

Using The Gray Wolf Optimization Algorithm for Community Detection

Malihe Ghasemzade¹, Mohammad Amin Ghasemzadeh²

1- Department of Metallurgical Engineering, Karaj Branch, Islamic Azad University, Karaj, Iran.

2- Master of Electrical Engineering, CQUniversity, Melbourne, Australia.

ABSTRACT:

In today's world, networks play a very important role in people's lives. One of the important issues related to networks is the issue of detecting communities. These communities are also called groups and clusters. Communities include nodes that are closely related to each other. Most of the nodes that are members of a community have common properties. In social networks, it is important to detect the community in order to analyze the network and it is a very important tool to understand the information of the network and its structure. Studying community detection has garnered significant interest in last few years, leading to the development of numerous algorithms in this area. This research, we used the gray wolf meta-heuristic algorithm and improved it with operators such as mutation, combination, and local search, and also improved the final solution of the gray wolf algorithm with the label propagation algorithm to detect communities. Experiments showed that the proposed method has high accuracy and also due to the applied techniques, the problem converges to the best solution very quickly.

KEYWORDS: Community Detection, Gray Wolf Algorithm, Label Propagation Algorithm, Optimization.

1. INTRODUCTION

Networks play an important role in today's human societies, and since the data in many different fields can be naturally converted into graph structure, it can be claimed that this network is present in all fields. [1]. Many of the above systems and networks can be modeled as a network, where the links of the networks represent the relationships between the internal components of the system (nodes). In different domains, network links can represent different types of relationships such as human friendship, organizational structure, physical proximity of animals, infrastructure connections, web links [2]. Technological networks, including the Internet, electricity networks, telephone networks and road networks are an important part of daily life. Some of the popular networks include social media and online social networking sites like Facebook.

In general, one of the things that we see a lot in nature and has existed since the beginning of creation, is the existence of different communities (groups). A community, also known as a module or cluster, is a collection of nodes that are almost interconnected and have an inherently network-specific structure. Identifying nodes that belong to the same community frequently exhibit similar characteristics, such as sharing common functions, interests, or goals. Therefore, the detection of communities is a significant consideration in network analysis [3]. Therefore, the study of various aspects of these networks has been taken into consideration by many researchers. An important topic in the analysis of social networks is the exploration of group identification. A significant aspect of social networks involves understanding their formation. This has led to the emergence of clusters at the graph level. A cluster is a collection of vertices with a higher number of internal edges than external edges linking these vertices to vertices in other clusters. [4].

So far, numerous algorithms have been suggested for identifying communities, and these algorithms can be broadly categorized as either general or local methods. [5]. General methods have high complexity [6]. Hence, the local group detection methods have been developed as a solution for the unavailability of social networks and the complexity of calculations. Local algorithms are less accurate than general algorithms due to the use of local information. Also, getting stuck in local minima (maxima) is one of the most important problems faced by this group of algorithms, and for this reason, these algorithms often do not recover all the vertices of a group and stop after recovering a limited number of vertices. Therefore, having an efficient method to overcome these shortcomings is strongly felt.

In this research, our goal is to present a meta-heuristic algorithm method and improve it with mutation and combination operators and local search and integrate it with the label propagation algorithm with the aim of faster convergence and better performance to detect communities in social networks.

2. LITERATURE

Studying community detection has garnered significant interest in last few years, leading to the development of numerous algorithms in this area. The first algorithm presented in community detection was the GN algorithm proposed by Girvan and Newman, which was based on the idea of graph partitioning. In this algorithm, it uses the concept of connection centrality to find adjacent edges. Despite the improvements, the algorithm has a high time complexity. [7]

