Xiao-xia Zhang, Qiang-hua Xiao, Bin Li; Sai Hu, Hui-jun Xiong, Bi-hai Zhao, 2015. Overlap maximum matching ratio (OMMR): a new measure to evaluate overlaps of essential modules. *Frontiers of Information Technology & Electronic Engineering*, **16**(4):293-300. [doi:10.1631/FITEE.1400282]

Overlap maximum matching ratio (OMMR): a new measure to evaluate overlaps of essential modules

Key words: Protein-protein interaction network, Essential protein modules, Overlap, Overlap maximum matching ratio

Contact: Bi-hai Zhao

E-mail:bihaizhao@163.com

ORCID: http://orcid.org/0000-0003-0870-7468

Introduction

- Protein complexes and their overlaps are helpful for the detection of essential proteins and others.

- Effective evaluation of overlapping complexes and their overlaps could improve the precision of identification of essential proteins.

- Overlaps supporting these algorithms (ClusterONE excluded) ignore the assessment of the identified overlapping module pairs. Even for ClusterONE, only overlap size distribution between generated complexes pairs is shown. Further study of evaluating the quality of overlaps between cluster pairs is lacking.

Emphasis on this work

Given a benchmark modules set BM, SBM= {sbm_1, sbm_2, ..., sbm_n} is a subset of BM, in which each module matches at least one protein by predicted protein complexes. PM is a predicted protein complexes set, and

$$\text{SPM}=\{\max_{1\le j\le|\text{PM}|}\text{OS}(\text{sbm}_1,\text{pm}_j),\ \max_{1\le j\le|\text{PM}|}\text{OS}(\text{sbm}_2,$$

$$\text{pm}_j),\ldots,\max_{1\le j\le|\text{PM}|}\text{OS}(\text{sbm}_n,\text{pm}_j)\}$$

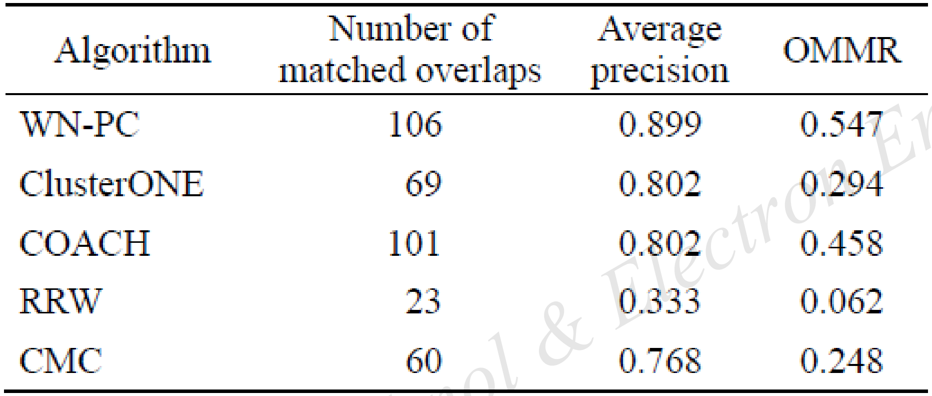
The OMMR can be defined as

$$\text{OMMR}=\frac{1}{|\text{SBM}\times\text{SBM}|}\sum_{i=1}^{|\text{SBM}|}\sum_{i=1}^{|\text{SBM}|}\text{spm}_i\cap\text{spm}_j.$$

Experimental results (1)

(1) OMMR

Table 3 Comparison of overlaps predicted by various algorithms



Algorithm	Number of matched overlaps	Average precision	OMMR
WN-PC	106	0.899	0.547
ClusterONE	69	0.802	0.294
COACH	101	0.802	0.458
RRW	23	0.333	0.062
CMC	60	0.768	0.248

