Network Analysis

Dr. Shaina Race
Institute for Advanced Analytics

Network Centrality

• • •

Measuring Influence in a Network

Centrality

- **Centrality** is a measure of how importance a vertex is to a graph.
- There are many ways to define importance.
- Measures usually normalized in [0,1] for comparison across networks
- Common measures of centrality:
  - **Degree Centrality**
  - **Betweenness Centrality**
  - **Closeness Centrality**
  - **Eigenvector Centrality**

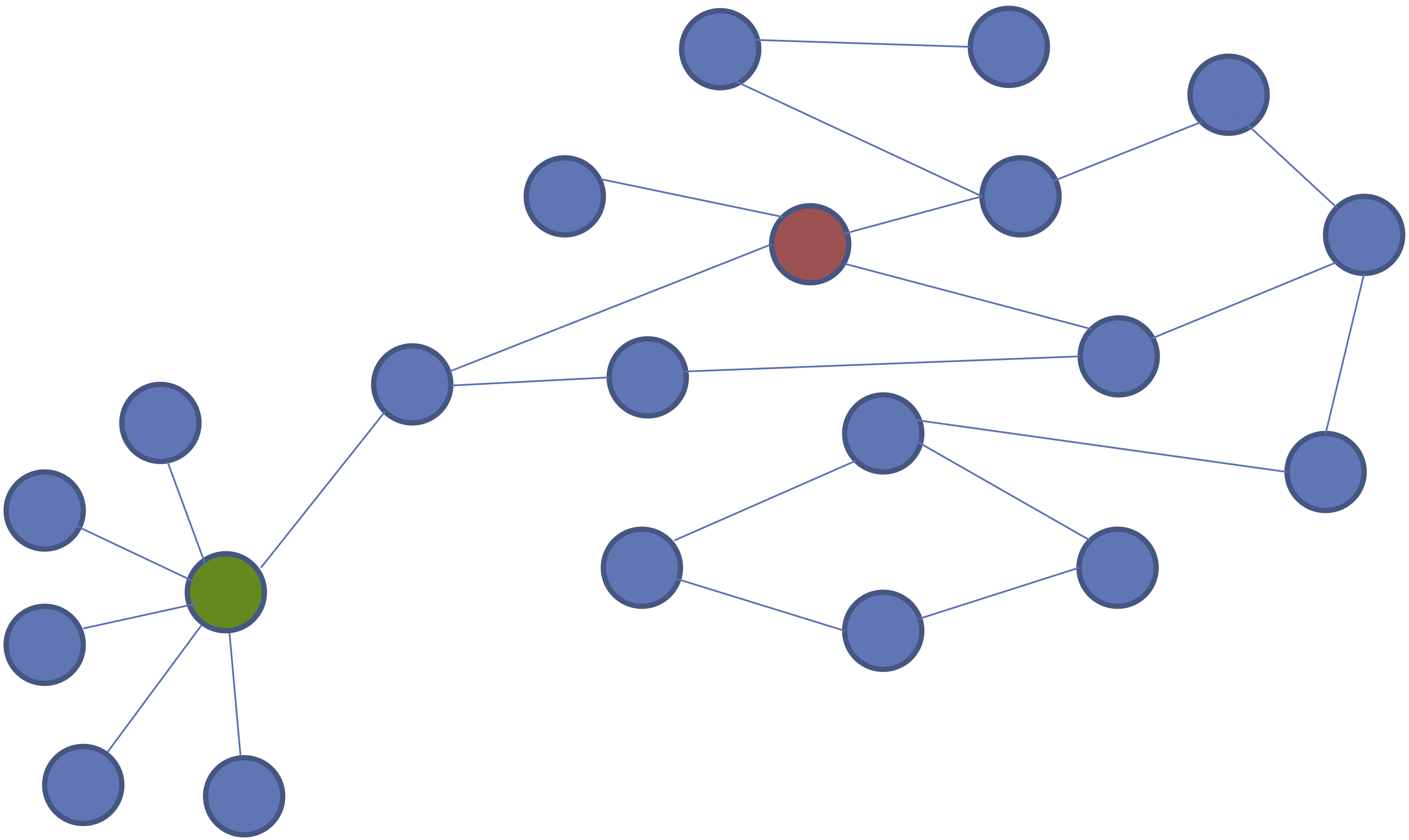
Centrality

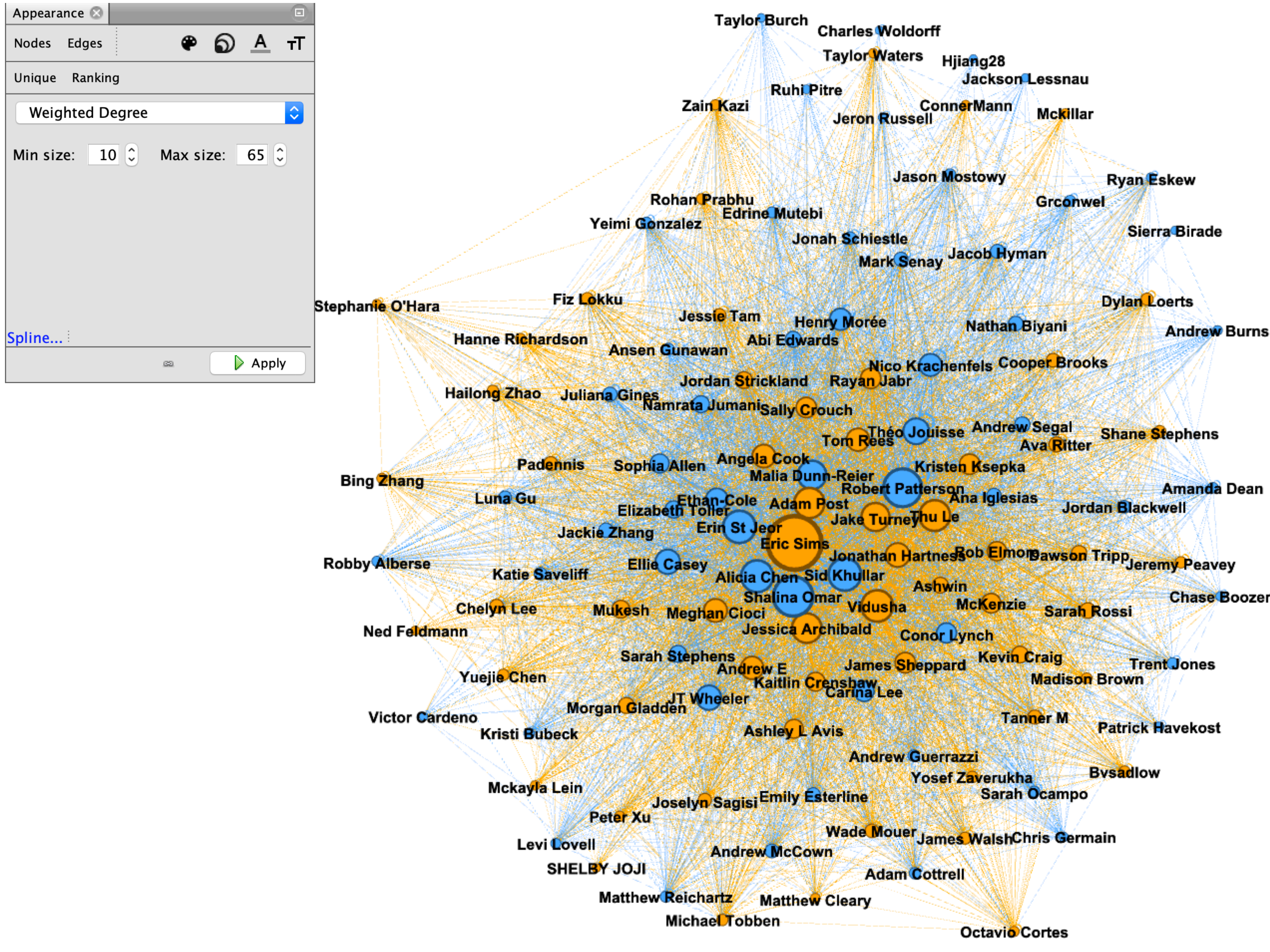
- **Centrality** is a measure of how importance a vertex is to a graph.
- There are many ways to define importance.
- Measures usually normalized in [0,1] for comparison across networks
- Common measures of centrality:
  - **Degree Centrality**
  - **Betweenness Centrality**
  - **Closeness Centrality**
  - **Eigenvector Centrality**

Degree Centrality

- **Measures the exposure of vertex to others in network**
- Degree centrality for vertex v is $c_D(v) = \dfrac{d_v}{n-1}$
  - d_v=degree of vertex v
  - n=number of vertices in network
- Max value is 1 (if vertex is connected to all other vertices).
- *Local measure*: can be deceiving

