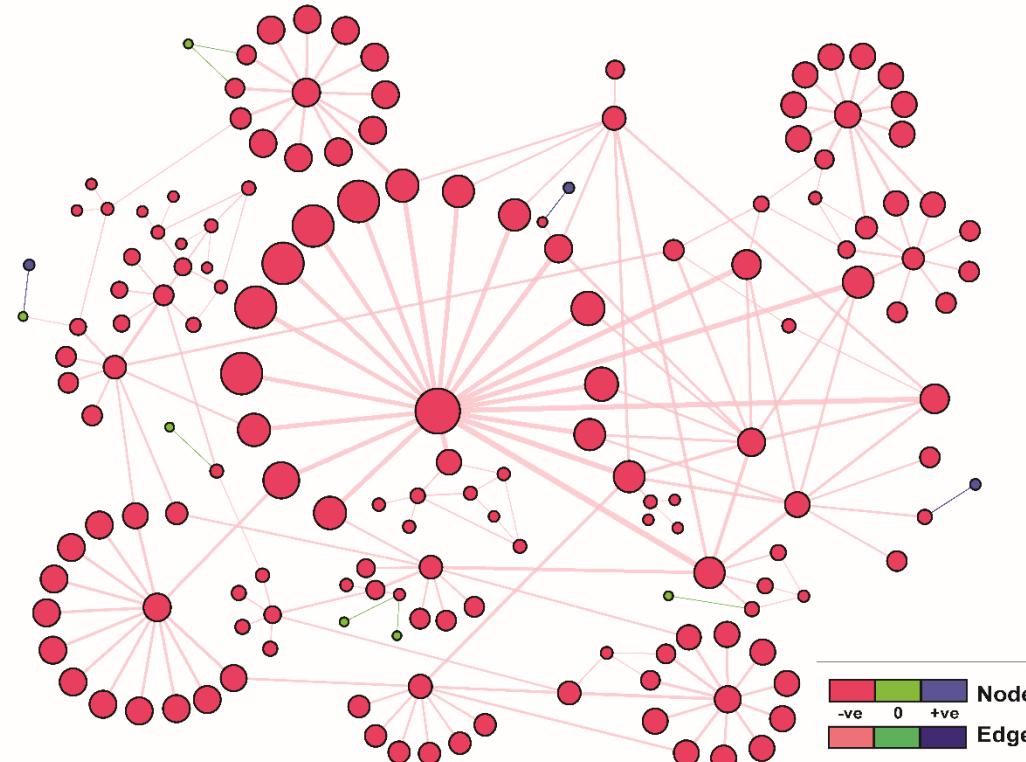


Introduction to Networks and Graph Theory



Areejit Samal

The Institute of Mathematical Sciences (IMSc)
Chennai



Why Networks?

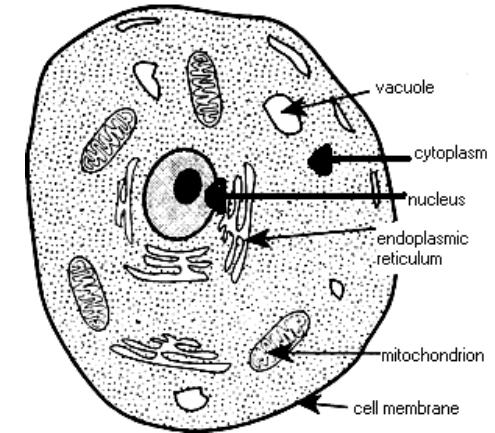
Complex systems pervade our everyday life



Transportation System



Society



Cell



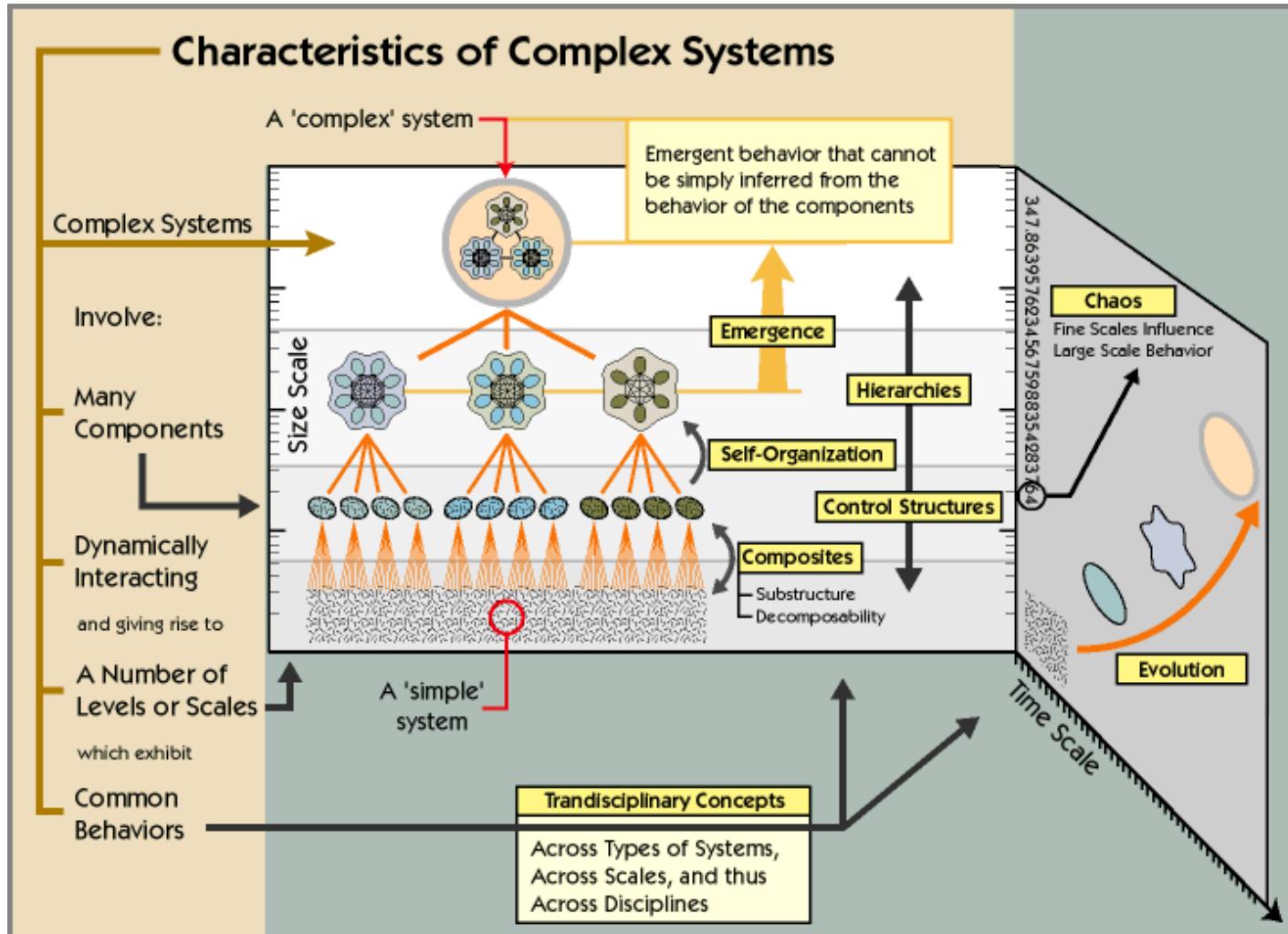
Economy

Any system of coupled elements can be represented as a 'network'

Image credits:

<http://www.jinnan.com.cn/en/ProductService/Multi-transport/6683.shtml>
<https://meridian-180.org/en/forums/forum-summary-sharing-economy>
<https://www.2020spaces.com/global-network-of-people/>
<http://regentsprep.org/>

Complex system: 'Whole is more than the sum of the parts'



To understand the design principles of *complex systems*, one requires an understanding of the design principles of underlying *complex networks*

Transportation Network



Nodes are cities and Links specify road connections

RP Sreejith et al JSTAT (2016)

Social Network



Nodes are individuals while links are relationships (professional or personal)

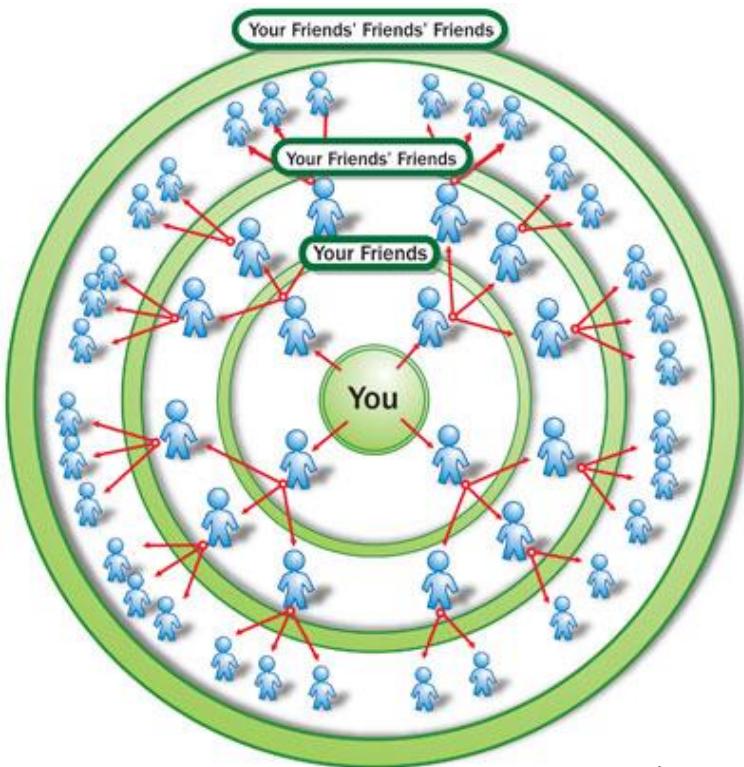
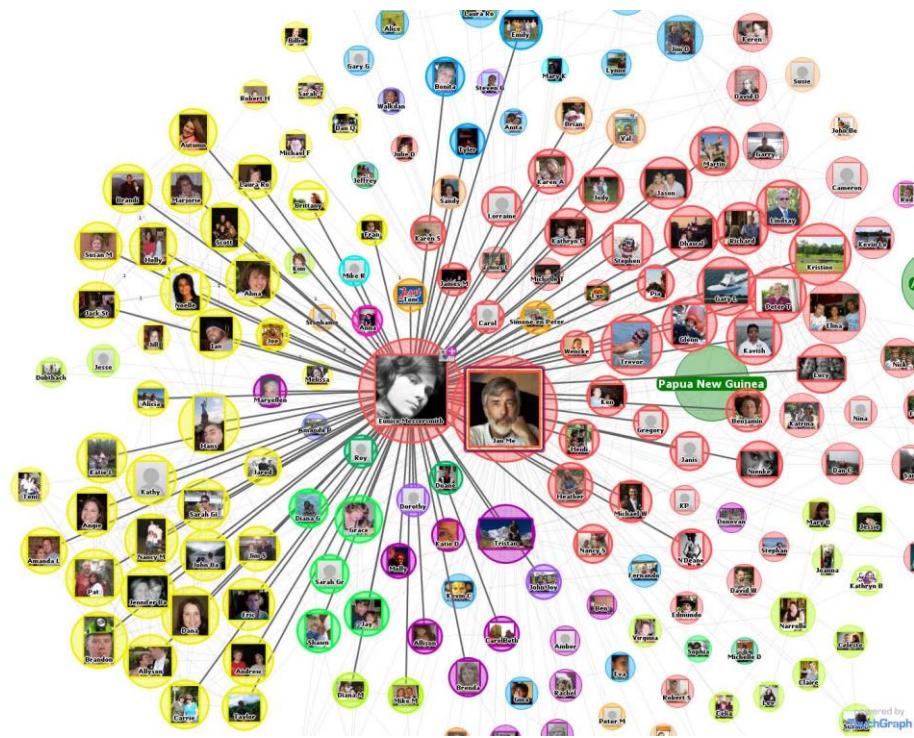


