Networks: October 19

36-720
Scribe: Jacqueline Mauro

November 2, 2016

Contents

1 From last time

Take a graph and its adjacency matrix as fixed. We are interested in the dynamics of some random field $X(t)$ with value $X_i(t)$ at node i. Contagion nodes can move from being susceptible (S) to infections when they meet an I. Focus on early stage of spread (I doesn't go back to R or S). Full mixing:

$$\frac{dI}{dt} = rI(N-I) \tag{1}$$

Solution is given by:

$$I(t) = I_0 \frac{e^{rt}}{N - I_0 + I_0 e^{rt}} \tag{2}$$

Solution is always a logistic.

2 Including Network Structure

Look carefully at how the contagion spreads and track the number of new infections due to each infected node.

Notation:

- $I(t) = Z(0) + Z(1) + \cdots + Z(t)$

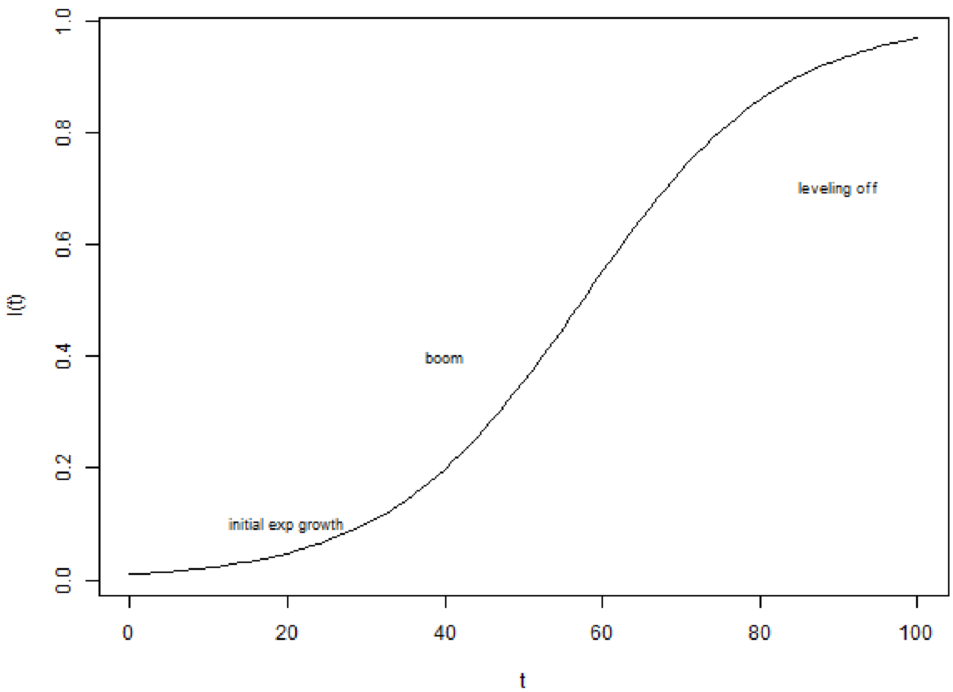

Figure 1: Exponential growth

- $Z(t)$ = new cases at time t
- $Z(t+1) = \sum_{i=1}^{Z(t)} \zeta_{i,t}$
- $\zeta_{i,t}$ = number infected by node i at time t

We neglect any overlap. This gives us a branching process.

Query: does $I(t) \to \infty$ (or at least large N)? Or does $I(t) \to$ constant $<< N$.

2.1 Basic reproductive number

Basic reproductive number: "R_0" = average new infections caused by inserting one more I into the population (in one step).

Folklore: $R_0 < 1$: contagion is limited. $R_0 > 1$: contagion spreads over an $O(n)$ part of the population, "epidemic".

Correct for random graphs with any degree distribution.

$$
\begin{aligned}
R_0 &= \mathbb{E}[\text{new nodes infected after making one node } I] \\
&= \mathbb{P}(\text{transmit to neighbor})\mathbb{E}[\text{available contacts}] \\
\mathbb{P}(\text{transmit to neighbor}) &= \tau
\end{aligned}
$$

Note $\mathbb{E}[\text{available contacts}] \neq$ average degree.