Deceptive Degree Centrality



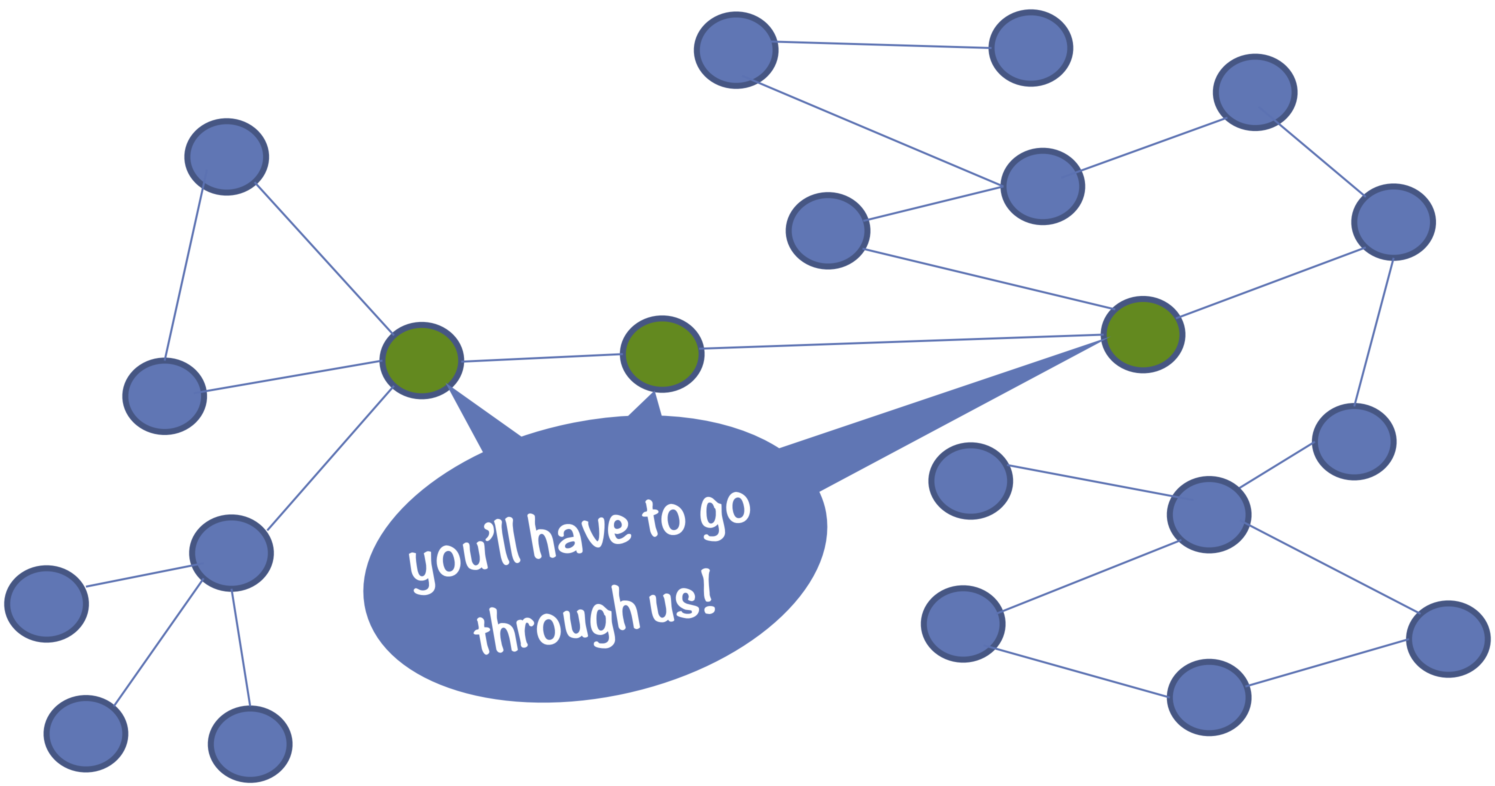

Appearance
Nodes
Edges
Unique
Ranking
Weighted Degree
Min size: 10
Max size: 65
Spline...
Apply
Taylor Burch
Charles Woldorff
Taylor Waters
Hjiang28
Jackson Lessnau
Ruhi Pitre
Zain Kazi
ConnerMann
Mckillar
Jeron Russell
Jason Mostowy
Ryan Eskew
Rohan Prabhu
Grconwel
Edrine Mutebi
Yeimi Gonzalez
Sierra Birade
Jonah Schiestle
Jacob Hyman
Mark Senay
Stephanie O'Hara
Fiz Lokku
Dylan Loerts
Jessie Tam
Henry Morée
Nathan Biyani
Andrew Burns
Hanne Richardson
Abi Edwards
Ansen Gunawan
Nico Krachenfels
Cooper Brooks
Jordan Strickland
Rayan Jabr
Hailong Zhao
Juliana Gines
Namrata Jumani
Sally Crouch
Théo Jouisse
Andrew Segal
Shane Stephens
Tom Rees
Ava Ritter
Angela Cook
Padennis
Sophia Allen
Kristen Ksepka
Malia Dunn-Reier
Bing Zhang
Robert Patterson
Amanda Dean
Luna Gu
Ethan-Cole
Ana Iglesias
Adam Post
Jordan Blackwell
Elizabeth Toller
Jake Turney
Thu Le
Erin St Jeor
Jackie Zhang
Eric Sims
Jonathan Hartness
Rob Elmore
Dawson Tripp
Jeremy Peavey
Robby Alberse
Ellie Casey
Katie Saveliff
Alicia Chen
Sid Khullar
Ashwin
Chase Boozer
Shalina Omar
Chelyn Lee
Mukesh
Meghan Cioci
Vidusha
McKenzie
Sarah Rossi
Ned Feldmann
Jessica Archibald
Conor Lynch
Sarah Stephens
Kevin Craig
Andrew E
James Sheppard
Trent Jones
Yuejie Chen
Kaitlin Crenshaw
Madison Brown
Carina Lee
JT Wheeler
Morgan Gladden
Victor Cardeno
Tanner M
Kristi Bubeck
Ashley L Avis
Patrick Havekost
Andrew Guerrazzi
Yosef Zaverukha
Bvsadlow
Mckayla Lein
Sarah Ocampo
Emily Esterline
Joselyn Sagisi
Peter Xu
Wade Mouer
James Walsh
Chris Germain
Levi Lovell
Andrew McCown
SHELBY JOJI
Adam Cottrell
Matthew Reichartz
Matthew Cleary
Michael Tobben
Octavio Cortes

Centrality

- **Centrality** is a measure of how importance a vertex is to a graph.
- There are many ways to define importance.
- Measures usually normalized in [0,1] for comparison across networks
- Common measures of centrality:
  - **Degree Centrality**
  - **Betweenness Centrality**
  - **Closeness Centrality**
  - **Eigenvector Centrality**

Betweenness Centrality

Measures **control** that each node has **over communication** between other nodes.



Betweenness Centrality

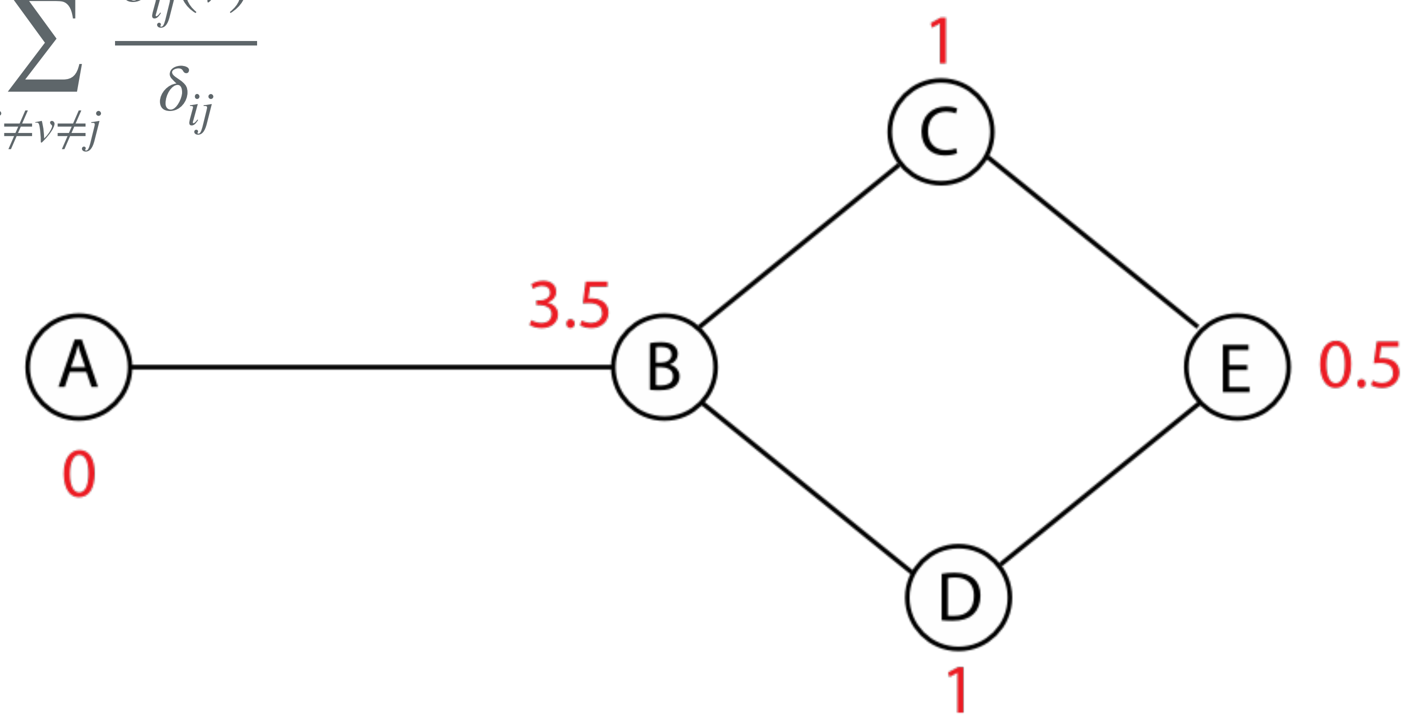
- Sum of proportions of shortest paths between 2 nodes that pass through the node of interest.

