MDPI

Article

An Effective Data Transmission Algorithm Based on Social Relationships in Opportunistic Mobile Social Networks

Yeqing Yan [1,2,†,‡] , **Zhigang Chen** [1,2,†,‡,*], **Jia Wu** [1,2,†,‡] **and Leilei Wang** [1,2,†,‡]

1   School of Software, Central South University, Changsha 410075, China; gretcheny@csu.edu.cn (Y.Y.); jiawu5110@163.com (J.W.); wll_1234@163.com (L.W.)
2   "Mobile Health" Ministry of Education-China Mobile Joint Laboratory, Changsha 410075, China
*   Correspondence: czg@csu.edu.com
†   Current address: Central South University, Changsha 410075, China.
‡   These authors contributed equally to this work.

check for
updates

Abstract: With the popularization of mobile communication equipment, human activities have an increasing impact on the structure of networks, and so the social characteristics of opportunistic networks become increasingly obvious. Opportunistic networks are increasingly used in social situations. However, existing routing algorithms are not suitable for opportunistic social networks, because traditional opportunistic network routing does not consider participation in human activities, which usually causes a high ratio of transmission delay and routing overhead. Therefore, this research proposes an effective data transmission algorithm based on social relationships (ESR), which considers the community characteristics of opportunistic mobile social networks. This work uses the idea of the faction to divide the nodes in the network into communities, reduces the number of inefficient nodes in the community, and performs another contraction of the structure. Simulation results show that the ESR algorithm, through community transmission, is not only faster and safer, but also has lower transmission delay and routing overhead compared with the spray and wait algorithm, SCR algorithm and the EMIST algorithm.

Keywords: opportunistic networks; opportunistic mobile social networks; community division; community structure contraction

1. Introduction

An opportunistic network [1] is a kind of delay-tolerant network which also has features of mobile ad hoc networks [2]. Unlike ad hoc networks, although the opportunistic network is also a multi-hop wireless network, there is never an end-to-end connection. However, there may be an end-to-end path in the opportunistic network at some point. Opportunistic networks are increasingly widely used, mainly in communication networks in military environments [3] and in wild animal tracking [4]. With an increasing number of vehicles equipped with wireless smart devices, vehicles driving on the road can communicate over short distances, which can form a dynamic, uneven density of nodes constituting a fast-moving wireless network of vehicles [5]. This is also an application scenario of the opportunistic network.

In opportunistic networks, the movement of nodes is random, so there is no persistent connection between any two nodes [6]. When information should be transferred in opportunistic networks, nodes should store information about data to be forwarded in their own cache, then forward the data to the nodes they encounter according to constraint rules. The nodes keep moving and forwarding messages until the target node obtains the information. This type of data transmission method is known

as the "storage-carry-and-forward" mechanism, which is the basic principle of data transmission and routing in opportunistic networks [7]. In the real world, people wear mobile phones, movement bracelets and smart watches which have mobile communication functions. The mobile terminals each person carries have a communication range within which information can be transmitted directly. However, because people are in motion constantly, the target node is not necessarily within the communication range of the current node. In this case, the node will keep information about the data, carry the data during movement, and pass the information to the node it thinks can contact the target node during the continuous movement.

With the advent of the era of big data and 5G, the amount of data that needs to be transmitted in the communication process has become very large, and social networks can no longer meet the demands of communication. We refer to this kind of opportunistic network in social network research as an opportunistic social network, which has both the social attributes of social networks and the data transmission characteristics of opportunistic networks. In this scenario, the routing needs "opportunity" for forwarding, e.g., people contacting each other by chance using mobile wireless devices, and here opportunistic social networks demonstrate their unique applicability. The message forwarding mechanism of opportunistic social networks does not require a stable end-to-end connection. Instead, we can choose the most appropriate node in the network structure to transmit information through. This means that an opportunistic network can also be established between strangers, where if the appropriate node appears, the information can be transferred. In other words, if people use Bluetooth or public Wi-Fi within a communication range, they can use the opportunistic social network to deliver messages. If there is no suitable node in the current communication range, the current node can save the information to the cache and wait to find an "opportunity" to transfer the information during movement.

As human beings take part in networked activities, the behavior of mobile nodes shows certain social features [8]. With the rapid popularization of portable mobile devices which have short distance communication functions, ever more users exploit these portable devices to contact and share data between each other. We define the mobile social network as a network structure that uses mobile devices to communicate and where the nodes in the network have social connections. References [9–11] point out that the social relationships that exist in human daily life are an important indicator of the performance of opportunistic routing. In the latest research on opportunistic networks, context information, node interests, and node social attributes are often used as variables to measure the performance of opportunistic network routing.

The biggest difficulty in opportunistic social networks is the problem of transmission delay and routing overhead when there is a large amount of information that needs to be forwarded. In many scenarios of opportunistic social mobile networks, data or messages take up lots of cache space in mobile equipment. This is because people use mobile devices during data transmission and there is no appropriate node around that can respond in a timely manner, which eventually caused transmission delay. Many existing algorithms choose routing based on the social characteristics of single nodes, and fail to consider that the aggregation of nodes in opportunistic social networks is particularly similar to that of communities in human life, which can reduce a large number of transmission delays and routing overhead.

To solve these problems, this work presents an effective data transmission algorithm based on social relationships in opportunistic social Networks (ESR). Because the load capacities of single nodes are limited in the big data environment, we propose a strategy to optimize the information transmission in the opportunistic social network by using the concept of the community [12], because the information transmission strategy through the community is stable, efficient and safe. In this algorithm, we divide the nodes into different communities in the network based on historical connection records [13] and propose a strategy of community structure reduction which divides the networks into several close-knit communities. The contraction of the community structures in the network can ensure fast and efficient information transmission within the community [14] and can also reduce the congestion caused by

popular nodes. In the process of information transmission, different numbers of data copies are allocated to each node through communities, which reduces network transmission delay and routing overhead, while improving the success rate of message transmission.

The contributions of this paper are listed as follows:

1. In opportunistic social networks, we propose an effective data transmission algorithm based on social relationships, which divides the network into several communities. By forwarding data through the community structures, the information transfer is faster and safer.
2. After dividing the nodes in the opportunistic social network into communities, we propose a method to reduce the community structure according to the features of nodes, to maintain the cohesion of the community and the high efficiency of data transmission.
3. According to the reduced community structure, we copy the information that needs to be transmitted several times and distribute different numbers of copies to different communities, which can reduce routing overhead while ensuring packet delivery ratio and transmission delay.
4. Using the simulation tool OMNET++, we analyze the performance of the ESR algorithm and compared it with some other algorithms. Our algorithm proposed in this paper has a higher packet delivery ratio, with lower transmission delay and routing overhead.

The rest of this paper is structured as follows. In Section 2, we describe and analyze related works. In Section 3, definitions and forwarding methods are presented for the improved Spray and wait algorithm in opportunistic social networks. The simulation results are presented in Section 4. The last section concludes the paper.

