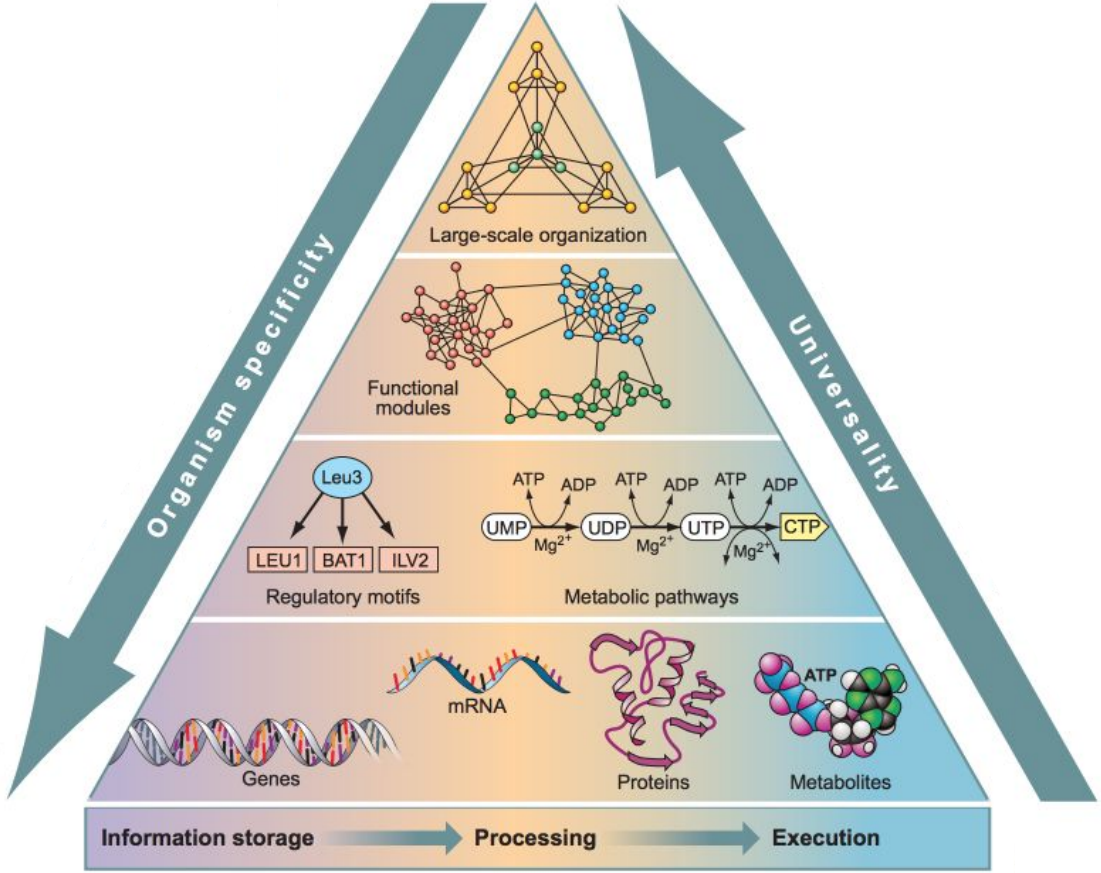
Lecture 14: Large-scale biological networks

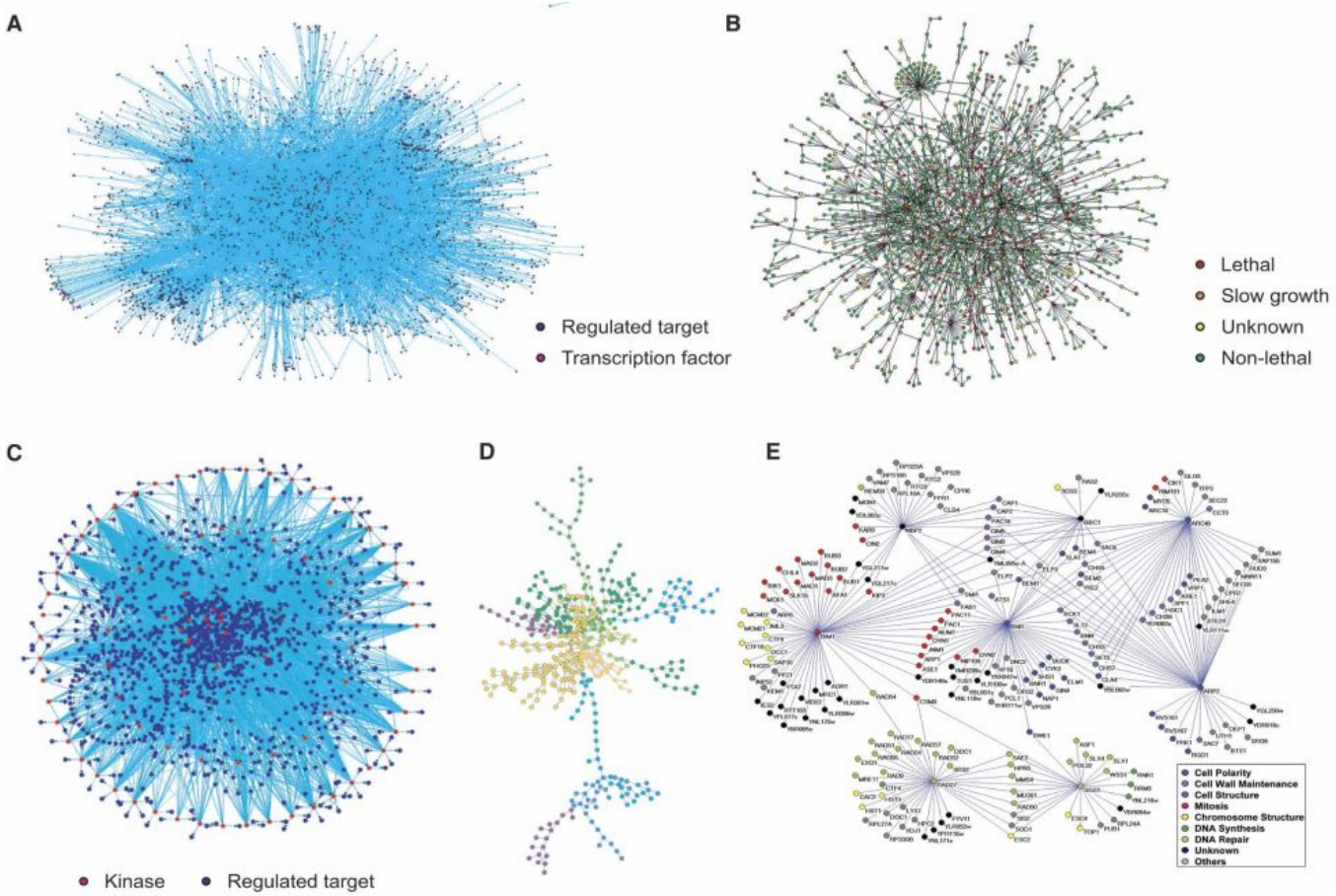
- Network topology
- Network motifs
- Condition-specific networks
- Network reconstruction

Life's complexity pyramid



Oltvai, Barabasi (2002) Science

Biological networks

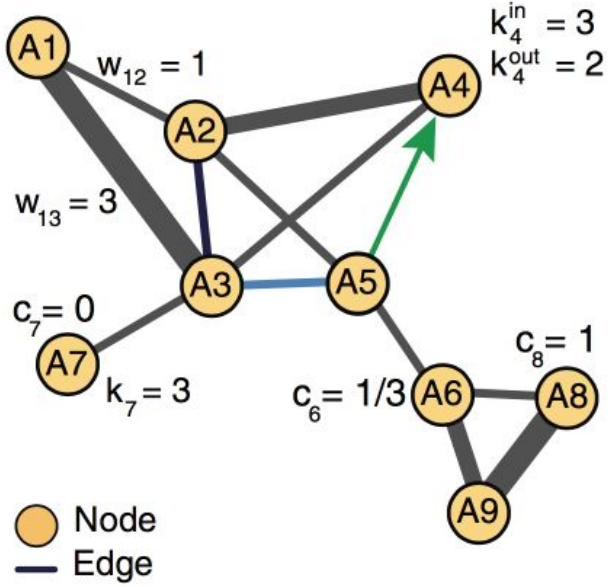


Zhu (2007) Genes & Dev.

