Local Algorithms For Predicting Epidemics

Yeganeh Alimohammadi

Stanford University

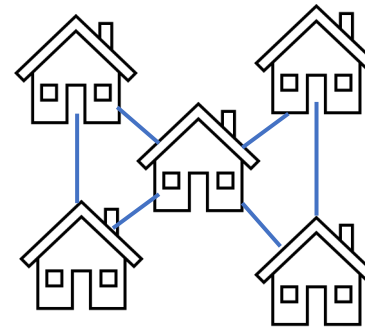# Motivation: Infections Spreading Over a Network

Infections spread from person-to-person contacts.

The dynamics of the infection spread depends on the contact network:

isolated cliques of households

connected network with super-spreaders

What information about the network do we need to predict epidemics?

The Epidemic Model

o State of nodes: Susceptible (S), Infectious (I), or Recovered (R).

o When a node is infected, it stays in state I for time T. Then it becomes R.

o Transmission rate from an infected node to its susceptible neighbor: β.
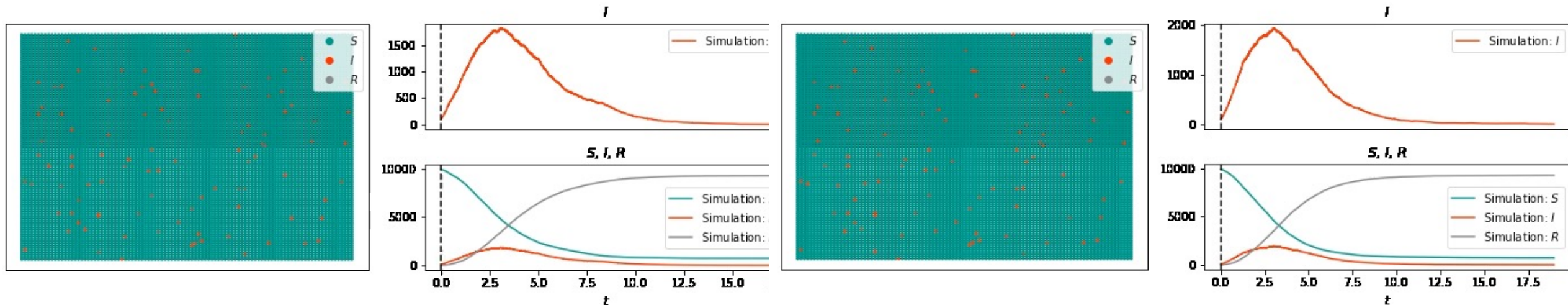Typical choices: β=const, T=const or $T \sim \exp(\gamma)$.
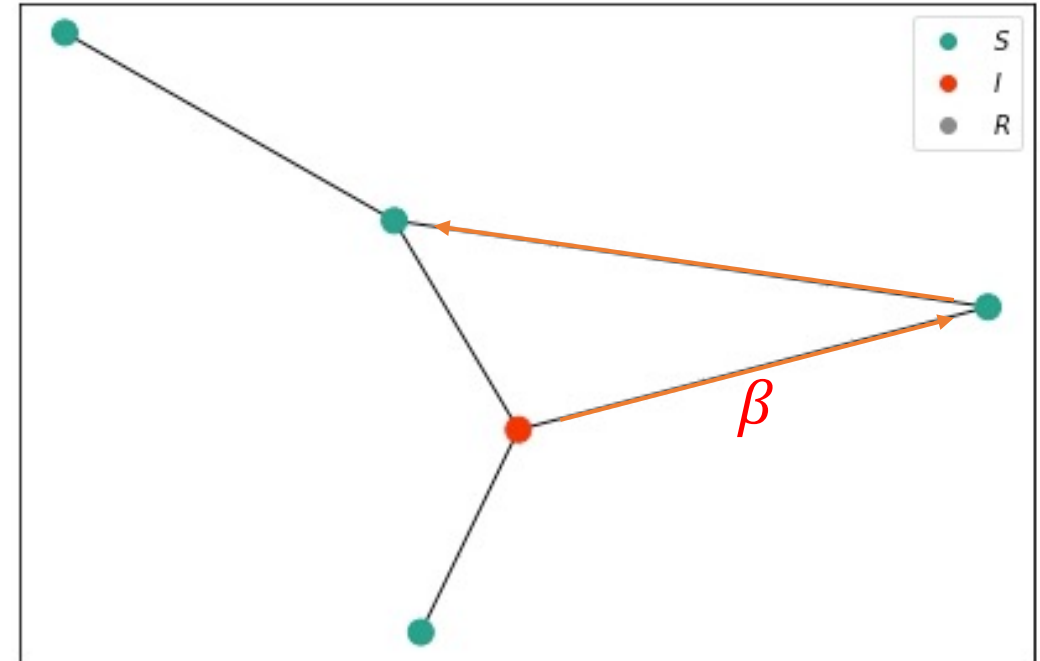
Epidemics in Parallel Worlds

A Thought experiment:

Two copies of the same network, with infection starting from a same set of nodes.

- Sometimes the infection dies out fast. What's the probability of an outbreak?

- If it leads to an outbreak, can we predict the size?

- Can we predict the time dynamic of S, I, R?

Predicting Epidemics Using Network Models

Step 1. Choose a network model to capture the interaction between individuals:

o   Erdos Renyi

o   Configuration Model [Molloy, Reed, Newman, Barabasi, Watts '11][Janson, Luczak, Windridge '14]

o   Preferential Attachment [Bollobás, Riordan '03]

o   Stochastic Block Model [Britton, Pardoux '18]

o   Household models [Ball, Sirl, Trapman. 2009, Hofstad, Leeuwaarden, Stegehuis. '15 -- for configuration model]
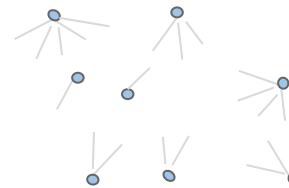
o   …

Predicting Epidemics Using Network Models

Step 1. Choose a network model to capture the interaction between individuals:
Erdos Renyi, Configuration Model, Preferential Attachment, Stochastic Block Model, Household Models, etc.

