Progress on three challenges in predicting dynamic protein complexes from a static protein interactome

Limsoon Wong

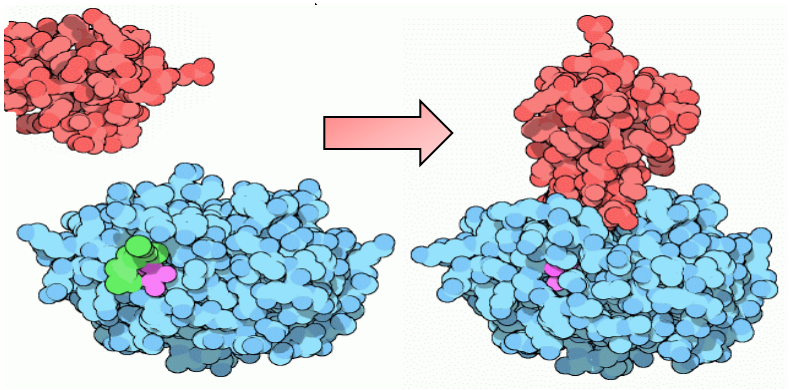
(Based on the work of my student Chern Han Yong)

Outline

- **Overview of protein-complex prediction**
- **Detecting overlapping complexes**
- **Detecting sparse complexes**
- **Detecting small complexes**

Overview of Protein-Complex Detection from PPIN

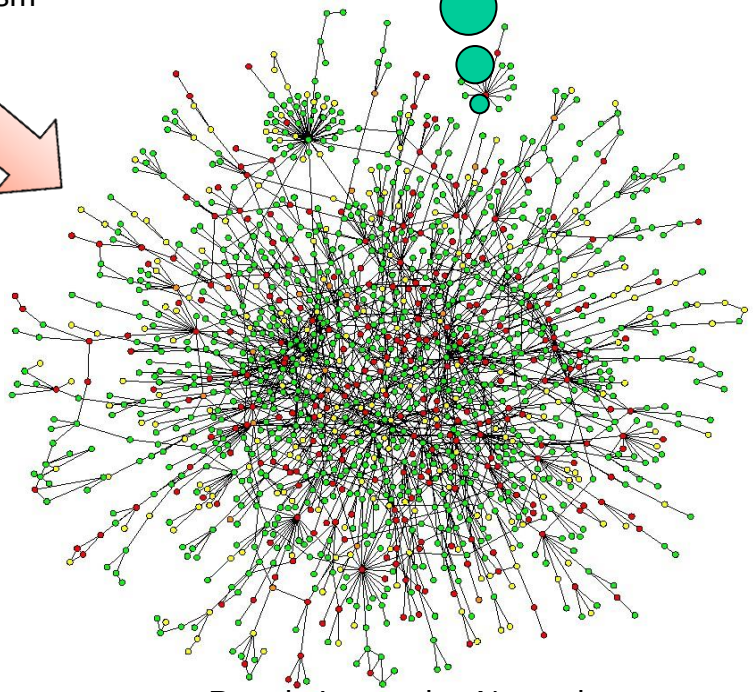
Protein-Interaction Networks



Individual proteins come together and interact

Collection of such interactions in an organism

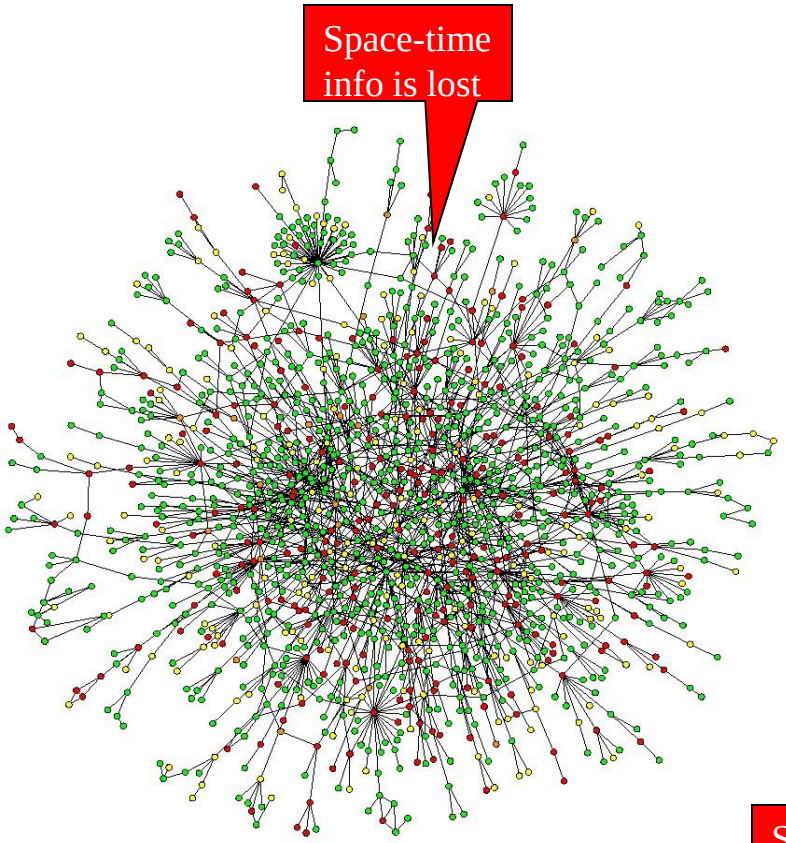
PPIN
Valuable source of knowledge



Protein Interaction Network

- **Proteins come together & interact**
- **The collection of these interactions form a Protein Interaction Network or PPIN**

Detection & Analysis of Protein Complexes in PPIN



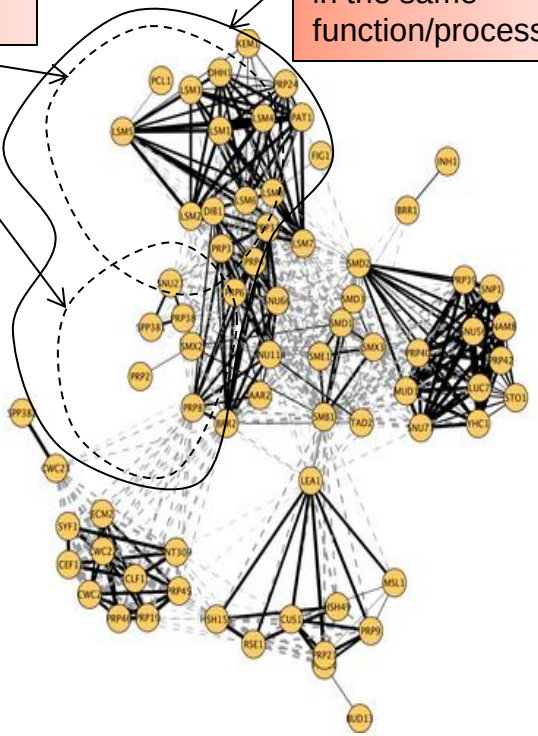
PPIN derived from several high-throughput expt

Identifying embedded complexes

Space-time info is "recovered"

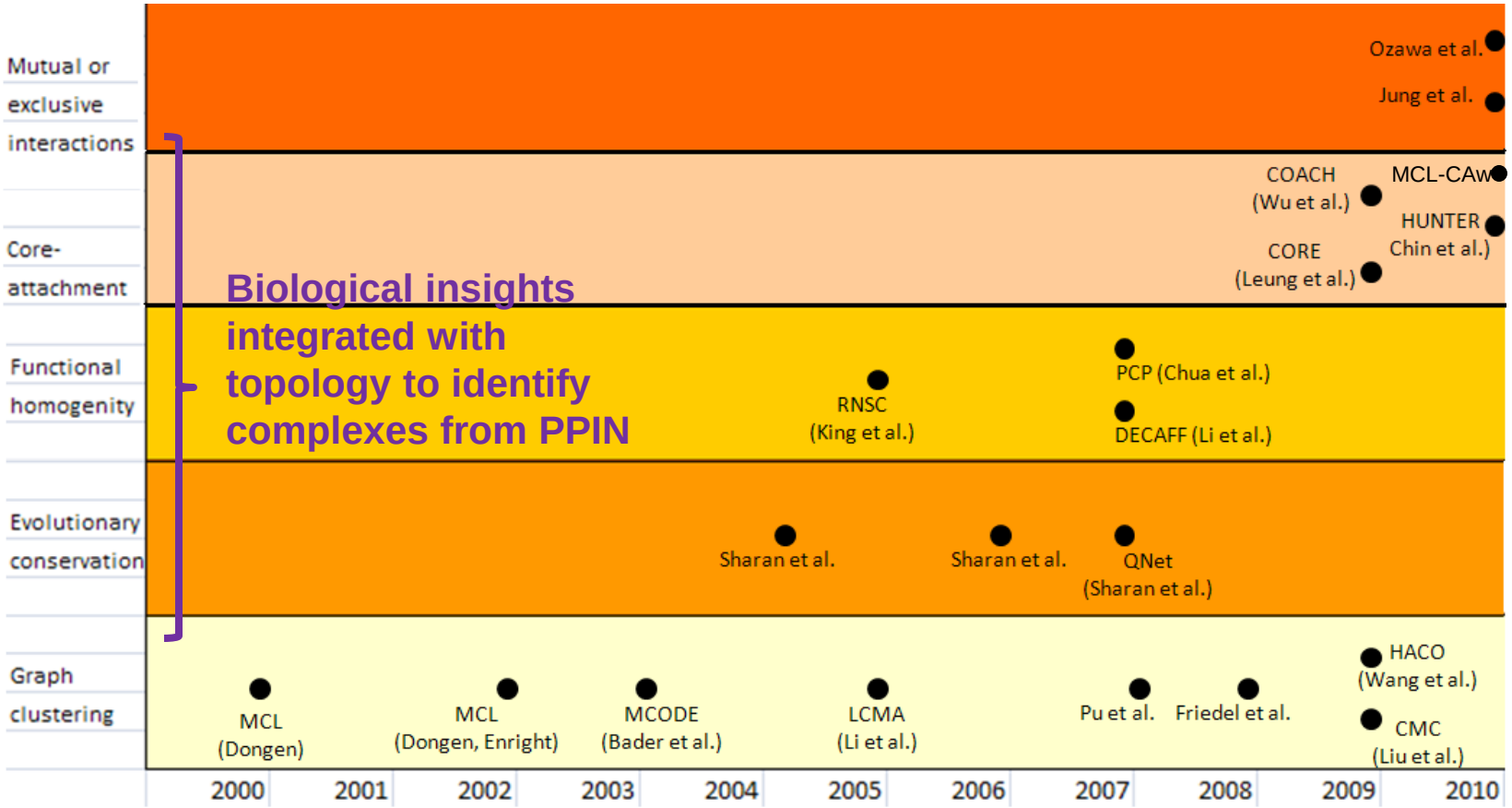
Individual complexes (Some might share proteins)

Entire module might be involved in the same function/process



Embedded **complexes** identified from PPIN

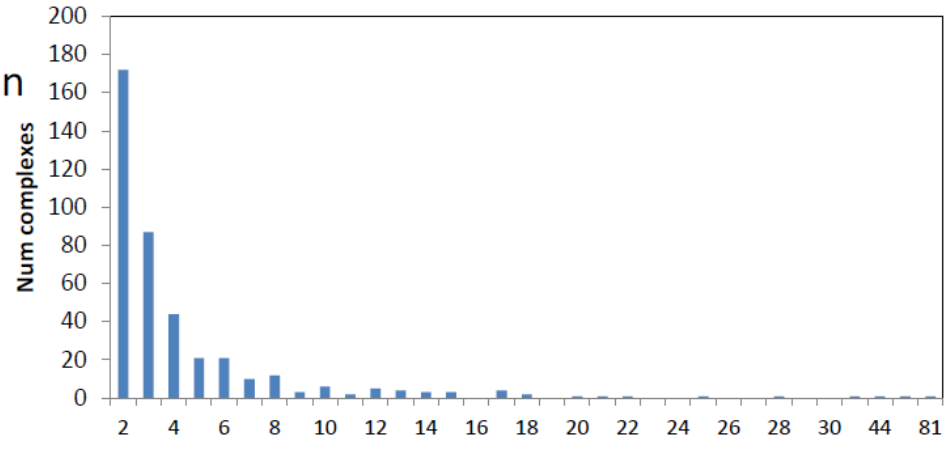
Chronology of Protein-Complex Prediction Methods

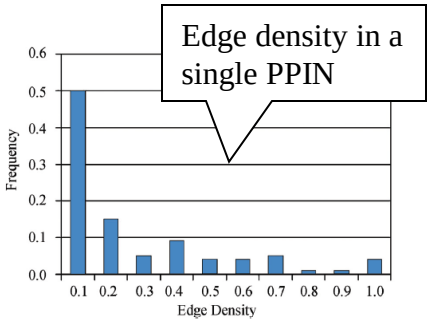


- **As researchers try to improve basic graph clustering techs, they also incorporate bio insights into the methods**

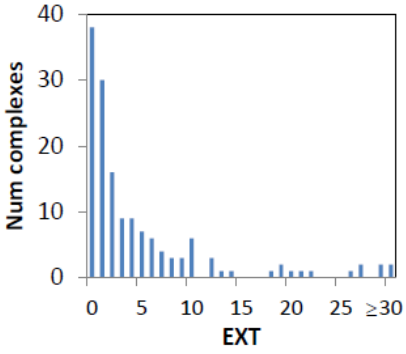
Statistics of Yeast Complexes

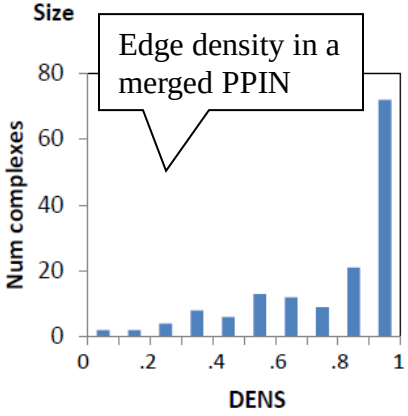
(a) Size distribution





(b) Large complexes





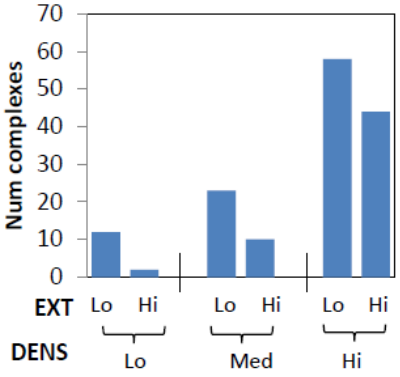


Figure 2.4: Statistics of the yeast reference complexes, from the CYC2008 database. (a) The size distribution of the complexes. (b)EXT (number of highly-connected external proteins) and DENS (density) distributions of large complexes.

What current methods do badly on

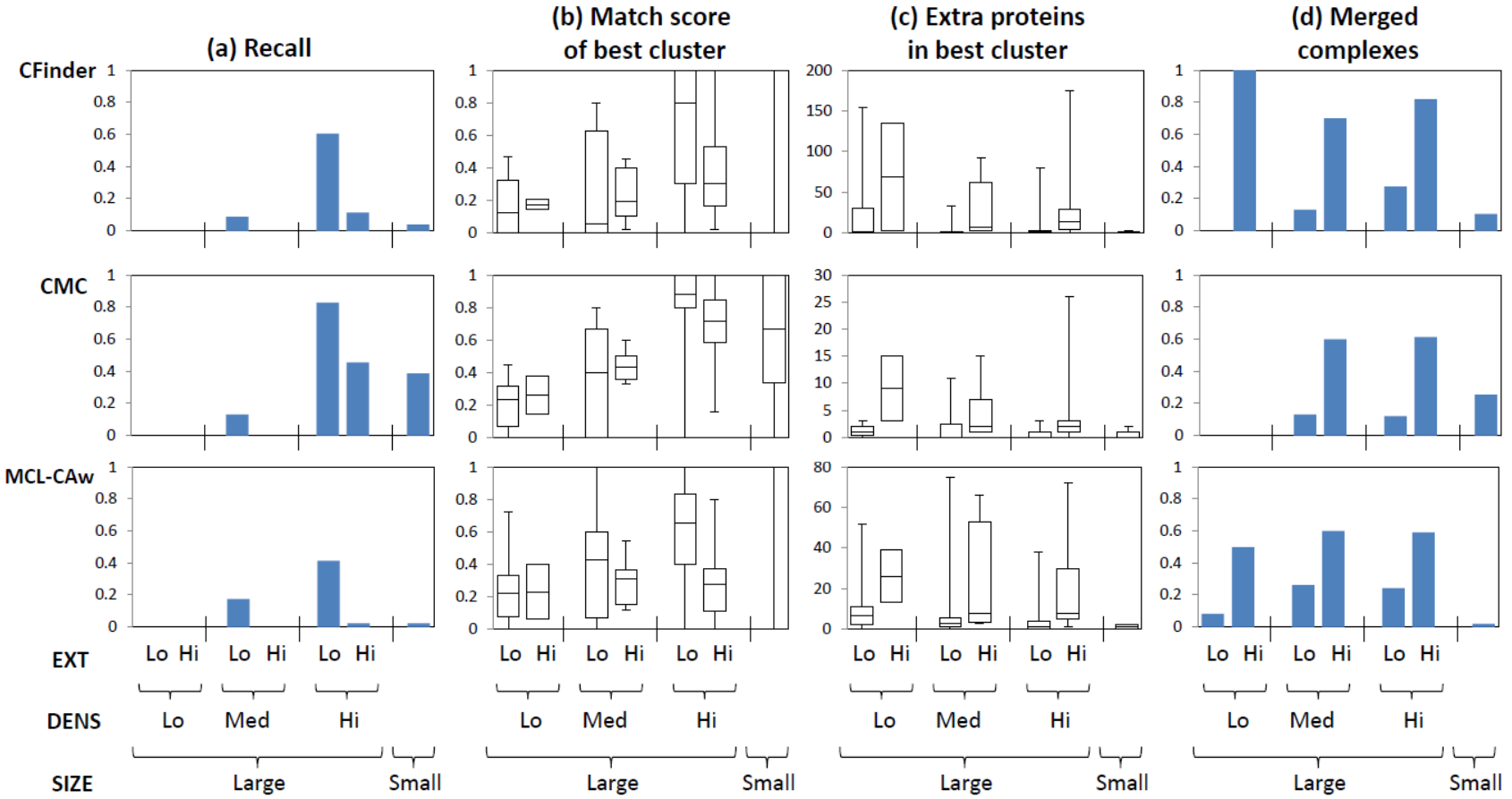


Figure 2.8: Performance of complex-discovery algorithms on yeast complexes, stratified by size, DENS, and EXT. The x-axis of each chart corresponds to the different stratified groups of complexes, given at the bottom of the figure.