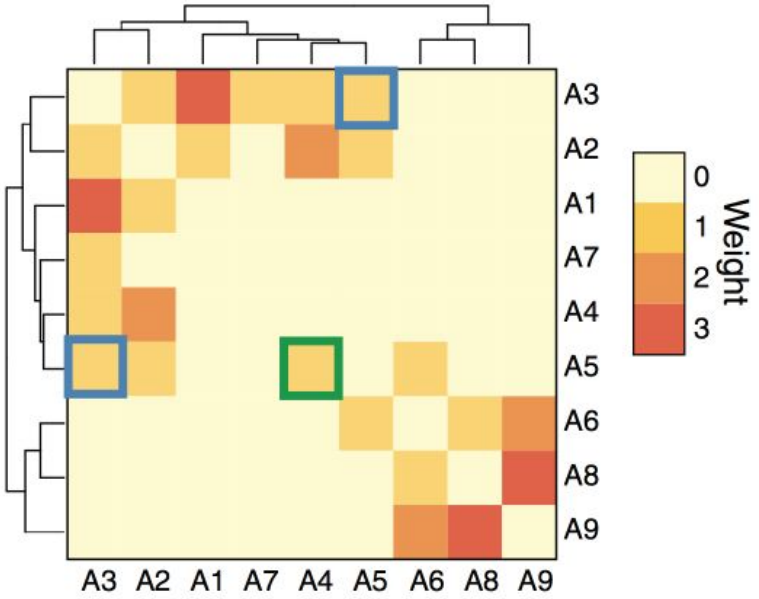
Network description and layout

Relationships	Optional weight
A1 ↔ A2	1
A1 ↔ A3	3
A2 ↔ A3	1
A2 ↔ A4	2
A2 ↔ A5	1
A3 ↔ A4	1
A3 ↔ A5	1
A3 ↔ A7	1
A5 → A4	1
A5 ↔ A6	1
A6 ↔ A8	1
A6 ↔ A9	2
A8 ↔ A9	3

Network

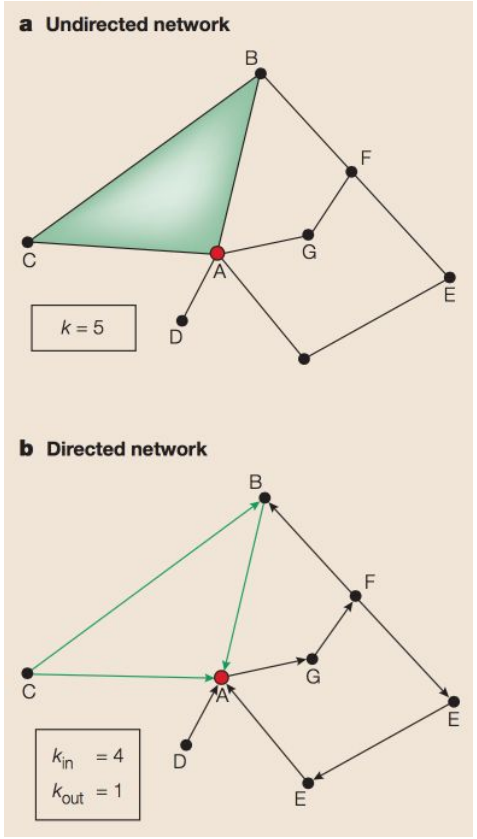


Heat map

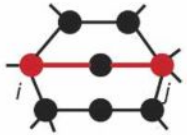


Merico (2009) Nat. Biotech.

Network topology

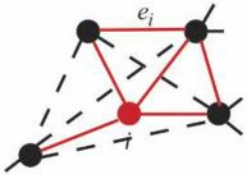


Degree k_i = number of links connected to node i



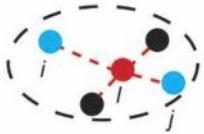
Distance d_{ij} = shortest path length between node i and j

Diameter $D = \max\{d_{ij} \mid i, j \in N\}$ N : all nodes in the network



Clustering Coefficient $c_i = \frac{2e_i}{k_i(k_i-1)}$

e_i : number of existing links (labeled in red) among the k_i nodes that connect to node i



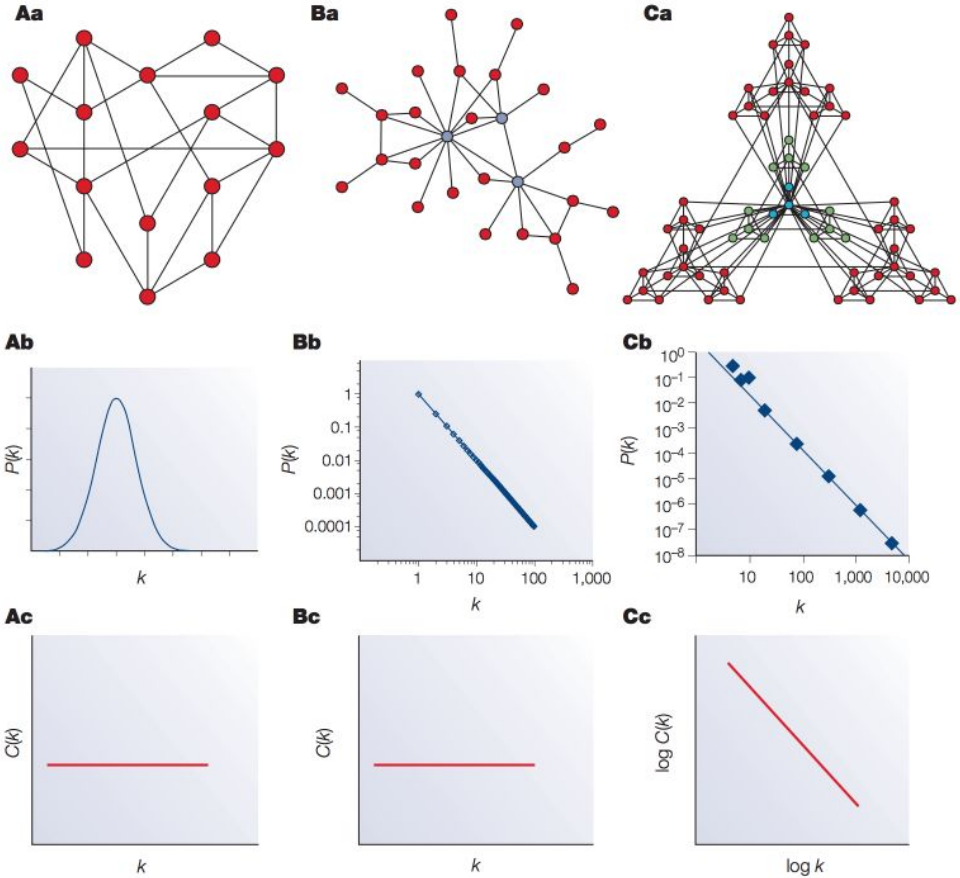
Betweenness $b_l = \sum_{ij} p_{ij}(l)/p_{ij}$

p_{ij} : number of shortest paths between i and j

$p_{ij}(l)$: number of shortest paths between i and j going through node l

Barabasi, Oltvai (2004) Nat. Rev. Genet., Zhu (2007) Genes & Dev.

