

# An Iterative Approach to Weighting PPI Networks and Its Impact on Complex Discovery

Guimei Liu      National University of Singapore  
Limsoon Wong    National University of Singapore  
Hon Nian Chua    Institute for Infocomm Research

1

## Outline

- An iterative weighting approach to identifying false positives and false negatives from high-throughput interaction data
  - Background
  - Existing weighting methods
  - The iterative weighting approach
  - Experiments
- Complex discovery from weighted PPI networks
  - Complex discovery method: maximal cliques
  - Experiments

2

## Background

- Protein-protein interactions play a critical role in most cellular processes and form the basis of biological mechanisms.
- High-throughput experimental techniques enable the study of protein-protein interactions at the proteome scale.
- However, high-throughput protein interaction data are often associated with high false positive and false negative rates
  - limitations of the associated experimental techniques
  - dynamic nature of protein interaction maps
  - ...

3

## Computational methods

- A weight is assigned to each interaction such that the higher the weight is, the more likely the interaction is true
- Various Information have been used
  - 3D protein structures
  - co-evolution
  - co-localization
  - gene fusion
  - literature
  - **network topology**
  - protein domains/motifs
  - ...

4

## Methods based on network topology

- Represents PPI networks as undirected graphs, where vertices are proteins, and edges represent interactions between proteins.
- IG1 [Saito et al. 2002]
  - The first one on evaluating the reliability of PPIs using solely PPI network topology
  - Mainly for PPI data generated by yeast-two-hybrid experiments
  - Based on the number of non-mutually interacting partners
- IG2 [Saito et al. 2002]
  - Uses 5 local network motifs
  - Performs better than IG1
- IRAP [Chen et al. 2005]
  - the collective reliability of the strongest alternative path between two proteins
  - Expensive to compute
- CD-distance [Brun et al. 2003] and FSWeight [Chua et al. 2006]
  - Based on the number of common neighbors of two proteins
  - Easy to compute
  - Outperforms the previous three methods on large PPI networks [Chen et al. 2006]

5

## CD-distance

- Given a pair of proteins ( $u, v$ ) in a PPI network  $G=(V, E)$ 
  - $N_u$ : the set of neighbors of  $u$  in  $G$
  - $N_v$ : the set of neighbors of  $v$  in  $G$
- $CD(u,v) = \frac{2|N_u \cap N_v|}{|N_u| + |N_v|}$
- Consider relative interaction size, not absolute interaction size
  - Case 1:  $|N_u| = 1, |N_v| = 1, |N_u \cap N_v| = 1, CD(u,v) = 1$
  - Case 2:  $|N_u| = 10, |N_v| = 10, |N_u \cap N_v| = 10, CD(u,v) = 1$

6

## FSWeight

- Try to overcome the weaknesses of CD-distance

$$\bullet \text{ FS}(u,v) = \frac{2|N_u \cap N_v|}{|N_u| + |N_u \cap N_v| + \lambda_u} \times \frac{2|N_u \cap N_v|}{|N_v| + |N_u \cap N_v| + \lambda_v}$$

Where  $\lambda_u$  and  $\lambda_v$  are used to penalize those proteins with very few neighbors

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

- Suppose the average degree is 4, then
  - Case 1:  $|N_u| = 1, |N_v| = 1, |N_u \cap N_v| = 1, \text{FS}(u,v) = 4/25 = 0.16$
  - Case 2:  $|N_u| = 10, |N_v| = 10, |N_u \cap N_v| = 10, \text{FS}(u,v) = 1$

7

## Differences between CD-distance and FSWeight

- Two differences between CD-distance and FSWeight
  - FSWeight penalizes those proteins with few neighbors
  - FSWeight assigns separate weight for each protein, and then takes the product of the two weights
- We consider another weighting method, which only penalize proteins with few neighbors

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

- Suppose the average degree is 4, then
  - Case 1:  $|N_u| = 1, |N_v| = 1, |N_u \cap N_v| = 1, \text{AdjustCD}(u,v) = 2/8 = 0.25$
  - Case 2:  $|N_u| = 10, |N_v| = 10, |N_u \cap N_v| = 10, \text{AdjustCD}(u,v) = 1$

8

## The iterative weighting approach

- Basic idea
  - The weight of an interaction reflects its reliability, so can we get better results if we use this weight to re-calculate the score of other interactions?
- Iterate CD-distance

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

$$\bullet w^0(u, v) = 1 \text{ if } (u, v) \in G, \text{ otherwise } w^0(u, v) = 0$$

$$\bullet w^1(u, v) = \frac{|N_u \cap N_v| + |N_u \cap N_v|}{|N_u| + |N_v|} = \text{CD}(u, v)$$

9

## Iterate AdjustCD

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

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

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

$$\bullet w^0(u, v) = 1 \text{ if } (u, v) \in G, \text{ otherwise } w^0(u, v) = 0$$

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

10

## Experiments

- PPI dataset: DIP yeast (dated 07-Oct-2007)
  - 4932 proteins and 17491 interactions
  - Core dataset: 6459 interactions
- Evaluation methods:
  - Functional homogeneity
    - Use Gene Ontology (GO) annotations
  - Localization coherence
    - use Gene Ontology (GO) annotations
  - 5-fold cross validation
    - DIP core dataset

11

## GO annotations

- Select only informative GO terms.
  - A GO term is informative if no less than 30 proteins are annotated with that term, and none of its descendant terms has at least 30 proteins
- 50 molecular function terms and 110 biological process terms
  - 3251 proteins and 11229 interactions have functional annotations.
- 42 cellular component terms
  - 1615 proteins and 4246 interactions have cellular component annotations

12

## Functional Homogeneity

- Given a set of protein pairs, its functional homogeneity is defined as

$$\frac{\text{\#protein pairs sharing same function annotation}}{\text{\#protein pairs that have function annotations}}$$

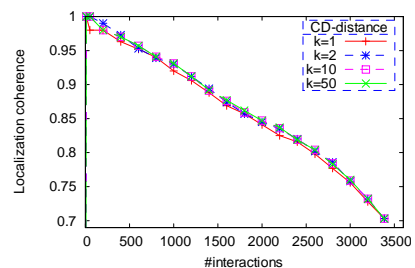
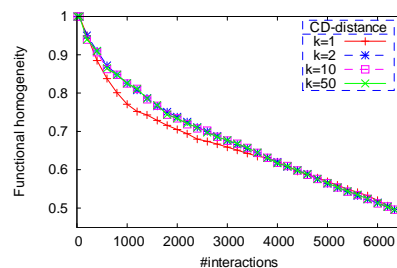
- Similarly, localization coherence is defined as

$$\frac{\text{\#protein pairs sharing same localization annotation}}{\text{\#protein pairs that have localization annotations}}$$

