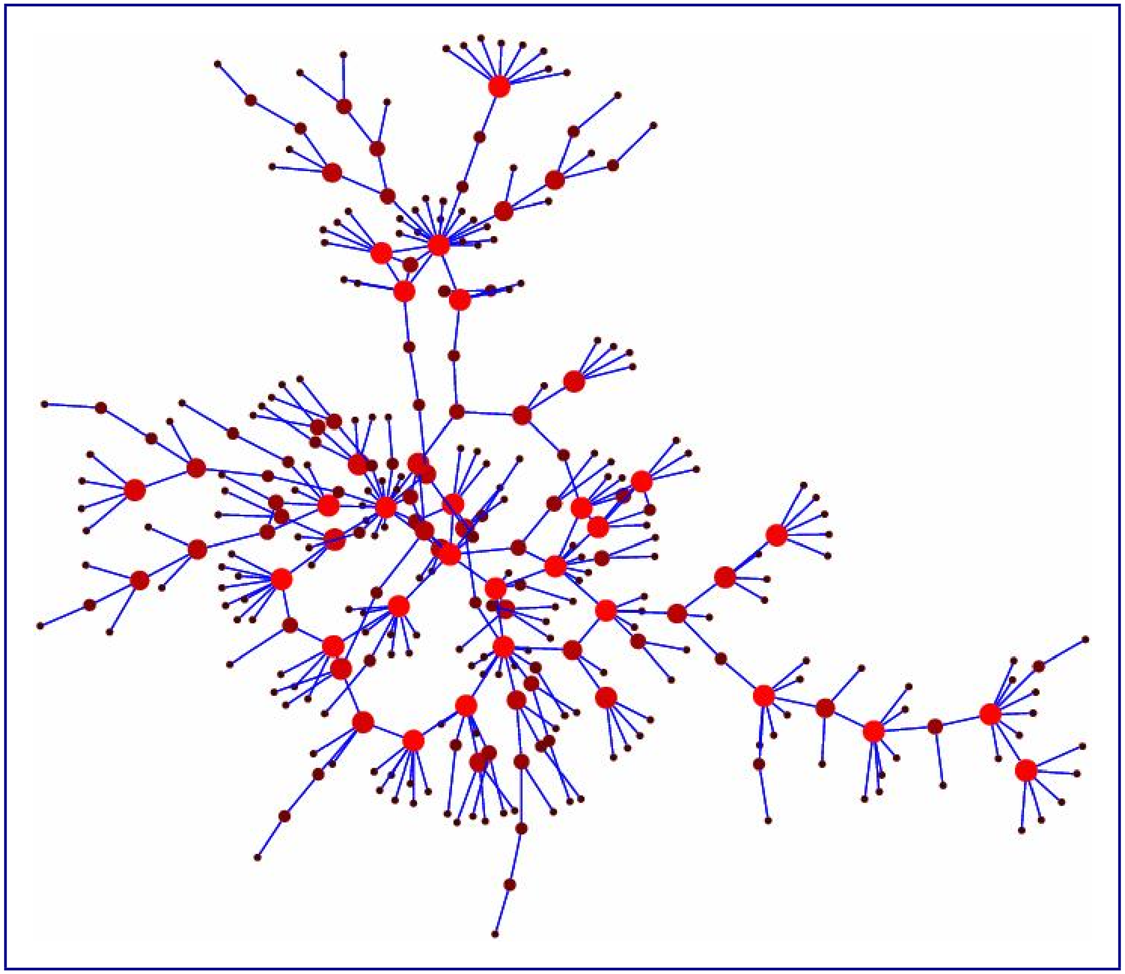
ECS 253 / MAE 253, Lecture 15

May 17, 2016

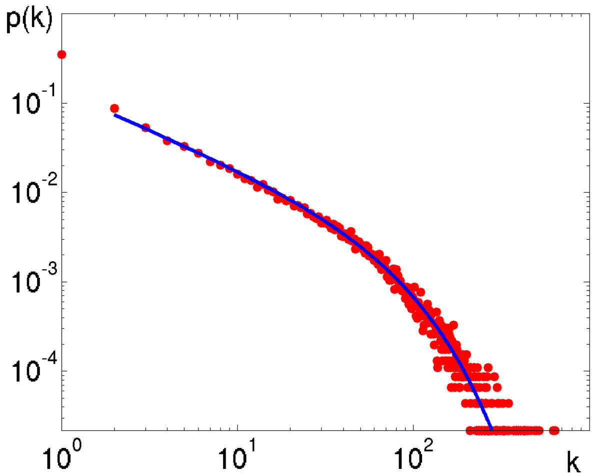


I. Probability generating function recap

Part I. Ensemble approaches

- A. Master equations (Random graph evolution, cluster aggregation)

- B. Network configuration model
  - Degree distribution, P_k



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

- C. Generating functions. Converting a discrete math problem into a function.