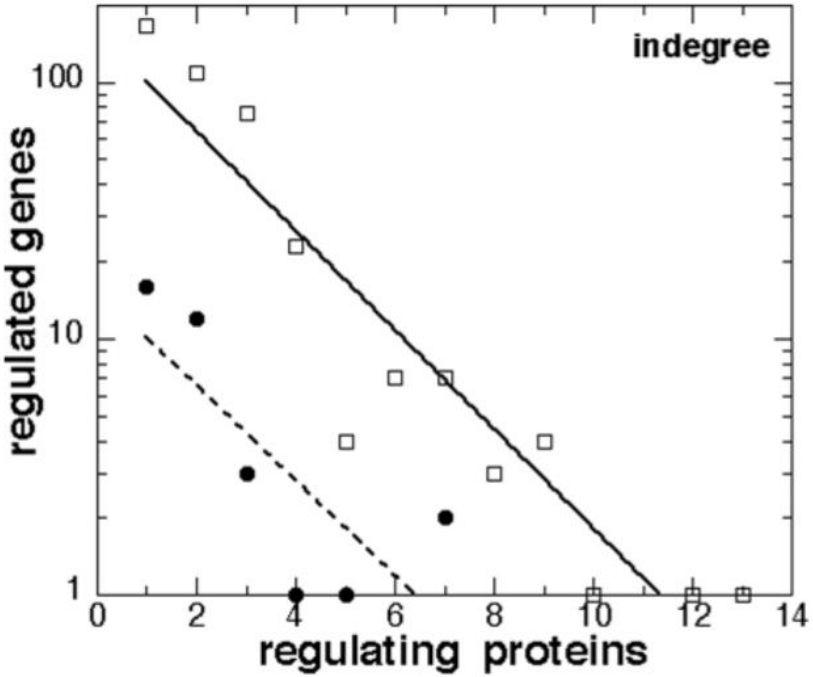
Network topology

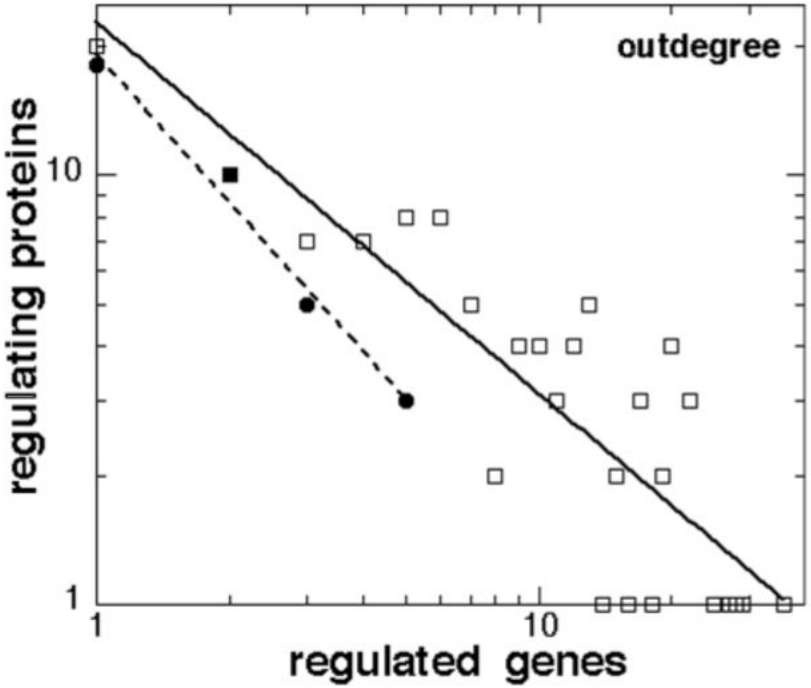


Barabasi, Oltvai (2004) Nat. Rev. Genet.

Network topology

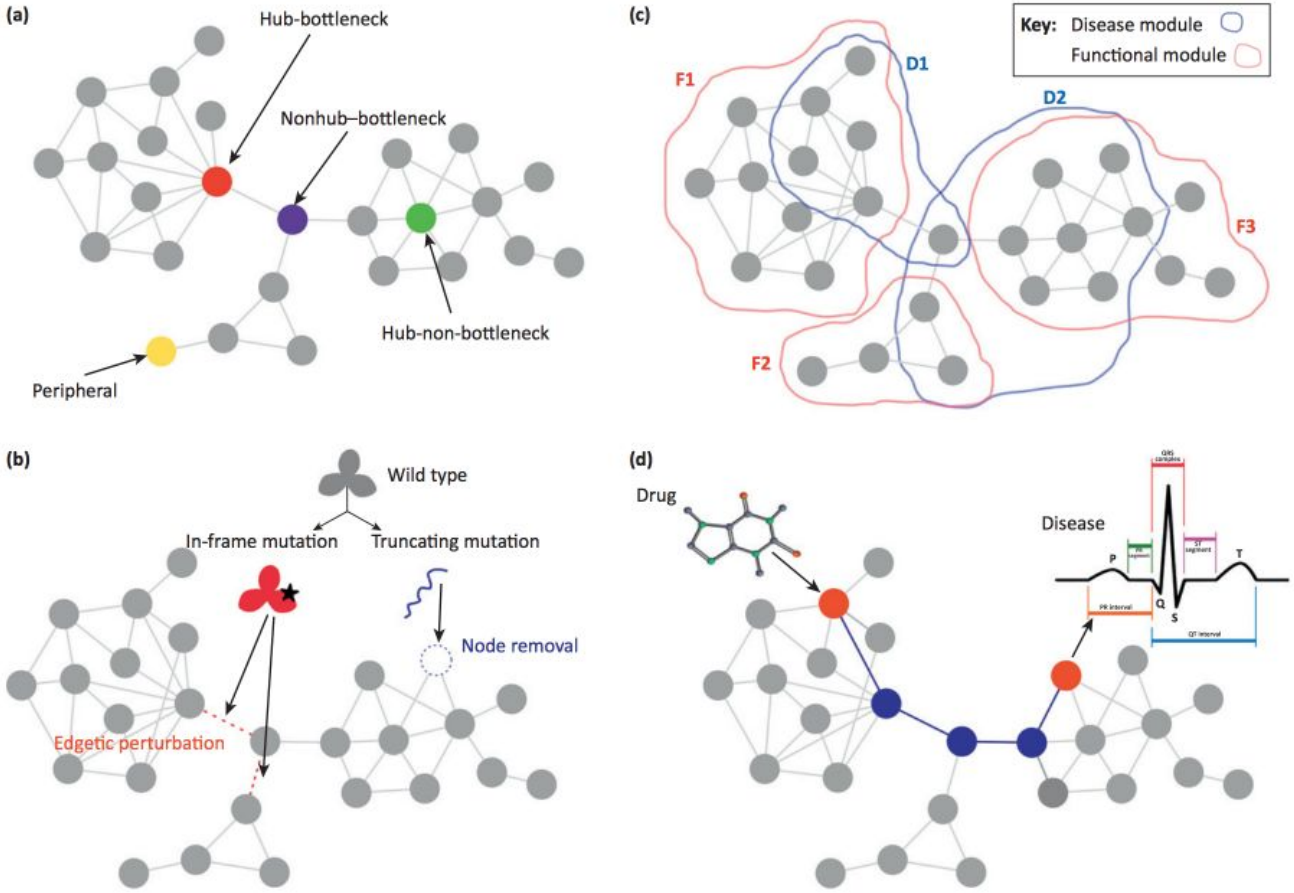
Yeast TF-target network





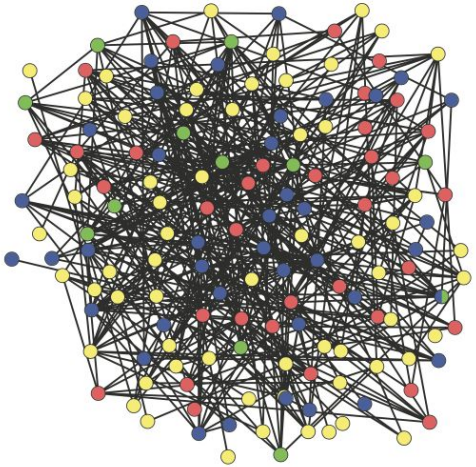
Oltvai, Barabasi (2002) Science

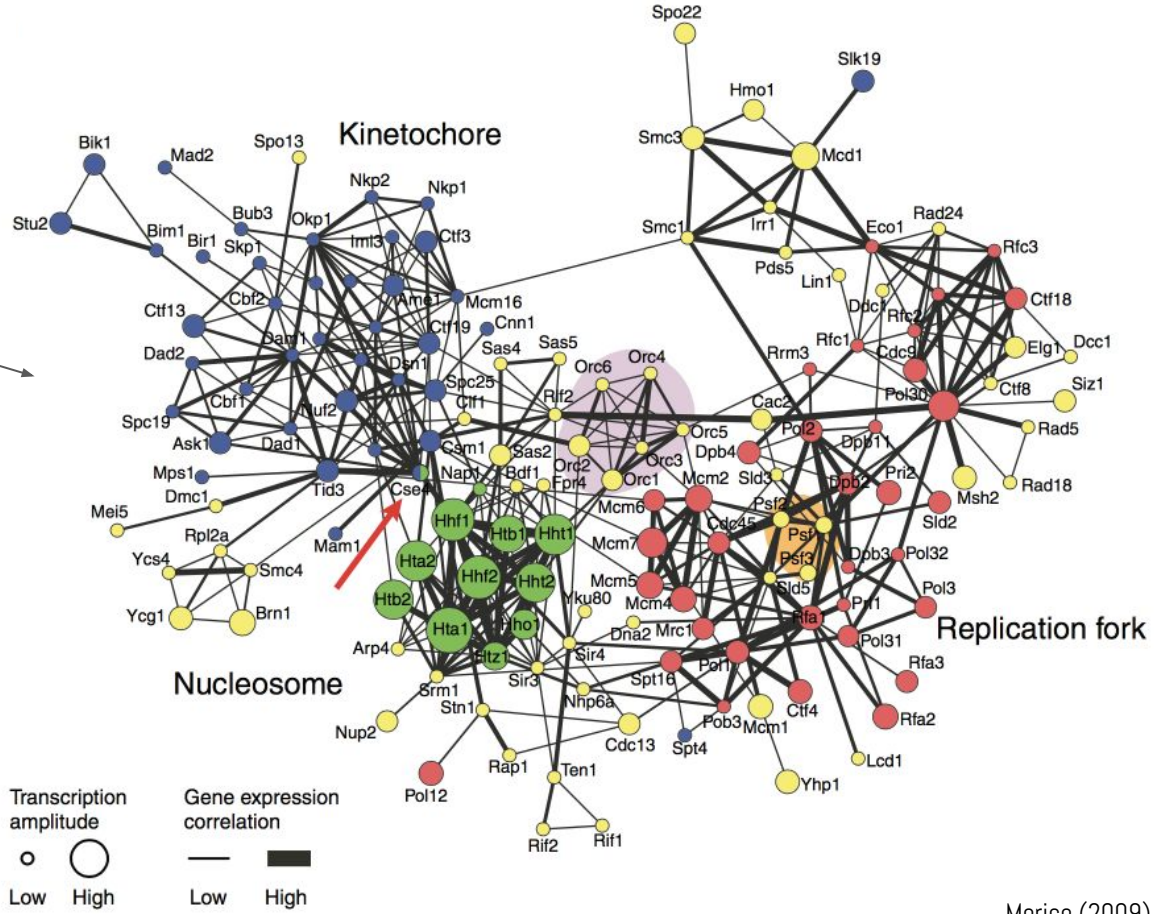
Network topology



Furlong (2013) Trends in Genet.