$$c_B(v) = \sum_{i \neq v \neq j} \frac{\delta_{ij}(v)}{\delta_{ij}}$$

  - δ_{ij}=number of shortest paths between nodes i and j
  - $\delta_{ij}(v)$=number of shortest paths between i and j that go through node u
- Can also consider *edge betweenness centrality* using paths that include a given edge.

Betweenness Centrality Example

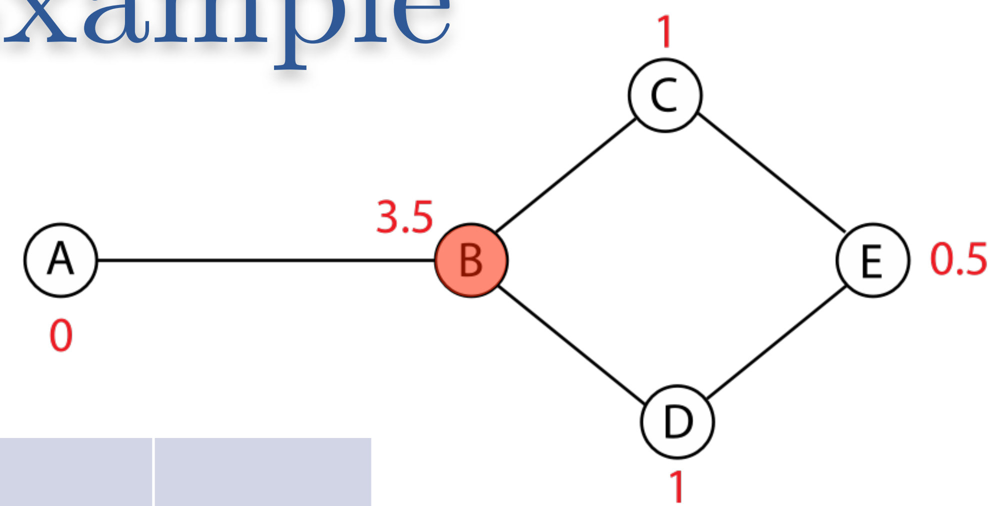
$$c_B(v) = \sum_{i \neq v \neq j} \frac{\delta_{ij}(v)}{\delta_{ij}}$$



<u>Note</u>: Be sure to use Brandes' Algorithm to compute betweenness! Developed in 2001, O(mn) for unweighted graphs and O(n^2logn+mn) for unweighted graphs

Betweenness Centrality Example

$$c_B(v) = \sum_{i \neq v \neq j} \frac{\delta_{ij}(v)}{\delta_{ij}}$$

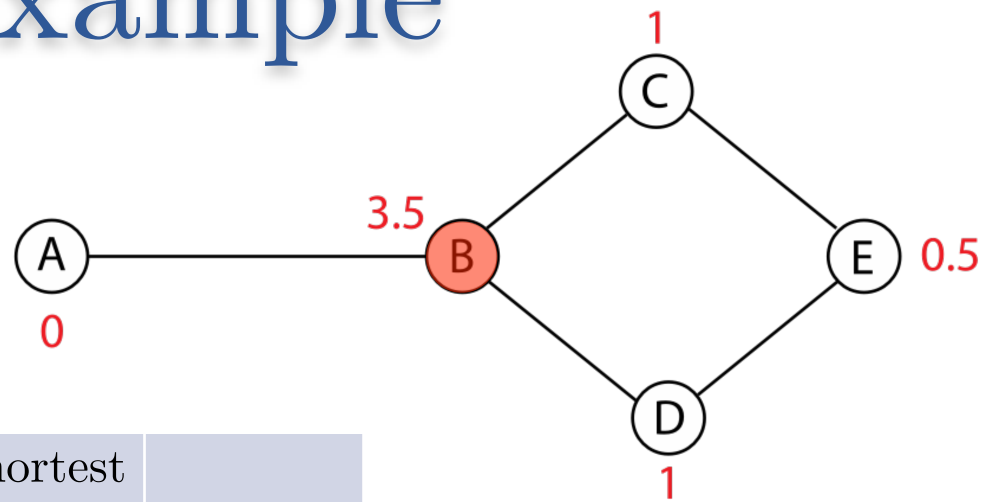


Source	Target		
A	C		
A	D		
A	E		
C	D		
C	E		
D	E		

Step 1: List all pairs of nodes, excluding the node of interest.

Betweenness Centrality Example

$$c_B(v) = \sum_{i \neq v \neq j} \frac{\delta_{ij}(v)}{\delta_{ij}}$$

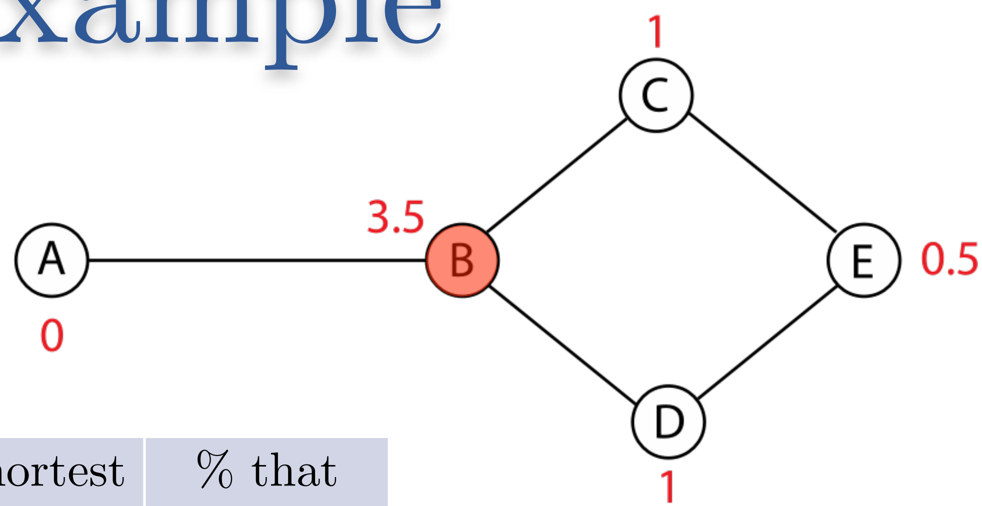


Source	Target	# of shortest paths	
A	C	1	
A	D	1	
A	E	2	
C	D	2	
C	E	1	
D	E	1	

Step 2: Compute the number of shortest paths between each pair.

Betweenness Centrality Example

$$c_B(v) = \sum_{i \neq v \neq j} \frac{\delta_{ij}(v)}{\delta_{ij}}$$

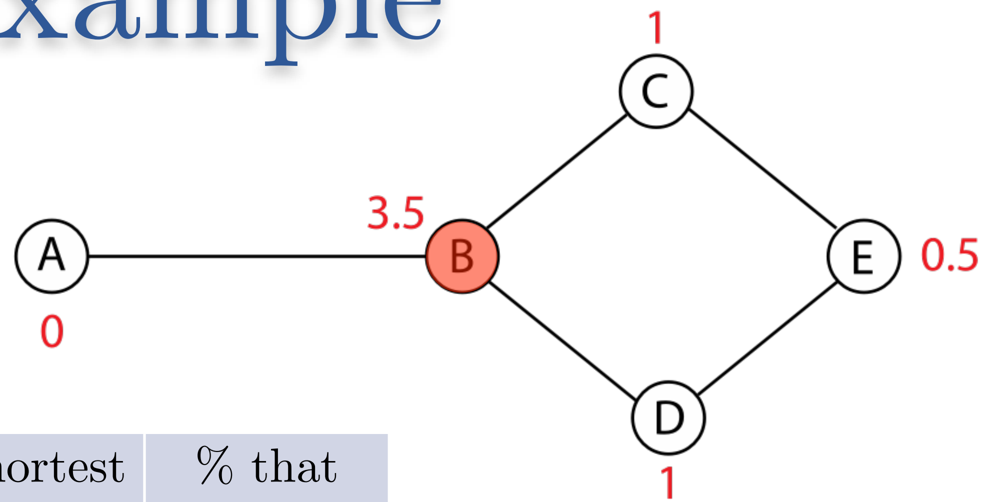


Source	Target	# of shortest paths	% that include B
A	C	1	1/1
A	D	1	1/1
A	E	2	2/2
C	D	2	1/2
C	E	1	0/1
D	E	1	0/1

Step 3: What proportion of those shortest paths contain the node of interest?

