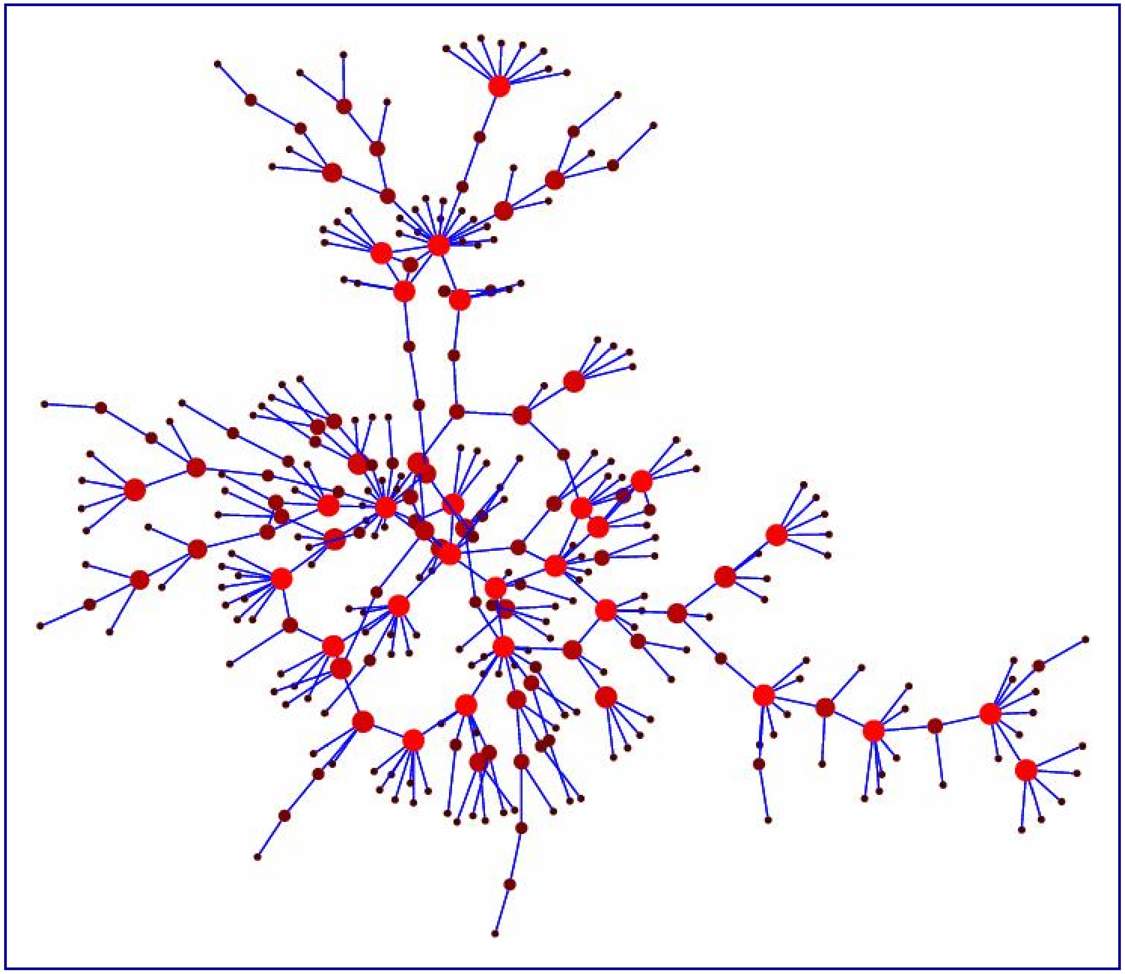
ECS 253 / MAE 253, Lecture 14

May 12, 2016



“Diffusion, Cascades and Influence”

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.

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

Generating functions / Master equations / giant components

- Contact processes / more similar to biological epidemic spreading
- Heterogeneity due to node degree (not due to different node preferences)
- Epidemic spreading:
- Social nets: Watts PNAS (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)

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

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

$$
\begin{aligned}
&= C \sum_j A_{ij}\phi_j - C\phi_i \sum_j A_{ij} \\
&= C \sum_j A_{ij}\phi_j - C\phi_i k_i \\
&= C \sum_j \left(A_{ij}\phi_j - \delta_{ij}k_i\right)\phi_j.
\end{aligned}
$$

- 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 $\mathbf{D}$ has zero entries except for diagonal with is degree of node:

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

The graph Laplacian

- L has real positive eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$.

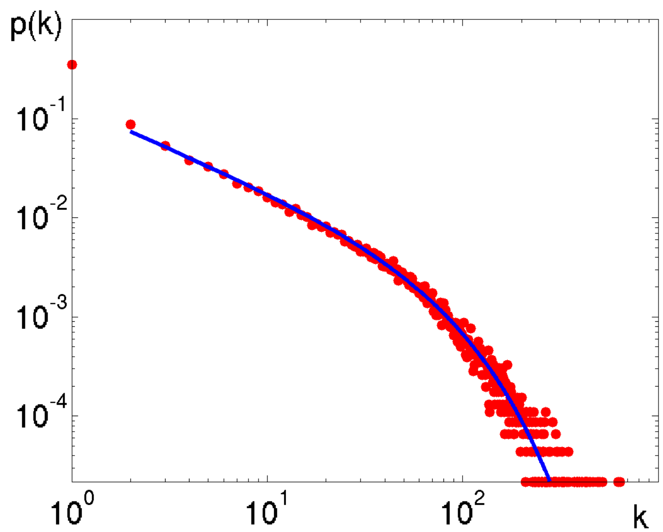
- Number of eigenvalues equal to 0 is the number of distinct, disconnected components of a graph

  (for the column-normalized state transition vector (i.e., random-walk), it is the number of λ's equal to 1).

