

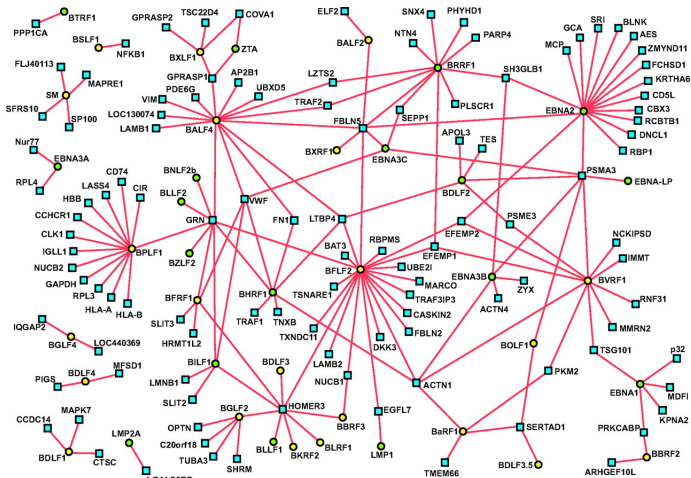
Collaboration-based Function Prediction in Protein-Protein Interaction networks

Hossein Rahmani

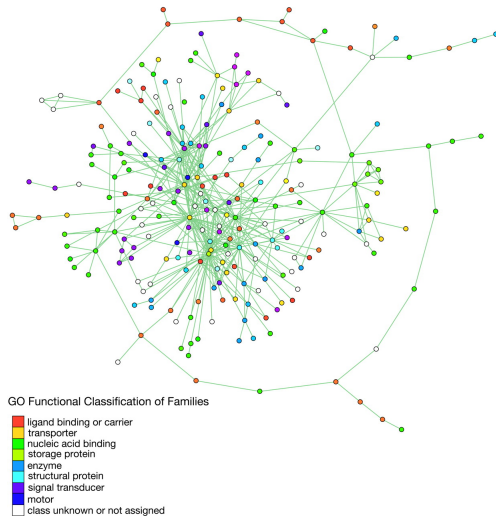
Joint work with: Hendrik Blockeel, Andreas Bender

October 2010

Protein-Protein Interaction (PPI) Networks



Function Prediction in PPI Networks



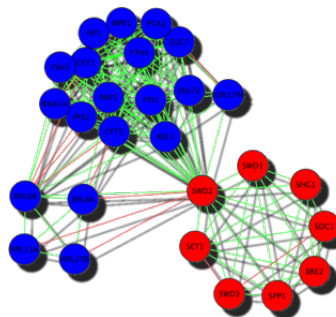
Outline

- Similarity based Function Prediction
- Proposed Methods:
 - Collaboration based Function Prediction
- Evaluation



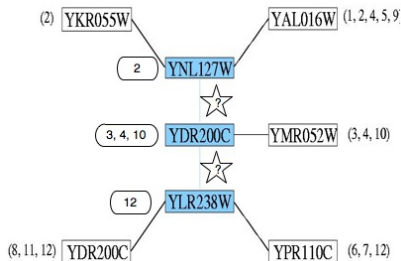
Similarity based Function Prediction Methods

- Assumption:
 - Interacting proteins have similar functions
- Optimization criteria:
 - Minimizing the number of interacting pairs of proteins with no common function
- Majority Rule
- Functional Clustering



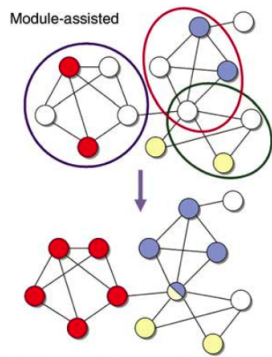
Majority Rule

- Predicted function: Most common function(s) among classified partners
- Problem: Links unclassified-unclassified proteins completely neglected



Functional Clustering

- Cluster the PPI network
- Predict the function of unclassified protein based on the cluster they belong to



Outline

- Similarity based Function Prediction
- **Proposed Methods:**
 - **Collaboration based Function Prediction**
- Evaluation

