Calculating Centralities and Detecting Communities with NetworkX

Artur Krochin
DATA SCIENTIST

linkedin.com/in/arturkrochin/

Overview

Why centrality measures?

Degree centrality

Closeness centrality

Betweenness centrality

Katz, Eigenvector, and PageRank centralities

Girvan-Newman community detection

Why Centrality Measures?

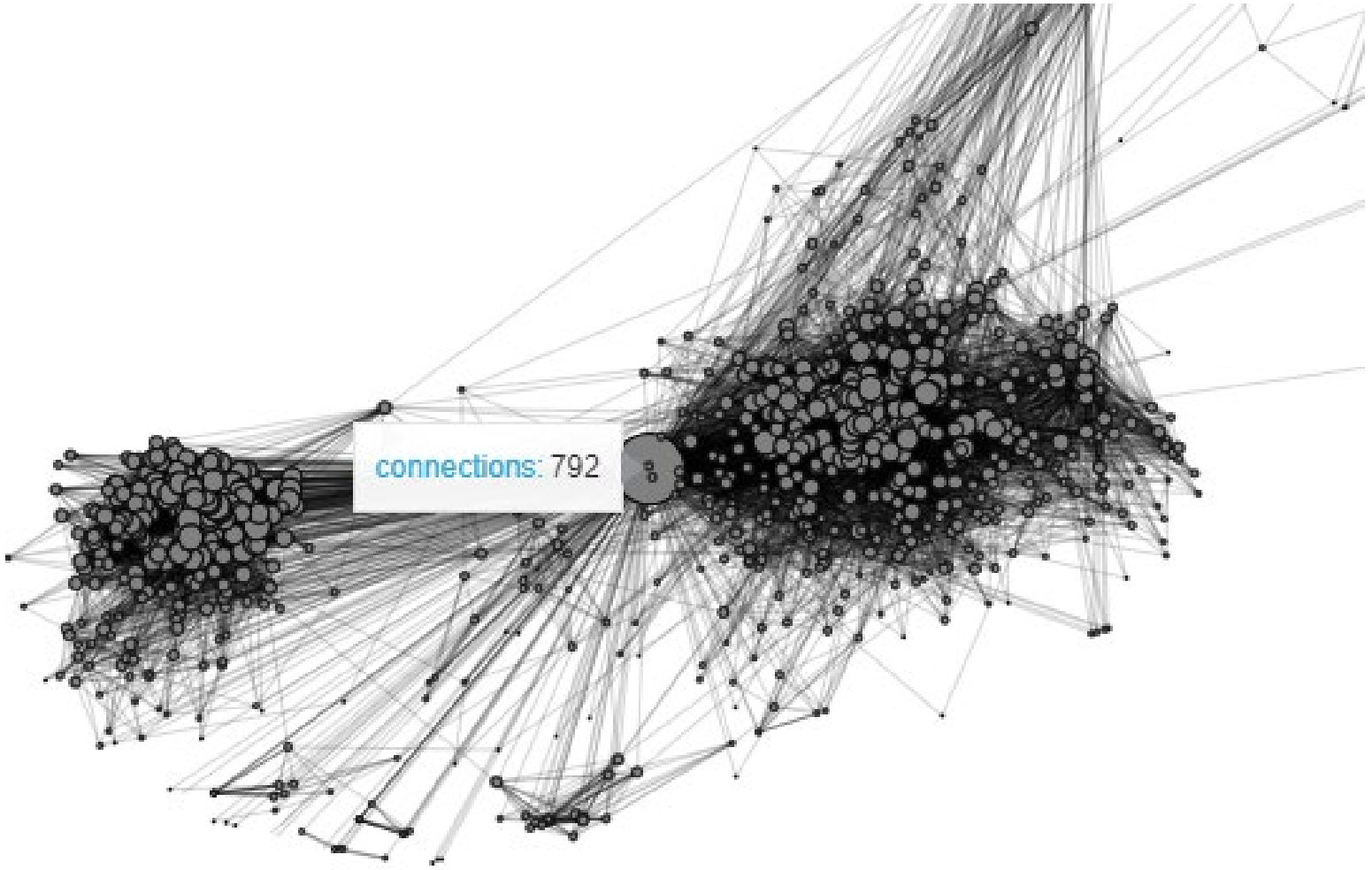
Centrality Use Cases

Influencer identification in social networks

Bottleneck identification in transportation networks

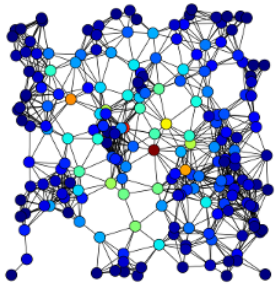
Pandemic prevention in virus spread networks

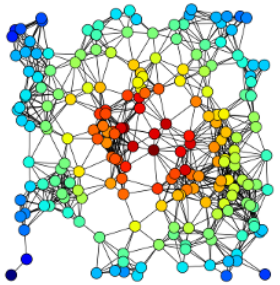
Degree Centrality



Different Centrality Measures

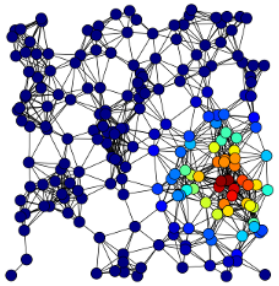
Betweenness

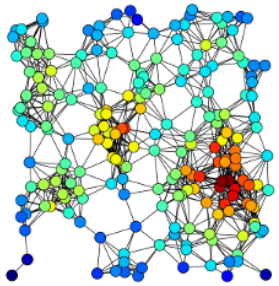




Closeness

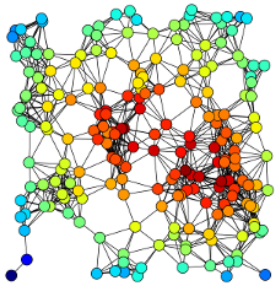
Eigenvector

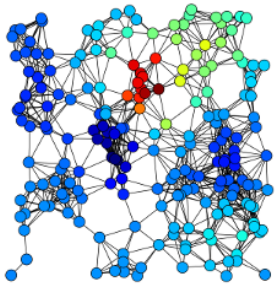




Degree

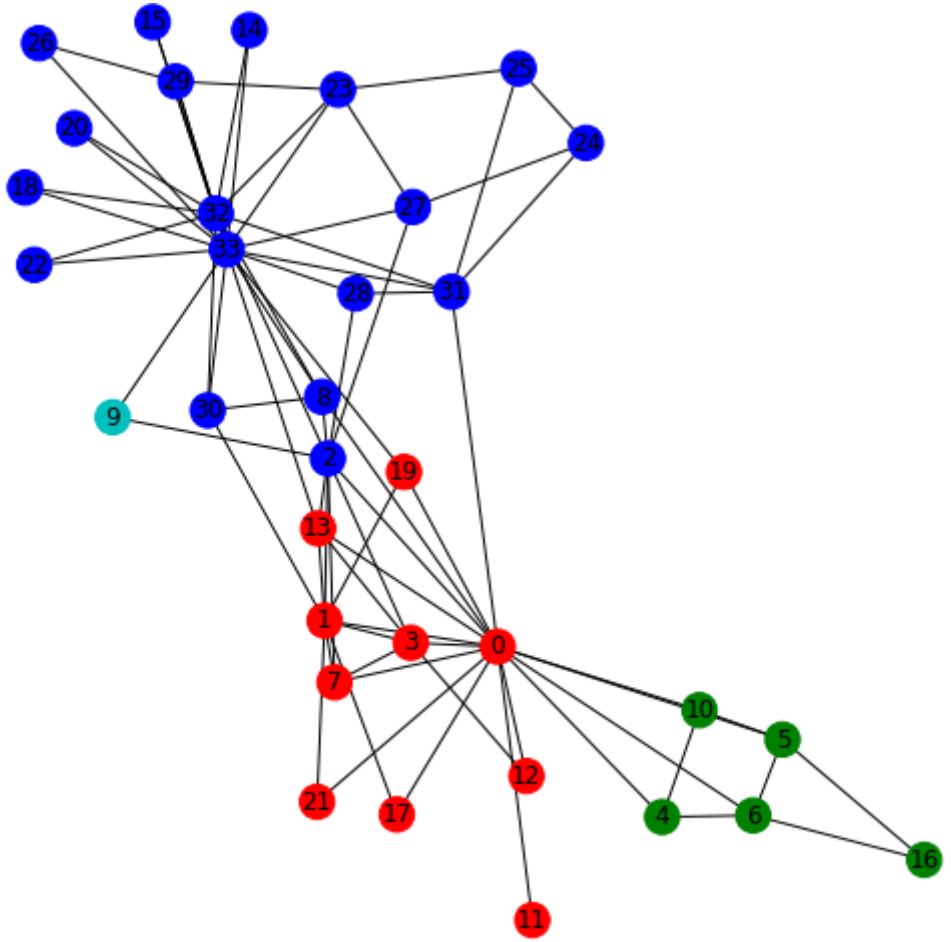
Harmonic





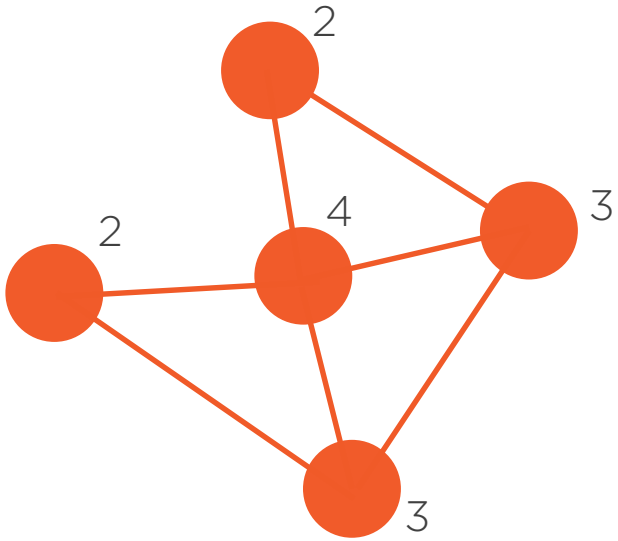
Katz

Girvan-Newman Algorithm

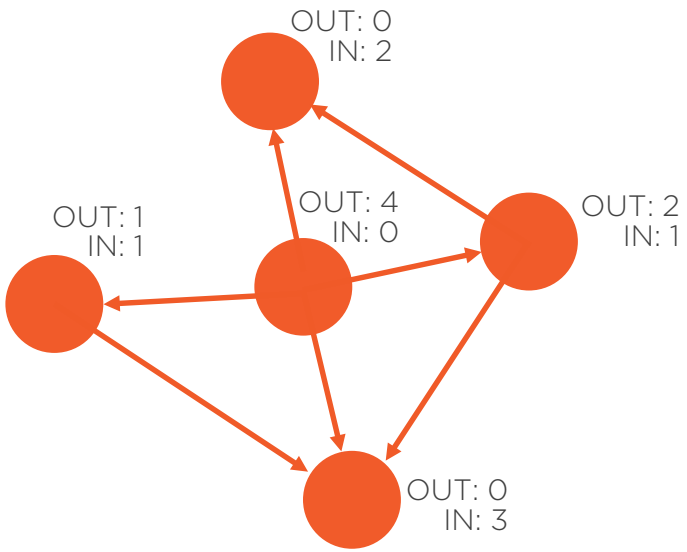


Degree Centrality

Undirected Degree



Directed Degree



Sources and Sinks

$$\Delta = \frac{D_{in}}{D_{out}}, D_{out} \neq 0$$

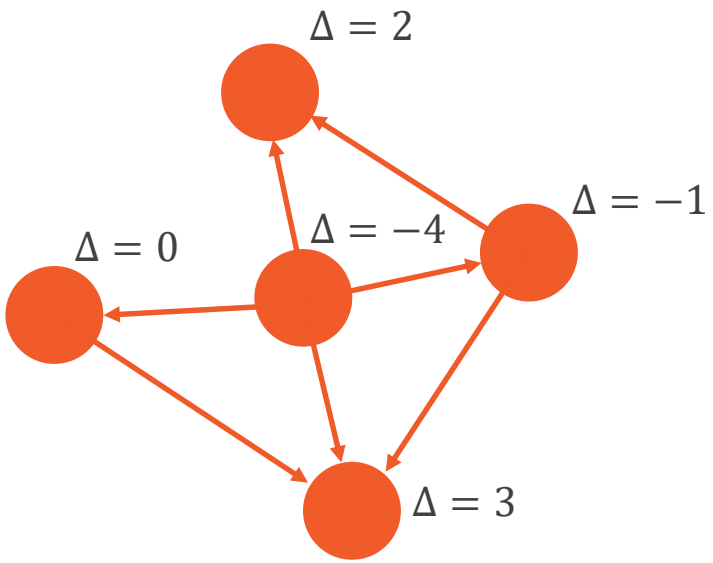
$$\Delta = D_{in} - D_{out}$$

Sources and Sinks

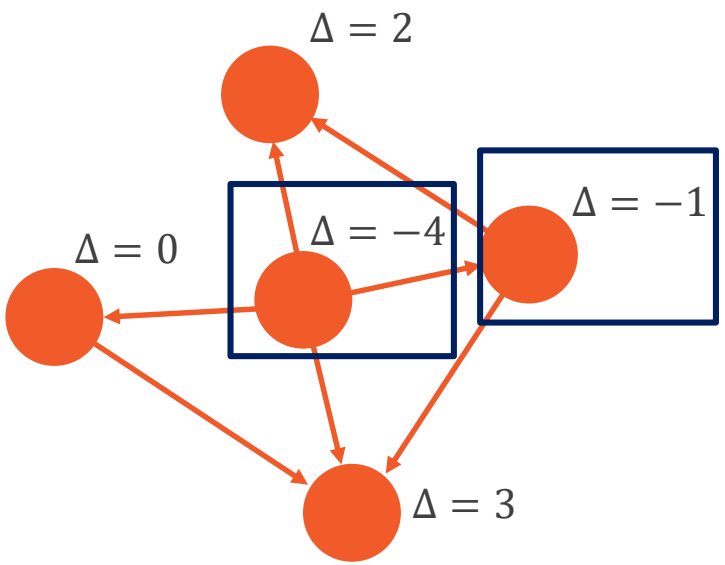
$$\Delta = \frac{D_{in}}{D_{out}}, D_{out} \neq 0$$

$$\boxed{\Delta = D_{in} - D_{out}}$$

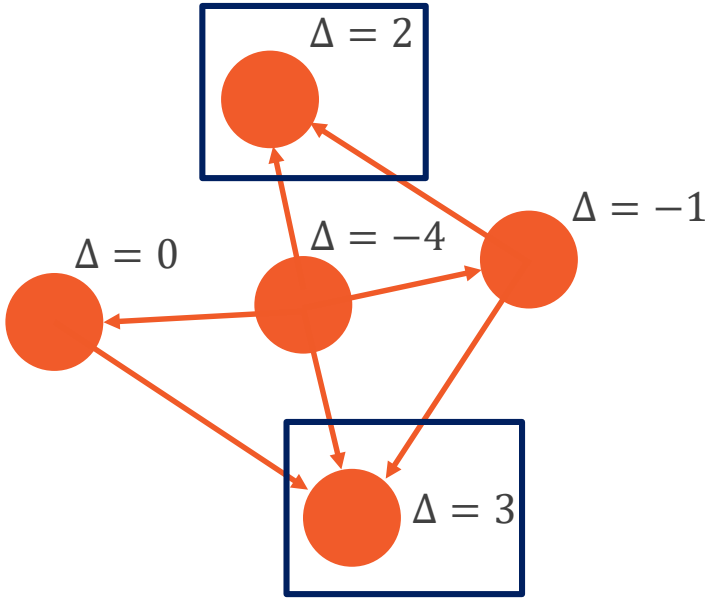
Sources and Sinks



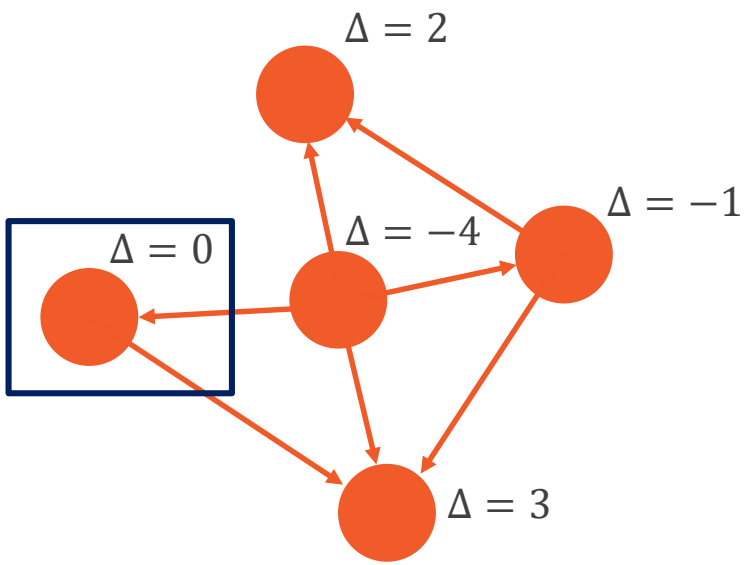
Sources and Sinks



Sources and Sinks



Sources and Sinks



```
nx.degree_centrality(G)

nx.in_degree_centrality(G)

nx.out_degree_centrality(G)
```

