Epidemic spreading on complex networks

Ljupco Kocarev

Macedonian Academy of Sciences and Arts
University of California San Diego
University Ss Cyril and Methodius, Skopje, Macedonia
IKT-labs, Skopje, Macedonia
(lkocarev@ucsd.edu)

CIDNET14, Dresden, Germany, June 2014

Epidemiology

- Epidemiology – patterns, causes, and effects of **health** and **disease** conditions in a population
    - literally meaning "the study of what is upon the people"
    - epi "upon, among"
    - demos "people, district"
    - logos "study, word, discourse"

- The distinction between "epidemic" and "endemic" was first drawn by Hippocrates
    – Diseases that are "visited upon" a population (epidemic)
    – Diseases that are "reside within" a population (endemic)

Epidemic models

- Mathematical models: study how infectious diseases progress and show the likely outcome of an epidemic in order to help inform public health interventions.
- The earliest account of mathematical modeling of spread of disease was carried out in 1766 by Daniel Bernoulli.
- A. G. McKendrick and W. O. Kermack: A Contribution to the Mathematical Theory of Epidemics (1927)
- Reed-Frost epidemic model (1928) – one of the simplest stochastic epidemic models

Several approaches to study processes on networks:

- Mathematics (stochastic, deterministic, dynamical systems approach)
- Physics (statistical physics, the theory of phase transitions and critical phenomena)
- Computer science (optimal solutions, computational complexity theory)

The problem of modeling how diseases spread among individuals has been intensively studied for many years.

SIS model

- Population of *N* individuals, connected in a network structure represented by a graph $G = (V, E)$ with node set *V* and edge set *E*
- Each node can be in one of two possible states: susceptible (S) and infective (I)
- $\mathbf{s}_i(t) = [s_i^S(t) \ s_i^I(t)]^T$ – status vector, an indicator vector containing a single 1 in the position corresponding to the present state, and 0 everywhere else
- $\mathbf{p}_i(t) = [p_i^S(t) \ p_i^I(t)]^T$ – probability mass-function (PMF) of node *i* at time *t*: $p_i^S(t) + p_i^I(t) = 1$

The evolution of SIS is described by the following equations:

$$
\begin{aligned}
p_i^I(t+1) &= s_i^S(t)f_i(t) + (1-\delta)s_i^I(t), \\
\mathbf{s}_i(t+1) &= MultiRealize[\mathbf{p}_i(t+1)].
\end{aligned}
$$

- *MultiRealize*[·] – performs a random realization for the PMF given with $\mathbf{p}_i(t+1)$
- The first term on the right hand side is the probability that a susceptible node i is infected $f_i(t)$ by at least a neighbor
- The second term stands for the probability that infected node i at time t does not recover

SIS model

$$f_i(t) = 1 - \prod_{j=1}^{N} \left[1 - \beta a_{ij} s_j^l(t) \right].$$

- In the SIS model individuals can exist in either of two possible states: "healthy" or "infected".
- Healthy individuals are infected when they come into contact with an infected individual with probability β.
- Infected individuals become once again susceptible with a recovery probability δ.

Deterministic model ($p_i \equiv p_i^I$):

$$
\begin{aligned}
p_i(t+1) &= [1 - p_i(t)] f_i(t) + (1 - \delta) p_i(t) \\
f_i(t) &= 1 - \prod_{j=1}^{N} \left[1 - \beta a_{ij} p_j(t) \right].
\end{aligned}
$$

- $p_i(t)$ is the expected probability that node i will be infected at time t
- $f_i(t)$ is the probability that node i receives the infection from at least one of its infected neighbors at time t

$$p_i(t+1) = [1 - p_i(t)] f_i(t) + (1 - \delta)p_i(t)$$
$$f_i(t) = 1 - \prod_{j=1}^{N} [1 - \beta a_{ij} p_j(t)].$$

- The origin $p_i = 0$ ($\forall i$) is a fixed point of the system
- The origin is stable when $1 - \delta + \beta\lambda_{1,A} < 1$, where $\lambda_{1,A}$ is the largest eigenvalue of the adjacency matrix
-
$$\frac{\beta}{\delta}\lambda_{1,A} > 1$$

– the disease will reach an endemic state

- What is the influence of the graph topology on virus (disease) spreading, more precisely, on the probability that given note is infective?
- How local properties constrain the interval of possible values of the probability that given note is infective?

