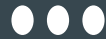


Network analysis

from Babelomics 5



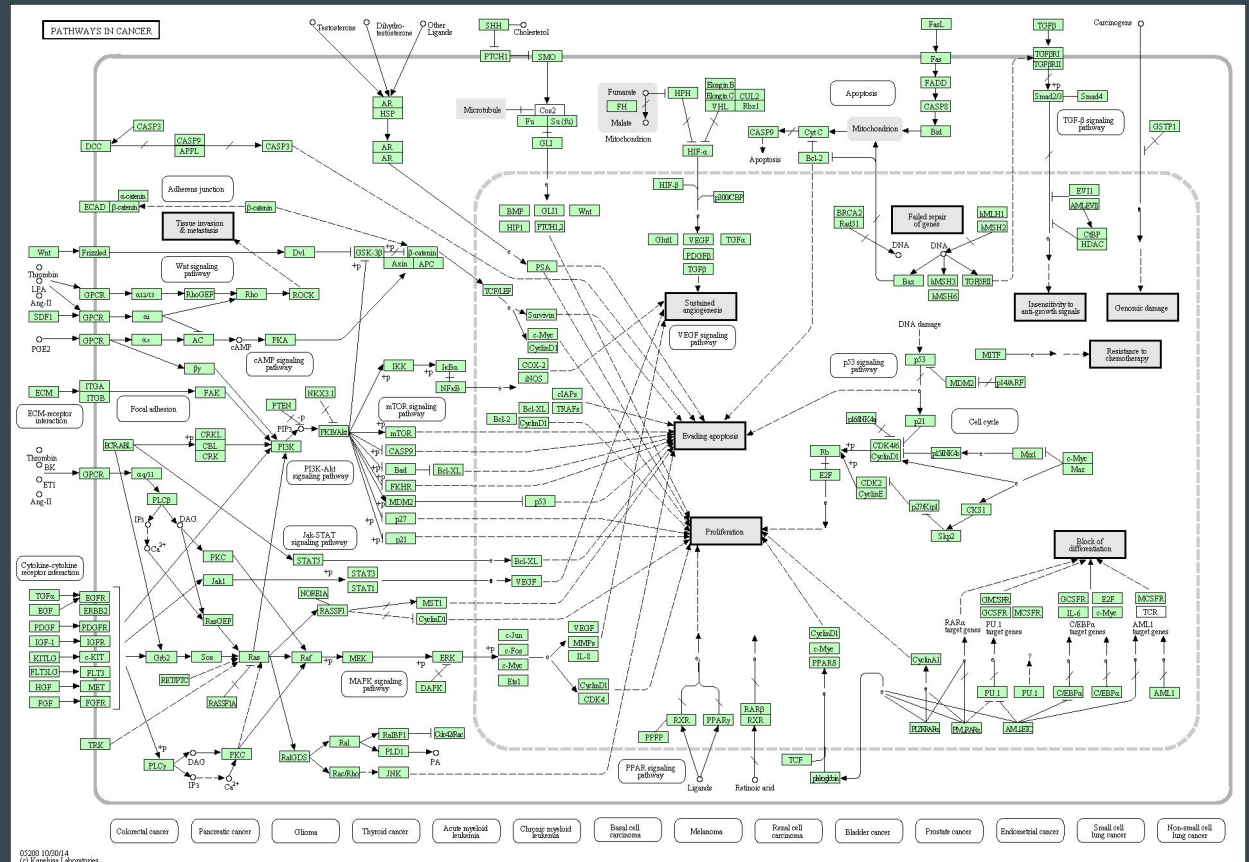
José Carbonell Caballero

Contents

- Networks in molecular biology
- Network enrichment (SNOW)
- Some exercises
- Gene set network enrichment (Network miner)
- Some exercises
- Exercises

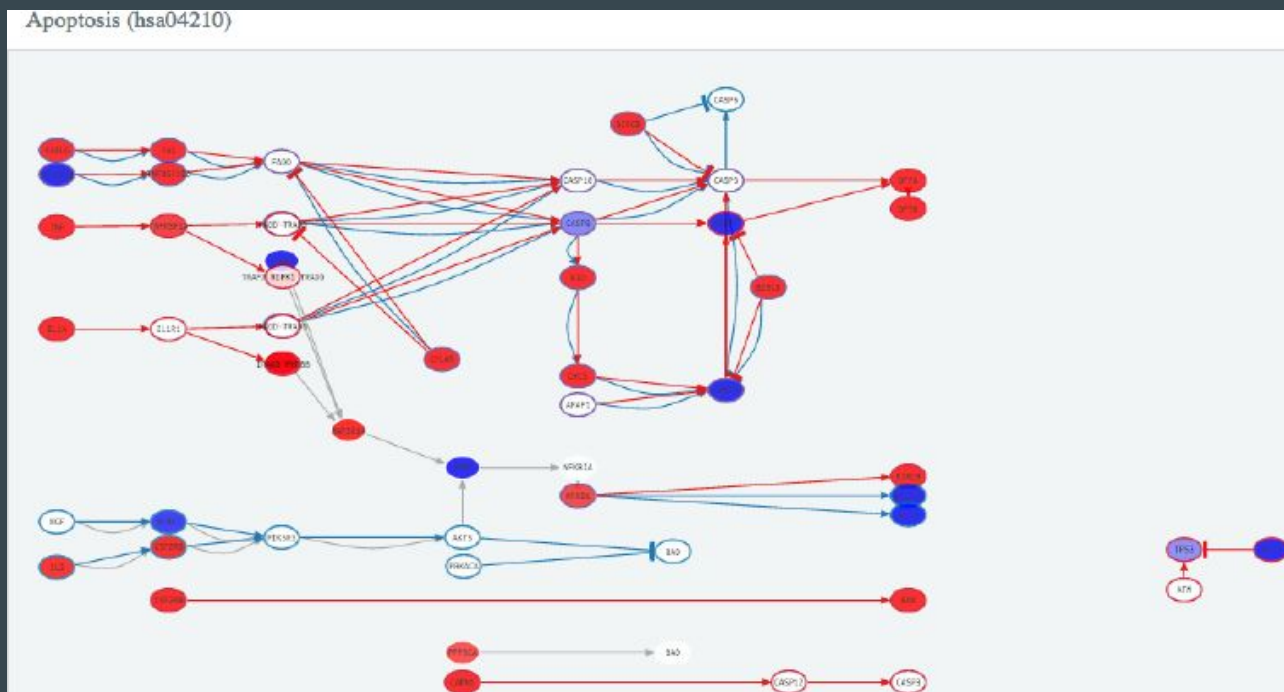
Networks in molecular biology

- Systems biology is the clue
 - Complex biological systems are mathematically modelled as a whole



