

Online Appendix

In the paper, we presented our main result for Inhomogeneous Random Networks (IRN), which capture as special instances Erdős-Renyi graphs, networks with power-law degree distributions and island models with homophily. A remaining question is how would results hold up for networks that have features not captured by the IRN model, e.g., networks with clustering, like “small-world” networks (Watts and Strogatz, 1998) or networks exhibiting power-law degree distributions *and* homophily).

In Section 4.3, we presented a generalization using the notion of graphs with **solar system structure**, showing that our results hold in any network models that admit a unique giant component with high probability or no giant component at all. For the sake of completeness, we include a description of this concept and the accompanying theorem in Section 1.

Here, we show that a broader class of networks enjoy the structure which is necessary for our results. Central to our extension is the notion of local convergence of graphs. Local convergence has recently become an essential tool for analyzing sparse graphs. It has been instrumental in determining the convergence and limits of various factors, such as clustering coefficients, the size of giant connected component, and the distribution of PageRank values in sparse random graphs; see Hofstad (2024) for an exposition.

We build on the recent work of Alimohammadi, Borgs and Saberi (2023) and Alimohammadi, Borgs and Saberi (2022) to extend our seeding results for certain families of graphs that exhibit local convergence. We incorporate several definitions and theorems from those works, but simplify statements in certain places to facilitate reading.

We show that the small-world or “Watts-Strogatz-type networks” satisfy the conditions for the main theorem of the paper to apply. We also discuss a natural approach for incorporating homophily into power-law networks and demonstrate that our results go through for the resulting model.

1 Networks with a solar system structure

Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of (possibly random) graphs in which the number of vertices of G_n is equal to n . We build the communication network $\mathcal{K}(G_n)$ by keeping each link in the graph independently and with probability c . Let C_i denote the set of vertices in the i 'th largest connected component of $\mathcal{K}(G_n)$. Finally, we use $\xrightarrow{\mathbb{P}}$ to denote convergence in probability with respect to both the random graph distribution and the process of forming the communication network.

Definition 1. *We say a sequence of graphs has the **solar system structure** if and only if there exists a threshold probability $0 < c_p < 1$ such that*

1. *If $c < c_p$ then,*

$$\frac{|C_1|}{n} \xrightarrow{\mathbb{P}} 0.$$

2. *If $c > c_p$, then there exists a function $\zeta : [0, 1] \rightarrow (0, 1]$*

$$\frac{|C_1|}{n} \xrightarrow{\mathbb{P}} \zeta(c),$$

3. Moreover, for all $c \in [0, 1]$, $\frac{|C_2|}{n} \xrightarrow{\mathbb{P}} 0$.

In a graph with solar system structure, either all connected components are of size $o(n)$ or there is a unique giant connected component that is of size linear in n . All the classes of random graphs considered in our paper have this property. In this appendix, we show that many other classes of graphs enjoy this property as well, including those that feature clustering and homophily.

Theorem 2. *Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of (possibly random) graphs with solar system structure and s be the number of seeds.*

- *If $c > c_p$, random seeding catches up to the omniscient seeding at an exponential rate in the number of extra seeds, i.e., for any x ,*

$$\lim_{n \rightarrow \infty} \frac{\mathbf{H}(\text{RAND}, s + x)}{\mathbf{H}(\text{OMN}, s)} = 1 - (1 - \alpha)^{s+x}.$$

- *If $c < c_p$, then any seeding strategy diffuses to only a vanishing fraction of the population:*

$$\lim_{n \rightarrow \infty} \mathbf{H}(\text{OMN}, s) = 0.$$

The solar system property lays bare the intuition driving the results in the paper. The random seeding is likely to reach the large component, if it exists, with sufficiently many seeds. If such a component does not exist, then a small number of seeds cannot diffuse to a sizable fraction of agents for large networks.

1.1 Definition of local convergence

Crucial to our formalization is the notion of local convergence, initiated independently by Aldous and Steele (2004) and Benjamini and Schramm (2001). A sequence of (possibly random) graphs is said to have a *local weak limit in probability*, if the distributions of the neighborhoods of randomly sampled vertices converge in probability. The limit is then a probability measure μ on the space of rooted, locally finite, connected graphs.

A *rooted graph* is simply a pair (G, o) where G is a graph and $o \in V(G)$ is a specific node. The graphs (G_1, o_1) and (G_2, o_2) are *isomorphic*, denoted as $(G_1, o_1) \simeq (G_2, o_2)$, if there exists a bijection $\phi: V(G_1) \mapsto V(G_2)$ such that $\phi(o_1) = o_2$ and $(u, v) \in E(G_1)$ if and only if $(\phi(u), \phi(v)) \in E(G_2)$. Denote the space of (potentially infinite) connected rooted graphs as $\mathcal{G}_\star$, where two rooted graphs are considered equivalent if they are isomorphic. Therefore, $\mathcal{G}_\star$ consists of equivalence classes of rooted graphs modulo isomorphism.