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

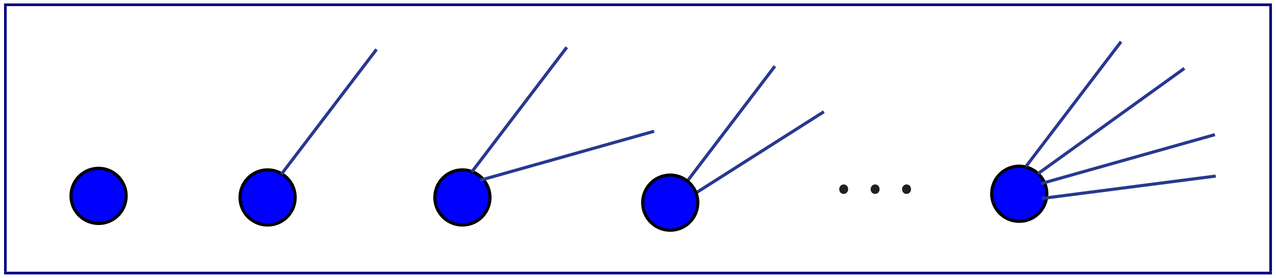
Part I. Ensemble approaches

- 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, $G_0(1) = \sum_k P_k = 1.$

- 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 $x = 1$, $G_0'(1) \equiv \frac{d}{dx} G_0(x)\Big|_{x=1} = \sum_k k P_k$ (the mean)

The moments of P_k can be obtained as a function of $G_0(x)$

- First moment, $\langle k \rangle = \sum_k k P_k = G_0'(1)$

  (And note $xG_0'(x) = \sum_k k P_k x^k$)

- Second moment, $\langle k^2 \rangle = \sum_k k^2 P_k$

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

  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

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

Generating functions for the giant component

Newman, Watts, Strogatz *PRE* 64 (2001)

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

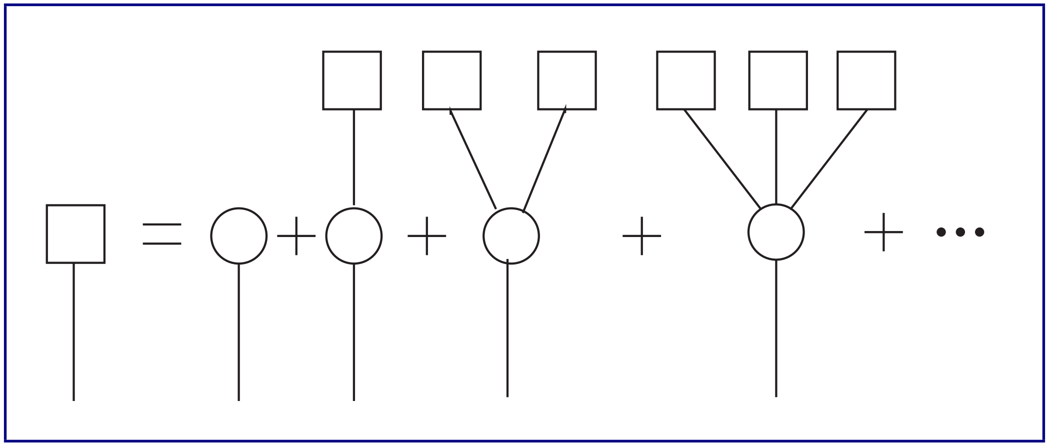
1. G.F. for connectivity of a node at edge of randomly chosen edge.

2. G.F. for size of the component to which that node belongs.

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:
$$q_k = kP_k / \sum_k kP_k.$$
- 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 q'_k.



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

For convenience, define the GF for random edge following

(Build up more complex from simpler)

- Recall prob of following random edge to node of degree k:

$$q_k = kP_k / \sum_k kP_k = kP_k / \langle k \rangle.$$

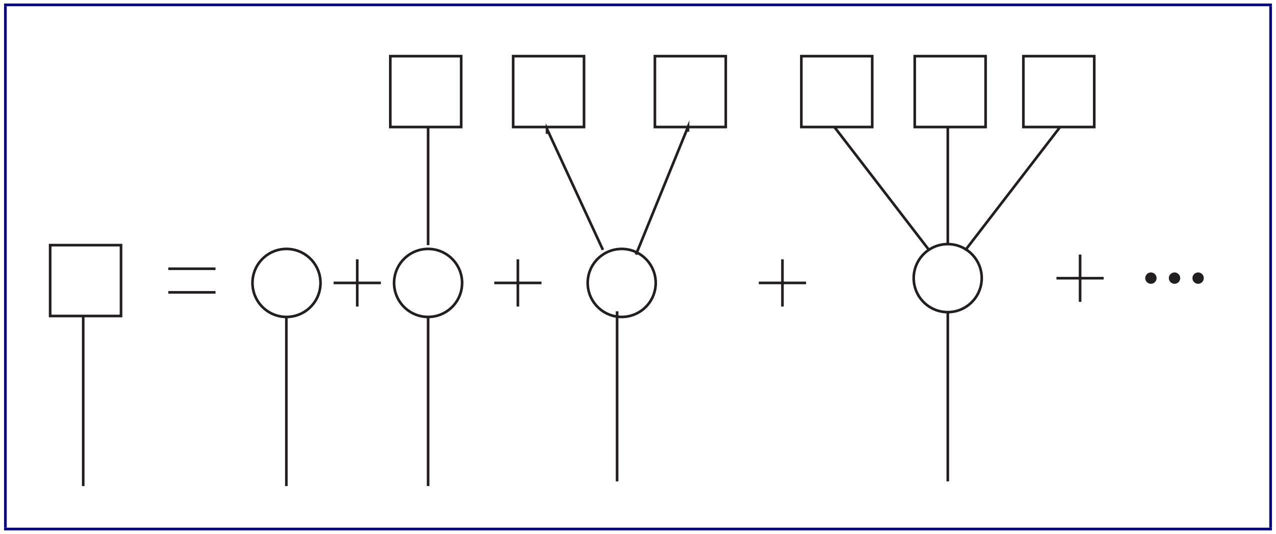
- Define a corresponding GF called $G_1(x)$:

$$\begin{aligned} G_1(x) &= \sum_k q_k x^k \\ &= \sum_k kP_k x^k / \langle k \rangle \\ &= x \tfrac{d}{dx} (\sum_k P_k x^k) / \tfrac{d}{dx} G_0(x) \Big|_{x=1} \\ &= x \tfrac{d}{dx} G_0(x) / \tfrac{d}{dx} G_0(x) \Big|_{x=1} \\ &\equiv x G_0'(x) / G_0'(1). \end{aligned}$$