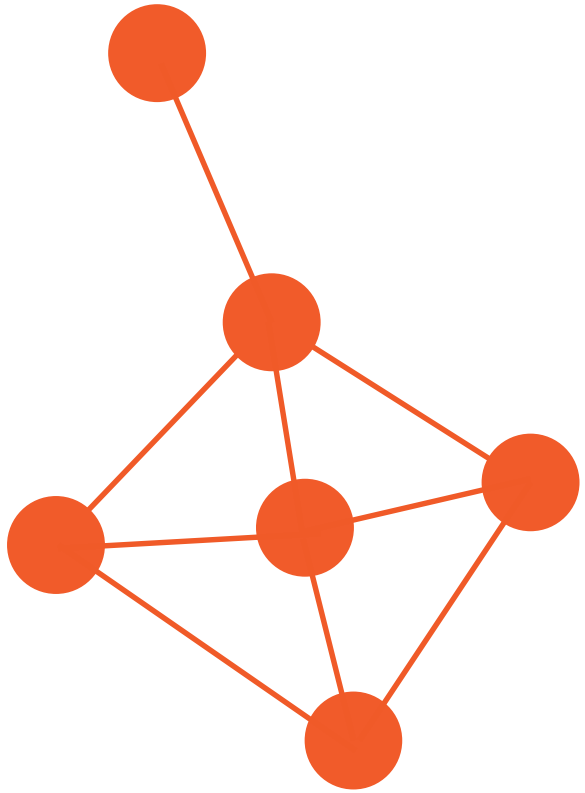
Degree Centrality in NetworkX

Closeness Centrality

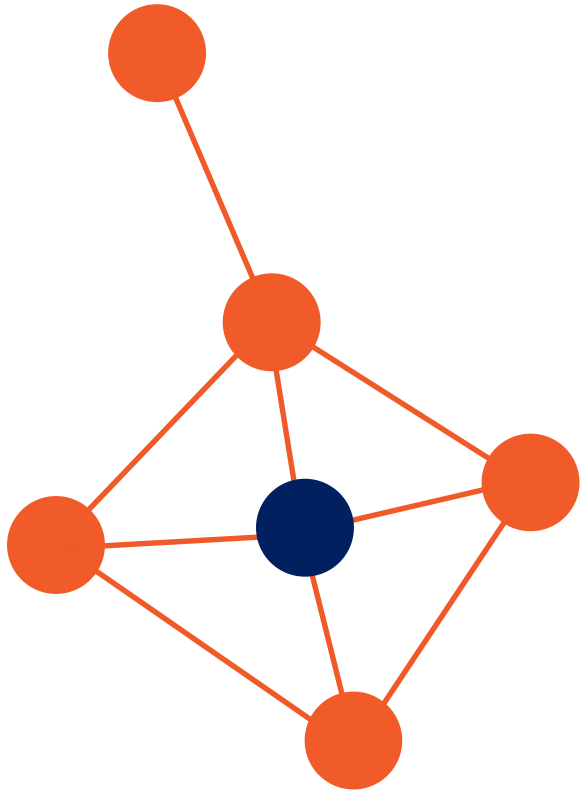
Closeness Centrality

Measures how close a node is, on average, to all the other nodes in the network.

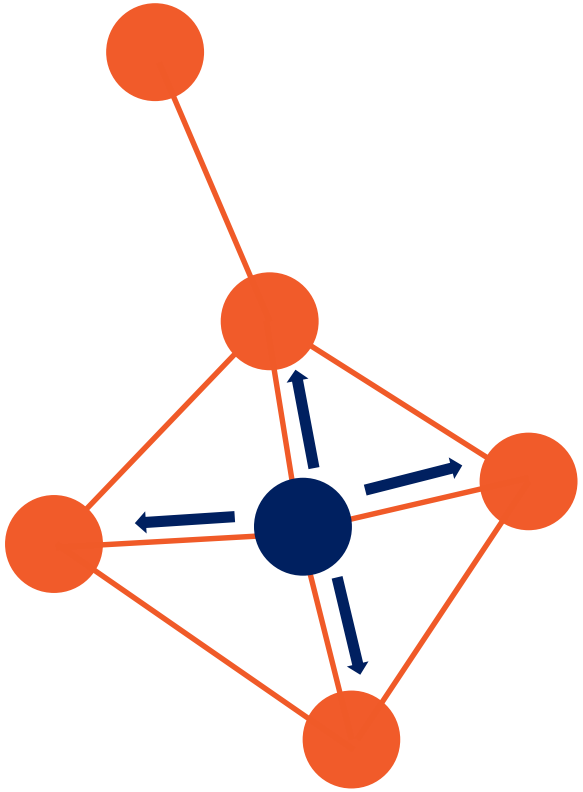
Closeness Example



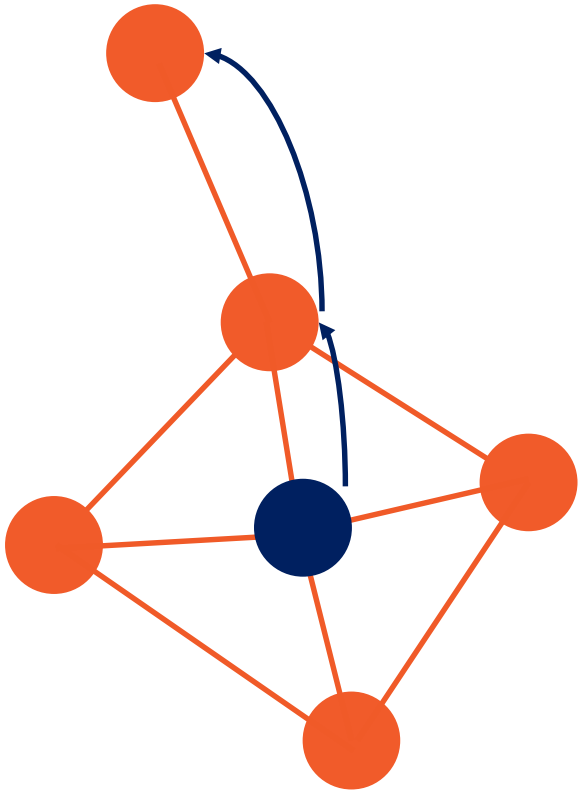
Closeness Example



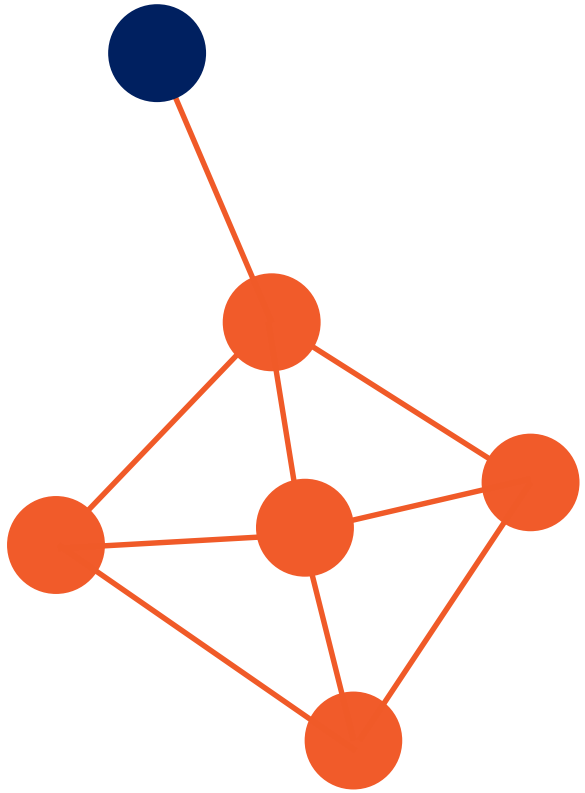
Closeness Example



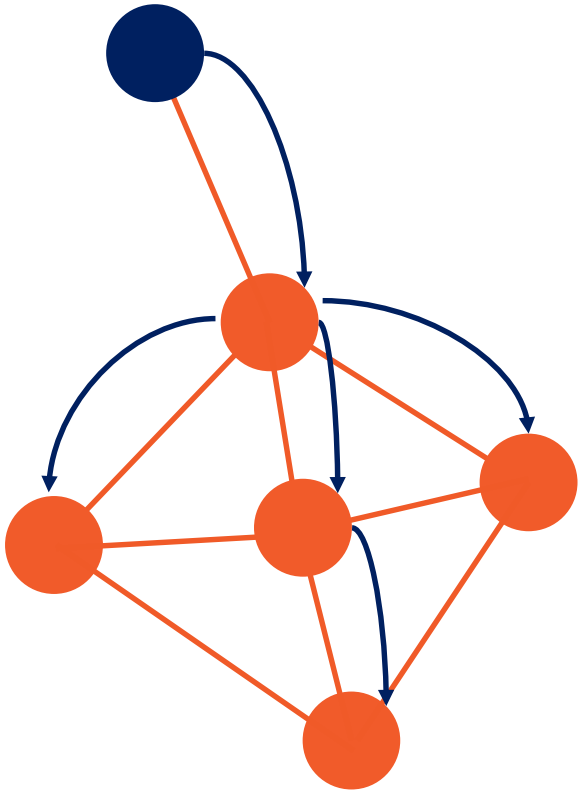
Closeness Example



Closeness Example



Closeness Example



Closeness Centrality Formulation

$$d(u, v) := \text{minimum number of hops}$$
$$\text{or sum of weights from node u to node v}$$

Closeness Centrality Formulation

$d(u, v) :=$ minimum number of hops
or sum of weights from node u to node v

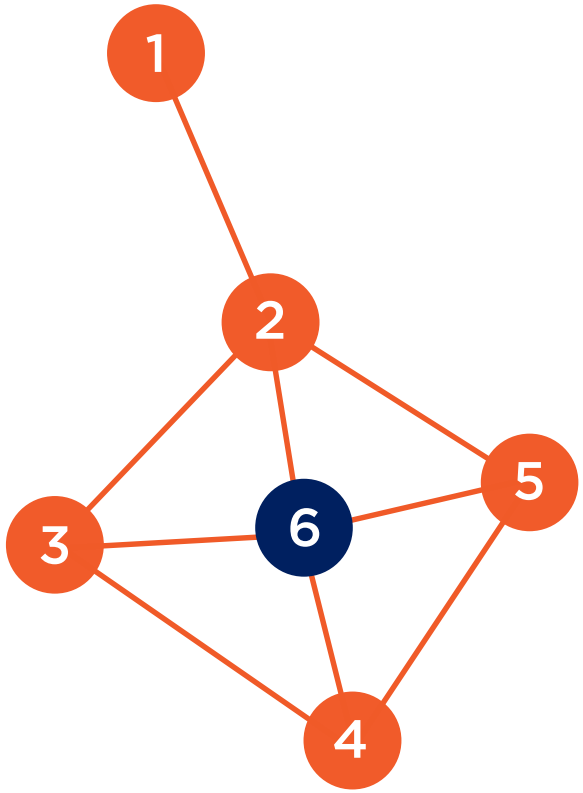
$$C'(u) = \frac{\sum_{v=1}^{n-1} d(u, v)}{n - 1}$$

