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

SUBMISSION X

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*. We prove that, under certain mild assumptions, the optimal seeding strategy is to put all the seeds in a single community, when the graph is not exceptionally sparse, in particular, when each edge appears with probability $\omega(n^{-(1+1/r)})$. This matches the intuition that it is beneficial to put seeds near each other to maximize their synergy in a nonsubmodular cascade model, which is in sharp contrast to the seeding strategy in submodular cascade models (e.g., the independent cascade model and the linear threshold model).

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 $\omega(n^{-(1+1/r)})$ or in $o(n^{-2})$.

Manuscript submitted for review to ACM Economics & Computation 2019 (EC '19).

1 INTRODUCTION

A *cascade* is a fundamental process on social networks: starting with some seed agents, and the infection may spread to neighbors. How to place a finite number of initial seeds, to maximize the spread of the resulting cascade, is a natural question known as influence maximization [5, 6, 18, 29]. For example, which students can most effectively be enrolled in an intervention to decrease student conflict at a school [32].

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 [3, 14, 24, 28, 33]. 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 positions for certain settings which we will remark on shortly. In these settings, the optimal seeding strategy is to put all the seeds in a single community. This is arrestingly different than in the submodular setting, where the seeds can only erode each others' efficacy. Thus, in the submodular case, the optimal solutions tends to spread out the seeds, lest they erode each others' influence. We demonstrate this in Sect. 4 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 proposed by Angell and Schoenebeck [1], that it is better to target one market to saturate first (act locally) and then to allow the success in this initial market to drive broader success (think globally) rather than initially attempt to a scatter-shot approach (act globally). It also cautions that we must understand the particular nature of the contagion before blindly applying influence maximization tools.

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

We consider *the stochastic hierarchical blockmodel* [35] networks which are a special case of the stochastic blockmodel [15, 20, 37] 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 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 [12, 17]. The stochastic hierarchical blockmodel is rather general and captures other well-studied models (e.g. the Erdos-Renyi 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 the r-complex contagion, under certain mild technical assumptions, the optimal seeding strategy is to put all the seeds in a single community, if the probability that each edge appears satisfies $p_{uv} = \omega(n^{-(1+1/r)})$. Notice that this assumption 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 results is a novel coupling argument in Proposition 3.7. 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.

Our result is in sharp contrast to Balkanski et al.'s observation. Balkanski et al. [4] 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, 13, 27, 36].

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 approximation influence maximization to within an $O(n^{1-\epsilon})$ factor of optimal [23], and the inapproximability results have been extended to several more restrictive nonsubmodular models [9, 26, 35]. The intrinsic reason why nonsubmodular influence maximization is hard is that one needs to take into account the potential synergy of multiple seeds. This is in sharp contrast to submodular influence maximization, where the submodularity enables a seed-picker to consider placing seeds one at a time in a myopic way, as it is in the greedy algorithm.

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: performing dynamic programming from the bottom to the root in the tree-like community structure. On the other hand, Schoenebeck and Tao [35] show that the $\Omega(n^{1-\epsilon})$ inapproximability results extend to the setting where the networks are stochastic hierarchical blockmodels.

Result 2: However, our observation that, 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 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 MODEL

We study complex contagions on social networks with community structure. This section defines the notions of complex contagions and our model for social network 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 *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 $\sigma_{r,\mathcal{G}}(S) = \mathbb{E}_{G\sim\mathcal{G}}\left[\sigma_{r,G}(S)\right]$ 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 nondeterministic 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 and $S \subseteq T \subseteq V$ and any 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 [35]. The stochastic hierarchical blockmodel is a special case of the *stochastic blockmodel* [20]. Specifically and intuitively, the stochastic blockmodel is a stochastic graph model dealing with networks with community structures, and the stochastic hierarchical blockmodel further assumes that the communities form a hierarchical structure. Our definition in this section follows closely to (the full version of) [35].

Definition 2.1. A *stochastic hierarchical blockmodel* is a distribution $\mathcal{G} = (V, T)$ of unweighted undirected graphs sharing the same vertex set V, and $T = (V_T, E_T, w, v)$ is a weighted tree T called a *hierarchy tree*. The third parameter is the weight function $w : V_T \mapsto [0, 1]$ 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 nodes $V(t) \subseteq V$, and the $V(t)$ sets partition the nodes in V. In general, if $t \notin L_T$, we denote $V(t) = \bigcup_{t' \in L_T : t' \text{ is an offspring of } t} V(t')$. Let the function $v : V_T \mapsto 2^V$ in the forth parameter denote such correspondence.

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 "node" and "vertex" to refer to the vertices in V_T and V respectively. In Definition 2.1, $V(t) \subseteq V$, corresponded by the node $t \in V_T$, represents a community in the social network, and $V(t_1), V(t_2), \ldots \subseteq V$, corresponded by the children $t_1, t_2, \ldots$ of t, partition $V(t)$ into sub-communities. Naturally, the relation between two nodes is stronger if they are in a same sub-community in a lower level. This justifies our assumption $w(t_1) < w(t_2)$ for any $t_1, t_2 \in V_T$ such that t_1 is an ancestor of t_2.

To consider the algorithmic aspect of influence maximization problem prior setting, because the graph G is not revealed, nodes with same label are symmetric, we can measure the complexity with respect to the complexity of stochastic hierarchical block (how complicate the tree structure T is), instead of the size of the graph (how large n is). As a result, we consider the following variant of the stochastic hierarchical block model.

Definition 2.2. 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 2.1, except for the followings.

(1) Instead of mapping a node t to a weight in $[0, 1]$, the weight function w is a function $w : V_T \mapsto \mathcal{F}$, where $\mathcal{F} = \{f \mid f : \mathbb{Z}^+ \mapsto [0, 1]\}$ is the space of functions mapping an integer 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) For each node $t \in V_T$, we do not record the subset of vertices $V(t) \subseteq V$ that t corresponds to. Instead, we record a real number $v(t) \in (0, 1]$ which denotes the fraction $|V(t)|/n$. 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 all node $t \in V_T$, we only consider the value of $v(t)$ that is independent of n. As a result, for each $t \in V_T$, $|V(t)| = v(t)n = \Theta(n)$ tends to infinity as $n \to \infty$.

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.

