A Letters Journal Exploring the Frontiers of Physics

July 2021

EPL, **135** (2021) 18001 www.epljournal.org
doi: 10.1209/0295-5075/135/18001

Multi-objective optimization for community detection in multilayer networks

Shihong Jiang[1], Xianghua Li[2], Xuejiao Chen[1], Zhen Wang[2(a)], Matjaž Perc[3,4,5] and Chao Gao[1,2(b)]

[1] *College of Computer and Information Science, Southwest University - Chongqing 400715, China*
[2] *School of Mechanical Engineering and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University - Xi'an 710072, Shaanxi, China*
[3] *Faculty of Natural Sciences and Mathematics, University of Maribor - Koroška cesta 160, 2000 Maribor, Slovenia*
[4] *Department of Medical Research, China Medical University Hospital, China Medical University Taichung 404332, Taiwan*
[5] *Alma Mater Europaea ECM - Slovenska ulica 17, 2000 Maribor, Slovenia*

received 10 April 2021; accepted in final form 10 June 2021
published online 6 September 2021

Abstract – Community detection in multilayer networks plays a key role in revealing the multiple aspects of information spreading and in comprehending the relationships and interactions within and between each layer. However, most existing algorithms are prone to local optimality, and they are also difficult to extend to high-dimensional networks. To address these challenges, we propose here a multi-objective algorithm for community detection that is based on the genetic algorithm. In particular, the modularity is introduced to optimize each network layer iteratively, and the local search is combined with genetic operations to overcome local optimality. Comparative benchmarks with other algorithms on artificial and real-world networks show that the proposed algorithm performs better, especially on high-dimensional networks.

Introduction. – Complex networks are a concise and valid methodology in representing the relevant structure among interacting units of a real complex system [1–4]. Community structure is a classic feature of networks [5], which can effectively reflect some typical characteristics, such as the involving regularity of network structures [6–8]. Although community detection algorithms of single-layer networks have made great achievements, an object in the real world tends to exhibit multi-dimensional attributes. For example, each person has multiple social accounts and can interact with each other by various social media platforms [9]. Each layer of a network depicts the connections among objects, and can reflect some properties of objects in different dimensions. Although each layer has its own unique community structure in multilayer networks, it cannot represent the composite structures of the whole network. How to use the complementary information provided by different layers to obtain the community structure of multilayer networks is one of the focuses of the current multilayer networks analysis, which has practical significance for a more real and more accurate understanding of the multi-dimensional structure of complex systems [10–14].

In recent years, due to the need for practical application, the research on multilayer networks community detection algorithms has attracted more attention [15]. The most intuitional methods are extended from the single-layer methods. Ma *et al.* divide these approaches into two categories, *i.e.*, single-analysis–based methods and multi-analysis–based methods [16]. The former one collapses the network into a single-layer network and the single-layer algorithms are used to explore the community division. The second kind of methods applies the single-layer method to every layer and mines the final community by using the consensus clustering approach. However, one shortcoming stands out according to which they fail to preserve the information of networks and ignore the interactions between different layers. To overcome the above problems, the multilayer network community detection is transformed into a multi-objective optimization problem.

In the multi-objective optimization method, multiple objectives are optimized simultaneously to ensure the integrity of network information [12]. However, how to select

(a) E-mail: zhenwang0@gmail.com
(b) E-mail: cgao@swu.edu.cn (corresponding author)

the optimal solution which adapts to different network structure from the *Pareto Sets* still remains a problem. Based on this, a novel GA-based multi-objective algorithm based on the NSGA_II multi-objective optimize framework, denoted as NSGAMOF, is proposed for detecting the multilayer networks community structure in this paper. The proposed method balances each layer to preserve the information of the multilayer network and different selection strategies are proposed to adapt to various networks.

Related work. – This section introduces formulations and algorithms of multilayer community detection.

Formulations of multilayer networks. A multilayer network is abstracted to $G = (\{G_1, \ldots, G_l\}, R)$, $G_l = (V_l, E_l)$ stands for a network at the l-th layer where V_l represents nodes and E_l the inter-layer links. $R = \{E_{ij} \subset V_i \times V_j, i, j \in 1, \ldots, l, i \neq j\}$ indicates connections among nodes of layer G_i and G_j. The elements of R represent inter-layer connections or crossed layer connections. That is, l subnetworks and the connections among these l networks make up a multilayer network together.

There are few indexes to directly assess the quality of the multilayer networks compound community. At present, the composite modularity (Q'_c) [17], redundancy index (R_c) [18], normalized mutual information (NMI) [19] and adjusted rand index (ARI) [20] are commonly used as the evaluation indicators.