Closeness Centrality Formulation

$d(u, v) :=$ minimum number of hops
or sum of weights from node u to node v

$$C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(u, v)}$$

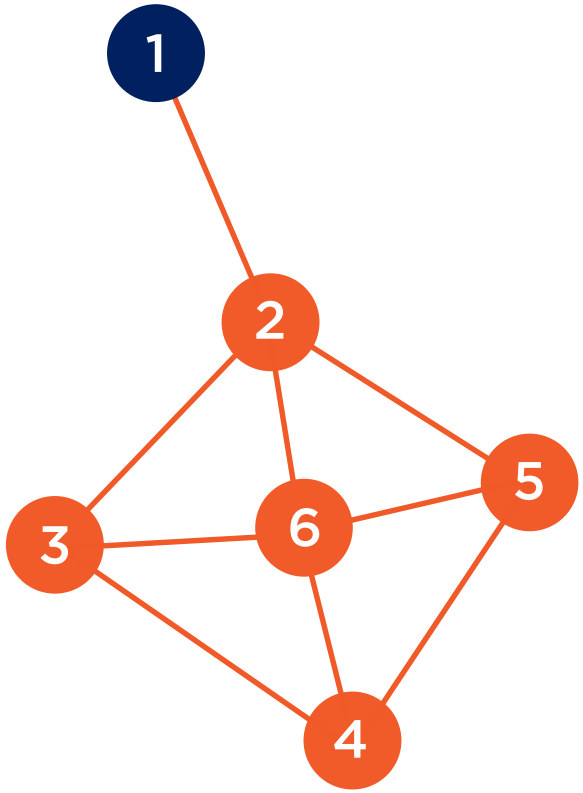
High Closeness



Path	Min. Distance
6 to 2	1
6 to 3	1
6 to 4	1
6 to 5	1
6 to 1	2

$$C(6) = \frac{6-1}{1+1+1+1+2} = 0.83$$

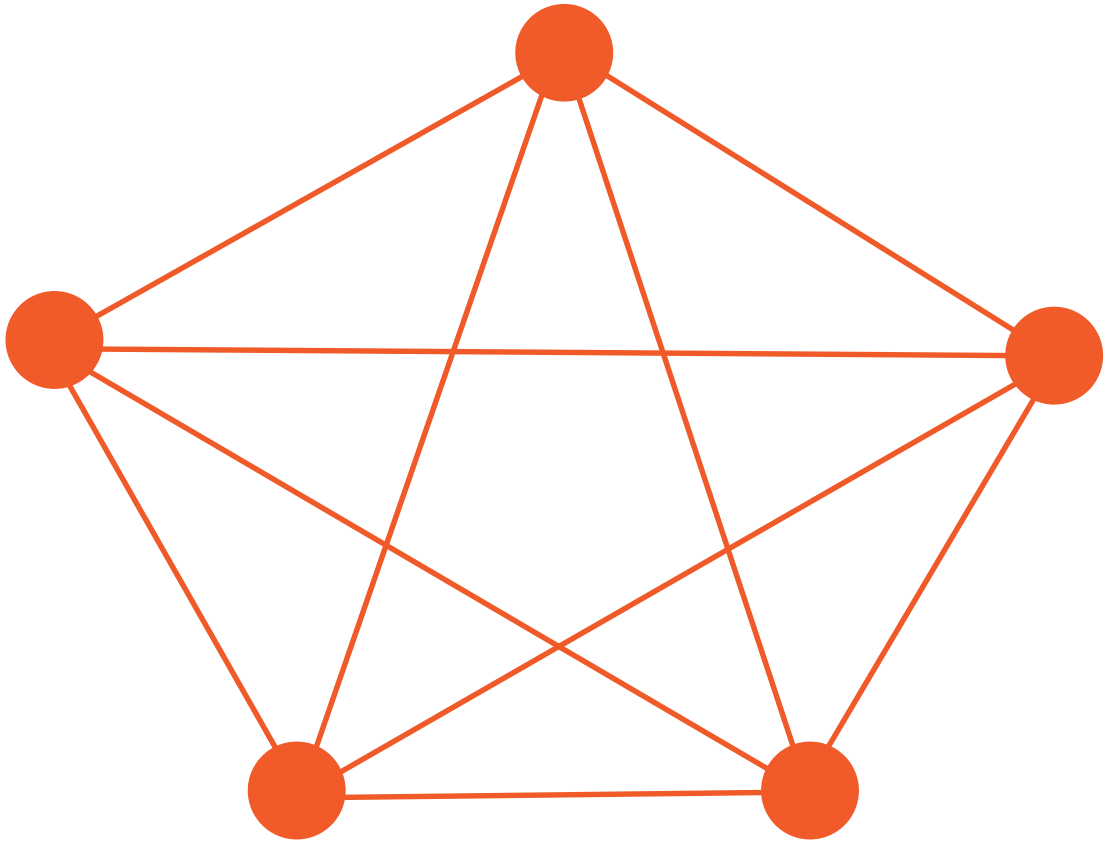
Low Closeness



Path	Min. Distance
1 to 2	1
1 to 3	2
1 to 5	2
1 to 6	2
1 to 4	3

$$C(1) = \frac{6-1}{1+2+2+2+3} = 0.5$$

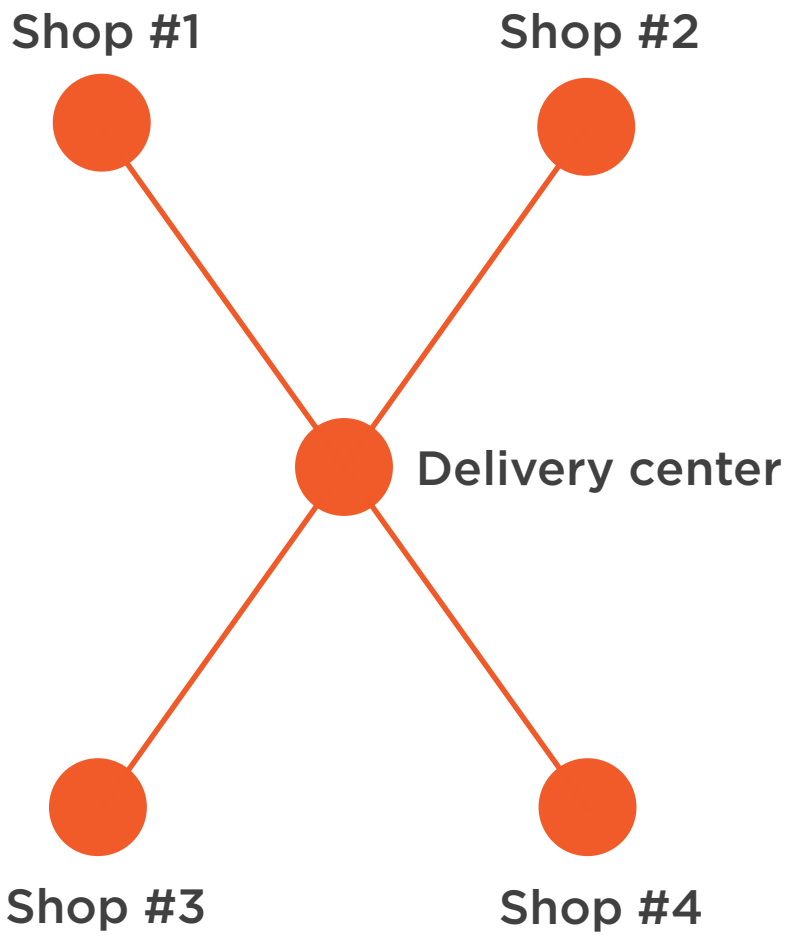
Closeness Caveat



K_5 graph

$$C(Any\ node) = \frac{5-1}{1+1+1+1} = 1$$

Delivery Center Placement



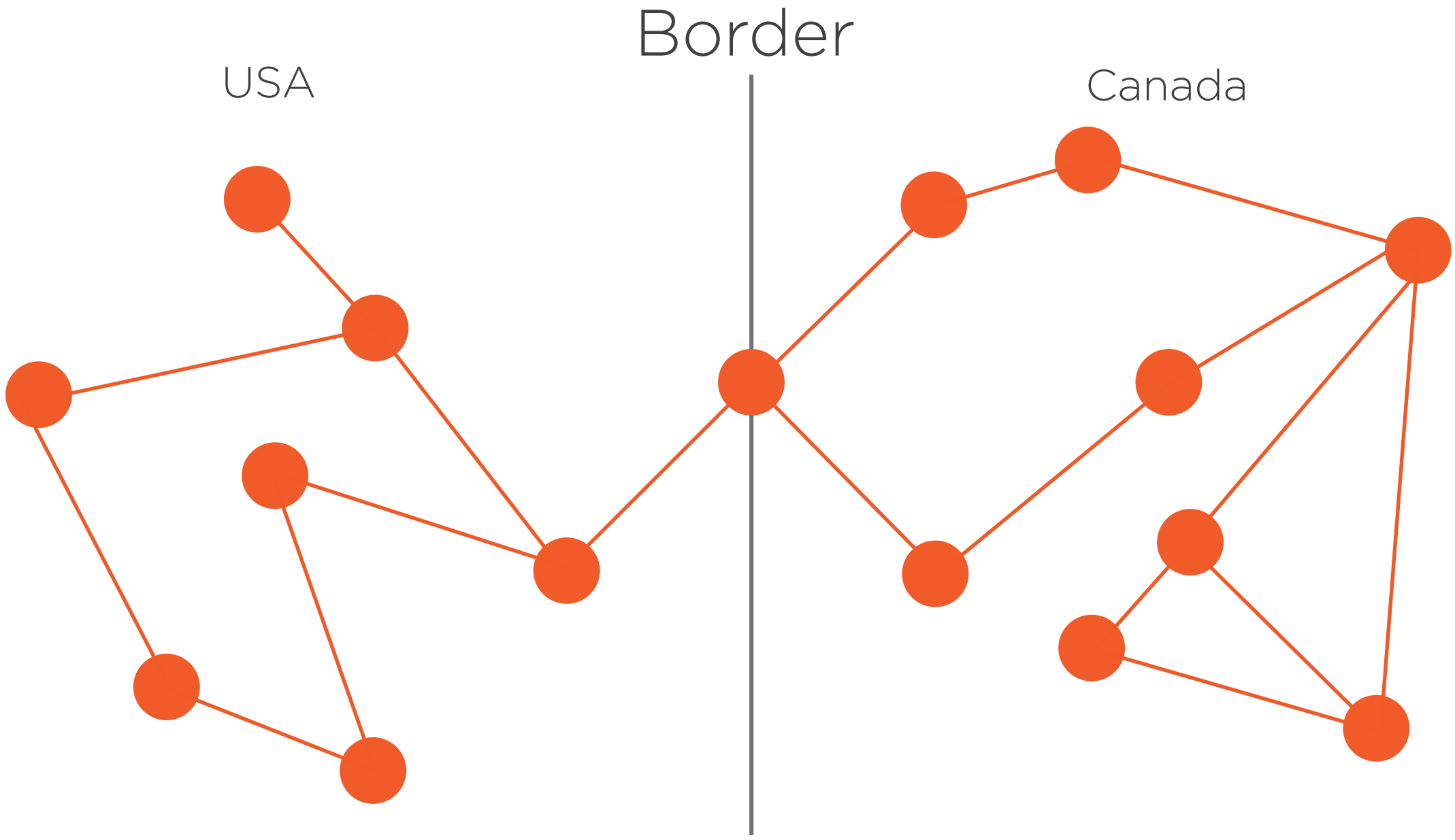
```
nx.closeness_centrality(G)
```

Closeness Centrality in NetworkX

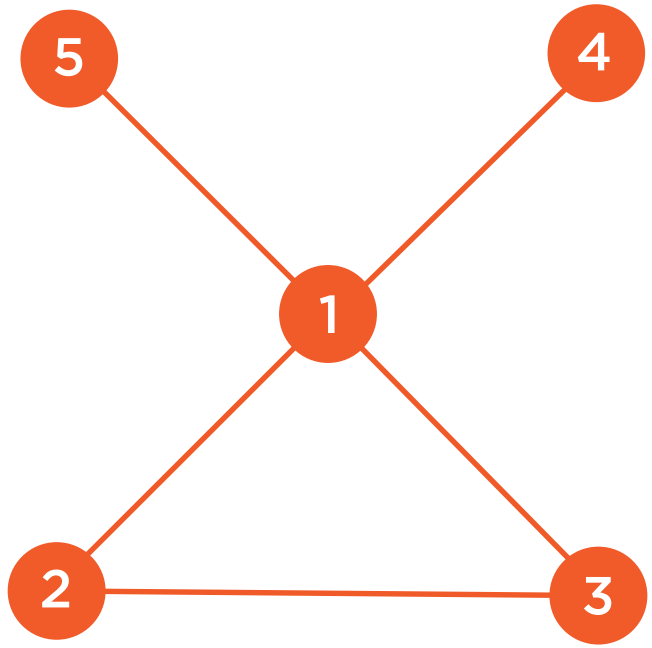
Betweenness Centrality

Betweenness Centrality

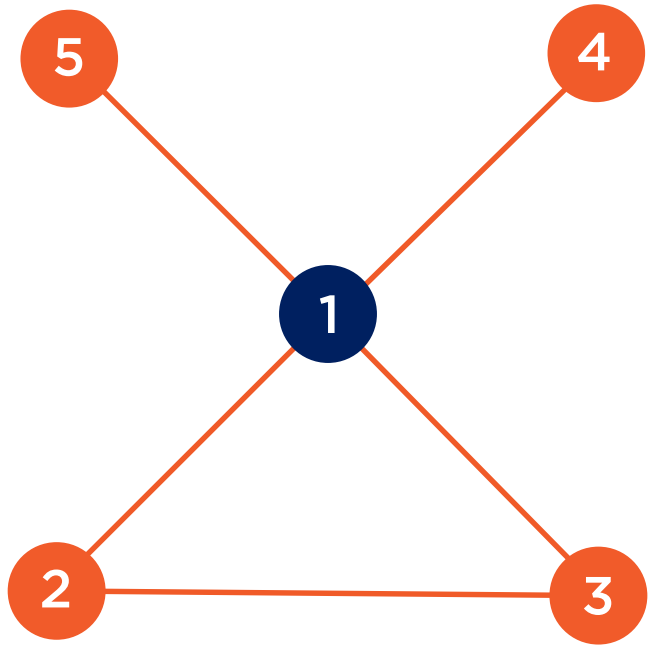
Measures the extent to which a node serves as a bridge or interchange hub to other nodes.


