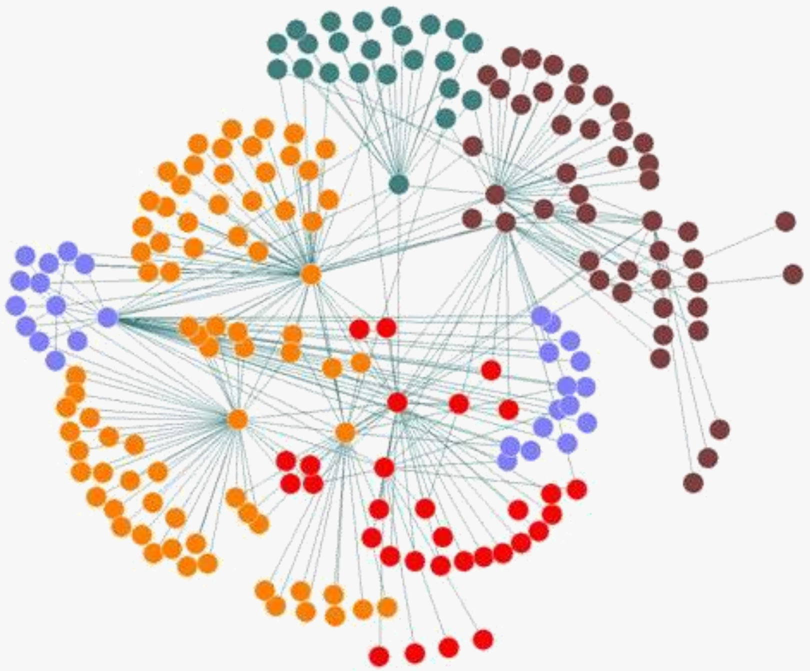
Social Network Analysis

By
Dr. Hussein Hazimeh



Lebanese University

Faculty of Information 1

Data Science Departement

3rd year – Social Network Analysis

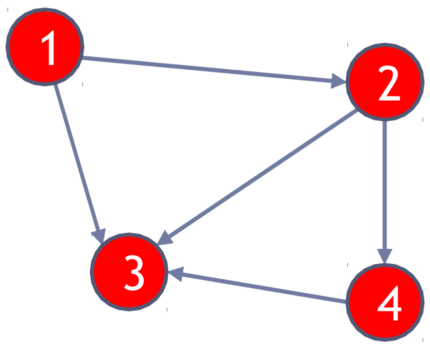
Spring – 2022 – Chapter 3

Agenda

- Network-to-matrix
- Whole & Ego Nets
- Network Centrality
- Centrality Measures
- Eigenvector
- Applications

Entering data on a directed graph

Graph (directed)



Edge list

Vertex	Vertex
1	2
1	3
2	3
2	4

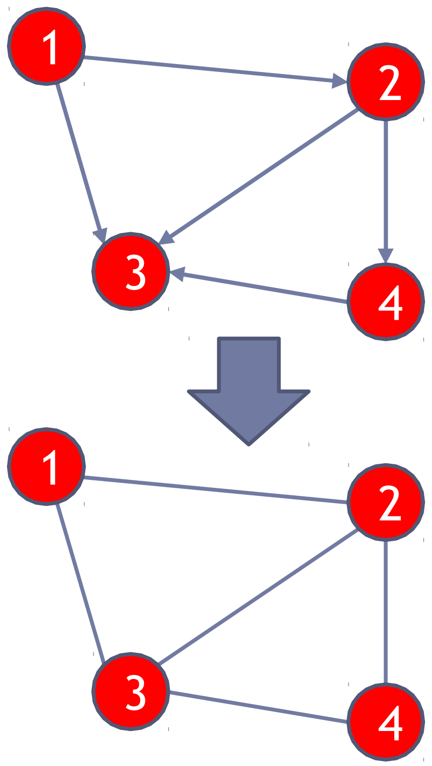
Adjacency matrix

Vertex	1	2	3	4
1	-	1	1	0
2	0	-	1	1
3	0	0	-	0
4	0	0	1	-

Representing an undirected graph

Directed

(who contacts whom)



Undirected

(who knows whom)

Edge list remains the same

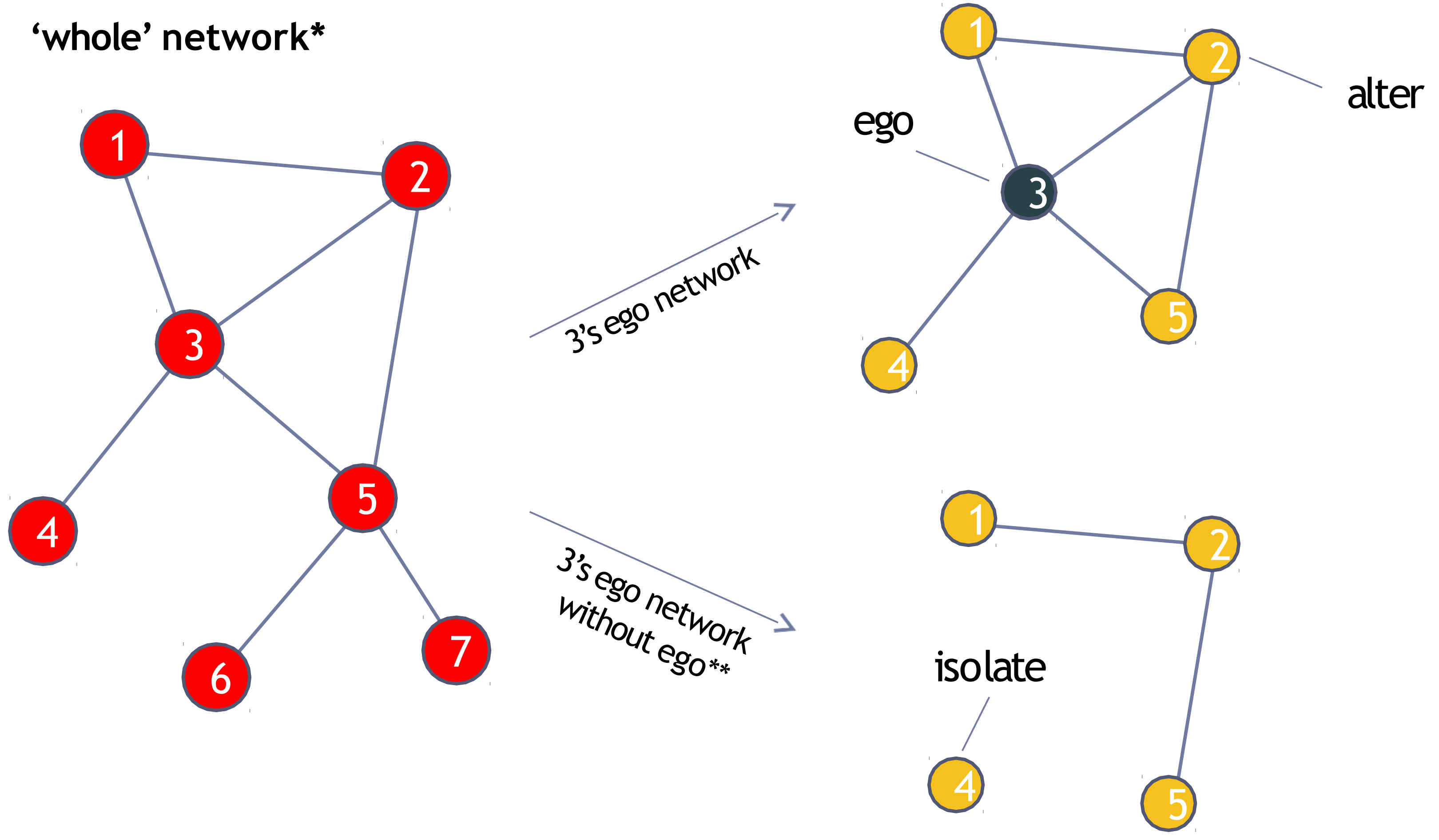
Vertex	Vertex
1	2
1	3
2	3
2	4
3	4

But interpretation is different now

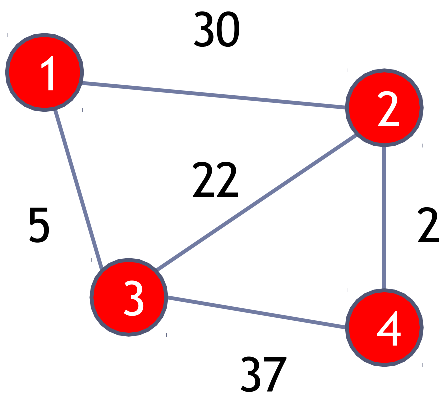
Adjacency matrix becomes symmetric

Vertex	1	2	3	4
1	-	1	1	0
2	1	-	1	1
3	1	1	-	1
4	0	1	1	-

Ego networks and 'whole' networks



Adding weights to edges (directed or undirected)



Weights could be:

- *Frequency of interaction in period of observation*
- *Number of items exchanged in period*
- *Individual perceptions of strength of relationship*
- *Costs in communication or exchange, e.g. distance*
- *Combinations of these*

Edge list: add column of weights

Vertex	Vertex	Weight
1	2	30
1	3	5
2	3	22
2	4	2
3	4	37

Adjacency matrix: add weights instead of 1

Vertex	1	2	3	4
1	-	30	5	0
2	30	-	22	2
3	5	22	-	37
4	0	2	37	-

What is centrality?!

