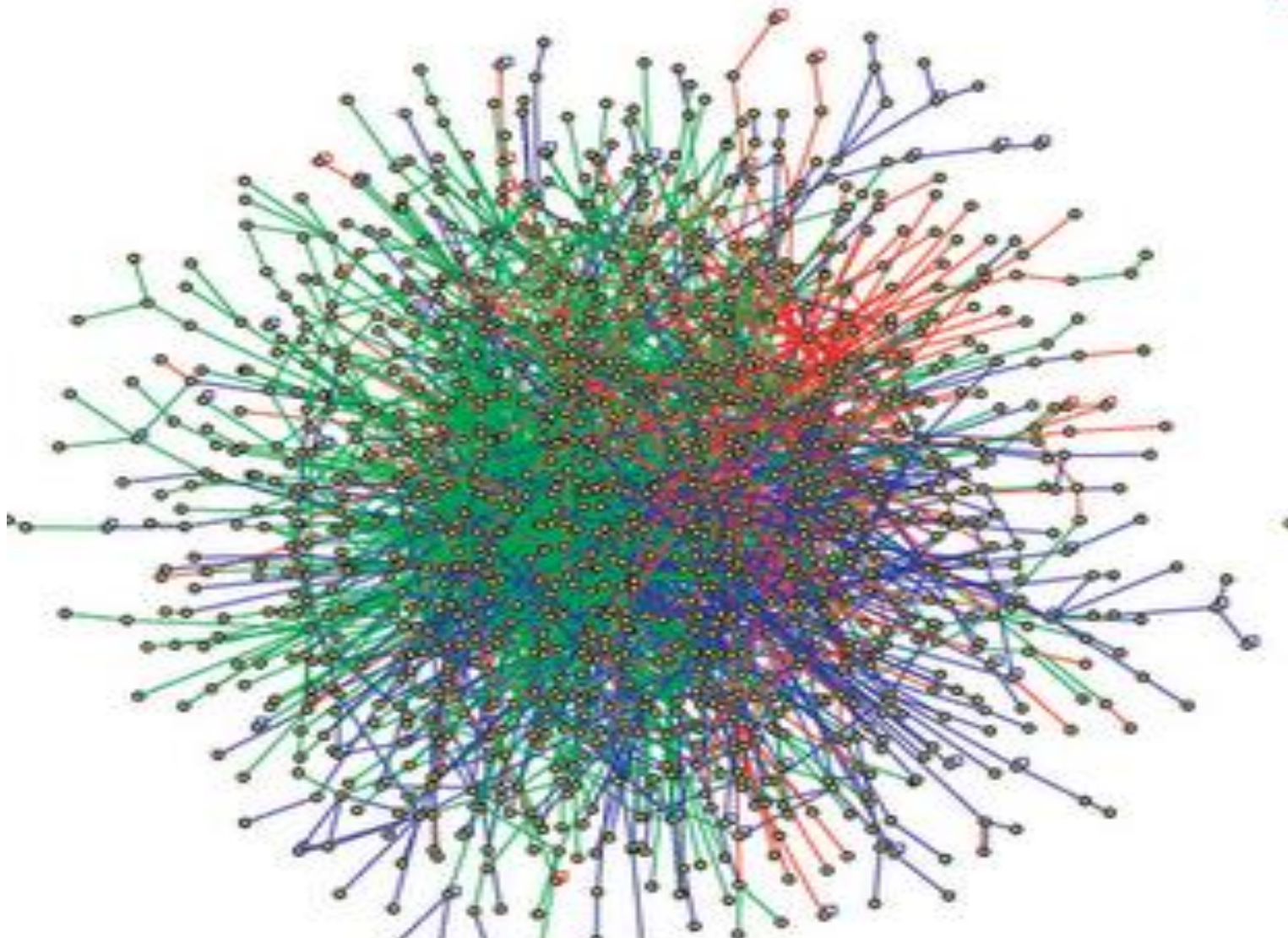


Networks

Day 6

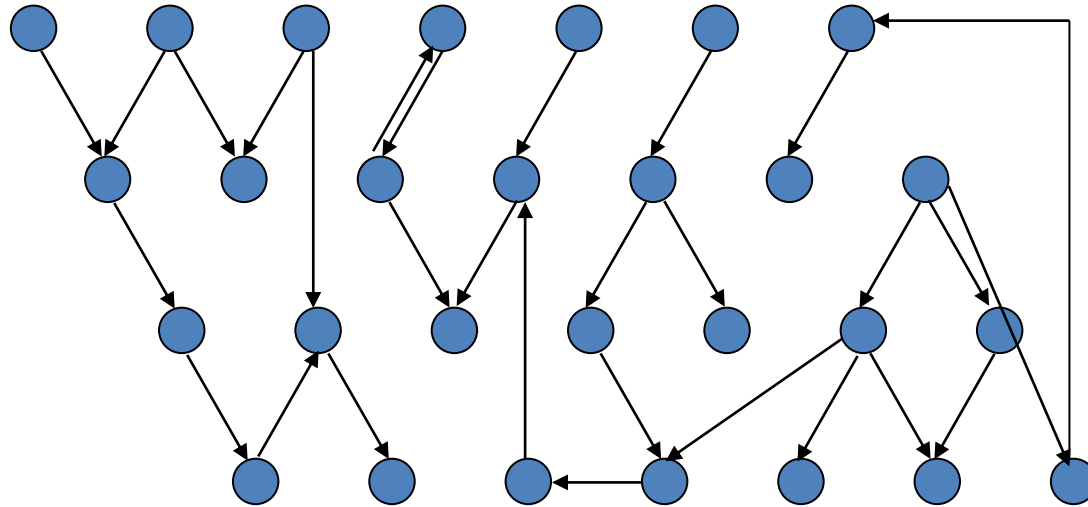
Human interactome

~130,000 binary interactions, most of which remain to be mapped



Networks

$$G = (V, E)$$



**Transcription
network**

**Signaling
network**

**Metabolic
network**

nodes V :



transcription factor
genes

(genes of) signaling
components

metabolites

edges (arcs)

E :



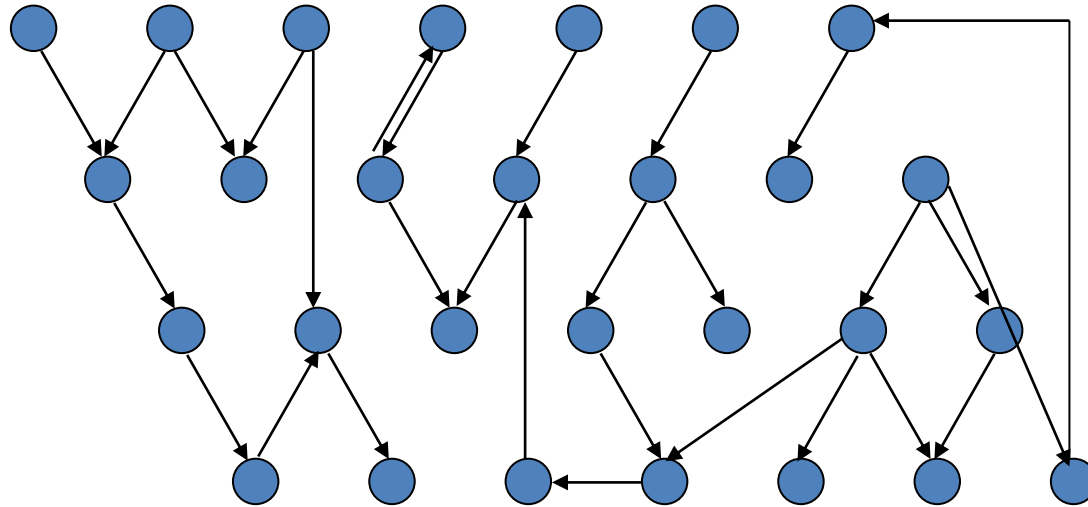
TF expression &
transcriptional
regulation

transmission
of a signal

enzymatically
catalyzed
conversions

Networks

$$G = (V, E)$$



**Transcription
network**

**Signaling
network**

**Metabolic
network**

nodes V :



transcription factor
genes

(genes of) signaling
components

metabolic enzyme
genes

edges (arcs)

E :



TF expression &
transcriptional
regulation

transmission
of a signal

forwarding a
metabolite

Networks

The sources for the existing knowledge:

Transcription network: **TRANSFAC[®]**

Signaling network: **TRANSPATH[®]**

Metabolic network: **KEGG**

Graphs

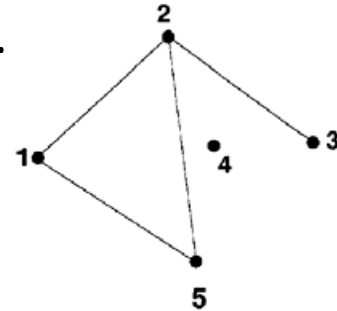
Graph - the object consisting of vertices and edges joining them.

