Contents lists available at SCCE

Journal of Soft Computing in Civil Engineering

Journal homepage: www.jsoftcivil.com

Providing SSPCO Algorithm to Construct Static Protein-Protein Interaction (PPI) Networks

Elham Azarm[1], Tarlan Motamedi Nia[2], Rohollah Omidvar[3*]

1. M.Sc., Department of Computer Engineering, Shiraz Branch, Islamic Azad University, Shiraz Branch, Iran
2. M.Sc., Institute for Management and Planning Studies of Tehran, Iran
3. Ph.D., Department of Computer Engineering, Shiraz Branch, Islamic Azad University, Shiraz Branch, Iran

Corresponding author: *r.omidvar.uni@gmail.com*

https://doi.org/10.22115/SCCE.2022.301262.1355

ARTICLE INFO

Article history:
Received: 24 August 2021
Revised: 06 April 2022
Accepted: 08 April 2022

Keywords:
P-P interaction networks;
Dynamic networks;
SSPCO;
Graph partitioning;
Gene expression.

ABSTRACT

Protein-Protein Inter-action Networks are dynamic in reality; i.e. Inter-actions among different proteins may be ineffective in different circumstances and times. One of the most crucial parameters in the conversion of a static network into a temporal graph is the well-tuning of transformation threshold. In this part of the article, using additional data, like gene expression data in different times and circumstances and well-known protein complexes, it is tried to determine an appropriate threshold. To accomplish this task, we transform the problem into an optimization one and then we solve it using a meta-heuristic algorithm, named Particle Swarm Optimization (SSPCO). One of the most important parts in our work is the determination of interestingness function in the SSPCO. It is defined as a function of standard complexes and gene co-expression data. After producing a threshold per each gene, in the following section we will discuss how using these thresholds, active proteins are determined and then temporal graph is created. For final assessment of the produced graph quality, we use graph clustering algorithms and protein complexes determination algorithms. For accomplishing this task, we use MCL, Cluster One, MCODE algorithms. Due to high number of the obtained clusters, the obtained results, if they have some special conditions, will filter out or be merged with each other. Standard performance criteria like Recal, Precision, and F-measure are employed. There is a new proposed criterion named Smoothness. Our experimental results show that the graphs produced by the proposed method outperform the previous methods.

How to cite this article: Azarm E, Motamedi Nia T, Omidvar R. Providing SSPCO algorithm to construct static protein-protein interaction (PPI) networks. J Soft Comput Civ Eng 2022;6(2):68–91. https://doi.org/10.22115/scce.2022.301262.1355

1. Introduction

Biology has been a source of inspiration for development of computational approaches to solve different optimization problems. The early attempts in computational biological systems have focused on different biological networks, like protein-protein interaction (PPI). Combining and understanding of information available in biological systems need the development of novel information gathering algorithms and methods, and solving these complex computational problems needs biological and computational research [1]. In the recent years, the systems biology is changed to an important interdisciplinary research domain which tries to understand the time/ place behavior of cellular components. The field deals with all the physicochemical aspects of life. The modern tendency toward cross-disciplinary research and the unification of scientific knowledge and investigation from different fields has resulted in significant overlap of the field of biology with other scientific disciplines. Modern principles of other fields—chemistry, medicine, and physics, for example: are integrated with those of biology in areas such as biochemistry, biomedicine, and biophysics.

In mathematical and computational models of biology, models behavior is determined during modeling according to some parameters. Some of the parameters don't be able to calculate experimentally. As a result, it needs to use computational methods to estimate the parameters. One of the conventional methods is the transformation of parameters estimation to an optimization issue which we can use optimization algorithms in finding the best congruous parameters. Protein-Protein Inter-action (PPI) is determined according to the collected experiments and sub-networks in different laboratories which as a result, the collected PPI includes some interactions which happen at different time and place frame. It means, it s not important in the current PPI networks whether the interactions are happened simultaneously or they are unique or not. While, the place information can be managed in details by interpretation of sub-cell positioning [2]. Life in all levels is a huge and complex system. Life from biologic prospective is a macro-molecular continuation which created cell and conveys information. One of the successful methods in recent decades is using network modeling. It means by modeling, to focus on systems units, whether it is protein or human, and identify the relation between them. The nature of time networks modeling is to ignore all the information except the couple of related units together and the time of this relation [3].

Many systems can be modeled by time networks such as cells process, social interaction, Internet, mobile network, and environmental networks (food sources system). The purpose of this research is to use the targeted framework of SSPCO algorithm to determine the appropriate threshold for converting static networks with the most accuracy.

2. Research significance

Available data collection for PPIare static and lacks time and place parameters for Protein-Protein Inter-action. As a result, the dynamic information about protein and protein complex inter-actions are ignored. It means the available networks are static actually and for utilizing the dynamic properties, other additional information to create different networks should be used.

Gene expression in different situations and different steps of cell cycle can return the dynamic presence of one protein. Gene expression dynamic levels, identifies the protein presence dynamic, but doesn't give a guarantee expression for balance dynamic.

In other words, although two proteins may present simultaneously, there is no guarantee to surely interact at the same time. Because, one of the proteins may be inactive at this time and isn't able to do any activity and interact with others. If the gene expression is lower than the threshold, it lacks the protein presence. But, there are some important proteins which have a few expression levels, and then it isn't possible to identify an especial threshold for all proteins to determine the presence or absence or activity or inactivity of them [4].

Constructing dynamic PPI networks is done by using modeling protein activity and collecting co-regulated proteins in each time point. The method based on differential co-expression correlation is presented for activating protein-protein inter-action networks [5].

Studies show that high positive co-expression proteins tend to create static module which appears all the time and there are some high levels hubs at the center of each one, which are called party hub.

Furthermore, some of the low co-expression proteins are interacted in especial time points, as a result of physical interaction, which the hubs are called date hub [5].

Another method is based on gene expression level variance by determining the time point peak expression for each protein. So, if a protein is in its peak, it can interact with its active neighbor. This supposes scored gene expression activity to be calculated by using a fixed threshold or systematic threshold [4]. Therefore, assembling these two aspects is vital in constructing dynamic network. In this paper, combination of two methods is used for constructing dynamic PPI network.

[6] Is the first to use topology-based local network alignment for predicting protein interactions, and the first to apply SSPCO to the network alignment problem itself? The close proximity of the proteins in the discovered topologically-similar patterns made them more likely to be biologically related.

In one of research present an heuristic optimization method, particle swarm optimization, which predicts protein-protein interaction by using the domain assignments information. Results are compared with another known method which predicts domain interactions by casting the problem of interactions inference as a maximum satiability (MAX-SAT) problem [7].

In a paper presents a new metaheuristic optimization algorithm called Honey Badger Algorithm (HBA). The proposed algorithm is inspired from the intelligent foraging behavior of honey badger, to mathematically develop an efficient search strategy for solving optimization problems. The dynamic search behavior of honey badger with digging and honey finding approaches are formulated into exploration and exploitation phases in HBA. Moreover, with controlled randomization techniques, HBA maintains ample population diversity even towards the end of the search process [8].

A novel bio-inspired algorithm, namely, Dingo Optimization Algorithm (DOA), is proposed for solving optimization problems. The DOA mimics the social behavior of the Australian dingo dog. The algorithm is inspired by the hunting strategies of dingoes which are attacking by persecution, grouping tactics, and scavenging behavior. In order to increment the overall efficiency and performance of this method, three search strategies associated with four rules were formulated in the DOA. These strategies and rules provide a fine balance between intensification (exploitation) and diversification (exploration) over the search space [9].

In an article propose a modified PSO algorithm for unbiased global minima search by integrating with density functional theory which turns out to be superior to the other evolutionary methods such as simulated annealing, basin hopping and genetic algorithm. The present PSO code combines evolutionary algorithm with a variational optimization technique through interfacing of PSO with the Gaussian software, where the latter is used for single point energy calculation in each iteration step of PSO. Pure carbon and carbon containing systems have been of great interest for several decades due to their important role in the evolution of life as well as wide applications in various research fields [10].

