

Real World and Synthetic Graphs

6.886

Andrew Xia

Feb 14th, 2018

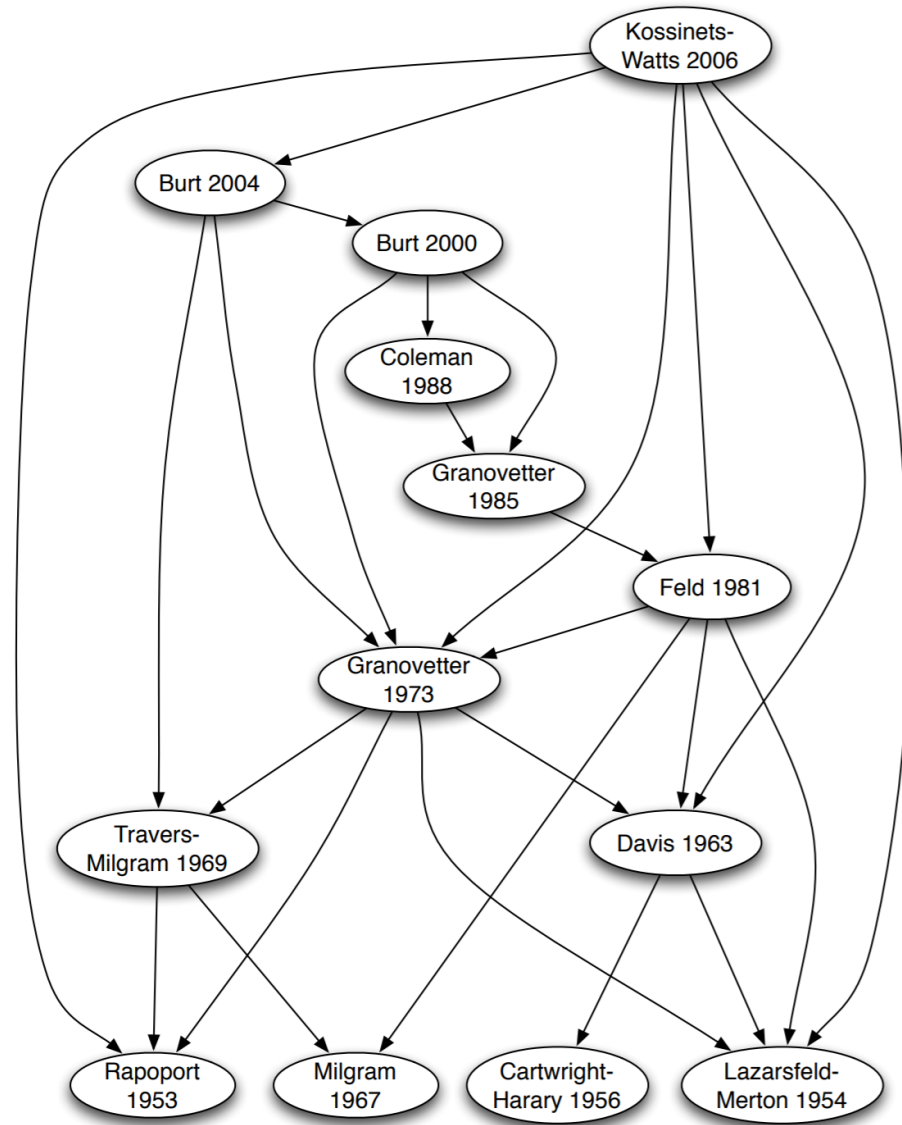
Ch 13: Structure of the Web

World Wide Web as a Network Metaphor

- Web pages
- Browser