The reason we define w and v in this seemly strange way is that we want to define $\mathcal{G}(n, T)$ in a way such that a same hierarchy tree $T = (V_T, E_T, w, v)$ is compatible with varying n. Thus, we need to encode each $w(t)$ and $v(t)$ in a way that is independent of V and n. The way we define w makes sure $w(t)$ is independent, and the large communities assumption above makes sure $v(t)$ is also independent. As we will see in the next subsection, we will take $n \to \infty$ when considering InfMax, and we would not like n or $|V|$ to be one of the inputs to the InfMax problem. As a result, T, being one of the inputs to the InfMax problem, should be independent of n. Notice that such property of the definition of InfMax enables us to consider graphs having exponentially many vertices.

2.3 The InfMax Problem

We study the r-complex contagion on the succinct stochastic hierarchical block model. 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 2.3. The *influence maximization problem* InfMax is an optimization problem which takes as inputs an integer r, a hierarchy tree $T = (V_T, E_T, w, v)$ in Definition 2.2, 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

$$\Sigma_{r,T}(\boldsymbol{k}) := \lim_{n \to \infty} \frac{\mathbb{E}_{G \sim \mathcal{G}(n,T)} \left[\sigma_{r,G}(S_{\boldsymbol{k}})\right]}{n},$$ [1]

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 1. In Definition 2.3, n is not the input of 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. For example, an algorithm runs in polynomial time would mean the running time is a polynomial of the input size, not of n. On the

[1] The purpose we divided the expected number of infected vertices by n is to avoid the infinite limit. However, as a result of this, our analysis naturally ignores lower order terms.

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, $K = \Theta_I(|V_T|^2)$ means the number of seeds is asymptotically the square of the tree size, while with respect to n we always have $r = \Theta(1)$, $K = \Theta(1)$ and $|V_T| = \Theta(1)$. As another example, the two asymptotic notions, $|V(t)| = v(t)n = \Theta(n)$ and $w(t_1) = o(w(t_2))$, appeared in the large communities and the proper separation assumptions are all in the scope of n, so we have not put the subscripts I under Θ and o.

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. For submodular InfMax, a simple greedy algorithm is known to achieve a $(1 - 1/e)$-approximation to the optimal influence [22, 23, 31].

2.4 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 2.4. The *Erdős-Rényi random graph* $\mathcal{G}(n, p)$ is a distribution of graphs with the same vertex set V with $|V| = n$ and we include an edge $(u, v) \in E$ with probability p for each pair of vertices u, v.

The InfMax problem in Definition 2.3 on $\mathcal{G}(n, p)$ is trivial, as there is only one possible allocation of those K seeds: allocate all the seeds to the single leaf node of T, which is the root. Therefore, $\sigma_{r,T}(\cdot)$ in Definition 2.3 depends only on the *number* of seeds $K = |\boldsymbol{k}|$, not on the seed allocation $\boldsymbol{k}$ itself. In this section, we slightly misuse the notation σ for 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 2.3), and let $\sigma_{r,\mathcal{G}(n,p)}(k)$ be the expected number of infected vertices after the cascade given k seeds. Correspondingly, let $\sigma_{r,G}(k)$ be the actual number of infected vertices after the graph G is sampled from $\mathcal{G}(n, p)$.

Theorem 2.5 (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 2.6 (Theorem 5.8 in [21]). *If $r \geq 2, p = \omega(n^{-1/r})$ and $k \geq r$, then $\Pr_{G\sim\mathcal{G}(n,p)}\left[\sigma_{r,G}(k) = n\right] = 1 - o(1)$.*

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 2.7 (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 $\xi_\ell \in \mathrm{Po}\left(\binom{\ell-1}{r-1}c^r\right)$, $\ell \geq 1$ 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 2.7, 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 2.8 (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 $\phi : \mathbb{Z}^+ \mapsto \mathbb{R}^+$ such that $\lim_{n\to\infty}\phi(n) = \infty$.

3 OUR MAIN RESULT

Before presenting our main result, we need the following definition which defines the *density* of a leaf.

Definition 3.1. Given a hierarchy tree $T = (V_T, E_T, w, v)$ and a leaf $t \in L_T$, the *density* of the leaf $\rho(t) = w(t) \cdot (v(t)n)^{1/r}$.

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 3.2. *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 $w(\text{root}) = \omega(1/n^{1+1/r})$. 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 $w(\text{root}) = \omega(1/n^{1+1/r})$ 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 $O(\log n)$. Therefore, if we were to use the stochastic hierarchical blockmodel to model the social networks in our daily life, we should expect that the root node has weight in $\Omega(1/n)$.

The remaining part of this section is dedicated to proving Theorem 3.2. We assume $w(\text{root}) = \omega(1/n^{1+1/r})$ in this section from now on. It is worthy noticing that, in many parts of this proof, and also in the proof of Theorem 5.2, 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 2.3.

Definition 3.3. Given $T = (V_T, E_T, w, v)$, a 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 the last section, 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 amount 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 2.7.

We say a 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 node is activated when $n \to \infty$. Notice that this is equivalent to at least one leaf being activated. The proof of Theorem 3.2 consists of two parts. We will first show that, $P_{\boldsymbol{k}}$ completely determines $\Sigma_{r,T}(\boldsymbol{k})$ (Lemma 3.4). Secondly, to maximize $P_{\boldsymbol{k}}$, an optimal seeding strategy is to put all seeds on a single leaf with the maximum density (Lemma 3.5).

Lemma 3.4. *Given any two seeding strategies $\boldsymbol{k}_1, \boldsymbol{k}_2$, if $P_{\boldsymbol{k}_1} \le P_{\boldsymbol{k}_2}$, then $\Sigma_{r,T}(\boldsymbol{k}_1) \le \Sigma_{r,T}(\boldsymbol{k}_2)$.*

Lemma 3.5. *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 3.4 and Lemma 3.5 imply Theorem 3.2.

3.1 Proof Sketch of Lemma 3.4

We sketch the proof here, and the full proof is in the appendix.

Proof (sketch). Let E be the event that at least one leaf (or node) is activated at the end of the cascade. Theorem 2.6 and Corollary 2.8 imply that the number of infected vertices in a critical or supercritical leaf t can only be either a constant or $v(t)n$, and Theorem 2.5 indicates that a subcritical leaf will not have $\omega(1)$ infected vertices by a constant number of seeds (as it is in our setting $K = \Theta(1)$). Therefore, if E does not happen, we only have $o(n)$ infected vertices in V, regardless of the seeding strategy.

