Interfaculty Master Program on Complex Systems and Networks

Aristotle University of Thessaloniki

The centrality – lethality rule in signed protein networks

Student

Savas Paragamian

Thessaloniki, May 2017

Commitee

Supervisor	Co-supervisor	Co-supervisor
Stefanos Sgardelis	Ioannis Antoniou	Christoforos Nikolaou
Prof. Biology AUTH	Prof. Mathematics AUTH	Ass. Prof. Biology UOC

The aim of this study is to extend the centrality – lethality rule to signed protein interaction networks in order to improve the prediction of essential genes / proteins.

Questions

1. Does the centrality – lethality rule holds in the signed protein interaction network of *D. melanogaster*?

2. What kinds of interactions are connecting essential proteins?

3. Is essentiality modular in *D. melanogaster*?

Gene essentiality

A gene/protein is essential if and only if its removal or disruption results in lethality or infertility of the organism

History

- Individual gene knock-out experiments

    - Lethal genes vs essential genes

- Laborious task

- Large scale gene essentiality screens in the 1990's

Tatum, E. L., & Lederberg, J. (1947). Gene Recombination in the Bacterium Escherichia coli. *Journal of Bacteriology*, *53*(6), 673–684.

Lu, L. J. (editor). (2015). *Gene Essentiality Methods and Protocols*. (L. J. Lu, Ed.). New York: Springer Science+Business Media.

Winzeler, E. A. (1999). Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. *Science*, *285*(5429), 901–906.

Origins and conservation

- Lower mutation rates

- Metabolic processes

- Cell cycle processes

- Non essentiality and redundancy

Jordan, I. K., Rogozin, I. B., Wolf, Y. I., & Koonin, E. V. (2002). Essential Genes Are More Evolutionarily Conserved Than Are Nonessential Genes in Bacteria. *Genome Research*, *12*, 962–968.

Zhang, Z., & Ren, Q. (2015). Why are essential genes essential? - The essentiality of Saccharomyces genes. *Microbial Cell*, *2*(8), 280–287.

Research interests

- Minimal gene sets and synthetic biology

- Human disease genes (impossible essentiality experiments)

- Possible drug targets for pathogens (laborious experiments)

Hutchison, C. A., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., … Venter, J. C. (2016). Design and synthesis of a minimal bacterial genome. *Science*, *351*(6280), 6253–6253.

Current challenges of gene essentiality

1. Evolvability
2. Differential/ conditional essentiality
3. **Modular essentiality in protein complexes, in general modular biology**
4. **Prediction**

D'Elia, M. A., Pereira, M. P., & Brown, E. D. (2009). **Are essential genes really essential?** *Trends in Microbiology*, *17*(10), 433–438.

Liu, G., Yong, M. Y. J., Yurieva, M., Srinivasan, K. G., Liu, J., Lim, J. S. Y., … Rancati, G. (2015). **Gene Essentiality Is a Quantitative Property Linked to Cellular Evolvability**. *Cell*, *163*(6), 1388–1399.

Ideker, T., & Krogan, N. J. (2012). **Differential network biology**. *Molecular Systems Biology*, *8*(565), 1–9.

Hartwell, L. H., Hopfield, J. J., Leibler, S., & Murray, A. W. (1999). **From molecular to modular cell biology**. *Nature*, *402*(6761 Suppl), C47–C52.

Koch, C. (2012). **Modular Biological Complexity**. *Science*, *337*(6094), 531–532.

Prediction of gene essentiality

- Data availability
- 1990's: Sequencing
  - Comparable genomics (DNA sequence homology tools, e.g BLAST)
  - Mushegian, A. R., & Koonin, E. V. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(19), 10268–10273.
- 2000's: Highthrouput experiments
  - Metabolomics
  - Proteomics
  - Network science emergence
  - Centrality – lethality rule
  - Types of networks : Protein interactions, metabolic, transcription regulation, gene expression

Jeong, H., Mason, S. P., Barabási, a L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41–42.

Tools and R packages

- Working enviroment
  - R Studio
  - Knitr
  - Rmarkdown
  - bookdown

- Data cleansing and manipulation
  - readxl
  - readr
  - reshape2
  - dplyr
  - tidyr

- Graphics and visualization
  - ggplot2
  - ggraph
  - scales
  - gridExtra

- Network analysis
  - Igraph
  - Bipartite
  - Vegan
  - RBGL

- Gene Ontology
  - AnnotationDbi
  - org.Dm.eg.db
  - GO.db
  - topGO

- Analysis
  - Rpart
  - C50
  - ROCR: for evaluation
  - Entropy

Signed PPI network of *D. melanogaster*
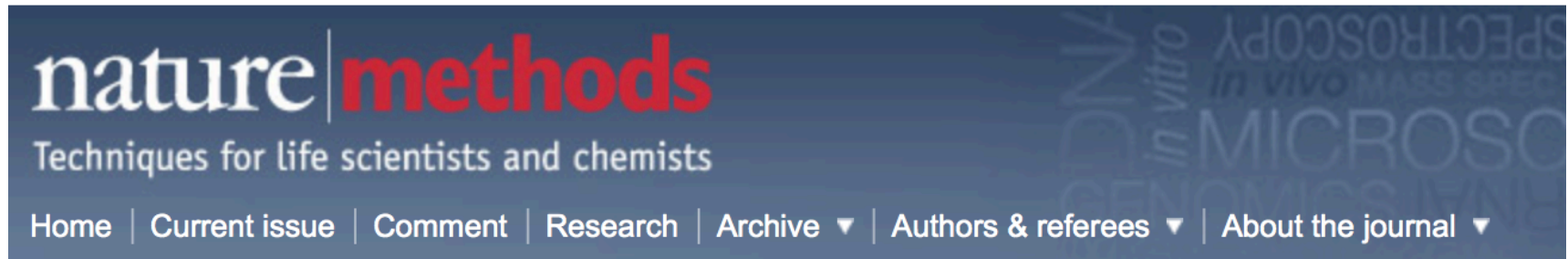
home ▸ archive ▸ issue ▸ article ▸ abstract

NATURE METHODS | ARTICLE

Integrating protein-protein interaction networks with phenotypes reveals signs of interactions

Arunachalam Vinayagam, Jonathan Zirin, Charles Roesel, Yanhui Hu, Bahar Yilmazel, Anastasia A Samsonova, Ralph A Neumüller, Stephanie E Mohr & **Norbert Perrimon**

Affiliations | **Contributions** | **Corresponding authors**

Signed PPI network of *D. melanogaster*

- **GOAL: enhance Protein interactions with activation/ inhibition information**
- From Protein – Protein interactions experiments
  - Yeast 2 hybrid
  - coAP-MS
  - Why directed?
- Interaction Databases
  - Biogrid, DroID, IntAct
- RNAi experiments
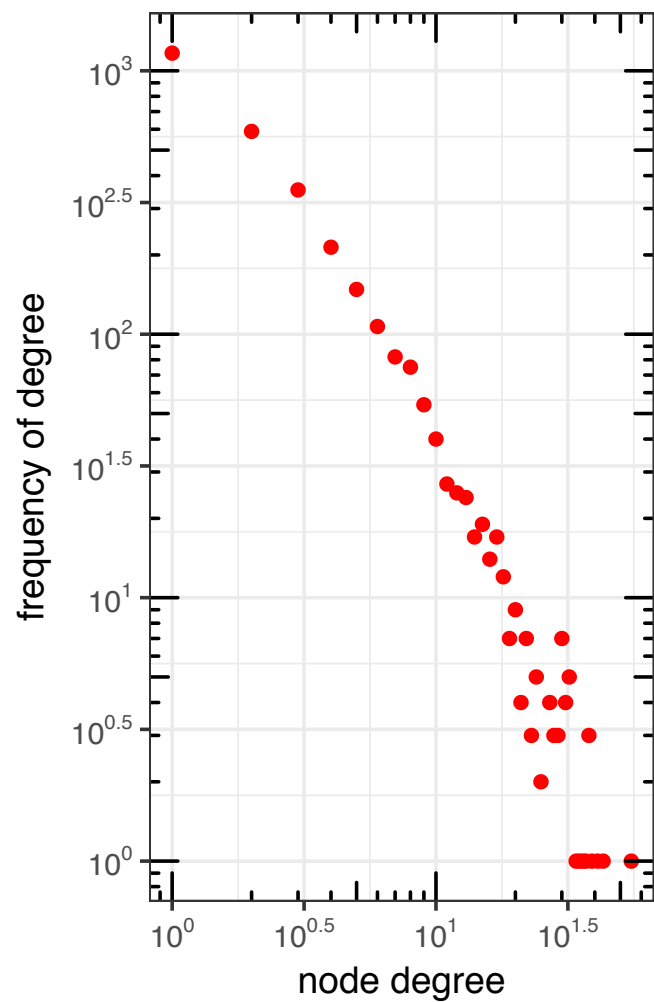 - Collection of phenotypes

- For each PI calculation of :

$$Sign\ score = \frac{Positive\ phenotypes\ - Negative\ phenotypes}{Total}\sqrt{Total}$$

- Apply threshold
- **Added value: Predicted relationships the were experimentally confirmed**

Signed PPI network of *D. melanogaster*
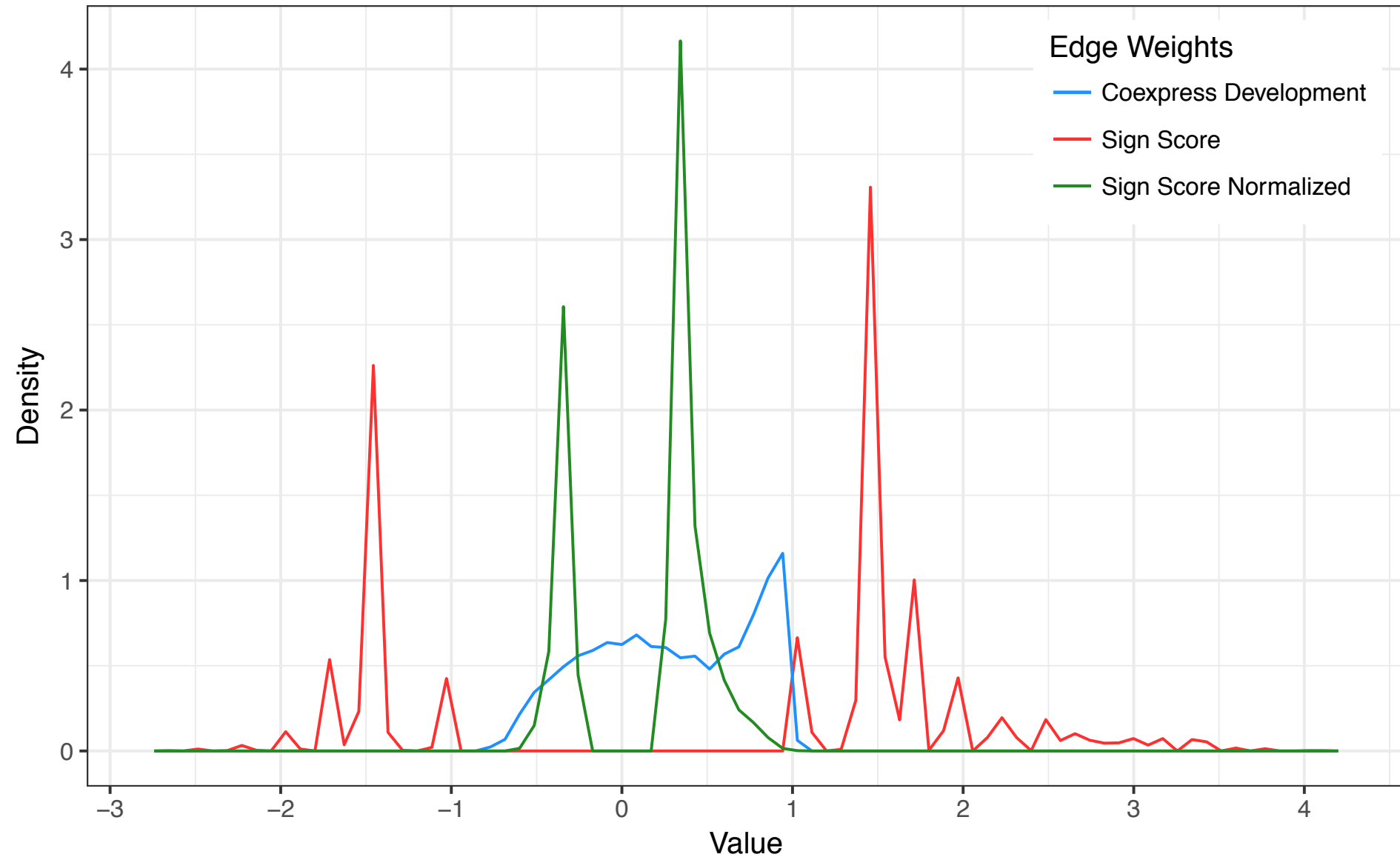
Signed PPI network of *D. melanogaster*

