

Epidemics on random intersection graphs

Pieter Trapman

Department of Mathematics, Stockholm University, Sweden

YEP VIII, EURANDOM Eindhoven, March 15, 2011

joint work with

Frank Ball and David Sirl (University of Nottingham)

arXiv:1011.4242v1

Overlapping cliques

Many real social networks are rather “clique-based”.

Individuals are part of households, in which everybody contacts everybody.

They are also part of work spaces, school classes, groups of friends etc.

To model SIR epidemics on networks incorporating this clique structure and obtain analytic results: bipartite graphs (Newman (2002)) or random intersection graphs (Britton, Deijfen, Lagerås & Lindholm (2008) and Deijfen & Kets (2009)).

Overlapping cliques

Many real social networks are rather “clique-based”.

Individuals are part of households, in which everybody contacts everybody.

They are also part of work spaces, school classes, groups of friends etc.

To model SIR epidemics on networks incorporating this clique structure and obtain analytic results: bipartite graphs (Newman (2002)) or random intersection graphs (Britton, Deijfen, Lagerås & Lindholm (2008) and Deijfen & Kets (2009)).

Drawbacks of existing models and analysis

- Analysis of epidemics on those networks with random infectious periods is not done yet
- Models are mainly studied as tools to create networks with clustering, not as models for real-life networks in itself

The graph

We use a bipartite random intersection-like graph:

- Two types of vertices: **persons** and **cliques**
- The number of persons is n and the number of cliques is m , where $m = \lfloor \alpha n \rfloor$
- Persons receive weights distributed as A and cliques receive weights distributed as B , with

$$\mu := \alpha \mathbb{E}(B) = \mathbb{E}(A) < \infty$$

- Create an intermediate graph \mathcal{A} by independently connecting persons with cliques. A person with weight a_i shares a Poisson number of edges with a clique of weight b_j with mean $a_i b_j / (\mu n)$

- A person of weight a has degree $\text{Poisson}(a \sum_{i=1}^{\lfloor \alpha n \rfloor} b_i / (\mu n))$
and a clique of weight b has degree $\text{Poisson}(b \sum_{i=1}^n a_i / (\mu n))$
- By the Law of Large Numbers, the degree distributions are close to $\text{Poisson}(a)$ and $\text{Poisson}(b)$
- Parallel edges and self-loops are sparse
- The graph G is obtained by projecting away the clique vertices:
i.e., connecting pairs of persons that have distance 2 in \mathcal{A} and then removing all the clique vertices.
The clique vertices in \mathcal{A} now correspond to complete subgraphs in G : **The cliques**

- A person of weight a has degree $\text{Poisson}(a \sum_{i=1}^{\lfloor \alpha n \rfloor} b_i / (\mu n))$ and a clique of weight b has degree $\text{Poisson}(b \sum_{i=1}^n a_i / (\mu n))$
- By the Law of Large Numbers, the degree distributions are close to $\text{Poisson}(a)$ and $\text{Poisson}(b)$
- Parallel edges and self-loops are sparse
- The graph G is obtained by projecting away the clique vertices:
i.e., connecting pairs of persons that have distance 2 in \mathcal{A} and then removing all the clique vertices.
The clique vertices in \mathcal{A} now correspond to complete subgraphs in G : **The cliques**

- A person of weight a has degree $\text{Poisson}(a \sum_{i=1}^{\lfloor \alpha n \rfloor} b_i / (\mu n))$ and a clique of weight b has degree $\text{Poisson}(b \sum_{i=1}^n a_i / (\mu n))$
- By the Law of Large Numbers, the degree distributions are close to $\text{Poisson}(a)$ and $\text{Poisson}(b)$
- Parallel edges and self-loops are sparse
- The graph G is obtained by projecting away the clique vertices:
i.e., connecting pairs of persons that have distance 2 in \mathcal{A} and then removing all the clique vertices.
The clique vertices in \mathcal{A} now correspond to complete subgraphs in G : **The cliques**

