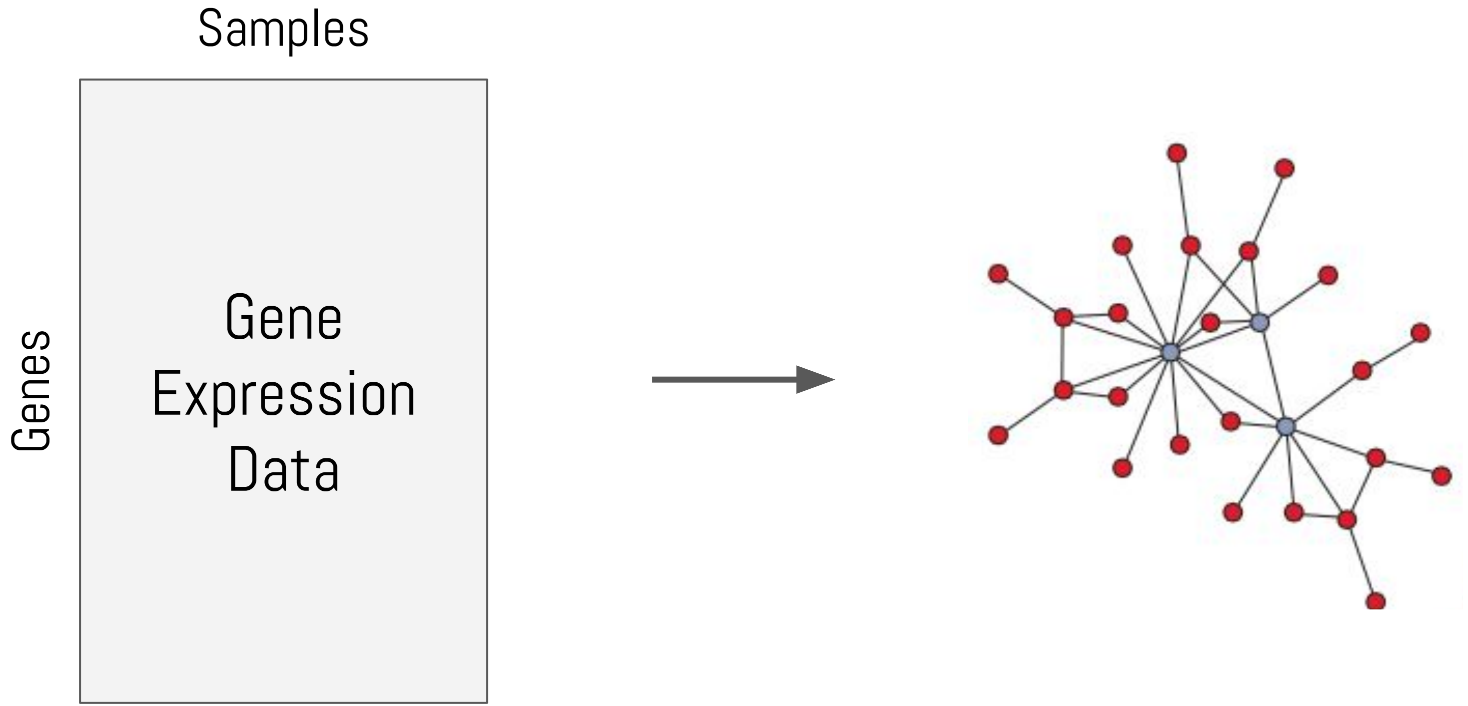
Week 15: Large-scale biological networks

