Received: 11 May 2023 | Revised: 9 September 2024 | Accepted: 9 September 2024

DOI: 10.1002/jgt.23184

ARTICLE

WILEY

Switchover phenomenon for general graphs

Dániel Keliger[1] | **László Lovász**[2] | **Tamás Ferenc Móri**[2] | **Gergely Ódor**[2,3]

[1]Department of Stochastics, Institute of Mathematics, Budapest University of Technology and Economics, Budapest, Hungary

[2]HUN-REN Alfréd Rényi Institute of Mathematics, Budapest, Hungary

[3]Department of Network and Data Science, Central European University, Vienna, Austria

Correspondence
László Lovász, HUN-REN Alfréd Rényi Institute of Mathematics,
H-1053 Budapest, Hungary.
Email: lovasz@renyi.hu

Funding information
European Research Council; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung; Dynasnet ERC Synergy, Grant/Award Number: ERC-2018-SYG 810115; Swiss National Science Foundation, Grant/Award Number: P500PT-211129

Abstract

We study SIR-type epidemics (susceptible-infected-resistant) on graphs in two scenarios: (i) when the initial infections start from a well-connected central region and (ii) when initial infections are distributed uniformly. Previously, Ódor et al. demonstrated on a few random graph models that the expectation of the total number of infections undergoes a switchover phenomenon; the central region is more dangerous for small infection rates, while for large rates, the uniform seeding is expected to infect more nodes. We rigorously prove this claim under mild, deterministic assumptions on the underlying graph. If we further assume that the central region has a large enough expansion, the second moment of the degree distribution is bounded and the number of initial infections is comparable to the number of vertices, the difference between the two scenarios is shown to be macroscopic.

KEYWORDS

graph percolation, SIR-type epidemics, switchover phenomenon

1 | INTRODUCTION

We study the propagation of a disease on a network, and in particular the "switchover" phenomenon established in [8, 9]. Informally, the phenomenon means the following. We have a network (describing the network of interactions of people in a country), which has a denser "central region" and a sparser "periphery." We compare the total number of nodes that get infected if a given number of seeds (initial infections) are distributed uniformly and randomly in the central region and in the whole graph, respectively. The switchover phenomenon means that for a low infection rate, an epidemic starting in the central region is worse (results in a larger epidemic), but this switches over so that the epidemics starting uniformly over the whole country are worse.

Ódor et al. [8] motivated the central seeding scenario by the emergence of a new pathogen that enters a country via international travel, and the uniform seeding scenario by a pathogen that is already present in various locations of the country, but causes a new epidemic wave due to changes in seasonality. Ódor et al. [8] have also shown by simulation that the switchover phenomenon occurs in many networks (not all), and established it rigorously for some very simple networks. In [9], some mathematical conditions were formulated (without proof), under which the switchover phenomenon occurs. The goal of this paper is to generalize those results and prove them mathematically.

Our model for the spread of infection is the SIR(1) model (which is one of the simplest). In this model, we have a finite graph G. A node can be in one of three states: susceptible (S), infected (I), or resistant (R). At each step, if a susceptible node has an infected neighbor, then it gets infected by this neighbor with probability β. If it has several infected neighbors, then the events that these infect the node are independent. The node becomes infected if at least one of its infected neighbors infects it. An infected node recovers deterministically after one step, and will be resistant from then on, which means that it does not infect and cannot be infected. If you think of a time scale where one step as a week, then this may be a reasonable assumption; every event (getting infected and then passing it on) is recorded on a weekly scale.

The main advantage of the SIR(1) model for us is that it is equivalent with a percolation problem. A proof of this simple observation was given in [9]. Briefly, it is not hard to see that we can decide about each edge in advance, independently and with probability β, whether it is going to pass on the infection, at any time when one of its endpoints is infected and the other one is susceptible. Our model guarantees that every edge has at most one chance to be in this situation. In other words, we keep every edge with probability β and delete the remaining edges; this way we get an edge-percolated graph G^β. For a seed set S, we denote by $G^\beta(S)$ the union of those components of G^β that contain at least one node of S. Then $|G^\beta(S)|$ nodes will be infected at one point during the epidemic in total.

Our goal is to compare the expectations of $|G^\beta(\mathbf{S}_1)|$ and $|G^\beta(\mathbf{S}_2)|$, where $\mathbf{S}_1$ is a random subset of the central region and $\mathbf{S}_2$ is a random subset of the whole node set. In Section 2.2, we show that under quite general conditions on G and for very small β, seeding the central region is worse, but for β very near to 1, seeding the whole graph uniformly is worse (Theorem 2.3). The conditions on G are intuitive and deterministically checkable, as they mainly ensure that the central region has a higher average degree than the entire graph, and exclude sparse cuts in G. However, such values of β are unlikely to occur in real life, and also the differences in epidemic sizes are minuscule. We call this "weak switchover," and we give its formal definition in Section 2.1.

In Section 2.3, we formulate conditions on the graph under which we can work with values of β in a more reasonable range, and we can establish that the difference between the sizes of the epidemics starting from $\mathbf{S}_1$ and $\mathbf{S}_2$ is of the same order of magnitude as the whole graph (we call this "strong switchover"). These conditions on the graph (Theorem 2.14) are tighter than for weak switchover, but they are still reasonable, and can be deterministically checked in real networks. Besides the condition that the central region has a higher average degree than the entire graph, we require the central region to satisfy an edge expansion property for large sets [1]. While most real networks are not expected to be expander graphs, we expect that the central region will have this property, depending on how it is selected. We do not make additional assumptions on the selection of the central region in this paper, but we note that [8] defined it as the nodes of the largest k-core in a commuting network between settlements of a country, which often turns out to be the capital city and its agglomeration—an almost fully connected network. Our assumptions are intended to be natural and minimal, but we suspect that similar results can be proven with slightly different assumptions. However, we expect that any set of assumptions that prove similar results must control the prevalence of small cuts, as the epidemic size is very sensitive to bottlenecks in the graph.

Finally, as an application of our results in Section 2.3, we prove that weak switchover occurs on Chung–Lu random graphs with power-law degree distribution [4] in Section 4. This result was stated in [8], along with a nonrigorous proof.

1.1 | Relationship with distribution-free graph models

Epidemics are often studied either theoretically or by simulation on random graph models [7]. In this paper, our goal is different: we aim to find deterministic conditions on the graph, which give rise to the switchover phenomenon (in expectation, where the randomness only comes from the epidemic or percolation process). Such combinatorial results, which are studied with a network science application in mind, are called *distribution-free* in the literature [6]. The main advantage of the distribution-free approach is that deterministic conditions can be verified on real networks, as opposed to the results on random graph distributions, where we can only hope that the results also apply to real networks. Moreover, one can go from results with deterministic conditions to results on random graphs relatively easily (as we do in Section 4), whereas going in the opposite direction seems much more difficult.

Proving facts that hold with high probability for random graphs for deterministic graphs with appropriate properties goes back (at least) to the study of quasirandom graphs [5]. In the network science setting, the study of distribution-free graph models was started by Fox et al. [6], and several papers followed. We refer to [10] for a review. The deterministic constraints studied in this topic include conditions on the triadic closure [6], on heterogeneous degree distributions [3] and on the expansion properties [2] of the graphs. While one of our main conditions is also a deterministic expansion property (a stronger one than in [2]), our conditions and proof techniques are different from all previous papers that we are aware of in this topic.

2 | RESULTS

2.1 | Notation and setup

Let $G = (V, E)$ be a simple graph on n nodes. We use the notation

$$|G| = |V| = n.$$

For a subset $S \subseteq V$, $G(S)$ denotes the union of connected components meeting S. As usual, we denote by $G[S]$ the subgraph induced by S. $e(K, L)$ stands for the number of edges between $K, L \subseteq V$. The average degree and the second moment of the set $K \subseteq V$ is denoted by

$$\overline{\deg}(K) := \frac{1}{|K|}\sum_{v \in K} \deg(v),$$
$$\overline{\deg^2}(K) := \frac{1}{|K|}\sum_{v \in K} \deg^2(v).$$

The path visiting vertices $v_1, v_2, \ldots, v_k \in V$ is denoted by $v_1 v_2 \ldots v_k$. For neighboring vertices $u, v \in V$ we write $u \sim v$ and for $K \subseteq V$, $\mathcal{N}(K)$ stands for $\{v \in V \backslash K | \exists\, u \in K : v \sim u\}$, that is, the neighborhood of K.