- Centrality measures address the question:
  "Who is the most important or central person in this network?"

- There are many answers to this question, depending on what we mean by importance.

- According to Scott Adams, the power a person holds in the organization is inversely proportional to the number of keys on his keyring.

  - A janitor has keys to every office, and no power.
  - The CEO does not need a key: people always open the door for him.

- There are a vast number of different centrality measures that have been proposed over the years.

Centrality Measures

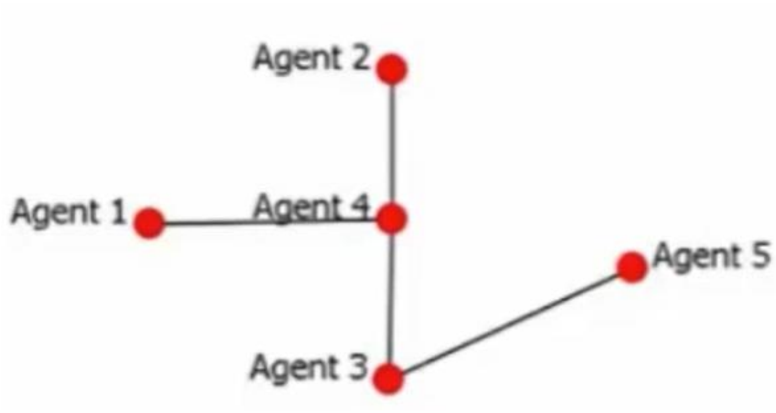
Centrality measure	Interpretation in social networks
Degree	How many people can this person reach directly?
Betweenness	How likely is this person to be the most direct route between two people in the network?
C loseness	How fast can this person reach everyone in the network?
Eigenvector	How well is this person connected to other well-connected people?

1

Degree Centrality

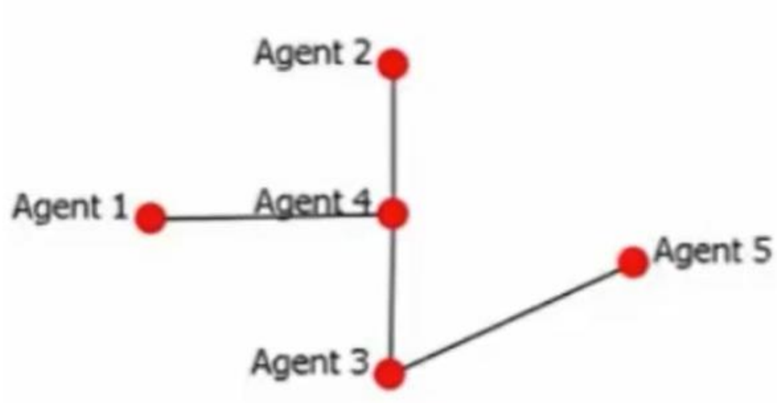
Degree Centrality

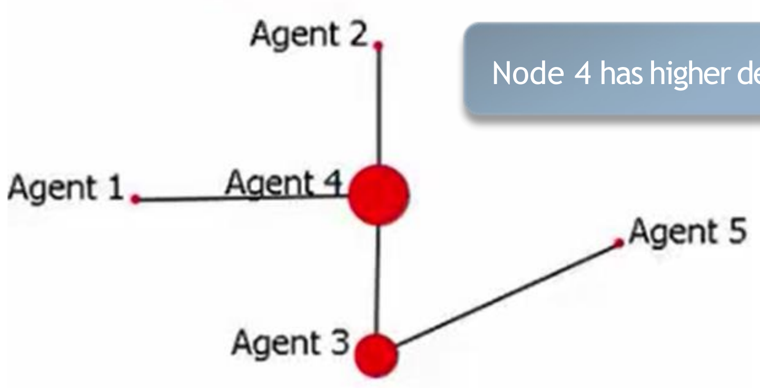
$$C_{D_i} = \frac{\sum_{j=1}^{n} a_{ij}}{n-1}$$



Degree centrality is defined as the number of links incident upon a node (i.e., the number of ties that a node has). If the network is directed (meaning that ties have direction), then two separate measures of **degree centrality** are defined, namely, indegree and outdegree.

Degree Centrality (cont.)

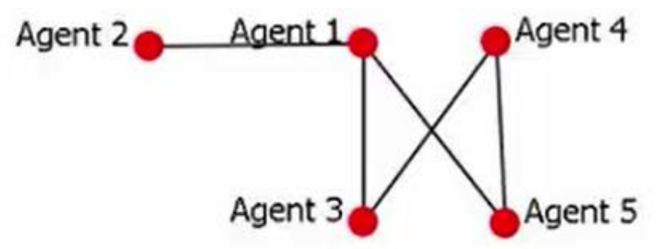


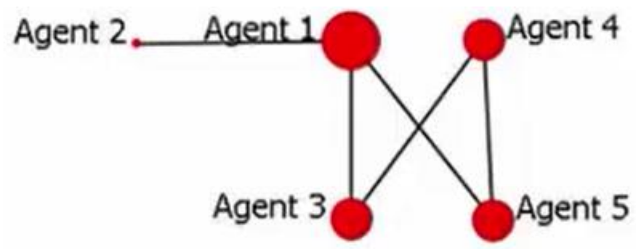


Node 4 has higher degree centrality than 3

A A	Agent 1	Agent 2	Agent 3	Agent 4	Agent 5	Sum	/	(N-1)	=	Degree
Agent 1		0	0	1	0	1		4		1/4
Agent 2	0		0	1	0	1		4		1/4
Agent 3	0	0		1	1	2		4		2/4
Agent 4	1	1	1		0	3		4		3/4
Agent 5	0	0	1	0		1		4		1/4

Degree Centrality (cont.)