Challenges

- **Recall & precision of protein complex prediction algo's have lots to be improved**
- **How to capture "high edge density" complexes that overlap each other?**
- **How to capture "low edge density" complexes?**
- **How to capture small complexes?**

Detecting Overlapping Protein Complexes from Dense Regions of PPIN

Complexes formed by Cdc28p

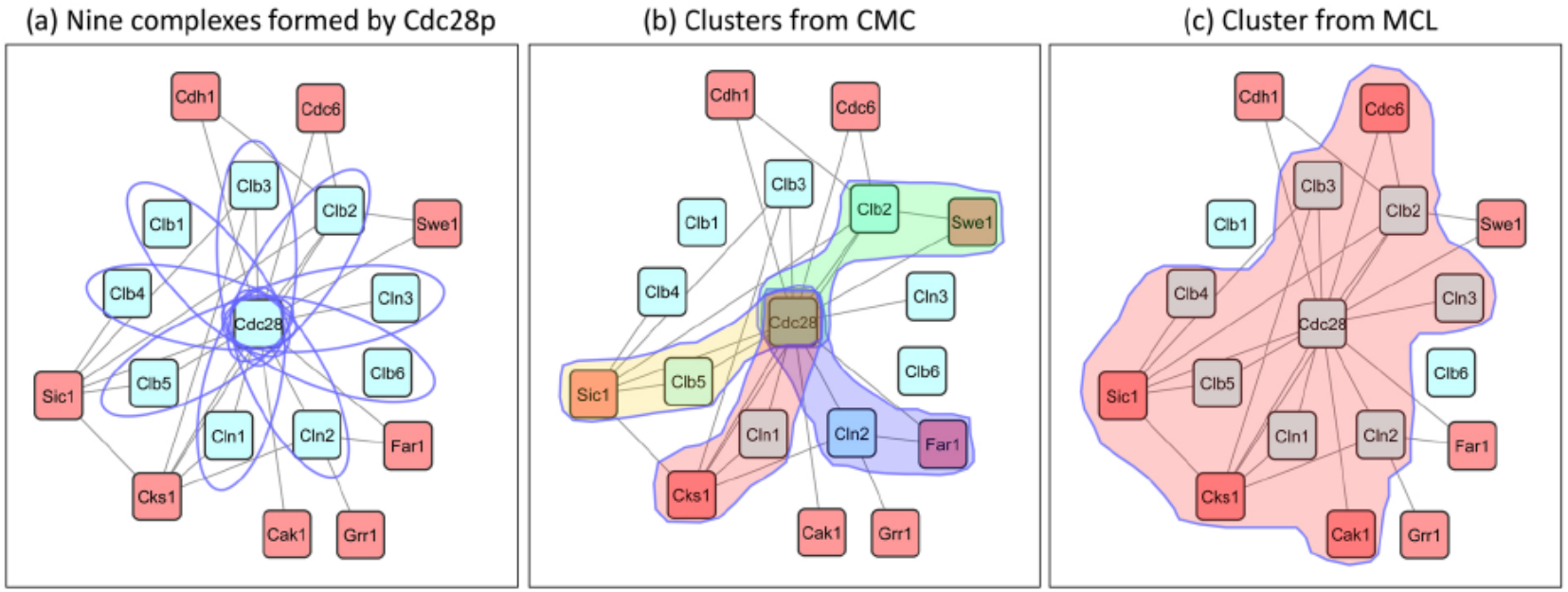


Figure 1.2: (a) Cdc28p is involved in nine distinct complexes, which overlap and have many extraneous edges. Three of the complexes are disconnected. (b) CMC includes extraneous proteins in its clusters. (c) MCL merges the complexes.

Overlapping Complexes in Dense Regions of PPIN

- **Dense regions of PPIN often contain multiple overlapping protein complexes**
- **These complexes often got clustered together and cannot be corrected detected**

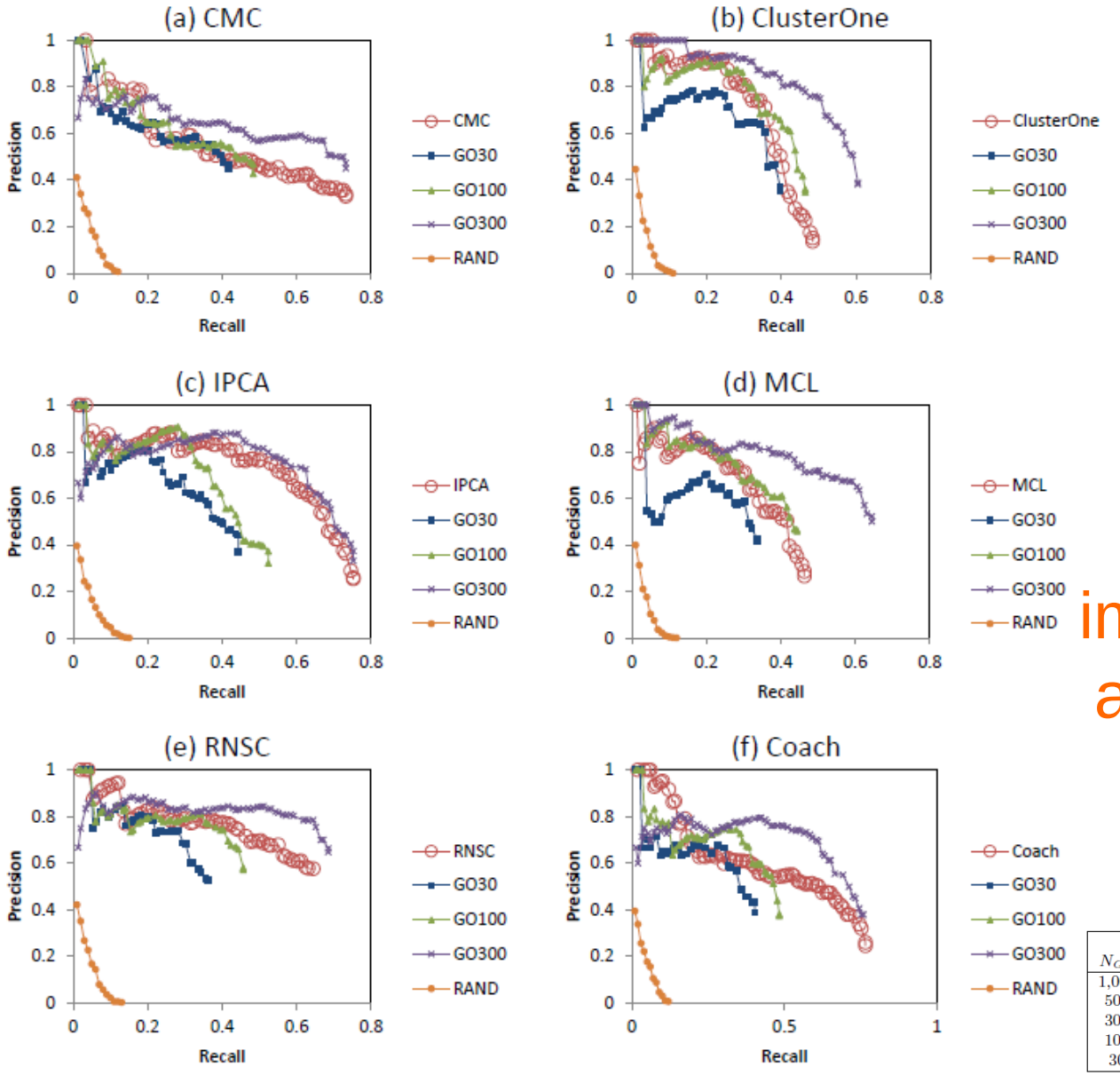
- **Two ideas to cleanse PPI network**
  - Decompose PPI network by localisation GO terms
  - Remove big hubs

Idea I: Split by Localization GO Terms

- **A protein complex can only be formed if its proteins are localized in same compartment of the cell**

⇒ **Use general cellular component (CC) GO terms to decompose a given PPI network into several smaller PPI networks**

- **Use "general" CC GO terms as it is easier to obtain rough localization annotation of proteins**
  - How to choose threshold N_{GO} to decide whether a CC GO term is "general"?



Precision & recall are improved when a suitable N_{GO} is used

N_{GO}	Yeast #GO terms	Yeast #prots discarded	Yeast #PPIs discarded
1,000	6	1,001	5,206
500	8	1,388	7,133
300	10	1,526	7,698
100	25	2,151	12,172
30	48	2,350	13,313

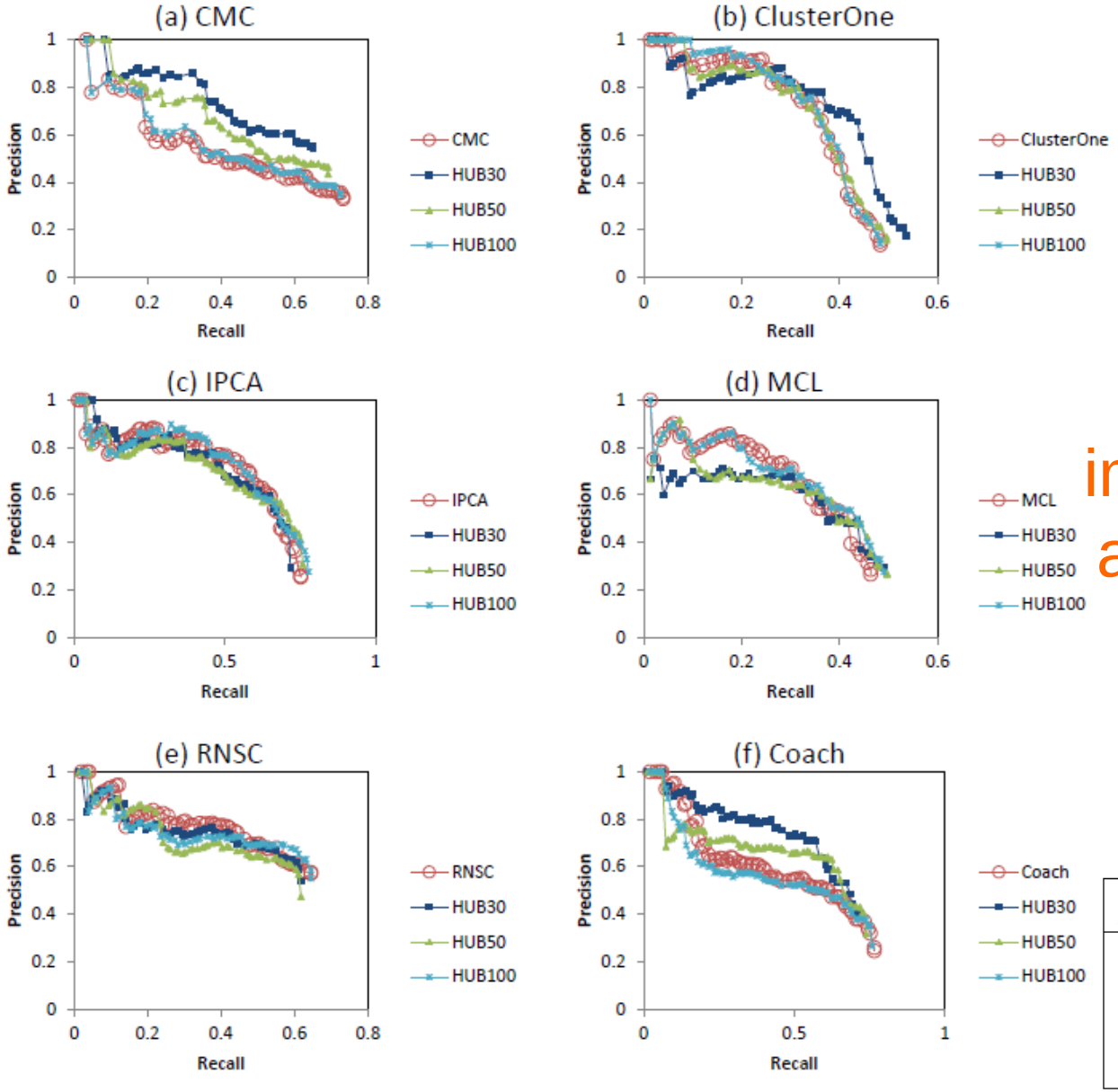
Figure 4.1: Precision-recall graphs for yeast complex prediction using GO decomposition at $N_{GO} = 30, 100, 300$, for the six clustering algorithms.

Idea II: Remove Big Hubs

- **Hub proteins are those proteins that have many neighbors in the PPI network**

- **Large hubs are likely to be "date hubs"; i.e., proteins that participate in many complexes**
  - Likely to confuse protein complex prediction algo

⇒ **Remove large hubs before protein complex prediction**
  - How to choose threshold N_{hub} to decide whether a hub is "large"?



Precision & recall are improved when a suitable N_{hub} is used

N_{hub}	Yeast #hub prots removed	Yeast #PPIs removed
200	2	568
150	3	739
100	9	1,408
75	44	3,920
50	126	7,178
30	340	12,238

Figure 4.3: Precision-recall graphs for yeast complex prediction using hub removal at $N_{hub} = 30, 50, 100$, for the six clustering algorithms.

Decomposition by GO terms and/or hub removal nearly doubles F-score and precision-recall AUC

	Match_thr	F-Score				Prec-Rec AUC			
		Orig	HUB50 GO300	HUB50	GO300	Orig	HUB50 GO300	HUB50	GO300
CMC	.5	.455	**.615**	.533	.557	.417	**.508**	.479	.470
	.75	.275	**.391**	.330	.347	.204	**.278**	.243	.251
ClusterOne	.5	.213	**.483**	.238	.468	.361	**.531**	.362	.514
	.75	.105	**.270**	.107	.255	.209	**.323**	.194	.310
IPCA	.5	.380	**.531**	.438	.460	.564	.560	.549	**.572**
	.75	.143	**.240**	.160	.220	.308	.310	.276	**.323**
MCL	.5	.338	.553	.345	**.563**	.326	.496	.315	**.514**
	.75	.192	.328	.162	**.336**	.170	.255	.104	**.280**
RNSC	.5	.606	.636	.536	**.665**	.500	.560	.455	**.564**
	.75	.355	.377	.321	**.422**	.239	.284	.209	**.305**
Coach	.5	.372	**.573**	.444	.506	.477	**.564**	.505	.536
	.75	.182	**.312**	.223	.262	.218	**.302**	.220	.265

Table 4.4: Performance statistics for yeast complex discovery.

