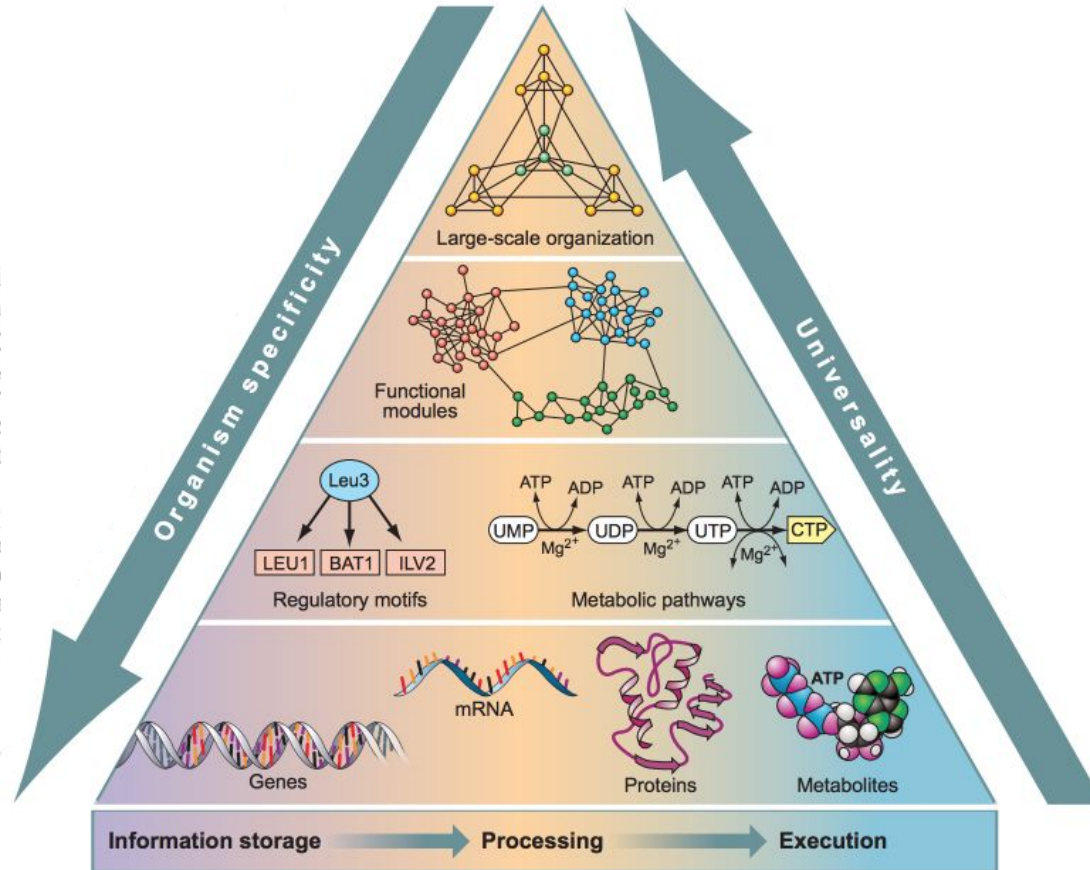


Lecture 14: Large-scale biological networks

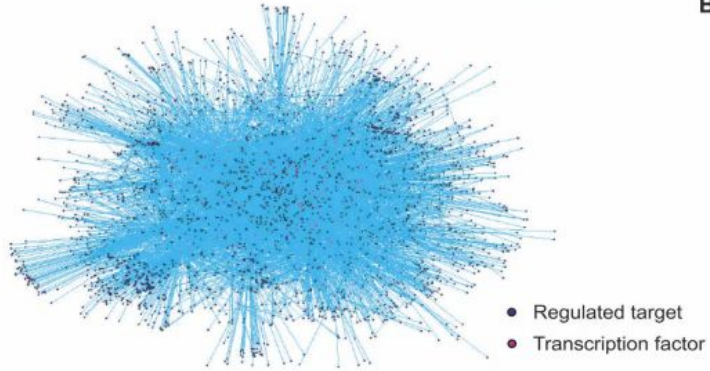
- Network topology
- Network motifs
- Condition-specific networks
- Network reconstruction

