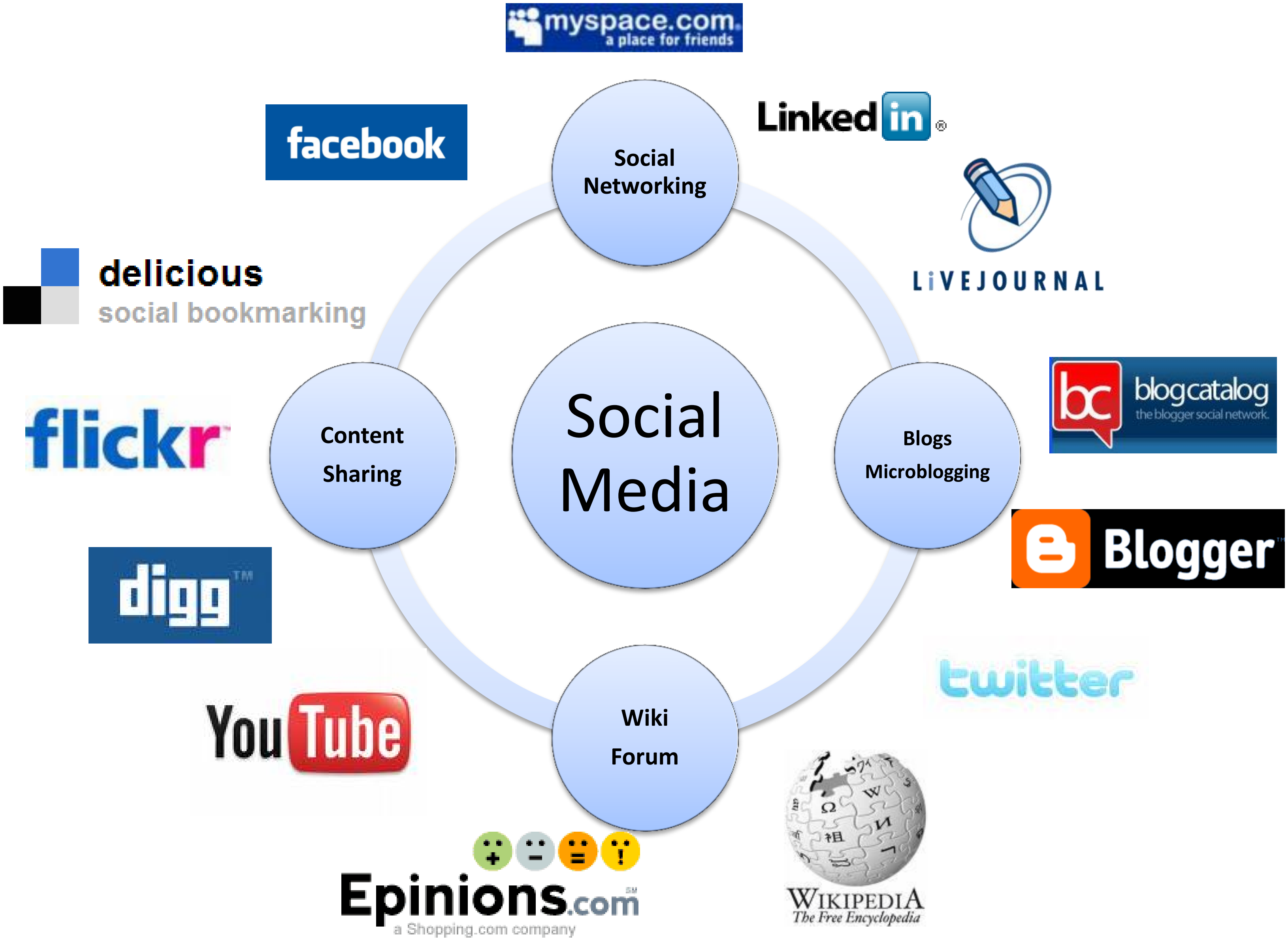
Social Networks and Social Media

香港科技大學
HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY

Social Media: Many-to-Many



Characteristics of Social Media

- "Consumers" become "Producers"
- Rich User Interaction
- User-Generated Contents
- Collaborative environment
- Collective Wisdom
- Long Tail