As defined in eq. (1), the composite modular Q'_c is used to estimate the community index. The greater Q'_c is, the better the quality of the multilayer network compound community performs. M denotes the amount of communication links and n indicates dimensions or communication layers of a network. A denotes the corresponding adjacent. d infers to the degree of a node. X_i indicates that node i belongs to community X. When $X_i = X_j$, $\delta(X_i, X_j) = 1$, 0 otherwise:

$$Q'_c = \frac{1}{2M} \sum_{i,j}^{n} \left(A'_{ij} - \frac{d_i \times d_j}{2M} \right) \delta(X_i, X_j). \tag{1}$$

The redundancy index R_c [18] denotes the calculated ratio of redundant connections of multilayer networks. The larger R_c is, the better the quality of the compound community is. Intuitively, a compound community should have links across multiple layers. $\|p\|$ represents the amount of communities. S'_i denotes a couple $\{c_1, c_2\}$ that can be connected at the lowest layer of the G community. If $\{c_1, c_2\} \in S'_i$, then $\beta(c_1, c_2, E_l) = 1$, otherwise $B(c_1, c_2, E_l) = 0$. The redundancy index is formulated in the following equation:

$$R_c = \frac{1}{t \times \|p\|} \sum_{G_\delta \in G} \sum_{\{c_1, c_2\} \in S'_i} \beta(c_1, c_2, E_l). \tag{2}$$

The normalized mutual information (NMI) can assess the comparability of the optimized community and the original one [19]. F' denotes a confusion matrix. F'_x (F'_y) denotes the amount of elements in the x-th row (or the y-th column) in F'. r_{c_1} (r_{c_2}) signifies the total amount of clustering in a partition c_1 (c_2). The range of NMI value is [0, 1]. The greater the NMI is, the higher the similarity between optimized and original networks is. Suppose that c_1 and c_2 are two partitions in a network, then the following equation is applied to calculate the value of $NMI(c_1, c_2)$:

$$NMI(c_1, c_2) = \frac{-2 \sum_{x=1}^{r_{c_1}} \sum_{y=1}^{r_{c_2}} F'_{xy} \log(F'_{xy} N / F'_{x.} F'_{.y})}{\sum_{x=1}^{r_{c_1}} F'_{x.} \log(F'_{x.}/N) + \sum_{y=1}^{r_{c_2}} F'_{.y} \log(F'_{.y}/N)}. \tag{3}$$

The adjusted Rand index (ARI) is used to calculate the similarity between real communities and clustering ones. e and h denote the number of node pairs. The former one denotes that located in the same community in the real partition (c_1) and the obtained one (c_2), the other in a disparate community in c_1 and c_2. f and g are for the quantity of node pairs. f denotes which located in the same community in c_1 and different community in c_2. g represents that located in the same community in c_2 and different community in c_1. The more similar the real partition and the obtained ones are, the greater the value of ARI is. ARI can be defined as follows [20]:

$$ARI = \frac{2(eh - fg)}{(e+f)(f+h) + (e+g)(g+h)}. \tag{4}$$

In this paper, the modularity is used as the optimization function, and NMI, R_c, ARI are used as evaluation indexes due to their capability of finding high-quality solutions in multilayer networks [11,21–24].

Multilayer networks community detection. Existing community detection algorithms for multiplex networks fall into four categories, *i.e.*, matrix-decomposition–based method, spectral-based method, information-based method, and modularity-based method.

Matrix-decomposition–based methods extract the community division by decomposing the matrix such as NMF [16,22], SNMF [11,25] and so on. Spectral-based methods compute the community division by employing eigendecomposition to Laplace matrices, such as MIMOSA [26], and SC-ML [27]. The performance of such kinds of algorithm stands out because they capture the global information across different layers. Information-diffusion–based methods integrate different layers of the multilayer network by employing the diffusion of networks [28]. For example, similarity network fusion (SNF) computes the fused matrix of all layers through a parallel interchanging diffusion process, and then explores the community division by employing the spectral clustering method to the fused matrix [21]. The generalized Louvain (GL), one of the most efficient modularity-based algorithms, achieves a great efficiency by optimizing their generalized modularity function [23]. However, GL suffers from great difficulties in mining the consensus community division of all layers.

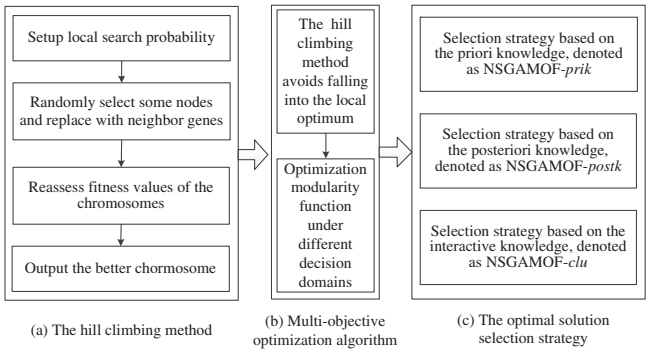


Fig. 1: The framework of NSGAMOF, which consists of (a) hill-climbing method, (b) the main body of multi-objective optimization and (c) the optimal solution selection strategy.

