Graph, Network

Network Structure (Static)

KENTECH
Korea Institute of Energy Technology
AI

- Characteristics of networks
  - Node degree distribution
  - Clusters
  - Diameter, sparse/dense, seachability
  - Attribute distributions
  - …
- How to generate networks with such properties?
  - Node generation
  - Edge generation

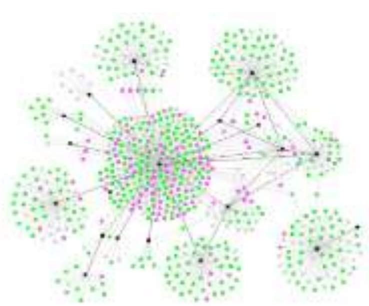
Dynamics of Networks

- How do nodes interact?
- Behaviors of nodes
  - Diffusion of information
  - Spread of virus
  - Vote
  - Reference (Citation)
  - …

Epidemiology

- Hippocrates
  - Epidemic & Endemic

Kermack, W.O. and McKendrick, A.G. (1927).
A contribution to the mathematical theory of epidemics I. *Proc.Roy.Soc. A* 115, 700-721.

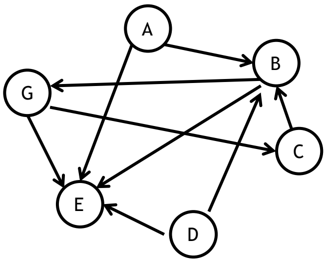


John Snow:
London Cholera, 1854

Representation

AI

- Text, table, …
  - Adam follows Bob and Eve
  - Bob follows Eve and Genie
  - Charlie follows Bob
  - Dave follows Bob and Eve
  - Genie follows Charlie and Eve
- Graphic



Matrix

AI

- Adjacency Matrix A: element A_{ij} is relation between node i and j

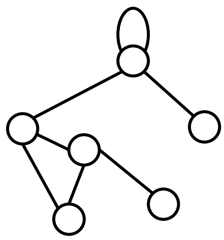
	A	B	C	D	F	G
A	0	1	0	0	1	0
B	0	0	0	0	1	1
C	0	1	0	0	0	0
D	0	1	0	0	1	0
E	0	0	0	0	0	0
G	0	0	1	0	0	1

Good for Analysis

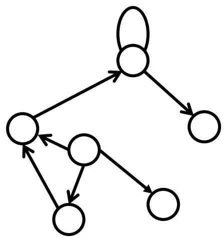
Directed/Undirected Networks

KENTECH Korea Institute of Energy Technology AI

- Undirected networks
  - Links have no direction
  - Co-authoring of papers



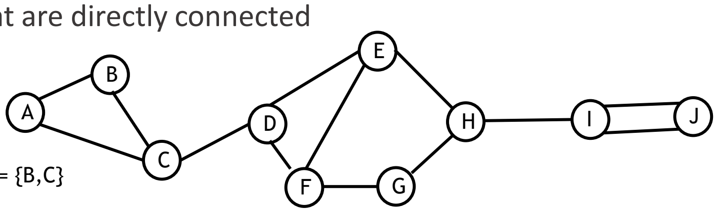
- Directed networks
  - Directed links (arcs)
  - Following
  - Refer



Neighbors

KENTECH Korea Institute of Energy Technology AI

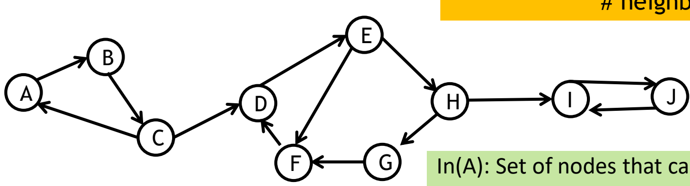
- Nodes that are directly connected



N(A) = {B,C}

Node degree: # edges connected to a node
neighbor nodes ≅ |N(A)|

- Directed graph



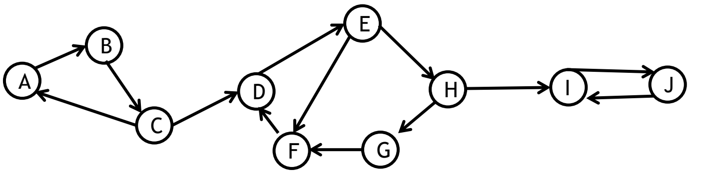
In(A): Set of nodes that can reach to node A

