Bioinformática y Tratamiento de Datos
(BIF) 2025

Tema 5

Redes Biológicas y Teoría de Grafos

Florencio Pazos (CNB-CSIC)

Florencio Pazos Cabaleiro
Computational Systems Biology Group (CNB-CSIC)
pazos@cnb.csic.es
@fpazos_bioinf
http://csbg.cnb.csic.es

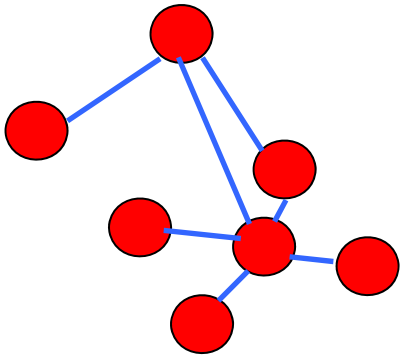
Biological Networks and Graph Theory

- Networks
- Biological and molecular networks
- Short intro to Graph Theory
- Characteristics of the main molecular networks
- Network applications to human pathologies

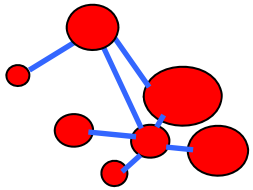
- *Cytoscape practical*

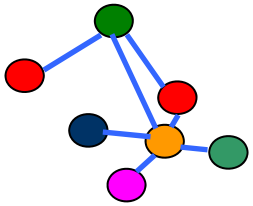
Network representation of a phenomenon

Some phenomena/data can be represented as entities ("nodes"/"vertices") linked by relationships ("edges") → Network

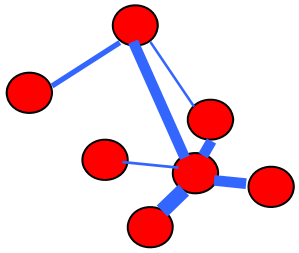


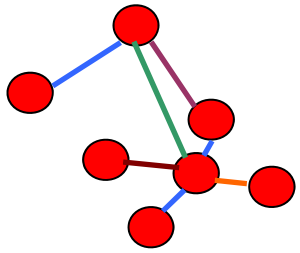
Node – Generic entity, physical or not.
(gene, protein, metabolite, cellular state, disease, …)
It can have associated features (quantitative or qualitative)

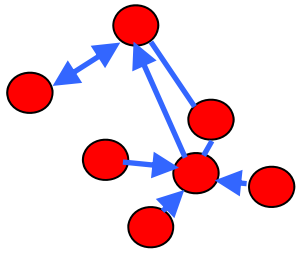




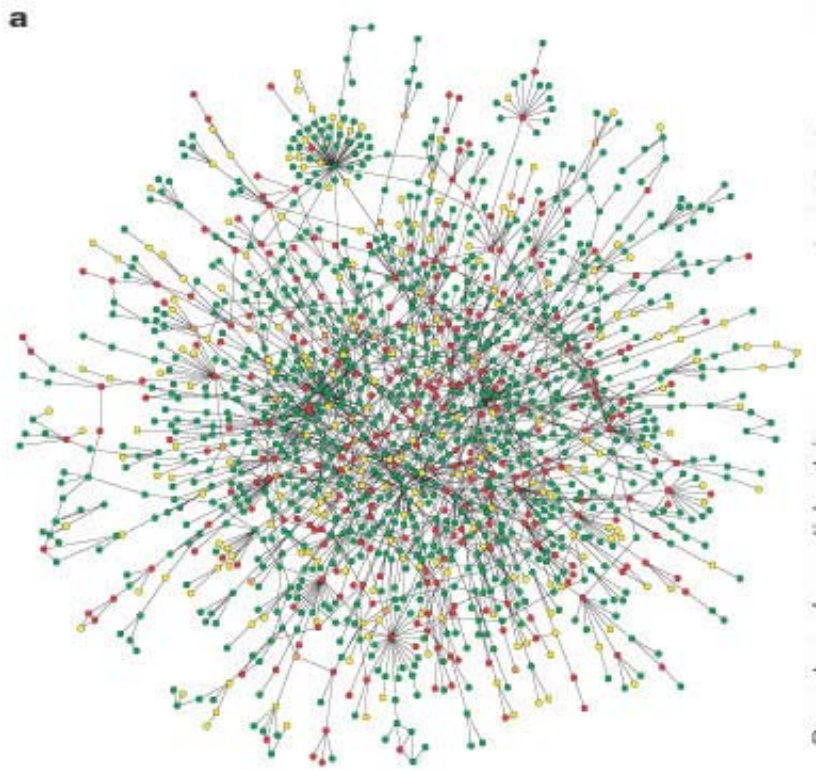
Edge – Generic relationship, in the broadest sense.
(interaction (physical or functional), transcriptional control, chemical transformation (reaction), …)
It can have associated features (quantitative or qualitative): weight, direction, ..







Large networks
Network theory / Graph theory

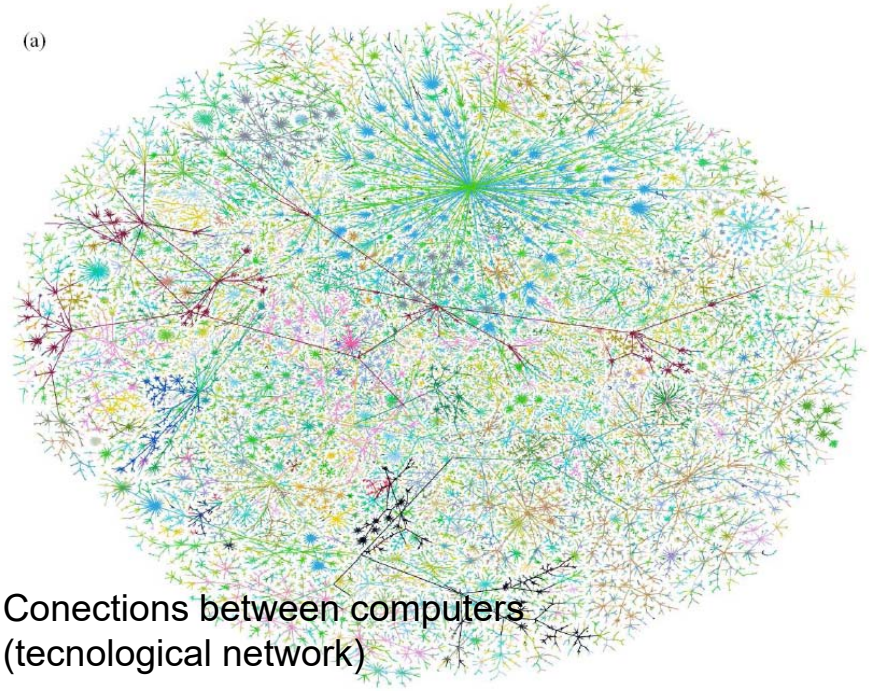


Once a given phenomenon is modelled as a **large** network, it can be studied using mathematical approaches (Graph Theory) in order to extract information hidden in its structure and topological patterns

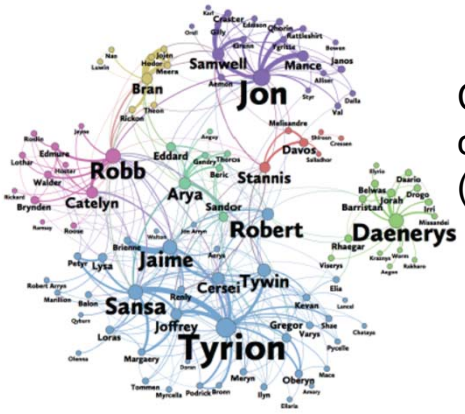
•Gavin, A.C., *et al.* (2002) Functional organisation of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**, 141-147.

Networks

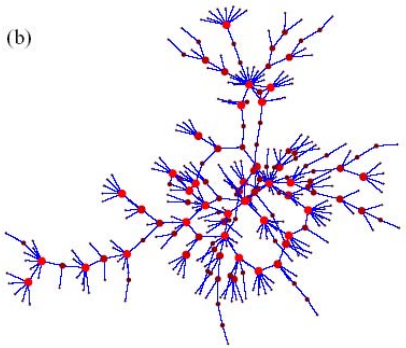
Networks have been used to model phenomena in all scientific disciplines



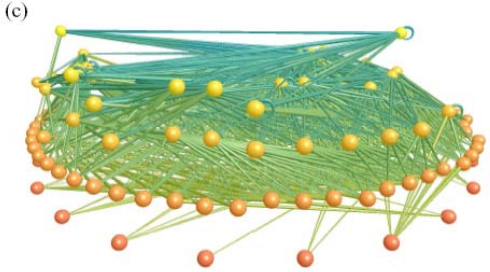
Conections between computers
(tecnological network)



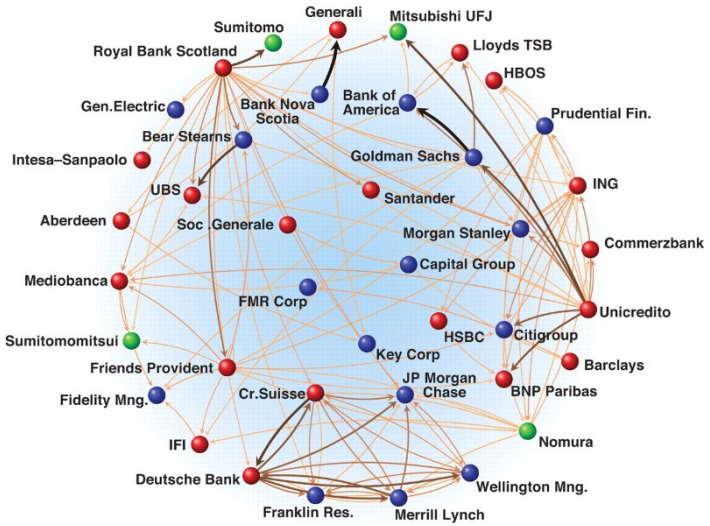
Co-mentioning of characters in a novel
(literature network)



Sexual contacts
(social network)



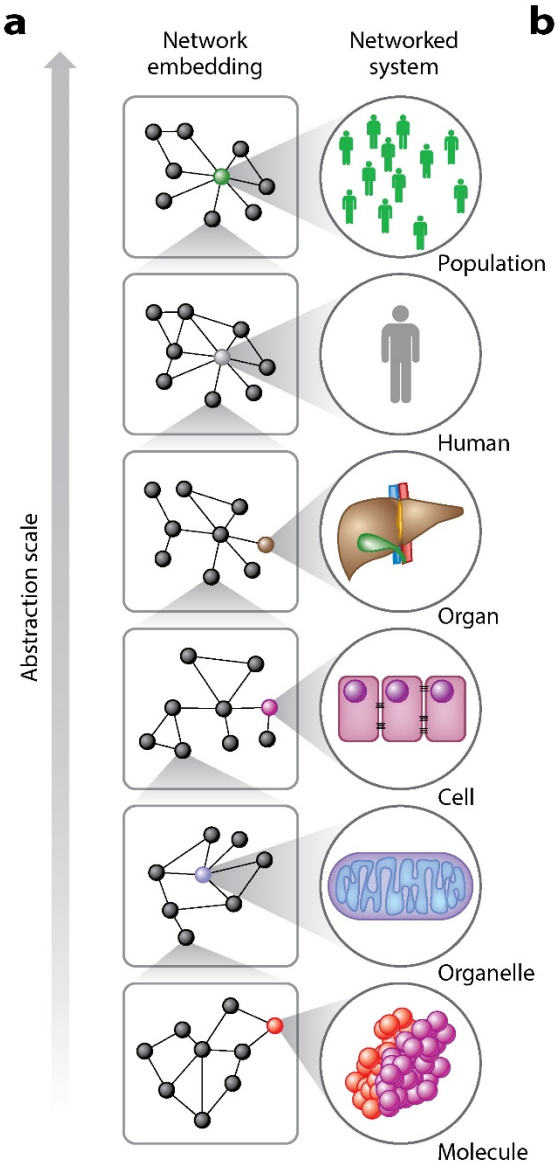
Predator-prey relationships
(food web)
(biological/ecological network)



Relationships between banks
(economic network)

M. E. J. Newman. (2003). The Structure and Function of Complex Networks. *SIAM Review*. **45:**2, 167-256

Biological networks



In Biology, networks have been used to model diverse phenomena at all biological levels

P McGillivray, D Clarke, W Meyerson, J Zhang, D Lee, M Gu, S Kumar, H Zhou, MB Gerstein (2018). Network Analysis as a Grand Unifier in Biomedical Data Science. *Annual Review of Biomedical Data Science* Vol. 1. 153-180

Molecular networks

Network approaches have been applied to the study of molecular phenomena as the data required to assemble these networks became available **(-omics techniques**).

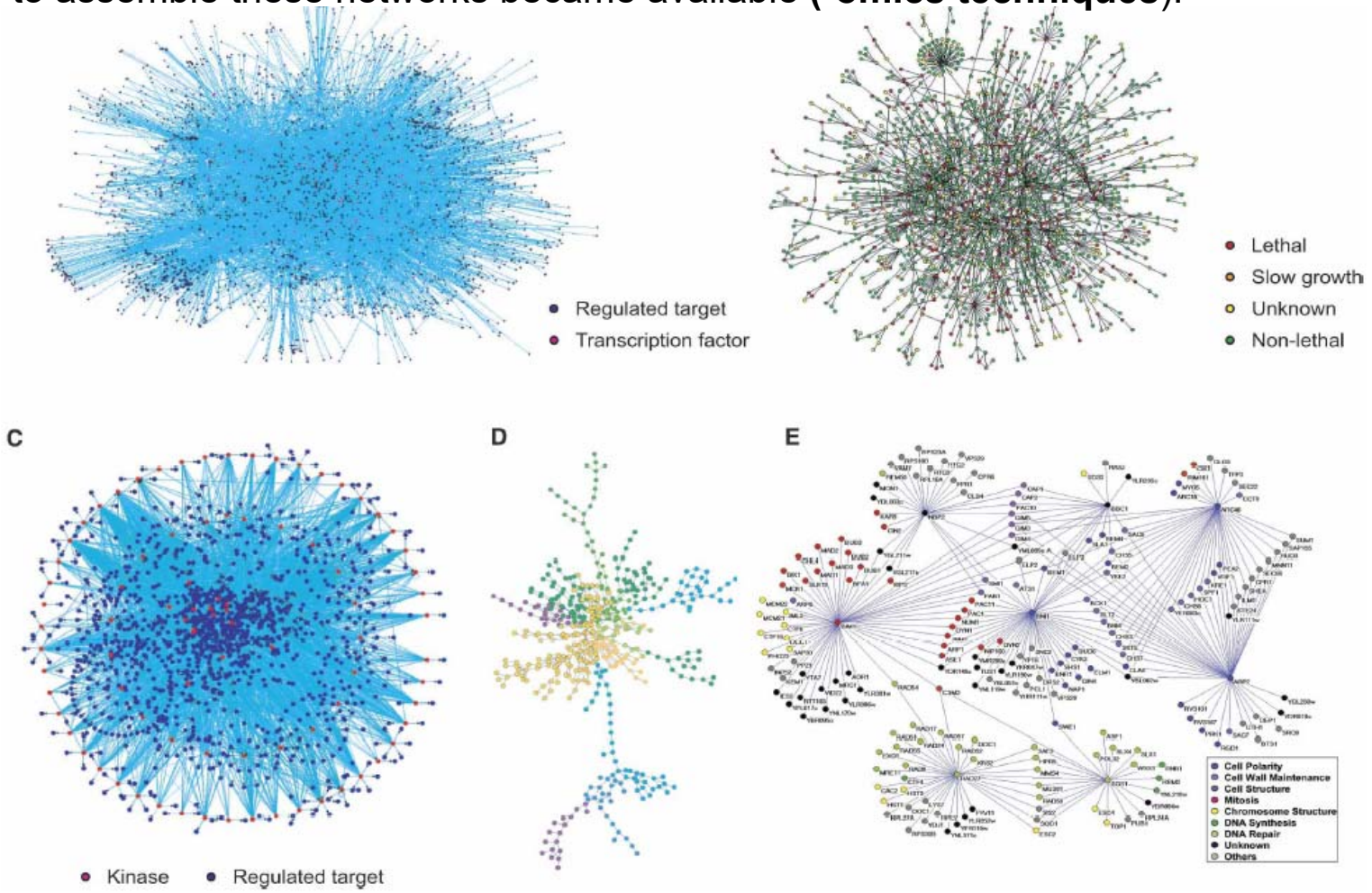
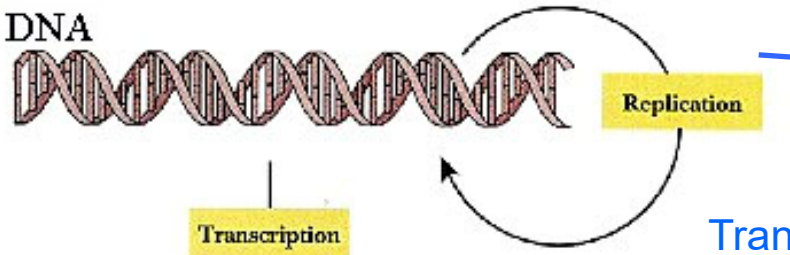
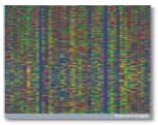


Figure 1. Examples of the five major biological networks. (*A*) A yeast transcription factor-binding network, composed of known transcription factor-binding data collected with large-scale ChIP–chip and small-scale experiments. This figure was generated with the program Pajek (de Nooy et al. 2005). (*B*) A yeast protein–protein interaction network, containing protein–protein interactions identified by yeast two-hybrid and protein complexes identified by affinity purification and mass spectrometry (Barabasi and Bonabeau 2003). (Reprinted by permission from Macmillan Publishers Ltd: *Nature* [Jeong et al. 2001], © 2001.) Nodes are colored according to the mutant phenotype. (*C*) A yeast phosphorylation network comprised primarily of in vitro phosphorylation events identified using protein microarrays (Ptacek et al. 2005). The figure was generated with Osprey 1.2.0. (Breitkreutz et al. 2003). (*D*) An *E. coli* metabolic network with 574 reactions and 473 metabolites colored according to their modules (Reprinted by permission from Macmillan Publications Ltd: *Nature* [Guimera and Nunes Amaral 2005], © 2005). (*E*) A yeast genetic network constructed with synthetic lethal interactions using SGA analysis on eight yeast genes (From Tong et al. 2001; reprinted with permission from AAAS). Nodes are colored according to their YPD cellular roles.

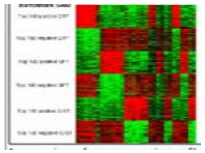
-omics and molecular networks in the context of the Central Dogma



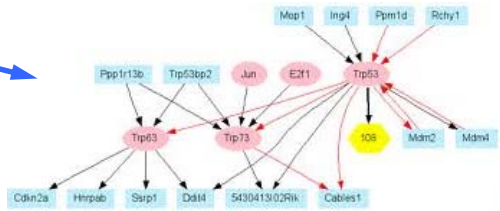
Genomics



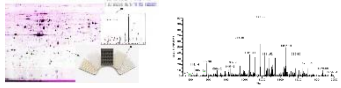
Transcriptomics



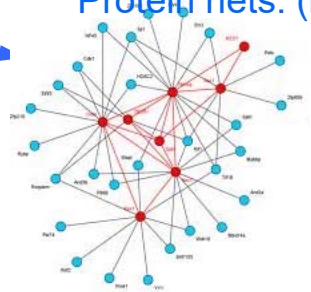
Gene regulation nets.

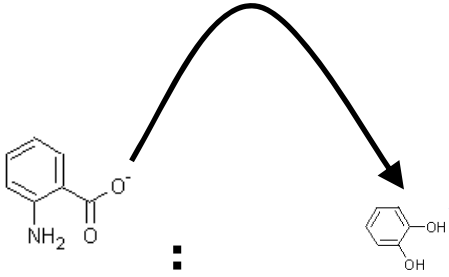


Proteomics

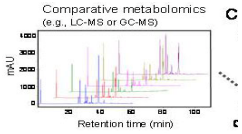


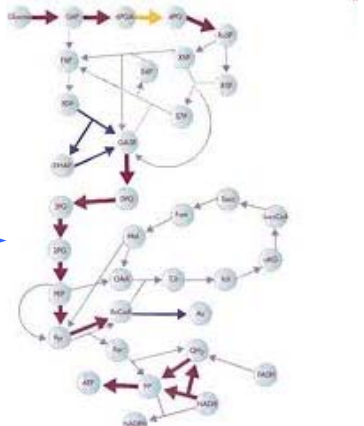
Protein nets. (interaction, fosforilation, …)



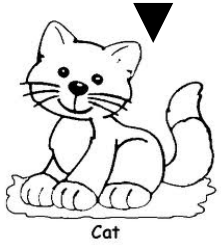


Metabolomics

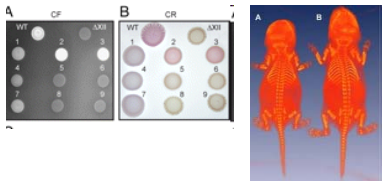


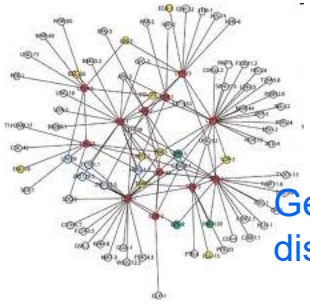


Metabolic nets.



Phenomics





Genetic nets, phenotipic nets, disease-related nets, …

Systems Biology

Systems biology is the study of biological systems whose behaviour cannot be reduced to the linear sum of their parts'. It is a biology-based interdisciplinary field of study that **focuses on complex interactions** within biological systems, using a holistic approach (holism instead of the more traditional reductionism) to biological research.

(nature.com, Wikipedia)

Is the reductionist approach of molecular biology enough?

Biological systems are a prototype of **complex systems**: systems characterized by a large number of components interacting non-linearly. They can not be (fully) modelled by first principles: a minimum difference in de details leads to a totally different outcome.

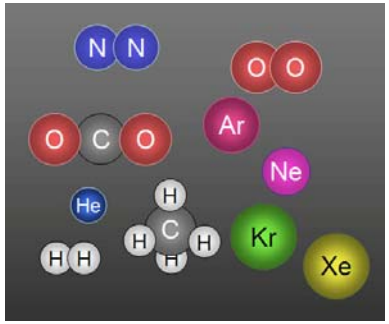
"The whole is more than the sum of the parts"

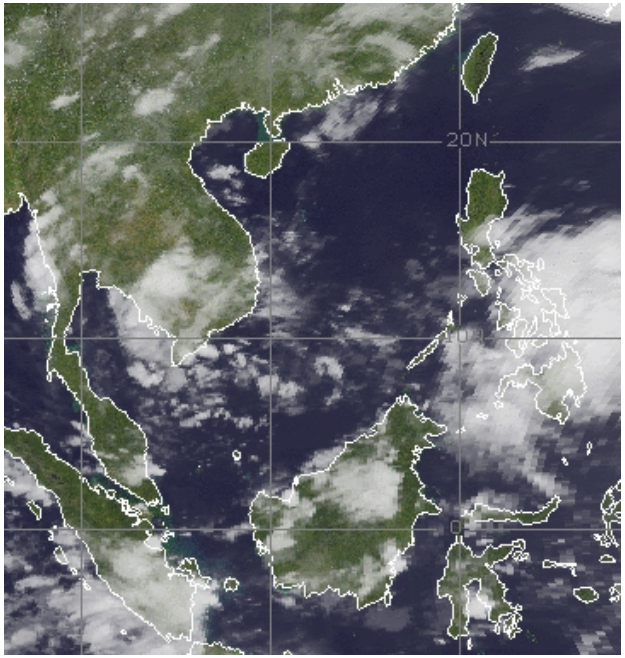
The main idea behind complex systems is that ***the ensemble behaves in ways not predicted by its components.*** *The* ***interactions*** ***matter more than the nature of the units****.*
Studying individual ants will almost never give us a clear indication of how the ant colony operates. For that, one needs to understand an ant colony as an ant colony, no less, no more, not a collection of ants. This is called an "emergent" property of the whole, by which parts and whole differ because what matters are the interactions between such parts. And interactions can obey very simple rules.

N. N. Taleb.
Skin in the game - Hidden asymmetries in daily life.

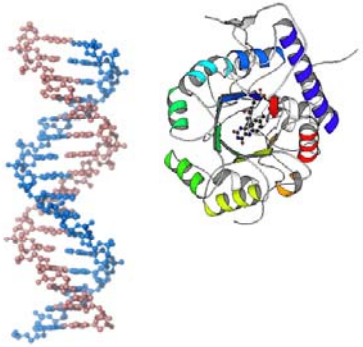
Reductionist vs. systemic approaches in other scientific areas

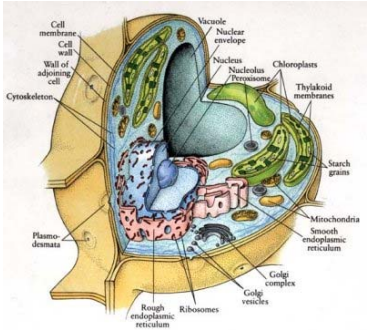
Σ





Σ





Graph

- A graph is a mathematical object used to represent entities and relationships between them (understanding entity and relationship in the broadest possible way)

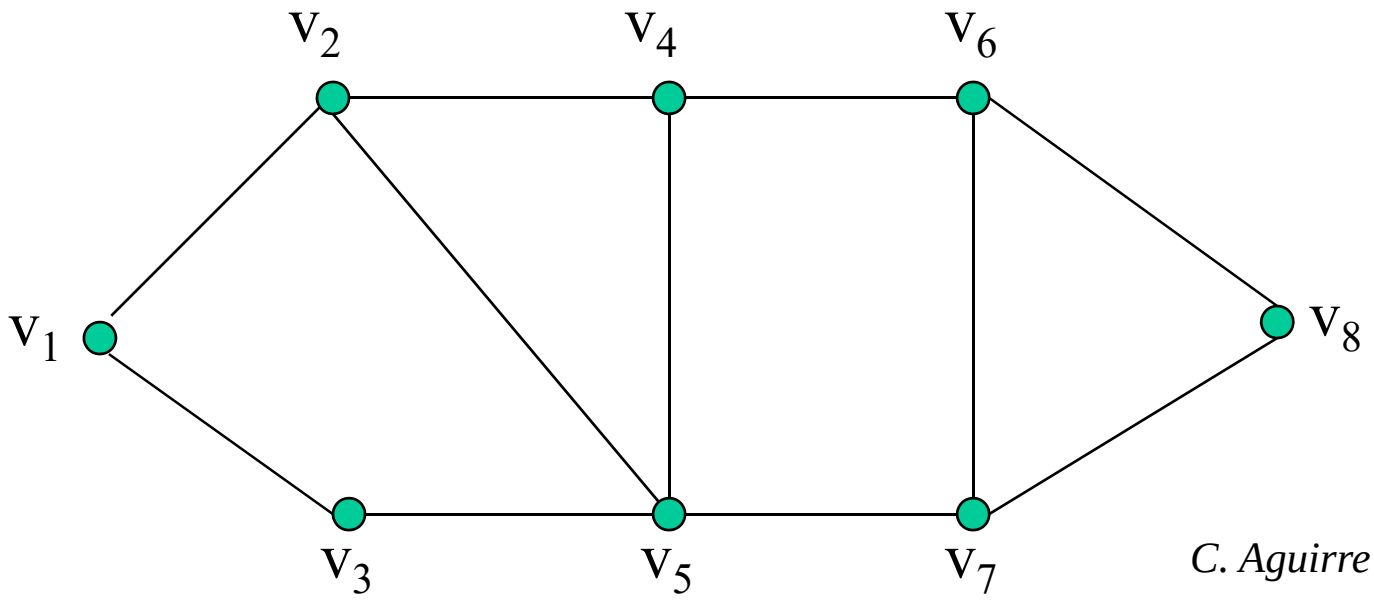
- Formally, graph G is a pair of sets (V,E)

  - $V=\{v_1,v_2,....v_n\}$ is the set of vertices/nodes
  - $E=\{(v_i,v_j),(v_{i'},v_{j'})\ldots\ldots\}$ is a set of edges (pairs formed by the elements in V).

