A Routing Algorithm Based on Node Utility and Energy in Opportunistic Networks

Peiyan Yuan[1,2](✉) and Xiaoyan Huang[1]

[1] College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China
peiyan@htu.cn

[2] Engineering Lab of Intelligence Business and Internet of Things, Xinxiang, Henan, China

Abstract. There does not exit a complete transmission path in the opportunistic network. In order to further improve the delivery rate and transmission delay, hybrid routing algorithms with node utility and redundancy was proposed, but they face the problem of higher network overhead. In addition, data transmission consumes energy while the energy of node is limited. Therefore, efficient nodes may lead to energy depletion due to excessive data transmission, aggravating the network disconnection. Considering this fact, a routing algorithm based on node utility and energy is proposed, which takes into account the influence of self-difference and dynamic variation of node relationship on routing packets, and makes full use of social relations to calculate the social utility of nodes, and synthesizes the node's residual energy to evaluate the node's forwarding capability, so as to make balance between communication overhead and energy consumption. Finally, compared with other algorithms, the proposed routing scheme can achieve better packet delivery rate and transmission delay, while network overhead and energy balance are greatly improved.

Keywords: Opportunistic network · Data forwarding · Social utility · Energy balance · Routing algorithm

1 Introduction

Nodes in traditional networks can find an end-to-end communication link before routing, and then complete the packet forwarding task according to the determined path. This means that the traditional network topology is connected in most of the time and there is at least one connected path. However, in some practical application scenarios, there is usually no end-to-end multi-hop wireless link between the source and destination node pairs, resulting in the traditional network routing strategy ineffectively. In order to solve the communication problem in disconnected environment, the opportunistic network has been widely concerned by researchers. The opportunistic network takes advantage of the encounter opportunities formed in the process of node movement and adopts the way of "store-carry-forward" for data transmission [1], which has a strong adaptability to the characteristics of node random movement, uncertain distribution density and limited communication range. Therefore, the

Q. Guo et al. (Eds.): WiSATS 2021, LNICST 410, pp. 533–553, 2022.
https://doi.org/10.1007/978-3-030-93398-2_49

opportunistic network has a wide application prospect in the fields of vehicle network, wildlife information collection, large gatherings (such as large sports events, concerts, etc.), remote areas and deep space communications.

The network topology of the opportunistic network changes frequently and the duration of links is uncertain, so the network is divided into several disconnected subregions, which can also be called "intermittently connected networks" [2]. The forwarding path cannot be determined in advance for data transmission in such a network environment. It is necessary to select node dynamically. Usually, the data packet will be sent to the destination after forwarding with multi-hop nodes. Obviously, routing decision-making problem has always been the focus of research in opportunistic networks, and its essence is how to select appropriate relay nodes to complete the data transmission task. So far, the routing protocols in opportunistic networks can be divided into two categories according to the way of selecting relays. One is zero-information routing, which does not need complex network information, and only uses the encounter opportunities generated by node movement to complete data transmission. The other is information-assisted routing, which employs additional information to make a forwarding decision, that is, calculate the utility value of the node according to the node information, and further select high-utility nodes to route packets, which is also called utility routing.

Zero-information routing does not consider the heterogeneity of nodes in the network, and uses a relatively single routing method to complete the exchange and delivery of data packets. Information-assisted routing is the focus of the current work, and the evaluation function judges the ability of nodes to route packets according to different types of parameter information. Literature [3–5] use the historical contact information of nodes to evaluate node's utility. The Prophet algorithm [3] is a classical probability algorithm based on the contact frequency of nodes. The IPRA algorithm [4] uses the historical contact information (contact times) of the direct encounter node and the two-hop neighbor node to calculate the contact probability, and then makes the forwarding decision. The HPR algorithm [5] calculates the probability that the intermediate node can successfully deliver the data packet to the destination node based on the encounter frequency and contact duration. Literature [6–11] uses the social information of nodes to evaluate node's utility. The PageRank algorithm further makes the forwarding decision by calculating the centrality of neighbors [6]. The PeopleRank algorithm [7] sorts the centrality of the nodes on the basis of the PageRank algorithm, and selects the node with high global centrality as the next-hop relay. The Bubblerap algorithm [8] uses the number of times that act as a relay to calculate the centrality of the node, and selects the appropriate relay node based on its centrality ranking in the local community and global environment. The CMTR algorithm [9] is a social awareness protocol based on throw-boxes, which uses the number of encounters of nodes to calculate the degree of centrality, and deploys static and dynamic throw-boxes to forward data between remote communities. The SAPC algorithm [10] calculates the social utility of nodes based on the degree of social activity and physical contact factors of nodes. The HiBOp algorithm [11] mainly uses the current and historical information of the node (including the node's own information and neighbor information) to calculate the similarity probability between the node and the destination node as the basis for forwarding. Literature [12–15] uses the location information of nodes to evaluate node's utility. The LOOP algorithm [12] learns the historical movement trajectory by building a Bayesian model and predicts the future movement of the node, and finally forwards the packet to the destination area rather than to a specific node. The EDR algorithm [13] dynamically determines the next-hop forwarding node by using the ratio of encounter parameters to distance parameters, and selects a better relay node by maximizing the number of encounters with the destination node and minimizing the distance between the packet and the destination node. The MLProph algorithm [14] uses decision tree and neural network to calculate the probability of successful delivery of nodes by training multiple state information such as position, energy, speed and so on. The Geo-social algorithm [15] uses the geographical location history of users to mine the similarity between users, and makes forwarding decisions according to the similarity between users. Literature [16–18] uses node energy information to evaluate node's utility. The routing algorithm proposed in reference [16] relies on the energy of each node to calculate the forwarding probability of the node in order to maximize the network lifetime under the energy consumption constraint of each node. The ESW algorithm [17] establishes an evaluation function based on the speed and residual energy of the node, and calculates the forwarding utility of the node. The ProphetEA algorithm [18] forwards according to the residual energy of the node and the delivery rate defined by the Prophet algorithm. Many routing protocols in opportunistic networks are intersected and related, such as the above literature [8, 9, 14]. The ultimate goal is to integrate the various information of the node and measure the forwarding ability of the node in order to improve the network performance according to the different network environment.

The nodes in the opportunistic network are in a state of frequent movement, and each node divides the network into several sub-networks with independent communication opportunities. When the node cannot transmit the data packet to the destination node in time, it can only forward the data between the intermediate nodes, which may cause a large packet transmission delay. In order to further improve the delivery rate and transmission speed of the network, on the basis of information-assisted routing, researchers propose a hybrid routing scheme of utility and redundancy, but this scheme may have high network overhead. In addition, the packet is always transmitted in the direction of the high-utility node, which will cause the energy of the high-utility node to be consumed too much. Especially in the case of limited energy, high-utility nodes often participate in data reception and forwarding, the greater the energy consumption. The energy of high-utility will even be exhausted, which will aggravate the network intermittence and lead to a sharp decline in network performance. It has been proved that better routing performance can be achieved by selecting relays based on the social attribute information of nodes, so the focus of this paper is to make full use of the social utility of nodes to guide routing and further reduce network overhead. Secondly, balance the energy consumption of nodes in the data transmission process, so as to avoid the rapid exhaustion of energy of high-utility nodes.

