Attribute Relationship Solving Method Based on Nodes and Communities in Opportunistic Social Networks

Jia WU (jiawu5110@163.com)
Central South University https://orcid.org/0000-0001-9013-0818

Xiaoming TIAN
Central South University

Research Article

Keywords: node, community, matrix, communication activities, social opportunistic networks

Posted Date: April 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-264362/v1

Attribute Relationship Solving Method Based on Nodes and Communities in Opportunistic Social Networks

Jia Wu[1*] and Xiaoming TIAN[2*]

1. School of Computer Science and engineering, Central South University, Changsha 410083, China;
Corresponding author: Jia Wu(Email: jiawu5110@163.com) and Xiaoming TIAN((Email: tianxiaoming1986@126.com)

Abstract: The penetration of the 5G Internet and big data communication into human society brings about the survival basis of the social opportunistic networks. Using mobile terminal devices for communication makes the communication of nodes in the social opportunistic network intermittent, because nodes may be in motion all the time. In social opportunistic networks, data communication activities can be recorded and analyzed by evaluating communication activities of human beings or determining their interest points. However, the identification of nodes with the same or similar types of attributes among a large number of user nodes, has become a research problem in the field of social opportunistic networks. How to find an effective method to classify nodes according to their social characteristics and similarity degree becomes the key point of social opportunistic network data forwarding process. In this study, we proposed a method of community mining by decomposition of node and community relationship matrix with large social network data attributes. By using the regular type and iterative community features among community-rule-meet nodes, the method is proved to be converged and yield a minimum solution. Experimental results show that the proposed method exhibits strong application value.

Keywords: node; community; matrix; communication activities; social opportunistic networks

---

1. Introduction

The explosion of mobile devices in recent years has fueled the development of high-speed communications and highly reliable networks [1-4]. With the popularization of communication devices, social interaction can be seen everywhere. People can share their interests anytime and anywhere through mobile phones, tablets, smart bracelets and other devices [13-15]. As a result, online social platforms like twitter and facebook have become an indispensable part of human life [5-8]. As human beings interact socially in life, information is stored in portable devices and connected and transmitted intermittently between devices along with human behaviors. Therefore, in the social opportunistic network, data transmission between nodes needs to find “opportunity”.

The information transfer process needs to look for opportunities meaning that only nodes considered reliable can participate in the communication. The “storage-carry-forward” mechanism is a transmission strategy of social opportunistic network [17-23]. The node stores the information to be transmitted in its own cache area, carries the information for movement, and does not send the information to the node until it meets the appropriate node. In the urban social scene, people with portable communication equipment represent nodes in the social opportunistic network, so the social attributes of these people will have a great impact on the data transmission strategy [25-32]. Data generated by human behaviors are of great significance for the selection and improvement of information transmission strategies, so they have become the research hotspot of social opportunistic networks.

However, a great number of online data that can be retrieved on the basis of human activities are complex and may require extensive calculations. Moreover, many mobile devices that carry the information may overload in big data online social opportunistic networks, they cannot receive or send any messages to others. This characteristic may affect traditional methods in wireless communication networks.

We face big data communication in social opportunistic networks, challenges for nodes are high delay, limited cache space, performing data update and improve delver ratio while appropriate neighbors can be selected by us [33-36]. How to evaluate transmission states between nodes and neighbors is very important. Data consume significant cache space and energy in devices when people use mobile devices during data transmission and no suitable transmission range target is responding, which eventually causes transmission delay [2]. Especially in big data social opportunistic network environment, where over-flooding and data redundancy are used to create transmission, devices must distribute considerable cache space to save messages [3]. A large number of awaiting information is stored in devices. Some information may be stored for a long time without user acceptance and response status.

To avoid over-consumption, the identification of nodes with the same or similar attributes among a large number of user nodes has become a research problem in the field of online social opportunistic networks. The resolve relationship matrix of node and community (RRMNC) method is established in this study. This method is used to conduct attribute decomposition of a large amount of social opportunistic network data by using regular and iterative in the community features of nodes with the minimum solutions after convergence. These nodes comply with community rules.

The main contributions of this study include the following:

(1) The rules for node iteration are established through the relationship matrix of nodes and communities.

(2) After demonstrating the convergence, the node that satisfies the minimum solution is identified. This node exhibits a strong correlation with the community.

(3) Numerous experiments show that the proposed method exhibits strong application value.

This paper is divided into five chapters.Chapter1 introduces the study. Chapter 2 presents the related works. Chapter 3 describes the system model. Chapter 4 indicates the experimental design, and Chapter 5 concludes the study.

2. Related work

In recent years, with the popularity of mobile devices, many researchers have invested in the study of opportunistic social networks. The research on opportunity social networks mainly focuses on routing strategies, making the opportunistic social network suitable for more application scenarios. Next, we will briefly introduce the current status of several methods related to the research of this paper.

According to the social attributes and mobile features of nodes, some researchers proposed community-based routing strategies. The communication between nodes is carried out through social relations, which has good transmission performance under specific application scenarios. Fu et al. [16] proposed a utility-oriented routing algorithm for community based opportunistic networks. The algorithm establishes a community model by combining the geographical location preference and time-variance behavior model of nodes. According to this model, messages are transmitted between nodes by comparing social relations and social degrees. Wu et al. [11] suggested a weight distribution and community reconstitution algorithm based on community communications. The research holds that the movement of nodes has regularity, showing repetitive and periodic changes. Nodes with the same social attributes have more contact opportunities. Therefore, the algorithm divides communities by the social attributes of nodes as the basis of messaging. Literature [17] proposed an effective data transmission strategy based on node socialization in opportunistic social networks. The method divides communities by social attributes of nodes, and then adopts the strategy of community reduction to remove inefficient nodes in the community, so as to improve the efficiency of data forwarding.

Park J et al. [18] proposed a forwarding scheme based on swarm intelligence and percolation centrality in opportunistic networks. This strategy establishes a routing scheme based on social relations of nodes by simulating the behavior characteristics of bees in artificial bee colonies. In this algorithm, a cell-based model is used to find out key nodes in the network, which can improve the transmission efficiency of messaging. However, the algorithm ignores that the node's cache is limited. A large number of transmission tasks will cause the network of key nodes to block. Zheng et al. [19] proposed an effective positive transmission routing algorithm based on social relationships in opportunistic social networks. By quantifying the trust degree and encounter strength, the strategy comprehensively evaluates the forwarding ability of nodes and then establishes a forwarding capability matrix about the network. Based on the model, messages can be transmitted along the direction of increasing forwarding capacity of nodes, which can effectively control the number of copies and reduce network load. Research[14] suggested a status estimation and cache management algorithm. The study established a node identification method for estimating probabilities to satisfy the higher priority messages stored. At the same time, a cache management strategy is established to improve the efficiency of transmission by utilizing collaborative sharing between nodes.

With the increasing computing power and memory of mobile devices, many researchers begin to apply complex mathematical and artificial intelligence models to routing algorithms of opportunistic social networks, such as graph theory, decision tree, markov chain etc. Nguyenet al. [20] proposed a context-aware and human-centric approach strategy in opportunistic mobile social networks. The scheme combines graph theory with vertex-covered problems and uses an approximation algorithm to find the coverage area of the node. At the same time, the algorithm designs a human-centered guiding strategy to distribute the sensing devices according to the social relationship of the nodes. Sharma et al. [21] suggested a machine learning-based protocol for efficient routing in opportunistic networks. The model trains data by various factors such as the node popularity, the power consumption and the position of nodes, and then predicts the transmission probability between the nodes. The algorithm can improve the efficiency of data transmission, but it increases the overhead of node cache to some extent. In order to enhance network coverage and reduce data redundancy, research [22] proposed an effective transmission strategy exploiting node preference and social relations in opportunistic social networks. This strategy not only considers the social attributes but also discusses the influence of the transmission preferences of nodes on the data delivery, which can effectively improve the success rate of data transmission. However, this algorithm frequently calculates the transmission preference between nodes, which will consume a lot of resources of nodes.