  - The number of nodes is called **order** of the graph.
  - The number of edges is called **size** of the graph.

Graph

Example of a graph (order: 8 , size: 11).



$V=\{v_1,v_2,v_3,v_4,v_5,v_6,v_7,v_8\}$

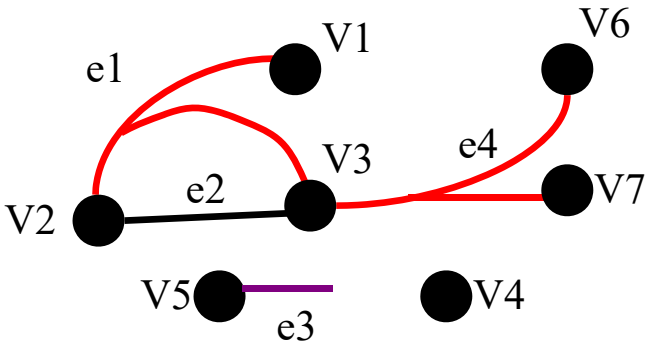
$E=\{(v_1,v_2),(v_1,v_3),(v_2,v_4),(v_3,v_5),(v_4,v_6),$
$(v_5,v_7),(v_6,v_8),(v_7,v_8),(v_2,v_5),(v_4,v_5),(v_6,v_7)\}$

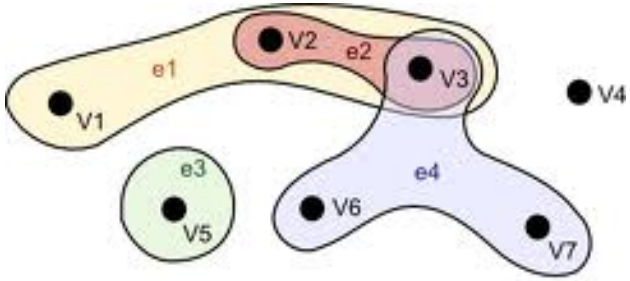
Hypergraph

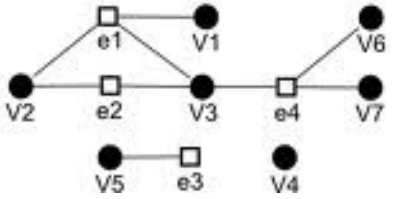
An **hyperedge** is an edge connecting **more** (or **less**) than two nodes. Graphs with hyperedges are called ***hypergraphs***

V={v1,v2,v3,v4,v5,v6,v7}

E={e1,e2,e3,e4}
 ={ {v1,v2,v3}, {v2,v3}, {v5}, {v3,v6,v7} }

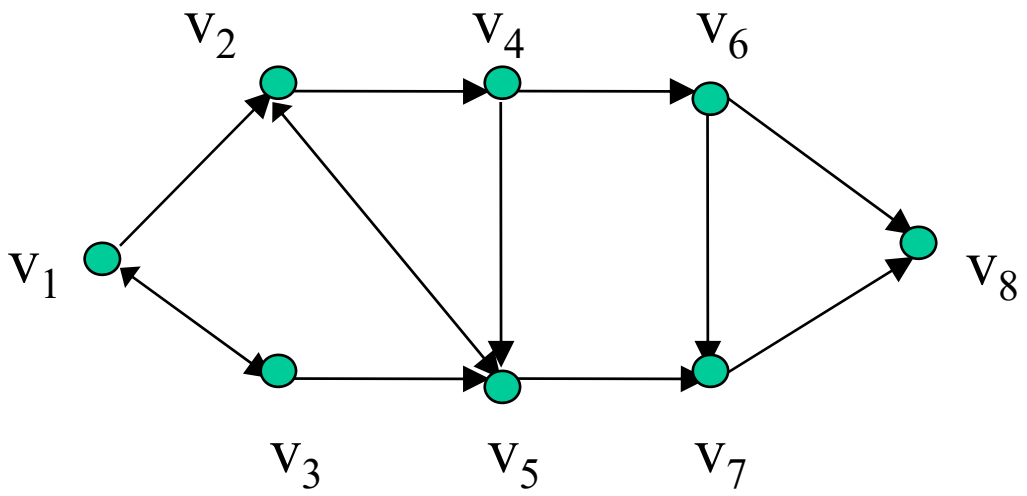






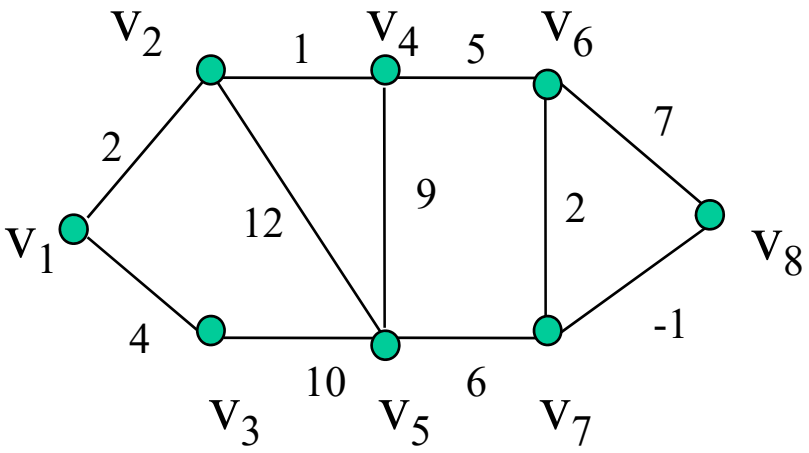
Directed graph

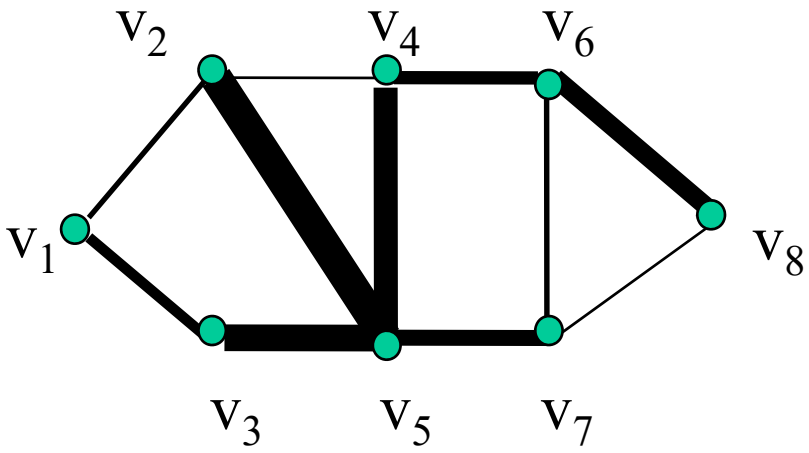
A directed graph (or ***digraph***) is a graph in which de edges have directionality: $(vi,vj) \neq (vj,vi)$



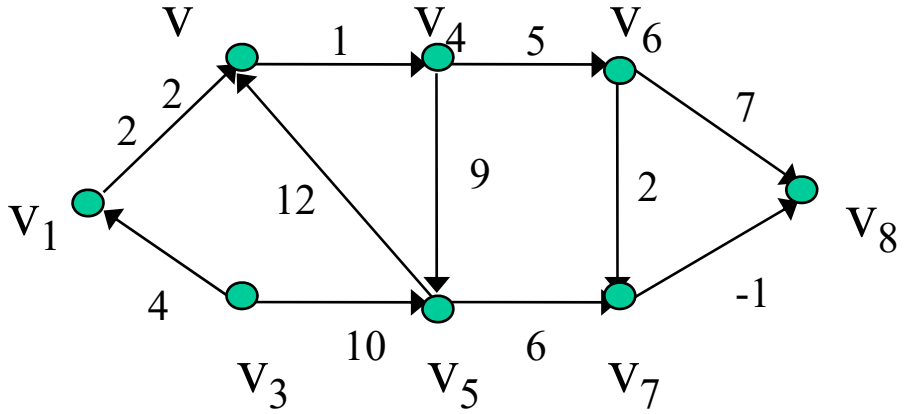
Weighted graph

A *weighted graph* is a graph in which the edges have associated numerical values (to quantify some characteristic of the relationship (importance, distance, capacity …)





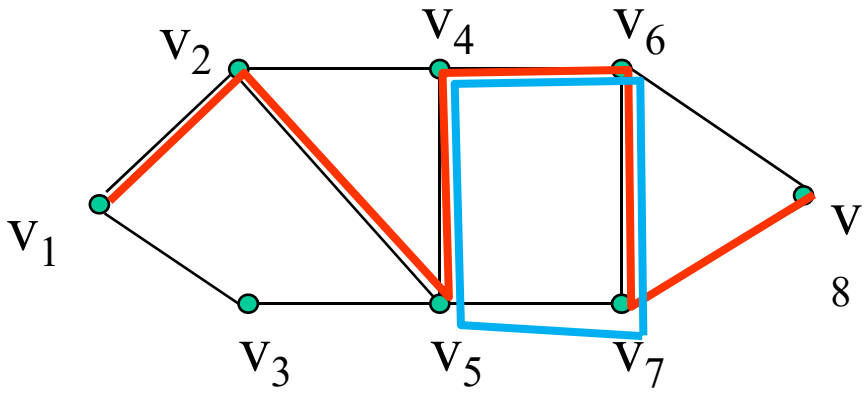
A *weighted directed graph…*



Path

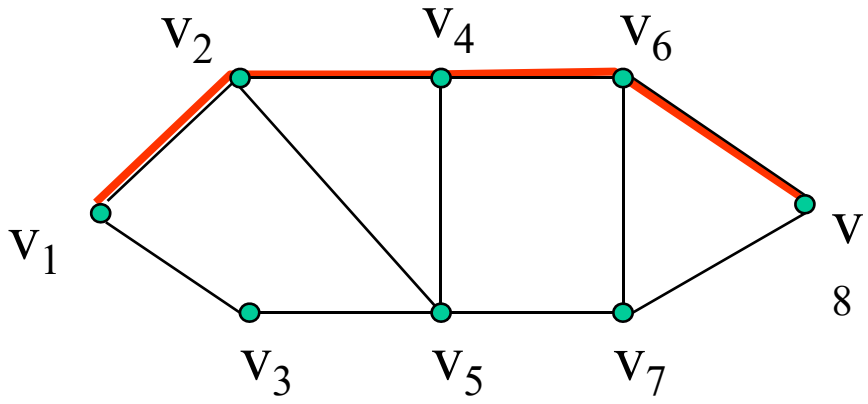
A **path** is a alternating sequence of nodes and edges connected in the network in which all nodes (and consequently all edges) are distinct.

A **closed path** (aka **cycle**) is a closed sequence of nodes and edges in which all nodes (and consequently all edges) are distinct.



$C=\{v_1,v_2,v_5,v_4,v_6,v_7,v_8\}$
$k=6$

The **shortest path** between two nodes…



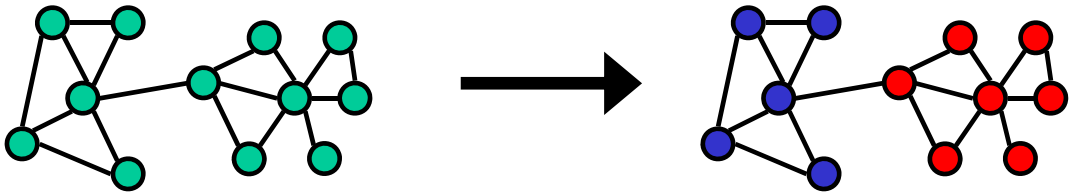
$C=\{v_1,v_2,v_4,v_6,v_8\}$
$k=4$

Distance between two nodes = length of shortest path connecting them

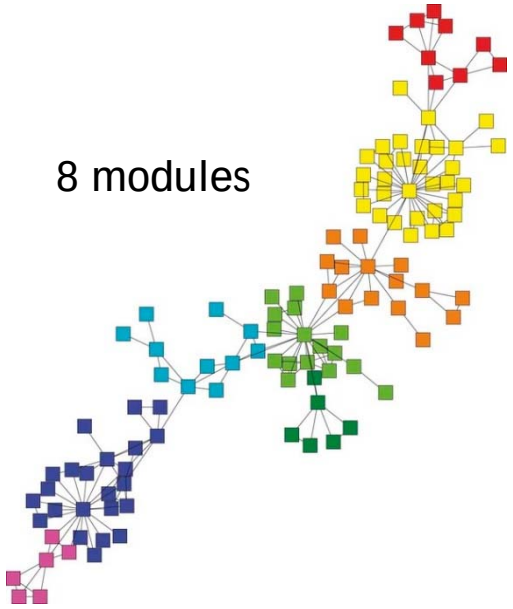
Distance(v1,v8)= 4

Clusters/modules/communities

Densely connected subgraphs = Subgraphs with many internal connections and few connections to the rest of the graph.



Two modules



8 modules

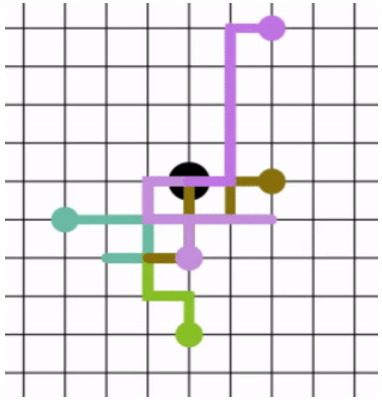
Signal diffusion in networks – aka “network propagation”

Detect the **network regions “affected” by a set of nodes** by propagating a signal from them.

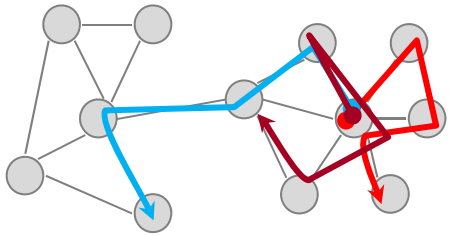
Examples:

Random walks

Simulates a set of *n* walkers falling in the node(s) of interest and moving *m* steps randomly following the network edges



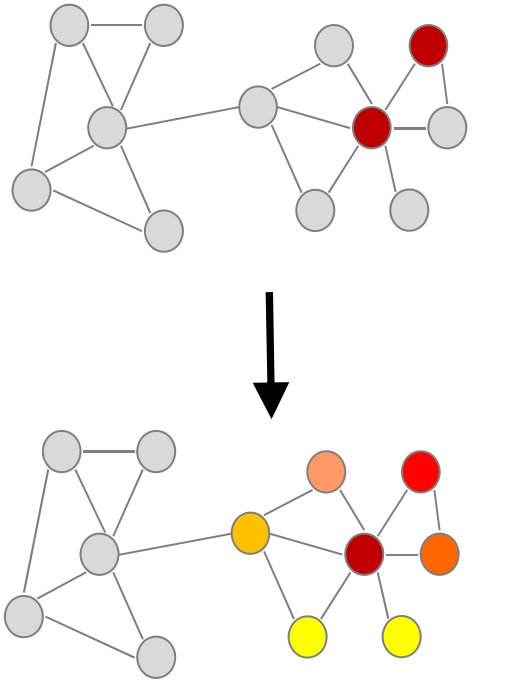
Five eight-step random walks from a central point in a 2D regular graph. Some paths appear shorter than eight steps where the route has doubled back on itself. (*Wikipedia*)



Three 4-step random walkers starting at the central node of right cluster

Heat diffusion

Simulates that the node(s) of interest are heat sources and the network edges wires able to transmit it.



HEAT

Graph Kernels

Mathematical operations with matrix representations of graphs that end up in a vector/matrix with the signal associated to each node and/or measures of “network distance” between nodes.

Main graph metrics

Local (node/edge) metrics

- Degree: k_i
- Clustering coeficient: $C_i=2\cdot n/(k^2-k)$
- Betweenness (node or edge): B_i= number of shortests paths using node/edge "i".
- Closeness: $C_i=1/\Sigma d_{ik}$

Global (graph) metrics

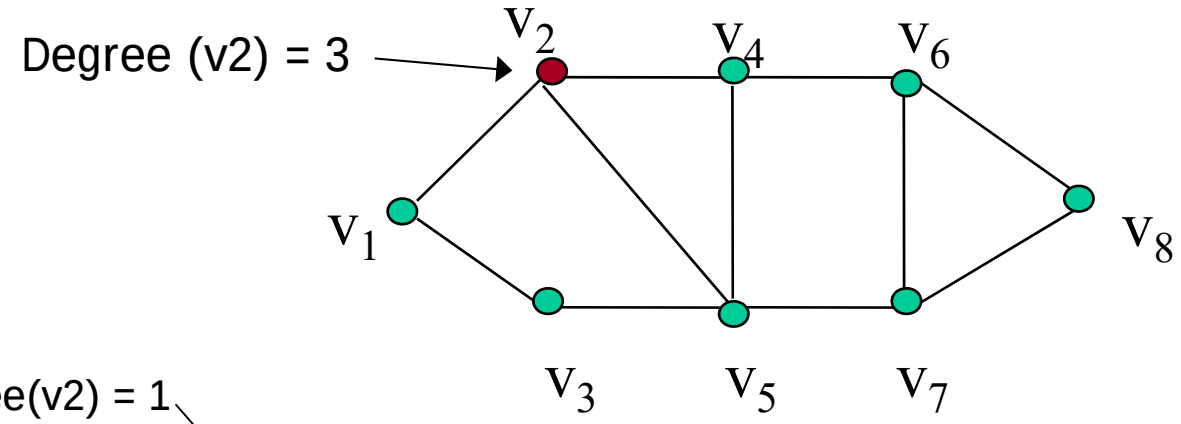
These metrics summarize the main topological features of graphs and can be used to classify them.

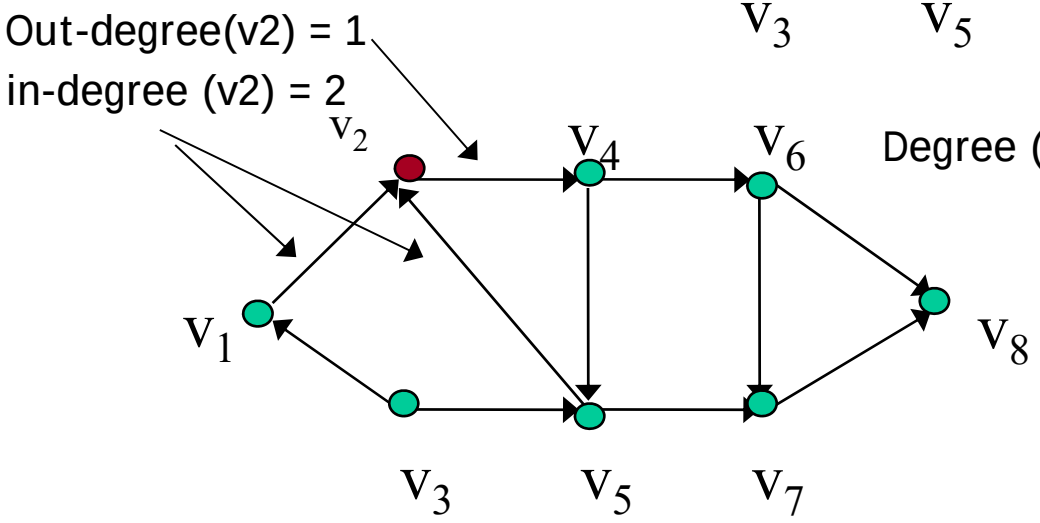
- Size |E| and order |V|
- Dispersion (|E|/|V|)
- Degree distribution: k vs. p(k)
- Average degree (<k>)
- Clustering coefficient (C)
- Average/characteristic path length (L)

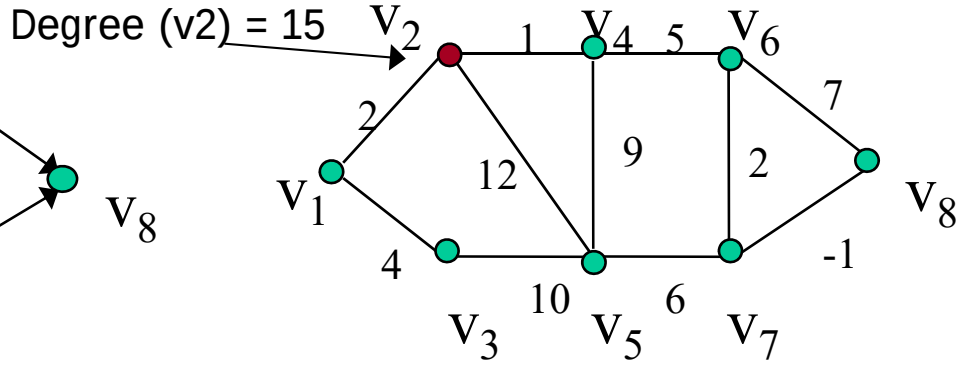
  The characteristic path length is the average of the lengths of the shortest paths (distances) between all nodes ***L*** = <***dij***>
- Diameter (D)

  The diameter of a graph is the maximum path length. ***D = max(dij)***

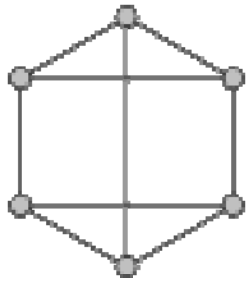
Node degree







regular graph: degree(i)=degree(j) ∀ i,j

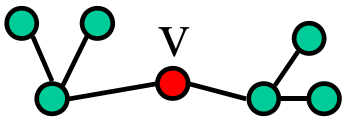


Node/Edge betweenness-centrality
And closeness-centrality

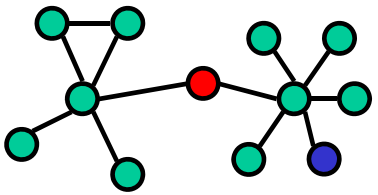
Betweenness-centrality: Number (or fraction) of shortest-paths passing through that node/edge

B(V) = 9 absolute or…
= 9/N relative to the number of shortest paths

N: #shortest paths
= #node pairs
= (7^2-7)/2



High B



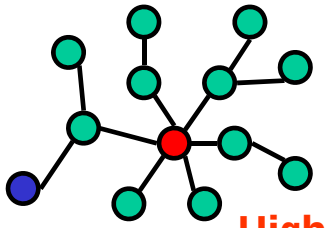
Low B

Tries to quantify the importance of a node/edge for the "transfer of information" between different parts of the network. Points to "bottlenecks", "bridges" in the network.

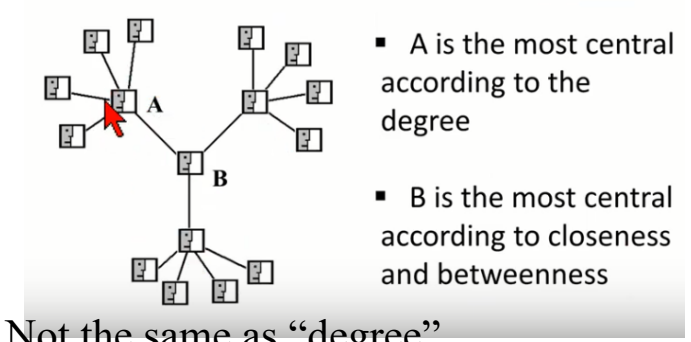
Closeness-centrality: Inverse of the sum of distances to all other nodes/edges

(graph above)
C(V) = 1/(1+2+2+1+2+2) = 1/10

Low C



High C



Not the same as "degree"

Points to central nodes close to most nodes in the network.

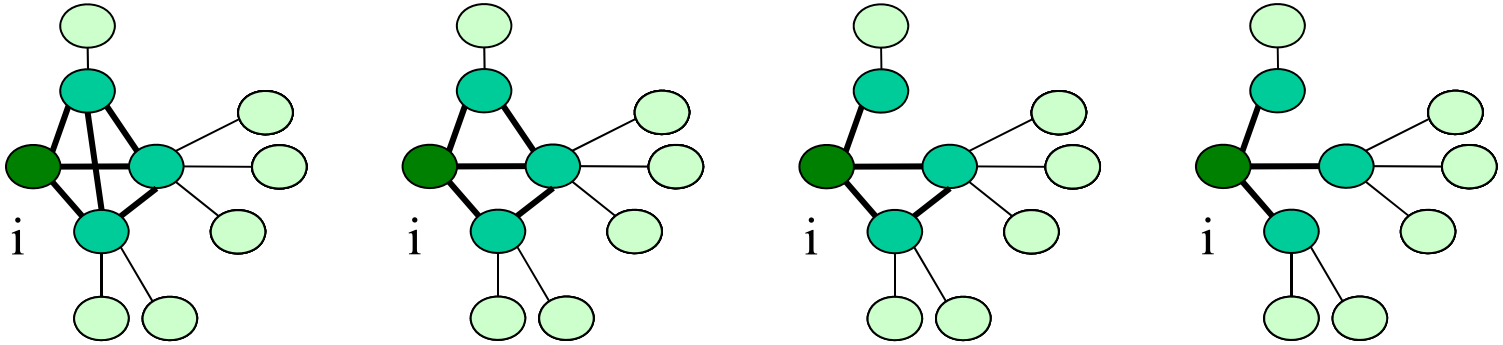
Clustering coefficient

The clustering coefficient of a node is the probability of finding a connection between its connected nodes (neighbors).

It is calculated as the number of connections between its neighbors (*n*) over the maximum number of possible connections between them.

Ci = n / (ki*(ki-1)/2) (being *ki* the degree of the node = number of neighbors)

It gives an idea how clustered or sparse is the neighborhood of a node.

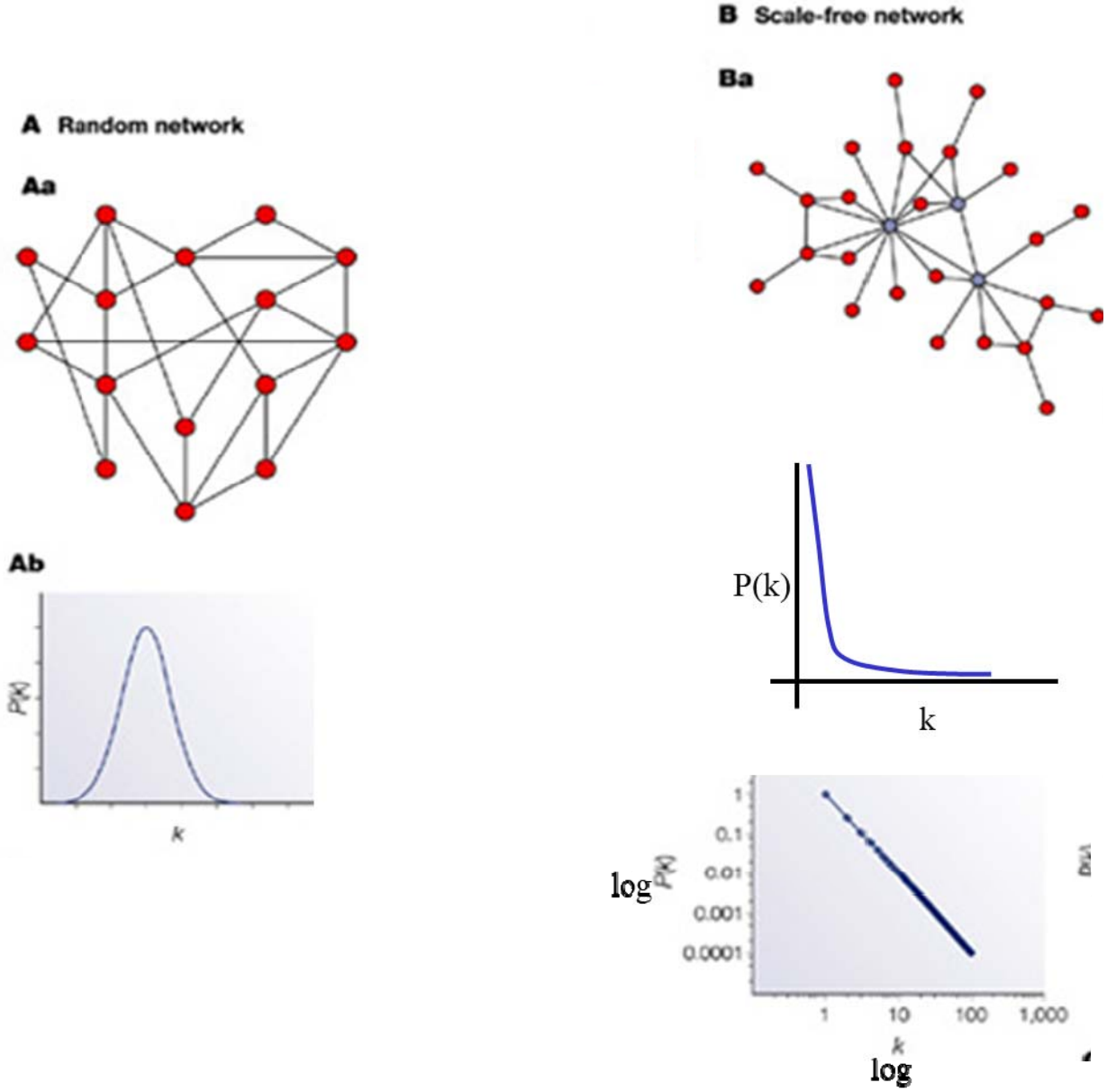


Ci=3/3 =1.0 Ci= 2/3 =0.67 Ci= 1/3 =0.33 Ci= 0/3 =0.0

The clustering coefficient of a graph is the average of the clustering coefficients of all its nodes. It gives an idea of how clustered/interconnected a graph is (vs. “star-like” graphs and trees)

C = <Ci>

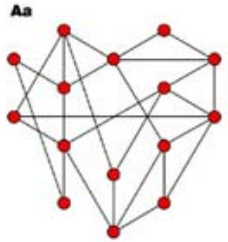
Global metrics - Degree distribution



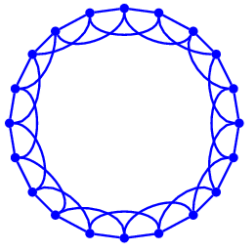
Barabasi, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet*, **5**, 101-113.