p-hop subgraph

- Let i be arbitrary node of the graph $G = (V, E)$, $i \in V$, and let $n_i = \max_x l(i, x)$. Let $V_i^0 = \{i\}$.
- We define a subgraph $G_i^p = (V_i^p, E_i^p)$ of $G = (V, E)$ as follows:

$$
\begin{aligned}
V_i^p &= \{x | x \in V, 0 \le l(i, x) \le p\} \\
E_i^p &= \{xy | xy \in E, x \in V_i^p, y \in V_i^{p-1}\},
\end{aligned}
$$

where $p = 1, \ldots, n_i + 1$.
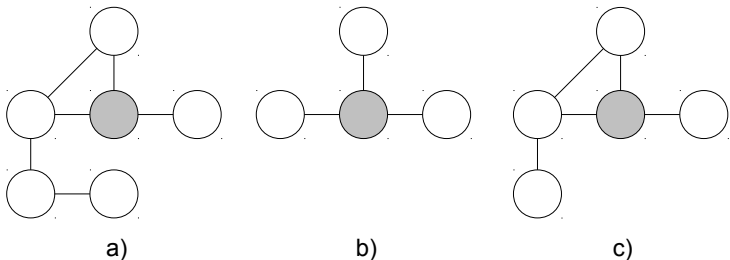- We say that G_i^p is a *p*-hop subgraph of G extracted by starting at node i.

Figure: b) and c) 1-hop and 2-hop subgraph of the graph shown in a) extracted by starting at the gray node.

-

$$l_i^1 < l_i^2 < \ldots < l_i^p \ldots \leq x_i^* < \ldots u_i^p < \ldots < u_1^2 < u_i^1$$

- The bounds l_i^1 and u_i^1 are obtained by considering only (first) neighbors of i.
- The bound u_i^1 depends on the degree of the node i, that is, the information contained in the 1-hop subgraph of G extracted by starting at node i, while for the bound l_i^1 one computes the SIS model on the subgraph G_i^1, which is the subgraph of neighbors of i.

- The bounds l_i^2 and u_i^2 are obtained by considering second neighbors of i (neighbors of the first neighbors).
- l_i^2 and u_i^2 reflect the topology of 2-hop subgraph of G extracted by starting at node i.
-
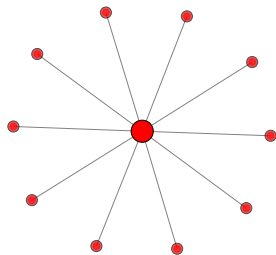$$d_i^p = u_i^p - l_i^p \qquad \Delta\rho_p = \sum_{i=1}^{N} \frac{d_i^p}{N},$$

Figure: 1-hop neighborhood for a node extracted from a real-world e-mail network with 33696 nodes. The node (largest in size) has 10 direct neighbors (medium sized). The probability of infection for the given node obtained after simulating a particular configuration of the SIS model was 0.373. The probability of infection for the node given the 1-hop neighborhood (node's degree) is calculated to be between 0.006 and 0.503.
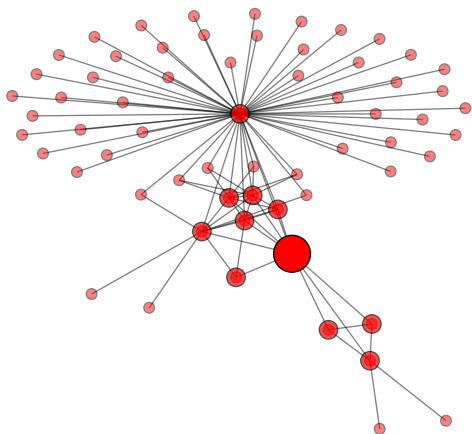
Figure: The 2-hop neighborhood contains 62 nodes and 92 edges.
Peripheral nodes are smallest in size and are two hops away from the
central node. The probability of infection for the node given the 2-hop
neighborhood topology is calculated to be between 0.297 and 0.416.