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

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

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



$$H_1(x) = xq_0 + xq_1 H_1(x) + xq_2[H_1(x)]^2 + xq_3[H_1(x)]^3 \cdots$$

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

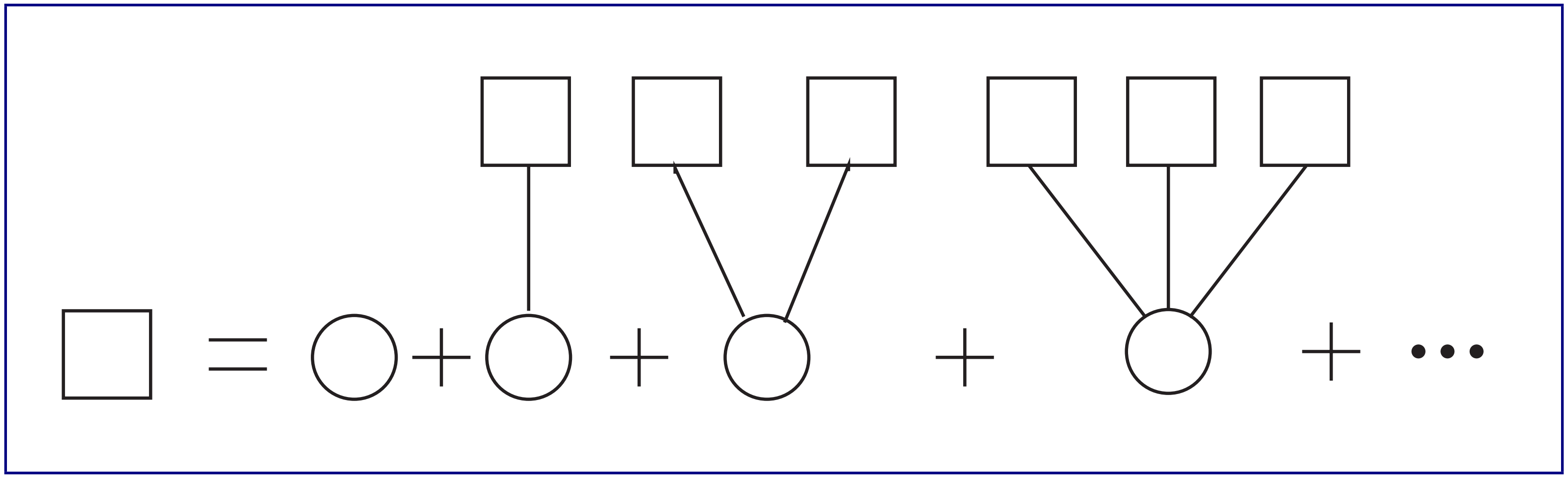
Note also that $H_1(x) = x \sum_k q_k [H_1(x)]^k = xG_1(H_1(x))$

Aside: Powers property

Note last slide we used the "Powers" property, that GF for the probability a random variable k summed over m independent realizations of the object is the mth power of GF for k, hence the $[H_1(x)]^m$ factors above.

- 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



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

$$= x\sum_k P_k[H_1(x)]^k = xG_0(H_1(x))$$

- Can take derivatives of $H_0(x)$ to find moments of component sizes!

- Note we have assumed a tree-like topology.

Expected size of a component starting from arbitrary node

- $\langle s\rangle = \frac{d}{dx}H_0(x)\Big|_{x=1} = \frac{d}{dx}xG_0(H_1(x))\Big|_{x=1}$

  $$= G_0(H_1(1)) + \frac{d}{dx}G_0(H_1(1)) \cdot \frac{d}{dx}H_1(1)$$

  Since $H_1(1) = 1$, (sum of the probabilities)

  $$\langle s\rangle = 1 + G_0'(1) \cdot H_1'(1)$$

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

  so $H_1'(1) = 1 + G_1'(1)H_1'(1) \implies H_1'(1) = 1/(1 - G_1'(1))$

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

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 $\langle k \rangle = \left(\langle k^2 \rangle - \langle k \rangle\right)$

- This means expected number of nearest neighbors $\langle k \rangle$, first equals expected number of second nearest neighbors $\left(\langle k^2 \rangle - \langle k \rangle\right)$.

- 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:

$$\boxed{\sum_k k\,(k-2)\,P_k = 0.}$$

*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.

Global Cascades on Random Networks
Watts PNAS 2002

- Each node can be in one of two states, say $\{+1, -1\}$.
- Start with almost all nodes in $\{-1\}$, but just one node (or a small fraction of nodes) in $\{+1\}$.
- Nodes update state asynchronously. For node j if the fraction of its neighbors in state $+1$ is greater than a threshold function Φ_j, j switches to $+1$ and stays in that state forever.
- The thresholds Φ_j are drawn at random from a distribution $f(\Phi)$ which is normalized in the usual way: $\int_0^1 f(\Phi) d\Phi = 1$.
- *Local dependence, fractional threshold* Φ_j, *heterogeneous degree* make this model differ from contact processes.

Global cascades?

- Question: for what kinds of networks and thresholds will a small perturbation (of even one node) cause a fraction of all nodes to flip? (i.e. a global cascade).

- Some terms:
  Innovator – The first node(s) flipped to $+1$.
  Early adopter / vulnerable – A neighbor of innovator who flips right away.

- Early adopter much have threshold $\Phi_j \leq 1/k_j$, or equivalently degree $k_j \leq K_j = \lfloor 1/\Phi_j \rfloor$

Using GFs can reduce a complicated dynamics to a static percolation problem

- As usual, degree distribution P_k.

- A node is *vulnerable / early adopter* if it's threshold $\Phi \leq 1/k$. The probability a given node of degree k is vulnerable is thus

$$\rho_k = P\left[\Phi \leq 1/k\right] = \int_0^{1/k} f(\Phi) d\Phi.$$

- The probability a node drawn uniformly at random from all nodes has 1) degree k, and 2) is vulnerable is thus: $\rho_k P_k$.

