Network Models

Thanks to Jure Leskovec, Stanford and Panayiotis Tsaparas, Univ. of Ioannina for slides

Agenda

- Erdös-Renyi Random Graph Model
- The Small-World Model
- The Configuration Model
- Power-law distributions
 - Exponential vs Power-law Distributions
 - Scale-free Networks
 - The anatomy of the long-tail
 - Consequence of Power-Law Degrees
- Preferential Attachment Model

Erdös-Renyi Random Graph Model

Simplest Model of Graphs

- Erdös-Renyi Random Graphs [Erdös-Renyi, '60]
 Two variants:
 - G_{n,p}: undirected graph on n nodes and each edge (u,v) appears i.i.d. with probability p
 - $G_{n,m}$: undirected graph with *n* nodes, and *m* uniformly at random picked edges

What kinds of networks does such model produce?

Random Graph Model

n and p do not uniquely determine the graph!

The graph is a result of a random process

 We can have many different realizations given the same *n* and *p*



Random Graph Model: Edges

- How likely is a graph on E edges?
- P(E): the probability that a given G_{np} generates a graph on exactly E edges:

$$P(E) = \begin{pmatrix} E_{\max} \\ E \end{pmatrix} p^{E} (1-p)^{E_{\max}-E}$$

where $E_{max} = n(n-1)/2$ is the maximum possible number of edges in an undirected graph of n nodes

P(E) is exactly the Binomial distribution >>> Number of successes in a sequence of E_{max} independent yes/no experiments



Degree distribution:P(k)Path length:hClustering coefficient:C

What are values of these properties for *G_{np}*?

Node Degrees in a Random Graph

What is expected degree of a node?

- Let X_v be a rnd. var. measuring the degree of node v
- We want to know: $E[X_v] = \sum_{i=0}^{n-1} j P(X_v = j)$
 - For the calculation we will need: Linearity of expectation
 - For any random variables $Y_1, Y_2, ..., Y_k$
 - If $Y = Y_1 + Y_2 + ... Y_k$, then $E[Y] = \sum_i E[Y_i]$

An easier way:

- Decompose X_v to $X_v = X_{v,1} + X_{v,2} + ... + X_{v,n-1}$
 - where X_{v,u} is a {0,1}-random variable which tells if edge (v,u) exists or not

$$E[X_{v}] = \mathop{\text{a}}\limits^{n-1}_{u=1} E[X_{vu}] = (n-1)p$$

How to think about this?

- Prob. of node *u* linking to node *v* is *p*
- *u* can link (flips a coin) to all other (*n*-1) nodes
- Thus, the expected degree of node u is: p(n-1)

Degree Distribution

Fact: Degree distribution of G_{np} is <u>Binomial</u>.
Let P(k) denote a fraction of nodes with degree k:





Mean, variance of a binomial distribution

$$k = p(n-1)$$

 $S^{2} = p(1-p)(n-1)$

$$\frac{S}{\overline{k}} = \stackrel{\acute{e}1 - p}{\acute{e}} \frac{1}{p} \frac{1}{(n-1)} \stackrel{\acute{u}^{1/2}}{\underset{i}{\check{u}}} \gg \frac{1}{(n-1)^{1/2}}$$

By the law of large numbers, as the network size increases, the distribution becomes increasingly narrow—we are increasingly confident that the degree of a node is in the vicinity of *k*.

Clustering Coefficient of G_{np}

• Remember:
$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

Where e_i is the number of edges between i's neighbors

• Edges in G_{np} appear i.i.d. with prob. p

So:
$$e_i = p \frac{k_i(k_i - 1)}{2}$$

Each pair is connected
with prob. p
Number of distinct pairs of
neighbors of node i of degree k_i
Then: $C_i = \frac{p \cdot k_i(k_i - 1)}{k_i(k_i - 1)} = p = \frac{\overline{k}}{n-1} \approx \frac{\overline{k}}{n}$

Clustering coefficient of a random graph is small. For a fixed avg. degree (that is p=1/n), *C* decreases with the graph size *n*.

Network Properties of G_{np}

Degree distribution:

Clustering coefficient:

Path length:

$$P(k) = \binom{n-1}{k} p^{k} (1-p)^{n-1-k}$$
$$C = p = \overline{k}/n$$

next!

