

ROLE OF SYNTHETIC GENETIC INTERACTIONS IN UNDERSTANDING FUNCTIONAL INTERACTIONS AMONG PATHWAYS

S. Mohammadi G. Kollias A. Grama

Department of Computer Science
Purdue University

The Pacific Symposium on Biocomputing, 2012

OUTLINE

1 PRELIMINARY

- Models and Definitions
- Datasets

2 MOTIVATION

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

3 CONSTRUCTING KEGG CROSSTALK MAP

- Overview
- Methods
- Results and Discussions

OUTLINE

1 PRELIMINARY

- Models and Definitions
- Datasets

2 MOTIVATION

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

3 CONSTRUCTING KEGG CROSSTALK MAP

- Overview
- Methods
- Results and Discussions

GENETIC INTERACTOME

Double mutants exhibit unexpected phenotypes, as compared to joint single mutations.

DEFINITION

- **Negative Interactions**: more severe phenotype than expected
 - Also known as *aggravating* or *synergistic*
- **Positive Interactions**: Less severe phenotype than expected
 - Also known as *alleviating* or *epistatic*

Most commonly used:

PHENOTYPE : Growth rate

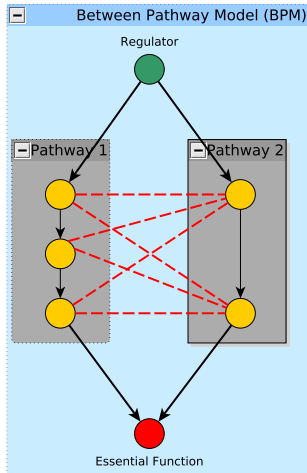
MODEL : Multiplicative null model

ORGANIZATION OF GENETIC INTERACTIONS

DEFINITION

- **Between-Pathway Model**
 - Among genes participating in redundant functions
- **Within-Pathway Model**
 - Among genes with additive effect
- **Indirect Effect**
 - Among genes with distant functions that are not directly related

BETWEEN-PATHWAY MODEL (BPM)



- Bi-clique structure
- Have been used to:
 - 1 Predict co-pathway membership of gene pairs
 - 2 Extract redundant pathways

OUTLINE

1 PRELIMINARY

- Models and Definitions
- Datasets

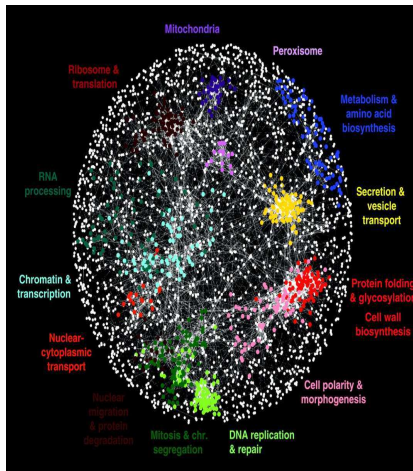
2 MOTIVATION

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

3 CONSTRUCTING KEGG CROSSTALK MAP

- Overview
- Methods
- Results and Discussions

THE GENETIC LANDSCAPE OF A CELL



Adopted from Costanzo et al., 2010

- Baker's yeast, *Saccharomyces cerevisiae*
- Synthetic Genetic Array (SGA)
- 1712 query genes
 - 1 1378 null alleles of non-essential genes
 - 2 334 hypomorphic or conditional alleles of essential genes
- 3885 array strains

FUNCTIONAL ANNOTATIONS



KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

KEGG2 PATHWAY BRUTE MODULE DISEASE DRUG GENES GENOME LIGAND DBGET

Select prefix Enter keywords

map Organism Go Help

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps (see new maps, change history, and last updates) representing our knowledge on the molecular interaction and reaction networks for:

- 0. Global Map
- 1. Metabolism
 - Carbohydrate Energy Lipid Nucleotide Amino acid Other amino acid Glycan Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Overview
- 2. Genetic Information Processing
- 3. Environmental Information Processing
- 4. Cellular Processes
- 5. Organismal Systems
- 6. Human Diseases

and also on the structure relationships (KEGG drug structure maps) in:

- 7. Drug Development

Pathway Mapping

- KEGG Pathway Database
- Annotations for 1026 genes in the experiment
- 96 Pathways
 - 80 pathways after filtering pathways with less than 10 genes.

OUTLINE

1 PRELIMINARY

- Models and Definitions
- Datasets

2 MOTIVATION

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

3 CONSTRUCTING KEGG CROSSTALK MAP

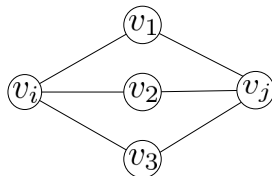
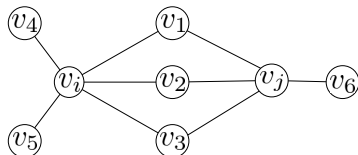
- Overview
- Methods
- Results and Discussions

LOCAL NEIGHBORHOOD SIMILARITY

A PREDICTOR OF CO-PATHWAY MEMBERSHIP

SIMILARITY PREDICTION METHODS

- 1 Number of Shared Neighbors
- 2 Congruence Score
- 3 Pearson Correlation of Interaction Profiles



Both v_i and v_j have three shared neighbors. However, in the first case their congruence score is almost 0.6, while in the second case it is approximately 2 (assuming a graph of size 10).