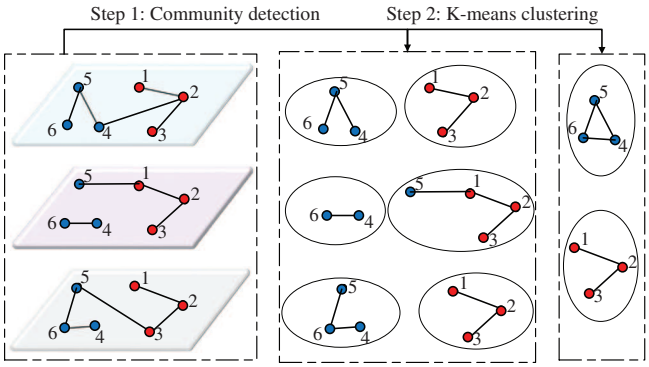


Fig. 2: An example of *NSGAMOF-clu* algorithm with three-layer networks and six nodes in each layer. The community structure in each layer is searched in Step 1 and the compound structure is obtained through k-means clustering in Step 2.

As for the methods mentioned above, they can handle almost all multilayer networks accurately and efficiently, but some drawbacks are unavoidable [11]. For example, the spectral-clustering–based method is good at extracting small-scale and tight communities, which may lead to some crucial information such as SC-ML being ignored. Many modularity-based algorithms are extended from the single-layer method that do not address the information across different layers. To overcome these problems, a novel multi-objective optimization algorithm is proposed to balance the community structure of every layer.

Formulation of *NSGAMOF*. – This section gives an elaborate explanation about the *NSGAMOF* optimization algorithm from the code scheme, genetic manipulation, local search and diverse optimal selection policy to prove the algorithm can preserve the information of networks and select the adaptive solution respectively. Figure 1 shows the framework of the *NSGAMOF* algorithm.

Code scheme. An appropriate code scheme of the solution can effectively reduce the computation and expedite the algorithm convergence. The label-based and locus-based representation play an important role in encoding methods for community detection. However, the label-based code scheme is redundant, which means that if there are t labels in the pattern, then $t!$ different chromosomes might be mapped to the same division [29]. In order to get the utmost output of the information contained in the pattern, this paper employs the locus-based adjacency presentation scheme. Suppose the solution of chromosomes in the population is set to $S = \{s_1, s_2, \ldots, s_N\}$. The length of the gene is N and each gene i can be arbitrary integer between 1 and N, namely, $1 < i < N$. The i-th gene value could be j, if i, j are linked at the lowest one layer in a network.

Genetic operators. The uniform crossover is applied for genetic operation in this paper. In fact, the uniform crossover pertains to multi-point crossover category, and becomes an efficient operator in evolutionary algorithms. First, the binary mask chromosomes with equal number of nodes are generated at random. According to the mask, the corresponding gene bits are selected from two parent chromosomes to form new chromosome as offspring chromosomes. To be specific, if the mask is 0, the operator selects the gene of the first parent chromosome; otherwise, the operator selects the second parent chromosome.

The mutation operator in chromosomes integrates the correlative information of the layer nodes neighbourhood, which makes random mutations in the form of the probability. The i-th gene in a chromosome is selected at random in the form of predefined probability, and then this gene mutates into the j-th neighbour of the i-th gene.

Local search operation. The NSGAMOF algorithm includes local search operations, namely the hill-climbing (HC) method. The first step is to define the neighbours of a chromosome. Given a chromosome $S_k = \{S_k^1, S_k^2, \ldots, S_k^n\}$, node S_k^i is selected from chromosome S_k at random. Then, the gene S_k^i is a substitute for other neighbour nodes S_k^j of the location i, and $S_k^j \neq S_k^i$, denoted by S_k'. The newly generated chromosome S_k' is used as a neighbour of chromosome S_k. In local search operations, a chromosome is chosen to be refined at random. Meanwhile, all possible neighbour chromosomes are identified. If the newly generated chromosome is better than the original one, we replace it with the new chromosome.

Optimal selection strategy. In MOOP (multi-objective optimization), a set of optimal solution sets will be obtained, which signify the optimal trade-off among optimization objectives. In order to obtain the final solution of *PS* in MOOP, we use the disparate optimal solution selection policy. There are three types of multi-criterion decision making: 1) *a posteriori-based*, 2) *a priori-based*, and 3) *interactive-based* [30].

The *a priori-based* policy leverages the value maximization of the objective function (*i.e.*, the modularity maximum) which is the optimal solution in the Pareto set. This policy using prior knowledge is called *NSGAMOF-prik*.

