ORIGINAL ARTICLE

Nature inspired link prediction and community detection algorithms for social networks: a survey

Srilatha Pulipati[1] · Ramasubbareddy Somula[2] · **Balakesava Reddy Parvathala[2]**

Abstract Social network analysis (SNA) has become a prominent research area in recent times. The popularity increased due to the rich information these networks possess. SNA is a domain of data analytics that practices graph theory to understand the social structures.Understanding and Analyzing the present links in the social network to predict the future possible links from the existing.The Network forms the interesting link problem and finding the similar groups in networks which is considered as the Community detection problem.The wide variety of applications of link prediction and community detection problems are recommendations.suggesting friends to users, predict the criminal association,inferences in biology networks and to analyze the trends especially in marketing. In various fields of science, the complexity of optimization problems increases as technology progresses. To find an optimum solution to a problem, there exists several approaches. In this regard swarm optimization techniques are more prominent. In SNA use of swarm optimization techniques has been employed in many aspects. Among these, the community detection and link prediction in SNA, is a key problem. Since with the growing network to find the similarity between the nodes in the network is a time consuming process to optimize the process many researchers using nature inspired algorithms to solve the link prediction and community detection problems. Apart from these problems nature inspired algorithms are widely used in many fields to solve constraint based optimization problems. In this paper, a review on swarm optimization techniques namely Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Artificial Bee Colony Optimization (ABC) and Firefly Algorithm (FA) along with its application in community detection have made in detail. A qualitative comparison is made among these methods to analyze and infer the nature of parameters involved in approaching an optimal solution.

Keywords Bio inspired computing · Social networks · Link prediction · Community detection

1 Introduction

In recent years, understanding, managing and analysing large or complex networks or social networks, has received significant interest from researchers in various domains and social network analysis (SNA) is considered as one of the most significant challenges that has to be addressed. Because of the fast advancement in Internet, several types of social networks like Facebook, Twitter and LinkedIn have emerged as online social networks with significantly large network size (due to the large number of users getting added on a daily basis). There are many features that can be interpreted by analyzing these social networks. Network analysis provides an effective tool to investigate statistical properties of the network such as small world

✉ Ramasubbareddy Somula
svramasubbareddy1219@gmail.com

Srilatha Pulipati
sreelatha.pulipati@gmail.com

Balakesava Reddy Parvathala
balakesavareddyp@gmail.com

[1] Department of Computer Science and Engineering, Anurag University, Hyderabad, Telangana 500 088, India

[2] Department of Information Technology, VNRVJIET, Hyderabad 500090, India

phenomenon (Watts and Strogatz 1998; Gandomi et al. 2013b; Yang et al. 2013) and scale free property (Wang and Chen 2003).

In Social Network Analysis (SNA), discovering communities has great significance. The social networks are usually denoted in graphs. In a social network, the community refers to a group of nodes that are tightly coupled to each other and loosely connected with other nodes. Nodes in graphs having the same community often share fascinating characteristics such as common interest, and common purpose. Hence community detection is one of the most important problems in SNA. Here, the important challenge is identifying connections or links in the communities (Wang and Liao 2014). Several surveys on identifying communities in networks can be found in Chintalapudi and Krishna Prasad (2015), Pourkazemi and Keyvanpour (2013). Deep auto encoders is another successful method to address the problem of community detection. this work proposed a novel deep autoencoder with Particle Swarm Optimization (PSO) and continuation algorithms to reveal community structures in complex networks (Al-Andoli 2021). The research of dynamic community detection where community structures change over time. In the paper an efficient and effective multi-objective method, namely DYN-MODPSO, is proposed, and where the traditional evolutionary clustering framework and the particle swarm algorithm are modified and enhanced (Ying Yin and YuhaiZhao 2020). Many significant challenges arises during community detection. The dynamic nature of networks, scalability issues, complexity of algorithm computation and heterogeneity of the networks like issues are arising with the increase in network size.Local structure based method of link prediction captures the similarity scores based on common neighbor which gives an accurate measure to know the link structure arising between nodes. As a result some interesting and potential links may be missed and also as a matter of fact it will be difficult and time consuming for computing similar score for all nodes in network. Similarly for global based methods for link prediction computes score paths separated by greater than 2. But calculating is again a time consuming process. In SNP, Most of the problems are NP-hard in nature (Cai et al. 2016). There is always a considerable amount of challenges available when it comes to optimization problems as the networks grow in wider as the number of users getting increased this can be treated as a scalable issue and need to have effective computation when working on networks related to various types and also many issues will come into picture as the network size grows. Many meta-heuristic algorithms have been investigated to solve the problems discussed above (Gandomi et al. 2013a; Mirjalili et al. 2017a). For instance, the challenges outlined above, have been successfully addressed using dynamics of many biological systems (Kotteeswaran and Rajesh 2014). The inherent characteristics such as adaptive, self organize and collective behaviour of biological systems are effectively applied to solve such complex problems. The most predominant and successful classes in the bio inspired computing are evolutionary based computing and swarm intelligence based computing. Hence, researchers are adopted to use nature-inspired algorithms to solve different problems in SNA. In this article, we are reviewing a number of bio-inspired algorithms used in community detection from different aspects in the following sections.

The major contributions in this paper are:

- Discussion on bio-inspired algorithms and their classifications.
- Discussion on the challenges that arise while analyzing the social graph, such as community analysis, connection analysis, and path analysis.
- The objective functions used in community detection research has been discussed. How the evolutionary algorithms,swarm-based algorithms helps to find out the solutions in community and connection analysis have been discussed.
- In swarm-based algorithms particularly, four well-known algorithms namely Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Artificial Bee Colony (ABC), and Firefly algorithm have been discussed.

Apart from solving the link prediction and community detection problems. More recently, a number of studies have reported on the success of such techniques for solving difficult problems on all key areas and computer science like optimization problems in data mining and rule extraction dynamic and multiple criteria website optimizations using genetic algorithms, in the field of data mining, ML algorithms using genetic programming and differential evolution in the field of structural engineering with evolution strategy, in the field of multi-modal biomedical imaging using PSO, ACO and many more. The organization of the paper is shown in Fig. 1. In Sect. 2, we summarized the nature-inspired algorithms. Sect. 3 highlights different challenges while analyzing the social network, that is performed on the graphs. In Sects. 3 and 4 addresses the different evolutionary approaches and Sect. 5 introduces the swarm intelligence based algorithms in social networks. Section 6 discusses the future scope of the work followed by a conclusion in Sect. 7.

Fig. 1 Taxonomy of the paper
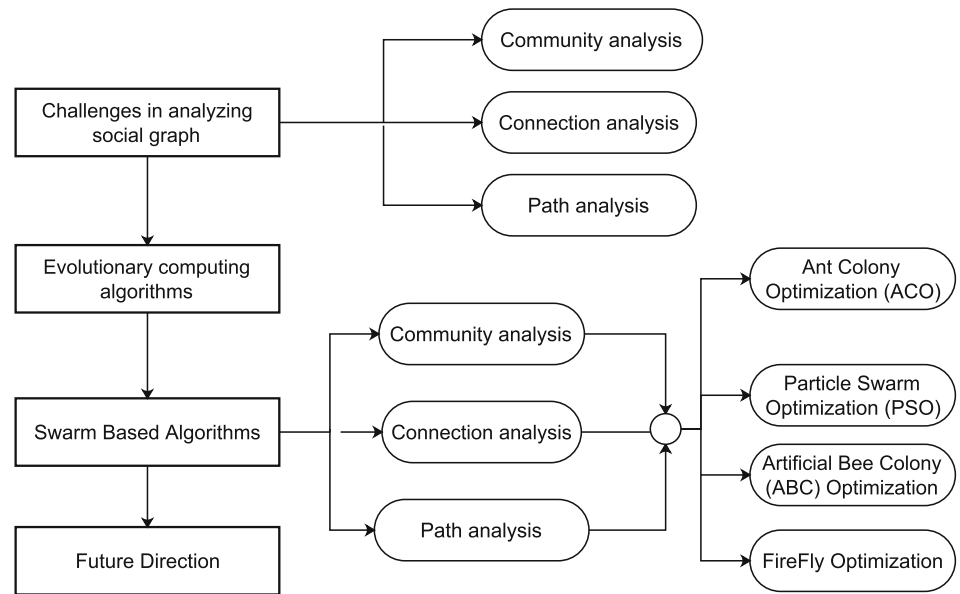
2 Background of nature inspired algorithms

Usually, real-world optimization problems are NP-hard in nature, which are difficult to solve using classical deterministic techniques (Rai and Tyagi 2013). In this aspect, the role of nature-inspired techniques have been proven to be great and have been applied to solve many such problems associated with different research domains. Last few decades the popularity and usage of nature-inspired algorithms have been increased. These algorithms are generally inspired by different biological swarm activities that happen in nature. Figure 2, shows the taxonomy of nature-inspired techniques.

The evolutionary algorithms (EAs) are the well-known and extensively used bio-inspired algorithms, for solving real-time problems to find optimum solutions for complex optimization problems. The Genetic algorithm (GA), Differential Evolution (DE), Evolutionary strategy (ES)are the most well-known algorithms in the class of EA. These techniques have their own properties. For instance, the GA is purely reliant on the genetic parameters. By choosing the appropriate values of mutation and crossover, GA may produce better optimization results.