2. Related Works

Opportunistic network routing has been extensively studied, and various types of data-forwarding methods have been proposed. These methods were addressed for "storage-carry-and-forward" data transmission, with the goal of reducing the routing overhead and improving the data delivery rate of networks. In opportunistic networks, the current opportunistic network routing algorithms can be divided into normal types and social types, according to whether social information is required in data forwarding.

The normal type of data-forwarding strategy is proposed in the early stages of the opportunistic network. The common feature of this kind of routing algorithm is that they only consider the effect of node movements on data forwarding and can be applied to various application scenarios of opportunistic networks. However, the three important metrics of packet delivery ratio, transmission delay and routing overhead can only satisfy one or two kinds. The Epidemic algorithm proposed in reference [15] is a typical diffusion propagation algorithm, and its data-forwarding process is similar to the propagation process of infectious diseases. In this routing algorithm, nodes which carry information pass their data to all the nodes they meet. This algorithm has a high packet delivery ratio and small transmission delay in some scenarios, but the routing overhead ratio is large, and the algorithm's adaptability and scalability are poor. Reference [16] proposes a passive waiting algorithm called direct transmission. This algorithm is a data transmission mechanism where the source node is always waiting for its target node. The direct transmission algorithm has the lowest routing overhead ratio of the existing algorithms, but also has a very low packet delivery ratio. In reference [17], a routing algorithm based on probabilities was proposed that uses the historical contact records of nodes. The ProPHET algorithm calculates the probability of forwarding through the history of node interaction and proposes the possible path of route forwarding. This algorithm has high packet delivery ratio, but its routing overhead is also high. Reference [18] proposed a multi-copy routing algorithm, the Spray and Wait algorithm. The Spray and wait algorithm divides the forwarding process for information into two steps. In the spray process, the source node copies the information n times. It then forwards half the number of data copies stored in the node cache to the encountered nodes in

each transmission process. This algorithm has lower transmission delay than the direct transmission algorithm, but it can easily cause excessive flooding and data redundancy.

The social type of opportunistic network routing has been proposed in recent years. Due to the influence of human activities on the opportunistic network, the interactivity of nodes is an important factor that needs to be considered. The social type of opportunistic routing takes the social attributes of nodes into account when forwarding data. In reference [19], a labeling strategy algorithm based on community structures for data forwarding is proposed. This algorithm uses the information about the community structure in the nodes history records to label each node with information about its community. Reference [20] proposed a community-based forwarding algorithm, the Bubble Rap algorithm. According to the historical records of interactive information between nodes, we can calculate the vitality of nodes. The algorithm ranks all nodes according to their vitality level and calculates global ranking and local ranking in the community. In the forward process, the nodes transfer data to the higher ranked nodes in the global ranking, until the data is forwarded to a node in the same community where the target node is located. Then, according to the local ranking, the data is forwarded in the community until the target node receives the messages. The Bubble Rap algorithm adopts a single-copy strategy, with low packet delivery ratio and high transmission delay.

The main idea of the PeopleRank algorithm proposed by reference [21] is that nodes with a higher "PeopleRank" value are usually more "centered" in the network. The idea is to use this more stable social information to increase available partial contact information, to provide efficient data routing in opportunistic networks. The PeopleRank algorithm evaluates the node's actual movement and social interaction. It is proven that the algorithm can deliver the message with a near optimal success rate (i.e., close to Epidemic Routing), and at the same time reduce the number of message re-transmissions by 50% compared to Epidemic Routing. Reference [22] stated that existing social-aware routing protocols for pocket switched networks use the status information of nodes (for example, the history of past encounters) to infer information about the social structure of the network to optimize routing. SANE (Social-Aware D2D Relay Networks for Stability Enhancement) is the first routing algorithm that combines the advantages of both social-aware and stateless approaches. This algorithm shows the superiority of our proposed social-aware, stateless routing and interest-casting approaches compared with existing routing approaches.

Reference [9] proposes an algorithm based on the social characteristics of nodes, such as online social connections, interests, and contact history. The hit ratio and transmission delay of the entire network are not affected. Reference [10] proposes a multi-layer social-network-based routing method, ML-SOR, which measures the forwarding capacity of nodes and selects an effective forwarding node in terms of center of nodes, connection strength and link prediction compared with the nodes encountered. Good routing performance can be achieved at a low cost. dLife, proposed by reference [11], is a routing protocol of DTN(Delay tolerant network), which uses two utility functions of time varying connection duration and node importance to implement forwarding decisions. Experiments show that buffer size and message size can affect the performance of dLife routing.

Due to frequent node movement and sparse networks, the networked physical system of vehicles has the characteristics of opportunistic networks, which are seriously affected by intermittent connections. Reference [23] proposes a path-based TDOR(A Trajectory-Driven Opportunistic Routing Protocol) protocol for sparse networks. In this paper, the relay node is selected according to the distance to the track, rather than considering the destination as the standard for next-hop selection. Compared with other geographic routing protocols, this next hop selection strategy provides reliable and efficient messaging with high delivery rates and low transmission overhead. TDOR is more tolerant of network intermittency due to the unbiased movement of intermediate nodes. The results show that TDOR performs better than known opportunistic routing protocols and can achieve lower routing overhead for the same message transmission rate. Reference [24] illustrates that existing geographical routing usually contains hidden assumptions to obtain accurate location information, but this assumption is not true in real life. The accuracy of vehicle location information can determine the efficiency and

extensibility of geographic routing. A location error elastic geographic routing (ler-gr) protocol is proposed in reference [24]. The position prediction and correction technology of Kalman Filtering was used to predict the position of adjacent vehicles, and the position error of adjacent vehicles was evaluated by Rayleigh distribution error calculation technology. Simulation results show that the position error resilience of ler-gr is strong in real application environments.

With the amount of data that need to be transmitted in the communication process becoming very large, adopting the community-based mode for data forwarding can increase the packet delivery ratio and reduce the routing pressure of single nodes. The effective data transmission algorithm based on social relationships in opportunistic mobile social networks proposed by us addresses a data-forwarding method of multiple copies. While the problem of large numbers of data copies exists in traditional algorithms, our research proposes a community-based forwarding strategy [25], which can effectively reduce the number of data copies. Compared with other algorithms, this work does not adopt the method of transmitting messages between single nodes but proposes a scheme of transmitting information through efficient communities. This data-forwarding strategy not only improves the speed of data forwarding, but information forwarding in a stable community shows its superiority in terms of security.

3. System Model Design

3.1. Community Division Strategy

In opportunistic networks, within a certain time frame, the distribution of nodes is very similar to that of nodes in social networks, which means that the nodes in the opportunistic networks can be divided into different communities within a certain time frame. The community is a small network structure based on social relationships. The community has a close connection with the nodes inside, and the nodes outside the community are sparse. The accurate division of community structure can help improve the transmission speed and packet delivery ratio of information transmission and data forwarding. In this paper, we record the connection status of nodes in the network periodically by analyzing the social characteristics, connection attributes, and context information of nodes, to divide the structure of the networks in certain time intervals.

