3814, 2017.

26 Brendan Lucier, Joel Oren, and Yaron Singer. Influence at Scale: Distributed Computation of Complex Contagion in Networks. In *ACM SIGKDD*, pages 735–744. ACM, 2015.

27 John S MacDonald and Leatrice D MacDonald. Chain migration ethnic neighborhood formation and social networks. *The Milbank Memorial Fund Quarterly*, 42(1):82–97, 1964.

28 Vijay Mahajan, Eitan Muller, and Frank M Bass. New product diffusion models in marketing: A review and directions for research. In *Diffusion of technologies and social behavior*, pages 125–177. Springer, 1991.

29 Elchanan Mossel and Sébastien Roch. Submodularity of Influence in Social Networks: From Local to Global. *SIAM J. Comput.*, 39(6):2176–2188, 2010. doi:10.1137/080714452.

30 Elizabeth Levy Paluck, Hana Shepherd, and Peter M Aronow. Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, 113(3):566–571, 2016.

31 Daniel M Romero, Brendan Meeder, and Jon Kleinberg. Differences in the Mechanics of Information Diffusion Across Topics : Idioms , Political Hashtags , and Complex Contagion on Twitter. In *WWW*, pages 695–704. ACM, 2011. URL: http://dl.acm.org/citation.cfm?id=1963503.

32 Grant Schoenebeck and Biaoshuai Tao. Beyond Worst-Case (In)approximability of Nonsubmodular Influence Maximization. In *International Conference on Web and Internet Economics*, pages 368–382. Springer, 2017.

33 Grant Schoenebeck and Biaoshuai Tao. Beyond worst-case (in) approximability of nonsubmodular influence maximization. *ACM Transactions on Computation Theory (TOCT)*, 11(3):12, 2019.

34 Grant Schoenebeck and Biaoshuai Tao. Influence Maximization on Undirected Graphs: Towards Closing the (1-1/e) Gap. In *Proceedings of the 2019 ACM Conference on Economics and Computation, EC 2019, Phoenix, AZ, USA, June 24-28, 2019.*, pages 423–453, 2019. doi:10.1145/3328526.3329650.

35 Gordon Tullock. Toward a theory of the rent-seeking society, chapter Efficient rent seeking,(pp. 112), 1980.

36 Harrison C White, Scott A Boorman, and Ronald L Breiger. Social structure from multiple networks. I. Blockmodels of roles and positions. *American journal of sociology*, 81(4):730–780, 1976.

A Optimal seeds in submodular InfMax

We have seen that putting all the K seeds in a single leaf is optimal for r-complex contagion, when the root node has weight $\omega(1/n^{1+1/r})$. To demonstrate the sharp difference between r-complex contagion and a submodular cascade model, we present a submodular InfMax example where the optimal seeding strategy is to put no more than one seed in each leaf. The hierarchy tree T in our example meets all the assumptions we have made in the previous sections, including large communities, proper separation, and $w(\text{root}) = \omega(1/n^{1+1/r})$, where r is now an arbitrarily fixed integer with $r \geq 2$.

We consider a well-known submodular cascade model, *the independent cascade model* [22], where, after seeds are placed, each edge (u, v) in the graph appears with probability p_{uv} and vertices in all the connected components of the resultant graph that contain seeds are infected. In our example, the probability p_{uv} is the same for all edges, and it is $p = 1/n^{1-\frac{1}{4r}}$. The hierarchy tree T contains only two levels: a root and K leaves. The root has weight $1/n^{1+\frac{1}{2r}}$, and each leaf has weight 1. After $G \sim \mathcal{G}(n, T)$ is sampled and each edge in G is sampled with probability p, the probability that an edge appears between two vertices from different leaves is $(1/n^{1-\frac{1}{4r}}) \cdot (1/n^{1+\frac{1}{2r}}) = o(1/n^2)$, and the probability that an edge appears between two vertices from a same leaf is $1 \cdot (1/n^{1-\frac{1}{4r}}) = \omega(\log n/n)$. Therefore, with probability $1 - o(1)$, the resultant graph is a union of K connected components, each of which corresponds to a leaf of T. It is then straightforward to see that the optimal seeding strategy is to put a single seed in each leaf.

B A dynamic programming algorithm

In this section, we present an algorithm which finds an optimal seeding strategy when all $w(t)$'s fall into two regimes: $w(t) = \omega(1/n^{1+1/r})$ and $w(t) = o(1/n^2)$. We will assume this for $w(t)$'s throughout this section. Since a parent tree node always has less weight than its children (see Definition 1), we can decompose T into *the upper part* and *the lower part*, where the lower part consists of many subtrees whose roots have weights in $\omega(1/n^{1+1/r})$, and the upper part is a single tree containing only tree nodes with weights in $o(1/n^2)$ and whose leaves are the parents of those roots of the subtrees in the lower part. We call each subtree in the lower part a *maximal dense subtree* defined formally below.

