Complex Networks

(CR 15)

Network characteristics

Class 1

Dr. Márton Karsai

ENS Lyon 2016

Practical matters

Reward: 4 ECTS

Lecturers:

- Márton Karsai
  - marton.karsai@ens-lyon.fr
  - Web: perso.ens-lyon.fr/marton.karsai
- Christophe Crespelle
  - christophe.crespelle@ens-lyon.fr

Tutorials (TD):

- Sophie Jacquin (sophie.jacquin@inria.fr)

Course web page:

- perso.ens-lyon.fr/marton.karsai/Marton_Karsai/complexnet.html

Slides:

- http://perso.ens-lyon.fr/marton.karsai/protected/complexnets/
- login: complexnet
- psw: cnet123

Lectures (24h):

- Before 27/10: Tuesday 8:00-10:00 (B1)
- After 27/10: Thursdays 10:15-12:15 (B1)
- No course on 17/11

Tutorials (6h):

- 20/10 15:45
- 25/10 15:45
- 17/11 10:15
- Not mandatory for M2IF students
- Technical conditions (computer and packages) will be mailed later

Evaluation:

- Lectures: writing exam
- Tutorials: projects during the semester

Outline

Class 1: Introduction and general network characteristics (MK)

Class 2: Properties and models (CC)

Class 3: Configuration model (CC)

Class 4: Betweenness centrality and algorithm (CC)

Class 5: Navigability of small-world networks (CC)

Class 6: Phase transitions in Erdös-Rényi random graphs (CC)

Class 7: Scale-free networks, Barabási-Albert model (MK)

Class 8: Motifs and communities (MK)

Class 9: Temporal networks (MK)

Class 10: Spatial networks (MK)

Class 11: Multiplex and interdependent networks (MK)

Class 12: Network sampling and visualisation (MK)

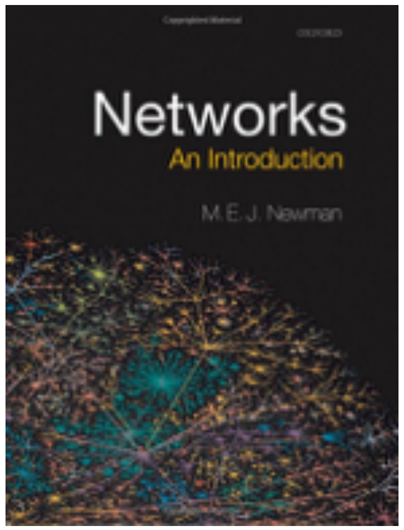
Course Targets

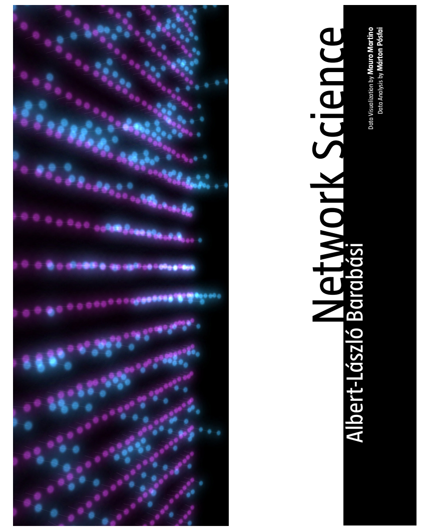
After the course, you should

- be able to interpret a complex system like a complex network
- know how to analyse and characterise networks
- understand the fundamental network models
- have insight into the evolution of networks
- detect community structure in networks
- be aware of state-of-the-art examples of real networks
- and much more…

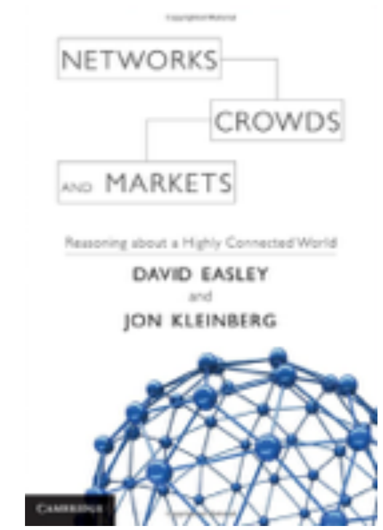
Materials

Lecture books





available free online



available free online

Reviews

SIAM REVIEW
Vol. 45, No. 2, pp. 167–256
© 2003 Society for Industrial and Applied Mathematics

The Structure and Function of Complex Networks*

M. E. J. Newman†

REVIEWS OF MODERN PHYSICS, VOLUME 74, JANUARY 2002

Statistical mechanics of complex networks

Réka Albert* and Albert-László Barabási
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

Characterization and Modeling of weighted networks

Marc Barthélemy[1], Alain Barrat[2], Romualdo Pastor-Satorras[3], and Alessandro Vespignani[2]

Physics Reports 486 (2010) 75–174
Contents lists available at ScienceDirect
ELSEVIER
Physics Reports
journal homepage: www.elsevier.com/locate/physrep

Community detection in graphs
Santo Fortunato*
Complex Networks and Systems Lagrange Laboratory, ISI Foundation, Viale S. Severo 65, 10133, Torino, I, Italy

Physics Reports 519 (2012) 97–125
Contents lists available at SciVerse ScienceDirect
ELSEVIER
Physics Reports
journal homepage: www.elsevier.com/locate/physrep

Temporal networks
Petter Holme[a,b,c,*], Jari Saramäki[d]
[a] IceLab, Department of Physics, Umeå University, 901 87 Umeå, Sweden
[b] Department of Energy Science, Sungkyunkwan University, Suwon 440–746, Republic of Korea
[c] Department of Sociology, Stockholm University, 106 91 Stockholm, Sweden
[d] Department of Biomedical Engineering and Computational Science, School of Science, Aalto University, 00076 Aalto, Espoo, Finland

Contents lists available at ScienceDirect
ELSEVIER
Physics Reports
journal homepage: www.elsevier.com/locate/physrep

Spatial networks
Marc Barthélemy*

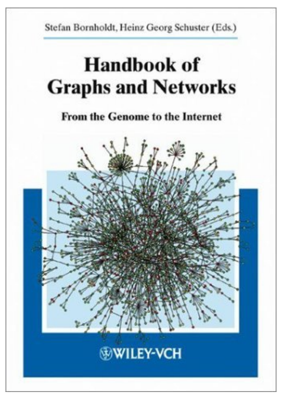
Contents lists available at ScienceDirect
ELSEVIER
Physics Reports
journal homepage: www.elsevier.com/locate/physrep

The structure and dynamics of multilayer networks
S. Boccaletti[a,b,*], G. Bianconi[c], R. Criado[d,e], C.I. del Genio[f,g,h], J. Gómez-Gardeñes[i], M. Romance[d,e], I. Sendiña-Nadal[j,e], Z. Wang[k,l], M. Zanin[m,n]

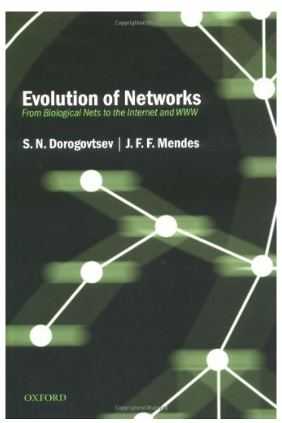
…and many more…all of them on arXiv.org!

Materials

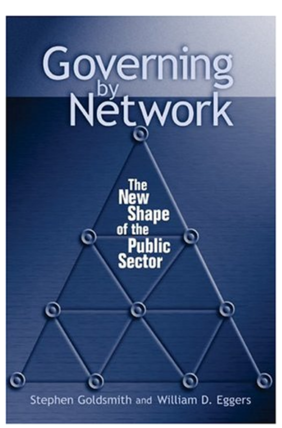
Related books



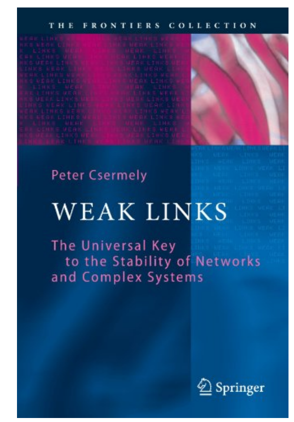
Handbook of Graphs and Networks: From the Genome to the Internet (Wiley-VCH, 2003).



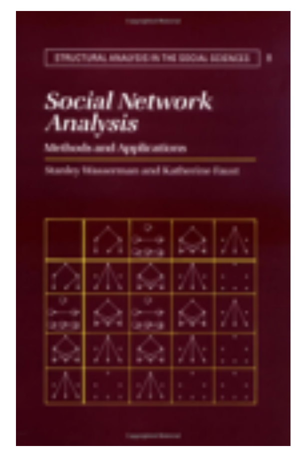
S. N. Dorogovtsev and J. F. F. Mendes, Evolution of Networks: From Biological Nets to the Internet and WWW (Oxford University Press, 2003).



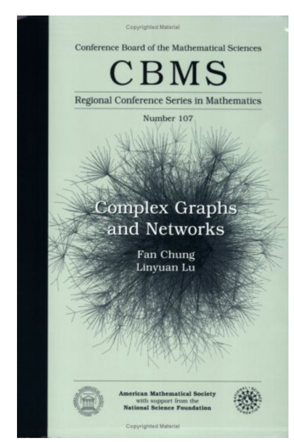
S. Goldsmith, W. D. Eggers, Governing by Network: The New Shape of the Public Sector (Brookings Institution Press, 2004).



P. Csermely, Weak Links: The Universal Key to the Stability of Networks and Complex Systems (The Frontiers Collection) (Springer, 2006), rst edn.



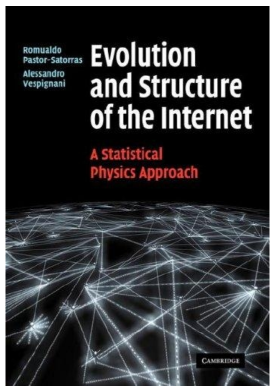
S. Wasserman and K. Faust
Social Network Analysis (Methods and Applications)
Cambridge University Press (1994)



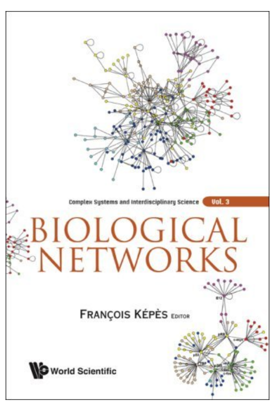
L. L. F. Chung, Complex Graphs and Networks (CBMS Regional Conference Series in Mathematics) (American Mathematical Society, 2006).

Materials

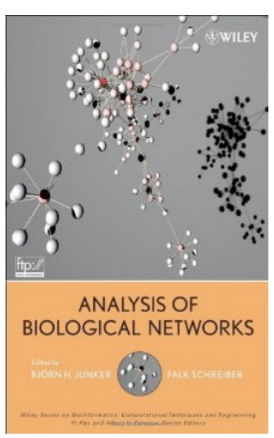
Related books



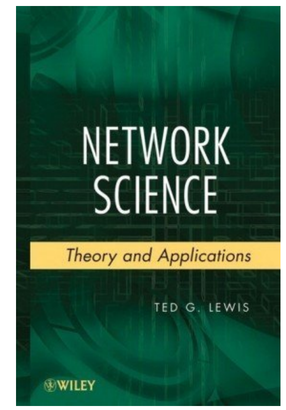
R. Pastor-Satorras, A. Vespignani, Evolution and Structure of the Internet: A Statistical Physics Approach (Cambridge University Press, 2007), rst edn.



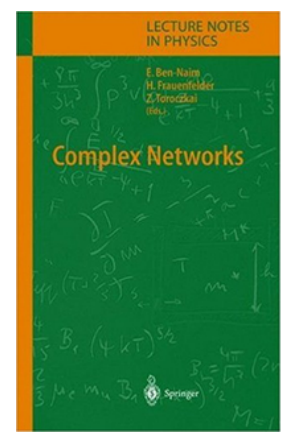
F. Kopos, Biological Networks (Complex Systems and Interdisciplinary Science) (World Scientific Publishing Company, 2007), rst edn.



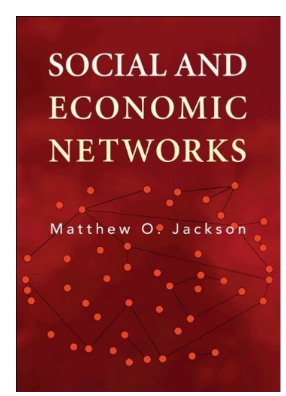
B. H. Junker, F. Schreiber, Analysis of Biological Networks (Wiley Series in Bioinformatics) (Wiley-Interscience, 2008).



T. G. Lewis, Network Science: Theory and Applications (Wiley, 2009).



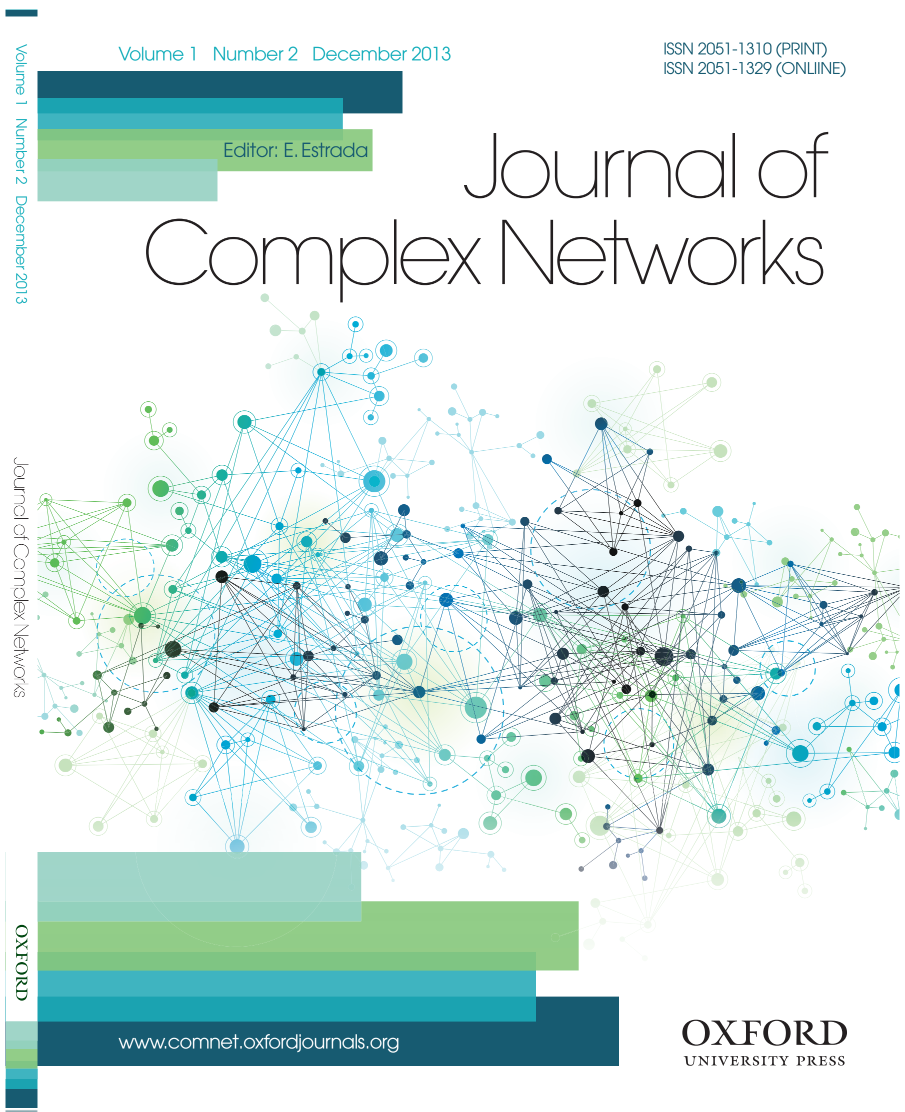
E. Ben Naim, H. Frauenfelder, Z.Torotzai, Complex Networks (Lecture Notes in Physics) (Springer, 2010), rst edn.