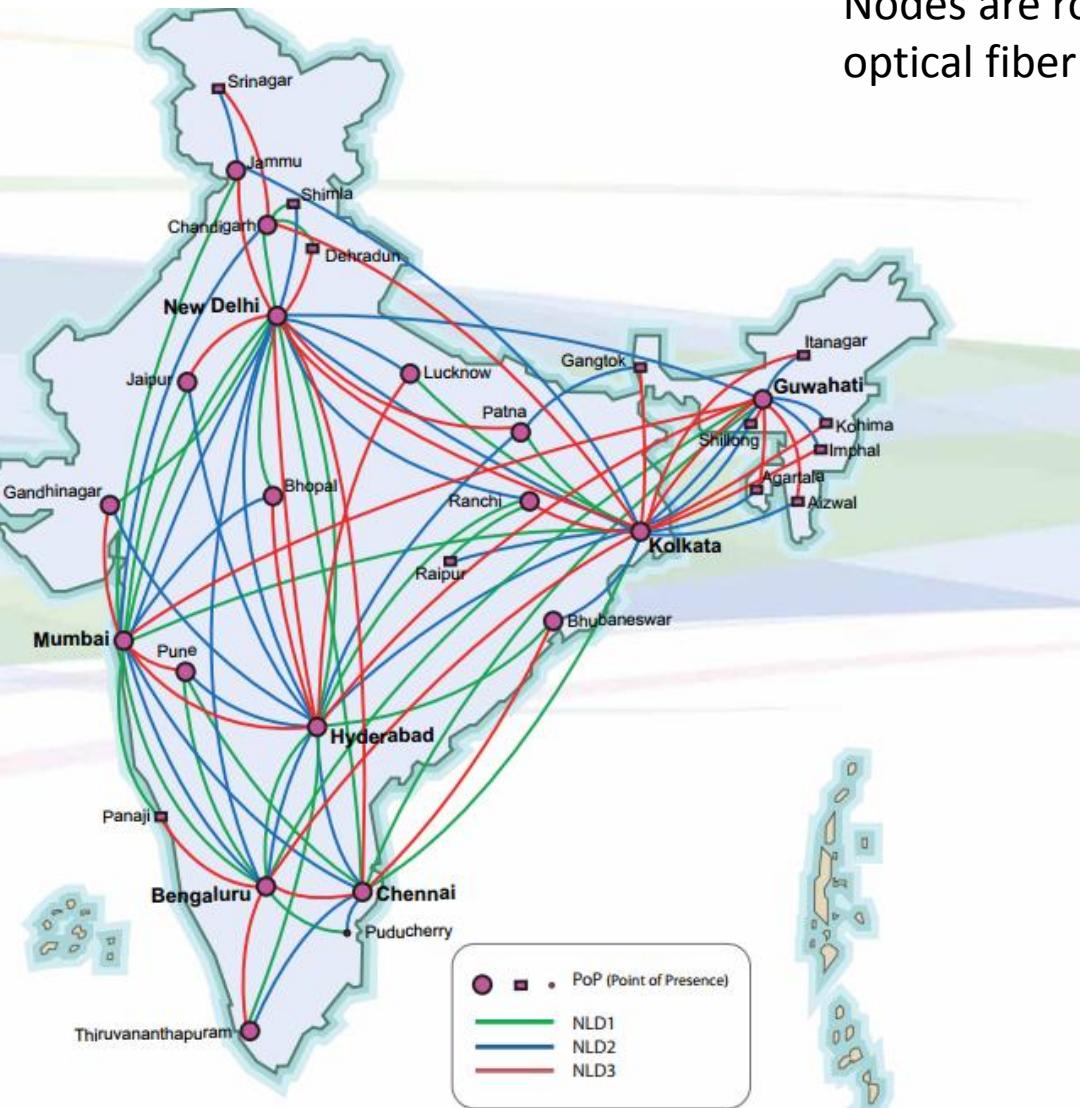
Image credits:

http://www.messersmith.name/wordpress/wp-content/uploads/2009/10/touchgraph_eunice_cluster.jpg
<http://computer.howstuffworks.com/internet/social-networking/networks/friendster3.htm>



Communication Networks - Internet

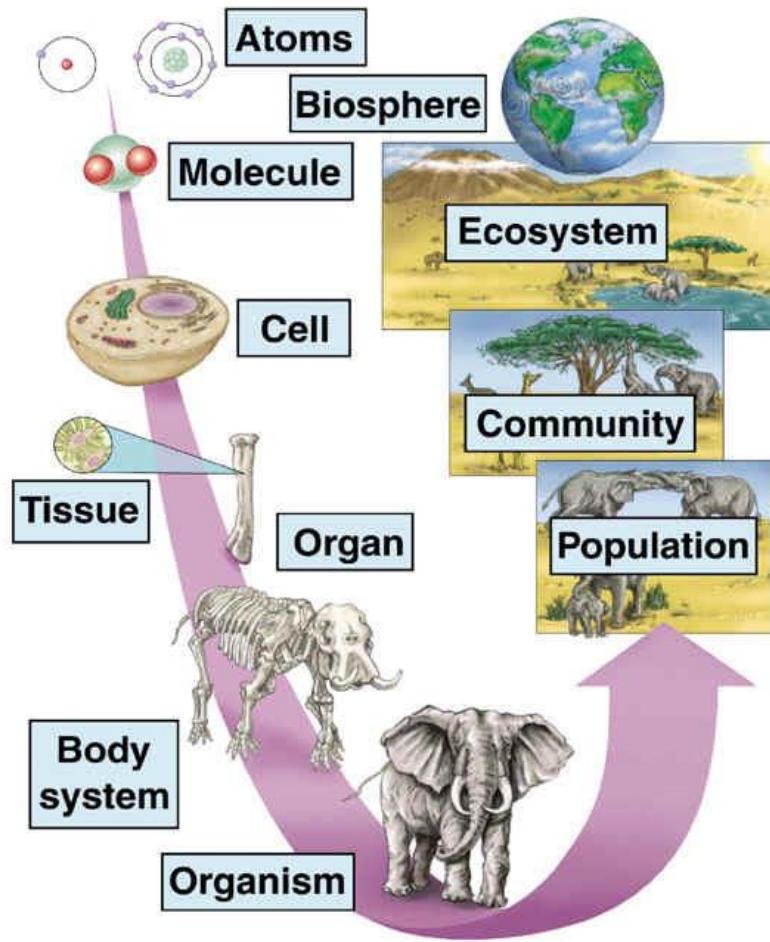
Nodes are routers while links are optical fiber cables linking the routers



National Knowledge Network (NKN)

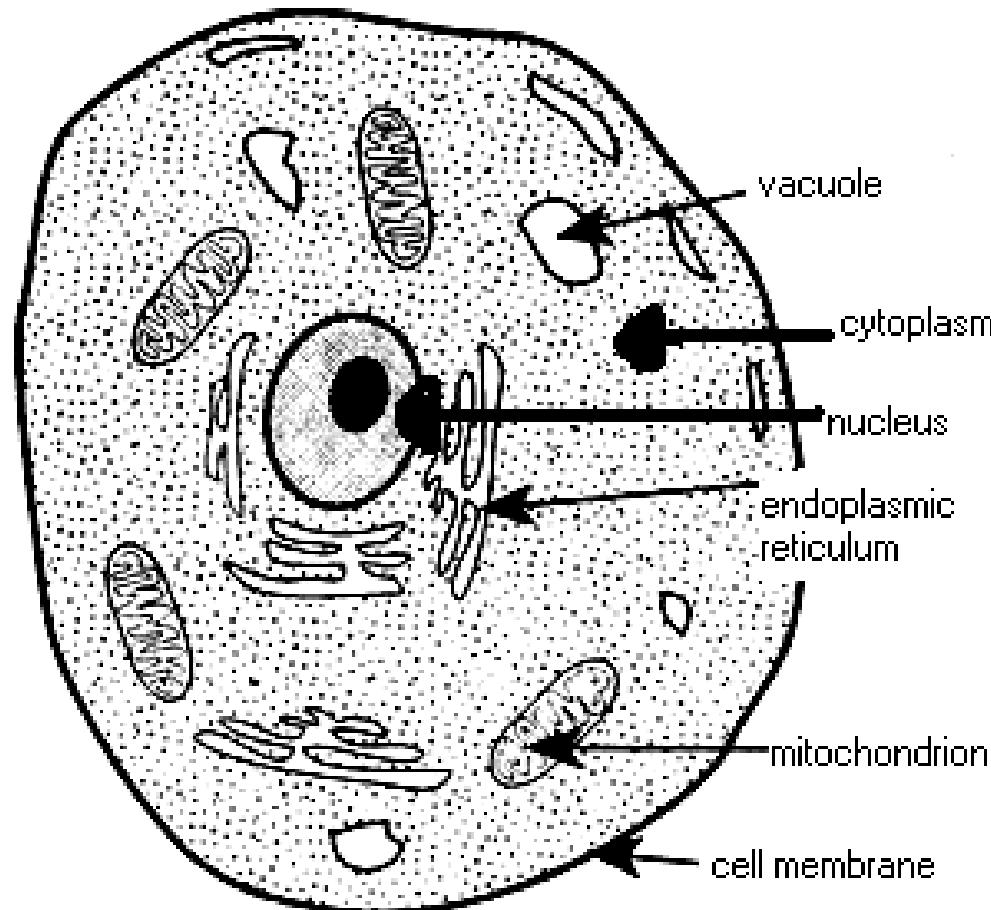
Complexity increases with the levels of biological organization

Raven/Berg, Environment, 3/e
Figure 4.1

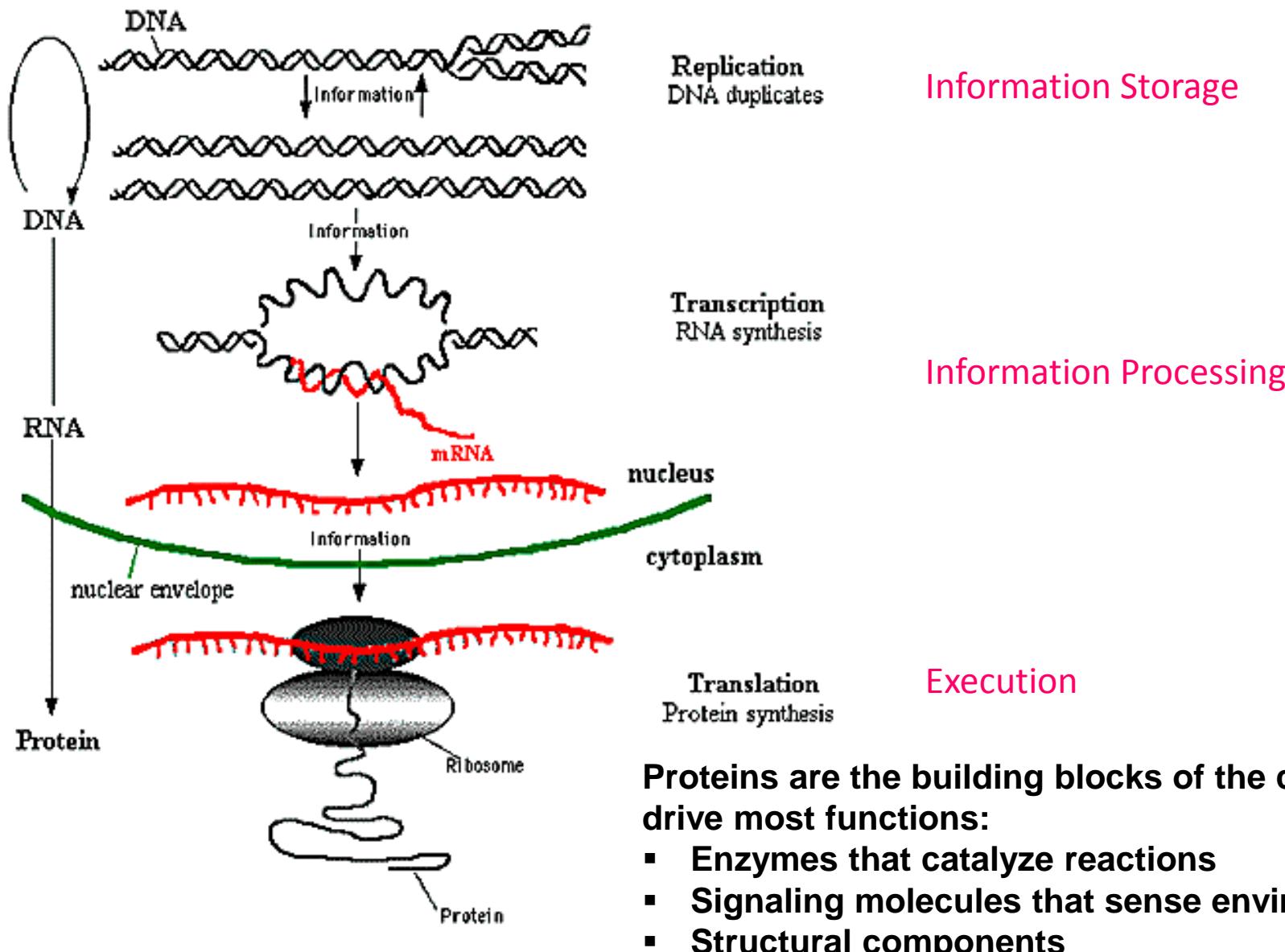


Harcourt, Inc.