13

## Experiment 1: the effect of k

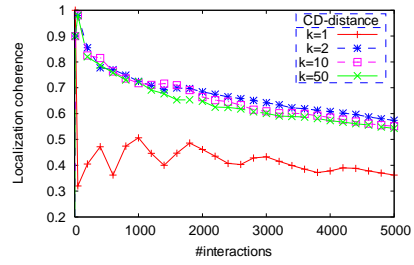
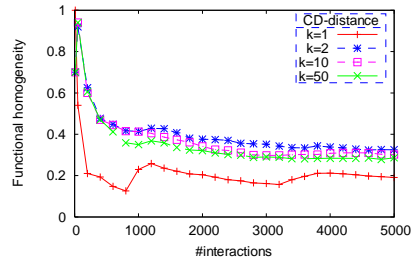
- CD-distance
- Assessing the reliability of PPIs in DIP dataset



14

## The effect of k

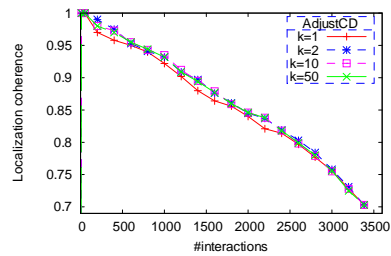
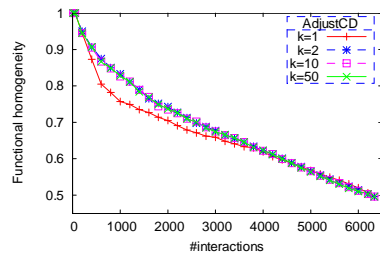
- CD-distance
- Predicting new PPIs



15

## The effect of k

- AdjustCD
- Assessing the reliability of PPIs in DIP dataset

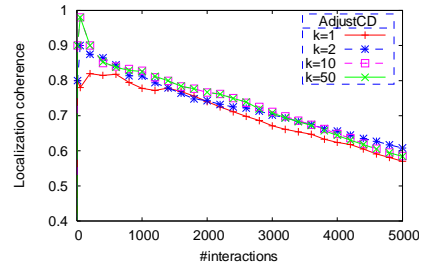
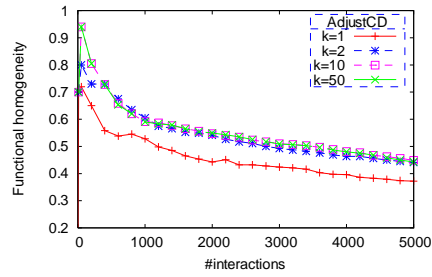


16



## The effect of k

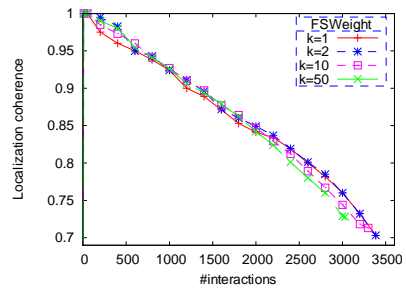
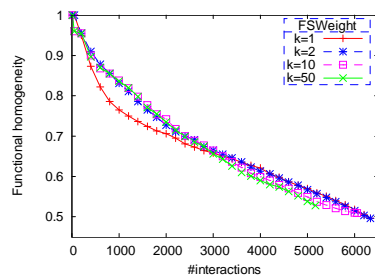
- AdjustCD
- Predicting new PPIs



17

## The effect of k

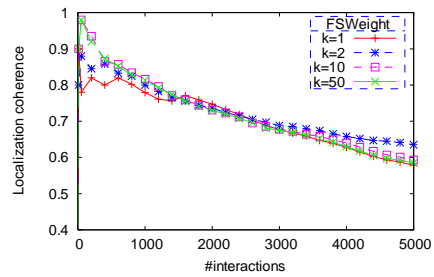
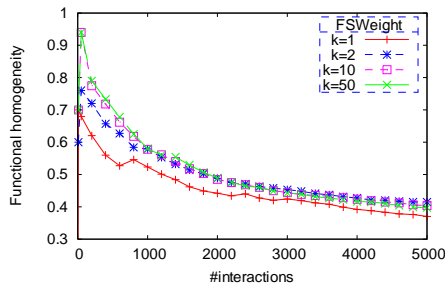
- FSWeight
- Assessing the reliability of PPIs in DIP dataset



18

## The effect of k

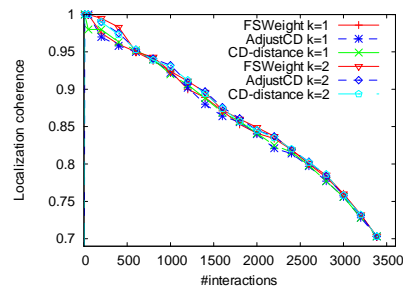
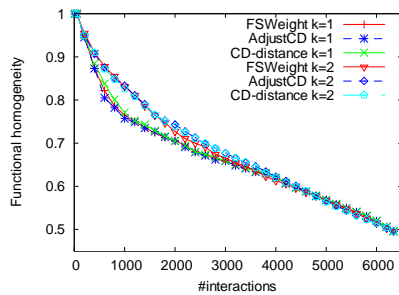
- FSWeight
- Predicting new PPIs



19

## Experiment 2: comparing different scoring methods

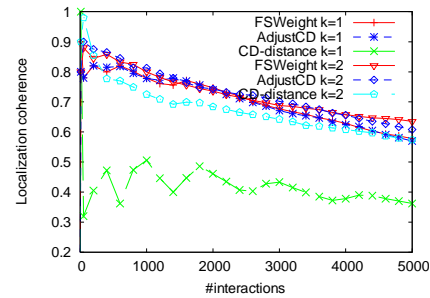
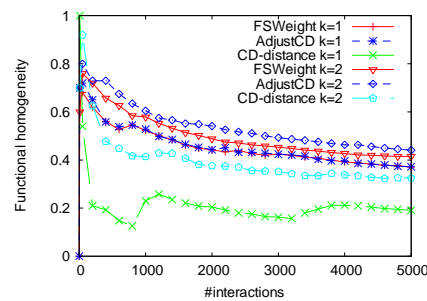
- Assessing the reliability of PPIs in DIP dataset



