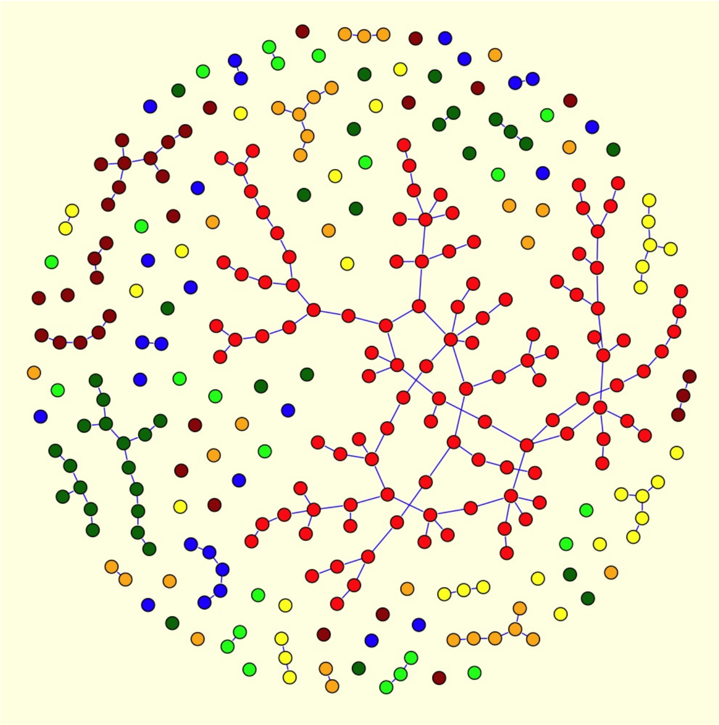
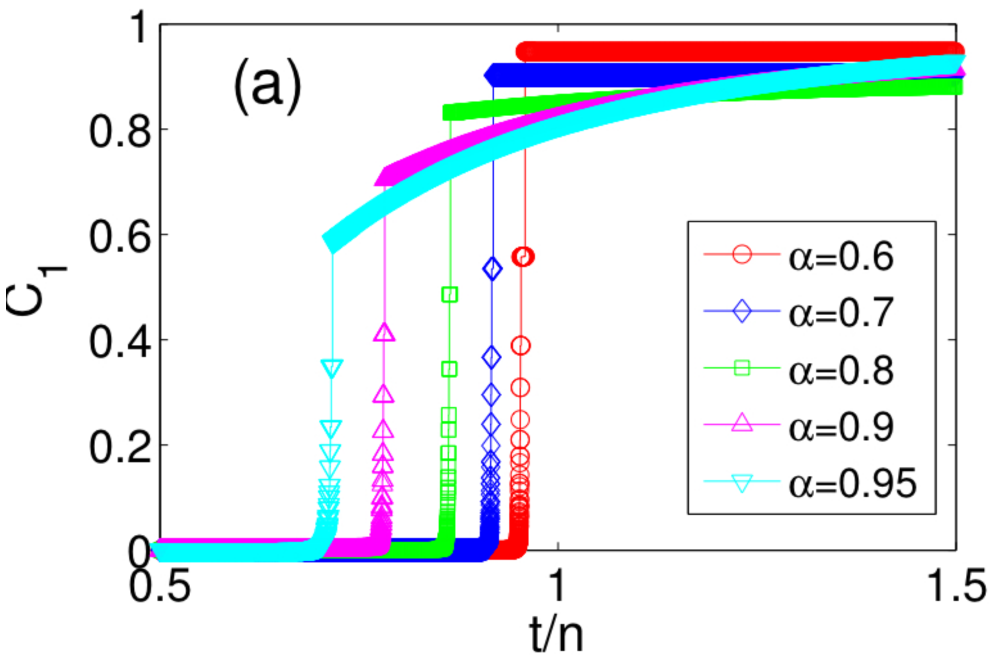
"Explosive" percolation transitions

(tomorrow: cascades on interdependent networks)





Raissa D'Souza

University of California, Davis

Dept of Mech. and Aero. Eng., Dept of CS

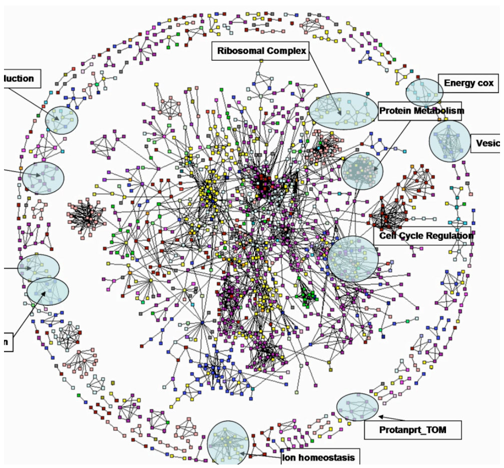
Complexity Sciences Center

External Professor, **Santa Fe Institute**

Networks are increasingly ubiquitous:

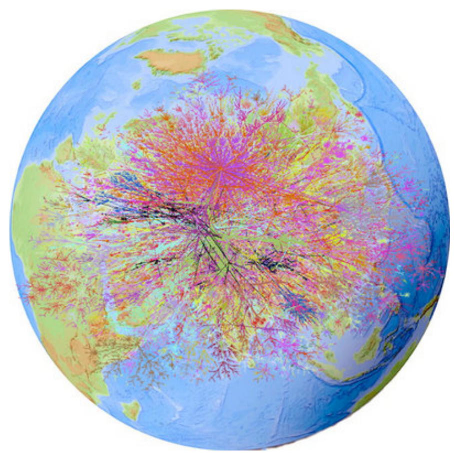
Networks:

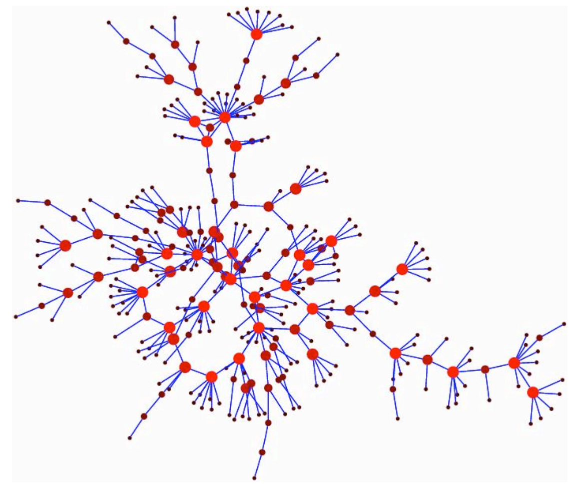
Transportation Networks/ Power grid
(distribution/ collection networks)



Biological networks
- protein interaction
- genetic regulation
- drug design

Computer networks



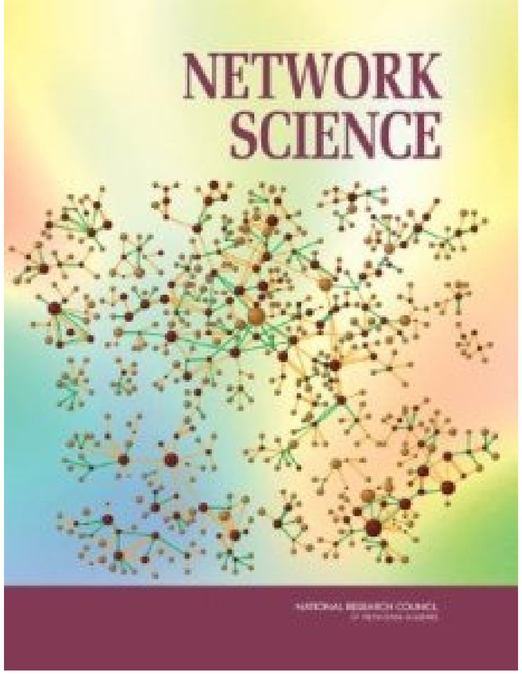


Social networks
- Immunology
- Information
- Commerce

(**Network**: a collection of discrete nodes/vertices connected to others by edges)

The past decade, a "Science of Networks":
(Physical, Biological, Social)

- **Geometric** versus **virtual** (Internet versus WWW).
- **Natural** /spontaneously arising versus **engineered** /built.
- Each network may **optimize** something unique.
- Fundamental **similarities** and **differences** to guide design/understanding/control.
- Interplay of **topology** and **function** ?
- Up until now, **studied largely as individual networks in isolation .**

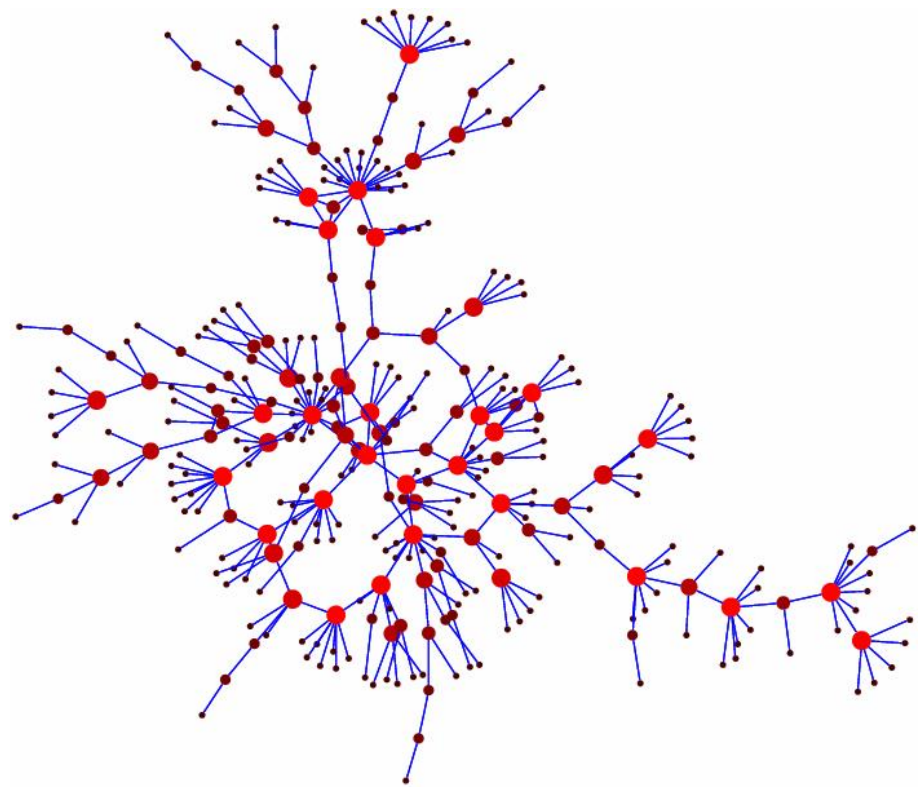


NRC, 2005

Achievements of Single Network View

(Goal : Intuition, prediction, design, control)

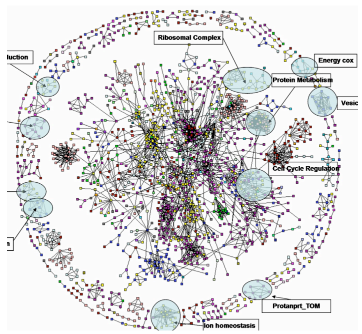
- **Power law** (broad scale) degree distributions ubiquitous.
- **Small world** effect (small diameter and local clusters).
- **Vulnerability** to "hub" removal **resilience** to random removal.
- **Percolation**, spreading and epidemics (phase transitions)
- **Cascades**.
- **Synchronization**.
- Random walks / **Page rank**.
- **Communities** / modules.