Life's complexity pyramid

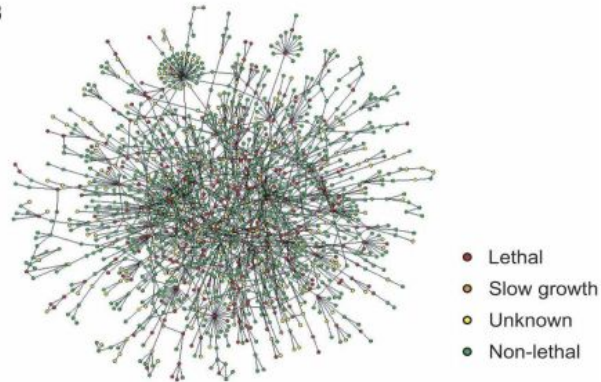


Biological networks

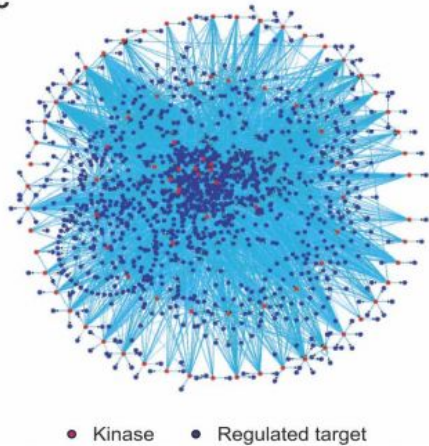
A



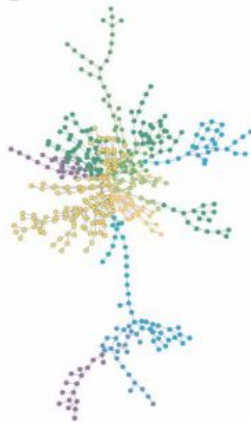
B



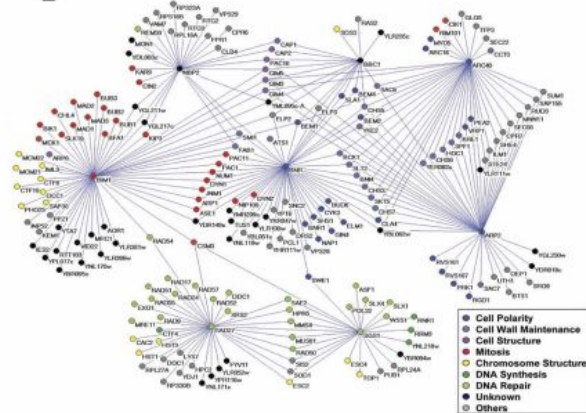
C



D



E



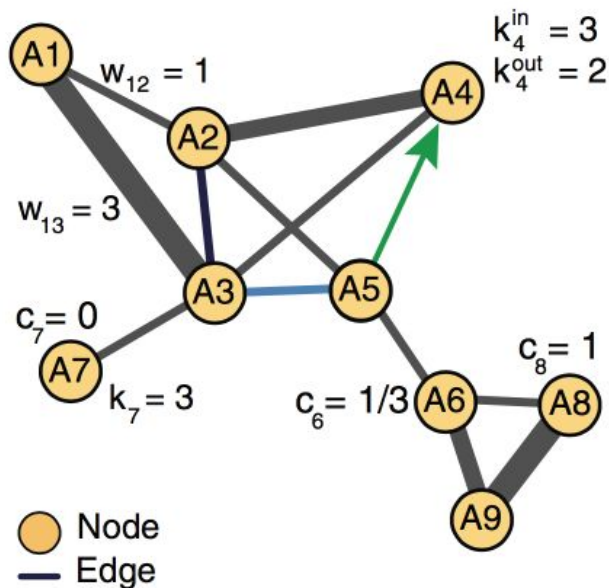
Network description and layout

Relationships

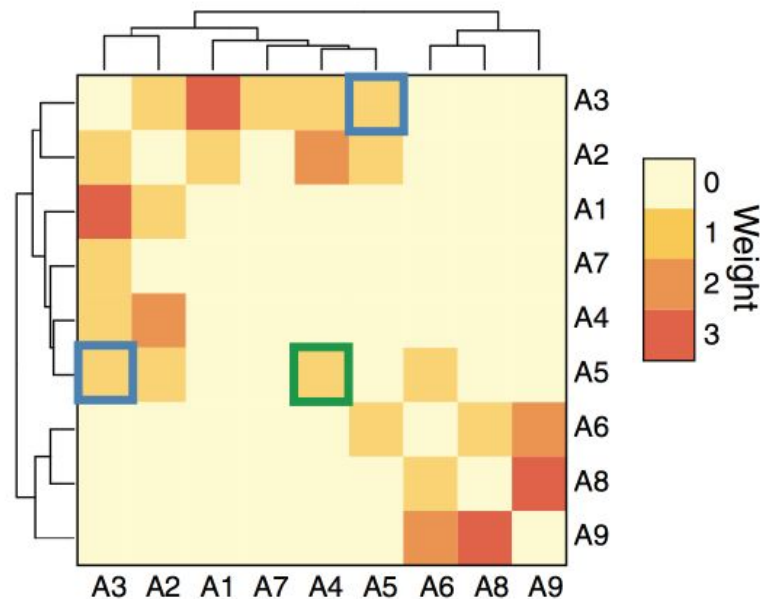
Optional weight

| | |
|---------|---|
| 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

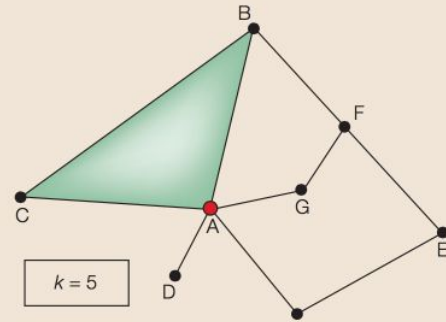


Heat map

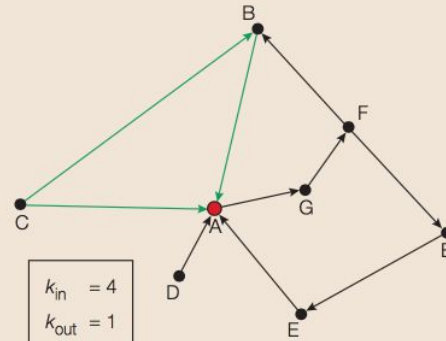


Network topology

a Undirected network

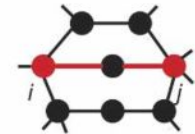


b Directed network



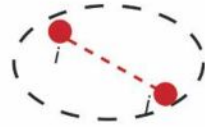
Degree

k_i = number of links connected to node i



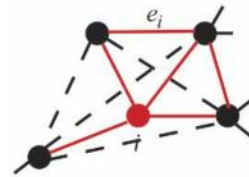
Distance

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



Diameter

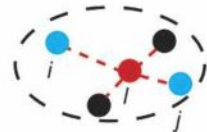
$D = \max \{ d_{ij} | i, j \in N \}$ N : all nodes in the network



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

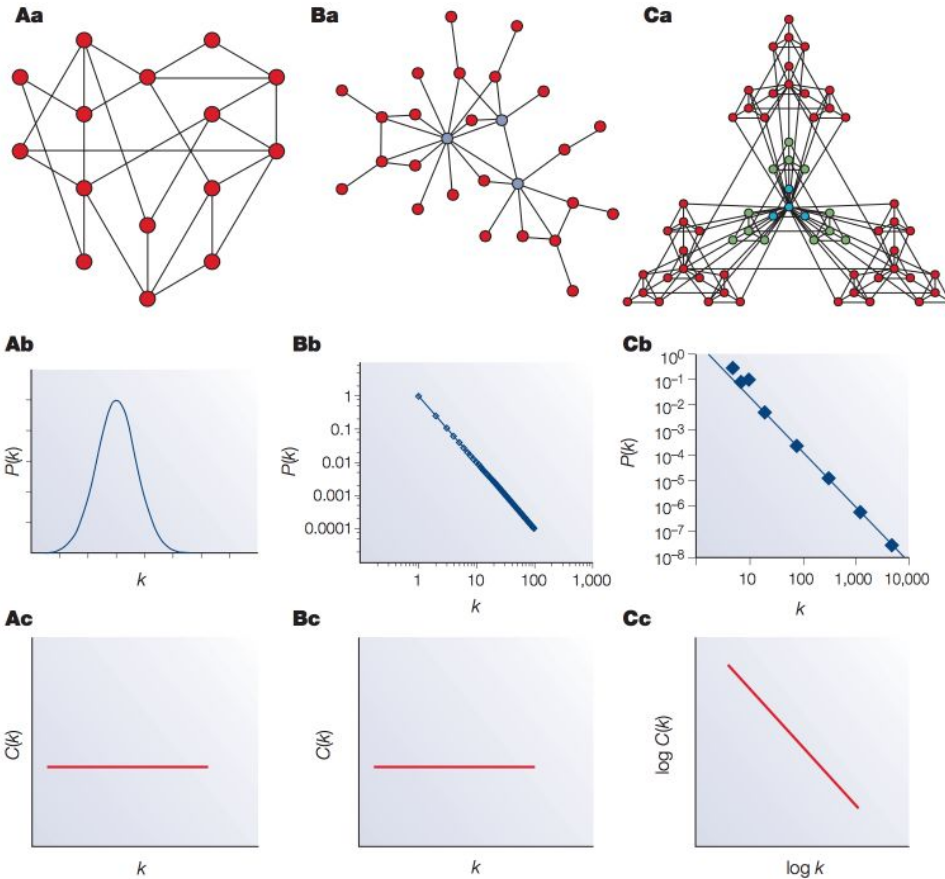
$p_{ij}(l)$: number of shortest paths between i and j going through node l

Barabasi, Oltvai (2004) Nat.

Rev. Genet.