$G=\{V, E\}$, where V - set of vertices (n), E – set of edges (m)

Graphs are cyclic and acyclic ,

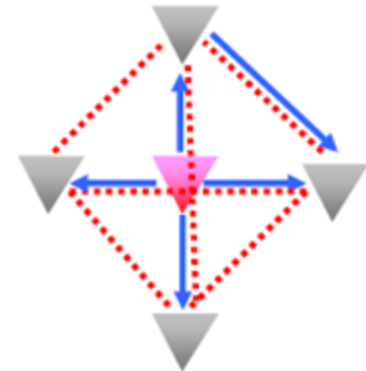
sparse if $m \sim n$, **dense** $m \sim n^2$, **complete** $m=n^2$

directed and undirected

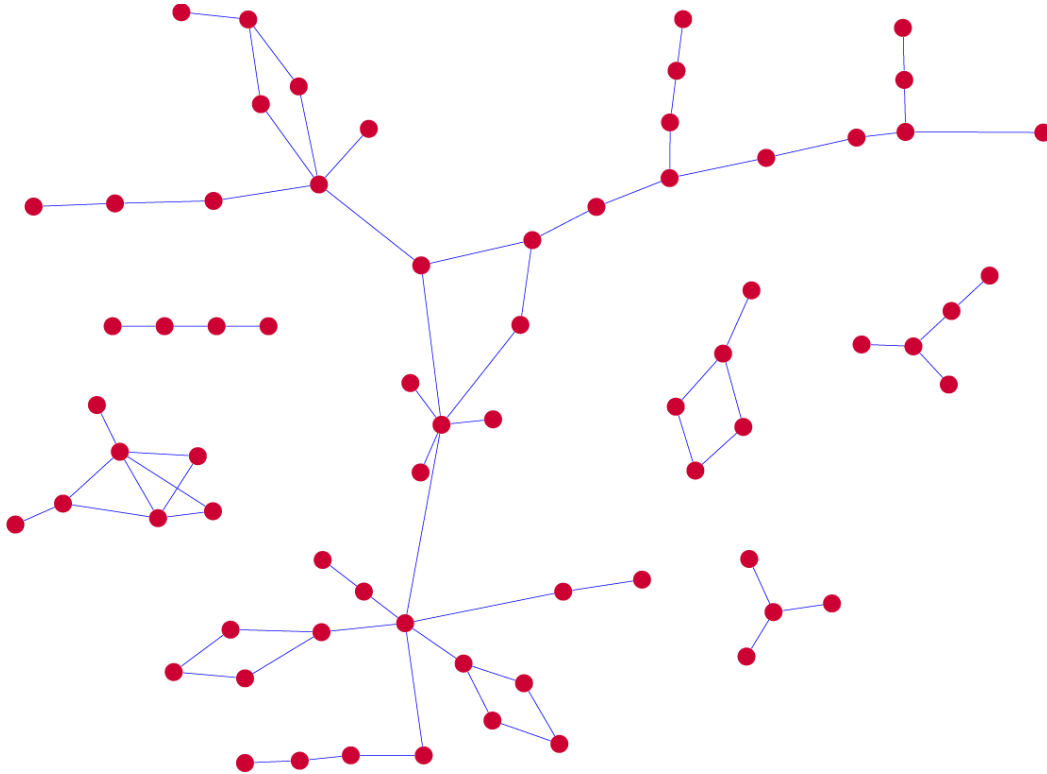


Characteristics of the topology

- Degree k_i
- Distribution of degree $P(k)$
- Average path length
- Network diameter
- Clustering coefficient