Main graph types

Random graphs
Generation: assign E edges randomly, from the $(N^2-N)/2$ possible. Equivalent to create each edge with probability p=E/(N2-N)/2
Degree distribution: poison.



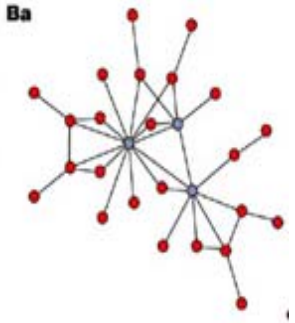
Regular graphs
Degree distribution: single value k.
Mathematical "curiosities". There are analytic expressions for all the metrics.
For small k (disperse graphs), C~0.75 (high) and L ~ N/k

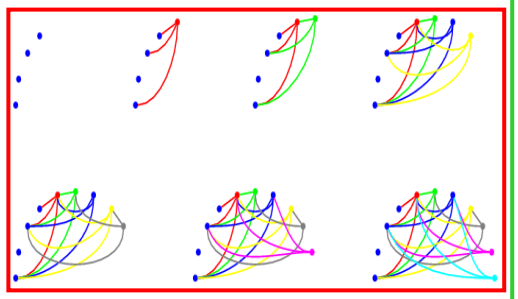


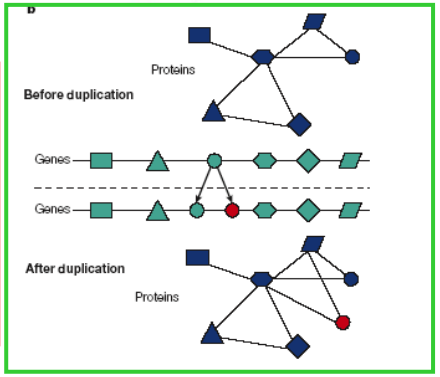
Scale-free graphs
Arise from various possible generation mechanisms:

- Preferential attachment: add nodes to an existing network connecting them PREFERENTIALLY to already highly connected nodes ("rich gets richer", …)
- Node duplication preserving links (biological networks)

Characterized by "**hubs**" (highly connected nodes, in low number)
"Small world" in some cases (hubs act as shortcuts).
Degree distribution: power-law: $p(k) = C \cdot k^{\gamma}$
γ characterizes the connectivity pattern of the network (proportion of hubs) and in real networks it ranges ~ 2.0 – 3.0
Adequately represent a number of real-word networks such as many biological networks and the Internet





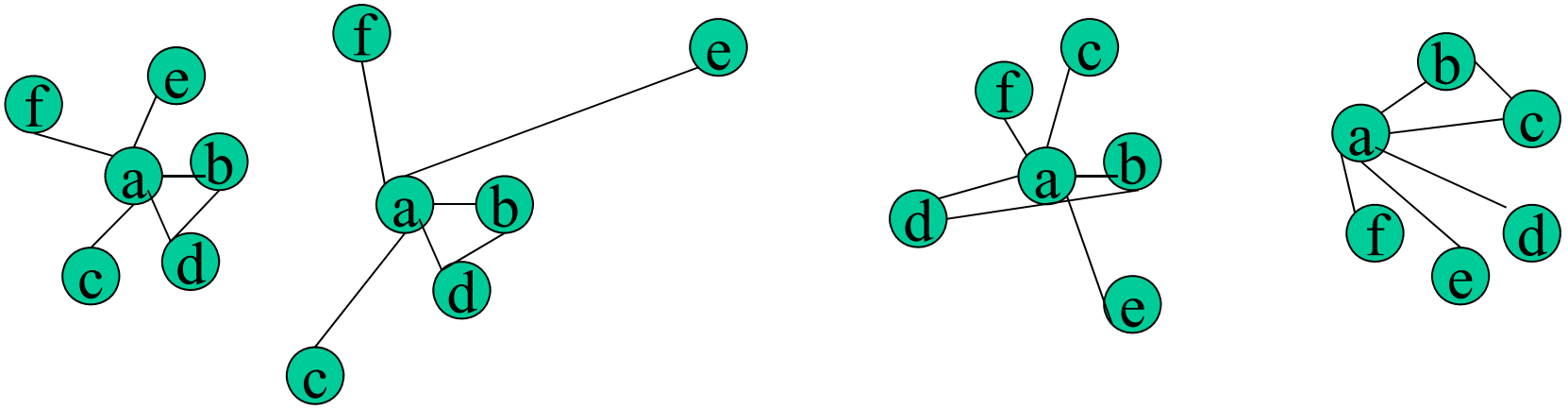


Graph layout

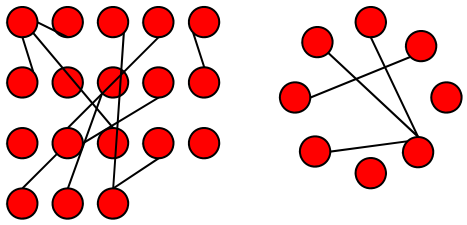
A “layout” is an arrangement of the graph nodes (and edges) in 2D or 3D which facilitates its visualization and/or makes more evident some properties of the graph.

The same graph can be visualized with many different layouts

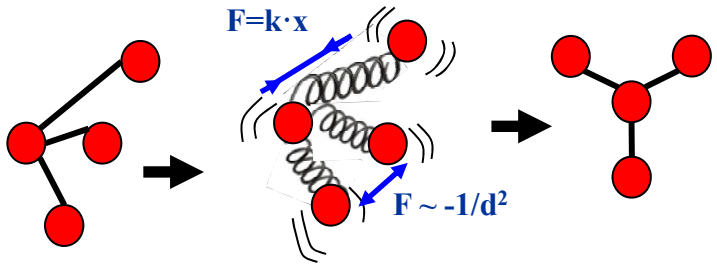
G= {a, b, c, d, e, f, (a,b), (a,c), (a,d), (a,e), (a,f), (b,d) }



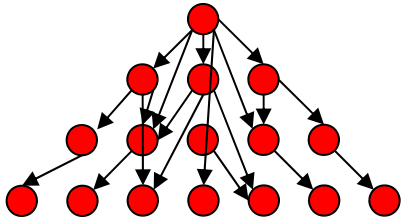
Regular arrangements



Based on “physics”

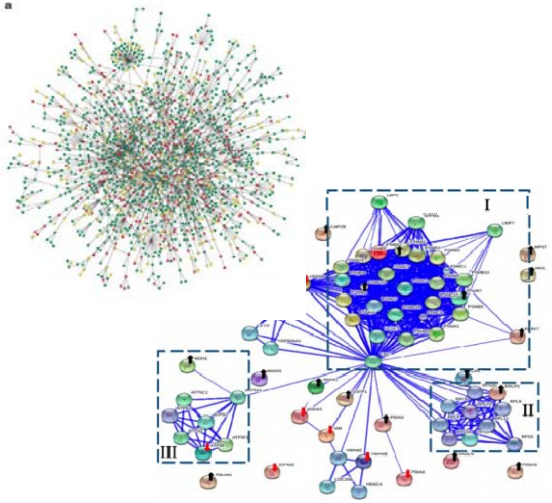


Based on topological properties (e.g. sort by k, to emphasize hierarchy)

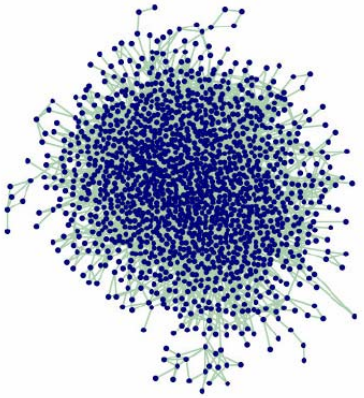


Main Biological Networks

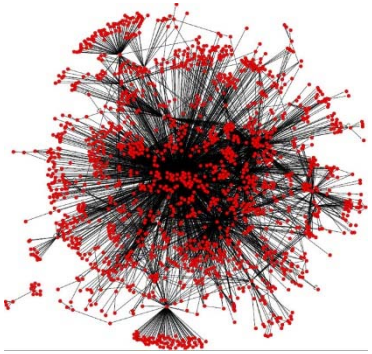
Interactome: Undirected network. Nodes=proteins, edges=protein interactions.
Scale-free structure (due to gene duplication)=> resistance to random perturbation while sensitivity to direct “attacks”. Hubs=conserved and related to essentiality. Small world. Modular (topological clusters=functional clusters (biological processes, macromolecular complexes, …).



Metabolome: Directed network. Nodes=metabolites, edges=chemical transformations.
Scale-free structure=> resistance to random perturbation while sensitivity to direct “attacks”. Small world (=> short and efficient chemical transformations). Modular, topological clusters~ classical metabolic pathways.

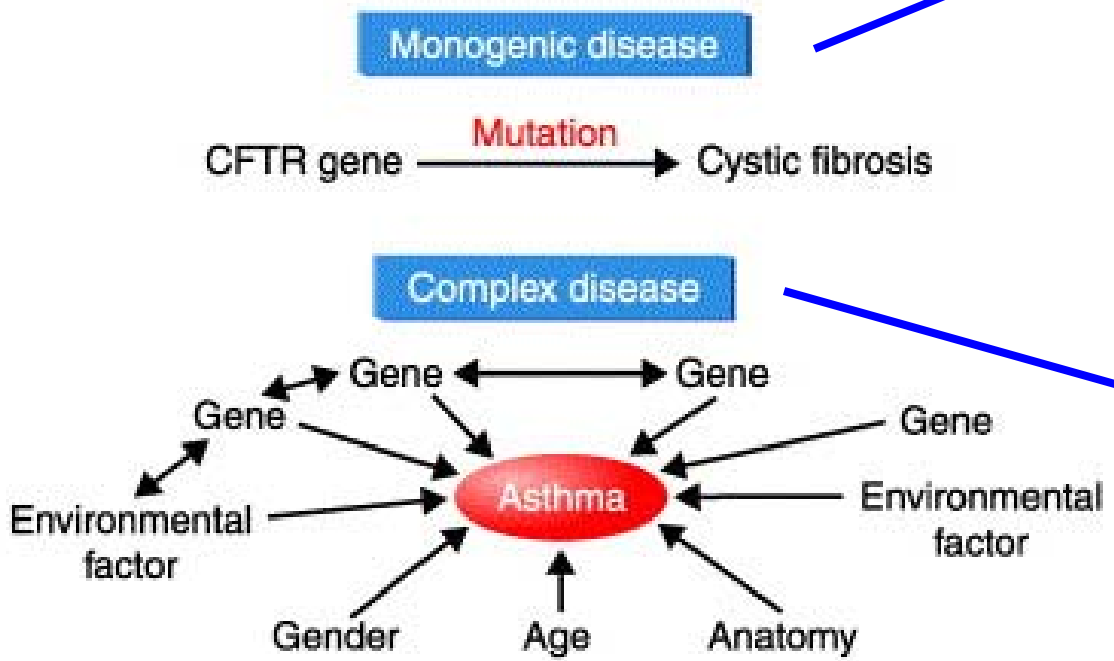


Regulome: Directed network. Nodes=genes/proteins, edges=gene regulatory relationships (TF->gene).
Scale free for out-degree, exponential for in-degree. Modular, topological clusters= biological processes. Bottlenecks more related to essentiality than hubs. Presence of overrepresented small functional motifs such as the “feed-forward loop”.



Other: genetic nets, phosphorilome, co-expression networks, miRNA-target networks, …

Monogenic vs. complex diseases

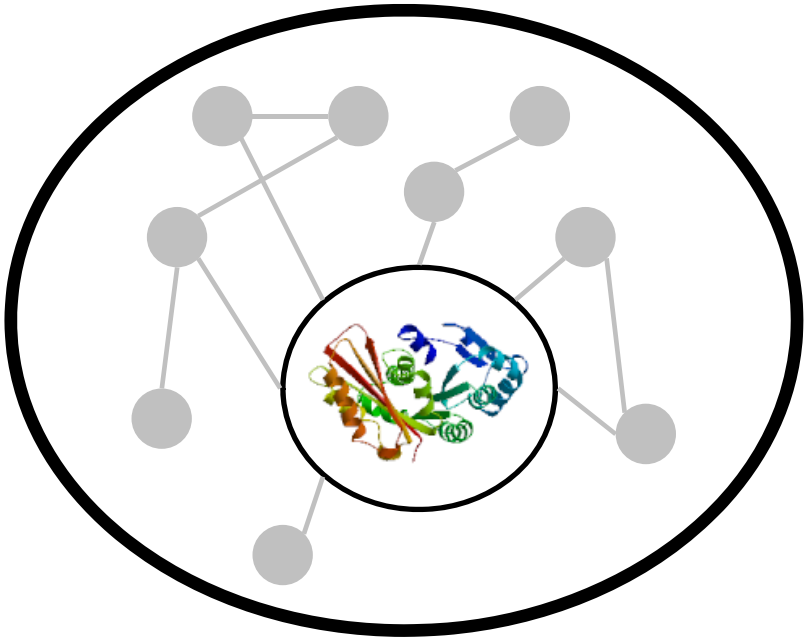


- Anemia falciforme
- Fibrosis quística
- Fenilcetonuria
- Enfermedad de Batten
- Hemocromatosis
- Deficiencia de alfa-1 antitripsina
- Enfermedad de Huntington
- Enfermedad de Marfan
- Distrofia muscular de Duchenne
- Síndrome de cromosoma X frágil
- Hemofilia A
- ...

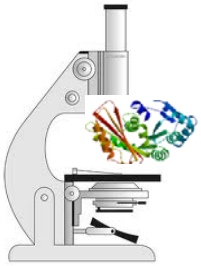
- Cáncer
- Autismo
- Diabetes
- Obesidad
- Alzheimer
- Asma
- Parkinson
- Esclerosis múltiple
- Osteoporosis,
- ...

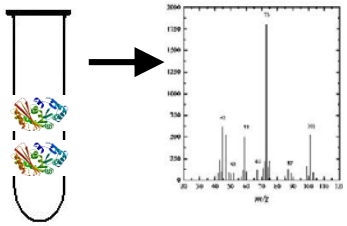
Whittaker, P. A. Genes for asthma: much ado about nothing? *Curr Opin Pharmacol* 3, 212–219 (2003).

Reductionist approaches to diseases

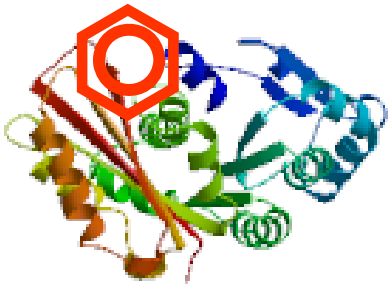


Diagnosis
(markers)





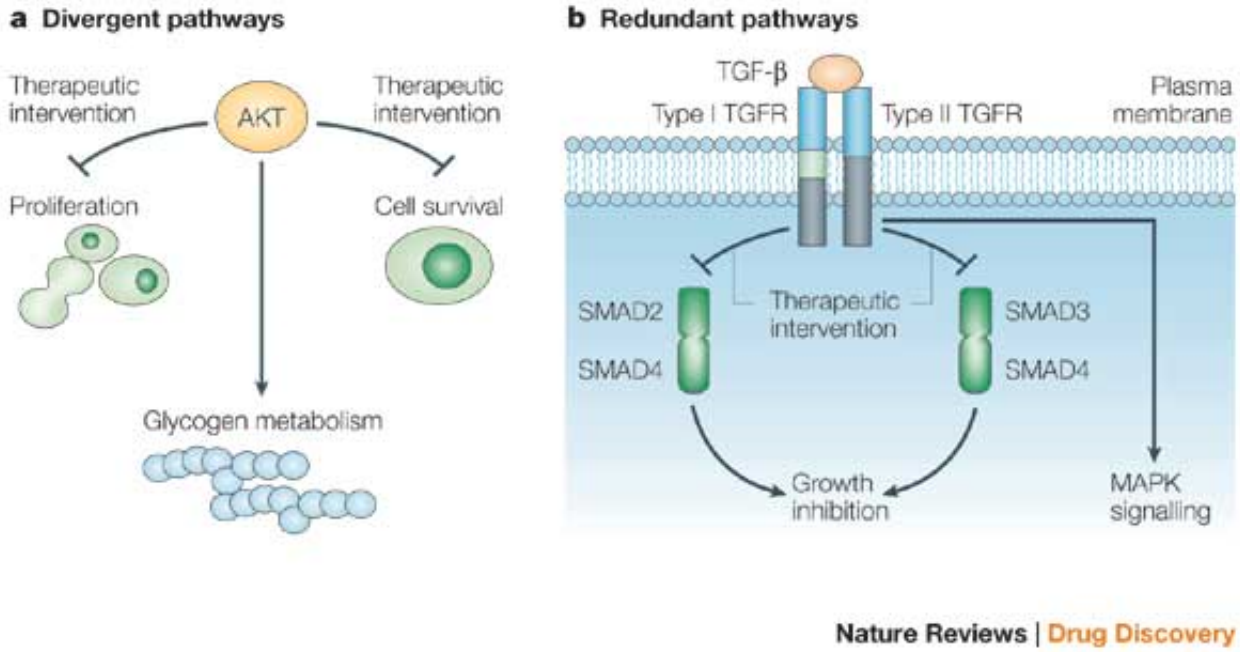
treatment



The traditional reductionist approach to diseases is based on locating THE gene associated to the disease, so that it can serve as marker for diagnosing the disease and, eventually, as target for curing it.

Networked systems require "networked" drugs

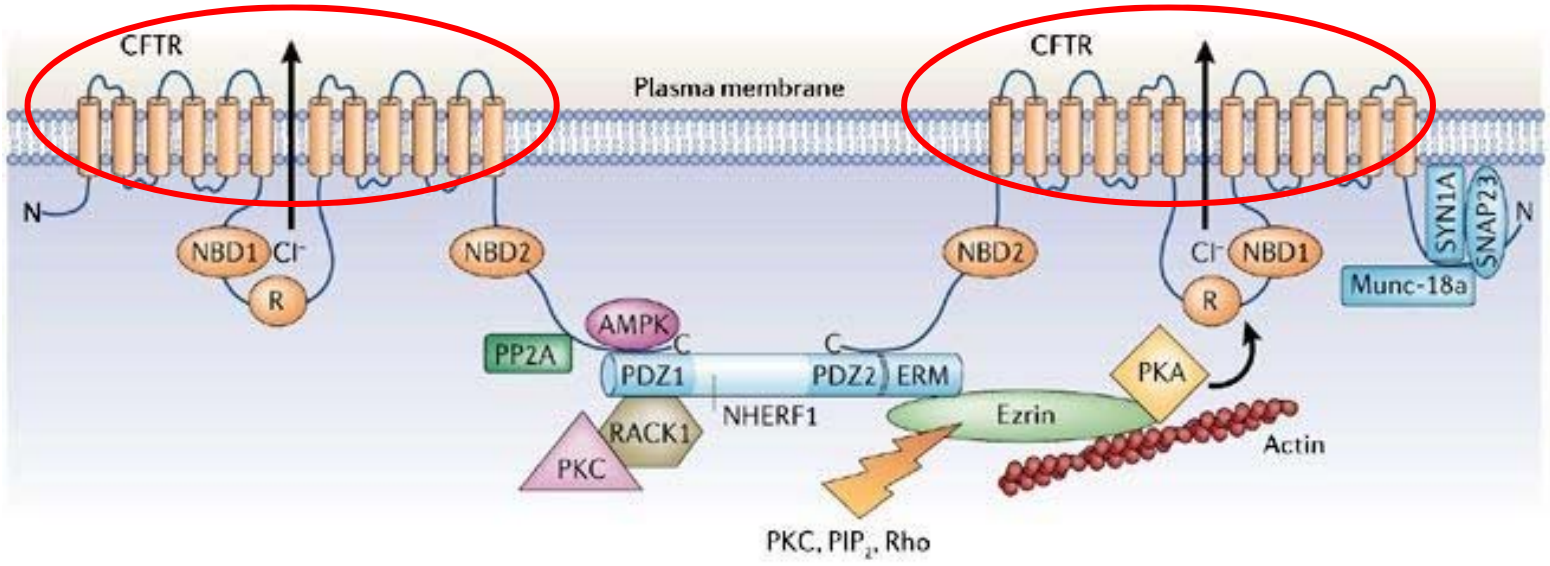
The intrinsic complexity of biological systems, reflected in the molecular networks is the main cause for the limitations of the reductionist approach



Networked systems might require multicomponent interventions to modulate signalling outputs. **a** | Targets at divergent pathway nodes might cause undesired side effects when acted on in isolation. For example, AKT regulates several downstream outputs, so inhibiting this protein on its own is not likely to achieve a separation of desired and undesired effects. If we want to inhibit cell-proliferation and cell-survival pathways, for example, without affecting glycogen metabolism, we would need multicomponent drugs to specifically inhibit these two downstream pathways rather than using a single AKT inhibitor. **b** | Redundant pathways can compensate for inhibition of another pathway. For example, SMAD2 and SMAD3 perform largely similar functions in tissue culture experiments. A small-molecule inhibitor of either SMAD2 or SMAD3 alone would therefore not be effective at blocking transforming growth factor- (TGF-) signalling if cells responded by upregulating a redundant SMAD. TGF- regulates several downstream outputs, so inhibiting this protein on its own could cause undesired effects by inhibiting SMAD-independent TGF- effects, such as activation of mitogen-activated protein kinase (MAPK) signalling. Using multicomponent interventions to simultaneously inhibit SMAD2 and SMAD3 would overcome both these problems by blocking SMAD-dependent TGF- effects without inhibiting SMAD-independent TGF- effects. TGFR, TGF- receptor.

Keith, C.T., Borisy, A.A. and Stockwell, B.R. (2005) Multicomponent therapeutics for networked systems. *Nat Rev Drug Discov.*, **4**, 71-78.

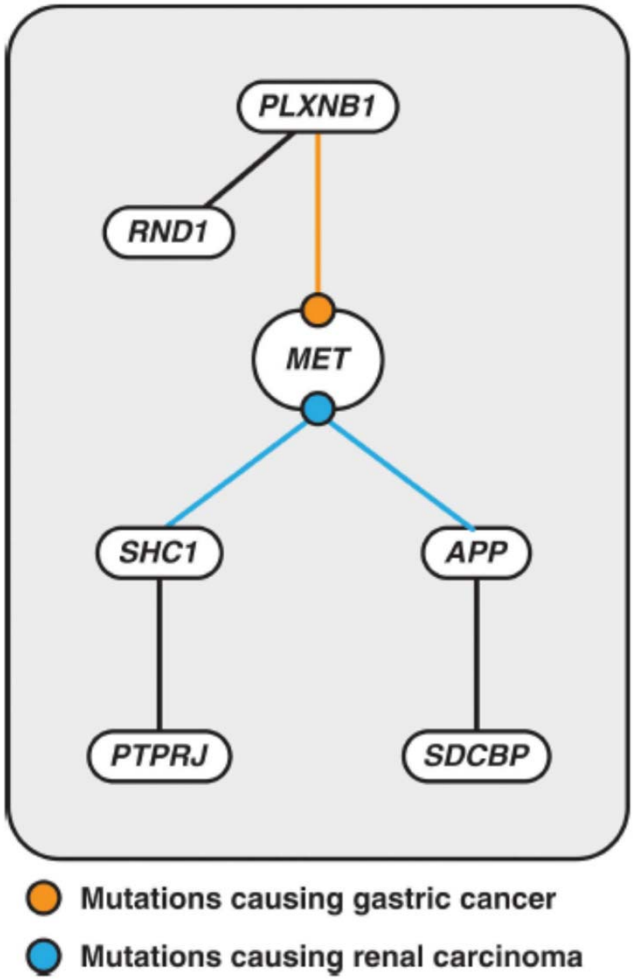
Classic monogenic diseases and networks



Even in “classic” monogenic diseases the causative gene(s) are inmmersed in molecular networks and hence, at least the severity of the disease depends on many other genes/mutations. Eg. Cistic fibrosis (CFTR gene).

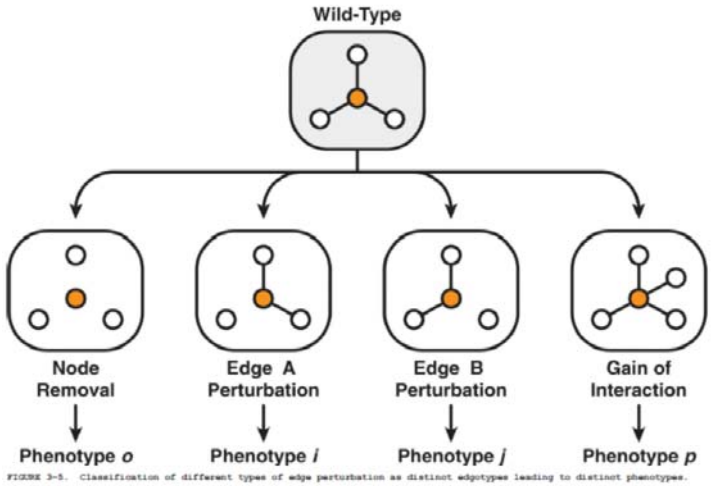
Guggino WB, Stanton BA. **2006**. New insights into cystic fibrosis: molecular switches that regulate CFTR. *Nat. Rev. Mol. Cell Biol.* 7(6): 426–36

Gene mutations and network context



Network context is fundamental for interpreting the differential effects of different mutations of the same gene.

Edgetic perturbation/disruption:
Disruption of a particular edge (e.g. interaction) in a biological network leaving the involved nodes functional (in their other functions/interactions).

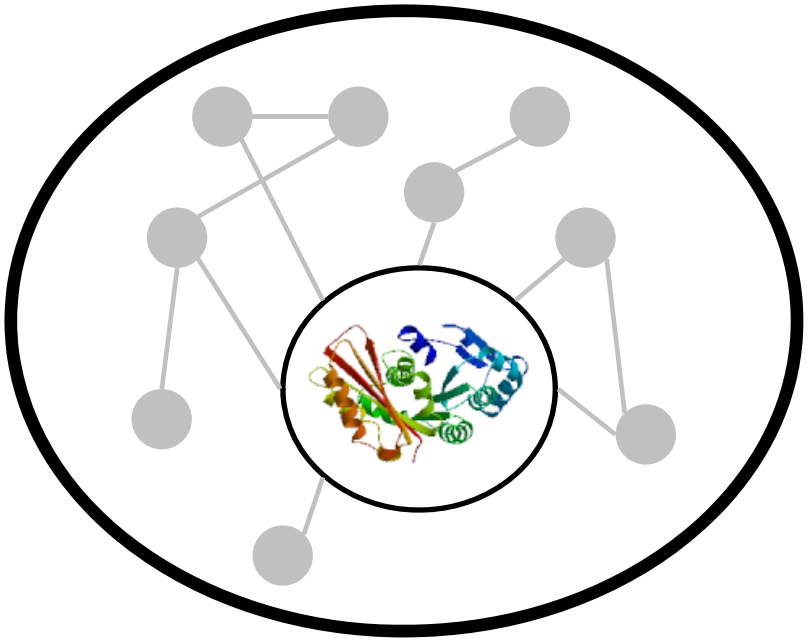


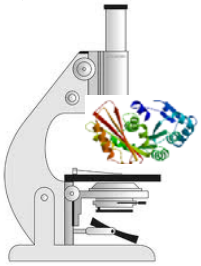
Joseph Loscalzo, Albert-László Barabási and Edwin K. Silverman (Eds) (2017). Network Medicine: Complex Systems in Human Disease and Therapeutics. Cambridge, Massachusetts : Harvard University Press.