Context-based routing algorithms have also attracted the attention of many researchers. This kind of algorithm measure the similarity of nodes through the historical information, and then evaluates the transmission relations between nodes, such as encounter probability and movement preference. Context-based algorithms require large amounts of data to be collected and frequently computed. Mayer et al. [23] proposed an algorithm supporting opportunities for context-aware social matching. Based on the simple similarity and proximity matching mechanism, this method establishes the relationship framework of the predictors of matching chance and maps the best message relay nodes. Wu et al. [12] proposed a sensor communication area and node extend routing algorithm in opportunistic networks. In this algorithm, the node can extend the sensor communication area through the context information and obtain the best relay node recommendation based on the historical interaction information. In literature，Yan et al. [24] suggested an effective data transmission algorithm based on social relationships in opportunistic mobile social networks. The algorithm takes into account the characteristics of the community in the mobile social network and uses the ideas of the faction to

divide communities. Then identify inefficient nodes to improve the performance of message transmission. In order to improve the efficiency and reliability of routing algorithm，Liet al. [23] proposed a cross-layer and reliable opportunistic routing algorithm. In the algorithm, a mechanism of logical reasoning and topology control is introduced. The node's relay priority is then determined based on the class of the node.

3. System model

Social opportunistic networks usually focus on user information, such as geographic location, time period, region, and keywords.

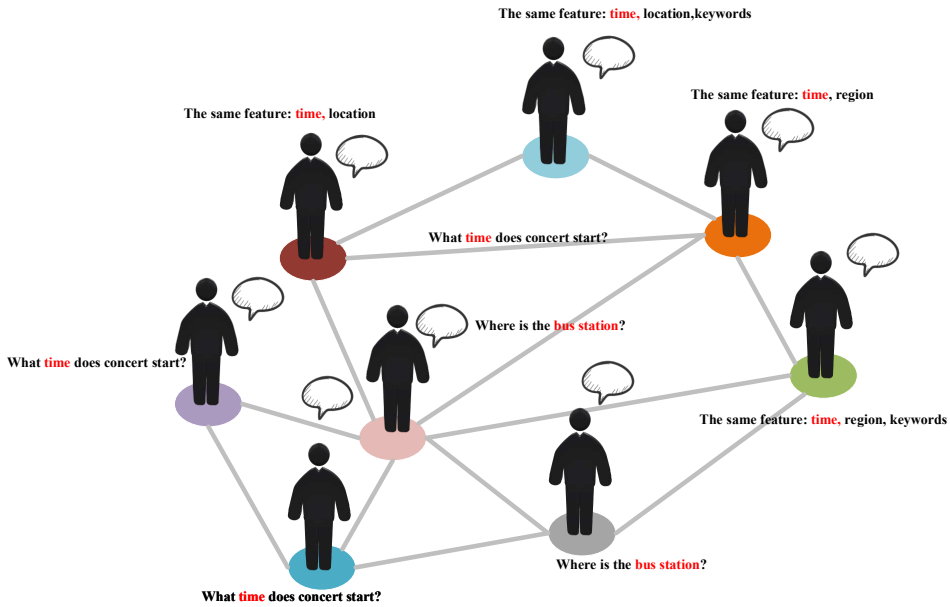


Figure 1. The characters and relationships in social opportunistic network

Figure 1 shows the characters and relationships in social opportunistic networks. For a person, he used to interest in the same characters when others mention similar topic. If a person establishes focusing on network, he has more ‘opportunity’ acquiring to ‘interest point’ or ‘help’ by mobile devices. He also found a good cooperation by neighbors when they have many similar characters.

In social opportunistic networks, if we only consider ‘interest point’ to choose neighbors and establish relationships, the over-consumption may limit ‘interest point’ nodes. So we must found some methods to solve this problem.

3.1. Problem description

The transmission environment of wireless network data can be expressed in the form of a graph structure. A complex network that contains linked information and attribute information can be represented as $G=(V,E,A)$, where V denotes the node set, E denotes the edge of a node, and A denotes the property of a node. For the matrix of nodes, we can define $X \in R_{+}^{n\times n}$, where the junctions i and j are connected. It can be expressed as $x_{ij}=x_{ji}=1$ or $x_{ij}=x_{ji}=0$. In opportunistic social networks, i and j can be defined as users. 1 can be shown as 'we are neighbors', else is 0. For the matrix X that is composed of x_{ij} and x_{ji}, it becomes a symmetric matrix. Matrix Y is characteristic parameter matrix with matrix X. Matrix Y concludes many ‘interest point’ in social networks.

We use $C=\{c_1,c_2,\cdots c_k\}$ to represent the set of communities in social network, k represents the number of nodes in a community, $H\in R_+^{n\times k}$ denotes the relationship matrix between nodes and communities in social opportunistic networks. Nodes maybe excavated according to the conditions of community membership, thereby proving that the existence of these nodes is the objective of our research.

3.2. Link matrix and attribute correlation matrix model analysis

In the problem description, each node in social opportunistic networks can be explained its connected matrix X and characteristic matrix Y. According to evaluate connection and characteristic, nodes can spend little consumption to select appropriate neighbors. Matrix X and Y are non-negative matrices to ensure that the community has many same close node links and attributes, and they can undertake joint decomposition of X and Y, and assume familiarity with the common breakdown factor matrix H. X and Y can be decomposed individually. For example, X can be transformed into three decomposed forms of $X=HSH^T$ in community mining, where H is the community affiliation matrix and S is the community connection strength matrix, it can judge relatedness with nodes in community. X is a symmetric matrix, and thus, S is also a symmetric matrix. Therefore, $H\leftarrow HS^{1/2}$ can be simplified into $X=HH^T$.

The joint decomposition model of X and Y is obtained based on their independent decomposition forms and by using the Frobenius norm to measure errors.

$$\min_{H\geq 0,W\geq 0}\{F(H,W)=\frac{1}{2}(\left\|X-HH^T\right\|_F^2+\left\|Y-WH^T\right\|_F^2+\lambda\left\|H\right\|_F^2+\varphi\left\|W\right\|_F^2)\} \tag{1}$$

$\left\|H\right\|_F^2$ and $\left\|W\right\|_F^2$ are regularized items that are used to improve the stability of the model, whereas λ and φ are regularized parameters. From the relationship between matrix trace and the Frobenius norm, the objective function of Formula (1) can be rewritten as

$$\begin{aligned}F=&\frac{1}{2}tr(XX^T)-2tr(XHH^T)+tr(HH^THH^T)+tr(YY^T)-2tr(YHW^T)+tr(WH^THH^T)\\&+\lambda tr(H^TH)+\varphi tr(W^TW)\end{aligned} \tag{2}$$

To minimize the objective function, F can obtain the approximate decomposition results of H and W given that $\forall h_{ij}\in H$, $\forall h_{ij}\geq 0$, $\forall w_{pq}\in W$, and $w_{pq}\geq 0$. The minimized F can be transformed into a typical constraint to solve the extremum problem using the Lagrange multiplier method. α_{ij} and β_{ij} are limited $h_{ij}\geq 0$ and $w_{pq}\geq 0$ that correspond to Lagrangian multipliers. $\alpha=[\alpha_{ij}]$, $\beta=[\beta_{ij}]$, and F corresponds to the Lagrangian multiplier function L, i.e., $L=F+tr(\alpha H^T)+tr(\beta H^T)$. The Karush–Kuhu–Tucker (KKT) condition can be introduced to optimize the solution function L.

$$\frac{\partial L}{\partial H}=-2XH+2HH^TH-Y^TW+HW^TW+\lambda H+\alpha=0 \tag{3}$$

$$\frac{\partial L}{\partial W} = -YH + WH^T H + \varphi W + \beta = 0 \tag{4}$$

In accordance with the smooth conditions of KKT, $\alpha_{ij} h_{ij} = 0$ and $\beta_{pq} w_{pq} = 0$ are obtained using Formulas (3) and (4).

$$(2XH + Y^T W)_{ij} h_{ij} - (2HH^T H + HW^T W + \lambda H)_{ij} h_{ij} = 0 \tag{5}$$

$$(YH)_{pq} w_{pq} - (WH^T H + \varphi W)_{pq} w_{pq} = 0 \tag{6}$$

The multiplicative iterative updating rules of h_{ij} and w_{pq} are obtained using Formulas (5) and (6), respectively.

$$h_{ij} = \frac{(2XH + Y^T W)_{ij}}{(2HH^T H + HW^T W + \lambda H)_{ij}} h_{ij} \tag{7}$$

$$w_{pq} = \frac{(YH)_{pq}}{(WH^T H + \varphi W)_{pq}} w_{pq} \tag{8}$$

Eq (7) and (8) as the objective function F constrained optimization of solving rules, the need to prove that the application of these iteration rules can guarantee the objective function F is a function. It can constantly achieve minimum convergence. The minimum value of convergence is the network node that satisfies the condition.