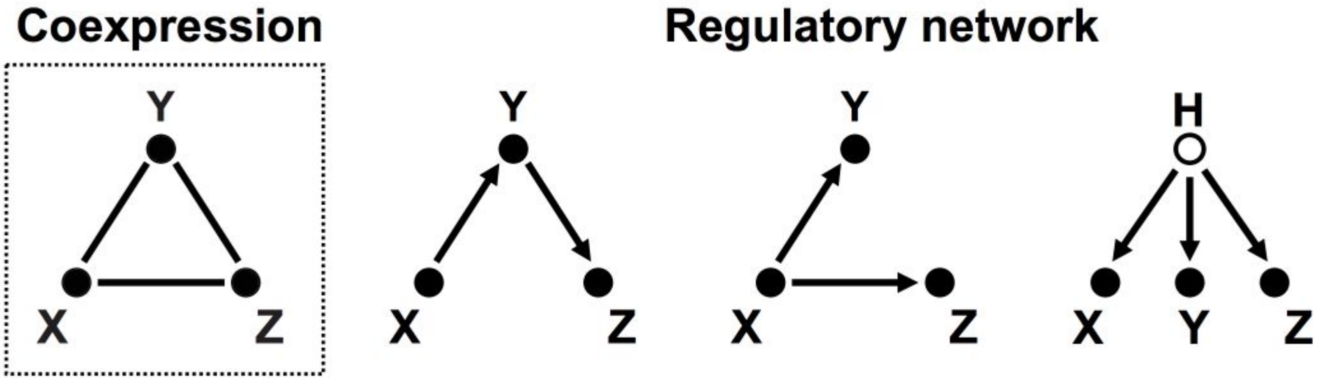
- Network topology
- Network motifs
- Condition-specific networks
- Network reconstruction
- Network propagation

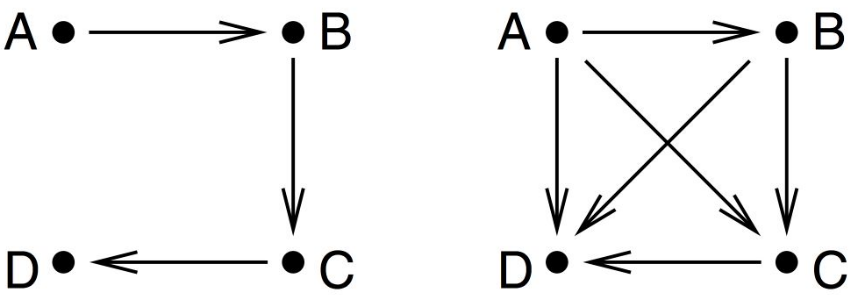
Reconstructing networks from observational data



Markowetz, Spang (2007) BMC Bioinfo.

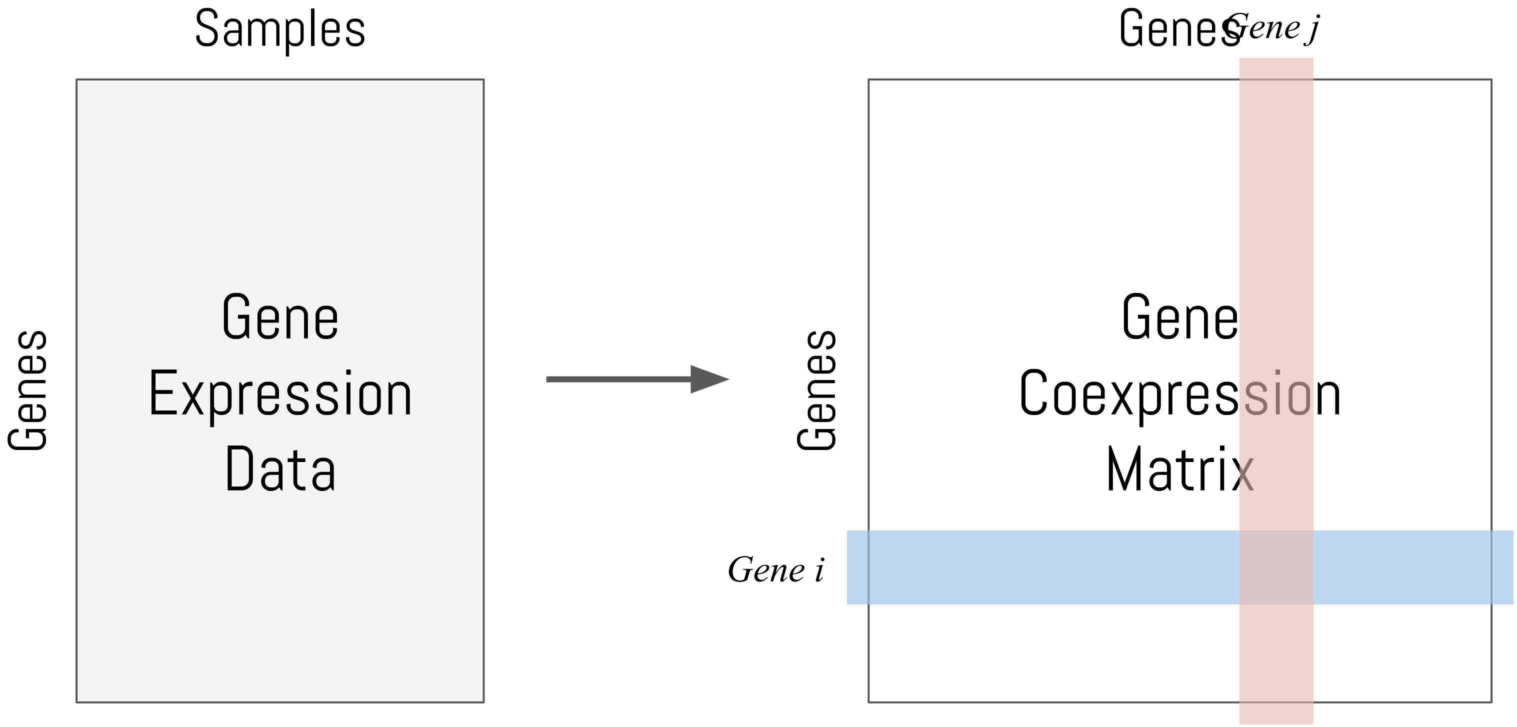
Reconstructing networks from observational data