Betweenness Centrality Example

$$c_B(v) = \sum_{i \neq v \neq j} \frac{\delta_{ij}(v)}{\delta_{ij}}$$



Source	Target	# of shortest paths	% that include B
A	C	1	1/1
A	D	1	1/1
A	E	2	2/2
C	D	2	1/2
C	E	1	0/1
D	E	1	0/1
			3.5

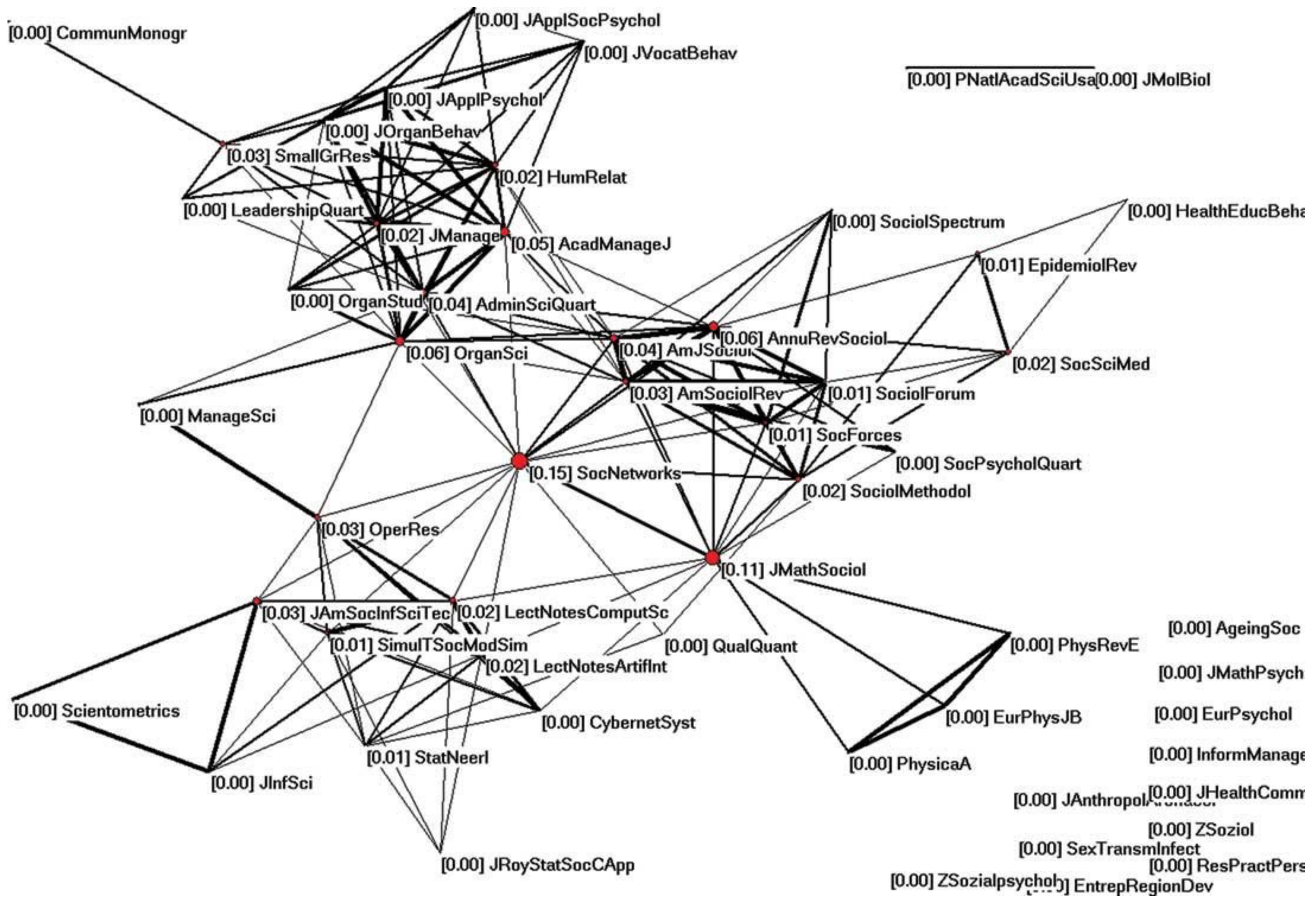
Step 4: Add.

Utility of Betweenness

- Identify potential **bottlenecks** in the network.
- **Bridge Betweenness**: Restrict pairs of nodes in previous table to a sources and targets from *different* communities.
- Teaching a robot how to learn new skills by first identifying which skill is most beneficial to master next.
-

Utility of Betweenness

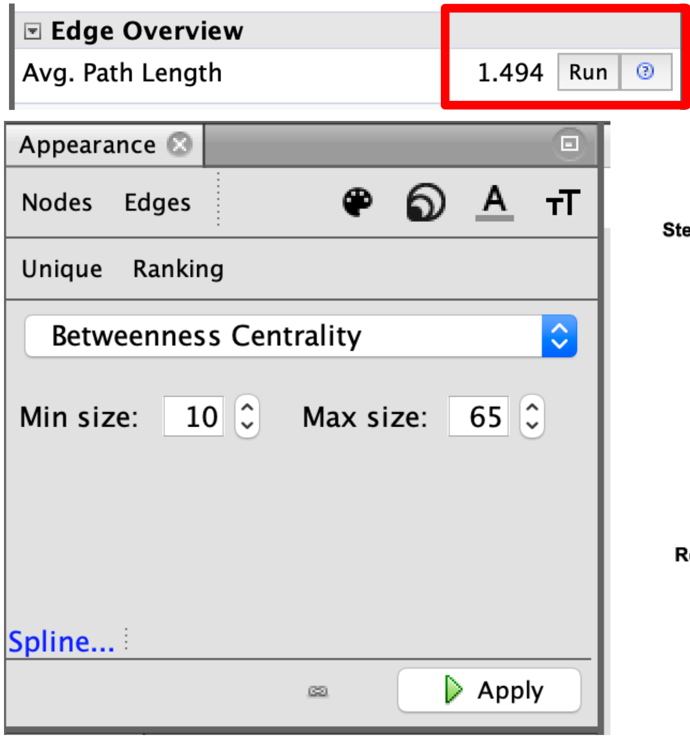
[Indicating the interdisciplinarity of scientific journals.]

