2008-09

SI 508 - Networks: Theory and Application, Fall 2008

Adamic, Lada

<http://hdl.handle.net/2027.42/64962>
http://hdl.handle.net/2027.42/64962
Why networks are interesting to study

Instructor: Lada Adamic
Outline

- the role of networks in life, nature, and research
- why model networks: structure & dynamics
  - models (structure):
    - Erdos-Renyi random graph
    - Watts-Strogatz small world model
    - Barabasi-Albert scale-free networks
  - implications (dynamics):
    - diffusion of disease and information
    - search by navigating the network
    - resilience
    - IR applications
examples: early social network analysis

- 1933 Moreno displays first sociogram at meeting of the Medical Society of the state of New York
  - article in NYT
  - interests: effect of networks on e.g. disease propagation

- Preceded by studies of (pre)school children in the 1920’s

examples: early social network analysis

- School kids – favorite (and captive) subjects of study
- These days much more difficult because need parental consent to gather social network data

Source: An Attraction Network in a Fourth Grade Class (Moreno, ‘Who shall survive?’, 1934).
**What are networks?**

Networks are collections of points joined by lines.

“Network” ≡ “Graph”

<table>
<thead>
<tr>
<th>points</th>
<th>lines</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>vertices</td>
<td>edges, arcs</td>
<td>math</td>
</tr>
<tr>
<td>nodes</td>
<td>links</td>
<td>computer science</td>
</tr>
<tr>
<td>sites</td>
<td>bonds</td>
<td>physics</td>
</tr>
<tr>
<td>actors</td>
<td>ties, relations</td>
<td>sociology</td>
</tr>
</tbody>
</table>
examples: Political/Financial Networks

- Mark Lombardi: tracked and mapped global financial fiascos in the 1980s and 1990s (committed suicide 2000)
- searched public sources such as news articles
- drew networks by hand (some drawings as wide as 10ft)
“I happened to be in the Drawing Center when the Lombardi show was being installed and several consultants to the Department of Homeland Security came in to take a look. They said they found the work revelatory, not because the financial and political connections he mapped were new to them, but because Lombardi showed them an elegant way to array disparate information and make sense of things, which they thought might be useful to their security efforts. I didn't know whether to find that response comforting or alarming, but I saw exactly what they meant.”

Michael Kimmelman
Webs Connecting the Power Brokers, the Money and the World
NY Times November 14, 2003
“Six degrees of Mohammed Atta”

Uncloaking Terrorist Networks, by Valdis Krebs
examples: boards of directors

Source: http://theyrule.net
examples: online social networks

- Friendster

examples: Networks of personal homepages

homophily: what attributes are predictive of friendship?
group cohesion

examples: internet

Source: Bill Cheswick http://www.cheswick.com/ches/map/gallery/index.html
examples: airline networks

Source: Northwest Airlines WorldTraveler Magazine
examples: railway networks

Source: TRTA, March 2003 - Tokyo rail map
other examples, e.g. natural language processing

- Wordnet

Source: http://wordnet.princeton.edu/man/wnlicense.7WN
examples: gene regulatory networks

- gene regulatory networks
  - humans have only 30,000 genes, 98% shared with chimps
  - the complexity is in the interaction of genes
  - can we predict what result of the inhibition of one gene will be?

Source: http://www.zaik.uni-koeln.de/bioinformatik/regulatorynets.html.en
examples: metabolic networks

- Citric acid cycle
- Metabolites participate in chemical reactions

Source: undetermined
modeling networks: random networks

- Nodes connected at random
- Number of edges incident on each node is Poisson distributed

Poisson distribution
Erdos-Renyi random graphs

What happens to the size of the giant component as the density of the network increases?

http://ccl.northwestern.edu/netlogo/models/run.cgi?GiantComponent.884.534
modeling networks: small worlds

- Small worlds
  - a friend of a friend is also frequently a friend
  - but only six hops separate any two people in the world

Arnold S. – thomashawk, Flickr; http://creativecommons.org/licenses/by-nc/2.0/deed.en
Small world models

Duncan Watts and Steven Strogatz

- a few random links in an otherwise structured graph make the network a small world: the average shortest path is short

Watts Strogatz Small World Model

- As you rewire more and more of the links and random, what happens to the clustering coefficient and average shortest path relative to their values for the regular lattice?

http://projects.si.umich.edu/netlearn/NetLogo4/SmallWorldWS.html
SIS models and small worlds

- SIS model: nodes return to “susceptible” state after being infected
- What is the role of random shortcuts in diffusion?

http://projects.si.umich.edu/netlearn/NetLogo4/SmallWorldWS.html
modeling networks: power law networks

- Many real world networks contain hubs: highly connected nodes
- Usually the distribution of edges is extremely skewed

![Graph showing the distribution of edges in power law networks]

- Fat tail: a few nodes with a very large number of edges
- No “typical” number of edges

- Many nodes with few edges
But is it really a power-law?

