

Detection of protein complex from protein-protein interaction network using Markov clustering

P J Ochieng^{1,2}, W A Kusuma^{1,2,3*} and T Haryanto^{1,3}

¹Department of Computer Science, Faculty of Mathematics & Natural Sciences, Bogor Agricultural University Dramaga, Bogor 16680, Bogor, Indonesia

²Kenyatta National Hospital, P O BOX 20723-00202, Upper hill, Nairobi, Kenya

³TropicalBiopharmaca Research Center, Bogor Agricultural University, Jl. Taman Kencana No. 3, Bogor 16128, Indonesia

E-mail: ananta@apps.ipb.ac.id

Abstract. Detection of complexes, or groups of functionally related proteins, is an important challenge while analysing biological networks. However, existing algorithms to identify protein complexes are insufficient when applied to dense networks of experimentally derived interaction data. Therefore, we introduced a graph clustering method based on Markov clustering algorithm to identify protein complex within highly interconnected protein-protein interaction networks. Protein-protein interaction network was first constructed to develop geometrical network, the network was then partitioned using Markov clustering to detect protein complexes. The interest of the proposed method was illustrated by its application to Human Proteins associated to type II diabetes mellitus. Flow simulation of MCL algorithm was initially performed and topological properties of the resultant network were analysed for detection of the protein complex. The results indicated the proposed method successfully detect an overall of 34 complexes with 11 complexes consisting of overlapping modules and 20 non-overlapping modules. The major complex consisted of 102 proteins and 521 interactions with cluster modularity and density of 0.745 and 0.101 respectively. The comparison analysis revealed MCL outperform AP, MCODE and SCPS algorithms with high clustering coefficient (0.751) network density and modularity index (0.630). This demonstrated MCL was the most reliable and efficient graph clustering algorithm for detection of protein complexes from PPI networks.

1. Introduction

In genomics studies, large-scale experiments often produce huge data sets of protein-protein interactions making it difficult to visualize and analyze the information contained in these data [1][2]. Therefore, application of computational methods can alleviate a lot of problems in this regard. Thus a general trend is to represent the interactions as a network/graph and to apply suitable graph algorithms to extract necessary information. In the post-genomic era, the most important issues are to find protein complexes from the protein-protein interaction (PPI) networks. Identification of those complex can help us to predict the functions of proteins [3], and they can also play an important role in understanding and explain certain biological processes.

* Corresponding author



The results obtained from different technologies for detection of high-throughput protein-protein interactions such as yeast two-hybrid assay (Y2H) and mass spectrometry of purified complexes say tandem affinity purification (TAP) and high-throughput mass spectrometric protein complex identification (HMS-PCI) show some variations[4][5]. For instance, the common PPI between the two different mass-spectrometry approaches stands at 1,728 pairs, which correspond to 27.5% of PPI detected by TAP and 19.2% of PPI detected by HMS-PCI. This implies that many variations in experimentally determined interaction might be false positive are incomplete. Hence, generation of protein complexes based on interaction networks of separate or combined data sets is helpful because the interactions that are involved in complexes are likely to be true.

Already a number of approaches have been proposed for detection of protein complexes in PPI networks[6]. The sequential constructive method of makes use of the concepts of clustering coefficient and k-core graphs [7] [8].

Another approach described in [9] use hierarchical clustering. However, they introduced the concept of secondary distances instead of considering the path length as the distance between a pair of proteins because of the fact that such distances among proteins are constrained and often cause distance ties[10]. The approach of [11] starts by composing an initial random clustering and then iteratively moving one node from one cluster to another in a randomized fashion to improve the clustering's cost [12]. Once the clusters are generated, they are filtered based on cluster size, density and functional homogeneity keeping in mind the criteria of the known biological complexes. Another approach related to analyzing protein complexes is super-paramagnetic clustering [13] [14]. Furthermore, module based approach such as the Molecular Complex Detection (MCODE), Affinity Propagation (AP) and Super Paramagnetic Clustering (SPC) have been proposed recently to detect protein complexes [15, 16,17].

Recently, AP have been proposed in the genomic study of complex proteins in PPI network since it identifies exemplars within the dataset by exchanging real-valued messages between all data points are then grouped with their most representative exemplar to give the final set of clusters [18][19]. Furthermore, Affinity Propagation (AP) have been applied gene identification from putative exons using microarray data, and have proven to be faster and more accurate than other clustering methods like K-Centers clustering algorithm. However, studies have demonstrated the algorithm function similar to Vertex Substitution Heuristic (VSH) algorithm which is often less reliable for protein complex detection [20, 21, 22]. Recently studies have suggested that clusters or locally dense regions of an interaction network might represent protein complexes [23]. However, the term "locally dense region" implies a very flexible concept. Some well-known clustering methods are k-core, k-block, k-plex and n-clan clustering where k-core is a maximal subgraph such that each node in the subgraph has at least degree k, however strategies are usually based on the number of node degrees or the number/length of paths between two nodes within the cluster which make it challenging for detection of complex protein in large PPI networks [24, 25, 26].

Therefore, in this paper we proposed Markov Clustering (MCL) to partition the PPI network to facilitate detection of complex proteins. Thus we assume that the interaction network is an undirected simple graph where a graph is undirected if its edges are not directed and a graph is simple if it has no parallel edge or self-loop. We also initially perform MCL parameter optimization before identification of protein complexes from PPI network based on the network modularity and density from the MCL cluster granules.