Swarm Intelligence (SI) is another field of studying and designing well-organized computational intelligent interactive solutions. The complex problems are being solved

Fig. 2 Taxonomy of Bio-Inspired Optimization Techniques
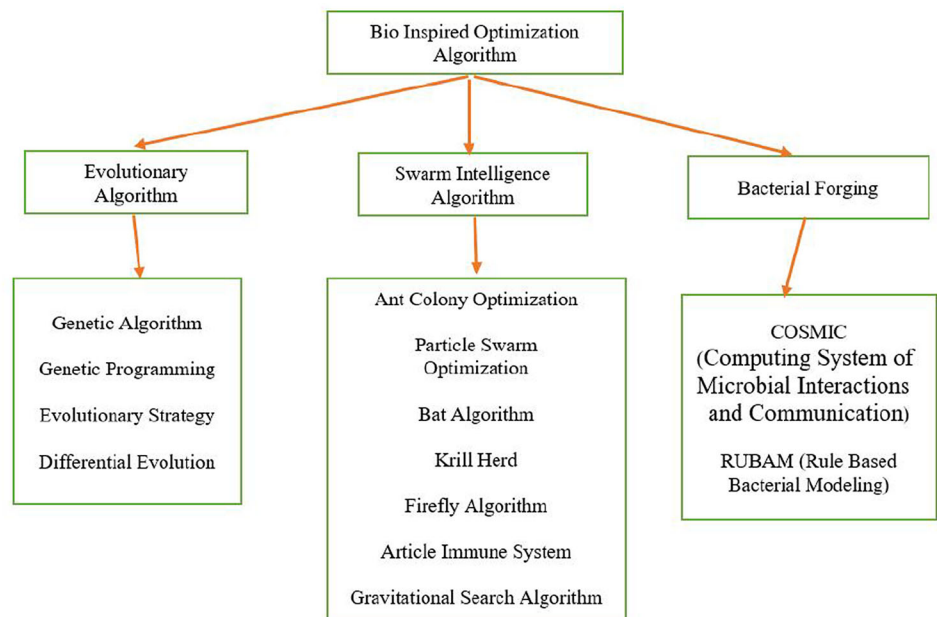
by using the behavior of living swarms such as birds, reptiles, fish, and ants. For instance, the Ant Colony Optimization (ACO) algorithm follows the behavior of ants' food searching pattern. Generally, ants like insects cooperate indirectly by an odorous chemical called pheromone trails. In a swarm, Each ant in a swarm represents an agent. They laid down the pheromone while searching the food source. In case any ant discovers the origin of food, it returns back by detecting a pheromone trail.

Bacterial Forging algorithm is an evaluation to study of developing a new bio-inspired optimization algorithm. COSMIC (Paton et al. 2004) and RUBAM algorithms are comes under bacterial forging.

3 Challenges in analyzing the social graphs

Some of the interesting challenges in analyzing the social networks are finding the interaction between the people, how they act, how diffusion happens in networks, finding the influential nodes in the network, finding the shortest paths between the nodes, and the evolution of communities in the networks. All these challenges are solved by analyzing the networks which are modeled as graphs. In this section, we review the different analysis techniques that can be performed on the graph along with the corresponding solution approaches using Bio-Inspired computing. Most of the analysis is related to graphs such as directed, undirected, signed, and unsigned graphs.

3.1 Community analysis

One of the main challenges in studying SNA is identifying communities in the network. Community analysis in a graph refers to the process of determining the groups that are structurally equivalent in the network. These structurally equivalent groups are usually termed as communities. Most of the conventional approaches to community discovery such as proposed by Girvan and Newman (2002), Wasserman and Faust (1994), and Kernighan and Lin (1970) etc. use the modularity measure to discover communities. But these methods do not reveal the overlaps among the communities and also need prior information of the size and number of the communities. However, there exists many real situations where a node belongs to multiple groups. For example, nodes in Facebook connect to multiple groups(i.e., a person can be a part of many groups on Facebook). In order to overcome these drawbacks, k – clique percolation method proposed by Palla et al. (2005) and several such methods has been developed (Ahn et al. 2010; Evans and Lambiotte 2009; Fortunato 2010) for detecting overlapped communities. But these methods do not scale well when the size of the network increases and

also results in high computational complexity. It has been proved that the optimization of objective functions such as community score (Clauset et al. 2004), community fitness (Lancichinetti et al. 2009), modularity (Brandes et al. 2008), minimum cut and normal cut (Shi and Malik 2000), which attempts to measure how well a partition divides the network into communities, are NP-Hard problems. To optimize such problems several approaches based on meta heuristics are proposed in the literature. Sections 3 and 4 briefly explains the biologically inspired approaches applied to perform community analysis in social networks.

3.2 Connection analysis

Connection Analysis in a graph is a process of determining the new connections among two vertices in a graph. This process is mentioned as link prediction (Lichtenwalter et al. 2010). Link prediction problems has variety of applications. Link prediction methods are classified into supervised and unsupervised techniques. Pujari and Kanawati (2012) proposed a technique which aggregate the rank of the nodes for link prediction in complex networks. Yl et al. (2015) proposed the link prediction algorithm by assigning weights to the nine structural related features and then aggregated their results to conclude final prediction scores. Another method proposed by De Sá and Prudêncio (2011) used weights to improve supervised link prediction. Several algorithms such as (Newman 2001; Adamic and Adar 2003; Zhou et al. 2009; Lü et al. 2009; Kumari et al. 2020) based on structural properties of a network have been used for link prediction. To avoid the high computation cost of these methods, some heuristic based methods are used in link prediction. In Sects. 3 and 4 several biologically inspired approaches are Discussed to predict the links in social networks.

3.3 Path analysis

Path Analysis in a graph is a process of determining the best path or the shortest path in the graph. Several applications based on the path analysing are discussed here.

1. for finding the communities in social networks which are known by calculating the number of shortest paths (Wasserman and Faust 1994; Freeman 1977) or density of random walks (Newman 2005),
2. for predicting the links in social networks such as SIM rank (Jeh and Widom 2002), superimposed random walk (Liu and Lü 2010), Random walk with restart (Liben-Nowell and Kleinberg 2007), local path index (Zhou et al. 2009)

3. for investigating statistical property of network called small world phenomenon which is based on finding average shortest path length

Analyzing the paths in the network has been computationally expensive when the size of the network increases. In order to optimize the efficiency of the methods in the literature, recently many metaheuristics are applied to the problem under consideration which are discussed in the upcoming sections.

and mutation. While the evolutionary strategies and evolutionary programming often use the vector of real numbers and put more emphasis on mutation operation. Reproduction of new solutions acts as key a factor for persevering population diversity. Crossover operators (also called recombination) and mutation operators are used to producing new solutions. Crossover operator combines two or more individuals to generate new individuals also called offsprings. The mutation operator is an unary operator operated as a single individual by introducing random permutations on the individual. The general framework for evolutionary algorithms is outlined below.

Algorithm 1 Evolutionary Computation Procedure

1: Population Initialization(P)
2: Population Evaluation (P)
3: $\partial_{best} \leftarrow$ Find Best of Population (P)
4: **while** (termination condition is not satisfied) **do**
5:     $p^1 \leftarrow$ Recombination (P)
6:     $p^{11} \leftarrow$ Mutation (P^1)
7:     Evaluate population (P^{11})
8:     $\partial' \leftarrow$ Find Best of population (P^{11})
9:     **if** $f(\partial)' < f(\partial_{best})$ **then**
10:         $\partial_{best} \leftarrow \partial'$
11:     **end if**
12: **end while**
13: **return** ∂_{best}

---

4 Approaches to evolutionary computation

4.1 Evolutionary computing algorithms

Evolutionary computing algorithms are inspired from biological models of natural evolution of genomes (Coello et al. 2002). The field of evolutionary computation encompasses several algorithms such as genetic algorithms which uses the genetic operators and evolutionary strategies which also includes differential evolution etc.

Evolutionary algorithms are population-based algorithms, i.e., the algorithm maintains a collection of feasible solutions to the problem under consideration, which is used to create new feasible solutions through the use of genetic operators. Based on the fitness value, solutions are chosen to be part of next-generation and new solutions are produced by applying genetic operators like reproduction, mutation, and recombination. This process repeats until some termination condition is satisfied. Differences among the different evolutionary computation algorithms depend on the representation of the solution and the way the genetic operators are implemented. For example, genetic algorithms use binary or discrete-valued variables to represent a solution and they favor the use of recombination

4.1.1 Community analysis

In the literature, many community detection algorithms have been proposed. Among them, one category is based on the principles of evolutionary computation inspired by biological mechanisms. Community mining methods determine clusters in a large network. A cluster is defined as a group of nodes in the network which share similar properties. Finding the clusters in large networks is considered a NP-hard problem. Evolutionary algorithms are considered as most popular methods for solving such NP-Hard problems either by considering a single objective function or by using multi-objective function that has to be optimized by using the EC algorithms or SI based algorithms. The single objective function that has to be optimized for community identification is modularity, proposed by Newman and Girvan (2003). Modularity is defined as strength of dividing a network in communities (Zhuang et al. 2017).

Modeled the network as a graph $G = (V, E)$ with V indicating the number of nodes in the network and E representing edges in the network, modularity(Q) is termed as follows (Newman 2006):

$$Q = \frac{1}{2|E|} \sum_{l,m}^{n} \left[A_{lm} - \frac{k_l k_m}{2|E|} \right] \delta(c_l, c_m) \qquad (1)$$

where $|E|$ represents the number of edges in the graph G, A_{lm} is the adjacency matrix of graph G and k is the degree distribution and $\delta(c_l, c_m)$ is a function that takes a value 1 when nodes l and m are in the same community.