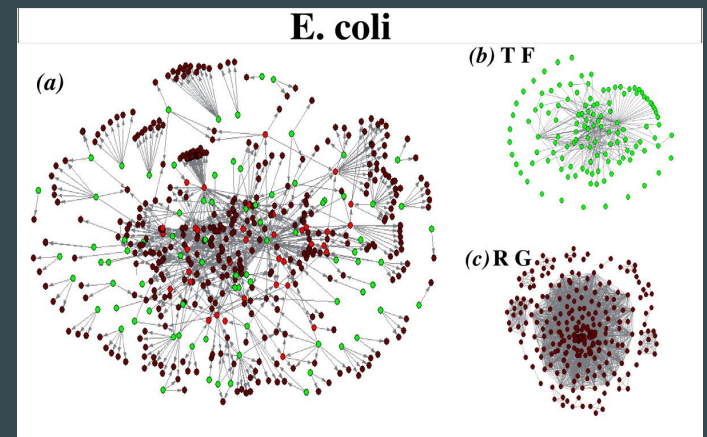
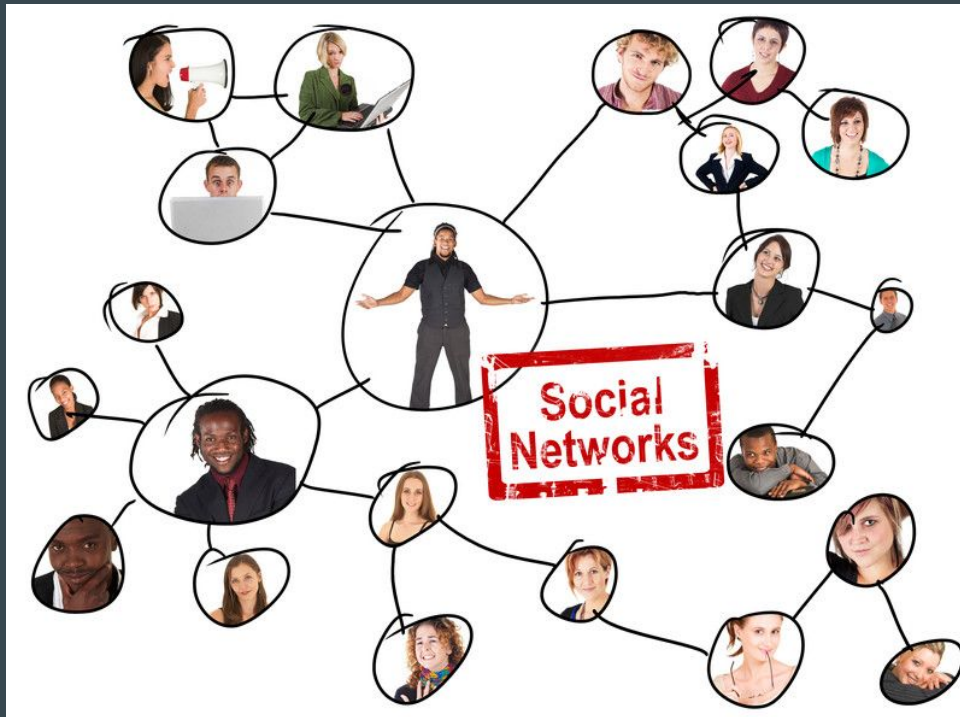
Networks in molecular biology

- ...some derived concepts
 - How the system is impaired rather than its elements
 - Explain origin of diseases
 - Face gene heterogeneity



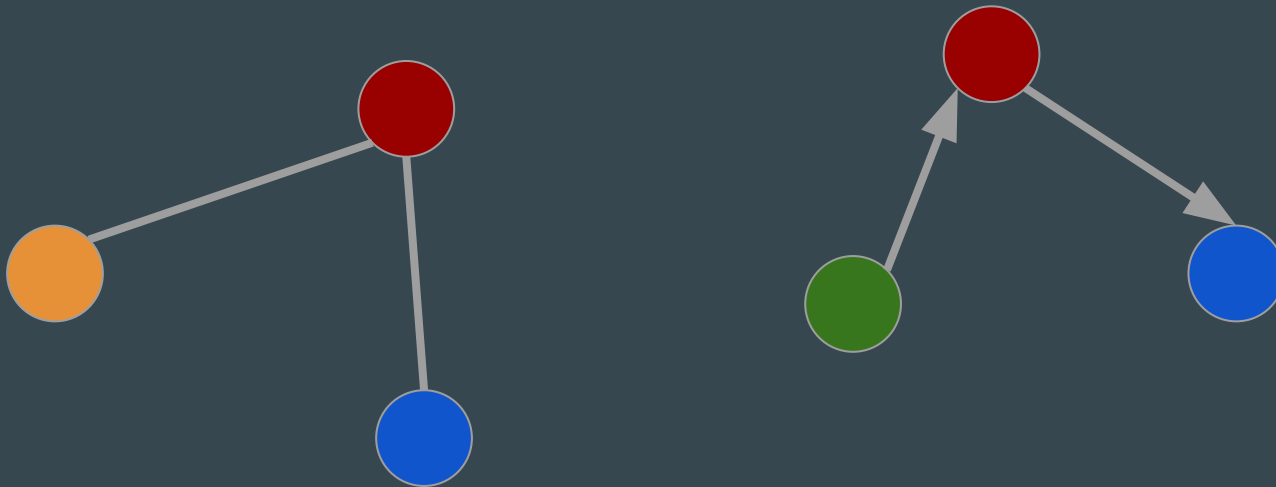
Networks in molecular biology

- Networks are extremely useful to represent different systems and their inner interactions elements



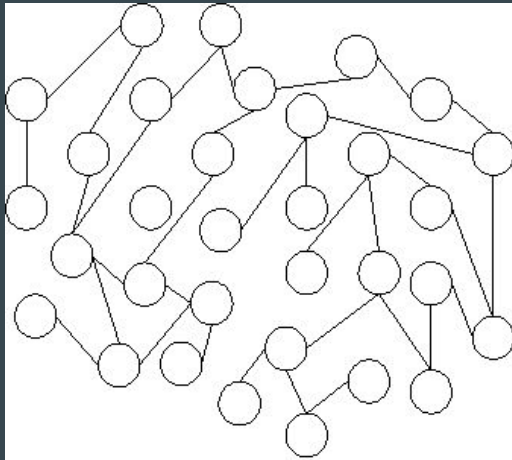
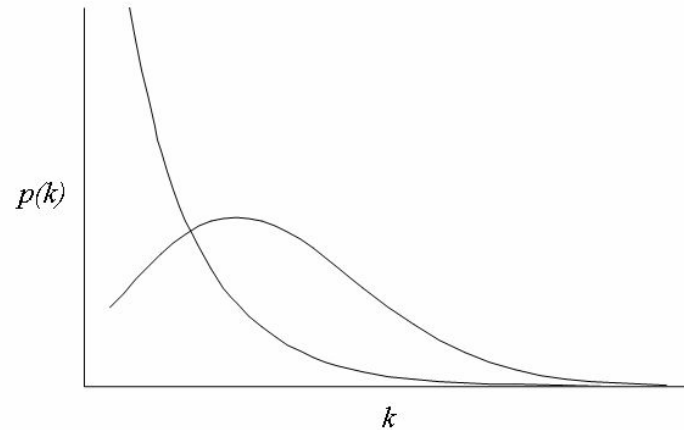
Networks in molecular biology

- Knowledge is represented as a graph
 - Proteins are the vertices
 - Interactions are the edges