Border
USA
Canada

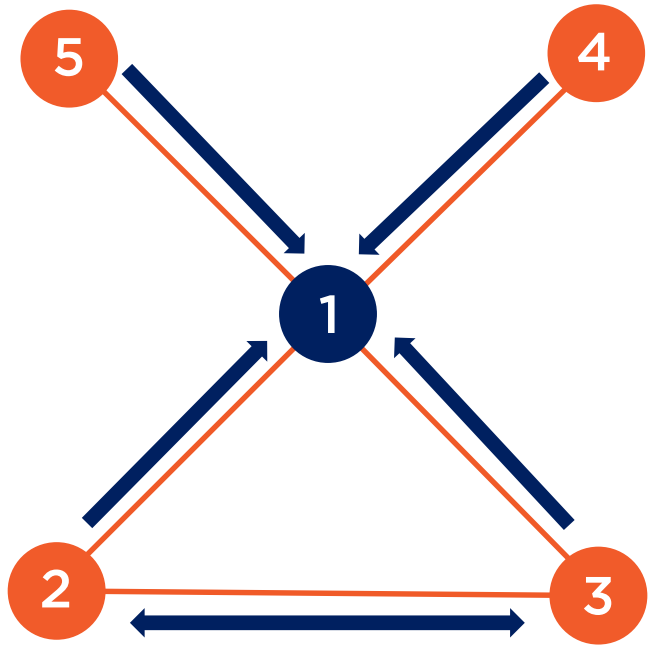
Betweenness Example



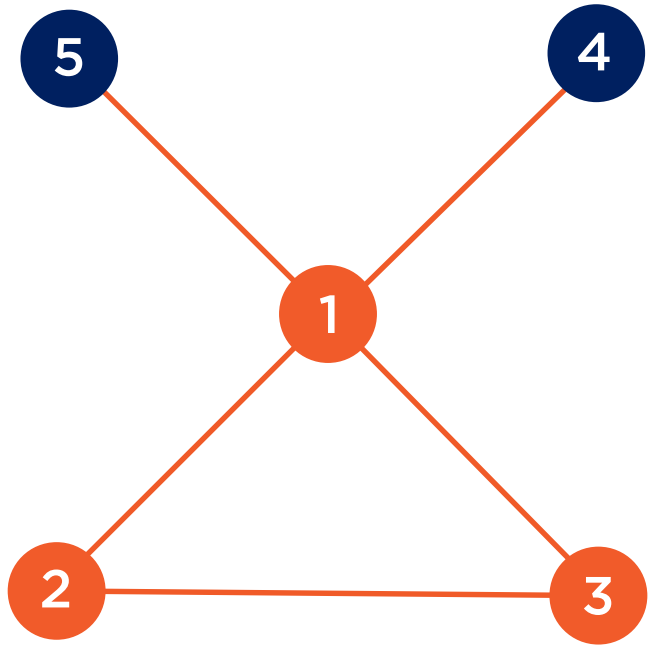
Betweenness Example



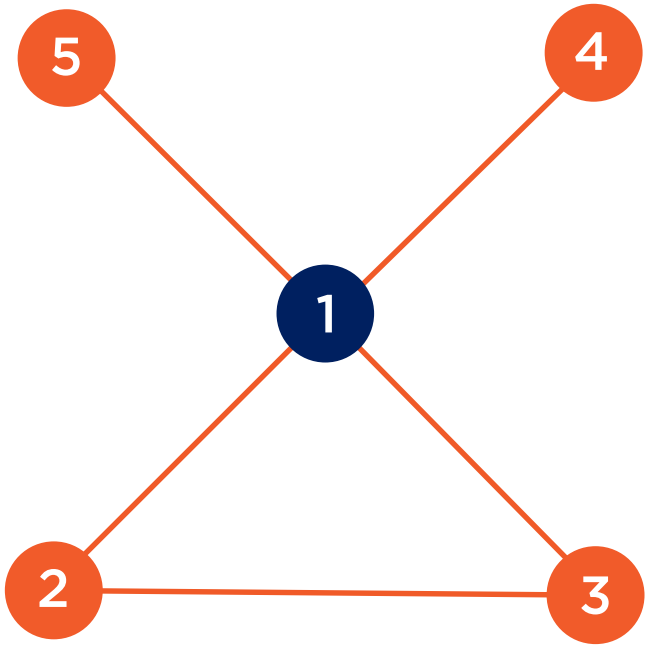
Betweenness Example



Betweenness Example

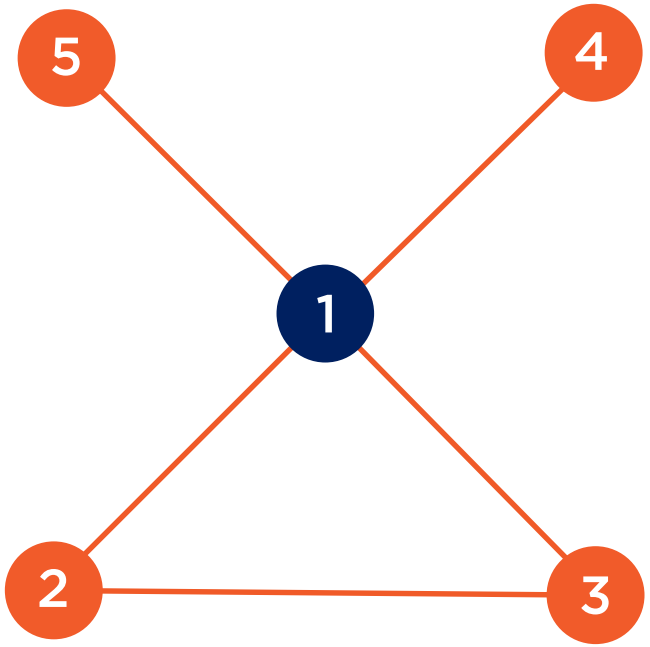


Betweenness Example

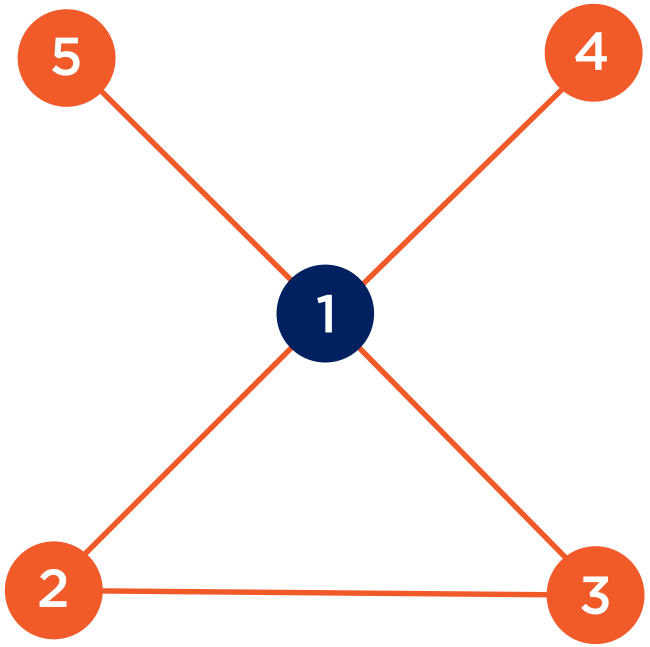


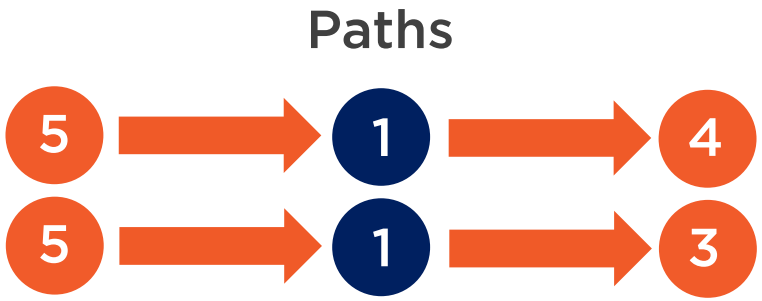
Paths

Betweenness Example

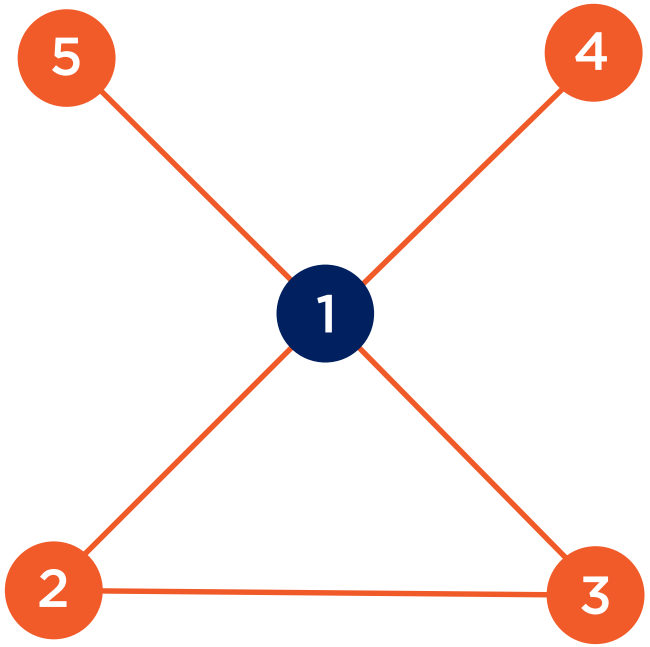


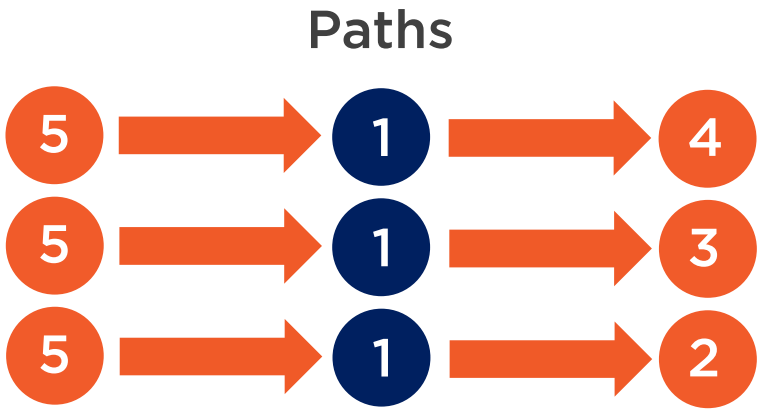
Betweenness Example



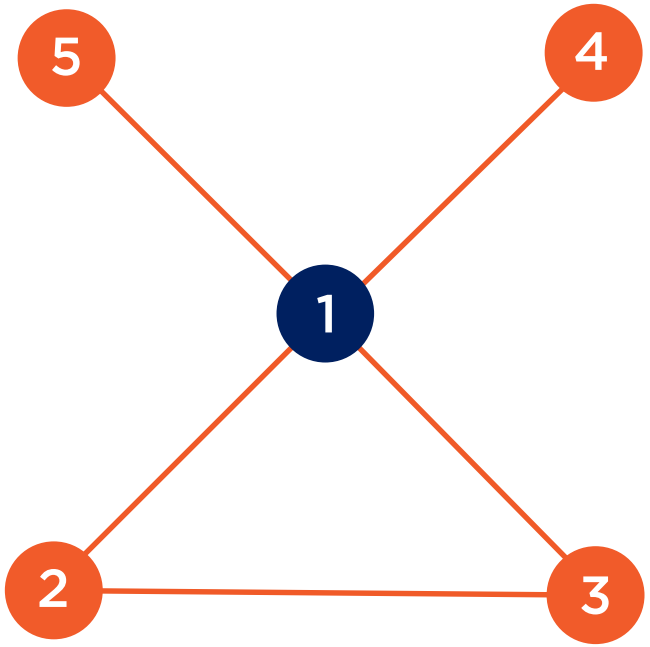


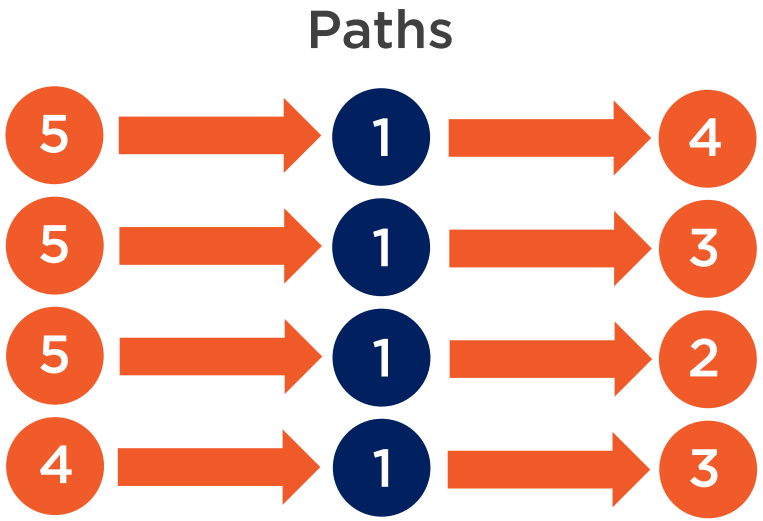
Betweenness Example



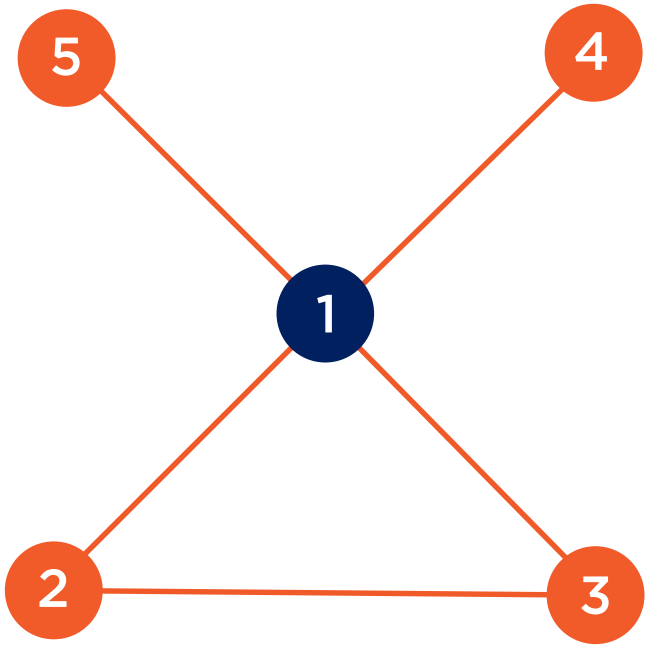


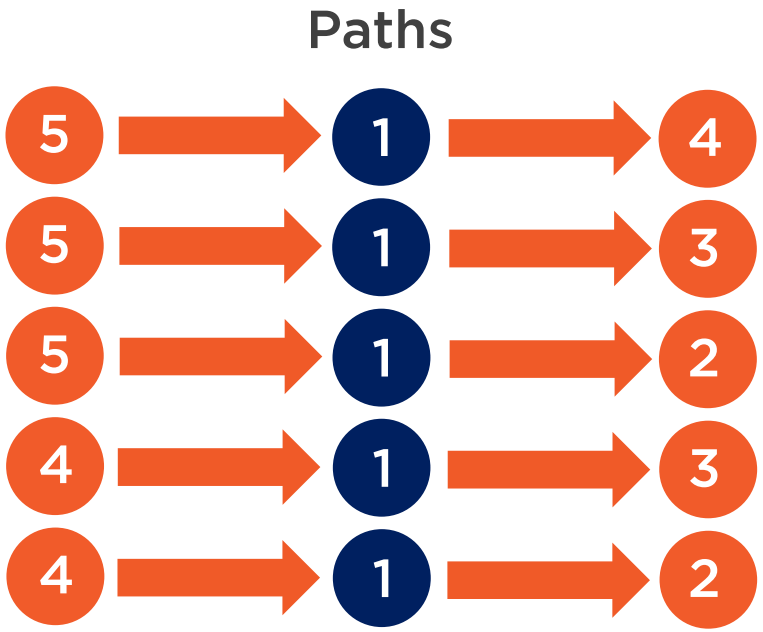
Betweenness Example



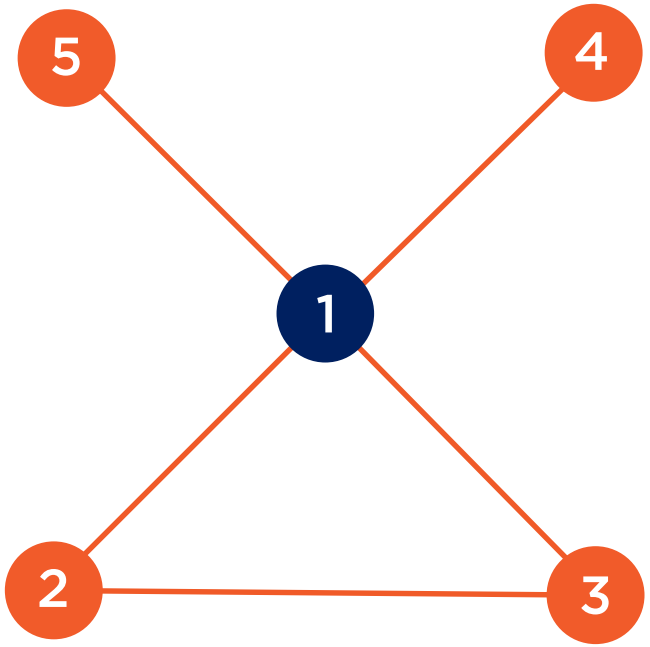


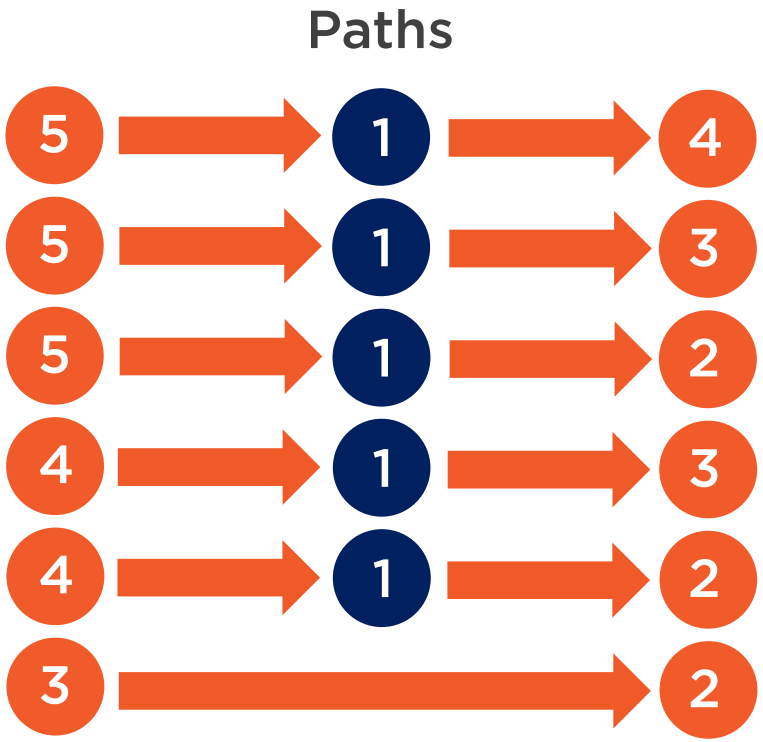
Betweenness Example



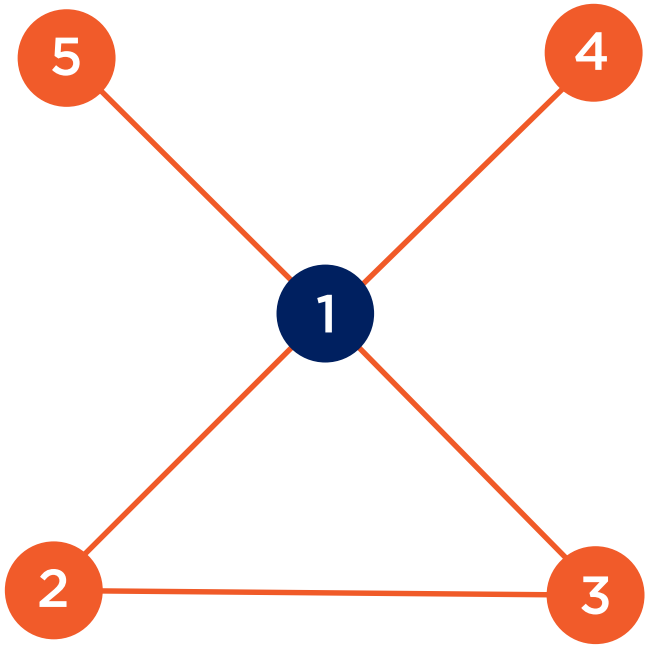


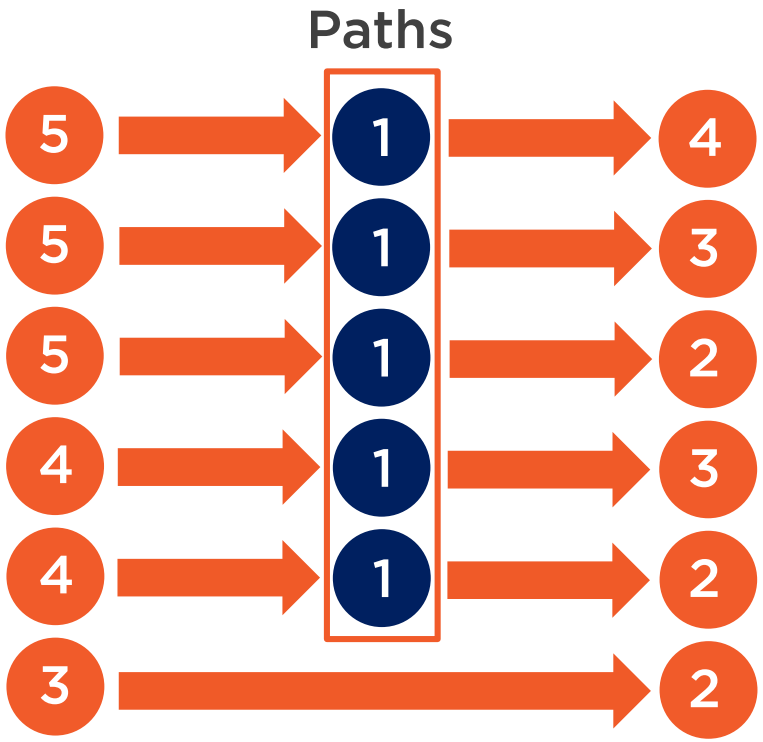
Betweenness Example





Betweenness Example



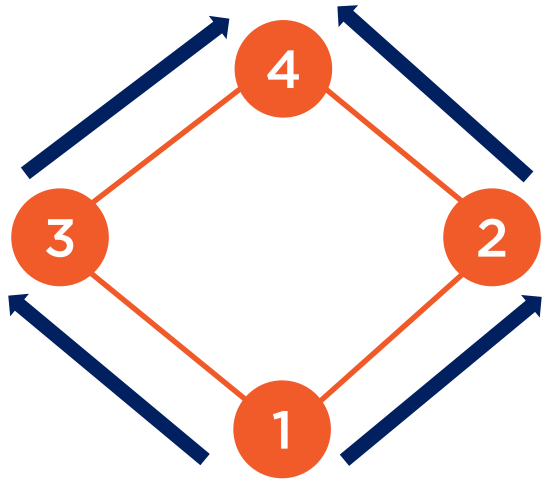


Betweenness Formulation

$$paths(s,t) \coloneqq number\ of\ \boldsymbol{shortest}\ paths\ between\ nodes\ s\ and\ t$$

Betweenness Formulation

$$paths(s,t) \coloneqq number\ of\ \boldsymbol{shortest}\ paths\ between\ nodes\ s\ and\ t$$



$$paths(1,4) = 2$$

Betweenness Formulation

$$paths(s,t) \coloneqq number\ of\ \boldsymbol{shortest}\ paths\ between\ nodes\ s\ and\ t$$

$$paths(s,t|u) \coloneqq number\ of\ \boldsymbol{shortest}\ paths\ between\ nodes\ s\ and\ t\ \boldsymbol{going\ through\ u}$$

Betweenness Formulation

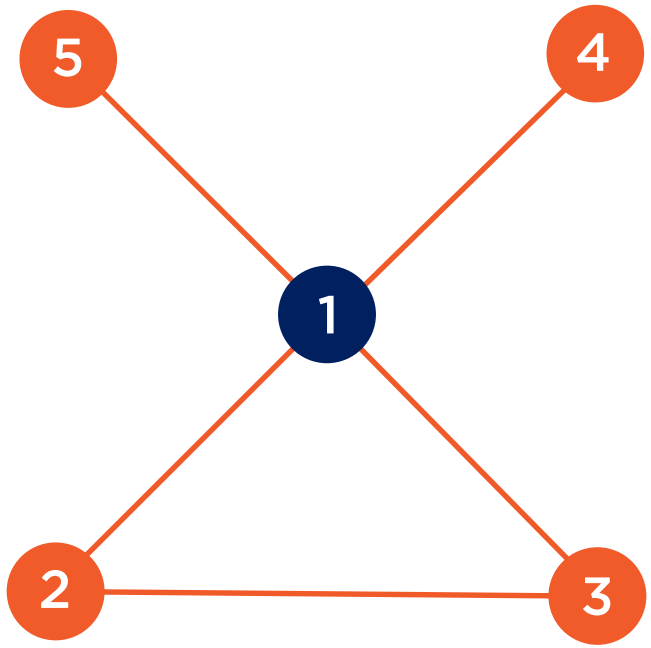
$$paths(s,t) \coloneqq number\ of\ \textbf{\textit{shortest}}\ paths\ between\ nodes\ s\ and\ t$$