In a paper, a new optimization algorithm called the search and rescue optimization algorithm (SAR) is proposed for solving single-objective continuous optimization problems. SAR is inspired by the explorations carried out by humans during search and rescue operations. The performance of SAR was evaluated on fifty-five optimization functions including a set of classic benchmark functions and a set of modern CEC 2013 benchmark functions from the literature. The obtained results were compared with twelve optimization algorithms including well-known optimization algorithms, recent variants of GA, DE, CMA-ES, and PSO, and recent meta-heuristic algorithms. The Wilcoxon signed-rank test was used for some of the comparisons, and the convergence behavior of SAR was investigated. The statistical results indicated SAR is highly competitive with the compared algorithms [11].

3. Methods

System analysis time protein complex doesn’t only improve the discover accuracy of complex proteins, but also strengthens our biological science about the process of formation of dynamic protein for organizing the cell. PPIs can be categorized according to their life time to static and transient PPI. On the other hand, transient PPIs are dependent on time and circumstances which provide a mechanism for quick reaction to external stimulations [12].

The analysis of time protein complexes can open a new dimension to dynamic gauge mechanism and improve our understanding of the diseases’ reasons. Although different time complexes occur in different time points, there are many protein complexes which formed static macro-molecules in order to perform an important biological function. Many static interactions which have a basic role for cell are appeared continuously in the preserved different time points and also their corresponding complexes in PPI static networks.

To protect cell’s suitability and stability and also to avoid undesirable disorder in cell’s basic function, these complexes should have mild and smooth changes during the time [13].

Meta-heuristic optimization methods like EA can find global optimization or close to global optimization in considerable time [14]. Researchers collected the applications of meta-heuristic algorithms such as Simulated Annealing, Genetic Algorithm, evolutionary programming, Deferential Evolutionary for estimating the models' parameters. To estimate the problem parameters in ERK signal paths, a multi-objective approach as a MOP Swarm based on Particle Swarm Optimization was used [15].

In article [16] bee colony optimization algorithm is used to predict proteins' structure which its results are compared to the results of simulated annealing method.

Rodrigues has used a meta-heuristic method to estimate parameters of static biological systems: also he used the combination of random and certain global optimization methods for decreasing time calculation [17].

In article [18], the combination of particle swarm algorithm and deferential evolutionary method was used to estimate biological non-linear parameters of signal paths. In article [19], to predict the protein structure of Lattice method in three-dimensional space, a method based on particle swarm algorithm was used. In this project, the particle swarm algorithm was used to identify different threshold.

SSPCO was first introduced by Omidvar et al in 2015. Due to its new mechanism, Algorithm SSPCO is designed in such a way that it can start from a scattered search in a problem space and reach good convergence around the optimal answer as soon as possible. The justification for this proper convergence is the population following the experience of a worthy population. In the reference paper, the degree and speed of convergence of the algorithm compared to other algorithms are well expressed. This is the justification for our use of this algorithm in this article. [20]. the main idea of this algorithm is taken from the behavior of the chicks of a type of bird called See-see partridge. These chicks when they feel insecure are located in a Purposeful queue at the reach a safe point and they start to move behind their mother. The biological reason for this movement is that each chick sets the criterion for its movement as a chick ahead of itself, which is one step ahead of it and has a better movement experience. According to Figure 1, each chick in the search space seeks to find a chick with the priority of a unit higher than itself and it tries to adjust its motion equation based on this chick.

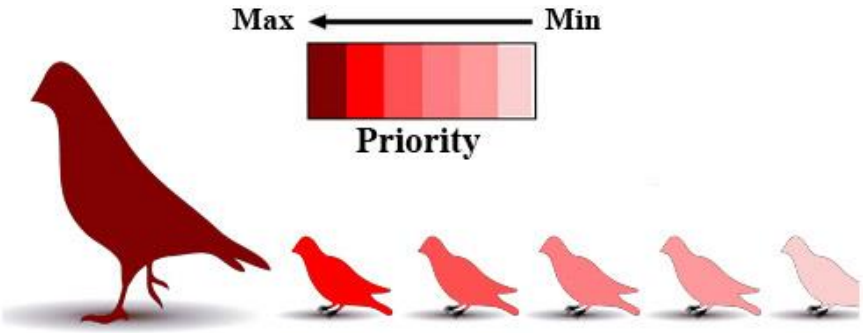

Fig. 1. Chicks motion in SSPCO algorithm.

In the algorithm, consider a variable for each particle entitled as priority variable. For particle i a priority variable defined. In every assessment, when a particle was better than the best personal experience or local optimum; a unit is added to the priority variable of that particle:

$$if \quad X_i.cost > P_{best} \rightarrow P_{best} = X_i.position \ and \quad X_i.priority = X_i.priority + 1 \tag{1}$$

$X_i.cost$ The cost of each particle in the benchmark, P_{best} is the best personal experience of each particle, and $X_i.position$ is the location of each particle. In every time of assessment, if the local optimum is better than the global optimum and vice versa, the particle's priority variable goes higher and a unit is added to it:

$$if \quad P_{best} > G_{best} \rightarrow G_{best} = P_{best} \ and \ X_i.priority = X_i.priority + 1 \tag{2}$$

G_{best} is the global optimum. The motion equation of each particle is set almost similar to the particle swarm algorithm in the form of equation 3:

$$X_i.position = X_i.position + X_i.velocity \tag{3}$$

$X_i.velocity$ is the velocity of each particle or chick. Then, Chickens sorted in array based on priority variable.

$$X_i.velocity = w * X_i.velocity + c * rand() * [position(X_{i+1}.priority)] - X_i.position \tag{4}$$

w is the coefficient impact of the previous velocity in the current velocity equation of particle, c is the coefficient impact of position of particle with upper priority in the current velocity equation of particle, $[position(X_{i+1}.priority)]$ is the location of the particle with one level higher priority than the current particle that the current particle tries to adjust its velocity according to the particle, $X_i.position$ is the current location of the particle.

Simulation dynamic behavior of nonlinear systems called chaos. It has raised enormous interest in different fields of sciences such as synchronization, chaos control, optimization theory, pattern recognition and so on. In optimization algorithms based on the chaos theory, the methods using chaotic variables instead of random variables are called chaotic optimization algorithm (COA). COA is a stochastic search methodology that differs from any of the existing swarm intelligence methods and evolutionary computation. COA can carry out overall searches at higher speeds than stochastic searches that depend on probabilities. There are several different chaotic sequences which the most commonly used such chaotic sequences are logistic maps that are considered in this paper. Logistic maps are frequently used chaotic behavior maps and chaotic sequences can be quickly generated and easily stored. For this reason, there is no need for storage of long sequences. In this study, we substitute the random parameters in PSO with sequences generated by the logistic map. The parameters random are modified by the logistic map based on the following equation:

$$Cr_{(t+1)} = k \times Cr_{(t)} \times \left(1 - Cr_{(t)}\right) \tag{5}$$

In Eq. (5), k =4 and for each independent run, Cr(0) is generated randomly, which Cr(0) not being equal to {0, 0.25, 0.5, 0.75, 1}..

Table 1
Pseudo code of Chaotic SSPCO algorithm.

