Dynamic processes on networks II: homophily vs. influence

Shengming Luo

19 October 2016

Contents

1 Contagion, cont'd

Recall:

Take graph and its adjacency matrix as fixed. We're interested in the dynamic of some random filed $X(t)$ with value $X_i(t)$ at node i.

Contagion: Nodes can move from being susceptible (S) to infectious (I) when they meet an I.

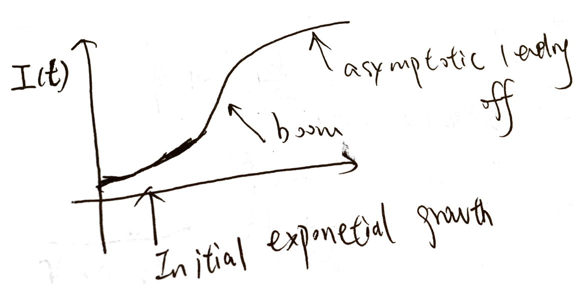
We focus on early stage at spread so no time for I to be removed (R) or go back to S.

Under full mixing,

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

and solution always a logistic curve.

Including network structure:



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

$$I(t) = Z(0) + Z(1) + Z(2) + \cdots + Z(t),$$

so, $Z(t) = \{\text{new cases at time } t\}$.

$Z(t+1) = \sum_{i=1}^{z(t)} \xi_{i,t}$, where $\xi_{i,t} = \#\{\text{infected by the } i\text{-th person to get it at time } t\}$.

Neglecting overlapping: we have a branching process.

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

2 Epidemic threshold

2.1 Basic reproductive remember

Basic reproductive number: "R_0"= average number of new infections caused by inserting one more I into the population directly.

Folklore:

- If $R_0 < 1 \Rightarrow$ contagion is limited
- If $R_0 > 1 \Rightarrow$ contagion spreads over an $O(n)$ part of the population, "epidemic".

This is correct for random networks with any degree distribution.

$$\begin{aligned} R_0 &= \mathbb{E}[\#\text{new noes infected after making one node an ``I''}] \\ &= \mathbb{P}(\text{transmitting disease to neighbours}) \cdot \mathbb{E}(\text{available contacts}). \end{aligned}$$

2.2 The branching process

It's worthwhile to notice $\mathbb{E}(\text{available contacts}) \neq$ average degree. We would show this later.

Suppose W.O.L.O.G. that

$$I(0) = Z(0) = 1$$

and

$$\xi_{1,0} = Z(1) = Binom\{\#\text{of neighbours of initial individual}, \tau\}.$$

Then,

$$\mathbb{E}(Z(1)) = \mathbb{E}(Binom\{\#\text{of neighbours}, \tau\}) = \tau \mathbb{E}(degree - 1).$$

2.3 Degree of nodes reached by the epidemic; the paradox of friendship

But, the second generations nodes are not random nodes. In fact, they are nodes we can reach by a link, which yields biases towards nodes with high degree.

Suppose degree distribution has p.m.f function p_k. The probability that a node of degree k is neighbour to initial is in proportional to k.

The degree distribution of nodes we reached by following an edge thus is in proportional to kp_k. To be specific,

$$q_k = \frac{kp_k}{\sum_k kp_k} = \frac{kp_k}{\mathbb{E}(K)},$$

where K denotes the degree for a randomly chosen node.

So, what about $Z(2)$?

$$Z(2) = \sum_{i=1}^{Z(1)} \xi_{i,2},$$

where each $\xi_{i,2} = -1 + Binom(\text{degree of node } i,\ \tau)$.

$$\mathbb{E}(\xi_{i,2}) = -1 + \tau\mathbb{E}(\text{degree of a node reached from a random node}).$$

We call the rightmost term as J.

Therefore,

$$\begin{aligned}
\mathbb{E}(J) &= \sum_{j=0}^{\infty} j q_j \\
&= \sum_{j=0}^{\infty} j^2 q_j / \mathbb{E}K \\
&= \frac{\mathbb{E}K^2}{\mathbb{E}K} \\
&= \frac{Var(K) + (\mathbb{E}K)^2}{\mathbb{E}K} \\
&= \mathbb{E}K + \frac{Var(K)}{\mathbb{E}K},
\end{aligned}$$

where the rightmost term is positive.

