

# Network Models: Pick your characteristic

*Elizabeth Bodine*

*CS 286B*

*2/26/09*

# Outline

- Motivation: Why do we care about models?
- A first random graph model
- Preferential Attachment (the rich get richer)
- Social Network Models (the highlights)
- Conclusions

# Why do we care?

- In general, why do we care about models?
  - It's easier
  - In theory, it gets us closer to the truth, or something like it
- Why do I care?
  - Real world datasets of social networks
    - HEAVY TAIL degree distribution
  - An experiment by the (in)famous Stanley Milgram in 1968
    - SMALL diameter, NAVIGABILITY, HIGH clustering
  - In general, social networks can help model themselves, influence (think marketing), epidemiology, and communications



# A first random graph model: When are we connected?

Proof of connectedness in ER graphs

# A first random graph model: Some properties

- If  $np < 1 \Rightarrow$  No connected component of size  $> O(\log n)$
- If  $np = 1 \Rightarrow$  Largest component size  $\sim n^{2/3}$
- If  $np = c > 1 \Rightarrow$  "Giant" component of size  $\sim O(n)$ , no other component has size  $> O(\log n)$
- If  $p < (1-\epsilon)\ln(n)/n \Rightarrow$  NOT connected
- If  $p > (1-\epsilon)\ln(n)/n \Rightarrow$  Connected!
  
- "Small-world effect"
  - $E[\# \text{ neighbors with distance } = k] = c^k$
  - To encompass entire network,  $c^k = n$
  - $k = \log n / \log c$  (typical distance through network – logarithmic!)

# A first random graph model: Some problems

- Low clustering coefficient:
  - $P(\text{connection}) = p$ , regardless of common neighbors
  - $C = p$
- Poisson degree distribution (NOT heavy-tailed)
- No correlation between degrees of adjacent vertices, no community structure, navigation impossible using distributed algorithms...
- So what's a girl to do?

# The Preferential Attachment model: Generation

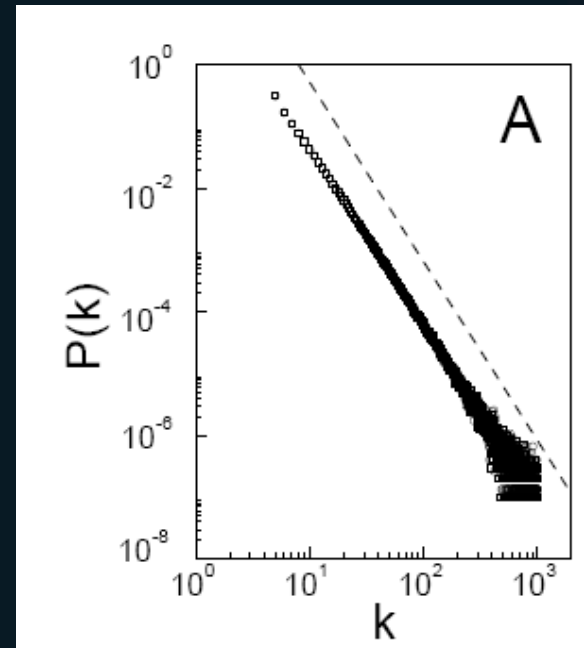
- Barabasi, Albert (1999)
- Vertices are added to the network with degree  $m$ , each edge is attached to another vertex with a probability proportional to the degree of that vertex
- Undirected
- Results in a power-law degree distribution!

[http://commons.wikimedia.org/wiki/File:Barabasi\\_Albert\\_model.gif](http://commons.wikimedia.org/wiki/File:Barabasi_Albert_model.gif)