In reality a collection of interacting networks:

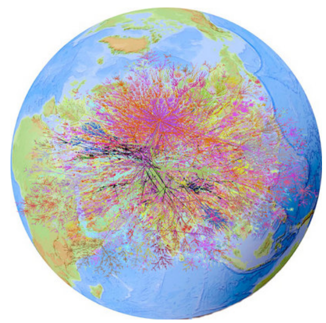
Networks:

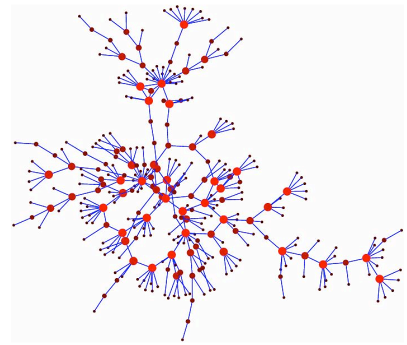
Transportation Networks/ Power grid
(distribution/ collection networks)



Biological networks
- protein interaction
- genetic regulation
- drug design

Computer networks





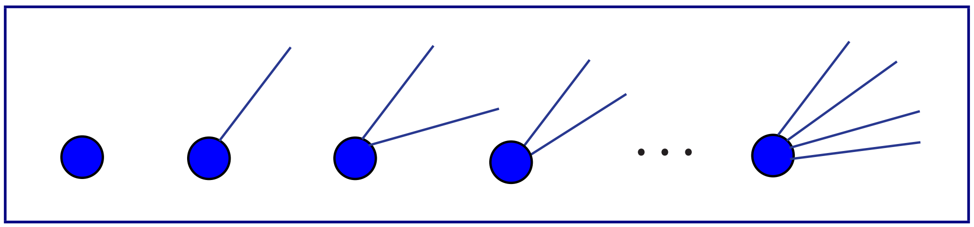
Social networks
- Immunology
- Information
- Commerce

- E-commerce $\rightarrow$ WWW $\rightarrow$ Internet $\rightarrow$ Power grid $\rightarrow$ River networks.
- Biological virus $\rightarrow$ Social contact network $\rightarrow$ Transportation networks $\rightarrow$ Communication networks $\rightarrow$ Power grid $\rightarrow$ River networks.

Modeling networks as random graphs

- Erdős and Rényi random graphs (1959, 1960). Phase transition.

- **Configuration models** (Bollobás 1980, Molloy & Reed *RSA* 1995).



Node degree is number of edges.

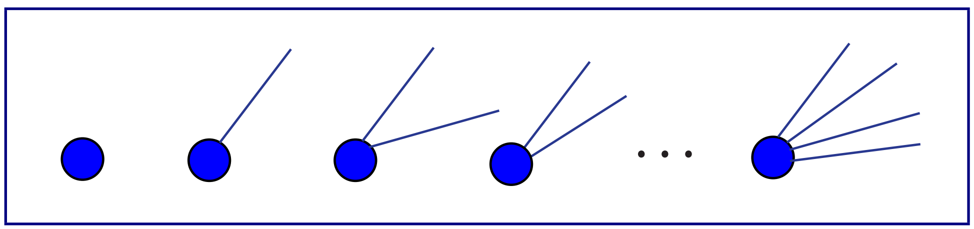
- Preferential attachment (Barbási-Albert 1999, etc.)

- Growth by copying (Kumar, Raghavan, Rajagopalan, Sivakumar, Tomkins, Upfal *FOCS* 2000), including duplication/mutation (Vazquez, Flammini, Maritan, Vespignani, *ComPlexUs* 2003)

- Random graphs analysis considers the **ensemble** of all graphs that can be constructed consistent with specified properties.

Configuration models

- (Bollobás 1980, Molloy & Reed *RSA* 1995).
- Enumerating over the **ensemble** of all networks with specified degree distribution. $\{p_k\}$ is fraction of nodes with degree k.
- **To generate an instance**: Begin with isolated nodes with half-edges and do a random matching. (Self-edges & multiple edges possible).



Node degrees sampled from p_k.

- **Probability generating functions** $G(x) = \sum_k p_k x^k$, allow us to calculate moments/properties of the ensemble.

c.f. Newman, Watts, Strogatz, "Random graphs with arbitrary degree distributions and their applications" *PRE* 2001.

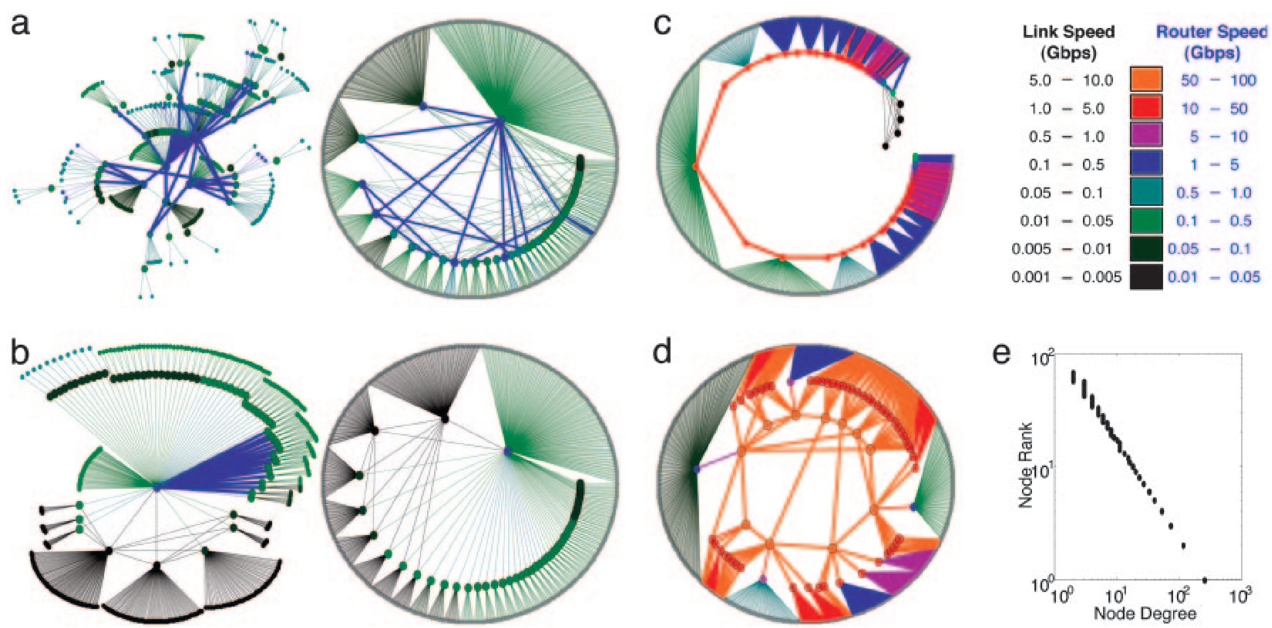
Does a random graph really model an individual engineered or biological system?

- **Ensemble (mean-field) not necessarily representative!**

Doyle, et. al., PNAS **102** (4)2005.

All these have

same deg dist, p_i:

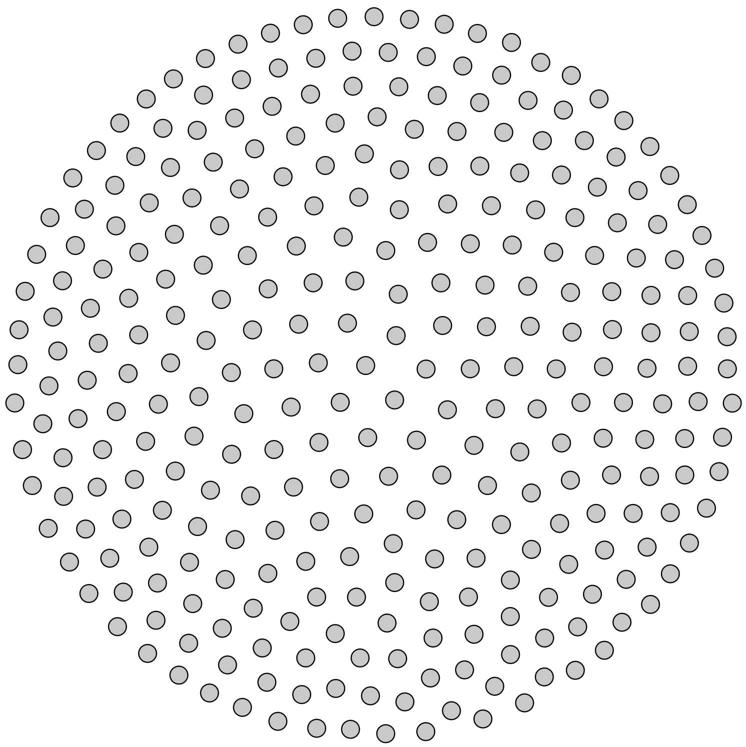


- Neglects design principles: Redundancy, degree correlations, local optimization (Although D'Souza, et. al. PNAS 2007), ...

- M. E. J. Newman *PRL* 103 (2009) – Augment degree by adding in small motifs (i.e., triangles). See also work by J. Gleeson.

The “classic” random graph, $G(n, p)$

- P. Erdős and A. Rényi, “On random graphs”, *Publ. Math. Debrecen.* 1959.
- P. Erdős and A. Rényi, “On the evolution of random graphs”, *Publ. Math. Inst. Hungar. Acad. Sci.* 1960.
- E. N. Gilbert, “Random graphs”, *Annals of Mathematical Statistics*, 1959.

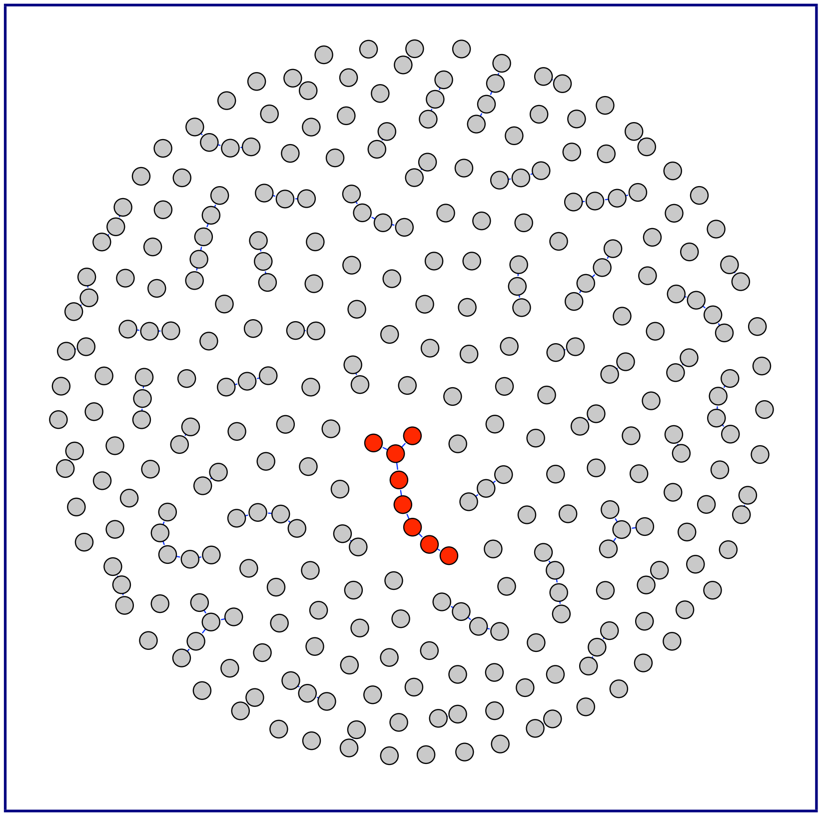


- Start with n isolated vertices.
- Consider each possible edge, and add it with probability p.

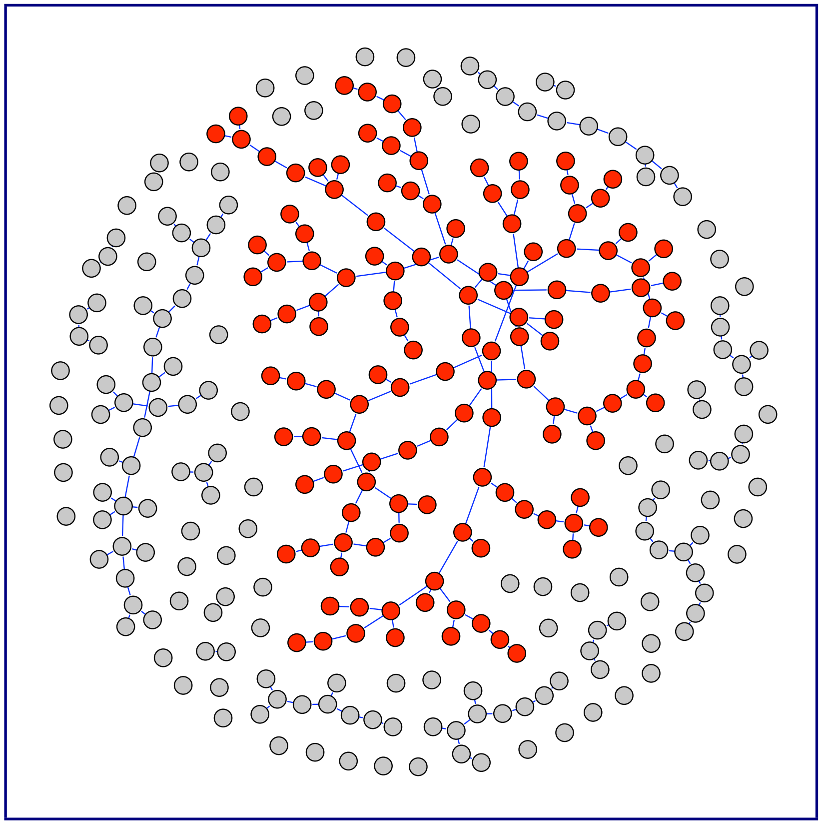
What does the resulting graph look like?