Out(A): Set of nodes that can be reached from node A

Connectivity of Directed Networks

KENTECH Korea Institute of Energy Technology AI

- Strongly connected
  - There are (directed) paths from one node to another node
- Weakly connected
  - Connected if edges are replaced by undirected edges



- **DAG** (Directed Acyclic Graph)
  - Directed graph without cycles
  - If there is a path from *u* to *v*, then there is no path from *v* to *u*

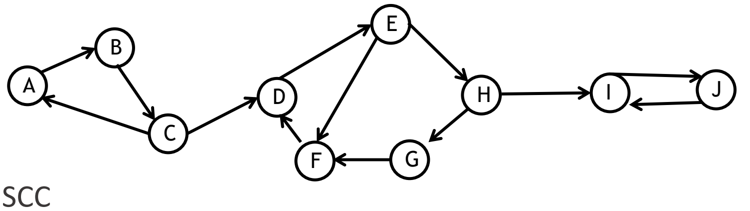
Korea Institute of Energy Technology

9

SCC (Strongly Connected Component)

KENTECH Korea Institute of Energy Technology AI

- A set *S* of nodes
  - Strongly connected
  - Largest set containing S with the strong connectivity property



- Find SCC
  - An SCC containing node A = In(A) ∩ Out(A)

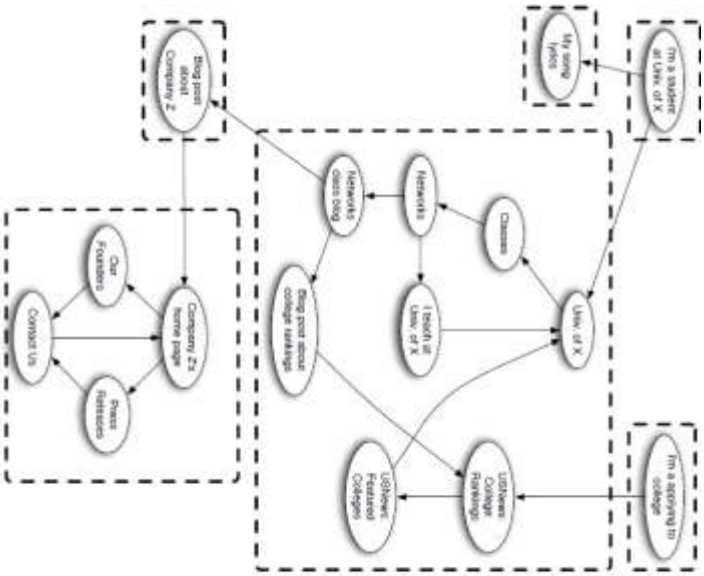
Note: A node belongs to only one SCC

Korea Institute of Energy Technology

10

The Structure of the Web

KENTECH Korea Institute of Energy Technology AI



Weighted, Multigraph

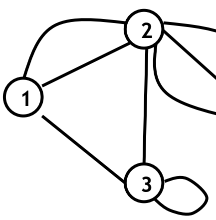
KENTECH Korea Institute of Energy Technology AI

- Weighted graph
  - Links have weights

$$E = \sum \sum A_{ij}$$

$$\bar{k} = {}^{2E}/_{N}$$

- Multigraph
  - More than one links between a node pair
  - Weight = # of links

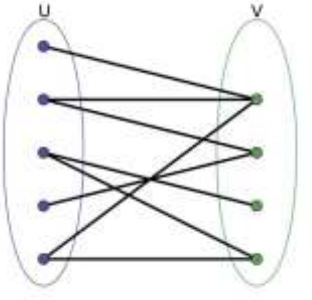


$$A_{ij} = \begin{bmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 1 & 0 \\ 0 & 3 & 0 & 0 \end{bmatrix}$$