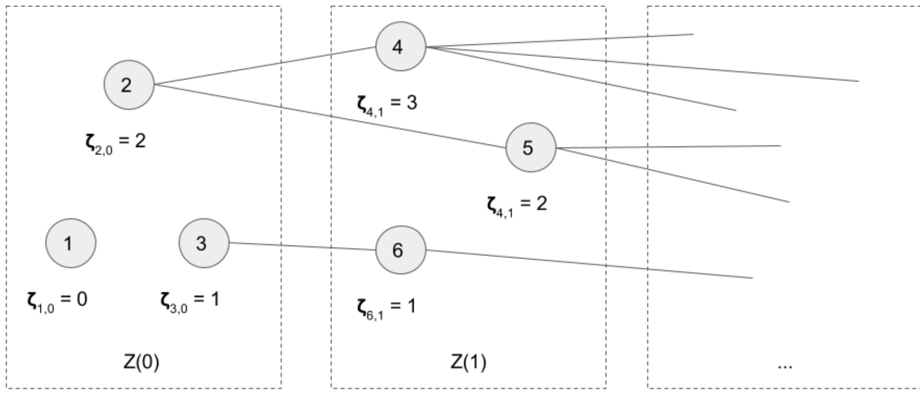

Figure 2: Branching Process

2.2 The branching process

Suppose WOLOG $I(0) = Z(0) = 1$. Then

$$\begin{aligned} \zeta_{1,0} &= Z(1) \\ &= Binom(\text{num neighbors of initial indiv}, \tau) \\ \Rightarrow \mathbb{E}\left[Z(1)\right] &= \mathbb{E}\left[Binom(\text{numn eighbors}, \tau)\right] \\ &= \tau\mathbb{E}(\text{degree}) \end{aligned}$$

2.3 The paradox of friendship

The 2nd generation nodes are not chosen uniformly at random. We can reach them by a link – so we are biased towards nodes with high degree. The degree distribution has pmf p_k, meaning the probability that a node of degree k is a neighbor to the initial node $\propto k$. Degree distribution of nodes we reach by following an edge is $\propto kp_k$.

$\Rightarrow$ degree distribution of reachable nodes $q_k \propto kp_k$. To normalize:

$$q_k = \frac{kp_k}{\sum_{k=0}^{\infty} kp_k} = \frac{kp_k}{\mathbb{E}(k)} \tag{3}$$

$\mathbb{E}(k)$ is the expected degree of a random node.

What about $Z(2)$? $Z(2) = \sum_{i=1}^{Z(1)} \zeta_{i,2}$. Each $\zeta_{i,2} = Binom(\text{degree of node } i - 1, \tau)$. Thus:

$$\mathbb{E}(\zeta_{i,2}) = \tau\mathbb{E}(\text{degree of a node reached from a random node } - 1) \tag{4}$$

Call the RHS J:

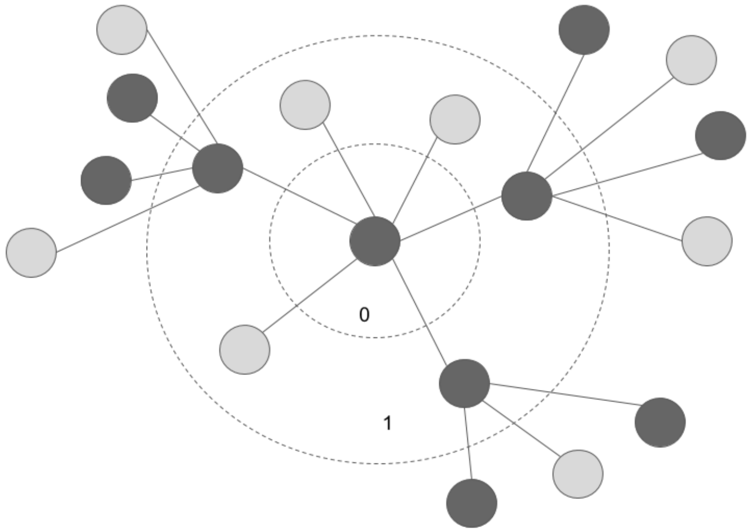

Figure 3: Infection Progression

$$
\begin{aligned}
\mathbb{E}(\zeta_{i,2}) &= \sum_{j=0}^{\infty} j q_j \\
&= \sum_{j=0}^{\infty} \frac{j j p_j}{\mathbb{E}(k)} \\
&= \sum_{j=0}^{\infty} \frac{j^2 p_j}{\mathbb{E}(k)} \\
&= \sum_{k=0}^{\infty} \frac{k^2 p_k}{\mathbb{E}(k)} \\
&= \frac{\mathbb{E}(k^2)}{\mathbb{E}(k)} \\
&= \frac{\mathbb{V}(k) + (\mathbb{E}(k))^2}{\mathbb{E}(k)} \\
&= \mathbb{E}(k) + \frac{\mathbb{V}(k)}{\mathbb{E}(k)} \\
&\geq \mathbb{E}(k)
\end{aligned}
$$

This is the "Paradox of Friendship": your friends have more friends than you do.