We consider graphs with a specified subset $C \subseteq V$ (modeling the *central region*) of size

$$|C| = r = cn.$$

Here, $0 < c < 1$ is considered to be “macroscopic,” and C will be denser than average in a sense to be defined later. Throughout, we use the notation $G_1 = G[C]$ and $G_2 = G \backslash E(G_1)$.

For $0 \le \beta \le 1$, G^β denotes the percolation of G with edge retention probability β; in other words, the graph is obtained by selecting each edge of G independently with probability β, and deleting the unselected edges.

Usually, the set $S \subseteq V$ represents a *deterministic* seed of initial infections. We will be interested in *random* seeds

$$\mathbf{S} \sim \mathrm{Uni}(L, k)$$

sampled uniformly from the k-subsets of a set $L \subseteq V$ for some

$$k = sn$$

($0 < s < c$). We think of L as a macroscopic subset; typical choices are $L = V$ and $L = C$. The corresponding random subsets for $L = C$ and $L = V$ are $\mathbf{S}_C \sim \mathrm{Uni}(C, k)$ and $\mathbf{S}_V \sim \mathrm{Uni}(V, k)$.

In our considerations, we generate the random graph G^β and the seed set $\mathbf{S}$ independently. We let $\mathbb{P}_\mathbf{S}$ and $\mathbb{E}_\mathbf{S}$ denote the probability and expectation if only the seed $\mathbf{S}$ is randomized, and define $\mathbb{P}_\beta$ and $\mathbb{E}_\beta$ analogously when only the graph G^β is randomized. We use no subscript if probability and expectation are taken over both random choices.

Now we come to our two main definitions.

Definition 2.1. We say the graph G exhibits a *weak switchover phenomenon* with seed size k ($1 \le k \le |C|$), if there are $\beta_1, \beta_2 \in (0, 1)$ such that for $\mathbf{S}_C \sim \mathrm{Uni}(C, k)$, and $\mathbf{S}_V \sim \mathrm{Uni}(V, k)$ we have

$$\mathbb{E}(|G^{\beta_1}(\mathbf{S}_C)|) > \mathbb{E}(|G^{\beta_1}(\mathbf{S}_V)|),$$

but

$$\mathbb{E}(|G^{\beta_2}(\mathbf{S}_C)|) < \mathbb{E}(|G^{\beta_2}(\mathbf{S}_V)|).$$

Note that Definition 2.1 only requires that there is some difference between $\mathbb{E}(|G^\beta(\mathbf{S}_C)|)$ and $\mathbb{E}(|G^\beta(\mathbf{S}_V)|)$, where this difference could be small, even vanishing as $n \to \infty$. In a more robust version, we require these differences to constitute a positive fraction of the whole population. To make an exact definition, we need to consider a sequence of graphs whose size tends to infinity:

Definition 2.2. We say the sequence of graphs (G_n, C_n) exhibits a *strong switchover phenomenon* with seed sizes k_n, if there are real numbers $\delta > 0$, $0 < \beta_1(n), \beta_2(n) < 1$ such that for $\mathbf{S}_{n,V} \sim \text{Uni}(V(G_n), k_n)$, and $\mathbf{S}_{n,C} \sim \text{Uni}(C_n, k_n)$ we have

$$\mathbb{E}(|G^{\beta_1}(\mathbf{S}_{n,V})|) \geq \mathbb{E}(|G^{\beta_1}(\mathbf{S}_{n,C})|) + \delta|V(G_n)|,$$

but

$$\mathbb{E}(|G^{\beta_2}(\mathbf{S}_{n,V})|) \leq \mathbb{E}(|G^{\beta_2}(\mathbf{S}_{n,C})|) - \delta|V(G_n)|$$

for large enough n.

2.2 | Weak switchover

We start by elementary remarks concerning the cases when $\beta \to 0$ and $\beta \to 1$. It is clear that if $\beta \to 0$, then $\mathbb{E}(|G^\beta(S)|) \to |S|$, while if $\beta \to 1$, then $\mathbb{E}(|G^\beta(S)|) \to n$ for every nonempty set S and connected G.

Note that in Section 2.2 all the $O(\cdot)$ notations depend on the fixed graph G and thus on n. For uniform bounds under stronger assumptions see Section 2.3

The case of small β is straightforward, since the seeds and those nodes reached in one step will dominate. The probability that a particular path of length 2 is retained in G^β is at most β^2, so with probability $1 - O(\beta^2)$, only neighbors of S get infected, and each such neighbor is infected by only one seed (here O refers to $\beta \to 0$). Hence for any subset $S \subseteq V$,

$$\mathbb{E}(|G^\beta(S)|) = |S| + \beta e(S, V\backslash S) + O(\beta^2). \tag{1}$$

The asymptotics at $\beta \to 1$ is more complicated. Assume that G has the (mild) property that

$(*)$ G has minimum degree d, it is not d-regular and the only edge-cuts in G with at most d edges are the stars of minimum degree nodes.

Let $Y \subseteq V$ be the set of nodes with degree d. Set $\gamma = 1 - \beta$. With probability at least $1 - O(\gamma^{d+1})$, at most d edges of G are missing in G^β. By $(*)$, in this case G^β is either a connected spanning subgraph of G, or it has a single isolated node in Y. The probability of the latter event is γ^d for any given node in Y.

This implies that with probability at least $1 - O(\gamma^{d+1})$, for every set $S \subseteq V$, $|S| \geq 2$, the infected graph $G^\beta(S)$ will miss at most one node in $Y \backslash S$. Hence

$$\mathbb{E}_\beta(|G^\beta(S)|) = n - |Y\backslash S|\gamma^d + O(\gamma^{d+1}). \tag{2}$$

Formulas (1) and (2) imply:

Theorem 2.3. *Let G be a connected graph, and $S_1, S_2 \subseteq V$, $|S_1| = |S_2|$.*

(a) *If $e(S_1, V \setminus S_1) > e(S_2, V \setminus S_2)$ and β is sufficiently close to* 0, *then* $\mathbb{E}(|G^\beta(S_1)|) > \mathbb{E}(|G^\beta(S_2)|)$.

(b) *If G has property* $(*)$, $|S_1 \cap Y| > |S_2 \cap Y|$, *and β is sufficiently close to* 1, *then* $\mathbb{E}(|G^\beta(S_1)|) < \mathbb{E}(|G^\beta(S_2)|)$.

Coming to random seed sets, it will be easy to derive from (1) and (2) the following.

Theorem 2.4. *Let G be a connected graph and $2 \le k < r$.*

(a) *If*

$$\frac{r-k}{r-1}\overline{\deg}(C) > \frac{n-k}{n-1}\overline{\deg}(V),$$

then $\mathbb{E}(|G^\beta(\mathbf{S}_C)|) > \mathbb{E}(|G^\beta(\mathbf{S}_V)|)$ *if β is sufficiently close to* 0.

(b) *If G has property* $(*)$ *and*

$$\frac{|Y \cap C|}{r} < \frac{|Y|}{n},$$

then $\mathbb{E}(|G^\beta(\mathbf{S}_C)|) < \mathbb{E}(|G^\beta(\mathbf{S}_V)|)$ *if β is sufficiently close to* 1.

Remark 2.5. For fixed c and small enough s it is enough to assume $\overline{\deg}(C) > \overline{\deg}(V)$ for part (a) of Theorem 2.4 as

$$1 \le \frac{r-1}{r-k}\frac{n-k}{n-1} = 1 + O(s).$$

Corollary 2.6. *If both conditions (a) and (b) above are satisfied, then G exhibits the weak switchover phenomenon for seed sets of size k.*

Note that both conditions say that C has larger degrees than average.

In conclusion, a weak switchover phenomenon occurs for all graphs under very mild hypotheses, but for unrealistically extreme values of β, and leading only to minuscule differences. Our goal in Section 2.3 is to exhibit a strong switchover with much more reasonable values of β.