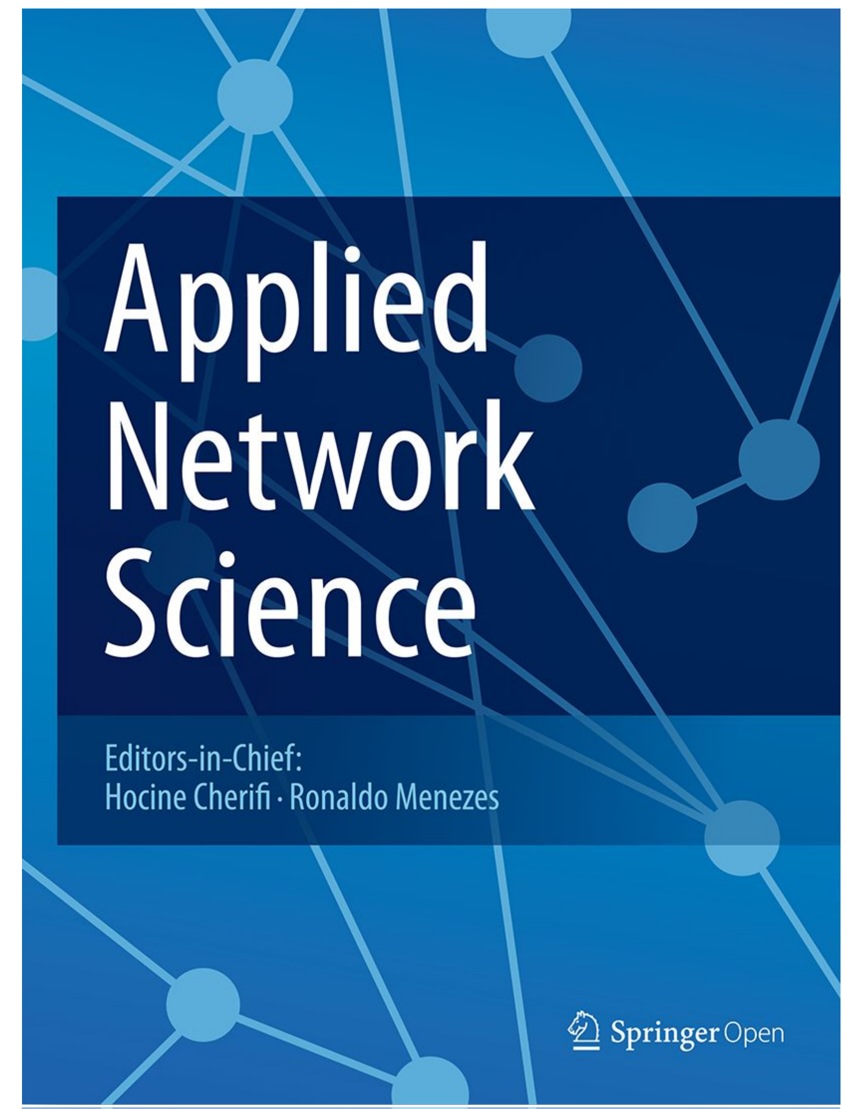


M. O. Jackson, Social and Economic Networks (Princeton University Press, 2010).

Materials

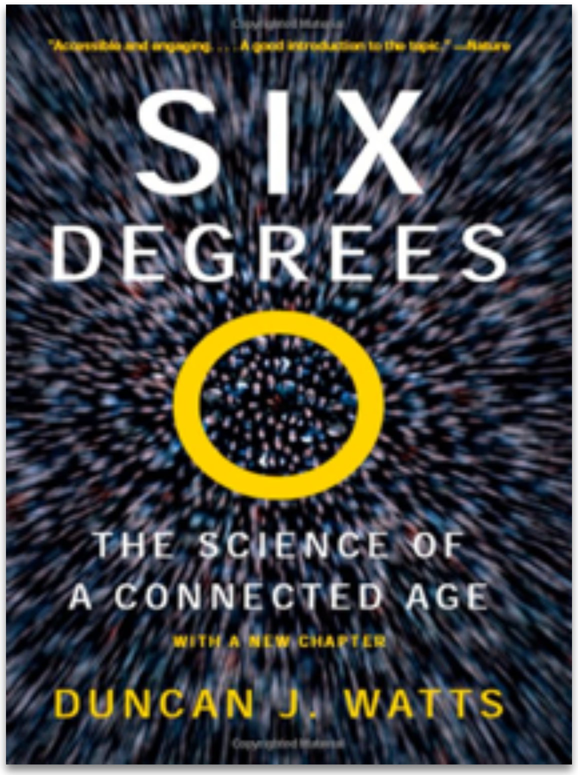
Journals

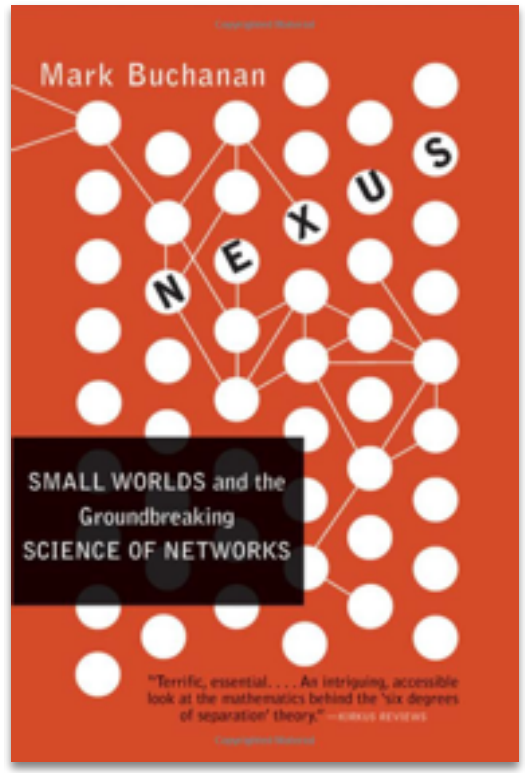


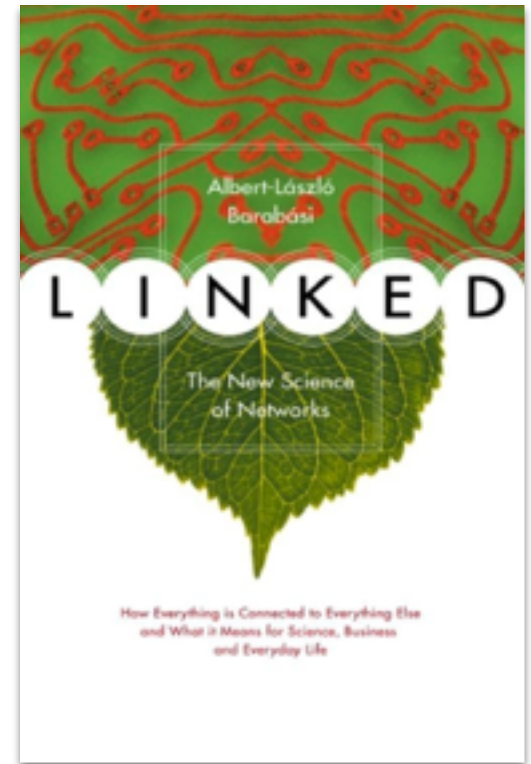


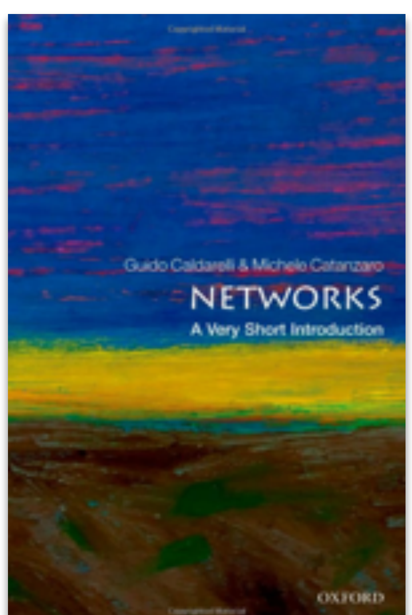
Materials

Pop-science books









Todays Schedule

1. Complexity and complex system
2. The network approach
3. Types of networks
4. Network characteristics

Complex

[adj., v. kuh m-pleks, kom-pleks; n. kom-pleks]
–adjective

1.
composed of many interconnected parts; compound; composite: a complex highway system.

2.
characterized by a very complicated or involved arrangement of parts, units, etc.: complex machinery.

3.
so complicated or intricate as to be hard to understand or deal with: a complex problem.

Source: Dictionary.com

Complexity, a **scientific theory** which asserts that some systems display behavioral phenomena that are completely inexplicable by any conventional analysis of the systems' constituent parts. These phenomena, commonly referred to as emergent behaviour, seem to occur in many complex systems involving living organisms, such as a stock market or the human brain.

Source: John L. Casti, Encyclopædia Britannica

Complexity

Complex Systems

- Self-organised
- Evolving
- Adaptive

- No central organising mind
- No conventional way of description

Complex Systems: how to approach

Statistical description

- Systems with random features
- One sample does not characterise the typical behaviour
- Statistical averages of quantities

Empirical data analysis

- How to detect patterns and structure in information?
- How to characterize the system instead of its building blocks?
- Multivariate methods etc

Analytical approach

- Write down (coupled) differential equations for interactions
- Attempt to solve
- Usually no closed-form solutions; numerical solutions, phase space analysis, etc

Simulations

- Postulate rules (e.g. the ant raids)
- Simulate and observe system behaviour
- Try to match empirical observations

OR

Complex Networks

…a way of mapping complexity

Each complex system can be interpreted as a complex network, which identifies the interactions between the interconnected components

The network approach

- Combines the elements of all the other approach
- Disregards (unnecessary) details of the system
- Focuses on the structure of interactions
- Statistical characterisation of system

The network approach

1. **Measuring** - make observations on Nature
2. **Modelling** - attempt to explain observations:
   2.1. Choose the right level of coarse-graining
   - Units: Vertices or nodes ⇔ interacting elements
   - Edges or links ⇔ interactions

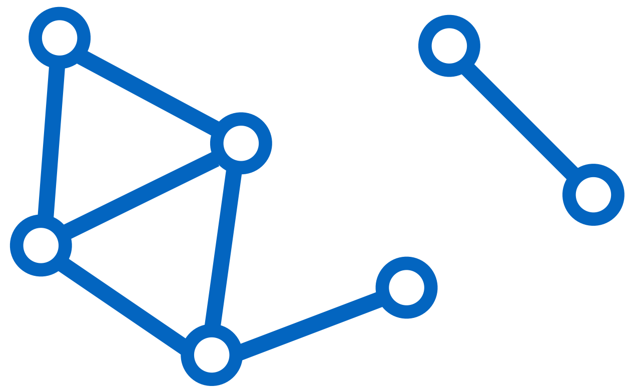
   2.2. Strip the problem to its simplest form
   - Interaction structure ⇔ evolution and behaviour of system

   2.3. Formulate the problem in mathematical terms
   - Statistical analysis of network structure
   - Dynamics of processes taking place on networks
3. **Validating** - check if calculations or simulations can
   - reproduce the observations
   - explain the observations
4. Go back to 1. & 2. and rethink

Types of Networks

Complex systems as networks

Networks are interpreted as graphs



$$G=(V, E)$$

- Components ⇔ vertices $v \in V$
- Interactions between components ⇔ edges $(u,v) \in E$
- Identification of vertices and edges defines the type of the actual network

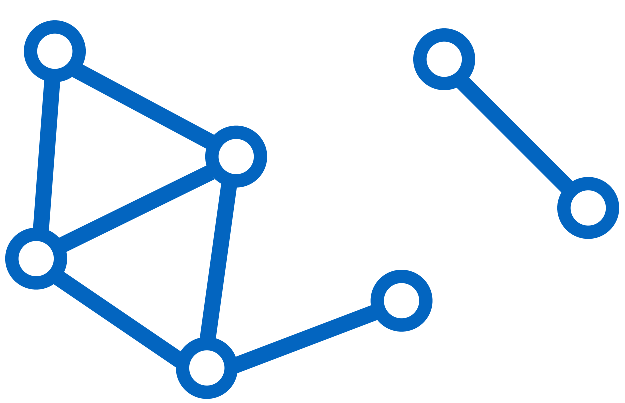
Vertex	Edge
person	friendship
neuron	synapse
WWW	hyperlink
company	ownership
gene	regulation

Undirected networks

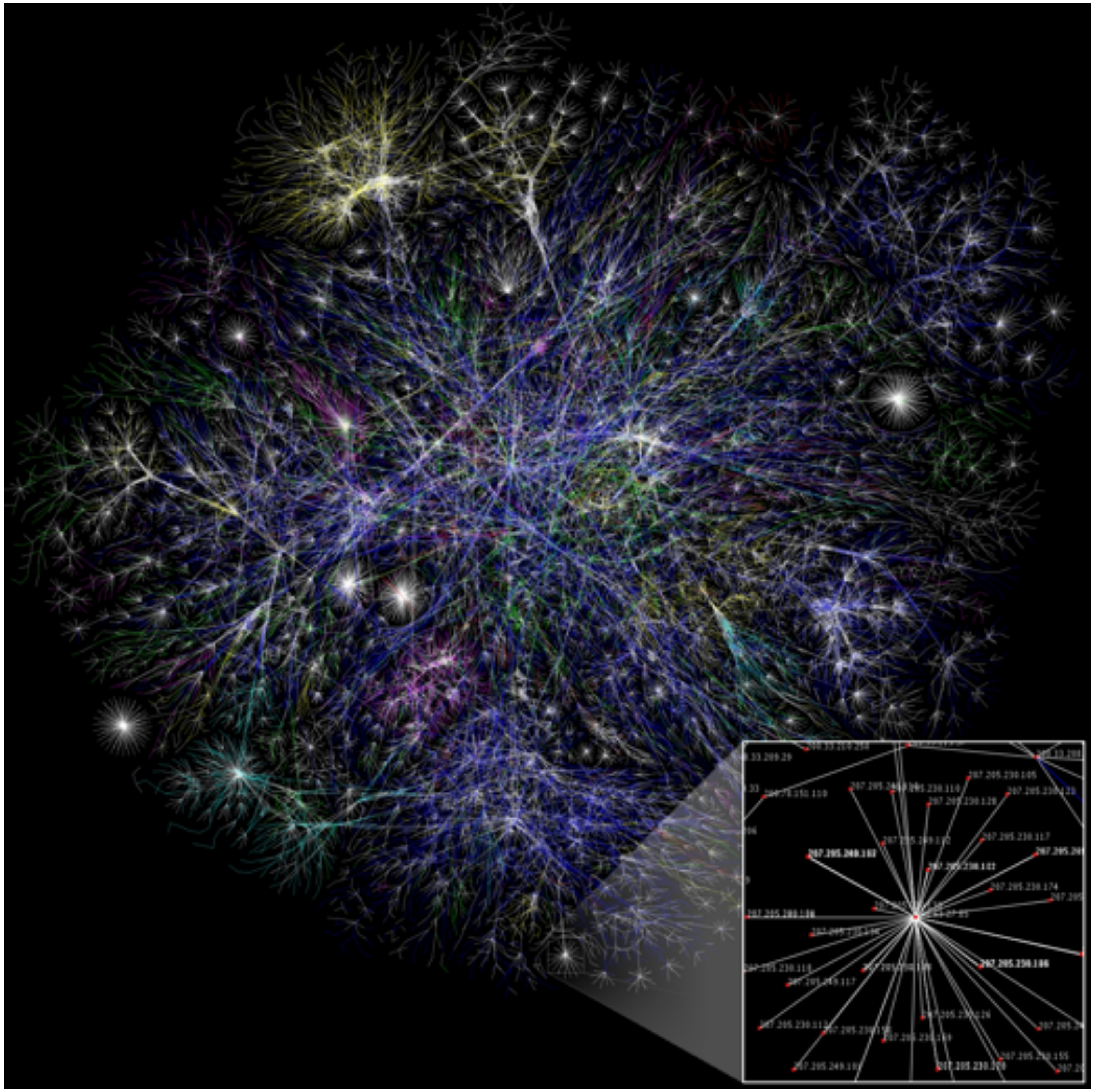
$G=(V, E)$

$(u,v) \in E \equiv (v,u) \in E$

- The directions of edges do not matter
- Interactions are possible between connected entities in both directions



Opte project



The Internet: Nodes - routers, Links - physical wires

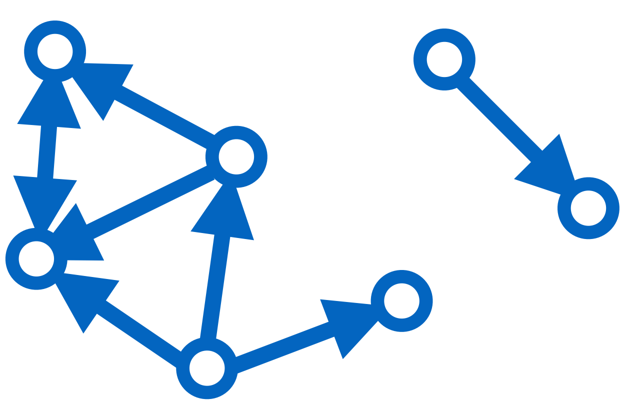
Directed networks

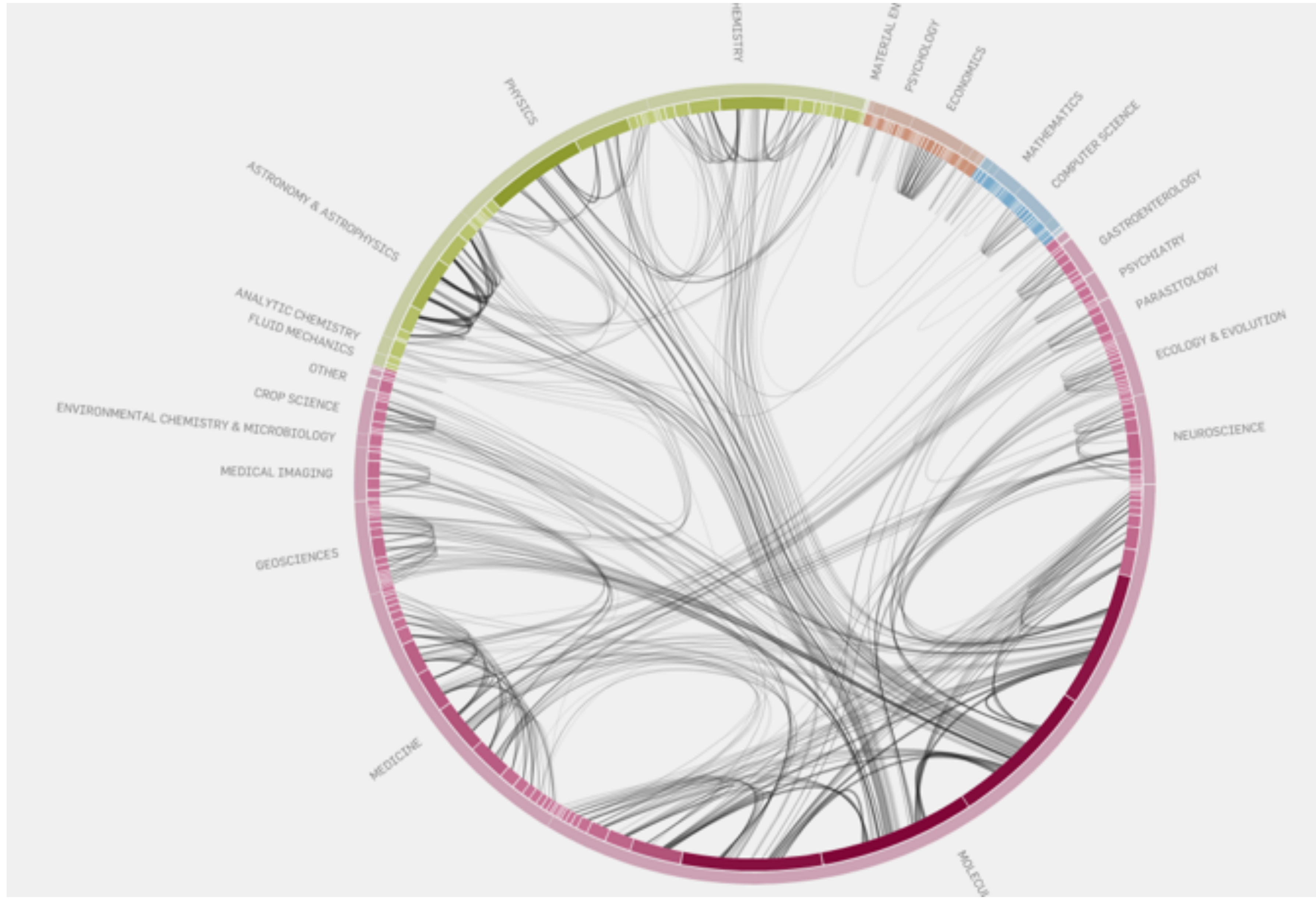
Moritz Stefaner, eigenfactor.com

$G=(V, E)$

$(u,v) \in E \not\equiv (v,u) \in E$

- The directions of edges matter
- Interactions are possible between connected entities only in specified directions

