

Article

# A Transmission Prediction Neighbor Mechanism Based on a Mixed Probability Model in an Opportunistic Complex Social Network

Genghua Yu <sup>1</sup>, Zhigang Chen <sup>2,\*</sup>, Jia Wu <sup>2,\*</sup> and Jian Wu <sup>1</sup>

<sup>1</sup> School of Information Science and Engineering, Central South University, Changsha 410075, China; cnygh821@163.com (G.Y.); cnhnwujian@163.com (J.W.)

<sup>2</sup> School of Software, Central South University, Changsha 410075, China

\* Correspondence: czg@mail.csu.edu.cn (Z.C.); jiawu5110@163.com (J.W.); Tel.: +86-133-8748-0797 (Z.C.); +86-182-7312-5752 (J.W.)

Received: 17 October 2018; Accepted: 1 November 2018; Published: 6 November 2018



**Abstract:** The amount of data has skyrocketed in Fifth-generation (5G) networks. How to select an appropriate node to transmit information is important when we analyze complex data in 5G communication. We could sophisticate decision-making methods for more convenient data transmission, and opportunistic complex social networks play an increasingly important role. Users can adopt it for information sharing and data transmission. However, the encountering of nodes in mobile opportunistic network is random. The latest probabilistic routing method may not consider the social and cooperative nature of nodes, and could not be well applied to the large data transmission problem of social networks. Thus, we quantify the social and cooperative relationships symmetrically between the mobile devices themselves and the nodes, and then propose a routing algorithm based on an improved probability model to predict the probability of encounters between nodes (PEBN). Since our algorithm comprehensively considers the social relationship and cooperation relationship between nodes, the prediction result of the target node can also be given without encountering information. The neighbor nodes with higher probability are filtered by the prediction result. In the experiment, we set the node's selfishness randomly. The simulation results show that compared with other state-of-art transmission models, our algorithm has significantly improved the message delivery rate, hop count, and overhead.

**Keywords:** Opportunistic complex social network; cooperative; neighbor node; probability model; social relationship

## 1. Introduction

With the mobile cellular network evolving to the fifth generation, huge amounts of information data are generated every day. Opportunistic complex social networks have become a common data delivery platform [1–3], and the popularity of mobile devices has enabled a variety of new services in social networks to be realized [4]. Many social tools, such as Google Plus [5], Facebook [6], and Twitter [7], have a large number of users and generate data in each moment. The traditional social network approaches to dealing with transmission and reception of big data diversification faces challenges due to the diversification of online data. Many wireless devices used to deliver information and then faced data overload, which would be a hindrance to information interworking and information sharing. In order to cope with the data transmission in 5G wireless networks, we need to design a model and more convenient transmission mode to implement data forwarding in a flexible manner, which is more suitable for complex network environments.

In opportunistic social networks, there is no complete end-to-end path between most mobile devices, so nodes communicate by hop by hop. One common method of data dissemination is a base station (BS) that can first deliver a message to a mobile device. The device of the user could carry the message and transmit it to other mobile devices via Device-to-Device (D2D) communication. The mobile device receives data and waits for an “opportunity” to deliver the message to the next mobile device [8]. For data forwarding problems, most existing work either assumes that users are completely unselfish, i.e., they are willing to help any user to pass information [9,10], or assume that users are absolutely selfish and that users need to be encouraged to participate in data transmission [11,12]. However, in real-life scenarios, users are not absolutely selfish or absolutely selfless. Their level of selfishness may be random or related to social relationships. Therefore, we have established a cooperative relationship to solve this problem.

However, the social network is a complex network. The challenge is it is hard to estimate the daily data sent and received by users [13,14]. It is difficult to satisfy a large amount of data decision, transmission, and storage by a personal PDA device alone. It may cause a low delivery ratio and an excessive energy consumption of the transmission device. The decision could be made by assistance devices, such as the base station or the edge. The delivery ratio can be improved and data transmission energy consumption can be reduced. In addition, this network is divided into multiple disconnected subnets due to the loss of nodes due to problems, such as device movement, failure, or non-cooperation [15,16]. Such nodes are called critical nodes. The loss of key nodes will cause the network to be disconnected, which leads to the loss of important data in the network [17,18]. The researcher proposed the concept of probabilistic routing [19]. Each node maintains a link probability table that reaches any other node in the network, and then it could determine the key nodes of message transmission by comparing probability sizes. However, this method is too simple to calculate probabilistic values in the case of an explosion of information. It is also too much trouble to perform calculations and decisions every time the data transmission node performs. Even the task of maintaining a routing table is too expensive. Therefore, we collected and used the node’s encounter data, cooperation information, node characteristics, and social relationship to assist decision-making to predict key nodes in the opportunistic social network, and propose a relatively simple transmission model based on probability prediction.

To solve the huge amount of information in the 5G network, the mobile device has too much energy consumption for decision-making, transmission, and storage of a large amount of data. We propose ways to assist decision-making information transmission by devices, such as base stations or edges. Since the opportunistic complex social network does not necessarily acquire information about all nodes in the entire network, the node information is incomplete and the characteristics of each node cannot be utilized more fully to filter the key nodes. Therefore, the collected network information may not be complete, and the resulting encounter probability matrix may be incomplete or even sparse. In response to this problem, we first calculated some node social relationship values, cooperation probability values, and node encounter probabilities by pre-processing some data collected over a period of time. Then, we establish three related feature matrices and the vector of residual energy of nodes, and propose an algorithm based on hybrid probability matrix decomposition to predict key nodes. Finally, we filter the key nodes more reasonably and effectively through the impact of the collected feature information on the message passing between nodes in the network. Specifically, hidden information between nodes can also be fully utilized through the model to maximize the value of the node. By adaptive message forwarding decisions, the purpose of selecting a relay node and improving network resource utilization is achieved.

Specifically, the main contribution of this paper can be summarized as the following three aspects:

1. Through the transmission of decision information assisted by base stations or edge devices, we propose a method of pre-processing collected device information in an opportunistic complex social network and calculating social relationship values and cooperation probability values, so that the new probability matrix can be trained by adding additional information.

2. We propose a hybrid probability matrix decomposition model to predict the probability of encountering a node. We add node encounter information, social relationships, and partnerships to form a hybrid model to predict the encounter cooperation (EC) values between nodes, and filter the key encounter nodes through EC values.
3. We have designed a simpler way to transfer information to share the large number of transmission tasks of the central equipment. The mobile device only needs to request the central node to encounter the probability table of the other node and the destination node. Then, according to the information in the table, the data is passed to the neighbor node that has a higher value than its own and the destination node. The simulation results show that the PNEC algorithm has excellent effects in the message transmission process between nodes, and maximizes the characteristics of mobile nodes in the network.

## 2. Related Work

In the form of 5G networks, base station construction is relatively dense. The base station needs to collect the soaring data generated by devices, and analyze, process and transmit the data further. As shown in [20], a practical system architecture is proposed to process and extract useful information, and through the data caching mechanism, to solve the user data requirements. This approach increases the complexity of data processing. In such cases, most of the work of data transmission and data sharing can be done by the mobile device. The node's transmission decision, calculation, and other tasks can be assigned to the base station or edge equipment to reduce the node energy consumption. Many researchers have proposed ways for mobile devices to deliver messages in opportunistic mobile complex social networks. In such a network, there is no guarantee that the full connection path between the source and destination exists at any time, which makes the traditional routing protocol unable to deliver messages better between nodes. Some researchers have proposed probabilistic routing protocols for the network [19]. The nodes predict the reachability probability between the nodes by storing their encounter information, and select the nodes with higher encounter probability as the relay node. Such a method will increase the energy consumption of nodes in the case of the heavy task of 5G network data transmission.

In mobile social networks, some researchers have proposed using community structure to make data forwarding decisions. They assume that people in the same community would have closer relationships and more opportunities to connect with each other. The concept of the probability of the source node to the destination node is proposed by tracking the encounter situations of the nodes to study the activity of the node and the probability of reaching the target node, using the Poisson model of the social network contact. On this basis, the contact-aware opportunity-based forwarding (CAOF) scheme [21] is proposed by calculating the global and local probability of two different transmission phases. It includes the forwarding scheme between the inter-community transmission phase and the intra-community transmission phase. In the inter-community phase, neighbor nodes with higher global activity and high probability of source node to target node are selected as relays. Furthermore, forwarding decisions are determined by local metrics in the internal phase of the community. In this program, social characteristics and cooperation relationships may not be considered for integration into the design of the scheme, and may not perform well when the nodes are selfish.

In addition, there is an interesting study on mobile social networks. Since the mobile social network consists of nodes with different social attributes, the connection is transmitted or shared by chance. Some researchers have proposed a routing algorithm based on social identity awareness (SIaOR) [22]. They believe that many socially aware routing algorithms ignore an important social attribute, social identity. Therefore, researchers propose an opportunity routing algorithm in mobile social networks by considering the social identity and social impact of mobile nodes in mobile social networks. The algorithm not only considers the multiple social identities of mobile nodes, but also their social impacts. However, it is difficult to measure the ability of nodes to forward data if they

simply consider the social identity impact with the target node. Moreover, nodes with strong social relationships may not be suitable nodes for cooperation.