Based on the above analysis, this paper proposes a multi-copy routing algorithm based on node utility and energy (ProEnergy algorithm), which makes full use of node utility to reduce network overhead and balance node energy consumption under the premise of guaranteeing better network performance. In Sect. 2, the design idea based on utility and energy algorithm is given and a network model based on social attributes is constructed. In Sect. 3, the calculation process of social utility of nodes is introduced. In Sect. 4, the main steps of the algorithm are given. In Sect. 5, the network performance of the four routing algorithms is compared by simulation. Finally, we summarize the full study in Sect. 6.

2 Design Idea of ProEnergy Algorithm

2.1 Motivation

The opportunistic network relies on the communication opportunities brought by the node movement to complete the data transmission task. In general, most of the nodes in the network are composed of intelligent terminal devices carried by people, and their social activities may follow some social characteristics [19]. Different routing protocols use different social metrics such as centrality, similarity, and community attributes to select appropriate relay nodes [20–22]. The relationship between people in the network is complex, the equipment is in a state of continuous movement with people's social activities, and the relationship between nodes has human characteristics. In the previous routing work, the influence of self-difference and dynamic variability of node relationship on routing packets is ignored, so a new utility metric is proposed to measure the social relations of nodes. In order to avoid the problem of excessive network overhead in multi-backup routing, the focus of the routing algorithm is to study how to make full use of node relationships to further reduce network overhead. In addition, in practical application scenarios, considering the fact that node communication needs energy support but node energy is limited, in the process of data forwarding, the social utility and residual energy of the node are jointly used to judge the forwarding capability of the node.

The design idea of the routing algorithm proposed in this paper is shown in Fig. 1. Firstly, the network model is constructed to maintain the contact information and information update mechanism of nodes, so as to provide sufficient preparation for data forwarding. Secondly, a routing scheme based on node utility and energy is designed, which mainly includes three steps: identifying intimate nodes, building relationship model and routing decision.

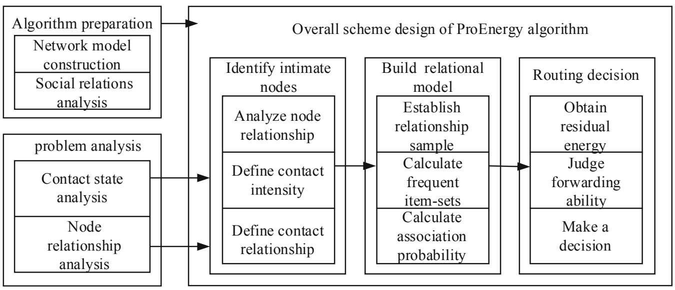


Fig. 1. Overall scheme design of ProEnergy algorithm

Each link of ProEnergy algorithm is described in detail in the following section. In Sect. 2.2, the node relationship is firstly analyzed, and then the contact relationship between nodes is defined to build a network model.

2.2 The Opportunistic Network Model Based on Social Attributes

In the opportunistic network based on social characteristics, the area of node activity is affected by human social activities, and human social activities are determined by social relations or the same social attributes. For example, nodes with certain social relationships (such as family, friends, colleagues) often appear in an environment with similar mobility and more contact time than unfamiliar relationships. Secondly, the node relies on social environmental life that there are many social relationships, and the relationship between nodes is a relative state. However, considering the influence of personality differences on the state of nodes, the boundaries of the relationship between nodes are not the same. In addition, with the increase of network time, the node relationship may change due to the influence of some force majeure factors (living environment and social pressure). Based on the above analysis of node relationship, the contact diagram with time attributes $G_t(V, E)$ is used to model the opportunistic network. Assuming that there are n nodes in the network, $V = \{v_1, v_2, \ldots, v_n\}$ represents the node set, and t_x represents a current moment when the network is running. The relevant definitions are as follows:

Definition 1 Neighbor Node Set: When the distance between nodes v_i and v_j is less than a certain threshold d, then the two nodes are neighbors of each other. The neighbor set of node v_i is represented as:

$$Nb(v_i) = \{v_j | D(v_i, v_j) < d\} \tag{1}$$

Where $D(v_i, v_j)$ is the communication distance between v_i and v_j, and d is the maximum communication distance between two nodes.

Definition 2 Intra and Inter Contact Time: The cumulative intra-contact time of nodes v_i and v_j is defined as $CT_{v_i}^{t_x}(v_j)$, and the cumulative inter-contact time of nodes v_i and v_j is defined as $OT_{v_i}^{t_x}(v_j)$.

The total duration of the network is represented by $T = \{t_0, t_1, t_2, \ldots, t_n\}$, where T is evenly divided into m different moments, the time gap between every two adjacent moments is Δt.$\boldsymbol{\delta}_{\Delta t}^{x}(v_i, v_j)$ represents the contact situation between v_i and v_j in the x-th time gap.

$$\boldsymbol{\delta}_{\Delta t}^{x}(v_i, v_j) = \begin{cases} 1 & v_j \in Nb(v_i) \\ 0 & v_j \notin Nb(v_i) \end{cases} \tag{2}$$

When $\boldsymbol{\delta}_{\Delta t}^{x}(v_i, v_j) = 1$, it means that v_i and v_j communicate with each other and can send data packets to each other. On the contrary, the communication link between them has been disconnected. $CT_{v_i}^{t_x}(v_j)$ and $OT_{v_i}^{t_x}(v_j)$ of nodes v_i and v_j are automatically updated according to the contact status of the node, that is:

$$\begin{cases} CT_{v_i}^{t_x}\left(v_j\right) = CT_{v_i}^{t_{x-1}}\left(v_j\right) + \Delta t & \delta_{\Delta t}^{x}\left(v_i, v_j\right) = 1 \\ OT_{v_i}^{t_x}\left(v_j\right) = OT_{v_i}^{t_{x-1}}\left(v_j\right) + \Delta t & \delta_{\Delta t}^{x}\left(v_i, v_j\right) = 0 \end{cases} \tag{3}$$

Definition 3 Contact Intensity: The contact intensity of nodes v_i and v_j is defined as the ratio of the intra-contact time to the inter-contact time.

$$CS_{v_i}^{t_x}\left(v_j\right) = \frac{CT_{v_i}^{t_x}\left(v_j\right)}{OT_{v_i}^{t_x}\left(v_j\right)} \tag{4}$$

The greater the contact intensity of the two nodes, the greater the chance for the two nodes to establish contact. It should be noted that because the contact between the two nodes is mutual, that is, $CS_{v_i}^{t_x}\left(v_j\right) = CS_{v_j}^{t_x}(v_i)$.