Theorem

Let Φ be the family of all possible simple and connected graphs $G = (V, E)$ with $|V| \geq 2$. Let $\mathbf{x}(G)^* = [x_1^* x_2^* \ldots x_N^*]$ be the stationary solution different from the origin and let i be arbitrary node of the graph $G = (V, E)$, $i \in V$. Let

$$u_i^n = \frac{1 - \prod_{j=1}^N (1 - \beta a_{ij} u_j^{n-1})}{1 - \prod_{j=1}^N (1 - \beta a_{ij} u_j^{n-1}) + \delta}$$

where $u_i^0 = 1/(1 + \delta)$. Then for all i, x_i^* is bounded by

$$x_i^* < \ldots < u_i^n < \ldots < u_i^1 < u_i^0$$

Theorem

Consider arbitrary node i of the graph $G = (V, E)$ and let $G_i^p = (V_i^p, E_i^p)$ be the p-hop subgraph of G extracted by starting at node i. Write $n = |V_i^p|$. Let $\mathbf{x}(G)^* = [x_1^* x_2^* \ldots x_N^*]$ and $\mathbf{x}(G_i^p)^* = [l_1^p l_2^p \ldots l_n^p]$ be the stationary solution different from the origin for the graphs $G = (V, E)$ and $G_i^p = (V_i^p, E_i^p)$, respectively. Then x_i^* is bounded by

$$l_i^1 < l_i^2 < \ldots < l_i^{n_i} < l_i^{n_i+1} = x_i^*$$

for all $i \in V$.

Average size of *p*-hop neighborhood

Table: Average size of *p*-hop neighborhood for the Enron e-mail network. $|E^p|$ is an average of $|E_i^p|$ over all nodes i and $|E|$ is the total number of edges in the network.

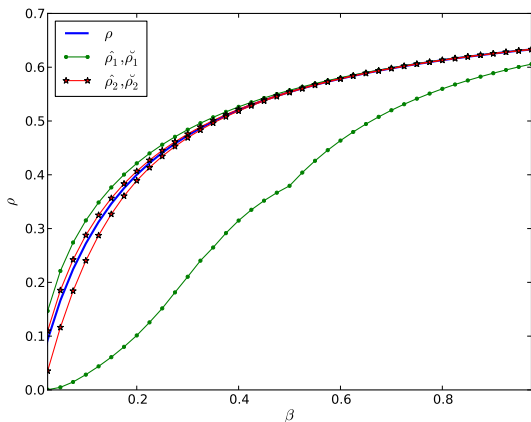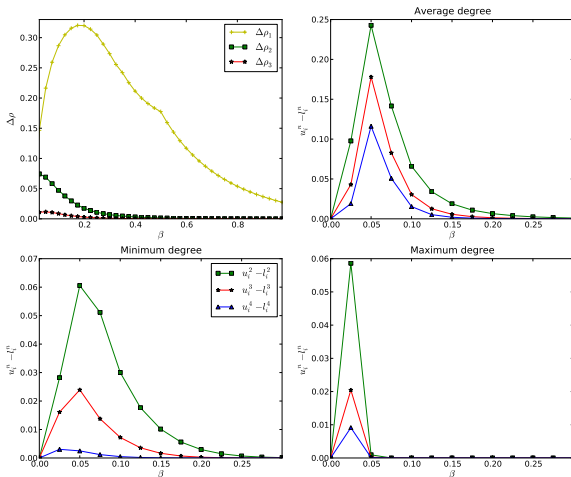| p	$|E^p|$	$|E^p| / |E|$
1	10	0.0003
2	1538	0.004
3	45067	0.125
4	207496	0.574

Figure: The density of infective nodes in the Enron e-mail network as the transmission parameter β is varied, and $\delta = 0.5$, along with the upper and lower bounds, $\hat{\rho}_p$ and $\check{\rho}_p$, on ρ using 1-hop and 2-hop topology information.

$$
\begin{aligned}
p_i(t+1) &= [1 - p_i(t)] \, f_i(t) + (1 - \delta_i) p_i(t) \\
f_i(t) &= 1 - \prod_{j=1}^{N} \left[1 - \beta_i a_{ij} p_j(t) \right].
\end{aligned}
$$

- What is the role of heterogeneity in the susceptibility of individuals?

- R_0 – the basic reproduction number: average number of other individuals each infected individual will infect in a population that has no immunity to the disease
- R_0 is the quintessential epidemiological parameter; a central question in epidemiology is under what conditions R_0 becomes greater than 1, ($R_0 > 1$).

- Homogenous network (each individual is in contact with k others):

$$R_0^h = \frac{k\beta}{\delta}$$

- Degree heterogenous networks with no degree-degree correlations and a degree distribution given by $P(k)$ [Pastor-Satorras and Vespignani, PRL 2001]:

$$R_0^{unc} = \frac{\beta}{\delta}\frac{\langle k^2 \rangle}{\langle k \rangle} = R_0^h \left(1 + \left[\frac{\sigma_k}{\langle k \rangle} \right]^2 \right)$$

- Correlated degree heterogenous networks [Boguna and Pastor-Satorras, PRE 2002]:

$$R_0^{corr} = \frac{\beta}{\delta}\lambda_{1,C}$$

$\lambda_{1,C}$ – the largest eigenvalue of the degree mixing matrix $C_{kk'} = kP(k'|k)$.