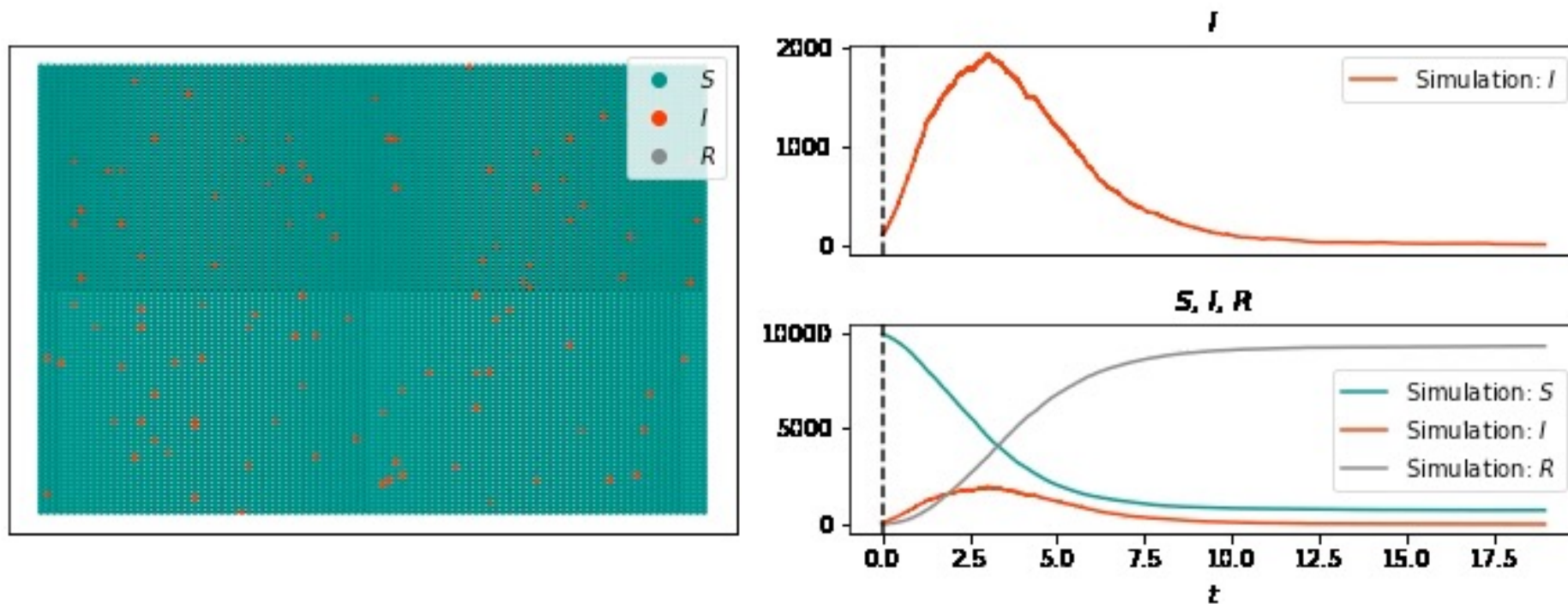**Step 2.** Estimate the relevant model parameters.

Erdös-Renyi: average degree

Configuration model: degree sequence

Can we have a model-free estimation of Epidemics?
e.g., can we predict probability and size of outbreak, or time evolution?

This Talk in a Nutshell:
Local information is enough. You don't need a network model!

Initial condition: each node is I with probability α, and S otherwise.

Under a general SIR process, and general underlying network:

Local information is enough to estimate the relative **size** of the outbreak and the **time evolution** of the epidemics.
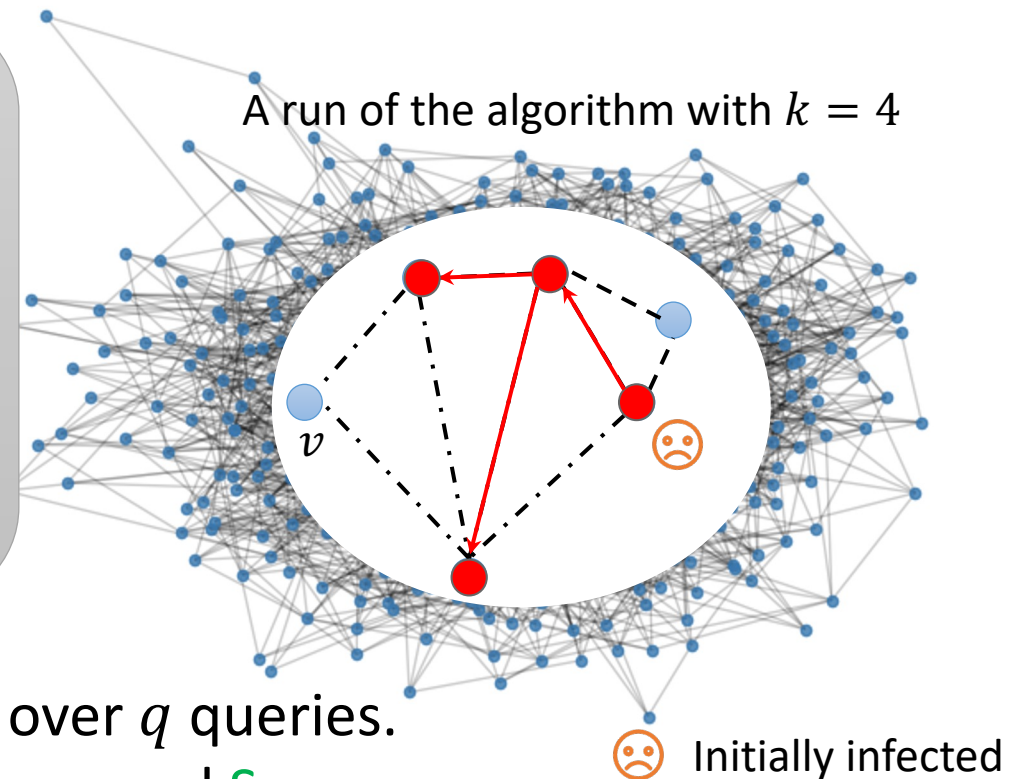
Local Estimator: Time Dynamics

Local SIR:

Input: a constant k, and time t.
1. Draw a uniform random node v.
2. Run the infection in k neighborhood up to time t.
return the status of v at time t

A run of the algorithm with $k = 4$

Estimator: average number of nodes in S, I, or R over q queries.
Initial condition: each node is is I with probability α, and S otherwise

☹ Initially infected

Local Estimator: Relative Size of an Outbreak

Backward Process

Input: a constant k.

1. Draw a uniform random node v.

2. Run the infection **backward** from v.

3. If it reaches to an infected node

within k neighborhood:

  return True.

otherwise:

  return False.

Estimator: average over q queries to the above algorithm.
Equivalently, we could run the local SIR up to time infinity.

A run of the algorithm with $k = 4$

☺ Initially not infected
☹ Initially infected
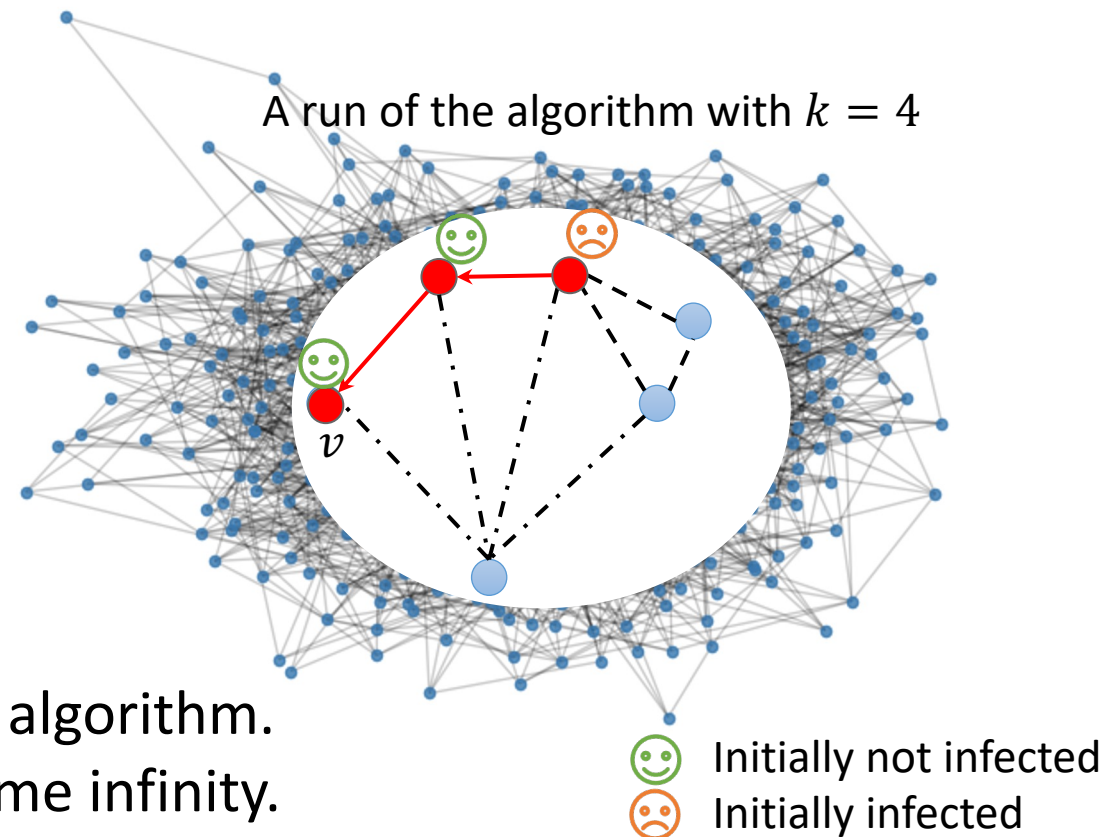
Main Result:

SIR process starting with each node being infected independently with probability α. Let $S_\alpha(t), I_\alpha(t), R_\alpha(t)$ be the number of susceptible, infectious and recovered at t.

Theorem. [**A.**, Borgs, Hofstad, Saberi ('22)]

For *convergent* sequence of graph in probability, and for any $\epsilon > 0$, there exist constants $q_\epsilon, k_\epsilon \geq 0$ such that whp:

1) q_ϵ queries to the backward process with input k_ϵ gives a $(1 - \epsilon)$- approximation of the final size of infection $(\frac{R_\alpha(\infty)}{n})$.

2) q_ϵ queries to the timed process with input k_ϵ, t gives a $(1 - \epsilon)$- approximation of $(\frac{S_\alpha(t)}{n}, \frac{I_\alpha(t)}{n}, \frac{R_\alpha(t)}{n})$.

The same results hold also when $\alpha \to 0$.

A Few Notes

The algorithms:
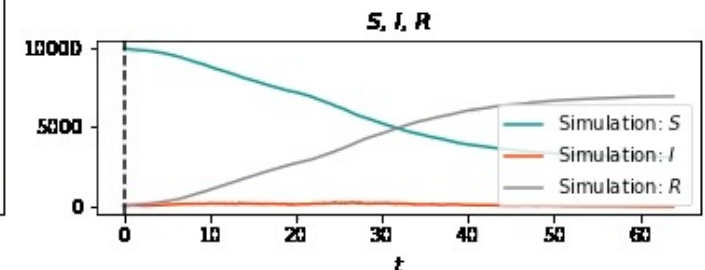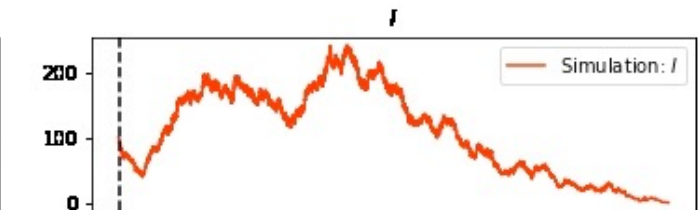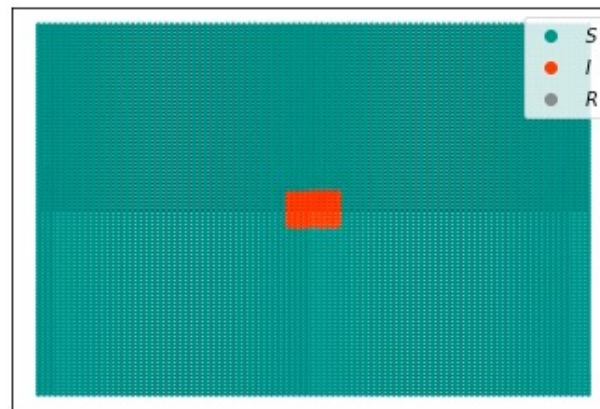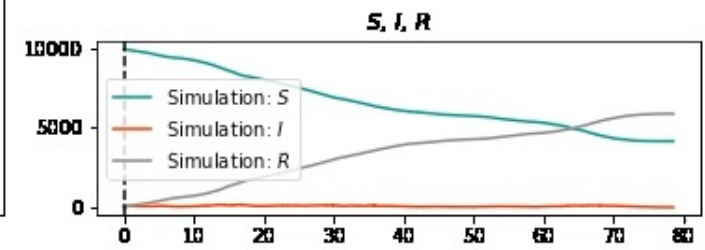
o   can be implemented in $O(1)$

o   is model-free

o   Preserves edge-differential privacy

o   is extendable to dynamic network models

o   Does not require local tree-like structures in the network

Epidemics Starting from One Node

Initial condition: A uniform random node is in I.

Can we still predict the time evolution of epidemics?
Not generally!
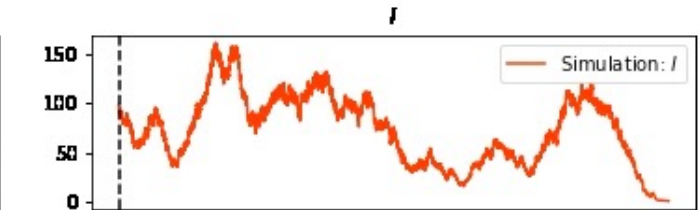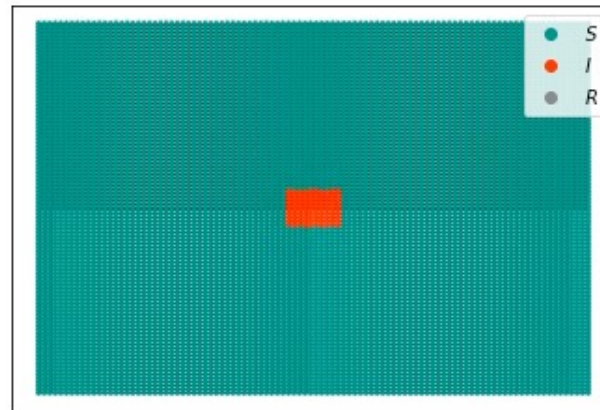
Same Local Structure but Different Outbreaks
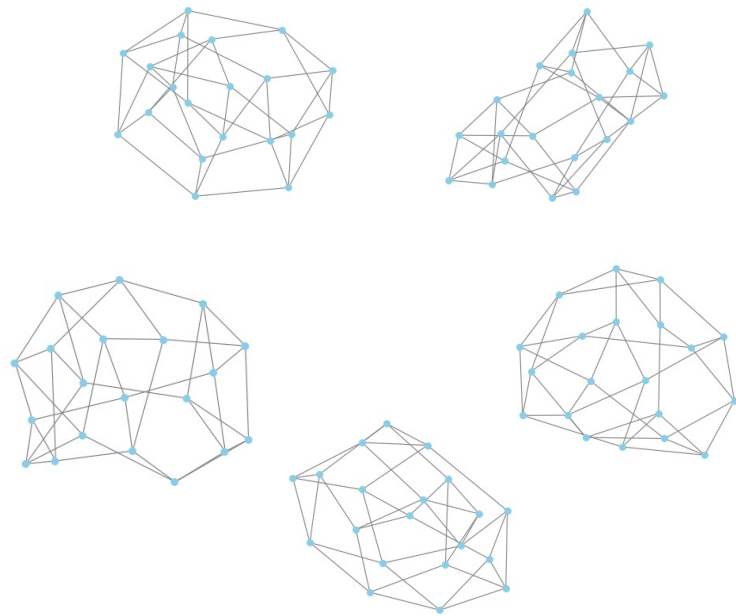
A collection of $\frac{n}{\log n}$ 4-regular random graphs, each of size $\log n$
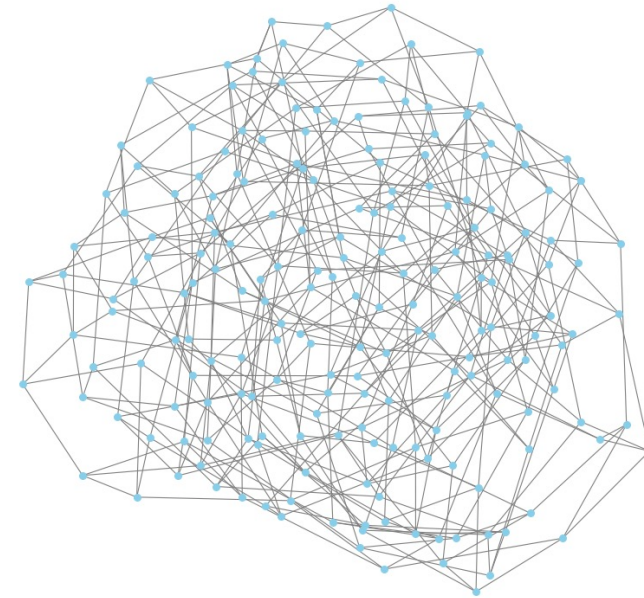
A 4-regular random graph of size n