Many bio-inspired metaheuristics have been presented to optimize single objective function i.e., Modularity (Q) (Tasgin et al. 2007; Pizzuti 2008; Hajeer et al. 2013; Sahoo et al. 2020). In Tasgin et al. (2007), the population is expressed as a set of chromosomes. Each chromosome is represented by a vector with n elements. The position value of the vector is a number assigned to the community and two-position values are the same if both nodes belong to the same community. In Pizzuti (2008), the author uses the adjacency representation for chromosome representation and uses the community scores as the fitness function to evaluate the new solutions. The drawbacks of these methods are maximizing Q is proved to be NP - Hard, this cannot detect communities in overlapped networks and also modularity maximization which is not able to find the similar group structures whose size falls below a certain resolution limit. To overcome these disadvantages many researchers have modeled multi-objective optimization algorithms (Fortunato 2010; Cheng et al. 2019). Pizzuti (2009) proposed multi-objective genetic algorithm in unsigned networks. In this method, two objective functions that need to be optimized are the Maximization of community fitness and Maximization of average degrees within the communities. Other optimization models such as proposed by Agrawal (2011) Maximize the function which is the combination of modularity and average degree measure within the communities. The drawback of this method is that it cannot be extended to other networks in the real world, where the relations can be positive and negative. For example Friend and Foe networks etc. Amelio and Pizzuti (2013) proposed novel multi-objective optimization for detecting communities in the signed networks. In order to detect the communities in the signed network, the authors introduce two objective functions. The first objective function is to maximize modularity Q and the second objective function is to minimize the sum of positive relationships between different communities and the sum of negative relationships within a community. Recently to detect communities in a signed network Shi et al. (2012) proposed a multi-objective evolutionary algorithm based method called MEA-SN. In this method author modeled the community detection task for signed networks, where weights used on the edge represent the positive or negative relationship. The weight > 0 represents a positive relationship and the weight < 0 represents a negative

relationship. The objective function designed for a signed network is based on social balance theory. The 2 objectives designed are maximization of the sum of positive structural similarity values within the community and also maximization of negative structure similarity values between the communities. To overcome the difficulty with the crossover operator and mutation operator Li and Song (2013) proposed a compact genetic algorithm that uses the probability distribution model using the MI algorithm instead of using traditional genetic operations. Shang et al. (2013) proposed an improved genetic algorithm that uses the simulated annealing method as the local search method to improve the ability in adjusting the parameters. In order to improve the search space and the performance of the algorithm, Guerrero et al. (2017) considers the safe and balanced initialization of the number of community structures that are defined using the community density vector that holds the number of nodes in each community. The evolutionary parameters used for analyzing the complex networks in order to detect the communities are listed in the Table. 1.

4.1.2 Connection analysis

Authors in Silva et al. (2010) proposed a friend recommendation system for online social networking sites like LinkedIn, Twitter, FaceBook etc. Their study analyzes the sub-graph, where a node in the sub-graph is connected to another node by at least 3 hops. In this method author constructs a linear combination of topological similarity measures like number of adjacent nodes, density of group formed by common friend of person i and person j and density of group formed by adjacent vertices of node n_i and node n_j and uses the genetic algorithm to calibrate the quotients w_i in relation to optimization function which is given by Eq. 2.

$$M(n, w) = I_{01}(n_c, n)w_{01} + I_{02}(n_c, n)w_{02} + I_{03}(n_c, n)w_{03} \qquad (2)$$

Weights are represented in binary strings. Crossover operator is used to generating new individuals after excluding the worst individual in the population according to the fitness function value and then mutation operation is applied to children. The algorithm terminates when no improvement in the fitness of the best individual is found. Naruchitparames et al. (2011) extended friend recommendation system proposed by Silva et al. (2010) by adding social features of an individual like shared friends, location, age, etc. In this method, the social genome is represented as a binary string of social features. The drawback of this method is that the algorithm will perform better only if the user information is complete and truthful. Recently

Table 1 Evolutionary parameters for community analysis

Evolutionary parameters	Community analysis
Population representation	$P = [g_1, g_2 \cdots g_p]$ Tasgin et al. (2007), Pizzuti (2008), Guerrero et al. (2017), Li and Song (2013)) $g_i \leftarrow$ represents ith chromosome $P \leftarrow$ Population of chromosomes $g_t = [g_{t1}, g_{t2} \cdots g_{tn}]$ $g_t \leftarrow$ Chromosome vector with n elements $g_{ti} \in \{1, 2, \cdots\}$ $r_m = [r_m^1, r_m^2, \cdots r_m^i \cdots r_m^N]$ Shang et al. (2013) $r_m \leftarrow$ Chromosome encoded as an integer string and mth chromosome in the population $r_m^i \leftarrow$ represents which class the i_{th} node belongs to. Location based adjacency representation Shi et al. (2012) where each genotype g consistes of n genes $g_{11}, g_{22}, \cdots g_{n2}$ and each g_{ii} can take one of adjacent nodes of ii (including nodes of ii itself).
Fitness function	$Q = \frac{1}{2m} \sum_{p,q} (A_{p,q} - \frac{h_p h_q}{2m}) \delta(C_p, C_q)$ Tasgin et al. (2007), Shang et al. (2013), Li and Song (2013) $m \leftarrow$ Total number of nodes $p, q \leftarrow$ Nodes of the given network $h_p, h_q \leftarrow$ degree of the pth and qth nodes $\delta(p, q) \leftarrow 1$ if p and q belongs to similar community $\delta(p, q) \leftarrow 0$ Otherwise $C(r) = \sum_p^r Q(S_p, r)$ Pizzuti (2008) $S_i \leftarrow$ Represents submatrix of order r $M(S, r) = Q(S, r) = \frac{\sum (a_{iJ})^r}{I} S[I, J] \leftarrow$ Submatrix $intra(c) = 1 - \sum_{c \in C} \frac{|E(c)|}{m}$ Shi et al. (2012) $inter(c) = \sum_{c \in C} \left(\frac{\sum_{v \in c} deg(v)}{2m} \right)^2$ $Q(c) = 1 - inter(c) - intra(c)$ $E(c) \leftarrow$ Number of links inside the community,c $m \leftarrow$ Total number of links in the network $C \leftarrow$ Set of partitions $\sum_{c \in C} \frac{|E(c)|}{m} \leftarrow$ Intra link strength of a partition Network degree abstraction by considering the interval [2,6] Guerrero et al. (2017)
Operators	Cross operator Tasgin et al. (2007), Pizzuti (2008), Guerrero et al. (2017) Mutation operator probability model based on Cluster MI algorithm Li and Song (2013) 2- Cross operator and Shang et al. (2013), Shi et al. (2012) mutation operation in combination with simulated annealing search

Bliss et al. (2014) proposed an evolutionary approach to link recommendation in dynamic networks. The author uses an evolution strategy to optimize weights which are used in linear combination of sixteen topological based similarity indexes. Initially from one potential solution algorithm generates a Gaussian cloud of feasible solutions using the covariance matrix. A population of feasible solutions is selected from the Gaussian cloud. The feasible solution with the best fitness value is selected and the process is repeated for some determined iterations. The algorithm gets trapped in local optima which is a drawback of this method. In Table. 2 evolutionary parameters for connection analysis is discussed.

5 Approaches to swarm based algorithms

5.1 Ant colony optimization(ACO)

Ant colony optimization is a swarm intelligent based algorithm that is derived from the natural behaviour of the ants (Newman 2001). In general, swarm intelligence is a discipline that deals with a system composed of many individuals that coordinates using decentralized and self organization of natural and artificial systems such as Ant colonies, flocks of fishes, or swarms of birds or bees (Gandomi and Alavi 2012). Such systems are typically composed of individuals that are relatively homogeneous and interaction among the individuals are local to each other and with their environment. Ant colony optimization is inspired by foraging behaviour of the ant colonies. In natural world, ants wander randomly in search of food, and upon locating food return to their nest while laying down the pheromone trails. Further ants follow the trial which are laid in the previous iteration. In this manner, whenan ant finds a decent way from the home to the food source, different ants are bound to follow a similar way. In the long run prompts all the ants following a solitary way.

The complete ACO calculation is portrayed as underneath (Dorigo and Stützle 2009).

Algorithm 2 S-ACO Procedure

1: Initialization of parameters like pheromone trails, heuristic information and number of ants.
2: **while** (termination condition not satisfied) **do**
3:     Construct Solutión();
4:     Daemon Actión();
5:     Update Pheromone();
6: **end while**.

5.1.1 Construct solution()

In this step ants build the solution by applying a step by step decision policy. The decision policy is given by

$$p_{ij}^k = \begin{cases} \dfrac{[\tau_{ij}(t)]^\alpha \times [n_{ij}]^\beta}{\sum_{l \in N_i^k} [\tau_{ij}(t)]^\alpha \times [n_{il}]^\beta} & \text{if } j \in N_i^k \\ 0 & \text{Otherwise} \end{cases} \quad (3)$$

The probability decision rule is a function of

1. Locally available pheromone tails (τ_{ij}) and the heuristic values that are associated with nodes $\tau_i n_i$ and with the edges ($\tau_{ij} n_{ij}$)
2. N_i^k is the neighborhood of the anti K when in node i. In S-ACO the neighborhood of a node i contains all the nodes that are not yet visited thereby avoiding the repetition of the nodes.
3. α and β are the adjustable parameters that controls the pheromone trails and the heuristic information attached to the edges.

