SOCIAL NETWORK ANALYSIS METRICS

LEARNING OBJECTIVES

- Understand how graph theory can be applied to social network analysis
- Understand the meanings of essential metrics used for the analysis
- Be able to analyse a social network

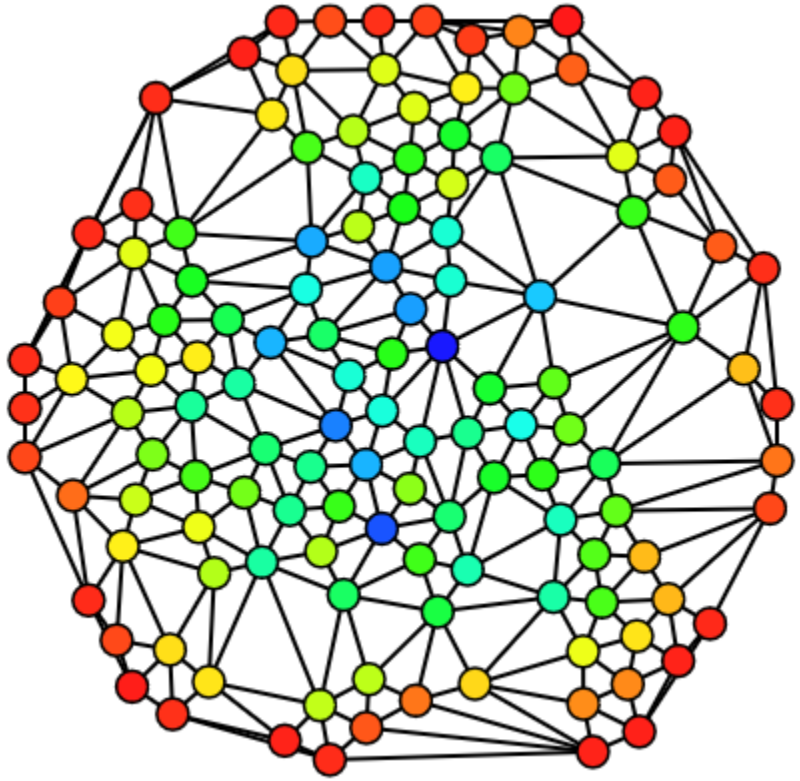
PART 1 – MOTIVATION

A NETWORK

- We can construct a network out of anything
- But a network may not mean much without metrics of analysis
  - Does this network tell you anything?

WHAT ARE THE QUESTIONS?

- Study mathematical properties of the network
  - Based on graph theory
- Questions:
  - How many friends have you got?
  - How close are you to everyone else?
  - How often do you connect with other people?
  - Are your friends acquainted with one another?



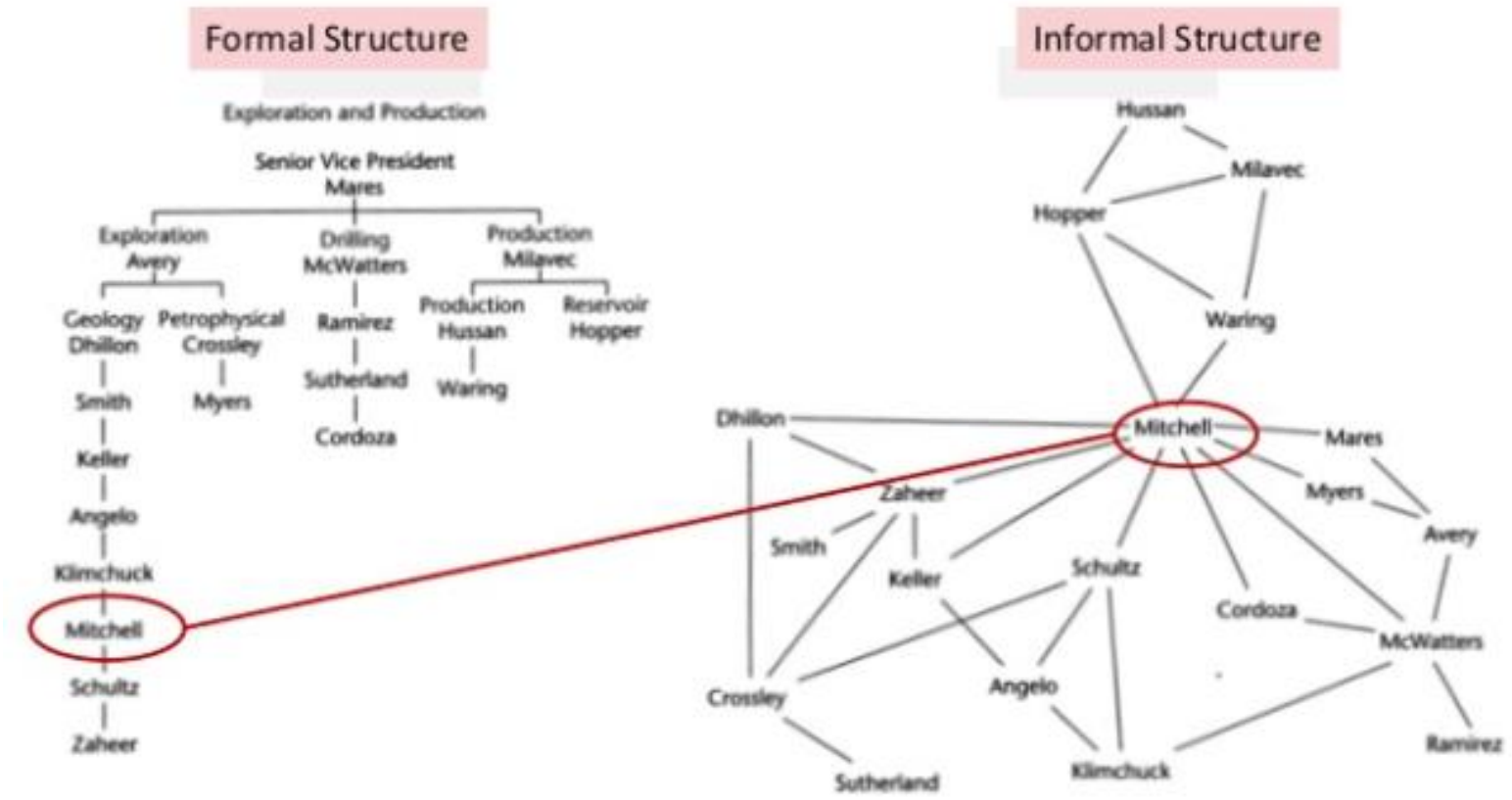
EXAMPLE - ORGANISATION

- Question:
  - Where are the most information flows?

From: *The Organizational Network Fieldbook*, Rob Cross et al, Jossey-Bass 2010

EXAMPLE - ORGANISATION

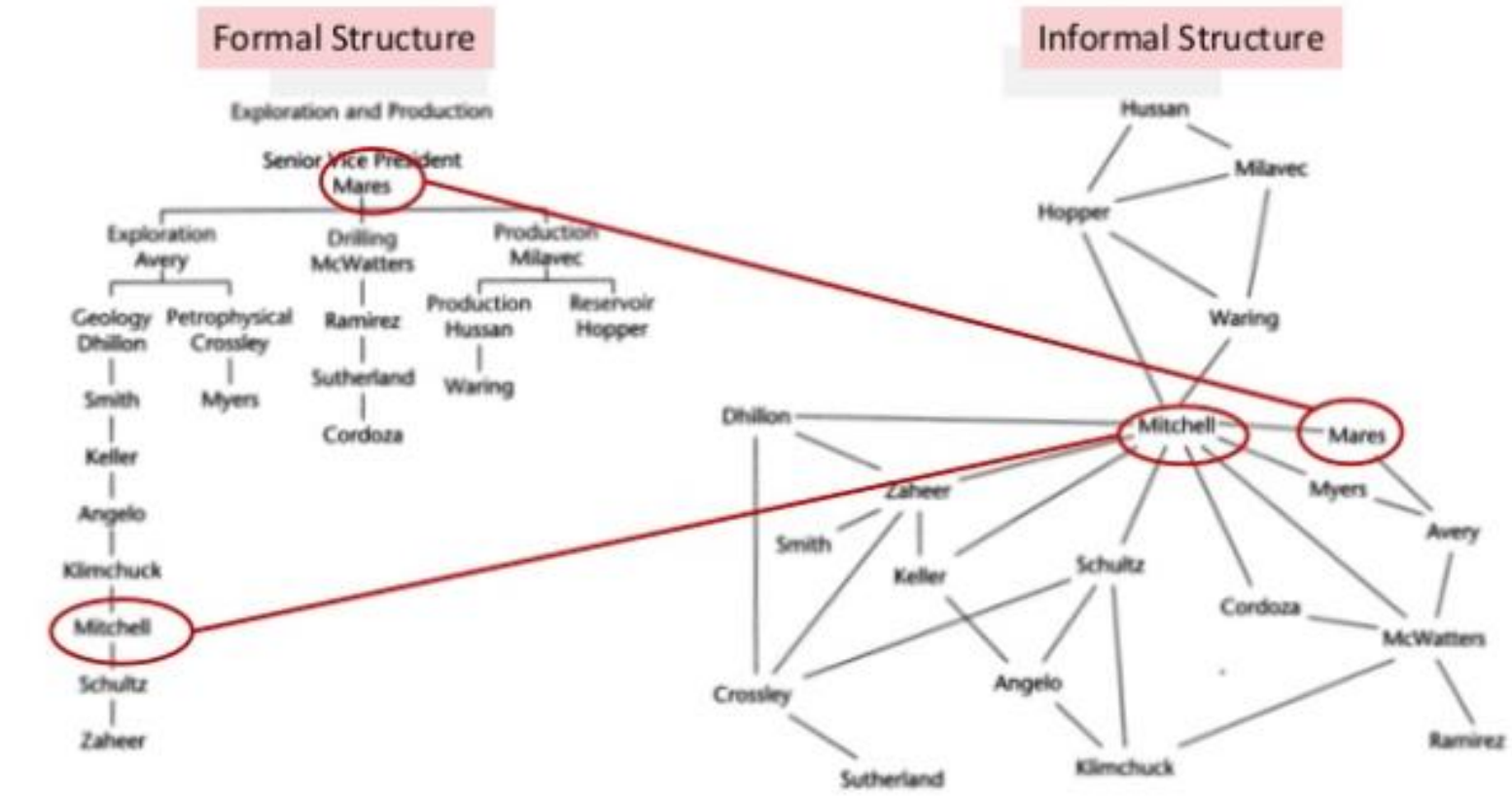
- Question:
  - Where are the most information flows?



rom: *The Organizational Network Fieldbook*, Rob Cross et al, Jossey-Bass 2010

EXAMPLE - ORGANISATION

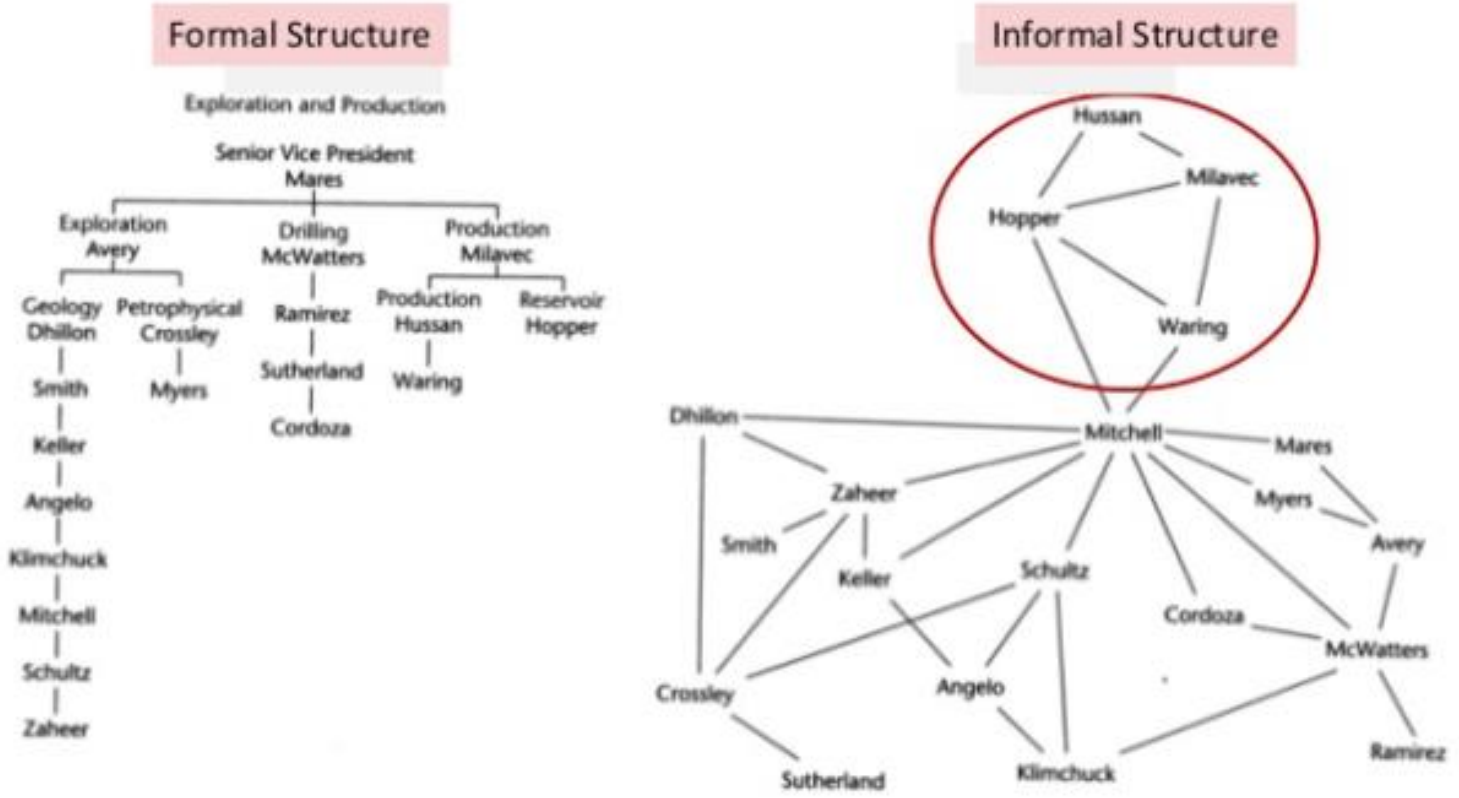
- Question:
  - Where are the most information flows?



