Network Biology

Dr. Martina Summer-Kutmon
Assistant professor
BIGCAT / MACSBIO

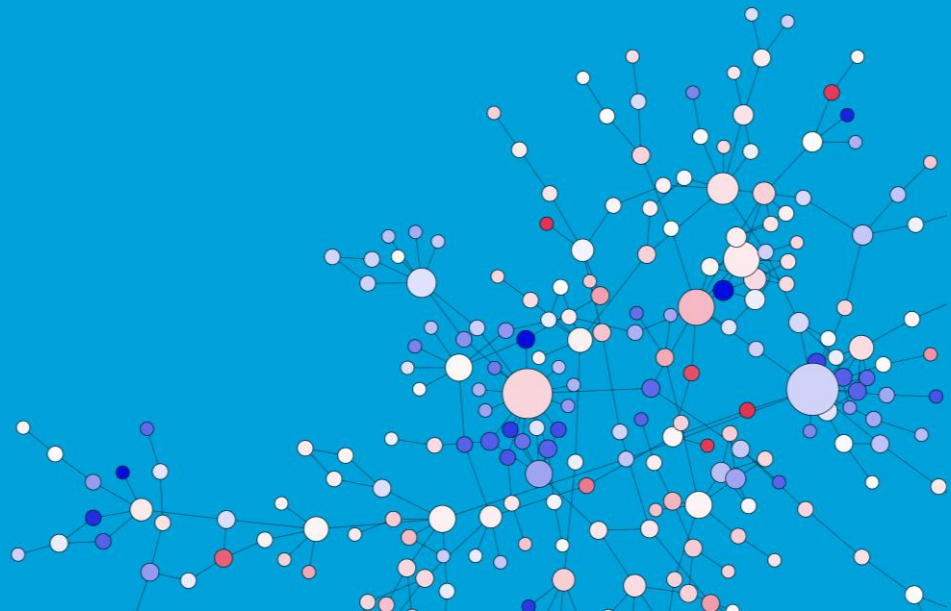
13 January 2021

HELIS course (online)

Maastricht University

Helis Academy

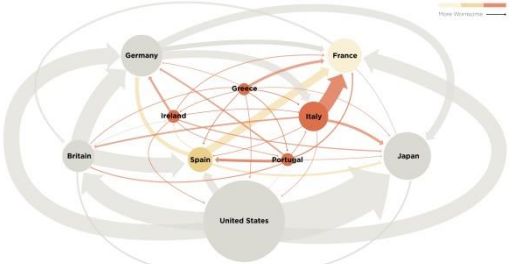
Interreg
Vlaanderen-Nederland
Europees Fonds voor Regionale Ontwikkeling



Networks are everywhere!

Images:
https://cdn.cnn.com/cnnnext/dam/assets/150205174625-internet-everywhere-large-169.jpg

Networks are everywhere!



Introduction

The basics of network science/ graph theory

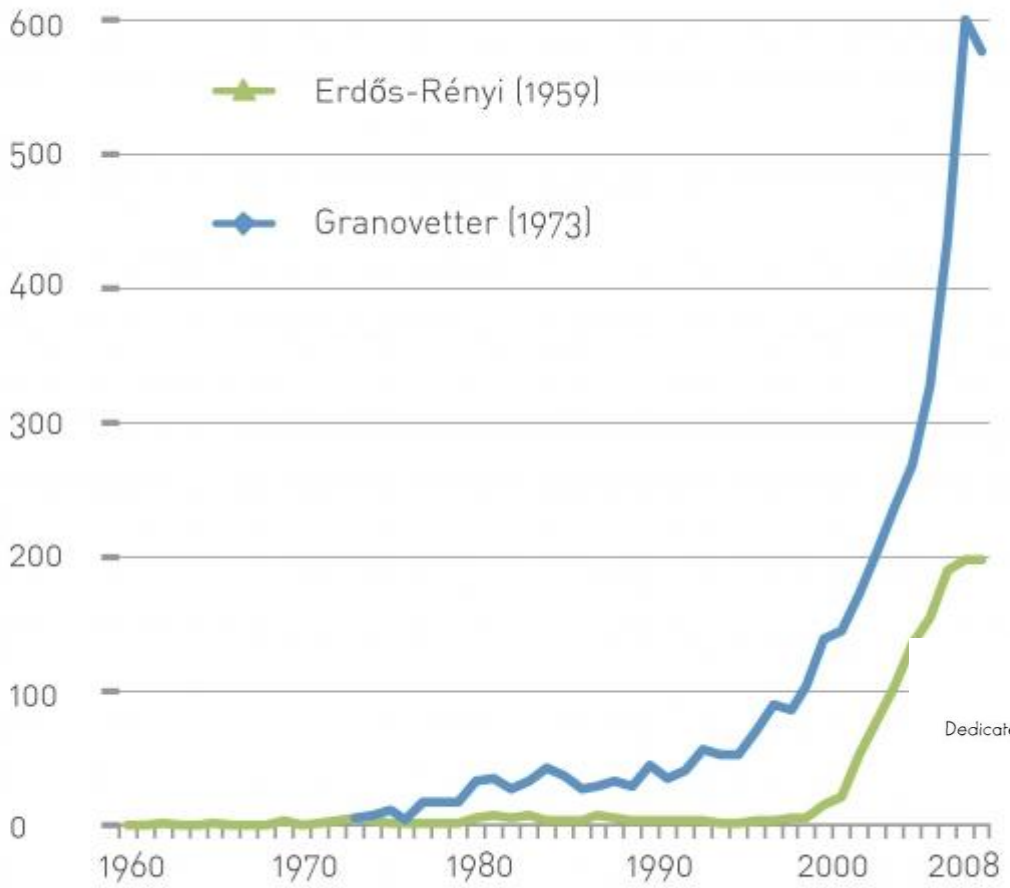
Images:
https://cdn.pixabay.com/photo/2016/12/16/15/47/network-1911678_960_720.png

How it all began…

- Leonard Euler in 1735
- Swiss mathematician
- Königsberg (now Kaliningrad, Russia)

Source: http://networksciencebook.com/images/ch-02/video-2-1.m4v

Key publications of the field



The Strength of Weak Ties[1]

Mark S. Granovetter
Johns Hopkins University

On random graphs I.

Dedicated to O. Varga, at the occasion of his 50th birthday.

By P. ERDŐS and A. RÉNYI (Budapest).

Why is network science only really thriving since the start of the 21st century?

Source: http://networksciencebook.com/images/ch-01/figure-1-3.jpg

Why is it a science of the 21st century?

- Networks are old…
  - Metabolic networks date back to the origins of life
  - Social networks are as old as humanity
  - Graph theory, a prolific subfield of mathematics, has explored graphs since 1735

- But…
  - Important publications were specialized and highly regarded within their discipline but with limited impact outside their field.

- Emergence of a new discipline – "network science" – interdisciplinary!

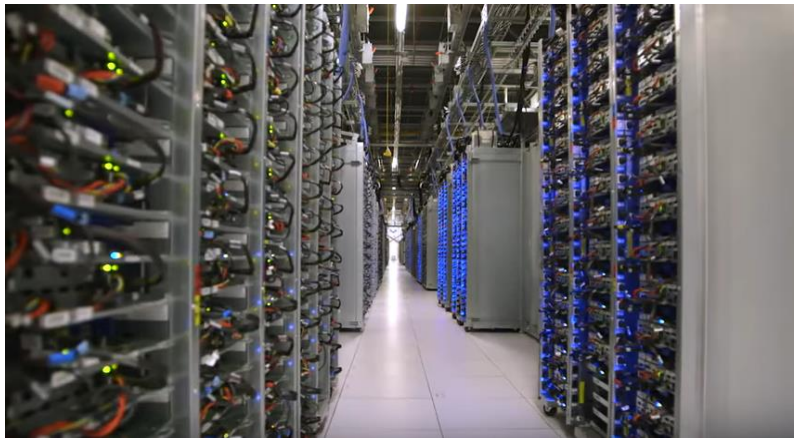
Network science

- Behind each complex system there is a network that defines the interactions between the components.
- We will never understand complex systems unless we map out and understand the network behind them.

Albert-Laszlo Barabasi, Network Science Book

Network science of the 21st century

- **Technological** advances



Inside a Google data center
https://www.youtube.com/watch?v=XZmGGAbHqa0

Microsoft underwater data center
https://www.cnbc.com

Facebook data center
http://www.datacenterknowledge.com/data-center-faqs/facebook-data-center-faq-page-4

Network science of the 21st century

- **Technological** advances
  - Internet revolution → effective and fast data sharing, cheap digital storage

- Availability of **tools to map and store** (large) networks
  - Keeping track of huge amount of data
  - Accessibility of data

Network science of the 21st century

- Key discovery:
  - Universality of network characteristics
  - Networks from different domains are very different

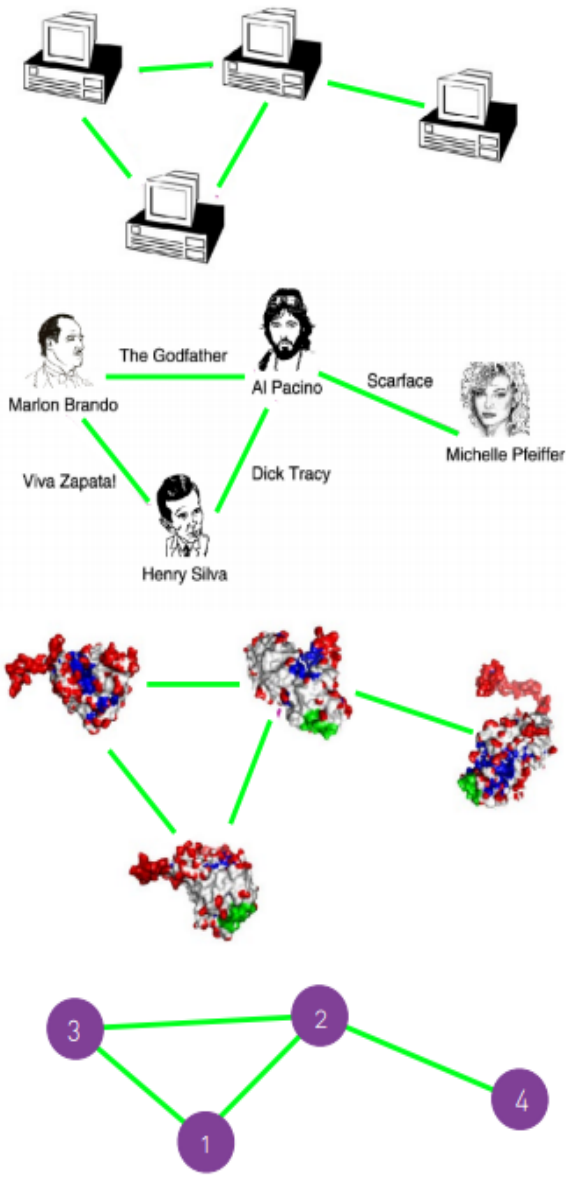
But…

* Architecture of the networks is very similar
* Same organizational principles
* Common set of mathematical tools to explore them

Network science of the 21st century

- Network as a universal concept
- Objects and relationships between them

- Same structure
- Same tools and methods to analyze networks in different disciplines



Source: http://networksciencebook.com/images/ch-02/figure-2-2.jpg

Network representation

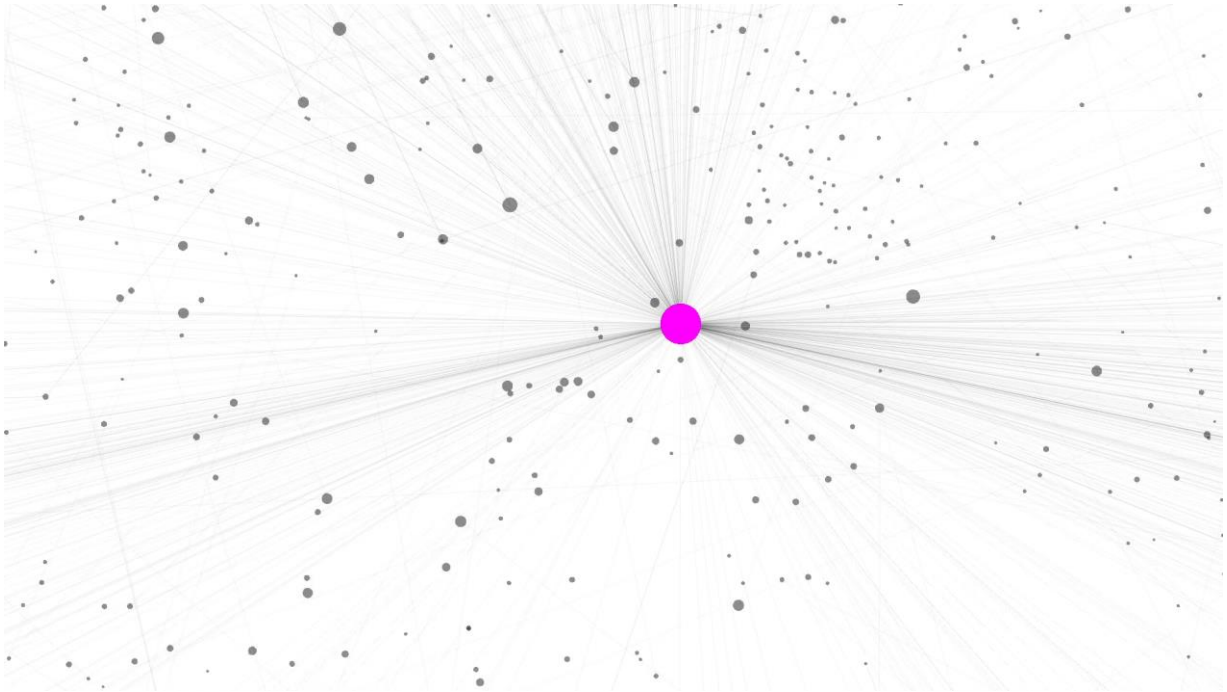
- Choice of **proper network representation** important for usability
- How we assign the links between nodes determines the **type of questions** we can study

  - Let's connect individuals based on their first name (all Peters connected to each other)
  - We will get a network and can do all the analyses that we will discuss
  - But what would you be exploring? What kind of questions can you answer?

Examples of network analysis in other fields?

