ECS 253 / MAE 253, Lecture 13 May 15, 2018



"Diffusion, Cascades and Influence" Mathematical models & generating functions

Last week: spatial flows and game theory on networks

- Optimal location of facilities to maximize access for all.
- Designing "optimal" spatial networks
 (collection/distribution networks subways, power lines, road networks, airline networks).
- Details of flows on actual networks make all the difference!
 - Users act according to Nash
 - Braess paradox

(removing edges may improve a network's performance!)

- The "Price of Anarchy"

(cost of worst Nash eqm / cost of system optimal)

This week: Diffusion and cascades in networks (Nodes in one of two states)

- Viruses (human and computer)
 - contact processes
 - epidemic thresholds
- Adoption of new technologies
 - Winner take all
 - Benefit of first to market
 - Benefit of second to market
- Political or social beliefs and societal norms

A long history of study, now trying to add impact of underlying network structure.

Simple diffusion

Diffusion of a substance ϕ on a network with adjacency matrix A.

- Let ϕ_i denote the concentration at node *i*.
- Diffusion: $\frac{d\phi_i}{dt} = C \sum_j A_{ij}(\phi_j \phi_i)$
- In steady-state, $\frac{d\phi_i}{dt} = 0 \implies \phi_j \phi_i$.
- In steady-state all nodes have the same value of ϕ .

• In opinion dynamics this is called **consensus**.

Simple diffusion: The graph Laplacian

•
$$\frac{d\phi_i}{dt} = C \sum_j A_{ij} (\phi_j - \phi_i)$$
$$= C \sum_j A_{ij} \phi_j - C \phi_i \sum_j A_{ij}$$
$$= C \sum_j A_{ij} \phi_j - C \phi_i k_i$$
$$= C \sum_j (A_{ij} - \delta_{ij} k_i) \phi_j.$$

(Note Kronecker delta: $\delta_{ij} = 1$ if i = j and $\delta_{ij} = 0$ if $i \neq j$)

• In matrix form: $\frac{d\phi}{dt} = C(\mathbf{A} - \mathbf{D})\phi = C\mathbf{L}\phi$

- From last page, matrix form: $\frac{d\phi}{dt} = C(\mathbf{A} \mathbf{D})\phi = C\mathbf{L}\phi$
- Graph Laplacian: $\mathbf{L} = \mathbf{A} \mathbf{D}$

where matrix **D** has zero entries except for diagonal with is degree of node:

 $D_{ij} = k_i$ if i = j and 0 otherwise.

The graph Laplacian

- *L* has real positive eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$.
- Number of eigenvalues equal to 0 is the number of distinct, disconnected components of a graph

(Compare this to the column-normalized state transition matrix from earlier in class (i.e., random-walk), where the number of λ 's equal to 1 is the number of components).

• If $\lambda_2 \neq 0$ the graph is fully connected. The bigger the value of λ_2 the more connected (less modular) the graph.

But people are not diffusing particles Opinion dynamics on networks

What drives social change?

Accelerating pace of social change

Speed of Change

Number of years from an issue's trigger point to federal action (all abortion years shown)



Bloomberg, April 26, 2015.

Collective phenomena in social networks How the online world is changing the game

J. Flack, R.D., editors, PIEEE (2014)



Past: Small, geographically localized social networks, concentrated power and influence

> Present: Digital footprint, massive online experimentation, global information, rapid rate of change.

"Re-computing the social sciences" Next step connecting these models with our digital footprints.

Mathematical models of social behavior

Analyze extent of epidemic spreading, product adoption, etc:

- Thresholds models
- Voter models
- Opinion dynamics (e.g. The Naming game)
- Percolation
- Game theory

INSIDE SCIENCE NEWS SERVICE

Zealots Help Sway Popular Opinions



Image credit: Gabriel Saldana via Flickr | http://bit.ly/1E9IjCQ Rights information: http://bit.ly/1dWcOPS

Enthusiasts can greatly influence the adoption of new ideas.

Originally published: Feb 19 2015 - 10:45am

By: Ker Than, Contributor

A. Waagen, G. Verma, K. Chan, A. Swami, R. D. PRE, 2015.

What mechanism makes an individual change their mind?