1.//initialize all chicken by $k \times Cr_{(t)} \times (1 - Cr_{(t)})$
2.Initialize by $k \times Cr_{(t)} \times (1 - Cr_{(t)})$
3.Repeat
4. For each chicken i
5. //update the chicken's best position and priority
6. If $f(x_i) > f(pbest_i)$ then
7. $pbest_i = x_i$
8. $prioirity_i = prioirity_i + 1$
9. End if
10. //update the global best position and priority
11. If $f(pbest_i) > f(gbest)$ then
12. $gbest = pbest_i$
13. $prioirity_i = prioirity_i + 1$
14. End if
15. End for
16. //update chicken's velocity and position
17. For each chicken i
18. For each dimension d
19. $X_i.velocity = w * X_i.velocity + c * rand() * [position(X_{i+1}.priority)] - X_i.position$
20. $x_{i,d} = x_{i,d} + v_{i,d}$
21. End for
22. End for
23. //advance iteration
24. itetation = itetation + 1
25.Until it > MaxIterations

Suppose the population size is N. For particle i $(1 \le i \le N)$ in D-dimension space, current position is $x_i = x_{i2}, x_{i2},, x_{iD}$ and velocity is $v_i = v_{i2}, v_{i2},, v_{iD}$. During the optimization process, velocity and position of each particle at each step is updated by (6,7):

$$X_i.velocity = w * X_i.velocity + c * rand() * [position(X_{i+1}.priority)] - X_i.position \quad (6)$$

$$X_i.position = X_i.position + X_i velocity \quad (7)$$

Where, x_{ij} is component j of particle i. $c1$ and $c2$ are acceleration coefficients and is constriction factor which has a fixed value less than one. R is a random number with uniform distribution in [0; 1]. P_i is the best individual experience of particle i and $G_i$is the best experience of swarm. SSPCO is an iterative algorithm and all particles update their velocity and position in each

performance iteration. In each iteration, after all particle positions are updated, $P_i$value of all particles and also $G_i$value of swarm are updated with respect to new positions.

Table 2
Pseudo code of the SSPCO algorithm.

Algorithm1: Particle Swarm Optimization

For each Particle i ε [1 .. N]
initialize xi , vi
Pi = xi
End for
$G = \arg\min_{P_i} f(P_1)$
repeat:
For each Particle i ε [1 .. N]
update vi using equation (5)
Check the velocity boundaries.
update x_i using equation (1)
If $f(x_i) \leq f(P_i)$
Then $P_i=x_i$
If $f(P_i) < f(G)$
Then $G=P_i$
End for
until stopping criterion is met

System analysis of time protein complexes does not only improve the discover accuracy of proteins' complexes, but also reinforces our biological knowledge of formation process of static protein for organizing the cell. We can create a sequence of static networks by using the recognition of transient and stable interactions by data collection of proteins' interaction and gene expression data. In different time points, the stable interaction is reserved as a network spinal cord of proteins' interaction. While, the existence of transient interaction in especial time point is related to specific activities and requested function from two related proteins. Usually, to identify stable interactions, the simultaneous expression coefficient is used.

In the next chapter, there is a short introduction of calculation methods of the simultaneous expression

The extents of gene expression for different samples are in vector form and as a result, calculating the simultaneous expression among genes is like calculating different criteria for two numerical vectors. There are four common criteria for constructing the simultaneous expression gene networks; Pearson's correlation coefficient, Mutual information, Spearman's rank correlation coefficient, Euclidian distance.

Euclidian distance calculates the geometric distance between two vectors' gene expression from two aspects of direction and size.

Mutual information gives the size of gene expression uncertainty by knowing the decrease amount of another gene expression.

Pearson's correlation coefficient measures two vectors' tendency to decrease or increase together.

Spearman's rank correlation coefficient calculates Pearson's correlation for gene expression of two vectors' rank [21].

Pearson's correlation coefficient is sensitive to outlier data.

For this reason, we use another correlation criterion in the proposed method as bicor which is resistant toward outlier data [22].

Equation 2 shows bicor calculation.

$$
\begin{aligned}
bicor(x,y) &= \frac{\sum_{i=1}^{m}\left(x_i - med(x)\right).w_i^{(x)}\left(y_i - med(y)\right).w_i^{(y)}}{\sqrt{\sum_{j=1}^{m}\left[\left(x_j - med(x)\right).w_j^{(x)}\right]^2}\sqrt{\sum_{k=1}^{m}\left[\left(y_k - med(y)\right).w_k^{(y)}\right]^2}} \\
w_i^{(x)} &= (1-u_i^2)^2.I(1-|u_i|) \\
w_i^{(y)} &= (1-v_i^2)^2.I(1-|v_i|) \\
I(1-|u_i|) &= \begin{cases} 1 & if\ 1-|u_i|>0 \\ 0 & o.w \end{cases} \\
u_i &= \frac{x_i - med(x)}{9*mad(x)} \\
v_i &= \frac{y_i - med(y)}{9*mad(y)}
\end{aligned}
\tag{8}
$$

We suppose G as a PPI static graph and GE as a gene expression matrix propotion to G proteins.

Static protein's interactions which supposed to be appeared in all time points are extracted from G according to concept of the even simultaneous gene's expression. For each protein interaction in$G$, Pearson's correlation coefficient based on their gene expression profile during all GE time points is calculated. Then, proteins' interaction with PPC more than the amount of especial slice (δ) is as an interaction and reserved in all times. S is a $N \times N$ symmetric matrix to show stable interactions in PPI networks which $S_{ij} = 1$ demonstrates the existence of stable interaction, while $e_{ij} \in E \quad PCC(e_{ij}) > \delta$ and $S_{ij} = 0$ demonstrates inexistence of stable interaction [12].

For isolating stable and transient interactions from PPC calculation, the amount of gene expression of related proteins in each mane in all (ije) PPC time points is used. Physical interactions with PPC more than an especial threshold δ, is described as a stable interaction. To identify cutting threshold, the PPC amount is used for all physical interactions and PPC distribution with two parameters distribution, one is adapted for stable interactions and the other for transient interactions. To estimate the proposed combined method parameters of Guisin (GMM), EM algorithm is used.

The static part of DPPI networks for each time points $(1 \leq t \geq T)$ t, is constructed from G and GE graphs as a transient interactions and based on the proteins being active and simultaneous [4]. In (t) time point, a Protein is active when the amount of its gene expression is more than a defined threshold (denoted by$AT(i)$). The threshold is defined as below:

$$AT(i) = u(i) + 3\sigma(i)(1 - F(i)) \tag{9}$$

$u(i)$ shows the average expression gene of ith protein.

$$F(i) = 1 \Big/ (1+\sigma^2(i)) \tag{10}$$

$F(i)$ Is a function of weight which react the fluctuations of expression gene of ith protein. One single edge is appeared in static network only when its two involved proteins are in active mode. This above method is known as 3δ. Tang et al. [23] used the stable amount of 0.7 instead of the high threshold for constructing. To evaluate the proposed method and the results comparison, these two methods are performed. Then, a proposed method for constructing dynamic networks describes by using meta-heuristic algorithms.

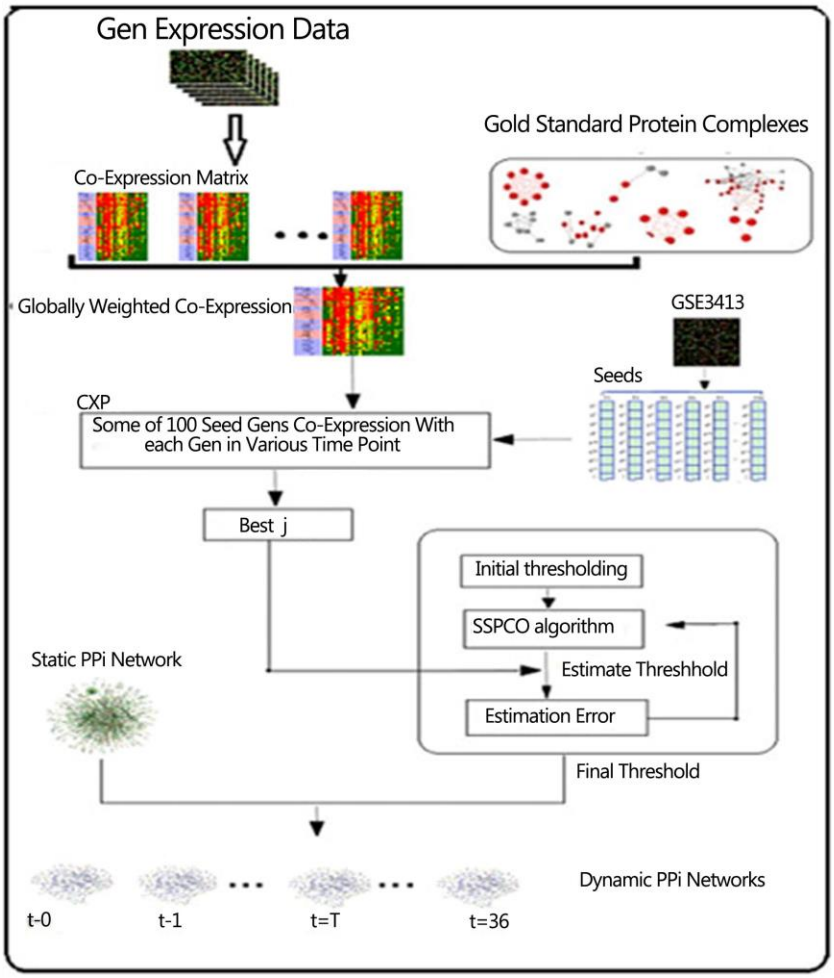

Fig. 2. Schematic chart of the proposed algorithm for constructing dynamic network.