- The number of cliques a vertex of weight a is part of, $X := X(a)$, is (asymptotically) characterized by the generating function

$$f_X(s; a) = \mathbb{E}(s^{X(a)}) = e^{-a(1-s)}$$

The generating function of the unconditional distribution of the number of cliques a vertex is part of is given by

$$f_X(s) = \mathbb{E}_A(e^{-A(1-s)})$$

- The number of vertices in a clique with clique-weight b , $Y := Y(b)$, is characterized by the generating function

$$f_Y(s; b) = \mathbb{E}(s^{Y(b)}) = e^{-b(1-s)}$$

The generating function of the unconditional distribution of the number of vertices in a clique is given by

$$f_Y(s) = \mathbb{E}_B(e^{-B(1-s)})$$

- The number of cliques a vertex of weight a is part of, $X := X(a)$, is (asymptotically) characterized by the generating function

$$f_X(s; a) = \mathbb{E}(s^{X(a)}) = e^{-a(1-s)}$$

The generating function of the unconditional distribution of the number of cliques a vertex is part of is given by

$$f_X(s) = \mathbb{E}_A(e^{-A(1-s)})$$

- The number of vertices in a clique with clique-weight b , $Y := Y(b)$, is characterized by the generating function

$$f_Y(s; b) = \mathbb{E}(s^{Y(b)}) = e^{-b(1-s)}$$

The generating function of the unconditional distribution of the number of vertices in a clique is given by

$$f_Y(s) = \mathbb{E}_B(e^{-B(1-s)})$$

- The number of other vertices in a clique of a given vertex is not distributed as $Y - 1$, but according to the size biased distribution, with generating function

$$f_{\bar{Y}}(s) = \frac{\mathbb{E}_B(Be^{-B(1-s)})}{\mathbb{E}(B)}.$$

- The degree distribution of the vertices is characterized by the generating function $f_X(f_{\bar{Y}}(s))$. This can be used to compute the expected degree and second moment of degree distribution,

$$\begin{aligned} \mathbb{E}(D) &= \mathbb{E}(X\bar{Y}) = \alpha\mathbb{E}(B^2) \\ \mathbb{E}(D(D-1)) &= \mathbb{E}(X\bar{Y}(\bar{Y}-1)) + \mathbb{E}(X(X-1))(\mathbb{E}(\bar{Y}))^2 \\ &= \alpha\mathbb{E}(B^3) + \mathbb{E}(A^2) \left[\frac{\mathbb{E}(B^2)}{\mathbb{E}(B)} \right]^2 \end{aligned}$$

- The number of other vertices in a clique of a given vertex is not distributed as $Y - 1$, but according to the size biased distribution, with generating function

$$f_{\bar{Y}}(s) = \frac{\mathbb{E}_B(Be^{-B(1-s)})}{\mathbb{E}(B)}.$$

- The degree distribution of the vertices is characterized by the generating function $f_X(f_{\bar{Y}}(s))$. This can be used to compute the expected degree and second moment of degree distribution,

$$\begin{aligned} \mathbb{E}(D) &= \mathbb{E}(X\bar{Y}) = \alpha\mathbb{E}(B^2) \\ \mathbb{E}(D(D-1)) &= \mathbb{E}(X\bar{Y}(\bar{Y}-1)) + \mathbb{E}(X(X-1))(\mathbb{E}(\bar{Y}))^2 \\ &= \alpha\mathbb{E}(B^3) + \mathbb{E}(A^2) \left[\frac{\mathbb{E}(B^2)}{\mathbb{E}(B)} \right]^2 \end{aligned}$$

- Triangles of vertices that are not all in the same clique are sparse
- The expected number of ordered triangles per clique is (by the Law of Large Numbers) $\mathbb{E}[Y(Y-1)(Y-2)] = \mathbb{E}(B^3)$.
- The expected number of ordered triples of which a given vertex is the center is

$$\mathbb{E}[D(D-1)] = \alpha\mathbb{E}(B^3) + \mathbb{E}(A^2) \left[\frac{\mathbb{E}(B^2)}{\mathbb{E}(B)} \right]^2$$

- The number of ordered triangles divided by the number of ordered triples in the graph converges to

$$\begin{aligned} c &= \frac{m\mathbb{E}(B^3)}{n(\alpha\mathbb{E}(B^3) + \mathbb{E}(A^2)[\mathbb{E}(B^2)/\mathbb{E}(B)]^2)} \\ &= \left[1 + \frac{\mathbb{E}(A^2)[\mathbb{E}(B^2)]^2}{\mathbb{E}(A)\mathbb{E}(B)\mathbb{E}(B^3)} \right]^{-1} \end{aligned}$$

