Spreading Processes — Network Analysis

we want to USE networks now :)

How does $\underbrace{\text{something}}_{\text{}}$ spread in a network?

rumors, ideas, disease, weather/temperature

All models origin from epidemology ♥

Network effects — we add knowledge on top of our network!

SI-model — S - susceptible to infection I - infected

β - infection rate/probability no healing

Effects:

1. The more connected/clustered, the faster the spread
   a. connected to degree distribution

SIS-model — Same as above, but including healing

μ - recovery rate

now the $\underbrace{\text{max \# of infected}}_{\text{Endemic state}}$ is less than 100%.

Endemic state — $G(n,p)$ = if $\lambda > \frac{1}{n+1}$ then endemic state

$\lambda = \frac{\beta}{\mu}$

Scale free can always be endemic no matter λ

SIR-model — Same as SIS, but once you're I, you cannot become S, but become Recovered or Removed DEAD

R-nodes will disconnect the network

Regular Cells SIR   Cancer cells SI

Triggering Mechanism

Contagion : $S \to I$

single- vs. multiple exposure

4 Contagion Models

1. single exposure
2. Threshold

   β is $p(S \to I)$, and $\underbrace{(1-(1-\beta)^n)}$ after n exposures is $p(S \to I)$
3. Cascade

   β is now min. fraction of neighbors

   you go $S \to I$ if $\frac{I}{T}$ of neighbors $> \beta$
4. Percolation

   no β, instead $k = \#$ ← number of neighbors to be I for you to become I

Reachability	Thresh	Casc	Perco
HUBS	Easy	Hard	Easy
Peripheral	Hard	Easy	Impossible ← usually ☺

Link Prediction

next edge? missing edges?

Classical Link Prediction

1. Observe Links
2. Hypotheses for link creation process
3. Operationalize it

Bias → high degree nodes

Preferential Attachment ← worst method

we 'prefer' high degree nodes

↳ the more links you have, the more likely you get new

↳ potential links are scored by product of nodes' degrees

Common Neighbors ← Triangles tend to close

we prefer other nodes for which we have common neighs

↳ potential links are scored by count of common neighbors

↳ possibly → % of set(neighbors) that are common :)

Adamic-Adar ← common neighbors correction

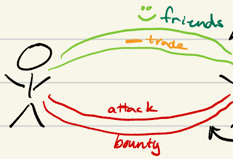
common neighbors are weighted → hubs penalized discounted

since hubs cannot neccessarily introduce all neighbors :(

↳ weight of neighbor is $\frac{1}{\log(d)}$ degree

↳ variations: $\frac{1}{d}$ etc.

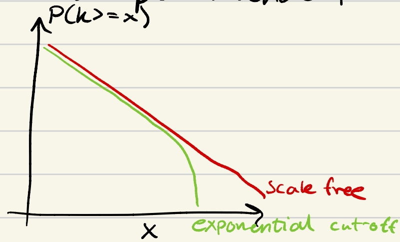
Signed Networks



good / positive links

bad / negative links

we can plot friendship and enemy degree dist



Funny:
Limit on # friends,
no limit on # enemies!

Link predictions

in signed networks

Social Balance Theory

undirected

1. Friend of my Friend → is a friend!
2. Enemy of my Friend → is my enemy!

balance

~ ? enemy of my enemy is my enemy?
%? friend of my friend is my enemy?

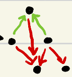
unbalance

This theory is used to determine sign of potential link

Social Status Theory

directed → endorsement

sending connections creates hierarchy ⟹



F.ex ratings/trust networks

Multilayer
Link Prediction

which two nodes? AND how? 'which layer?'

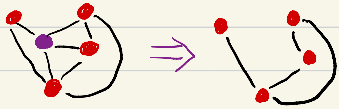
Layer Independence
Multilayer score

The Mesoscale

Dividing our network structure into smaller substructures

Ego Network

Only consider nodes that have incident edges to "chosen" node, and all edges between them.

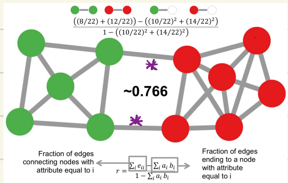


Edge Creation Tendencies

Homophily edges connect similar nodes
Heterophily edges connect different nodes

Quantifying homophily
assortativity

Measuring the probability of connecting 2 nodes that are similar:



Granovetter

Strength of weak ties*
Weak ties are actually more beneficial

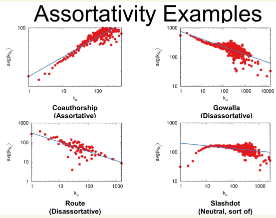
knowlede sharing!