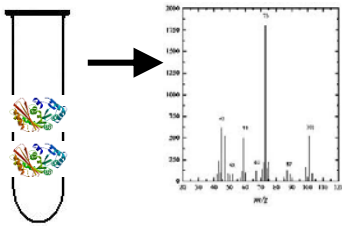
Reductionist vs. systemic approaches to diseases

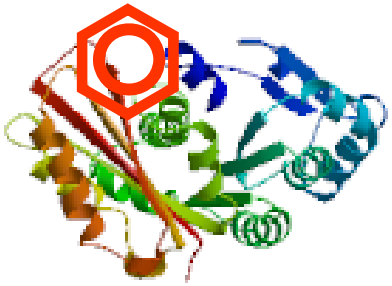
Diagnosis
(markers)

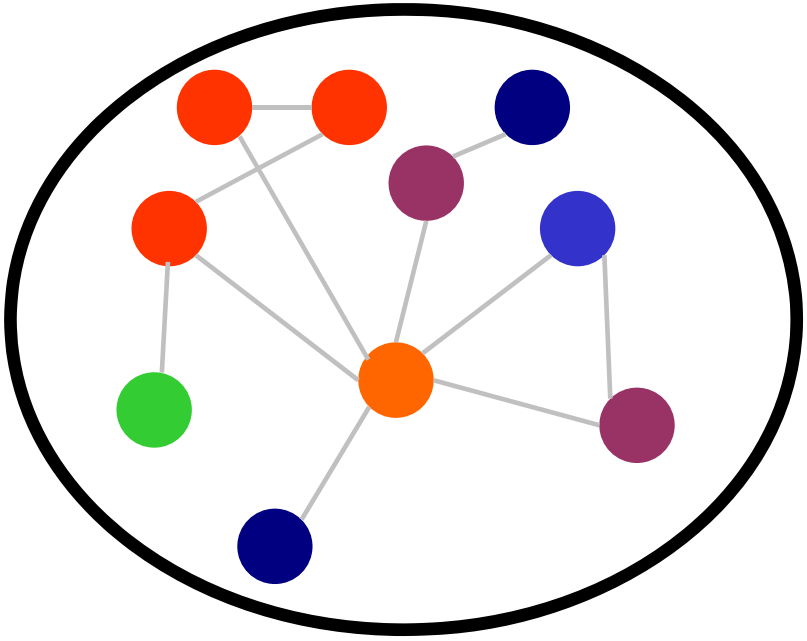
treatment

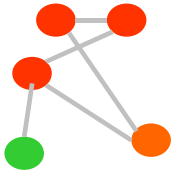


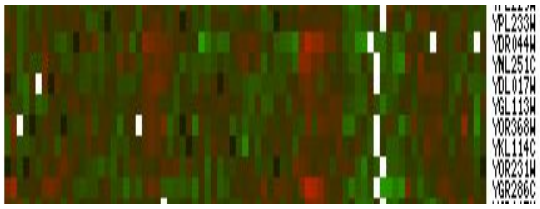


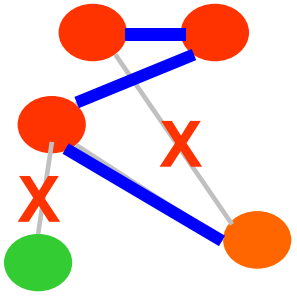












Perturbations in cellular networks might explain phenotype-genotype relationships

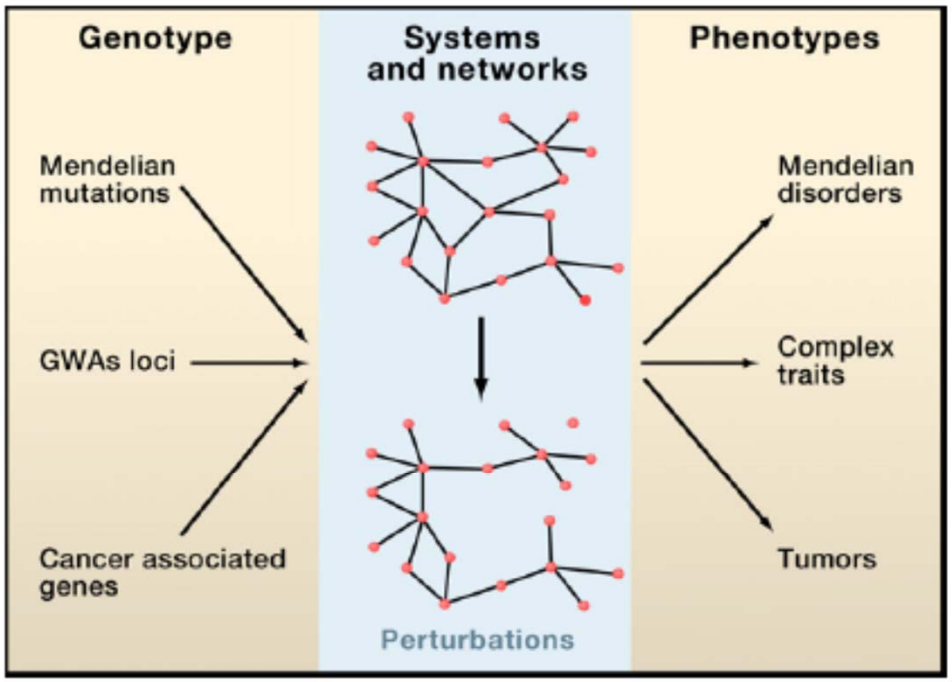
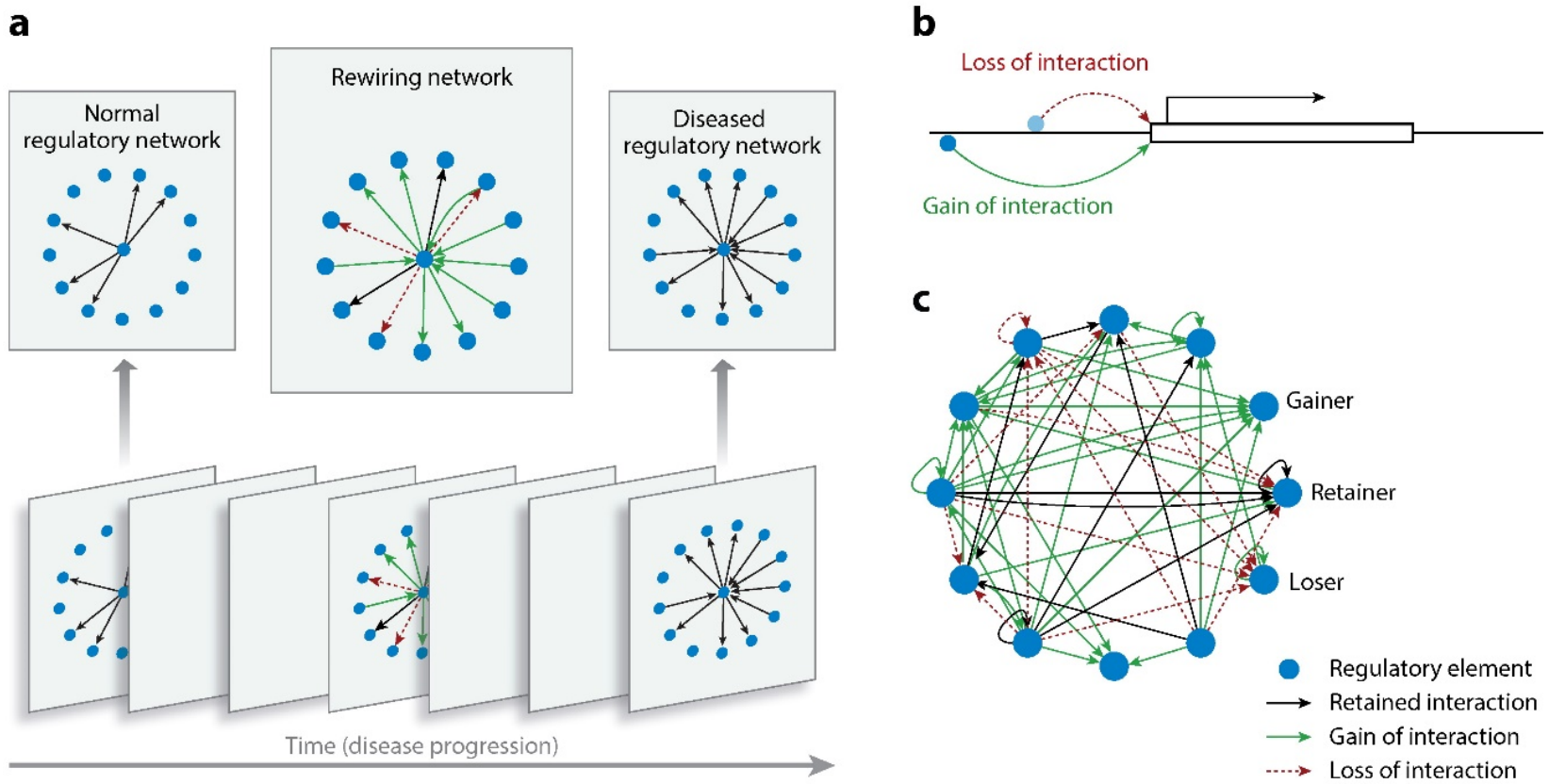


Figure 1. Perturbations in Biological Systems and Cellular Networks May Underlie Genotype-Phenotype Relationships

By interacting with each other, genes and their products form complex cellular networks. The link between perturbations in network and systems properties and phenotypes, such as Mendelian disorders, complex traits, and cancer, might be as important as that between genotypes and phenotypes.

Vidal M, Cusick ME, Barabási AL. (2011). Interactome networks and human disease. *Cell.* **144(6):**986-98.

Diseases as re-wiring of molecular networks



For the network/systemic approaches to diseases, these are caused by perturbations (e.g. re-wiring) of large networks, instead of single genes.

McGillivray, P., Clarke, D., Meyerson, W., Zhang, J., Lee, D., Gu, M., Kumar, S., Zhou, H. and Gerstein, M.B. (2018) Network Analysis as a Grand Unifier in Biomedical Data Science. *Annual Review of Biomedical Data Science*.

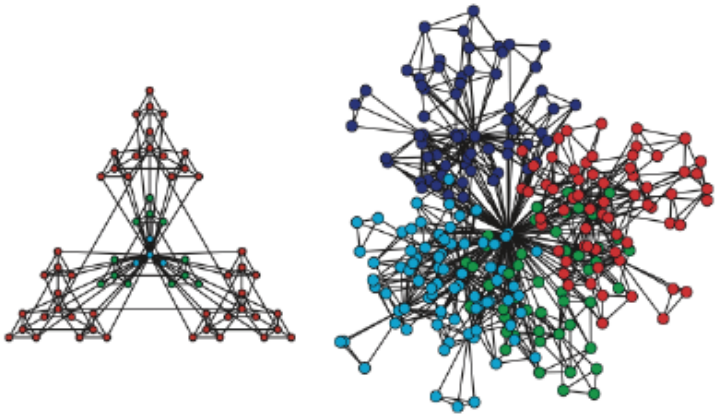
Human molecular networks

-Protein interactions

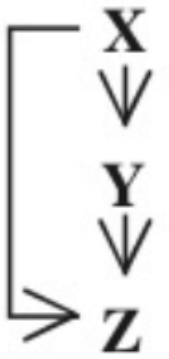
- Metabolic network

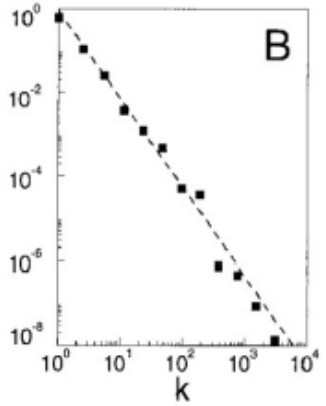
- Gene regulatory network

- Others… co-expression nets, phosphorilome, RNA nets, combinations of the above …



Associated dynamical data:
- Gene expression
- Metabolomics





Main characteristics of the human networks:

- Modularity

- Scale free structure => hubs

- Small world

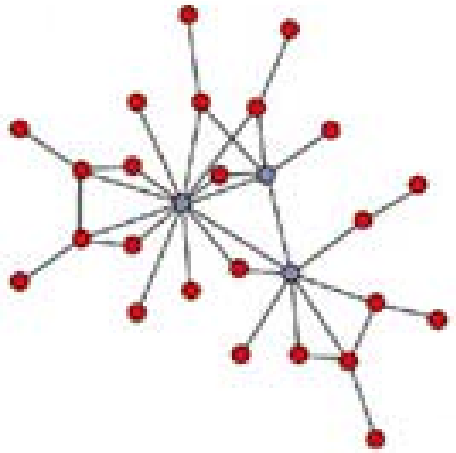
- Presence of small motifs

Hubs

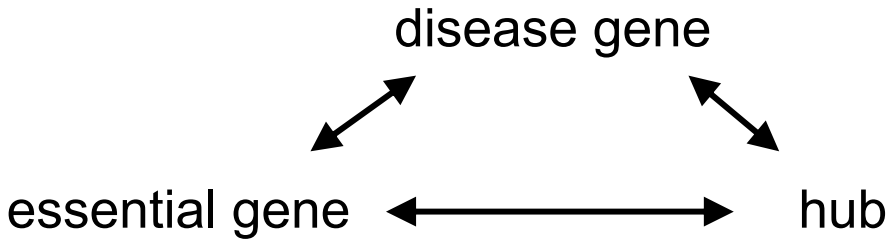
- Upregulated genes associated to lung carcinoma tend to have higher degree than unchanging ones.

-346 proteins involved in Cancer have twice as many interactors as “non-cancer” proteins

-disease proteins in the OMIM Morbid Map have more protein–protein interactions than do non-disease proteins

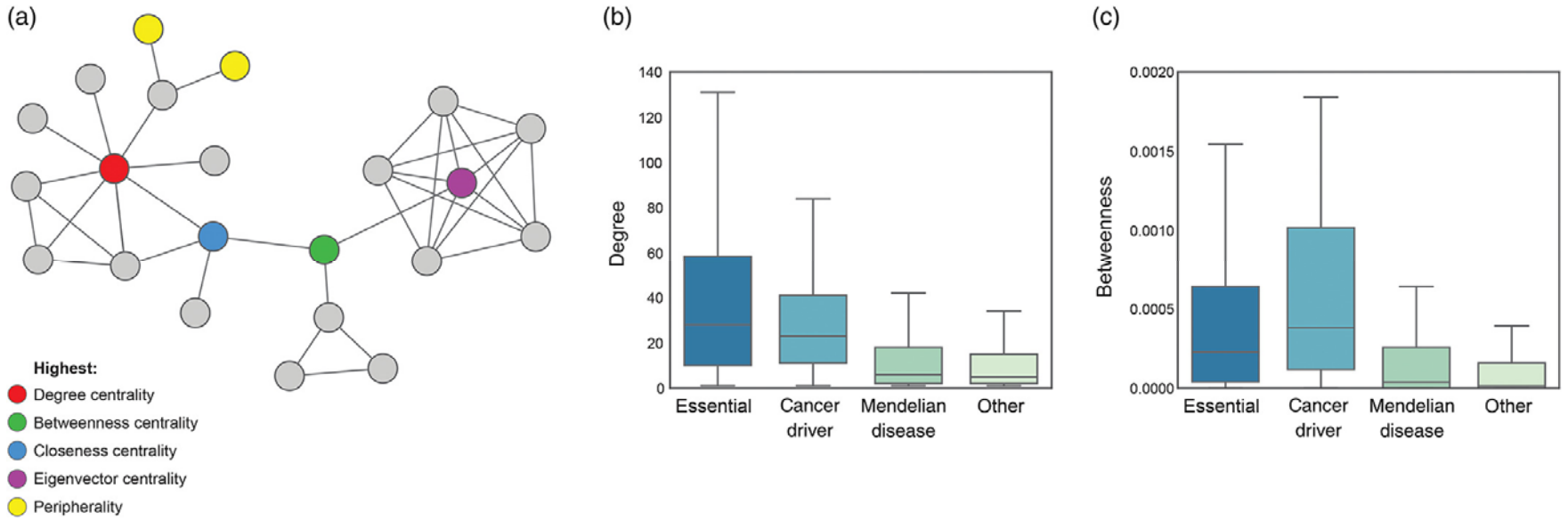


- Nevertheless, disease genes ≠ essential genes (not viable => no disease)



Barabási AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. Nat Rev Genet 12(1): 56–68.

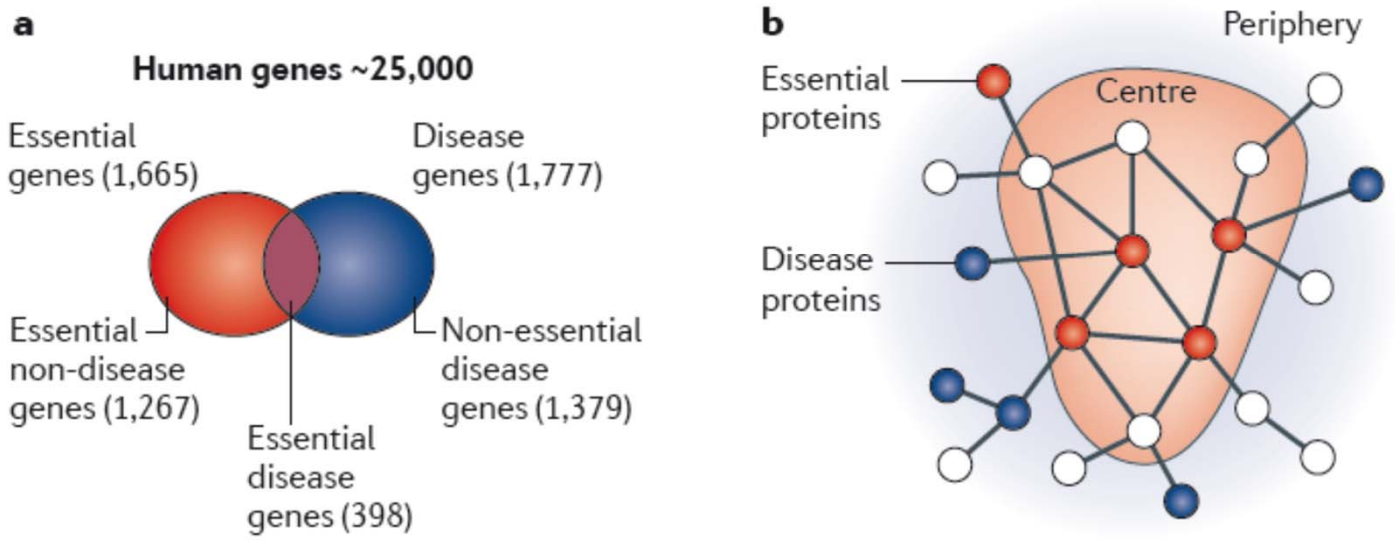
Disease-related and essential genes in the interactome



"Hubs" are more related to essentiality than to disease.
"Bottlenecks" are more related to diseases, at least in Cancer.

Capriotti, E., Ozturk, K. and Carter, H. (2019) Integrating molecular networks with genetic variant interpretation for precision medicine. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine,* **11(3),** e1443.

Disease-related and essential genes in the interactome



Disease-related proteins tend to have more interactors than the average but they are not hubs. Hubs tend to be essential proteins and hence do not lead to disease but to (embryonic) death.

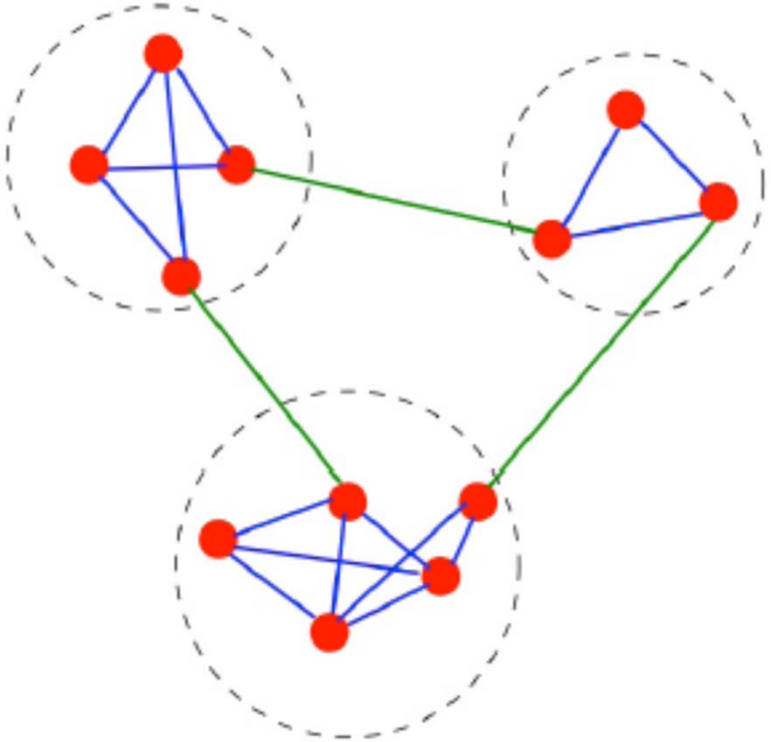
Barabási AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. Nat Rev Genet 12(1): 56–68.

Network modules

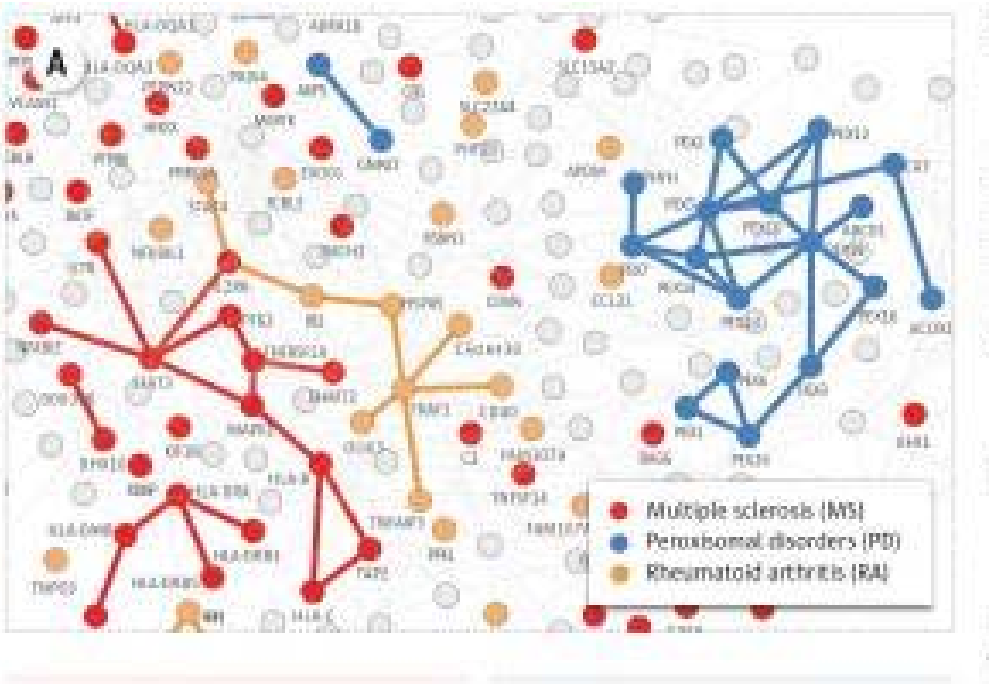
Module: group of nodes highly connected among themselves and poorly connected to the rest of the network

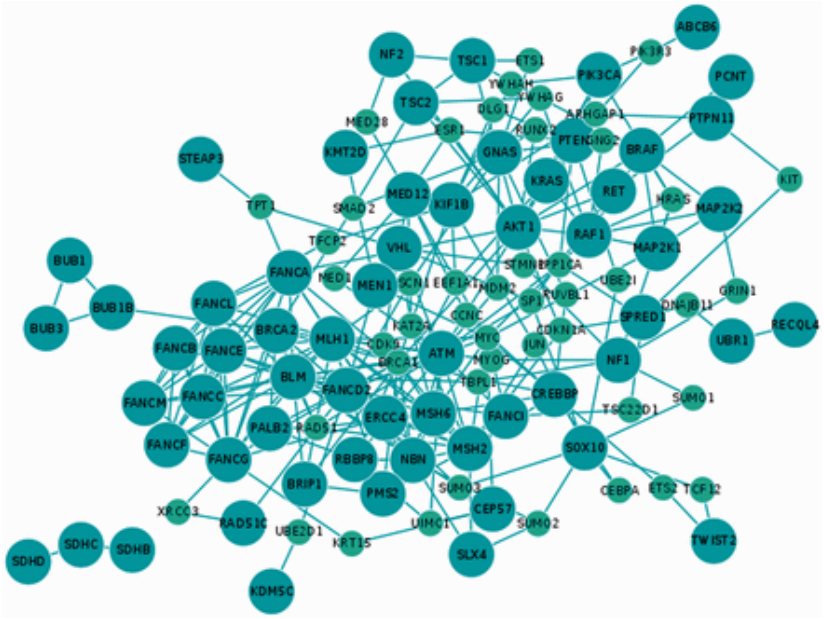
In general… function/role separable from the rest of the network

In biological nets… correspondence with "functional" modules/pathways



Diseases and symptoms are related to Network modules





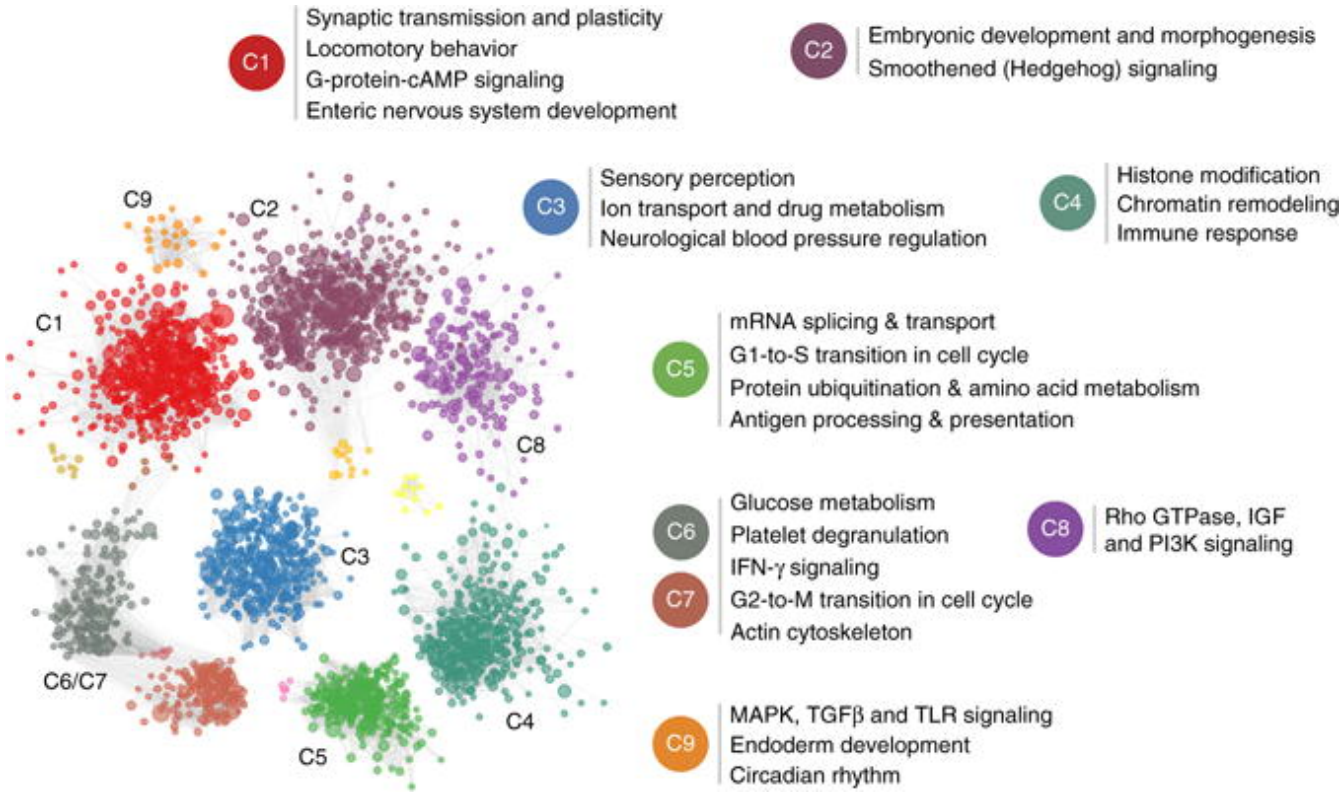
Example of a modular clinical sign: *café-au-lait* spots.

