Link prediction in protein interaction networks

supervision: Lionel Tabourier (LIP6 - CNRS/SU), Élodie Laine (LCQB - CNRS/SU)
location: Sorbonne Université – 4 place Jussieu, Paris

`stages@complexnetworks.fr`, `http://complexnetworks.fr`

Scientific project:

PPI (*protein-protein interaction*) networks represent interactions between proteins within a living organism. PPI network maps are incomplete because checking the existence of each relationship demands specific experiments, it is therefore desirable to have means to select the most probable interactions. Recent works, such as [KLS+19] or [DMC20] brought to light the fact that link prediction approaches are relevant to detect interactions between proteins.
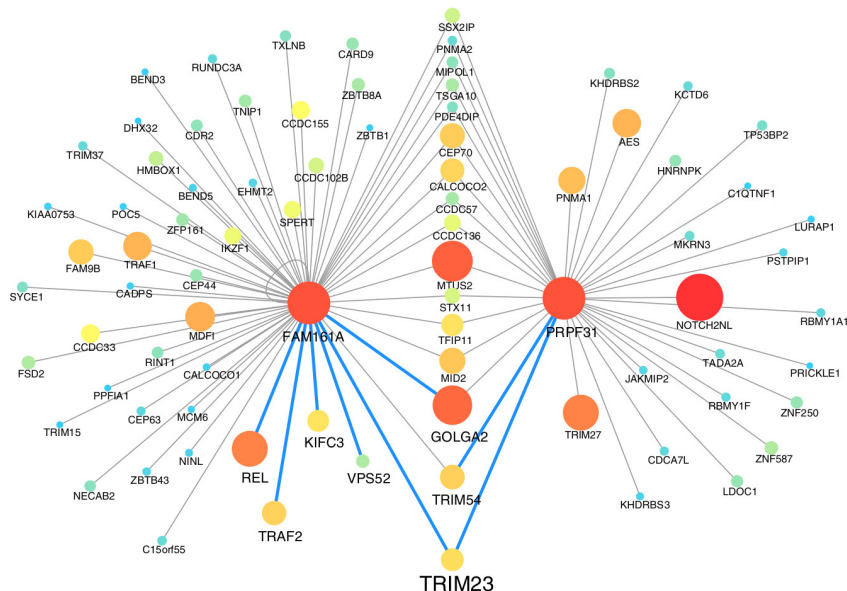

Figure 1: Representation of a part of the PPI network involved in *retinitis pigmentosa* disease, grey edges were already known while blue edges have been detected by link prediction methods and then validated experimentally (image extracted from [KLS+19]).

The approaches in question are unsupervised, however there exist supervised methods which have been designed for analogous problems in other contexts (*e.g.*, [PK12, TBLL19]). We think that it is possible to adapt such methods to the context of PPI networks. By defining adequate graph features – particularly specific graph motifs – in order to achieve the learning, it would be possible to improve significantly the predictive power of these methods. The purpose of the internship is to design and apply such prediction methods.

The developed methods will be trained and validated using several networks comprised of 5 000 - 18 000 proteins (nodes) establishing between 20 000 and more than 2 million experimentally validated interactions (edges) coming from reference PPI resources, namely the STRING database (`https://string-db.org`), the BioGRID (`https://thebiogrid.org`), and the Human Reference Interactome (HuRI, `http://www.interactome-atlas.org`).

Profile:

This internship is preferably directed at Master students with a background in computer science or bioinformatics. Good coding skills are requested for the internship, knowledge of a widely-used language in learning, such as python, is preferable but not mandatory. An open-mind to interdisciplinary applications is certainly a plus.

Details about the position:

The remuneration of the internship is based on the standard rate of academic internship in France (3.9 euros/h, 35 h/week) with an additional compensation for food and transport. The intern will be located in Paris on the Jussieu Campus of Sorbonne Université and hosted at the Complex Networks team of the LIP6 (Computer Science Laboratory) but arrangements can be discussed to be hosted at the LCQB (Laboratory of Computational and Quantitative Biology). The internship will be supervised by Lionel Tabourier (LIP6) and Élodie Laine (LCQB).

References

[DMC20] Kapil Devkota, James M Murphy, and Lenore J Cowen. Glide: combining local methods and diffusion state embeddings to predict missing interactions in biological networks. *Bioinformatics*, 36(Supplement_1):i464–i473, 2020.

[KLS+19] István A Kovács, Katja Luck, Kerstin Spirohn, Yang Wang, Carl Pollis, Sadie Schlabach, Wenting Bian, Dae-Kyum Kim, Nishka Kishore, Tong Hao, et al. Network-based prediction of protein interactions. *Nature communications*, 10(1):1–8, 2019.

[PK12] Manisha Pujari and Rushed Kanawati. Tag recommendation by link prediction based on supervised machine learning. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 6, pages 547–550, 2012.

[TBLL19] Lionel Tabourier, Daniel F Bernardes, Anne-Sophie Libert, and Renaud Lambiotte. Rankmerging: a supervised learning-to-rank framework to predict links in large social networks. *Machine Learning*, 108(10):1729–1756, 2019.