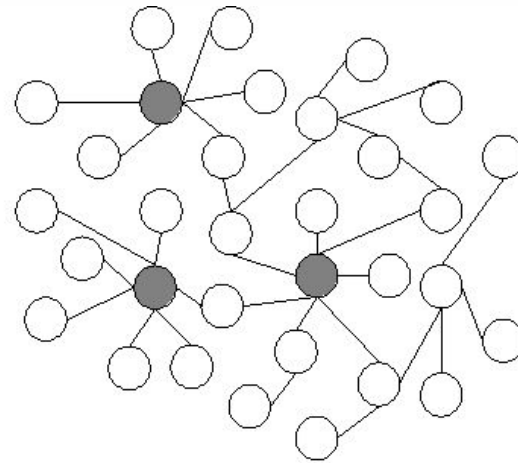


Networks in molecular biology

- Biological networks
 - Scale-free (power law)



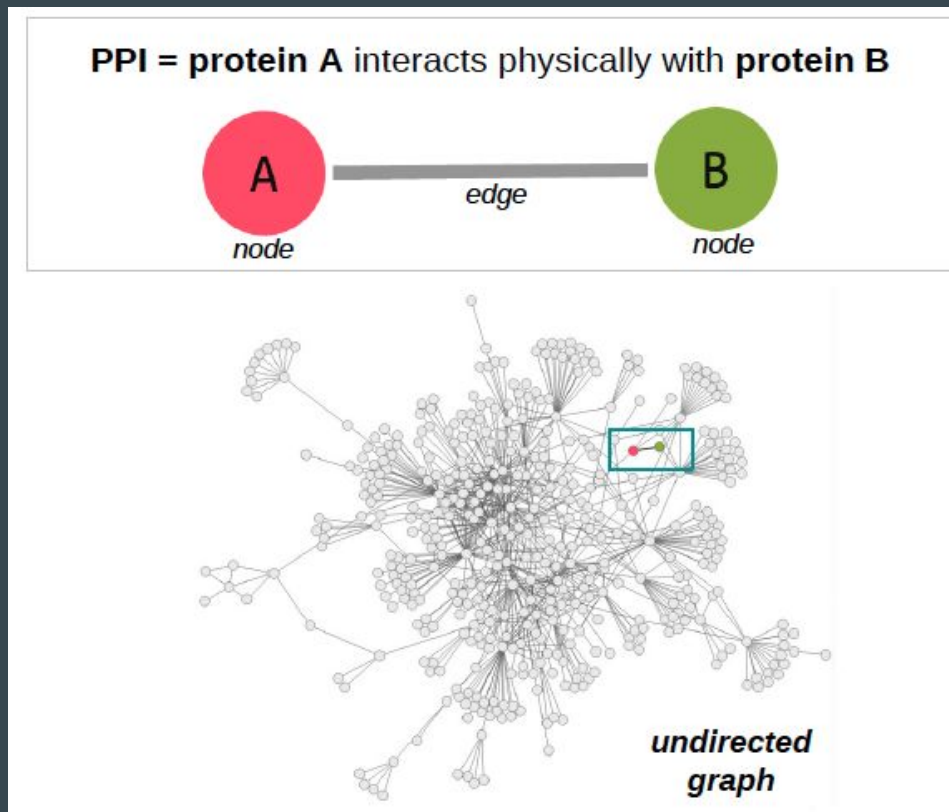
(a) Random network



(b) Scale-free network

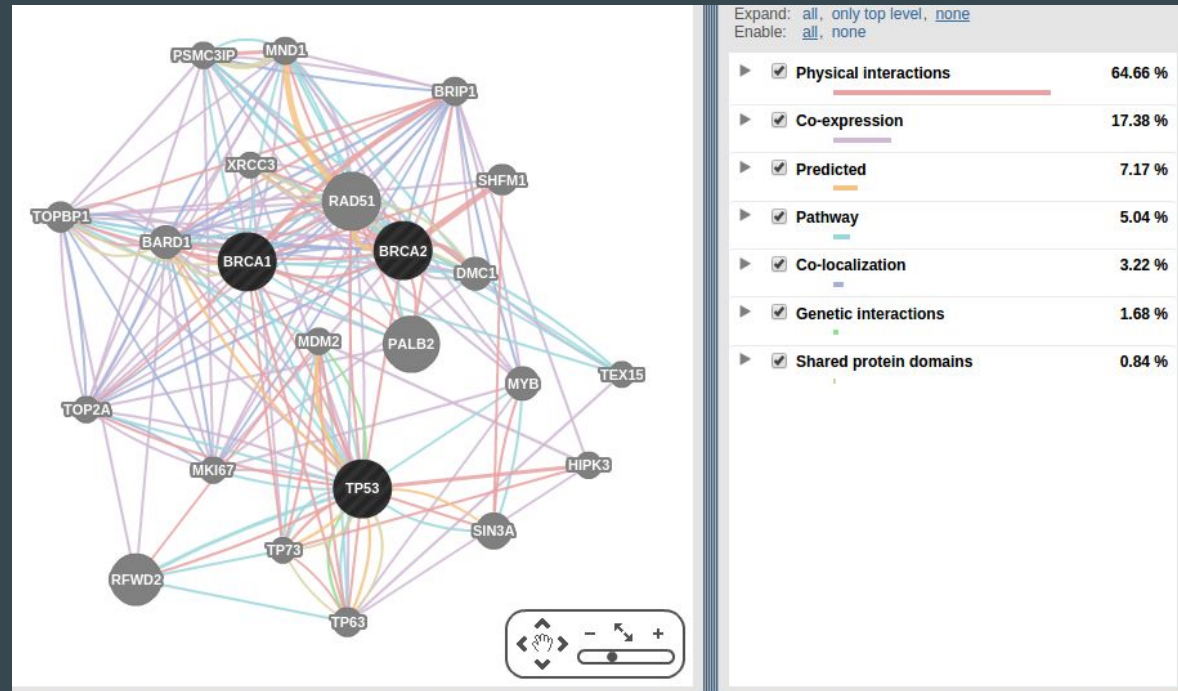
Networks in molecular biology

- Types of networks (aka interactomes)
 - Protein-protein interactions



Networks in molecular biology

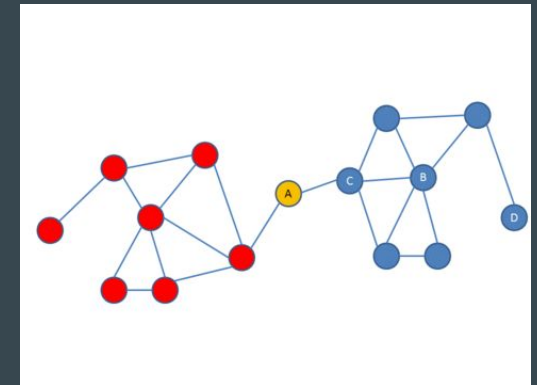
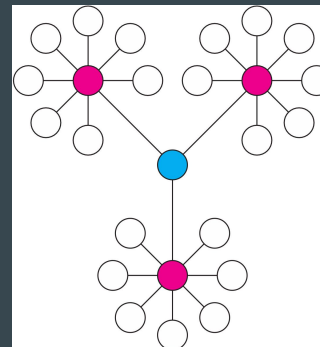
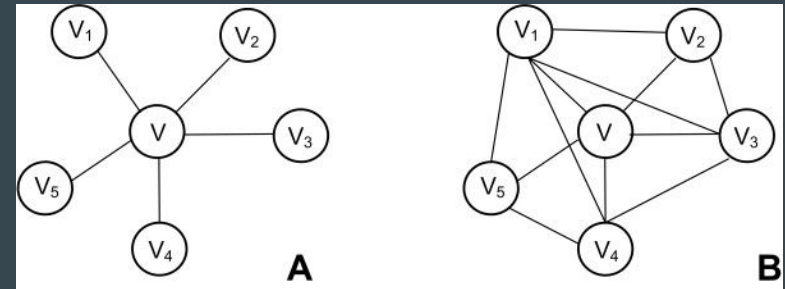
- Types of networks (aka interactomes)
 - Post-translational modifications
 - Coexpression
 - Functional terms
 - ...



