ProtEGOnist | Exploration of Protein-Protein Interactions using EGO-Graph Networks

N. Brich, T. Harbig, M. Witte Paz, S. Hackl, C. Jachmann , M. Schäfer, M. Krone, K. Nieselt

Institute for Bioinformatics and Medical Informatics, Eberhard Karls Universität Tübingen

Motivation

Protein-Protein Interaction (PPI) networks are vital for understanding cellular processes. However, they require a large amount of information that hinders their interpretable visualization. For example, the challenge's dataset [1] consists of 8,392 human proteins and 66,396 PPIs acquired from the STRING database [2].

The resulting network visualization for the redesign challenge was greatly cluttered and key proteins, e.g., drug targets identified by the authors, cannot be easily visually located. Moreover, an additional bar chart was required to transmit the main message of the authors.

Design Requirements

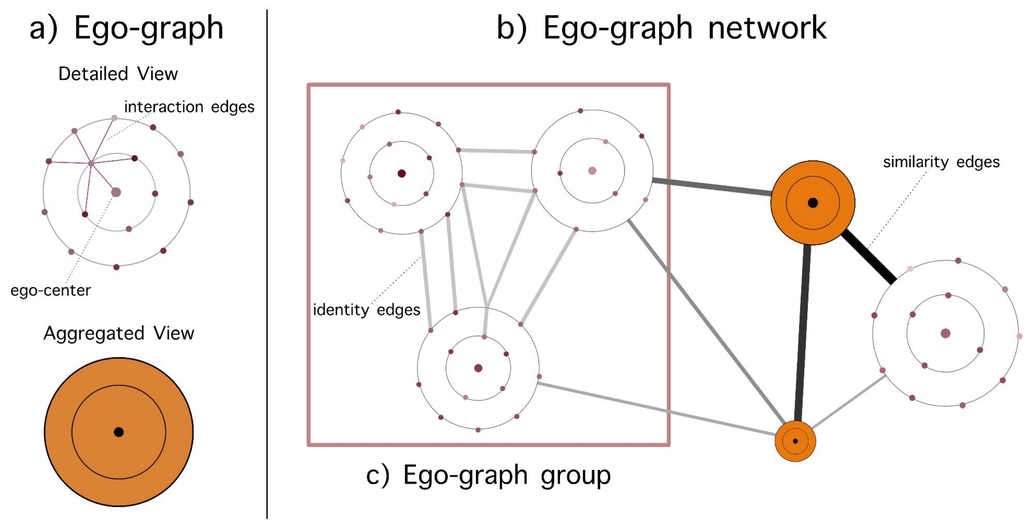
To tackle these deficiencies and to meet the challenge's tasks, we defined requirements for our redesign approach:

1. **Reduce** the PPI network to a set of proteins of interest, which represent the protein-drug associations identified in [2].
2. Provide an **overview** of the PPI network, while offering detailed exploration for subnetworks of drug-protein associations.
3. **Compare** subnetworks of drug-protein associations.
4. Incorporate functional information for drug-protein associations

Approach

Build ego-graphs for proteins (neighborhood with edge distance ≤2) and calculate a similarity value (Jaccard index). This allows for **comparing** a single ego-graph to many others, building an ego-graph network. Different views (see right Figure):

a) Detailed ego-graphs inspired by egoComp [3]: Nodes distributed in two circular layers around the ego-center, showing interacting proteins.
Aggregated ego-graphs: Circular glyph with area encoding ego-neighborhood size. These are connected to other graphs via similarity edges
b) Ego-graphs are visualized in an ego-graph network.
c) When adjacent ego-graphs are expanded, ego-graph groups are formed. Identical proteins are connected via identity edges.