In fact, there are no independent community structures in the real world. This means that networks cannot be divided into several separate communities but are instead made up of a lot of overlapping areas of associations. A community can be seen as a collection of interconnected "small fully coupled networks" which can be called factions. We use the factional filtering algorithm to divide the communities. If we find the largest full-coupling subgraph of each part of the network, it can be used to find the connected subgraph of the faction, and accordingly divides the networks into many communities according to the factional divisions.

Definition 1. *The size of the fully coupled network. When we are dividing communities by the factional filtering algorithm, N is the probable size of the largest fully coupled network. The result of the final community division will be impressed by the value of N.*

Definition 2. *Fully coupled network. If there is a path for each node in a network to any other node, the network is called a fully coupled network.*

Definition 3. *Overlap matrix. Analogous to the adjacency matrix. Each column of the matrix corresponds to a faction. The value on the diagonal corresponds to the corresponding faction size, and the element on the non-diagonal represents the number of common nodes of the two factions.*

Given network $G = (V, E)$, $V = \{V_1, ..., V_n\}$ is a node-set including n nodes. Set N as the size of a possible fully coupled network. Starting from a set V_i, find all the N-size communities which contain node V_i, then delete node V_i and the edges that connect the node V_i. After we find all the

N-size factions, set $N = N - 1$ until N equals 2. So far, we have found the community structure of all the different factions in the network. In this paper, we use a 12-node diagram as an example to illustrate the community partition based on the factional filtering algorithm, with a value of N of 4.

When we use the factional filtering algorithm, an iteration regression method is usually adopted to find the factions in the network structure. If a vertex v is adjacent to all vertices of faction K, vertex v is an adjacent vertex of faction K. If the adjacent vertex of the faction is no longer contiguous to the other adjacency point of the faction, the adjacent node is referred to as the first type of adjacency node. If the adjacent vertex of the faction is connected to the other adjacent vertices of the faction, it is called the adjacent vertex of the second type.

For an initial node V_i, define two sets of nodes A and B, where set A is a set of nodes that contain node V_i and all nodes are connected to each other; Nodes in B connect to all nodes in A, and set B contains the first class adjacency node and the second type adjacency node. First, start iterating from node V_i, and initialize the set $A = \{V_i\}$, $B = neighbor [V_i]$. Move a node from set B to set A, then check that all the elements in B are still connected to all nodes in A, and delete nodes that are no longer connected to all nodes in A. Continuously move the node and delete some nodes which are not meeting the requirements. In the continuous process, if B is already empty before the size of A increase to N, or A or B are a set or subset that already exists, then stop counting and return to the first step—find another initial node V_i and start a new iteration. If B is not empty before the size of A reaches N and A is not a subset of existing factions, we get a new faction. Record the faction and then return to the first step to continue the search for new factions. We take this 12-node network as an example as shown in Figure 1 .
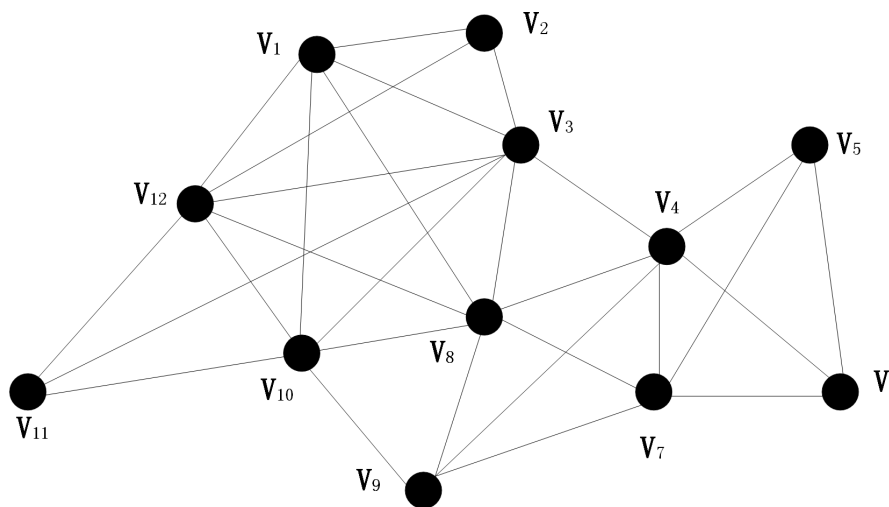

Figure 1. A 12-node network as an example, to show the division of nodes based on faction.

In this 12-node network structure, we set the faction size to 4. Suppose the initial node is V_1, and the set $B = neighbor [V_1] = \{V_2, V_3, V_8, V_{10}, V_{11}, V_{12}\}$. If we move node V_3 into set A, now $B = \{V_2, V_8, V_{10}, V_{11}, V_{12}\}$. At this point, V_1, V_2 are connected. Then, we move the node V_8 into the set A, and the nodes in the set B are still connected to the nodes in A, with only nodes V_{10} and V_{12} left. Because B is not empty when the size of the set A is less than N, this process can still go on. However, when we prepare to move V_3 to set A, we can find that there already exist a set or subset, so we do not save this group as a faction.

After finding all fully coupled network factions of size 4 and faction size N minus one, we continue to find all the appropriate factions until N is 2. In this way, we have all the fully-coupled networks within the network so far.

With the idea of iterative regression, all nodes in the network are traversed and the seven factions of this 12-node network are obtained as follows. In the Figure 2 below (top to bottom, left to right):

$G_1 = \{V_1, V_2, V_3, V_{12}\}$; $G_2 = \{V_1, V_3, V_8, V_{10}, V_{12}\}$; $G_3 = \{V_3, V_{10}, V_{11}, V_{12}\}$; $G_4 = \{V_4, V_5, V_6, V_7\}$; $G_5 = \{V_4, V_5, V_8, V_9\}$; $G_6 = \{V_3, V_4, V_8\}$; $G_7 = \{V_8, V_9, V_{10}\}$.
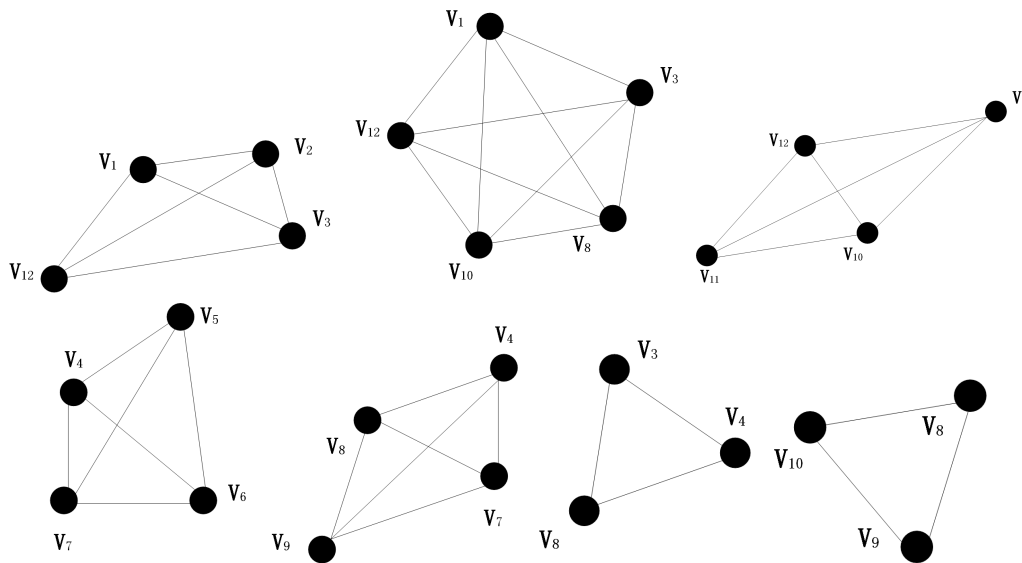
Figure 2. These seven graphs show the results of the division of factions after all nodes in the network have been traversed.