Different from the *a priori-based* policy, the *a posteriori-based* strategy takes the overall information of the network

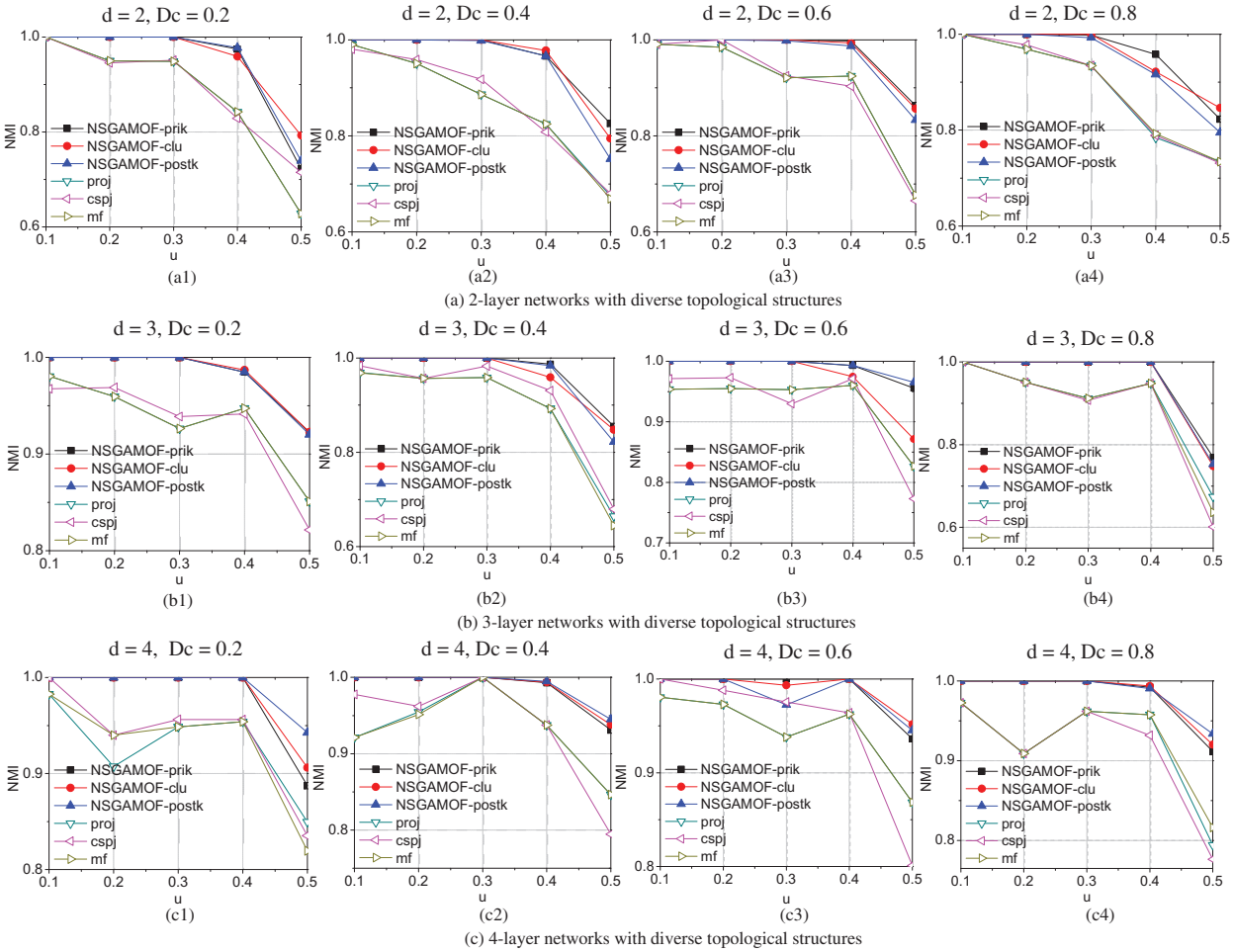


Fig. 3: NMI results with the changes of network topologies (u and Dc codetermine the network structure) and the layers increasing for mLFR networks. The results indicate that the network structure and total number of layers marginally affect the *NSGAMOF* (*i.e.*, *NSGAMOF-prik*, *NSGAMOF-clu*, *NSGAMOF-postk*). By comparison, the performance of proposed *NSGAMOF* is superior to that of *MLMaOP*c (*i.e.*, *proj*, *cspa*, *mf*).

into the consideration. It calculates the average modularity of each layer and the chromosome with the highest average value is the optimal solution. In this case, the method is called *NSGAMOF-postk*.

The last one is the *interactive-based* strategy, which mainly uses the k-means clustering method to classify the collection of data got by disparate models and enhance the quality of community detection. Such a strategy which selects the fittest solution by using the clustering method in the Pareto solution is called *NSGAMOF-clu*. Figure 2 shows an example of *NSGAMOF-clu* with three-layer networks and six nodes per layer.