This space of rooted graphs, $\mathcal{G}_\star$, can be endowed with a metric structure denoted as d_{loc} . Let $B_r(G, o)$ denote the subgraph of G comprised of all nodes at a graph distance of at most r from o . The metric d_{loc} between two rooted graphs (G_1, o_1) and (G_2, o_2) is defined as:

$$d_{loc}((G_1, o_1), (G_2, o_2)) = \frac{1}{1 + \inf_k \{k : B_k(G_1, o_1) \not\simeq B_k(G_2, o_2)\}}.$$

Two rooted graphs are closer according to d_{loc} if they are isomorphic for a larger neighborhood of their respective roots.

Note that this metric endows $\mathcal{G}_*$ with the natural σ -algebra of Borel sets, allowing us in particular to consider measures μ on $\mathcal{G}_*$.

Definition 3 (Local convergence in probability). *Let μ be a measure on $\mathcal{G}_*$. A sequence of graphs $(G_n)_{n \geq 1}$ converges locally in probability to a limit $(G, o) \sim \mu$ if and only if for every $r \geq 0$ and $H_* \in \mathcal{G}_*$,*

$$\frac{1}{|V(G_n)|} \sum_{v \in V(G_n)} \mathbb{1}\{B_r(G_n, v) \simeq H_*\} \xrightarrow{\mathbb{P}} \mu(B_r(G, o) \simeq H_*). \quad (1)$$

This definition implies that the proportions of subgraphs in the random graph G_n converges in probability to those given by μ . If vertex $o_n \in V(G_n)$ is chosen uniformly at random, then the above expression is equivalent to $p_r^{(G_n)}(B_r(G_n, o_n) \simeq H_*) \xrightarrow{\mathbb{P}} \mu(B_r(G, o) \simeq H_*)$.

We also need to define a notion of graph expansion, which is a measure of connectivity of graphs.

Definition 4 (Large-set edge expansion). *Given a graph $G = (V, E)$ and a constant $\epsilon < 1/2$, define*

$$\phi(G, \epsilon) = \min_{A \subset V: \epsilon|V| \leq |A| \leq |V|/2} \frac{e(A, V \setminus A)}{|A|} \quad (2)$$

where $e(A, V \setminus A)$ is the number of edges joining A to its complement.

In words, ϕ measures how connected a group of nodes is nodes outside of it, relative to the size of this group. It takes the minimum of this ratio over all sufficiently large sets of nodes (hence the name ‘‘large-set expansion’’).

Call a graph G an $(\alpha, \epsilon, \bar{d})$ large-set (edge) expander if the average degree of G is at most $\bar{d}$ and $\phi(G, \epsilon) \geq \alpha$. A sequence of possibly random graphs $\{G_n\}_{n \in \mathbb{N}}$ is called a *large-set (edge) expander sequence with bounded average degree*, if there exists $\bar{d}$ and $\alpha > 0$ such that for any $\epsilon \in (0, 1/2)$ the probability that G_n is an $(\alpha, \epsilon, \bar{d})$ large-set (edge) expander goes to 1 as $n \rightarrow \infty$.

1.2 Locally convergent graphs have a solar-system structure

One can define the communication network $\mathcal{K}(G)$ for the (infinite) limit graph G as well. Given a probability measure μ on $\mathcal{G}_*$, we define

$$\zeta(c) = \mathbb{E}_\mu[\mathbb{P}_{\mathcal{K}(G)}(|C(o)| = \infty)] \quad (3)$$

where o is the root in $(G, o) \sim \mu$ and $C(o)$ is the connected component of o in $\mathcal{K}(G)$, and we define the percolation threshold $c_p(\mu)$ of μ as

$$c_p(\mu) = \inf_c \{c \in [0, 1] : \zeta(c) > 0\}. \quad (4)$$

Theorem 5 (Theorem 1.1 in Alimohammadi et al. (2023), rephrased). *Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of (possibly random) large-set expanders with bounded average degree converging locally in probability to $(G, o) \in \mathcal{G}_*$. Then G_n has a **solar system structure**. Specifically,*

1. If $c < c_p$ then,

$$\frac{|C_1|}{n} \xrightarrow{\mathbb{P}} 0.$$

2. If $c > c_p$, then for $\zeta : [0, 1] \rightarrow (0, 1]$

$$\frac{|C_1|}{n} \xrightarrow{\mathbb{P}} \zeta(c),$$

with $\xrightarrow{\mathbb{P}}$ denoting convergence in probability with respect to both μ and percolation.