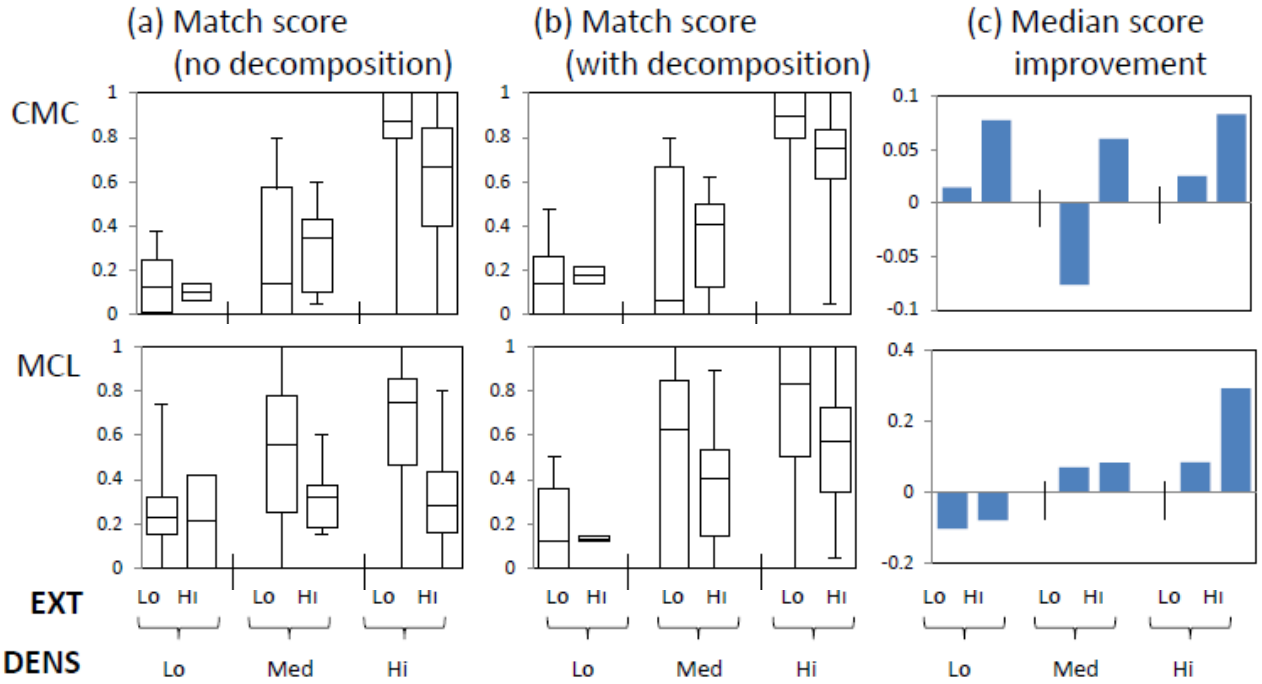
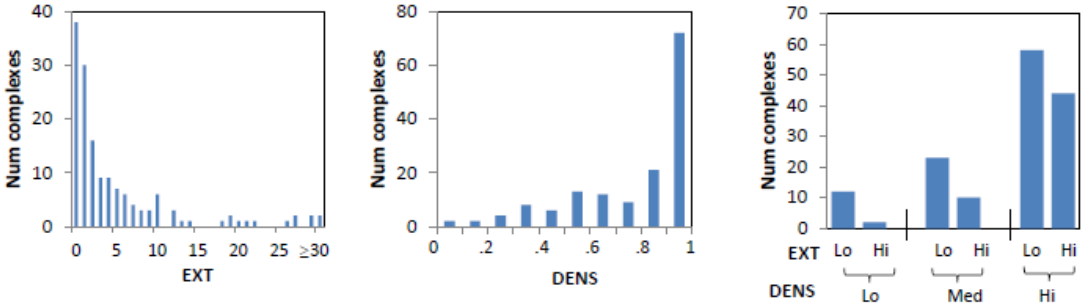


Figure 4.8: Match scores of the best clusters to yeast complexes in the six analysis strata, (a) without PPI decomposition, and (b) with PPI decomposition, generated by various clustering algorithms. (c) shows the improvements score medians.

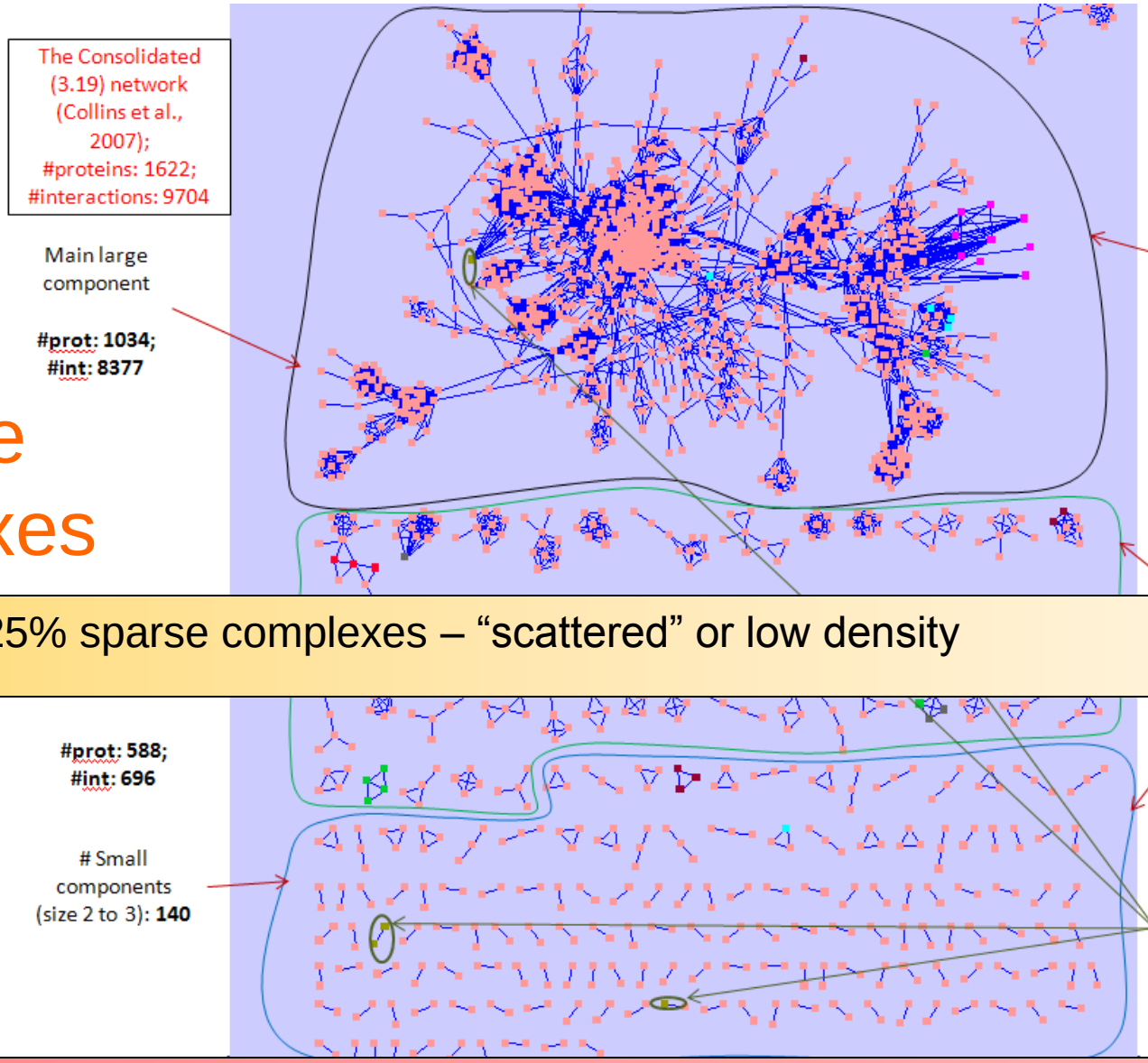
Decomposition is effective in improving prediction of overlapping protein complexes

Distribution of large yeast complexes



Detecting Protein Complexes from Sparse Regions of PPIN

Sparse Complexes

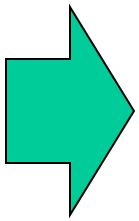


~ 25% sparse complexes – "scattered" or low density

ANY algorithm based solely on topological will miss these sparse complexes!!

- **Key idea to deal with sparseness**

  Augment physical PPI network with other forms of linkage that suggest two proteins are likely to integrate



Supervised Weighting of Composite Networks (SWC)

- **Data integration**
- **Supervised edge weighting**
- **Clustering**

Overview of SWC

1. **Integrate diff data sources to form composite network**
2. **Weight each edge based on probability that its two proteins are co-complex, using a naïve Bayes model w/ supervised learning**
3. **Perform clustering on the weighted network**

- **Advantages**
  - Data integration increases density of complexes
    - **co-complex proteins are likely to be related in other ways even if they do not interact**
  - Supervised learning
    - **Allows discrimination betw co-complex and transient interactions**
  - Naïve Bayes' transparency
    - **Model parameters can be analyzed, e.g., to visualize the contribution of diff evidences in a predicted complex**

1. Integrate Multiple Sources

- **Composite network: Vertices represent proteins, edges represent relationships between proteins**
- **There is an edge betw proteins u, v, if and only if u and v are related according to any of the data sources**

		YEAST			HUMAN		
Data source	Description	# pairs	# distinct proteins	% complex edges	# pairs	# distinct proteins	% complex edges
PPIREL	PPIs, scored by reliability	48,286	5,030	13.6%	44,636	9,535	10.8%
PPITOPO	Topological score of PPI edges	274,277	5,469	3.4%	298,399	9,771	6.1%
STRING	Predicted functional association	175,712	5,964	5.7%	311,435	14,784	3.1%
PubMed	Literature co-occurrence	161,213	5,109	4.9%	91,751	10,659	4.3%
All		518,417	6,099	2.1%	636,966	17,945	3.4%

Coverage ~98%

Coverage ~49%

2. Supervised Edge-Weighting

- **Treat each edge as an instance, where features are data sources and feature values are data source scores, and class label is "co-complex" or "non-co-complex"**

PPI	L2 PPI	STRING	Pubmed	Class
0	0.56	451	0	"co-complex"
0.1	0	25	0	"non-co-complex"

- Supervised learning:
  1. Discretize each feature (Minimum Description Length discretization[7])
  2. Learn maximum-likelihood parameters for the two classes:

$$P(F = f|co-comp) = \frac{n_{c,F=f}}{n_c} \qquad P(F = f|non-co-comp) = \frac{n_{\neg c,F=f}}{n_{\neg c}}$$

  for each discretized feature value f of each feature F

- Weight each edge e with its posterior probability of being co-complex:

$$\begin{aligned} & weight(e) \\ &= P(co-comp|F_1 = f_1, F_2 = f_2, \ldots) \\ &= \frac{P(F_1 = f_1, F_2 = f_2, \ldots|co-comp)P(co-comp)}{Z} \\ &= \frac{\prod_i P(F_i = f_i|co-comp)P(co-comp)}{Z} \\ &= \frac{\prod_i P(F_i = f_i|co-comp)P(co-comp)}{\prod_i P(F_i = f_i|co-comp)P(co-comp) + \prod_i P(F_i = f_i|non-co-comp)P(non-co-comp)} \end{aligned}$$

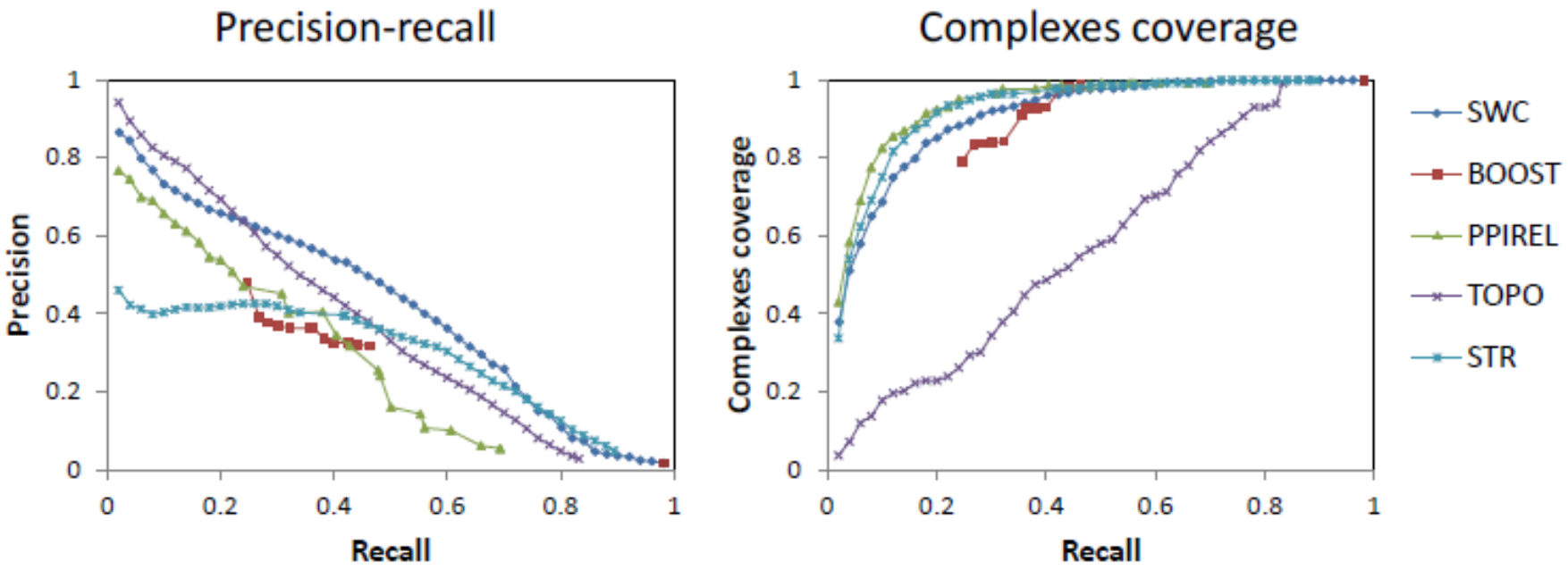
3. Complex Discovery

- **Weighted composite network used as input to clustering algorithms**
  - CMC, ClusterONE, IPCA, MCL, RNSC, HACO
- **Predicted complexes scored by weighted density**

- **The clustering algo's generate clusters with low overlap**
  - Only 15% of clusters are generated by two or more algo's

⇒ **Voting-based aggregative strategy, COMBINED:**
  - Take union of clusters generated by the diff algo's
  - Similar clusters from multiple algo's are given higher scores
    - **If two or more clusters are similar (Jaccard >= 0.75), then use the highest scoring one and multiply its score by the # of algo's that generated it**

Co-Complex Edge Prediction



- **Precision-recall and complex coverage graphs for classification of co-complex edges for yeast**
- **Only TOPO has higher precision than SWC, but its edges are clustered in very few complexes**

Yeast Complex Prediction

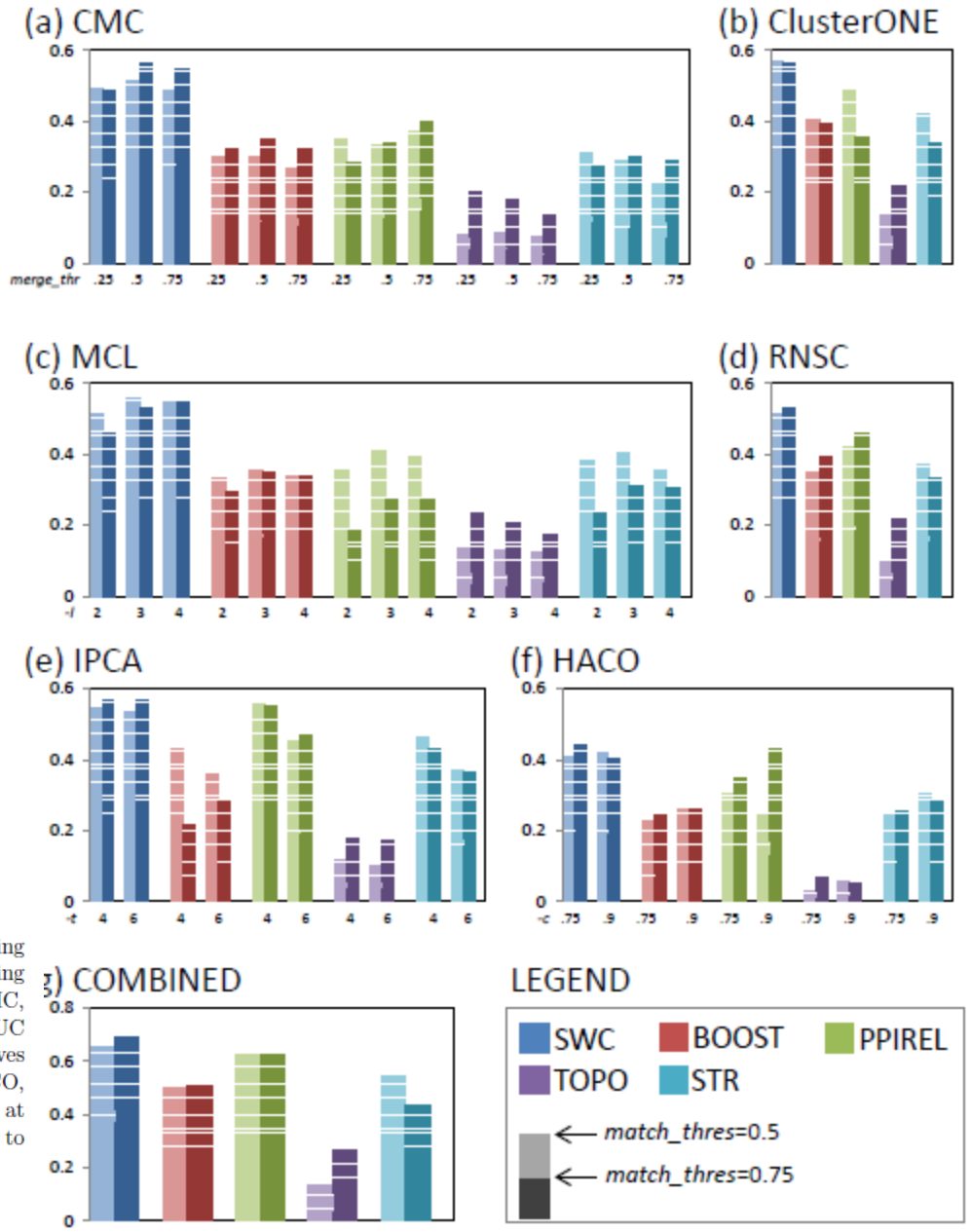


Figure 3.3: Precision-recall AUC for yeast complex prediction, using the five weighting approaches for each of the six clustering algorithms and the COMBINED clustering strategy, for $k = 10000$ (lighter shade), and $k = 20000$ (darker shade). For CMC, MCL, IPCA, and HACO, different sets of clustering parameters are tried. The AUC for *match_thres* $= 0.5$ and *match_thres* $= 0.75$ are shown in each bar. SWC achieves highest precision-recall AUC for all clustering algorithms except IPCA and HACO, where it performs about evenly with PPIREL at *match_thres* $= 0.5$ but better at *match_thres* $= 0.75$. The COMBINED strategy achieves higher AUC compared to using any single clustering algorithm alone.