2.3 | Strong switchover

To establish the case of small β for strong switchover is similar to the analogous case for weak switchover: again seeds and their neighbors will play the main role. We have to do more careful estimates, involving the spectrum of G. Our main tool is the following refined version of (2).

Lemma 2.7. *Let $L \subseteq V$, $m = |L|$, and let* $\mathbf{S}$ *be a random k-subset of L. Then*

$$\mathbb{E}(|G^{\beta}(\mathbf{S})|) = k + k\left(\overline{\deg}(L) - \frac{k-1}{m-1}\frac{1}{m}e(L,L)\right)\beta + R,$$

where

$$|R| \leq \overline{\deg^2}(V)\beta^2 n.$$

Applying this lemma with $L = V$ and $L = C$, we will get that for an appropriate β, seeding the central region is substantially more dangerous than seeding the whole node set. More exactly:

Corollary 2.8. *Assume that*

$$\frac{r-k}{r-1}\overline{\deg}(C) - \frac{n-k}{n-1}\overline{\deg}(V) \geq c_1 > 0. \tag{3}$$

Let

$$0 < \beta \leq \frac{1}{4}\frac{c_1}{\overline{\deg^2}(V)}s. \tag{4}$$

Then

$$\mathbb{E}(|G^{\beta}(\mathbf{S}_C)|) - \mathbb{E}(|G^{\beta}(\mathbf{S}_V)|) \geq \frac{1}{2}c_1\beta s n.$$

Remark 2.9. Similarly to part (a) of Theorem 2.4 it is enough to ensure that $\overline{\deg}(C) > \overline{\deg}(V)$ uniformly for (3) when s is small enough.

Ensuring the large β case for strong switchover is more involved and requires further assumptions regarding the graph G. More precisely, we assume edge expansion of the central region instead of large average degree.

Definition 2.10. We say that a graph $G = (V, E)$ has *edge-expansion* (a, q) with some $a > 0$ and $0 < q < \frac{1}{2}$, if for every set $X \subset V$, $qn < |X| \leq n/2$, the number of edges between X and $V \backslash X$ is at least $a|X|$.

Remark 2.11. Note that the parameters a, q might not be optimal. If $a_1 \leq a_2$, $q_1 \leq q_2$ and G has edge-expansion (a_2, q_1), then it is also true that G has edge-expansion (a_1, q_2).

A very similar property was studied in the context of epidemic spreading in [1] with the notable difference that we do not assume the expander property on the entire graph, but only on the central region. The following lemma shows that if the central region has a large enough expansion, the epidemic will produce more infections from a uniform seeding when β is close to 1.

Lemma 2.12. *Let b be the average degree of nodes of $V\setminus C$ in G, and assume G_1 has edge expansion (a,q) with $q<1/3$. Then*

$$\begin{aligned} &\mathbb{E}(|G^{\beta}(\mathbf{S}_V)|)-\mathbb{E}(|G^{\beta}(\mathbf{S}_C)|)\\ \geq\ & s(1-c)(1-\beta)^b n-\frac{c}{c-s}qn-\left(1+\frac{c}{c-s}\right)n\rho^r-n\mathrm{e}^{-2ck/3}, \end{aligned} \tag{5}$$

where $\rho:=\left(\frac{e(1-\beta)^a}{q}\right)^q$.

Remark 2.13. When G_1 has edge expansion (a,q) with $a>b$ and $q=(1+\varepsilon)e(1-\beta)^a$ for some $\varepsilon>0$ we end up with $0\leq\rho<1$ resulting in

$$\left(1+\frac{c}{c-s}\right)n\rho^r+n\mathrm{e}^{-2ck/3}=o(n)$$

for all fixed $0<\beta<1$. Furthermore, as $0\leq b<a$ it is possible to set $0<\beta<1$ to a value for which $q<\frac{1}{3}$ and

$$s(1-c)(1-\beta)^b>\frac{c}{c-s}(1+\varepsilon)e(1-\beta)^a=\frac{c}{c-s}q,$$

thus, there is a $\delta>0$ such that $\mathbb{E}(|G^{\beta}(\mathbf{S}_V)|)>\mathbb{E}(|G^{\beta}(\mathbf{S}_C)|)+\delta n$ for large enough n.

Our main result concerning strong switchover will easily follow by a combination of Lemmas 2.7 and 2.12.

Theorem 2.14. *Let $(G_m:m=1,2,\ldots)$ be a sequence of such that $n_m=|V(G_m)|\to\infty$. Let $C_m\subseteq V(G_m)$ so that $|C_m|=c_m n_m$ and let b_m denote the average degree in G_m of nodes in $V(G_m)\setminus C_m$ with some uniform bound $b_m\leq b_{max}$.*

Assume there is a $\varepsilon>0$ such that:

- $c_m\leq 1-\varepsilon$,
- $\varepsilon\leq s_m\leq(1-c_m)c_m/2$,
- *For any $0<q<\frac{1}{3}$, $G_m[C_m]$ has edge-expansion $(b_{max}+\varepsilon,q)$ when m is large enough.*

Also, assume the second moments of the degrees are uniformly bounded.

Then the graph sequence $((G_m,C_m):m=1,2,\ldots)$ exhibits the strong switchover phenomenon for seed sizes $s_m n_m$.

3 | PROOFS

3.1 | Overview or the proofs

In this section we present the main idea of the proofs.

For weak switchover we use the fact that for small β merely the neighborhood of the seed gets infected while for large β most vertices end up being in that state and the vertices which are most likely to stay susceptible are the ones with the lowest degree. Therefore, if the central region has high average degree and the low degree vertices are in the periphery, we expect switchover to occur.

The small β case of the strong switchover proof is similar to the weak switchover proof, however, we have to upper-bound the error coming from infections reaching distance 2 or further. This error can be controlled via the second moment of the average degree as the number of length 2 paths going through vertex v is $\binom{d(v)}{2}$.

More involved techniques are required for large values of β. First, we show that via the Chernoff–Hoeffding bound that assuming $G_1 = G[C]$ has a large set expansion for sets of size $qr \leq |X| \leq \frac{r}{2}$ then the largest component H^C in the percolated graph $G^\beta[C]$ will have size at least $(1-q)r$ with probability exponentially close to 1. (The convergence is faster when β is closer to 1.) H^C is the main contributor for the number of infections when the seeding is from the central region. However, with large probability at least one seed will be placed there during the uniform seeding as well, while uniform seeding can also catch vertices which are not on C and isolated in G^β giving it an advantage over the seeding from the central region.

Lastly, in the application of our results on the Chung–Lu graph, we show large set expansion with high expansion rate a in the unusual setting when we only focus on an $o(n)$ sized subset of the highest degree nodes—the central region in this model. To our knowledge, this setting has not been studied before. The proof here, once again, relies on the Chernoff–Hoeffding bound.

3.2 | Weak switchover

We start with a simple identity, which will imply that when choosing a "small" random seed $\mathbf{S}$ compared to V, it becomes unlikely that two vertices in $\mathbf{S}$ are adjacent to each other, and hence

$$\mathbb{E}(e(\mathbf{S}, V\backslash\mathbf{S})) \approx \mathbb{E}(e(\mathbf{S}, V)) = k\,\overline{\deg}(L).$$

Lemma 3.1. *Let* $\mathbf{S}$ *be a random* k*-element subset of* L*. Then*

$$\mathbb{E}_{\mathbf{S}}(e(\mathbf{S}, V\backslash\mathbf{S})) = k\left(\overline{\deg}(L) - \frac{k-1}{m-1}\frac{1}{m}e(L,L)\right).$$

The last (error) term can be estimated as

$$\frac{k-1}{m-1}\frac{1}{m}e(L,L) < \frac{k}{m}\overline{\deg}(L).$$

Proof.

$$\begin{aligned}\mathbb{E}(e(\mathbf{S}, V\backslash\mathbf{S})) &= \mathbb{E}(e(\mathbf{S}, V)) - \mathbb{E}(e(\mathbf{S}, \mathbf{S}))\\ &= \frac{k}{m}e(L,V) - \frac{k(k-1)}{m(m-1)}e(L,L) = k\left(\overline{\deg}(L) - \frac{k-1}{m-1}\frac{1}{m}e(L,L)\right).\end{aligned}$$