Clauset et al introduced a fast greedy algorithm that works with modularity. [8] At first, each node is assigned to a cluster. Then, the amount of modularity change ΔQ_{ij} in the combination of clusters is calculated. The combinations that have the highest increase are merged and become a new cluster. Modularity change is calculated only for clusters that have been affected. Once again, the combination with the highest value is merged. This process continues until all nodes become several large communities. This process is both memory efficient and fast in terms of time. This is because the modularity change is only updated for the affected clusters. The time complexity of the algorithm is close to linear. Zhao et al presented the CL Anet method, which is used for modularity maximization along with a local limiter using learning automata. [9]. An article by Panizo et al. [10] titled "Genetic Algorithm with Local Search Based on the label Propagation for detecting dynamic communities" was published. The idea used in this article is to combine the label propagation algorithm with the genetic algorithm. In 2022, an article [11] titled "An algorithm based on gray wolf optimization and balanced modularity for community detection in social networks" was published by Ehsan Jokar et al. In this paper, to improve the gray wolf algorithm, they used the label propagation algorithm to generate the initial population and applied the local search operator to the best solution of the algorithm. Fatemeh Besharatnia et al. [12] used the gray wolf algorithm to detect communities in static networks. In their method, they used the modularity criterion as the objective function. The comparison of the results of their method with other famous algorithms showed that their method has good performance and accuracy.

Raghavan [13] was the first one who used the label propagation algorithm for the problem of community detection. The purpose of this algorithm is to divide the network without knowing the size and number of communities. The steps of the standard algorithm are as follows: First, all the nodes are given a unique initial label. Then a random visit list is generated for all nodes. The label of each node is updated according to the label of neighboring nodes. The tag that has the highest number of repetitions in the neighbors is tied to it, and if there are several tags with the same number of repetitions, the tag is randomly selected from among them. This operation of updating labels will continue until the label of each node is equal to the label of most of its neighbors. Finally, the nodes that have the same label are placed in a community. The time complexity of the label propagation algorithm is close to linear, and hence it is a suitable candidate for detecting communities in social networks.

3. RESEARCH METHODOLOGY

Our goal in this research is to introduce a new model for detecting communities using the meta-heuristic algorithm of the gray wolf optimizer in social networks and improve it with mutation and combination operators and local search and combine it with the label propagation algorithm that has linear time complexity [14] and is one of the famous and fast algorithms in the issue of community detection with the aim of detecting common nodes. In the following, the performance of this algorithm has been evaluated by applying it to the problem of community detection and evaluating the quality of communities.

4. GRAY WOLVES METAHEURISTIC ALGORITHM

Inspired by the social life and hunting of gray wolves, The gray wolf algorithm employs four categories of wolves to model hierarchical leadership patterns. Gray wolves demonstrate a highly structured social hierarchy, with the alpha pair consisting of one male and one female serving as leaders. The alpha is primarily tasked with making decisions related to hunting, resting locations, and wake-up times, and their directives are followed by the rest of the group. However, instances of democratic behavior have been noted, wherein the alpha defers to other pack members. During group assemblies, the alpha's status is affirmed as the rest of the pack holds its tail down in recognition. Because the group is required to comply with the alpha's commands, only alpha wolves have the privilege of selecting a mate within the pack. It is intriguing to note that the alpha is not always the most physically dominant group member, but rather excels in group management. This serves to demonstrate that the organization and discipline within a group hold greater significance than sheer power. [15, 18].

The gray wolf hierarchy's second tier is occupied by the beta, who serves as the advisor to the alpha and organizes the group's activities. The beta then ensures the implementation of the alpha's directives across the group and provides feedback to the alpha. Within the gray wolf structure, the omega holds the lowest rank and assumes the role of the victim, being the last among the wolves to partake in eating. Any wolf not classified as an alpha, beta, or omega is referred to as a subordinate (or delta in certain sources). The delta wolf is required to report to the alpha and beta, but holds dominance over the omega.