2. Material and Methods

2.1. Data preparation

In this study, 2053 protein datasets associated with type 2 diabetes diseases obtained from Human Reactom Protein Database (HRPD) [27]. The dataset was pre-processed to eliminate any artefact and check for missing data.

2.2. Construction of PPI Network

Protein-protein interaction network (graph) was initially built to link protein pair that share similar biological function. Where G represented the graph (network), and the vertex (V) be the node (protein) and the edge (E) be the interaction (protein interaction). Therefore, we represented graph $G = (V, E)$ where mathematically it consist of finite nonempty sets V of nodes (proteins) where, a set $E \subseteq V \times V$ of edges consisting of unordered pairs of vertices (proteins) with connecting edges (E). To extract complex protein network we build a subgraph (network) in this case we let subgraph (H) be $H = (V_H, E_H)$ of protein-protein $G = (V, E)$ if $V_H \subseteq V$ and $E_H \subseteq E$. We also say that G is a supergraph of H . Given a subset of the vertices $V' \subseteq V$. In this case, the induced subgraph $G' = (V', E')$ consists exactly of all the edges (protein interaction) present in G between vertices in V' . In this case, for all $v_i, v_j \in V'$, $(v_i, v_j) \in E' \iff (v_i, v_j) \in E$. In other words, two nodes are adjacent in G' (protein complex) if and only if they are adjacent in G (protein network). From the constructed network, it is easy to identify the neighbour distribution and connectivity, therefore the next step is to partition the network via MCL to detect the sub-network (complexes) with highest interaction and degree distribution.

2.3. Network partitioning via MCL algorithm

To partition the PPI network we subject the network to Markov Clustering (MCL). In this case, we let G be a graph on n nodes and $M = M_G$ be the associated matrix. We then denote graph G associated with Markov matrix by $\mathcal{T}_G$ which we formally defined by letting its q^{th} column be the q^{th} column of M normalized. Since our graph is undirected, we let $M = M_G$ be the associated matrix and $\mathcal{T} = \mathcal{T}_G$ be the associated Markov matrix. Again, to this end, we used $\mathbf{d}$ to denote the diagonal matrix that has diagonal entries the column M , thus $d_{kk} = \sum_i M_{ik}$, and $d_{ij} = 0$; $i \neq j$ while $\mathcal{T}_G$ is defined by (1).

$$\mathcal{T}_G = M_G \mathbf{d}^{-1} \quad (1)$$

Then the number of positive, negative and zero eigenvalues are the same for T and M . In this case the matrix $M \in \mathbb{R}^{k \times l}$, $M \geq 0$, and a real nonnegative number r , the resultant matrix from rescaling each of the columns of M with power coefficient r is represented by $\Gamma_r M$, where Γ_r is the inflation operator with power coefficient r . Thus, theoretically, the action of $\Gamma_r: \mathbb{R}^{k \times l} \rightarrow \mathbb{R}^{k \times l}$ is defined by (2).

$$(\Gamma_r M)_{pq} = (M_{pq})^r / \sum_{i=1}^k (M_{iq})^r \quad (2)$$

In this case, we tune the inflation parameter (r) in (2) to generate the optimal number of clusters. The aim of tuning the inflation operator is to make it stronger, to enhance cluster granularity or tightness. Therefore, by setting maximum iteration for expansion and inflation we enhance the graph partitioning into different segments until equilibrium is achieved hence we interpret the collections of segments as protein complexes. Below is representation of the proposed MCL algorithm (Figure 1).

MCL Algorithm

G is a graph
 add loops to **G**
 set Γ to some value # affects granularity
 set **M_1** to be the matrix of random walks on **G**

```

while (change) {
M_2 = M_1 * M_1           # expansion
M_1 =  $\Gamma$ (M_2)             # inflation
  change = difference(M_1, M_2)
}

```

set CLUSTERING as the components of **M_1**

Figure 1. MC algorithm

2.4. Parameterization

To optimize the clustering performance of the proposed MCL we initially perform flow simulation by setting the inflation value (r) within a range of 1.4 to 2.5 power coefficient this was according to [28]. The ultimate goal was to determine the optimal inflation value for maximum cluster granularity in other words efficient network partitioning. Thereby, in this study flow simulation acts as preliminary MCL optimization for accurate detection of protein complexes.

2.5. Detection of protein complexes from MCL clusters

In the current study protein complexes from the network were calculate based modularly score and density of the protein community (cliques) cluster granules generated by MCL. Thereby we used the formulation in (2.5.1) and (2.5.2) to determine the significant protein complex.

