Networks

Elsa Arcaute

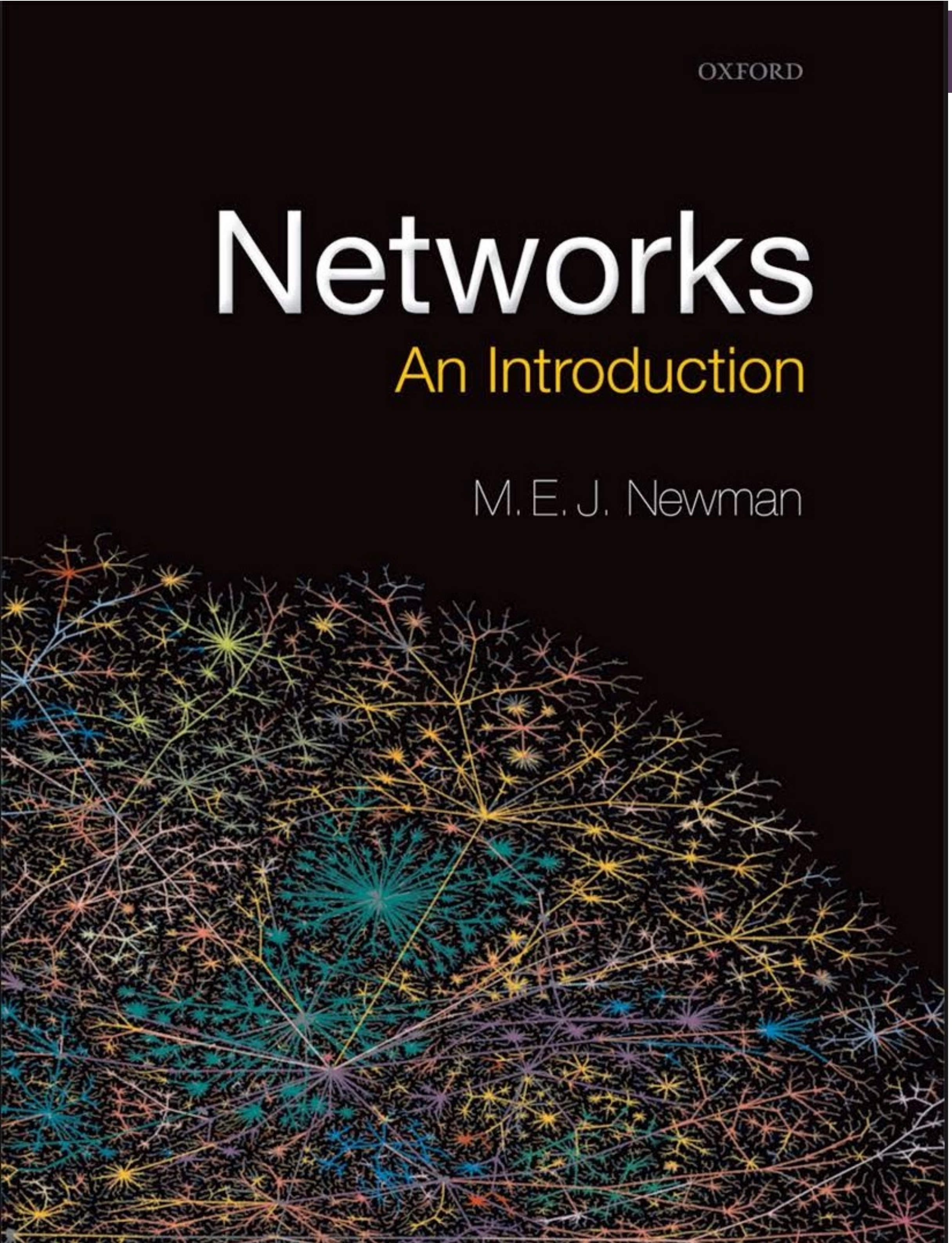
BENVGSC5: Urban Simulation

Week 6: 19th February 2024

Outline

1. Definitions: node, link, degree, etc.
2. Undirected, weighted, directed networks
3. Adjacency matrix, paths and connectivity
4. Centrality measures: closeness, betweenness, etc.
5. Clustering, similarity and modularity
6. Degree distribution
7. Scale-free networks
8. Preferential attachment
9. Random graphs
10. Small-world
11. Community detection
12. Spatial networks

Today



Course based on Newman’s book: Networks: An Introduction (2010)

Many slides are taken from Barabasi’s book and prepared slides

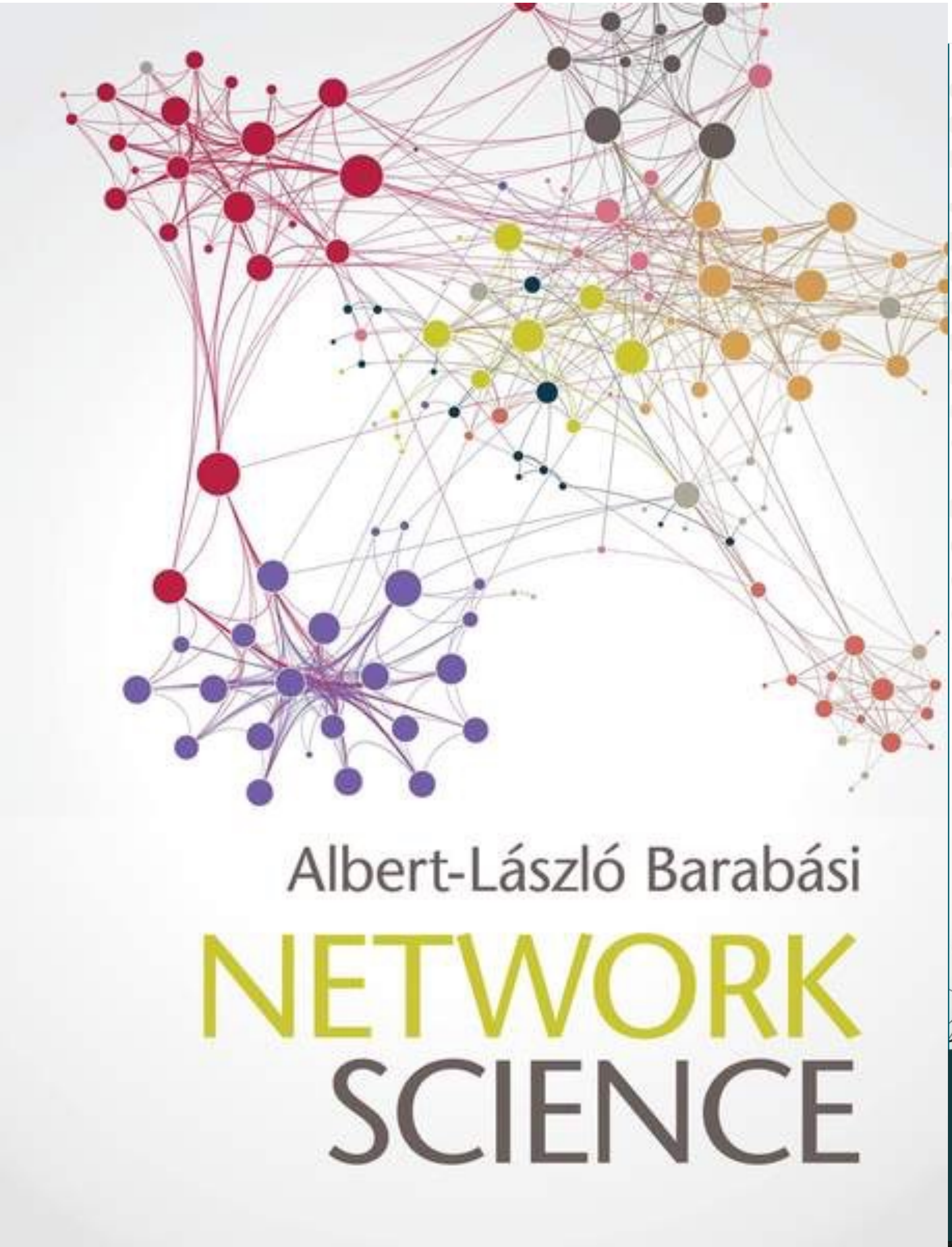
http://barabasi.com/networksciencebook/
https://www.barabasilab.com

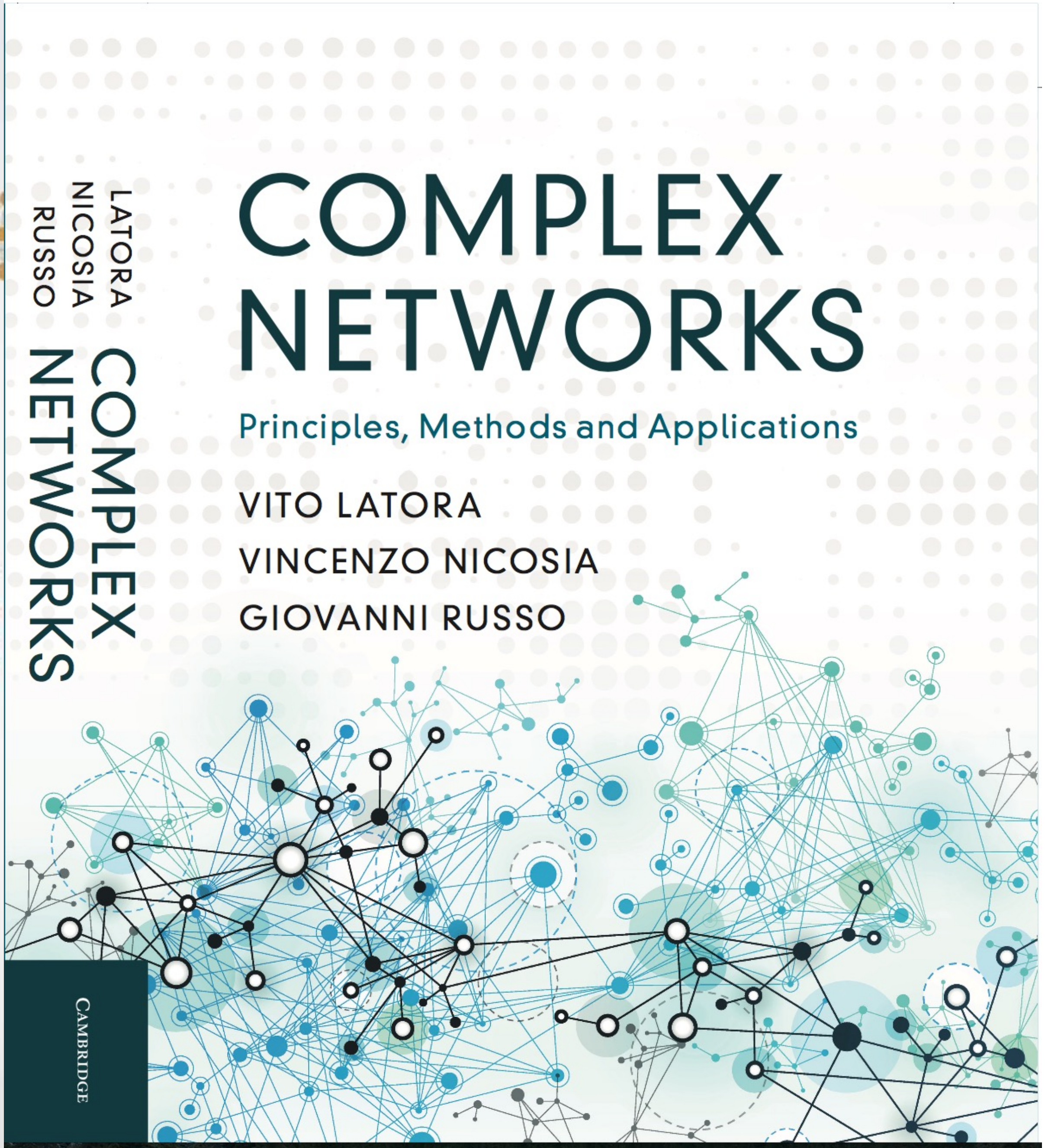
Vito Latora course on networks at QMUL

http://www.maths.qmul.ac.uk/~latora/index.html

Other books for reference:

http://barabasi.com/networksciencebook/ http://www.maths.qmul.ac.uk/~latora/index.html





Motivation

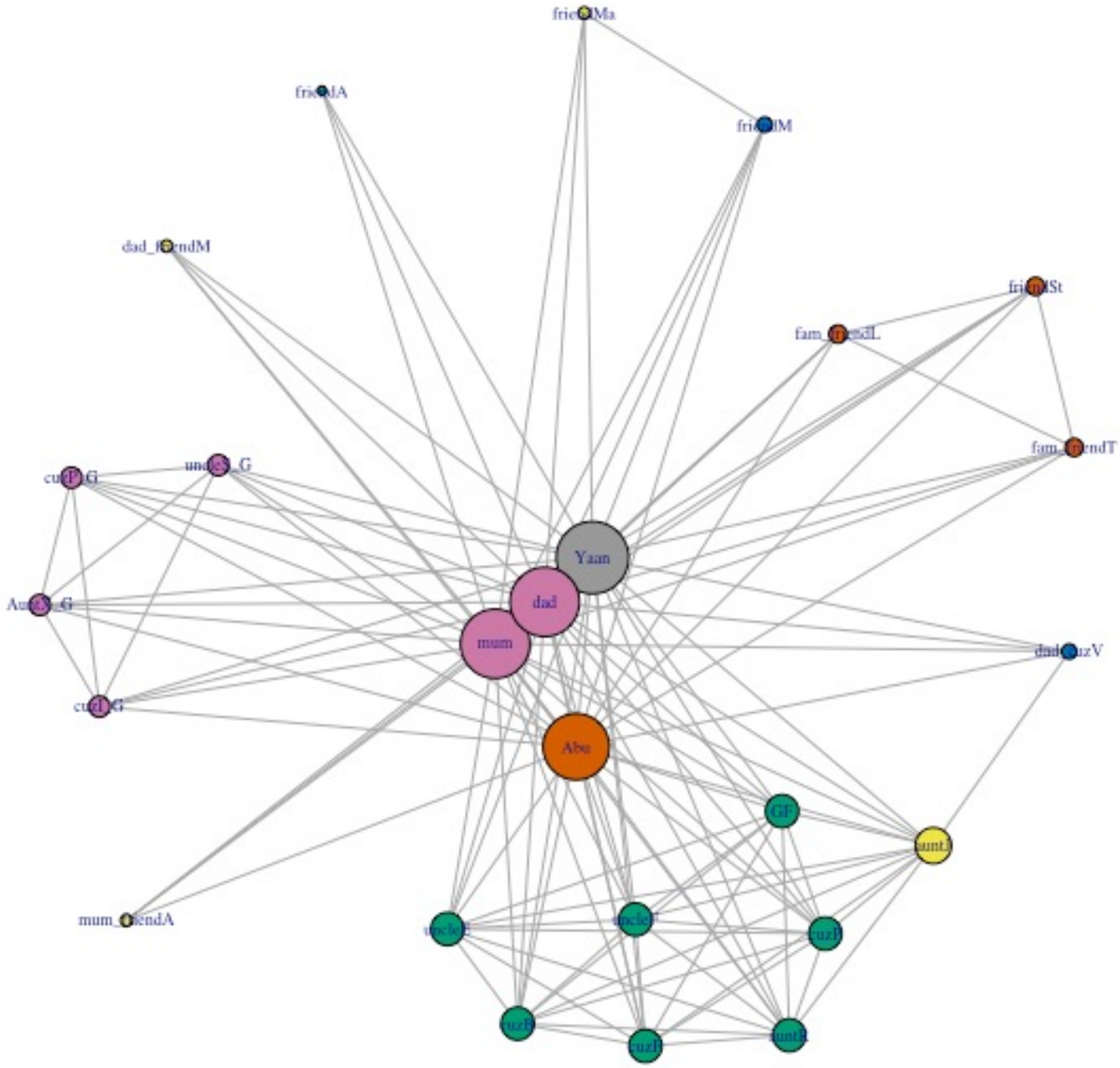
Generic way to encode agents and interactions in complex systems:

→ if we understand the dynamics of one system, can extrapolate to a different system that behaves in the same way: generic mechanisms

e.g.

