## Centrality Analysis

We're using centrality as a measure of importance

- Used to measure the **importance** (or "centrality") as in how "central" a node is in the graph of various nodes in a graph.
- Each node could be **important from an angle depending on how "importance"** is defined.
- So, **centrality comes in different flavors** and each flavor or a metric defines importance of a node from a different perspective

Now the meaning of importance comes from the angle we're looking at
diff methods of measuring centrality

## Normalized Degree Centrality

- Normalized by the **maximum possible degree**

$$C_d^{norm}(v_i) = \frac{d_i}{n-1}$$

how my degree is compared to highest possible degree
look at everyone else in
subtract 1 cuz this is me
no. of nodes
max that's supposed to be possible

- Normalized by the **maximum degree**

my degree vs the highest degree in my network.

$$C_d^{max}(v_i) = \frac{d_i}{\max_j d_j}$$

- Normalized by **the degree sum**

The first one is **meso-level analysis by Freeman** i.e. express this as a fraction of centrality in star network (which is the maximum)

$$C_d^{sum}(v_i) = \frac{d_i}{\sum_j d_j} = \frac{d_i}{2|E|}$$

my degree over the sum of all possible degrees.

## Interpretation of normalized degree centrality

Linton Freeman (one of the authors of UCINET) developed basic measures of the **centrality of actors**

1. **based on their degree** — no. of friends I have.
2. **the overall centralization of graphs.** — how popular I am

As indicated earlier, this can be calculated based on out degree, in degree or ignoring direction.

If interested in comparing across networks, this can be standardized as a % of nodes (no of nodes in the network -1 ) i.e. ego is excluded

**Graph centralisation** measures : since **star network is the most centralized network,** express the **degree of inequality or variance** in our **network as a percentage of that of a perfect star network of the same size.**

This is like a normalized degree centrality where normalization done wrt maximum possible degree.

star connected network.
Hence we sub 1
highest degree
that's $\frac{d_i}{n-1}$
me vs best case
star network

More on degree centrality
① Doesn't look @ direction of lines — idc who goes where.
② Binary symmetric data — cuz you're looking @ undirected adj. matrix.
③ Boohoo edge weights — introduces skew.

- A clear disadvantage of degree centrality is it can only be **used to make comparison between actors on the same network**
  - If you want to compare between **actors of two different networks** , they **have to be of the same size as measures**

degree centrality sucks cuz it only works in the network you're a part of
I can't compare

**same network**

- If you want to compare between **actors of two different networks** , they **have to be of the same size as measures are based on size of the network**
- **Hence, Freeman** converted centrality scores into proportions i.e. **normalized centrality**

So to normalize the centrality — Freeman!

you're a part of
I can't compare two completely diff networks
like a BITS 9 GPA vs a BMS 9 GPA.

- Should you compare networks that differ **greatly in size** ?
- **NO**
  - **An actor in a small network**, by virtue of small number of actors in the network, will have a **greater chance of connecting to other actors**
  - **Hence higher normalized degree centrality** !
  - **Smaller networks then tend to have higher density** scores !!

Same BITS vs BMS GPA stuff

fish in pond thinks it's the king of the world
but fish goes to ocean a big diff.

So close, higher density
again, higher connectivity.

normalize even more

It is all **about neighborhood** !!

A node with a high degree centrality may be **capable of affecting a lot of neighbors in its neighborhood** at once, but **we cannot say anything about the opportunities for global outreach**.

कूप मण्डूक

If you perceive the direction of ties are important, you may prefer **indegree and outdegree centrality**

Now sometimes, I may need to know in which directions the relationships go

In this case, we have to look @ indegree & outdegree centrality

- These are **calculated on digraph**
- Indegree centrality : no of ties received by an actor
- Outdegree centrality : no of ties given by an actor
- **Indegree centrality** : measure of **prestige** or **popularity**
- **Outdegree centrality** : measure of **expansiveness** or **gregariousness**

how many people want me

how many people I want.

how many wanna be friends w/ me

how many I wanna be friends with

ofc, we need digraph for this

how much I put myself out there

how many people wanna be w/ me
POPULARITY



Namita Achyuthan

## Degree Centrality in Digraph

- In directed graphs, we can either use the in-degree, the out-degree, or the combination as the degree centrality value:

$$C_d(v_i) = d_i^{in} \qquad (prestige),$$
$$C_d(v_i) = d_i^{out} \qquad (gregariousness),$$
$$C_d(v_i) = d_i^{in} + d_i^{out}.$$

popularity

"fond of company"
greg = group
gregarious = belonging to flock.

how much I put myself out there.

$d^{out}_i$ is the number of outgoing links for vertex $v_i$

# Betweenness Centrality

for when I compare myself on global scale

SOCIAL COMPUTING

### Intuition

If I have to, HEAVY normalize

Can't compare actors from diff networks

That's a drawback isn't it?

Degree centrality does **NOT consider the rest of the network** ( only look at immediate ties or **ego network** )

- Sometimes it is **not important how many people you know but rather where you are placed in the network**
- In communication networks, betweenness centrality measures how **much potential control an actor has over the flow of information**
- More specifically, **it measures how many times an actor sits on the "geodesic" ( shortest path) between two other actors** together

So it's not always how imp. I am, it's where I stand in the big picture (network)

Betweenness centrality
how much control I can have over info flow

How many times I sit in way of info flow

## Betweenness centrality

In other words, we are measuring how central $v_i$'s role is in connecting any pair of nodes **s** and **t**. This measure is called betweenness centrality. This is a normalized definition

$$C_b(v_i) = \sum_{s \neq t \neq v_i} \frac{\sigma_{st}(v_i)}{\sigma_{st}}$$

no. of shortest paths from s to t that pass thro' me

no. of shortest paths (info flow path) b/w s,t. If 0, that's not a path there, discard

how often I sit in the shortest possible path & control info flow

i.e. how central I am in connecting two nodes.