Epidemics in Well-Connected Network

Initial condition: one uniform random node is I

Simple epidemic model: constant recovery time.

Outbreak: $\omega(n)$ nodes eventually getting infected.

Local information is enough to estimate the relative **size** and **probability** of an outbreak on **well-connected** networks.

Local Estimator: Probability of Outbreak

A run of the algorithm with $k = 4$

Forward Process

Input: a constant k.

1. Draw a uniform random node v.

2. Simulate an infection starting from v.

3. If v can lead to infecting k others:

      return 1.

otherwise:

      return 0.
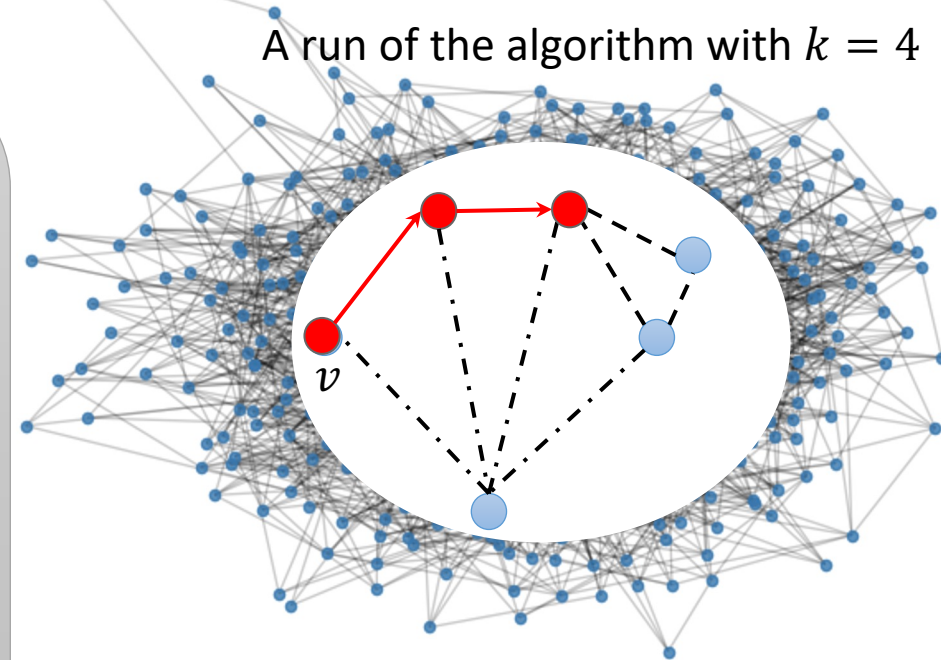
Estimator: average over q queries to the above algorithm.

Outbreak: linear number of nodes eventually getting infected.

Relative Size and Probability of Outbreak

In SIR with constant recovery time, transmission happens over each edge independently with the same probability p.

Theorem 1. [**A.,** Borgs, Saberi '21]

Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of large-set expanders with bounded average degree converging locally in probability to (G, o) with non-random distribution μ. Let $R(\infty)$ be final infection size. Then for $p \neq p_c(\mu)$

$$\frac{R(\infty)}{n} \xrightarrow{\mathbb{P}} \chi_p, \qquad \chi_p = \begin{cases} 0. & \text{with prob } 1 - \zeta(p) \\ \zeta(p). & \text{with prob } \zeta(p) \end{cases}.$$

$\xrightarrow{\mathbb{P}}$: convergence in probability in percolation and μ.

$\zeta(\mathrm{p}) := \mathbb{E}_{(G,o) \sim \mu}[\mathbb{P}_{G(p)}(|\text{connected component of } o| = \infty)]$.

$p_c(\mu) = \inf\{p \in [0,1] : \zeta(p) > 0\}$.

Takeaways:
1. With high probability, the final infection size is either either $O(1)$ or $\zeta(\mathrm{p})n + o(n)$.
2. Size and probability of outbreak are the same in SIR with constant recovery time.

Local Estimator Predicts Size and Probability of Outbreak

A run of the algorithm with $k = 4$

Input: a constant k.
1. Draw a uniform random node v.
2. Simulate an infection starting from v.
3. If v can lead to infecting k others:

   return 1.

otherwise:

   return 0.
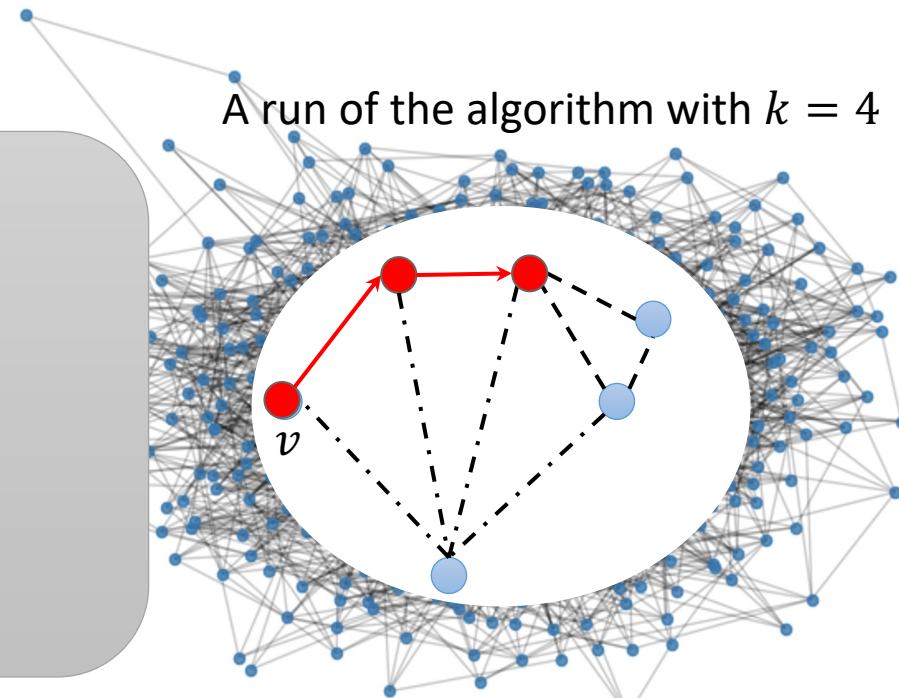
Theorem 2. [**A.**, Borgs, Saberi (SODA'22)]
Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of convergent large-set expanders. Then for any $\epsilon > 0$, there exist constants q_ϵ, $k_\epsilon \geq 0$, such that whp q_ϵ queries to the above algorithm with input k_ϵ **is a $(1 - \epsilon)$-approximation of** $\zeta(p)$ (the relative size / probability of an outbreak).

