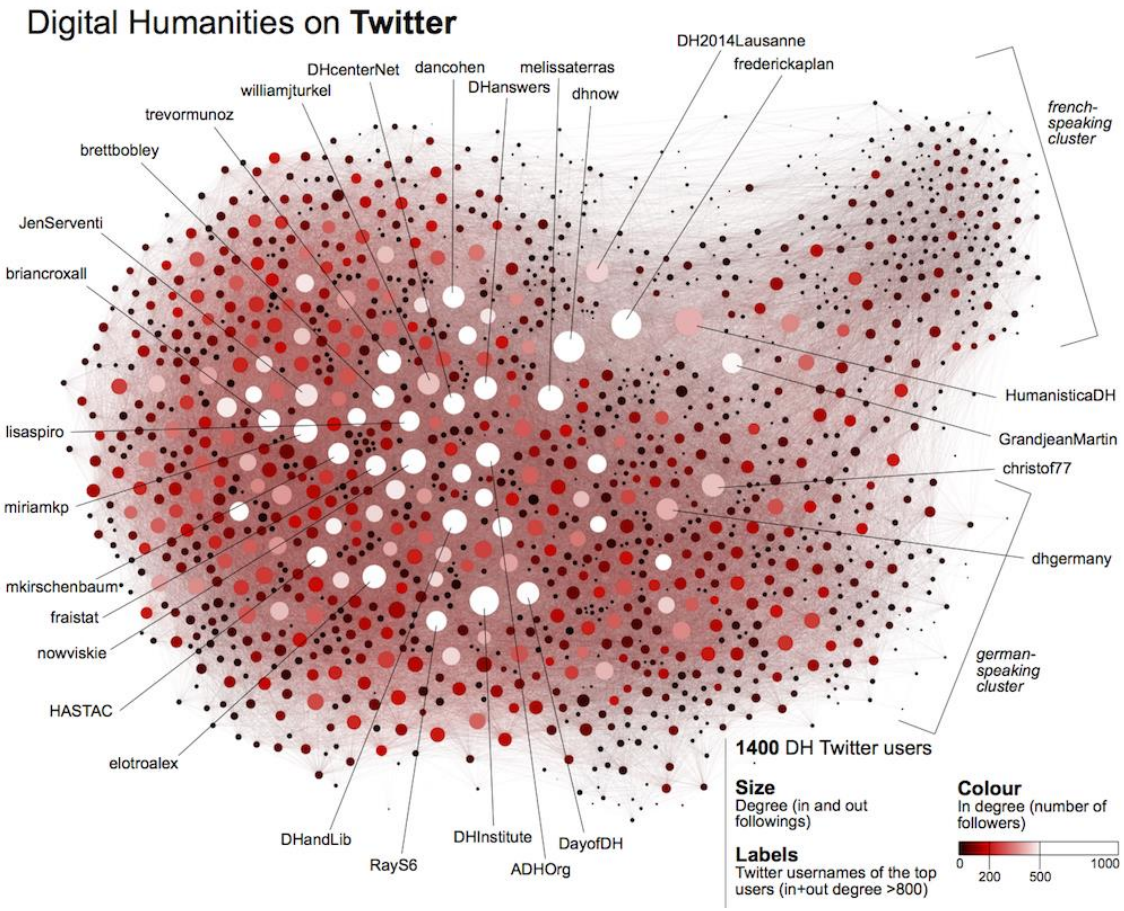
Mining & Learning on Graphs

Network Analysis

Yu Wang, Ph.D.
Assistant Professor
Computer and Information Science
University of Oregon
CS 410/510 - Fall 2024

Network Analysis - Motivation

Why network analysis?