3.3. Iterative proof process of node conditional convergence

To prove the minimum value of convergence in a community, we must prove Formulas (7) and (8) for the objective function F.

The auxiliary function is imported to prove the method. First, for $\forall h_{ij} \in H$, the function $F_{h_{ij}}(h)$ is used to represent the first partial derivative of F with respect to according to Formula (1):

$$F_{h_{ij}}'(h) = \frac{\partial F}{\partial h_{ij}} = \left(-2XH + 2HH^T H - Y^T W + HW^T W + \lambda H\right)_{ij} \tag{9}$$

$F_{h_{ij}}'(h)$ can be calculated based on $F_{h_{ij}}(h)$ in the second rate $F_{h_{ij}}''(h)$.

$$F_{h_{ij}}''(h) = \frac{\partial(-2XH)_{ij}}{\partial h_{ij}} + \frac{\partial(2HH^T H)_{ij}}{\partial h_{ij}} + \frac{\partial(HW^T W)_{ij}}{\partial h_{ij}} + \lambda \tag{10}$$

Given that $\frac{\partial(-2XH)_{ij}}{\partial h_{ij}} = -2X_{ij}$,

$$\frac{\partial(2HH^T H)_{ij}}{\partial h_{ij}} = 2(HH^T)_{ij} + h_{ij} + (\sum_k k^2{}_{kj}) \tag{11}$$

$$\frac{\partial (HW^T W)_{ij}}{\partial h_{ij}} = (W^T W)_{ij} \tag{12}$$

Formulas (11) and (12) are obtained.

$$F_{h_{ij}}''(h) = -2X_{ij} + (2(HH^T)_{ij} + h_{ij} + (\sum_k k^2{}_{kj})) + (W^T W)_{ij} + \lambda \tag{13}$$

From $F_{h_{ij}}'(h)$, $F_{h_{ij}}(h)$ can be obtained on $F_{h_{ij}}^{(3)}(h) = 12h_{ij}$, $F_{h_{ij}}^{(4)}(h) = 12$, and the fourth derivative h_{ij}. Thus, $F_{h_{ij}}(h)$, which is the other high derivatives of h_{ij}, is $F_{h_{ij}}^{(n)}(h) = 0$ and $n \geq 5$.

Assume that $F_{h_{ij}}{}^t(h)$ represents the h_{ij} value of the t interaction update. Then, the $F_{h_{ij}}(h)$ Taylor expansion at $h^t{}_{ij}$ is expressed as

$$\begin{aligned} F_{h_{ij}}(h) &= F_{h_{ij}}(h^t{}_{ij}) + F_{h_{ij}}'(h^t{}_{ij})(h - h^t{}_{ij}) + \frac{1}{2}F_{h_{ij}}''(h^t{}_{ij})(h - h^t{}_{ij})^2 + \frac{1}{6}F_{h_{ij}}^{(3)}(h^t{}_{ij})(h - h^t{}_{ij})^3 \\ &+ \frac{1}{24}F_{h_{ij}}^{(4)}(h^t{}_{ij})(h - h^t{}_{ij})^4 \end{aligned} \tag{14}$$

Theorem 1: Define function, $G_{h_{ij}}(h, h^t{}_{ij})$ as an auxiliary function of $F_{h_{ij}}(h)$ can be obtained as follow:

$$\begin{aligned} G_{h_{ij}}(h, h^t{}_{ij}) &= F_{h_{ij}}(h^t{}_{ij}) + F_{h_{ij}}'(h^t{}_{ij})(h - h^t{}_{ij}) + \frac{1}{2}[\frac{(2HH^T H + HW^T W + \lambda H)_{ij} + 4(h_{ij})^3}{h^t{}_{ij}}] \times (h - h^t{}_{ij})^2 \\ &+ \frac{1}{6}F_{h_{ij}}^{(3)}(h^t{}_{ij})(h - h^t{}_{ij})^3 + \frac{1}{24}F_{h_{ij}}^{(4)}(h^t{}_{ij})(h - h^t{}_{ij})^4 \end{aligned}$$

Proof: When $h^t{}_{ij} = h$, $G_{h_{ij}}(h, h^t{}_{ij}) = F_{h_{ij}}(h)$ is used. We should show that when $h^t{}_{ij} \neq h$, $G_{h_{ij}}(h, h^t{}_{ij}) \geq F_{h_{ij}}(h)$,Thus, we will prove that this assumption is true.

$$\frac{(2HH^T H + HW^T W + \lambda H)_{ij} + 4(h_{ij})^3}{h^t{}_{ij}} \geq F_{h_{ij}}(h) \tag{15}$$

Because $h^t{}_{ij} \geq 0$, $w_{ij} \geq 0$ and $X_{ij} \geq 0$,

$$\begin{aligned} \frac{(2HH^T H + HW^T W + \lambda H)_{ij} + 4(h_{ij})^3}{h^t{}_{ij}} &= 2(HH^T)_{ij} + 2\sum_{k \neq i}(h_{kj})^2 + 4(h_{ij})^2 + (W^T W)_{ij} + \lambda \\ &\geq -2X_{ij} + 2((HH^T)_{ij} + h^2{}_{ij} + \sum_{k \neq i}(h_{kj})^2) + (W^T W)_{ij} + \lambda = F_{h_{ij}}''(h) \end{aligned} \tag{16}$$

Therefore, the inequality is set up and the proof ends. Theorem $G_{h_{ij}}(h, h^t{}_{ij})$ is the auxiliary function of $F_{h_{ij}}(h)$.

We must prove that the iterative solution for $F_{h_{ij}}(h)$ is consistent with the proof of Theorem 1.

Proof: From Theorem 1, the local minimum point of $G_{h_{ij}}(h, h^t{}_{ij})$, which is the local minimum point of $F_{h_{ij}}(h)$, can be obtained. The local minimum point of $G_{h_{ij}}(h, h^t{}_{ij})$ can be obtained by solving equation $G_{h_{ij}}{}'(h, h^t{}_{ij}) = 0$

$$\begin{aligned} G_{h_{ij}}{}'(h, h^t{}_{ij}) &= F'_{h_{ij}}(h^t{}_{ij}) + [\frac{(2HH^TH + HW^TW + \lambda H)_{ij} + 4(h_{ij})^3}{h^t{}_{ij}}] \times (h - h^t{}_{ij})^2 \\ &+ \frac{1}{2} F_{h_{ij}}{}^{(3)}(h^t{}_{ij})(h - h^t{}_{ij})^3 + \frac{1}{6} F_{h_{ij}}{}^{(4)}(h^t{}_{ij})(h - h^t{}_{ij})^4 \end{aligned} \tag{17}$$

$G_{h_{ij}}{}'(h, h^t{}_{ij})$ is the Taylor expansion of h. Therefore, $G_{h_{ij}}{}'(h, h^t{}_{ij}) = 0$ can be solved via Newton's iteration.

$$h^{t+1}{}_{ij} = h^t{}_{ij} - \frac{G_{h_{ij}}{}'(h^t{}_{ij}, h^t{}_{ij})}{G_{h_{ij}}{}''(h^t{}_{ij}, h^t{}_{ij})} \tag{18}$$

$$G_{h_{ij}}{}'(h^t{}_{ij}, h^t{}_{ij}) = \left(-2XH + 2HH^TH - Y^TW + HW^TW + \lambda H\right)_{ij}$$

$$G_{h_{ij}}{}''(h^t{}_{ij}, h^t{}_{ij}) = \frac{(2HH^TH + HW^TW + \lambda H)_{ij} + 4(h_{ij})^3}{h^t{}_{ij}} \tag{19}$$

According to (18) and (19)

$$h^{t+1}{}_{ij} = \frac{(2XH + Y^TW)_{ij} + 4(h_{ij})^3}{(2HH^TH + HW^TW + \lambda H)_{ij} + 4(h_{ij})^3} h^t{}_{ij} \tag{20}$$

The formula for Newton's iteration guarantees $G_{h_{ij}}{}'(h^t{}_{ij}, h^t{}_{ij})$ convergence. $G_{h_{ij}}(h, h^t{}_{ij})$ can obtain the local minimum point. From Theorem 1, the local minimum point can be obtained as $F_{h_{ij}}(h)$. In Formula (1) for the iteration rules with a new H all the elements, it can make the objective function F converged, thereby finally obtaining the local minimum.