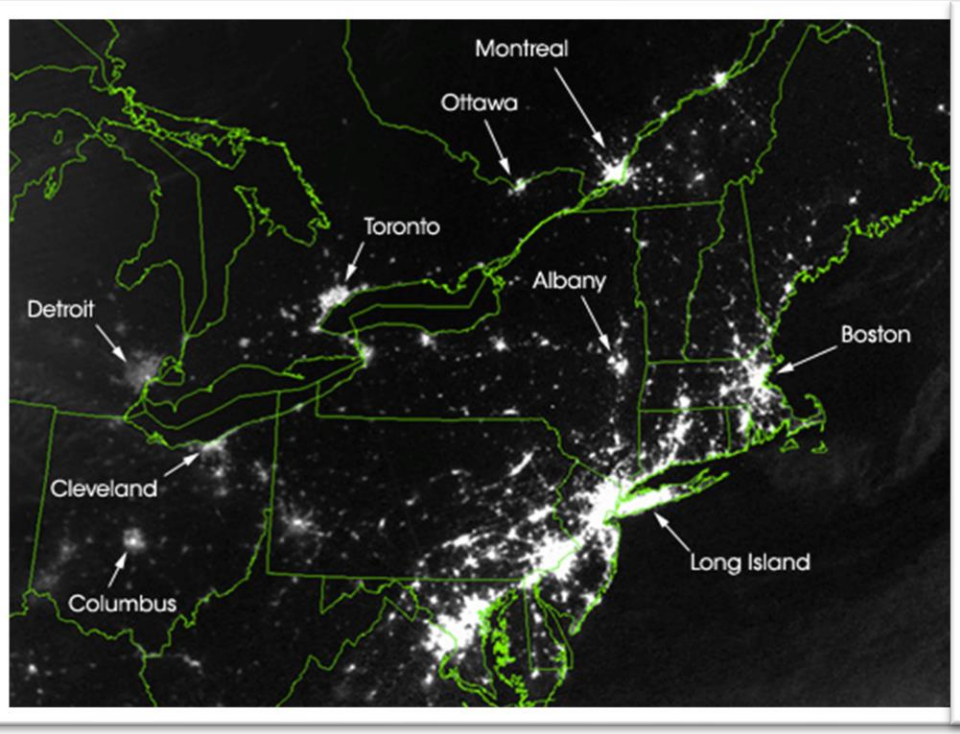
Internet

- 192,244 nodes (routers)
- 609,066 edges (connections)

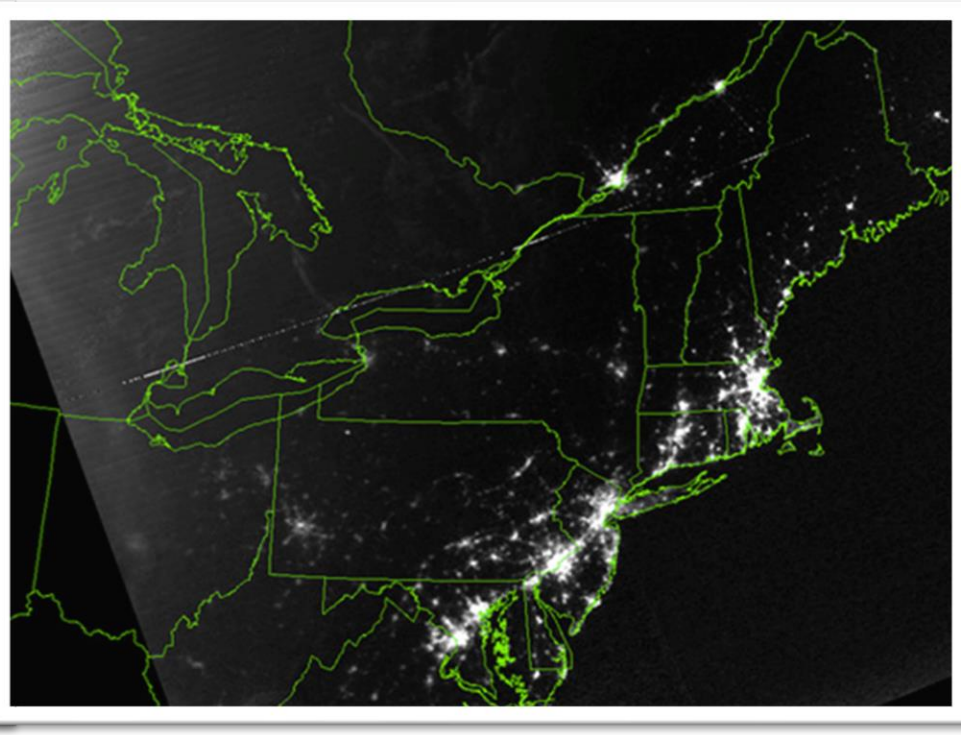


Source: http://networksciencebook.com/images/ch-04/video-4-1.mov, https://www.vaultnetworks.com/wp-content/uploads/2015/03/internet-03.jpg

Power grid



August 14, 2003: 9:29pm EDT
20 hours before



August 15, 2003: 9:14pm EDT
7 hours after

Source: http://networksciencebook.com/images/ch-01/figure-1-1.jpg

Power grid

- We must understand how network structure affects the robustness of a complex system
- System failures follow reproducible laws that can be quantified and even predicted using the tools of network science

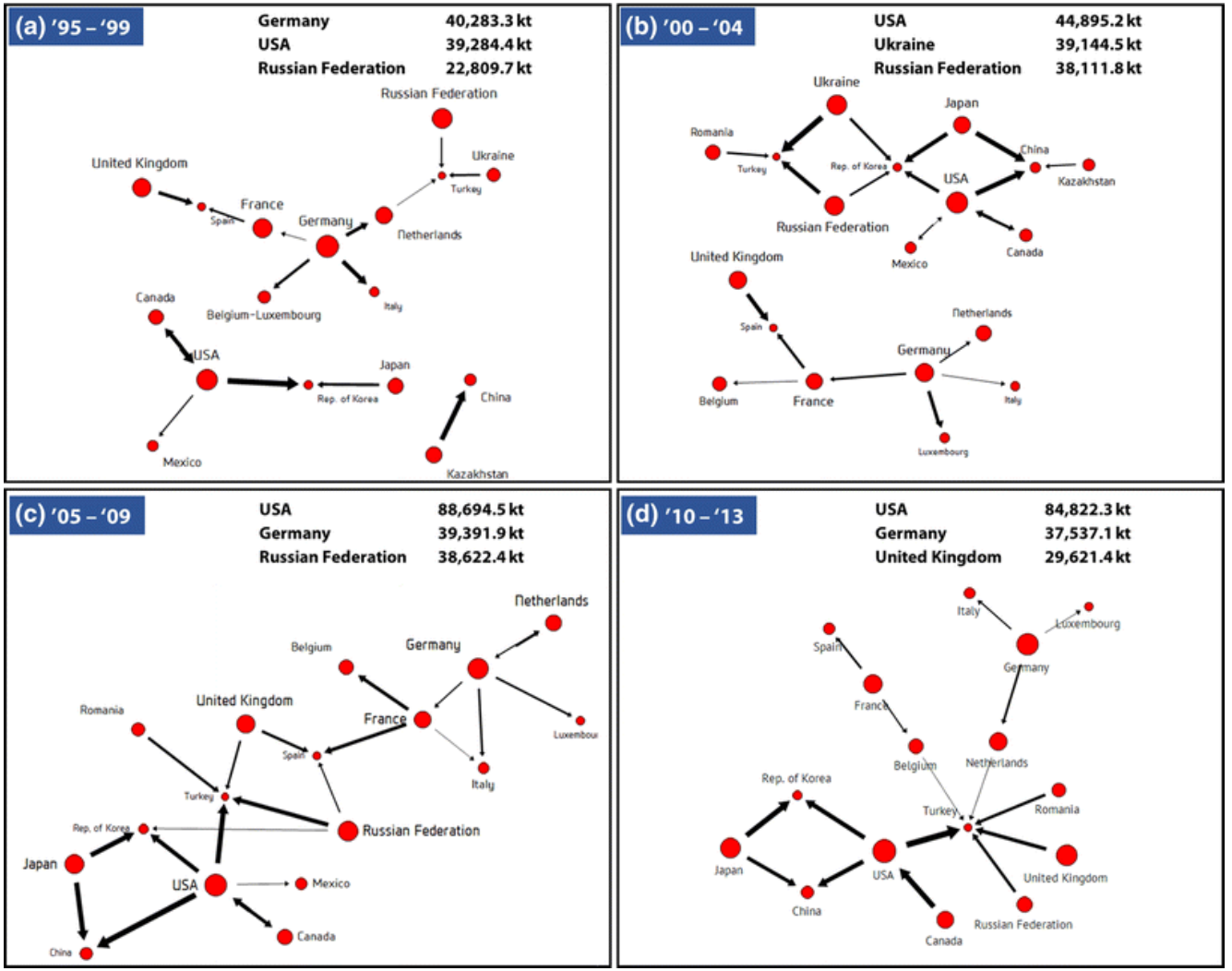
Transportation

- Dutch railway

Source: http://www.projectmapping.co.uk/Europe%20World/Resources/NS-spoorkaart-2018.pdf

Economics

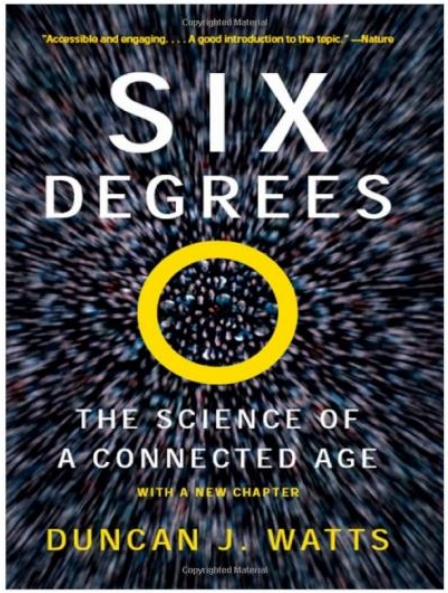
- Trading networks
- Steel scrap

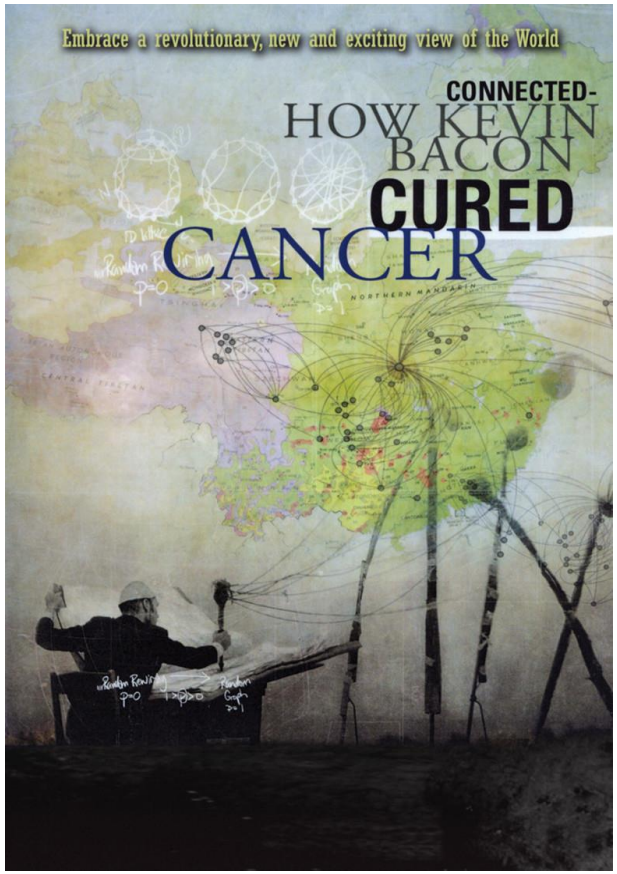


Source: https://link.springer.com/article/10.1007/s40831-015-0007-7

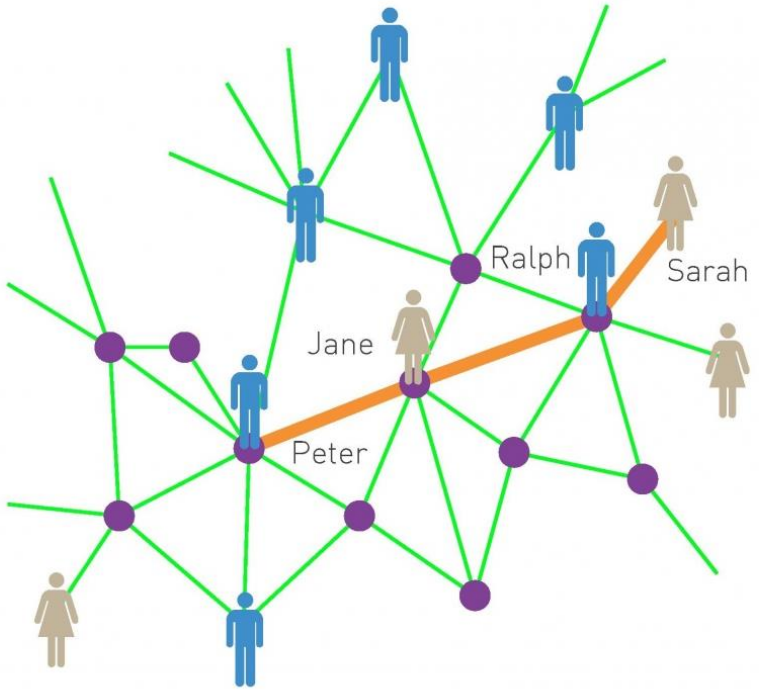
Social networks

- Six degrees of separation
- Small world phenomenon



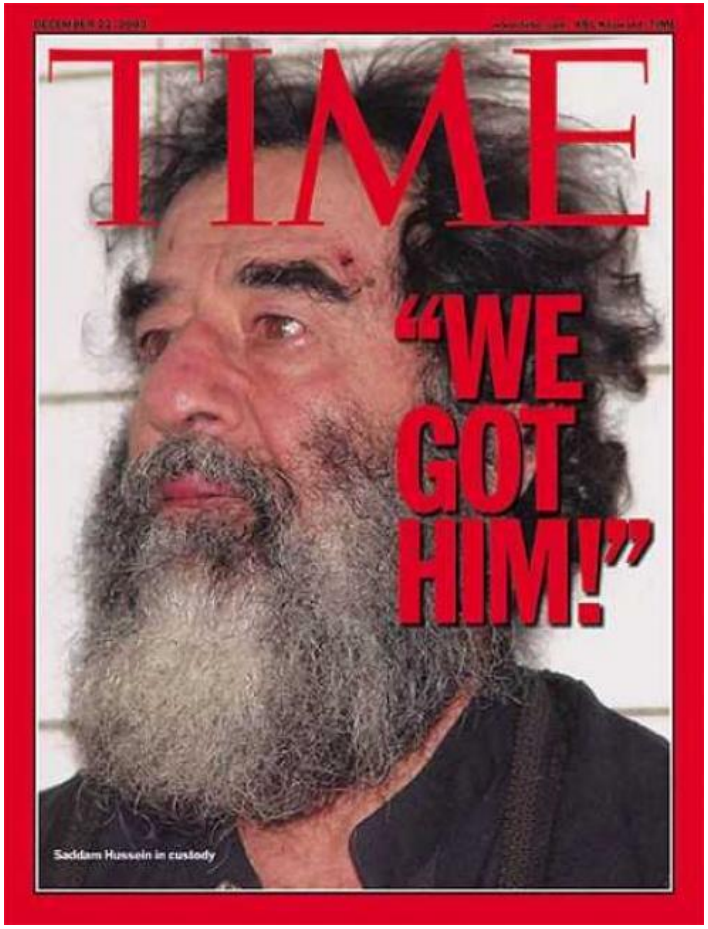


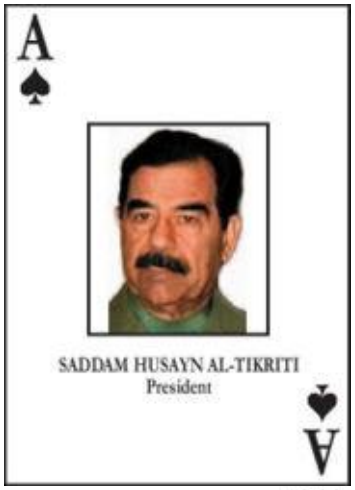
Movie: Connected: The Power of Six Degrees (2008)



Source: http://networksciencebook.com/images/ch-03/figure-3-10.jpg, http://networksciencebook.com/images/ch-01/figure-1-11.jpg

