Supplementary Material

# Identification of Protein Complexes by Integrating Multiple Alignment of Protein Interaction Networks

Cheng-Yu Ma, Yi-Ping Phoebe Chen, Bonnie Berger, and Chung-Shou Liao

## 1 Further comparison with ClusterONE

We further compared the performance between our algorithm and ClusterONE under the same evaluations proposed in [2], from the website of ClusterONE (http://www.paccanarolab.org /clusterone/). Nepusz *et al.* assessed the quality of the predicted complexes mainly by the following three scores: fraction of matched known protein complexes, maximum matching ratio (MMR), and geometric accuracy measure. The first one is the fraction of the known complexes that match exactly one predicted complex among all known complexes. The MMR score is defined to be the average edge weight of the maximal matching in a bipartite graph, in which the two disjoint sets of nodes represent the predicted complexes and reference (known) complexes, respectively, and each edge between a predicted complex and a reference complex is associated with a weight defined by their overlapping score. This score reflects how accurately the predicted complexes match the reference (known) complexes. The third one is defined as the geometric means of the clustering-wise sensitivity (SN) and the positive predictive value (PPV) [3]. Because the SN score can be inflated by putting every protein in the same predicted complex and the PPV can be maximized by putting every protein in its own cluster, the two measures are usually balanced by their geometric means, i.e., the geometric accuracy measure. Finally, they used the *composite score*, which is the sum of these three scores, to show the overall performance. Readers can refer to the supplementary of Nepusz *et al.*'s work [2] for more details of the evaluations.

| Tools | Fraction of matched known complexes | MMR | SN | PPV | Accuracy | Composite score |
|---|---|---|---|---|---|---|
| Human | | | | | | |
| NEOComplex | 0.30 | 0.07 | 0.37 | 0.13 | 0.22 | **0.58** |
| ClusterONE | 0.17 | 0.04 | 0.31 | 0.20 | 0.25 | 0.46 |
| Yeast | | | | | | |
| NEOComplex | 0.28 | 0.08 | 0.41 | 0.14 | 0.24 | **0.60** |
| ClusterONE | 0.20 | 0.06 | 0.39 | 0.18 | 0.26 | 0.53 |
| Fly | | | | | | |
| NEOComplex | 0.19 | 0.05 | 0.31 | 0.12 | 0.20 | **0.43** |
| ClusterONE | 0.15 | 0.04 | 0.28 | 0.18 | 0.22 | 0.41 |

Table S1. Comparison between NEOComplex and ClusterONE under our reference set: Calculated by Nepusz *et al.*'s evaluation script under our reference set

Table S1 shows the comparison results between our algorithm and ClusterONE in the three species under our reference set. Our algorithm still outperforms ClusterONE in the composite score for all three species. In particular, our algorithm derives better fraction of matched known complexes and MMR,

| Tools | Fraction of matched known complexes | MMR | SN | PPV | Accuracy | Composite score |
|---|---|---|---|---|---|---|
| Human | | | | | | |
| NEOComplex | 0.31 | 0.09 | 0.47 | 0.20 | 0.31 | **0.71** |
| ClusterONE | 0.19 | 0.05 | 0.40 | 0.31 | 0.35 | 0.59 |
| Yeast | | | | | | |
| NEOComplex | 0.41 | 0.15 | 0.34 | 0.40 | 0.37 | **0.93** |
| ClusterONE | 0.32 | 0.15 | 0.37 | 0.44 | 0.40 | 0.88 |

Table S2. Comparison between NEOComplex and ClusterONE under the MIPS reference set: Calculated by Nepusz *et al.*'s evaluation script under the MIPS reference set

while attaining a similar performance in geometric accuracy to that of ClusterONE. This demonstrates that our predicted complexes can more completely represent the known complexes in the reference set.

Moreover, we also referred to the MIPS data [4], which was used as the reference set in [2]. We retrieved the data of protein complexes in human and yeast from MIPS [4, 5]. Note that there are 1139 and 203 protein complexes for human and yeast, respectively, in the MIPS reference set, where there is no fly data. Table S2 presents the comparison results under the MIPS reference set. The performance of our algorithm is stable and still achieves a better composite score than ClusterONE for both human and yeast.

