




Inspire…Educate…Transform.

9. Social Network Analysis

Dr. Manish Gupta

Sr. Mentor – Academics, INSOFE

Adapted from

http://www.stanford.edu/class/cs224w/slides/01-intro.pdf

http://temporalweb.net/2011/files/kumar-twaw2011.pdf

http://www.stanford.edu/class/cs224w/slides/02-gnp.pdf

http://www.stanford.edu/class/cs224w/slides/03-smallworld.pdf

Course Content

- Collection of three main topics of high recent interest.
  - Search engines (Crawling, Indexing, Ranking)
    - Language Modeling
    - Text Indexing and Crawling
    - Relevance Ranking
    - Link Analysis Algorithms
  - Text Processing (NLP, NER, Sentiments)
    - Natural Language Processing
    - Named Entity Recognition
    - Sentiment Analysis
    - Summarization
  - Social networks (Properties, Influence Propagation)
    - **Social Network Analysis**
    - Influence Propagation in Social Networks

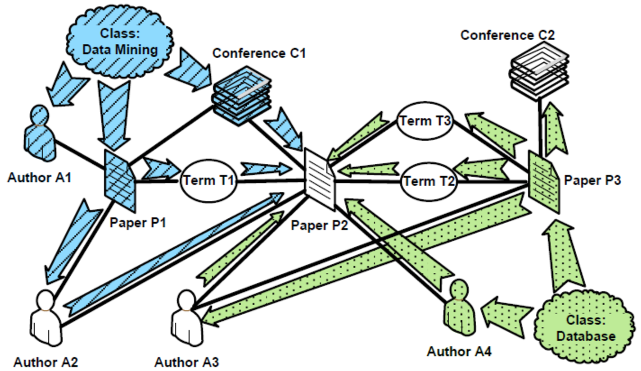
Today's Agenda

- **Introduction to Social Network Analysis**

Lots of Social Networks Today

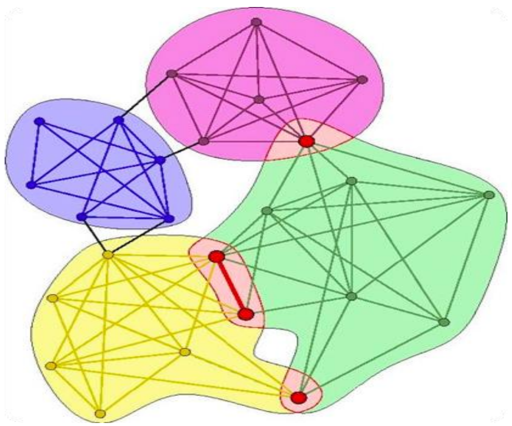
CSE 7306c

Network Analysis Tasks

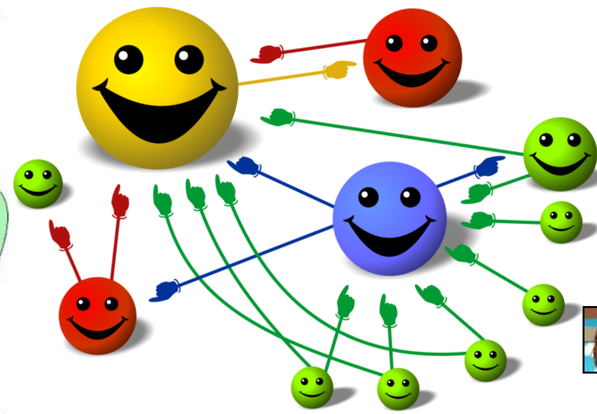


Classification

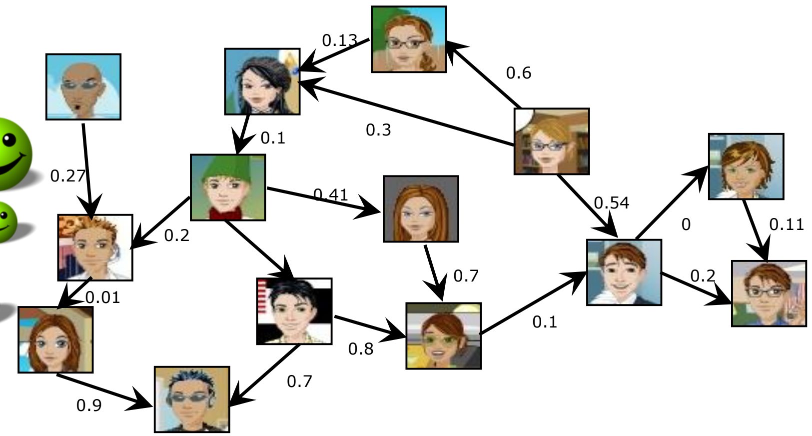
Link Prediction



Community Detection



PageRank

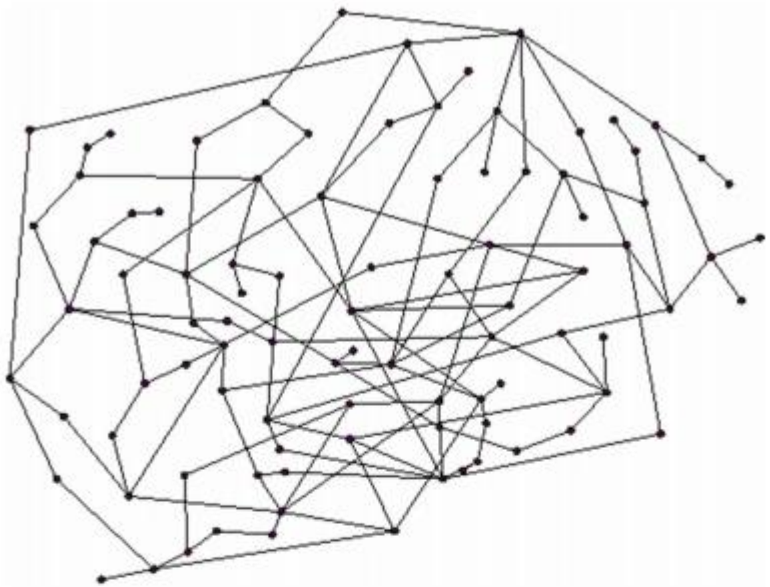


Influence Propagation

Studying Social Networks

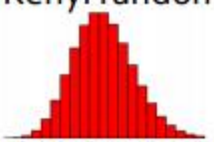