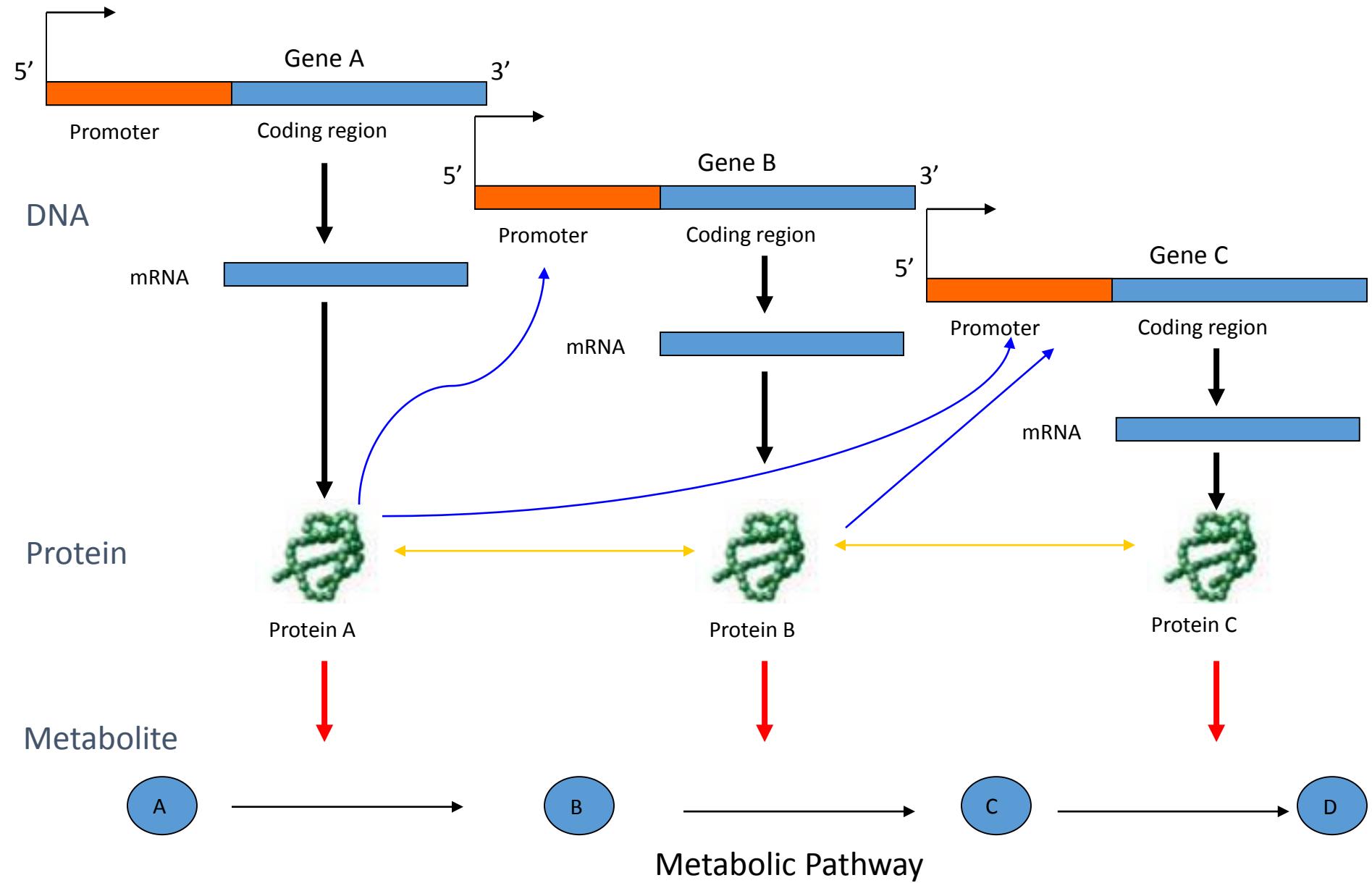
Cell – Basic unit of life



Central dogma of molecular biology



Cell as a ‘network of networks’

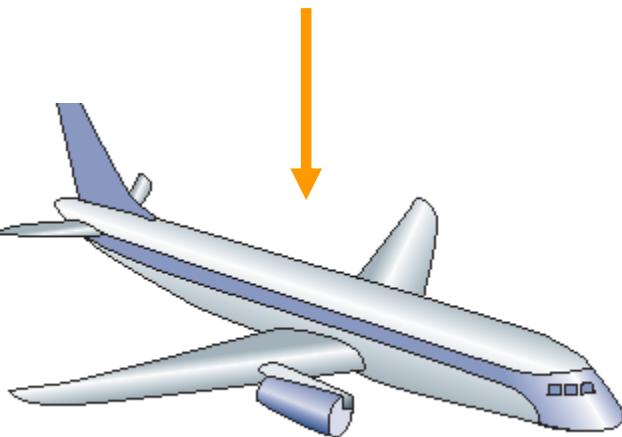


Cellular (Biological) Networks

Cell can be viewed as an overlay of at least three different networks:

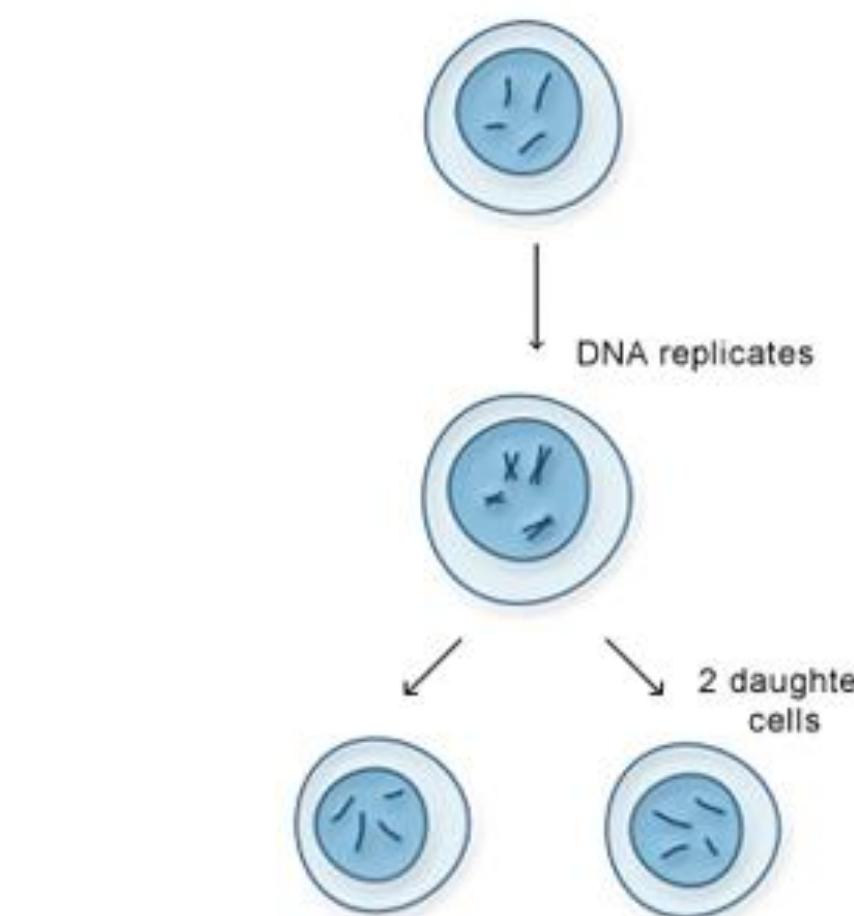
- Protein–DNA (Transcriptional regulatory network)
- Protein–Protein (Protein interaction network)
- Protein-Metabolite (Metabolic network)

What are the design principles of living systems?



Product of airplane factory is an airplane
The airplane (product) cannot produce
the factory.

The design of the two factories must be different.



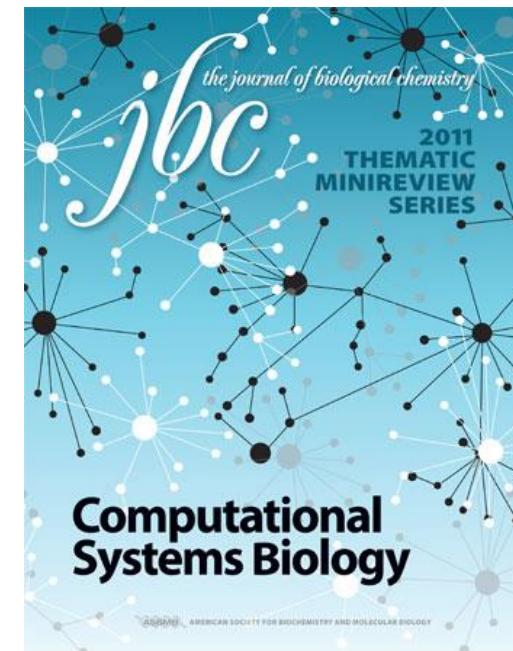
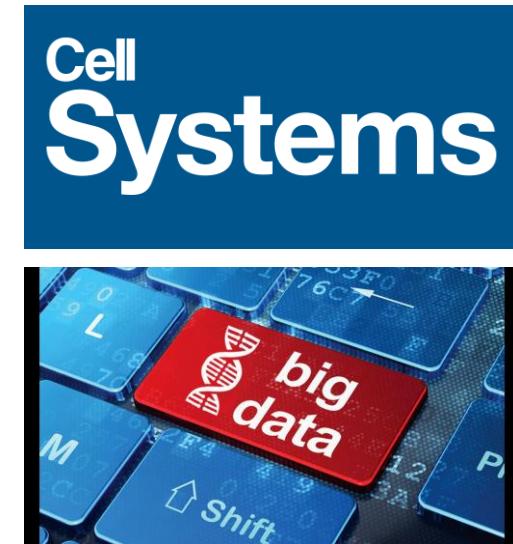
Product of yeast is another yeast. The factory is
itself the product including workers and managers!

Systems Biology

- Understand the biological complexity
- Quantitatively predict the effect of changes in one part of the system (network) on other parts of the system (network)
- Need to chart the maps of various interaction networks inside the cell
- Need information on strength of interactions between various components

Systems Biology ~ Network Biology

- For a systems-level understanding:
 - Structure
 - Dynamics - How system behaves over time
- Tools of graph theory has been extensively used to study the structure of real biological networks



Whole genome sequences provide the ‘parts list’ for an organism

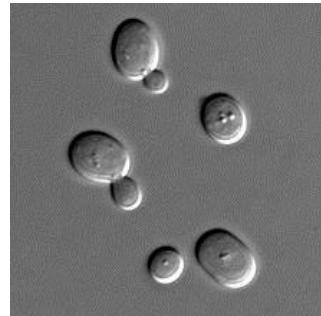
- Towards the end of last century, completion of genome sequencing projects for several organisms revolutionized biology
- Complete genome sequences are now available for several organisms including humans



The Complete Genome Sequence of *Escherichia coli* K-12

E. coli

Frederick R. Blattner,* Guy Plunkett III,* Craig A. Bloch, Nicole T. Perna, Valerie Burland, Monica Riley, Julio Collado-Vides, Jeremy D. Glasner, Christopher K. Rode, George F. Mayhew, Jason Gregor, Nelson Wayne Davis, Heather A. Kirkpatrick, Michael A. Goeden, Debra J. Rose, Bob Mau, Ying Shao