Networks in molecular biology

Network are evaluated through some parameters:

- At a network level:
 - Number of components
 - Clustering coefficient
- At a gene level:
 - Closeness centrality
 - Betweenness centrality
 - Degree



Networks in molecular biology

Disease-gene prioritization approaches

ARTICLE

Walking the Interactome for Prioritization of Candidate Disease Genes

Sebastian Köhler,^{1,2} Sebastian Bauer,^{1,2} Denise Horn,¹ and Peter N. Robinson^{1,*}

The identification of genes associated with hereditary disorders has contributed to improving medical care and to a better understanding of gene functions, interactions, and pathways. However, there are well over 1500 Mendelian disorders whose molecular basis remains unknown. At present, methods such as linkage analysis can identify the chromosomal region in which unknown disease genes are located, but the regions could contain up to hundreds of candidate genes. In this work, we present a method for prioritization of candidate genes by use of a global network distance measure, random walk analysis, for definition of similarities in protein-protein interaction networks. We tested our method on 110 disease-gene families with a total of 783 genes and achieved an area under the ROC curve of up to 98% on simulated linkage intervals of 100 genes surrounding the disease gene, significantly outperforming previous methods based on local distance measures. Our results not only provide an improved tool for positional-cloning projects but also add weight to the assumption that phenotypically similar diseases are associated with disturbances of subnetworks within the larger protein interactome that extend beyond the disease proteins themselves.

Bioinformatics Advance Access published July 30, 2014

Walking the interactome for candidate prioritization in exome sequencing studies of Mendelian diseases

Damian Smedley^{1,†}, Sebastian Köhler^{2,†}, Johanna Christina Czeschik³, Joanna Amberger⁴, Carol Bocchini⁴, Ada Hamosh⁴, Julian Veldboer^{2,5}, Tomasz Zemojtel^{2,6}, and Peter N Robinson^{2,5,7,8*}

¹The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK, ²Institute for Medical Genetics and Human Genetics, Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany, ³Genome Informatics, Institute of Human Genetics, University Hospital Essen, University of Duisburg-Essen, Hufelandstr. 55, 45122 Essen, Germany, ⁴Baltimore, MD, USA, ⁵Institute for Bioinformatics, Department of Mathematics and Computer Science, Freie Universität Berlin, Takustr. 9, 14195 Berlin, ⁶Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland ⁷Berlin-Brandenburg Center for Regenerative Therapies, Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany, ⁸Max Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany

BRIEFINGS IN FUNCTIONAL GENOMICS, VOL 10, NO 5, 280–293

doi:10.1093/bfgp/bt024

Network-based methods for human disease gene prediction

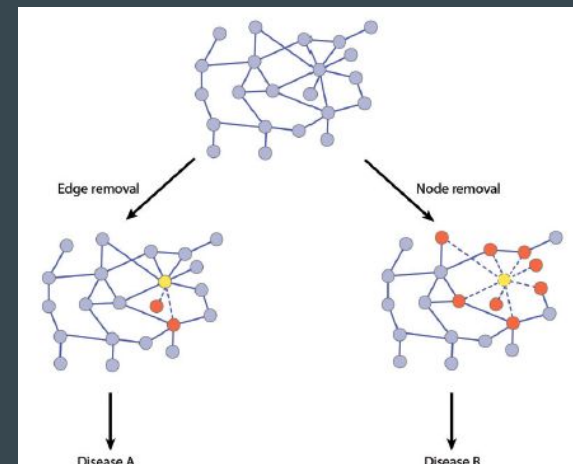
Xiujuan Wang, Natali Gulbahce and Haiyuan Yu

Advance Access publication date 15 July 2011

Abstract

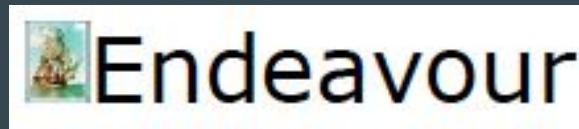
Despite the considerable progress in disease gene discovery, we are far from uncovering the underlying cellular mechanisms of diseases since complex traits, even many Mendelian diseases, cannot be explained by simple genotype–phenotype relationships. More recently, an increasingly accepted view is that human diseases result from perturbations of cellular systems, especially molecular networks. Genes associated with the same or similar diseases commonly reside in the same neighborhood of molecular networks. Such observations have built the basis for a large collection of computational approaches to find previously unknown genes associated with certain diseases. The majority of the methods are based on protein interactome networks, with integration of other large-scale genomic data or disease phenotype information, to infer how likely it is that a gene is associated with a disease. Here, we review recent, state of the art, network-based methods used for prioritizing disease genes as well as unraveling the molecular basis of human diseases.

Keywords: human diseases; disease network; disease gene prediction; protein–protein interaction; molecular network

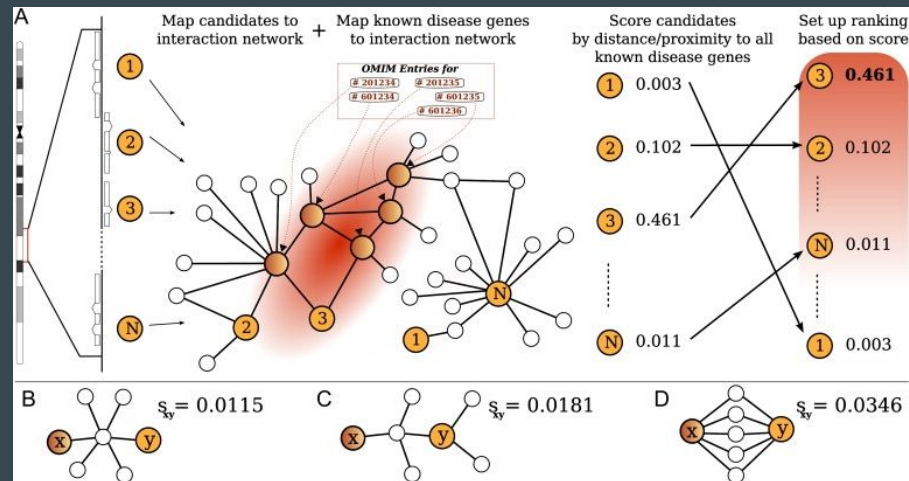


Networks in molecular biology


Gene prioritization approaches






exomewalker




Network enrichment (SNOW)


 Babelomics 5


Expression ▾ Genomics ▾ Cancer ▾ Functional ▾

Upload  profile  jobs 

Snow 

Examples

Downregulated in fibroblasts from old individuals, compared to young 

Upregulated by induction of exogenous BRCA1 in EcR-293 cells 


Define your input data

☒ One list ☐ Comparing two lists

Select your input files

List 1:

☒ File ☐ Text area

The files must be on the server to select them.
You can upload files using the button  inside file browser.

