A107 · REPEL - Random Embedding Perturbation for Enhanced Learning of Protein Function

Presenter: Di Zhou — Tufts University

Authors: Di Zhou, Lenore Cowen, Kaiyi Wu, Xiaozhe Hu, Donna K. Slonim

Session: Day 2 · Fri Oct 2, 2026 · 2:15–4:15 PM

Protein function prediction from multiplex protein-protein association networks is a crucial approach to extending functional annotation. Current methods use embeddings of the heterogeneous network data that aim to place related proteins near each other in embedding space. However, such embeddings suffer from spurious protein proximity as well, reducing function prediction accuracy. Because heterogeneous input networks often have very different structures, it is hard to confidently declare proteins to be dissimilar using the network structure or the resulting embeddings. Here we address this problem with REPEL, a function prediction tool using a random graph augmentation method that applies a uniform weak force to push nodes apart. We assess this method on simulated networks with planted overlapping communities, as well as on real multiplex yeast and E.coli protein association networks. Surprisingly, we find that this method consistently improves protein function prediction over competing methods Mashup, deepNF, and BIONIC. The random repelling nature of the augmented graphs has a denoising effect on the learning process, distancing node pairs with spurious proximity while preserving true functional connections, thus increasing robustness. This graph augmentation principle may generalize to denoising and improving robustness in other graph-based learning algorithms.

REPEL - Random Embedding Perturbation for Enhanced Learning of Protein Function

Di Zhou[1], Lenore Cowen[1], Kaiyi Wu[2], Xiaozhe Hu[2], Donna K. Slonim[1]

[1] *Department of Computer Science, Tufts University, Medford, MA 02155, USA*
[2] *Department of Mathematics, Tufts University, Medford, MA 02155, USA*
Email: di.zhou@tufts.edu, cowen@cs.tufts.edu, kaiyiwu0124@gmail.com, xiaozhe.hu@tufts.edu, slonim@cs.tufts.edu

Protein function prediction from multiplex protein-protein association networks is a crucial approach to extending functional annotation. Current methods use embeddings of the heterogeneous network data that aim to place related proteins near each other in embedding space. However, such embeddings suffer from spurious protein proximity as well, reducing function prediction accuracy. Because heterogeneous input networks often have very different structures, it is hard to confidently declare proteins to be dissimilar using the network structure or the resulting embeddings. Here we address this problem with REPEL, a function prediction tool using a random graph augmentation method that applies a uniform weak force to push nodes apart. We assess this method on simulated networks with planted overlapping communities, as well as on real multiplex yeast and *E. coli* protein association networks. Surprisingly, we find that this method consistently improves protein function prediction over competing methods Mashup, deepNF, and BIONIC. The random repelling nature of the augmented graphs has a denoising effect on the learning process, distancing node pairs with spurious proximity while preserving true functional connections, thus increasing robustness. This graph augmentation principle may generalize to denoising and improving robustness in other graph-based learning algorithms.

Keywords: Protein Function Prediction; Protein-Protein Interaction, Denoising, Noise Injection for Regularization, Embedding, Multiplex Networks, Robustness

Code and data availability: https://github.com/TuftsBCB/REPEL.git