A new sensing method is the Internet of Things in the Mobile Crowd Sensor Network (MCSN). There are two existing transmission mechanisms: One is to transmit data through a cellular network, the other is an opportunistic transmission method through short-range wireless communication technology. Some researchers have proposed that the cost of cellular network transmission is high and it is not conducive to increasing user participation. Therefore, researchers focus on the application of wireless communication technology in MCSN by constructing a new opportunistic propagation model. They proposed an opportunistic data transmission mechanism based on the Socialization Susceptible Infected Susceptible Epidemic Model (SSIS) [23]. The mechanism uses the SSIS model to obtain a social relationship table by analyzing the social relationships of mobile nodes. The source node that performs the sensing task through the combination of the social relationship table and the spray and wait mechanism selectively propagates the data to other nodes until it reaches the destination node that can send the data to the platform. Using the spray and wait method will make the transmission mode more complicated and the node energy consumption increase.

Some researchers have proposed a new mobile opportunistic network routing protocol, MLProph [24], through machine learning and through further research on the mobile social network. The model uses decision trees and neural networks to train various factors, such as the predicted value inherited from the PROPHET routing scheme, node popularity, node's power consumption, speed, and location, to further calculate the probability of successful delivery of information. The algorithm trains based on historical information to obtain an equation for calculating the probability to detect whether the node can send the message to the destination node. This is an interesting way to make probabilistic predictions through machine learning methods, however, this method may only consider the nodes themselves, without considering the characteristics between the nodes.

According to the discussion on those methods, there is no complicated decision-making scheme considering the cooperative relationship and social relationship of the nodes. Additionally, the key neighbors are calculated and decided by nodes. When the amount of data transmission is large, calculation and decision by nodes may increase energy consumption. We need more complicated decision-making methods to solve data transmission problems more conveniently when the amount of data skyrockets in 5G networks. A good and effective decision-making mode determines whether the information transmission between mobile devices can achieve the desired effect. Many researchers have proposed a probabilistic routing method to calculate the probability of a nodes' final successful delivery to the destination by training the probabilistic equation. However, these methods could not consider whether the selected relay node is willing to cooperate in forwarding messages. Furthermore, nodes move randomly and socially, and it is worth considering whether such social relations will affect the activity rules of nodes. Therefore, we designed a model based on probabilistic prediction by quantifying social relationships and cooperative relationship values. We need to analyze and process and make decisions on a large amount of data, and improve the user satisfaction through active caching of edge devices to meet the needs of users in the 5G network [20]. In this case, we can make transmission calculations and decisions through small base stations or edge devices to reduce the workload of mobile nodes. Consequently, we have collected the characteristic information of nodes in the region through the base station and quantified the collected information to form the encounter matrix, cooperative matrix, and social relation matrix, respectively. Then, we used the improved probability matrix decomposition method to integrate the matrix information to update the encounter probability matrix and predict the probability value of the encounter and cooperation between two nodes without the encounter history information and updated the existing value. Finally, the neighbor node is filtered by the size of EC in the updated probability matrix. Experiments show that the proposed model reduces the network overhead of non-cooperative nodes, optimizes the path of messages to the destination nodes, and enables messages to be transmitted along the cooperative nodes with a high probability of encountering the destination nodes.

### 3. Model Design

#### 3.1. Node Data Collection and Transmission

The base station (BS) collects information of all nodes in the area and trains the encounter matrix at period, T, in the area. As shown in Figure 1, the base station collects the information of all nodes in the region over a period of time and trains the final probability matrix according to the designed model through these features. If the social relationship and the cooperative relationship cannot be obtained in the initial stage of the model application, the final probability matrix is the historical encounter probability matrix. When a node has a transmission task, the probability table of known nodes and destination nodes is requested from the base station. Specifically, when a new node enters the area, its characteristic information is sent to the BS. If it carries a probability forecast table, the BS obtains the table and updates its own matrix based on the probability forecast table obtained from the BS. If it does not carry a probability forecast table, it is sent to the matrix trained by the BS in the area. After the T period, the BS trains the matrix, M, according to Algorithm 1 (refer to Section 3.5), and transmits the matrix information according to the node information collected in the past T periods. During the period, the matrix, M, is updated by Algorithm 2 (refer to Section 3.5) according to the obtained node's probability forecast table and communication between BS before the T period. The main update idea is to add a record without some node information. The exchange of information when the mobile devices meet is also the same as updating the records carried. The message carried by the mobile devices is sent to the encounter devices whose probability of encounter is greater than the mobile device's own and destination mobile device's probability of encounter. If there is no record of the destination node, it is sent to nodes in the encounter node that have a higher probability of encountering mobile devices in another area.

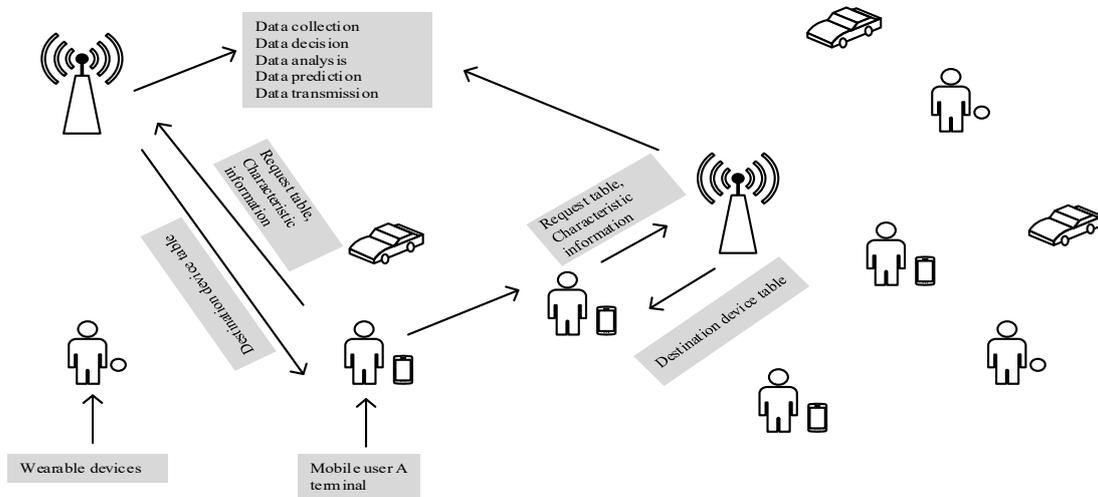


Figure 1. Example diagram of node information exchange in the area.

#### 3.2. Encounter Probability and Social Relationship Decomposition

##### 3.2.1. Encounter Relationship Value Calculation

First, we give a definition of computing the encounter probability. To facilitate the description of related problems, in this paper,  $m_{ij}$  is the probability of an encounter between node  $i$  and node  $j$  over a period of time. That is, the probability that node  $i$  and node  $j$  meet within the perceived range. The definition is as follows:

$$m_{ij} = \frac{w_{ij}}{\sum_{adj}^n w_{i,adj}} \tag{1}$$

where  $i$  denotes the source node, and  $j$  denotes the target node.  $w_{ij}$  is the number of historical encounters between node  $i$  and node  $j$  within a certain period of time.  $w_{i,adj}$  indicates that the number of times that node  $i$  has met with other nodes is within a certain period of time.

### 3.2.2. Social Relationship Value Calculation

The number of encounters between nodes simply reflects the node's encounters over a period of time. We cannot just rely on the number of node history encounters to predict an encounter in the future. Here, we consider the quantitative factors of mobile node social relations.

#### (1). Node Similarity

We believe that the closer the social relationship is, the more likely it is to meet in the future. We first consider the number of shared neighbor nodes in the node's social relationship. The similarity between nodes is defined as the number of common neighbors between two nodes. This equation can be defined as:

$$sim_{ij} = \frac{|C_i \cap C_j|}{|C_i \cup C_j|} \quad (2)$$

where  $sim_{ij}$  represents the similarity between node  $i$  and node  $j$ .  $|\cdot\cdot\cdot|$  represents the number of nodes in the collection.  $C_i$  is the set of neighbor nodes connected to node  $i$  at the current time, and  $C_j$  is the set of neighbor nodes connected to node  $j$  at the current time. Considering that two nodes share a similar set of neighbor nodes, two nodes are more likely to pass data through a common neighbor. Therefore, we consider that the relationship between the two nodes is higher and the probability of data transmission is higher.

#### (2). Devices' Mobility

Considering other factors that affect the value of total social relationships, node mobility change and node connection transformation are important factors that must be considered in the value of the overall social relationship. In a mobile opportunistic network, since the nodes are constantly moving, the connection between the nodes and the neighbors will change over time. In a continuous  $T$  time interval, the degree of change in the movement of such a node relative to another node is defined as the degree of motion transformation:

$$Move_{ij} = \frac{|(C'_i \cup C'_j) \cup (C_i \cup C_j)| - |(C'_i \cup C'_j) \cap (C_i \cup C_j)|}{|(C'_i \cup C'_j) \cup (C_i \cup C_j)|} \quad (3)$$

where  $Move_{ij}$  is the moving degree of the node at the current time.  $C'_i$  represents the set of neighbor nodes of node  $i$  before the  $T$  time interval, as is the case with  $C'_j$ .  $C_i$  represents the set of neighbor nodes of node  $i$  at the current time, and the same is true for  $C_j$ . From Equation (3), we can deduce that the higher the frequency of the movement of node  $j$  relative to node  $i$ , the larger the transformation of its neighbor node set relative to  $i$ , and the greater the degree of motion transformation.