Bipartite Graph

- Two types of nodes
- Edge between nodes of different types
  - U, V are independent sets
- Examples
  - Actor-Movie
  - Author-paper
  - Person-Community
- Incident matrix

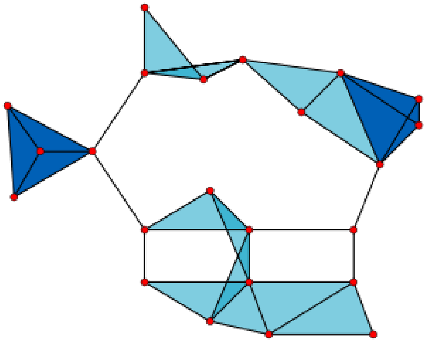
  Aij = 1 if member j belongs to group i

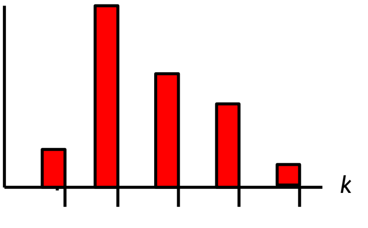


Network Property: Degree Dist.

- *p(k)* = Prob. that a node degree = *k*

$$N_k = \#\ of\ nodes\ with\ degree = k$$
$$p(k) = {}^{N_k}/_{N}$$





Power Law Distribution

Path

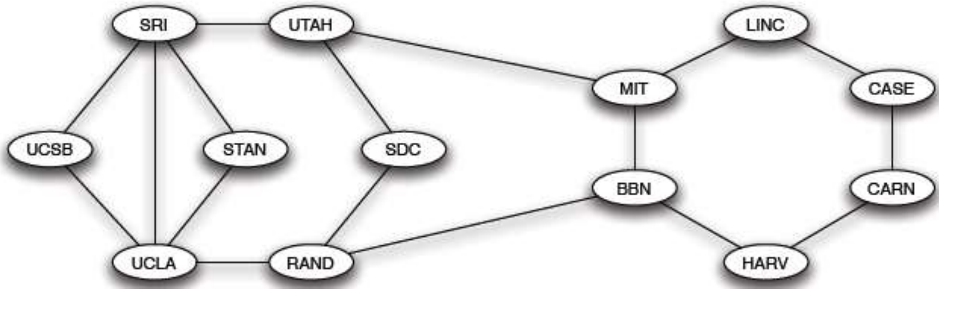
KENTECH Korea Institute of Energy Technology AI

- Definition:
  - Sequence of nodes that are directly connected
- Path length, h
  - Number of hops (≡ # of edges ≡ # of nodes -1)
- Directly connected node pair (u, v)
  - Path length =1
  - $A_{uv} = 1$
- u, v are connected by a path of length h
  - ${A^h}_{uv}$=1
- Shortest path
  - Path of smallest length

BFS (Breath First Search)

KENTECH Korea Institute of Energy Technology AI

- Finding shortest paths in weighted graph
  - Dijkstra algorithms, Bellman-Ford, Floyd-Warshall, …
- In unweighted graph
  - BFS

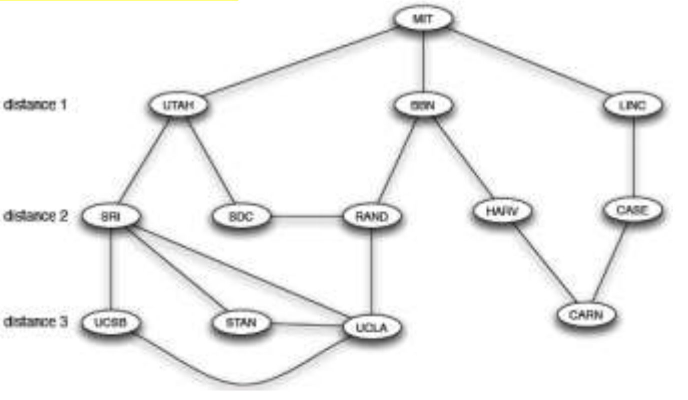


BFS

KENTECH Korea Institute of Energy Technology AI

Shortest path tree from node u
Mark u, let h(u) = 0, enqueue(u)

