KU LEUVEN

Introduction to network analysis

11.05.2021

Lectio Roundtable ‘Linking People. Network Analysis and intellectual History’

Margherita Fantoli

margherita.fantoli@kuleuven.be

Presentation structure

Part I: Definition and characterization of a network

- Nodes
- Edges
- Encoding the information

Part II: Using a network

- Visualization
- Analysis

Part I

Definition and characterization of a network

What is a network?

Formal definition

Graph: ordered pair constituted by a set of nodes and a set of edges.

$G = (V, E)$

What is a network?

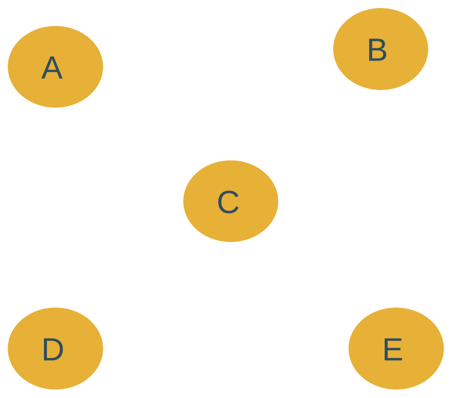
Formal definition

Graph: Ordered pair constituted by a set of nodes and a set of edges.

G= (V, E)

Nodes: the entities you deal with (people, institutions, texts, words...)

V= {A, B, C, D, E}



What is a network?

Formal definition

Graph: Ordered pair constituted by a set of nodes and a set of edges.

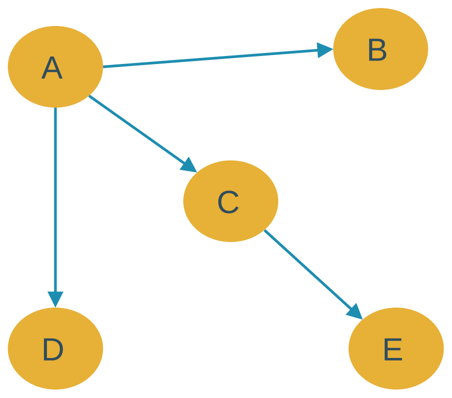
G= (V, E)

Nodes: the entities you deal with (people, institutions, texts, words...)

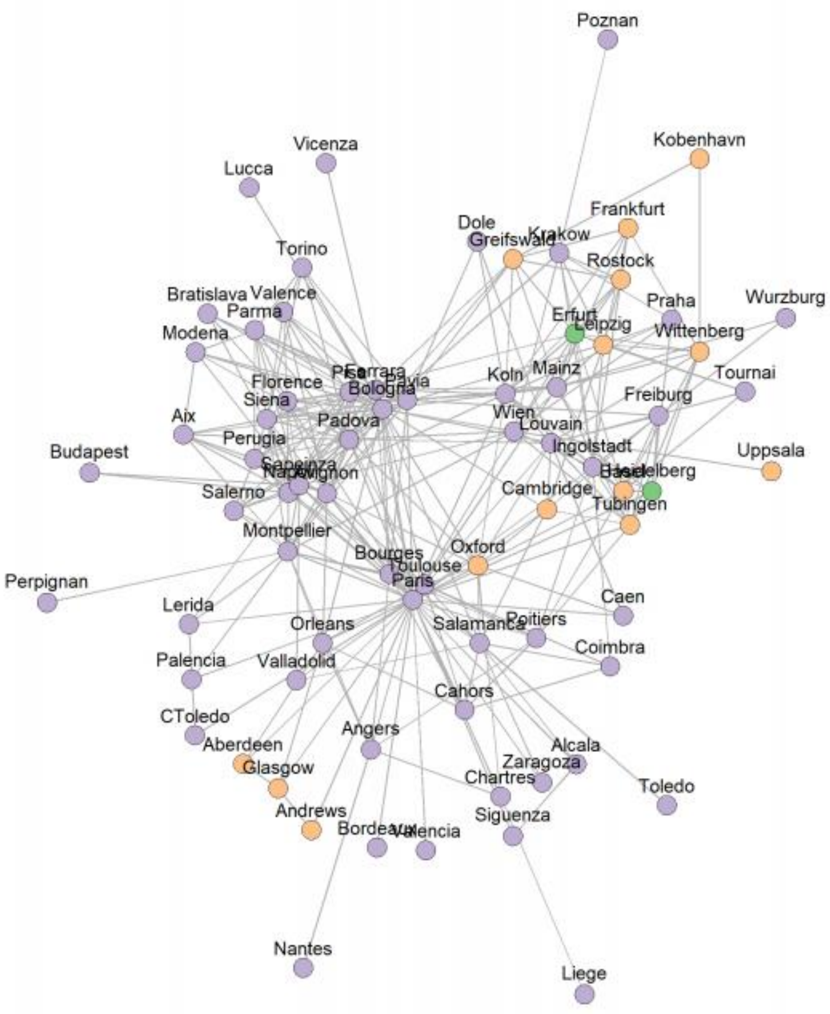
V= {A, B, C, D, E}

Edges: the links between the nodes. Formally defined as pair of nodes.