- More and more of all interactions are happening online
- The resulting data is a goldmine
  - Massive amounts
  - Study phenomena at different scales (eg, interaction of people in focused groups of different sizes, overall structure of the network)
  - Ability to measure, record, and track individual activities at the finest resolution (eg, user befriending another, user buying a dvd, user tagging a photo, user tweeting — when, how, why)
- A double-edged sword
  - Data is inherently noisy
  - Large scale of data leads to algorithmic challenges
- Graph-theoretic analysis has led to significant impact
  - Link analysis in web search
  - Sophisticated recommendation systems
- Interplay of measurements, modeling, and algorithms

Random vs. Scale-free Networks

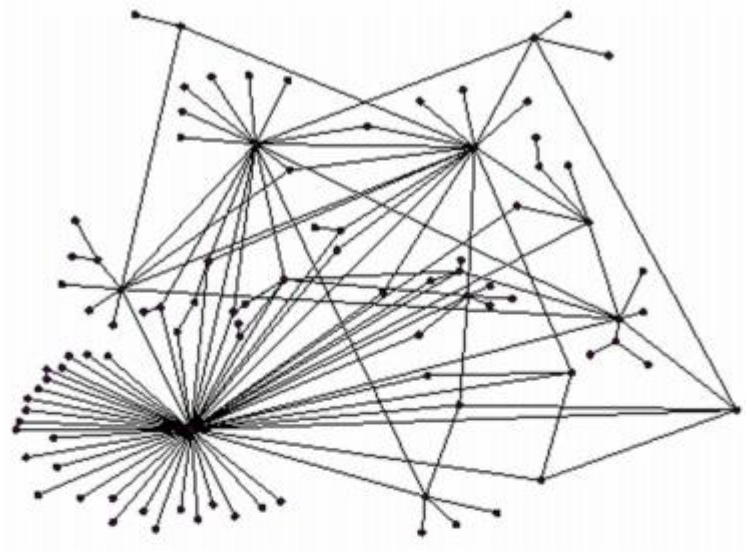


Random network

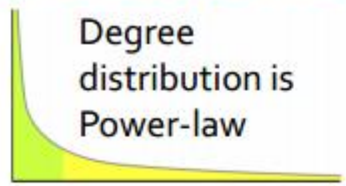
(Erdos-Renyi random graph)



Degree distribution is Binomial



Scale-free (power-law) network



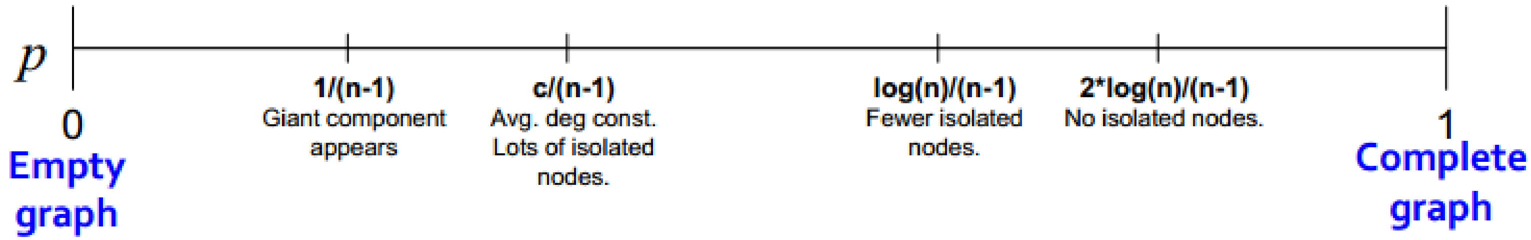
Today's Agenda

- Introduction to Social Network Analysis
- **Erdös-Renyi Model**

The G_{np} Model

- Erdös-Renyi Random Graphs [Erdös-Renyi, ‘60]
- Undirected graph on n nodes and each edge (u,v) appears with probability p
- Expected degree of a node? $(n-1)p$
- Degree distribution of G_{np} is Binomial(n,p)
- Clustering coefficient = p
- Path length = O(log n)