Markowetz, Spang (2007) BMC Bioinfo.

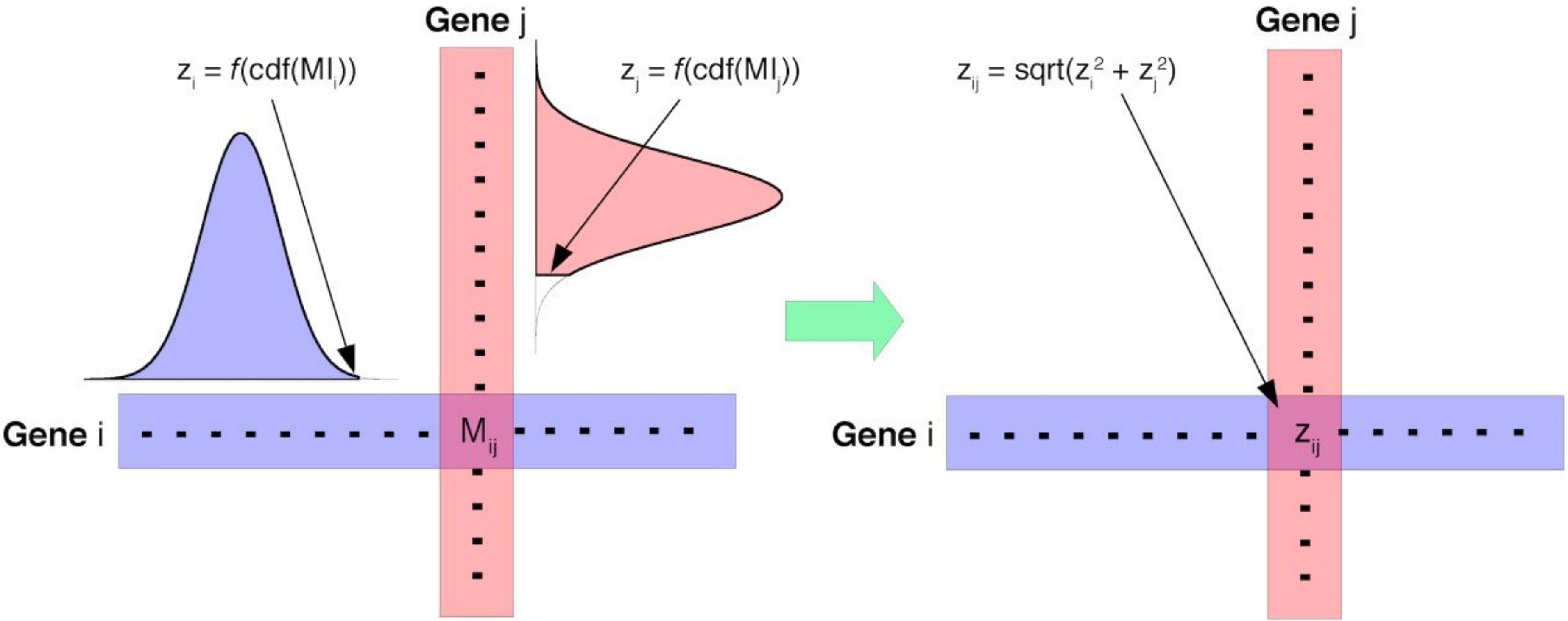
Reconstructing networks from observational data



Markowetz, Spang (2007) BMC Bioinfo.