After we find all the factions in the network, we can find the overlapping matrix of these factions as shown in (1). Each column of the matrix corresponds to a faction, the value on the diagonal line corresponds to the size of corresponding faction, and the elements on the non-diagonal line represent the number of common nodes of each two factions.

$$
M_O = \begin{bmatrix}
4 & 3 & 2 & 0 & 0 & 1 & 0 \\
3 & 5 & 3 & 0 & 1 & 2 & 2 \\
2 & 3 & 4 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 4 & 1 & 1 & 0 \\
0 & 1 & 0 & 1 & 4 & 2 & 2 \\
1 & 2 & 1 & 1 & 2 & 3 & 1 \\
0 & 2 & 1 & 0 & 2 & 1 & 3
\end{bmatrix}
\tag{1}
$$

In this overlapping matrix, if the elements in the diagonal of the matrix are less than N or the elements not on the diagonal are less than N minus 1, set these elements equal to 0 and the rest of the elements to 1. The resulting matrix is the adjacency matrix of the N factional community structure as shown in (2). In this matrix, these remaining connected parts represent the community structure that meets the requirements.

$$
M_A = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\tag{2}
$$

According to the connectivity analysis of (2), the structure of the three communities in the network is shown in the Figure 3:
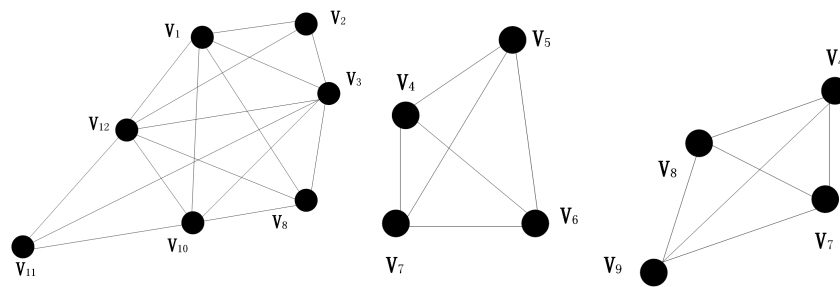
Figure 3. This picture shows the result of community division of the network structure shown in Figure 1.

We can use the above community partitioning strategy to divide the network into several community structures.

3.2. Community Structure Contraction Strategy

In opportunistic social networks, information transmission relies on the movements of the nodes and the importance of measuring the significance of nodes is pivotal. After dividing the network into several communities, the nodes in the opportunistic networks coalesce from scattered nodes into individual community structures. However, in these community structures, the significance of each node is also different.

It is important to measure the importance of nodes in community structures. If you do not measure the importance of popular nodes, it will probably lead to network congestion. Therefore, we measure the importance of nodes after the division of communities to reduce the number of inefficient nodes. The measurement of node importance is generally divided into two aspects, including the influence of nodes on information flow and the interaction between nodes.

The influence of nodes on information flow in networks is defined as the ratio of the number of paths through node V_i in all shortest paths in the network, to all shortest paths. If the network has S nodes, the influence of node on information flow is expressed as:

$$S_b\left(v_i\right) = \sum_{m<n} \frac{\gamma_{mn}\left(v_i\right)}{l_{mn}} \tag{3}$$

In the formula, $\gamma_{mn}\left(v_i\right)$ is defined to be the ratio of the number of paths through node v to all shortest paths in the network. l_{mn} represents the number of shortest paths between node v_m and v_n. In this network, for a given node v_i, the maximum influence of nodes on information flow is $[N-1][N-2]/2$, so the index can be normalized to:

$$S_B\left(v_i\right) = S_b\left(v_i\right)/\left[(N-1)\left[N-2\right]/2\right] \tag{4}$$

The importance of nodes is not only related to the number of edges connected, but also to the importance of connected nodes. Nodes can enhance their importance in the network by connecting to nodes that are more important. In a network of S nodes, A represents the network connection matrix, and $\alpha_1, \alpha_2, ..., \alpha_S$ represent the N characteristic values of A. Suppose α_1 is the main characteristic value of the matrix A. The corresponding characteristic vector is $e = (e_1, e_2, ..., e_s)^T$, and then:

$$\alpha e_i = \sum_{j=1}^{S} a_{ij} e_j \tag{5}$$

The characteristic vector indices of nodes can be defined as:

$$S_c\left(v_i\right) = \alpha^{-1} \sum_{j=1}^{S} a_{ij} e_j \tag{6}$$

The overall importance of the nodes is expressed as:

$$S_{IMP}(v_i) = \beta S_B(v_i) + \delta S_c(v_i) \tag{7}$$

In this formula, β and δ are variable parameters, representing the weight of $S_B(v_i)$ and $S_C(v_i)$ when we calculate the importance of nodes, respectively. A threshold is set here to measure whether the importance of nodes is within an acceptable range. If the node is less important than the threshold value τ, directly remove the node from its community. According to the actual application scenario, if the degree of bias in two aspects is reasonably evaluated, then determine the values of parameter β and δ, $\beta + \delta = 1$ (in this paper, $\beta = 0.5$, $\delta = 0.5$).

After removing the less important nodes, the community structure changes. However, the community can continue to shrink to reduce the size of the network. In this case, we use the cohesion of the nodes to recontract the community structure. The cohesion of nodes is an important index to measure whether the network structure can be recontracted, which is represented as:

$$\varphi[G] = \frac{1}{S * D} \tag{8}$$

$$D = \frac{\sum_{i \neq j \in V} d_{ij}}{S(S-1)} \tag{9}$$

When the structure shrinks, the cohesion of the network depends on the connectivity between nodes in the network and the number of nodes in the network. The connectivity between nodes can be measured by the average distances, that is, the average distances between all node pairs. By formula (5) and (6), the cohesion degree $\varphi[G]$ can be expressed as:

$$\varphi[G] = \frac{S-1}{\sum_{i \neq j} D_{ij}} \tag{10}$$

$S \geq 2$, D_{ij} represent the distance between node V_i and V_j. When there is only one node in the network, $\varphi[G] = 1$. $G * V_i$ represents the graph obtained after the node is contracted. The criticality of nodes in the network can be calculated according to the cohesion.

$$IM(v_i) = 1 - \varphi[G] / \varphi[G * V_i] \tag{11}$$

Under the same conditions, the larger the connectivity of nodes is, the smaller the number of nodes and the greater the network cohesion. If there is a node in the key position, the shortest path between many nodes will pass through this node, and the average network distances will be greatly reduced after the network shrinks. By (7) and (8), formula (11) can be expressed as:

$$IM(v_i) = \frac{S * D(G) - (S - k_i) * D(G_v)}{S * D(G)} \tag{12}$$

Now we calculate the criticality $IM(v_i)$ of each node after contraction according to formula (11). If the criticality $IM(v_i)$ of node V_i is bigger than the control parameter ε, it shows that the network structure is more compact after structure contraction. At this point, we can use a new node V_i' to replace the original node V_i and its associated nodes.