E= {(A, B), (A, D), (A, C), (C, E)}

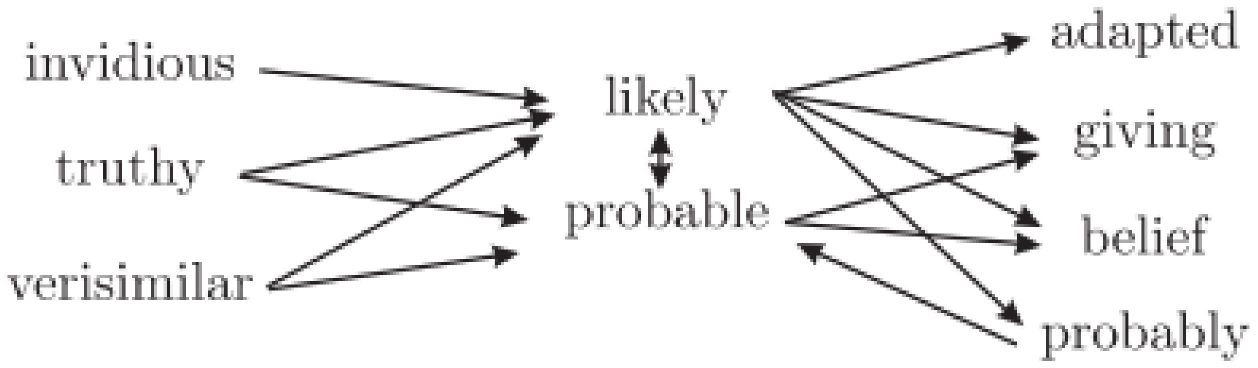


Nodes: institutions



Network of Universities before 1527
Source: [2020029.pdf (uclouvain.be)]() (David De La Croix).
Links: existence of shared scholars
Colors: religious affiliation
Goal: Historical evolution of university collaborations

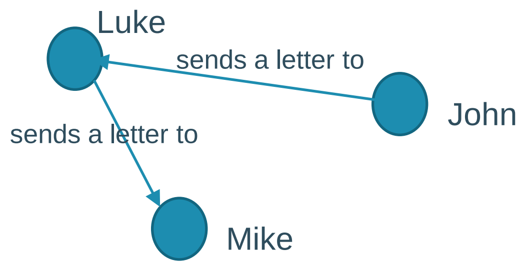
Nodes: words



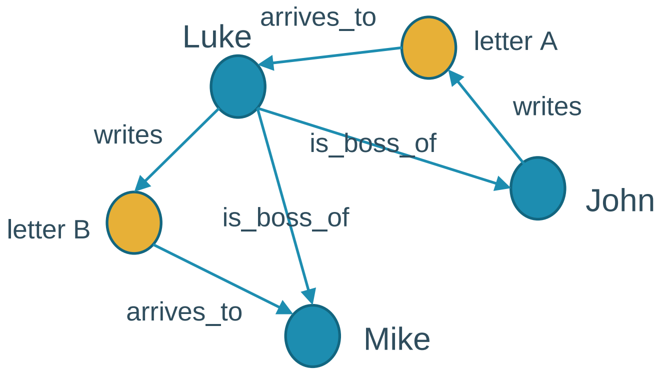
"Dictionary network" of the term likely
Source [4159 (uclouvain.be)](https://) (Blondel et al.)
Links: if *v* appears in the definition of *u*, *u*→*v*
Goal: synonyms detection

Nodes: unimodal or multimodal network

- Unimodal: only one kind of nodes (previous examples). Generally used for running metrics.



- Multimodal: different kind of entities (and relations). Particularly frequent case: bimodal network (two kinds of nodes). Useful to represent complex information and queries.



Edges

- Links indicate a relationship between nodes: communication, shared activity, cooccurrence, etc.
- Directed/undirected
- Weighted/unweighted

Edges: directed/undirected

Directed (arc): (A,B) ≠ (B,A)

- Asymmetric relations: "parent of", "writes to", "appears in the definition"

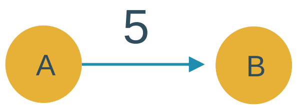
Undirected: (A,B)=(B,A)

- Symmetric relations: "is friend on FB", "is married to", "cooccurs with"

Edges: weighted/unweighted

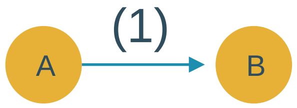
Weighted:

- Quantitative property of edges
- Registers the strength of the link
- Number of emails exchanged, number of times in which two words appear in the same sentence, probabilities, scalar evaluations.



Unweighted:

- Authorship network, genealogical networks etc.



How can we represent a network? List of edges and nodes

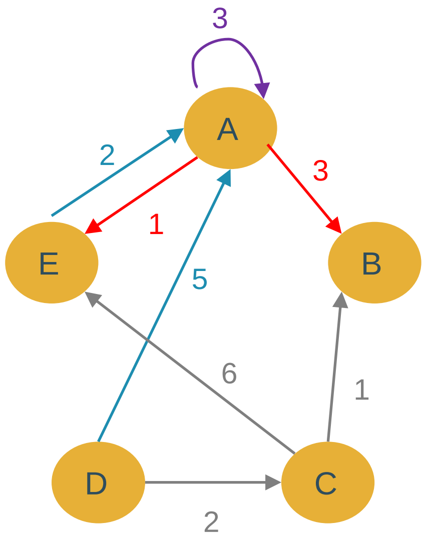
nodes

Id	Label
0	ABRAMSON, G
1	KUPERMAN, M
2	ACEBRON, J
3	BONILLA, L
4	PEREZVICENTE, C
5	RITORT, F
6	SPIGLER, R
7	ADAMIC, L
8	ADAR, E
9	HUBERMAN, B
10	LUKOSE, R
11	PUNIYANI, A
12	AERTSEN, A
13	GERSTEIN, G
14	HABIB, M
15	PALM, G
16	AFRAIMOVICH, V
17	VERICHEV, N
18	RABINOVICH, M
19	AGRAWAL, H
20	AHUJA, R

edges

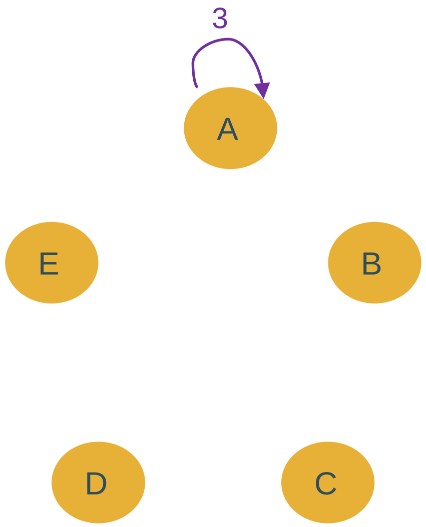
Source	Target	Type	Id
1	0	Undirected	2742
3	2	Undirected	2743
4	2	Undirected	2744
4	3	Undirected	2745
5	2	Undirected	2746
5	3	Undirected	2747
5	4	Undirected	2748
6	2	Undirected	2749
6	3	Undirected	2750
6	4	Undirected	2751
6	5	Undirected	2752
8	7	Undirected	2753
9	7	Undirected	2754
10	7	Undirected	2755
10	9	Undirected	2756
11	7	Undirected	2757
11	10	Undirected	2758
11	9	Undirected	2759
13	12	Undirected	2760
14	12	Undirected	2761
14	13	Undirected	2762
15	12	Undirected	2763