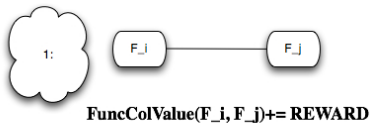
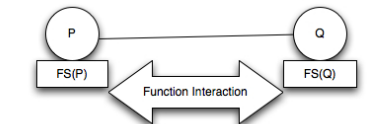


Collaboration based Function Prediction Method

- Main Idea: A biological process is the aggregation of each individual protein's functions
- Assumption: Topologically close proteins tend to have *collaborative functions*
- Collaborative functions: Pairs of functions that frequently interface with each other in different interacting proteins
- A Reinforcement Based Function Predictor (RL)
- SOM Based Function Predictor
- protein p :
 - { Function Set : $FS_p; FS_p(f_i)$
 - { Neighborhood Function Vector : $NB_p; NB_p(f_j)$



A Reinforcement Based Function Predictor



A Reinforcement Based Function Predictor

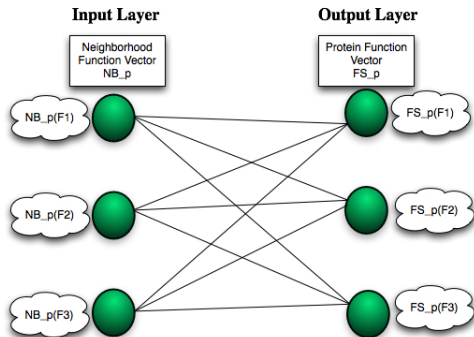
■ Prediction Time:

- Select candidate functions
- Rank candidate functions based on how well they collaborate with the neighborhood of unclassified protein p
- Formula (1) assigns a collaboration score to each candidate function f_c :

$$Score(f_c) = \sum_{\forall f_j \in F} NB_p(f_j) * FuncColVal(f_j, f_c) \quad (1)$$

SOM Based Function Predictor

- Self Organizing Map (SOM)
- $|inputNeurons| = |outputNeurons| = |F|$
- $inputNeuron(i) = NB_p(f_i)$
- $outputNeuron(i) = FS_p(f_i)$
- Tune Parameters
- Predict Functions



Outline

- Similarity based Function Prediction
- Proposed Methods:
 - Collaboration based Function Prediction
- Evaluation
 - Data sets
 - Parameter Tuning
 - Similarity V.S Collaboration



Datasets

■ Three Yeast Datasets: Krogan, VonMering and DIP-Core

	Number of Proteins	Number of Interactions
Von Mering	2401	22000
Krogan	2708	14246
DIP-Core	2388	4400

Table: Statistical information of datasets.

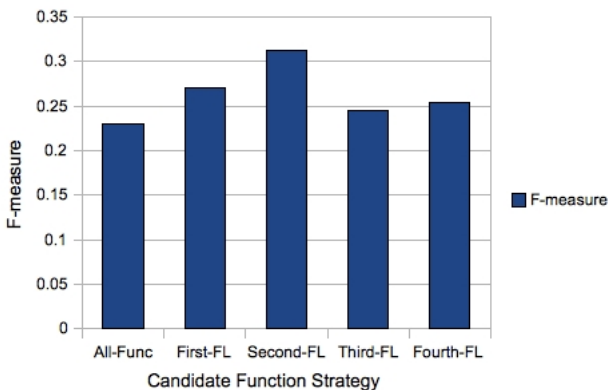
Parameter Tuning

- SOM method
 - Candidate Function Strategy
 - Decreasing Learning Rate
 - Termination Criteria
- Majority Rule
 - Wider Neighborhood Level



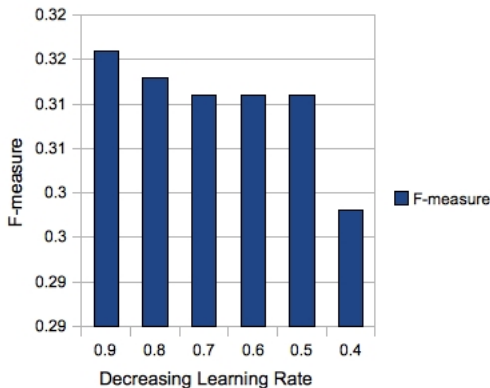
“Candidate Function Strategy” in SOM method