# The Preferential Attachment Model: Proof of degree distribution

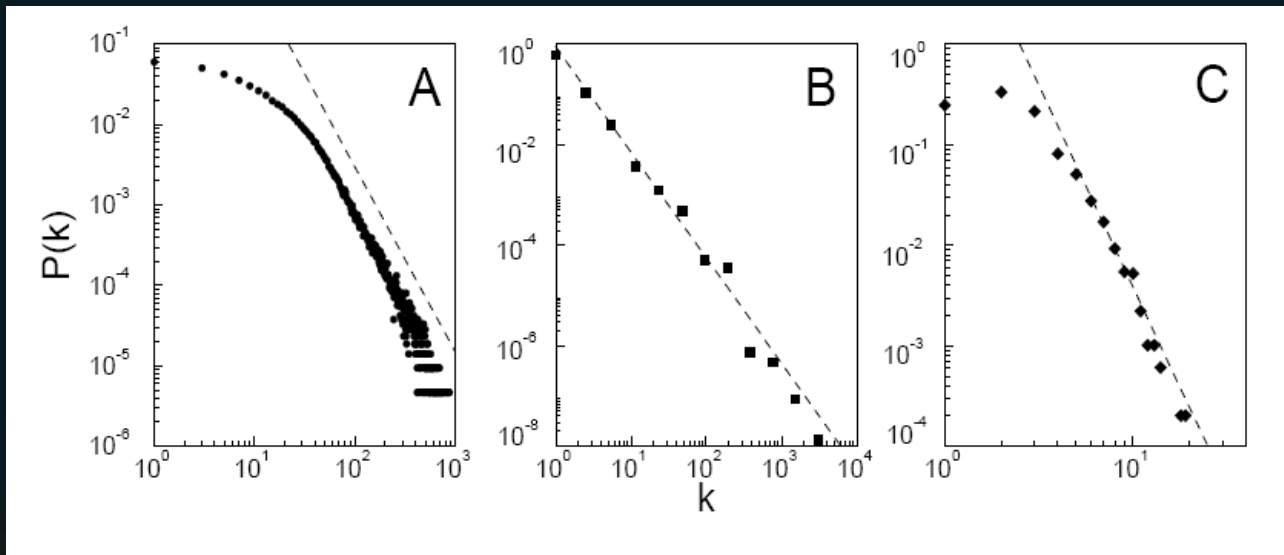
# The Preferential Attachment Model: Some properties

- Heavy-tailed:  $\gamma \sim 2.9 \pm .1$
- Correlation between age of vertex and its degree (or is this bad?)
- Correlation between degrees of adjacent vertices
- To get HT distribution that is independent of time, need both GROWTH and PREFERENTIAL ATTACHMENT



# The Preferential Attachment Model: Some properties

- In the real world, we have the following power-laws:
  - Movie actors:  $\gamma \sim 2.3 \pm .1$
  - WWW:  $\gamma \sim 2.1 \pm .1$
  - Power grid of US:  $\gamma \sim 4$
  - Citation pattern:  $\gamma \sim 3$



# Let's be sociable!



“Fred Jones of Peoria, sitting in a sidewalk café in Tunis, and needing a light for his cigarette, asks the man at the next table for a match. They fall into conversation; the stranger is an Englishman who, it turns out, spent several months in Detroit studying the operation of an interchangeable-bottlecap factory. “I know it’s a foolish question,” says Jones, “but did you ever by chance run into a fellow named Ben Arkadian? He’s an old friend of mine, manages a chain of supermarkets in Detroit...”

“Arkadian, Arkadian,” the Englishman mutters. “Why, upon my soul, I believe I do! Small chap, very energetic, raised merry hell with the factory over a shipment of defective bottlecaps.”

