Useful for homework:

create BFS() function

BFS(sources, direction, omissions)

- sources → where we're searching from
- direction → forward / backward / both
- omissions → vertices we ignore

For forward-back algo to get SCC, IN, OUT

```
sources = verticesInMaxSCC()

FW = BFS(sources, "forward", {})
BW = BFS(sources, "backward", {})

SCC = FW.intersection(BW)
IN = BW - SCC
OUT = FW - SCC

TUBE_BW = BFS(OUT, "back", SCC)
TUBE_FW = BFS(IN, "forward", SCC)
TUBES = TUBE_BW.intersection(
            TUBE_FW)
```

Centrality (vertex centrality)

Centrality (vertex centrality)

Centrality is a measure of "importance" or "influence" for some vertex in a network

Note: Usually defined topologically locally or globally (sp?)

Who has high centrality?

Social nets: popular or important people

Road/infrastucture: important intersections or roads (edges)

Epidemiology: someone with a lot of interactions OR

someone with a lot of weak ties to disparate clusters

Types of centrality

Types of centrality

Degree centrality:

Defined as the degree or number of connections from a vertex

→ $d(v)$ or $|N(v)|$

For directed we might consider

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

out degree, in degree

Pros:

Easy to calculate

# of connections gives good insight into quick spread of a diffusive process

Useful in various contexts

Cons:

Easy to falsify (spam networks)

Not super useful for determining how a diffusive process spreads in the whole network

↳ very local measurement

↳ very local measurement

Closeness Centrality

How "close" a vertex is to all other vertices in a network

→ (Avg. shortest path length)$^{-1}$

Diffusion: How many hops to reach most of the rest of the network

Pros:

Loosely correlates with how fast diffusion occurs from a given vertex

Cons:

Computationally expensive (though quick for a subset of vertices)

Betweenness Centrality

The amount of diffusion through a single vertex (or edge)

Ratio of shortest u,v-paths for all

Ratio of shortest u,v-paths for all u,v that go through some vertex

Pros:

Determines "key" vertices for diffusive processes in a network

Cons:

Also pretty expensive

(approximation algos are possible)

Eigenvector Centrality

Note: we'll talk much more on Thursday (Pagerank)

Basically: a vertex is important if its neighbors are important

Consider Adjacency matrix A

solve for x: $Ax = \lambda x$

(↑ eigenvector, ↑ eigenvalue)

Pros:

Easier to calculate than the above via power-iteration via power-iteration

Gives useful results for various applications

Cons:

Difficult for our feable (sp?) human minds to comprehend

Directed Graphs

Note: A lot of the above measures are directly applicable to digraphs

E.g., by considering directed paths or degrees or adj matrix

One specific measure:

Reachability (Harmonic centrality)

↙

either IN or OUT

How many vertices can I reach?

How many vertices can reach me?

Pro:

Easy to calculate and useful

Con:

Doesn't differentiate many vertices

Diffusive Processes

We've already discussed how diffusion might be affected by weak ties
↳ edge centrality measure

Now: We'll consider how vertex centrality might impact diffusion

How: We're going to statistically measure how diffusion is impacted in a epidemiological model

---

Epidemiology

Defined as the study of how disease spreads

(—

disease spread

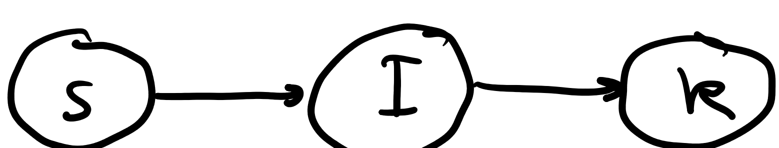
→ For us, how disease spreads on some network topology

Note: the spread of disease is a diffusive process

Here: we'll create an explicit model, randomness comes into play

Note 2: while "disease" has a specific definition, the general ideas here apply a lot more broadly (adoption of tech. spread of memes, etc.)

Our model: SIR



susceptible ↓ | infectuous ↓ | removed ↓

not be infected

↓ vertex is not infected but can be

↓ can infect other verts

↓ can not be infected and can not infect

How our model will run:

- Initializing some population as I, R
- Rest are in S
- p = probability of transmission per interaction
- Iterate until $I = \phi$

Our Network:

- Vertices are individuals
- Edges are interaction

Assumptions:

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

What we'll be looking for:

How does centrality impact diffusion dynamics diffusion dynamics

↳ specifically in our selection of initial I, R vertices