- Hypertext: edges in a directed graph

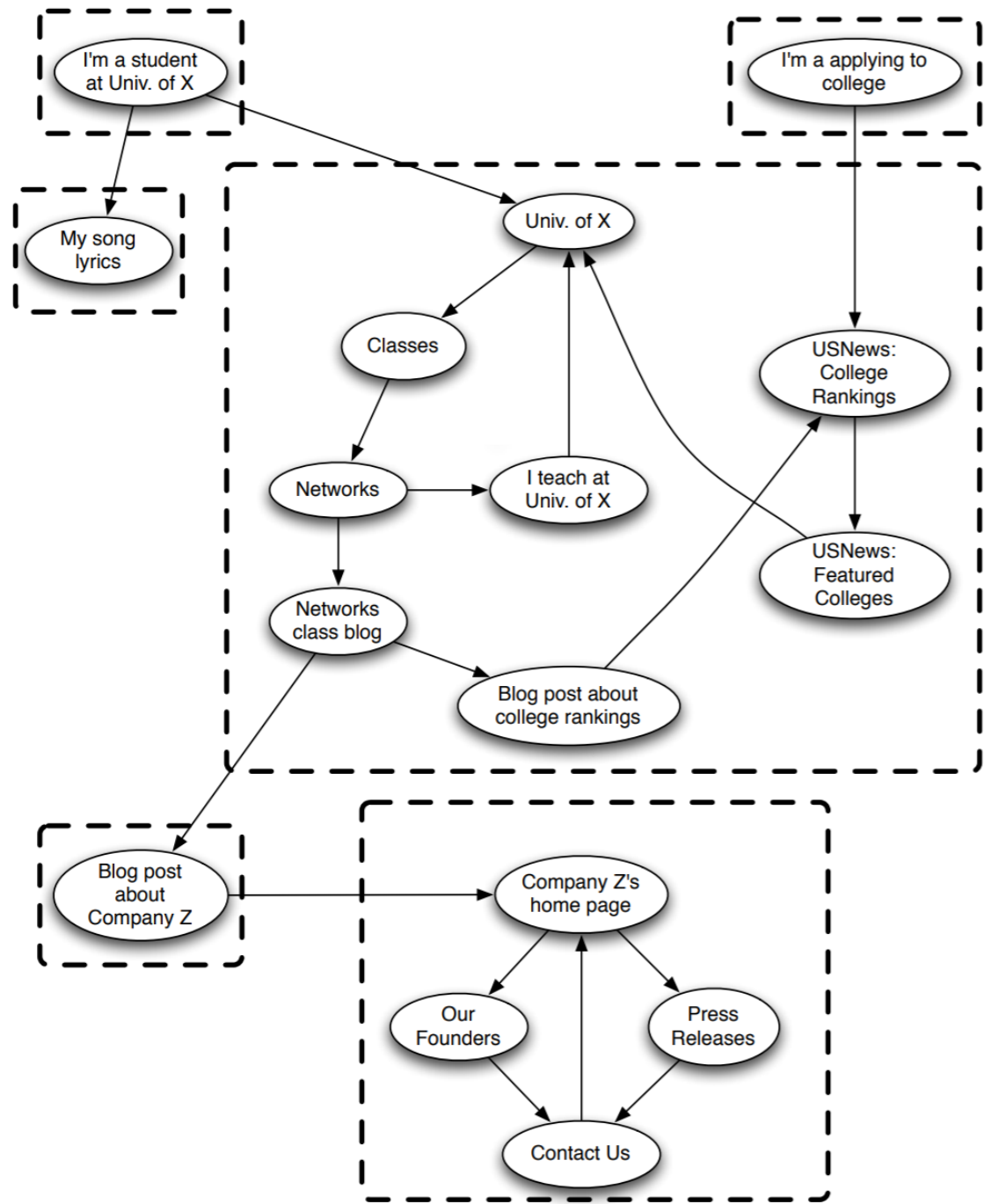
Citation Network

- Arrow of time

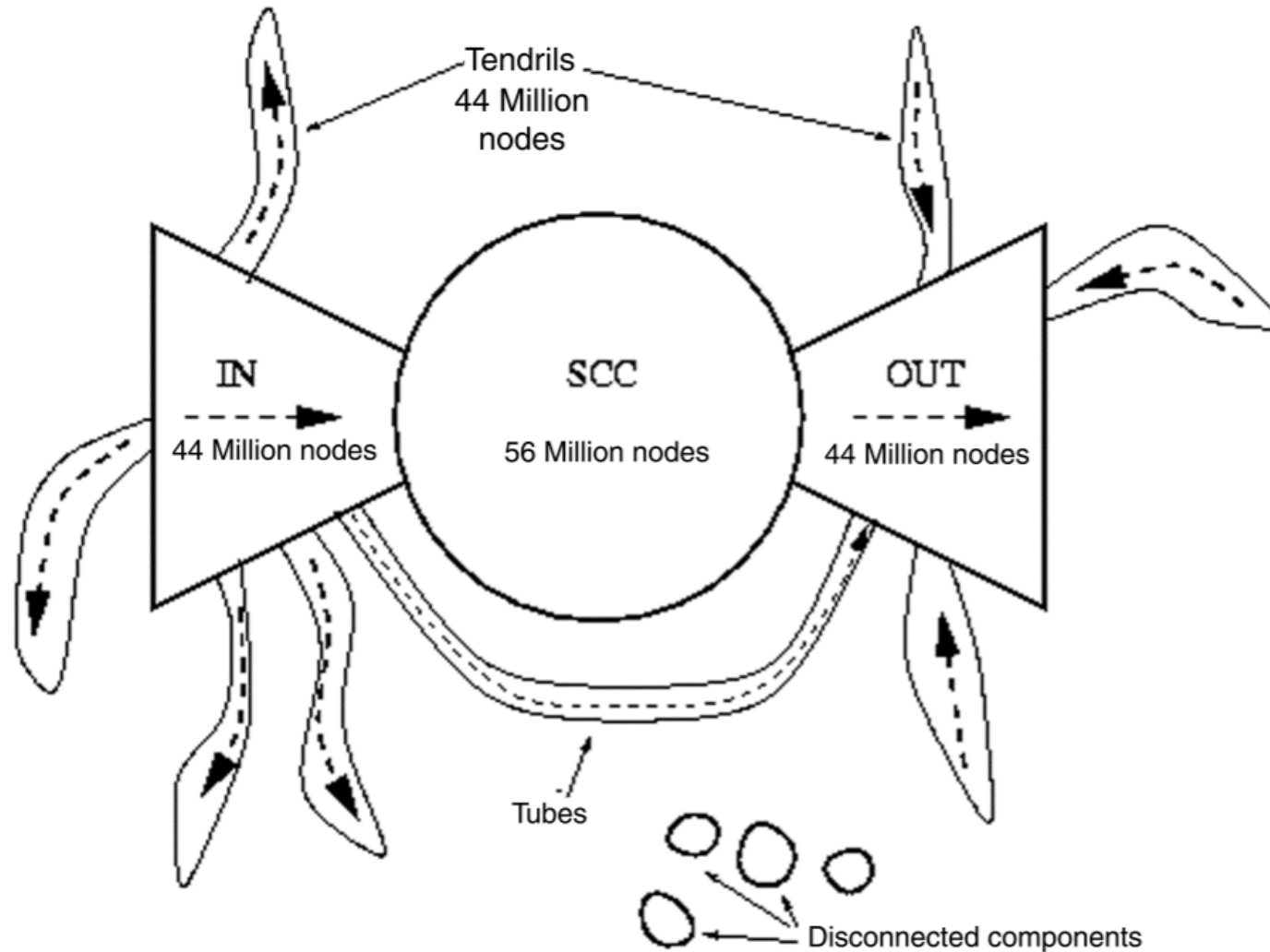


Web as a Directed Graph

- Path: two nodes A, B are linked if there exists a set of edges that lead A to B
- Connected: all pairs are linked via paths
 - Strongly connected
- Reachability
- Connected components
 - Strongly connected components (DAG)