So this gives:

$$
\mathbb{E}(\zeta_{i,2}) = \tau \left[\mathbb{E}(k) + \frac{\mathbb{V}(k)}{\mathbb{E}(k)} - 1 \right] \tag{5}
$$

Note in a Poisson graph, $\mathbb{V}(k) = \mathbb{E}(k)$ so $\mathbb{E}(\zeta_{i,2}) = \tau \mathbb{E}(k)$.

2.4 The epidemic threshold

Same story after 2nd generation. In generation $t \geq 2$:

$$\mathbb{E}(\zeta_{i,2}) = \tau \left[\mathbb{E}(k) + \frac{\mathbb{V}(k)}{\mathbb{E}(k)} - 1 \right] \tag{6}$$

because 3rd generation etc nodes are also reached via edges and have degree pmf q_k:

$$\tau \left[\mathbb{E}(k) + \frac{\mathbb{V}(k)}{\mathbb{E}(k)} - 1 \right] = \tau \left(\mathbb{E}(J) - 1 \right) \tag{7}$$

In a branching process, if $\mathbb{E}(\zeta_{i,t}) < 1$, extinction happens with probability 1 (subcritical). If $\mathbb{E}(\zeta_{i,t}) > 1$, positive probability of living forever (supercritical):

$$\mathbb{E}\left[I(t) \right] = \left(\mathbb{E}(\zeta_{i,t}) \right)^t \tag{8}$$

This $\mathbb{E}(\zeta_{i,t})$ is the "basic reproductive number" (after 1st generation).

Moral: In a random graph with arbitrary degree distribution, contagions become epidemics when $\tau \left[\mathbb{E}(k) + \frac{\mathbb{V}(k)}{\mathbb{E}(k)} - 1 \right] > 1$, or equivalently $\tau > \left[\mathbb{E}(k) + \frac{\mathbb{V}(k)}{\mathbb{E}(k)} - 1 \right]^{-1}$.

The critical level of transmissibility $\rightarrow 0$ as $\mathbb{E}(k) \rightarrow \infty$ or $\mathbb{V}(k) \rightarrow \infty$ with $\mathbb{E}(k)$ fixed.

2.5 "The web of human sexual contacts"

Observation: if degree distribution is a power law $\propto k^{-\alpha}$, then $\mathbb{E}(k^2) = \sum_{k=0}^{\infty} k^2 C_\alpha k^{-\alpha} = \infty$ if $\alpha \leq 3$. Liljeros *et al.* (2001): the data says that for human sexual contacts, $\hat{\alpha} = 3.2 \pm 0.3$. $\Rightarrow \tau_c = 0$ unless we "destroy the hubs".

Some caveats:

- not a good estimation of degree – not a power law (Handcock and Jones, 2004)
- data a self-reporting study of Swedish adults

2.6 Epidemics as giant components in thinned graphs

Flip a coin and thin the graph if "tails" with probability $1 - \tau$. Condition for epidemic = new graph has expected degree greater than 1 $\Rightarrow$ new graph has a giant connected component.

3 Do fat friends make you fat?

Christakis and Fowler (2007): Obesity is contagious.

- $X_i(t) = 0$ if not obese at time t
- $X_i(t) = 1$ if become obese at time t

Logistic regression of $X_i(t)$ on $X_j(t-1)$ controlling for all sorts of i-covariates C_i and $X_i(t-1)$ where $A_{ij} = 1$.

Finding: having an obese friend increases risk of becoming obese $\approxeq 7\times$.