From: *The Organizational Network Fieldbook*, Rob Cross et al, Jossey-Bass 2010

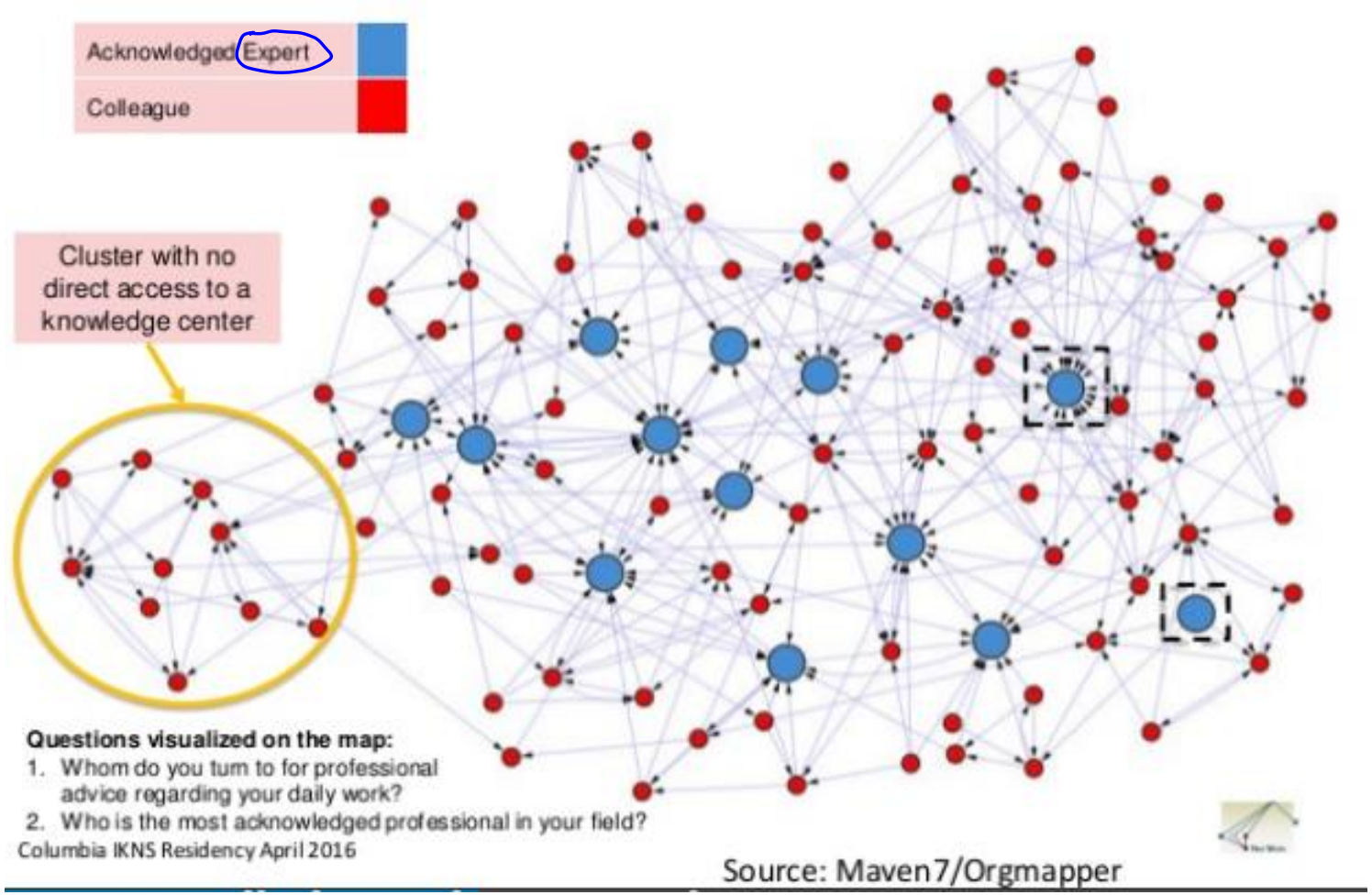
EXAMPLE - ORGANISATION

- Question:
  - Where are the most information flows?

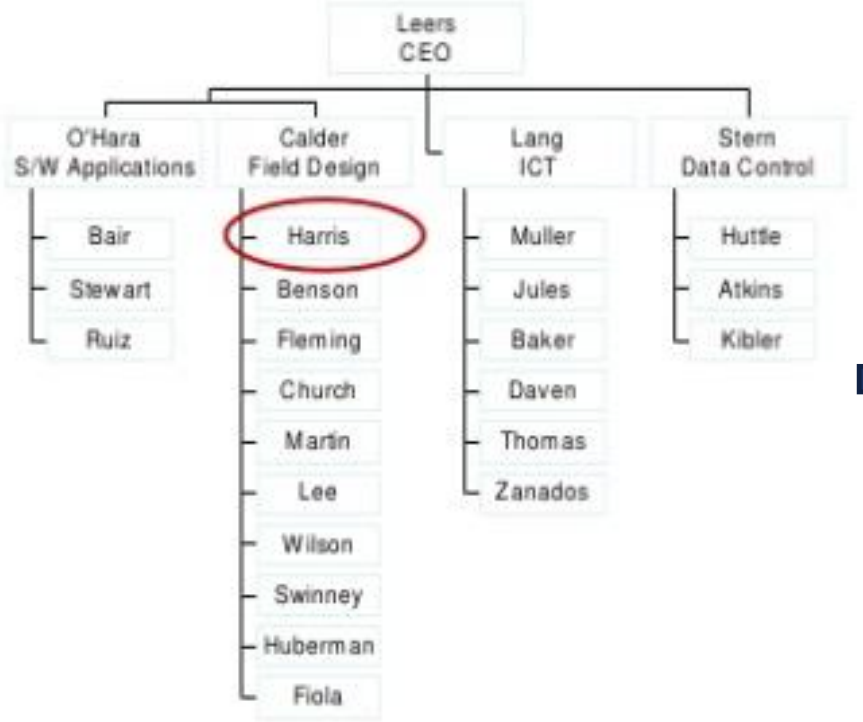


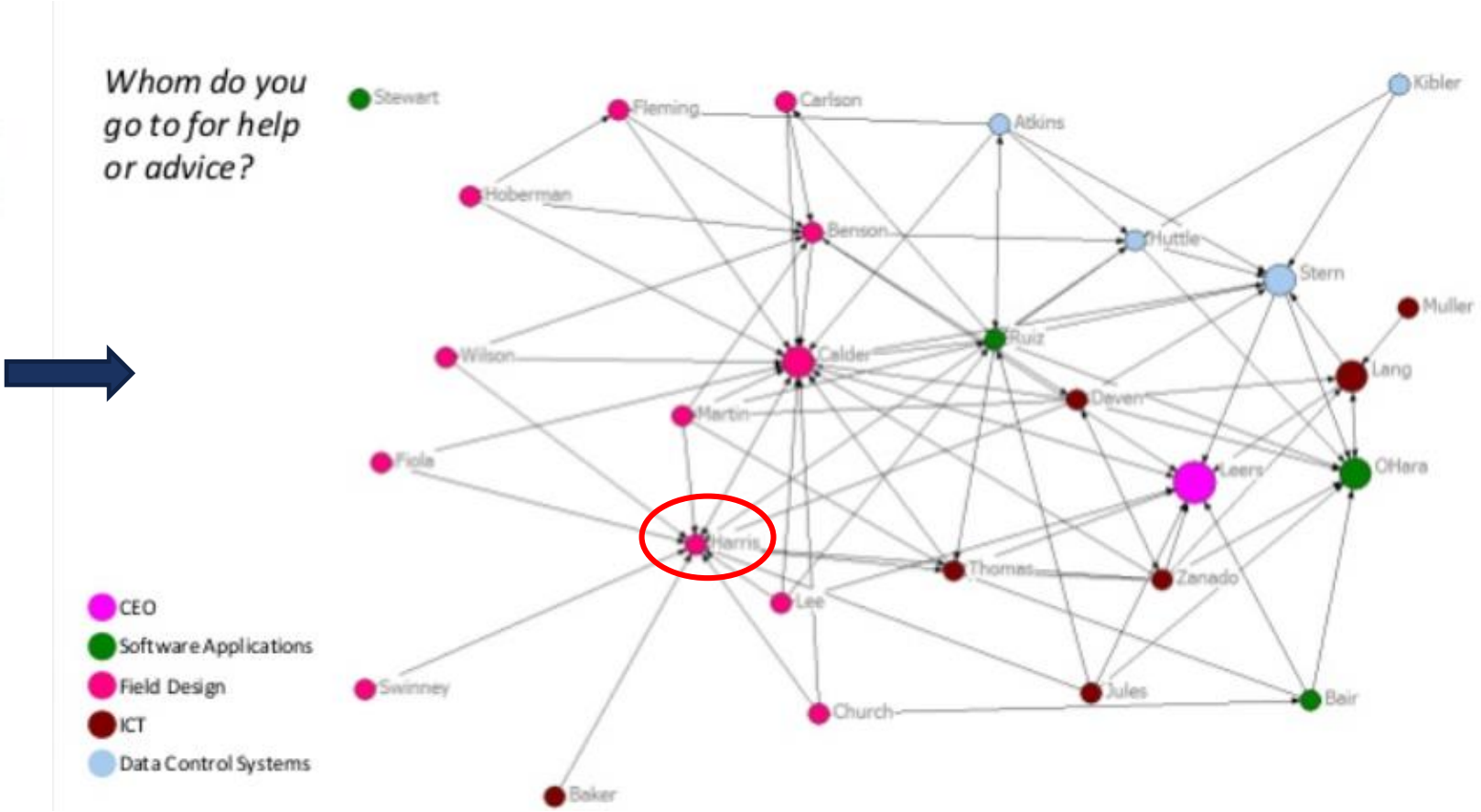
From: *The Organizational Network Fieldbook*, Rob Cross et al, Jossey-Bass 2010

EXAMPLE - RESEARCHERS

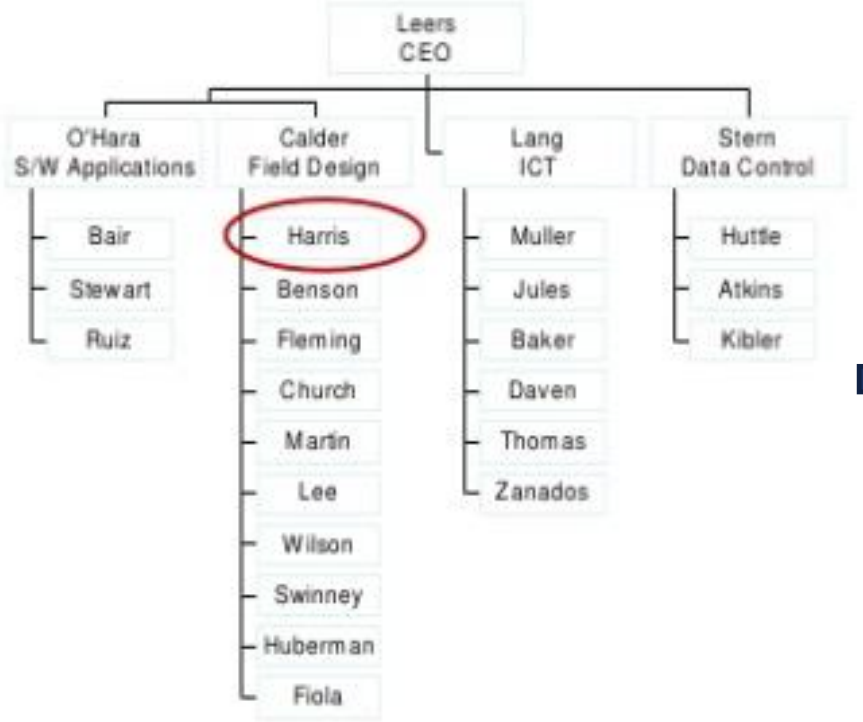


EXAMPLE – CHOOSING A TEAM LEADER

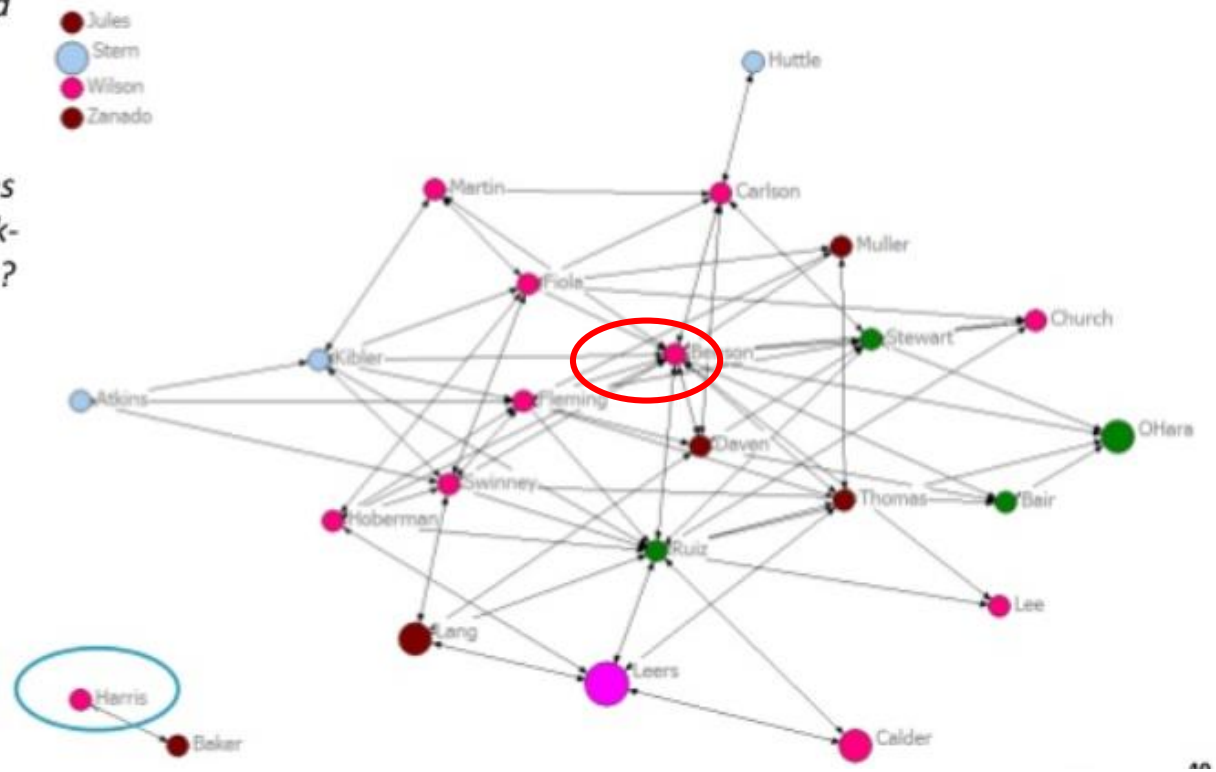




EXAMPLE – CHOOSING A TEAM LEADER



Whom would you trust to keep in confidence your concerns about a work-related issue?