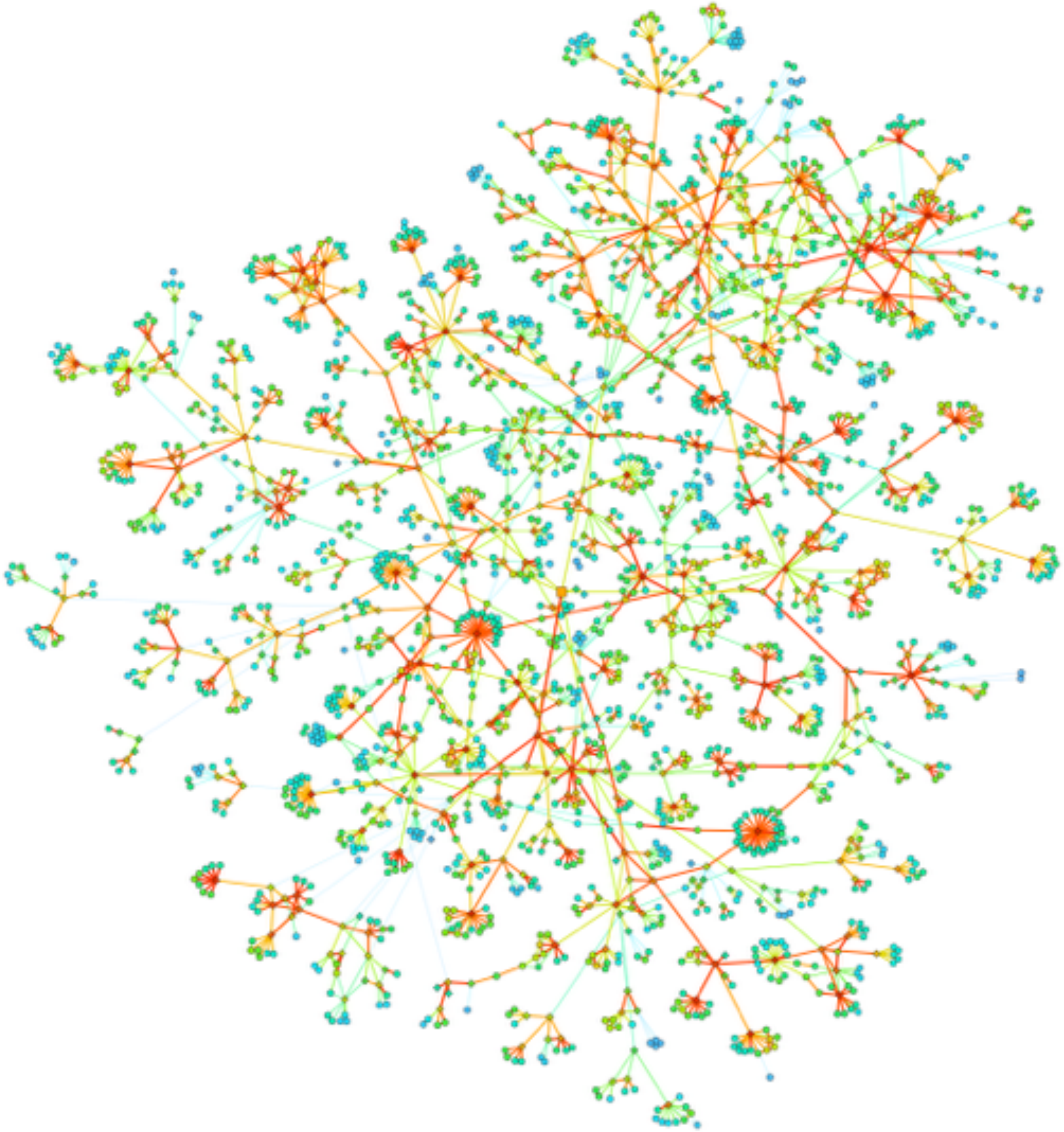




Citation network: Nodes - publications, Links - references

Weighted networks

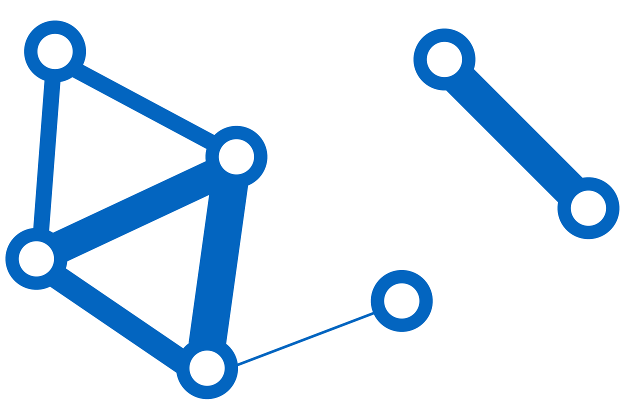
Onnela et.al. New Journal of Physics 9, 179 (2007).



$G=(V, E, w)$

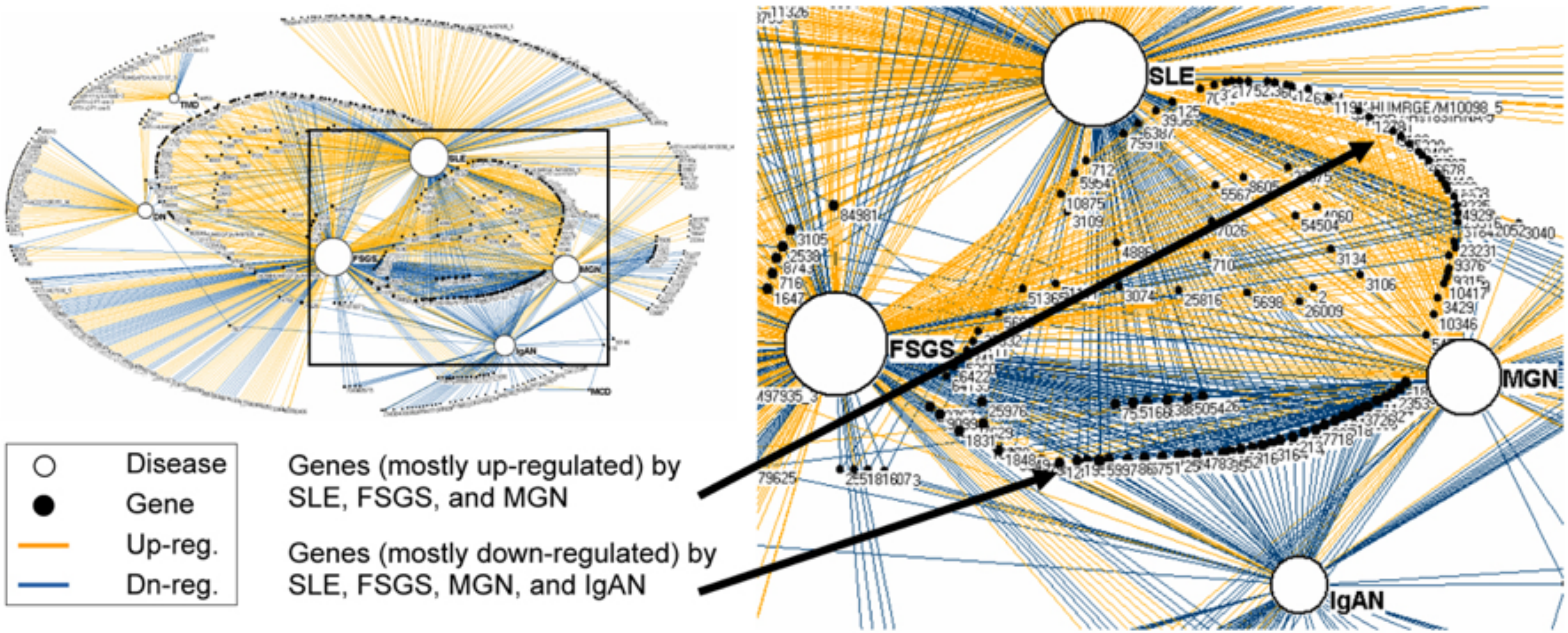
$w: (u,v) \Rightarrow R$

- Strength of interactions are assigned by the weight of links



Social interaction network: Nodes - individuals
Links - social interactions

Bipartite network

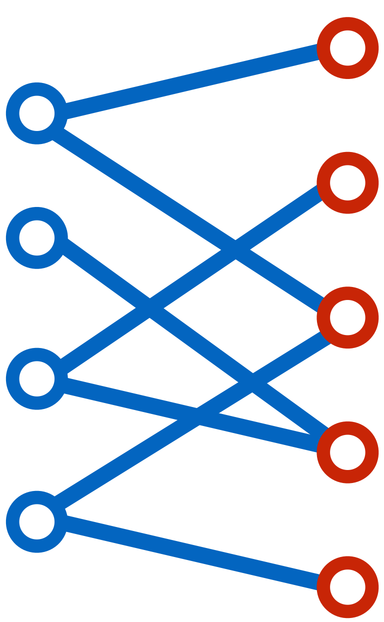


Bhavnani et.al. BMC Bioinformatics 2009, **10**(Suppl 9):S3

Gene-desease network:

Nodes - Desease (7)&Genes (747)

Links - gene-desease relationship



$$G=(U, V, E)$$

$$U \cap V = \emptyset$$

$$\forall (u,v) \in E,\ u \in U \text{ and } v \in V$$

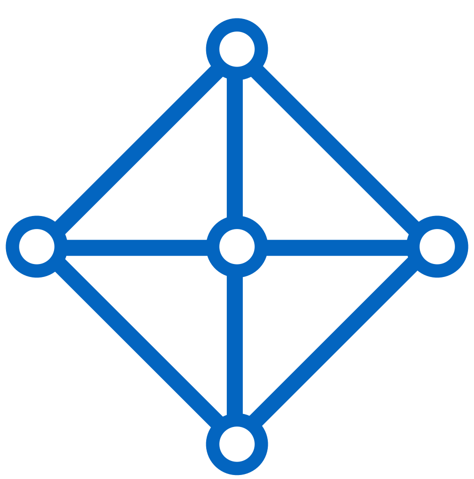
Planar networks

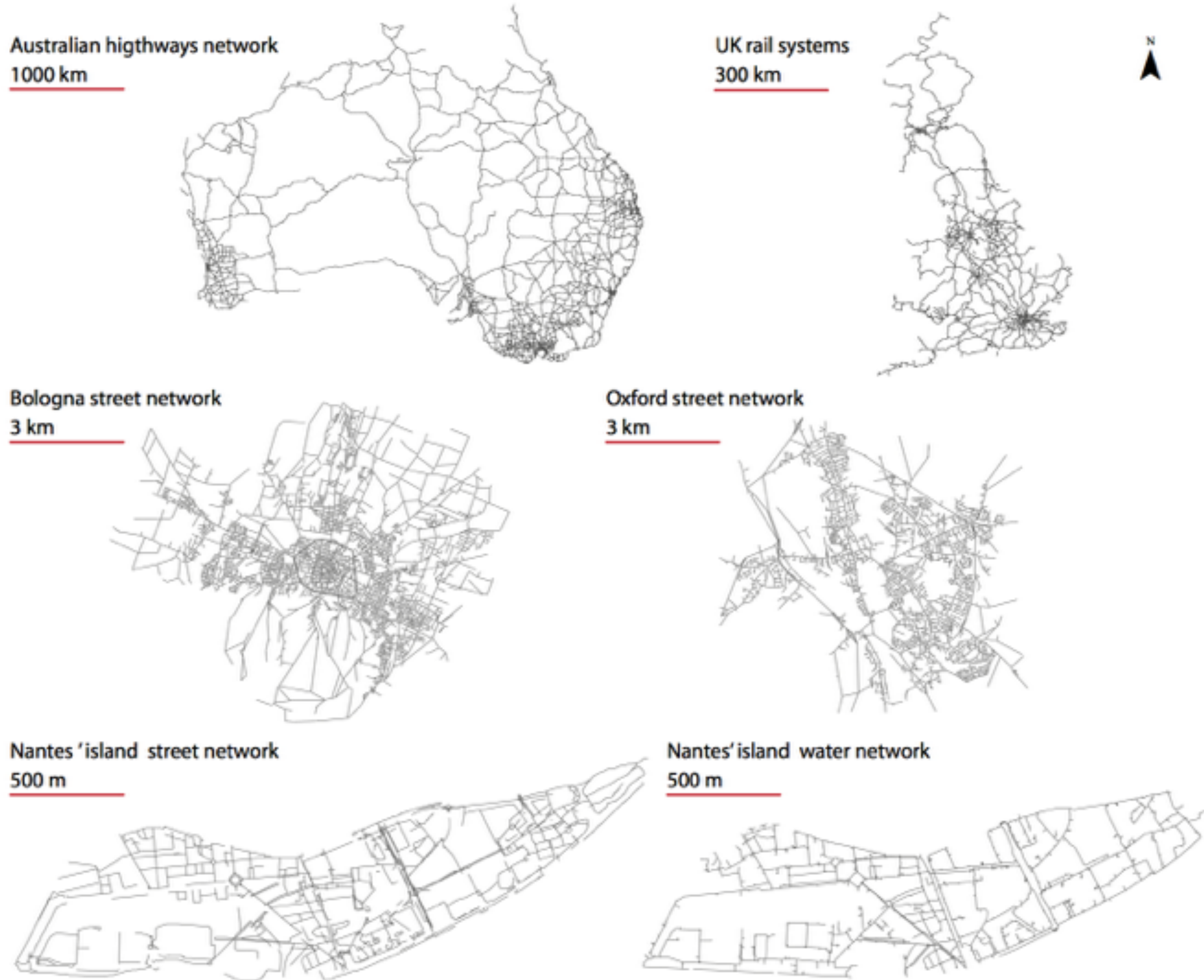
Viana et.al. Nature Scientific Reports 3:3495 (2013)

$G=(V, E, loc)$

$loc: v \Rightarrow (x,y)$

- Nodes can be embedded in a plane
- Geo-localised networks
- Spatial networks





Street networks:

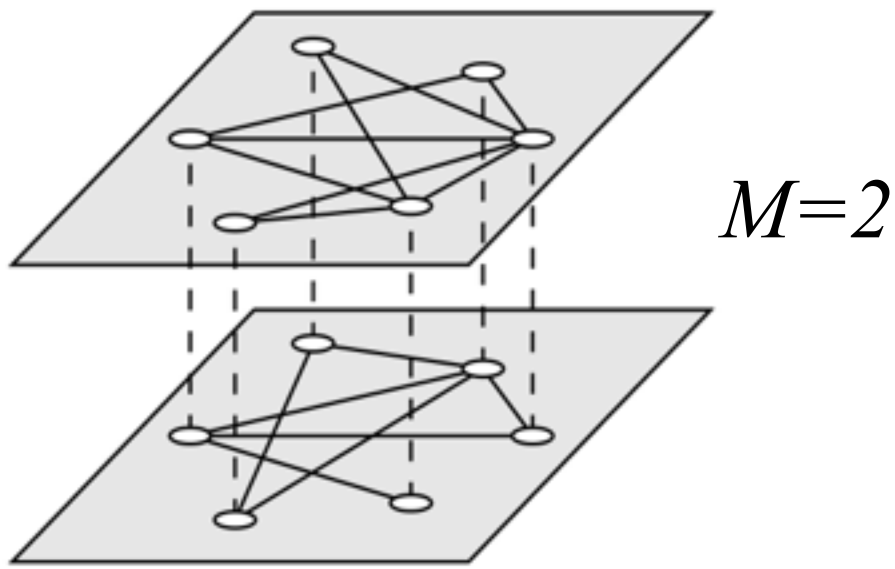
Nodes - junctions, Links - streets

Multiplex and multilayer networks

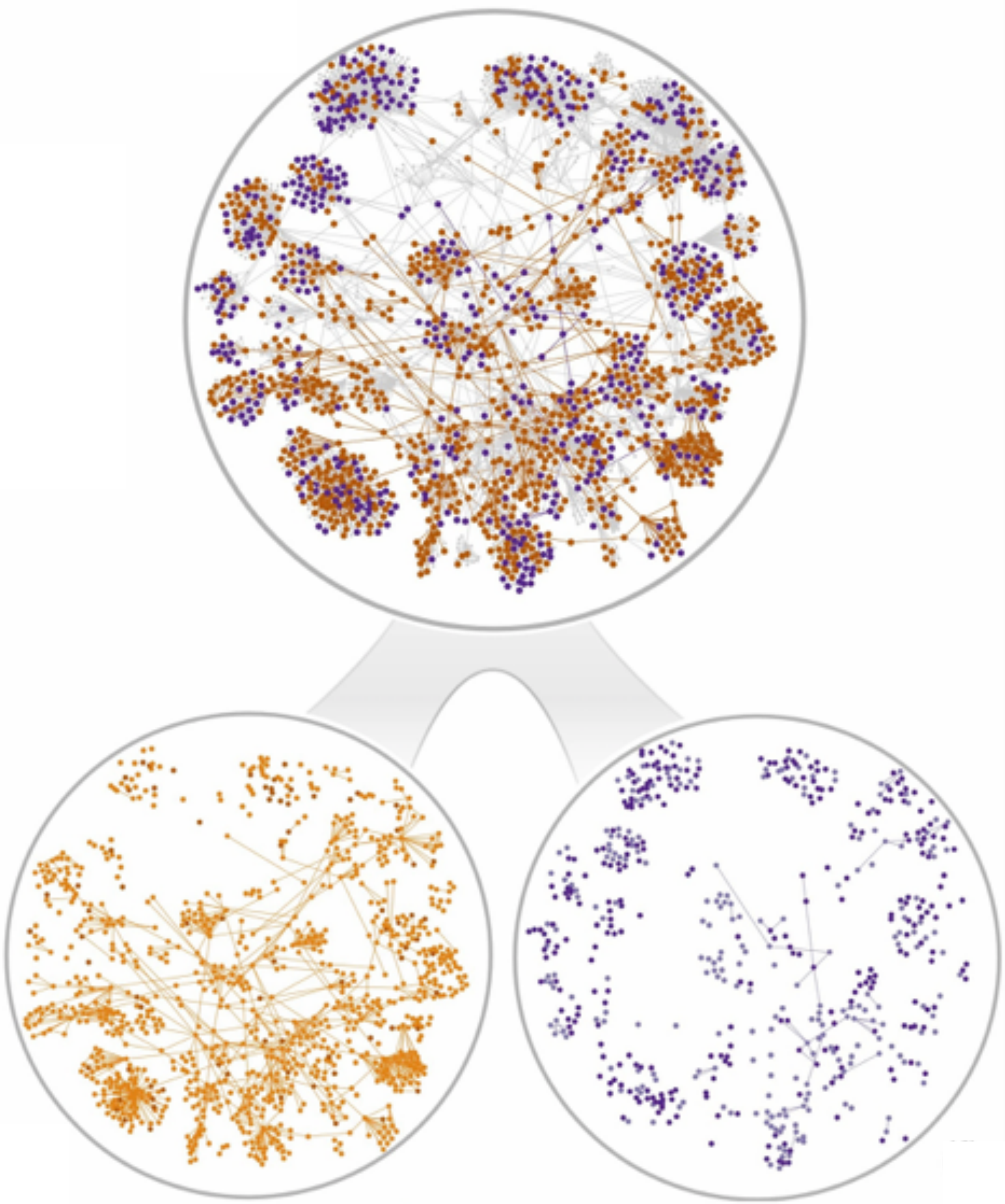
Karsai et.al. (submitted)

$G=(V, E_i),\ i=1 \ldots M$

- Nodes can be present in multiple networks simultaneously
- These networks are connected (can influence each other) via the common nodes



Gomes et.al. Phys. Rev. Lett. 110, 028701 (2013)



Skype adoption network

Nodes - users, Links - social ties,

Colours - service adoption/termination

Temporal and evolving networks

$G=(V, E_t), (u,v,t,d) \in E_t$

t - time of interaction (u,v)

d - duration of interaction (u,v,t)

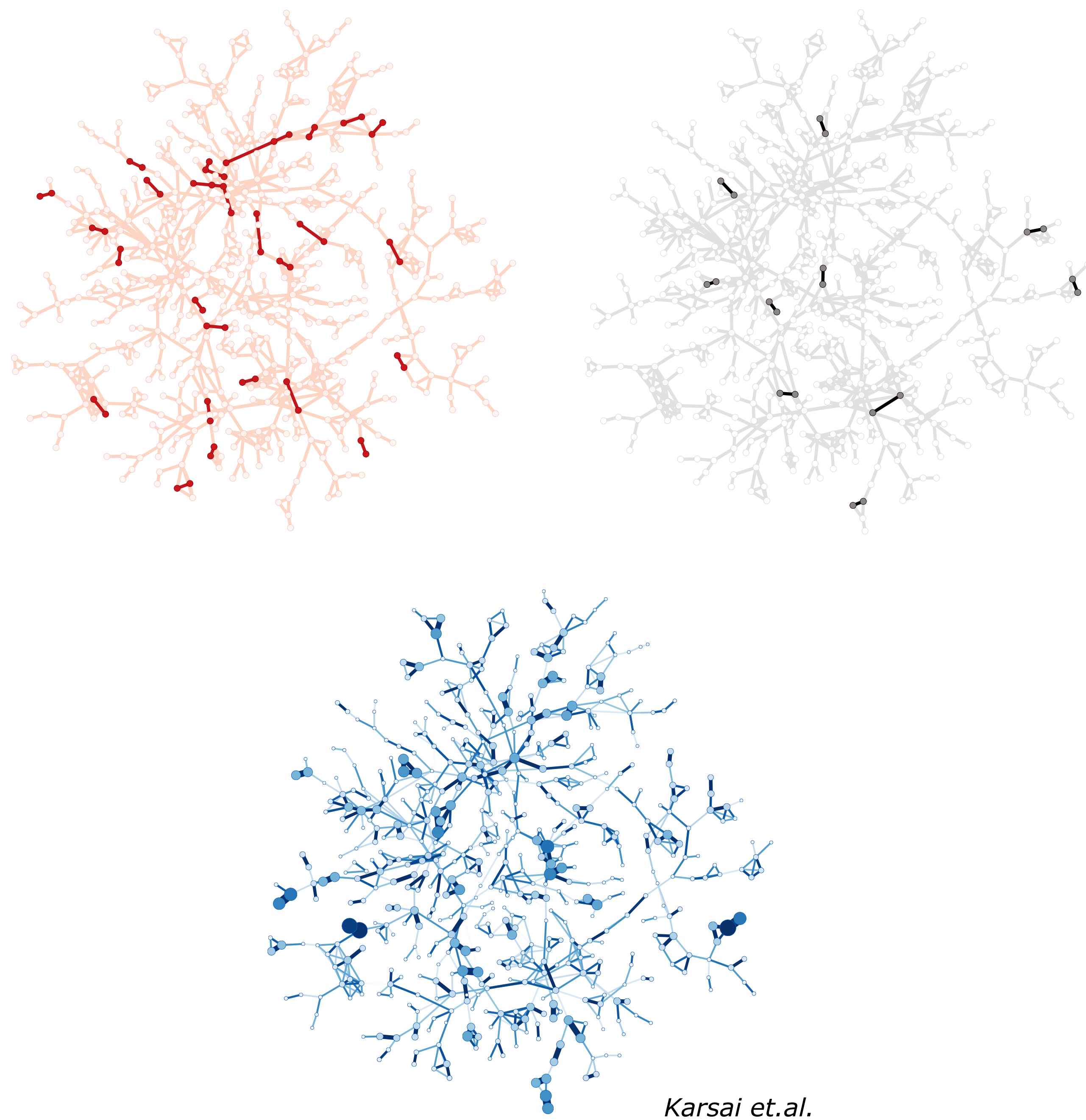
- Temporal links encode time varying interactions

$G=(V_{t'}, E_{t'})$

$v(t) \in V_{t'}$

$(u,v,t) \in E_{t'}$

- Dynamical nodes and links encode the evolution of the network
- Usually t<<t'



Karsai et.al.

Mobile communication network

Nodes - individuals

Links - calls and SMS

WHY

and

WHY NOW?

Why?

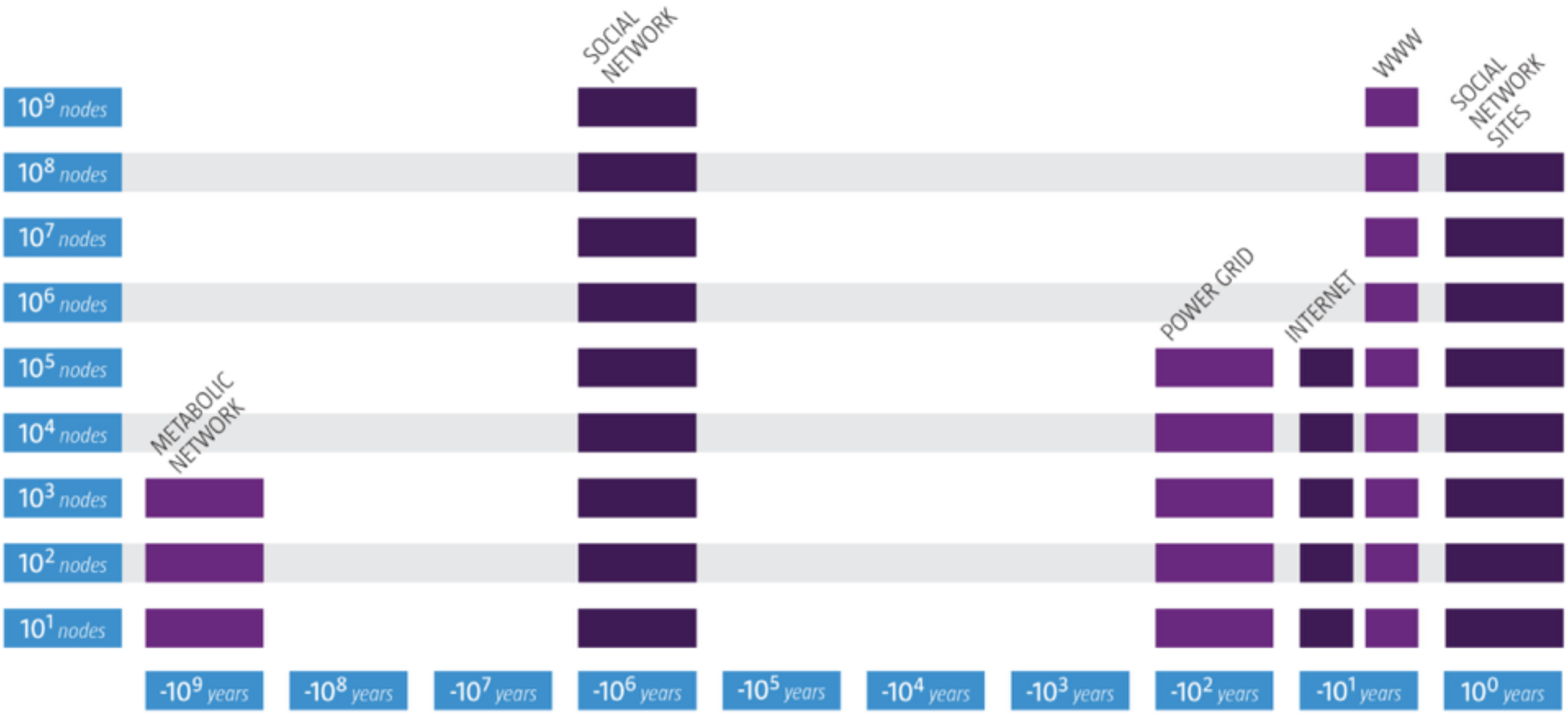
- A common framework applicable to many systems
- Different systems can be studied with same methods
- A “birds-eye” view on the system

MANY NETWORKS SHARE SIMILAR CHARACTERISTICS

- Similar processes shape the networks

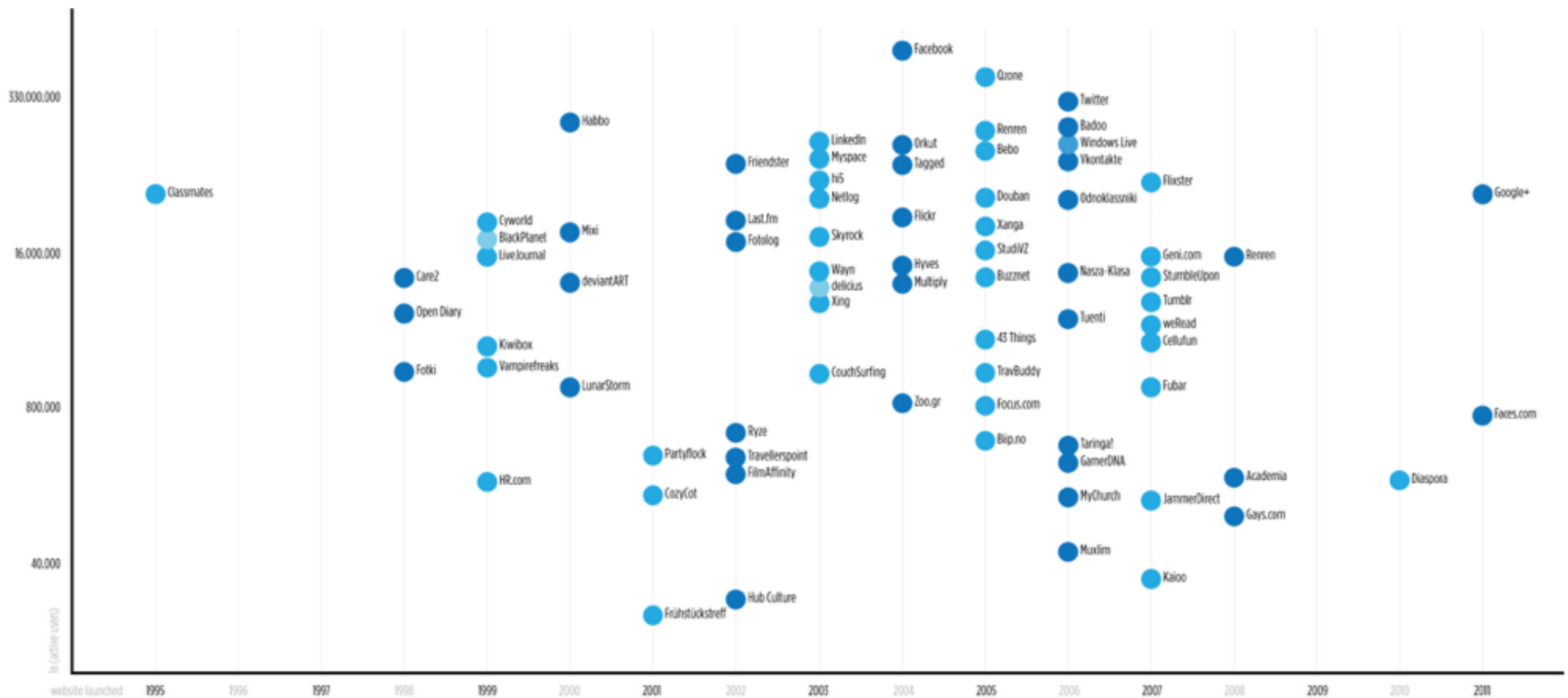
WE WILL NEVER UNDERSTAND COMPLEX SYSTEMS UNLESS WE MAP OUT AND UNDERSTAND THE NETWORKS BEHIND THEM *AL Barabási*

Why now?



Slide from CCNR Complex Networks Course
A. L. Barabási 2012

Why now?



Slide from CCNR Complex Networks Course
A. L. Barabási 2012

Why now?

- Data availability - the Big Data Revolution
- Universality - similar features of very similar systems
- Urgent need to understand complexity
  - Economic impact
  - Drug design, metabolic engineering
  - Human decease network
  - Fighting, terrorism and military
  - Epidemic forecast
  - Brain research

Network representation and properties

(Large) Network representation

	Advantage	Disadvantage	
Adjacency matrix	Direct access	Memory hungry	
Edge list	Memory friendly	Slow search	
Neighbour list	Memory optimised	Fast search	

Adjacency matrix

	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	0	0	0	0	0
2	1	0	1	1	0	0	0	0	0	0
3	0	1	0	1	0	0	0	0	0	0
4	0	1	1	0	1	0	1	0	0	0
5	0	0	0	1	0	1	0	1	0	0
6	0	0	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	1	0

Edge list

1	2
2	3
2	4
3	4
4	5
4	7
5	6
5	8
9	10

Neighbour list

1	2
2	1 3 4
3	2 4
4	2 3 5 7
5	4 6 8
6	5
7	4
8	5
9	10
10	9

(Large) Network representation

Directed networks

- Non-symmetric adjacency matrix
- Order of adjacency link list

	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0	0
4	0	1	1	0	0	0	1	0	0	0
5	0	0	0	1	0	1	0	1	0	0
6	0	0	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	1	0

```
1    2
2    1 3 4
3    2
4    2 3 5 7
5    4 6 8
6
7    4
8
9    10
10   9
```

```
1 2
2 3
2 4
3 4
4 5
4 7
5 6
5 8
9 10
```

Weighted networks

- Matrix element assign the weight of links
- Edges: non-zero elements
- (e,w) tuples
- (b,e,w) triplets

	1	2	3	4	5	6	7	8	9	10
1	0	6	0	0	0	0	0	0	0	0
2	6	0	1	3	0	0	0	0	0	0
3	0	1	0	3	0	0	0	0	0	0
4	0	3	3	0	6	0	4	0	0	0
5	0	0	0	6	0	1	0	3	0	0
6	0	0	0	0	1	0	0	0	0	0
7	0	0	0	4	0	0	0	0	0	0
8	0	0	0	0	3	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	4
10	0	0	0	0	0	0	0	0	4	0

```
1    (2,6)
2    (1,3) (3,1) (4,3)
3    (2,1) (4,3)
4    (2,3) (3,3) (5,6) (7,4)
5    (4,6) (6,1) (8,3)
6    (5,1)
7    (4,4)
8    (5,3)
9    (10,4)
10   (9,4)
```