(Typical member of the ensemble)

G(n=300,p)

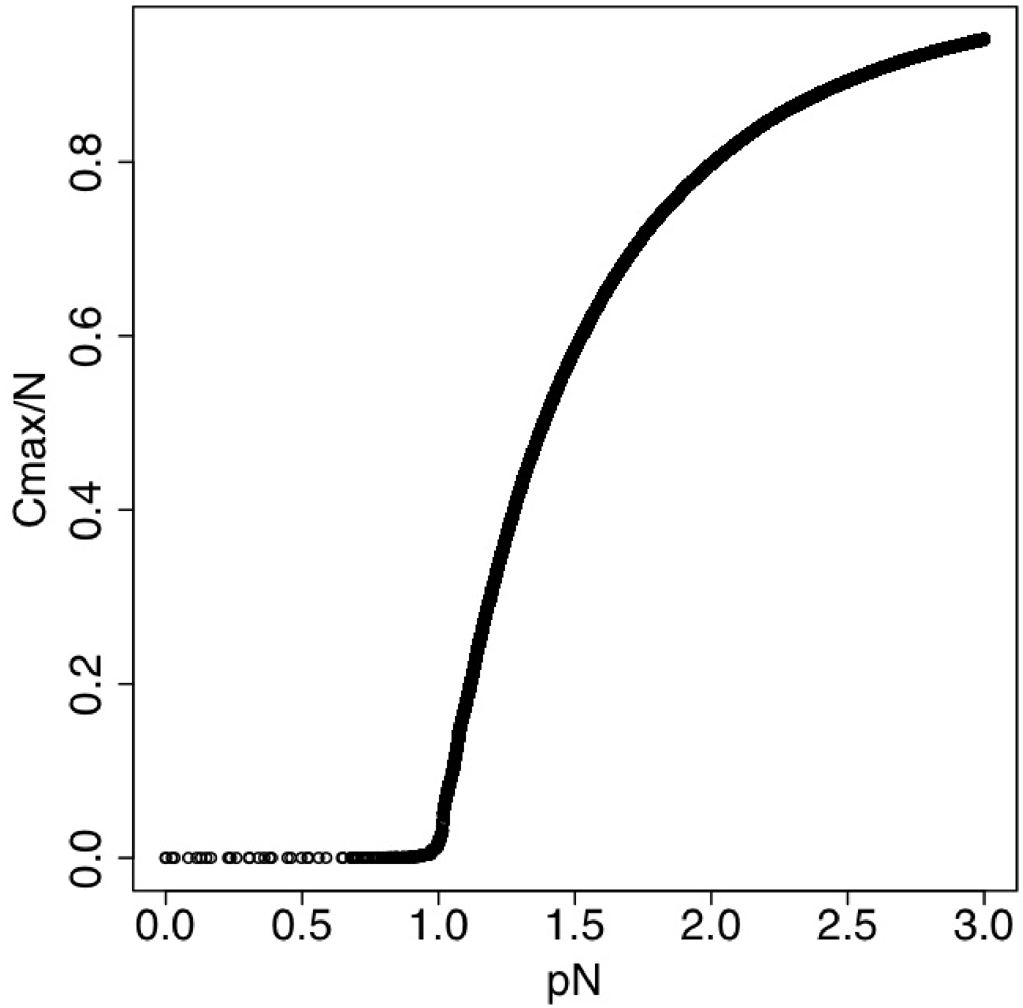


$p = 1/400 = 0.0025$



$p = 1/200 = 0.005$

Emergence of a *unique* "giant component"
Phase transition in connectivity



- $p_c = 1/n$.
- $p < p_c$, $C_{\max} \sim \log(n)$
- $p = p_c$, $C_{\max} \sim n^{2/3}$
- $p > p_c$, $C_{\max} \sim A \cdot n$

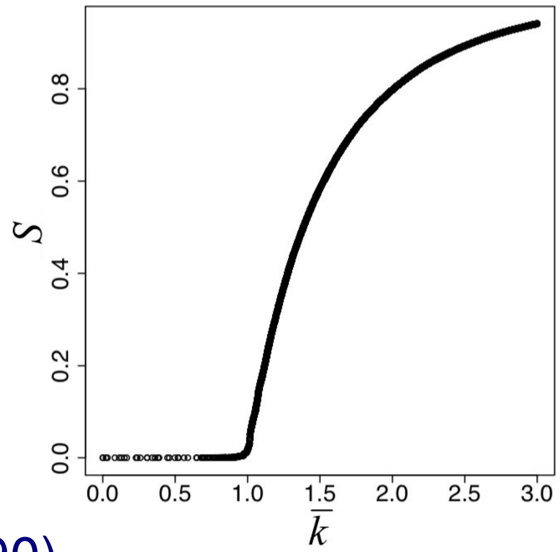
Expected # of edges per node

$$t = e/n = p(n-1)/2$$

so $t_c = 1/2$

Erdős-Rényi: unique "giant component"

- $t < 1/2, \quad C_{\max} \sim \mathrm{O}(\ln n)$
- $t = 1/2, \quad C_{\max} = n^{2/3}$
- $t > 1/2, \quad C_{\max} \sim An,$ with $A > 1$



- The critical window

  Bollobás, *Trans. Amer. Math. Soc.*, **286** (1984).

  Luczak, *Random Structures and Algorithms*, **1** (1990).

$$t = 1 + \lambda n^{-1/3} \qquad \text{(where } t = 2e/n\text{)}$$

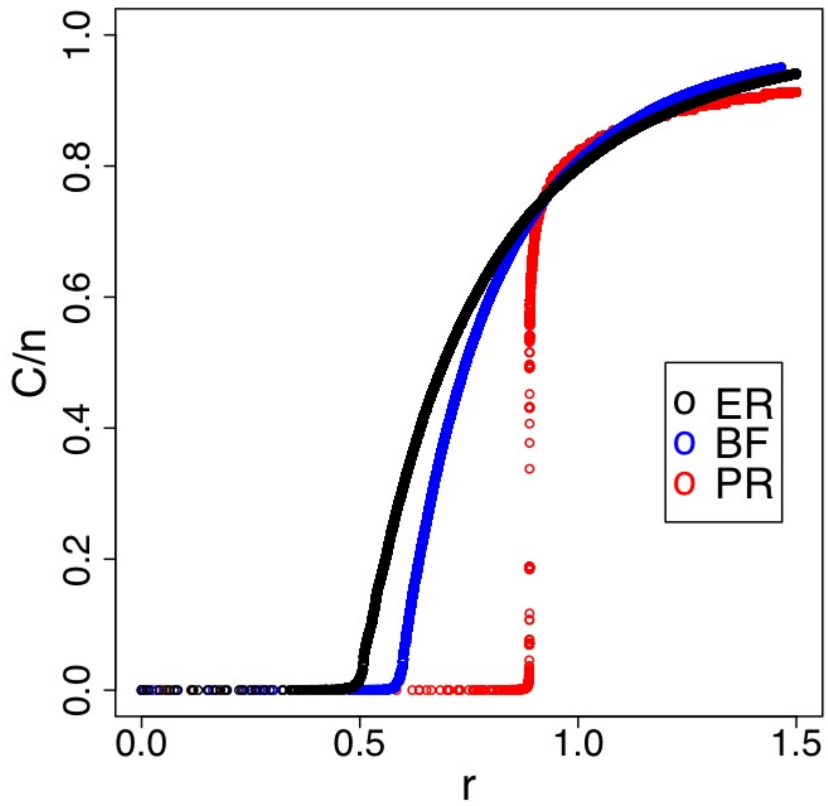
- Mean field critical exponents

  e.g., Grimmett, *Percolation*. 2nd Edition. Springer-Verlag. 1999.

$$\chi \sim (t_c - t)^{-\gamma}, \quad \text{with } \gamma = 1.$$

where χ is the expected size of the component to which an arbitrarily chosen vertex belongs.

Is connectivity a good thing?
(Context dependence)



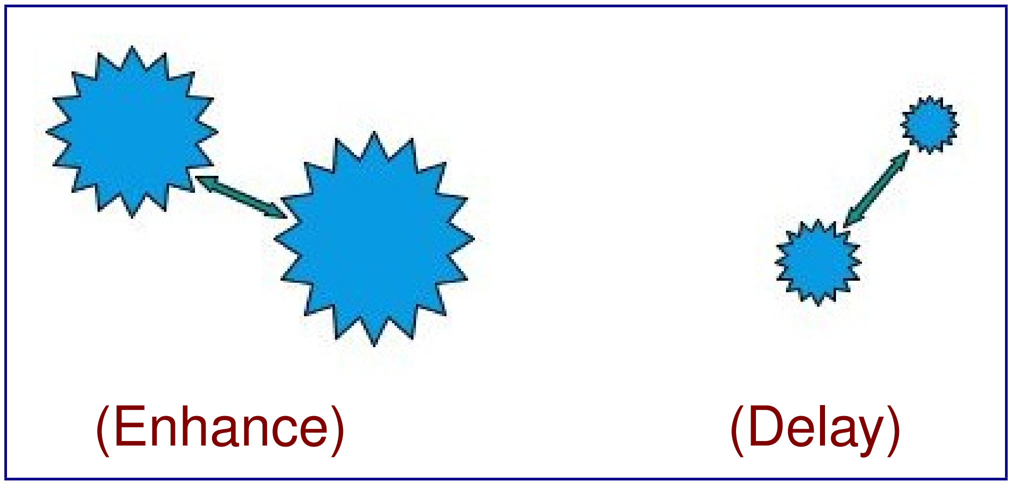
- Communications, Transportation, Synchronization, ... versus
- Spread of human or computer viruses

Can any **limited perturbation** change the phase transition?

[Bohman, Frieze, *RSA* **19**, 2001]

[Achlioptas, D'Souza, Spencer, *Science* **323**, 2009]

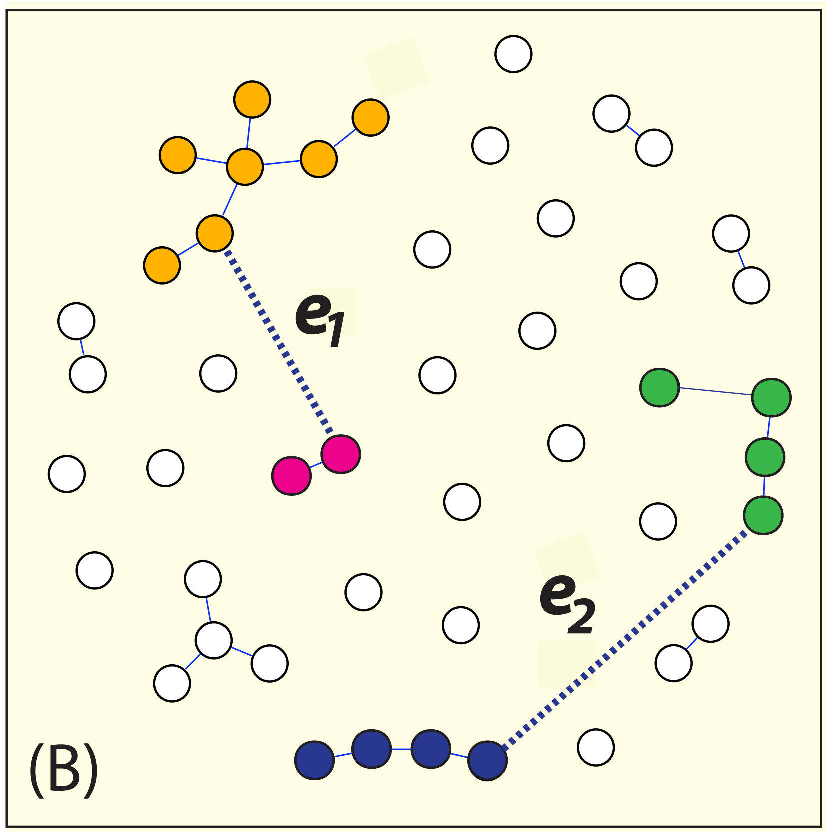
- Possible to **Enhance** or **Delay** the onset?
- The "**Product Rule**"
  - Choose ***two*** edges at random each step.
  - Add only the desirable edge and discard the other.