Digital Humanities practitioners connect and interact on Twitter

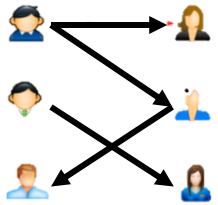
Key influencers

Community

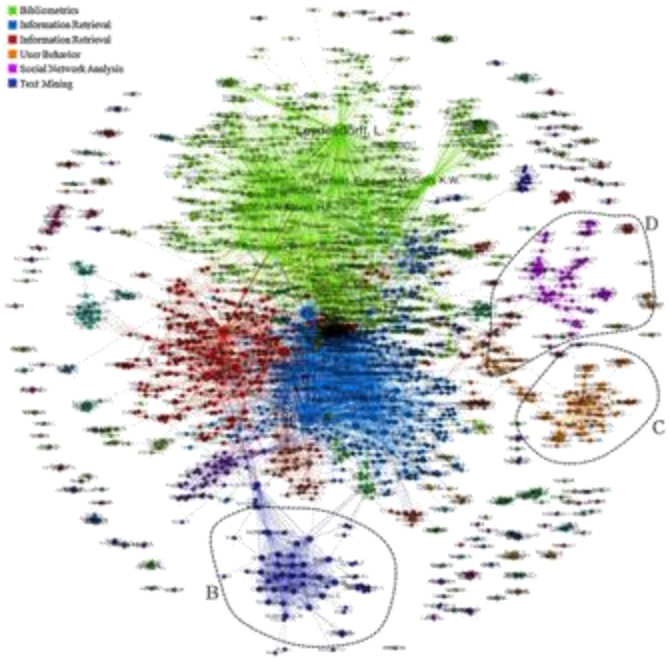
Language-based Cluster

Why network analysis?

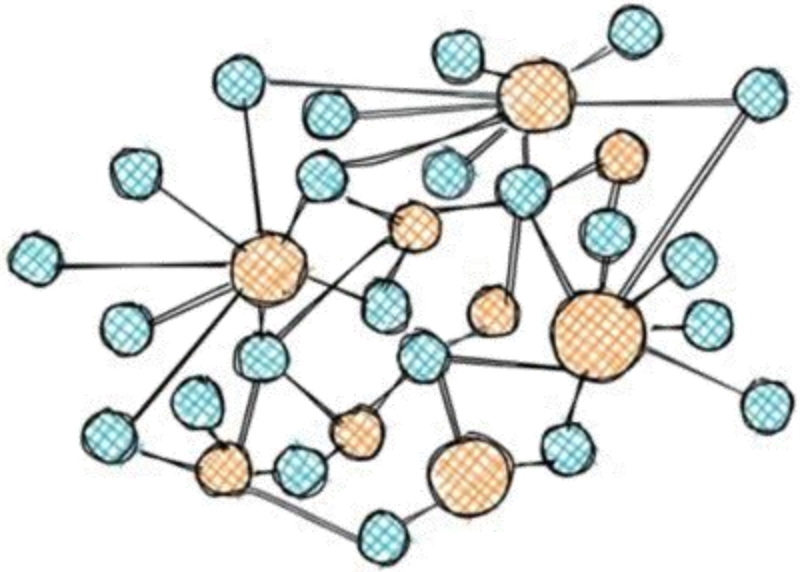
Birds of a feather flock together



Dating Network



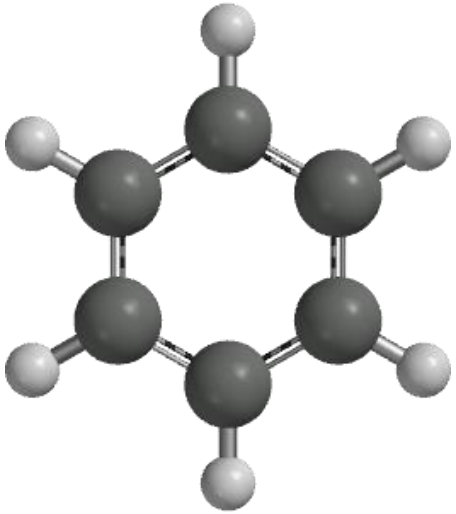
Homophily



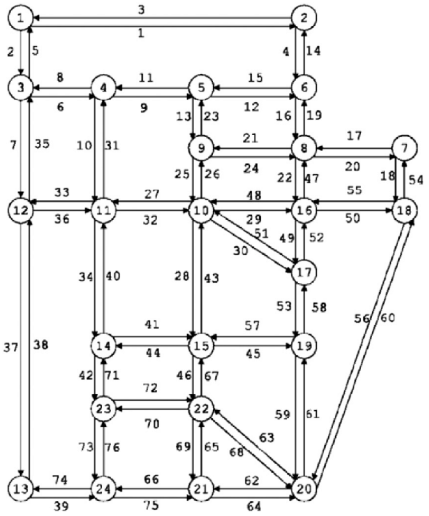
Heterophily

Network Analysis - Motivation

Why network analysis?



