

# Protein Hypernetworks

Johannes Köster<sup>\*,†</sup>

Eli Zamir<sup>†</sup>

Sven Rahmann<sup>\*,‡</sup>

<sup>\*</sup>Genome Informatics, Institute of Human Genetics, Faculty of Medicine, University of Duisburg-Essen, Essen, Germany.

<sup>†</sup>Department of Systemic Cell Biology, Max Planck Institute of Molecular Physiology, Dortmund, Germany

<sup>‡</sup>Bioinformatics for High-Throughput Technologies, Algorithm Engineering, Computer Science 11, TU Dortmund, Germany

## Protein Networks

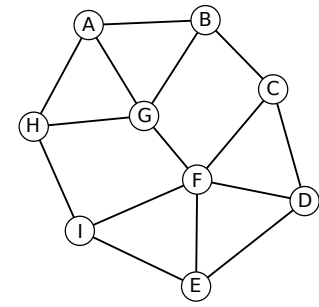
Proteins are the building blocks of cells. By interacting they form complex networks, which execute most of the cellular functions.

Models for protein networks range between two extremes:

- Differential equations

$$\frac{d[C]}{dt} = k_{on}[A][B] - k_{off}[C]$$

- Interaction graphs



↑ accuracy  
↓ scalability

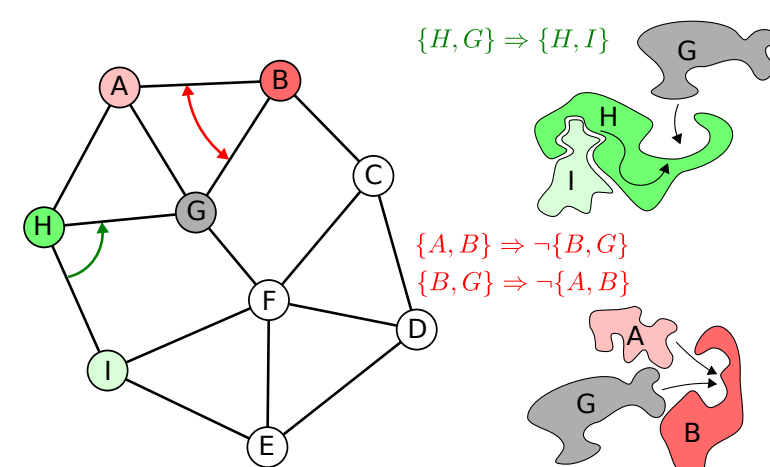
We introduce *protein hypernetworks* [3], a new model that stands in between.

## Protein Hypernetworks

Rationale: The functional versatility of protein networks emerges from dependencies between the interactions, e.g. allosteric regulation, competition on binding domains, and post-translational modifications (e.g. phosphorylation).

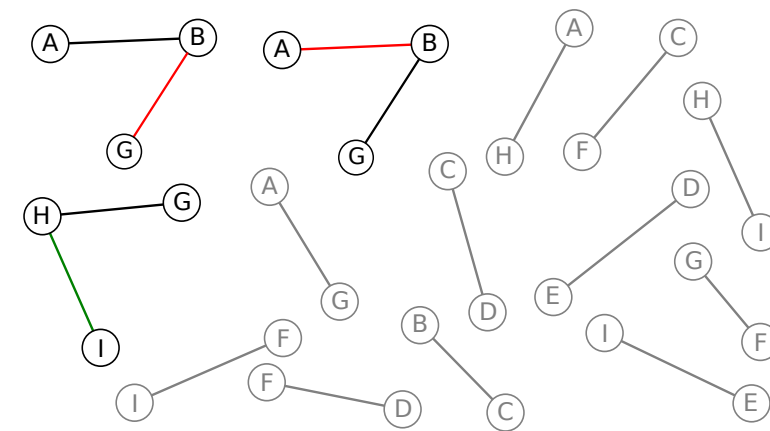
Protein Hypernetwork  $(P, I, C)$ :

- Use a plain undirected graph  $(P, I)$  for proteins and interactions.
- Add a set  $C$  of propositional logic constraints modeling interaction dependencies.



## Minimal Network States

Under the given constraints  $C$ , we are interested in all minimal sufficient network configurations that allow the occurrence of an interaction or protein  $q \in P \cup I$ .

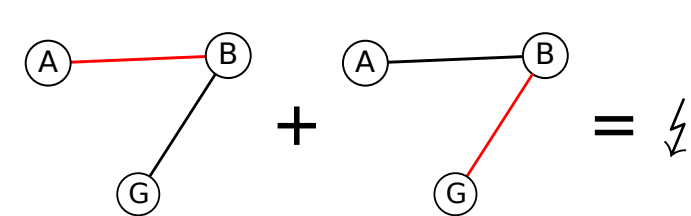


Minimal network states can be calculated by enumerating satisfying solutions for

$$q \wedge \bigwedge_{c \in C} c$$

using the propositional logic tableau calculus [4].