Life with 6000 Genes

Yeast

A. Goffeau,* B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, S. G. Oliver



The Sequence of the Human Genome

J. Craig Venter,^{1*} Mark D. Adams,¹ Eugene W. Myers,¹ Peter W. Li,¹ Richard J. Mural,¹ Granger G. Sutton,¹ Hamilton O. Smith,¹ Mark Yandell,¹ Cheryl A. Evans,¹ Robert A. Holt,¹ Jeannine D. Gocayne,¹ Peter Amanatides,¹ Richard M. Ballew,¹ Daniel H. Huson,¹ Jennifer Russo Wortman,¹ Qing Zhang,¹ Chinnappa D. Kodira,¹ Xiangqun H. Zheng,¹ Lin Chen,¹ Marian Skupski,¹ Gangadharan Subramanian,¹ Paul D. Thomas,¹ Jinghui Zhang,¹ George L. Gabor Miklos,² Catherine Nelson,³ Samuel Broder,¹ Andrew G. Clark,⁴ Jon Nadeau,⁵ Victor A. McKusick,⁶ Norton Zinder,⁷ Arnold J. Levine,⁷ Richard J. Roberts,⁸ Mel Simon,⁹ Carolyn Slayman,¹⁰ Michael Hunkapiller,¹¹ Randall Bolanos,¹ Arthur Delcher,¹ Ian Dew,¹ Daniel Fusalo,¹ Michael Flanigan,¹ Liliana Florea,¹ Aaron Halpern,¹ Sridhar Hannenhalli,¹ Saul Kravitz,¹ Samuel Levy,¹ Mark Mobarry,¹ Knut Reinert,¹ Karin Remington,¹ Jane Abu-Threideh,¹ Ellen Beasley,¹ Kendra Biddick,¹ Vivien Bonazzi,¹ Rhonda Brandon,¹ Michele Cargill,¹ Ishwar Chandramouliswaran,¹ Rosane Charlab,¹

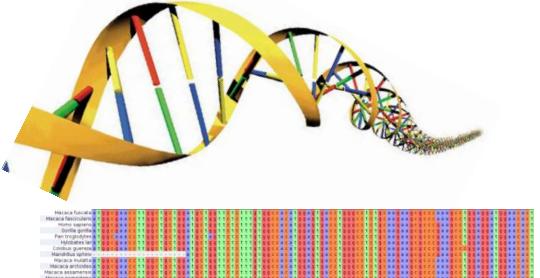
Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

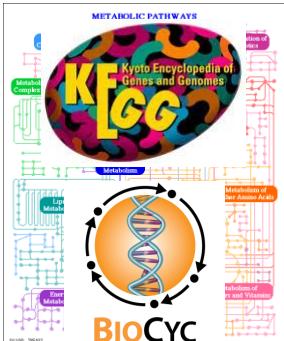
Image credit: Wikipedia

Network reconstruction – ‘Parts list’ to ‘ interactome’

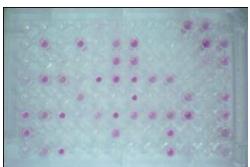
Genome sequence with annotation



Biochemical Knowledge



Physiology



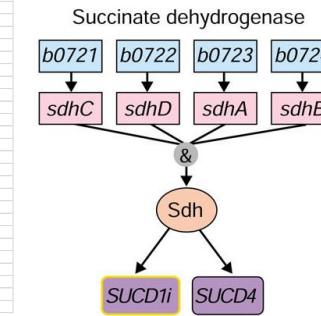
ESCHERICHIA COLI AND SALMONELLA TYPHIMURIUM
CELLULAR AND MOLECULAR BIOLOGY

Genome-scale reconstructed metabolic network

List of metabolic reactions that can occur in an organism along with gene-protein -reaction (GPR association)

Abbreviation	OfficialName	Equation [mete] and [ene] at the beginning refer to t: Subsystem	ProteinClassDescription
ALATA_L	L-alanine transaminase	[c]ala-L <=> glu-L + pyr	Alanine and aspartate metabolism EC-2.6.1.1
ALAR	alanine racemase	[c]ala-L <=> ala-D	Alanine and aspartate metabolism EC-5.1.1.1
ASNN	L-asparaginase	[c]asn-L + h2o >> asp-L + nh4	Alanine and aspartate metabolism EC-3.5.1.1
ASN2	asparagine synthetase	[c]asp-L + atp + nh4 >> amp + asn-L + h + ppi	Alanine and aspartate metabolism EC-6.3.1.1
ASN51	asparagine synthetase (glutamine-hydrolyzing)	[c]asp-L + atp + gln-L + h2o >> amp + asn-L + glu-L + h	Alanine and aspartate metabolism EC-6.3.5.4
ASP1	aspartate transaminase	[c]asp-L > jum + nh4	Alanine and aspartate metabolism EC-4.3.1.1
VPAWT	Valine-syntrota aminotransferase	[c]ala-L + atp + glu-L > jum + oxa	Alanine and aspartate metabolism EC-4.3.1.1
DAAD	D-Amino acid dehydrogenase	[c]3mmba + ala-L > pyr + val-L	Alanine and aspartate metabolism EC-2.6.1.65
ALAR	alanine racemase (irreversible)	[c]ala-D + fad + h2o > fad2 + nh4 + pyr	Alanine and aspartate metabolism EC-1.4.99.1
FFSD	beta-fructofuranosidase	[c]ala-L > jum + ala-D	Alanine and aspartate metabolism EC-5.1.1.1
ASPISO	arabinose-5-phosphate isomerase	[c]h2o + suc6p > fru + gip	Alternate Carbon Metabolism EC-3.2.1.26
MME	methylmalonyl-CoA epimerase	[c]mmcoa-R <=> mmcoa-S	Alternate Carbon Metabolism EC-5.3.1.13
MKD	2-methylcitrate dehydratase	[c]2mccan + h2o > mci	Alternate Carbon Metabolism EC-4.2.1.99
ALD019	alcohol dehydrogenase (glutaral)	[c]glyal + nad > jum + glyc + nad	Alternate Carbon Metabolism EC-1.1.1
LOAD	lactate dehydrogenases	[c]h2o + lact-L + nad > (2) h + lac-L + nadh	Alternate Carbon Metabolism EC-1.2.1.22
TGBPA	Tagatose-bisphosphate aldolase	[c]tagdp-D >=> dhap + p3o	Alternate Carbon Metabolism EC-4.1.2.40
LCAD	lactate/dehydrogenase	[c]h2o + laht-L + nad > (2) h + lac-L + nadh	Alternate Carbon Metabolism EC-1.2.1.22
ALDD2x	aldehyde reductase (acetaldehyde, NAD)	[c]acalac + h2o + nad > ac + (2) h + nadh	Alternate Carbon Metabolism EC-1.2.1.3
ARAI	L-arabinose isomerase	[c]araal-L <=> rbl-L	Alternate Carbon Metabolism EC-5.3.1.4
RBL_1	L-rubulokinase (L-rubulose)	[c]catp + rbl-L > adp + h + ru5p-L	Alternate Carbon Metabolism EC-2.7.1.16
RBP44	L-rubulose-5-phosphate 4-epimerase	[c]ru5p-L <=> xulbp-D	Alternate Carbon Metabolism EC-5.3.1.4
ACO2	Acetyl-CoA:acyl-CoA acyltransferase	[c]acetyl-CoA + acyl-CoA > acetyl-CoA + ac	Alternate Carbon Metabolism EC-2.8.3.8
BUTOT	Acetyl-CoA:butyrate-CoA transferase	[c]decetyl-CoA + h > brcos	Alternate Carbon Metabolism EC-3.2.1.86
AB6PGH	Adenosin 5'-phosphate glucohydrolase	[c]carb6p + h2o > gip + hgn	Alternate Carbon Metabolism EC-5.4.2.8
PMAMM	phosphomannomutase	[c]man1p <=> man6p	Alternate Carbon Metabolism EC-5.4.2.7
PPM2	phosphoenolpyruvate 2 (deoxyribose)	[c]2driP <=> 2driP	Alternate Carbon Metabolism EC-5.4.2.7

GPR Associations



Genome-scale reconstructed metabolic model

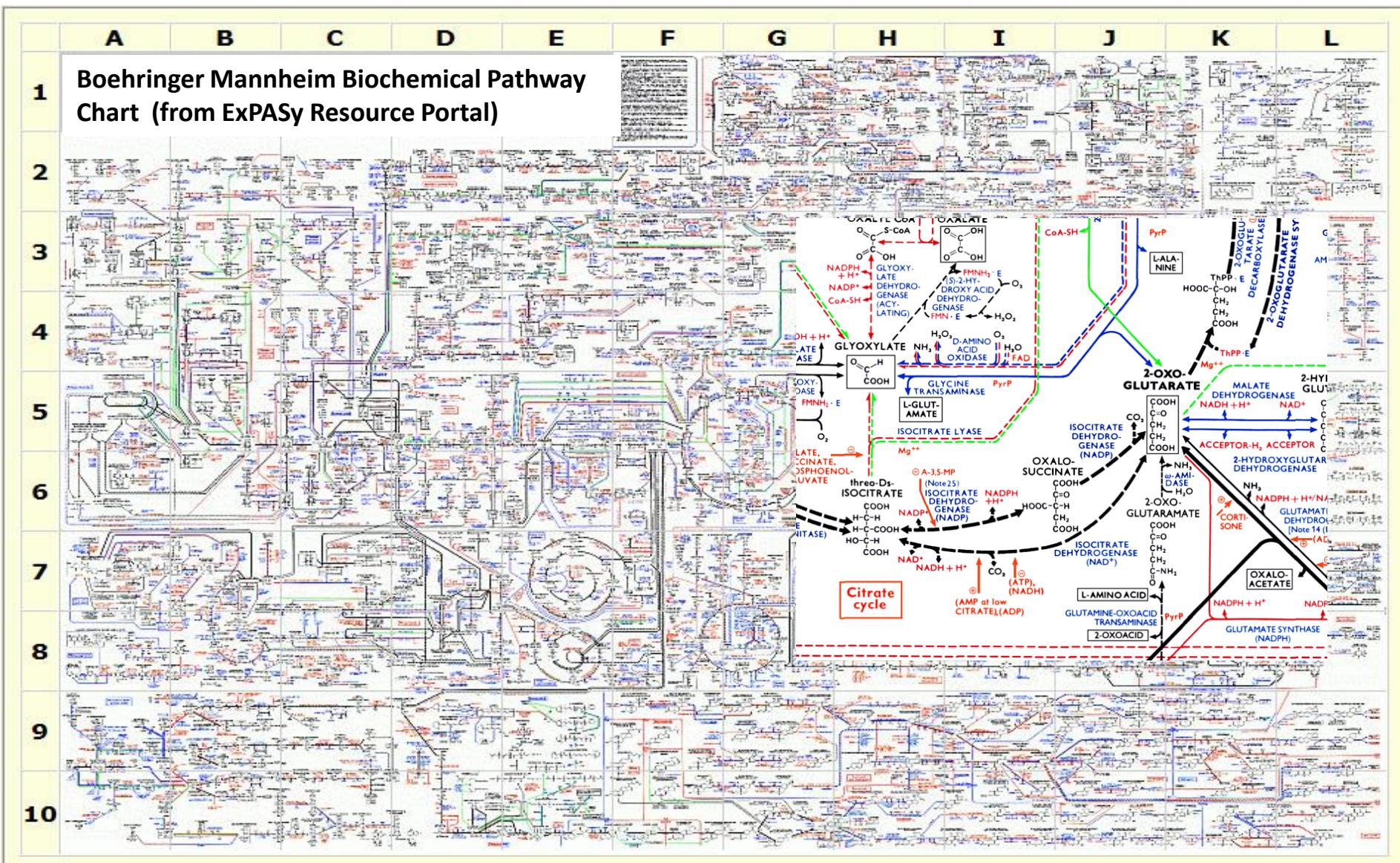
Exchange and transport reactions defining possible nutrients

