

Short workshop on Co-Expression Network Analysis

Katja Nowick
Freie Universität Berlin
Katja.nowick@fu-berlin.de



Goals for today

- Analysis of co-expression networks, more specifically weighted topological overlap (wTO) networks
- Comparison of networks (CoDiNA)
- Using R

Networks in Biology

Biological systems are complex → can well be understood as networks

Networks in cells (molecular networks):

- Metabolic Networks
- Gene regulatory networks
- Protein-Protein-Interaction networks

Networks between cells:

- Neural networks
- Immune system

Networks in ecosystems:

- Food networks
- Cooperation/Symbiosis

Social networks:

- Friendships
- Epidemiology

Networks in Biology

Biological systems are complex → can well be understood as networks

Networks in cells (molecular networks):

- Metabolic Networks
- **Gene regulatory networks**
- Protein-Protein-Interaction networks

Networks between cells:

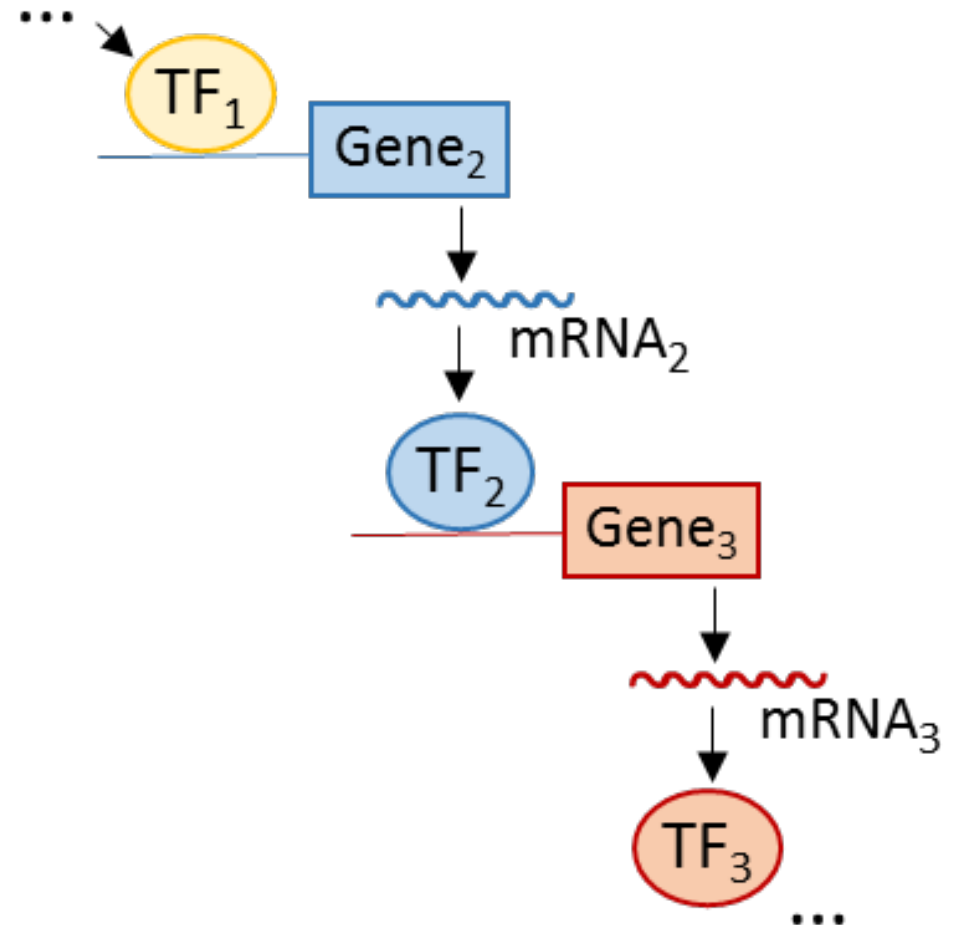
- Neural networks
- Immune system

Networks in ecosystems:

- Food networks
- Cooperation/Symbiosis

Social networks:

- Friendships
- Epidemiology



Networks in Biology

Biological systems are complex → can well be understood as networks

Networks in cells (molecular networks):

- Metabolic Networks
- Gene regulatory networks
- **Protein-Protein-Interaction networks**

Networks between cells:

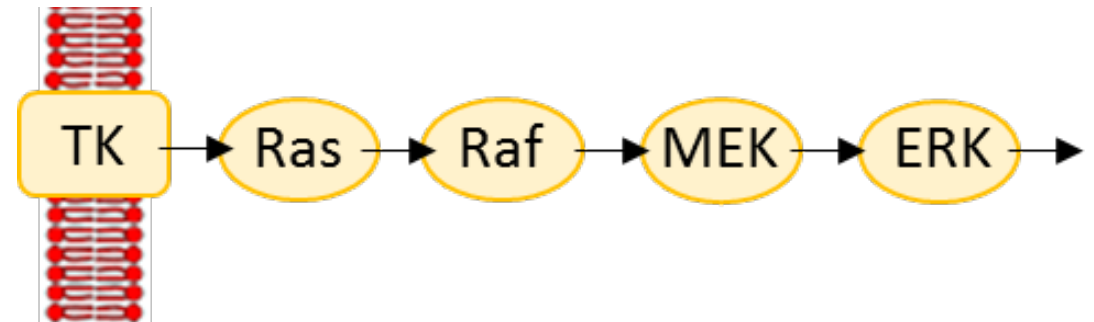
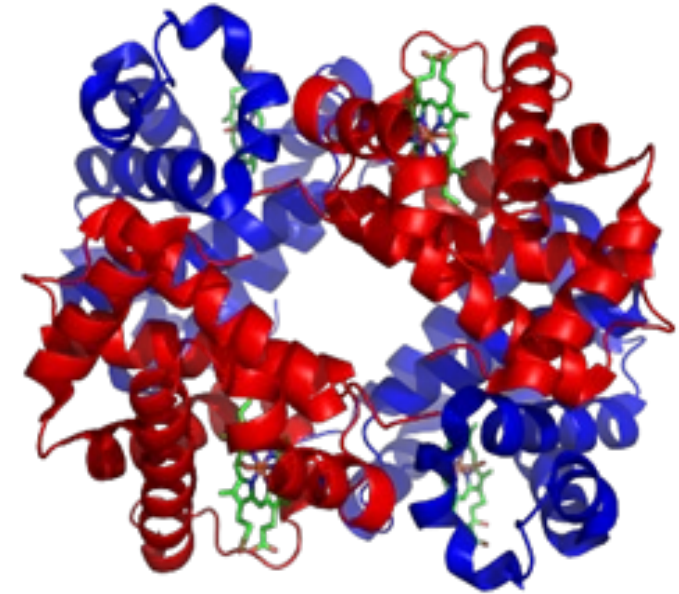
- Neural networks
- Immune system

Networks in ecosystems:

- Food networks
- Cooperation/Symbiosis

Social networks:

- Friendships
- Epidemiology



Networks in Biology

Biological systems are complex → can well be understood as networks

Networks in cells (molecular networks):

- Metabolic Networks
- Gene regulatory networks
- Protein-Protein-Interaction networks

Networks between cells:

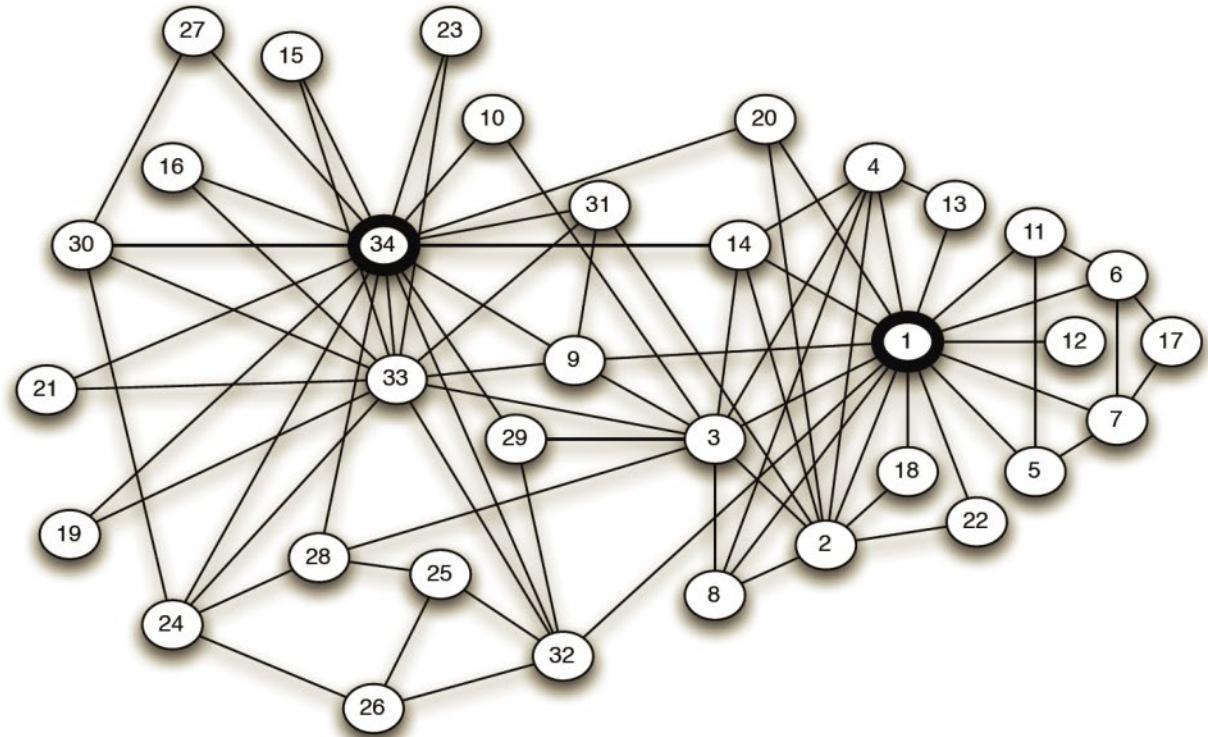
- Neural networks
- Immune system

Networks in ecosystems:

- Food networks
- Cooperation/Symbiosis

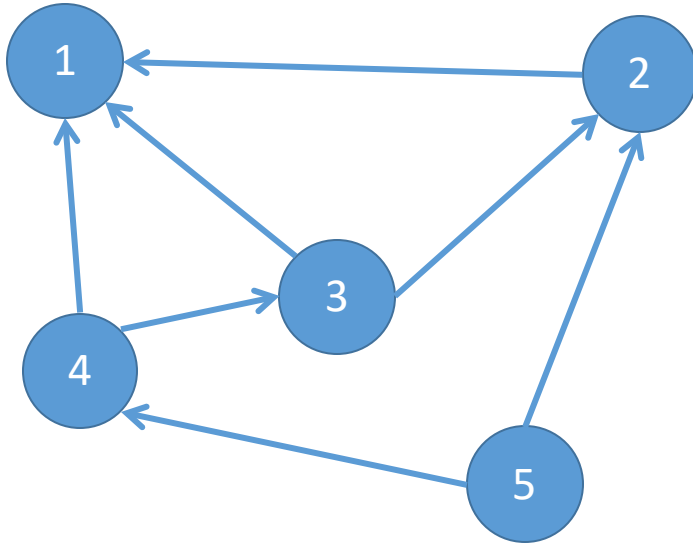
Social networks:

- **Friendships**
- Epidemiology

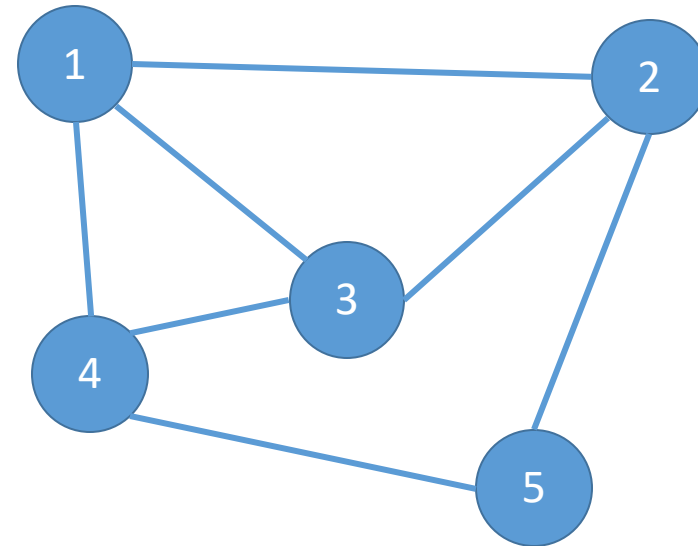


Networks in Biology

- Identity of nodes and links depends on the investigated network
- Links can be directed or undirected