This proves the lemma. □

Proof of Theorem 2.4. We start with the small β case.

$$\begin{aligned}&\mathbb{E}(|G^{\beta}(\mathbf{S}_C)|)-\mathbb{E}(|G^{\beta}(\mathbf{S}_V)|)\\ =\ &(\mathbb{E}(e(\mathbf{S}_C, V\backslash\mathbf{S}_C))-\mathbb{E}(e(S_V, V\backslash\mathbf{S}_v)))\beta+O(\beta^2).\end{aligned}$$

Due to Lemma 3.1 the leading term can be bounded as

$$\begin{aligned}&\mathbb{E}(e(\mathbf{S}_C, V\backslash\mathbf{S}_C))-\mathbb{E}(e(\mathbf{S}_V, V\backslash\mathbf{S}_V))\\ =\ &k\left(\overline{\deg}(C)-\frac{k-1}{r-1}\frac{1}{r}e(C,C)-\overline{\deg}(V)+\frac{k-1}{n-1}\frac{1}{n}e(V,V)\right)\\ \geq\ &k\left(\frac{r-k}{r-1}\overline{\deg}(C)-\frac{n-k}{n-1}\overline{\deg}(V)\right)>0,\end{aligned}$$

making $\mathbb{E}(|G^{\beta}(\mathbf{S}_V)|)>\mathbb{E}(|G^{\beta}(\mathbf{S}_C)|)$ for sufficiently small β.

As for small $\gamma=1-\beta$

$$\begin{aligned}\mathbb{E}(|G^{\beta}(\mathbf{S}_V)|)-\mathbb{E}(|G^{\beta}(\mathbf{S}_C)|)&=(\mathbb{E}(|Y\backslash\mathbf{S}_C|)-\mathbb{E}(|Y\backslash\mathbf{S}_V|))\gamma^d+O(\gamma^{d+1})\\ &=k\left[\left(1-\frac{|Y\cap C|}{|C|}\right)-\left(1-\frac{|Y|}{|V|}\right)\right]\gamma^d+O(\gamma^{d+1})\\ &=k\left(\underbrace{\frac{|Y|}{|V|}-\frac{|Y\cap C|}{|C|}}_{>0}\right)\gamma^d+O(\gamma^{d+1}),\end{aligned}$$

implying $\mathbb{E}(|G^{\beta}(\mathbf{S}_V)|)>\mathbb{E}(|G^{\beta}(\mathbf{S}_C)|)$ when β is close to 1. □

3.3 | Strong switchover

3.3.1 | Lemmas for small β

We are going to prove Lemma 2.7 in the following slightly stronger form:

Lemma 3.2. *Let $L\subseteq V$, $m=|L|$, and let $\mathbf{S}$ be a random k-subset of L. Then*

$$\mathbb{E}(|G^{\beta}(\mathbf{S})|)=k+k\left(\overline{\deg}(L)-\frac{k-1}{m-1}\frac{1}{m}e(L,L)\right)\beta+R,$$

where

$$-\frac{1}{2}(\overline{\deg^2}(V)-\overline{\deg}(V))\beta^2n\leq R\leq(\overline{\deg^2}(V)-\overline{\deg}(V))\beta^2n.$$

Proof. For ease of notation introduce $\deg_{\mathbf{S}}(v) := e(\{v\}, \mathbf{S}))$ representing the number of neighbors of vertex $v \in V$ from $\mathbf{S} \sim \mathrm{Uni}(L, k)$. For the lower bound on R, it suffices to count nodes in $\mathbf{S}$ and their neighbors:

$$\begin{aligned}\mathbb{E}_{\mathbf{S}}(|G^{\beta}(\mathbf{S})|) &\geq k + \sum_{v \in \mathcal{N}(\mathbf{S})} \left(1 - (1-\beta)^{\deg_{\mathbf{S}}(v)}\right) \\ &\geq k + \sum_{v \in \mathcal{N}(\mathbf{S})} \left(\beta \deg_{\mathbf{S}}(v) - \beta^2 \binom{\deg_{\mathbf{S}}(v)}{2}\right) \\ &\geq k + \beta e(\mathbf{S}, V \backslash \mathbf{S}) - \beta^2 \sum_{v \in V} \binom{\deg_{\mathbf{S}}(v)}{2}.\end{aligned} \tag{6}$$

The probability that the random set $\mathbf{S}$ contains two given nodes in L is $\frac{k(k-1)}{m(m-1)}$, hence

$$\begin{aligned}\beta^2 \sum_{v \in V} \mathbb{E}\left[\binom{\deg_{\mathbf{S}}(v)}{2}\right] &= \frac{k(k-1)}{m(m-1)} \beta^2 \sum_{v \in V} \binom{\deg_L(v)}{2} \\ \leq \frac{1}{2} \beta^2 \sum_{v \in V} \deg(v)(\deg(v) - 1) &= \frac{1}{2}(\overline{\deg^2}(V) - \overline{\deg}(V))\beta^2 n.\end{aligned}$$

For the upper bound notice

$$\mathbb{E}_{\mathbf{S}}(|G^{\beta}(\mathbf{S})|) = k + \sum_{v \in \mathcal{N}(\mathbf{S})} \mathbb{P}_{\mathbf{S}}(v \in G^{\beta}(\mathbf{S})) + \sum_{v \in V \backslash (\mathbf{S} \cup \mathcal{N}(\mathbf{S}))} \mathbb{P}_{\mathbf{S}}(v \in G^{\beta}(\mathbf{S})).$$

Let $\deg_{\mathbf{S}}^{\beta}(v)$ denote the number of neighbors of $v \in V$ from $\mathbf{S}$ in the percolated graph G^{β}. Clearly, for $v \in \mathcal{N}(\mathbf{S})$

$$\begin{aligned}\mathbb{P}_{\mathbf{S}}(v \in G^{\beta}(\mathbf{S})) &\leq \mathbb{P}_{\mathbf{S}}\left(\deg_{\mathbf{S}}^{\beta}(v) > 0\right) + \mathbb{P}_{\mathbf{S}}\left(v \in G^{\beta}(\mathbf{S}) | \deg_{\mathbf{S}}^{\beta}(v) = 0\right) \\ &= \underbrace{1 - (1-\beta)^{\deg_{\mathbf{S}}(v)}}_{\leq \beta \deg_{\mathbf{S}}(v)} + \mathbb{P}_{\mathbf{S}}\left(v \in G^{\beta}(\mathbf{S}) | \deg_{\mathbf{S}}^{\beta}(v) = 0\right).\end{aligned}$$

Let H denote the graph where the edges between v and $\mathbf{S}$ are deleted. Since edge retention happens independently

$$\mathbb{P}_{\mathbf{S}}\left(v \in G^{\beta}(\mathbf{S}) | \deg_{\mathbf{S}}^{\beta}(v) = 0\right) = \mathbb{P}_{\mathbf{S}}(v \in H^{\beta}(\mathbf{S})).$$

We will call a length 2 path *vuw proper* if v and w has distance two, or in other words, v, u, w does not form a triangle. $\mathcal{A}_v(G)$ denotes the event that there is no proper path starting from v in the percolated graph G^{β} which used to be a proper path in G. (We are not interested in triangles becoming proper paths due to edge percolation as these "artificial" proper paths do not lead to distance 2 after two steps in the original graph G.)