Genes known to be associated to diseases and clinical signs tend to cluster in molecular networks.

Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabási AL. (2015). Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science*. **347(6224):**1257601.

Monica Chagoyen, Florencio Pazos (2016). Characterization of clinical signs in the human interactome, *Bioinformatics*. **32(12):** 1761–1765,

Diseases and symptoms are related to network modules (=functional pathways)

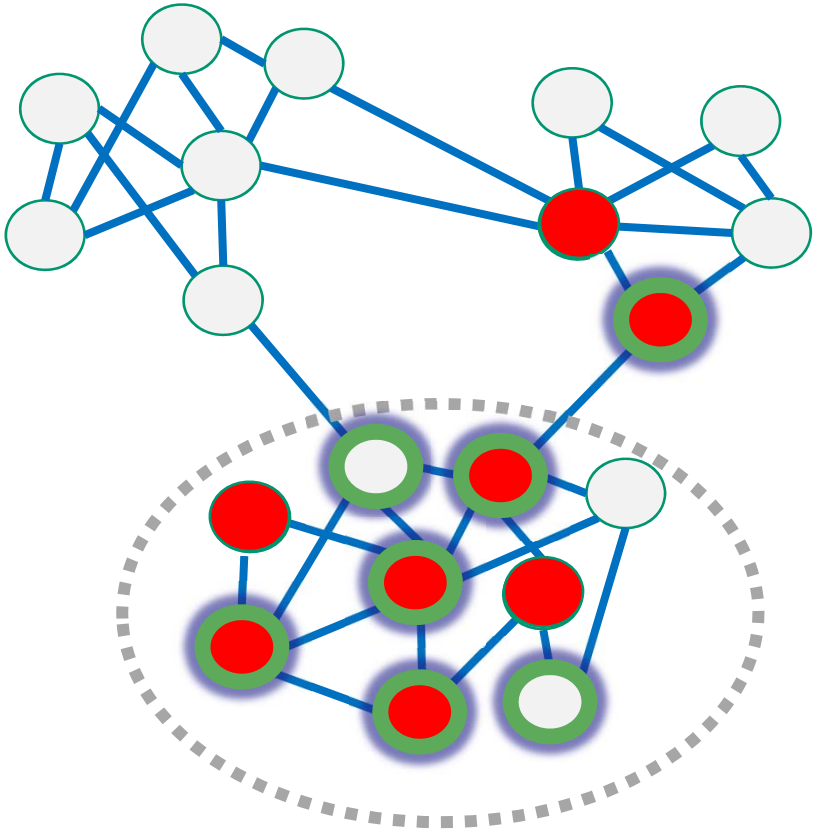


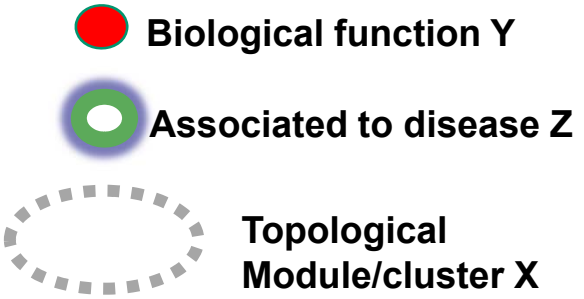
Even in very complex diseases involving hundreds/thousands of genes, these tend to concentrate in a reduced number of modules/pathways

---

Krishnan A, Zhang R, Yao V, Theesfeld CL, Wong AK, et al. (2016). Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat. Neurosci.* **19(11):** 1454–62

Relationship between network topological modules, functional pathways and diseases



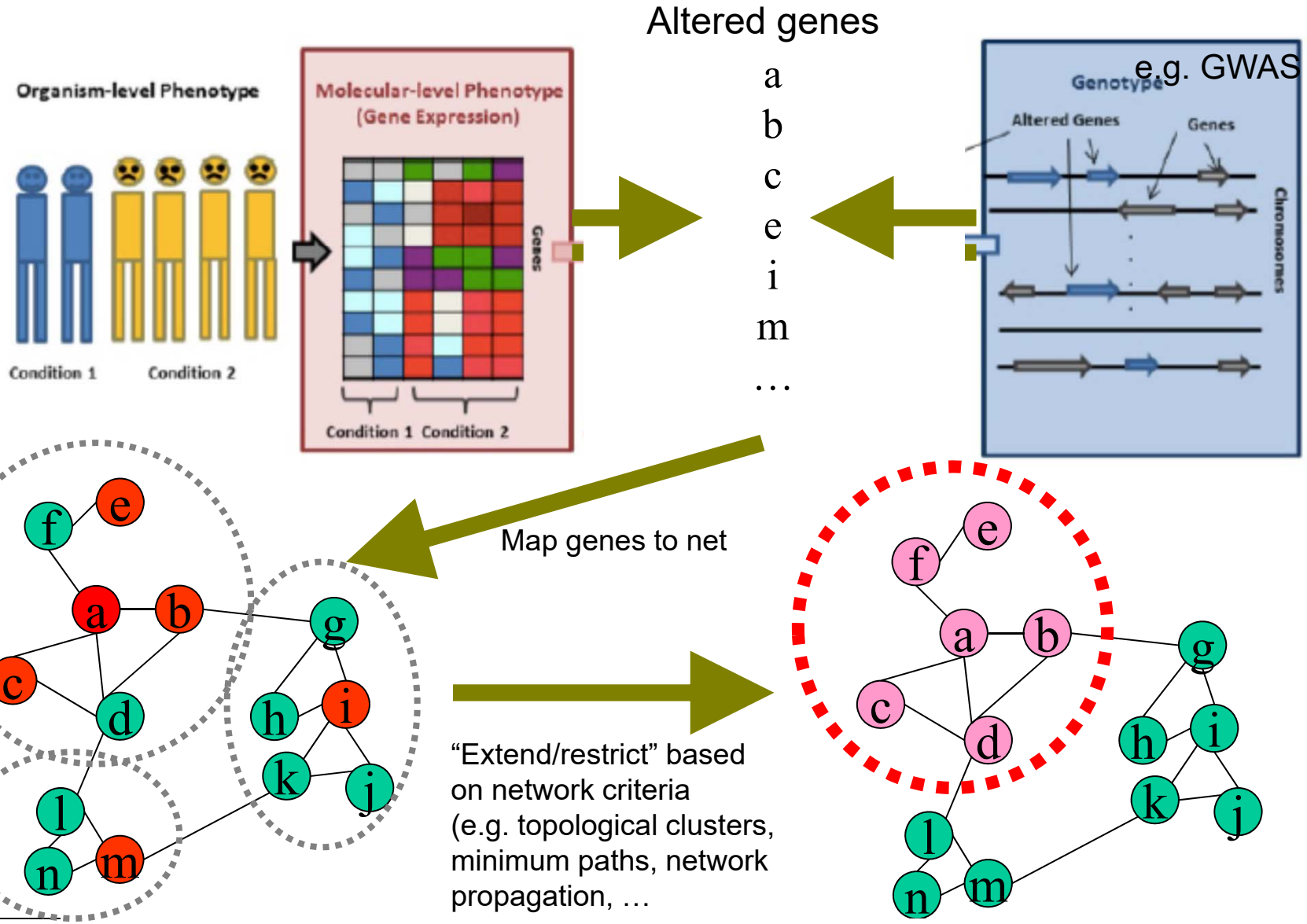


Working model of network-based approaches to diseases:

"Disease **Z** is due to a malfunctioning in biological system **Y**, whose function is carried out by a group of proteins working together (reflected in topological module **X**)"

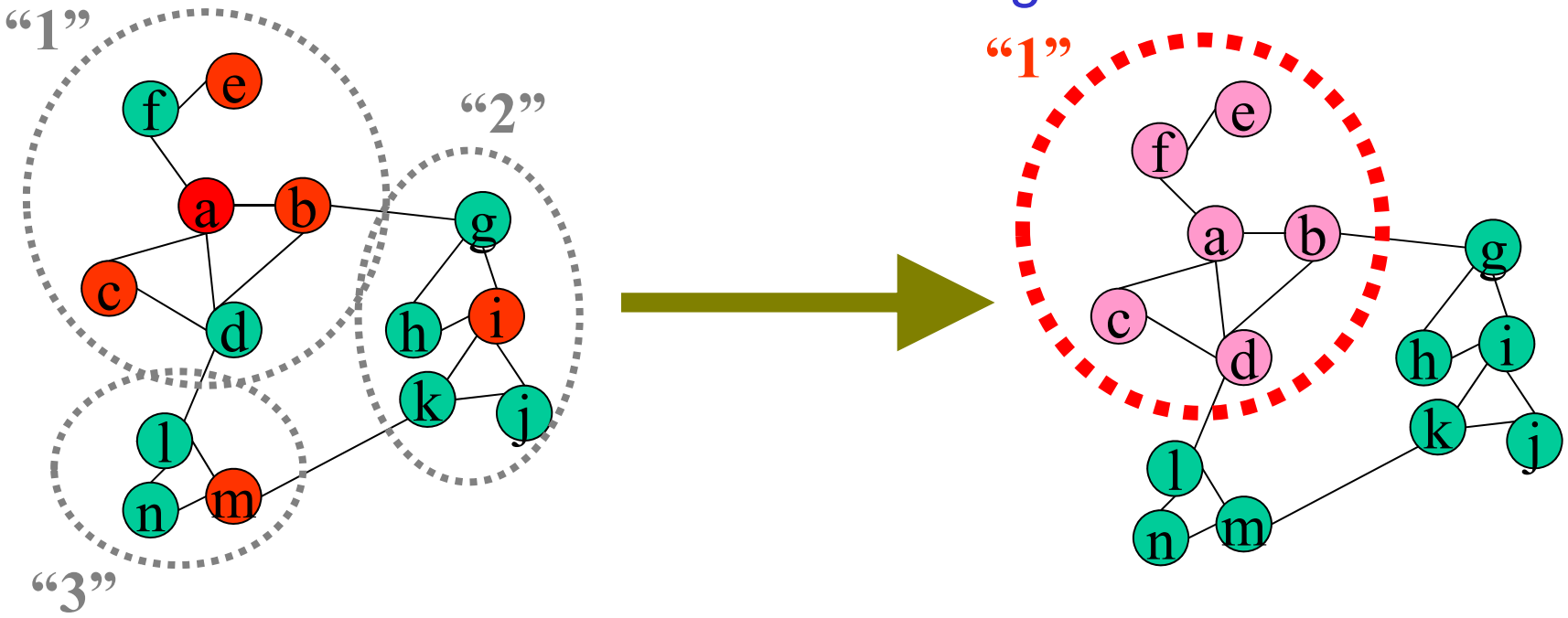
Monica Chagoyen, Juan A G Ranea and Florencio Pazos. (2019). Applications of molecular networks in biomedicine. *Biology Methods and Protocols*. **4(1)**:bpz012.

Identifying disease-related modules
General strategy



Kim, Y.-A. & Przytycka, T. M. Bridging the Gap between Genotype and Phenotype via Network Approaches. Front. Genet. 3,

Identifying disease-related modules
General advantages



- Identify genes not altered (or not passing threshold for being selected as "altered") but eventually important. E.g. "f" (connecting a and e)

- Discard "unrelated" genes (experimental errors, among other things). E.g "l".

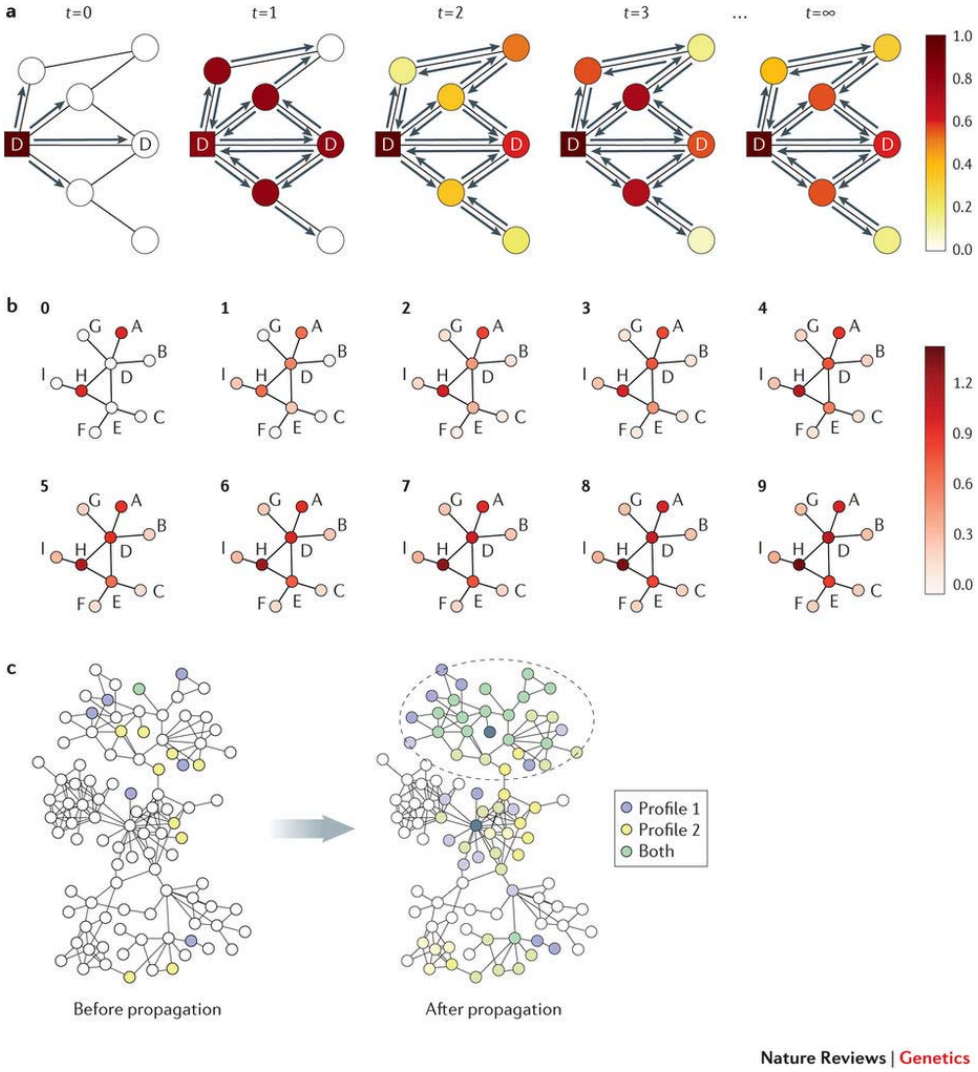
- Provide additional molecular information on the disease/alteration. E.g. Disease related to pathway "1"

- Identify other potential targets eventually more "drugeable" (e.g f, d)

- Design re-wiring strategies for recovering, for example, a malfunctioning module.

Cho D-Y, Kim Y-A, Przytycka TM (2012) Chapter 5: Network Biology Approach to Complex Diseases. PLoS Comput Biol 8(12): e1002820.

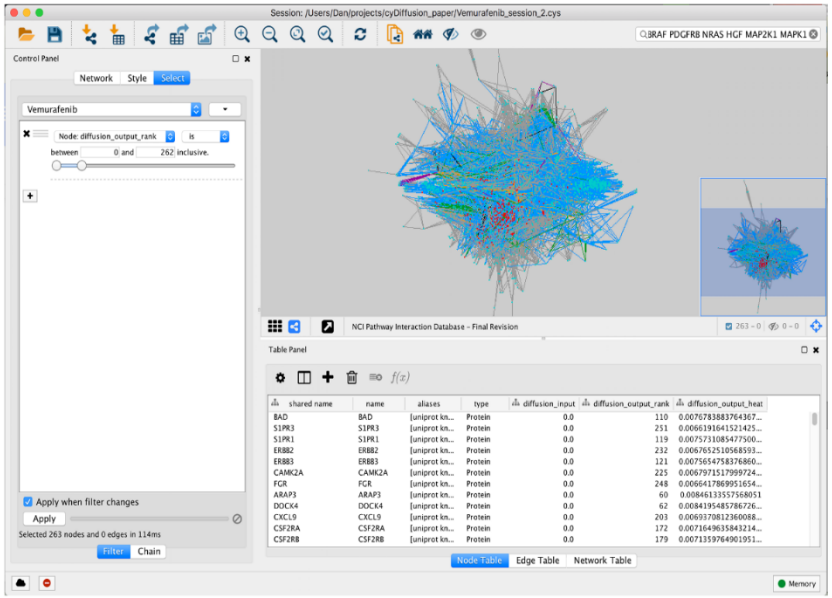
Identifying disease-related modules
Network propagation – General strategy

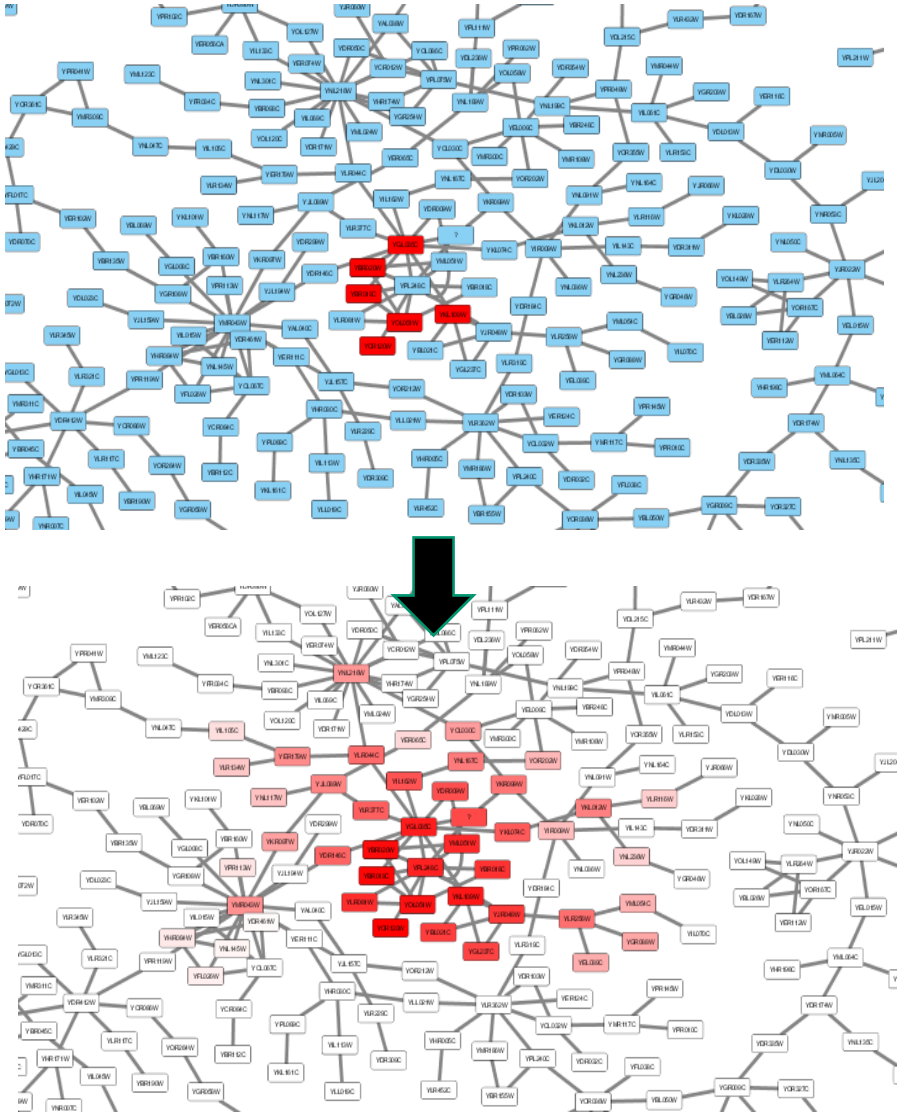


Cowen, L., *et al.* (2017) Network propagation: a universal amplifier of genetic associations. *Nature Rev Genetics.* **18,** 551.

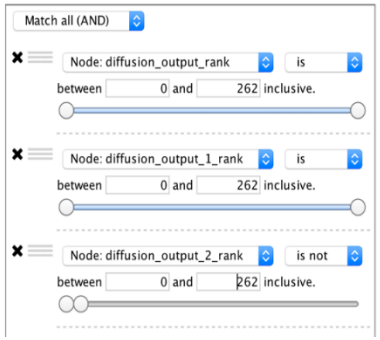
Identifying disease-related modules
Network propagation with Cytoscape

a.

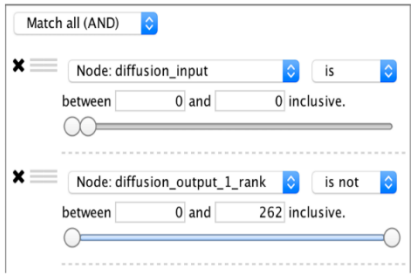




b.



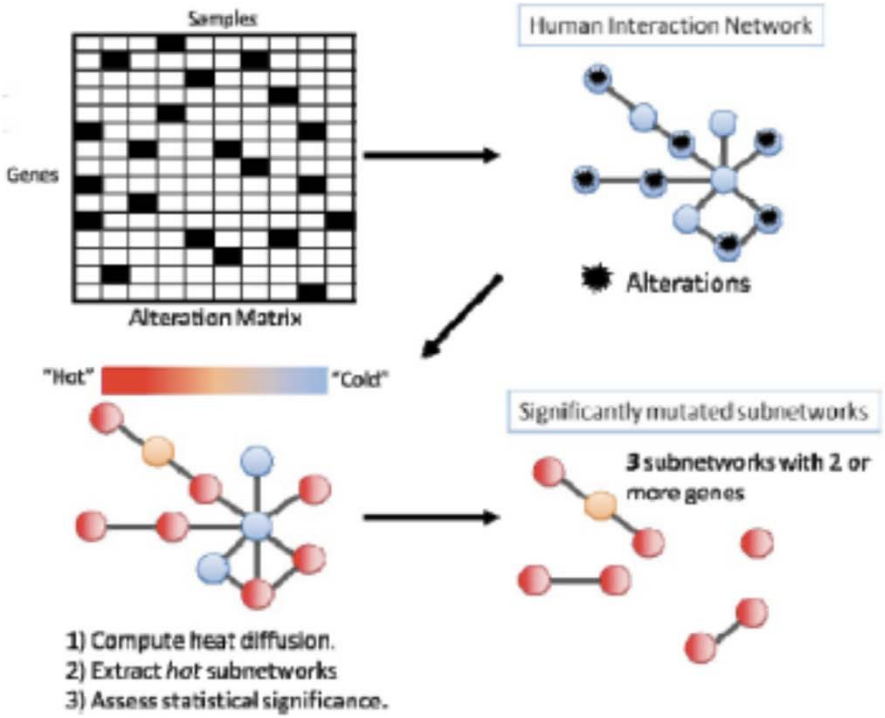
c.

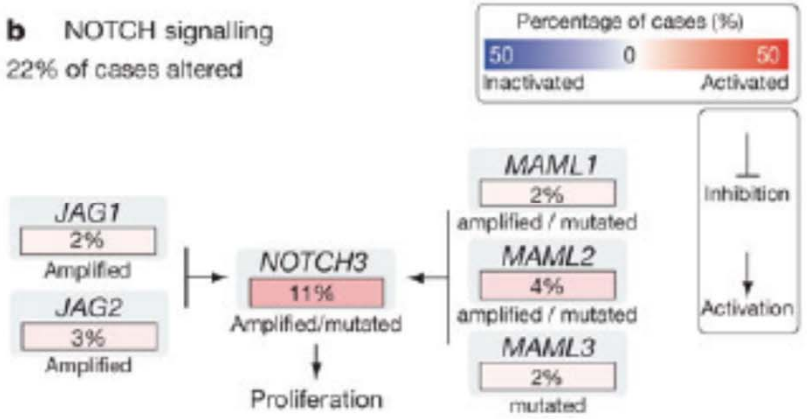


Carlin, D.E., et al. (2017) Network propagation in the cytoscape cyberinfrastructure. *PLOS Computational Biology*, **13(10),** e1005598.

Identifying disease-related modules
Network propagation

"HOTNET"

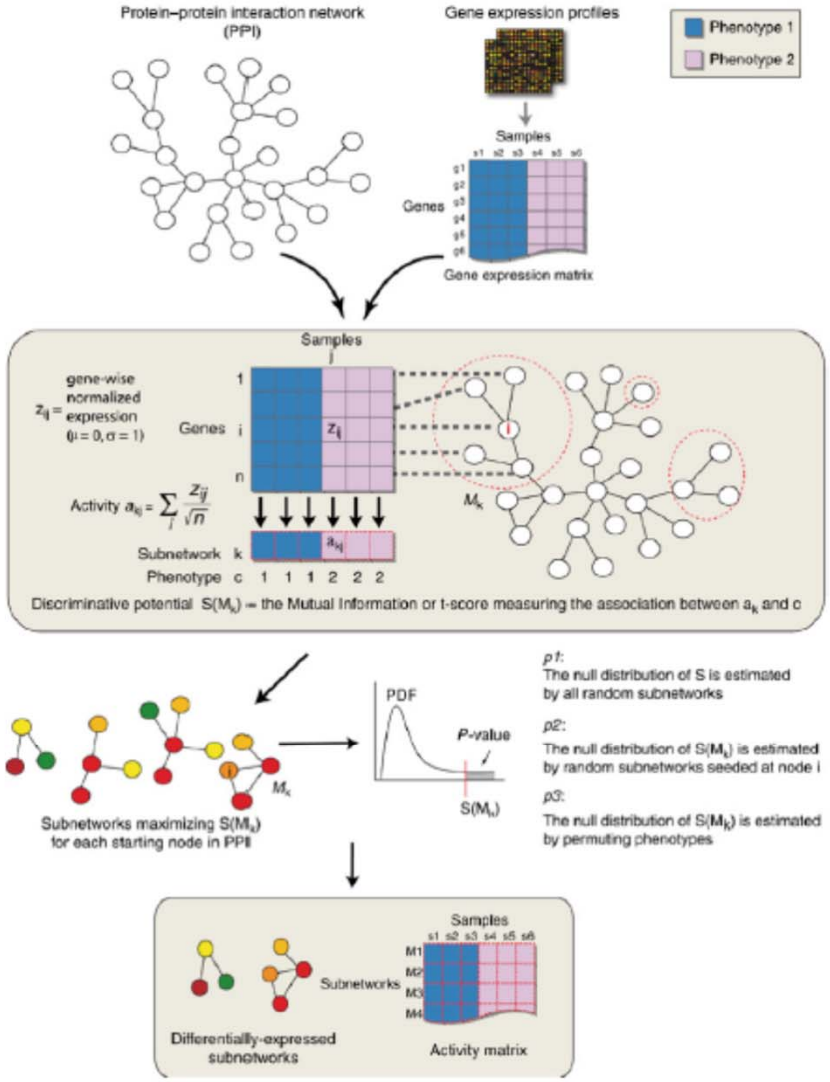


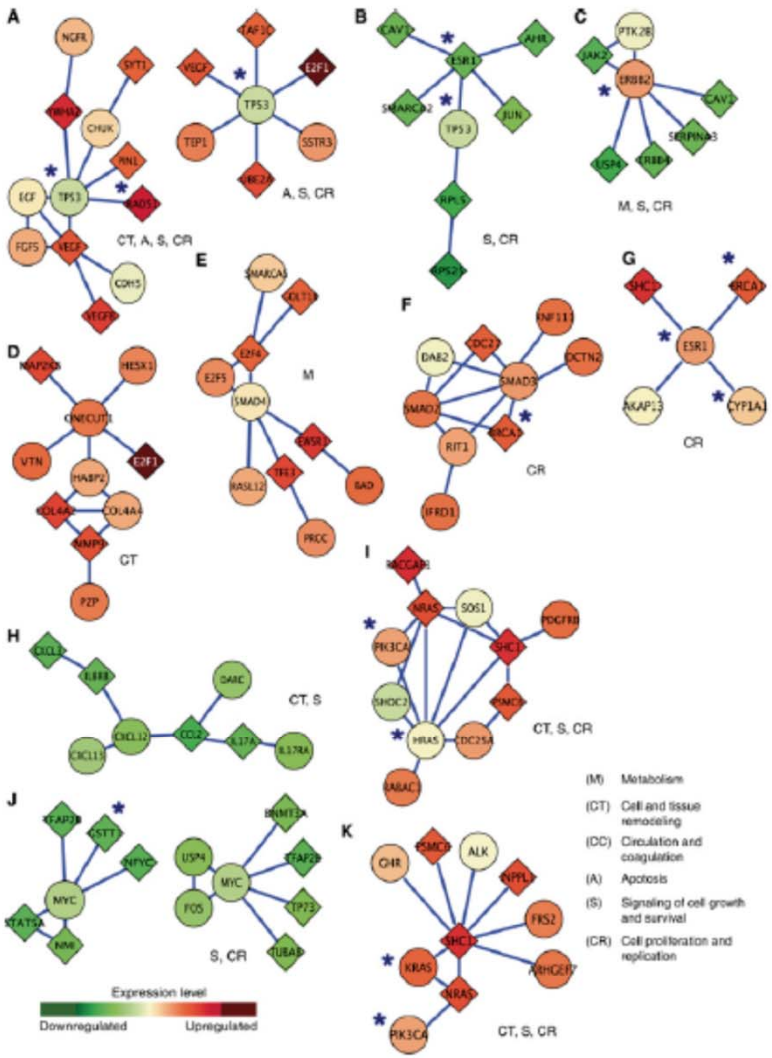


Vandin, F., Upfal, E. & Raphael, B. J. Algorithms for detecting significantly mutated pathways in cancer. J. Comput. Biol. 18, 507–522 (2011).

Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature 474, 609–615 (2011).

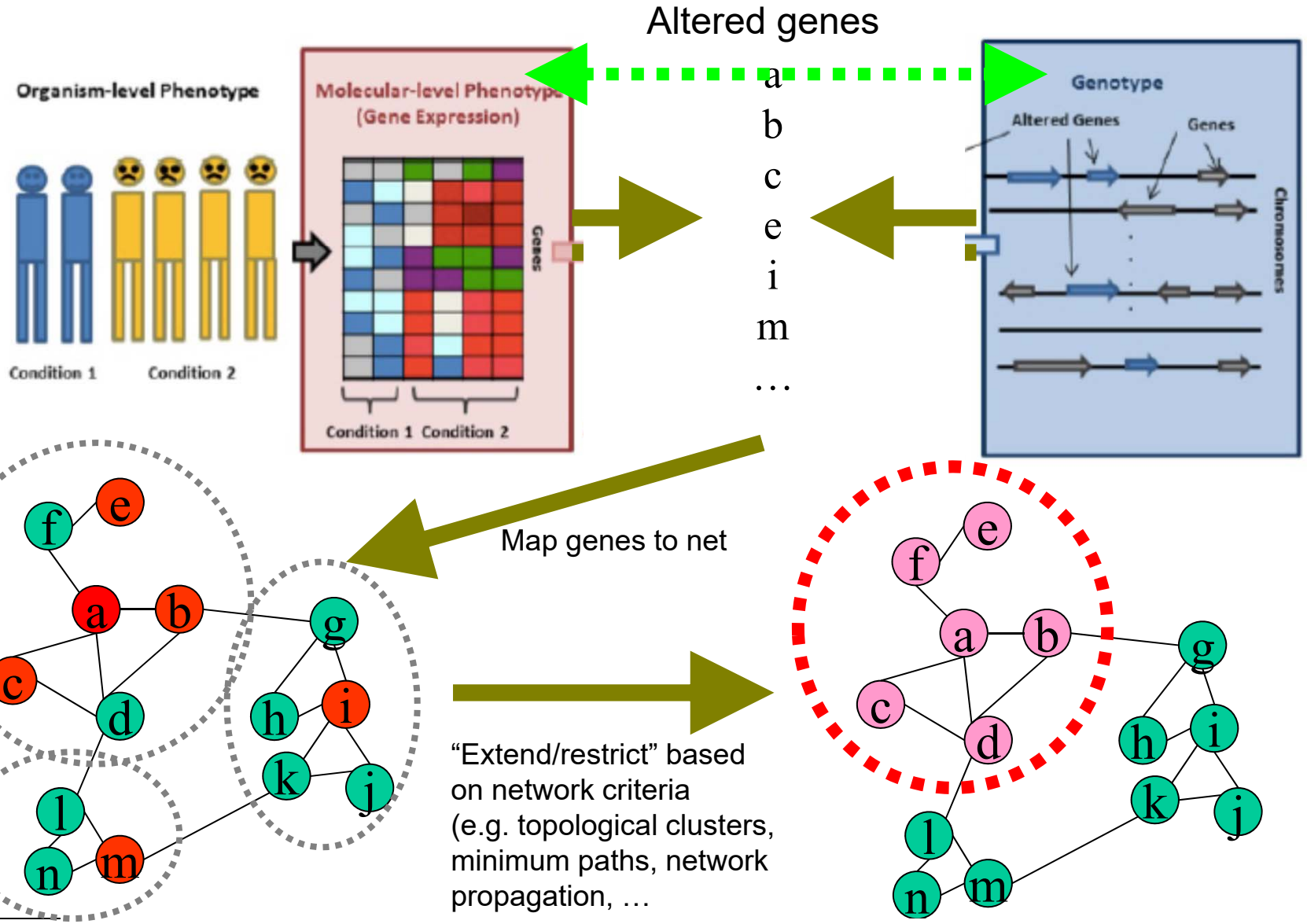
Identifying disease-related modules





Chuang, H.-Y., Lee, E., Liu, Y.-T., Lee, D. & Ideker, T. Network-based classification of breast cancer metastasis. Mol Syst Biol 3, 140 (2007).

Identifying disease-related modules
Combining genotypic and phenotypic alterations

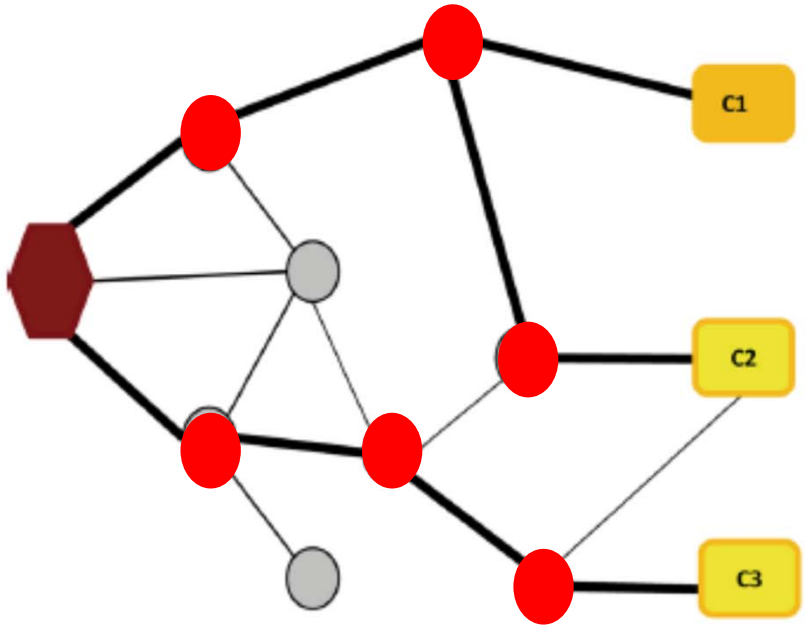


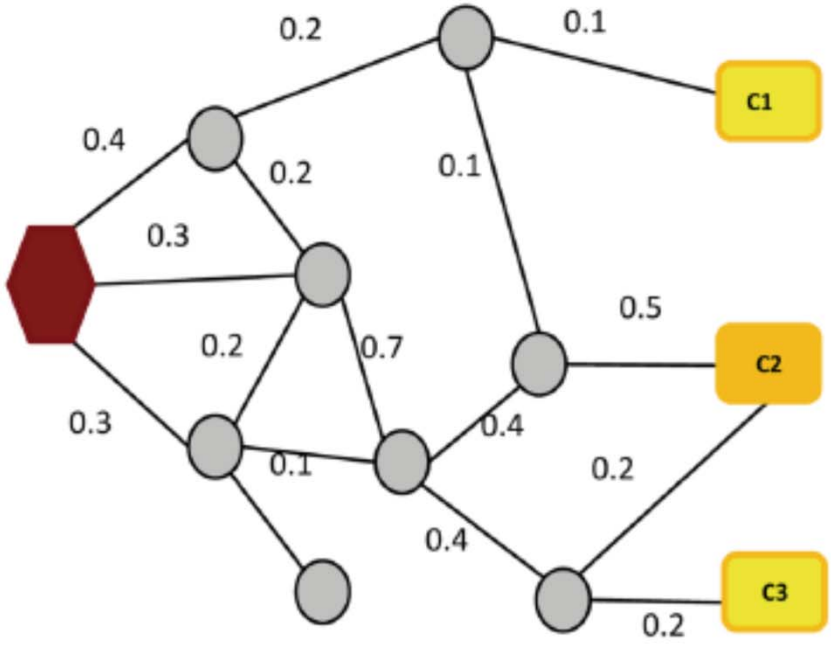
Kim, Y.-A. & Przytycka, T. M. Bridging the Gap between Genotype and Phenotype via Network Approaches. Front. Genet. 3,

Identifying disease-related modules
Combining genotypic and phenotypic alterations

Genotypic variation

Phenotypic variation



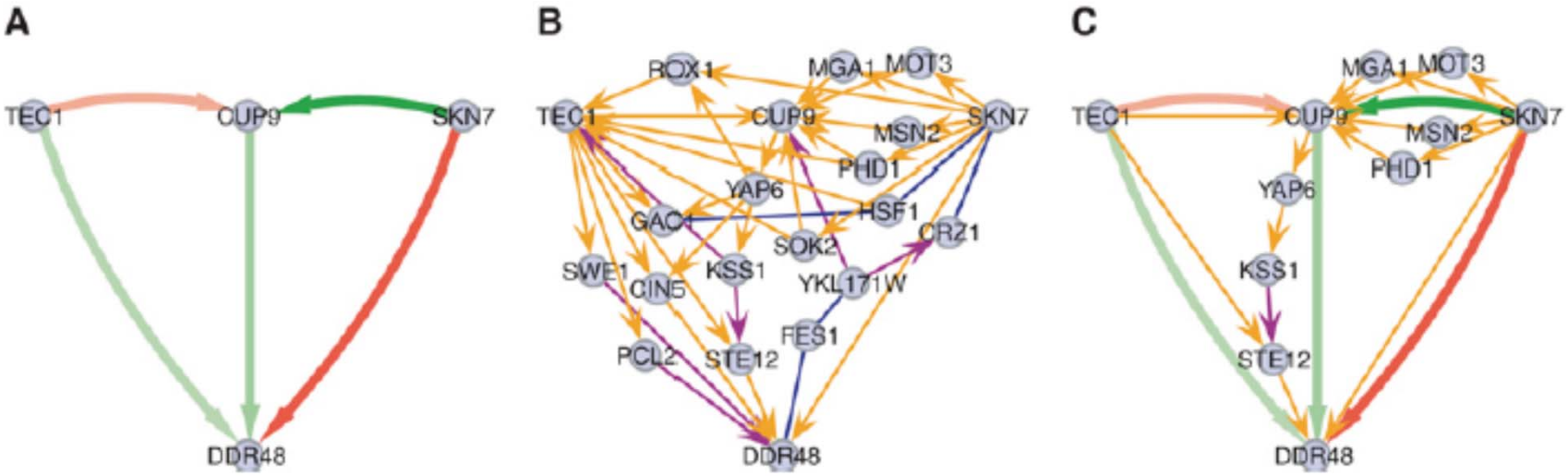


Distance-based

Information/"electric current"/flow -based

Cho D-Y, Kim Y-A, Przytycka TM (2012) Chapter 5: Network Biology Approach to Complex Diseases. PLoS Comput Biol 8(12): e1002820.

Identifying disease-related modules
Combining genotypic and phenotypic alterations
Distance-based approaches

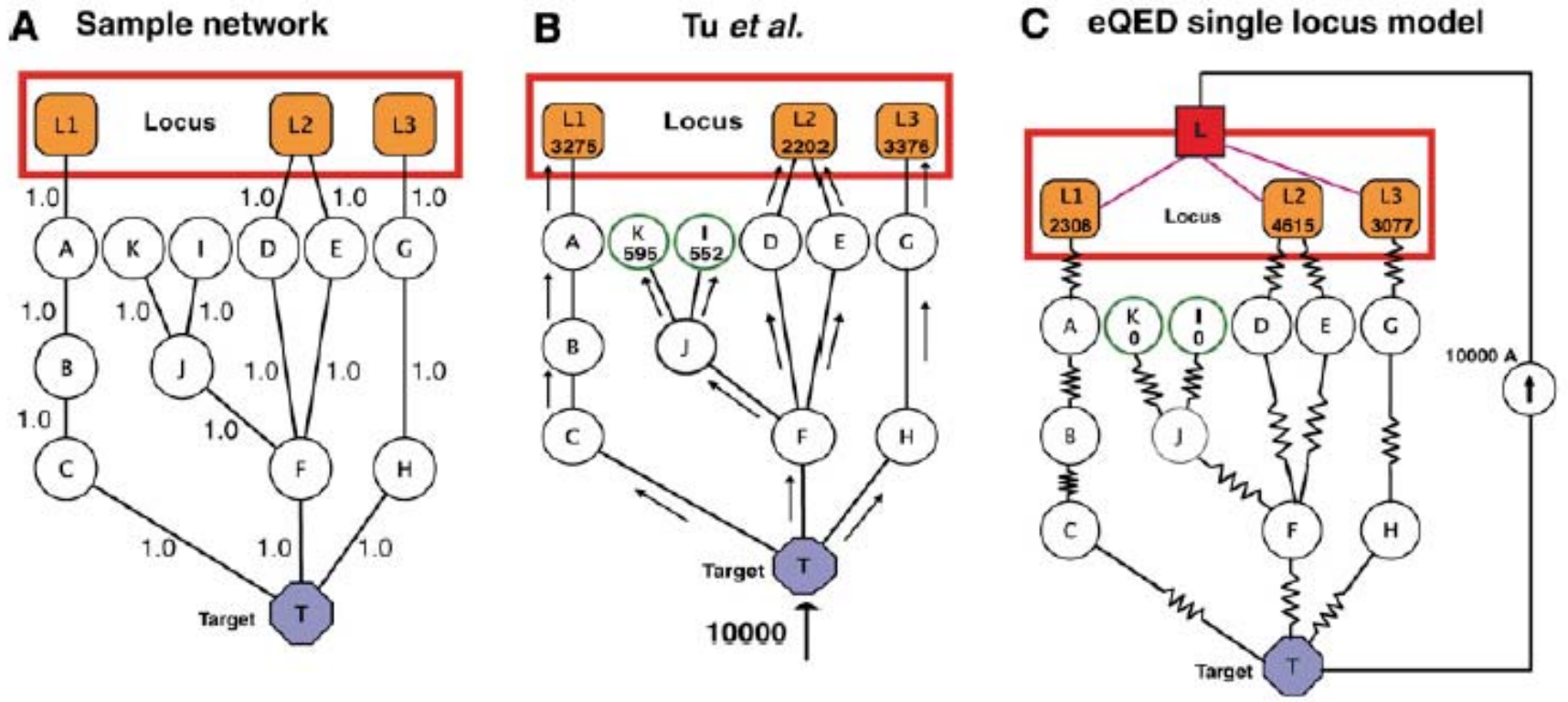


Look for (minimum) pathways of protein/genes which best explain the long-range relationships observed

---

Carter, G. W. et al. Prediction of phenotype and gene expression for combinations of mutations. Mol Syst Biol 3, 96 (2007).

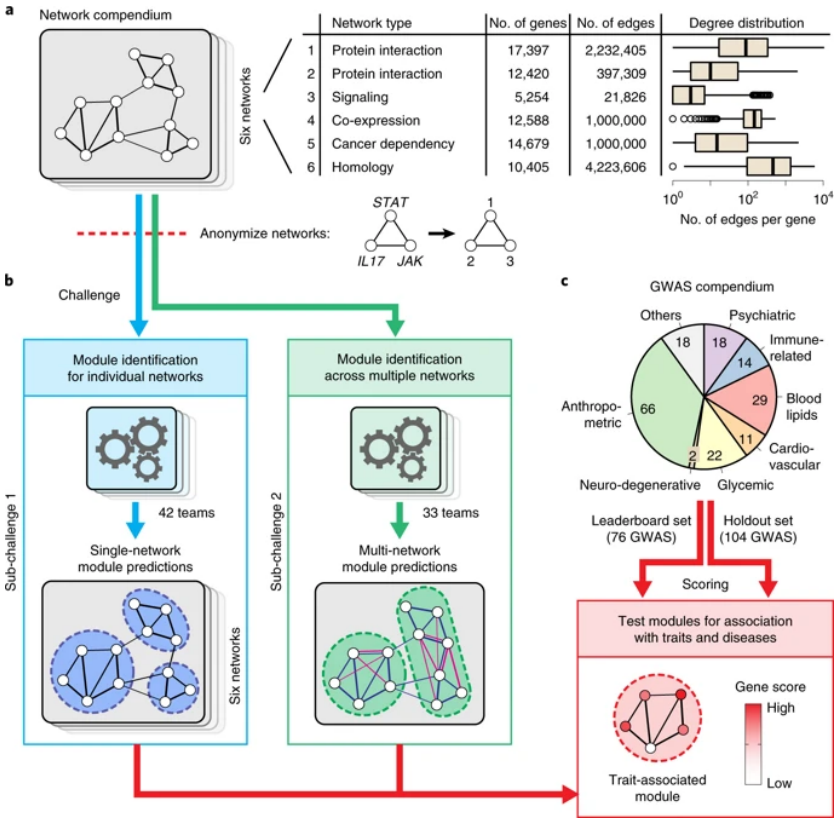
Identifying disease-related modules
Combining genotypic and phenotypic alterations
Flow-based approachess

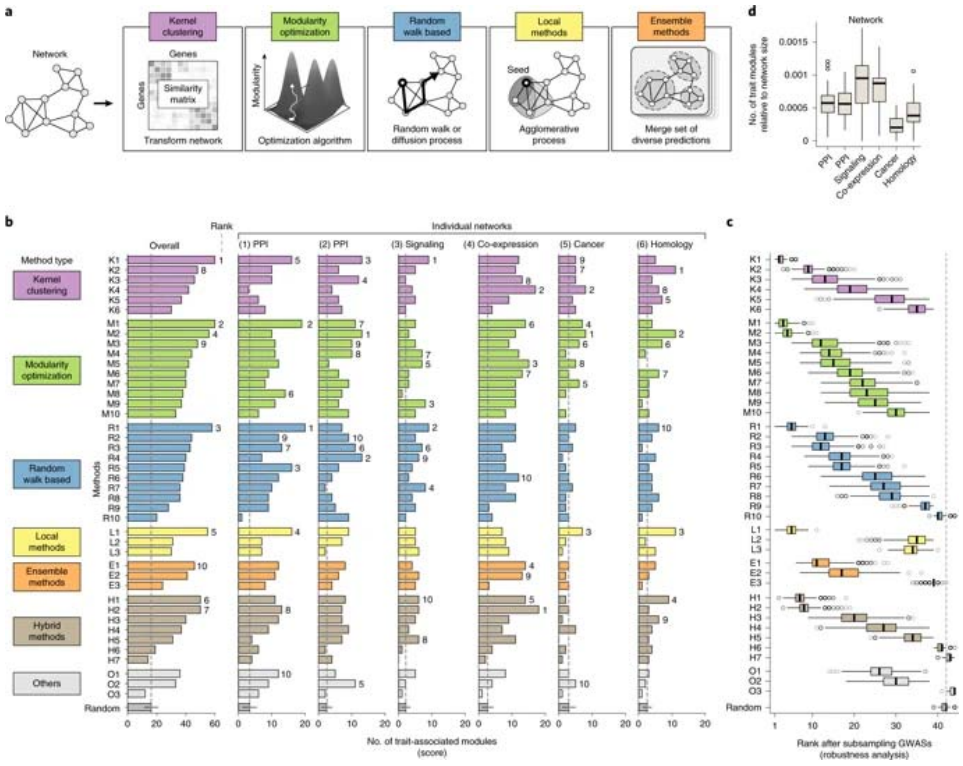


Tu, Z., Wang, L., Arbeitman, M. N., Chen, T. & Sun, F. An integrative approach for causal gene identification and gene regulatory pathway inference. Bioinformatics 22, e489–96 (2006).

Suthram, S., Beyer, A., Karp, R. M., Eldar, Y. & Ideker, T. eQED: an efficient method for interpreting eQTL associations using protein networks. M Syst Biol 4, 162 (2008).

Benchmarking disease module identification approaches

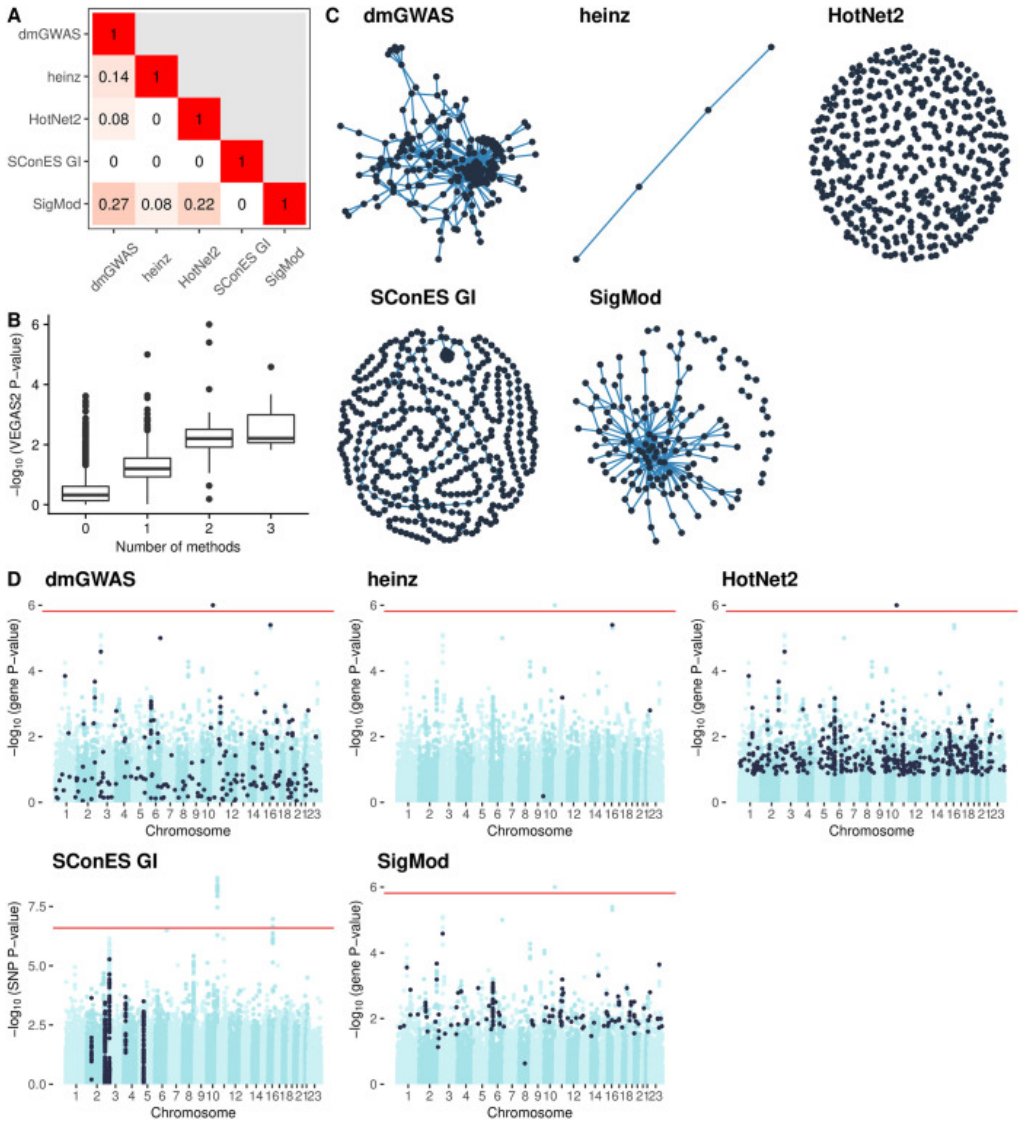




As new methods for identifying disease-associated network modules show up, different initiatives appeared to benchmark and compare them in a fair way, based on the same datasets, etc.

Choobdar, S., *et al.* (2019). Assessment of network module identification across complex diseases. *Nature Methods.* **16(9):**843-852.

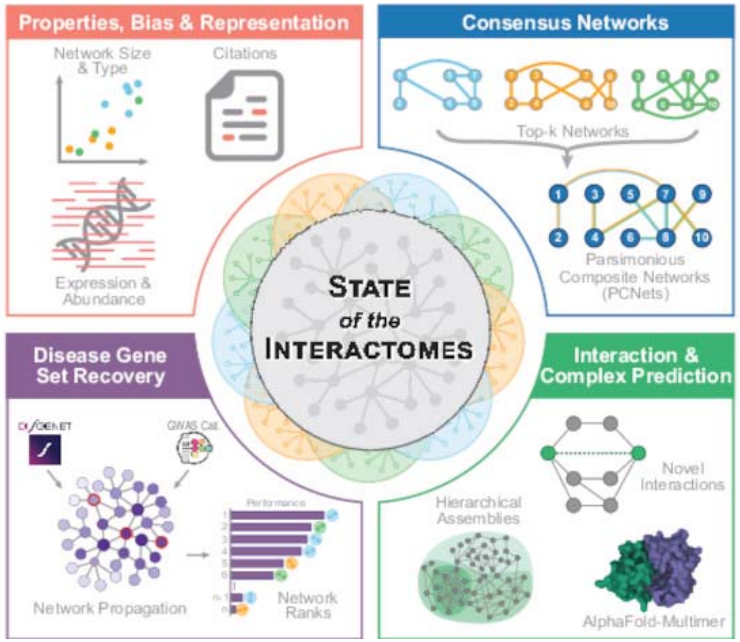
Benchmarking network-based approaches for filtering GWAS data

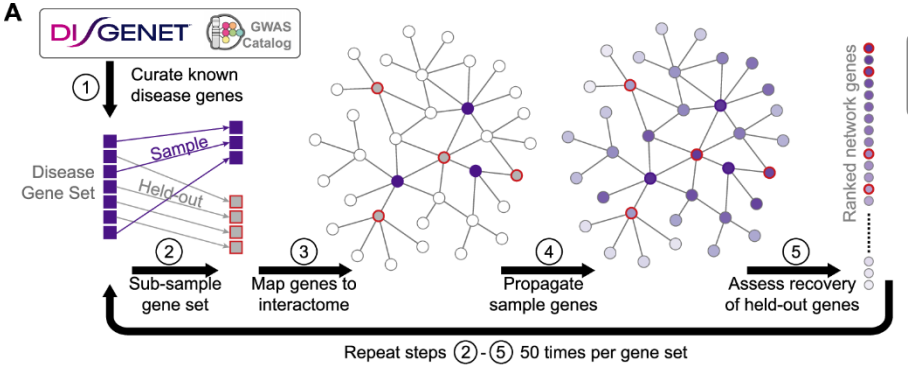


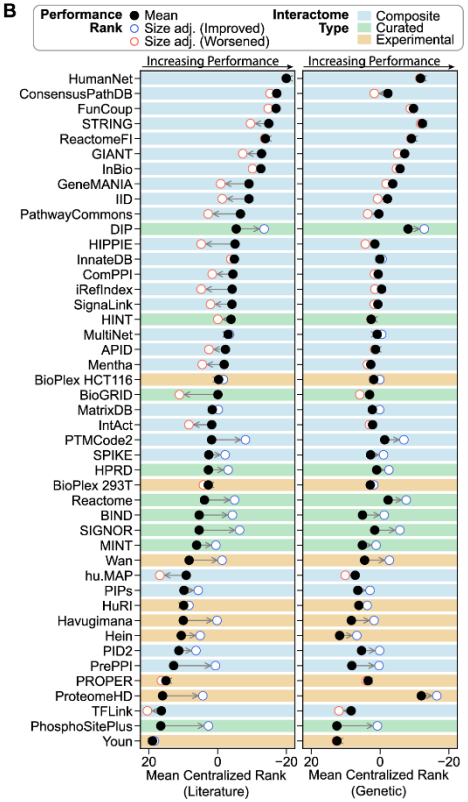
- Network methods recovered more interpretable results than a standard GWAS.
- Consensus among methods yield better results

Climente-González H, Lonjou C, Lesueur F; GENESIS study group; Stoppa-Lyonnet D, Andrieu N, Azencott CA. (2021). Boosting GWAS using biological networks: A study on susceptibility to familial breast cáncer. *PLoS Comput Biol.* **17(3):** e1008819.