Once an ant has completed the tour or while building the tour, the ant evaluates the complete tour or partial tour by the update pheromone procedure to decide how much pheromone to be deposited.

5.1.2 Update pheromone trails

After all the ants completing the tour, the pheromone trails are updated. First, pheromone values are lowered by a constant factor and then add pheromone on the edges the ants have visited in their tours. Pheromone evaporation is given by

$$\tau_{ij}(t) \leftarrow (1-p)\tau_{ij}(t) \quad \text{for all} \quad i,j \in L \quad (4)$$

where P With $O \leq P < 1$ is the pheromone trail evaporation rate which helps to avoid quick convergence and also helps the ants to continue their search process. Each ant K, deposits an amount $\Delta\tau_{ij}^k$ on each edge it has visited and is given by

$$\tau_{ij}(t) \leftarrow \tau_{ij}(t) + \sum_{k=1}^n \Delta\tau_{ij}^k(t) \quad \text{for all} \quad (i,j) \in L \quad (5)$$

$$\Delta\tau_{ij}^k = \begin{cases} \dfrac{1}{C^k(t)} & \text{if } arc(i,j) \in \tau^k \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where $C^k(t)$ is the path length traveled by the kth ant and is computed as the sum of the length of the edges belonging to τ^K. In the next section challenges in social network analysis are discussed that are based on ACO and its variants.

5.1.3 Connection analysis

Predicting connections between the nodes has become popular issue as the size of the online social network increases. Recently ACO, a population based meta heuristic, has been used for fixing NP Hard problems which are visualized in the form of graph. Chen and Chen (2014) proposed supervised link prediction algorithm. In this

Table 2 Evolutionary parameters for connection analysis

Evolutionary parameters	Connection analysis
Individual food source representation	Each candidate solution is given by a vector $\vec{v}$ holds the set of real values between 0 and 1, taken from a uniform random distribution and
	is represented by $\vec{v} = [v_1, v_2, \cdots v_{16}] \in R^{16}$. A population of candidate solutions is selected from a multivariate gaussian cloud. Bliss et al. (2014) Each individual is a central node in the network and each node individual is associated with a social genome which is represented by binary string of 10 social features. Population is a set of all potential friendships with related to each individual. Naruchitparames et al. (2011) Population of each solution is composed of 3 bytes. Weight of each index uses d by each and it ranges from 0 to 255. Silva et al. (2010)
Fitness function	Link prediction error is defined by Bliss et al. (2014) $Fitness = \frac{No.of incorrect Links}{No.of predicted links}$ Fitness of social genome and the individual is calculated as length of the social genome which is directly proportional to the number of 1's in the genome vector. Naruchitparames et al. (2011) Fitness of the individual(central user) is calculated by considering the candidate set of all friends connected to the central user. Silva et al. (2010)
Probability decision rule	CrossOver and Mutation operators. Bliss et al. (2014), Silva et al. (2010) Elitist replacement strategy for exploration ofthe search space and uses the single point crossover and bit flip mutation operators. Naruchitparames et al. (2011)

method a logical graph is formulated by considering all the missing edges. Initially Ants are placed randomly on the selected nodes. Ants construct their tour by selecting the next node with a probability P_{ij}, which is function of pheromone trail on the edge and the set of common neighbours as heuristic information represented by (n_{ij}) on the edge between the nodes (V_i, V_j). After all the ants complete their forward tour, the pheromone values on each edge that has visited by the ant will be updated. Pheromone values are updated based on quality of the path travelled by the ant. Finally, the algorithm terminates at certain iterations or when the pheromone trials are stabilized. Final values of the pheromones on the edges are used as the similarity score between the nodes.But the algorithm converges quickly when the large amount of pheromone is deposited on the edges. Later Sherkat et al. (2015) proposed unsupervised link prediction in their approach, the solution construction is done initially by keeping each ant randomly on the selected graph node. The key objective for these ants is to find triangles (or a triad), i.e., a sub graph with 3 nodes and to update the pheromone of all edges of that triangle. In this method, ants construct the tour by selecting the next node which is inversely proportional to the pheromone trial laid on the edge that enable ants to explore more paths. After finding the triangles pheromones on the triangle edges are increased by one unit. In this way ants explore the triangles till all ants dies or until certain number of iterations is reached. This algorithm works better when the network is strongly connected. The ACO parameters used for analyzing the complex networks inorder to find the links between 2 nodes are listed in the Table. 3.

5.1.4 Path analysis

Based on the small world phenomenon, random networks has been investigated (Lawrence and Latha 2015). The proposed algorithm provides an approach to find shortest path of length six between source and the destination using the foraging behaviour of the ants.In this method the author used the greedy policy inorder to find the next decision node to build the shortest path. Rivero et al. (2011) proposed an algorithm for finding the relationships in social networks. In order to find relationships, the author investigated the chain of two people that are separated by 2 hops. The ants find the destination node by considering the path that goes through node that has higher centrality, i.e., through a person with lot of friends. Initially pheromone is deposited on edges, food is deposited on the selected nodes. Food odour helps the ant to find the path. The selection of the next node from the valid list of nodes depends on the pheromones on the edges. As a result, the

time required to diffuse food odour and different values of threshold used for food odour intensity depends on the mentioned parameters. Honghao et al. (2013) proposed an algorithm to detect communities based on the path analysis. In this method the author detects the communities by using random walk as a heuristic rule. Each ant in the algorithm move to the next node randomly with in its neighbourhood. After completion of each iteration, collective behaviour of all the ants is used to produce the current solution and this process continues till the algorithm converges. The ACO parameters used for analyzing the complex networks inorder to detect the links in the network are listed in the Table. 4.

5.1.5 Community analysis

Mandala et al. (2013) proposed the first approach to detect the community in social approach using min – max ant system. Their study relies on the idea of finding maximum cliques, i.e., the sub-graph where the maximum nodes are connected to each other. But once the clique is found, ants continue their travel on the graph to find other cliques by maintaining a tabu list for the visited nodes as well as the construction cliques. In order to avoid quick convergence and to get global optimum solution initial pheromone values are initialized to upper pheromone trail limit and the pheromone update is measured based on the best tour travelled by the ant in the current iteration. Later, Romdhane et al. (2013) proposed a method that detects the communities by optimizing modularity measure. But the algorithm will be able to partition the network only into two groups. Most of the existing approaches were devoted to undirected network like facebook etc. Thus, Poli et al. (2007) Proposed an algorithm for community mining in directed social networks. In order to maximize the modularity value i.e., the measure used to find the quality of the community, their study uses ant colony optimization. Their study starts with an empty solution i.e., a set of computed empty partitions. Then initialize the community by placing the ant on the most influential vertex and the ants construct the solution by selecting the vertex from the set free direct neighbours with a probability P_{ij} subjected to the importance of the vertex j and the purity of the vertex j belonging to the community C_i. Finally the algorithm terminates when all the nodes are visited. Mandala et al. This method forms communities that have direct connections between the nodes but ignores the nodes with multiple connections. Mandala et al. (2013) proposed method for clustering dynamic graphs. Their methodology utilizes various ant colonies, every province has hive at its middle. Calculation introduces K hubs with the most noteworthy hub degrees at hive position. In every cycle ants will investigate graph

Table 3 ACO parameters for connection analysis

ACO parameters	Connection analysis
Pheromone values	Pheromone $\forall$ edges $= 1$ Sherkat et al. (2015) $\tau_{(ij)} = \lambda \times (a_{(ij)} + \varepsilon)$ Chen and Chen (2014) $n_{(ij)} = \gamma \times \Gamma(i,j)$ $\Gamma(i,j) \leftarrow$ Common neighbours $\lambda, \gamma, \varepsilon \leftarrow$ positive Constants $\tau_{(ij)} = \delta$ Cao et al. (2018) $n_{(ij)} = \sum_{z \in \Gamma(i) \cap \Gamma(j)} \frac{1}{k(z)^l} + B$ $B = w \sum_{x,y \in l_{i \rightarrow j}} \frac{1}{|k(x)k(y)|^l}$ $\mu_j = \frac{1}{|\alpha_j|^2}$ $\alpha_i = \frac{1}{1.44}, \beta_i = \frac{1}{1.44}$, w $= 0.01$ $\Gamma(i), \Gamma(j) \leftarrow$ Set of neighbours $k(x), k(y), k(z) \leftarrow$ degree of nodes x, y and z $l_{i \rightarrow j} \leftarrow$ Node set in the paths from i to j
Fitness function	$Q(S) = C \times \frac{1}{n} \sum_{i=1}^{n} d(V_i)$ Cao et al. (2018), Sherkat et al. (2015), Chen and Chen (2014) $d(V_i) \leftarrow$ degree of the vertex i
Update rule	Increase the pheromone of the edges by one unit Sherkat et al. (2015) $\tau_{(ij)}(t+1) = \rho \times \tau_{(ij)}(t) + \Delta\tau_{(ij)}(t)$ Chen and Chen (2014) $\Delta\tau_{(ij)}(t) = \sum_{k=1}^{m} \Delta\tau^k_{(ij)}(j)$ $\Delta\tau^k_{(ij)}(t) = Q(S)$ $\tau^{new}_{(ij)} = \rho \times \tau^{old}_{(ij)} + \Delta\tau(i,j)$ Cao et al. (2018) $\Delta\tau(i,j) \leftarrow$ variation of pheromone link $\Delta\tau(i,j) = \sum_{k=1} m\Delta\tau_k(i,j)$ $\Delta\tau_k(i,j) \leftarrow$ Pheromone released by the ant k on the link
Probability rule	$P^k_{(ij)} = \frac{\left(\frac{1}{\tau_{(ij)}}\right)^\alpha}{\sum_{l \in N^k_i}\left(\frac{1}{\tau_{(il)}}\right)^\alpha}$ Sherkat et al. (2015) $P^k_{(ij)} = \frac{\tau^\alpha_{(ij)} \times n^\beta_{(ij)}}{\sum_{i=1}^{n} \tau^\alpha_{(ij)} \times n^\beta_{(ij)}}$ Chen and Chen (2014) $P^k_{(ij)} = \frac{P^k_{(ij)}}{\sum_l P^k_{(il)}}$ Cao et al. (2018) $P^k_{(ij)} = \frac{\tau^\alpha_{(ij)} \times n^\beta_{(ij)} \times \mu^\gamma_{(j)}}{\sum_l \tau^\alpha_{(il)} \times n^\beta_{(il)} \times \mu^\gamma_{(l)}}$