The optimal selection strategy mentioned above can adapt to different network structures, which ensures a high performance. For example, the *prik-based* method adapts to networks which have uneven information distribution for each layer (*i.e.*, most of the information exists in one layer). However, for the network with the uniform information distribution, the *prik-based* approach may lose some information but the *postk-based* strategy can address the problem. In general, the proposed strategies can improve the accuracy for multilayer networks with various structures.

Experiments. – This section compares *NSGAMOF* with the other advanced methods in both real and synthetic multilayer datasets. The result is the average value got by executing the method for 100 times. The population size and iteration number are set as 200 and 100. We set crossover and mutation as 0.8 and 0.2, respectively.

Datasets. The synthetic network m-LFR128 function has 128 nodes in each layer [31]. The actual partition structure is known. To change parameters allows to control the total edges among communities and the difference of node degrees among layers. Therefore, 2, 3 and 4 layers of the network are generated respectively and these networks have different network topologies. The mixture parameter u signifies the part of the connection between one node and every other node in a community. The quality of community detecting usually becomes more dissatisfying as u gets larger. The degree of each node in different

Table 1: The structural information of the real-world networks. The *Layers* and *Nodes* denote the number of layers and nodes of multilayer networks, respectively.

	KTS	CAS	CoRA	MPD	WBN	WTN	SND(o)	SND(s)
Layers	4	5	2	3	10	14	3	3
Nodes	39	61	1662	87	279	183	71	71

network layers is determined by *Dc*, that is, the degree change chance. The greater the parameter *Dc* is, the more diverse the nodes of various layers might be. Moreover, eight real-world network datasets are used in this paper, *i.e.*, *Kapferer Tailor Shop, KTS*[1], *Cs-Aarhus Social, CAS* [32], *Bibliographic Data, CoRA* [33], *Mobile Phone Data, MPD* [34], *Worm Brain Networks, WBN*[2], *Word Trade Networks, WTN* [35] and *Social Network Nata, SND* including two different ground truth divisions, *i.e.*, *SND(o)* and *SND(s)* [36] (see table 1).

Experiments on synthetic networks. Figure 3 compares three strategies of *NSGAMOF* with the *MLMaOP* algorithm (*i.e.*, *proj*, *cspa*, *mf*) [14] in terms of *NMI* based on different parameters (*i.e.*, *d*, *Dc* and *u*) in 12 network structures from mLFR dataset. Results show that each strategy the *NSGAMOF* algorithm performs better than all strategies in *MLMaOP* algorithm. Moreover, the increasing number of layers scarcely influences the performance of *NSGAMOF* algorithms. Although the parameter *Dc* affects the network architecture, *NSGAMOF* can still find out the optimal division beneath diverse network architectures. We can conclude that *NSGAMOF* algorithms (*i.e.*, *NSGAMOF-prik*, *NSGAMOF-clu*, *NSGAMOF-postk*) perform better than *MLMaOP* algorithms (*i.e.*, *proj*, *cspa*, *mf*) in the diverse optimal selection policy, network architecture and network layers.

Experiments on real-world networks. Comparative results of *NSGAMOF-prik*, *NSGAMOF-clu*, *NSGAMOF-postk*, BGLL [37], the MOEA-MultiNet algorithms [11] in the *KAPFERER TAILOR SHOP* and *CS-AARHUS SOCIAL* multilayer networks are shown in table 2. More specifically, L_1, L_2, L_3 and L_4 of the one-layer strategies show that the simplified composite modular Q'_c and redundancy *Rc* are calculated by utilizing each community returned by BGLL algorithms in each layer of the networks. MOEA-MultiNet and NSGAMOF of multilayer strategies calculate the corresponding simplified composite modular Q'_c and redundancy *Rc*.

Results show that the multilayer strategy is superior to the one-layer strategy in solving multilayer network community detection. And the proposed *NSGAMOF-postk* algorithms of multilayer strategy outperforms obviously the several other algorithms, especially in simplified composite modular metric. In general, we can conclude that the compound community structure acquired by the proposed algorithms *NSGAMOF-prik*, *NSGAMOF-clu*, *NSGAMOF-postk*) have better performance than the BGLLs [37] and the MOEA-MultiNet algorithm [11] in the single layer.

Table 2: The Q'_c and R_c metric comparative results.

