Social Networks

Jean Mark Gawron

Linguistics 572
San Diego State University

November 10, 2016

Outline

Why social networks?

A long-standing area of study

Seminal papers	Milgram (1967), Granovetter (1973), Zachary (1977), Freeman (1979), . . .
Texbooks/Overviews	Wasserman (1994), Borgatti et al. (2013)

1. Seminal work in the 90s (discussed below)
2. Explosion of world wide web
3. Fresh interest because of social media
4. Success in understanding networks of other kinds provides insight into social networks Newman (2010)

Why networks?

Complex systems

Large entities with many interacting parts that can function cooperatively (note the word **can**) to get things done that none of the parts could get done alone.

1. Network: A view of a system with parts that enter into recurring relations
2. The pattern of connections may be complex, but **the way** in which system elements are connected is uniform

Examples

System	Element	Connection
CNS	neuron	neural connection
Internet	websites	hyperlink
Power grid	power stations	electrical transmission lines
Company	employees	management relations
Food web	species	predator/prey relation
metabolic network	organic compound	metabolyte/product relation

Emergent properties

Why study a complex system as a whole? Why not break it into parts?

Predictability

In a complex system, important properties of the whole may be difficult to predict from the properties of the parts, or even from significant subsets of the parts. They arise from complex interactions of the parts (**Especially true** of biological and social systems).

Modules

Sometimes the correct analysis of the system into functionally meaningful parts (**modules**) requires careful analysis of global properties (e.g., community finding algorithms in social networks).

Food web example

Killing seals

In the late 1990s, local Canadian fishing communities began killing seals in an attempt to bring back the cod population.

David Lavigne's Predator/prey interactions

Lavigne (1996)

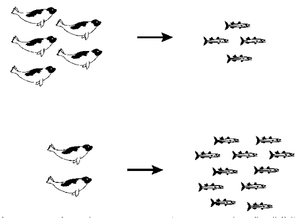


Figure 2 A simple, two-component model of marine ecosystems, where marine mammals eat commercially important fish (upper). A reduction in the number of marine mammals can only result in more fish for fishermen (lower).

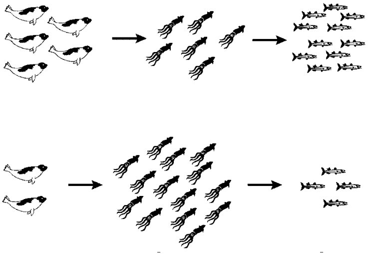
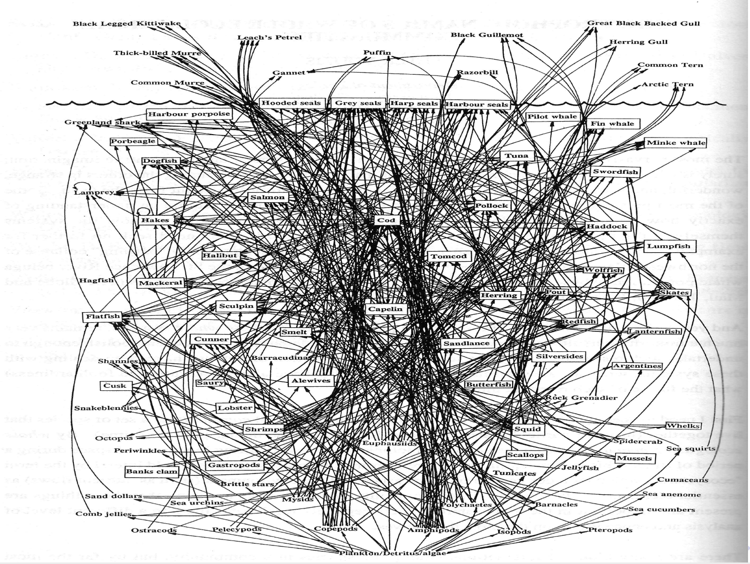


Figure 3 A simple, three-component model of marine ecosystems, where marine mammals eat the predators (or competitors) of a commercially important fish (upper). A reduction in the number of marine mammals results in fewer fish for fishermen (lower).

Lavigne (1996) North Atlantic food web



Yodzis (1994) study survey

Remove predator of species A

Effects are unpredictable. In some cases, Species A populations went up; in some cases, they went down.