Network Properties of G_{np}

Degree distribution:

Clustering coefficient:

Path length:

$$P(k) = \binom{n-1}{k} p^{k} (1-p)^{n-1-k}$$
$$C = p = \overline{k}/n$$

$$O(\log n)$$



Paul Erdös

G_{np} is so cool! Let's also look at its evolution

"Evolution" of a Random Graph

• Graph structure of G_{np} as p changes:



- Emergence of a Giant Component: avg. degree k=2E/n or p=k/(n-1)
 - $k=1-\varepsilon$: all components are of size $\Omega(\log n)$
 - $k=1+\varepsilon$: 1 component of size $\Omega(n)$, others have size $\Omega(\log n)$

G_{np} Simulation Experiment





Fraction of nodes in the largest component

• G_{np} , *n*=100,000, *k*=*p*(*n*-1) = 0.5 ... 3

Back to MSN vs. G_{np}



Real Networks vs. G_{np}

Are real networks like random graphs?

- Average path length: ③
- Giant connected component: ③
- Clustering Coefficient: S
- Problems with the random network model:
 - Degree distribution differs from that of real networks
 - Giant component in most real networks does NOT emerge through a phase transition
 - No "local" structure clustering coefficient is too low

Most important: Are real networks random?

The answer is simply: NO!

Real Networks vs. G_{np}

If G_{np} is wrong, why did we spend time on it?

- It is the reference model for the rest of the class
- It will help us calculate many quantities, that can then be compared to the real data
- It will help us understand to what degree is a particular property the result of some random process

So, while G_{np} is WRONG, it will turn out to be extremly USEFUL!

The Small-World Model

Can we have high clustering while also having short paths?



Six Degrees of Kevin Bacon

Origins of a small-world idea:The Bacon number:

- Create a network of Hollywood actors
- Connect two actors if they co-appeared in the movie
- Bacon number: number of steps to Kevin Bacon
- As of Dec 2007, the highest (finite) Bacon number reported is 8
- Only approx. 12% of all actors cannot be linked to Bacon



Elvis Presley has a Bacon number of 2.





Find out your Erdos number: http://www.ams.org/mathscinet/collaborationDistance.html

The Small-World Experiment

- What is the typical shortest path length between any two people?
 - Experiment on the global friendship network
 - Can't measure, need to probe explicitly
- Small-world experiment [Milgram '67]
 - Picked 300 people in Omaha, Nebraska and Wichita, Kansas
 - Ask them to get a letter to a stock-broker in Boston by passing it through friends
- How many steps did it take?





[Milgram, '67]

The Small-World Experiment

64 chains completed:

(i.e., 64 letters reached the target)

It took 6.2 steps on the average, thus
 "6 degrees of separation"

Further observations:

- People who owned stock had shorter paths to the stockbroker than random people: 5.4 vs. 6.7
- People from the Boston area have even closer paths: 4.4



Milgram: Further Observations

- Boston vs. occupation networks:Criticism:
 - Funneling:



- 31 of 64 chains passed through 1 of 3 people as their final step → Not all links/nodes are equal
- Starting points and the target were non-random
- There are not many samples (only 64)
- People refused to participate (25% for Milgram)
 - Not all searches finished (only 64 out of 300)
- Some sort of social search: People in the experiment follow some strategy instead of forwarding the letter to everyone. They are not finding the shortest path!
- People might have used extra information resources

[Dodds-Muhamad-Watts, '03]

Columbia Small-World Study

- In 2003 Dodds, Muhamad and Watts performed the experiment using e-mail:
 - 18 targets of various backgrounds
 - 24,000 first steps (~1,500 per target)
 - 65% dropout per step
 - 384 chains completed (1.5%)



Avg. chain length = 4.01 Problem: People stop participating Correction factor: $n^*(h) = \frac{n(h)}{\prod_{i=0}^{h-1} (1-r_i)}$ $r_i \dots$ drop-out rate at hop *i*

[Dodds-Muhamad-Watts, '03]

Small-World in Email Study

• After the correction:

Typical path length h = 7

Some not well understood phenomena in social networks:



- Funneling effect: Some target's friends are more likely to be the final step
 - Conjecture: High reputation/authority
- Effects of target's characteristics: Structurally why are high-status target easier to find
 - <u>Conjecture</u>: Core-periphery network structure



What is the structure of a social network?