Benzene:
6-cycle



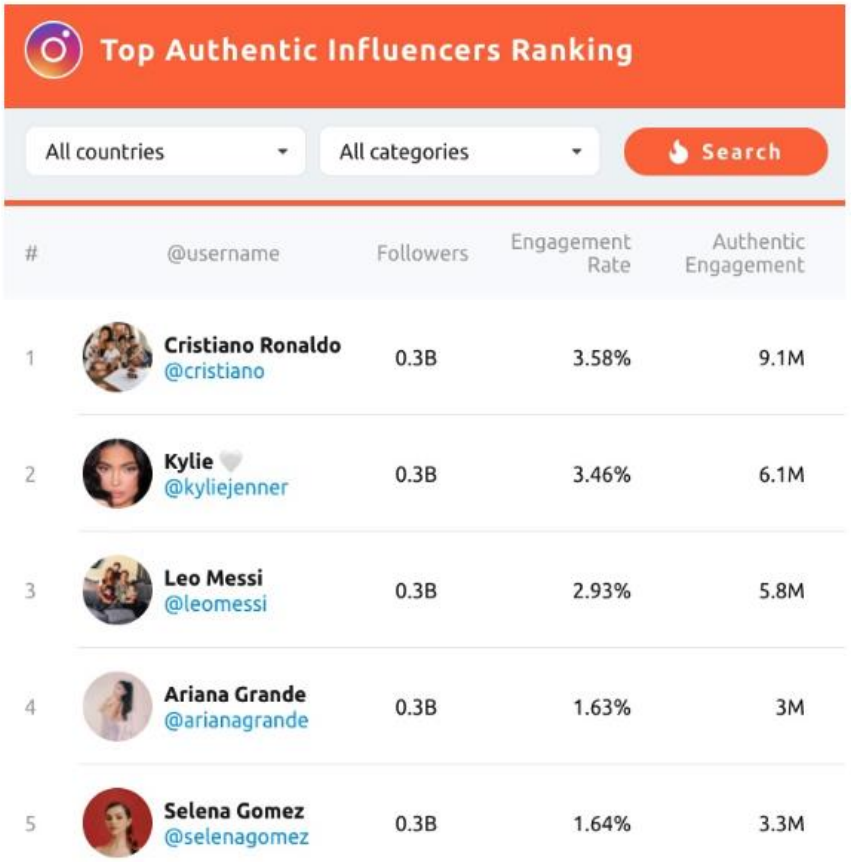
Road:
4-cycle

Network Centrality

- Which nodes in the graph are “important”?

Motivation

- Which users in a social network are more influential?