From Hussein to network theory

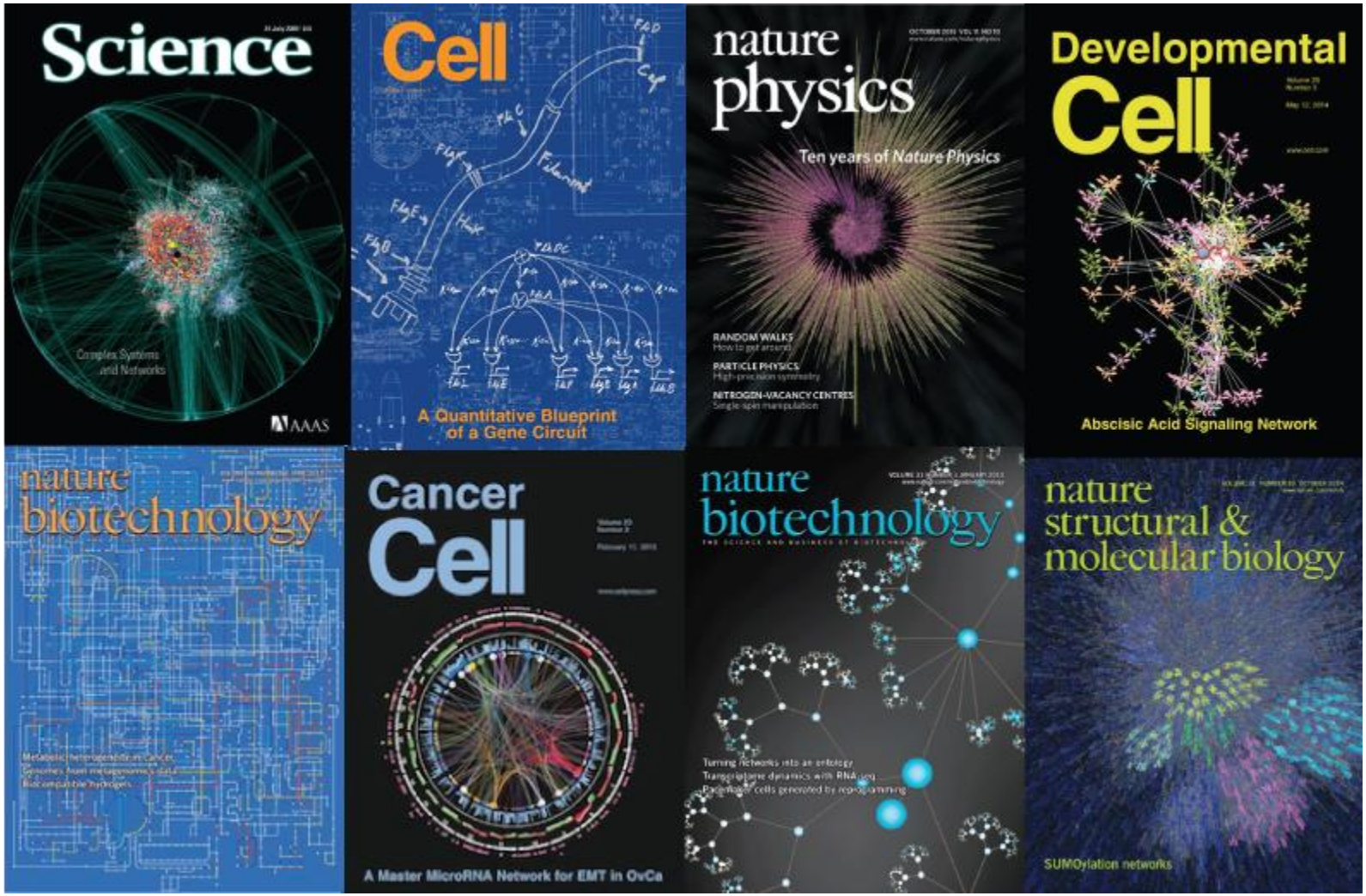




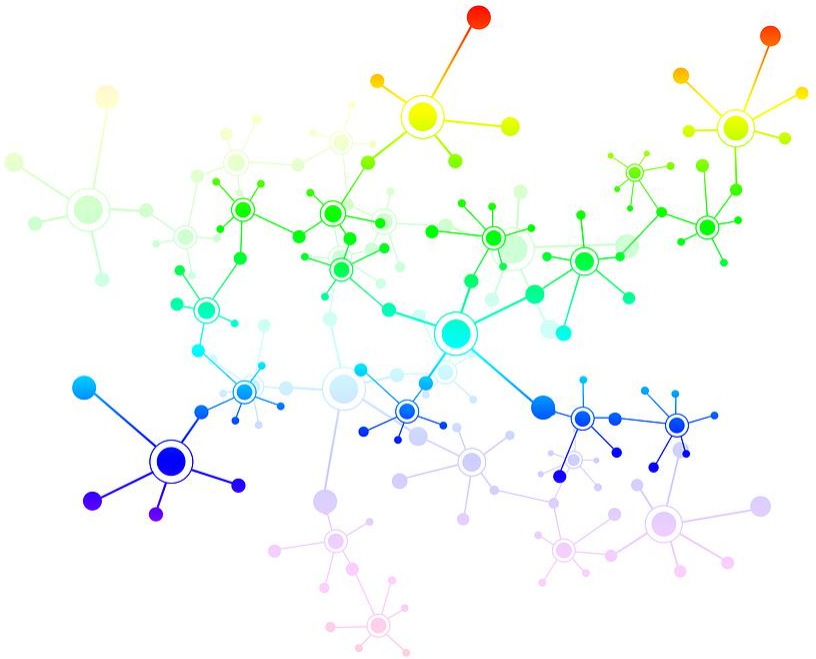
From Hussein to network theory

- The capture of Saddam Hussein:
  - Shows the predictive power of networks
  - Underlies the need to obtain accurate network maps (and how difficult mapping processes often are)
  - Demonstrates remarkable stability of networks (not fresh intelligence but photos from an old family album)
  - Choice of network we focus on make a huge difference (hierarchical official organization of Iraqi government was of no use)

Networks in science



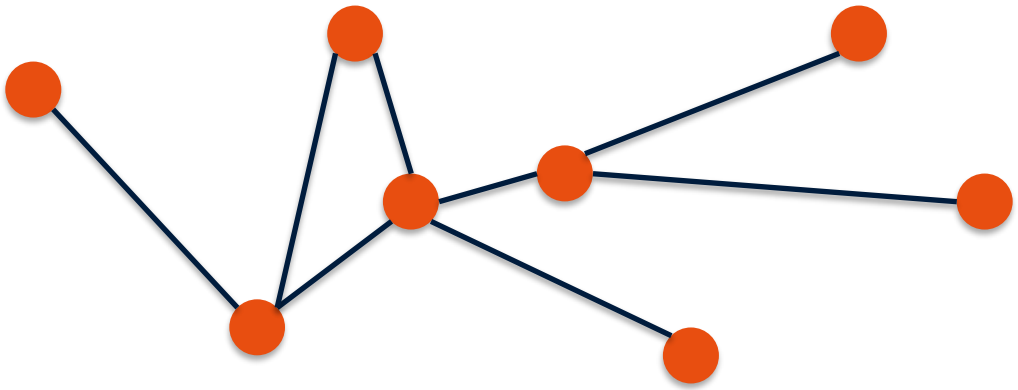
Let’s get into the basics of networks



Images:
https://cdn.pixabay.com/photo/2016/12/16/15/47/network-1911678_960_720.png

Graph vs. network

- In the scientific literature the terms network and graph are used interchangeably



components:	nodes, vertices	N
interactions:	links, edges	L
system:	network, graph	(N,L)

Graph vs. network

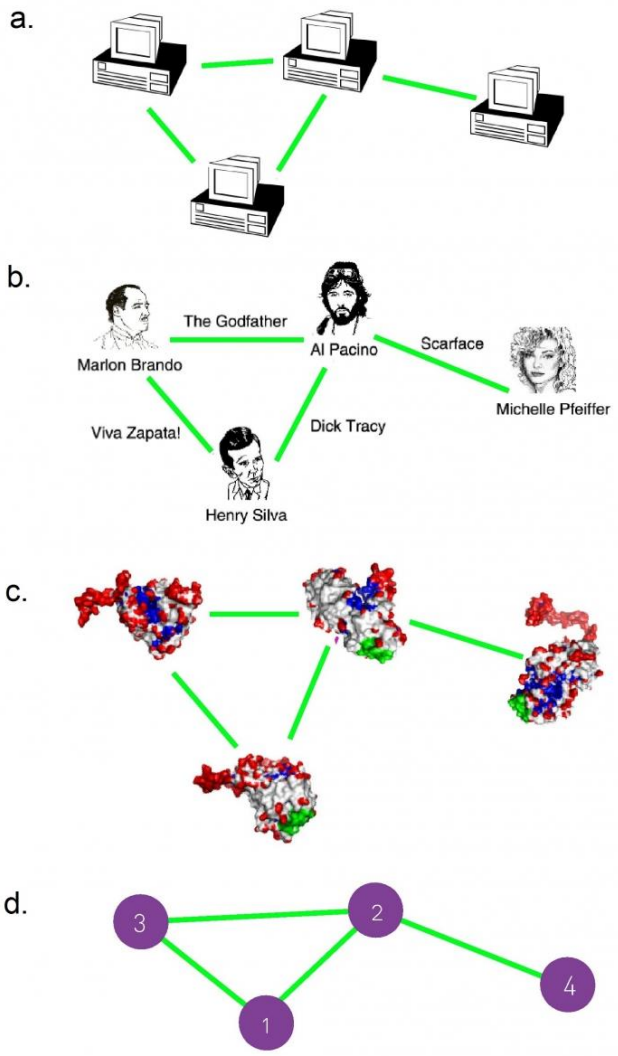
- **Network** often refer to real systems
- **Graphs** are the mathematical representation

- In most cases, we will use the two terms interchangeably.

Graph vs. network

- N: number of nodes
- L: number of links (edges)

- N = 4, L = 4
- Same graph / different network

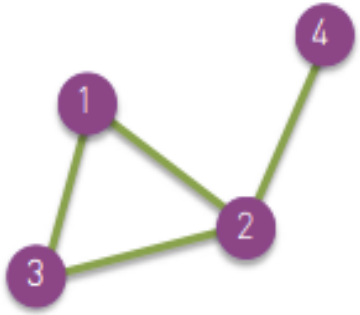


Source: http://networksciencebook.com/images/ch-02/figure-2-2.jpg

Network properties

Undirected

Links: undirected (symmetrical)

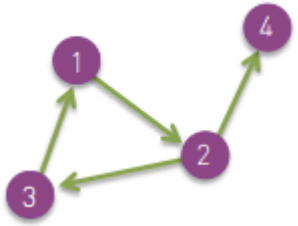


Examples:

Co-authorship, actor network, protein interactions

Directed

Links: directed (arcs)



Examples:

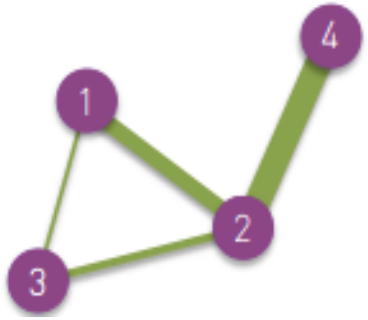
URLs (internet), phone calls, metabolic reactions

Source: http://networksciencebook.com/images/ch-02/figure-2-5.jpg

Network properties

Weighted

Links: weight, strength, flow



Examples:

Route planning, correlation, metabolic flux

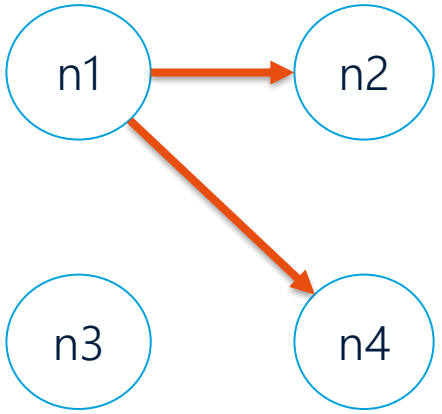
Adjacency matrix

Mathematical representation

$A_{ij} = 1$ there is an edge between node i and j

$A_{ij} = 0$ there is no edge between node i and j

$$A_{ij} = \begin{array}{c} \\ n1 \\ n2 \\ n3 \\ n4 \end{array} \begin{array}{c} \begin{array}{cccc} n1 & n2 & n3 & n4 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix} \end{array}$$



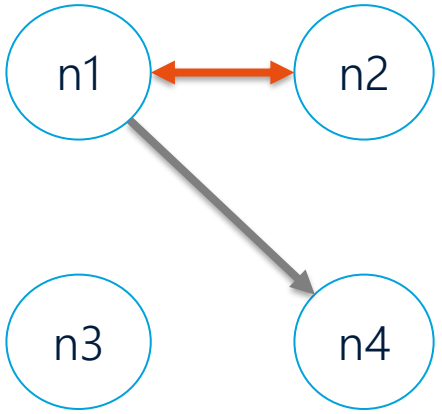
Adjacency matrix

Mathematical representation

$A_{ij} = 1$ there is an edge between node i and j

$A_{ij} = 0$ there is no edge between node i and j

$$A_{ij} = \begin{array}{c} \\ n1 \\ n2 \\ n3 \\ n4 \end{array} \begin{array}{c} \begin{array}{cccc} n1 & n2 & n3 & n4 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix} \end{array}$$



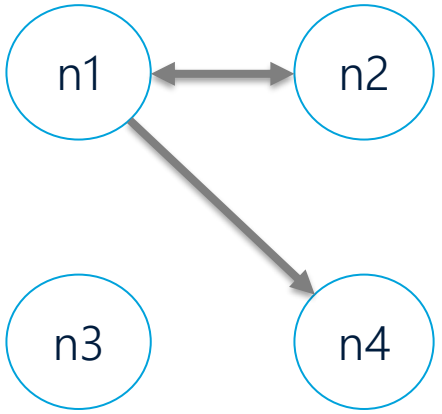
Adjacency matrix

Mathematical representation

$A_{ij} = 1$ there is an edge between node i and j

$A_{ij} = 0$ there is no edge between node i and j

$$A_{ij} = \begin{array}{c} \\ n1 \\ n2 \\ n3 \\ n4 \end{array} \begin{array}{c} \begin{array}{cccc} n1 & n2 & n3 & n4 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix} \end{array}$$



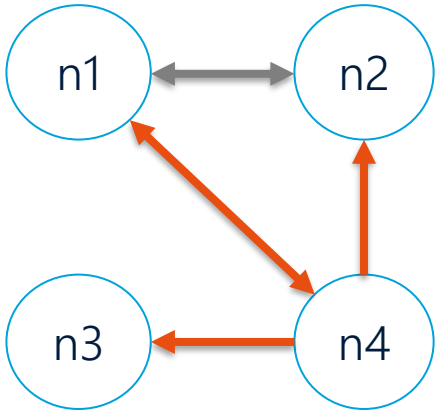
Adjacency matrix

Mathematical representation

$A_{ij} = 1$ there is an edge between node i and j

$A_{ij} = 0$ there is no edge between node i and j

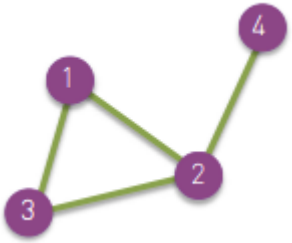
$$A_{ij} = \begin{array}{c} \\ n1 \\ n2 \\ n3 \\ n4 \end{array} \begin{array}{c} \begin{array}{cccc} n1 & n2 & n3 & n4 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix} \end{array}$$



Adjacency matrix

Undirected

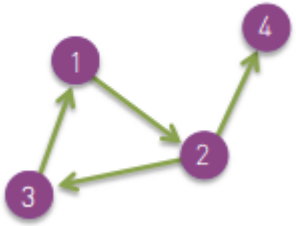
Links: undirected (symmetrical)



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Directed

Links: directed (arcs)

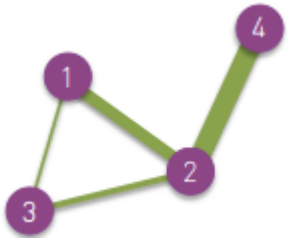


$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Adjacency matrix

Weighted

Links: weight, strength, flow

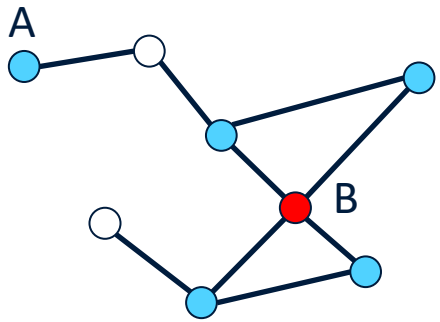


$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

Degree

- k: Degree
  - number of links a node has to other nodes

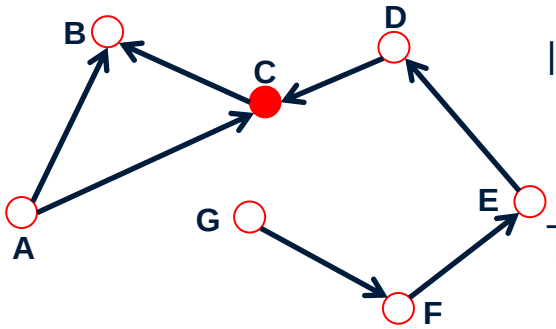
Undirected



$k_A = 1 \qquad k_B = 4$

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i$$

Directed



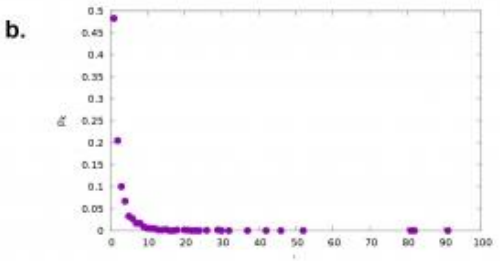
In *directed networks* we can define

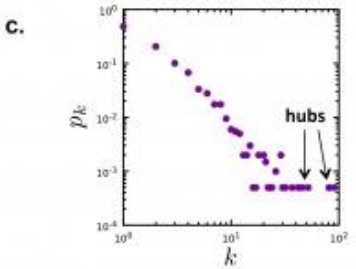
in- and out-degree.

