

# Topology of PPI Networks: Applications and Questions

**Limsoon Wong**  
(Works with Hon Nian Chua & Guimei Liu)



## Plan



- **Past works of my group on PPI networks**
  - PPI network cleansing based on PPI topology
  - PPI-based protein complex prediction
  - PPI-based protein function prediction
- **New directions?**
  - Application to lipid biology
  - Application to drug response/escape

# PPI Network Cleansing Based on PPI Topology



4



## Noise in PPI Networks

Experimental method category*	Number of interacting pairs	Co-localization <sup>b</sup> (%)	Co-cellular-role <sup>b</sup> (%)
All: All methods	9347	64	49
A: Small scale Y2H	1861	73	62
A0: GY2H Uetz <i>et al.</i> (published results)	956	66	45
A1: GY2H Uetz <i>et al.</i> (unpublished results)	516	53	33
A2: GY2H Ito <i>et al.</i> (core)	798	64	40
A3: GY2H Ito <i>et al.</i> (all)	3655	41	15
B: Physical methods	71	98	95
C: Genetic methods	1052	77	75
D1: Biochemical, <i>in vitro</i>	614	87	79
D2: Biochemical, chromatography	648	93	88
E1: Immunological, direct	1025	90	90
E2: Immunological, indirect	34	100	93
2M: Two different methods	2360	87	85
3M: Three different methods	1212	92	94
4M: Four different methods	570	95	93

Sprinzak et al., *JMB*, 327:919-923, 2003

Large disagreement betw methods

- High level of noise
- ⇒ Need to clean up before making inference on PPI networks

## Measures that correlate with function homogeneity and localization coherence



- Two proteins participating in same biological process are more likely to interact
- Two proteins in the same cellular compartments are more likely to interact



- CD-distance
- FS-Weight

CD-distance & FS-Weight: Based on concept that two proteins with many interaction partners in common are likely to be in same biological process & localize to the same compartment

Talk at Barbados Workshop, Bellairs, 19-25 April 2009. Copyright © 2009 by Limsoon Wong

## Iterated CD-Distance (Liu et al, 2008)



- Variant of CD-distance that penalizes proteins with few neighbors

$$wL(u,v) = \frac{2 |N_u \cap N_v|}{|N_u| + \lambda_u + |N_v| + \lambda_v}$$

$$\lambda_u = \max\left\{0, \frac{\sum_{x \in G} |N_x|}{|V|} - |N_u|\right\}, \lambda_v = \max\left\{0, \frac{\sum_{x \in G} |N_x|}{|V|} - |N_v|\right\}$$

- Suppose average degree is 4, then
  - Case 1:  $|N_u| = 1, |N_v| = 1, |N_u \cap N_v| = 1, wL(u,v) = 0.25$
  - Case 2:  $|N_u| = 10, |N_v| = 10, |N_u \cap N_v| = 10, wL(u,v) = 1$

Talk at Barbados Workshop, Bellairs, 19-25 April 2009. Copyright © 2009 by Limsoon Wong

## A thought...

$$wL(u,v) = \frac{2 | N_u \cap N_v |}{| N_u | + \lambda_u + | N_v | + \lambda_v}$$

- **Weight of interaction reflects its reliability**

⇒ **Can we get better results if we use this weight to re-calculate the score of other interactions?**

## Iterated CD-Distance (Liu et al, 2006)

- $wL^0(u,v) = 1$  if  $(u,v) \in G$ , otherwise  $wL^0(u,v) = 0$

$$\bullet \quad wL^1(u,v) = \frac{| N_u \cap N_v | + | N_u \cap N_v |}{| N_u | + \lambda_u + | N_v | + \lambda_v}$$

$$\bullet \quad wL^k(u,v) = \frac{\sum_{x \in N_u \cap N_v} wL^{k-1}(u,x) + \sum_{x \in N_u \cap N_v} wL^{k-1}(v,x)}{\sum_{x \in N_u} wL^{k-1}(u,x) + \lambda_u^k + \sum_{x \in N_v} wL^{k-1}(v,x) + \lambda_v^k}$$