→ the internet
→ relationships between individuals
→ food webs: prey-predator webs
→ protein-protein interaction
→ transport networks
→ etc.

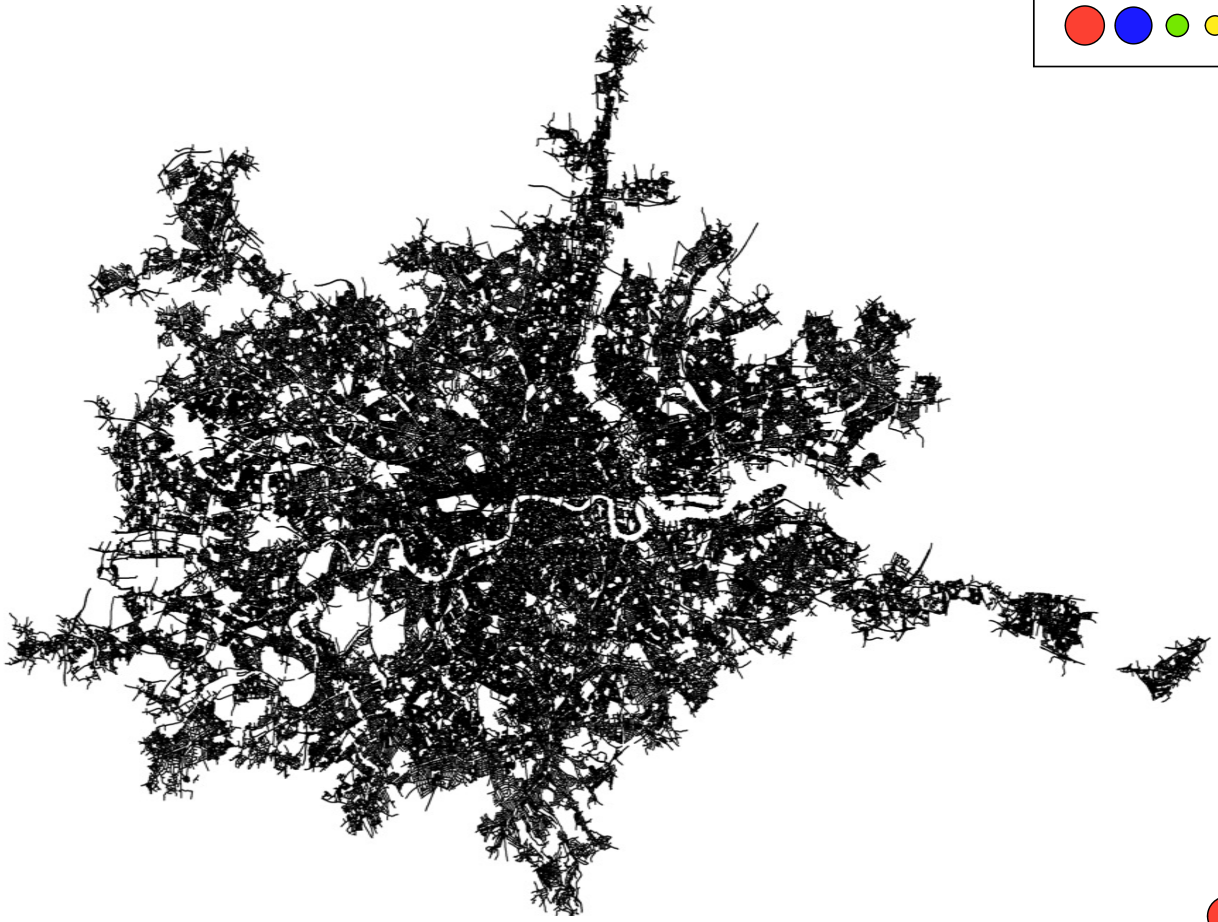
Smallest social network of Yaan’s dad, perceived by him



Networks extracted from surveys in the social sciences: **ego-networks**

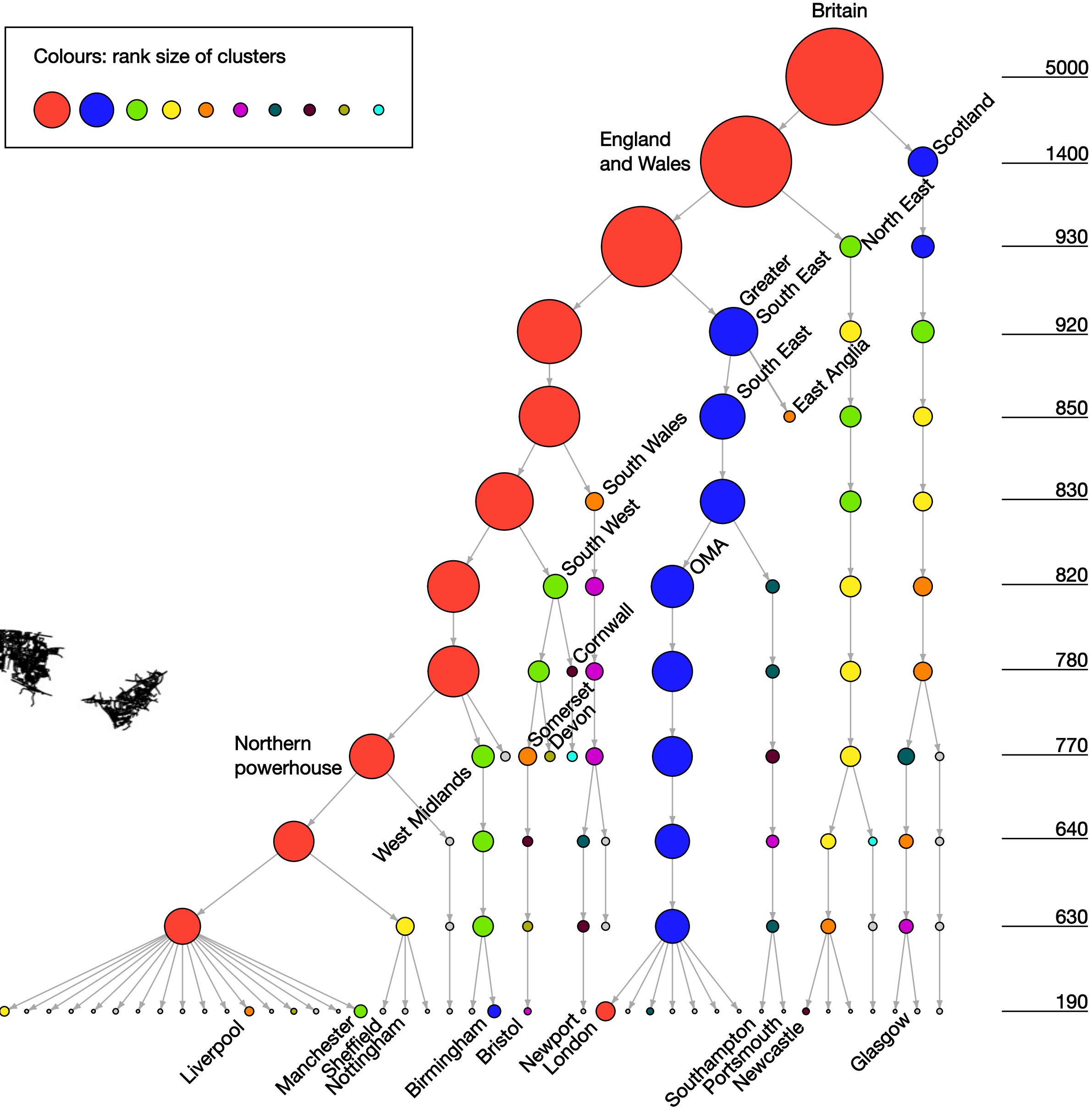
E.g. of networks

Street Network

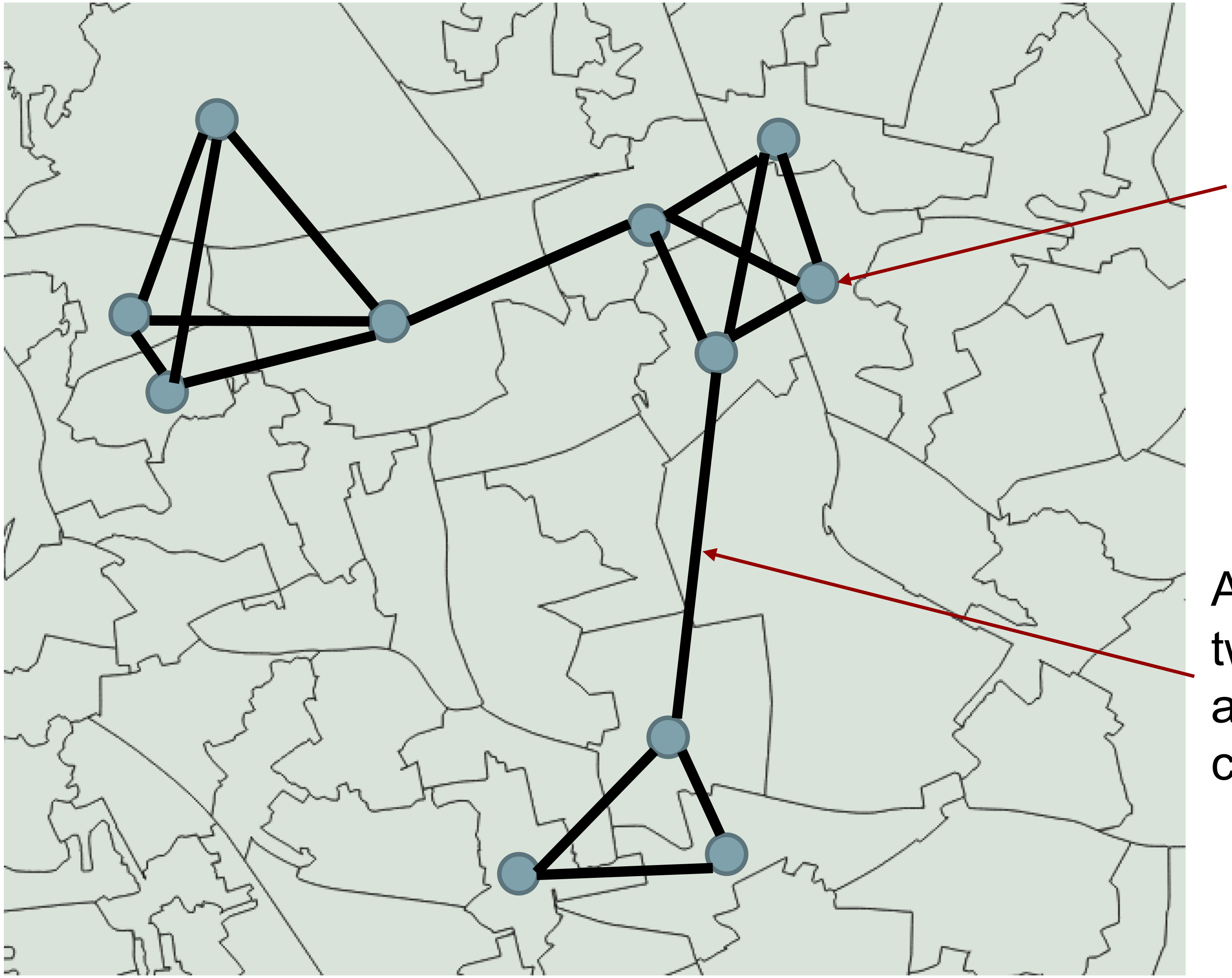


Nodes: street intersections
Links: streets

Hierarchical organisation of Britain: Infrastructural connectivity



Network of relationships between places



Each area can be represented by its centroid, which will be considered as a node in the network

A link between the two areas is defined according to desired characteristics

Definitions: mathematical representation of a network (*n,m*)

→ Most of the time a system might be represented in various ways through networks, depending on what aspect is to be analysed.