ACALDt	acetaldehyde reversible transport	acald[e] <=> acald[j]	Transport, Extracellular
GUAT	Guanine transport	guat[e] <=> guat[c]	Transport, Extracellular
HYXNt	Hypoxanthine transport	hxan[e] <=> hxan[c]	Transport, Extracellular
XANT	xanthine reversible transport	xan[e] <=> xan[c]	Transport, Extracellular
NACUP	nicotinic acid uptake	nac[e] <=> nac[i]	Transport, Extracellular
ASNabc	L-asparagine transport via ABC system	asn-L[e] + atp[c] + h2o[c] >> adp[c] + asn-L[c] + h[c] + p	Transport, Extracellular
ASN2r	L-asparagine reversible transport via proton symport	asn-L[e] + h[e] <=> asn-L[c] + h[c]	Transport, Extracellular
DAPabc	M-diaminopimelic acid ABC transport	26dap-M[e] + atp[c] + h2o[c] >> 26dap-M[c] + adp[l] + h	Transport, Extracellular
CYSabc	L-cysteine transport via ABC system	atp[c] + cys-L[e] + h2o[c] >> adp[c] + cys-L[c] + h[c] + p	Transport, Extracellular
ACT2r	acetate reversible transport via proton symport	acel[e] + h[e] <=> acd[e] + h[e]	Transport, Extracellular
ETOH2r	ethanol reversible transport via proton symport	etoh[e] + h[e] <=> etho[h] + h[c]	Transport, Extracellular
PYR12r	pyruvate reversible transport via proton symport	pyr[e] + pyr[f] <=> [h]o + pyr[f]	Transport, Extracellular
O2t	o2 transport (diffusion)	o2[e] <=> o2[c]	Transport, Extracellular
CO2t	CO2 transporter via diffusion	co2[e] <=> co2[c]	Transport, Extracellular
H2Ot	H2O transport via diffusion	h2o[e] <=> h2o[c]	Transport, Extracellular
DHAt	Dihydroxyacetone transport via facilitated diffusion	dha[e] <=> dha[c]	Transport, Extracellular
NH3t	ammonia reversible transport	nh4[e] <=> nh4[c]	Transport, Extracellular
ARB12r	L-arabinose transport via proton symport	arab-L[e] + h[e] <=> arab-L[c] + h[c]	Transport, Extracellular

Cellular Objective: Biomass Reaction

(0.05) 5mthf + (5.0E-5) accoa + (0.488) ala-L + (0.0010) amp + (0.281) arg-L + (0.229) asn-L + (0.45.7318) atp + (1.29E-4) clpn_EC + (6.0E-6) coa + (0.126) ctp + (0.087) cys-L + (0.0247) datp + (0.0254) dctp + (0.0254) dtgp + (0.0247) dtpp + (1.0E-5) fad + (0.25) glu-L + (0.25) glu-L + (0.582) gly + (0.154) glycogen + (0.203) gtp + (45.5608) h2o + (0.09) his-L + (0.276) ile-L + (0.428) leu-L + (0.0084) lps_EC + (0.326) lys-L + (0.146) met-L + (0.00215) nad + (0.4E-5) nadh + (4.0E-4) nadph + (0.001935) pe_Ec + (0.0276) peptide_Ec + (4.64E-4) pg_Ec + (0.176) phe-L + (0.21) pro-L + (5.2E-5) ps_Ec + (0.035) ptc + (0.205) ser-L + (0.0070) spmd + (3.0E-6) succoa + (0.241) thr-L + (0.054) trp-L + (0.0030) ufdg + (0.136) utp + (0.402) val-L -> (45.5608) adp + (45.56035) h + (45.5628) pi + (0.7302) ppi

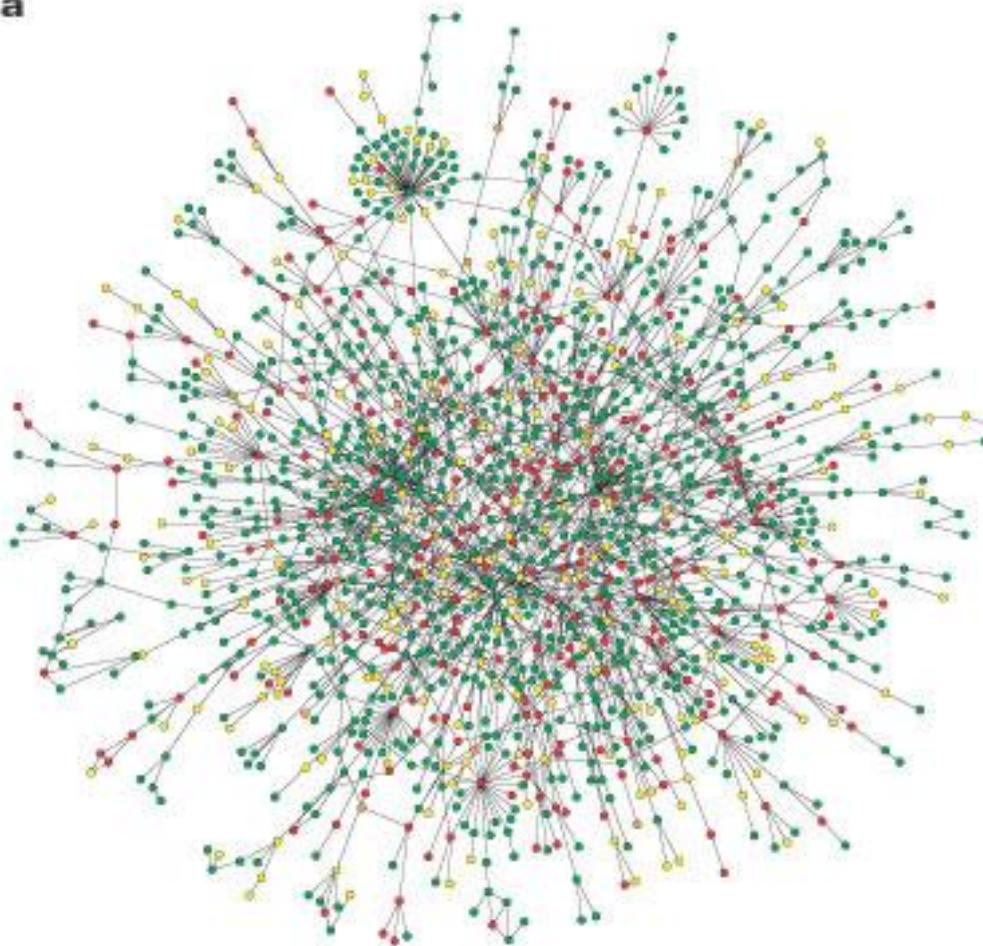
Metabolic network



Nodes are metabolites (compounds) and links are reactions connecting metabolites.

Protein – Protein interaction network

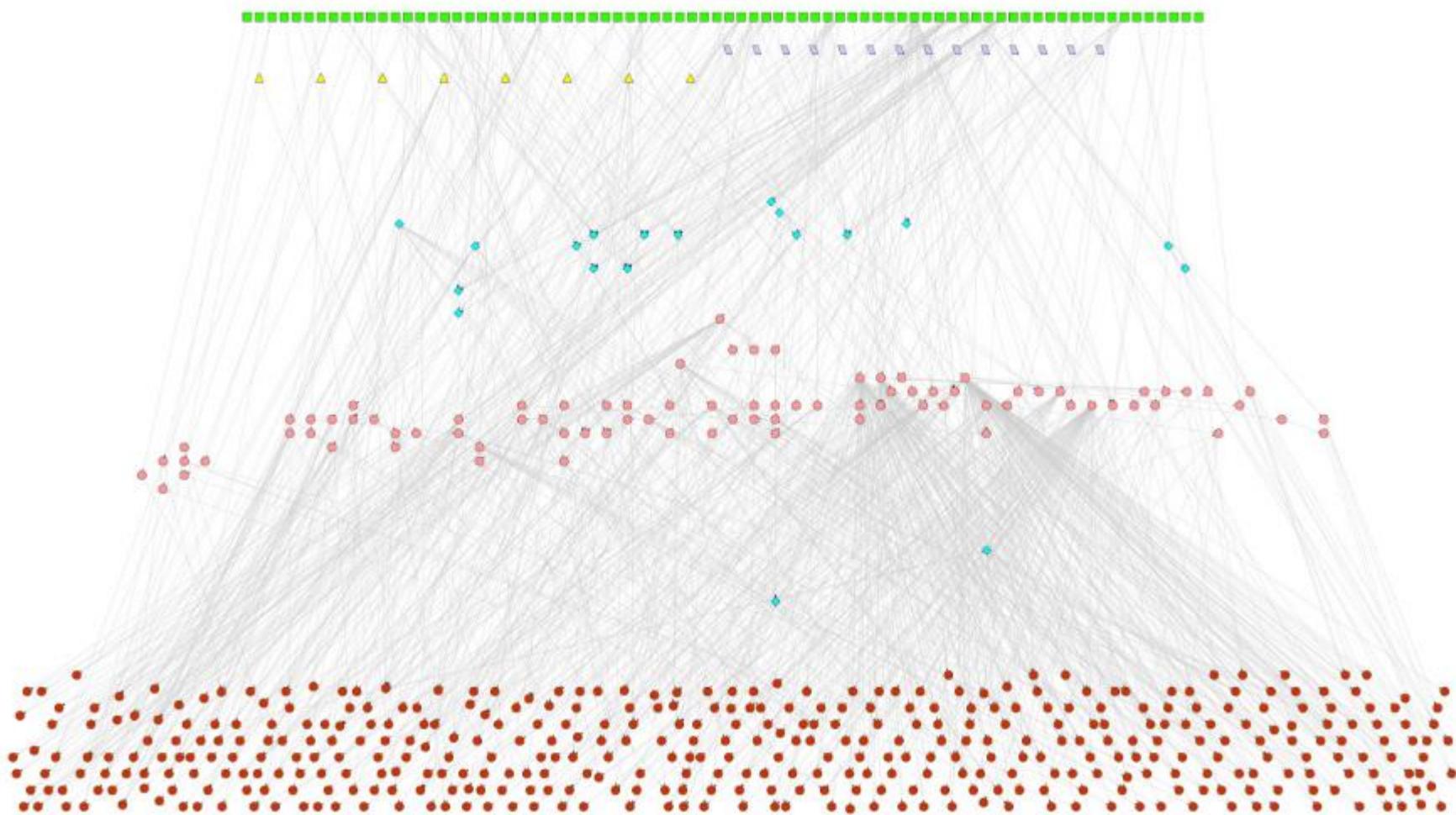
a



Yeast Protein Interaction Network

Nodes are proteins and links are interactions between them

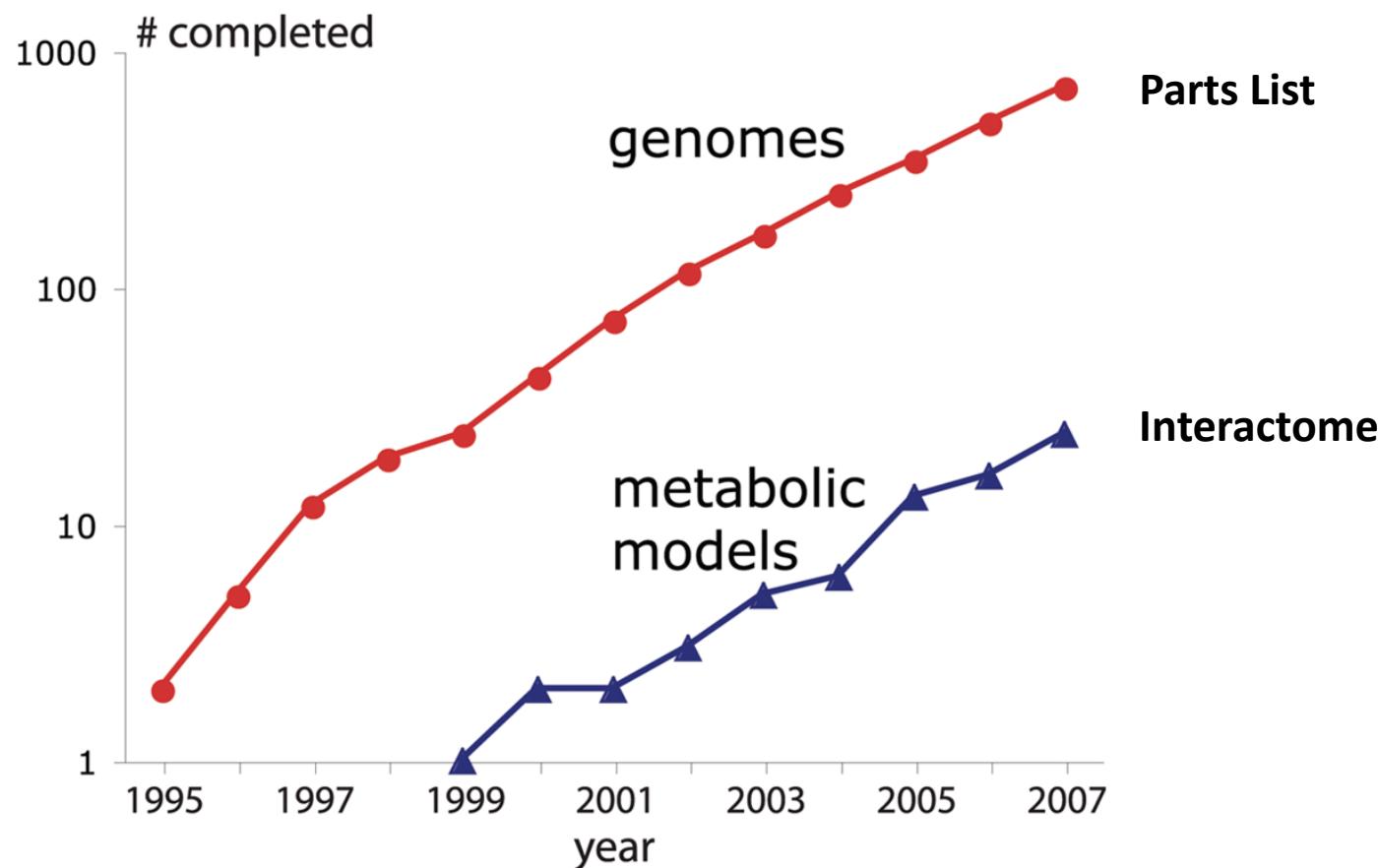
Transcription regulatory network



E. coli transcriptional regulatory network

Nodes are genes or transcription factors and links are regulatory interactions

Mapping of biological networks lags far behind sequencing efforts



Metabolic reconstruction effort versus progress in sequencing effort

Human Interactome - Metabolic network

[Home](#) > Current Issue > vol. 104 no. 6 > Natalie C. Duarte, 1777–1782, doi: 10.1073/pnas.0610772104