2.5.1 Modularity score. To detect protein complexes from the MCL networks we computed the modularity for each network using global modularity calculation according to [29] and [30]. This measure provides a composite modularitiescore across all clusters and is defined as:

$$\text{Modularity} = \sum_{c \in C} \left[\frac{E_{cIn}}{E_{total}} - \left(\frac{2E_{cIn} + E_{cOut}}{2E_{total}} \right) \right] \quad (3)$$

where, for each cluster c in the set of all clusters C , E_{cIn} , E_{cOut} and E_{total} represent the number of edges within the cluster, the number of edges leading out of the cluster, and the total edges in the network, respectively. We note that while the global modularity score only considers clusters that are contained within the main graph component, in practice this does not significantly affect the results because few or no clusters in the networks we consider them isolated from the main component. In some case we calculate local modularity for each cluster using (4).

$$C_{mod} = E_{in}/E_{out} \quad (4)$$

2.5.2 Density. Calculation of cluster densities plays an important role in identifying the major complexes in the partitioned PPI Network. Here be considered density d_k of any cluster k as the ratio of the number of edges present in the cluster ($|E_k|$) and the maximum possible number of edges in the cluster ($|E_k|_{max}$) and is represented by (5).

$$d_k = \frac{|E_k|}{|E_k|_{max}} = \frac{2 \times |E_k|}{|N_k| \times (|N_k| - 1)} \quad (5)$$

Thus, $|N_k|$ is the size of the cluster, i. e. the number of nodes in the cluster however it should be noted that density of a given cluster is a real number ranging from 0 to 1.

3. Results

The prime goal of the current study was to efficiently detect protein complexes from PPI network via MCL partitioning and network topological calculation, therefore we initially generated the PPI network (3.1) then perform the flow simulation (3.2) to determine optimal parameters values for MCL this was then followed by analysis of overall quality of complex (3.3) finally evaluation of MCL performance results (3.4).

3.1. Original PPI network

The unweighted PPI network was built 2254 protein pairs of proteins. The resultant network (Figure 2) consist of 482 nodes (proteins) with 2054 edges (interactions) with network density = 0.08 and

average neighbour connectivity= 8.51 demonstrating that there is strong transitive relationship (interaction) between protein pairs in the network.

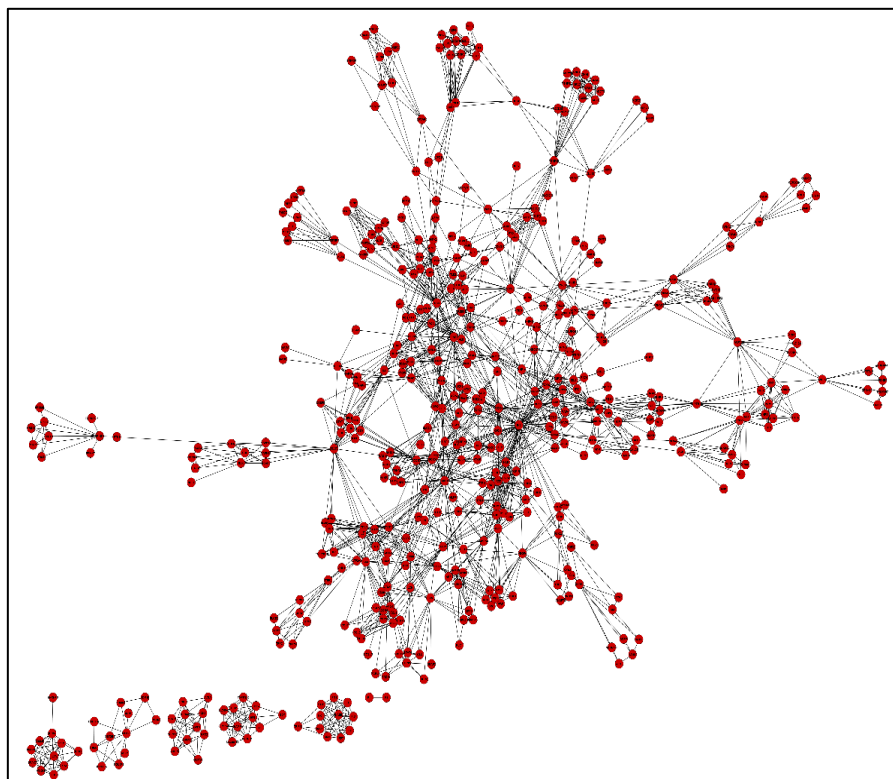


Figure 2. Original unweighted PPI network comprising of 482 proteins and 2054 interaction generated from protein interaction pairs associated with type 2 diabetes.

3.2. Flow simulation

In this study we initially perform flow simulation to search optimal inflation value for maximum cluster granularity, therefore we set the inflation value within a range of 1.4 to 2.5 simply because the inflation parameter is responsible for both strengthening and weakening the current (strengthen the already strong currents and weaken the already weak currents). Furthermore, by setting the inflation range we are able to obtain the optimal inflation value with high strengthening or weakening property in the end influences the granularity of clusters. From our simulation adjusting the inflation influence the average size of cluster and the maximum size of cluster as this had the over impact on the number of clusters generated from the original PPI network (Figure 3). Flow simulation clearly demonstrate setting inflation parameter at $R = 1.4$ resulted in 13 protein complexes (clusters granules) likewise when we set the maximum value $R = 2.5$ number of the identified protein complexes is extremely high (49 clusters) leading to over-representation due to high strengthening of current this result in high cluster granularity and this may not reflect the true protein complexes within the PPI network.