- Triangles of vertices that are not all in the same clique are sparse
- The expected number of ordered triangles per clique is (by the Law of Large Numbers) $\mathbb{E}[Y(Y-1)(Y-2)] = \mathbb{E}(B^3)$.
- The expected number of ordered triples of which a given vertex is the center is

$$\mathbb{E}[D(D-1)] = \alpha \mathbb{E}(B^3) + \mathbb{E}(A^2) \left[\frac{\mathbb{E}(B^2)}{\mathbb{E}(B)} \right]^2$$

- The number of ordered triangles divided by the number of ordered triples in the graph converges to

$$\begin{aligned} c &= \frac{m \mathbb{E}(B^3)}{n(\alpha \mathbb{E}(B^3) + \mathbb{E}(A^2) [\mathbb{E}(B^2)/\mathbb{E}(B)]^2)} \\ &= \left[1 + \frac{\mathbb{E}(A^2) [\mathbb{E}(B^2)]^2}{\mathbb{E}(A) \mathbb{E}(B) \mathbb{E}(B^3)} \right]^{-1} \end{aligned}$$

SIR epidemic model

- Neighbors in the graph make contacts according to independent Poisson processes with intensity 1
- When a **S**usceptible contacts an infective, it becomes **I**nfectious immediately
- Infectious vertices stay so for a random time, distributed as \mathcal{I} , the lengths of infectious periods are independent
- After the infectious period vertices **R**ecover and stay immune forever

Remark: The randomness of the infectious period makes analysis of the epidemic non-trivial

SIR epidemic model

- Neighbors in the graph make contacts according to independent Poisson processes with intensity 1
- When a **S**usceptible contacts an infective, it becomes **I**nfectious immediately
- Infectious vertices stay so for a random time, distributed as \mathcal{I} , the lengths of infectious periods are independent
- After the infectious period vertices **R**ecover and stay immune forever

Remark: The randomness of the infectious period makes analysis of the epidemic non-trivial

Locally dependent percolation

- Start with a given directed graph $G' = (V, E')$, where all edges in G are replaced by two directed edges in opposite directions
- Assign i.i.d. “(pseudo-)infectious periods” to the vertices $\{\mathcal{I}_v; v \in V\}$
- An edge starting at v is open with probability $1 - e^{-\mathcal{I}_v}$ and conditioned on $\{\mathcal{I}_v; v \in V\}$, the states of edges are independent
- The law of the vertices that can be reached by an open path from the origin is the same as the law of vertices ultimately recovered vertices

Locally dependent percolation

- Start with a given directed graph $G' = (V, E')$, where all edges in G are replaced by two directed edges in opposite directions
- Assign i.i.d. “(pseudo-)infectious periods” to the vertices $\{\mathcal{I}_v; v \in V\}$
- An edge starting at v is open with probability $1 - e^{-\mathcal{I}_v}$ and conditioned on $\{\mathcal{I}_v; v \in V\}$, the states of edges are independent
- The law of the vertices that can be reached by an open path from the origin is the same as the law of vertices ultimately recovered vertices

Locally dependent percolation

- Start with a given directed graph $G' = (V, E')$, where all edges in G are replaced by two directed edges in opposite directions
- Assign i.i.d. “(pseudo-)infectious periods” to the vertices $\{\mathcal{I}_v; v \in V\}$
- An edge starting at v is open with probability $1 - e^{-\mathcal{I}_v}$ and conditioned on $\{\mathcal{I}_v; v \in V\}$, the states of edges are independent
- The law of the vertices that can be reached by an open path from the origin is the same as the law of vertices ultimately recovered vertices

Locally dependent percolation

- Start with a given directed graph $G' = (V, E')$, where all edges in G are replaced by two directed edges in opposite directions
- Assign i.i.d. “(pseudo-)infectious periods” to the vertices $\{\mathcal{I}_v; v \in V\}$
- An edge starting at v is open with probability $1 - e^{-\mathcal{I}_v}$ and conditioned on $\{\mathcal{I}_v; v \in V\}$, the states of edges are independent
- The law of the vertices that can be reached by an open path from the origin is the same as the law of vertices ultimately recovered vertices

