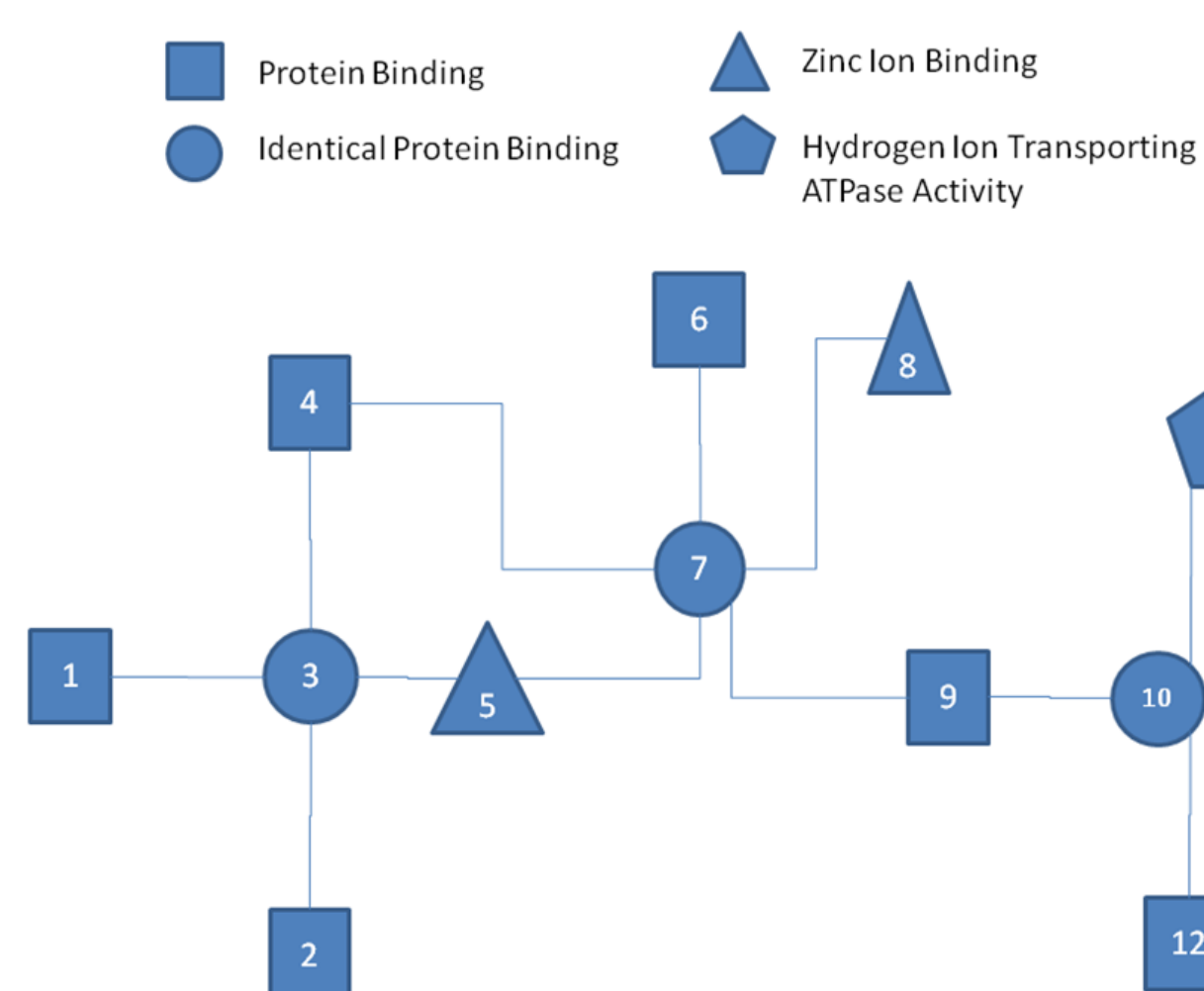


Overview

- Determining protein function is one of the most important problems in proteomics.
- High-throughput data collection is messy, and often has missing data.
- Therefore, many proteomes are missing functional annotations of their proteins.
- We present a pattern-matching based algorithm to predict values for these missing data