SOCIAL NETWORK ANALYSIS METRICS

มาตร

SNA

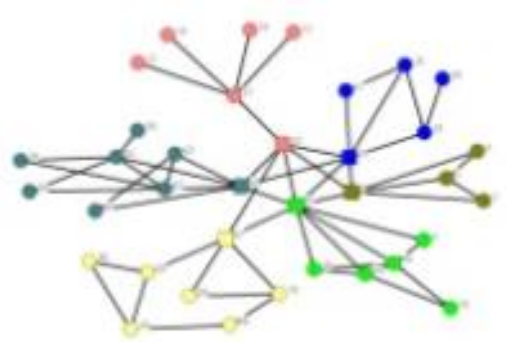
Graph theory

statistic

ของคุณสมบัติ

1 Structural (Network) Metrics

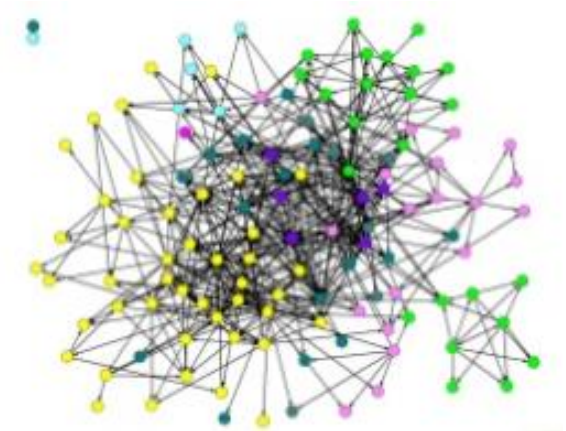
- Look at the whole network and its components



2 Nodal Metrics

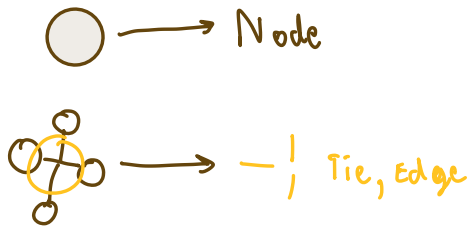
Node

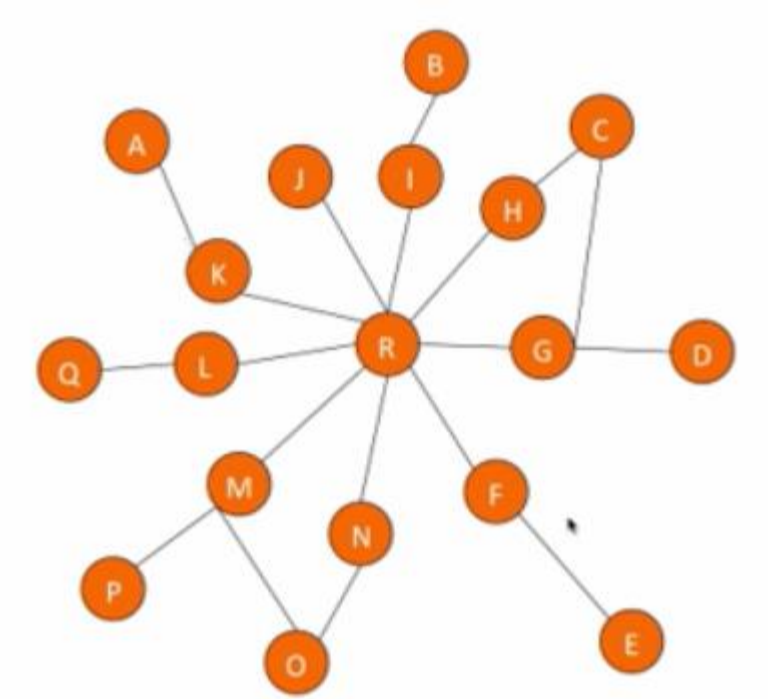
- Look at the positions of individuals in the network



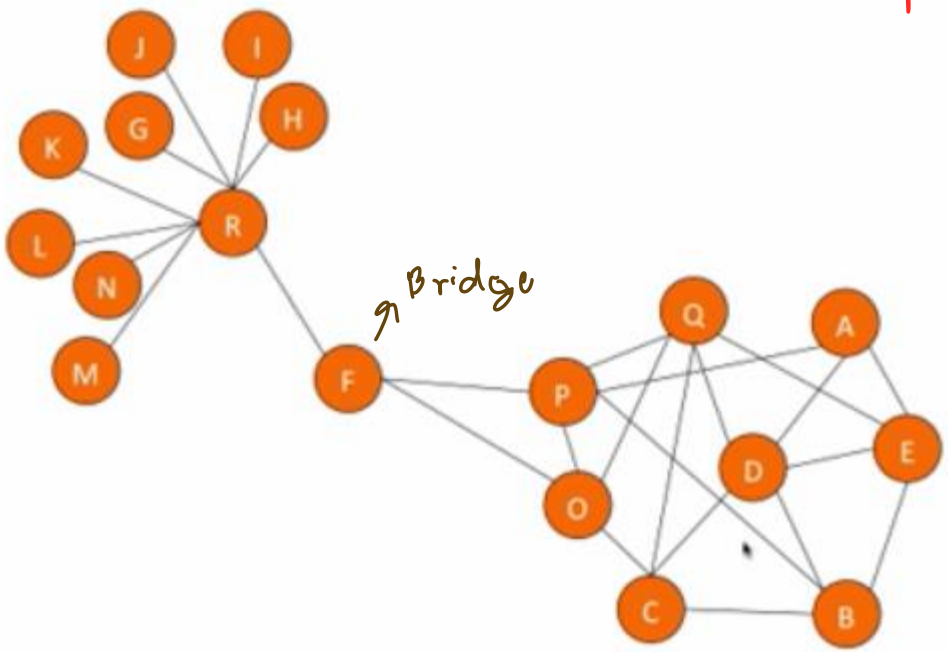
WHICH NODE IS THE MOST IMPORTANT?

หา Node ที่สำคัญที่สุด





WHICH NODE IS THE MOST IMPORTANT?



Graph theory → Centrality Measures : Metrics.

Numbers → Compare

Nodal Metrics

PART 2 – DEGREE CENTRALITY

the number of edges (connections)

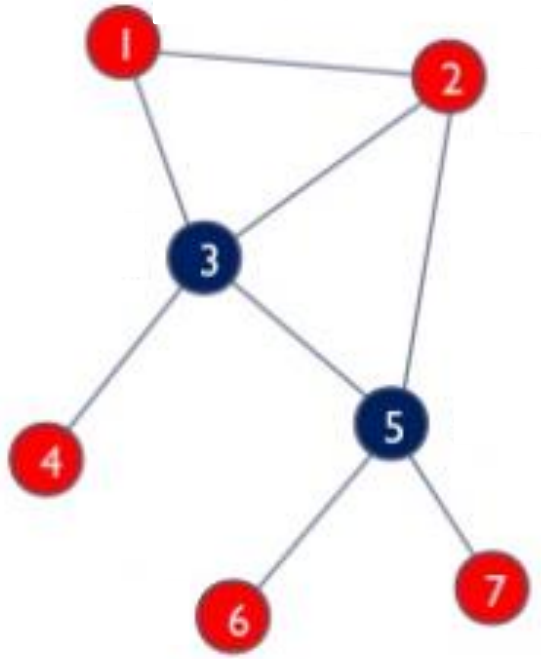
จำนวนของความสัมพันธ์ (connection)

edges / connections

NODAL METRICS – DEGREE FOR AN UNDIRECTED GRAPH

Metrics ที่บอกคุณลักษณะของแต่ละ Node

- We count the number of edges that are connected to a node
- What is the degree for each node in the network?



This is an undirected graph.

นับจำนวน Edges

list ออกมาก่อน →

Degree (1) = 2

Degree (2) : 3

Degree (3) : 4

Degree (4) : 1

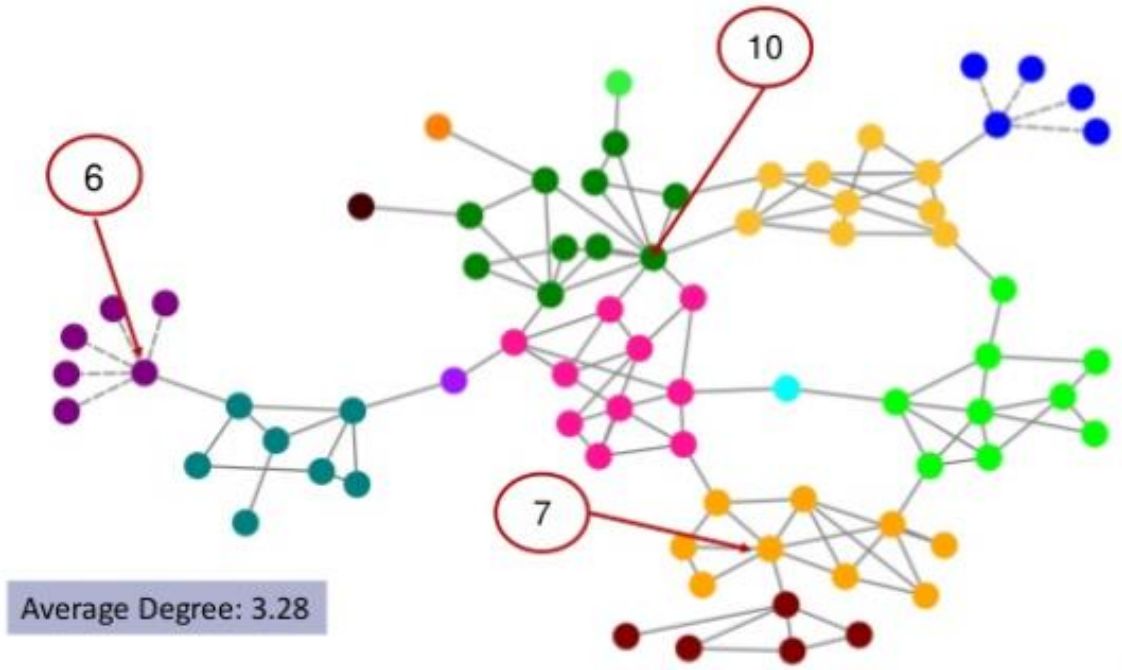
Degree (5) : 4

Degree (6) : 1

Degree (7) : 1

NODAL METRICS – DEGREE FOR AN UNDIRECTED GRAPH

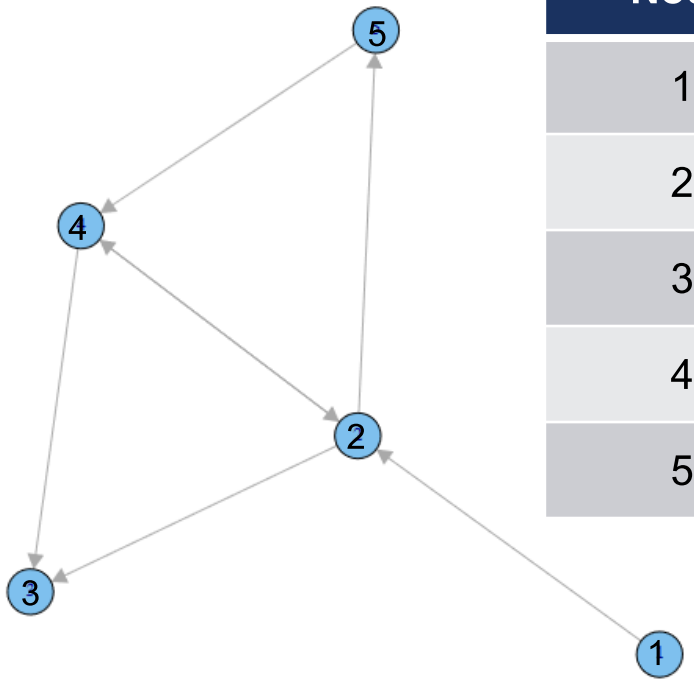
- Average Degree = 2L/N
  - L is the number of edges
  - N is the number of nodes



A node that has a degree larger than the average degree is called a hub.

NODAL METRICS – DEGREE FOR A DIRECTED GRAPH

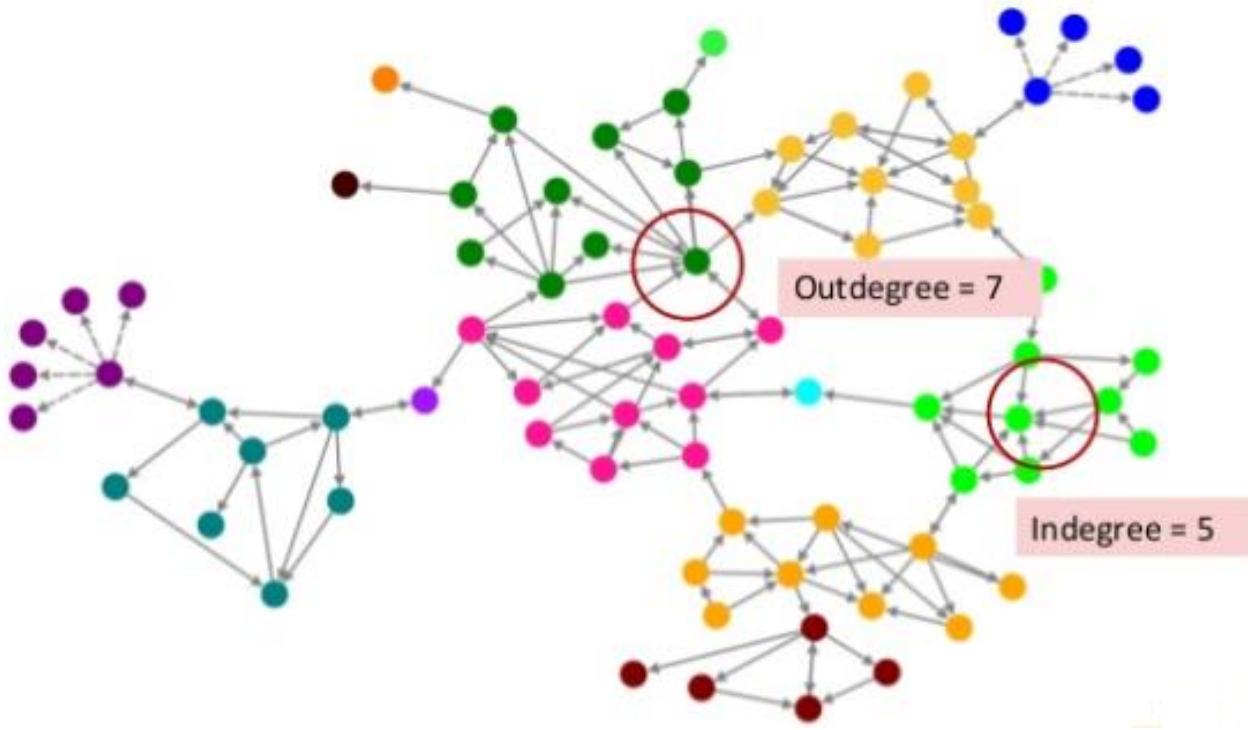
- Difference variance of degree
  - In-degree and Out-degree
- What is the in-degree for each node?
- What is the out-degree for each node?