- A power-law will appear as a straight line on a log-log plot:

- A deviation from a straight line could indicate a different distribution:
  - exponential
  - lognormal
network growth & resulting structure

- random attachment: new node picks any existing node to attach to
- preferential attachment: new node picks from existing nodes according to their degrees

http://projects.si.umich.edu/netlearn/NetLogo4/RAndPrefAttachment.html
What implications does this have?

- Robustness
- Search
- Spread of disease
- Opinion formation
- Spread of computer viruses
- Gossip
How do we search?

Who could introduce me to Richard Gere?

Mary

Bob

Jane

Richard Gere – spaceodissey, Flickr; http://creativecommons.org/licenses/by/2.0/deed.en

Friends collage – luc, Flickr; http://creativecommons.org/licenses/by/2.0/deed.en
power-law graph

number of nodes found

<table>
<thead>
<tr>
<th>Number of Nodes Found</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>94</td>
<td>White</td>
</tr>
<tr>
<td>67</td>
<td>Light Blue</td>
</tr>
<tr>
<td>63</td>
<td>Light Purple</td>
</tr>
<tr>
<td>54</td>
<td>Purple</td>
</tr>
<tr>
<td>42</td>
<td>Pink</td>
</tr>
<tr>
<td>31</td>
<td>Red</td>
</tr>
</tbody>
</table>

The diagram illustrates a power-law graph with varying numbers of nodes found, color-coded to indicate the frequency or intensity of their connections within the network.
Power-law networks are robust to random breakdown
But are especially vulnerable to targeted attack

- Targeting and removing hubs can quickly break up the network
In social networks, it's nice to be a hub
But it depends on what you’re sharing...
The role of hubs in epidemics

- In a power-law network, a virus can persist no matter how low its infectiousness.

- Many real world networks do exhibit power-laws:
  - needle sharing
  - sexual contacts
  - email networks
Spread of computer viruses can be affected by the underlying network.
SI models & network structure

- Will random or preferential attachment lead to faster diffusion?

random growth  preferential growth

http://projects.si.umich.edu/netlearn/NetLogo4/BADiffusion.html
resilience: power grids and cascading failures

- Vast system of electricity generation, transmission & distribution is essentially a single network
- Power flows through all paths from source to sink (flow calculations are important for other networks, even social ones)
- All AC lines within an interconnect must be in sync
- If frequency varies too much (as line approaches capacity), a circuit breaker takes the generator out of the system
- Larger flows are sent to neighboring parts of the grid – triggering a cascading failure

Cascading failures

- **1:58 p.m.** The Eastlake, Ohio, First Energy generating plant shuts down (maintenance problems).
- **3:06 p.m.** A First Energy 345-kV transmission line fails south of Cleveland, Ohio.
- **3:17 p.m.** Voltage dips temporarily on the Ohio portion of the grid. Controllers take no action, but power shifted by the first failure onto another power line causes it to sag into a tree at 3:32 p.m., bringing it offline as well. While Mid West ISO and First Energy controllers try to understand the failures, they fail to inform system controllers in nearby states.
- **3:41 and 3:46 p.m.** Two breakers connecting First Energy’s grid with American Electric Power are tripped.
- **4:05 p.m.** A sustained power surge on some Ohio lines signals more trouble building.
- **4:09:02 p.m.** Voltage sags deeply as Ohio draws 2 GW of power from Michigan.
- **4:10:34 p.m.** Many transmission lines trip out, first in Michigan and then in Ohio, blocking the eastward flow of power. Generators go down, creating a huge power deficit. In seconds, power surges out of the East, tripping East coast generators to protect them.

(dis) information cascades

- Rumor spreading
- Urban legends
- Word of mouth (movies, products)

- Web is self-correcting:
  - Satellite image hoax is first passed around, then exposed, hoax fact is blogged about, then written up on urbanlegends.about.com

Source: undetermined
Actual satellite images of the effect of the blackout

20 hours prior to blackout
7 hours after blackout

Source: NOAA, U.S. Government
IR applications: online info retrieval

- It's in the links:
  - Links to URLs can be interpreted as endorsements or recommendations.
  - The more links a URL receives, the more likely it is to be a good/entertaining/provocative/authoritative/interesting information source.
  - But not all link sources are created equal:
    - A link from a respected information source.
    - A link from a page created by a spammer.

An important page, e.g., slashdot.

If a web page is slashdotted, it gains attention.

Many webpages scattered across the web.
Ranking pages by tracking a drunk

- A random walker following edges in a network for a very long time will spend a proportion of time at each node which can be used as a measure of importance.

- Various eigenvalue metrics yield variations of importance measures.
Wrap up

- Networks are everywhere and can be used to describe many, many systems.
- By modeling networks we can start to understand their properties and the implications those properties have for processes occurring on the network.