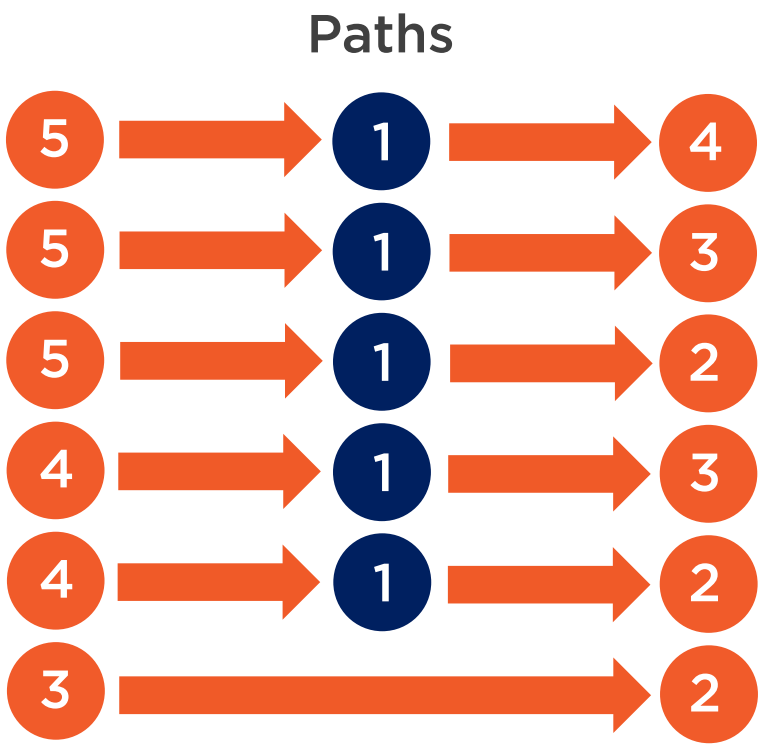
$$paths(s,t|u) \coloneqq number\ of\ \textbf{\textit{shortest}}\ paths\ between\ nodes\ s\ and\ t\ \textbf{\textit{going through}}\ \boldsymbol{u}$$

$$C(u) = \sum_{For\ every\ pair\ of\ nodes\ s\ and\ t} \frac{paths(s,t|u)}{paths(s,t)}$$

Betweenness Example

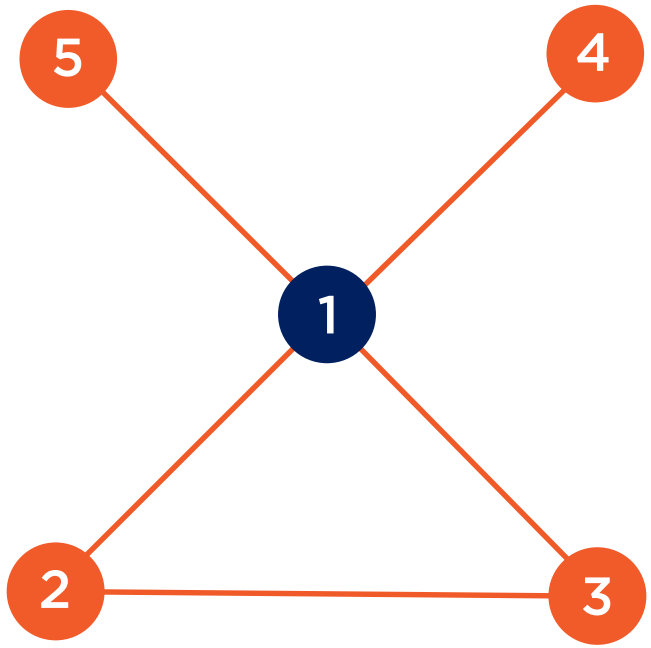
$$\frac{paths(s,t|1)}{paths(s,t)}$$



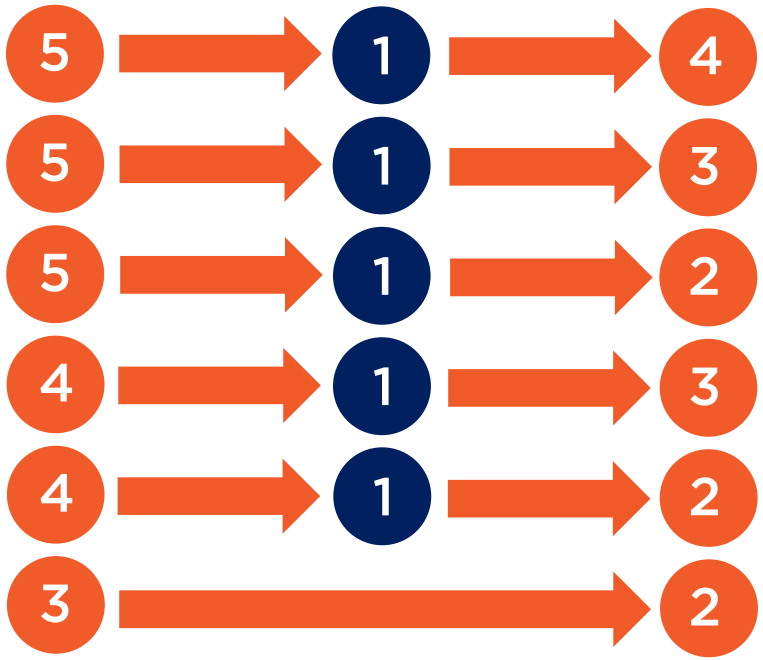


Betweenness Example

$$\frac{paths(s,t|1)}{paths(s,t)}$$



Paths

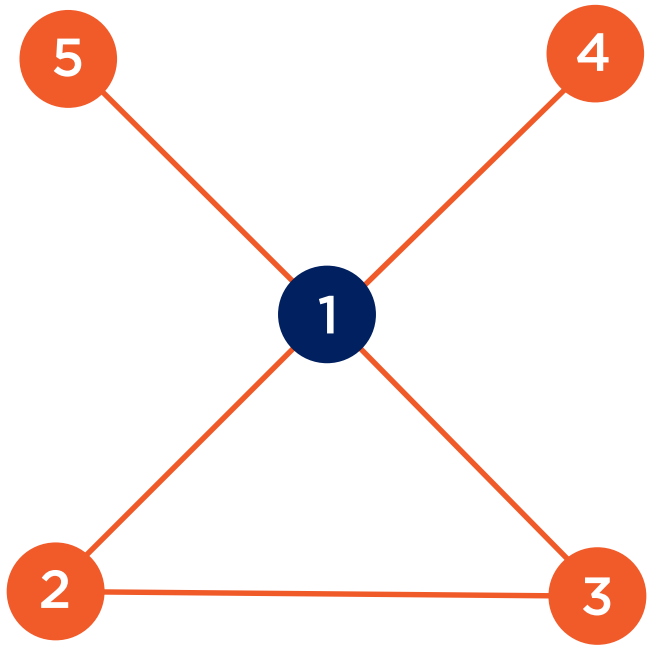


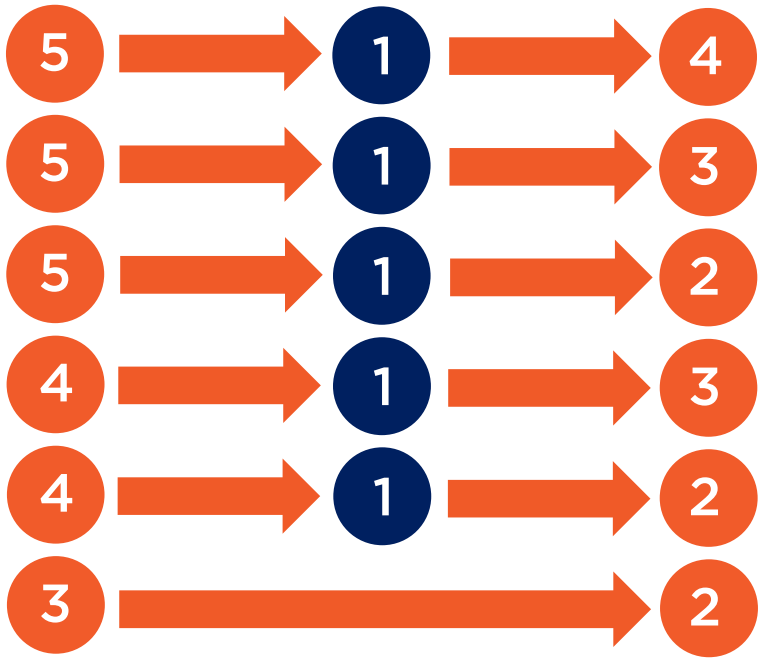
Contribution

1

Betweenness Example

$$\frac{paths(s,t|1)}{paths(s,t)}$$

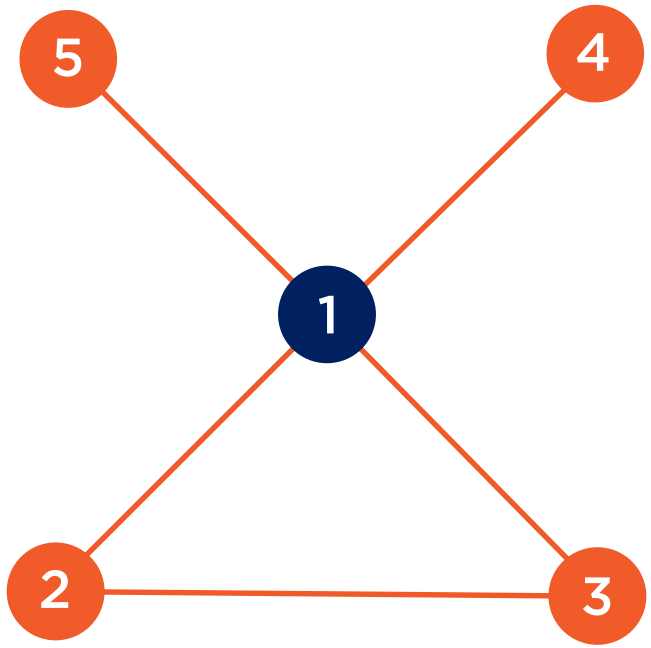


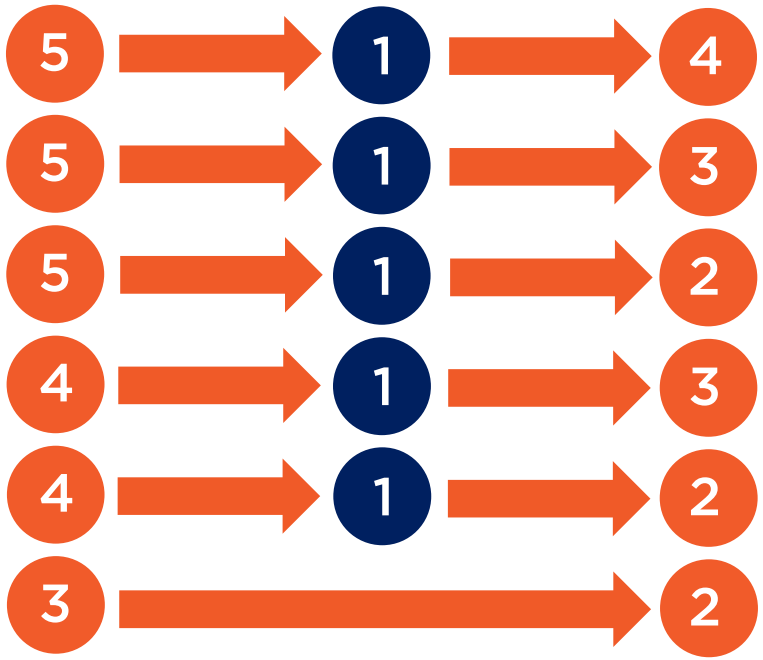


Contribution
1
1
1
1
1

Betweenness Example

$$\frac{paths(s,t|1)}{paths(s,t)}$$

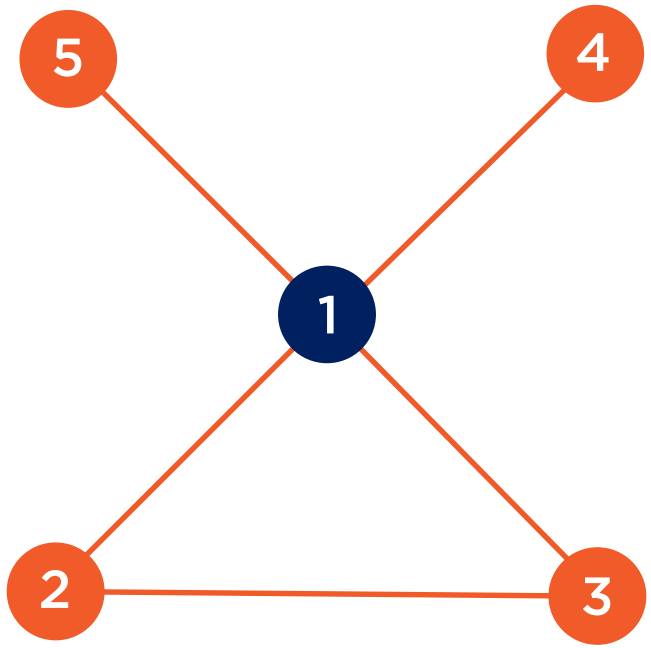


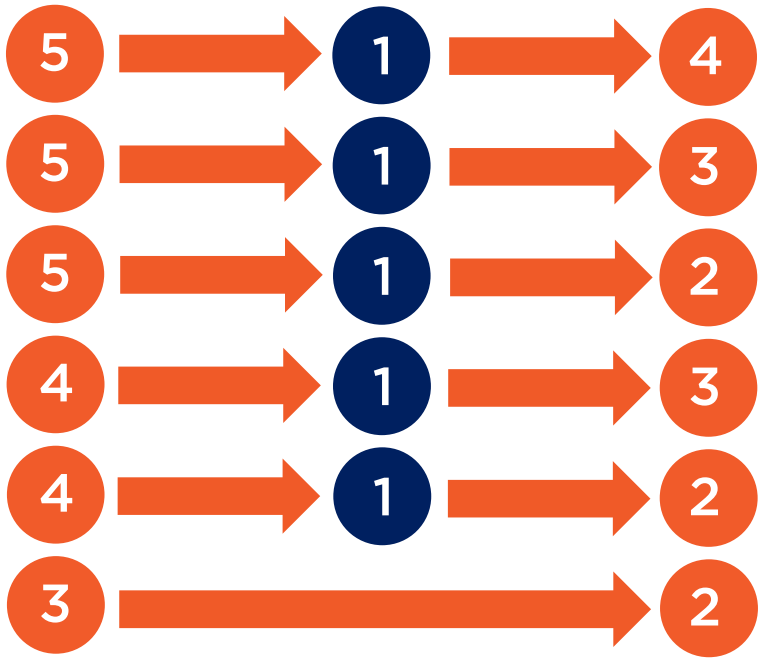


Contribution
1
1
1
1
1
0

Betweenness Example

$$\frac{paths(s,t|1)}{paths(s,t)}$$





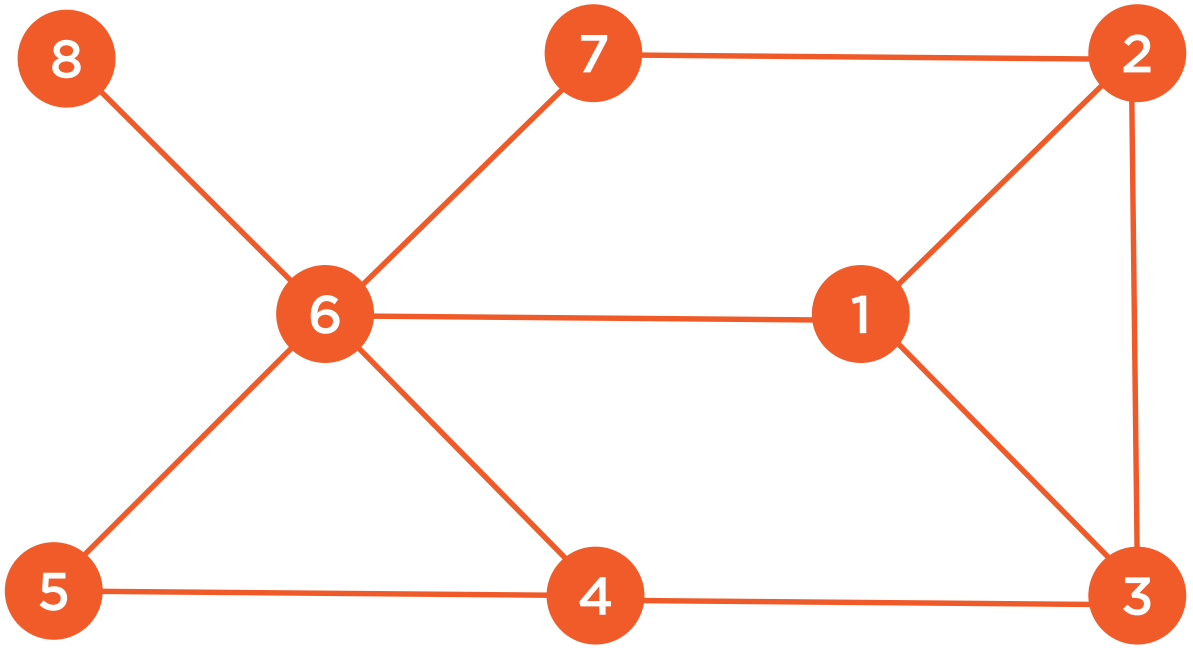
	Contribution
	1
	1
	1
	1
	1
	0
	=
$C(1) =$	5

```
nx.betweenness_centrality(G)
```