Broadcast Media
Filter, then Publish

→

Social Media
Publish, then Filter

Top 20 Websites at USA

1	Google.com	11	Blogger.com
2	Facebook.com	12	msn.com
3	Yahoo.com	13	Myspace.com
4	YouTube.com	14	Go.com
5	Amazon.com	15	Bing.com
6	Wikipedia.org	16	AOL.com
7	Craigslist.org	17	LinkedIn.com
8	Twitter.com	18	CNN.com
9	Ebay.com	19	Espn.go.com
10	Live.com	20	Wordpress.com

40% of websites are social media sites

What is Social Network and Social Media

- Social Network
  - The networks formed by individuals
- Social Media
  - social network + media
    - media = content of twitter, tag, videos, photos ....

Statistical Properties of Social Networks

Why do statistics

- To understand the networks
  - Understand their topology and measure their properties
  - Study their evolution and dynamics
  - Create realistic models
  - Create algorithms that make use of the network structure

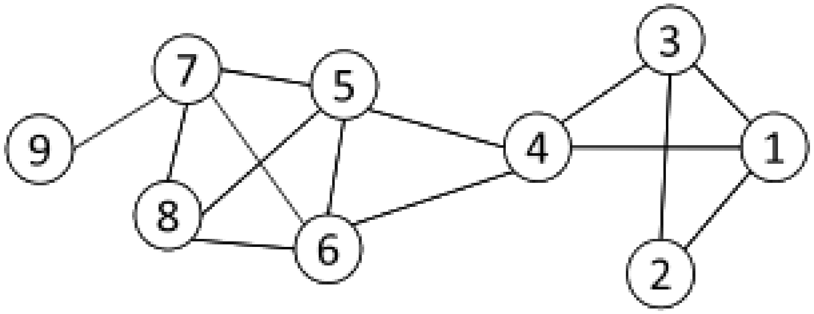
Interesting Questions: demonstration

- Some interesting questions
  - What do social networks look like, on a large scale?
  - How do networks behave over time?
  - How do the non-giant weakly connected components behave over time?
  - What distributions and patterns do weighted graphs maintain?

Networks and Representation

Social Network: A social structure made of nodes (individuals or organizations) and edges that connect nodes in various relationships like friendship, kinship etc.

- Graph Representation
- Matrix Representation



Node	1	2	3	4	5	6	7	8	9
1	-	1	1	1	0	0	0	0	0
2	1	-	1	0	0	0	0	0	0
3	1	1	-	1	0	0	0	0	0
4	1	0	1	-	1	1	0	0	0
5	0	0	0	1	-	1	1	1	0
6	0	0	0	1	1	-	1	1	0
7	0	0	0	0	1	1	-	1	1
8	0	0	0	0	1	1	1	-	0
9	0	0	0	0	0	0	1	0	-

Basic Concepts

- A: the adjacency matrix
- V: the set of nodes
- E: the set of edges
- v_i: a node v_i
- $e(v_i, v_j)$: an edge between node v_i and v_j
- N_i: the neighborhood of node v_i
- d_i: the degree of node v_i
- geodesic: a shortest path between two nodes
  - geodesic distance

Statistical Properties

- Static analysis
  - Static snapshots of graphs
- Dynamic analysis
  - A series of snapshots of graphs

Some famous properties

1. ‘Small-world’ phenomenon
   – An Experiment by Milgram (1967)
      - Asked randomly chosen “starters” to forward a letter to the target
      - Name, address, and some personal information were provided for the target person
      - The participants could only forward a letter to a single person that he/she knew on a first name basis
      - Goal: To advance the letter to the target as quickly as possible

The Milgram Experiment (Wikipedia)

Detailed procedure

1. Milgram typically chose individuals in the U.S. cities of Omaha, Nebraska and Wichita, Kansas to be the starting points and Boston, Massachusetts to be the end point of a chain of correspondence
   – because they were thought to represent a great distance in the United States, both socially and geographically.
2. Information packets were initially sent to "randomly" selected individuals in Omaha or Wichita. They included letters, which detailed the study's purpose, and basic information about a target contact person in Boston.
   – It additionally contained a roster on which they could write their own name, as well as business reply cards that were pre-addressed to Harvard.