Collective phenomena: Phase transitions





Cusp bifurcation/catastrophe

• $\frac{dx}{dt} = -x^3 + x + a.$

- Percolation
- Contact processes
- Epidemic spreading

- Abrupt shift as slow-time parameter varies
 - e.g., Vinyl records vs digital music

Phase transitions depend on the underlying details

• The network structure

- Degree distribution (variation in connectivity)
- Modular structure

• The model of human behavior

- Simple contact process / percolation / epidemic spreading
 - * Thresholds (critical mass) versus diminishing returns
 - * Influential versus susceptible individuals
- Voter models
- Opinion dynamics / consensus
 - * The role of zealots
- Strategic interactions / Nash equilibrium (decentralized solutions)

Simplest model of human behavior: Binary opinion dynamics

Each individual can be in one of two states $\{-1, +1\}$

- "Infected" or "healthy" (relevant to both human and computer networks)
- Holding opinion "A" or "B"
- Adopting new product, or sticking with status quo
- Many other choices....

But what causes opinion to change?

I. Diminishing returns versus thresholds





k = number of friends adopting

Critical mass?

Watts, Dodds e.g. *PNAS* 2002.

"Hill climbing" / best response Percolation & generating functions Algorithms for influential seed nodes Susceptibles vs influentials/mavens (Depends on active vs passive influence.)

II: The Voter model, "Tell me what to think"

V. Sood, S. Redner, Phys. Rev. Lett. 94, 2005.

- At each time step in the process, pick a node at random.
- That node picks a random neighbor, and adopts the opinion of the neighbor.
- Ultimately, only one opinion prevails. The high degree nodes (hubs) win.



• Invasion percolation(the "bully" model) yields the opposite: leaf nodes propagate opinions.

III: "The Naming Game" / open minded individuals

Steels, Art. Life 1995; Barrat et al., Chaos 2007; Baronchelli et al., Int. J. Mod. Phys. 2008.

- Originally introduced for linguistic convergence. Two opinions, A and B.
- And each individual can hold A, B, or $\{A, B\}$.
- Exchange opinions with neighbors and update





The impact of Zealots

Committed individuals who will never change opinions

$$x' = z\left(x + z + \frac{p}{2}\right) - y(x - p),$$
$$y' = z\left(y + z + \frac{q}{2}\right) - x(y - q),$$
$$z = 1 - x - y.$$

- p is fraction of A zealots; q is fraction of B zealots.
- Voter model: A finite *number* of zealots can sway the outcome.
- Naming game: A small *fraction* of zealots can sway the outcome.
- Naming game with multiple choices, k:

Operating systems; cell phones; political parties; etc

- Zealots of only one kind: Quickly obey the zealot.
- Equal fractions of zealots of all kinds: Quickly reach stalemate.

More formal analysis

Diffusion, Cascade behaviors, and influential nodes Part I: Ensemble models

Generating functions / Master equations / giant components

- Contact processes / similar to biological epidemic spreading
- Heterogeneity due to node degree (not due to different node preferences)
- Epidemic spreading
- Opinion dynamics
- Social networks: Watts PNAS 2001 (threshold model; no global cascade region)

Diffusion, Cascade behaviors, and influential nodes Part II: Contact processes with individual node preferences

- Long history of empirical / qualitative study in the social sciences (Peyton Young, Granovetter, Martin Nowak ...; diffusion of innovation; societal norms)
- Recent theorems: "network coordination games" (bigger payout if connected nodes in the same state) (Kleinberg, Kempe, Tardos, Dodds, Watts, Domingos)
- Finding the influential set of nodes, or the *k* most influential. Often NP-hard and not amenable to approximation algorithms
- Key distinction:
 - thresholds of activation (leads to unpredictable behaviors)
 - diminishing returns (submodular functions nicer)

Diffusion, Cascade behaviors, and influential nodes Part III: Markov chains and mixing times

- New game-theoretic approaches (general coordination games)
 - Results in an Ising model.
 - Montanari and Saberi PNAS 2010.
- Studies using techniques from Parts I and II suggest:
 - Innovations spread quickly in highly connected networks.
 - Long-range links benefit spreading.
 - High-degree nodes quite influential (enhance spreading).
- Studies using techniques from Part III suggest:
 - Innovations spread quickly in locally connected networks.
 - Local spatial coordination enhances spreading (having a spatial metric; graph embeddable in small dimension).
 - High-degree nodes slow down spreading.

Part I. Ensemble approaches

- A. Master equations (Random graph evolution, cluster aggregation)
- B. Network configuration model
- C. Generating functions
 - Degree distribution (fraction of nodes with degree k, for all k)



- Degree sequence (A realization, N specific values drawn from P_k)

A. Network Configuration Model Degree sequence given



- Bollobas 1980; Molloy and Reed 1995, 1998.
- Build a random network with a specified degree sequence.
- Assign each node a degree at the beginning.
- Random stub-matching until all half-edges are partnered. (Make sure total # edges even, of course.)
- Self-loops and multiple edges possible, but less likely as network size increases.