SWC gives better precision at similar or better recall

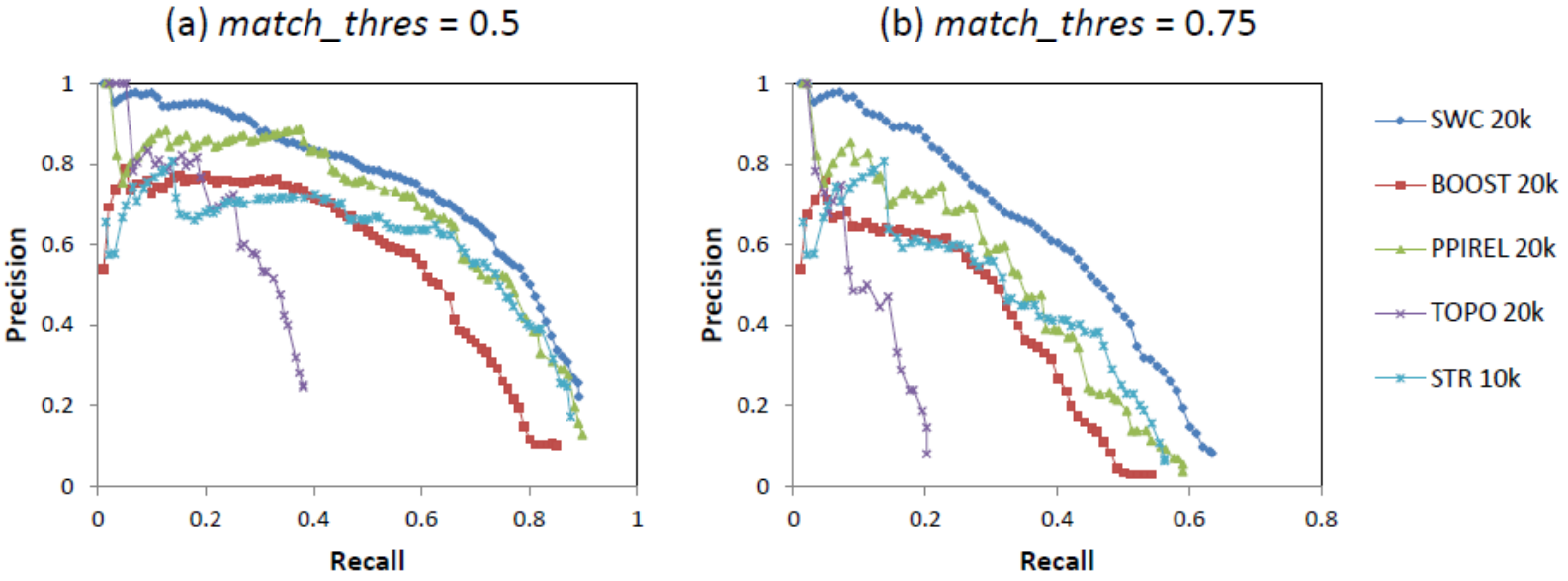


Figure 3.5: Precision-recall graphs for yeast complex prediction using the five weighting approaches with the COMBINED clustering strategy, using $k = 20000$ for SWC, BOOST, PPIREL, and TOPO, and $k = 10000$ for STR, at (a) $match\_thres = 0.5$, (b) $match\_thres = 0.75$. At $match\_thres = 0.5$, SWC achieves similar recall as BOOST, PPPIREL, and STR, but with the higher precision at almost all recall levels. At the stricter $match\_thres = 0.75$, SWC achieves the highest recall with the highest precision at almost all recall levels. Thus it outperforms all other weighting approaches, especially at predicting complexes with fine granularity.

NUS National University of Singapore

SWC is successful in improving prediction of sparse complexes

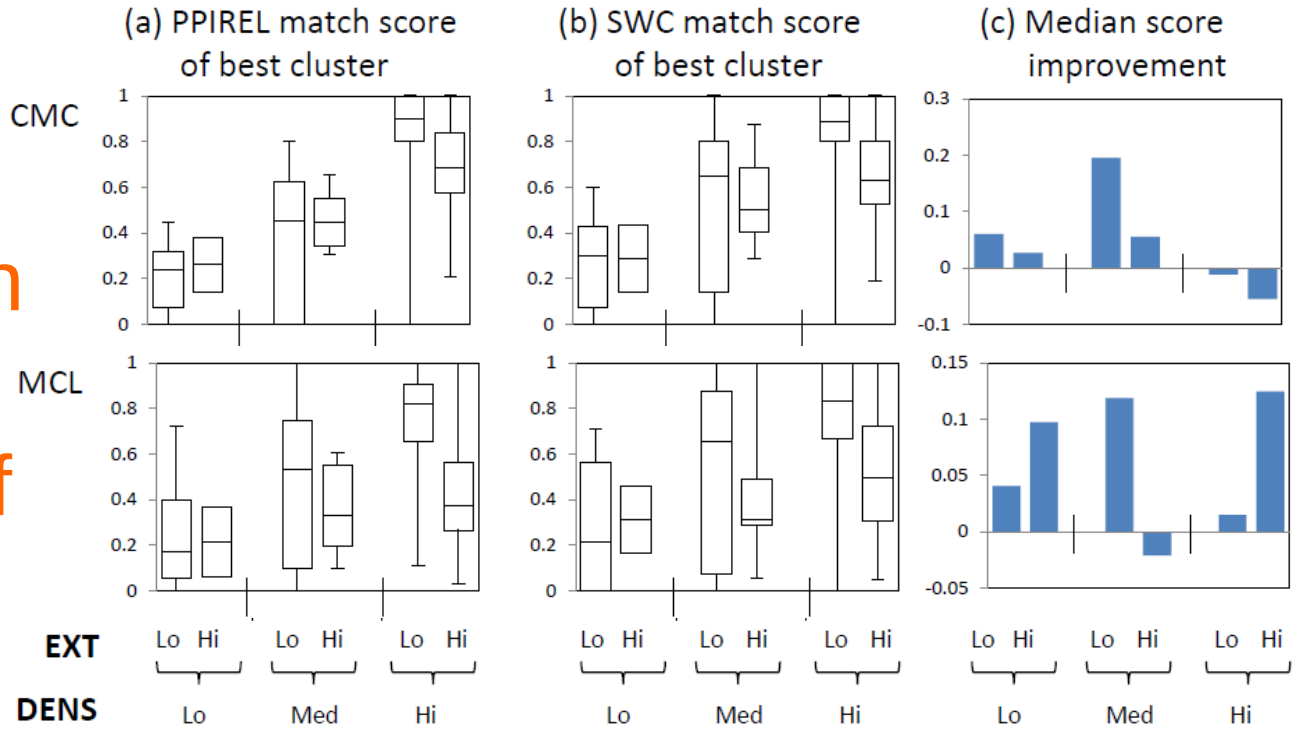
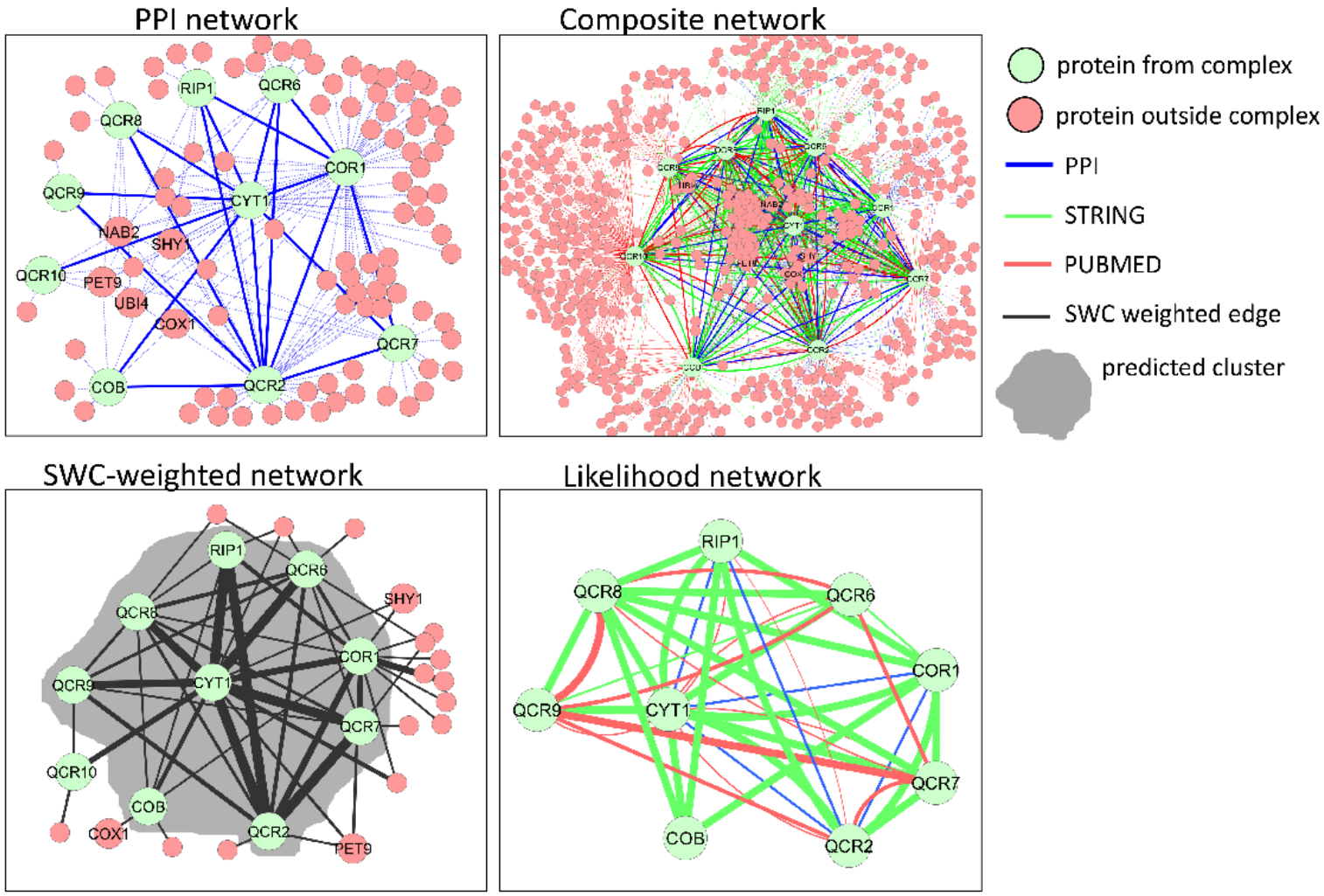


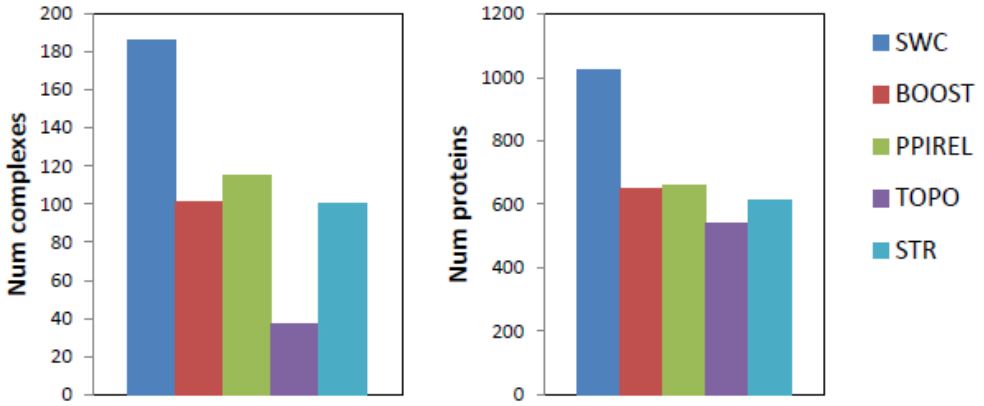
Figure 3.9: Match scores of the best clusters to yeast complexes in the six analysis strata, using (a) PPIREL, and (b) SWC, generated by various clustering algorithms. (c) shows the improvements score medians. SWC gives bigger improvements among low- and medium-density complexes for most clustering algorithms.

Yeast BC1 Complex



Novel Predicted Complexes

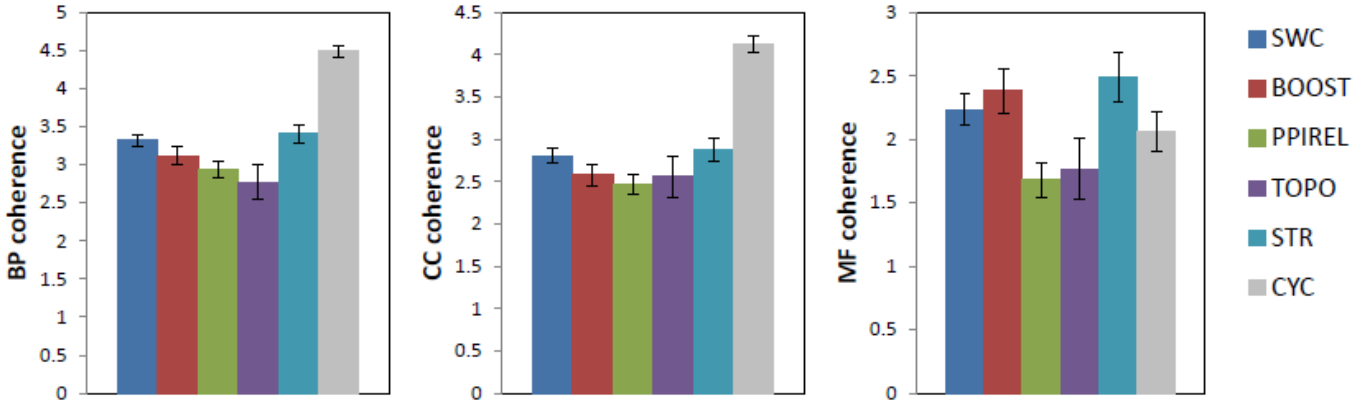
(a) Number of unique, high-confidence, novel predicted yeast complexes



Novel Predicted Yeast Complexes

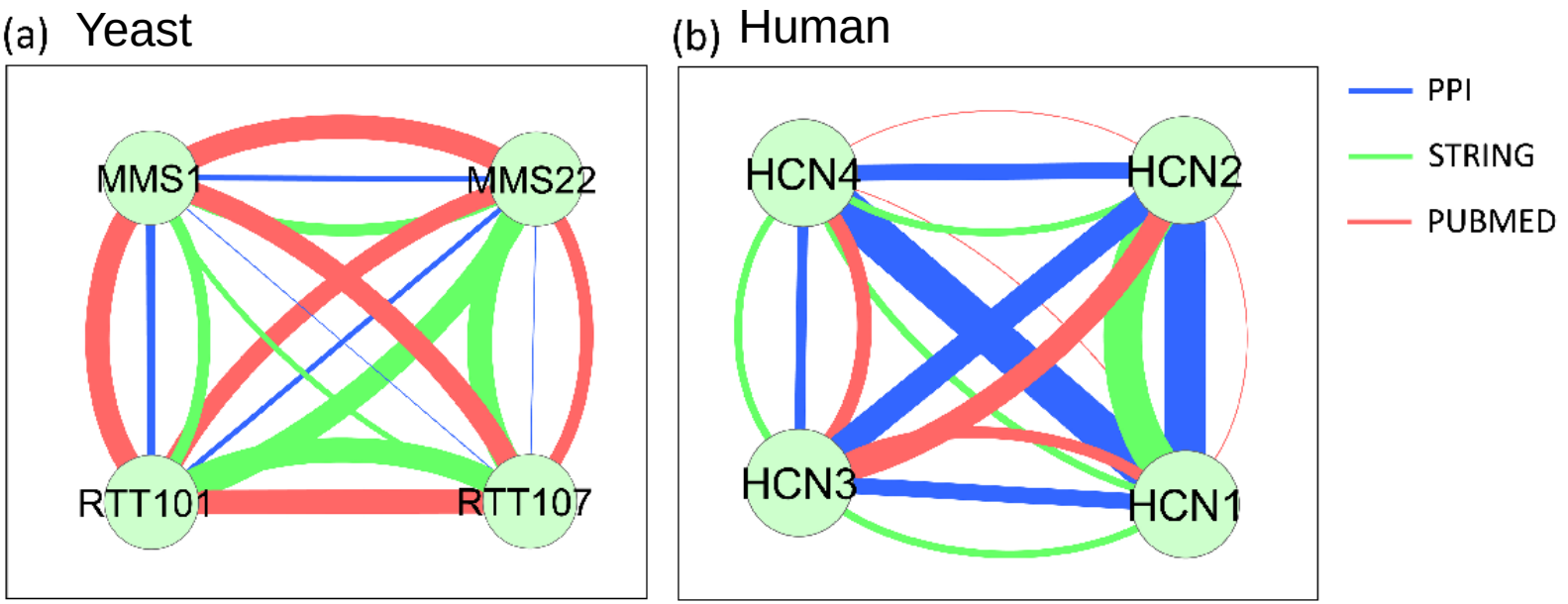
Biological process	# complexes
Protein metabolic process	39
RNA metabolic process	25
DNA metabolic process	9
Small molecule metabolic process	16
Regulation of metabolic process	20
Regulation of gene expression	13
Organelle organization	33
Transport	44
Response to stress	16
Response to chemical stimulus	5
Cell cycle process	8

