Lecture 6 - Vertex Centrality

Thursday, February 5, 2026 9:22 AM

Plan for today:

- Ashvin presents: Study on centrality measures in social networks: a survey
- Talkin' proj
- Quick review (triadic and strong triadic closure)
- Vertex centrality
- Diffusion and epidemiology
- CODE MODE (SIR model+centrality)

Talkin' proj

- proposal presentation in a couple weeks
- order will be random
- grad students: also must give lit review as part of your proposal

IMPORTANT → must use or consider graph data and use some kind of "graph mining" technique

(Note: theory studies prob. OK, but

(Note: theory studies prob. OK, but double check with me)

Topic examples:

* studying real-world network properties, especially with novel data
* prediction → link prediction, metadata labels, edge direction
* classification → determining properties of verts | edges
* clustering → classification method, finding communities
* diffusion / simulation → how information "flows" or propogates through network topology
* data creation → creating a novel graph dataset
* visualization → enabling the visual study of graph data

visual study of graph data
and data + above GM stuff

* etc. → ask me

<u>Centrality</u> → vertex centrality

↳ a measure of "importance" or "influence" for a vertex in a network

Note: usually defined topologically locally or glabally

Who has high centrality:

Social networks: popular or important people

Road / infrastructure: important intersections or junctions

Epidemiology: someone with a lot of contact with others OR limited contact

others OR limited contact
but with important people

Types of centrality

Degree centrality:

Defined as the degree or # of connections for a vertex

$\rightarrow d(v)$ or $|N(v)|$

For directed graphs:

$d^+(v)$, $d^-(v)$, or both

out degree, in degree

Pros:

- Easy to calculate
- Gives good insight into local spread of a diffusive process
- Overall useful in various contexts

Cons:

- Easy to falsify (spam, bot nets)
- Local measurement → not useful

Local measurement → not useful
for diffusion over entire graph

Closeness centrality

Average shortest paths lengths
↳ How "close" a vertex is to
all other vertices

Diffusion: how quickly is the rest
of the network be impacted

Pros: More useful global measure
for diffusive processes
Captures a lot of the same
usefulness as degree centrality

Cons: Computationally expensive
to compute for *all* vertices
(quick for a small subset)
Can be a skewed measure
if attached to a hub

Betweenness Centrality

↳ ratio of shortest u,v-paths that contain some vertex w over all u,v-paths

→ The amount of diffusion that travels through a vertex

Pros: Global measure, very useful for understanding diffusive processes

Cons: Also pretty expensive (can approximate if needed)

Eigenvector centrality

Note: more to come Monday (Pagerank)

→ Basically: a vertex is important if its neighbors are

Consider adjacency matrix A

solve for x: $Ax = \lambda x$

eigenvector   eigenvalue

Pros: Easier to calculate (power iteration)

Useful in ways that pure path-based centrality measures are not

Cons: Doesn't exactly/nicely correlate with diffusions

Difficult to intuitively infer

Directed graphs

Note: many of the above measures can apply directly to digraphs

↳ consider directed paths, degrees, or adjacency matrix

One specific measure:

Harmonic centrality

Harmonic centrality

↳ Reachability: How many can I reach? How many can reach me? And how quickly for both?

Pros: easy to calculate

captures similar information to degree / closeness centrality

Cons: doesn't differentiate many vertices

Diffusive processes

Previously: we focused on strength or importance of edges

(explitly defined)

Now: we'll consider how vertex centrality impact diffusion

HOW: Simulate an SIR model
and infect / remove vertices
based on centrality

Epidemiology

↳ defined as the study of
how diseases spread

For GM: How a disease spreads on
a network topology

Disease spreading: diffusive process

↳ we'll create an explicit model,
run it on a network, and
observe average results

Our model: SIR

$$(S) \longrightarrow (I) \longrightarrow (R)$$

susceptible   infected   removed
↓

susceptible → not infected but can be

infected → can infect others

removed → can't be infected, can't infect others

Our approach:

- Initialize some populations as I, R
- Rest are in S
- p = probability of transmission per I, S interaction
- Iterate until $I = \emptyset$

Assumptions:

- p is fixed
- interactions are fixed
- duration of infection is fixed

What we're looking for:

How do our centrality measures correlate with diffusion dynamics

→ Given initial I, R sets, what total # of people infected and

total # of people infected and
how long does the epidemic last?