ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Robustness in Protein-Protein Interaction Networks: A Link Prediction Approach

Alessandro Dipalma, Domenico Tortorella, Alessio Micheli

University of Pisa - Department of Computer Science
Largo B. Pontecorvo 3, 56127 Pisa - Italy

Abstract. Protein-protein interaction networks (PPINs) are indispensable in exploring complex biological systems, facilitating advancements in fields like drug discovery, protein function annotation, and disease mechanism elucidation. So far, predicting the dynamical properties of biochemical pathways has relied on costly numerical simulations. In this paper, we propose exploiting the topological information in PPINs to restate the problem of predicting pathway robustness as a link prediction task. Our experiments show that the PPIN topology can supply information on inter-pathway relationships, significantly improving predictions of the graph-agnostic baseline relying only on protein sequence embeddings.

1 Introduction

Protein-Protein Interaction Networks (PPINs) serve as a foundational framework for understanding cellular processes by mapping the *physical*, *genetic*, and *predicted* interactions among proteins. Interactomes, i.e. the set of all species-specific protein interactions, are humongous dynamical systems, driven by internal and external factors. Due to the incomplete knowledge of biochemical process underlying interactions, and the computational infeasibility of simulating the whole interactome, so far detailed study of dynamics has been restricted to isolated cellular processes, known as *biochemical pathways* (BPs). In this context, BPs' simulations have been extensively utilized to measure the impact of perturbations on the nominal behavior of the biological process [1, 2]. Measuring Dynamical Properties (DPs) provides insights that can be employed, for example, to advance knowledge on disease mechanisms and identify drug targets.

Deep learning on graphs has emerged as a powerful approach to address biological network analysis, thanks to its ability to deal with incomplete knowledge, and automatically learn a hierarchy of meaningful representations directly from data, while at the same time considerably offsetting the computational cost of numerical simulations [3].

In this work, we propose to model *concentration robustness* estimation as a learning task on the whole PPI network. Our approach bridges the gap between dynamic property inference and static network analysis, allowing to explore long range functional dependencies in PPINs. We show that a Deep Graph Network (DGN) can efficiently produce accurate predictions for any pair of proteins in the interactomes with high performances by leveraging the PPI network topology.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

2 Background

BPs represent the intricate series of reactions among molecular species. BPs are often modeled with formalisms that allow to simulate their dynamical behavior, such as Ordinary Differential Equations (ODEs). Simulations via exact or stochastic approaches enable the computation of Dynamical properties (DPs) such as stability, robustness and monotonicity to measure system's responses to perturbations. Introduced in [4], α-*robustness* evaluates whether the steady-state concentrations of reactants remain within predefined acceptable bounds under varying input conditions. The measure is particularly useful in analyzing BPs, where perturbations to one molecular species can propagate through the network and influence downstream interactions. An output species s_{out} is deemed α-robust if its concentration demonstrates a bounded response to perturbations of the concentration of an input species s_{in}.

Simulation based methods can provide deep insights, but they suffer from several limitations. First, accurate ODE modeling of BPs requires extensive knowledge of kinetic parameters and reactions, which are often unavailable or incomplete. Second, the computational complexity of simulations scales poorly with network size. As pointed out in [5], DPs can often be inferred from network topology alone without requiring explicit simulations. These ideas have seen further development in [6, 7, 3], where the authors have shown that is possible to predict DPs from BPs employing DGNs reaching high accuracies representing the BP as a bipartite graph, leaving out stoichiometric details.

PPI databases provide a static representation of the interactome, focusing on the presence or absence of interactions without capturing their temporal dynamics. However, there is a strong relationship between PPINs and BPs: arcs in PPINs represent interactions that are integral to biochemical reactions, such as protein modifications, complex formation, or signaling cascades. By mapping the connectivity patterns in PPINs to DPs observed in BPs, researchers can study how BPs modifications can affect neighboring processes. Current interactomes are or are close to being completed [8], and as most biological networks they exhibit complex topological features (high clustering coefficient, hierarchical structure, small-worldness) that should allow to study their dynamics [9]. A lot of work in predictive tasks in PPINs has focused in the prediction of missing interaction links, or to compute meaningful protein representations [10, 11], but there is an unfilled gap in predictive methodologies to extract insights about PPIN evolving behaviour and indirect influences between distant proteins.