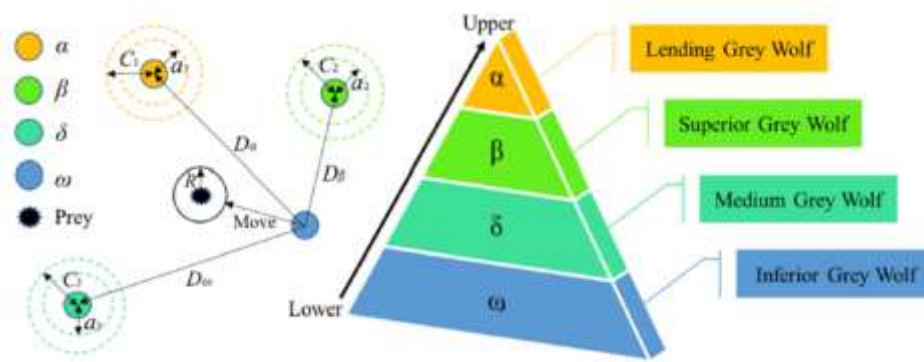


Fig. 1. Schematic diagram of GWO. [19].

In addition to the social hierarchy, gray wolf hunting has three stages: tracking, chasing and approaching the prey. To model the social hierarchy of wolves, we consider alpha as the best answer and beta and delta as the second and third among the best solutions. We consider the rest of the candidate solutions to be Omega. Optimization is driven by alpha, beta and delta and the fourth group follows these three groups. The modeling of wolves' siege behavior uses relations 1 and 2:

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)| \tag{1}$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \tag{2}$$

Where, t is the current iteration number, A and C are coefficient vectors, X_p is the position vector of the prey and X is the position vector of a wolf. To calculate vectors A and C , relations 3 and 4 are used.

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \tag{3}$$

$$\vec{C} = 2 \cdot \vec{r}_2 \tag{4}$$

The vector a decreases from 2 to 0 linearly during the iteration period in both exploration and exploitation phases. R is a random vector between 0 and 1. Due to the randomness of the vectors r_1 and r_2 , wolves can change their position in the space containing the prey randomly and using relations 5 and 6. This idea can be expanded to apply to an n -dimensional search space, where the gray wolves navigate around the optimal solution found in a greater number of dimensions compared to the dimensions of a cube.

Alpha wolves typically lead gray wolf hunts, with occasional participation from beta and delta wolves. To emulate this behavior, we store the three best solutions acquired and compel other search agents to adjust their position based on the location of the top search agents according to equation 7.

In this algorithm, the implementation of the exploit or attack phase, which happens when the prey is stopped, is done by reducing the value of the variable a from 2 to 1. The value of A is also dependent on a , so it decreases.

$$\begin{aligned} \vec{D}_\alpha &= |\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}| \\ \vec{D}_\beta &= |\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}| \\ \vec{D}_\delta &= |\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}| \end{aligned} \tag{5}$$

$$\begin{aligned} \vec{X}_1 &= \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha \\ \vec{X}_2 &= \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta \\ \vec{X}_3 &= \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta \end{aligned} \tag{6}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{7}$$

As the value of A decreases, wolves are forced to attack prey. An identification phase is also provided to avoid getting stuck in the local minimum trap. Wolves move away from each other to search for prey and come close to each other to attack and cooperate. To simulate this divergence, we use vector A with random values larger than 1 or smaller than -1 as shown in Fig. 2.

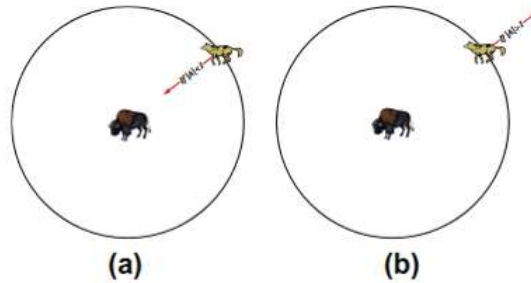


Fig 2. Attacking prey versus searching for prey.