While queue not empty
dequeue(v), retrieve its unmarked neighbor, w
h(w) = h(v)+1
mark w
enqueue(w)



Network Diameter

KENTECH Korea Institute of Energy Technology AI

- Diameter Definition 1:
  - The length of the longest path in a graph
- Diameter Definition 2:
  - Average path length of connected pairs

$$\bar{h} = \frac{1}{N(N-1)} \sum \sum h_{ij}$$

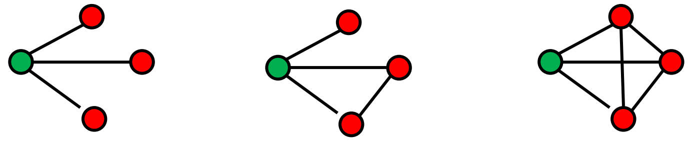
Complex
Use sampling to reduce complexity

Clustering Coefficient

AI

triangles

- Clustering coefficient
  - Proportion of directly connected neighbor pairs
  - $C_i = \frac{2\,e_i}{k_i(k_i-1)}$ where k_i is the degree of node i and e_i is the number of directed connected neighbor pairs



- $Ci = 0$ $Ci = 1/3$ $Ci = 1$

- Average clustering coefficient

$$C = \frac{1}{N}\sum C_i$$

Random Network & Basic Graph Theory

C. Kim

Erdös-Renyi Random Graph

- Erdös-Renyi - 1960

  http://en.wikipedia.org/wiki/Paul_Erd%C5%91s

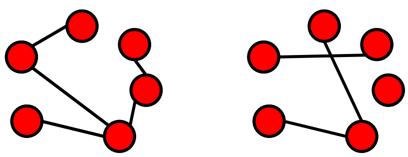
  http://en.wikipedia.org/wiki/File:Alfred_Renyi.jpg

- Two variants
  - $G_{n,p}$: Undirected graph on n nodes each edge (u,v) appears i.i.d. with probability p
  - $G_{n,m}$: Undirected graph with n nodes and m uniformly and randomly picked edges

Random Graph, $G_{n,p}$

AI

- Construction
  - Connect each n·(n-1)/2 node pairs with prob. p



Two examples of N=6, p=0.3

Random networks vs Real networks

Properties of Random Graph

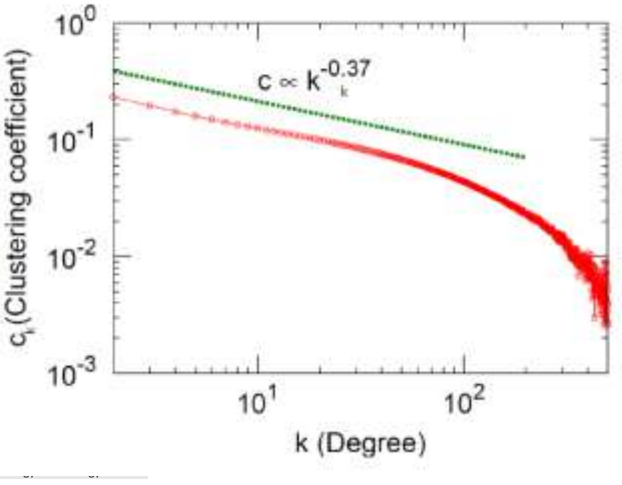
KENTECH Korea Institute of Energy Technology AI

- Node degree distribution
- Clustering coefficient
- Size of giant components, prob. of isolated nodes
- Diameter
- …