almost a hive position by adding the hubs to same hive with most noteworthy part of pheromone. After the completion of one cycle, pheromones will get refreshed and the hive will reassigned so that hub with the minimal average briefest way length to all the hubs alloted to a similar hive. This strategy utilizes the nearby pursuit which prompts the neighborhood ideal and time utilization in genuine networks. The ACO boundaries utilized for investigating the complex networks inorder to identify the network communities are recorded in the Table. 5.

5.2 Particle swarm optimization(PSO)

Particle Swarm Optimization is a population based stochastic optimization technique that simulates the behaviour of bird flocking or fish schooling. It is proposed by Kennedy and Eberhart (1995). PSO has been successfully applied to many engineering problems because of its unique nature in searching mechanism, efficiency in computation and easy implementation. Instead of chromosomes, PSO has particles that form its population called swarm. Unlike in genetic algorithm, there is no "survival

Table 4 ACO parameters for path analysis

ACO parameters	Path analysis
Pheromone values	Centroid of graphs Rivero et al. (2011) Heuristic info: Honghao et al. (2013) $C(i,j) = \frac{\sum_{v_l \in V}(A_{il} - \mu_i)(A_{jl} - \mu_j)}{n\sigma_i\sigma_j}$ $n_{ij} = \frac{1}{1+e^{-(ij)}}$ $\mu_i \leftarrow \sum_l \left(\frac{A_{ij}}{n}\right)$ $\sigma_i \leftarrow \sqrt{\sum_l \left(\frac{A_{il} - \mu_i^2}{n}\right)}$
Fitness Function	$Q = \frac{1}{2m}\sum_{i,j}^{n}\left(A_{ij} - \frac{k_i.k_j}{2m}\right)\delta(C_i, C_j)$ Honghao et al. (2013) $C_i \leftarrow$ community index of vertex i $k_i \leftarrow$ degree of vertex i
Update Rule	$O(n_i) = O(n_i) - k \times w_{ij}$ Rivero et al. (2011) $\tau_{ij} \leftarrow \rho\tau_{ij} + \Delta\tau_{ij}$ Honghao et al. (2013) $\rho \leftarrow 0 \leq \rho < 1$ $\Delta\tau_{ij} \leftarrow$ pheromone edge $\Delta\tau_{ij} = M(S^{ib})$ (if edge (i,j) is in S^{ib}) $\Delta\tau_{ij} = 0$ otherwise
Probability Rule	$P[n_i, n_j] = \frac{\tau_{ij}}{\sum_{k \in node_i}(\tau_{ij})}$ Rivero et al. (2011) $P^k_{(ij)} = \frac{\tau^\alpha_{(ij)} \times n^\beta_{(ij)}}{\sum_{i=1}^{n} \tau^\alpha_{(ij)} \times n^\beta_{(ij)}}$ Honghao et al. (2013)

of the fittest" selection process for the particles, but rather PSO uses mutation strategy where each particle is simply moved from one place to another place. The PSO algoritm utilizes potential arrangements called particles, that fly through the problem space by following the current ideal particles. At every cycle, the speed of every particle is stochastically refreshed by the best arrangement that the particle accomplished up until now and the local best position. The simple usage and brisk computational assembly makes PSO a famous enhancement technique for taking care of continuous and discrete improvement issues. The PSO algorithm is outlined as below:

et al. (2010), and PSO algorithm. Kernigh-Lin algorithm discover communities by determining the maximum value of modularity function. Genetic algorithms detect communities by optimizing single or multiple objective functions. However these both methods need prior information about the size and number of communities and too many parameters that need to be initialized. Basically PSO algorithms are designed for solving continuous optimization problem (Shi and Eberhart 1998, 2001). Later for updating the individual velocity and position, the researchers redefined the equation by using the swap operator (Mitrović and Tadić 2009), crisp - set meth-

---

Algorithm 3 PSO Procedure

1: Initialize particle position and velocity vectors.
2: **while** (Termination condition not satisfied) **do**
3:     For each particle calculate fitness value. Compare the individual fitness value with its P_{best}.
4:     **if** current value is improved than the best value (Pop_{best}) in history **then**
5:         set current value as the new Pop_{best}
6:     **else**
7: Find the particle with the good fitness value of all the particles as the Glo_{best}.
8:     **end if**
9:     Update the velocity and position of each population.
10: **end while**

In the following analysis of social networks are outlined based on Particle Swarm Optimization and its variance.

5.2.1 Community analysis

Many optimization methods are applied for solving the community detection problem which is considered as a NP – Hard problem. Typical clustering techniques includes Kernigh-Lin algorithm [57], genetic algorithm Leskovec

od (Chen et al. 2010), fuzzy - matrix method (Liao et al. 2007). Because of its quick convergence and easy implementation they are extended for solving discrete optimization problem. Cai et al. (2015) proposed a discrete PSO algorithm for finding the communities in large scale social networks. In this method, the authors have carefully redefined the equations to generate the discrete values and also used greedy method for updating particle velocity and position. Chen and Qiu (2013) proposed a novel PSO based

Table 5 ACO parameters for community analysis

ACO Parameters	Community analysis
Pheromone values	$\tau_i = 0.5$ Mandala et al. (2013) Vertex in the graph Romdhane et al. (2013)
Fitness function	$max(x) = \frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}x_i x_j$ Mandala et al. (2013) where $x \in \{0,1\}^n$ $DP(p) = \frac{Density(p)+Purity(p)}{2}$ Romdhane et al. (2013)
	$Density(P) = \frac{1}{|P|}\sum_{C_i \in P} Density(C_i)$ $Density(C) = \frac{1}{|C|}\sum_{S_i \in C} Density(S_i, C)$ $Density(S,C) = \frac{Degree_{in}(S,C)}{Degree(S)}$ $Purity(C) = \frac{1}{|c|}\sum_{S_i} I(S_i)$
Update rule	$S^* = \sum_{s \in S}\frac{Q(S)}{\sum_{s \in S} Q(S)} \times S$ Mandala et al. (2013) $\tau \leftarrow \tau + \rho(S^* - \tau)$ Update pheromone with same color label as S. Romdhane et al. (2013)
Probability rule	$P(S_i^{kt} = 1) = \tau_i$ Mandala et al. (2013) $P(S_i^{kt} = 0) = 1 - \tau_i$ $S_i^{kt} \rightarrow$ potential solution of ant k in iteration t $P_{ij} = I(j) + purity(j, C_j)$ Romdhane et al. (2013)

on modularity optimization for detecting the clusters in social networks. The algorithm uses modified PSO to optimize the modularity in new constructed weighted network which is compressed network taken from the original network. Later, Cao et al. (2015) proposed PSO for discrete optimization problem by considering the connections in the network when initializing the population and there is

determined utilizing the wellness esteems gave by worker honey bees. At that point apply ravenous determination between new arrangement and past arrangement found by spectator honey bee by Memorizing the best one. This cycle proceeds until some end standards is met. Outline of ABC algorithm:

In the following sections analysis of social networks are

Algorithm 4 ABC Procedure

1: Initialization phase
2: **while** (**do**Termination condition not met)
3:    Employee/worker bees phase
4:    Spectator bees phase
5:    Scout bees phase
6:    Memorize the best arrangement achieved so far
7: **end while**

---

no need of prior information about the size and number of communities. In this method, authors used modulo operator and different operators which generates an integer value in $IDPSO_{RO}$ algorithm. In order to increase the accuracy of the algorithm, the author introduce a new method called community correction strategy for updating individual velocity and individual position with best solution at 1/2, 1/2, 3/4 of total iterations. The PSO parameters used for analyzing the complex networks inorder to detect the communities are listed in the Table. 6.