(b) Coherence of predicted yeast complexes



Based on COMBINED

Two Novel Predicted Complexes



- **Novel yeast complex: Annotated w/ DNA metabolic process and response to stress, forms a complex called Cul8-RING which is absent in our ref set**
- **Novel human complex: Annotated w/ transport process, Uniprot suggests it may be a subunit of a potassium channel complex**

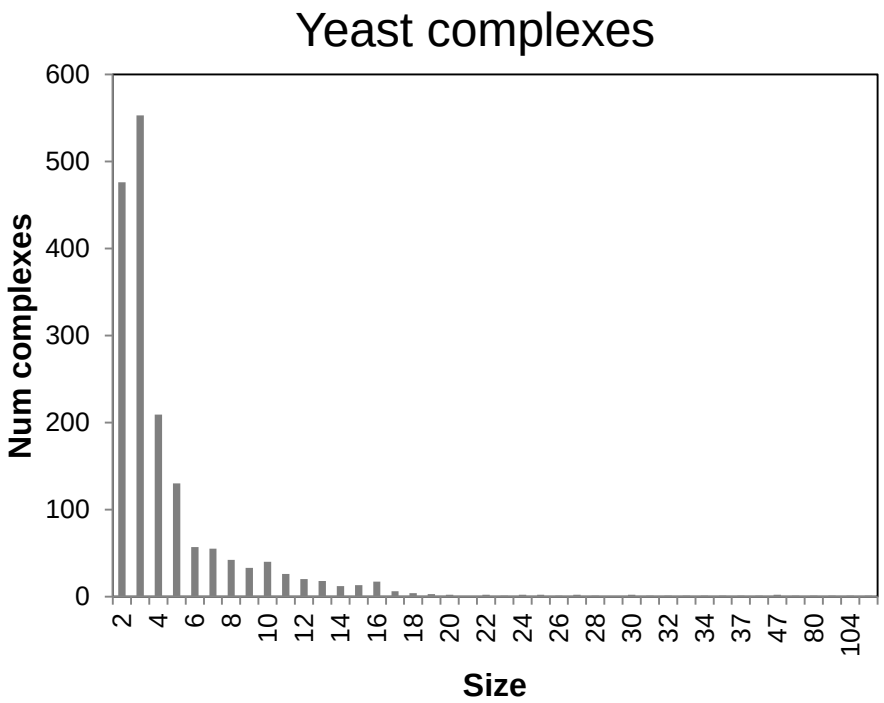
Conclusions

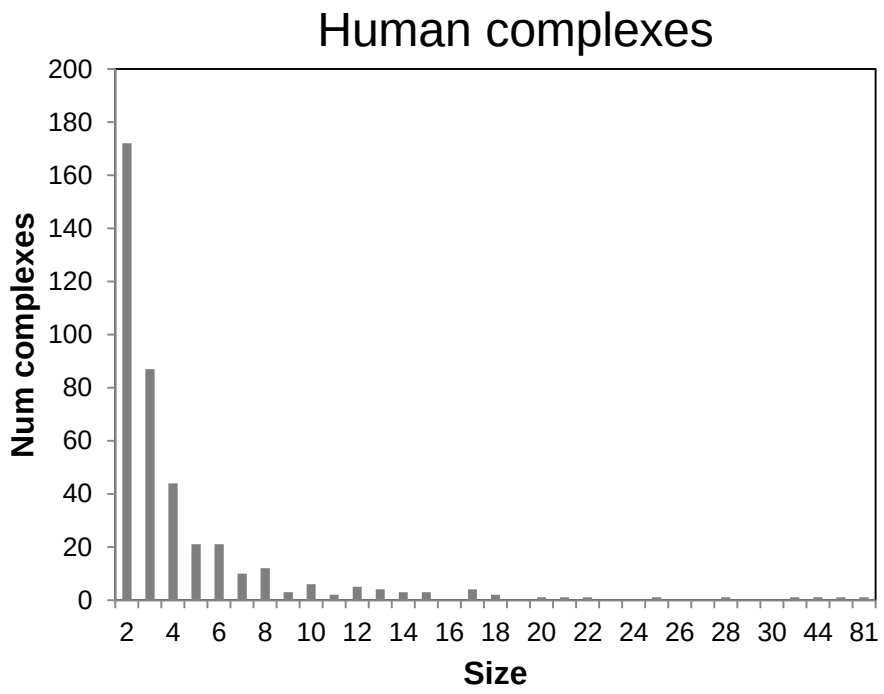
- **Naïve-Bayes data-integration to predict co-complexed proteins**
  - Use of multiple data sources increases density of complexes
  - Supervised learning allows discrimination betw co-complex and transient interactions
- **Tested approach using 6 clustering algo's**
  - Clusters produced by diff algo's have low overlap, combining them gives greater recall & precision
  - SWC is successful in improving sparse complexes prediction

Detecting Small Protein Complexes

Motivation

- **Size of protein complexes follows a power-law distribution, meaning that most complexes are small (ie. 2 or 3 distinct proteins)**





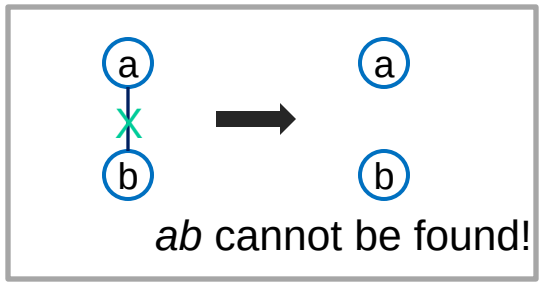
Small Complexes, Big Challenges

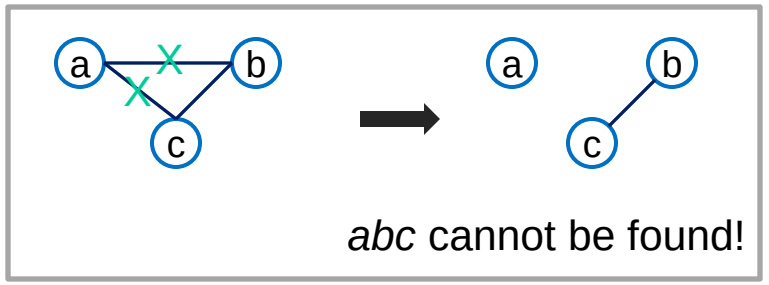
- **Traditionally, complexes are predicted by searching for dense clusters in a PPI network**

- **For small complexes, topological characteristics like density are problematic**
  - A fully-dense size-2 complex is an edge
  - A fully-dense size-3 complex is a triangle
  - But there are many edges and triangles in the PPI network that are not complexes

Small Complexes, Big Challenges

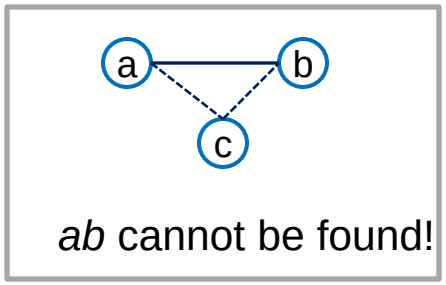
- **Sensitive to missing edges**
  - One missing edge disconnects a size-2 complex
  - Two missing edges disconnect a size-3 complex

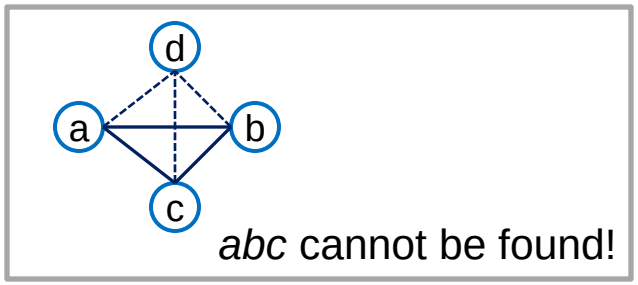




Small Complexes, Big Challenges

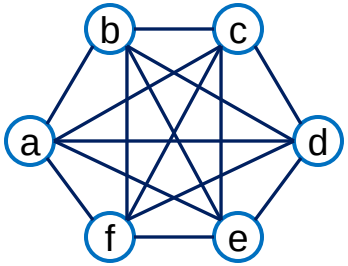
- **Sensitive to extraneous edges**
  - Two extraneous edges embed a size-2 complex in a size-3 clique
  - Three extraneous edges embed a size-3 complex in a size-4 clique





Small Complexes, Big Challenges

- **Predicted complexes are scored using their internal weights to give them some reliability measure, eg. using weighted density. This reliability is averaged out over the internal weights of the candidate complex**



Size-6 complex: Score is averaged over 15 edge weights

- **Scores of small complexes are sensitive to the correct edge weights, since only one or three edges weights are used**

Size-2 complex: Score depends on just 1 edge weight. It is very sensitive to its value

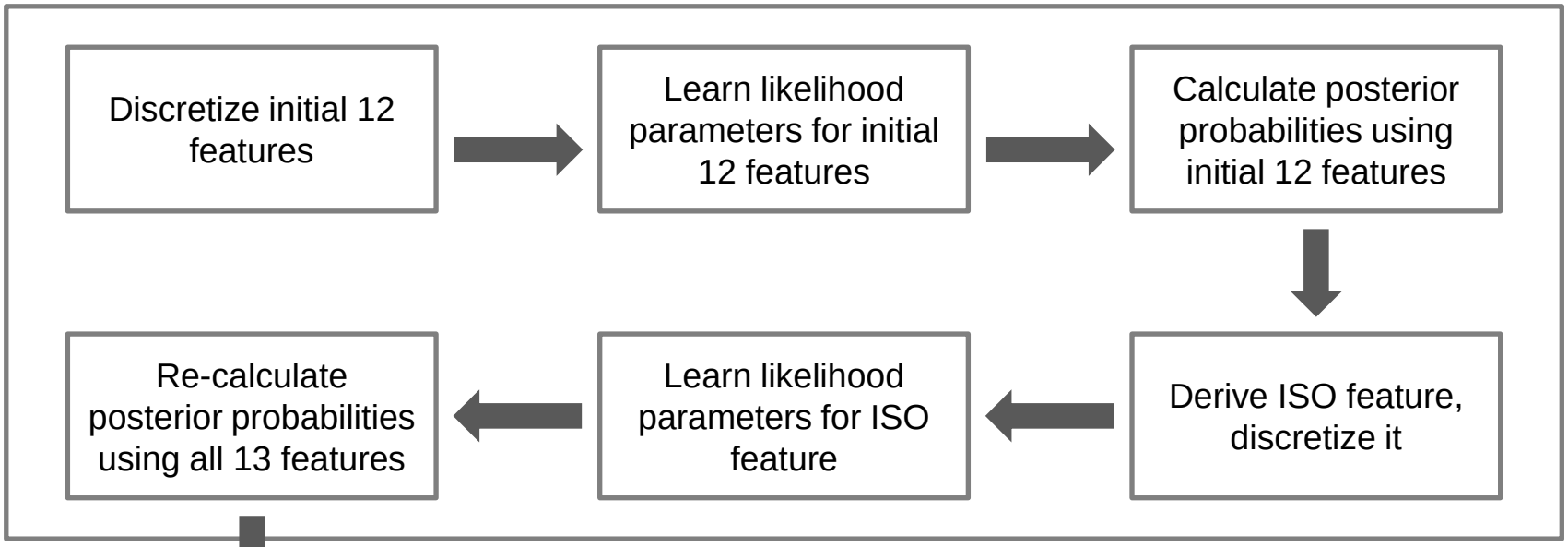
Small Complexes, Big Challenges

- **Previously used data integration and supervised learning successfully for predicting large complexes (SWC2)**

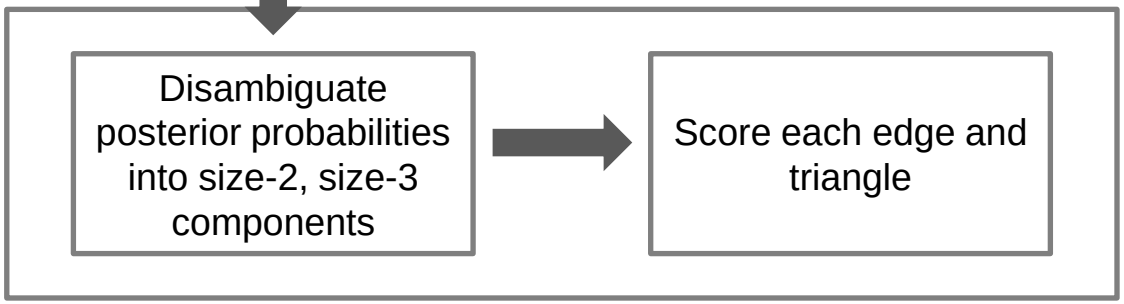
- **It does not work well for small complexes**
  - Small complexes have different topological features compared to large complexes
  - Learned model corresponds to large complexes, not small complexes, as large complexes have much more edges