Complications (reprised)

a. Predator of Species A may also be a predator of another predator of Species A.
b. Predator of Species A may also be predator some competitor of of Species A

Social links

We seek **social ties** that help explain social dynamics or divisions (e.g., efficiencies in resource flow or patterns of conflict)

Tie type	Sample elicitation	Comments
Role-based relations (e.g., friend; kin; teachers; sexual partner)	Please name all the people in Hartford that you consider a friend	Respondents may have different definitions of “friends”
Interactions (e.g., communicating with; going to movies with)	Who are the people you have talked to about health-related matters at the gym over the last two weeks?.	Respondents more accurate answering questions about who they usually talk to than who they talked to during a given period
Affective ties (e.g., like; dislike)	Who are the people in your office that you feel particularly close to?	Respondent is the sole authority on their feelings
Exchanges & Flows (e.g., help around the house; borrow from)	Who are the people who helped you by giving financial aid after the storm?	Respondents typically feel these are easy questions to answer
Cognitive ties (e.g., is an acquaintance of, know what they eat for breakfast)	You've been in here a month now. Who have you met?	Responses can be voluminous if not well circumscribed

Adapted from **Analyzing Social Networks** Borgatti et al. (2013)

Practical goals of network analysis

1. Find/eliminate/exploit structures that promote efficient transmission of effects through the network
   1. Market analysis
   2. Epidemiology
   3. Ecology (keystone species)
2. Find/eliminate/exploit weaknesses in the network
   1. Network failure analysis (internet, power grid)
   2. Analysis of the effect of infrastructure failure on economic productivity

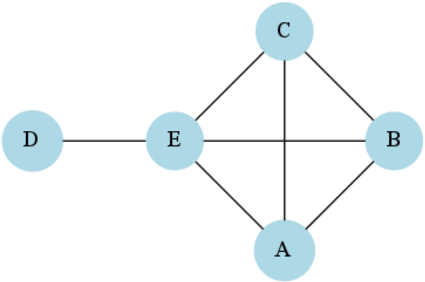
Keystone species

As was described by Dr. Robert Paine in his classic 1966 paper, some sea stars (e.g., **Pisaster ochraceus***) may prey on sea urchins, mussels, and other shellfish that have no other natural predators. If the sea star is removed from the ecosystem, the mussel population explodes uncontrollably, driving out most other species, while the urchin population annihilates coral reefs.*

Wikipedia, "Keystone species"

The mathematical object

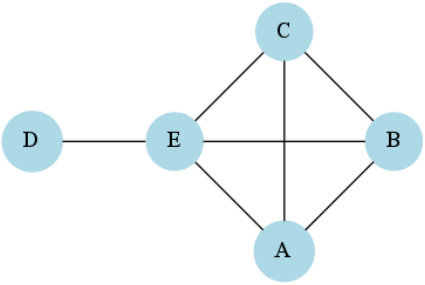
Points	**Lines**	
vertices	edges, arcs	Math
nodes	links, edges	Computer Science
sites	bonds	Physics
actors	ties, relations	Sociology



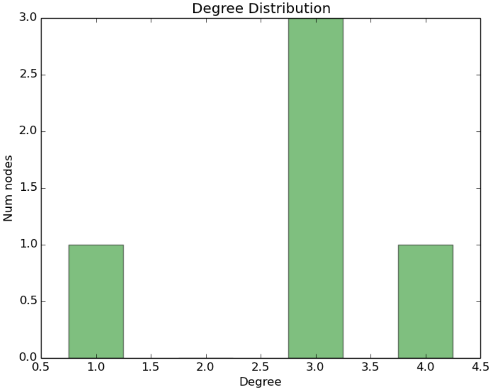
Shortest path

4 paths connecting A and D

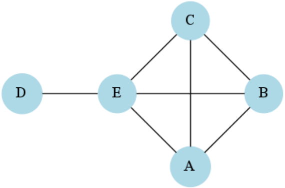
1. D E A
2. D E B A
3. D E C A
4. D E C B A



Degree distribution



Node	Degree
A	3
B	3
C	3
D	1
E	4



Clustering coefficient

For a node n, a measure of how likely two nodes are to be neighbors given they are neighbors of n. In a friendship network: how likely two people are to know each other, given they are n's friends. For a graph $\mathcal{G}$, a measure of how tightly knit the communities are.

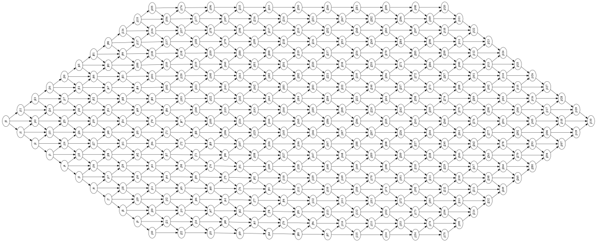
Definition for a node n

The ratio of the number of links connecting ns neighbors to each other to the maximum possible number of such links.