Global reconstruction of the human metabolic network based on genomic and bibliomic data

Natalie C. Duarte, Scott A. Becker, Neema Jamshidi, Ines Thiele, Monica L. Mo, Thuy D. Vo, Rohith Srivastava, and Bernhard Ø. Palsson*

Recon 1: 1496 genes and 3743 reactions

A community-driven global reconstruction of human metabolism

Ines Thiele, Neil Swainston, Ronan M T Fleming, Andreas Hoppe, Swagatika Sahoo, Maike K Aurich, Hulda Haraldsdottir, Monica L Mo, Ottar Rolfsson, Miranda D Stobbe, Stefan G Thorleifsson, Rasmus Agren, Christian Bölling, Sergio Bordel, Arvind K Chavali, Paul Dobson, Warwick B Dunn, Lukas Endler, David Hala, Michael Hucka, Duncan Hull, Daniel Jameson, Neema Jamshidi, Jon J Jonsson, Nick Juty + et al.

[Affiliations](#) | [Contributions](#) | [Corresponding author](#)

Nature Biotechnology 31, 419–425 (2013) | doi:10.1038/nbt.2488

Received 07 September 2012 | Accepted 19 December 2012 | Published online 03 March 2013

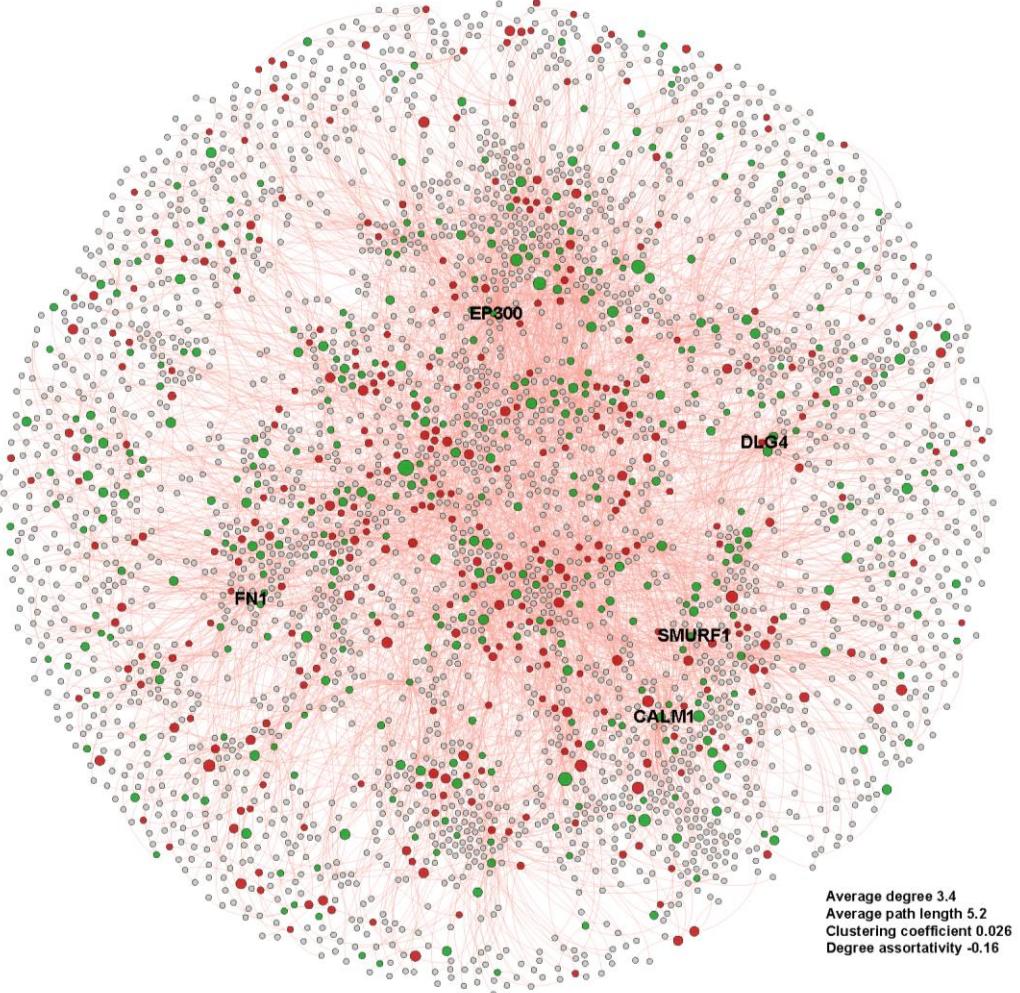
Recon 2: 1789 genes and 7440 reactions

Valuable knowledgebase that can be used to shed insights on metabolic disorders such as Diabetes and Cancer



Work in progress!

Human Interactome – Protein interaction network



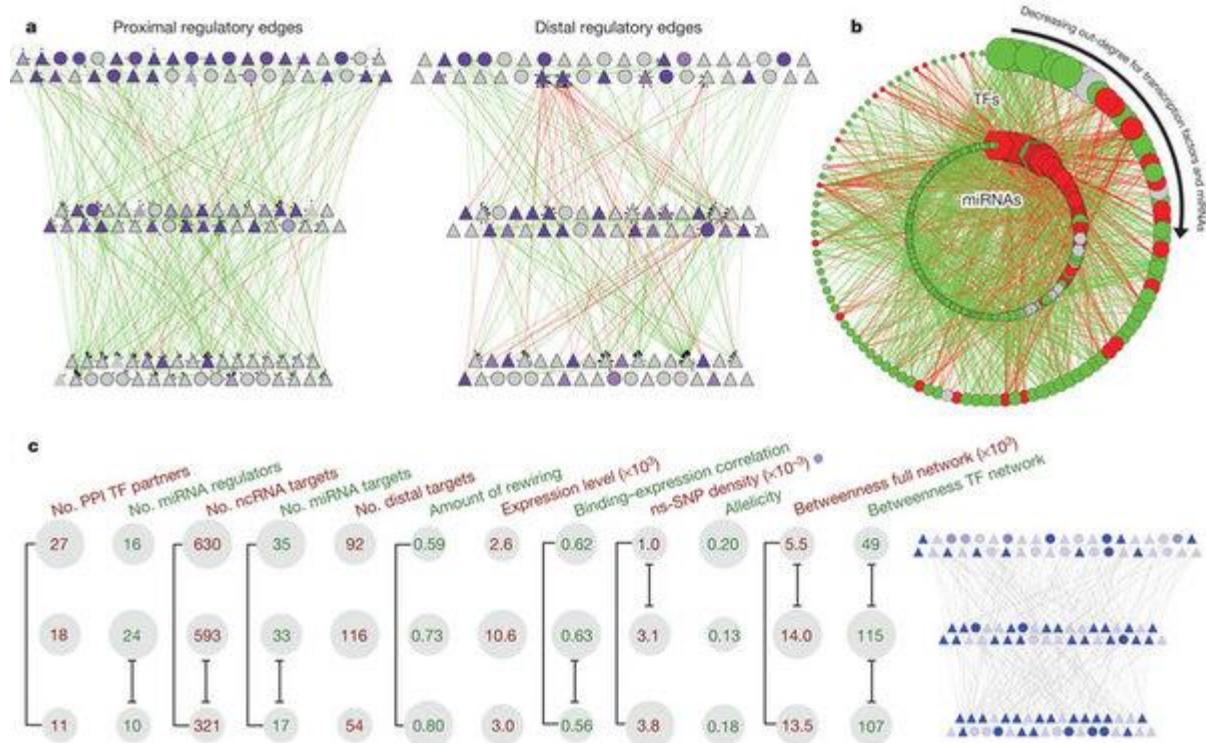
Several databases devoted to maps of protein interaction networks:

- STRING
- HPRD
- BIOGRID

Valuable knowledgebase that can be used to shed insights on diseases

Work in progress!

Human Interactome – Transcriptional Regulatory Network



Architecture of the human regulatory network
derived from ENCODE data

Mark B. Gerstein, Anshul Kundaje, Manoj Hariharan, Stephen G. Landt, Koon-Kiu Yan,
Chao Cheng, Xinmeng Jasmine Mu, Ekta Khurana, Joel Rozowsky, Roger Alexander,
Renqiang Min, Pedro Alves, Alexej Abzyov, Nick Addleman, Nitin Bhardwaj, Alan P. Boyle,
Philip Cayting, Alexandra Charos, David Z. Chen, Yong Cheng, Declan Clarke, Catharine
Eastman, Ghia Euskirchen, Seth Frietze, Yao Fu + et al.

[Affiliations](#) | [Contributions](#) | [Corresponding authors](#)

Nature 489, 91–100 (06 September 2012) | doi:10.1038/nature11245

Received 09 December 2011 | Accepted 22 May 2012 | Published online 05 September 2012

Work in progress!

Size of the Interactome of a Human Cell

- Estimates based on the human genome sequence:
 - ~ 25000 proteins
 - Several thousand RNAs
 - Several thousand metabolites
- Estimated number of different components in a human cell is ~ 100000 or 10^5
- Number of possible pair-wise interactions between different components of a human cell $\sim 10^{10}$
- Even if 10^{-4} of the possible interactions are realized, the interactome in a human cell will contain 10^7 interactions

Computationally challenging to model the human cellular interactome!

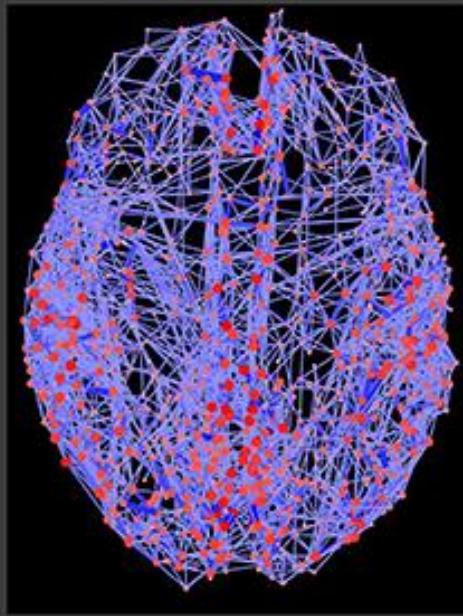
Map of the Human Brain

The Human Connectome



Anatomy

Klingler's method for fiber tract dissection uses freezing of brain matter to spread nerve fibers apart. Afterwards, tissue is carefully scratched away to reveal a relief-like surface in which the desired nerve tracts are naturally surrounded by their anatomical brain areas.



Connectome

Shown are the connections of brain regions together with "hubs" that connect signals among different brain areas and a central "core" or backbone of connections, which relays commands for our thoughts and behaviors.