5.3 Artificial bee colony (ABC) algorithm

Artificial Bee Colony is an as of late created swarm knowledge based algorithm proposed by Karaboga and Basturk in 2006 to formulate NP-Hard problems Karaboga and Basturk (2007). It was enlivened by self organizing and aggregate intelligent behaviour of honey bee colonies. In ABC, artificial bees colony contains three kind of bees : Employee/worker honey bees, spectator honey bees and scout honey bees. Every worker honey bee fly around in multi dimensional quest space for deciding new arrangements with the area of the current one and assess them. The calculation utilizes avaricious choice cycle between the past one and the upgraded one. The worker honey bee whose arrangement has been disposed of will turn into a scout honey bee, at that point the scout honey bee caries out looking for additional new arrangements and constrained by a boundary called "limit". When the investigation of new arrangements is finished by worker honey bees, they return to hive and offer their data about gainfulness of the food source with the passerby honey bees in the hive. Presently the passerby honey bee pick an answer contingent upon bigger productivity esteem which are

outlined based on ABC algorithm.

5.3.1 Connection analysis

Friend recommendation is a fundamental problem in online social networks that aims to suggest new links usually done by analyzing the links connected between the nodes. Several algorithms are developed based on the structural properties of the network. But when the network grows, friend recommendation problem becomes NP Hard problem. Typically Bio Inspired algorithm works better than the conventional algorithm. One such Category is ABC algorithm. Akbari et al. (2013) recommends new friends using the structural properties of network. In the proposed approach ABC is applied to optimize the weight coefficients assigned to the topological features of the network. The fitness function for finding the quality of each solution is the error function which is defined as sum of differences between the actual value and expected values. This method effectively recommends new links over other methods such as KNN, SVM, MLP neural network and topological GA based approach. The ABC parameters used for analyzing the complex networks inorder to predict the connections among the nodes in the network are listed in the Table. 7.

5.3.2 Community analysis

Network community discovery can be considered as optimization issue in which a target quality function called modularity, wellness score and so forth are picked to be improved. Hafez et al. (2014) to speak to a network structure, have applied locus based adjacency encoding plan proposed by Shi et al. (2009) and to speak to the every individual food source he utilizes a vector $\overrightarrow{X}$ containing

Table 6 PSO parameters for community analysis

PSO parameters	Community analysis
Particle representation	$p_i = (p_{i,1}, p_{i,2} \cdots p_{i,n})$ Chen and Qiu (2013)
	Position: Cai et al. (2015)
	$X_i = \{X_i^1, X_i^2, \cdots X_i^n\}$
	Where $X_i^j \in [1, n]$ is an integer
	$X_i^j \leftarrow$ label
	identifier which carries cluster information
	$X_i^j = X_i^k \leftarrow$ nodes j and k belongs to same cluster
	Velocity:
	$V_i = \{V_i^1, V_i^2, \cdots V_i^n\}$
	$V_i^j \in \{0\} \leftarrow$ position vector contains original state
	$V_i^j \in \{1\} \leftarrow$ position vector will be changed
	Particle(paticle_id, ismaped, attribute info, neighbouring, clustering info) Chaitanya et al. (2018)
Update strategy	$V_{i,j}(t+1) = wV_{i,j}(t) + PB + GB$ Chen and Qiu (2013)
	$PB = c_1 r_1 (X_{i,j}^{(p)} - X_{i,j}(t))$
	$GB = c_2 r_2 (X_{i,j}^{(g)}(t) - X_{i,j}(t))$
	$X_{i,j}^{(p)} \leftarrow personal\,best\,position$
	$X_{i,j}^{(g)} \leftarrow global\,best\,position$
	$C_1, C_2 \leftarrow learning\,factors$
	$r_1, r_2 \leftarrow random\,values\,in\,the\,region[0, 1]$
	$V_i = wV_i \oplus PB + GB$ Cai et al. (2015)
	$PB = C_1 r_1 \times Pbest_i X_i$
	$GB = C_1 r_2 \times Gbest_i X_i$
	$X_i = X_i \otimes V_i$
Fitness function	$Q = \frac{1}{m} \sum_{i,j} (A_{i,j} - \frac{k_i k_j}{2m}) \delta(C_i, C_j)$ Chen and Qiu (2013)
	$S_{i,j} > \beta$ Chaitanya et al. (2018)
	$S_{i,j} = \frac{1}{k} \sum_{i,j}^{k} S_{ij}$

components that take a worth j in the range[1, n]. A worth j put away in the ith position is deciphered as connection between hub i and hub j, i.e., the creator instates the food sources from its hub neighborhood. The wellness of the arrangement is determined utilizing be diverse target functions, for example, modularity, network score, wellness score and so on. The presentation of the calculation utilized relies upon the quality function utilized. Further ABC, is applied in numerous domain, for example, information mining for performing grouping proposed in (Fathian et al. 2007; Marinakis et al. 2007). The ABC boundaries utilized for investigating the complex networks inorder to distinguish the networks are recorded in the Table. 8.

5.4 Firefly algorithm

Firefly algorithm is an yet another important population based intelligence algorithm inspired by the flashing of fireflies and is developed by Yang (2010). The key idea of Fire fly algorithm is that firefly will be attracted to other mates that has higher intensity and attractiveness. Intensity of each firefly is proportional to quality of the solution and the attractiveness depends on the distance moved towards a more brighter firefly . Outline of Firefly algorithm is given below

---

Algorithm 5 FO Procedure

1: Initialize n fire flies to random positions.
2: **while** Termination condition not met **do**
3:     **for** i: 0 to n-1 **do**
4:         **for** j: 0 to n-1 **do**
5:             **if** Intensity(i) < intensity(j) **then**
6:                 Compute attractiveness
7:                 Move firefly(i) toward firefly(j)
8:                 Evaluate new solutions & Update firefly(i) intensity
9:             **end if**
10:         **end for**
11:     **end for**
12:     Sort the fireflies and find the current best
13: **end while**
14: Return best position found

5.4.1 Community analysis

Amiri et al. (2013) designed multi objective functions to discover communities in large scale networks. In this method, firefly algorithm is used to optimize multiple objective functions. The two objective functions were introduced in this paper, maximization of internal links and minimization of external links. Amiri et al. (2013) consider random movement of firefly as in Caponetto et al. (2003) and the absorption parameter is considerd as in Apostolopoulos (2010). The complexity of algorithm is high because it involves several heuristics such as external repository for storing on dominated Pareto optimal solutions, Niching strategy to find the best global solution from the list of Pareto optimal solutions, fuzzy based clustering and chaotic based tuning of the algorithm. However, the proposed algorithm cannot detect the overlapped communities and also not suitable for dynamic networks. The Firefly parameters used for analyzing the complex networks inorder to detect the communities are listed in the Table. 9.

6 Future directions

Some of the future directions to work on in the area of social network analysis is discussed in this section:

6.1 Employing new metaheuristic measures

Due to recent research in the field of meta heuristics lot of new promising optimization techniques are introduced in the literature such as bat algorithm, cuckoo search, salp swarm algorithm (Mirjalili et al. 2017b), krill herd algorithm (Gandomi and Alavi 2012), Whale Optimization Algorithm (Faramarzi et al. 2020) and they are popular in many diverse applications (Gill and Buyya 2019; Mirjalili et al. 2017b). Using these algorithms in the context of social network analysis problems particularly connection analysis, path analysis and community analysis seems to be fruitful.

6.2 Secure computation on social networks

Social Networks possess lot of vital information and for analysis only anonymized data or network is used. In order to compute data in real world without revealing identity of the users and without affecting privacy of the users is of at

Table 7 ABC parameters for Conection analysis

ABC parameters	Connection analysis
Individual food source representation	Weights in the range [0,1] Akbari et al. (2013)
Fitness function	$nectar_k = \frac{1}{q}\sum_{p=1}^{q} |T_{ij}^p - O_{ij}^p|$ Akbari et al. (2013) $q \leftarrow$ Number of all instances of the data $O_{ij}p \leftarrow$ Ordered value for the lth instance $T_{ij}p \leftarrow$ Target value for the lth instance $k \leftarrow k$th bee
Probability decision rule	Local Search rule Akbari et al. (2013)

Table 8 ABC parameters for community analysis

ABC parameters	Community analysis
Individual food source representation	$\overrightarrow{x} = \{x[1], x[2], \cdots x[n]\}$ Hafez et al. (2014); Belkhiri et al. (2017); Aung et al. (2019) $\overrightarrow{x_r}[i] = RC(i)$ $RC(i) \leftarrow$ Random function that selects randomly a node j from the node i's neighbours
Fitness function	$fit_r(\overrightarrow{x_r}) = \begin{cases} \dfrac{1}{1 + f_r(\overrightarrow{x_r})} & \text{if } f_r(\overrightarrow{x_r}) \geq 0 \\ 1 + abs(f_r(\overrightarrow{x_r})) & \text{if } f_r(\overrightarrow{x_r}) < 0 \end{cases}$ Hafez et al. (2014) $f_r(\overrightarrow{x_r}) \leftarrow$ Objective Function value of solution $\overrightarrow{x_m}$ $\overrightarrow{x_m} \leftarrow$ Neighbourhood Vector $Q = \frac{1}{2m}\sum_{i,j}(A_{i,j} - \frac{h_i h_j}{2m})\delta(C_i, C_j)$ Belkhiri et al. (2017); Aung et al. (2019) $m \leftarrow$ Total number of nodes $i, j \leftarrow$ Nodes in the network $h_i, h_j \leftarrow$ degree of the ith and jth nodes $\delta(i,j) \leftarrow 1$ if i and j belongs to same community $\delta(i,j) \leftarrow 0$ Otherwise
Probability decision rule	$P_r = \dfrac{0.9 \times fit_r(\overrightarrow{x_r})}{fit_{best}} + 0.1$ Hafez et al. (2014) Local search using Flip algorithm Aung et al. (2019); Belkhiri et al. (2017)