6-Degrees: Should We Be Surprised?

- Assume each human is connected to 100 other people Then:
 - Step 1: reach 100 people
 - Step 2: reach 100*100 = 10,000 people
 - Step 3: reach 100*100*100 = 1,000,000 people
 - Step 4: reach 100*100*100*100 = 100M people
 - In 5 steps we can reach 10 billion people
- What's wrong here?
 - 92% of new FB friendships are to a friend-of-a-friend [Backstom-Leskovec '11]





Clustering Implies Edge Locality

 MSN network has 7 orders of magnitude larger clustering than the corresponding G_{np}!
 Other examples:

Actor Collaborations (IMDB): N = 225,226 nodes, avg. degree $\overline{k} = 61$ Electrical power grid: N = 4,941 nodes, $\overline{k} = 2.67$ Network of neurons: N = 282 nodes, $\overline{k} = 14$

Network	\mathbf{h}_{actual}	h_{random}	C _{actual}	C _{random}
Film actors	3.65	2.99	0.79	0.00027
Power Grid	18.70	12.40	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

- h ... Average shortest path length
- C ... Average clustering coefficient
- "actual" ... real network
- "random" ... random graph with same avg. degree

The "Controversy"

Consequence of expansion:

Short paths: O(log n)

- This is "best" we can do if we have a constant degree
- But clustering is low!
- But networks have "local" structure:
 - Triadic closure:

Friend of a friend is my friend

- High clustering but diameter is also high
- How can we have both?



Low diameter Low clustering coefficient



High clustering coefficient High diameter

Small-World: How?

- Could a network with high clustering be at the same time a small world?
 - How can we at the same time have high clustering and small diameter?





- Clustering implies edge "locality"
- Randomness enables "shortcuts"

[Watts-Strogatz, '98]

Solution: The Small-World Model

Small-world Model [Watts-Strogatz '98] Two components to the model:

- (1) Start with a low-dimensional regular lattice
 - (In our case we use a ring as a lattice)
 - Has high clustering coefficient
- Now introduce randomness ("shortcuts")

(2) Rewire:

- Add/remove edges to create shortcuts to join remote parts of the lattice
- For each edge with prob. p move the other end to a random node



[Watts-Strogatz, '98]

The Small-World Model



Rewiring allows us to "interpolate" between a regular lattice and a random graph

The Small-World Model



Small-World: Summary

- Could a network with high clustering be at the same time a small world?
 - Yes! You don't need more than a few random links
- The Watts Strogatz Model:
 - Provides insight on the interplay between clustering and the small-world
 - Captures the structure of many realistic networks
 - Accounts for the high clustering of real networks
 - Does not lead to the correct degree distribution
 - Does not enable navigation (offline lecture)

How to Navigate a Network?

• (offline) What mechanisms do people use to navigate networks and find the target?


The Configuration Model

Intermezzo: Configuration Model

Goal: Generate a random graph with a given degree sequence k₁, k₂, ... k_N
 Configuration model:



Useful as a "null" model of networks

We can compare the real network G and a "random" G' which has the same degree sequence as G

Another Example

Suppose that the degree sequence is

- Create multiple copies of the nodes
- Pair the nodes uniformly at random

4 1 3 2 • • • •

Generate the resulting network



Other Properties

The giant component phase transition for this model happens when

$$\sum_{k=0}^{\infty} k(k-2)p_k = 0$$

p_k: fraction of nodes with degree **k**

The clustering coefficient is given by

$$C = \frac{1}{n} \frac{\left(\left\langle k^2 \right\rangle - \left\langle k \right\rangle\right)^2}{\left\langle k \right\rangle^3}$$

The diameter is logarithmic

Power-Law Degree Distributions

Thanks to Jure Leskovec, Stanford and Panayiotis Tsaparas, Univ. of Ioannina for slides

Agenda

- Power-law distributions
 - Exponential vs Power-law Distributions
 - Scale-free Networks
 - The anatomy of the long-tail
- Mathematics of Power-laws
- Estimating Power-law Exponent Alpha
- Consequence of Power-Law Degrees