Definition 4 Average Contact Intensity: The average contact intensity of the node v_i at t_x is expressed as the average of the sum of the current contact intensity with the historical communication node, which represents the average level of the contact condition of the node at t_x.

$$ACS_{v_i}^{t_x} = \frac{1}{|N(v_i)|} \sum_{v_j \in N(v_i)} CS_{v_i}^{t_x}\left(v_j\right) \tag{5}$$

where $N(v_i)$ and $|N(v_i)|$ respectively represent the historical contact neighbor node set of v_i and the number of neighbor nodes in the set. $N(v_i)$ updates according to the contact condition of v_i at t_x.

$$N(v_i) = N(v_i)_{old} \cup \left\{ v_j | D(v_i, v_j) < d \cap v_j \notin N(v_i)_{old} \right\} \tag{6}$$

$N(v_i)_{old}$ represents the neighbor node set of v_i at t_{x-1}. Therefore, the historical neighbor node set of the node is composed of the historical neighbor node set of the previous moment and the newly joined neighbor nodes at the current time.

Definition 5 Contact Relationship: when $CS_{v_i}^{t_x}\left(v_j\right) > ACS_{v_i}^{t_x}$, it indicates that there is a strong contact relationship between v_i and v_j. On the contrary, when $CS_{v_i}^{t_x}\left(v_j\right) < ACS_{v_i}^{t_x}$, it indicates that there is a weak contact relationship between v_i and v_j.

3 Relation Model

3.1 Contact Intensity Prediction

The contact intensity of nodes is constantly updated with the increase of network time, and this value is a statistical value under the current time of the network. According to the average contact level of nodes in the network, the nodes with more contact opportunities with themselves are further identified. Based on the global time, the contact intensity between different nodes can be compared better, but it cannot better reflect the instantaneous change of the contact intensity between a single node pair. Therefore, we try to predict the change trend of contact intensity between single node pairs, excavate the change rule of contact intensity in a short time, and predict the future contact intensity of the two nodes, so as to guide the routing work more pertinently.

GM (1,1) prediction model is a small sample prediction model, which can use a small amount of historical data to predict the data of the next moment, and it has a better prediction effect on uncertainty [23–25]. Due to the mobility of nodes, the contact between nodes is constantly changing, and the contact or disconnection state of nodes is random. Figure 2 shows the contact situation of a node pair in the network. It is assumed that the node pair is in the contact state at t_x, while it may be in the contact state or disconnected state at t_{x-1}, and has no relation with the contact or disconnection state of the node at the past t_{x-1}, t_{x-2}, t_{x-3},…, t_{x-n}. Due to the sudden change and irregularity of node movement over a long period of time, with the increase of network time, the contact intensity of the two nodes at time t_{x-n} and t_{x+n} may change greatly. However, the contact intensity of the node in a short period of time fluctuates slightly, which may show a certain regularity. Therefore, using the changing law of contact intensity in a short period of time, predict the contact intensity in a short period of time in the future.

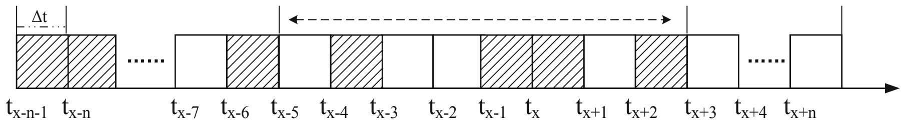


Fig. 2. The change diagram of contact and disconnection between nodes over time

The prediction process of contact intensity based on GM (1,1) is as follows:

(a) The contact intensity of the current time t_x and its previous history $w - 1$ time is obtained, and the total contact intensity of w time is obtained, from which the original contact intensity sequence is given.

$$CS^{(0)}_{v_i,v_j} = \left\{CS^{(0)}_{v_i,v_j}(1), CS^{(0)}_{v_i,v_j}(2), \cdots, CS^{(0)}_{v_i,v_j}(w)\right\} \tag{7}$$

(b) Accumulate the original contact intensity sequence to generate $CS^{(1)}_{v_i,v_j}$ sequence, and call $CS^{(1)}_{v_i,v_j}$ the 1-AGO (Accumulated Generating Operation) sequence of $CS^{(0)}_{v_i,v_j}$.

$$CS^{(1)}_{v_i,v_j} = \left\{CS^{(1)}_{v_i,v_j}(1), CS^{(1)}_{v_i,v_j}(2), \cdots, CS^{(1)}_{v_i,v_j}(w)\right\} \tag{8}$$

Where $CS_{v_i,v_j}^{(1)}(k) = CS_{v_i,v_j}^{(1)}(1) + CS_{v_i,v_j}^{(1)}(2) + \cdots + CS_{v_i,v_j}^{(1)}(k)$. A new sequence $Z_{v_i,v_j}^{(1)}$ is generated by processing the nearest neighbor mean of the cumulative sequence $CS_{v_i,v_j}^{(1)}$.

$$Z_{v_i,v_j}^{(1)} = \left\{ Z_{v_i,v_j}^{(1)}(2), Z_{v_i,v_j}^{(1)}(3), \cdots, Z_{v_i,v_j}^{(1)}(w) \right\} \tag{9}$$

$Z_{v_i,v_j}^{(1)}$ is called background value, $z_{v_i,v_j}^{(1)}(k) = \frac{z_{v_i,v_j}^{(1)}(k) + z_{v_i,v_j}^{(1)}(k-1)}{2}$.

(c) The first-order differential equation of the GM(1,1) model is established for the accumulated contact intensity sequence, and the corresponding whitening differential equation is as follows:

$$\frac{dCS_{v_i,v_j}^{(1)}}{dt} + aCS_{v_i,v_j}^{(1)}(t) = b \tag{10}$$

Where a is called the development coefficient, and b is called the gray effect, both of which are constant. The least square method is used to obtain the values of a and b, which satisfy the following equation:

$$[a, b]^T = \left(B^T B\right)^{-1} B^T Y \tag{11}$$

The value of Y and B is:

$$Y = \begin{bmatrix} CS_{v_i,v_j}^{(0)}(2) \\ CS_{v_i,v_j}^{(0)}(3) \\ \vdots \\ CS_{v_i,v_j}^{(0)}(w) \end{bmatrix}, B = \begin{bmatrix} -Z_{v_i,v_j}^{(1)}(2) & 1 \\ -Z_{v_i,v_j}^{(1)}(3) & 1 \\ \vdots & \vdots \\ -Z_{v_i,v_j}^{(1)}(w) & 1 \end{bmatrix} \tag{12}$$

(d) Expand the matrix to get the expression of a and b.