Dataset	Strategy	Algorithm	Q'_c	R_c
KAPFERER TAILOR SHOP NETWORK	One-layer	BGLL/L_1	0.2179	0.3964
		BGLL/L_2	0.2006	0.4717
		BGLL/L_3	0.1380	0.2657
		BGLL/L_4	0.0932	0.4094
	Multilayer	MOEA-MultiNet	0.2094	0.4735
		NSGAMOF-prik	0.4343	0.3705
		NSGAMOF-clu	0.4698	0.3511
		NSGAMOF-postk	0.4810	0.3134
CS AARHUS NETWORK	One-layer	BGLL/L_1	0.4685	0.2852
		BGLL/L_2	0.1672	0.0472
		BGLL/L_3	0.0832	0.1205
		BGLL/L_4	0.2893	0.1611
		BGLL/L_5	0.4115	0.2715
	Multilayer	MOEA-MultiNet	0.4010	0.3186
		NSGAMOF-prik	0.2316	0.3703
		NSGAMOF-clu	0.2315	0.3617
		NSGAMOF-postk	0.2287	0.3847

Tables 3 and 4 further compare the proposed strategies of NSGAMOF with other algorithms (*i.e.*, CSNMF, CP-NMF, CSNMTF, CSsNMTF [11], SNF [21], SC-ML [27], CGC [22], GL [23] and Infomap [38]) on five different real multilayer datasets in terms of *NMI* and *ARI*, respectively. Obviously, at any rate one of the presented methods is superior to other algorithms, especially in *NMI* metrics in most real-world datasets. And the *ARI* index value of the proposed algorithm has significant advantages over other algorithms. In addition, to validate the performance of the proposed algorithm roundly, the results on the HBN network, which contains 13281 nodes and 8 layers, are shown in table 5. The results denote that the proposed algorithm achieves an acceptable result.

Analysis of selection strategies. Although the ground truth represents consensus division of a kind, the information and importance of each layer are different. To identify the varying frequency of the different importance of each layer, three strategies are proposed to better adapt to different network structures. Here, two datasets (*i.e.*, MPD and SND(o)) are used to verify the validity of the proposed strategies. Figure 4 plots the *NMI* and *ARI* of each layer.

For the results shown in tables 3 and 4, the interactive-based strategy (postk) performs well when run on the SND(o) dataset and the *a priori*-based strategy acquires

[1] `http://vlado.fmf.uni-lj.si/pub/networks/data/UciNet/UciData.htm#kaptail`.

[2] `https://www.wormatlas.org/`.

Table 3: The *NMI* comparison results on real-world multilayer networks datasets. prik, clu and postk denote NSGAMOF-prik, NSGAMOF-clu and NSGAMOF-postk, respectively. The dash means that MIMOSA does not work on such dataset.

	CoRA	MPD	WBN	SND(o)	SND(s)	WTN
prik	0.716	0.554	0.372	0.578	0.221	0.266
clu	0.834	0.464	0.387	0.553	0.163	0.266
postk	0.575	0.468	0.363	0.691	0.058	0.291
CSNMF	0.514	0.504	0.463	0.681	0.053	0.322
CPNMF	0.480	0.451	0.432	0.685	0.053	0.172
CSNMTF	0.346	0.458	0.404	0.773	0.276	0.154
CSsNMTF	0.39	0.521	0.424	0.678	0.034	0.155
SNF	0.449	0.395	0.425	0.689	0.057	0.073
SC-ML	0.48	0.495	0.079	0.681	0.030	0.226
CGC	0.389	0.457	0.370	0.673	0.078	0.072
GL	0.418	0.467	0.398	0.618	0.097	0.426
Infomap	0.373	0.410	0.355	0.001	0.001	0.273
MIMOSA	0.011	0.096	0.020	0.132	0.046	–
S2-jNMF	0.796	0.516	0.072	0.582	0.037	0.157
Comclus	0.471	0.421	0.362	0.555	0.073	0.374
GMC	0.519	0.451	0.242	0.597	0.051	0.204

Table 4: *ARI* comparison results. prik, clu and postk denote NSGAMOF-prik, NSGAMOF-clu and NSGAMOF-postk, respectively. The dash means that MIMOSA does not work on such dataset.

	CoRA	MPD	WBN	SND(o)	SND(s)	WTN
prik	0.777	0.409	0.214	0.462	0.257	0.196
clu	0.879	0.373	0.246	0.435	0.177	0.174
postk	0.645	0.35	0.313	0.544	0.058	0.203
CSNMF	0.491	0.394	0.291	0.493	0.059	0.160
CPNMF	0.47	0.368	0.233	0.503	0.059	0.094
CSNMTF	0.279	0.346	0.237	0.811	0.234	0.035
CSsNMTF	0.288	0.422	0.225	0.484	0.031	0.088
SNF	0.47	0.28	0.211	0.515	0.058	0.005
SC-ML	0.485	0.379	0.001	0.493	0.021	0.133
CGC	0.296	0.357	0.211	0.472	0.092	0.002
GL	0.334	0.372	0.185	0.460	0.089	0.068
Infomap	0.016	0.115	0.200	0.012	0.001	0.093
MIMOSA	0.001	0.010	0.005	0.086	0.001	–
S2-jNMF	0.813	0.396	0.076	0.452	0.023	0.069
Comclus	0.447	0.365	0.251	0.481	0.086	0.269
GMC	0.426	0.248	0.012	0.428	0.034	0.015

a better solution when run on the MPD dataset. For SND(o), the best solution is extracted by choosing the maximum average of all layers since complementary information can improve the accuracy, but for MPD, the information of other layers reduces the accuracy so the prik strategy get the best solution. The experimental result demonstrates that the proposed strategies can improve the performance by adapting to various data structures.