Definition for a Graph

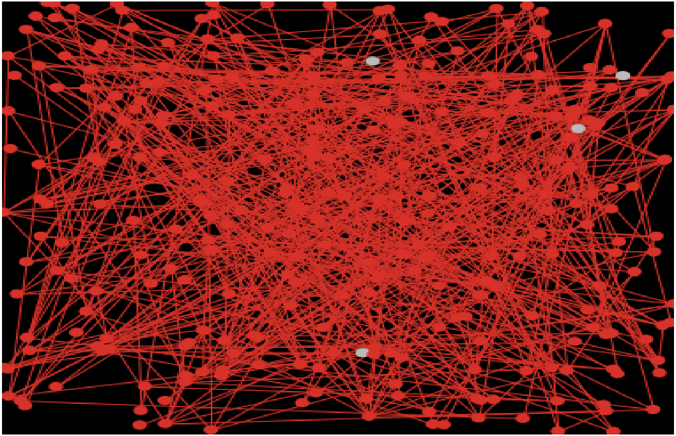
The average clustering coefficient of nodes in $\mathcal{G}$.

Why random graphs?



1. General approach to complexity: Characterize the properties that create order.
2. Real world networks fall somewhere between random and the crystalline order of the above hexagon network.
3. Erdos and Renyi (1959). A series of papers 1959-1968.
4. How do social networks differ from randomness and crystalline order?

A random graph



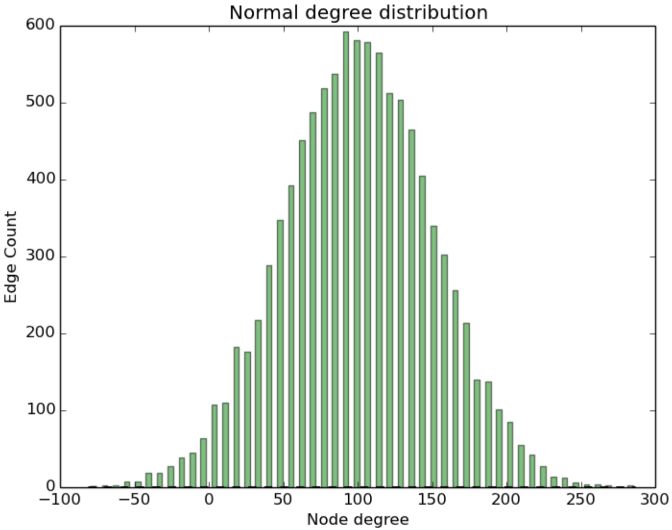
Key properties of random graphs

Property	Value	Shared with social networks?
Average Path Length	Low	Yes
Clustering Coefficient	Low	No
Giant component?	True	Yes
Degree distribution	Normal	No

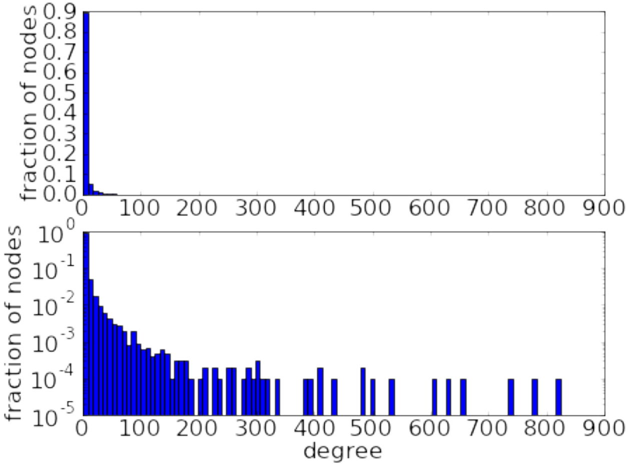
Note: The average path length is the average length of the shortest path between nodes in the graph

Critical point (link with percolation theory): Giant component netlogo demo

The degree distribution of a random graph

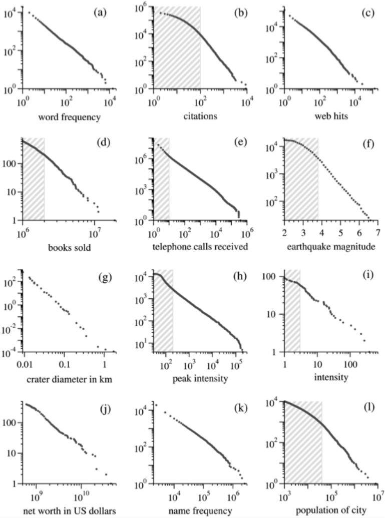


A power law degree distribution



The average value is not the peak of the distribution.