File browser

WorkSpace/

List nature

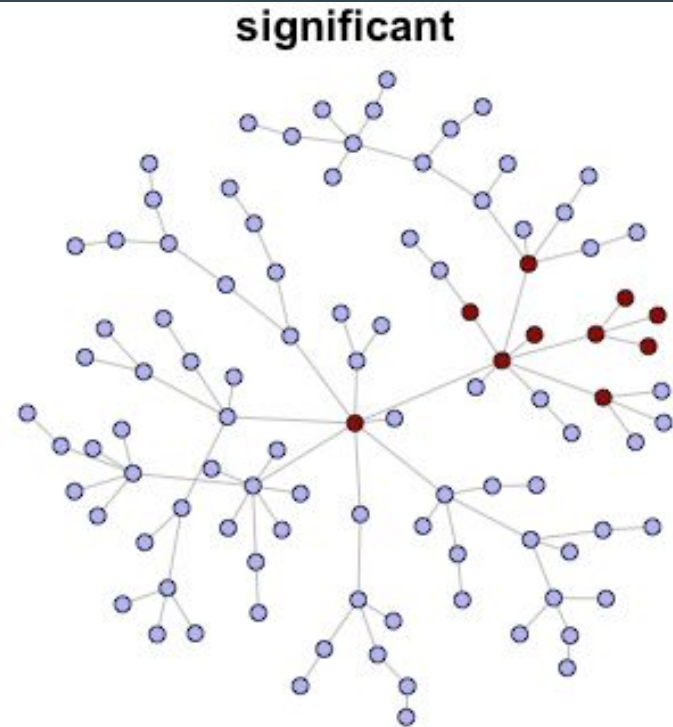
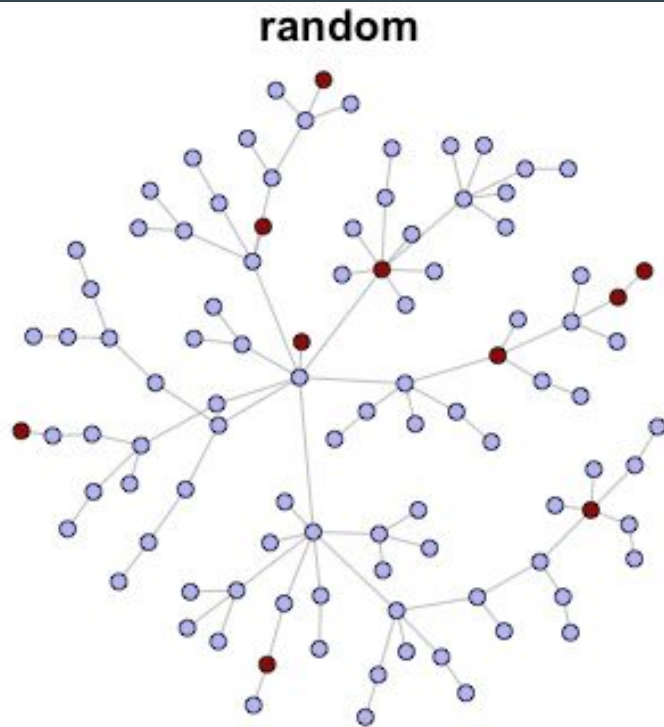
☐ Transcripts ☒ Proteins ☐ Genes

Species

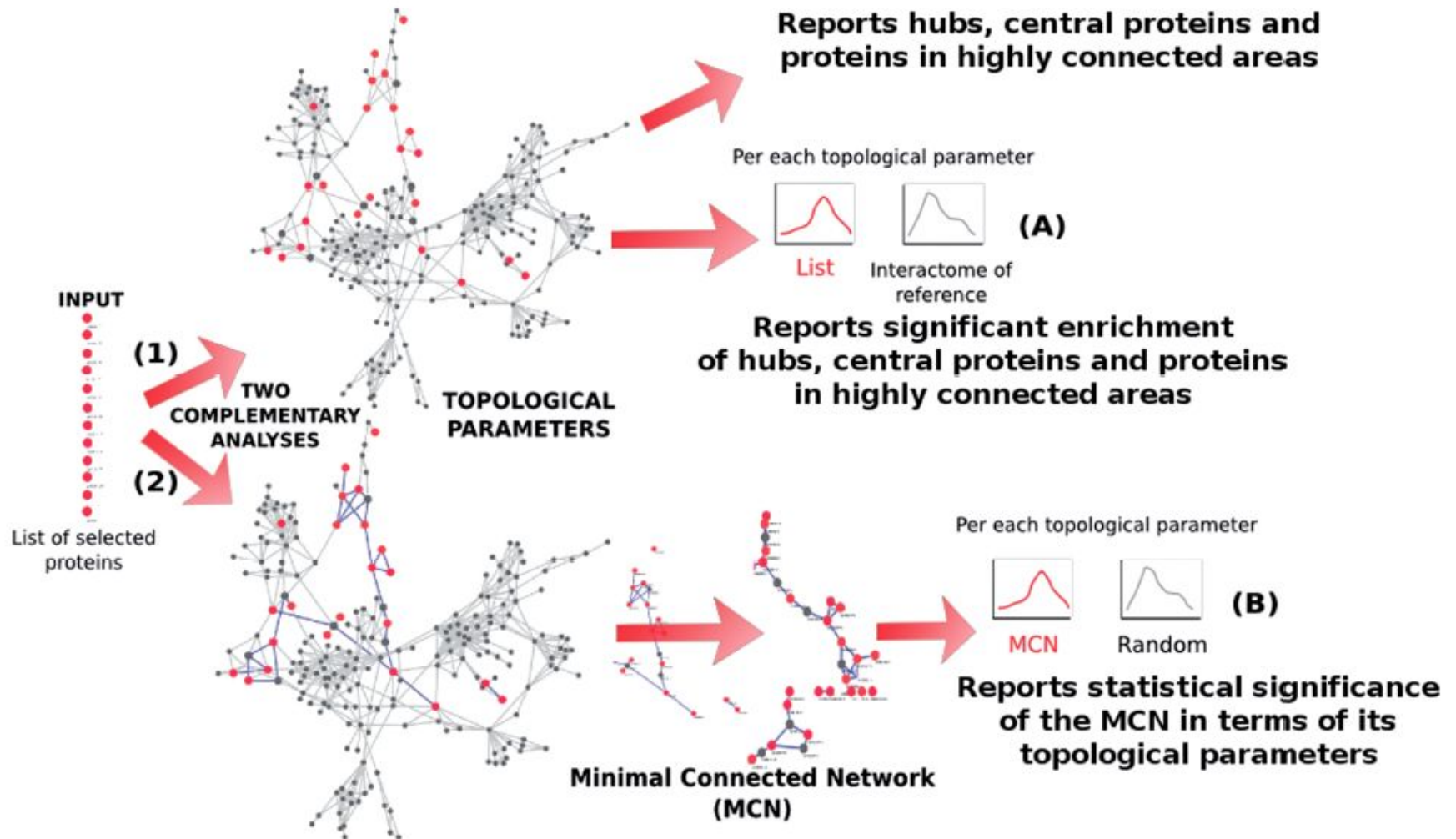
Homo sapiens ▾

Network enrichment (SNOW)

- GOAL: it determines whether a set of input genes represents a biologically meaningful network



Network enrichment (SNOW)



Network enrichment (SNOW)

