

Epidemic on Reed-frost Random Intersection Graph with Tunable Degree Distribution and Clustering

¹Dr. T. Vasanthi and ²Miss. S. Subasri

¹*Associate Professor, A.D.M. College for women, Nagapattinam.*

²*Assistant Professor, Thiru. Vi. Ka. Govt. Arts College, Thiruvarur.*

Abstract

A branching processes approximation for the spread of a Reed-Frost epidemic on a network with tunable clustering is derived. The approximation gives rise to expressions for the epidemic threshold and the probability of a large outbreak in the epidemic. A random intersection graph is constructed by assigning independently to each vertex a subset of a given set and drawing an edge between two vertices if and only if their respective subsets intersect. The distribution of the degree of a given vertex is characterized and is shown to depend on the weight of the vertex. Moreover, we provide a deterministic approximation of the bivariate process of susceptible and infective individuals, valid when the number of initially infective individuals is large. These results are used in order to derive the basic reproduction number and the asymptotic final epidemic size of the process. The model is described in the framework of random graphs.

Keywords: Epidemics, random graphs, clustering, degree distribution, branching processes, epidemic threshold, Reed-Frost.

1. Introduction

Let us consider a simple stochastic discrete time model of a so-called SIR epidemic, that is, an epidemic where individuals receive lifelong immunity after having recovered from the disease. Imagine a closed population consisting of n individuals, where each individual has a random number of acquaintances. Let the i th individual have D_i friends, the variables D_i being identically distributed and almost independent, and suppose that two friends rarely have other friends in common. Now introduce an

infectious disease into the population by infecting a individuals. If a susceptible individual meets an infective acquaintance at time t , then she will become infective at time $t+1$, and the infective will have recovered at this time point, now being immune to the disease. For instance, they typically have power law degree sequences, that is, the fraction of vertices with degree k is proportional to $k^{-\tau}$ for some exponent $\tau > 1$. Furthermore, many networks are highly clustered, meaning roughly that there is a large number of triangles and other short cycles. A related explanation is that human populations are typically divided into various subgroups- working places, schools, associations etc which gives rise to high clustering in the social network, since members of a given group typically know each other.

Real-life networks are generally very large, implying that it is a time-consuming task to collect data to delineate their structure in detail. This makes it desirable to develop models that capture essential features of the real networks. A natural candidate to model a network is a random graph, and, to fit with the empirical observations, such a graph should have a heavy-tailed degree distribution and considerable clustering. The model makes it possible to obtain arbitrary prescribed values for the clustering and to control the mean and the tail behavior of the degree distribution.

2. Reed-Frost Model

The Reed-Frost model can easily be adapted to incorporate this type of heterogeneity by introducing a graph to represent the social structure in the population and then stipulating that infective individuals can only infect their neighbors in the social network. This modification makes the analysis of the model two-fold. Firstly, one wants to find a realistic model for the underlying social network, and, secondly, one wants to study the behavior of the epidemic on this graph. If the relation between the number of individuals and the number of groups is chosen appropriately, this leads to a graph where the amount of clustering can be tuned by adjusting the parameters of the model.

3. Random Intersection Graph

Let us consider $\mathcal{V} = \{1, \dots, n\}$ be a set of n vertices and \mathcal{A} a set of m elements. For $p \in [0, 1]$, construct a bipartite graph $B(n, m, p)$ with vertex sets \mathcal{V} and \mathcal{A} by including each one of the nm possible edges between vertices from \mathcal{V} and elements from \mathcal{A} independently with probability p . The random intersection graph $G(n, m, p)$ with vertex set \mathcal{V} is obtained by connecting two distinct vertices $i, j \in \mathcal{V}$ if and only if there is an element $a \in \mathcal{A}$ such that both i and j are adjacent to a in $B(n, m, p)$. When the vertices in \mathcal{V} are thought of as individuals and the elements of \mathcal{A} as social groups, this gives rise to a model for a social network in which two individuals are joined by an edge if they share at least one group. We frequently borrow the terminology from the field of social networks and refer to the vertices as individuals and the elements of