- **Graph structure of G_{np} as p changes:**



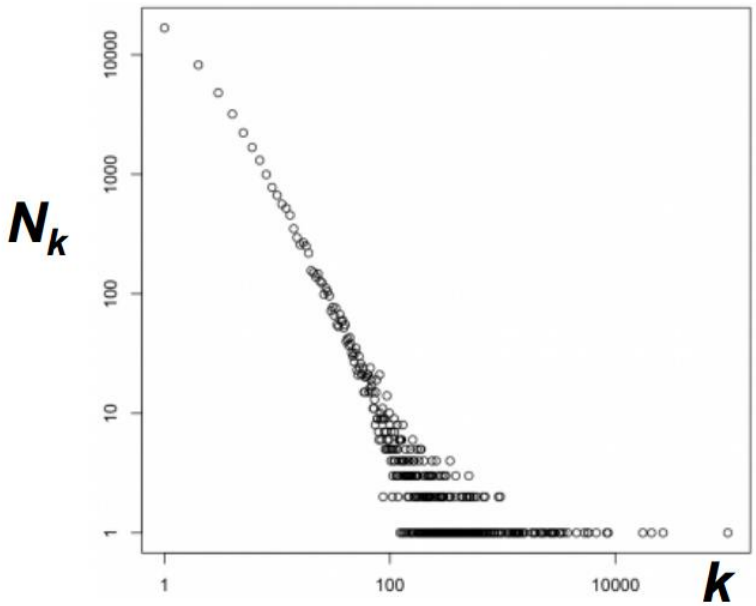
CSE 7306c

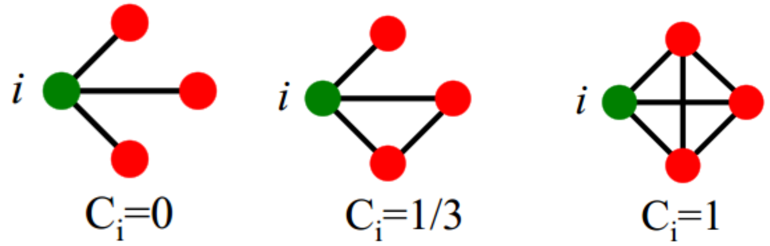
Properties of Graphs

- Degree distributions
  - Heavy tail
- Clustering
  - High clustering coefficient
- Communities and dense subgraphs
  - Abundance; locally dense, globally dense; spectrum
- Connected components
  - Distribution; bow-tie structure
- Connectivity
  - Low diameter; small-world properties

Properties of Graphs

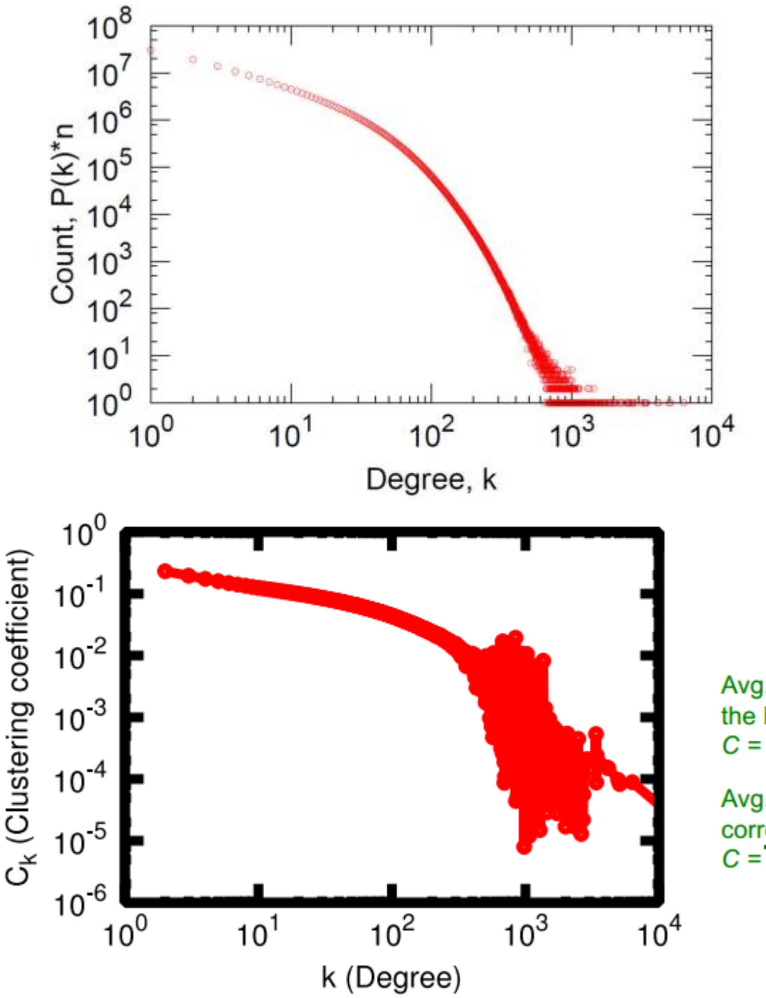
- Degree distribution P(k): Probability that a randomly chosen node has degree k
  - N_k= # nodes with degree k
- Network Diameter
  - the maximum (shortest path) distance between any pair of nodes in a graph
- Clustering coefficient of a node i
  - What fraction of i's neighbors are connected?
  - Node i with degree k_i
    - $C_i = \frac{2e_i}{k_i(k_i-1)}$
    - $C_i \in [0,1]$
- Average clustering coefficient
  - $C = \frac{1}{N}\sum_{i=1}^{N} C_i$





How well does G_{np} correspond to Real Networks?