Takeaway: Size of the outbreak, can be estimated with a constant number of queries with constant k as input. The result is robust to non-exact estimation of the infection probability (p)

Main Technique: Graph Convergence

Definition. (Local Convergence in Probability [Benjamini, Schramm '01])
A sequence of finite graphs $\{G_n\}_{n\in\mathbb{N}}$ converges locally in probability to μ if for any bounded continuous function $f: \mathcal{G}_* \to \mathbb{R}$,

$$\mathbb{E}_{\mathcal{P}_n}[f|G_n] \xrightarrow{\mathbb{P}} \mathbb{E}_\mu[\mathrm{f}],$$

where in $\mathbb{E}_{\mathcal{P}_n}[f|G_n]$, we take expectation with respect to the uniform random root in G_n.

Takeaway: the distribution of the neighborhood of a typical node converges.
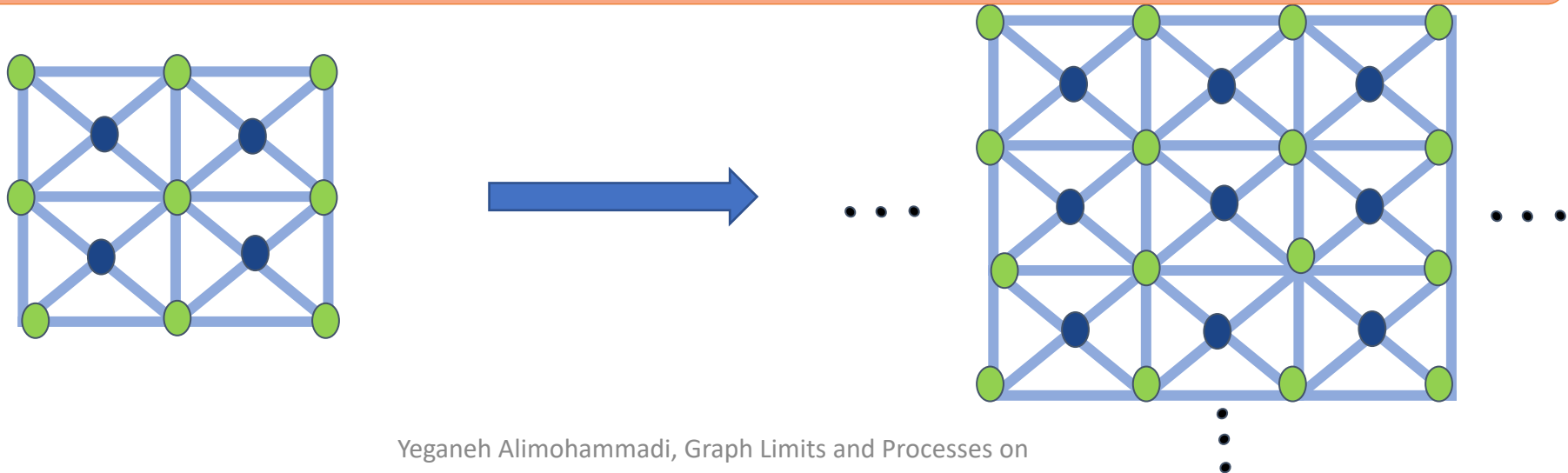
Most Network Locally Converge

o Erdos Renyi → Branching Process

o Configuration Model → Branching Process

o Preferential Attachment → Polya Point Process

o Household models → Households on the limiting external graph
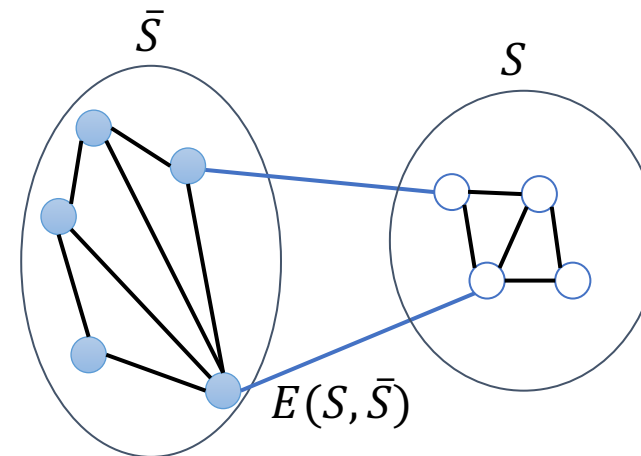
Condition on Network: Expansion

Definition. (Expanders)
G is α-expander if $\phi(G) \geq \alpha$, where

$$\phi(G) = \min_{S \subseteq V(G)} \frac{E(S,\bar{S})}{\min(|S|,|\bar{S}|)}$$

Takeaway: If you want to isolate a large community from the rest of the town, you need to remove many connections.

Large-set Expansion

Definition. (Expander)
G is α-expander if $\phi(G) \geq \alpha$, where

$$\phi(G) = \min_{S \subseteq V(G)} \frac{E(S, \bar{S})}{\min(|S|, |\bar{S}|)}$$

Definition. (Large-set Expander)
G with average degree bounded by d is
(α, ϵ, d) large-set expander if $\phi_\epsilon(G) \geq \alpha$, where

$$\phi_\epsilon(G) = \min_{\substack{S \subseteq V(G) \\ |S| \geq \epsilon n}} \frac{E(S, \bar{S})}{\min(|S|, |\bar{S}|)}$$

Definition. (Sequence of Large-set Expander)
A sequence of possibly random graphs $\{G_n\}_{n \in \mathbb{N}}$ is called a large-set expander sequence with bounded average degree, if there exists $\bar{d} < \infty$ and $\alpha > 0$ such that for all $\epsilon \in (0, .5)$, the probability that G_n is an (α, ϵ, d) large-set expander goes to 1 as $n \to \infty$.