20

## Comparing different scoring methods

- Predicting new PPIs



21

## Experiment 3: 5-fold cross-validation

- Use the DIP core dataset as the golden standard
- Divide the proteins in the DIP full yeast dataset into 5 disjoint groups.
- For each group of proteins
  - Training data: remove the interactions between proteins in the group, and use the remaining interactions as training data
  - Testing data: all the protein pairs within this group
  - Correct answer PPIs: the pairs of proteins in the group that are in the core dataset

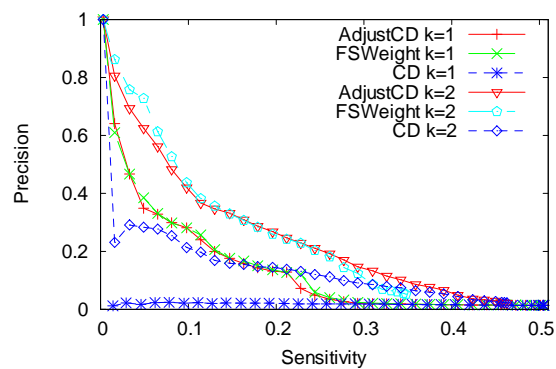
22

## 5-fold cross-validation

- Average number of proteins in 5 groups: 986
- Average number of interactions in 5 training datasets: 16723
- Average number of interactions in 5 testing datasets: 486591
- Average number of correct answer interactions: 307
- Measures:
  - sensitivity =  $TP / (TP + FN)$
  - specificity =  $TN / (TN + FP)$ 
    - #negatives  $\gg$  #positives, specificity is always very high
    - $>97.8\%$  for all scoring methods
  - precision =  $TP / (TP + FP)$

23

## 5-fold cross-validation



24

## Roadmap

- An iterative weighting approach to identifying false positives and false negatives from high-throughput interaction data
  - Background
  - Existing weighting methods
  - The iterative weighting approach
  - Experiments
- **Complex discovery from weighted PPI networks**
  - Complex discovery method: maximal cliques
  - Experiments

25

## Protein complexes

- Protein complexes are fundamental for understanding the organization of PPI networks
- With more and more PPI data available, it is possible to predict protein complexes from PPI networks.
- However, PPI networks are often noisy and incomplete, which makes it difficult to predict complexes accurately
- ⇒ Predicting complexes from cleaned and weighted PPI networks (Chua et al. 2008, JBCB)
- Our objective:
  - Study whether the iterative weighting approach can improve the performance of complex prediction algorithms

26

## Predicting complexes using maximal cliques

- Given a PPI network  $G$ , we first generate all the maximal cliques from  $G$  using the CLIQUES algorithm (Tomita et al. 2006)
- Calculate the score of each clique generated and rank cliques in descending order of their score
  - $\text{Score}(C) = \sum_{u,v \in C} w(u,v) / (|C| \cdot (|C| - 1) / 2)$
- Remove and merge highly overlapped cliques
  - Given two cliques  $C_1, C_2$  and  $\text{score}(C_1) \geq \text{score}(C_2)$ , if  $|C_1 \cap C_2| / |C_2| \geq \text{overlap\_thres}$ 
    - Case 1:  $\text{InterConnect}(C_1, C_2) \geq \text{merge\_thres}$ , then merge  $C_1$  and  $C_2$
    - Case 2:  $\text{InterConnect}(C_1, C_2) < \text{merge\_thres}$ , then remove  $C_2$

27

## Inter-connectivity score

- The inter-connectivity score of  $C_1$  and  $C_2$  measures the connectivity between the non-overlapping part of  $C_1$  and  $C_2$

$$\text{InterConnec}(C_1, C_2) = \sqrt{\frac{\sum_{x \in C_1 - C_2, y \in C_2} w(x, y)}{|C_1 - C_2| \times |C_2|} \times \frac{\sum_{x \in C_2 - C_1, y \in C_1} w(x, y)}{|C_2 - C_1| \times |C_1|}}$$

where  $w(x, y)$  is the weight of the edge between  $x$  and  $y$

28

## Experiments

- PPI dataset: union of six datasets
  - Gavin et al. 2002
  - Gavin et al. 2006
  - Ho et al. 2002
  - Krogan et al. 2006
  - Ito et al. 2001
  - Uetz et al. 2000
- #proteins: 4673
- #interactions: 20461
- #interactions with common neighbors: 11487
- We select the top 6000 interactions

29

## Reference complex sets

- Reference complex set (consider only size $\geq 4$ )
  - MIPS (dated 18-May-2006)
  - Aloy
    - P. Aloy, B. Bottcher, H. Ceulemans, C. Leutwein, C. Mellwig, S. Fischer, A.-C. Gavin, P. Bork, G. Superti-Furga, L. Serrano, and R. B. Russell. *Structure-based assembly of protein complexes in yeast*. *Science*, 303(5666): 2026–2029, 2004.

	#cplxes	#proteins	max size	avg size	median size	avg density	median density
MIPS	162	1171	95	14.93	9	0.408	0.318
Aloy	63	544	34	9.22	7	0.747	0.833

30

## Recall & precision at complex level

- Given a true complex  $C$  and a predicted complex  $S$ 
  - $\text{score}(S, C) = |S \cap C| / |S \cup C|$  (Jaccard index)
  - If  $\text{score}(S, C) \geq \text{match\_thres}$ , then we say  $S$  matches  $C$
- Given a set of complexes  $\{C_1, C_2, \dots, C_n\}$  and a set predicted complexes  $\{S_1, S_2, \dots, S_m\}$ , recall and precision at complex level are defined as

- $\text{Recall}_{\text{Cmplx}} = \frac{|\{C_i \mid \exists S_j, S_j \text{ matches } C_i\}|}{n}$

- $\text{Precision}_{\text{cmplx}} = \frac{|\{S_j \mid \exists C_i, C_i \text{ matches } S_j\}|}{m}$