Observe that vertices $v \in V \backslash (\mathbf{S} \cup \mathcal{N}(\mathbf{S}))$ in graph G at least 2 steps away from the set $\mathbf{S}$. Since without proper paths one cannot go further than distance 1 from vertex v

$$v \in V \backslash (\mathbf{S} \cup \mathcal{N}(\mathbf{S})) \Rightarrow \mathbb{P}_{\mathbf{S}}(v \notin G^{\beta}(\mathbf{S})) \geq \mathbb{P}(\mathcal{A}_v(G)).$$

The argument is similar in the $v \in \mathcal{N}(S)$ case. Assuming $\mathcal{A}_v(G)$ we can only reach vertices that are at distance 1 away from v. On H every edge between $\mathbf{S}$ and v are deleted, hence,

$$v \in \mathcal{N}(\mathbf{S}) \Rightarrow \mathbb{P}_{\mathbf{S}}(v \notin H^{\beta}(\mathbf{S})) \geq \mathbb{P}(\mathcal{A}_v(G)).$$

Together, they make bound

$$\mathbb{E}_{\mathbf{S}}(|G^{\beta}(\mathbf{S})|) \leq k + \beta e(\mathbf{S}, V \backslash \mathbf{S}) + \sum_{v \in V \backslash \mathbf{S}} (1 - \mathbb{P}(\mathcal{A}_v(G))).$$

Fix v and u. Let $\delta_v(u)$ denote the number of vuw proper paths with some other vertex w. Clearly, $\delta_v(u) \leq \deg(u) - 1$.

Note that two proper paths $vuw, vu'w'$ can only share an edge in their vu, vu' segment when $u = u'$, the second segment is always disjoint. ($w' = u, u' = w$ would make u, v, w a triangle.) This means any two proper paths are kept independently when $u \neq u'$.

$$\mathbb{P}\left(\bigcup_{w:\, w \sim u} \{vuw \text{ is a proper path in } G \text{ and } G^{\beta}\} \right) = \beta \left(1 - (1-\beta)^{\delta_v(u)}\right),$$

$$\begin{aligned}\mathbb{P}(\mathcal{A}_v(G)) &= \prod_{u:\, u \sim v} \left[1 - \beta\left(1 - (1-\beta)^{\delta_v(u)}\right)\right] \overset{*}{\geq} 1 - \sum_{u:\, u\sim v} \beta\left(1 - (1-\beta)^{\delta_v(u)}\right) \\ &\geq 1 - \beta^2 \sum_{u:\, u \sim v} \delta_v(u) \geq 1 - \beta^2 \sum_{u:\, u\sim v} (\deg(u) - 1),\end{aligned}$$

$$\begin{aligned}\sum_{v \in V \backslash \mathbf{S}} (1 - \mathbb{P}(\mathcal{A}_v(G))) &\leq \beta^2 \sum_{v \in V} \sum_{u:\, u \sim v} (\deg(u) - 1) = \beta^2 \sum_u \deg(u)(\deg(u) - 1) \\ &= (\overline{\deg^2}(V) - \overline{\deg}(V))\beta^2 n.\end{aligned}$$

Note that at step $*$ we used the union bound for independent events. ☐

Proof of Corollary 2.8. Let R_C, R_V be the remainder terms in Lemma 2.7 when $L = C, V$. Since $\frac{n-k}{n-1} \geq \frac{r-k}{r-1}$, (3) implies $\overline{\deg}(C) \geq \overline{\deg}(V)$. Thus,

$$|R_C|, |R_V| \leq \overline{\deg^2}(V) \beta^2 n.$$

This results in the bound

$$
\begin{aligned}
&\mathbb{E}(|G^{\beta}(\mathbf{S}_C)|) - \mathbb{E}(|G^{\beta}(\mathbf{S}_V)|) \\
&= \left(\overline{\deg}(C) - \frac{k-1}{r-1}\frac{1}{r}e(C,C) - \overline{\deg}(V) + \frac{k-1}{n-1}\frac{1}{n}e(V,V)\right)\beta k + R_C - R_V \\
&\geq \left(\frac{r-k}{r-1}\overline{\deg}(C) - \frac{n-k}{n-1}\overline{\deg}(V)\right)\beta k - 2\overline{\deg^2}(V)\beta^2 n \\
&\geq \frac{1}{2}c_1\beta k = \frac{1}{2}c_1\beta s n.
\end{aligned}
$$

□

3.3.2 | Lemmas for large β

Lemma 3.3. *Let G be a graph with n nodes and edge-expansion (a, q), where $a > 1$ and $q < 1/3$. Let $0 < \beta < 1$, and let H be a largest connected component of G^{β}. Then*

$$\mathbb{P}(|H| \leq (1-q)n) \leq \rho^n,$$

where

$$\rho = \left(\frac{e(1-\beta)^a}{q}\right)^q. \tag{7}$$

For the bound to be nontrivial, we need that $q > e(1-\beta)^a$.
We start with an elementary observation.

Claim 1. If the largest connected component of G^{β} has at most $n - t$ nodes, where $t \leq n/3$, then there is a set $X \subseteq V$ such that $t \leq |X| \leq n/2$ and no edge of G^{β} connects X and $V\setminus X$.

Proof of Claim 1. Indeed, let H_1 be the node set of the largest connected component of G^{β}. Then $|H| \leq n - t$ by hypothesis. If $|H| \geq n/2$, then $X = V\setminus H$ satisfies the conditions in the claim. So suppose that $|H| < n/2$. If $t \leq |H|$, then H satisfies the conditions in the claim. So suppose that $|H| < t$. Let us add further connected components to H as long as it remains at most $n/2$ in cardinality, to get a set X. If $|X| \geq t$ then we are done, so suppose that $|X| < t$. Adding any other connected component, we get a set X' with $|X'| > n/2$ and $|X'| < |X| + t$. If $|X'| \leq n - t$, then $V\setminus X'$ satisfies the conditions in the claim. So suppose that $|X'| > n - t$. But then $n - t < |X'| \leq |X| + t \leq 2t$, and so $t > n/3$, contrary to the hypothesis. □

Proof of Lemma 3.3. Let $z = (1-\beta)^a$. For a fixed k-subset X ($qn \leq k \leq n/2$), the graph G has at least ak edges between X and $V\setminus X$, and the probability that none of them is selected is at most $(1-\beta)^{ak} = z^k$. So the probability that there is a set $X \subseteq V$ with $qn \leq |X| \leq n/2$ and having no edges between X and $V\setminus X$ is at most

$$\sum_{k=\lceil qn\rceil}^{\lfloor n/2\rfloor}\binom{n}{k}z^k.$$

Let $p = z/(1+z)$ and let ξ be a $\mathrm{Binom}(n, p)$ distributed random variable. Then, by the well-known Chernoff–Hoeffding bound,

$$\begin{aligned}\sum_{k=\lceil qn\rceil}^{\lfloor n/2\rfloor}\binom{n}{k}z^k &= (1+z)^n\sum_{k=\lceil qn\rceil}^{\lfloor n/2\rfloor}\binom{n}{k}p^k(1-p)^{n-k}\\ &\le (1+z)^n\,\mathbb{P}(\xi \ge qn) \le (1+z)^n\left[\left(\frac{p}{q}\right)^q\left(\frac{1-p}{1-q}\right)^{1-q}\right]^n\\ &= \left[\left(\frac{z}{q}\right)^q\frac{1}{(1-q)^{1-q}}\right]^n.\end{aligned} \tag{8}$$

Here

$$\frac{1}{(1-q)^{1-q}} = \left(1+\frac{q}{1-q}\right)^{1-q} < \mathrm{e}^q,$$

hence, by (8) and Claim 1

$$\mathbb{P}(|H| \le (1-q)n) < \left(\frac{ez}{q}\right)^{qn},$$

proving the lemma. ☐

Proof of Lemma 2.12. Let H_1 denote the component of G^β with the largest number of nodes in C, and let $H_2, \ldots, H_m$ be the other components. Let H^C be the largest component in $G^\beta(C)$. There is a component H_i of G^β such that $H^C \subseteq H_i$. Since $H^C \subseteq C$ this makes

$$|H^C| \le |C \cap H_i| \le |C \cap H_1| \le |H_1|.$$

Let $h_j = |H_j|$ and $c_j = |C \cap H_j|$. Let p_j and q_j denote the probability that H_j is not infected by $\mathbf{S}_V$ and $\mathbf{S}_C$, respectively. Then

$$p_j = \prod_{i=0}^{k-1}\left(1 - \frac{h_j}{n-i}\right),$$

where the last equality holds whenever $h_j \le n - k + 1$; else, $p_j = 0$. Similarly,

$$q_j = \prod_{i=0}^{k-1}\left(1 - \frac{c_j}{r-i}\right),$$

where again the last equality holds if $c_j \le r - k + 1$ and 0 otherwise. Then

$$\begin{aligned}\mathbb{E}_\beta(|G^\beta(\mathbf{S}_V)|) - \mathbb{E}_\beta(|G^\beta(\mathbf{S}_C)|) &= \sum_{j=1}^m h_j(1-p_j) - \sum_{j=1}^m h_j(1-q_j)\\ &= \sum_{j=1}^m h_j(q_j - p_j).\end{aligned} \tag{9}$$