$\bar{S}$

S

$E(S, \bar{S})$
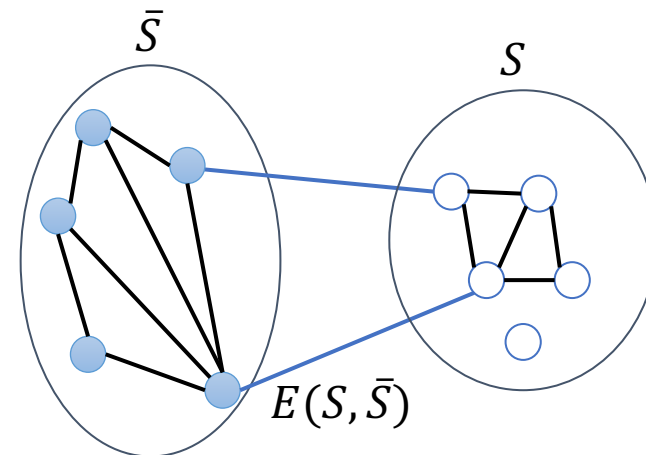
Examples of Convergent Large-set Expanders

o **Configuration Model** [Molloy, Reed, Newman, Barabasi, Watts '11]
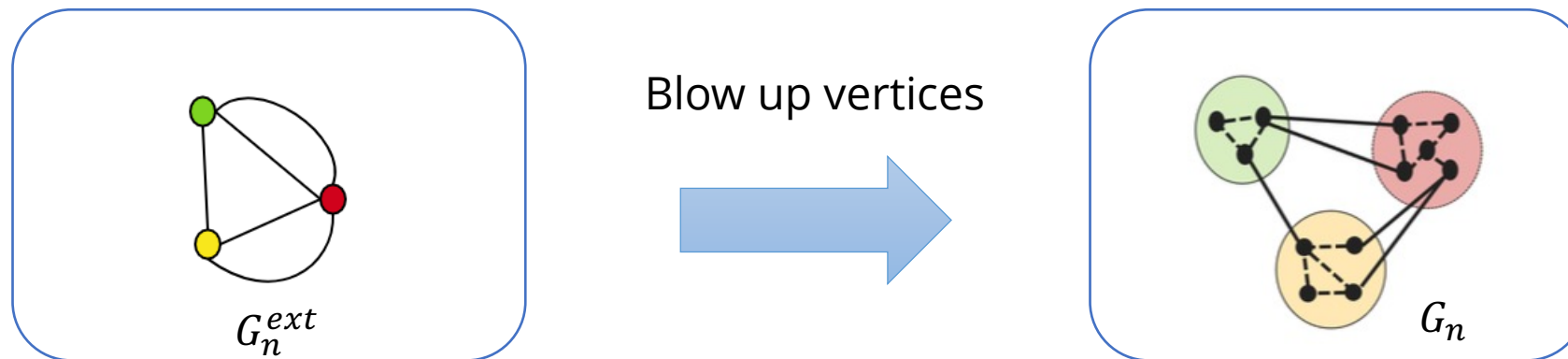
o **Preferential Attachment** [Bollobás, Riordan '03]

o **Household models** [Ball, Sirl, Trapman. 2009, Hofstad, Leeuwaarden, Stegehuis. '15 -- for configuration model]

Blow up vertices

G_n^{ext}

G_n

Informal Lemma. If $\{G_n^{ext}\}_{n \in \mathbb{N}}$ are convergent large-set expanders then $\{G_n\}_{n \in \mathbb{N}}$ is as well.

Does the Algorithm Work on Real-world Graphs?

o Copenhagen dataset
o GPS data of 700 students
o Edge exists if distance <6 ft

Probability of Outbreak: as Measured by K Neighborhoods vs Entire Population
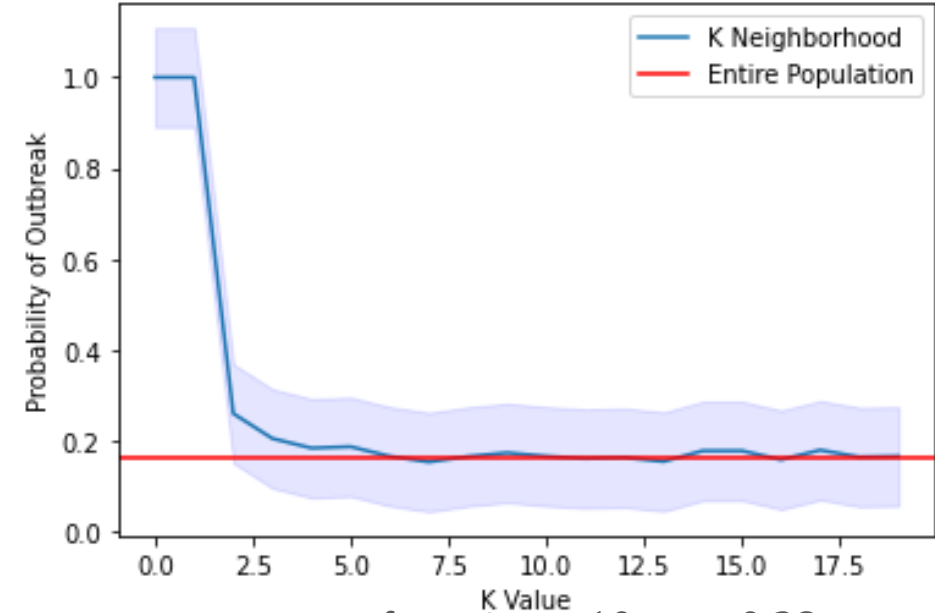
Num of queries $= 10, p = 0.28,$
95% confidence interval is highlighted

Data: "Interaction data: Copenhagen Networks Study"
[Sapiezynski, Stopczynksi, Lassen, & Lehman, Nature '19]

Proof Ideas:

SIR with constant recovery time and percolation

Percolation: keep each edge with probability p (call this graph $\mathrm{G}(p)$).

SIR with constant recovery time can be coupled to percolation.

Largest component (giant) in $\mathrm{G}(p)$ corresponds to an outbreak!

Instead, we can study the size of the giant and its uniqueness.

Relative Size of the Giant in Expanders

Theorem 1. [A., Borgs, Saberi '21]
Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of large-set expanders with bounded average degree converging locally in probability to $(G, o) \in \mathfrak{G}_*$ with non-random distribution μ. Let C_i be the ith largest component. If $p \neq p_c(\mu)$,