Network ⇔ Graph: *G(n,m)*
Node ⇔ vertex: *n*
Link ⇔ edge: *m*

Edge list: list of links between nodes (*i*,*j*)

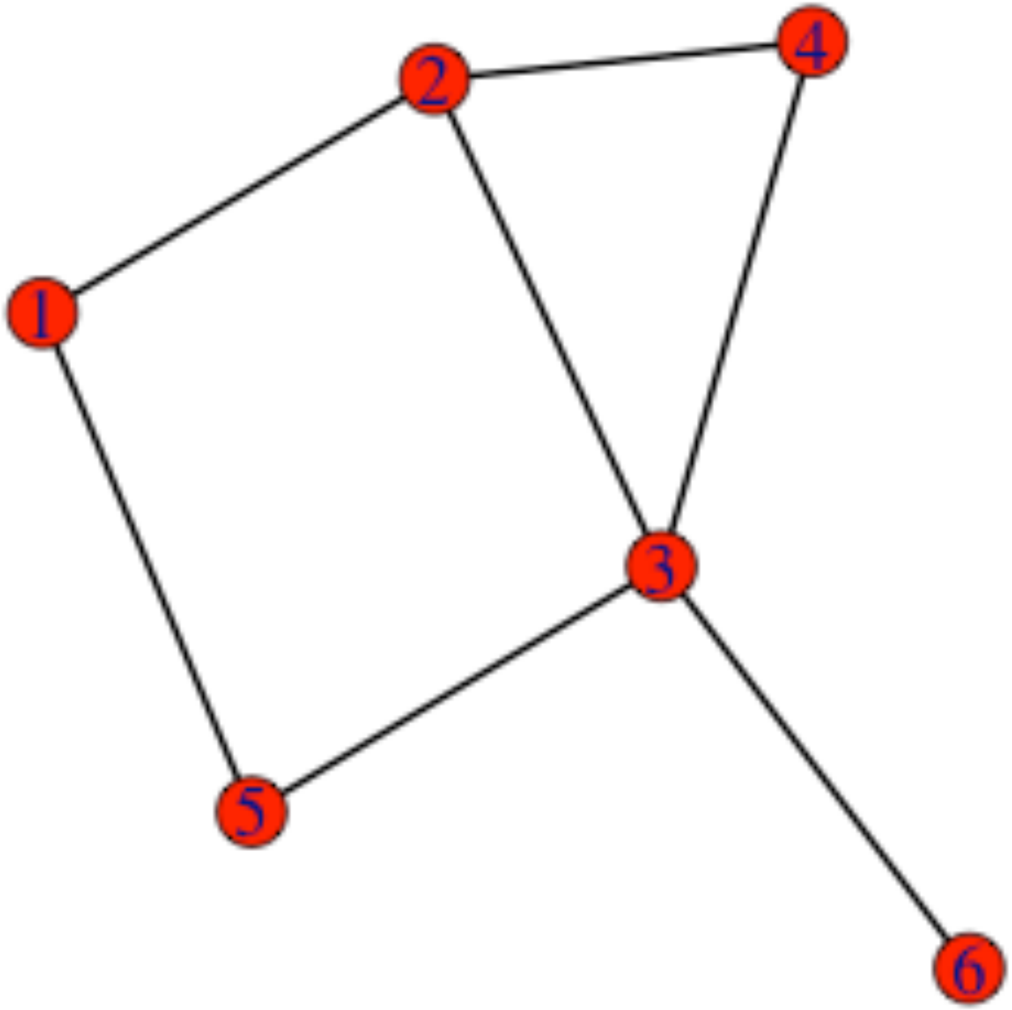
or

Adjacency matrix: ***A*** gives the value of the links. Simplest unweighted case:

$$A_{ij} = \begin{cases} 1 & \text{If nodes } i \text{ and } j \text{ are connected} \\ 0 & \text{If there's no link between them} \end{cases}$$

Example: undirected simple network

n=6: n. of nodes (vertices)
m=7: n. of links (edges)



Edge list: how many pairs do I need listed?

***m*=7**

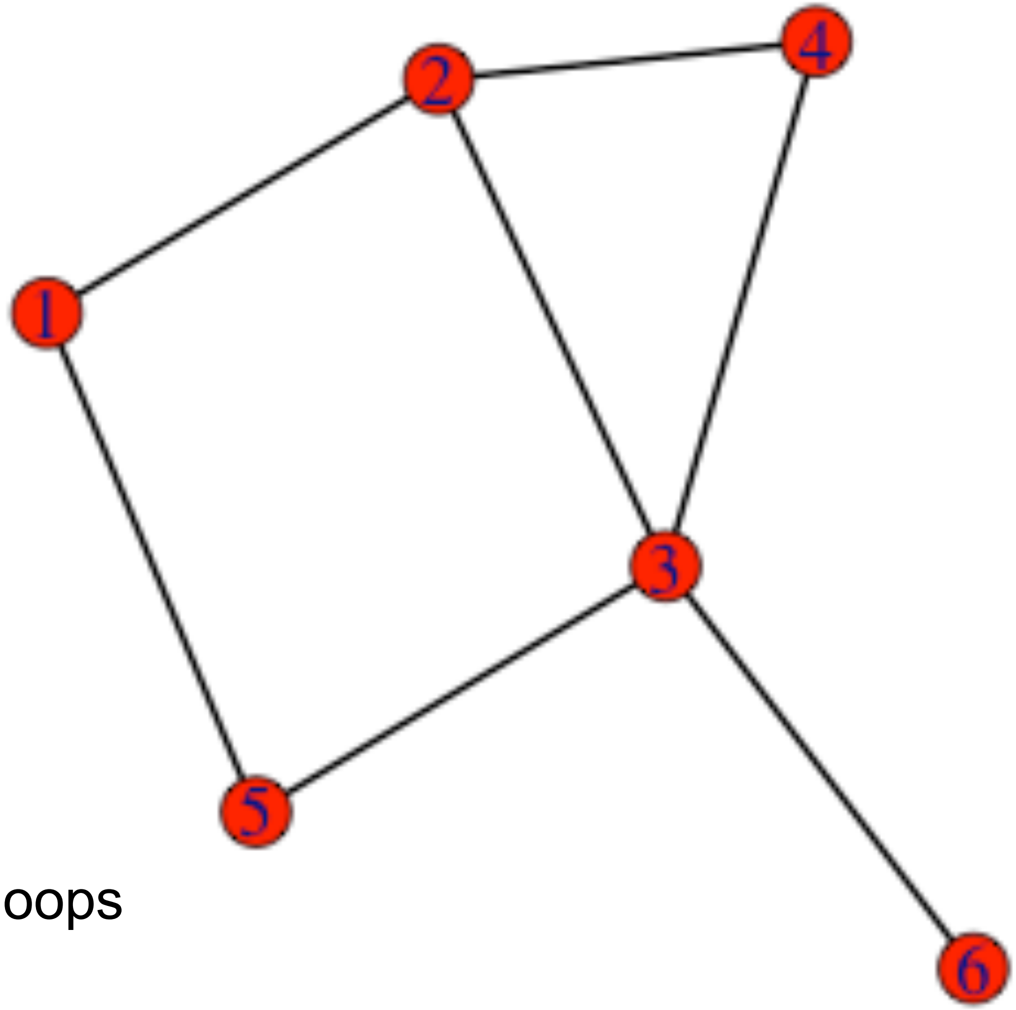
Edge list: (1,2), (1,5), (2,3), (2,4), (3,4), (3,5), and (3,6)

Example: undirected simple network

Adjacency matrix:

$$\begin{array}{rcccccc|l}
\text{Nodes} \rightarrow & 1 & 2 & 3 & 4 & 5 & 6 & \text{Nodes} \downarrow \\
 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
 & 1 & 0 & 1 & 1 & 0 & 0 & 2 \\
 & 0 & 1 & 0 & 1 & 1 & 1 & 3 \\
A = & 0 & 1 & 1 & 0 & 0 & 0 & 4 \\
 & 1 & 0 & 1 & 0 & 0 & 0 & 5 \\
 & 0 & 0 & 1 & 0 & 0 & 0 & 6
\end{array}$$

No self-loops



→ **symmetrical!!!**

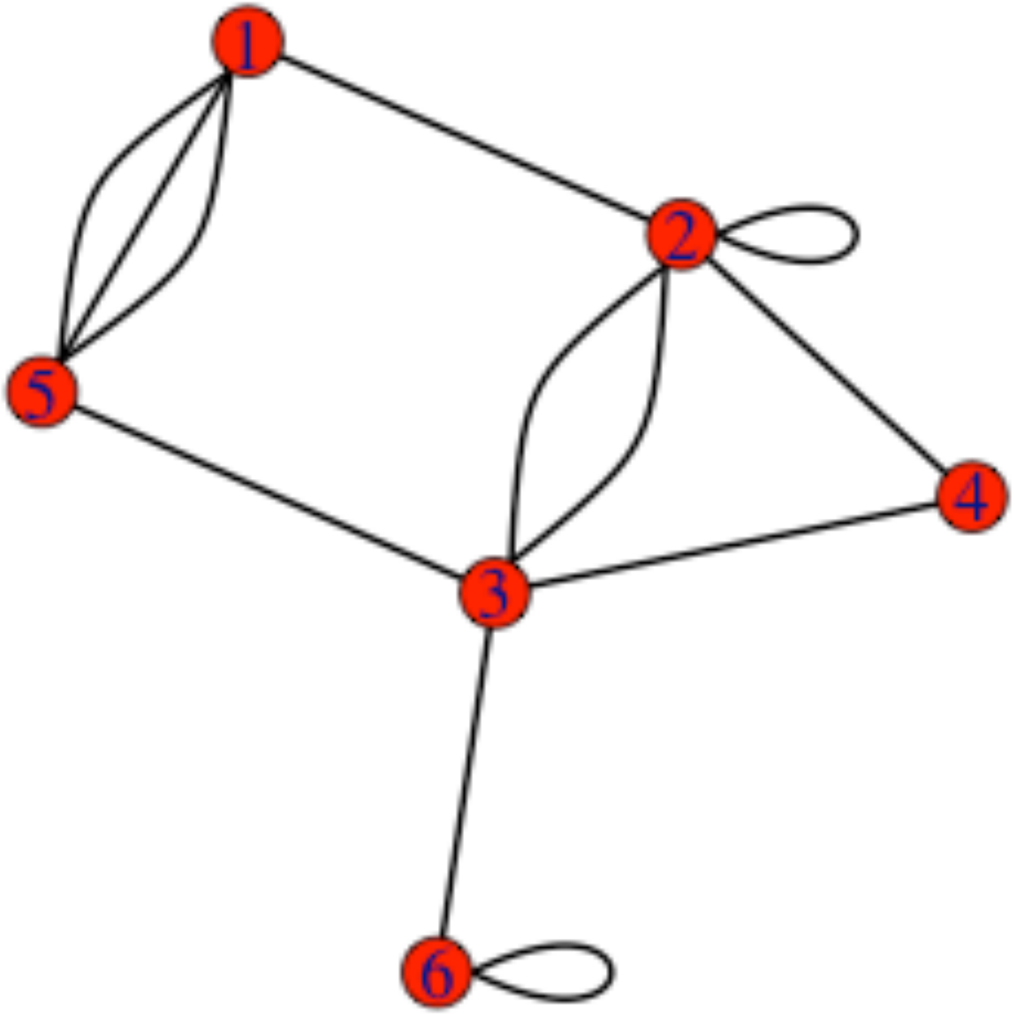
→ **Only 1s and 0s since no weights considered**

Example: undirected multi-edge network

n=6: n. of nodes (vertices)

How many links (edges) do I have now?

m**=12**



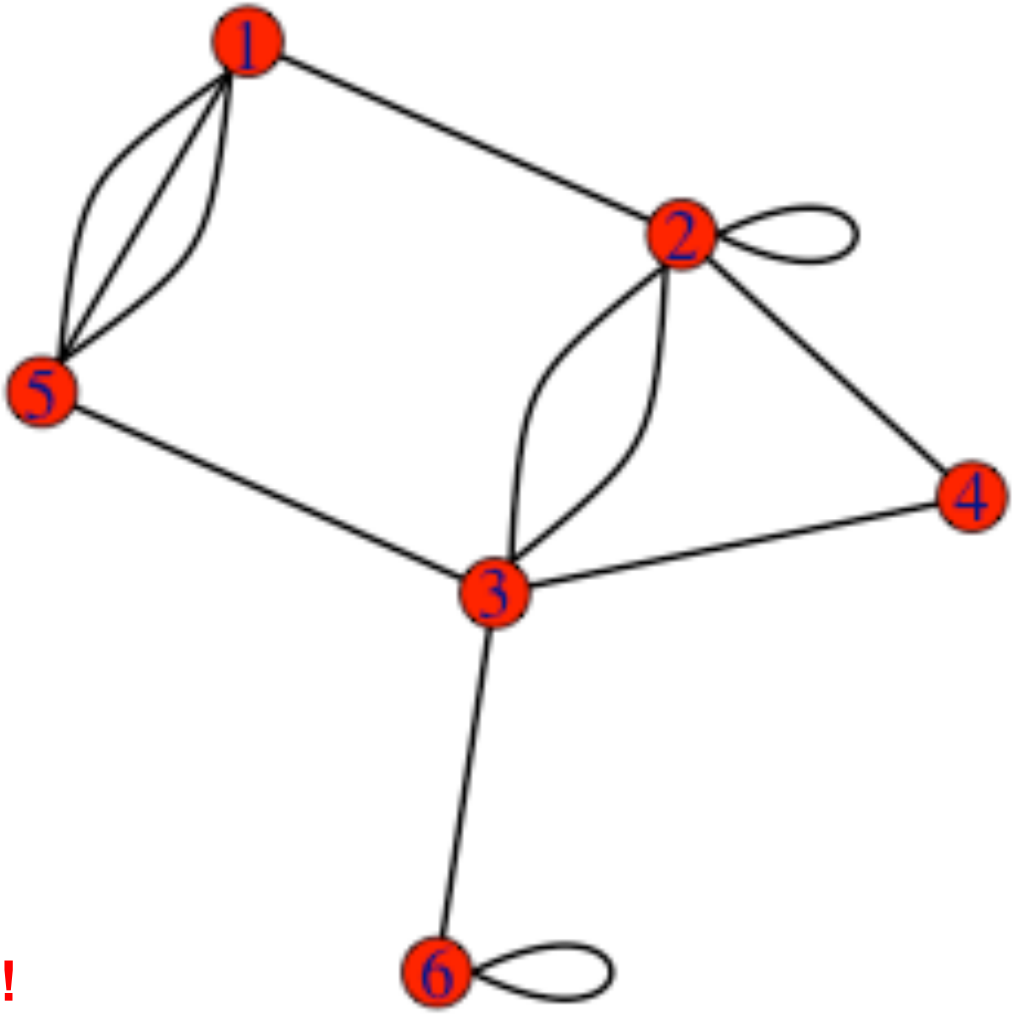
→ Different links (edges) between nodes can represent e.g. different relationships between agents, multi-modal transport networks, etc.

→ Particular care must be taken if loops are present

Example: undirected multi-edge network

Adjacency matrix:

$$\begin{array}{cccccccc} \text{Nodes} \rightarrow & 1 & 2 & 3 & 4 & 5 & 6 & \text{Nodes} \downarrow \\ & 0 & 1 & 0 & 0 & 3 & 0 & 1 \\ & 1 & 2 & 2 & 1 & 0 & 0 & 2 \\ & 0 & 2 & 0 & 1 & 1 & 1 & 3 \\ A = & 0 & 1 & 1 & 0 & 0 & 0 & 4 \\ & 3 & 0 & 1 & 0 & 0 & 0 & 5 \\ & 0 & 0 & 1 & 0 & 0 & 2 & 6 \end{array}$$



→ **Value for loops needs to be set to 2!!**

→ **Matrix symmetrical since undirected**

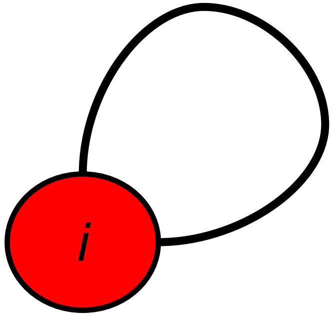
→ **No weights for links, value only corresponds to number of links**

Loops

Why diagonal for undirected graphs containing loops are 2s or 0s?