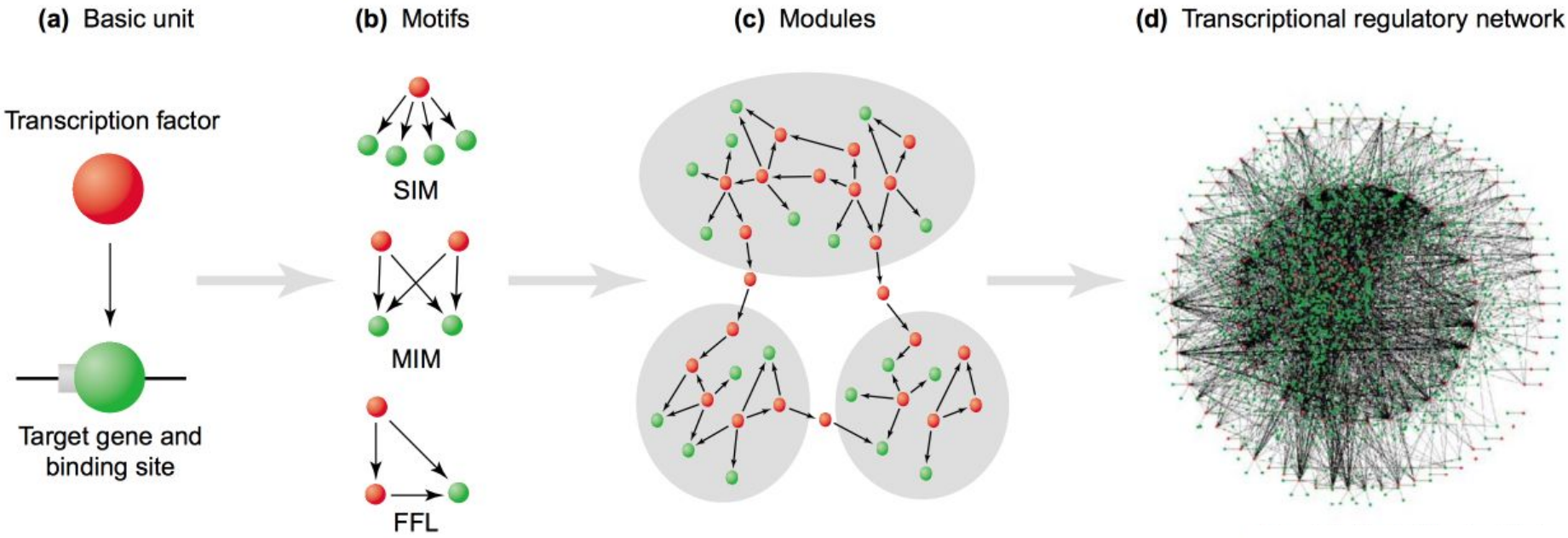
Network description and layout





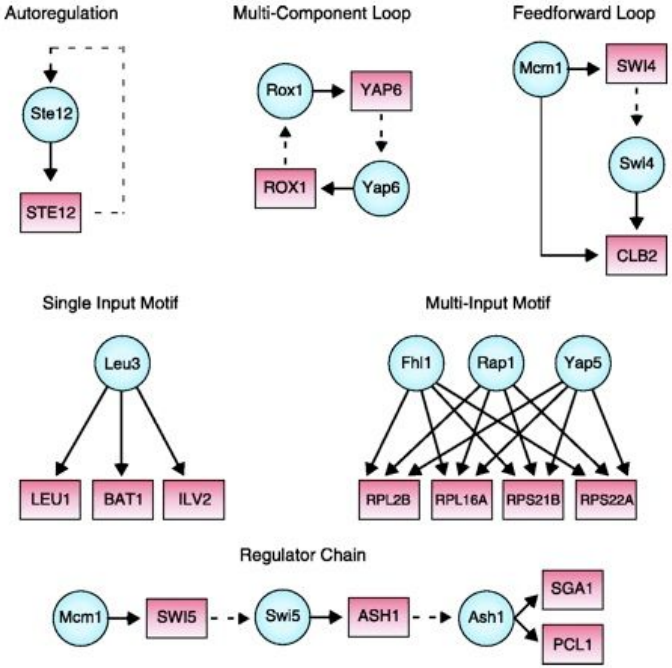
Merico (2009) Nat. Biotech.