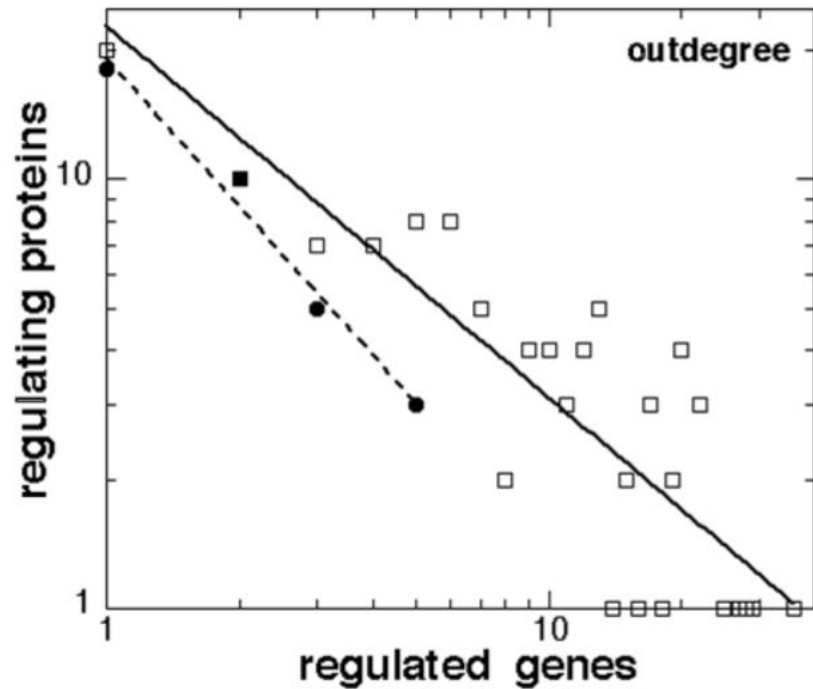
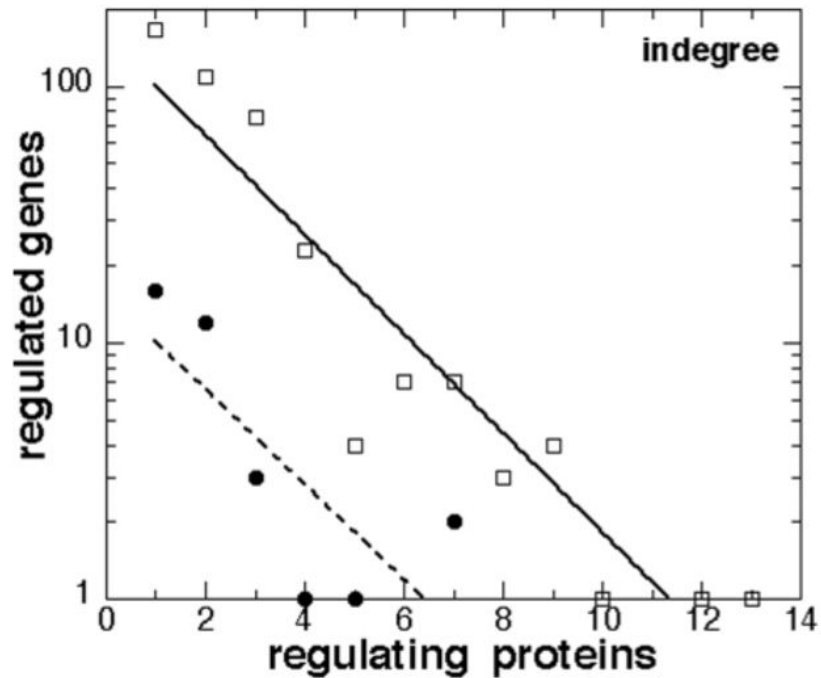
Zhu (2007) Genes & Dev.

Network topology

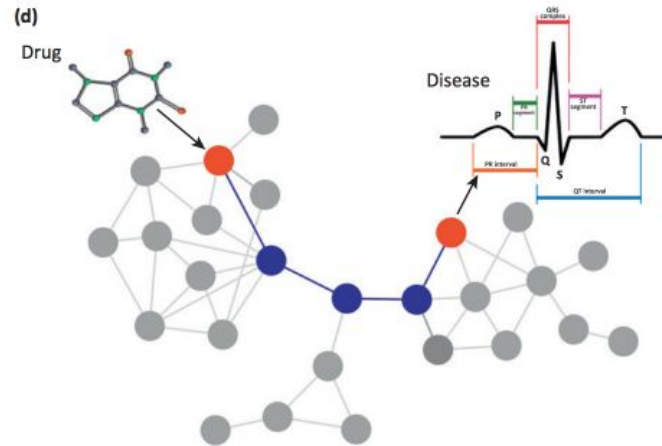
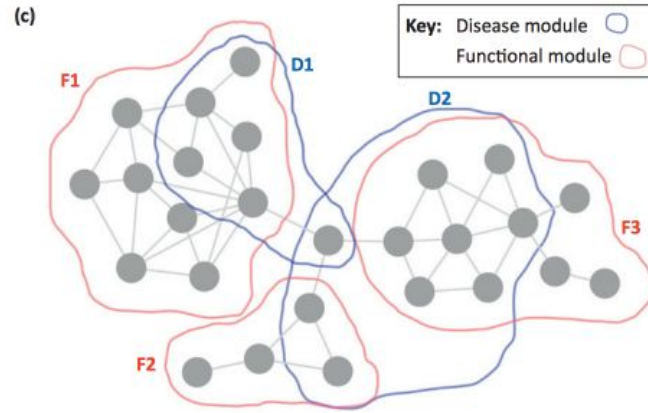
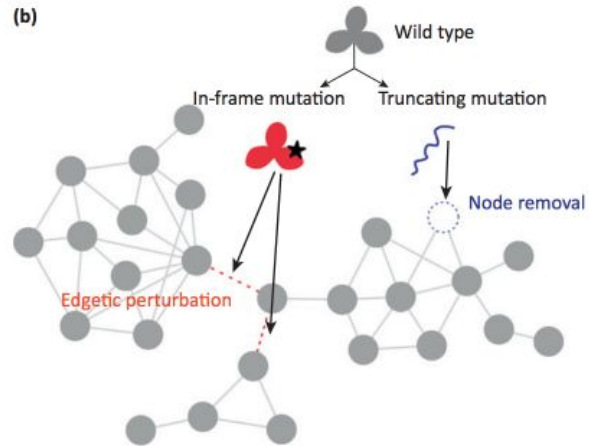
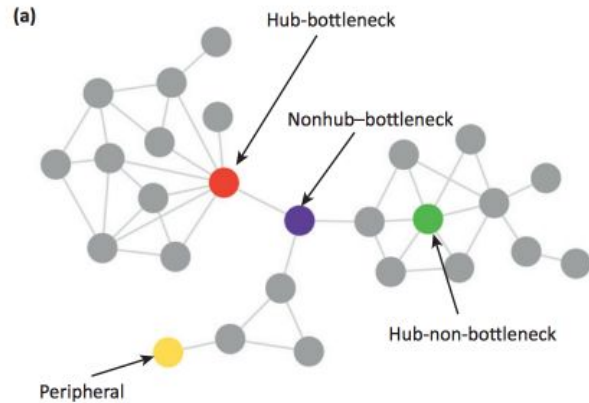