→ If a link between i and j exists $A_{ij}=A_{ji}=1$. This term will appear twice in the adjacency matrix.

→ In the limit of *j* becoming closer and closer to *i*, we would get $j \sim i$, and a loop will emerge. In this case the node *i* will have not only one end, as was the case of $A_{ij}=1$, but two ends! Hence $\mathbf{A_{ii}=2}$.



Weighted networks

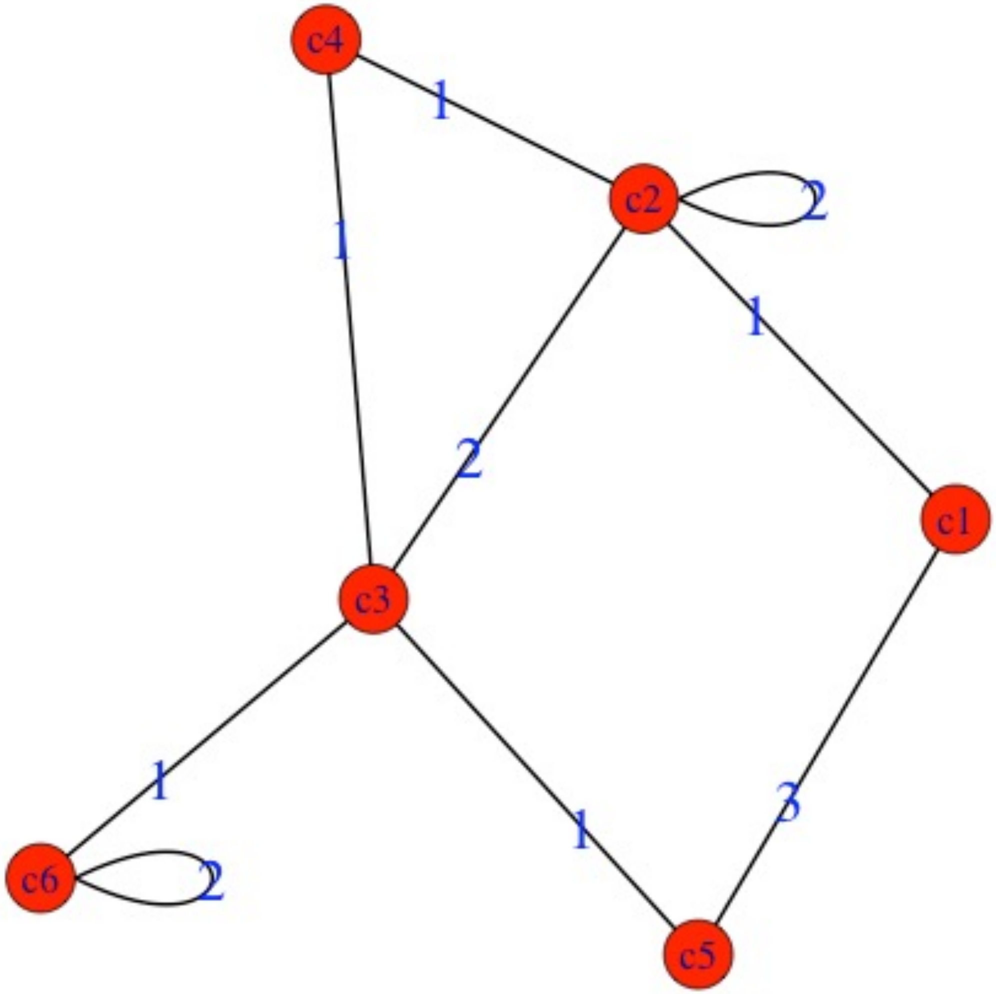
Adjacency matrix:

Nodes →

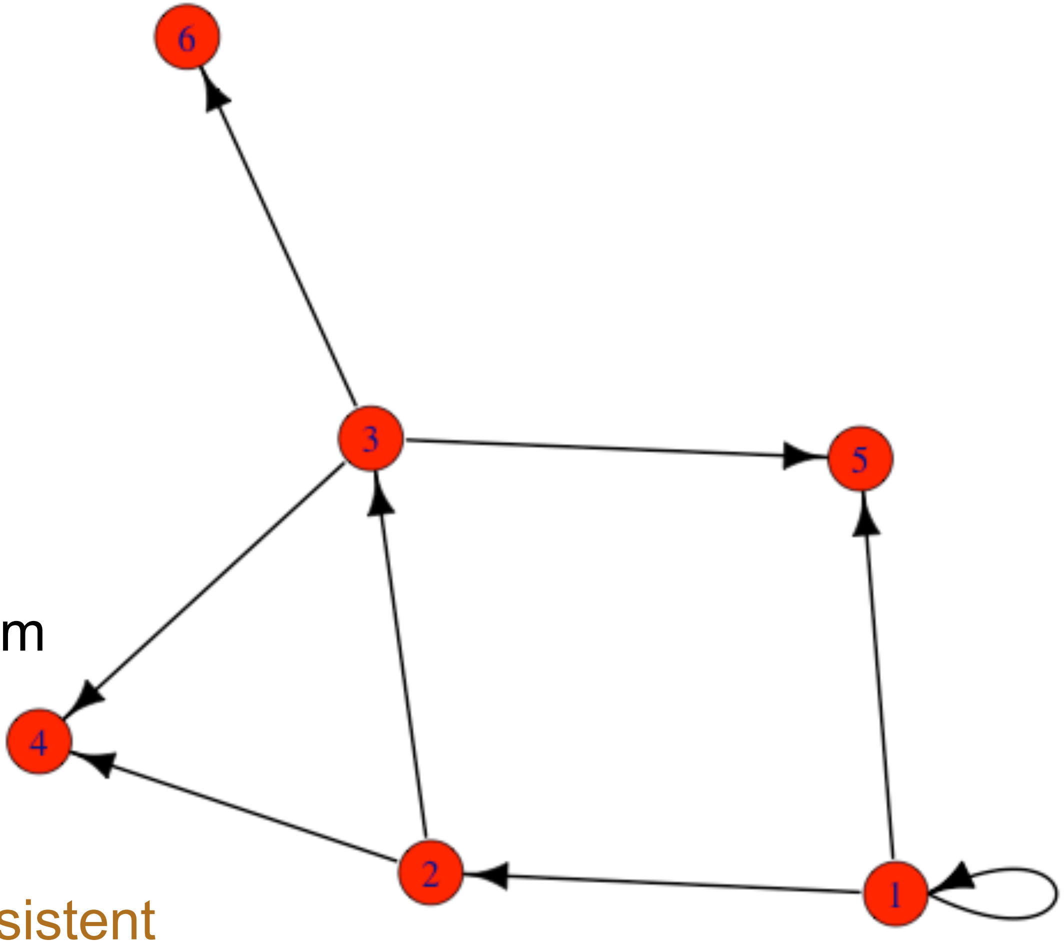
Nodes ↓

$$A = \begin{array}{cccccc|c} c1 & c2 & c3 & c4 & c5 & c6 & \\ 0 & 1 & 0 & 0 & 3 & 0 & c1 \\ 1 & 2 & 2 & 1 & 0 & 0 & c2 \\ 0 & 2 & 0 & 1 & 1 & 1 & c3 \\ 0 & 1 & 1 & 0 & 0 & 0 & c4 \\ 3 & 0 & 1 & 0 & 0 & 0 & c5 \\ 0 & 0 & 1 & 0 & 0 & 2 & c6 \end{array}$$

→ Still symmetrical



Directed network



Links have now a direction

→ ATTENTION!!! Notation for elements of adjacency matrix $\boldsymbol{A_{ij}}$:

$$A_{ij} = \begin{cases} 1 & \text{If there's an edge from } j \text{ to } i \\ 0 & \text{If there's no link between them} \end{cases}$$

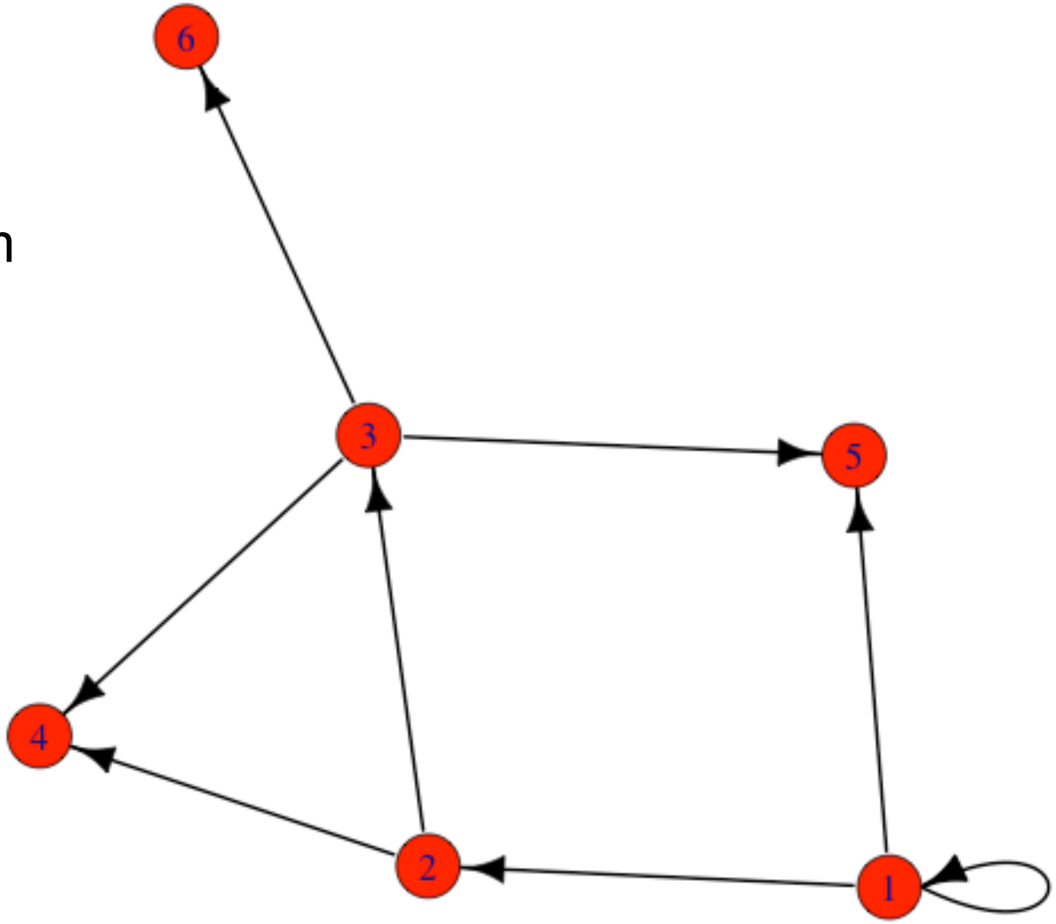
This is just a convention, just need to be consistent

→ This convention is different to the convention used for the OD matrices in Spatial Interaction models and for NetworkX and iGraph

Directed network

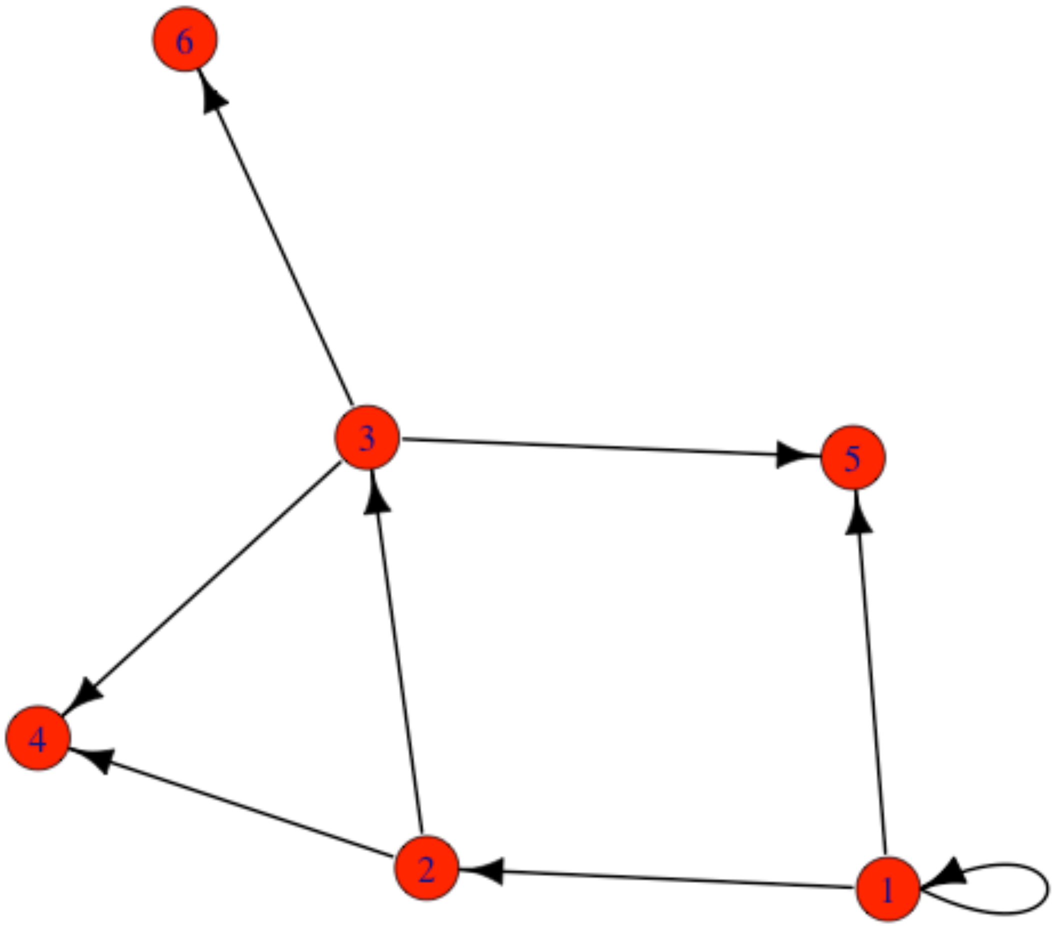
$$A_{ij} = \begin{cases} 1 & \text{If there's an edge from } j \text{ to } i \\ 0 & \text{If there's no link between them} \end{cases}$$