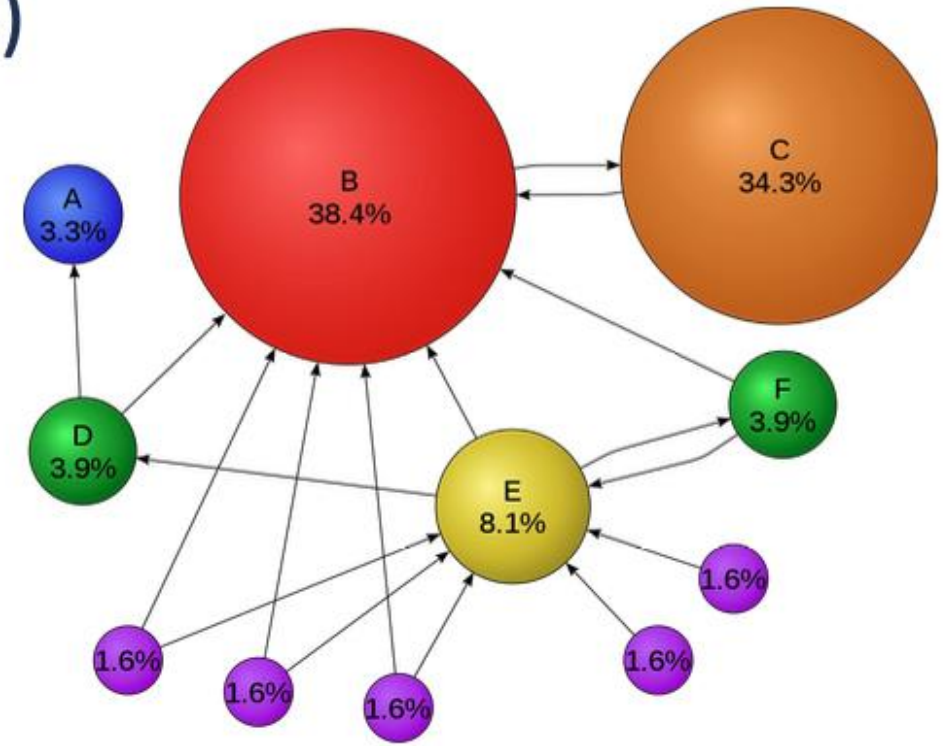


Note: Influence maximization will be a dedicated topic later

Motivation

- Which employees are the most central to a company?
- Which web pages are more likely of interest/important to show you?

Is a measure that ranks the nodes of a graph to answer the mentioned types of questions (and more!)



Note that today we will focus on undirected graphs

Network Centrality

- Basic measures of centrality/prestige
  - Mostly from a social perspective
- Path-based
  - Closeness, Betweenness, Katz
- Eigenvector-based
  - Eigenvector, Katz, PageRank
- Others
  - Hubs and Authorities, Goodness and Fairness

Network Centrality

- Basic measures of centrality/prestige
  - Mostly from a social perspective
- Path-based
  - Closeness, Betweenness, Katz
- Eigenvector-based
  - Eigenvector, Katz, PageRank
- Others
  - Hubs and Authorities, Goodness and Fairness

Network Centrality – Sociology Perspective

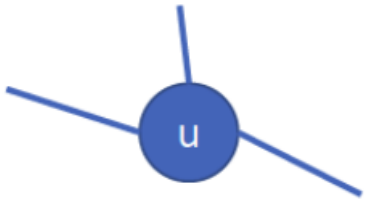
- For a social network let us consider the graph of "actors" (i.e., nodes) and then defines "importance" measures accordingly

- Actor prestige
  - Motivation: The more ties an actor has the higher their importance
  - Note in a directed network would this be incoming or outgoing?
- Actor centrality
  - Motivation: Involvement with other actors, how many ties, who are they linked with, etc.

Undirected Degree Centrality

- Actor prestige (i.e., Degree Centrality)
  - Motivation: The more ties an actor has the higher their importance

$C_D(v) = 6$



$C_D(u) = 3$

Directed Degree Centrality

- Actor prestige (i.e., Degree Centrality)
  - Motivation: The more ties an actor has the higher their importance
  - Note in directed networks we use incoming links (and not outgoing)

$C_D(v) = 1$

$C_D(u) = 2$

E.g., Twitter

- Why we might prefer using incoming (i.e., "Followers") here, but this still isn't taking full information into account when deciding a ranking

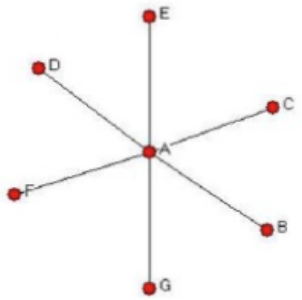
Problems: Degree Centrality

- Undirected Degree Centrality: number of neighbors 1-hop away from the node

$$C_D(i) = k_i = \sum_j \mathbf{A}_{ij} = \sum_j \mathbf{A}_{ji}$$

How to compare with other networks?

$$C_{D^*}(i) = \frac{1}{n-1} C_D(i)$$



Problems: Degree Centrality

$$C_D(i) = k_i = \sum_j \mathbf{A}_{ij} = \sum_j \mathbf{A}_{ij}$$

$$C_{D^*}(i) = \frac{1}{n-1} C_D(i)$$

Which users have high centrality?

Those with direct contact to many others

Which users have low centrality?