- A significant improvement over the heterogenous mean-field theory; it has been shown that the basic reproductive number takes the form:

$$R_0 = \frac{\beta}{\delta}\lambda_{1,A}$$

  $\lambda_{1,A}$ – the largest eigenvalue of the adjacency matrix A.
- Wang, Chakrabarti, Wang, and Faloutsos, SRDS 2003

- SIS model:

$$p_i(t+1) = (1 - p_i(t))f_i(t) + (1 - \delta)p_i(t)$$

- The probability $f_i(t)$ has the form:

$$f_i(t) = 1 - \prod_{j=1}^{N}(1 - \beta_i a_{ij} p_j(t)).$$

- What is the role of heterogeneity in the susceptibility of individuals, that is, under what conditions $R_0 > 1$

Differential susceptibility

- We linearize the system around $\mathbf{p} = 0$:

$$p_i(t+1) = (1-\delta)p_i(t) + \sum_{j=1}^{N} \beta_i a_{ij} p_j(t) = \sum_{j=1}^{N} m_{ij} p_j(t).$$

- $M = [m_{ij}]$
  $m_{ij} = \beta_i a_{ij} + \Delta_{ij}(1-\delta)$, Δ_{ij} – Kronecker delta
- When $\lambda_{1,M} < 1$, the infection will die out exponentially fast with a rate determined by $\lambda_{1,M}$.

- $M = R + (1 - \delta)I$,
  $R = [r_{ij}]$, $r_{ij} = \beta_i a_{ij}$, I – identity matrix
- We can define the threshold at which epidemics begin to spread through the largest eigenvalue of R and generalize R_0 to:

$$R_0 = \lambda_{1,R} + 1 - \delta.$$

Correlated networks

- Constant β ($\beta_i = \beta$):

$$\frac{\beta}{\delta} \lambda_{1,A} > 1$$

- Differential susceptibility:

$$\frac{1}{\delta} \lambda_{1,R} > 1$$

- Assume that r_{ij} is a random variable following an arbitrary distribution and use a mean-field approximation $r_{ij} \approx \langle r_{ij} \rangle$ where $\langle r_{ij} \rangle$ is the expected value of r_{ij} over all possible network realizations.

- In a network where the susceptibility β_i is assigned independently of the topology: $\langle r_{ij} | a_{ij} \rangle = \langle \beta | a_{ij} \rangle a_{ij} = \langle \beta \rangle a_{ij}$ where $\langle x | y \rangle$ is the expected value of x given y.

-

$$R_0^{ind} = \langle \beta \rangle \lambda_{1,A} + 1 - \delta \qquad \frac{\langle \beta \rangle}{\delta} \lambda_{1,A} > 1$$

- For uncorrelated networks with heterogenous degrees, $\langle r_{ij} \rangle$ can be relaxed to the expected number of links between nodes i and j

-

$$R_0^{unc} = \frac{\langle \beta k^2 \rangle}{\langle k \rangle} + 1 - \delta = \frac{\langle \beta \rangle \langle k^2 \rangle}{\langle k \rangle} + \frac{\rho \sigma_\beta \sigma_{k^2}}{\langle k \rangle} + 1 - \delta.$$

$-1 \leq \rho \leq 1$ – Pearson correlation coefficient

- Small correlations between the susceptibility and the degree can lead to significant over- or underestimation of R_0 when the variation in connectivity, as measured by σ_{k^2}, is large compared to the average connectivity $\langle k \rangle$.

Uncorrelated networks

- Constant β ($\beta_i = \beta$):

$$\frac{\beta}{\delta} \frac{\langle k^2 \rangle}{\langle k \rangle} > 1$$

- Differential susceptibility:

$$\frac{\langle \beta \rangle}{\delta} \frac{\langle k^2 \rangle}{\langle k \rangle} + \frac{\rho \sigma_\beta \sigma_{k^2}}{\delta \langle k \rangle} > 1$$