Type	All PI of *D.melanogaster*	Signed PI of *D.melanogaster*	Giant component
Proteins	9107	3352	3058
Interactions	47293	6094	5930
Positive	0	4109	3998
Negative	0	1985	1932

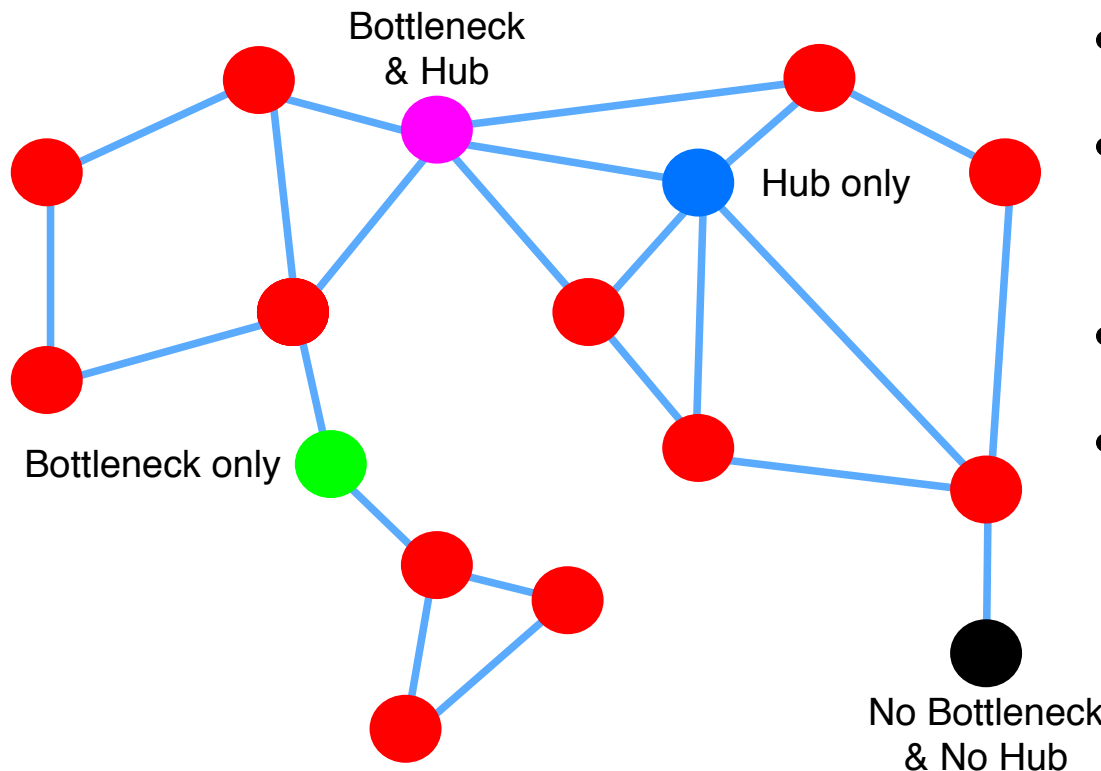# OGEE: an online gene essentiality database 🔓

Wei-Hua Chen, Pablo Minguez, Martin J. Lercher, Peer Bork ✉

Consensus	All *D.melanogaster*	Complete network	Giant component
Nonessential	13373	3009	2737
Essential	267	154 + 33 conditional	146 + 29 conditional
Conditional	141	33	29
NA	0	156	146
Total	13781	3352	3058

Centralities : Network topology

- High Degree : Hubs
- High Betweenness : bottlenecks
- They are correlated...
- New centralities are constantly introduced

Joy, M. P., Brock, A., Ingber, D. E., & Huang, S. (2005). High-betweenness proteins in the yeast protein interaction network. *Journal of Biomedicine and Biotechnology*, *2005*(2), 96–103.

Yu, H., Kim, P. M., Sprecher, E., Trifonov, V., & Gerstein, M. (2007). The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. *PLoS Computational Biology*, *3*(4), 713–720.

Other network approaches

- Use of other data: gene expression, metabolic interactions...

- Naïve Bayes

- Weighted k-nearest neighbors

- Logistic regression

- Decision trees

- Etc...

- Integration of methods : Best results!

Zhang, X., Acencio, M. L., & Lemke, N. (2016). Predicting essential genes and proteins based on machine learning and network topological features: A comprehensive review. *Frontiers in Physiology*, *7*(MAR), 1–11