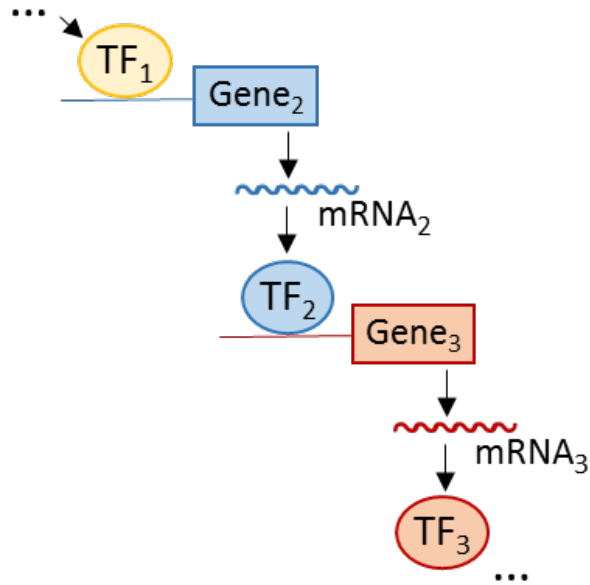
Directed network



Undirected network

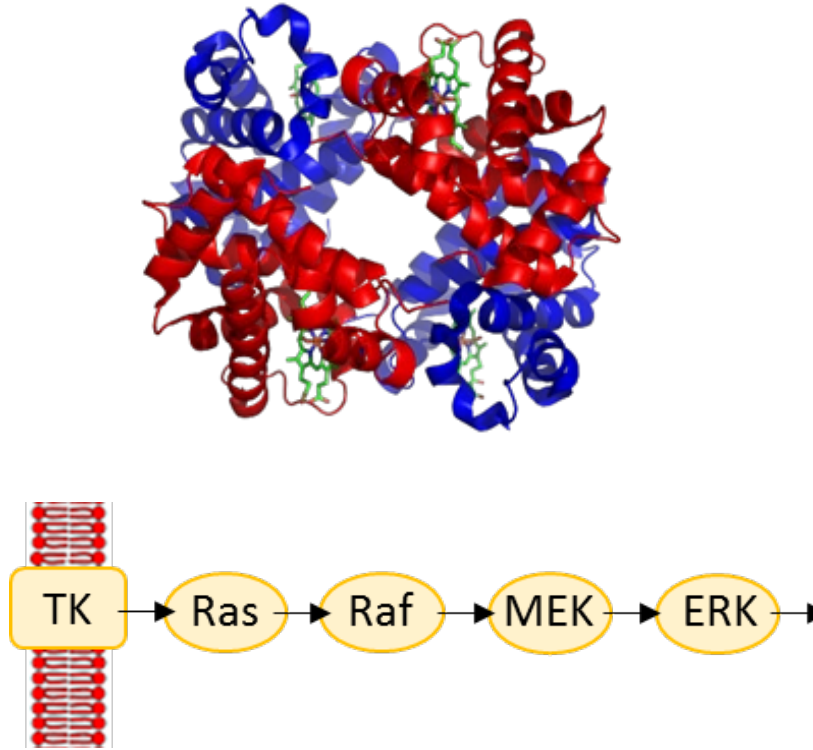
Networks in Biology

Gene regulatory networks



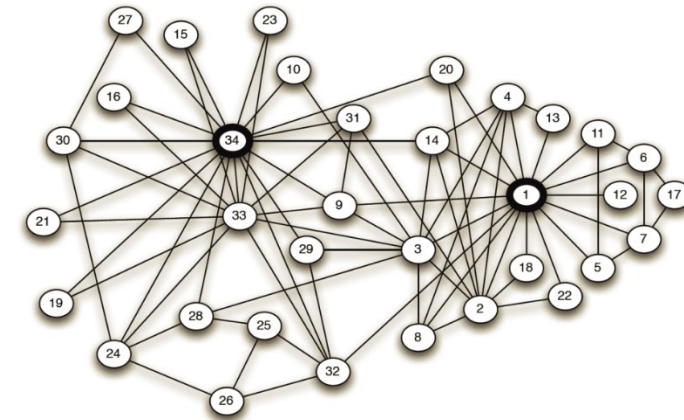
Nodes: Gene und gene regulatory factors
Links: Activation or repression
directed

Protein-Protein-Interaction networks



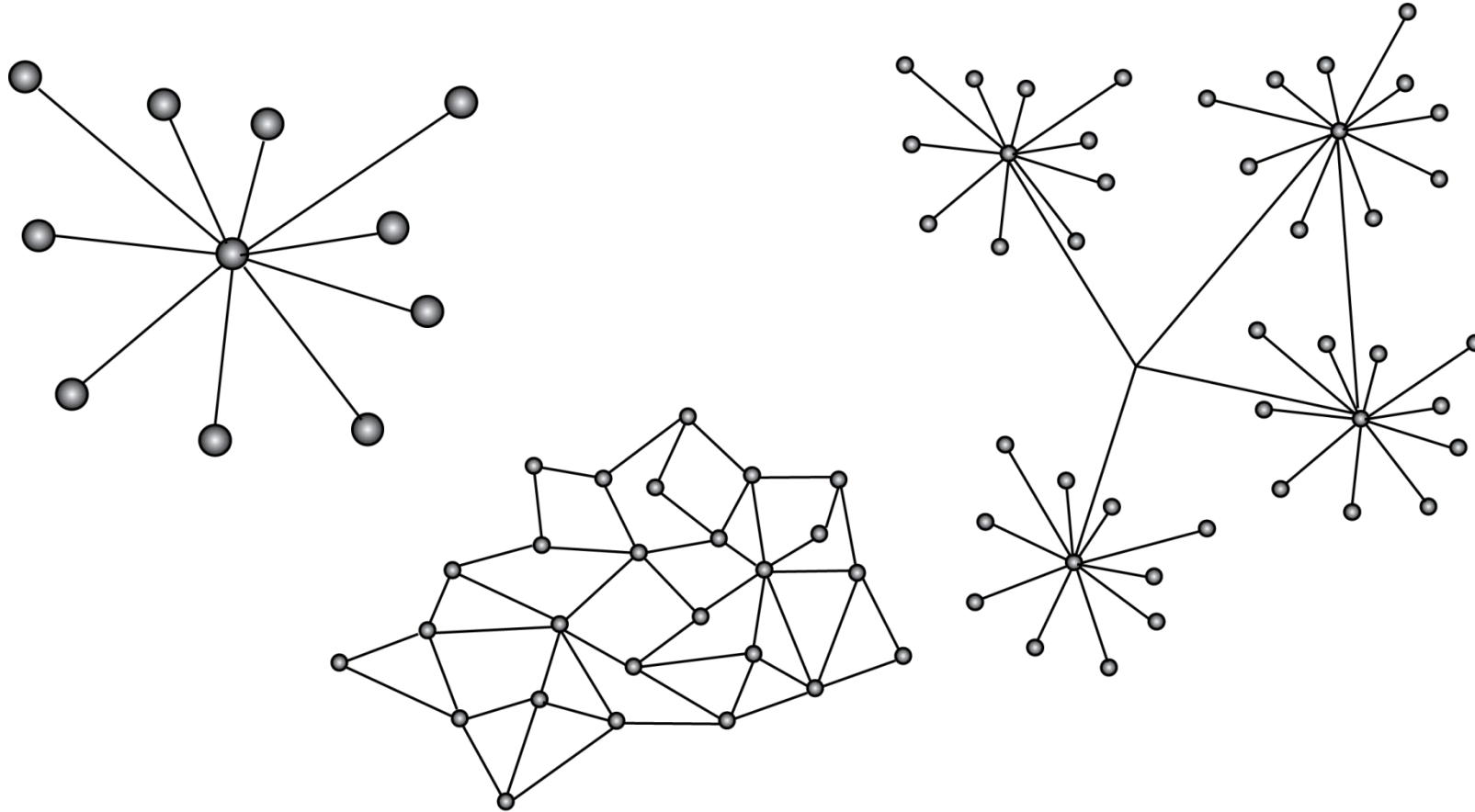
Nodes: Proteins
Links: Binding; activation or inactivation
undirected; directed

Social networks



Nodes: Individuals
Links: e.g. friendships
usually undirected

Which nodes are the most important nodes?

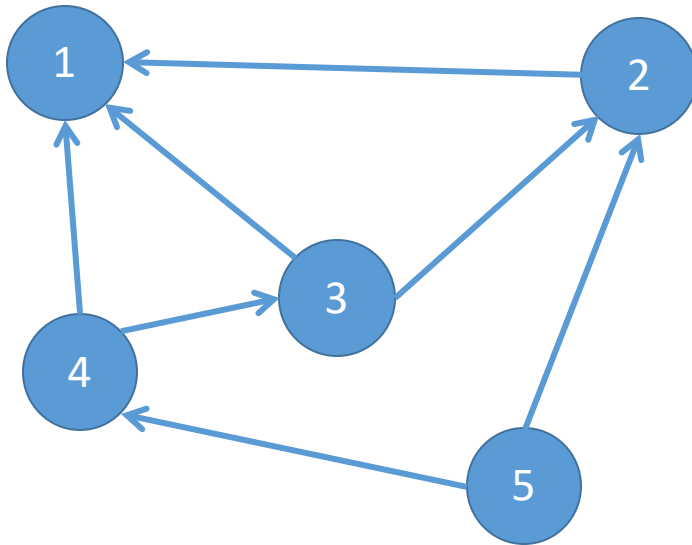


Which nodes are the most important nodes?

- Degree-

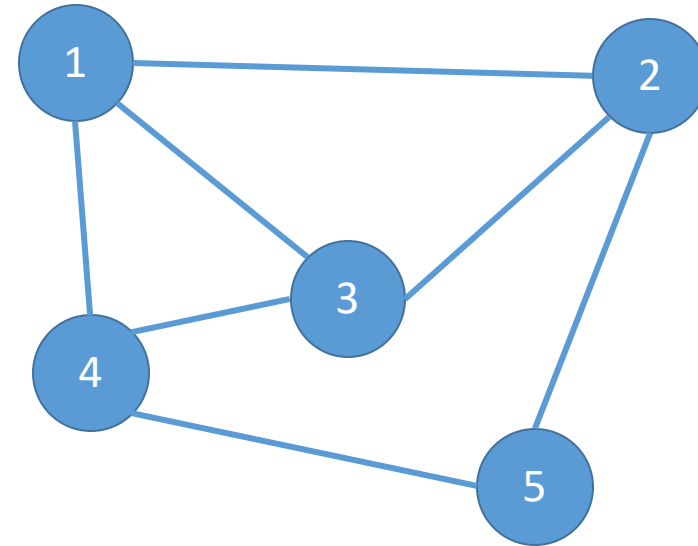
Degree = number of links a node has

Hub = node with high degree



Directed network

In-degree of a node = its number of incoming links
Out-degree of a node = its number of outgoing links

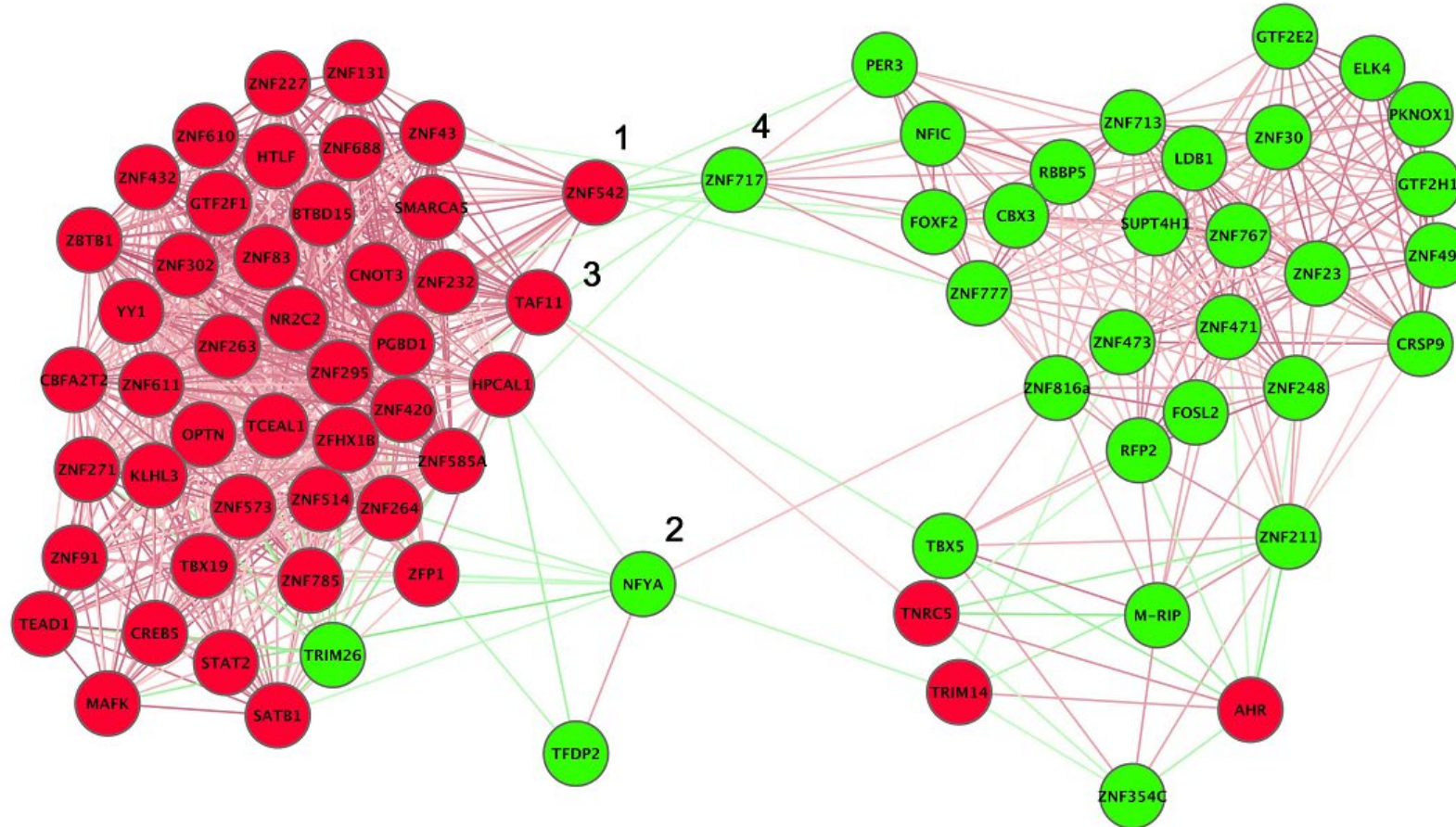


Undirected network

Degree of a node = its number of links

Which nodes are the most important nodes?
- Betweenness Centrality-

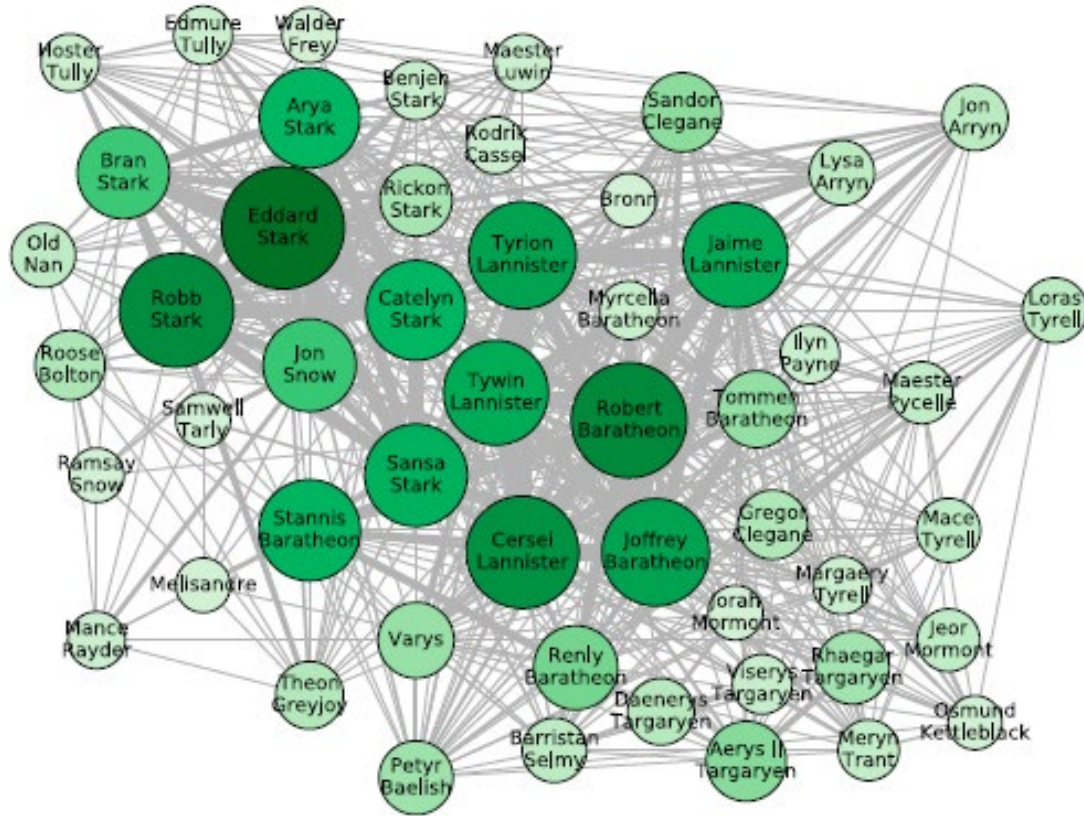
Betweenness Centrality (BC) = number of shortest paths going through that node
“Gatekeeper”



Labeled are the four nodes with highest BC
They tend to **connect the two modules/communities**

Which nodes are the most important nodes?