#### (3). Connection Transformation

The degree to which a shared neighbor node of a node's connection with respect to another node changes and the dynamics of the neighbor node connection between them are referred to as the connection degrees of transition. In successive  $T$  time intervals, the degree of change in the connection of such a node relative to another node is defined as the connection transformation:

$$Conn_{ij} = \frac{|(C'_i \cap C'_j) \cup (C_i \cap C_j)| - |(C'_i \cap C'_j) \cap (C_i \cap C_j)|}{|(C'_i \cap C'_j) \cup (C_i \cap C_j)|} \quad (4)$$

From the formula (4),  $Conn_{ij}$  is the degree of change in the shared neighbor node set connected to the node. It can be seen that the greater the change in the set of shared neighbor nodes of the connection is, the higher the value of  $Conn_{ij}$ . By observing and analyzing the common neighbor nodes of the two nodes, we can see the motion changes of the two nodes in the current network. When the common neighbor node with node  $i$  associated with node  $j$  increases, the connection probability between node  $i$  and node  $j$  will increase, and then the possibility of data exchange will be improved.

Through the analysis of node similarity and node relative change, we can quantify the social relationship values between two nodes as follows:

$$S_{ij} = \alpha sim_{ij} + \beta(1 - Move_{ij}) + \gamma(1 - Conn_{ij}) \quad (5)$$

Among them, the smaller the degree of transformation, the greater the value of social relations.  $\alpha$ ,  $\beta$ ,  $\gamma$  are the coefficients of node similarity, moving transform, and connection transform, respectively. They represent the weighting factors that influence the value of social relationships by different factors, and  $\alpha + \beta + \gamma = 1$ .

### 3.2.3. Decomposition Method

We represent the probability of encounter and social relationship values between the nodes described above in a matrix. Wherein, let  $M = \{m_{ij}\}$  denote the matrix of  $n \times n$ ; that is, the encounter probability matrix. For a pair of nodes,  $n_i$  and  $n_j$ ,  $m_{ij} \in [0, 1]$  denote the historical encounter probability of node  $i$  to node  $j$ . Let  $S = \{s_{ij}\}$  denote the social relationship matrix of  $n \times n$ . Two nodes with strong social relationships affect each other's probability of encountering the same node. We also believe that nodes are more willing to be close to the nodes with high social relations. Because the devices in the opportunistic network are mostly carried by people, it is significant to analyze the probability of node encounters through the social relations of nodes. It is assumed that node,  $n_a$ , knows nothing about a node,  $n_c$ , and it meets node,  $n_b$ , and node,  $n_c$ , with a high relation value of node,  $n_a$ . So, node,  $n_a$ , meets it because it is close to node,  $n_b$ , and then node,  $n_a$ , and node,  $n_c$ , also meet. From the above analysis, we can summarize the above social process as:

$$\tilde{M}_{ij} = \frac{\sum_{z \in \kappa(i)} M_{zj} S_{iz}}{|\kappa(i)|} \quad (6)$$

where  $\tilde{M}_{ij}$  is the prediction of the probability that user  $u_i$  meets user  $u_j$ ,  $M_{ij}$  is the probability that user  $u_i$  meets user  $u_j$ ,  $\kappa(i)$  is the neighbor's set that user  $u_i$  relations and  $|\kappa(i)|$  is the number of related users of user  $u_i$  in the set,  $\kappa(i)$ .  $|\kappa(i)|$  can be merged into  $S_{ij}$ , since it is the normalization term of the relation evaluation of the estimate.

Then, we can infer the prediction of the probability that user  $u_i$  for all users is as follows:

$$\begin{pmatrix} \tilde{M}_{i1} \\ \tilde{M}_{i2} \\ \vdots \\ \tilde{M}_{in} \end{pmatrix} = \begin{pmatrix} S_{i1} & S_{i2} & \cdots & S_{in} \end{pmatrix} \begin{pmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ M_{21} & M_{22} & \cdots & M_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{n1} & M_{n2} & \cdots & M_{nn} \end{pmatrix} \quad (7)$$

Consequently, for all users, we can infer as:

$$\tilde{M} = M^T S \quad (8)$$

where  $SM$  can be interpreted as the probability forecasting purely based on the user relevance.

The idea of the probability matrix decomposition is to derive the  $l$ -dimensional features of the high-quality representative node (or user),  $N$ , of the node (or user) based on the analysis of the probability values of the encounter between the nodes. Let  $N \in M^{l \times n}$  and  $X \in M^{l \times n}$  be the latent

nodes and social factors. The feature matrix, the column vectors,  $N_i$  and  $X_j$ , represent the potential feature vectors of the node and social factor, respectively. We define the conditional distribution of the observed encounter probability as:

$$p(M|N, X, S, \sigma_M^2) = \prod_{i=1}^n \prod_{j=1}^n \mathcal{N} \left[ (m_{ij} | g(\sum_{z \in \kappa(i)} S_{iz} N_z^T X_j), \sigma_M^2) \right]^{I_{ij}^M} \tag{9}$$

where  $\mathcal{N}(x|u, \sigma^2)$  is the probability density function of the Gaussian distribution with a mean of  $u$  and a variance of  $\sigma^2$ , and  $I_{ij}^M$  is an exponential function. If node  $i$  meets node  $j$ , it is equal to 1, otherwise it is equal to 0. The function,  $g(x) = 1/(1 + \exp(-x))$ , is a logarithmic function, which allows the function value corresponding to  $N_i^T X_j$  to fall in the interval  $[0, 1]$ . We use zero-mean spherical Gaussian priors [24,25] on the node and social factor feature vectors:

$$p(N, \sigma_N^2) = \prod_i \mathcal{N}(N_i|0, \sigma_N^2 \mathbf{I}) p(X, \sigma_X^2) = \prod_j \mathcal{N}(X_j|0, \sigma_X^2 \mathbf{I}) \tag{10}$$

So, through the node and social factors features with a simple Bayesian inference [26] in Formulas (9) and (10), we have:

$$\begin{aligned} & p(N, X|M, S, \sigma_M^2, \sigma_N^2, \sigma_X^2) \\ & \propto p(M|S, N, X, \sigma_M^2) p(N|\sigma_N^2) p(X|\sigma_X^2) \\ & = \prod_{i=1}^n \prod_{j=1}^n \mathcal{N}[(m_{ij} | g(\sum_{z \in \kappa(i)} S_{iz} N_z^T X_j), \sigma_M^2)]^{I_{ij}^M} \\ & \quad \times \prod_{i=1}^n \mathcal{N}(N_i|0, \sigma_N^2 \mathbf{I}) \times \prod_{j=1}^n \mathcal{N}(X_j|0, \sigma_X^2 \mathbf{I}) \end{aligned} \tag{11}$$

We can assume that  $S$  is independent with the low-dimensional matrices,  $N$  and  $X$ , then this Equation (11) can be changed to:

$$\begin{aligned} & p(N, X|M, S, \sigma_M^2, \sigma_N^2, \sigma_X^2) \\ & = \prod_{i=1}^n \prod_{j=1}^n \mathcal{N}[(m_{ij} | g(\lambda N_i^T X_j + (1 - \lambda) \sum_{z \in \kappa(i)} S_{iz} N_z^T X_j), \sigma_M^2)]^{I_{ij}^M} \\ & \quad \times \prod_{i=1}^n \mathcal{N}(N_i|0, \sigma_N^2 \mathbf{I}) \times \prod_{j=1}^n \mathcal{N}(X_j|0, \sigma_X^2 \mathbf{I}) \end{aligned} \tag{12}$$

### 3.3. Cooperation Probability and Energy Decomposition

#### 3.3.1. Cooperative Probability Calculation

##### (1). Node Preference

$Pref_{ij}$  is the degree of preference between the two nodes,  $i$  and  $j$ . The node preference is different from the node encounter frequency, which refers to the time occupation ratio between two nodes establishing a connection and transmitting a message within a period of time. The specific calculation formula is as follows:

$$Pref_{ij} = \frac{\sum_{k=1}^n t_{i,j}^k}{\sum_{k=1}^n t_{i,other}^k} \tag{13}$$

$t_{i,j}^k$  represents the total time taken for message delivery when node  $i$  and node  $j$  establish a connection for the  $k$ th time.  $t_{i,other}^k$  represents the total time taken for message delivery when node  $i$  and other nodes establish a connection for the  $k$ th time.

## (2). Node Relationship Stability

$Rel_{ij}$  is the correlation between the average interval time delivery messages of the two nodes,  $i$  and  $j$ . Node relationship stability refers to the average interval between messages sent between nodes. Through it we can predict the possibility of future nodes establishing connections for a long period of time. This means that the shorter the connection interval in the most recent period, the more likely it is that the connection will be established in the future. In a period of time, the more the number of encounters, the longer the interval. Which indicates that the node has stable cooperation for a long time, as described below:

$$Rel_{ij} = \frac{T_{ij}^k - T_{ij}^1}{k \times T} \quad (14)$$

In the above equation,  $k$  is the number of message passes in  $T$  time.  $T_{ij}^k$  and  $T_{ij}^1$  are the time of the  $k$ th transmission of the message and time of the first transmission of the message, respectively, in  $T$  time.