3 Materials and methods

The robustness is defined over BPs, which involve molecules that can be complexes containing multiple proteins. Therefore, we must first rephrase the problem of its prediction from BP species to pairs of proteins on the PPIN. Once transposed in graph form, the robustness prediction problem can be addressed as a graph learning task, namely predicting whether a pair of input-output proteins

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

of the PPIN is robust or not to concentration perturbation.

Robustness from BPs to PPINs We first simulate all the BPs in the Biomodels (1060) repository [12] up to the steady state, following the methodology of [3]. Input species concentrations are varied sampling 32 points within ±20% of the reference input concentration, for a total of 315K simulations. We consider an output species s_{out} robust to perturbation on the input species s_{in} if the variation of the output species concentration remains within ±20% of its reference value.

As BP species can correspond to complexes constituted of multiple proteins, we consider a protein p_{out} robust to the perturbation of protein p_{in} if and only if all biochemical species s_{out} containing p_{out} are robust to all biochemical species s_{in} containing p_{in}. We use the biomodel annotations from the UniPROT database [13] to map BP species to their constituent proteins.

Finally, we construct our training graph collecting all the PPI, independently from the host organism, from the BioGRID database [14]. We have now obtained a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ having as set of nodes $\mathcal{V} = \{\mathsf{p}_i\}_i$ the proteins, and as the set of directed edges $\mathcal{E} = \{(\mathsf{p}_i, \mathsf{p}_j)\}_{ij}$ the interactions of the PPIN. We further add as node features $\mathbf{x}_i \in \mathbb{R}^{128}$ the encoding of protein sequences obtained by ProtT5 [15] and compressed with principal component analysis. The robustness between a input-output pair $\varrho(\mathsf{p}_{\text{in}}, \mathsf{p}_{\text{out}})$ is thus expressed as a relationship between pairs of nodes in $\mathcal{G}$, that is as a link prediction task.

Learning robustness with DGNs A plethora of models have been proposed to perform Deep Learning on graphs [16]. Within the the class of Deep Graph Networks (DGNs), convolutional architectures are based on a stack of L layers that learn a hierarchy of node representations $\mathbf{h}_i^{(\ell)} \in \mathbb{R}^H$ by aggregating neighborhood information, progressively relying on larger contexts and discovering longer-range relationships between nodes. For our model, we adopt the convolutional layers of GraphSAGE [17],

$$\mathbf{h}_i^{(\ell)} = \text{ReLU}\left(\mathbf{W}_1^{(\ell)}\,\mathbf{h}_i^{(\ell-1)} + \tfrac{1}{|\mathcal{N}_i|}\textstyle\sum_{j\in\mathcal{N}_i}\mathbf{W}_2^{(\ell)}\,\mathbf{h}_j^{(\ell-1)}\right), \tag{1}$$

where $\mathcal{N}_i$ is the set of neighbors of node i, and $\mathbf{W}_1^{(\ell)}, \mathbf{W}_2^{(\ell)} \in \mathbb{R}^{H\times H}$ are trainable parameters. For the first layer $\ell = 1$, the input representations $\mathbf{h}_i^{(0)}$ are the sequential embeddings $\mathbf{x}_i$ of the protein p_i. The final layer's representations are used to predict the robustness between pairs of proteins via the readout

$$\hat{\varrho}(\mathsf{p}_i, \mathsf{p}_j) = \sigma\left(\bar{\mathbf{W}}_{\text{in}}\,\mathbf{h}_i^{(L)} + \bar{\mathbf{W}}_{\text{out}}\,\mathbf{h}_j^{(L)}\right), \tag{2}$$

whose parameters are trained end-to-end with the convolutional layers by minimizing the binary cross-entropy loss.