Interaction networks: Game of Thrones



Node size proportional to the number of chapters in which the characters interact
Link thicknesses represent the numbers of times that corresponding pair of characters interact

Who is important?

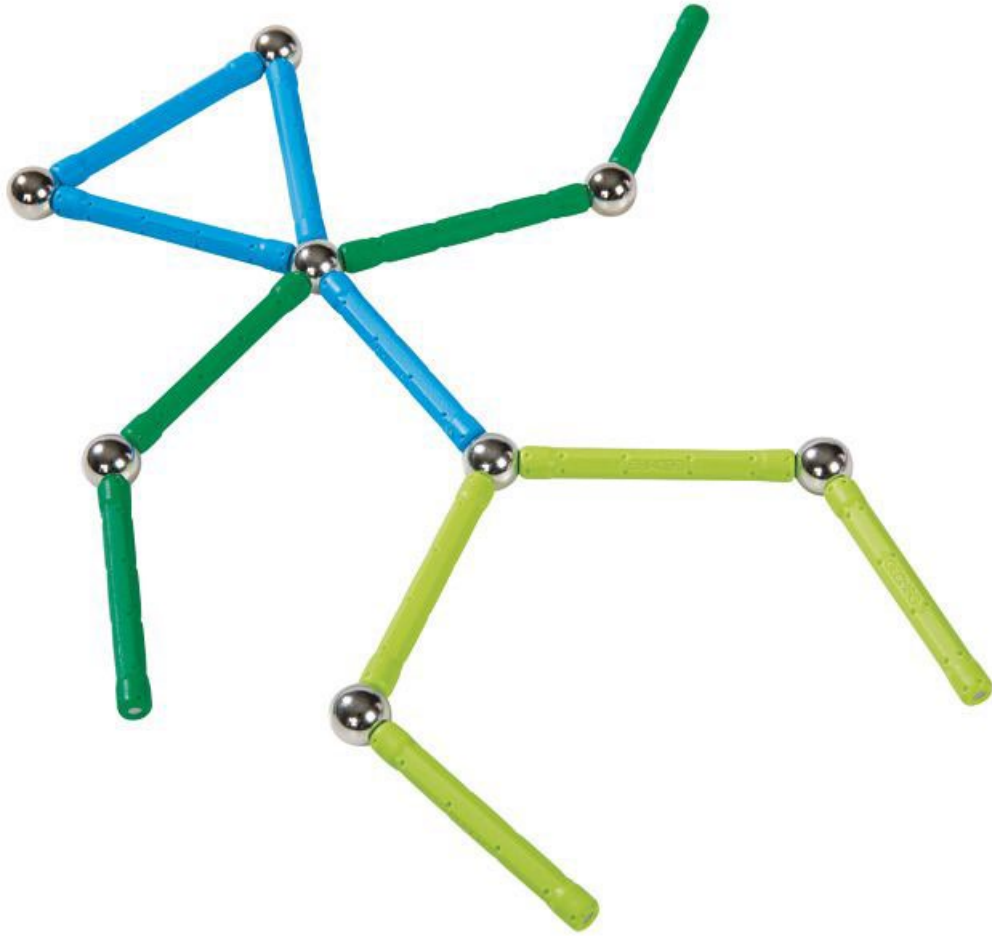
Table 1. Characters ranked by various network attributes

Degree	Betweenness centrality
Full network	
1. Jon Snow (214)	1. Jon Snow (0.0889)
2. Jaime Lannister (212)	2. Barristan Selmy (0.0831)
3. Tyrion Lannister (209)	3. Arya Stark (0.0777)
4. Catelyn Stark (204)	4. Tyrion Lannister (0.0700)
5. Arya Stark (192)	5. Theon Greyjoy (0.0671)
6. Theon Greyjoy (175)	6. Jaime Lannister (0.0606)
7. Cersei Lannister (161)	7. Catelyn Stark (0.0568)
8. Robb Stark (158)	8. Stannis Baratheon (0.0519)
9. Sansa Stark (156)	9. Tywin Lannister (0.0356)
10. Barristan Selmy (156)	10. Eddard Stark (0.0351)
12. Eddard Stark (140)	12. Sansa Stark (0.0275)
16. Brienne of Tarth (108)	13. Cersei Lannister (0.0250)
17. Bran Stark (106)	14. Brienne of Tarth (0.0236)
19. Daenerys Targaryen (104)	17. Samwell Tarly (0.0207)
20. Samwell Tarly (103)	18. Bran Stark (0.0202)
51. Davos Seaworth (72)	21. Daenerys Targaryen (0.0185)
	25. Davos Seaworth (0.0167)

Degree = number of links
→ Numbers of
interaction partners

BC = number of shortest paths
going through that node
“Gatekeeper”

Build a network



Which nodes are most important?

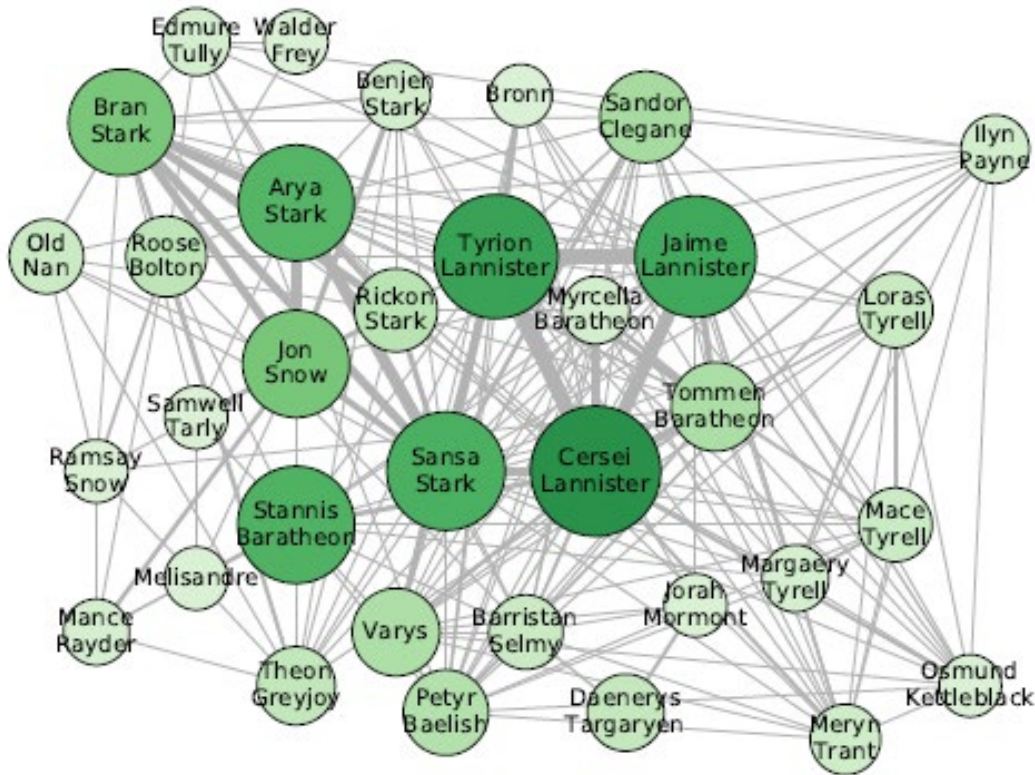
- Remove* one of the least important nodes
- Remove* one of the most important nodes

How big is the impact on structure and function?

* This could happen due to a mutation in a gene

Networks change over time

Interaction networks: Game of Thrones

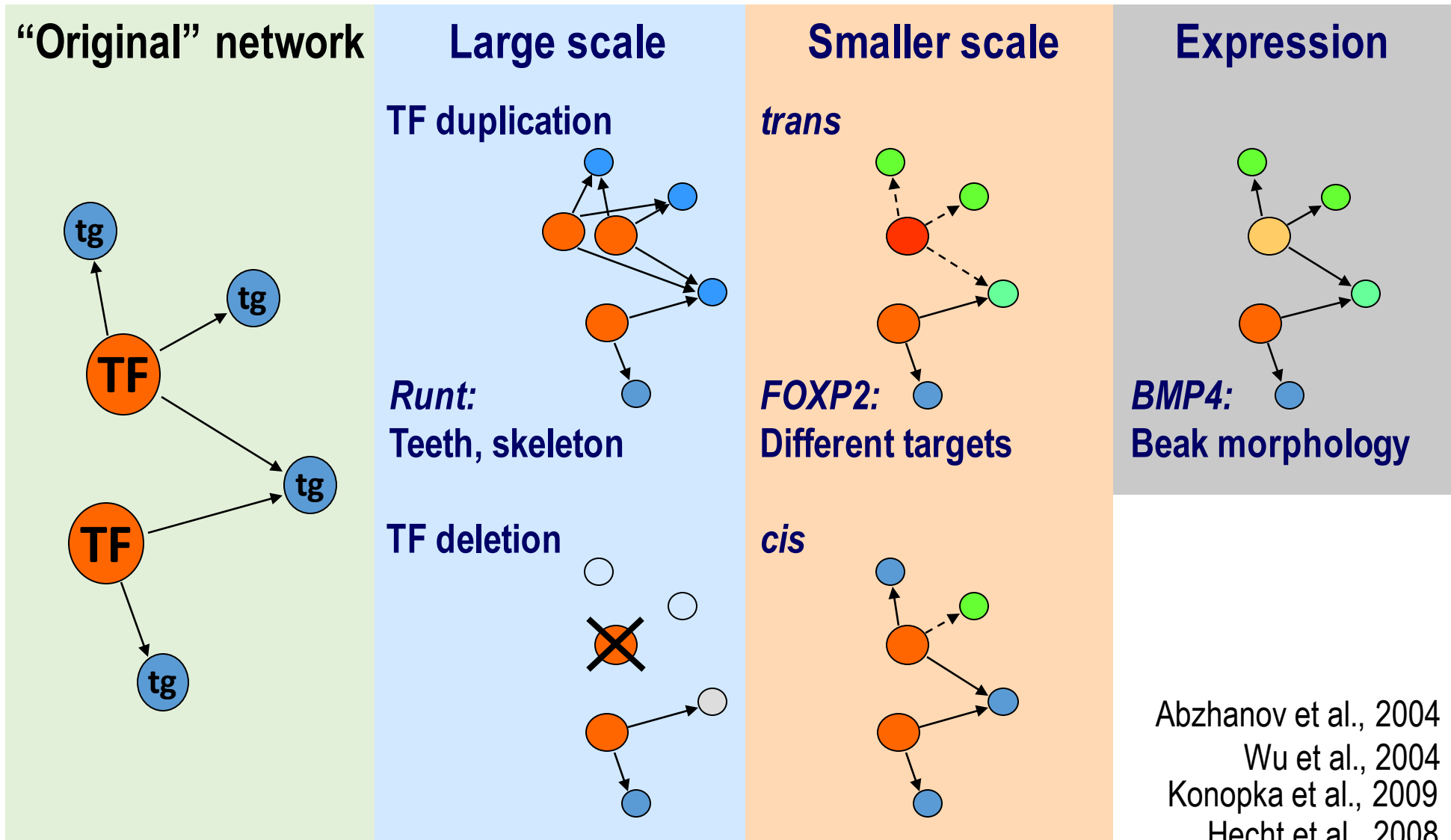


Shown is the survivor network, i.e., the characters still alive at the end of the fifth book

Table 1. Characters ranked by various network attributes

Degree	Betweenness centrality
Full network	
1. Jon Snow (214)	1. Jon Snow (0.0889)
2. Jaime Lannister (212)	2. Barristan Selmy (0.0831)
3. Tyrion Lannister (209)	3. Arya Stark (0.0777)
4. Catelyn Stark (204)	4. Tyrion Lannister (0.0700)
5. Arya Stark (192)	5. Theon Greyjoy (0.0671)
6. Theon Greyjoy (175)	6. Jaime Lannister (0.0606)
7. Cersei Lannister (161)	7. Catelyn Stark (0.0568)
8. Robb Stark (158)	8. Stannis Baratheon (0.0519)
9. Sansa Stark (156)	9. Tywin Lannister (0.0356)
10. Barristan Selmy (156)	10. Eddard Stark (0.0351)
12. Eddard Stark (140)	12. Sansa Stark (0.0275)
16. Brienne of Tarth (108)	13. Cersei Lannister (0.0250)
17. Bran Stark (106)	14. Brienne of Tarth (0.0236)
19. Daenerys Targaryen (104)	17. Samwell Tarly (0.0207)
20. Samwell Tarly (103)	18. Bran Stark (0.0202)
51. Davos Seaworth (72)	21. Daenerys Targaryen (0.0185)
	25. Davos Seaworth (0.0167)
Survivor network	
1. Tyrion Lannister (162)	1. Tyrion Lannister (0.0972)
2. Jon Snow (150)	2. Barristan Selmy (0.0952)
3. Jaime Lannister (149)	3. Arya Stark (0.0923)
4. Arya Stark (135)	4. Theon Greyjoy (0.0909)
5. Sansa Stark (122)	5. Jon Snow (0.0871)
6. Cersei Lannister (120)	6. Stannis Baratheon (0.0812)
7. Theon Greyjoy (115)	7. Jaime Lannister (0.0805)
8. Barristan Selmy (103)	8. Sansa Stark (0.0408)
9. Stannis Baratheon (86)	9. Samwell Tarly (0.0320)
10. Brienne of Tarth (83)	10. Cersei Lannister (0.0310)
12. Samwell Tarly (79)	12. Brienne of Tarth (0.0274)
18. Daenerys Targaryen (69)	13. Bran Stark (0.0248)
20. Bran Stark (68)	17. Davos Seaworth (0.0184)
38. Davos Seaworth (54)	33. Daenerys Targaryen (0.0093)

How do networks evolve?