Jalili, M., Salehzadeh-Yazdi, A., Gupta, S., Wolkenhauer, O., Yaghmaie, M., Resendis-Antonio, O., & Alimoghaddam, K. (2016). Evolution of Centrality Measurements for the Detection of Essential Proteins in Biological Networks. *Frontiers in Physiology*, *7*(August), 375.
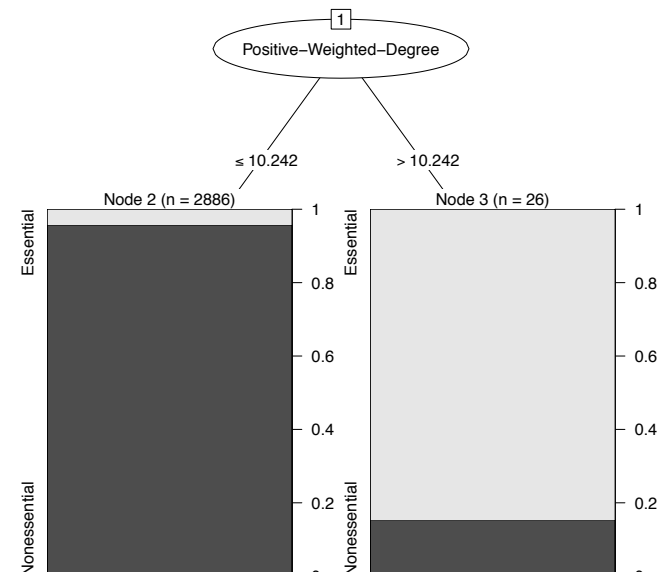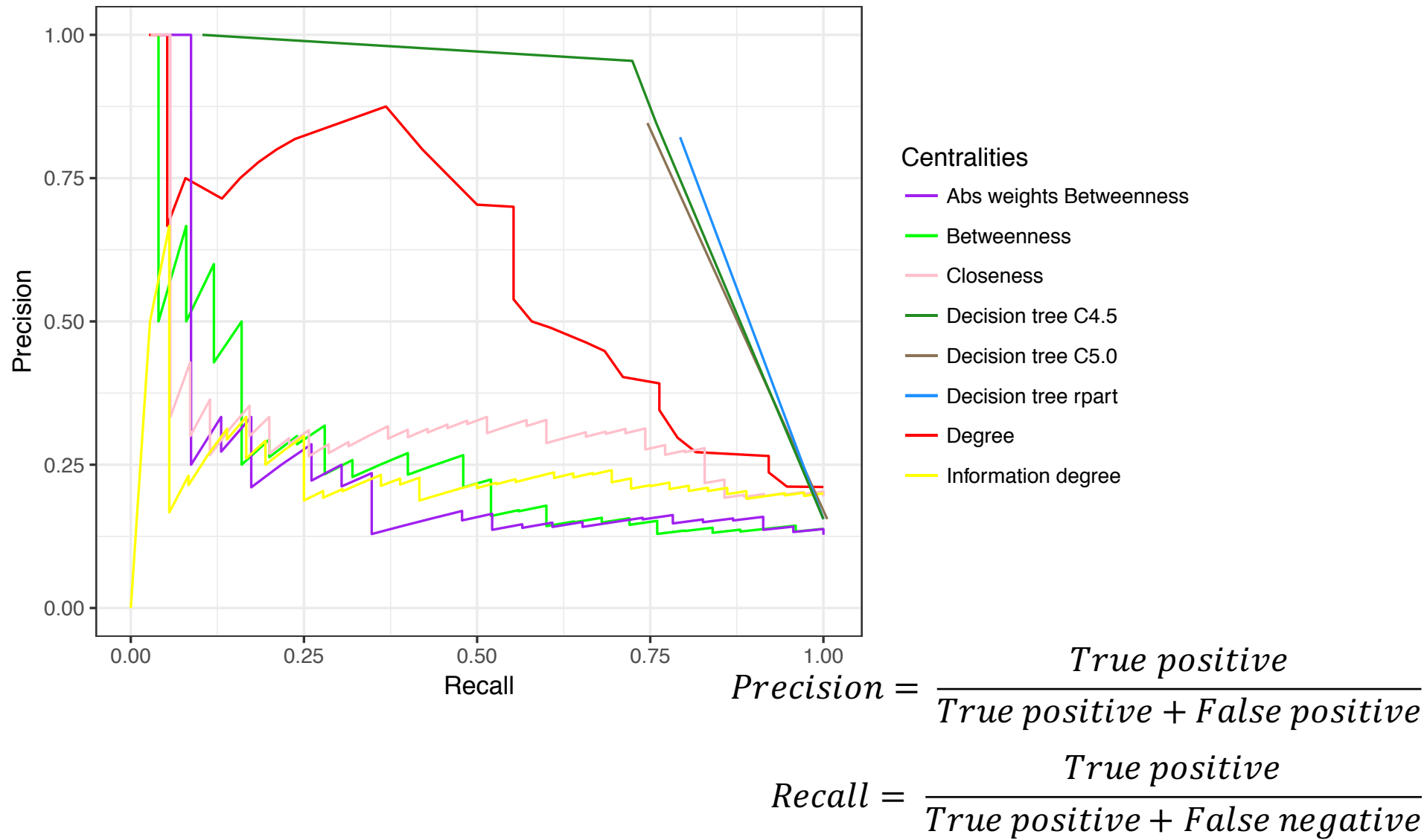
Decision trees
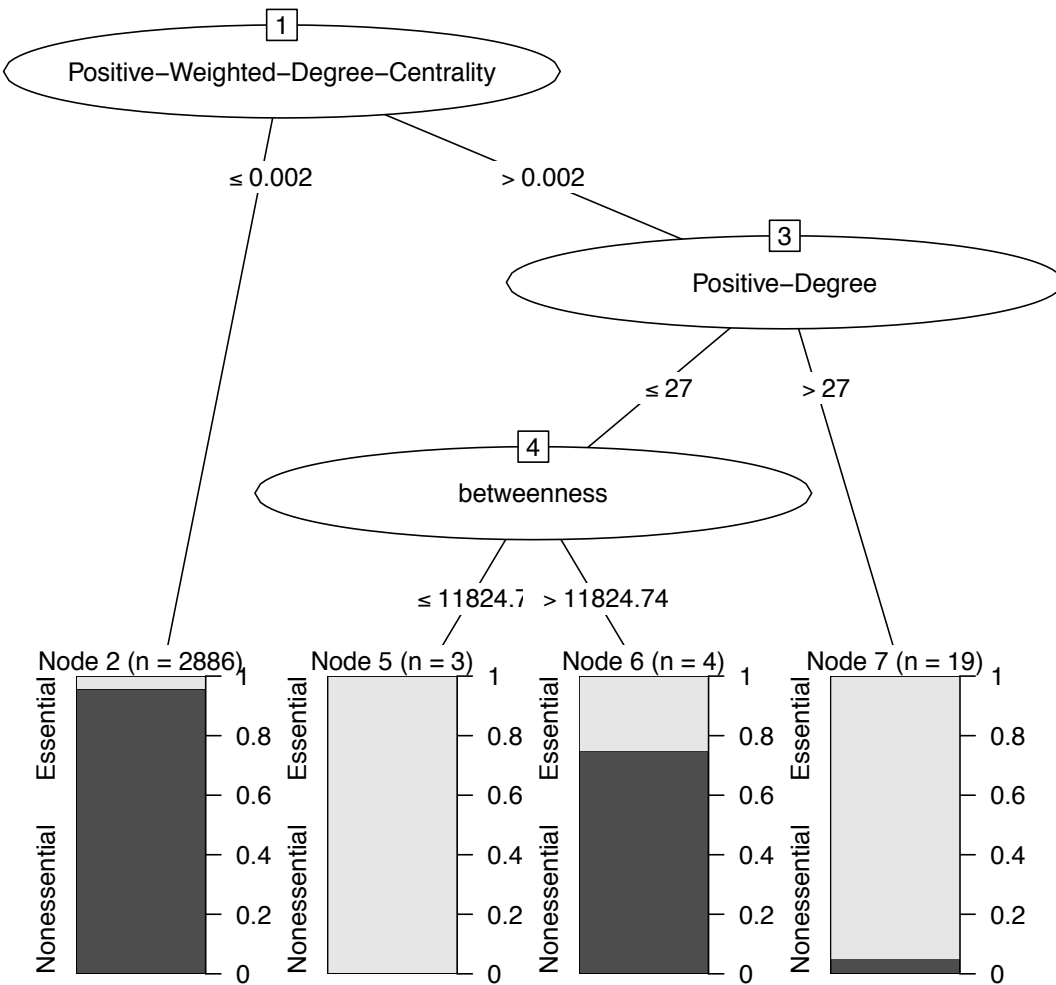
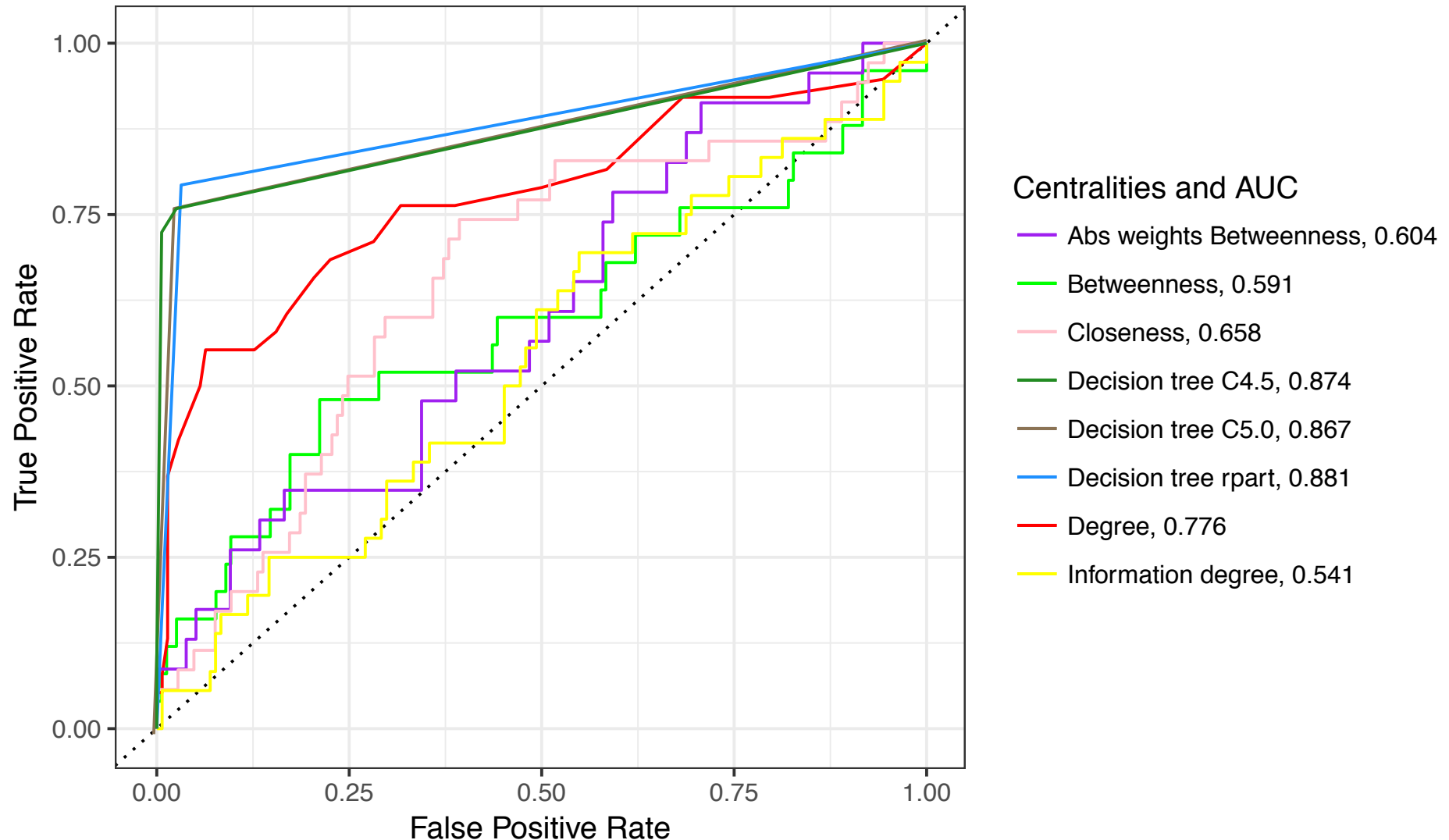| Predicted condition	C5.0	rpart	C4.5 RWeka
True positives	22	23	21
False negatives	124	123	125
False positives	4	5	1
True negatives	2762	2761	2765
Precision	0.846	0.821	0.955
Recall	0.151	0.158	0.144

C4.5 RWeka

C5.0

Centrality – Lethality rule : Results

$$Precision = \frac{True\ positive}{True\ positive + False\ positive}$$

$$Recall = \frac{True\ positive}{True\ positive + False\ negative}$$

Centrality – Lethality rule : Results

ROC Curve

Centralities and AUC

- Abs weights Betweenness, 0.604
- Betweenness, 0.591
- Closeness, 0.658
- Decision tree C4.5, 0.874
- Decision tree C5.0, 0.867
- Decision tree rpart, 0.881
- Degree, 0.776
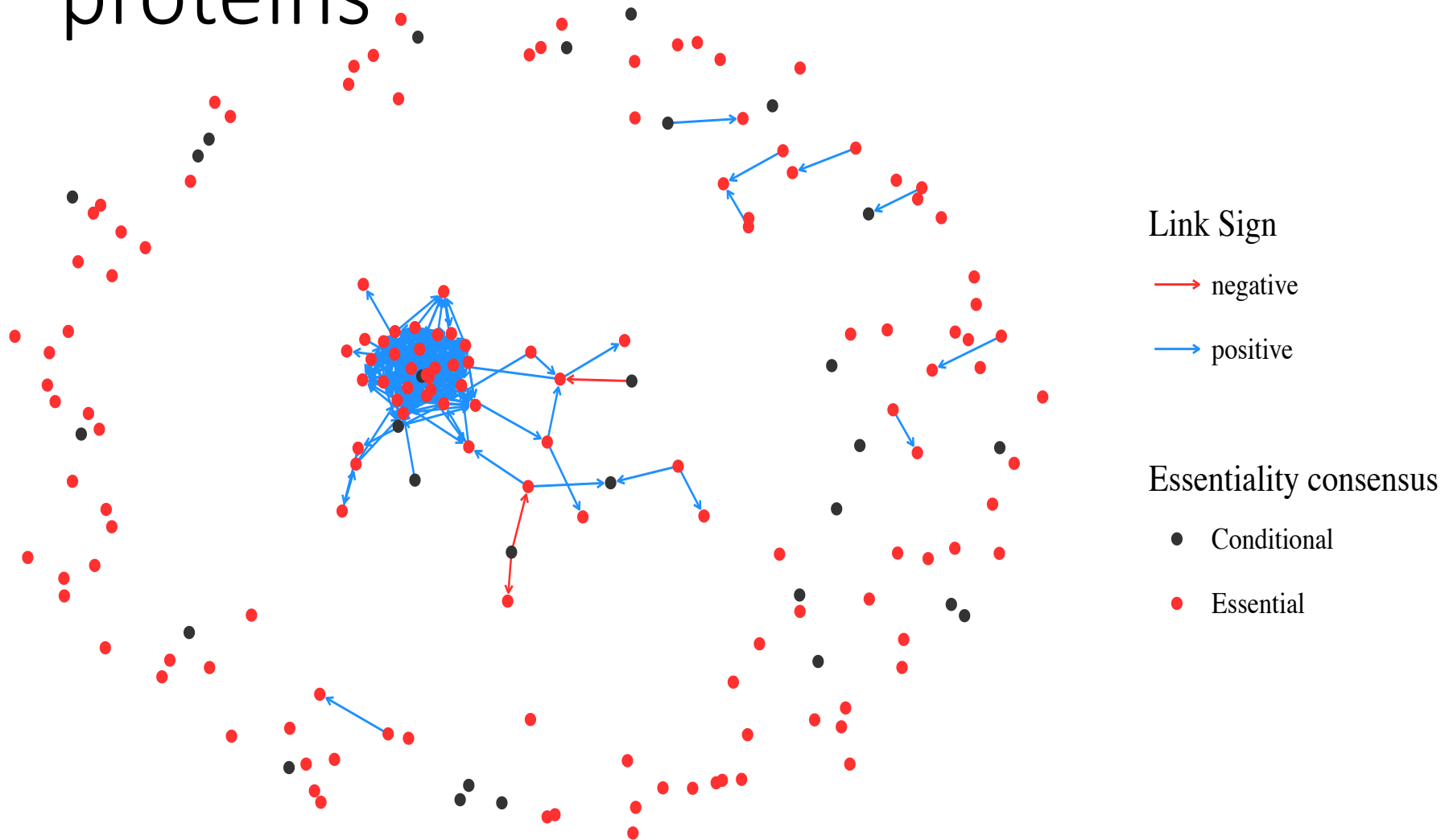- Information degree, 0.541