Properties of $G_{n,p}$ – Clustering Coefficient

KENTECH Korea Institute of Energy Technology AI

- Expected clustering coefficient ?
- Real networks



Avg. clustering of the MSN: C = 0.1140

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

Properties of $G_{n,p}$ – Node Degree

KENTECH Korea Institute of Energy Technology AI

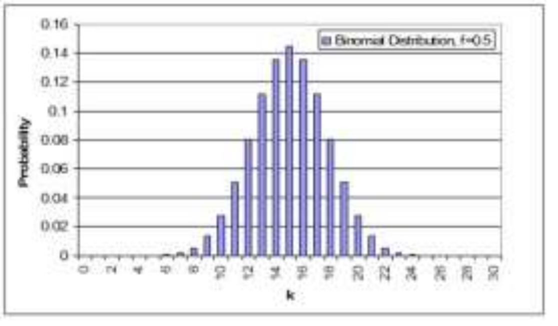
- Prob. that $G_{n,p}$ generates a graph with E edges

$$P(E) = \binom{E^{max}}{E} p^E (1-p)^{E^{max}-E}$$

➔ Binomial distribution

- Pr. that a node has k edges

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



Real Networks – Degree Distribution

KENTECH Korea Institute of Energy Technology AI

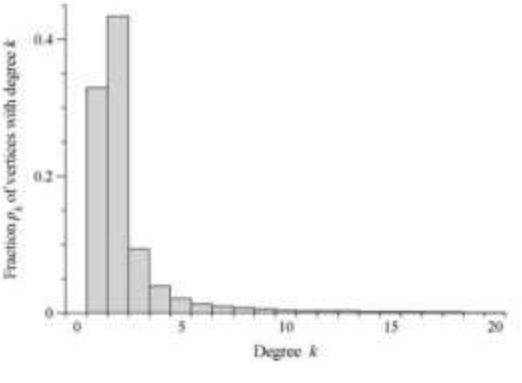


Figure 8.3: The degree distribution of the Internet. A histogram of the degree distribution of the vertices of the Internet graph at the level of autonomous systems.

Real Networks – Degree Distribution

KENTECH Korea Institute of Energy Technology AI

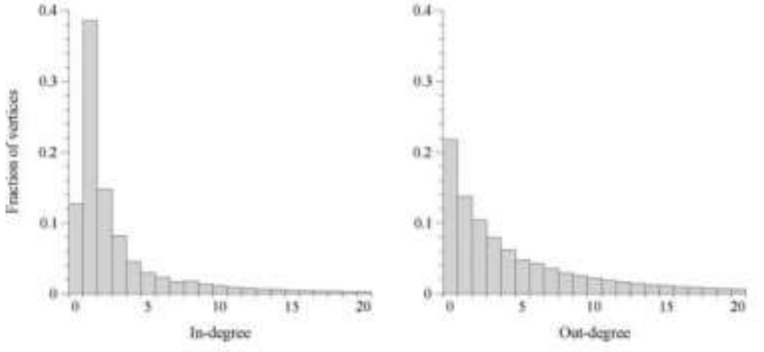


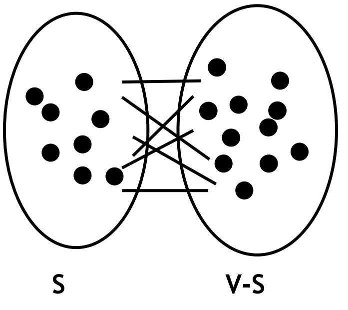
Figure 8.4: The degree distributions of the World Wide Web. Histograms of the distributions of in–and out-degrees of pages on the World Wide Web. Data are from the study by Broder *et al.* [56].

Properties of $G_{n,p}$ - Diameter

KENTECH Korea Institute of Energy Technology AI

- First define expansion
- Graph G(V, E) has expansion *a*
  - If $\forall S \subseteq V$, # of edges leaving S ≥ a * min(|S|, |V-S|)

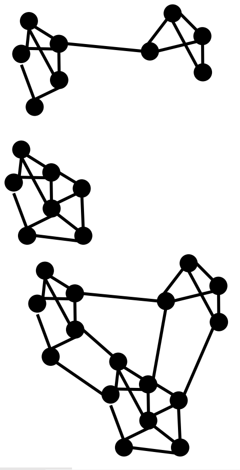
$$\rightarrow a = \min \frac{\#\, of\ edges\ leaving\ S}{\min(|S|,|V-S|)}$$