Current Methods

- General idea is to look at attribute values in the neighborhood of a query protein and use most frequently occurring values to determine attribute values for the query protein.
- Assumes highly assortative networks



Example Sub-Network

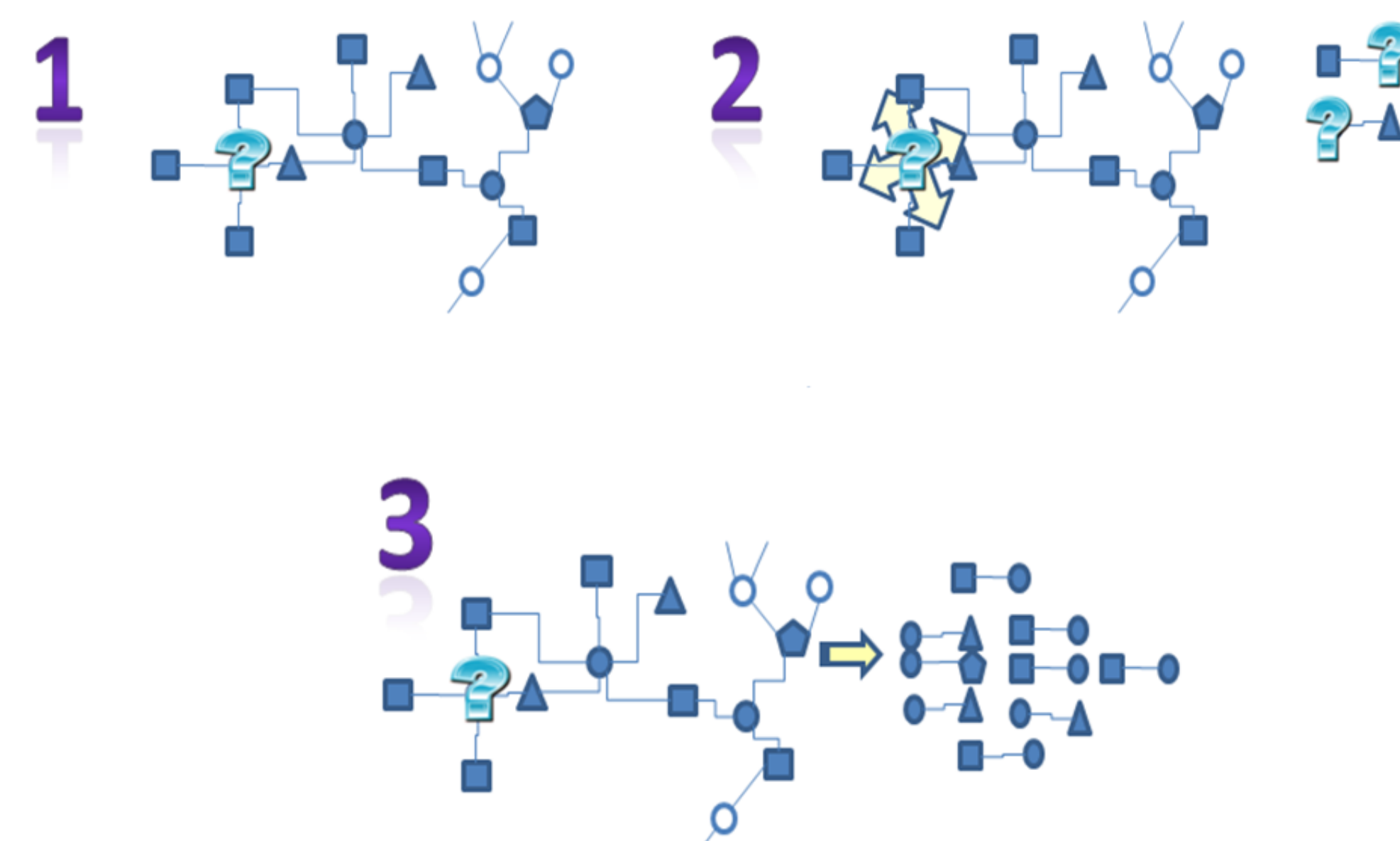
- Shows a protein-protein interaction sub-network from the earthworm proteome
- Let us assume that the protein labeled 3 is missing its annotation.
- If we only take into account annotation frequencies, we mistakenly guess protein 3 is a square
- The weakness of these methods is they don't adapt to structural features of the graph