By showing that it conforms to the node of community features, the minimum point occurs after convergence. It can be the node to the community and make the node with the same or similar characteristics and form a new community, with a strong correlation between community models.

4. Experiments

In the experiment, we adopt real network data sets to comprehensively evaluate the performance of the proposed algorithm. To compare the advantages and disadvantages of RRMNC, this study compares and analyses with three algorithms: the Agent Unsigned Network [28], SAC[29]and BAGC [30]. In the experiment, the tool used for execution is NS-3. Experimental results show that the proposed algorithm has the characteristics of feasibility and efficiency.

To verify the effectiveness of the research method, four typical complex network data sets, including link and attribute information, are selected for the experiment. The specific information of each data set is described as follows：

(1) Political blog data set [13]. This data set contains 1490 nodes and 19,190 edges, where each node represents a blog page about American politics and each edge represents the hyperlink relationship between web pages. considering LANMF modeling complex network in the case of an undirected graph so ignore the hyperlinks to sex, the last remaining 16175 side. Each node is associated with an attribute that indicates the political orientation of the web page of the blog, which can either be liberal or conservative

(2) Citeseers data set [14]. This data set contains 3312 nodes and 36,141 edges.. Each node represents a piece of literature of science and technology, each edge represents a reference to the relationship between science and technology literature, and each node is associated with a class attribute. The general category attribute value number is 6.

(3) CORA data set [15]. This data set is a Citeseer citation network data set with science and technology literature. It contains 2708 nodes and 56,417 edges. Each node is associated with a class attribute, and the general category attribute value number is 7.

In this study, community $C=\{c_1,c_2,\cdots c_k\}$ is adopted to evaluate the entropy of community chain density and community attributes. Their definitions are:

$$Desity(c_i)=\sum_{i=1}^{k}\frac{\{(v_p,v_q)\,|\,v_p,v_q\in c_i,(v_p,v_q)\in E\}}{|E|} \tag{21}$$

$$Entropy(c_i)=\sum_{i=1}^{m}\sum_{j=1}^{k}\frac{|c_j|}{|V|}entropy(\mathrm{a}_i,c_j) \tag{22}$$

Among，

$$entropy(a_i,c_j)=-\sum_{n=1}^{n_i}s_{ijn}log_2 s_{ijn} \tag{23}$$

where s_{ijn} represents the percentage of vertices in community j, in which s_{in} indicates the value of attribute a_i. Entropy represents the measure of uncertainty of information. In this paper, it is used to quantify the entropy weight of n communities in all attributes.

For the analysis of the ideal community, the result shows that as density increases, entropy decreases, and the higher the degree of the dominant node that will be found in the community.

4.1Community Quality Evaluation

We analyze four data sets according to several types of analysis method. The results are as follows.

Table 1. This Community analysis of Political Blog data set

algorithm	k=3		k=5		k=7		k=9	
	Density	Entropy	Density	Entropy	Density	Entropy	Density	Entropy
Agent	0.6754	0.9012	0.6671	0.9147	0.6279	0.7865	0.6721	0.7612

SAC	0.8541	0.6413	0.7732	0.4415	0.6941	0.2047	0.6356	0.3729
BAGC	0.7052	0.6211	0.7589	0.3924	0.7158	0.4756	0.7668	0.3755
RRMNC	0.9121	0.0147	0.8733	0.0221	0.8644	0.0412	0.8411	0.0579

Table 2 Community analysis of Citeseers data set

algorithm	k=5		k=10		k=15		k=20	
	Density	Entropy	Density	Entropy	Density	Entropy	Density	Entropy
Agent	0.2845	6.3321	0.2917	6.2571	0.2378	7.1544	0.1175	7.9865
SAC	0.3914	5.4123	0.3926	6.1452	0.2722	6.8411	0.1798	6.7815
BAGC	0.3456	5.9711	0.3968	5.1489	0.2477	6.4189	0.1591	5.7458
RRMNC	0.4095	2.5623	0.4311	2.4785	0.3721	2.5691	0.4123	2.7439

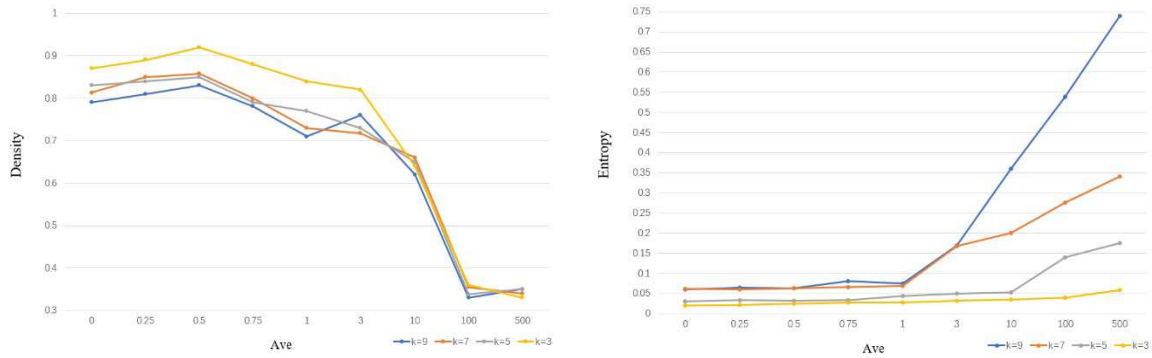
Table 3 Community analysis of CORA data set

algorithm	k=5		k=12		k=15		k=20	
	Density	Entropy	Density	Entropy	Density	Entropy	Density	Entropy
Agent	0.2364	7.3521	0.2247	6.9921	0.2533	6.9957	0.2371	8.6415
SAC	0.2655	6.4185	0.2886	5.4588	0.3108	6.9171	0.2782	7.4581
BAGC	0.2984	7.2151	0.2543	6.8471	0.3157	6.0412	0.2417	8.1459
RRMNC	0.3985	2.6213	0.4124	2.6225	0.4552	2.6051	0.4435	2.6171

Three methods are listed in Tables 1 to 3. The community evaluation results of the data sets can be found in Table 1, considering only community mining methods in the link information of Agent and BAGC because no integration occurs using attribute information. However, with the increase in k value, the density values become stable at 0.6 and above. The entropy value is relatively large, thereby indicating that the Agent node and the BAGC method mining community members have large attribute value differences and a low mining community quality. The RRMNC method is superior to Agent and the BAGC method for community mining, and it contains link information and attributes information.