Expansion

AI

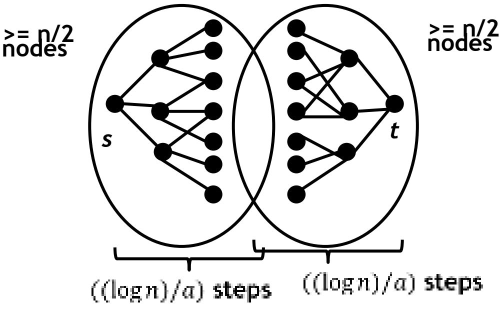
- Measure of Robustness
  - To partition *m* nodes, should cut at least *a***m* edges



Expansion & Diameter - 1

AI

- Theorem: The diameter of a graph on *n* nodes with expansion a is $O((\log n)/a)$
- Proof
  - To show the length of a shortest path between any node pair (s, t) is $O((\log n)/a)$



Expansion & Diameter - 2

KENTECH
Korea Institute of Energy Technology
AI

- Let S_j be a set of all nodes found within j steps of BFS from s

  → $|S_{j+1}| \geq |S_j| + \frac{a * |S_j|}{k}$

  $= |S_j|\left(1 + \frac{a}{k}\right)$

  $= (1 + \frac{a}{k})^{j+1}$

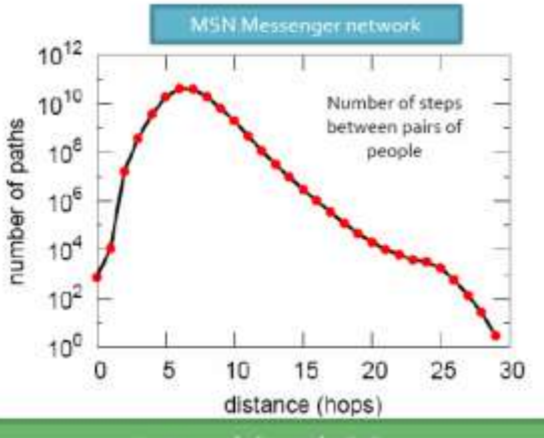
How many BFS steps are required to cover > n/2 nodes?

$(1 + \frac{a}{k})^j$ **>= n/2**

→ **Let** $j = \frac{k\, log_2 n}{a}$ **, then** $(1 + \frac{a}{k})^{\frac{k\, log_2 n}{a}}$ **>= n**

Real Network - Diameter

KENTECH
Korea Institute of Energy Technology
AI



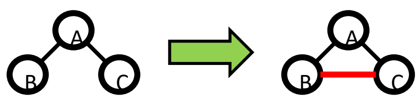
Avg. path length 6.6
90% of the people can be reached in < 8 hops