Two-Stage Approach

1. Size-specific supervised weighting (SSS)

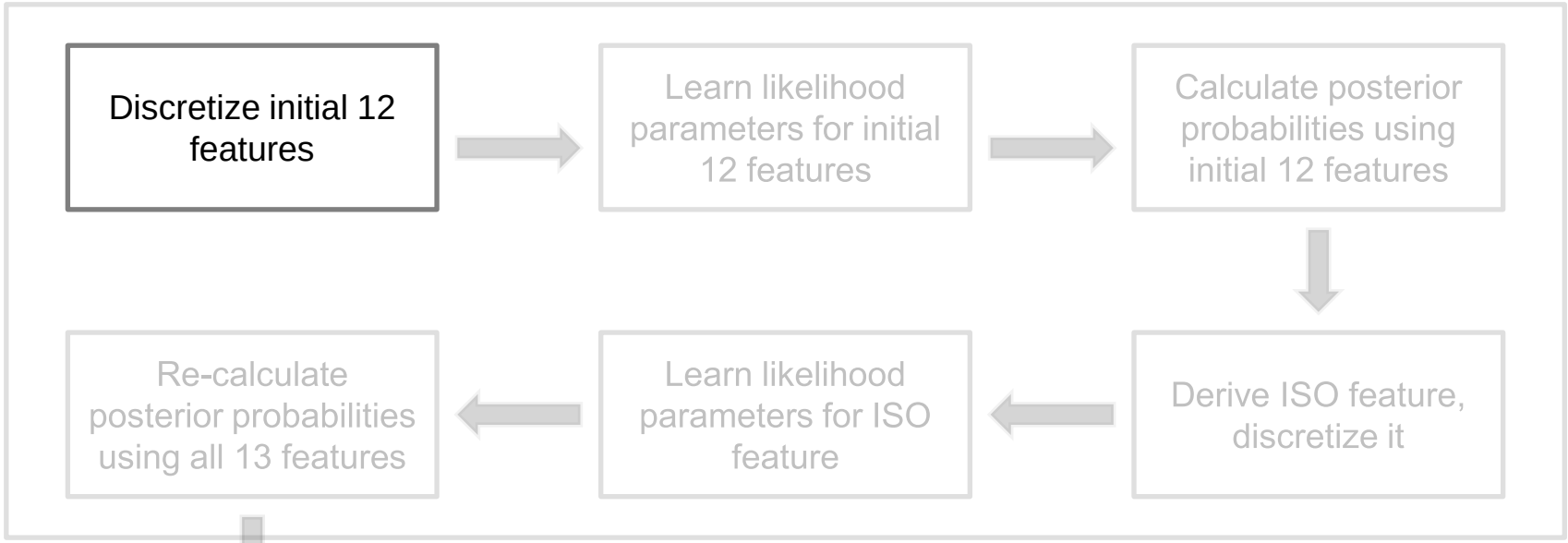


2. Extract

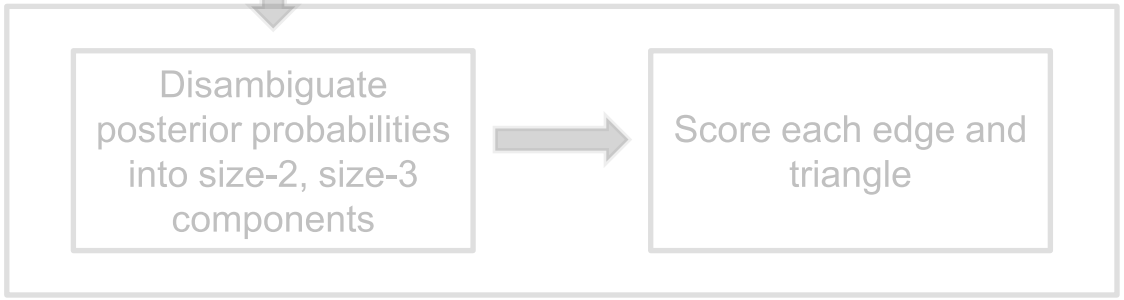


Stage 1: SSS

1. Size-specific supervised weighting (SSS)



2. Extract



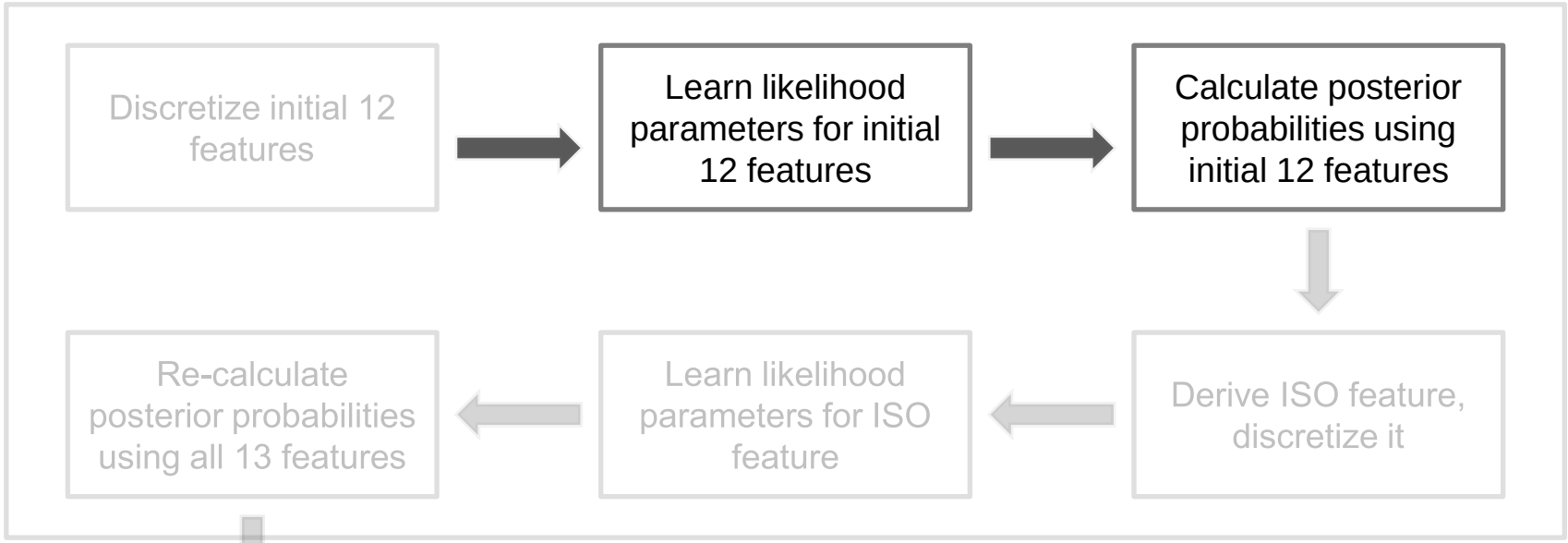
Discretize initial 12 features

- **Each edge in PPIN is cast as a data instance, with 12 initial features**
  - 3 data sources
    - **PPI (BioGrid + IntAct + MINT)**
    - **Functional associations (STRING)**
    - **Co-occurrence in literature (PUBMED)**
  - 3 topological characteristics for each data source
    - **Degree**
    - **Neighbourhood connectivity**
    - **Shared neighbours**

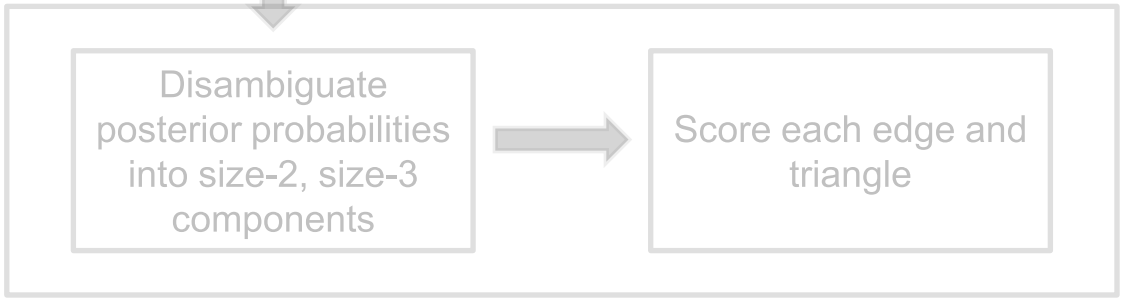
- **Discretize based on Minimum Description Length (MDL)**

Stage 1: SSS

1. Size-specific supervised weighting (SSS)



2. Extract



NUS National University of Singapore

Learn likelihood parameters for initial 12 features

- **Likelihood models for 3 classes (small co-complex, large co-complex, non co-complex)**

$$P(F = f|sm\text{-}comp) = \frac{n_{sm,F=f}}{n_{sm}}$$

$$P(F = f|lg\text{-}comp) = \frac{n_{lg,F=f}}{n_{lg}}$$

$$P(F = f|non\text{-}comp) = \frac{n_{non,F=f}}{n_{non}}$$

Calculate posterior probabilities using initial 12 features

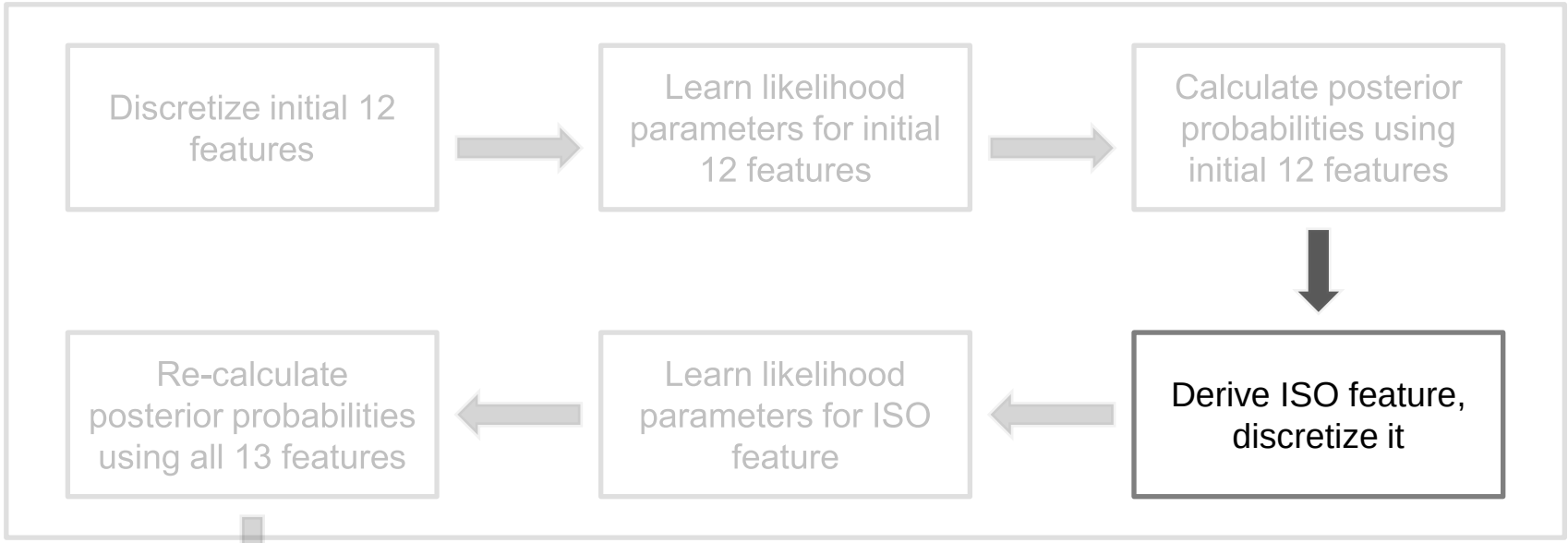
- **Weight each edge with its posterior probability of being small co-complex, large co-complex, or non co-complex, using the naïve-Bayes formulation**
  - Eg., probability that edge (a,b) is small co-complex

$$= \frac{P((a,b)\ is\ sm\text{-}comp|F_1 = f_1, F_2 = f_2, \ldots) \prod_i P(F_i = f_i|(a,b)\ is\ sm\text{-}comp)P(sm\text{-}comp)}{\sum_{class \in \{sm\text{-}comp, lg\text{-}comp, non\text{-}comp\}} \prod_i P(F_i = f_i|(a,b)\ is\ class)P(class)}$$

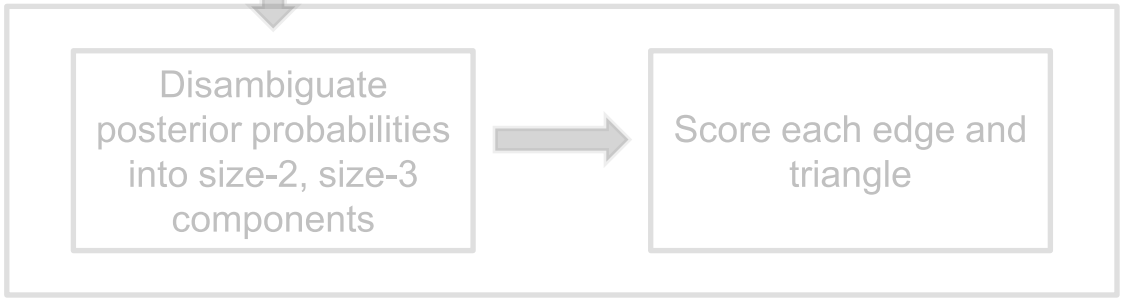
- **These three probabilities are abbreviated as**
  - $P_{(a,b),sm}$
  - $P_{(a,b),lg}$
  - $P_{(a,b),non}$

Stage 1: SSS

1. Size-specific supervised weighting (SSS)



2. Extract



Derive ISO feature

- **For each edge, derive a new feature, Isolatedness**
  - Prob that the edge is isolated, or is part of an isolated triangle
  - Uses posterior prob calculated previously

$$ISO(a,b) = ISO2(a,b) + ISO3(a,b)$$

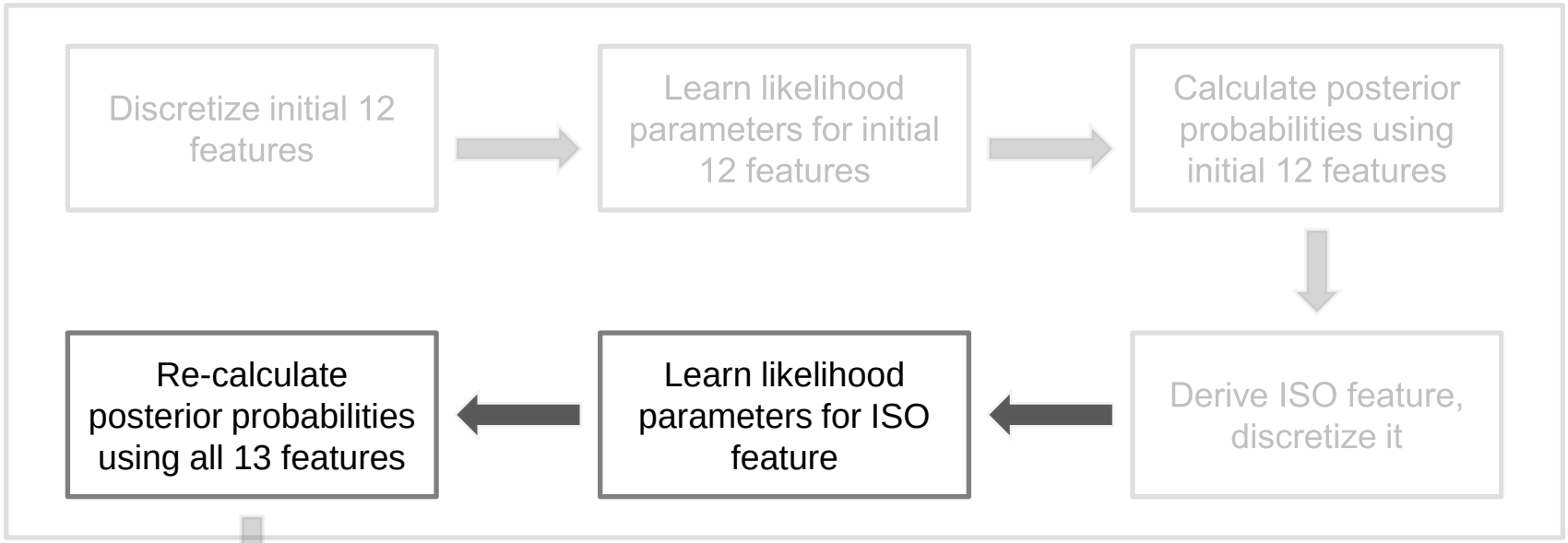
$$ISO2(a,b) = P_{(a,b),sm} \prod_{x \in \{a,b\}, y \in N_{a,b}} P_{(x,y),non}$$

$$ISO3(a,b) = \sum_{c \in N_a \cap N_b} \left(P_{(a,b),sm} P_{(a,c),sm} P_{(b,c),sm} \prod_{x \in \{a,b,c\}, y \in N_{a,b,c}} P_{(x,y),non} \right)$$

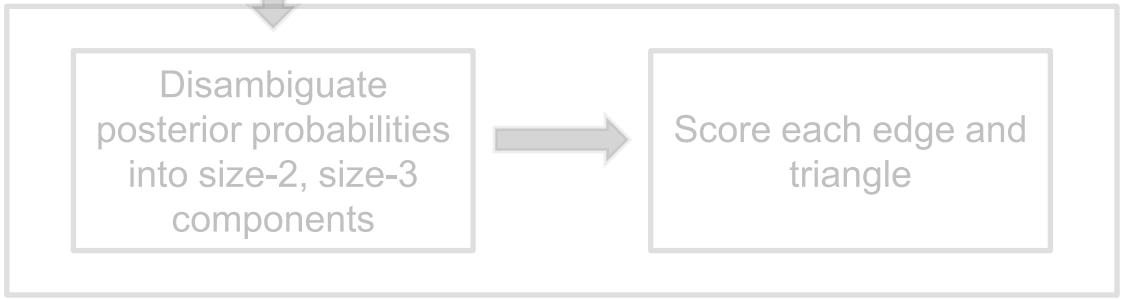
- **This feature is also discretized using MDL**

Stage 1: SSS

1. Size-specific supervised weighting (SSS)



2. Extract



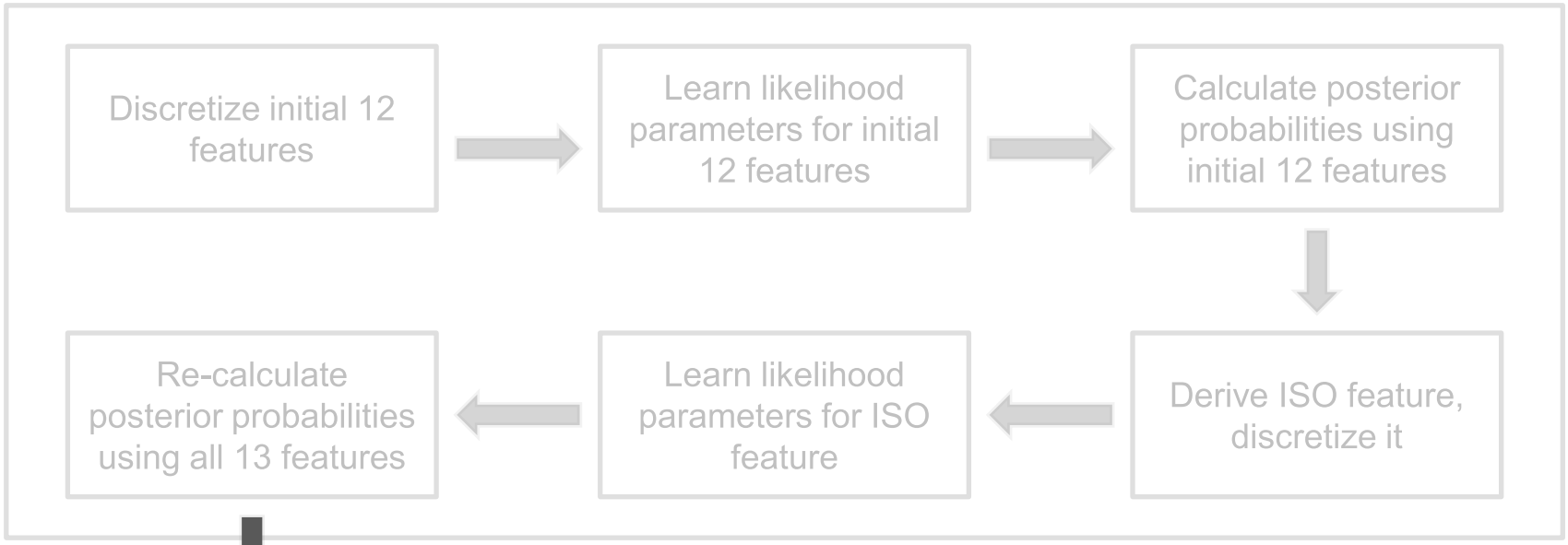
Learn likelihood parameters for ISO feature & Recalculate posterior prob using all 13 features

- **Likelihood parameters are learned for the ISO feature in the same way as with the previous features**

- **Posterior prob are re-calculated as before, this time incorporating the new ISO feature**
  - P(a,b),sm = prob that (a,b) is small co-complex
  - P(a,b),lg = prob that (a,b) is large co-complex
  - P(a,b),non = prob that (a,b) is non co-complex