In other words,

$$\mathbb{E}(\xi_{i,2}) = \tau(\mathbb{E}K + \frac{Var(K)}{\mathbb{E}K} - 1). \tag{2.1}$$

Now we have the paradox of friendship: Your friends have more friends than you.

In a Poisson graph, $Var(K) = \mathbb{E}K$, so just

$$\mathbb{E}(\xi_{i,2}) = \tau\mathbb{E}K. \tag{2.2}$$

So it's the same thing after second generation in Poisson graph. But in other graphs, it's different.

2.4 Back to the branching process

In generation, when $t \geq 2$,

$$\mathbb{E}(\xi_{i,t}) = \tau(\mathbb{E}K + \frac{Var(K)}{\mathbb{E}K} - 1), \tag{2.3}$$

because third etc. generation nodes are also reached via edges and have degree p.m.f q_k.

In a branching process, if $\mathbb{E}(\xi_{i,t}) < 1$, then extinction with probability 1 (subcritical). If $\mathbb{E}(\xi_{i,t}) > 1$, then with some positive probability of $\to \infty$, $\mathbb{E}(I_t)$ is $\mathbb{E}(\xi_{i,t})^t$.(super-critical, which means positive probability of living.)

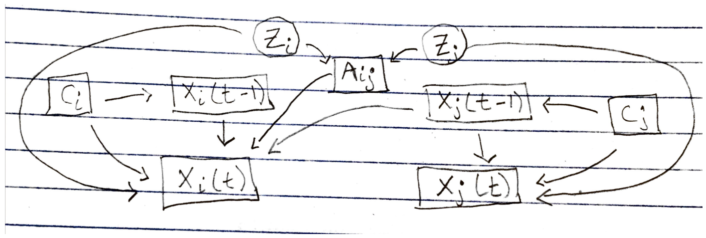


Figure 1: Graph of how the variables in their system are related.

2.5 Epidemic threshold on random networks

Moral:

In a random graph with arbitrary degree distribution, contagion become epidemics when

$$\tau(\mathbb{E}K + \frac{Var(K)}{\mathbb{E}K} - 1) > 1,$$

Or

$$\tau > \frac{1}{(\mathbb{E}K + \frac{Var(K)}{\mathbb{E}K} - 1)},$$

where τ is critical level of transmissibility $\to 0$ as $\mathbb{E}K \to \infty$ or $Var(\mathbb{E}K) \to \infty$ with $\mathbb{E}K$ fixed.

3 "The web of human sexual contacts".

Observation: if degree distribution is power law in proportional to $k^{-\alpha}$, then

$$\mathbb{E}(K^2) = \sum_{k=0}^{+\infty} k^2 C_\alpha k^{-\alpha} = +\infty (\text{ if } \alpha \leq 3).$$

Liljeros *et al.* (2001): Data says that for the "web of human sexual contacts": $\hat{\alpha} = 3.2 \pm 0.3 \Rightarrow \tau = 0$, unless we "destroy the hubs".

Counter-point: (1) Data are from one survey, of one country, of self-reported sexual contacts. (2) As pointed out by Handcock and Jones (2004), there is no way even that data is really a power law.

4 Do fat friends make you fat?

Christakis and Fowler (2007)): obseity is contagious.

Define $X_i(t) = 0$ if no obesity at time t and $X_i(t) = 1$ if obese at time t.

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

Finding:

Having an obese friend increases risk for becoming obese ≈ 7 folds.

Crucial problem: Irena and Joey are not assigned an edge randomly, since it's observational data.

If we select only neighbours, we get confounding of contagion and homophily. If we don't select the neighbours, we do not get andy evidence about social effects. (See above figure.) Shalizi and Thomas (2011) has details.

Hence, "social networks are machines for creating selection bias".

Ways out:

- Actually measure what matters, so control for Z_i.
- Randomize the network, i.e. assign people at random to sections like colleges, to roommates, prisons, etc.
- Inject random variation into $X_j(t-1)$(facebook, social engineer, "instrumental" variables.)
- Measure Z_i indirectly from the whole network.

Given $A_{ij} = 1$, $Z(i)$ is extra likely to be close to $Z(j)$(homophily), so the network should have clusters of tightly-linked nodes with similar $Z_i's \Rightarrow$ Use community discovery to get $\hat{Z}_i$ and control for it (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.