How can we represent a network? Adjacency matrix



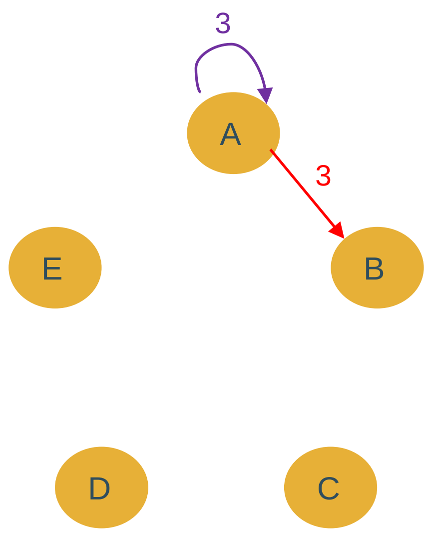
	A	B	C	D	E
A	3	3	0	0	1
B	0	0	0	0	0
C	0	1	0	0	6
D	5	0	2	0	0
E	2	0	0	0	0

How can we represent a network? Adjacency matrix



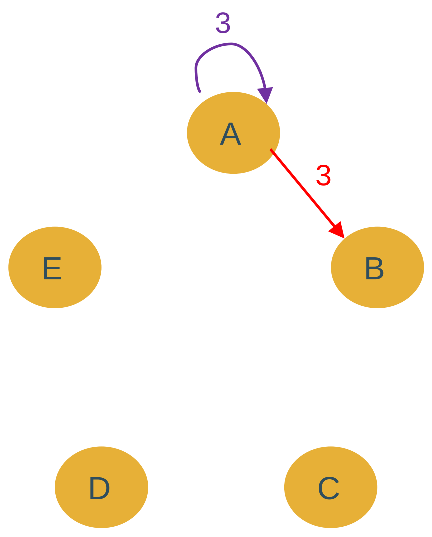
	A	B	C	D	E
A	3				
B					
C					
D					
E					

How can we represent a network? Adjacency matrix



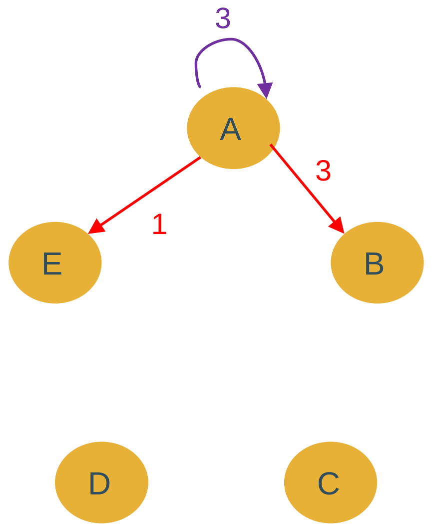
	A	B	C	D	E
A	3	3			
B					
C					
D					
E					

How can we represent a network? Adjacency matrix



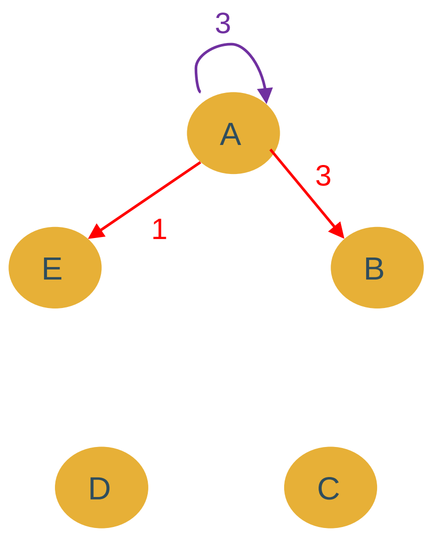
	A	B	C	D	E
A	3	3	0	0	
B					
C					
D					
E					

How can we represent a network? Adjacency matrix



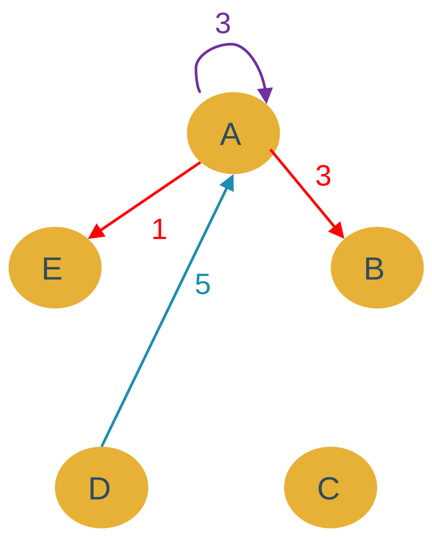
	A	B	C	D	E
A	3	3	0	0	1
B					
C					
D					
E					

How can we represent a network? Adjacency matrix



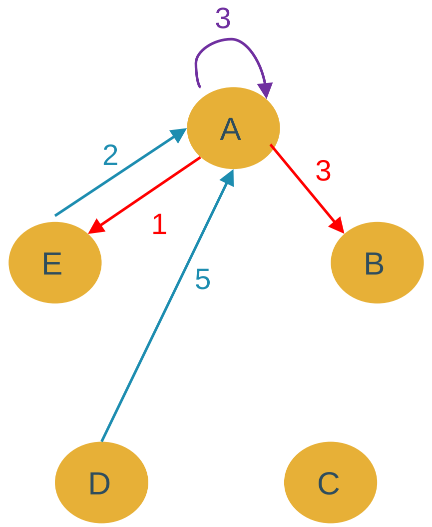
	A	B	C	D	E
A	3	3	0	0	1
B	0				
C	0				
D					
E					