Network Formation Processes

What do we observe that needs explaining

- Small-world model?
 - Diameter
 - Clustering coefficient
- Preferential Attachment:
 - Node degree distribution
 - What fraction of nodes has degree k (as a function of k)?
 - Prediction from simple random graph models: p(k) = exponential function of k
 - Observation: Often a power-law: $p(k) = k^{-\alpha}$



Degree Distributions



Node Degrees in Networks

Take a network, plot a histogram of P(k) vs. k



[Leskovec et al. KDD 'o8]

Node Degrees in Networks

Plot the same data on *log-log* scale:



Node Degrees: Faloutsos³

Internet Autonomous Systems [Faloutsos, Faloutsos and Faloutsos, 1999]



Node Degrees: Web

The World Wide Web [Broder et al., 2000]



Node Degrees: Barabasi&Albert

Other Networks [Barabasi-Albert, 1999]



Exponential vs. Power-Law



Above a certain x value, the power law is always higher than the exponential!

[Clauset-Shalizi-Newman 2007]

Exponential vs. Power-Law

Power-law vs. Exponential on log-log and semi-log (log-lin) scales



Exponential vs. Power-Law



Power-Law Degree Exponents

- Power-law degree exponent is typically 2 < α < 3
 - Web graph:
 - α_{in} = 2.1, α_{out} = 2.4 [Broder et al. 00]
 - Autonomous systems:
 - α = 2.4 [Faloutsos³, 99]
 - Actor-collaborations:
 - α = 2.3 [Barabasi-Albert 00]
 - Citations to papers:
 - α ≈ 3 [Redner 98]
 - Online social networks:
 - $\alpha \approx 2$ [Leskovec et al. 07]



Scale-Free Networks

Definition:

Networks with a power-law tail in their degree distribution are called "scale-free networks"

Where does the name come from?

- Scale invariance: There is no characteristic scale
- Scale-free function: $f(ax) = a^{\lambda}f(x)$

• Power-law function: $f(ax) = a^{\lambda}x^{\lambda} = a^{\lambda}f(x)$

Log() or Exp() are not scale free! $f(ax) = \log(ax) = \log(a) + \log(x) = \log(a) + f(x)$ $f(ax) = \exp(ax) = \exp(x)^a = f(x)^a$

[Clauset-Shalizi-Newman 2007]

Power-Laws are Everywhere



Many other quantities follow heavy-tailed distributions

[Chris Anderson, Wired, 2004]

Anatomy of the Long Tail

ANATOMY OF THE LONG TAIL

Online services carry far more inventory than traditional retailers. Rhapsody, for example, offers 19 times as many songs as Wal-Mart's stock of 39,000 tunes. The appetite for Rhapsody's more obscure tunes (charted below in yellow) makes up the so-called Long Tail. Meanwhile, even as consumers flock to mainstream books, music, and films (right), there is real demand for niche fare found only online.

6.100



THE NEW GROWTH MARKET: **OBSCURE PRODUCTS YOU CAN'T GET ANYWHERE BUT ONLINE**



Consequence of Power-Law Degrees

Random vs. Scale-free network





Random network

(Erdos-Renyi random graph)



Scale-free (power-law) network

Degree distribution is Power-law

Consequence: Network Resilience

- How does network connectivity change as nodes get removed? [Albert et al. 00; Palmer et al. 01]
- Nodes can be removed:
 - Random failure:



- Remove nodes uniformly at random
- Targeted attack:

Remove nodes in order of decreasing degree

This is important for robustness of the internet as well as epidemiology

Network Resilience



- Real networks are resilient to <u>random failures</u>
 G_{np} has better resilience to <u>targeted attacks</u>
 - Need to remove all pages of degree >5 to disconnect the Web
 - But this is a very small fraction of all web pages

Preferential Attachment Model

Exponential vs. Power-Law Tails



Model: Preferential attachment

Preferential attachment:

[de Solla Price '65, Albert-Barabasi '99, Mitzenmacher '03]