- MSN Messenger activity in June 2006:
  - 245 million users logged in
  - 180 million users engaged in conversations
  - More than 30 billion conversations
  - More than 255 billion exchanged messages



Avg. clustering of the MSN: $C = 0.1140$

Avg. clustering of corresponding G_{np}: $C = \bar{k}/n \approx 8 \cdot 10^{-8}$

CSE 7306c

Real Networks vs. G_{np}

- Are real networks like random graphs?
  - Giant connected component, and Average path length match
  - Clustering Coefficient and degree distribution are different
- Problems with the random network model
  - Giant component in most real network does NOT emerge through a phase transition
  - Degree distribution differs from real networks
  - No local structure – clustering coefficient is too low

Today’s Agenda

- Introduction to Social Network Analysis
- Erdös-Renyi Model
- **Small World Model and Kleinberg's Model**

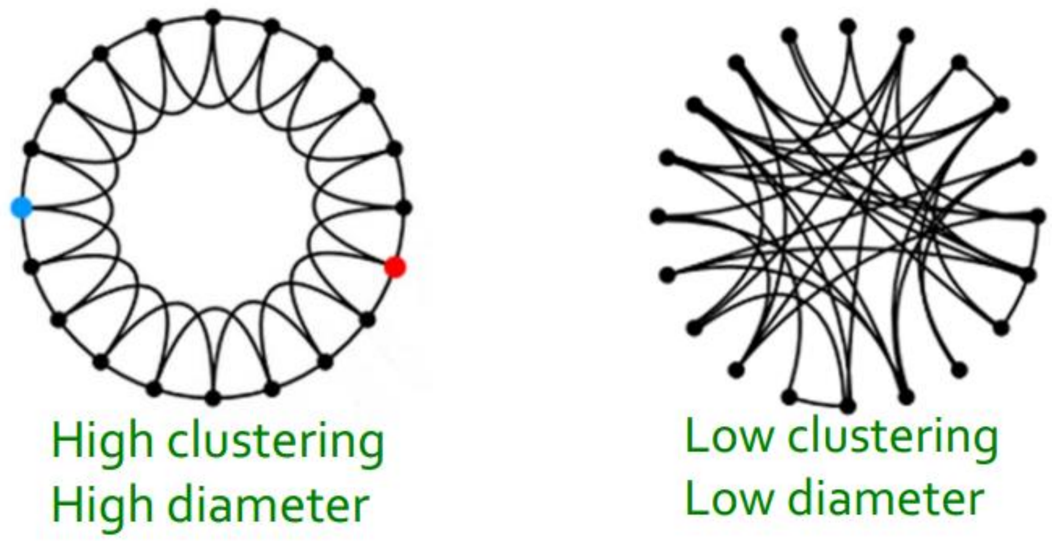
CSE 7306c

Small World Experiment [Milgram '67]

- What is the typical shortest path length between any two people?
- Experiment on the global friendship network
- 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?
  - 64 chains completed (i.e., 64 letters reached the target)
  - It took 6.2 steps on the average,thus "6 degrees of separation“
- In 2003 Dodds, Muhamad and Watts performed the experiment using e-mail
  - Typical path length h = 7

High Clustering vs. Small Diameter

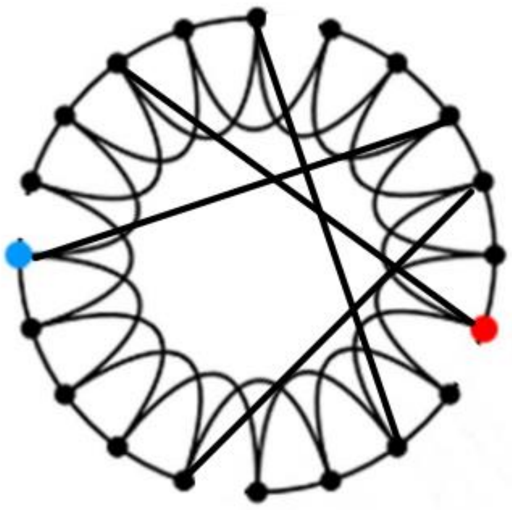
- MSN network has 7 orders of magnitude larger clustering than the corresponding G_{np}
- Real networks show clustering behavior
- Clustering implies high diameter (distance between nodes in different clusters could be high)
- Small diameter means there exist shortcuts which bridge clusters
- Could a 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?