How can we represent a network? Adjacency matrix



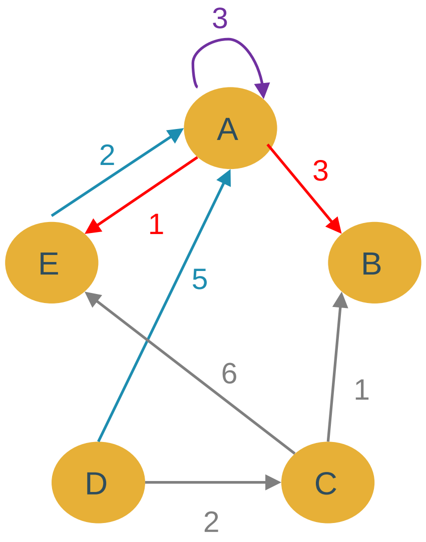
	A	B	C	D	E
A	3	3	0	0	1
B	0				
C	0				
D	5				
E					

How can we represent a network? Adjacency matrix



	A	B	C	D	E
A	3	3	0	0	1
B	0				
C	0				
D	5				
E	2				

How can we represent a network? Adjacency matrix



	A	B	C	D	E
A	3	3	0	0	1
B	0	0	0	0	0
C	0	1	0	0	6
D	5	0	2	0	0
E	2	0	0	0	0

Part II

Using a network

What can we do with a network?

Visualization

Analysis → global

Analysis → local

Visualization

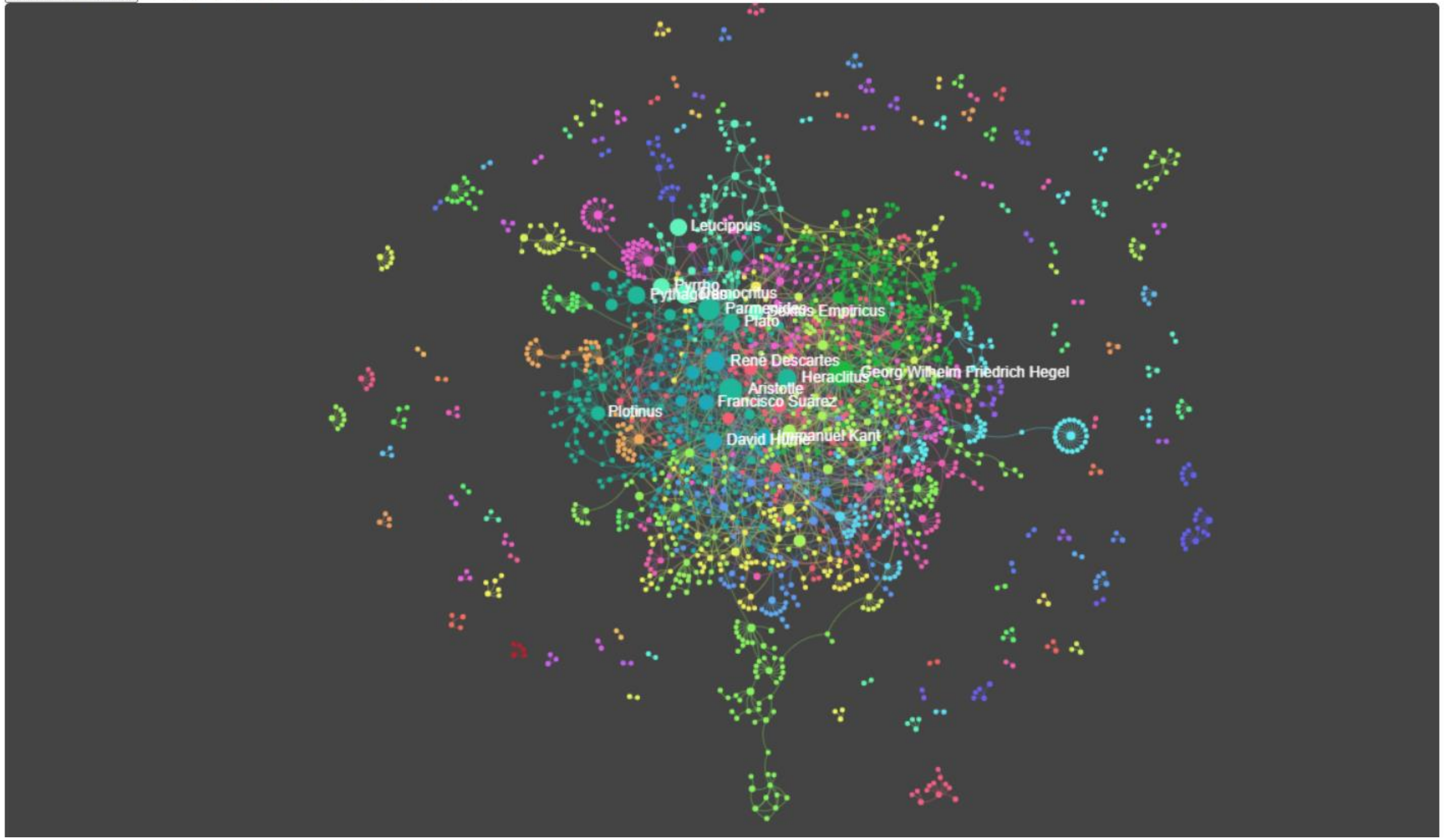
“In Ersilia, to establish the relationships that sustain the city's life, the inhabitants stretch strings from the corners of the houses, white or black or gray or black-and-white according to whether they mark a relationship of blood, of trade, authority, agency.[...] Thus, when traveling in the territory of Ersilia, you come upon the ruins of abandoned cities, without the walls which do not last, without the bones of the dead which the wind rolls away: spiderwebs of intricate relationships seeking a form.” (I. Calvino, *The invisible cities,* 1972)

Visualization

A Zoomable Graph of the History of Philosophy

(Is a bit of a mess.) Back to Design & Analytics. Or try this other one.

Rescale Graph Go ahead, zoom and click around.