- Second function level produces the best result



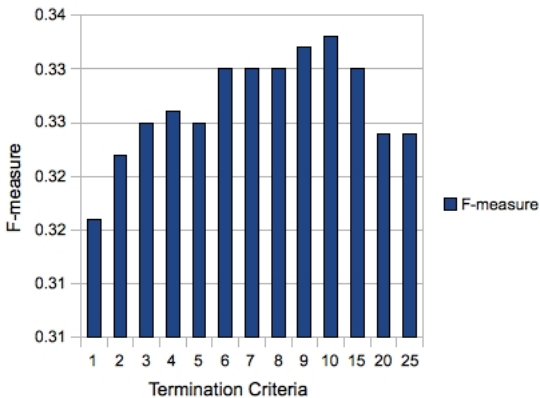
“Decreasing Learning Rate” in SOM method

- Fmeasure values maximize when DecLR equals to 0.9



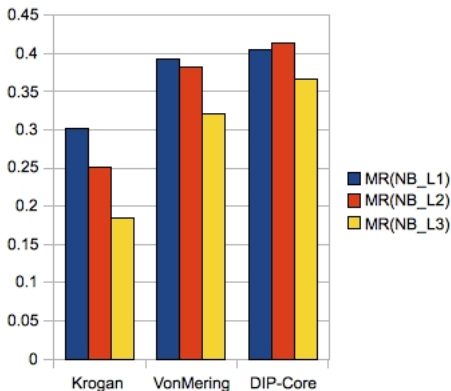
“Termination Criteria” in SOM method

- TC = 10 produces the best result



“Wider Neighborhood Level” in Majority Rule method

- NB-Li represents the 1-, 2- or 3-neighborhood of the protein
- Mostly, no improvement by considering wider neighborhood



Similarity V.S Collaboration

- Compare Collaboration based methods (SOM and RL) with Similarity based methods (MR and FC)
- Five different function levels
 - 11.02.01 (rRNA synthesis) Vs 11.02.03 (mRNA synthesis)

01 METABOLISM

01.01 amino acid metabolism

01.01.03 assimilation of ammonia, metabolism

01.01.03.01 metabolism of glutamine

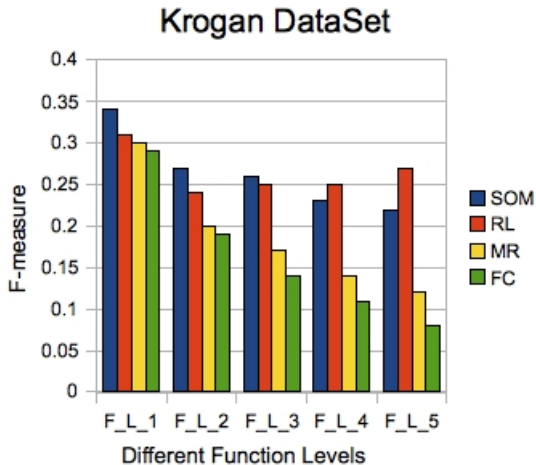
01.01.03.01.01 biosynthesis of glutamine