CSE 7306c

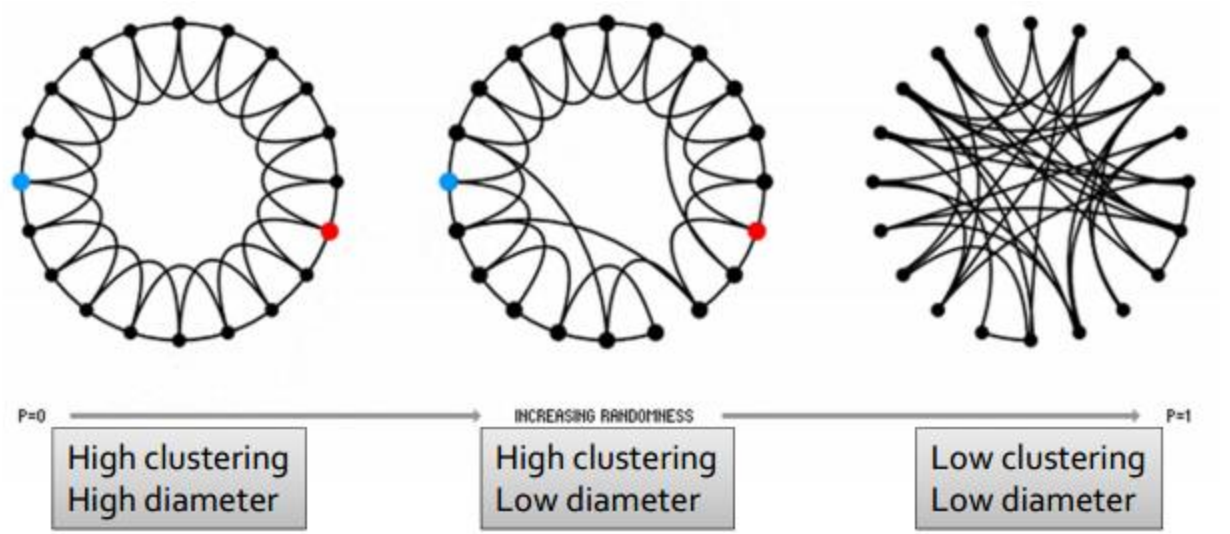
Small World Model [Watts-Strogatz, '98]

- 2 components to the model
  - Start with a low-dimensional regular lattice
    - Has high clustering coefficient
    - Now introduce randomness ("shortcuts")
  - 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



The Small World Model

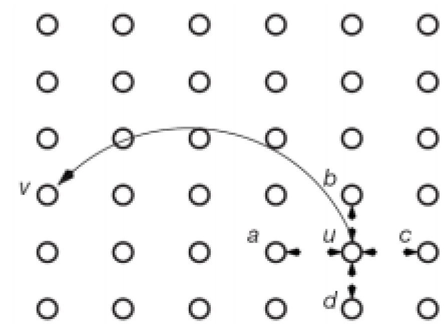
- Rewiring allows us to “interpolate” between a regular lattice and a random graph
- It takes a lot of randomness to ruin the clustering, but a very small amount to create shortcuts.
- Diameter is log n
- Accounts for the high clustering of real networks
- Does not lead to the correct degree distribution
- Does not enable navigation



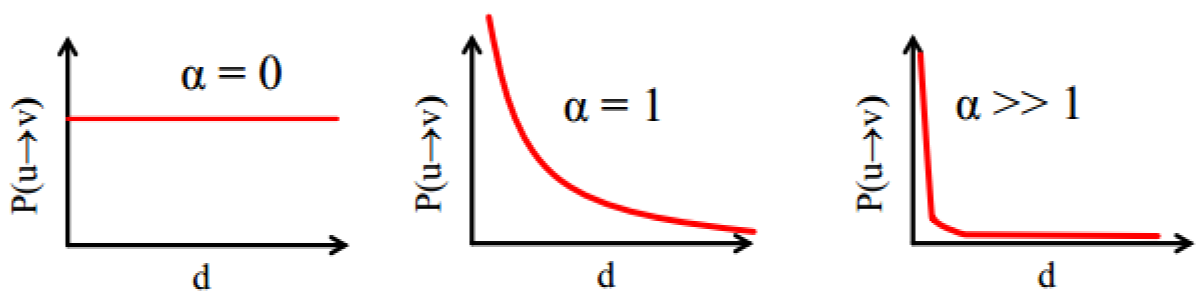
Regular Network　Small World Network　Random Network

Kleinberg's Model

- Watts-Strogatz graphs are not searchable
- How do we make a searchable small-world graph?
- Intuition
  - Our long range links are not random
  - They follow geography!
- Model
  - Nodes on a grid
  - Node has one long range link
  - Prob. of long link to node v: $P(u \to v) \sim d(u,v)^{-\alpha}$
    - d(u,v) is the grid distance between u and v



$$P(u \to v) = \frac{d(u,v)^{-\alpha}}{\sum_{w \neq u} d(u,w)^{-\alpha}}$$



Today’s Agenda

- Introduction to Social Network Analysis
- Erdös-Renyi Model
- Small World Model and Kleinberg's Model
- **Power Laws**

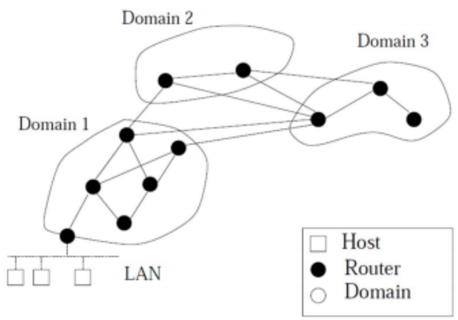
CSE 7306c

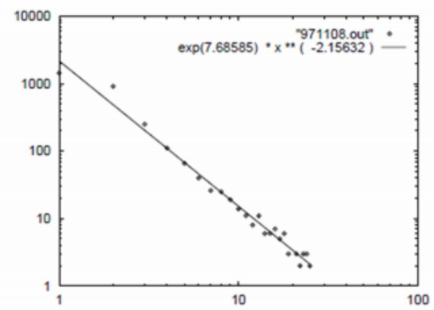
What do we Want?