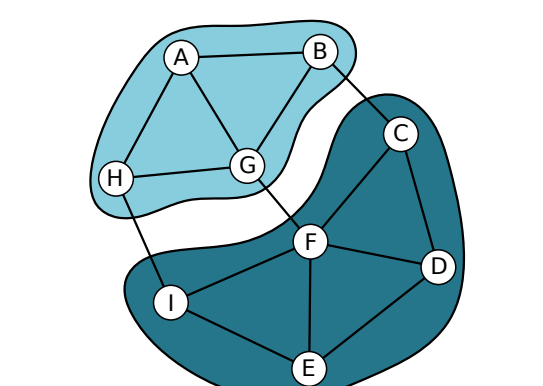
If all minimal network states of two interactions clash, they are not possible simultaneously:



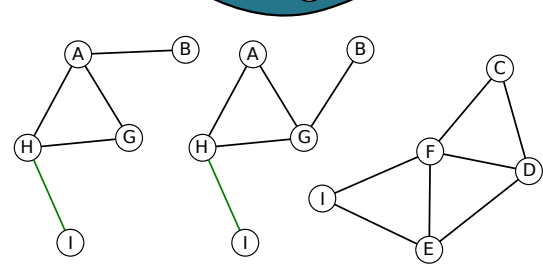
## Prediction of Protein Complexes

Protein complexes are expected to be dense subnetworks. Protein hypernetworks can be used to improve the prediction of protein complexes by taking constraints into account.

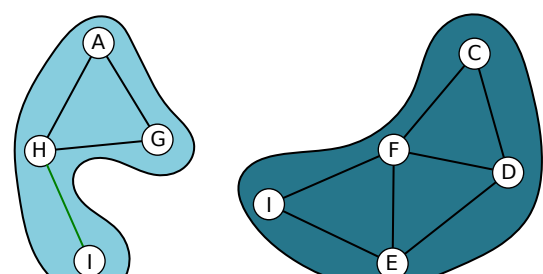
- Predict protein complexes on the plain graph



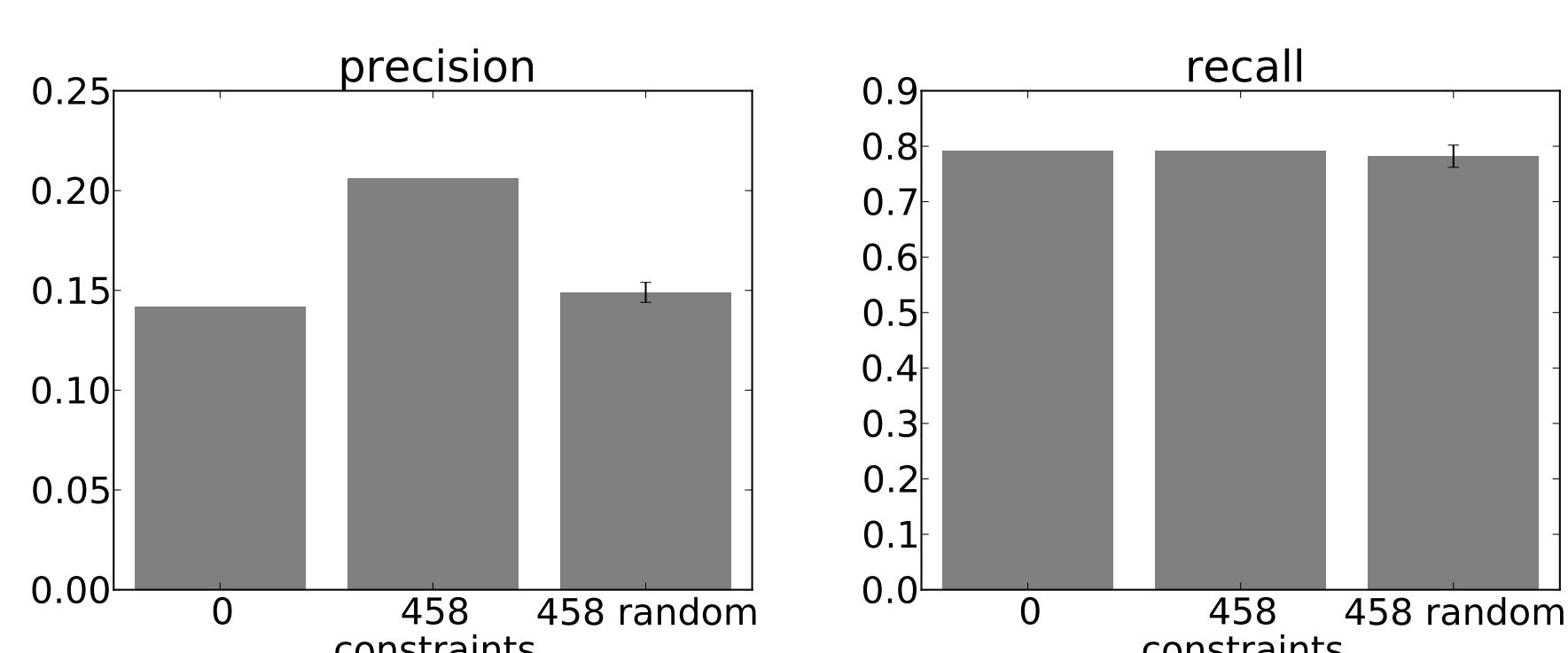
- Calculate maximal combinations of minimal network states