## (3). Node Activeness

We use the number of times the node forwards the message for a period of time as a measure of node activeness. The higher the node activity, the higher the degree of cooperation is to some extent. We define it as follows:

$$Act_i = \frac{re_i}{\sum_{i=1}^n re_i} \quad (15)$$

where  $re_i$  indicates that the number of times the node  $i$  forwards the message is within a certain period of time,  $T$ .

## (4). Data Demand

We consider that nodes are more willing to cooperate to deliver data with common needs. At the time of message delivery, the node gives a rating feedback of the message demand. Whether a node meets and connects with other nodes within a certain period of time, the node prefers to receive and forward data useful to itself. Therefore, if the score of node feedback is higher for a period of time, it indicates that the probability that the neighbor node is willing to cooperate to forward the message is larger. Increasing this parameter can help predict the likelihood of future cooperation between nodes.

$$Need_{ij} = \frac{\sum_{\tau=1}^k q_{ij}^{\tau}}{Q \times k} \quad (16)$$

where  $q_{ij}^{\tau}$  represents the data demand feedback score given by node  $j$  when node  $i$  passes a message to node  $j$ .  $Q$  represents the highest score.

Through the degree of node preference, relationship stability, activity, and data demand analysis, we can quantify the cooperation probability values between two nodes as follows:

$$R_{ij} = aPref_{ij} + bRel_{ij} + cAct_j + dNeed_{ij} \quad (17)$$

Among them,  $a$ ,  $b$ ,  $c$ ,  $d$  is the coefficient of node preference, relationship stability, activity, and data demand. They respectively represent the weighting factors that influence the value of cooperation and  $a + b + c + d = 1$ .

### 3.3.2. Residual Energy Ratio

We consider the residual energy of the current moment of the mobile node. When the remaining energy of the mobile node is greater, the node is more willing to cooperate to forward the data.

$$E_i = \frac{E_r}{E_d} \quad (18)$$

where  $E_r$  represents the energy remaining in the device and  $E_d$  is the energy required for the packet to be transmitted.

### 3.3.3. Cooperation and Energy Decomposition

We consider the cooperation probability matrix. Suppose we have  $n$  nodes. The probability of forwarding messages from  $n$  nodes to different nodes is different. In other words, the selfishness of nodes for different nodes is different. We use the above method to calculate the cooperation value to form a cooperative forwarding probability matrix. Considering that the cooperation relationship between nodes is related to the ratio of the remaining energy of nodes, we believe that the greater the residual energy ratio of nodes, the more stable the cooperative relationship of nodes. It is significant to analyze the probability of node cooperation through the residual energy ratio of nodes. From the above analysis, we can define this as:

$$\tilde{R} = E^T \circ R \quad (19)$$

where  $E$  is a vector with a residual energy ratio of  $n$  nodes, and the  $\circ$  symbol is defined as each element corresponding to the  $i$ -th column element multiplied by the  $i$ th column of  $R$ . The new matrix we get is still represented by  $R$ , which is  $R = \tilde{R}$ .

Therefore, we let  $r_{ik}$  represent the probability of the message node,  $n_k$ , forwarding sent by node,  $n_i$ . We set  $N \in R^{l \times n}$  and  $Y \in R^{l \times n}$  as the feature matrix of potential nodes and cooperation factors, and the column vectors,  $N_i$  and  $Y_k$ , respectively represent potential feature vectors of the node and cooperation factor. We define the conditional distribution of the observed cooperative probabilities as:

$$p(R|N, Y, \sigma_R^2) = \prod_{i=1}^n \prod_{k=1}^n \mathcal{N} \left[ (r_{ik} | g(\sum_{z \in \kappa(i)} S_{iz} N_z^T X_{jk}), \sigma_R^2) \right]^{I_{ij}^R} \quad (20)$$

where  $I_{ij}^R$  is an exponential function. If node  $n_i$  has the probability of forwarding the message sent by node  $n_k$ , it is equal to 1, otherwise it is equal to 0. We also place zero-mean spherical Gaussian priors [24,25] on the node and cooperation factor feature vectors:

$$p(N, \sigma_N^2) = \prod_i \mathcal{N}(N_i | 0, \sigma_N^2 \mathbf{I}) p(Y, \sigma_Y^2) = \prod_k \mathcal{N}(Y_k | 0, \sigma_Y^2 \mathbf{I}) \quad (21)$$

Therefore, through the node and cooperation factors features with a simple Bayesian inference [26] in Formulas (20) and (21), we have:

$$\begin{aligned} & p(N, Y | R, \sigma_R^2, \sigma_N^2, \sigma_Y^2) \\ & \propto p(R|N, Y, \sigma_R^2) p(N | \sigma_N^2) p(Y | \sigma_Y^2) \\ & = \prod_{i=1}^n \prod_{k=1}^n \mathcal{N}[(r_{ik} | g(N_i^T Y_j), \sigma_R^2)]^{I_{ik}^R} \times \prod_{i=1}^n \mathcal{N}(N_i | 0, \sigma_N^2 \mathbf{I}) \times \prod_{k=1}^n \mathcal{N}(Y_k | 0, \sigma_Y^2 \mathbf{I}) \end{aligned} \quad (22)$$

### 3.4. Hybrid Model Solution

Through the above processes, to reflect the probability that the historical encounter between nodes will affect the judgment that the node is more likely to meet a certain node, we use the graphical model shown in Figure 2 to model the problem of probability prediction. The model combines the

encounter probability matrix, social relationship matrix, and node cooperative probability matrix into a consistent and compact feature representation. According to Figure 2, and Equations (11) and (22), the logarithm of the posterior distribution of node encounters and cooperation probability matrix are given by:

$$\begin{aligned} & \ln p(N, X, Y | M, R, \sigma_M^2, \sigma_R^2, \sigma_N^2, \sigma_X^2, \sigma_Y^2) = \\ & -\frac{1}{2\sigma_M^2} \sum_{i=1}^n \sum_{j=1}^n I_{ij}^M (m_{ij}^* - g(\lambda N_i^T X_j + (1-\lambda) \sum_{z \in \kappa(i)} S_{iz} N_z^T X_j))^2 \\ & -\frac{1}{2\sigma_R^2} \sum_{i=1}^n \sum_{k=1}^n I_{ik}^R (R_{ik} - g(N_i^T Y_k))^2 - \frac{1}{2\sigma_N^2} \sum_{i=1}^n N_i^T N_i - \frac{1}{2\sigma_X^2} \sum_{j=1}^n X_j^T X_j - \frac{1}{2\sigma_Y^2} \sum_{k=1}^n Y_k^T Y_k \\ & -\frac{1}{2} \left( \left( \sum_{i=1}^n \sum_{j=1}^n I_{ij}^M \right) \ln \sigma_M^2 + \left( \sum_{i=1}^n \sum_{k=1}^n I_{ik}^R \right) \ln \sigma_R^2 \right) - \frac{1}{2} (n \ln \sigma_N^2 + n \ln \sigma_X^2 + n \ln \sigma_Y^2) + \varepsilon \end{aligned} \tag{23}$$

where  $\varepsilon$  is a constant that does not depend on parameters. The hyperparameters (i.e., observed noise variance and a priori variance) are kept fixed after maximizing the logarithmic values of the three potential features, which is equivalent to the use of quadratic regularization terms to minimize the following sum of squared error objective functions:

$$\begin{aligned} & L(R, M, N, X, Y, S, T) \\ & = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n I_{ij}^M (m_{ij} - g(\lambda N_i^T X_j + (1-\lambda) \sum_{z \in \kappa(i)} S_{iz} N_z^T X_j))^2 \\ & + \frac{\lambda_R}{2} \sum_{i=1}^n \sum_{k=1}^n I_{ik}^R (r_{ik} - g(N_i^T Y_k))^2 + \frac{\lambda_N}{2} \|N\|_F^2 + \frac{\lambda_X}{2} \|X\|_F^2 + \frac{\lambda_Y}{2} \|Y\|_F^2 \end{aligned} \tag{24}$$

About formula (24),  $\lambda_R = \sigma_M^2 / \sigma_R^2$ ,  $\lambda_N = \sigma_M^2 / \sigma_N^2$ ,  $\lambda_X = \sigma_M^2 / \sigma_X^2$ ,  $\lambda_Y = \sigma_M^2 / \sigma_Y^2$ ,  $\|\cdot\|_F^2$  is the F norm. The local minimum of the objective function given by Equation (24) can be found by performing gradient descent in  $N_i$ ,  $X_j$ , and  $Y_k$ .

$$\begin{aligned} \frac{\partial L}{\partial N_i} &= \alpha \sum_{j=1}^n I_{ij}^M g'(\alpha N_i^T X_j + (1-\alpha) \sum_{z \in \kappa(i)} S_{iz} N_z^T X_j) X_j \\ &\quad \times (g(\alpha N_i^T X_j + (1-\alpha) \sum_{z \in \kappa(i)} S_{iz} N_z^T X_j) - m_{ij}) \\ &\quad + (1-\alpha) \sum_{h \in \varphi(i)} \sum_{j=1}^n I_{hj}^M g'(\alpha N_h^T X_j + (1-\alpha) \sum_{z \in \kappa(h)} S_{hz} N_z^T X_j) \\ &\quad \times (g(\alpha N_h^T X_j + (1-\alpha) \sum_{z \in \kappa(h)} S_{hz} N_z^T X_j) - m_{hj}) S_{hi} X_j \\ &\quad + \lambda_R \sum_{k=1}^n I_{ik}^R g'(N_i^T Y_k) (g(N_i^T Y_k) - r_{ik}) Y_k + \lambda_N N_i \\ \frac{\partial L}{\partial X_j} &= \sum_{i=1}^n I_{ij}^M g'(\lambda N_i^T X_j + (1-\lambda) \sum_{z \in \kappa(i)} S_{iz} N_z^T X_j) \\ &\quad \times \left( g(\lambda N_i^T X_j + (1-\lambda) \sum_{z \in \kappa(i)} S_{iz} N_z^T X_j) - m_{ij}^* \right) \times \left( \lambda N_i + (1-\lambda) \sum_{z \in \kappa(i)} S_{iz} N_z^T \right) + \lambda_X X_j \\ \frac{\partial L}{\partial Y_k} &= \lambda_R \sum_{i=1}^n I_{ik}^R g'(N_i^T Y_k) (g(N_i^T Y_k) - r_{ik}) N_i + \lambda_Y Y_k \end{aligned} \tag{25}$$

where  $\varphi(i)$  is the set of trust nodes for node  $n$ , and  $g'(x)$  is the derivative of the logarithmic function,  $g'(x) = \exp(-x) / (1 + \exp(-x))^2$ .