Table 5: The results on large-scale datasets. The R_c and Q'_c are chosen as the metrics to evaluate the performance of the algorithm when running on the large-scale datasets.

	prik	clu	postk	Comclus
Q_m	0.124	0.128	0.087	0.1775
R_c	0.057	0.065	0.053	0.001

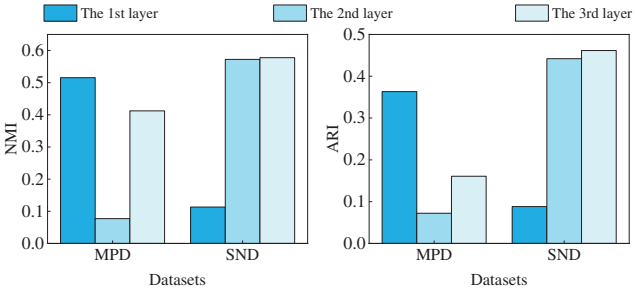


Fig. 4: The *NMI* and *ARI* of each layer. There exist relative differences in different layers for a multilayer network. Different strategies can adapt to different multilayer networks to improve the accuracy of algorithms.

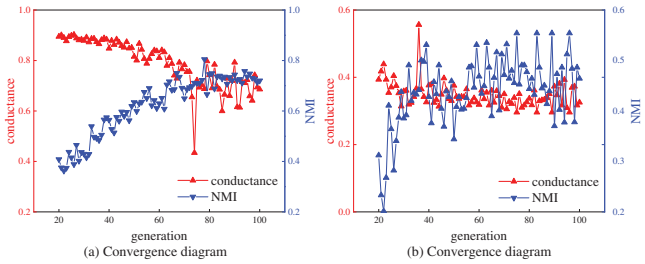
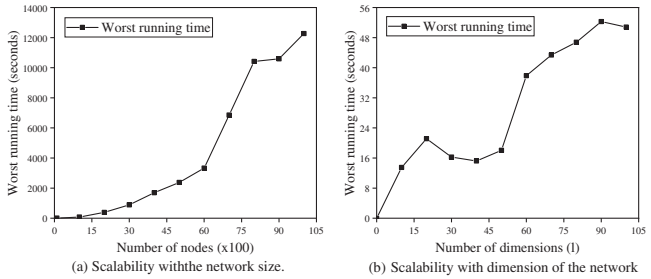


Fig. 5: The results of conductance and *NMI* running on (a) CoRA and (b) MPD. The red line denotes the conductance. The smaller the index value is, the better the performance is; the blue line denotes *NMI*, and the greater the value is, the better the performance is.

Convergence analysis. To demonstrate the convergence performance of *NSGAMOF*, *NMI* and conductance indices are calculated to observe the convergence of the *NSGAMOF* algorithm. As shown in fig. 5, the *NSGAMOF* algorithm converges slowly at the beginning, and then the convergence starts to accelerate after 55 iterations. Finally, the *NSGAMOF* algorithm stably converges within 90 times. In the whole operation process of the *NSGAMOF* algorithm, the curve of the *NMI* index generally rises and tends to be stable after some reasonable fluctuation, which proves that *NSGAMOF* algorithm has strong convergence.

Scalability analysis. Figure 6 shows the result of scalability as the network nodes and dimension increase. The experiments employ 10 diverse datasets with 2 dimensions. In fig. 6(a), the curve exhibits the characteristics of a quadratic equation as the networks increase in



(a) Scalability withthe network size. (b) Scalability with dimension of the network

Fig. 6: (a) With the growth of node numbers, the worst running time changes. It can reflect that the scalability of the algorithm varies with the network size. (b) With the growth of dimensions, the worst running time changes. The scalability of the algorithm varies with the dimension of networks.

size. Overall, it is reasonable for the execution time of the *NSGAMOF-prik* algorithm. Figure 6(b) depicts the worst running time of the proposed algorithm as the network dimension increases. According to this result, the algorithm presents the scalability for the high-dimensional network.

Conclusion. – In order to effectively balance the community structure of every layer to obtain a high-quality compound community, we transform the multilayer network community detection into a MOOP and present a new GA-based multi-objective optimization algorithm *NSGAMOF* for multilayer network community detection. The concept of modularity and local search are introduced to optimize every layer of a network iteratively. The diverse optimal selection policy is applied to ensure the relatively optimal compound community structure.