- Generating function for this (our base GF)

$$G_0(x) = \sum_k \rho_k P_k x^k.$$

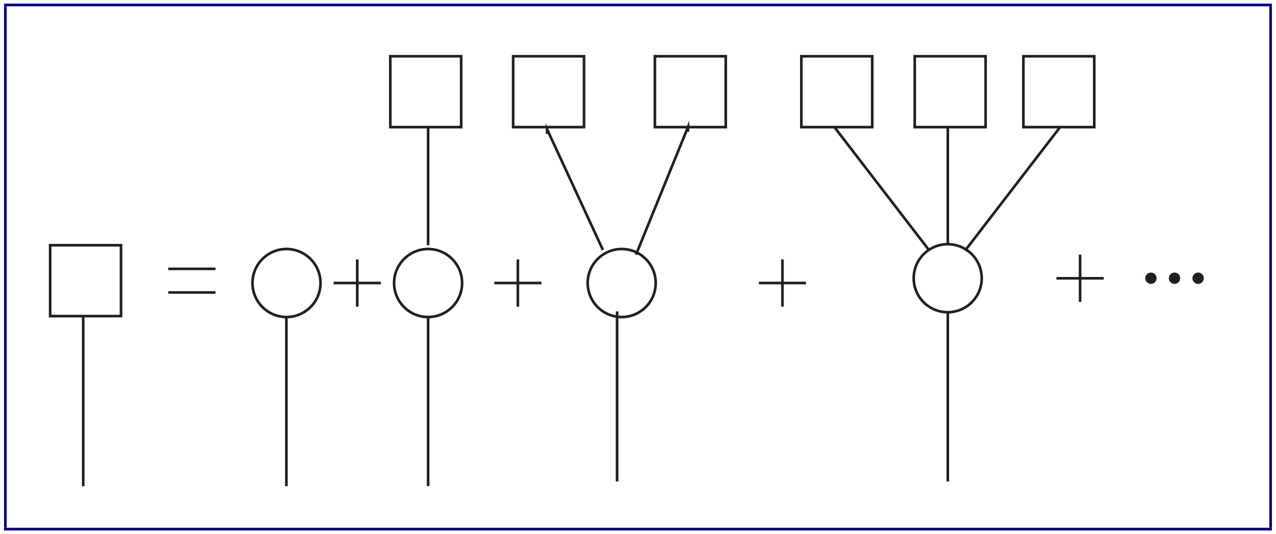
“Propagation” of a cascade is edge following from a vulnerable node

- As with the basic framework, probability of following edge to node of degree k is proportional to k.

- GF for following a random edge to a *vulnerable* node of degree k. (Again, observe building up process.):

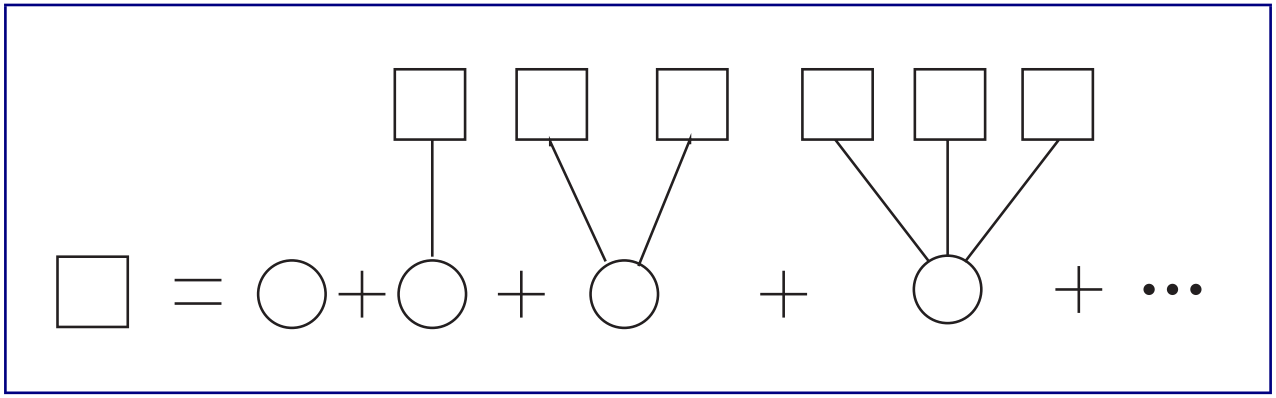
$$G_1(x) = \sum_k (k\rho_k P_k) / \sum_k kP_k = G_0'(x) / \langle k \rangle$$
$$= G_0'(x)/G_0'(1)$$

GF for size of component made of vulnerable nodes found by following initial edge:



$$H_1(x) = [1 - G_1(1)] + xG_1(H_1(x)).$$

GF for size of component made of vulnerable nodes found by choosing arbitrary node:



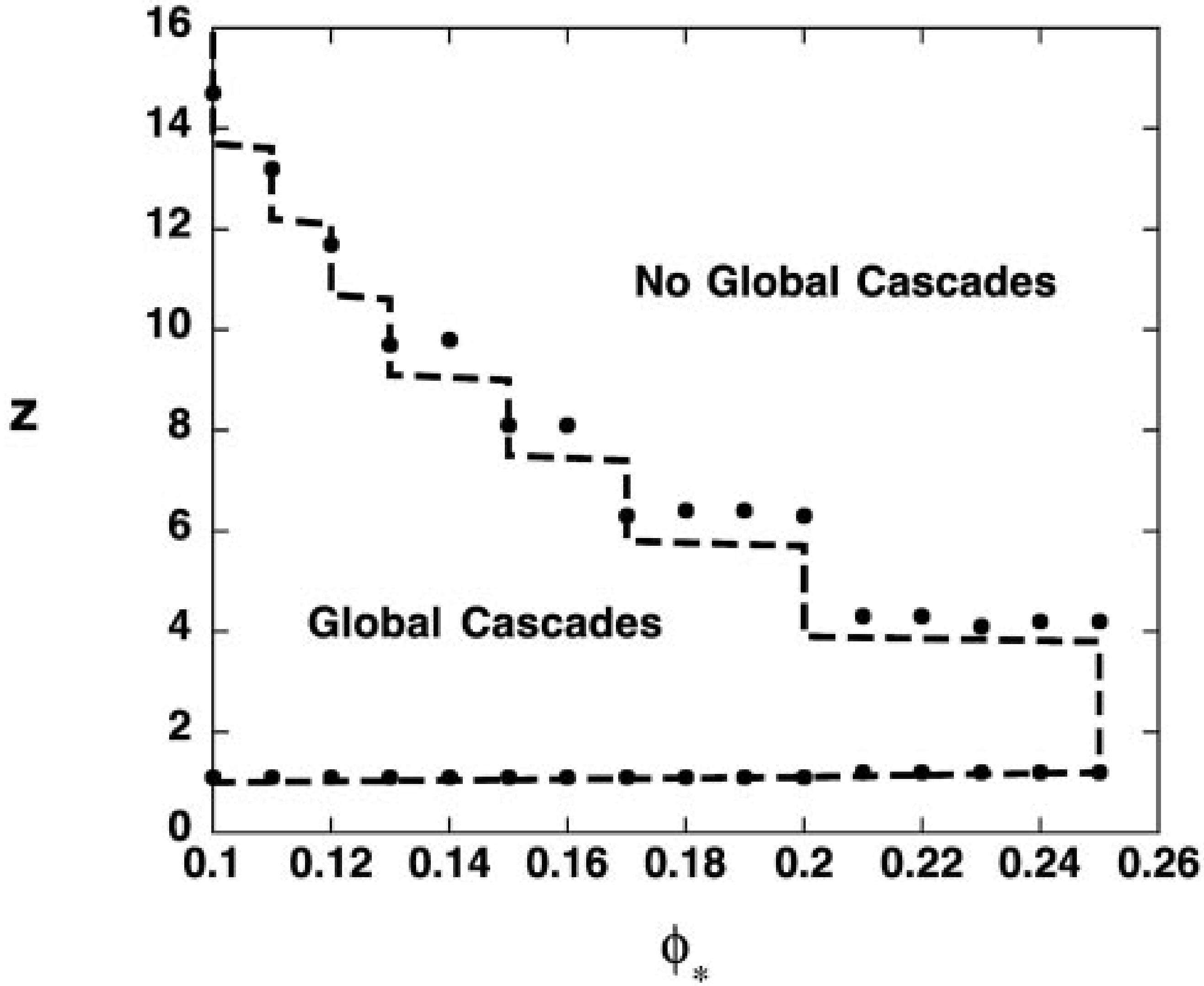
$$H_0(x) = [1 - G_0(1)] + xG_0(H_1(x)).$$