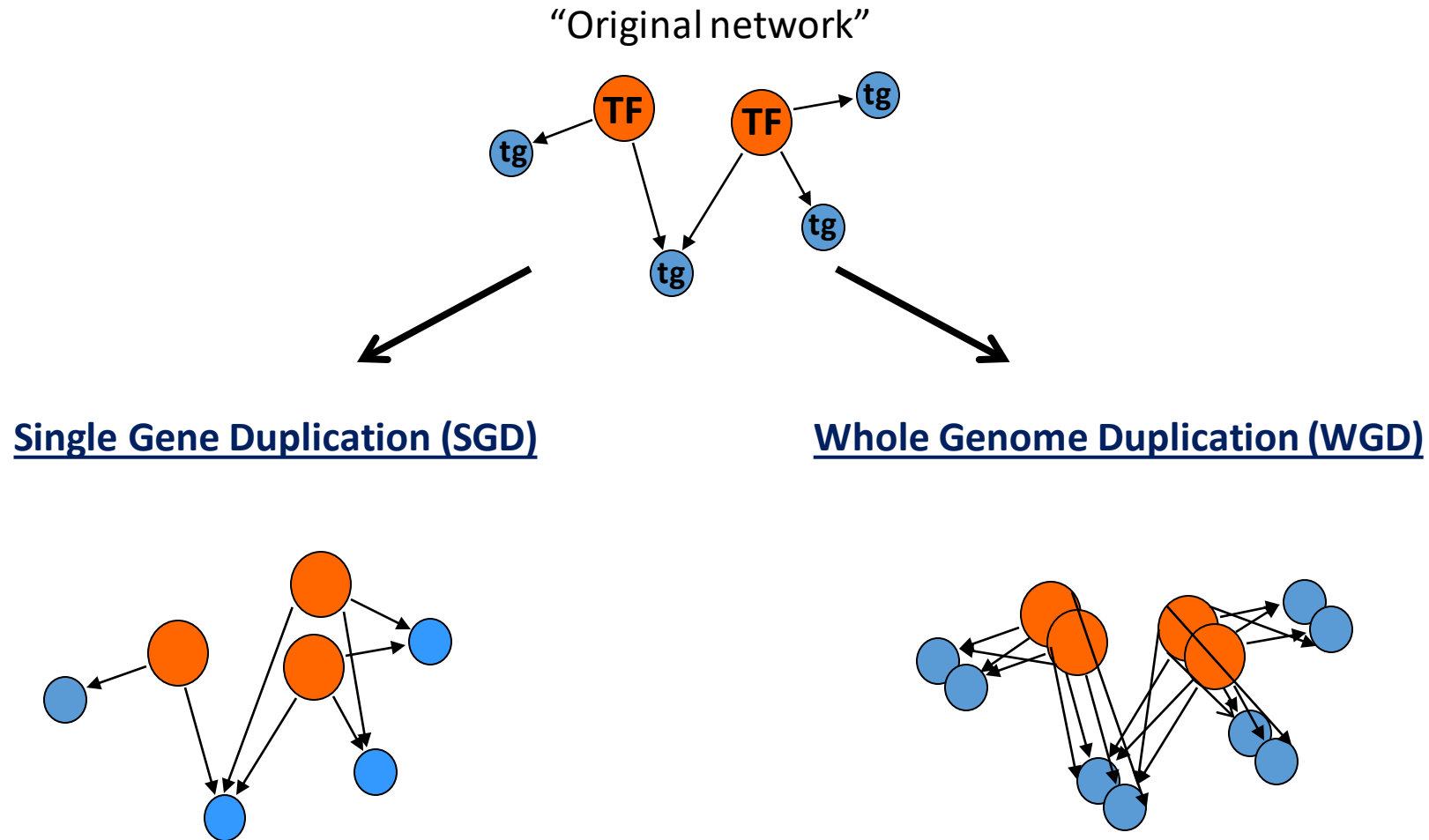


Abzhanov et al., 2004
Wu et al., 2004
Konopka et al., 2009
Hecht et al., 2008

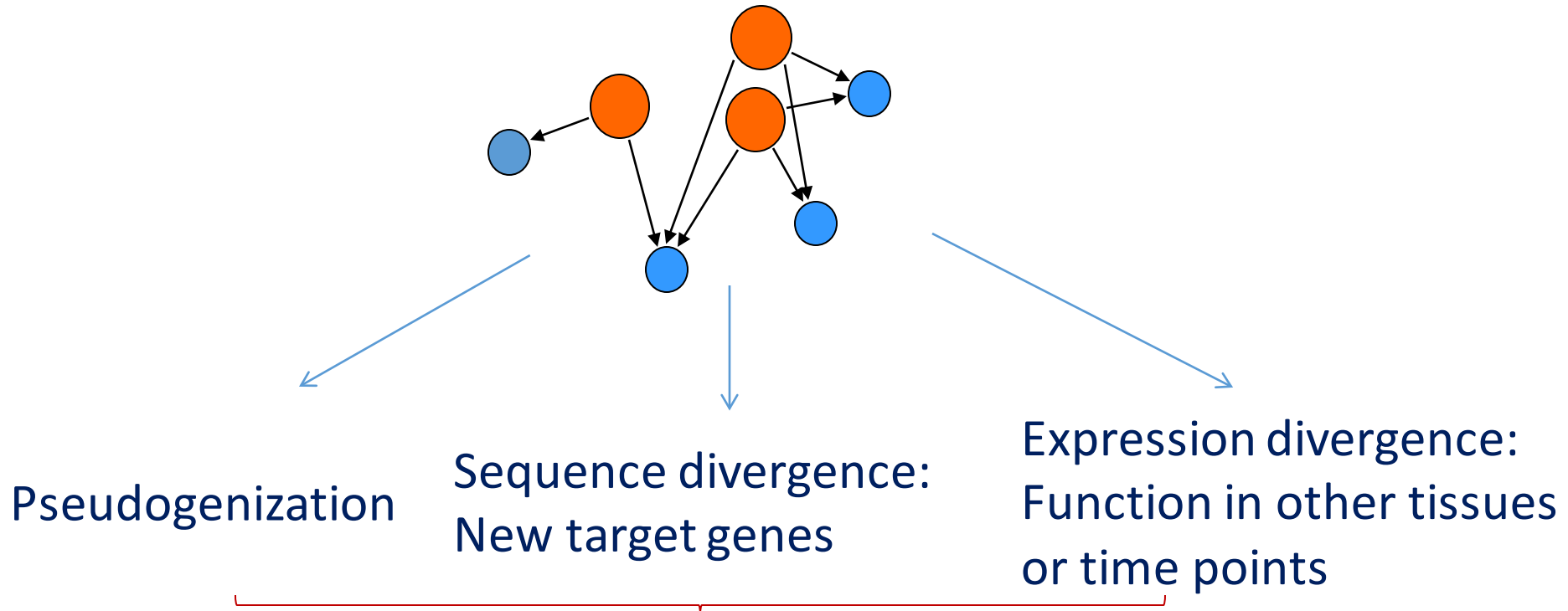
→TFs have key role in evolution and species differences

Fate of duplicates

Genomes grow by duplication → Addition of nodes to the molecular networks



Divergence of Single Gene Duplications



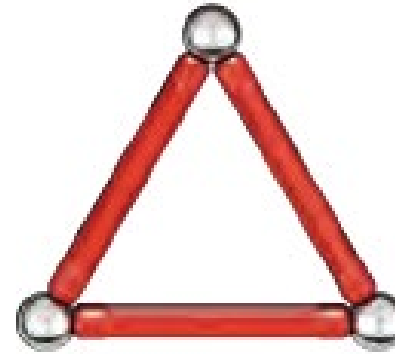
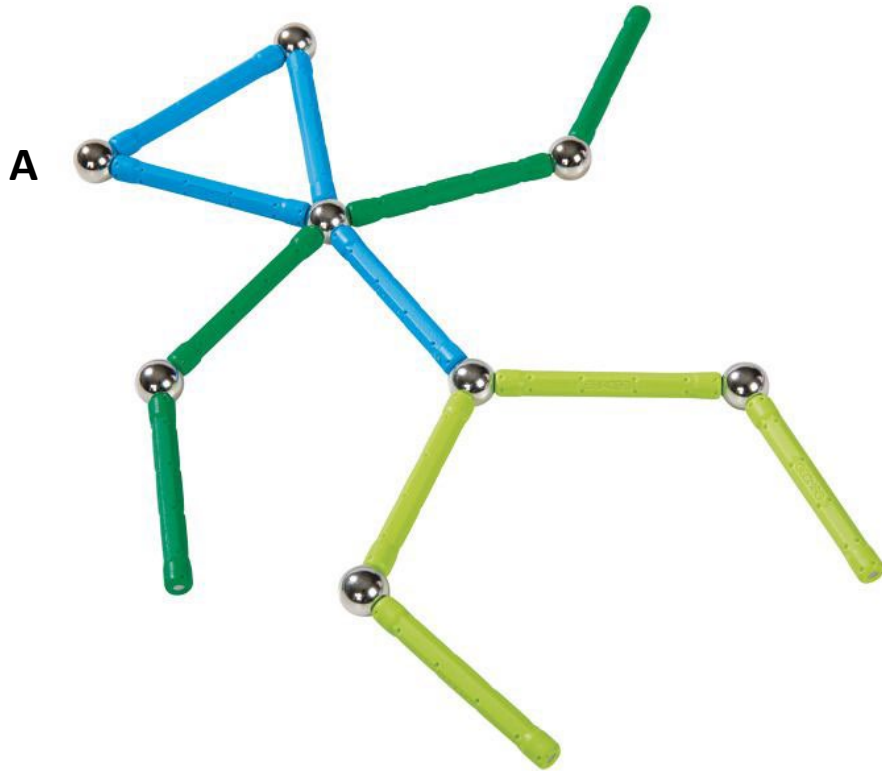
Very rapid, within a few million years!

One copy diverges: neo-functionalization

Both copies diverge: sub-functionalization

Often **asymmetric divergence**: one duplicate retains much more links than the other (although both will have more links in common than expected by chance)

Evolve your network



Duplicate node A

What does the network look like after

- Neo-functionalization?
- Sub-functionalization?

Create either the triangle or square

What does the network look like

- Immediately after whole genome duplication?
- Some million years later?

Co-expression Networks

Can be built from gene expression data

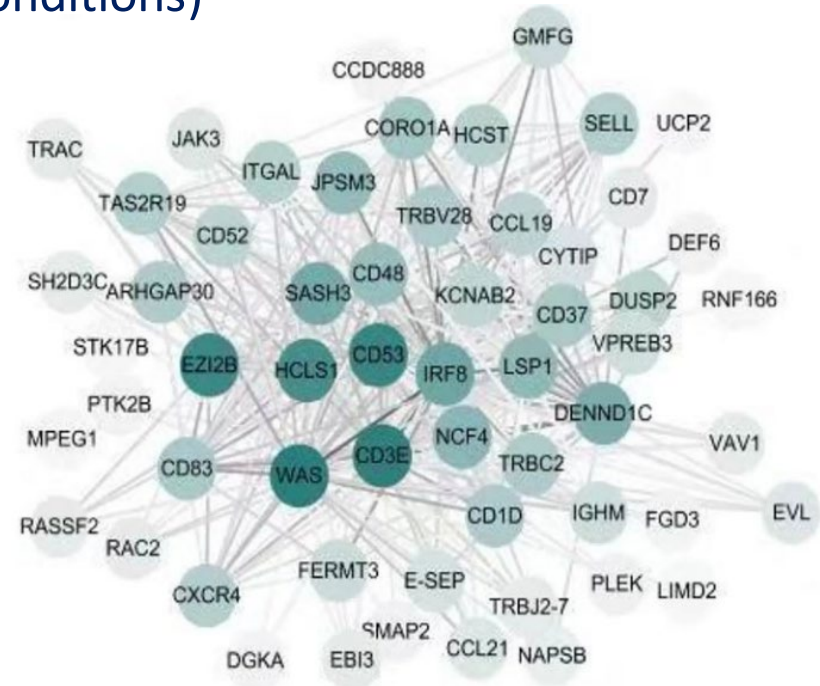
Are based on gene expression correlation (→ undirected, no causation!)