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & A_{16} \\ A_{21} & A_{22} & A_{23} & \cdots & \cdots & A_{26} \\ A_{31} & A_{32} & \cdots & \cdots & \cdots & \cdots \\ A_{41} & \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{51} & \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{61} & \cdots & \cdots & \cdots & \cdots & A_{66} \end{pmatrix}$$



Directed network

$$\begin{array}{c|cccccc|c}
\text{Nodes} \rightarrow & A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & A_{16} & \text{Nodes} \downarrow \\
\hline
 & 1 & 0 & 0 & 0 & 0 & 0 & A_{11} \ldots \\
 & 1 & 0 & 0 & 0 & 0 & 0 & A_{21} \ldots \\
A = & 0 & 1 & 0 & 0 & 0 & 0 & A_{31} \ldots \\
 & 0 & 1 & 1 & 0 & 0 & 0 & A_{41} \ldots \\
 & 1 & 0 & 1 & 0 & 0 & 0 & A_{51} \ldots \\
 & 0 & 0 & 1 & 0 & 0 & 0 & A_{61} \ldots
\end{array}$$

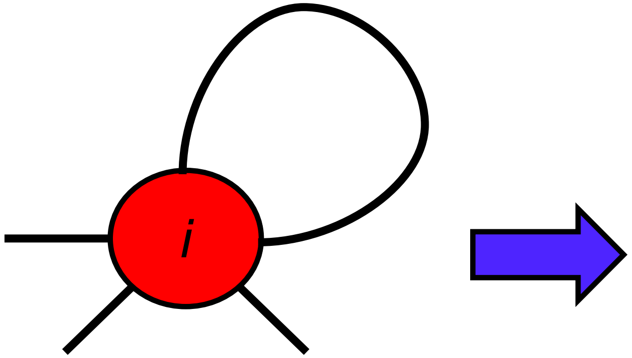


→ **No longer symmetrical!!!**

→ **Loop has value 1 in this case!!!**

Degree of a node: undirected case

The degree of a node is given by the number of links attached to it.
→ The degree of a node *i* is denoted as k_i



k_i=5

$$\boxed{k_i = \sum_{j=1}^{n} A_{ij}}$$

e.g.

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & A_{16} \\ A_{21} & A_{22} & A_{23} & \dots & \dots & A_{26} \\ A_{31} & A_{32} & \dots & \dots & \dots & \dots \\ A_{41} & \dots & \dots & \dots & \dots & \dots \\ A_{51} & \dots & \dots & \dots & \dots & \dots \\ A_{61} & \dots & \dots & \dots & \dots & A_{66} \end{pmatrix} \longrightarrow k_5 = \sum_{j=1}^{6} A_{5j} = A_{51} + \dots + A_{56}$$

Mean degree, number of links: undirected case

In an undirected graph, each link has associated two end points.
→ Each link contributes to two nodes' degree
→ Total sum of all degrees = twice the number of links

$$2m = \sum_{i=1}^{n} k_i \quad \Rightarrow \quad m = \frac{1}{2}\sum_{i=1}^{n} k_i = \frac{1}{2}\sum_{ij} A_{ij}$$

Mean degree <*k*> of a node, recall m=number of links and n=number of nodes

$$\boxed{< k >= \frac{1}{n}\sum_{i=1}^{n} k_i = \frac{2m}{n}}$$

Degree directed case: in/out-degree

$$k_i^{in} = \sum_{j=1}^{n} A_{ij}$$

$$k_j^{out} = \sum_{i=1}^{n} A_{ij}$$

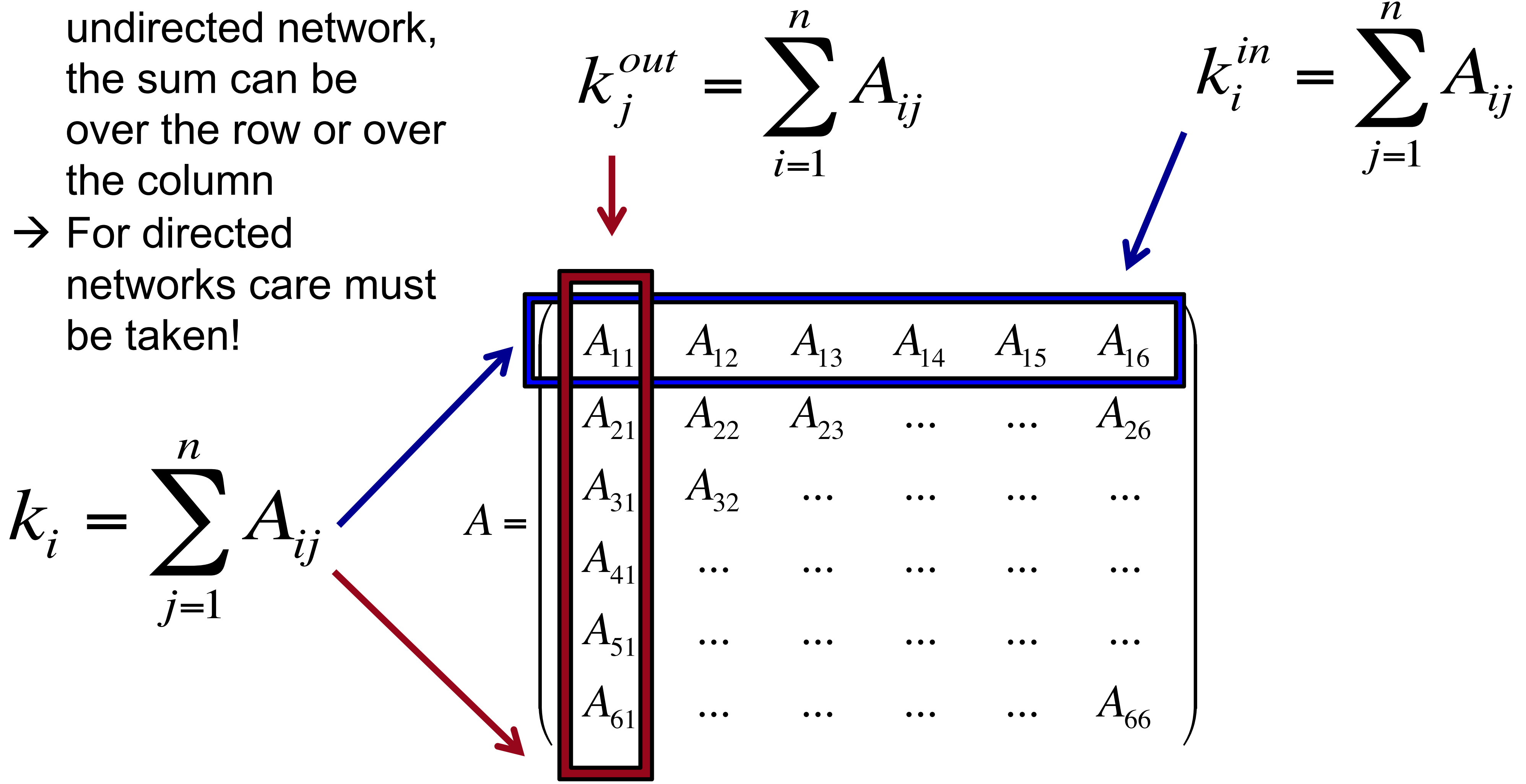
Number of links pointing to you

Number links you point to others

For a directed graph you need to consider these two degrees separately

In/out-degree

- → Since the matrix is symmetric for an undirected network, the sum can be over the row or over the column
- → For directed networks care must be taken!



$$k_j^{out} = \sum_{i=1}^{n} A_{ij}$$

$$k_i^{in} = \sum_{j=1}^{n} A_{ij}$$

$$k_i = \sum_{j=1}^{n} A_{ij}$$

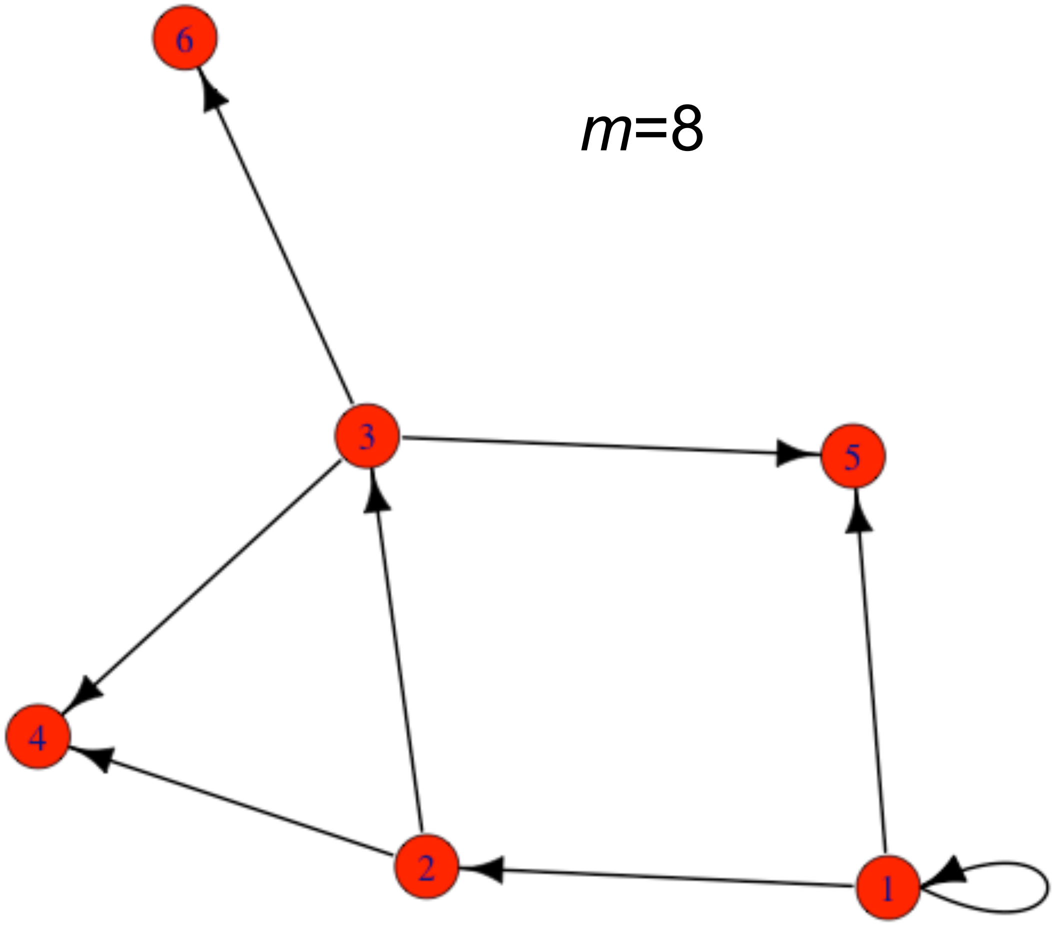
$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & A_{16} \\ A_{21} & A_{22} & A_{23} & \cdots & \cdots & A_{26} \\ A_{31} & A_{32} & \cdots & \cdots & \cdots & \cdots \\ A_{41} & \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{51} & \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{61} & \cdots & \cdots & \cdots & \cdots & A_{66} \end{pmatrix}$$

Mean degree, number of links: directed case

In a directed graph, each link contributes to ONLY ONE node's degree: **in OR out**
→ Total sum of all **in or out degrees** = total number of links

$$m = \sum_{i=1}^{n} k_i^{in} = \sum_{j=1}^{n} k_j^{out} = \sum_{ij} A_{ij}$$

$$A = \begin{array}{cc} \begin{array}{cccccc} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & A_{16} \end{array} & \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} & \begin{array}{c} A_{11} \ldots \\ A_{21} \ldots \\ A_{31} \ldots \\ A_{41} \ldots \\ A_{51} \ldots \\ A_{61} \ldots \end{array} \end{array}$$



Mean degree, number of links: directed case

In a directed graph, each link contributes to ONLY ONE node's degree: **in OR out**
→ Total sum of all **in or out degrees** = total number of links

$$m = \sum_{i=1}^{n} k_i^{in} = \sum_{j=1}^{n} k_j^{out} = \sum_{ij} A_{ij}$$

Mean in/out-degree of a node: $<k^{in}> = <k^{out}> = <k>$

$$< k^{in} >= \frac{1}{n}\sum_{i=1}^{n} k_i^{in} = \frac{1}{n}\sum_{j=1}^{n} k_j^{out} =< k^{out} >$$

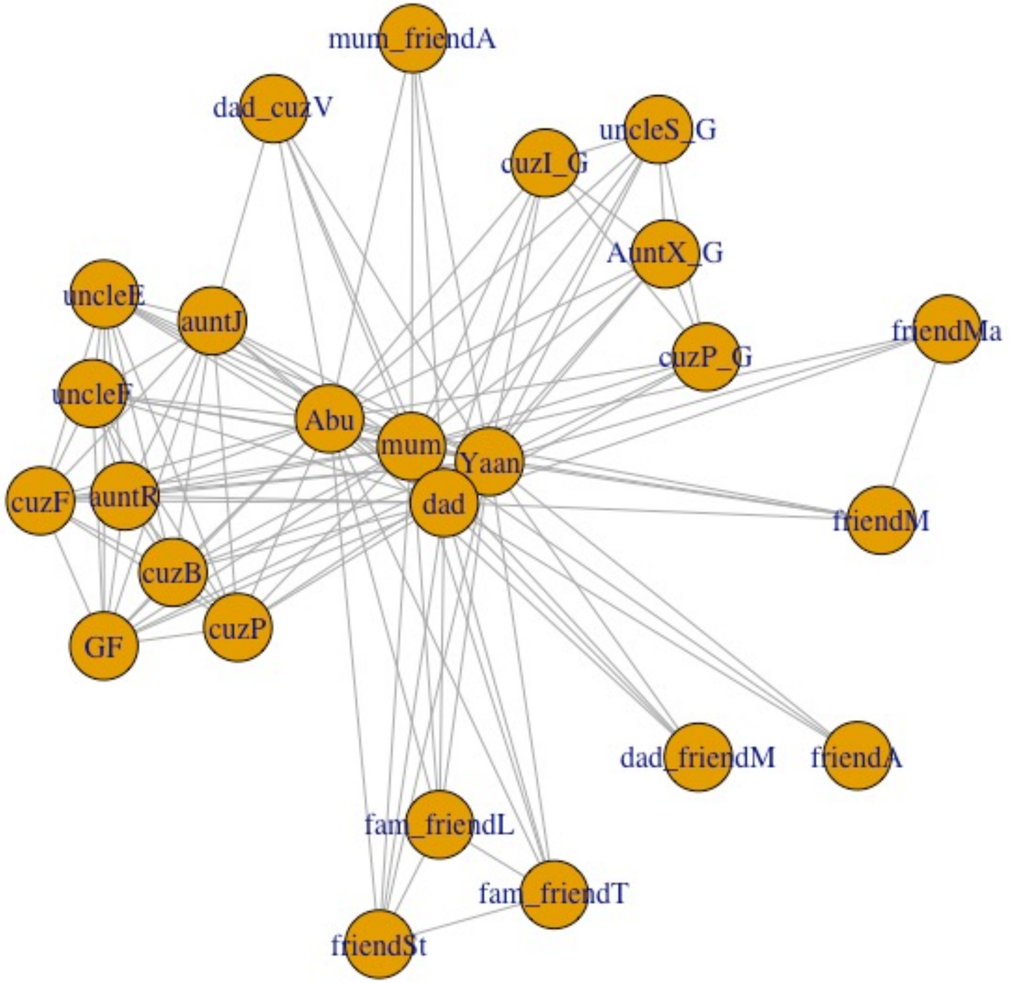
$$\boxed{< k >= \frac{m}{n}}$$

NetworkX and iGraph convention!!!!!