The Milgram Experiment (cont.)

3. Upon receiving the invitation to participate, the recipient was asked whether he or she personally knew the contact person described in the letter.
   – If so, the person was to forward the letter directly to that person. For the purposes of this study, knowing someone "personally" was defined as knowing them on a first-name basis.
4. In the more likely case that the person did not personally know the target, then the person was to think of a friend or relative they know personally that is more likely to know the target.
   – A postcard was also mailed to the researchers at Harvard so that they could track the chain's progression toward the target.

The Milgram Experiment

5. When and if the package eventually reached the contact person in Boston, the researchers could examine the roster to count the number of times it had been forwarded from person to person.
   – Additionally, for packages that never reached the destination, the incoming postcards helped identify the break point in the chain.

Result of the Experiment

- However, a significant problem was that often people refused to pass the letter forward, and thus the chain never reached its destination.
- In one case, 232 of the 296 letters never reached the destination.[3]
- However, 64 of the letters eventually did reach the target contact.
- Among these chains, the average path length fell around 5.5 or six.

Some famous properties con’t

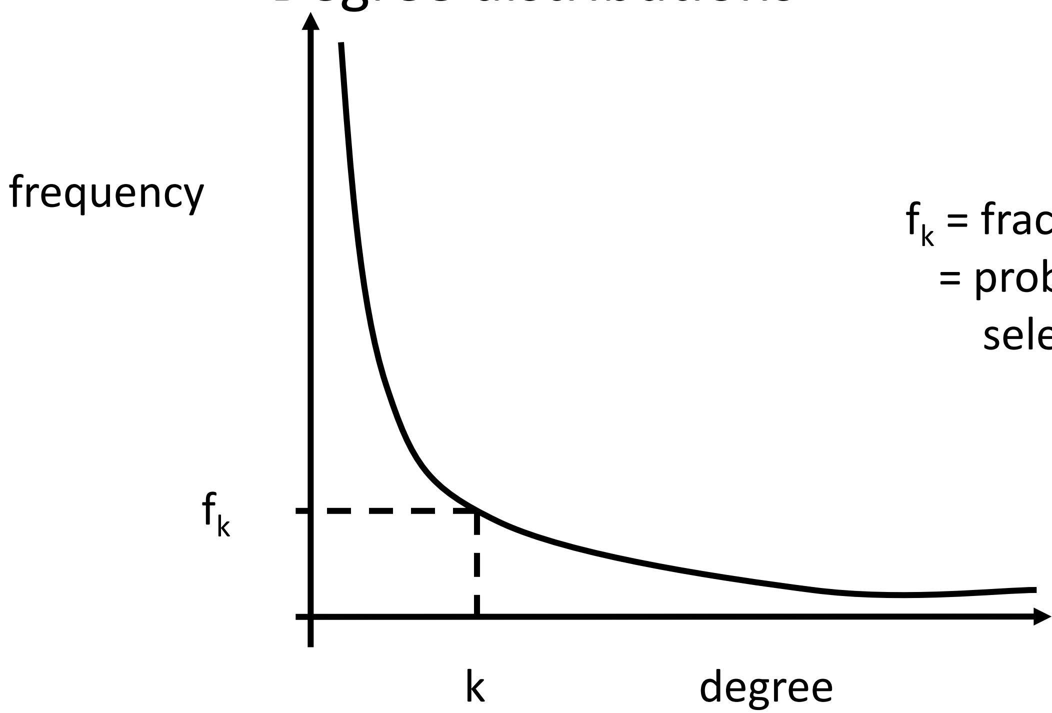
1. ‘Small-world’ phenomenon
   – Property
     - Any two people can be connected within 6 hops

       six degrees of separation

   – Verified on a planetary-scale IM network of 180 million users (Leskovec and Horvitz 2008)
     - The average path length is 6.6