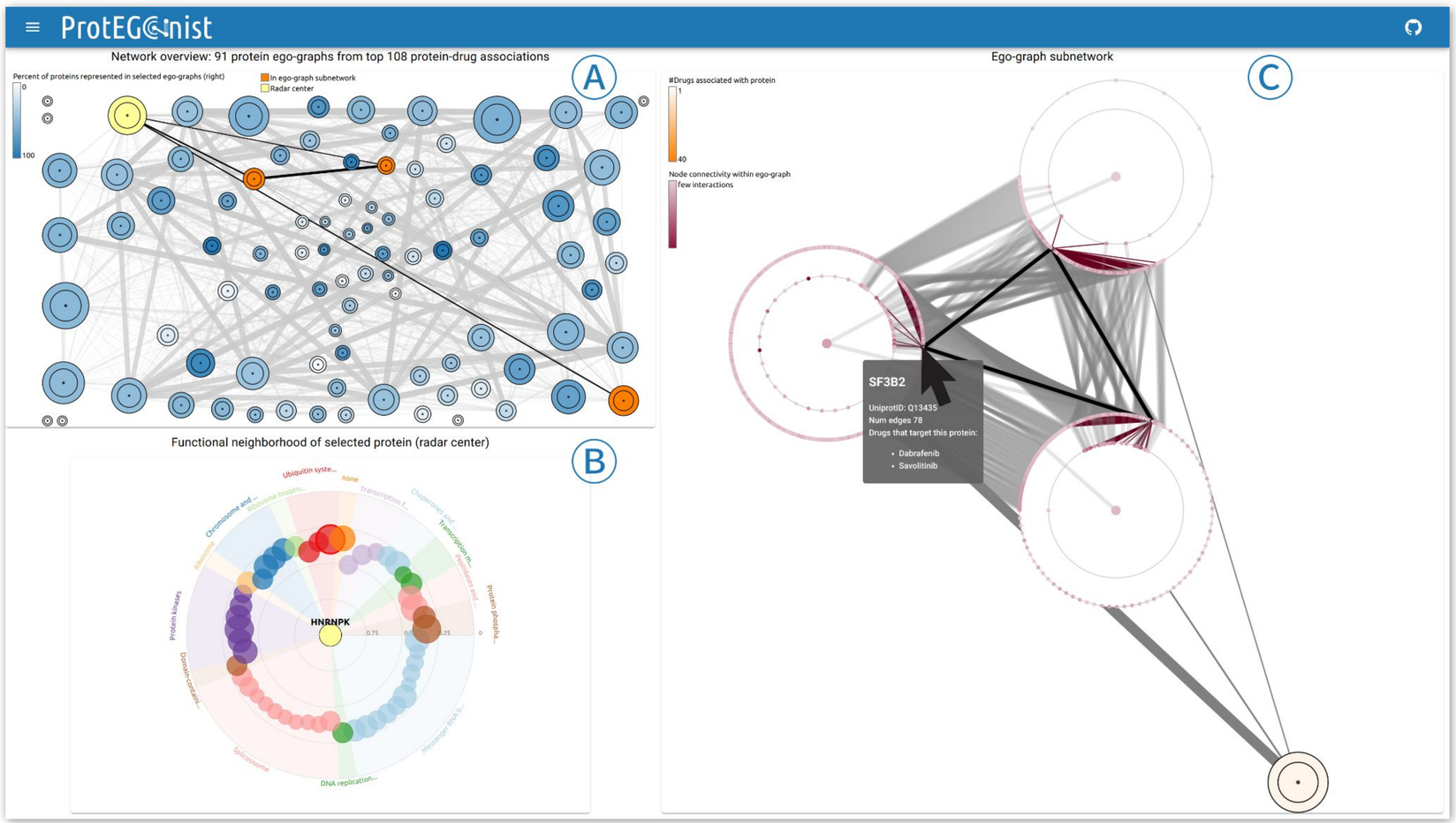
ProtEGOnist Dashboard

A - Network Overview

Network of 91 ego-graphs visualized as glyphs. Edges are weighted by the Jaccard index of the two resp. ego-graphs.

B - Radar Chart

Radar chart showing the similarity between a selected ego-graph (center, yellow node in A) and its neighbor ego-graphs. Neighbor ego-graphs are grouped (same color) by the function of their central protein using the BRITE hierarchy [4]. Radial distance encodes similarity.



C - Ego-subnetwork

Ego-graph subnetwork of ego-graphs selected from the overview (orange nodes in A) for more details. Glyph nodes are colored by #drugs associated with the respective ego-center. Three ego-graph nodes are decollapsed to show a detailed ego-graph group (see Figure b) above).

Views in the Dashboard

- are complemented by a selection table
- are linked
- provide details on hover (e.g., tooltips)
- geometrical zoom in ego-graph subnetwork

Example Use Case

First, all proteins associated with a single drug, here drug Ara-G, are selected from the network overview. From the resulting ego-subnetwork consisting of four ego-graphs, three highly connected ego-graphs are decollapsed. From this view, one can interact with one single protein that is common in all three ego-graphs, such as seen in the figure in ego-graph subnetwork.

The analysis of the function of the ego-center proteins using the radar chart reveal prominently represented protein classes. All three decollapsed ego-graphs were highly associated with the class "spliceosome". From this point, users could, e.g., continue analyzing the effects of drug Ara-G on spliceosome-related proteins.

Outlook

- Addition of other data types:
  - Expression and drug response
- Abstraction to less specific experiment utilizing the STRING database or PPI networks in their workflow

References

[1] D. Szklarczyk, A. Franceschini, et al.: *STRING v10*. Nucleic Acids Res., 43(Database):D447–452, 2015.
[2] E. Gonçalves, R. C. Poulos, et al.: *Pan-cancer proteomic map of 949 human cell lines*. Cancer Cell, 40(8):835–849.e8, 2022.
[3] D. Liu, F. Guo, et al. egoComp. Inf. Vis., 16(3):179–189, 2017.
[4] M. Kanehisa, M. Araki, et al.: *KEGG for linking genomes to life and the environment*. Nucleic Acids Res., 36(Database): D480–D484, 2007.