One of the challenging subjects in constructing static networks of protein-protein interaction by using gene expression data is the identification of suitable threshold for defining active proteins in time points. As it explains in the last chapter, two current methods for activating PPI networks used 3δ and 0.7 relations. Experimental results at the end of the report shows that those two dynamic methods aren't appropriate enough. In this part, a method based on Particle Swarm

Optimization for identifying threshold is introduced**.** To increase the ability of the proposed method, different biological data for identifying the appropriate threshold are integrated.

The whole framework of the proposed method for constructing dynamic networks of protein-protein interaction shows in diagram (1). As you can see, there are independent inputs of this algorithm, stable networks of protein-protein interaction, different data of gene expression profile and the complexes of golden standard. Other algorithm's necessaries such as simultaneous expression matrixes are created during execution and according to inputs.

3.1. Constructing the matrix of simultaneous gene expression

At first, according to different gene expression data such as $gse3431$, $gse7645$, $gse4987$, …, some co-expression matrix (cox_i) are created by using $bicor$ criteria. Since, the $gp190$ framework is used in this project; each matrix of simultaneous expression has $9335 * 9335$ dimensions. Nine gene expression datasets are used for extraction of simultaneous expression; consequently the final result will include 9 matrices of size$9335 * 9335$. We called this general matrix, COX. We will use it in the next stages. For appropriate usage of these gene expression matrices in Particle Swarm Optimization algorithm, there is a need for assembling these matrixes in one united matrix.

For constructing unit co-expression matrix from different cox_i which is FCOX from available data in protein complexes of standards golden sets such as SGD, CYC2008, and MIPS are used. In fact, giving weight to each cox_i happens according to different proteins which come together in protein complexes. Two available proteins in a complex should have high co-expression coefficient, then each cox_i which has the higher amount of protein couple, gives more weight. This is considered certainly after calculating and collecting all proteins couple; All complexes whether complete or incomplete are employed. Finaly weighted sum of 9 matrices of simultaneous expression is calculated and saved in FCOX matrix.

To identify a function for evaluation and changing the problem into optimization problem, the points with high gene expression are chosen as seeds. Seeds are identified according to gse3431 amount in each time point. Actually, the seeds are contemplated as active proteins. The amount of 100 is supposed for seeds.

Sum of simultaneous expression complex of each protein is calculated by 100 protein centers in FCOX and one n*36 matrix saved as a CXP. Table (1) shows CXP matrix structure where n is number genes available in our datasets. Entries of this matrix are calculated for each gse3431 time points. Now, a threshold should be introduced for each protein. In identifying the threshold, available amount for each protein in CXP matrix and each protein expression in each time points are important and it will be affected according to its amount and the amount of gene expression at that time point on the identification of threshold.

Actually by employing the above levels, the problem is changed to the increased problem. i.e. the threshold should be identified in a way which by employing it, the whole amount on the corresponding line to each protein becomes maximum. If all CXPs are positive, we can normalize it or decrease the stable amount.

Table 3
$\boldsymbol{cxp_i^k}$:That is the complex of simultaneous expression of i[th] protein in k time point with 100 seed protein of k time point.

$\boldsymbol{CXP^1}$	$\boldsymbol{CXP^2}$		$\boldsymbol{CXP^k}$		$\boldsymbol{CXP^{35}}$	$\boldsymbol{CXP^{36}}$
cxp_1^1	cxp_1^2		cxp_1^k		cxp_1^{35}	cxp_1^{36}
cxp_2^1	cxp_2^2		cxp_2^k		cxp_2^{35}	cxp_2^{36}
cxp_3^1	cxp_3^2		cxp_3^k		cxp_3^{35}	cxp_3^{36}
cxp_i^1	cxp_i^2		cxp_i^k		cxp_i^{35}	cxp_i^{36}
cxp_{n-1}^1	cxp_{n-1}^2		cxp_{n-1}^k		cxp_{n-1}^{35}	cxp_{n-1}^{36}
cxp_n^1	cxp_n^2		cxp_n^k		cxp_n^{35}	cxp_n^{36}

3.2. Identification of disceret threshold for every protein

The thresholds presence of the proteins should identify in such a way that the proteins with high simultaneous expression by seeds protein has the high possibility of presents. As a result, the threshold tries to identify in such a way that the protein expressions with its seed proteins has high simultaneous expression. It should also inactivate proteins with negative or low simultaneous expression with their seed proteins. Therefore, we should use a related criterion to the set of positive simultaneous expression of proteins with seeds proteins for creating a threshold. This criterion will be based on the introduced best vector by 11 and 12 relations:

$$\underbrace{pos_j}_{j=1,2,\dots,|gpl90|} = \left\{ cxp_j^i \middle| cxp_j^i > 0 \quad , i = 1,2,\dots,|gse3431| \right\} \tag{11}$$

$$\underbrace{best_j}_{j=1,2,\dots,|gpl90|} = \sum_{i=1}^{|pos_j|} pos_j^i \tag{12}$$

In the above relations, cxp_j^is which have positive amount, means the possible j protein expression in i time point is high by its seed protein and the threshold of this protein (j) should identify in a way which this protein in time point which $cxp_j^i is$ positive is active and in the other points (negative points) are inactives. As a result, the set of positive simultaneous expression of protein with seeds protein in different time points are calculated and positioned in the best n number vector.

The threshold identification in this situation is changed to an optimization problem. The purpose of the increasing is the activation number of the proteins with positive simultaneous expression of protein with seeds protein. To evaluate this situation and relate it with activation of proteins with high activity, we used the set of positive simultaneous expression of protein with seeds protein criterion in time points instead of the number. The routine is that first, we start by random

threshold and identify the active proteins. To evaluate the wellness of this threshold, we calculate the set of simultaneous expression of activated proteins with 100 seed protein and compare to available amount in best vector.

If the identified threshold is ideal, the proteins with positive simultaneous expression with seed proteins pass and make a barrier for others and cause the similarity between the amount of set before and after threshold.

The ideal situation isn't possible in reality, because every amount for the threshold will cause the presence of proteins with the amount of negative simultaneous expression. So, our purpose is the identification of the threshold in a way that the amount of error among best vector and the gained vector from this threshold is decreased. As a result, we can use the optimization methods for calculating the amount of different threshold.

$$\underbrace{mask\_thresh_j^i}_{j=1,2,\dots,|gpl90|} = \begin{cases} 1, & gse3431_j^i < threshold_j \\ 0, & o.w \geq 0 \end{cases} \tag{13}$$

$$masked\_cxp_j^i = mask\_thresh_j^i * cxp_j^i \tag{14}$$

$$\underbrace{estim\_sum_j}_{j=1,2,\dots,|gpl90|} = \sum_{i=1}^{|gse3431|} masked\_cxp_j^i \tag{15}$$

If the threshold identifies well, the protein with negative CXP isn't mentioned and the proteins with positive amounts are mentioned.