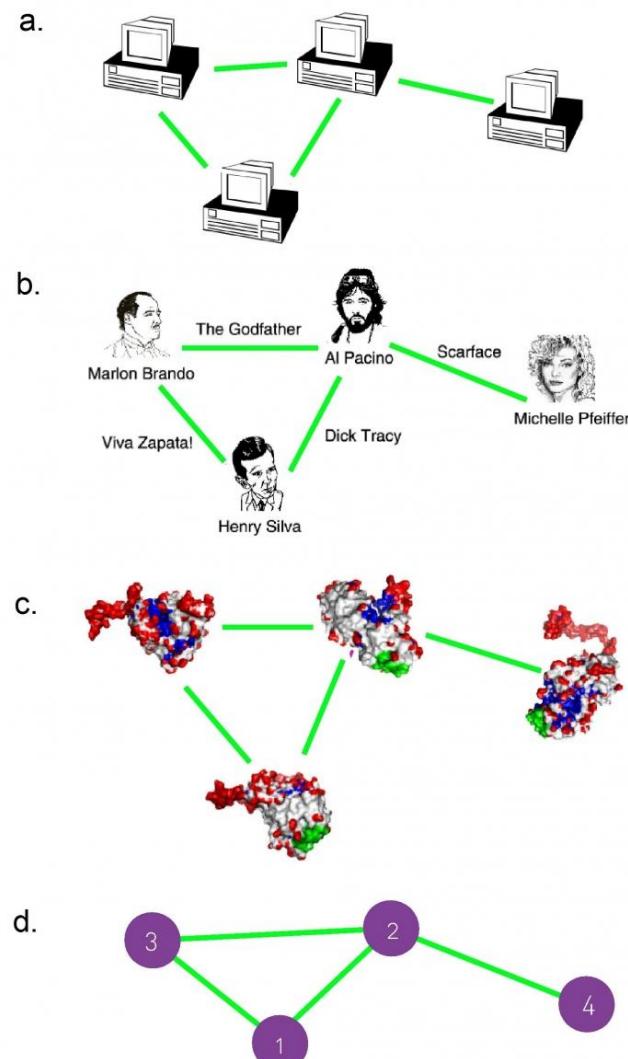
Neuronal Pathways

A new MRI technique called diffusion spectrum imaging (DSI) analyzes how water molecules move along nerve fibers. DSI can show a brain's major neuron pathways and will help neurologists relate structure to function.

Graph theory: Mathematical language to study networks

Graphs

- Graph theory has been a branch of mathematics for more than 200 years.
- A graph or network is a set of nodes (or vertices) connected by links (or edges).
- Size of network is given by number of nodes N and number of links L .



Adjacency Matrix

The adjacency matrix of a network of N nodes has N rows and N columns, its elements being:

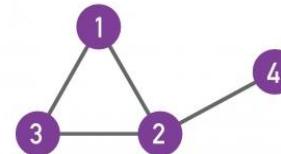
- $A_{ij} = 1$ if there is a link pointing from node j to node i
- $A_{ij} = 0$ if nodes i and j are not connected to each other

Adjacency matrix of an undirected network is symmetric!

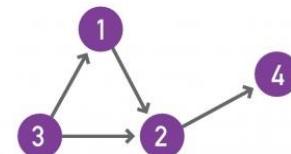
a. Adjacency matrix

$$A_{ij} = \begin{matrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{matrix}$$

b. Undirected network



c. Directed network

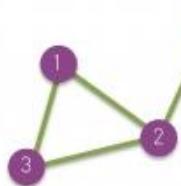


$$A_{ij} = \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$

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

Different types of graphs

a. Undirected

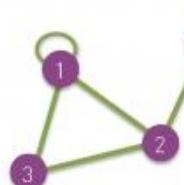


$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \quad A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1}^N A_{ij} \quad \langle k \rangle = \frac{2L}{N}$$

b. Self-loops

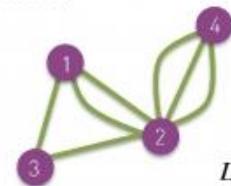


$$A_{ij} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$$\exists i, A_{ii} \neq 0 \quad A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1, i \neq j}^N A_{ij} + \sum_{i=1}^N A_{ii} \quad ?$$

c. Multigraph
(undirected)

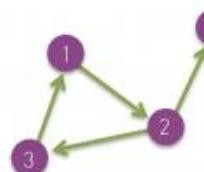


$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \quad A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1}^N A_{ij} \quad \langle k \rangle = \frac{2L}{N}$$

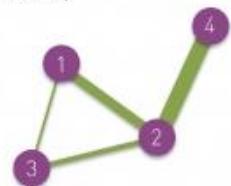
d. Directed



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

$$A_{ij} \neq A_{ji} \quad L = \sum_{i,j=1}^N A_{ij} \quad \langle k \rangle = \frac{L}{N}$$

e. Weighted
(undirected)

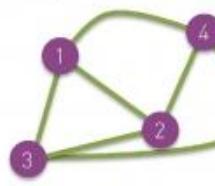


$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \quad A_{ij} = A_{ji}$$

$$\langle k \rangle = \frac{2L}{N}$$

f. Complete Graph
(undirected)



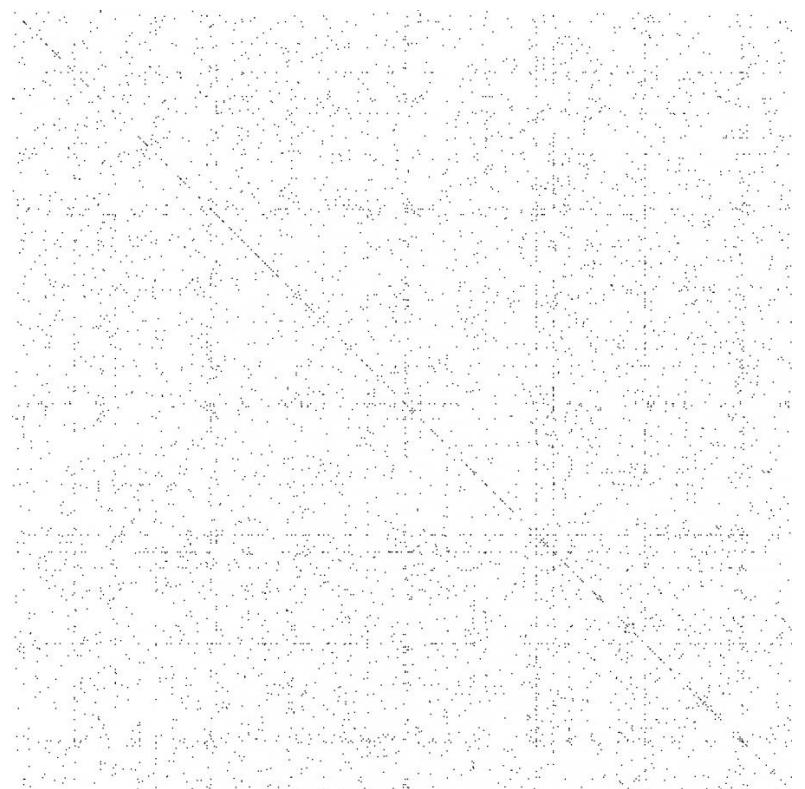
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \quad A_{i \neq j} = 1$$

$$L = L_{\max} = \frac{N(N-1)}{2} \quad \langle k \rangle = N-1$$

Real networks are sparse

a



Adjacency matrix of Yeast protein-protein interaction network

Birth of graph theory

- Euler in 1736 was able to solve the Königsberg bridge problem using graph theory
- Contemporary puzzle: Could a pedestrian walk around Königsberg crossing each bridge once?
- Euler cast the puzzle as a graph theory problem where nodes are separate land masses and links are bridges connecting land masses.
- Graph associated with Königsberg is a multi-graph and Euler used the graph to show that such a path is not possible.
- In 1875, a new bridge was created in Königsberg to make such a path possible.

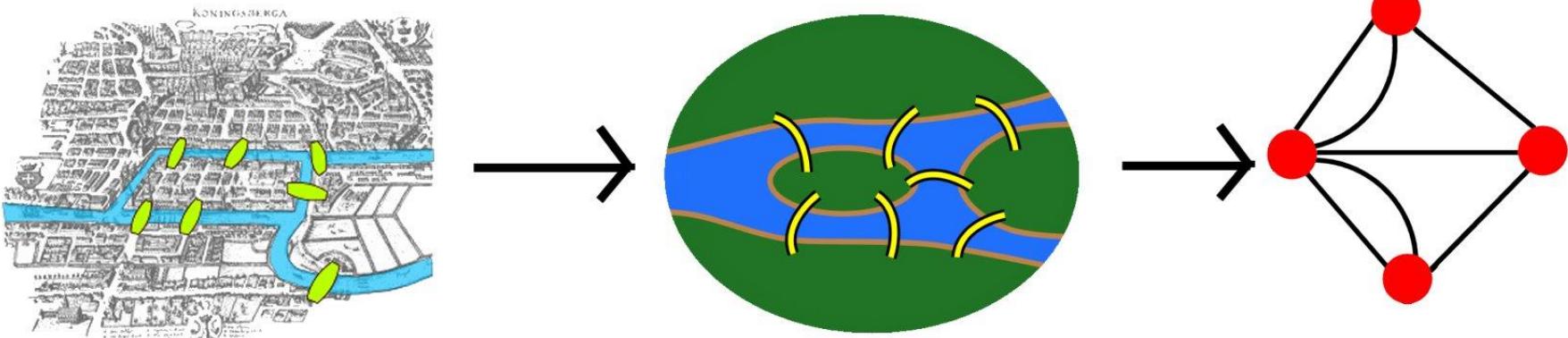
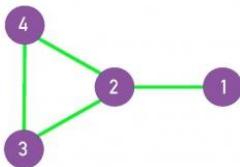


Image credit: <https://simanaitissays.files.wordpress.com/2014/08/progression.jpg>

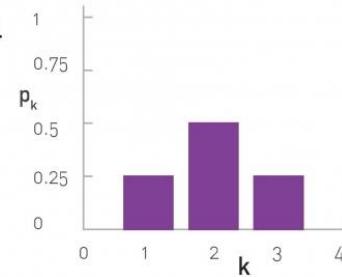
Degree and Degree distribution

- A key property of each node is its *degree*, representing the number of links it has to other nodes.
- Average degree of a network is the mean degree across nodes of a network.
- Degree distribution p_k provides the probability that a randomly selected node in the network has degree k .

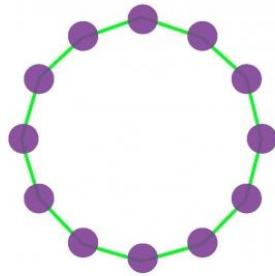
a.



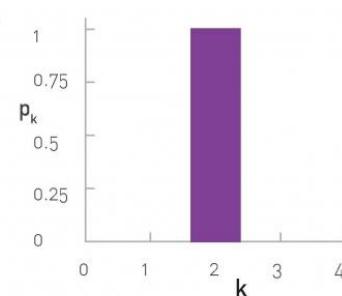
b.



c.



d.



Paths and distances

Path

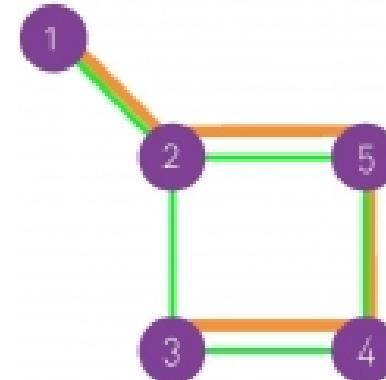
A sequence of nodes such that each node is connected to the next node along the path by a link. Each path consists of $n+1$ nodes and n links. The length of a path is the number of its links, counting multiple links multiple times.

For example, the $1 \rightarrow 2 \rightarrow 5 \rightarrow 4 \rightarrow 3$ covers a path of length four.

Shortest Path (Geodesic Path, d)

The path with the shortest distance d between two nodes. We also call d the distance between two nodes.

Note that the shortest path does not need to be unique: between nodes 1 and 4 we have two shortest paths, $1 \rightarrow 2 \rightarrow 3 \rightarrow 4$ and $1 \rightarrow 2 \rightarrow 5 \rightarrow 4$, having the same length $d_{1,4} = 3$.



Diameter (d_{max})

The longest shortest path in a graph, or the distance between the two furthest nodes. In the graph shown here the diameter is between nodes 1 and 4, hence $d_{max}=3$.

Average Path Length ($\langle d \rangle$)

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

For the graph shown on the left we have $\langle d \rangle=1.6$.

Paths and distances

Cycle

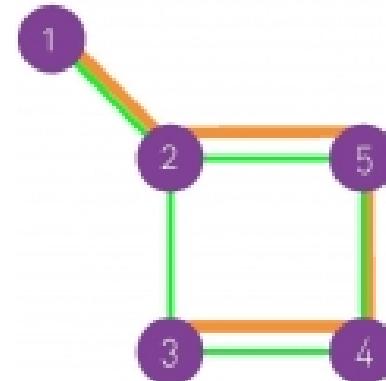
A path with the same start and end node. In the graph shown on the left we have only one cycle.