- Link level correlations – tendency of individuals to connect to other individuals with similar characteristics leads to non-trivial mixing patterns
- We approximate r_{ij} with $\langle r_{ij} \rangle$, by using the expected number of links from node with degree k_i and susceptibility β_i to node with degree k_j and susceptibility β_j.
- The expected number of links are proportional to the two-point conditional probability $P(k', \beta'|k, \beta)$:

$$r_{ij} \approx \langle r_{ij} \rangle = \frac{k_i \beta_i P(k_j, \beta_j | k_i, \beta_i)}{N P(k_j, \beta_j)}.$$

- Let d=$|\{k, \beta\}|$ be the number of different combinations of degree (k) and susceptibility (β) that a node can have in the network.
- The entire information about the system (network topology and the susceptibilities of nodes) is compressed into a coarsened $d \times d$ matrix D:

$$R_0 = \lambda_{1,R} + 1 - \delta \approx \lambda_{1,D} + 1 - \delta.$$

- Example: each individual can be assiged to one of 5 degree classes, $k = 1, 2, 4, 8, 16$ and to one of 3 susceptibility classes, $\beta = \beta_{low}, \beta_{avg}, \beta_{high}$ corresponding to low, average and high susceptibility respectively.

- Assume independence between the degree of a node and its susceptibility, i.e. $P(k', \beta'|k, \beta) = P(k'|k)P(\beta'|\beta)$
- Let $\lambda_{1,C}$ and $\lambda_{1,B}$ be the largest eigenvalues of the matrices $C_{kk'} = kP(k'|k)$ and $B_{\beta\beta'} = \beta P(\beta'|\beta)$ respectively
- $\lambda_{1,R} = \lambda_{1,C}\lambda_{1,B}$
- Note that when there is no degree mixing, $\lambda_{1,C} = \langle k^2 \rangle / \langle k \rangle$. On the other hand, when there are no mixing patterns in susceptibility, $\lambda_{1,B} = \langle \beta \rangle$.

Segregation

- The contact networks that underlie the spread of a disease are not only heterogeneous in terms of degree and susceptibility, but also segregated.
- Strong effects of segregation dynamics on R_0 are illustrated by running a variant of Schelling's segregation process on a real-world contact network obtained from face-to-face proximity between students and teachers.
- Mild level of segregation can drastically increase the critical reproductive number in a network where individuals differ in their susceptibility

$$\langle k \rangle \left[\frac{1-s}{N_\beta} + s \right] \beta_{max} + 1 - \delta \leq R_0 \leq \langle k \rangle \left[(1-s) \langle \beta \rangle + s\beta_{max} \right] + 1 - \delta$$

- $s \in [0, 1]$ models the segregation in the network
- $s = 0$ – segregation is minimal (the probability that a node will connect to others is independent of the susceptibility)
- $s = 1$ – segregation is maximal (nodes share links only with others that have the same susceptibility)

Summary

- One can estimate the probability of being infective using only local information (considering only $n-$hop local topology, for small n), without knowing the whole network.
- From this local information one can also estimate the density of being infective on the whole network, as well as assess the extend to which the topology affects the outcome of the infection on macroscopic level.
- The results are extendable to other ergodic models (such as SIRS, for example) and are related to all types of spreading (idea, failure, rumor).

- Basic reproductive numbers estimated from compartmental models will be systematically over or underestimated if the mathematical models of epidemic spreading do not include the effects of heterogenous susceptibility.
- Mild level of segregation can drastically increase the critical reproductive number in a network where individuals differ in their susceptibility

References

- D. Trpevski, W. K. S. Tang, and L. Kocarev, Model for rumor spreading over networks, Physical Review E 81, 056102 (11 pages) 2010
- D. Smilkov and L. Kocarev, Influence of the network topology on epidemic spreading, Physical Review E 85, 016114 (10 pages), 2012
- I. Tomovski and L. Kocarev, Simple Algorithm for Virus Spreading Control on Complex Networks, IEEE Transactions on Circuits and Systems I: Regular Papers, Volume: 59, Issue: 4, Page(s): 763 – 771, 2012

References

- I. Tomovski, I. Trpevski, L. Kocarev, "Topology independent SIS process: An engineering viewpoint", Communications in Nonlinear Science and Numerical Simulation, 19, 627 – 637, 2014

- D. Smilkov, C. A. Hidalgo, and L. Kocarev, "Beyond network structure: How heterogeneous susceptibility modulates the spread of epidemics", Scientific Reports 4, Article number: 4795, 2014

- A. Stanoev, D. Trpevski, and L. Kocarev, "Modeling the Spread of Multiple Concurrent Contagions on Networks", PLOS One, DOI: 10.1371/journal.pone.0095669, Published: June 12, 2014.