Network topology

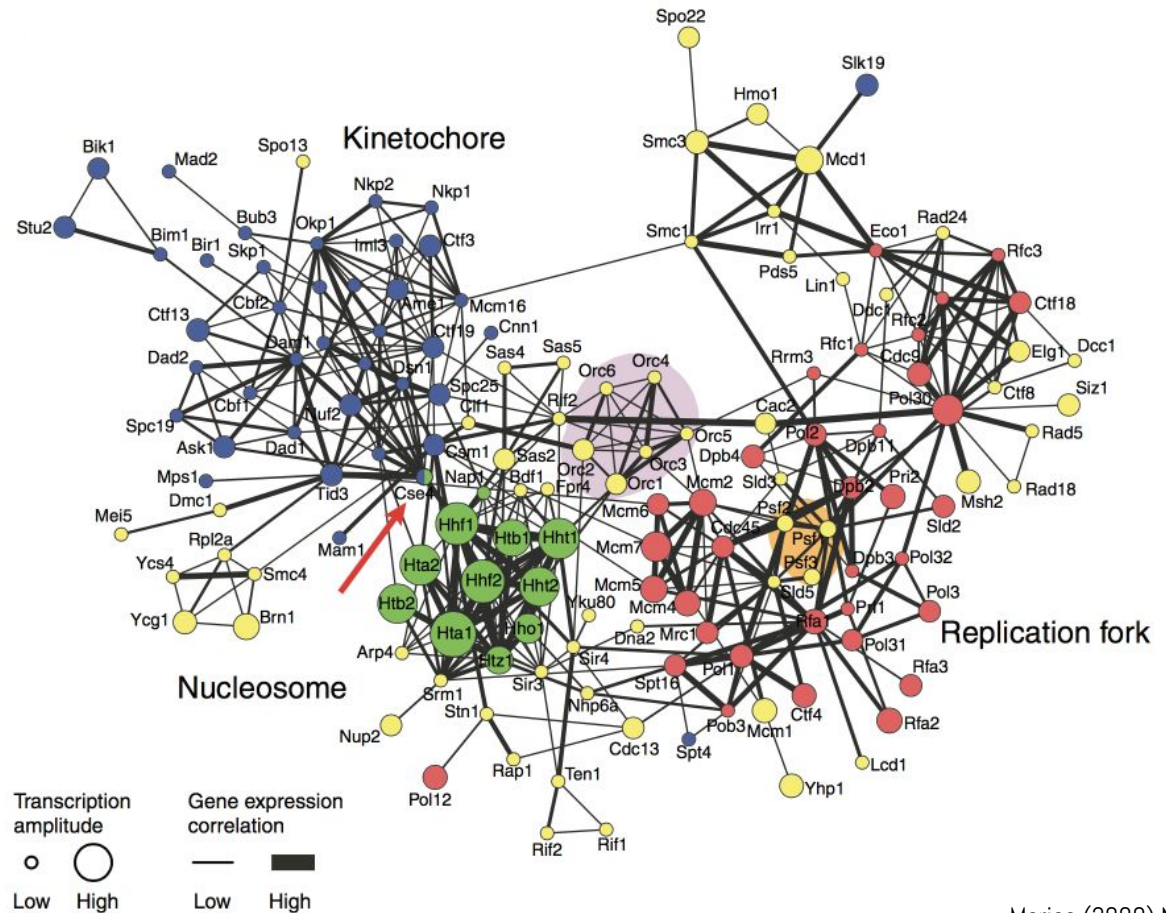
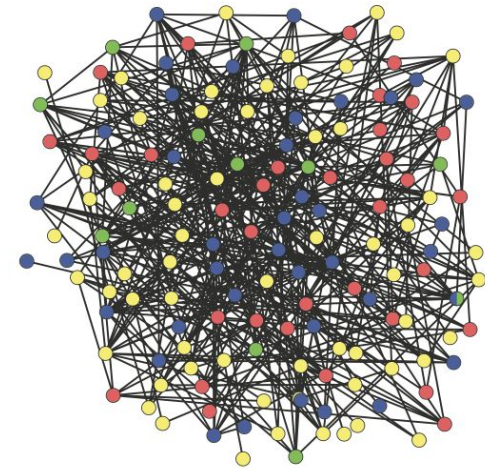
Yeast TF-target network



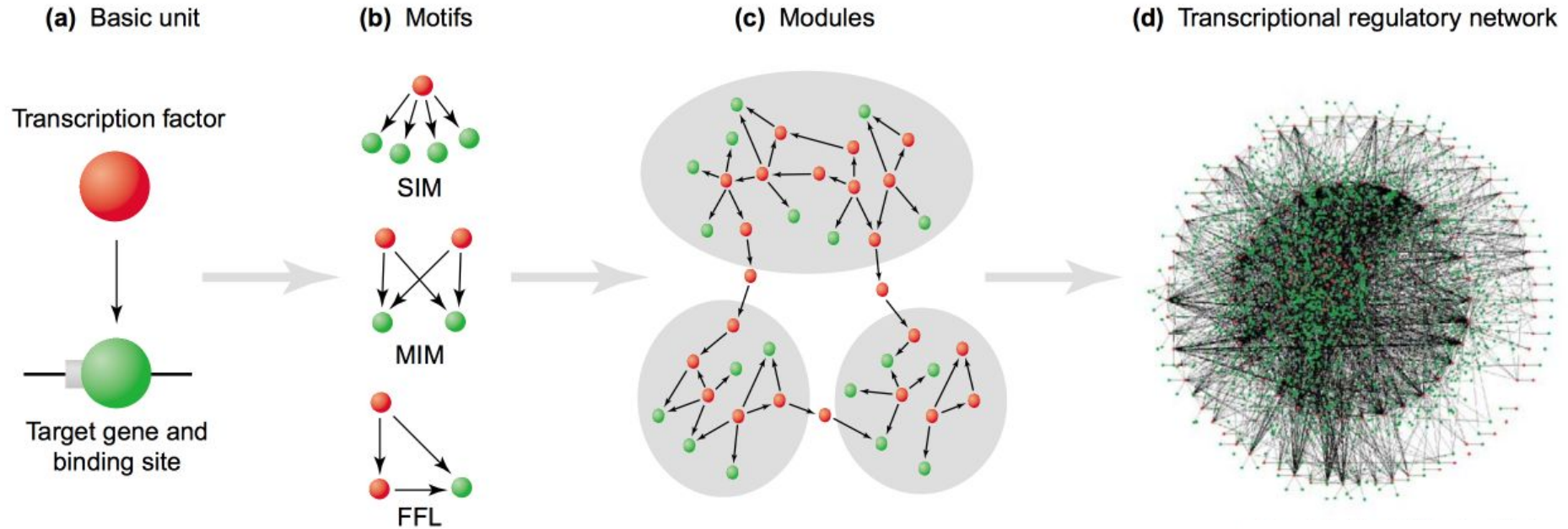
Network topology



Network description and layout

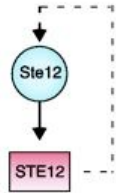


Network motifs

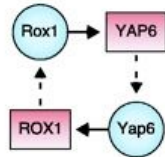


Network motifs

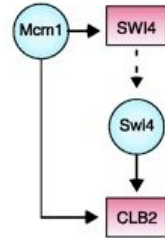
Autoregulation



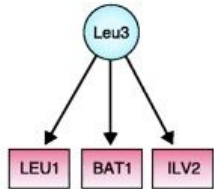
Multi-Component Loop



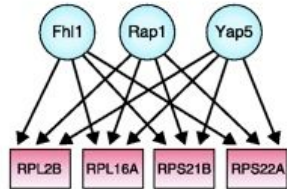
Feedforward Loop



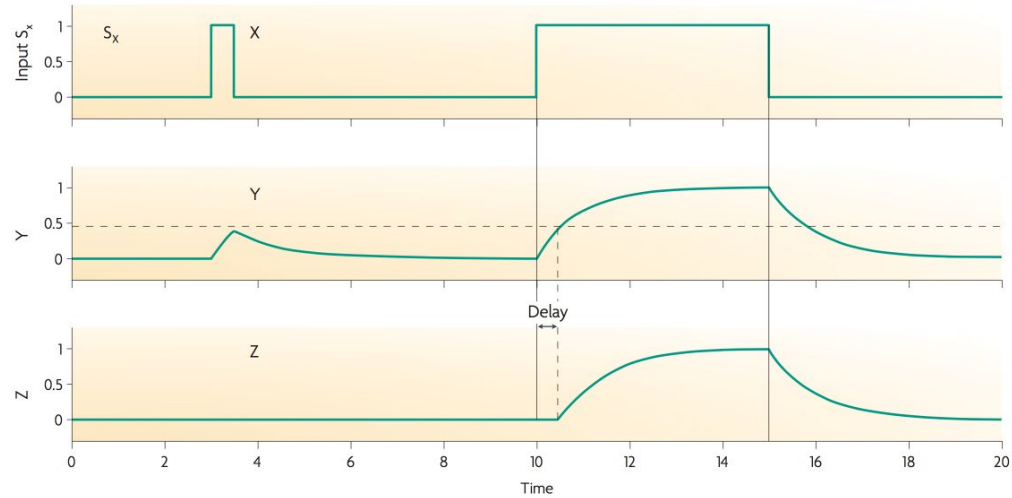
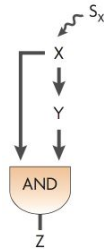
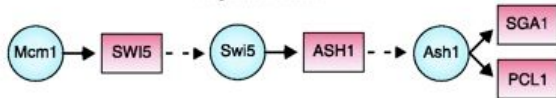
Single Input Motif