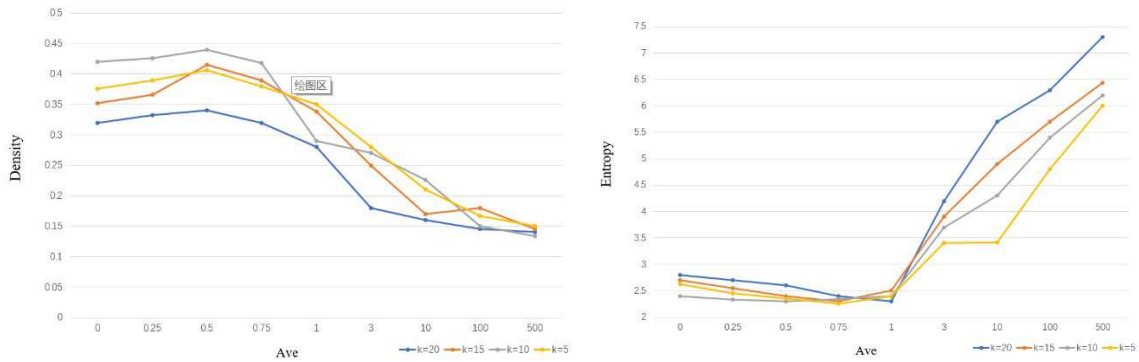
In the RRMNC method, the density values are greater than 0.8 with an increase in k value and the entropy value is less than 0.06, thereby indicating that the RRMNC community has not only close internal node links but also high attributes similarity ratios. This result is better than those of the other two methods (Agent and BAGC). In addition, the attribute value has a relatively higher number of SAC data sets. With an increase in k value, the two kinds of evaluation indexes of RRMNC method are within the ideal scope and of the SAC method are sufficiently stable, with differences in evaluation results. This finding shows that the RRMNC method exhibits before stability and reliability. For example, RRMNC is adopted because it can unite the integrated processing link and the joint matrix decomposition model of attribute information. In addition, Agent and BAGC must deal with the use of an isolated model, with no guarantee that the community member balance of the link and the attribute of the unified. Another significant advantage of the RRMNC method is that it can identify nodes and community ownership by directly approximating H matrix decomposition results. It does not need to use other clustering algorithms for the secondary processing of the obtained community mining results, and thus, it is more direct and effective than the Agent, BAGC, and SAC methods.

For the analysis of the social opportunistic network model according to the regularization process, the hope curve can be smoothed using λ and φ. The identification of the smooth curve in the regularization process is the key to the analysis of community quality. Given that the regularization parameters are equally important, $\lambda = \varphi = Ave$ is selected uniformly. The effects of the regularization parameters on the community models are as follows.



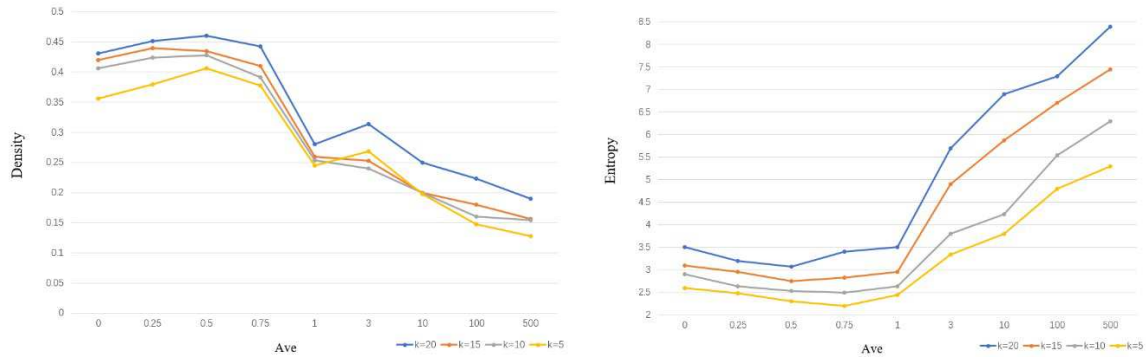
(a) Density (b) Entropy

Figure 2. Ave in Political Blogs



(a) Density (b) Entropy

Figure 3. Ave in Citeseers



(a) Density (b) Entropy

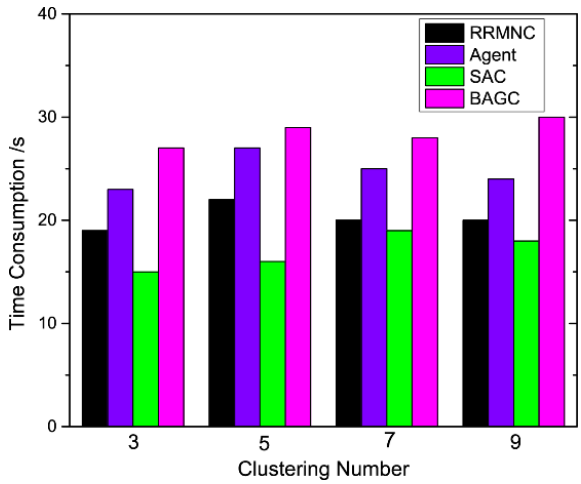
Figure 4. Ave in CORA

From the results that $\lambda = \varphi$ values are within the range of [0,1], stable and better evaluation results of community excavation quality can be obtained, along with an increase in value. The results of the community excavation quality evaluation become worse.

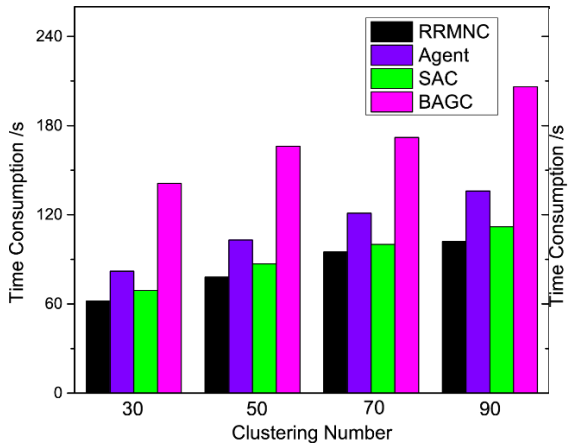
From the analysis of the experiments, the choice of community chain density and community attribute information entropy in different environments is important for the community discovery node. Simultaneously, selecting reasonable parameters using the regularization method is helpful in adjusting the community model.

4.2 Clustering Efficiency Evaluation

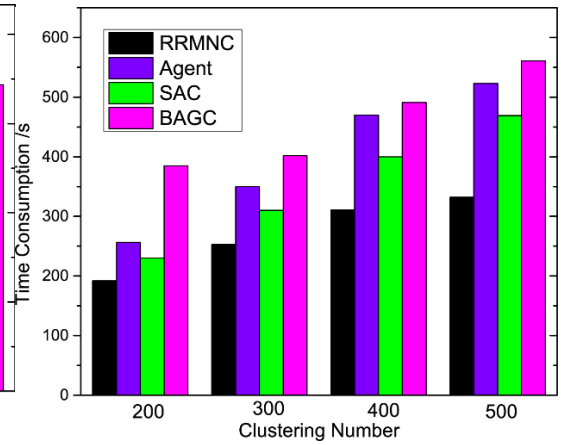
In this section, we compare the proposed algorithm with the other three algorithms on the efficiency of clustering in the same experimental environment. The clustering efficiency of the algorithms is quantified by clustering number and time consumption.



(a) Political blog data set



(b)Citeseers data set



(c) CORA data set

Figure 5. Clustering Efficiency

As shown in Figures5(a-c), the performance of several clustering algorithms under different clustering numbers. It can be seen that the time consumed by all methods in figure 5(a) is relatively short, which is due to the low clustering number and low data complexity in Political blog data set. In Figure 5 (b) and (c), RRMNC always has small time consumption, because the analytical matrix method of nodes and communities is established. This method makes use of the community characteristics of nodes with the

smallest convergent solution, and adopts the rule iteration method to decompose the properties of a large number of social opportunistic network data, which has a relatively low time complexity compared with other algorithms..

5. Conclusion

In the study, we established the resolve relationship matrix of node and community. This method is used to conduct attribute decomposition of a large amount of social opportunistic network data by using regular and iterative in the community features of nodes with the minimum solutions after convergence. These nodes comply with community rules. In the future work, we may focus on big data research and solve resource schedule and cache majorization methods when node can select the next transmit neighbors to keep and deliver messages. It is good to improve cooperation relationship between nodes.

Author Contributions: All authors designed the project and drafted the manuscript, collected the data, wrote the code and performed the analysis. All participated in finalizing and approved the manuscript.

