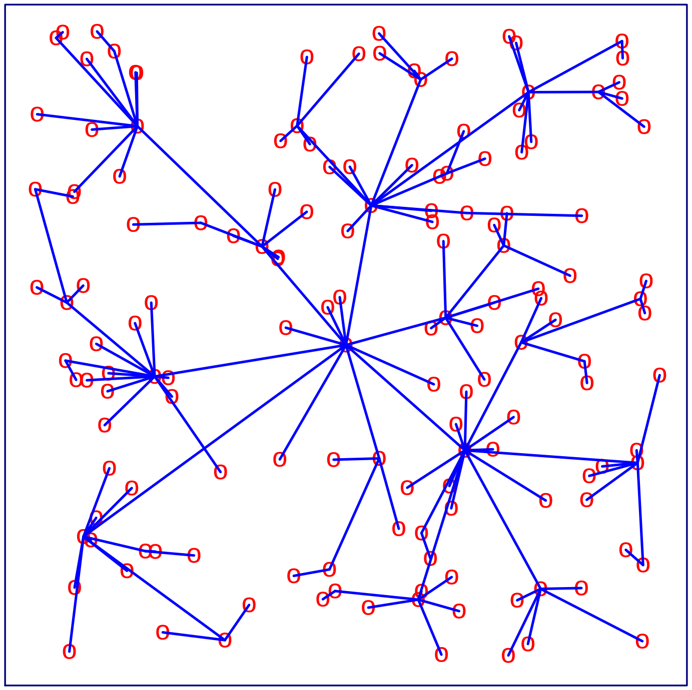
MAE 298, Lecture 2

April 4, 2006



"Random graphs"

Networks

1. Nodes (also called vertices).
2. Edges (also called connections).

- Edges can be directed or undirected.
- Networks can be geometric or be geometry-free. (i.e., the vertices have a geometric location).

Random graphs

What does a “typical” graph with n vertices and m edges look like?

- P. Erdös and A. Rényi, “On random graphs”, *Publ. Math. Debrecen.* **6**, 1959.
- P. Erdös and A. Rényi, “On the evolution of random graphs”, *Publ. Math. Inst. Hungar. Acad. Sci.* **5**, 1960.
- E. N. Gilbert, “Random graphs”, *Annals of Mathematical Statistics* **30**, 1959.

Papers which started the field of graph theory.

Erdös-Rényi random graphs

- Consider a *labelled* graph. Each vertex has a label ranging from $[1, 2, 3, \cdots n]$, for a set of n vertices. (This will make counting and analysis easier.)

- Let E denote the total number of edges possible:

$$E = \binom{N}{2} = \frac{N!}{2!(N-2)!} = \frac{N(N-1)}{2}$$

(If directed edges, we would not divide by 2).

Two formulations

- 1) $\mathcal{G}(n, p)$: The *ensemble* of graphs constructed by putting in edges with probability p, independent of one another. (An edge is present with probability p and absent with probability $[1 - p]$.)

  Let $G(n, p)$ denote a random realization of $\mathcal{G}(n, p)$.

- 2) $\mathcal{G}(n, m)$: The ensemble of all graphs with n nodes and exactly m edges.

  Let $G(n, m)$ denote a random realization of $\mathcal{G}(n, m)$.

- The two are almost interchangeable if $m = pE$. (Recall, E is total number of edges possible).

- We will focus on $G(n, p)$.

$G(n,p)$

- We can build a realization of $G(n,p)$ by the following graph process:
- Start with n isolated vertices.
- At each discrete time step, add one edge chosen at random from edges not yet present on the graph.
- At "time" t (i.e., at the addition of t edges), we have built a realization of $G(n,p)$ where $p = t/E$.
- This is a Markov process (build graph at time $t+1$ from graph at time t).

Illustration of $G(n, p)$ generation process

Component

A component is a subset of vertices in the graph each of which is reachable from the other by some path through the network.

Behavior for small p

- Consider a realization $G(n,p)$ for $0 < p < 1$ and $n \to \infty$. (A number of interesting properties of random graphs can be proven in this limit).