Reconstructing networks from observational data

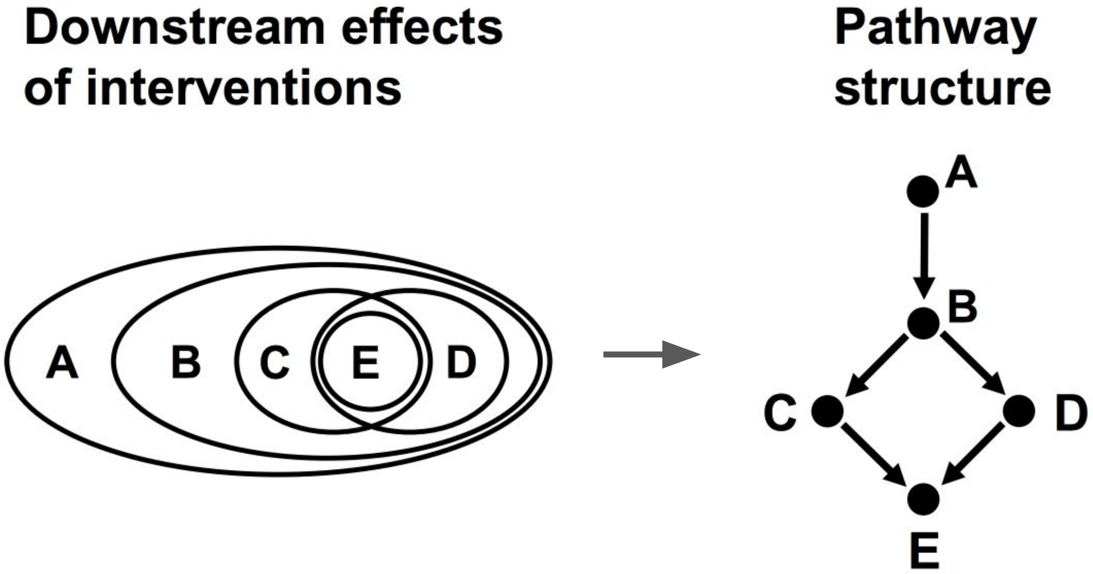
Context likelihood relatedness



Faith (2007) PLoS Biology

Reconstructing networks from observational data

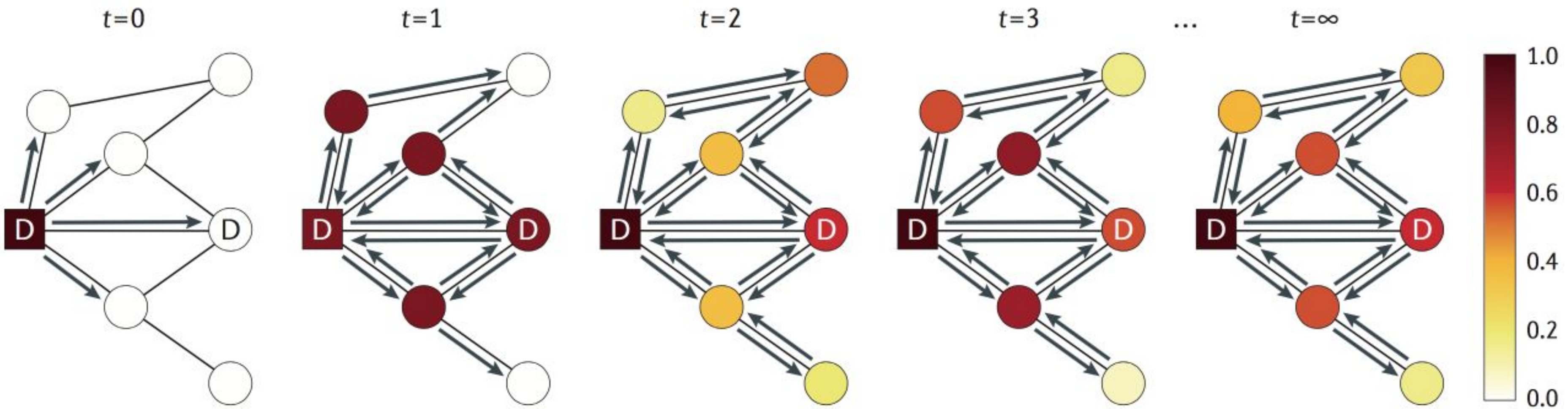
Single
gene
perturbations



Markowetz, Spang (2007) BMC Bioinfo.

Network propagation

Tracing the flow of information through a network over time.