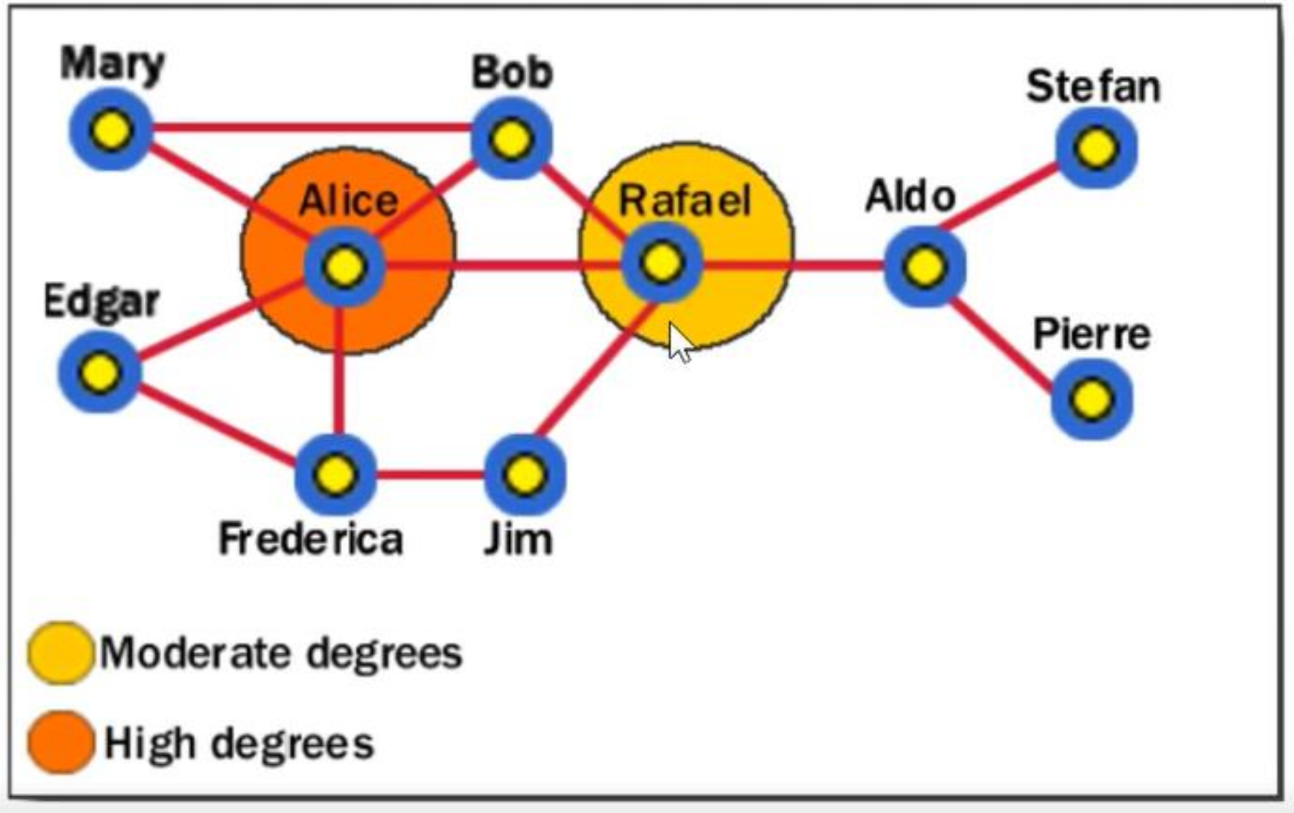


Node 1 has higher degree centrality than 3

A A	Agent 1	Agent 2	Agent 3	Agent 4	Agent 5	Sum	/	(N-1)	=	Degree
Agent 1		1	1	0	1	3		4		3/4
Agent 2	1		0	0	0	1		4		1/4
Agent 3	1	0		1	0	2		4		2/4
Agent 4	0	0	1		1	2		4		2/4
Agent 5	1	0	0	1		2		4		2/4

Degree Centrality (cont.)

Application

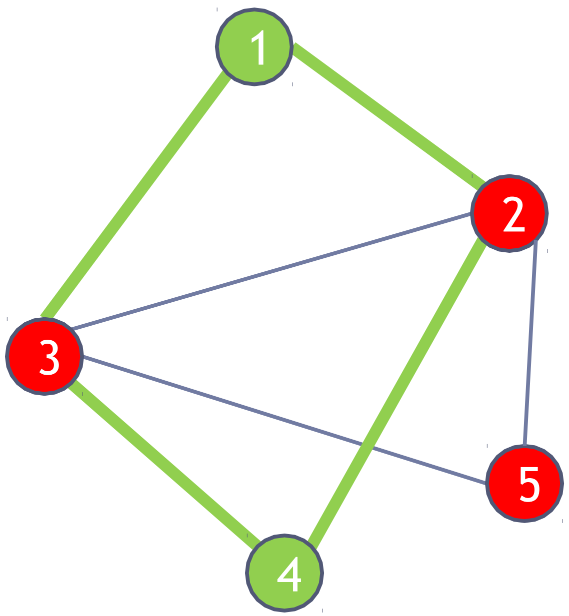


Paths and shortest paths

- A *path* between two nodes is any sequence of non-repeating nodes that connects the two nodes
- The *shortest path* between two nodes is the path that connects the two nodes with the shortest number of edges (also called the *distance* between the nodes)
- In the example to the right, between nodes 1 and 4 there are two shortest paths of length 2: {1,2,4} and {1,3,4}
- Other, longer paths between the two nodes are {1,2,3,4}, {1,3,2,4}, {1,2,5,3,4} and {1,3,5,2,4} (the longest paths)
- Shorter paths are desirable when speed of communication or exchange is desired (often the case in many studies, but sometimes not, e.g. in networks that spread

disease)

Hypothetical graph



2

Closeness Centrality

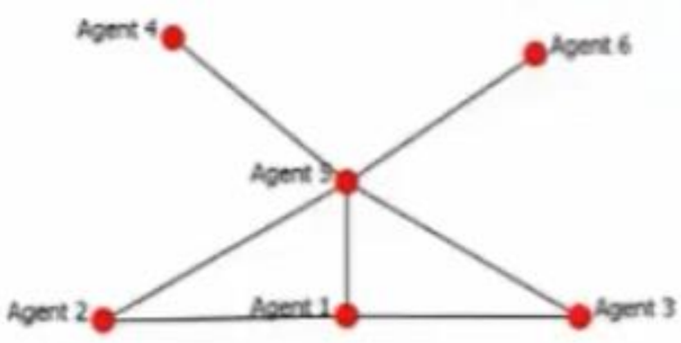
Closeness Centrality

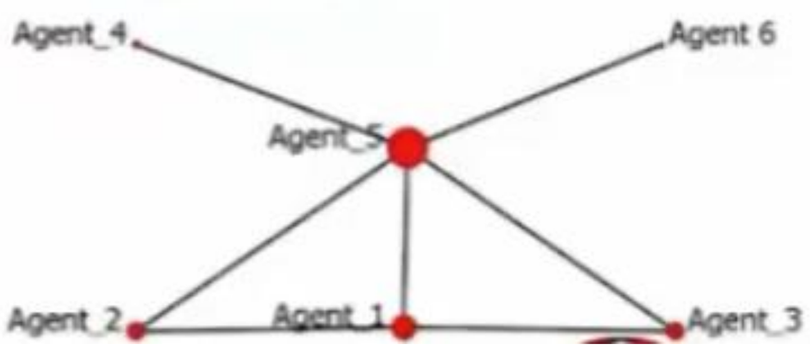
» **Closeness centrality** is a way of detecting nodes that are able to spread information very efficiently through a graph.

» Nodes with a high **closeness** score have the shortest distances to all other nodes.

Closeness Centrality

$$C(x) = \frac{N-1}{\sum_y d(y,x)}$$





	Agent 1	Agent 2	Agent 3	Agent 4	Agent 5	Agent 6
Agent 1		1	1	2	1	2
Agent 2	1		2	2	1	2
Agent 3	1	2		2	1	2
Agent 4	2	2	2		1	2
Agent 5	1	1	1	1		1
Agent 6	2	2	2	2	1	

SUM	/	(n-1)	=	Closeness
7		5		5/7
8		5		8/5
8		5		5/8
9		5		5/9
5		5		5/5
9		5		5/9

2

Betweenness Centrality

Why betweenness centrality?

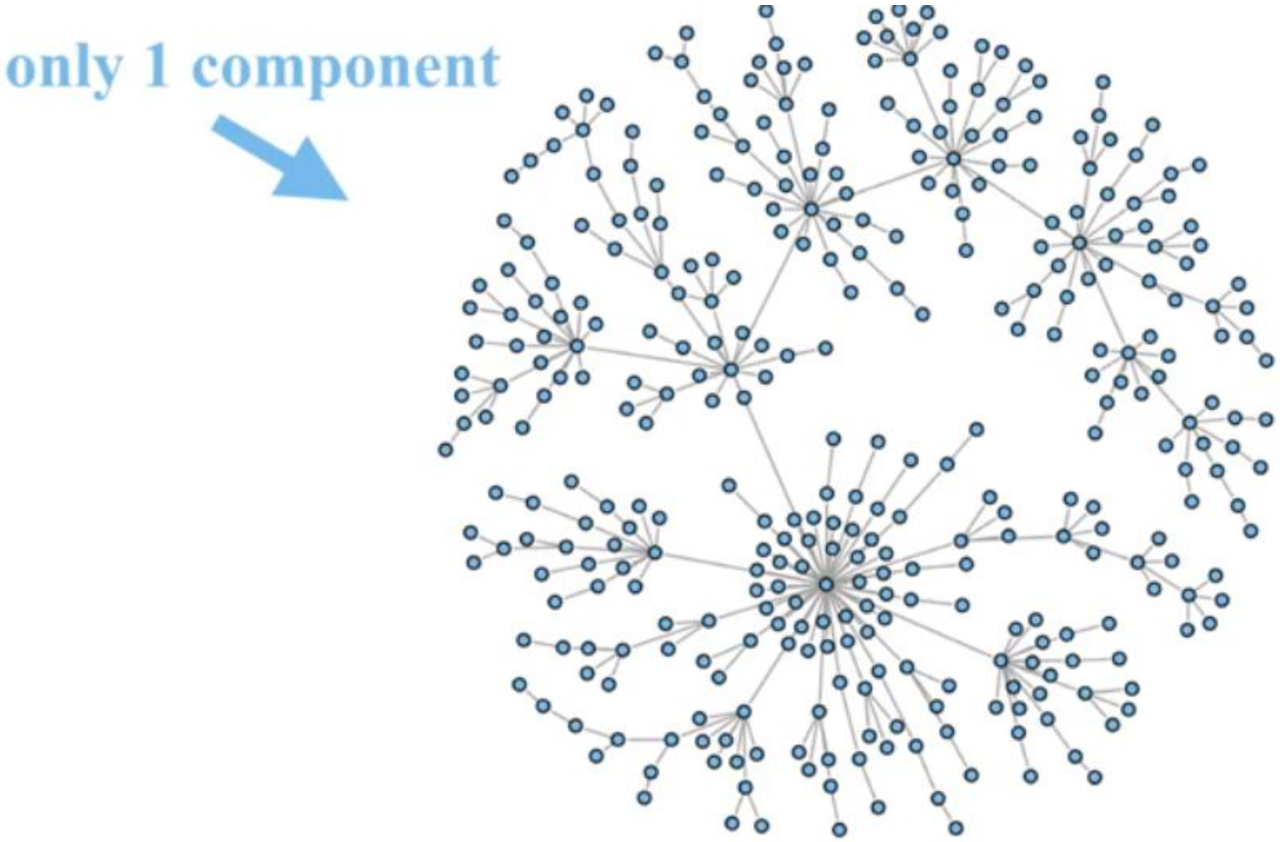
A handy benefit to betweenness centrality