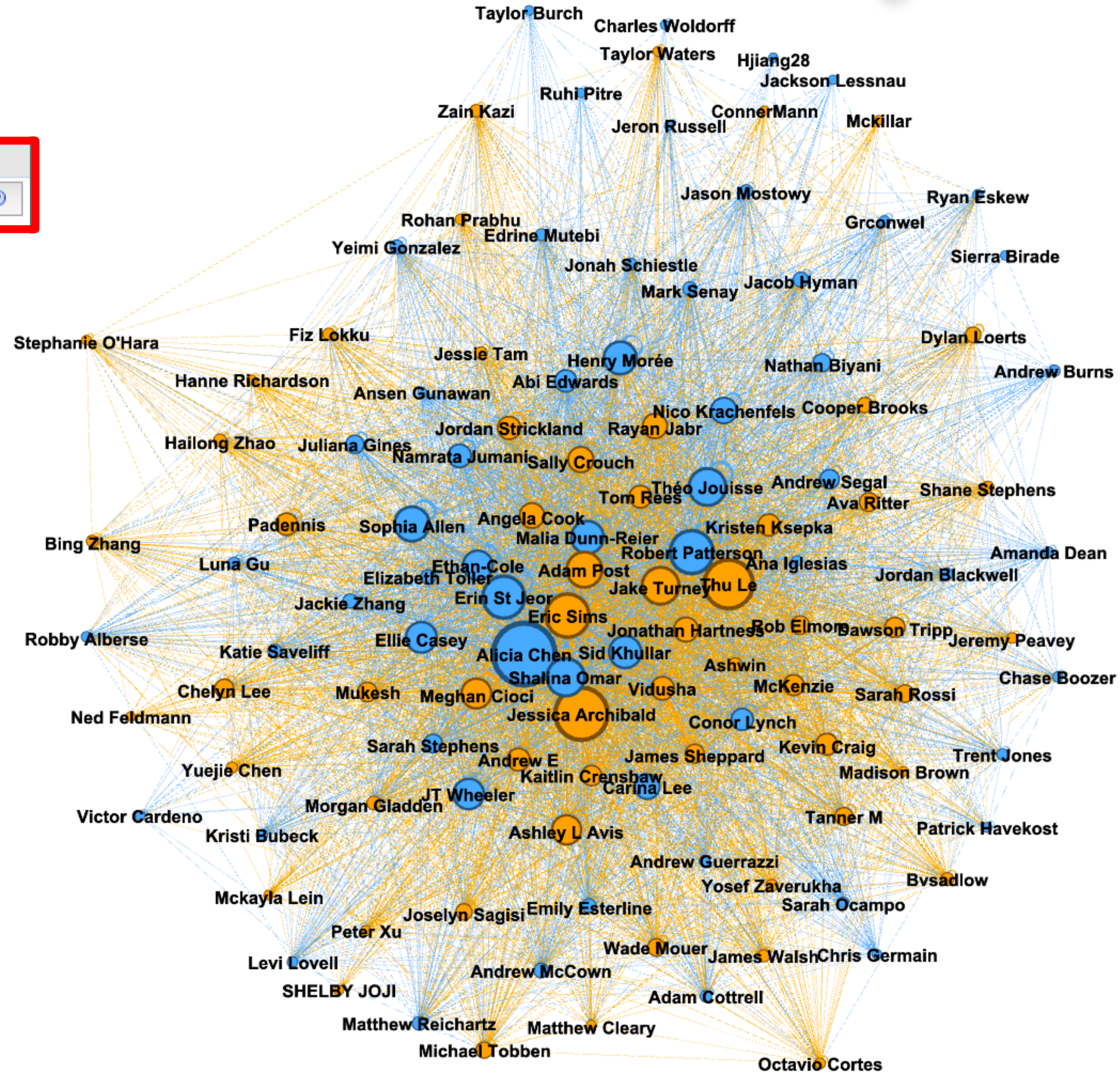


Betweenness Centrality Normalization

- Normalize each value by the maximum possible centrality score = the number of pairs of nodes excluding the given node.
  - For directed graph, normalize by (n-1)(n-2)
  - For undirected graph, normalize by (n-1)(n-2)/2
- Again, **normalization just allows for comparison across networks.**

Betweenness Centrality





Centrality

- **Centrality** is a measure of how importance a vertex is to a graph.
- There are many ways to define importance.
- Measures usually normalized in [0,1] for comparison across networks
- Common measures of centrality:
  - **Degree Centrality**
  - **Betweenness Centrality**
  - **Closeness Centrality**
  - **Eigenvector Centrality**

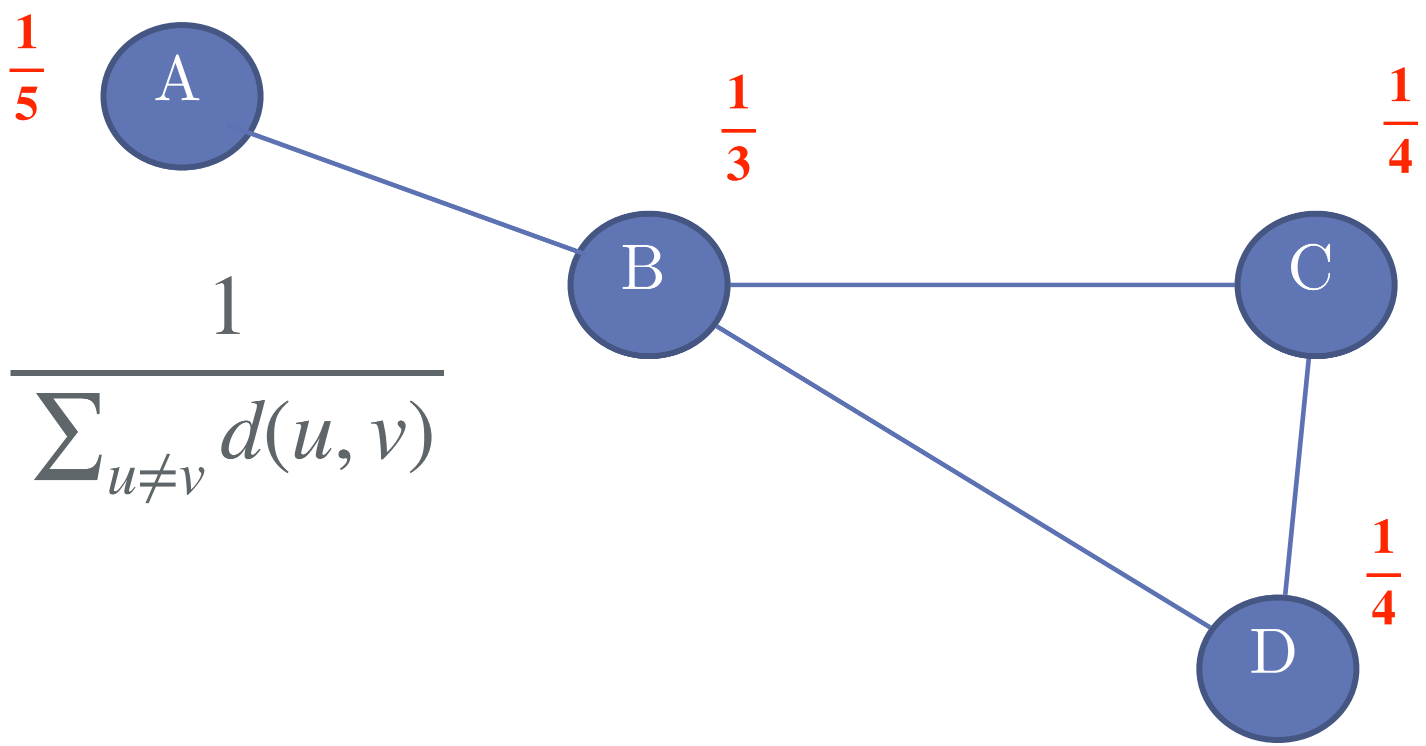
Closeness Centrality

- Measures how quickly information will spread from one node to *all* other nodes.

- Node is important if close to all other nodes.

- **Farness** defined as the sum of distances to all other nodes. Closeness is then the inverse of **farness**.

Closeness Centrality



$$c_c(v) = \frac{1}{\sum_{u \neq v} d(u, v)}$$

Utility of Closeness

- "In an auditory lexical decision task participants responded more quickly to words with high closeness centrality [in a phonological network]."
  —Goldstein, Vitevitch. U Kansas. 2017
- Identifying a group of nodes to optimally spread information.
- City planning
- "Too interconnected to fail" - examination of centrality of institutions in network of financial institutions. Closeness implies transmission of failure to many in a few steps.

Closeness Centrality

$$c_c(v) = \frac{1}{\sum_{u \neq v} d(u, v)}$$

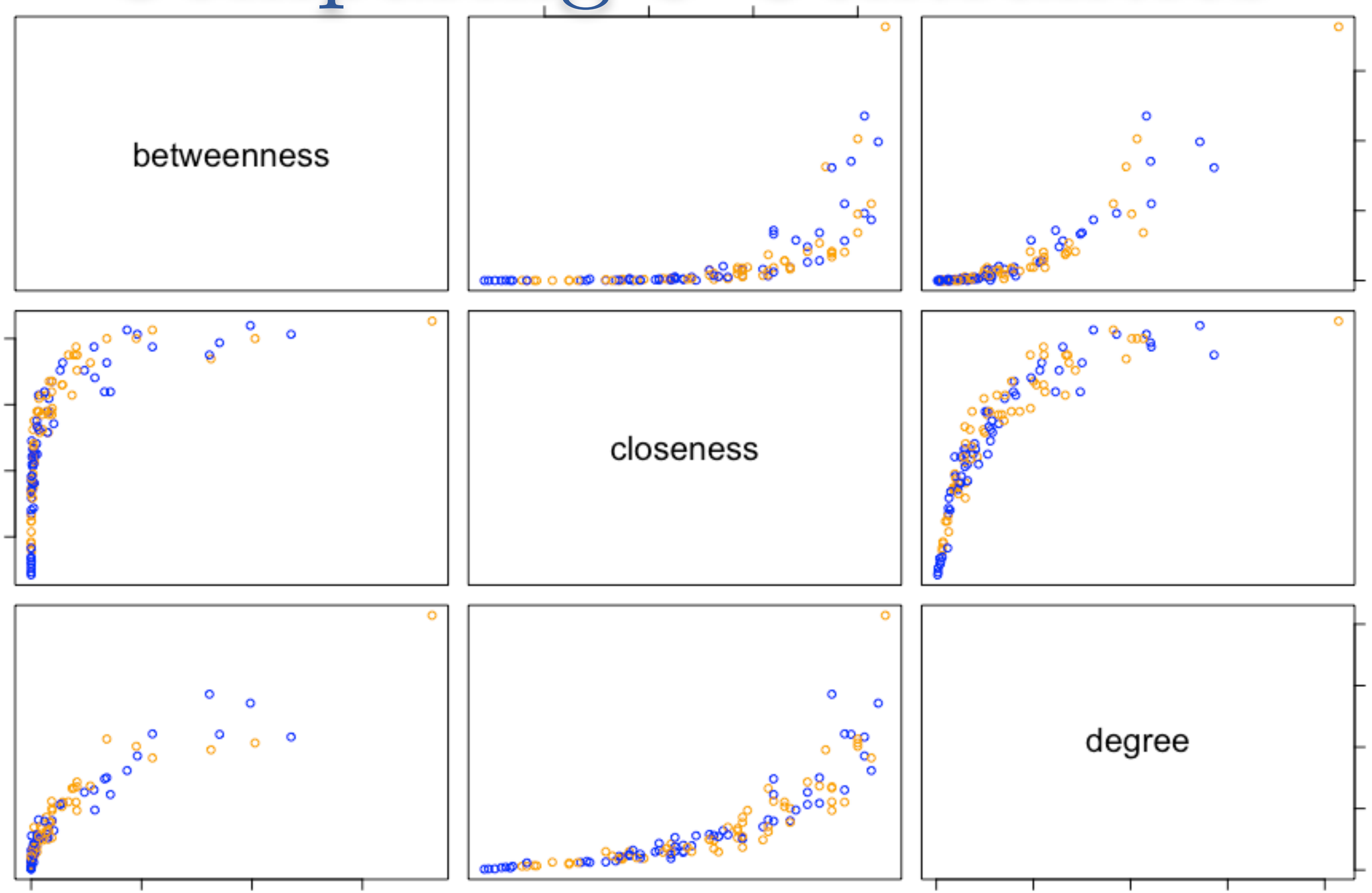
- Normalize by maximum possible score, which would be 1/(n-1) when the node is connected to all others.
  - i.e. multiply by (n-1) to normalize.
- **Only defined for connected graphs**