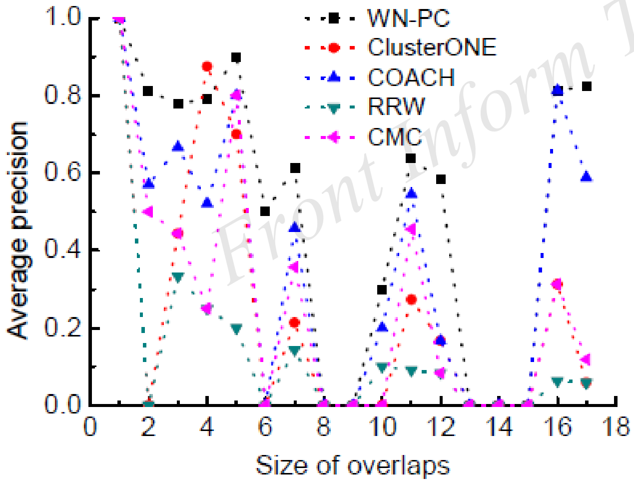


Fig. 1 **Comparison of average precision according to the size of overlaps**

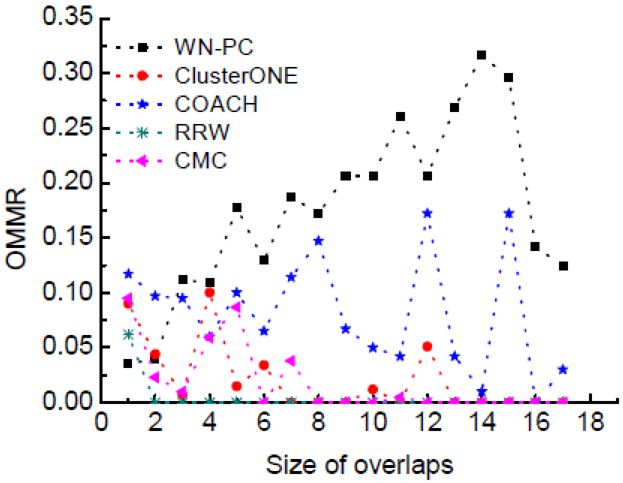


Fig. 2 **Comparison of OMMR according to the size of overlaps**

Experimental results (2)