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

Moment generating functions

- Base: $G_0(1) = \sum_k P_k = 1$ (it is the sum of probabilities).

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

  (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 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.

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.

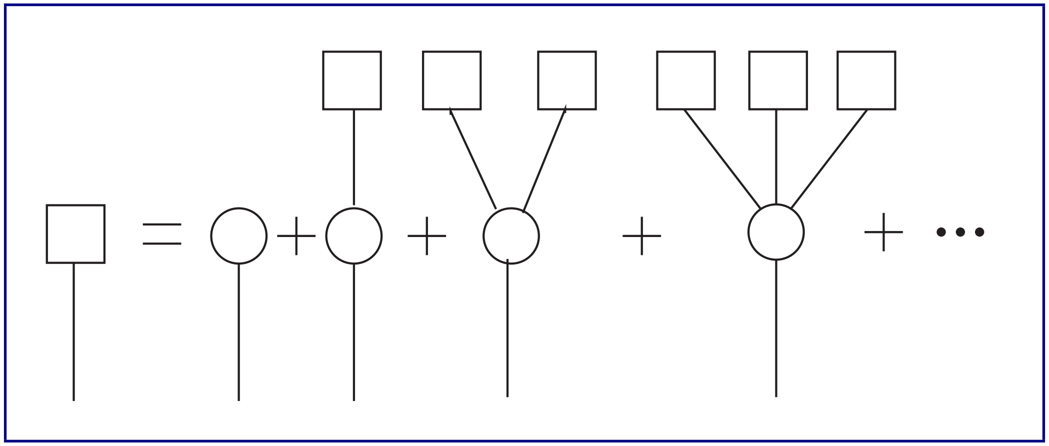
Following a random edge

- 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:

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

- 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 PGF 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

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

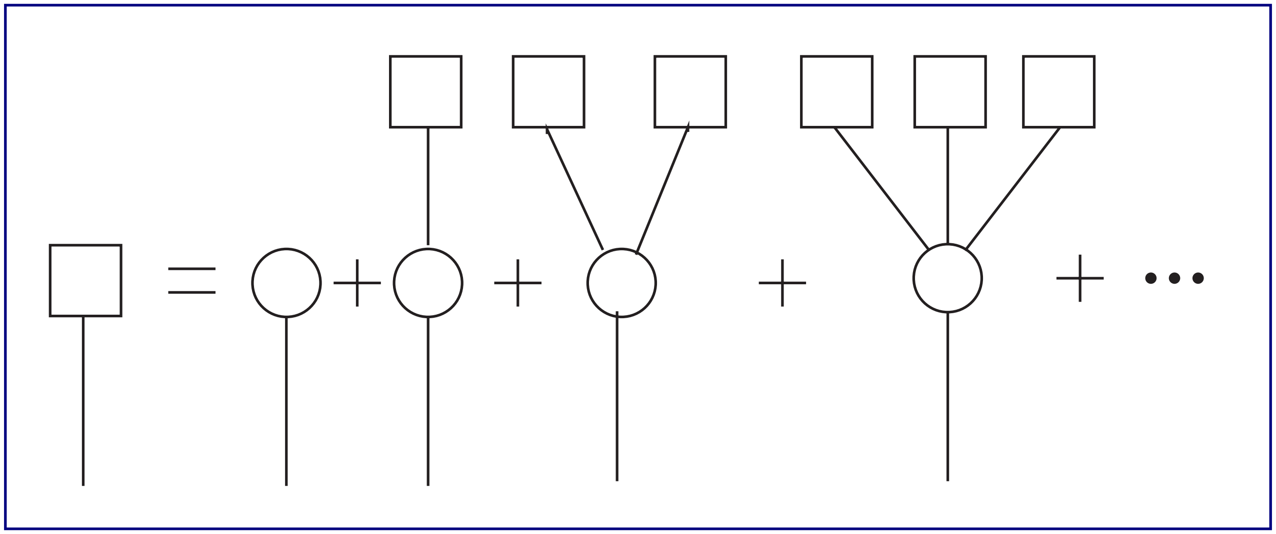
$$= \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} k P_k x^{k-1}$$