For evaluation, we use the set of positive amounts in best and the set of whole mentioned proteins in masked-cxp. It is one of the available challenges in constructing dynamic networks to identify the threshold for every protein.

3.3. Utilizing the heuristic algorithm of SSPCO algorithm for threshold identification

In this part, by using the SSPCO algorithm the optimized threshold is extracted. The related settings of different parts are explained later. Higher and lower amounts of the threshold contemplate the equal high and low amount of gene expression in 36 time zone gse3431. In this part, available amounts in gse3431 data are used without normalization. One of the effective parameters in the result of algorithm is the number of chosen seeds which is the result and total suitability for different calculated seeds and the best amount for this variable are identified. (100 and 50) best amount of calculation according to the set of simultaneous expression of proteins with seeds protein is done in this part. One of the other changeable parameters in algorithm is the initial amount of the threshold for beginning the FA algorithm performance. As a result, the average gene expression is used in 36 time points as an initial amount.

$$Initial\_thresh(i) = mu(i) \tag{16}$$

3.4. Using cost function in SSPCO algorithm

Thirty-six masks, i.e. 36 temporal networks, are created using the initial thresholds. In each temporal network, some proteins will be active and the others will be inactive. In fact, our problem is converted into an optimization problem for classification. There are negative and positive classes in this problem. Their samples are weighted and the thresholds should be determined in a way that the overall weight of the classified samples is maximized as the positive class. Some positives can be considered as negative and some samples of the negative class may be considered as positive samples (TP, FP, FN, TN). Equation (11) shows the error calculation method with respect to the activation of different proteins using the obtained thresholds. Estim_sum denotes the whole samples classified as positive. In this version of the algorithm, a total of the positive and negative samples existing in the class is classified as positive.

$$error = \sum_{i=1}^{|gpl90|} \left|estim\_sum_j - best_j\right| \quad (17)$$

4. Results

4.1. Graph clustering by MCL

MCL: Markov Clustering [24] is a graph clustering method using flow simulation. Including two operators called expansion and inflation, this method acts using random walk simulation in a graph and definitely calculates the possibility of random walk in the sequence of similar subgraphs.

"Expansion" intensifies strong flows inside the strongly-connected areas, whereas "*Inflation*" removes weak connections in connected areas. These are performed repeatedly to partition a graph into some distinct clusters. Many researchers have proved that MCL is very resistant to noise. Flow-based approaches need complex procedures to simulate the stochastic behavior of a system.

4.2. Graph clustering by MCODE

MCODE: Bader *et al*. [25] presented a method for finding molecular complexes. In this method, each node is weighted by the density of local neighbors, and heavier nodes are selected as the core of initial clusters. Later, other nodes are added to these clusters. MCODE has two pre-processing steps including filtering non-dense clusters and creating overlapping clusters. This method never guarantees to find the subgraphs of necessarily close to each other. However, due to its polynomial time complexity, it is suitable for large-scale networks.

4.3. Graph clustering by cluster one

Cluster ONE has just been presented by Nepusz *et al*. to detect overlapping protein complexes in PPI networks. [26] Its major function is based on the development of overlapping neighbors. The algorithm consists of three steps. In the first step, high-cohesion groups are developed out of selected seed proteins. First, the protein with the largest number of connections (highest degree) is selected as a seed and a cohesive group develops through a greedy method. After completion

of the development of a group, the algorithm selects the following seed with the highest degree. Selection is performed among all proteins which are not in any of the protein complexes. The whole operation continues as long as there is no protein for investigation. In the second step of the algorithm, local optimal coherent groups with significant overlap are merged. In the third and final step, the candidates with fewer than three proteins and those with a density of lower than the given threshold (δ) are discarded. The density of a protein complex with N proteins is calculated by dividing the sum of its internal weights by (N* (N - 1) / 2).

With the temporal points rising, the number of sub-networks and the predicted clusters and protein complexes increases considerably. This becomes problematic while assessing and comparing it with the limited number of complexes, known as "the gold standard". Therefore, an algorithm is needed to reduce the number of protein complexes. Protein complexes can be reduced through combining different complexes and/or ignoring similar complexes [27].

The following procedure is used for combining/ignoring complexes:

All complexes are sorted based on their size. Then, any complex is compared with another one. If their similarities exceed the threshold, the smaller complex will be ignored and deleted. The complexity among complexes is calculated using the following equations:

$$sim(C_1, C_2) = {|C_1 \cap C_2|}\Big/{Max(|C_1|, |C_2|)} \tag{18}$$

$$sim(C_1, C_2) = {|C_1 \cap C_2|}\Big/{(|C_1| + |C_2|)} \tag{19}$$

Table (2) shows the impact of different thresholds of similarity criteria on the number of the known complexes and different assessment criteria.

Table 4
The result of the application of reduction strategy on the results.

Dm	cm	TOTAL	filtered(2,0.65)	filtered(2,0.5)	merg(2,0.5)
seven	mcl	6348	1703	1362	1409
seven	mcode	2537	1800	1590	1604
seven	clusterone	2652	1923	1457	1538
3s	mcl	4357	1203	976	992
3s	mcode	2585	1750	1476	1502
3s	clusterone	2672	1739	1292	1384
U_v10	mcl	9893	738	645	647
U_v10	mcode	15779	7813	4955	5494
U_v10	clusterone	14496	2274	1588	1925

5. Datasets

BioGRID: [6], This database is an integrated and continuously updating collection of physical and general interactions.

This collection consists of over 544,000 interactions and more than 27 different organisms. With over 544,000 non-repetitive interactions of yeast, it is the largest PPI collection for this organism.

GEO: [28], The gene expression levels are measured in different temporal points and are stored in some datasets. GEO database stores these series of gene expression under certain platforms (GPLxxx) for various samples (GSMxxx) under unique names (GSExxx). For instance, series GSE3431 of platform GPL90 was measured at 12 temporal points for the temporal distances of 25 minutes and 6777 gene expression levels were measured.

The PPI data used in this project were obtained from *BioGRID* datasets. The gene expression and co-expression data were also extracted from GEO.

Table 5
A selected list of gene expression series**.**

Name of series	Number of series
Gse26169	210
Gse25582	151
Gse18121	42
Gse15254	72
Gse11452	170
Gse9482	40
Gse7645	48
Gse3431	36
Gse3076	96

Quality evaluating of the SPINs made using different method is among the major issues. Comparison of the predictions using SPIN and the known biological knowledge has a limited assessment capability. On the one hand, the topological features of SPINs should be calculated and the scale of SPINs sub-networks should be absolute according to the studies on gene expression and on the other hand, the biological interpretation of SPINs during their quality assessment encompasses large areas. At any temporal point or under any condition, the proteins and the interaction among them have not been selected randomly and they are involved in certain biological processes. Therefore, the whole sub-network may be under the influence of certain functions. Intuitively, each sub-network can be considered as a whole structure and its biological function is analyzed.

Selection of biological knowledge display to check compatibility and stability of different SPINs is a convenient method for measuring the quality of SPINs. A quality SPIN may be helpful for discovering proteins and detecting interactions with high reliability.

SPINs provide all temporal, spatial, and qualitative data. Therefore, protein complexes and the biomarkers with changes including dynamic features can be used for detecting protein performance modules.

Functional interactions of proteins in dynamic networks are revealed more accurately than static networks. The better the manufacturing method of a dynamic network, the better results will be provided for analyzing functional modules. Comparison of the results of clustering conventional algorithms in determining modules in any created dynamic network is used for evaluating the creation method of dynamic networks. Some papers used the protein complexes created in CYC2008 datasets as a gold standard set for evaluation of results.