If E happens, we can show that the expected total number of infected vertices does not vary significantly by different seeding strategies. If a leaf t_1 is activated, the probability that a vertex $v \in V(t_2)$ is infected due to the influence of $V(t_1)$ is at least

$$\binom{v(t_1)n}{r} w(t)^r (1 - w(t))^{v(t_1)n - r} = \omega\left(n^r \left(\frac{1}{n^{1+\frac{1}{r}}}\right)^r \cdot 1\right) = \omega\left(\frac{1}{n}\right),$$

where in the above equality we have further assumed $w(t) = o(1/n)$ (which can only further reduce the probability if $w(t)$ were in $\Omega(1/n)$) so that $(1 - w(t))^{v(t_1)n - r} = \omega((1 - 1/n)^{v(t_1)n}) = \omega(1)$. Thus, there are $\omega(1/n) \cdot \Theta(n) = \omega(1)$ infected vertices in $V(t_2)$. Theorem 2.6 and Corollary 2.8 show that t_2 will be activated if t_2 is critical or supercritical. Therefore, when E happens, all the critical and supercritical will be activated. As for subcritical leaves, the number of infected vertices may vary, but Theorem 2.5 intuitively suggests that adding a constant number of seeds is insignificant (we will handle this rigorously in the full proof). Therefore, the expected total number of infections equals to the number of vertices in all critical and supercritical leaves, plus the expected number of infected vertices in subcritical leaves which does not significantly depend on the seeding strategy $\boldsymbol{k}$.

In conclusion, the number of infected vertices only significantly depends on whether or not E happens. In particular, we have a fixed fraction of infected vertices whose size does not depend on $\boldsymbol{k}$ if E happens, and a negligible number of infected vertices if E does not happen. Therefore, $P_{\boldsymbol{k}}$ characterizes $\Sigma_{r,T}(\boldsymbol{k})$, and a larger $P_{\boldsymbol{k}}$ implies a larger $\Sigma_{r,T}(\boldsymbol{k})$. □

3.2 Proof of Lemma 3.5

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 $K \ge r$ seeds in this leaf, by Theorem 2.6, 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 $K \geq r$ and all the leaves are either critical or subcritical. Notice that, by the proper separation assumption, all internal nodes of T are subcritical.

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

Proposition 3.6 shows that maximizing $P_{\boldsymbol{k}}$ is equivalent to maximizing the probability that a leaf is activated in Phase 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 3.6. *If no leaf is activated after Phase I, then with probability* $1 - o(1)$ *no vertex will be infected in Phase II, i.e., the cascade will end after Phase I.*

We sketch the proof here, and the full proof is omitted due to the space limit.

Proof (sketch). Consider any critical leaf t and an arbitrary vertex $v \in V(t)$ that is not infected after Phase I. Let K_{in} be the number of infected vertices in $V(t)$ after Phase I, and K_{out} be the number of infected vertices in $V \setminus V(t)$. If no leaf is activated after Phase I, Theorem 2.5 and Corollary 2.8 suggest that $K_{in} = O(1)$ and $K_{out} = O(1)$. The probability that v is connected to any of the K_{in} infected vertices in $V(t)$ can only be less than $w(t) = \Theta(n^{-1/r})$ conditioning on the cascade inside $V(t)$ does not carry to v, so the probability that v has a infected neighbors in $V(t)$ is $O(n^{-a/r})$. On the other hand, the probability that v has $r - a$ neighbors among the K_{out} outside infected vertices is $o(n^{-(r-a)/r})$. Therefore, the probability that v is infected in the next iteration is $\sum_{a=0}^{r-1} O(n^{-a/r}) \cdot o(n^{-(r-a)/r}) = o(1/n)$, and the expected total number of vertices infected in the next iteration after Phase I is $o(1)$. The proposition follows from the Markov's inequality. □

Since Theorem 2.5 suggests that any constant number of seeds will not activate a subcritical leaf, we should only consider putting seeds in critical leaves. In Proposition 3.7, 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 3.7 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 effect is maximized, which intuitively implies Lemma 3.5.

Proposition 3.7. *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 *be the event that seeding the first* k *vertices does not make all the* n *vertices infected. We have* $\Pr(E_{k+2} \mid E_{k+1}) < \Pr(E_{k+1} \mid E_k)$ *for any* $k \geq r - 1$.

Proof. Since the event E_{k+1} implies E_k, we have

$$\Pr(E_{k+1} \mid E_k) = \frac{\Pr(E_{k+1} \wedge E_k)}{\Pr(E_k)} = \frac{\Pr(E_{k+1})}{\Pr(E_k)},$$

and the inequality we are proving is equivalent to

$$\Pr(E_{k+2}) \Pr(E_k) < \Pr(E_{k+1}) \Pr(E_{k+1}). \tag{2}$$

We adopt the inhomogeneous random walk interpretation in Theorem 2.7, and view E_k as the following process: the random walk starts at $x = k$; in the i-th iteration, x moves to the right by 1 unit, and moves to the left by $\alpha_i \sim \mathrm{Po}\left(\binom{i-1}{r-1} c^r\right)$ units; $\Pr(E_k)$ is then the probability that the random walk reaches $x = 0$. In this proof, we let $\lambda_i = \mathrm{Po}\left(\binom{i-1}{r-1} c^r\right)$, and in particular, $\lambda_0 = \lambda_1 = \cdots = \lambda_{r-1} = 0$.

To prove (2), we consider two random walks in $\mathbb{Z}^2$, starting at $(k+2,k)$ and $(k+1,k+1)$ respectively. In each iteration i, we move from (x,y) to $(x-1+\alpha_i, y-1+\beta_i)$ where α_i and β_i are sampled from $\mathrm{Po}(\lambda_i)$ independently. If we hit the x-axis after a certain iteration t, then we stick to the x-axis, i.e., for any $i > t$, the update in the i-th iteration is from $(x,0)$ to $(x-1+\alpha_i, 0)$; similarly, we stick to the y-axis and update $(0, y-1+\beta_i)$ if we hit the y-axis after certain iteration. Then, $\Pr(E_{k+2})\Pr(E_k)$ (or $\Pr(E_{k+1})\Pr(E_{k+1})$) is the probability that the random walk starting from $(k+2,k)$ (or $(k+1,k+1)$) reaches $(0,0)$. Let A be the random walk starting from $(k+2,k)$, and B be the random walk starting from $(k+1,k+1)$. To prove (2), we define a coupling between the two random walks such that, 1) whenever A reaches $(0,0)$, B also reaches $(0,0)$, and 2) with positive probability, B reaches $(0,0)$ but A fails to reach $(0,0)$.