Branching process approximations

Ordinary use of branching process approximations for the epidemic process is hard:

- **Problem:** The cliques cause many short cycles in the graph:
Solution: Declare all vertices affected by a local epidemic, children of the first infectious vertex in the clique
- **New problem:** The infectious period of a vertex and its number of siblings are dependent, which implies that the number of siblings and the number of children of a vertex are dependent
Solution: We use a multi (uncountable)-type branching process (cf. Bollobás, Janson and Riordan (2007 & 2010))
The type of a vertex is its infectious period

Branching process approximations

Ordinary use of branching process approximations for the epidemic process is hard:

- **Problem:** The cliques cause many short cycles in the graph:
Solution: Declare all vertices affected by a local epidemic, children of the first infectious vertex in the clique
- **New problem:** The infectious period of a vertex and its number of siblings are dependent, which implies that the number of siblings and the number of children of a vertex are dependent
Solution: We use a multi (uncountable)-type branching process (cf. Bollobás, Janson and Riordan (2007 & 2010))
The type of a vertex is its infectious period

Branching process approximations

Ordinary use of branching process approximations for the epidemic process is hard:

- **Problem:** The cliques cause many short cycles in the graph:
Solution: Declare all vertices affected by a local epidemic, children of the first infectious vertex in the clique
- **New problem:** The infectious period of a vertex and its number of siblings are dependent, which implies that the number of siblings and the number of children of a vertex are dependent
Solution: We use a multi (uncountable)-type branching process (cf. Bollobás, Janson and Riordan (2007 & 2010))
The type of a vertex is its infectious period

Branching process approximations

Ordinary use of branching process approximations for the epidemic process is hard:

- **Problem:** The cliques cause many short cycles in the graph:
Solution: Declare all vertices affected by a local epidemic, children of the first infectious vertex in the clique
- **New problem:** The infectious period of a vertex and its number of siblings are dependent, which implies that the number of siblings and the number of children of a vertex are dependent

Solution: We use a multi (uncountable)-type branching process (cf. Bollobás, Janson and Riordan (2007 & 2010))
The type of a vertex is its infectious period

Branching process approximations

Ordinary use of branching process approximations for the epidemic process is hard:

- **Problem:** The cliques cause many short cycles in the graph:
Solution: Declare all vertices affected by a local epidemic, children of the first infectious vertex in the clique
- **New problem:** The infectious period of a vertex and its number of siblings are dependent, which implies that the number of siblings and the number of children of a vertex are dependent
Solution: We use a multi (uncountable)-type branching process (cf. Bollobás, Janson and Riordan (2007 & 2010))
The type of a vertex is its infectious period

Branching process approximations

Ordinary use of branching process approximations for the epidemic process is hard:

- **Problem:** The cliques cause many short cycles in the graph:
Solution: Declare all vertices affected by a local epidemic, children of the first infectious vertex in the clique
- **New problem:** The infectious period of a vertex and its number of siblings are dependent, which implies that the number of siblings and the number of children of a vertex are dependent
Solution: We use a multi (uncountable)-type branching process (cf. Bollobás, Janson and Riordan (2007 & 2010))
The type of a vertex is its infectious period

epidemic within a clique

Use special case of results by Ball and O'Neill (1999):

- Initially there are a infectious vertices, and n susceptibles. All have infectious period distribution \mathcal{I} , if infected
- Let $\mathbf{u} := \mathbf{u}(f) := (u_0, u_1, \dots)$ be an infinite vector with $u_k = \mathbb{E}(e^{-k\mathcal{I}} f(\mathcal{I}))$, where f is some function
- Gontcharoff polynomials are defined by the recursive relation:

$$\frac{x^n}{n!} = \sum_{k=0}^n \frac{(u_k)^{n-k}}{(n-k)!} G_k(x|\mathbf{u})$$

Theorem (Ball and O'Neill (1999))

$$\mathbb{E}\left(\prod_{i \in S} f(\mathcal{I}_i)\right) = \sum_{k=0}^n \frac{n!}{(n-k)!} (u_k)^{n-k+a} G_k(1|\mathbf{u})$$

epidemic within a clique

Use special case of results by Ball and O'Neill (1999):