You don’t need a **fully connected** graph or **component** to calculate it (unlike closeness centrality).

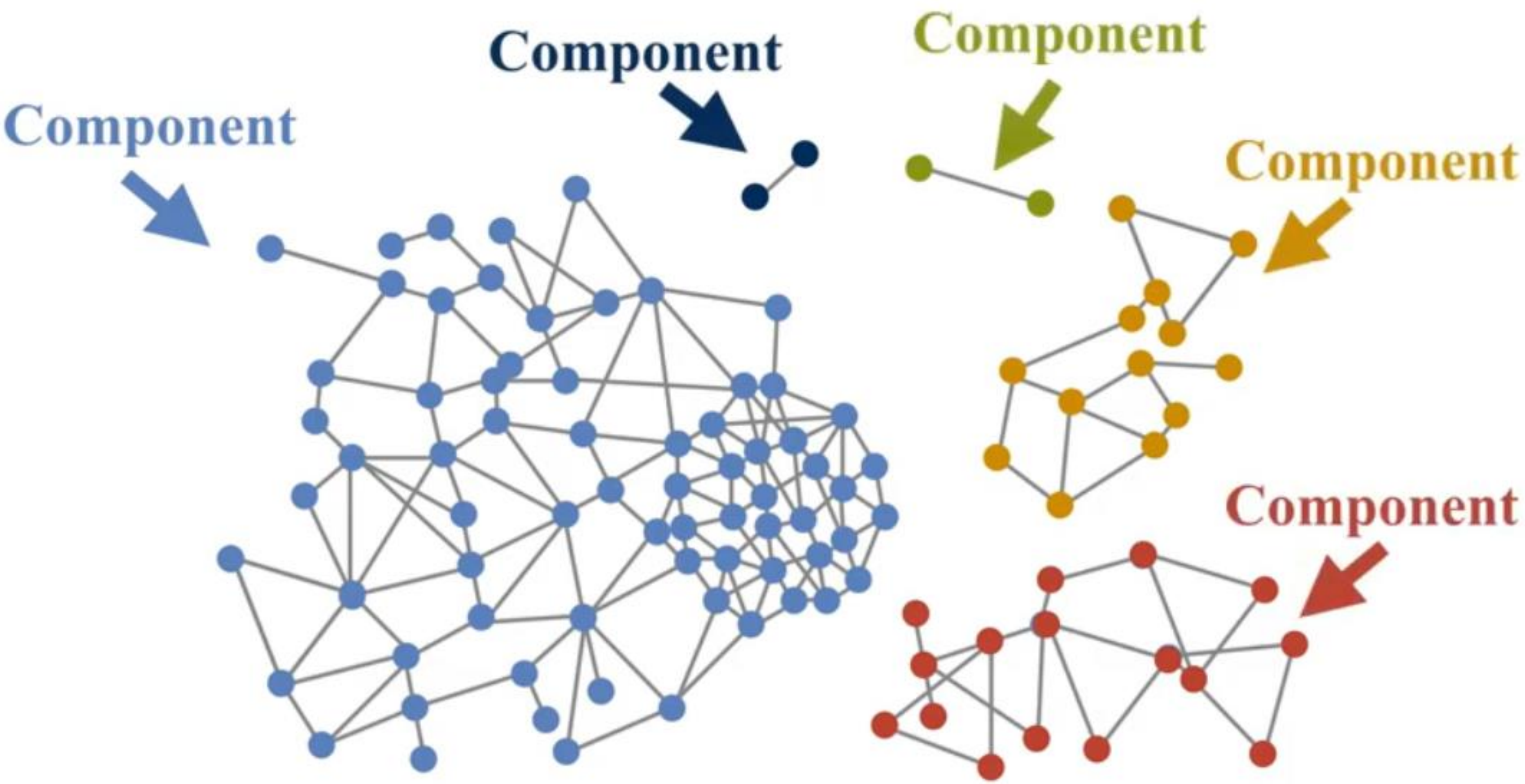
Dr Hussein Hazimeh

Component

» A component is a group of nodes connected to each other.

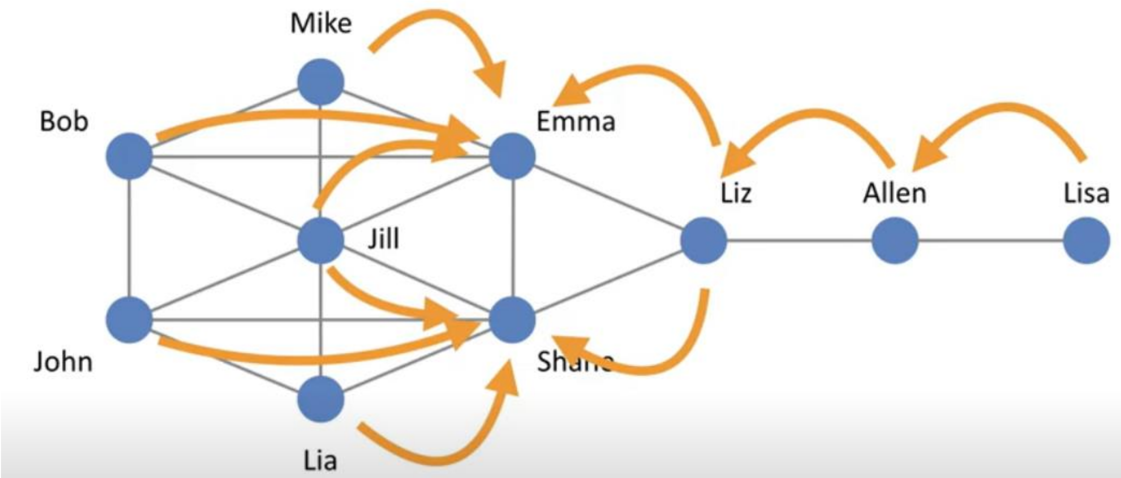


Component (cont.)