Power law distributions (Newman 2005)



a) Words in *Moby Dick* (Zipf's Law)
b) Citations for scientific papers
c) Web site hits
d) Bestseller sales
e) Calls per day (AT&T)
f) Earthquake magnitudes
g) Moon crater diameters
h) Solar flare intensity
i) Battle deaths per war
j) Aggregate net worth in dollars of the rich (Pareto's Law)
k) Family name frequency
l) U.S. city populations

Milgram (1967)

The experiment

In 1967, Stanley Milgram of Harvard University conducted the following experiment. He asked 300 people chosen at random to send a letter through friends to a stockbroker near Boston we will call the *target*. If they had never met the target, they were asked to pass the letter, along with the instructions, on to a friend whom they felt might be "closer to" the target.

The result

A whopping 64 of the original 300 letters arrived safely. The average length of the recipient chain for those 64 letters was 5.5 persons.

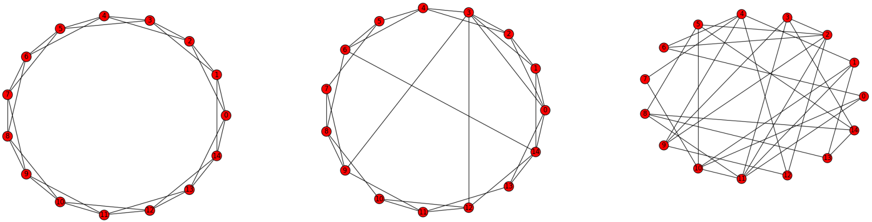
Six degrees of separation

Playwright John Guare coined the term **six degrees of separation** for this new conception of how linked together we all are in the modern world.

Watts and Strogatz

W&S definition of a small world network

A network with a short average path length and a high clustering coefficient is called **a small world.** Watts and Strogatz (1998)



$p = 0$ Increasing randomness $p = 1.0$

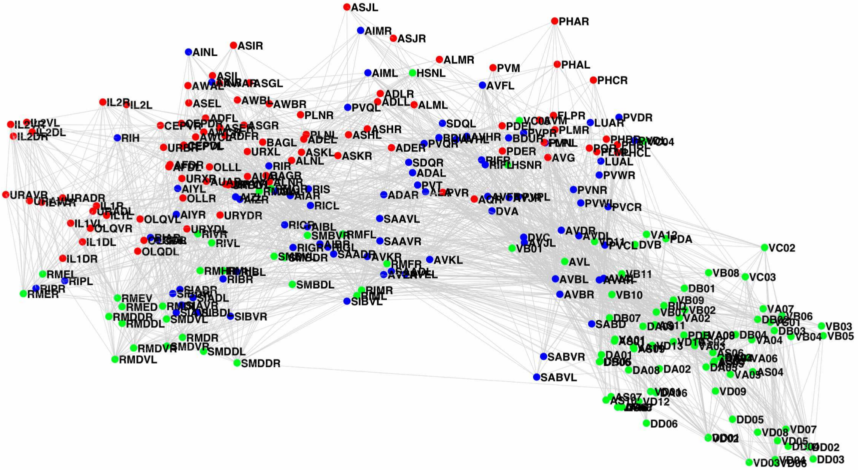
Watts & Strogratz Netlogo demo

Small world networks

C. elegans

1. Actor collaboration network
2. U.S. power grid
3. Neural network of *C. elegans*
4. Condensed matter physicist collaborations
5. The internet (either as physical connections or as hyperlinks)

The neural network of *C. elegans*

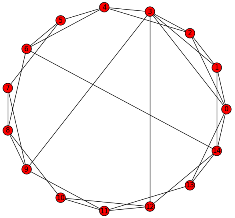


Watts and Strogatz (1998)

Properties of movie actors network

Movie actors		Random graph ($p = .1$)
Size	225,226	225,000
Average degree	61	22,500
Average path length	3.65	1.7
Clustering coefficient	0.79	.1

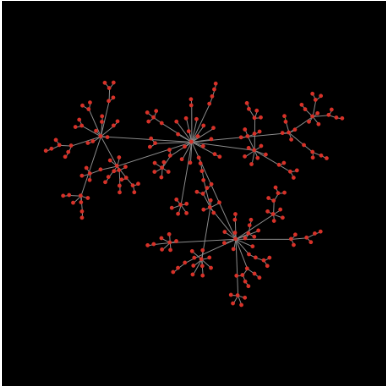
The strength of weak ties (Granovetter 1973)



Did you get your job through a friend?