The (total) degree is the sum of in- and out-degree.

$$k_C^{in} = 2 \qquad k_C^{out} = 1 \qquad k_C = 3$$

Degree distribution





PPI of yeast

Source: http://networksciencebook.com/images/ch-02/figure-2-4.jpg

Relevance

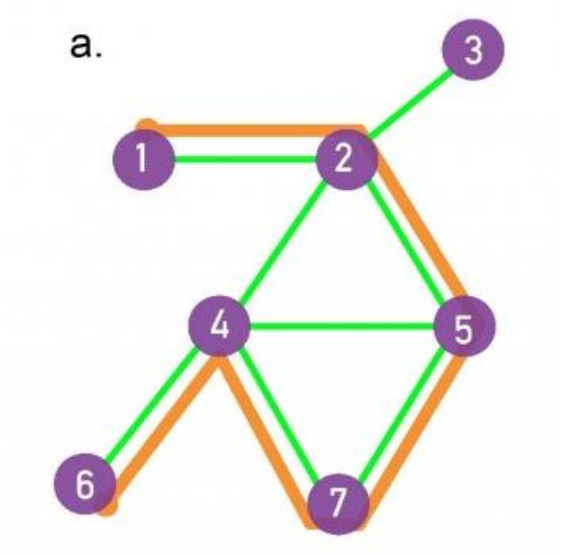
- Hubs (nodes with high degree) tend to be essential nodes in the network

- Degree distribution:
  - Real network have few nodes with a high degree (hubs) and many nodes with a low degree (scale-free topology)
  - Determines many network phenomena like robustness

- Real networks are sparse

Betweenness Centrality

- Load on a node / edge in the network
- How many of the shortest paths pass through the node/edge?



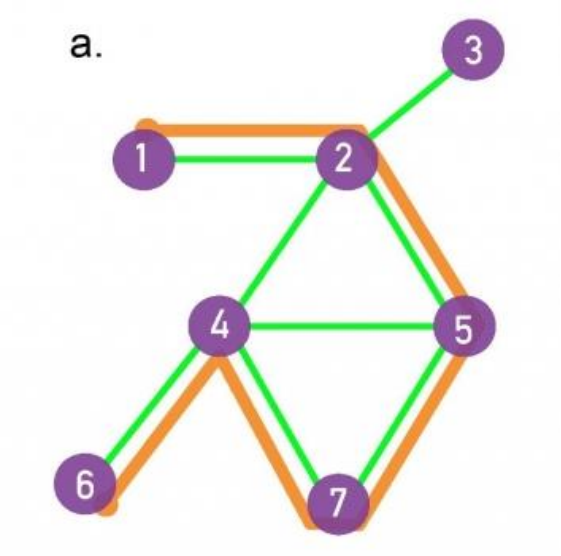
Paths between 1 and 6?

- 1-2-5-7-4-6
- 1-2-4-6
- 1-2-5-4-6

Images:
http://networksciencebook.com/

Betweenness Centrality

- Load on a node / edge in the network
- How many of the shortest paths pass through the node/edge?



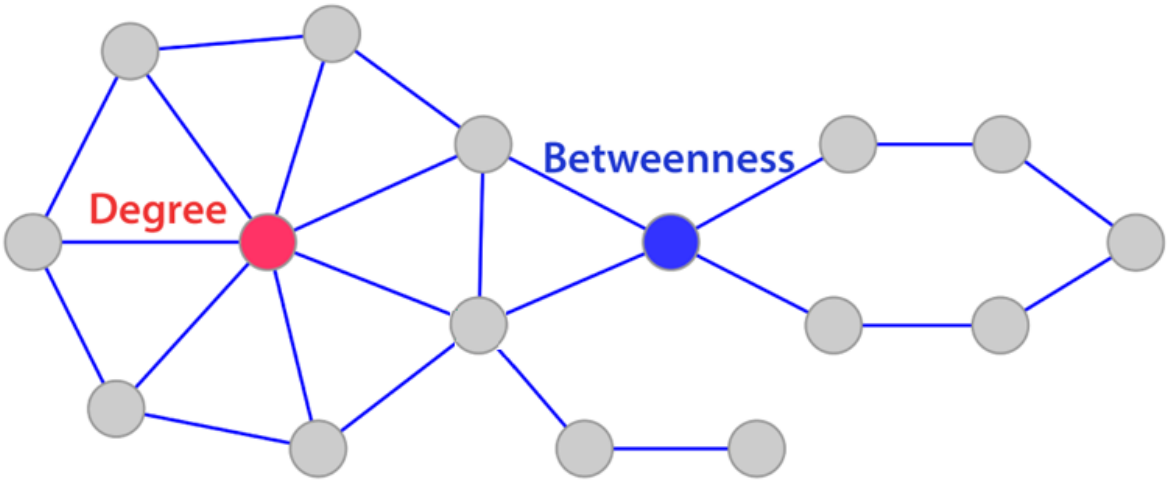
Paths between 1 and 6?

- 1-2-5-7-4-6 d = 5
- 1-2-4-6 d = 3
- 1-2-5-4-6 d = 4

Images:
http://networksciencebook.com/

Relevance

- Nodes/edges with a high betweenness are important for the connectivity of the network
- Nodes/edges connecting two subnetworks
- Bottleneck nodes/edges



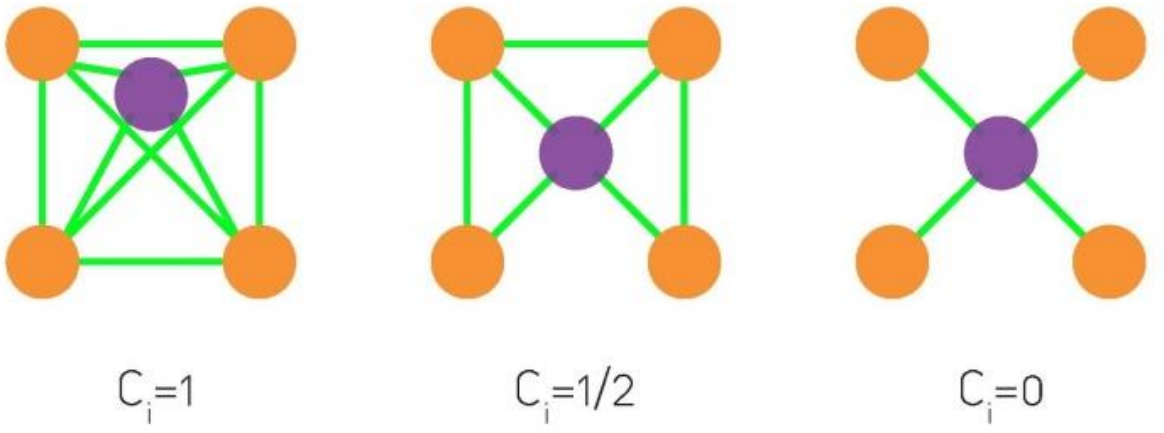
Degree	BetweennessCentrality
7	0.29047619
5	0.42380952
4	0.4952381

Clustering coefficient

- What fraction of a node's neighbors are connected?

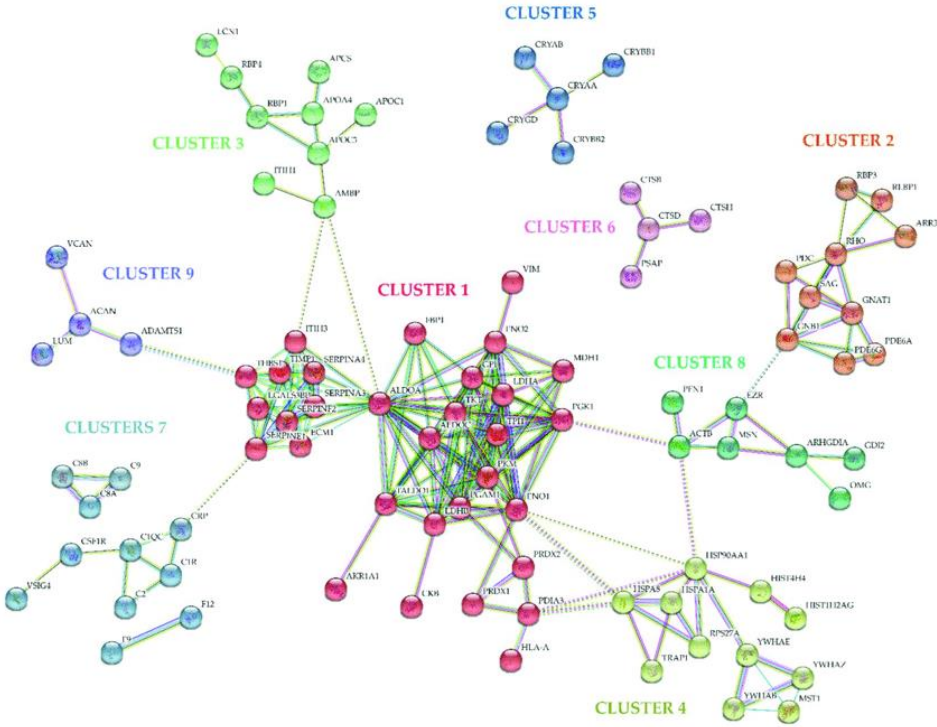
$$C_i = \frac{2L_i}{k_i(k_i-1)}$$

L_i represents the number of links between the k_i neighbors



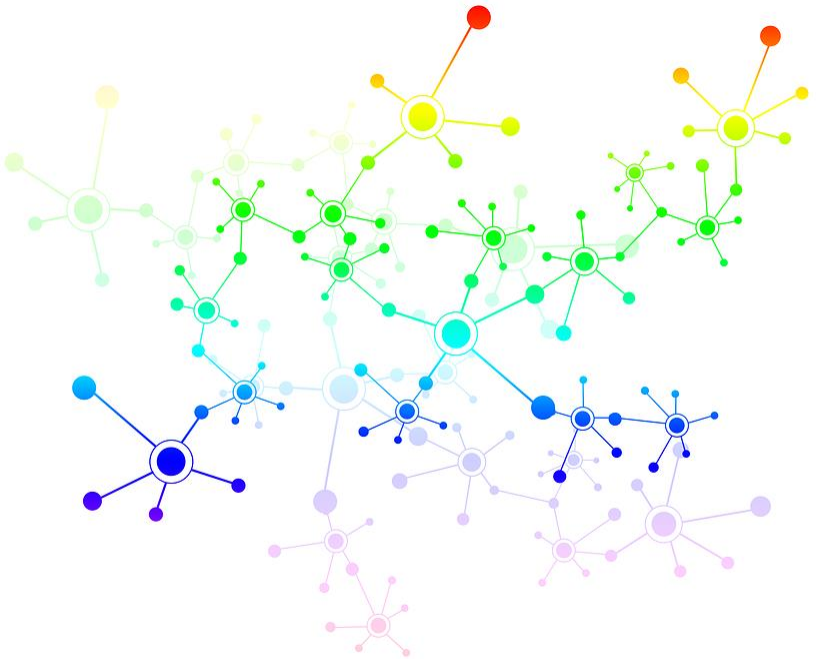
Relevance

- Biological networks have a significantly higher average clustering coefficient compared to random networks
- They are modular by nature
- Protein clusters
- Regulatory patterns



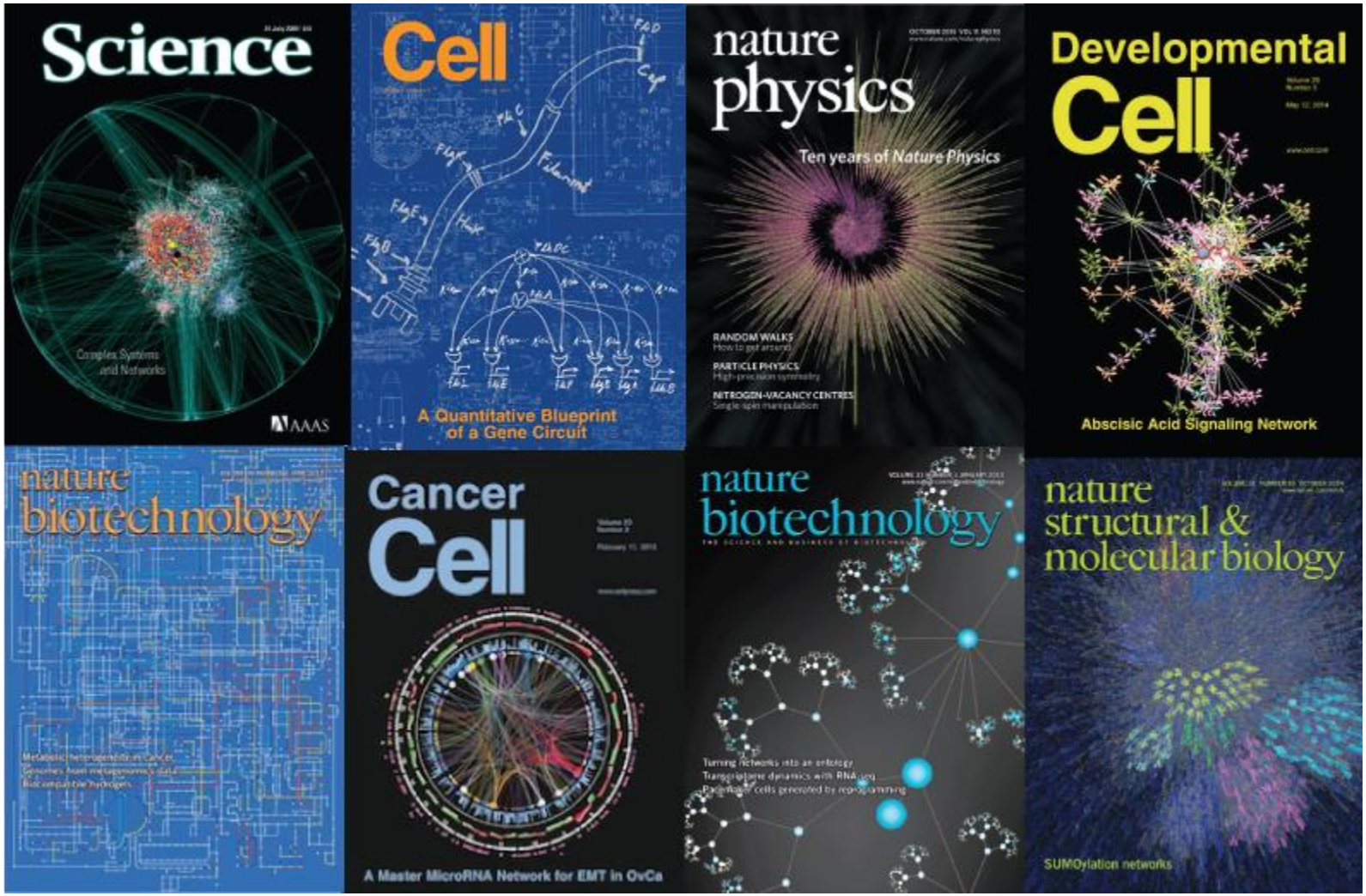
Images:
http://doi.org/10.3390/ijms19041157