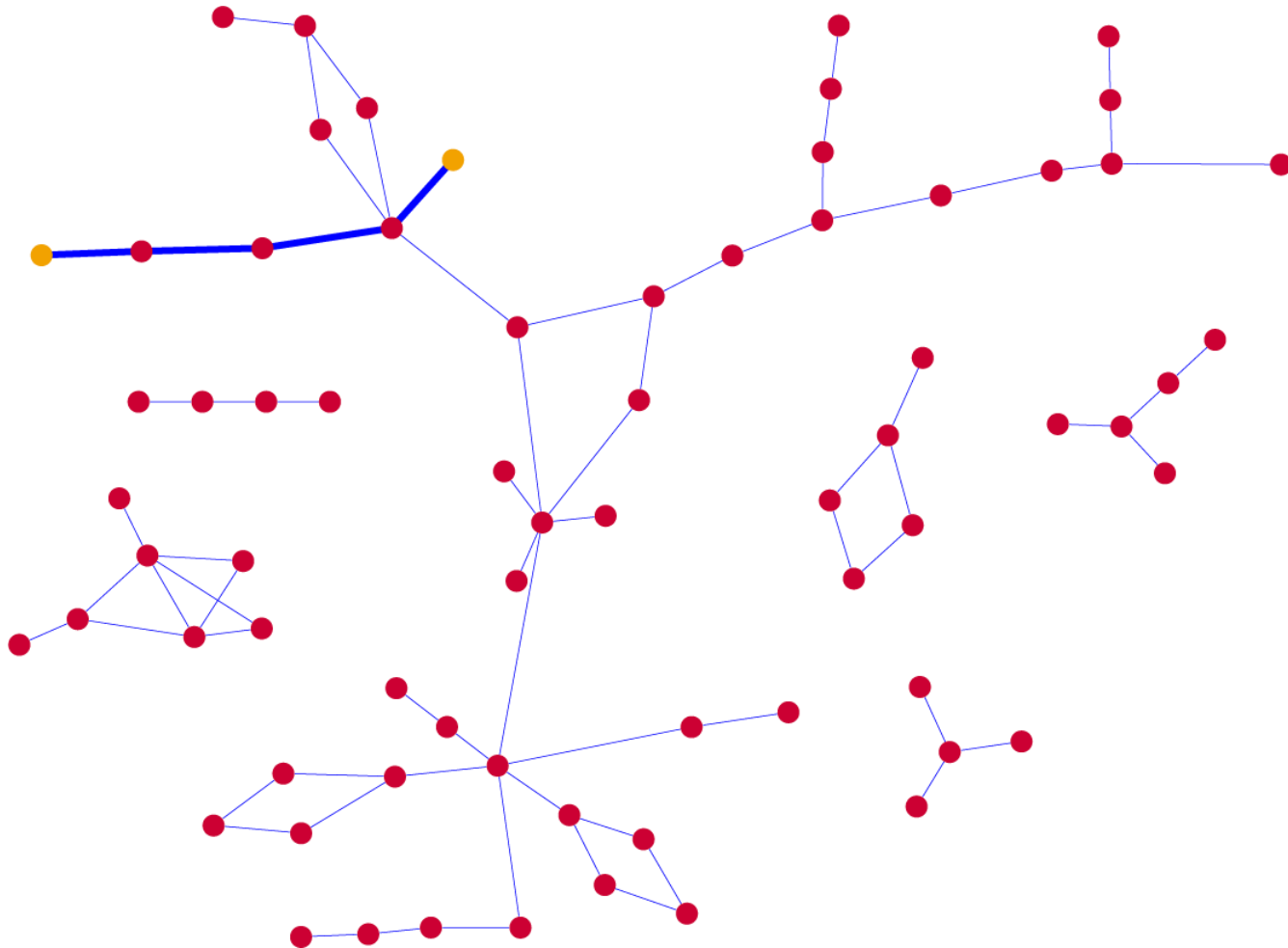


Undirected graph

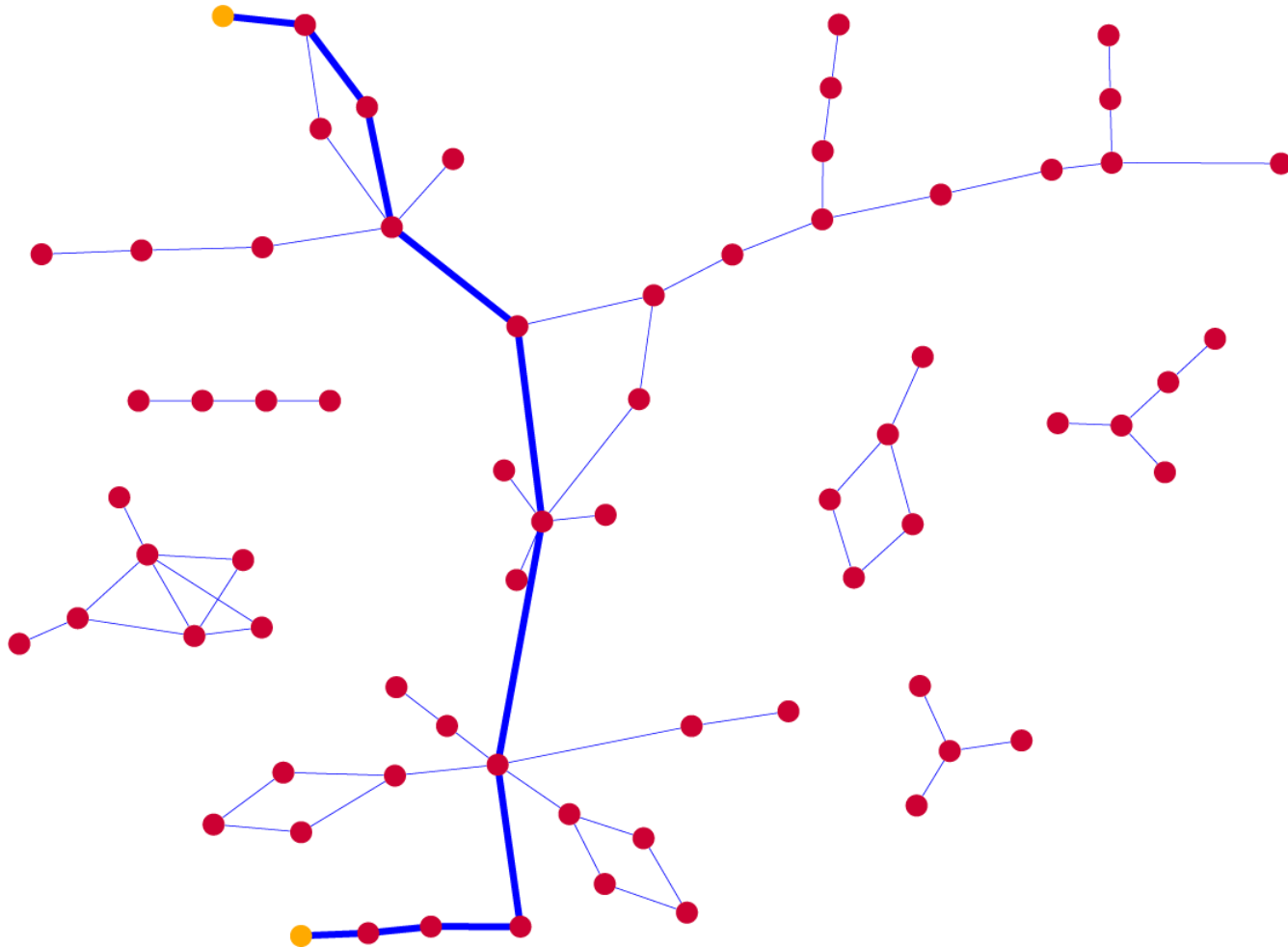


- $G(V,E)$
- $|V| = 69$
- $|E| = 71$

The shortest path between vertices

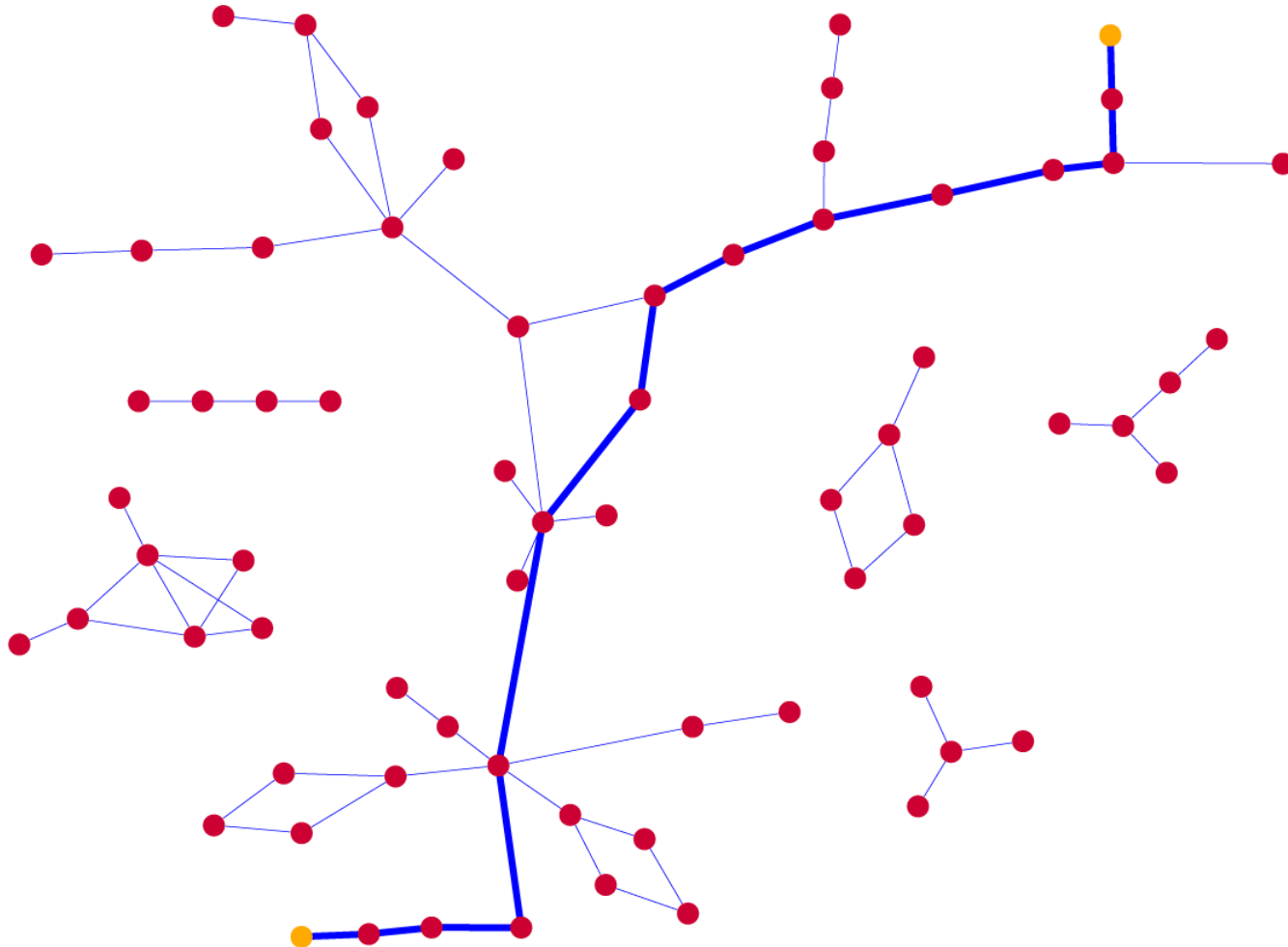


The shortest path between vertices



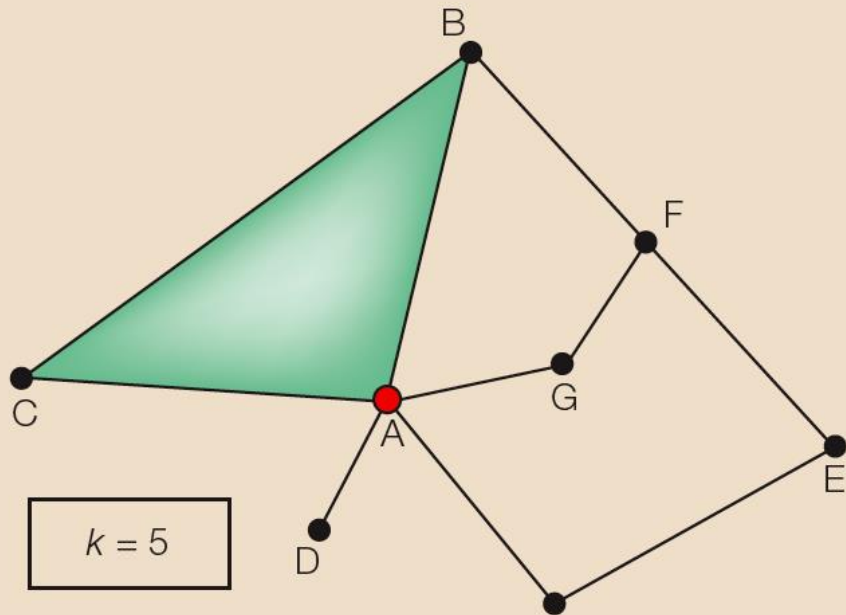
Diameter - the maximum distance in the network.

Longest Shortest-Path

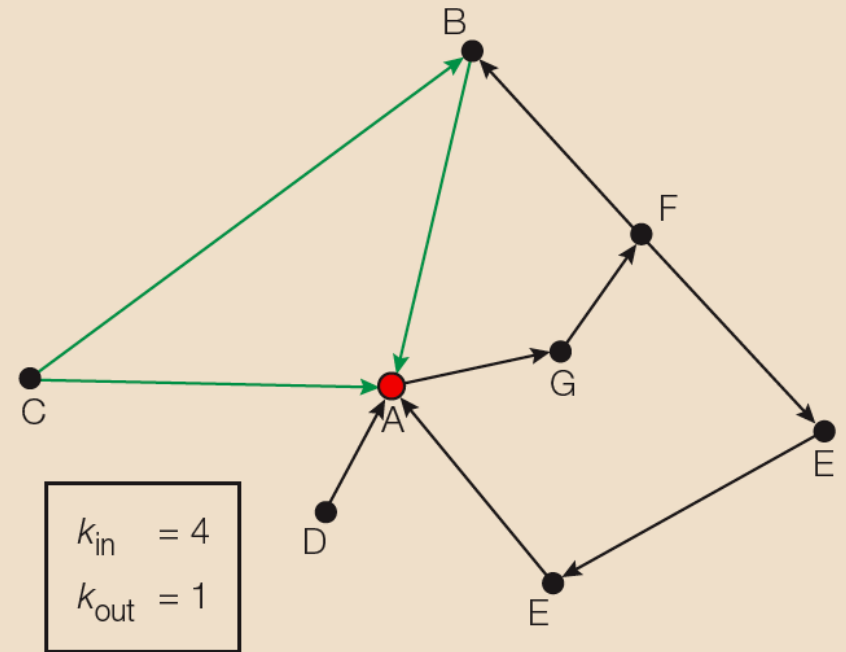


Characteristics of the network - the degree of vertex

a Undirected network



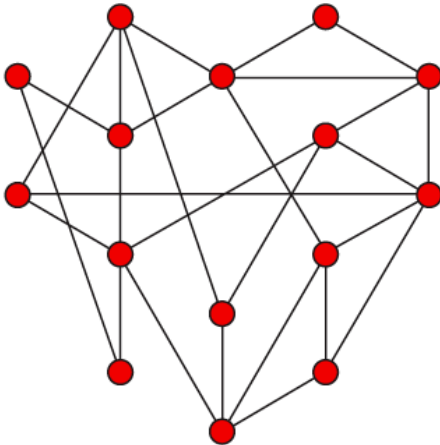
b Directed network



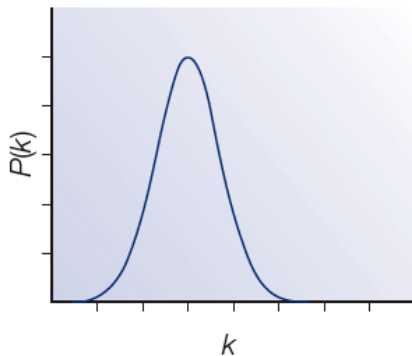
Distribution of degree

A Random network

Aa

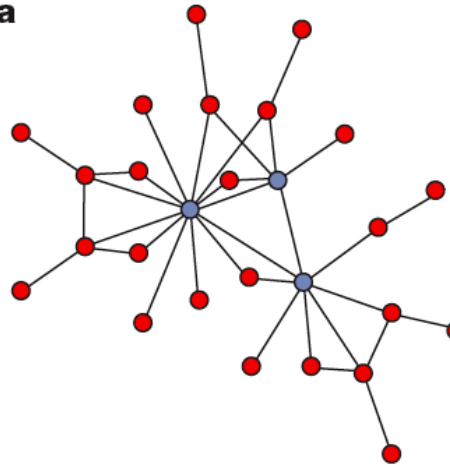


Ab

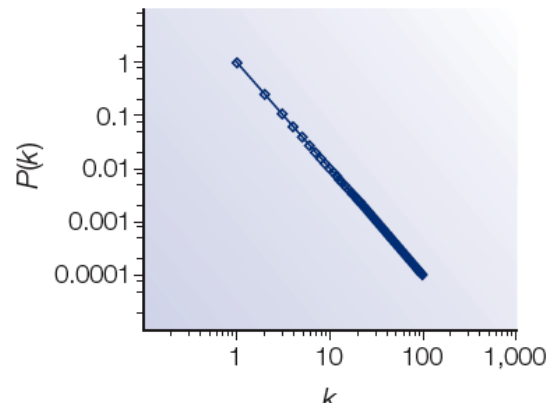


B Scale-free network

Ba



Bb



$P(k)$ is probability of each degree k , i.e fraction of nodes having that degree.