Bow-Tie Structure of Graph



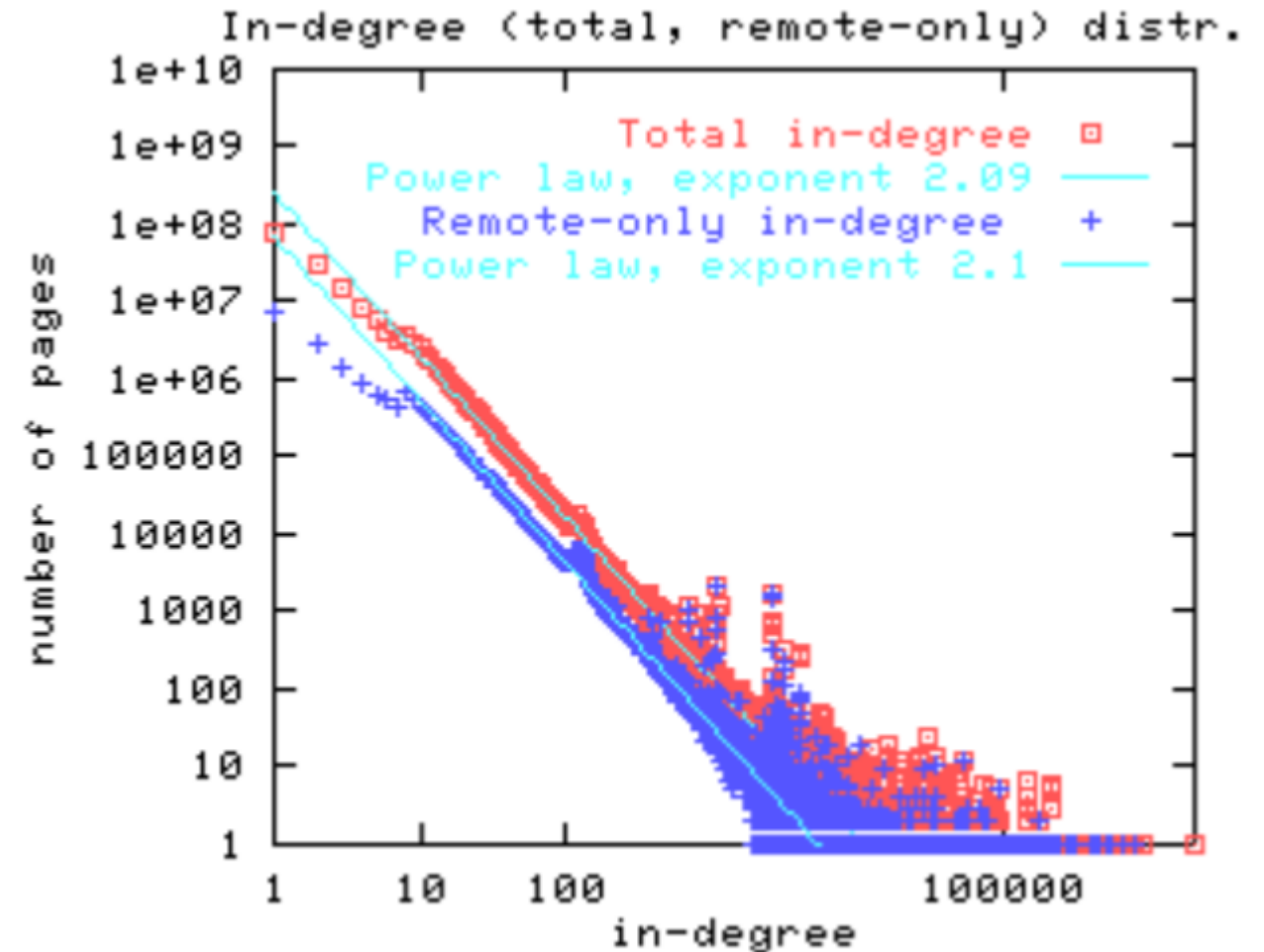
Web 2.0

- Software that gets better the more people use it
- The wisdom of crowds

Ch 18: Power Laws, Rich-Get-Richer Phenomena

In Links

- In link: full set of links pointing to a given webpage
- *What fraction of web pages have k in-links?*
 - Proportional to $1/k^2$
 - $F(k) = a/k^c$
 - $\text{Log } F(k) = \text{log}(a) - c \text{ log}(k)$



Rich Get Richer Model

- Assume that pages are created in order $1, \dots, N$
- When page j created...
 - With probability p , point to page $i < j$
 - With probability $1-p$, choose page i , choose a page that page i points to
 - Repeat for multiple links
- Preferential Attachment
- Shortcomings
 - Only consult one prior page
 - Imitation, but not rational decision-making