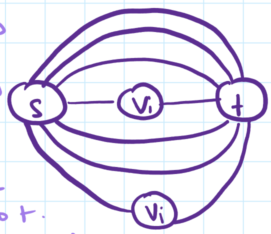
$\sigma_{st}$ **the number of shortest paths from vertex s to t – a.k.a. information pathways**

$\sigma_{st}(v_i)$ **the number of shortest paths from s to t that pass through $v_i$. Terms where $\sigma_{st}(v_i)$ =0, are not considered, since this means that s and t belong to different components**

$$C_b(v_i) = \sum \frac{\sigma_{st}(v_i)}{\sigma_{st}}$$

no. of times $v_i$ sits in the way

no. of shortest paths from s to t. (info flow pathways)




* if $\sigma_{st}(v_i) = 0$, st belong to diff. components.

Best case? $v_i$ is everywhere — sits on all paths of info flow.
$\Rightarrow \sigma_{st}(v_i) = 1$.

Namita Achyuthan

Best case? $v_i$ is everywhere — sits on all paths of info flow.

$\Rightarrow \frac{\sigma_{st}(v_i)}{\sigma_{st}} = 1.$

∴ Max value:

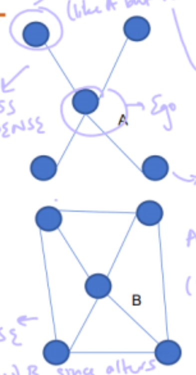
$$C_b(v_i) = \sum 1 = (n-1)(n-2) \quad \text{ie } \underbrace{2}_{2\times}\underbrace{\binom{n-1}{2}}_{\text{from } n-1 \text{ select } 2}$$

So for normalized betweenness centrality:

$$C_b^{norm}(v_i) = \frac{C_b(v_i)}{(n-1)(n-2)}$$

## Betweenness centrality and structural hole

- **Ronald Burt** coined term "structural holes"
- In the figures, A's alters are unconnected but B's alters are completely connected
- A is getting diversity of advice whereas B is getting same information mostly . This structural holes will benefit A more than B
- B has a **lower betweenness** centrality score compared to A.
  - So, **high betweenness centrality** is characterized by **more structural holes** in ego network !!
- A's ego network density score is 0.00 and B has 0.6667
  - So, for ego network, **less density score** means **more structural holes !!**




A's alters (like A but not exact)

Bird's of a feather principle.

LESS DENSE

Ego

Alters not connected → A's getting more new info (like a cult leader / disconnect followers) → High between-ness cuz all new info to A

MORE DENSE

But w/ B, since alters are all connected, not much betweenness cuz he's not the alpha

uncontrolled subordinates

17

Less density ⇒ more structural holes
↓
more diversity of info
↓
more beneficial for A.

# Closeness Centrality

## Closeness centrality

- The intuition is that **influential and central nodes can quickly reach other nodes.** It shows **how close the node is to the rest of the graph**.
- This centrality (normalized) is also in the range from **0 (the node has no neighbours; it is severed from the rest of the network) to 1 (the node is the hub of the global star and is one hop away from any other node).**
- Most central nodes should have a **smaller average shortest path length to other nodes** i.e. defined as the **reciprocal mean distance** (length of the geodesics) from a node to all other reachable nodes in the network → Help
- **The smaller the average shortest path length (geodesic) i.e. distance, the higher the centrality i.e. closeness for the node.** makes sense cuz all popular kids are one call away from info.

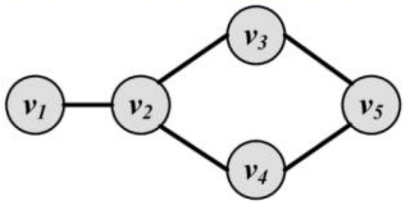
Closeness centrality: $C_c(v_i) = \frac{1}{\bar{l}_{v_i}}$, where $\bar{l}_{v_i} = \frac{1}{n-1}\sum_{v_j \neq v_i} l_{i,j}$



Namita Achyuthan

## Calculate

Closeness centrality = $\dfrac{1}{\left[\dfrac{\text{no. of hops to other nodes}}{n}\right]}$




$C_c(v_1) = \dfrac{1}{\left[\dfrac{1+2+2+3}{4}\right]} = 0.5$

$C_c(v_2) = \dfrac{1}{\left[\dfrac{1+1+1+2}{4}\right]} = 0.8$ → the avg. dist is low if high centrality → so we reciprocate

$C_c(v_3) = \dfrac{1}{\left[\dfrac{2+1+2+1}{4}\right]} \sim 0.66$

You get the point.

$$
\begin{aligned}
C_c(v_1) &= 1/((1+2+2+3)/4) = 0.5,\\
C_c(v_2) &= 1/((1+1+1+2)/4) = 0.8,\\
C_c(v_3) = C_b(v_4) &= 1/((1+1+2+2)/4) = 0.66,\\
C_c(v_5) &= 1/((1+1+2+3)/4) = 0.57
\end{aligned}
$$

## Harmonic Centrality

- In closeness centrality, we **did reciprocal of the mean distance**
- Another way to quantify the sense of closeness is to look at the **mean reciprocal distance** (as opposed to the **reciprocal mean distance** - the order of the sum and reciprocal operations reverses). Such measure is called **harmonic centrality.**

Reciprocal mean distance = closeness centrality = $(\text{mean dist})^{-1}$

Mean reciprocal distance = $\sum \frac{1}{n} \times (\text{distance})^{-1}$