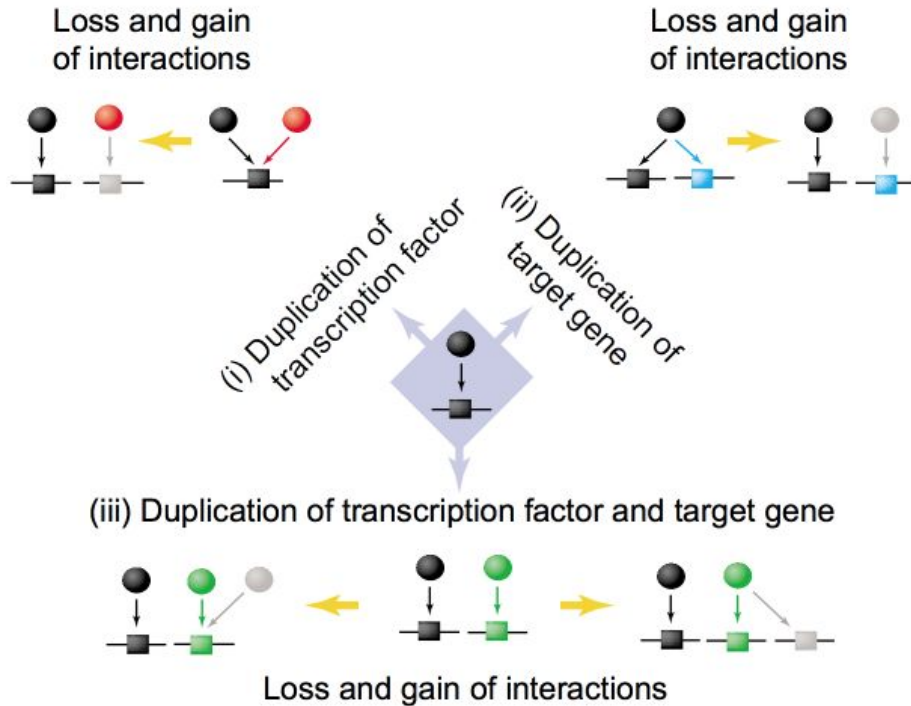
Multi-Input Motif



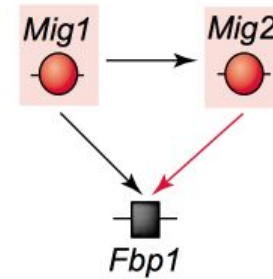
Regulator Chain



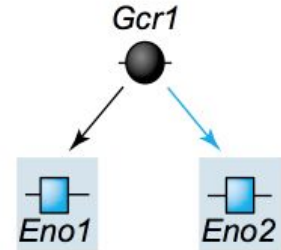
Network evolution



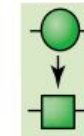
(i) Duplication of transcription factor



(ii) Duplication of target gene

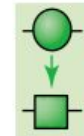


Aro80



Aro9

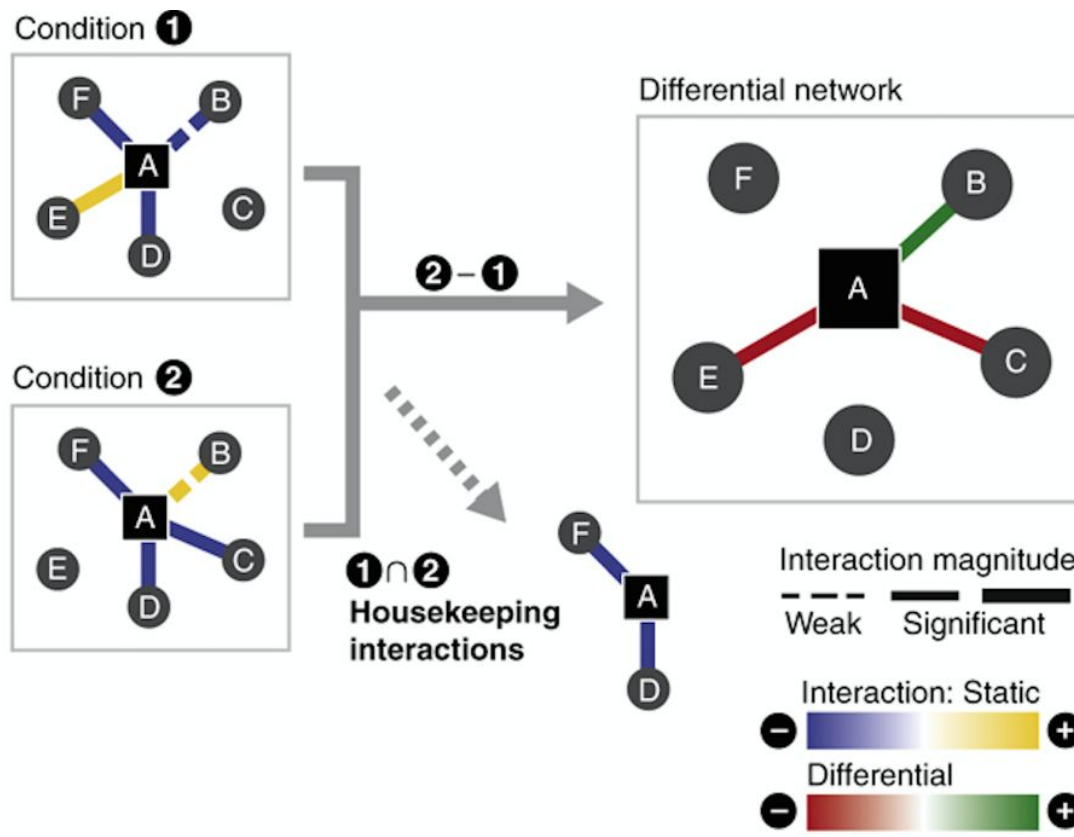
Dal81



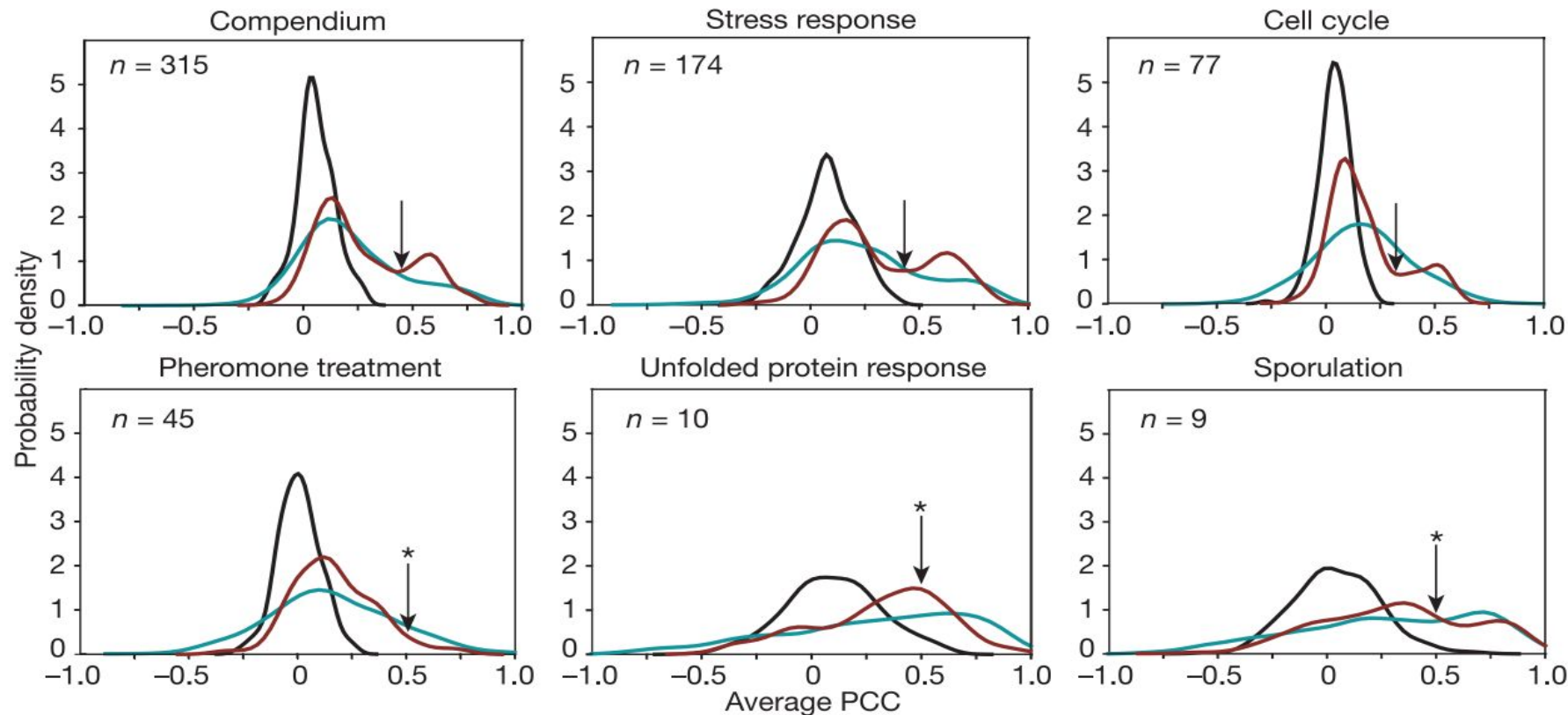
Uga1

(iii) Duplication of transcription factor and target gene

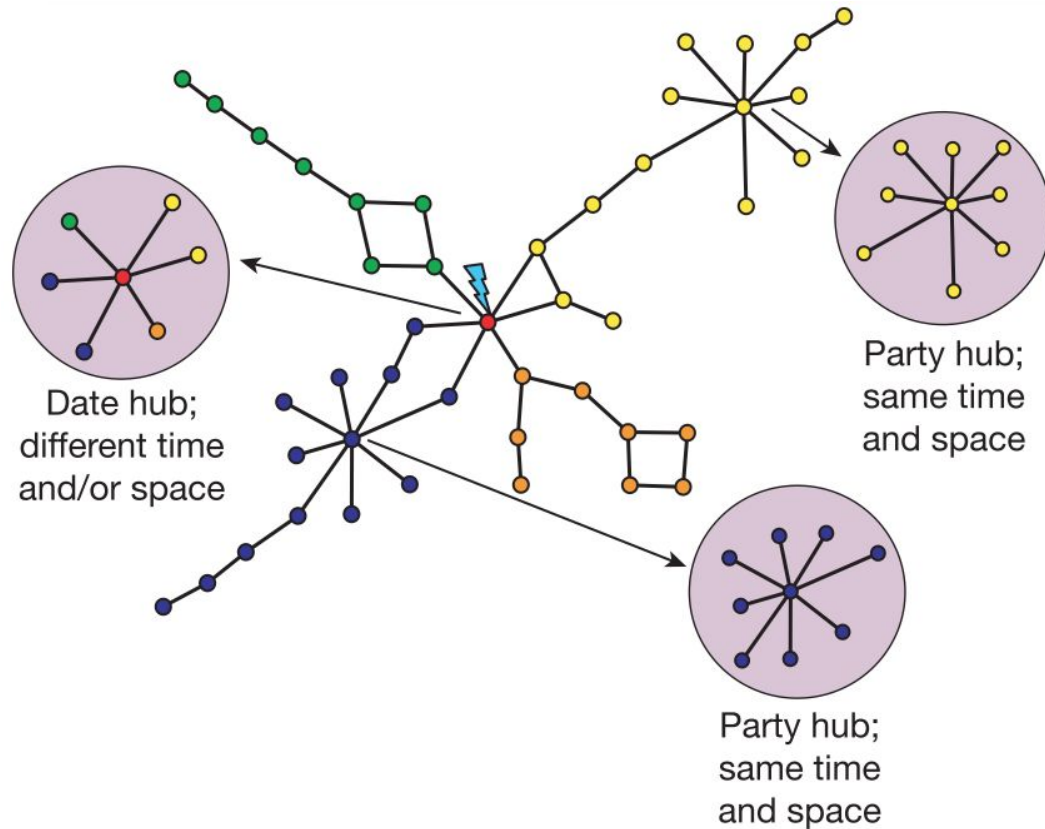
Network interactions are condition-specific



Network interactions are condition-specific