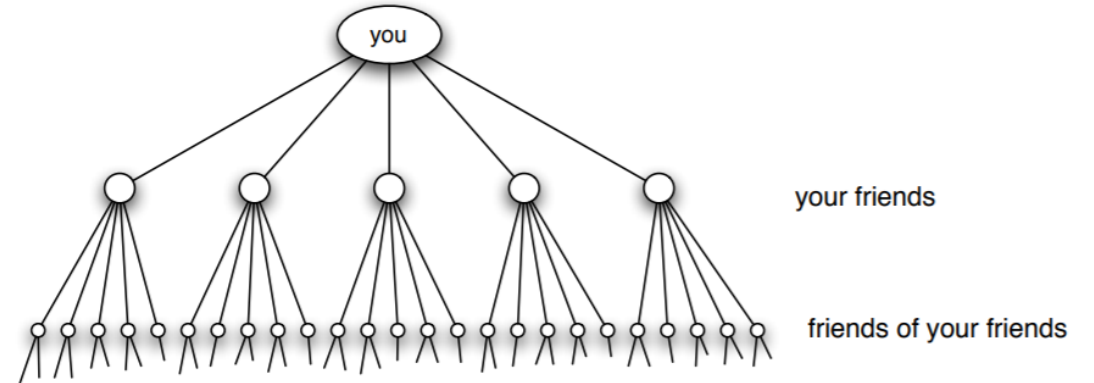
The Long Tail

- *Are most sales generated by small, popular set of items, or by larger population of items individually less popular?*
- *What number of items have popularity k ?*
- Search tools, recommendation systems

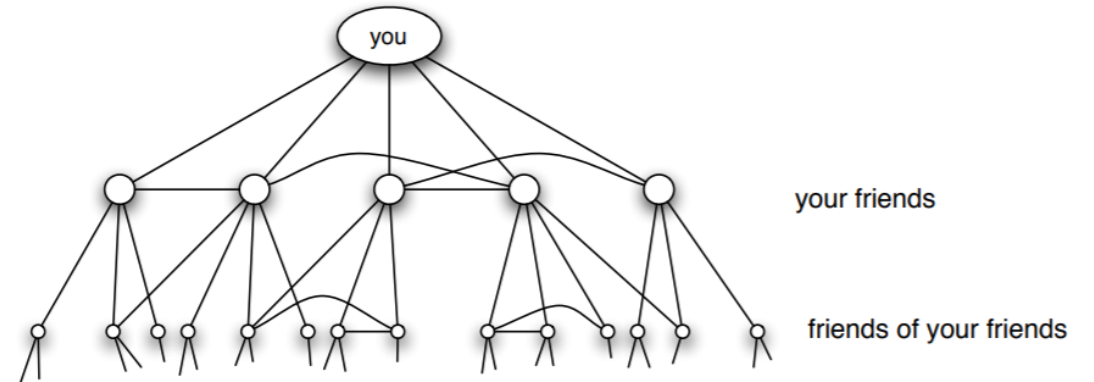
Ch 20: Small World Phenomena

Stanley Milgram: Six Degrees of Separation

- Short Paths are in Abundance
- People are effective at finding shortest path even with only local information
- Triadic closure in network