This leads to the cascade condition:

$$\boxed{\sum_k k(k-1)\rho_k P_k > \langle k \rangle}$$

Results: Theory and simulation

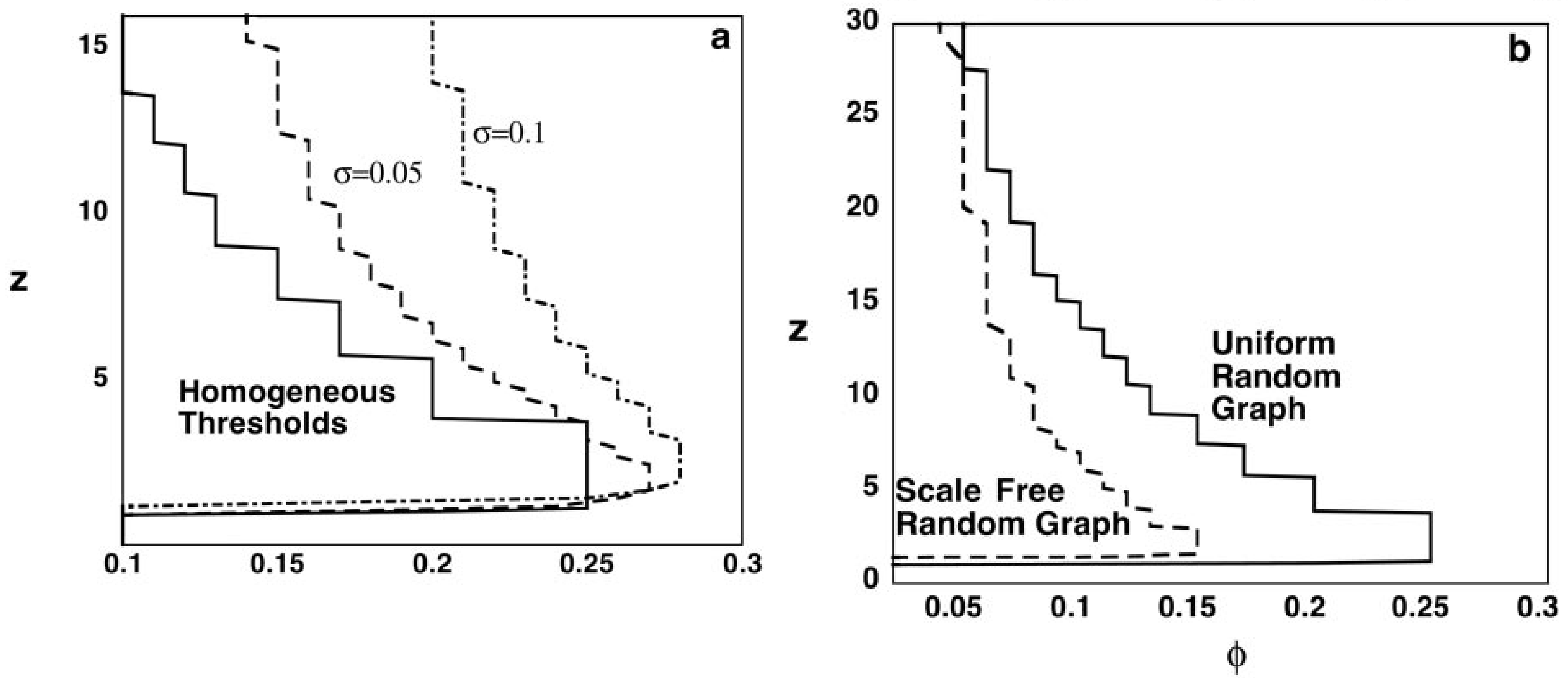
Uniform thresholds on random graph where all nodes have degree z (a z-regular random graph)



Results: Theory and simulation

(a) Normally distributed thresholds with std dev σ, on z-regular random graph.

(b) Uniform threshold on regular vs power law random graph.



- Heterogeneous thresholds seem to enhance global cascades.
- Heterogeneous node degrees seem to reduce global cascades.

Generating function approach to adoption of new behavior: Watts PNAS (2002)

- All nodes, except one, start in "inactive" state, $\{-1\}$
- Fractional threshold model (Φ_i).
  – Node "activated" once a fraction of it's neighbors $\geq \Phi_i$ are active.
  – A *vulnerable* node is one that needs only a single neighbor to be active before it flips (i.e., $\Phi_i \leq 1/k$).
  – Use generating functions to calculate the expected size of clusters of vulnerable nodes.
  – A "Global cascade" corresponds to a giant component
- Results
  – Heterogeneity in thresholds (Φ_i) **enhances** global cascades.
  – Heterogeneity of degree (P_k) **suppresses** global cascades.