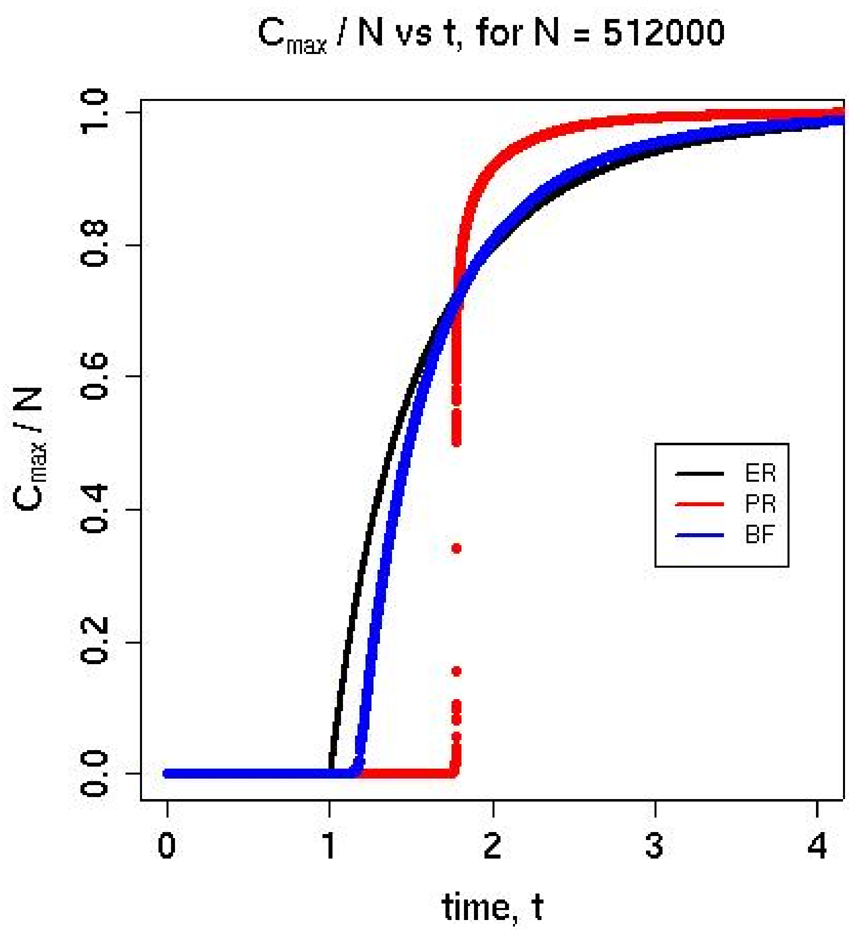
- Consider the size of the largest component of $G(n,p)$ as a function of p, $C_{max}(p)$.

- For small p, few edges on the graph. Almost all vertices disconnected. The components are small, with size $O(\log n)$, independent of p.

- Keep increasing p (or equivalently t in our model). At $p = 1/n$ (i.e. $t = E/n$), something surprising happens:

Emergence of the Giant Component

- For $p = 1/n$ (or equivalently $t = pE = E/n$), suddenly the largest component contains a finite fraction F of the total number of vertices, $C_{max} = Fn$, instead of a logarithmic fraction. All other components remain of size $O(\log n)$.



A Phase Transition!

An abrupt sudden change in one or more physical properties, resulting from a small change in a external control parameter.

Examples from physical systems:

- Magnetization
- Superconductivity
- Liquid/Gas
- Bose-Einstein condensation

Phase transition in connectivity

- Below $p = 1/n$, only small disconnected components.

- Above $p = 1/n$, one large component, which quickly gains more mass. All other components remain sub-linear.

- Note the average node degree, z:

$$
\begin{aligned}
z &= (2 \times \#edges)/\#vertices \\
&= pE/n = pn(n-1)/n = (n-1)p \approx np.
\end{aligned}
$$

  (Factor of 2 since each edge contributes degree to two vertices – each end of the edge contributes).