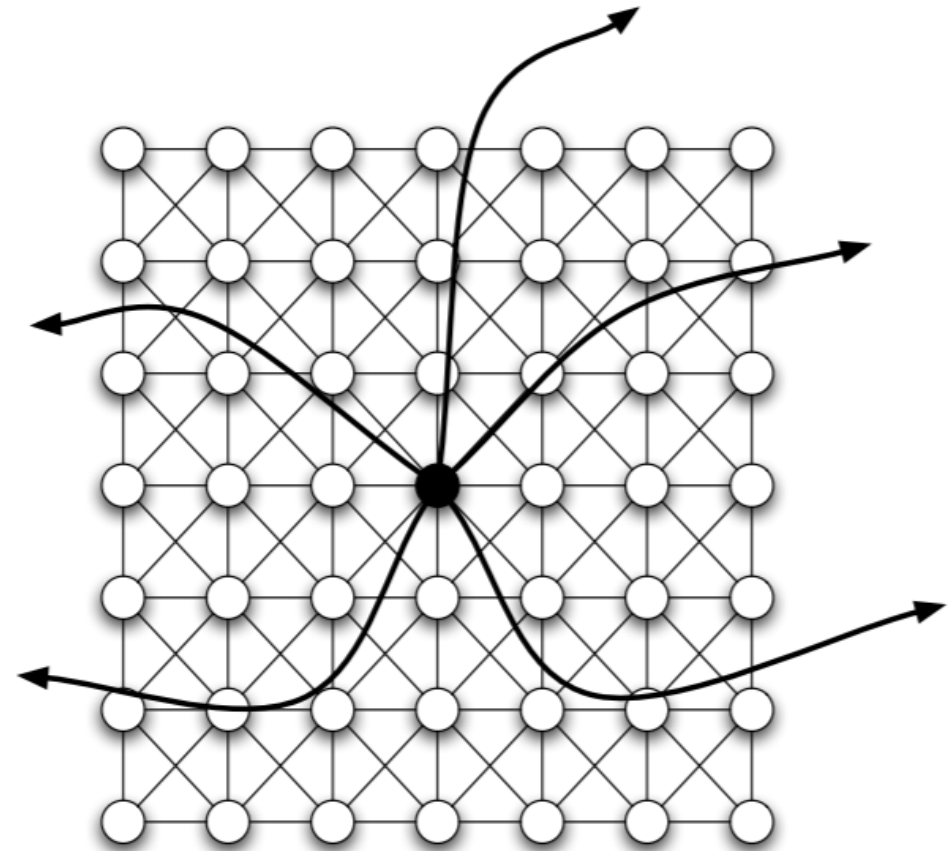
(a) *Pure exponential growth produces a small world*



(b) *Triadic closure reduces the growth rate*

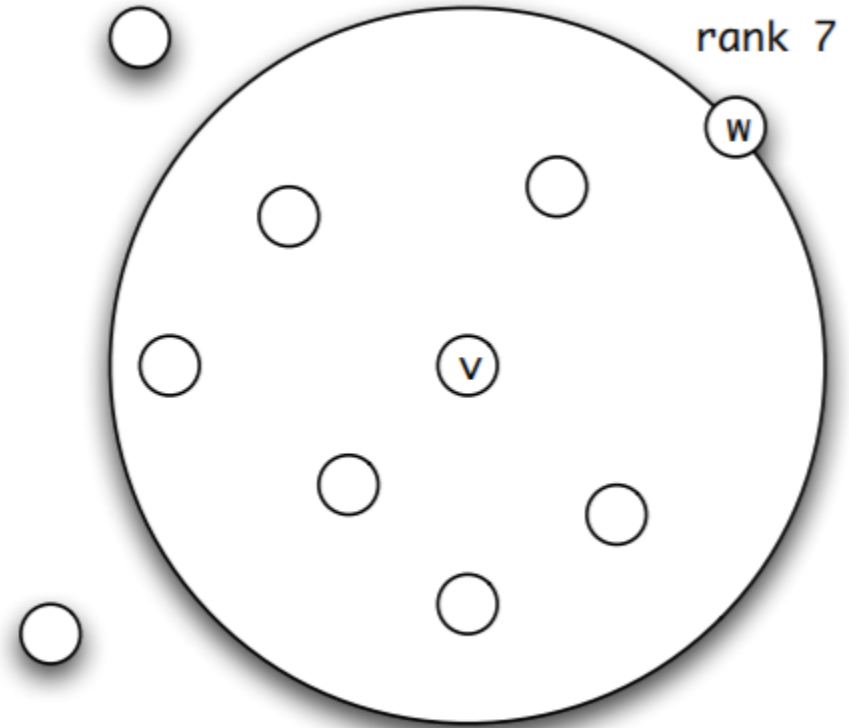
Watts-Strogatz Model

- Social Network is highly clustered
- **Homophily:** connected to all other nodes up to r grid steps away
- **Weak ties:** link to k other nodes selected from the grid
- **Grid step:** manhattan distance
- **Delivery time:** expected number of steps to reach target