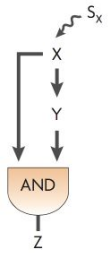
Network motifs

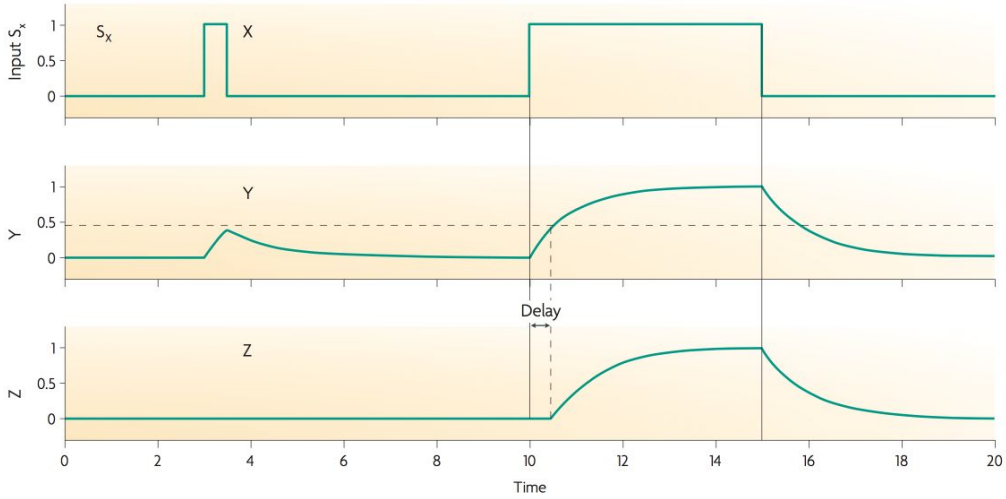


Babu (2004) Curr. Opin. Struct. Biol

Network motifs



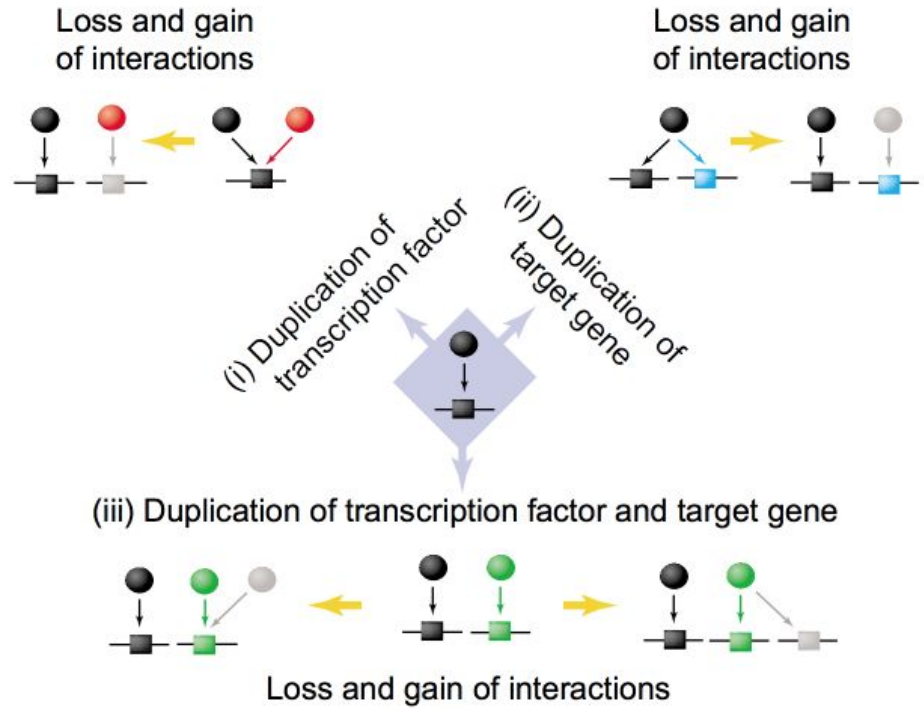


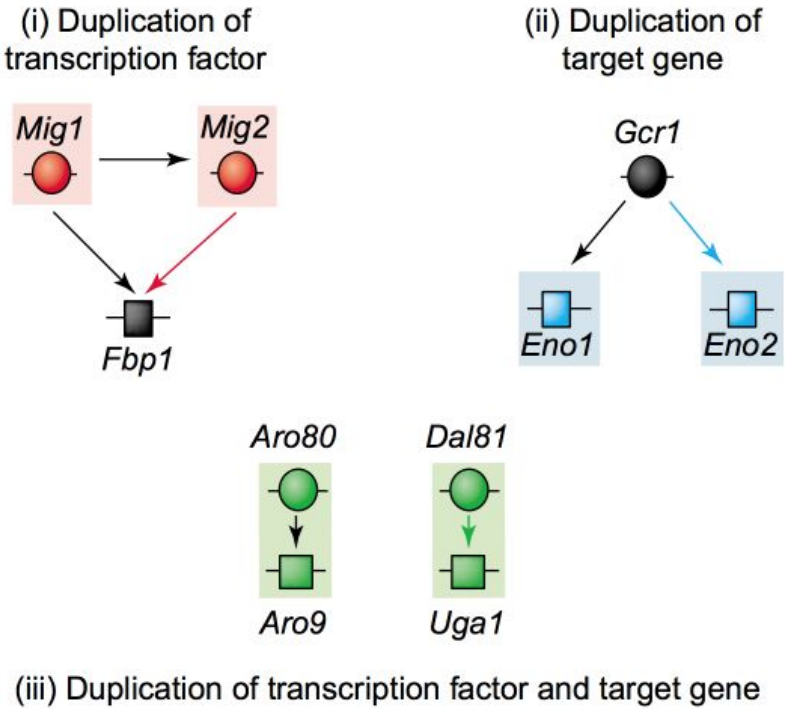


Lee (2002) Science

Alon (2007) Nat. Rev. Genet.