Before we define the coupling, we reinterpret the random walk a little bit by breaking down each *iteration* i into T *steps*:

- at step 0 of iteration i, we update (x,y) to $(x-1,y-1)$ (or $(x-1,y)$ if $y=0$, $(x,y-1)$ if $x=0$);
- at each step j for $j = 1,\ldots,T$, sample two Bernoulli random valuables independently, $\alpha_{ij} \sim \mathrm{Be}(\lambda_i/T)$, $\beta_{ij} \sim \mathrm{Be}(\lambda_i/T)$, and update from (x,y) to $(x+\alpha_{ij}, y+\beta_{ij})$ (notice that in each step we have at most 1 unit movement in each direction, and we can only move further away from both the x-axis and y-axis);
- at the end of iteration i, if $x=0$ (or $y=0$), the random walk is stuck to the y-axis (or the x-axis) forever (notice that we only do this at the end of an iteration, i.e., after step T; in particular, if a random walk hits one of the x-axis or y-axis after step 0 in an iteration i, say, the x-axis, it will only stick to the x-axis if $\beta_{ij}=0$ for all $j=1,\ldots,T$).

Standard results from Poisson process indicate that, when $T\to\infty$, the effect of the T steps from 1 to T is equivalent as sampling $\alpha_i \sim \mathrm{Po}(\lambda_i)$ and $\beta_i \sim \mathrm{Po}(\lambda_i)$ (see, for example, Definition 8.4 and Theorem 8.7 in [30]).

Now we are ready to describe the coupling. Figure illustrations are available in Fig. 1 and Fig. 2. Let (x_{ij}^A, y_{ij}^A), and (x_{ij}^B, y_{ij}^B) be the coordinates for A and B respectively after iteration i step j. Similarly, let α_{ij}^A, and α_{ij}^B denote the x-direction movements of both walks, and β_{ij}^A, and β_{ij}^B denote the y-direction movements. The coupling consists of two phases.

Phase I A and B move in exactly the same way, i.e., $\alpha_{ij}^A = \alpha_{ij}^B$ and $\beta_{ij}^A = \beta_{ij}^B$, until one of the following two events happens.

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

Event (b) A hits the x-axis *at the end of an iteration*. Notice that this means A is then stuck to the x-axis forever. When (b) happens, we move on to Phase II(b)

It is important to notice that A is always below the line $y=x$ before (a) happens, so A will never hit the y-axis in Phase I. To see this, A can only have four types of movements in each step: lower-left $(x,y)\mapsto(x-1,y-1)$, up $(x,y)\mapsto(x,y+1)$, right $(x,y)\mapsto(x+1,y)$, and upper-right $(x,y)\mapsto(x+1,y+1)$. It is easy to see that, 1) A will never step across $y=x$ in one step, and 2) if A ever reaches the $y=x$ at (w,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 (a) already happens.

Phase II(a) A and B move in a way that is symmetric to the line $y=x$: $\alpha_{ij}^A = \beta_{ij}^B$ and $\beta_{ij}^A = \alpha_{ij}^B$. In this phase, by symmetry, A hits the x-axis if and only if B hits the y-axis, and A hits the y-axis if and only if B hits the x-axis. The coupling carries on when one of these happens. For example, if A hits the x-axis and B hits the y-axis, then we always have $\alpha_{ij}^A = \beta_{ij}^B$ and

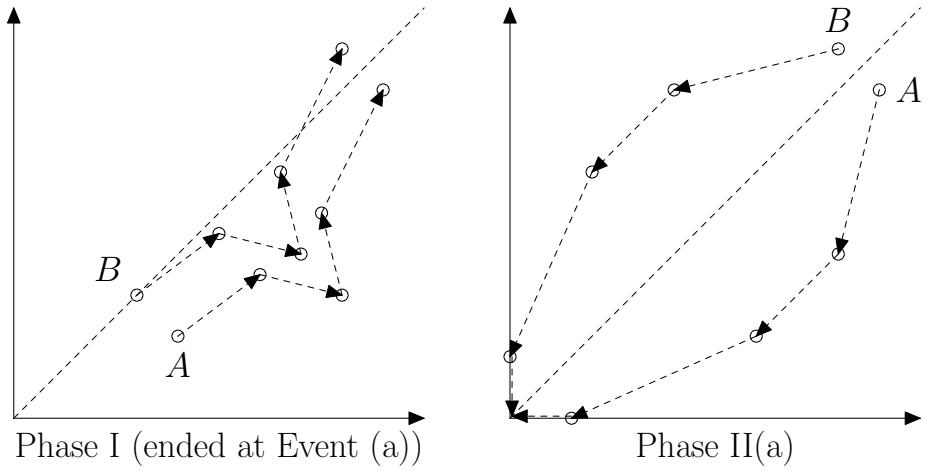

Fig. 1. The coupling with Phase I ended at Event (a)

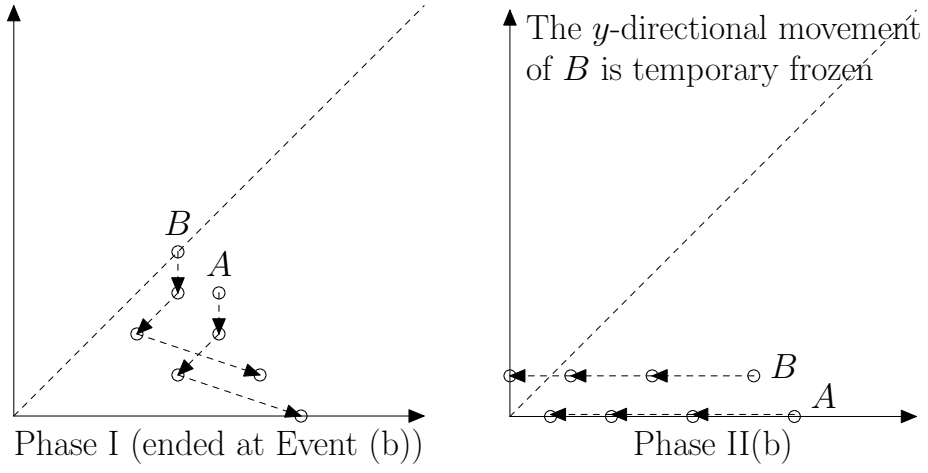

Fig. 2. The coupling with Phase I ended at Event (b)

$\beta^A_{ij} = \alpha^B_{ij} = 0$. Notice that, in Phase II(a), A may cross $y = x$, after which A is above $y = x$ while B is below.

