

Complex Contagions Models in Opportunistic Mobile Social Networks

Yunsheng Wang

Dept. of Computer Science
Kettering University, Flint, MI 48504

Jie Wu

Dept. of Computer and Info. Sciences
Temple University, Philadelphia, PA 19122

Abstract—Information contagions is one of the key applications in opportunistic mobile social networks (OMSNs). Most of the recent work only considers the simple contagions, in which the “infected” node will infect each of its opinion-free neighbors through simple contact. However, when the behaviors, beliefs, or preferences spread through the social contact, the willingness to participate may require independent affirmation or reinforcement from multiple sources. In this paper, we first review two complex contagions schemes proposed recently: *ring lattice-based complex contagions (RLCC)* and *grid-based complex contagions (GCC)* approaches. However, these two approaches do not consider the hierarchical structure in social network. Then, we present three novel hierarchical complex contagions models: *tree-based complex contagions (TCC)*, which is a binary tree model; *clique-based complex contagions (CCC)*, in which there is a k -clique and other nodes connect to every node in the k -clique, and *hypercube-based complex contagions (HCC)*, which is a balanced hypercube model. Extensive analysis and simulations are conducted in comparison to these approaches. CCC has the smallest delivery delay when there is only one message, while HCC has the best performance when the number of messages increases.

Keywords—Bisection width, complex contagions, diameter, hypercubes, node degree, opportunistic mobile social networks, trees.

I. INTRODUCTION

Most social information, collective behaviors, and diseases spread through social contacts. In an opportunistic mobile social network (OMSN), users walk around and communicate with each other when they are in each other’s communication range. Due to the uncertainty and intermittent connectivity of OMSNs, the technologies applied in traditional Internet or wireless networks are not suitable for OMSNs.

Information dissemination is one of the key applications in OMSNs. However, most of the recent work [1–3] only considers simple contagions in OMSNs, which means that the disease infection or information propagation only requires one “activated” source to infect the opinion-free neighbors. However, in the real world, spreading of beliefs or behaviors is slower, the willingness to participate may require independent affirmation or reinforcement from multiple sources. For example, when a new product is released, you may need to get information from different sources (e.g., friends, neighbors, reviewers, etc) to convince yourself to purchase it. Therefore, a new concept, *complex contagions* [4], is introduced. In the spread of complex contagions, preferences, behaviors, and beliefs spread via social contact with multiple adopters.

Figure 1 illustrates the concept of complex contagions in a simple OMSN. Fig. 1 shows an a -contagion network, where a is the number of activated nodes required to trigger the adoption. When $a = 1$, it is a simple contagions. If node

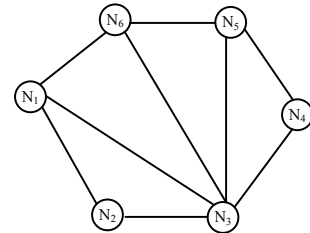


Fig. 1. An illustration of an a -complex contagions in a social network.

N_1 is the initial activated node, it only takes 2 steps to active all other nodes. In the first step, N_2 , N_3 , and N_6 can receive the information from N_1 . Then, in the second step, N_4 and N_5 can be activated by N_3 . In a 2-complex contagions, if the initial activated nodes are N_1 and N_2 , it needs 4 steps to spread the information to all other nodes. N_3 has two activated neighbors N_1 and N_2 . Hence, after the first step, N_3 becomes active. In the second step, N_6 can be activated by N_1 and N_3 . In the third step, the information can be spread to N_5 by N_3 and N_6 . Finally, N_3 and N_5 can disseminate the information to N_4 . In a 3-complex contagions with the initial activated nodes: N_1 , N_2 , and N_3 , it is impossible to activate any other node. Therefore, information diffusion in complex contagions, which requires independent affirmation from multiple sources, is more challenging than simple contagions.

Recently, the researchers proposed two complex contagions models in social networks: *ring lattice-based complex contagions (RLCC)* [4] and *grid-based complex contagions (GCC)* [5]. In [4], Centola and Macy show the analytic and simulation results based on the Watts and Strogatz’s small world model [6], which is a ring lattice-based approach. In this model, complex contagions spread using mostly the local links (or strong ties), and are going to be significantly slower in many settings, which requires a substantially large number of random ties to even create one single ‘bridge’ to diffuse the contagion. In [5], Ghasemiesfeh, Ebrahimi, and Gao analyze the complex contagions problem in the Kleinberg’s small world model, which is a grid-based approach. The complex contagions’ speed depends heavily on the way long-range links (or weak ties) distributed in a network. However, the random generated long-range links may not be efficient in helping complex contagions in both approaches.