OUTLINE

1 PRELIMINARY

- Models and Definitions
- Datasets

2 MOTIVATION

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

3 CONSTRUCTING KEGG CROSSTALK MAP

- Overview
- Methods
- Results and Discussions

EVALUATING RANKING METHODS

Given a pathway P_A and a cut size (target set) l .

DEFINITION

$$\begin{aligned} P - \text{value}(X = k) &= \text{Prob}(k \leq X) \\ &= \text{HGT}(k|N, N_A, l) \\ &= \sum_{x=k}^{\min(N_A, l)} \frac{C(l, x) C(N - l, N_A - x)}{C(N, N_A)} \end{aligned}$$

X : Random variable denoting the number of true positives in a random sample, N : Total number of gene pairs, N_A : Number of gene pairs in pathway A , l : Size of target set

MINIMUM HYPERGEOMETRIC (mHG) SCORE

Target size unknown:

DEFINITION

The **Minimum HyperGeometric (mHG)** score

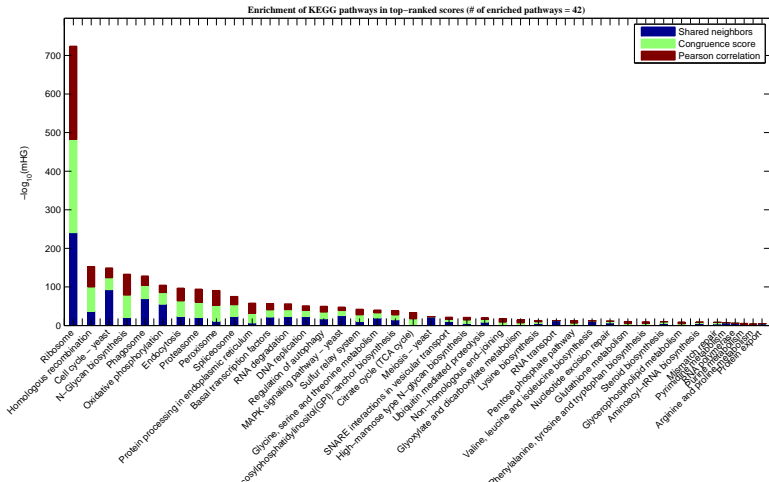
$$mHG(\lambda) = \min_{1 \leq l \leq N} HGT(b_l(\lambda); N, N_A, l),$$

where $b_l(\lambda) = \sum_{i=1}^l \lambda_i$

λ_i is 1 if both of the genes in the i^{th} ranked gene pair are members of P_A , and 0 otherwise.

- mHG Adjusted for Multiple Comparison

PREDICTIONS ARE NOT EQUALLY ACCURATE IN DIFFERENT KEGG PATHWAYS



OUTLINE

1 PRELIMINARY

- Models and Definitions
- Datasets

2 MOTIVATION

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

3 CONSTRUCTING KEGG CROSSTALK MAP

- Overview
- Methods
- Results and Discussions

HIGHLIGHTS

BASIC IDEA

Heterogeneous
performance of
co-pathway membership
predictions



Existence of specific
structure around
enriched pathways

- Decomposing neighborhood of each pathway
- Inferring lethal crosstalk among pathways

OUTLINE

1 PRELIMINARY

- Models and Definitions
- Datasets

2 MOTIVATION

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

3 CONSTRUCTING KEGG CROSSTALK MAP

- Overview
- **Methods**
- Results and Discussions

MODIFIED CONGRUENCE SCORE (MCS)

EVALUATING NEIGHBORHOOD OVERLAP OF GENE PAIRS WITH RESPECT TO A GIVEN PATHWAY

DEFINITION

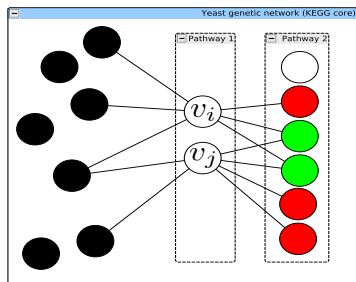
$$\begin{aligned}
 P - value(X = k_{ij}^B) &= Prob(k_{ij}^B \leq X) \\
 &= HGT(k_{ij}^B | n_B, d_i^B, d_j^B) \\
 &= \sum_{x=k_{ij}^B}^{\min(d_i^B, d_j^B)} \frac{C(d_j^B, x) C(n_B - d_j^B, d_i^B - x)}{C(n_B, d_i^B)}
 \end{aligned}$$

MCS is defined as $-\log_{10}$ of the P-value.