Now let's get to Biological Networks



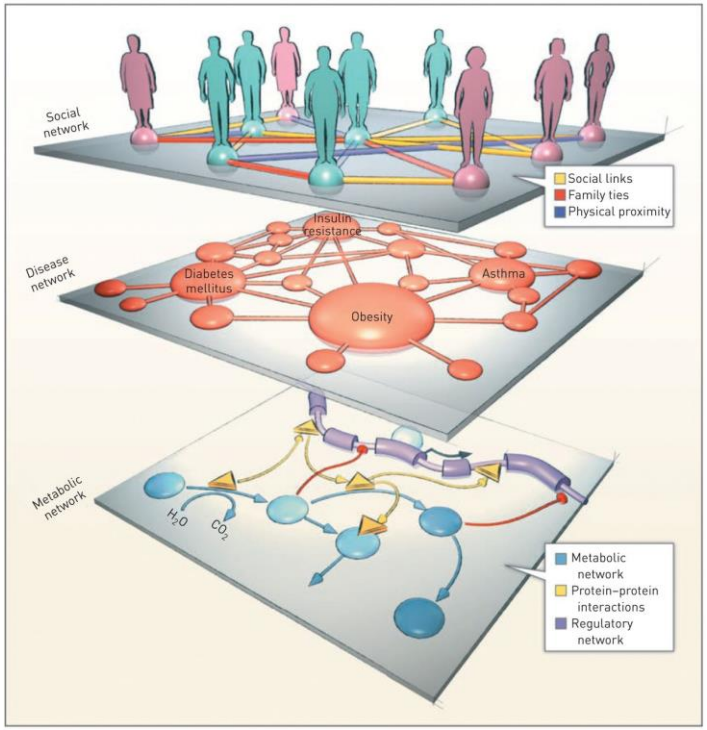
Images:
https://cdn.pixabay.com/photo/2016/12/16/15/47/network-1911678_960_720.png

Networks in Science



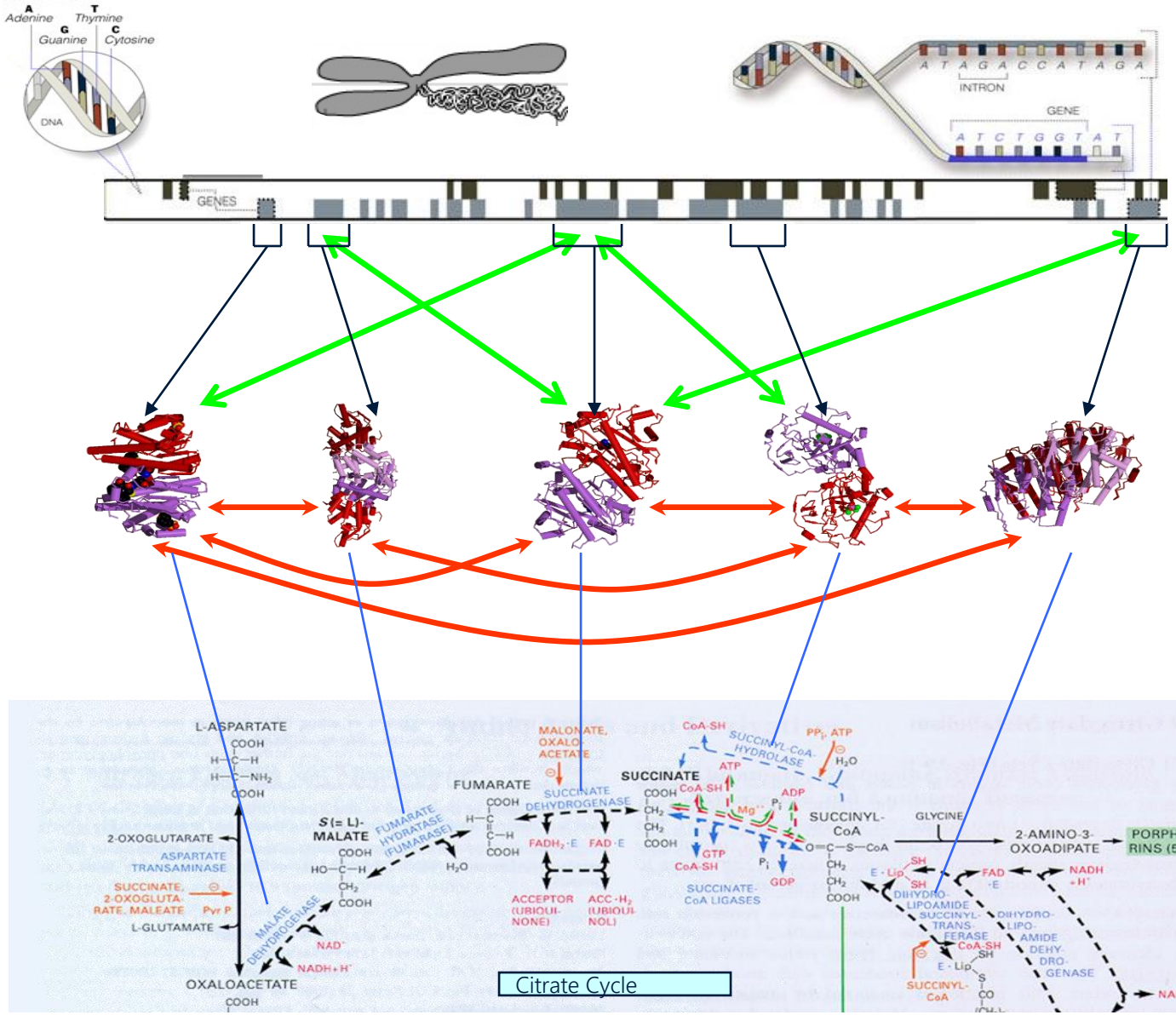
Why networks in biology?

- Understanding biological complexity at all levels
  - Reduce complexity
  - Integrate diverse types of data
  - Find relations & patterns
  - Visualize complex data
  - Identify key pathways and drivers
  - Find functional signatures of interventions
  - Find perturbed paths



Images:
https://erj.ersjournals.com/content/erj/44/3/775/F2.large.jpg

Molecular networks



GENOME

protein-gene interactions

PROTEOME

protein-protein interactions

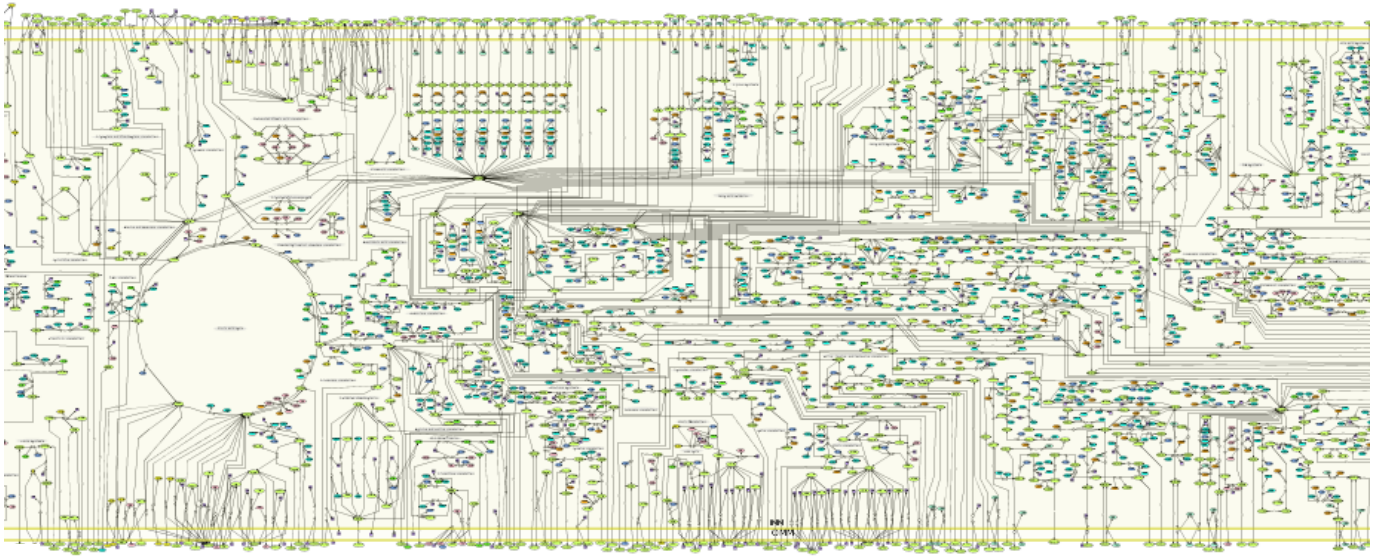
METABOLISM

bio-chemical reactions

Metabolic networks

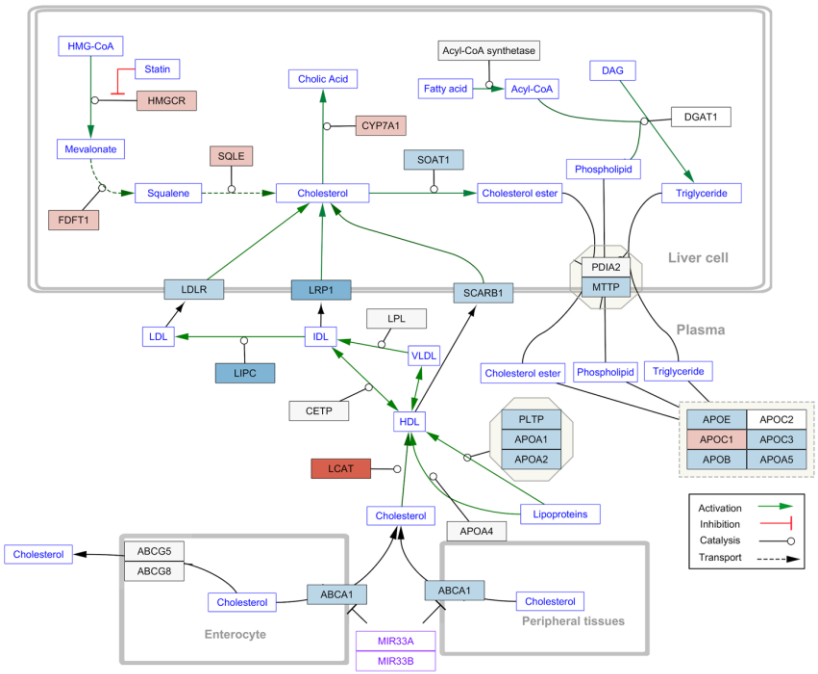
- Collection of biochemical reactions which convert one compound into another
- Reactions are catalyzed by enzymes

- Recon maps → genome-scale metabolic models



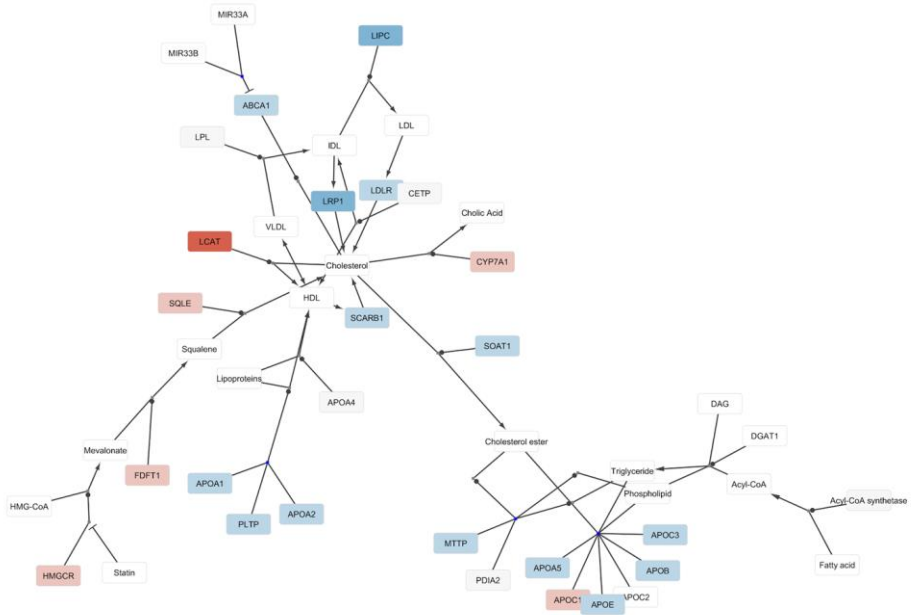
Images:
https://www.vmh.life/#reconmaps/

Pathways as networks



PathVisio

Cytoscape

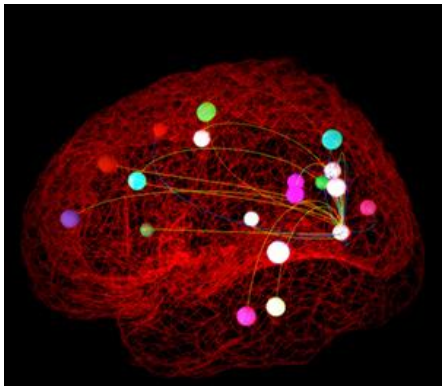


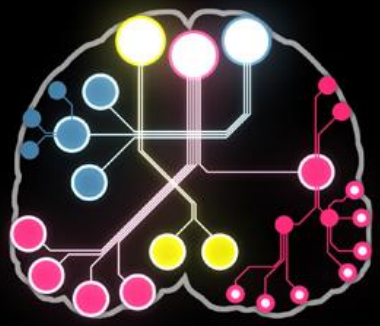
Cytoscape

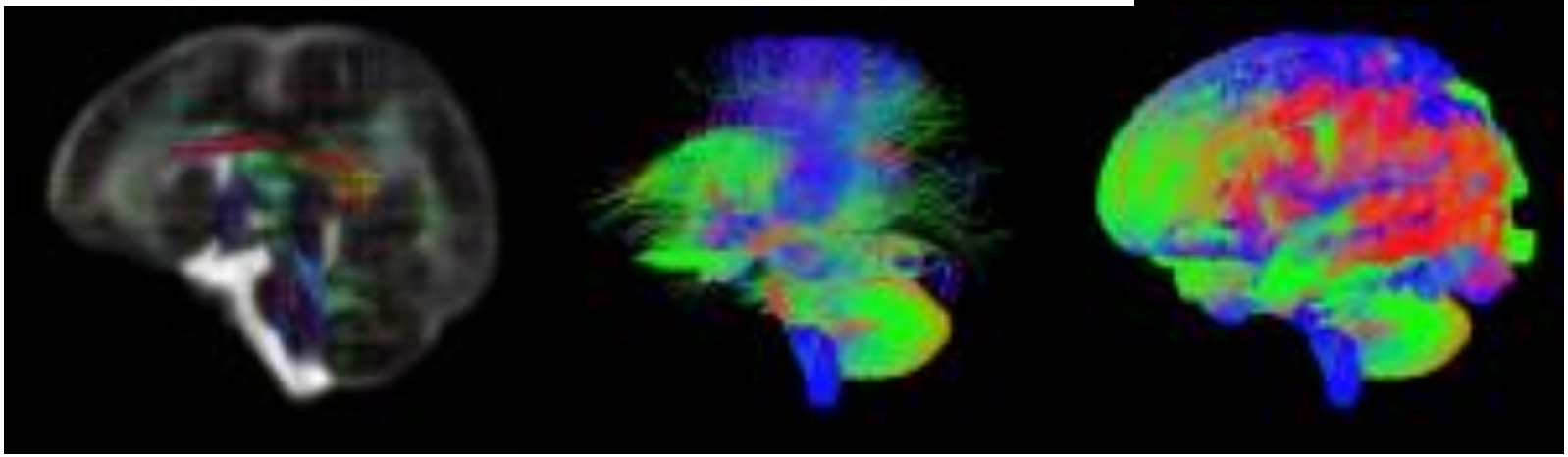
WikiPathways App for Cytoscape: Making biological pathways amenable to network analysis and visualization (2014) https://doi.org/10.12688/f1000research.4254.2