3. Moreover, for all $c \in [0, 1]$, $\frac{|C_2|}{n} \xrightarrow{\mathbb{P}} 0$.

1.3 Extension to directed communication networks

The above theorem can be further extended to the case where the communication network is directed. More formally, let $D_G(c)$ be the random digraph obtained from a graph G by first replacing each edge $\{u, v\}$ by two oriented edges uv and vu and then keeping each oriented edge independently with probability c .

Next, given a digraph, let $C^+(v)$ (and $C^-(v)$) be the set of nodes w that can be reached by a directed path from v to w (from w to v). We refer to these sets as the fan-out (and fan-in) of v . As usual, the set $SCC(v) = C^+(v) \cap C^-(v)$ is called the strongly connected component of v . For a strongly connected component SCC , we use the symbol SCC^+ for the set of nodes $SCC^+ = \bigcup_{v \in SCC} C^+(v)$ and the symbol SCC^- for the set of nodes $SCC^- = \bigcup_{v \in SCC} C^-(v)$.

Finally, as usual, we say that a sequence of events, $(\mathcal{E}_n)$, holds with high probability if the probability of $\mathcal{E}_n$ goes to 1 as $n \rightarrow \infty$.

Theorem 6 (Theorem 1.3 in Alimohammadi et al. (2023), rephrased). *Let $c \in (0, 1]$ and $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of (possibly random) graphs satisfying the conditions of Theorem 5. Also, let SCC_i be the i^{th} largest strongly connected component in $D_{G_n}(c)$. Then*

1. if $c < c_p$, for all $v \in V(G_n)$

$$\frac{|C^+(v)|}{n} \xrightarrow{\mathbb{P}} 0, \quad \frac{|C^-(v)|}{n} \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \frac{|SCC_1|}{n} \xrightarrow{\mathbb{P}} 0.$$

2. If $c > c_p$ then

$$\frac{|SCC_1|}{n} \xrightarrow{\mathbb{P}} \zeta(c).$$

Furthermore, for all $v \in V(G_n)$, with high probability the following holds:

- If $v \notin SCC_1^-$, then $|C^+(v)| = o(n)$.
- If $v \in SCC_1^-$ then $|C^+(v)| = \zeta'(c)n + o(n)$.

Given the above structure of the percolated graph, it is easy to show that the main statement of the paper in fact holds for any family of graphs satisfying the conditions of Theorem 5.

Proposition 7. *Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of (possibly random) graphs satisfying the conditions of Theorem 5. Consider the directed communication model and denote the probability of successful transmission from a vertex to its neighbor by c . Let s be the number of seeds.*

If $c > c_p$, random seeding catches up to the omniscient seeding at an exponential rate in the number of extra seeds, i.e. for any x ,

$$\lim_{n \rightarrow \infty} \frac{\mathbf{H}(\text{RAND}, s+x)}{\mathbf{H}(\text{OMN}, s)} = 1 - (1 - \alpha)^{s+x}.$$

If $c < c_p$, then any seeding strategy diffuses to only a vanishing fraction of the population:

$$\lim_{n \rightarrow \infty} \mathbf{H}(\text{OMN}, s) = 0.$$

1.4 Application to networks in preferential and configuration models

Theorems 5 and 6 are applicable to a wide range of network models, including random graphs in both configuration model and preferential attachment models, as well as Watts-Strogatz small-world networks.

In order to apply our results to the configuration model, we first need to describe the conditions for the existence of the local weak limits. To formulate it, let D_n be the degree of a uniform random node in the degree sequence $\mathbf{d}_n$. Recall the definition of the empirical distribution F_n of the sequence $\mathbf{d}_n$,

$$F_n(x) = \frac{1}{n} \sum_{j \in [n]} \mathbf{1}_{d_j \leq x}.$$

Condition 1.1. *Let $\mathbf{d}_n = (d_1, \dots, d_n)$ be a graphical degree sequence. Then there exists a random variable D such that*

1. $\mathbb{P}(D \geq 3) = 1$ and as $n \rightarrow \infty$, $D_n \xrightarrow{d} D$, and $\mathbb{E}[D_n] \rightarrow \mathbb{E}[D] < \infty$.
2. The empirical distribution F_n satisfies, $[1 - F_n](x) \leq c_F x^{-\tau-1}$, for some $c_F > 0$ and $\tau \in (2, 3)$.

For the preferential attachment model $\{PA\}_{m,n}$, we follow the set up of Alimohammadi et al. (2023). The model has a parameter $m \in \mathbb{N}$, and is defined as follows. Starting from a connected graph G_{t_0} on at least m vertices, a random graph G_t is defined inductively: given G_{t-1} and its degree sequence $d_i(t-1)$, we form a new graph by adding one more vertex, v_t , and connect it to m distinct vertices $w_1, \dots, w_m \in V(G_{t-1})$ by first choosing $w_1, \dots, w_m \in V(G_{t-1})$ i.i.d with distribution $\mathbb{P}(w_s = i) = \frac{d_i(t-1)}{2|E(G_{t-1})|}$, $s = 1, \dots, m$, and then conditioning on all vertices being distinct (thus avoiding multiple edges).