Some famous properties con’t

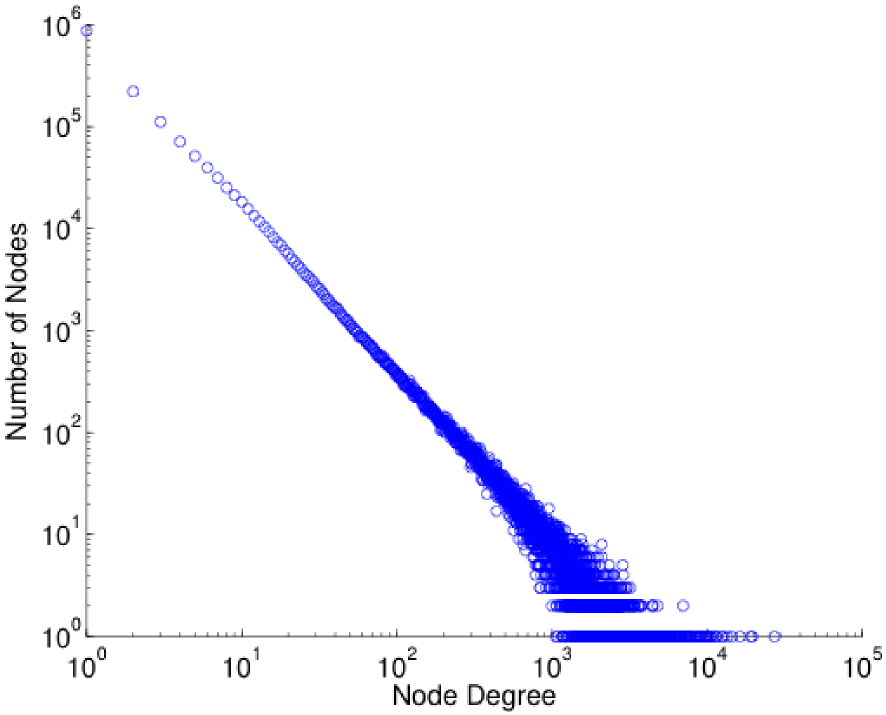
2. ‘Power-law’ degree distributions
   - Degree distributions



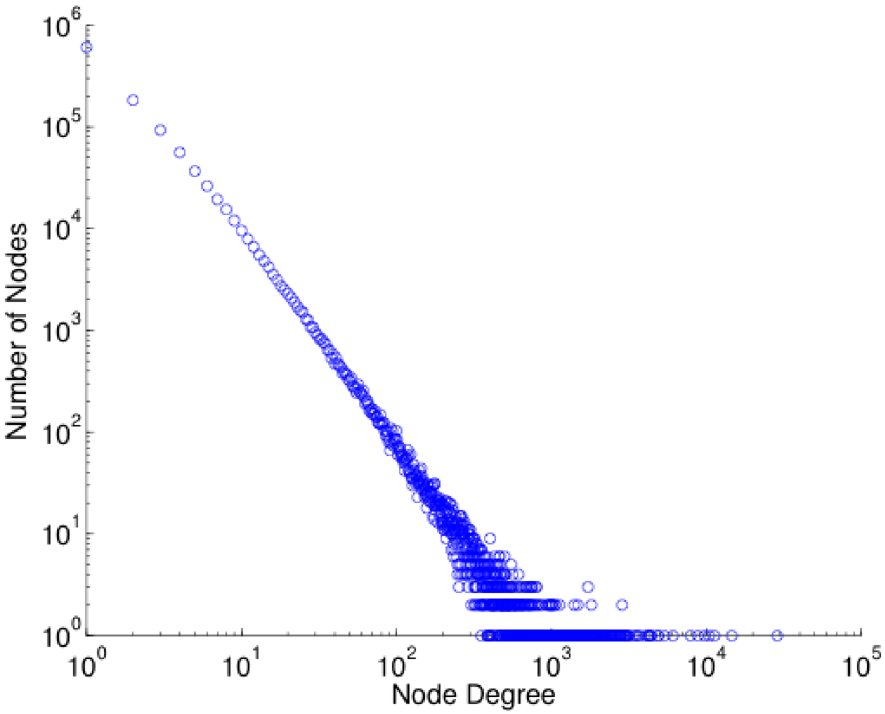
f_k = fraction of nodes with degree k
= probability of a randomly selected node to have degree k