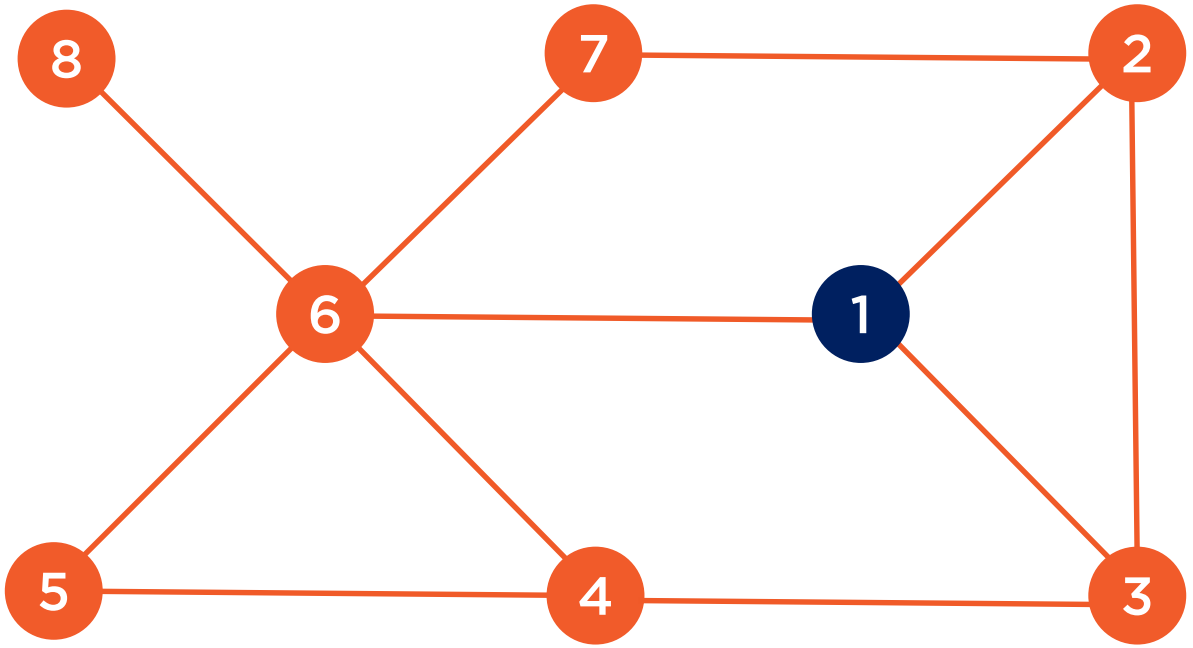
Betweenness Centrality in NetworkX

Katz, Eigenvector, and PageRank Centralities

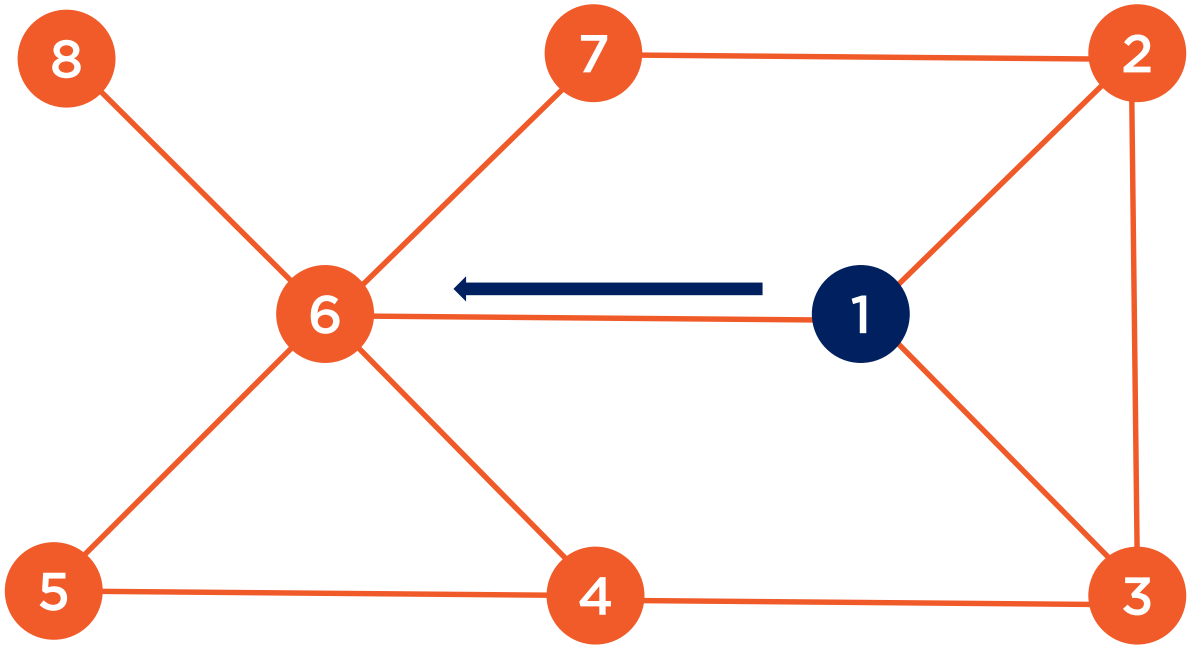
Katz Centrality



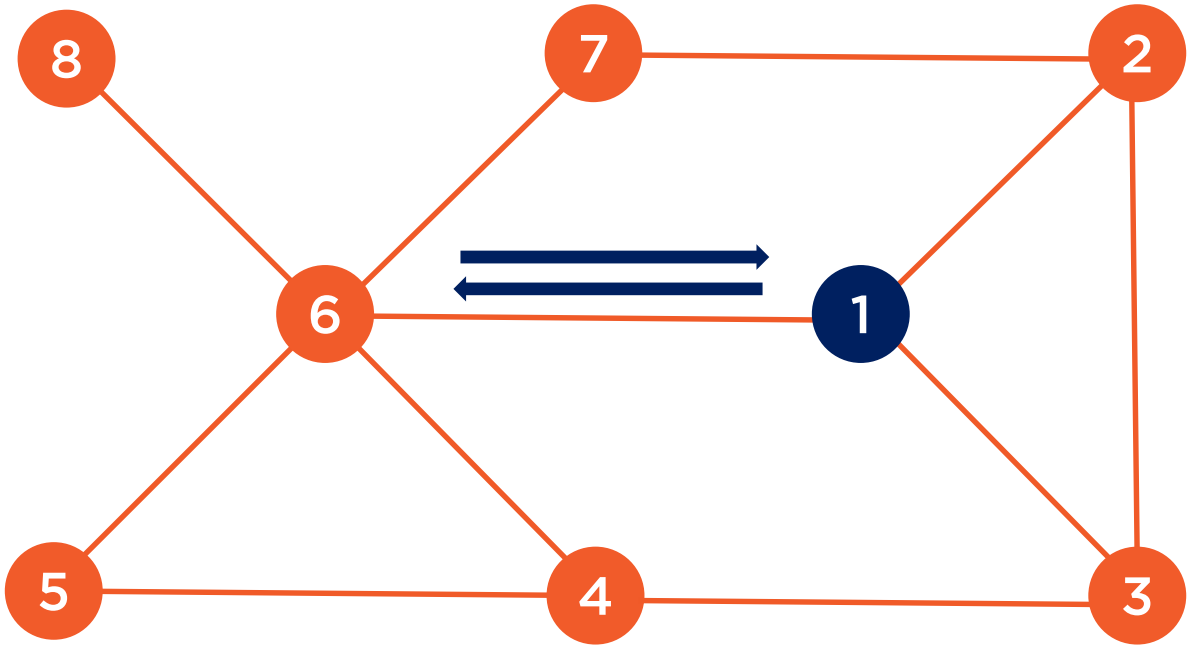
Katz Centrality



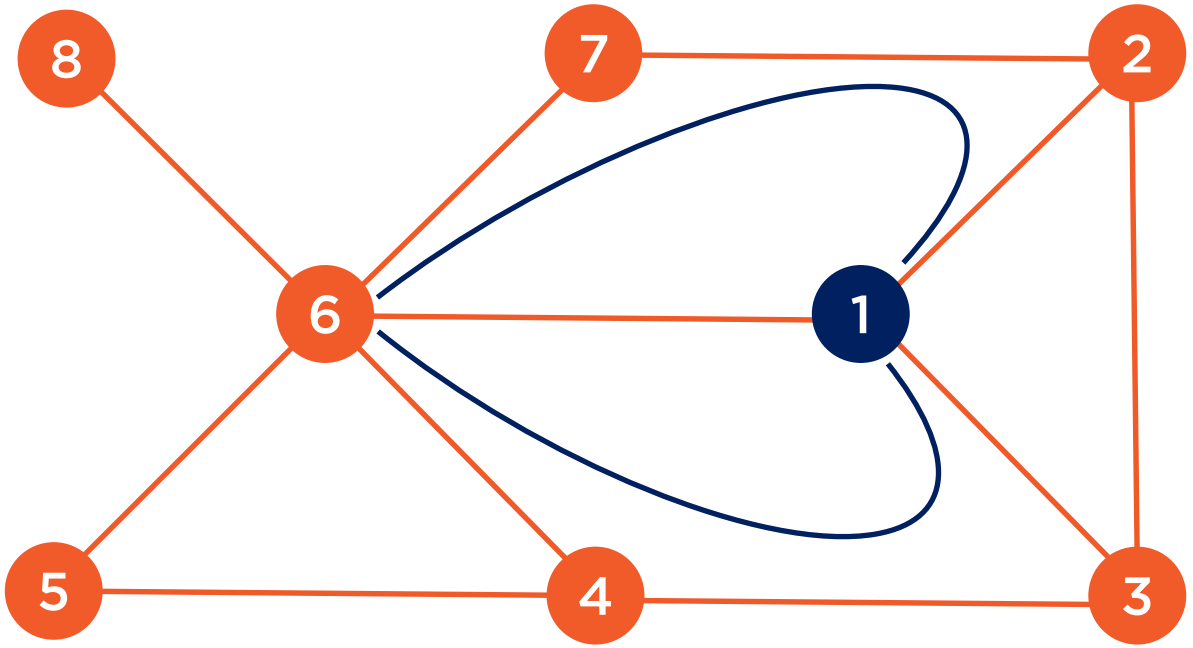
Katz Centrality



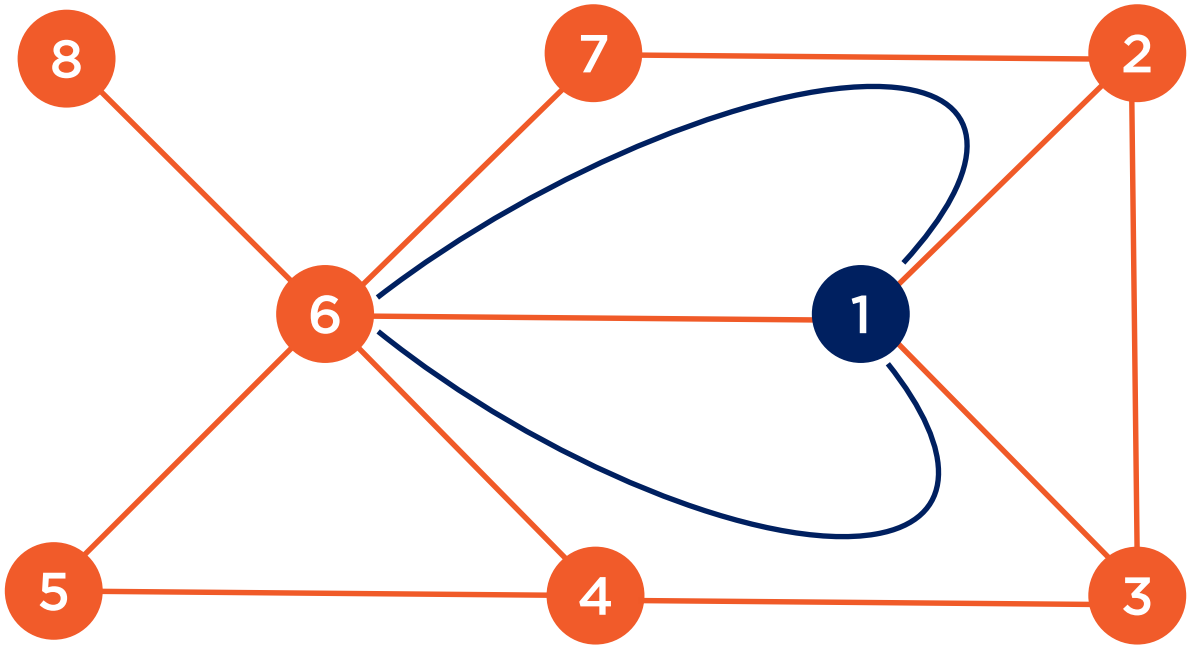
Katz Centrality



Katz Centrality



Katz Centrality

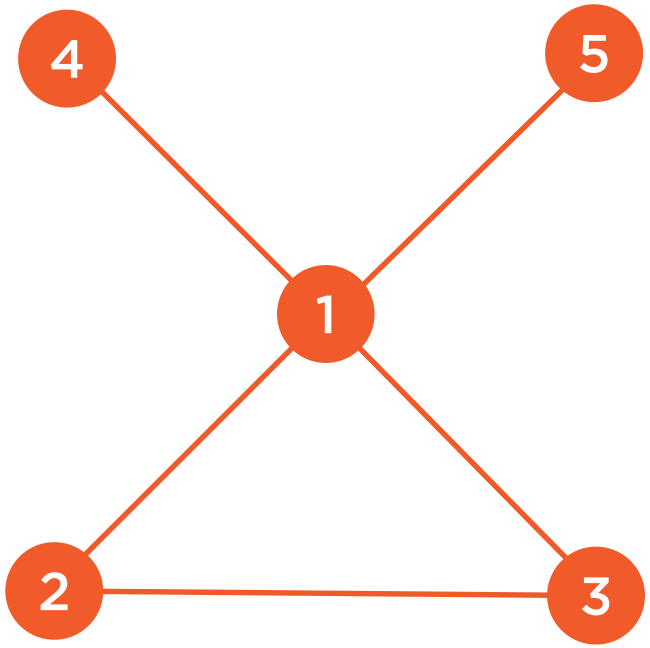


$$Attenuation\ factor\ \alpha = 0.5$$

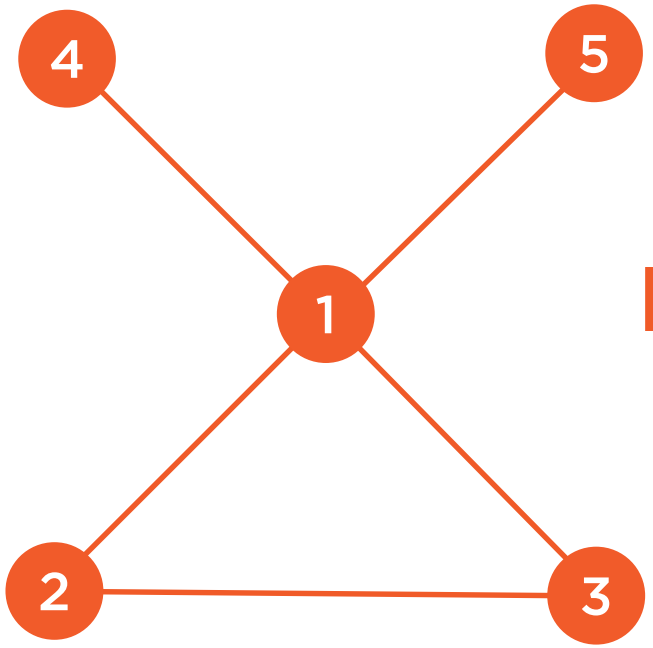
```
nx.katz_centrality(G, alpha=0.5)
```

Katz Centrality in NetworkX

Eigenvector Centrality



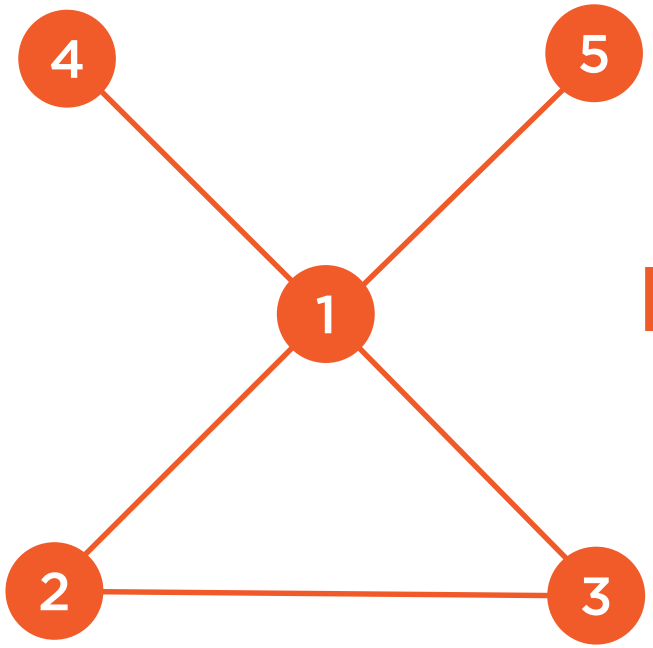
Eigenvector Centrality



Adjacency matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

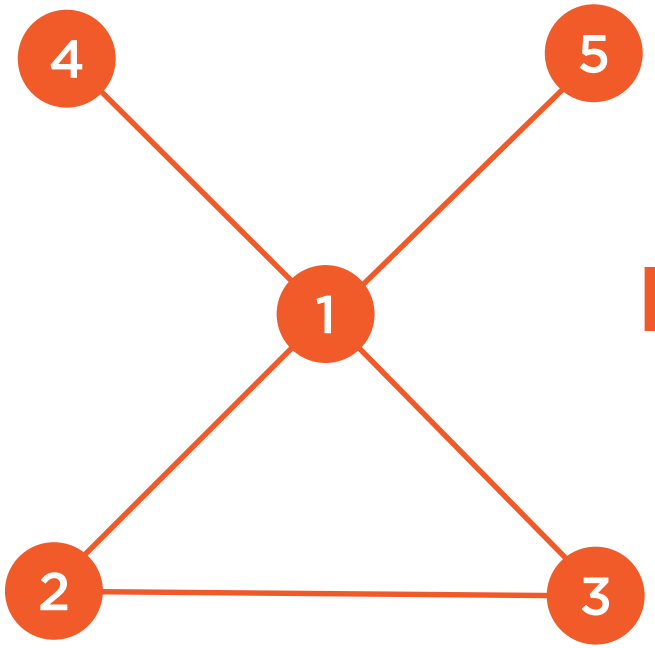
Eigenvector Centrality



Adjacency matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Eigenvector Centrality