Homophily in Networks

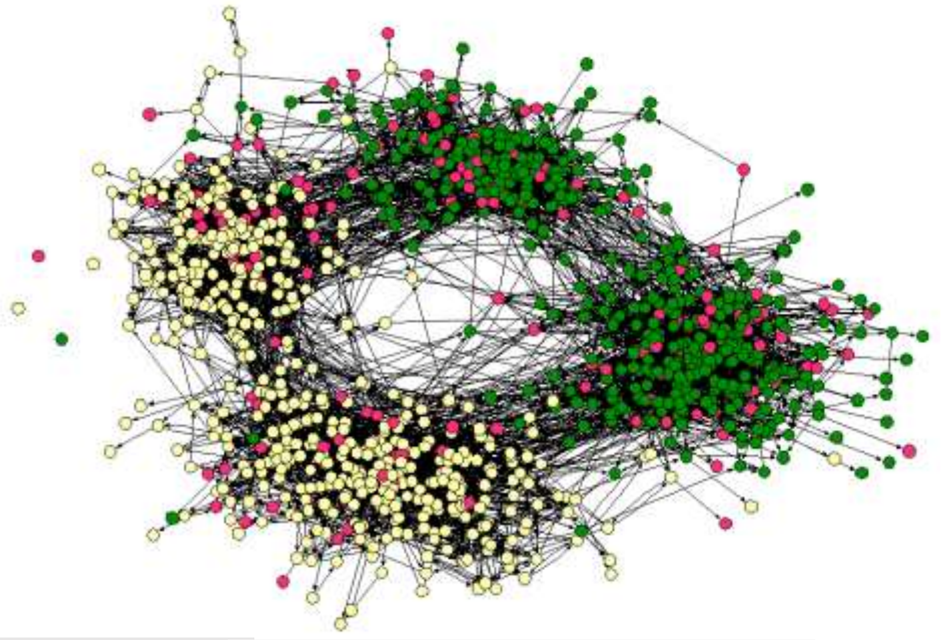
C. Kim

Homophily

- We observed triadic closures
  - People with common friends become friend
- Why?
  - Intrinsic to the network
    - Introduction by a common friend
  - Surrounding contexts (Property outside of network)
    - Same school, same region, …
    - A-B, A-C are similar → B-C are similar too



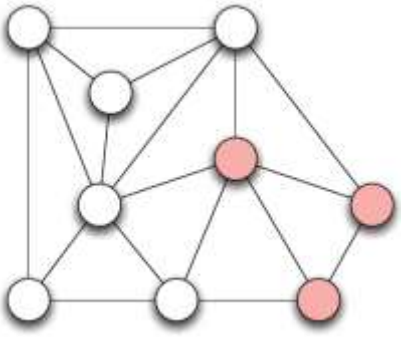
- Surrounding Context affects network formation
- Homophily:
  - Links are formed between nodes of similarity and compatibility
  - Plato "Similarity begets friendship"



Homophily Test

- Test if nodes of the same context form links
- Let populations of attribute A and B are p and q, resp.
  - If no homophily, then
    A-A link : B-B link : A-B link = $p^2 : q^2 : 2pq$
- Test
  - The fraction of A-B links < 2pq,
    then homophily exist
  - The fraction of A-B links > 2pq,
    reverse homophily

P=2/3, q=1/3
➔ 2pq= 4/9
|E|=18
Heterogeneous edge= 5



Selection & Social Influence

- Note: Links connect people with similar characteristics
- Why?

1. Selection: Select friends with similar characteristics
  - Individual characteristics drive the formation of links

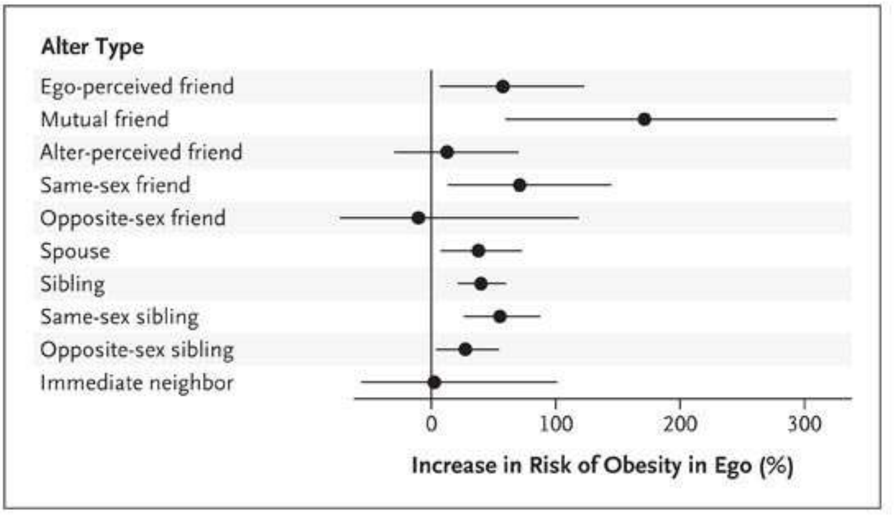
2. Social influence: Modify behaviors adapting to friends' behavior
  - Links shape characteristics

- What factor is more dominant?
- Longitudinal study
  - Social connections and behaviors within a group are traced over time