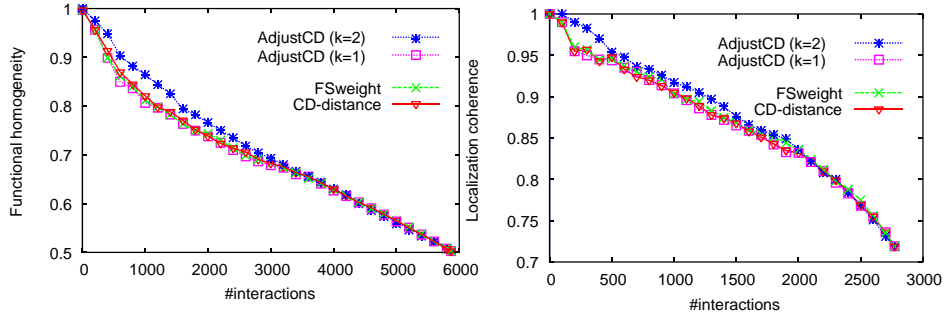
$$\bullet \quad \lambda_u^k = \max\left\{0, \frac{\sum_{x \in V} \sum_{y \in N_x} wL^{k-1}(x,y)}{|V|} - \sum_{x \in N_u} wL^{k-1}(u,x) \right\}$$

$$\bullet \quad \lambda_v^k = \max\left\{0, \frac{\sum_{x \in V} \sum_{y \in N_x} wL^{k-1}(x,y)}{|V|} - \sum_{x \in N_v} wL^{k-1}(v,x) \right\}$$



# Identifying False Positive PPIs

Cf. ave localization coherence of protein pairs in DIP < 5%  
ave localization coherence of PPI in DIP < 55%

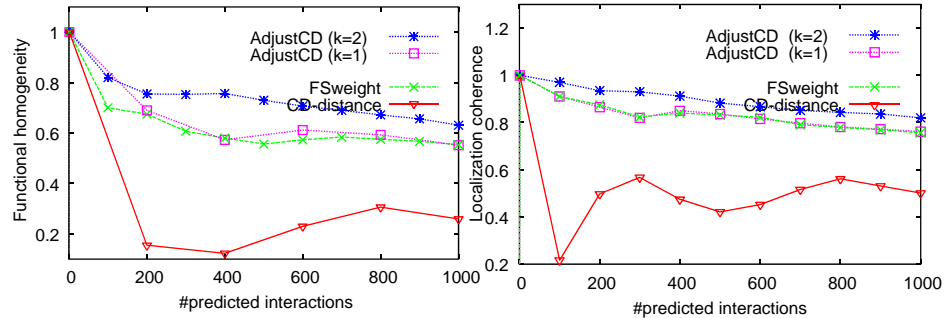


- Iterated CD-distance is an improvement over previous measures for assessing PPI reliability



# Identifying False Negative PPIs

Cf. ave localization coherence of protein pairs in DIP < 5%  
ave localization coherence of PPI in DIP < 55%



- Iterated CD-distance is an improvement over previous measures for predicting new PPIs

# PPI-Based Protein Complex Prediction

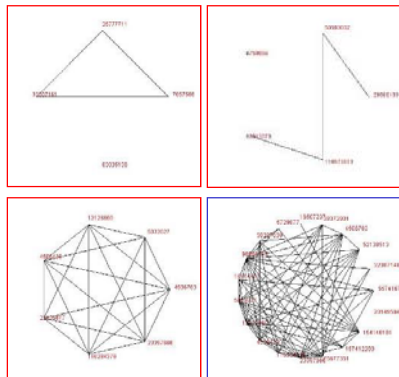


12

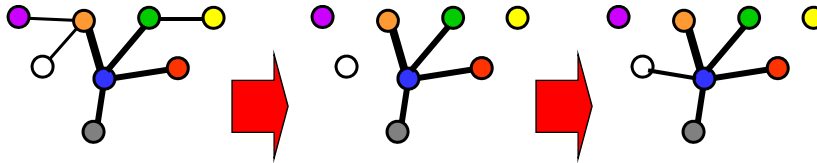
## PPI-Based Complex Prediction



- Recall & precision of protein complex prediction also have lots to be improved
- Does a “cleaner” PPI network help?
- How to capture non-ball-like complexes?
  - ⇒ Clique merging?
  - ⇒ Relative density?
  - ⇒ Core-n-attachment?