Centrality – Lethality rule : Results

Jackknife curves of essentiality prediction methods
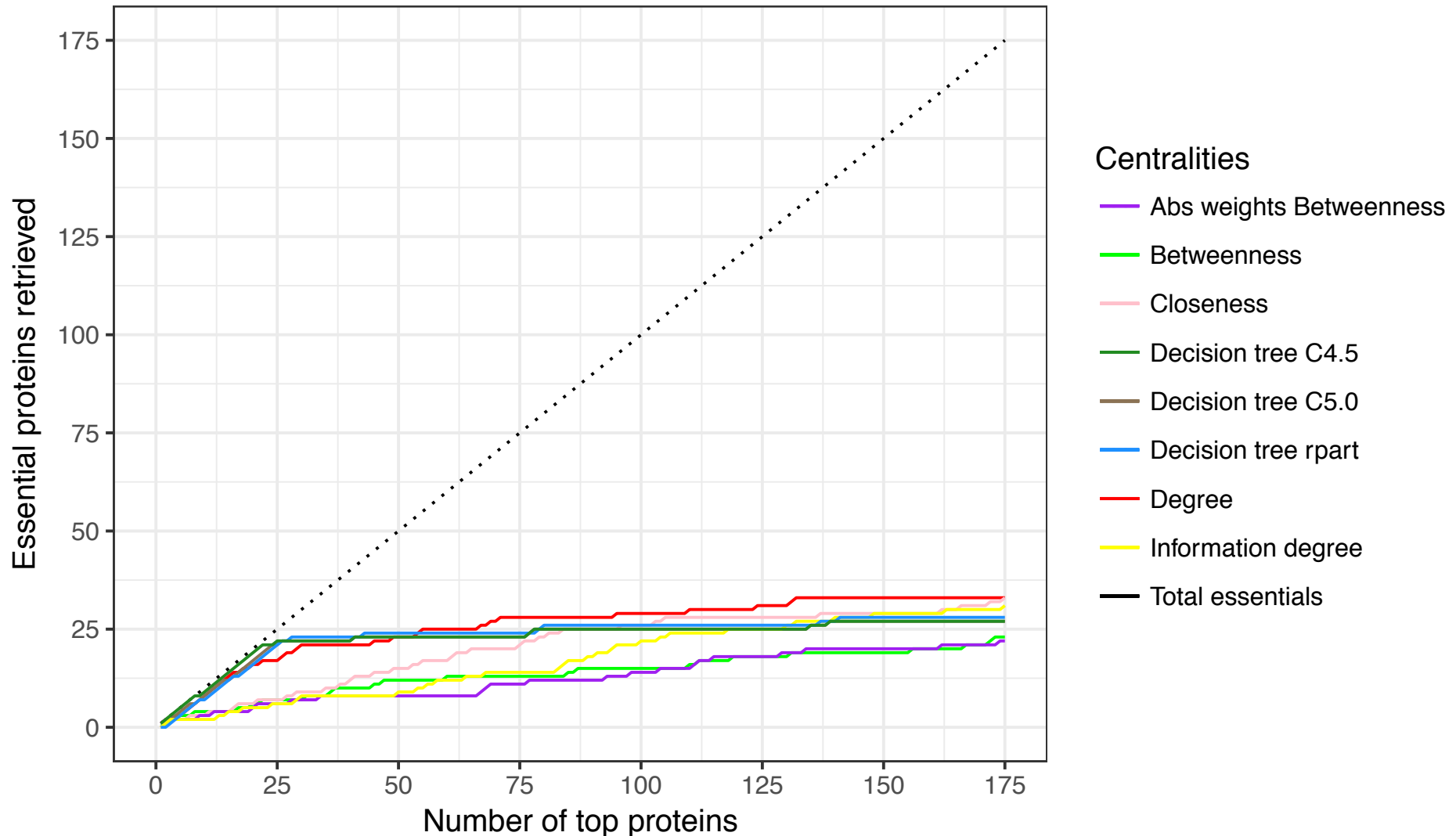
Centrality – Lethality rule : Challenges

- More data, more enriched data
  - Gene expression, activation/inhibition, phenotype/genotype, transcription regulation, metabolic processes

- Tools that can make use of that data
  - Best results -> Complex centrality : centrality index that incorporates protein complex information
  - Centralities for signed networks

- Is this an artifact of the extensive work on essential genes?

Jalili, M., Salehzadeh-Yazdi, A., Gupta, S., Wolkenhauer, O., Yaghmaie, M., Resendis-Antonio, O., & Alimoghaddam, K. (2016). Evolution of Centrality Measurements for the Detection of Essential Proteins in Biological Networks. *Frontiers in Physiology*, *7*(August), 375.

Interactions between essential proteins

Link Sign

→ negative

→ positive

Essentiality consensus

• Conditional

• Essential

Zotenko, E., Mestre, J., O'Leary, D. P., & Przytycka, T. M. (2008). Why do hubs in the yeast protein interaction network tend to be essential: Reexamining the connection between the network topology and essentiality. *PLoS Computational Biology*, *4*(8).

Perron – Frobenius decomposition

- **Definition**: For $n \geq 2$, an *n* x *n* complex matrix *A* is reducible if there exists an *n* x *n* permutation matrix *P* such that
$$PAP^{-1} = \begin{bmatrix} A_{1,1} & A_{1,r} \\ 0 & A_{r,r} \end{bmatrix}$$

- **Definition:** A directed graph with *n* nodes is strongly connected if, for any ordered pair *(Pi,Pj)* of nodes, with $1 \leq i, j \leq n,$ there exists a directed path

- **Theorem:** *An n x n complex matrix A is irreducible if and only if its directed graph G(A) is strongly connected.*

- **Theorem:** *Let $A \geq 0$ be an irreducible n x n matrix. Then, A has a positive real eigenvalue equal to its spectral radius.*

- **Definition:** Let $A \geq 0$ be an irreducible *n* x *n* matrix, and let *k* be the number of eigenvalues of *A* of modulus *p(A) (spectral radius).* If *k* = 1, then *A* is primitive. If *k* > 1, then *A* is cyclic of index *k.*

Tarjan, R. (1971). Depth-first search and linear graph algorithms. *12th Annual Symposium on Switching and Automata Theory (Swat 1971), 1*(2), 146–160.

Carey V, Long L and Gentleman R (2017). *RBGL: An interface to the BOOST graph library*. R package version 1.52.0, http://www.bioconductor.org.

Varga, R. S. (2000). *Matrix Iterative Analysis*. (H. Yserentant, R. Bank, R. L. Graham, J. Stoer, & R. Varga, Eds.) (2nd edition). Heidelberg: Springer-Verlag.

Perron – Frobenius decomposition

Components

- Outside component
- Strongly connected component
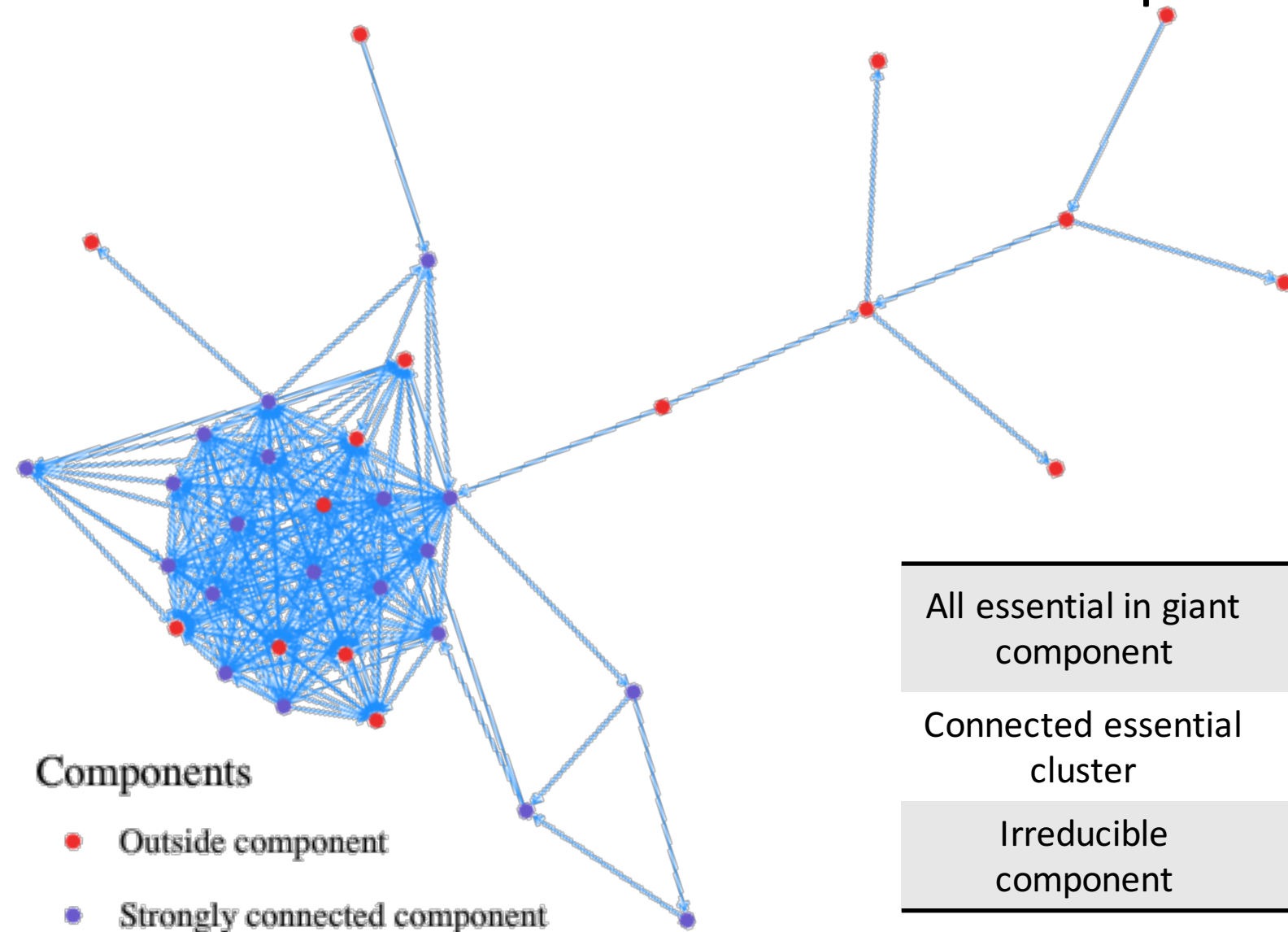
All essential in giant component	146 + 29 conditional
Connected essential cluster	36 proteins, 243 links
Irreducible component	20 proteins, 118 links