$$a = \frac{\sum_{k=2}^{n} CS_{v_i,v_j}^{(0)}(k) \sum_{k=2}^{n} Z_{v_i,v_j}^{(1)}(k) - (w-1) \sum_{k=2}^{n} CS_{v_i,v_j}^{(0)}(k) Z_{v_i,v_j}^{(1)}(k)}{(w-1) \sum_{k=2}^{n} (Z_{v_i,v_j}^{(1)}(k))^2 - (\sum_{k=2}^{n} Z_{v_i,v_j}^{(1)}(k))^2} \tag{13}$$

$$b = \frac{\sum_{k=2}^{n} CS_{v_i,v_j}^{(0)}(k) \sum_{k=2}^{n} (Z_{v_i,v_j}^{(1)}(k))^2 - \sum_{k=2}^{n} Z_{v_i,v_j}^{(1)}(k) \sum_{k=2}^{n} Z_{v_i,v_j}^{(1)}(k) CS_{v_i,v_j}^{(0)}(k)}{(w-1) \sum_{k=2}^{n} (Z_{v_i,v_j}^{(1)}(k))^2 - (\sum_{k=2}^{n} Z_{v_i,v_j}^{(1)}(k))^2} \tag{14}$$

(e) Solve the differential equation and get the discrete solution of the GM(1,1) model:

$$CS^{(1)}_{v_i,v_j}(k+1) = (CS^{(0)}_{v_i,v_j}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} \tag{15}$$

(f) Reverted to the original sequence, the prediction model is:

$$CS^{(0)}_{v_i,v_j}(k+1) = CS^{(1)}_{v_i,v_j}(k+1) - CS^{(1)}_{v_i,v_j}(k) \tag{16}$$

Bring Eq. 15 into Eq. 16 to get:

$$CS^{(0)}_{v_i,v_j}(k+1) = (1 - e^{a})(CS^{(0)}_{v_i,v_j}(1) - \frac{b}{a})e^{-ak}, k = 1, 2, \cdots, w \tag{17}$$

When $k = w$, $CS^{(0)}_{v_i,v_j}(k+1)$ is the predicted value of the contact intensity at the time t_{x+1} in the future.

The prediction method of contact intensity takes full account of the motion characteristics of nodes and is based on the motion rule of nodes in a short time. Mining the dynamic characteristics of contact intensity between nodes with the passage of time is more suitable for predicting the contact intensity between nodes in a short time in the future. GM (1,1) prediction model has high prediction accuracy for data samples with small scale, small fluctuation and persistence in the short term.

3.2 Build Relationship Samples

Nodes rely on the social environment to survive rather than independent individuals, and each node will have a relatively close social relationship. There may be more opportunities for contact between nodes and nodes with close relationships. Therefore, we believe that nodes and nodes with strong contact relationships have relatively close social relationships with a higher probability. In order to further analyze the node relationship to help reduce network overhead, first use the node contact relationship to establish the node relationship matrix.

Definition 6 Relational Model: Maintain a relationship matrix R based on the average contact intensity of the nodes, which is as follows:

$$\begin{pmatrix} r_1^{t_x} & r_{2\to1} & \cdots & r_{n\to1} \\ r_{1\to2} & r_2^{t_x} & \cdots & r_{n\to2} \\ \vdots & \vdots & r_j^{t_x} & \vdots \\ r_{n\to1} & r_{n\to2} & \cdots & r_n^{t_x} \end{pmatrix}$$

The diagonal position of R is $r_i^{t_x}$, where $r_i^{t_x}$ represents the average contact intensity $\widehat{ACS}_{v_i}^{t_x}$ of node v_i calculated according to the predicted contact intensity at t_x. The intersection of the i-th row and the j-th column represents the contact relationship $r_{j\to i}$ between node v_j and node v_i, and its value is as follows:

$$r_{j\to i} = \begin{cases} 1 & CS_{v_i}^{t_x}(v_j) > ACS_{v_i}^{t_x} \\ 0 & CS_{v_i}^{t_x}(v_j) < ACS_{v_i}^{t_x} \end{cases} \tag{18}$$

Where, $CS_{v_i}^{t_x}(v_j)$ is the predicted contact intensity between v_i and v_j at t_x. When $r_{j\to i} = 1$, it means that v_i and v_j have some kind of intimate relationship. When $r_{j\to i} = 0$, it means that there is no intimate relationship between v_i and v_j. It should be noted that the relational matrix R is an asymmetric matrix, $r_{i\to j} \neq r_{j\to i}$, the value of $r_{i\to j}$ is based on $ACS_{v_j}^{t_x}$, i.e.,

$$r_{i\to j} = \begin{cases} 1 & CS_{v_j}^{t_x}(v_i) > ACS_{v_j}^{t_x} \\ 0 & CS_{v_j}^{t_x}(v_i) < ACS_{v_j}^{t_x} \end{cases} \tag{19}$$

3.3 Relational Model

The Apriori algorithm was proposed by Agrawal in 1994. It is a classic algorithm for mining data association rules [26]. The most famous case is the ‘supermarket shopping basket’ case, which optimizes the placement of supermarket items by analyzing and mining supermarket shopping data to find out the set of items that people frequently buy [27]. Figure 3 shows the workflow of the model. It is mainly divided into two major steps. The first step is to find frequent item-sets. The purpose is to find frequent item-sets from the dataset. The second step is to discover association rules, that is, use conditional probabilities to find association rules in frequent item sets, and use potential rules to make predictions. Scan is called dataset selection process, and its main function is to filter the support, delete item-sets that do not meet the minimum support, and retain item-sets that meet the minimum support.

Aiming at the network environment proposed in this paper, a relationship model is established based on the idea of Apriori algorithm. The relationship matrix R of the node is used as the input of the model, and the social forwarding utility of the node is finally obtained to help routing to make forwarding decisions. The following describes the process of mining potential relationships between nodes based on the relationship model shown in Fig. 3. According to the input relationship matrix R, the model saves the close relationship of each node in the corresponding record of the relationship transaction sample Ω, that is, the name of the node with the close relationship is saved in the record. Each element in the candidate item-sets (C_1 and C_2) and frequent item-sets (L_1 and L_2) is a set, and the elements in the set are nodes in the network. Each element in C_1 and L_1 is called a 1-item-set. In the same way, each element in C_2 and L_2 is called a 2-item-set. 1-item-sets and 2-item-sets are also collectively called item-sets.