It is expected for any module detection method, predefined cluster (p_c) and reference complex (R_c) to be compatible as much as possible. To determine the adapted complexes, the overlapping score is used as the following equation (14):

$$OL(P_c, R_c) = \frac{|V_{P_C} \cap V_{R_C}|^2}{|V_{R_C}| * |V_{P_C}|} \tag{20}$$

In the above equation, $|VP_C|$ indicates the size of the predefined cluster, $|V\ R_C|$ is the size of the recognized complex and $|VP_C \cap VR_C|$ is number of the common items of the predefined cluster and the recognized complex. If the overlapping score of OL is greater than a threshold (sigma), $P_C$and R_C comply with each other.

Conformity of complex with the set of standard complexes is used for assessing the quality of the generated complexes. “Precision” and “recall” are the common criteria for assessing the performance of the methods to predict protein complexes.

The precision of a fraction of the predicted complexes is exactly in proportion to all the discovered complexes, whereas recall (sensitivity) of a fraction of the discovered standard complex is in proportion to all the standard complexes.

TP is the number of the correctly predicted complexes adapted to OS more than one value of threshold T and FP is the total number of the predicted complexes subtracted by TP, while TN is the number of the standard complexes predicted by OS more than a threshold value of T. FN is the number of the unpredicted standard complexes of T.

T is the predefined threshold and it is usually determined 0.2. F-measure measurement criterion, the harmonic mean of precision and sensitivity, is another criterion for assessing the performance of a method.

The equations for assessment criteria are as follows:

$$Precision = {TP}/{(TP + FP)} \tag{21}$$

$$Recall = {TP}/{(TP + FN)} \tag{22}$$

$$F - measure = \frac{2 * Precision * Recall}{(Precision + Recall)} \tag{23}$$

Tables (4), (5), (6), and (7) show the assessment results in different modes and with various algorithms.

The assessment criteria for the dynamization methods of protein-protein interaction networks were calculated in 4 different modes using clustering algorithms.

In the first mode, the complexes obtained from each clustering method are filtered using clusters reduction mechanism and the complexes with lower than 3 proteins and/or similar complexes are deleted. In this mode, CYC08 standard set is also filtered by the above-mentioned method. Figure 4 shows the results obtained in this mode.

Table 6
The results obtained using filtered similar clusters and filtered CYC08.

Dm	Cm	recall	Precision	f_measure
0.7	Mcl	0.292373	0.054022	0.091194
0.7	Mcode	0.309322	0.06	0.100505
0.7	clusterone	0.495763	0.092564	0.156001
3sigma	Mcl	0.186441	0.0532	0.08278
3sigma	Mcode	0.317797	0.076	0.122665
3sigma	clusterone	0.411017	0.097182	0.157196
FA_thr	Mcl	0.237288	0.074526	0.113427
FA_thr	Mcode	0.504237	0.047997	0.087651
FA_thr	clusterone	0.622881	0.119613	0.200688

In the second mode, the complexes obtained using any clustering method are filtered using clusters reduction mechanism and the complexes with lower than 3 proteins are deleted and similar complexes are merged. In this mode, CYC08 standard set is used without change and in a complete manner. Figure 5 shows the results in this mode.

Table 7
The results obtained using similar merged clusters and unfiltered CYCO8.

Dm	Cm	recall	precision	f_measure
0.7	Mcl	0.201	0.067424	0.1009734
0.7	Mcode	0.228	0.073566	0.1112327
0.7	clusterone	0.328	0.105982	0.1602518
3sigma	Mcl	0.11	0.064516	0.0814111
3sigma	Mcode	0.194	0.073901	0.1069742
3sigma	clusterone	0.282	0.114884	0.1632356
FA_thr	Mcl	0.186	0.103555	0.1331103
FA_thr	Mcode	0.331	0.055333	0.094811
FA_thr	clusterone	0.368	0.137662	0.2003175

In the third mode, the complexes obtained from any clustering method and the CYC08 standard set is assessed without any changes. Table 6 shows the results in this mode.

Table 8
Results obtained from clusters and unfiltered CYC08.

dm	Cm	recall	Precision	f_measure
0.7	Mcl	0.607843	0.135676	0.221836
0.7	Mcode	0.333333	0.107213	0.162243
0.7	Clusterone	0.39951	0.150075	0.218189
3sigma	Mcl	0.367647	0.122332	0.183579
3sigma	Mcode	0.286765	0.114507	0.163662
3sigma	Clusterone	0.328431	0.172156	0.2259
FA_thr	Mcl	0.681373	0.126049	0.212742
FA_thr	Mcode	0.558824	0.092655	0.158954
FA_thr	Clusterone	0.556373	0.215853	0.311035

In the fourth mode, the complexes obtained from any clustering method are filtered using clusters reduction mechanism and the complexes with lower than 3 proteins and/or the similar complexes are deleted. However, the CYC08 standard set is used without any change and in a complete manner. Table 7 shows the results in this mode.

Table 9
The results obtained by filtering similar clusters and unfiltered CYC08**.**

dm	cm	Recall	Precision	f_measure
0.7	mcl	0.218137	0.066353	0.101754936
0.7	mcode	0.242647	0.075556	0.115230564
0.7	clusterone	0.387255	0.119085	0.18215503
3sigma	mcl	0.127451	0.05985	0.081451508
3sigma	mcode	0.213235	0.082857	0.119341564
3sigma	clusterone	0.323529	0.121334	0.17648177
FA_thr	mcl	0.181373	0.093496	0.12338698
FA_thr	mcode	0.431373	0.060028	0.105390567
FA_thr	clusterone	0.504902	0.157432	0.240022913

The increase of temporal points raises sub-networks and increases predicted clusters and protein complexes considerably. This will be problematic while assessing and comparing with some limited known gold standard complexes.

Figure 3 shows the comparative diagram of the results obtained using MCODE, Cluster One, and MCL clustering algorithms, and different methods. As stated earlier, the proposed method was assessed using three algorithms and it was compared with two basic methods of protein-protein interaction networks. As Figure 3 and the relevant tables show, the results of the proposed method outperformed the earlier methods with respect to the recall criteria in three clustering algorithms and in all comparisons. Precision criterion and following that F-measure criterion depend on the number of the clusters obtained by each algorithm. Therefore, if there are many algorithm clusters, the precision criterion may diminish. As the amount of protein in temporal

networks created by the proposed method exceeds that of ones made with previous methods, the number of clusters obtained will be consequently more. Nevertheless, the proposed method in Cluster One and MCL algorithms has a higher precision and F-measure compared to the basic methods of 0.7 and 3sigma.