Perron – Frobenius decomposition
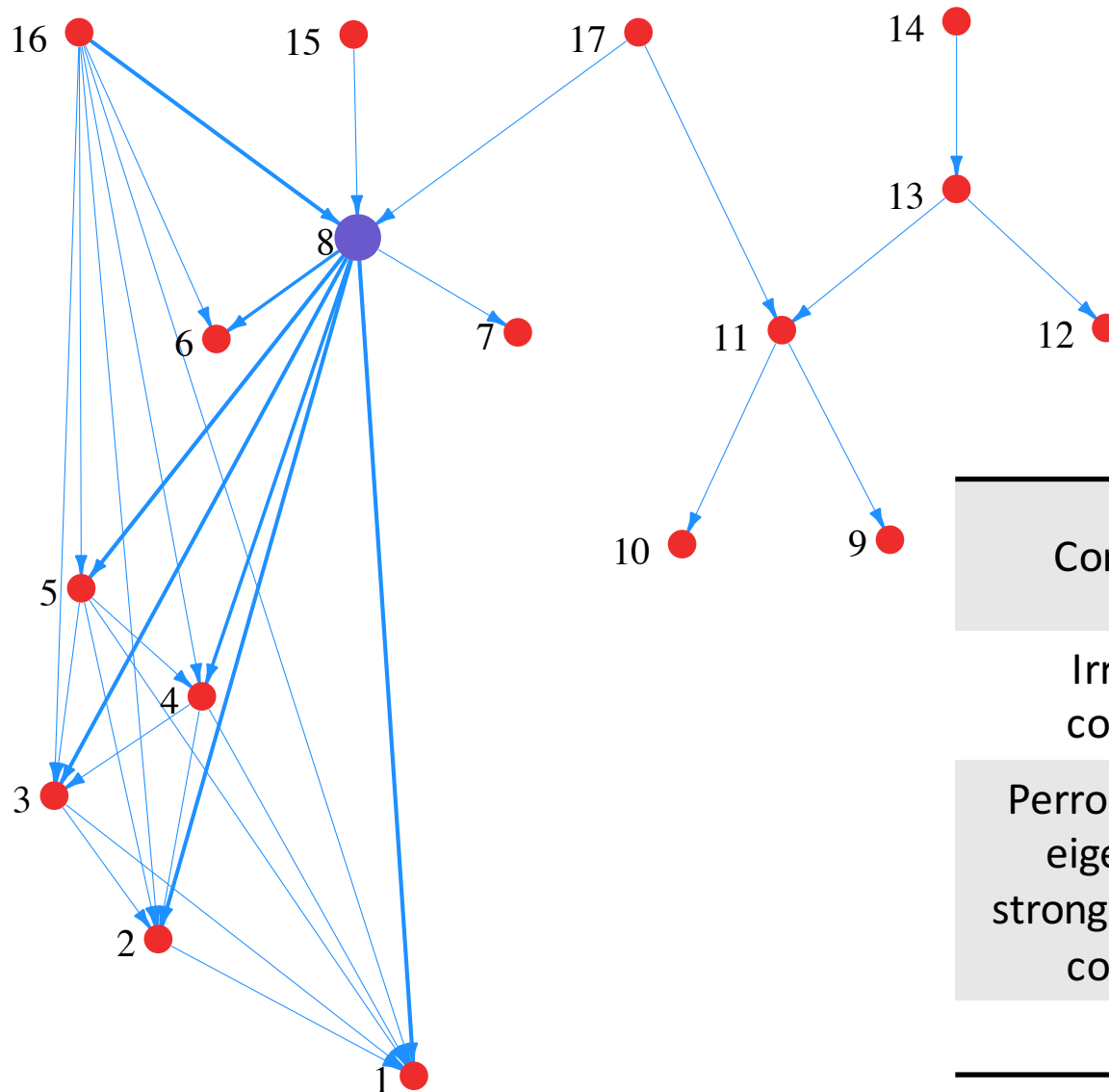
Components	17 (16 singular)
Irreducible component	20 proteins, 118 links
Perron–Frobenius eigenvalue of strongly connected component	4.0210
k	1

Singular enrichment analysis
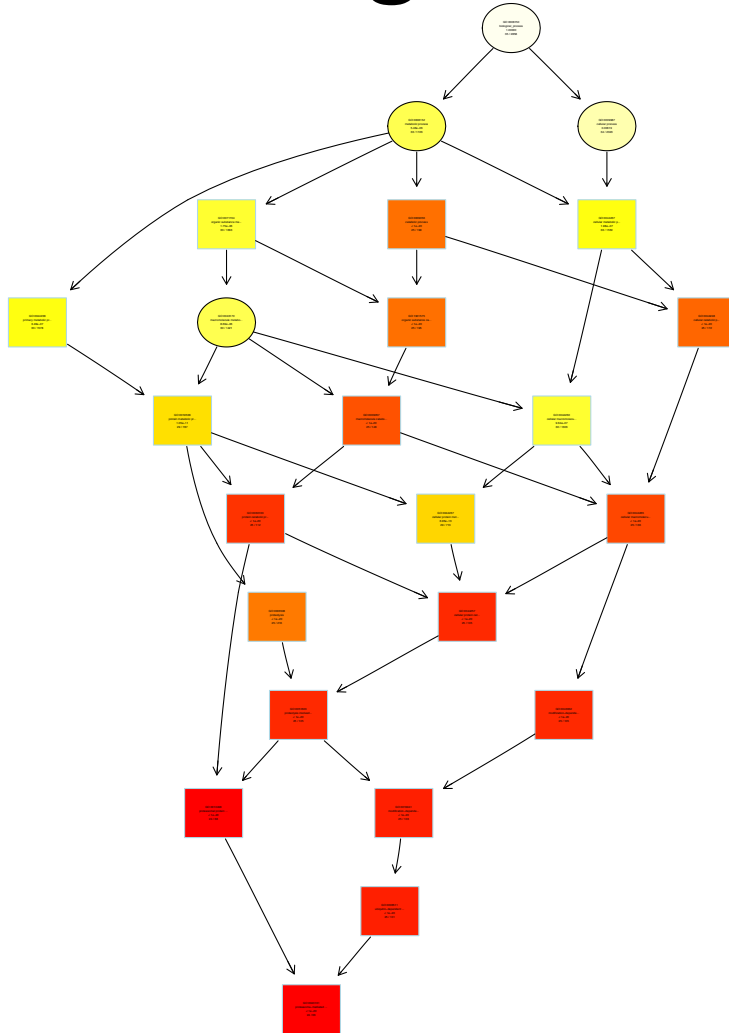
- Gene universe : Network proteins

- Fishers' exact test

- 58 significant GO terms
  - 23 with FDR
  - 21 with Bonferroni correction

- Controversy:
  - Genes probabilities are not equal
  - Heavy tail distribution of the genes' GO term membership

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29.

Adrian Alexa and Jorg Rahnenfuhrer (2016). topGO: Enrichment Analysis for Gene Ontology. R package version 2.26.0.

Glass, K., & Girvan, M. (2014). Annotation enrichment analysis: an alternative method for evaluating the functional properties of gene sets. *Scientific Reports*, *4*, 4191.

Functional enrichment analysis
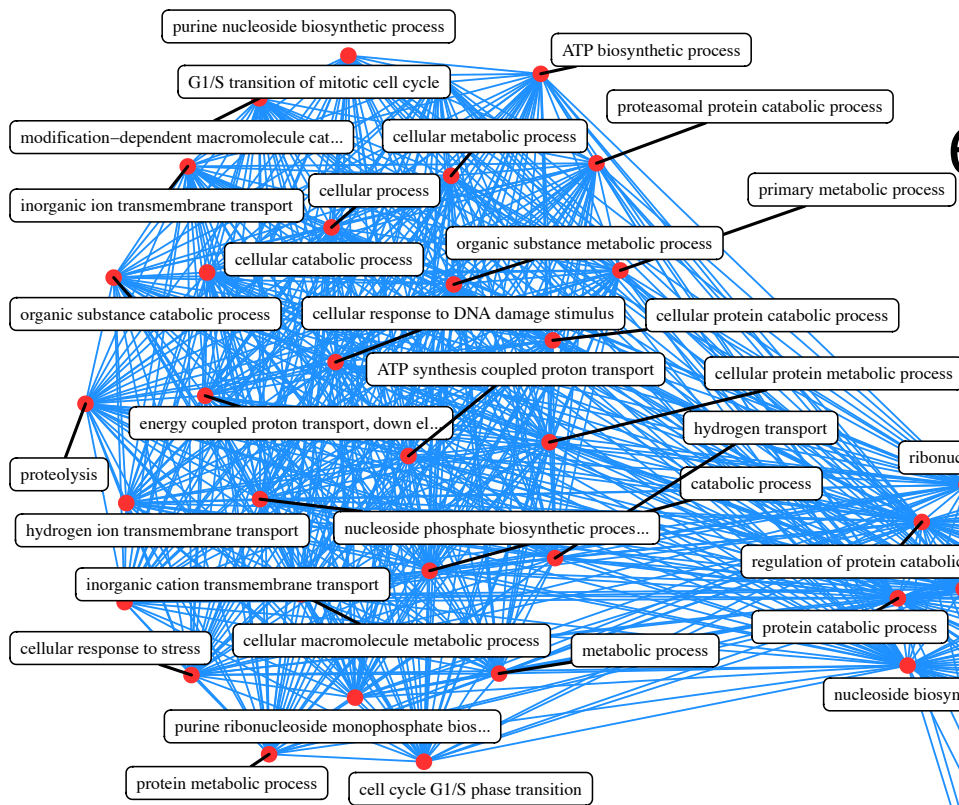
4 processes

- Protein catabolism
  - Proteasome subunits
  - ATPase
  - Ubiquitin action
  - Response to stress
- ATP biosynthesis
- Hydrogen membrane transport
- Cell cycle G1/S transition
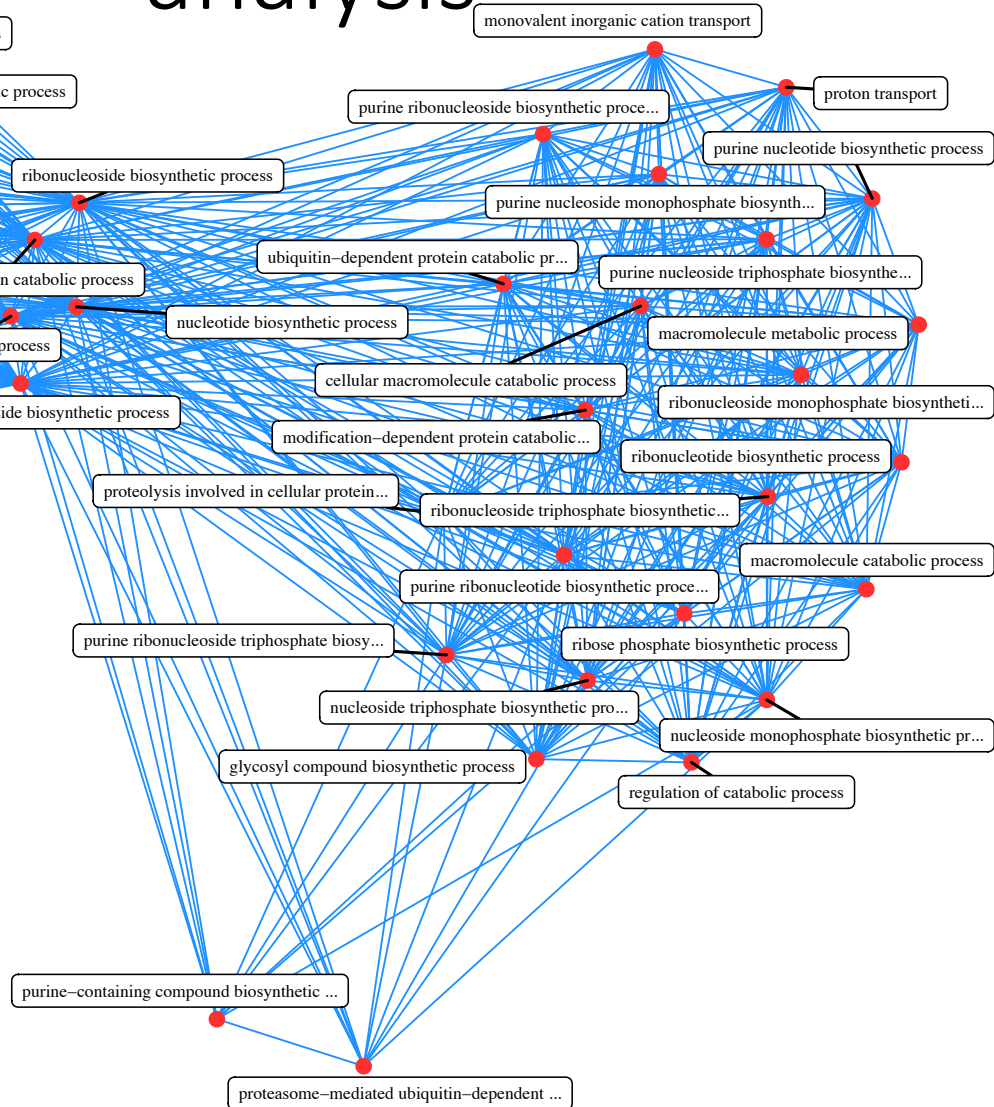
Cooperation – Synergy

- Essential proteins tend to form highly connected clusters rather than function independently (Zotenko et al. 2008).