▶ **Definition 22.** *Given a hierarchy tree $T = (V_T, E_T, w, v)$, a subtree rooted at $t \in V_T$ is a* maximal dense subtree *if $w(t) = \omega(1/n^{1+1/r})$, and either t is the root, or $w(t') = O(1/n^{1+1/r})$ where t' is the parent of t.*

Since we have assumed either $w(t) = \omega(1/n^{1+1/r})$ or $w(t) = o(1/n^2)$, $w(t') = O(1/n^{1+1/r})$ in the definition above implies $w(t') = o(1/n^2)$.

The idea of our algorithm is the following: firstly, after the decomposition of T into the upper and lower parts, we will show that the weights of the tree nodes in the upper part, falling into $w(t) = o(1/n^2)$, are negligible so that we can treat the whole tree T as a forest with only those maximal dense subtrees in the lower part (that is, we can remove the entire upper part from T); secondly, Theorem 11 shows that when we have decided the number of seeds to be allocated for each maximal dense subtree, the optimal seeding strategy is to put all the seeds together in a single leaf that has the highest density, where the density of a leaf $t \in L_T$ is defined in Definition 3; finally, the only problem remaining is how to allocate the K seeds among those maximal dense subtrees, and we decide this allocation by a dynamic programming approach.

Now, we are ready to describe our algorithm, presented in Algorithm 1.

The correctness of Algorithm 1 follows immediately from Theorem 23 (below) and Theorem 11. Recall Theorem 23 shows that we can ignore the upper part of T and treat T as the forest consisting of all the maximal dense subtrees of T when considering the InfMax problem. Theorem 11 shows that for each subtree T_i and given the number of seeds, the optimal seeding strategy is to put all the seeds on the leaf with the highest density.

▶ **Theorem 23.** *Given $T = (V_T, E_T, w, v)$, let $\{T_1, \ldots, T_m\}$ be the set of all T's maximal dense subtrees and let T^- be the forest consisting of $T_1, \ldots, T_m$. For any seeding strategy $\boldsymbol{k}$ and any $r \geq 2$, we have $\Sigma_{r,T}(\boldsymbol{k}) = \Sigma_{r,T^-}(\boldsymbol{k})$.*

Proof. Let $V(T_i)$ be the set of vertices corresponding to the subtree T_i. Since the total number of possible edges between those $V(T_i)$'s is upper bounded by n^2 and each edge appears with probability $o(1/n^2)$, the expected number of edges is $o(1)$. By Markov's inequality the probability there exists edges between those $V(T_i)$'s is $o(1)$. Therefore, we have

$$\frac{\mathop{\mathbb{E}}_{G\sim\mathcal{G}(n,T)}[\sigma_{r,G}(\boldsymbol{k})]}{n} = \frac{o(1)O(n) + (1-o(1))\mathop{\mathbb{E}}_{G\sim\mathcal{G}(n,T')}[\sigma_{r,G}(\boldsymbol{k})]}{n}.$$

Taking $n \to \infty$ concludes the proof. ◀

Finally, it is straightforward to see the time complexity of Algorithm 1, in terms of the number of evaluations of $\Sigma_{r,\mathcal{G}(n,T)}(\cdot)$.

▶ **Theorem 24.** *Algorithm 1 requires $O_I(|V_T|K^2)$ computations of $\Sigma_{r,\mathcal{G}(n,T)}(\cdot)$.*

Algorithm 1 The InfMax algorithm.

Input: $r \in \mathbb{Z}$ with $r \geq 2$, $T = (V_T, E_T, w, v)$, and $K \in \mathbb{Z}^+$
Find all maximal dense subtrees $T_1, \ldots, T_m$, and let $r_1, \ldots, r_m$ be their roots (Definition 22).
For each T_i and each $k = 0, 1, \ldots, K$, let $\boldsymbol{s}_i^*(k)$ be the seeding strategy that puts k seeds in the leaf $t \in L_{T_i}$ with the highest density, and let

$$h(T_i, k) = \lim_{n\to\infty} \frac{\mathbb{E}_{G\sim\mathcal{G}(v(r_i)\cdot n, T_i)}[\sigma_{r,G}(\boldsymbol{s}_i^*(k))]}{n}$$