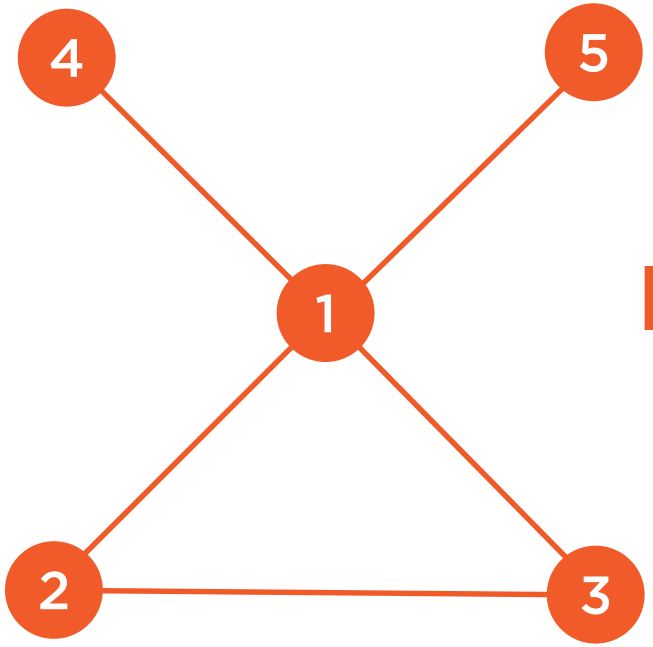
Adjacency matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Eigenvalues

$$(2.34, -1.81, 1, 0.47, 0)$$

Eigenvector Centrality



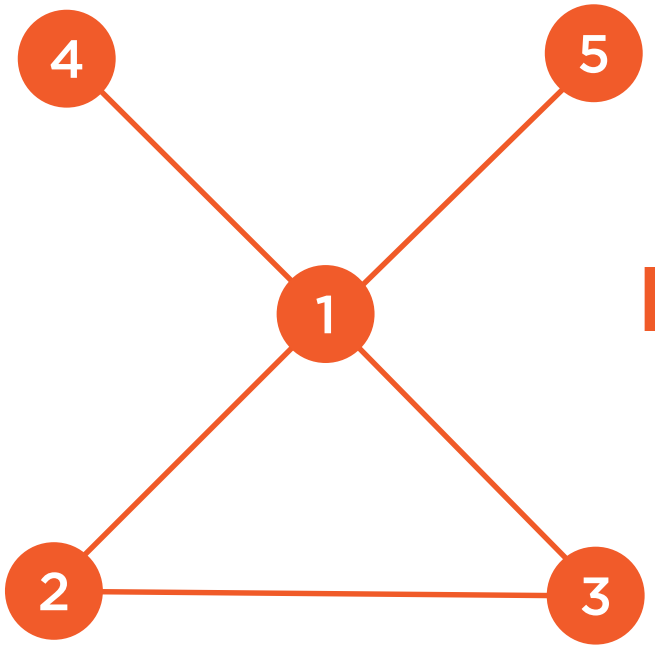
Adjacency matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Eigenvalues

$$(2.34, -1.81, 1, 0.47, 0)$$

Eigenvector Centrality



Adjacency matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

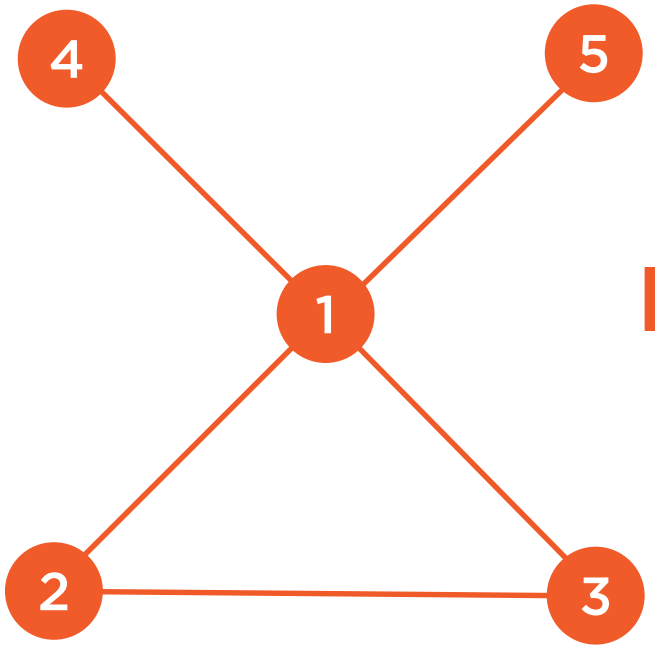
Eigenvalues

$$(2.34, -1.81, 1, 0.47, 0)$$

Principal Eigenvector

$$\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{pmatrix} 2.34 \\ 1.74 \\ 1.74 \\ 1 \\ 1 \end{pmatrix}$$

Eigenvector Centrality



Adjacency matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

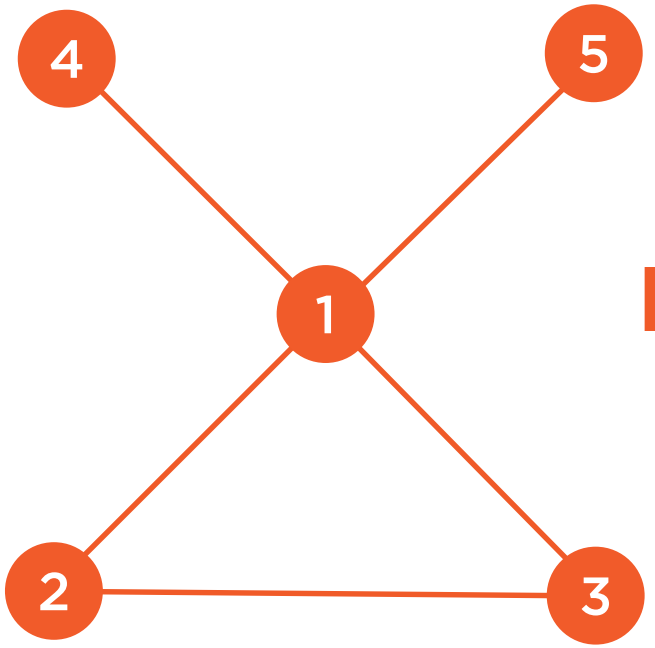
Eigenvalues

$$(2.34, -1.81, 1, 0.47, 0)$$

Principal Eigenvector

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{pmatrix} 2.34 \\ 1.74 \\ 1.74 \\ 1 \\ 1 \end{pmatrix}$$

Eigenvector Centrality



Adjacency matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

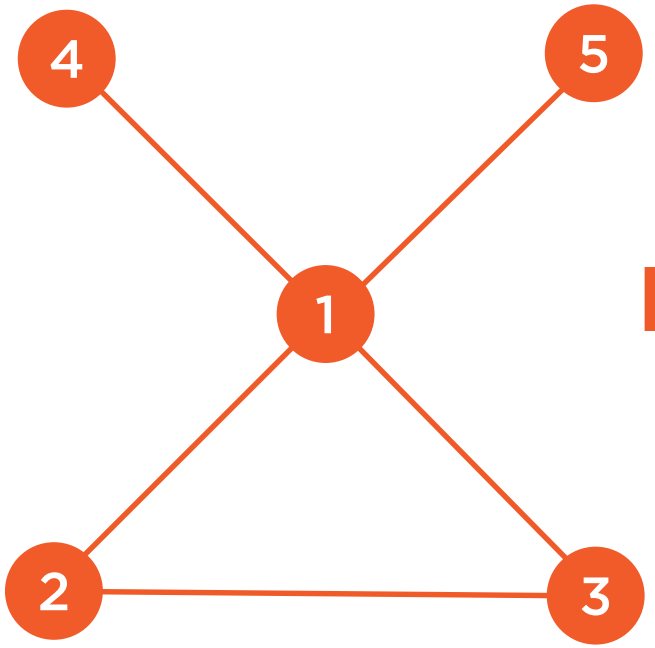
Eigenvalues

$$(2.34, -1.81, 1, 0.47, 0)$$

Principal Eigenvector

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{pmatrix} 2.34 \\ 1.74 \\ 1.74 \\ 1 \\ 1 \end{pmatrix}$$

Eigenvector Centrality



Adjacency matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Eigenvalues

$$(2.34, -1.81, 1, 0.47, 0)$$

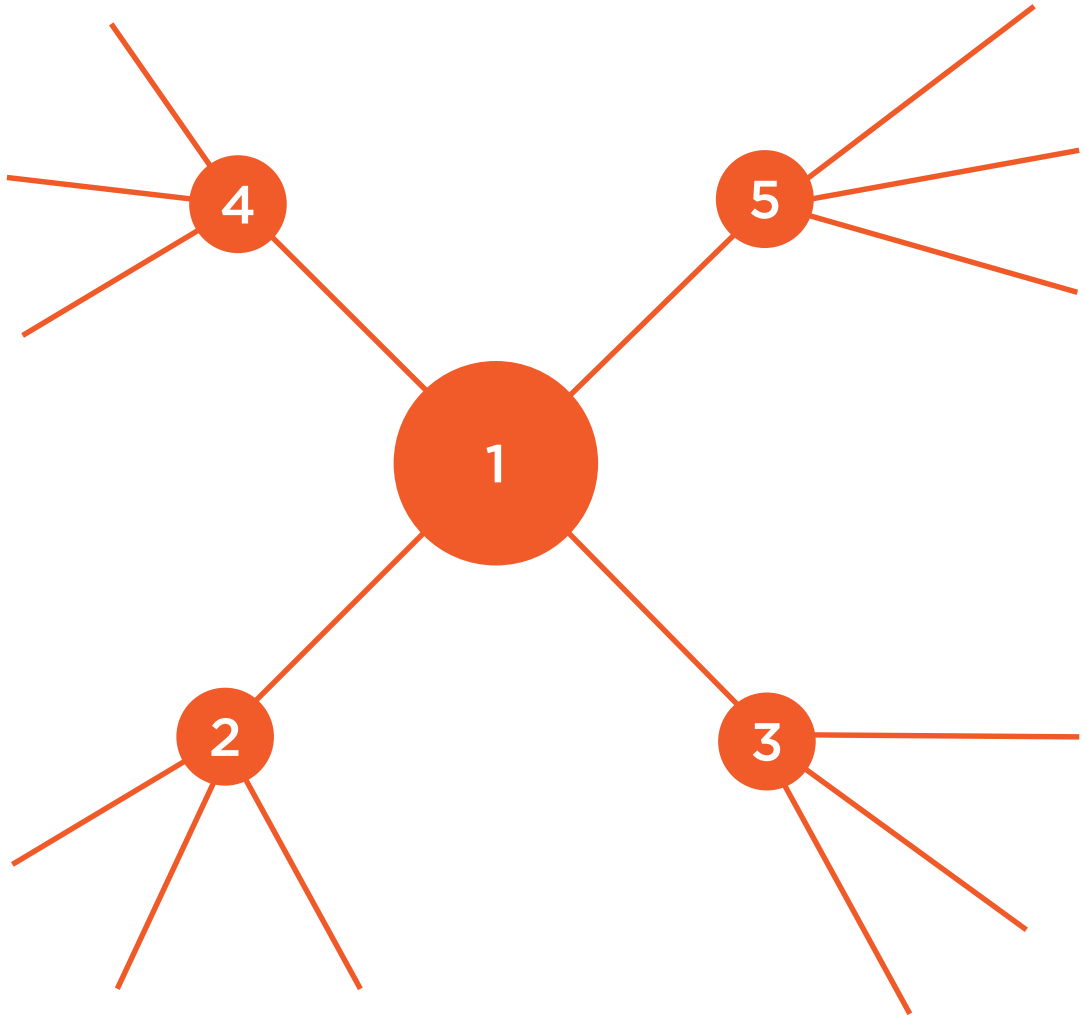
Principal Eigenvector

$$\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{pmatrix} 2.34 \\ 1.74 \\ 1.74 \\ 1 \\ 1 \end{pmatrix}$$

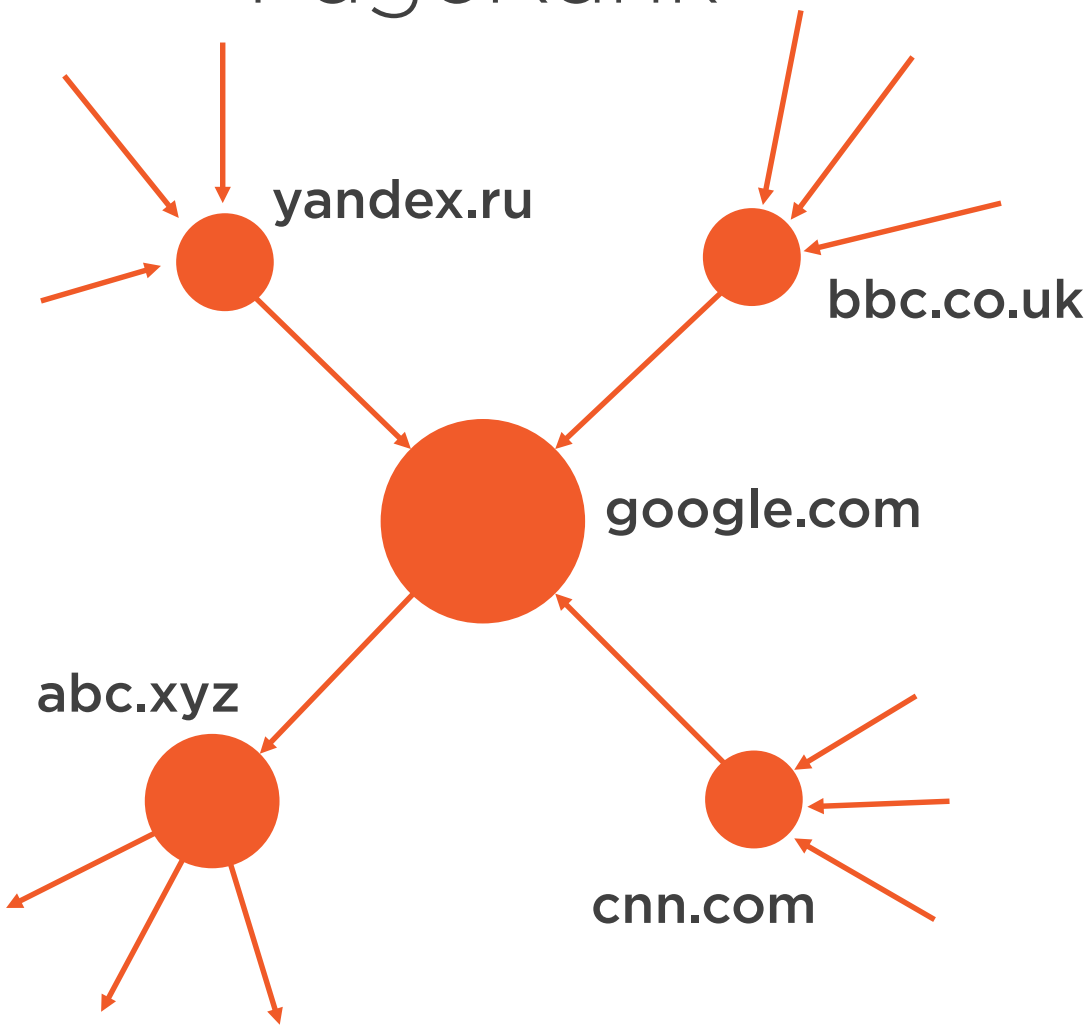
Eigenvector vs. Katz

$$\alpha \approx \frac{1}{\lambda_{principal}}$$

Alternative Definition



PageRank



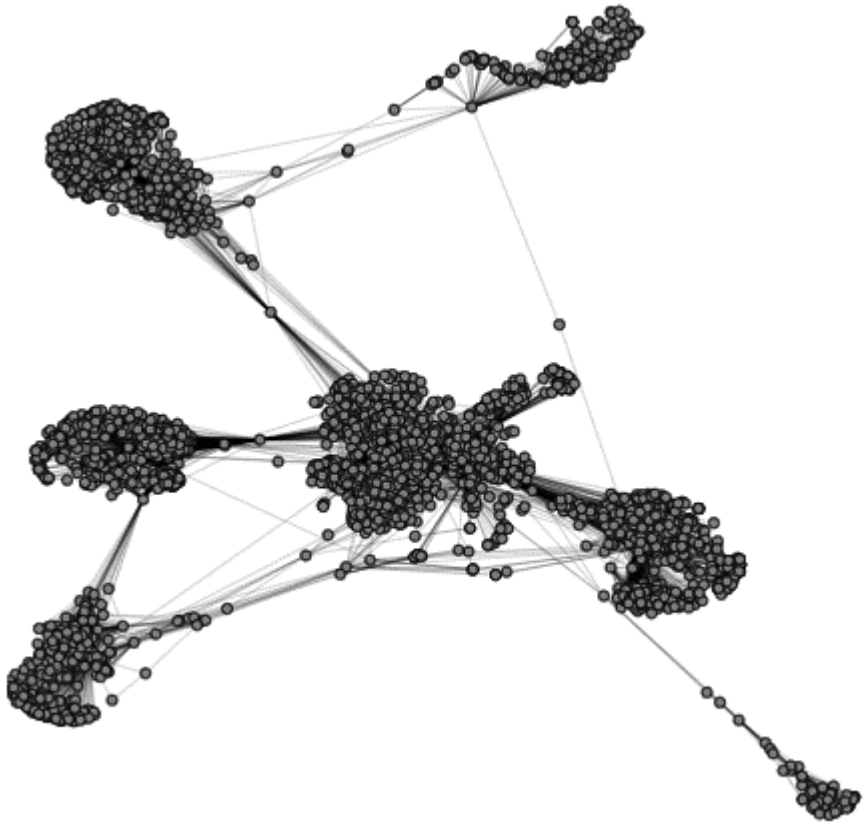
```
nx.eigenvector_centrality(G)

nx.pagerank_centrality(G)
```