- The Power of Two Choices in randomized algorithms.

  Azar; Broder; Mitzenmacher; Upfal; Karlin;

ProdRule: Explicit example

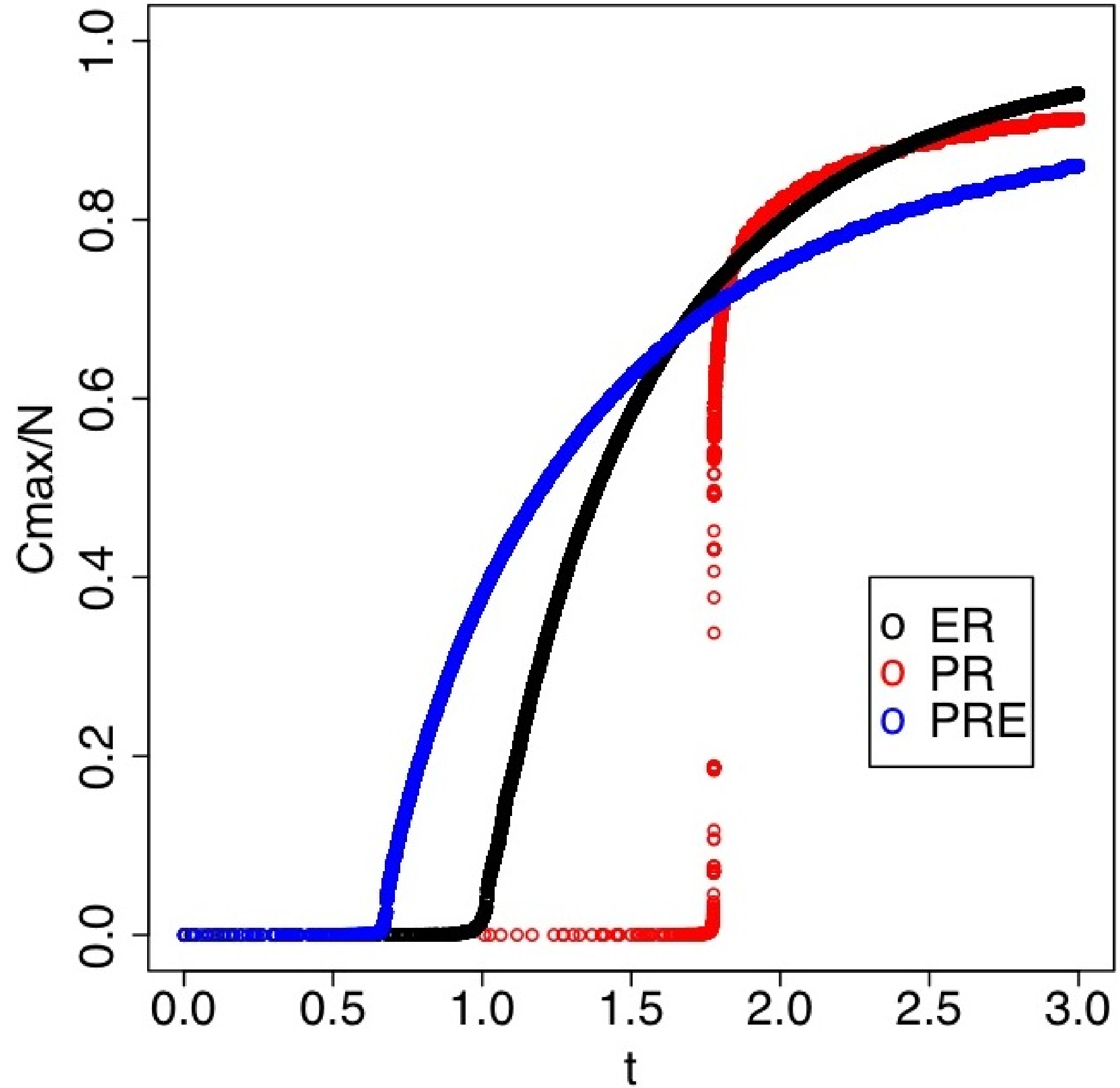


- Prod $e_1 = (7) \times (2) = 14$
- Prod $e_2 = (4) \times (4) = 16$
- To *enhance* choose e_2. To *delay* choose e_1.

Product Rule

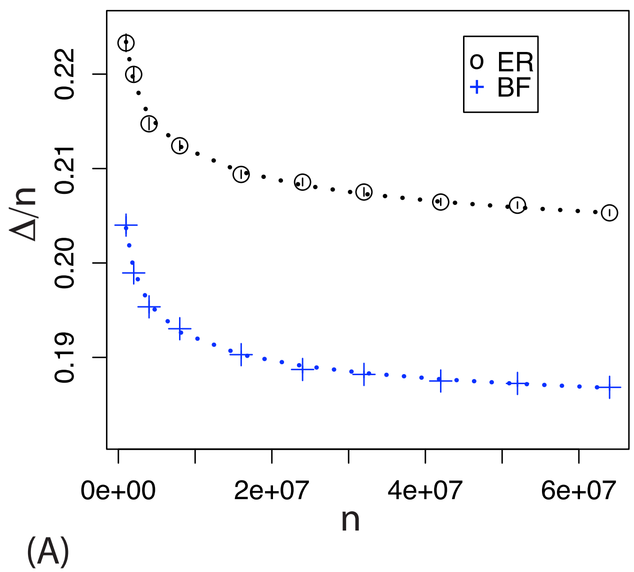
- **Enhance** – similar to **ER** but with earlier onset.

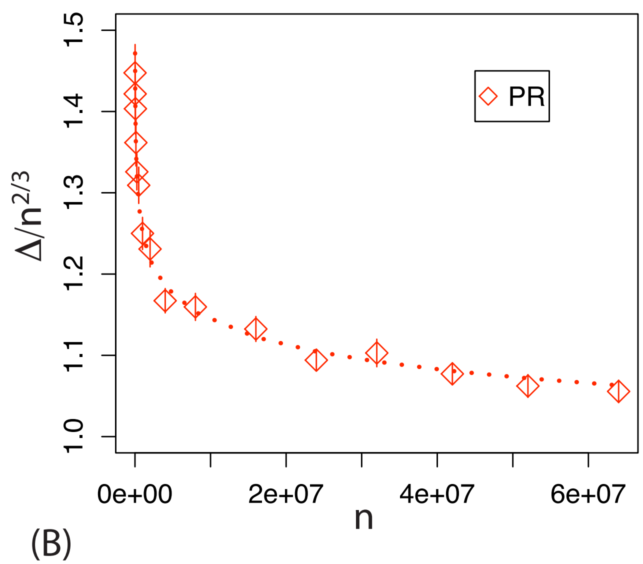
- **Delay** – Extremely abrupt



The scaling window, Δ from $n^{1/2}$ to $0.5n$

- Let e_0 denote the last edge added for which $C_{max} < n^{1/2}$. (Recall ER has $n^{2/3}$ at t_c.)
- Let e_1 denote the first edge added for which $C_{max} > 0.5n$.
- Let $\Delta = e_1 - e_0$.





ER (and BF) $\Delta \sim n$

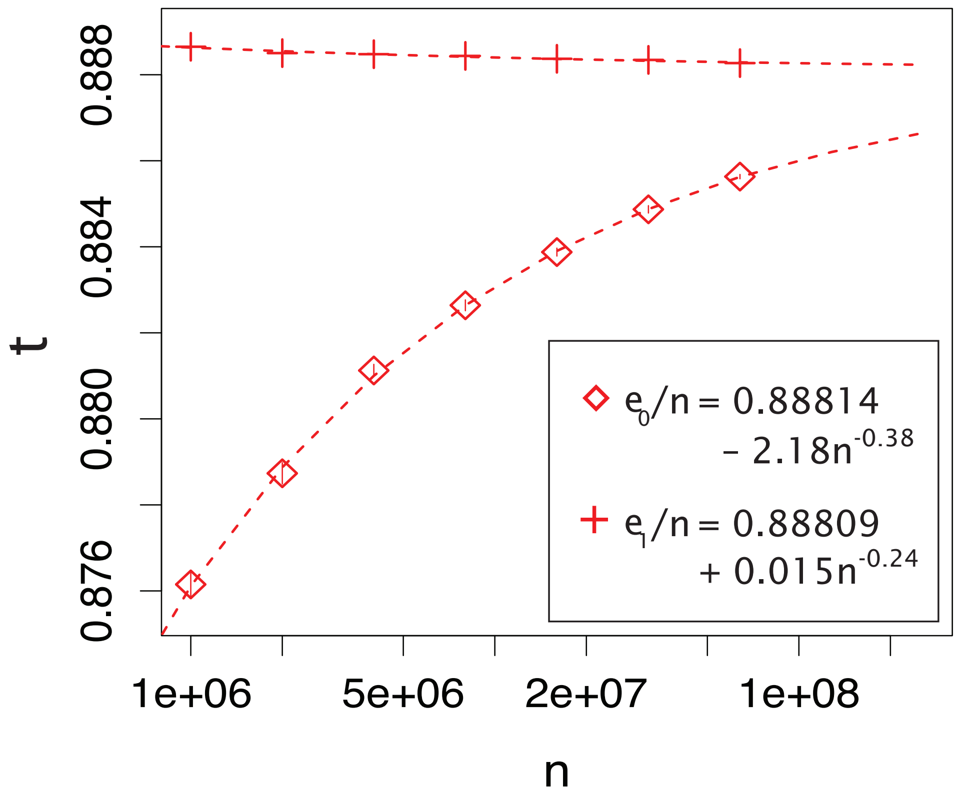
PR $\Delta \sim n^{2/3}$.

PR From $n^{1/2}$ to $0.5n$ in number of edges that is sublinear in n.

In terms of edge density or "time", t_c, where $t = e/n$

(Note, for ER, $t_c = 1/2$)

- For $t < t_c$, $C_{\max} < n^{1/2}$.
- For $t > t_c, C_{\max} > 0.5n$.



Jumps "instantaneously" from $C_{\max}$ = $n^{1/2}$ to $0.5n$.

Why this is surprising

Percolation theory on networks and lattices serves as a theoretical underpinning for :

- Onset of epidemic spreading
- Flow through porous media / random transport
- Vulnerability and resilience of networks

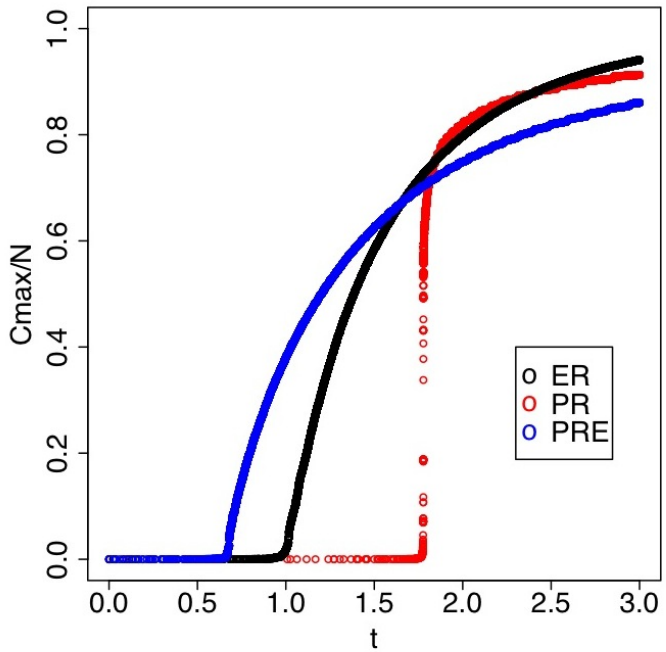
– Many prior variants (bond, site, directed, ...) on many types of networks and lattices; All continuous transitions.

– Continuous phase transitions are accompanied by *critical scaling* which can provide warning signs.