31

## Recall & precision at protein level

- Given a set of complexes  $\{C_1, C_2, \dots, C_n\}$  and a set predicted complexes  $\{S_1, S_2, \dots, S_m\}$ , recall and precision at protein level are defined as

- $\text{Recall}_{\text{prtn}} = \frac{\sum_{i=1}^n \max\{|C_i \cap S_j| \mid j = 1, 2, \dots, m\}}{\sum_{i=1}^n |C_i|}$

- $\text{Precision}_{\text{prtn}} = \frac{\sum_{j=1}^m \max\{|S_j \cap C_i| \mid i = 1, 2, \dots, n\}}{\sum_{j=1}^m |S_j|}$

32



## Localization coherence

- Complexes can be formed only if its proteins have the same location in the cell
- Let  $\{L_1, L_2, \dots, L_k\}$  be a set of localization groups, where each location group contains proteins at the same location.
  - The co-localization score of a complex  $C$  is defined as the maximal fraction of proteins in  $C$  that are in the same location.
  - The co-localization score of a set of complexes  $\{C_1, C_2, \dots, C_n\}$  is defined as

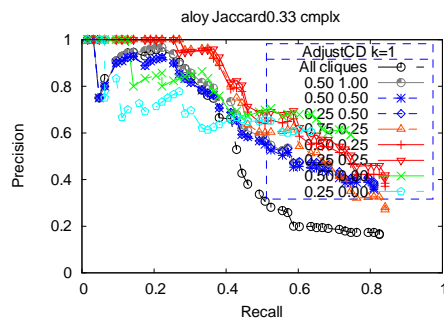
$$\text{loc\_score}(\{C_1, C_2, \dots, C_n\}) = \frac{\sum_{i=1}^n \max\{C_i \cap L_j \mid j = 1, 2, \dots, k\}}{\sum_{i=1}^n |\{p \in C_i \text{ and } p \text{ has annotations}\}|}$$

33

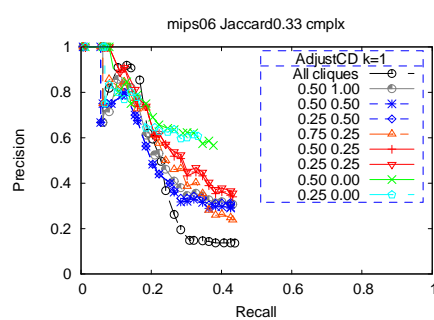
- Use informative localization GO terms

## Experiment 1: the effect of overlap\_thres and merge\_thres

- AdjustCD k=1
- Recall & precision at complex level: match\_thres=0.33



Aloy

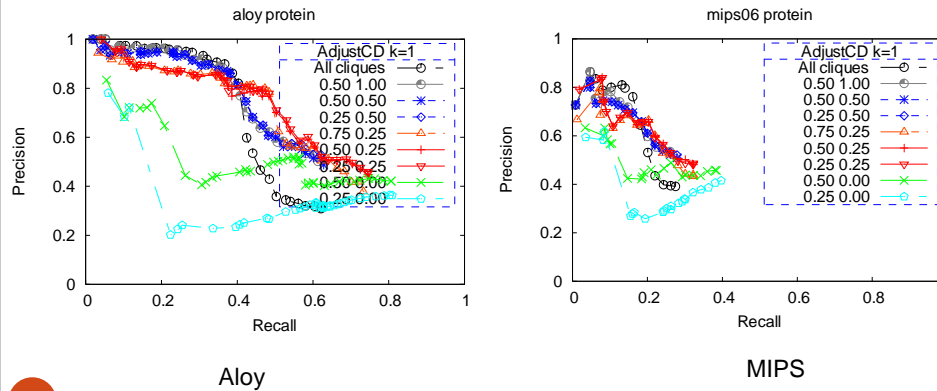


MIPS

34

## The effect of overlap\_thres and merge\_thres

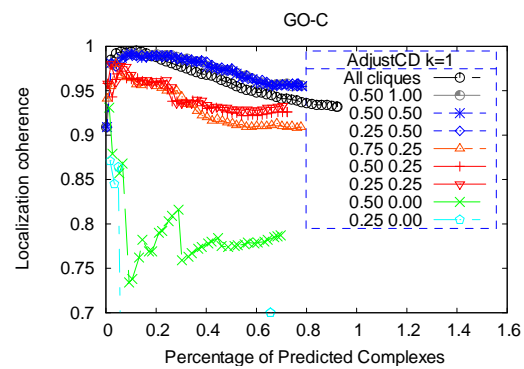
- AdjustCD k=1
- Recall & precision at protein level



35

## The effect of overlap\_thres and merge\_thres

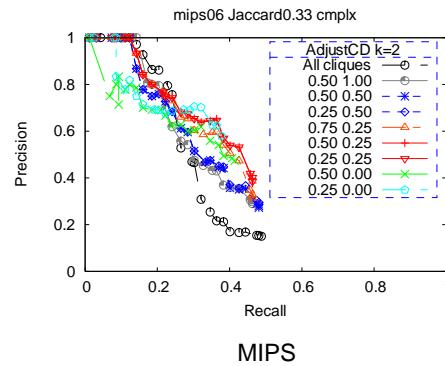
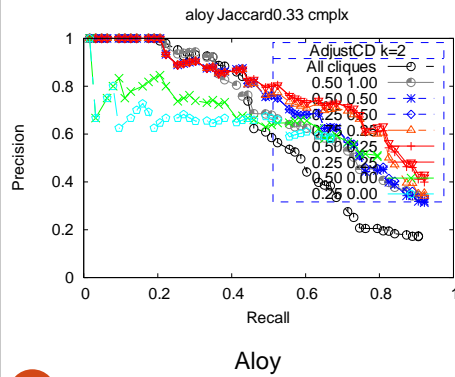
- AdjustCD k=1
- Localization coherence