## 2 Effect of weighted edge density thresholds

Observe that our algorithm and ClusterONE both used *weighted edge density* as a threshold to filter out the proteins that have a weak connection with the other proteins in a candidate complex. Such criteria can guarantee the quality of predicted complexes; that is, better settings of the threshold may lead to better performance. Therefore, we investigate the effect of adjusting the parameters on weighted edge density as follows. Figure S1 illustrates the performance of our algorithm and ClusterONE, i.e., precision, recall, and F-measure, under different settings of weighted edge density. Note that the weighted edge density used by our algorithm was defined by NECC (see Section 2.2), but ClusterONE considered the weighted edge density in a weighted PPI network, where each edge weight may represent the interaction reliability or probability (ranged from zero to one) between the pair of proteins. For instance, the NECC weighted edge density of the PPI network for human is $5.61 \times e^{-6}$, and the weighted edge density used by ClusterONE is $1.21 \times e^{-3}$ (for yeast, $2.12 \times e^{-5}$ vs. $3.91 \times e^{-3}$, and for fly, $1.32 \times e^{-5}$ vs. $1.13 \times e^{-3}$). We thus select parameter values from 0.1 to 0.9 for ClusterONE, and from 0.005 to 0.013 for our algorithm. As shown in Figure S1, our algorithm maintains good performance even under the variation of the parameter. Moreover, when the weighted edge density is 0.009, our algorithm can achieve the best F-measure, i.e., the best balance between precision and recall for the three species. More importantly, our algorithm has better F-measure (as well as recall) than that of ClusterONE under its best setting for all three species. Note that ClusterONE has the best F-measure when its density threshold is set to 0.6. In addition, when the density threshold is larger than 0.6, the precision of ClusterONE increases, but its recall and F-measure fall down fast. On the other hand, when the weighted edge density is set to be larger than 0.009 for our algorithm, the recall and F-measure are decreasing slowly, while the precision is increasing.

We remark that Nepusz *et al.* showed that ClusterONE performs well in *weighted* PPI networks of yeast. In contrast, our study focused on unweighted PPI data. Further, when testing on the PPI network of yeast retrieved from BioGRID used in [2], our algorithm still outperforms ClusterONE in composite score (1.14 vs. 1.10), fraction of matched known protein complexes (0.50 vs. 0.47), and MMR (0.22 vs. 0.19). Note that the threshold parameter may affect the excution time of our algorithm and ClusterONE,

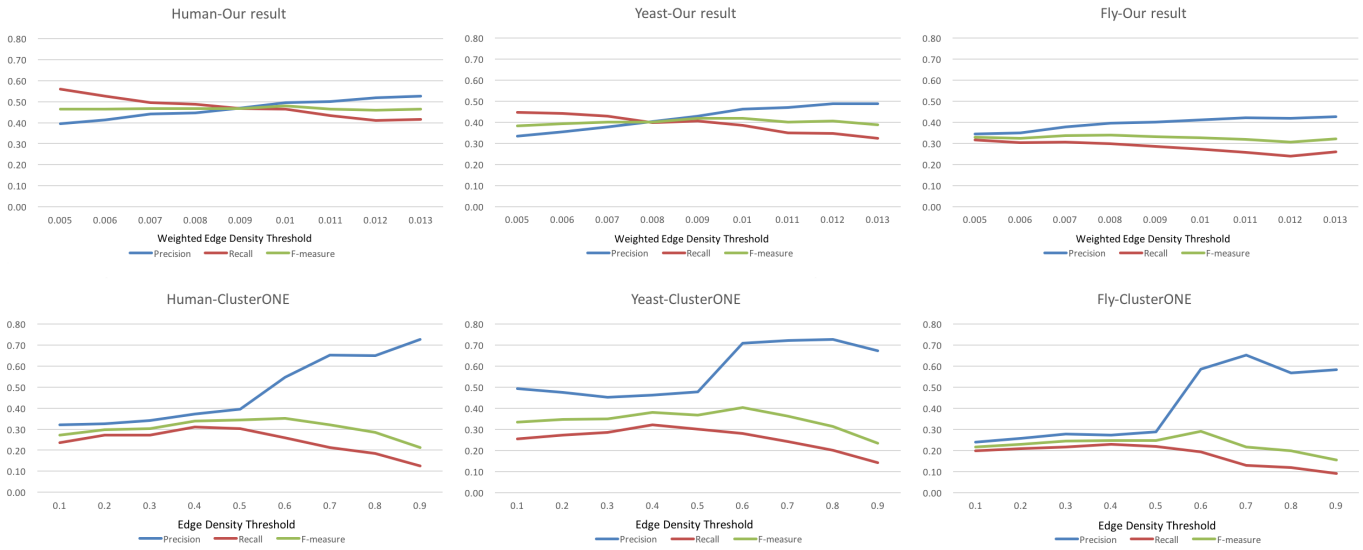


Fig. S1. Effect of weighted edge density thresholds

although both of them run in polynomial time. Briefly speaking, our algorithm counts the number of the common neighbors of the common neighbors. That means, for two end nodes $u$ and $v$, of each edge, it first calculates the number of the common neighbors between $u$ and $v$. (The set of the common neighbors is denoted by $S$.) Next, it counts the common neighbors between $u$ and each node in $S$ and then the common neighbors between $v$ and each node in $S$, respectively. Thus, the theoretical upper bound on the running time of our algorithm is $O(n^2)$, where $n$ is the number of total nodes in a given network.