Random walk

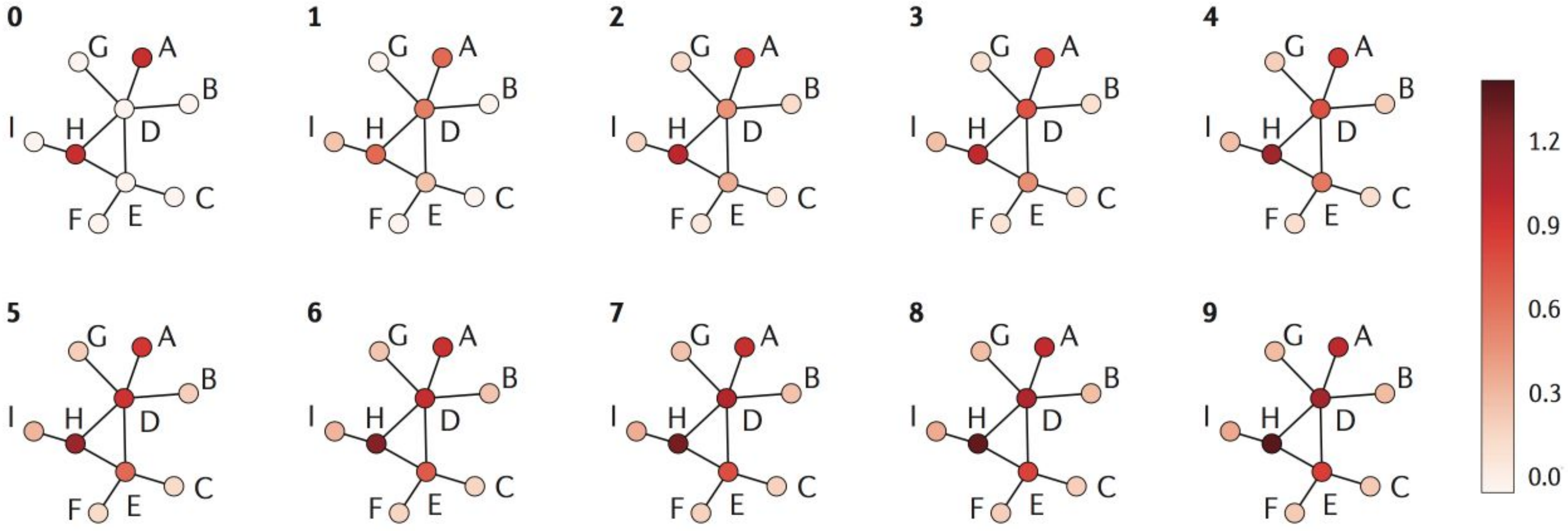
A mathematical formalization of the paths resulting from successive random steps a 'walker' takes from **one node to another** with a probability that is **proportional to the weight of the edge** connecting the nodes.

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)