Comparing interactomes used for network-based priorization







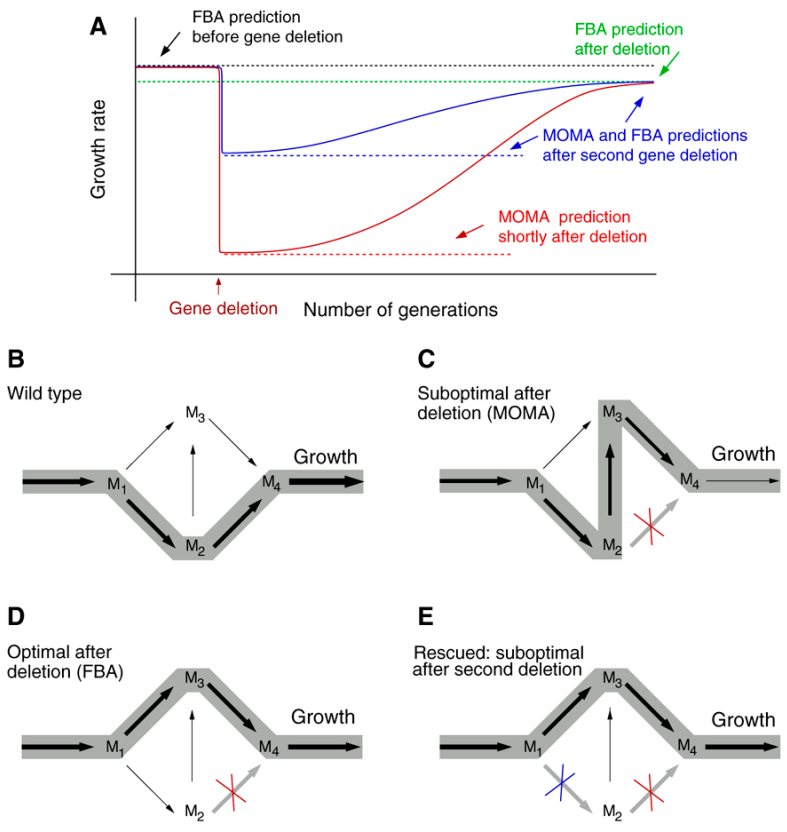
Not only the methods. The interactomes used for network-based priorization have to be benchmarked too.

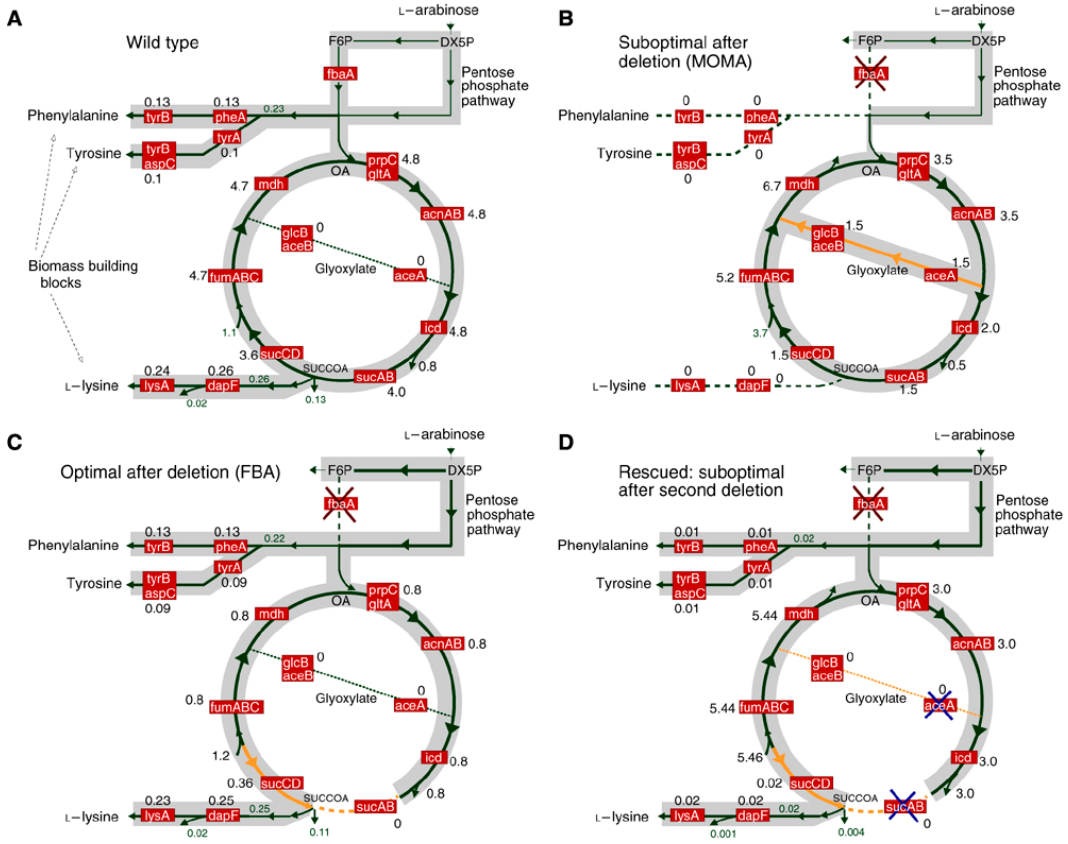
Large <u>composite</u> networks (e.g STRING) remain the most powerful for disease gene prioritization

---

Wright SN, Colton S, Schaffer LV, Pillich RT, Churas C, Pratt D, Ideker T. (2025). State of the interactomes: an evaluation of molecular networks for generating biological insights. *Mol Syst Biol*. **21(1):**1-29

Design therapeutic strategies with networks in mind



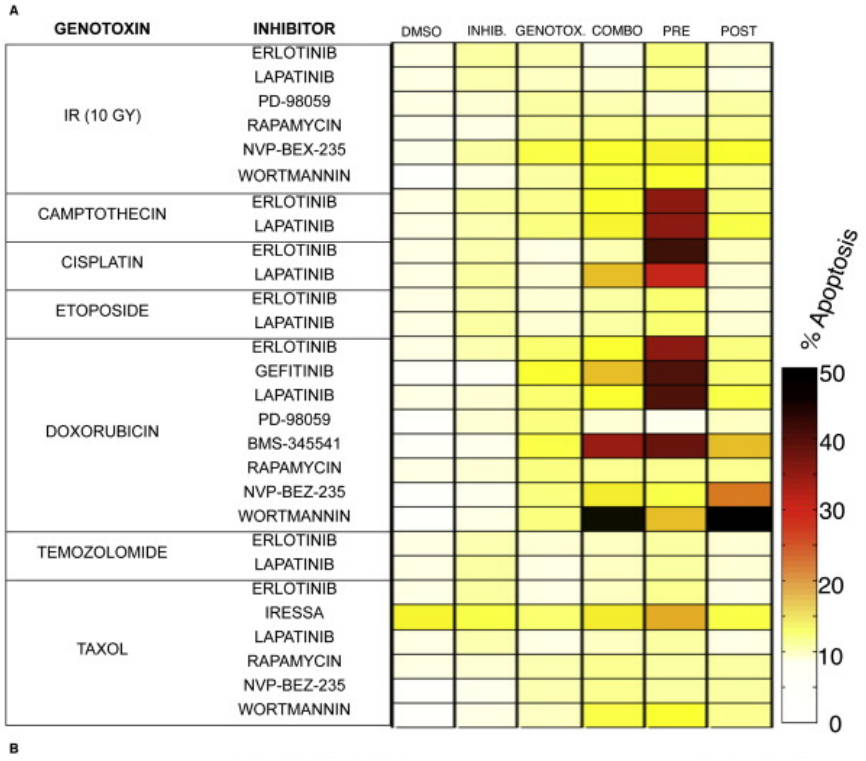


Sometimes it is better to break than trying to repair

Motter, A.E., Gulbahce, N., Almaas, E. and Barabasi, A.L. (2008) Predicting synthetic rescues in metabolic networks. *Mol Syst Biol.*, **4**, 168.

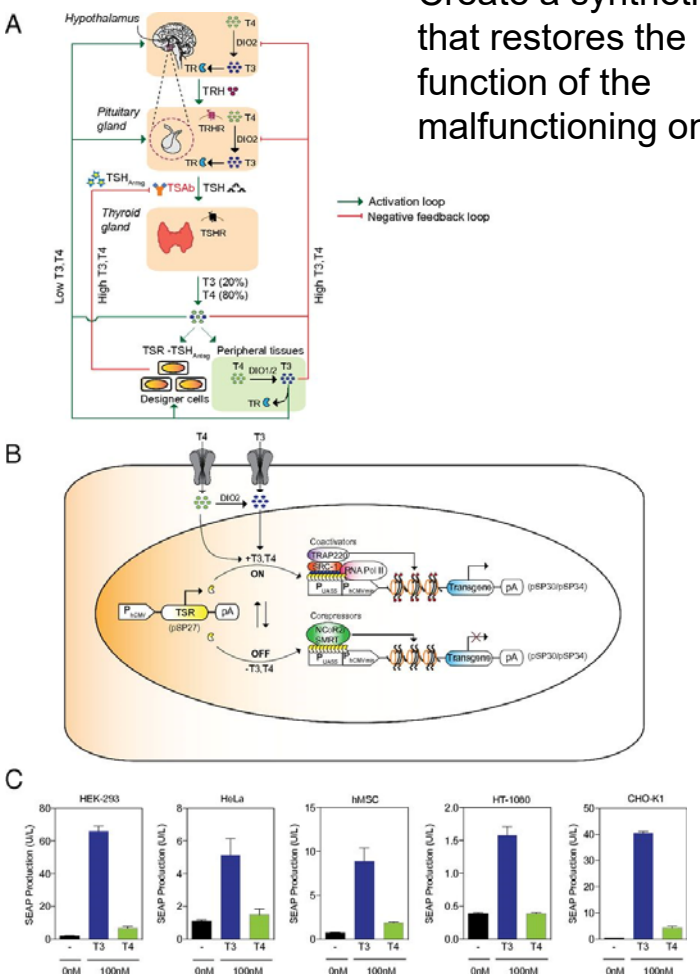
Design therapeutic strategies with networks in mind

Select which nodes to touch (inhibit with drug) in the network to get the desired re-wiring



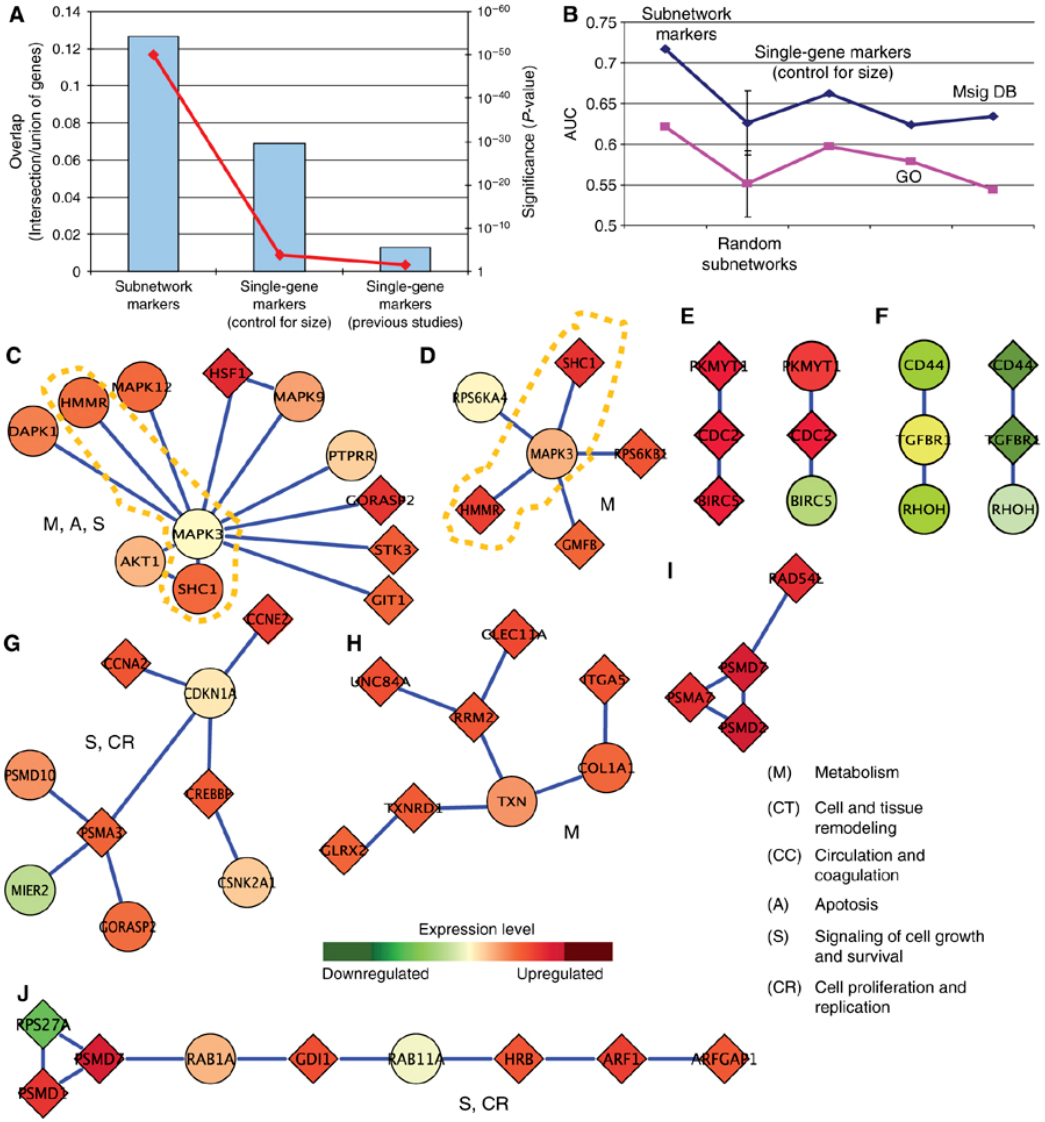
Lee MJ, Ye AS, Gardino AK, Heijink AM, Sorger PK, et al. **2012**. Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* 149(4): 780–94

Create a synthetic net that restores the function of the malfunctioning one



Saxena P, Charpin-El Hamri G, Folcher M, Zulewski H, Fussenegger M. **2016**. Synthetic gene network restoring endogenous pituitary-thyroid feedback control in experimental Graves' disease. *PNAS* 113(5): 1244–49

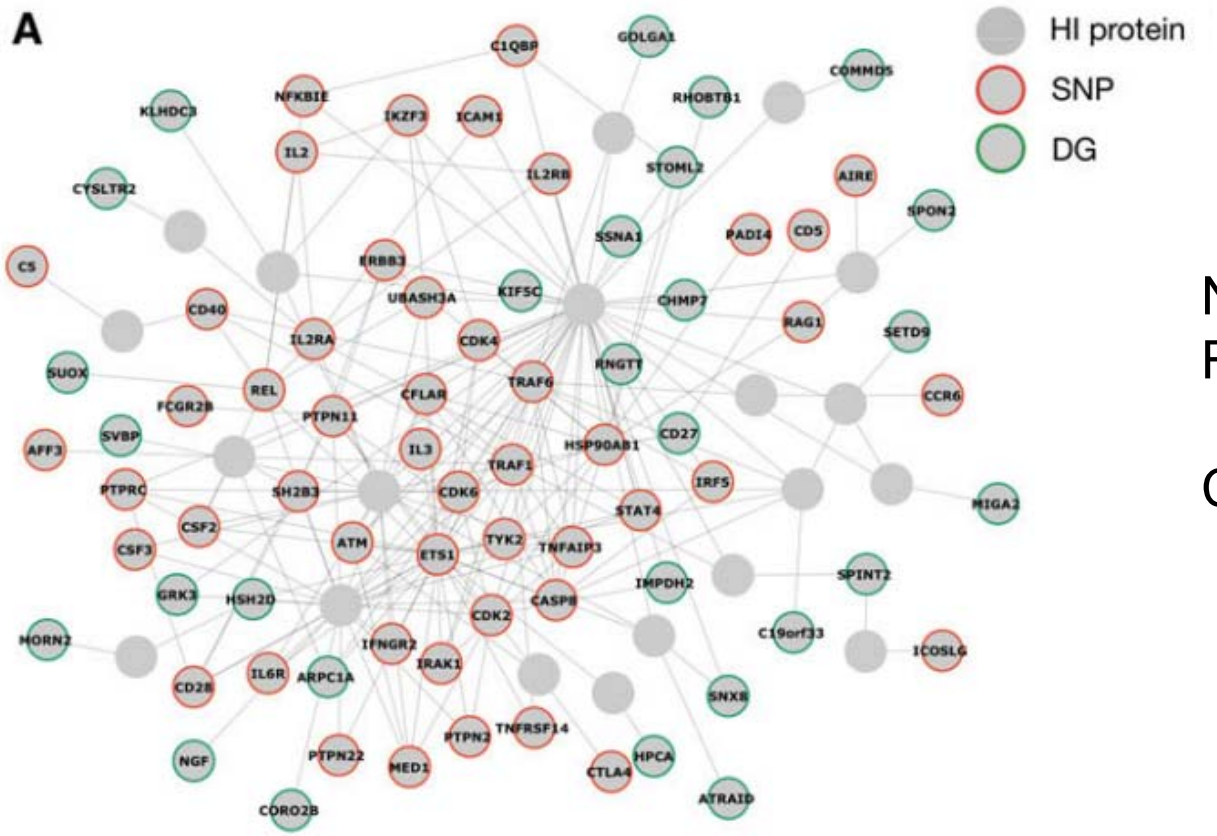
Biomarkers, for diagnosis/prognosis



Networks are betters markers of an alteration/disease than single genes or combinations of them

Chuang, H.Y., Lee, E., Liu, Y.T., Lee, D. and Ideker, T. (2007) Network-based classification of breast cancer metastasis. *Mol Syst Biol.*, **3**, 140.

Biomarkers, for diagnosis/prognosis

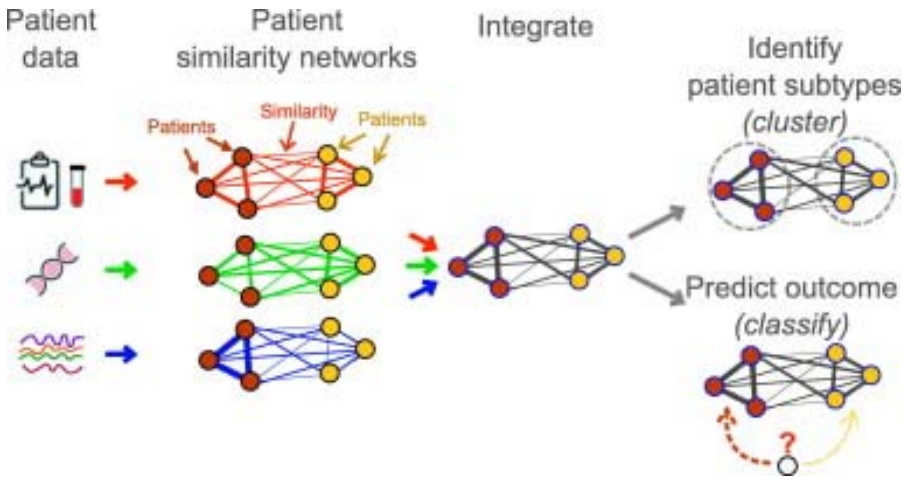


Network-based test for Rheumatoid arthritis.

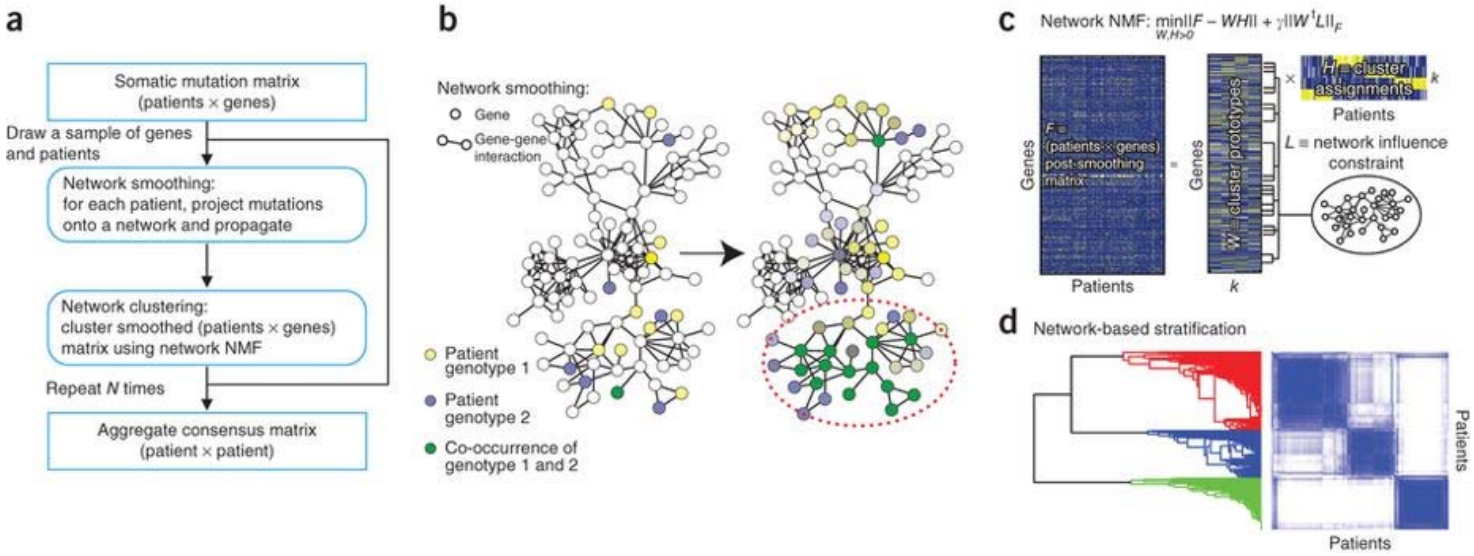
Clinically tested

Mellors, Theodore, et al. Clinical validation of a blood-based predictive test for stratification of response to tumor necrosis factor inhibitor therapies in rheumatoid arthritis patients. (2020). *Network and Systems Medicine.* **3.1:** 91-104.

Network-based patient stratification (personalized medicine)



Patient networks can be used to identify patient subgroups which, for example, present different responses to drugs.



Pai S, Bader GD. Patient Similarity Networks for Precision Medicine (2018). *J Mol Biol*. 430 (18 Pt A):2924-2938.

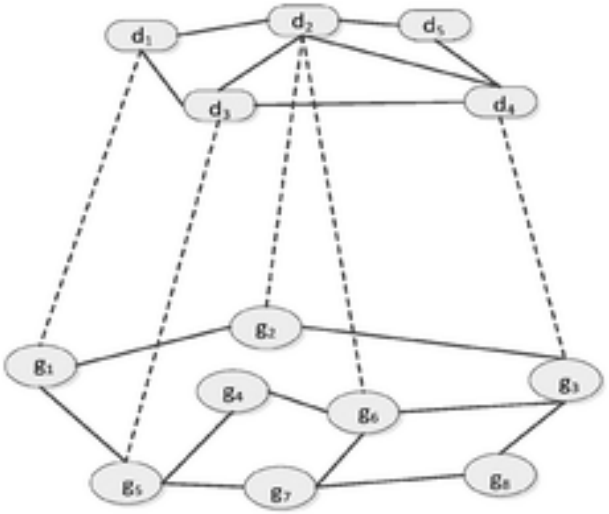
Hofree M, Shen JP, Carter H, Gross A, Ideker T. (2013). Network-based stratification of tumor mutations. *Nat. Methods* 10(11): 1108–15

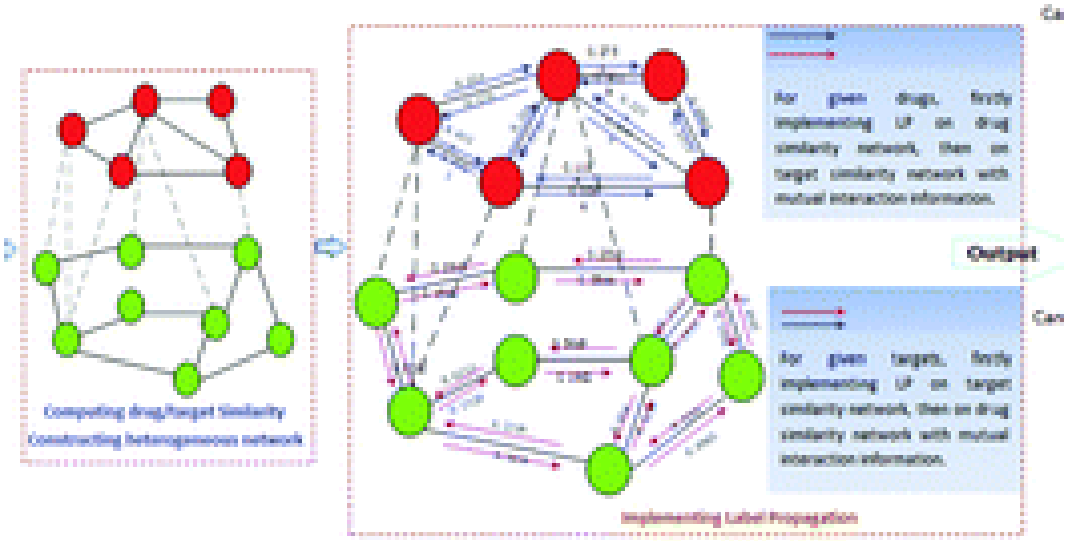
Drug repositioning

It often takes 10–15 years of research and 0.8–1.5 billion dollars to bring a drug from abstract concept to market-ready product [1]. Every year, ~90% of drugs fail during FDA evaluations, preventing their use in actual therapy.

Drug repositioning (DR, aka d. repurposing, redirecting, retargeting, …) seeks to find new uses for existing drugs, with established and demonstrated human safety. Usually this involves finding new targets for approved drugs.

Computational methods for DR include molecular docking, machine learning, literature mining, … and network-based approaches.