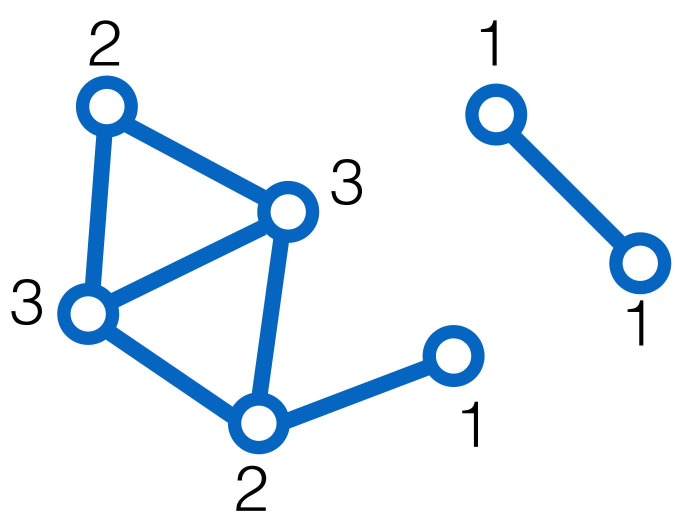
```
1 2 6
2 3 1
2 4 3
3 4 3
4 5 6
4 7 4
5 6 1
5 8 3
9 10 4
```

Node degree

Number of connections of a node

	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	0	0	0	0	0
2	1	0	1	1	0	0	0	0	0	0
3	0	1	0	1	0	0	0	0	0	0
4	0	1	1	0	1	0	1	0	0	0
5	0	0	0	1	0	1	0	1	0	0
6	0	0	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	1	0

- Undirected network



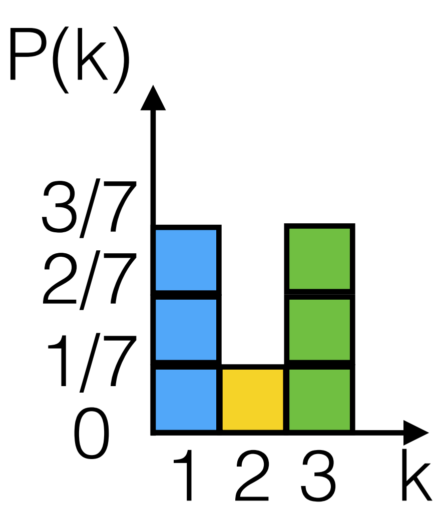
$$k_i = A_{i1} + A_{i2} + \ldots + A_{iN} = \sum_j^N A_{ij}$$

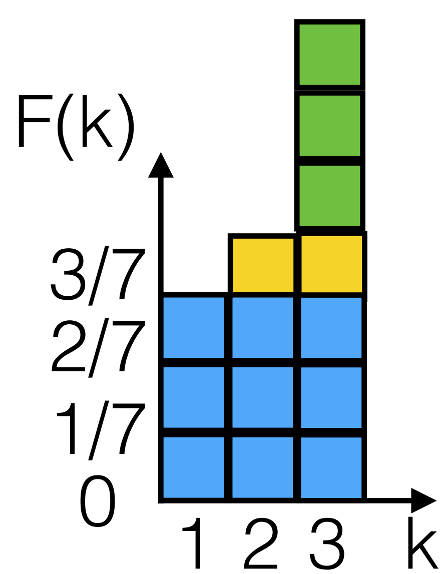
$$m = \frac{\sum_i k_i}{2} \quad \text{where} \quad m = |E|$$

mean degree $$\langle k \rangle = \frac{1}{N}\sum_i^N k_i$$

- Degree distribution: PDF of k or CDF of k

$$P(k) = \frac{n_k}{\sum_k n_k} \qquad F(k) = \sum_{k_i=1}^{k} P(k_i)$$





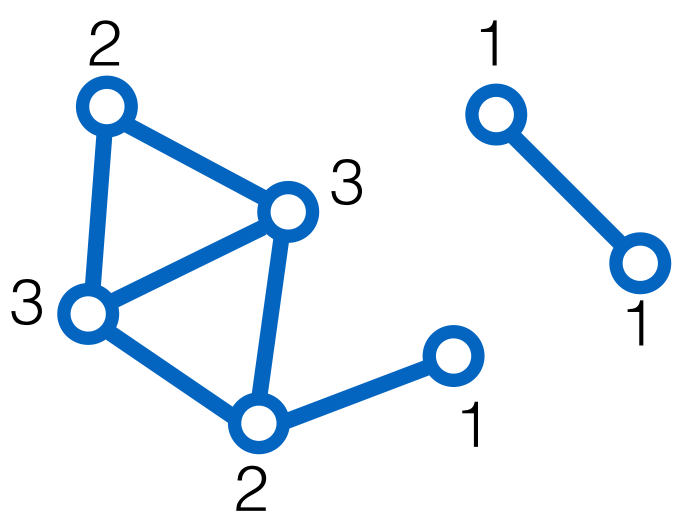
- normalisation condition

$$\sum_{k=1}^{\infty} P(k) = 1 \qquad \lim_{k\to\infty} F(k) = 1$$

Node degree

Number of connections of a node

- Undirected network



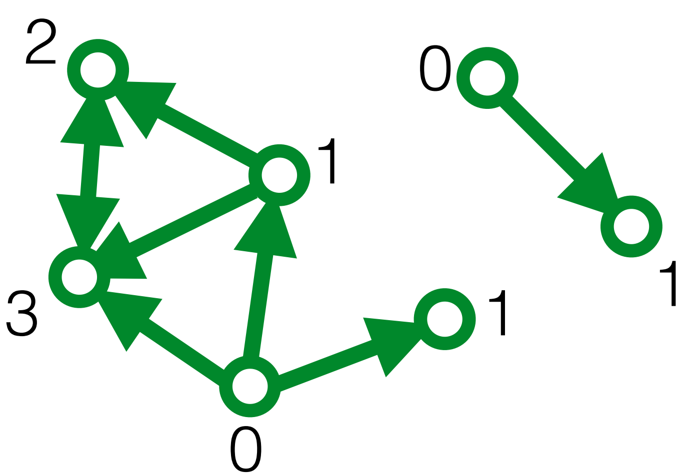
$$k_i = A_{i1} + A_{i2} + \ldots + A_{iN} = \sum_j^N A_{ij}$$

$$m = \frac{\sum_i k_i}{2} \quad \text{where} \quad m = |E|$$

mean degree $$\langle k \rangle = \frac{1}{N} \sum_i^N k_i$$

	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	0	0	0	0	0
2	1	0	1	1	0	0	0	0	0	0
3	0	1	0	1	0	0	0	0	0	0
4	0	1	1	0	1	0	1	0	0	0
5	0	0	0	1	0	1	0	1	0	0
6	0	0	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	1	0

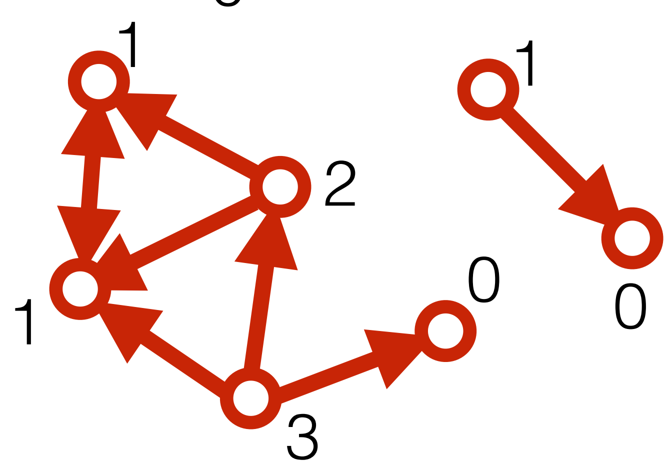
- Directed network



In degree $$k_i^{in} = \sum_j^N A_{ij}$$

Out degree $$k_j^{out} = \sum_i^N A_{ij}$$

	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	0	0	0	0	0
2	1	0	1	1	0	0	0	0	0	0
3	0	1	0	1	0	0	0	0	0	0
4	0	1	1	0	1	0	1	0	0	0
5	0	0	0	1	0	1	0	1	0	0
6	0	0	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	1	0



$$m = \sum_i^N k_i^{in} = \sum_j^N k_j^{out} = \sum_{ij} A_{ij}$$

$$\langle k^{in} \rangle = \frac{1}{N} \sum_{i=1}^N k_i^{in} = \frac{1}{N} \sum_{j=1}^N k_j^{out} = \langle k^{out} \rangle$$

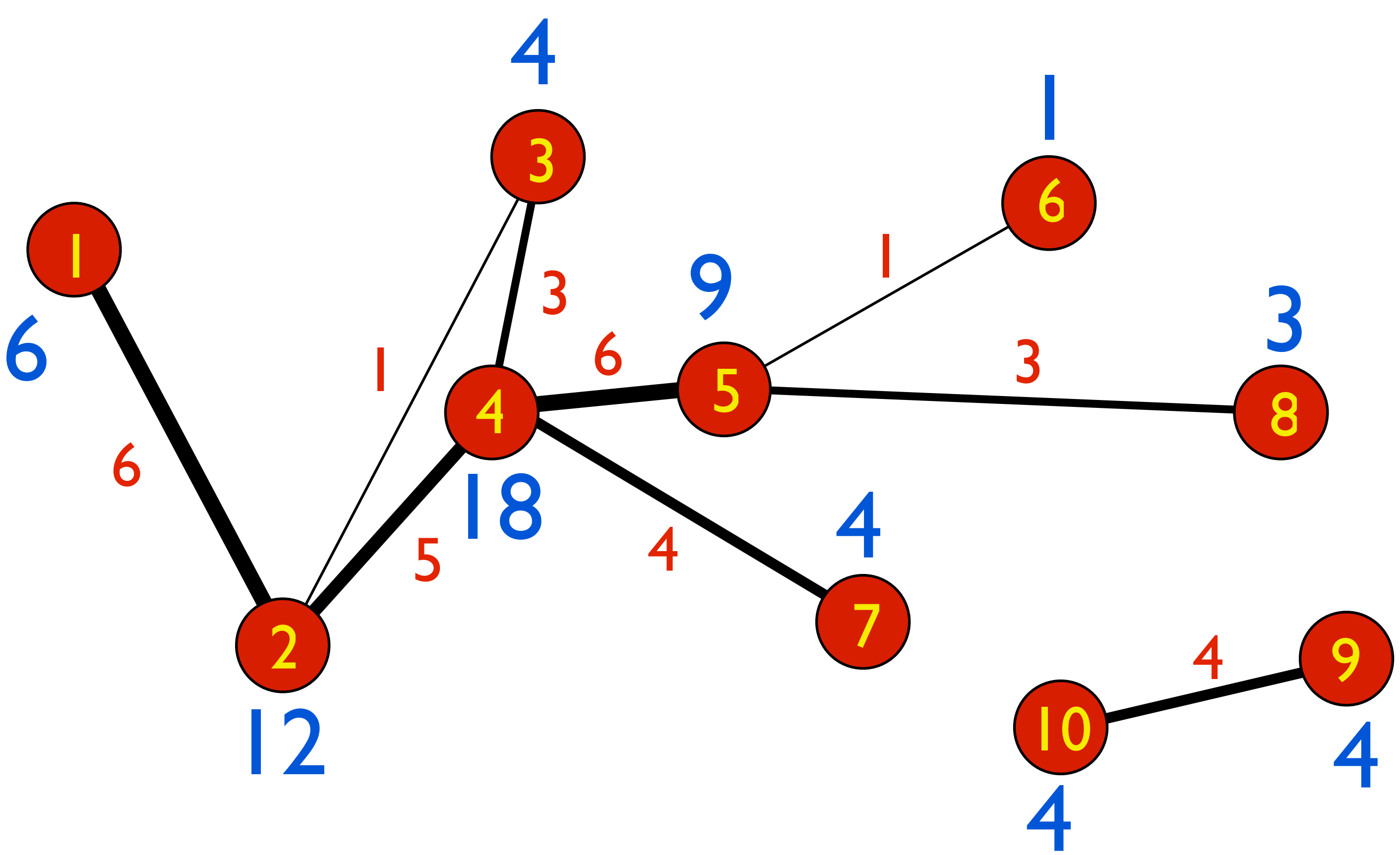
	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	0	0	0	0	0
2	1	0	1	1	0	0	0	0	0	0
3	0	1	0	1	0	0	0	0	0	0
4	0	1	1	0	1	0	1	0	0	0
5	0	0	0	1	0	1	0	1	0	0
6	0	0	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	1	0

Weighted degree: strength

- Weighted networks

The sum of the weights of links connected to node i

$$s_i = w_{i1} + w_{i2} + \ldots + w_{iN} = \Sigma_j\, w_{ij}$$



	1	2	3	4	5	6	7	8	9	10
1	0	6	0	0	0	0	0	0	0	0
2	6	0	1	3	0	0	0	0	0	0
3	0	1	0	3	0	0	0	0	0	0
4	0	3	3	0	6	0	4	0	0	0
5	0	0	0	6	0	1	0	3	0	0
6	0	0	0	0	1	0	0	0	0	0
7	0	0	0	4	0	0	0	0	0	0
8	0	0	0	0	3	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	4
10	0	0	0	0	0	0	0	0	4	0

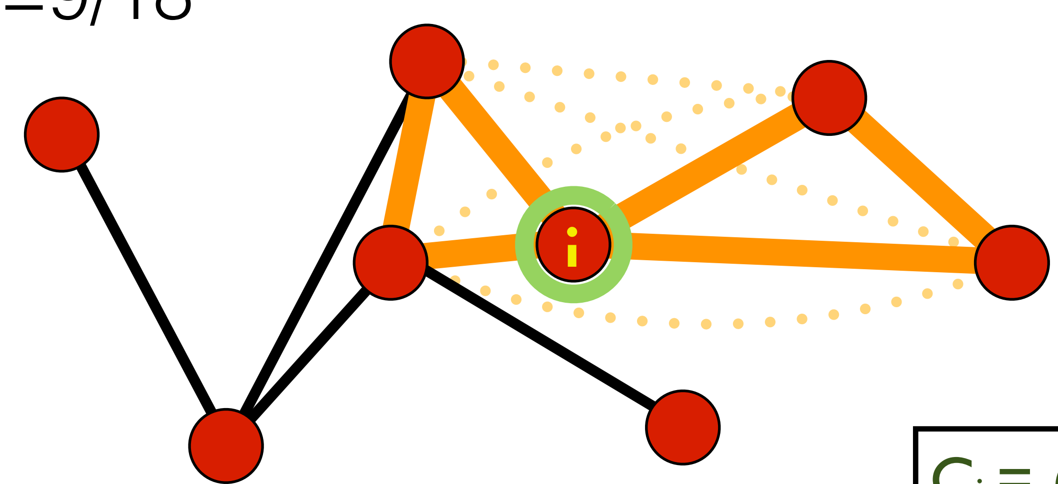
Node clustering coefficient

- Measure of interconnectivity
- What portion of neighbours of a node are connected to each other?

Global clustering coefficient

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of vertices}} = \frac{\text{number of closed triplets}}{\text{number of connected triples of vertices}}.$$

C=9/18



$C_i = (2x2)/(4x3) = 1/3$

Local clustering coefficient

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

- e_i - number of links between the neighbours of node i
- $(k_i(k_i-1))/2$ - maximum number of triangles

Average local clustering coefficient

$$\langle C_{WS} \rangle = \frac{\sum_i C_i}{N}$$

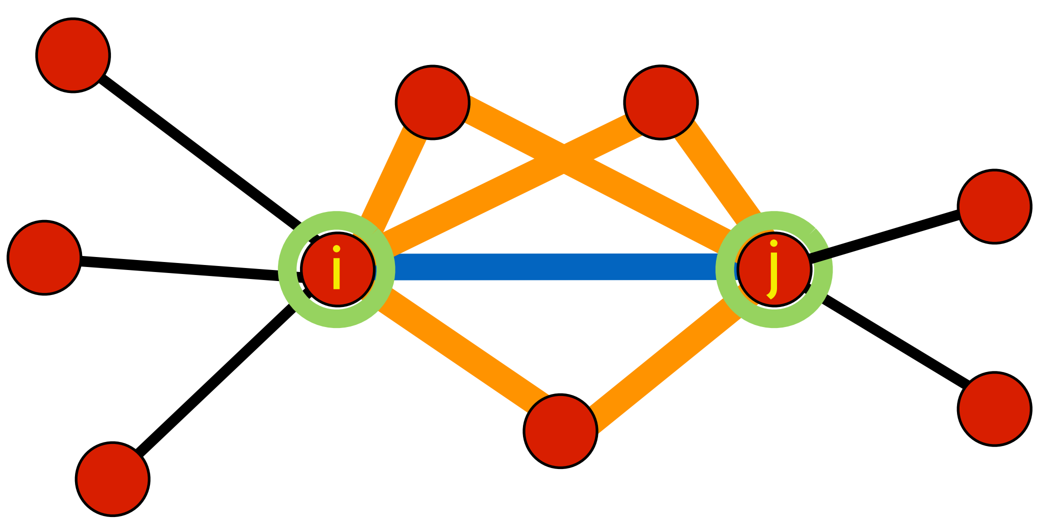
Definition: Watts and Strogatz 2002

Link clustering coefficient: Overlap

- Link property
- Fraction of common neighbours of a connected pair
- Jaccard index of common neighbours

$$O_{ij} = \frac{n_{ij}}{(k_i - 1) + (k_i - 1) - n_{ij}}$$

- n_i - number of common neighbours of nodes i and j
- $(k_i\text{-}1)+(k_j\text{-}1)\text{-}n_{ij}$ maximum number possible triangles between nodes i and j



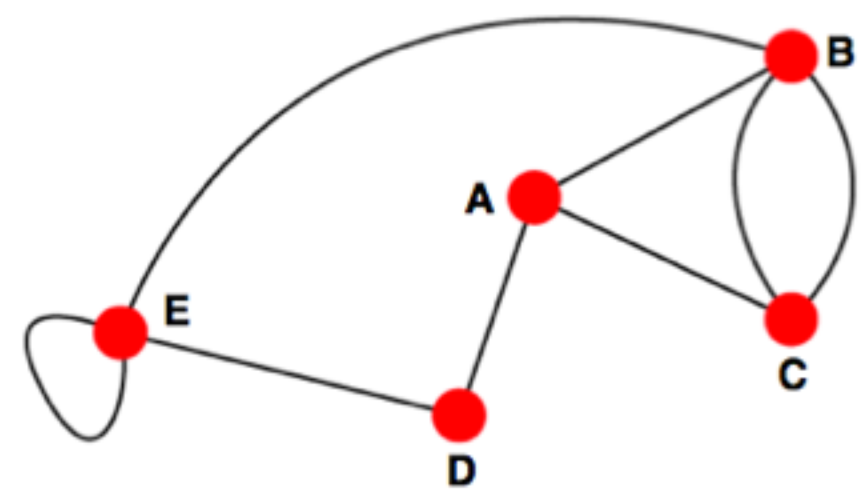
O_{ij} = 3/(6+5-3) = 3/8

Path length

A ***path is*** a sequence of nodes in which each node is adjacent to the next one

$P_{i0,in}$ of length ***n*** between nodes i_0 and i_n is an ordered collection of ***n+1*** nodes and ***n*** links

$$P_n = \{i_0, i_1, i_2, \ldots, i_n\} \qquad P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{n-1}, i_n)\}$$

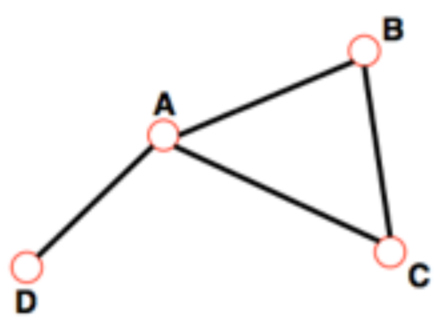


- A path can intersect itself and pass through the same link repeatedly. Each time a link is crossed, it is counted separately

- A legitimate path on the graph on the right:
ABCBCADEEBA

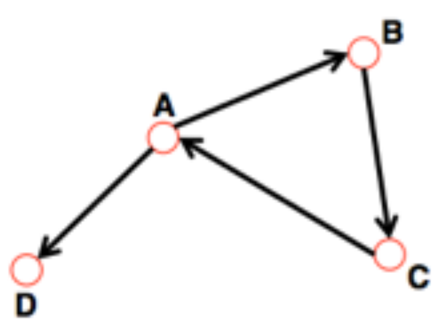
- In a directed network, the path can follow only the direction of an arrow.