Node	In-Degree	Out-Degree
1	0	1
2	2	3
3	2	0
4	2	2
5	1	1

NODAL METRICS – DEGREE FOR A DIRECTED GRAPH

- Average Degree = L/N
  - L is the number of edges
  - N is the number of nodes



WHY DO WE CARE ABOUT DEGREE?

- Degree is interesting for several reasons.
  - The ones who have connections to many others might have more influence or more access to information
  - The degree is the immediate risk of a node for catching whatever is flowing through the network (such as the coronavirus)

NODAL METRICS – DEGREE FOR AN UNDIRECTED GRAPH

Scaling Method : 0-1 สามารถเปรียบเทียบ Metrics ที่อยู่ใน scal เดียวกันได้

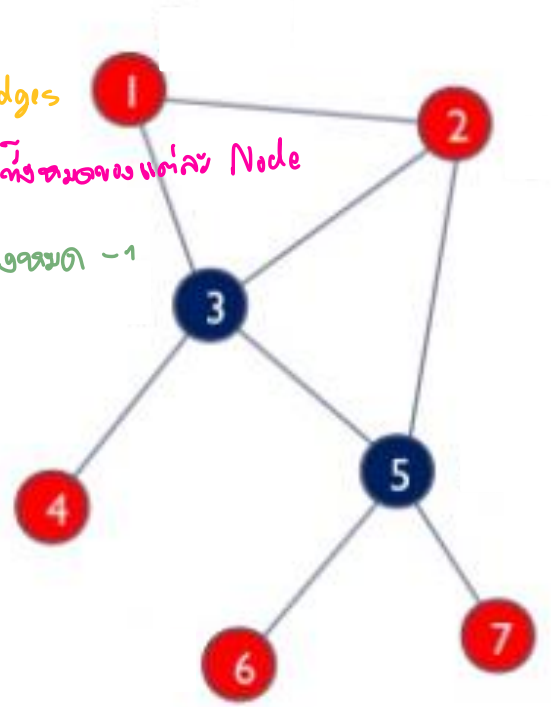
- Another method for computing degree centrality

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

$\sum_{j=1}^{n} a_{ij}$ จำนวน all edges

จำนวน edges ที่เชื่อมของแต่ละ Node

$n-1$ จำนวน Node ทั้งหมด -1

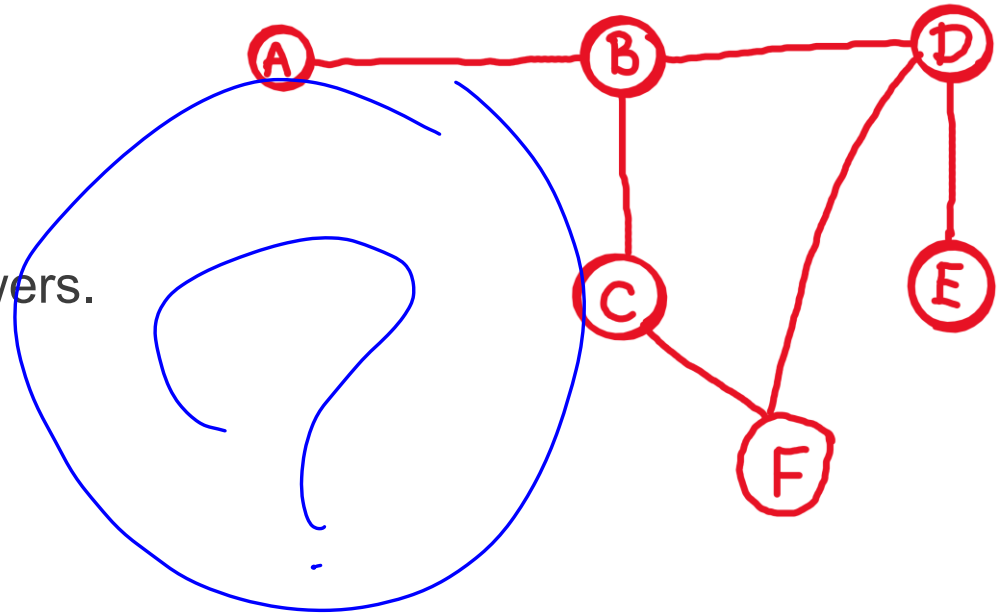


	1	2	3	4	5	6	7	Sum	N-1	Degree	Degree
1	-	1	1	0	0	0	0	2	6	2/6	0.34
2	1	-	1	0	1	0	0	3	6	3/6	0.50
3	1	1	-	1	1	0	0	4	6	4/6	0.67
4	0	0	1	-	0	0	0	1	6	1/6	0.17
5	0	1	1	0	-	1	1	4	6	4/6	0.67
6	0	0	0	0	1	-	0	1	6	1/6	0.17
7	0	0	0	0	1	0	-	1	6	1/6	0.17

Do this for scaling purpose.

DEGREE CENTRALITY EXERCISE

- Determine the degree centrality for all nodes in the following social graph using
  - The non-scaling method
  - The scaling method
- Please complete the exercise.
  - You will be given marks for the correct answers.



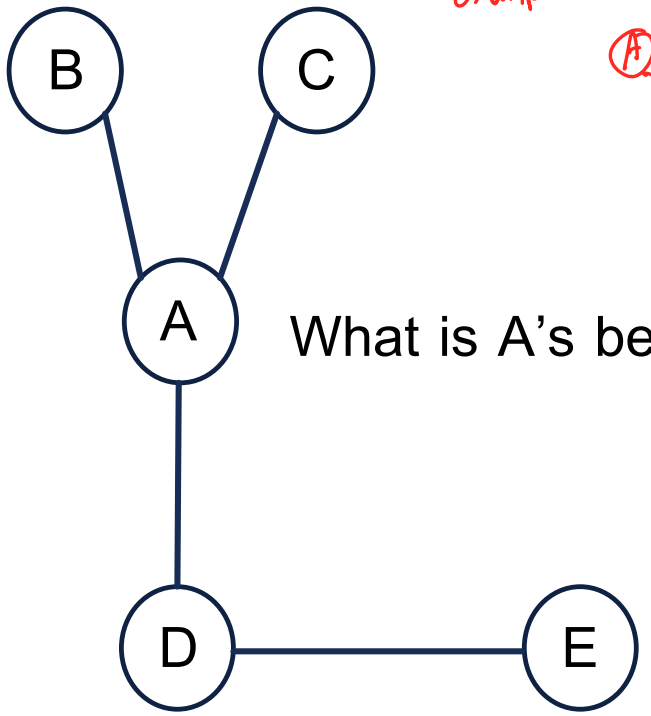
https://forms.gle/mGRFKrpKuDHhyEy16

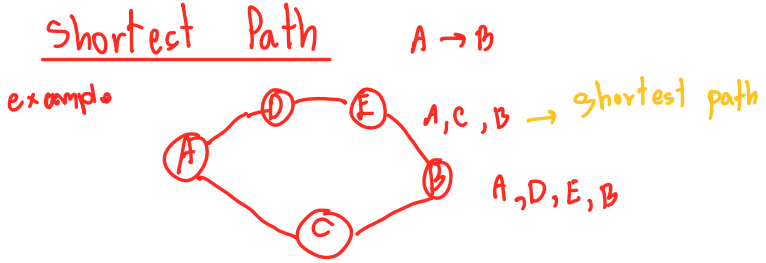
Nodal Metric

PART 3 – BETWEENNESS CENTRALITY

NODAL METRICS – BETWEENNESS CENTRALITY

- How many paths does a node lie on?
- For a given node v, the number paths between i and j that pass through v
- In the middle of the path





What is A's betweenness centrality?

NODAL METRICS – BETWEENNESS CENTRALITY

A

From	To
~~A~~	~~A~~
A	B
A	C
A	D
A	E

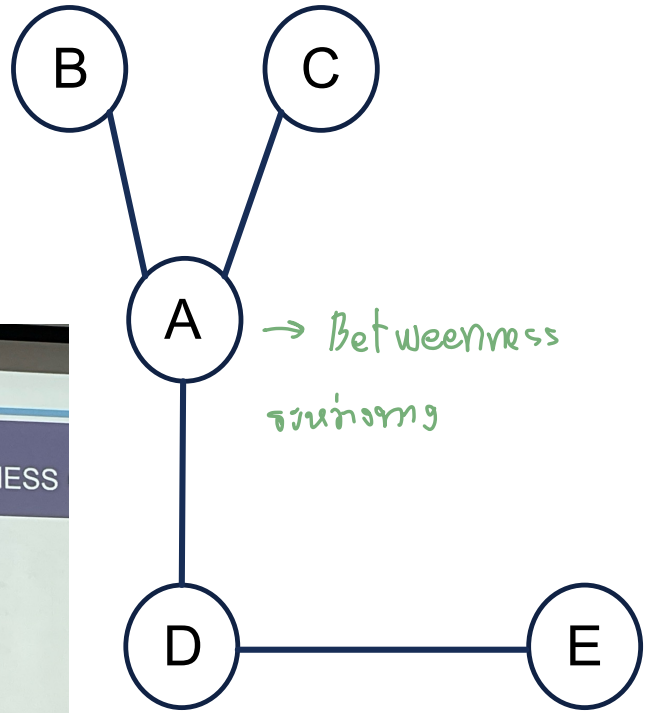
B

From	To
~~B~~	~~A~~
~~B~~	~~B~~
B	C
B	D
B	E

C

From	To
~~C~~	~~A~~
~~C~~	~~D~~
~~C~~	~~C~~
C	D
C	E

* ตารางแสดงความสัมพันธ์ของ Nodes



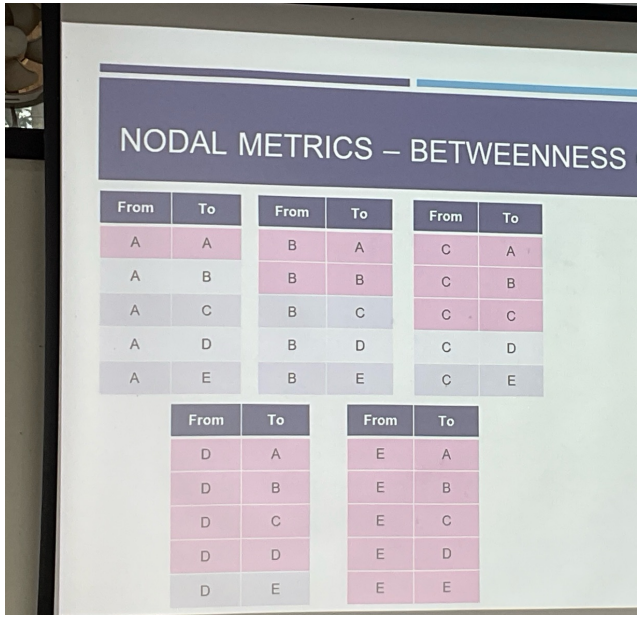
→ Betweenness ระหว่าง

D

From	To
~~D~~	A
~~D~~	~~B~~
~~D~~	~~C~~
~~D~~	~~D~~
~~D~~	~~E~~

E

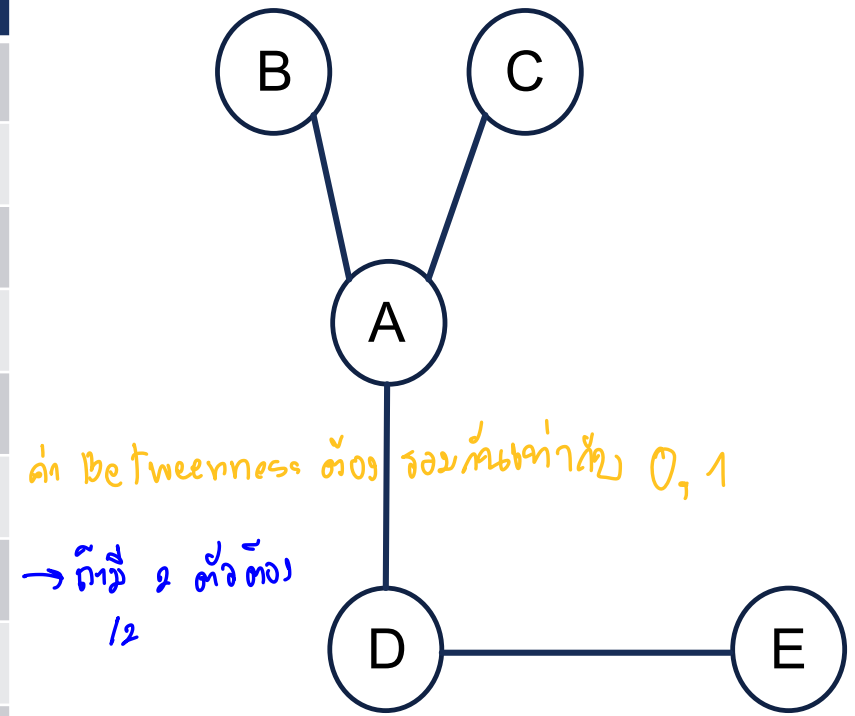
From	To
~~E~~	~~A~~
E	B
E	C
E	~~D~~
~~E~~	E