Phase II(b) We temporary freeze B's movement in y-direction, and let the x-direction movements of A and B be the same. To be specific, suppose Phase II(b) starts at iteration i_0. In each iteration $i \geq i_0$ and step j, we couple $\alpha^A_{ij} = \alpha^B_{ij}$, and we do not sample $\beta^B_{ij} \sim \mathrm{Be}(\lambda_i/T)$ (β^A_{ij} is always 0 now, as A is stick to the x-axis). Till now, the relative position of A and B is always the same: $x^A_{ij} = x^B_{ij} + 1$ and $y^A_{ij} = y^B_{ij} - 1$. If B hits the y-axis at the end of an iteration i_1, we terminate the coupling. Notice that the current positions for A and B can only be that $(x^A, y^A) = (1, 0)$ and $(x^B, y^B) = (0, 1)$.

To prove Eqn. (2), it suffices to show that

(1) if the coupling moves to Phase II(a), A reaches $(0, 0)$ if and only if B reaches $(0, 0)$;
(2) with positive probability, the coupling moves to Phase II(b) and Phase II(b) terminates;

(3) if the coupling moves to Phase II(b) and Phase II(b) terminates, then B has strictly higher chance to reach $(0,0)$ than A.

By symmetry, 1 is trivial. With positive probability $\beta_{ij}^A = 0$ for all $i = 1,\dots,k$ and all $j = 1,\dots,T$ (this is the probability that we independently sample k Poisson random variables with means $\lambda_1,\dots,\lambda_k$ respectively, and obtain 0 for all of them), in which case event (b) happens, so 2 is also trivial. To see 3, by the time Phase II(b) terminates, A is stuck to the x-axis and one step away from $(0,0)$, while B is stuck to the y-axis and also one step away from $(0,0)$. Thus, we only need to consider a one-dimensional random walk for both A and B. In each future iteration $i_1 + t$ (for $t = 1,2,\dots$), A moves one step closer to $(0,0)$ and moves further by a distance sampled from $\mathrm{Po}(\lambda_{i_1+t})$, and meanwhile, taking into account that we have frozen B's y-direction movements between iterations i_0 and i_1, B also moves one step closer to $(0,0)$ and moves further by a distance sampled from $\mathrm{Po}(\lambda_{i_0+t})$. Since we have assumed $k \geq r-1$, it takes at least $r-1$ steps for A to reach the x-axis, so $i_0 \geq r-1$. Since λ_i is strictly increasing for $i \geq r-1$, we have $\lambda_{t_0+t} < \lambda_{t_1+t}$ for all $t \geq 0$. Therefore, in each future iteration after Phase II(b) terminates, the distance that B moves away from the destination is sampled from a Poisson distribution with mean strictly less than the mean of the Poisson distribution from which the distance that A moves away is sampled. This implies that B is strictly more likely to reach $(0,0)$ than A, which implies 3.

From 1, 2 and 3, we learn that the probability that B reaches $(0,0)$ is strictly larger than that of A, which implies (2) and concludes the proof. □

Equipped with Proposition 3.7, to show Lemma 3.5, 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 be the event that the first k seeds on t_1 fail to activate t_1 and F_k be the event that the first k seeds on t_2 fail to activate t_2. By Proposition 3.7, we have

$$\Pr(E_{K_1+1} \mid E_{K_1}) < \Pr(E_{K_1} \mid E_{K_1-1})$$

and

$$\Pr(F_{K_2+1} \mid F_{K_2}) < \Pr(F_{K_2} \mid F_{K_2-1}),$$

which implies

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

Therefore, we have either $\frac{\Pr(E_{K_1+1})\Pr(F_{K_2-1})}{\Pr(E_{K_1})\Pr(F_{K_2})}$ or $\frac{\Pr(E_{K_1-1})\Pr(F_{K_2+1})}{\Pr(E_{K_1})\Pr(F_{K_2})}$ 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 makes it less likely that none of t_1, t_2 is activated, or the strategy putting K_1-1 seeds on t_1 and K_2+1 seeds on t_2 makes it less likely that none of t_1, t_2 is activated, which implies that the strategy putting K_1 and K_2 seeds on t_1, t_2 respectively cannot be optimal. Therefore, in an optimal strategy, we 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 2.7, 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 3.5.

4 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.

5 A DYNAMIC PROGRAMMING ALGORITHM

In this section, we present an algorithm which finds an optimal seeding strategy when $w(t)$ satisfies either $w(t) = \omega(1/n^{1+1/r})$ or $w(t) = o(1/n^2)$ for each $t \in V_T$, and we will assume this for $w(t)$ throughout this section. Since a parent node always has less weight than its children (see Definition 2.1), 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 nodes with weight 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 5.1. 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 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 3.2 shows that when we have decide 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 3.1; 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 5.2 (below) and Theorem 3.2. Recall Theorem 5.2 shows that we can ignore the upper part of T and treat T as the forest consisting

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.
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 each $k = 0, 1, \ldots, K$, set $S[1, k] = \boldsymbol{s}_i^*(k)$ and $H[1, k] = h(T_1, k)$.
for each $i = 2, \ldots, m$, **do**
    for each $k = 0, 1, \ldots, K$, **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 $H[i, k] = H[i-1, k-k_i] + h(T_i, k_i)$;
    endfor
endfor
Output: the seeding strategy $S[m, K]$.

of all the maximal dense subtrees of T when considering the InfMax problem. Theorem 3.2 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 5.2. *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 consists 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

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

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

6 CONCLUSION

In this paper, we presented an influence maximization algorithm which finds optimal seeds for the stochastic hierarchical blockmodel, assuming the weights of tree nodes do not fall into a narrow regime between $\Omega(1/n^2)$ and $O(1/n^{1+1/r})$. As a crucial observation behind the algorithm, when the

root of the tree has weight $\omega(1/n^{1+1/r})$, our results show that the optimal seeding strategy is to put all the seeds together. Our results provide a formal verification for the intuition that one should put the seeds close to each other to maximize the synergy effect in a nonsubmodular cascade model.

Removing Limitations. One obvious future direction is to extend our algorithm such that it works for node weights between $\Omega(1/n^2)$ and $O(1/n^{1+1/r})$ as well. Related to this, Schoenebeck and Tao [35] shows that InfMax for the complex contagion on the stochastic hierarchical blockmodel is NP-hard to approximate to within factor $n^{1-\varepsilon}$ if vertices have non-homogeneous thresholds, i.e., each vertex v has a individual threshold $r_v \in \mathbb{Z}^+$ such that v is infected when it has at least r_v infected neighbors. It is unknown whether this inapproximability result carries over to the homogeneous case where all agents have the same threshold.