- Initially there are a infectious vertices, and n susceptibles. All have infectious period distribution \mathcal{I} , if infected
- Let $\mathbf{u} := \mathbf{u}(f) := (u_0, u_1, \dots)$ be an infinite vector with $u_k = \mathbb{E}(e^{-k\mathcal{I}} f(\mathcal{I}))$, where f is some function
- Gontcharoff polynomials are defined by the recursive relation:

$$\frac{x^n}{n!} = \sum_{k=0}^n \frac{(u_k)^{n-k}}{(n-k)!} G_k(x|\mathbf{u})$$

Theorem (Ball and O'Neill (1999))

$$\mathbb{E}\left(\prod_{i \in \mathcal{S}} f(\mathcal{I}_i)\right) = \sum_{k=0}^n \frac{n!}{(n-k)!} (u_k)^{n-k+a} G_k(\mathbf{1}|\mathbf{u})$$

- Assume there is one initial infectious vertex in a clique of weight b . The number of susceptibles in the clique is Poisson distributed with mean b
- Let a be the random number of vertices infected directly by an initial infective vertex. If the infectious period of the initial infective is x , then a is $\text{Poisson}([1 - e^{-x}]b)$
- Let n be the random number initial susceptibles that is not directly infected by the initial infectious vertex. n is $\text{Poisson}(e^{-x}b)$
- Conditioned on x , the random variables a and n are independent

- Assume there is one initial infectious vertex in a clique of weight b . The number of susceptibles in the clique is Poisson distributed with mean b
- Let a be the random number of vertices infected directly by an initial infective vertex. If the infectious period of the initial infective is x , then a is $\text{Poisson}([1 - e^{-x}]b)$
- Let n be the random number initial susceptibles that is not directly infected by the initial infectious vertex. n is $\text{Poisson}(e^{-x}b)$
- Conditioned on x , the random variables a and n are independent

- Assume there is one initial infectious vertex in a clique of weight b . The number of susceptibles in the clique is Poisson distributed with mean b
- Let a be the random number of vertices infected directly by an initial infective vertex. If the infectious period of the initial infective is x , then a is $\text{Poisson}([1 - e^{-x}]b)$
- Let n be the random number initial susceptibles that is not directly infected by the initial infectious vertex. n is $\text{Poisson}(e^{-x}b)$
- Conditioned on x , the random variables a and n are independent

- the weight of a clique, containing a given vertex is distributed as \bar{B}
- using Fubini's theorem and

$$\mathbb{E} \left(\prod_{i \in \mathcal{S}} f(\mathcal{I}_i) \right) = \sum_{k=0}^n \frac{n!}{(n-k)!} (u_k)^{n-k+a} G_k(\mathbf{1}|\mathbf{u})$$

for a clique with initially a infectives and n susceptibles, we obtain that

$$\mathbb{E} \left(\prod_{i \in \mathcal{S}} f(\mathcal{I}_i) | \mathcal{I}_0 = x \right) = \mathbb{E}_{\bar{B}} \left(\sum_{k=0}^{\infty} e^{-xk} \bar{B}^k e^{-\bar{B}(1-u_k)} G_k(\mathbf{1}|\mathbf{u}) \right)$$

Threshold parameter

- Filling in $f(x) = s$ gives the generating function of the final size in the clique
- Assume initial infective in clique has infectious period x and all n' other persons are susceptible
- Let $T(n'; x)$ denote the random number of ultimately recovered vertices out of the n' initial susceptibles
- An infected vertex is part of \bar{A} cliques from which it is not infected, and such a clique contains \bar{B} other vertices
- Whether a vertex gets infected is independent of its infectious period
- $R_* = \mathbb{E}(\bar{A})\mathbb{E}(T(\bar{B}, \mathcal{I}))$

Threshold parameter

- Filling in $f(x) = s$ gives the generating function of the final size in the clique
- Assume initial infective in clique has infectious period x and all n' other persons are susceptible
- Let $T(n'; x)$ denote the random number of ultimately recovered vertices out of the n' initial susceptibles
- An infected vertex is part of \bar{A} cliques from which it is not infected, and such a clique contains \bar{B} other vertices
- Whether a vertex gets infected is independent of its infectious period
- $R_* = \mathbb{E}(\bar{A})\mathbb{E}(T(\bar{B}, I))$