36

## The effect of overlap\_thres and merge\_thres

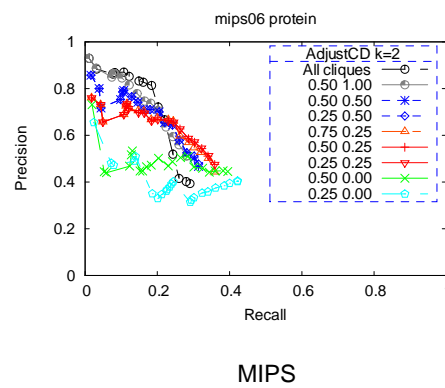
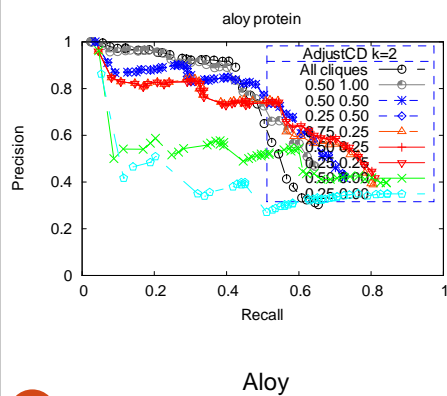
- AdjustCD k=2
- Recall & precision at complex level: match\_thres=0.33



37

## The effect of overlap\_thres and merge\_thres

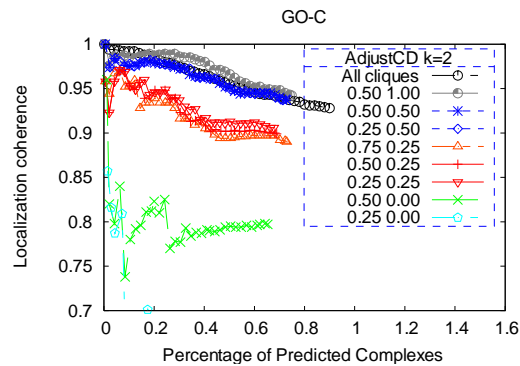
- AdjustCD k=2
- Recall & precision at protein level



38

## The effect of overlap\_thres and merge\_thres

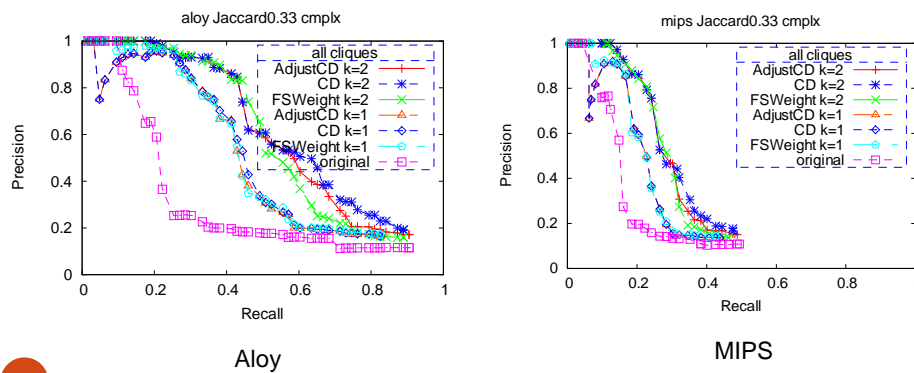
- AdjustCD k=2
- Localization coherence



39

## Experiment 2: comparing different PPI weighting methods

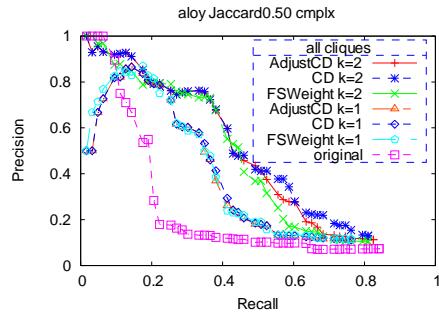
- All maximal cliques (without removing and merging overlapped cliques)
- Recall & precision at complex level: match\_thres=0.33



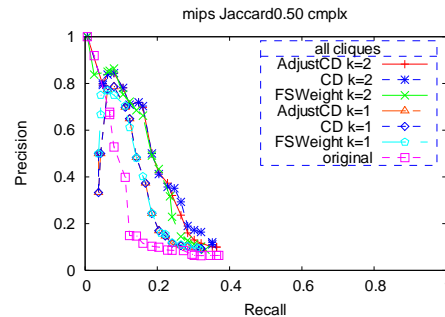
40

## Comparing different PPI weighting methods

- All maximal cliques (no removing and merging overlapped cliques)
- Recall & precision at complex level: match\_thres=0.50



Aloy

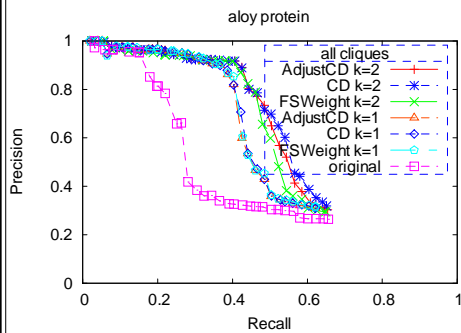


MIPS

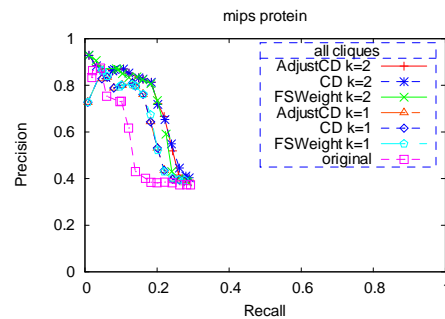
41

## Comparing different PPI weighting methods

- All maximal cliques (no removing and merging overlapped cliques)
- Recall & precision at protein level



Aloy

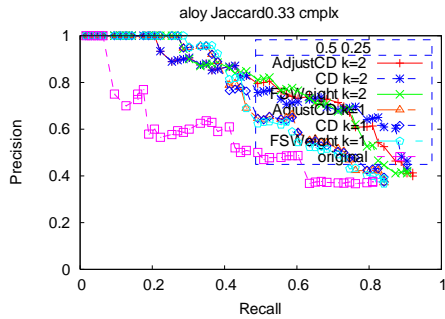


MIPS

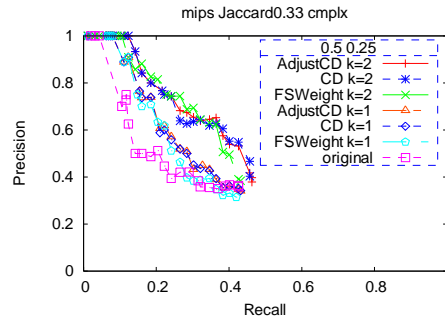
42

## Comparing different PPI weighting methods

- overlap\_thres=0.50, merge\_thres=0.25
- Recall & precision at complex level: match\_thres=0.33