\mathcal{A} as groups, with the understanding that the model is of course much more general. The number of groups $m = \lfloor n^\alpha \rfloor$ for some $\alpha > 0$. The probability that two individuals do not share a group in $B(n, m, p)$ is $(1 - p^2)^m$. It follows that the edge probability in $G(n, m, p)$ is

$$\begin{aligned} & 1 - (1 - p^2)^m \text{ and hence the expected degree is} \\ \mathbf{E}[D_i] &= (n - 1) (1 - (1 - p^2)^m) \\ &= (n - 1) (mp^2 + o(m^2p^4)) \end{aligned} \quad (3.1)$$

The expected degree bounded as $n \rightarrow \infty$, let $p = \gamma n^{-(1+\alpha)/2}$ for some constant $\gamma > 0$. We then have that $\mathbf{E}[D_i] \rightarrow \gamma^2$.

Description of the model

A generalization of the original random intersection graph where the edge probability p is random and depends on weights associated with the vertices. The model is defined as follows:

Let us consider n be a positive integer, and define $m = \lfloor \beta n^\alpha \rfloor$ with $\alpha, \beta > 0$. As before, take $\mathcal{V} = \{1, \dots, n\}$ to be a set of n vertices and \mathcal{A} a set of m elements. Also, let $\{W_i\}$ be an i.i.d. sequence of positive random variables with distribution F , where F is assumed to have mean 1 if the mean is finite. Finally, for some constant $\gamma > 0$, set

$$p_i = \gamma W_i n^{-(1+\alpha)/2} \wedge 1. \quad (3.2)$$

Now construct a bipartite graph $B(n, m, p)$ with vertex sets \mathcal{V} and \mathcal{A} by adding edges to the elements of \mathcal{A} for each vertex $i \in \mathcal{V}$ independently with probability p_i . The random intersection graph $G(n, m, F)$ is obtained as before by drawing an edge between two distinct vertices $i, j \in \mathcal{V}$ if and only if they have a common adjacent vertex $a \in \mathcal{A}$ in $B(n, m, F)$.

Proposition 3.1.1.

Let D_i be the degree of vertex $i \in \mathcal{V}$ in a random intersection graph $G(n, m, F)$ with $m = \lfloor \beta n^\alpha \rfloor$ and p_i as in (3.2). If F has finite mean, then, for all values of $\alpha > 0$, we have that $\mathbf{E}[D_i | W_i] \rightarrow \beta \gamma^2 W_i$ as $n \rightarrow \infty$.

Proof:

To prove the claim for vertex $i = 1$. Let us define,

$$W_j' = W_j \cdot \mathbf{1}_{\{W_j \leq n^{1/4}\}} \text{ and } W_j'' = W_j \cdot \mathbf{1}_{\{W_j > n^{1/4}\}}, \quad (3.3)$$

and let D' and D'' denote the degree of vertex 1 when $\{W_j\}_{j \neq 1}$ are replaced by $\{W_j'\}$ and $\{W_j''\}$ respectively, that is, D' is the number of neighbors of 1 with weight

smaller than or equal to $n^{1/4}$ and D'' is the number of neighbors with weight larger than $n^{1/4}$. Write p'_j and p''_j for the analog of (3.2) based on the truncated weights.