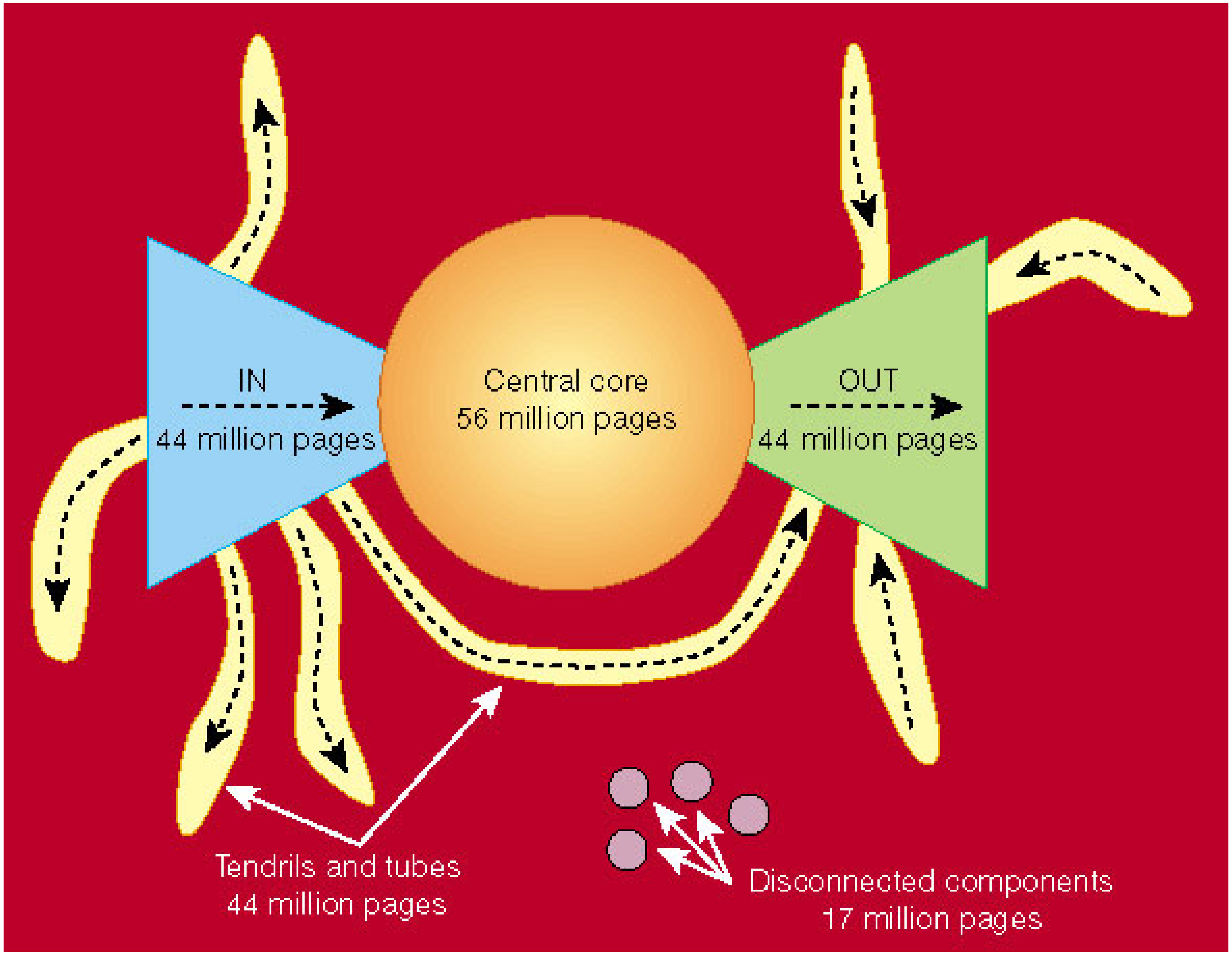
- At the phase transition, $z = np = 1$. The phase transition occurs when the average vertex degree is one!

Giant component observed in real-world networks

- Formation reminiscent of many real-world networks.
  "Gain critical mass".