The main idea of the proof is that we partition the index set $K = \{1, \ldots, m\}$ into four sets:

$$\begin{aligned} K_1 &:= \{1\},\\ K_2 &:= \{j \in K : h_j = 1, c_j = 0\},\\ K_3 &:= \{j \in K \backslash K_1 : h_j \le c_j/(c-s)\},\\ K_4 &:= K \backslash (K_1 \cup K_2 \cup K_3). \end{aligned}$$

Let $V_i = \cup_{j \in K_i} H_j$. Recall $G_1 = G[C]$ and $G_2 = G \backslash E(G_1)$. We lower bound the sum in Equation (9) using a different estimate over each set K_j. There are two sets where uniform seeding is more dangerous (K_2 and K_4), one set where the two seedings are essentially equally dangerous (K_1, the giant component) and one set where the central seeding is more dangerous (K_3), but this set K_3 only contains components which have a relatively large part in C compared to $V \backslash C$, and since the giant component is quite large in G_1, the components in K_3 cannot have too much weight We make this intuition precise in the computation below.

First, we fix the percolation G^β, and estimate the expectations over the choice of seed sets. We start with K_1, which only contains the index of the component with the largest number of nodes in C. This component will have a nonempty intersection with both $\mathbf{S}_V$ and $\mathbf{S}_C$ with high probability. More precisely,

$$\begin{aligned} \mathbb{E}_\beta(|V_1 \cap G^\beta(\mathbf{S}_V)|) - \mathbb{E}_\beta(|V_1 \cap G^\beta(\mathbf{S}_C)|) &= h_1(q_1 - p_1) \ge -h_1 p_1 \\ &\ge -h_1\left(1 - \frac{h_1}{n}\right)^k \ge -n\left(1 - \frac{h_1}{n}\right)^k \ge -n\mathrm{e}^{-h_1 k/n}. \end{aligned} \tag{10}$$

Next we consider K_2, the index set of those components of G^β that are isolated nodes of $V \backslash C$. Clearly $q_j = 1$ and $p_j = \prod_{i=0}^{k-1}\left(1 - \frac{1}{n-j}\right) = \frac{n-k}{n}$ for $j \in K_2$. So

$$\mathbb{E}_\beta(|V_2 \cap G^\beta(\mathbf{S}_V)|) - \mathbb{E}_\beta(|V_2 \cap G^\beta(\mathbf{S}_C)|) = \sum_{j \in K_2} h_j(q_j - p_j) = \frac{k}{n}|V_2|. \tag{11}$$

For K_3 we use the lower bound

$$\begin{aligned} \mathbb{E}_\beta(|V_3 \cap G^\beta(\mathbf{S}_V)|) - \mathbb{E}_\beta(|V_3 \cap G^\beta(\mathbf{S}_C)|) &= \sum_{j \in K_3} h_j(q_j - p_j) > -\sum_{j \in K_3} h_j \\ &\ge -\sum_{j \in K_3} \frac{c_j}{c-s} \ge -\frac{1}{c-s}|C \backslash V_1|. \end{aligned} \tag{12}$$

Finally, if $j \in K_4$, then it must satisfy

$$1 - \frac{h_j}{n} \le 1 - \frac{c_j}{r-k},$$

which implies that for these components $q_i \geq p_i$ and so

$$\mathbb{E}_\beta(|V_3 \cap G^\beta(\mathbf{S}_V)|) - \mathbb{E}_\beta(|V_3 \cap G^\beta(\mathbf{S}_C)|) \geq 0. \tag{13}$$

Summing (10)–(13), we get

$$\mathbb{E}_\beta(|G^\beta(\mathbf{S}_V)|) - \mathbb{E}_\beta(|G^\beta(\mathbf{S}_C)|) \geq -n\mathrm{e}^{-|V_1|k/n} + \frac{k}{n}|V_2| - \frac{1}{c-s}|C \backslash V_1|. \tag{14}$$

To compute the expectation of this over the percolation, let us denote the degree of the ith node in $V \backslash C$ by b_i. Then by Jensen's inequality (since $(1-\beta)^x$ is convex),

$$\mathbb{E}_\beta(|V_2|) = \sum_{i=1}^{n-r}(1-\beta)^{b_i} \geq (n-r)(1-\beta)^b. \tag{15}$$

Recall H^C being the largest component of $G^\beta(C)$. By applying Lemma 3.3 to G_1, we have $|V_1| = |H_1| \geq |H^C| > (1-q)r$ with probability at least $1-\rho^r$, where ρ is defined by (7). Hence,

$$\mathbb{E}\left(\mathrm{e}^{-|V_1|k/n}\right) \leq \mathrm{e}^{-(1-q)ck} + \rho^r \leq \mathrm{e}^{-2ck/3} + \rho^r$$

and

$$\mathbb{E}(|H^C|) \geq (1-q)r\mathbb{P}(|H^C| > (1-q)r) \geq (1-q)(1-\rho^r)r$$

resulting in

$$\begin{aligned} \mathbb{E}(|C \backslash V_1|) \leq \mathbb{E}(|C \backslash H^C|) &= r - \mathbb{E}(|H^C|) \\ &\leq r[1-(1-q)(1-\rho^r)] \leq qr + r\rho^r. \end{aligned} \tag{16}$$

To sum up,

$$\begin{aligned} &\mathbb{E}(|G^\beta(\mathbf{S}_V)|) - \mathbb{E}(|G^\beta(\mathbf{S}_C)|) \\ \geq\ & -n\mathrm{e}^{-2ck/3} + \frac{k}{n}\left(1 - \frac{k}{2n}\right)(n-r)(1-\beta)^b - \frac{1}{c-s}qr - \left(n + \frac{r}{c-s}\right)\rho^r \\ \geq\ & s(1-c)(1-\beta)^b n - \frac{c}{c-s}qn - \left(1 + \frac{c}{c-s}\right)n\rho^r - n\mathrm{e}^{-2ck/3}. \end{aligned}$$

This proves the lemma. □

Proof of Theorem 2.14. We start with the small β case. We need the following fact:

Claim 2. Let H be a graph with N nodes and edge-expansion (a, q) $(a \geq 1, 0 < q < 1/2)$. Then the average degree in H is at least $2a\frac{N-1}{N+1}$.

Proof of Claim 2. We check this for the case when $|V(H)| = 2m + 1$ is odd (the even case is similar). For every m-subset $S \subseteq V$, there are at least am edges between S and $V \backslash S$. This gives $am\binom{2m+1}{m}$ edges. Each edge is counted $2\binom{2m-1}{m-1}$ times, hence

$$|E(H)| \geq am\frac{\binom{2m+1}{m}}{2\binom{2m-1}{m-1}} = a\frac{(2m+1)m}{m+1}.$$

Thus the average degree is

$$\overline{\deg}(H) = \frac{2|E(H)|}{2m+1} = 2a\frac{m}{m+1}.$$

This proves the claim. ☐

Consider a graph G_m from the given sequence. In the rest of this proof, we omit the indices m, to make the arguments more readable. The claim above implies that

$$\overline{\deg}(C) \geq 2a\frac{r-1}{r+1} \sim 2a,$$

therefore

$$\begin{aligned}
\frac{r-k}{r-1}\overline{\deg}(C) - \frac{n-k}{n-1}\overline{\deg}(V) &\sim \left(1 - \frac{s}{c}\right)\overline{\deg}(C) - (1-s)\overline{\deg}(V) \\
&\geq \left(1 - \frac{s}{c}\right)\overline{\deg}(C) - (c\overline{\deg}(C) + (1-c)b) \\
&= \left(1 - c - \frac{s}{c}\right)\overline{\deg}(C) - (1-c)b \overset{s \leq \frac{1}{2}c(1-c)}{\geq} \\
&\quad (1-c)\left(\frac{1}{2}\overline{\deg}(C) - b\right) \gtrsim (1-c)(a-b) \geq \varepsilon^2 \\
&\Rightarrow \frac{r-k}{r-1}\overline{\deg}(C) - \frac{n-k}{n-1}\overline{\deg}(V) \geq \varepsilon^2 - o(1) \geq \frac{1}{2}\varepsilon^2 =: c_1 > 0,
\end{aligned}$$

when n is large enough. This means the conditions of Corollary 2.8 are satisfied when β is small enough.