Now, conditional on the weights, the probability that there is an edge between 1 and j is $1 - (1 - p_1 p_j)^m$. To see that $\mathbf{E}[D''] \rightarrow 0$ as $n \rightarrow \infty$, we observe that

$$1 - (1 - p_1 p''_j)^m \leq m p_1 p''_j = \beta \gamma W_1 n^{(\alpha-1)/2} p''_j, \quad (3.4)$$

Summing the expectation of the right-hand side over $j \neq 1$, keeping W_1 fixed, gives

$$\mathbf{E}[D''] \leq \beta \gamma n^{(\alpha-1)/2} \mathbf{E}[p''_k] \quad (3.5)$$

$$\leq \beta \gamma (\gamma \mathbf{E}[W''_k] + n^{(1+\alpha)/2} \mathbf{P}(\gamma W_k \geq n^{(1+\alpha)/2})), \quad (3.6)$$

Where both terms on the right hand side converge to 0 as $n \rightarrow \infty$ since F has finite mean. As for D' , we have

$$1 - (1 - p_1 p'_j)^m = \beta \gamma^2 W_1 W'_j n^{-1} + O(W_1^2 (W'_j)^2 n^{-2}), \quad (3.7)$$

The sum over $j \neq 1$ of the expectation of the first term equals $\beta \gamma^2 W_1 \mathbf{E}[W'_k]$, where $\mathbf{E}[W'_k] \rightarrow \mathbf{E}[W_k] = 1$ and the sum of the expectation of the second term converges to 0. Since $D_0 = D' + D''$.

4. The Epidemic Model and an Approximating Branching Process

Let us consider a closed homogeneous population consisting of n individuals, labeled v_1, v_2, \dots, v_n , with a social structure represented by a random intersection graph $\mathcal{G}^{(n)}$. The Reed-Frost dynamics to assumed to be fixed an infection in this population. The social graph $\mathcal{G}^{(n)}$ is assumed to be fixed throughout the spread of the infection. Furthermore, for simplicity, let us start with one single randomly selected infective individual at time 0, the rest of the population being susceptible. An individual that is infective at time t ($t=0,1,\dots$) contacts each one of its neighbors in $\mathcal{G}^{(n)}$ independently with some probability p , and if a contacted neighbor is susceptible, it becomes infective at time $t+1$. The individuals that were infective at time t are removed from the epidemic process at time $t+1$ and take no further part in the spread of the infection.

Lemma 4.1.1

Let $k > 0$ be such that $1/k > 2 \log(\beta \gamma^2)$. As $n \rightarrow \infty$, the probability that the subgraph of $\mathbf{B}^{(n)}$ induced by $C^{(n)}(\lfloor k \log n \rfloor)$, is a tree, tends to 1.

Proof

Let $C^{(n)}(t)$ be a sequence $\{\mathcal{D}^{(n)}(t) : t \geq 0\}$, constructed in such a way that $C^{(n)}(t) = \bigcup_{0 \leq s \leq t} \mathcal{D}^{(n)}(s)$. For odd t , the set $\mathcal{D}^{(n)}(t)$ will consist of groups and for even t by

individuals. By definition, let us have $\mathcal{C}^{(n)}(0) = \{\mathcal{V}_1\}$, so necessarily $\mathcal{D}^{(n)}(0) = \mathcal{C}^{(n)}(0)$. For odd t , the set $\mathcal{D}^{(n)}(t)$ is then constructed by choosing, independently for each individual in $\mathcal{D}^{(n)}(t-1)$, a Binomial $(m, \gamma/n)$ distributed number of distinct groups in \mathbf{A} , and, likewise, for even t , let us construct $\mathcal{D}^{(n)}(t)$ by choosing, independently for each group in $\mathcal{D}^{(n)}(t-1)$, a Binomial $(n, \gamma/n)$ distributed number of distinct individuals in \mathcal{V} . Let $X^{(n)}$ be a compound binomial random variable with generating function

$$g(s) = \mathbf{E}[s^{X^{(n)}}] = (1 - \frac{\gamma}{n} + \frac{\gamma}{n} (1 - \frac{\gamma}{n} + \frac{\gamma}{n} s)^n)^m, \quad (4.1)$$

And let $\{X^{(n)}(t) : t \geq 0\}$ be a branching process with offspring distribution $X^{(n)}$. Furthermore, write $Y^{(n)}(t) = \sum_{s=0}^t X^{(n)}(s)$ for the total progeny of the branching process at time t . Then, for even t , the number of individuals that have been chosen in the construction of the process $\mathcal{C}^{(n)}(t)$, has the same distribution as $Y^{(n)}(t/2)$, and the number of groups that have been chosen is strictly smaller than $Y^{(n)}(t/2)$.