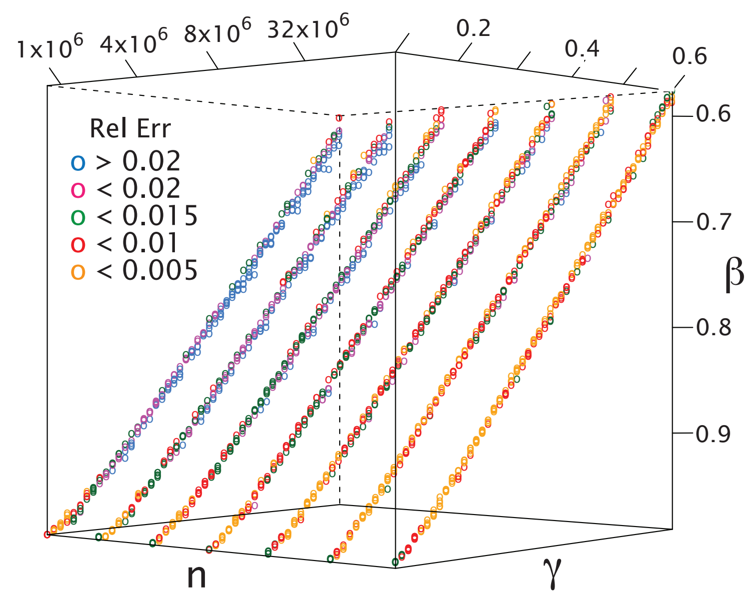
"Explosive Percolation in Random Networks"

From n^γ to greater than $0.6n$ "instantaneously"
(Compelling evidence that the transition is discontinuous)

$C_{\max}$ jumps from sublinear n^γ to $\geq 0.5n$ in n^β edges, with $\beta, \gamma < 1$.



Nontrivial Scaling behaviors
$\gamma + 1.2\beta = 1.3$ for $A \in [0.1, 0.6]$



Achlioptas, D'Souza, Spencer, *Science*, **323** (5920), 2009

Many more EP systems and mechanisms now discovered

(Condensed list here)

Lattice percolation, power law graphs, cluster aggregation:

- R. Ziff, *Phys. Rev. Lett.* 103, 045701 (2009).
- Y. S. Cho, J. S. Kim, J. Park, B. Kahng, D. Kim, *Phys. Rev. Lett.* 103, 135702 (2009).
- F. Radicchi, S. Fortunato, *Phys. Rev. Lett.* 103, 168701 (2009).
- E. J. Friedman, A. S. Landsberg, *Phys. Rev. Lett.* 103, 255701 (2009).
- Y.S. Cho, B. Kahng, D. Kim, *Phys. Rev. E* (R), 2010.
- R. M. D'Souza, M. Mitzenmacher, *Phys. Rev. Lett.* 104, 195702 (2010).
- Araújo, Andrade Jr, Ziff, Herrmann, *Phys. Rev. Lett.* 106, 095703 (2011).
- Hooyberghs, Van Schaeybroeck, Phys. Rev. E 83, 032101 (2011).
- Gomez-Gardenes, Gomez, Arenas, Moreno, *Phys. Rev. Lett. in press*.

Observed in real world:

- Rozenfeld, Gallos, Makse; *Eur. Phys. J. B*, 75, 305-310, (2010). (PHN)
- Pan, Kivelä, Jari Saramäki, Kaski, Kertész, Phys. Rev. E 83, (2011). (Communities)
- Y. Kim, Y.-k. Yun, and S.-H. Yook, Phys. Rev. E 82, 061105 (2010). (Nanotubes)
- Growth of Wikipedias (Bounova, personal communication.)

Alternate mechanisms (with out competition):

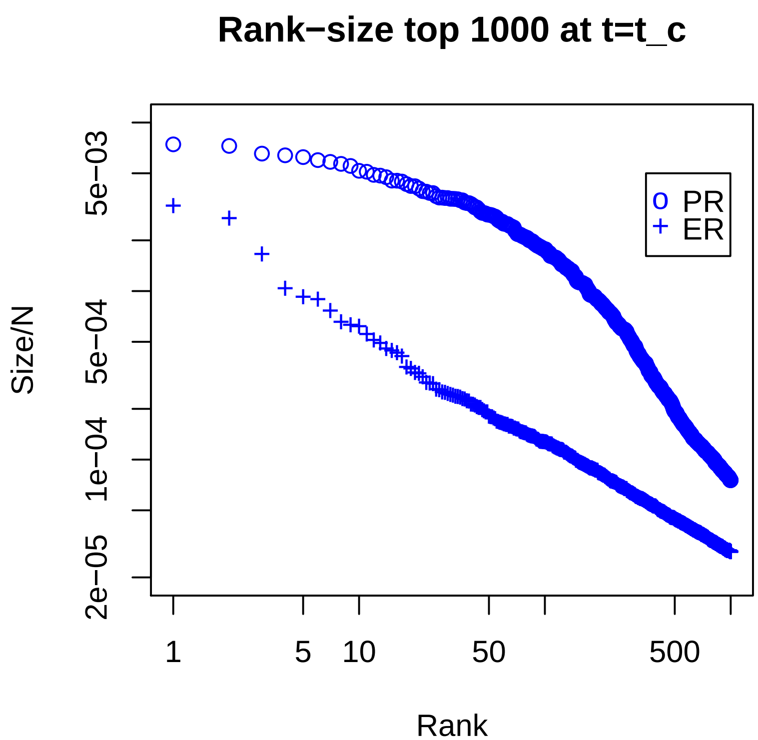
- Araújo, Herrmann, *Phys. Rev. Lett.* 105, 035701 (2010).
- W. Chen, R. M. D'Souza, *Phys. Rev. Lett.* 106, 115701 (2011).

Beyond "Product Rule": Models with fixed choice

- "**Achlioptas process**": examine fixed number of edges, add the one that optimizes a pre-set criterion.

  "Sum rule", Adjacent edge, Triangle rule, k-clique rule, etc., all also work.

- **Novel subcritical behavior : components are similar in size; many almost linear size components**



- **Applications**: Community detection, Minimizing interference in wireless networks, Wikipedia growth....

"Explosive Percolation": Some caveats

- "Weakly discontinuous" :

  $\Delta C_{\max}$, the biggest change in C_1 due to **addition of a single edge**, decays with system size. (Nagler, et. al, *Nature Physics*, 2011).

- In limit $n \to \infty$, **fixed choice rules** are continuous!

  – da Costa, Dorogovtsev, Goltsev, Mendes, *Phys. Rev. Lett.* 105, (2010).

  – Riordan and Warnke, *Science* 333, (2011).

- Infinite choice : if number of choices $k \to \infty$ as number of nodes $n \to \infty$, this is sufficient for discontinuous transition.

  e.g. $k = \log(n)$.

- As $n \to \infty$, jump $\Delta C_{\max} \to 0$, but for $n \sim 10^{18}$, $\Delta C_{\max}$ can be of size 0.1n.

  The $n \to \infty$ limit is not the regime of real-world networks.

  e.g., social networks $n \le 10^{10}$

Percolation as cluster aggregation models

- Excellent review on percolation as cluster aggregation:

  D. J. Aldous, “Deterministic and stochastic models for coalescence (aggregation and coagulation): A review of the mean-field theory for probabilists”, *Bernoulli*, 5(1): 348, 1999.

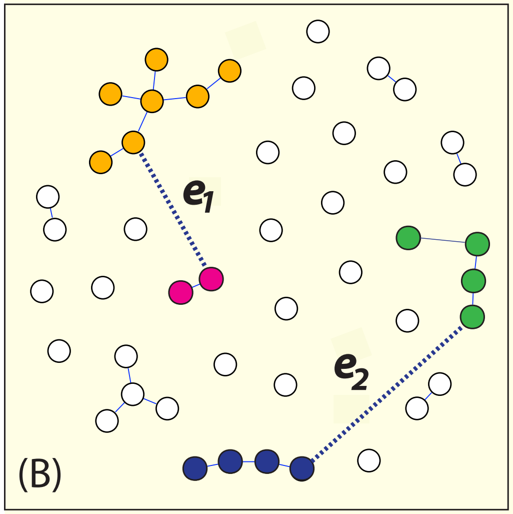
  (Scientific Modeling (SM) mathematics rather than Theorem-Proof (TP) mathematics.)

- Assume each edge merges two previously distinct components, with probability of connecting a component of size x and one of size y, proportional to **kernel** $K(x, y)$.

$K(x, y) = 1$ uniform attachment / size independent

$K(x, y) = xy$ “gravitational attraction” / this is Erdős-Rényi.

$$(F_{\text{gravity}} = -M_1 M_2 / r_{12}^2)$$



Smoluchowski family of coagulation equations

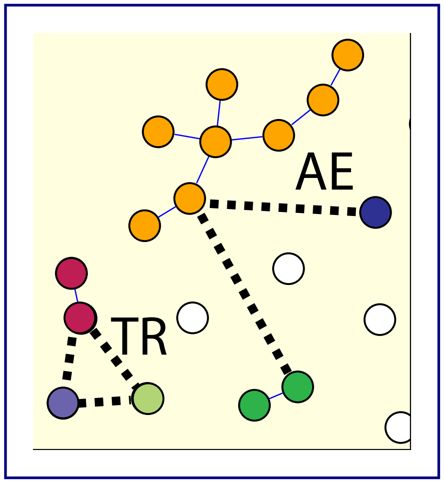
- Given kernel $K(x, y)$
- Evolution of $n(x, t)$, the expected number of clusters of size x at time t.
- Mean-field over all graphs (ensemble properties)

$$\frac{d}{dt}n(x,t) = \frac{1}{2}\sum_{y=1}^{x-1} K(y, x-y)n(y,t)n(x-y,t) - n(x,t)\sum_{y=1}^{\infty} K(x,y)n(y,t)$$

Smoluchowski approach to “Explosive Percolation”

- Y.S. Cho, B. Kahng, D. Kim; *Phys. Rev. E* 81, 030103(R), 2010. “Cluster aggregation model for discontinuous percolation transition”

- R.D. and M. Mitzenmacher, “Local cluster aggregation models of explosive percolation”, *Phys. Rev. Lett.*, 104, 2010.



Adjacent edge: Let $x_i = in(i, t)$ (fraction of nodes)

$$\frac{dx_i}{dt} = -ix_i - i(2x_iS_i - x_i^2) + i\sum_{j+k=i} x_j(2x_kS_k - x_k^2)$$

- S. S. Manna and Arnab Chatterjee “A new route to Explosive Percolation”, Physica A 390, 177182 (2011).

- R. A. da Costa, S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, “‘Explosive Percolation’ Transition is Actually Continuous”, Phys. Rev. Lett. 105, 255701 (2010).

da Costa, et al *PRL* 2010

- Define $P(s,t) = sn(s,t)/\langle s \rangle$, distribution of finite component sizes to which a randomly chosen vertex belongs.

- Use a (mean-field) Smoluchowski-type eqn:

$$\frac{\partial P(s,t)}{\partial t} = s \sum_{u+v=s} Q(u,t)Q(v,t) - 2sQ(s,t)$$

- Size of largest component, $S(t) = 1 - \sum_i P(s,t) \cong 1 - \sum_{i=1}^{10^6} P(s,t)$.

- If assume $P(s, t_c)$ is distributed according to a power law, obtain the main result: critical behavior, $S(t) \sim (t - t_c)^\beta$, with $\beta = 0.0555 \approx 1/18$.

- Jump: $\Delta S = S(t_c^+) - S(t_c) = S(t_c^+) - o(n) \sim (t_c^+ - t_c)^\beta = (1/n)^\beta$

- If $n = 10^{18}$, jump = 0.1 n ... ten percent of system!

Are any real social or technological networks of size $n \sim 10^{18}$?

Riordan and Warnke, *Science* 2011

- Rigorous proof: Any *fixed choice* process ultimately continuous!

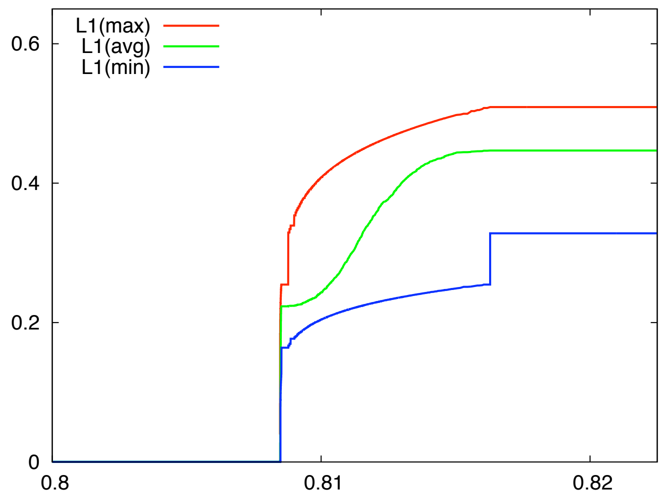
- Proof by contradiction. ("The vanishing 'powder keg'")

- Δ, the scaling window from our PR simulations, will ultimately crossover to linear in n, but no estimate of crossover length from these arguments.