(2) Comparison of essential proteins in overlaps

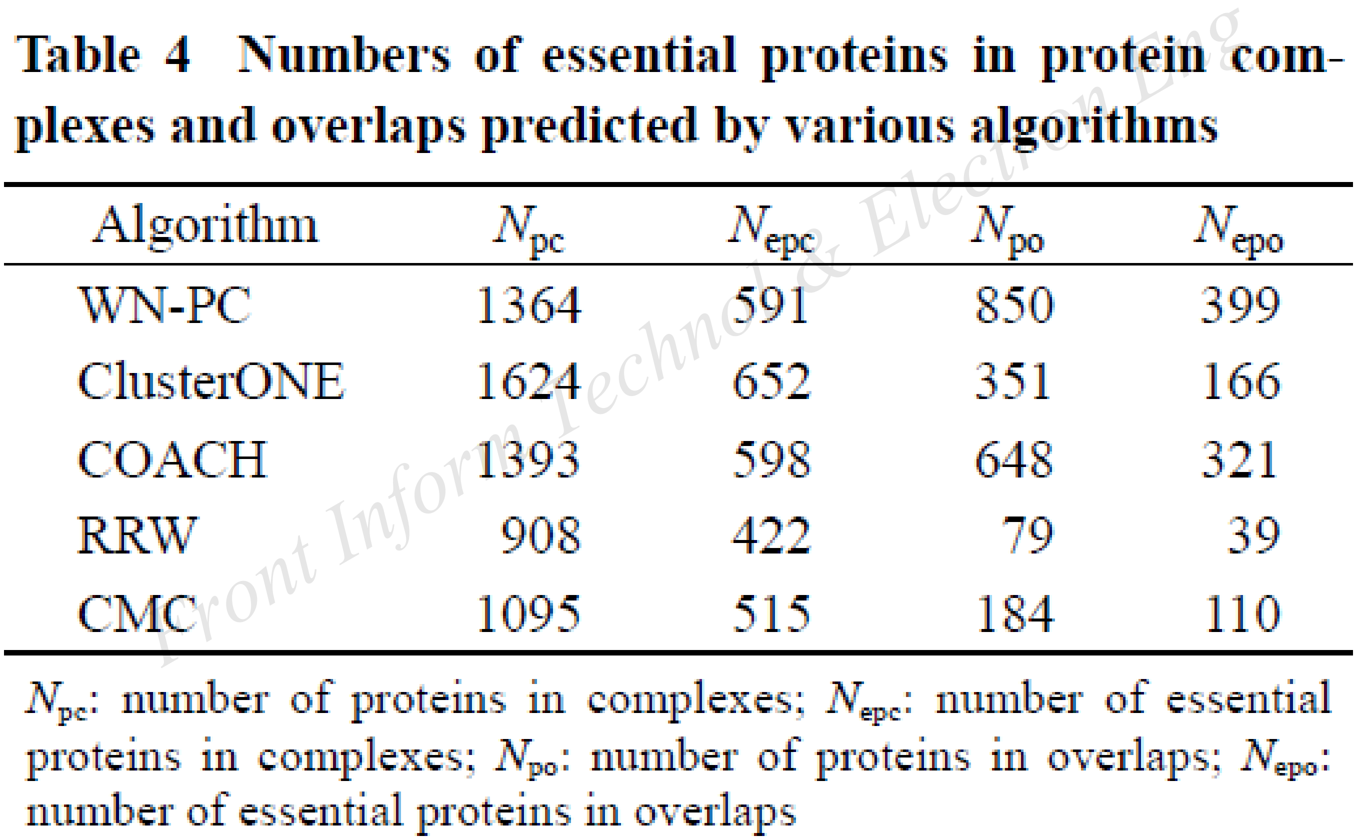


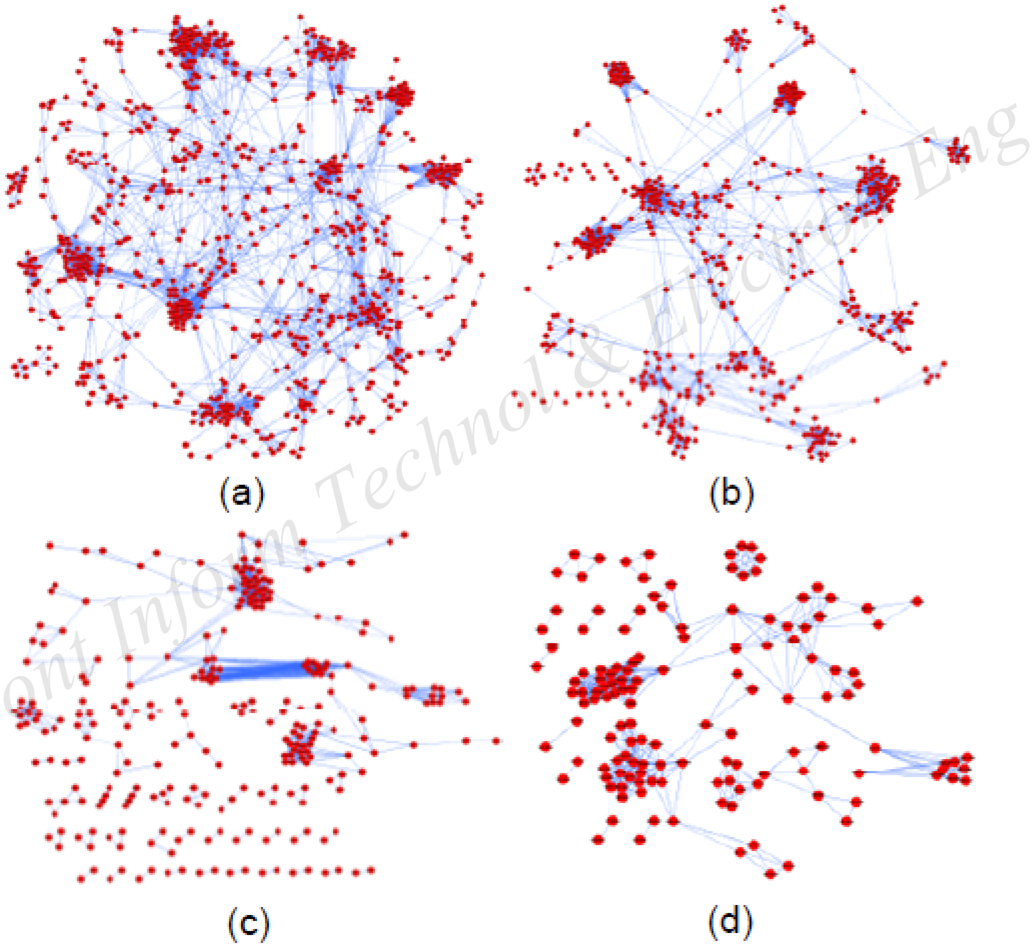
Table 4 Numbers of essential proteins in protein complexes and overlaps predicted by various algorithms

Algorithm	N_{pc}	N_{epc}	N_{po}	N_{epo}
WN-PC	1364	591	850	399
ClusterONE	1624	652	351	166
COACH	1393	598	648	321
RRW	908	422	79	39
CMC	1095	515	184	110

N_{pc}: number of proteins in complexes; N_{epc}: number of essential proteins in complexes; N_{po}: number of proteins in overlaps; N_{epo}: number of essential proteins in overlaps

Experimental results (3)

(3) Visualization



(a) (b) (c) (d)

Fig. 3 Visualization of overlaps by various algorithms

(a) WN-PC; (b) COACH; (c) ClusterONE; (d) CMC

Conclusions

Some overlapping protein complexes prediction methods have been developed. Our previous research shows that overlapping protein complexes, especially their overlaps, play important roles in identifying essential proteins. In this paper, we propose a measure called OMMR to evaluate overlaps of essential modules. The experimental results indicate the importance of overlaps and reveal the relationship between overlaps and identification of essential proteins.