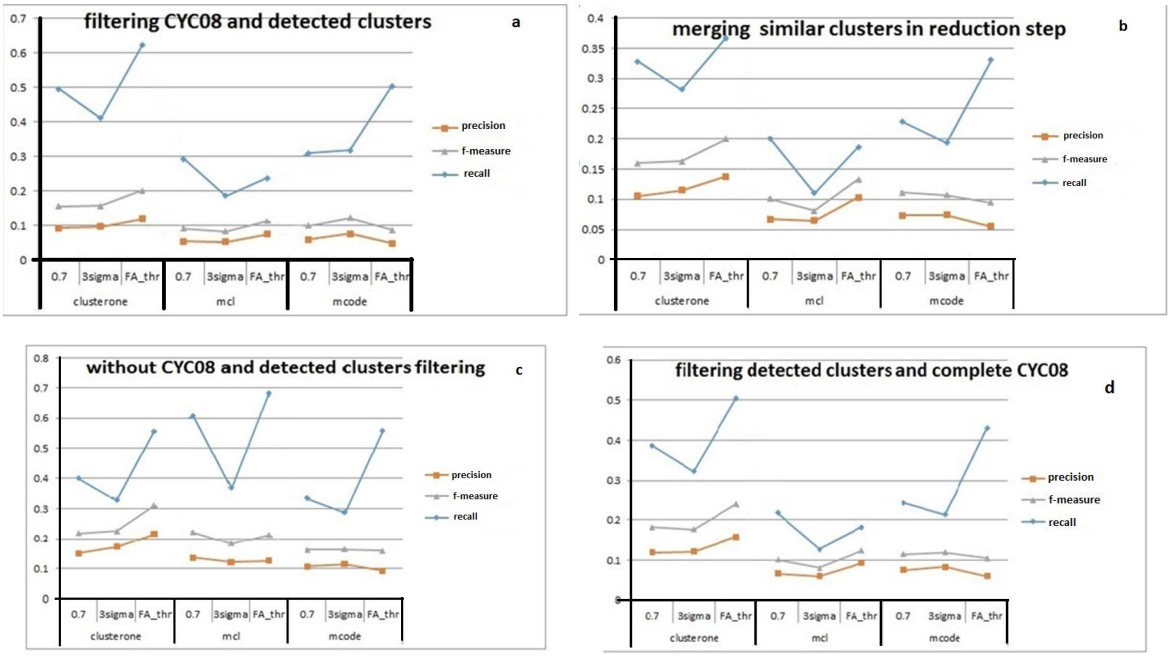

Fig. 3. The comparative diagram for the results obtained using MCODE, ClusterOne, and MCL clustering algorithms and different methods (A): The diagram is related to the data of Table (4) in which the identified similar clusters and CYC08 were filtered and deleted at clusters reduction step (B): The diagram related to the data of Table (5) in which similar clusters were merged at clusters reduction step, but similar clusters of CYC08 were used without filter, (C) The diagram related to the data of Table (6) in which clusters reduction step was not applied, (D) The diagram related to the data of Table (7) in which similar clusters were deleted at clusters reduction step, but the clusters similar to CYC08 were used without filter.

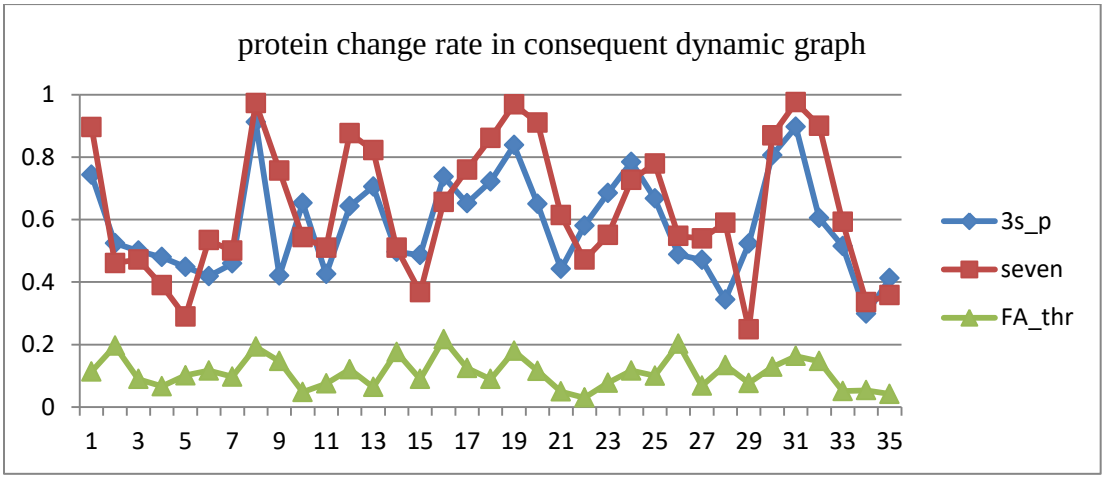

Fig. 4. Variations in proteins in a sequence of time points (degree of variations at the time point t in comparison with t-1).

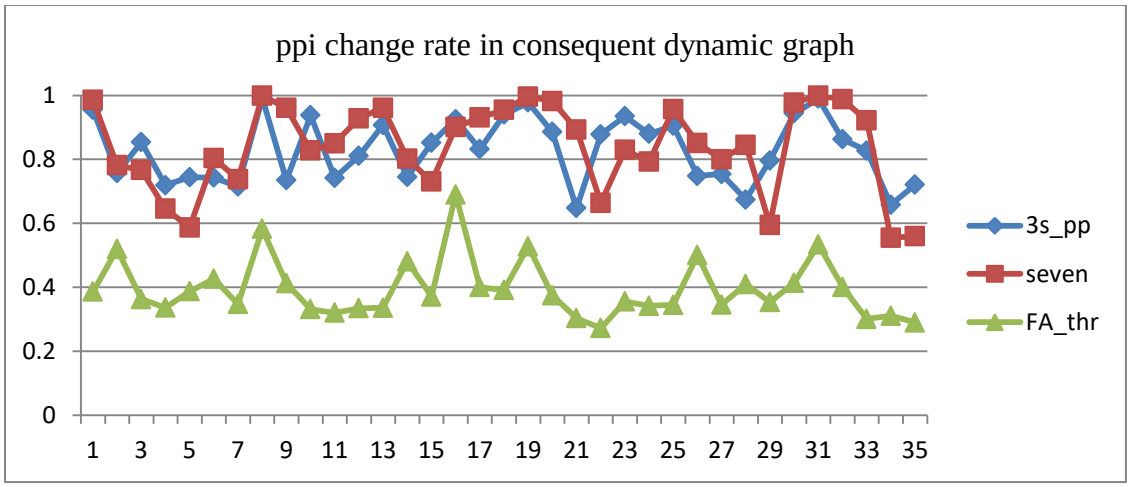

Fig. 5. Variations in protein interactions in a sequence of time points (degree of variations at time t in comparison with time t-1).

The changes of dynamic networks were measured as compared with the temporal points before them as far as the changes of protein presence and the interaction among them are concerned. Figures 4 and 5 show the results.

As Figure 4 shows, the rate of changes in 0.7 and 3sigma methods is very high and it even approaches one at some temporal points. For instance, the rates of changes of temporal point 8 as compared with temporal point 7 in 0.7, 3sigma, and the proposed method are 0.972403, 0.91182, and 0.192931, respectively.

The rate of change of 0.972403 indicates that approximately all the active proteins in a temporal point are new and no protein remained from the earlier step; it means fundamental changes in a cell, which is inconsistent with the concepts and biology. For maintaining the fitness and stability of the cell and avoiding unfavorable disorder in its basic performance, the complexes should have smooth and mild changes over time. The results of the proposed method exhibit smooth and mild changes. Figure 4 shows the rate of changes of the interactions among active proteins. It proves that the changes of the proposed method are smoother than the ones of the earlier methods.

One of the positive and interesting points of the proposed method is that the determination of stable interactions during different temporal points is implicit. That is, this value was set to zero or a value close to zero for some proteins while determining the threshold value. This way, all genes expression in temporal points exceeded this value and the proteins were always active. The interaction between them is specified as stable interactions.

We have to correct the results of the proposed method in order to ensure that a test is carried out on the results. Friedman test, a nonparametric test, is an analysis of variance with repeated measures and is equivalent to that of the comparison between the K variables used in the average rating. The test status variables are assessed in several related cases. More information about Friedman's test is available [29,30]. We have to consider the validity of the results of the proposed algorithm. For test in 4 different iterations of the proposed algorithm that is specified in

various iterations is similar. The main samples taken from Friedman test show this on results. The final answer of this test is 0.502, because it is more indicative of the value of 0.50; this is the natural course that answers the same level and between different repetitions compliance on each of the result, and the results are reliable.

Table 10
Friedman test for 4 times.

Descriptive Statistics								
	N	recall	Precision	Minimum	Maximum	Percentiles		
						25th	50th(Median)	75th
Var00001	13	10.7154	5.41148	3.31	16.66	4.4850	13.5600	15.6400
Var00002	13	10.7646	5.67621	3.35	19.33	4.8200	12.0500	15.7800
Var00003	13	11.6962	7.56971	3.09	27.22	4.3750	9.5500	17.5800
Var00004	13	10.9092	5.36789	4.00	18.51	5.5200	13.2900	15.2100

Rank		Test Statistics	
Var00001	2.08	N	13
Var00002	2.54	Chi-Square	2.354
Var00003	2.54	df	3
Var00004	2.85	Asymp. Sig.	.502

6. Conclusions

This paper discussed determination of appropriate thresholds to convert static networks into dynamic ones as one of the challenges in systemic biology and provided a new method for threshold determination. Determination of a unique threshold for any gene is one of the important points in the proposed method; the whole thresholds do not use a fixed formula and equation for all genes. Meanwhile, merging particle swarm optimization Meta-heuristic algorithm using genes co-expression concepts and gold standard datasets are among the other prominent points of this project.