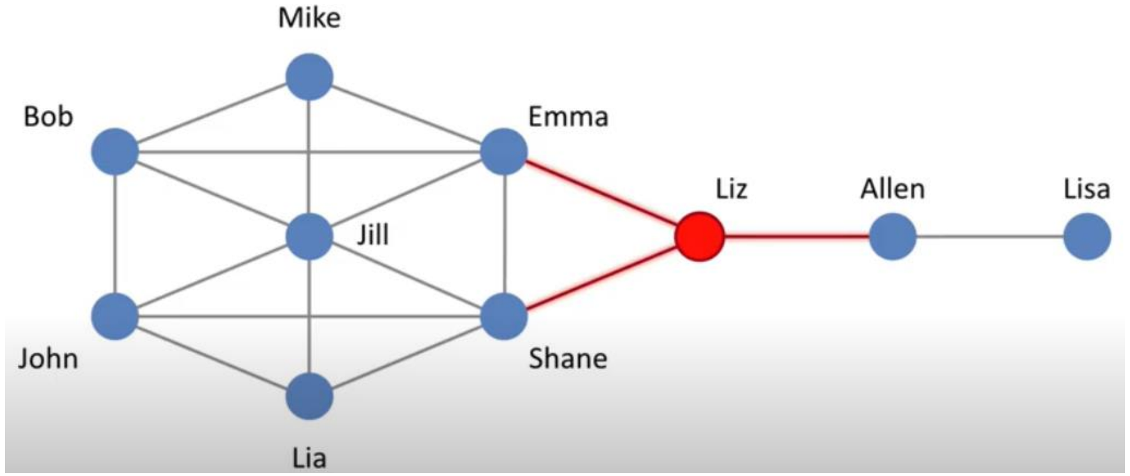


Betweenness Centrality

- Betweenness centrality (4) measures the fraction of shortest paths passing through a vertex.
- Nodes with high betweenness centrality are often important **controllers** of **power** or **information**



Emma & Shane have highest closeness centrality.



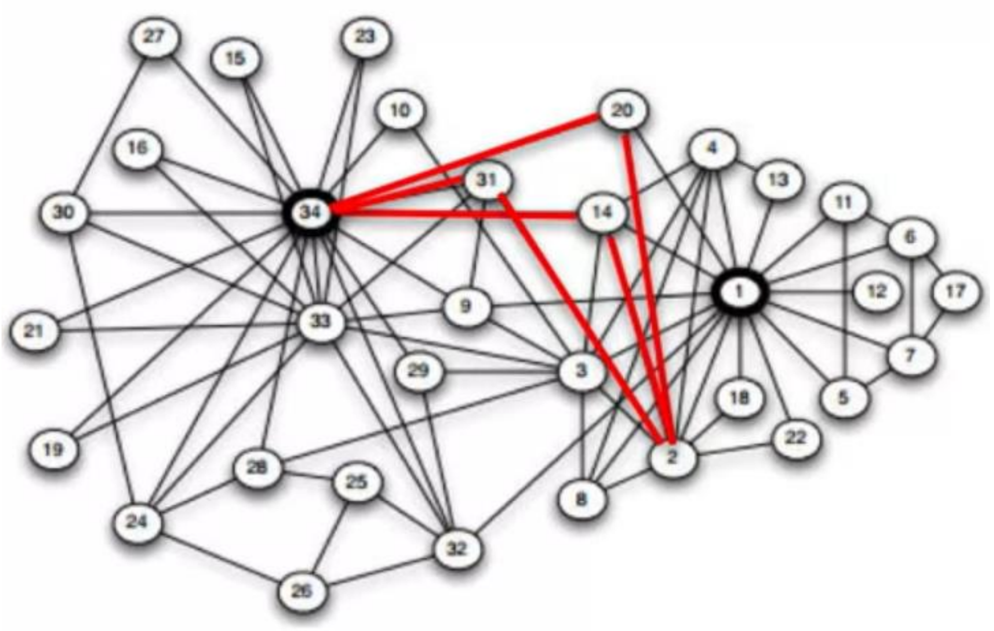
Liz has highest betweenness centrality.

Dr Hussein Hazimeh

Betweenness Centrality (cont.)

- **Assumption:** important nodes connect other nodes.
- **Ex.** the distance between nodes 34 and 2 is 2:
- Path 1: 34-31-2
- Path 2: 34-14-2
- Path 3: 34-20-2

- Nodes 31,14,and 20 are in a shortest path of between nodes 34 and 2.



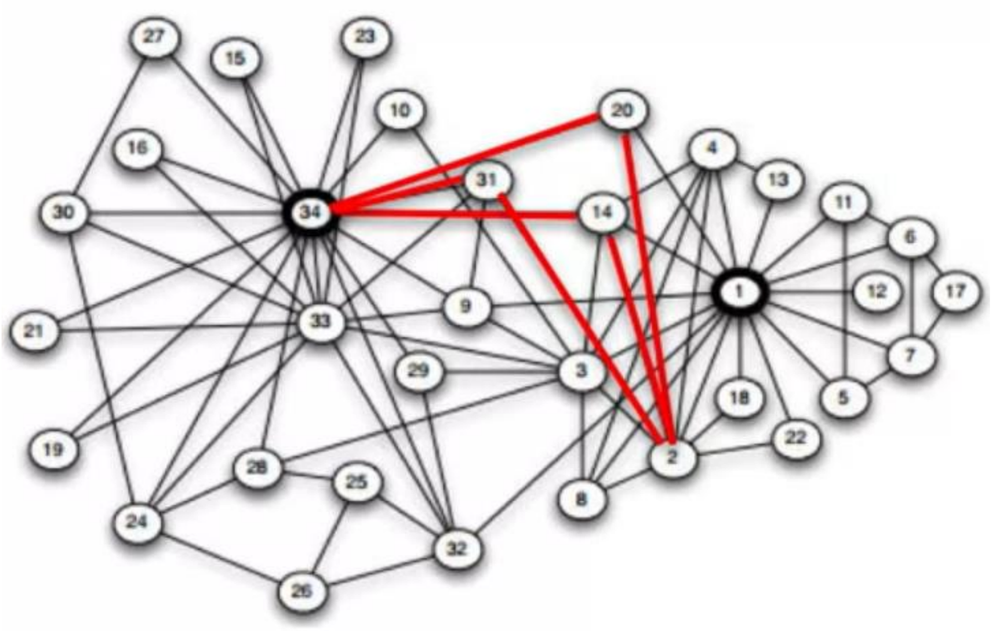
Betweenness Centrality (cont.)

» **Assumption:** important nodes connect other nodes.

$$C_{btw}(v) = \sum_{s,t \in N} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}$$, where

$\sigma_{s,t}$ = the number of shortest paths between nodes s and t.

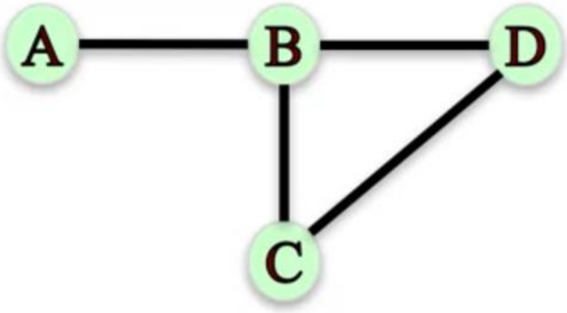
$\sigma_{s,t}(v)$ = the number shortest paths between nodes s and t *that pass through node* v.



Betweenness Centrality (cont.)

Assumption: important nodes connect other nodes.

$$C_{btw}(v) = \sum_{s,t \in N} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}$$



Endpoints: we can either include or exclude node v as node s and t in the computation of $C_{btw}(v)$.

Ex. If we exclude node v, we have:

$$C_{btw}(B) = \frac{\sigma_{A,D}(B)}{\sigma_{A,D}} + \frac{\sigma_{A,C}(B)}{\sigma_{A,C}} + \frac{\sigma_{C,D}(B)}{\sigma_{C,D}} = \frac{1}{1} + \frac{1}{1} + \frac{0}{1} = 2$$

If we include node v, we have:

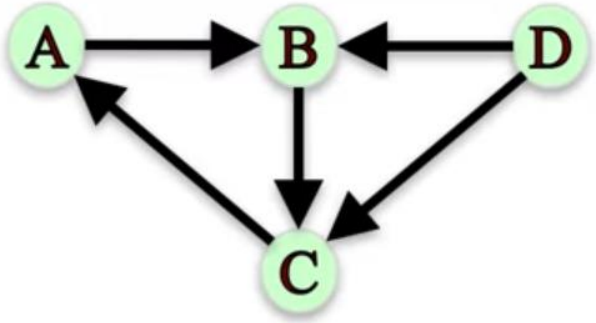
$$C_{btw}(B) = \frac{\sigma_{A,B}(B)}{\sigma_{A,B}} + \frac{\sigma_{A,C}(B)}{\sigma_{A,C}} + \frac{\sigma_{A,D}(B)}{\sigma_{A,D}} + \frac{\sigma_{B,C}(B)}{\sigma_{B,C}} + \frac{\sigma_{B,D}(B)}{\sigma_{B,D}} + \frac{\sigma_{C,D}(B)}{\sigma_{C,D}} = \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{0}{1} = 5$$

Betweenness Centrality (cont.)