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

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

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

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

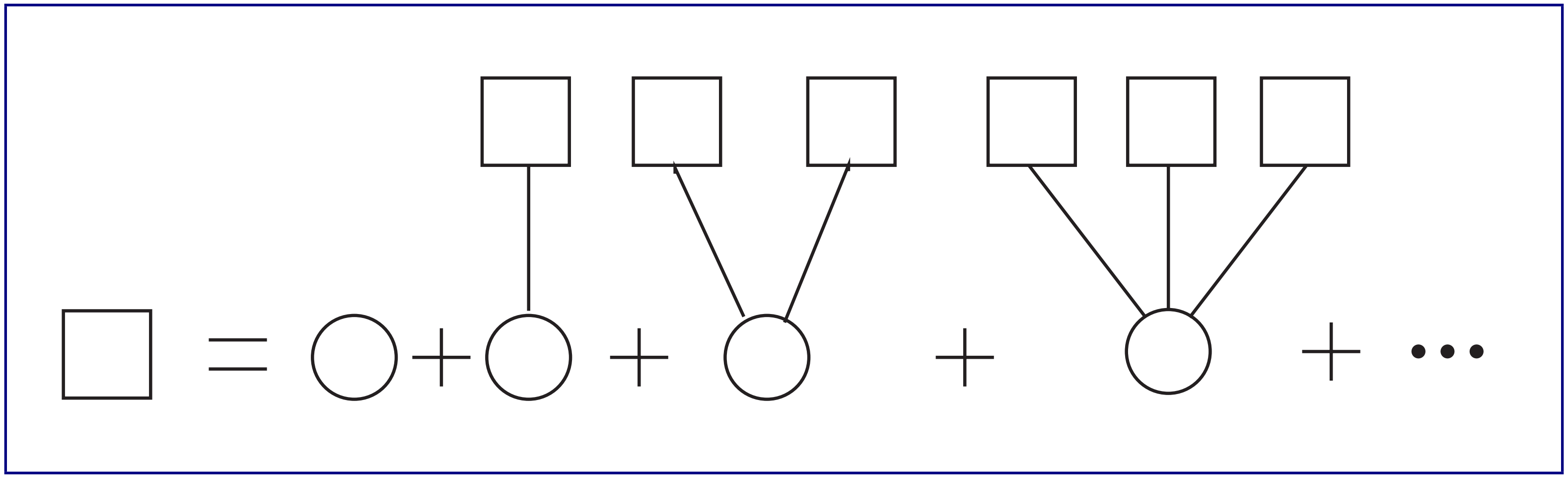


$$H_1(x) = xq_0 + xq_1 H_1(x) + xq_2[H_1(x)]^2 + xq_3[H_1(x)]^3 \cdots$$

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

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

$H_0(x)$, 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 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) $H_1(x) = x G_1(H_1(x))$

  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:

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

$$= \frac{1}{\langle k \rangle}\sum_k k(k-1)P_k x^{(k-2)}$$

- Evaluate at $x = 1$

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

Expected size of a component starting from arbitrary node

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

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

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:

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

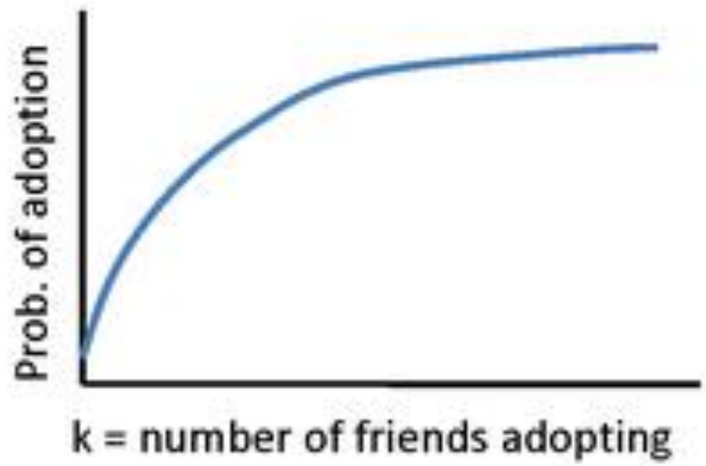
*GF approach is easier than Molloy Reed!

