Prediction of Protein-Protein Interactions on the Human and Rice Interactomes

Nicolas A. Lopez-Rozo[*], Jorge Finke and Camilo Rocha

[*]Correspondence:
nicolaslopez@javerianacali.edu.co
Department of Electronics and
Computer Science, Pontificia
Universidad Javeriana, Cali, CO
Full list of author information is
available at the end of the article

Abstract

Background: Abstract A recent study in network-based prediction of protein-protein interactions (PPIs) reveals that two proteins are more likely to interact, the higher the number of paths of length 3 between them (normalized by the geometric average of their interactions). This paper extends previous work on mapping binary interactions by taking into account the learning of features (embeddings) of the PPI network. In particular, we implement a gradient boosted decision tree model (XGBoost) using handcrafted features (including the normalized measure) and embeddings from an algorithm that generates a low-dimensional representation of nodes (node2vec).

Results: Our main result shows that while the measure remains an important feature for predicting interactions, better performance is achieved when in addition embedding features are considered. The proposed approach is validated for the human and rice interactomes. For both cases, the combination of both types of features yield higher AUC values.

Conclusions: As found on this study on both human and rice, when information from handcrafted features based on neighborhood is enhanced with vector representations from random walks, the prediction power of the model improves. Besides, a supervised learning model can be trained for predicting unknown interactions based on such information. Finally, the developed framework can also be applied to interactomes of other organisms for which PPI networks have recently become available.

Keywords: PPI; Prediction; Protein Interaction; Machine Learning; Python; XGBoost; node2vec

Introduction

Text for this section...

Sub-heading for section
Text for this sub-heading...

Sub-sub heading for section
Text for this sub-sub-heading...

Sub-sub-sub heading for section Text for this sub-sub-sub-heading...

In this section we examine the growth rate of the mean of Z_0, Z_1 and Z_2. In addition, we examine a common modeling assumption and note the importance of considering the tails of the extinction time T_x in studies of escape dynamics. We

will first consider the expected resistant population at vT_x for some $v > 0$, (and temporarily assume $\alpha = 0$)

$$E\big[Z_1(vT_x)\big] = \int_0^{v \wedge 1} Z_0(uT_x) \exp(\lambda_1) \, du.$$

If we assume that sensitive cells follow a deterministic decay $Z_0(t) = xe^{\lambda_0 t}$ and approximate their extinction time as $T_x \approx -\frac{1}{\lambda_0} \log x$, then we can heuristically estimate the expected value as

$$
\begin{aligned}
E\big[Z_1(vT_x)\big] \\
= \frac{\mu}{r} \log x \int_0^{v \wedge 1} x^{1-u} x^{(\lambda_1/r)(v-u)} \, du.
\end{aligned}
\tag{1}
$$

Thus we observe that this expected value is finite for all $v > 0$ (also see [1, 2, 3, 4, 5, 6]).

Appendix
Text for this section...

Abbreviations
Text for this section...

Availability of data and materials
Text for this section...

Ethics approval and consent to participate
Text for this section...

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Text for this section...

Authors' contributions
Text for this section ...

Authors' information
Text for this section...

Author details
Department of Electronics and Computer Science, Pontificia Universidad Javeriana, Cali, CO.

References
1. Koonin, E.V., Altschul, S.F., Bork, P.: Brca1 protein products: functional motifs. Nat. Genet. **13**, 266–267 (1996)
2. Jones, X.: Zeolites and synthetic mechanisms. In: Smith, Y. (ed.) Proceedings of the First National Conference on Porous Sieves: 27-30 June 1996; Baltimore, pp. 16–27 (1996)
3. Margulis, L.: Origin of Eukaryotic Cells. Yale University Press, New Haven (1970)
4. Schnepf, E.: From prey via endosymbiont to plastids: comparative studies in dinoflagellates. In: Lewin, R.A. (ed.) Origins of Plastids, 2nd edn., pp. 53–76. Chapman and Hall, New York (1993)
5. Kohavi, R.: Wrappers for performance enhancement and obvious decision graphs. PhD thesis, Stanford University, Computer Science Department (1995)
6. ISSN International Centre: The ISSN register (2006). http://www.issn.org Accessed Accessed 20 Feb 2007

Figure 1 Sample figure title

Figure 2 Sample figure title

Table 1 Sample table title. This is where the description of the table should go

	B1	B2	B3
A1	0.1	0.2	0.3
A2	...	..	.
A3	..	.	.

Figures
Tables
Additional Files
Additional file 1 — Sample additional file title
Additional file descriptions text (including details of how to view the file, if it is in a non-standard format or the file extension). This might refer to a multi-page table or a figure.

Additional file 2 — Sample additional file title
Additional file descriptions text.