It is also interesting to see if our main result Theorem 3.2 still holds without the proper separation assumption. We only use this assumption in the proof of Proposition 3.6. To remove the proper separation assumption, more insight is needed on the behavior of the cascade in the critical leaves. As a next step for this, one might consider the case when leaves t_1 and t_2 have weights $c_1 n^{-1/r}$ and $c_2 n^{-1/r}$ respectively, and their parent t has weight $dn^{-1/r}$ with $d < c_1$ and $d < c_2$; it is an interesting open problem to see that if it is still optimal to put all the seeds in one of t_1, t_2. We conjecture this is true.

Future Work. One way to extend our results is to relax the assumption that the network is known. For example, can the network be learned from observing previous cascades, or by experimenting with them? Or, can they be elicited from agents with limited, local knowledge? Another direction would be to leverage these results to create heuristics that work well on real-world networks. A final direction would be more careful empirical studies (particularly experiments) about the nature of various cascades (e.g. submodular versus nonsubmodular.

REFERENCES

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

[2] Richard Arratia, Larry Goldstein, and Louis Gordon. 1990. Poisson approximation and the Chen-Stein method. *Statist. Sci.* (1990), 403–424.

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

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

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

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

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

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

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

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

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

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

[13] Edith Cohen, Daniel Delling, Thomas Pajor, and Renato F Werneck. 2014. 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*. ACM, 629–638.

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

[15] Paul DiMaggio. 1986. Structural analysis of organizational fields: A blockmodel approach. *Research in organizational behavior* (1986).

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

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

[18] Jacob Goldenberg, Barak Libai, and Eitan Muller. 2001. 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 (2001), 1–18.

[19] Mark Granovetter. 1978. Threshold Models of Collective Behavior. *Amer. J. Sociology* 83, 6 (1978), 1420–1443. http://www.journals.uchicago.edu/doi/abs/10.1086/226707

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

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

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

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

[24] Jure Leskovec, Lada A. Adamic, and Bernardo A. Huberman. 2006. The dynamics of viral marketing. In *EC*. 228–237.

[25] David A Levin and Yuval Peres. 2017. *Markov chains and mixing times*. Vol. 107. American Mathematical Soc.

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

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

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

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

[30] Michael Mitzenmacher and Eli Upfal. 2005. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge university press.

[31] Elchanan Mossel and Sébastien Roch. 2010. Submodularity of Influence in Social Networks: From Local to Global. *SIAM J. Comput.* 39, 6 (2010), 2176–2188.

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

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

[34] Nathan Ross et al. 2011. Fundamentals of Stein's method. *Probability Surveys* 8 (2011), 210–293.

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

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

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

A FULL PROOFS

A.1 Proof of Lemma 3.4

Proposition A.1. *Suppose the root of T has weight $\omega(1/n^{1+1/r})$ and consider a leaf t. If there are $\Theta(n)$ infected vertices in $V \setminus V(t)$, then these infected vertices outside $V(t)$ will infect $\omega(1)$ vertices in $V(t)$ with probability $1 - o(1)$.*

Proof. Let $X = \Theta(n)$ be the number of infected vertices in $V \setminus V(t)$. For each $u \in V(t)$ and $v \in V \setminus V(t)$, we assume that the probability p_{uv} that the edge (u, v) appears satisfies $p_{uv} = \omega(1/n^{1+1/r})$ and $p_{uv} = o(1/n)$, where $p_{uv} = \omega(1/n^{1+1/r})$ holds since the root of T has weight $\omega(1/n^{1+1/r})$, and assuming $p_{uv} = o(1/n)$ may only decrease the number of infected vertices in $V(t)$ if the least common ancestor of the two leaves containing u and v has weight $\Omega(1/n)$. Let p be the minimum probability among those p_{uv}'s, and we further assume that each edge (u, v) appears with probability p, which again may only reduce the number of infected vertices in $V(t)$.

For each vertex $u \in V(t)$, by only accounting the probability that it has exactly r neighbors among those X outside infected vertices, the probability that u is infected is at least

$$\rho := \binom{X}{r} p^r (1-p)^{X-r} = \omega\left(n^r \cdot \left(\frac{1}{n^{1+1/r}}\right)^r \left(1 - \frac{1}{n}\right)^n\right) = \omega\left(\frac{1}{n}\right),$$

and the expected number of infected vertices in $V(t)$ is $v(t)n \cdot \rho = \omega(1)$.

Let Y be the number of vertices in $V(t)$ that are infected due to the influence of $V \setminus V(t)$, so we have $\mathbb{E}[Y] = v(t)n\rho$. Applying Chebyshev's inequality,

$$\begin{aligned}\Pr\left(Y \le \frac{1}{2}v(t)n\rho\right) &\le \Pr\left(|Y - \mathbb{E}[Y]| \ge \frac{1}{2}v(t)n\rho\right) \\ &\le \frac{\mathrm{Var}(Y)}{(\frac{1}{2}v(t)n\rho)^2} = \frac{v(t)n\rho(1-\rho)}{\frac{1}{4}v(t)^2n^2\rho^2} = o(1),\end{aligned}$$

where we have used the fact that $n\rho = \omega(1)$ and the variance of the Binomial random variable with parameter n, p is $np(1-p)$. Therefore, with probability $1 - o(1)$, the number of infected vertices in $V(t)$ is at least $\frac{1}{2}v(t)n\rho = \omega(1)$. □

Let E be the event that at least one leaf (or node) is activated at the end of the cascade. By our definition, $P_{\boldsymbol{k}} = \lim_{n\to\infty} \Pr(E)$. Given a seeding strategy $\boldsymbol{k}$, let $\sigma(\boldsymbol{k}) := \mathbb{E}_{G\sim\mathcal{G}(n,T)}[\sigma_{r,G}(\boldsymbol{k})]$ be the expected number of infected vertices, $\sigma(\boldsymbol{k} \mid E) := \mathbb{E}_{G\sim\mathcal{G}(n,T)}[\sigma_{r,G}(\boldsymbol{k}) \mid E]$ be the expected number of infected vertices conditioning on event E happens, and $\sigma(\boldsymbol{k} \mid \neg E) := \mathbb{E}_{G\sim\mathcal{G}(n,T)}[\sigma_{r,G}(\boldsymbol{k}) \mid \neg E]$ be the expected number of infected vertices conditioning on E does not happen. We have

$$\sigma(\boldsymbol{k}) = \Pr(E) \cdot \sigma(\boldsymbol{k} \mid E) + (1 - \Pr(E)) \cdot \sigma(\boldsymbol{k} \mid \neg E),$$

and

$$\begin{aligned}\Sigma_{r,T}(\boldsymbol{k}) &= \lim_{n\to\infty} \frac{\sigma(\boldsymbol{k})}{n} \\ &= P_{\boldsymbol{k}} \cdot \lim_{n\to\infty} \frac{\sigma(\boldsymbol{k} \mid E)}{n} + (1 - P_{\boldsymbol{k}}) \cdot \lim_{n\to\infty} \frac{\sigma(\boldsymbol{k} \mid \neg E)}{n}. \qquad (3)\end{aligned}$$