NODAL METRICS – BETWEENNESS CENTRALITY

From	To	Path	A	B	C	D	E
A	B	(A, B)	0	0	0	0	0
A	C	(A, C)	0	0	0	0	0
A	D	(A, D)	0	0	0	0	0
A	E	(A, D, E)	0	0	0	1	0
B	C	(B, A, C)	1	0	0	0	0
B	D	(B, A, D)	1	0	0	0	0
B	E	(B, A, D, E)	0.5	0	0	0.5	0
C	D	(C, A, D)	1	0	0	0	0
C	E	(C, A, D, E)	0.5	0	0	0.5	0
D	E	(D, E)	0	0	0	0	0
			4	0	0	2	0

Sum



ค่า Betweenness ต้องอยู่ระหว่าง 0, 1

→ ถ้ามี 2 ตัวต้อง 1/2

: ค่า Betweenness แบบ Non-Scaling

NODAL METRICS – BETWEENNESS CENTRALITY

A	B	C	D	E	
4	0	0	2	0	Sum
6	6	6	6	6	Denominator
0.67	0	0	0.34	0	Betweenness

ค่าหาร เท่ากันทุก Node

Scaling

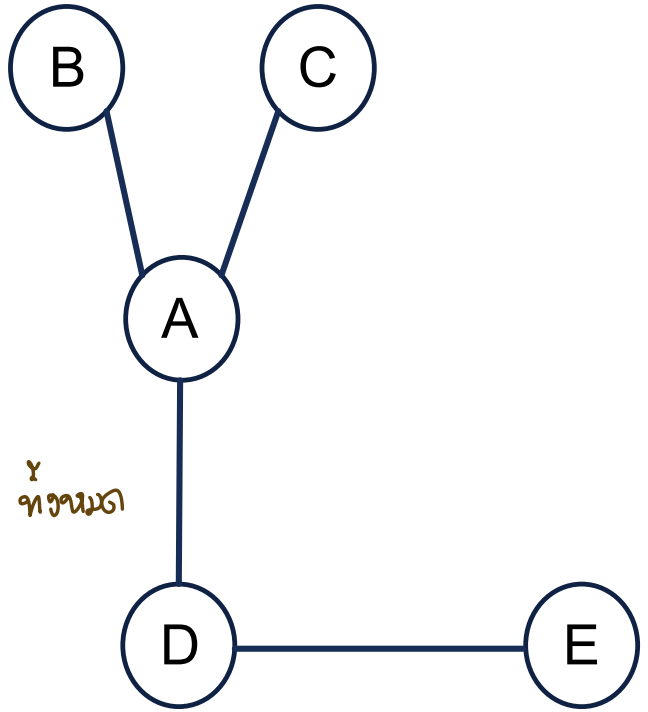
0 - 1

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

n คือ จน. Node ทั้งหมด

n = 5

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



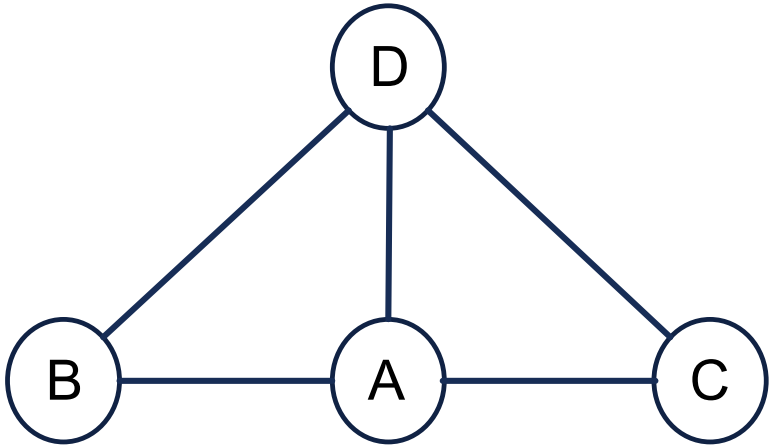
NODAL METRICS – BETWEENNESS CENTRALITY

From	To
A	A
A	B
A	C
A	D

From	To
B	A
B	B
B	C
B	D

From	To
C	A
C	B
C	C
C	D

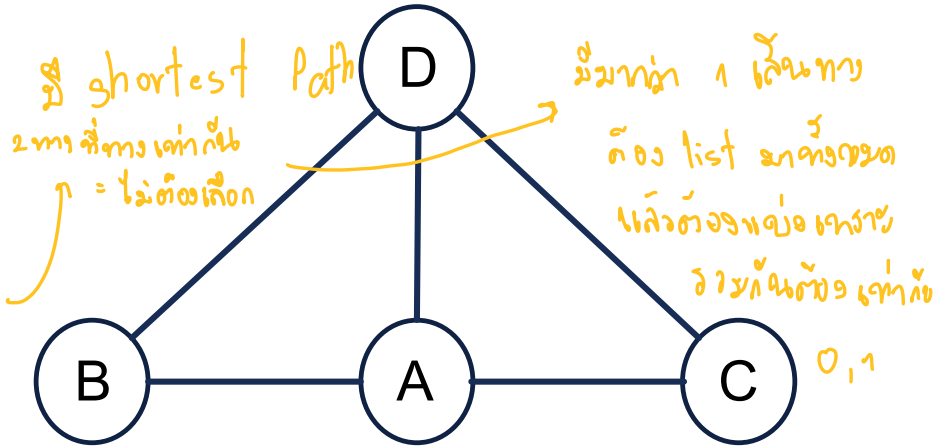
From	To
D	A
D	B
D	C
D	D



NODAL METRICS – BETWEENNESS CENTRALITY

From	To	Path	A	B	C	D
A	B	(A, B)	0	0	0	0
A	C	(A, C)	0	0	0	0
A	D	(A, D)	0	0	0	0
B	C	(B, A, C), (B, D, C)	0.5	0	0	0.5
B	D	(B, D)	0	0	0	0
C	D	(C, D)	0	0	0	0
			0.5	0	0	0.5

Sum

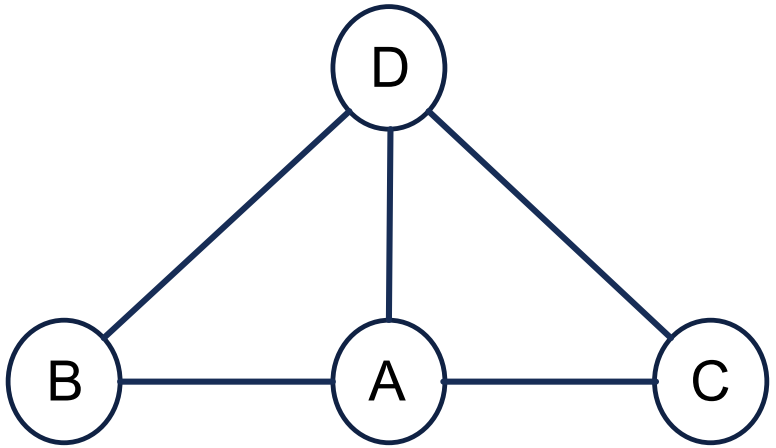


NODAL METRICS – BETWEENNESS CENTRALITY

A	B	C	D	
0.5	0	0	0.5	Sum
3	3	3	3	Denominator
0.17	0	0	0.17	**Betweenness**

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

$\frac{(3)(2)}{2} = 3$



Min-Max scaling Method

ANOTHER SPECIAL CASE

B → D : (B, A, C, D)

(B, A, E, D)

4

$\frac{1}{4}$ = 0.25

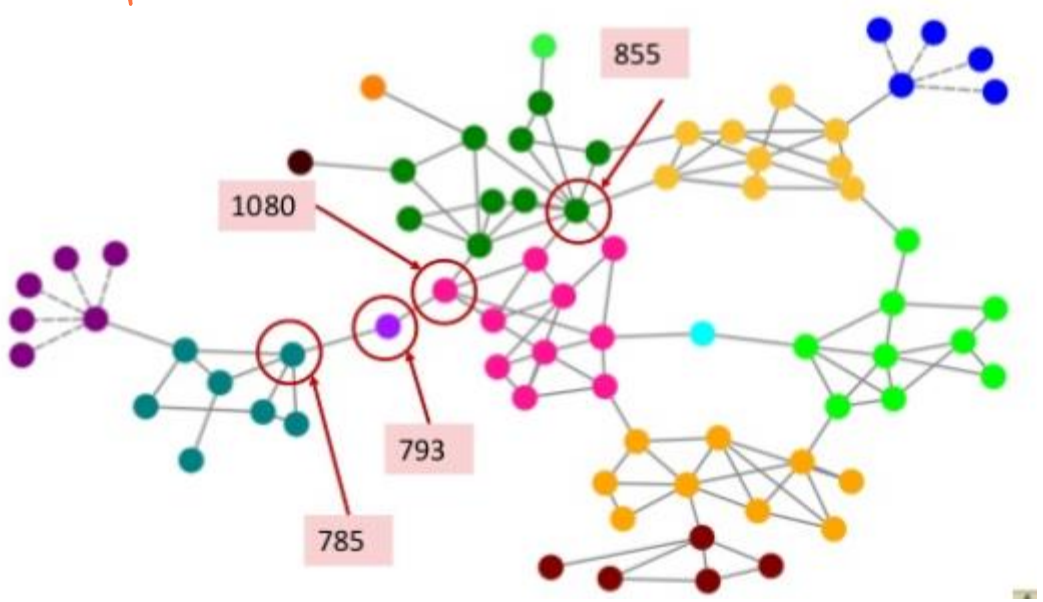
A	B	C	D	E
0.5		0.25		0.25

A ที่ เท่า 0.25 / 2 = 0.5

NODAL METRICS – BETWEENNESS CENTRALITY

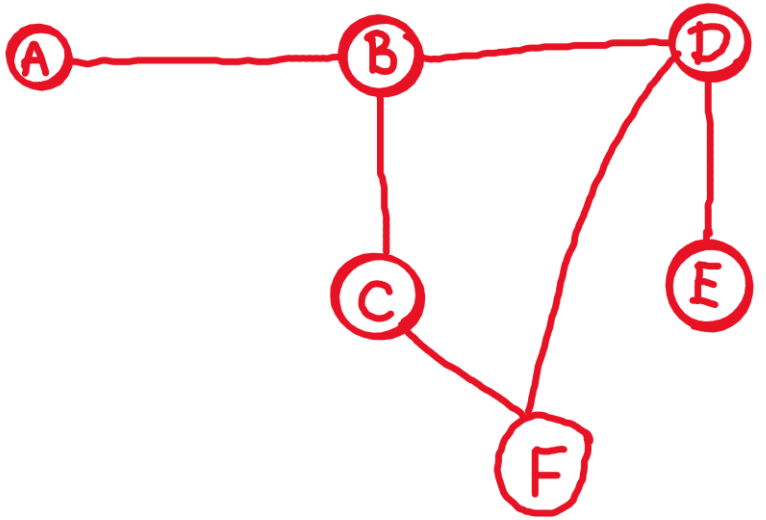
- Shows which nodes are more likely to be in communication paths between other nodes
- Useful in determining points where the network would break apart

Nodes 가 in Betweenness 값이 : Important



BETWEENNESS CENTRALITY EXERCISE

- Determine the betweenness centrality for all nodes in the following social graph, using the scaling method
- Please complete the exercise
  - You will be given marks for the correct answers



https://forms.gle/TYWuggpCg42cFgag8

TO BE CONTINUED

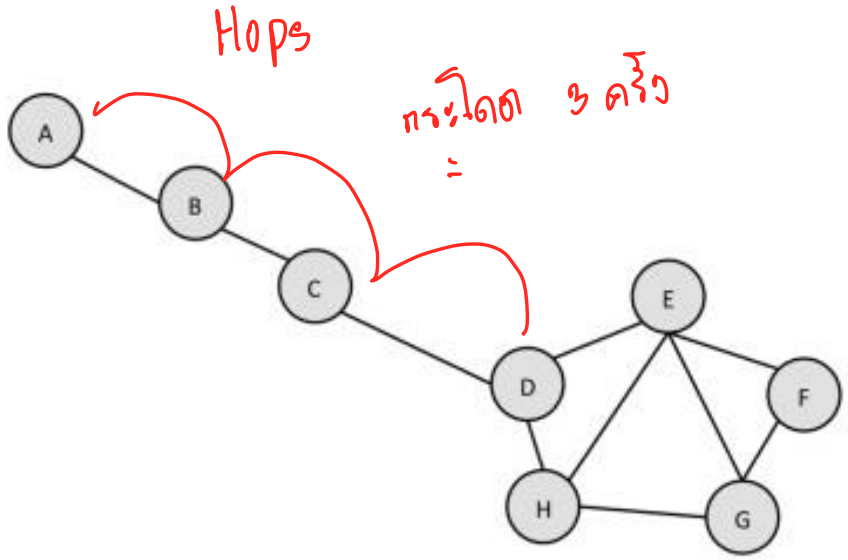
SOCIAL NETWORK ANALYSIS METRICS

PART 4 – CLOSENESS CENTRALITY

NODAL METRICS – CLOSENESS

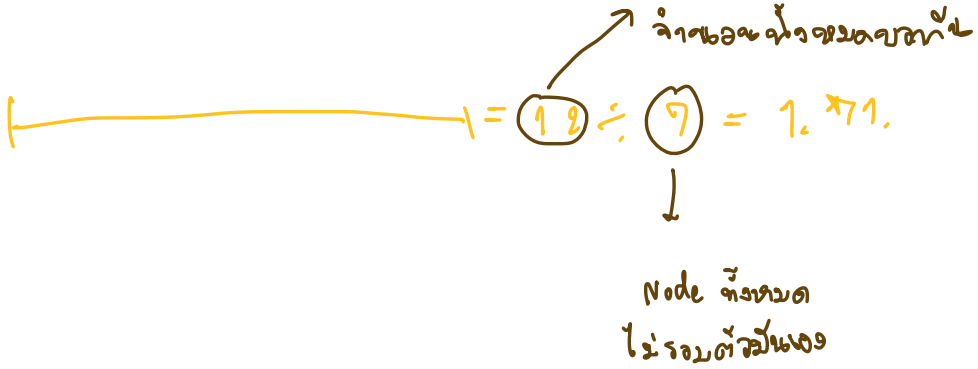
- Closeness or Closeness Centrality (Distance)
  - Indicates how close a node is to all other nodes in the network
  - Looks for the node that is closest to all other nodes
- The average length of all the shortest paths from that one node to every other node in the network

NODAL METRICS – CLOSENESS CENTRALITY

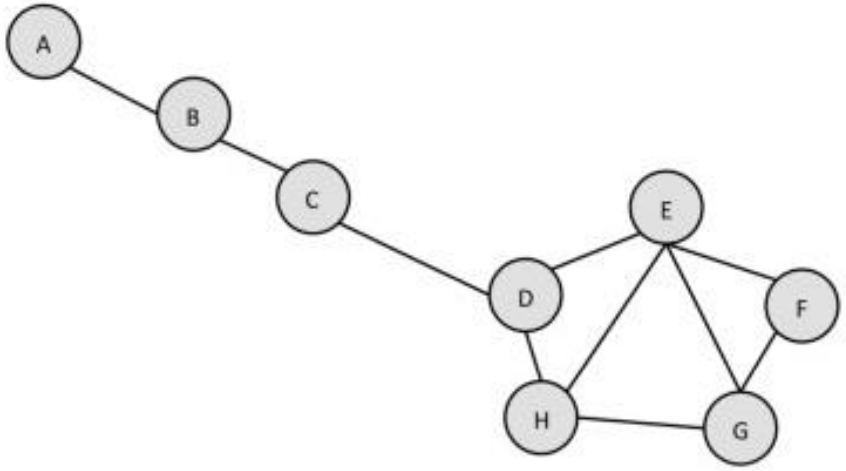


Node	Shortest Path from D
A	3
B	2
C	1
E	1
F	2
G	2
H	1

Let's determine the average distance for node D.