Comparing 3 Centralities

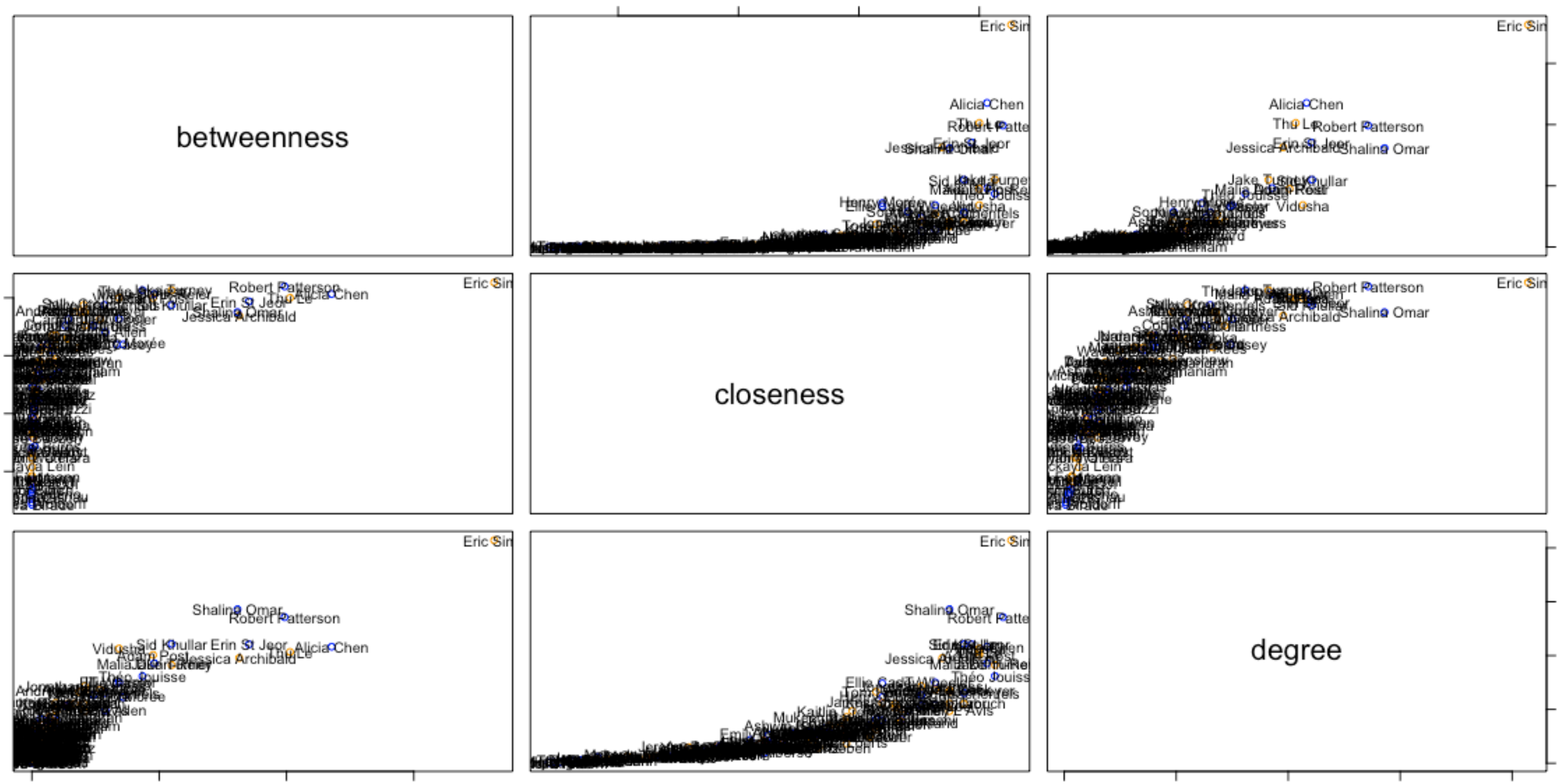
- The centrality measures should be correlated.
- If not, that *might* mean something interesting.

	Low Degree	Low Closeness	Low Betweenness
High Degree		In cluster far from rest of network	Connections are redundant. Communication bypasses node
High Closeness	Key player, tied to important, active alternatives		Prob. multiple paths everywhere in network. Near many, just like everyone else
High Betweenness	A few edges are crucial for network flow	RARE. Node monopolizes ties from few to many	

Comparing 3 Centralities



Comparing 3 Centralities



Code for Previous Slides

```
library(igraph)

c.b=betweenness(slack, v=V(slack), directed=T)
c.c=closeness(slack, v=V(slack), mode="total")
c.d=degree(slack, v=V(slack))

centralities = data.frame(betweenness = c.b,
                          closeness = c.c,
                          degree = c.d,
                          cohort = vertex_attr(slack, "Cohort"))
cor(centralities[,1:3])
pairs(centralities[,1:3],
      col = c('blue','orange')[as.factor(centralities$cohort)])
```

Centrality

- **Centrality** is a measure of how importance a vertex is to a graph.
- There are many ways to define importance.
- Measures usually normalized in [0,1] for comparison across networks
- Common measures of centrality:
  - **Degree Centrality**
  - **Betweenness Centrality**
  - **Closeness Centrality**
  - **Eigenvector Centrality**

Eigenvector Centrality

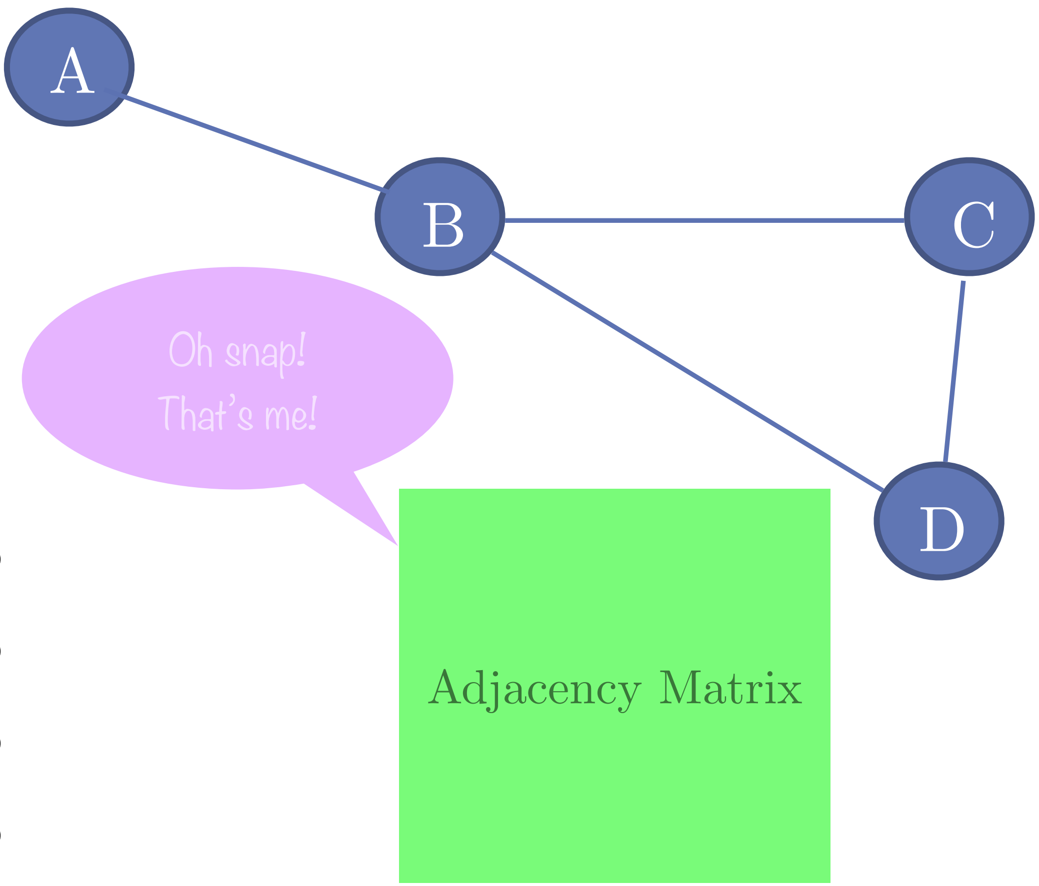
- Would you rather have influence over the Provost and the Chancellor or all of the university's graduate TAs?
- The previous definitions of centrality have a problem: They don't take into account the importance of your contacts.
- Previous measures don't necessarily measure *influence*.