Connectome

- Mapping the brain
- Neuronal connectivity
- Human connectome project

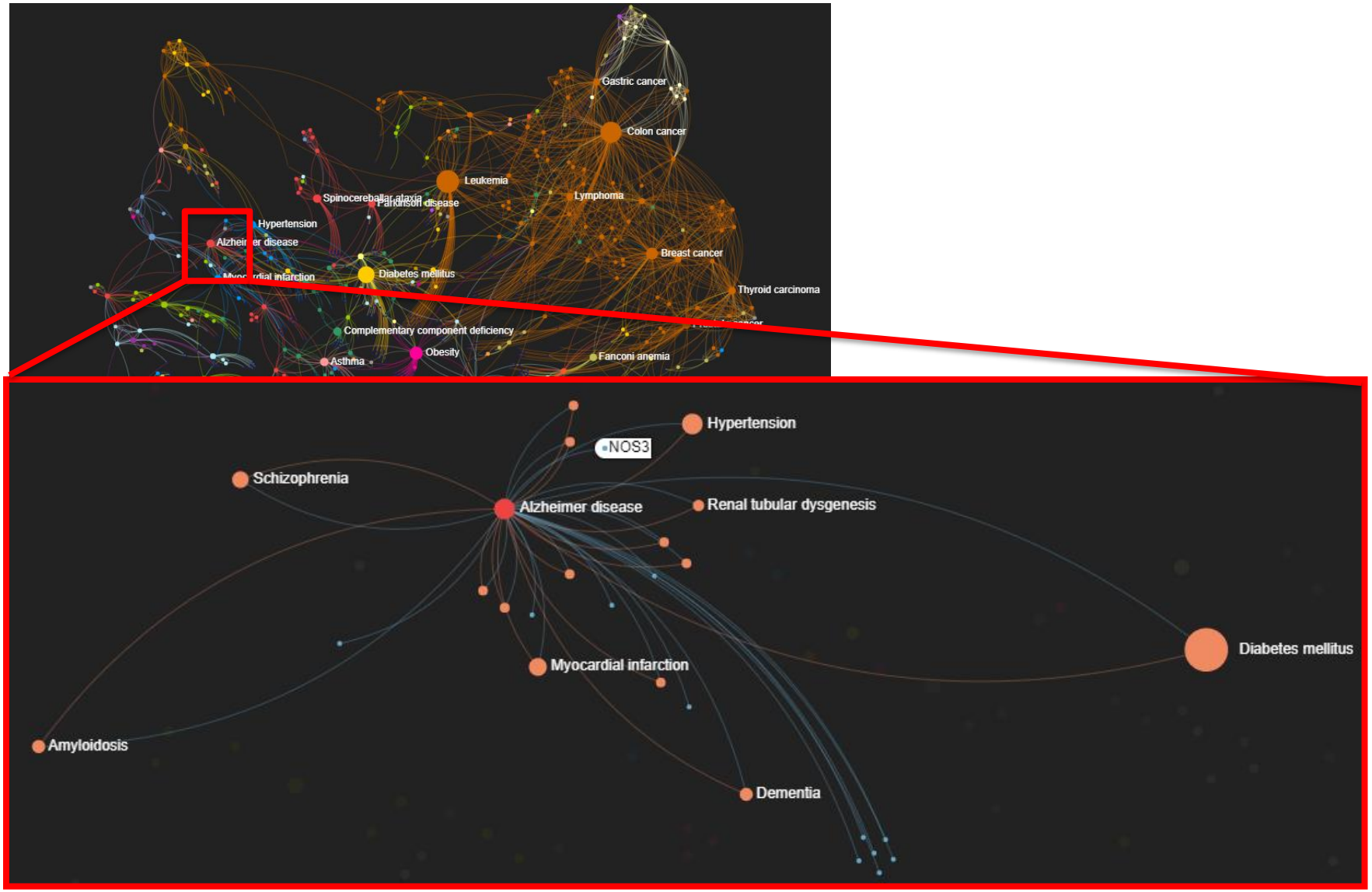




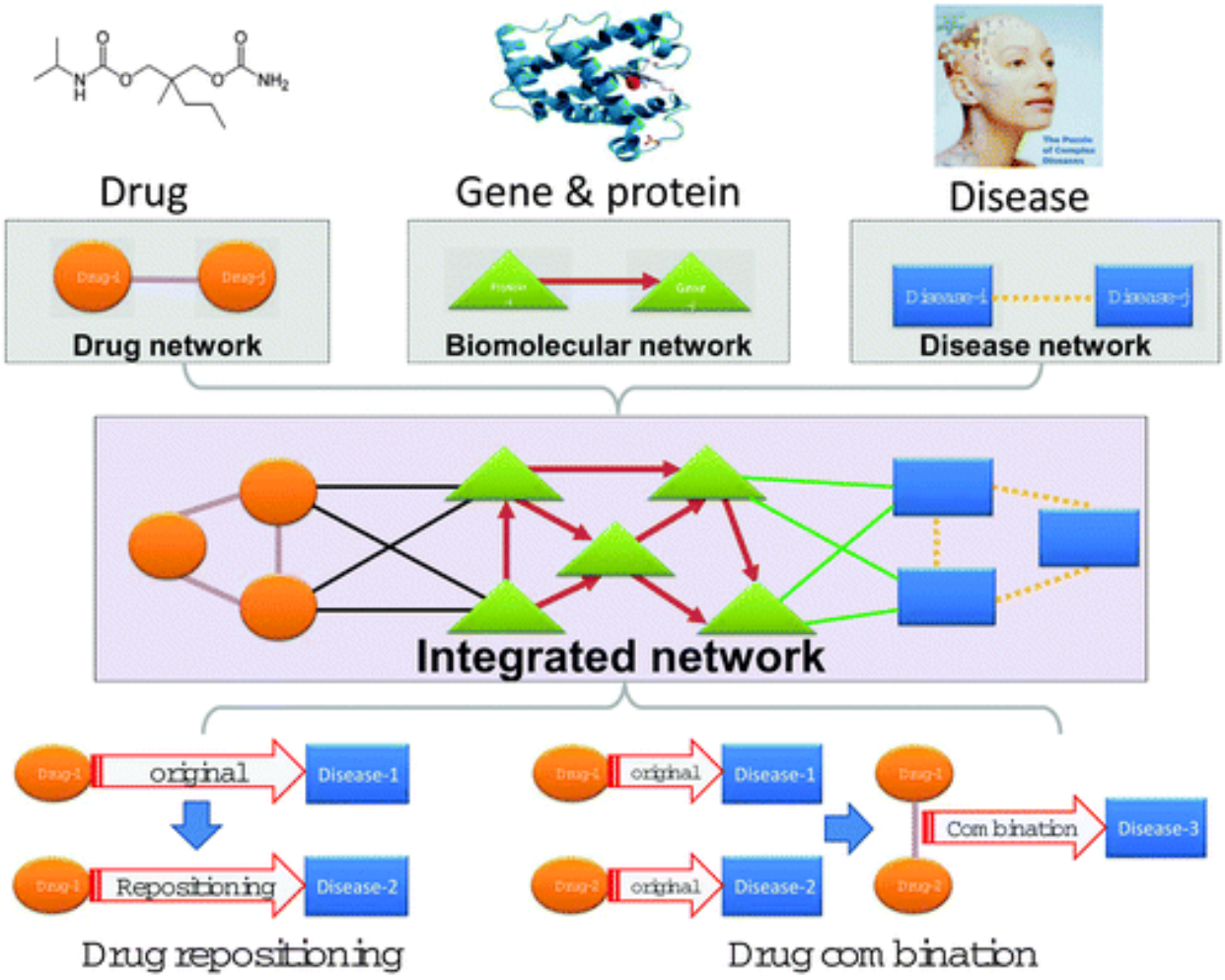


Source: http://www.developingconnectome.org/wp-content/uploads/sites/70/2017/05/DHCP-3-768x220.png, http://www.humanconnectomeproject.org/wp-content/uploads/2011/10/data_Sidebar_Long3.jpg

Human disease network

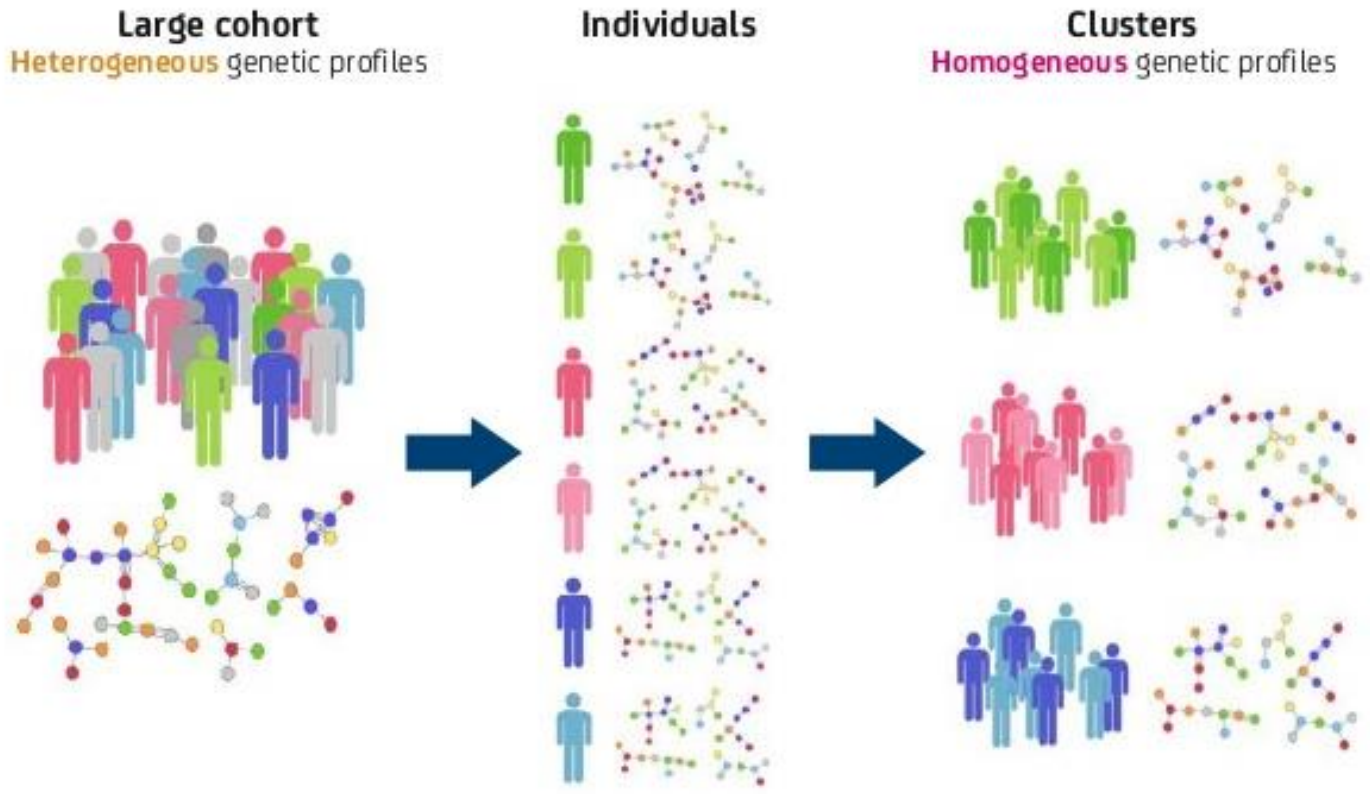


Drug repositioning/repurposing



Images:
https://doi.org/10.1039/C3MB25382A

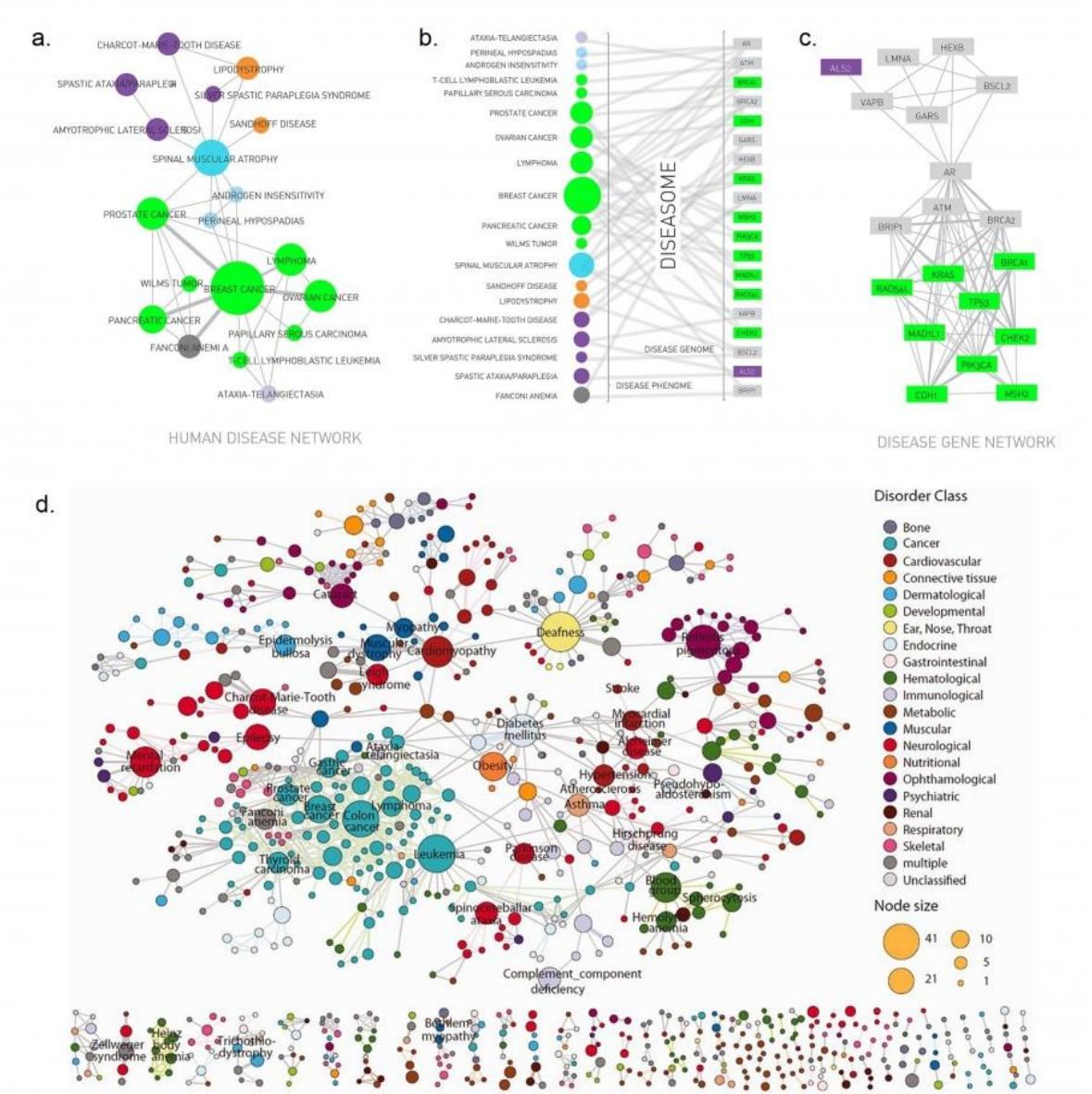
Network-based stratification



Images:
https://www.slideshare.net/gdumas/normative-modeling-patients-stratifications-dealing-with-dimensions-categories-in-modern-psychiatry-83789123