Stage 2: Extract

1. Size-specific supervised weighting (SSS)

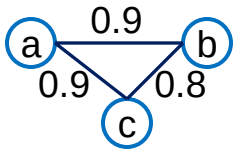


2. Extract

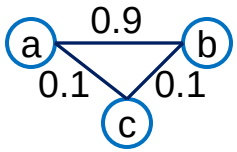
Disambiguate $P_{(a,b),sm}$, the prob that (a,b) is small co-complex, into size-2 and size-3 components

- **If (a,b) is part of a high-weighted triangle, then it is likelier to be part of a size-3 complex, so reduce its size-2 component**

$$P'_{(a,b),sm2} = P_{(a,b),sm} - \sum_{x \in N_a \cap N_b} P_{(a,b),sm} P_{(a,x),sm} P_{(b,x),sm}$$



(a,b) likelier to be part of a size-3 complex *abc* than a size-2 complex *ab*

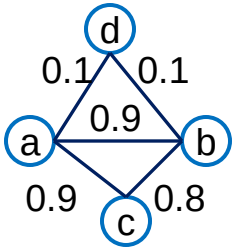


(a,b) likelier to be a size-2 complex than size-3 complex *abc*

Disambiguate $P_{(a,b),sm}$, the prob that (a,b) is small co-complex, into size-2 and size-3 components

- **If (a,b) is part of a high-weighted triangle, and is part of another low-weighted triangle, then it is likelier to be in a complex with the first triangle**

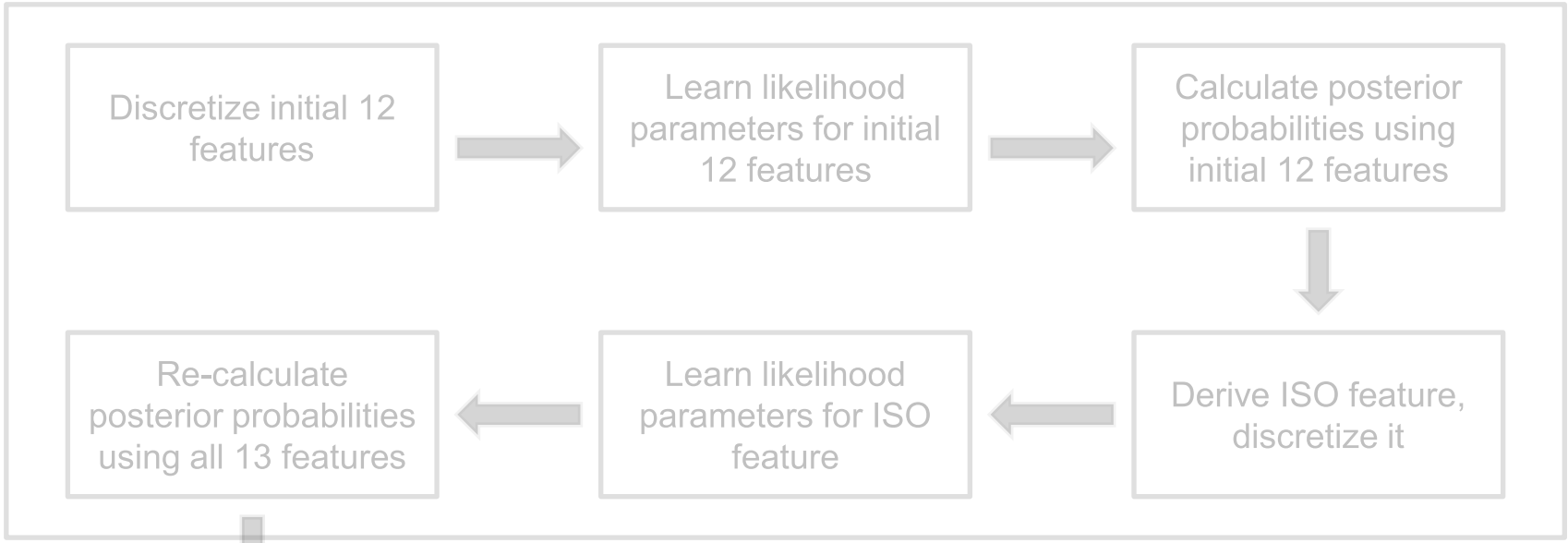
$$P'_{(a,b),sm3,abc} = P_{(a,b),sm} - \sum_{x \in N_a \cap N_b \setminus \{c\}} P_{(a,b),sm} P_{(a,x),sm} P_{(b,x),sm}$$



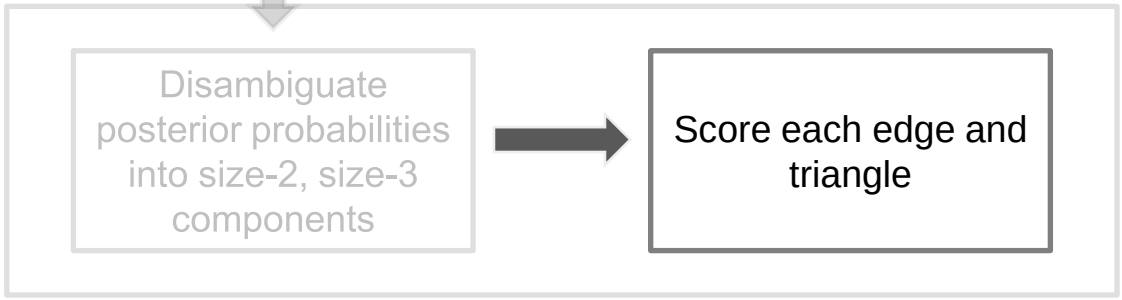
(a,b) likelier to be part of a size-3 complex *abc*, than complex *abd*

Stage 2: Extract

1. Size-specific supervised weighting (SSS)



2. Extract



Score each edge and triangle

- **Every edge / triangle is taken as candidate size-2 / -3 complexes**

- **Score each candidate complex, using edges inside the complex, as well as outgoing edges from the complex**
  - For each candidate complex, its score is its cohesiveness multiplied by its weighted density

- **Cohesiveness:**

$$\frac{\sum edge\ weights\ inside\ cluster}{\sum edge\ weights\ inside\ cluster + \sum outdoing\ edge\ weights\ from\ cluster}$$

The cohesiveness of a size-2 cluster (a, b) and a size-3 cluster (a, b, c) respectively are:

$$Coh(a,b) = \frac{P'_{(a,b),sm2}}{P'_{(a,b),sm2} + \sum_{x\in\{a,b\},y\in Na,b} (P_{(x,y),sm} + P_{(x,y),\lg})}$$

$$Coh(a,b,c) = \frac{P'_{(a,b),sm3,abc} + P'_{(a,c),sm3,abc} + P'_{(b,c),sm3,abc}}{P'_{(a,b),sm3,abc} + P'_{(a,c),sm3,abc} + P'_{(b,c),sm3,abc} + \sum_{x\in\{a,b,c\},y\in Na,b,c} (P_{(x,y),sm} + P_{(x,y),\lg})}$$

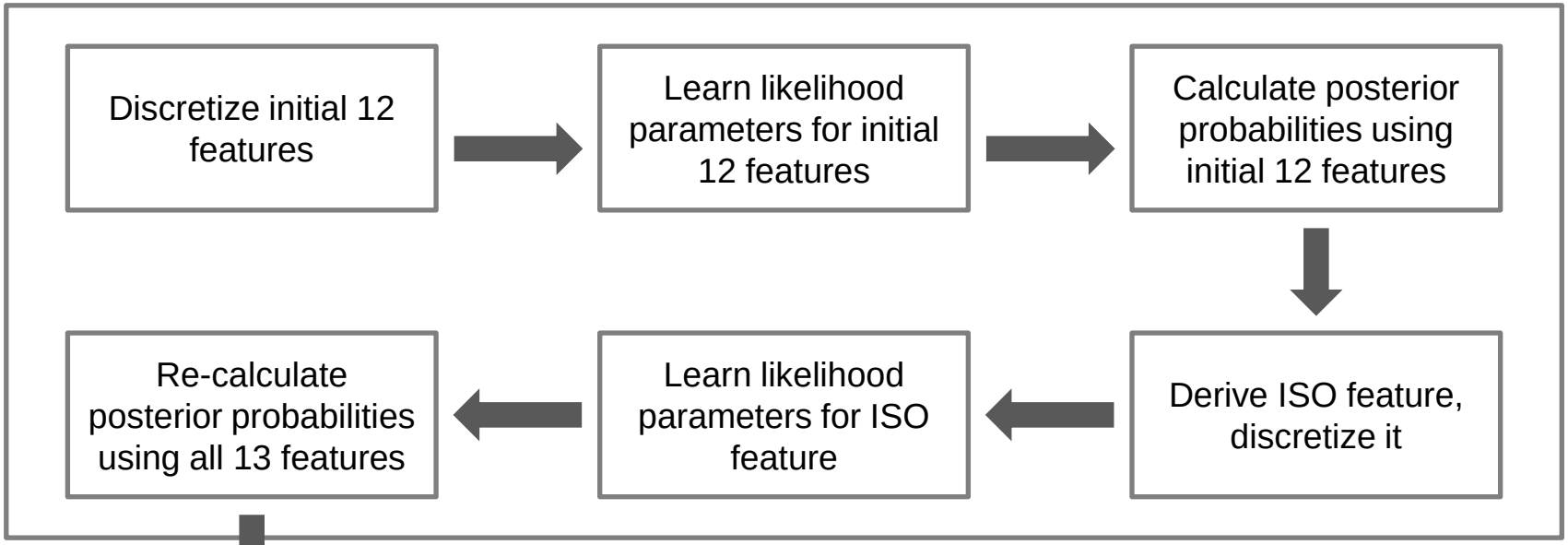
We then define the score of a cluster as its cohesiveness-weighted density, or the product of its weighted density and its cohesiveness. The score of a size-2 cluster (a, b), and a size-3 cluster (a, b, c) respectively are:

$$score(a,b) = Coh(a,b)P'_{(a,b),sm2}$$

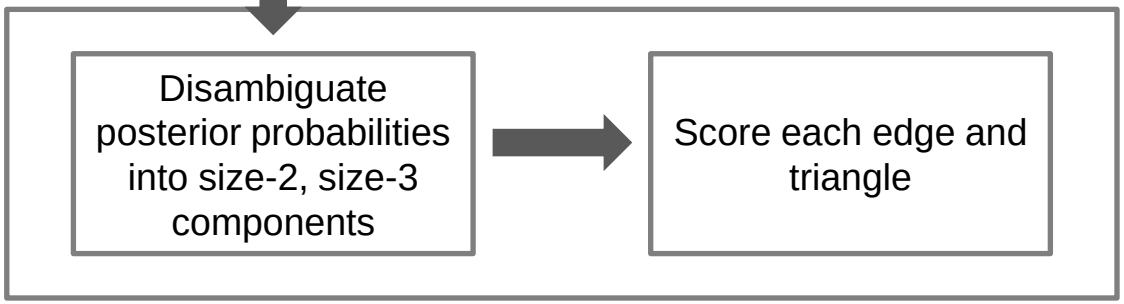
$$score(a,b,c) = Coh(a,b,c)\frac{(P'(a,b),sm3,abc + P'(a,c),sm3,abc + P'(b,c),sm3,abc)}{3}$$

Two-Stage Approach

1. Size-specific supervised weighting (SSS)



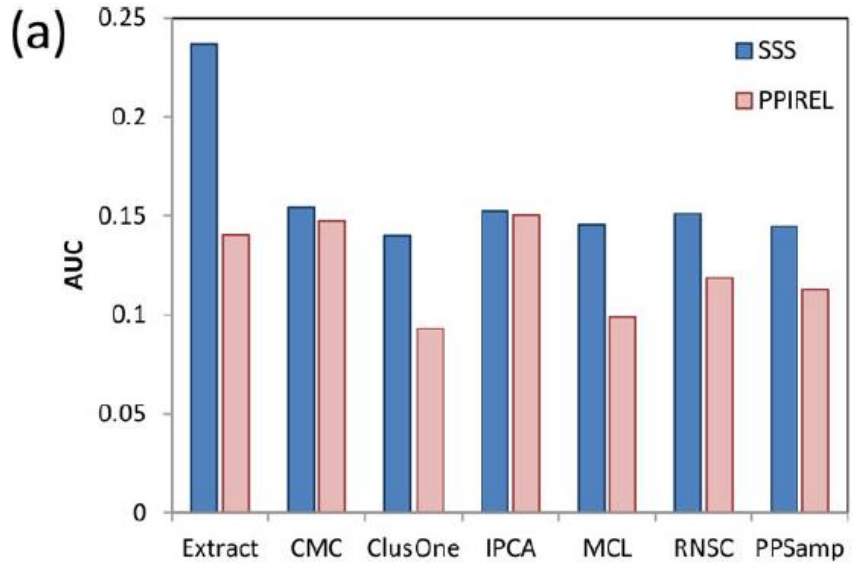
2. Extract

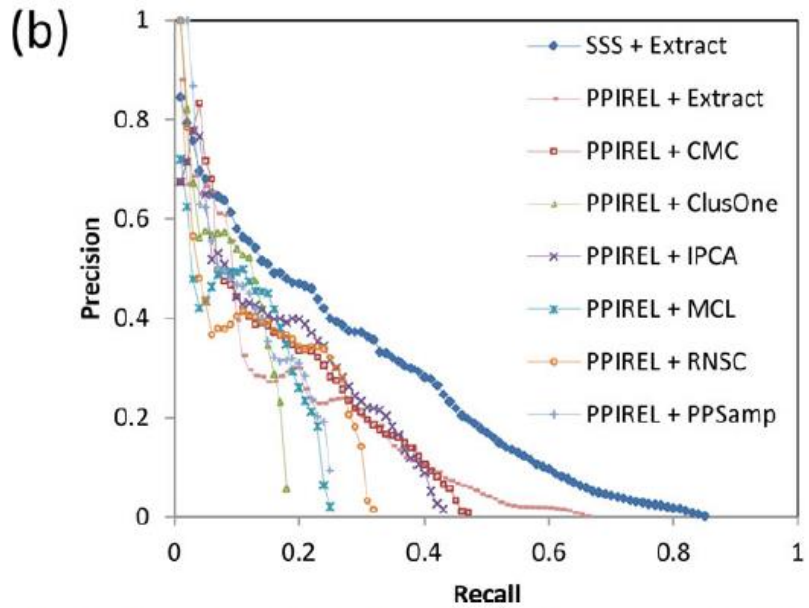


Benefits

- **Groups of proteins may take on small-complex topological characteristics in PPIN by chance**
  - ⇒ Use multiple data sources & their topological features
    - **Unlikely that all data sources share small-complex characteristics by chance**
- **Small-complex prediction is sensitive to noise in PPIN**
  - ⇒ Reduce noise by data integration with supervised learning
- **Other supervised-weighting complex-prediction approaches learn features of large complexes**
  - – Do not perform well for small complexes
  - ⇒ Size-specific weighting
- **Scoring candidate small complexes is sensitive to correct edge weights (very few edge weights used for scoring)**
  - ⇒ Use also outgoing edges from candidate complex during scoring