Local level analysis

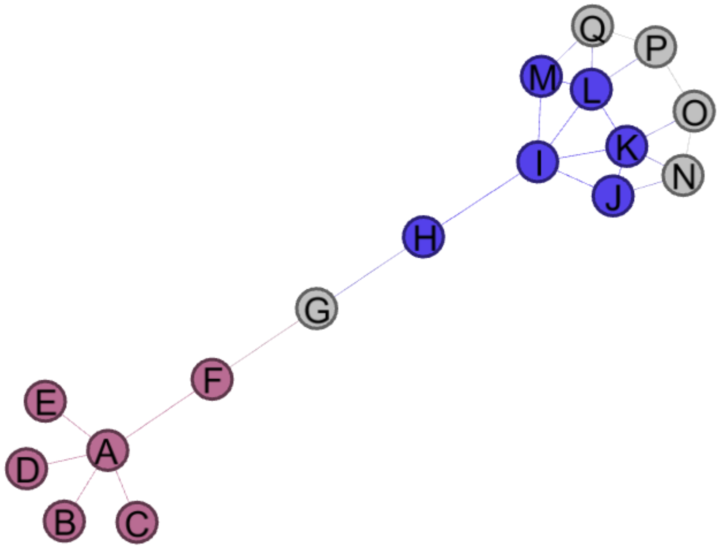
- Degree
- Centrality
- Community detection and clustering

Local level analysis: **degree**

- **Number of edges touching a node**
- In case of directed graph, we can distinguish **in-degree** (number edges whose TARGET is the node), and **out-degree** (number of edges whose SOURCE is the node)
- In case of weighted graphs: weighted (in or out) degree is the weighted sum of (incoming or outgoing) edges
- If we take the matrix representations this comes down to summing along columns or rows in the matrix

Local level analysis: degree (undirected)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
A	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	5
B	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
C	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
D	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
E	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
F	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	
G	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	
H	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	
I	0	0	0	0	0	0	0	1	0	1	1	1	1	0	0	0	0	
J	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	
K	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	0	
L	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	1	
M	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	
N	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	
O	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	
P	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	
Q	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	
	5																	

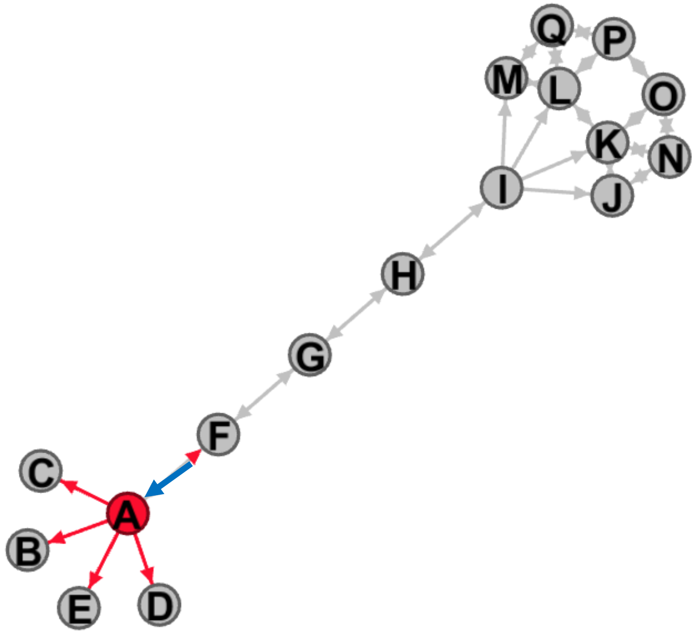


Local level analysis: degree (directed)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
A	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	1	0	1	1	1	1	0	0	0	0
J	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
K	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0
L	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1
M	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
N	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0
O	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0
P	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1
Q	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0

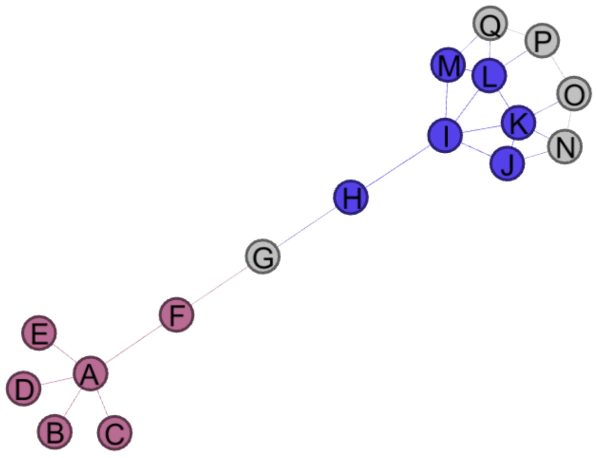
5

1



Local level analysis: degree centrality

- The degree is the simplest form of centrality measure: nodes with high degree are central to the graph in the sense that they have a high number of relations.
- However, this doesn't explain what "role" do they play in the structure of the network
- In this example, A, I and L have all the same degree (5), but are likely to have different functions in a "real network"



Local level analysis: centrality

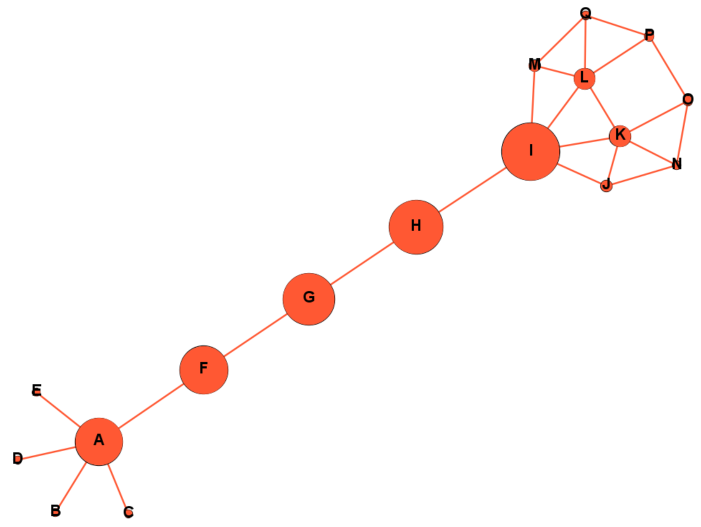
- Measuring the impact of different kinds of roles in the network
- Betweenness centrality
- Closeness centrality
- Eigenvector centrality (similar to Google pagerank)