MODIFIED CONGRUENCE SCORE (MCS)

CONTINUED

EXAMPLE



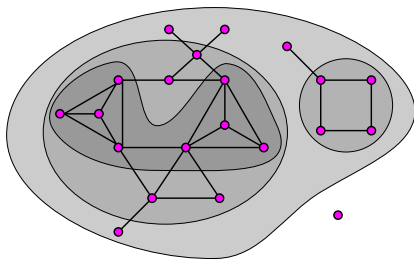
A sample neighborhood configuration for v_i and v_j . Here $n = 15$, $D_i = 6$, $D_j = 5$, $n_B = 6$, $d_i = 3$, $d_j = 4$ and $k = 2$.

CONSTRUCTING NEIGHBORHOOD OVERLAP GRAPH FOR A GIVEN PATHWAY PAIR

DEFINITION

The neighborhood overlap graph (NOG) of a given pathway P_A with respect to pathway P_B , denoted by $H_{A \rightarrow B} = (V_H, E_H)$, is an unweighted, undirected graph defined over same vertices as P_A . In this graph, there is a link between vertices v_i and v_j if the network structure around them with respect to P_B is statistically significant .

PRUNING NEIGHBORHOOD OVERLAP GRAPH, FINDING COHESIVE SUBGRAPHS, AND IDENTIFYING INTERACTION PORTS



Adopted from Batagelj and Zaversnik, 2002

- 1 Iterative peeling of K-shells
Pruning hairy components
- 2 Connected components in each core
- 3 Evaluating the significance of components
 - Evaluating significance using ER random graph model

OUTLINE

1 PRELIMINARY

- Models and Definitions
- Datasets

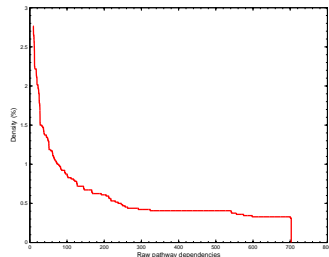
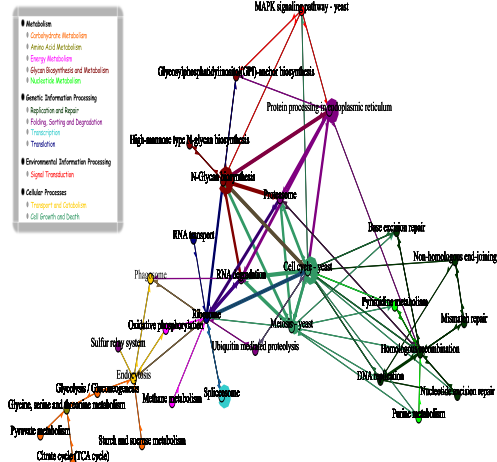
2 MOTIVATION

- Functional Similarity of Gene Pairs
- Performance Evaluation and Identifying Shortcomings

3 CONSTRUCTING KEGG CROSSTALK MAP

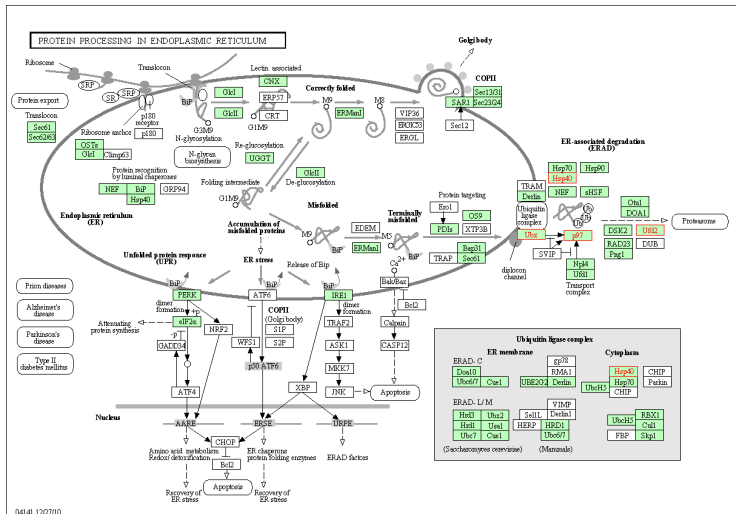
- Overview
- Methods
- Results and Discussions

KEGG CROSSTALK MAP



INTERACTION PORT CASE STUDY

CROSSTALK BETWEEN PROTEIN PROCESSING IN ER AND PROTEASOME



SUMMARY

- The **local neighborhood similarity** gives heterogeneous performance in predicting co-pathways membership of gene pairs.
- This phenomena is due to the specific structure around enriched pathways
- **Decomposing the neighborhood** around each pathway sheds light on the cellular machinery.
- Future works:
 - Analysing the hierarchy of ports instead of the most significant interaction port.
 - Using our methodology to uncover dependencies among functional pathways.

FOR FURTHER READING I



M. Costanzo et al.

The Genetic Landscape of a Cell
Science, 425 2010.



SJ. Dixon et al.

Systematic Mapping of Genetic Interaction Networks
Annual review of genetics, 43, 601 2009.



D. Segre et al.

Modular Epistasis in Yeast Metabolism
Nature Genetics, 37, 77 2005.



C.L. Tucker and S. Fields

Lethal Combinations
Nature Genetics, 35, 204 2003.