Table 9 Firefly parameters for community analysis

Firefly parameters	Community analysis
Fireflies	Each individual of the population consists of N variables $x_1, x_2, \cdots x_N$ and for each variable there is a set of possible range of values based on the adjacency matrix. Entire population is calculated as $x_{i,mfirefly} = \mu \times x_{i,m-1firefly} \times (1 - x_{i,m-1firefly})$ where $m = 2, \cdots N_{firefly}$ Amiri et al. (2013)
Fitness function	$F(x_{i,m}^k) = max_{q=1,2}[\mu_{Fq}^{ref} - \mu_{Fq}(x_{i,m}^k)]$ Amiri et al. (2013) $m \leftarrow$ firefly $k \leftarrow$ Iteration $\mu_q^{ref} \leftarrow$ Membership value of each firefly
Movement of the fireflies	$x_{i,mfly}^k = \begin{cases} x_{i,mfly}^k + \beta^k A + \alpha^k B & \text{if } F(x_{i,nfly}) < F(x_{i,mfly}) \\ x_{i,nfly}^k & \text{Otherwise} \end{cases}$ Amiri et al. (2013) $A \leftarrow (x_{i,nfirefly}^k - x_{i,mfirefly}^k)$ $B \leftarrow (rand - \frac{1}{2})$
Brightness of the firefly	$\beta^k = (\beta_{max} - \beta_{min}e^{-\gamma(r_{mn}^k)^2} + \beta_{min}$ Amiri et al. (2013) $r_{mn}^k \leftarrow$ Euclidean distance between $x_{i,nfly}^k, x_{i,mfly}^k$ $\beta_{min} \leftarrow 0.2$

most importance. Privacy preserving computation, multi party computation or similar techniques have to be developed over social networks so that link prediction and community analysis can be performed over real time data. Performing link prediction and community analysis over real time data has significance in e–commerce and related applications apart from understanding the nature of evolution.

learning area especially for updating the gradients in designing the deep neural network models. Further, due to the problem of convergence speed which can be encountered in solving real challenges.These bio-inspired algorithms could be hybridized with other approaches and methods such as quantum computing and choas theory to enhance the performance of nature inspired optimization algorithms

7 Conclusions

Swarm based optimization techniques used in social network analysis are compared in this paper with respect to community analysis, connection analysis and path analysis. From the comparison we can infer that firefly algorithm provides better result by exploring the solution space efficiently without getting struck at the global optima or local minima. Further it is interesting to explore the application of firefly algorithm in social networks with respect to different computational applications.As a future scope swarm based optimization techniques can be extended to the deep

References

Adamic LA, Adar E (2003) Friends and neighbors on the web. Soc Netw 25(3):211–230

Agrawal R (2011) Bi-objective community detection (bocd) in networks using genetic algorithm. In: Aluru S, Bandyopadhyay S, Catalyurek UV, Dubhashi DP, Jones PH, Parashar M, Schmidt B (eds) International conference on contemporary computing. Springer, Berlin, Heidelberg, pp 5-15

Ahn YY, Bagrow JP, Lehmann S (2010) Link communities reveal multiscale complexity in networks. Nature 466(7307):761–764

Akbari F, Tajfar AH, Nejad AF (2013) Graph-based friend recommendation in social networks using artificial bee colony. In:

Dependable, autonomic and secure computing (DASC), 2013 IEEE 11th international conference on, IEEE, pp 464–468

Al-Andoli M, Cheah WP, Tan SC (2021) Deep auto encoder-based community detection in complex networks with particle swarm optimization and continuation algorithms. J Intell Fuzzy Syst 40(3):4517–4533

Amelio A, Pizzuti C (2013) Community mining in signed networks: a multiobjective approach. In: Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining, ACM, pp 95–99

Amiri B, Hossain L, Crawford JW, Wigand RT (2013) Community detection in complex networks: Multi-objective enhanced firefly algorithm. Knowl-Based Syst 46:1–11

Apostolopoulos T and Vlachos A (2010) Application of the firefly algorithm for solving the economic emissions load dispatch problem. Int J Combinatorics

Aung TT, Nyunt TTS, Cho PPW (2019) Community detection in social graph using nature-inspired based artificial bee colony algorithm with crossover and mutation. In: 2019 IEEE 4th international conference on computer and communication systems (ICCCS), IEEE, pp 213–217

Belkhiri Y, Kamel N, Drias H, Yahiaoui S (2017) Bee swarm optimization for community detection in complex network. In: World conference on information systems and technologies, Springer, pp 73–85

Bliss CA, Frank MR, Danforth CM, Dodds PS (2014) An evolutionary algorithm approach to link prediction in dynamic social networks. J Comput Sci 5(5):750–764

Brandes U, Delling D, Gaertler M, Görke R, Hoefer M, Nikoloski Z, Wagner D (2008) On modularity clustering. Knowl Data Eng IEEE Trans 20(2):172–188

Cai Q, Gong M, Ma L, Ruan S, Yuan F, Jiao L (2015) Greedy discrete particle swarm optimization for large-scale social network clustering. Inf Sci 316:503–516

Cai Q, Ma L, Gong M, Tian D (2016) A survey on network community detection based on evolutionary computation. Int J Bio-Inspired Comput 8(2):84–98

Cao C, Ni Q, Zhai Y (2015) A novel community detection method based on discrete particle swarm optimization algorithms in complex networks. In: Evolutionary computation (CEC). IEEE Congress on, IEEE, pp 171–178

Cao Z, Zhang Y, Guan J, Zhou S (2018) Link prediction based on quantum-inspired ant colony optimization. Sci Rep 8(1):13,389

Caponetto R, Fortuna L, Fazzino S, Xibilia MG (2003) Chaotic sequences to improve the performance of evolutionary algorithms. Evolut Comput, IEEE Trans 7(3):289–304

Chaitanya K, Somayajulu D, Krishna PR (2018) A pso based community detection in social networks with node attributes. In: 2018 IEEE congress on evolutionary computation (CEC), IEEE, pp 1–6

Chen B, Chen L (2014) A link prediction algorithm based on ant colony optimization. Appl Intell 41(3):694–708

Chen WN, Zhang J, Chung HS, Zhong WL, Wu WG, Shi YH (2010) A novel set-based particle swarm optimization method for discrete optimization problems. Evolut Comput IEEE Trans 14(2):278–300

Chen Y, Qiu X (2013) Detecting community structures in social networks with particle swarm optimization. In: Frontiers in internet technologies, Springer, pp 266–275

Cheng J, Su X, Yang H, Li L, Zhang J, Zhao S (2019) Chen X (2019) Neighbor similarity based agglomerative method for community detection in networks. Complexity

Chintalapudi SR, Krishna Prasad M (2015) A survey on community detection algorithms in large scale real world networks. In: Computing for sustainable global development (INDIACom), 2015 2nd international conference on, IEEE, pp 1323–1327

Clauset A, Newman ME, Moore C (2004) Finding community structure in very large networks. Phys Rev E 70(6):066,111

Coello CAC, Van Veldhuizen DA, Lamont GB (2002) Evolutionary algorithms for solving multi-objective problems, vol 242. Springer, Berlin

De Sá HR, Prudêncio RB (2011) Supervised link prediction in weighted networks. In: Neural Networks (IJCNN), The 2011 international joint conference on, IEEE, pp 2281–2288

Dorigo M, Stützle T (2009) Ant colony optimization: overview and recent advances. Techreport, IRIDIA, Universite Libre de Bruxelles

Evans T, Lambiotte R (2009) Line graphs, link partitions, and overlapping communities. Phys Rev E 80(1):016,105

Faramarzi A, Heidarinejad M, Mirjalili S, Gandomi AH (2020) Marine predators algorithm: a nature-inspired metaheuristic. Expert Syst Appl 152:113377

Fathian M, Amiri B, Maroosi A (2007) Application of honey-bee mating optimization algorithm on clustering. Appl Math Comput 190(2):1502–1513

Fortunato S (2010) Community detection in graphs. Phys Rep 486(3–5):75–174

Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry 40(1):35–41

Gandomi AH, Alavi AH (2012) Krill herd: a new bio-inspired optimization algorithm. Commun Nonlinear Sci Numer Simul 17(12):4831–4845

Gandomi AH, Yang XS, Alavi AH (2013a) Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems. Eng Comput 29(1):17–35

Gandomi AH, Yang XS, Talatahari S, Alavi AH (2013b) Metaheuristic algorithms in modeling and optimization. In: Metaheuristic applications in structures and infrastructures, pp 1–24

Gill SS, Buyya R (2019) Bio-inspired algorithms for big data analytics: a survey, taxonomy, and open challenges. In: Dey N, Das H, Naik B, Behera HS (eds) Big data analytics for intelligent healthcare management. Academic Press, pp 1–17

Girvan M, Newman ME (2002) Community structure in social and biological networks. Proc National Acad Sci 99(12):7821–7826

Guerrero M, Montoya FG, Baños R, Alcayde A, Gil C (2017) Adaptive community detection in complex networks using genetic algorithms. Neurocomputing 266:101–113