Hypothesis

- Our hypothesis is that looking at the interactions of the query protein will give us a better estimate of its functions than simply by looking at its neighborhood.
- We explore interaction pair similarities to interpolate missing function.
- We treat the graph as the raw data for deducing similarities and supports for interpolation by performing a Bayesian inference analysis on interaction pairs

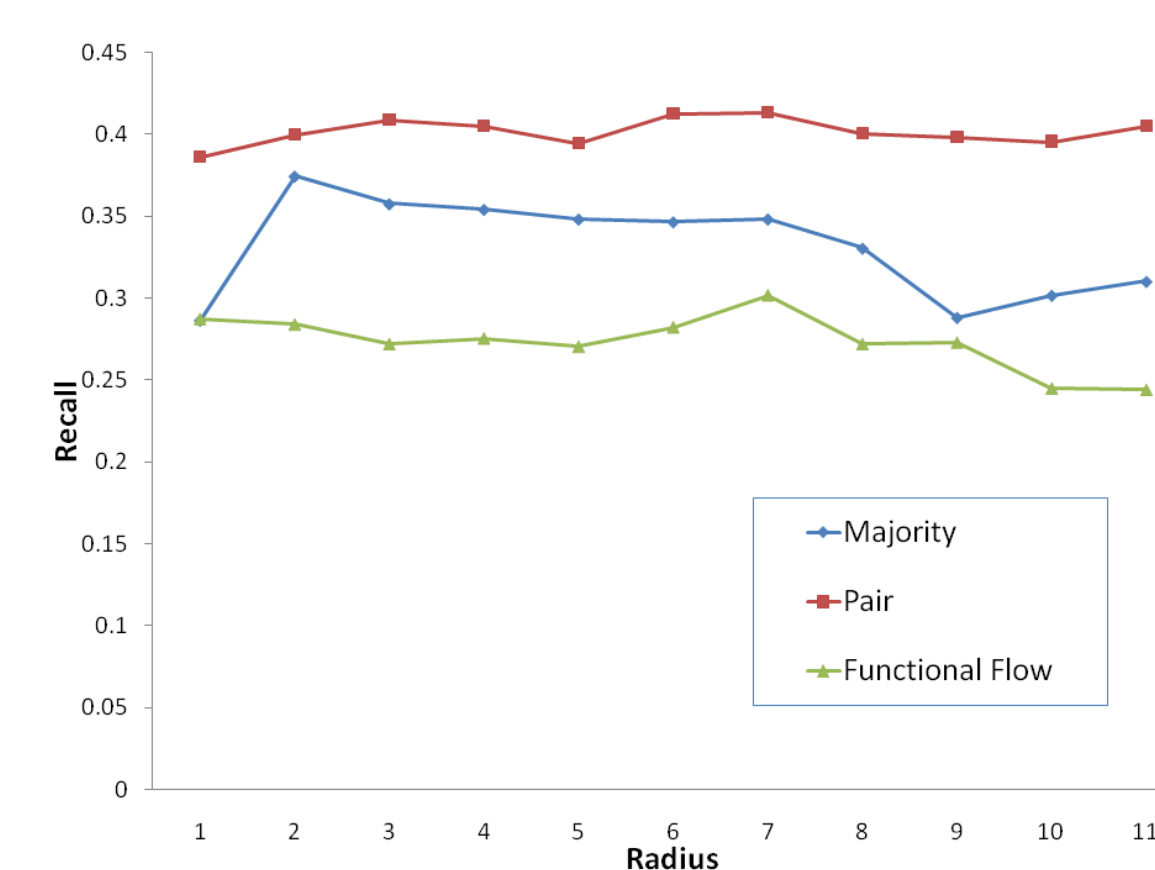
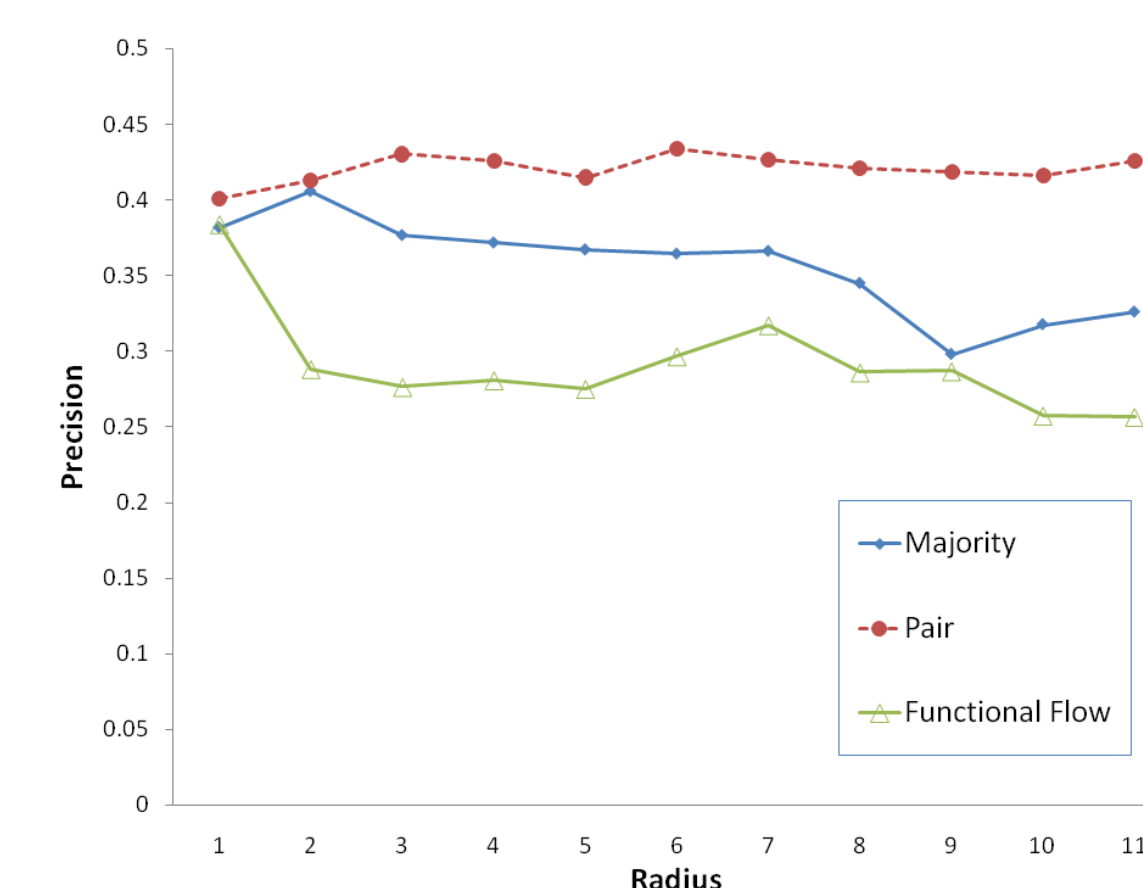
Our Method

- First we find all interaction pairs in which the query protein participates
- Then we look in the neighborhood of the query node to find other interaction pairs
- We use Bayesian inference analysis on the interaction pairs collected in the neighborhood to predict values for the missing function
- We store the prior and conditional probabilities of the observations seen in the neighborhood interaction pairs
- Before we store counts of the values seen, we flatten the interaction pairs



Experimental Method

- We use 2-fold cross validation
- Implementation in Java
- We compare with majority (Schwikowski et al.) and function flow methods (Nabieva et al.)
- Yeast Protein-Protein Interaction. The attribute we interpolate is function (48 possibilities).



Acknowledgements

This work was supported by National Institutes of Health: Grant #U54-DA021519