Aloy

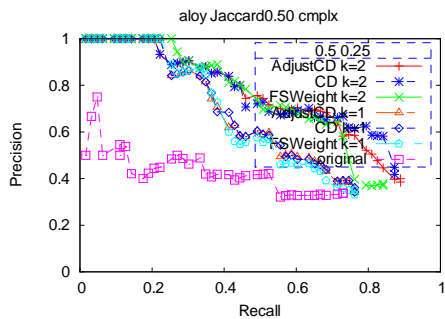


MIPS

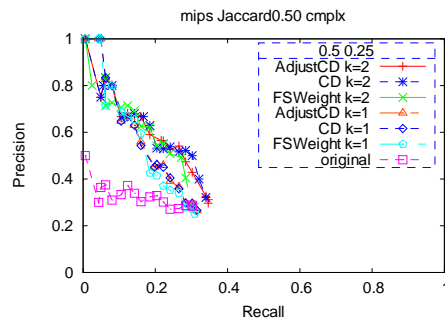
43

## Comparing different PPI weighting methods

- overlap\_thres=0.50, merge\_thres=0.25
- Recall & precision at complex level: match\_thres=0.50



Aloy

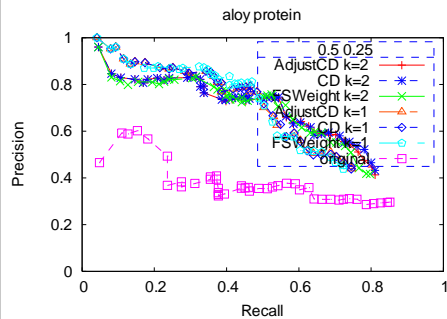


MIPS

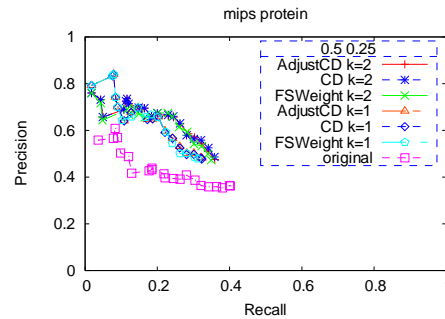
44

## Comparing different PPI weighting methods

- overlap\_thres=0.50, merge\_thres=0.25
- Recall & precision at protein level



Aloy

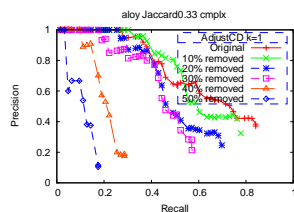


MIPS

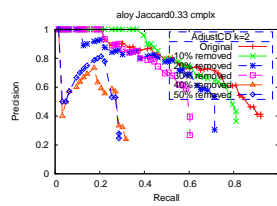
45

## Experiment 3: Robustness

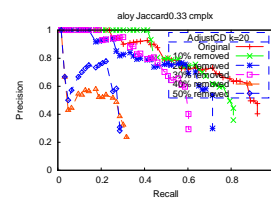
- AdjustCD
- overlap\_thres=0.50, merge\_thres=0.25
- Random removal of interactions
- Reference complex set: Aloy
- Recall & precision at complex level: match\_thres=0.33



k=1



k=2

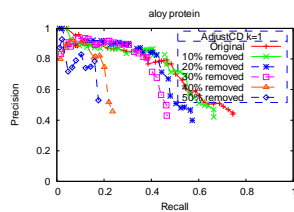


k=20

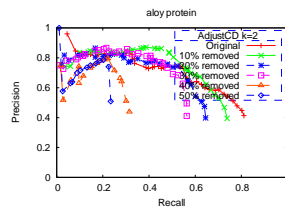
46

## Robustness: random removal

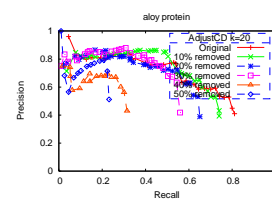
- AdjustCD
- overlap\_thres=0.50, merge\_thres=0.25
- Reference complex set: Aloy
- Recall & precision at protein level



k=1



k=2

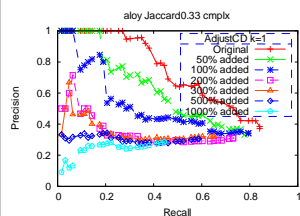


k=20

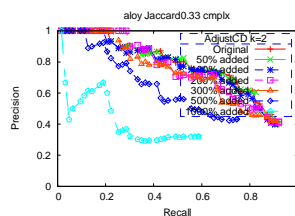
47

## Robustness: random addition

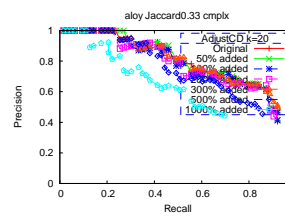
- AdjustCD
- overlap\_thres=0.50, merge\_thres=0.25
- Reference complex set: Aloy
- Recall & precision at complex level: match\_thres=0.33



k=1



k=2



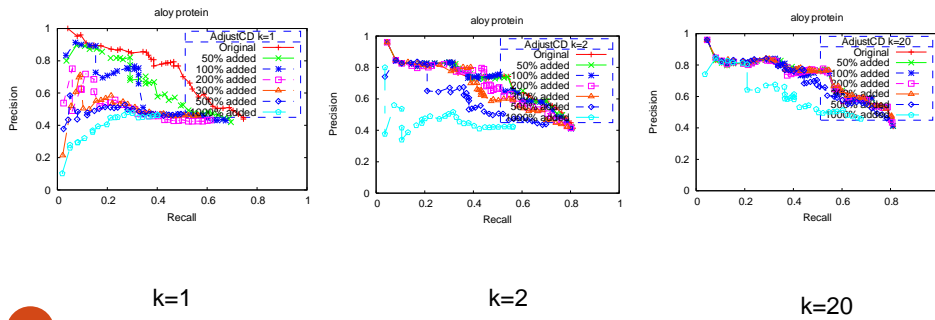
k=20

48



## Robustness: random addition

- AdjustCD
- overlap\_thres=0.50, merge\_thres=0.25
- Reference complex set: Aloy
- Recall & precision at protein level