Another influencing factor is the process of identifying the C value. The value of this random numerical vector is in the interval of [0, 2] and this random value intensifies ($C > 1$) or weakens ($C < 1$) the influence of the position of the prey in determining the distance. This vector can also be considered as the effect of obstacles that prevent approaching the prey in nature.

```

Initialize the grey wolf population  $X_i$  ( $i = 1, 2, \dots, n$ )
Initialize  $a$ ,  $A$ , and  $C$ 
Calculate the fitness of each search agent
 $X_\alpha$  = the best search agent
 $X_\beta$  = the second best search agent
 $X_\delta$  = the third best search agent
while ( $t < \text{Max number of iterations}$ )
  for each search agent
    Update the position of the current search agent by equation (3.7)
  end for
  Update  $a$ ,  $A$ , and  $C$ 
  Calculate the fitness of all search agents
  Update  $X_\alpha$ ,  $X_\beta$ , and  $X_\delta$ 
   $t = t + 1$ 
end while
return  $X_\alpha$ 

```

Fig. 3. Pseudo code of the GWO algorithm.

5. THE PROPOSED ALGORITHM

Detecting communities is very important and vital in extracting the network format. A large number of proposed algorithms are known to date, one of their most important problems is the scalability of these types of algorithms. In this research, gray wolf meta-heuristic algorithm has been used along with applying techniques such as: mutation, combination and local search operators and its combination with label propagation algorithm, so that our proposed method converges to the best possible solution with better speed and accuracy than other algorithms.

6. EVALUATION CRITERIA

The quality of communities obtained by the algorithm is obtained using the modularity criterion provided by Girvan-Newman.

$$Q(C) = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j)$$

Where A is the adjacency matrix of the network, k_i represents the degree of node i , and m is the number of edges in the network. The function δ has a value of one for two vertices inside a community and zero otherwise. If the number

of outcluster edges is as many random graphs, then Q will be zero. Q values close to 1 indicate a strong community structure. In practice, this value is between 0.3 and 0.7 for the structure of strong communities. [16,20].

7. DATASET

The specifications of the dataset used in this research are as follows.

Table 1. Test network specifications

Oriented	The number of communities	The number of edges	The number of nodes	Dataset
No	۲	۱۵۹	۶۲	Dolphin Social Network

8. DOLPHIN BOTTLENOSE NETWORK

The Bottleneck Dolphins network consists of 62 nodes and 159 edges that reflect the social behavior between dolphins. The network was created by David Lusseau, a biologist who spent 7 years scrutinizing dolphin behavior and mining data for the network. In this way, if a connection is established between two dolphins, a connection is created. The figure below shows the basic template of the Dolphin Network community [17].

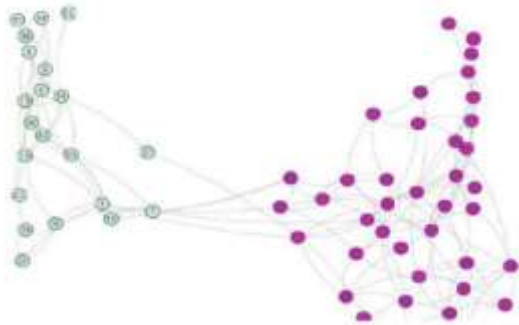


Fig. 4. Dolphin network graph.

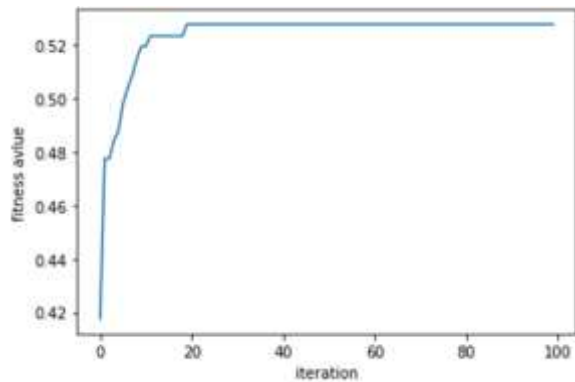


Fig. 6. Algorithm test result.