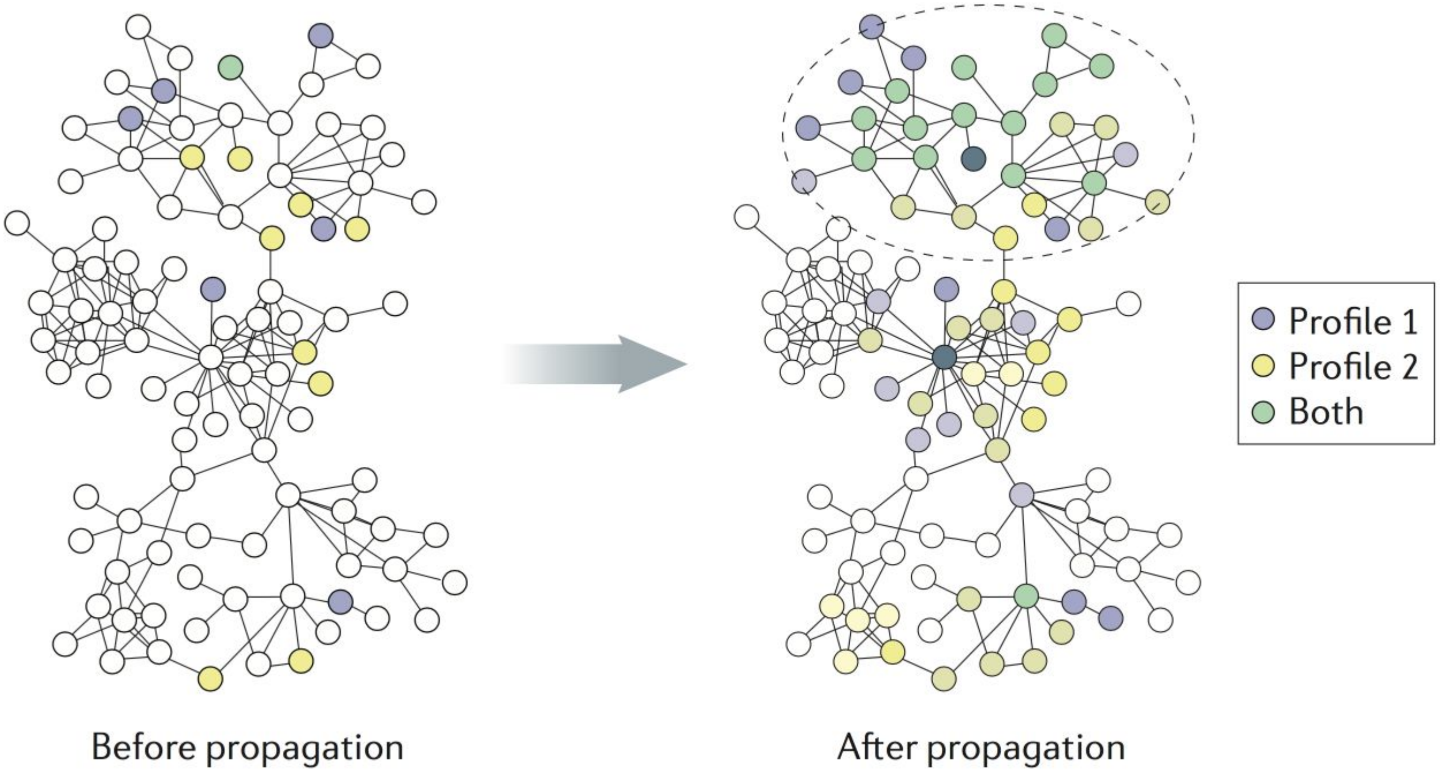


Convergence...

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)

$$p_0(v)$$

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

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

W: normalized adjacency matrix (stochastic).

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

Cowen (2017) Nat. Rev. Genet.

Network propagation

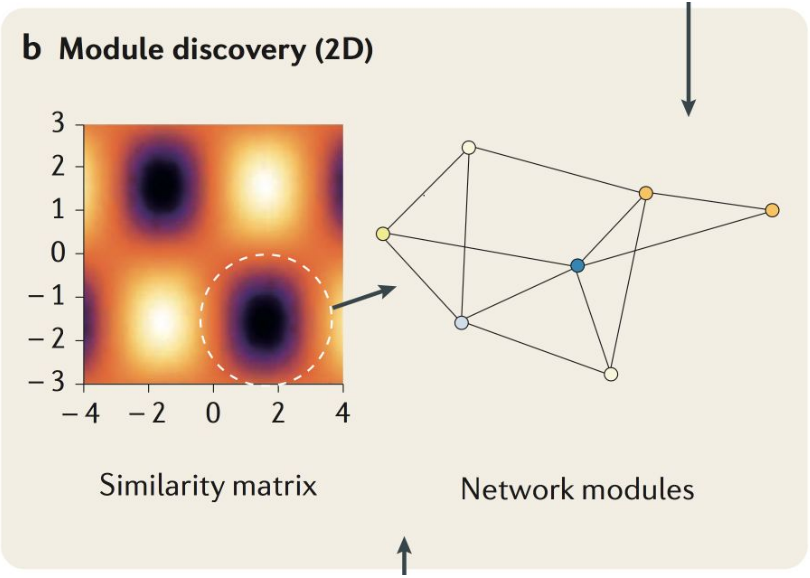
Random Walk & RWR

p: steady-state distribution of node scores.