For random networks, $P(k)$ is normally distributed.

For real networks the distribution is often a power-law:

$$P(k) \sim k^{-\gamma}$$

Such networks are said to be **scale-free**

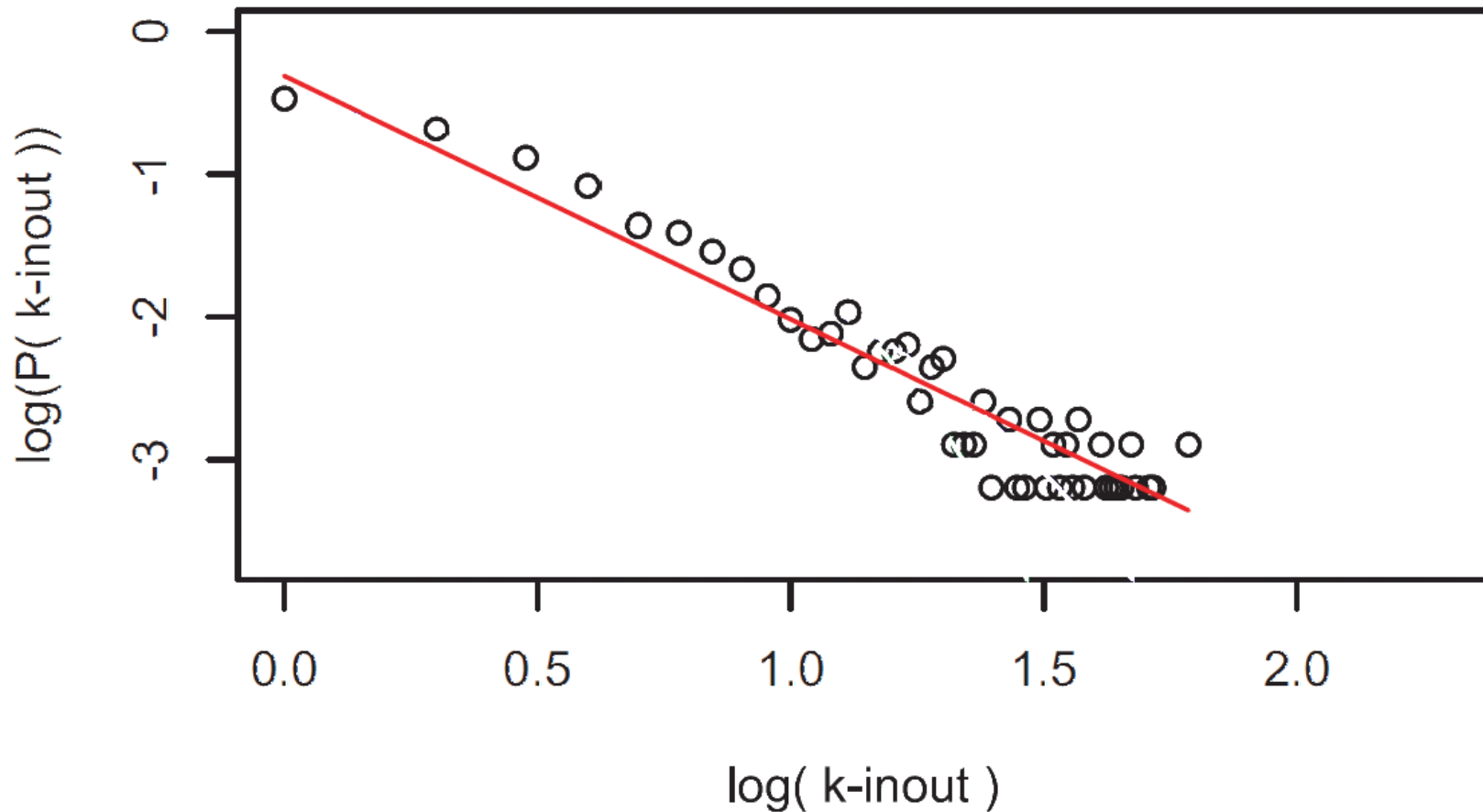
Mammalian network analysis

Scale-freeness

$$\gamma = 1.88$$

Signaling network

inout-degree

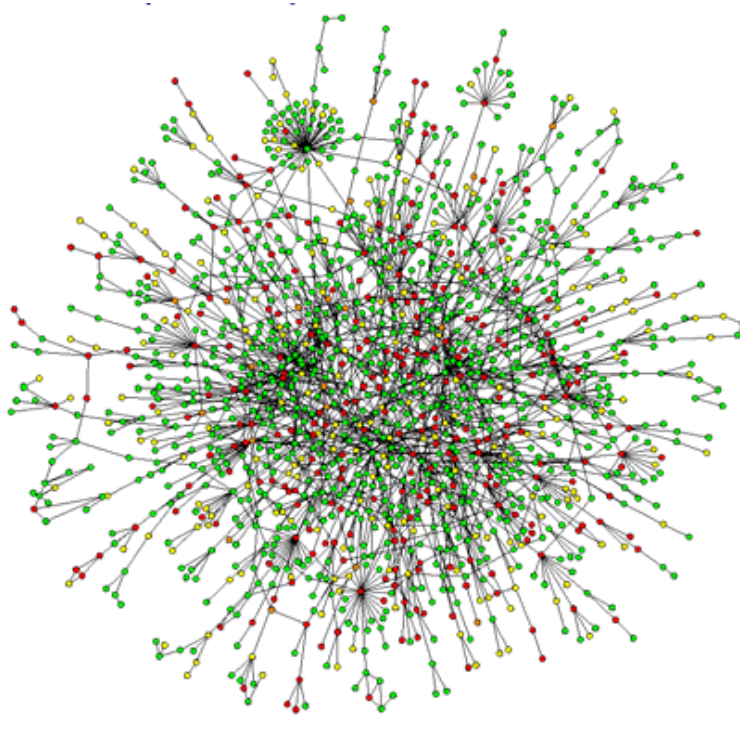


$$P(k) \propto k^{-\gamma}$$

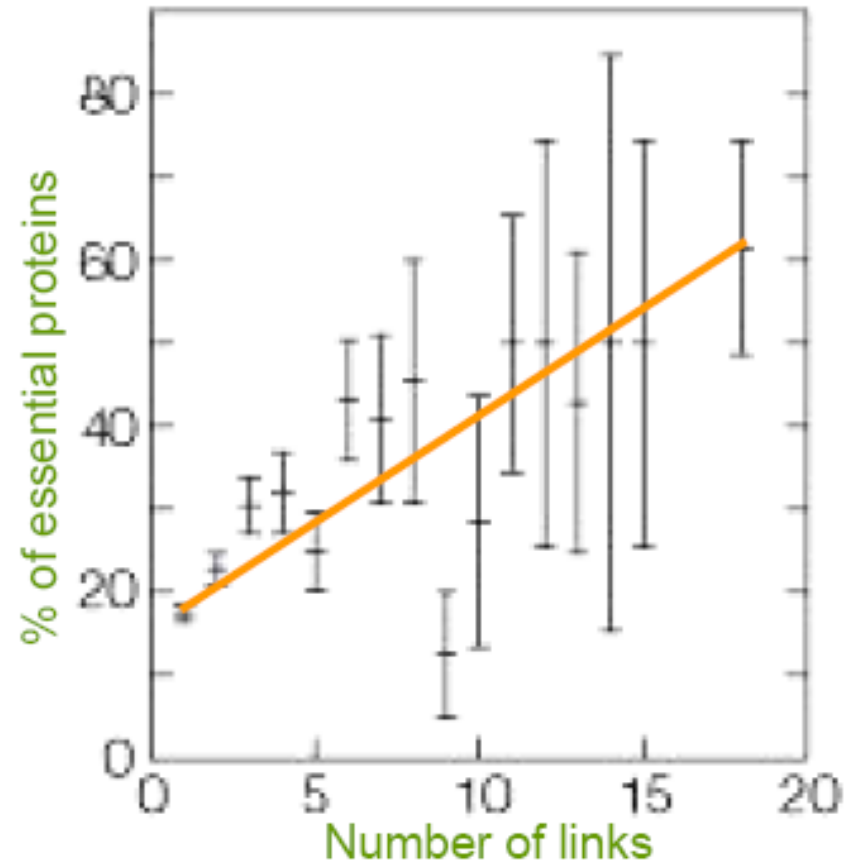
Scalable networks are stable (Robust)

- Complex systems (cell, internet, social networks), are resilient to component failure
- Network topology plays an important role in this robustness
 - Even if ~80% of nodes fail, the remaining ~20% still maintain network connectivity
- *Attack vulnerability* if hubs are selectively targeted
- In yeast, only ~20% of proteins are lethal when deleted, and are 5 times more likely to have degree $k > 15$ than $k < 5$.

