

Many real world complex systems that are available to complex network, such as social networking, biological networks, power network, Web network, etc. by means of two simple mapping, the mapping of the object to the network node to the object side of the mapping relationship between can be represented as a complex network graph model. complex network research is being attracted from physics, biology, sociology and computer science the attention of scholars in different fields in addition to well-known small-world and scale-free and other features, complex networks also has a very important module, namely implicit in complex networks rich community structure model according to the literature on online community description, can loosely be defined as having some common characteristics of interconnected information carrier collection For example, under a particular topic collection of Web pages, having some common interests microblogging group composed of micro, etc. From the point of view of network topology, an online community is a densely connected sub-network diagram figure, in this sub-graph of nodes connected between the internal node density is higher than sub-graph density of connections with external nodes. complex network research community discovery has been successfully applied to terrorist organizations such as the identification, protein function prediction, public opinion analysis and processing and many other areas in which [1].

Online Community discovery research is attracting wide attention from researchers of complex networks in recent years, the emergence of a number of methods, the literature [2] for a systematic analysis of these methods and gives a preliminary classification from data mining level, network discover the essence of community-based network link clustering, the goal is the network node set is divided into a number of internal links and external links closely sparse clusters from the clustering of learning perspective, the quality of online communities discovery algorithm is large extent depends on network community structure index for judging the quality of design ideas and optimization strategies. Currently, the network community discovery algorithm, the objective function (network community structure quality evaluation index) for solving optimization strategies can be broadly divided into two categories: basic heuristics heuristic methods and ultraformer complex network community discovery problem into predefined heuristic rules of design problems, according to various community characteristics index for judging the quality of design optimization strategies; latter using a variety of ultra-heuristic operator network community discovery problem space quality evaluation indicators for community optimization.

Community index for judging the quality of the basic heuristic methods can be divided into direct and indirect greedy greedy method. Direct greedy method of thinking is very simple, is to initialize the network of $|V|$ communities, iterative process until the algorithm terminates the following conditions are met: calculate the side with respect to the degree of modularity of the information gain, select index for judging the quality of the community's largest while adding incremental value, in order to achieve community merge directly representative algorithms greedy method is the module level index value Q optimization algorithm, the original Q optimization algorithm The time complexity is $O((m+n)n)$ or $O(n^2)$ [3]. To improve the efficiency and effectiveness of the algorithm, proposed a series of improved methods: Clauset, who designed the data structure max-heaps will algorithm time complexity is reduced to $O(m \log n)$, Danon, who proposed the Q value of the increment is normalized to discover the difference with the size of the community has a large community structures [4], Wakita et al proposed the use of combined ratio (consolidation ratio) the Q value of the incremental algorithm is weighted to improve scalability [5], Blondel et al proposed merger in the iterative process allows multiple communities rather than merging the two communities merge [6]. Indirect basic greedy method idea is: the entire network as a community, then loop the following procedure until the community quality evaluation index value Q satisfy the given conditions: choose a side and remove certain characteristics to achieve the online community discovery of the method of choice strategies include: Edge mediated neutral (betweenness) greater priority [7], while clustering coefficient (clustering coefficient) Small preferred [8], while the information center degrees (information centrality) greater priority [3] and edge stability factor (stability coefficient) greater priority [9] In addition to right edges greedy methods, Gan Wen Yan et al proposed a potential community-based topology discovery algorithm, the topology of each community as potential field local high potential areas, through potential values for local maxima greedy nodes to achieve network optimization community division [10].

The basic idea is simple and intuitive heuristic method, easy to implement, however, such methods require a priori knowledge by means of a recursive definition of termination condition, does not have the automatic identification of the total capacity of the network community, which largely restrict such optimization methods in the real complex network community discovery application.

In order to overcome the lack of basic heuristic methods, researchers have proposed a class is used to optimize the community index for judging the quality of super-heuristic methods, including optimization algorithm based on a single objective-based multi-objective optimization algorithm. Tasgin other people through

h the use of GA algorithm optimization community modularity Q function to achieve optimal partitioning of the network approximation [11]. Pizzuti given first for judging the quality of a community network is divided fraction (community score) is defined, and then optimize the use of GA-Net network segmentation [12] Taking into account the mass of social networks, Lipczak et al [13] proposed a community-based is sufficiently small and limited number of assumptions ACGA Community algorithms: a community encoded as an individual, based on the potential to improve the quality of community capacity to select individuals genetic manipulation. Duan Xiaodong, who introduced PSO network iterative two points of Web communities found [14]. CDPSO algorithm [15] using a node-based neighbor ordered list of particles encoding by PSO global search to mining communities. Gog et al proposed a mechanism for sharing information based on individual co-evolutionary algorithms for network community structure optimization [16], combined with the local search algorithm GA variant CCGA [17] and LGA [18] by optimizing the quality evaluation index community