log-log plot

- Power law distribution becomes a straight line if plot in a log-log scale



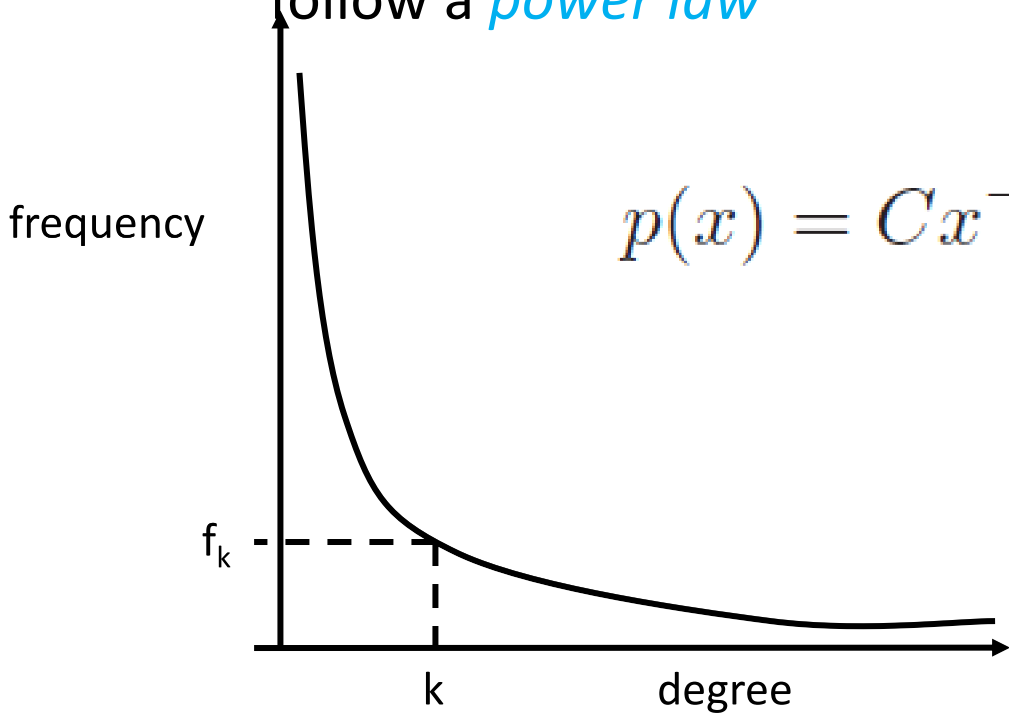
Friendship Network in Flickr



Friendship Network in YouTube

Some famous properties con't

2. 'Power-law' degree distributions
   - The degree distributions of most real-life networks follow a *power law*