01.01.03.01.02 degradation of glutamine

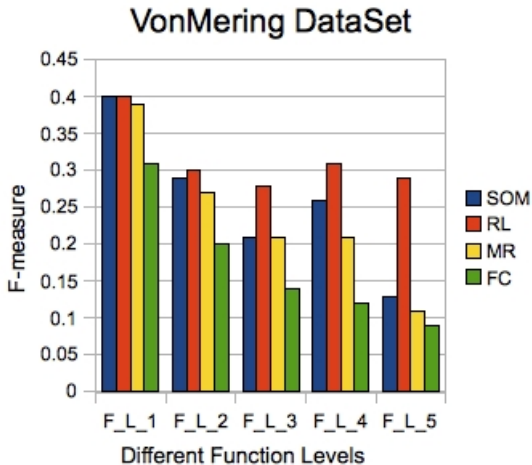
01.01.03.02 metabolism of glutamate

01.01.03.02.01 biosynthesis of glutamate

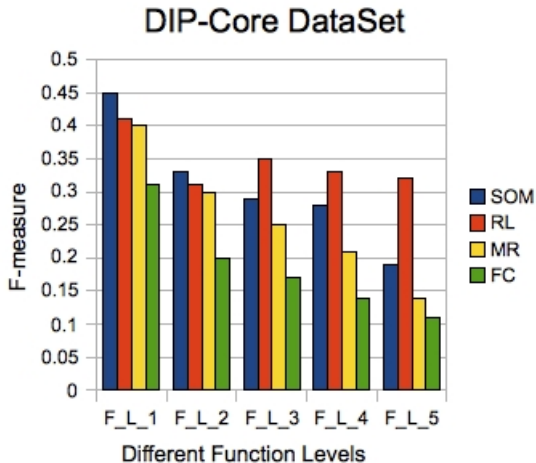
Similarity V.S Collaboration



Similarity V.S Collaboration

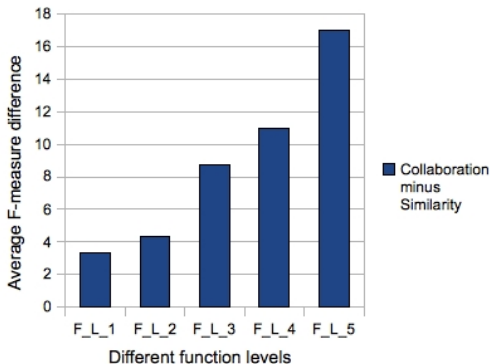


Similarity V.s Collaboration



Similarity V.s Collaboration

- In all three datasets, collaboration methods predicts functions more accurately than similarity based methods
- More detailed functions level \rightarrow More difference in performance



Conclusions

- Function prediction in PPI networks
- Similarity based Approaches
- Collaboration based Approaches
 - Reward-Punish (RL)
 - Self Organizing Map (SOM)
- Similarity V.s Collaboration
 - 3% to 17% improvement in F-measure values



Thanks!

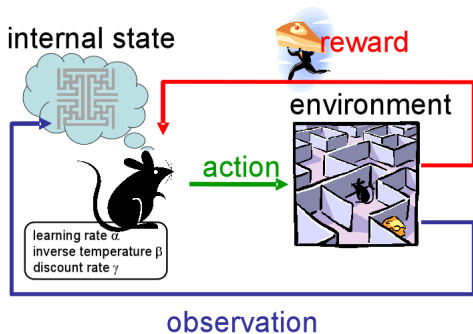


Reinforcement based Function Predictor

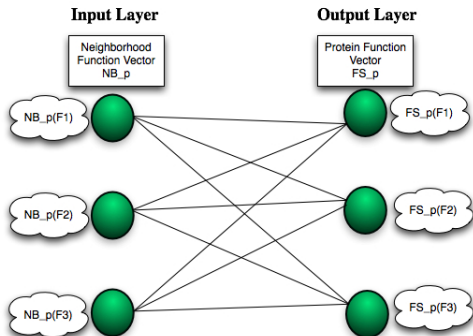
- $\forall f_i \in FS_p :$

$$FuncColVal(f_i, f_j) += \frac{NB_p(f_j) * R}{support(f_j)}$$
- $\forall f_i \in FS_p :$

$$FuncColVal(f_i, f_j) -= \frac{P}{support(f_j)}$$
- Candidate Function Strategies:
 - First Function Level Strategy
 - Second/Third/Fourth Function Level Strategy



SOM based Function Predictor



Algorithm 4.1: SOM TRAINING PHASE($LR, DecLR, TC$)

```

procedure SOM-TRAINING( $LR, DecLR, TC$ )
   $maxChangeInNetworkWeights \leftarrow 0;$ 
  repeat
    for each classified protein  $p \in P$ 
       $winnerNeuronSet = \{\}$ 
      build  $NB_p$ 
      for each  $f_i \in F$ 
        do  $inputNeuron(i) = NB_p(f_i)$ 
      for each  $f_j \in FS_p$ 
        do  $winnerNeuronSet = \bigcup_{f_i \in FS_p} \{f_i\}$ 
      apply Formula (2).
      update  $maxChangeInNetworkWeights$ 
       $LR \leftarrow LR * DecLR$ 
  until ( $maxChangeInNetworkWeights < TC$ )
  
```