- Repredict refined complexes



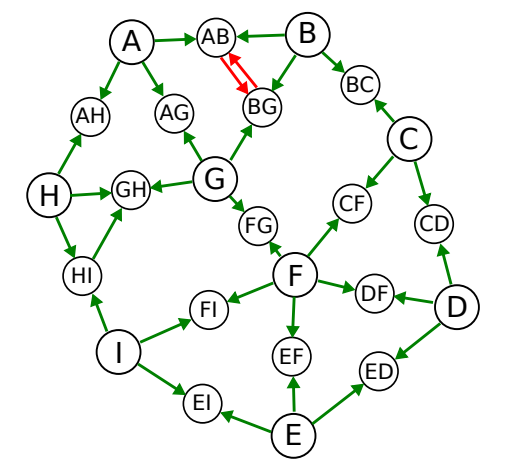
Results on the yeast protein network (4579 proteins, 12576 interactions [1], 458 constraints [2]):



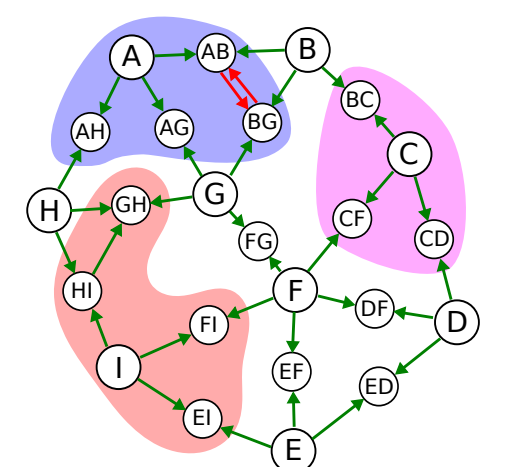
## Prediction of Functional Importance

Using protein hypernetworks we can simulate the impact of a protein perturbation (i.e. deletion). A protein with a high perturbation impact is expected to be important.