Those not very active or peripheral users

Network Centrality

- Basic measures of centrality/prestige
  - Mostly from a social perspective
- Path-based
  - Closeness, Betweenness, Katz
- Eigenvector-based
  - Eigenvector, Katz, PageRank
- Others
  - Hubs and Authorities, Goodness and Fairness

Path-based Centrality

- Built on the idea that the centrality of a node is in relation to how they are connected with others in the network

- Closeness Centrality:
  - How close an actor is to all other actors in the network

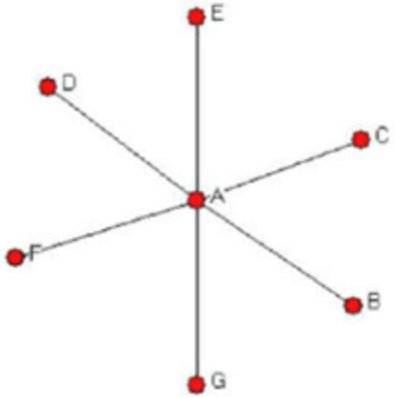
- Betweenness Centrality:
  - The # of shortest paths that flow through that actor

Closeness Centrality

Closeness Centrality: Alex Bavelas, 1948

- How close an actor is to all other actors in the network

$$C_C(i) = \frac{1}{\sum_j d(i,j)}$$



Problem: How to compare across graphs?

Normalized Closeness Centrality

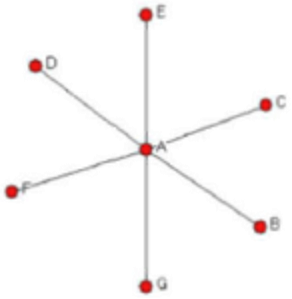
Closeness Centrality:

- How close an actor is to all other actors in the network

$$C_C(i) = \frac{1}{\sum_j d(i,j)}$$

Normalized Closeness Centrality:

$$C_{C^*}(i) = (n-1)C_C(i)$$



Problem: What if graph is disconnected?

Harmonic Centrality

Closeness Centrality:

$$C_C(i) = \frac{1}{\sum_j d(i,j)}$$

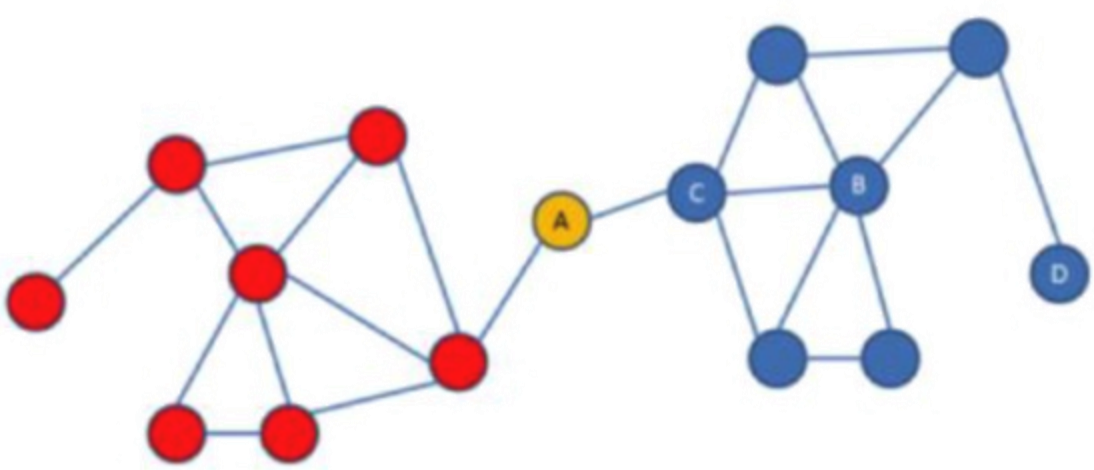
Harmonic Centrality:

Take the harmonic mean by inverting all distances

$$C_H(i) = \sum_j \frac{1}{d(i,j)}$$

Betweenness Centrality

- Betweenness Centrality: Linton Freeman, 1977
  - The # of shortest paths σ that flow through that actor