## 3 Performance of different functional orthology relationships

As discussed, our algorithm can integrate any type of orthology relationships between multiple species into protein complex identification. In this study, we used IsoRankN to obtain functional orthology information across species. To investigate the effect of functional orthology relationships obtained by different multiple network alignment tools, we used four recent global network aligners: SMETANA, BEAMS, NetCoffee, and multiMAGNA++ [1,6–8]. SMETANA and BEAMS derive many-to-many alignments of multiple networks, while NetCoffee and multiMAGNA++ provide one-to-one MNA mappings. SMETANA uses a semi-Markov random walk model to compute the node correspondence scores and output the maximum expected alignment of given networks. BEAMS extracts the cliques of k-partite graph as seeds based on pairwise sequence similarity, and then iteratively merges these seeds to maximize the total alignment score of the whole cluster output set. NetCoffee uses simulated annealing on a set of weighted bipartite graphes to maximize a target function for seeking a global alignment match. multiMAGNA++ is a GA-based search method that optimizes both node and edge conservation at the same time while the alignment is constructed. We conducted the experiments with the following parameter settings. For SMETANA, we used the recommended setting. For BEAMS, we set alpha to 0.2 and beta to 0.4. For NetCoffee, we used the recommended setting and set alpha to 0.2. For multiMAGNA++, we set alpha to 0.8, population size to 100, number of generation to 10000.

For the results of using the orthology information obtained by IsoRankN, SMETANA, BEAMS, NetCoffee, and multiMAGNA++, we briefly call the result of IsoRankN, SMETANA, BEAMS, NetCoffee, and multiMAGNA++ respectively, for short. Table S3 shows that IsoRankN has better performance than the other four approaches in precision, recall, and F-measure in the three species. Moreover, IsoRankN

extracted more *sparse* complexes; The number of sparse complexes obtained by IsoRankN, SMETANA, BEAMS are 17, 0, and 11 in human, 49, 7, and 24 in yeast, 43, 18, and 29 in fly, respectively. Moreover, we note that NetCoffee and multiMAGNA++ which provided one-to-one mappings derived very few sparse complexes in the three species. The result demonstrates that such sparse protein complexes benefit from the assistance of the orthology information obtained by IsoRankN. The major difference which could make IsoRankN more flexible and tolerant to missing interactions is that the tool incorporated topology similarity information when calculating the functional similarity score of each pair of proteins between different species, and then used a star aligned approach for multiple alignment. By contrast, SMETANA and BEAMS considered the topological information while maximizing the overall alignment score of the whole output cluster set. That is, the edge loss in PPI data may affect their performance. We remark that, the results obviously shows that the one-to-one MNA algorithms, such as NetCoffee and multiMAGNA++, are not appropriate for finding protein complexes because their output clusters contains at most one protein from each species. That is our algorithm cannot construct candidate complexes based on such few orthology proteins when incorporating their orthology information.

| Tools | Precision | Recall | F-measure | #Sparse complexes |
|---|---|---|---|---|
| Human | | | | |
| IsoRankN | 0.47 | 0.47 | **0.47** | **17** |
| SMETANA | 0.46 | 0.44 | 0.45 | 0 |
| BEAMS | 0.15 | 0.20 | 0.17 | 11 |
| NetCoffee | 0.44 | 0.45 | 0.44 | 0 |
| multiMAGNA++ | 0.39 | 0.45 | 0.41 | 0 |
| Yeast | | | | |
| IsoRankN | 0.43 | 0.41 | **0.42** | **49** |
| SMETANA | 0.43 | 0.36 | 0.39 | 7 |
| BEAMS | 0.35 | 0.41 | 0.38 | 24 |
| NetCoffee | 0.34 | 0.35 | 0.35 | 3 |
| multiMAGNA++ | 0.29 | 0.34 | 0.32 | 0 |
| Fly | | | | |
| IsoRankN | 0.40 | 0.29 | **0.33** | **43** |
| SMETANA | 0.38 | 0.24 | 0.29 | 18 |
| BEAMS | 0.24 | 0.20 | 0.22 | 29 |
| NetCoffee | 0.35 | 0.23 | 0.28 | 0 |
| multiMAGNA++ | 0.33 | 0.23 | 0.27 | 0 |

Table S3. Performance comparison of our algorithm by using different functional orthology relationships between species: #Sparse complexes represents the number of complexes that can only be found by our algorithm.