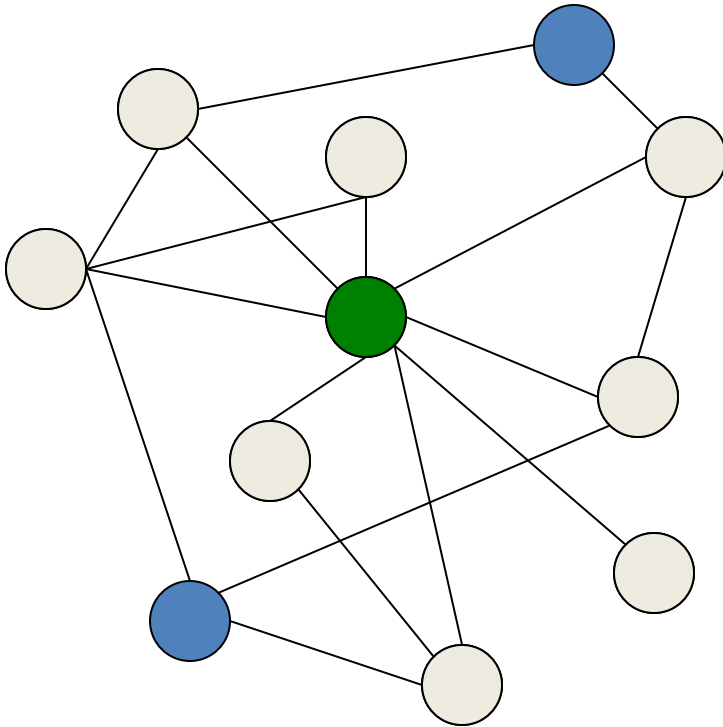
Lethality and connectedness are positively correlated



- Lethal
- Slow-growth
- Non-lethal
- Unknown



Clusters



The center node has 8 (grey) neighbors

There are 4 edges between the neighbors

$$C = 2 \cdot 4 / (8 \cdot (8-1)) = 8/56 = 1/7$$

Clustering coefficient of vertex i - is the ratio of the number of edges E_i neighbors to the number of possible edges that might exist in neighbors. Measure of network connectivity. Clustering coefficient for networks C - is the average ratio of all vertices C_i

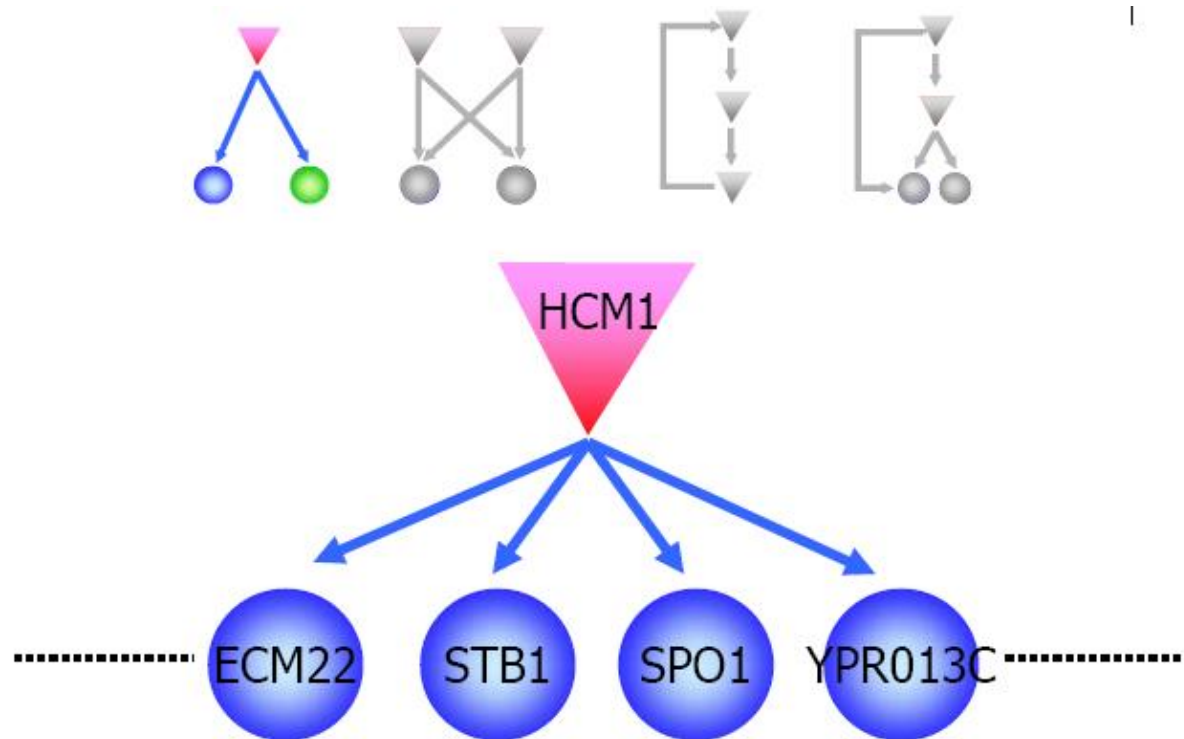
$$C_i = \frac{n_i}{\binom{k}{2}} = \frac{2n_i}{k \cdot (k-1)}$$