Funding: Hunan Provincial Natural Science Foundation of China (2018JJ3299, 2018JJ3682).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wu J. &Chen, Z. "Human Activity Optimal Cooperation Objects Selection Routing Scheme in Opportunistic Networks Communication", Wireless Personal Communications, vol.95, no. 3, pp. 3357-3375, 2017. DOI: 10.1007/s11277-017-4001-1
2. Cheong M, Lee V C S. A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. Information Systems Frontiers, 2011, 13(13):45-59.Author 1, A.; Author 2, B. Book Title, 3rd ed.; Publisher: Publisher Location, Country, 2008; pp. 154–196.
3. Jia WU, Zhigang CHEN and Ming ZHAO, “Effective information transmission based on socialization nodes in opportunistic networks”,Computernetworks,Volume 129, Part 1, 24 December 2017, Pages 297-305. DOI:10.1016/j.comnet.2017.10.005
4. Phua C, Smith-Miles K, Lee V, et al. Resilient Identity Crime Detection. IEEE Transactions on Knowledge & Data Engineering, 2010, 24(3):533-546.
5. Aliotta J M, Pereira M, Johnson K W, et al. “Microvesicle entry into marrow cells mediates tissue-specific changes in mRNA by direct delivery of mRNA and induction of transcription”. Experimental hematology, vol. 38, no.3, pp. 233-245,March, 2012.
6. Jia WU, Xiaoming TIAN and Yanlin TAN, "Hospital evaluation mechanism based on mobile health for IoT system in social networks", Computers in Biology and Medicine, Volume 109, June 2019, Pages 138-147.
7. Jia WU, Peiyuan GUAN and Yanlin TAN, "Diagnosis and data probability decision based on Non-small cell lung cancer in medical system", IEEE ACCESS, Volume: 7, Issue:1, pp. 44851 - 44861. 2019. doi: 10.1109/ACCESS.2019.2909538
8. Lokhov A Y, Mézard M, Ohta H, et al. “Inferring the origin of an epidemic with a dynamic message-passing algorithm”. Physical Review E, vol. 90, no.1, pp.012801-012801, January 2014. Author 1, A.B. Title of Thesis. Level of Thesis, Degree-Granting University, Location of University, Date of Completion.
9. Wu J, Chen Z, Zhao M (2018) Information cache management and data transmission algorithm inopportunistic social networks. Wirel Netw. https://doi.org/10.1007/s11276-018-1691-6
10. Misra S, Ojha T, Mondal A. “Game-Theoretic Topology Controlfor Opportunistic Localization in Sparse Underwater Sensor Networks”. IEEE Transactions on Mobile Computing, vol.14, pp.990-1003, May, 2015.
11. Wu J, Chen Z, ZhaoM(2018)Weight distribution and community reconstitution based on communitiescommunications in social opportunistic networks. Peer-to-Peer Netw Appl. https://doi.org/10.1007/s12083-018-0649-x

12. Jia WU and Zhigang CHEN, “Sensor communication area and node extend routing algorithm in opportunistic networks”, Peer-to-Peer Networking and Applications. 2018. 11(1):90–100. DOI：10.1007/s12083-016-0526-4
13. Wu, J.; Yin, S.; Xiao, Y.; Yu, G. Effective Data Selection and Management Method Based on Dynamic Regulation in Opportunistic Social Networks. *Electronics.* 2020, 9, 1271.DOI: 10.3390/electronics9081271.
14. Jia WU and Zhigang CHEN and Ming ZHAO, “SECM: Status Estimation and Cache Management algorithm in opportunistic networks”, The Journal of Supercomputing. pp. 1–12, 2018. DOI: 10.1007/s11227-018-2675-0
15. Nicosia V, Mangioni G, Carchiolo V, et al. Extending the definition of modularity to directed graphs with overlapping communities[J]. Journal of Statistical Mechanics Theory & Experiment, 2008, 2009(3):3166 - 3168.
16. Xiaoli LI and Jia WU, "Node-oriented secure data transmission algorithm based on IoT system in social networks", *IEEE Communications Letters*, 17th Augest,2020. DOI:10.1109/LCOMM.2020.3017889.
17. Y. Yan, Z. Chen, J. Wu, L. Wang, K. Liu and Y. Wu, "Effective DataTransmission Strategy Based on Node Socialization in OpportunisticSocial Networks," in IEEE Access, vol. 7, pp. 22144-22160, 2019.
18. Park, Jiho,et al. "A forwarding scheme based on swarm intelligence and percolation centrality in opportunistic networks." Wireless Networks 22.8(2016):2511-2521.
19. Jingwen LUO, Jia WU and Yuzhou WU,“Advanced Data Delivery Strategy Base on Multi-Perceived Community with IoT in Social Complex Networks”,Complexity. Volume 2020 (2020), Article ID 3576542, 15 pages. DOI: 10.1155/2020/3576542 .
20. Nguyen, P.; Nahrstedt, K. Crowdsensing in Opportunistic Mobile SocialNetworks: A Context-awareand Human-centric Approach. arXiv, 2017,arXiv:1704.08598.
21. Sharma D K, Dhurandher S K, Woungang I, et al. A Machine Learning-Based Protocol for Efficient Routing in Opportunistic Networks[J]. IEEESystems Journal, 2016, PP(99):1-7
22. Jia WU, Zhigang CHEN and Ming ZHAO, “Community recombination and duplication node traverse algorithm in opportunistic social networks”, Peer-to-Peer Networking and Applications. pp 1–8, 2020. DOI: 10.1007/s12083-019-00833-0
23. Mayer, J.M.; Hiltz, S.R.; Barkhuus, L.; Vaananen, K.; Jones, Q. Supporting Opportunities for Context-AwareSocial Matching: An Experience Sampling Study. In Proceedings of the 2016 CHI Conference on HumanFactors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 2430–2441.
24. Xiao Y, Wu J. Data Transmission and Management Based on Node Communication in Opportunistic Social Networks. *Symmetry*. 2020; 12(8):1288. https://doi.org/10.3390/sym12081288.
25. Jia WU, Genghua YU and Peiyuan GUAN, "Interest characteristic probability predicted method in social opportunistic networks", IEEE ACCESS, Volume: 7, Issue:1, pp. 59002 - 59012. 2019. doi: 10.1109/ACCESS.2019.2915359.
26. Peiyuan GUAN and Jia WU, "Effective data communication based on social community in social opportunistic networks", IEEE ACCESS, Volume: 7, Issue:1, pp. 12405-12414, 2019. DOI: 10.1109/ACCESS.2019.2893308.
27. N. Li, J. F. Martínez-Ortega, and V. H. Diaz, ‘‘Cross-layer and reliableopportunistic routing algorithm for mobile ad hoc networks,’’ IEEE Sensors J., vol. 18, no. 3, pp. 5595–5609, Jul. 2018.
28. Xia, Feng, and Jianhua Ma. "Building smart communities with cyber-physical systems." In Proceedings of 1st international symposium on From digital footprints to social and community intelligence, pp. 1-6. ACM, 2011.
29. Weth, Christian Von Der, Ashraf M. Abdul, and Mohan Kankanhalli. "Cyber-physical social networks." ACM Transactions on Internet Technology (TOIT) 17, no. 2 (2017): 17.
30. Behfarnia, Ali, and Ali Eslami. "Error correction coding meets cyber-physical systems: message-passing analysis of self-healing interdependent networks." IEEE Transactions on Communications 65, no. 7 (2017): 2753-2768.
31. Liu J, Timsina P, El-Gayar O. A comparative analysis of semi-supervised learning: The case of article selection for medical systematic reviews. Information Systems Frontiers, 2018, 20(4):1-13.
32. Pourmasoumi A, Kahani M, Bagheri E. Mining variable fragments from process event logs. Information Systems Frontiers, 2017:1-21.
33. Wu D, Zhang F, Wang H, Wang R. Security-oriented opportunistic data forwarding in mobile social networks. Futur e Generation Computer Systems. 2018 Oct 1;87:803-15.
34. Yu, G., Wu, J. Content caching based on mobility prediction and joint user Prefetch in Mobile edge networks. Peer-to-Peer Netw. Appl. 13, 1839–1852 (2020). https://doi.org/10.1007/s12083-020-00954-x.