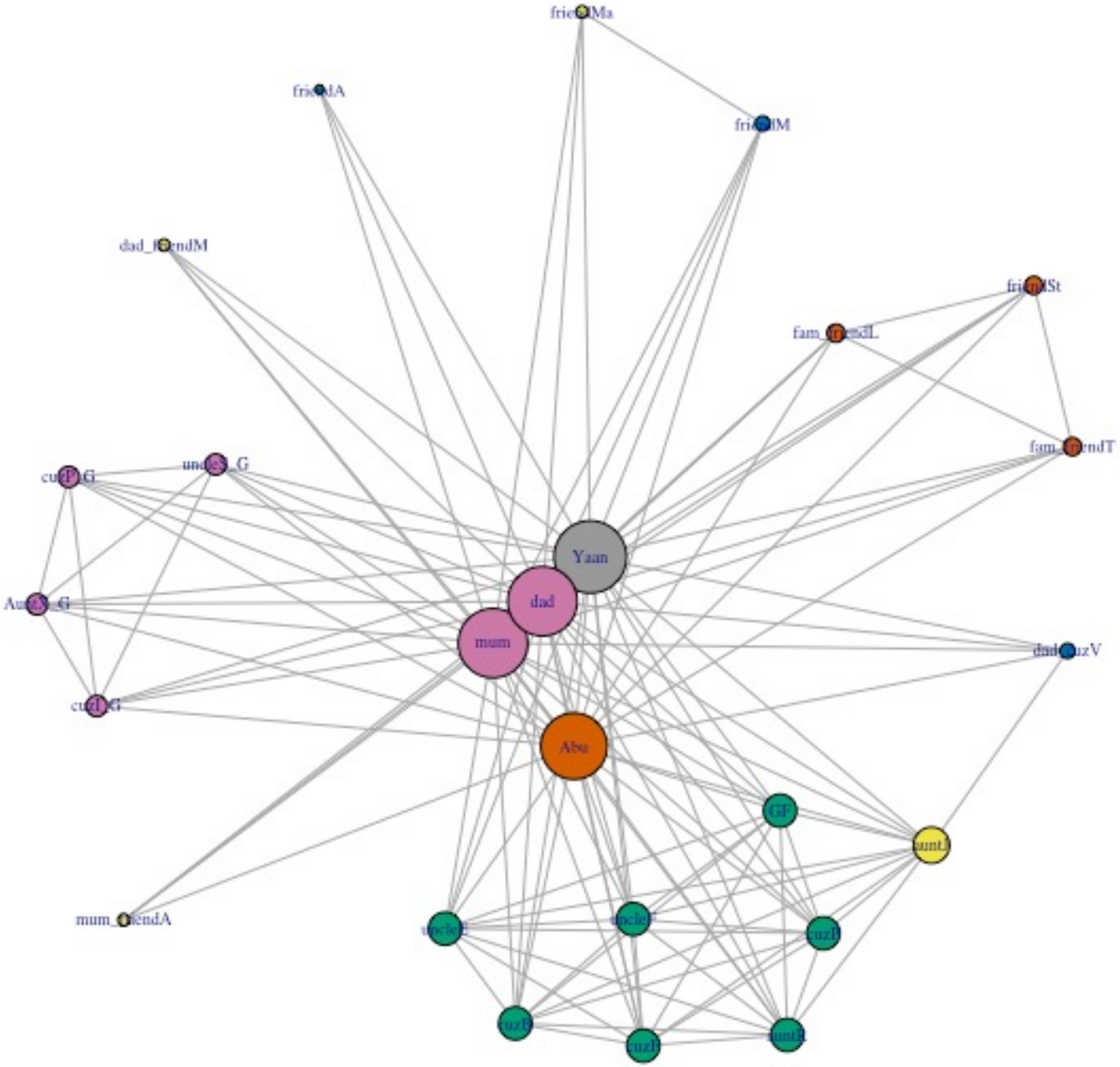
The notation given here is consistent with **Newman**'s book. **NEVERTHELESS**, each software and individual define their own notation so care must be taken.

In the exercises note that the adjacency matrix defined here corresponds to the transposed: A^T in **NetworkX** and in **iGraph**. For undirected graphs: $A^T=A$

Original network



Node size given by degree



OK, so far I know my degree, I know who are the important people around, but can I talk to them?

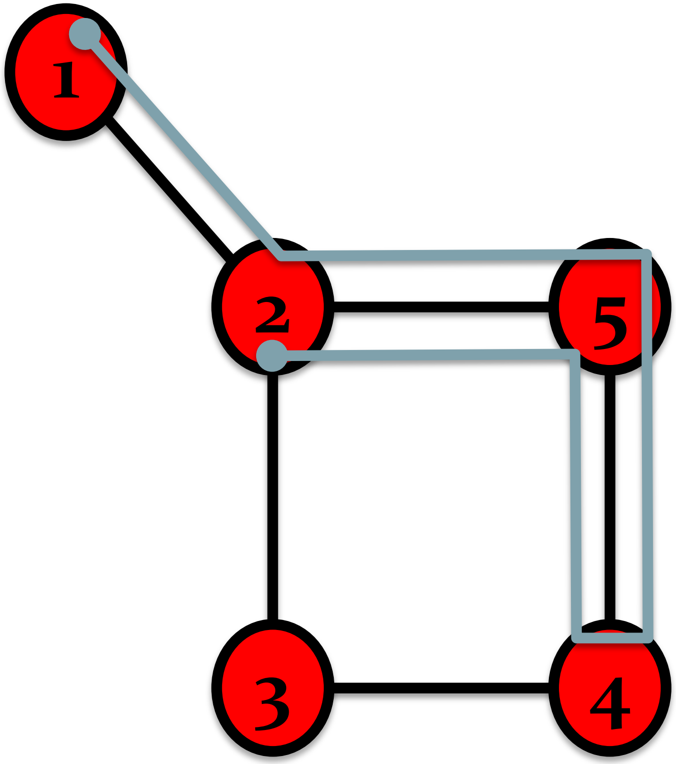
Or can I reach that piece of information?

Or how likely I am to get infected from that disease?

How well connected is my network?

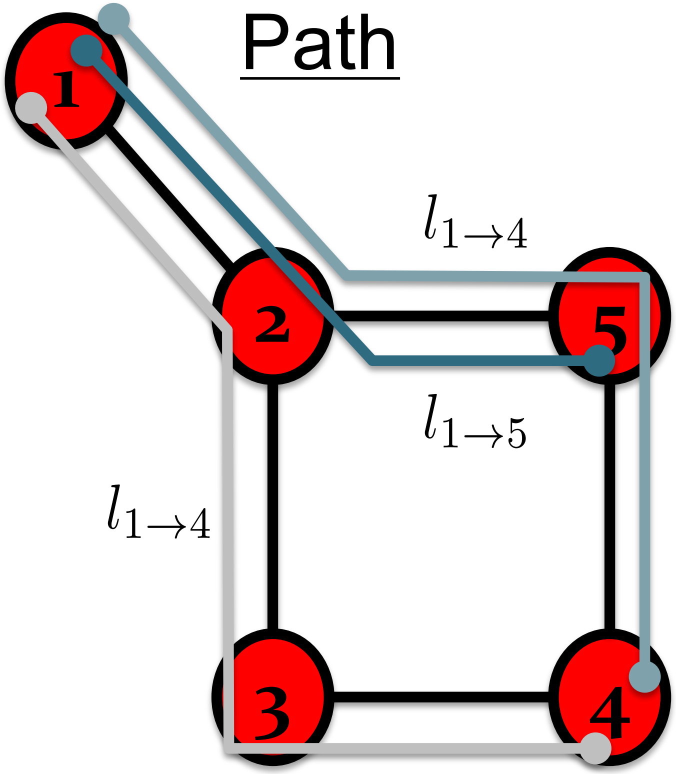
Slide from Barábasi lectures
https://www.barabasilab.com/course

Path



A sequence of nodes such that each node is connected to the next node along the path by a link.

Shortest Path

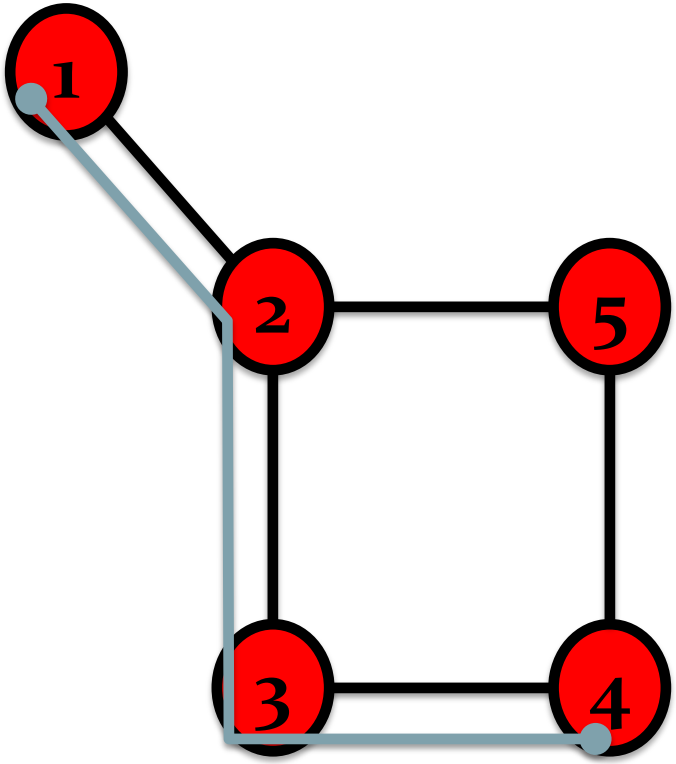


$$l_{1\to4} = 3$$

$$l_{1\to5} = 2$$

The path with the shortest length between two nodes (distance).

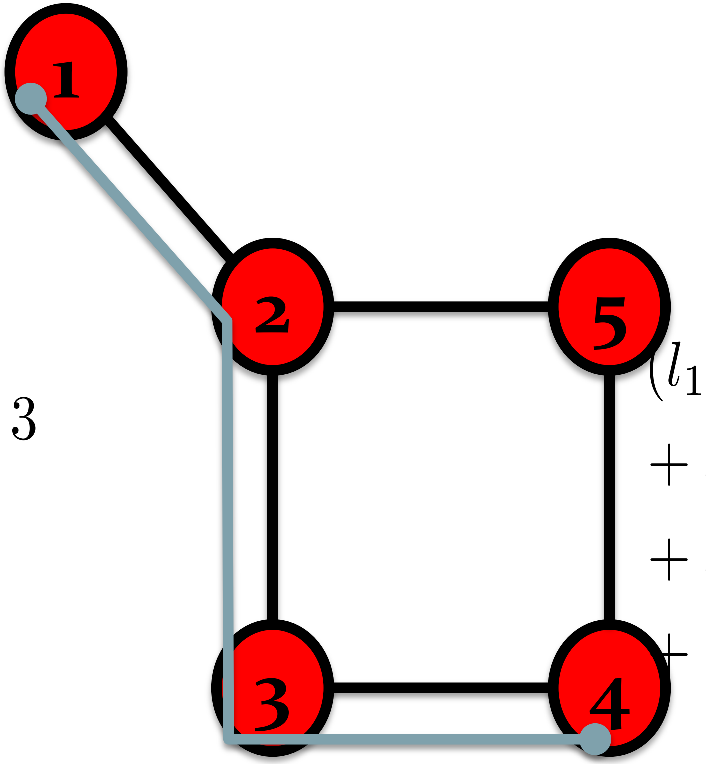
Diameter



$$l_{1\to4} = 3$$

The longest shortest path in a graph

Average Path Length

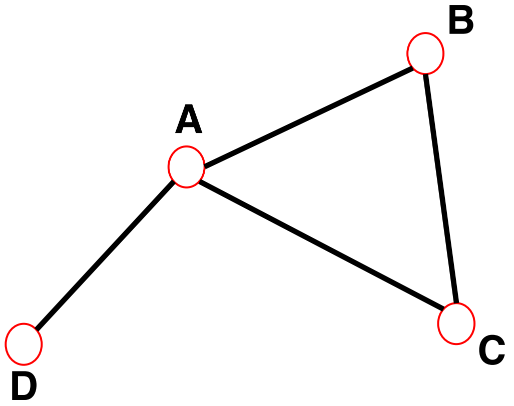


$$\begin{gathered}(l_{1\to2} + l_{1\to3} + l_{1\to4} + \\ + l_{1\to5} + l_{2\to3} + l_{2\to4} + \\ + l_{2\to5} + l_{3\to4} + l_{3\to5} + \\ + l_{4\to5})/10 = 1.6\end{gathered}$$

The average of the shortest paths for all pairs of nodes.

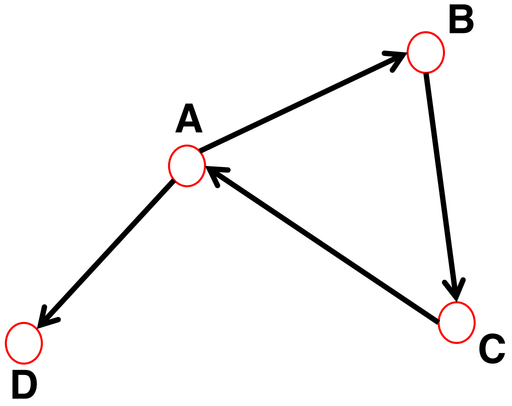
Slide from Barábasi lectures
https://www.barabasilab.com/course

Shortest Path, Geodesic Path



The *distance (shortest path, geodesic path)* between two nodes is defined as the number of edges along the shortest path connecting them.