Let us consider $\mu_n := \mathbf{E}[X^{(n)}] = \gamma^2 m/n$ and note that $\beta\gamma^2(1 - \frac{1}{n}) \leq \mu_n \leq \beta\gamma^2$, so $\mu_n \rightarrow \mu := \beta\gamma^2$ as $n \rightarrow \infty$. In branching process theory, we have that $\mu_n^{-t} X^{(n)}(t) \rightarrow W^{(n)}$ almost surely as $t \rightarrow \infty$, where $W^{(n)}$ is a random variable with $W^{(n)} \equiv 0$ if and only if $\mu_n \leq 1$. Furthermore, $W^{(n)} \rightarrow W$ in distribution as $n \rightarrow \infty$, where W is the corresponding limiting random variable for the branching process $\{X(t) : t \geq 0\}$ with offspring generating function $\mathbf{E}[s^{X(1)}] = \exp\{\beta\gamma(e^{\gamma(s-1)} - 1)\}$. Thus,

$$X^{(n)}(t) = \mu_n^t (W^{(n)} + o_t(1)) \quad (4.2)$$

$$= \mu_n^t (W + o_n(1) + o_t(1)) \quad (4.3)$$

$$\leq \mu^t (O_n(1) + O_t(1)), \quad (4.4)$$

where $o_x(\cdot)$ and $O_x(\cdot)$ is the usual order notation when $x \rightarrow \infty$.

It follows that $Y^{(n)}(t) \leq \mu^t (O_n(1) + O_t(t))$, and, $t = \lfloor k \log n \rfloor$ with $\frac{1}{k} > \max\{2 \log \mu, 0\}$, (4.5)

We get,

$$Y^{(n)}(\lfloor k \log n \rfloor) \leq e^{\log \mu \lfloor k \log n \rfloor} O_n(\log n) \quad (4.6)$$

$$= o_n(\sqrt{n}). \quad (4.7)$$

Now note that, if all individuals and groups that have been chosen in the construction of $\mathcal{C}^{(n)}(t)$ are distinct, then clearly the subgraph of $\mathbf{B}^{(n)}$ induced by $\mathcal{C}^{(n)}(t)$ is a tree. Thus the probability in the statement of the lemma is greater than

$$\prod_{k=1}^{Y^{(n)}(\lfloor k \log n \rfloor)} \left(1 - \frac{k}{n}\right) \left(1 - \frac{k}{m}\right) = \exp \left\{ \sum_{k=1}^{o(\sqrt{n})} \left(\log\left(1 - \frac{k}{n}\right) + \log\left(1 - \frac{k}{m}\right) \right) \right\} \quad (4.8)$$

$$= \exp \left\{ - \left(1 + \frac{1}{\beta}\right) \sum_{k=1}^{o(\sqrt{n})} \left(\frac{k}{n} + O\left(\frac{k^2}{n^2}\right) \right) \right\} \quad (4.9)$$

$$= \exp\{o(1)\} \rightarrow 1.$$

5. Final Outcome of the Epidemic

The social network is a random intersection graph with $\alpha = 1$. There are no rigorous results concerning the component structure in a random intersection graph with $\alpha = 1$. The size of the largest component in a random graph are derived by heuristic means and it is observed that the relative final size of the giant component seems to decrease as the clustering in the graph increases.

Let us consider an arbitrary graph with n vertices and $k = O(n)$ edges and assume that the clustering equals 1. This implies that all subgraphs are complete. Hence, with n_{max} denoting the size of the largest subgraph, we have that the number of edges in the maximal subgraph is $\binom{n_{max}}{2}$. It follows that $n_{max} \leq O(\sqrt{k}) = O(\sqrt{n})$, that is, the relative size of the largest component tends to zero.