» Disconnected nodes

$$C_{btw}(v) = \sum_{s,t \in N} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}$$

What if not all nodes can reach each other?



Node D cannot be reached by any other node.

Hence, $\sigma_{A,D} = 0$, making the above definition undefined.

Ex. What is the betweenness centrality of node B, without including it as endpoint?

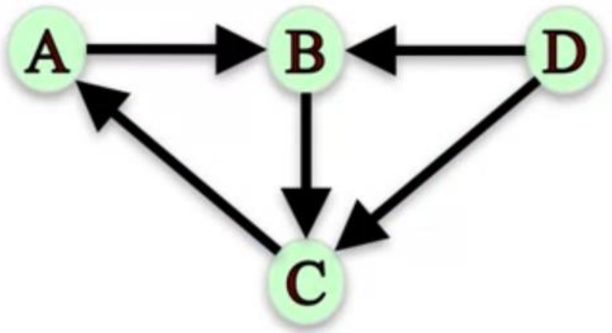
$$C_{btw}(B) = \frac{\sigma_{A,C}(B)}{\sigma_{A,C}} + \frac{\sigma_{C,A}(B)}{\sigma_{C,A}} + \frac{\sigma_{D,C}(B)}{\sigma_{D,C}} + \frac{\sigma_{D,A}(B)}{\sigma_{D,A}} = \frac{1}{1} + \frac{0}{1} + \frac{0}{1} + \frac{0}{1} = 1$$

Betweenness Centrality (cont.)

- Disconnected nodes

$$C_{btw}(v) = \sum_{s,t \in N} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}$$

What if not all nodes can reach each other?



Node D cannot be reached by any other node.

Hence, $\sigma_{A,D} = 0$, making the above definition undefined.

Ex. What is the betweenness centrality of node C, without including it as endpoint?

$$C_{btw}(C) = \frac{\sigma_{A,B}(C)}{\sigma_{A,B}} + \frac{\sigma_{B,A}(C)}{\sigma_{B,A}} + \frac{\sigma_{D,B}(C)}{\sigma_{D,B}} + \frac{\sigma_{D,A}(C)}{\sigma_{D,A}} = \frac{0}{1} + \frac{1}{1} + \frac{0}{1} + \frac{1}{1} = 2$$

Betweenness Centrality - Normalization

- Normalization concept

Normalization: betwenness centrality values will be larger in graphs with many nodes. To control for this, we divide centrality values by the number of pairs of nodes in the graph (excluding v):

$\frac{1}{2}(|N|-1)(|N|-2)$ in undirected graphs

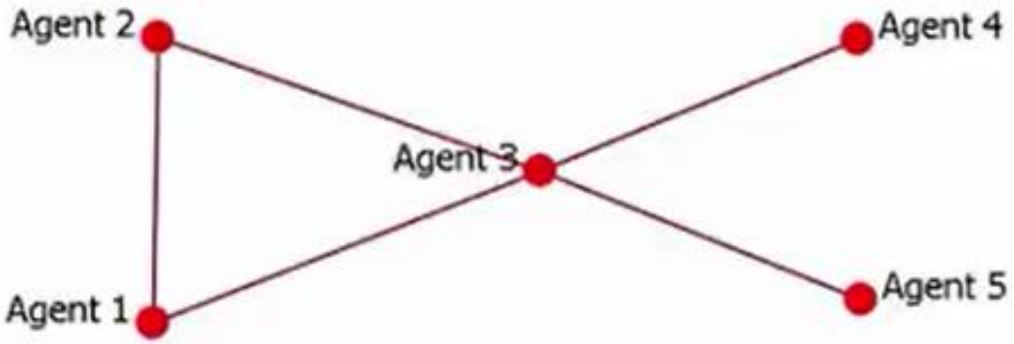
$(|N|-1)(|N|-2)$ in directed graphs

Betweenness Centrality

$$nC_2 = \frac{n(n-1)}{2}$$

$$\frac{5(5-1)}{2} = \frac{5(4)}{2} = \frac{20}{2} = 10$$

paths possible



Agent 1

From	To
1	1
1	2
1	3
1	4
1	5

Agent 2

From	To
2	1
2	2
2	3
2	4
2	5

Agent 3

From	To
3	1
3	2
3	3
3	4
3	5

Agent 4

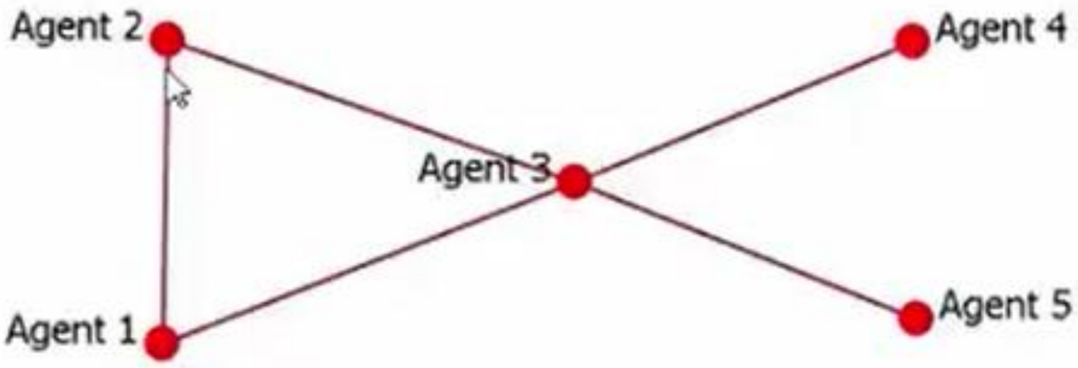
From	To
4	1
4	2
4	3
4	4
4	5

Agent 5

From	To
5	1
5	2
5	3
5	4
5	5

Betweenness Calculation

From	To	Geodesic
1	2	(1,2)
1	3	(1,3)
1	4	(1,3,4)
1	5	(1,3,5)
2	3	(2,3)
2	4	(2,3,4)
2	5	(2,3,5)
3	4	(3,4)
3	5	
4	5	



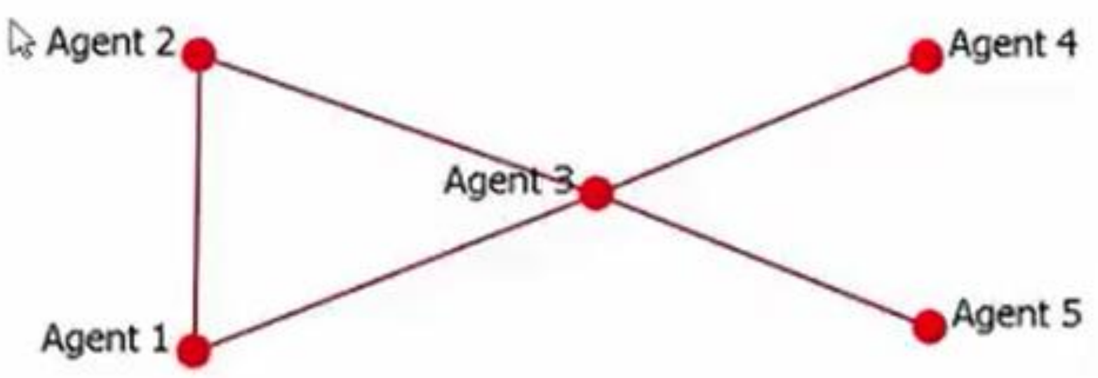
Betweenness Calculation (cont.)

From	To	Geodesic	1	2	3	4	5
1	2	(1,2)	0	0	0	0	0
1	3	(1,3)	0	0	0	0	0
1	4	(1,3,4)	0	0	1	0	0
1	5	(1,3,5)	0	0	1	0	0
2	3	(2,3)	0	0	0	0	0
2	4	(2,3,4)	0	0	1	0	0
2	5	(2,3,5)	0	0	1	0	0
3	4	(3,4)	0	0	0	0	0
3	5	(3,5)	0	0	0	0	0
4	5	(4,3,5)	0	0	1	0	0
		Sum	0	0	5	0	0

Betweenness Calculation (cont.)