- Small world
  - Small diameter
  - High Clustering Coefficient
- Searchability
- Power law degree distribution
  - Preferential Attachment
  - $P(k) = k^{-\alpha}$
  - Appears as a straight line on log-log plot

Power Laws Everywhere

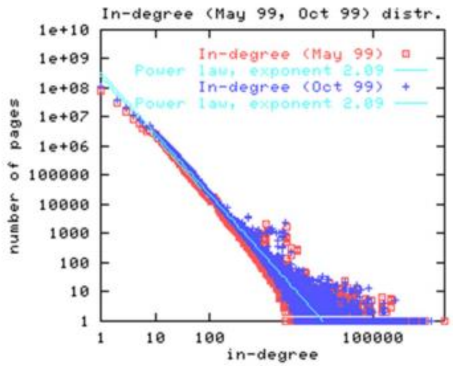
- **Internet Autonomous Systems** [Faloutsos, Faloutsos and Faloutsos, 1999]

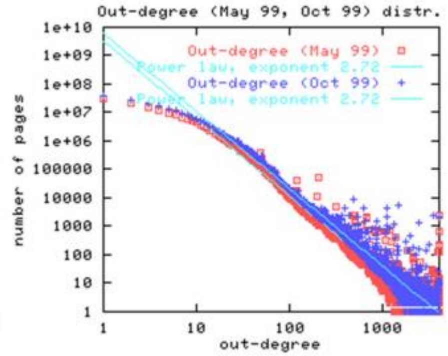




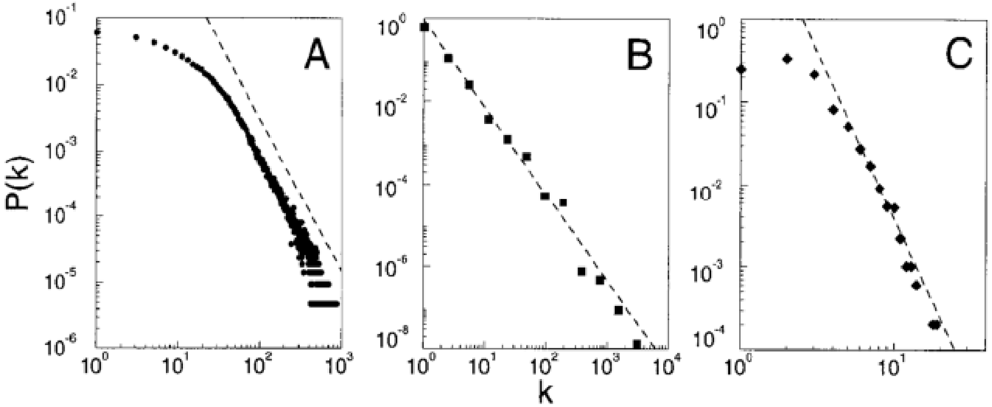
Internet domain topology

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





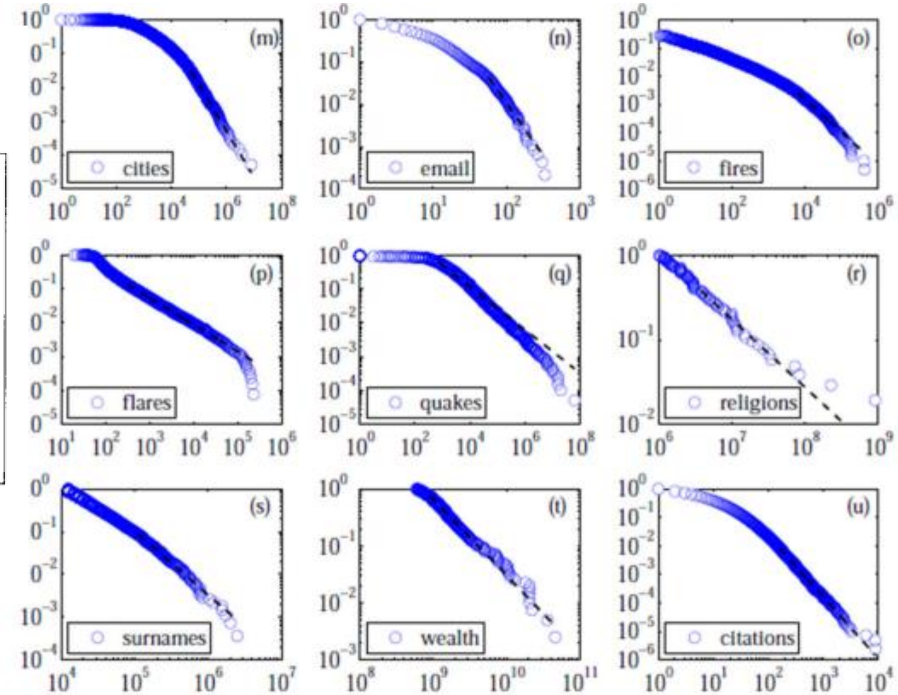
- **Other Networks** [Barabasi-Albert, 1999]