Eigenvector Centrality

What if each node's centrality was the sum of the centralities of the nodes to which it is connected?

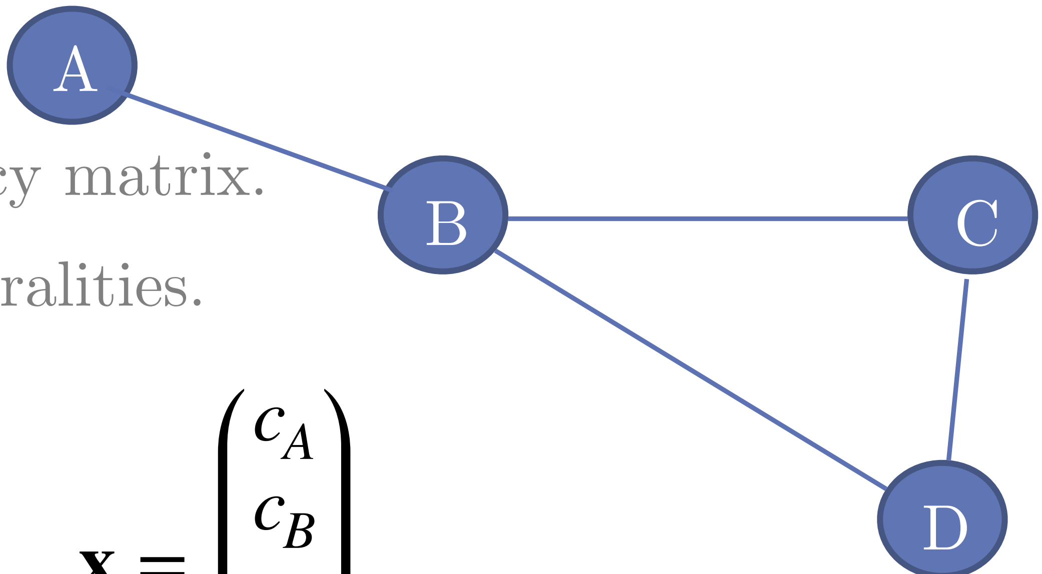
$$c_A = c_B$$
$$c_B = c_A + c_C + c_D$$
$$c_C = c_B + c_D$$
$$c_D = c_B + c_C$$

$$c_A = 0c_A + 1c_B + 0c_C + 0c_D$$
$$c_B = 1c_A + 0c_B + 1c_C + 0c_D$$
$$c_C = 0c_A + 1c_B + 0c_C + 1c_D$$
$$c_D = 0c_A + 1c_B + 1c_C + 0c_D$$



Eigenvector Centrality

What if each node's centrality was the sum of the centralities of the nodes to which it is connected?



- Let **A** be a binary adjacency matrix.
- Let **x** be the vector of centralities.

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \qquad \mathbf{x} = \begin{pmatrix} c_A \\ c_B \\ c_C \\ c_D \end{pmatrix}$$

Then what we want is: $\mathbf{A}\mathbf{x} = \mathbf{x}$

Eigenvector Centrality

$$\mathbf{Ax} = \mathbf{x}$$

- Of course, this may not always be possible, so we'll add in a constant — each node's centrality *is proportional to* the sum of its neighbors.

$$\mathbf{Ax} = \lambda\mathbf{x}$$

- Look familiar?? The ever-useful eigenvector equation

Eigenvector Centrality

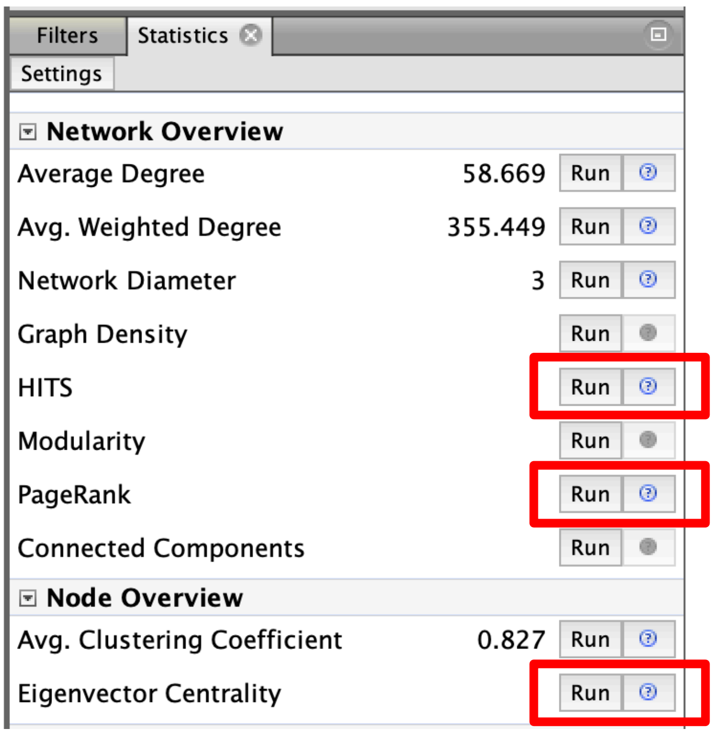
- But *which* eigenvalue??
- We want the centralities to be positive.
- The **first** eigenvector of an adjacency matrix is guaranteed to have all positive elements.
  - (Perron-Frobenius Theorem of Nonnegative Matrices)

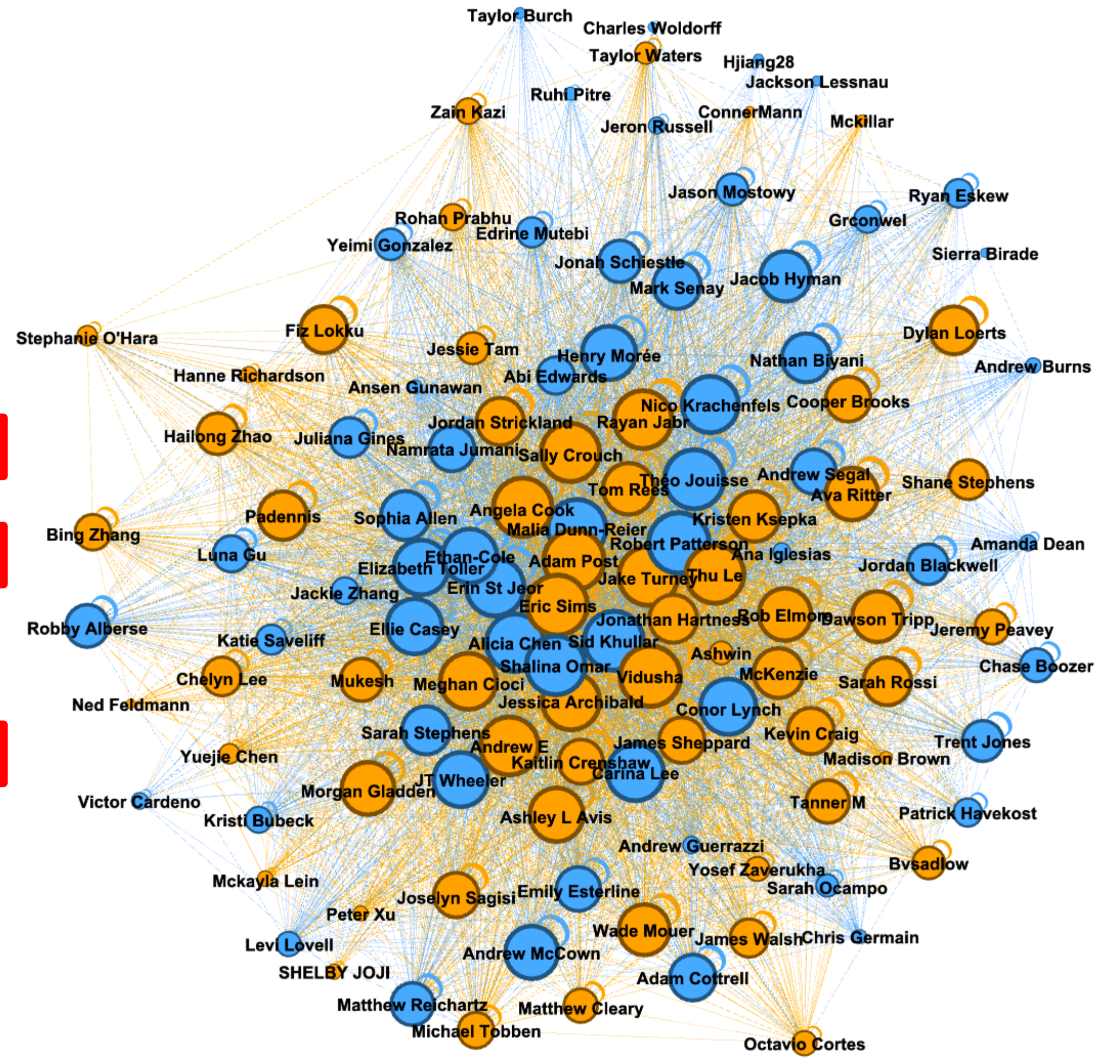
- Google's PageRank is a variant of eigenvector centrality, using the hyperlink structure of the net.
  - See original HITS algorithm