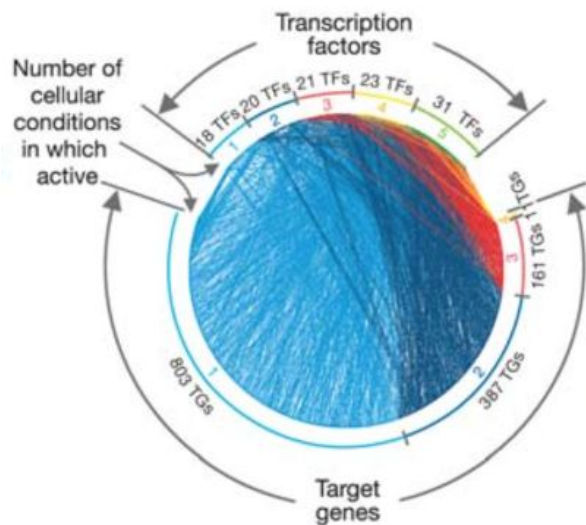


Network interactions are condition-specific



Network interactions are condition-specific

Static



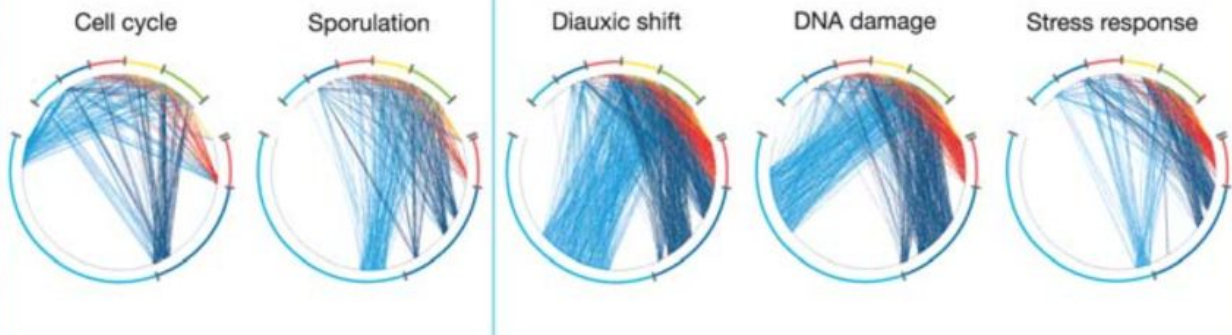
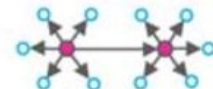
Endogenous

- Complex transcription factor combinations
- Few targets per transcription factor
- Long path lengths
- Highly inter-connected transcription factors
- Many feed-forward loops



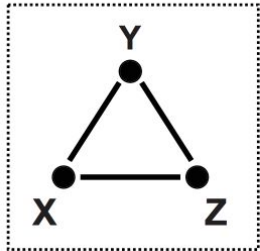
Exogenous

- Simple transcription factor combinations
- Many targets per transcription factor
- Short path lengths
- Few inter-connected transcription factors
- Many single input motifs