HW 4b – build a configuration model and analyze percolation and spreading.

B. Generating functions:

Properties of the ensemble of configuration model RGs

Determining properties of the ensemble of all graphs with a given degree distribution, P_k .

- The basic generating function: $G_0(x) = \sum_k P_k x^k$ Note, evaluate at x = 1: $G_0(1) = \sum_k P_k = 1$.
- The moments of *P_k* can be obtained from derivatives of *G*₀(*x*):
 First derivative:

$$G'_0(x) \equiv \frac{d}{dx}G_0(x) = \sum_k k P_k x^{(k-1)}$$

Evaluate at x = 1, $G'_0(1) \equiv \frac{d}{dx}G_0(x)|_{x=1} = \sum_k kP_k$ (the mean)

Calculating moments

- **Base:** $G_0(1) = \sum_k P_k = 1$ (it is the sum of probabilities).
- First moment, $\langle k \rangle \equiv \sum_k k P_k = G'_0(1)$

(And note $xG'_0(x) = \sum_k kP_k x^k$)

• Second moment, $\left< k^2 \right> \equiv \sum_k k^2 P_k$

$$\frac{d}{dx}(xG_0'(x)) = \sum_k k^2 P_k x^{(k-1)}$$

So
$$\frac{d}{dx}(xG'_0(x))\Big|_{x=1} = \sum_k k^2 P_k$$

(And note $x \frac{d}{dx}(xG'_0(x)) = \sum_k k^2 P_k x^k$)

• The n-th moment

$$\langle k^n \rangle \equiv \sum_k k^n P_k = \left(x \frac{d}{dx} \right)^n G_0(x) \Big|_{x=1}$$

Generating functions for the giant component of a random graph

Newman, Watts, Strogatz PRE 64 (2001)

With the basic generating function in place, can build on it to calculate properties of more interesting distributions.

- 1. G.F. for connectivity of a node at edge of randomly chosen edge. Which enables calculating:
- 2. G.F. for size of the component to which that node belongs. Which enables calculating:
- 3. G.F. for size of the component to which an arbitrary node belongs.

Following a random edge

• k times more likely to follow edge to a node of degree k than a node of degree 1. Probability random edge is attached to node of degree k: of degree k: $m_k = k P_k / \sum_{i=k} k P_k / (k)$

$$m_k = kP_k / \sum_k kP_k = kP_k / \langle k \rangle$$

- There are k 1 other edges outgoing from this node. (Called the "excess degree")
- Each of those leads to a node of degree k' with probability m'_k .



(Circles denote isolated nodes, squares components of unknown size.)

What is the GF for the excess degree?

(Build up more complex from simpler)

- Let q_k denote the probability of following an edge to a node with excess degree of k: $q_k = \left[(k+1)P_{k+1} \right] / \langle k \rangle$
- The associated GF

$$G_1(x) = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} (k+1) P_{k+1} x^k$$

$$=\frac{1}{\langle k\rangle}\sum_{k=1}^{\infty}kP_kx^{k-1}$$

$$=\frac{1}{\langle k\rangle}G_0'(x)$$

• Recall the most basic GF: $G_0(x) = \sum_k P_k x^k$

$H_1(x)$, Generating function for probability of size of component reached by following random edge

(Note: subscript 0 on GF denotes node property, 1 denotes edge property)



 $H_1(x) = xq_0 + xq_1H_1(x) + xq_2[H_1(x)]^2 + xq_3[H_1(x)]^3 \cdots$

(A self-consistency equation. We assume a tree network.)

Note also that $H_1(x) = x \sum_k q_k [H_1(x)]^k = x G_1(H_1(x))$

Aside 1: Self-consistency equations Graphical solution

• See HW 1b: Self-consistency for ER giant component

$$S = 1 - e^{-\langle k \rangle S}$$

• Solve for $S(\langle k \rangle)$ (see Fig a) and plot result in Fig b.



(Barabasi book)

Figure 3.18 Graphical Solution

Aside 2: Powers property

The PGF for the sum of m instances of random variable k is the PGF for k to the m'th power.