- Using the minimal network states, generate a directed dependency graph



- Perform a breadth-first-search (BFS) from each node

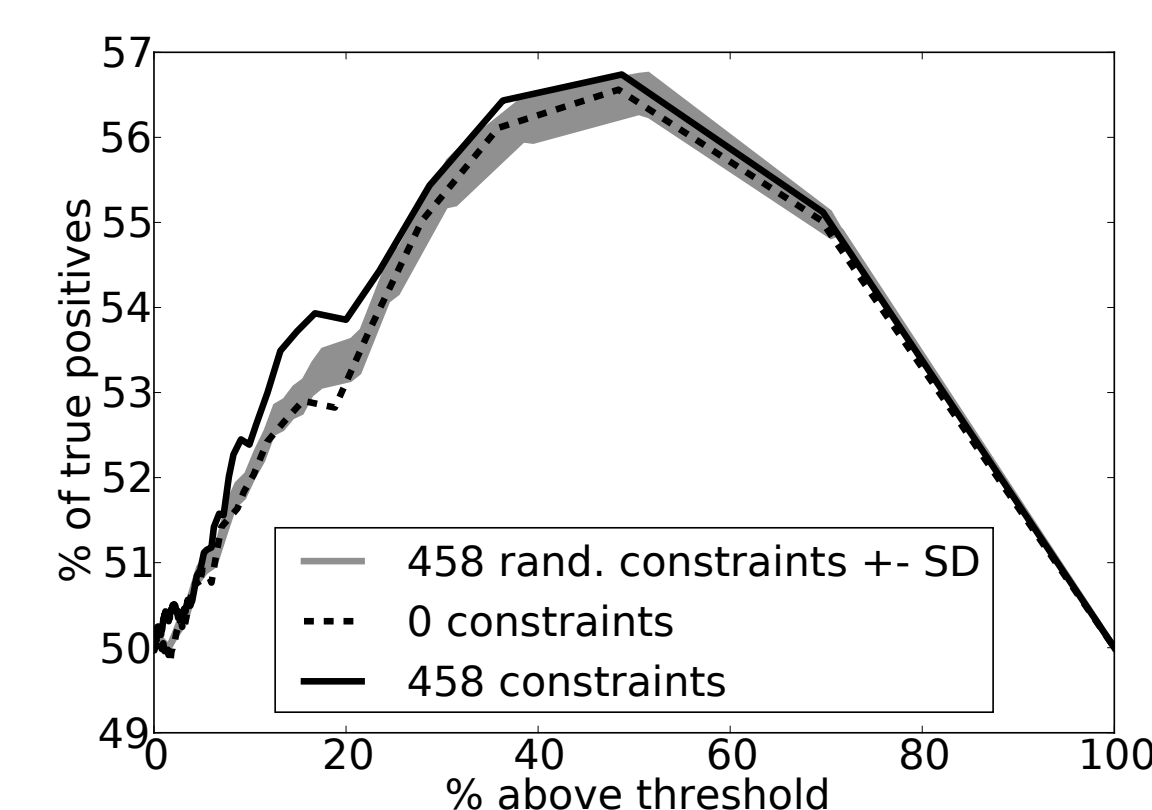


Then the perturbation impact score

$$\sum_{q \in reach_{q_i}^{BFS}} dist_{q_i}^{BFS}(q)$$

is a measure for the functional importance of a protein or interaction  $q_i \in P \cup I$ .

Results on the yeast protein network:



## Text-Mining for Interaction Dependencies

In contrast to protein interactions, that can be found in online databases like CYGD [1] or BioGrid [5], interaction dependencies are so far hidden as natural language statements in scientific publications. We propose a linear-time two-step algorithm to mine large collections of publications for interaction dependencies.

Given a text, e.g.

... It is also possible that the interaction is indirect, via a Cbl-associated protein. Shishido and colleagues proposed that binding of Abl induces a conformational change in Cbl that allows binding of Src via its SH3 domain (Shishido et al., 2000); association of endogenous Src could explain the weak phosphorylation of Cbl occasionally induced by kinase-inactive Abl (Figure 2c). ...

Miyoshi-Akiyama et al. (2001) Oncogene

the algorithm performs the following two steps:

- Translate the text into a sequence of tokens. Each token represents a relevant class of words.  $p . i p a p i p d i p p p .$
- On the sequence of tokens, search for given regular expression patterns. Matches hint to interaction dependencies.  $p . i p a p i p d i p p p .$

... binding of Abl induces a conformational change in Cbl that allows binding of Src ...

For the human adhesome protein network [6] we extracted 144 matches from 16,419 publications. From these, we manually curated 41 interaction dependencies.

dependency type	precision	reported dependencies
allosteric	0.38	6
mutually exclusive	0.62	35
complete	0.59	41

## References

- U. Güldener, M. Münsterkötter, G. Kastenmüller, N. Strack, J. van Helden, C. Lemer, J. Richelles, S. J. Wodak, J. García-Martínez, J. E. Pérez-Ortín, H. Michael, A. Kaps, E. Talla, B. Dujon, B. André, J. L. Souciet, J. De Montigny, E. Bon, C. Gaillardin, and H. W. Mewes. CYGD: the comprehensive yeast genome database. *Nucleic Acids Research*, 33(Supplement 1):D364–D368, 2005.
- S. H. Jung, B. Hyun, W.-H. Jang, H.-Y. Hur, and D.-S. Han. Protein complex prediction based on simultaneous protein interaction network. *Bioinformatics*, 26(3):385–391, 2010.
- J. Köster, E. Zamir, and S. Rahmann. Protein Hypernetworks: a Logic Framework for Interaction Dependencies and Perturbation Effects in Protein Networks. *arXiv preprint*, June 2011.
- R. M. Smullyan. *First-Order Logic*. Dover Publications, Jan. 1995.
- C. Stark, B.-J. J. Breitkreutz, A. Chatr-Aryamontri, L. Boucher, R. Oughtred, M. S. Livstone, J. Nixon, K. Van Auken, X. Wang, X. Shi, T. Reguly, J. M. Rust, A. Winter, K. Dolinski, and M. Tyers. The BioGRID Interaction Database: 2011 update. *Nucleic acids research*, 39(Database issue):D698–D704, Jan. 2011.
- R. Zaidel-Bar, S. Itzkovitz, A. Ma'ayan, R. Iyengar, and B. Geiger. Functional atlas of the integrin adhesome. *Nature cell biology*, 9(8):858–867, Aug. 2007.