- Nodes arrive in order 1,2,...,n
- At step j, let d_i be the degree of node i < j</p>
- A new node j arrives and creates m out-links
- Prob. of *j* linking to a previous node *i* is proportional to degree *d_i* of node *i*

$$P(j \to i) = \frac{d_i}{\sum_k d_k}$$



Rich Get Richer

New nodes are more likely to link to nodes that already have high degree

Herbert Simon's result:

Power-laws arise from "Rich get richer" (cumulative advantage)

Examples

- Citations [de Solla Price '65]: New citations to a paper are proportional to the number it already has
 - Herding: If a lot of people cite a paper, then it must be good, and therefore I should cite it too
- Sociology: Matthew effect
 - Eminent scientists often get more credit than a comparatively unknown researcher, even if their work is similar
 - <u>http://en.wikipedia.org/wiki/Matthew_effect</u>

Node j

The Exact Model

We will analyze the following model:

- Nodes arrive in order 1,2,3, ..., *n*
- When node *j* is created it makes a single out-link to an earlier node *i* chosen:
 - 1) With prob. *p*, node *j* links to *i* chosen uniformly at random (from among all earlier nodes)
 - 2) With prob. 1 p, node j chooses i uniformly at random & links to a random node l that i points to
 - This is same as saying: With prob. 1 p, node j links to node l with prob. proportional to d_l (the in-degree of l)
 - Our graph is directed: Every node has out-degree 1

The Model Gives Power-Laws

Claim: The described model generates networks where the fraction of nodes with in-degree k scales as:

$$P(d_i = k) \propto k^{-(1+\frac{1}{q})}$$

where q=1-p

So we get power-law degree distribution with exponent:



Preferential attachment: Good news

- Preferential attachment gives power-law degrees!
- Intuitively reasonable process
- Can tune p to get the observed exponent
 - On the web, *P[node has degree d]* ~ *d*^{-2.1}
 - $2.1 = 1 + 1/(1-p) \rightarrow \underline{p} \sim 0.1$

Preferential Attachment: Bad News

- Preferential attachment is not so good at predicting network structure
 - Age-degree correlation
 - Solution: Node fitness (virtual degree)
 - Links among high degree nodes:
 - On the web nodes sometime avoid linking to each other
- Further questions:
 - What is a reasonable model for how people sample through network node and link to them?
 - Short random walks

Generating Power-Law Values

- A simple trick to generate values that follow a power-law distribution:
 - Generate values r uniformly at random within the interval [0,1]
 - Transform the values using the equation $x = x_{min}(1-r)^{-1/(\alpha-1)}$
 - Generates values distributed according to powerlaw with exponent α

Many models lead to Power-Laws

- Copying mechanism (directed network)
 - Select a node and an edge of this node
 - Attach to the endpoint of this edge
- Walking on a network (directed network)
 - The new node connects to a node, then to every first, second, ... neighbor of this node
- Attaching to edges
 - Select an edge and attach to both endpoints of this edge

Node duplication

- Duplicate a node with all its edges
- Randomly prune edges of new node

Distances in Preferential Attachm

Ultra
small
world
$$\left\{ \begin{array}{l} const & \alpha = 2 \\ \frac{\log \log n}{\log(\alpha - 1)} & 2 < \alpha < 3 \end{array} \right.$$

 $\overline{h} = \left\{ \begin{array}{l} \frac{\log n}{\log(\alpha - 1)} & 2 < \alpha < 3 \\ \frac{\log n}{\log\log n} & \alpha = 3 \end{array} \right.$
Small
world $\left\{ \begin{array}{l} \log n & \alpha > 3 \end{array} \right.$
Avg. path Degree
exponent

h

Size of the biggest hub is of order O(N). Most nodes can be connected within two steps, thus the average path length will be independent of the network size.

The average path length increases slower than logarithmically. In G_{np} all nodes have comparable degree, thus most paths will have comparable length. In a scalefree network vast majority of the path go through the few high degree hubs, reducing the distances between nodes.

Some models produce $\alpha = 3$. This was first derived by Bollobas et al. for the network diameter in the context of a dynamical model, but it holds for the average path length as well.

The second moment of the distribution is finite, thus in many ways the network behaves as a random network. Hence the average path length follows the result that we derived for the random network model earlier.

Summary: Scale-Free Network