The aforementioned models do not follow the hierarchical property of social network [7, 8]. Therefore, in this paper, we first introduce a *tree-based complex contagions (TCC)* model, which is one of the most intuitive representations of the hierarchical structure. In TCC, we convert the OMSN into a complete binary tree, in which the children nodes under the same parent connect with each other, and the children nodes also connect with the cousin nodes of their parent nodes. However, TCC has the bottleneck problem, which may prevent

the complex contagions. Then, we propose another hierarchical structure: *clique-based complex contagions* (CCC). In CCC, there is a k -clique in the center; other nodes connect to every node in this k -clique, which forms a star structure. The CCC structure can accelerate the complex contagions' speed. However, it increases the number of links dramatically. Finally, we introduce a load-balanced structure: *hypercube-based complex contagions* (HCC). In HCC, we convert the OMSN into an m -dimensional balanced hypercube, in which each node has a backup (matching) node that shares the same set of neighboring nodes. HCC is a bipartite graph, which guarantees the fault tolerance. Thus, there is no bottleneck problem in HCC.

The major contributions of our work are as follows: (1) We point out the limitations of ring lattice-based and grid-based complex contagions approaches. (2) We present three efficient hierarchical complex contagions schemes: tree-based, clique-based, and hypercube-based. The tree-based approach uses a complete binary tree structure. Clique-based complex contagions scheme utilizes the center k -clique and forms a star structure. Finally, hypercube-based complex contagions is based on the balanced hypercube structure. (3) We analyze characteristics of these complex contagions schemes in number of links, diameter, node degree, and bisection width. (4) We evaluate the proposed schemes in synthetic social network models. We compare the performance in both one information 2-complex contagions and multiple information 2-complex contagions. (5) Analysis and simulation results show that CCC has the smallest delivery time when there is only single message, and HCC has the best performance when the number of messages increases.

II. RELATED WORK

Opportunistic mobile social network, a new type of DTN, becomes more and more interesting, due to the widespread use of smart phones. Researchers study OMSNs from a social networking point of view [3, 9]. There has been some work on data dissemination in OMSNs [1, 2]. In [1], Ning et al. proposed a credit-based incentive-aware data dissemination scheme in DTN. Their scheme effectively tracks the value of a message, which highly depends on its probability of being delivered by an intermediate node. Gao and Cao proposed a user-centric data dissemination in [2]. Their approach was based on a social centrality metric, which considers the social contact patterns and interests of mobile users simultaneously, and thus ensures effective relay selection. These approaches studied the simple contagions problem. In this paper, we study the complex contagions to spread behaviors and beliefs.

Recently, complex contagions models have been studied for the spread of ideas across a social network. The research of complex contagions can provide crucial insights into social influence and behavior-adoption cascades on networks. In [4], Centola and Macy study the complex contagions based on Watts and Strogatz's small world model [6]. The authors point out that the weak ties are effective for simple contagions, but they will prevent the complex contagions. In [5], the authors analyze the complex contagions' diffusion speed in three social network models: Newman-Watts model [10], Kleinberg's small world model [11], and Kleinberg's hierarchical network model [12]. The authors show the upper and lower bounds

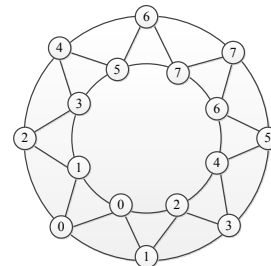


Fig. 2. An illustration of ring lattice-based complex contagions (RLCC) in Watts and Strogatz's small world model.

of the 2-complex contagions speed in these social network models. Then, they generalize these results to a -complex contagions.

III. PRELIMINARIES

A. Simple Contagions vs Complex Contagions

The similarity among different kinds of contagions invites generalization of the small world principle from the spread of information and disease to the spread of collective behaviors and beliefs. For the information and disease spreading, it requires only one influenced source, which likes epidemics. However, the requirement for behaviors, beliefs, and preferences spreading is more complex, which is more costly, controversial, or risky for the individuals to participate.