PGFs widely used in network theory

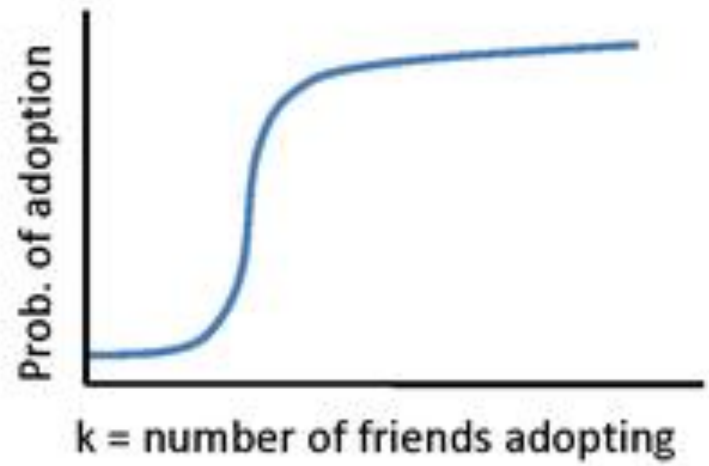
- Fragility of Power Law Random Graphs to targeted node removal / Robustness to random removal
  - Callaway PRL 2000
  - Cohen PRL 2000
- Onset of epidemic threshold:
  - C Moore, MEJ Newman, Physical Review E, 2000
  - MEJ Newman - Physical Review E, 2002
  - Lauren Ancel Meyers, M.E.J. Newmanb, Babak Pourbohlou, Journal of Theoretical Biology, 2006
  - JC Miller - Physical Review E, 2007
- **Cascades on random networks**
  Watts PNAS 2002.
  Susceptible agents drive social change

What really drives a node to activate?

- Basis for models:
  - Probability of adopting new behavior depends on the number of friends who have adopted [Bass '69, Granovetter '78, Shelling '78]
- What's the dependence?



Diminishing returns?



Critical mass?

(from Leskovec talk)