Susceptibles versus influentials

- A long debate
- Malcolm Gladwell vs Duncan Watts
- Aral and Walker, *Science*, 2012.

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)

Part II. Network Coordination Games

- The most basic model: Reviewed in Kleinberg "Cascading Behavior in Networks: Algorithmic and Economic Issues", Chap 24 of *Algorithmic Game Theory*, (Cambridge University Press, 2007).
- Again each node in one of two states, say $\{-1, +1\}$.
- Play a game with each connected neighbor independently. Total payout is sum over all games.
- Assume neighbor(s) of j in fixed state while j updates.
- Positive payout if connected nodes i and j adopt the same state. No payout if they differ. And -1 can have different payout that +1 coordinated behavior.

Payout matrix:

q	0
0	(1-q)

How each node operates

- Again assume all other nodes fixed while node j updates.
- It has k_j^A nodes in state -1, and k_j^B nodes in state $+1$.
- If it chooses state -1, payout of qk_j^A.
- If it chooses state $+1$, payout of $(1-q)k_j^B$.
- Chooses -1 if $qk_j^A > (1-q)k_j^B$.
- Substitute in $k_j = k_j^A + k_j^B$ and rearrange:
  Criteria: choose -1 if $k_j^B < qk_j$ and $+1$ if $k_j^B > qk_j$.
- A **threshold** model! Adopt $+1$ if a fraction q of your neighbors have state $+1$.

Contagion threshold and cascades

- Start all nodes in -1. And all nodes update synchronously at discrete time steps.

- **Key question:** When is there a small set of nodes S, that when set to $+1$ convert all (or almost all) of the population?

- A set S is *contagious* if every other node is converted by S.

- Easier for S to be contagious if the threshold q is small.

- Define the *contagion threshold* of a graph G to be the maximum q for which there exists a finite contagous set.

- (Like with generating functions, here no notion of how long it takes for the full network to be activated. Just a final steady-state answer.)

Progressive vs. non-progressive processes

- The model thus far is non-progressive: nodes can flip from -1 to $+1$ and back to -1.
- This makes the situation less stable. Consider a line of all -1 at the start with a single $+1$ in the center, and $q = 1/2$. At next time steps neighbors of the $+1$ flip, but the $+1$ switches back to -1! And the whole system ends up “blinking”.
- Progressive: Once you flip, always stay in that state.
- The line above now all flips to $+1$ in a wavefront moving right and left-wards.

Theorem:
The Contagion Threshold for any Graph is at most 1/2.

- (Recall the contagion threshold is the maximum value of q for which a finite contagious set exists.)
- Independent of progressive vs non-progressive.
- "A behavior cant spread very far if it requires a strict majority of your friends to convince you to adopt it."
- This means if $q > 1/2$ on any graph, it cannot support a cascade and the full graph will not be activated.
- This is for any graph: uniform degree, power law, etc.

Extending this simple model

So far all nodes have same fractional threshold q, and all neighbors contribute equally in calculation of fraction.

- **The General Linear Threshold Model**
  - Directed graphs (not reciprocal influence necessarily).
  - Each node has a threshold chosen uniformly at random between $[0, 1]$.
  - Each neighbor exerts a non-negative weight. The only constraint is that sum over all the weights is less than or equal to 1.
  - Note we now have diversity of influence (e.g., spouse/relative can exert stronger weight than coworker/friend).

Finding the influential nodes
Motivation

- Viral marketing – use word-of-mouth effects to sell product with minimal advertising cost.

- Design of search tools to track news, blogs, and other forms of on-line discussion about current events

Finding the influential nodes: formally

- The minimum set $S \in V$ that will lead to the whole network being activated.

- The optimal set of a specified size $k = |S|$ that will lead to largest portion of the network being activated.

Due to thresholds/ critical mass

- In general NP-hard to find optimal set S.

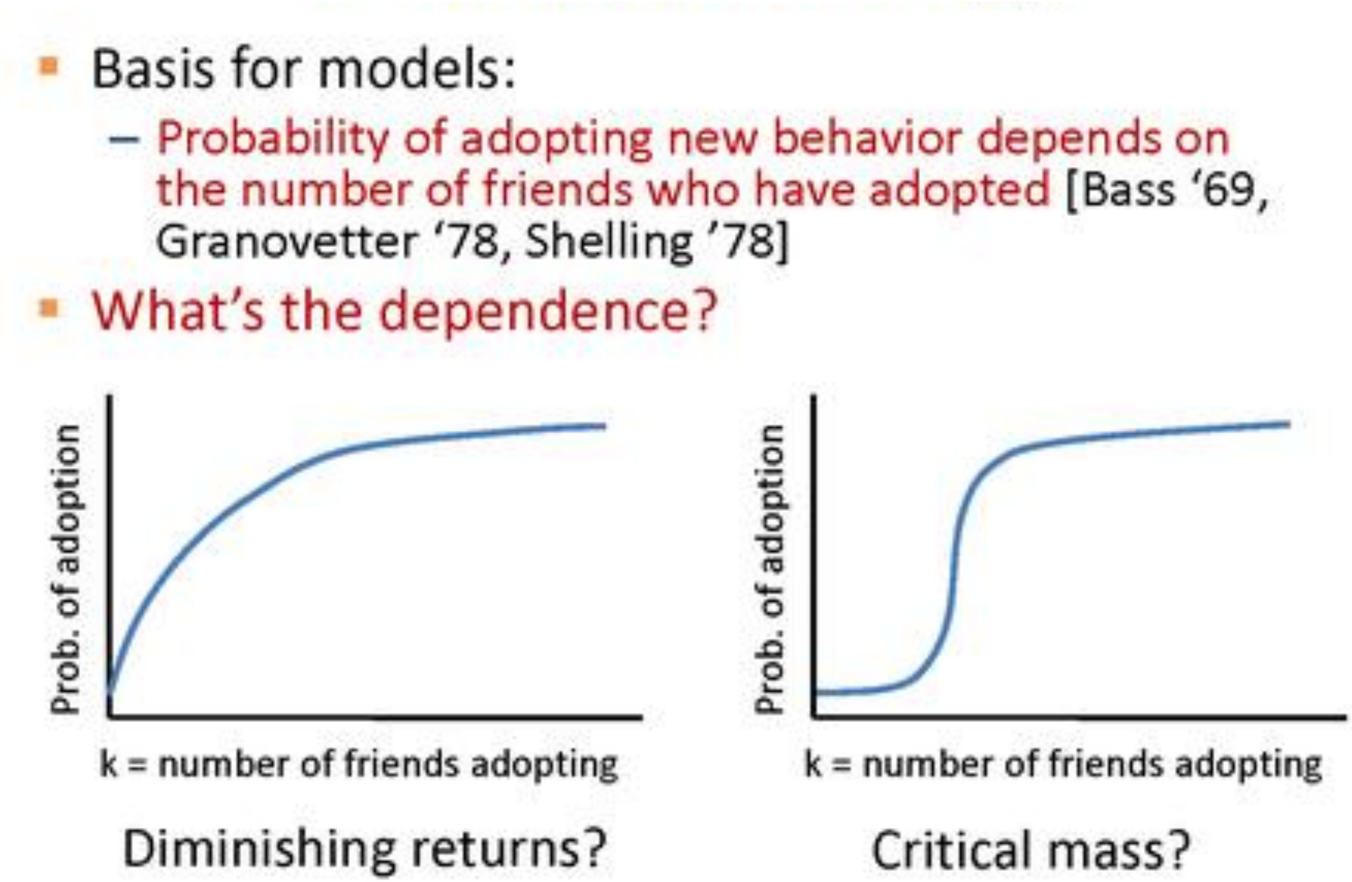
- NP-hard to even find a approximate optimal set (optimal to within factor $\eta^{1-\epsilon}$ where n is network size and $\epsilon > 0$.) ("inapproximability")