## 4 Functional enrichment and consistency

In this section, we use $p$-value which can represent the probability of co-occurrence of proteins with common GO terms [9] to evaluate the predicted complexes derived by each MNA approach. We use GOTERMFINDER [10] to calculate the $p$-value and select significant predicted complexes. We consider the complexes that have $p$-value smaller than 0.01 to be significant complexes. Table S4 shows the

statistics of significant complexes for each MNA approach. At least 43 percent of the complexes derived by our algorithm are considered significant. Note that MCODE and COACH have a higher proportion of significant complexes because the former generated a very low number of output protein complexes (i.e. low recall) and the latter, conversely, derived a large set of predicted complexes with low precision. Moreover, we used *entropy* to measure the consistency of GO terms of our predicted complexes and further compared with MCODE and COACH. The entropy was first exploited in systems biology by Liao *et al.* [11]. The entropy of a given complex $S_v^*$ is :

$H(S_v^*) = H(p_1, p_2, ..., p_d) = \sum_{i=1}^{d} p_i log p_i$

where $p_i$ is the fraction of $S_v^*$ with GO group ID $i$. A complex has lower entropy if its GO annotations are more within-cluster consistent. Table S5 shows the mean entropy of our algorithm, MCODE, and COACH. The result demonstrates that the complexes derived by our algorithm are more consistent than what MCODE and COACH derived in GO annotation. Note that we only consider the Biological Process category of GO terms when we calculate $p$-value and entropy.

| Tools | Human (%) | Yeast (%) | Fly (%) |
|---|---|---|---|
| NEOComplex | 69.18 | 43.29 | 60.96 |
| MCODE | 88.75 | 63.44 | 100 |
| COACH | 93.73 | 43.20 | 89.85 |
| CMC | 54.60 | 13.85 | 37.61 |
| MCL | 43.33 | 16.41 | 38.55 |
| ClusterONE | 66.17 | 33.49 | 74.60 |

Table S4. Percentage of significant complexes: we consider the protein complexes that have $p$-value smaller than 0.01 to be significant complexes.

| Tools | Human | Yeast | Fly |
|---|---|---|---|
| NEOComplex | 1.15 | 1.40 | 1.58 |
| MCODE | 2.45 | 2.65 | 2.47 |
| COACH | 2.35 | 1.80 | 3.14 |

Table S5. Average entropy of predicted complexes

## References


[1] Alkan, F. and Erten, C. (2014) BEAMS: backbone extraction and merge strategy for the global many-to-many alignment of multiple PPI networks. *Bioinformatics*, **30(4)**, 531–539.

[2] Nepusz, T., Yu, H., and Paccanaro, A. (2012) Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods*, **9**, 471–472.

[3] Brohee, S. and van Helden, J. (2006) Evaluation of clustering algorithms for protein-protein interaction networks. *BMC Bioinformatics*, **7**, 488.

[4] Mewes, H.W., *et al.* (2004) MIPS: analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.*, **32**, D41–D44.

[5] Ruepp, A., *et al.* (2008) CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res.*, **36**, D646-D650.

[6] Sahraeian, S.M.E. and Yoon, B.J. (2013) SMETANA: Accurate and Scalable Algorithm for Probabilistic Alignment of Large-Scale Biological Networks. *PLoS ONE*, **8(7)**, e67995.

[7] Hu, J., Kehr, B., and Reinert, K. (2014) NetCoffee: a fast and accurate global alignment approach to identify functionally conserved proteins in multiple networks. *Bioinformatics*, **30(4)**, 540–548.

[8] Vijayan, V. and Milenković, T. (2016) Multiple network alignment via multiMAGNA++. *In Proceedings of the 15th International Workshop on Data Mining in Bioinformatics (BIOKDD) at the 22nd ACM SIGKDD 2016 Conference on Knowledge Discovery & Data Mining (KDD)*, San Francisco, CA, USA, August 13-17.

[9] Wu, M., Li, X., Kwoh, C.K., and Ng, S.K. (2009) A core-attachment based method to detect protein complexes in PPI networks. *BMC Bioinformatics*, **10**, 169.

[10] Boyle, E.I., *et al.* (2004) GO::TermFinderXopen source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes *Bioinformatics*, **20(18)**, 3710–3715.

[11] Liao, C.S., *et al.* (2009) IsoRankN: spectral methods for global alignment of multiple protein networks. *Bioinformatics*, **25**, i253–i258.