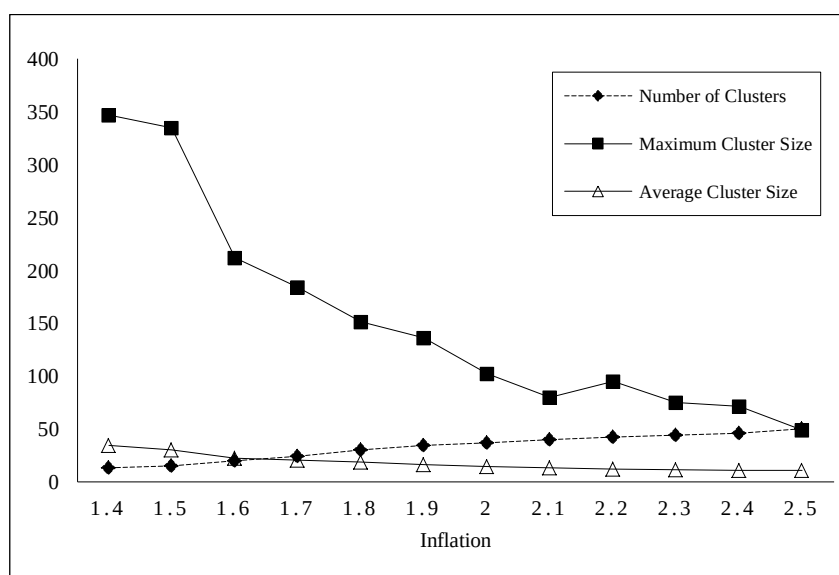


Figure 3. Effect of inflation adjustment on average cluster size, number of cluster granules and maximum cluster size.

This implies by setting inflation $R=1.4$ the resultant output is a set of nested overlapping clustering's on the other hand extremely high inflation value leads to high cluster granularity resulting to under-representation of the true clusters. Interestingly, it can also be evidence from the result as the number of inflation increase gradually the number of detected protein complexes also increases. This is very satisfactory, as one expects clusters at different levels of granularity to be related to each other. It holds that when inflation parameter is set sufficiently low value, the network will be clustered into a single cluster. However, when the dealing with large networks where the diameter is extremely large, the algorithm is likely to experience time complexity since random network are built on distinctly different probability distributions several expansion steps at optimal inflation parameter is required equalize the matrix. Thus in this analysis the optimal inflation values obtained at value 1.9 which gave 34 complexes (clusters) with a maximum cluster size of 136 (Figure3) and network density of 0.14 as well as the modularity score of 0.61 (Figure 4).

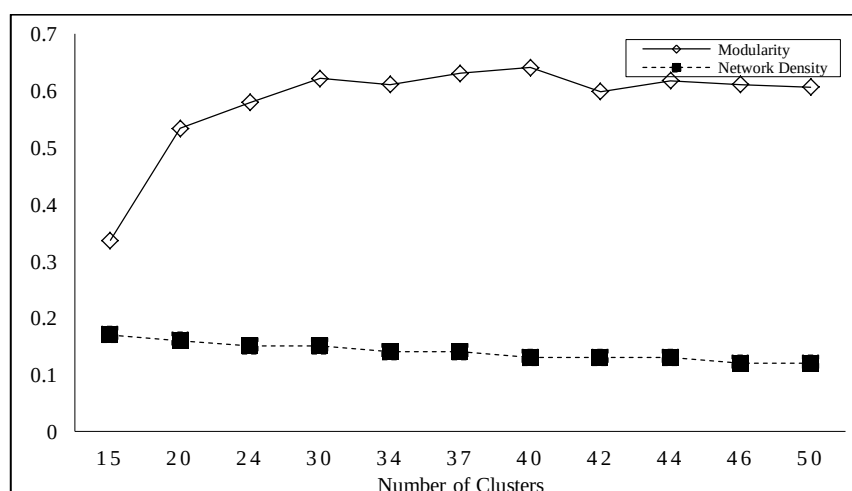


Figure 4. Effect of change in number of clusters to network modularity and density.

We further assess the effect inflation value on modularity and network density (Figure 5), from the simulation we discovered that an modularity score increase with rise in inflation value whereas the network density decrease as we increase the inflation values, in network perspective cluster modularity and density represents the level of connectivity within a group of nodes relative to the group's connections to the rest of the network. Previously (2.5.1) we defined modularity as the ratio of the number of edges between nodes in a cluster (in-degree, E_{in}) to the number of edges between members of the cluster and any neighbors not designated as members of the cluster (out-degree, E_{out}) therefore by performing the flow simulation. In our case, high modularity score indicated that a cluster is very isolated from the rest of the network (Figure 4).

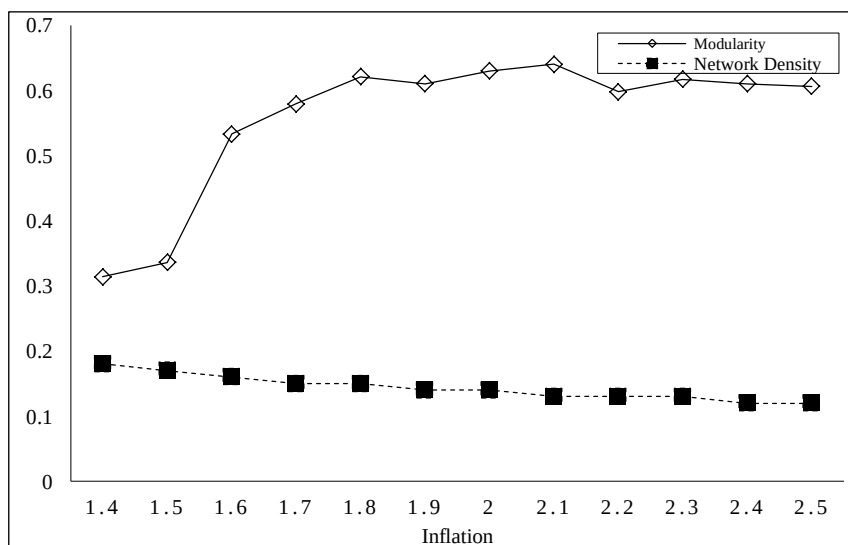


Figure 5. Effect of adjusting inflation parameter on PPI network modularity and density