Betweenness Centrality

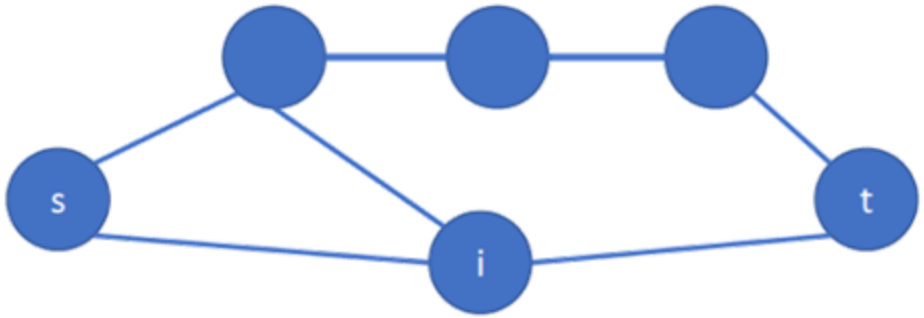
- **Betweenness Centrality:** Linton Freeman, 1977
  - The # of shortest paths σ that flow through that actor

$$C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

where

$\sigma_{st}(i)$ number of shortest paths through node I

σ_{st} number of shortest paths



$\sigma_{st}(i) = 1$

$\sigma_{st} = 1$

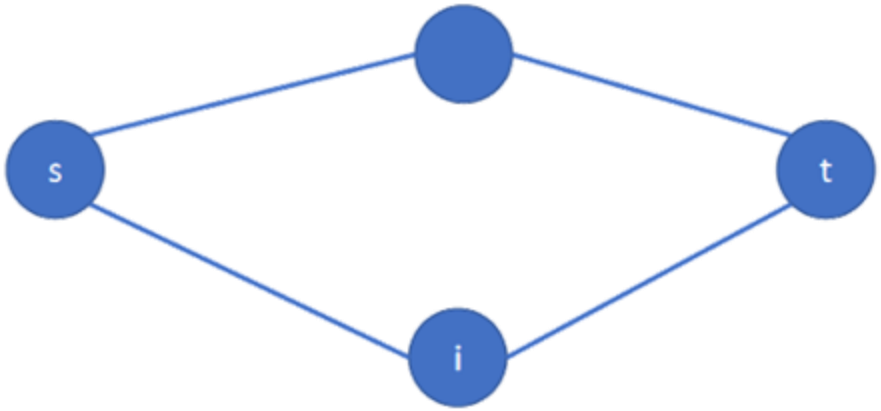
Betweenness Centrality

- Betweenness Centrality: Linton Freeman, 1977
  - The # of shortest paths σ that flow through that actor

$$C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

$\sigma_{st}(i) = 1$

$\sigma_{st} = 2$



Betweenness Centrality

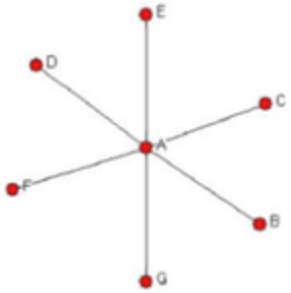
- **Betweenness Centrality:** Linton Freeman, 1977
  - The # of shortest paths σ that flow through that actor