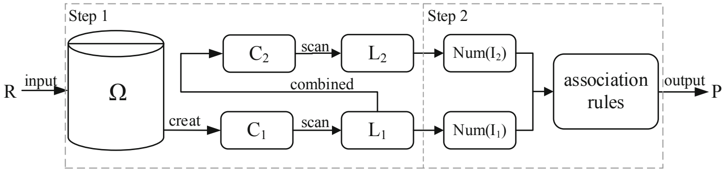


Fig. 3. Workflow chart of relationship mining model based on Apriori algorithm

It can be seen from Fig. 3 that the candidate 1-item-set is generated according to Ω, then $C_1 = \{\{v_1\}, \{v_2\}, \cdots, \{v_n\}\}$, which contains all the node items in the transaction sample, that is, contains a total of n 1-item-sets. The Scan process is divided into two stages. The first stage is to calculate the item-set frequency, which is based on candidate item-set C. Assuming that the item-set is represented by I, and the frequency of item-set I at t_x is defined as:

$$Fre_I^{t_x} = \partial^{t_x}(I) \tag{20}$$

$$\partial^{t_x}(I) = \sum_{k=1}^{n} F(I,k) \tag{21}$$

Where $F(I,k)$ is expressed as the relation function between the k-th node in the network and the nodes contained in the item set I. I is divided into 1-item-set I_1 and 2-item-set I_2. When $I = I_1$, the frequency of 1-item-set is calculated to obtain the number of intimate relationships of nodes at t_x. Assume that $I_1 = \{v_i\}$,where v_i is the i-th node in the network, then $F(I,k) = F(I_1,k)$ and $k \neq i$, its value is as follows:

$$F(I_1,k) = \begin{cases} 1, & R[k][i] = 1 \\ 0, & R[k][i] = 0 \end{cases} \tag{22}$$

When $I = I_2$, the frequency of 2-item-set is calculated to obtain the number of common close nodes of the two nodes in the item-sets at t_x. Assume that $I_2 = \{v_i, v_j\}$, where v_i and v_j are respectively the i-th and j-th nodes in the network, then $F(I,k) = F(I_2,k)$ and $k \neq i$ or $k \neq j$, its value is as follows:

$$F(I_2,k) = \begin{cases} 1, & R[k][i] = 1, R[k][j] = 1 \\ 0, & otherwise \end{cases} \tag{23}$$

In the first stage of the Scan process, the frequency of all candidate item-sets is finally obtained. In the second stage, the frequent item-sets L is obtained according to the minimum support T_M. At the current network time t_x, according to the relationship between the frequency of the item-set $Fre_I^{t_x}$ in the candidate item-sets C and the minimum support T_M, the frequent set L is obtained, and the basis is as follows:

$$\begin{cases} I \in L^{t_x} & Fre_I^{t_x} > T_M \\ I \notin L^{t_x} & Fre_I^{t_x} < T_M \end{cases} \tag{24}$$

L^{t_x} includes $L_1^{t_x}$ and $L_2^{t_x}$, which represent frequent 1-item-sets and frequent 2-item-sets at the current time t_x. The nodes included in the frequent 1-item-sets are called active nodes in the network. If a node has a close relationship with many nodes, then the node is more likely to contact other nodes, so the node can transmit the message to the destination node with a higher probability. In this process, T_M helps to reduce unnecessary data transmission and increase the number of valid data packets received by neighbor nodes. The corresponding two nodes included in the frequent 2-item-sets are called frequent contact node pairs in the network. If two nodes have more intimacy in common, the more likely it is to establish a connection between the two nodes. Therefore, the two nodes can establish a connection through a common intimacy with a higher probability. In this process, T_M helps to eliminate the node pairs with a lower degree of association in the network, and finally make the data packets flow to the node that has a higher probability of establishing contact with the destination node, achieving the purpose of further reducing network overhead and improving network performance.

The first step of the Apriori model is to finally obtain a pair of nodes in the network that can establish potential connections with a higher probability. Therefore, in the second step, it is necessary to further calculate the probability that a node can establish contact through a common close node. Assuming that at t_x, 2-item-set $\{v_i, v_j\} \in L_2^{t_x}$, it is necessary to calculate the probability that nodes v_i and v_j can establish a connection through the potential relationship. Let $p^{t_x}_{(v_i \Rightarrow v_j)}$ denote the probability that the node v_i establishes a connection with the node v_j through the latent relationship at t_x. The formula is defined as follows:

$$p^{t_x}_{(v_i \Rightarrow v_j)} = \frac{Num^{t_x}(v_i \cup v_j)}{Num^{t_x}(v_i)} = \frac{Fre^{t_x}_{\{v_i, v_j\}}}{Fre^{t_x}_{\{v_i\}}} \tag{25}$$

where $Fre^{t_x}_{\{v_i, v_j\}}$ represents the frequency of 2-item-set $\{v_i, v_j\}$ in $L_2^{t_x}$, and $Fre^{t_x}_{\{v_i\}}$ represents the frequency of 1-item-set $\{v_i\}$ in $L_1^{t_x}.Num^{t_x}(v_i \cup v_j)$ is the number of transactions that include v_i and v_j in the transaction sample, and $Num^{t_x}(v_i)$ is the number of transactions that include node v_i in the transaction sample. $p^{t_x}_{(v_i \Rightarrow v_j)}$ represents the conditional probability of node v_j in the transaction containing node v_i at t_x. This probability represents the possibility that the node v_i and the node v_j will establish a connection through a potential relationship, which can also be called the social forwarding utility of v_i being able to forward the data packet to v_j. It is worth noting that, according to the relational model, at t_x, the social forwarding utility $p^{t_x}_{(v_j \Rightarrow v_i)}$ of node v_j that can forward the data packet to node v_i and the social forwarding utility $p^{t_x}_{(v_i \Rightarrow v_j)}$ of node v_i that can forward the data packet to node v_j are not equal. The formula is as follows:

$$p^{t_x}_{\left(v_j \Rightarrow v_i\right)} = \frac{Num^{t_x}\left(v_i \cup v_j\right)}{Num^{t_x}\left(v_j\right)} = \frac{Fre^{t_x}_{\{v_i, v_j\}}}{Fre^{t_x}_{\{v_j\}}} \tag{26}$$

That is, the conditional probability of the occurrence of node v_i in the transaction containing node v_j is used as the forwarding utility of node v_j being able to forward the data packet to node v_i. When the 2-item-set $\{v_i, v_j\} \notin L^{t_x}_2$, then $p^{t_x}_{\left(v_i \Rightarrow v_j\right)} = 0$, it is considered that the probability of establishing potential association between node v_i and v_j is small and can be ignored.