First, we show that $\sigma(\boldsymbol{k} \mid \neg E) = o(n)$, so the second term in (3) is always 0. If there is no critical or supercritical leaf in T, given that the total number of seeds $K = \Theta(1)$ is a constant, Theorem 2.5 suggests that there can be at most $2K = \Theta(1)$ infected vertices. To be specific, we can take the maximum weight $w^*(t)$ over all the leaves, and assume the entire graph is the Erdős-Rényi graph $\mathcal{G}(n, w^*(t))$. This makes the graph denser, so the expected number of infected vertices increases. However, even in this case, Theorem 2.5 implies that the total number of infected vertices is less than $2K$. If there is at least one critical or supercritical leaf t, for the sake of contradiction we

assume the total number of infected vertices is $\Theta(n)$ and E does not happen. Since the number of leaves is a constant, there exists $t' \in L_T$ such that the number of infected vertices in $V(t')$ is $\Theta(n)$. Theorem 2.6 and Corollary 2.8 indicate that, with probability $1 - o(1)$, the number of infected vertices in $V(t)$ is either a constant or $v(t)n$. Therefore, if $t' = t$, with probability $o(1)$, those $\Theta(n)$ infected vertices in $V(t)$ will not activate t, and this probability becomes 0 after taking the limit $n \to \infty$, which makes no contribution to the second term in (3). If $t' \neq t$, let $X = \Theta(n)$ be such that with probability $1 - o(1)$ the number of infected vertices in $V(t')$ is more than X, then the total number of vertices in $V(t)$ that are infected by those X vertices in $V(t')$ is $\omega(1)$ according to Proposition A.1. Theorem 2.6 and Corollary 2.8 suggest that those $\omega(1)$ infected vertices in $V(t)$ will further spread and activate t, which again contradicts to that E does not happen.

Secondly, to conclude the proof, it remains to show that the first term in (3) only depends on $P_{\boldsymbol{k}}$, or $\sigma(\boldsymbol{k} \mid E) = cn + o(n)$ for some constant c which does not depend on $\boldsymbol{k}$. As an intuitive argument, Proposition A.1, Theorem 2.6 and Corollary 2.8 suggest that, when E happens, a single activated leaf will activate all the critical and supercritical leaves, and the number of vertices corresponding to all the critical and supercritical leaves is fixed and independent of $\boldsymbol{k}$; based on the tree structure and the number of infected outside vertices, the number of infected vertices in a subcritical leaf may vary; however, we will see that the seeding strategy $\boldsymbol{k}$, adding only a constant number of infections, is too weak to affect the number of infected vertices in a subcritical leaf.

To break it down, we first show that all critical and supercritical leaves will be activated if E happens. This is straightforward: Proposition A.1 shows that an activated leaf can cause $\omega(1)$ infected vertices in every other leaf, and Theorem 2.6 and Corollary 2.8 indicate that those critical and supercritical leaves will be activated by those $\omega(1)$ infected vertices.

Lastly, assuming all critical and supercritical leaves are activated, we show that the number of infected vertices in any subcritical leaf does not depend on $\boldsymbol{k}$. We do not need to worry about those seeds that are put in the critical or supercritical leaves, as all vertices in those leaves will be infected later. As a result, we only need to show that a constant number of seeds in subcritical leaves has negligible effect to the cascade.

We say a subcritical leaf t is *vulnerable* if there exists a criticial or supercritical leaf t' such that the least common ancestor of t and t' has weight $\Omega(1/n)$, and we say t is *not-very-vulnerable* otherwise. It is easy to see that a vulnerable leaf t will always be activated, even if no seed is put into it. Since each $v \in V(t)$ is connected to one of the $v(t')n$ vertices in $V(t')$ with probability $\Omega(1/n)$, the number of infected neighbors of v follows a Binomial distribution with parameter $(v(t')n, p)$ where $p = \Omega(1/n)$. We only consider $p = \Theta(1/n)$, as there can only be more infected vertices if $p = \omega(1/n)$. If $p = \Theta(1/n)$, the Binomial distribution becomes a Poisson distribution with a constant mean λ for $n \to \infty$. In this case, with constant probability $e^{-\lambda}\frac{\lambda^r}{r!}$, v has r infected neighbors. Therefore, v will be infected with constant probability, and $V(t)$ has $\Theta(n)$ vertices that are infected by $V(t')$ outside. The second part of Theorem 2.5 suggests that, these $\Theta(n)$ infected vertices will further spread and activate t. Therefore, the seeds on those vulnerable subcritical leaves have no effect, since vulnerable subcritical leaves will be activated regardless the seeding strategy.

Let $t_1, \ldots, t_M$ be all the not-very-vulnerable subcritical leaves. Suppose we are at the stage of the cascade process where all those critical, supercritical and vulnerable subcritical leaves have already been activated (as they will with probability $1 - o(1)$ since we assumed that E has happened) and we are revealing the edges between $V \setminus \bigcup_{m=1}^{M} V(t_m)$ and $\bigcup_{m=1}^{M} V(t_m)$ to consider the cascade process in $\bigcup_{m=1}^{M} V(t_m)$. For each $i = 0, 1, \ldots, r-1$ and each $m = 1, \ldots, M$, let χ_i^m be the number of vertices in $V(t_m)$ that have *exactly* i infected neighbors among $V \setminus \bigcup_{m=1}^{M} V(t_m)$, which can be viewed as a random variable. For each $m = 1, \ldots, M$, let χ_r^m be the number of vertices in $V(t_m)$ that have

at least r infected neighbors. If there are K_m seeds in $V(t_m)$, we increase the value of χ_r^m by K_m. Let $\boldsymbol{\chi}^m = (\chi_0^m, \chi_1^m, \ldots, \chi_r^m)$. Since $(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)$ completely characterize the expected number of infected vertices in the subcritical leaves, we let $\sigma(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)$ be the total number of infected vertices in the subcritical leaves, given $(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)$. We aim to show that adding $K_1, \ldots, K_M$ seeds in $V(t_1), \ldots, V(t_M)$ only changes the number of infected vertices by $o(n)$. Let $(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)$ correspond to the case where no seed is added, and $(\bar{\boldsymbol{\chi}}^1, \ldots, \bar{\boldsymbol{\chi}}^M)$ correspond to the case where K_m seeds are added to t_m for each $m = 1, \ldots, M$. The outline of the proof is that, we first show that the total variation distance of the two distributions $(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)$ and $(\bar{\boldsymbol{\chi}}^1, \ldots, \bar{\boldsymbol{\chi}}^M)$ is $o(1)$; then we show that $\sigma(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)$ and $\sigma(\bar{\boldsymbol{\chi}}^1, \ldots, \bar{\boldsymbol{\chi}}^M)$ can only differ by $o(n)$ in expectation.