Proposition 8. *The conditions of Theorem 5 holds for the following models:*

- $\{G_n\} = \{PA\}_{m,n}$, for $m \geq 2$, that is networks of size n in the preferential attachment model

- $\{G_n\}_{n \in \mathbb{N}} = \{CM^*(\mathbf{d}_n)\}_{n \in \mathbb{N}}$, where $\{\mathbf{d}_n\}_{n \in \mathbb{N}}$ is a graphical degree sequence satisfying Condition 1.1.

It is possible to generalize the above theorem to “household” or “motif-based” preferential or configuration models in which every vertex is replaced with a small subgraph sampled from a give distribution. We refer the reader to Alimohammadi et al. (2022) for a more detailed discussion.

1.5 Application to Watts-Strogatz models

Using the above formalization, we can show that our results hold for Watts-Strogatz small-world graphs. Not surprisingly, the answer is positive. Let us start with a natural formalization for such networks following Barbour and Reinert (2006).

Definition 9 (Small-world network). *A Small-World network SW_n of size n is comprised of $n = Ld$ vertices around a circle, where each vertex is connected to all its neighbors within distance d by an undirected edge. Shortcuts or long-range links are formed by adding edges between every pair of non-adjacent vertices with probability q/n .*

The local limit of the family $\{SW_n\}$ is shown to be a two-type branching process as in Barbour and Reinert (2006). These graphs are known to be expanders as well Flaxman (2007). Therefore, we have:

Proposition 10. *Small-world networks have a solar system structure. Therefore, they satisfy the conditions of Theorem 2 and Theorem 6.*

1.6 Application to power-law networks with homophily

There are different ways to introduce homophily in preferential attachment networks. We build on the island model discussed in the paper. Specifically, we adapt the preferential attachment model with global communities as discussed in Hofstad (2024), (section 9.3.4),

Start with an initial graph G_{n_0} where each vertex v has out-degree m and a type $\sigma(v) \in [t]$. At time $n+1$, a new vertex $n+1$ arrives with type $\sigma(n+1)$ chosen iid from $[t]$ with distribution $\mu(s) = P(\sigma(n+1) = s)$. Vertex $n+1$ has m out-half edges, each labeled with $\sigma(n+1)$. An affinity matrix $\kappa : [t] \times [t] \rightarrow [0, \infty)$ assigns weights to half-edges where half-edge x with label r has weight $\kappa(\sigma(n+1), r)$. Each of the m half-edges of $n+1$ chooses an existing half-edge to connect to proportional to the weights $\kappa(\sigma(n+1), r)$. This process is repeated indefinitely to grow graph G_n .

Now, consider the above process with an affinity matrix κ whose entries are positive and is diagonally dominant to support homophily. The resulting network has a power-law degree sequence, as proved in Theorem 9.14 in Hofstad (2024). The local limit and expansion of these models can be derived similar to the argument in Alimohammadi et al. (2022) for preferential attachment networks.

Proposition 11. *Power-law networks with homophily as defined above have a solar system structure. Therefore, they satisfy the conditions of Theorem 2 and Theorem 6.*

References

- Aldous, D. and J.M. Steele**, “The objective method: probabilistic combinatorial optimization and local weak convergence,” in “Probability on discrete structures,” Vol. **110** of *Encyclopaedia Math. Sci.*, Springer, Berlin, 2004, pp. 1–72.
- Alimohammadi, Yeganeh, Christian Borgs, and Amin Saberi**, “Algorithms using local graph features to predict epidemics,” in “Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)” SIAM 2022, pp. 3430–3451.
- , —, and —, “Locality of random digraphs on expanders,” *The Annals of Probability*, 2023, **51** (4), 1249–1297.
- Barbour, Andrew and Gesine Reinert**, “Discrete small world networks,” *Electronic Journal of Probability [electronic only]*, 2006, **11**, 1234–1283.
- Benjamini, I. and O. Schramm**, “Recurrence of distributional limits of finite planar graphs,” *Electron. J. Probab.*, 2001, **6**, no. 23, 13 pp. (electronic).
- Flaxman, Abraham D**, “Expansion and lack thereof in randomly perturbed graphs,” *Internet Mathematics*, 2007, **4** (2-3), 131–147.
- van der Hofstad, R.**, “Random graphs and complex networks. Vol. 2,” 2024. In preparation, see <http://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf>.
- Watts, D.J. and S.H. Strogatz**, “Collective dynamics of small-world networks,” *Nature*, 1998, **393** (6684), 440–442.