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

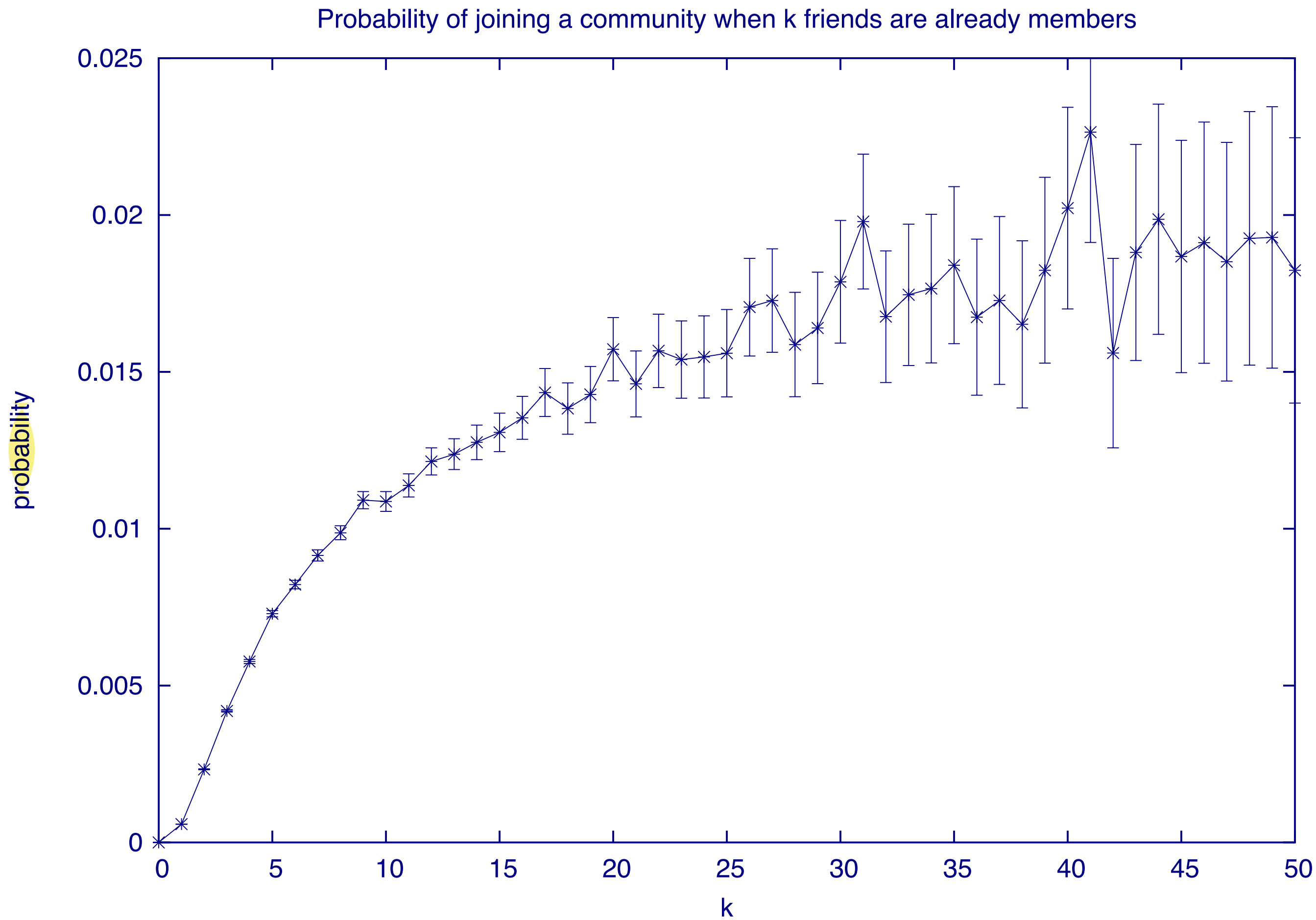
Finding the influential nodes: formally

- 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.

Influentials

- Critical mass model
  - NP-hard problem.
  - 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”)
- Diminishing returns
  - Greedy algorithms (e.g. “hill-climbing” within $(1-1/e) \sim 63\%$ of optimal)

Joining Livejournal: on online bulletin board network

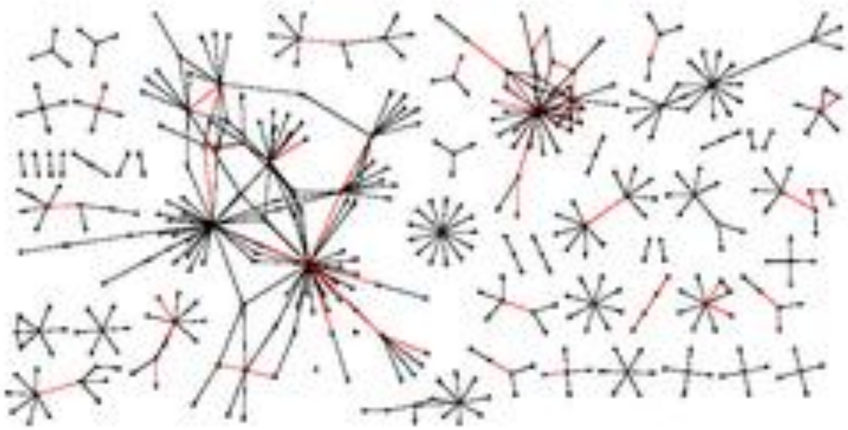


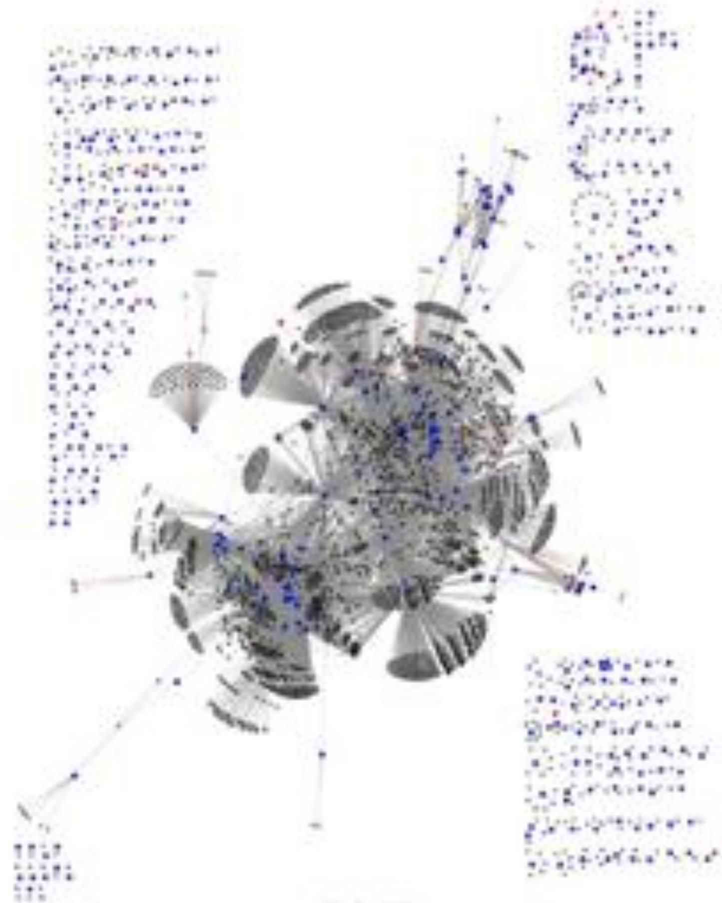
- Diminishing returns only sets in once $k > 3$.

- 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)