

A Web Graphical Tool for the Integration of Unbalanced Biomolecular Networks

P. Perlasca¹, M. Mesiti^{1*}, M. Notaro¹, A. Petrini¹,
J. Gliozzo², G. Valentini¹, M. Frasca¹

¹ AnacletoLab – Dipartimento di Informatica, Università degli Studi di Milano

² Fondazione IRCCS Ca' Granda - Ospedale Maggiore Policlinico, Università degli Studi di Milano

*To whom correspondence should be addressed: mesiti@di.unimi.it

1 MOTIVATION

The automated protein function prediction problem (AFP) is mainly characterized by the unbalance between annotated and unannotated genes and the integration of multiple data sources. The “informativeness” of each network/source may depend on the considered protein function, and neglecting the unbalance between annotated and unannotated proteins may thereby lead to strong decay in performance. Recently, the UNIPred algorithm [1] was proposed to integrate in a function-specific fashion the input networks by automatically handling the data imbalance. A relevant challenge in this context is the appropriate visualization and interpretation of the resulting network. Indeed, the network size can be extremely big and their simple visualization, with off-the-shelf graphical visualization tools (e.g. GraphViz, GeneMania) produces a cloud of points hard to interpret and handle within a browser (lack of memory). To face the problem, here we propose a web-tool implementing UNIPred and introducing an approximate visualization of the graph. Having the system embedded different levels of abstraction, the user can both decide the part of the graph to explore, and click on the corresponding part to obtain a new enhanced visualization.

2. METHODS

The different input networks have been represented and stored according to the relational model. Efficient PL/SQL procedures calculate subgraphs centered on a vertex and with a given radius. An R software gathers the networks from the database, integrate them according to the UNIPred algorithm and store the results back in the database. Relying on this infrastructure, a web graphical tool has been implemented that offers different facilities to the user for the network management, their integration, visualization and exploration. In particular, both “vertex-centric” exact and approximate visualizations are provided. With “vertex-centric” [2] visualization approach we mean that the user can specify a vertex, named *target*, he wishes to explore the result of the integration, and the size of the subnetwork to extract. The extracted network can be shown to the user when its size is small and the available canvas is big enough for its visualization. Otherwise, approximate visualization techniques are used. For example, the target node can be connected with bubbles of different sizes that distribute the nodes of the subgraph

relying on the weight of the outgoing edges and the distance from the target. Clicking on one of the bubbles, the visualization can be then expanded by showing other bubbles or single nodes.

3 RESULTS

We have realized a Web tool offering different facilities for the exploration of protein networks, their efficient integration using the UNIPred algorithm in a Web-based environment, and the exploration of the resulting network by means of a vertex-centric visualization approach. The visualization can be exact or approximate depending on the size of the network and of the drawing canvas. In the left part of Figure 1 a target is shown in the center and then four bubbles that partition the vertex at distance 1 from the target depending on the weight of the outgoing edges. This representation points out how much the co-functionality confidence degree has been propagated (or not) from the target node. Moreover, by clicking on one of the bubbles, it is split again in 4 parts, thus allowing multiple “vertex-centric” views at different resolution levels.

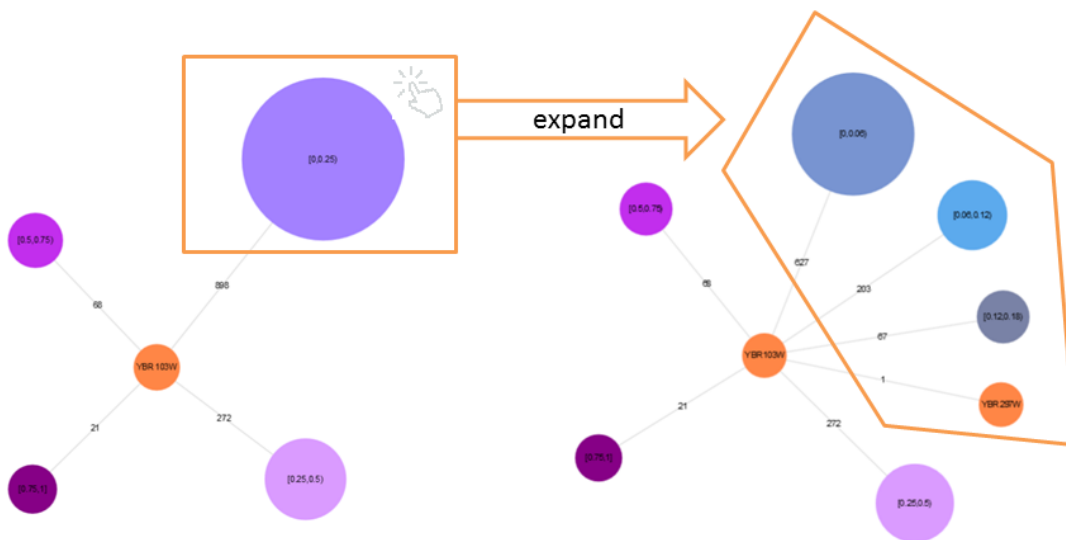


Figure 1 Vertex centric approximate visualization of an integrated network and the result of the expansion of a bubble

References

- [1] M. Frasca, A. Bertoni, G. Valentini UNIPred: Unbalance-aware Network Integration and Prediction of protein functions, *Journal of Computational Biology*, 22(12): 1057-1074, 2015
- [2] M. Mesiti, M. Re, G. Valentini Think globally and solve locally: secondary memory-based network learning for automated multi-species function prediction, *GigaScience*, 3:5, 2014