Can represent regulatory relationships or protein-protein interactions, and indirect relationships

Transcription factors (TFs) are often hubs)

Needs multiple related samples (e.g. individuals, experimental conditions)

with gene expression profiles (e.g. RNA-Seq data)

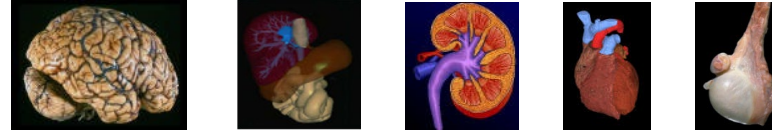


Nodes: Genes

Links: positive or negative correlation of expression values
undirected

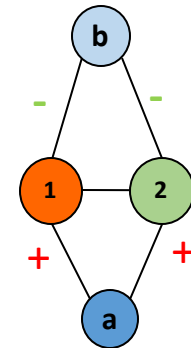
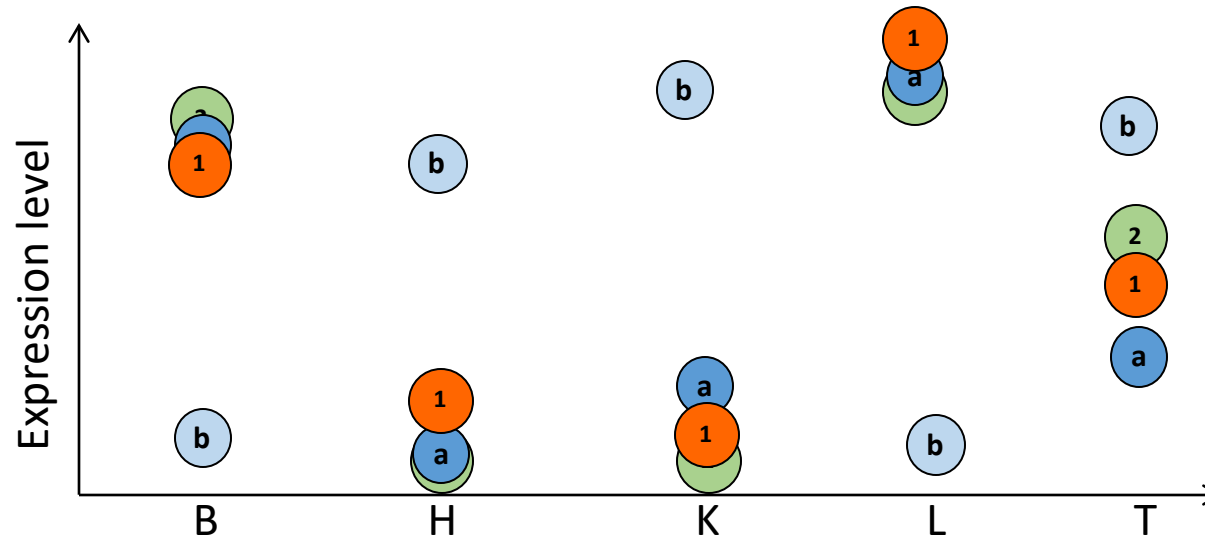
Correlation of gene expression patterns

Expression profiles



Co-expression patterns

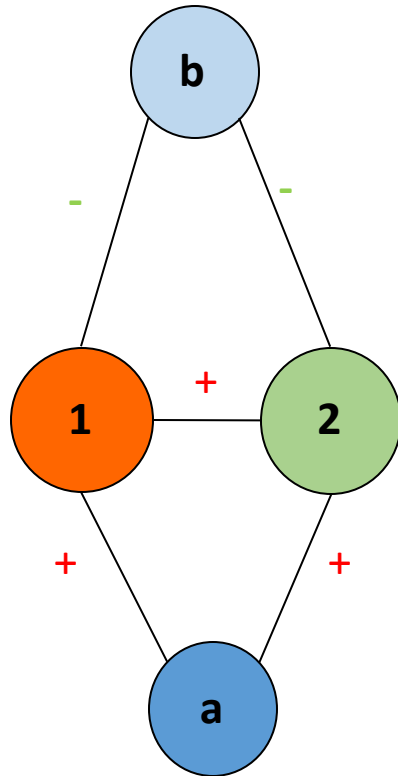
Significant Spearman correlations



→ Co-expression network

“Guilt by association”

Co-expression networks

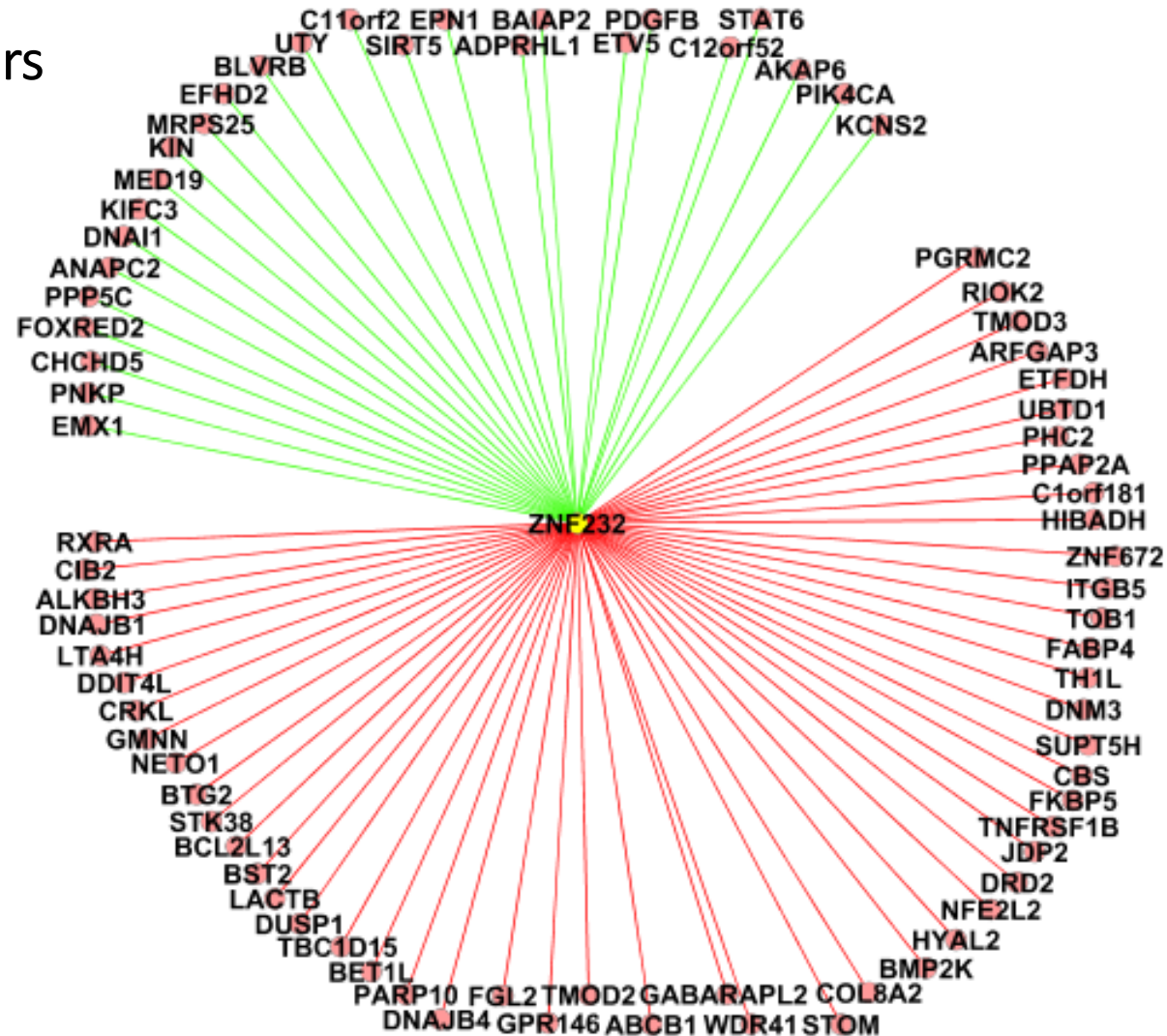


Nodes = genes/proteins
TFs and targets

Links = relationship between the genes
correlated in expression
positive or negative
undirected!

Co-expressed genes of a TF

- TFs are often hubs
- Potential target genes
- Potential interaction partners

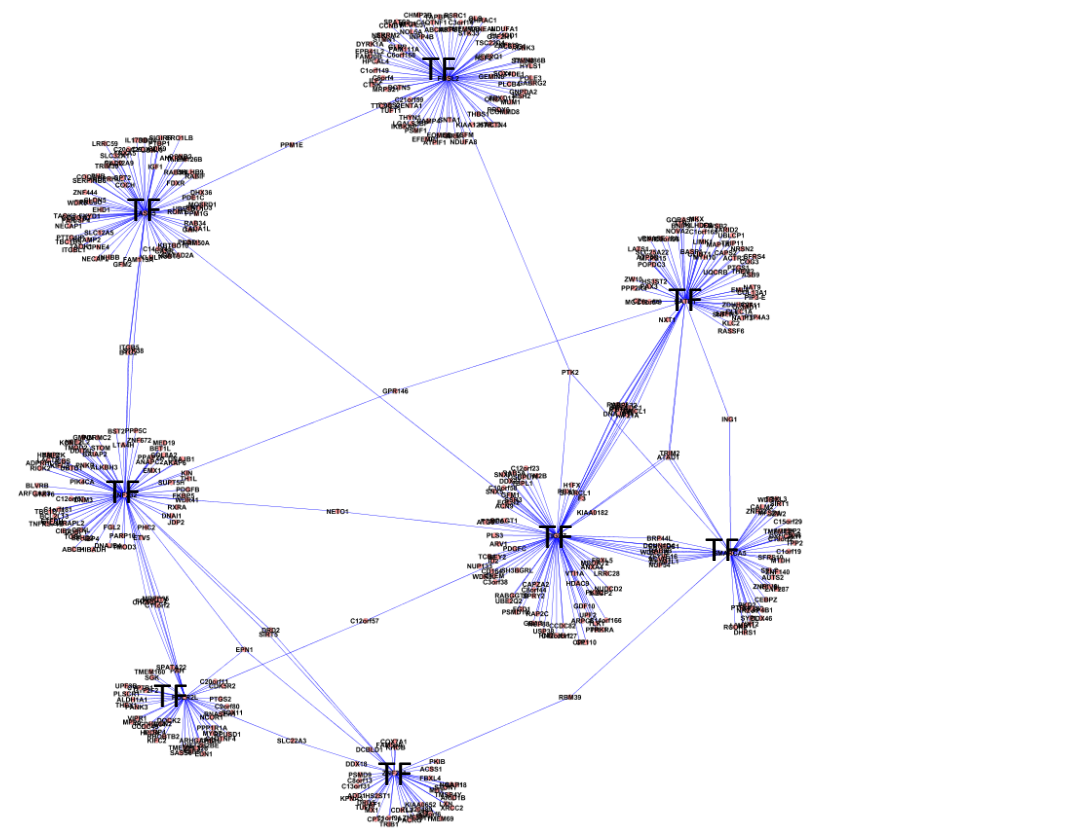


Prediction of targets and interaction partners

Genes correlated with the TFs (Spearman Rank correlation)

Assumption: Genes that are expressed together, function in the same molecular pathways

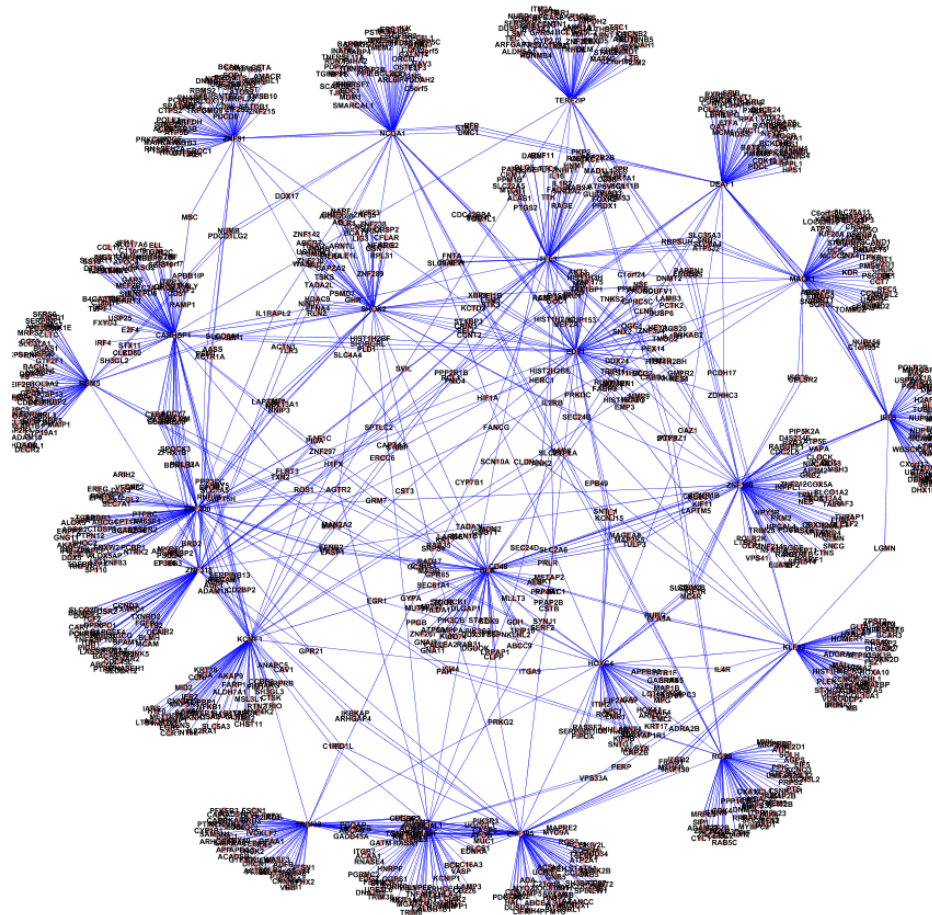
- Putative target genes
- Putative interaction partners



(Simplified representation of TFs and their co-expressed genes)