* * *

This work was supported by the National Key R&D Program of China (No. 2019YFB2102300), National Natural Science Foundation of China (Nos. 61976181, 11931015, U1803263), Fok Ying-Tong Education Foundation, China (No. 171105), Key Technology Research and Development Program of Science and Technology-Scientific and Technological Innovation Team of Shaanxi Province (No. 2020TD-013), and by the Slovenian Research Agency (Nos. P1-0403 and J1-2457).

Data availability statement: All data that support the findings of this study are included within the article (and any supplementary files).

REFERENCES

[1] Zhu P. C. *et al.*, *EPL*, **126** (2019) 48001.
[2] Li H. J. *et al.*, *New J. Phys.*, **22** (2020) 063035.
[3] Zhang Z. P. *et al.*, *IEEE J. Sel. Areas Commun.*, **38** (2020) 845.
[4] Xia C. Y. *et al.*, to be published in *IEEE Syst. J.*, DOI: 10.1109/JSYST. 2021.3065378 (2021).
[5] Li S. D. *et al.*, *Appl. Math. Comput.*, **401** (2021) 126012.
[6] Wang Z. *et al.*, *Sci. China: Inform. Sci.*, **63** (2020) 212205.
[7] Zitnik M. and Leskovec J., *Nat. Commun.*, **9** (2018) 1.
[8] Zhu P. *et al.*, *Chaos, Solitons Fractals*, **109** (2018) 231.
[9] Gao C. *et al.*, *Commun. ACM*, **62** (2019) 61.
[10] Zhu P. *et al.*, *Appl. Math. Comput.*, **359** (2019) 512.
[11] Gligorijevic V. *et al.*, *IEEE Trans. Pattern Anal. Mach. Intell.*, **41** (2018) 928.
[12] Chen X. J. *et al.*, *Proceedings of the 12th International Conference on Knowledge Science, Engineering and Management* (Springer) 2019, p. 197.
[13] Gao C. and Liu J. M., *IEEE Trans. Syst. Man Cybern.: Syst.*, **47** (2017) 171.
[14] Pizzuti C. and Socievole A., *Proceedings of the IEEE Congress on Evolutionary Computation* (IEEE) 2017, p. 411.
[15] Zhai X. M. *et al.*, *Sci. Rep.*, **8** (2018) 1.
[16] Ma X. K. *et al.*, *IEEE Trans. Knowl. Data Eng.*, **31** (2018) 273.
[17] Ma L. J. *et al.*, *Swarm Evol. Comput.*, **39** (2018) 177.
[18] Shahmoradi M. R. *et al.*, *Future Gener. Comput. Syst.*, **101** (2019) 221.
[19] Rossetti G. and Cazabet R., *ACM Comput. Surv.*, **51** (2018) 1.
[20] Schutze H. *et al.*, *Introduction to Information Retrieval* (Cambridge University Press, Cambridge) 2008.
[21] Wang B. *et al.*, *Nat. Methods*, **11** (2014) 333.
[22] Cheng W. *et al.*, *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM) 2013, p. 320.
[23] Mucha P. J. *et al.*, *Science*, **328** (2010) 876.
[24] Gao C. *et al.*, *Phys. Life Rev.*, **29** (2019) 1.
[25] Ni J. C. *et al.*, *IEEE Trans. Knowl. Data Eng.*, **30** (2017) 435.
[26] Chen P. Y. and Hero A. O., *IEEE Trans. Signal Inf. Process. Netw.*, **3** (2017) 553.
[27] Dong X. W. *et al.*, *IEEE Trans. Signal Process.*, **62** (2014) 905.
[28] Wang H. *et al.*, *IEEE Trans. Knowl. Data Eng.*, **32** (2019) 1116.
[29] Pizzuti C., *IEEE Trans. Evol. Comput.*, **22** (2017) 464.
[30] Purshouse R. C. *et al.*, *Proceedings of the IEEE Congress on Evolutionary Computation* (IEEE) 2014, p. 1147.
[31] Lancichinetti A. *et al.*, *Phys. Rev. E*, **78** (2008) 046110.
[32] Liu W. F. *et al.*, *Proceedings of the IEEE Congress on Evolutionary Computation* (IEEE) 2017, p. 443.
[33] Sen P. *et al.*, *AI Mag.*, **29** (2008) 93.
[34] Dong X. W. *et al.*, *IEEE Trans. Signal Process.*, **60** (2012) 5820.
[35] De Domenico M. *et al.*, *Nat. Commun.*, **6** (2015) 1.
[36] Snijders T. A. B. *et al.*, *Sociol. Methodol.*, **36** (2006) 99.
[37] Lancichinetti A. and Fortunato S., *Sci. Rep.*, **2** (2012) 1.
[38] De Domenico M. *et al.*, *Phys. Rev. X*, **5** (2015) 011027.