Threshold parameter

- Filling in $f(x) = s$ gives the generating function of the final size in the clique
- Assume initial infective in clique has infectious period x and all n' other persons are susceptible
- Let $T(n'; x)$ denote the random number of ultimately recovered vertices out of the n' initial susceptibles
- An infected vertex is part of \bar{A} cliques from which it is not infected, and such a clique contains \bar{B} other vertices
- Whether a vertex gets infected is independent of its infectious period
- $R_* = \mathbb{E}(\bar{A})\mathbb{E}(T(\bar{B}, \mathcal{I}))$

Threshold parameter

- Filling in $f(x) = s$ gives the generating function of the final size in the clique
- Assume initial infective in clique has infectious period x and all n' other persons are susceptible
- Let $T(n'; x)$ denote the random number of ultimately recovered vertices out of the n' initial susceptibles
- An infected vertex is part of \bar{A} cliques from which it is not infected, and such a clique contains \bar{B} other vertices
- Whether a vertex gets infected is independent of its infectious period
- $R_* = \mathbb{E}(\bar{A})\mathbb{E}(T(\bar{B}, \mathcal{I}))$

Threshold parameter

- Filling in $f(x) = s$ gives the generating function of the final size in the clique
- Assume initial infective in clique has infectious period x and all n' other persons are susceptible
- Let $T(n'; x)$ denote the random number of ultimately recovered vertices out of the n' initial susceptibles
- An infected vertex is part of \bar{A} cliques from which it is not infected, and such a clique contains \bar{B} other vertices
- Whether a vertex gets infected is independent of its infectious period
- $R_* = \mathbb{E}(\bar{A})\mathbb{E}(T(\bar{B}, \mathcal{I}))$

The probability of survival

$$\begin{aligned}\Phi(h)(x) &:= 1 - \mathbb{E}\left(\prod_{i \in \mathcal{S}} (1 - h(\mathcal{I}_i)) \mid \mathcal{I}_0 = x\right) \\ &= 1 - \mathbb{E}_{\bar{B}}\left(\sum_{k=0}^{\infty} e^{-xk} \bar{B}^k e^{-\bar{B}(1-u_k)} G_k(\mathbf{1}|\mathbf{u})\right)\end{aligned}$$

- Fill in $h(x) = \rho(x)$, where $\rho(x)$ is the probability that the offspring of a vertex with infectious period x , infected in the early stages of an epidemic survives
- $\rho_1(x) := \Phi(\rho)(x)$, is the probability of survival of an epidemic started by a vertex with infectious period x , which is part of one clique

- $\rho(x) = 1 - \mathbb{E}(e^{-\bar{A}(\rho_1(x))})$, So we know that ρ_1 satisfies:

$$\rho_1(x) = 1 - \mathbb{E}_{\bar{B}} \left(\sum_{k=0}^{\infty} e^{-xk} \bar{B}^k e^{-\bar{B}(1-u_k)} G_k(\mathbf{1}|\mathbf{u}) \right),$$

where $u_k = \mathbb{E}(e^{-kI} e^{-\bar{A}(\rho_1(I))})$

- If $R_* > 1$, (An ugly prove gives that) $\rho_1(x)$ is the unique positive solution, satisfying this functional equation

- $\rho(x) = 1 - \mathbb{E}(e^{-\bar{A}(\rho_1(x))})$, So we know that ρ_1 satisfies:

$$\rho_1(x) = 1 - \mathbb{E}_{\bar{B}} \left(\sum_{k=0}^{\infty} e^{-xk} \bar{B}^k e^{-\bar{B}(1-u_k)} G_k(\mathbf{1}|\mathbf{u}) \right),$$

where $u_k = \mathbb{E}(e^{-kI} e^{-\bar{A}(\rho_1(I))})$

- If $R_* > 1$, (An ugly prove gives that) $\rho_1(x)$ is the unique positive solution, satisfying this functional equation