- Moreover, AP's can be nonconvergent (no scaling limit). (arXiv.1111.6177)

  Typically **assume** $\lim_{n\to\infty} C_1 = A(t)n$ once $t > t_c$

  (That there is a function $A(t)$ that describes the growth of C_1 in the supercritical regime.)



translated into physics terminology:

"Achlioptas processes are not always self-averaging", to appear *PRE*

Beyond choice and competition:
Discontinuous percolation other mechanisms

- **Control only of the largest cluster**

  – Araujo, N. A. M. & Herrmann, H. J. Explosive percolation via control of the largest cluster. Phys. Rev. Lett. 105, 035701 (2010).

  – Araujo, et. al. **Tricritical point** in explosive percolation. Phys. Rev. Lett. 106, (2011). (‘tri-critical” points separate region of 1st order (discontinuous) from 2nd order (continuous) transitions).

  – W. Chen and R.D. *Phys. Rev. Lett.* 83 (2011).

- **Cooperative phenomena**

  – Bhizani, Paczuski, Grassberger ”Discontinuous percolation transitions in epidemic processes, surface deppining in random media and Hamiltonian graphs”. in press PRE

- **Correlated percolation**

  – L. Cao, J. M. Schwarz, ”Correlated percolation and tricriticality”, arXiv:1206.1028

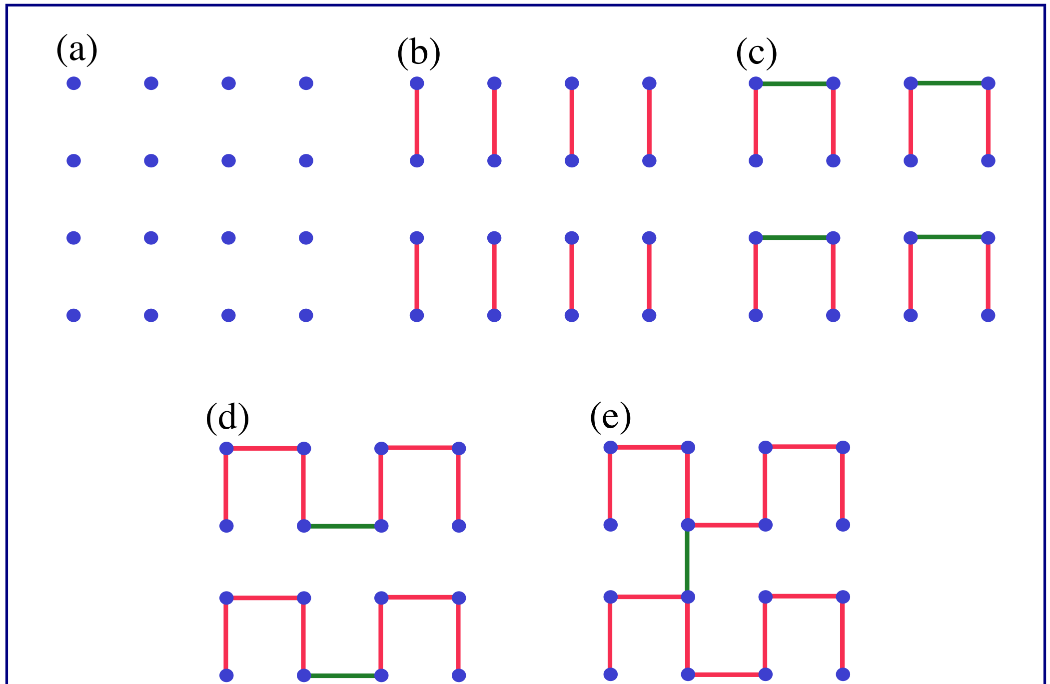
- **Dressing up a simple structure** (one-dim lattice with hierarchy of long-range bonds) Boettcher, Singh, Ziff, *Nature Communications*, 3:787 (2012).

- **Restricted Erdős-Rényi**: Choose one node at random, one from restricted set. Panagiotou, et. al. *Elec. Notes. Disc. Math.* 2011.

A deterministic model

Friedman, Landsberg *PRL* (2009); Rozenfeld, et. al. *EPJB* (2010);
Nagler, Levina, Timme, *Nature Phys.* (2011)

- (a) Phase $k = 2$, merge all isolated nodes into pairs.
- (b) Phase $k = 4$, merge pairs into size 4 components.
- (c) Phase $k = 8$, merge pairs of 4's into 8's.
- etc.



(From Rozenfeld, et. al.)

- At edge $e = n$ (time $t = 1$) one giant of size n emerges

(Giant emerges when only one component remains)

Re-visiting the Bohman Frieze Wormald model (BFW)

(*Random Structures & Algorithms*, 25(4):432-449, (2004))

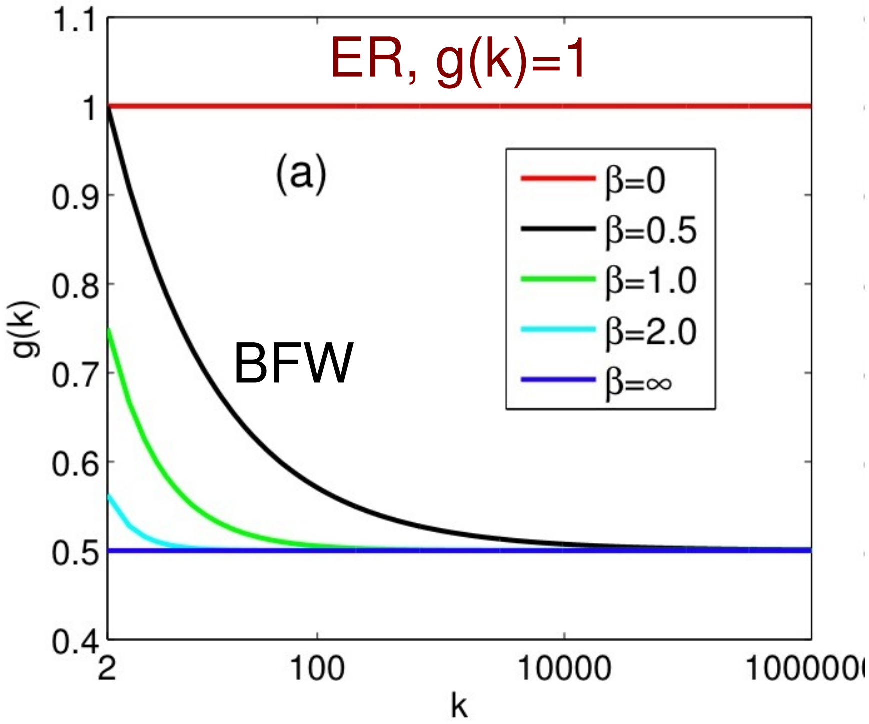
- A **stochastic model**, which exams a **single-edge at a time**.
- Like deterministic, start with n isolated vertices, and stage $k = 2$.
- Sample edges uniformly at random from the complete graph on n nodes.
- Can *reject* edges provided the fraction of accepted remains greater than a function decaying with phase k. Let:

u be number of edges sampled,
t be the number accepted:

Fraction of accepted edges,

$$t/u \geq g(k) = 1/2 + (2k)^{-1/2}$$

(Note: $\lim_{k\to\infty} g(k) \to 1/2$)

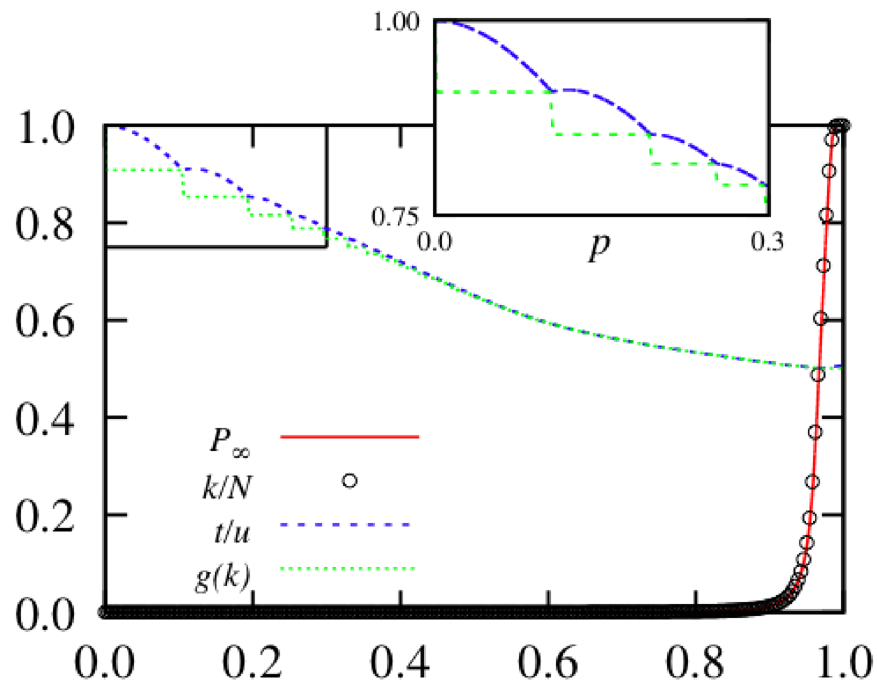


The BFW model

- Start with n isolated vertices, and cap on maximum component set to $k = 2$.
- Examine an edge selected uniformly at random from the complete graph:

1. If the resulting component size $\leq k$, accept the edge.
2. Otherwise reject that edge if possible (meaning the fraction of accepted edges $t/u \geq g(k)$).
3. Else augment $k \to k + 1$, and repeat (1) and (2), with (3) if necessary.

   (Step 3 executes for “troubling edge”)



When troubling edge encountered, $k \to k + 1$ until either:
– The edge can be rejected due to sufficient decrease of $g(k)$
– The edge can be accepted due to k large enough.

The BFW model stated formally

- Initially n isolated nodes with cap on maximum size set to $k = 2$.
- Let u denote the total number of edges sampled
- A the set of accepted edges (initially $A = \emptyset$)
- $t = |A|$ the number of accepted edges.

At each step u, select edge e_u uniformly at random from complete graph, and apply the following loop:

Set $l =$ maximum size component in $A \cup \{e_u\}$

  if $(l \leq k)\,\{$

    $A \leftarrow A \cup \{e_u\}$

    $u \leftarrow u + 1\ \}$

  else if $(t/u < g(k))\,\{\ k \leftarrow k + 1\ \}$

  else $\{\ u \leftarrow u + 1\ \}$

• If the edge e_u is troubling and $t/u < g(k)$, augment k repeatedly until either:

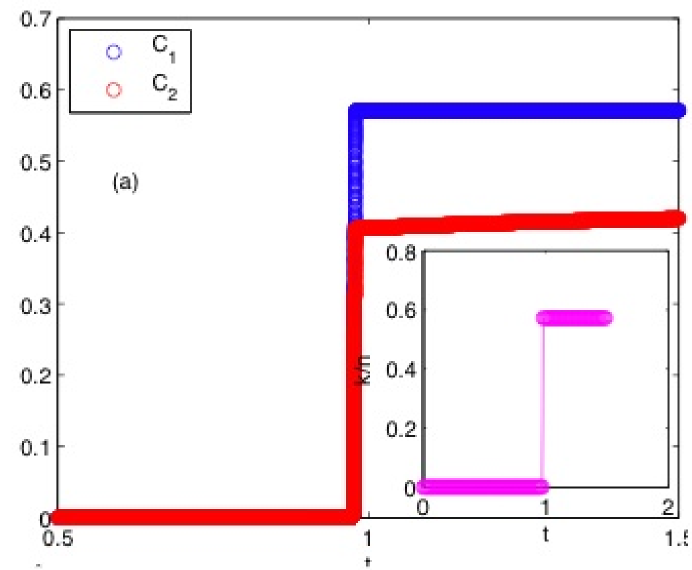
(i) k increases sufficiently that e_u is accepted or