4 ProEnergy Routing

4.1 Energy Weight Factor

Considering that the relay node has an association relationship with the destination node of multiple data packets at the same time, it can be called the key node in the network. The node needs to continuously receive and forward data, which will inevitably consume a lot of energy. In practical application scenarios, the communication between nodes requires energy support and the energy of the nodes is limited. If the energy factor of the node is not considered, it will stop working when the energy of the node is exhausted, resulting in the all paths through the node to be interrupted, which will greatly increase the transmission delay of the message. Therefore, in the ProEnergy routing scheme, the initial value of the node energy is set to 500. It is assumed that the node will consume one unit of energy when receiving and forwarding data packets, and the energy consumption required by the node when it is moving is ignored. Determine whether to participate in data forwarding according to the remaining energy of the node. When the energy of the node is less than a certain threshold, it only receives data packets that need to be transmitted to itself. Only when the energy of the node is greater than the threshold, the social forwarding utility of the node and the remaining energy need to be integrated to calculate the forwarding capacity of the node, as shown in the following formula:

$$P\left(v_i, v_j\right) = \lambda \times p^{t_x}_{\left(v_i \Rightarrow v_j\right)} + (1 - \lambda) \times C_i \tag{27}$$

Where $\lambda(\lambda \in [0, 1])$ is expressed as a weighting factor, $p^{t_x}_{\left(v_i \Rightarrow v_j\right)}$ is the social forwarding utility from v_i to v_j and $p^{t_x}_{\left(v_i \Rightarrow v_j\right)} < 1$. C_i is the remaining energy of node v_i. In addition, $P\left(v_i, v_j\right) \in (0, 1)$ and $p^{t_x}_{\left(v_i \Rightarrow v_j\right)} \in (0, 1)$, in order to eliminate the dimensional influence between the indicators, the dispersion standardization is used to standardize the remaining energy C_i of the node v_i, as shown in the following formula:

$$C_i = \frac{C_i - C_{min}}{C_{max} - C_{min}} \tag{28}$$

Where C_{min} represents the minimum remaining energy of the node in the network, and C_{max} represents the maximum remaining energy of the node in the network. This method maps the remaining energy of the node to [0,1] by linearly transforming the remaining energy of the node. After the energy is standardized, the forwarding capacity of the intermediate node is jointly calculated based on the remaining energy and social utility to guide the completion of the routing work. In this way, the energy consumption of nodes can be balanced to a certain extent, and nodes with too low energy can be prevented from participating in data packet forwarding. In order to better balance the problems of network performance and resource consumption, the value of the weighting factor λ will be explained in detail in the experimental part.

4.2 ProEnergy Routing Algorithm

The main steps of the ProEnergy routing algorithm proposed in this paper are shown in Table 1.

Table 1. Main steps of ProEnergy routing algorithm

ProEnergy Algorithm
Input: The packet M, the destination node d_M, the contact intensity of node pair GM(1,1)
Output: Whether to forward M
1. **FOR** each node $a, b, d_M \in V$ **DO**
2. **IF** a and b communicate with each other and a is the sender of M **THEN**
3. calculate $CS_a^{t_x}(b)$ based on $CT_a^{t_x}(b)$ and $OT_a^{t_x}(b)$,update GM(1,1)
4. **IF** b does not carry the copy of M **THEN**
5. **IF** the energy of a and b are greater than 50 **THEN**
6. **IF** b is the destination node of M **THEN**
7. update packet information, a forwards a copy of M to b
8. **ELSE**
9. calculate $CS_a^{t_x}(d_M)$ and $CS_b^{t_x}(d_M)$, update GM(1,1), calculate $CS_a^{t_x}(d_M)$, $CS_b^{t_x}(d_M)$, $CS_a^{t_x}(b)$, ACS_a , ACS_b , ACS_{d_M} , update R
10. calculate L_1 and L_2 based on R, calculate $p_{(a \Rightarrow d_M)}^{t_x}$ and $p_{(b \Rightarrow d_M)}^{t_x}$, normalize C_a and C_b , calculate $P(a,d_M)$ and $P(b,d_M)$
11. **IF** $P(a,d_M) < P(b,d_M)$ **THEN**
12. update packet information, a forwards a copy of M to b
13. **END IF**
14. **END IF**
15. **END IF**
16. **END IF**
17. **END IF**
18. **END FOR**

5 Simulation Experiment

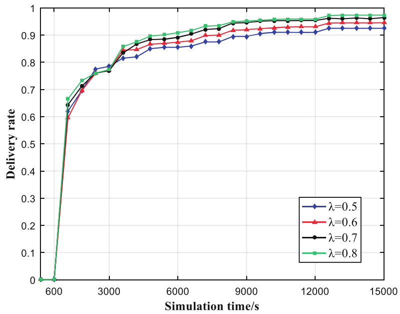
5.1 Simulation Environment and Parameter Setting

For performance evaluation, the experiment in this paper is based on the mobile opportunistic network simulator platform, which is described in detail in [28, 29]. The movement trajectory of the node uses the KAIST data set, which is provided by the Korea Institute of Science and Technology. The detailed introduction of the data set can be found in [30]. Other simulation parameters are shown in Table 2. This paper is based on the same parameters for simulation experiments. Where the Apriori algorithm is used to calculate the associated probability, the T_M value is set to 45. We use packet delivery rate, transmission delay, network overhead, energy consumption, and energy consumption to evaluate the performance of the routing algorithm, and compares it with the Prophet algorithm, the PageRank algorithm, and the GeoSocial algorithm.

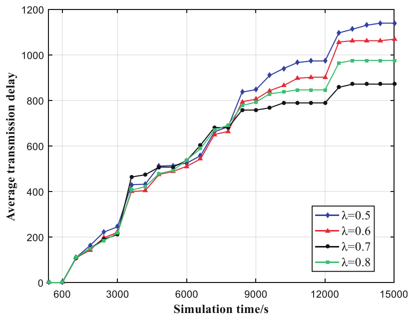
Table 2. Experimental parameter setting

Parameter	Value
Simulation field size (m^2)	600 × 600
Simulation time (s)	15000
Number of packets	200
Number of nodes	90
Maximum communication distance (m)	250
Send packets rate (ms)	100
Neighborhood search cycle (ms)	100

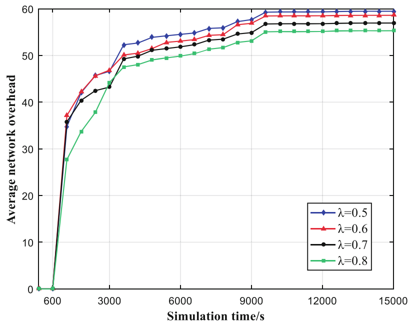
The influence of λ on network performance is analyzed in first. Figure 4(a) shows the delivery rate of λ with different values from 0.5 to 0.8. Observation shows that the delivery rate trends in the four cases are similar, and the delivery rate from λ = 0.8 to λ = 0.5 decreases sequentially, because the social utility influence of nodes gradually decreases. ProEnergy uses a multi-copy transmission method, so it has a higher transmission probability under different λ values. Figure 4(b) shows the transmission delay of λ with different values from 0.5 to 0.8. Observation shows that the transmission delay is the lowest when λ = 0.7. When λ = 0.5, the higher transmission delay is because when the value is small, more data packets are transmitted to nodes with relatively high energy but low correlation probability, which causes the delay to increase. Figure 4(c) shows the network overhead of λ with different values from 0.5 to 0.8. Observation shows that the network overhead of nodes from λ = 0.5 to λ = 0.8 decreases sequentially. When λ increases, the chances of those nodes with lower correlation probability but higher energy carrying data packets will be smaller, so the number of network copies will decrease. Through comprehensive comparative analysis, we believe that based on the above network parameters, when λ = 0.7, routing can obtain the best network performance.