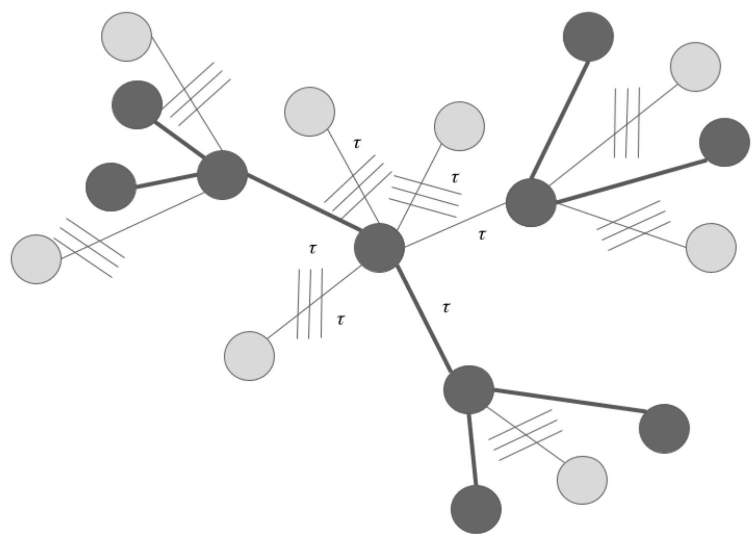

Figure 4: Thinning the graph

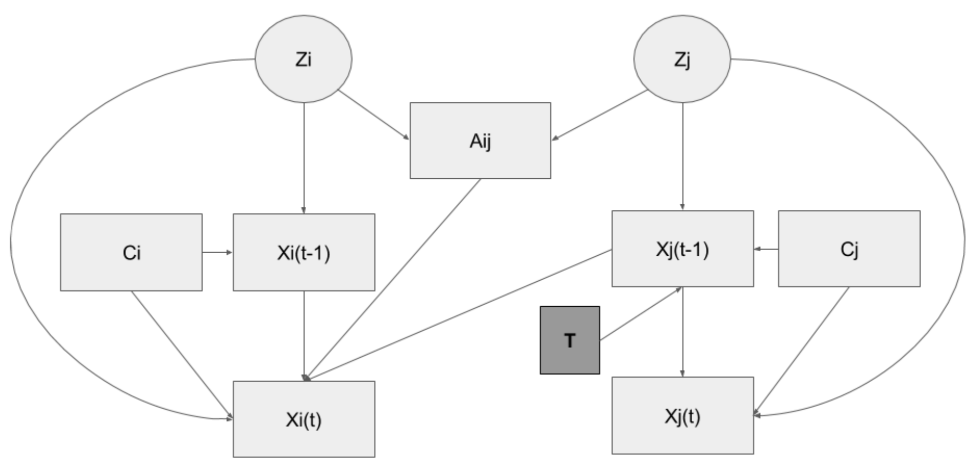

Figure 5: Causation

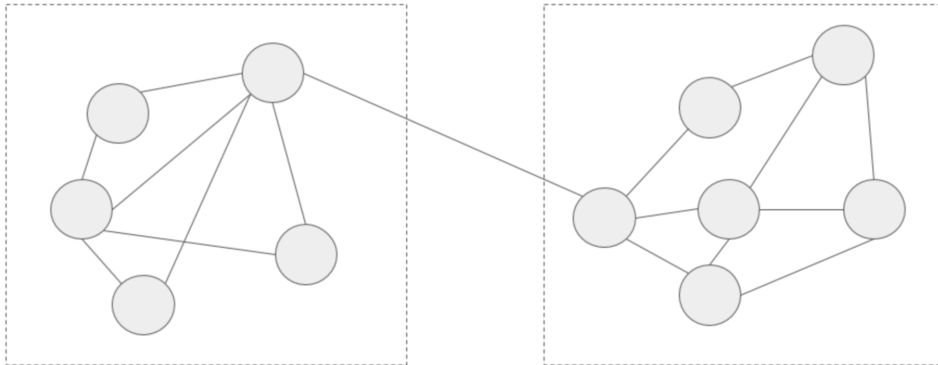

Figure 6: Community Discovery

If we only select neighbors ($A_{ij} = 1$) we get confounding of contagion with homophily. If we don't select neighbors, we don't get any evidence about social effects. (Shalizi and Thomas, 2011)

"Social networks are machines for creating selection bias" – Unknown

Ways out:

- Actually measure what matters (control for Z_i)
- Randomize the network (colleges, prisons, maybe Facebook)
- Inject random variation into $X_j(t-1)$ ie treat some nodes (Facebook, instrumental variables)
- Measure Z_i indirectly from the whole network: Given $A_{ij} = 1$, Z_j is extra likely to be close to Z_j (homophily). So the network should have clusters of tighly linked nodes with similar Z_i's. Use community discovery to get $\hat{Z}_i$ amd control for it. This will work under some assumptions (Shalizi and McFowland, 2016).

References

Christakis, Nicholas A. and James H. Fowler (2007). "The Spread of Obesity in a Large Social Network over 32 Years." *The New England Journal of Medicine*, **357**: 370–379. URL http://content.nejm.org/cgi/content/abstract/357/4/370.

Handcock, Mark S. and James Holland Jones (2004). "Likelihood-based inference for stochastic models of sexual network formation." *Theoretical Population Biology*, **65**: 413–422. URL http://www.csss.washington.edu/Papers/wp29.pdf.

Liljeros, F. C. R., L. A. N. Amaral, H. E. Stanley and Y. Aberg (2001). "The Web of Human Sexual Contacts." *Nature*, **411**: 907–908. URL http://arxiv.org/abs/cond-mat/0106507. doi:10.1038/35082140.

Shalizi, Cosma Rohilla and Edward McFowland, III (2016). "Controlling for Latent Homophily in Social Networks through Inferring Latent Locations." arxiv:1607.06565. URL http://arxiv.org/abs/1607.06565.

Shalizi, Cosma Rohilla and Andrew C. Thomas (2011). "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies." *Sociological Methods and Research*, **40**: 211–239. URL http://arxiv.org/abs/1004.4704. doi:10.1177/0049124111404820.