NODAL METRICS – CLOSENESS CENTRALITY



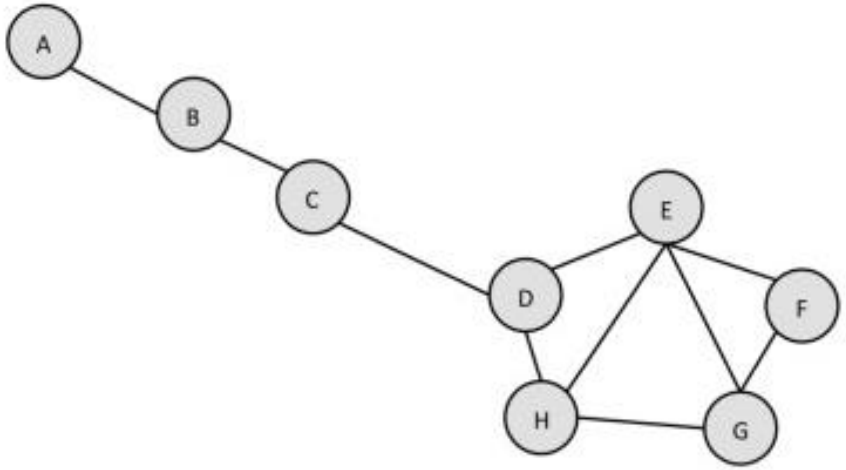
Node	Shortest Path from D
B	1
C	2
D	3
E	4
F	5
G	5
H	4

Let's determine the average distance for node A.

ระยะทางเฉลี่ย

24 ÷ 7 = 3.43

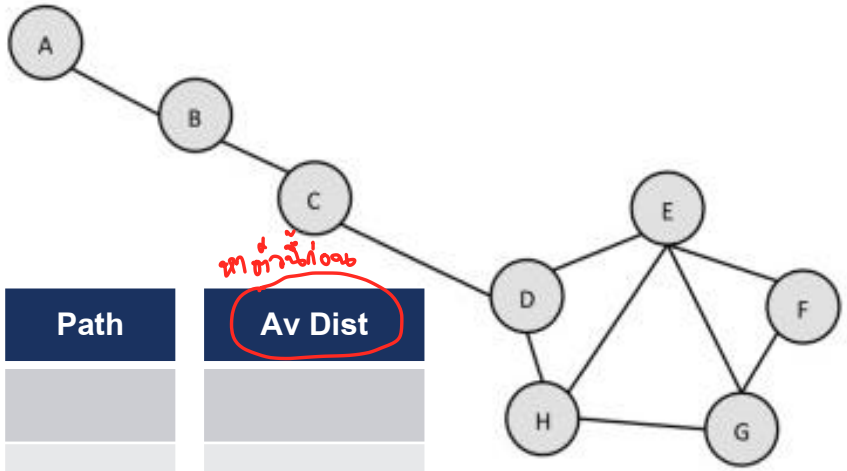
NODAL METRICS – CLOSENESS CENTRALITY



Let's determine the average distance for all other nodes.

NODAL METRICS – CLOSENESS CENTRALITY

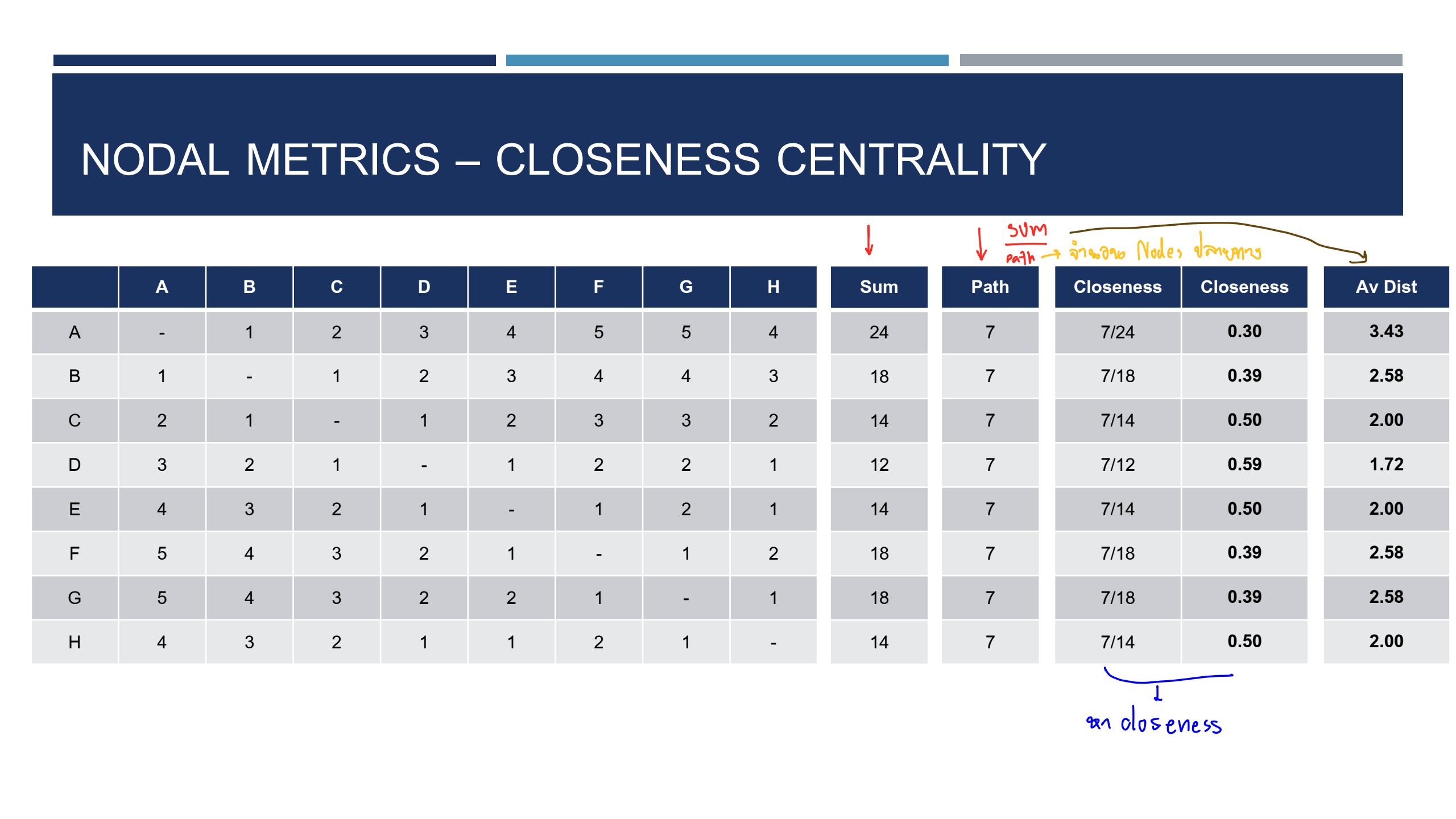
Let's determine the closeness centrality for all nodes.



	A	B	C	D	E	F	G	H	Sum	Path	Av Dist
A											
B											
C											
D											
E											
F											
G											
H											

NODAL METRICS – CLOSENESS CENTRALITY

	A	B	C	D	E	F	G	H	Sum	Path	Closeness	Closeness	Av Dist
A	-	1	2	3	4	5	5	4	24	7	7/24	**0.30**	**3.43**
B	1	-	1	2	3	4	4	3	18	7	7/18	**0.39**	**2.58**
C	2	1	-	1	2	3	3	2	14	7	7/14	**0.50**	**2.00**
D	3	2	1	-	1	2	2	1	12	7	7/12	**0.59**	**1.72**
E	4	3	2	1	-	1	2	1	14	7	7/14	**0.50**	**2.00**
F	5	4	3	2	1	-	1	2	18	7	7/18	**0.39**	**2.58**
G	5	4	3	2	2	1	-	1	18	7	7/18	**0.39**	**2.58**
H	4	3	2	1	1	2	1	-	14	7	7/14	**0.50**	**2.00**



NODAL METRICS – CLOSENESS CENTRALITY

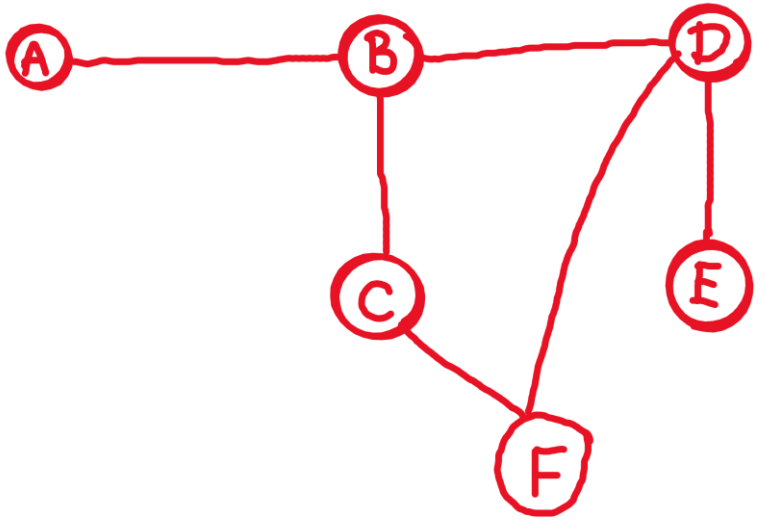
- For closeness centrality
  - Larger values mean the node is more central
  - It takes fewer steps to get to other nodes
- For average shortest distance
  - Smaller values mean the node is more central
  - It takes fewer steps to get to other nodes

NODAL METRICS – CLOSENESS CENTRALITY

- Good for
  - Finding the individuals who are best placed to influence the entire network most quickly.

CLOSENESS CENTRALITY EXERCISE

- Determine the average distance and closeness centrality for all nodes in the following social graph
- Please complete the exercise
  - Marks will be given for the correct answers



https://forms.gle/8nUFCVs5fAUmFhWY9

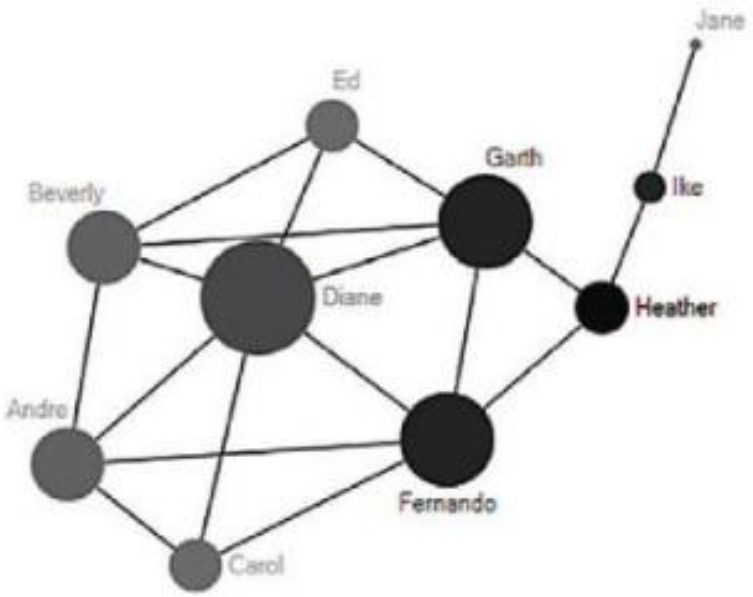
Nodal Metric

PART 5 – EIGENVECTOR CENTRALITY

NODAL METRICS – EIGENVECTOR CENTRALITY

- In many cases, a connection to a popular individual is more important than a connection to a loner.
- The Eigenvector Centrality network metric takes into consideration not only how many connections a node has (i.e., its degree), but also the degree of the nodes that it is connected to.
- Useful for determining who is connected to the most connected node

NODAL METRICS – EIGENVECTOR CENTRALITY



- Both Heather and Ed have a degree of 3
- Ed is directly connected to Diane (the most popular)
- Heather is connected to Ike (the least popular)
- The Eigenvector Centrality metric for Heather is lower than it is for Ed

A node is important if it is linked to other important nodes.

INTERPRETATION OF NODAL METRICS

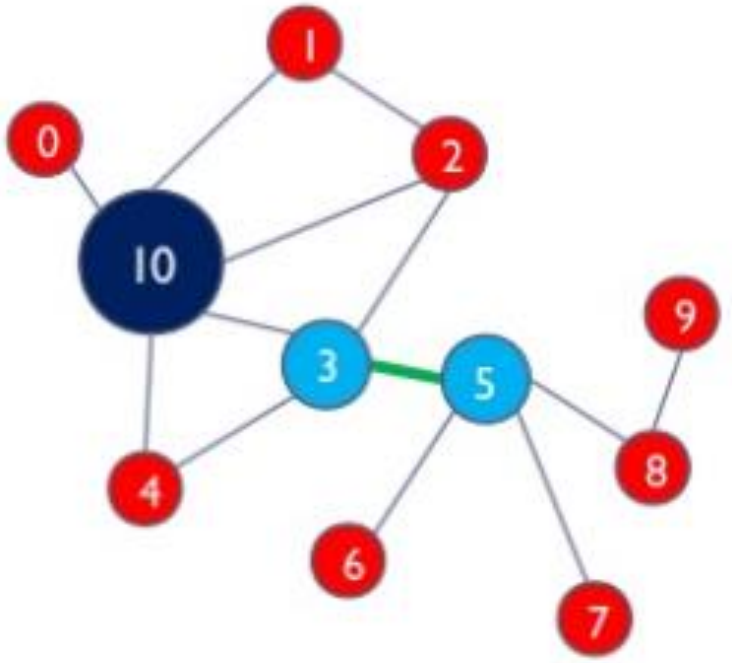
Degree	• How many people can this person reach directly?
Betweenness	• How likely is this person to be the most direct route between two people in the network?
Closeness	• How fast can this person reach everyone in the network?
Eigenvector	• How well is this person connected to other well-connected people?