*If the two nodes are disconnected, the distance is infinity.



In directed graphs each path needs to follow the direction of the arrows.

Thus in a digraph the distance from node A to B (on an AB path) is generally different from the distance from node B to A (on a BCA path).

Connectance or density ρ: undirected simple graph

ρ: refers to fraction of the maximum number of possible edges in a simple graph that are present.

Max. possible n. of edges: $m_{\max} = \binom{n}{2} = \frac{1}{2} n(n-1)$

$$\boxed{\rho = \frac{m}{m_{\max}} = \frac{2m}{n(n-1)} = \frac{\langle k \rangle}{n-1}}$$

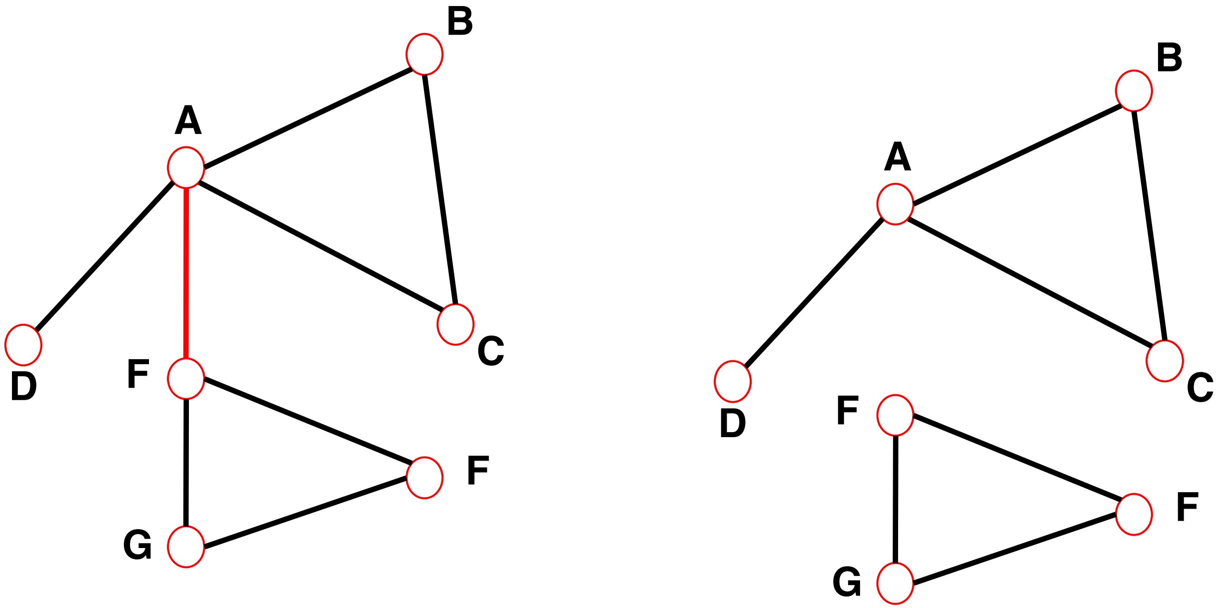
Note on the binomial coefficient:
→ It gives you the n. of ways you can choose k elements out of a set of n.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Slide from Barábasi lectures
http://barabasilab.neu.edu/courses/phys5116/

CONNECTIVITY OF UNDIRECTED GRAPHS

Connected (undirected) graph: any two vertices can be joined by a path.
A disconnected graph is made up by two or more connected components.



Largest Component: **Giant Component**

The rest: **Isolates**

Bridge: if we erase it, the graph becomes disconnected.

CONNECTIVITY OF UNDIRECTED GRAPHS

Adjacency Matrix

The adjacency matrix of a network with several components can be written in a block-diagonal form, so that nonzero elements are confined to squares, with all other elements being zero:

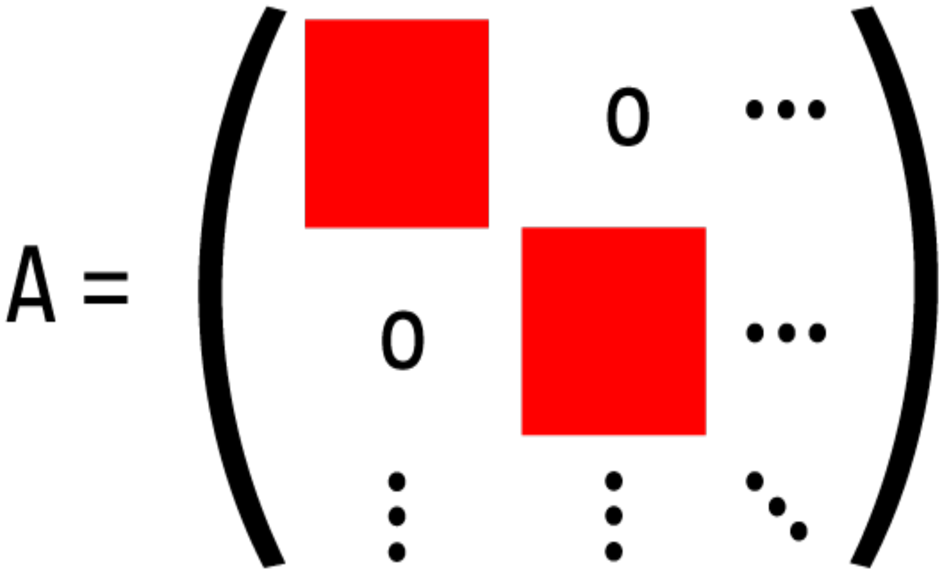


Figure after Newman, 2010

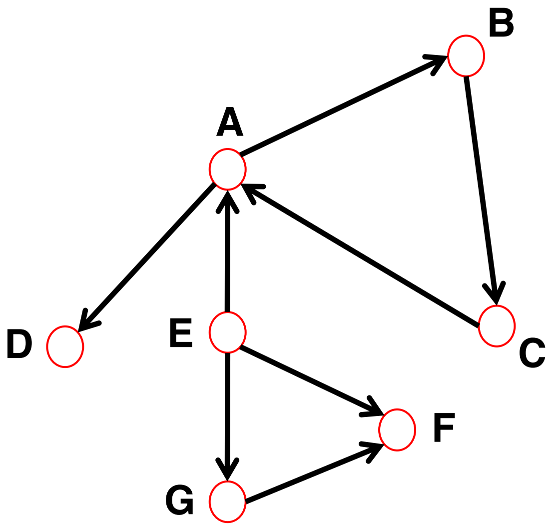
Slide from Barábasi lectures
https://www.barabasilab.com/course

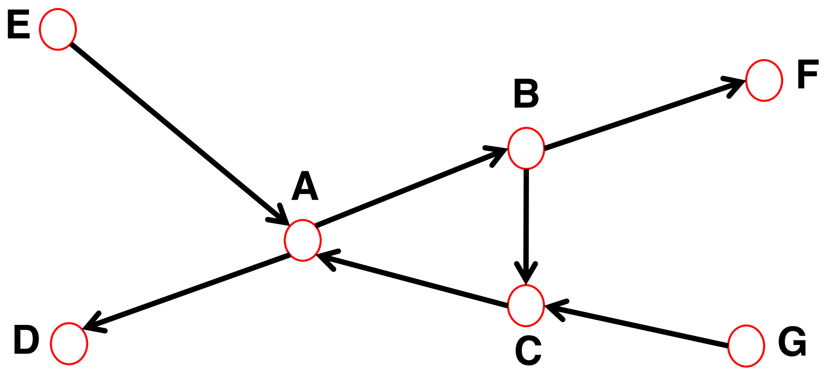
CONNECTIVITY OF DIRECTED GRAPHS

Strongly connected directed graph: has a path from each node to every other node and vice versa (e.g. AB path and BA path).

Weakly connected directed graph: it is connected if we disregard the edge directions.

Strongly connected components can be identified, but not every node is part of a nontrivial strongly connected component.





In-component: nodes that can reach the scc.

Out-component: nodes that can be reached from the scc.

How do we now find the main influencer or broker of the system?

- Which are the main characteristics a so called “influencer” has?
- And those of a “broker”?

Centrality Measures

→ Key measures for social network analysis of topological properties of networks.
→ Measures can be interpreted in many non-social contexts as well, such as biology.

→ E.g. identify key individuals: hubs, brokers, etc.
→ Adjacency matrix essential!!! A_{ij}

- **Degree** ⟺ **degree centrality**: recall for directed networks need to compute in- and out-degree. They do not convey the same information!!!!

  → Individual with high out-degree and low in-degree is sending information to many people, but not receiving much back: particular role in community or spammer!

Eigenvector centrality: x

→ It's not about how many connections you have, but *how important* those connections are!

- Initially we do not know how important anybody is according to this measure, so set at t=0: $x_i(0)=1$; where $x_i(t)$ gives the measure for node *i* at time *t*
- Can encode the measures for all *n* nodes in a vector of dimension *n:* $\mathbf{x}$, with elements x_i
- To find the effect of being connected to relevant nodes, we take all the neighbours of node *i* given through the adjacency matrix:

$$x'_i = \sum_j A_{ij} x_j \quad \Leftrightarrow \quad \mathbf{x}' = A\mathbf{x}$$

- Repeat process *t* steps to get better estimate:

$$\mathbf{x}(t) = A^t \mathbf{x}(0)$$

Linear algebra

- Any *n*-dimensional vector can be expressed in a basis defined by a linear combination of *n* linearly independent vectors.

- An *n x n* matrix can have a max of *n* eigenvectors, if it has *n* different eigenvalues. These can be used as a basis. They can be found through the following equation:

$$A\mathbf{v}=\kappa\mathbf{v}$$

$$\Updownarrow$$

$$(A-\kappa I)\ \mathbf{v}=0\ \ ;\ I:\ n\ x\ n\ \text{identity matrix}$$

$$\Updownarrow$$

$$\det(A-\kappa I)=0\ \ ;\ \text{if}\ \mathbf{v}\neq 0$$

Eigenvector centrality: x

- Let $\mathbf{v}_i$ be the eigenvectors, and κ_i the eigenvalues of the adjacency matrix A, thus:

$$\vec{x}(0) = \sum_i c_i \vec{\mathrm{v}}_i$$

$$\vec{x}(t) = \mathrm{A}^t \sum_i c_i \vec{\mathrm{v}}_i = \sum_i c_i \kappa_i^t \vec{\mathrm{v}}_i = \kappa_1^t \sum_i c_i \left[\frac{\kappa_i}{\kappa_1}\right]^t \vec{\mathrm{v}}_i$$

Where κ_1 is the largest eigenvalue: $\kappa_i/\kappa_1 < 1$, for all $i \neq 1$.
In the limit of $t \rightarrow \infty$:

$$\mathrm{A}\mathbf{x} = \kappa_1 \mathbf{x}$$

$$\boxed{x_i = \kappa_1^{-1} \sum_j A_{ij} x_j}$$

Eigenvector centrality: x

- ✧ Encounter issues with directed networks

  → which direction should we use?

⇩

use nodes pointing at you!

Note adjacency matrix non-symmetric for directed graphs:

$$\boxed{x_i = \kappa_1^{-1} \sum_j A_{ij} x_j}$$

Recall A_{ij}=1 if j points to i

Katz centrality: x

However!!!! What if the only neighbour pointing at you only has out degree links???

→ it will have null centrality

→ If this is your only neighbour: your centrality will also be null!

→ your contribution to other centralities will also be null, etc.

→ Oh NO!!! The whole network could have null centrality!!!!

Katz centrality (often called α-centrality): x

→ solution: add a bit of centrality "for free" to each node **β**

$$x_i = \alpha \sum_j A_{ij} x_j + \beta$$

$$\mathbf{x} = \alpha\, \mathbf{A}\mathbf{x} + \beta\, \mathbf{I}$$

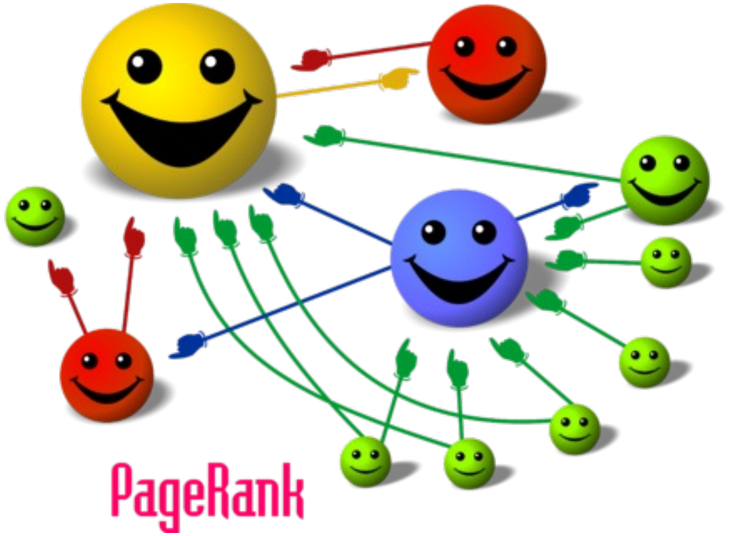
$$\Downarrow$$

$$\boxed{\mathbf{x} = (\mathbf{I} - \alpha\, \mathbf{A})^{-1}\, \mathbf{1}}$$

→ Setting β=1

→ Note that $\alpha \leq 1/k_1$. If this is very close to = then it gives you the eigenvector centrality with non-zero terms for very low centrality.

Page Rank: Google used this algorithm (derived by Brin and Page) for the ranking of websites in a search



Source: Wikipedia

→ What if a prestigious node points towards many other nodes?

→ The importance of this node pointing at you gets diluted, but it wouldn’t seem so according to the previous centrality measure.

→ Divide the contribution of neighbour to your centrality by their out-degree

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{out}} + \beta$$

Set all out-degrees that are =0 to be =1, since A_{ij}=0 in those cases anyway, and so there’s now contribution from those nodes.