In this part, we propose a strategy for the contraction of community structures. We first remove some of the inefficient nodes in the community structure through node attributes. Then, the network is further contracted according to the influence of nodes on network cohesion, and we can get several close-knit community structures, which can improve the efficiency of information transmission and data forwarding. The simplified pseudo-code for this part is as follows (See Algorithm 1):

Algorithm 1 Community structure contraction algorithm

Input: subgraph G_n
Output: adjusted subgraph G_n
　1: For each group do //delete inefficient node
　2:　　　For each node V_i do
　3:　　　　　Calculate the importance $IM(v_i)$ of node V_i
　4:　　　　　　If $IM(v_i) < \tau$
　5:　　　　　　　Delete node V_i
　6:　　　　　End If
　7:　　　End For
　8: End for
　9: For each node pair $e = (v_i, v_j)$ // structure-shorten
　10:　　　Calculate distance D_{ij}
　11:　　　Calculate initial network cohesion $\varphi[G]$ and $\varphi[G * V_i]$
　12:　　　　If the value of V_i' is bigger than ε then
　13:　　　　　replace V_i and its neighbor with a new node V_i'
　14:　　　End If
　15: End For

3.3. Community-Based Data-Forwarding Algorithm

In the ESR algorithm, the data-forwarding process is divided into two phases. In the first stage, we will make several copies of the data we need to forward and send these copies to all the adjacent nodes. If an adjacent node is in the same community structure, the node propagates data to its own community structure, otherwise it will spread the data copies to all its adjacent nodes. In the second process, the information forwarding process ends when the target node receives the message and sends a receive-confirmation message.

Our algorithm proposed in this paper is based on the structure of the community, and we propose a method to reduce the structure of a community. Then, we copy a certain number of messages. After the contraction of the structure, the network scale decreases, and the nodes in the network structure have very close relationships. When information is transmitted and data is forwarded, the number of copies of the forwarded data can be determined according to the sociality of the nodes. The sociality of node is defined as:

$$C_i = \sum_{j=1, i \neq j}^{n} link(i, j) \tag{13}$$

In the formula, $link(i, j)$ is defined to be the connection status of node v_i in the community. If there is a node v_j connected to node v_i, then $link(i, j) = 1$, otherwise $link(i, j) = 0$. If node v_i is not in a defined community, the nodes that the link function compares is all the nodes in the entire network. Such a measure strategy distinguishes between nodes in the community and nodes outside the community.

In the ESR algorithm, the minimum number of copies required to achieve the expected time delay is independent of network size N and propagation range K, only depending on the number of nodes M in the network. Many studies have shown that the number of copies required only depends on the number of nodes, and the number of messages is an important index affecting the algorithm. Reference [26] gives the following formula, where M is the total number of nodes, and $H_n^r = \sum_{i=1}^{n} \frac{1}{i^r}$.

$$\left(H_M^3 - 1.2\right) L^3 + \left(H_M^2 - \frac{\pi^2}{6}\right) L^2 + \left(\alpha - \frac{2M - 1}{M(M - 1)}\right) L = \frac{M}{M - 1} \tag{14}$$

In the first process, sending n copies of the data to each node can easily lead to oversizing. Aiming at this problem, this paper proposes a copy-allocation method based on community division,

a reduction strategy of copy, and copy distribution based on the following formula, with L as the number of available messages that the node currently has:

$$
\begin{cases}
T_i = (L-1) \frac{c_i}{c_i + c_j} \\
T_j = (L-1) \frac{c_j}{c_i + c_j}
\end{cases}
\tag{15}
$$

In the second process, the source node sends information to all its neighbors. If the neighboring node is within the community, the copies of the data are propagated through the community. If not, the data information is forwarded to the meeting node. All data copies are allocated according to formula (15), and the forwarding process ends when the target node gets a copy of the data.

Because of the social feature of the opportunistic social network, in the first process, we distribute copies of data through the community. The number of data copies will not cause overflow, and the expected packet delivery ratio and transmission delay can also be achieved. In the second process, the data-forwarding speed and communication link stability in the community are very good. Relative to the transmission of information between single nodes, the ESR algorithm is transmitted through the community, which can improve the speed of information transmission. At the same time, since relatively stable community structures can be thought of as trusted institutions, message transmission within communities is safe.

The simplified pseudo-code for this part is as follows (see Algorithm 2):

Algorithm 2 Community based data-forwarding algorithm

Input: subgraph G_n
Output: adjusted subgraph G_n
  1: **For** each group **do** //delete inefficient node
  2:       For each node V_i do
  3:             Calculate the importance $IM(v_i)$ of node V_i
  4:                   If $IM(v_i) < \tau$
  5:                         Delete node V_i
  6:                   End If
  7:             End For
  8: End for
  9: For each node pair $e = (v_i, v_j)$ // structure-shorten
 10:       Calculate distance D_{ij}
 11:       Calculate initial network cohesion $\varphi[G]$ and $\varphi[G * V_i]$
 12:             If the value of V_i' is bigger than ε then
 13:                   replace V_i and its neighbor with a new node V_i'
 14:             End If
 15: End For

In this algorithm, time complexity is O(n). The information is transmitted by nodes through the reduced structure of communities, and the algorithm not only has fast transmission speed, but also has high transmission efficiency. We may then compare with the Spray and wait algorithm. In the spray process, the node continuously copies the data and assigns it to its neighbors. Its time complexity is O(n^2). It is the same as the SCR algorithm, because nodes are selected by social relationships to form a local cluster. In the EMIST algorithm, nodes transmit messages and save them as a list in their cache, so while its time complexity is O(n), it has higher routing overhead compared with our work.

4. Simulation

We implement the proposed ESR scheme in the OMNET++ simulator [27] and evaluate ESR by performance comparison with the EIMST algorithm, Spray and wait algorithm and SCR algorithm. The simulation time is 12 h, and the simulation area is 4300 × 1200 m in the communication area. Our background city is St Paul. The number of involved nodes is 1000. The node distribution is

random, and the transmission mode is broadcast. The maximum transmission distance of each node is 10 m, and the transmission speed of a node is 0.5∼1.5 m per second. The cache of each node is 10 MB, and the Packet size is 500 KB∼1 MB. The specific setting of this simulation scenario is shown in Table 1, and the setting of algorithm parameters is shown in Table 2.

Table 1. Environmental parameter of simulation settings.