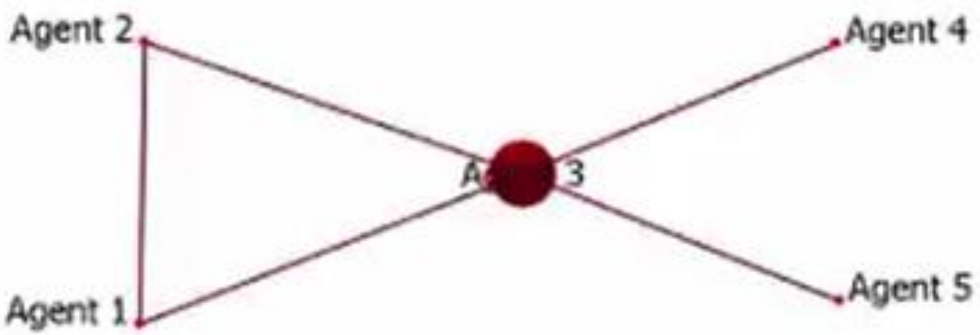
$$\frac{(n-1)(n-2)}{2} = \text{Denominator}$$

$$\frac{(5-1)(5-2)}{2} = \frac{(4)(3)}{2} = 6$$



0	0	5	0	0	Numerator
6	6	6	6	6	Denominator

0/6	0/6	5/6	0/6	0/6	Betweenness



3

Eigenvector Centrality

1. Eigenvector calculation
2. Eigenvector application on SNA

1. Eigenvector calculation

Eigenvector centrality

» Eigenvector is a word which is derived from the German word "eigen," which means self, and the word "vector" which means vector.

» We'll call the importance of each node its *centrality score*, and to measure this we'll want the centrality score to be proportional to the sum of the scores of all nodes which are connected to it. This way, if a node is connected to many important node, it will also be an important node, and if it is connected to only a few unimportant nodes then it won't be important.

Calculation process

- How to find the eigenvector and eigenvalues for n x n matrix
- to solve the eigenvalues, λi, and the corresponding eigenvectors, $\bar{x}_i$ of an n x n matrix A, do the following:

1. Multiply an n x n identity matrix by the scalar λ.
2. Subtract the identity matrix multiple from the matrix A.
3. Find the determinant of the matrix and the difference.
4. Solve for the values of λ that satisfy the equation det(A- λI) = $\bar{0}$
5. Solve the corresponding vector to each λ.

Calculation process (cont.)

» Find th eigenvectors of the matrix:

» A= $\begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix}$

» **1.** $\lambda I = \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$

» **2.** $A - \lambda I = \begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 7-\lambda & 3 \\ 3 & -1-\lambda \end{bmatrix}$

» **3.** $\det \begin{bmatrix} 7-\lambda & 3 \\ 3 & -1-\lambda \end{bmatrix} = (7-\lambda)(-1-\lambda)-(3)(3) = -7- 7\lambda + \lambda + \lambda^2 - 9 = \lambda^2 - 6\lambda - 16$

» **4.** $\lambda^2 - 6\lambda - 16 = 0$

» $(\lambda - 8)(\lambda + 2) = 0$

» **λ = 8** and **λ = -2** ⇐ **Eigenvalues**

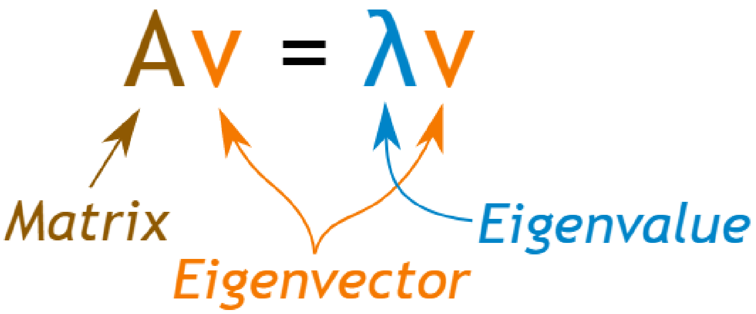
Calculation process (cont.)

- **5.** $\begin{bmatrix} 7-\lambda & 3 \\ 3 & -1-\lambda \end{bmatrix}$
- For λ = 8:
- $\begin{bmatrix} 7-8 & 3 \\ 3 & -1-8 \end{bmatrix} = \begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix}$ ⇐ B
- Solve: $B\bar{x} = \bar{0}$
- $\begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix} \begin{bmatrix} x1 \\ x2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
- x_1 x_2
- $\begin{bmatrix} -1 & 3 & |0| \\ 3 & -9 & |0| \end{bmatrix} \rightarrow \begin{bmatrix} -1 & 3 & |0| \\ 0 & 0 & |0| \end{bmatrix}$
- 3R1 + R2 → R2
- $-x_1 + 3x_2 = 0$
- $3x_2 = x_1$ → let $x_2 = 1$ ➔ $x_1 = 3$

- For eigenvalue **λ = 8**:
- Eigenvector = $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$

Calculation process (cont.)

- Check if your eigenvector is correct:
- $\begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 8 \begin{bmatrix} 3 \\ 1 \end{bmatrix}$
- $\begin{bmatrix} 24 \\ 8 \end{bmatrix} \stackrel{?}{=} \begin{bmatrix} 24 \\ 8 \end{bmatrix}$ ✓



Example: For this matrix

$$\begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix}$$

an eigenvector is:

$$\begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

with a matching eigenvalue of 6

Let's do some matrix multiplies to see what we get.

Av gives us:

$$\begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} -6\times1+3\times4 \\ 4\times1+5\times4 \end{bmatrix} = \begin{bmatrix} 6 \\ 24 \end{bmatrix}$$

λv gives us :

$$6 \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 6 \\ 24 \end{bmatrix}$$

Yes they are equal! So Av = λv as promised.

Calculation process (cont.)

- For λ = -2:
- $\begin{bmatrix} 7-(-2) & 3 \\ 3 & -1-(-2) \end{bmatrix} = \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix}$
- **x_1 x_2**
- $\left[\begin{matrix} 9 & 3 \\ 3 & 1 \end{matrix} \middle| \begin{matrix} 0 \\ 0 \end{matrix}\right] \rightarrow \left[\begin{matrix} 0 & 0 \\ 3 & 1 \end{matrix} \middle| \begin{matrix} 0 \\ 0 \end{matrix}\right]$
- -3R2 + R1 → R1
- $3x_1 + x_2 = 0$
- $x_2 = -3x_1$ → let **x_1** = 1 ➔ **x_2** = -3

- For eigenvalue **λ = -2**:
- Eigenvector = $\begin{bmatrix} \mathbf{1} \\ \mathbf{-3} \end{bmatrix}$

Eigenvector applications

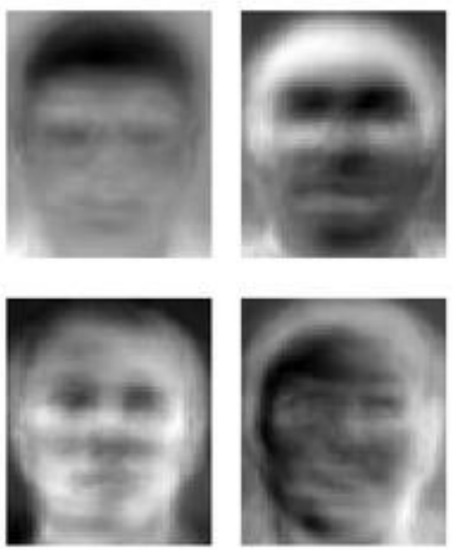
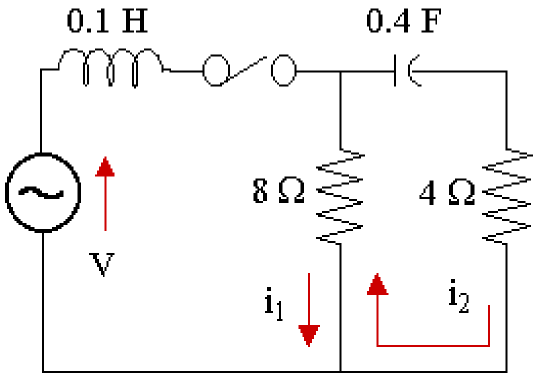
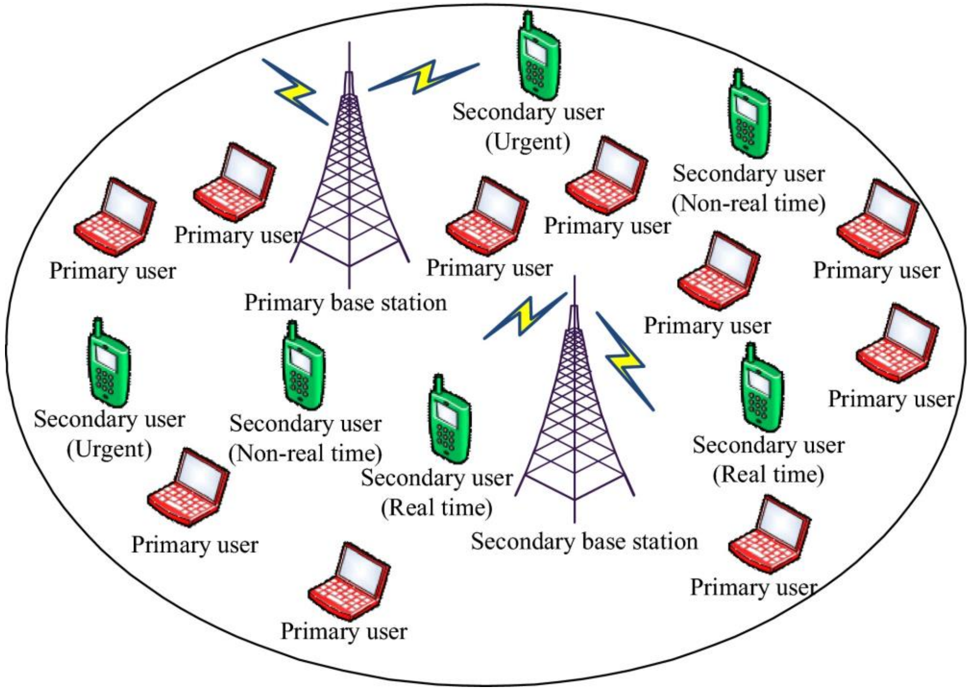