$$p(x) = Cx^{-\alpha}, \;\; x \geq x_{min}, \;\; \alpha > 1$$

Some other properties

3. Triangle Power Law

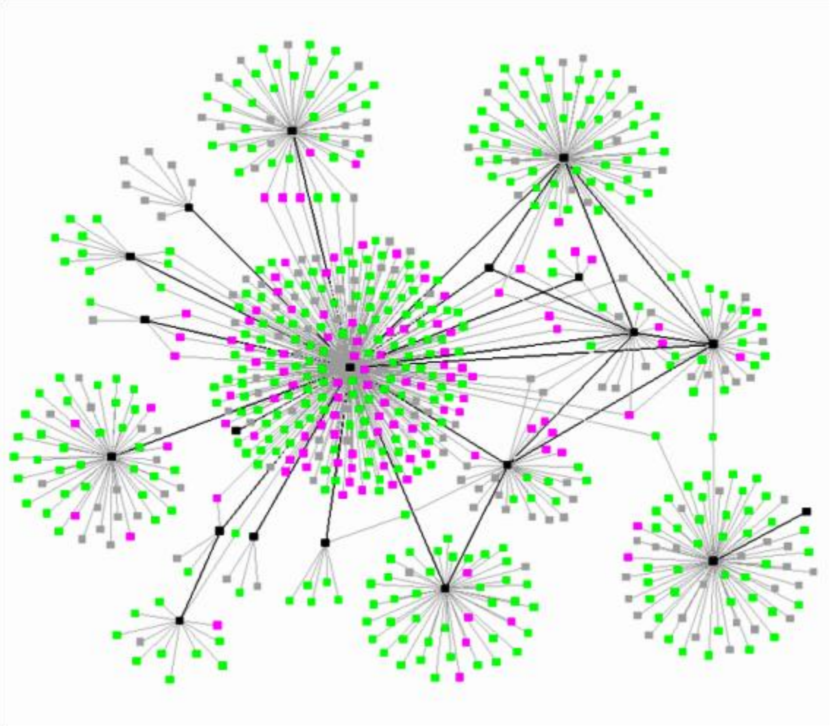
$$f(\Delta) \propto \Delta^{\alpha} \qquad \alpha < 0$$

Some other properties

4. Eigenvalue Power Law
   - The 20 or so largest eigenvalues of the adjacency matrix are power law distributed
   - This is consequence of the “Degree Power Law”

Some other properties

5. Community Structure
   – Social networks are modular
     • i.e nodes form communities

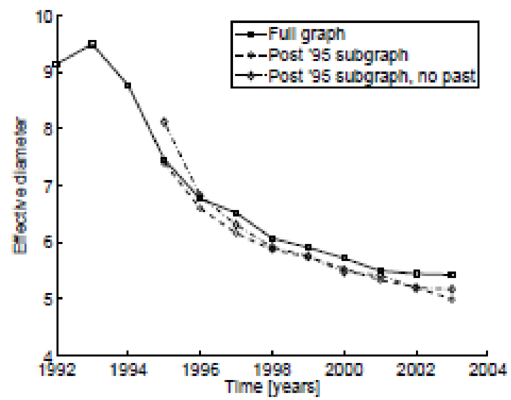


Statistical Properties

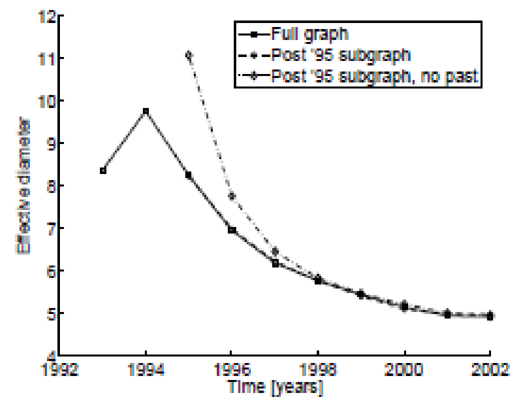
- Static analysis
  - Static snapshots of graphs
- Dynamic analysis
  - A series of snapshots of graphs