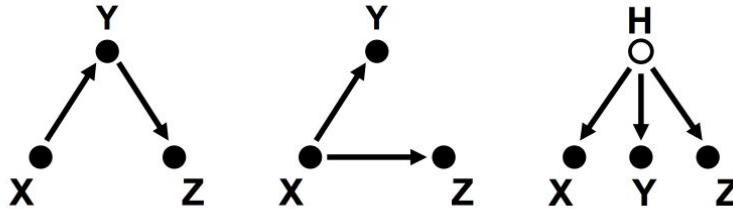


Reconstructing networks from observational data

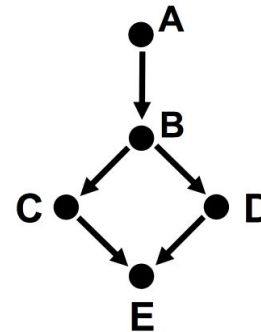
Coexpression



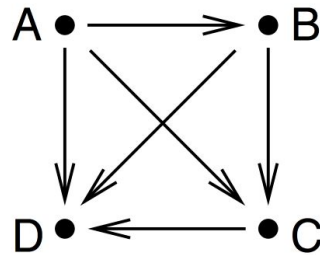
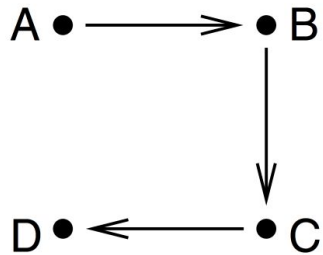
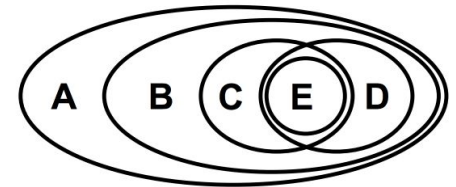
Regulatory network



Pathway structure

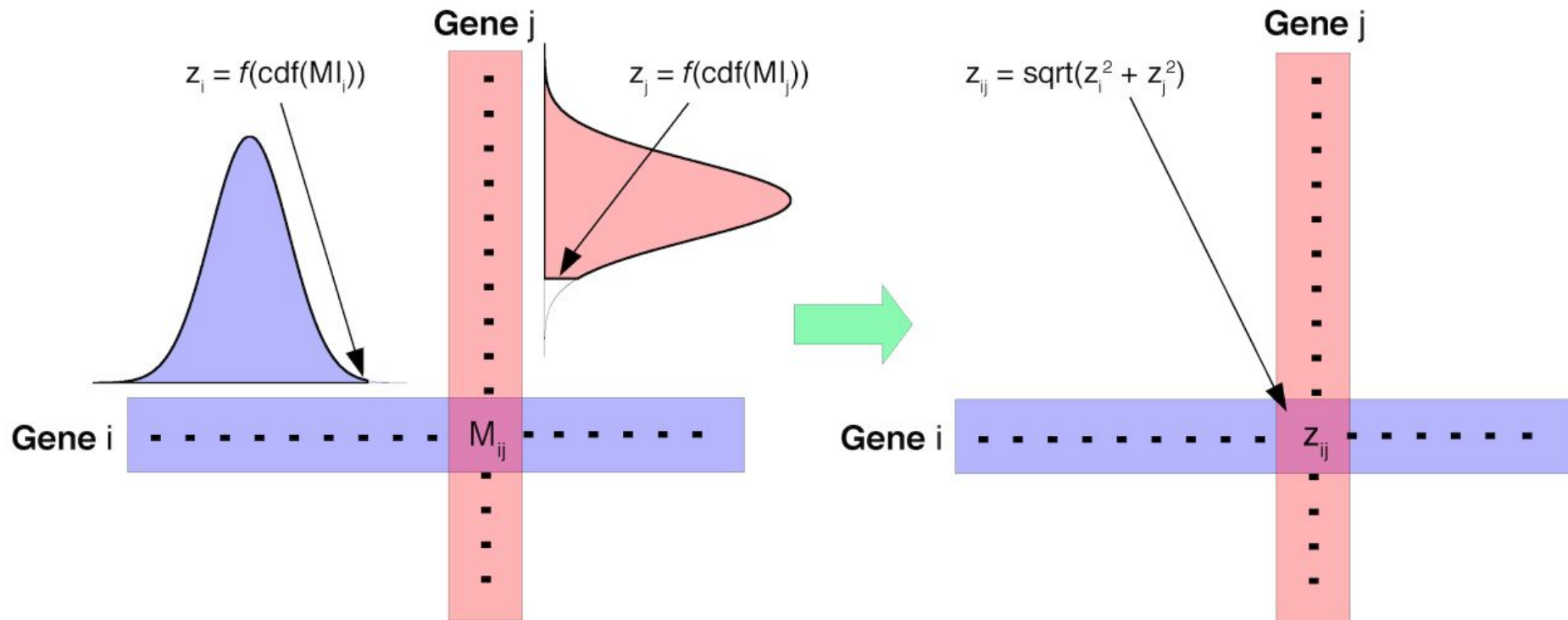


Downstream effects of interventions



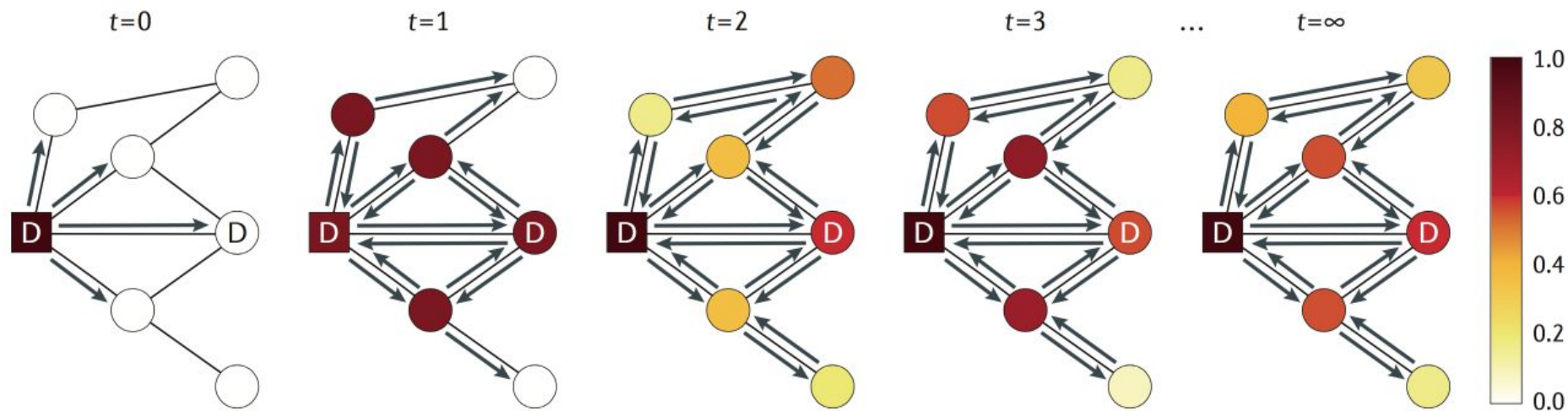
Reconstructing networks from observational data

Context likelihood relatedness



Network propagation

Tracing the flow of information through a network over time.