6. Clustering Coefficient

The adjacency relations between actors in real networks are not statistically independent events. Often, chances of a link $v' \sim v''$ increase as we learn that actors v' and v'' have a common neighbor, say, v . As a theoretical measure of such a statistical dependence, one can use the conditional probability

$$\alpha = \mathbf{P}(v' \sim v'' \mid v' \sim v, v'' \sim v), \quad (6.1)$$

The empirical estimates of the conditional probability α ,

$$\hat{\alpha} = n^{-1} \sum_{v \in V} \frac{N_3(v)}{N_2(v)} \text{ and } \hat{\alpha} = \frac{\sum_{v \in V} N_3(v)}{\sum_{v \in V} N_2(v)} \quad (6.2)$$

are called the clustering coefficient and the global clustering coefficient, respectively. Here n denotes the number of vertices of a graph, $N_3(v)$ is the number of unlabeled triangles having vertex v , $N_2(v)$ is the number of unlabelled 2-stars with the central vertex v . The term clustering coefficient is used exclusively for the conditional probability α .

In the random graph $G_s(n, m, P)$, the conditional probability α does not depend on the choice of v, v', v'' . It does not depend on n either. We write $\alpha = \alpha_s(m, p)$ in order to indicate the dependence on s, m and P . The uniform random intersection graph

$G_s(n, m, \delta_x)$ where, for large m , the asymptotics of the clustering coefficient is simple and transparent.

6.1 Clustering coefficient and degree

Let us consider a sequence of sparse random intersection graphs $\{G_s(n, m, P)\}_m$ with nonvanishing clustering coefficient and nondegenerate asymptotic vertex degree distribution. The clustering coefficient

$$\alpha^* = \alpha^*(n, m, P) = \mathbf{P}(w_2 \sim w_3 \mid w_1 \sim w_2, w_1 \sim w_3), \quad (6.3)$$

of a passive random intersection graph $G_1^*(n, m, P)$. In particular, that

$$\alpha^*(n, m, P) = \frac{\beta_*^2 m^{-1} (\mathbf{E}(X_1)_2)^3 + \mathbf{E}(X_1)_3}{\beta_* (\mathbf{E}(X_1)^2 + \mathbf{E}(X_1)_3)} + O(1), \quad (6.4)$$

$\beta_* := nm^{-1}$, provided that $\mathbf{E}(X_1)_2 > 0$ and $\mathbf{E}(X_1)_2 = o(m^2 n^{-1})$ as $m, n \rightarrow \infty$.

7. Clustering for a Power Law Weight Distribution

When $\alpha = 1$, the clustering is given by

$$c(G) = \mathbf{E}[(1 + \beta\gamma W_k)^{-1}] \quad (7.1)$$

Here we investigate this expression in more detail for the important case that F is a power law. More precisely, we take F to be a Pareto distribution with density

$$f(x) = \frac{(\lambda-2)^{\lambda-1}}{(\lambda-1)^{\lambda-2}} x^{-\lambda} \text{ for } x \geq \frac{\lambda-2}{\lambda-1} \quad (7.2)$$

When $\lambda > 2$, this distribution has mean 1, as desired. The asymptotic clustering $c(G)$ is given by the integral

$$\frac{(\lambda-2)^{\lambda-1}}{(\lambda-1)^{\lambda-2}} \int_{\frac{\lambda-2}{\lambda-1}}^{\infty} (1 + \beta\gamma x)^{-1} x^{-\lambda} dx \quad (7.3)$$