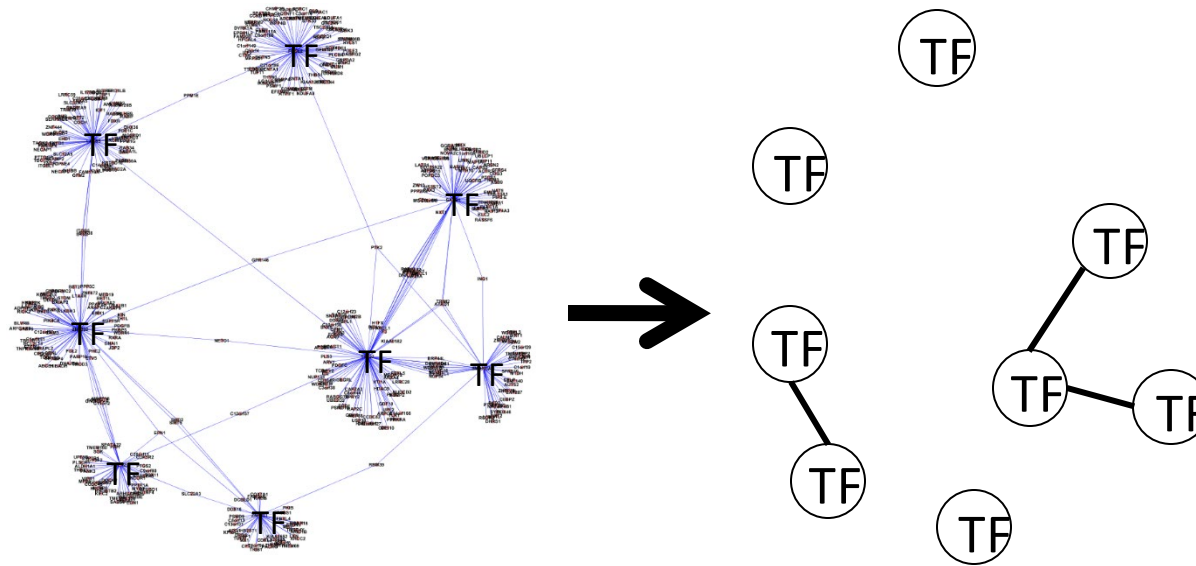
TFs Form Interaction Networks

→ Representation of TFs and their co-expressed genes



→ Many TFs share correlated genes

Weighted Topological Overlap (wTO) network for capturing TF interactions

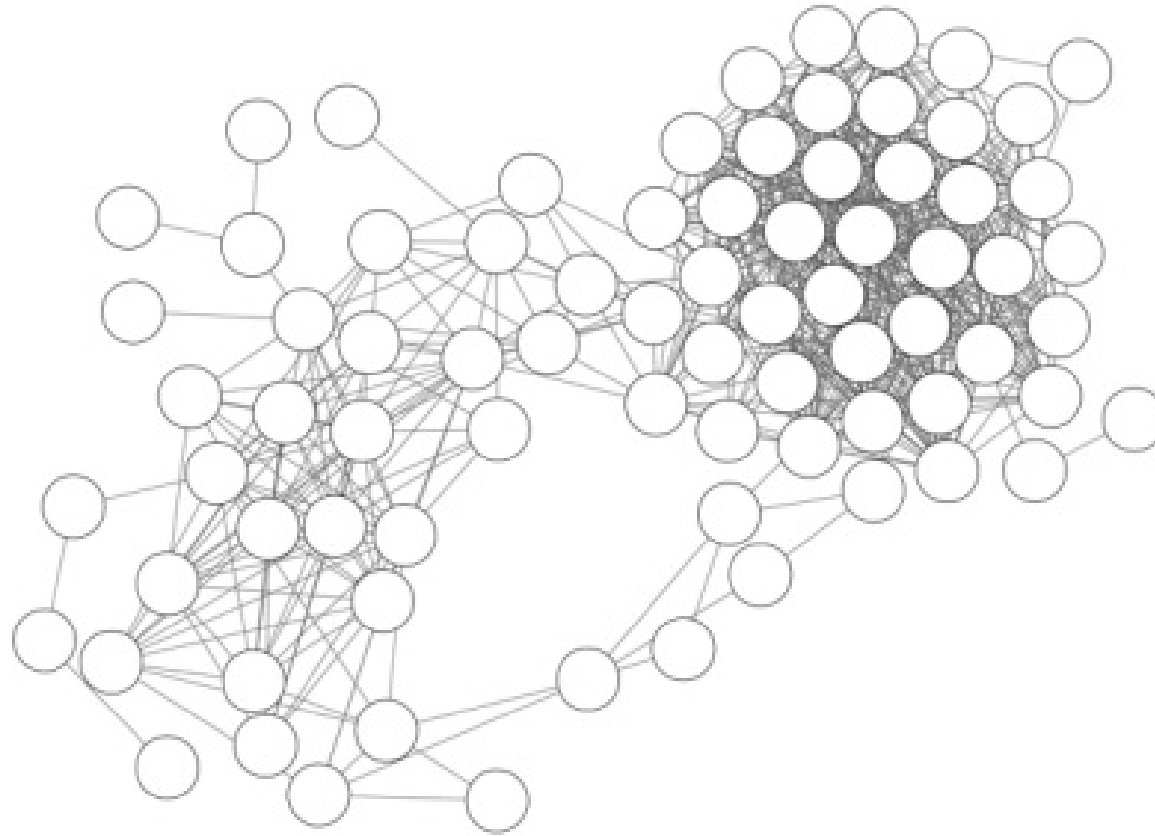


i, j
 u
 a_{ij}
 a_{iu}
 K_i

TFs, nodes in the network
genes correlated with the TFs
rho of the Spearman rank correlation between expression values of TFs i and j
rho of the Spearman rank correlation between expression values of TF i and gene u
connectivity of TF i , $\sum_i a_{ij}$

$$\omega_{ij} = \frac{\sum_u a_{iu} a_{uj} + a_{ij}}{\min(K_i, K_j) + 1 - |a_{ij}|'}$$

wTO Network Representation



Nodes = TFs

links = commonality of TFs in correlated genes

Other weighted correlation network methods

WGCNA (weighted gene co-expression network analysis) developed by Steve Horvath and colleagues

Both are based on gene expression correlations and weighting the topological overlap

→ More robust representation of gene relationships by not focusing on individual correlations

wTO explicitly considers negative correlations

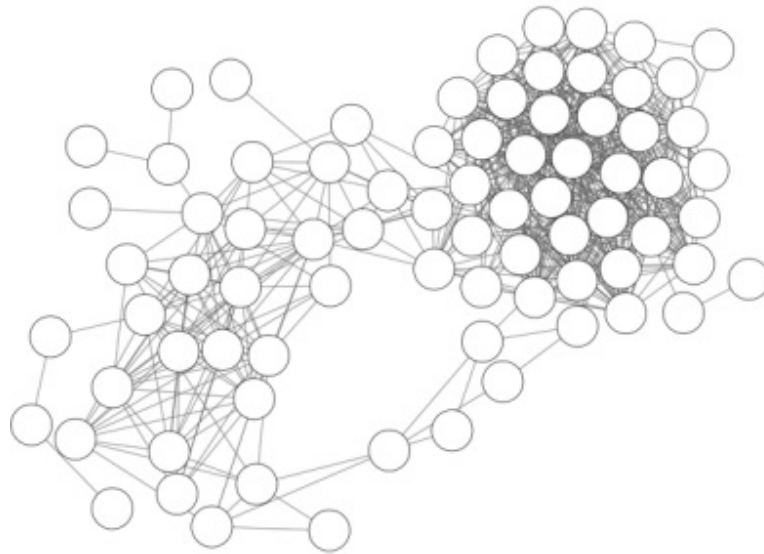
- provides p-values for links

- option to focus on a set of genes (e.g. TFs, cell cycle genes, invasion genes, positively selected genes ...)