- Network population evolution model showed that selection favors cooperation interactions (Jain and Krishna, 2001)

- Autocatalytic sets theory: An ACS is a subgraph, each of whose nodes has at least one incoming link from a node belonging to the same subgraph (S. A. Kauffman 1993)

- In (Solé 2011): Cooperation has been suggested as an essential step toward the emergence of complex, self-organized chemical systems (Eigen 1971)

Jain, S., & Krishna, S. (2001). A model for the emergence of cooperation, interdependence, and structure in evolving networks. *Pnas*, *98*(2), 543–7.
Kauffman, S. A. (1993). *The Origins of Order Self-Organization and Selection in Evolution*. New York: Oxford University Press.
Sole, R. V. (2011). *Phase Transitions*. Princeton: Princeton University Press.
Corning, P. A. (1995). Synergy and Self-Organization. *Systems Research*, *12*(2), 89–121.

Interactions between essential proteins

- Examine their signed interactions in other organisms

- Other ideas for analysis…

- Maybe there are missing inhibition interactions…

Modular essentiality

- Proteins form complexes to function

- Databases: Compleat, DPiM, CYT2008, MIPS

- What is the distribution of essential proteins in complexes?
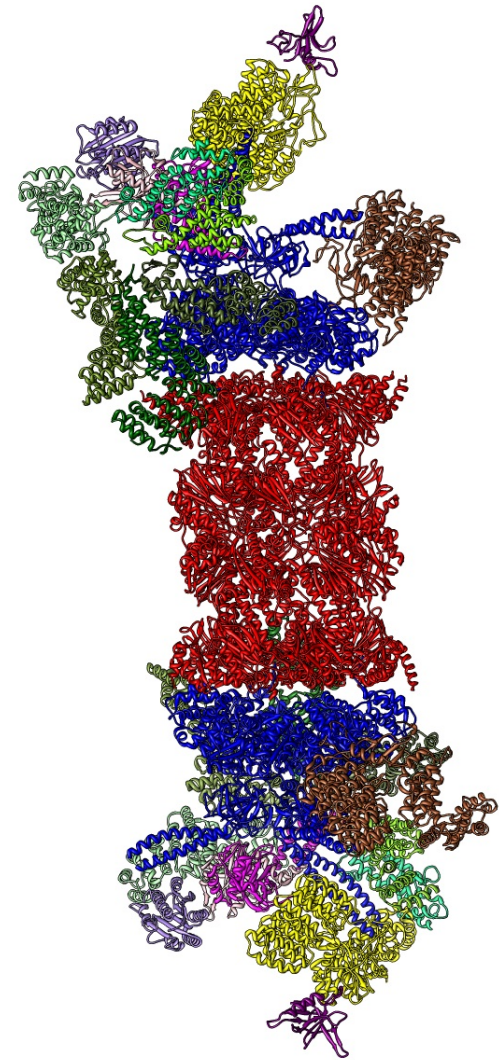
Pu, S., Wong, J., Turner, B., Cho, E., & Wodak, S. J. (2009). Up-to-date catalogues of yeast protein complexes. *Nucleic Acids Research*, *37*(3), 825–831.

Guruharsha, K G; Rual, JF; Zhai, B; Mintseris, J; Vaidya, N; Beekman, C; Wong, C; Rhee, DY; Cenaj, O; McKilli[, E; Shah, S; Spatleton, M; Wan, KH; Yu, CH; Artavanis-Tsakonas, S. (2011). A Protein Complex Network of Drosophila melanogaster. *Cell*, *147*(3), 690–703.

Mitra, K., Carvunis, A.-R., Ramesh, S. K., & Ideker, T. (2013). Integrative approaches for finding modular structure in biological networks. *Nature Reviews Genetics*, *14*(10), 719–732.

Protein complexes as communities

Non-overlapping : Modularity
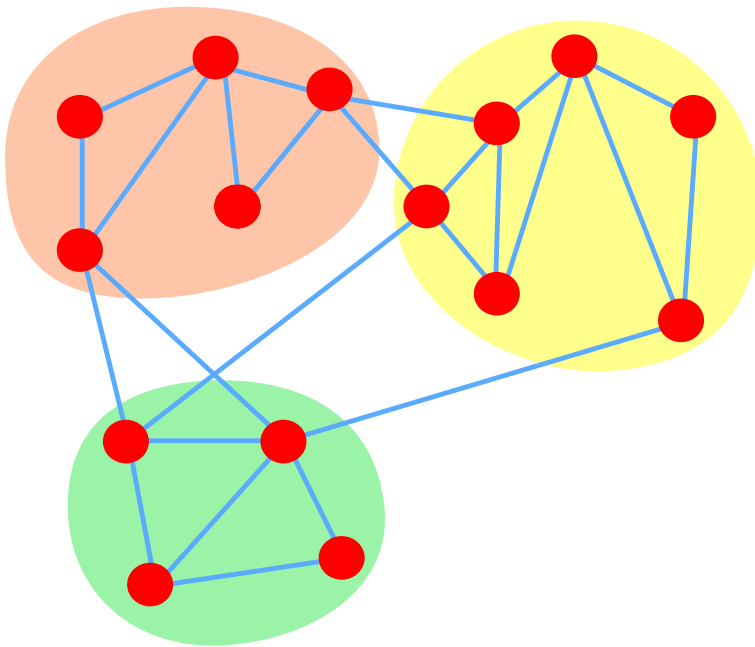
Overlapping : Link clustering
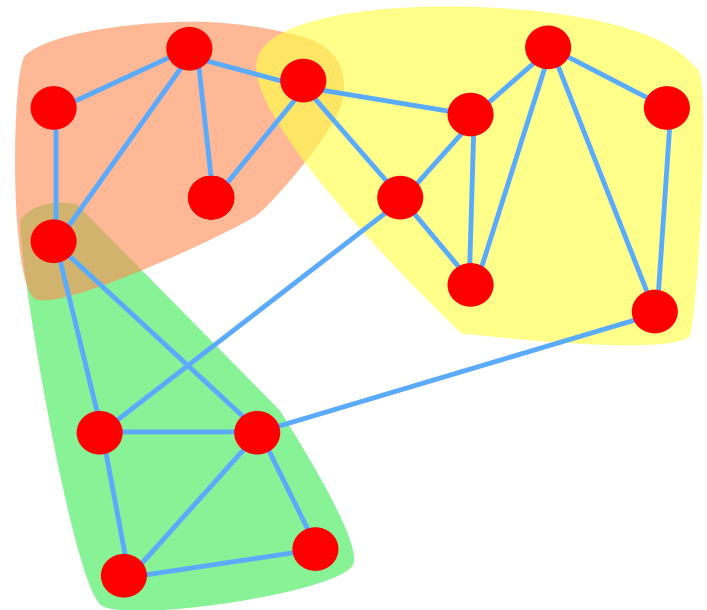
Ahn, Y.-Y., Bagrow, J. P., & Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. *Nature*, *466*(7307), 761–764.

Kalinka, A. T., & Tomancak, P. (2011). linkcomm: An R package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. *Bioinformatics*, *27*(14), 2011–2012.

Nepusz, T., Yu, H., & Paccanaro, A. (2012). Detecting overlapping protein complexes in protein-protein interaction networks. *NATURE METHODS*, *9*(5)

Ou-Yang, L., Dai, D. Q., & Zhang, X. F. (2015). Detecting Protein Complexes from Signed Protein-Protein Interaction Networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *12*(6), 1333–1344.

Compleat database
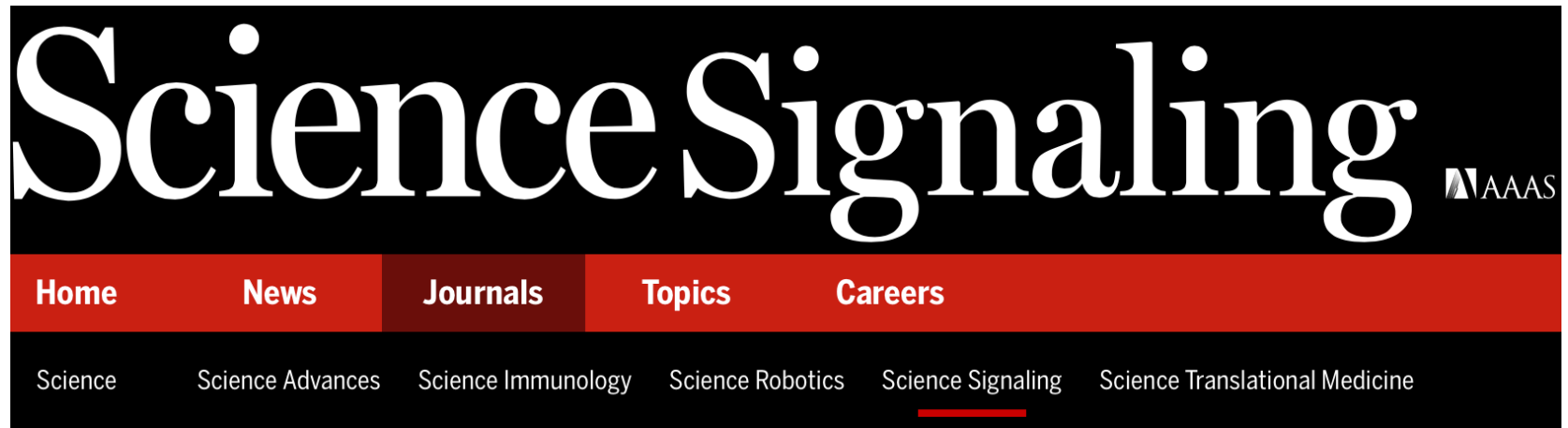
SHARE

RESEARCH RESOURCE | SYSTEMS BIOLOGY

Protein Complex–Based Analysis Framework for High-Throughput Data Sets

Arunachalam Vinayagam[1,*], Yanhui Hu[1,2,†], Meghana Kulkarni[1,††], Charles Roesel[2,3], Richelle Sopko[1], Stephanie E. Moh...