## Cleaning PPI Network

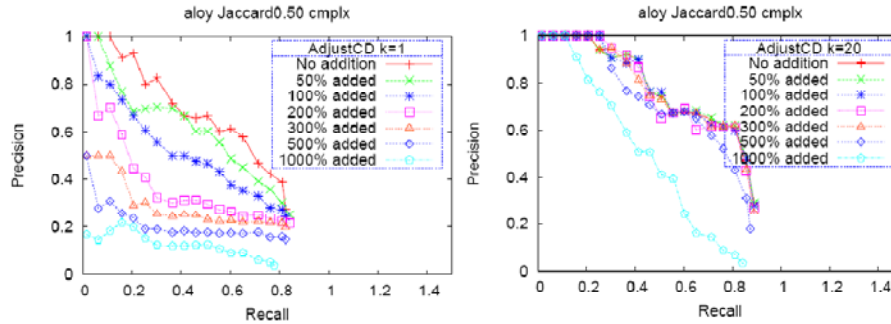


- **Modify existing PPI network as follow**
  - Remove interactions with low weight
  - Add interactions with high weight
- **Then run RNSC, MCODE, MCL, ..., as well as our own method CMC**

## CMC: Clustering of Maximal Cliques

- **Remove noise edges in input PPI network by discarding edges having low iterated CD-distance**
- **Augment input PPI network by addition of missing edges having high iterated CD-distance**
- **Predict protein complex by finding overlapping maximal cliques, and merging/removing them**
- **Score predicted complexes using cluster density weighted by iterated CD-distance**

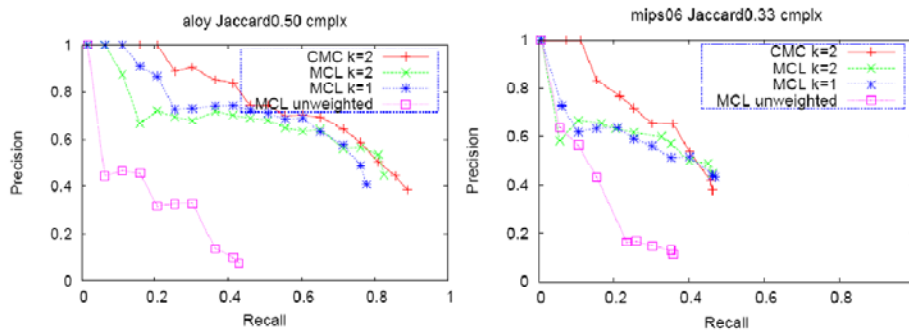
## Noise Tolerance of CMC



- If cleaning is done by iterating CD-distance 20 times, CMC can tolerate up to 500% noise in the PPI network!

Talk at Barbados Workshop, Bellairs, 19-25 April 2009. Copyright © 2009 by Limsoon Wong

## Effect of Cleansing on MCL



- MCL benefits significantly from cleansing too
- Ditto for other protein complex prediction methods

Talk at Barbados Workshop, Bellairs, 19-25 April 2009. Copyright © 2009 by Limsoon Wong



## Characteristics of Unmatched Clusters

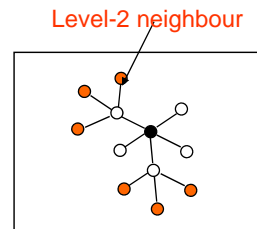
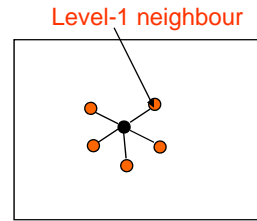
- At  $k = 2 \dots$
  - 85 clusters predicted by CMC do not match complexes in Aloy and MIPS
  - Localization coherence score  $\sim 90\%$
  - 65/85 have the same informative GO term annotated to  $> 50\%$  of proteins in the cluster
- ⇒ Likely to be real complexes

Talk at Barbados Workshop, Bellairs, 19-25 April 2009. Copyright © 2009 by Limsoon Wong