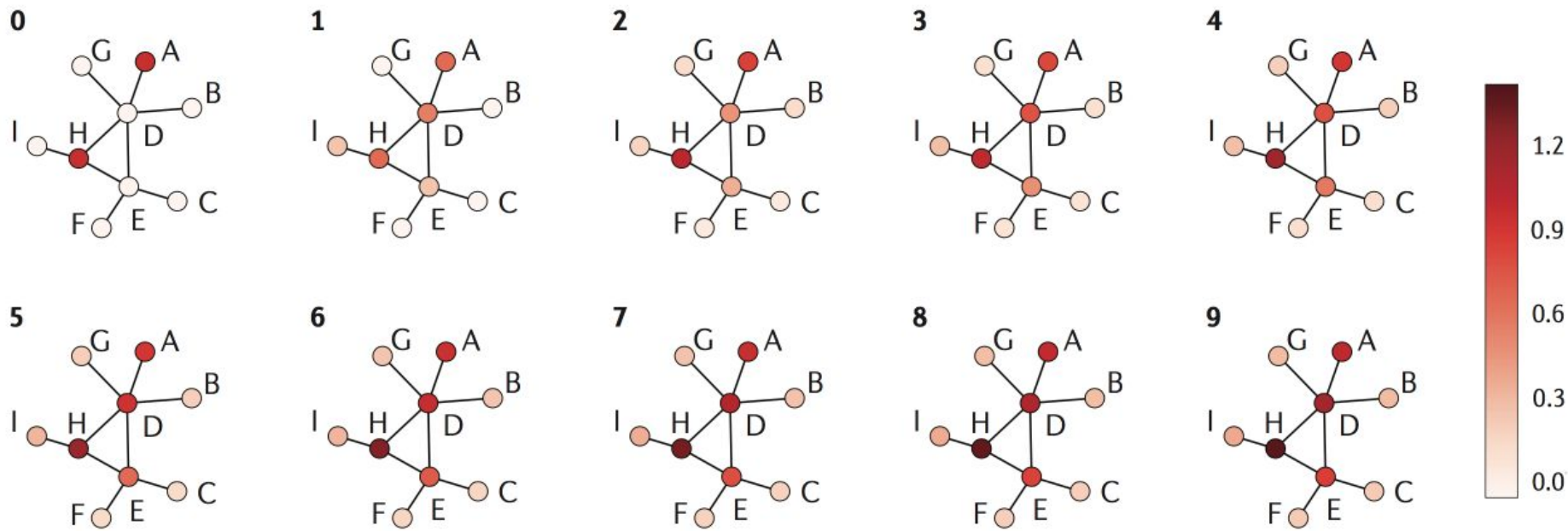
Random walk

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.

Network propagation

Tracing the flow of information through a network over time.

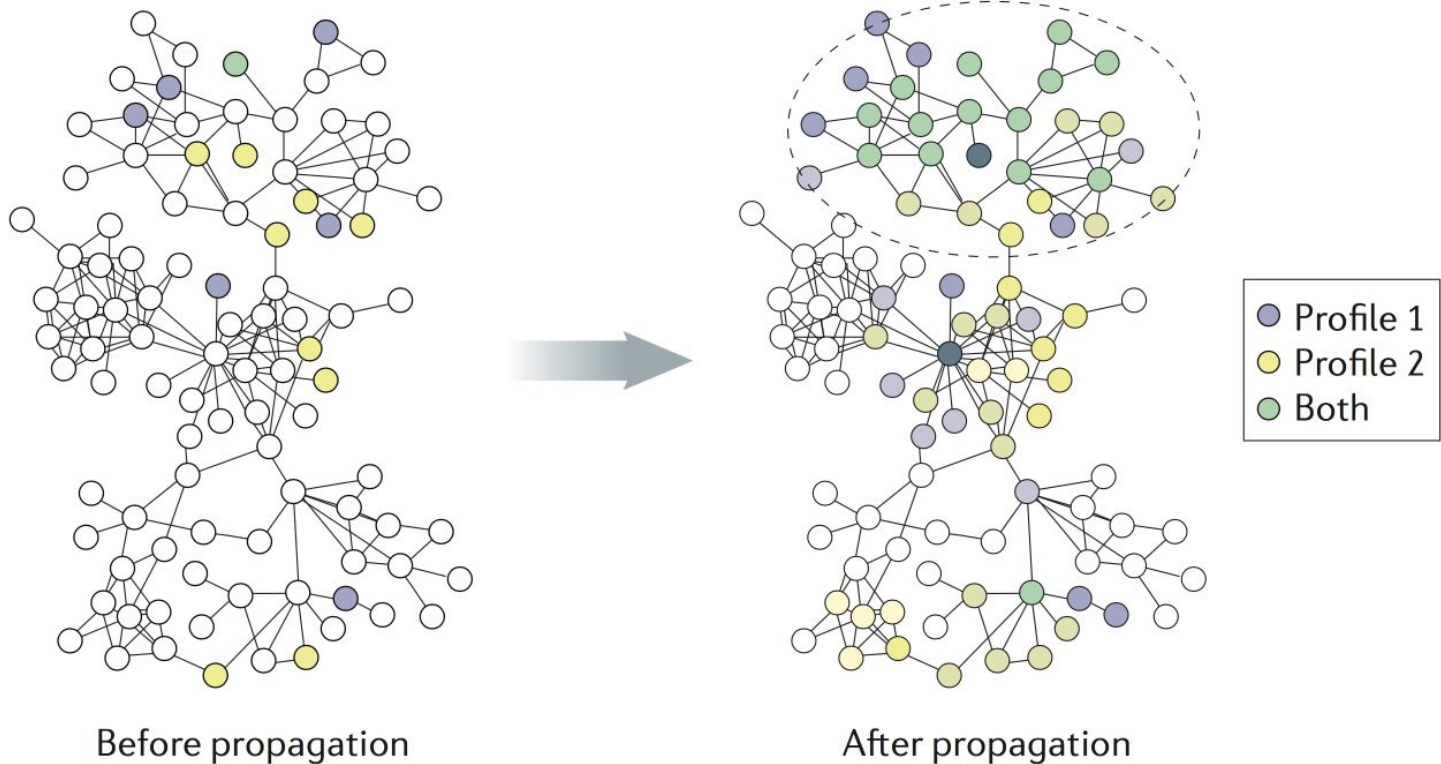
Initial node scores... (e.g. expression in a condition or association with a disease)



Convergence...

Network propagation

Tracing the flow of information through a network over time.



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)

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

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

W : normalized adjacency matrix (stochastic).

$$p_0(v)$$

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

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

$$p_k = W^k p_0$$

Network propagation

Random Walk with Restart (RWR)

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

$p_k(\mathbf{v})$: node scores at time-step k .

\mathbf{W} : normalized adjacency matrix (stochastic).

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

$$p_0(\mathbf{v})$$

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

Network propagation

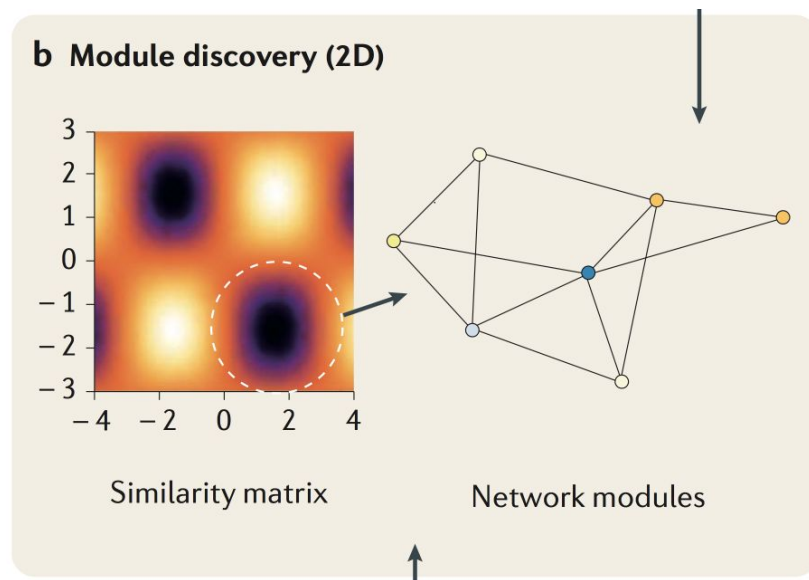
Random Walk & RWR

p : steady-state distribution of node scores.

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.

$$p = Sp_0$$



Network propagation