To show the first claim, noticing that M is a constant and $\boldsymbol{\chi}^{m_1}$ is independent of $\boldsymbol{\chi}^{m_2}$ for any m_1, m_2 (the appearances of edges between $V(t_{m_1})$ and $V \setminus \bigcup_{m=1}^M V(t_m)$ are independent of the appearances of edges between $V(t_{m_2})$ and $V \setminus \bigcup_{m=1}^M V(t_m)$), it is sufficient to show that the total variation distance between $\boldsymbol{\chi}^m$ and $\bar{\boldsymbol{\chi}}^m$ is $o(1)$. For each vertex $v \in V(t_m)$, it is connected to an arbitrary vertex in a critical or supercritical leaf with probability between $\omega(1/n^{1+1/r})$ (since the root has weight $\omega(1/n^{1+1/r})$) and $o(1/n)$ (otherwise t_m is vulnerable). Since the number of infected vertices in $V \setminus \bigcup_{m=1}^M V(t_m)$ is $\Theta(n)$, the number of v's infected neighbors follows a Binomial distribution, $\mathrm{Bin}(n, \theta)$, with mean $n\theta$ between $\omega(1/n^{1/r})$ and $o(1)$, we can use Poisson distribution $\mathrm{Po}(n\theta)$ to approximate it. Formally, the total variation distance is $d_{TV}(\mathrm{Bin}(n, \theta), \mathrm{Po}(n\theta)) \le n\theta^2 = o(1/n)$. Thus, this approximation only changes the total variation distance of $\boldsymbol{\chi}^m$ by $o(1)$. Observing these, the proposition below shows the total variation distance between $\boldsymbol{\chi}^m$ and $\bar{\boldsymbol{\chi}}^m$ is $o(1)$.

Proposition A.2. *Let λ be such that $\lambda = \omega(1/n^{1/r})$ and $\lambda = o(1)$. Let $Y_1, \ldots, Y_n \in \mathbb{Z}$ be n independently and identically distributed random variables where each Y_i is sampled from a Poisson distribution with mean λ, Let $Z_1, \ldots, Z_n \in \mathbb{Z}$ be n random variables where $Z_1 = \cdots = Z_K = r$ with probability 1, and $Z_{K+1}, \ldots, Z_n$ are independently sampled from a Poisson distribution with mean λ. For $i = 0, 1, \ldots, r-1$, let χ_i be the numbers of random variables in $\{Y_1, \ldots, Y_n\}$ that have value i, and $\bar{\chi}_i$ be the numbers of random variables in $\{Z_1, \ldots, Z_n\}$ that have value i. Let χ_r be the numbers of random variables in $\{Y_1, \ldots, Y_n\}$ that have values at least r, and $\bar{\chi}_r$ be the numbers of random variables in $\{Z_1, \ldots, Z_n\}$ that have values at least r. The total variation distance between $\boldsymbol{\chi} = (\chi_0, \chi_1, \ldots, \chi_r)$ and $\bar{\boldsymbol{\chi}} = (\bar{\chi}_0, \bar{\chi}_1, \ldots, \bar{\chi}_r)$ is $d_{TV}(\boldsymbol{\chi}, \bar{\boldsymbol{\chi}}) = o(1)$ if $K = \Theta(1)$.*

To show this for random vectors χ and $\bar{\chi}$ that have small total variation distance, we use some straightforward computations and the Poisson approximation [2, 34]. We first decouple the correlation between $\boldsymbol{\chi} = (\chi_0, \chi_1, \ldots, \chi_r)$, and consider $r+1$ coordinate-wise independent Poisson $\boldsymbol{\zeta} = (\zeta_0, \zeta_1, \ldots, \zeta_r)$ with the same expectation $\mathbb{E}[\boldsymbol{\chi}] = \mathbb{E}[\boldsymbol{\zeta}]$. Then we define $\bar{\boldsymbol{\zeta}}$ similarly. Finally the total variational distance between two coordinate-wise independent Poisson vectors is well studied in the literature. The full proof of this proposition is omitted due to the space limit.

To show the second claim, notice that the range of the function $\sigma(\cdot)$ falls into the interval $[0, n]$. The total variation distance of $(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)$ and $(\bar{\boldsymbol{\chi}}^1, \ldots, \bar{\boldsymbol{\chi}}^M)$ being $o(1)$ implies that

$$\left| \mathop{\mathbb{E}}_{(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)} [\sigma(\boldsymbol{\chi}^1, \ldots, \boldsymbol{\chi}^M)] - \mathop{\mathbb{E}}_{(\bar{\boldsymbol{\chi}}^1, \ldots, \bar{\boldsymbol{\chi}}^M)} [\sigma(\bar{\boldsymbol{\chi}}^1, \ldots, \bar{\boldsymbol{\chi}}^M)] \right| = o(n),$$

by a standard property of total variation distance (see, for example, Proposition 4.5 in [25]).

This concludes that the seeds on subcritical leaves can only affect $o(n)$ infections.

Adding together, $\sigma(\boldsymbol{k} \mid E)$ equals to the number of vertices in all critical and supercritical leaves which is independent of the seeding strategy, plus the expected number of infected vertices in those subcritical leaves for which different seeding strategies only affect a value in $o(n)$. This implies that $\lim_{n\to\infty} \frac{\sigma(\boldsymbol{k}|E)}{n}$ in the first term of (3) does not depends on $\boldsymbol{k}$. Therefore, Eqn. (3) reveals that $\Sigma_{r,T}(\boldsymbol{k})$ is proportional to $P_{\boldsymbol{k}}$, which implies the lemma.