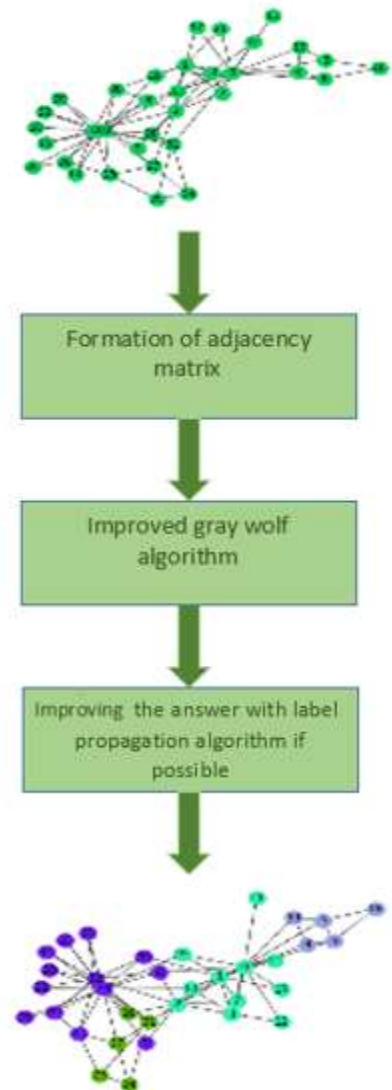


Fig. 5. Summary of the proposed method.

9. RESEARCH FINDINGS

In this article, by using the gray wolf meta-heuristic algorithm and applying several techniques on it to improve the algorithm, we were able to achieve better performance and modularity than other methods.

In the implementation of our proposed method (GWO-LP), we created the gray wolf algorithm along with the changes we considered to improve it in the form of a function called GWO. Before calling the GWO function, it is necessary to create the adjacency matrix of the desired graph in the form of a two-dimensional matrix. After creating the adjacency matrix of the desired data set, it is enough to call the GWO function to perform the community detection operation. After the gray wolf algorithm is fully implemented, we apply the label propagation algorithm to the labels obtained by the gray wolf algorithm, with the aim of increasing the modularity.

In this step, we traverse all the nodes of the graph by the label propagation algorithm. In the label propagation algorithm, we relabel nodes that overlap, that is, nodes that can belong to multiple communities, if possible, based on the importance of neighboring nodes. After changing the labels by the label propagation algorithm, if we reach a higher modularity, we accept the labels of the label propagation algorithm as the final solution.

The results obtained from running the algorithm on the Dolphin dataset are reported in the table below. Also, the images of the convergence diagram and the discovered communities are specified below.

Table 2. Results of running the algorithm on the dolphin dataset

Modularity	The number of gray wolves	The number of repetitions	Dataset
0.5277	80	100	dolphin

Table 3. Test results

GWO-LP	BGLL	MODPSO	CNM	Ga-net	NLPSO-D	Newman	
0.5277	0.526	0.5268	0.4950	0.4946	0.5144	0.465	Dolphins

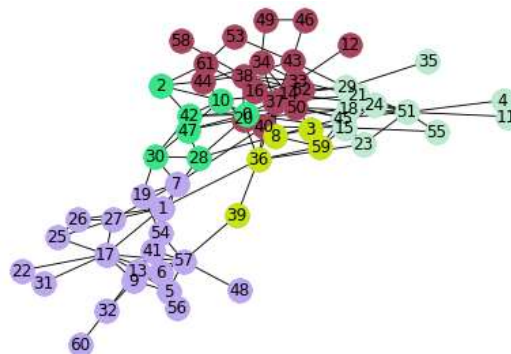


Fig. 7. The structure of communities detected in Dolphin graph.

10. CONCLUSION

So far, many algorithms have been presented for the topic of community detection, most of them have problems such as unstable results, lack of scalability, long execution time, etc. In this research, we proposed an improved gray wolf algorithm for community detection. We also used the label propagation algorithm, which is considered a fast algorithm, to improve the solution obtained by the gray wolf algorithm. By examining the results, we have come to the conclusion that the presented model achieves higher modularity than other meta-heuristic algorithms. Also, due to the use of existing edges, between graph nodes in the construction of the initial population and the use of various operators such as mutation, combination and local search, etc., the algorithm has faster convergence than other methods.

