Theory and Applications of Complex Networks

Class Three

College of the Atlantic

David P. Feldman

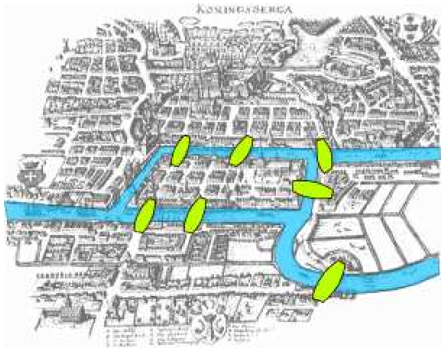
18 September 2008

http://hornacek.coa.edu/dave/

1. A very brief history of networks research
2. Introduction to random graphs
3. Properties of random graphs

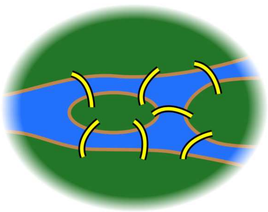
The Beginning of Graph Theory

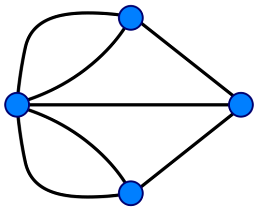
- Leonhard Euler wonders, can I walk through the city of Königsberg and cross each bridge once and only once?



- Figure Source: http://en.wikipedia.org/wiki/Image:Konigsberg_bridges.png.
- In 1736, Euler answers this question with a theorem.

Eulerian paths





- Figure sources: http://en.wikipedia.org/wiki/Image:7_bridges.svg and http://en.wikipedia.org/wiki/Image:Konigsburg_graph.svg.
- Euler recognized this as a graph problem, as shown above.
- He then proved the following general theorem: *An Eulerian path exists on a graph if and only if there are exactly zero or two nodes with odd degree.*
- An *Eulerian path* is a path that uses each edge in the graph exactly once.

Graph Theory

- Euler's 1736 theorem is the first result in the field of graph theory.
- Subsequently, there was a fair amount of work done in this area.
- This sort of work is generally considered to belong to the branch of math known as *combinatorics*.
- This work was mostly confined to pure mathematics and, much later, theoretical computer science.
- In this line of work, graphs were generally viewed as fixed, static quantities. They were not viewed as random variables, nor were the statistics of graphs studied.
- Wikipedia seems to have some good, thorough pages on graph theory and its history.

Random Graphs

- Rapoport (1957) and Erdős and Rényi (1959) introduce random graph models.
- These are, in a sense, maximally random—like flipping coins.
- Erdős and Rényi rigorously prove a number of properties of random graphs.
- These results are probabilistic in nature.
- The basic form of this model is now known as the Erdős and Rényi model.
- Much more about the E-R model later today.
- In general, math for very ordered thing and totally disordered things is “easy.”

Highly Schematic Picture of Order and Disorder

ORDER ←	→ DISORDER
Crystal Structures	Ideal Gases
Exact Symmetries	Tossing Coins (IID Processes)
Group Theory	Unpredictability
Abstract Algebra	Chaos, Mixing, etc.
Regular Graphs, Lattices	Erdos−Renyi Model, Random Graphs

- There are well understood mathematical techniques for studying the extremes of order and disorder.
- Intermediate regions are harder. Often one starts at one extreme and then perturbs or expands off that extreme to get approximate solutions.

Networks and Graphs after Erdős and Rényi

- A fair amount of work in sociology, social networks, economics, etc.
- Also work on computer and technological networks, engineering, etc.
- Then, in 1998, Duncan Watts and Steven Strogatz publish Collective Dynamics of ’Small-world” networks, Nature 393:440–442.
- This paper sparks a remarkable surge of interest in networks.
- Watts and Strogatz’s paper has been cited over 6000 times.
- In 1999, Barabasi and Albert (re)-discover power laws in networks.
- Their paper, Emergence of Scaling in Random Networks, Science 286:509 has now been cited over 3000 times.

Why this Sudden Surge in Networks Research?

In my opinion, this is due to a number of factors.

- Electronic data became available that wasn’t available before.
- Advances in computing power.
- The idea of networks resonates with increased attention to connection, links, globalization, etc.
- Watts and Strogatz’s model was very elegant and simple mathematically, and captured the imagination of a great many people.
- Once physicists became aware of networks, it was quickly realized that they were very well suited to a physics style of analysis.
- Arguably, there wasn’t that much interesting and exciting going on in other areas of physics.
- Complex networks are a natural extension of chaos and complex systems, areas that had attracted considerable attention in the 1980’s and 90’s.

The Erdős Rényi Model

- The Model:
  1. Start with N nodes.
  2. Connect each pair of nodes with probability p.
- Questions:
  - Is the graph connected?
  - What is the degree distribution?
  - What is the size of the graph?
  - What is the clustering coefficient?
- Why might we care?
  - In science, we frequently need to ask, Could this have happened randomly, by chance?
  - In order to answer this question, we need to know about random graphs.

ER Analysis: Preliminaries

Suppose a coin comes up heads with probability p.

- $\mathrm{P}(HHHHH) = p^5$.
- $\mathrm{P}(HHHTT) = p^3(1-p)^2$.
- $\mathrm{P}(HTHHT) = p^3(1-p)^2$.