+ See all authors and affiliations

Article   Figures & Data   Info & Metrics   eLetters   📄 **PDF**

Compleat database distributions
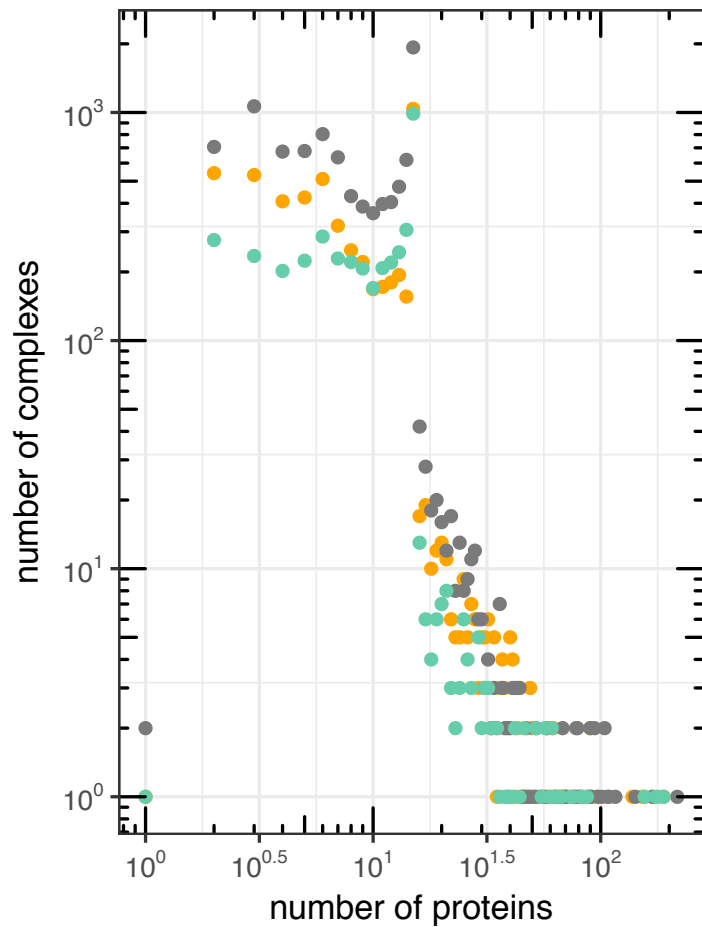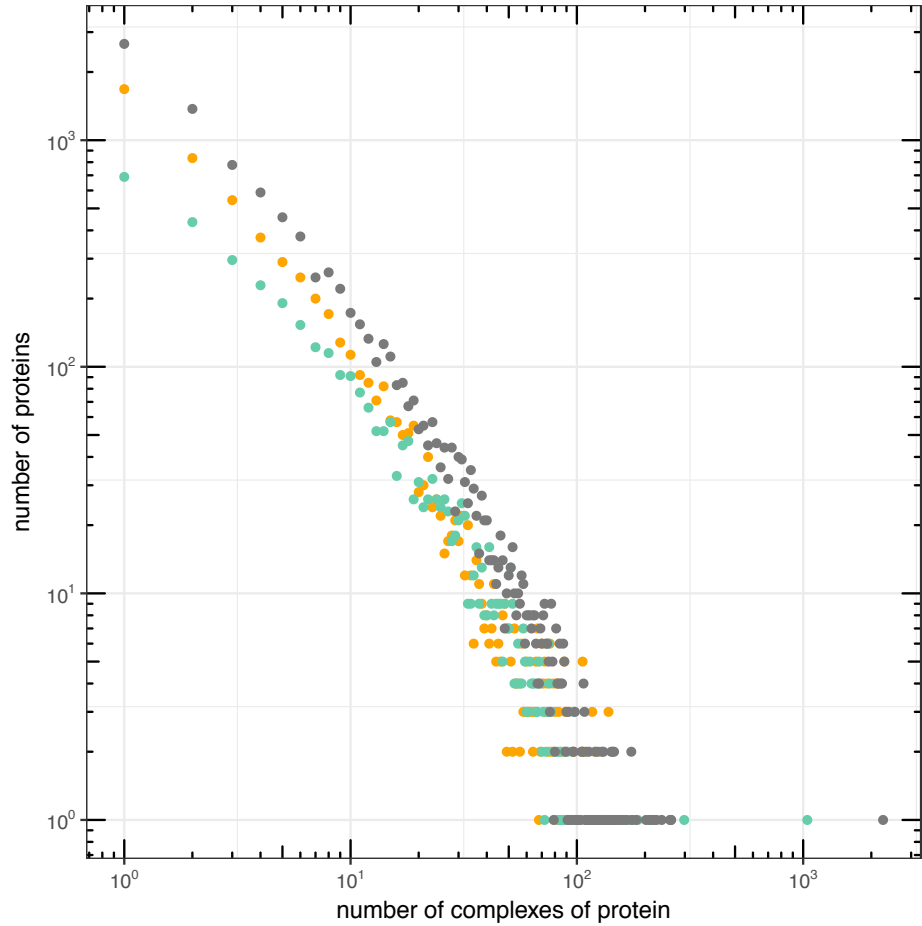
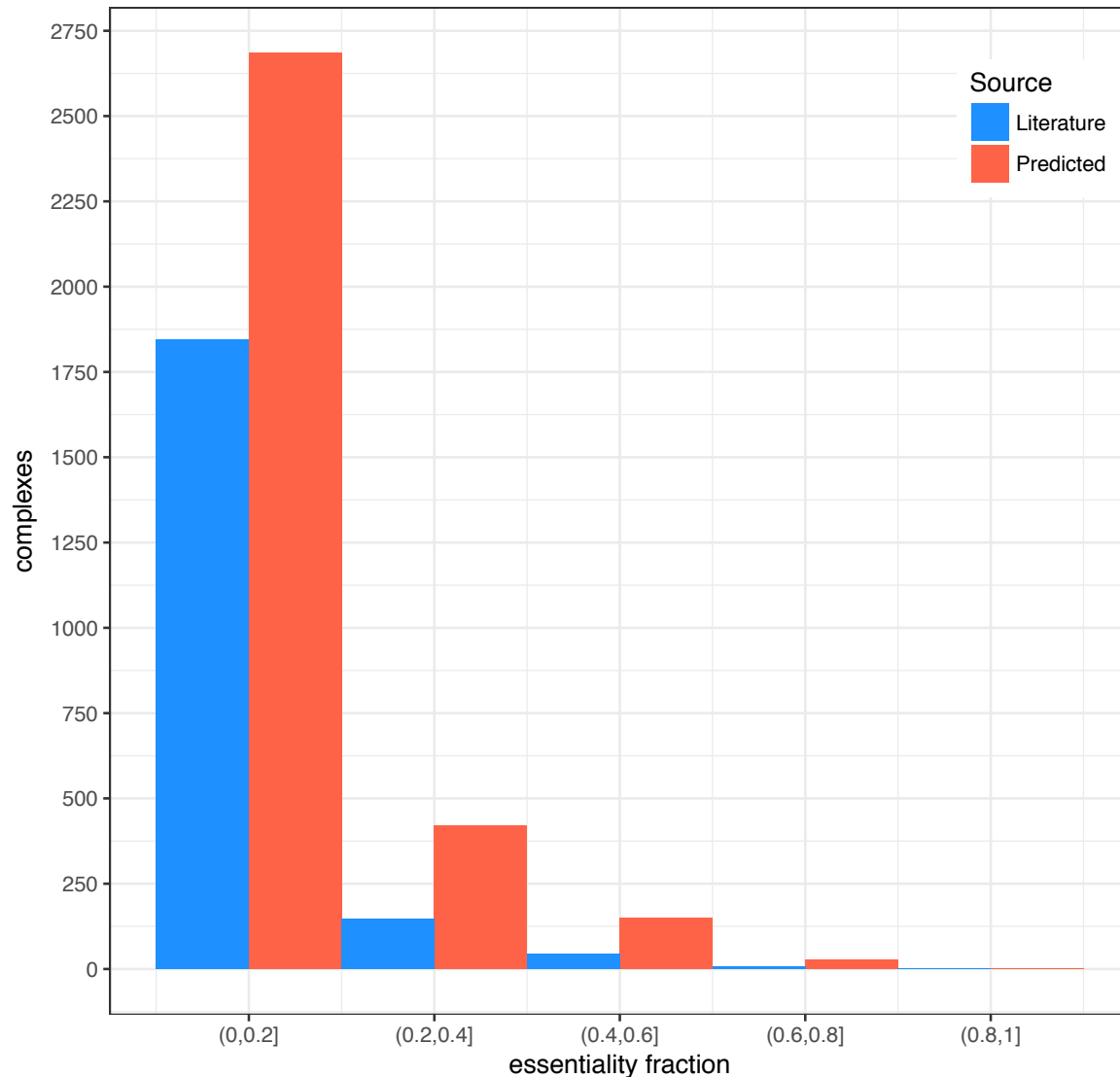# t database *D. melanogaster*
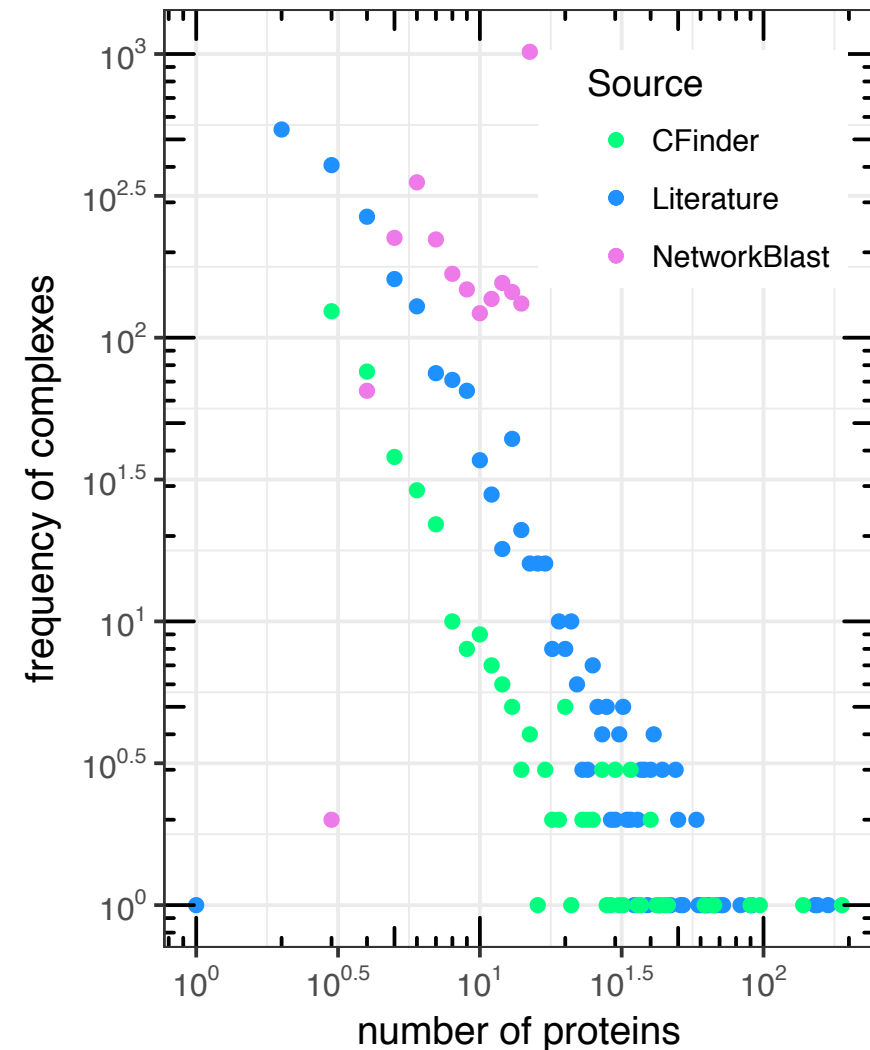
Source	Complexes	Proteins
Literature (326 distinct experiments)	2045	4501
NetworkBlast	2893	3525
CFinder	389	1362
Total	5327	5786