(a) Delivery rate



(b) Transmission delay



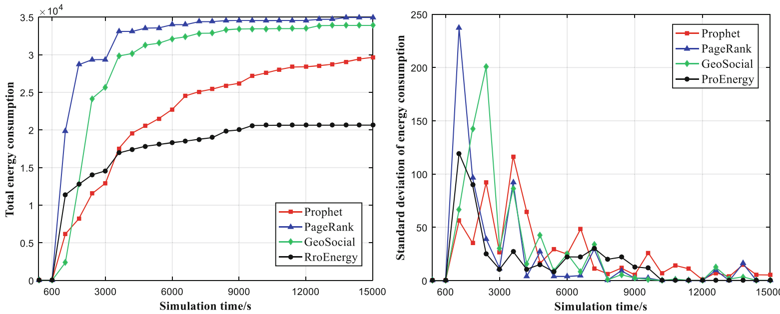
(c) Network overhead

Fig. 4. Experimental results under different λ

5.2 Experimental Results Analysis

Figure 5(a) shows the total energy consumption of 90 nodes in the network. It can be seen from the figure that the energy consumption of the four algorithms increases with the increase of simulation time. Before the simulation time is 3200 s, the total energy consumption of ProEnergy algorithm is greater than Prophet algorithm. Because the number of contacts of the nodes is limited at the beginning of the simulation, the Prophet algorithm only calculates the utility of the nodes within the communication range for forwarding. The ProEnergy algorithm calculates the social utility of nodes based on historical communication records. Due to the unstable contact between nodes in the initial period of the network, and the sufficient energy of the nodes, more nodes participate in data forwarding, resulting in higher energy consumption. However, as the simulation time increases, the total energy consumed by the ProEnergy algorithm increases slowly and tends to stabilize, while the Prophet algorithm continues to increase. At the end of the simulation, the energy consumption of the PageRank algorithm reached about 35,000, which was the highest consumption among the routing algorithms. The energy consumption of the GeoSocial algorithm is about 34,000, that of the Prophet algorithm is about 29,000, and the ProEnergy algorithm is about 21,000.

Figure 5(b) shows the variation of the standard deviation of node energy consumption as the simulation time increases. The remaining energy of the 90 nodes is output every 600 s, and then the standard deviation of the energy consumption of the 90 nodes in each 600 s is finally calculated. It can be seen from the figure that all algorithms have large fluctuations in the standard deviation at the beginning of the simulation (0 ~ 3000 s), which is caused by the uneven distribution of messages in the network. However, after 3000 s, it can be clearly seen that the standard deviation of ProEnergy algorithm is smaller than the other three routing algorithms, and the fluctuations are small and relatively stable, which indicates that ProEnergy algorithm can balance the energy consumption of the nodes, and does not overuse a certain node for data forwarding during data transmission, thus avoiding premature death of some efficient nodes in the network due to energy exhaustion.



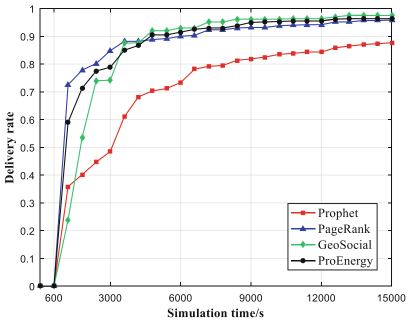
(a) Total energy consumption (b) Standard deviation

Fig. 5. Energy experiment results

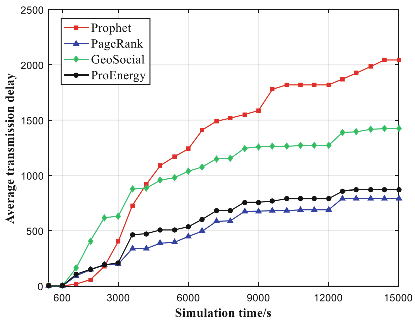
The delivery rate of the four algorithms is shown in Fig. 6(a). All of the four algorithms complete routing work in the way of multiple copies, so they all get higher delivery rates. It can be seen that PageRank algorithm is obviously higher than the other three algorithms before 7000 s. At the end of the simulation experiment, the delivery rate of PageRank is 0.98, GeoSocial is 0.97, and the ProEnergy algorithm is 0.96. The delivery rates of the three are similar. The delivery rate of Prophet algorithm is 0.89.

The average transmission delay of the four algorithms is shown in Fig. 6(b). It can be seen that the transmission delay increases with the increasing simulation time. In the initial simulation stage (0 ~ 3200 s), the transmission delay of Geosocial algorithm is the largest, but then the transmission delay of the Prophet algorithm increases with the increasing of the simulation time, because the Prophet algorithm updates the node utility when the node is in contact, and the high-utility node carries the data packet, which does not consider the energy problem of the node. After transferring the node to the high-utility node, the high-utility node may consume too much energy and aggravate the network intermittence, resulting in high network overhead. At the end of the simulation, the delay of Prophet is the largest, followed by Geosocial. ProEnergy is about 870 s. Compared with Prophet and Geosocial, the ProEnergy algorithm can effectively reduce the transmission delay.

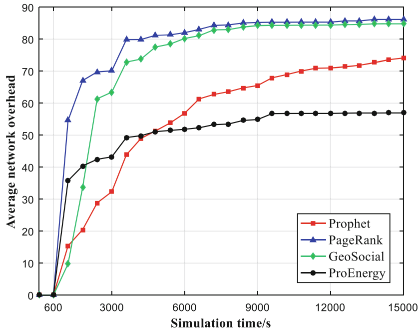
The average network overhead is shown in Fig. 6 (c). The average network overhead represents the ratio of the number of copies of all data packets generated in the network to the total number of packets. In the simulation experiment before 3200 s, the network overhead of ProEnergy algorithm is higher than Prophet algorithm, but after that, the ProEnergy algorithm is gradually stable, and the reason is similar to energy. At the end of the experiment, the number of copies of PageRank, GeoSocial and Prophet algorithms is higher than ProEnergy. Because PageRank, GeoSocial, and Prophet algorithms are based on different utility indicators, they forward when they encounter nodes with higher utility, and they do not consider the energy factor of the node, so it brings a higher number of data packet copies. In the ProEnergy algorithm, the relational utility of the node is used to mine the association probability of the node, and the energy of the node is also considered as the forwarding factor to effectively reduce the number of copies of data packets.