INTERPRETATION OF NODAL METRICS

Degree	• How many people has this person collaborated with?
Betweenness	• Who is the person that information is most likely to go through?
Closeness	• How fast will the virus spread from this person to the rest of the network?
Eigenvector	• How is the person that is connected to other well connected people?

IDENTIFYING KEY ACTORS

- In the network to the right, node 10 is the most central according to degree centrality
- But nodes 3 and 5 together will reach more nodes
- The tie between them is critical; if severed, the network will break into two isolated sub-networks



SUMMARY

- Seeing nodes and edges as a graph is not enough
- Need to be able to analyse and interpret the network in different dimensions
  - Degree centrality
  - Betweenness centrality
  - Closeness centrality
  - Eigenvector centrality

PART 6 – SOCIAL NETWORK DENSITY

STRUCTURAL METRICS – NETWORK DENSITY

- Number of connections that exist out of all possible connections
  - Actual connections vs Potential connections

จำนวน connection ที่เป็นไปได้

- Calculation:

Node ทั้งหมด

Potential Connections:

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

Network Density:

$$\frac{\text{Actual Connections}}{\text{Potential Connections}}$$

Actual connection (AC) = 5

Potencial connection (PC) = 10

Density = $\frac{AC}{PC} = \frac{5}{10} = 0.5$

  - Closer to 1 means it is a dense network.

หนาแน่น

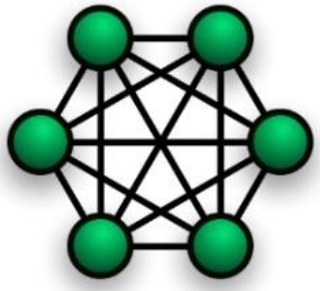
  - Otherwise, it is a sparse network.

ไม่หนาแน่น = ไม่มีการเชื่อมต่อ

  - When the density is 1, the network is called a clique.

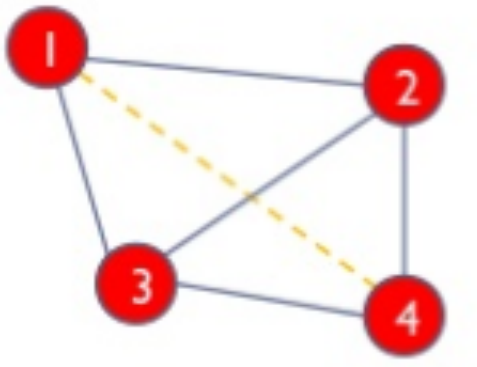
ทุก node มีการเชื่อมต่อกันหมด

A **CLIQUE** is a network in which *everyone is connected to everyone else.*



NETWORK DENSITY – EXAMPLE

- What is the density of this network?
  - What is the value of PC?
  - What is the value of actual connection?
  - What is the density of this network?
  - Density of 5/6 = 0.83 (Fairly dense)



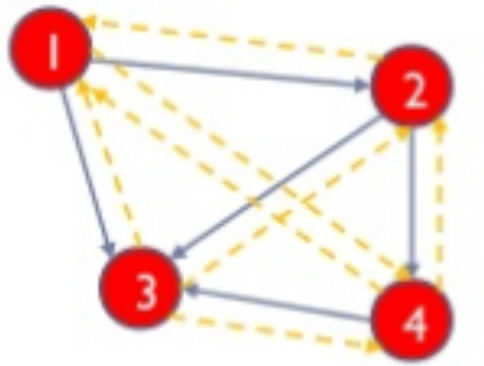
$Pc = \frac{n * (n-1)}{2}$

This is an undirected graph.

$\text{Density} = \frac{Ac}{Pc} = \frac{5}{6} = 0.83$

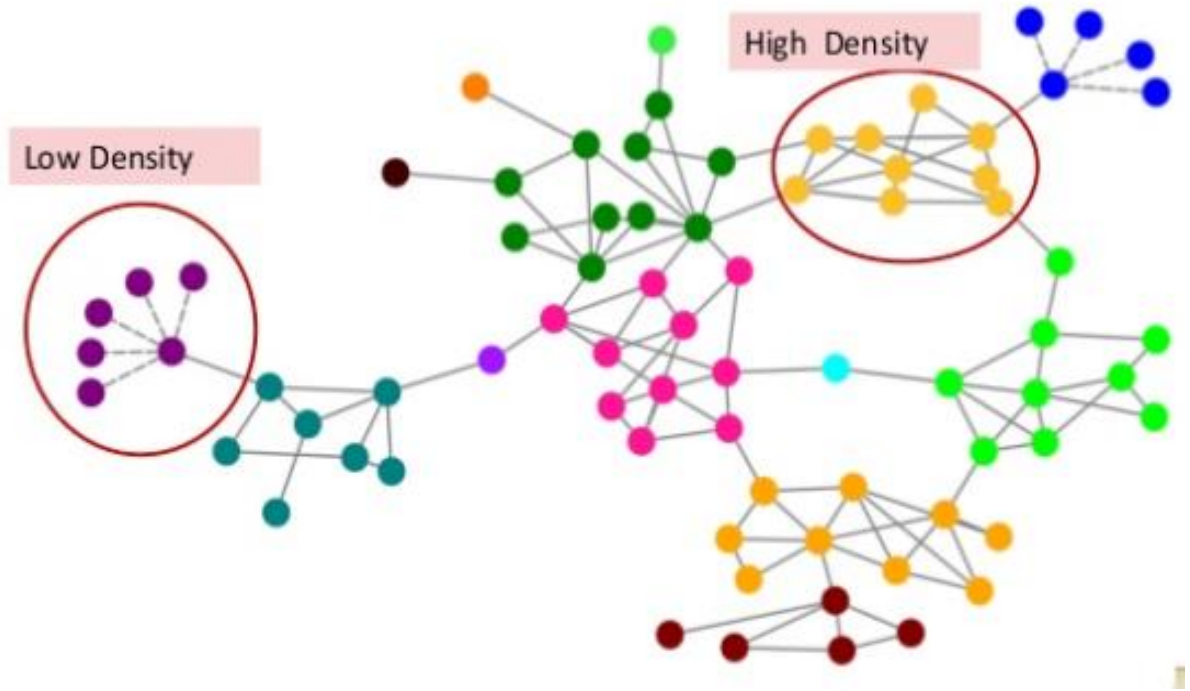
NETWORK DENSITY – EXAMPLE

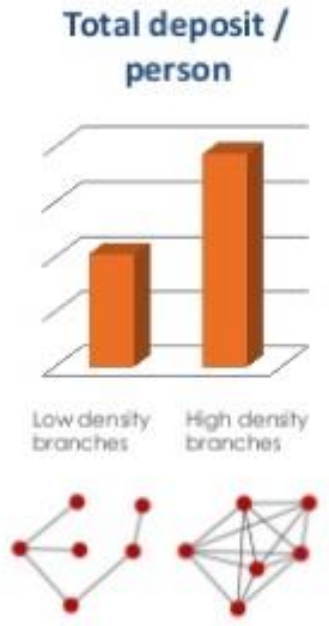
- In case of a directed graph
  - The density is halved of its undirected equivalent
  - There are twice as many possible edges
  - That is, potential connections = n*(n-1)



Density = 5/12 = 0.42

NETWORK DENSITY – EXAMPLE





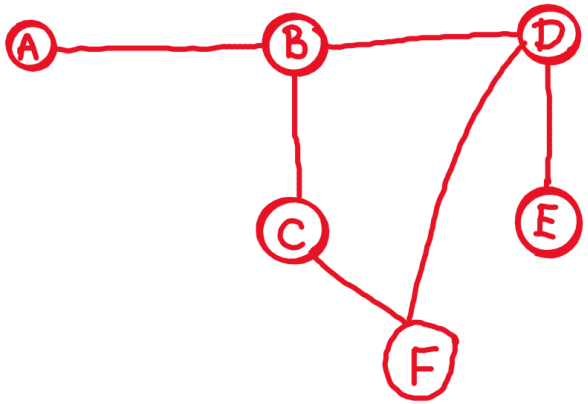
Difference in performance in bank branches based on the density of their relationships

NETWORK DENSITY EXERCISE

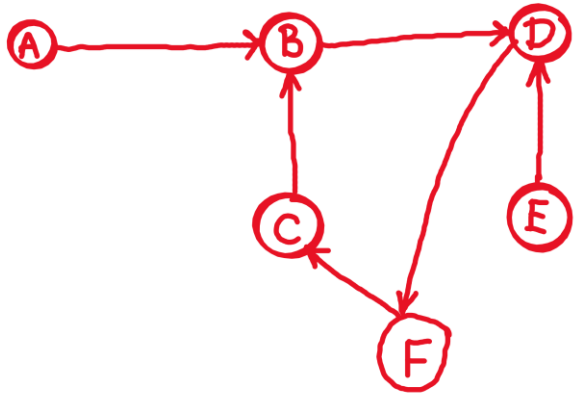
- Determine the density of the following social graphs
- Please complete the exercise
  - Marks will be given for the correct answers

https://forms.gle/Vu8eKuGLxdaUxncUA

1.



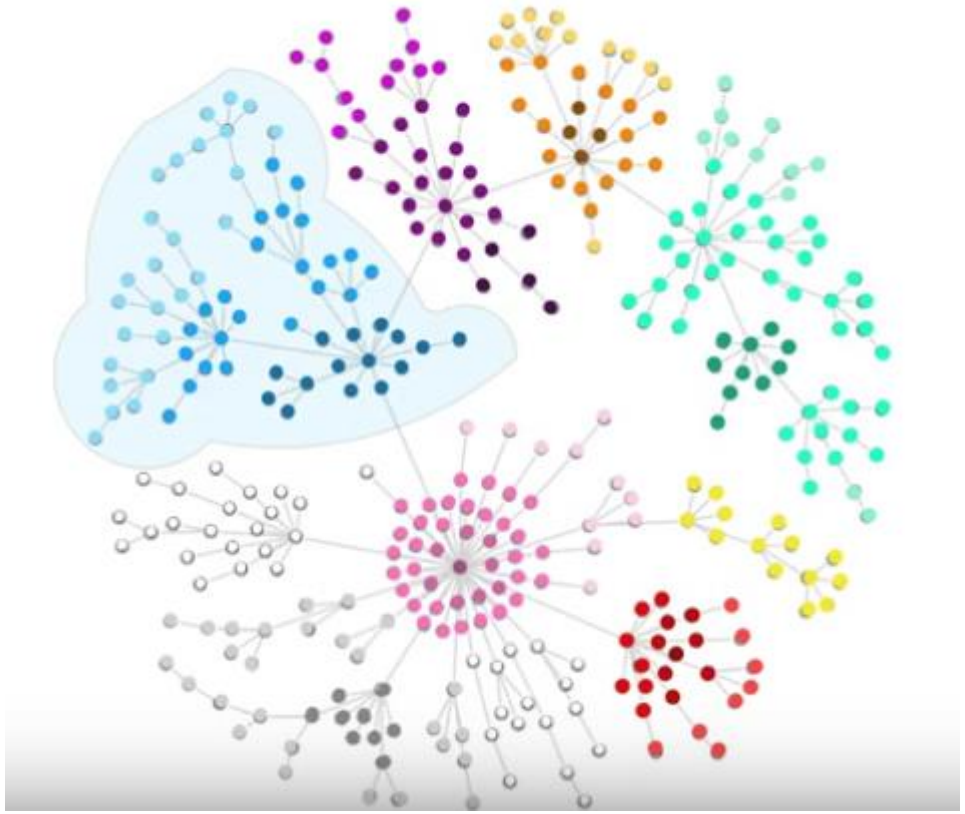
2.