4 Experiments and discussion

The graph $\mathcal{G}$ accounts for 60K nodes and 2M edges, with average node degree 67. As all PPIs have been collected independently from the host organism, the

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

	Input data			Metric (%)		
Model	Emb.	Graph	Direct.	AUROC	ACC	F1
Null	-	-	-	50.00	57.34	0.00
MLP	✓	✗	-	90.28±0.25	82.69±0.11	79.36±0.26
DGN	✗	✓	✗	81.91±1.27	73.36±1.04	68.25±0.72
DGN	✗	✓	✓	82.63±0.39	74.49±0.51	69.19±0.84
DGN	✓	✓	✗	<u>93.58</u>±0.29	**87.44**±0.25	**84.86**±0.34
DGN	✓	✓	✓	**93.66**±0.27	<u>87.32</u>±0.38	<u>84.70</u>±0.25

Table 1: Final classification scores, mean and deviation over 3 different weights initializations. *Null* is a model that always predicts the majority class. Bold and underline hihglight the best and second bestperformance for each metric.

graph is constituted by 321 connected components, the largest of whom is the human interactome, presenting average shortest path length 5.6 and diameter of 9. The 15366 robustness target links are split in training/validation/test sets with 60:20:20 proportions. The best model hyper-parameters are selected according to the validation F1 score.

We consider a multilayer perception (MLP) fed with the concatenation of the input and output protein sequence embeddings $(\mathbf{x}_i, \mathbf{x}_j)$ as a baseline model that does not take into account graph structure in making robustness predictions. On the DGN model, we further perform an ablation study to assess the importance of protein sequence embeddings and of interaction directionality in the PPI network, by considering variants of the input graph without node features and with undirected edges. For all models, we search the best architecture varying the number of hidden or convolutional layers $L \in \{1, ..., 8\}$, and the number of units per layer between 128 and 512; training is halted with early stopping with a patience of 200 epochs. Following [17], neighbor sampling is performed during training, sampling 50 neighbors for each node, which is slightly less than the average degree.

In Tab. 1 we report the results of our experiments, including AUROC and accuracy, with average and standard deviation over 3 different random seeds for weights initialization. Overall, the results show that we can predict robustness with high score for all the considered metrics. The best performances are achieved using the most complete information set available, i.e. graph structure (PPIN) with protein sequence embeddings. No statistically significant difference is observed by removing the interaction direction, an outcome that can be related to the fact that nodes are tightly connected, so that information propagation is not affected to a great degree. We must consider also that edge direction in PPINs often depends more on the experimental setting than the real interaction [14]. The high performance of the MLP baseline (80% F1) shows that the sequence embeddings have a good correlation with the robustness label, even though they are significantly worse than the DGN results (-5% F1). This aspect suggests the importance of having useful protein representa-

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

tions, even task-agnostic ones. One of the most interesting results is that the robustness property can be learned by DGN even without any information about the protein sequences. Without input embeddings, the DGN is forced to learn from exclusively from PPIN topology; we have observed in our experiments that performances consistently increase with layers L up to the network diameter. Furthermore, in this case the DGN seems to take more advantage of edge directionality (+1% F1), as it does not have information shortcuts provided by protein sequence embeddings. Even though performances are significantly lower (-8% F1) than the MLP baseline with embeddings, we can safely state that PPINs carry important dynamical information about underlying BPs.

As a final remark, we stress that the inference time of DGN is 100 ms, a 2 to 3 orders of magnitude speed-up compared to computing robustness via BP simulations, which on average require 46 seconds for each of the 315K simulations.