49

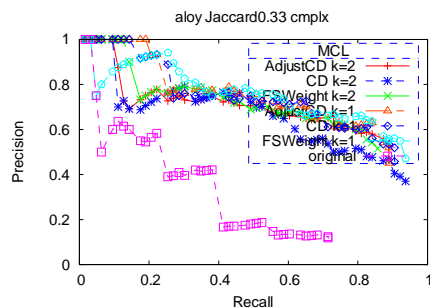
## Experiment 4: Impact of PPI weighting methods on MCL

- MCL (Stijn van Dongen, 2000)
  - Flow simulation
  - Do not allow overlap among clusters
  - Parameter setting:
    - Inflation: 1.8
    - Other parameters: default

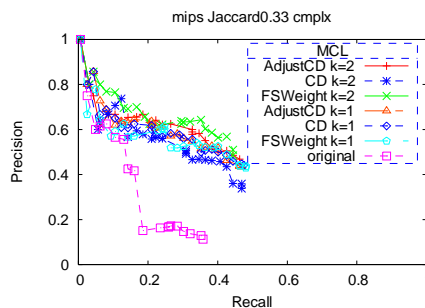
50

## Impact of PPI weighting methods to MCL

- Top 6000 interactions
- Recall & precision at complex level: match\_thres=0.33



Aloy

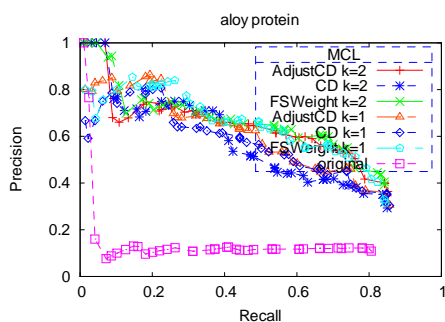


MIPS

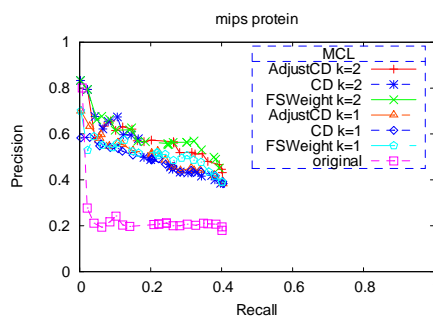
51

## Impact of PPI weighting methods to MCL

- Recall & precision at protein level



Aloy

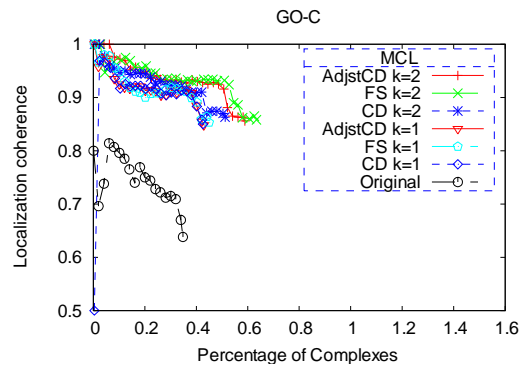


MIPS

52

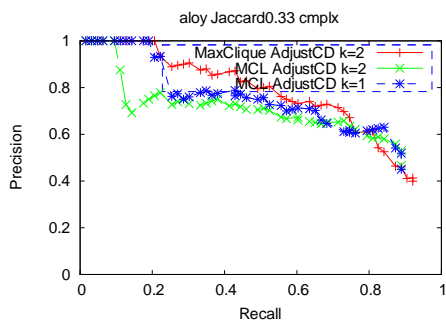
# Impact of PPI weighting methods to MCL