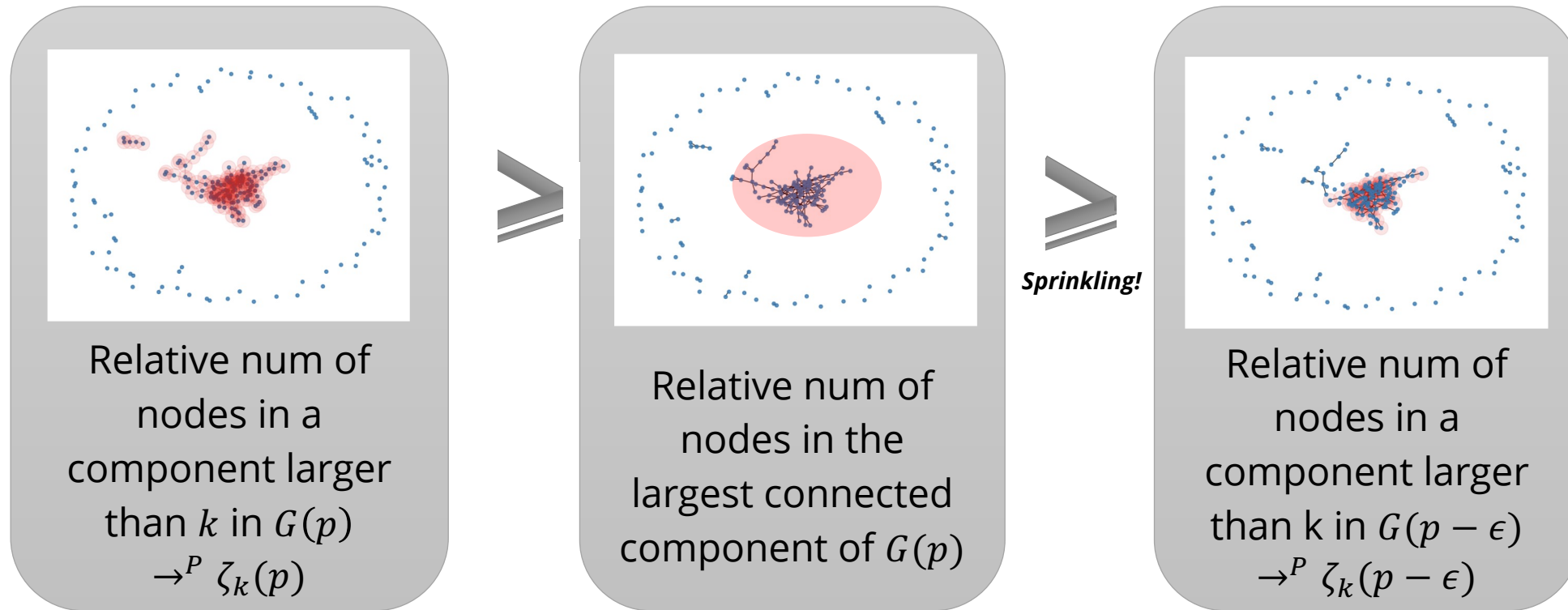$$\frac{|C_1|}{n} \xrightarrow{\mathbb{P}} \zeta(p),$$

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

$\xrightarrow{\mathbb{P}}$: convergence in probability in percolation and μ.
$\zeta(p) := \mathbb{E}_{(G,o) \sim \mu}[\mathbb{P}_{G(p)}(|\text{connected component of } o| = \infty)].$

Takeaway: Giant in convergent expanders is unique, and its size converges to its limit.

Proof Sketch: Size of the Giant Converges

Relative num of nodes in a component larger than k in $G(p)$
$\to^P \zeta_k(p)$

$\geq$

Relative num of nodes in the largest connected component of $G(p)$

$\geq$

Sprinkling!

Relative num of nodes in a component larger than k in $G(p - \epsilon)$
$\to^P \zeta_k(p - \epsilon)$

$$\zeta_k(\mathrm{p}) := \mathbb{E}_{(G,o)\sim\mu}[\mathbb{P}_{G(p)}(|\text{connected component of } o| \geq k)].$$

$$\lim_{k\to\infty} \zeta_k(\mathrm{p}) = \zeta(p).$$

Proof Sketch: Size of the Giant Converges

Relative num of
nodes in a
component larger
than k in $G(p)$
$\to^P \zeta_k(p)$

Converges to $\zeta(p)$

$\geqq$

Relative num of
nodes in the
largest connected
component of $G(p)$

Sprinkling!

$\geqq$

Relative num of
nodes in a
component larger
than k in $G(p-\epsilon)$
$\to^P \zeta_k(p-\epsilon)$

Converges to $\zeta(p-\epsilon)$
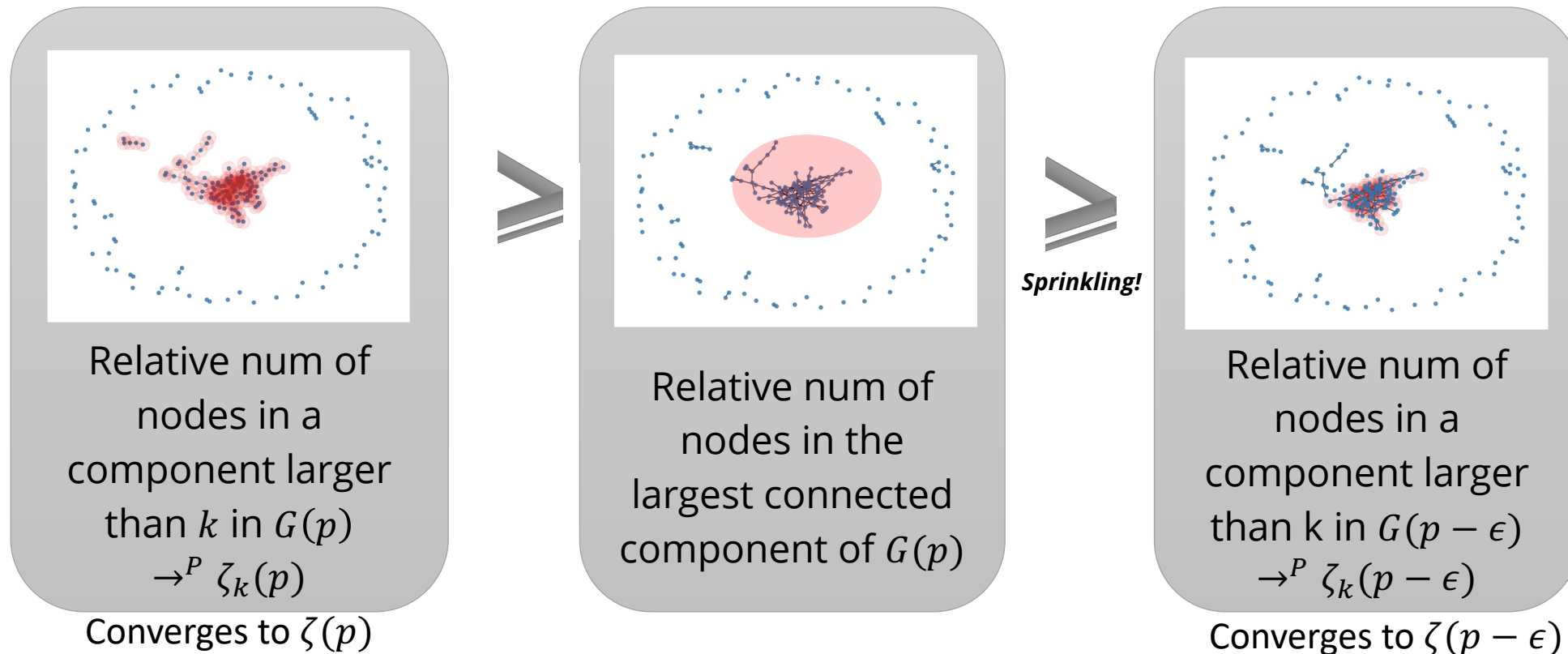
Lemma. For a sequence of graphs satisfying the assumptions of Theorem 2, $\zeta(p)$ is continuous for all $p \neq p_c(\mu)$. Equivalently, the limit μ is ergodic.

(Sourav Sarkar proved this lemma for deterministic sequence of convergent expanders in 2018.)