Environmental Parameter	Settings
Simulation time/h	12
Simulation area	4300 × 1200 m
Background city	St Paul
Number of nodes	1000
Velcocity of a node/(m/s)	0.5 1.5
Transmit speed(KB/s)	250
Maximum transmission distance/m	10
Transmission mode	broadcast
Buffer size/MB	10
Packet size	500 KB 1 MB
Packet sending intervals/s	25 35

Table 2. The optimal parameter value of our algorithm.

Parameter	Value	Description
N	16	Faction setting parameter
τ	0.17	A threshold value that measures whether a node is important within a controllable range
ε	0.42	Parameter decides whether the community structure needs to shrink

We compared the performance of different routing algorithms in the same scenario and analyzed the effect of varying parameters in the ESR algorithm. In this study, ESR is compared with spray and wait [15], SCR [4], and EMIST [3], which are a classic routing algorithm and two new algorithms published in 2017, respectively. EMIST is an information transmission strategy based on socialization nodes, and SCR is an effective social-based clustering and routing scheme in opportunistic networks. These algorithms were analyzed by three performance indicators, namely packet delivery ratio, transmission delay and routing overhead [28]. The experiment shows that the effective data transmission algorithm based on social relationships in opportunistic mobile social networks has the best comprehensive performance when N = 16, τ = 0.17, and ε = 0.42.

4.1. Community Division Interval

In a community-based opportunistic social network, the role of the community division interval is important. In a period of time, it can be considered that the status of a node and communities in the network do not change, and data transmission can be stable. If the division interval is relatively small, because the community result is not stable enough, the guiding effect of community division on the routing algorithm is not good, and the packet delivery ratio is low. As the division interval increases, the packet delivery ratio increases correspondingly as shown in Figure 4. When the interval is selected as 900 s, the optimal packet delivery ratio is 80%. As the division interval continues to increase, the packet delivery ratio decreases. At 2700 s, the packet delivery rate is 70%, and the algorithm performance is worse than that at 900 s. This is because when the interval of community division is too long, the true node distribution state is not consistent with the assumed stable state, and previous community segmentation results do not represent the latest community structure well.
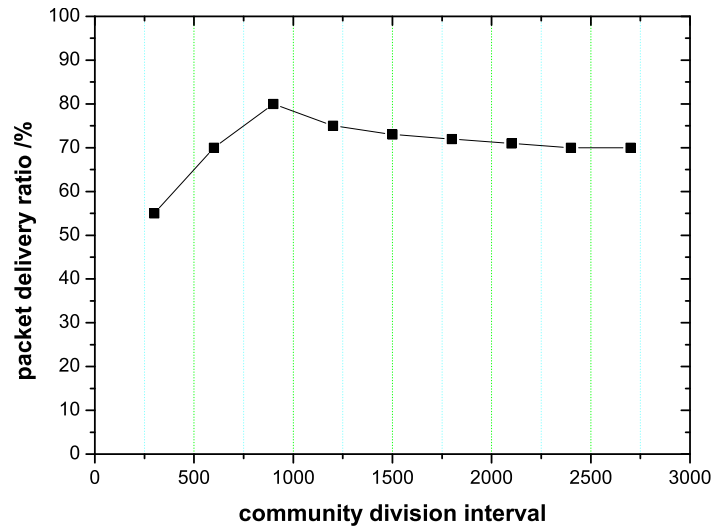
Figure 4. The influence of community division interval on packet delivery rate.

4.2. Effect of Node Cache and Node Numbers on Routing Algorithms

The ESR algorithm is compared with the three types of classical algorithm mentioned above to verify its performance. This study focuses on the following parameters:

1. **Packet delivery ratio**: This parameter refers to the probability that a packet sent from the source successfully reaches the target within a certain time.
2. **Average end-to-end on delay**: This parameter comprehensively evaluates the delay caused by packet routing, waiting delay and transmission delay.
3. **Routing overhead**: This parameter shows the overhead between two nodes when information is transmitted.

Figure 5 shows the packet delivery ratio of the ESR algorithm, EIMST algorithm, Spray and wait algorithm and the SCR algorithm when the node cache sizes and number of nodes change. Compared with the three other algorithms, the packet delivery ratio in ESR is the best. The delivery ratio in ESR is more than 0.93 when cache is 40 M. The reason is that nodes transfer information through communities in the ESR algorithm, and message transmission speed in communities is faster than between single nodes. As the process of forwarding information ends if you can find the community with the target node, we can quickly complete the data transmission. In the EIMST algorithm, nodes calculate the encounter probability in the network. This approach uses cooperative nodes when transferring information but does not establish a good management mechanism in the cache space. Thus, the packet delivery ratio (0.8 average) is lower than for ESR. Spray and wait takes a multi-copy approach, sending many copies of data to the communication area, and then not having enough cache to store messages. When there is some new information starting to be transferred, it is easy to cause the discarding of previous information, and so the packet transmission rate is low. Therefore, delivery ratio in this algorithm is low (0.4–0.48). Due to the formation of clusters in the SCR algorithm requiring a process which needs lots of nodes and takes a long time, the packet delivery ratio of SCR is lower than for ESR. Above all, the ESR scheme improves the delivery ratio over the traditional algorithms, on average.
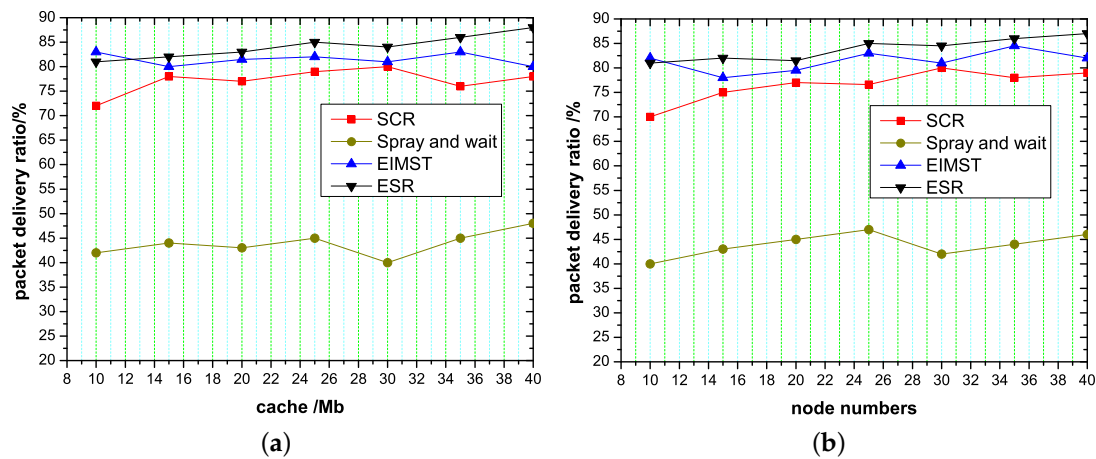
Figure 5. The packet delivery ratio of the ESR algorithm, EIMST algorithm, Spray and wait algorithm and the SCR algorithm at different node cache sizes and numbers of nodes. (**a**) The effect of node cache on packet delivery ratio; (**b**) The effect of node numbers on packet delivery ratio.