- Due to thresholds (esp if each node can have its own) might have a tiny activated final set of nodes but it jumps abruptly if just a few more nodes or, moreover, the right nodes activated.

- Kleinberg calls this abrupt response the "Knife edge" property

Diminishing returns
(No longer a threshold, but a concave function)

- Each additional friend who adopts the new behavior enhances your chance of adopting the new behevaior, but with less influence for each additional friend



(from Leskovec talk)

Diminishing returns
(Submodular / concave function)

- The benefit of adding elements decreases as the set to which they are being added grows.

- So no longer get to have more influence from family or other special nodes. (Instead its the first nodes exert more influence.)

- Since no longer have special nodes easy to build up optimal set S of k nodes.

- **Hill climbing** – add one at the time nodes to the set S that cause maximum impact.

Hill climbing

An Approximation Result

- Diminishing returns: $p_v(u,S) \geq p_v(u,T)$ *if* $S \subseteq T$
- Hill-climbing: repeatedly select node with maximum marginal gain
- Performance guarantee: hill-climbing algorithm is within (1-1/e) ~63% of optimal [Kempe et al. 2003]

(from Leskovec talk)

Submodular and hill climbing more formally:

An Approximation Result

- Analysis: diminishing returns at individual nodes implies diminishing returns at a "global" level
  - Cascade size *f(S)* grows slower and slower with *S*.

    f is submodular: if $S \subseteq T$ then

    $$f(S \cup \{x\}) - f(S) \geq f(T \cup \{x\}) - f(T)$$
  - <u>Theorem</u> [Nehmhauser et al. '78]:
    If *f* is a function that is monotone and submodular, then *k*-step hill-climbing produces set *S* for which *f(S)* is within *(1-1/e)* of optimal.

(from Leskovec talk)

Empirical observations

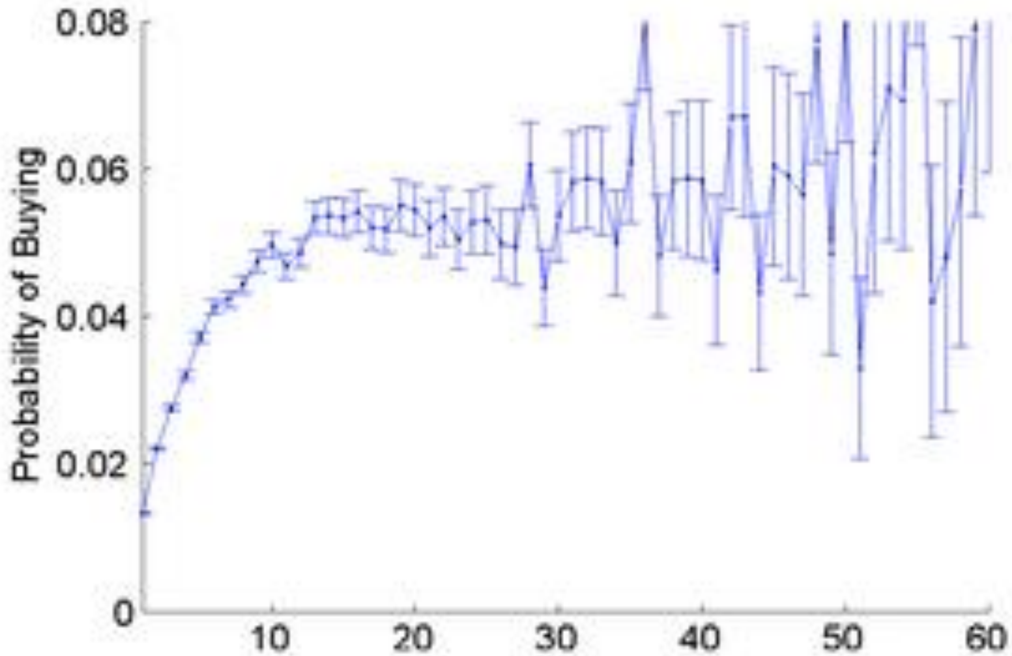
Part 2: Empirical Analysis

- What do diffusion curves look like?
- How do cascades look like?
- Challenge:
  - Large dataset where diffusion can be observed
  - Need social network links and behaviors that spread
- We use:
  - Blogs: How information propagates? [Leskovec et al. 2007]
  - Product recommendations: How recommendations and purchases propagate? [Leskovec-Adamic-Huberman 2006]
  - Communities: How community membership propagates? [Backstrom et al. 2006]

(from Leskovec talk)

How do diffusion curves look like?