Hafez AI, Zawbaa HM, Hassanien AE, Fahmy AA (2014) Networks community detection using artificial bee colony swarm optimization. In: Proceedings of the fifth international conference on innovations in bio-inspired computing and applications IBICA 2014, Springer, pp 229–239

Hajeer MH, Singh A, Dasgupta D, Sanyal S (2013) Clustering online social network communities using genetic algorithms. arXiv preprint arXiv:13122237

He Yl, Liu JN, Yx Hu, Wang Xz (2015) Owa operator based link prediction ensemble for social network. Exp Syst Appl 42(1):21–50

Honghao C, Zuren F, Zhigang R (2013) Community detection using ant colony optimization. In: 2013 IEEE congress on evolutionary computation, IEEE, pp 3072–3078

Jeh G, Widom J (2002) Simrank: a measure of structural-context similarity. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp 538–543

Karaboga D, Basturk B (2007) A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm. J Global Optim 39(3):459–471

Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of ICNN'95-international conference on neural networks, IEEE, vol 4, pp 1942–1948

Kernighan BW, Lin S (1970) An efficient heuristic procedure for partitioning graphs. Bell Syst Tech J 49(2):291–307

Kotteeswaran C, Rajesh A (2014) A survey of diverse nature bio-inspired computing models. In: Current Trends in Engineering and Technology (ICCTET), 2014 2nd international conference on, IEEE, pp 120–124

Kumari A, Behera RK, Sahoo KS, Nayyar A, Kumar Luhach A, Prakash Sahoo S (2020) Supervised link prediction using structured-based feature extraction in social network. Concurrency and Computation: Practice and Experience p e5839

Lancichinetti A, Fortunato S, Kertész J (2009) Detecting the overlapping and hierarchical community structure in complex networks. New J Phys 11(3):033,015

Lawrence EE, Latha R (2015) Analysis of six degrees of separation in facebook using ant colony optimization. In: Circuit, power and computing technologies (ICCPCT), 2015 international conference on, IEEE, pp 1–5

Leskovec J, Lang KJ, Mahoney M (2010) Empirical comparison of algorithms for network community detection. In: Proceedings of the 19th international conference on World wide web, ACM, pp 631–640

Li J, Song Y (2013) Community detection in complex networks using extended compact genetic algorithm. Soft Comput 17(6):925–937

Liao CJ, Tseng CT, Luarn P (2007) A discrete version of particle swarm optimization for flowshop scheduling problems. Comput Oper Res 34(10):3099–3111

Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. J Am Soc Inf Sci Technol 58(7):1019–1031

Lichtenwalter RN, Lussier JT, Chawla NV (2010) New perspectives and methods in link prediction. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp 243–252

Liu W, Lü L (2010) Link prediction based on local random walk. EPL (Europhysics Letters) 89(5):58,007

Lü L, Jin CH, Zhou T (2009) Similarity index based on local paths for link prediction of complex networks. Phys. Rev. E 80(4):046,122

Mandala SR, Kumara SR, Rao CR, Albert R (2013) Clustering social networks using ant colony optimization. Oper Res 13(1):47–65

Marinakis Y, Marinaki M, Matsatsinis N (2007) A hybrid clustering algorithm based on honey bees mating optimization and greedy randomized adaptive search procedure. In: Maniezzo V, Battiti R, Watson J-P (eds) International conference on learning and intelligent optimization. Springer, Berlin, Heidelberg, pp 138-152

Mirjalili S, Gandomi AH, Mirjalili SZ, Saremi S, Faris H, Mirjalili SM (2017a) Salp swarm algorithm: A bio-inspired optimizer for engineering design problems. Adv Eng Softw 114:163–191

Mirjalili S, Gandomi AH, Mirjalili SZ, Saremi S, Faris H, Mirjalili SM (2017b) Salp swarm algorithm: A bio-inspired optimizer for engineering design problems. Adv Eng Softw 114:163–191

Mitrović M, Tadić B (2009) Spectral and dynamical properties in classes of sparse networks with mesoscopic inhomogeneities. Phys Rev E 80(2):026,123

Naruchitparames J, Gunes MH, Louis SJ (2011) (2011) Friend recommendations in social networks using genetic algorithms and network topology. Evolutionary computation (CEC). IEEE congress on, IEEE, pp 2207–2214

Newman ME (2001) Clustering and preferential attachment in growing networks. Phys Rev E 64(2):025,102

Newman ME (2005) A measure of betweenness centrality based on random walks. Soc Netw 27(1):39–54

Newman ME (2006) Modularity and community structure in networks. Proc National Acad Sci 103(23):8577–8582

Newman ME, Girvan M (2003) Mixing patterns and community structure in networks. In: Pastor-Satorras R, Rubi M, Diaz-Guilera A (eds) Statistical mechanics of complex networks. Springer, Berlin, Heidelberg, pp 66–87

Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435(7043):814–818

Paton R, Gregory R, Vlachos C, Saunders J, Wu H (2004) Evolvable social agents for bacterial systems modeling. IEEE Trans Nanobiosci 3(3):208–216

Pizzuti C (2008) Ga-net: A genetic algorithm for community detection in social networks. In: Parallel problem solving from nature–PPSN X, Springer, pp 1081–1090

Pizzuti C (2009) A multi-objective genetic algorithm for community detection in networks. In: Tools with artificial intelligence, 2009. ICTAI'09. 21st international conference on, IEEE, pp 379–386

Poli R, Kennedy J, Blackwell T (2007) Particle swarm optimization. Swarm Intell 1(1):33–57

Pourkazemi M, Keyvanpour M (2013) A survey on community detection methods based on the nature of social networks. In: Computer and knowledge engineering (ICCKE), 2013 3th international econference on, IEEE, pp 114–120

Pujari M, Kanawati R (2012) Supervised rank aggregation approach for link prediction in complex networks. In: Proceedings of the 21st international conference companion on world wide web, ACM, pp 1189–1196

Rai D, Tyagi K (2013) Bio-inspired optimization techniques: a critical comparative study. ACM SIGSOFT Softw Eng Notes 38(4):1–7

Rivero J, Cuadra D, Calle FJ, Isasi P (2011) A bio-inspired algorithm for searching relationships in social networks. In: Computational aspects of social networks (CASoN), 2011 international conference on, IEEE, pp 60–65

Romdhane LB, Chaabani Y, Zardi H, Group MR et al (2013) A robust ant colony optimization-based algorithm for community mining in large scale oriented social graphs. Exp Syst Appl 40(14):5709–5718

Sahoo KS, Tripathy BK, Naik K, Ramasubbareddy S, Balusamy B, Khari M, Burgos D (2020) An evolutionary svm model for ddos attack detection in software defined networks. IEEE Access 8:132,502-132,513

Shang R, Bai J, Jiao L, Jin C (2013) Community detection based on modularity and an improved genetic algorithm. Phys A: Stat Mech Appl 392(5):1215–1231

Sherkat E, Rahgozar M, Asadpour M (2015) Structural link prediction based on ant colony approach in social networks. Phys A: Stat Mech Appl 419:80–94

Shi C, Wang Y, Wu B, Zhong C (2009) A new genetic algorithm for community detection. In: International conference on complex sciences, Springer, pp 1298–1309

Shi C, Yan Z, Cai Y, Wu B (2012) Multi-objective community detection in complex networks. Appl Soft Comput 12(2):850–859

Shi J, Malik J (2000) Normalized cuts and image segmentation. Pattern Anal Mach Intell IEEE Trans 22(8):888–905

Shi Y, Eberhart R (1998) A modified particle swarm optimizer. In: Evolutionary computation proceedings, 1998. IEEE World congress on computational intelligence., The 1998 IEEE international conference on, IEEE, pp 69–73

Shi Y, Eberhart RC (2001) Fuzzy adaptive particle swarm optimization. In: Evolutionary Computation, 2001. Proceedings of the 2001 Congress on, IEEE, vol 1, pp 101–106

Silva NB, Tsang IR, Cavalcanti GD, Tsang IJ (2010) (2010) A graph-based friend recommendation system using genetic algorithm. Evolutionary Computation (CEC). IEEE Congress on, IEEE, pp 1–7

Tasgin M, Herdagdelen A, Bingol H (2007) Community detection in complex networks using genetic algorithms. arXiv preprint arXiv:07110491

Wang T, Liao G (2014) A review of link prediction in social networks. In: Management of e-Commerce and e-Government (ICMeCG), 2014 international conference on, IEEE, pp 147–150

Wang XF, Chen G (2003) Complex networks: small-world, scale-free and beyond. Circuits Syst Mag IEEE 3(1):6–20

Wasserman S, Faust K (1994) Social network analysis: Methods and applications, vol 8. Cambridge University Press, Cambridge

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393(6684):440–442

Yang XS (2010) Nature-inspired metaheuristic algorithms. Luniver Press, Beckington

Yang XS, Cui Z, Xiao R, Gandomi AH, Karamanoglu M (2013) Swarm intelligence and bio-inspired computation: theory and applications. Newnes

Ying Yin, Hc YuhaiZhao (2020) Multi-objective evolutionary clustering for large-scale dynamic community detection. Information Sciences Elsevier

Zhou T, Lü L, Zhang YC (2009) Predicting missing links via local information. Eur Phys J B 71(4):623–630

Zhuang D, Chang JM, Li M (2017) Dynamo: Dynamic modularity-based community detection in evolving social networks. arXiv preprint arXiv:170908350