k : neighbors of i

n_i : edges between node i 's neighbors

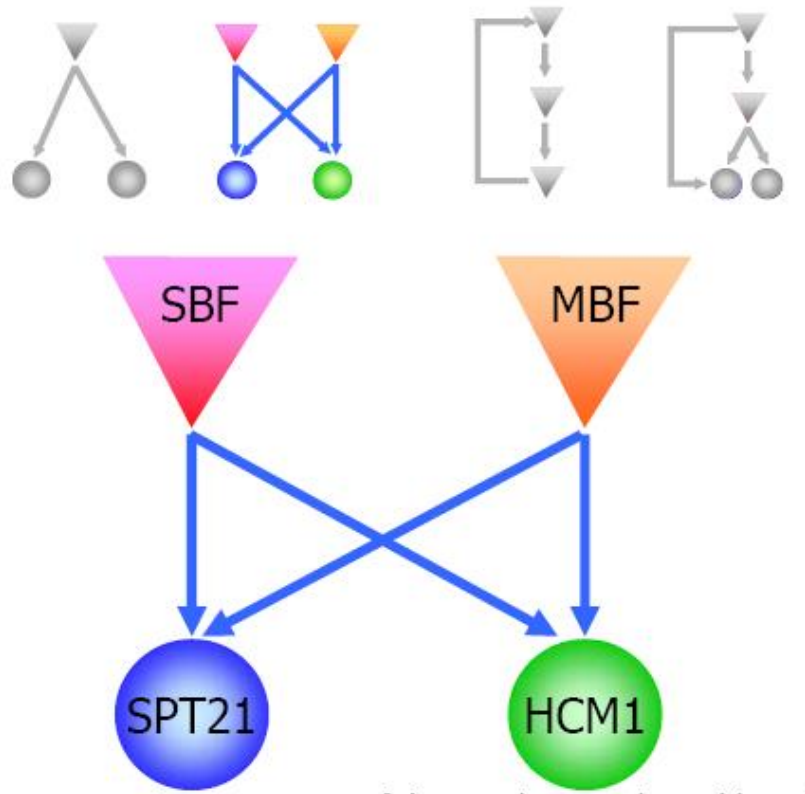
Regulatory motifs

Single input motifs



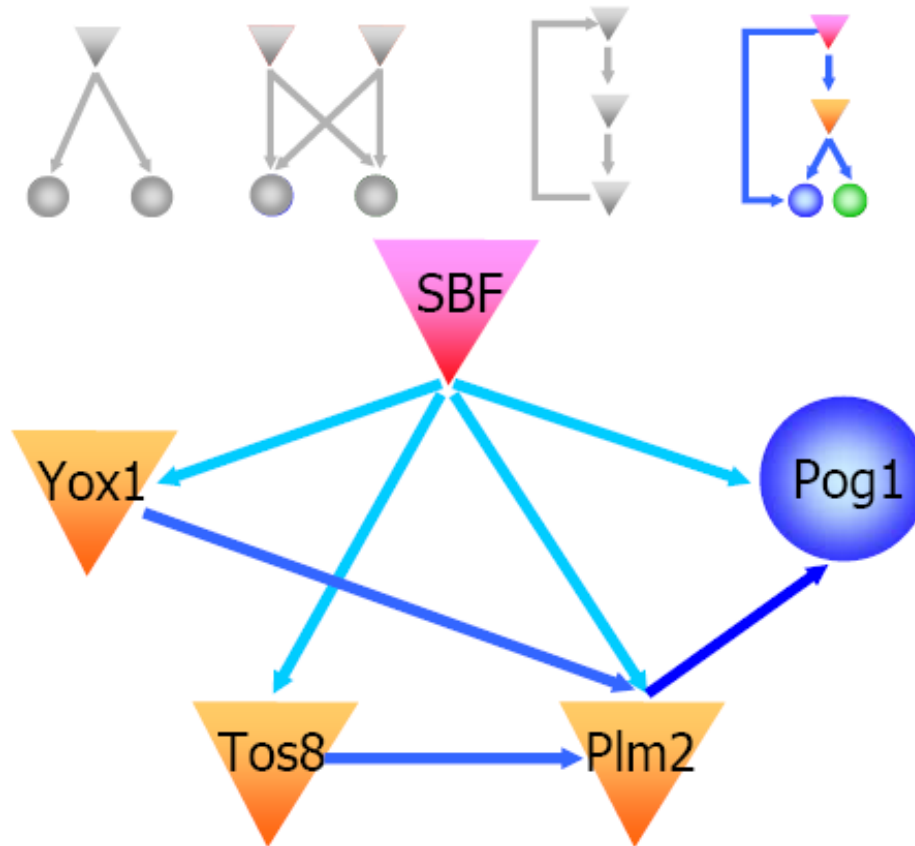
Regulatory motifs

Multiple output motifs



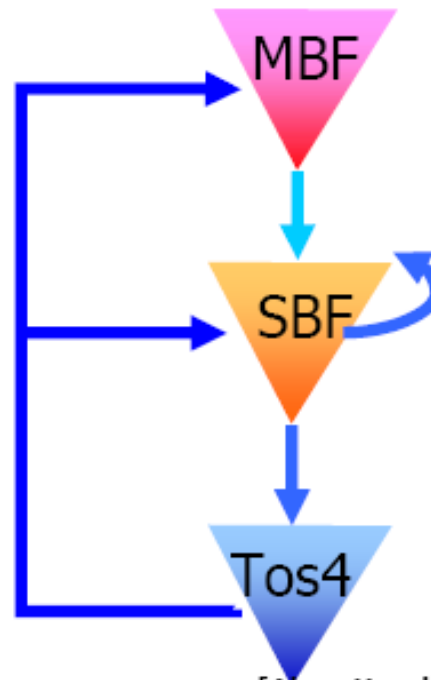
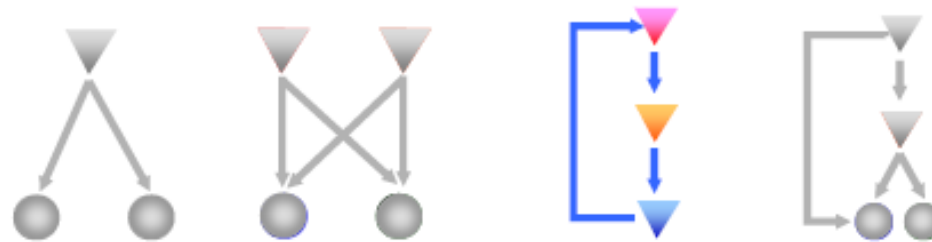
Regulatory motifs in the networks

Feed-forward loops



Regulatory motifs

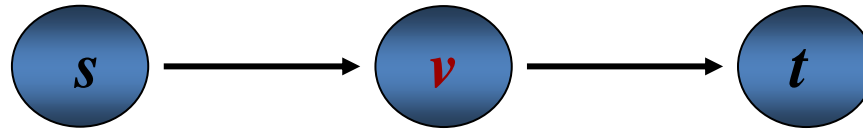
Feed-back loops



[New York: Lucchini et al (2003), *Genes &*

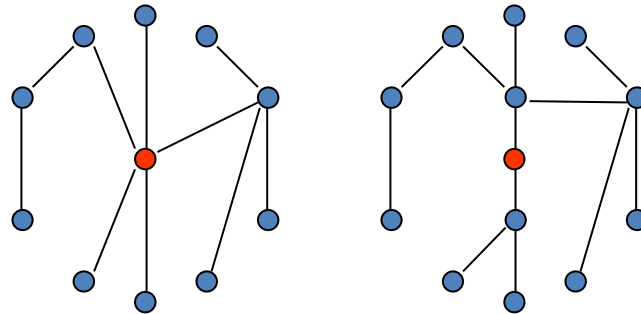
Network analysis: methodology

Betweenness centrality of a vertex v



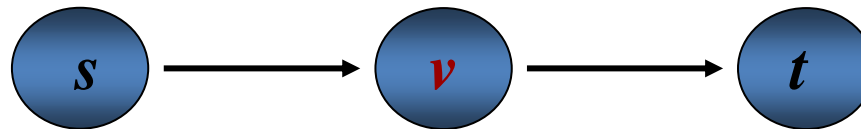
$$B_c(v) = \frac{1}{(n-1)(n-2)} \sum_{s \neq t \neq v \in V} \frac{\delta_{st}(v)}{\delta_{st}}$$

Fraction of those shortest paths between all pairs of vertices s and t that pass through vertex v . Allows to quantify how influential a given gene/protein in a whole network is.



Network analysis: methodology

Betweenness centrality of a vertex v

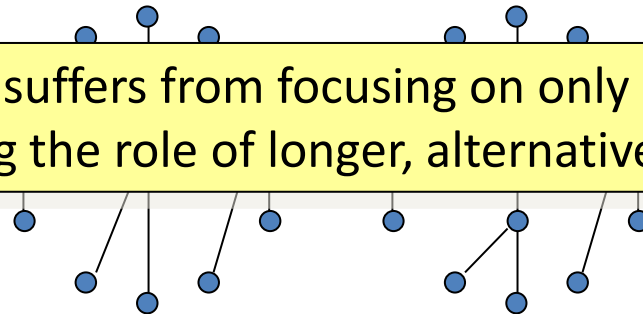


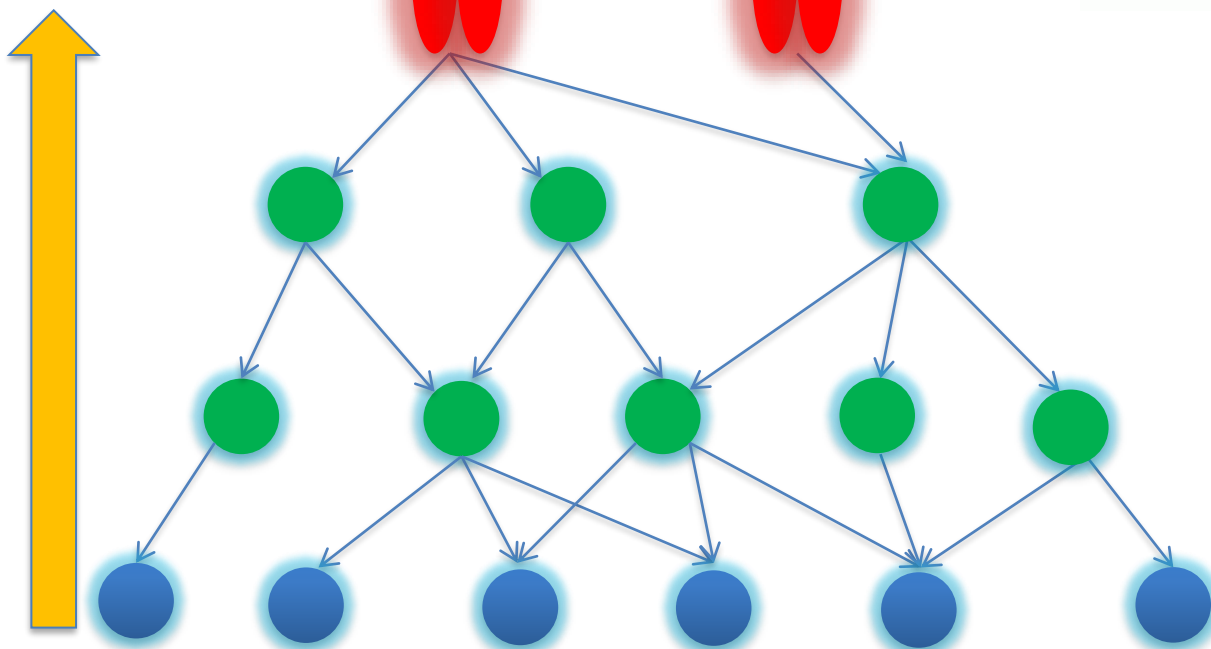
$$\frac{1}{\delta(s) \delta(t)}$$

Betweenness centrality may point to important nodes in the network, e. g. proto-oncogenes in the transcriptional network.

Fraction of those shortest paths between all pairs of vertices s and t that pass through vertex v . Allows to quantify how influential a given TF-gene in a whole network is.

However, it suffers from focusing on only shortest path and ignoring the role of longer, alternative paths.





Score:

$$S(X) = \sum_{r=1}^R \frac{M(X,r)}{M_{max}(r)} \cdot \frac{1}{1 + pN(X,r)/N_{max}(r)}$$

Where:

R - Max radius (input parameter)

p - Penalty (input parameter)

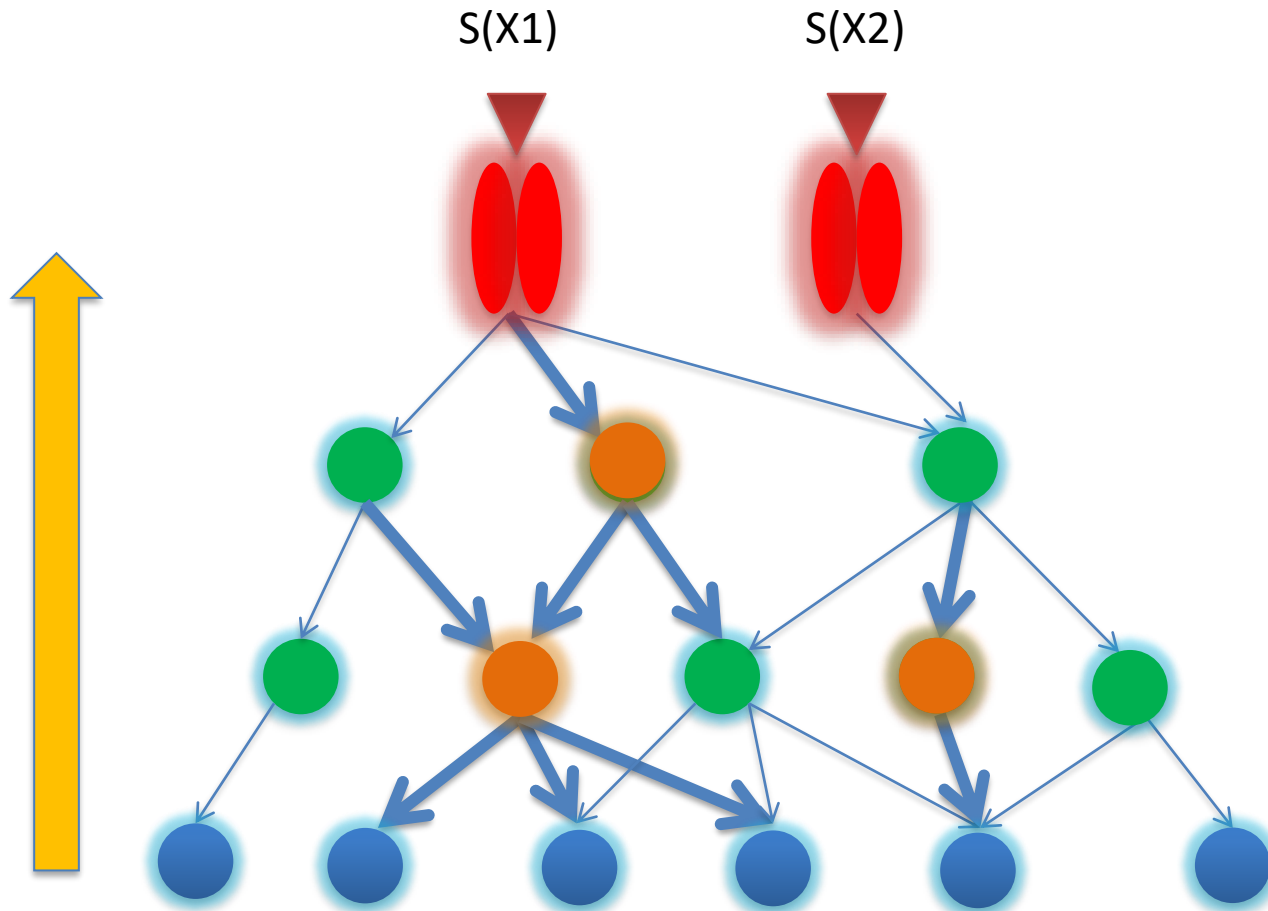
$N(X,r)$ - total number of molecules reachable from key molecule X within the radius r .