1) Getting a job
2) Starting a new business
3) Finding a mate
4) Learning new information
5) Spreading a social innovation, such as a new kind of technology, or a political movement.

Preferential attachment: Rich get richer (Barabási and Albert 1999)



NetLogo Preferential Attachment

1. Start with a network with two connected nodes
2. Add nodes
3. Attach a new node *n* to an existing node *e* according to the following rule: The more "popular" *e* is, the more likely it is that *n* connects to *e*.

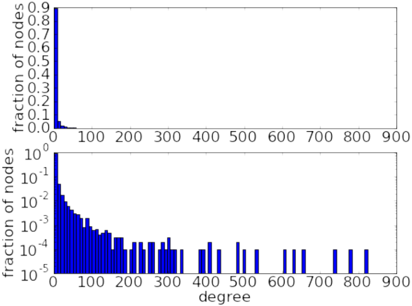
$$p_i = \frac{d_i}{\sum_j d_j}$$

Power law degree distribution

$$P(k) \propto k^{-\gamma}$$

P(k) is the proportion of nodes having some given degree k. P(k) is proportional to k raised to some negative exponent γ (typically $2 < \gamma < 3$).

So P(k) falls rapidly as k grows.



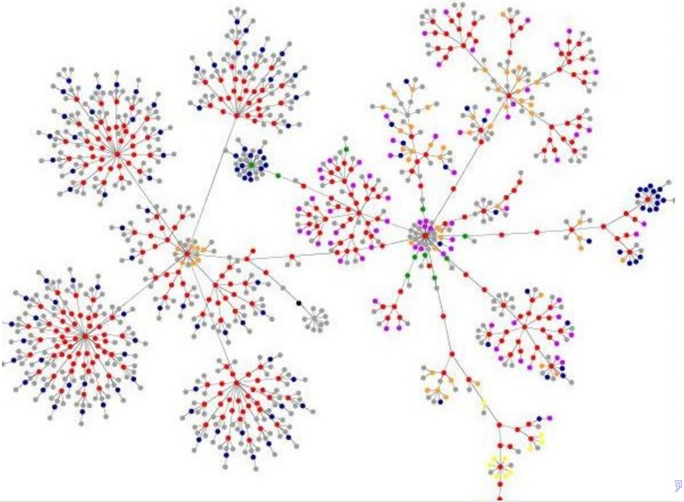
Scale-free graphs: Small worlds?

Networks generated by the Barabási-Albert procedure are small worlds, but not all small worlds are scale-free.

Power law	Non power law
Actor collaboration network	*C. elegans* neural network
U. S. Power grid	
Physicist collaborations	
Internet	

In a scale-free distrubution, we expect most nodes to have few links, but we also look for a few nodes with many links. For example, in the actor's collaboration network, 41% of the actors have fewer than 10 links, John Carradine has 4000 links. Robert Mitchum has 2905. Scuh extreme outliers are extremely improbable in a normal distribution.

The key feature of power law networks: Hubs



References I

Barabási, Albert-László, and Réka Albert. 1999.
Emergence of scaling in random networks.
science 286(5439):509–512.

Borgatti, Stephen P, Martin G Everett, and Jeffrey C Johnson. 2013.
Analyzing social networks.
SAGE Publications Limited.

Freeman, Linton C. 1979.
Centrality in social networks conceptual clarification.
Social networks 1(3):215–239.

Granovetter, Mark S. 1973.
The strength of weak ties.
American journal of sociology 1360–1380.

References II

Lavigne, DM. 1996.
Ecological interactions between marine mammals, commercial fisheries, and their prey: unravelling the tangled web.
High-latitude seabirds 4:59–71.

Milgram, S. 1967.
The small world problem.
Psychology today 1:60–67.

Newman, Mark. 2010.
Networks: an introduction.
Oxford University Press.

References III

Wasserman, Stanley. 1994.
Social network analysis: Methods and applications.
Vol. 8.
Cambridge university press.

Watts, D., and S. Strogatz. 1998.
The small world problem.
Collective dynamics of 'small-world' networks 393:440–442.

Yodzis, Peter. 1994.
Predator-prey theory and management of multispecies fisheries.
Ecological Applications 51–58.

Zachary, W. W. 1977.
An information flow model for conflict and fission in small groups.
Journal of Anthropological Research 33:452–473.