Yeast Small-Complex Prediction





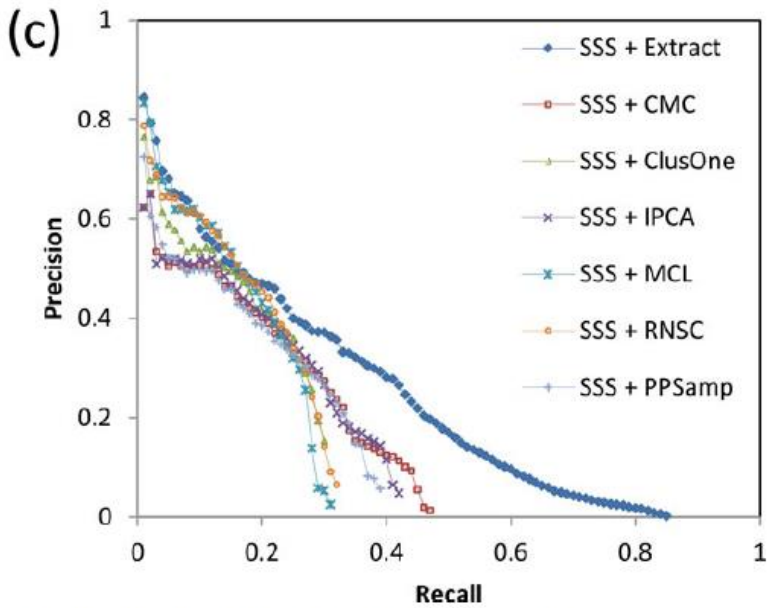
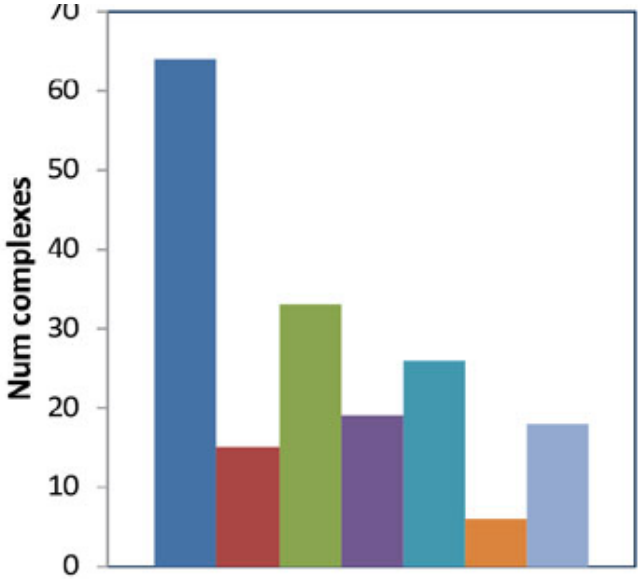
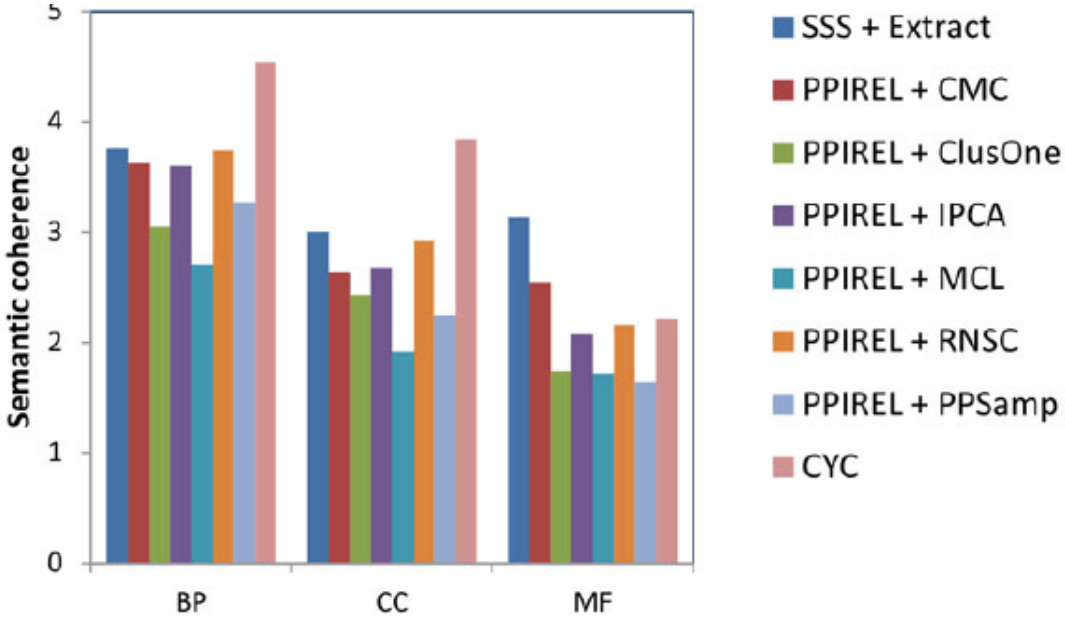


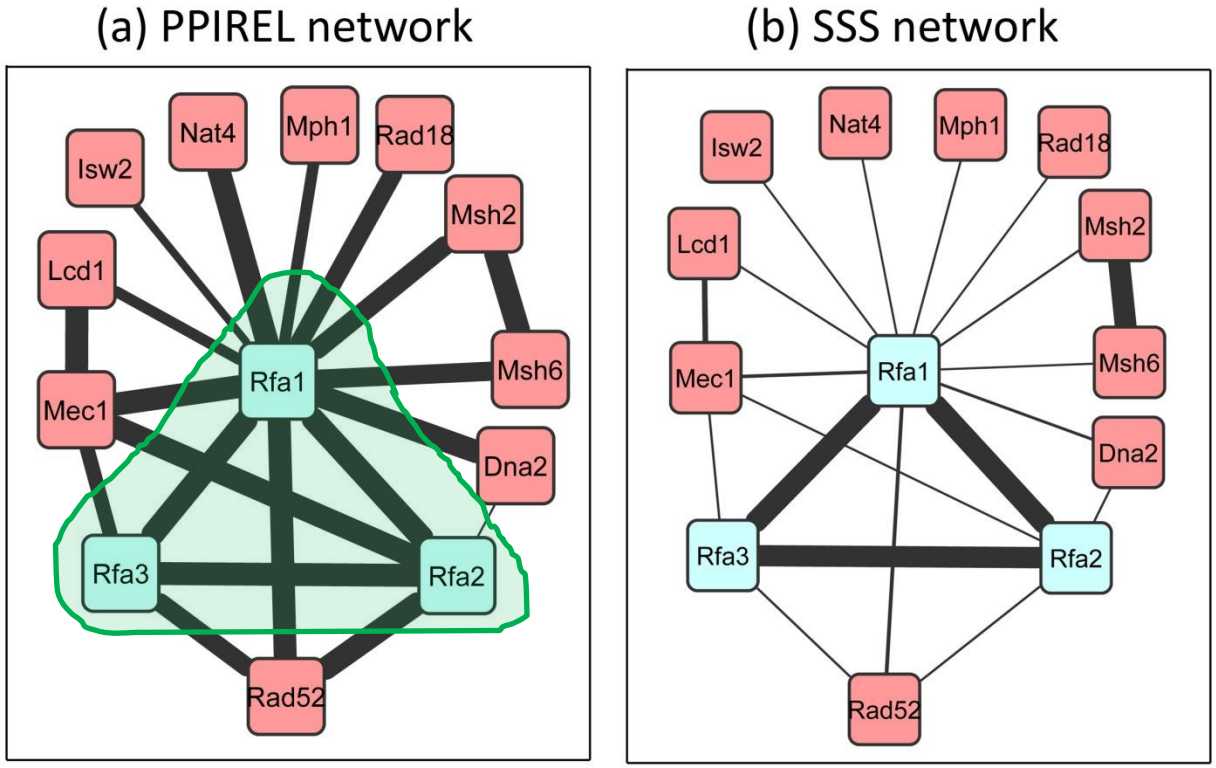
Figure 2 Performance of small complex prediction in yeast, (a) precision-recall AUC, (b) and (c) precision-recall graphs.

Novel Yeast Complexes Predicted





Yeast DNA Replication Factor A



- **DNA replication factor A consists of 3 proteins**
- **Cannot be found by standard clustering algorithms on the PPI network**
  - Embedded within two size-4 cliques
  - Also part of many other size-3 cliques
- **After weighting by SSS, the internal weights of the complex remain high, while extraneous weights are lowered → Can be found in all cross-validation rounds**

Conclusion

- **Most complexes are small, so small-complex prediction is an impt part of complex prediction**

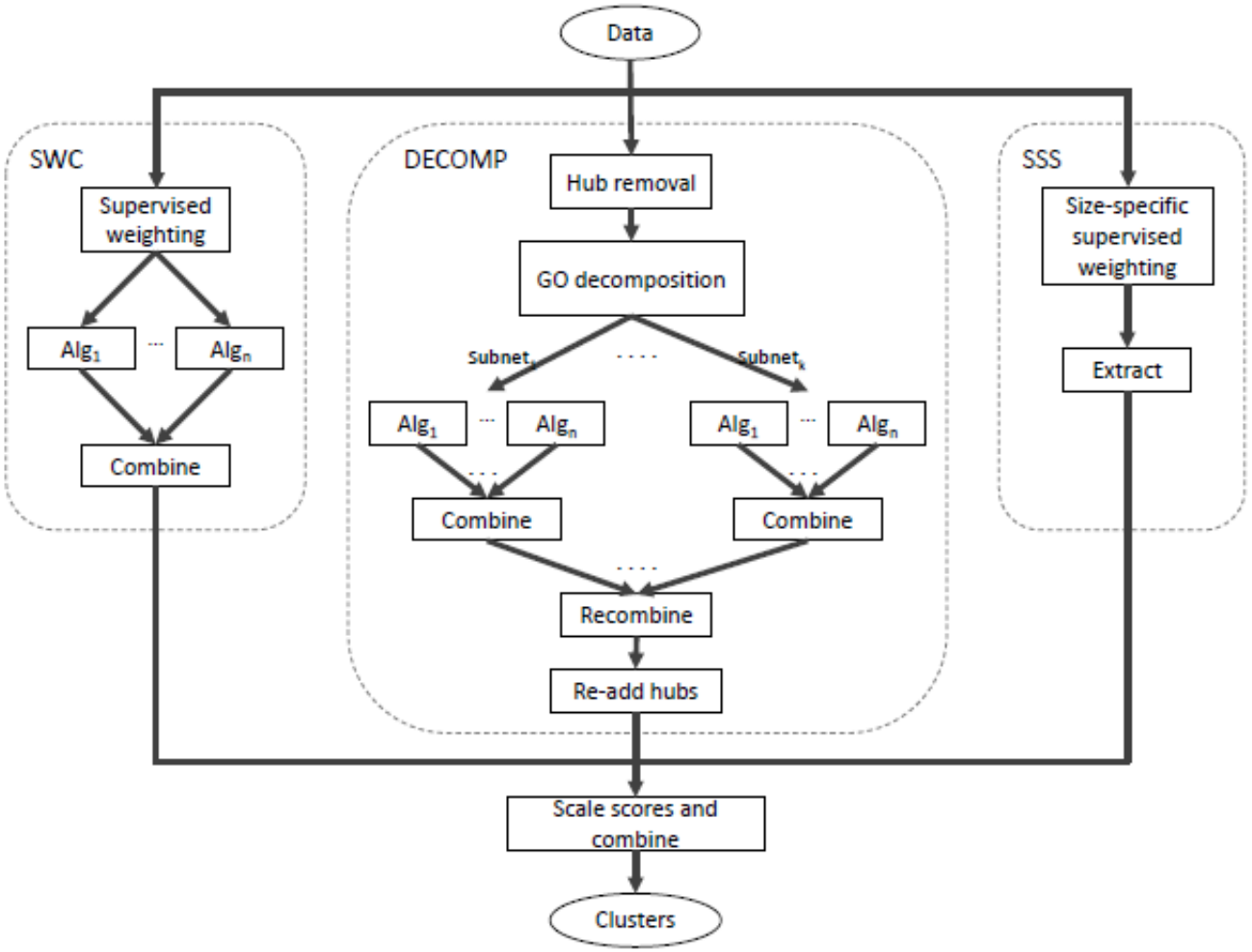
- **Many challenges in small-complex prediction**
  - Searching for dense clusters is ineffectual
  - Sensitive to noise
  - Scoring candidate complexes is sensitive to edge weights

- **SSS + Extract**
  - Integrate 3 data sources w/ their topological features
  - Size-specific edge weighting by supervised learning
  - When scoring candidate complexes, incorporates outgoing edges from clusters as well

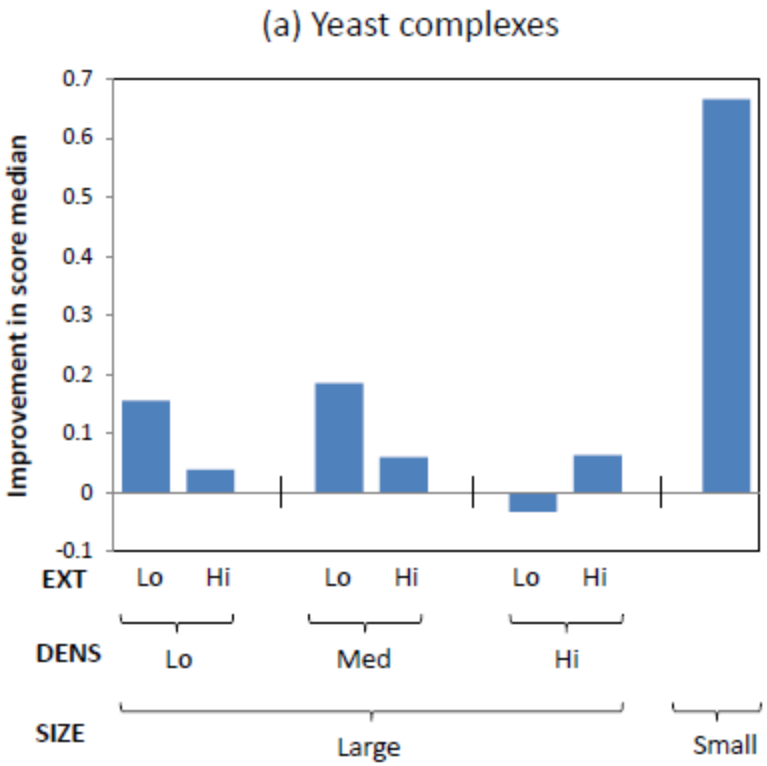
$\Rightarrow$ **Much improved performance in yeast and human**

Putting Everything Together

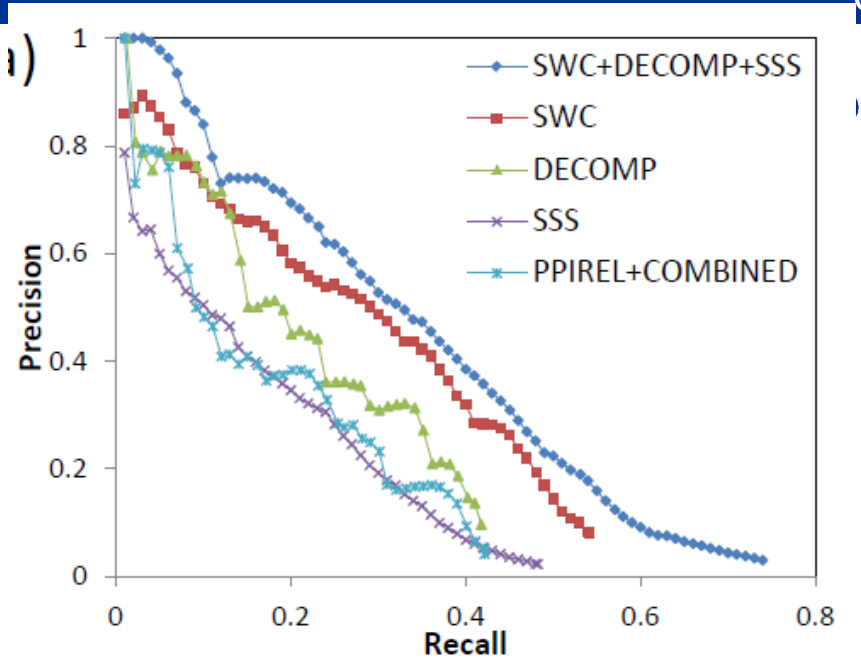
Integrated System of SWC, Decomposition, & SSS

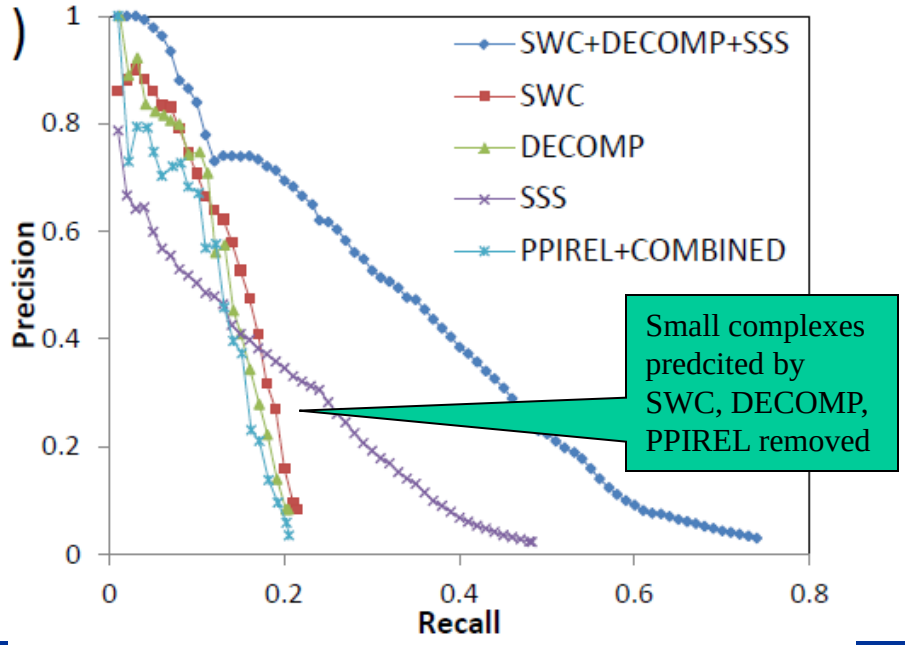


Performance of Integrated System

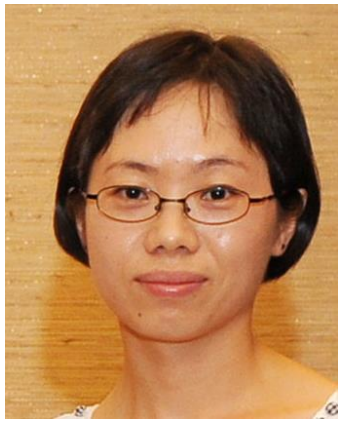


Match scores of top 500 predictions of SWC+DECOMP+SSS vs PPIREL+COMBINED





Acknowledgements



Lui Guimei **Yong Chern Han**

- Liu et al. **Decomposing PPI Networks for Complex Discovery.** *Proteome Science*, 9(Suppl. 1):S15, 2011
- [SWC] Yong et al. **Supervised maximum-likelihood weighting of composite protein networks for complex prediction.** *BMC Systems Biology*, 6(Suppl 2):S13, 2012
- [SSS] Yong et al. **Discovery of small protein complexes from PPI networks with size-specific supervised weighting.** *BMC Systems Biology*, 8(Suppl 5):S3, 2014