Local level analysis: betweenness centrality

- Betweenness centrality measures the rate of shortest path passing through the node. It measures the function of the node as “bridge” in the network
- The betweenness of vertex A in a graph G:= (V,E), is computed as follows:

1. for each pair of vertices (s, t) compute the shortest paths between them
2. For each pair of vertices (s, t) determine the fraction of shortest paths that pass through the vertex A
3. Sum this fraction over all pairs of vertices (s, t)

source Wikipedia

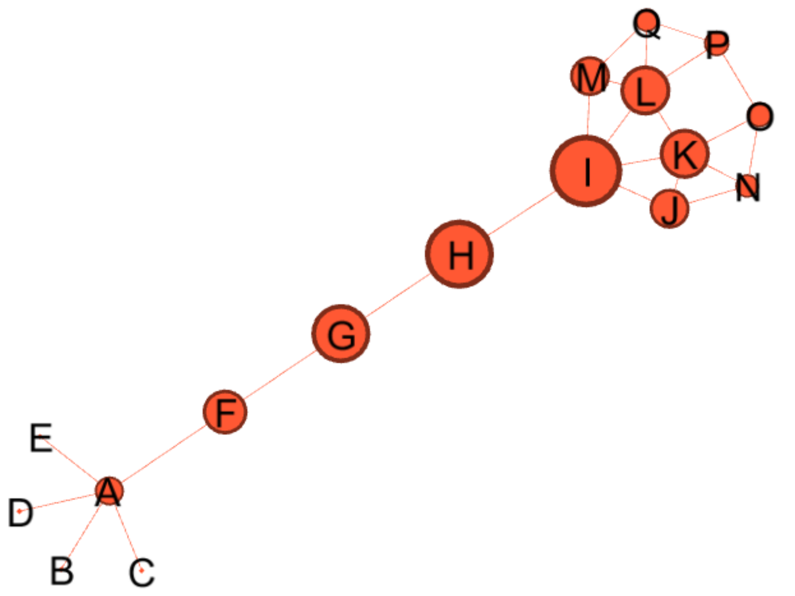


$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Local level analysis: closeness centrality

- Closeness centrality measures how close is one node to the rest of the network. It measures how "easy" it is for each node to reach the others
- The closeness centrality of vertex *A* in a graph *G*:= (V,E), is computed as follows:

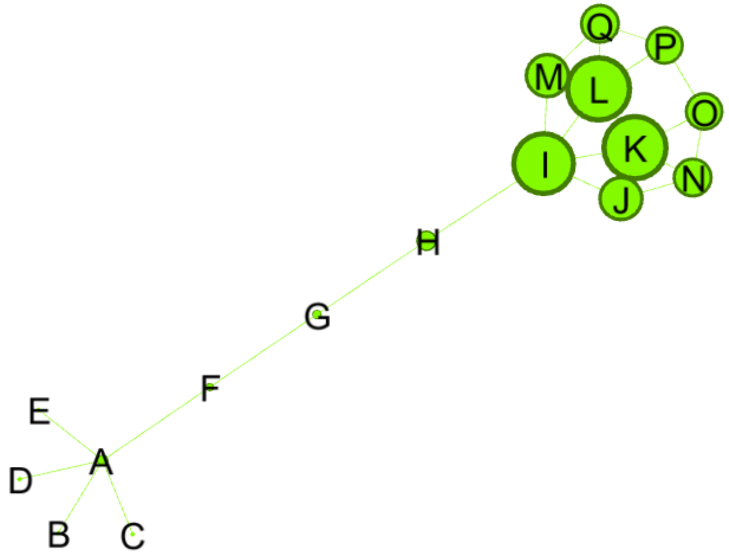
1. For each node *t* present in the graph, you compute the shortest path from *A* to *t*, and thus calculate the distance.
2. You then add-up the inverse of the distance *d* of each of these nodes (1/*d*)
3. You can normalize with respect to the size of the graph



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

Local level analysis: eigenvector centrality

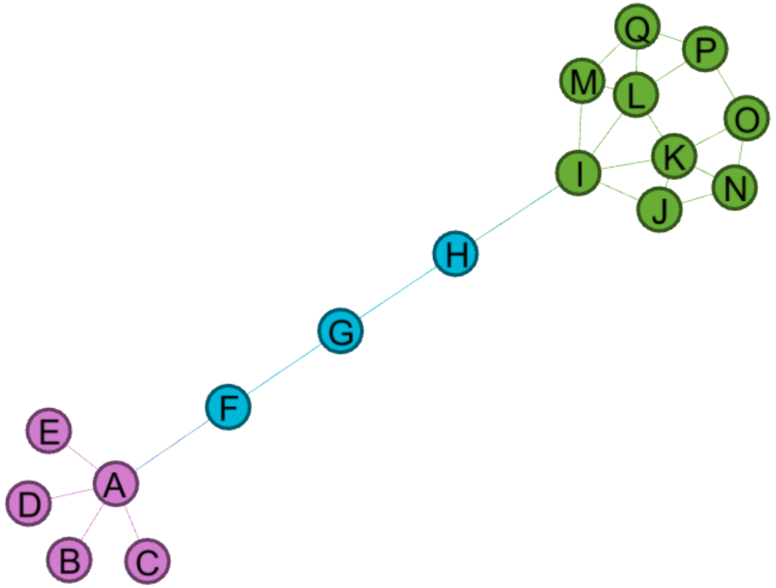
- Eigenvector centrality measures the number and quality of the connections of a node. A node has high centrality if its connection are also very influent in the network.
- Computation spared!