- Localization coherence

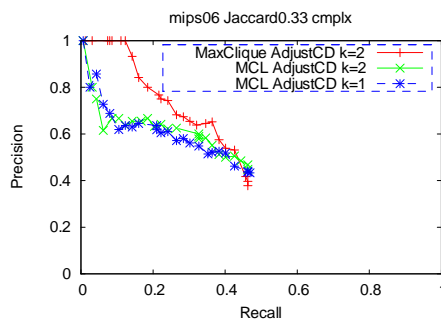


# Experiment 5: comparing different clustering methods

- Recall & precision at complex level: match\_thres=0.33



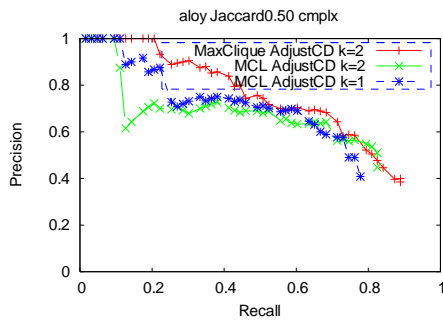
Aloy



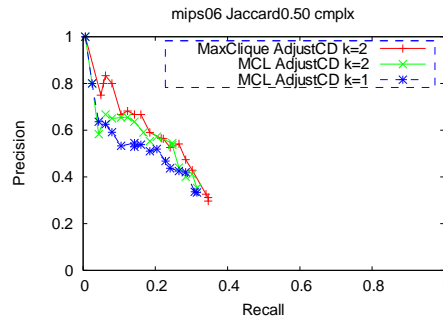
MIPS

# Comparing different clustering methods

- Recall & precision at complex level: match\_thres=0.50



Aloy

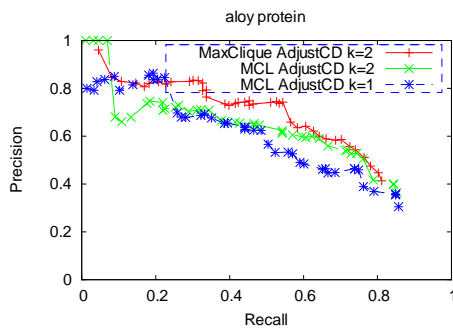


MIPS

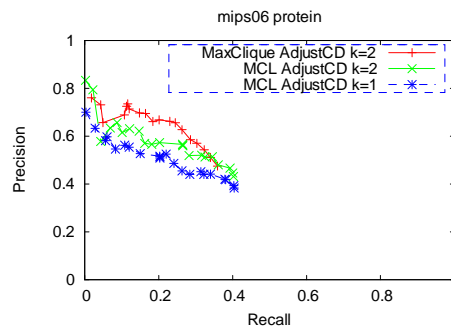
55

# Comparing different clustering methods

- Recall & precision at protein level



Aloy

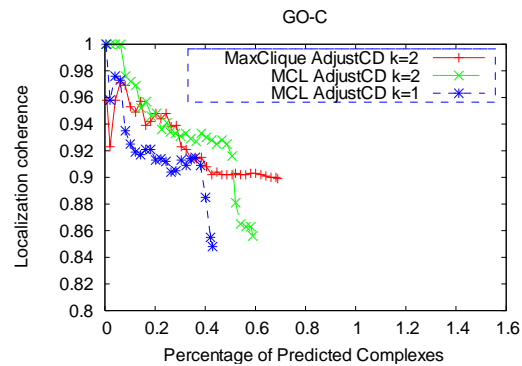


MIPS

56

## Comparing different clustering methods

- Localization coherence



57

## Summary and Conclusion

- Assessing the reliability of PPIs
  - CD-distance, FSWeight, and AdjustCD show similar performance
  - The iterative approach can improve the performance slightly
- Predicting new interactions
  - FSWeight and AdjustCD perform much better than CD-distance
  - The iterative approach can improve the performance significantly

58

## Summary and Conclusion

- Complex discovery
  - Dense complexes are easier to detect than sparse complexes
  - For the maximal clique based algorithm, the iterative approach improves the performance considerably
  - For the MCL algorithm, the iterative approach has no obvious improvement
  - The iterative approach is more robust to random noise

## References

1. P. Aloy, B. Bottcher, H. Ceulemans, C. Leutwein, C. Mellwig, S. Fischer, A.-C. Gavin, P. Bork, G. Superti-Furga, L. Serrano, and R. B. Russell. *Structure-based assembly of protein complexes in yeast*. *Science*, 303(5666):2026–2029, 2004.
2. Brun C, Chevenet F, Martin D, Wojcik J, Guenoche A, and Jacq B, *Functional classification of proteins for the prediction of cellular function from a protein-protein interaction network*. *Genome Biology*, 5(1):R6, 2003.
3. Chen J, Hsu W, Lee ML, and Ng SK, *Discovering reliable protein interactions from high-throughput experimental data using network topology*. *Artificial Intelligence in Medicine*, 35(1-2):37–47, 2005.
4. Chen J, Chua HN, Hsu W, Lee ML, Ng SK, Saito R, SungWK, andWong L, *Increasing confidence of protein-protein inteactomes*. In *Proc. of 17th International Conference on Genome Informatics*, pp. 284–297, 2006.
5. Chua HN., Sung WK., and Wong L., *Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions*. *Bioinformatics*, 22(13):1623-30, 2006.
6. Chua HN, Ning K, Sung WK, Leong HW, Wong L. *Using indirect protein-protein interactions for protein complex predication*. *Journal of Bioinformatics and Computational Biology*, 6(3):435--466, June 2008.

7. Stijn van Dongen, *Graph Clustering by Flow Simulation*. PhD thesis, University of Utrecht, May 2000.
8. A.-C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dumpelfeld, and et al. *Proteome survey reveals modularity of the yeast cell machinery*. *Nature*, 440:631–636, 2006.
9. A.-C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A.-M. Michon, C.-M. Cruciat, and et al. *Functional organization of the yeast proteome by systematic analysis of protein complexes*. *Nature*, 415:141–147, 2002.
10. Y. Ho, A. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S.-L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, and et al. *Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry*. *Nature*, 415:180–183, 2002.
11. T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki. *A comprehensive two-hybrid analysis to explore the yeast protein interactome*. *Proceedings of the National Academy of Sciences of the United States of America*, 98(8):4569–4574, 2001.
12. N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, and et al. *Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae**

13. Saito R, Suzuki H, and Hayashizaki Y, *Interaction generality, a measurement to assess the reliability of a protein-protein interaction*. *Nucleic Acids Research*, 30(5):1163–1168, 2002.
14. Saito R, Suzuki H, and Hayashizaki Y, *Construction of reliable protein-protein interaction networks with a new interaction generality measure*. *Bioinformatics*, 19(6):756–763, 2002.
15. E. Tomita, A. Tanaka, H. Takahashi. *The worst-case time complexity for generating all maximal cliques and computational experiments*. *Theoretical Computer Science* 363(1) (2006) 28–42
16. P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, and et al. *A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae**. *Nature*, 403:623–627, 1999.

## Q&A

- Thank you for your attention 😊

63

## Rank difference and score difference

- Given an interaction  $(u,v)$ , the rank difference of  $(u,v)$  at  $k$ -th iteration is

$$\text{rank\_diff}^k(u,v) = |\text{rank}^k(u,v) - \text{rank}^{k-1}(u,v)|$$

- Given a set of interactions  $E$ , the average ranking difference of all the interactions in  $E$  at  $k$ -th iteration is defined as

$$\text{avg\_rank\_diff}^k(E) = \frac{\sum_{(u,v) \in E} |\text{rank}^k(u,v) - \text{rank}^{k-1}(u,v)|}{|E|}$$

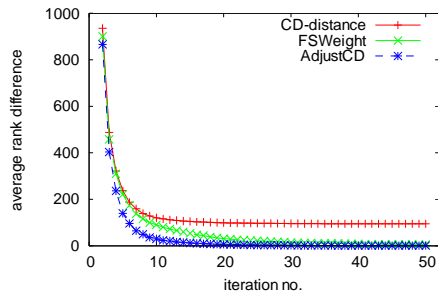
- Similarly, we can define average score difference of all the interactions in  $E$  at  $k$ -th iteration

$$\text{avg\_score\_diff}^k(E) = \frac{\sum_{(u,v) \in E} |\text{score}^k(u,v) - \text{score}^{k-1}(u,v)|}{|E|}$$

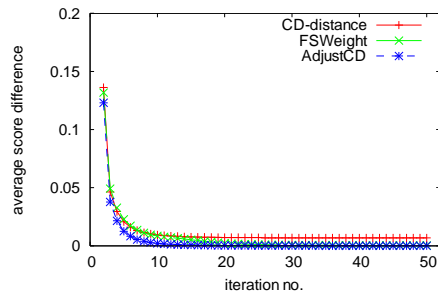
64



# Results on the DIP dataset



Rank difference



Score difference