Closeness centrality

Let d_{ij} be the geodesic between *i* and *j*. The mean geodesic distance is:

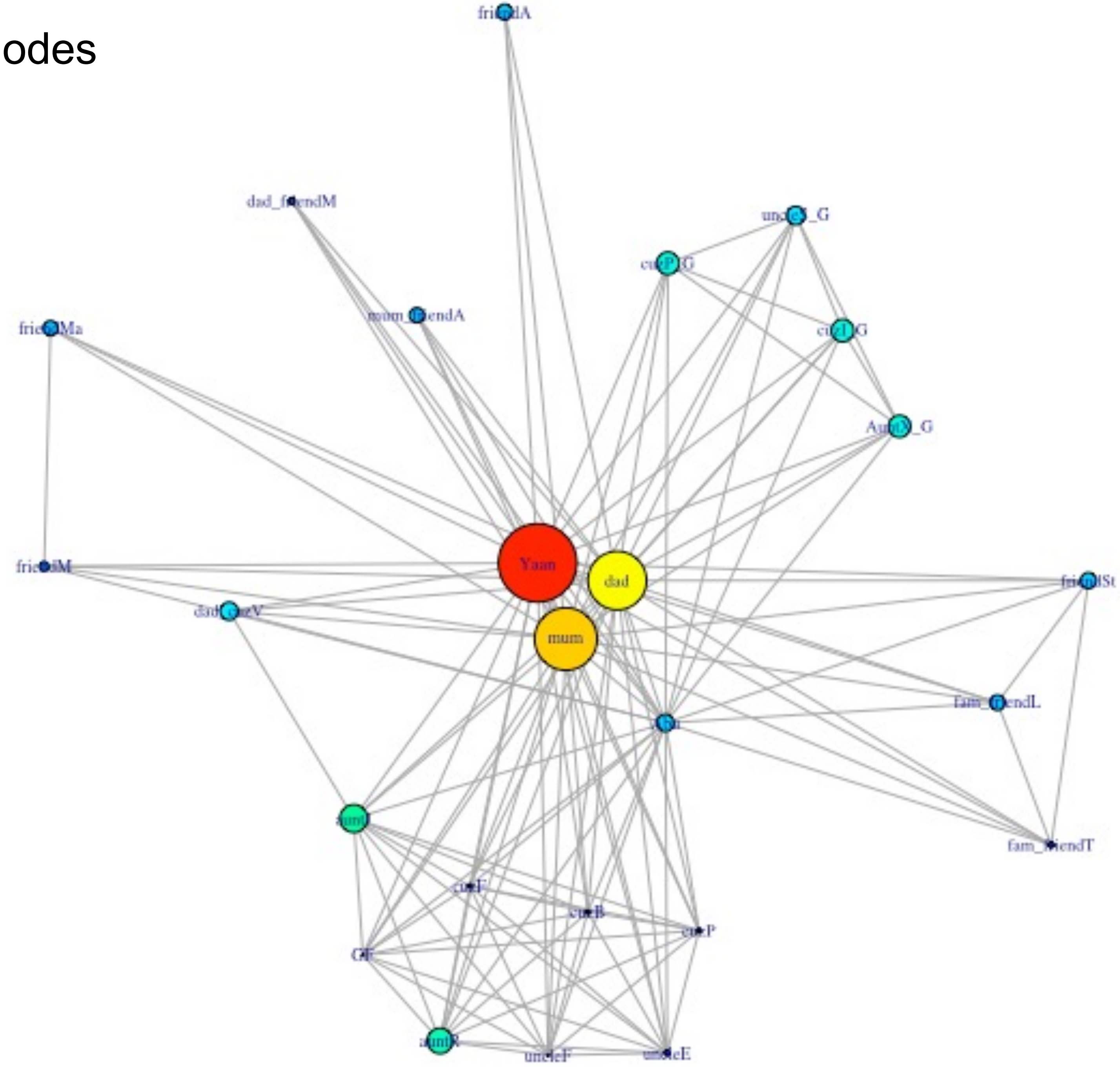
$$l_i = \frac{1}{n}\sum_j d_{ij}$$

n, is the total number of nodes.

A person that is very close to most nodes, and has hence low mean geodesic, will be influential: we define closeness centrality as

$$\boxed{C_i = \frac{1}{l_i} = \frac{n}{\sum_j d_{ij}}}$$

Colour and size of nodes given by value of closeness centrality



Harmonic closeness centrality

Consider that if network is disconnected, d_{ij} → infinity.

$$\boxed{C_i' = \frac{1}{n-1} \sum_{j(\neq i)} \frac{1}{d_{ij}}}$$

We need to exclude $i=j$ since $d_{ii}=0$. If $d_{ij} = \infty$ we are OK now.

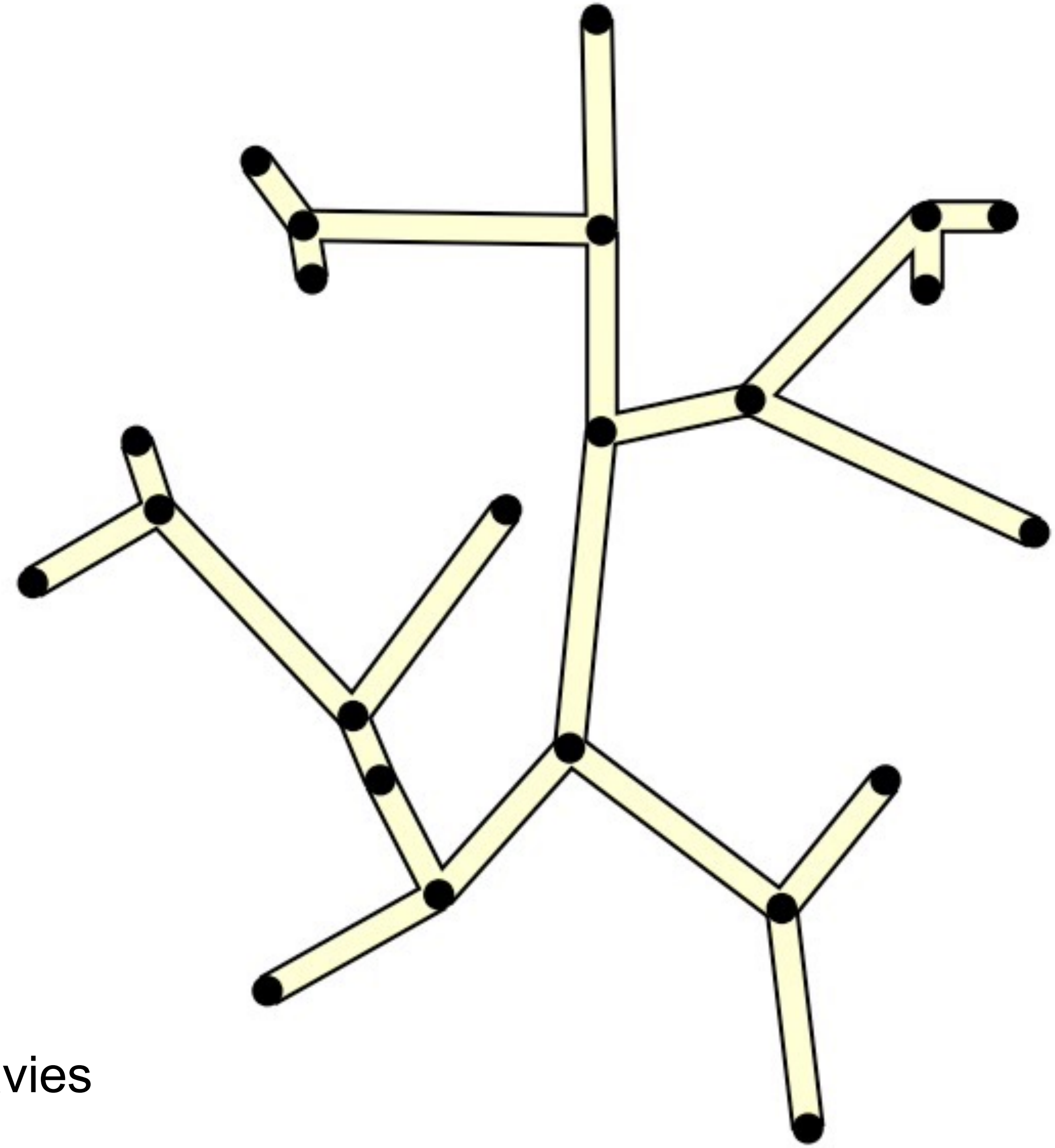
Betweenness centrality

Vertex betweenness

This is a measure of the number of shortest paths between all pairs of nodes passing though that vertex.

Edge Betweenness

This is a measure of the number of shortest paths between all pairs of nodes passing though that edge.



Courtesy of Toby Davies

Constructing betweenness centrality

Let us denote by n^i_{st}

$$n^i_{st} = \begin{cases} 1 & \text{if vertex } i \text{ lies on geodesic path from } s \text{ to } t \\ 0 & \text{otherwise} \end{cases}$$

The betweenness centrality can be roughly defined as initially as

$$x_i = \sum_{st} n^i_{st}$$

If more than one geodesic passes through that vertex, need to weight its contribution

$$x_i = \sum_{st} \frac{n^i_{st}}{g_{st}}$$

where g_{st} is the total number of geodesics from s to t

Betweenness centrality

Let us normalise the measure so that it lies between 0 and 1

$$\boxed{x_i = \frac{1}{n^2}\sum_{st}\frac{n^i_{st}}{g_{st}}}$$

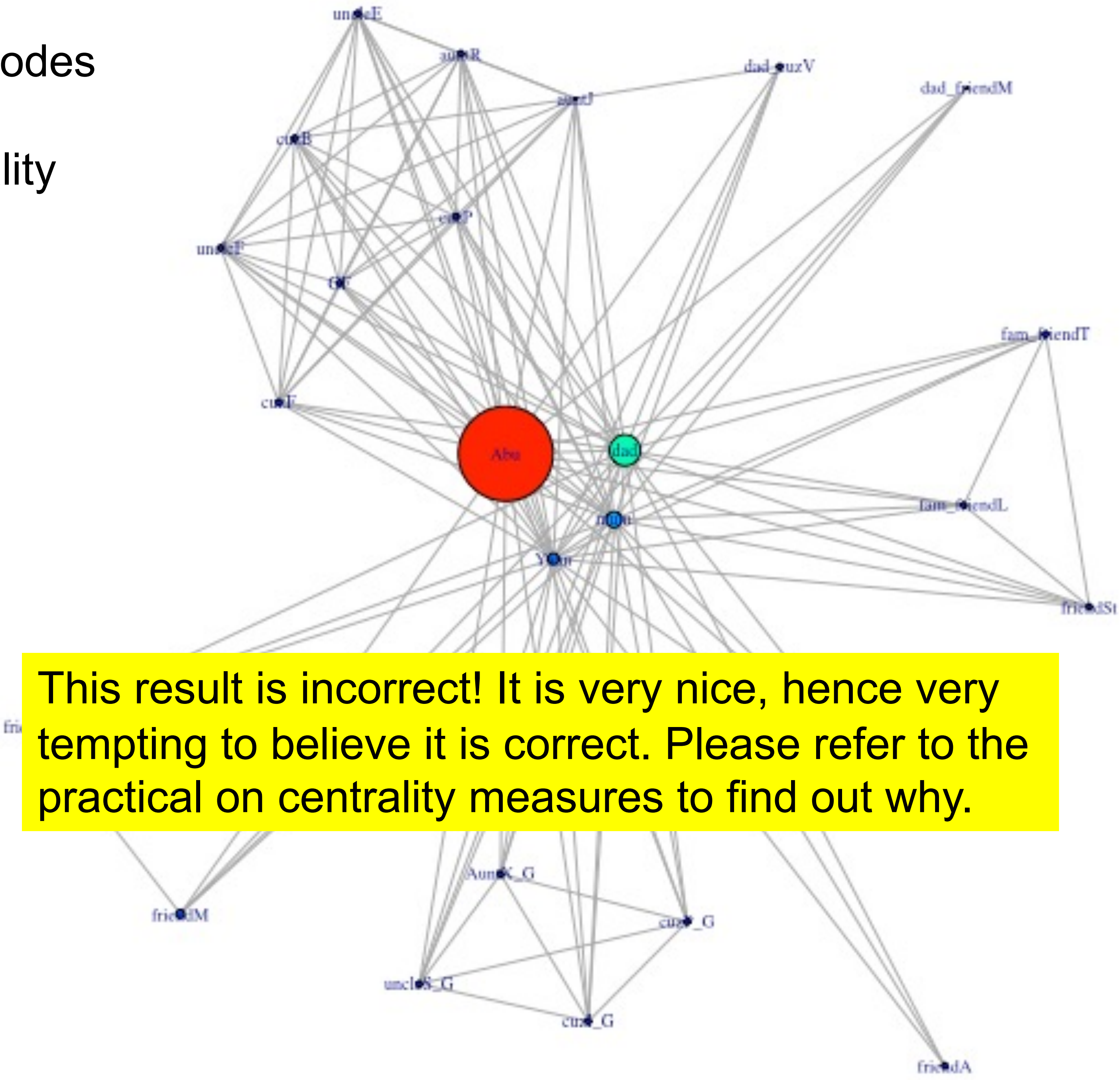
→ Flow of information or any other sort of traffic assuming that this takes the shortest path!!!!

→ In real systems we might have to modify this betweenness according to the more realistic behaviour of the system, where the other extreme would be a random walk.

→ High betweenness indicates the role of a *broker* in the system

Colour and size of nodes given by value of betweenness centrality

Grandma is the broker of the system!!!!



This result is incorrect! It is very nice, hence very tempting to believe it is correct. Please refer to the practical on centrality measures to find out why.

Delta-centrality (Vito Latora and Massimo Marchiori)

Latora, V. and Marchiori, M., 2007. A measure of centrality based on network efficiency, *New Journal of Physics 9, 188.*

Centrality of node *i* with respect to its contribution to the cohesiveness of the network *G*. How to do this?

→ Observe changes to the network once the node is removed.

$$C_i^{\Delta} = \frac{(\Delta P)_i}{P} = \frac{P[G] - P[G']}{P[G]}$$

Where *P*[*G*] is the **performance** of a graph *G*, *G*' is the new graph after removing *i* and *P*[*G*'] is the performance of the new graph *G*';
$(\Delta P)_i$ is the variation of the performance after deactivation of node *i*

How to choose P? It must satisfy: $(\Delta P)_i \geq 0$

Delta-centrality (Vito Latora and Massimo Marchiori)

$$C_i^{\Delta} = \frac{(\Delta P)_i}{P} = \frac{P[G] - P[G']}{P[G]}$$

E.g. If $P[G]=m$ the number of links, if node i is removed, then $(\Delta P)_i = k_i$ the degree of the node removed.

What about efficiency?

$$E = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1, j\neq 1}^{N} \frac{1}{d_{ij}}$$

N is the total number of nodes.

→ If the efficiency between two nodes i and j is $1/d_{ij}$, E is the average over all pairs. The drop in efficiency gives a measure of *information centrality*.

Latora, V. and Marchiori, M., 2001. Efficient Behavior of Small-World Networks, *PRL 87, 198701.*