$$p = Sp_0$$

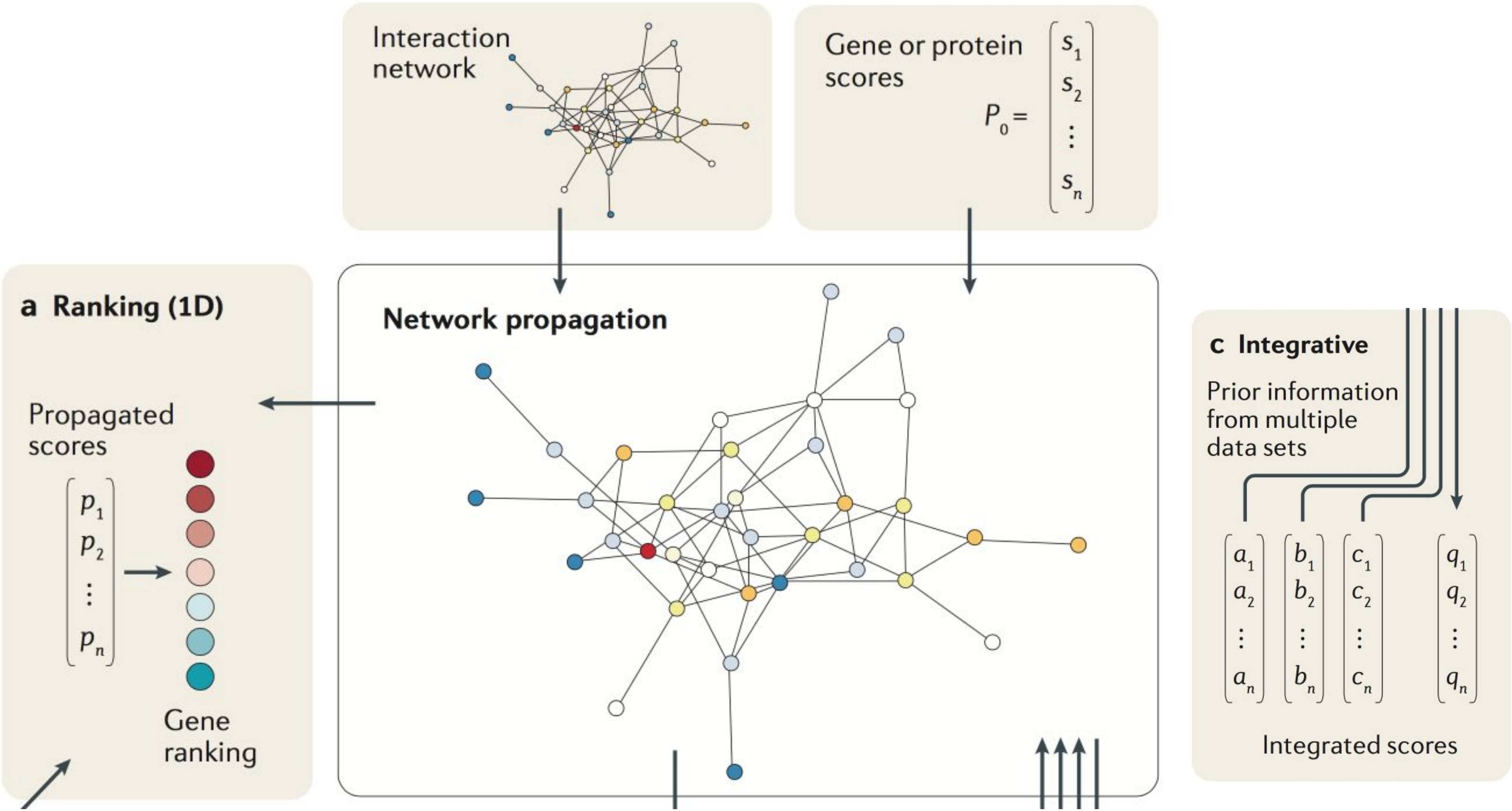
S: Can be interpreted as a similarity matrix.