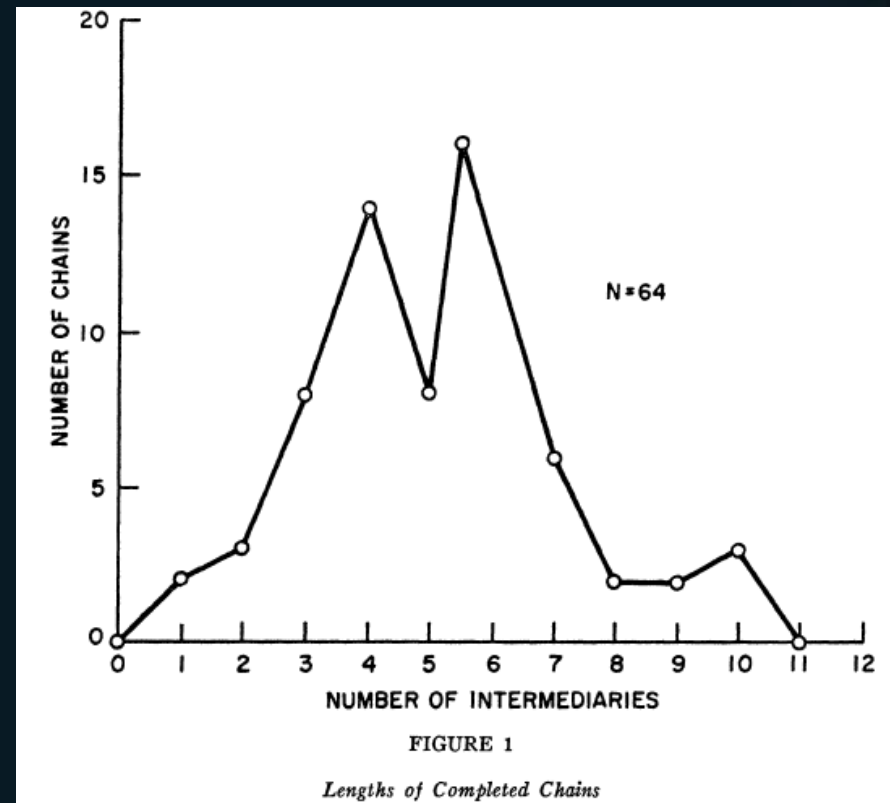
“No kidding!” Jones explains in amazement.

“Good lord, it’s a small world, isn’t it?”

Milgram, Stanley. "The Small World Problem". *Psychology Today*, 1(1), May 1967. pp 60 - 67

# Milgram's 6 degrees of separation

- Milgram's experiment: "Given any two people in the world, person X and person Z, how many intermediate links are needed before X and Z are connected?"
- Target person: stockholder in Sharon, MA
- 296 volunteers, 196 residents of Nebraska sent letters, asking them to move a message towards the target person, using only a chain of friends and acquaintances
- Result: 29% of letters reached target



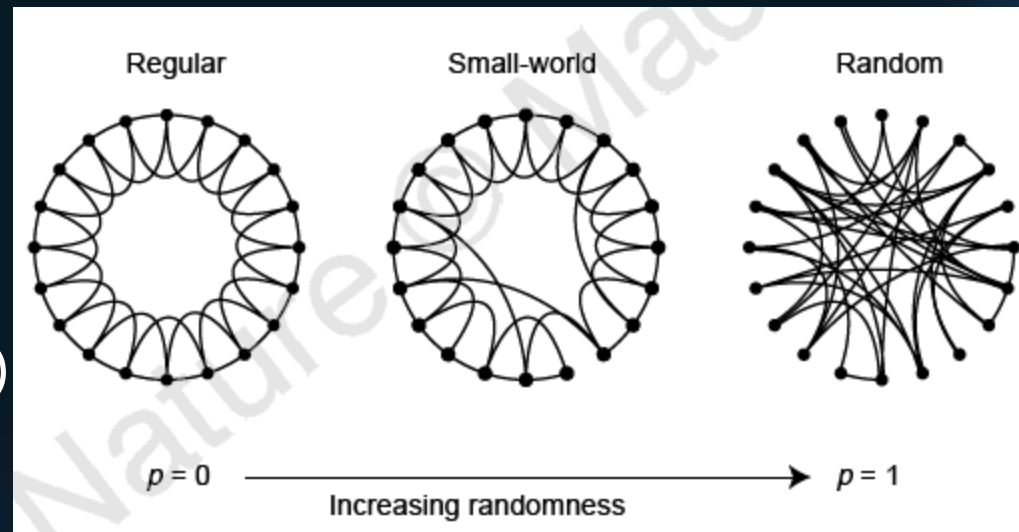
# Models of Social Networks: What do we care about?

- HEAVY-TAILED degree distributions
  - $P(d) \sim x^{-\alpha}$
- SMALL diameter
  - Average (or maximum) distance between nodes is poly-logarithmic
- HIGH clustering
  - Average clustering coefficient higher than in ER graphs
- NAVIGABILITY
  - Can we find each other?
  - More precisely, given a limited amount of knowledge, can any distributed algorithm locate a poly-log length path?