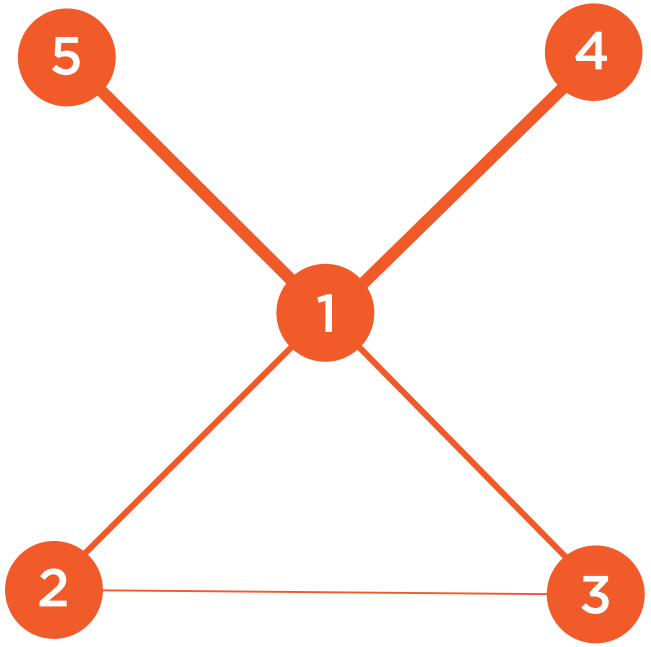
Eigenvector and PageRank Centrality in NetworkX

Community Detection: Girvan-Newman Algorithm

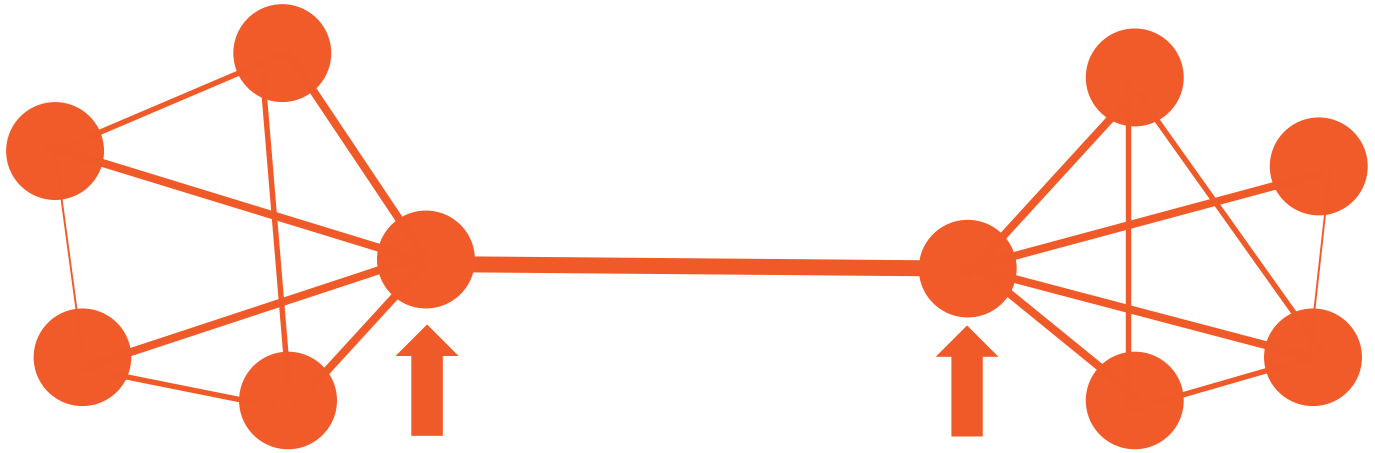
Communities



Edge Betweenness Example



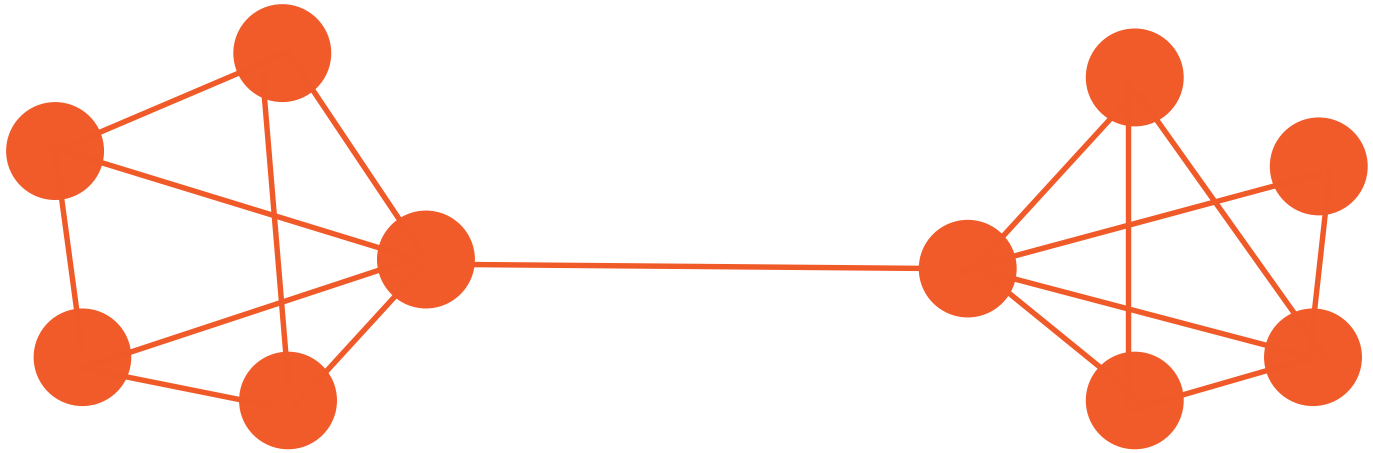
Girvan-Newman Intuition



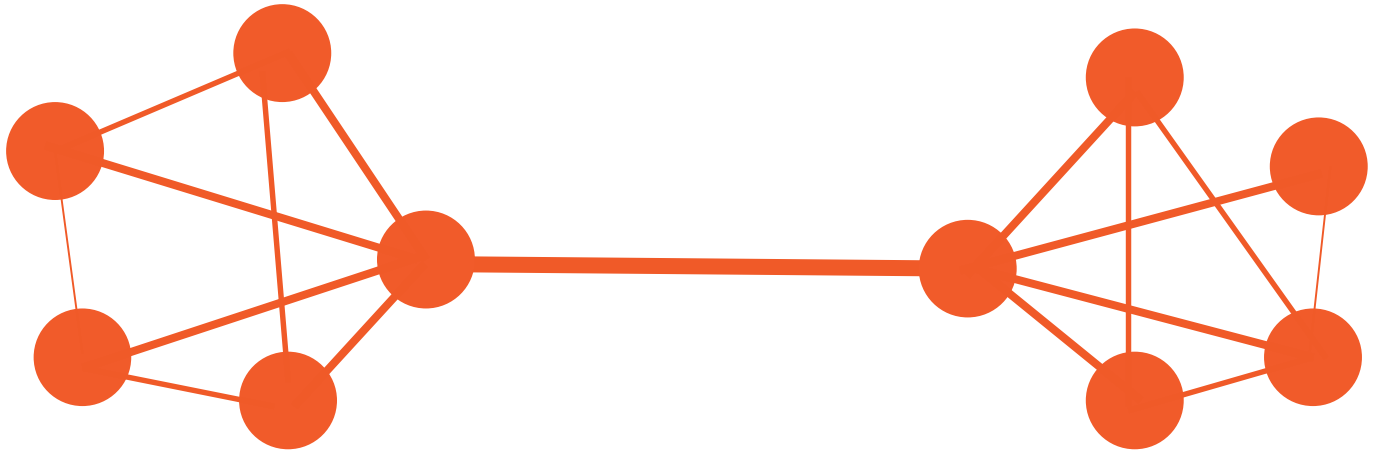
Assumption: if two nodes are connected by an edge with high betweenness centrality, those nodes are likely to be part of separate communities.

Conversely, if they are connected by an edge with low betweenness centrality, they are likely to be part of the same community.

Girvan-Newman Intuition

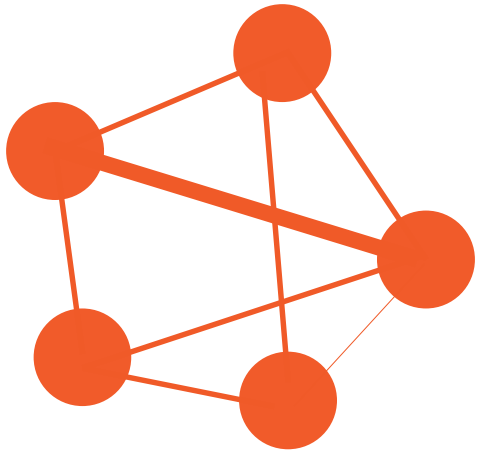


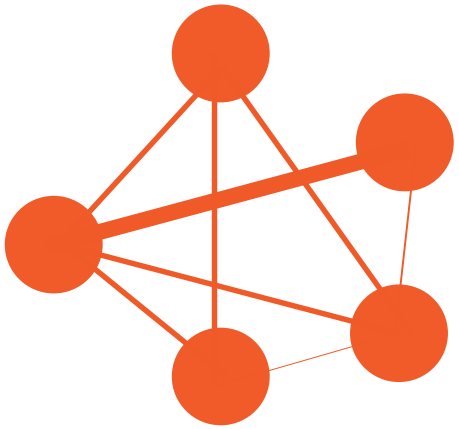
Girvan-Newman Intuition



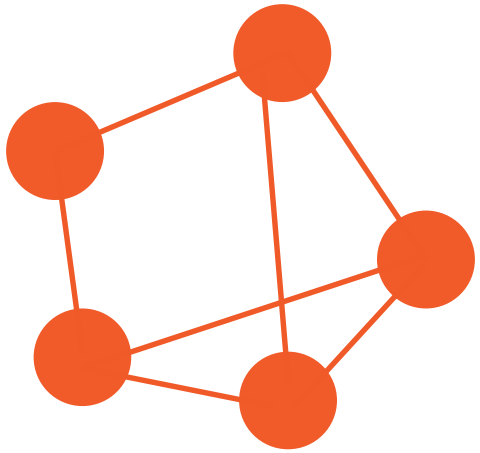
Girvan-Newman Intuition

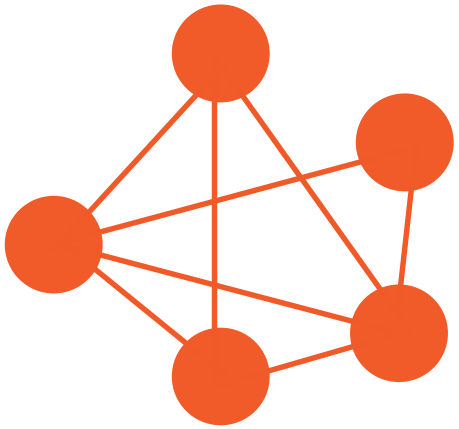
Girvan-Newman Intuition





Girvan-Newman Intuition





Girvan-Newman Intuition

Girvan-Newman Intuition

Community	Members
A	0
B	1
C	2
D	3
E	4
F	5
G	6
H	7
I	8
J	9

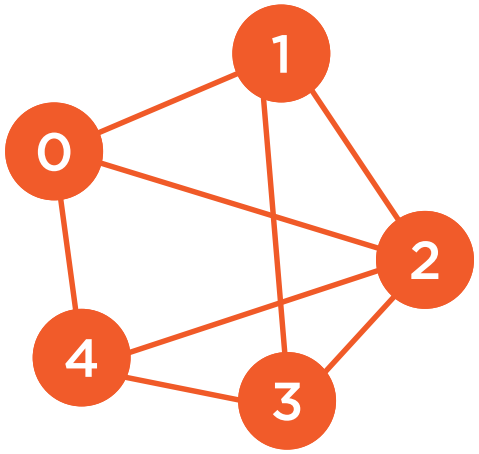
Girvan-Newman Intuition

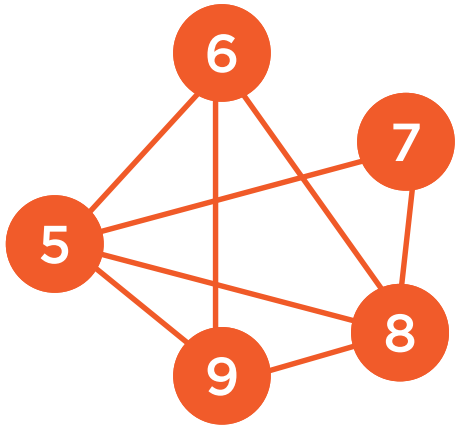
Community	Members
AE	0,4
B	1
C	2
D	3
F	5
G	6
H	7
I	8
J	9

Girvan-Newman Intuition

Community	Members
AE	0,4
B	1
C	2
D	3
F	5
G	6
HI	7,8
J	9

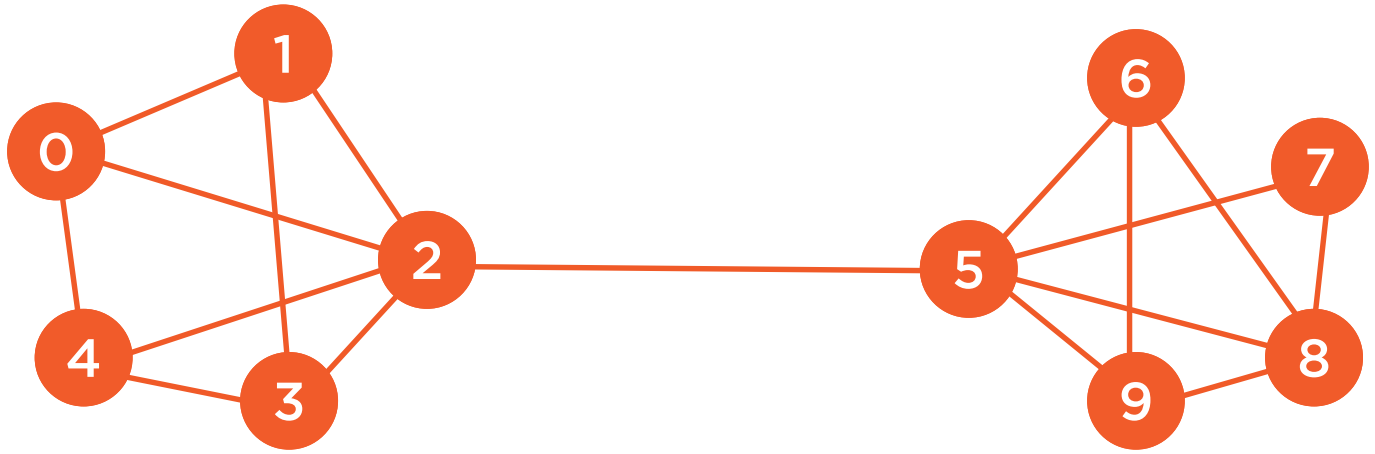
Girvan-Newman Intuition





Community	Members
AECDB	0,4,2,3,1
HIGFJ	7,8,6,5,9

Girvan-Newman Intuition

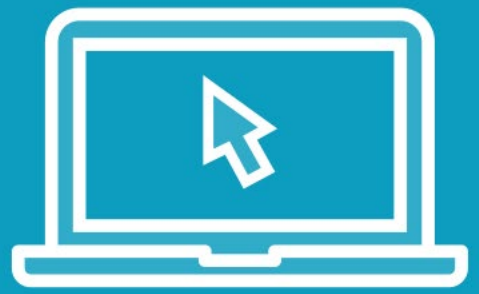


Community	Members
AECDBHIGFJ	0,4,2,3,1,7,8,6,5,9

Girvan-Newman Time Complexity

$$O(E^2N)$$

Demo



Detecting communities in NetworkX

Summary

Many centrality measures because many notions of “center” in a network

- Degree centrality for identifying local leaders
- Closeness centrality for identifying nodes on average closest to all other nodes
- Betweenness centrality for identifying bridges and communication hubs

Katz and Eigenvector are great for identifying global influence

- Katz gives more control, but Eigenvector has reasonable defaults

PageRank is similar to Eigenvector but adjusted for directed networks

Girvan-Newman used to detect levels of communities algorithmically