The large beta case is an easy consequence of Remark 2.13. ☐

4 | APPLICATION: CHUNG–LU MODEL WITH POWER-LAW DEGREE DISTRIBUTION

In this section, we apply Lemma 2.12 to rigorously prove the previous claim of [8], that the uniform seeding can be more dangerous in random graphs with power-law degree distribution with exponent $\tau \in (2, 3)$ if

$$\frac{1}{n}\beta^{-\frac{1}{|\tau-3|}} \ll s \ll \beta^{\frac{\tau-1}{3-\tau}}. \tag{17}$$

In [8], this claim appears as an if and only if statement, however, in this section we only address the "if" part. Our proof strategy has already been outlined in a previous paper [9], but this is the first time when we give a fully rigorous proof.

We start by defining the random graph distribution in the focus of this section, which is one of the most standard models of networks with a power-law degree distribution.

Definition 4.1. Let us denote by $\mathrm{C}\mathcal{L}(\tau)$ the *distribution of Chung–Lu random graphs with exponent* $\tau \in (2,3)$, where the nodes $v_i \in V$ are indexed from 1 to n, and v_i and v_j are connected by an edge independently with probability $p_{ij} = \min\left\{\frac{d_i d_j}{D}, 1\right\}$, where $d_i = \left(\frac{n}{i}\right)^{\frac{1}{\tau-1}}$ and $D = \sum_{k=1}^{n} d_k$.

Notice that the expected degree of a node with index i in a graph sampled from $\mathrm{C}\mathcal{L}(\tau)$ is $d_i + o(1)$. Moreover, for d_i to be less than some integer degree d, we need to have $\frac{i}{n} \le d^{1-\tau}$, which hints that the exponent of the cumulative degree distribution is expected to be around $1-\tau$, and therefore the degree distribution is expected to follow a power-law with exponent $\tau \in (2,3)$. The average degree of the distribution is expected to be a constant, because

$$\sum_{k=1}^{n} d_k \sim \int_1^n \left(\frac{n}{x}\right)^{\frac{1}{\tau-1}} dx = \frac{\tau-1}{\tau-2}\left(n - n^{\frac{1}{\tau-1}}\right) = \Theta(n). \tag{18}$$

Chung–Lu random graphs were introduced in [4], we refer to this paper and follow-up works for more precise statements on interpreting Definition 4.1. Here, we continue by stating an elementary result about the edge expansion of Chung–Lu random graphs, which may have already appeared in the literature in a similar form, but we are not aware of it.

Lemma 4.2. *If G is sampled from $\mathrm{C}\mathcal{L}(\tau)$ with $\tau \in (2,3)$, and if C is the set of vertices of G with index $i \le cn$, with $\frac{1}{\sqrt{n}} \ll c \ll (\log(n))^{-\frac{\tau-1}{3-\tau}}$, then $G[C]$ has edge expansion $\left(\frac{n}{4D}c^{-\frac{3-\tau}{\tau-1}}, 0\right)$ asymptotically almost surely.*

Before proving Lemma 4.2, let us show how we can apply it to rigorously prove the claim in Equation (17), at least partially.

Corollary 4.3. *Let us consider a sequence of graphs G_n sampled from $\mathrm{C}\mathcal{L}(\tau)$ with $\tau \in (2,3)$, and let us define the central region C as the $\lfloor cn \rfloor$ nodes with the largest expected degree (i.e., index) in G_n. Let us assume that the size of the seed set satisfies $s = \Theta(c)$ and $c - s = \Theta(c)$. Under the mild assumption $1-\beta = \Theta(1)$, if*

$$\sqrt{\frac{\log(n)}{n}} \ll s \ll \left(\frac{\beta}{\log(n)}\right)^{\frac{\tau-1}{3-\tau}} \tag{19}$$

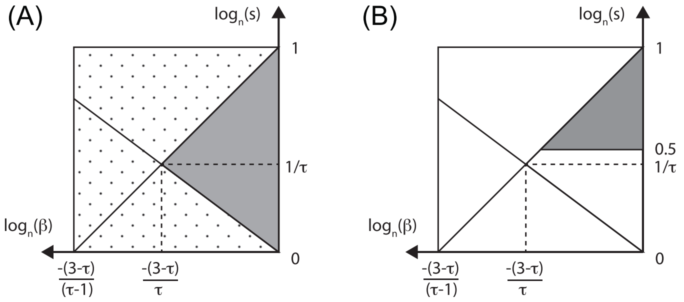

FIGURE 1 Phase diagrams for the switchover phenomenon on Chung–Lu random graphs with power-law degree distribution ($\tau \in (2, 3)$). (A) With gray we show the region where the central area is more dangerous, as claimed in [8], and with the dotted pattern we show the region where the central area is more dangerous by Equation (17), as claimed in [8]. (B) With gray we show the region where the central area is more dangerous by Corollary 4.3.

hold then the uniform seeding creates a larger epidemic than the central seeding (G_n has the weak switchover property) with probability tending to 1 *as* $n \to \infty$.

Notice that with this choice of parameters, the number of seeds is linear only the size of the central region, but sub-linear in the size of the graph.

As shown in Figure 1, the parameter ranges set by Equation (19) form a subset of the parameters in Equation (17), therefore, Corollary 4.3 is weaker than the claim in [8]. To generalize Corollary 4.3 for the remaining parameter ranges, different proof methods are necessary.

Proof of Corollary 4.3. We will use Lemma 2.12 for Chung–Lu random graphs with

$$q = \frac{s(c-s)(1-c)(1-\beta)^b}{2c}. \tag{20}$$

For Lemma 2.12 to be applicable, we need to make sure that for our choice

$$q = \frac{s(c-s)(1-c)(1-\beta)^b}{2c} > (1+\varepsilon)e(1-\beta)^a, \tag{21}$$

and for Equation (5) to imply that the uniform seeding can be more dangerous, we need that

$$\frac{1}{2}s(1-c)(1-\beta)^b > \left(1 + \frac{c}{c-s}\right)\rho^r + \mathrm{e}^{-2ck/3}, \tag{22}$$

keeping in mind that c and s are not constants anymore. Condition $q < \frac{1}{3}$ is trivially satisfied for large enough n as $q = \Theta(s)$.

Recall, that we chose $s = \Theta(c)$ and $c - s = \Theta(c)$. Notice that we can apply Lemma 4.2, because the condition $\frac{1}{\sqrt{n}} \ll c \ll (\log(n))^{-\frac{\tau-1}{3-\tau}}$ holds by (19). Then, since we also know $b = \Theta(1)$ by Equation (18), we can show that Equation (21) holds if

$$\log(c) \gg \log(1-\beta)c^{-\frac{3-\tau}{\tau-1}},$$

which is implied by Equation (19) as

$$\log(1-\beta)c^{-\frac{3-\tau}{\tau-1}} \leq -\beta c^{-\frac{3-\tau}{\tau-1}} = -\left(\frac{c}{\beta^{\frac{\tau-1}{3-\tau}}}\right)^{-\frac{3-\tau}{\tau-1}} \overset{(19)}{\ll} -\log(n)$$
$$= 2\log\left(\frac{1}{\sqrt{n}}\right) \ll \log(c).$$

Similarly, Equation (22) holds if

$$\rho^{r} + \mathrm{e}^{-2ck/3} \ll s.$$

By Equation (18) and substituting the definition of ρ from Equation (7), we get

$$(1+\varepsilon)^{-qr} + \mathrm{e}^{-2ck/3} \ll s,$$

which must hold because Equation (19) implies

$$\Theta(qr) = \Theta(ck) = \Theta(s^2 n) \overset{(19)}{\gg} \log(n).$$

Therefore, Lemma 4.2 implies that for these parameter ranges, the uniform seeding can be more dangerous, and weak switchover occurs. □

We conclude the section by providing the proof of Lemma 4.2.