Properties

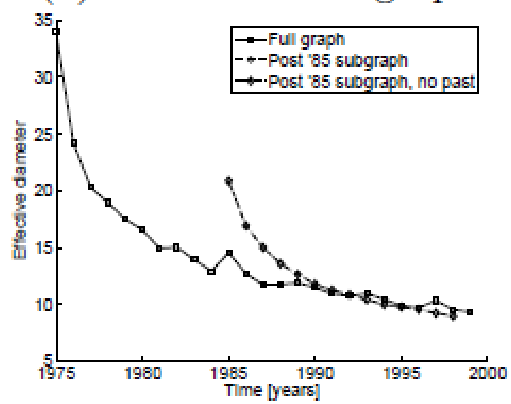
1. Shrinking Diameter



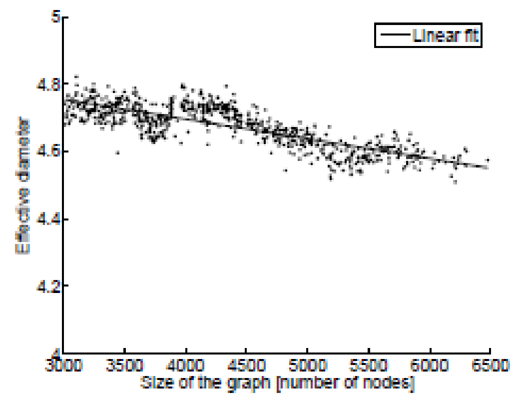
(a) arXiv citation graph



(b) Affiliation network



(c) Patents

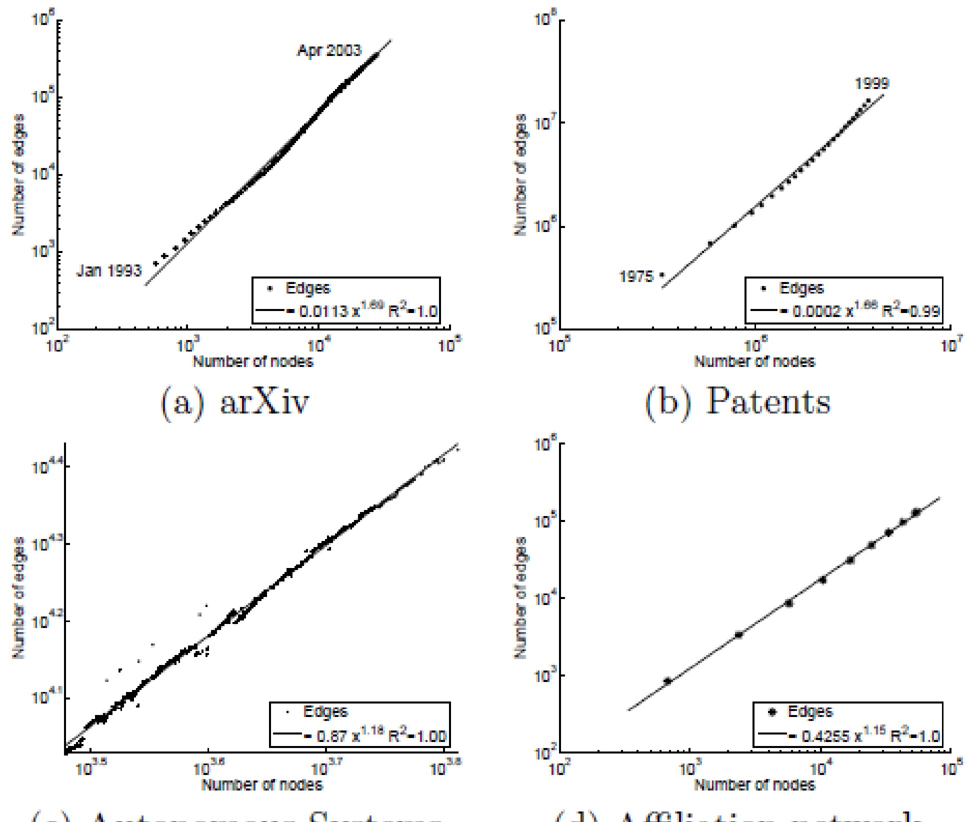


(d) AS

Properties con't

2. Densification Power Law (DPL)
   - E(t): the number of edges
   - N(t): the number of nodes

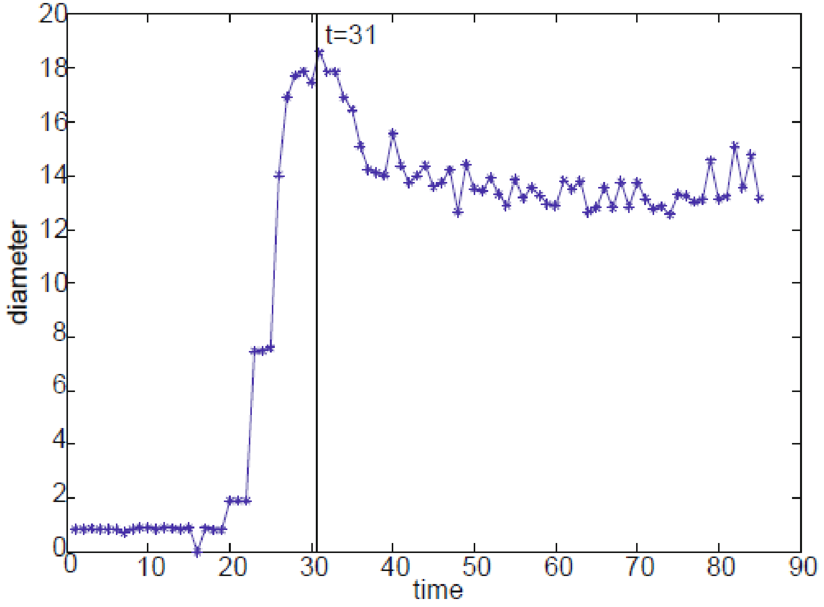
$$E(t) \propto N(t)^{\beta}$$



(a) arXiv (b) Patents

Properties con't

3. Diameter-plot and Gelling point
   - Graph forming
     - Establishment period
     - Stable period



(a) Diameter(t)