Bias due to authors' implementation
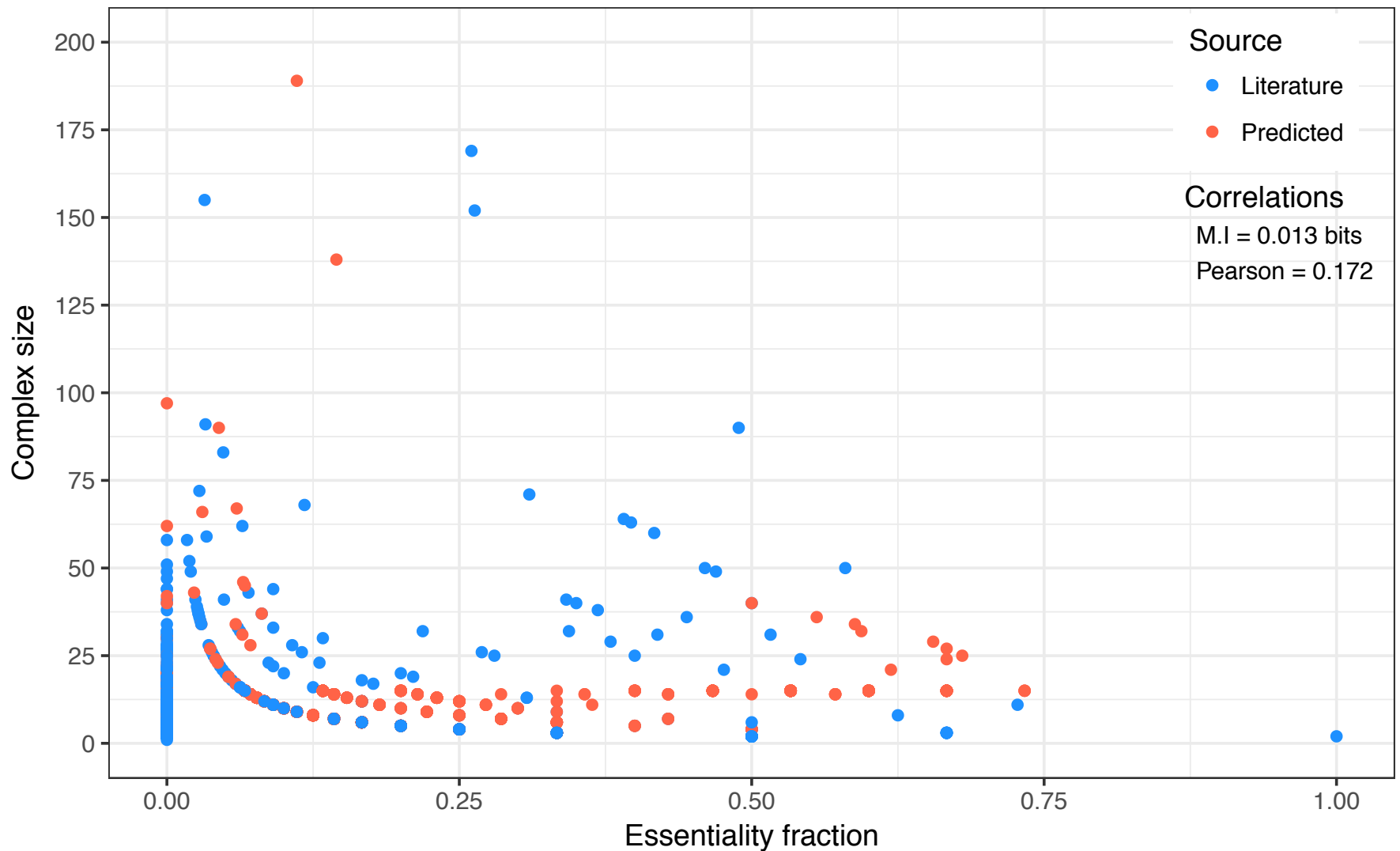or
due to NetworkBlast algorithm?
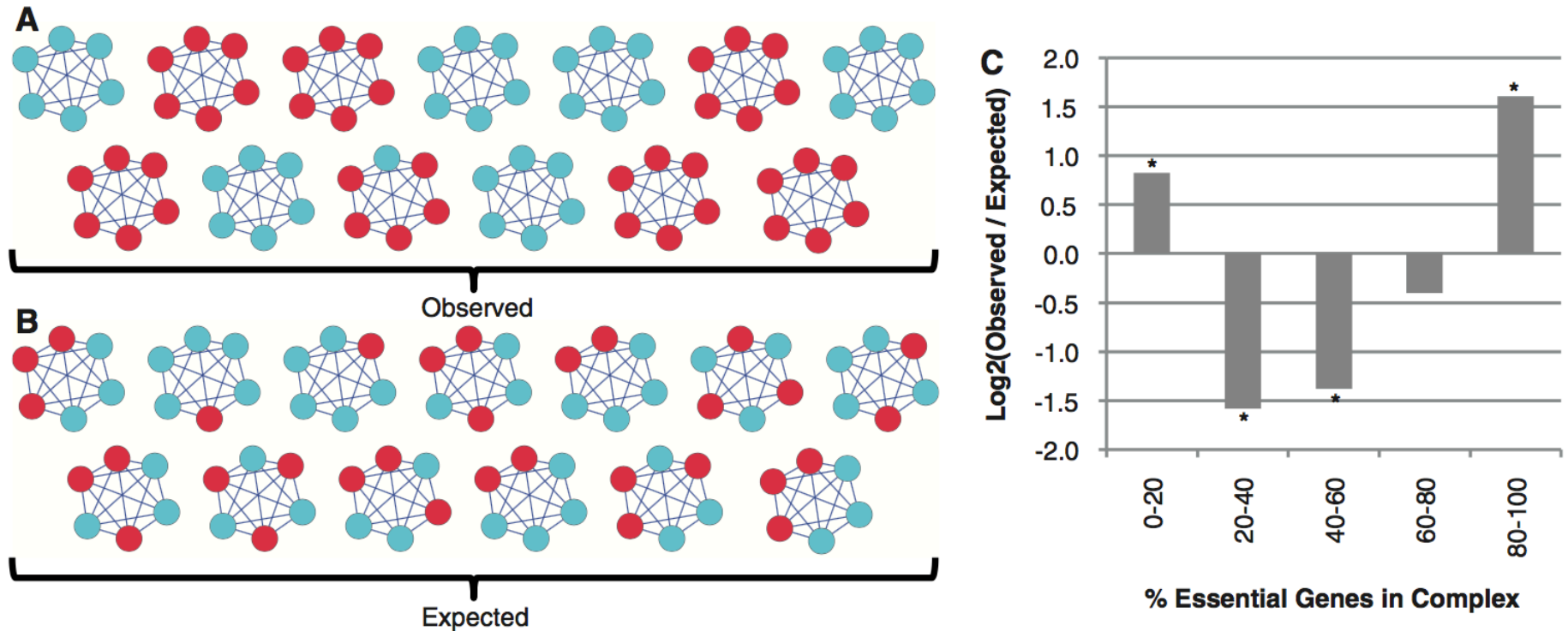
Complex essentiality fraction

$$\text{Essentiality fraction} = \frac{essential\ proteins}{total\ proteins}$$

Essentiality fraction	Literature Complexes
(0,0.2]	1907
(0.2,0.4]	90
(0.4,0.6]	37
(0.6,0.8]	9
(0.8,1]	1

Size and essentiality fraction

All or nothing hypothesis : Yeast

Hart, G. T., Lee, I., & Marcotte, E. R. (2007). A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality. *BMC Bioinformatics*, *8*, 236.
Ryan, C. J., Krogan, N. J., Cunningham, P., & Cagney, G. (2013).
All or nothing: Protein complexes flip essentiality between distantly related eukaryotes.
Genome Biology and Evolution, *5*(6), 1049–1059.

All or nothing hypothesis : *D. melanogaster*
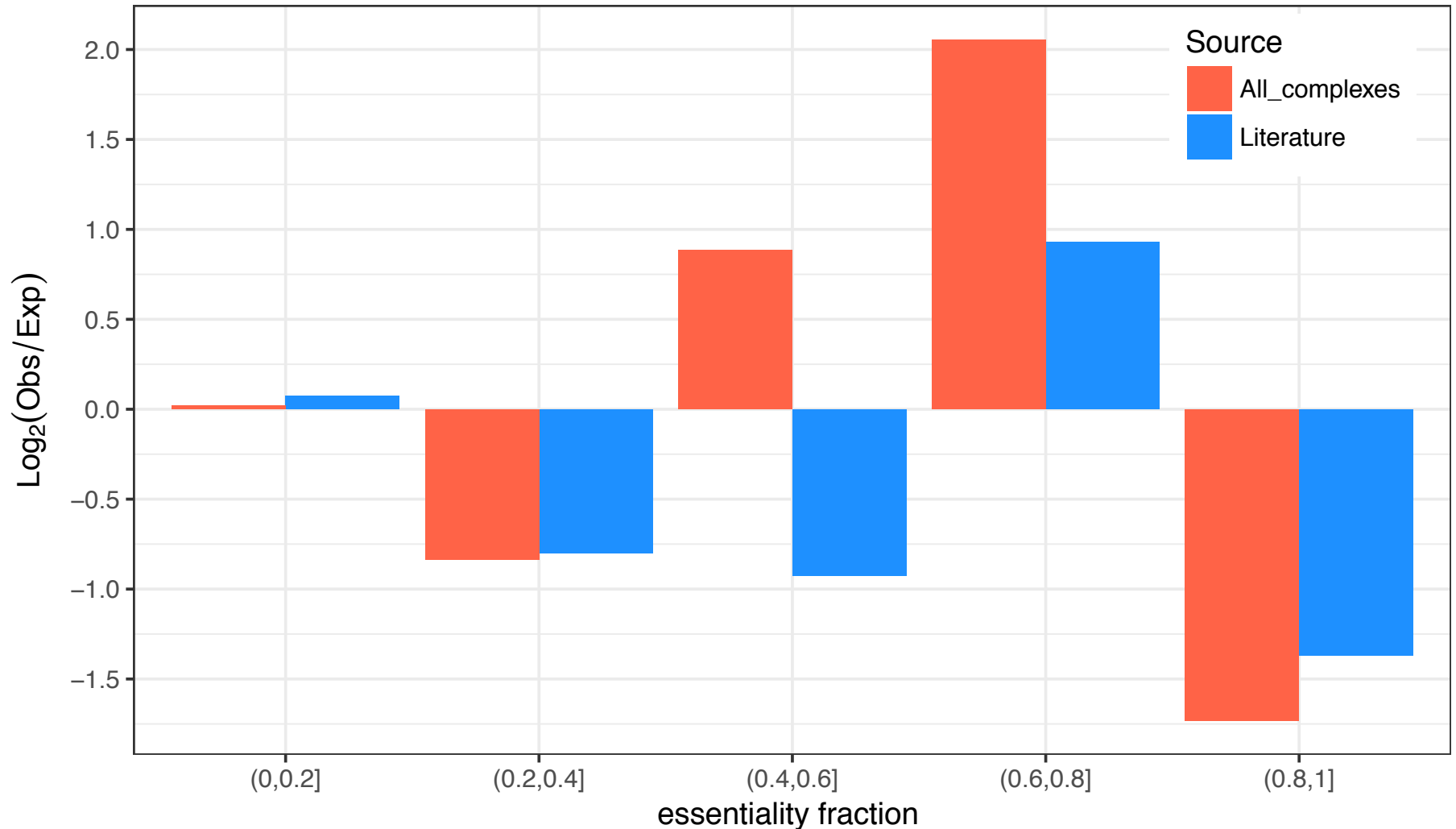
Challenges

- Network contraction with complexes to represent biological processes in networks

- Better complex prediction algorithms

- Advancement of experimental methods for better and cheaper complex detection

Summary

- Centrality – lethality is valid (although weaker) in the signed PPI network of *D. melanogaster*
- Essential proteins form a cluster with activation (positive) interactions with functions:
  - ATP synthesis
  - Proteolysis
  - Hydrogen membrane transport
  - cell cycle G1/S phase transition
- Modular essentiality is statistically significant in *D. melanogaster*
- COMPLEAT database has bias towards big complexes
- Data and method integration provides better results

Next steps

- Generalization of the centrality indices for signed networks

- Construction of large scale activation / inhibition protein networks for more organisms

- Advancement of experimental and computational methods for the construction of complexes interaction network

- Network dynamics and temporality of link activation