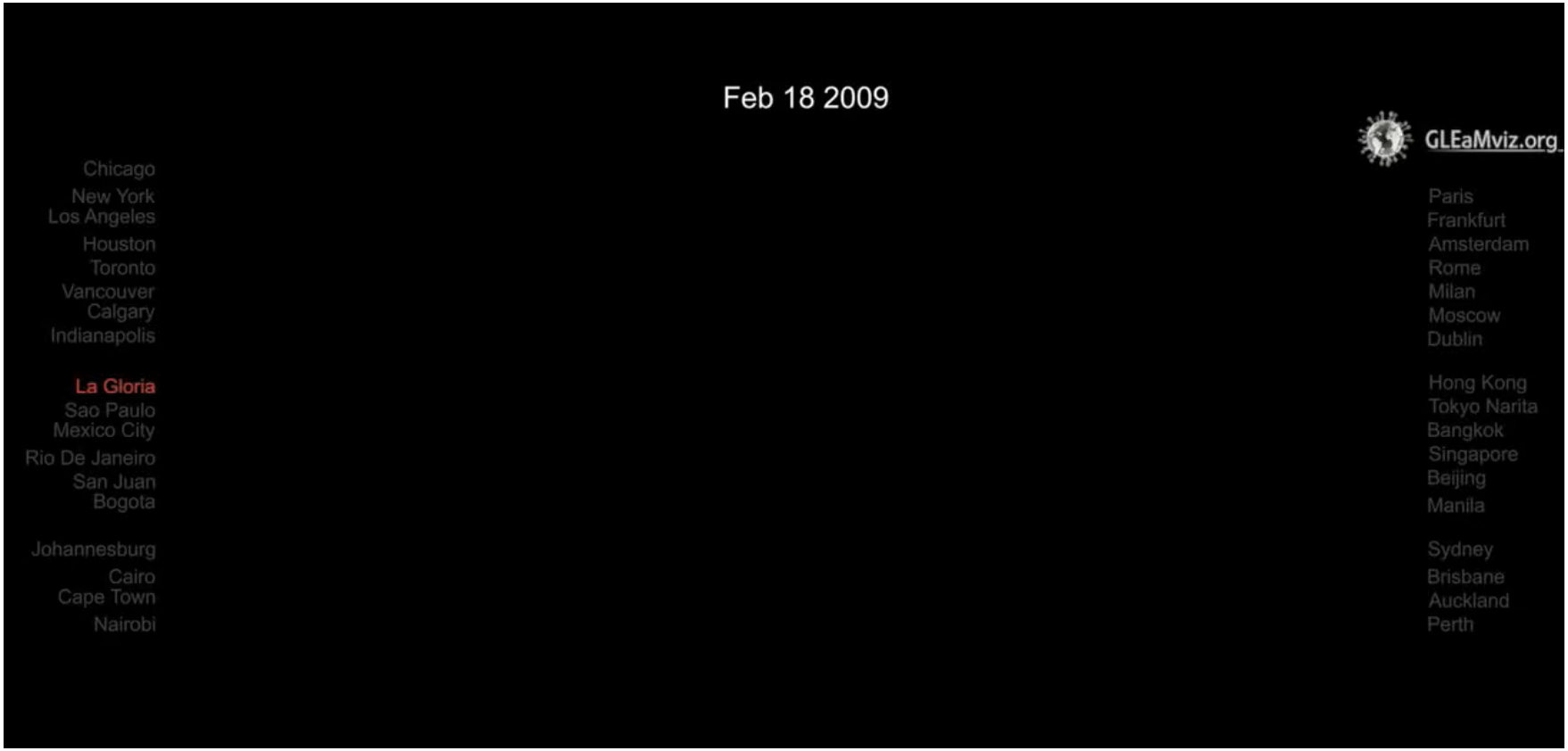
Human-disease networks

- Full diseasome:
  - 1,283 disorders
  - 1,777 genes



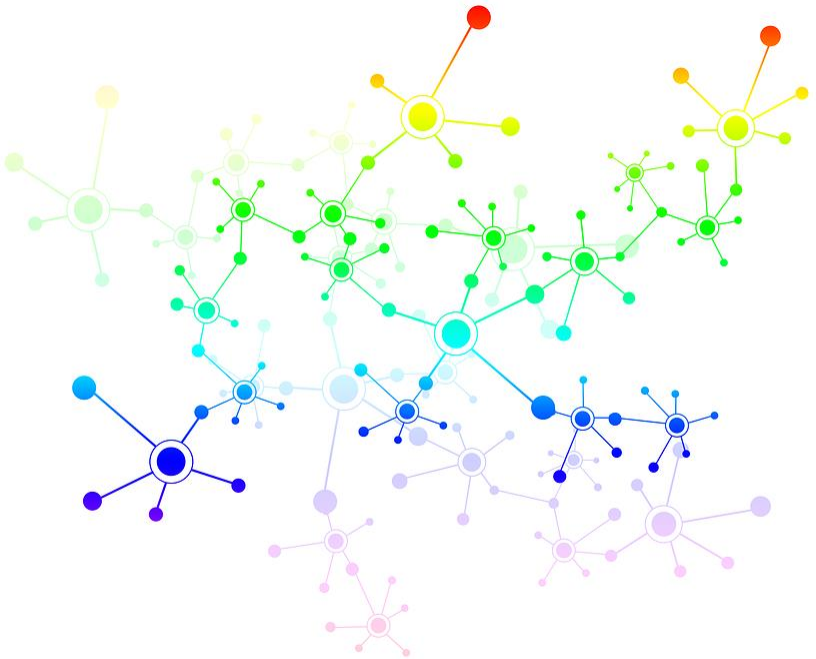
Source: http://networksciencebook.com/images/ch-02/figure-2-10.jpg

Epidemics – H1N1 simulations



Source: http://networksciencebook.com/images/ch-01/video-1-1.m4v

Where can I find the network?
(focus on molecular networks)



Images:
https://cdn.pixabay.com/photo/2016/12/16/15/47/network-1911678_960_720.png

Where do I find *the* network?

- There is no such thing!
- Data-driven vs. knowledge-driven

- >700 different interaction databases

www.pathguide.org

Knowledge resources

- PPI – STRING / IntAct
- Drugs – DrugBank / ChEMBL
- microRNAs – miRTarBase / TargetScan
- Pathways – WikiPathways / Reactome
- Gene-disease – DisGenNET

- .....

- Network sharing
  - NDEx (Network Data Exchange)

Data-driven approaches

- Network inference
  - Finding interactions and connections between (biological) components
  - Often requires large datasets and is computational expensive

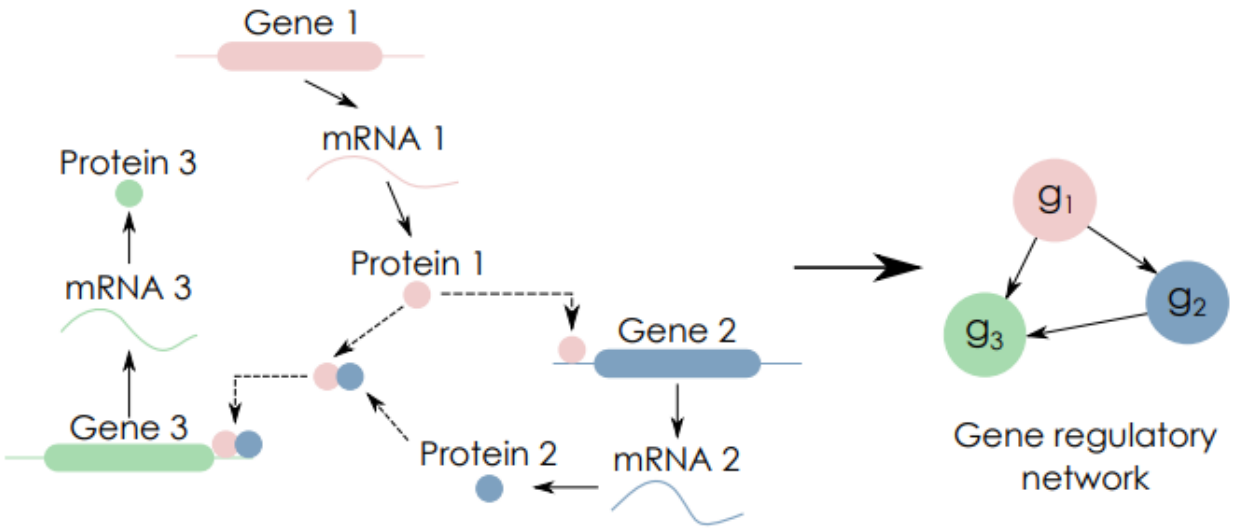


Fig. 1 A cartoon schematic of a gene regulatory network. A complex biophysical model describes the interaction between three genes, involving both direct regulation (gene 2 by gene 1) and combinatorial regulation via complex formation (gene 3 by genes 1 and 2). The abstracted structure of twork on the right.

Images:
https://arxiv.org/abs/1801.04087

Correlation networks

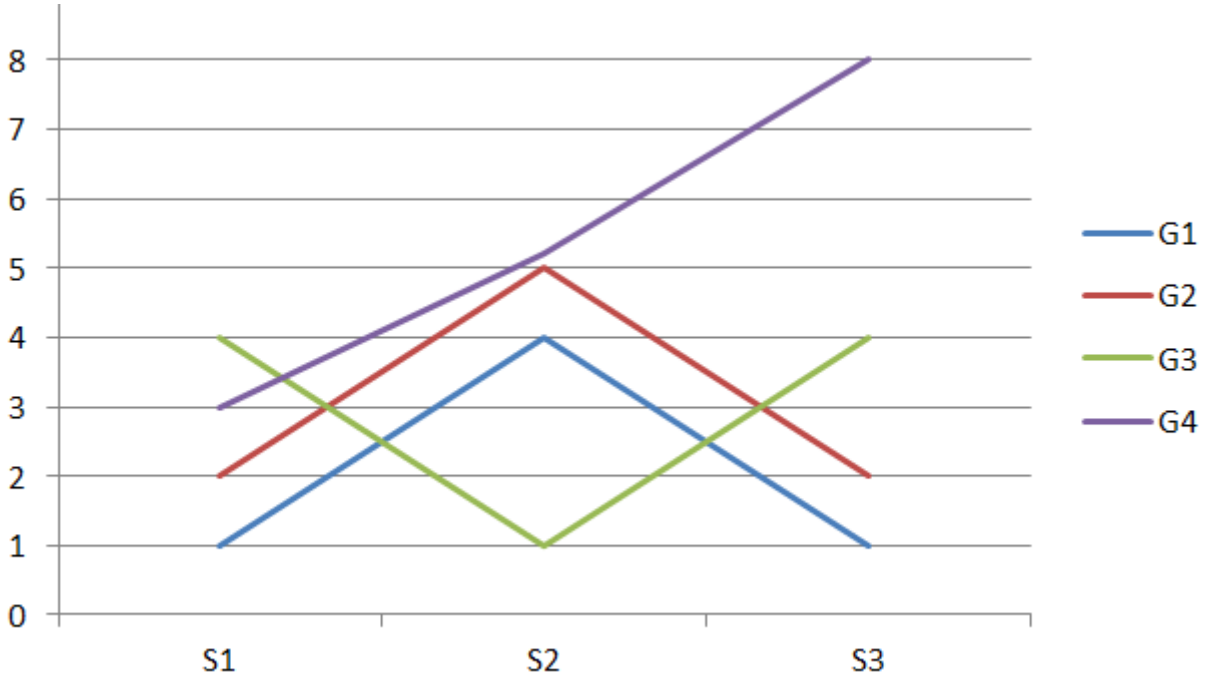
- How strongly are two components related to one another?
- Select certain correlation cut-off to decide if an edge is present or not

- Example: Co-expression networks

Correlation networks

	S_1	S_2	S_3
G_1	43.26	40.89	5.05
G_2	166.6	41.87	136.65
G_3	12.53	39.55	42.09
G_4	28.77	191.92	236.56
G_5	114.7	79.7	99.76
G_6	119.1	80.57	114.59
G_7	118.9	156.69	186.95
G_8	3.76	2.48	136.78
G_9	32.73	11.99	118.8
G_{10}	17.46	56.11	21.41

Gene expression values



Correlation networks

	S_1	S_2	S_3
G_1	43.26	40.89	5.05
G_2	166.6	41.87	136.65
G_3	12.53	39.55	42.09
G_4	28.77	191.92	236.56
G_5	114.7	79.7	99.76
G_6	119.1	80.57	114.59
G_7	118.9	156.69	186.95
G_8	3.76	2.48	136.78
G_9	32.73	11.99	118.8
G_{10}	17.46	56.11	21.41

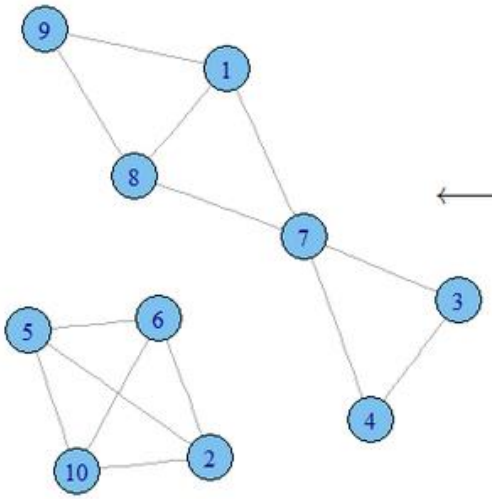
Gene expression values

$|r(G_i, G_j)|$

Pearson correlation

	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9	G_{10}
G_1	1.00	0.23	0.61	0.71	0.03	0.35	**0.86**	**1.00**	**0.97**	0.37
G_2	0.23	1.00	0.63	0.52	**0.98**	**0.99**	0.29	0.30	0.46	**0.99**
G_3	0.61	0.63	1.00	**0.99**	0.77	0.53	**0.93**	0.56	0.41	0.51
G_4	0.71	0.52	**0.99**	1.00	0.69	0.41	**0.97**	0.66	0.52	0.40
G_5	0.03	**0.98**	0.77	0.69	1.00	**0.95**	0.48	0.09	0.27	**0.94**
G_6	0.35	**0.99**	0.53	0.41	**0.95**	1.00	0.17	0.41	0.57	**1.00**
G_7	0.86	0.29	**0.93**	**0.97**	0.48	0.17	1.00	**0.83**	0.72	0.16
G_8	**1.00**	0.30	0.56	0.66	0.09	0.41	0.83	1.00	**0.98**	0.42
G_9	**0.97**	0.46	0.41	0.52	0.27	0.57	0.72	**0.98**	1.00	0.58
G_{10}	0.37	**0.99**	0.51	0.40	**0.94**	**1.00**	0.16	0.42	0.58	1.00

Similarity (Co-expression) score



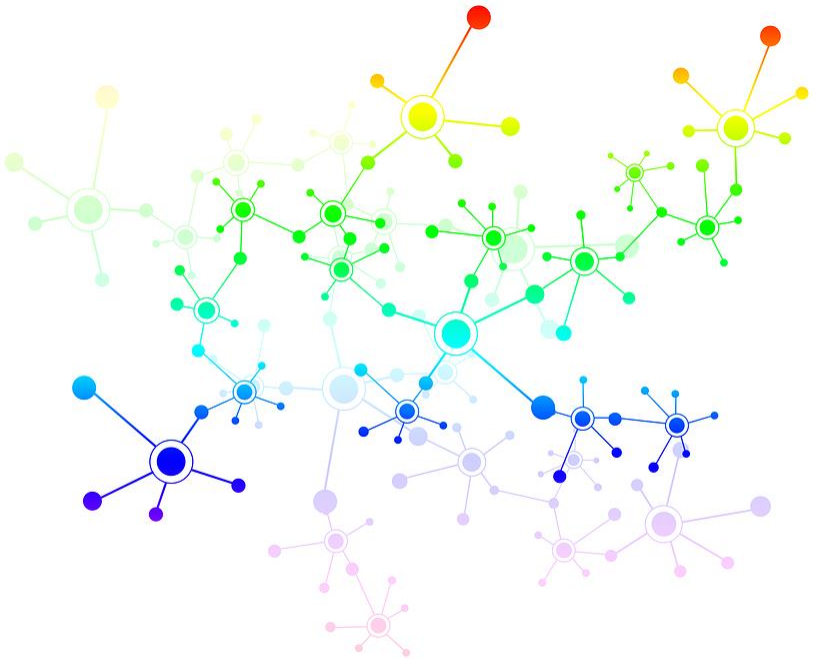
	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9	G_{10}
G_1	0	0	0	0	0	0	1	1	1	0
G_2	0	0	0	0	1	1	0	0	0	1
G_3	0	0	0	1	0	0	1	0	0	0
G_4	0	0	1	0	0	0	1	0	0	0
G_5	0	1	0	0	0	1	0	0	0	1
G_6	0	1	0	0	1	0	0	0	0	1
G_7	1	0	1	1	0	0	0	1	0	0
G_8	1	0	0	0	0	0	1	0	1	0
G_9	1	0	0	0	0	0	0	1	0	0
G_{10}	0	1	0	0	1	1	0	0	0	0