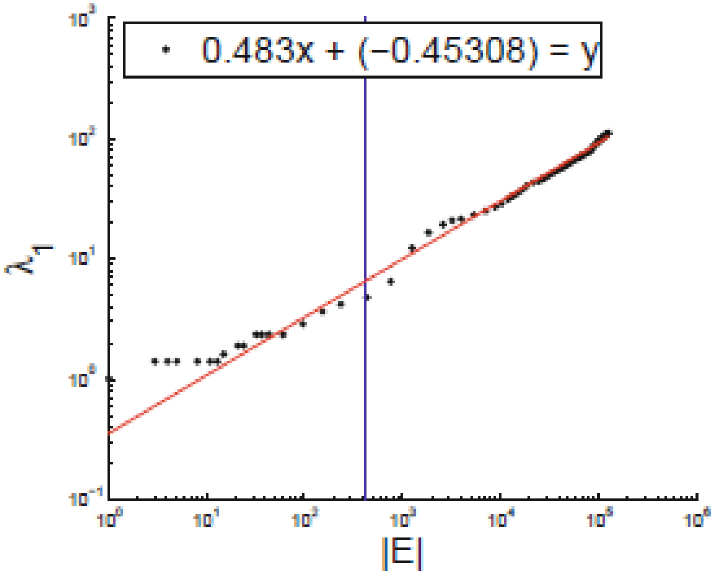
Properties con’t

4. Constant/Oscillating Connected Components (CC)
   – Largest: constant
   – Second/Third: oscillation

Properties con't

5. Principal eigenvalue over time
   – E(t): the number of edges
   – $\lambda_1(t)$: the largest eigenvalue

$$\lambda_1(t) \propto E(t)^{\alpha} \qquad \alpha \leq 0.5$$



(b) Blog Network

Conclusion

- Usefulness of the statistical properties
  - Understanding human behaviors
  - Anomalous graphs/subgraphs detection
  - Identifying authorities and search algorithms
  - Prepare resources based on the prediction
  - …