The problem is that, while all contagions have a minimum threshold of one, the range of nonzero thresholds can be quite large. For communicable diseases and information, the threshold is always exactly one. These are examples of simple contagions, in which contact with a single source is sufficient for the target to become informed or infected. While information and disease are archetypes of simple contagions, some collective behaviors can also spread through simple contact. However, many collective behaviors involve complex contagions that require social affirmation or reinforcement from multiple sources. For complex contagions to spread, multiple sources of activation are required, since contact with a single active neighbor is not enough to trigger adoption. A contagion is complex if its transmission requires an individual to have contact with two or more sources of activation. Depending on how contagious the disease, infection may require multiple exposures to carriers, but it does not require exposure to multiple carriers.

B. Watts and Strogatz's Small World Model and Ring Lattice-based Complex Contagions

The classic formalization of the small world model comes from Watts and Strogatz [6], in which n nodes are placed on a ring lattice, and nodes within ring distance 2 are connected by a *local link* (or *strong tie*). The *long-range links* (or *weak ties*) are the links randomly rewired in this ring lattice. The authors demonstrate that the rate of propagation on a clustered network can be dramatically increased by randomly rewiring a few local links (within a cluster), and making them into bridges between clusters that reduce the mean distance between arbitrarily chosen nodes in the network. They used a ring lattice to demonstrate the small world effect for a simple contagion, as shown in Fig. 2.

In [4], Centola and Macy remarked that the strength of weak ties is not always so significant in helping diffusions. Specifically, it does not help with the diffusion of complex contagions. Information or diseases are simple contagions. They could be spread through a single contact. Hence, a single long-range link can help affect remote regions, which greatly speeds up the diffusion. In some other cases, however, diffusion requires multiple confirmations, or multiple contacts with affected nodes, to accumulate sufficient influence. Complex contagions appear due to strategic complementarity, credibility, legitimacy, and emotional exchange, as explained by Centola and Macy [4]. While long-range links can carry information across long social distances, they are not as effective in spreading complex contagions, simply due to the lack of multiple, collective contacts. Thus, fast diffusion of complex contagion requires not only long bridges, but also “wide” ones, which may or may not exist. Therefore, complex contagions spread using mostly the local connections, and are going to be significantly slower in many settings. Analytic results and simulation results have been shown based on the Watts and Strogatz’s small world model [6]. The authors introduce a *critical width* (W_c) of bridges, which is the minimum number of nonredundant ties required for a contagion to propagate to an unactivated neighborhood. For simple contagions, $W_c = 1$, regardless of network topology. On a ring lattice, for minimally complex contagions ($a = 2$), $W_c = 3$. The critical width also determines the minimum number of links that need to be rewired to create a shortcut across the ring.

While a single random tie is sufficient to promote the spread of simple contagions, complex contagions require more rewiring in order to benefit from randomization. The number of links that need to be randomly rewired increases exponentially with the number required to form a bridge, and the number of ties needed to form a bridge, in turn, increases exponentially with the required number of activated neighbors [4]. Even for this minimally complex contagion on this very small ring (with only 16 nodes), as shown in Fig. 2, the probability that three random links will form a bridge is close to zero. Therefore, we believe that the ring lattice model is not efficient for complex contagions. In this paper, we propose three novel efficient complex contagions approaches. The details will be discussed later.

C. Kleinberg’s Small World Model and Grid-based Complex Contagions

Kleinberg claims that Watts and Strogatz’s small world model lacks the essential navigability property in [11]: without global knowledge of the network, none of the short paths can be computed efficiently. He shows that navigability can be obtained by adjusting the amount of randomness to the underlying metric. Precisely, he introduces an grid-based graph model, in which a constant number of random additional long ties will be given to each node with the harmonic distribution. In this model, the local knowledge at each node is the underlying metric of the grid (which can be viewed as the geographic locations of the nodes) and the positions on the grid of its long-range neighbors. Note that a global knowledge would be the set of positions of all the long-range neighbors on the grid.