Interactome construction

1 Interactions retrieval

BioGRID

MINT

IntAct

Annotated using the
Molecular Interaction
ontology vocabulary



2 All physical PPIs

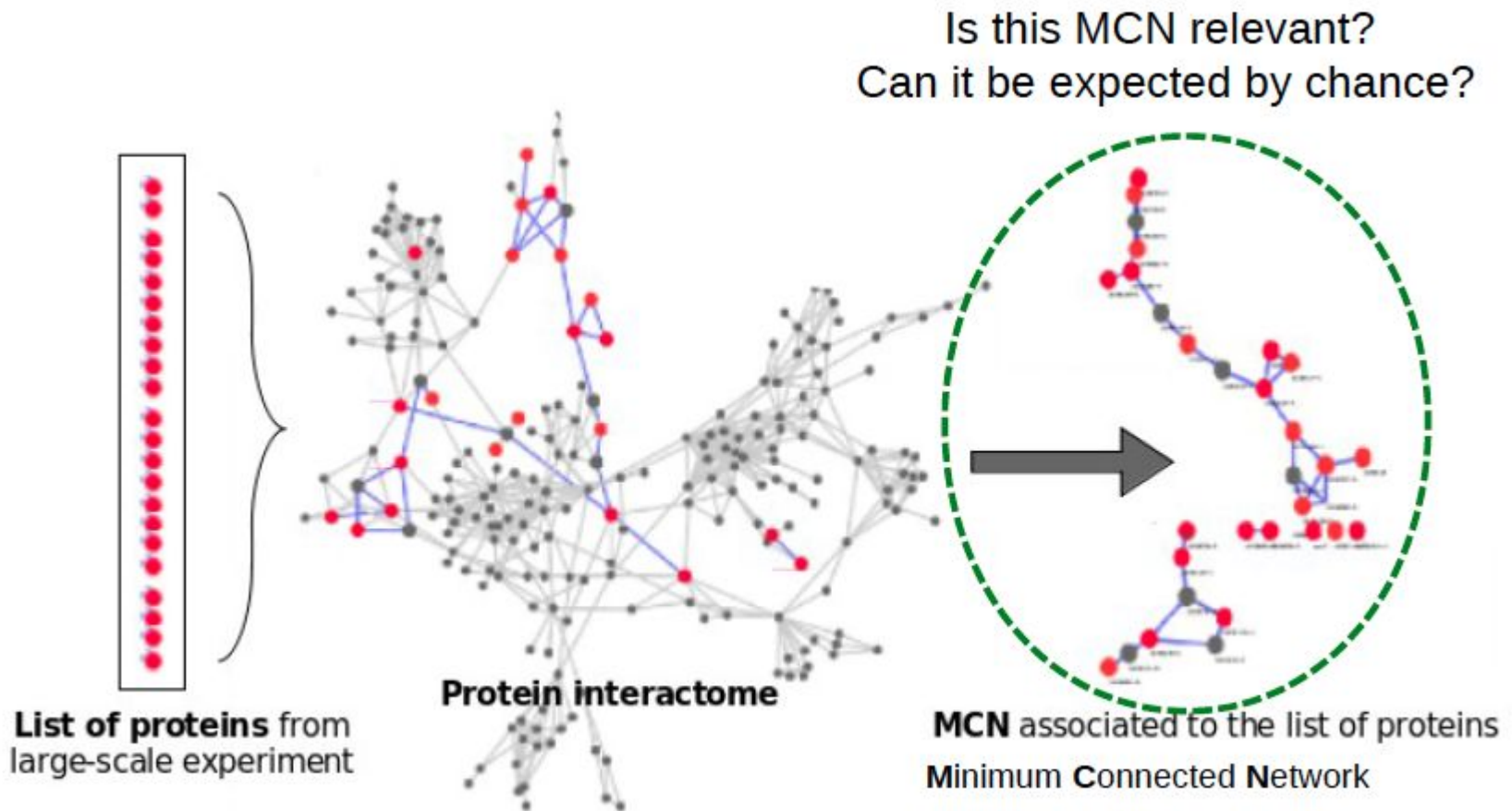
- [-] molecular interaction
 - [+] alias type
 - [+] attribute name
 - [+] biological role
 - [+] cooperative interaction
 - [+] cross-reference type
 - [+] curation content
 - [+] curation quality
 - [+] database citation
 - [+] experimental preparation
 - [+] experimental role
 - [+] feature detection method
 - [+] feature range status
 - [+] feature type
 - [+] interaction confidence
 - [+] interaction detection method
 - [+] interaction type
 - [-] association
 - [+] physical association
 - [-] colocalization
 - [+] genetic interaction
 - [+] predicted interaction
 - [+] interactor type
 - [+] parameter type
 - [+] parameter unit
 - [+] participant identification method

3 Curated PPIs

- [-] molecular interaction
 - [+] alias type
 - [+] attribute name
 - [+] biological role
 - [+] cooperative interaction
 - [+] cross-reference type
 - [+] curation content
 - [+] curation quality
 - [+] database citation
 - [+] experimental preparation
 - [+] experimental role
 - [+] feature detection method
 - [+] feature range status
 - [+] feature type
 - [+] interaction confidence
 - [+] interaction detection method
 - [-] experimental interaction detection
 - [+] biochemical
 - [+] biophysical
 - [+] genetic interference
 - [+] imaging technique
 - [+] phenotype-based detection assay
 - [+] post transcriptional interference
 - [+] protein complementation assay
 - [+] inference
 - [+] interaction prediction
 - [+] unspecified method

2 different methods
criteria

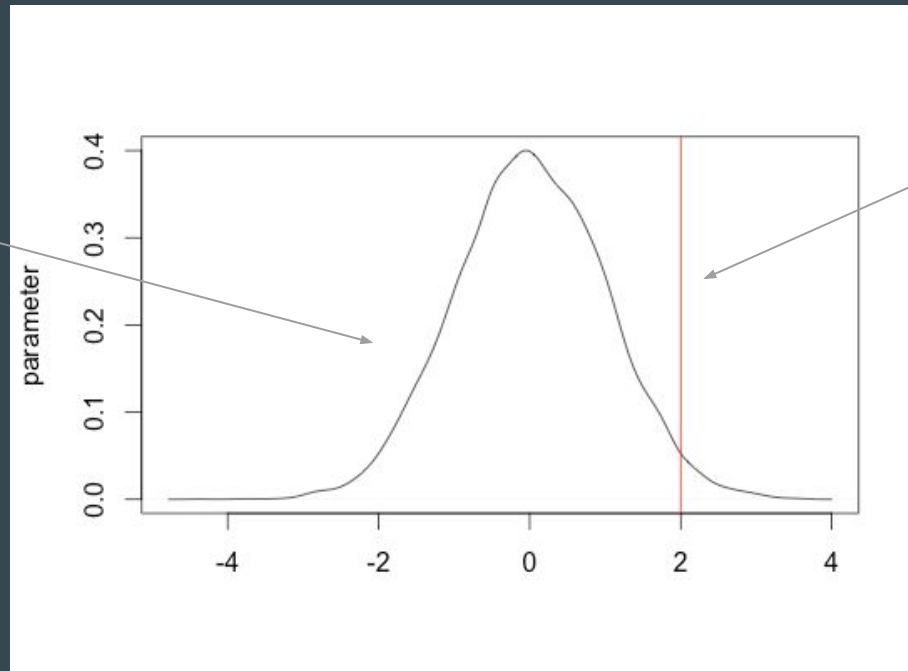
Network enrichment (SNOW)