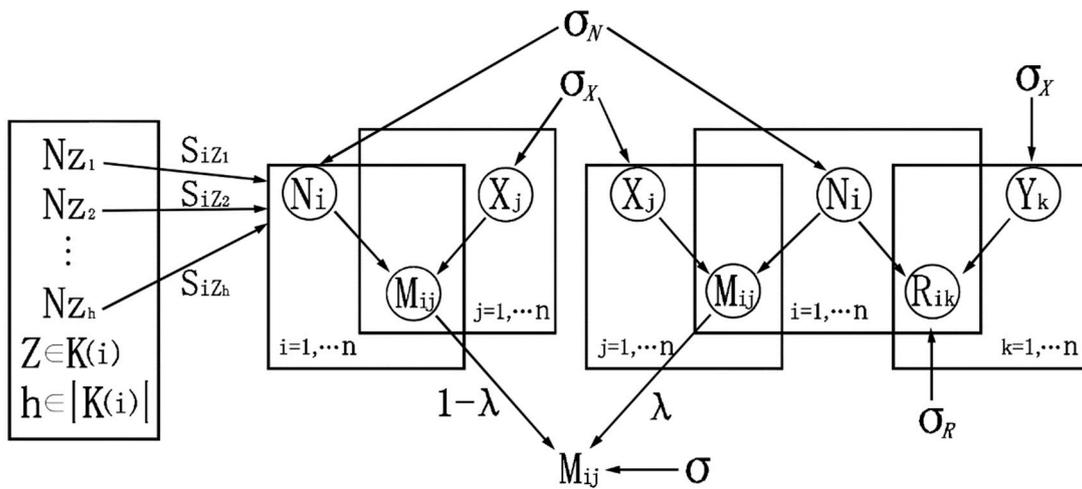


Figure 2. Probability prediction model.

### 3.5. Algorithm

The new probability matrix is obtained by using the matrix of social relation transformation and the matrix of cooperation forwarding relation. Because meeting the probability matrix is obtained by node meet history information, which is often incomplete, we need to fill in the missing value with additional information or update the encounter value with the cooperation degree and social of the devices. We called the newly updated probability matrix the encounter cooperation matrix containing the encounter cooperation (EC) values. Our algorithm is shown as follows:

---

**Algorithm 1:** PEBN (predict the probability of encounters between nodes).

---

**Input:**  $M$ : The encounter matrix formed by the node encounter probability map in the  $T$  period

$Sim, Move, Conn, Pref, Rel, Act, Need, E$

$T$  period

- 1: **Let**  $t = 0$ ; //  $t$  is used to record the elapsed time after the new matrix is updated;  
**Let**  $A []$ ; // Record all nodes that carry the message.
  - 2:  $S \leftarrow getSocialMatrix(Sim, Move, Conn, T)$  // get the social relations matrix
  - 3:  $R \leftarrow getCooperationMatrix(Pref, Rel, Act, Need, E, T)$  // get the cooperation relations matrix
  - 4:  $\tilde{M} \leftarrow getMeetProbabilityMatrix(M, S, R, T)$  // The new probability matrix of meeting is calculated based on the historical meeting, social relations and cooperation relations of the past  $T$  time period
  - 5:  $M^T \leftarrow getUpdateMatrixAccordAlgorithm2(M)$
  - 6: **For each**  $i \in (1, A.length]$  **do**
  - 7: Create a set of neighbor nodes  $N_{adj}$  // Devices in  $A$  looks for the neighbor within the communication range after  $\Delta t$  time.
  - 8: **If**  $N_{adj}$  contains a destination node; **break**; // If neighbors contains a destination node, the information transmission is successful and the loop ends.
  - 9: **Else If**  $M^T$  has the column vectors of the destination node **Then**
  - 10: **For each**  $z \in [1, N_{adj}.length]$
  - 11: **If** the EC value of the node  $n_z$  in  $N_{adj}$  and destination node > the EC value of node  $n_i$  and destination node // The neighbor with the highest EC Value of the target node is selected as the next hop node.
  - 12: **Choose** next hop nodes  $n_z$
  - 13: **A.add** ( $n_z$ ); // The selected devices are used as a new message-carrying devices.
  - 14: **End if**
  - 15: **End for**
  - 16: **End if**
  - 17: **Else If**  $M^T$  does not have the column vectors of the destination node **Then**
  - 18:  $M' \leftarrow M^T.getRemoveColVector(M)$  // The nodes contained in  $M$  are the area. We need to find nodes with high probability of encountering other area and remove the column vectors corresponding to the area nodes.
  - 19:  $N[] \leftarrow N_{adj}.getChooseColMaxMR(M', M'.getLine(N_{adj}, A.get(i)))$  // If the corresponding EC value of node  $n_i$  is the largest, this column is invalidated, and the nodes corresponding to the maximum EC value of each column are selected in turn.
  - 20: **A.add**( $N$ )
  - 21: **End if**
  - 22: **End for**
  - 23:  $t = getTime()$ ;
  - 24: **If** ( $t \geq T$ )
  - 25:  $M^T \leftarrow getUpdateNewMeetMatrix(M^T, R^T, S^T, T)$  // After the  $\Delta t$  time information is passed, the updated target encounter probability matrix is used as the input for calculating the encounter probability matrix in the next round of information transfer.
- Return** Step 5;
- 26: **End if**
-

**Algorithm 2:** Matrix update algorithm**Input:**  $M$  The matrix trained by Algorithm 1 $M^{t_1} \cdots M^{t_n}$  The encounter matrixes carried by the nodes entering the area that are received at different time intervals.**Output:**  $M^T$  Complete the updated matrix

- 1: **Let**  $k \in [1, n]; i, j \in [1, z]$  //  $n$  is the total number of encounter matrixes transmitted by nodes in the area at different time periods,  $z$  is the maximum number of rows or columns of the matrix  $M^{t_k}$
- 2: **For**  $k = n$  to 1 **do** // The later the matrix arrives, the more it will be processed first.
- 3: **If** matrix  $M$  does not contain the node  $i$  in the matrix  $M^{t_k}$  **do** // If the matrix  $M$  does not contain a row or a column.
- 4:  $M.add(M^{t_k}.getRow(i), M^{t_k}.getColumn(i))$  // Add this row or column directly
- 5: **Else if**  $M_{ij}$  is null and  $M_{ij}^{t_k}$  is not null // If the value of the  $i$ -th row and the  $j$ -column in the matrix  $M$  is null, and the value of the  $i$ -th row and the  $j$ -column of the matrix  $M^{t_k}$  reached at time  $t_k$  is not null, the value is filled into the matrix  $M$ .
- 6:  $M_{ij}.fill(M_{ij}^{t_k})$
- 7: **Else if**  $M_{ij}$  is not null and  $M_{ij}^{t_k}$  is not null // If the value of the  $i$ -th row and the  $j$ -th column in the matrix is not null, and the value of the  $i$ -th row and the  $j$ -th column of the matrix reached at time  $t_k$  is not null
- 8: **If**  $i \notin M.getHisArea()$  and  $j \in M.getHisArea()$  // If node  $i$  does not belong to the area and  $j$  belongs to the area, the value of  $M_{ij}$  is updated to  $M_{ij}^{t_k}$ .
- 9:  $M_{ij}.update(M_{ij}^{t_k})$
- 10: **End if**
- 11: **End if**
- 12: **End for**
- 13: **If**  $(i = j)$  // The matrix diagonal value is set to 1.
- 14:  $M_{ij} = 1$
- 15: **End if**
- 16: **Output**  $M^T$  //  $M^T$  is the updated encounter probability matrix.