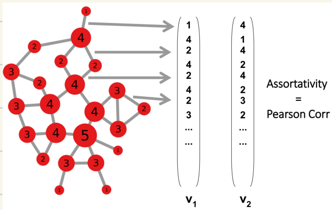
Echo Chamber

New edges with similar attributes,
Remove edges with different attributes.

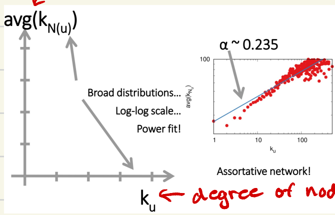
Degree assortativity

Nodes are generally connected with nodes of a similar degree.



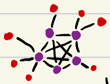


avg. degree of neighbors



← degree of node

Core-Periphery



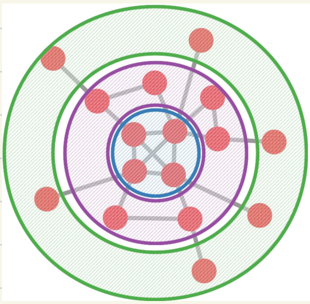
Core - periphery

edges either; C-C OR C-p

Periphery is low-degree nodes, Core is high-degree densely connected

Pure C-p is too strict and rare, Instead;

Continuous model



Core

Outer Core

Periphery

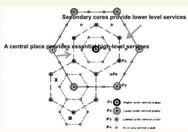
Community Discovery

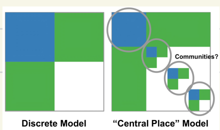
must be different since CP only views 1 community... But each community has a Core-Periphery model within itself..

Hotelling's law :

Service-Rich center; Copenhagen city center

mobility link; People outside travel in





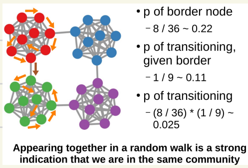
Community Discovery

Definition

A community is densely connected within itself, and sparsely connected to the outside...

Random walk approach

In a random walk you have;
- ↳ way higher probability of choosing nodes that are part of same community
- ↳ walking an edge indicates with high probability being part of same community



Label Propagation

Every node in own community
Look at all neighbors
adopt majority community, ties broken randomly

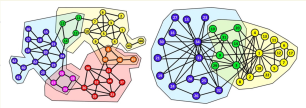
Modularity
to evaluate quality of partition
BUT!*

1. Build null-expectation (Random Graph)
2. Observe & subtract null-expectation

Disassortative -1 ... 0 No community ... 1 community

$$\underbrace{\frac{1}{2|E|}}_{\text{normalised}} \sum_{vu} \Big[\overbrace{A_{vu}}^{\text{connected? 1 or 0}} - \frac{\overbrace{k_v k_u}^{\text{degrees}}}{\underbrace{2 \cdot |E|}_{\#\text{edges}}} \Big] \underbrace{\delta(c_v, c_u)}_{\text{in same community?}}$$

* Not always... M tends to prefer communities of a given size... Does not like small ones ☹ $\sim \sqrt{|E|}$
↳ merges 2 cliques when small
Also might give high M for disconnected cliques

Temporal communities	What if we have a time dimension? ↳ communities can grow, shrink, merge, born, die
Hierchical Community "Not all levels will have valid communities for our initial assumption"	When we consider communities nested within a higher grain community. ↳ higher grain community might have lower internal density, but can still offer meaning Bottom-up: start with network. Do community discovery. Communities are now nodes in new network. Repeat! Top-down: everything in one → smaller High edge-betweenness nodes are connecting communities by def.
Overlapping Communities 	Nodes belong to several communities ↳ we can split node in 2 (1 for each community) CONGA: using Pair edge betweeness; Merge nodes with min. score +sum scores repeat until $\|E\| = 1$ ↳ we can cluster edges (edges considered as a group in bipartite network) then: weighted projection on edges, community discovery, and move back to graph!

Bipartite Network	CD on projection $\Rightarrow$ information loss, unclear mapping Solution: Bi-clique Percolation

TA Guidance

Helpful Measures	Measure	Networkx
many connections?	Density	density(G)
triangles close?	Clustering coeff	(average-)clustering(G)
	Diameter	diameter(G)
	Path length dist	

Purpose

What do these measures say about our particular network?

Use gephi → experiment w. layout types
→ plot with positions

General Focus areas: explain method
explain application
explain results

Remember Limitations:
Randomness?
Our assumptions?