Network enrichment (SNOW)

Comparison against N random networks (of similar size)

distribution of random
networks



value of input network

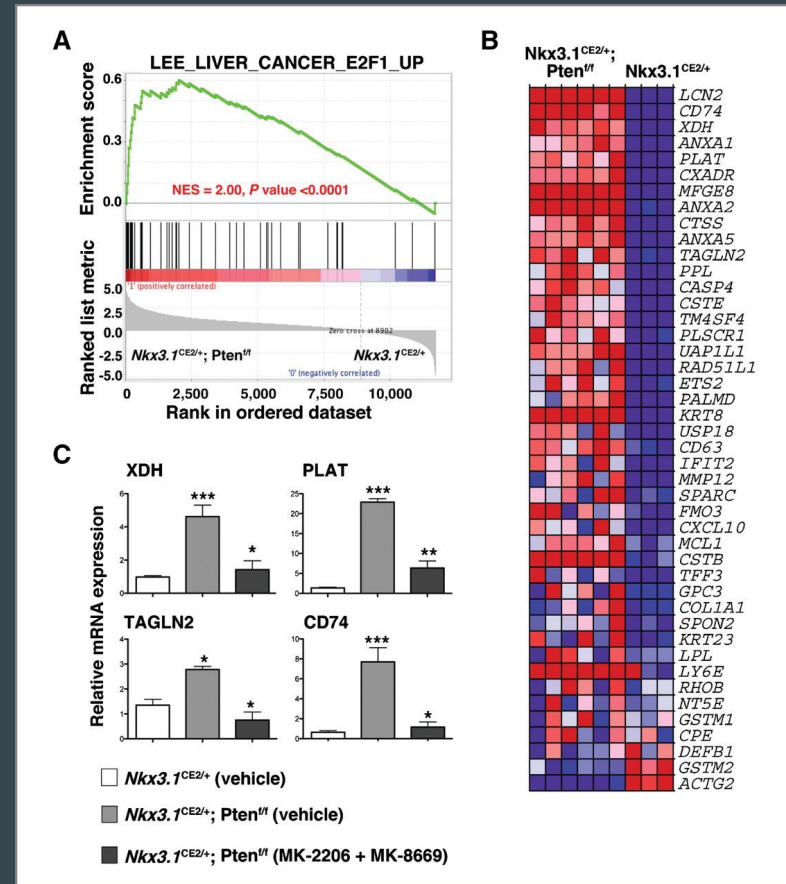
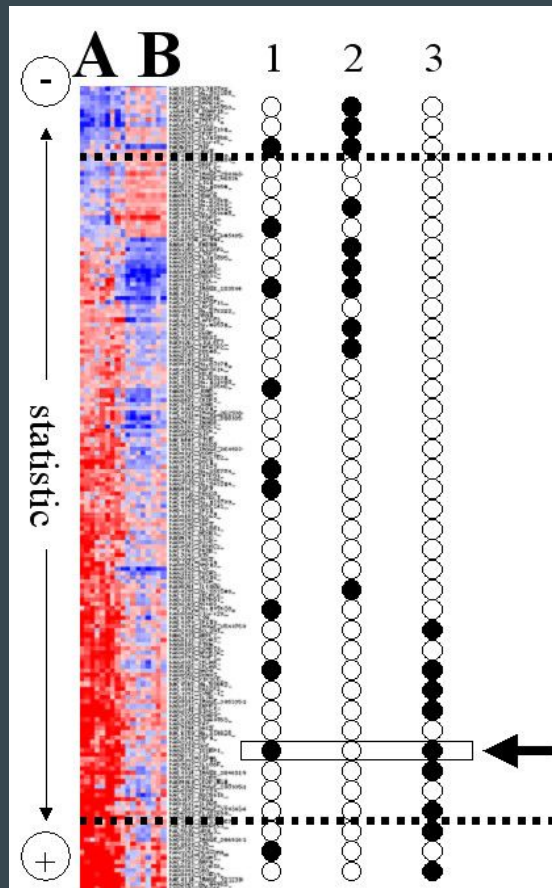
Some exercises

Worked examples of SNOW in Babelomics 5 wiki


[https://github.com/babelomics/babelomics/wiki/network-enrichment-\(snow\)](https://github.com/babelomics/babelomics/wiki/network-enrichment-(snow))




Gene set network enrichment (Network miner)

- Gene set enrichment







Gene set network enrichment (Network miner)

 Babelomics 5


Processing ▼ Expression ▼ Genomics ▼ Cancer ▼ Functional genomics ▼ Upload  profile  jobs 

Network Miner

Examples

Essential genes in cancer cell line K562 	Genes Down-regulated in Fanconi Anemia 
Essential genes in cancer cell line JURKAT 	Genome-Wide Association Study in Bipolar Disorder 

Select your data


The files must be on the server to select them.
You can upload files using the button  inside file browser.

File browser

WorkSpace/

Select your seed list (optional)

☒ File ☐ Text area

The files must be on the server to select them.
You can upload files using the button  inside file browser.