Utility of Eigenvector Centrality

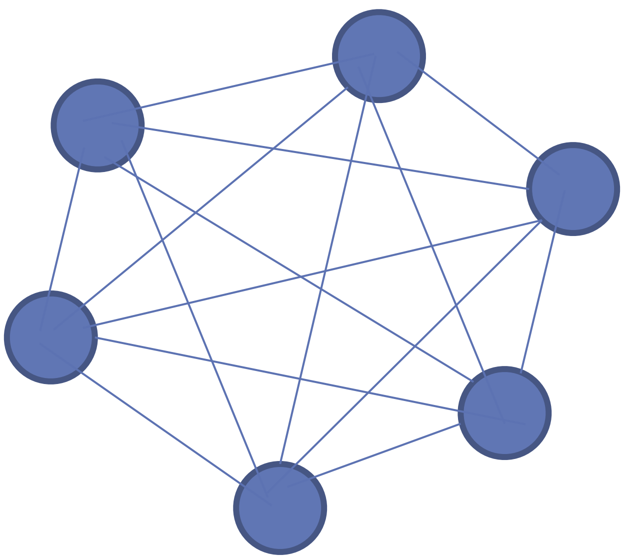
- It's the **most** useful measure because it contains a notion of influence of neighbors.
- Google
- Examining Volleyball game flow
- Cooperative Streaming - timely and efficient distribution of content in a communication network
- Feature Selection!

Eigenvector Centrality

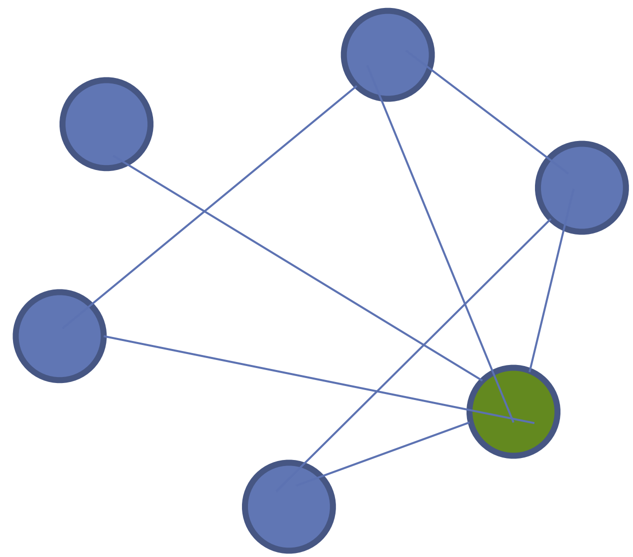




When is Centrality Interesting?



vs.



Centralization of Network

- How much variation in centrality scores?
- How central is maximum centrality compared to others?

$$0 < \text{Centralization} < 1$$

All nodes have the same centrality

One node has maximal centrality, all other nodes have minimal centrality

Centralization of Network

- Freeman's formula for **degree centralization**:

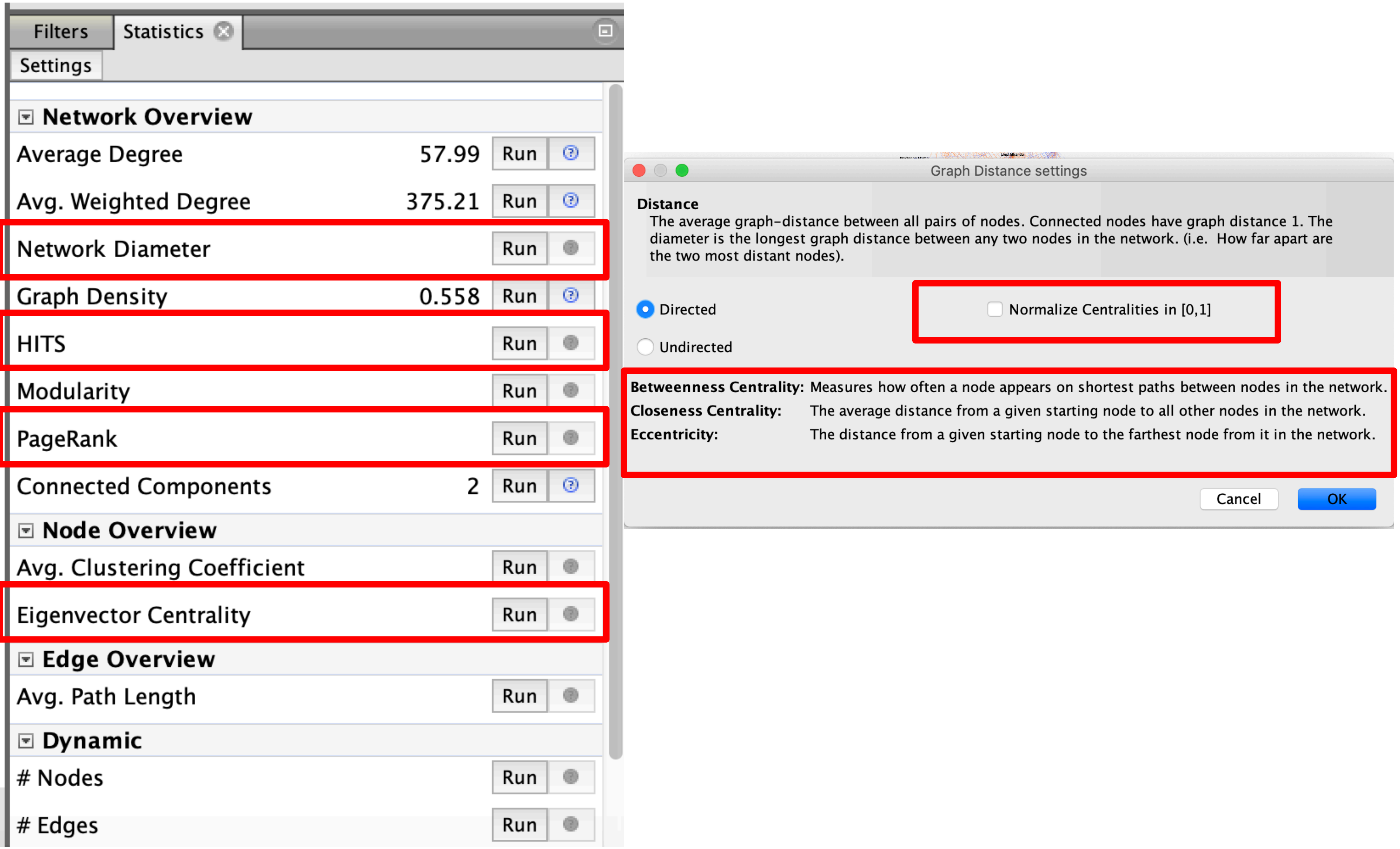
$$C_D = \frac{\sum_{i=1}^{n} \left(max(c_D) - c_D(i) \right)}{(n-1)(n-2)}$$

- Can be adopted for other types of centrality just by changing denominator and centrality function.

Calculating Centrality Scores in Gephi

• • •

Try: Filtering + Rerunning measures.



Filters
Statistics
Settings
Network Overview
Average Degree 57.99 Run
Avg. Weighted Degree 375.21 Run
Network Diameter Run
Graph Density 0.558 Run
HITS Run
Modularity Run
PageRank Run
Connected Components 2 Run
Node Overview
Avg. Clustering Coefficient Run
Eigenvector Centrality Run
Edge Overview
Avg. Path Length Run
Dynamic
Nodes Run
Edges Run
Graph Distance settings
Distance
The average graph-distance between all pairs of nodes. Connected nodes have graph distance 1. The diameter is the longest graph distance between any two nodes in the network. (i.e. How far apart are the two most distant nodes).
Directed
Undirected
Normalize Centralities in [0,1]
Betweenness Centrality: Measures how often a node appears on shortest paths between nodes in the network.
Closeness Centrality: The average distance from a given starting node to all other nodes in the network.
Eccentricity: The distance from a given starting node to the farthest node from it in the network.
Cancel
OK