REFERENCES

- [1] Newman, M.E., “**The structure and function of complex networks**”. *SIAM review*, 2003. 45(2): p. 167-256.
- [2] Ellison, N.B., “**Social network sites: Definition, history, and scholarship**”. *Journal of Computer-Mediated Communication*, 2007. 13(1): p. 210-230.
- [3] Xu, T., et al., “**Generative models for evolutionary clustering**”. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2012. 6(2): p. 7.
- [4] Barber, M.J., “**Modularity and community detection in bipartite networks**”. *Physical Review E*, 2007. 76(6): p. 066102.
- [5] Plantié, M. and M. Crampes, “**Survey on social community detection**”, in *Social media retrieval*. 2013, Springer. p. 65-85.
- [6] Bagrow, J.P. and E.M. Bollt, “**Local method for detecting communities**”. *Physical Review E*, 2005. 72(4): p. 046108.
- [7] Girvan, M. and M.E. Newman, “**Community structure in social and biological networks**”. *Proceedings of the national academy of sciences*, 2002. 99(12): p. 7821-7826.
- [8] Clauset, A., M.E. Newman, and C. Moore, “**Finding community structure in very large networks**”. *Physical review E*, 2004. 70(6): p. 066111.
- [9] Zhao, Y., Jiang, W., Li, S., Ma, Y., Su, G., & Lin, X. (2015). “**A cellular learning automata based algorithm for detecting community structure in complex networks**”. *Neurocomputing*, 151, 1216- 1226
- [10] Panizo, A., G. Bello-Orgaz, and D. Camacho. “**A genetic algorithm with local search based on label propagation for detecting dynamic communities**”. in *Intelligent Distributed Computing XII*. 2018. Springer.
- [11] Jokar, E., M. Mosleh, and M. Kheyrandish, “**GWBM: an algorithm based on grey wolf optimization and balanced modularity for community discovery in social networks**”. *The Journal of Supercomputing*, 2022. 78(5): p. 7354-737.
- [12] F. Basharatnia, A. Talebpour, and S. Ali Akbari, “**Community detection in static social networks using gray wolf optimization algorithm**”. *Bi-quarterly Journal of Soft Engineering and Computer Engineering/Vol. 4, No. 1, Spring and Summer 1394*.
- [13] Raghavan, U.N., R. Albert, and S. Kumara, “**Near linear time algorithm to detect community structures in large-scale networks**”. *Physical review E*, 2007. 76(3): p. 036106.
- [14] Zhu, X. and Z. Ghahramani, “**Learning from labeled and unlabeled data with label propagation**”, 2002.
- [15] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “**Grey wolf optimizer**,” *Advances in engineering software*, vol. 69, pp. 46-61, 2014.
- [16] Newman, M.E. and M. Girvan, “**Finding and evaluating community structure in networks**”. *Physical review E*, 2004. 69(2): p. 026113.
- [17] Lusseau D. “**The emergent properties of a dolphin social network**”[J]. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 2003, 270(Suppl 2): S186- S188.
- [18] A. Mohammadzadeh, et al., “**Presenting an improved gray wolf optimization algorithm for Workflow scheduling in cloud computing environment**”, *Scientific Research Journal of Soft Computing and Information Technology*, Vol. 8, No. 4, 2018
- [19] Jian Zhou, Yuxin Chen, “**Hybridizing five neural-metaheuristic paradigms to predict the pillar stress in bord and pillar method**”, *Frontiers in PublicHealth*, 2023,
- [20] Maliheh Ghasemzadeh, Mohsen Ashourian, “**Identifying Communities on Static Social Networks**”, *Majlesi Journal of Telecommunication Devices*, Vol. 8, No. 3, September 2019.