- Viral marketing – DVD purchases:

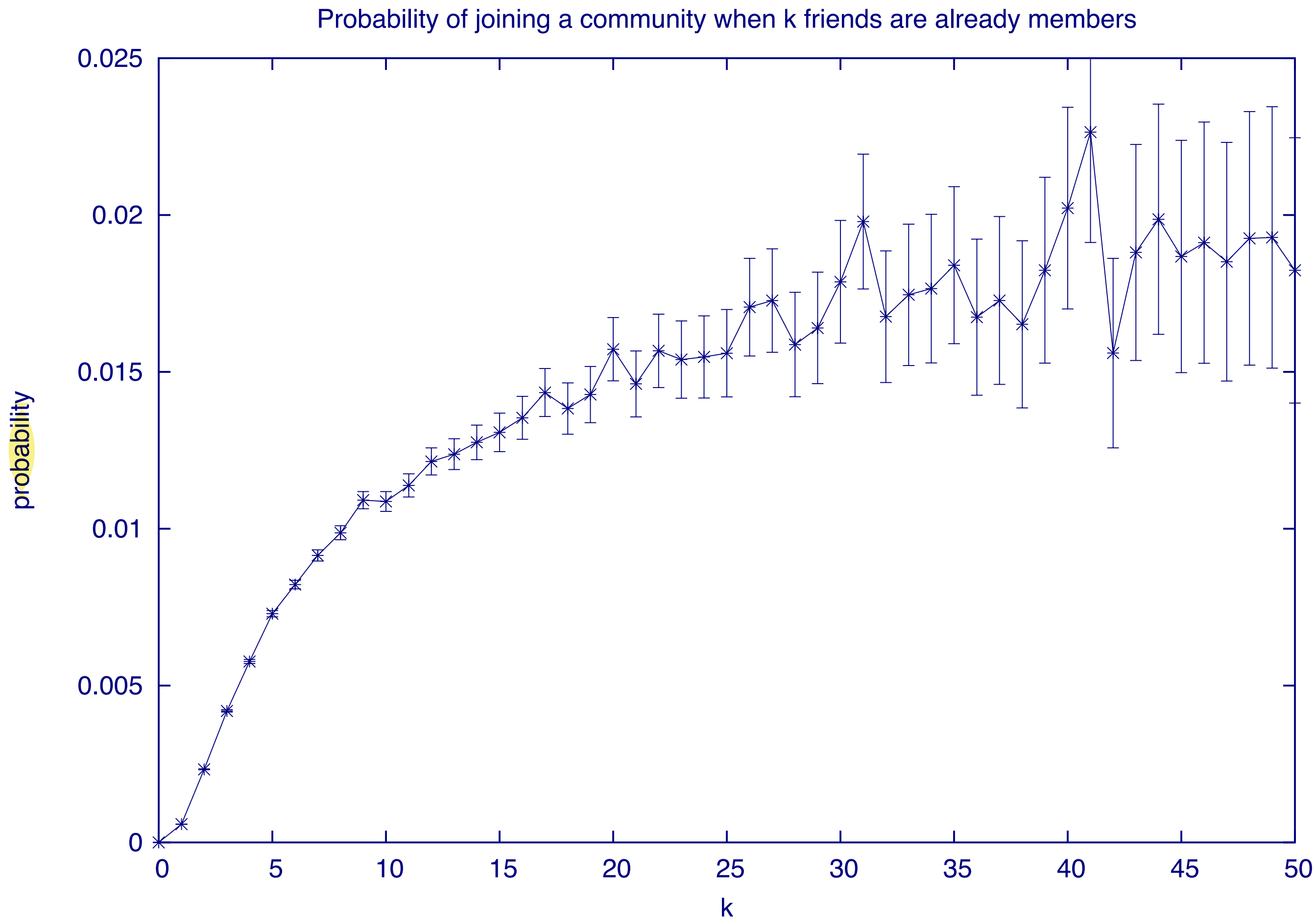


k (number of in-recommendations)

  - Mainly diminishing returns (saturation)
  - Turns upward for k = 0, 1, 2, ...

(from Leskovec talk)

Joining Livejournal: on online bulletin board network



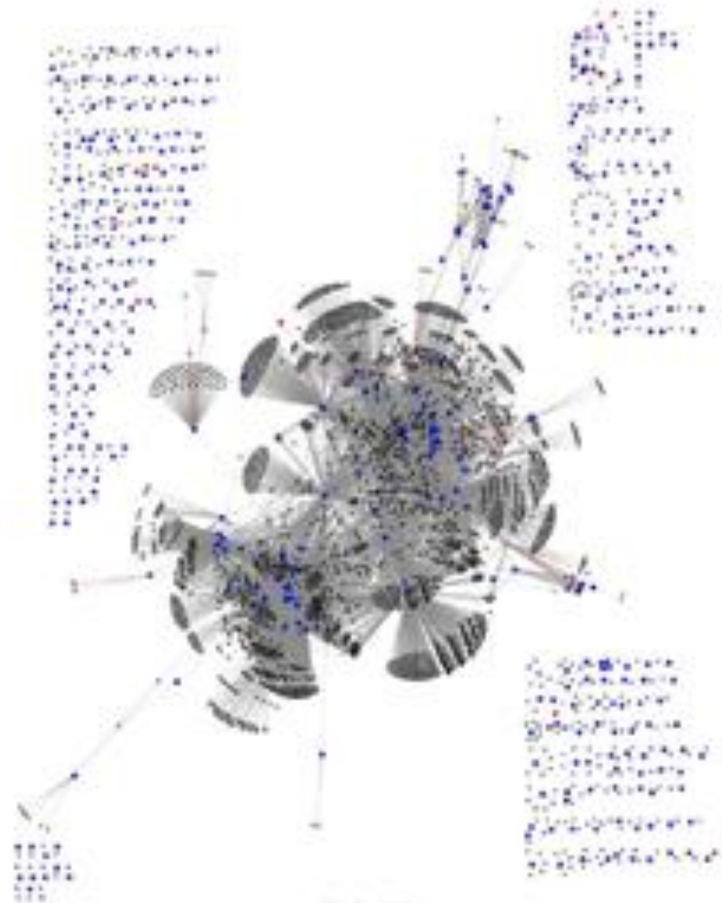
- Diminishing returns only sets in once $k > 3$.

- Network effect not illustrated by curve: If the k friends are highly clustered, the new user is more likely to join.

How Do Cascades Look Like?

- How big are cascades?
- What are the building blocks of cascades?

Medical guide book



DVD

(from Leskovec talk)