SOCIAL + INFORMATION NETWORK ANALYSIS

CS584: Data Mining Fall 2016 Huzefa Rangwala, Ph.D. http://www.cs.gmu.edu/~hrangwal

Why are networks interesting to study?

Know your thoughts!

 What is meant by social network analysis ? What are the end results of such analysis ? and why do we care ?

Example: Internet



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

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

Source: The New York Times (April 3, 1933, page 17).

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).

Today Social Networks



Facebook social graph 4-degrees of separation [Backstrom-Boldi-Rosa-Ugander-Vigna, 2011]

What are networks?

 Networks are collections of points joined by lines.

	o node edge	
$\mathbf{\hat{v}}$	points	lir
	vertices	ec
	nodes	lin
	sites	bo

"Network" ≡ "Graph"

points	lines	
vertices	edges, arcs	math
nodes	links	computer science
sites	bonds	physics
actors	ties, relations	sociology

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)
- Book: Hobbs, Robert. Mark Lombardi :global networks /Robert Hobbs.. New York : Independent Curators International, c2003..





Understanding through visualization

 "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

examples: terrorist networks



"Six degrees of Mohammed Atta"

Uncloaking Terrorist Networks, by Valdis Krebs

examples: boards of directors



Source: http://theyrule.net

Citation Networks



[Börner et al., 2012]

Co-Authorship Networks



examples: Networks of personal homepages



homophily: what attributes are predictive of friendship? group cohesion

Source: Lada A. Adamic and Eytan Adar, 'Friends and neighbors on the web', *Social Networks*, 25(3):211-230, July 2003.

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/wnlicens.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 provide 's'hat result of the inhibition of one cone 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



Source: Roche Applied Science, http://www.expasy.org/cgi-bin/show_thumbnails.pl

Food Web



Brain Networks





Social Network Impact



flickr.com/photos/stabilo-boss/101793493/ licensed under cc-attribution-noncommercial-sharealike

PREVIEW OF SNA

modeling networks: random networks

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



modeling networks: small work

- 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



regular lattice: my friend's friend is always my friend small world: mostly structured with a few random connections

random graph: all connections random

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 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
- Sually the distribution of edges is extremely skewed many nodes with few edges fat tail: a few nodes with a very large nu of edges fat tail: a few nodes with a very large number number of edges

no "typical" number of 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





Poisson graph



Power-law networks are robust to random breakdown





• Targeting and removing hubs can quickly break up the network

In social networks, it's nice to be a hub





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



SI models & network structure

• Will random or preferential attachment lead to faster diffusion?





random growth

preferential growth

(dis) information cascades

- Urban legends
- Word of mouth (movies, products)
- Web is selfcorrecting:
 - Satellite image hoax is first passed around, then exposed, hoax fact is blogged about, then written up on urbanlegends.about.com



Source: undetermined

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



- 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