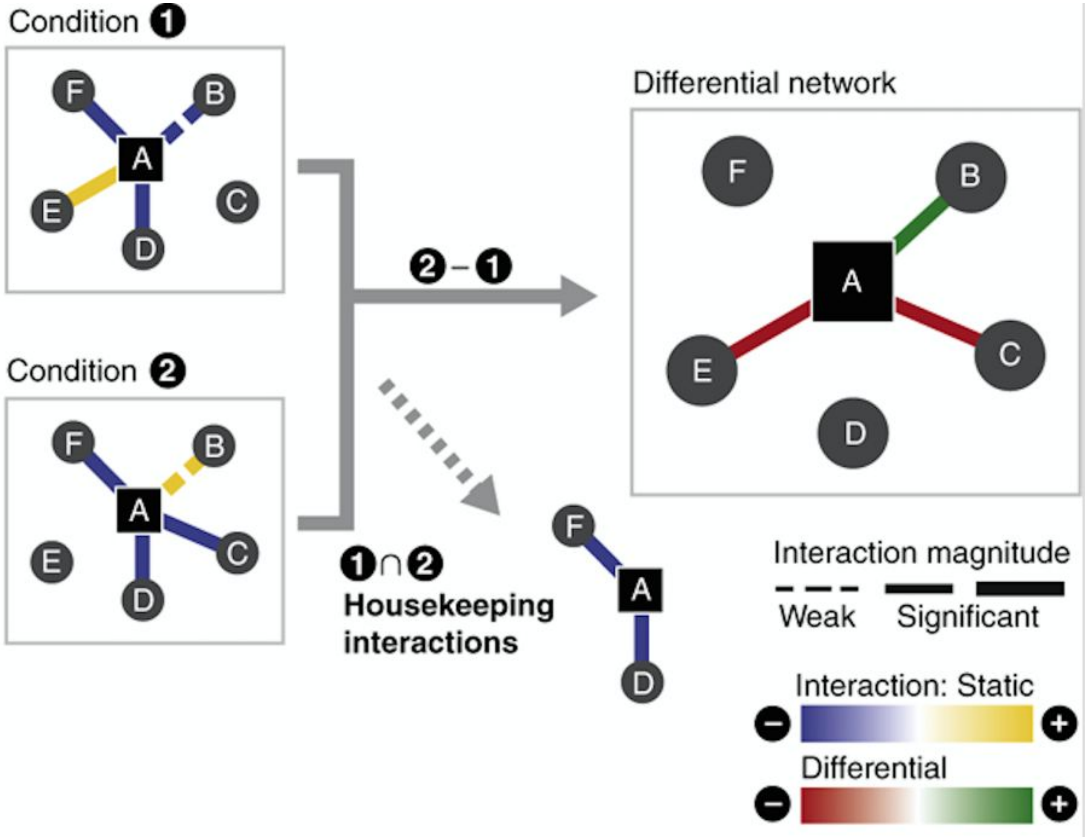
Network evolution





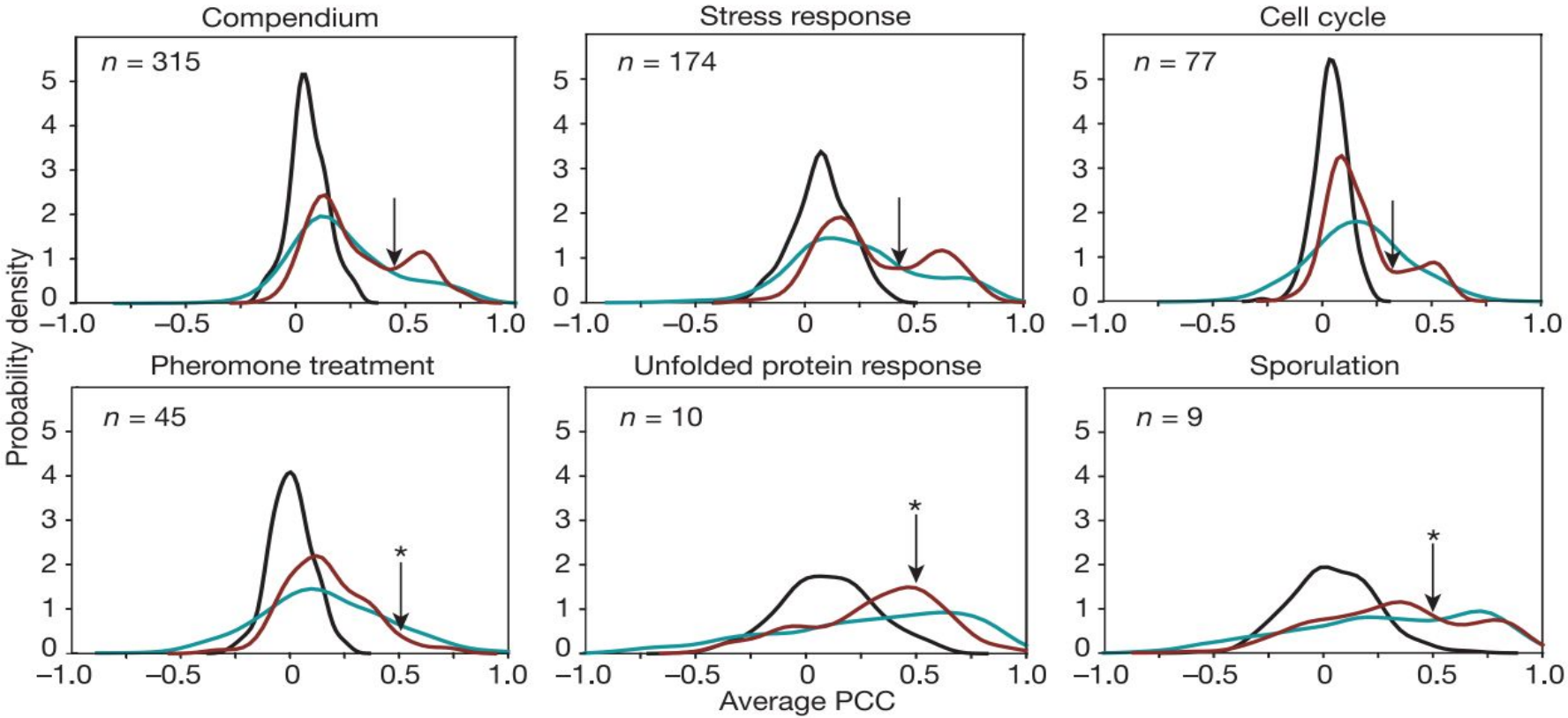
Babu (2004) Curr. Opin. Struct. Biol

Network interactions are condition-specific



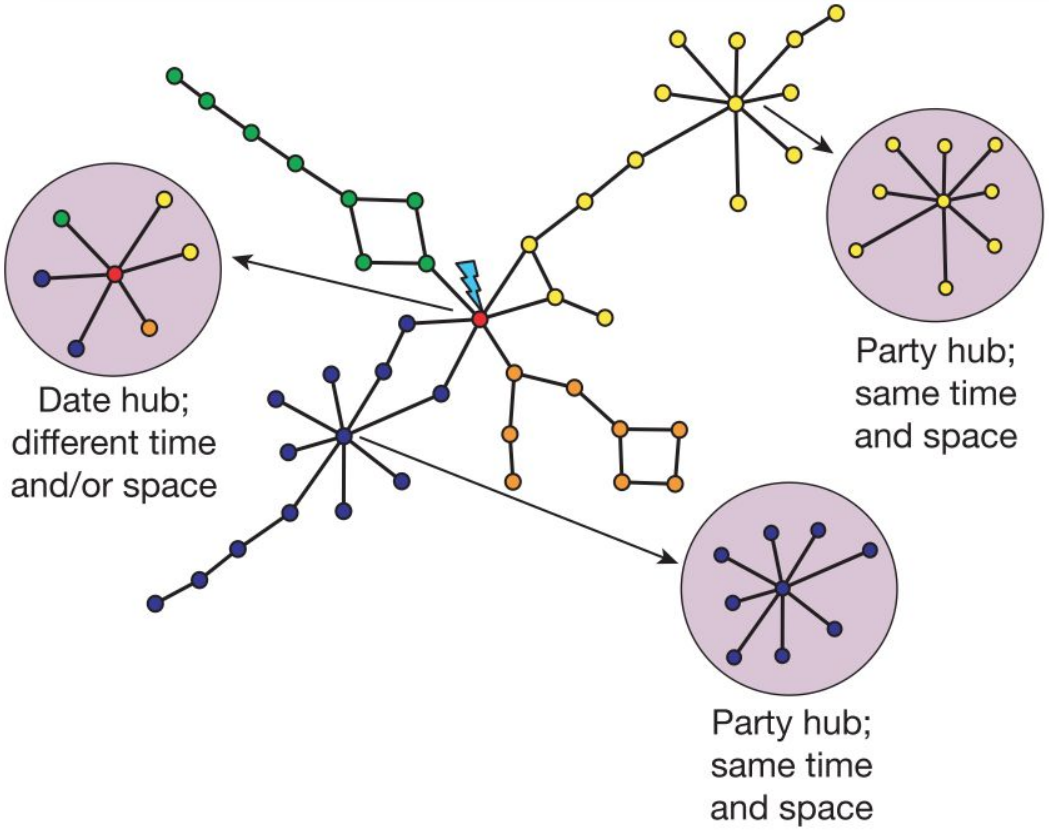
Ideker, Krogan (2007) Mol. Sys. Biol.

Network interactions are condition-specific



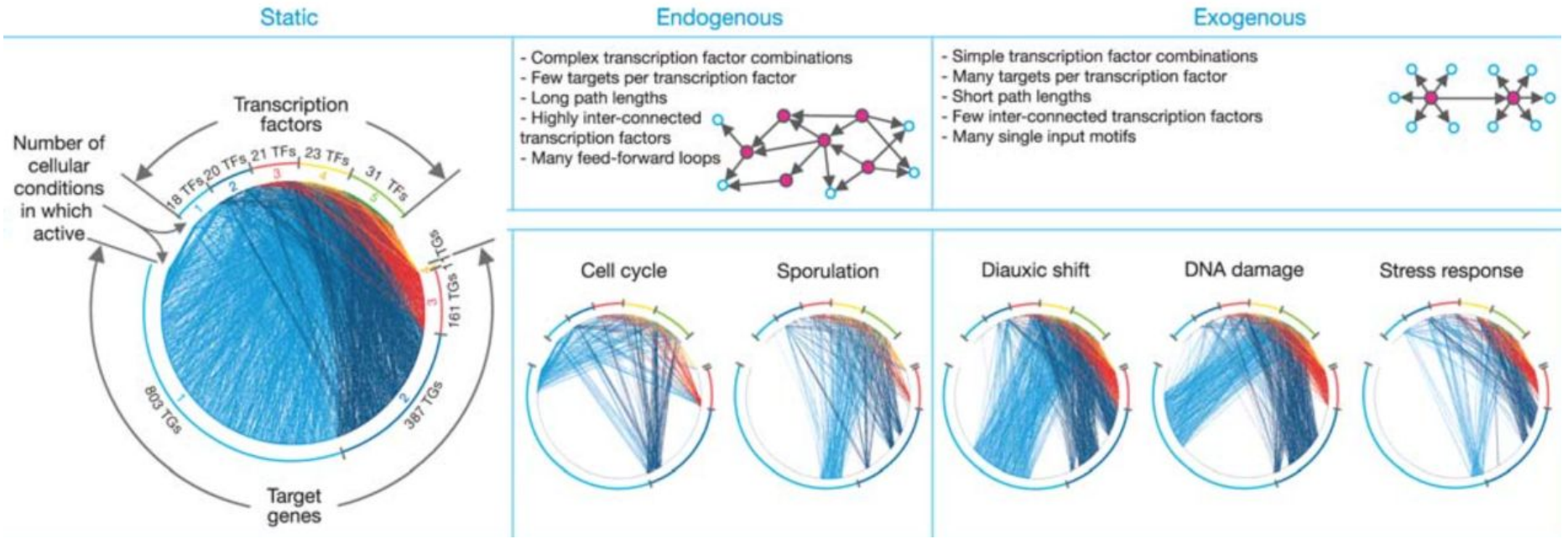
Han (2004) Nature

Network interactions are condition-specific



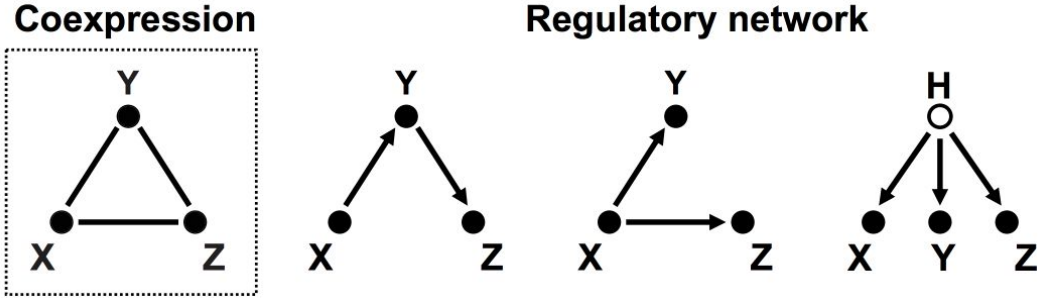
Han (2004) Nature

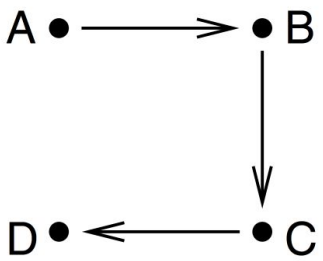
Network interactions are condition-specific

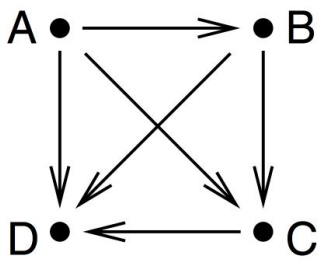


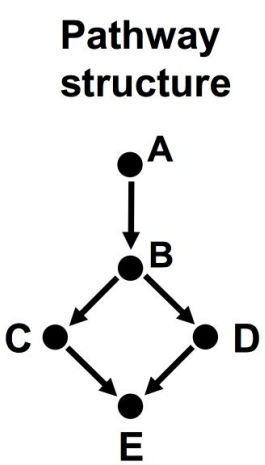
Luscombe, Babu (2004) Nature

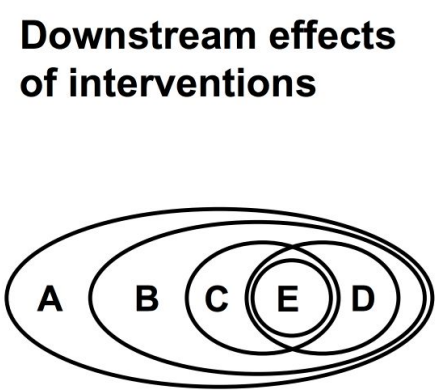
Reconstructing networks from observational data