Performing flow simulation enable us choose the optimal inflation value for MCL algorithm for maximum networktherefore set optimal inflation value MCL generate 34 clusters modules (protein complex) with the major complex consisting of 103 proteins and 521 interactions (Figure 6) likewise smallest protein complex consisted of 3 protein and 4 interactions (Figure 7). Generally, average clusters modularity was 0.63 with the clustering coefficient of 0.742 and density of 0.077. This demonstrated that majority of the clusters highly connected probably due to optimal inflation value, furthermore the majority of complex identified had 100% shortest path between nodes within cluster this indicated strong intra-clusters. This played an important role discovery of proteins sharing similar biological function.

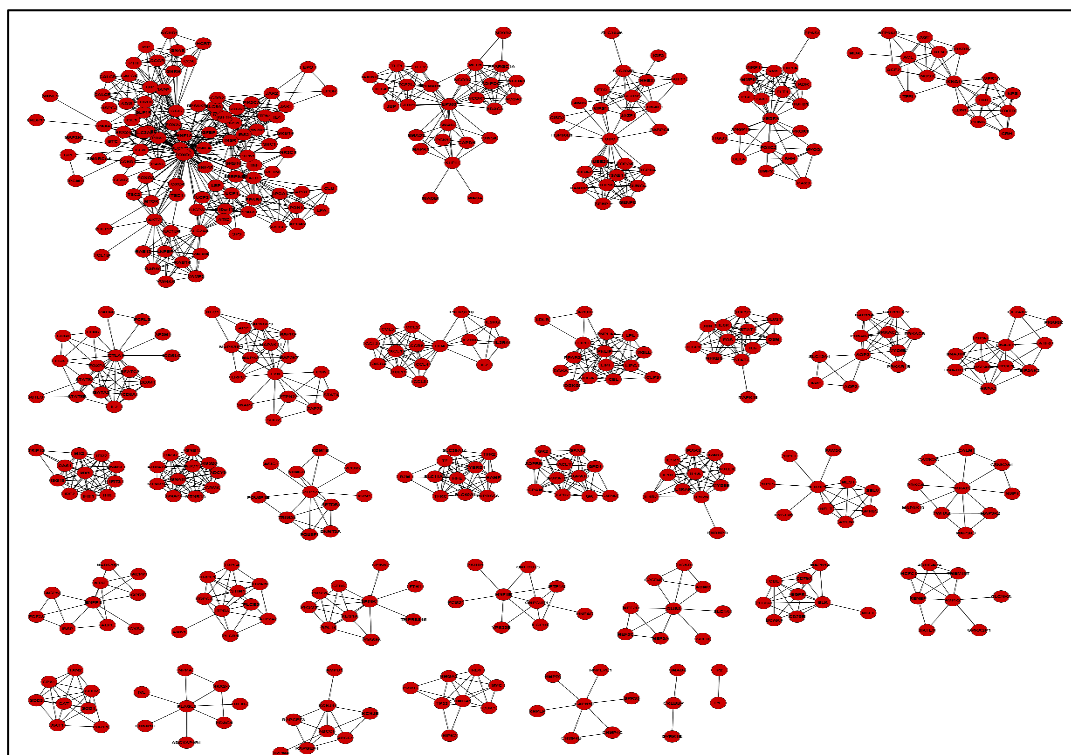


Figure 6. MCL clusters of original PPI network (3.1). 34 cluster modules are generated from PPI network with five major complexes consisting of >20 protein interaction.

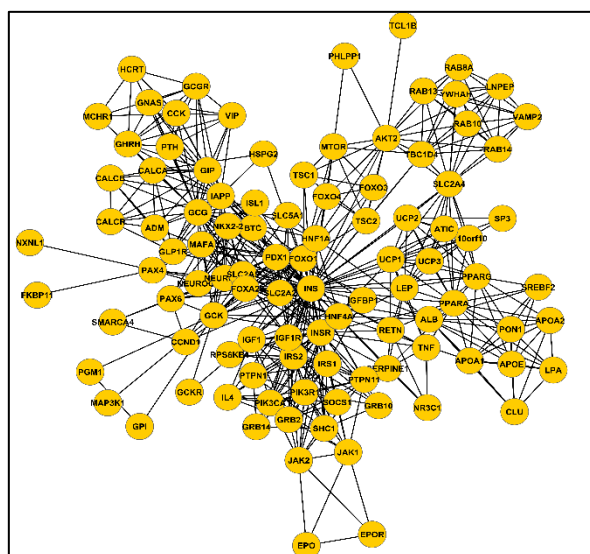


Figure 7. Major complex extracted from MCL clusters. It is the subnetwork with the highest number interaction (edge) and node (vertex) connectivity.