5 Conclusion and future directions

In this preliminary work, we have demonstrated the potential of addressing the prediction of dynamical properties in biological pathways via Deep Learning on graphs. Our experiments have confirmed that combining the information on protein sequences with the relationships between protein pairs of the protein-protein interaction network allows Deep Graph Networks to learn the robustness dynamical property as a link prediction task. The ablation study has further demonstrated the advantage of leveraging both protein structural information and PPIN topological information. When compared from the point of view of computational efficiency, our approach is $450\times$ faster than biological pathways simulations, unleashing the potential of unexploited dynamical information richness available in large scale biological networks. In future works, we will explore information from protein folding in addition to protein sequences, and more advanced DGN model architectures that have exhibited similar performances in preliminary experiments, to refine predictions also over species-specific PPINs and to broaden the range of target dynamical properties, offering a valid instrument to increase our understanding of complex biological systems.

Acknowledgments Research partly supported by PNRR, PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 1, and PNRR-M4C2 1.5, Ecosistema dell'Innovazione [ECS00000017], "Tuscany Health Ecosystem"- CUP [I53C22000780001]-Spoke 6, both funded by European Commission under the NextGeneration EU programme.

References

[1] Miguel A. Garcia-Campos et al. "Pathway Analysis: State of the Art". In: *Frontiers in Physiology* 6 (2015).

[2] Luca Marchetti et al. *Simulation Algorithms for Computational Systems Biology*. Texts in Theoretical Computer Science. An EATCS Series. Cham: Springer International Publishing, 2017.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

[3] Michele Fontanesi et al. "Exploiting the structure of biochemical pathways to investigate dynamical properties with neural networks for graphs". In: *Bioinformatics* 39.11 (2023).

[4] Lucia Nasti et al. "Formalizing a Notion of Concentration Robustness for Biochemical Networks". In: *Software Technologies: Applications and Foundations*. Ed. by Manuel Mazzara et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, pp. 81–97.

[5] Marc Santolini and Albert-Laszlo Barabasi. "Predicting perturbation patterns from the topology of biological networks". In: *Proceedings of the National Academy of Sciences of the United States of America* 115.27 (2018), E6375–E6383.

[6] Pasquale Bove et al. "Prediction of Dynamical Properties of Biochemical Pathways with Graph Neural Networks:" in: *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies*. Valletta, Malta: SCITEPRESS, 2020, pp. 32–43.

[7] Marco Podda et al. "Classification of Biochemical Pathway Robustness with Neural Networks for Graphs". In: *Biomedical Engineering Systems and Technologies*. Ed. by Xuesong Ye et al. Communications in Computer and Information Science. Cham: Springer International Publishing, 2021, pp. 215–239.

[8] Georgios N. Dimitrakopoulos et al. "How Far Are We from the Completion of the Human Protein Interactome Reconstruction?" In: *Biomolecules* 12.1 (2022).

[9] Hawoong Jeong et al. "The large-scale organization of metabolic networks". In: *Nature* 407.6804 (2000), pp. 651–654.

[10] Lun Hu et al. "A survey on computational models for predicting proteinâprotein interactions". In: *Briefings in Bioinformatics* 22.5 (2021), bbab036.

[11] Farzan Soleymani et al. "Protein-protein interaction prediction with deep learning: A comprehensive review". In: *Computational and Structural Biotechnology Journal* 20 (2022), pp. 5316–5341.

[12] Rahuman S Malik-Sheriff et al. "BioModels - 15 years of sharing computational models in life science". In: *Nucleic Acids Research* 48 (D1 2020), pp. D407–D415.

[13] The UniProt Consortium. "UniProt: the Universal Protein Knowledgebase in 2023". In: *Nucleic Acids Research* 51 (D1 2023), pp. D523–D531.

[14] Rose Oughtred et al. "The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions". In: *Protein Science: A Publication of the Protein Society* 30.1 (2021), pp. 187–200.

[15] Ahmed Elnaggar et al. *ProtTrans: Towards Cracking the Language of Life's Code Through Self-Supervised Deep Learning and High Performance Computing*. 2021. arXiv: `2007.06225[cs,stat]`.

[16] Davide Bacciu et al. "A gentle introduction to deep learning for graphs". In: *Neural Networks* 129 (2020), pp. 203–221.

[17] William L. Hamilton et al. "Inductive representation learning on large graphs". In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017, pp. 1025–1035.