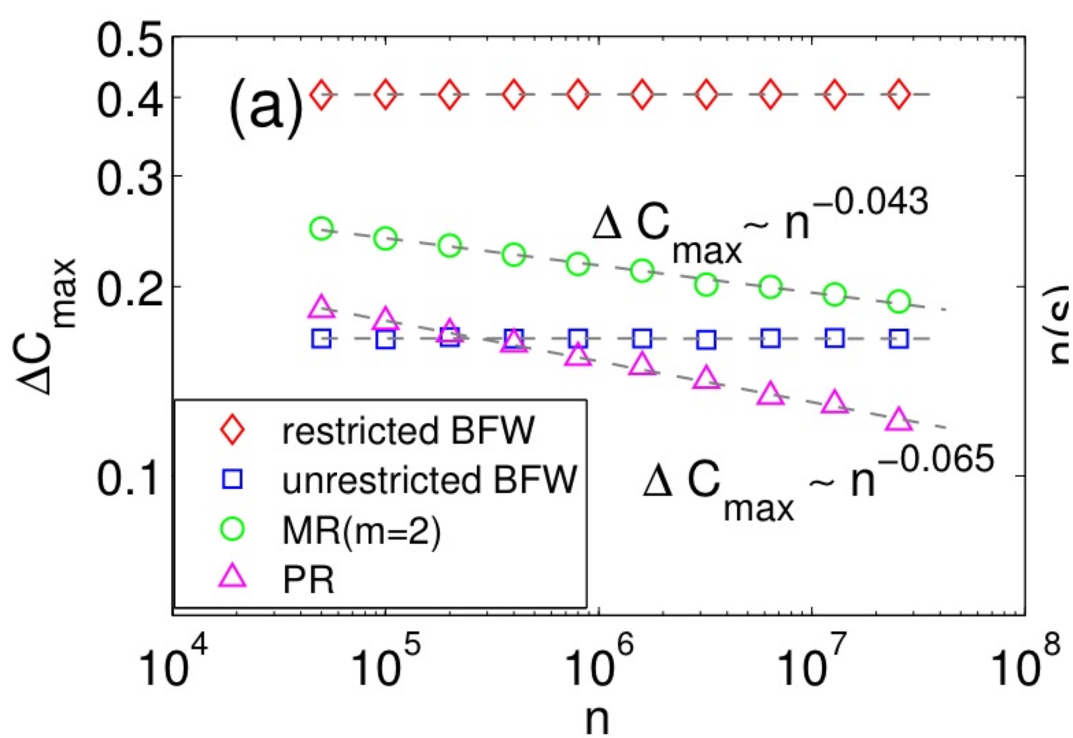
(ii) $g(k)$ decreases sufficiently that e_u is rejected.

Simultaneous emergence of multiple stable giants in a strongly discontinuous transition

(Wei Chen and R.D. *Phys. Rev. Lett.* 83 (2011).)

- Two stable giants!
  ($C_1 = 0.570, C_2 = 0.405.$)
  - Fraction of internal cluster edges $> 1/2$.
  - (If restrict to sampling only edges that span clusters, only one giant ultimately.)





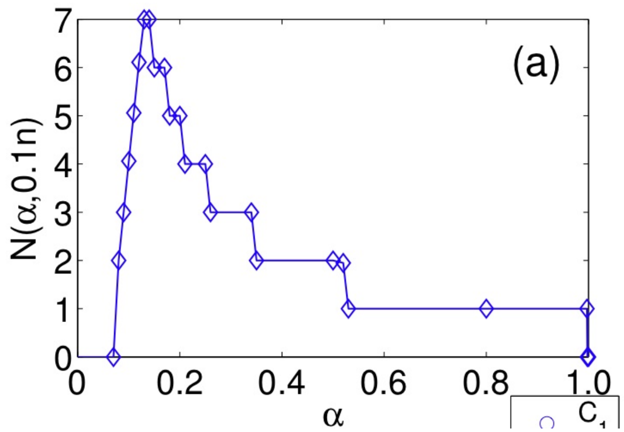
"Strongly" discontinuous (gap independent of n)

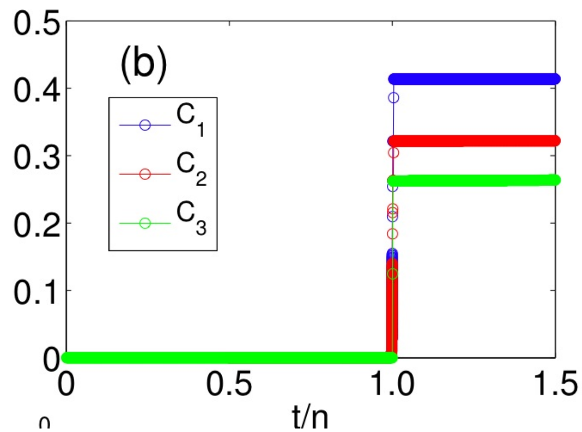
$\Delta C_1 \approx 0.165$

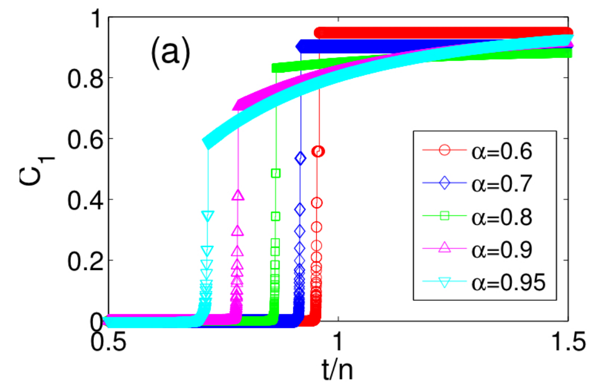
Tuning the number of stable giants

(Wei Chen and R.D. *Phys. Rev. Lett.* 83 (2011).)

- Now let $g(k) = \alpha + (2k)^{-1/2}$. Smaller α more edges can be rejected. α determines number of stable giants!





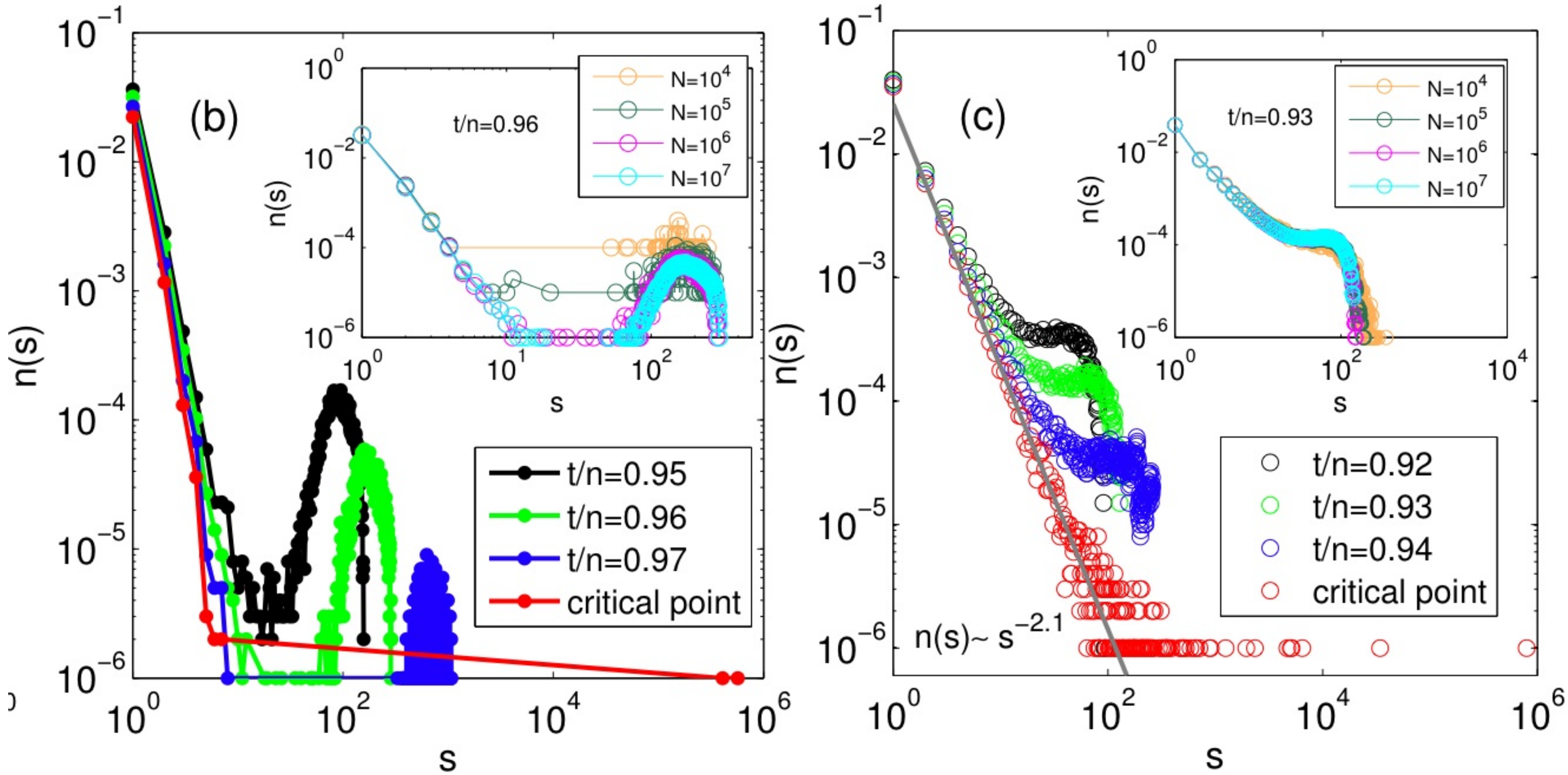


- Multiple stable giants, not anticipated. ("uniqueness of the giant component" / gravitational coalescence of Smoluchowski kernel $K(x, y) = xy$)

- Applications for multiple giants? (Communications, epidemiology, building blocks for modular networks, polymerization (Krapivsky, Ben-Naim)...)

Evolution of component density for BFW

$\beta = 0.5$ $\beta = 2.0$



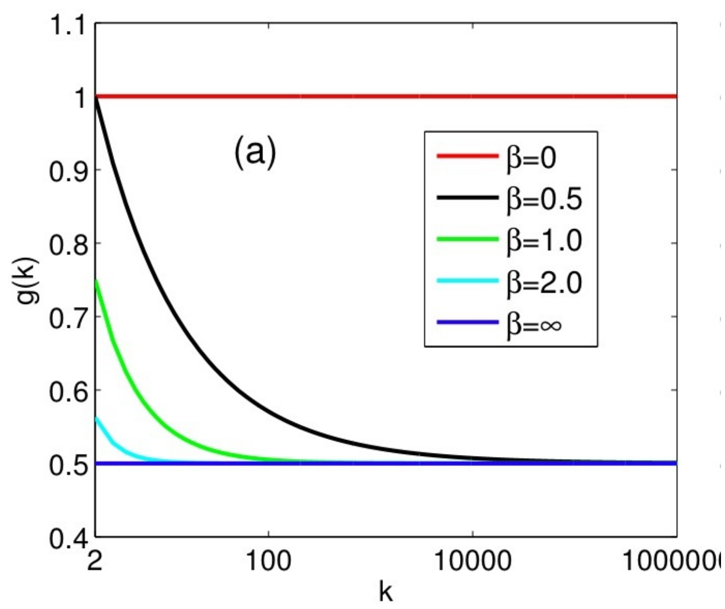
- For $\beta = 0.5$ no scaling. Separates into components of size $O(n)$ and $< \log(n)$.
- For $\beta = 0.5$ and $\beta = 2.0$ no finite size effects in the location of the "hump" (inset), unlike for PR where location depends on n. (c.f. Lee, Kim, Park: data collapse)
- No scaling, no "early warning signs" (Scheffer, et. al. *Nature* (2009).