### 3.6. Complexity Analysis

The main calculation method of the gradient method is to calculate the objective function,  $L$ , and its gradient according to the variables. The computational complexity of the objective function,  $L$ , is  $O(\rho_M l + \rho_M \bar{v} l + \rho_R l)$ , where  $\rho_M$  is the number of non-zero terms in the matrix,  $M$ , and  $\bar{v}$  is the average number of nodes that have a social relationship with the node. Since there are not many users in a region who have a social relationship value with a user, this indicates that the value of  $\bar{v}$  is relatively small. The computational complexities for the gradients,  $\frac{\partial L}{\partial N_i}$ ,  $\frac{\partial L}{\partial X_j}$ , and  $\frac{\partial L}{\partial Y_k}$ , in Equation (25) are  $O(\rho_M \bar{s} l + \rho_M \bar{s} \bar{v} l + \rho_R l)$ ,  $O(\rho_M l + \rho_M \bar{v} l)$  and  $O(\rho_R l)$ , respectively. Among them,  $\bar{s}$  is the average number of devices with social relationships that are close to one device,  $\rho_R$  is the number of non-zero terms in the matrix,  $R$ . Actually, in a complex social network, the value of  $\bar{s}$  is approximately equal to the value of  $\bar{v}$ . Therefore, the total computational complexity of an iteration is  $O(\rho_M \bar{s} l + \rho_M \bar{s} \bar{v} l + \rho_R l)$ , which indicates the calculation time of our method is linear with the amount of information collected by the node. If the obtained matrices are all sparse matrices, the algorithm will execute rapidly with a time complexity of  $O(n)$ . It also applies when the number of nodes is large. However, if the obtained matrix is a dense matrix, the complexity in the communication range of big data with dense nodes and complex relations is difficult to estimate. Since the complexity of the CAOP, SSIS, and SIaOR algorithms are both  $O(n^2)$ , the complexity analysis shows that the proposed method is simpler than them.

### 4. Experiment Analysis

In this section, we evaluate the performance of the proposed model, PEBN, in OMNets. The OMNets is a free, open source multi-protocol network simulation software, which plays an important role in network simulation.

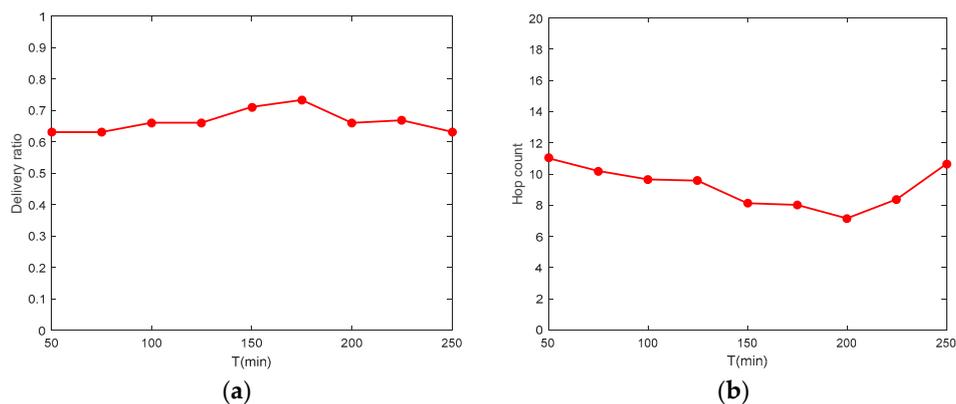
#### 4.1. Parameter Settings

The article uses the open source simulation tool, OMNets, to make the simulation. The network has a size of 3400 m \* 4500 m and 100–1300 nodes. Each node is moving randomly. Their moving speeds are, respectively, 1 m/s and 10 m/s. After reaching its destination, the node stays there for a random time. The sensing radius of the node is 50 m. The message generation interval is 25 to 35 s, the transmission interval is 17–18 min. The Time to Live is set to 100 min. The matrix update interval is 180 min. In the above simulation environment, we compared the performance of CAOP, SSIS, and SIaOR algorithms and our proposed routing algorithm (PEBN) in the delivery rate, network overhead rate, and hop count with the change of the number of nodes and simulation time.

#### 4.2. Parameter Analysis

The performance of PEBN is affected by some time-dependent parameters. First, we analyzed the parameters before comparing the algorithms. In order to improve the performance of the algorithm, we set the reasonable parameter values by experimentally analyzing the update period of the matrix and the node message delivery time interval.

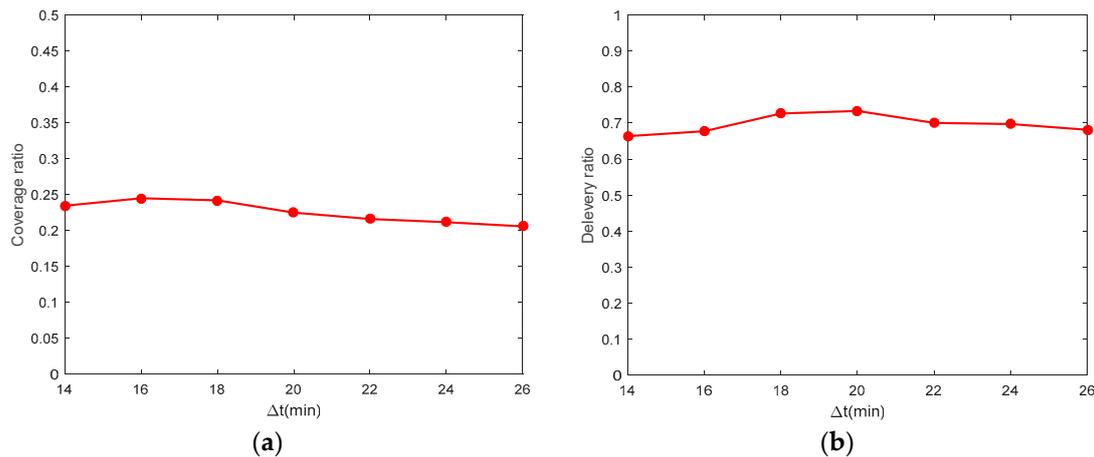
Different time periods,  $T$ , have different effects on the results of model training. Too short a time period will increase the network resource consumption. If the time period is too long, it will not be able to timely capture the latest dynamic changes of the node, reducing the accuracy of prediction, and affecting the transmission success rate. Therefore, we need to select the appropriate time period to update the predicted target matrix. Through analysis, as shown in Figure 3a, we can observe the fluctuation of the curve. As the time period increases, the transmission success rate increases first and then decreases. The transmission success rate reaches a large value at 175 min. Due to the change of the time period, the average transmission hop count is also affected. If the time period is set too long, it will affect the value of the model training and will be not accurate enough, resulting in more hops. Therefore, it can be seen from Figure 3b that as the time period increases, the hop count decreases and then increases. The phenomenon is because the time period is too short to collect enough information to reduce the error. When the time period reaches 200 min, the hop count is small. We can use the tradeoff method to determine that the period,  $T$ , is 180 min and increase the number of hops to ensure the delivery ratio.



**Figure 3.** The relationship between delivery ratio and average hop count with time periods  $T$ : (a) delivery ratio with different  $T$ ; (b) average hop count with different  $T$ .

We set the node to pass a message at  $\Delta t$  intervals. When a node transmits a message, a copy of the message is sent from a node within the messaging range that filters EC values greater than those of itself and the destination node. If the interval,  $\Delta t$ , is too short, it will cause the nodes in the network to send the same message copies to the same neighbor node more than once. The neighbor node receives the same message copy and refuses to receive the message, resulting in unnecessary network overhead. If the time interval,  $\Delta t$ , is too long, it will cause the message in the network to spread slowly and

increase the transmission delay of the message. From Figure 4b, we can observe that as  $\Delta t$  increases, the number of nodes covered by message copies transmitted by the node also decreases. During the 16 min to 18 min time period, the number of nodes covered by the message copy is large. However, the transmission of messages after  $\Delta t$  interval also affects the transmission success rate of the nodes. Just because a larger number of nodes are covered by the message does not mean that the delivery ratio can be improved; it may cause route congestion and cause information loss. Therefore, we can observe from Figure 4a that when  $\Delta t$  reaches 20 min, our delivery rate reaches the highest value. Since the value of  $\Delta t$  is between 16 min and 18 min, the node coverage does not fluctuate, and the transmission success rate fluctuates less in the range of 18–20 min. Therefore, we set  $\Delta t$  to 18 min.



**Figure 4.** The relationship between delivery ratio and node coverage ratio with different time interval  $\Delta t$ : (a) node coverage ratio different  $\Delta t$ ; (b) delivery ratio with different  $\Delta t$ .

To sum up, in the model with consideration of the node's transmission success rate and transmission delay variation, we set  $T = 180$  min and  $\Delta t = 18$  min for the following experiments.

### 4.3. Analysis of Simulation Results

#### 4.3.1. Methodology

In a real environment, each node is selfish. To make the simulation experiment closer to the real scene, we randomly set the selfishness of the node relative to other nodes in the experiment. When a node forwards a message, it forwards the message according to a certain cooperation probability that removes the selfishness. It should be emphasized that we aim to select nodes that have a high probability of encountering the target node and are willing to cooperate to forward the message. In the experiment, the cooperation matrix of the algorithm was not set manually, but was based on the cooperative probability calculated by the number of times the node helps a node to forward the message in the previous  $T$  period.

The evaluation of the model mainly comes from two aspects: (1) Effectiveness analysis. The experiment compares the difference between our presented model and other existing models in optimizing the network structure and improving the information delivery ratio. (2) Adaptability analysis: The probability matrix is periodically updated by various dynamic change information, and the nodes that can cooperate and have a high probability of encountering with the destination node are filtered to transmit data, thereby increasing the delivery ratio and reducing the average hop count. As a reference, we set a list of performance metrics, which is used for comparison, as follows:

1. Delivery ratio is the ratio of the number of delivered messages to the total number of generated messages.
2. Overhead is the average number of intermediate nodes used for one delivered message.