$N_{max}(r)$ - maximal value of $N(X,r)$ over all key molecules X found for this radius.

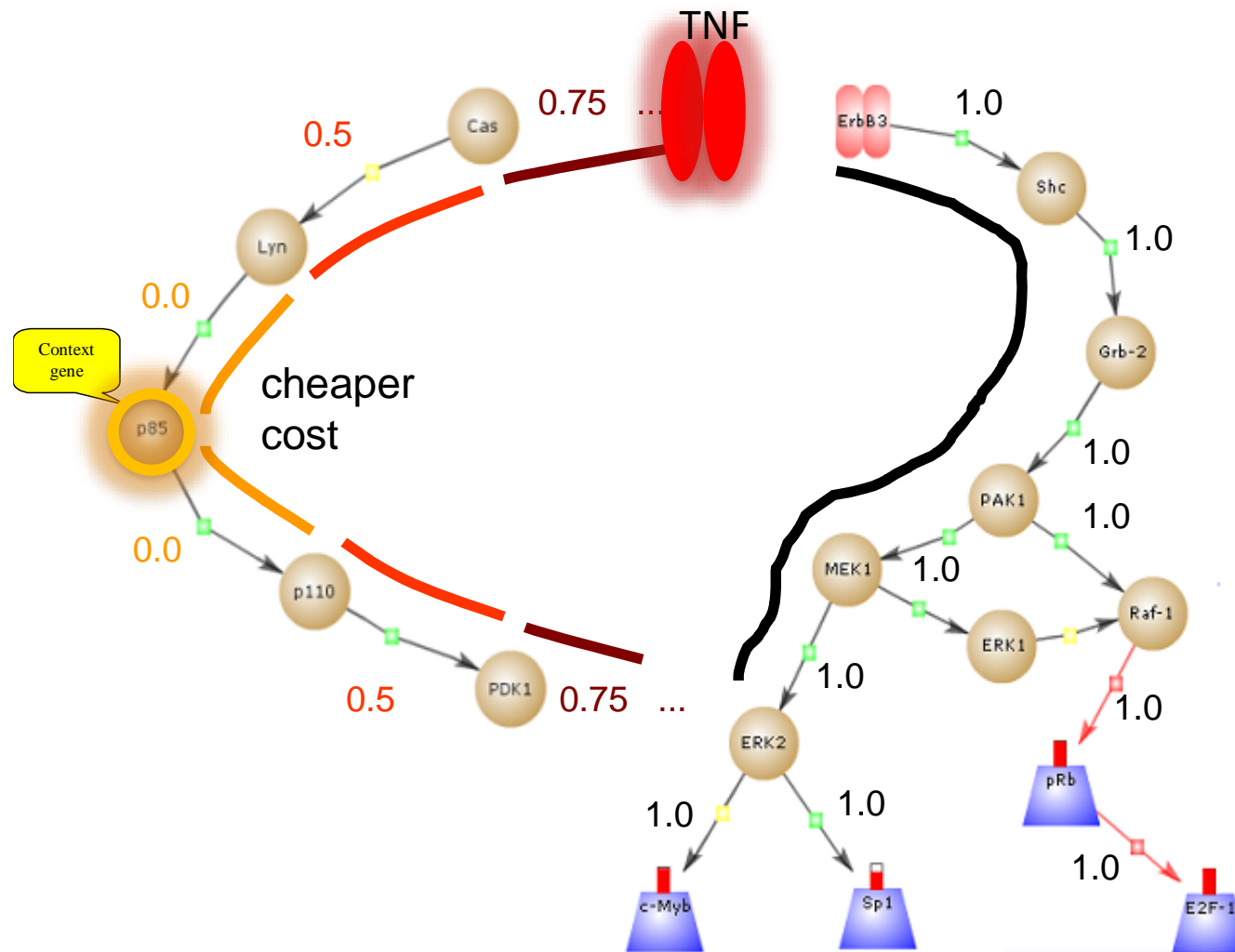
$M(X,r)$ - sum of $w(X)$ for all hits reachable from key molecule X within the radius r , where $w(X)$ - weight of hit X .

$M_{max}(r)$ - maximal value of $M(X,r)$ over all key molecules X found for this radius.

Use of context proteins (orange nodes) to calculate weights/costs of the edges.

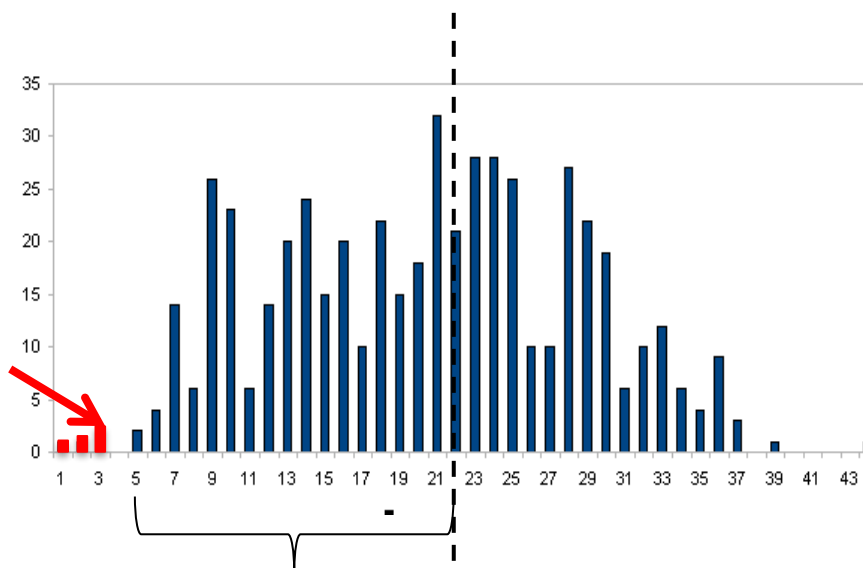


The key-node search is attracted to the paths containing context proteins





Master molecule name	Maximal radius	Reached from set	Reachable total	Score	FDR	Z-Score	Ranks sum	Hits names
betaARK-1(h)	9.925	148	23946	0.57997	0	4.76362	39	14-3-3eta(h), AKT-3-isoform1(h), AKT-3-isoform2(h), AKT-3-p51(h), Amphiphysin-BIN1(h), (more)
PDGFRbeta(h)	9.795	154	24088	0.49885	0	4.72898	45	14-3-3eta(h), AKT-3-isoform1(h), AKT-3-isoform2(h), AKT-3-p51(h), Amphiphysin-BIN1(h), (more)
RPTPalph-isoform3(h)	9.82	151	24484	0.47905	0	4.38887	57	14-3-3eta(h), AKT-3-isoform1(h), AKT-3-isoform2(h), AKT-3-p51(h), Amphiphysin-BIN1(h), (more)
DDTDalnh(h)	9.82	151	24484	0.47905	0	4.35105	61	14-3-3eta(h), AKT-3-isoform1(h), AKT-3-isoform2(h), AKT-3-p51(h), Amphiphysin-BIN1(h), (more)