Eulerian Path

A path that traverses each link exactly once.

Hamiltonian Path

A path that visits each node exactly once.



Paths based on Adjacency matrix

The number of shortest paths, N_{ij} , and the distance d_{ij} between nodes i and j can be calculated directly from the adjacency matrix A_{ij} .

$d_{ij} = 1$: If there is a direct link between i and j, then $A_{ij} = 1$ ($A_{ij} = 0$ otherwise).

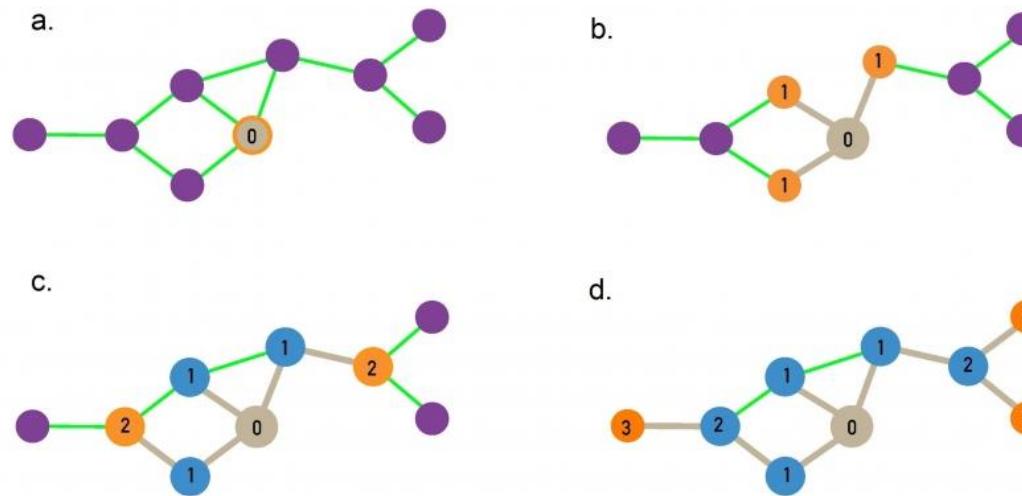
$d_{ij} = 2$: If there is a path of length two between i and j, then $A_{ik}A_{kj}=1$ ($A_{ik}A_{kj}=0$ otherwise).

One can determine shortest paths by taking the product of adjacency matrices.

Breadth-First-Search (BFS) algorithm

The identification of the shortest path between node i and j follows the following steps:

1. Start at node i , that we label with “0”.
2. Find the nodes directly linked to i . Label them distance “1” and put them in a queue.
3. Take the first node, labeled n , out of the queue ($n = 1$ in the first step). Find the unlabeled nodes adjacent to it in the graph. Label them with $n + 1$ and put them in the queue.
4. Repeat step 3 until you find the target node j or there are no more nodes in the queue.
5. The distance between i and j is the label of j . If j does not have a label, then $d_{ij} = \infty$.



Clustering coefficient

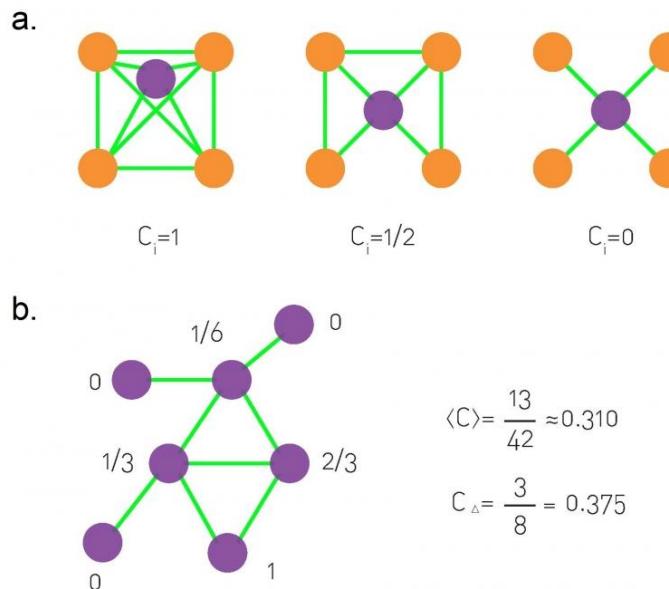
In many networks, if node 1 has links to nodes 2 and 3, then it is highly probable that 2 and 3 also have a link between each other. E.g. Friendship networks

This phenomenon can be quantified using the **clustering coefficient**

$$C_I = \frac{2n_I}{k_I(k_I-1)},$$

where n_I is the number of mutual links between the k_I neighbors of node I in the graph, while $k_I(k_I-1)/2$ is the maximum number possible of such links.

The **average clustering coefficient**, $\langle C \rangle$, characterizes the overall tendency of nodes to form clusters or groups.

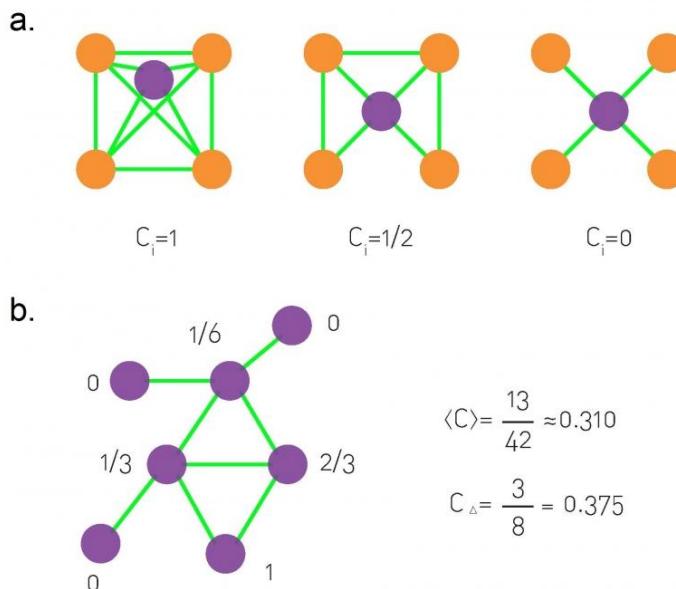


Global clustering coefficient

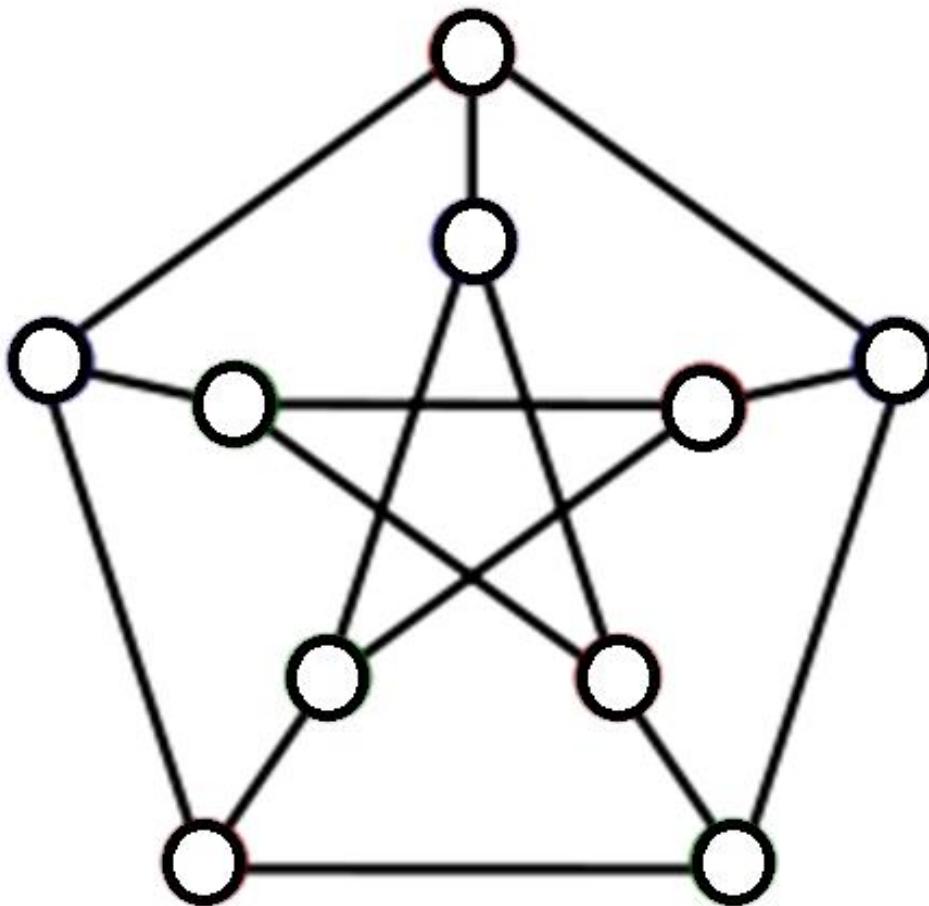
The *global clustering coefficient* is defined as follows:

$$C_{\Delta} = \frac{3 \times \text{Number Of Triangles}}{\text{Number Of Connected Triples}}$$

where a *connected triplet* is an ordered set of three nodes ABC such that A connects to B and B connects to C. For example, an A, B, C triangle is made of three triplets, ABC, BCA and CAB.



Clustering coefficient of Petersen Graph



Clustering coefficient of each node in this graph is 0, and thus, this may not be a good benchmark to compare against real networks.

Network science versus graph theory

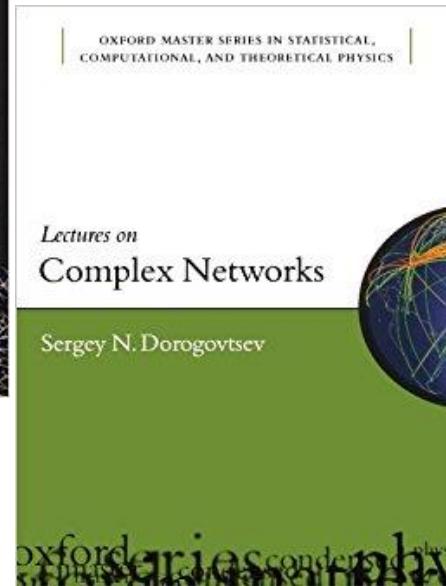
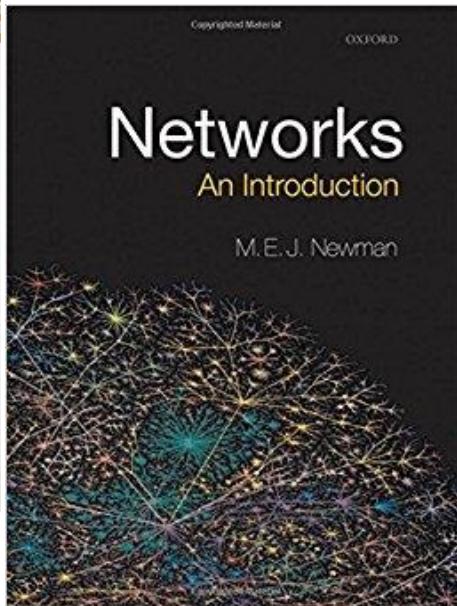
- Interdisciplinary nature
- Empirical and data driven
- Quantitative in nature
- Computational nature

Note that physicists have studied simple networks such as regular lattices for many years while graph theorists have studied model networks for more than 200 years.

Reference Books



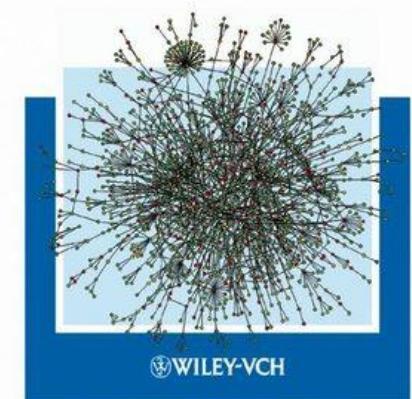
Albert-László Barabási
**NETWORK
SCIENCE**



[Stefan Bornholdt, Heinz Georg Schuster \(Eds.\)](#)

**Handbook of
Graphs and Networks**

[From the Genome to the Internet](#)



WILEY-VCH