$$C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

- **Normalized Betweenness Centrality:**

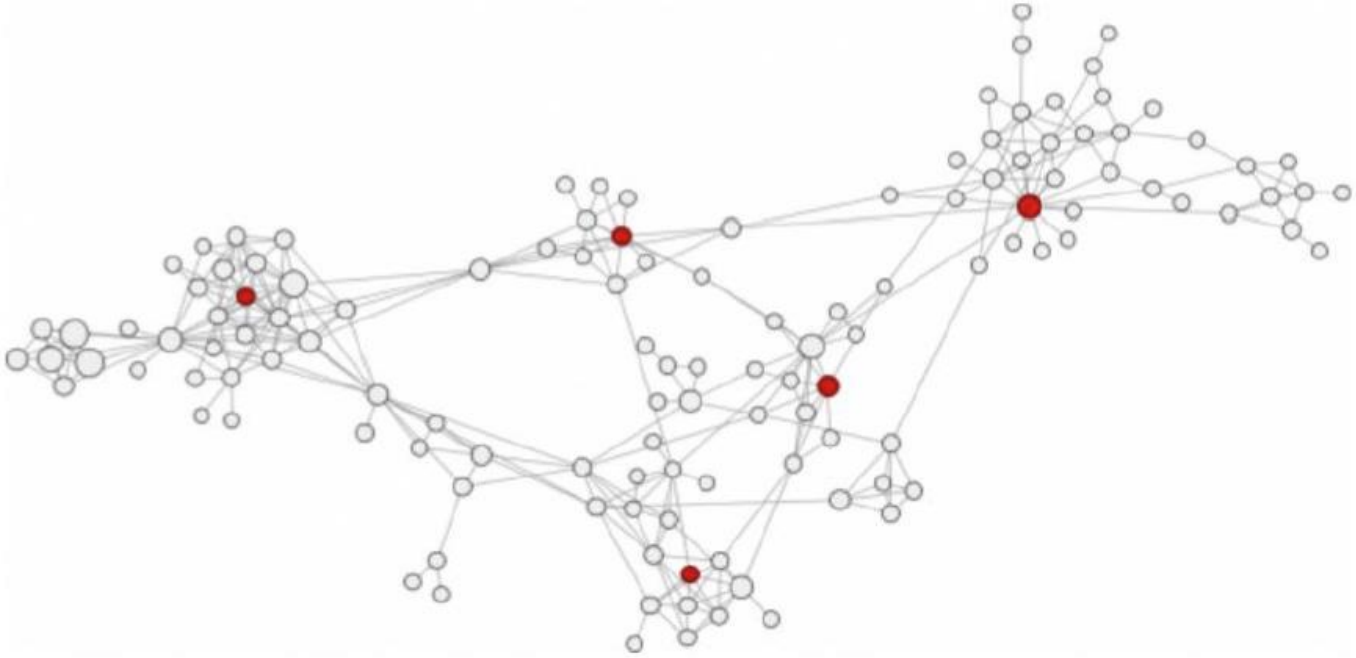
$$C_{B^*}(i) = \frac{2}{(n-1)(n-2)} C_B(i)$$

How did we get this normalization?



Recap early path-based

- Closeness Centrality



Recap early path-based

- Betweenness Centrality

Katz Centrality

Leo Katz, 1953

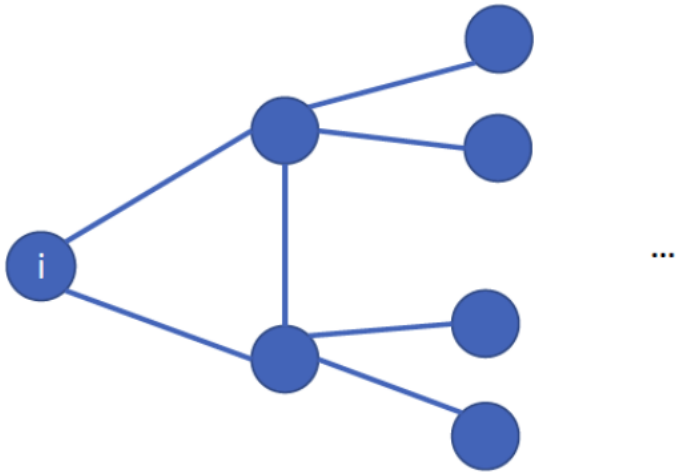
- Degree Centrality was based on counting neighbors
- What if we want to consider how many neighbors' neighbors, etc.

- Weighted count of all paths connecting to that node with a discount on longer paths
  - Weight of path of length n is β^n
  - [Note that $\beta < \frac{1}{\lambda_1}$ but we will not discuss this]

$$c_i = \beta \sum_j \mathbf{A}_{ij} + \beta^2 \sum_j \mathbf{A}_{ij}^2 + \beta^3 \sum_j \mathbf{A}_{ij}^3 + \ \ldots$$

Katz Centrality

$$c_i = \beta \sum_j \mathbf{A}_{ij} + \beta^2 \sum_j \mathbf{A}_{ij}^2 + \beta^3 \sum_j \mathbf{A}_{ij}^3 + \ \ldots$$