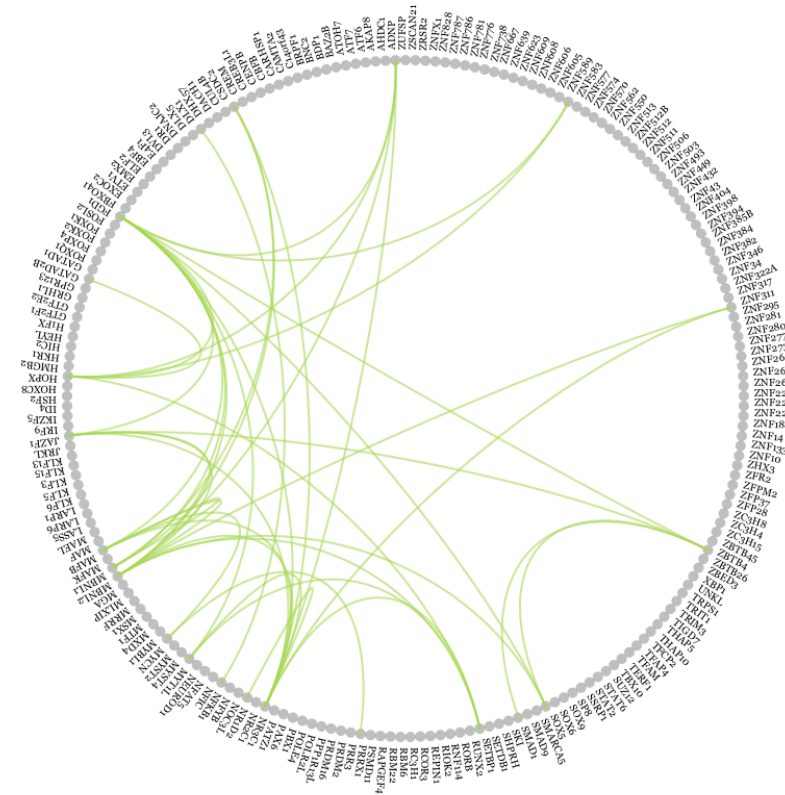
- allows for Spearman rank correlations

wTO Network Representation

Structural organization of the network



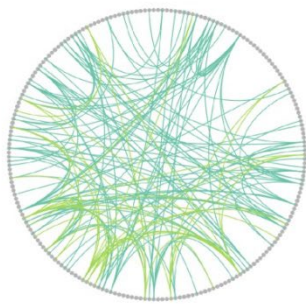
Comparison of links in the network



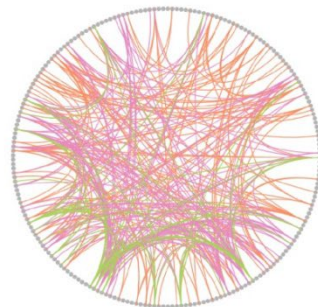
Nodes = TFs

links = commonality of TFs in correlated genes

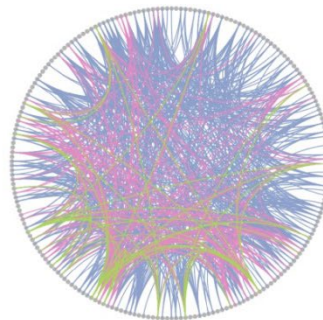
TF Network Evolution – changes of links



Rhesus macaque



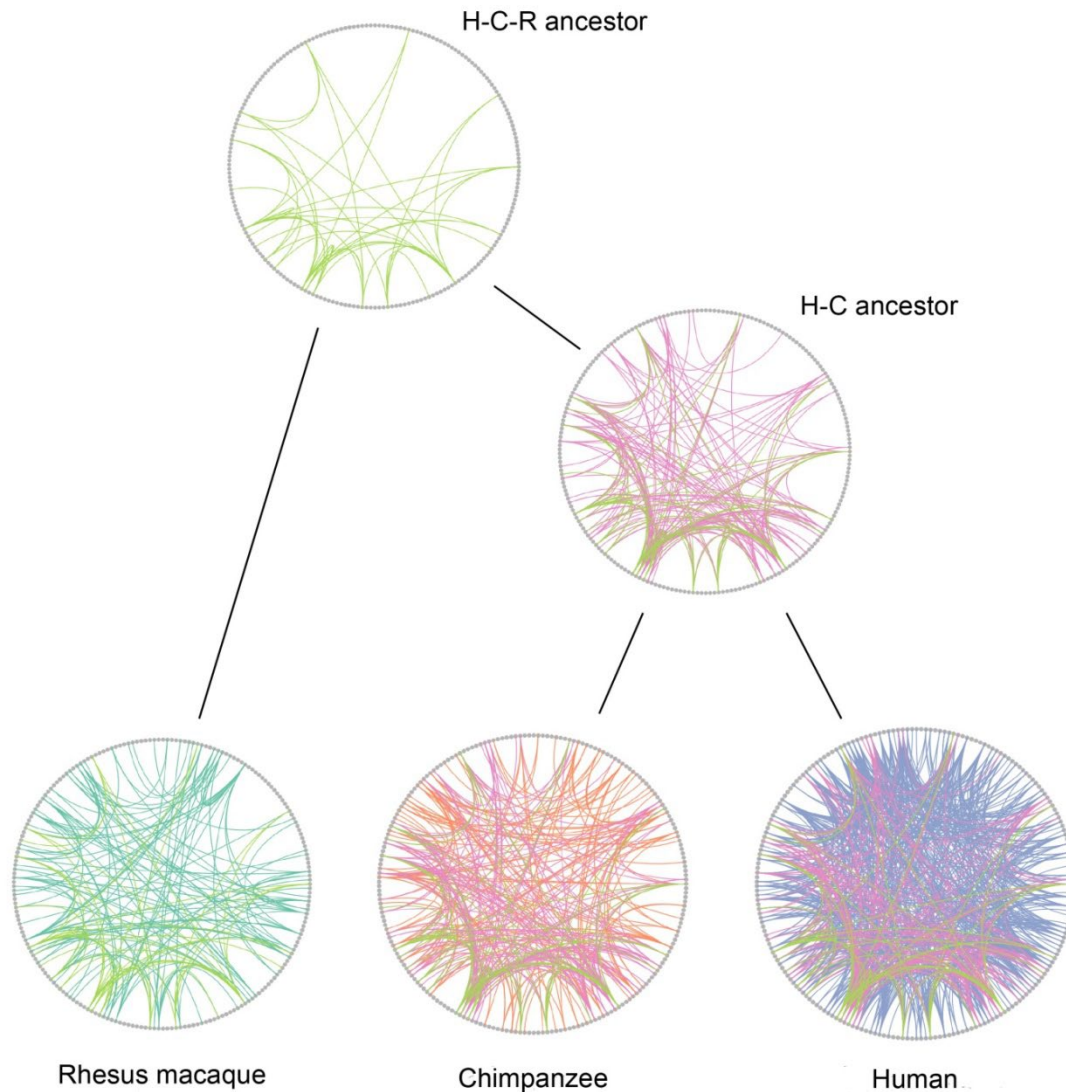
Chimpanzee



Human

wTO > 0.5

Network Evolution – changes of links



**Network of ancestor of humans, chimpanzees, and rhesus macaques:
Links present in the networks of all species**

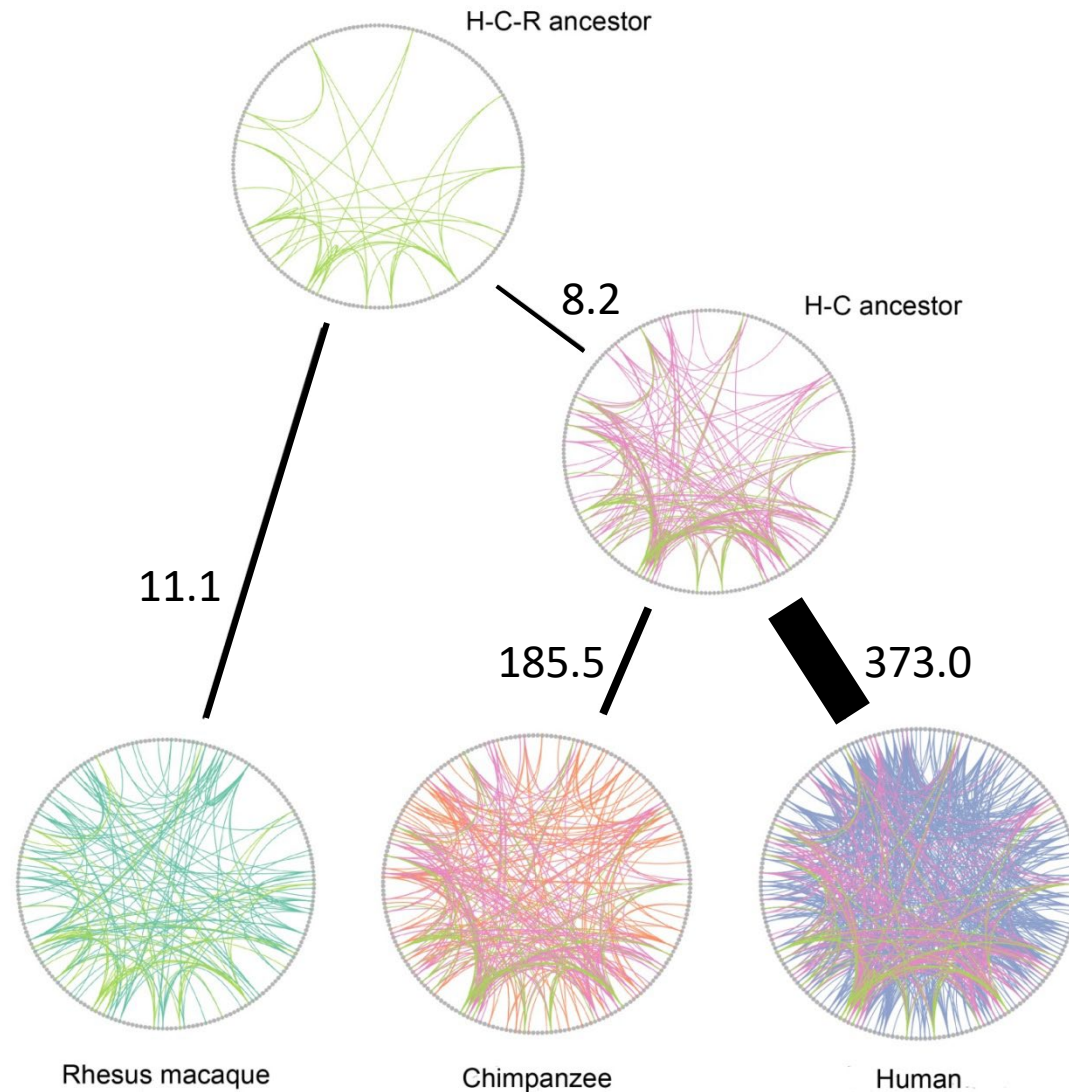
**Network of human-chimpanzee ancestor
Links present in the networks
of humans and chimpanzees**



→ Many species-specific links

wTO > 0.5

Network Evolution – changes of links



Network of ancestor of humans, chimpanzees, and rhesus macaques:
Links present in the networks of all species

Network of human-chimpanzee ancestor
Links present in the networks of humans and chimpanzees



- Many species-specific links
- Most links gained on human lineage
- Highest connectivity in humans (Wilcoxon test, $p=2.2 \times 10^{-16}$)

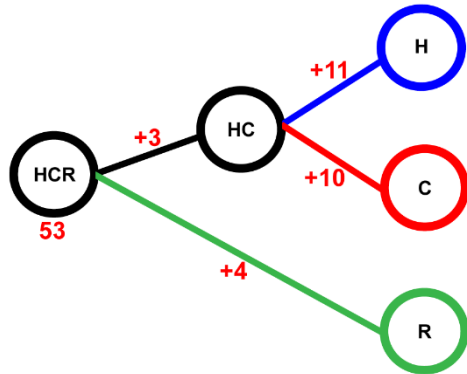
wTO > 0.5

Berto & Nowick 2018

Network Evolution – changes of links

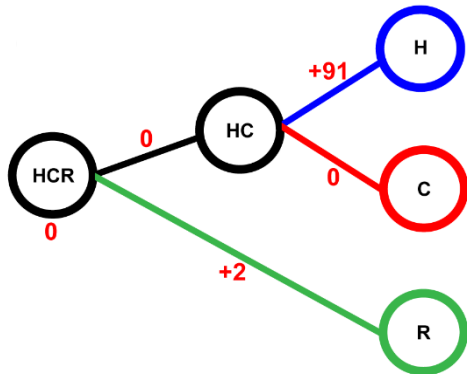
BBX

Development of neocortex and hippocampus



CC2D1A

Non-syndromic intellectual disability



Network of ancestor of humans, chimpanzees, and rhesus macaques:
Links present in the networks of all species

Network of human-chimpanzee ancestor
Links present in the networks of humans and chimpanzees

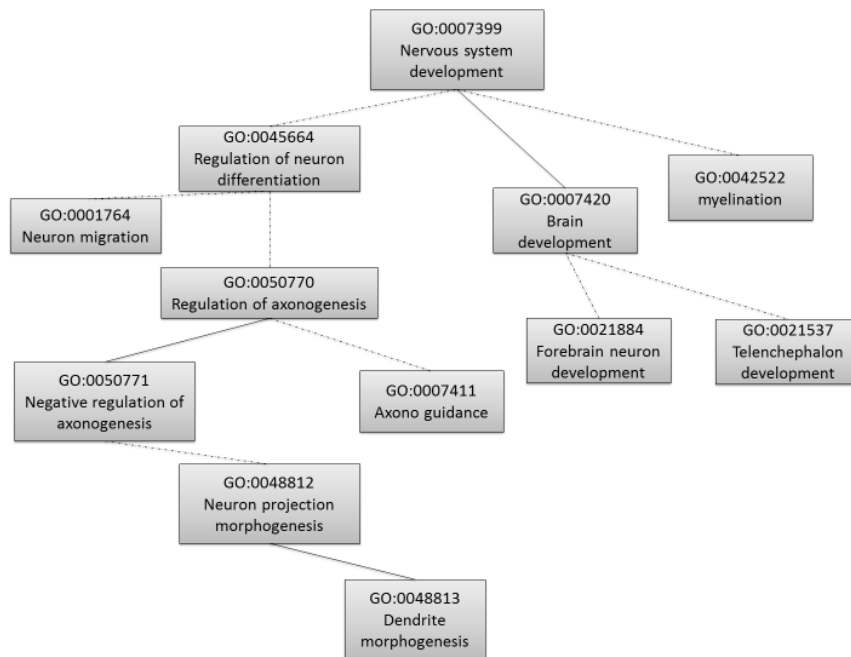


- Many species-specific links
- Most links gained on human lineage
- Highest connectivity in humans (Wilcoxon test, $p=2.2 \times 10^{-16}$)

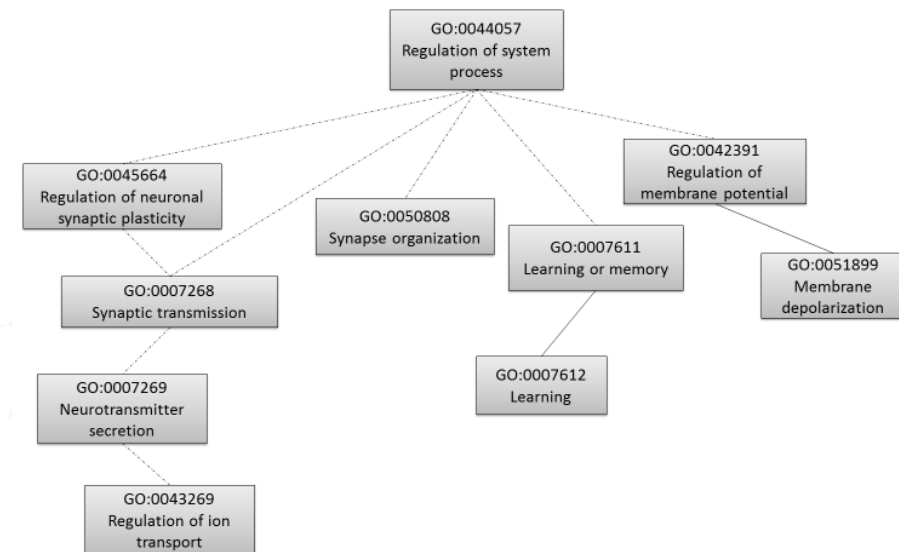
Functions of the Network

Gene Ontology enrichment among genes correlated with the TFs in the network

Nervous system development



Synaptic transmission



→ TF network involved in brain development and functions experienced increase in connectivity and rewiring during human evolution

Two R packages: wTO and CoDiNA



wTO

(Weighted Topological Overlap)

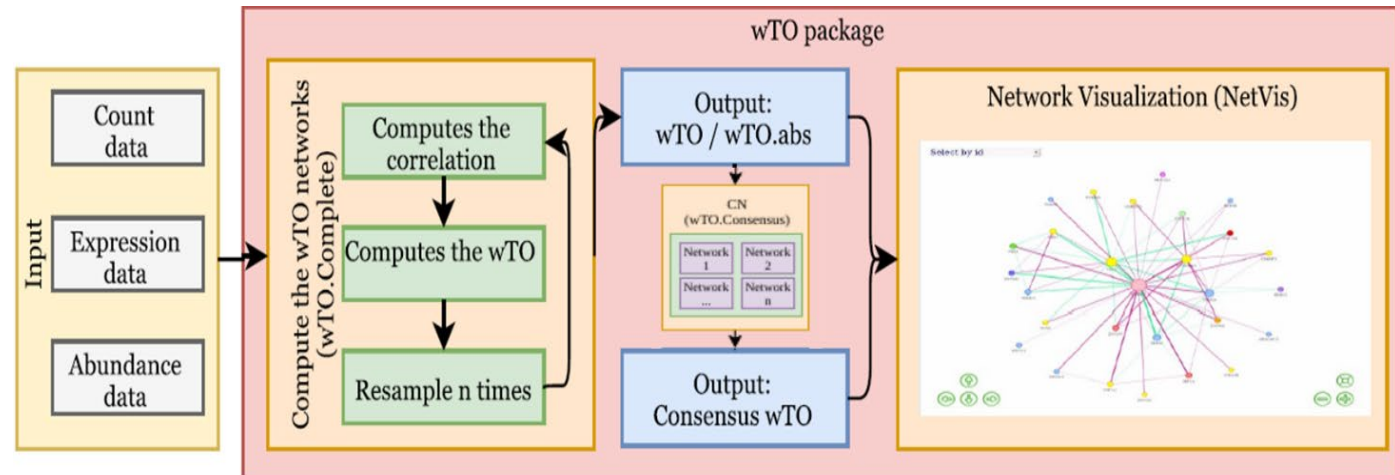
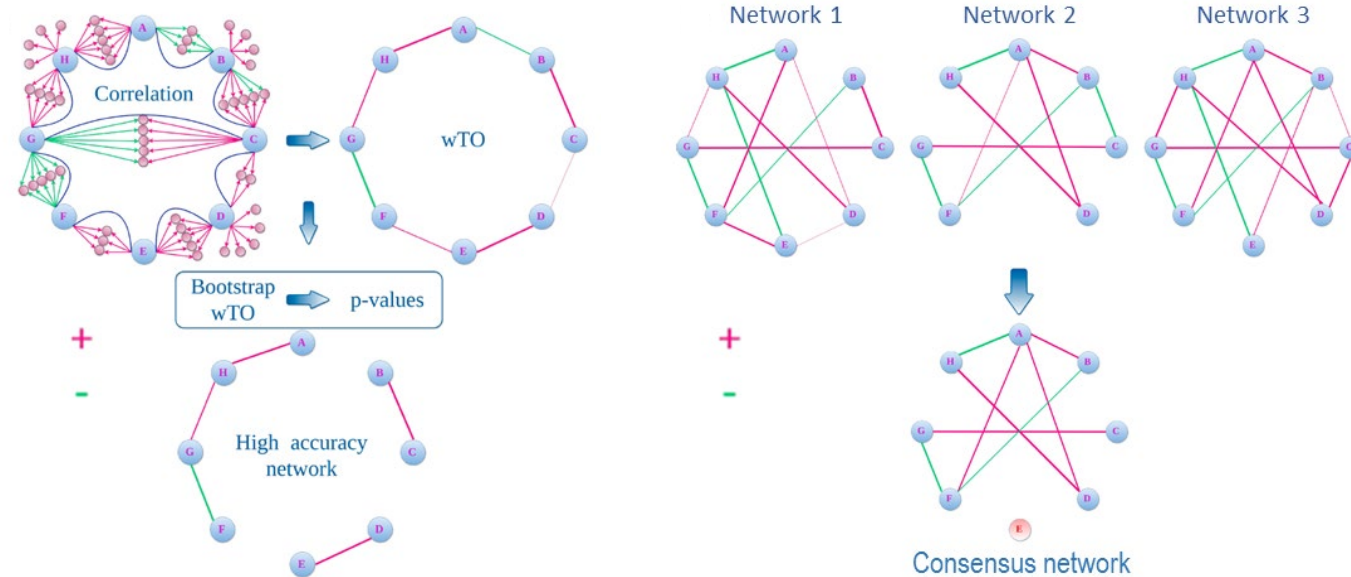
In CRAN and published:

Gysi et al. *BMC Bioinformatics* (2018)

<https://cran.r-project.org/web/packages/wTO/index.html>

Focal gene (nodes):

- TFs
- Cell cycle genes
- Inversion genes
- ...