Actor collaborations

Web graph

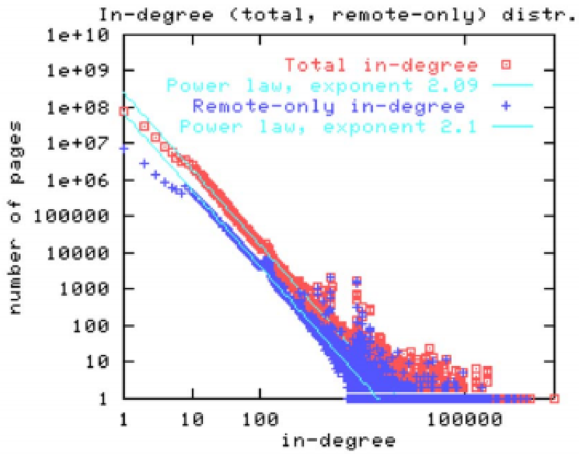
Power-grid

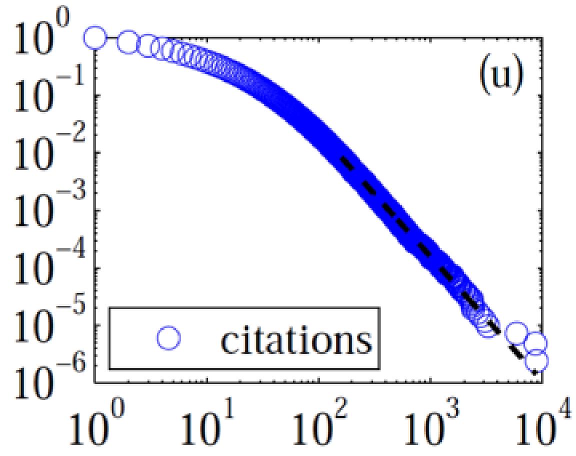


Power Law Degree Exponents

Power-law degree exponent is typically 2 < α < 3

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





CSE 7306c

Today’s Agenda

- Introduction to Social Network Analysis
- Erdös-Renyi Model
- Small World Model and Kleinberg's Model
- Power Laws
- **Preferential Attachment Model**

Preferential Attachment Model (1)

- Nodes arrive in order 1,2,…,n
- At step j, let d_i be the degree of node $i < j$
- A new node j arrives and creates m outlinks
- Probability of j linking to previous node i is proportional to the degree d_i of node i. $P(j \longrightarrow i) = \frac{d_i}{\sum_k d_k}$
- 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)

CSE 7306c

Preferential Attachment Model (2)

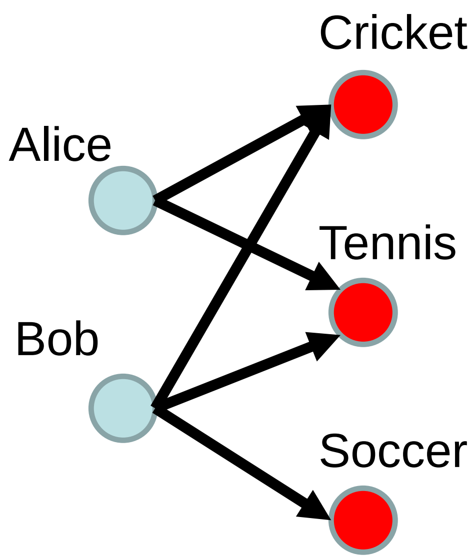
- Other network formation mechanisms that generate scale-free networks
  - Random surfer model [Blum-Mugizi]
  - Copying model [Kleinberg et al.]
  - Forest Fire model [Leskovec et al.]

Today’s Agenda

- Introduction to Social Network Analysis
- Erdös-Renyi Model
- Small World Model and Kleinberg's Model
- Power Laws
- Preferential Attachment Model
- **Copying Model, Forest Fire Model**

Communities in Social Networks

- Edges usually imply endorsement or interest in a topic or a person
  - Users link to pages they care about
    - Two users with similar interest need not know each other
  - Friendship links in social networks
- Communities are dense subgraphs or dense bipartite subgraphs
- Web and social networks are abundant in communities

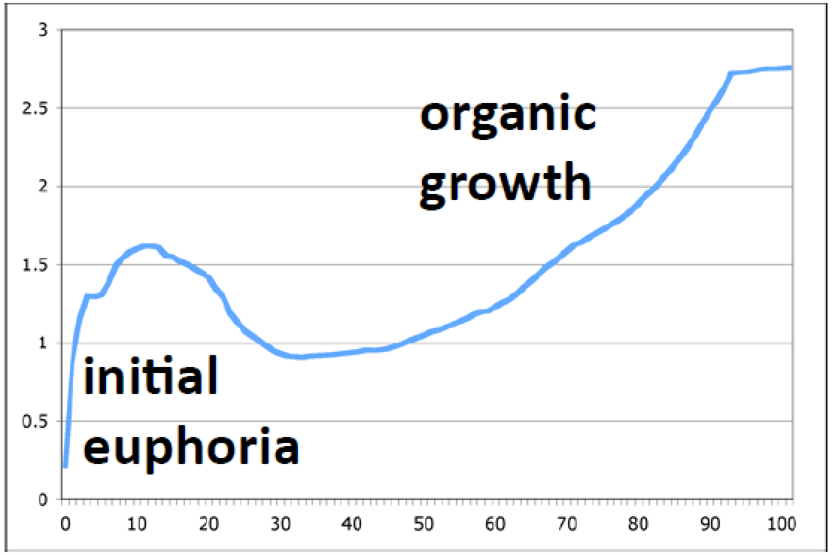


Copying Model [Kumar et al., 2000]

