Team 1: Prioritizing disease-associated genes using graph neural networks

Jędrzej Kubica

PhD student at University Grenoble Alpes, France

BioAI Hackathon at CeNT, University of Warsaw

May 12-15, 2025

Network medicine

Interactions underlie the cellular behavior

Biological networks

- protein-protein interactions

- gene regulatory networks

- disease-gene associations

- ...

Disease gene prioritization

- Known disease genes + biological networks

- How to discover **new disease-gene associations**?

Network propagation

The state-of-the-art methods:

- Random Walk with Restart (RWR)

- PageRank

- …

Challenges in network propagation

- Bias towards highly connected artificial protein "hubs"

- How to combine network topology with known diseases genes?

- Imbalanced datasets (network size vs. the number of disease genes)

- Validation of the highest-scoring genes

- Multi-omic data integration

Graph neural networks

Input layer   Hidden layers   Output layer

a graph

predicted node features
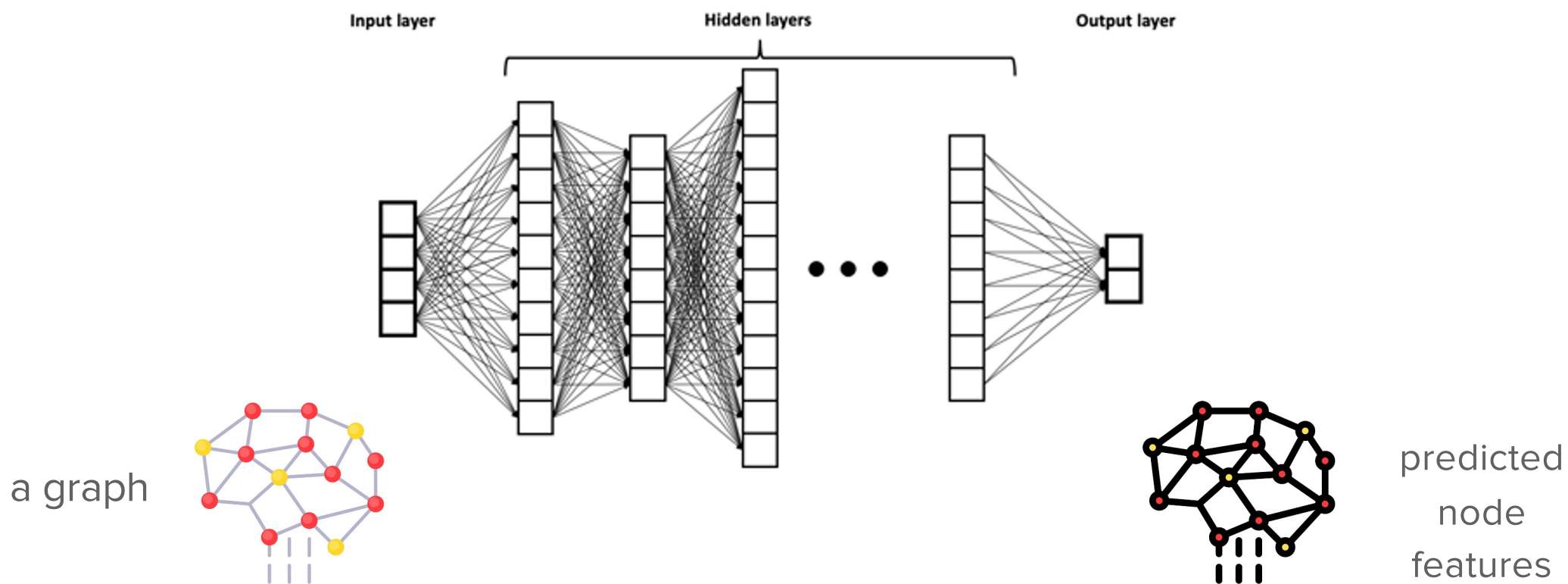
Graph neural networks

Example software:

- XGDAG (Mastropietro et al., 2023)

  https://github.com/GiDeCarlo/XGDAG
- geneDRAGNN (Altabaa et al., 2022)
  https://github.com/geneDRAGNN/geneDRAGNN

Challenges in graph neural networks

- Computationally expensive

- Model architecture design (e.g., number of layers)

- Validation of the highest-scoring disease genes

The aim of the hackathon project

Explore the **advantages** and **limitations** of graph neural networks compared to network propagation methods.

Methods

Input:

- biological network (protein-protein interactions)
- list of known disease genes
- *other data?

Desired outcome:

- gene scores
- comparison of performance: GNN vs. network propagation methods
- use XAI to compute explanation subgraphs

Have a great BioAI hackathon!

Tasks

- Compare XGDAG: https://github.com/GiDeCarIo/XGDAG and geneDRAGNN: https://github.com/geneDRAGNN/geneDRAGNN
- Compare GNNs (XGDAG & geneDRAGNN) with SoA network propagation methods
- **Learn about the advantages and limitations of graph neural networks compared to network propagation methods**
- Use Interactome_human.gml and known disease genes (Human Phenotype Ontology)

Future directions

- Use XAI strategies to compute explanation subgraphs
- Construct multi-layer networks for GNN imputation
- Improve the usability of GNNs for non-CS researchers

References

Altabaa, A., Huang, D., Byles-Ho, C., Khatib, H., Sosa, F., & Hu, T. (2022). GeneDRAGNN: Gene disease prioritization using graph neural networks. 2022 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 1–10. https://doi.org/10.1109/cibcb55180.2022.986304

Mastropietro, A., De Carlo, G., & Anagnostopoulos, A. (2023). XGDAG: explainable gene-disease associations via graph neural networks. Bioinformatics (Oxford, England), 39(8). https://doi.org/10.1093/bioinformatics/btad482

Shi, Y., Huang, Z., Wang, W., Zhong, H., Feng, S., & Sun, Y. (2020). Masked label prediction: Unified message passing model for semi-supervised classification. In arXiv [cs.LG]. arXiv. http://arxiv.org/abs/2009.03509