Figure 6 shows the delay on average in these algorithms. As shown in Figure 6, the average delay in the spray and wait algorithm is high (more than 105), because the spray and wait algorithm is based on the policy of multiple copies, which can easily cause data redundancy and loss. The SCR algorithm analyzes the social attributes of nodes to form a local cluster. As the node cache and node number increases, the cluster process becomes better, and the transmission delay decreases. However, it is still higher than the ESR algorithm. The EIMST algorithm is better than the spray and wait algorithm, because all neighbors are selected by nodes, reducing data copies by a large number. However, the transmission delay is still large. In the ESR algorithm the average delay is 50, because data is delivered among nodes through communities. Therefore, the data can share the cache of nodes in the same community during transmission. The proposed ESR algorithm reduces transmission delay compared with the Spray and wait algorithm and the SCR algorithm, on average.
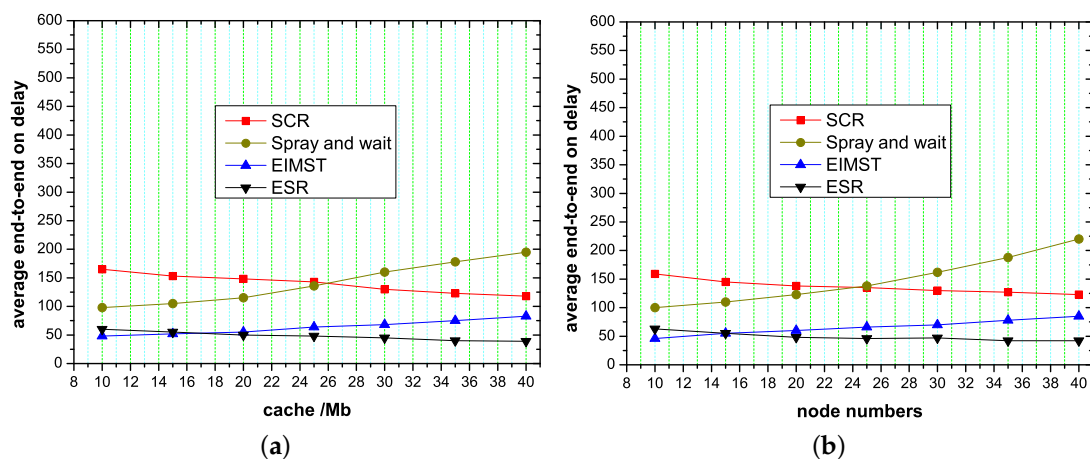
Figure 6. The average end-to-end transmission delay of ESR algorithms, EIMST algorithm, Spray and wait algorithm and SCR algorithm when at different node cache sizes and number of nodes. (**a**) The effect of node cache on transmission delay; (**b**) The effect of node numbers on transmission delay.

Figure 7 explains the overhead on average. According to Figure 7, the overhead ratio of the spray and wait algorithm is the highest. In SCR, each node only forwards a copy of the message to a node that has the target node as a member of the cluster, but the clustering process needs time. Therefore, although the routing overhead is decreased, the routing overhead is still greater than that of the ESR algorithm. The routing overhead of ESR and EIMST remains at 25–65 because nodes spray a lot of

redundant data. As the cache and the number of nodes increases, this state will not change. In the ESR algorithm, all communities can help to share the routing pressure, which is less expensive than the above three algorithms.
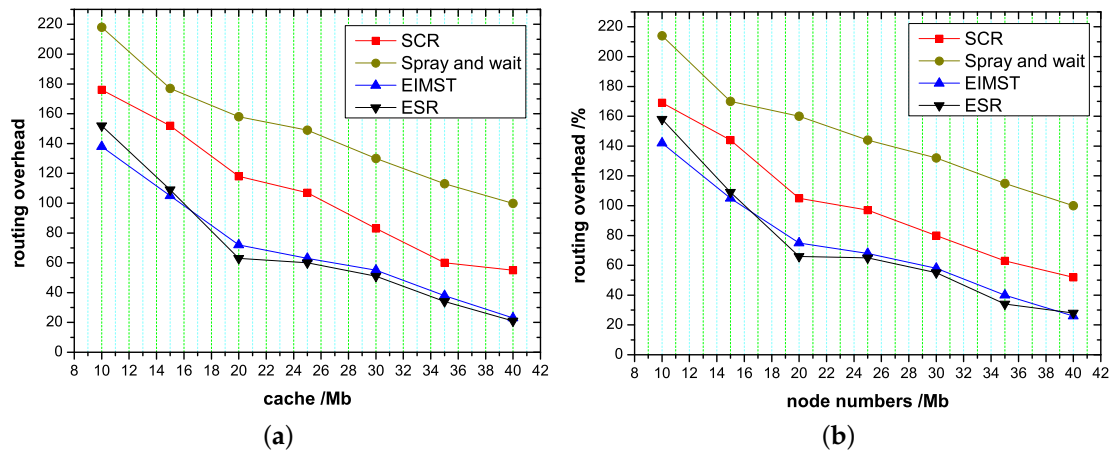
Figure 7. The average routing overhead of the ESR algorithm, EIMST algorithm, Spray and wait algorithm and SCR algorithm when at different node cache sizes and number of nodes. (**a**) The effect of node cache on routing overhead; (**b**) The effect of node cache on routing overhead.

5. Conclusions

In this paper, we learned some characteristics of opportunistic social networks and proposed an effective data transmission algorithm based on social relationships. In this algorithm, we use the social relationships of nodes to divide nodes into several communities based on the connection attributes between nodes, then judge the importance of nodes according to the influence of nodes on the network and delete some inefficient nodes. Then, the node's criticality is used to further reduce the community structure after reducing inefficient nodes. After the network is divided into several tight communities, we use the characteristics of the various communities in the network to allocate the number of data copies and transmit them through the community. This algorithm divides the nodes in the network by means of social relations and can effectively reduce the inefficient nodes in the network. Through close community information transmission, the efficiency of information transmission is ensured and the goal of low delay is realized. At the same time, to avoid congestion caused by popular nodes, the inefficient nodes are evaluated through the influence of nodes on information flow and the interaction between nodes, which reduces the network overhead. The routing and transmission of opportunistic social networks will still be the focus of this field in present and future research. In our future work, we will improve our data-forwarding algorithm based on more social network behaviors and further study security and privacy in social routing.

Author Contributions: Conceptualization, Y.Y. and J.W.; Methodology, Y.Y.; Software, J.W.; Validation, Y.Y.; Formal Analysis, Y.Y. and J.W.; Investigation, Y.Y. and L.W.; Resources, Z.C. and J.W.; Data Curation, Z.C. and J.W.; Writing—Original Draft Preparation, Y.Y.; Writing—Review & Editing, J.W. and L.W.; Visualization, J.W.; Supervision, Z.C.; Project Administration, J.W.; Funding Acquisition, Z.C.