Huang Faliang etc.: Based on the multi-objective optimization online community discovery methods Q to achieve large-scale complex network community discovery. Zhu and Wang proposed mining online community parallel genetic algorithm PGA [19]. 3

Despite these based on a single objective optimization algorithm [12? 19] has a good time efficiency and can dig out of the network to meet a specific target community structures, however, the practical application of the Internet community find problems often need to balance multiple objectives, and these goals may be conflicting. Obviously, a community based on a single objective optimization methods can not meet this discovery application requirements. Therefore, based on multi-objective optimization community discovery began to receive attention. GONG Mao et al proposed a fruit used for network community discovery methods based on mathematical programming and evolutionary algorithm combining multi-objective optimization algorithm, while the internal links and external links density density optimization [20]. multi-objective optimization algorithm (NNIA-Net [21], MOGA-Net [22], MOHSA [23] and SCAH-MOHSA [24]) are selected Community Rating (community score) and community fitness (community fitness) as the optimization objective to implement network community mining, the difference is used metaheuristic Methods: NNIA-Net using immune algorithm, MOGA-Net using GA algorithm, MOHSA adaptive hybrid multi-objective harmony search algorithm, SCAH-MOHSA in MOHSA add a spectral clustering on the basis of pre-processing operator.

With a single objective optimization algorithm compared to these multi-objective optimization algorithm can judge the quality of a variety of community indicators into consideration, you can find more higher-quality network community structures due to community-based multi-objective optimization discovery research is in its infancy there are some deficiencies, such as the current algorithms are hypothetical index for judging the quality of the network between the community may not be consistent, but did not give up on the assumption theoretical proof or experimental verification, there is no index for judging the quality of research networks in the community nature of the relationship between addition, the existing community-based multi-objective optimization methods found in almost all based on genetic algorithms (genetic algorithm, referred GA), the multi-objective particle swarm optimization algorithm for community discovery research reports have not yet seen, The studies show that multi-objective particle swarm optimization algorithm has excellent ability of global optimization [25].

Firstly, from the perspective of experimental verification of the online community quality evaluation index coupling relationship with the existence of data dependencies, which are derived for multi-objective optimization of the community discovery necessity, then the multi-objective optimization problems in the form of online communities depiction, and then propose a multi-objective particle swarm optimization based online community discovery algorithm MOCD-PSO, the method by simultaneously optimizing multiple online communities generate Pareto optimal quality evaluation index divided communities set, the user can choose according to specific needs most satisfying community structures. experiments show that both single-objective optimization method (GN and GA-Net), compared with the multi-objective optimization algorithm (MOGA-Net and SCAH-MOHSA) comparison, MOCD-PSO algorithm can not prior information to dig out under the conditions of higher quality online community.

Section 1 of this online community quality evaluation of existing indicators Introduction Part 2 analyzed experimentally derived index for judging the quality of community coupled with data-dependent relationship, and formalize two properties described in Section 3 to a network with multi-objective optimization in the form of community-related definitions - Part 4 MOCD-PSO algorithm gives a detailed description and performance analysis. Section 5 gives the real network data and artificial network data and compare the results of the test - Part 6 summary of the full text.

A related work

As the "community about what is" lack of strict uniform definition, which researchers from different application scenarios and theoretical angles, define a variety of different forms of community, and community-based loosely defined, you can put these kinds of communities can grouped into three broad categories: internal links with the largest density connected subgraph with minimal external links connected subgraph density, but also has the maximum density and minimum internal links external links connected subgraph density researchers to dig out from the network each define different communities, made a variety of online communities for the quality measure quantitative indicators in this section gives the formal definition of network community structures (Definition 1), based on the list of commonly used community quality evaluation index (Table 2), Table 1 is a variety of metrics related symbols more community quality evaluation index reference [26].