Local level analysis: community detection and clustering

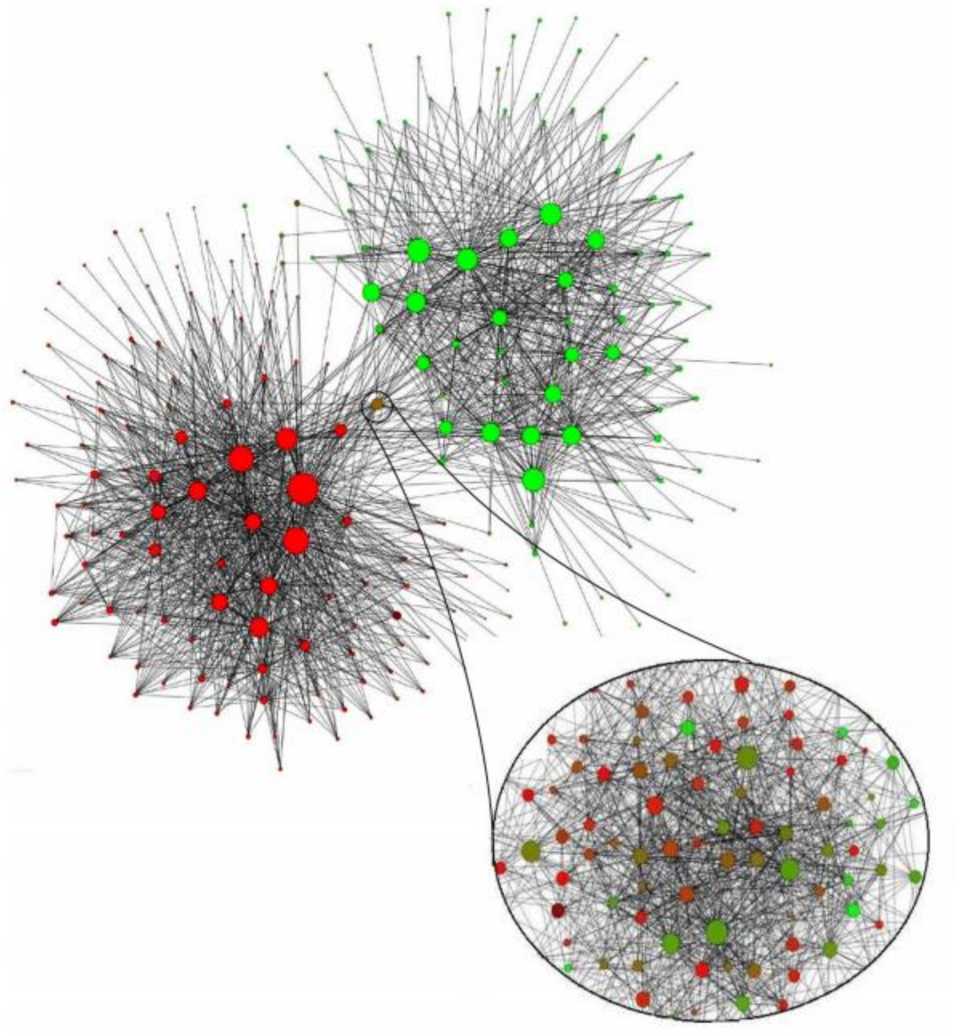
- Goal: identifying clusters of nodes that are internally strongly connected, and loosely connected with the other communities. Very useful to identify group of actors or words that tend to interact/cooccur a lot.

- Computation spared!



Local level analysis: community detection and clustering

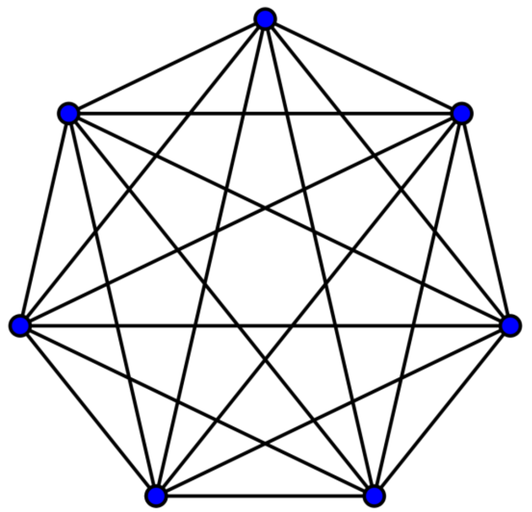
- Application of "Louvain algorithm"
- Source: 0803.0436.pdf (arxiv.org) (p. 6)
- Belgian mobile phone network: Flemish (green) and Wallonian (red) communities



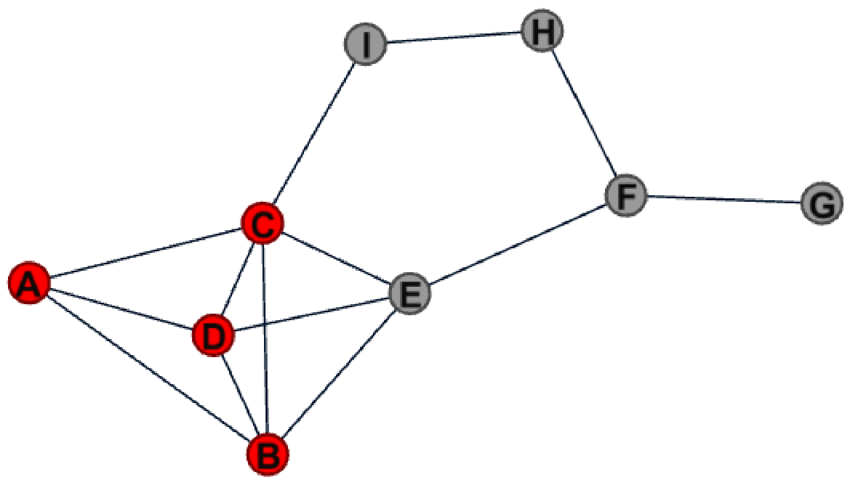
Global level analysis

- Focus on general structure of the network and in its partition in large sub-graphs
- Fully connected/components
- Density and degree distribution
- Diameter