Two R packages: wTO and CoDiNA



CoDiNA

(Co-expression Differential Network Analysis)

In CRAN and arXiv:

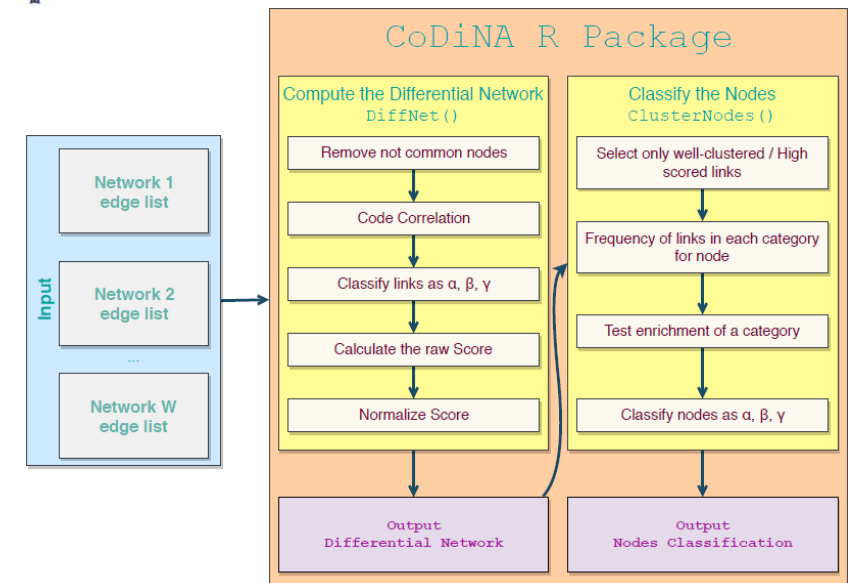
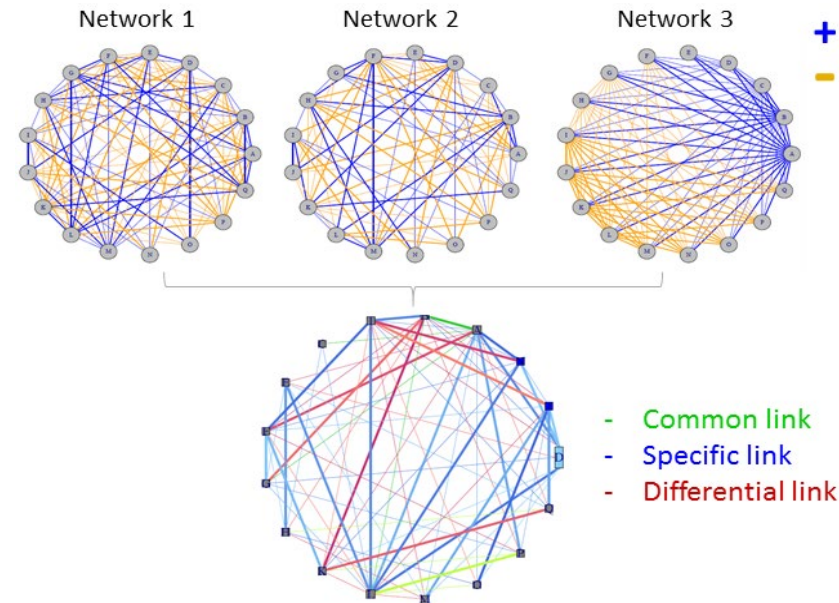
Gysi et al. PLoS One (2020)

<https://cran.r-project.org/web/packages/CoDiNA/index.html>

For as many networks as desired

- Differences between tissues
- Differences between diseases
- Differences between species
- Differences between treatments
- ...

Package downloads:
CoDiNA: 15363 times
wTO: 28575 times



Correlations

e.g. between the physical statures of parents and their offspring
between the demand for a product and its price

Correlation does not imply causation!

e.g. one may observe a correlation between an ordinary alarm clock ringing and daybreak, though there is no direct causal relationship between these events

Direction often not clear

e.g. correlation between mood and health in people: Does improved mood lead to improved health, or does good health lead to good mood?

Dataset for Network part

Social status alters immune regulation and response to infection in macaques

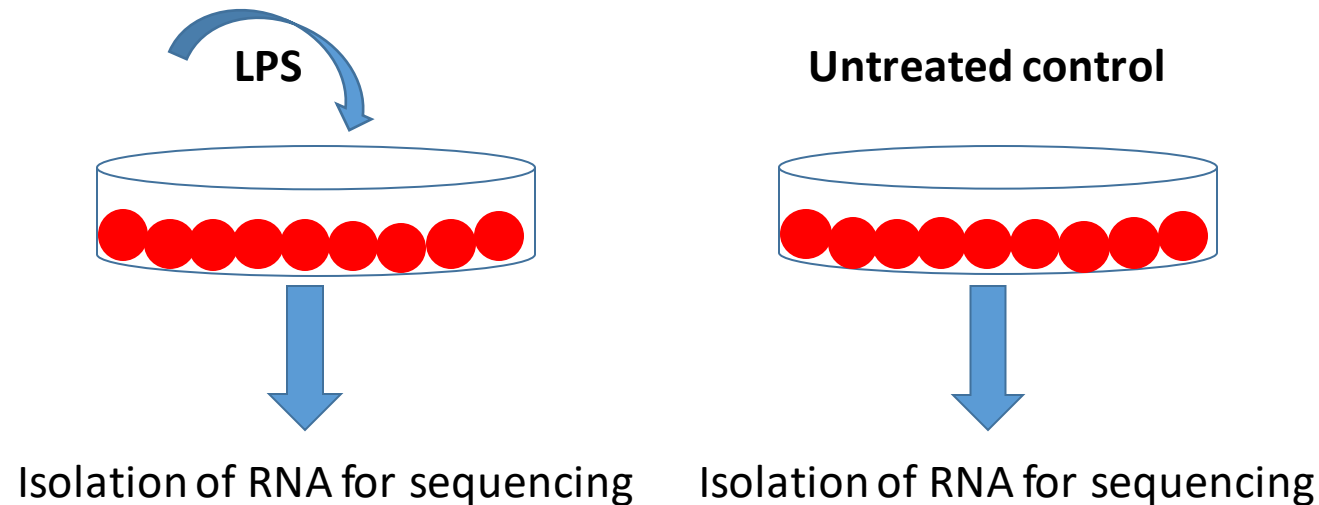
Noah Snyder-Mackler,^{1,2*} Joaquín Sanz,^{3,4*} Jordan N. Kohn,⁵ Jessica F. Brinkworth,^{3,6} Shauna Morrow,¹ Amanda O. Shaver,^{1†} Jean-Christophe Grenier,⁴ Roger Pique-Regi,^{7,8} Zachary P. Johnson,^{5,9‡} Mark E. Wilson,^{5,10} Luis B. Barreiro,^{4,11§||} Jenny Tung^{1,12,13,14§||}



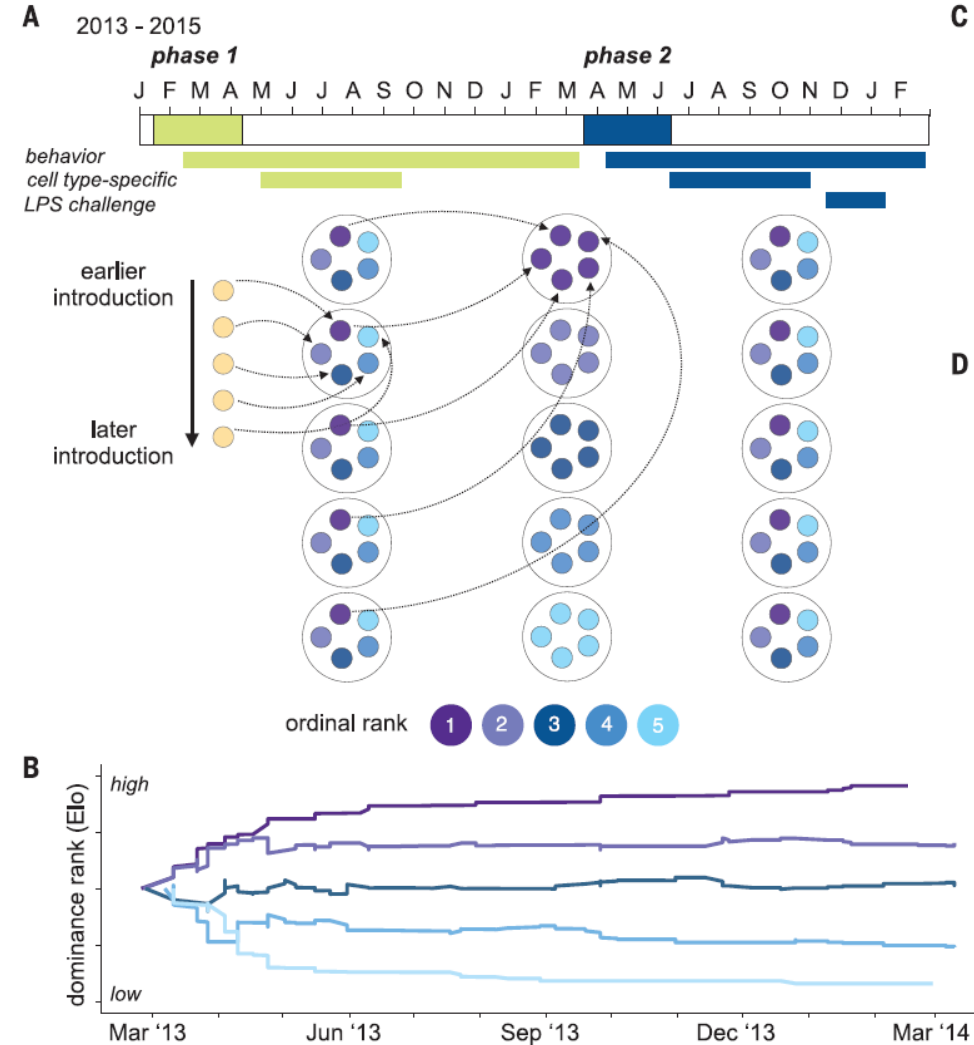
Dataset for Network part

Social status alters immune regulation and response to infection in macaques

- Lower ranking individuals suffer more stress
- Isolated blood cells from each individual
- Treated cells with LPS to activate the immune system



Dataset for the exercise



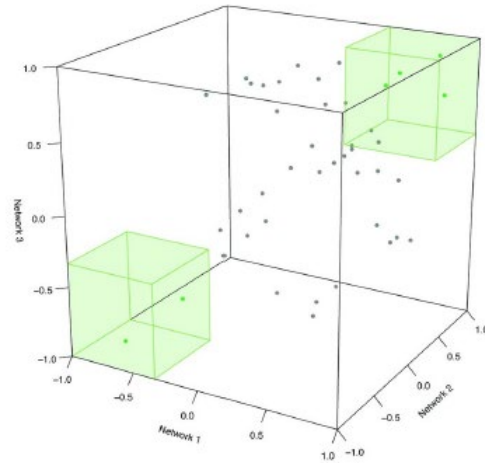


Tutorial on Gene Expression and Network Analysis

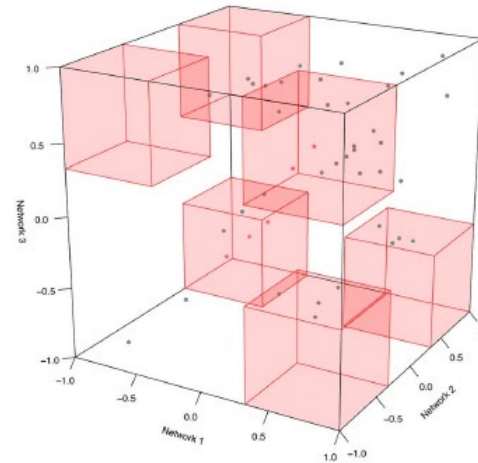
Let's run some network analyses

CoDiNA: Representation of a comparison of three networks as a cube

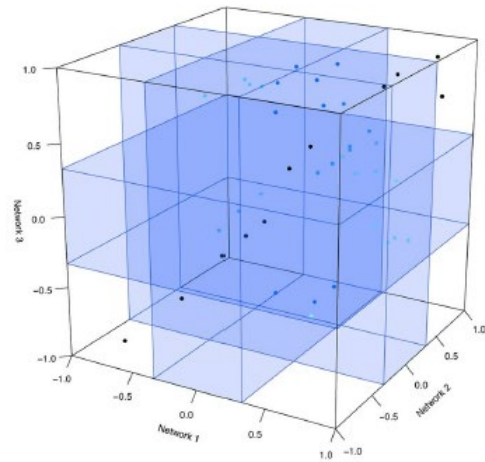
α -links



β -links



γ -links



α -, β -, and γ -links

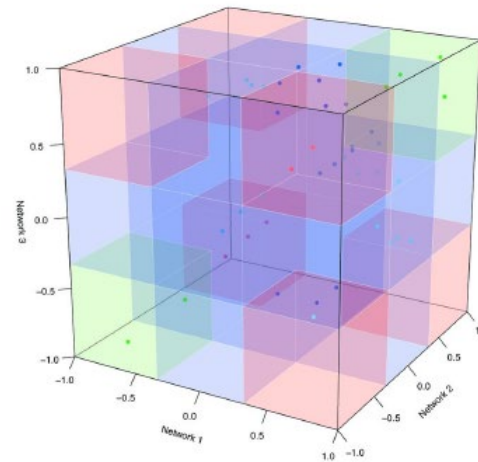
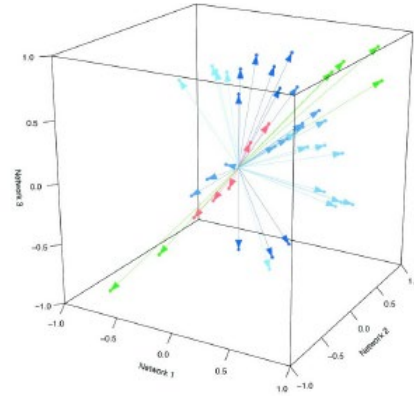


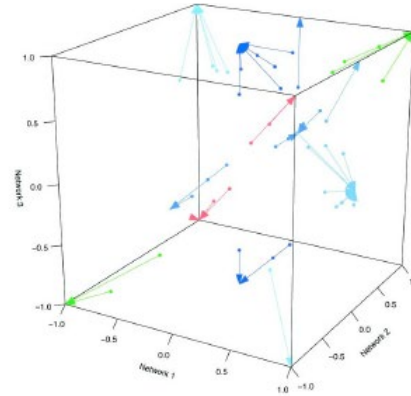
Fig 2. Visual representation of the CoDiNA method for a 3-network comparison: Categories definition. 2a represents where the α links lie in the 3D space. 2b and 2c represent the locations of β and γ links, respectively. The complete set of Φ and $\bar{\Phi}$ positions is shown in 2d.

CoDiNA: Selecting the “best” links

Strength score

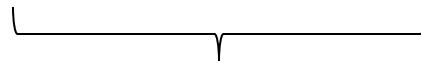


Internal score



Euclidian distance from the center of the cube
The longer, the stronger (positive or negative) the link

Distance of the link's weight to the categorical weight of α , β , and γ
The smaller, the better defined is a link



Pick links that are strong and well defined in a category
High ratio of strength score/internal score

Any Questions or further Interests?

Katja Nowick
Freie Universität Berlin
Katja.nowick@fu-berlin.de