PGF for $\sum_{m} k$ is $[H_1(x)]^m$

• Easiest to see if m = 2 (sum over two realizations)

•
$$[G_0(x)]^2 = \left[\sum_k P_k x^k\right]^2$$

= $\sum_{jk} p_j p_k x^{j+k}$
= $p_0 p_0 x^0 + (p_0 p_1 + p_1 p_0) x + (p_0 p_2 + p_1 p_1 + p_2 p_0) x^2 + \cdots$

• The coefficient multiplying power n is the sum of all products $p_i p_j$ such that i + j = n.

$H_0(x)$, Generating function for distribution in component sizes starting from arbitrary node



 $H_0(x) = xP_0 + xP_1H_1(x) + xP_2[H_1(x)]^2 + xP_3[H_1(x)]^3 \cdots$

$$= x \sum_{k} P_{k} [H_{1}(x)]^{k} = x G_{0}(H_{1}(x))$$

• Can take derivatives of $H_0(x)$ to find moments of component size distribution!

Note we have assumed a tree-like topology.

Expected size of a component starting from arbitrary node

•
$$\langle s \rangle = \frac{d}{dx} H_0(x) \Big|_{x=1} = \frac{d}{dx} x G_0(H_1(x)) \Big|_{x=1}$$

= $G_0(H_1(1)) + \frac{d}{dx} G_0(H_1(1)) \cdot \frac{d}{dx} H_1(1)$
Since $H_1(1) = 1$, (i.e., it is the sum of the probabilities)
 $\langle s \rangle = 1 + G'_0(1) \cdot H'_1(1)$

(Recall $\langle k \rangle = G'_0(1))$

• Recall (three slides ago) $H_1(x) = xG_1(H_1(x))$

so $H'_1(1) = 1 + G'_1(1)H'_1(1) \implies H'_1(1) = 1/(1 - G'_1(1))$

And thus, $\langle s \rangle = 1 + \frac{G'_0(1)}{1 - G'_1(1)}$

• Now evaluating the derivative:

$$G_1'(x) = \frac{d}{dx} \frac{1}{\langle k \rangle} G_0'(x) = \frac{1}{\langle k \rangle} \frac{d}{dx} \sum_k k P_k x^{(k-1)}$$
$$= \frac{1}{\langle k \rangle} \sum_k k(k-1) P_k x^{(k-2)}$$

• Evaluate at x = 1

$$G_1'(1) = \frac{1}{\langle k \rangle} \sum_k k(k-1) P_k = \frac{1}{\langle k \rangle} \left[\left\langle k^2 \right\rangle - \left\langle k \right\rangle \right]$$

Expected size of a component starting from arbitrary node

•
$$\langle s \rangle = 1 + \frac{G'_0(1)}{1 - G'_1(1)}$$

- $G'_0(1) = \langle k \rangle$
- $G'_1(1) = \frac{1}{\langle k \rangle} \left[\left\langle k^2 \right\rangle \left\langle k \right\rangle \right]$

$$\left\langle s \right\rangle = 1 + \frac{G_0'(1)}{1 - G_1'(1)} = 1 + \frac{\left\langle k \right\rangle^2}{2\left\langle k \right\rangle - \left\langle k^2 \right\rangle}$$

Emergence of the giant component

- $\langle s \rangle \to \infty$
- This happens when: $2\langle k \rangle = \langle k^2 \rangle$, which can also be written as $\langle k \rangle = (\langle k^2 \rangle \langle k \rangle)$
- This means expected number of nearest neighbors $\langle k \rangle$, first equals expected number of second nearest neighbors $(\langle k^2 \rangle \langle k \rangle)$.
- Can also be written as $\langle k^2 \rangle 2 \langle k \rangle = 0$, which is the famous Molloy and Reed criteria^{*}, giant emerges when:

$$\sum_{k} k \left(k - 2 \right) P_k = 0.$$

*GF approach is easier than Molloy Reed!

GFs widely used in "network epidemiology"

- Fragility of Power Law Random Graphs to targeted node removal / Robustness to random removal
 - Callaway PRL 2000
 - Cohen PRL 2000
- Onset of epidemic threshold:
 - C Moore, MEJ Newman, Physical Review E, 2000 MEJ
 Newman Physical Review E, 2002
 - Lauren Ancel Meyers, M.E.J. Newmanb, Babak Pourbohlou,
 - Journal of Theoretical Biology, 2006
 - JC Miller Physical Review E, 2007
- Information flow in social networks
 F Wu, BA Huberman, LA Adamic, Physica A, 2004.
- Cascades on random networks Watts PNAS 2002.