3. Hop count is the average hops of successfully delivered messages.

We use these performance metrics to analyze the effects of the model and compare them with other researchers' proposed solutions. To compare the performance of the PEBN protocol with other schemes, the delivery ratio, overhead ratio, and hop count were measured.

#### 4.3.2. Impacts on Delivery Ratio

Figures 5 and 6 shows the delivery performance of the PEBN protocol when the number of nodes and simulation time are different, respectively. From Figure 5 we can see that the delivery ratio of messages by the PEBN protocol is higher than other schemes due to better choices while sending the messages. This is because our solution sets a different degree of selfishness to each node for other nodes. Other schemes can easily lead to message loss, because they cannot fully consider the cooperating factors and social relationship factors of the nodes, and even cannot get more information about the nodes. Therefore, our algorithm can obtain more accurate predictions to have a better choice in the process of screening nodes. Moreover, as the number of nodes increases, the training of the prediction matrix is more extensive and accurate due to the increase of node social information. Thus, obtaining more unknown information between nodes, increasing the number of candidate nodes in the information transmission process, and improving the accuracy of node selection, thus makes it more likely to select the optimal node. From Figure 6, it can be observed that the delivery performance of the PEBN protocol is better than the other scheme. Because we not only consider the probability of encountering the target node, but also the path of the target node, it is unlikely to lose the packet. In addition, as the simulation time increases, the message delivery rate also increases, and the message will have more time to wait for the next best node.

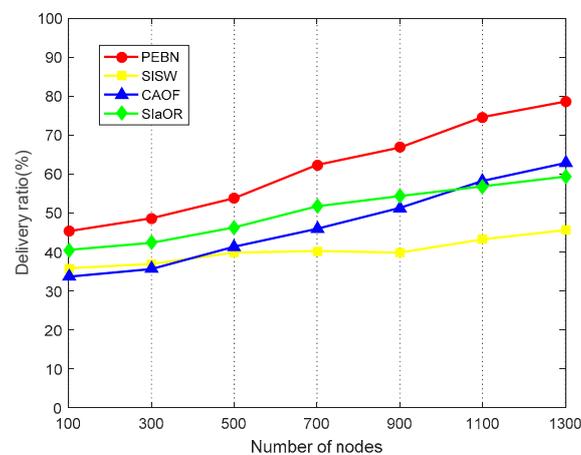


Figure 5. The relationship between the number of nodes and the delivery ratio.

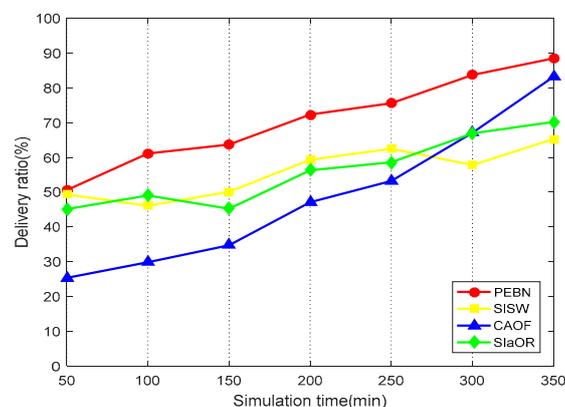


Figure 6. The relationship between time and delivery ratio.

### 4.3.3. Impacts on Average Hop Count

Figures 7 and 8 shows the hop count of the PEBN protocol when the number of node and simulation time varies, respectively. In Figure 7, the average hop count of the protocols is compared with others schemes. The PEBN protocol has the smallest average hop count in the compared scheme because it optimizes the routing of message transmissions more efficiently than other schemes. It can be seen in Figure 7 that as the number of nodes increases, the average hop count of the node increases and then gradually stabilizes. Our solution considers many factors, and as the number of nodes increases, the relationship between nodes becomes more complex, and more hidden features between nodes are mined through probabilistic prediction models, thus reducing the number of hops from the source node to the destination node. However, other methods may not have many parameters to enhance the adaptability of the model. Therefore, our model performs better. From Figure 8, it can be observed that our proposed scheme has fewer hops than other schemes. This is because our solution collects and processes node feature information, so we can get as much data as possible to more accurately predict the probability of encounters between nodes. As the simulation time increases, the node can capture more information about the movement of the nodes, which can add more additional information to the process of selecting the key nodes, making the model more robust. Therefore, the average hop count of the node is gradually reduced and stabilized.

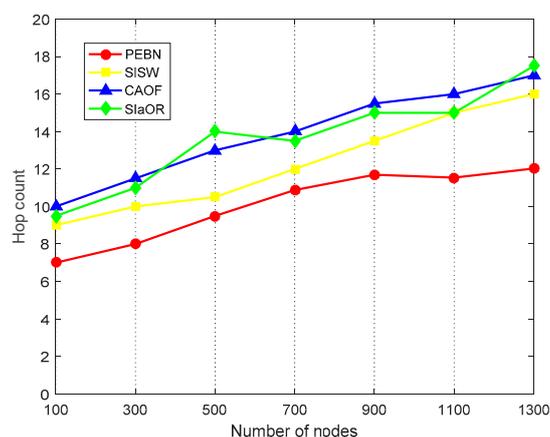


Figure 7. The relationship between the number of nodes and hop count.

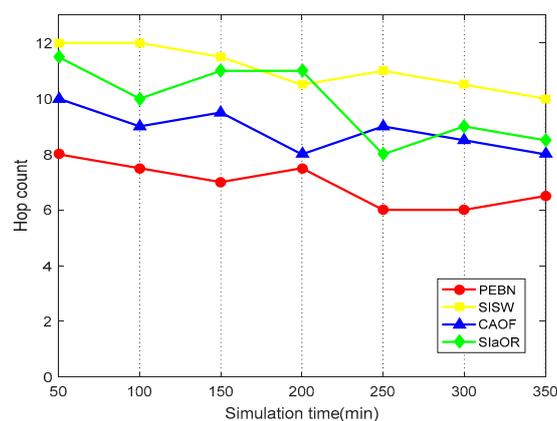
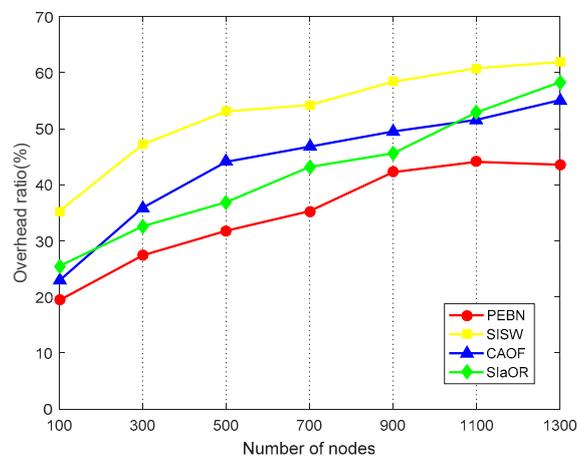


Figure 8. The relationship between time and hop count.

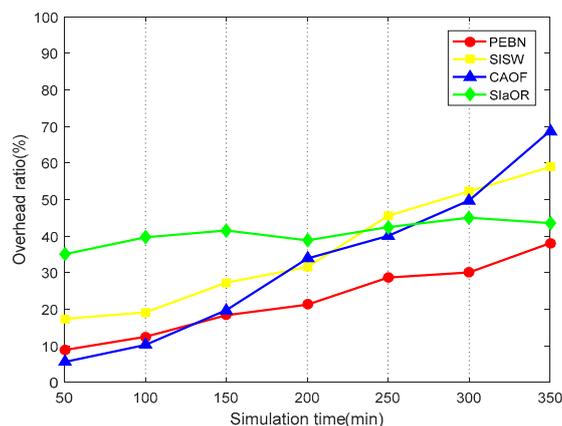
### 4.3.4. Impacts on Routing Overhead

Figures 9 and 10 shows the routing overhead of the PEBN protocol when the number of nodes and simulation time are varied, respectively. From Figure 9, it can be observed that the routing overhead of our scheme is compared with other schemes. Compared with the result, PEBN has a better

performance than other models. Since we apply the node cooperation to our proposed model, we can also perform well in scenarios where the node has selfish features. PNEC can predict transmitting neighbors better than other models; the message will be forwarded to a node with a high probability of satisfying the destination in the transmission route, which effectively reduces the cost loss of sending the message to the non-cooperative node. In addition, we can also observe that as the number of nodes increases, the amount of routing overhead also increases. An increase in the number of nodes means the number of candidate nodes for message selection increases. Messages need to be passed to more nodes, thereby increasing the overhead of information being propagated over the network. Moreover, this same observation can be seen in Figure 10, which shows that routing overhead increases as the simulation time is increased. Because the nodes do not need to continuously calculate and decide through the proposed model during the information transmission process, the transmission task of nodes becomes simpler. Furthermore, we have more complex probability models to choose neighbors, so our overhead increase ratio is relatively smaller than other models with the increase of simulation time.