File browser

WorkSpace/

List nature

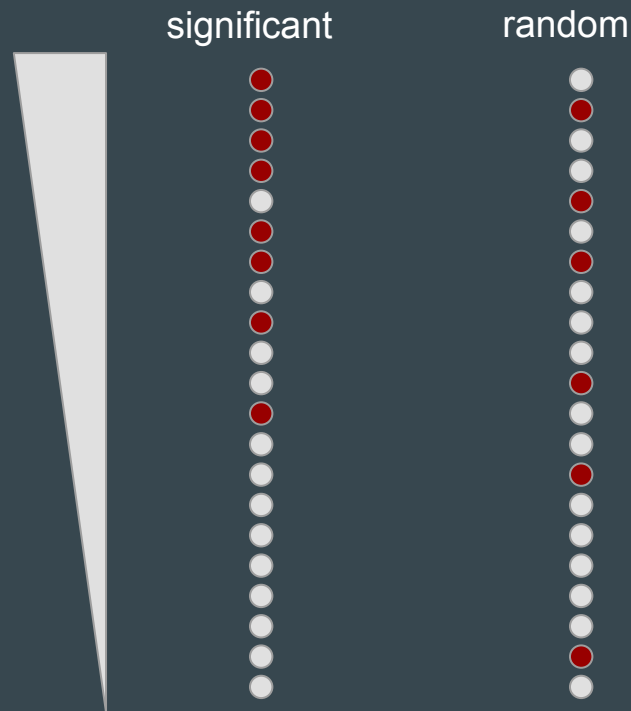
☐ Transcripts ☒ Proteins ☐ Genes

Species

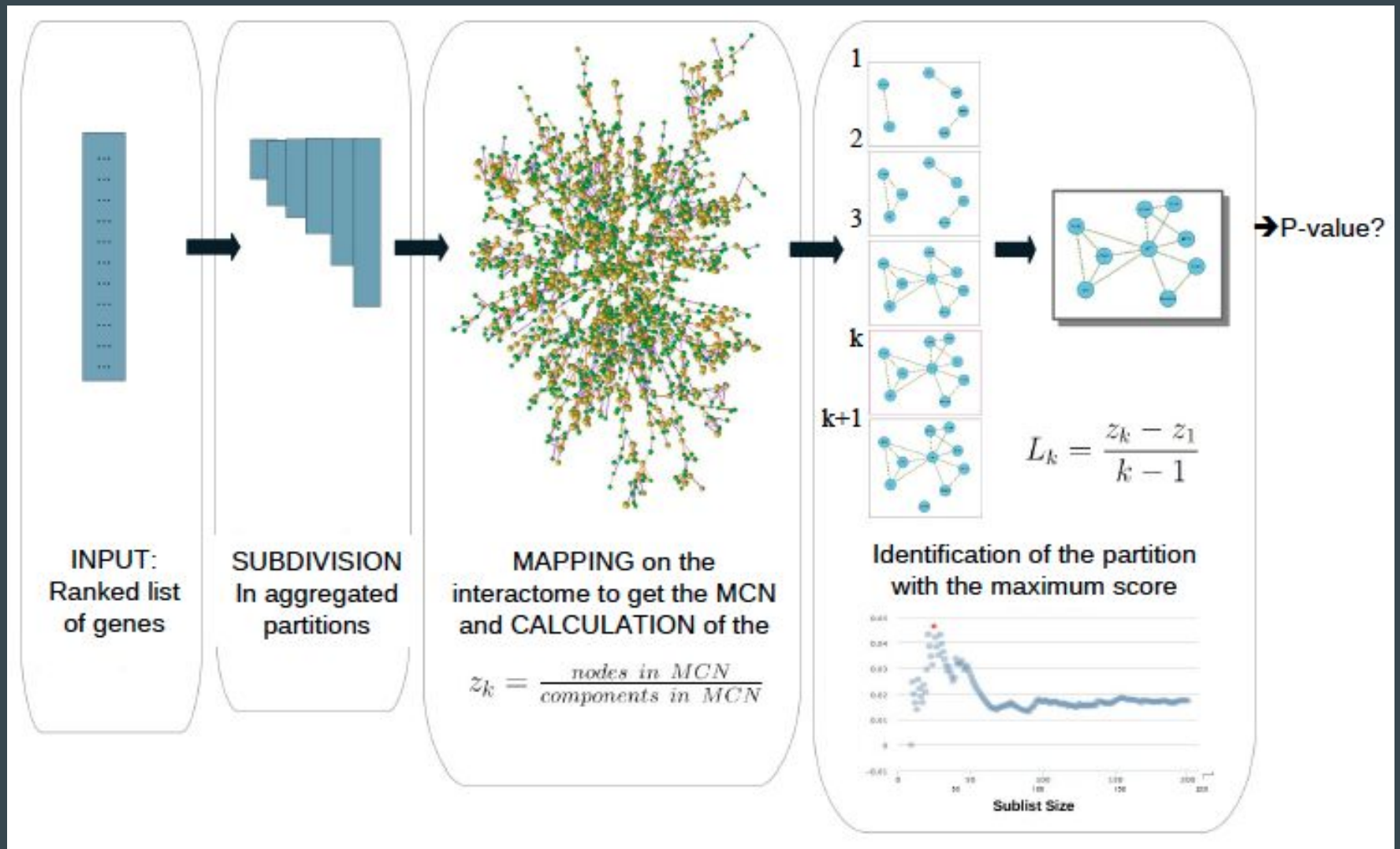
Homo sapiens ▼

Gene set network enrichment (Network miner)

- GOAL: Is there a latent sub-network related to ranking criteria?
- input is a ranked list (diff. expression, GWAS, ...)



Gene set network enrichment (Network miner)



Gene set network enrichment (Network miner)

Enrichment score

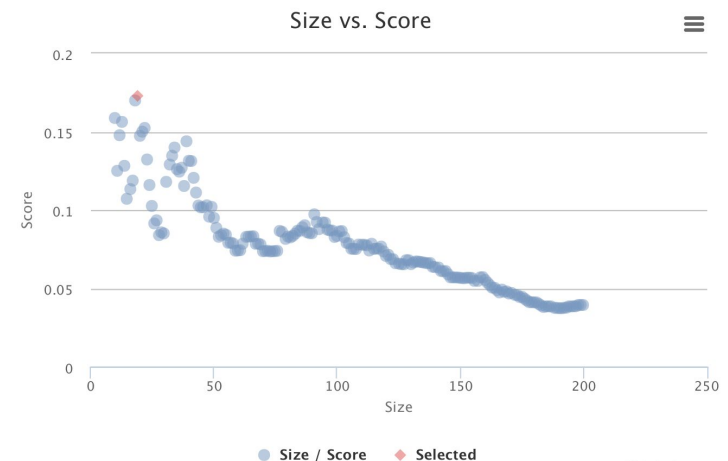
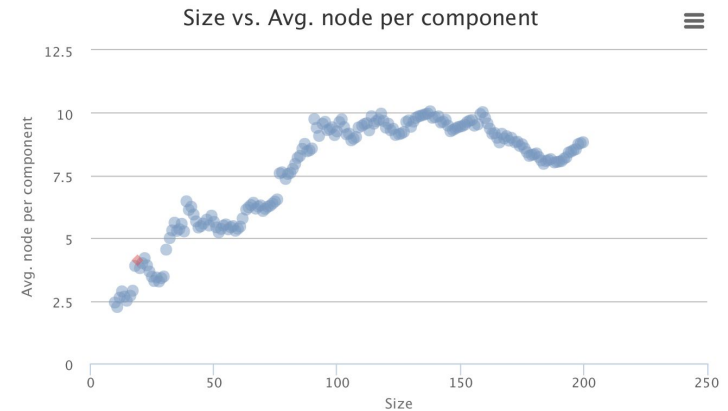
- First, ordering the parameter of interest z_k according to the ranked list, all relative maxima are identified. The partitions so selected (S_k^{max}) represent situations where a new protein capable of connecting to the previous ones is added to the previous partitions.
- Second, the score L_k is computed as $L_k = (z_k - 1)/(k - 1)$ for all the selected partitions S_k^{max} . The score can be seen as a balance between the increase in connected nodes and the distance to the top of the ranked list ($k = 1$).
- Third, we choose the partition S_{best} and index k_{best} corresponding to the highest L_k computed in b) form the S_k^{max} chosen in (a).

Results: Minimum Connected Network selected

All results

MCN selected pval: **<0.0035**

MCN selected size: **19**



All results [result_all.txt](#)

Some exercises...

Worked examples of NM on Babelomics 5 wiki

[https://github.com/babelomics/babelomics/wiki/Gene%20Set%20Network%20Enrichment%20\(Network%20Miner\)](https://github.com/babelomics/babelomics/wiki/Gene%20Set%20Network%20Enrichment%20(Network%20Miner))

Some conclusions

- Systems biology provides us a global view of our living system
- Networks are extremely useful to represent inner interactions
- Networks can be measured by different topological parameters

The background of the slide is a dark blue-grey color. Overlaid on this is a complex network graph. The graph consists of numerous small, dark grey circular nodes connected by thin, light grey lines representing edges. A specific path or set of connections is highlighted in a darker blue color. Along this highlighted path, several nodes are colored red, while the others remain dark grey. The overall structure of the graph is dense and interconnected, with the highlighted path weaving through various clusters of nodes.

Any question?