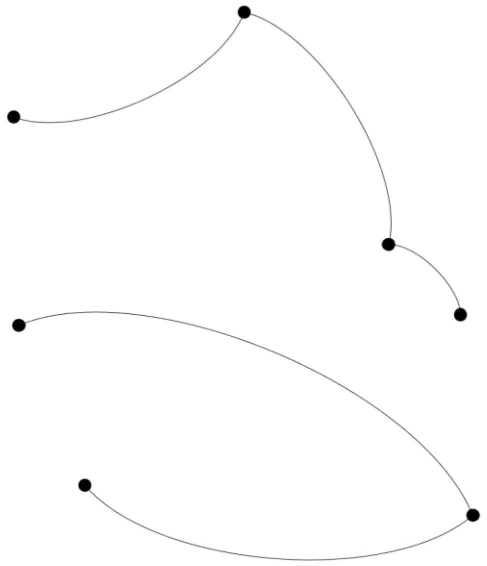
Global level analysis: cliques and components



Fully connected graph



Complete subgraph: clique



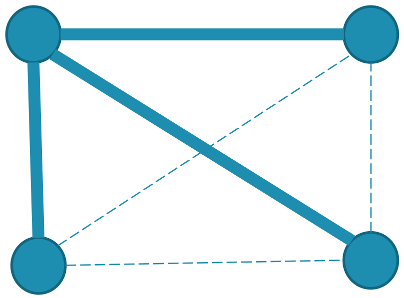
Treated as distinct graphs

Global level analysis: density

- Density: $\frac{m}{\frac{n(n-1)}{2}}$

with m, number or edges, and n number of nodes: number of edges over number of possible edges

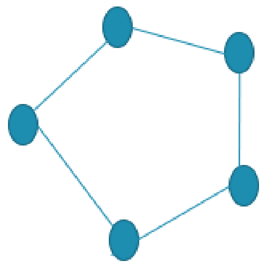
- $\frac{3}{\frac{4(4-1)}{2}} = \frac{3}{6} = 0.5$



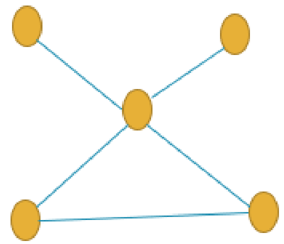
Global level analysis: average degree and degree distribution

- Average degree: average number of links per node.
  - Undirected graphs: 2E/N
  - Directed graphs: E/N

where E is the number of edges, and N the number of nodes

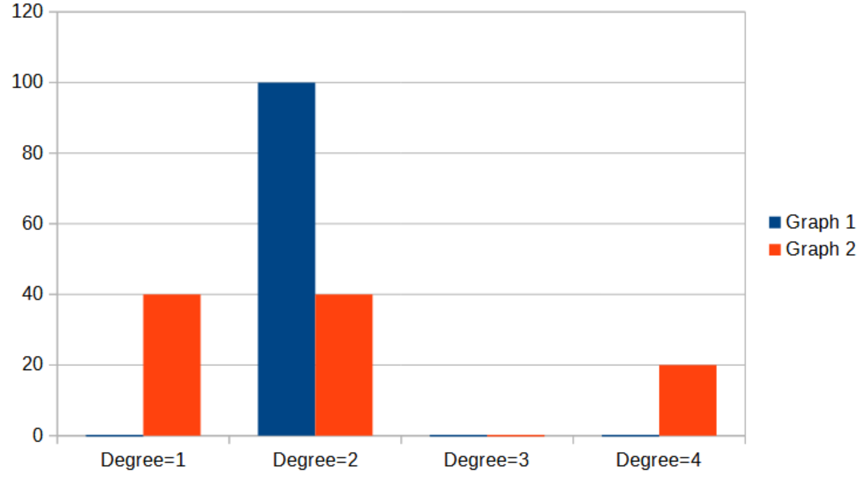


Average degree: 2



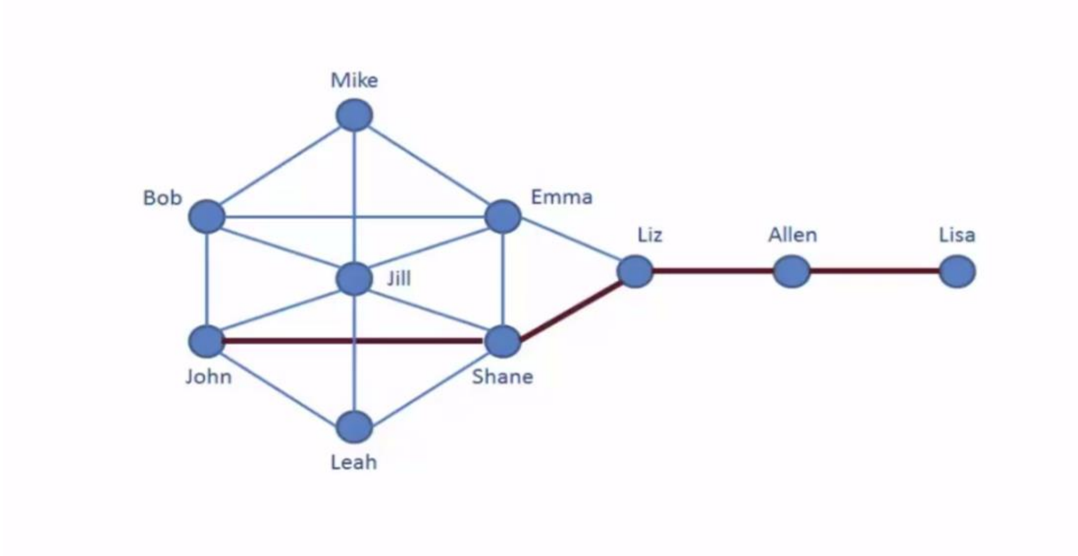
Average degree: 2

- Degree distribution: percentage of nodes having degree *k*



Global level analysis: diameter

- Shortest distance between the two most distant nodes in the network.
- You calculate the shortest path between each pair of nodes, and then pick the highest value.



diameter= 4

friendsarena: Image (wordpress.com)

The end

Thank you very much for the attention!