Obesity & Social Network

- Christakis & Fowler, "The spread of obesity in a large social network over 32 years", NEJM, 2007

Obesity & Social Network - 2



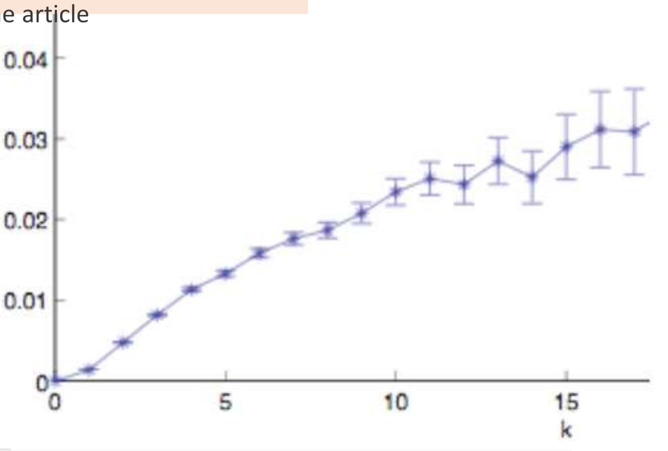
Context, Characteristics

- Immutable
  - Race, Ethnicity
- Mutable
  - Behavior, activity, interest, belief
  - Modify one's characteristics to align with neighbors

Network Evolution –Membership Closure

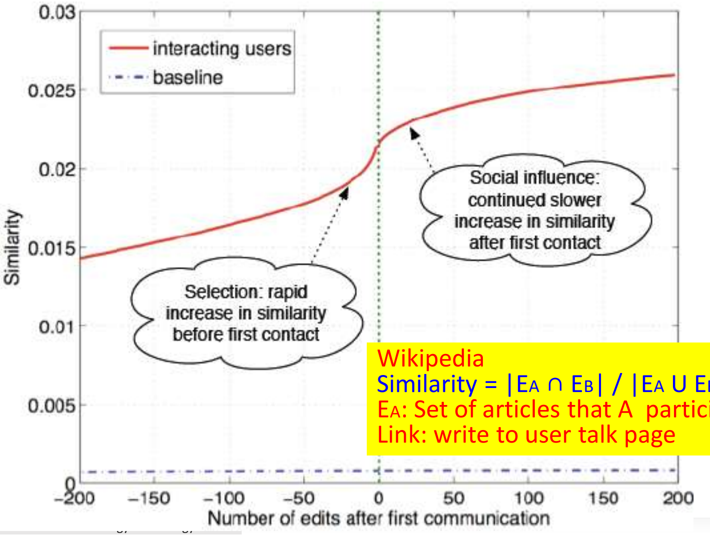
AI

Leskovec
Wikipedia
Friendship: User talk page
Foci: Edit the same article



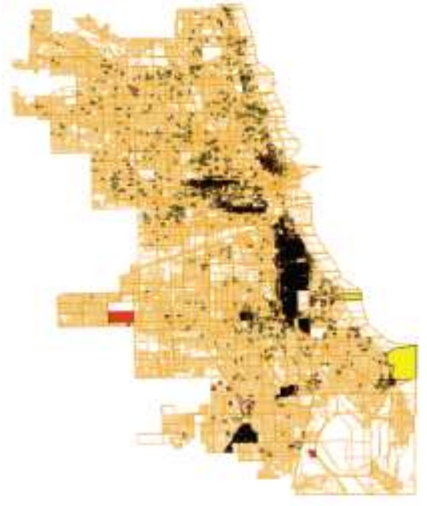
Selection OR Social Influence

AI



Wikipedia
Similarity = |EA ∩ EB| / |EA U EB|
EA: Set of articles that A participates in editing
Link: write to user talk page

Spatial Segregation


(a) Chicago, 1940


(b) Chicago, 1960

Spatial Segregation

- **Segregation** is a result of **homophily**
- Segregation
  - Global pattern
- Homophily
  - Local decision (individual desire)
- Shelling explained the relation using a simple model

SN SCONE lab.