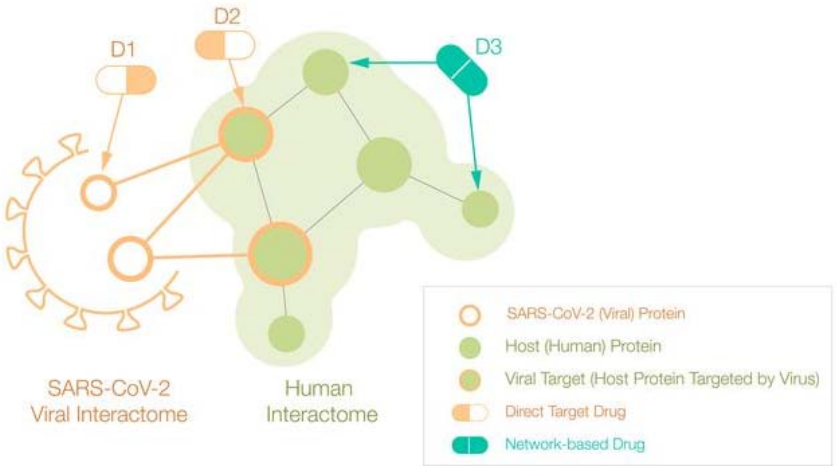




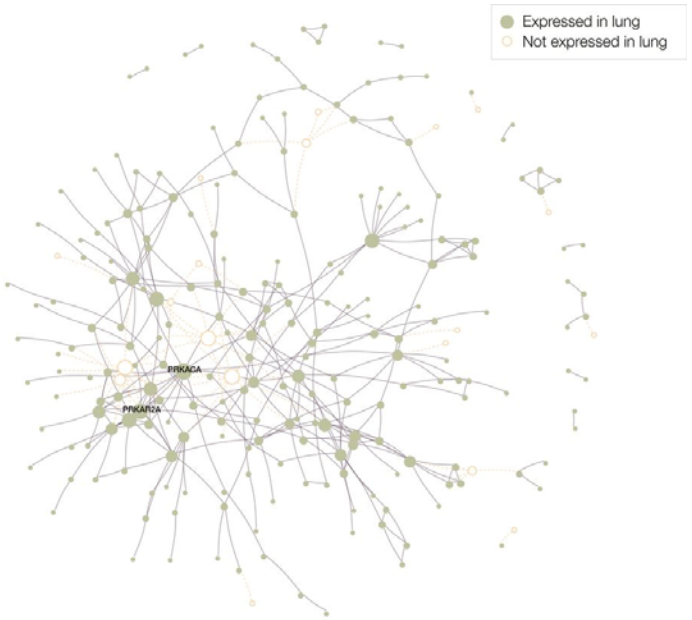
Maryam Lotfi Shahreza, Nasser Ghadiri, Sayed Rasoul Mousavi, Jaleh Varshosaz, James R Green. (2018). A review of network-based approaches to drug repositioning, *Briefings in Bioinformatics*, **19(5):** 878–892,

Xiao-Ying Yan,ab Shao-Wu Zhang and Song-Yao Zhanga (2016). Prediction of drug–target interaction by label propagation with mutual interaction information derived from heterogeneous network. *Mol. BioSyst*.,**12,** 520-531

Drug repositioning COVID-19



A

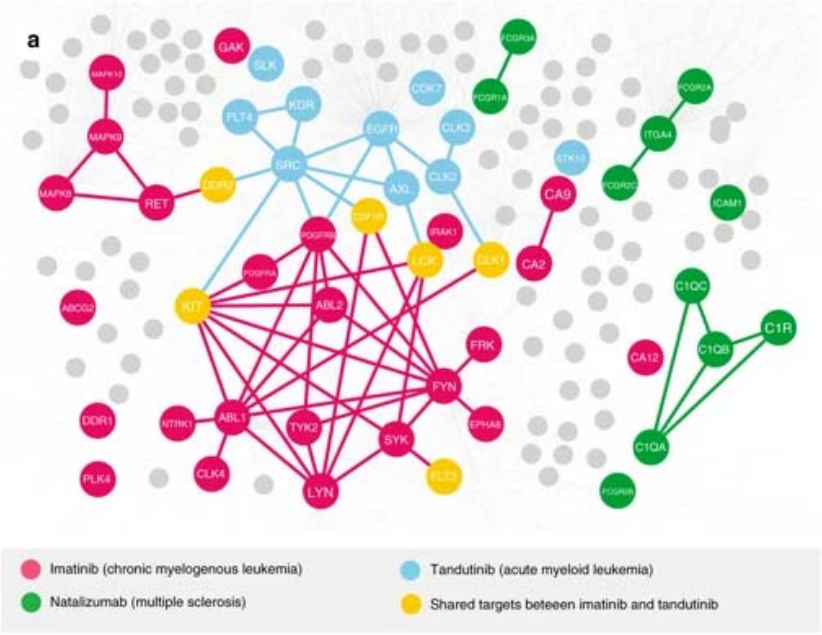


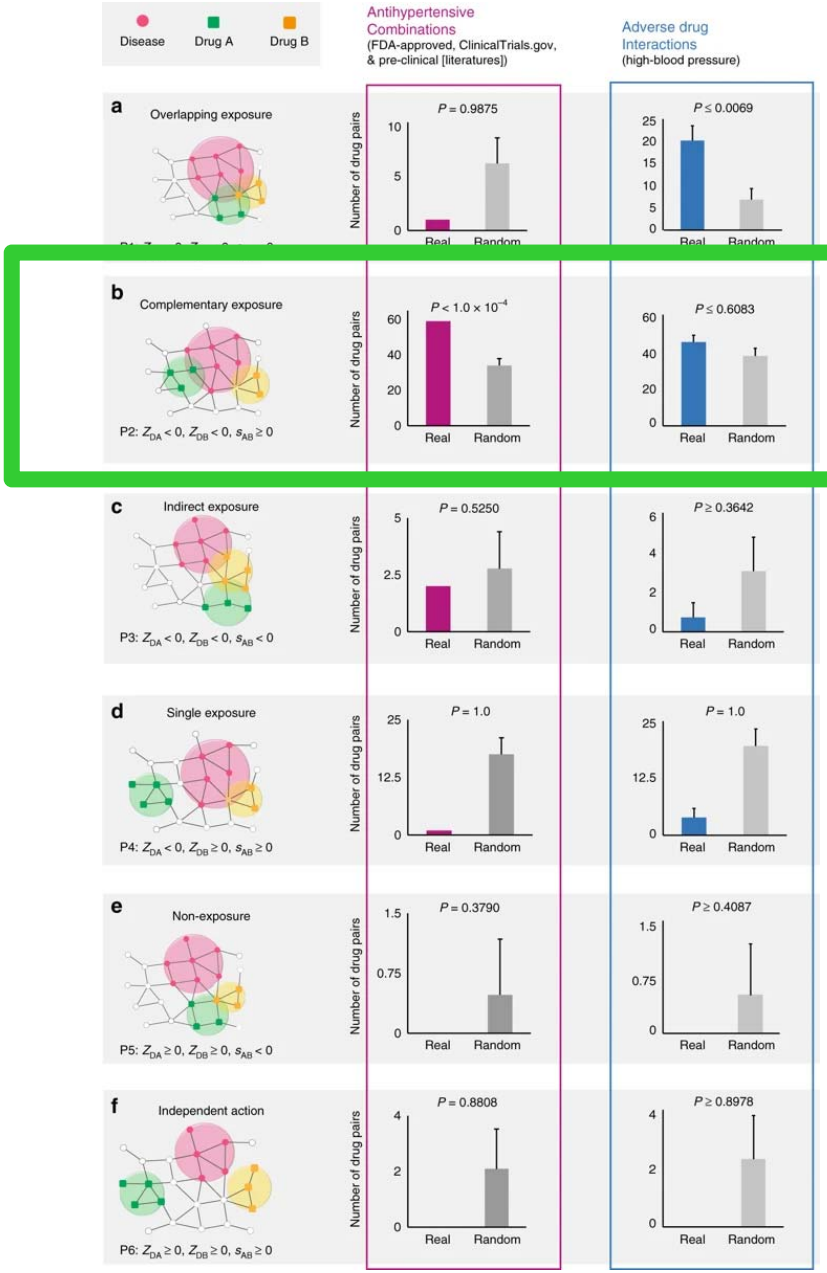
"We deployed algorithms relying on artificial intelligence, network diffusion, and network proximity […] to rank 6,340 drugs for their expected efficacy against SARS-CoV-2. We screened in human cells the top-ranked drugs, obtaining a 62% success rate. Of the six drugs that reduced viral infection, four could be directly repurposed to treat COVID-19. We also found that 76 of the 77 drugs that successfully reduced viral infection do not bind the proteins targeted by SARS-CoV-2, **indicating that these network drugs rely on network-based mechanisms that cannot be identified using docking-based strategies**.

Morselli Gysi, Deisy; do Valle, Ítalo; Zitnik, Marinka; Ameli, Asher; Gan, Xiao; Varol, Onur; Ghiassian, Susan Dina; Patten, J J; Davey, Robert A; Loscalzo, Joseph; Barabási, Albert-László. (2021). Network medicine framework for identifying drug-repurposing opportunities for COVID-19. *Proc Natl Acad Sci U S A.* **118(19):**e2025581118.

Drug combinations

Use networks to predict pairs of (approved) drugs with synergistic effects





"for hypertension and cancer, we find that only one of the six classes correlates with therapeutic effects: if the targets of the drugs both hit disease module, but target separate neighborhoods."

Cheng, F., Kovács, I.A. and Barabási, A.-L. (2019) Network-based prediction of drug combinations. *Nature communications*, **10(1),** 1197.

Other Networks used for studying human pathologies

Networks specifically generated for studying human pathologies

- Drug-target interactions (DTI)
- Drug-drug interactions (based on chemical similarity, biological effect similarity, target(s) similarity, …)
- Drug-disease associations
- Drug-side effect associations
- Disease-disease associations
- Integrated (combined) networks
- ….

Apart from the “generic” molecular networks (interactome, metabolome, gene regulation, …) some networks were generated to specifically represent pathology-related information

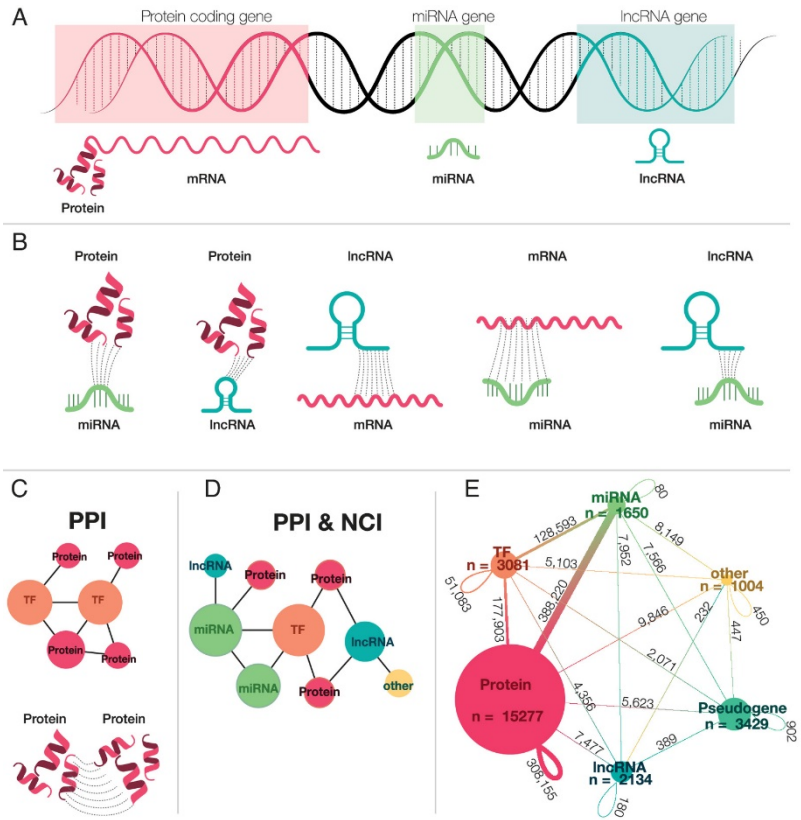
Other networks used for studying human pathologies

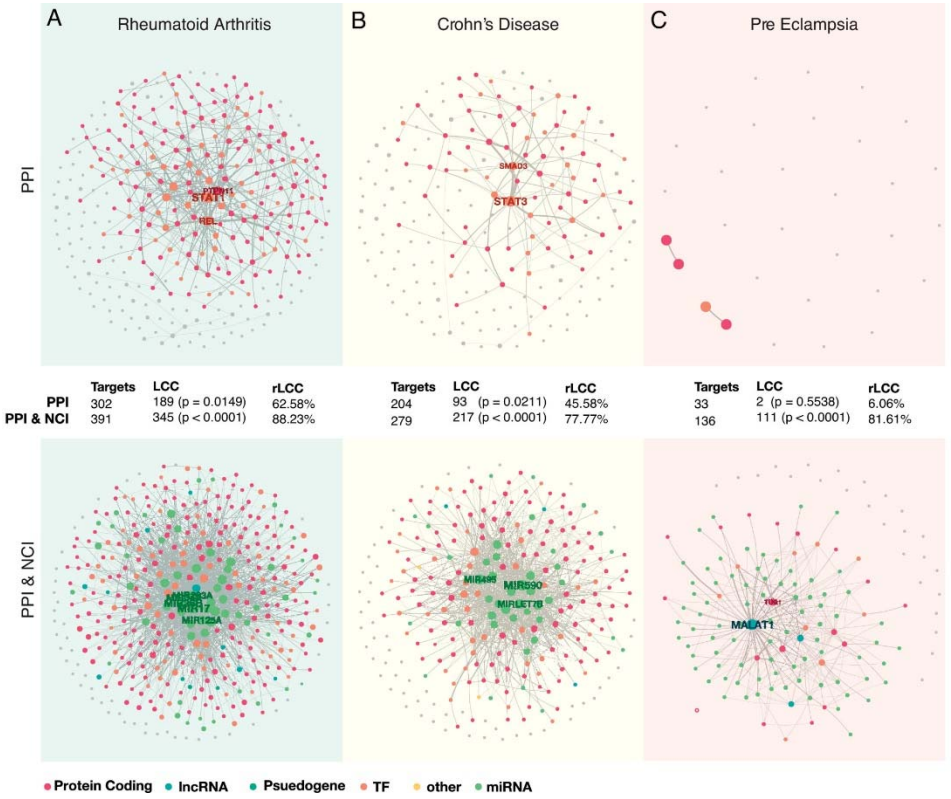
- Social Networks
- Technological networks
- ….

Maryam Lotfi Shahreza, Nasser Ghadiri, Sayed Rasoul Mousavi, Jaleh Varshosaz, James R Green. (2018). A review of network-based approaches to drug repositioning, *Briefings in Bioinformatics*, **19(5):** 878–892,

Integrated networks (aka. multipartite)

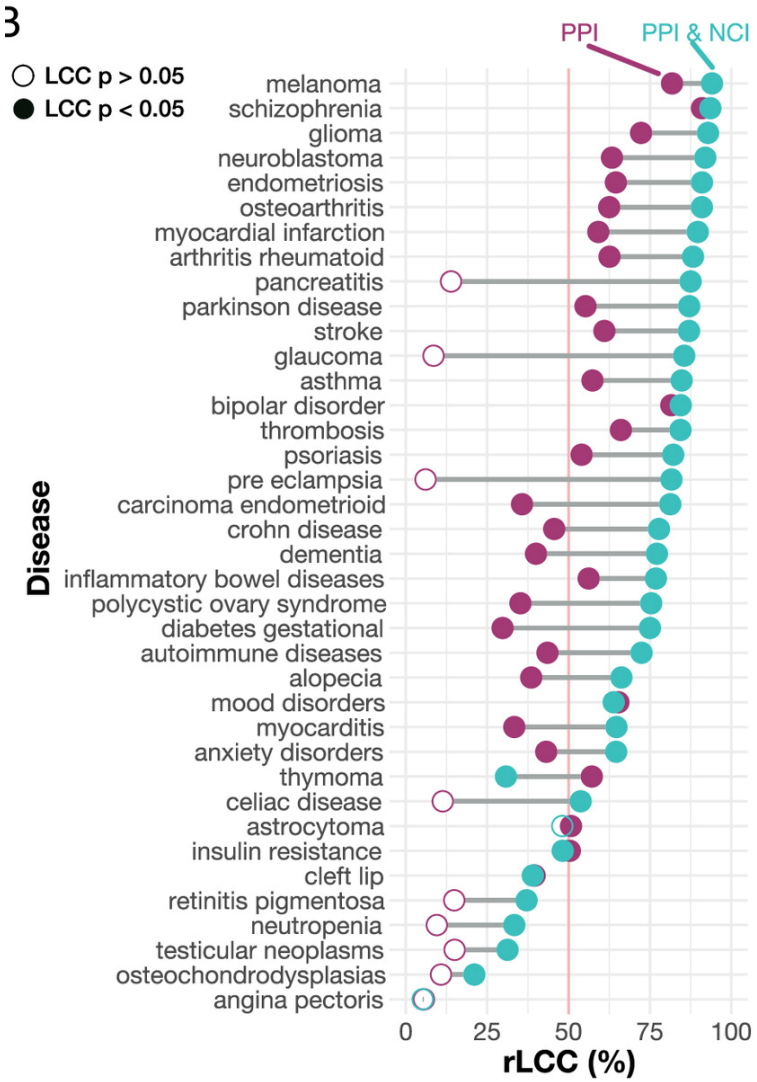
Including non-coding RNAs (ncRNA) in the network





Deisy Morselli & Albert-László Barabási. (2023). Noncoding RNAs improve the predictive power of network medicine. *Proc. Natl. Acad. Sci. USA*.**120(45):**e2301342120

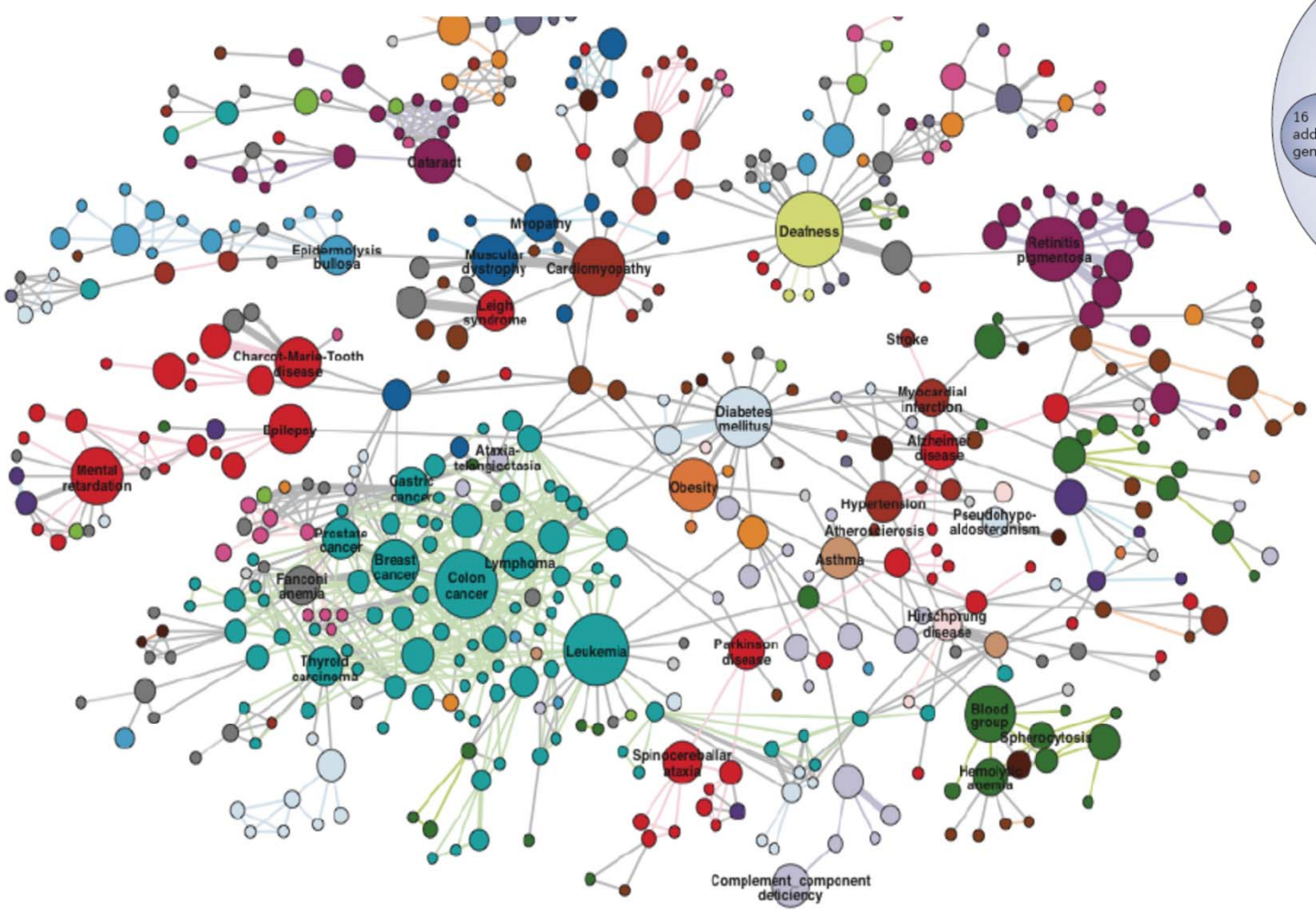
Integrated networks

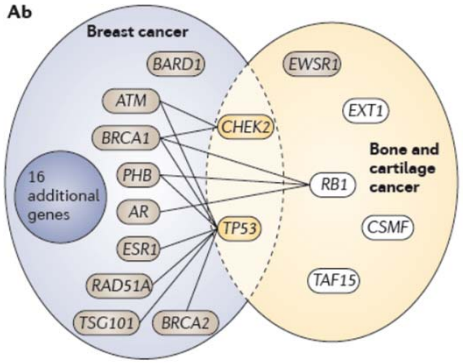


Including non-coding RNAs (ncRNA) in the network improve the identification of disease-related modules and genes

Deisy Morselli & Albert-László Barabási. (2023). Noncoding RNAs improve the predictive power of network medicine. *Proc. Natl. Acad. Sci. USA*.**120(45):**e2301342120

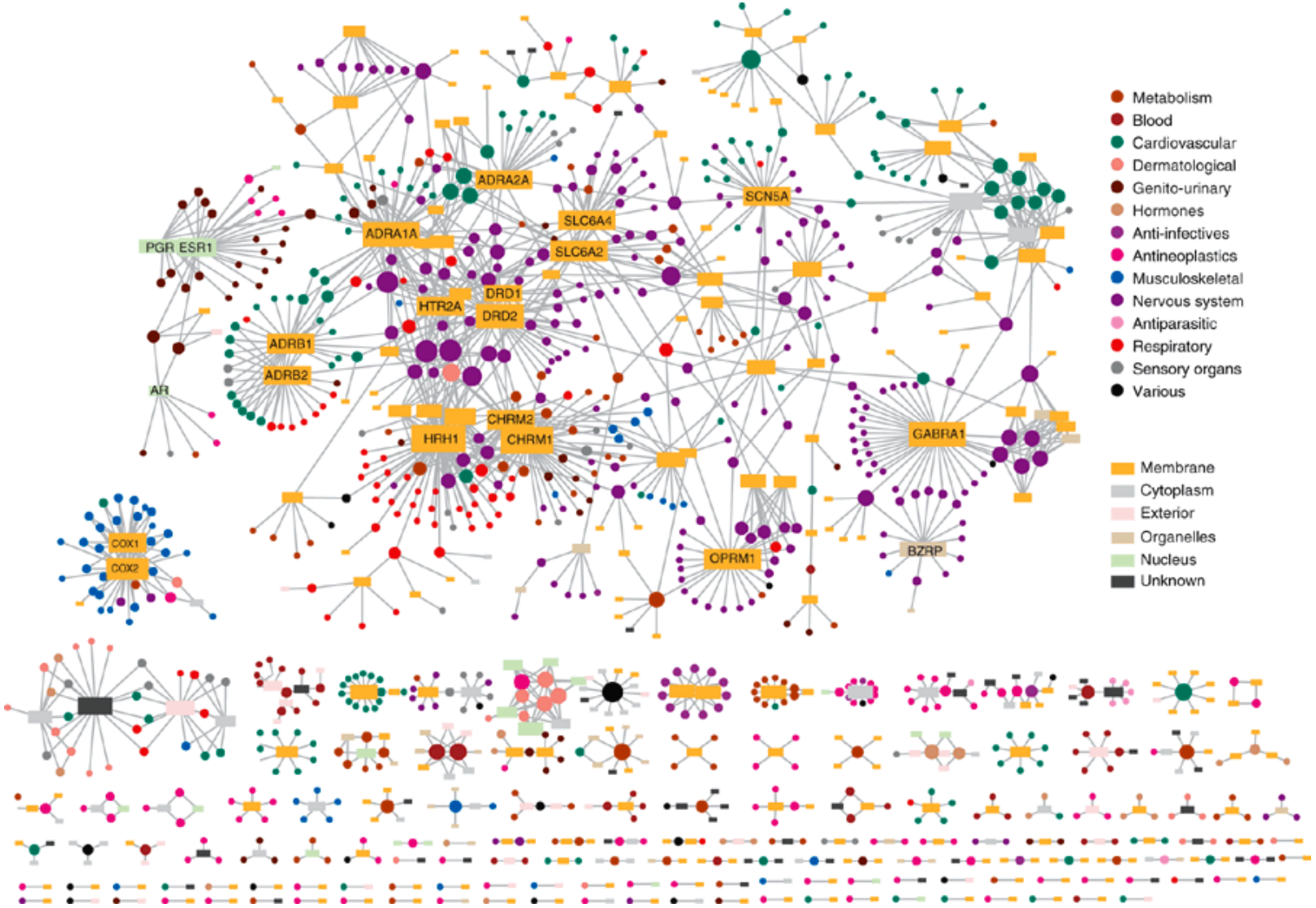
Disease network





Goh, K.-I. *et al.* The human disease network. *Proc Natl Acad Sci USA* **104**, 8685–8690 (2007).

“network pharmacology”
Drug-target network



Yildirim, M.A., Goh, K.I., Cusick, M.E., Barabasi, A.L. and Vidal, M. (2007) Drug-target network. *Nat Biotechnol.*, **25**, 1119-1126.

Summary

• Any phenomenon that can be modelled as entities linked by relationships can be represented as a network and studied with the tools of Graph Theory.

• Many phenomena in different disciplines have been studied from this point of view, including molecular phenomena as soon as the data required to assemble these networks became available.

• Biological (molecular) networks are the prototypic subject of study of Systems Biology, that focus on the complex network of relationships between components instead of the components themselves, hence complementing the reductionist approach of Molecular Biology.

• From the topological features of molecular networks, biological information related to their functioning, resistance to perturbations, ability to transmit information, etc. has been extracted.

• These network approaches are now being used to study and treat human pathologies.

Bibliography

- van Regenmortel, M.H. (2004) Reductionism and complexity in molecular biology. Scientists now have the tools to unravel biological and overcome the limitations of reductionism. *EMBO Rep,* **5(11),** 1016-1020.
- Xia, Y., *et al.* (2004) Analyzing cellular biochemistry in terms of molecular networks. *Annu Rev Biochem,* **73,** 1051-1087.
- Barabasi, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet,* **5(2),** 101-113.
- Zhu, X., Gerstein, M. and Snyder, M. (2007) Getting connected: analysis and principles of biological networks. *Genes Dev.,* **21(9),** 1010-1024.

• **Joseph Loscalzo, Albert-László Barabási and Edwin K. Silverman (Eds) (2017). Network Medicine: Complex Systems in Human Disease and Therapeutics. Cambridge, Massachusetts : Harvard University Press.**

• Cho DL, Kim YA, Przytycka TM (2012). Chapter 5: Network Biology Approach to Complex Diseases. *PLoS Comp Biol.* **8(12):**e1002820

• Barabási AL, Gulbahce N, Loscalzo J (2011). Network medicine: a network-based approach to human disease. *Nat Rev Genet* **12(1):** 56–68.

• McGillivray, P., *et al.* (2018) Network Analysis as a Grand Unifier in Biomedical Data Science. *Annual Review of Biomedical Data Science.* **1:**153-180.

• Cowen, L., et al. (2017) Network propagation: a universal amplifier of genetic associations. *Nature Rev Genetics.* **18,** 551.

PDFs available at **http://csbg.cnb.csic.es/pazos/cursos/UMA_BIF/**

Florencio Pazos *@fpazos_bioinf*
pazos@cnb.csic.es *http://csbg.cnb.csic.es*