**Figure 9.** The relationship between the number of nodes and overhead.



**Figure 10.** The relationship between time and overhead.

The conclusion can be drawn by comparing the three indicators of several schemes with the number of nodes and the simulation time. The PEBN algorithm adds social relationships and cooperation relationships when predicting probabilities so that a node can deliver a message to a cooperating node that has a high probability of encountering a destination node. Experiments show that the PEBN algorithm outperforms other algorithms in terms of the transmission success rate, routing overhead, and hop count. Specifically, the algorithm not only improves the data transmission efficiency in the network, but also adopts the updated information to adapt to the current network environment when the network topology changes dynamically. In addition, handling complex

transmission decisions through base stations or edge devices reduces the node overhead. By combining the social and cooperative relationships between nodes with a probabilistic prediction model, a better transmission path can be selected to reduce the number of hops to the destination node. Therefore, our scheme can improve performance in the transmission environment.

## 5. Conclusions

We proposed a routing decision method based on an improved probability model combined with a quantitative social relationship value and cooperative value to filter neighbor nodes. The algorithm combines multiple feature information between nodes and uses this feature information to quantify social relationship values and partnership values. Then, the prediction matrix was obtained by matrix decomposition and gradient descent, and the relay nodes were filtered according to the predicted probability values. In our model, we first quantified the node social relationship value and the cooperative relationship value based on the collected information to form the social relationship matrix and the cooperation relationship matrix. Then, we used them to update and predict the probability of encountering cooperation between nodes in the way we proposed. Finally, in the transmission phase, the node requested a probability table associated with the destination node, and selected a node with a high probability of encountering the destination node as the next hop node. The simulation results show that the protocol performs better than the SISW, CAO, and SIAOR transmission models in the transmission success rate, average hop count, and overhead. For a single node, the model optimizes the path from the source node to the target node. For the entire network, the performance of the network is improved to accommodate large-scale data transmission in the 5G. In the future, we will use real data sets to simulate real-world scenarios and explore other more efficient ways to improve information transmission.

**Author Contributions:** G.Y., Z.C. and J.W. (Jia Wu) conceived the idea of the paper. G.Y., Z.C., J.W. (Jia Wu) and J.W. (Jian Wu) drafted the manuscript and collected the data, wrote the code and performed the analysis; Z.C. contributed reagents/materials/analysis tools; G.Y. wrote and revised the paper.

**Funding:** This research was funded by [The Major Program of National Natural Science Foundation of China] grant number [No. 71633006]; [The National Natural Science Foundation of China] grant number [No. 616725407] [No. 61379057]; [China Postdoctoral Science Foundation funded project] grant number [2017M612586]; [The Postdoctoral Science Foundation of Central South University] grant number [185684].

**Acknowledgments:** This work was supported partially by “Mobile Health” Ministry of Education—China Mobile Joint Laboratory.

**Conflicts of Interest:** The authors declare that they have no competing interests.

## References

1. Wu, J.; Chen, Z.; Zhao, M. Effective information transmission based on socialization nodes in opportunistic networks. *Comput. Netw.* **2017**, *129*, 297–305. [[CrossRef](#)]
2. Wu, J.; Chen, Z.; Zhao, M. Information cache management and data transmission algorithm in opportunistic social networks. *Wirel. Netw.* **2018**, 1–12. [[CrossRef](#)]
3. Wu, J.; Chen, Z.; Zhao, M. Weight Distribution and Community Reconstitution Based on Communities Communications in Social Opportunistic Networks. *Peer Peer Netw. Appl.* **2018**, 1–9. [[CrossRef](#)]
4. Zhao, Y.; Song, W.; Han, Z. Social-Aware Data Dissemination via Device-to-Device Communications: Fusing Social and Mobile Networks with Incentive Constraints. *IEEE Trans. Serv. Comput.* **1939**. [[CrossRef](#)]
5. Lokhov, A.Y.; Mézard, M.; Ohta, H.; Zdeborová, L. Inferring the Origin of an Epidemic with Dynamic Message-Passing Algorithm. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2014**, *90*, 012801. [[CrossRef](#)] [[PubMed](#)]
6. Korkmaz, T.; Krunz, M. *A Randomized Algorithm for Finding a Path Subject to Multiple QoS Requirements*; Elsevier North-Holland, Inc.: Amsterdam, The Netherlands, 2001.
7. Aliotta, J.M.; Pereira, M.; Johnson, K.W.; de Paz, N.; Dooner, M.S.; Puente, N.; Ayala, C.; Brilliant, K.; Berz, D.; Lee, D.; et al. Microvesicle Entry into Marrow Cells Mediates Tissue-Specific Changes in Mrna by Direct Delivery of Mrna and Induction of Transcription. *Exp. Hematol.* **2010**, *38*, 233–245. [[CrossRef](#)] [[PubMed](#)]

8. Wu, J.; Zhao, M.; Chen, Z. Small Data: Effective Data Based on Big Communication Research in Social Networks. *Wirel. Pers. Commun.* **2018**, *99*, 1–14. [[CrossRef](#)]
9. Chuang, Y.J.; Lin, C.J. Cellular Traffic Offloading through Community-Based Opportunistic Dissemination. In Proceedings of the Wireless Communications and NETWORKING Conference, Shanghai, China, 1–4 April 2012.
10. Cheng, H.H.; Lin, C.J. Source Selection and Content Dissemination for Preference-Aware Traffic Offloading. *IEEE Trans. Parallel Distrib. Syst.* **2015**, *26*, 3160–3174. [[CrossRef](#)]
11. Ning, T.; Yang, Z.; Wu, H.; Han, Z. Self-Interest-Driven Incentives for Ad Dissemination in Autonomous Mobile Social Networks. In Proceedings of the IEEE INFOCOM, Turin, Italy, 14–19 April 2013.
12. Liu, G.; Ji, S.; Cai, Z. Strengthen Nodal Cooperation for Data Dissemination in Mobile Social Networks. *Pers. Ubiquitous Comput.* **2014**, *18*, 1797–1811. [[CrossRef](#)]
13. Tsiropoulou, E.E.; Mitsis, G.; Papavassiliou, S.; Tsiropoulou, E.E.; Mitsis, G.; Papavassiliou, S. Interest-Aware Energy Collection & Resource Management in Machine to Machine Communications. *Ad Hoc Netw.* **2018**, *68*, 48–57.
14. Tsiropoulou, E.E.; Paruchuri, S.T.; Baras, J.S. Interest, Energy and Physical-Aware Coalition Formation and Resource Allocation in Smart Iot Applications. In Proceedings of the Information Sciences and Systems, Baltimore, MD, USA, 22–24 March 2017.
15. Tsiropoulou, E.E.; Koukas, K.; Papavassiliou, S. A Socio-Physical and Mobility-Aware Coalition Formation Mechanism in Public Safety Networks. *EAI Endorsed Trans. Future Internet* **2018**, *4*, 154176. [[CrossRef](#)]
16. Lin, C.R.; Gerla, M. Adaptive Clustering for Mobile Wireless Networks. *IEEE J. Sel. Areas Commun.* **1997**, *15*, 1265–1275. [[CrossRef](#)]
17. Liu, W.; Gerla, M. Routing in Clustered Multihop Mobile Wireless Networks with Fading Channel. In Proceedings of the IEEE SICON, Singapore, 14–17 April 1997.
18. Gerla, M.; Tsai, T.C. Multicluster, Mobile, Multimedia Radio Network. *Wirel. Netw.* **1995**, *1*, 255–265. [[CrossRef](#)]
19. Wang, G.; Zheng, L.; Yan, L.; Zhang, H. Probabilistic routing algorithm based on transmission capability of nodes in DTN. In Proceedings of the 2017 11th IEEE International Conference on Anti-counterfeiting, Security, and Identification (ASID), Xiamen, China, 27–29 October 2017; pp. 146–149.
20. Zeydan, E.; Bastug, E.; Bennis, M.; Kader, M.A.; Karatepe, I.A.; Er, A.S.; Debbah, M. Big Data Caching for Networking: Moving from Cloud to Edge. *IEEE Commun. Mag.* **2016**, *54*, 36–42. [[CrossRef](#)]
21. Tao, J.; Wu, H.; Shi, S.; Hu, J.; Gao, Y. Contacts-aware opportunistic forwarding in mobile social networks: A community perspective. In Proceedings of the 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, 15–18 April 2018; pp. 1–6.
22. Wang, R.; Wang, X.; Hao, F.; Zhang, L.; Liu, S.; Wang, L.; Lin, Y. Social Identity-Aware Opportunistic Routing in Mobile Social Networks. *Trans. Emerg. Telecommun. Technol.* **2018**, *29*, e3297. [[CrossRef](#)]
23. Jia, B.; Zhou, T.; Li, W.; Xu, Z. An Opportunity Transmission Mechanism in Mobile Crowd Sensing Network based on SSIS Model. In Proceedings of the 2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design (CSCWD), Nanjing, China, 9–11 May 2018; pp. 695–700.
24. Sharma, D.K.; Dhurandher, S.K.; Woungang, I.; Srivastava, R.K.; Mohananey, A.; Rodrigues, J.J.P.C. A Machine Learning-Based Protocol for Efficient Routing in Opportunistic Networks. *IEEE Syst. J.* **2016**, *12*, 2207–2213. [[CrossRef](#)]
25. Salakhutdinov, R.; Mnih, A. Probabilistic Matrix Factorization. In Proceedings of the International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2007.
26. Dueck, D.; Frey, B.J. *Probabilistic Sparse Matrix Factorization*; University of Toronto Technical Report Psi; University of Toronto: Toronto, ON, Canada, 2004.