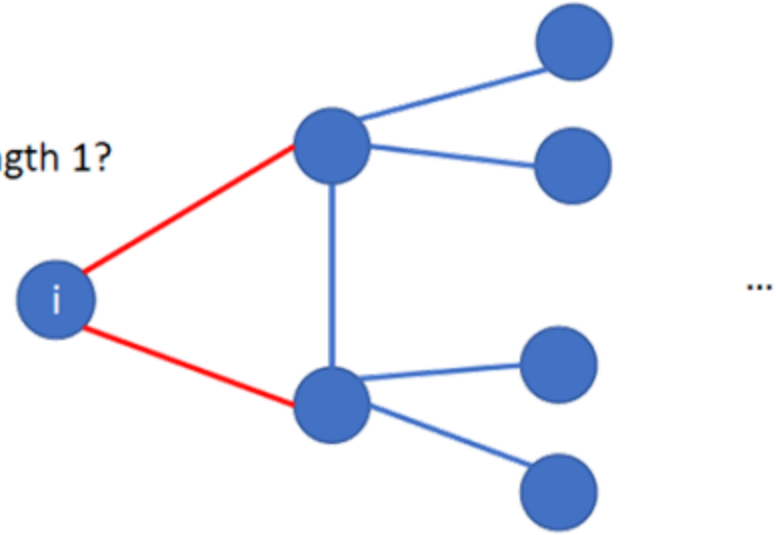


Katz Centrality

$$c_i = \beta \sum_j \mathbf{A}_{ij} + \beta^2 \sum_j \mathbf{A}^2_{ij} + \beta^3 \sum_j \mathbf{A}^3_{ij} + \ \ldots$$

How many paths of length 1?

2

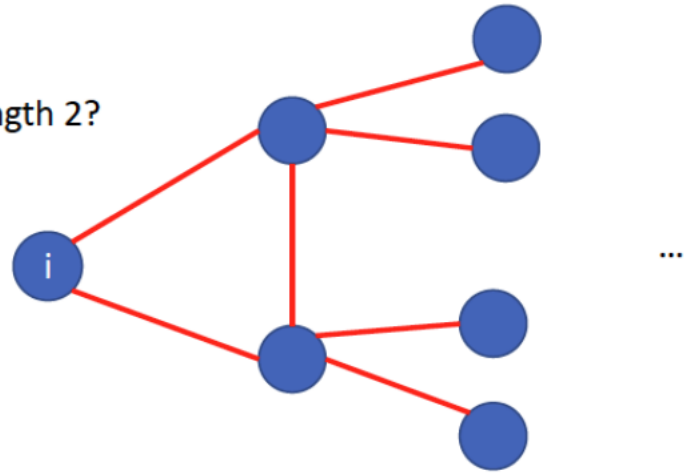


…

Katz Centrality

$$c_i = \beta \sum_j \mathbf{A}_{ij} + \beta^2 \sum_j \mathbf{A}_{ij}^2 + \beta^3 \sum_j \mathbf{A}_{ij}^3 + \ \ldots$$

How many paths of length 2?



...

Katz Centrality

$$c_i = 1 + \beta \sum_j \mathbf{A}_{ij} + \beta^2 \sum_j \mathbf{A}^2_{ij} + \beta^3 \sum_j \mathbf{A}^3_{ij} + \ \ldots$$

$$\mathbf{c} = (\mathbf{I} + \beta\mathbf{A} + \beta^2\mathbf{A}^2 + \beta^3\mathbf{A}^3 + \ \ldots)\mathbf{1}$$

$$\mathbf{c} = \sum_{n}^{\infty} (\beta^n \mathbf{A}^n)\mathbf{1}$$

$$\mathbf{c} = (\mathbf{I} - \beta A)^{-1}\mathbf{1}$$

Intuition:
Taylor series expansion

$$\frac{1}{1-x} \approx 1 + x + x^2 + \cdots$$

Modified Katz

- Newman 2010

$$\mathbf{c}_i = \alpha \sum_j \mathbf{A}_{ij} \boldsymbol{c}_j + \beta_i$$

$$c = (\mathbf{I} - \alpha \mathbf{A})^{-1} \beta$$

- Bonacich 2001 (Alpha-Centrality)

$$\mathbf{c}_i = \alpha \sum_j \mathbf{A}_{ij} \boldsymbol{c}_j + 1$$

$$c = (\mathbf{I} - \alpha \mathbf{A})^{-1} \mathbf{1}$$

1. Log in to DuckWeb
2. On the Main Menu, click the Course Surveys link
3. Click "Open the Course Surveys site"
4. Choose the course and send in your feedback

Any Question?