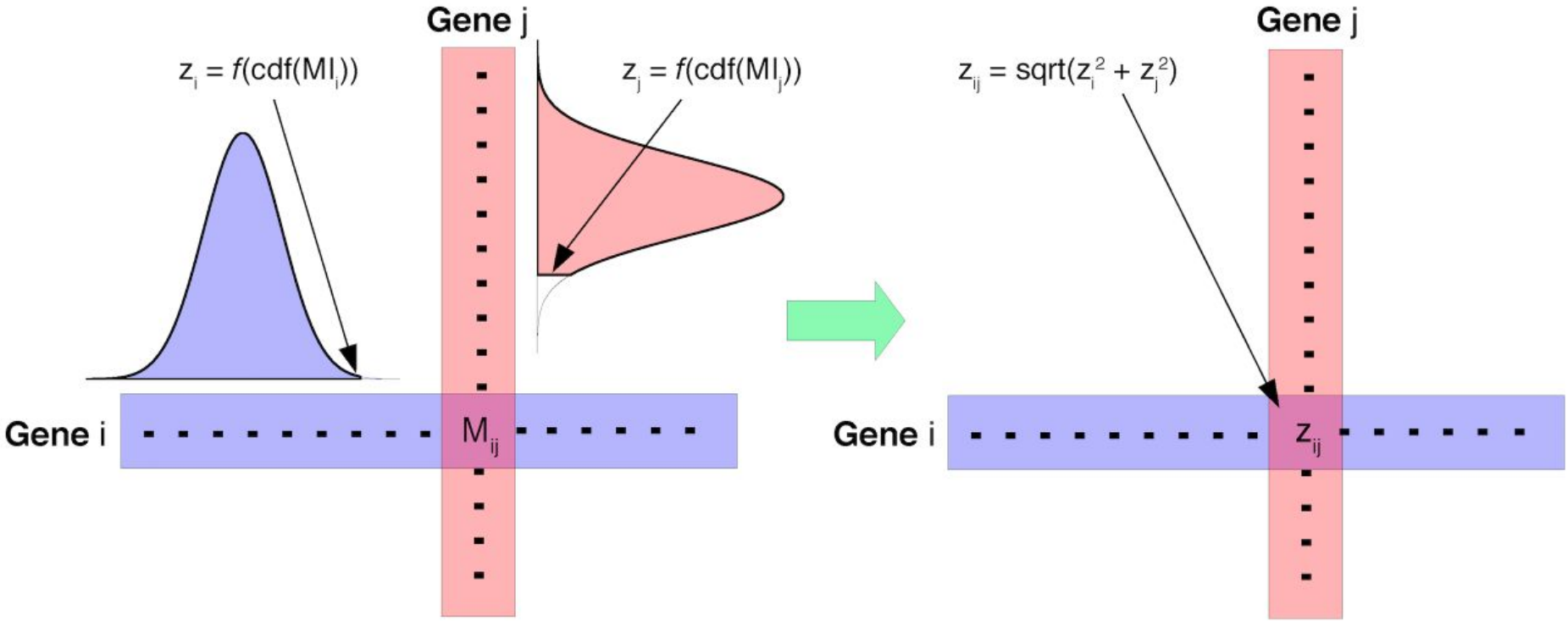




Markowetz, Spang (2007) BMC Bioinfo.

Reconstructing networks from observational data

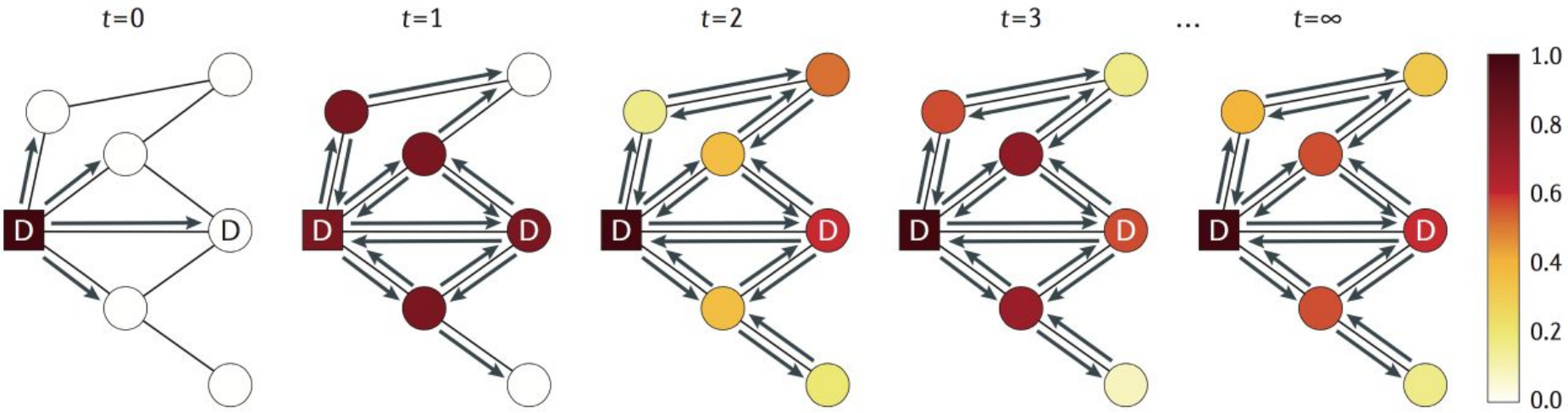
Context likelihood relatedness



Faith (2007) PLoS Biology

Network propagation

Tracing the flow of information through a network over time.



Random walk

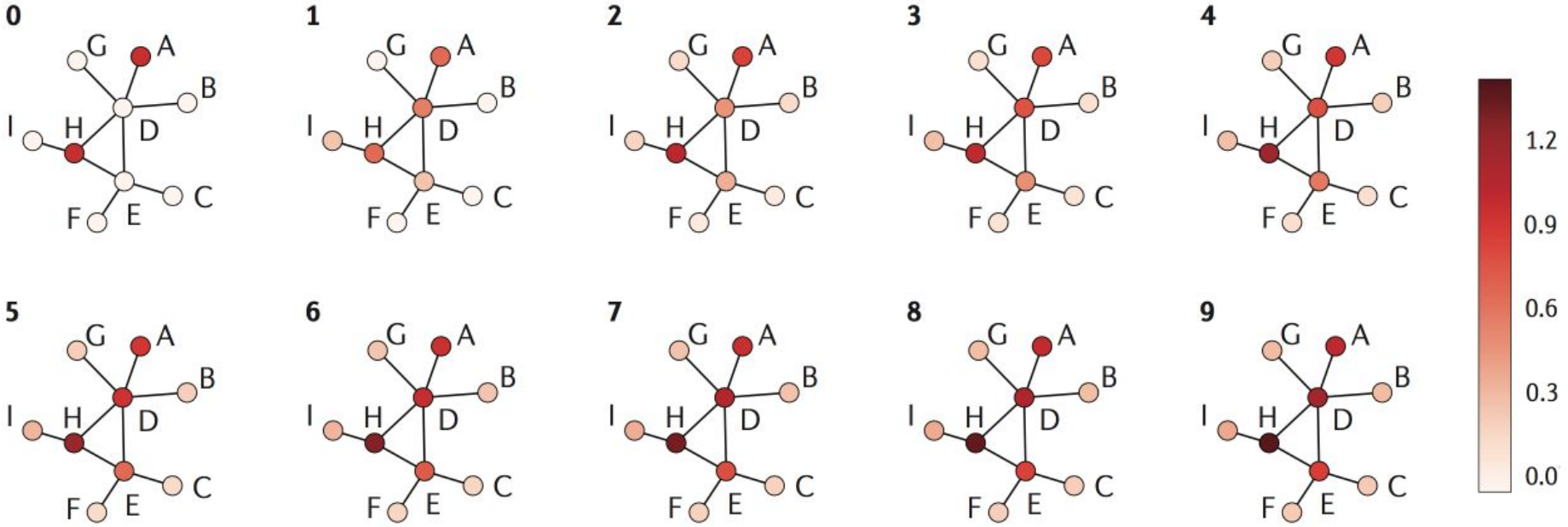
A mathematical formalization of the paths resulting from successive random steps a 'walker' takes from **one node to another** with a probability that is **proportional to the weight of the edge** connecting the nodes.

Cowen (2017) Nat. Rev. Genet.

Network propagation

Tracing the flow of information through a network over time.

Initial node scores... (e.g. expression in a condition or association with a disease)