Deriving the underlying mechanism:
Slow decay of $g(k)$ leads to growth by overtaking

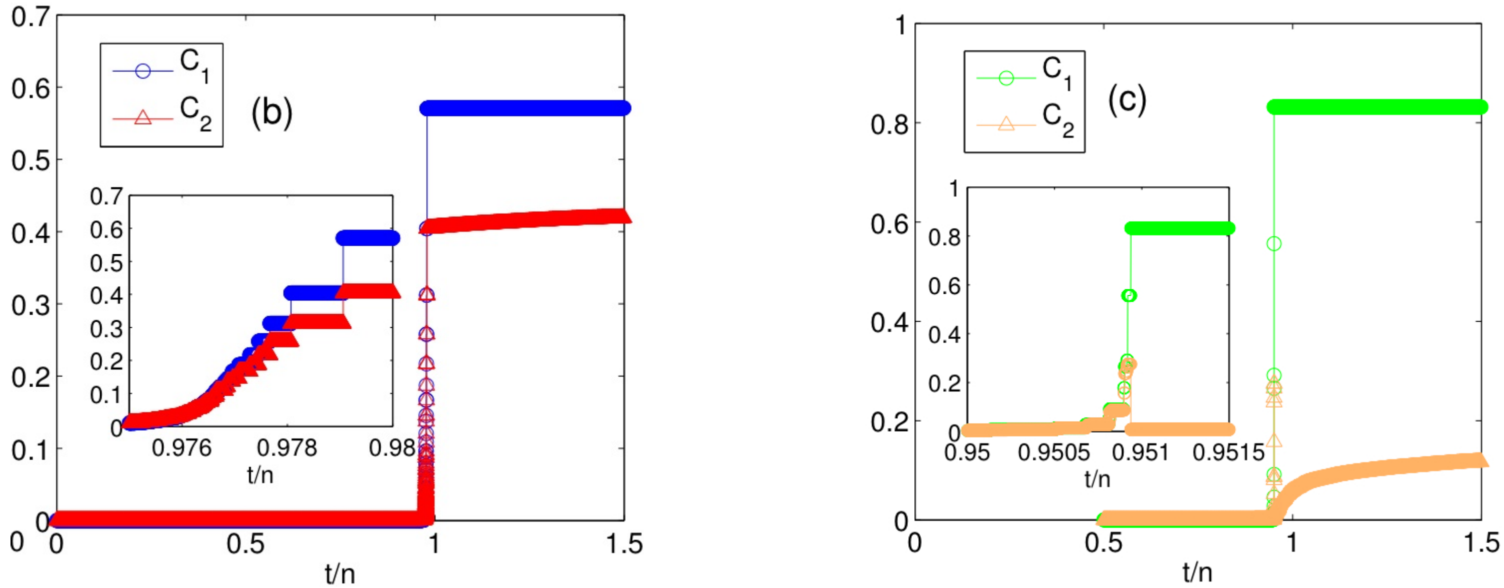
(Wei Chen and R.D, arXiv:1106.2088)

- Instead of $g(k) = 1/2 + (2k)^{-1/2}$ now let $g(k) = 1/2 + (2k)^{-\beta}$

- Procedure: analyze by how much k must grow before $g(k)$ would decrease sufficiently to reject troubling edge.



- For $\beta \in (0.5, 1]$, an increase in $k \sim n^{\beta}$ is always sufficient to reject a troubling edge. Slow increase in k means:
  - Growth by overtaking*: two smaller components merge becoming new C_1.
  - Multiple components of size $O(n)$ before the largest jump.

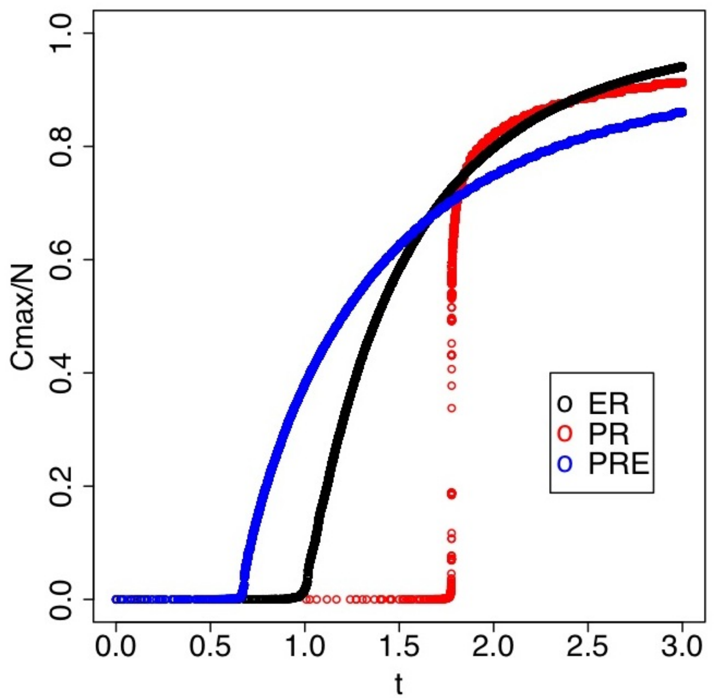


- For $\beta > 1$, once stage $k = n^{1/\beta}$, troubling edges **must** be accepted at times, leading to large direct growth of C_1, and a weakly discontinuous transition.

* Consistent with Nagler, et. al., *Nature Phys* (2011), for direct growth forbidden.

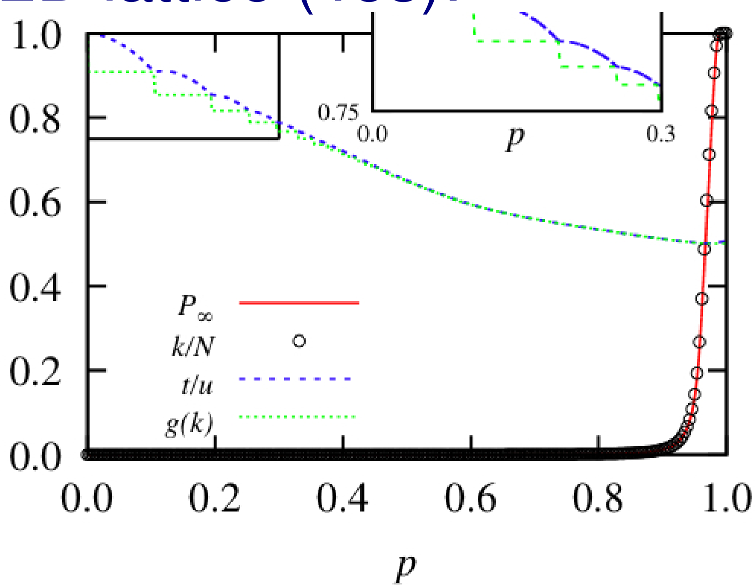
More generally, macroscopic jump means: Multiple giants coexist in critical window

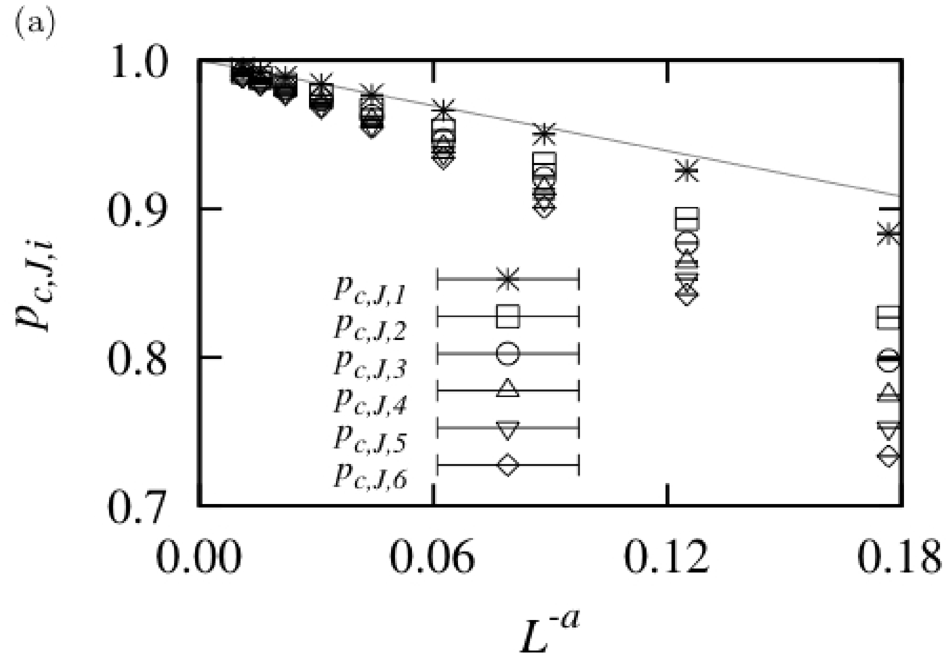
- Note, we define as the critical point t_c, the single edge who's addition causes the biggest change, ΔC_1. (Recall C_1 is the *fraction* of nodes in the largest component.)
- If $\Delta C_1 > 0$ there necessarily existed another macroscopic component. e.g. If $\Delta C_1 = 0.1$ that means C_1 merged with a component of size $|C_j| = 0.1n$.
- Let t'_c denote emergence of giant.
- Let t_c denote largest jump in C_1
- Is $t_c = t'_c$??



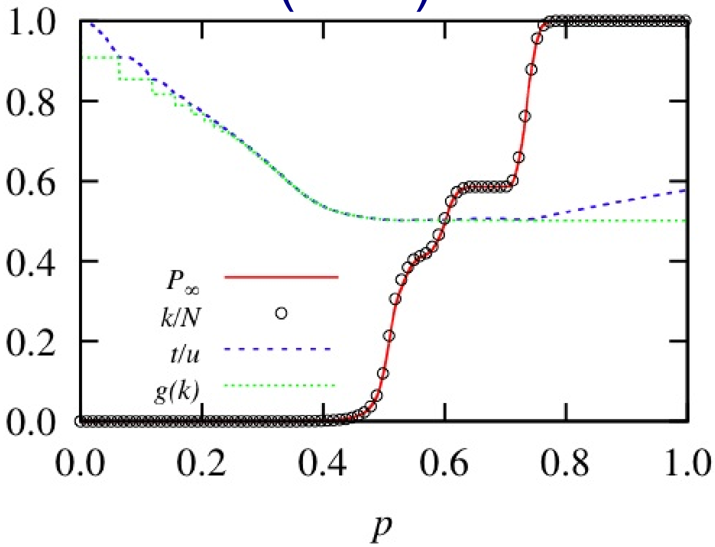
Is $t_c = t'_c$?

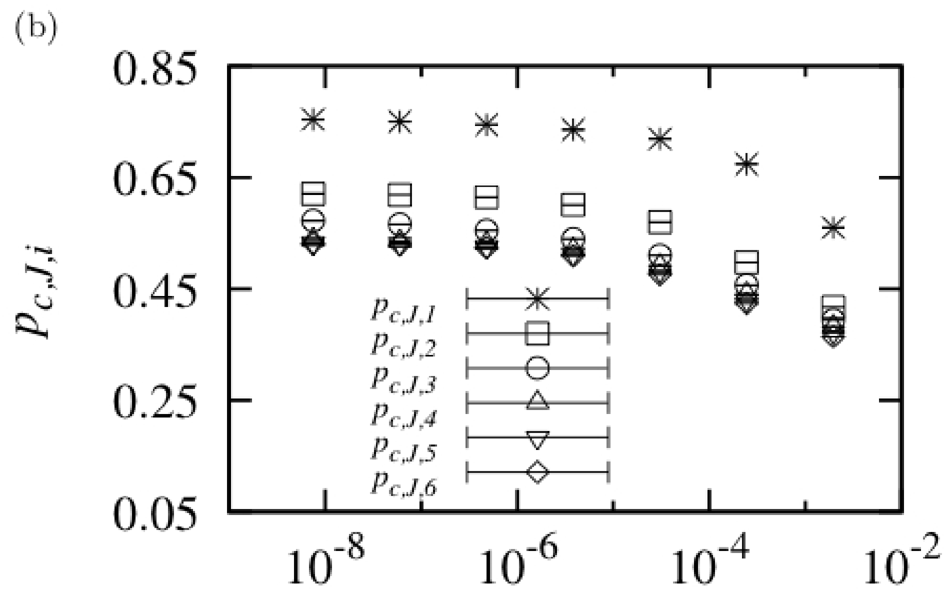
2D lattice (Yes):





3D lattice (NO!):





(Schrenk, Felder, Deflorin, Araujo, D'Souza, Herrmann, PRE 2012)

"Explosive Percolation"
Conclusions & Future Directions:

- Delaying percolation leads to abrupt connectivity transition.

- Finite choice results in continuous transition for $n \to \infty$. But large jumps (e.g., 0.2n to 0.5n) for sizes of real-world networks (n=10^{10})
  Can we develop a rigorous finite size scaling theory?

- Is $t_c = t'_c$?

- Mechanisms:
  – $\log(n)$ choices (i.e. infinite choice)
  – evolving cap on largest component,
  – cooperation / correlations
  – specialized structures (e.g., hierarchical small world 1-D lattices, restricted Erdős-Rényi)

- Applications based on keeping clusters distributed in space and of similar size — community structure detection, wireless networks, going viral through local community growth....

Tomorrow?

Methods

- Probability generating functions / configuration models
- Cluster aggregation evolution equations / Smoluchowski equations
- Multitype branching processes

Models

- Cascades on interconnected networks