Image processing



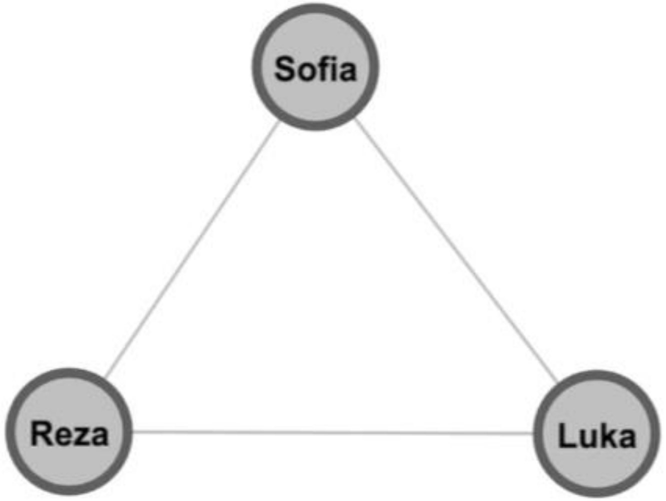
Electronics



Telecommunications

2. Eigenvector application on SNA

Eigenvector application on SNA

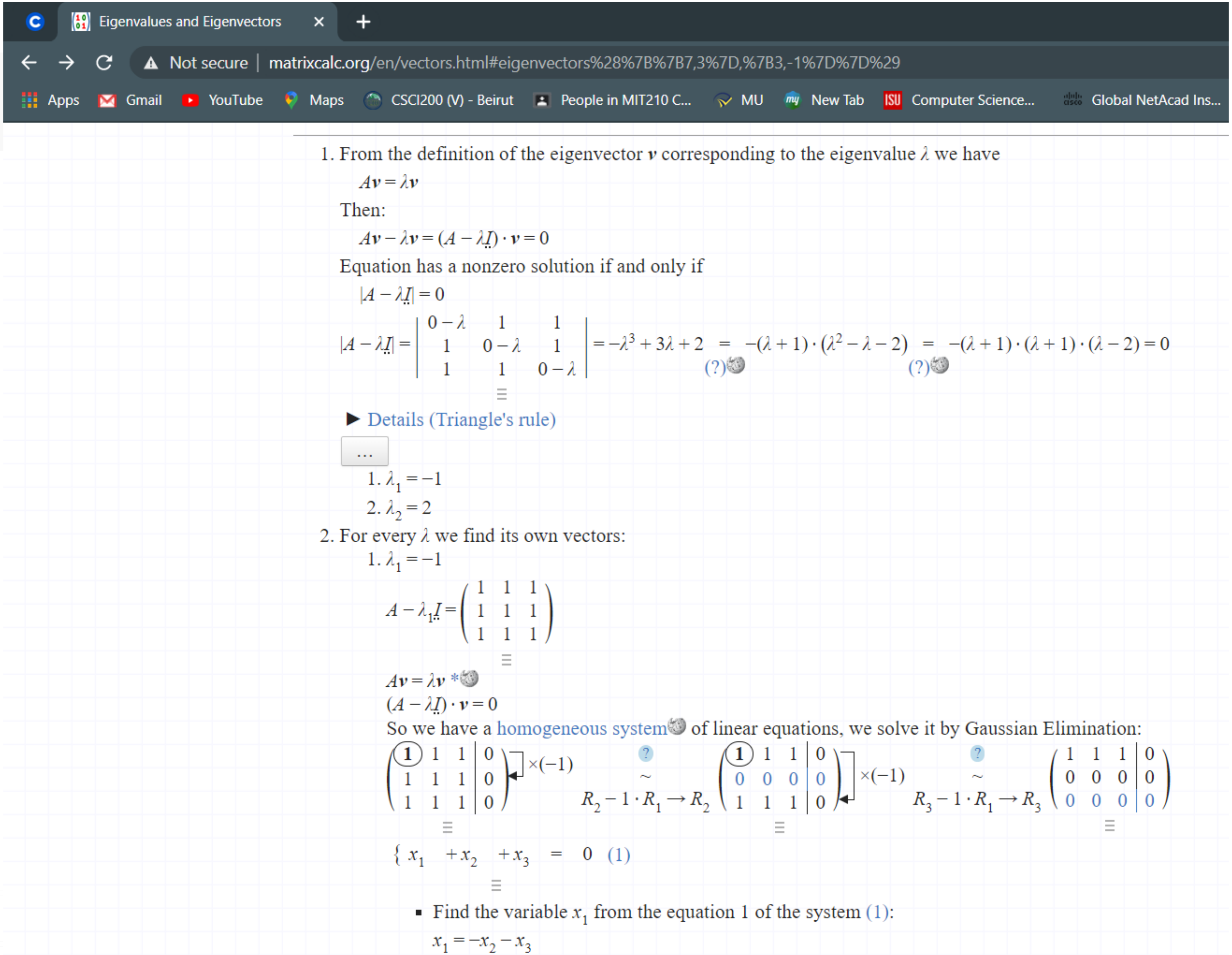


		From		
		Reza	Sofia	Luka
To	Reza	0	1	1
	Sofia	1	0	1
	Luka	1	1	0

What about large networks?

Eigenvector application on SNA

- » http://matrixcalc.org/en/vectors.html
- » The above website is an online tool that helps you to calculate and find the eigenvalues and eigenvectors of a matrix.



1. From the definition of the eigenvector v corresponding to the eigenvalue λ we have

$Av = \lambda v$

Then:

$Av - \lambda v = (A - \lambda I) \cdot v = 0$

Equation has a nonzero solution if and only if

$|A - \lambda I| = 0$

$$|A - \lambda I| = \begin{vmatrix} 0-\lambda & 1 & 1 \\ 1 & 0-\lambda & 1 \\ 1 & 1 & 0-\lambda \end{vmatrix} = -\lambda^3 + 3\lambda + 2 \overset{(?)}{=} -(\lambda+1)\cdot(\lambda^2 - \lambda - 2) \overset{(?)}{=} -(\lambda+1)\cdot(\lambda+1)\cdot(\lambda-2) = 0$$

▶ Details (Triangle's rule)

...

1. $\lambda_1 = -1$
2. $\lambda_2 = 2$

2. For every λ we find its own vectors:

1. $\lambda_1 = -1$

$$A - \lambda_1 I = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$Av = \lambda v$ *

$(A - \lambda I) \cdot v = 0$

So we have a homogeneous system of linear equations, we solve it by Gaussian Elimination:

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{array}\right) \times(-1) \underset{R_2 - 1 \cdot R_1 \to R_2}{\sim} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{array}\right) \times(-1) \underset{R_3 - 1 \cdot R_1 \to R_3}{\sim} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

$\{ x_1 + x_2 + x_3 = 0 \quad (1)$

- Find the variable x_1 from the equation 1 of the system (1):

$x_1 = -x_2 - x_3$