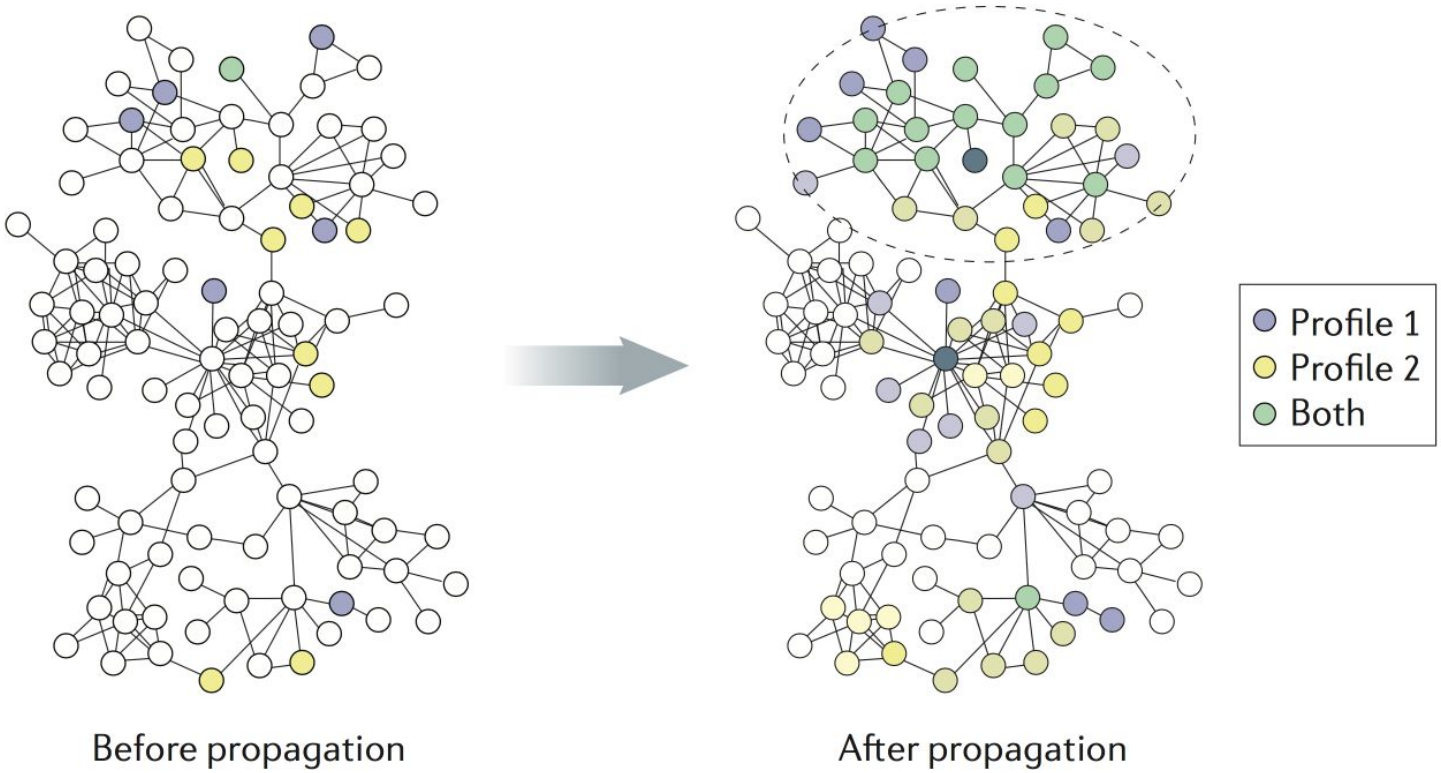


Convergence...

Cowen (2017) Nat. Rev. Genet.

Network propagation

Tracing the flow of information through a network over time.



Cowen (2017) Nat. Rev. Genet.

Network propagation

Random Walk

$p_0(v)$: Vector of initial node scores representing experimental measurements or our prior knowledge (e.g. expression in a condition or association with a disease)

$$p_0(v)$$

$p_k(v)$: node scores at time-step k.

$$p_k(v) = \sum_{u \in N(v)} p_{k-1}(u) w(u, v)$$

$w(u,v)$: (normalized) weight or the confidence of the interaction between u and v.

$$p_k = W p_{k-1}$$

W: normalized adjacency matrix (stochastic).

$$p_k = W^k p_0$$

Cowen (2017) Nat. Rev. Genet.

Network propagation

Random Walk with Restart (RWR)

$p_0(v)$: Vector of initial node scores representing experimental measurements or our prior knowledge (e.g. expression in a condition or association with a disease)

$p_k(v)$: node scores at time-step k.

W: normalized adjacency matrix (stochastic).

α: user-defined parameter that specifies the trade-off between prior information and network smoothing

$$p_0(v)$$

$$p_k = \alpha p_0 + (1-\alpha) W p_{k-1}$$

Cowen (2017) Nat. Rev. Genet.

Network propagation

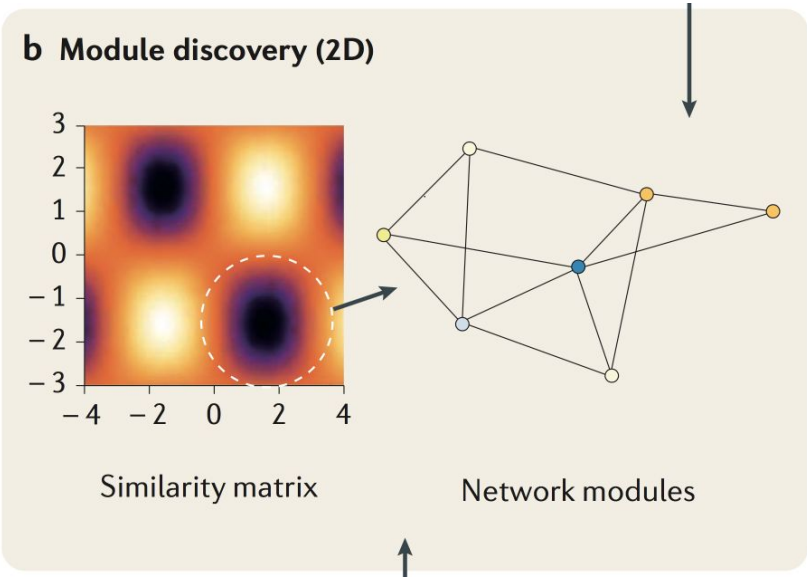
Random Walk & RWR

p: steady-state distribution of node scores.

$$p = Sp_0$$

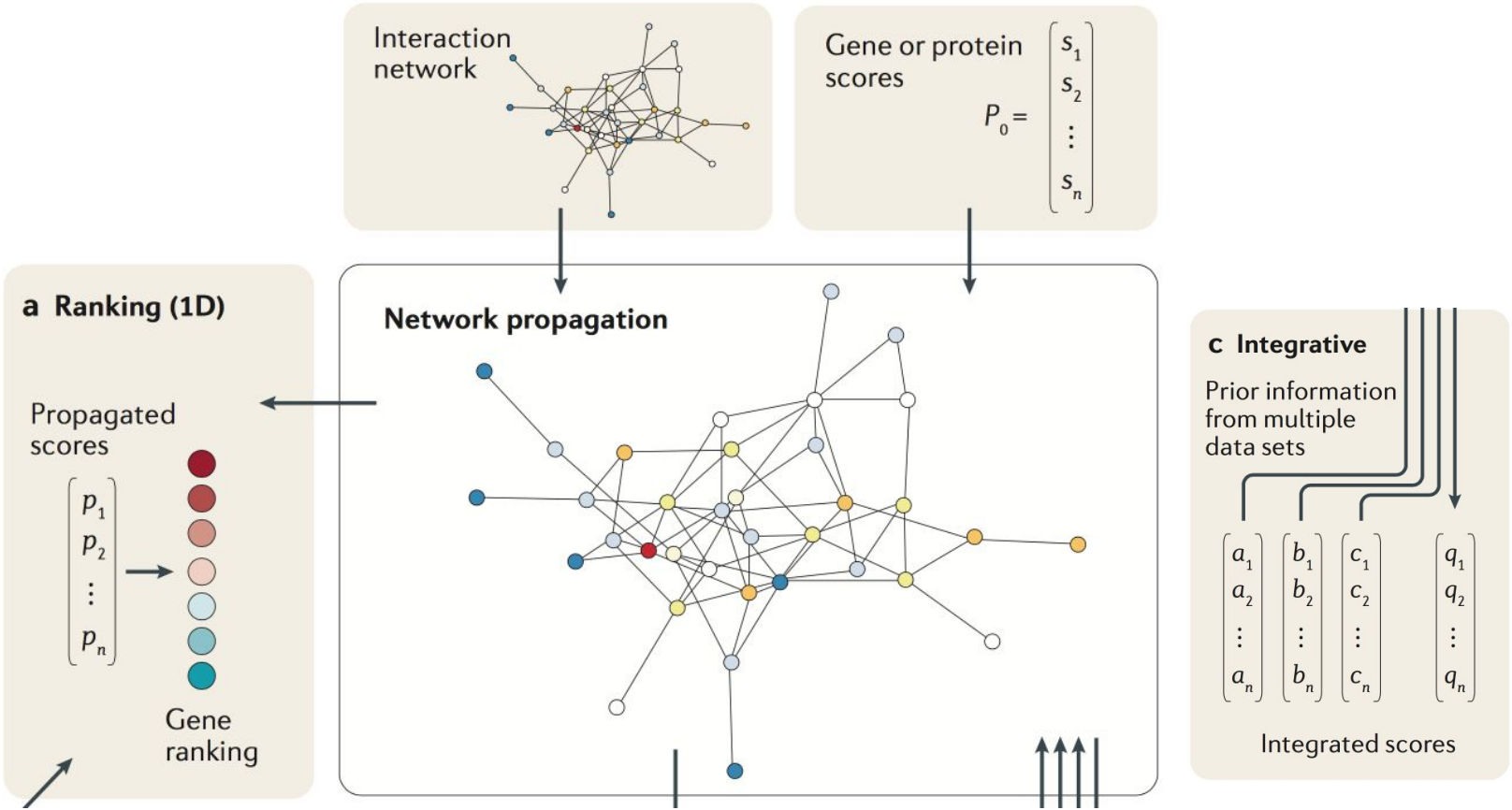
S: Can be interpreted as a similarity matrix.

- S_{ij}: the amount of information propagated to node i, given that the initial ranking p_0 is an elementary vector with 1 at entry j and 0 elsewhere.



Cowen (2017) Nat. Rev. Genet.

Network propagation



Cowen (2017) Nat. Rev. Genet.

Network-based gene function prediction

- Label propagation
- Machine learning