3.3. Overall quality of the detected complexes

We discovered that 78% of 2054 interactions (involving a total of 482 proteins) occurred between pairs from the original PPI network (Figure 1) at least shared a common function. Though the majority

of the interactions are between similar function protein pairs, there are many instances of interactions between proteins of different functions (overlapping modules). Therefore, it was reasonable to assume that interactions that forms the protein complex are between similar function protein pairs. In other words, it can be said that the quality of detected complex is good if it contains mostly similar function proteins. Thus, to check the quality of the detected complexes, we estimate the relative amount of interactions that are between similar function protein pairs out of intra-complex interactions for complexes using equation below (6).

$$RA = \frac{\sum_{i=0}^n SFI_i}{\sum_{i=0}^n AI_i} \quad (6)$$

where, n is the number of complexes of size ≥ 3 in a set, SFI_i is the number of interactions of cluster i , that are between protein pairs of identical functional class, and AI_i is the number of all interactions in cluster i of the corresponding set. Figure. 6 shows the relation between RA and d_{in} . There is a sudden rise from $d_{in} = 0.101$ to $d_{in} = 0.268$. Thus the complexes having density 0.101 or less have high statistical significance. When complexes are generated using $d_{in} = 0.268$, many of the complexes consist of more than three proteins and two interactions on the other hand when $d_{in} = 0.101$ is used, the aforementioned type of 3-protein complexes are excluded. Therefore, it may be concluded that many of the interactions contained in 3-protein complexes of density 0.268 are not interactions between similar function proteins pairs since the functions of all the proteins involved in the network or complexes are not yet known. It is noticeable that the percentage of interactions between similar function protein pairs is higher in high-density complexes and this percentage might increase if the functions of all the proteins were known (Figure 8). Hence it can be concluded that the interactions that form high-density clusters in PPI networks represent functional complexes and can be considered as true interactions with high chances.

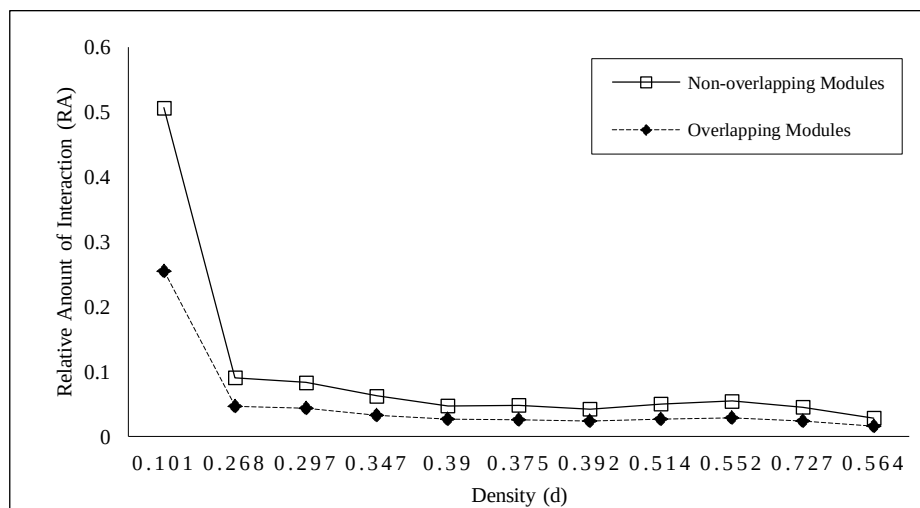


Figure 8. Relative amount of interactions involving non-overlapping and overlapping modules in MCL 34 clusters (protein complexes) against corresponding density values

3.4. Evaluation of MCL performance

To assess the robustness of MCL in the detection of protein complexes in the PPI network we further compared MCL performance with MCODE, SCPS, and AP using similar protein dataset (Figure 9). Thus the included the adjustable parameters for the three tested algorithms (MCODE, SCPS, and AP)

and evaluated across a wide spectrum of their settings. The performance of all algorithms initially assessed in terms of density, modularity, and clustering coefficient of identified clusters with respect to a number of detected clusters (protein complexes). MCL and SCPS algorithm generated high modularity (0.630 and 0.652 respectively) and cluster coefficient (0.751 and 0.731 respectively) with the both algorithm generating clusters with similar density (0.014). The three algorithms also tend to produce a much greater variation in the total number of clusters identified across their parameter settings, while still producing clusters of lower modularity; this is particularly striking for MCODE (0.304). Furthermore, cluster granule produced by both AP and MCODE were of low density implying that there was less interaction with protein in each of those complexes this explain why both MCODE and AP had low clustering coefficient (0.573 and 0.595 respectively).

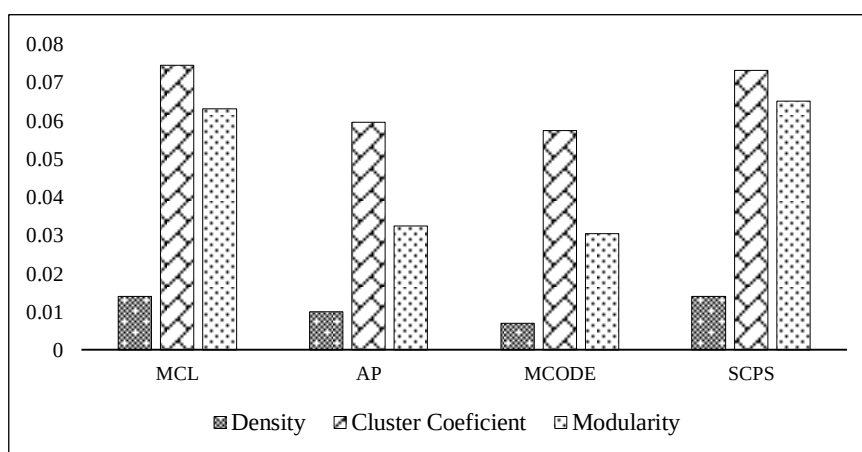


Figure 9. Comparison of clustering performance of MCL with AP, MCODE and SCPS algorithm based on cluster coefficients, modularity and cluster density.