Proof of Lemma 4.2. For $S \subset V$ with $|S| \leq |C|/2$, let X_S be the number of edges between S and $C \backslash S$. In the first part of the proof, we show that $\mathbb{E}(X_S) \geq \frac{n}{2D} c^{-\frac{3-\tau}{\tau-1}} |S|$ for every S, and in the second part we prove that the variables X_S are all well-concentrated around their expectation.

Note that since we assumed $c \gg \frac{1}{\sqrt{n}}$, and since we know $D = \Theta(n)$, we have that $\min\left\{d^2_{\lfloor cn \rfloor}, D\right\} = d^2_{\lfloor cn \rfloor} \geq c^{-\frac{2}{\tau-1}}$. Then, we compute the expectation of X_S as

$$\mathbb{E}(X_S) = \sum_{i\in S}\sum_{i\in C\backslash S} \min\left\{\frac{d_i d_j}{D}, 1\right\} = \sum_{i\in S}\sum_{i\in C\backslash S} \frac{d_i d_j}{D} \geq \sum_{i\in S}\sum_{i\in C\backslash S} \frac{d^2 \lfloor cn \rfloor}{D}$$
$$\geq \frac{|S|(|C|-|S|)}{D} c^{-\frac{2}{\tau-1}} \overset{|S|\leq\frac{|C|}{2}}{=} \frac{n}{2D} c^{-\frac{3-\tau}{\tau-1}} |S|.$$

Next, we use the union bound, and well-known multiplicative Chernoff bounds on binomial random variables, to prove that the random variables X_S are concentrated around their expectation. We bound

$$\begin{aligned}\mathbb{P}\left(\exists\, S\subset C, |S|\le \frac{|C|}{2}\text{ with } X_S\le \frac{1}{2}\mathbb{E}(X_S)\right) &\le \sum_{\substack{S\subset C\\ |S|\le\frac{|C|}{2}}} \mathbb{P}\left(X_S\le\frac{1}{2}\mathbb{E}(X_S)\right)\\ &\le \sum_{\substack{S\subset C\\ |S|\le\frac{|C|}{2}}} \mathrm{e}^{-\frac{1}{8}\mathbb{E}(X_S)}\\ &= \sum_{\substack{S\subset C\\ |S|\le\frac{|C|}{2}}} \mathrm{e}^{-\frac{n}{16D}c^{-\frac{3-\tau}{\tau-1}}|S|}.\end{aligned}$$

Set $\eta=\frac{n}{16D}c^{-\frac{3-\tau}{\tau-1}}$. Let us change the indexing of the sum to the size of the set S, and apply a standard upper bound on binomial coefficients to obtain

$$\sum_{\substack{S\subset C\\ |S|\le\frac{|C|}{2}}} \mathrm{e}^{-\eta|S|}=\sum_{k=1}^{\frac{\lfloor cn\rfloor}{2}}\binom{\lfloor cn\rfloor}{k}\mathrm{e}^{-\eta k}\le\sum_{k=1}^{\frac{\lfloor cn\rfloor}{2}}\left(\frac{enc}{k}\right)^{k}\mathrm{e}^{-\eta k}\le\sum_{k=1}^{\frac{\lfloor cn\rfloor}{2}}\left(cn\mathrm{e}^{1-\eta}\right)^{k}.$$

Notice that we have arrived at a geometric series with a common ratio $cn\mathrm{e}^{1-\eta}$, which tends to zero as long as $\eta=\frac{n}{16D}c^{-\frac{3-\tau}{\tau-1}}\gg\log(n)$; an asymptotic inequality that holds by the assumption $c\ll(\log(n))^{-\frac{\tau-1}{3-\tau}}$. Therefore, we arrived to the equation

$$\mathbb{P}\left(\exists\, S\subset C, |S|\le\frac{C}{2}\text{ with } X_S\le\frac{n}{4D}c^{-\frac{3-\tau}{\tau-1}}|S|\right)\to 0,$$

which completes the proof of the lemma. □

5 | CONCLUDING REMARKS

In this paper, we gave the first fully rigorous proofs of the switchover phenomenon, introduced in [8], for general classes of graphs. We showed that weak switchover exists under mild conditions on the graph, and we also showed sufficient conditions for strong switchover.

One limitation of the current paper is that, in the case of the strong switchover, the size of the seed set was assumed to be fairly large, of size $\Omega(n)$. Although for the Chung–Lu model in Section 4 we did study smaller seed sets, and we did use the machinery of the strong switchover proofs, we were only able to show the existence of the weak switchover phenomenon. This agrees with the simulations and heuristic derivations of the previous work [8], which also claimed that Chung–Lu models exhibit weak switchover, but not strong switchover. However, Ódor et al. [8] also showed that the strong switchover phenomenon occurs with much smaller seed sets on geometric graphs, notably on the commuting network of Hungary constructed from real data, and also random graph models with an underlying geometry. Unfortunately, our current results do not say much about such geometric graphs.

As in this paper, proving the existence of strong switchover with small seed sets would boil down to the distribution of component sizes in the percolated graph G^{β}. However, contrary to this paper, we need the existence of at least medium size components also in the periphery, because if most of the peripheral nodes in G^{β} are contained in bounded-size components, then we need $\Omega(n)$ seeds to have strong switchover (even to have an epidemic of size $\Omega(n)$). Finding appropriate conditions for such medium size components in the periphery which could lead to the existence of strong switchover with small seed sets (say, of size $\sqrt{n}$ or even $\log n$) is an interesting future direction.

ACKNOWLEDGMENTS

The authors are thankful to Marianna Bolla for her insightful remarks. This work has been supported by the Dynasnet ERC Synergy project (ERC-2018-SYG 810115). Gergely Ódor was supported by the Swiss National Science Foundation, under grant number P500PT-211129.

REFERENCES

1. Y. Alimohammadi, C. Borgs, and A. Saberi, *Algorithms using local graph features to predict epidemics*, Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA) (J. Naor and N. Buchbinder, eds.), SIAM, 2022.
2. G. Barmpalias, N. Huang, A. Lewis-Pye, A. Li, X. Li, Y. Pan, and T. Roughgarden, *The idemetric property: When most distances are (almost) the same*, Proc. Roy. Soc. A. **475** (2019), 20180283.
3. M. Borassi, *Algorithms for metric properties of large real-world networks from theory to practice and back*, PhD Thesis, IMT School for Advanced Studies, Lucca, 2016.
4. F. R. K. Chung and L. Lu, *Connected components in random graphs with given expected degree sequences*, Ann. Comb. **6** (2002), 125–145.
5. F. R. K. Chung, R. L. Graham, and R. M. Wilson, *Quasi-random graphs*, Combinatorica. **9** (1989), 345–362.
6. J. Fox, T. Roughgarden, C. Seshadhri, F. Wei, and N. Wein, *Finding cliques in social networks: A new distribution-free model*, SIAM J. Comput. **49** (2020), 448–464.
7. M. Newman, *Networks*, Oxford University Press, Oxford, 2018.
8. G. Ódor, D. Czifra, J. Komjáthy, L. Lovász, and M. Karsai, *Switchover phenomenon induced by epidemic seeding on geometric networks*, Proc. Nat. Acad. Sci. **118** (2021), no. 41, e2112607118.
9. G. Ódor, D. Czifra, J. Komjáthy, L. Lovász, and M. Karsai, *Longer-term seeding effects on epidemic processes: A network approach*, Sci. Secur. **2** (2022), no. 4, 409–417.
10. T. Roughgarden and C. Seshadhri, *Distribution-free models of social networks*, ArXiv Preprint ArXiv:2007.15743. (2020).

How to cite this article: D. Keliger, L. Lovász, T. F. Móri, and G. Ódor, *Switchover phenomenon for general graphs*, J. Graph Theory. **108** (2025), 560–581. https://doi.org/10.1002/jgt.23184