Defining $u := (\lambda - 2)/(x(\lambda - 1))$, we obtain

$$c(G) = \frac{1}{\beta\gamma} \frac{(\lambda-1)^2}{(\lambda-1)} \int_0^1 u^{\lambda-1} \left(1 + \frac{u}{\beta\gamma} \left(\frac{\lambda-1}{\lambda-2}\right)\right)^{-1} du, \quad (7.4)$$

$$=: \frac{1}{\beta\gamma\lambda} \frac{(\lambda-1)^2}{(\lambda-2)} {}_2F_1\left(1, \lambda; 1 + \lambda; -\frac{1}{\beta\gamma} \left(\frac{\lambda-1}{\lambda-2}\right)\right), \quad (7.5)$$

where ${}_2F_1$ is the hypergeometric function. For $\beta\gamma \geq (\lambda - 1)/(\lambda - 2)$, a series expansion of the integrand yields that

$$c(G) = \frac{1}{\beta\gamma} \frac{(\lambda-1)^2}{(\lambda-2)} \sum_{k=0}^{\infty} \left(-\frac{1}{\beta\gamma} \left(\frac{\lambda-1}{\lambda-2} \right) \right)^k \frac{1}{k+\lambda}, \quad (7.6)$$

$$=: \frac{1}{\beta\gamma} \frac{(\lambda-1)^2}{(\lambda-2)} \Phi \left(-\frac{1}{\beta\gamma} \left(\frac{\lambda-1}{\lambda-2} \right), 1, \lambda \right), \quad (7.7)$$

where Φ is the Lerch transcendent. Furthermore, when λ is an integer, we get

$$c(G) = \frac{(\lambda-2)^{\lambda-1}}{(\lambda-1)^{\lambda-2}} \left[(-\beta\gamma)^{\lambda-1} \ln \left(1 + \frac{\lambda-1}{\beta\gamma(\lambda-2)} \right) + \sum_{l=1}^{\lambda-1} \frac{(-\beta\gamma)^{\lambda-1-l}}{l} \left(\frac{\lambda-1}{\lambda-2} \right)^l \right]. \quad (7.8)$$

The clustering depends on λ and $\beta\gamma$ respectively. For any $c \in (0,1)$ and a given tail exponent λ , we can find a value of $\beta\gamma$ such that the clustering is equal to c .

Many real networks are “small worlds”, meaning roughly that the distances between vertices remain small also in very large networks. It is interesting to study the relation between the distances between vertices, the degree distribution and the clustering in the current model.

References

- [1] Britton, T., Deijfen, M., Lageras, A. And Lindholm, M. (2008). Epidemics on random graph with tunable Clustering, *Journal of Applied Probability* 45, 743-756.
- [2] Stark, D. (2004). The vertex degree distribution of random intersection graphs, *Random Structures & Algorithms* 24, 249-258.
- [3] Ball, F., Mollison, D. And Scalia-Tomba, G. (1997). Epidemics with two levels of mixing, *Annals of Applied Probability* 7, 46-89.
- [4] Deijfen, M. And Kets, W. (2007). Random intersection graphs with tunable degree distribution and clustering, preprint, available at www.math.su.se/~mia.
- [5] Neal, P. (2006). Multityperandomized Reed-Frost epidemics and epidemics upon graphs, *Annals of Applied Probability* 16, 1166-1189.
- [6] Andersson, H. (1998). Limit theorems for a random graph epidemic model. *Annals of Applied Probability* 8, 1331-1349.
- [7] Mindaugas Bloznelis¹ (2013). Degree and clustering coefficient in sparse random intersection graphs, *The Annals of Applied Probability* 23, 1254-1289.
- [8] Mindaugas Bloznelis And Julius Damarackas (2013) Degree distribution of an inhomogeneous random Intersection graph.
- [9] Jaworski, J., Karoriski, M. And Stark, D. (2006). The degree of a typical vertex in generalized random intersection graph models, *Discrete Mathematics* 306, 2152-2165.
- [10] Newman, M. (2003). Spread of epidemic disease on networks, *Phys. Rev. E* 66 016128, MR1919737..