The final size of an epidemic

- The final size of an epidemic is the number of ultimately recovered vertices
- The susceptibility set of vertex v is the set of all vertices that if they were the initial infectious, then v would be ultimately recovered (The in-component of the vertex v)
- If there is a giant outbreak (the approximating branching process in the forward direction survives), then the expected fraction of vertices that is ultimately recovered is equal to the probability that a uniformly at random chosen vertex has a giant susceptibility set

The final size of an epidemic

- The final size of an epidemic is the number of ultimately recovered vertices
- The susceptibility set of vertex v is the set of all vertices that if they were the initial infectious, then v would be ultimately recovered (The in-component of the vertex v)
- If there is a giant outbreak (the approximating branching process in the forward direction survives), then the expected fraction of vertices that is ultimately recovered is equal to the probability that a uniformly at random chosen vertex has a giant susceptibility set

The final size of an epidemic

- The final size of an epidemic is the number of ultimately recovered vertices
- The susceptibility set of vertex v is the set of all vertices that if they were the initial infectious, then v would be ultimately recovered (The in-component of the vertex v)
- If there is a giant outbreak (the approximating branching process in the forward direction survives), then the expected fraction of vertices that is ultimately recovered is equal to the probability that a uniformly at random chosen vertex has a giant susceptibility set

- The susceptibility set of a vertex can be explored by a single type branching process. Again all “progeny” within a clique is considered as children of the initial vertex (in the exploration process) in that clique
- Using Gontcharoff polynomials allows to compute the distribution of the number of children per vertex in this backward process

- The susceptibility set of a vertex can be explored by a single type branching process. Again all “progeny” within a clique is considered as children of the initial vertex (in the exploration process) in that clique
- Using Gontcharoff polynomials allows to compute the distribution of the number of children per vertex in this backward process

- Complications might arise, because it is not nice if the initial infective is in the early generations of the susceptibility set of the vertex under consideration or if a large susceptibility set does not overlap with a large epidemic
- To overcome this problem we first ignore all vertices and cliques with weight exceeding $\log[n]$
- Because both the vertex and clique weights have finite expectation, the weight of the ignored vertices is $o(n)$
- Run the forward exploration process generation by generation up to generation $\lceil \log \log[n] \rceil$
- Show that the total weight of vertices and cliques affected by this forward process is $o(n^{1/3})$ (in probability)

- Complications might arise, because it is not nice if the initial infective is in the early generations of the susceptibility set of the vertex under consideration or if a large susceptibility set does not overlap with a large epidemic
- To overcome this problem we first ignore all vertices and cliques with weight exceeding $\log[n]$
- Because both the vertex and clique weights have finite expectation, the weight of the ignored vertices is $o(n)$
- Run the forward exploration process generation by generation up to generation $\lceil \log \log[n] \rceil$
- Show that the total weight of vertices and cliques affected by this forward process is $o(n^{1/3})$ (in probability)

- Complications might arise, because it is not nice if the initial infective is in the early generations of the susceptibility set of the vertex under consideration or if a large susceptibility set does not overlap with a large epidemic
- To overcome this problem we first ignore all vertices and cliques with weight exceeding $\log[n]$
- Because both the vertex and clique weights have finite expectation, the weight of the ignored vertices is $o(n)$
- Run the forward exploration process generation by generation up to generation $\lceil \log \log[n] \rceil$
- Show that the total weight of vertices and cliques affected by this forward process is $o(n^{1/3})$ (in probability)

- Complications might arise, because it is not nice if the initial infective is in the early generations of the susceptibility set of the vertex under consideration or if a large susceptibility set does not overlap with a large epidemic
- To overcome this problem we first ignore all vertices and cliques with weight exceeding $\log[n]$
- Because both the vertex and clique weights have finite expectation, the weight of the ignored vertices is $o(n)$
- Run the forward exploration process generation by generation up to generation $\lceil \log \log[n] \rceil$
- Show that the total weight of vertices and cliques affected by this forward process is $o(n^{1/3})$ (in probability)

- Run the exploration process of the susceptibility set in the graph where all vertices and cliques in the already explored set of the epidemic are temporarily ignored
- Once this exploration process is finished, we check whether the susceptibility set connects to the “still active” vertices of the epidemic exploration process
- If this susceptibility set is small, it will not connect to the epidemic with probability tending to 1