Currently, the most widely used indicator is the community judge the quality of Newman and Girvan's Q-value function [27], it is by comparing the network subgraphs corresponding zero random graph model to define the connection density, Q value is larger, the network The higher the quality the community, but there is the indicator of the "limited size" defect [28], Li et al [29] to improve the modularity density index Q Li. evaluation index MinMaxCut [30] attempts to maximize the community similar to the node communities while minimizing the degree of similarity between the nodes, MinMaxCut smaller the value, the higher the quality of the network community. evaluation index silhouette [31] is a reference to a graphic display of the pixels within the same cluster similar views to measure the value of online communities quality, the index of $a_i = \text{avg}_{j \in V_i} (A_{ij})$, $i \in V_i$ expressed within the community

$b_i = \frac{1}{|V_i|} \sum_{j \in V_i} A_{ij}$ represents the community V_i with other nodes within a node i and the other nodes in the community average V_i similarity, $b_i = \max_{k \in V} \sum_{j \in V_k} A_{kj} / |V_k|$

The maximum average similarity community, silhouette larger index value, the higher the quality of the network community.

Definition 1 (network community structures) Given an undirected network $G = (V, E)$, the set of network nodes V , the network edge set is $E = \{e = (u, v) \mid u \in V, v \in V\}$, G a

4 Journal of Software Software Technology

A size $|V| \times |V|$ of the matrix A to that if the edge $e = (i, j) \in E$, then $A_{ij} = 1$; otherwise, $A_{ij} = 0$. Internet community structure is the network node set V is an m division scheme $P = (V_1, V_2, \dots, V_m)$, wherein, V_i four conditions must be satisfied: $V_i \cap V_j = \emptyset$ ($i = 1, 2, \dots, m$), $\cup_{i=1}^m V_i = V$ and $V_i \cap V_j = \emptyset$ ($i \neq j$).

$i = 1, 2, \dots, m$

Table 1 Symbols and notations

Table 1 related symbols Symbol Remark

$d_i = \sum_{j \neq k, k \in V} A_{jk}$ Community V_i node j within the internal degrees

d_i Community V_i node j within the external degree

d_{ij} Within the community V_i V_k node j , the degree of association with the community

d_{in} Community V_i internal degrees

d_{out} Community V_i external degrees

d_{ij} Community V_i and V_j community correlation $d_{out} = \sum_{j \neq k, k \in V} A_{jk}$ $d_{in}(j, V_k) = \sum_{i \in V} A_{ji}$ $d_{in}(V_i) = \sum_{j \neq k, j \in V, k \in V} A_{jk}$ $d_{out}(V_i) = \sum_{j \neq k, j \in V, k \in V} A_{jk}$ $d_{in}(V_i, V_j) = \sum_{i' \in V, j' \in V} A_{i'j'}$

Table 2 Common metrics to measure network community goodness

Table 2 common quality evaluation index Metrics mFormula community

Q

QLiMinMaxCut

silhouette

ductance [32]

Expansion [33]

NCut [34]

$[23] m? Din(V)? Dout(V)? 2? Q = \sum? I?? I?? I = 1? D(V)? D(V)???? Inoutm? D(V_i) ? d(V_i)? QLi = \sum?? | V_i | i = 1?? MMC = \sum GS = dout(V_i) i = 1d(V_i) mlk? lai? bi??? \sum \sum j = 1? vi \in V_j? k | V | \max(b, a) ij i?? dout(V_i) \sum 2d(V) + d(V) i = 1iimdout(V_i) \sum | V | i = 1m \sum 2d_{in}(V) + d_{iout}(V) + 2(|E|? d_{in}(V)) i + dout(V) i = 1iimidout(V) dout(V) rcommunity score? kin???? 1j(V_i)? A?? \sum \sum jk??? | V | j \in V? | V |?? i = 1? j, k \in Viii???? i? k????$

community fitness [23] $\sum \sum (k_{in}(V) + k_{out}(V)) i = 1j \in V_{ij} j i k_{in} j (V_i)$

2 Community index for judging the quality of nature

Community quality evaluation index values are maximized optimal, also minimized the best, the two can be transformed into each other without loss of generality, this article discusses only maximize the optimal s

ituation. Below, we first experiment to explore the online community Quality evaluation index coupling relationship with the existence of data dependencies, and then gives the corresponding form definition.

Experiment 1. Inspection coupling index for judging the quality of the community association. Experimental data: Karate network; algorithm framework: CDPSO [15]; algorithm framework embedded community quality evaluation index: modularity Q, silhouette value of the GS.

Algorithm (CDPSO + Q) is a community-quality evaluation index modularity Q is a single target online community structure optimization algorithm, and algorithm (CDPSO + GS) is a community-quality evaluation index silhouette is a single target network optimization community structures from Figure 1 can be seen : in the number of iterations

[Star library wxphp.com](http://www.wxphp.com) contains summary report, file downloads, test data, teaching and research, office documents, professional literature, tutorials Raiders, IT computers, teaching materials, Humanities and Social Sciences, tourist attractions, as well as 86 network-based multi-objective optimization community discovery methods and so on.