Path length



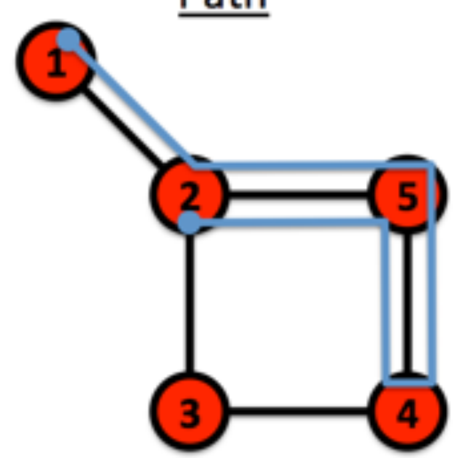
The ***distance (shortest path, geodesic path)*** between two nodes is defined as the number of edges along the shortest path connecting them.

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



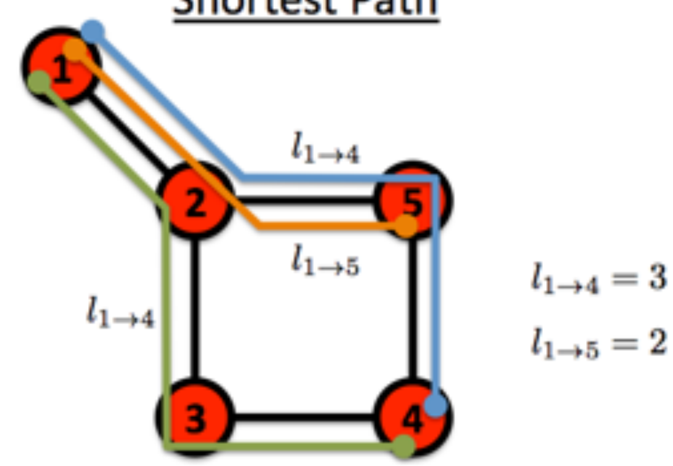
In directed graphs each path needs to follow the direction of the arrows.
Thus in a digraph the distance from node A to B (on an AB path) is generally different from the distance from node B to A (on a BCA path).

Path



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

Shortest Path



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

Slide from CCNR Complex Networks Course
A. L. Barabási 2012

Path length

N_{ij},number of paths between any two nodes *i* and *j*:

Length n=1: If there is a link between ***i*** and ***j***, then A_{ij}=1 and A_{ij}=0 otherwise.

Length n=2: If there is a path of length two between ***i*** and ***j***, then $A_{ik}A_{kj}$=1, and $A_{ik}A_{kj}$=0 otherwise.
The number of paths of length 2:

$$N_{ij}^{(2)} = \sum_{k=1}^{N} A_{ik}A_{kj} = [A^2]_{ij}$$

Length n: In general, if there is a path of length ***n*** between ***i*** and ***j***, then $A_{ik} \ldots A_{lj}$=1 and $A_{ik} \ldots A_{lj}$=0 otherwise.
The number of paths of length ***n*** between ***i*** and ***j*** is*

$$N_{ij}^{(n)} = [A^n]_{ij}$$

****holds for both directed and undirected networks.***

- Finding shortest path: Breadth-first search algorithm

Slide from CCNR Complex Networks Course
A. L. Barabási 2012

Path length

- d_{max} diameter- the maximum distance between any pairs of nodes
- $\langle d \rangle$ average path length - for directed graphs

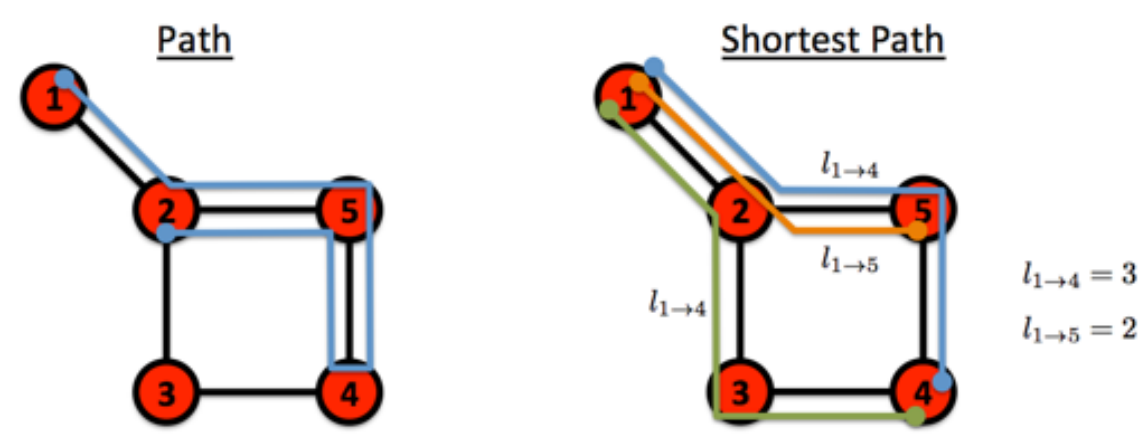
$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}$$

- where d_{ij} is the shortest distance between nodes *i* and *j*
- multiplicative is *(2 x max number of links)*
- distance between unconnected nodes is 0

- $\langle d \rangle$ average path length - for un-directed graphs

$$\langle d \rangle = \frac{2}{N(N-1)} \sum_{i < j} d_{ij}$$

- since $d_{ij} = d_{ji}$
- multiplicative is *(max number of links)*



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

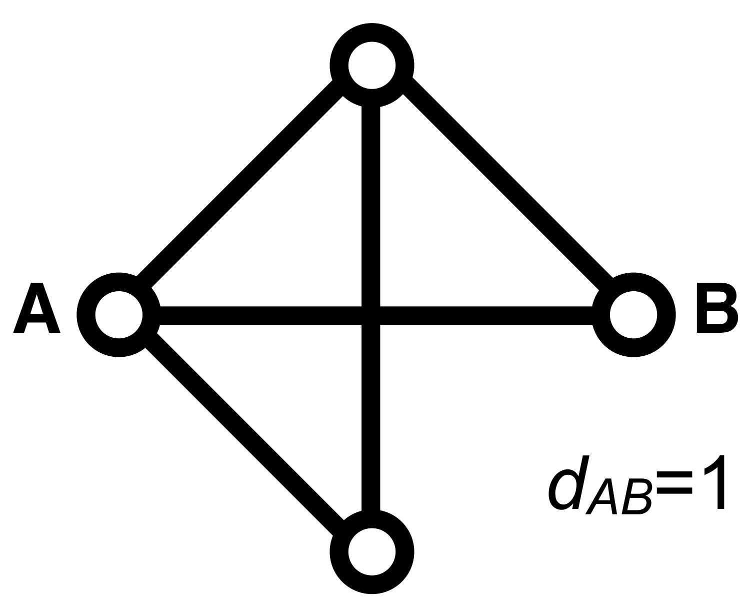
The path with the shortest length between two nodes (distance).

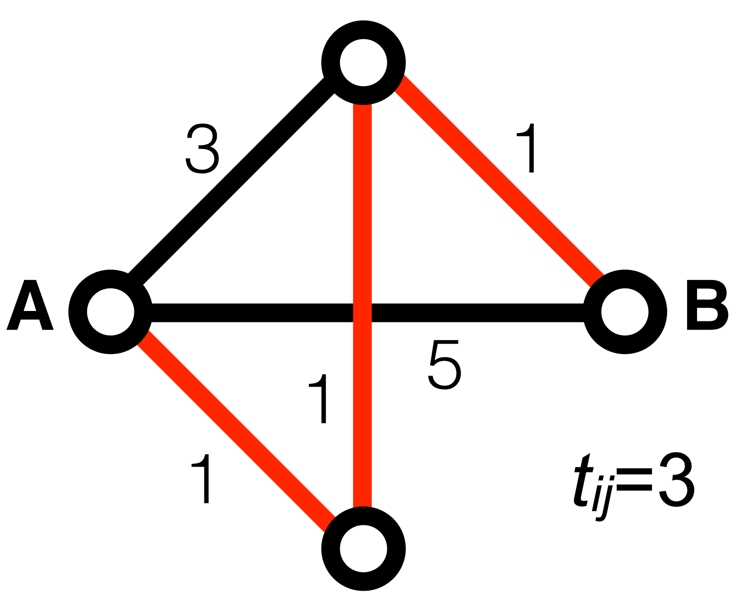
Weighted path length

length of a shortest path $P(i \to j)$ ≠ length of a weighted shortest path $P(i \to j)$

$$d_{ij} = \sum_{e_{mn} \in P(i \to j)} A_{mn} \qquad t_{ij} = \sum_{e_{mn} \in P(i \to j)} w_{mn}$$

Shortest path ≠ Weighted shortest path





Central quantities in network analysis

- Degree distribution: $P(k)$
- Clustering coefficient: C
- Average path length: $<d>$

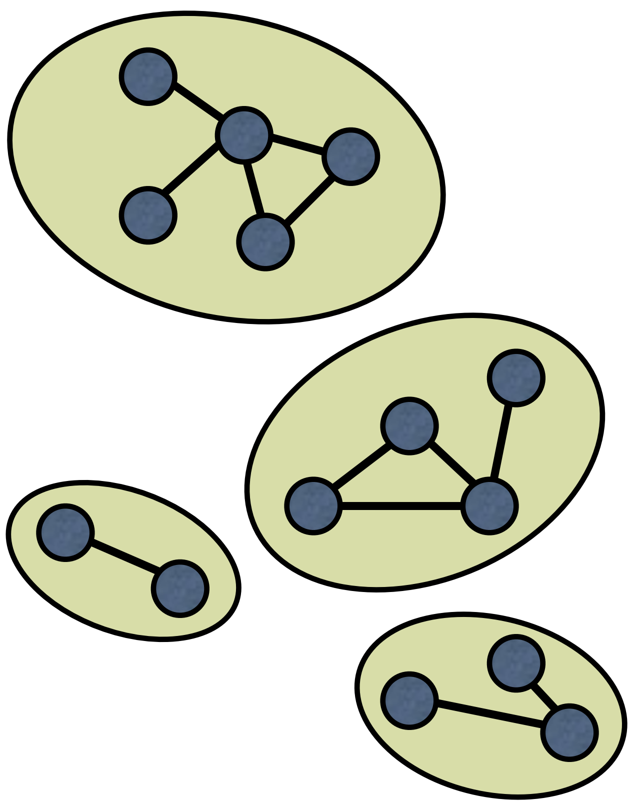
Network density

$$\rho = \frac{m}{\binom{N}{2}} = \frac{2m}{N(N-1)}$$

- Ratio of the numbers of existing edges m and possible number of edges $(N(N\text{-}1))/2$ in a network of size N
- If the network is fully connected: $\rho=1$
- If $\rho \rightarrow 0$ (or tends to a constant) as network size $N \rightarrow \infty$ the network is sparse
- Real networks are usually sparse

Connectivity and components

- A connected component is a subset of vertices with at least one path connecting each of them
- A network may consist of a single connected component (a connected network) or several of those
- Distances between nodes in disjoint components are not defined (infinite)
- Bridge: if we erase it, the graph becomes disconnected.
- 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



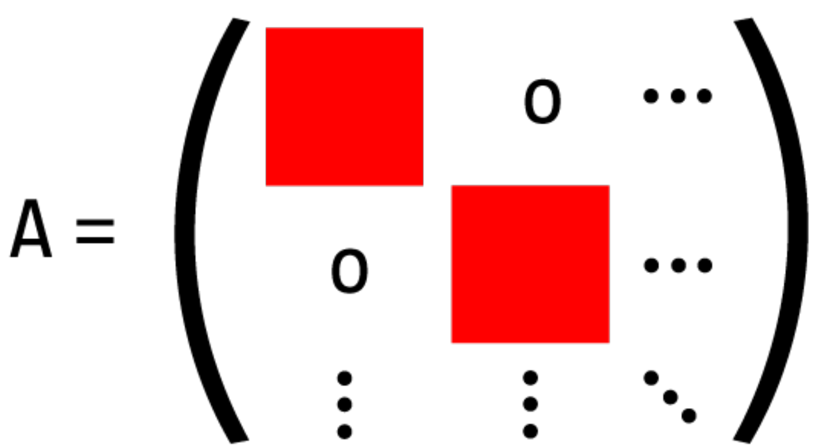


Figure after Newman, 2010

Connectivity and components - directed networks

- Strongly connected component (SCC): has a path from each node to every other node in the component
- Weakly connected component (WCC): it is connected if we disregard the directions
- In-component: nodes that can reach the SCC
- Out-component: nodes that can be reached from SCC

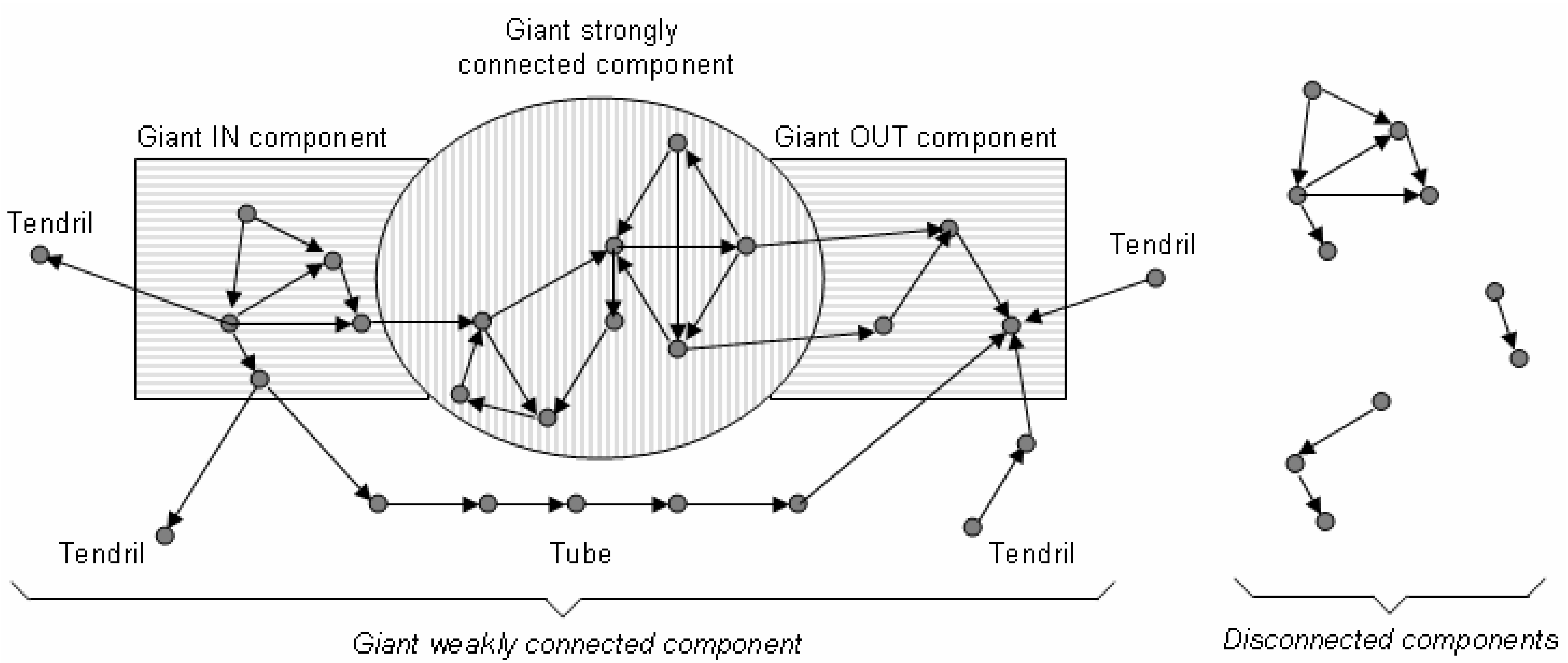
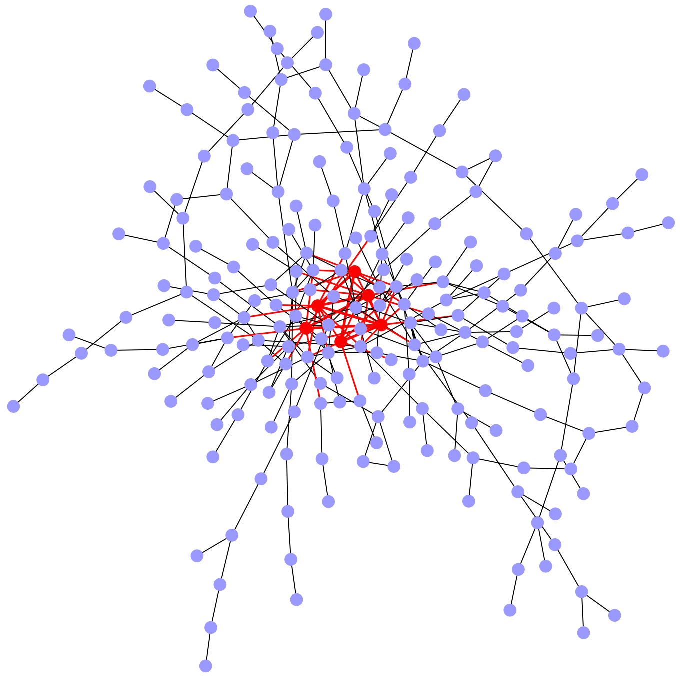