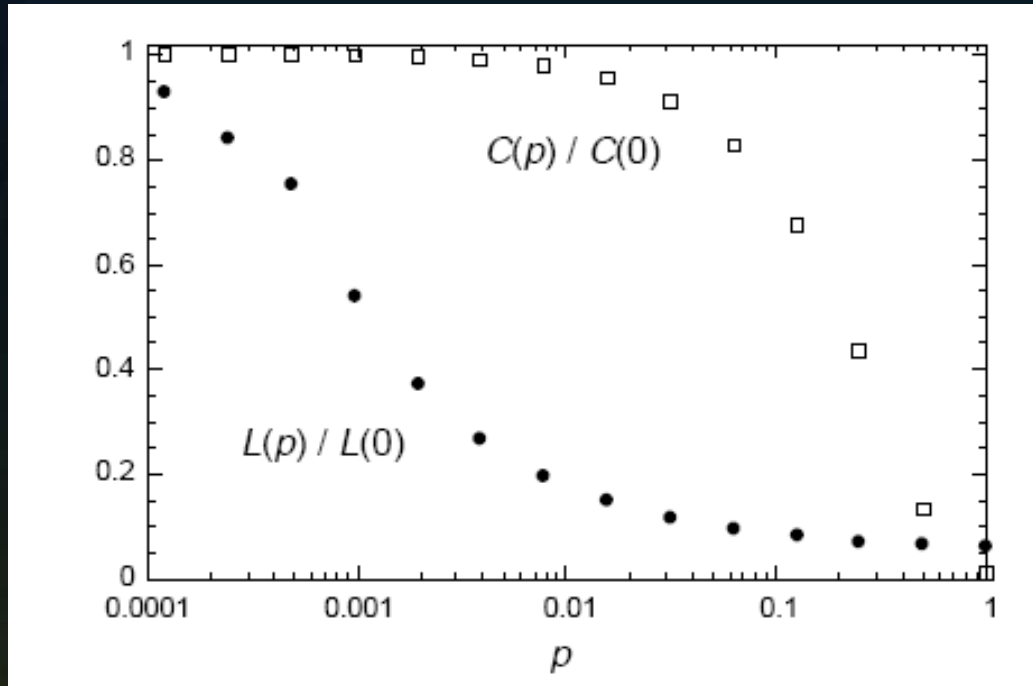
# Watts and Strogatz Model: Generation

- Start with a ring lattice with  $n$  vertices and  $k$  edges per vertex, rewire each edge with probability  $p$
- Length ( $L$ ): # of edges in the shortest path between two vertices, averaged over all pairs of vertices
- Clustering ( $C$ ): Average clustering coefficient

- As  $p \rightarrow 0$   
 $L \sim n/2k \gg 1$   
 $C \sim 3/4$
- As  $p \rightarrow 1$ ,  
 $L \approx L_{\text{random}} \sim \ln(n)/\ln(k)$   
 $C \approx C_{\text{random}} \sim k/n \ll 1$



# Watts and Strogatz Model: High clustering, small diameter



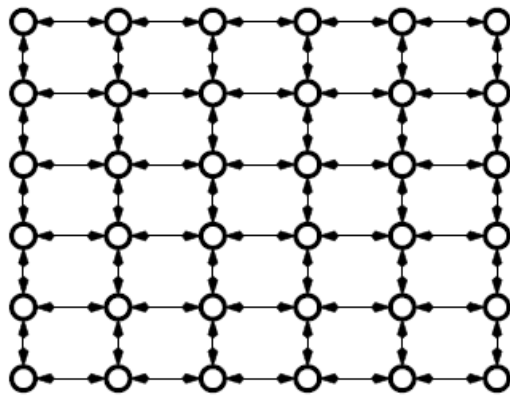
- There exists a large interval over which  $L(p)$  is almost as small as  $L_{\text{random}}$ , yet  $C(p) \gg C_{\text{random}}$
- “small-world network”

- We get: HIGH clustering, SMALL Diameter
- We lose: Power-law degree distribution, distributed navigability
- Examples: collaboration of film actors, power grid of US, neural network of *C. elegans*

# The Kleinberg model: Generation

- Start with an  $n \times n$  lattice, all nodes connected within distance  $p$
- Add  $q$  long range links to each node, with probability  $\sim d(u,v)^{-2}$

A)



B)