Probability $\mathrm{P}(3)$ that we get three heads out of five tosses:

$$\mathrm{P} = (\text{\# of ways to get 3 heads}) \times p^3(1-p)^2 \,. \tag{1}$$

After some careful counting, we see that:

$$\mathrm{P} = (10) \times p^3(1-p)^2 \,. \tag{2}$$

What's the general formula?

ER Analysis: Preliminaries

- How many ways can we choose k objects out of a total of N?

$$\binom{N}{k} = \frac{N!}{k!(N-k)!} \,. \tag{3}$$

- So, if we toss a coin N times, the probability $\mathrm{P}(k)$ that we get k heads is:

$$\mathrm{P}(k) = \binom{N}{k} p^N (1-p)^{N-k} \,, \tag{4}$$

  where p is the probability of heads.
- This is an extremely versatile result.
- Try typing “5 choose 2” into a google search.

ER Analysis: Degree Distribution

- How many links does a node have? Each node gets $N-1$ potential links, and each chance yields a link with probability p.
- Thus, the degree distribution $\mathrm{P}(k)$ is:

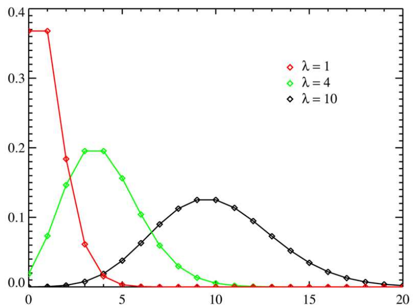
$$\mathrm{P}(k) = \binom{N-1}{k} p^{N-1}(1-p)^{N-1-k} \,. \tag{5}$$

- For large N, this equation becomes well approximated by:

$$\mathrm{P}(k) \approx \frac{z^k e^{-z}}{k!} \,, \tag{6}$$

- Where $z = p(n-1)$ is the mean degree.
- This is known as the Poisson distribution. It arises in many different applications, not just networks. More about its origins in a few week.

Poisson Distribution Properties



- Plot of Poisson Distribution $\mathrm{P}(x) = \frac{\lambda^k e^{-\lambda}}{k!}$ for three different λ values.
- Figure Source: `http://en.wikipedia.org/wiki/Image:Poisson_distribution_PMF.png`.
- Variance $=$ Mean $=$ λ.
- The distribution is discrete. It is only defined for integer k.

ER Analysis: Degree Distribution Conclusion

- **For an Erdős-Rényi Random Graph, the degree distribution is Poisson.**
- The distribution $P(k)$ decays *extremely* rapidly as k gets large—much faster than exponential!
- This means that well connected nodes are astronomically unlikely.
- Example: If $N = 100$ and p = 0.1, then $\lambda = 10$ and $P(50) = 0.000000012$.
- We will see that the degree distribution of empirical networks are very rarely Poisson.

ER Analysis: Clustering Coefficient

- The cluster coefficient is the fraction of your friends that are friends.
- Link probabilities in the ER model are independent.
- Thus, the probability that your friends are friends is just p.
- Hence, $C = p$.
- Conclusion: **Erdős-Rényi** graphs have small clustering coefficients.
- Almost all real-world graphs have clustering coefficients larger than would be expected for comparable ER graphs.

ER Analysis: Characteristic Path Length

- Let $z = np$ be the average degree.
- The number of nodes a distance d from any node is approximately x^d.
- How big must d be so that it includes all of the nodes in the graph? This value of d is ℓ, the characteristic path length:

$$z^\ell = n \quad \longrightarrow \quad \ell = \frac{\log n}{\log z} = \frac{\log n}{\log p + \log n} . \tag{7}$$

- Thus, ER graphs are "small-world," since ℓ grows logarithmically with n.
- Many real-world graphs have the small-world property.

ER Analysis: Is the Graph Connected?

- Roughly speaking, the graph undergoes a phrase transition as p is increased from being a collection of small connected fragments to a graph which has a giant connected component.
- A giant connected component is a connected component that is proportional to n in the large n limit.
- The critical parameter at which this occurs is, not surprisingly, $z = 1$.

ER Analysis: Connectivity Phase Transition

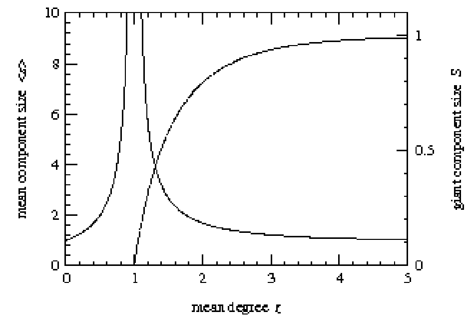


FIG. 10 The mean component size (solid line), excluding the giant component if there is one, and the giant component size (dotted line), for the Poisson random graph, Eqs. (20) and (21).

- Figure Source: M.E.J. Newman, The Structure and Function of Complex Networks, SIAM Reviews, 45(2):167–256, 2003.

Summary of Properties of Erdős-Rényi Model

- Degree distribution is Poisson
- Very low clustering
- Highly connected, “Small-world”
- Connectivity properties change discontinuously

Erdős-Rényi Model Conclusions

- Simple, tractable model of random graphs
- Not a realistic model, but a useful “straw man” or null model
- Does capture the small-world feature common in real-world networks
- Also has discontinuous changes, suggesting that other, more realistic models, might also have sharp thresholds
- Gives us intuition about what to expect from more complicated and realistic models