- Run the exploration process of the susceptibility set in the graph where all vertices and cliques in the already explored set of the epidemic are temporarily ignored
- Once this exploration process is finished, we check whether the susceptibility set connects to the “still active” vertices of the epidemic exploration process
- If this susceptibility set is small, it will not connect to the epidemic with probability tending to 1

- Run the exploration process of the susceptibility set in the graph where all vertices and cliques in the already explored set of the epidemic are temporarily ignored
- Once this exploration process is finished, we check whether the susceptibility set connects to the “still active” vertices of the epidemic exploration process
- If this susceptibility set is small, it will not connect to the epidemic with probability tending to 1

- If the susceptibility set is large, it will have total weight of both vertices and cliques $\theta(n)$ with probability tending to 1
- If the epidemic process survives up to generation $\lceil \log \log[n] \rceil$ and the susceptibility set of v has weight $\theta(n)$, then the two processes can be glued together and v will be ultimately recovered
- If we start with 2 uniformly at random chosen vertices and explore their susceptibility sets, we get a bound for the variance of the final size, which is $o(n^2)$. That is the variance of the fraction of ultimately recovered vertices converges to 0

- If the susceptibility set is large, it will have total weight of both vertices and cliques $\theta(n)$ with probability tending to 1
- If the epidemic process survives up to generation $\lceil \log \log[n] \rceil$ and the susceptibility set of v has weight $\theta(n)$, then the two processes can be glued together and v will be ultimately recovered
- If we start with 2 uniformly at random chosen vertices and explore their susceptibility sets, we get a bound for the variance of the final size, which is $o(n^2)$. That is the variance of the fraction of ultimately recovered vertices converges to 0

- If the susceptibility set is large, it will have total weight of both vertices and cliques $\theta(n)$ with probability tending to 1
- If the epidemic process survives up to generation $\lceil \log \log[n] \rceil$ and the susceptibility set of v has weight $\theta(n)$, then the two processes can be glued together and v will be ultimately recovered
- If we start with 2 uniformly at random chosen vertices and explore their susceptibility sets, we get a bound for the variance of the final size, which is $o(n^2)$. That is the variance of the fraction of ultimately recovered vertices converges to 0

- Similar results are expected to hold for configuration model in which the number of cliques a vertex is part of is distributed as A and the clique sizes are distributed as B
- Here the survival probability of the progeny of a vertex of infectious period x and which is part of one clique is given by

$$\rho_1(x) = 1 - \sum_{k=0}^{\infty} e^{-kx} G_k(1|U) f_{B-1}^{(k)}(u_k)$$

where

- $u_k = \mathbb{E}(e^{-kI}(1 - \rho(I)))$
- $\rho(x) = 1 - f_{A-1}(\rho_1(x))$
- $f_X(s)$ is the generating function of X

- Similar results are expected to hold for configuration model in which the number of cliques a vertex is part of is distributed as A and the clique sizes are distributed as B
- Here the survival probability of the progeny of a vertex of infectious period x and which is part of one clique is given by

$$\rho_1(x) = 1 - \sum_{k=0}^{\infty} e^{-kx} G_k(1|U) f_{\bar{B}-1}^{(k)}(u_k)$$

where

- $u_k = \mathbb{E}(e^{-kI}(1 - \rho(I)))$
- $\rho(x) = 1 - f_{\bar{A}-1}(\rho_1(x))$
- $f_X(s)$ is the generating function of X

References

BALL, F.G. AND O'NEILL, P.D. (1999), *Journal of Applied Probability* **36(2)** 473-491.

BOLLOBÁS, B.; JANSON, S. AND RIORDAN, O (2007), *Random Structures and Algorithms* **31** 3-122.

BOLLOBÁS, B.; JANSON, S. AND RIORDAN, O (2010), to appear in *Random Structures and Algorithms*.

BRITTON, T., DEIJFEN, M., LAGERÅS, A. N. AND LINDHOLM, M. (2008), *J. Appl. Probab.* **45** 743–756.

DEIJFEN, M. AND KETS, W. (2009), *Probability in the Engineering and Informational Sciences* **23** 661-674

NEWMAN, M.E.J. (2002), *Physical Review E* **66(1)** 016128.