In [5], Ghasemiesfeh et al. study the spread speed of com-

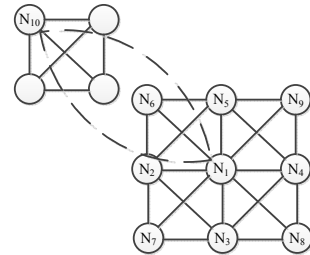


Fig. 3. An illustration of grid-based complex contagions (GCC) in Kleinberg’s small world model.

plex contagions based on the Kleinberg’s small world model we discuss above, which is in a $\sqrt{n} \times \sqrt{n}$ grid with n nodes, as shown in Fig. 3. In order to eliminate the boundary effect, the authors wrap up the grid into a torus. They consider two types of edges in this network. The local links are the edges that connect two nodes on the grid with Manhattan distance 2 or smaller. In addition, each node generates 2 random outgoing edges that are considered as long-range links. The probability that node i chooses j as a neighbor through a long-range link follows harmonic distribution, which is proportional to $1/|ij|^\alpha$, where $\alpha \geq 0$ is a parameter and $|ij|$ is the Manhattan distance between nodes i and j . The authors claim that if $\alpha = 2$, the complex contagions’ spread speed is $O(\log^{3.5} n)$ and $\Omega(\log n / \log \log n)$ with high probability. For $0 < \alpha < 2$, the complex contagions’ spread speed is $O(n^{\frac{4-2\alpha}{10-4\alpha}} \sqrt{\log^3 n})$ and $\Omega(\log n / \log \log n)$ with high probability.

Essentially, it was shown that when the contagion is merely minimally complex, i.e., requiring two active neighbors to be affected instead of one, it would require a substantially large number of random long-range outgoing edges to even create one single ‘bridge’ to diffuse the contagion. Therefore, in the GCC, the additional long-range links may not be significant in helping diffusion of complex contagions.

IV. MOTIVATION

As discussed in Sections III-B and III-C, long-range links are not effective in complex contagions. Following arguments originally proposed by Mark Granovetter’s seminal 1973 paper, *The Strength of Weak Ties*, the majority of influence in novel information contagions is generated by weak ties [13]. However, in the complex contagions scenario, weak ties (or long-range links) can impede diffusion. Most of the weak ties are local bridges, which means there is no adjacent node between two neighbors on the weak tie. Since the complex contagions need more than one activated neighbor to trigger the adoption, the long-range links will prevent the behaviors, beliefs, and presences propagation.

We design a simple simulation to illustrate the participants of the long-range links in complex contagions in three real traces: Infocom 06 conference trace [14], MIT reality mining campus trace [15], and Intel lab trace [16]. Here, we apply a simple epidemic forwarding approach. We change the number of required activated neighbors (a) to see the percentage of involved forwarding long-range links. As shown in Fig. 4, we can see that when $a = 1$, which means simple contagions, most of the contagions are through long-range links. This confirms the conclusions from Granovetter’s paper: *The Strength of Weak Ties* [13]. However, when the number of required activated neighbors increases, the percentage of the

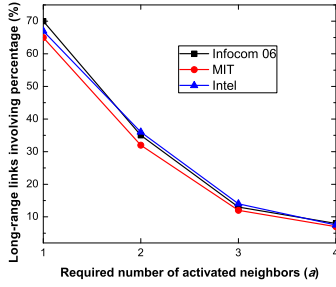


Fig. 4. An illustration of the percentage of the involved long-range links in complex contagions in real traces.

forwardings through long-range links decreases dramatically. When a is 4, only 10% of the forwarding ties involved in the epidemic process are long-range links, in Fig. 4.

Long-range links are not effective in complex contagions, sometimes they may even impede diffusion. From the simple simulation in Fig. 4, there are not many long-range links involved in complex contagions. Therefore, in order to design efficient complex contagions schemes in OMSNs, we only consider the local links and propose tree-based, clique-based, and hypercube-based complex contagions approaches in the following sections. At the same time, the proposed models follow the hierarchical property of social networks, while RLCC and GCC do not have the hierarchical property.

V. TREE-BASED COMPLEX CONTAGIONS (TCC)

Complex contagions is a complex broadcasting process. In network science, tree-based structure is one of the most intuitive representations to study the broadcast process. In a number of settings, nodes represent objects that can be classified according to a hierarchy or taxonomy; nodes are more likely to form links if they belong to the same small subtree in the hierarchy, indicating they are more closely related. Here, we introduce a tree-based complex contagions (TCC) for beliefs or behaviors spreading.