Appropriate thresholds are determined for dynamization of static networks and stable interactions in all temporal points are achieved implicitly using the available additional data such as gene expression in different periods and conditions and the series of gold standard protein complexes. Dynamic proteins are specified and temporal graphs are created for making dynamic networks using the threshold created specifically for any gene.

The MCL, Cluster ONE, and MCODE graph clustering algorithms were used for the final assessment of the performance of the created graphs. The set of CYC2008 gold standard complexes was used for the final assessment. The standard assessment criteria of recall, precision, f-measure and a new criterion called “smoothness” were calculated. The experimental results on BioGRID data show that the results of the graphs created by the innovative method outperformed the earlier methods. For future research, we can focus on the time component and change the mechanism of the article to improve time.

References

[1] Trewavas A. A Brief History of Systems Biology. Plant Cell 2006;18:2420–30. https://doi.org/10.1105/tpc.106.042267.

[2] Chen B, Fan W, Liu J, Wu F-X. Identifying protein complexes and functional modules--from static PPI networks to dynamic PPI networks. Brief Bioinform 2014;15:177–94. https://doi.org/10.1093/bib/bbt039.

[3] Holme P, Saramäki J. Temporal Networks as a Modeling Framework, 2013, p. 1–14. https://doi.org/10.1007/978-3-642-36461-7_1.

[4] Wang J, Peng X, Li M, Luo Y, Pan Y. Active Protein Interaction Network and Its Application on Protein Complex Detection. 2011 IEEE Int. Conf. Bioinforma. Biomed., IEEE; 2011, p. 37–42. https://doi.org/10.1109/BIBM.2011.45.

[5] Zhang Y, Du N, Li K, Feng J, Jia K, Zhang A. Critical protein detection in dynamic PPI networks with multi-source integrated deep belief nets. 2013 IEEE Int. Conf. Bioinforma. Biomed., IEEE; 2013, p. 29–36. https://doi.org/10.1109/BIBM.2013.6732606.

[6] Ahmed H, Glasgow J. 3PI: A Novel Method for Predicting High-Confidence Protein-Protein Interactions using Particle Swarm-based Network Alignment, 2013. https://doi.org/10.13140/2.1.2630.9764.

[7] Iqbal M, Freitas AA, Johnson CG. Protein Interaction Inference Using Particle Swarm Optimization Algorithm. Evol. Comput. Mach. Learn. Data Min. Bioinforma., Berlin, Heidelberg: Springer Berlin Heidelberg; n.d., p. 61–70. https://doi.org/10.1007/978-3-540-78757-0_6.

[8] Hashim FA, Houssein EH, Hussain K, Mabrouk MS, Al-Atabany W. Honey Badger Algorithm: New metaheuristic algorithm for solving optimization problems. Math Comput Simul 2022;192:84–110. https://doi.org/10.1016/j.matcom.2021.08.013.

[9] Peraza-Vázquez H, Peña-Delgado AF, Echavarría-Castillo G, Morales-Cepeda AB, Velasco-Álvarez J, Ruiz-Perez F. A Bio-Inspired Method for Engineering Design Optimization Inspired by Dingoes Hunting Strategies. Math Probl Eng 2021;2021:1–19. https://doi.org/10.1155/2021/9107547.

[10] Jana G, Mitra A, Pan S, Sural S, Chattaraj PK. Modified Particle Swarm Optimization Algorithms for the Generation of Stable Structures of Carbon Clusters, Cn (n = 3–6, 10). Front Chem 2019;7. https://doi.org/10.3389/fchem.2019.00485.

[11] Shabani A, Asgarian B, Gharebaghi SA, Salido MA, Giret A. A New Optimization Algorithm Based on Search and Rescue Operations. Math Probl Eng 2019;2019:1–23. https://doi.org/10.1155/2019/2482543.

[12] Ou-Yang L, Dai D-Q, Li X-L, Wu M, Zhang X-F, Yang P. Detecting temporal protein complexes from dynamic protein-protein interaction networks. BMC Bioinformatics 2014;15:335. https://doi.org/10.1186/1471-2105-15-335.

[13] Byrum S, Smart SK, Larson S, Tackett AJ. Analysis of Stable and Transient Protein–Protein Interactions, 2012, p. 143–52. https://doi.org/10.1007/978-1-61779-477-3_10.

[14] Jianyong Sun, Garibaldi JM, Hodgman C. Parameter Estimation Using Metaheuristics in Systems Biology: A Comprehensive Review. IEEE/ACM Trans Comput Biol Bioinforma 2012;9:185–202. https://doi.org/10.1109/TCBB.2011.63.

[15] X. G. Systems biology approaches to the computational modelling of trypanothione metabolism in Trypanosoma brucei. thesis, University of Glasgow. 2010.

[16] Fonseca R, Paluszewski M, Winter P. Protein Structure Prediction Using Bee Colony Optimization Metaheuristic. J Math Model Algorithms 2010;9:181–94. https://doi.org/10.1007/s10852-010-9125-1.

[17] Rodriguez-Fernandez M, Egea JA, Banga JR. Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. BMC Bioinformatics 2006;7:483. https://doi.org/10.1186/1471-2105-7-483.

[18] Abdullah A, Deris S, Anwar S, Arjunan SN V. An Evolutionary Firefly Algorithm for the Estimation of Nonlinear Biological Model Parameters. PLoS One 2013;8:e56310. https://doi.org/10.1371/journal.pone.0056310.

[19] Maher B, Albrecht A, Loomes M, Yang X-S, Steinhöfel K. A Firefly-Inspired Method for Protein Structure Prediction in Lattice Models. Biomolecules 2014;4:56–75. https://doi.org/10.3390/biom4010056.

[20] Omidvar R, Parvin H, Rad F. SSPCO Optimization Algorithm (See-See Partridge Chicks Optimization). 2015 Fourteenth Mex. Int. Conf. Artif. Intell., IEEE; 2015, p. 101–6. https://doi.org/10.1109/MICAI.2015.22.

[21] Dehmer M, Emmert-Streib F, Graber A, Salvador A, editors. Applied Statistics for Network Biology. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2011. https://doi.org/10.1002/9783527638079.

[22] Song L, Langfelder P, Horvath S. Comparison of co-expression measures: mutual information, correlation, and model based indices. BMC Bioinformatics 2012;13:328. https://doi.org/10.1186/1471-2105-13-328.

[23] Tang X, Wang J, Liu B, Li M, Chen G, Pan Y. A comparison of the functional modules identified from time course and static PPI network data. BMC Bioinformatics 2011;12:339. https://doi.org/10.1186/1471-2105-12-339.

[24] V D, S. M. Graph Clustering by Flow Simulation. PhD thesis, Center for Math. and Computer Science (CWI). 2000.

[25] Bader GD, Hogue CW V. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003;4:1–27.

[26] Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. Nat Methods 2012;9:471–2. https://doi.org/10.1038/nmeth.1938.

[27] Wang J, Peng X, Li M, Pan Y. Construction and application of dynamic protein interaction network based on time course gene expression data. Proteomics 2013;13:301–12. https://doi.org/10.1002/pmic.201200277.

[28] http://www.ncbi.nlm.nih.gov/geo/, last accessed: 10/29/2014 n.d.

[29] García S, Molina D, Lozano M, Herrera F. A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 Special Session on Real Parameter Optimization. J Heuristics 2009;15:617–44. https://doi.org/10.1007/s10732-008-9080-4.

[30] Villegas JG. Using nonparametric test to compare the performance of metaheuristics. EU/ME-The Metaheuristics Community, Internet 2011.