Network adjacency matrix

$|r(G_i, G_j)| >= 0.8$

Significance threshold

Network analysis



Images:
https://cdn.pixabay.com/photo/2016/12/16/15/47/network-1911678_960_720.png

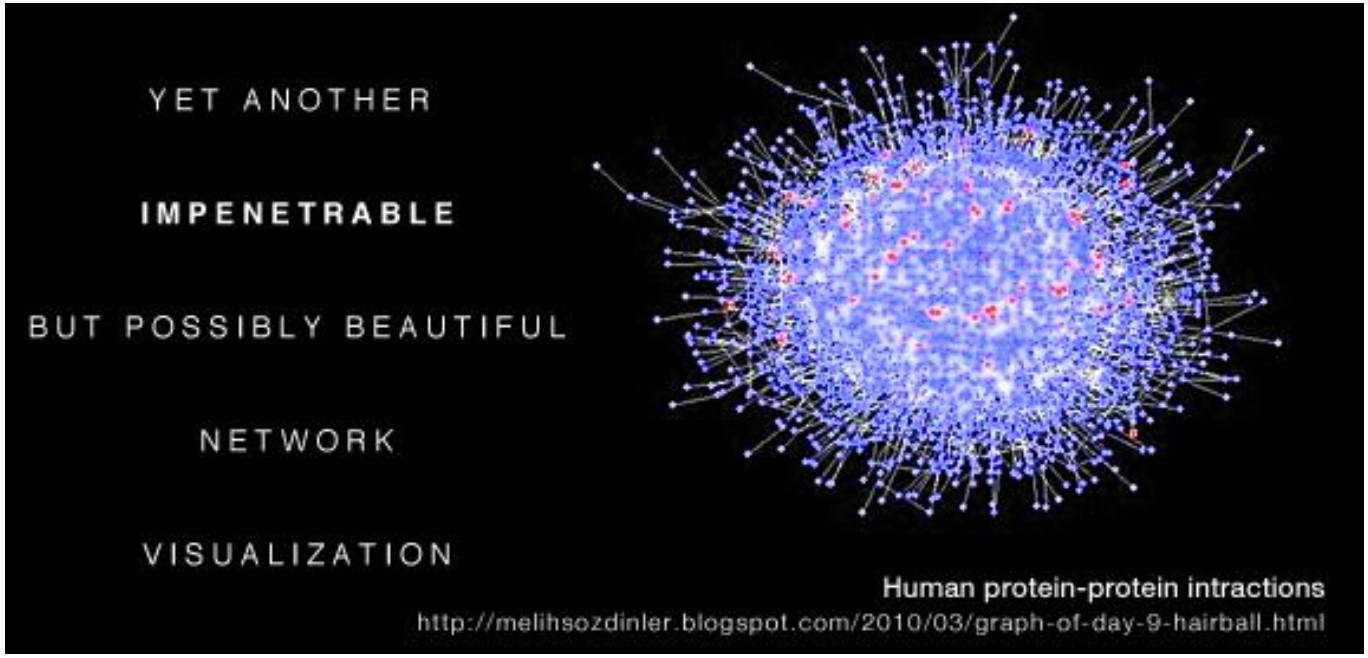
Network algorithms

- Not just integration and visualization

- Algorithms can be used to analyze and study the network – facilitate understanding and interpretation
  - Find hub nodes
  - Find clusters
  - Find patterns / motifs
  - Find modules

Network visualization

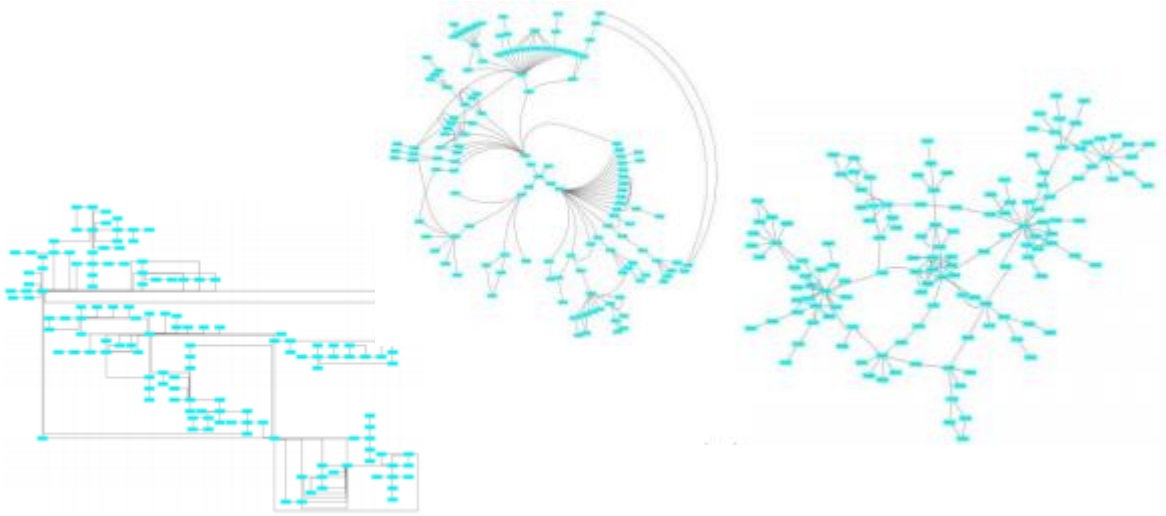
- Large networks become quickly crowded and hard to grasp → hairball syndrome



Images:
http://www.hiveplot.com/img/hiveplot-hairball-1.png

Network visualization

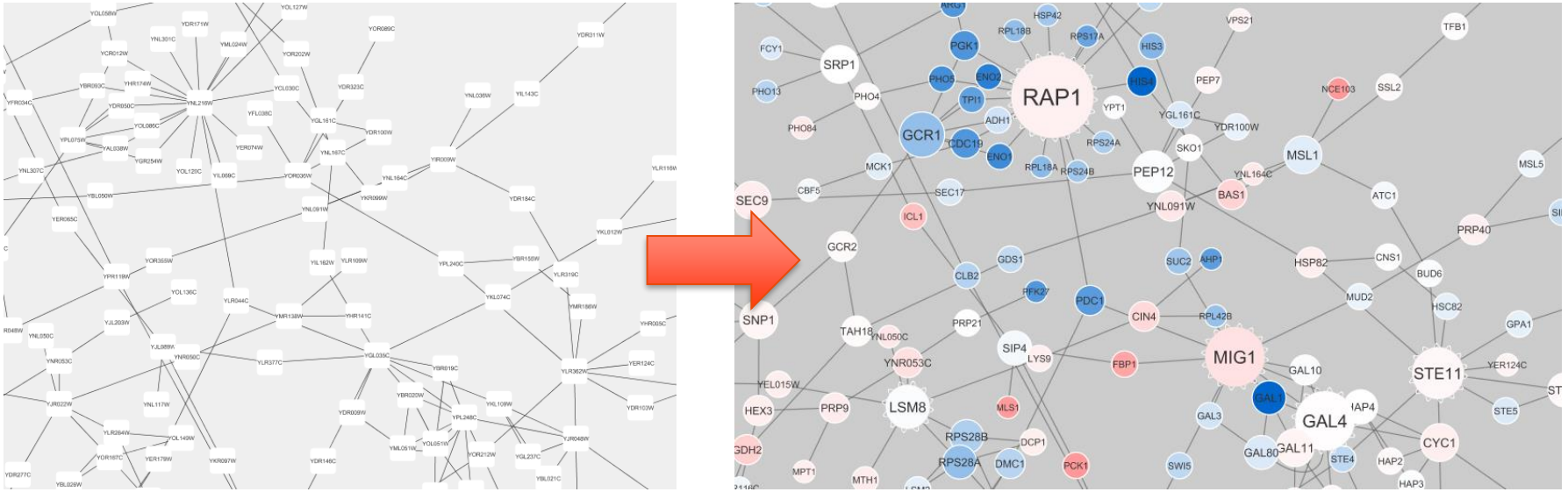
- Layouts
  - Try different algorithms to arrange the nodes of the network
    - circular
    - hierarchical
    - Tree
    - orthogonal
    - forced-directed
    - radial
    - organic
    - etc.



  - Emphasize topology and/or properties of the network
  - Illustrate high/low traffic areas in the network

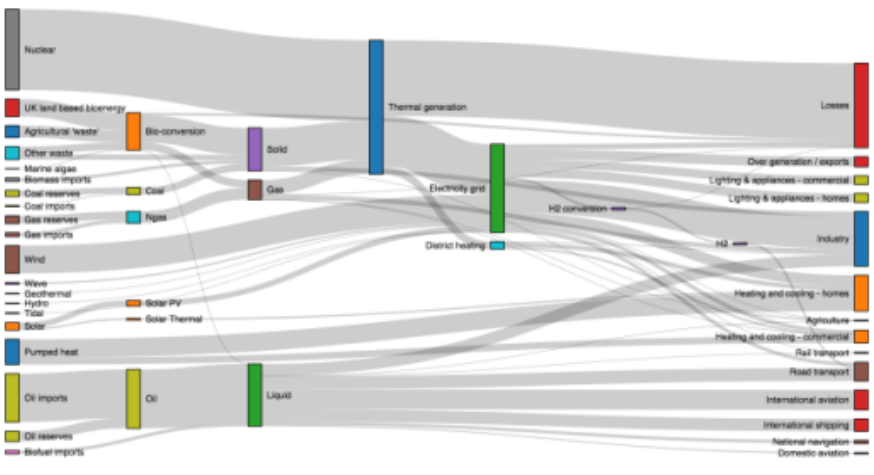
Network visualization

- Data visualization
  - Node shape, size, color, transparency
  - Edge type, shape, width, color, transparency

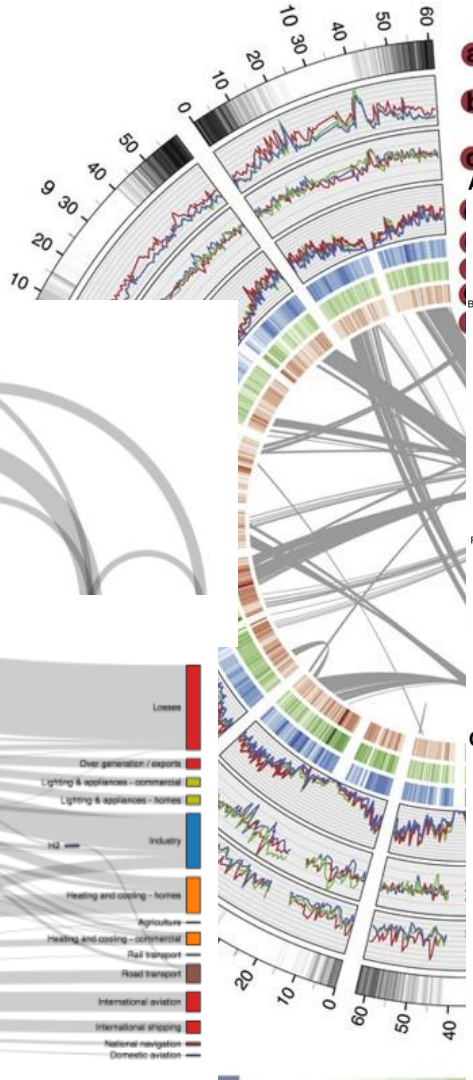


Advanced Visualization

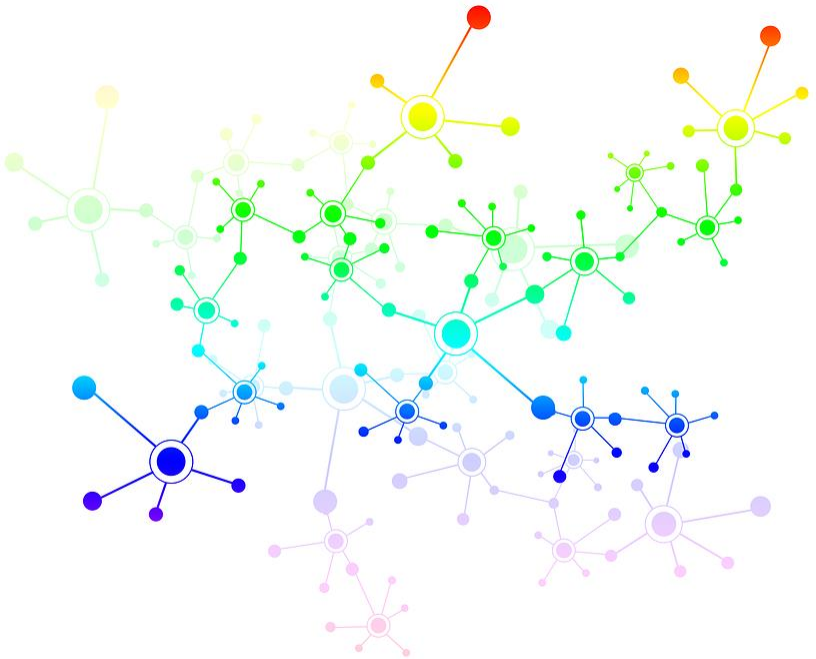
- Circular plots
- Arc diagram
- Sankey plots
- Hive plots



d3 Sankey - Github



Hands-on session



Images:
https://cdn.pixabay.com/photo/2016/12/16/15/47/network-1911678_960_720.png

Hands-on

- This is a basic introduction to network analysis!
- Quick Cytoscape demo
- Hands-on session
  - Cytoscape user interface
  - Automation in R (prior R programming experience required)

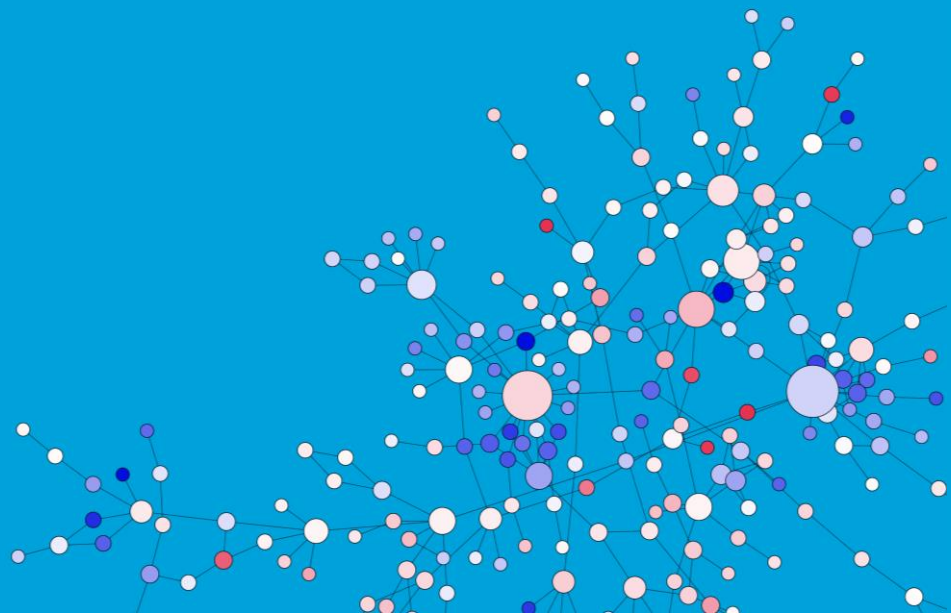
Questions?

Contact:
Martina Summer-Kutmon
martina.kutmon@maastrichtuniversity.n

Maastricht University

Helis Academy

Interreg
Vlaanderen-Nederland
Europees Fonds voor Regionale Ontwikkeling



Helis Academy Partners

BIOTECH TRAINING FACILITY

DTL DUTCH TECHCENTRE FOR LIFE SCIENCES

flanders.bio

health house EXPERIENCE THE FUTURE

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

vives

Hyphen Projects

Maastricht University

TU Delft

VIB

Helis Academy Data Analysis and Data Stewardship Funding

Provincie Antwerpen

Ministerie van Economische Zaken en Klimaat

provincie limburg

Interreg Vlaanderen-Nederland
Europees Fonds voor Regionale Ontwikkeling

Provincie Noord-Brabant

Interreg Vlaanderen-Nederland
Europees Fonds voor Regionale Ontwikkeling

Maastricht University

Helis Academy