35. Wu, J., Chen, Z. & Zhao, M. Community recombination and duplication node traverse algorithm in opportunistic social networks. *Peer-to-Peer Networking and Applications.* 13, 940–947 (2020).https://doi.org/10.1007/s12083-019-00833-0.
36. Wu, J., Chen, Z. & Zhao, M. An efficient data packet iteration and transmission algorithm in opportunistic social networks. *Journal of Ambient Intelligence and Humanized Computing*. **11,** 3141–3153 (2020). https://doi.org/10.1007/s12652-019-01480-2.

Jia WU received the Ph.D. Degrees in software engineering Central South University, Chang-sha, Hunan, P.R.China, in 2016. He is associate professor in School of computer and engineering, Central South University and Faculty of IT, Monash University. Since 2010, he has been Algorithm engineer in IBM company in Seoul, Republic of Korea and in Shang-hai, P.R.China. He is a senior member of CCF(China Computer Federation), a member of IEEE and ACM. His research interests include wireless network, big data research, Medical Informatics.

Tian Xiaoming, senior engineer, associate researcher. She was received her Ph.D from Beijing Forestry University in 2013. She is a senior member of the China Forestry Society. Her research interests include the mechanism of action of medicinal plants and the bioinformatics analysis of important metabolic pathways of medicinal plants.

Figures

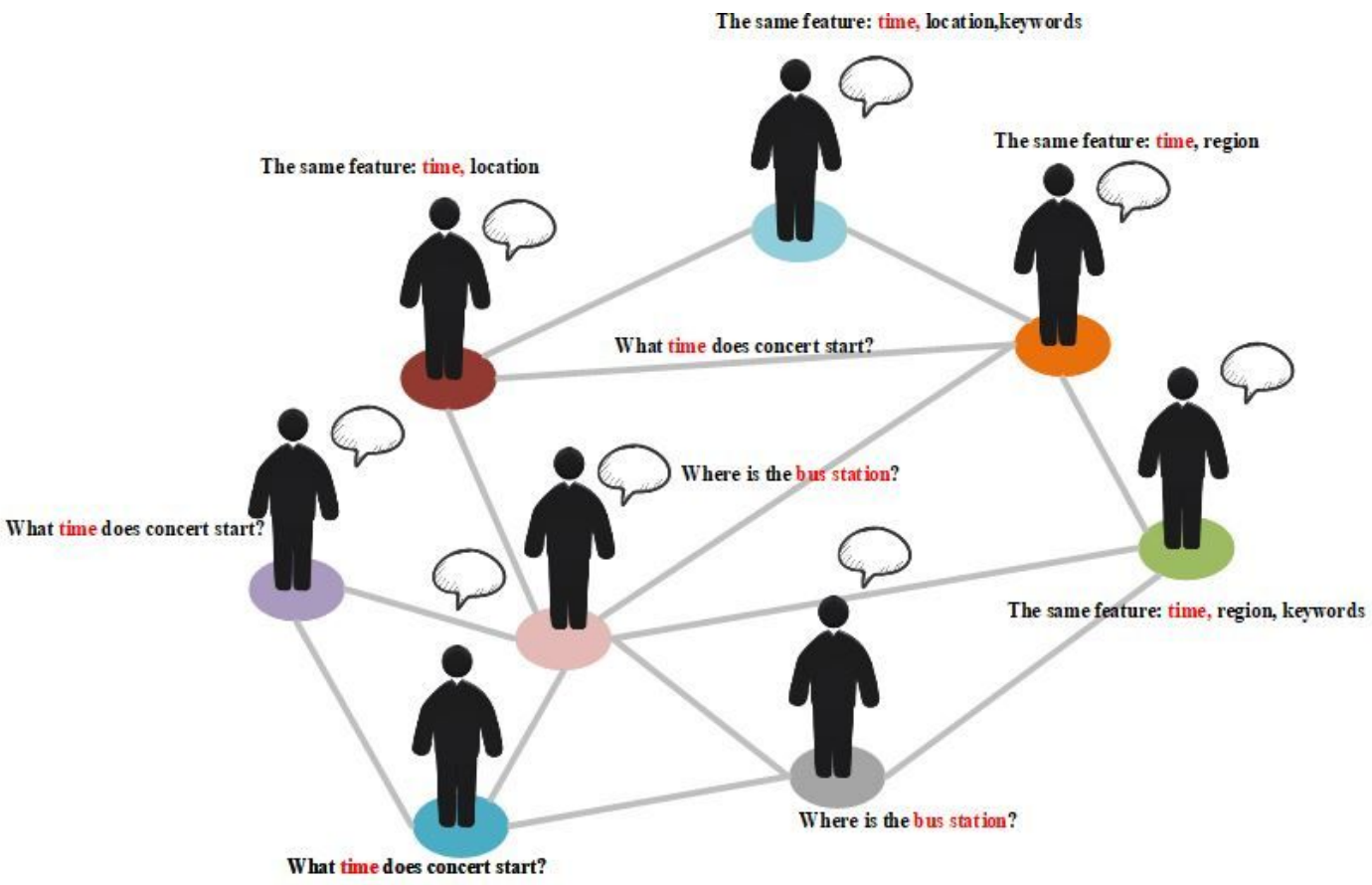
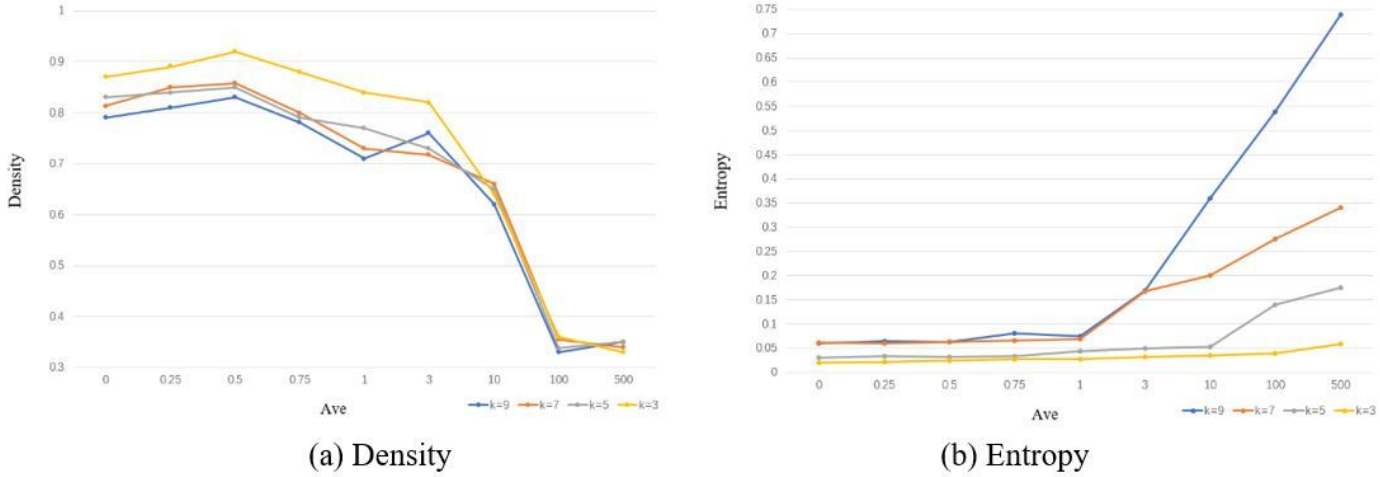