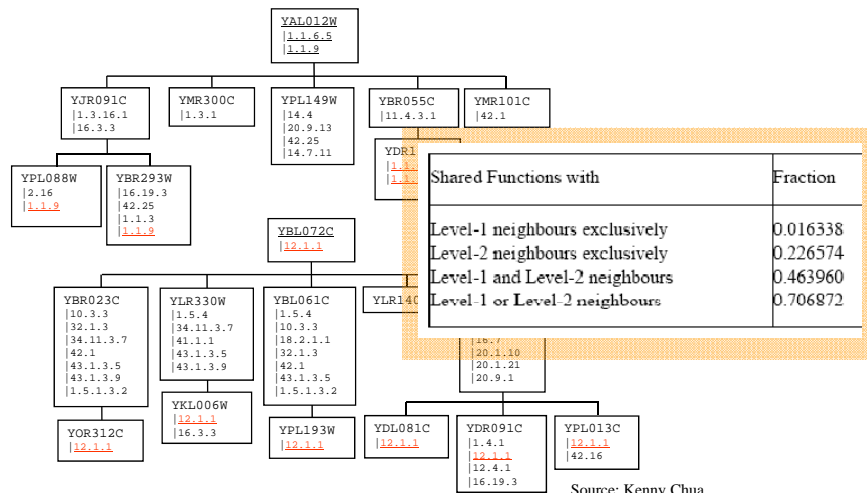
## PPI-Based Protein Function Prediction

## Functional Association Thru Interactions

- **Direct functional association:**
  - Interaction partners of a protein are likely to share functions w/ it
  - Proteins from the same pathways are likely to interact
- **Indirect functional association**
  - Proteins that share interaction partners with a protein may also likely to share functions w/ it
  - Proteins that have common biochemical, physical properties and/or subcellular localization are likely to bind to the same proteins



## Freq of Indirect Functional Association



## Functional Similarity Estimate: FS-Weighted Measure



- FS-weighted measure

$$S(u, v) = \frac{2|N_u \cap N_v|}{|N_u - N_v| + 2|N_u \cap N_v|} \times \frac{2|N_u \cap N_v|}{|N_v - N_u| + 2|N_u \cap N_v|}$$

- $N_k$  is the set of interacting partners of  $k$
- Greater weight given to similarity

⇒ Rewriting this as

$$S(u, v) = \frac{2X}{2X + Y} \times \frac{2X}{2X + Z}$$

## Functional Similarity Estimate: FS-Weighted Measure with Reliability



- Take reliability into consideration when computing FS-weighted measure:

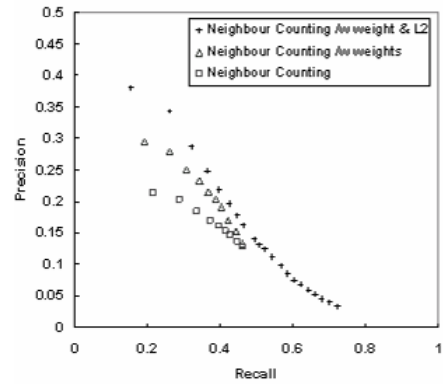
$$S_R(u, v) = \frac{2 \sum_{w \in (N_u \cap N_v)} r_{u,w} r_{v,w}}{\left( \sum_{w \in N_u - N_v} r_{u,w} + \sum_{w \in (N_u \cap N_v)} r_{u,w} (1 - r_{v,w}) \right) + 2 \sum_{w \in (N_u \cap N_v)} r_{u,w} r_{v,w}} \times \frac{2 \sum_{w \in (N_u \cap N_v)} r_{u,w} r_{v,w}}{\left( \sum_{w \in N_v - N_u} r_{v,w} + \sum_{w \in (N_u \cap N_v)} r_{v,w} (1 - r_{u,w}) \right) + 2 \sum_{w \in (N_u \cap N_v)} r_{u,w} r_{v,w}}$$

- $N_k$  is the set of interacting partners of  $k$
- $r_{u,w}$  is reliability weight of interaction betw  $u$  and  $v$