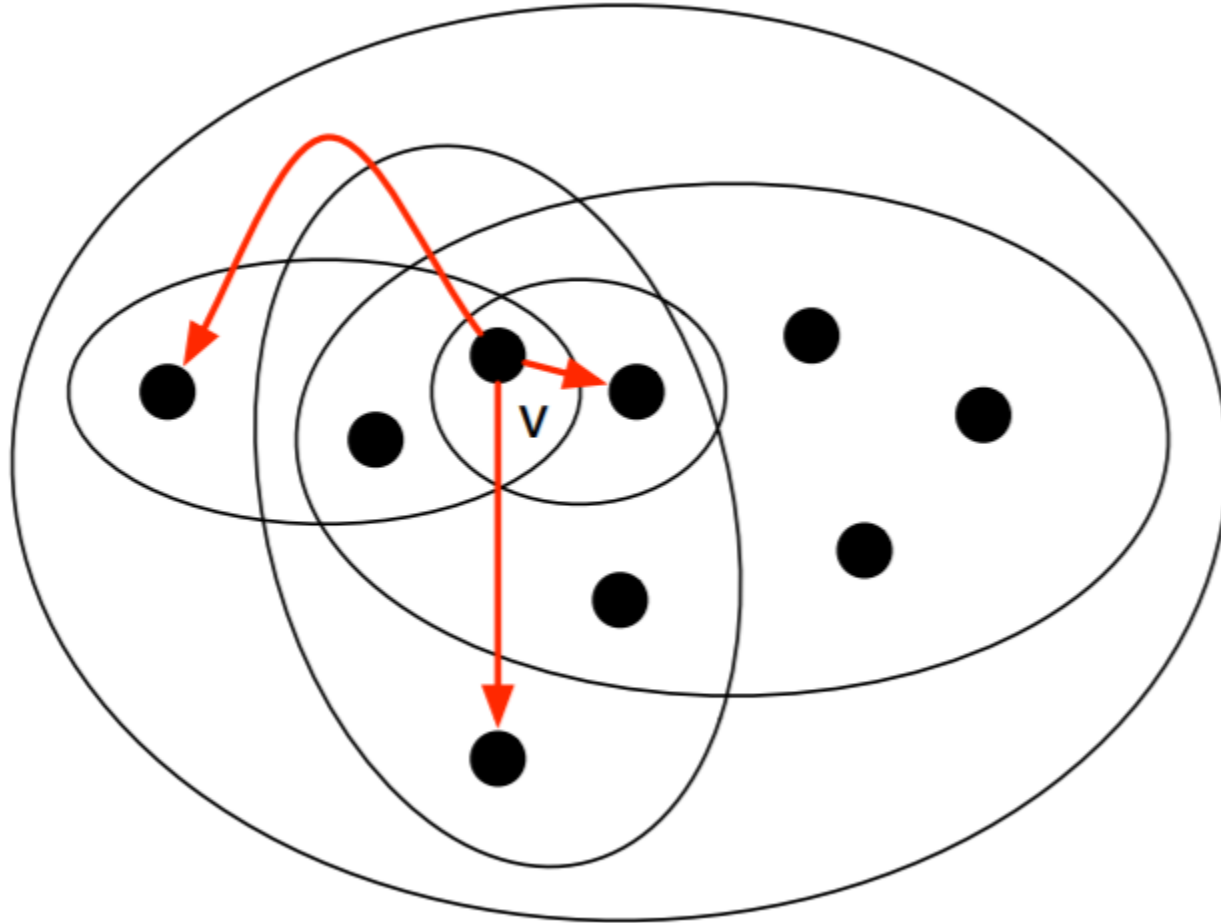


Clustering Exponent

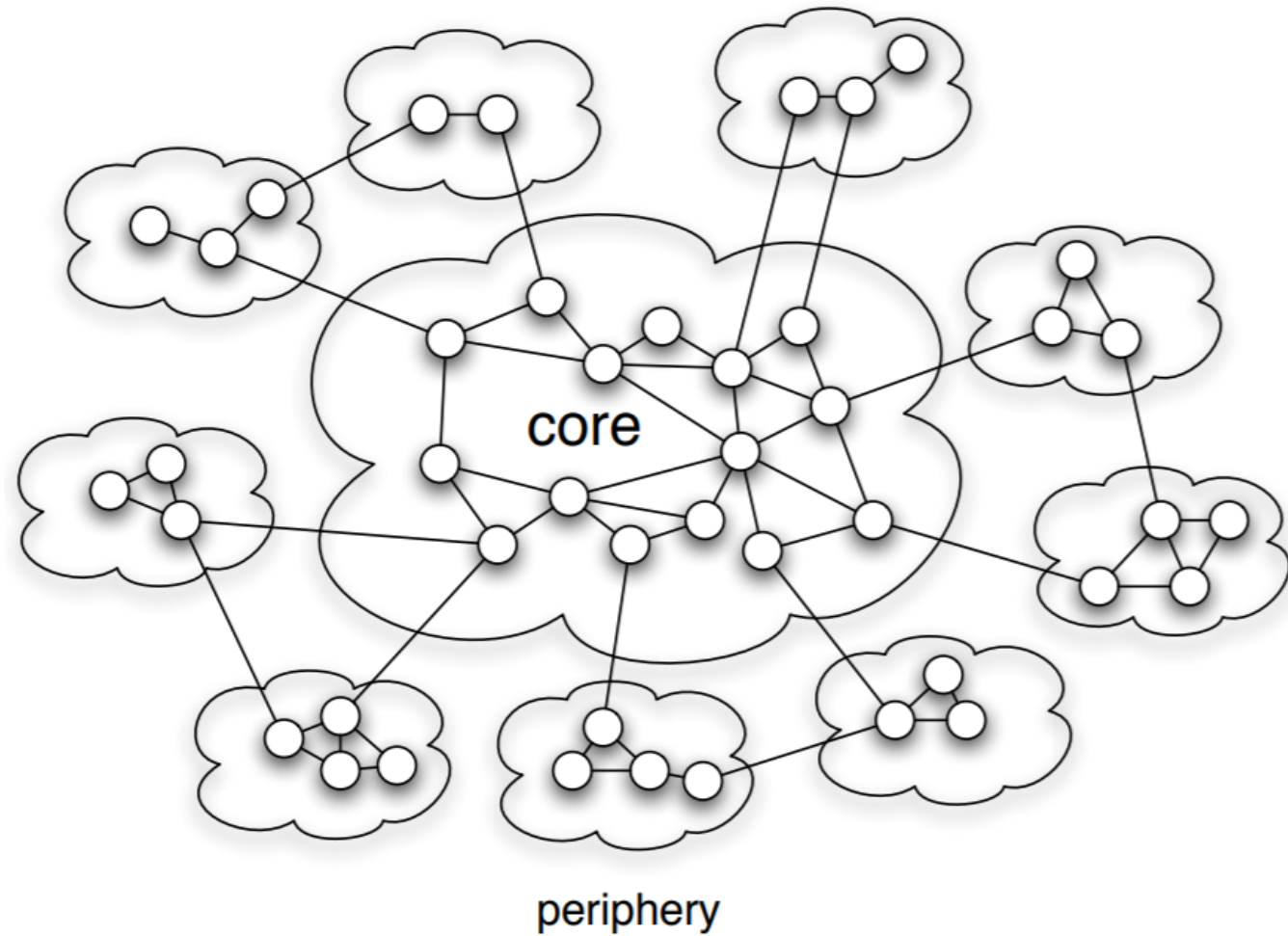
- Generate k random edges such that probability decays with distance
- Clustering exponent q
- $\text{Pr} = d(v,w)^{-q}$
- Most efficient when $q = 2$
 - Intuition: total number of nodes within distance d is d^2
- *Rank-based friendship*
 - Instead of geographical distance, rank by order of distance



Social Foci



Core-Periphery Structure



References

- D. Easley, J. Kleinberg, Networks, Crowds, and Markets: Reasoning About a Highly Connected World, Cambridge University Press, Cambridge, UK, 2010

Thanks!