Brief History of Sprinkling

[Erdös, Rényi'60]
[Posa'76][Ajtai, Kolmós, Szemerédi '82]
[Bollobás, Riordan '01] [Alon, Benjamini, Stacey '02]
[Borgs, Chayes, van der Hofstad, Slade, Spencer '07]
[Benjamini, Nachmias, Peres '09]
[Janson, Rucinski'10] [van der Hofstad, Nachmias '17]
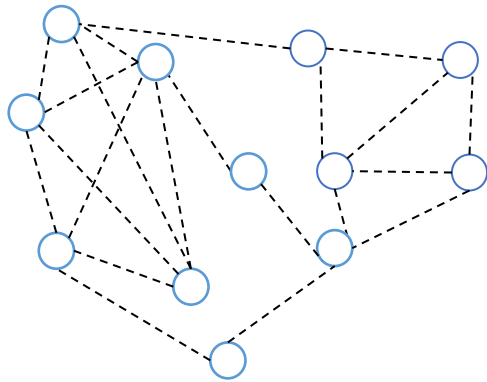[Krivelevich, Sudakov '17]
[Dudek, C. Reiher, A. Rucínski, and M. Schacht '20]
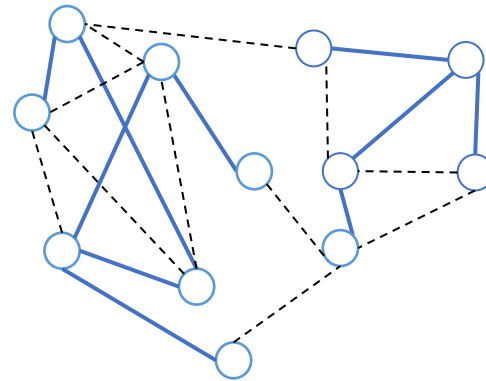[Nenadov, Trujic '21][Easo, Hutchcroft '21]

[**A.,** Borgs, Saberi '21+]

Proof: the Lower Bound

Step 0: For some $\epsilon > 0$ let $p_1 = p_c(\mu) + \epsilon$ be such that $1 - p = (1 - p_1)(1 - \epsilon)$.

Consider two copies of percolation $G_n(p_1)$ and $G_n(\epsilon)$. The union of them gives an instance of $G_n(p)$.

The original graph G_n $G_n(p_1)$ $G_n(\epsilon)$
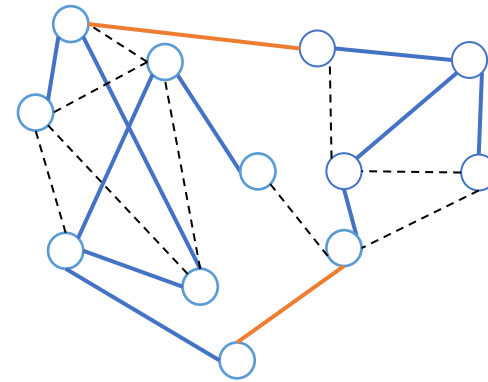
Proof: the Lower Bound

Step 0: For some $\epsilon > 0$ let $p_1 = p_c(\mu) + \epsilon$ be such that $1 - p = (1 - p_1)(1 - \epsilon)$.
Consider two copies of percolation $G_n(p_1)$ and $G_n(\epsilon)$. The union of them gives an instance of $G_n(p)$.

Step 1: There exists some $\delta > 0$ such that for all $K > 0$, whp there are δn nodes with component larger than K in $G_n(p_1)$.

Step 2 (Sprinkling): Let $Z_K = \{$nodes with component larger than $K\}$.
There is a path in $G_n(\epsilon)$ between any two large partition of components in Z_K:

$$\mathbb{P}_{G_n(\epsilon)}\left(\exists\, A, B \subseteq 2^{Z_K}:\ A\,, B \text{ disconnected in } G_n(\epsilon) \text{ and } G_n(p_1),\ |A|, |B| \geq \frac{\delta n}{3} \mid G_n(p_1)\right)$$
$$\leq \exp(-n c_{\{\alpha, \delta, d, \epsilon\}})$$

Step 3: $\mathbb{P}_{G_n(p)}\left(\text{contains a component of size } \frac{\delta n}{3}\right) \to 1,\ \text{as } n \to \infty.$

Step 2: Sprinkling

Step 2 (Sprinkling): There is a path in $G_n(\epsilon)$ between any two large partition of components in Z_K:
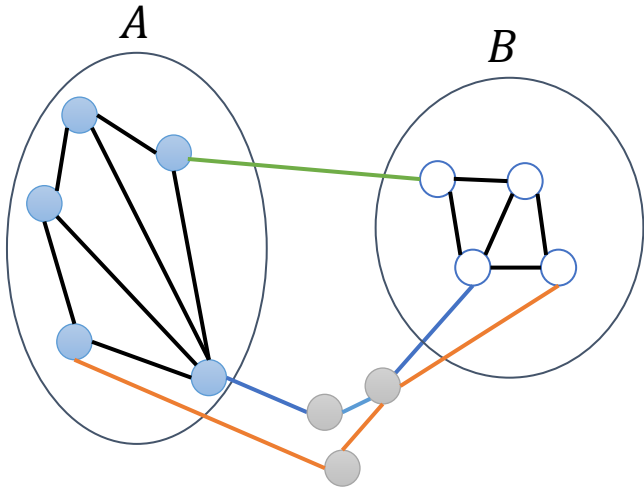
$$\mathbb{P}_{G_n(\epsilon)}\left(\exists\, A, B \subseteq 2^{Z_K}: A, B \text{ disconnected in } G_n(\epsilon) \text{ and } G_n(p_1), |A|,|B| \geq \frac{\delta n}{3} \mid G_n(p_1)\right) \leq \exp(-nc_{\{\alpha,\delta,d,\epsilon\}})$$

A

B

Menger's Theorem. Let G be a finite undirected graph and A and B two disjoint set of vertices. Then the minimum edge-cut between A and B is equal to the number of pairwise <u>edge-independent paths</u> from A to B.

There are $\frac{\delta\alpha n}{3}$ edge-disjoint paths in G_n between A and B (expansion).
Since the average degree is bounded by d, the length of half of these paths is bounded by $\ell = \frac{6d}{\delta\alpha}$. (# paths $= \frac{\delta\alpha n}{6}$)
Each path appear in $G_n(\epsilon)$ with probability ϵ^ℓ.
The probability that non of the paths appear in $G_n(\epsilon)$: $\left(1 - \epsilon^\ell\right)^{\#paths}$
Number of A, B partitions in $G_n(p_1)$: $2^{\frac{n}{K}}$
Finally: $2^{\frac{n}{K}}\left(1 - \epsilon^{\frac{6d}{\delta\alpha}}\right)^{\frac{\delta\alpha n}{6}} \leq \exp\left(n(\frac{1}{K} - \frac{\delta\alpha}{6}\epsilon^{\frac{6d}{\delta\alpha}})\right)$

Takeaways: Local Information Goes a Long Way!

Initial condition: α fraction infected.
Local information is enough to estimate the **time evolution** of the epidemics.

Initial condition: one node infected.
Local information is enough to estimate the **probability** and relative **size** of an **outbreak** for large class of networks under a simple infection spread**.**

Graph limits enabled us to analyze global quantities with local structures. Can we find more applications?

References

Alimohammadi, Borgs, Saberi, "*Algorithms Using Local Graph Features to Predict Epidemics*" (SODA' 2022)

Alimohammadi, Borgs, Saberi, "*Locality of Random Diagraphs on Expanders*" (arxiv, 2021+)

Alimohammadi, Borgs, van der Hofstad, Saberi, "*Epidemics on Networks is Local.*" (working paper, soon on arxiv)

Thank You

yeganeh@stanford.edu