(a) Delivery rate



(b) Transmission delay



(c) Network overhead

Fig. 6. Experimental results of network performance

6 Conclusion

The nodes in the opportunistic network are in a state of continuous movement. When they cannot be delivered to the destination node, the packet will be transmitted between the intermediate nodes, which may bring high transmission delay. In order to further improve the network delivery rate and reduce the transmission delay, researchers proposed a hybrid routing mechanism of utility and redundancy, but this mechanism has the problem of high network overhead. In addition, the routing mechanism is based on the multi-copy transmission mode. In the actual environment, the energy of the node is limited, and the high-utility node will experience premature energy exhaustion due to excessive data transmission. In order to solve the above problems, the goal of the routing algorithm based on utility and energy proposed in this paper is to reduce the network overhead and balance the energy consumption of nodes while ensuring better network performance. First of all, the relationship between nodes is analyzed, the network model is constructed, and the Apriori algorithm is introduced to establish the relationship model to calculate the social forwarding utility of nodes. Secondly, add energy factors to balance the energy consumption of nodes and optimize routing performance. Finally, the proposed algorithm is simulated. The experimental results show that while ensuring a higher delivery rate and a lower average delay, it effectively reduces network overhead and balances the energy consumption of nodes, which provides a reference for future research on multiple backup routing algorithms.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China under Grants U1804164, 62072159 and U1404602.

References

1. Wu, D., Zhang, F., Wang, H., et al.: Security-oriented opportunistic data forwarding in mobile social networks. Future Gener. Comput. Syst. **87**(10), 803–815 (2017)
2. Yuan, P., Fan, L., Liu, P., et al.: Recent progress in routing protocols of mobile opportunistic networks: a clear taxonomy, analysis and evaluation. J. Network Comput. Appl. 62(C), 163–170 (2016)
3. Patel, D., Shah, R.: Improved PROPHET routing protocol in DTN. Int. Res. J. Eng. Technol. 503–509 (2016)
4. Pan, H., Chaintreau, A., Scott, J., et al.: Pocket switched networks and human mobility in conference environments. In: Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-Tolerant Networking, pp. 244–251. ACM, Philadelphia (2005)
5. Wang, X., Lin, Y., Zhang, S., et al.: A social activity and physical contact-based routing algorithm in mobile opportunistic networks for emergency response to sudden disasters. Enterp. Inf. Syst. 1–30 (2015)
6. Yuan, P., Ma, H., Fu, H.: Hotspot-entropy based data forwarding in opportunistic social networks. Pervasive Mob. Comput. **16**, 136–154 (2015)
7. Ayyat, S.A., Harras, K.A., Aly, S.G.: Interest aware PeopleRank: towards effective social-based opportunistic advertising. In: IEEE Wireless Communications and Networking Conference. IEEE (2013)

8. Hui, P., Crowcroft, J., Yoneki, E.: BUBBLE Rap: social-based forwarding in delay-tolerant networks. IEEE Trans. Mob. Comput. **10**(11), 1576–1589 (2011)
9. Qirtas, M.M., Faheem, Y., Rehmani, M.H.: A cooperative mobile Throwbox-based routing protocol for social-aware delay tolerant networks. Wirel. Netw. **2**, 1–13 (2020)
10. Wang, X., Lin, Y., Zhang, S., et al.: A social activity and physical contact-based routing algorithm in mobile opportunistic networks for emergency response to sudden disasters. Enterp. Inf. Syst. **11**(1–5), 597–626 (2015)
11. Boldrini, C., Conti, M., Jacopini, J., et al.: HiBOp: a history based routing protocol for opportunistic networks. In: IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks (2007)
12. Lindgren, A., Doria, A.: Probabilistic routing in intermittently connected networks. ACM SIGMOBILE Mobile Comput. Commun. Rev. **7**(3), 19 (2003)
13. Dhurandher, S.K., et al.: EDR: an encounter and distance based routing protocol for opportunistic networks. In: IEEE International Conference on Advanced Information Networking and Applications IEEE (2016)
14. Sharma, D.K., Dhurandher, S.K., Woungang, I., et al.: A machine learning-based protocol for efficient routing in opportunistic networks. IEEE Syst. J. **12**, 2207–2213 (2016)
15. Ying, Z., Zhang, C., Li, F., et al.: Geo-social: routing with location and social metrics in mobile opportunistic networks. In: IEEE International Conference on Communications, pp. 3405–3410 (2015)
16. Jang, K., Lee, J., Kim, S.K., et al.: An adaptive routing algorithm considering position and social similarities in an opportunistic network. Wirel. Netw. **22**(5), 1537–1551 (2016)
17. Jia, W.U., Chen, Z.: Reducing energy consumption priority selection of node transmission routing algorithm in opportunistic network. Adv. Inf. Sci. Serv. Sci. (2014)
18. Sobin, C.C., Raychoudhury, V., Saha, S.: An energy-efficient and buffer-aware routing protocol for opportunistic smart traffic management. In: Proceedings of the 18th International Conference on Distributed Computing and Networking, pp. 1–8 (2017)
19. Basaras, P., Iosifidis, G., Katsaros, D., et al.: Identifying influential spreaders in complex multilayer networks: a centrality perspective. IEEE Trans. Netw. Sci. Eng. **6**(1), 31–45 (2017)
20. Freeman, L.C.: Centrality in social networks conceptual clarification. Soc. Netw. **1**(3), 215–239 (1978)
21. Wei, K., Xiao, L., Ke, X.: A survey of social-aware routing protocols in delay tolerant networks: applications, taxonomy and design-related issues. IEEE Commun. Surv. Tutor. **16** (1), 556–578 (2014)
22. Ahmad, T., Li, X.J., Seet, B.C., et al.: Social network analysis based localization technique with clustered closeness centrality for 3D wireless sensor networks. Electronics **9**(5), 738 (2020)
23. Ding, S., Hipel, K.W., Dang, Y.G.: Forecasting China's electricity consumption using a new grey prediction model. Energy **149**(4), 314–328 (2018)
24. Yu, Q., Lyu, J., Jiang, L., et al.: Traffic anomaly detection algorithm for wireless sensor networks based on improved exploitation of the GM (1,1) Model. Int. J. Distrib. Sens. Netw. **12**(7), 2181256 (2016)
25. Kun, G., Qishan, Z.: Privacy preserving method based on GM (1,1) and its application to clustering. Grey Syst. Theor. Appl. **2**(2), 157–165 (2012)
26. Pratima, G., Pardasani, K.R.: A fast algorithm for mining multilevel association rule based on Boolean matrix. Int. J. Comput. Sci. Eng. **2**(3), 746–752 (2010)
27. Soni, A., Saxena, A., Bajaj, P.: A methodological approach for mining the user requirements using Apriori algorithm. J. Cases Inform. Technol. **22**(4), 1–30 (2020)

28. Yuan, P., Song, M.: MONICA one simulator for mobile opportunistic networks. In: 11th EAI International Conference on Mobile Multimedia Communications, pp. 21–32 (2018)
29. Ma, H., Zhao, D., Yuan, P.: Opportunities in mobile crowd sensing. Infocommunications J. **7** (2), 32–38 (2015)
30. Rhee, I., Shin, M., Hong, S., et al.: On the levy-walk nature of human mobility. IEEE/ACM Trans. Netw. (TON) **19**(3), 630–643 (2011)