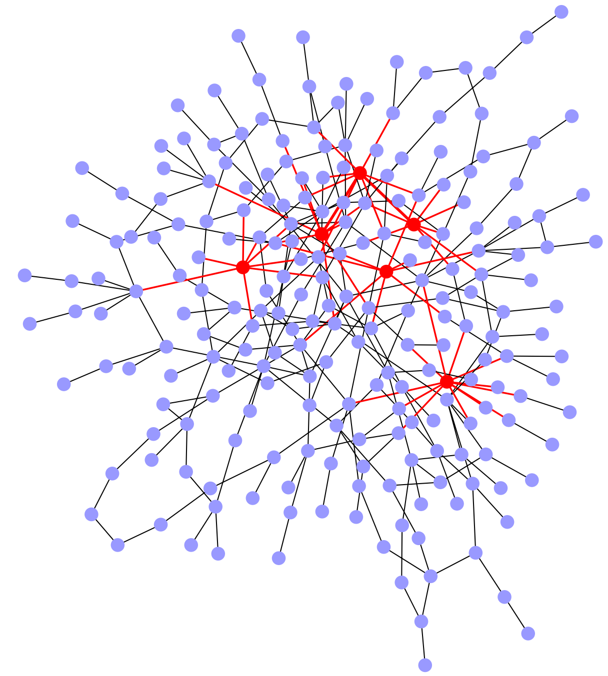
Figure from Broder et. al. (2000)

Measures of degree correlations



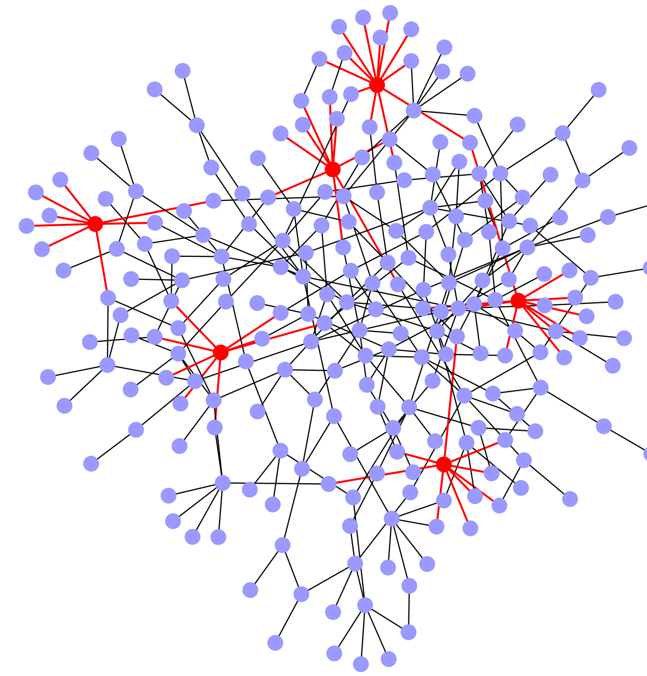
Assortative:

hubs show a tendency to link to each other.



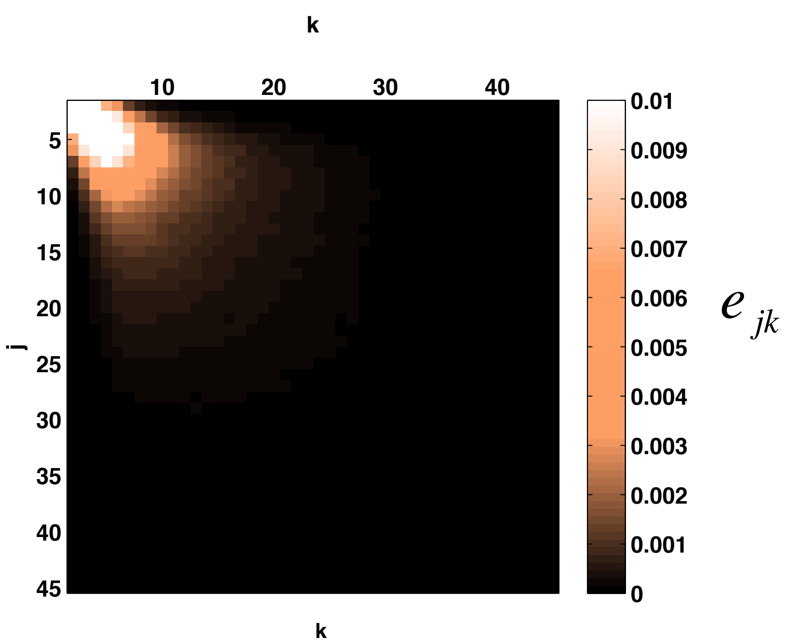
Neutral:

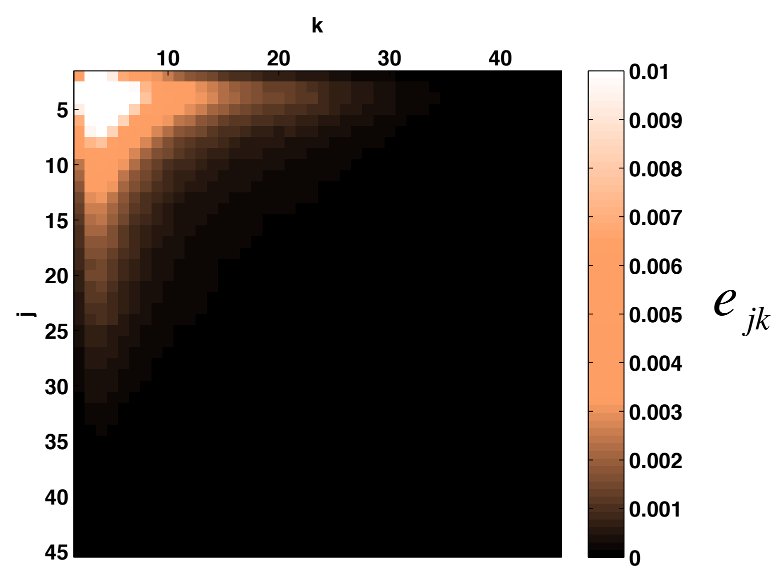
nodes connect to each other with the expected random probabilities.

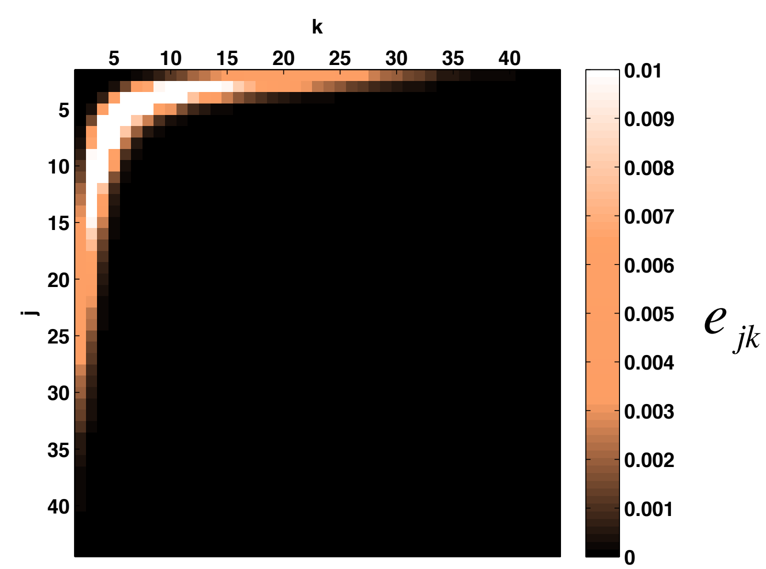


Disassortative:

Hubs tend to avoid linking to each other.







Figures from CCNR Complex Networks Course
A. L. Barabási 2012

Pearson degree-degree correlation

- q_k: the probability having degree k of a node in the end of a randomly selected link

$$q_k = \frac{kp_k}{\langle k \rangle}$$

  - it is biased towards high degree nodes

- e_{jk}: probability to find a node with degree j and degree k at the two ends of a randomly selected edge:

$$\sum_{jk} e_{jk} = 1, \qquad \sum_{j} e_{jk} = q_k$$

  - if the network has no degree correlations: $e_{jk} = q_j q_k$
  - Any deviation from this assigns correlations ⟹ calculate the degree-degree correlation function

$$r = \frac{1}{\sigma_q^2} \sum_{jk} jk(e_{jk} - q_j q_k) \qquad -1 \le r \le 1$$

$$\sigma_r^2 = \max \sum_{jk} jk(e_{jk} - q_j q_k)$$

normalised by the maximum r value

- if
  - $r < 0 \Rightarrow$ disassortative network
  - $r = 0 \Rightarrow$ neutral network
  - $r > 0 \Rightarrow$ assortative network

M. E. J. Newman, Phys. Rev. Lett. **89,** 208701 (2002)

Average nearest-neighbour degree

R. Pastor-Satorras, A. Vázquez, A. Vespignani, Phys. Rev. E 65, 066130 (2001)

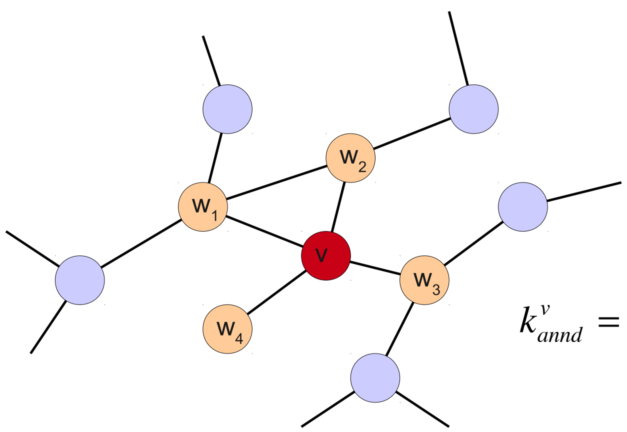
- Easier characterisation of degree-degree correlations
- Average degree of the first neighbours of node v

$$\bar{k}^v_{nn} = \frac{1}{k_v} \sum_j k_{v,j}$$

- Average nearest neighbour degree of nodes with degree k

$$\bar{k}_{nn}(k) = \frac{1}{N_k} \sum_i k_{nn,i} \delta_{k_i,k}$$

  - where $\delta_{i,j}=1$ if $i=j$ and $\delta_{i,j}=0$ otherwise



$$k^v_{annd} = \frac{4+3+3+1}{4}$$

- k_{nn} can be written in another form:

$$\bar{k}_{nn}(k) = \sum_{k'} k' P(k'|k)$$

  - where $P(k'|k)$ is the conditional probability that and edge of a node with degree k points to a node with degree k'

Figures from CCNR Complex Networks Course
A. L. Barabási 2012

Nearest neighbour degree

- k_{nn} can be written once in another form:

$$k_{annd}(k) = \sum_{k'} k' P(k' \mid k) = \frac{\sum_{k'} k' e_{kk'}}{\sum_{k'} e_{kk'}}$$

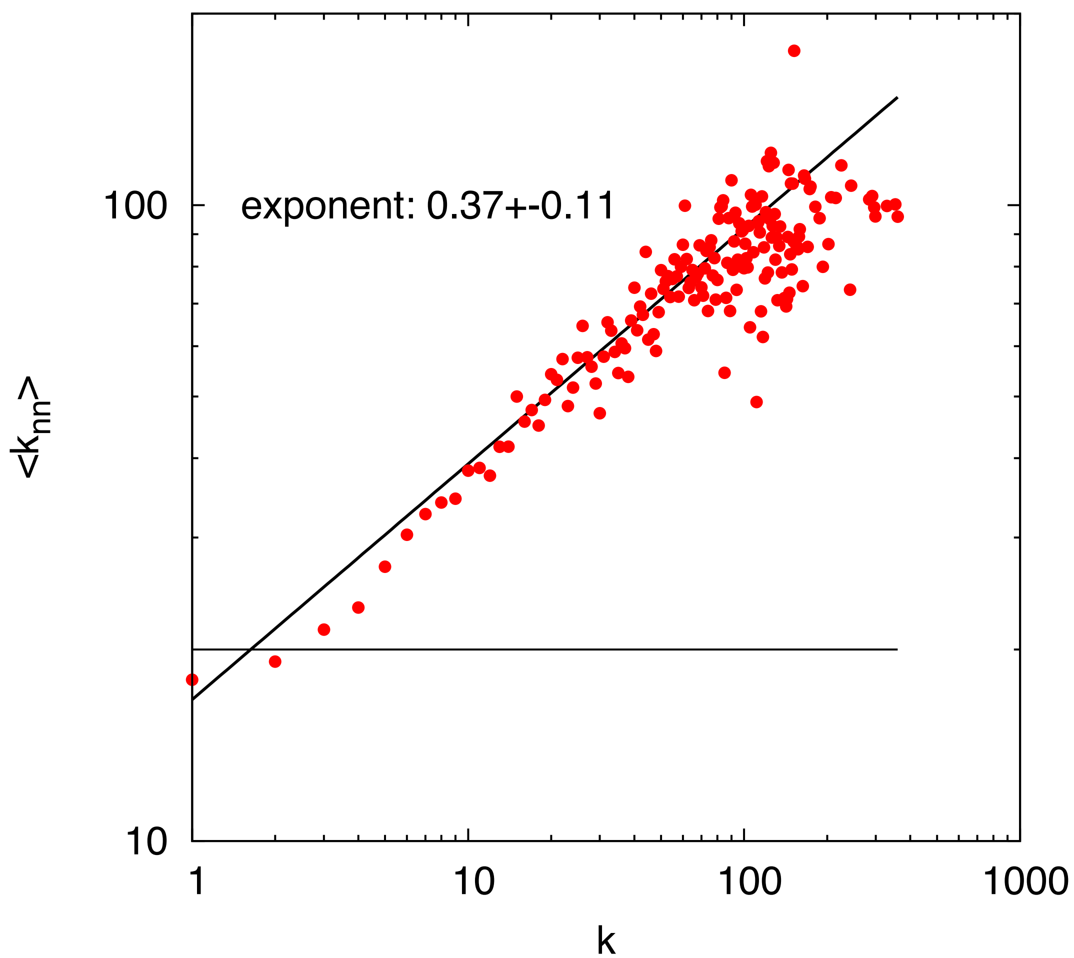
- If there are no degree correlations:

$$k_{annd}(k) = \frac{\sum_{k'} k' e_{kk'}}{\sum_{k'} e_{kk'}} = \frac{\sum_{k'} k' q_k q_{k'}}{q_k} = \sum_{k'} k' q_{k'} = \sum_{k'} k' \frac{k' p(k')}{< k >} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

  - k_{nn} is independent of k

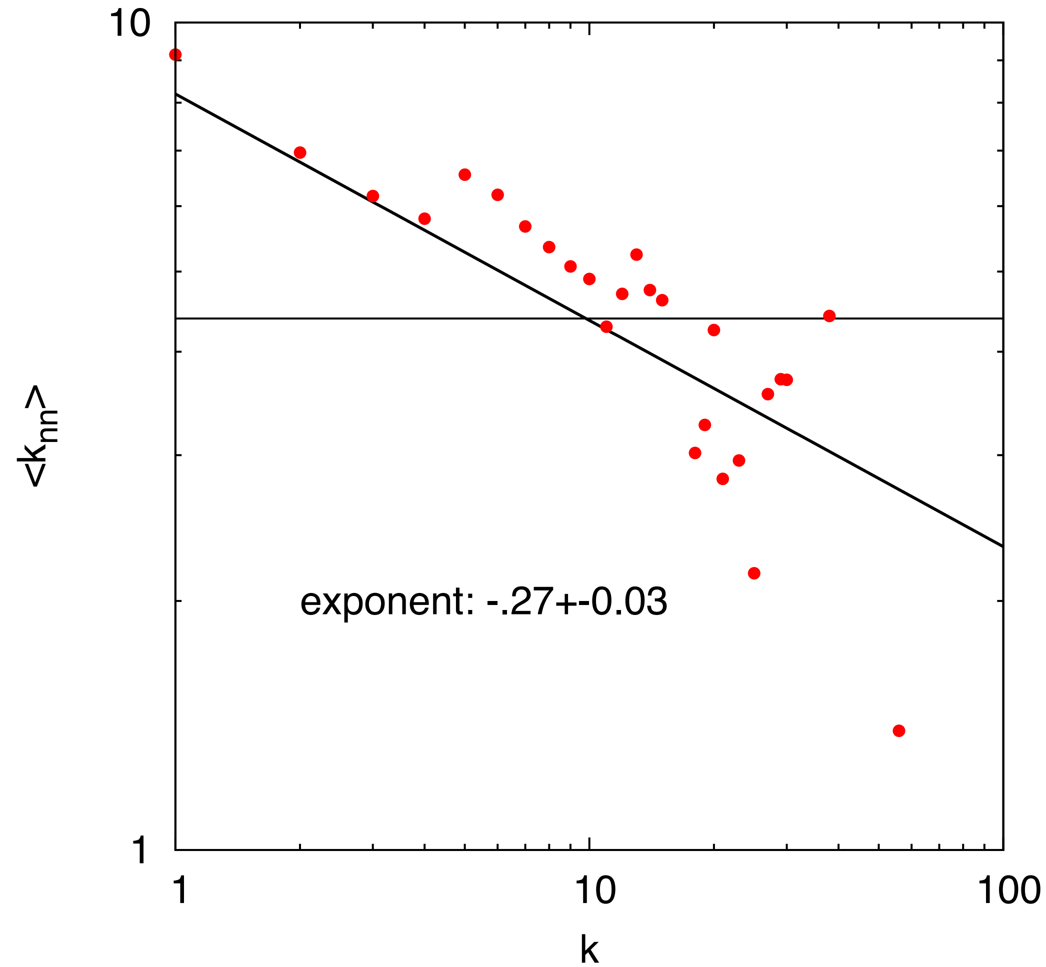
- If the network is assortative $k_{nn}(k)$ is a positive function
- If the network is disassortative $k_{nn}(k)$ is a negative function