We compared the performance of the proposed MCL with other three algorithms (MCODE, SCPS, and AP) in terms of the number of identified complexes (clusters), average cluster size and minimum cluster size (Table 1). From the analysis it was quite evidenced that MCL outperform MCODE, SCPS and AP algorithm, the MCL successfully detect 34 complexes (clusters) with the largest component comprise of 136 proteins with 711 interactions. Again, it is evidence that the MCL algorithm is able to archive strong cluster-wise separation with an average cluster size of 14.176. In addition, MCL present the highest inter-cluster range with 99 the algorithm also gives quite impressive modularity index (0.63) meaning that the identified components consist of highly interacting proteins. MCODE performs quite better than SCPS and AP by yielding 29 clusters with average cluster size being 11.103. Despite MCODE yielding better inter-cluster range 72, the algorithm record the lowest modularity index (0.304) may probably explain the weak cluster-wise separation when compared to MCL algorithm. On the other hand, SCPS algorithm only yields 26 clusters with highest an average cluster size of 18.538 and modularity index (0.64). Nevertheless, the AP performs the worst when subjected to the unweighted graph, in this analysis we had to set arbitrarily preference weight of -1.0 to facilitate the clustering. Although AP gave, the highest number of clusters (60) there was biases due to cluster over-representation. Therefore, it still holds that MCL outperform AP when clustering unweighted graphs.

Table 1. Evaluation of clustering performance by optimal parametrization.

	Parameter	Optimization values	Number Clusters	Average Cluster size	Maximum Size	Minimum Size
MCL	Inflation (r)	1.9	34	14.176	103	3
	Maximum iteration	100				
	Cut-off	1.0				
SCPS	Epsilon (ϵ)	-0.1	26	18.538	63	4
MCODE	Depth	100	29	11.103	75	3
	Haircut	TRUE				
	fluff	FALSE				
AP	Lambda (λ)	0.5	60	8.033	110	1
	Edge cut-off					
	Arbitrarily weight	1				
	preference	-0.1				

4. Discussion

Detection of protein complexes using existing clustering algorithm poses a great the challenges in the detection of significant biological interactions contained in the PPI networks. The MCL algorithm is a fast and scalable unsupervised clustering algorithm. It is one of the most widely used algorithms and it is based on simulating stochastic flows in networks. The MCL algorithm can detect cluster structures in graphs by taking advantage of a mathematical bootstrapping procedure. The process is trying to perform random walks through a graph and deterministically compute their probabilities to find the best paths. It does so by using stochastic Markov matrices. The algorithm works by alternating the inflation parameter, which iteratively calculate the set of transition probabilities. The inflation operator implements a stochastic matrix transformation to emphasize larger probabilities and deemphasize smaller ones [31]. Affinity propagation is an unsupervised algorithm and thus the number of clusters are automatically calculated. The idea behind this algorithm is to find sub-paths, which allow easy message exchanges between nodes. It takes as input a similarity matrix, which keeps the distances between all possible pairs of data points whereas it initially considers all data points as potential “exemplars”. In later steps, real-valued messages are exchanged between the nodes until a set of exemplars and corresponding clusters emerges with high quality. The main parameter of this algorithm is the ‘preference’, which controls how many data points are selected as exemplars [32]. Spectral clustering tries to detect clusters in a graph, where nodes are connected with highly-similarity. The algorithm also tries to find connections between such areas that should be weak, constituted by edges of low similarity. The aim is to identify highly connected clusters and, at a later stage, filter the inter edges within the cluster. The only parameter required is the user-defined number of clusters [33]. In the current study, we used MCL to partition PPI network in order to detect protein complexes. In MCL clustering occasionally there is overlap at several levels of cluster granularity this always corresponds network asymmetry for undirected graphs which are always characterized by overlapping modules. We realised global cluster modularity plays a significant role in the detection of major complexes in partitioned PPI network since by computing the global modularity score from a composite of the local modularity across all clusters we accounts for edges inside each cluster, edges connecting each cluster to the rest of the network, and the total number of edges in the network. When evaluated over a range of parameters, we realised the proposed MCL algorithm produce cluster granularity from PPI network with higher modularity than other methods (AP, MCODE, and SCPS) this is because MCL perform random walk within the PPI network taking into account the expansion and inflation this make it more efficient for detection of protein complexes than other methods, with the exception of a single

setting of epsilon value for SCPS, which result in higher modularity however with less number of identified clusters. On other hand AP and MCODE produced large number clusters however with low modularity and density due to less interaction within the clusters. Furthermore, AP yields a high number of complexes, the algorithm seems to be the most sensitive to the unweighted graph thus making it less reliable for detection of protein complexes. Therefore, the proposed MCL algorithm can be considered as the most stable one and performs best when partitioning both unweighted and weighted graphs.