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

References

1. Wu, J.; Chen, Z. Sensor communication area and node extend routing algorithm in opportunistic networks. *Peer-to-Peer Netw. Appl.* **2016**, *11*, 90–100. [CrossRef]
2. Roy, A.; Deb, T. Performance Comparison of Routing Protocols in Mobile Ad Hoc Networks. *Int. J. Eng. Sci. Technol.* **2018**, *2*, 279.
3. Mukherjee, A.; Basu, S.; Roy, S.; Bandyopadhyay, S. Developing a coherent global view for post disaster situation awareness using opportunistic network. In Proceedings of the 2015 7th International Conference on Communication Systems and Networks, Bangalore, India, 6–10 January 2015; pp. 1–8.
4. Ghumare, S.S.; Labade, R.P.; Gagare, S.R. Rare Wild Animal Tracking in the Forest area with Wireless Sensor Network in Network Simulator-2. *Int. J. Comput. Appl.* **2016**, *133*, 1–4.
5. Kaiwartya, O.; Abdullah, A.H.; Cao, Y.; Altameem, A.; Prasad, M.; Lin, C. T.; Liu, X. Internet of Vehicles: Motivation, Layered Architecture, Network Model, Challenges, and Future Aspects. *IEEE Access* **2017**, *4*, 5356–5373. [CrossRef]
6. Jia, W.U.; Chen, Z.; Zhao, M. Effective information transmission based on socialization nodes in opportunistic networks. *Comput. Netw.* **2017**, *129*, 297–305.
7. Zeng, F.; Zhao, N.; Li, W. Effective Social Relationship Measurement and Cluster Based Routing in Mobile Opportunistic Networks. *Sensors* **2017**, *17*, 1109. [CrossRef] [PubMed]
8. Wu, J.; Chen, Z. Human Activity Optimal Cooperation Objects Selection Routing Scheme in Opportunistic Networks Communication. *Wirel. Pers. Commun.* **2017**, *95*, 3357–3375. [CrossRef]
9. Ciobanu, R.I.; Marin, R.C.; Dobre, C.; Cristea, V.; Mavromoustakis, C.X. ONSIDE: Socially-aware and Interest-based dissemination in opportunistic networks. In *Network Operations and Management Symposium*; IEEE: Piscataway, NJ, USA, 2014; pp. 1–6.
10. Socievole, A.; Yoneki, E.; De Rango, F.; Crowcroft, J. ML-SOR: Message routing using multi-layer social networks in opportunistic communications. *Comput. Netw.* **2015**, *81*, 201–219. [CrossRef]
11. Kumiawan, Z.H.; Yovita, L.V.; Wibowo, T.A. Performance analysis of dLife routing in a delay tolerant networks. In Proceedings of the 2016 International Conference on International Conference on Control, Electronics, Renewable Energy and Communications, Bandung, Indonesia, 13–15 September 2017; pp. 41–46.
12. Wang, X.; Chen, M.; Kwon, T.; Jin, L.; Leung, V. Mobile traffic offloading by exploiting social network services and leveraging opportunistic device-to-device sharing. *Wirel. Commun. IEEE* **2014**, *21*, 28–36. [CrossRef]
13. Wu, J.; Chen, Z.; Zhao, M. Weight distribution and community reconstitution based on communities communications in social opportunistic networks. *Peer-to-Peer Netw. Appl.* **2018**, 1–12.. [CrossRef]
14. Yuan, L.; Qin, L.; Zhang, W.; Chang, L.; Yang, J. Index-based Densest Clique Percolation Community Search In Networks. *IEEE Trans. Knowl. Data Eng.* **2018**, *30*, 922–935. [CrossRef]
15. Neena, V.V.; Rajam, V.M.A. Performance analysis of epidemic routing protocol for opportunistic networks in different mobility patterns. In Proceedings of the 2013 International Conference on International Conference on Computer Communication and Informatics, Coimbatore, India, 4–6 January 2013; pp. 1–5.
16. Kim, J.B.; Lee, I.H. Non-Orthogonal Multiple Access in Coordinated Direct and Relay Transmission. *IEEE Commun. Lett.* **2015**, *19*, 2037–2040. [CrossRef]
17. Wan, B. An optimized Prophet delay tolerant network routing algorithm based on social environment. *Wirel. Internet Technol.* **2017**.
18. Sisodiya, S.; Sharma, P.; Tiwari, S.K. A new modified spray and wait routing algorithm for heterogeneous delay tolerant network. In Proceedings of the 2017 International Conference on International Conference on I-Smac, Coimbatore, India, 10–11 February 2017; pp. 843–848.
19. Han, X.; Yun, L.; Zhang, Z.; Li, J.; Xiong, F. A multi-label community discovery algorithm based on the community kernel. In Proceedings of the 11th International Knowledge Management in Organizations Conference on The changing face of Knowledge Management Impacting Society, Hagen, Germany, 25–28 July 2016; pp. 1–5.
20. Jaimini, P.; Patel, R. Efficient Routing using Bubble Rap in Delay Tolerant Network. *Int. J. Comput. Appl.* **2016**, *137*, 16–19. [CrossRef]
21. Du, W. Research on Individual Influence in Social Networking Services Based on MapReduce. *J. Inf. Comput. Sci.* **2015**, *12*, 4715–4723. [CrossRef]

22. Zhang, M.; Chen, X.; Zhang, J. Social-aware relay selection for cooperative networking: An optimal stopping approach. In Proceedings of the 2014 IEEE International Conference on Communications, Sydney, Australia, 10–14 June 2014; pp. 2257–2262.

23. Cao, Y.; Han, C.; Zhang, X.; Kaiwartya, O.; Zhuang, Y.; Aslam, N.; Dianati, M. A Trajectory-Driven Opportunistic Routing Protocol for VCPS. *IEEE Trans. Aerosp. Electron. Syst.* **2018**. [CrossRef]

24. Kasana, R.; Kumar, S.; Kaiwartya, O.; Yan, W.; Cao, Y.; Abdullah, A.H. Location error resilient geographical routing for vehicular ad-hoc networks. *IET Intell. Transp. Syst.* **2017**, *11*, 450–458. [CrossRef]

25. Zhang, X.K.; Ren, J.; Song, C.; Jia, J.; Zhang, Q. Label propagation algorithm for community detection based on node importance and label influence. *Phys. Lett. A* **2017**, *381*, 2691–2698. [CrossRef]

26. Spyropoulos, T.; Psounis, K.; Raghavendra, C.S. Spray and wait: An efficient routing scheme for intermittently connected mobile networks. In Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-Tolerant Networking, Philadelphia, PA, USA, 22–26 August 2005; pp. 252–259.

27. Hornig, R. An overview of the OMNeT++ simulation environment. In Proceedings of the International Conference on Simulation TOOLS and Techniques for Communications, Networks and Systems & Workshops, Marseille, France, 3–7 March 2008; p. 60.

28. Wu, J.; Chen, Z.; Zhao, M. Information Transmission Probability and Cache Management Method in Opportunistic Networks. *Wirel. Commun. Mob. Comput.* **2018**. [CrossRef]