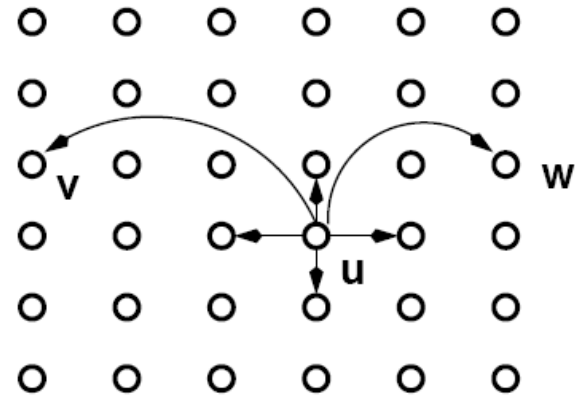


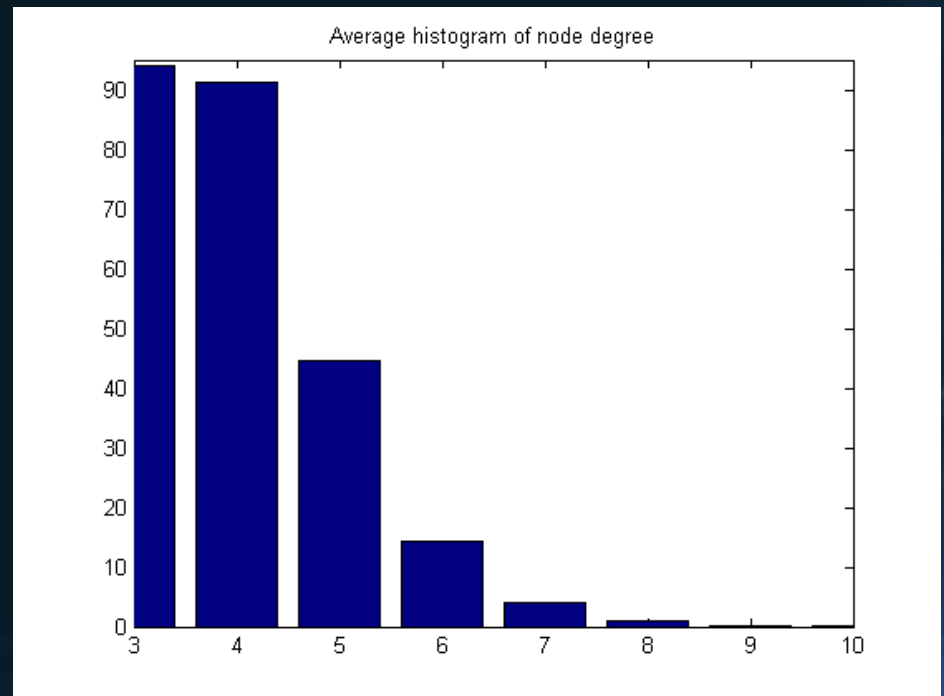
Figure 1: (A) A two-dimensional grid network with  $n = 6$ ,  $p = 1$ , and  $q = 0$ . (B) The contacts of a node  $u$  with  $p = 1$  and  $q = 2$ .  $v$  and  $w$  are the two long-range contacts.

# The Kleinberg model:

## Main result

*In the  $k$ -dimensional case, a decentralized algorithm can construct paths of length polynomial in  $\log(n)$  if and only if  $r = k$ .*

- We get: SEARCHABILITY
- We lose: Power-law degree distribution  
(In fact,  $E[e^{sV}] = e^{-(1-x)} = e^{-1}(1+x+x^2/2!+\dots)$ )



# And many more...

- Kleinberg: generalized to trees, groups
- Hierarchical (Watts, Newman, Dodd):
  - Seems to be like real life
- Kronecker graphs: generate probabilistic adjacency matrix from initiator matrix and Kronecker multiplication

# So what have we learned?

- Models are good for analysis, but not perfect
  - Erdos-Reyni: easy!
  - Preferential Attachment: Power-law degree distribution
  - Watts-Strogatz: High clustering, small diameter
  - Kleinberg: navigability
- No model captures all of the effects of a social network
- In the end, the definition of “small-world” is flexible, so you can choose a model that fits the characteristic you care about