- Observation: People copy their friend's webpage when creating a new one or copy their friend's contacts when joining a social network
- When a new node arrives, it copies edges from a pre-existing node with probability 1 – a and links to the destination of the edge
  - With prob a, it performs preferential attachment.
- The degree distribution is a power-law
- Can explain communities: The number of dense bipartite cliques in this model is large

CSE 7306c

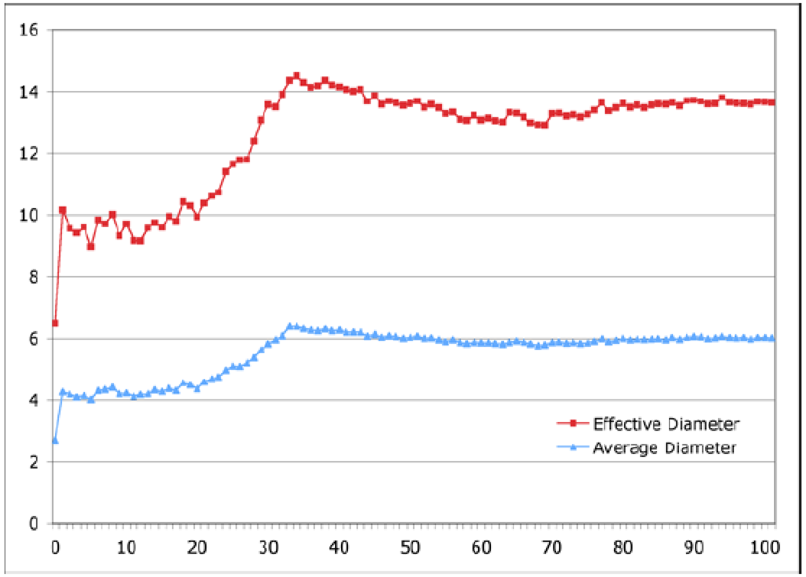
Flickr: Density, Diameter over Time



Density increases over time

Shrinking diameters and densification in citation graphs: Leskovec, Kleinberg, Faloutsos **2005**

Diameter shrinks over time



Forest Fire Model [Leskovec, Kleinberg, Faloutsos 2005]

- Observation: Copying happens beyond one step
- When a new node arrives, it
  - copies an edge from a pre-existing node with prob. 1 - a
  - copies an edge from the destination of the edge
  - …
- An iterated version of the copying model
- In addition to the above, leads to densification and shrinking diameters, in empirical simulations

Take-away Messages

- More and more of all interactions are happening online. Hence social network analysis is important.
- Social network data is complex because of its structure, both static and temporal, and scale.
- Web graph looks like a bow-tie
- We recognized a few important properties of graphs
- We studied random graph model, small world model, Kleinberg's model
- We also studied Power laws
- Preferential attachment model has power law degree distribution but cannot explain communities well.
- Copying model can explain communities
- Forest fire model leads to densification and diameter shrinkage

Further Reading

- http://www.stanford.edu/class/cs224w/handouts.html
- https://www.coursera.org/course/sna
- Book by David Easley and Jon Kleinberg
  - http://www.cs.cornell.edu/home/kleinber/networks-book/
- Barabási, A.-L.; R. Albert (1999). "Emergence of scaling in random networks". *Science* **286** (5439): 509–512
- R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. 2000. Stochastic models for the Web graph. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science* (FOCS '00). IEEE Computer Society, Washington, DC, USA

Further Reading

- Barabási, A.-L.; R. Albert (1999). "Emergence of scaling in random networks". Science 286 (5439): 509–512
- R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. 2000. Stochastic models for the Web graph. In Proceedings of the 41st Annual Symposium on Foundations of Computer Science (FOCS '00). IEEE Computer Society, Washington, DC, USA
- Jure Leskovec, Deepayan Chakrabarti, Jon M. Kleinberg, Christos Faloutsos: Realistic, Mathematically Tractable Graph Generation and Evolution, Using Kronecker Multiplication. PKDD 2005: 133-145
- Jure Leskovec, Jon M. Kleinberg, Christos Faloutsos: Graphs over time: densification laws, shrinking diameters and possible explanations. KDD 2005: 177-187
- Jure Leskovec, Jon M. Kleinberg, Christos Faloutsos: Graph evolution: Densification and shrinking diameters. TKDD 1(1) (2007)
- Paolo Boldi, Sebastiano Vigna: The webgraph framework I: compression techniques. WWW 2004: 595-602

International School of Engineering

Plot 63/A, 1st Floor, Road # 13, Film Nagar, Jubilee Hills, Hyderabad - 500 033

For Individuals:	+91-9502334561/63 or 040-65743991
For Corporates:	+91-9618483483
Web:	http://www.insofe.edu.in
Facebook:	https://www.facebook.com/insofe
Twitter:	https://twitter.com/Insofeedu
YouTube:	http://www.youtube.com/InsofeVideos
SlideShare:	http://www.slideshare.net/INSOFE
LinkedIn:	http://www.linkedin.com/company/international-school-of-engineering