Nearest neighbour degree



Astrophysics co-authorship network

Assortative



Yeast PPI

Disassortative

Figures from CCNR Complex Networks Course
A. L. Barabási 2012

A social experiment

Average degree

- Average degree of a node: $\langle k \rangle = \sum_k kP(K)$
- Average degree of a neighbour of a node: $\sum_k k \frac{kP(K)}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$

$$\frac{\langle k^2 \rangle}{\langle k \rangle} - \langle k \rangle = \frac{\langle k^2 \rangle - \langle k \rangle^2}{\langle k \rangle} = \frac{\sigma^2(k)}{\langle k \rangle} > 0$$

"Your friends have more friends than you have"

Experiment:

- Count the number of friends you have on Facebook
- Select 10 friends randomly and count the average number of friends they have
- Send these to count by mail to: marton.karsai@ens-lyon.fr

The study anonymised, no personal information will be undisclosed

Rich-club coefficient

- How well connected are the well connected among themselves
- It is calculated on a non-decreasing degree ranked node list as

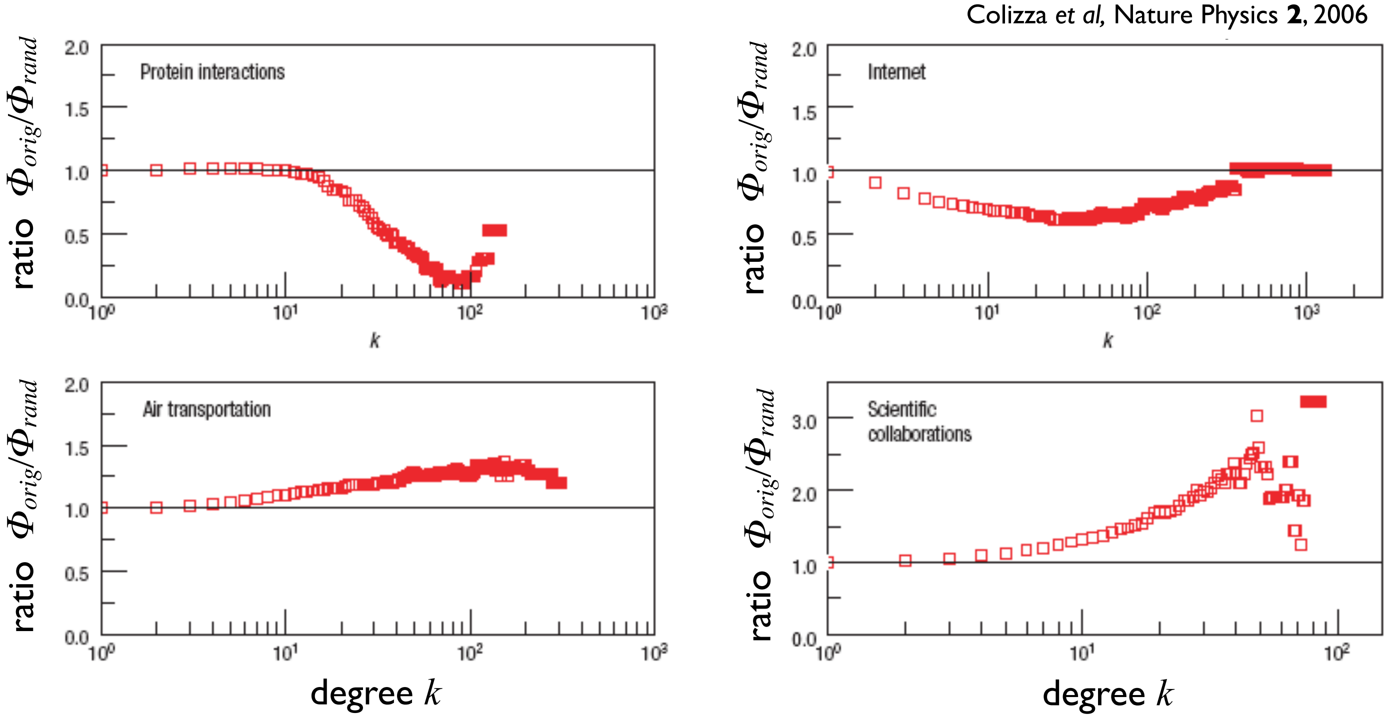
$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k}-1)}$$

- $N_{>k}$ denotes the number of nodes with degree k or larger than k
- $E_{>k}$ assigns the number of links between them
- Results are usually compared to random references
  - configuration model of equivalent synthetic network
  - configuration model of the empirical network

Algorithm

- rank nodes by degree
- remove nodes in an ascendant degree order
- measure the density of the remaining network

Colizza *et al*, Nature Physics **2**, 2006



k-core decomposition

- To identify dense cores of networks to study their importance
- Algorithm:
  1. Take a directed or undirected network
  2. Remove nodes with degree k(=1) and all of those which degree became k(=1) because of the removal process
  3. Repeat step 2 for k=2,3,… until no node can be removed
- Nodes removed in the k^{th} turn are in the k-shell and the remaining nodes form the k-core

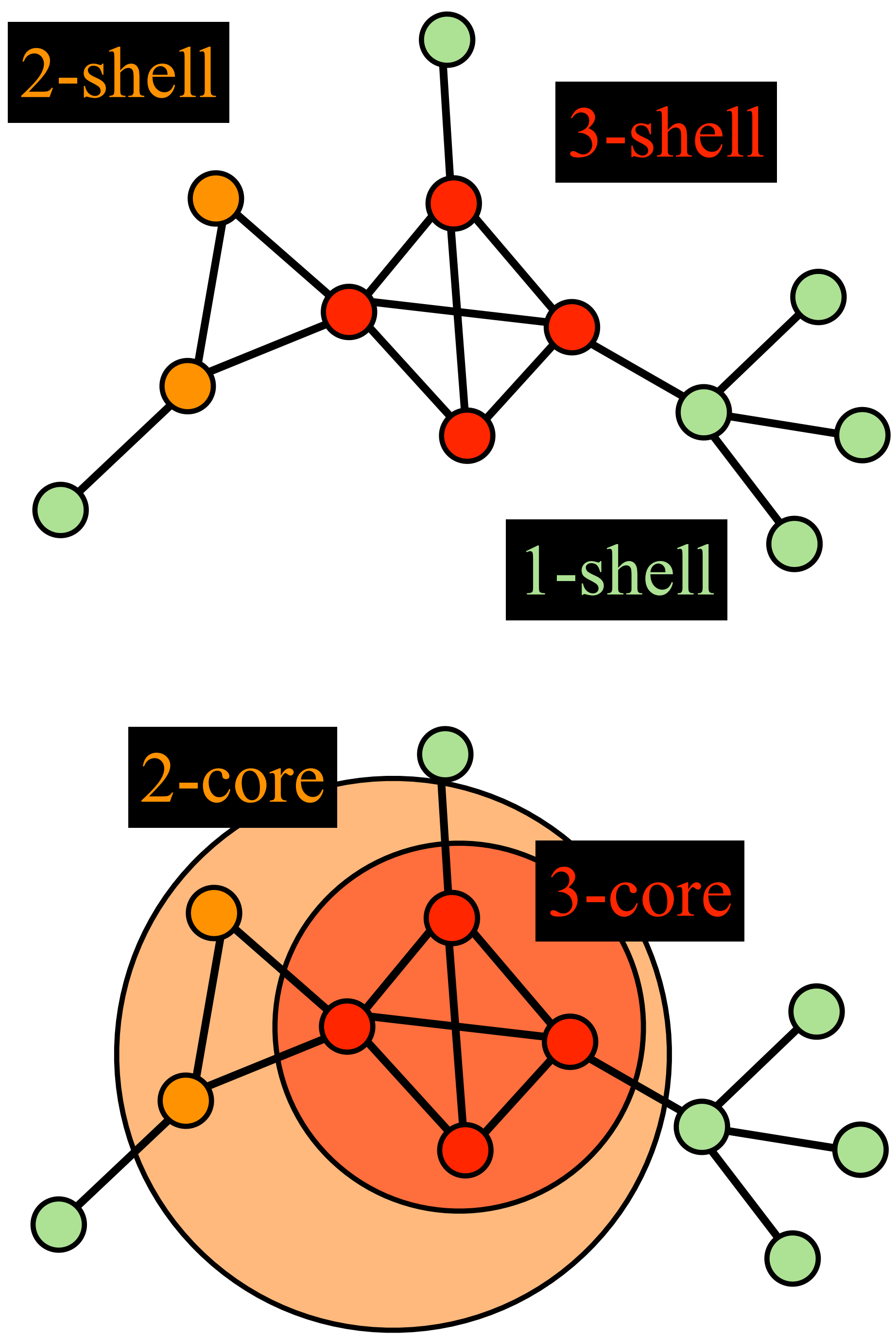
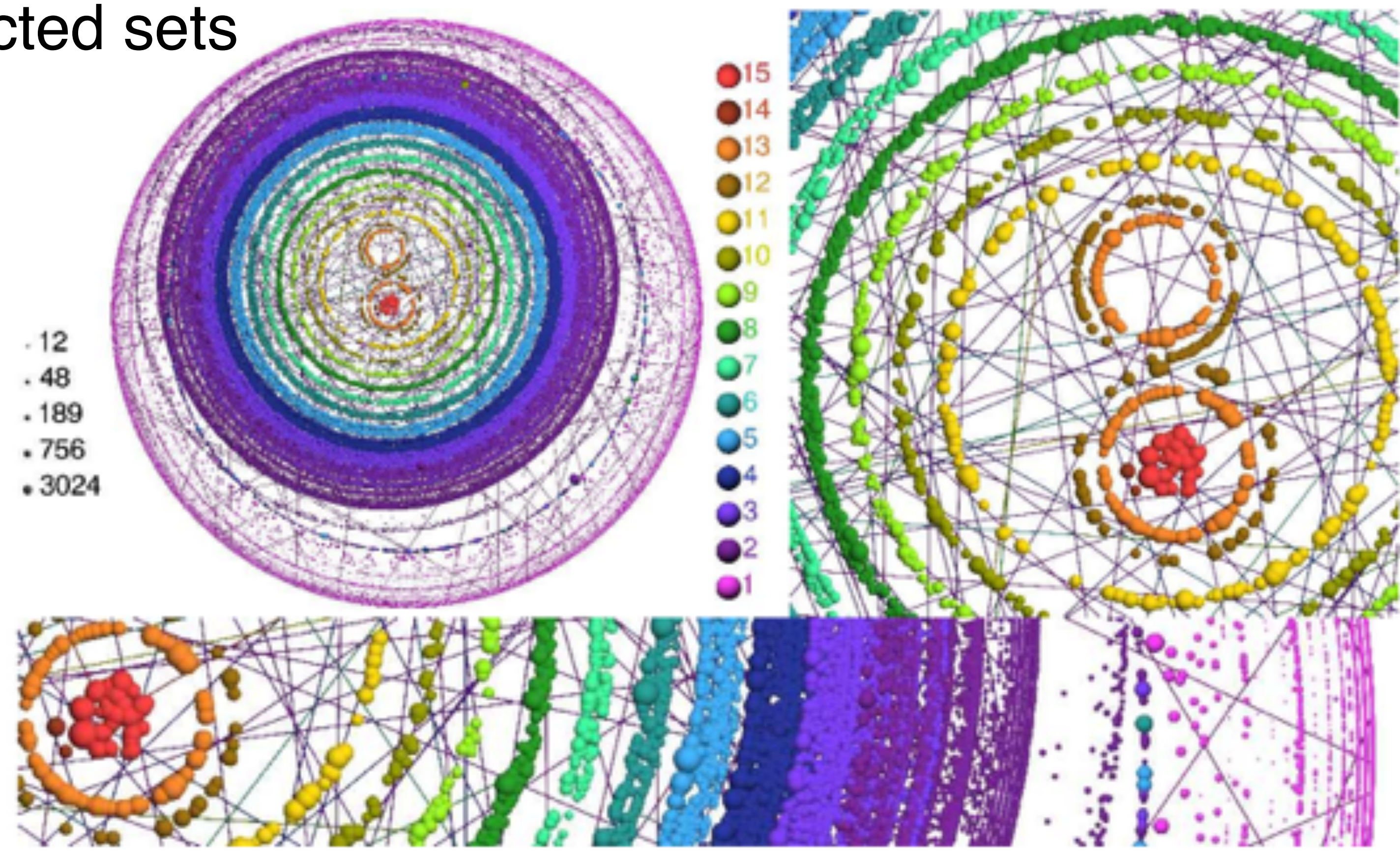


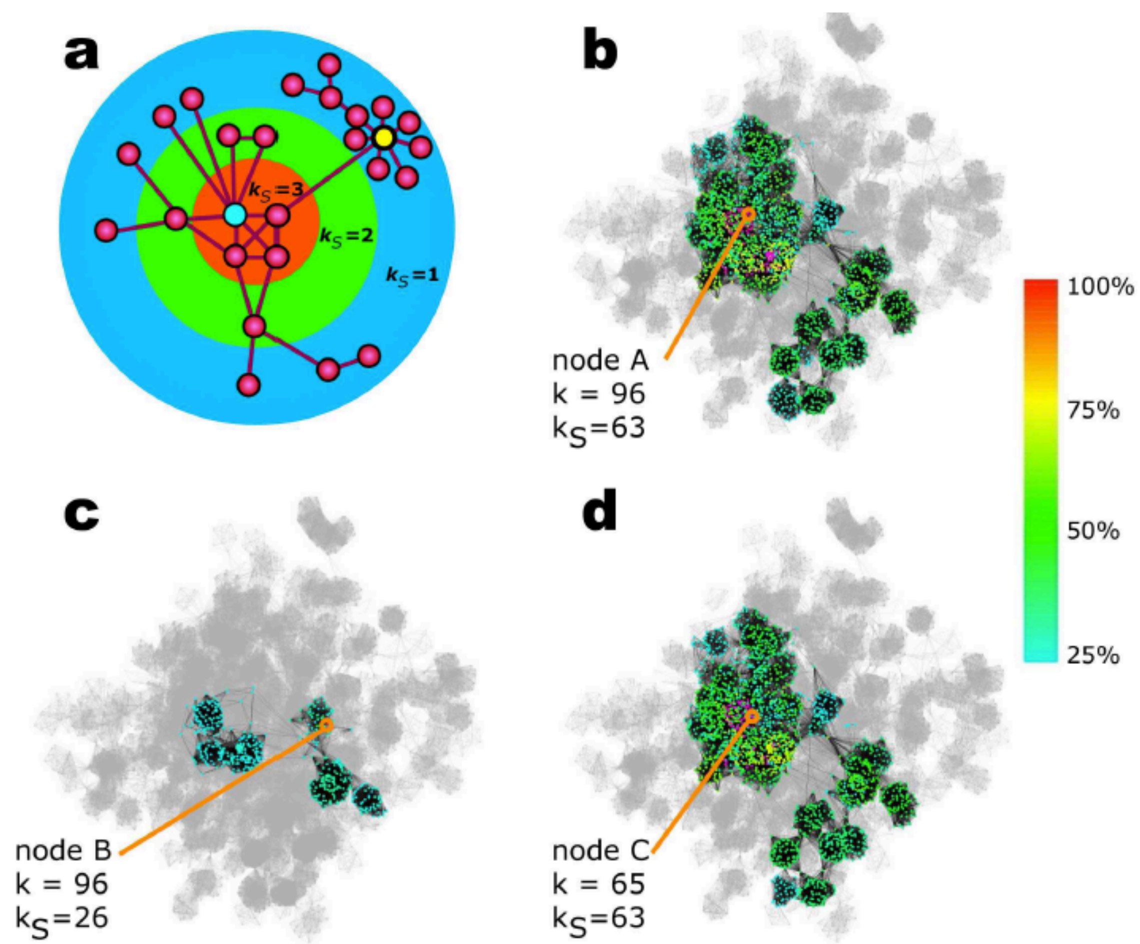
Figure from BECS Complex Networks Course
J Saramäki 2013

k-core decomposition

- Centrality measure: the k-shell index
- High k-shell index assigns central position of the node in the network within densely connected sets of nodes
- They play important role in spreading and contagion processes

Alvarez-Hamelin et.al. (2006)





Kitsak *et al.* (2010)

- outcome of a spreading process when initiated on nodes with different degrees and k-shell indices