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



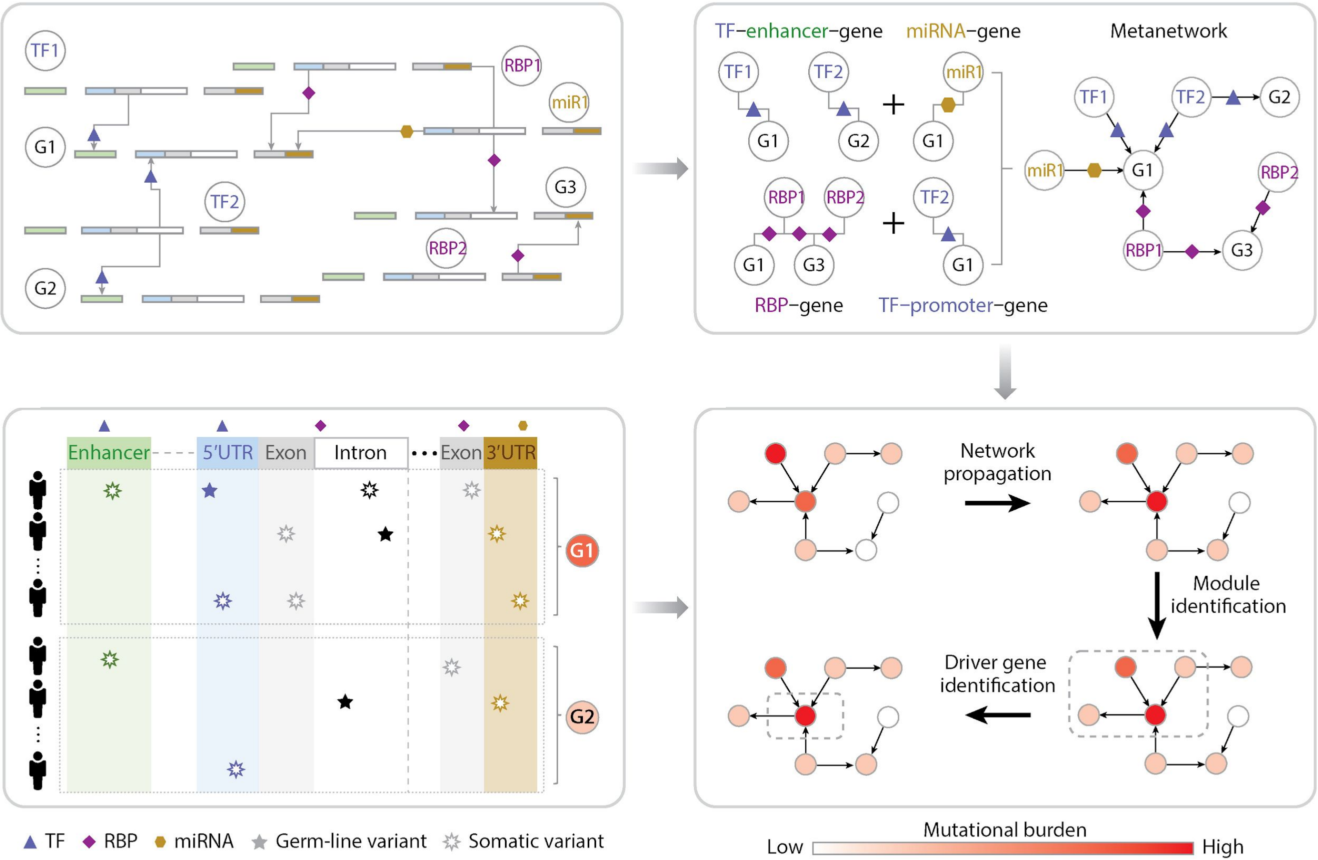
Cowen (2017) Nat. Rev. Genet.

Network propagation



Cowen (2017) Nat. Rev. Genet.

Network propagation



McGillivray (2018) Annu. Rev. Biomed. Data Sci.