Figure 1

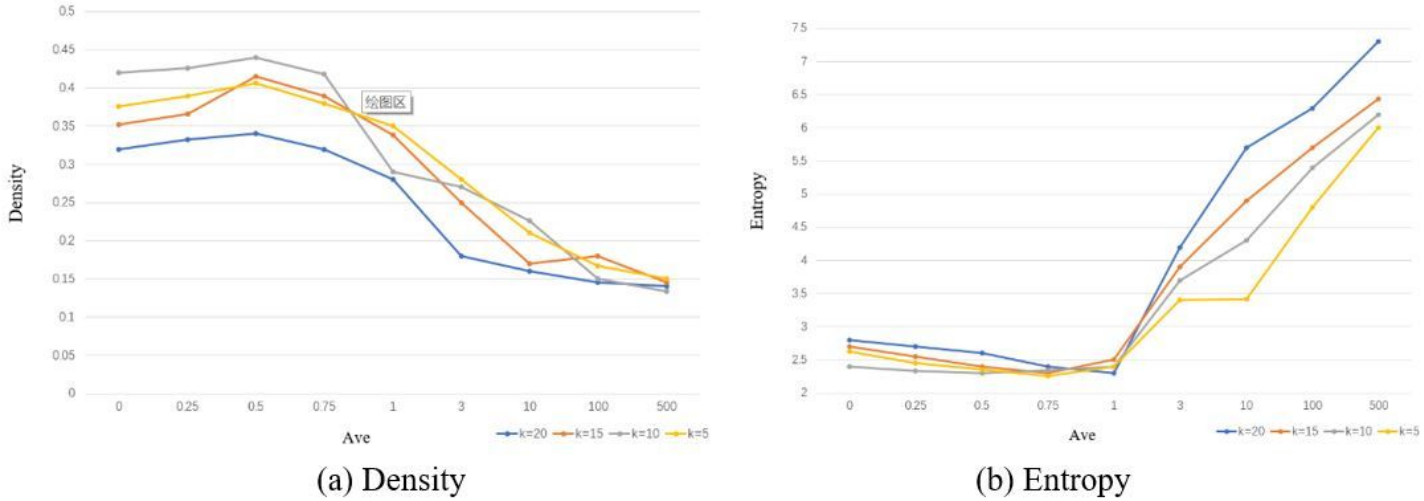
The characters and relationships in social opportunistic network



(a) Density (b) Entropy

Figure 2

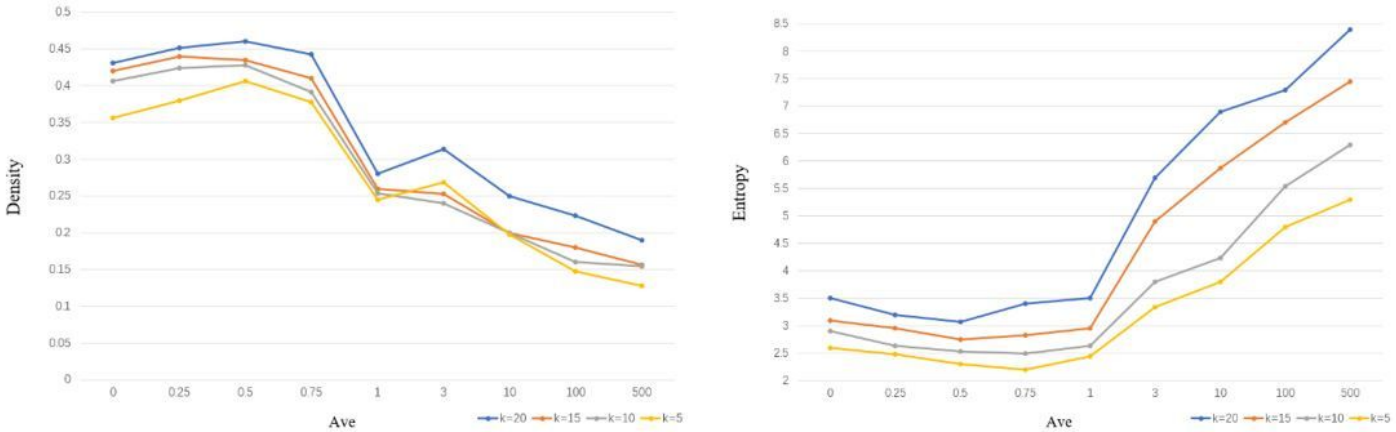
Ave in Political Blogs



(a) Density (b) Entropy

Figure 3

Ave in Citeseers



(a) Density (b) Entropy

Figure 4

Ave in CORA

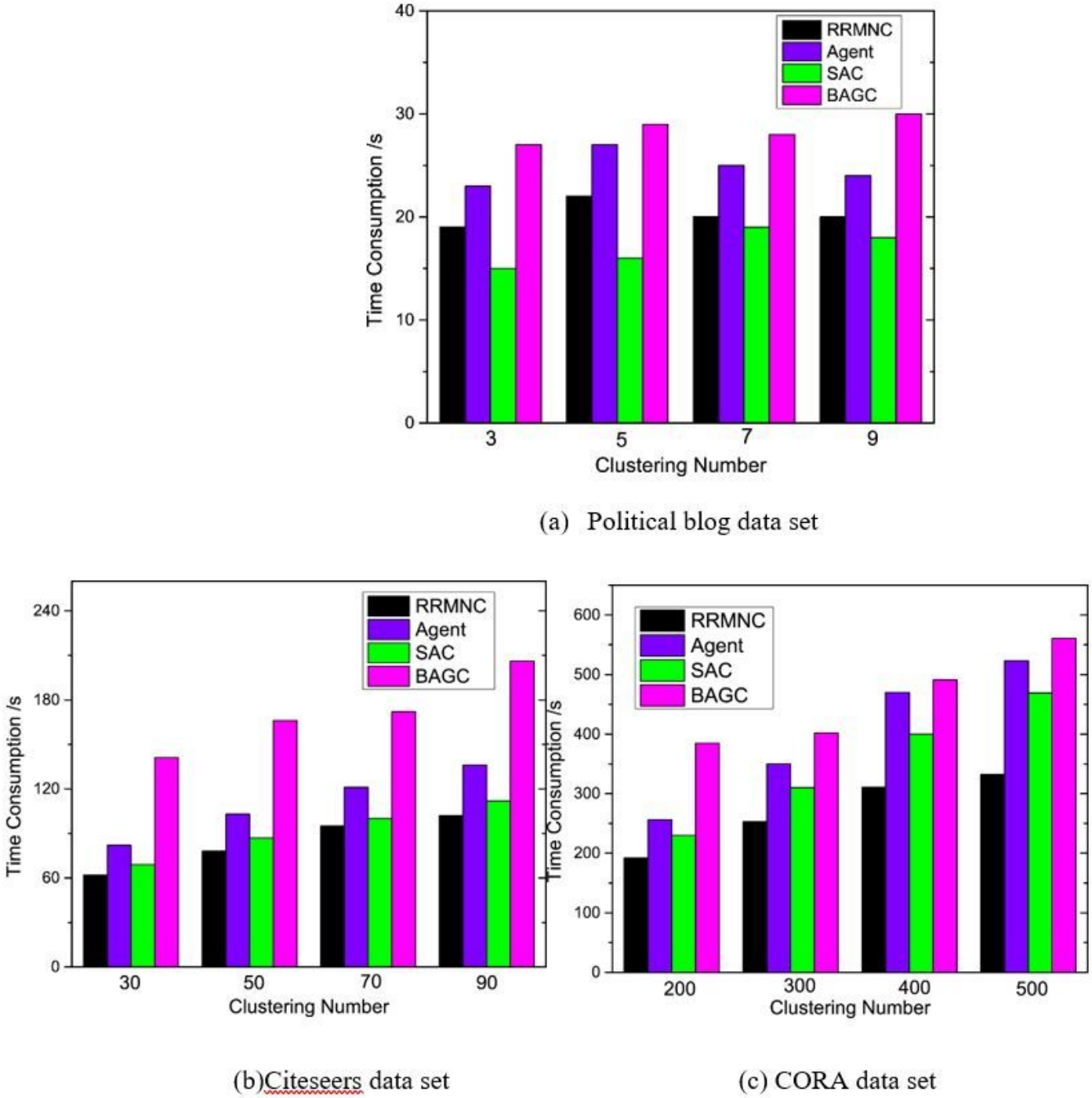


Figure 5

Clustering Efficiency