Rank of the key node

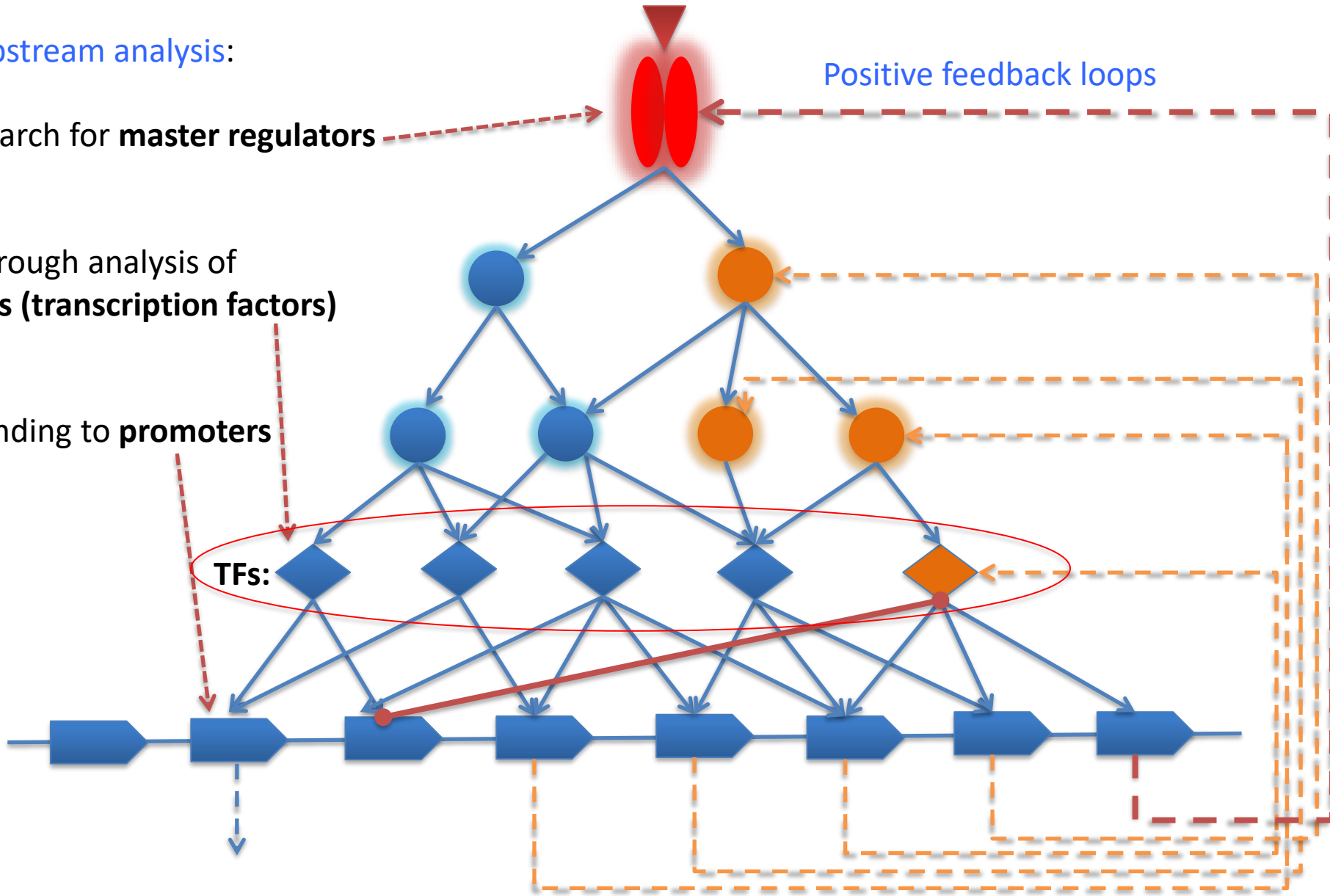
$$Z = \frac{X - \mu}{\sigma}$$

Upstream analysis:

Search for master regulators

through analysis of
TFs (transcription factors)

binding to promoters



Master regulators in networks

Firefox

geneXplain 2.4.1 web edition x Mitotic regulation of the human ana... x cdg627 6598..6609 - cdg627.pdf x O компании | DevelopmentOnTheEd... x +

platform.genexplain.com/bioulweb/#

Log out

Research: PNaS13

Databases Data Analyses Users

data

- Examples
- Projects
 - Anatolij_Potapov
 - Colon Cancer Public datasets
 - GerontoShield
 - PNaS13
 - Data
 - Brain Tumor
 - Models
 - Optimisation
 - BIOMD0000000228
 - Journal
 - tmp
 - alexander.kel2@googlemail.com
 - alexander.kel@genexplain.com
 - test
 - Public

Start page

- [Normalize data](#)
- [Detect differentially expressed genes](#)
- [Discover functional enrichment](#)
- [Identify master regulators in networks](#)
- [Analyze regulatory genome regions](#)
- [Find potential targets](#)
- [Analyze ChIP-Seq data](#)
- [Functional analysis of genomic variations](#)

[Create your own workflow](#)

[About geneXplain](#)

Search Info Default

ID: Data
Size: 4
Size on disk: 7.3Mb (7,606,235 bytes)
Complete name: data/Projects/PNaS13/Data

My description Graph search Script Clipboard Tasks

Put your comment here - press Edit button above

Master regulators in networks

Firefox

Top 3 regulators, PDGFRbeta(h) - gen... x Mitotic regulation of the human ana... x cdg627 6598..6609 - cdg627.pdf x O компании | DevelopmentOnTheEd... x +

platform.genexplain.com/bioulmweb/#de=data/Examples/Brain Tumor GSE1825%2C Affymetrix HG-U133A micro ☆ ▾ ↻ 🔍 Conduit Search

Log out Home P ?

Databases Data Analyses Users

- Experiment normalized (MAS5) vs Co
- Experiment normalized (MAS5) vs Co
- Find Common Effectors
- Find master regulator_Geneways
- Find Master regulator_Transpath
- Downregulated_Entrez
 - Regulator genes
 - Regulator genes annot
 - Regulator proteins
 - Regulators upstream 10
 - Top 3 regulators, betaARK-1
 - Top 3 regulators, PDGFRb
 - Top 3 regulators, RPTPalph
 - Transpath peptides
- Upregulated_Entrez
- Mapping to ontologies_Proteome
- Master Regulator_Cyclic
- Upstream analysis
- Control normalized (MAS5)
- Experiment normalized (MAS5)
- Experiment normalized (MAS5) annot

Start page Top 3 regulators, PDGFRb... X

Users
alexander.kel2@googlemail.com

Chat

Search Info Default

« Overview Layout Expression mapping My description »

Layouter: Hierarchic layout

- Vertical orientation ☐
- Hoist nodes ☐

ID: Top 3 regulators, PDGFRbeta(h)
Created: Tue Jan 22 11:01:38 UTC 2013
Modified: Mon Jul 15 04:26:45 UTC 2013
Title: Top 3 regulators, PDGFRbeta(h)
Size: 660

Master regulators in networks

Firefox

Top 3 regulators, PDGFRbeta(h) - gen... x Mitotic regulation of the human ana... x cdg627 6598..6609 - cdg627.pdf x O компании | DevelopmentOnTheEd... x +

platform.genexplain.com/bioulweb/#de=data/Examples/Brain Tumor GSE1825%2C Affymetrix HG-U133A micro ☆ ▾ ↻ 🔍 Conduit Search

Log out Home P ?

Databases Data Analyses Users

Experiment normalized (MAS5) vs Co
Experiment normalized (MAS5) vs Co
Find Common Effectors
Find master regulator_Geneways
Find Master regulator_Transpath
Downregulated_Entrez
Regulator genes
Regulator genes annot
Regulator proteins
Regulators upstream 10
Top 3 regulators, betaARK-1
Top 3 regulators, PDGFRb
Top 3 regulators, RPTPalph-
Transpath peptides
Upregulated_Entrez
Mapping to ontologies_Proteome
Master Regulator_Cyclic
Upstream analysis
Control normalized (MAS5)
Experiment normalized (MAS5)

Start page Top 3 regulators, PDGFRb... X

Users
alexander.kel2@googlemail.com
Chat

Search Info Default

ID: MO000104360
Title: PDGFRbeta
Size: 0
Role: VariableRole: \$MO000104360=0.0 null;
de=Compartment:PDGFRbeta
Attributes:

Overview Layout Expression mapping My description

GridX	5
GridY	5
Delta X	75
Delta Y	75

Master regulators in networks

The screenshot shows the GeneXplain web interface. On the left is a navigation tree with categories like 'Databases', 'Data', 'Analyses', and 'Users'. The main area displays a table titled 'Regulators upstream 10'. The table contains 10 rows of data. The first row is highlighted in blue.

ID	Master molecule name	Maximal radius	Reached from set	Reachable total	Score	FDR	Z-Score	Ranks sum
MO000078267	betaARK-1(h)	9.925	148	23946	0.57997	0	4.76362	39
MO000104360	PDGFRbeta(h)	9.795	154	24088	0.49885	0	4.72898	45
MO000080313	RPTPalphaisoform3(h)	9.82	151	24484	0.47905	0	4.38887	57
MO000219258	RPTPalph(h)	9.82	151	24484	0.47905	0	4.35195	61
MO000080314	RPTPalphaxbb2(h)	9.82	151	24484	0.47905	0	4.31239	63
MO000219259	RPTPalphaisoform2(h)	9.82	151	24484	0.47905	0	4.23961	67
MO000103238	tec(h)	9.885	144	23238	0.38849	0.009	4.61671	81
MO000018780	calmodulin(h)	9.945	108	14188	0.35495	0	6.1317	87
MO000016694	NRG2(h)	9.99	134	20862	0.33859	0	9.91269	90
MO000036416	Cdk5(h):CDK5R1p35(h)	9.93	149	20593	0.35421	0	5.72554	92
MO000256934	prlrisoform8(h)	9.885	154	23758	0.36884	0.003	4.57343	93
MO000081201	14-3-3beta(h)	9.76	130	22654	0.43636	0	3.79013	94
MO000256928	prlrisoform2(h)	9.885	154	23758	0.36884	0.004	4.47373	94
MO000256933	prlrisoform7(h)	9.885	154	23758	0.36884	0.003	4.49017	95
MO000105205	Tiam-1(h)	9.99	130	18310	0.32568	0	7.63778	99
MO000213896	LAMA5(h)	10	123	17183	0.29847	0	9.21436	101
MO000256929	prlrisoform3(h)	9.885	154	23758	0.36884	0.003	4.41942	102
MO000256930	prlrisoform4(h)	9.885	154	23758	0.36884	0.005	4.38913	102
MO000256932	prlrisoform6(h)	9.885	154	23758	0.36884	0.006	4.42015	102