PART 7 – CLUSTERS

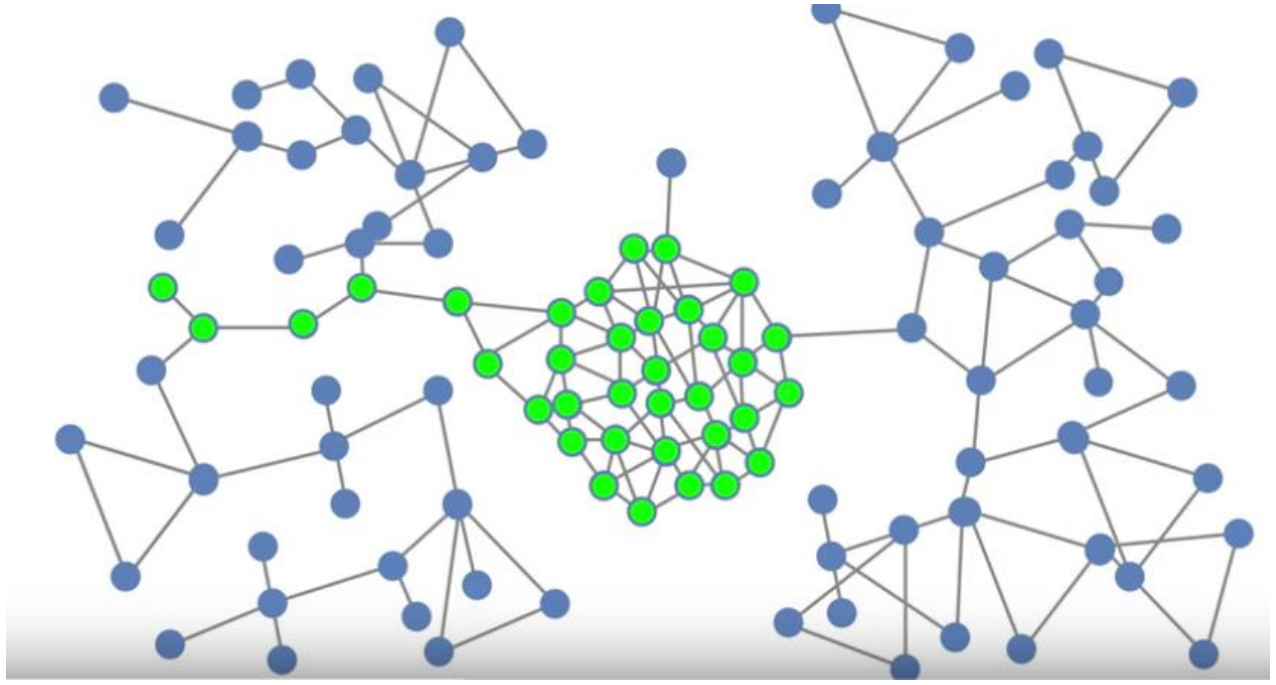
STRUCTURAL METRICS – CLUSTERS

- A cluster is a group of nodes more connected to each other than to other nodes



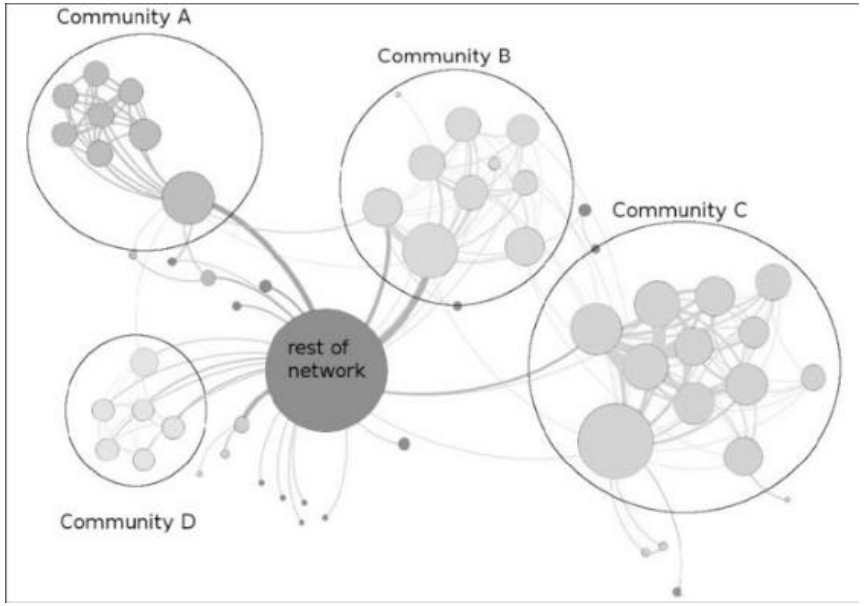
STRUCTURAL METRICS – CLUSTERS

- Could create a bottle neck for information or diffusion process



STRUCTURAL METRICS – CLUSTERS

- We can do community detection to identify clusters
- How different parts of the network interact with each other
- Any community that we might be interested in
- Sometimes people in a cluster do have unique features



https://www.researchgate.net/publication/332207478_The_Distance_Between_Us/figures?lo=1

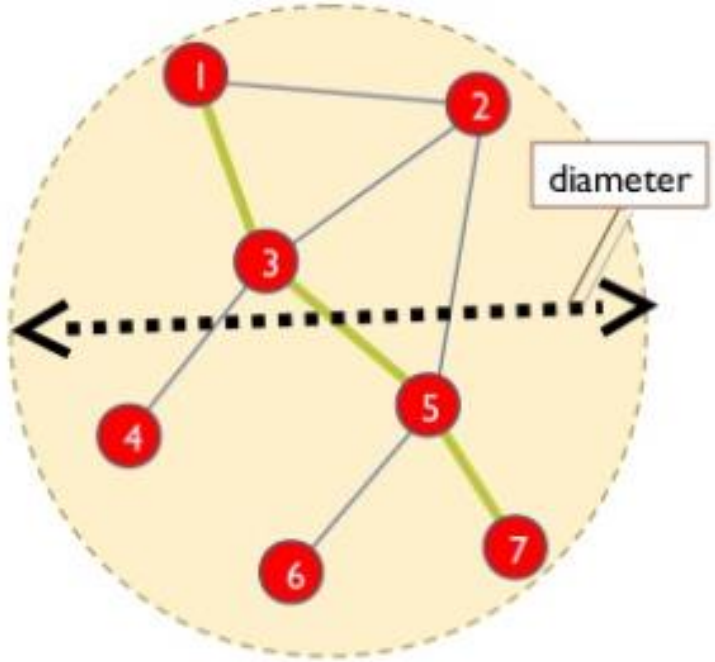
STRUCTURAL METRICS – MODULARITY CLASS

- Designed to measure the strength of division of a network into clusters
- High modularity
  - Dense connections between nodes within cluster
  - Sparse connections between nodes in different clusters

PART 8 – NETWORK DIAMETER

STRUCTURAL METRICS – NETWORK DIAMETER

- The longest shortest path between any two nodes is called the network's diameter.
- It indicates how long it will take at most to reach any node in the network
- Sparser networks tend to have larger diameters

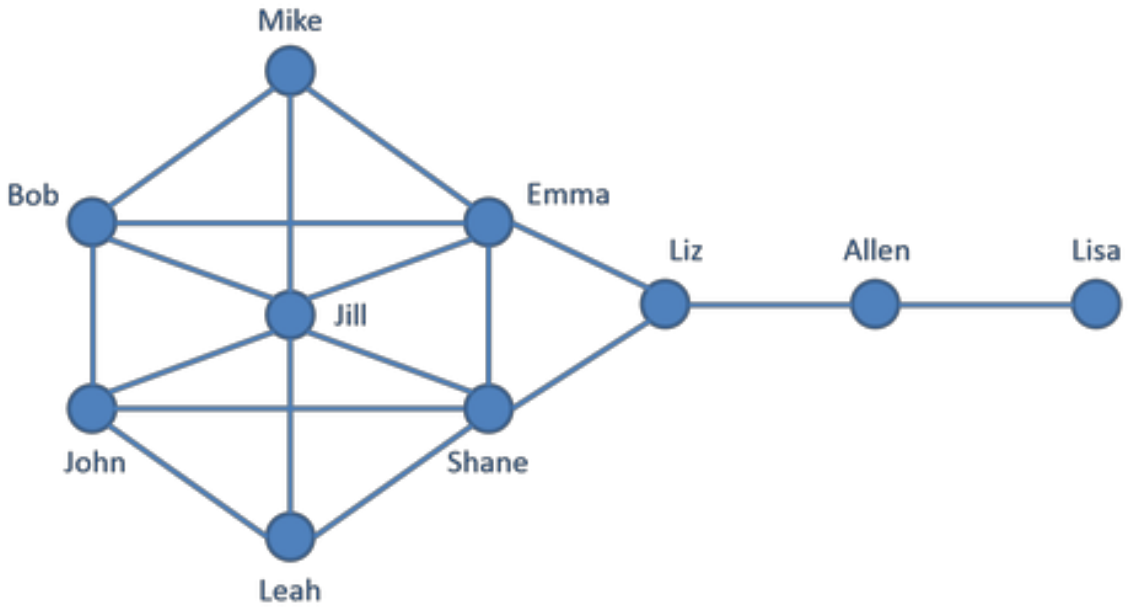


This network's diameter is 3.

https://www.slideshare.net/bhadauriapradeep/social-network-analysis-basics

STRUCTURAL METRICS – NETWORK DIAMETER

- What is the diameter of this social graph?



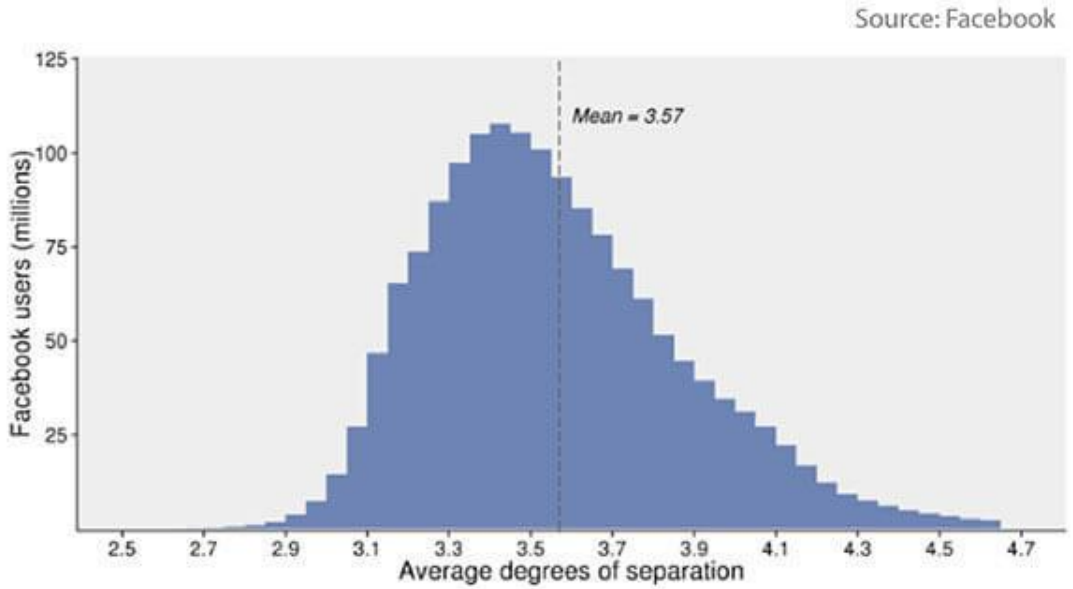
https://nauczanki.wordpress.com/2014/11/10/social-network-analysis-basics-in-300-words-dalmooc-assignment-week-3/

STRUCTURAL METRIC – SMALL WORLD

- A small world social network
  - Short average path length
  - High modularity
    - Many nodes know one another
- Facebook when it had approximately 721 million users with 69 billion friendship connections between them
  - The average path length turned out to be just 4.74 hops

STRUCTURAL METRIC – SMALL WORLD

- In their latest research published in February 2016, this number dropped down to 3.57, with more than 1.59 billion people active on Facebook.

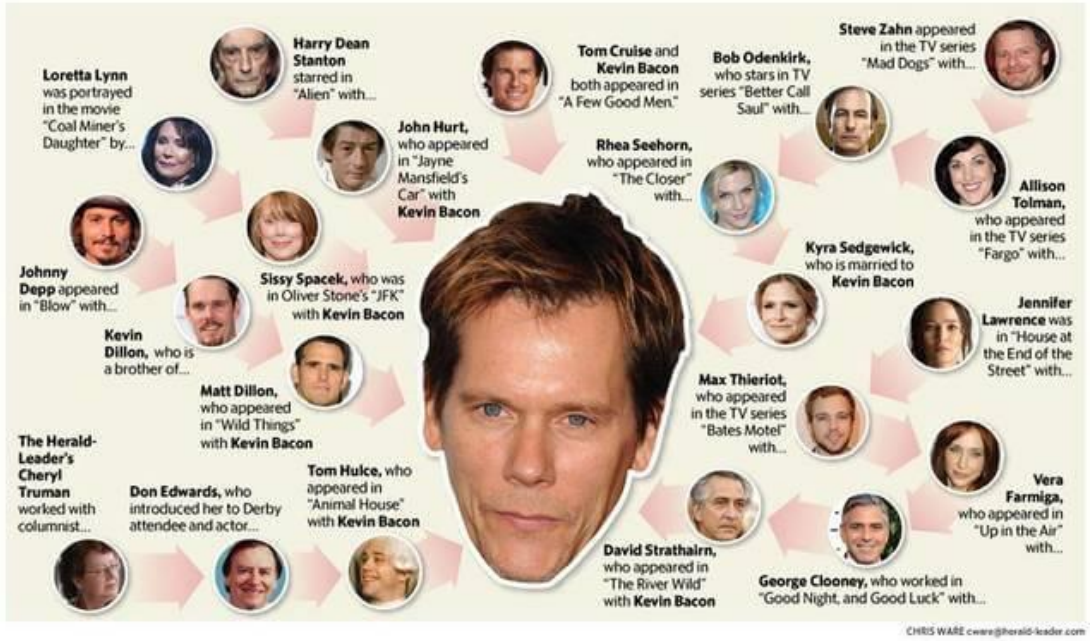


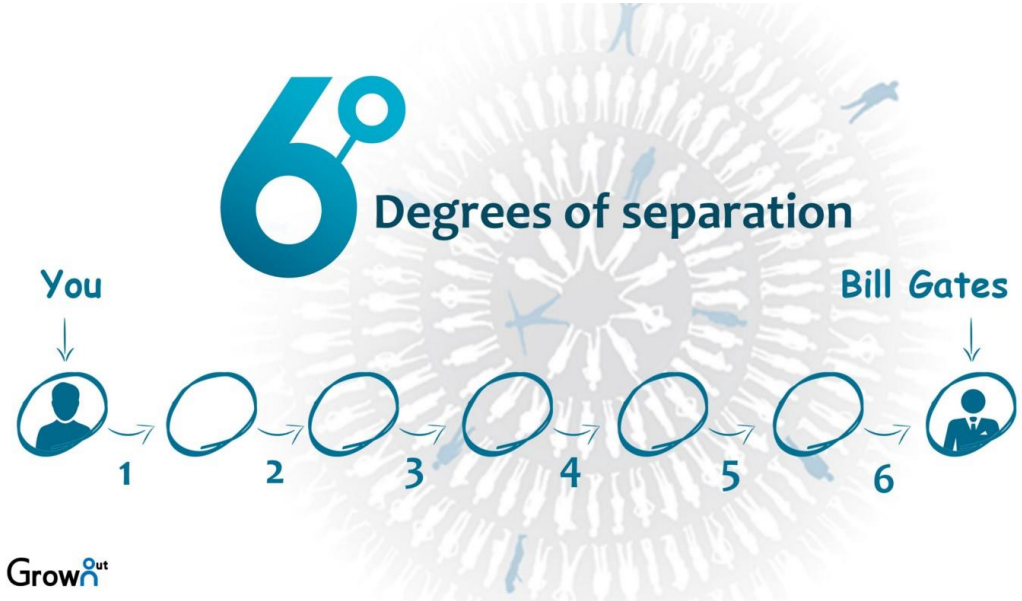
https://www.hackerearth.com/blog/developers/theory-six-degrees-of-separation/

STRUCTURAL METRIC – SMALL WORLD

- Despite a large number of nodes, it is possible to find short communication paths between them
- Six degrees of separation
  - This concept tells that everything in this world are six or fewer steps away from each other

SIX DEGREES OF SEPARATION





SUMMARY

- Two more nodal metrics
  - Closeness centrality
  - Eigenvector centrality
- Analyse social network at macro level
  - Structural metrics
    - Density
    - Clusters
    - Diameters
  - Small world networks