be the expected number of infected vertices in the subgraph defined by T_i, normalized by the total number of vertices in the whole graph.
Let $S[i,k]$ store a seeding strategy that allocates k seeds in the first i subtrees $T_1, \ldots, T_i$, and let $H[i,k]$ be the expected total number of infected vertices corresponding to $S[i,k]$, divided by n.
for $k = 0, 1, \ldots, K$ **do**
  set $S[1,k] = \boldsymbol{s}_i^*(k)$ and $H[1,k] = h(T_1, k)$.
end for
for each $i = 2, \ldots, m$ **do**
  for $k = 0, 1, \ldots, K$ **do**
    $k_i = \operatorname{argmax}_{k_i \in \{0,1,\ldots,k\}} H[i-1, k-k_i] + h(T_i, k_i)$;
    set $S[i,k]$ be the strategy that allocates $k - k_i$ seeds among $T_1, \ldots, T_{i-1}$ according to $S[i-1, k-k_i]$ and puts the remaining k_i seeds in the leaf of T_i with the highest density;
    set $H[i,k] = H[i-1, k-k_i] + h(T_i, k_i)$;
  end for
end for
Output: the seeding strategy $S[m, K]$.

C Proof of Proposition 15

By Theorem 7 and Corollary 10, if no leaf is activated by the local seeds, then there can be at most constantly many infected vertices. Consider an arbitrary vertex v that is not infected, and let t be the leaf such that $v \in V(t)$. Let K_{in} be the number of infected vertices in $V(t)$ after Stage I and K_{out} be the number of infected vertices outside $V(t)$. By our assumption, $K_{in} = O(1)$ and $K_{out} = O(1)$. We compute an upper bound on the probability that v is infected in the next cascade iteration. Let X_v be the number of v's infected neighbors in $V(t)$ and Y_v be the number of v's infected neighbors outside $V(t)$.

Since the probability that v is connected to each of those K_{out} vertices is $o(n^{-1/r})$, we have

$$\Pr(Y_v \geq r - a) \leq \binom{K_{out}}{r-a} \left(o(n^{-1/r})\right)^{r-a} = o\left(n^{-(r-a)/r}\right)$$

for each $a \in \{0, 1, \ldots, r-1\}$.

Ideally, we would also like to claim that

$$\Pr(X_v \geq a) \leq \binom{K_{in}}{a} w(t)^a = O\left(n^{-a/r}\right), \tag{3}$$

so that putting together we have,

$$\Pr(v \text{ is infected}) \leq \sum_{a=0}^{r-1} \Pr(X_v \geq a) \Pr(Y_v \geq r-a) = r \cdot O\left(n^{-a/r}\right) \cdot o\left(n^{-(r-a)/r}\right) = o\left(\frac{1}{n}\right).$$

and conclude that the expected number of infected vertices in the next iteration is $o(1)$, which implies the proposition by the Markov's inequality.

However, conditioning on the cascade in $V(t)$ stopping after K_{in} infections, there is no guarantee that the probability an edge between v and one of the K_{in} infected vertices is still $w(t)$. Moreover, for any two vertices u_1, u_2 that belong to those K_{in} infected vertices, we do not even know if the probability that v connects to u_1 is still independent of the probability that v connects to u_2. Therefore, (3) does not hold in a straightforward way. The remaining part of this proof is dedicated to proving (3).

Consider a different scenario where we have put K_{in} seeds in $V(t)$ (instead of that the cascade in $V(t)$ ends at K_{in} infections), and let $\bar{X}_v$ be the number of edges between v and those K_{in} seeds (where v is not one of those seeds). Then we know each edge appears with probability $w(t)$ independently, and (3) holds for $\bar{X}_v$:

$$\Pr(\bar{X}_v \geq a) \leq \binom{K_{in}}{a} w(t)^a = O\left(n^{-a/r}\right).$$

Finally, (3) follows because $\bar{X}_v$ stochastically dominates X_v (i.e., $\Pr(\bar{X}_v \geq a) \geq \Pr(X_v \geq a)$ for each $a \in \{0, 1, \ldots, r-1\}$), which is easy to see:

$$\Pr\left(X_v \geq a\right) = \Pr\left(\bar{X}_v \geq a \mid \bar{X}_v \leq r-1\right) = \frac{\Pr(a \leq \bar{X}_v \leq r-1)}{\Pr(\bar{X}_v \leq r-1)}$$

$$= \frac{\Pr(\bar{X}_v \geq a) - \Pr(\bar{X}_v \geq r)}{1 - \Pr(\bar{X}_v \geq r)} \leq \Pr\left(\bar{X}_v \geq a\right),$$

where the first equality holds as $\Pr\left(\bar{X}_v \geq a \mid \bar{X}_v \leq r-1\right)$ exactly describes the probability that v has at least a infected neighbors among K_{in} conditioning on v not yet being infected.