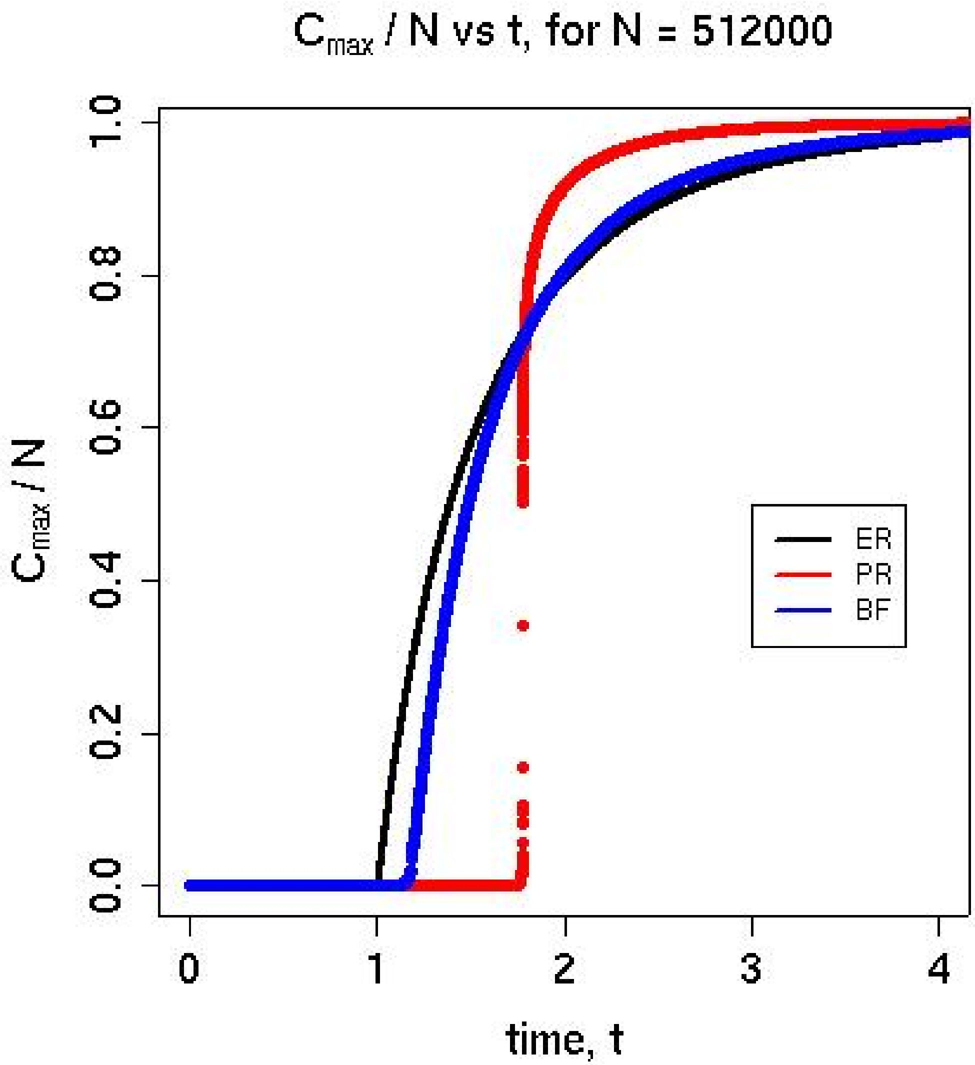
- The giant component/Strongly Connected Component used extensively to categorize networks.

The giant component/Strongly Connected Component of the WWW



From “The web is a bow tie” Nature **405**, 113 (11 May 2000)

"On-line" algorithms for suppressing the emergence of the Giant Component



Back to Erdös-Rényi random graphs

Degree distribution of a graph

- The degree of a node is how many edges connect that node to others.

- If edges are *directed*, a node has a distinct in-degree and out-degree. (Edges in $G(n, p)$ are undirected, so don't have to make that distinction here).

  The degree distribution of the graph is the distribution over all the degrees of all the nodes.

Degree distribution of $G(n, p)$

- Now consider $G(n, p)$ for a fixed value of p and the large n limit.
- The mean degree $z = (n - 1)p$ is constant.
- The absence or presence of an edge is independent for all edges.
  - Probability for node i to connect to all other n nodes is p^n.
  - Probability for node i to be isolated is $(1 - p)^n$.
  - Probability for a vertex to have degree k follows a binomial distribution:

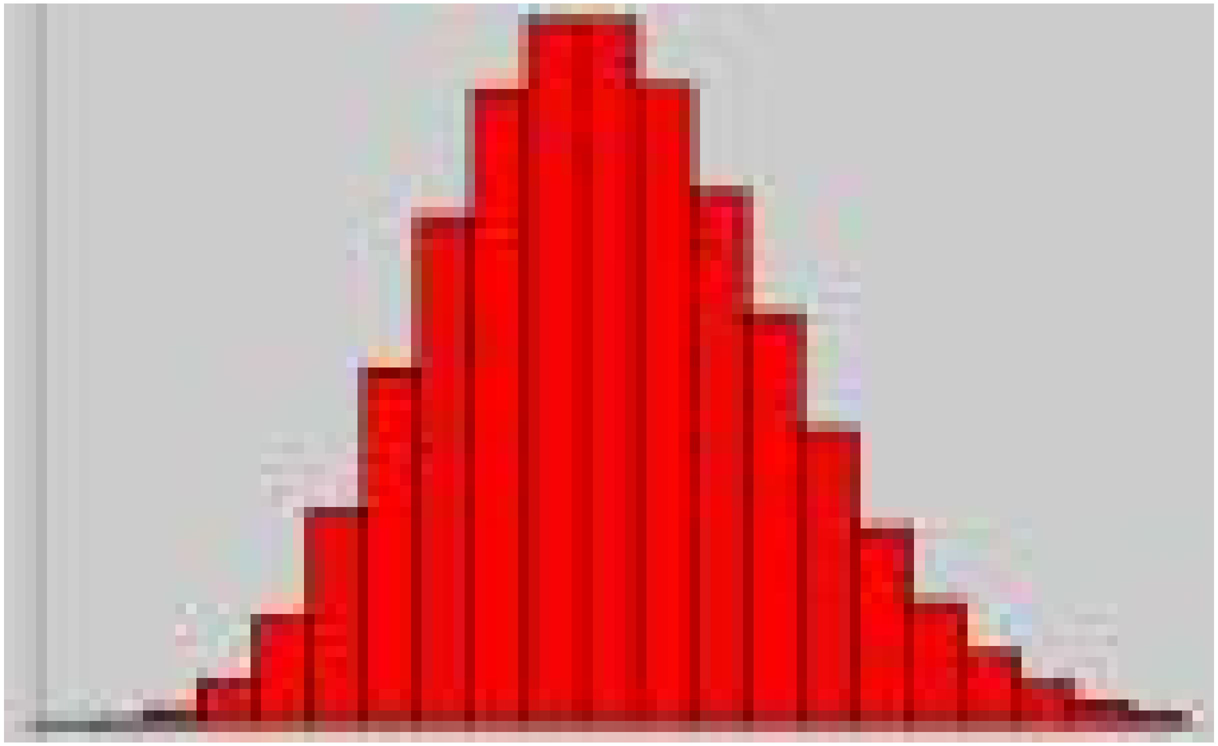
$$p_k = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Binomial converges to Poisson as $n \to \infty$

- Recall that $z = (n-1)p = np$ (for large n).
- $$\begin{aligned} \lim_{n\to\infty} p_k &= \lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} (z/n)^k (1-z/n)^{n-k} \\ &= z^k e^{-z}/k! \end{aligned}$$

For more details see for instance: http://en.wikipedia.org/wiki/Poisson_distribution

Poisson Distribution



Diameter

The diameter of a graph is the *maximum* distance between any two connected vertices in the graph.

- Below the phase transition, only tiny components exist. In some sense, the diameter is infinite.
- Above the phase transition, all vertices in the giant component connected to one another by some path.
- The mean number of neighbors a distance l away is z^l. To determine the diameter we want $z^l \approx n$. Thus the typical distance through the network, $l \approx \log n / \log z$.
- This is a small-world network: diameter $d \sim O(\log N)$.

Clustering coefficient

A measure of transitivity: If node A is known to be connected to B and to C, does this make it more likely that B and C are connected?

(i.e., The friends of my friends are my friends)

- In E-R random graphs, all edges created independently, so no clustering coefficient!

Properties of Erdös-Rényi random graphs:

1. Phase transition in connectivity at average node degree, $z = 1$ (i.e., $p = 1/n$).

2. Poisson degree distribution, $p_k = z^k e^{-z}/k!$.

3. Diameter, $d \sim \log N$, a small-world network.

4. Clustering coefficient; none.

How well does $G(n,p)$ model common real-world networks?

1. Phase transtion: Yes! We see the emergence of a giant component in social and in technological systems.

2. Poisson degree distribution: NO! Most real networks have much broader distributions. (See handout).

3. Small-world diameter: YES! Social systems, subway systems, the Internet, the WWW, biological networks, etc.

4. Clustering coefficient: NO!

Well then, why are random graphs important?

- Much of our basic intuition comes from the study of random graphs.

- Phase transition and the existence of the giant component. Even if not a giant component, many systems have a dominate component much larger than all others.

Generalized random graph

Much effort has gone into thinking about how to make a random graph have a degree distribution different from Poisson.

The configuration model (1970's)

- Specify a degree distribution p_k, such that p_k is the fraction of vertices in the network having degree k.
- We chose an explicit *degree sequence* by sampling in some unbiased way from p_k. And generate the set of n values for k_i, the degree of vertex i.
- Think of attaching k_i "spokes" or "stubs" to each vertex i.
- Choose pairs of "stubs" (from two distinct vertices) at random, and join them. Iterate until done.

Summary: Terms introduced today

- Component
- Phase transition
- Degree distribution
- Graph diameter

Further reading on random graphs

- M. E. J. Newman review, pages 20-25. (Heuristic arguments)
- R. Durrett book, Chaps 1 and 2. (Technical proofs)
- B. Bollobás, *Random Graphs*, 2nd Edition, Cambridge U Press, 2001 (the seminal text on the mathematics of random graphs).