5. Conclusion

In this study it's clearly evidenced that the proposed MCL algorithm is highly tuneable graph-clustering algorithm for protein complexes detection in dense and highly interconnected PPI networks. Performing flow simulation enable us obtained optimal inflation value for MCL enabling the algorithm to generate maximum cluster granularity. This facilitate detection proteins that share common biological functional within the clusters. Moreover, by computing cluster modularity and density we identified major protein complex with high protein interaction from partitioned the PPI networks. Further empirical comparison indicates MCL to be robust and reliable for detection of protein complexes as opposed to AP, SCPS and MCODE algorithm especially when we set optimal parameters of each algorithm although, we discovered under most conditions, AP and MCL outperform MCODE and SCPS. It still holds MCL is remarkably robust to variations in the choice of parameters, whereas the other algorithms require appropriate tuning in order to yield relevant results as we witness with AP one need to set arbitrary preference weight to facilitate clustering. This confirmed the general superiority of MCL over the three tested algorithms making suitable graph partitioning algorithm for detection of complex proteins in the PPI network.

References

- [1] Bader G, Hogue C 2003 *BMC Bioinf.* **4** 27
- [2] Schwikowski B, Uetz P and Fields S 2000 *Nat. Biotechnol.* **18** 1257-61
- [3] Ho Y, Gruhler A, Heilbut A, Bader G D, Moore L, Adams S L, Millar A, Taylor P, Bennett K, Boutilier K, et al 2002 *Nature* **415** 180-3
- [4] Peng J, Elias, Thoreen C, Licklider L and Gygi P 2002 *J Proteome Res.* **10** 1021
- [5] Uetz P, Giot L, Cagney G, Mansfield TA and Judson RS, et al 2000 *Nature* **403** 623-7.
- [6] Edachery J, Sen A and Brandenburg J 1999 *Lect Notes Comput Sc.* **1731** 98-106
- [7] Matula D 1972 *SIAM J. Appl. Math.* **22** 459-80
- [8] Arnau V, Mars S and Marin I 2005 *Bioinformatics* **21** 364-78
- [9] King A D, Pržuli N and Jurisica I 2004 *Bioinformatics* **20** 3013-20
- [10] Spirin V and Mirny L A 2003 *Proc. Natl. Acad. Sci. USA.* **100** 12123-8
- [11] Jeong H, Mason S P, Barabasi A L, Oltvai Z N 2001 *Nature* **411** 41-2
- [12] Oyama T, Kitano K, Satou K, and Ito T 2002 *Bioinformatics* **18** 705-14
- [13] Pereira-Leal J B, Enright A J and Ouzounis C A 2004 *Proteins.* **54** 49-57
- [14] Saito R, Suzuki H and Hayashizaki Y 2003 *Bioinformatics* **19** 756-63
- [15] Barabasi A L, Gulbahce N and Loscalzo J 2011 *Nat. Rev. Genet.* **12** 56-68
- [16] Cantini L, Caselle M, Fortunato S and Medico E 2015 *Sci Rep.* **5** 17386
- [17] Li W, Dai C, Liu C C and Zhou X J 2012 *J Comput Biol.* **19** 710-73
- [18] Schonbach C 2009 Molecular Biology of Protein-Protein Interactions for Computer Scientists *Biological Data Mining in Protein Interaction Networks* (Japan: Kyoto university) Chapter 1 pp 1-13
- [19] Ito T, Chiba T, Hattori M, Ozawa R, Sakaki Y and Yoshida M 2001 *Proc. Natl. Acad. Sci. USA.* **98** 4569-74
- [20] Li X, Foo C and Ng S 2007 *Comput Syst Bioinformatics Conf.* **6** 157-68
- [21] Watts DJ and Strogatz SH 1998 *Nature* **393** 440-2
- [22] Bader G and Hogue C 2003 *BMC Bioinformatics* **4** 2

- [23] Wuchty S and Almaas E 2005 *BMC Evol. Biol.* **5** 24.
- [24] Pereira-Leal J B, Enright A J and Ouzounis C A 2004 *Proteins Struct. Funct. Bioinf.* **54** 49–57
- [25] Brohee S and van Helden J 2006 *BMC Bioinf.* **7** 488
- [26] Enright A J, Van Dongen S, Ouzounis C A 2002 *Nucleic Acids Res* **30** 1575-84
- [27] KeshavaT S *et al.* 2009 *Nucleic Acids Res.* **37** 767–72
- [28] Van Dongen S 2000 Graph Clustering by Flow Simulation (Thesis) (Netherlands : University of Utrecht)
- [29] Yoon J, Blumer A and Lee K. 2006 *Bioinformatics* **22** 3106–8
- [30] Liu G M, Chua H N and Wong L 2009 *Bioinformatics.* **25** 1891–7
- [31] Frey B J and Dueck D 2007 *Science.* **315** 972-6
- [32] Tong A H *et al.* 2004 *Science.* **303** 808-13
- [33] Paccanaro A, Casbon J A and Saqi M A 2006 *Nucleic Acids Res.* **34** 1571-80

Acknowledgment

The authors express their gratitude to Bioinformatics research group, Computer Science department for registration of this paper and search related to this study.