⇒ Rewriting

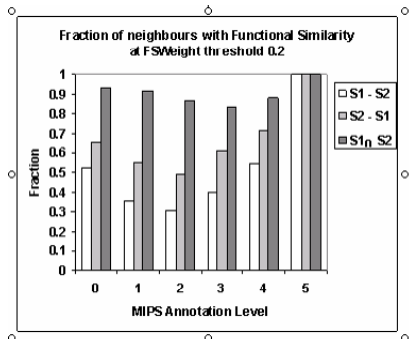
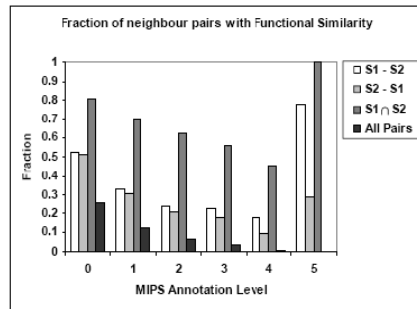
$$S(u, v) = \frac{2X}{2X + Y} \times \frac{2X}{2X + Z}$$

## Improvement to Prediction Power by Majority Voting



Considering only neighbours w/ FS weight > 0.2

## Improvement to Over-Rep of Functions in Neighbours



## Use L1 & L2 Neighbours for Prediction

- **FS-weighted Average**

$$f_x(u) = \frac{1}{Z} \left[ \lambda r_{int} \pi_x + \sum_{v \in N_u} \left( S_{TR}(u, v) \delta(v, x) + \sum_{w \in N_v} S_{TR}(u, w) \delta(w, x) \right) \right]$$

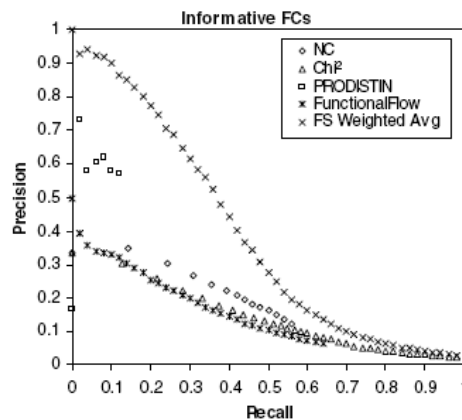
- $r_{int}$  is fraction of all interaction pairs sharing function
- $\lambda$  is weight of contribution of background freq
- $\delta(k, x) = 1$  if  $k$  has function  $x$ , 0 otherwise
- $N_k$  is the set of interacting partners of  $k$
- $\pi_x$  is freq of function  $x$  in the dataset
- $Z$  is sum of all weights

$$Z = 1 + \sum_{v \in N_u} \left( S_{TR}(u, v) + \sum_{w \in N_v} S_{TR}(u, w) \right)$$

Talk at Barbados Workshop, Bellairs, 19-25 April 2009. Copyright © 2009 by Limsoon Wong

## Performance of FS-Weighted Averaging

- **LOOCV comparison with Neighbour Counting, Chi-Square, PRODISTIN**



Talk at Barbados Workshop, Bellairs, 19-25 April 2009. Copyright © 2009 by Limsoon Wong

## Future Directions



28

### What we are planning next...



- **Lipid biology**
  - How to expand PPI networks with more info?
  - How to use it to infer proteins & complexes involved in lipid metabolism?
- **Drug response & escape**
  - How to augment PPI networks of microbacteria?
  - How to infer drug-response/escape routes?
  - How to cut off drug-escape routes?

## Readings

- H. N. Chua, et al. "Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions", *Bioinformatics*, 22:1623-1630, 2006
- H.N. Chua, et al. "Using Indirect Protein-Protein Interactions for Protein Complex Prediction", *JBCB*, 6(3):435--466, 2008
- H. N. Chua, L. Wong. "Increasing the Reliability of Protein Interactomes", *Drug Discovery Today*, 13(15/16):652--658, 2008
- G. Liu, et al. "Assessing and predicting protein interactions using both local and global network topological metrics", *Proc GIW2008*
- G. Liu, et al. "Complex Discovery from Weighted PPI Networks", submitted.

Any Question?