As shown in Fig. 5, it is a binary tree structure. In TCC, the children nodes under the same parent connect with each other, and the children nodes also connect with the cousin nodes of their parent nodes. Therefore, the initial active nodes can be any two linked nodes in this binary tree.

In a tree-based 2-complex contagions, if the initial activated nodes have at least one common neighbor, then in our proposed TCC model, the whole network will be activated. In TCC, the spread speed is the height of the tree, which can achieve efficient complex contagions. The major difference between TCC and the previous approaches – RLCC and GCC is that, in TCC, without the random generated long-range links, the complex behaviors or beliefs can also be spread in a short time to the whole network. The disadvantage of TCC is the bottleneck problem. If the high level ‘parent’ nodes are not activated, their children nodes cannot be activated.

VI. CLIQUE-BASED COMPLEX CONTAGIONS (CCC)

In social networks, there is a small amount of individuals, who are more popular than other people. According to preference attachment algorithm [7] and rich club phenomenon [8], we realize that there exists a popular individual community in the social networks. When a company plans to release a

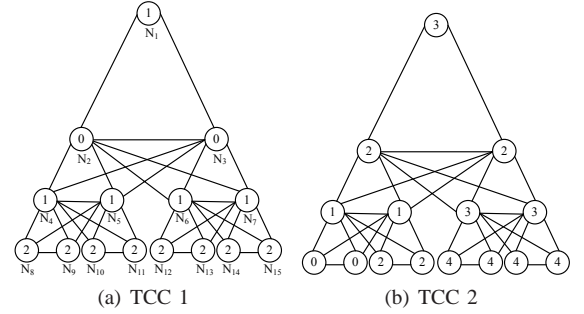


Fig. 5. An illustration of tree-based complex contagions (TCC).

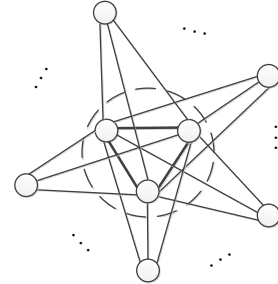


Fig. 6. An illustration of clique-based complex contagions (CCC).

new product, they will invite some people to try the new product, and complex contagions to other people. It is obvious that choosing the individuals from the popular community can increase the spreading speed.

In this section, we propose a clique-based complex contagions (CCC) scheme, which is also a hierarchical architecture. Here, we consider the set of the popular people as a k -clique. Other individuals connect to all nodes in this k -clique, which forms a star structure. As shown in Fig. 6, the center is a 3-clique. For a 2-complex contagions, any two initial activated nodes can spread the behaviors or beliefs to all nodes in the network. If the initial activated nodes are in the clique, the spread speed will increase dramatically.

VII. HYPERCUBE-BASED COMPLEX CONTAGIONS (HCC)

In the social network, many pairs of nodes have multiple common friends. For 2-complex contagions, it requires 2 sources of activation to trigger adoption. Therefore, if the pair of nodes are activated, eventually, their common friends will be activated. In this section, we leverage the matching pair property of the balanced hypercube to achieve efficient complex contagions.

An m -dimensional hypercube is a graph having 2^m nodes. Two nodes are joined by an edge if their addresses, as binary integers, differ in exactly one bit position. Balanced hypercube is a special hypercube, which has been studied in computer system [17]. Balanced hypercube is a load-balanced graph.

Definition 1: An m -dimensional balanced hypercube consists of 2^{2m} nodes $(a_0, a_1, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_{m-1})$, where a_0 and $a_i \in \{0, 1, 2, 3\}$ ($1 \leq i \leq m-1$). Every node $(a_0, a_1, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_{m-1})$ connects the following $2m$ nodes:

- 1) $((a_0 + 1) \bmod 4, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_{m-1})$,
 $((a_0 - 1) \bmod 4, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_{m-1})$, and
- 2) $((a_0 + 1) \bmod 4, \dots, (a_i + (-1)^{a_0}) \bmod 4, \dots, a_{m-1})$,
 $((a_0 - 1) \bmod 4, \dots, (a_i + (-1)^{a_0}) \bmod 4, \dots, a_{m-1})$.

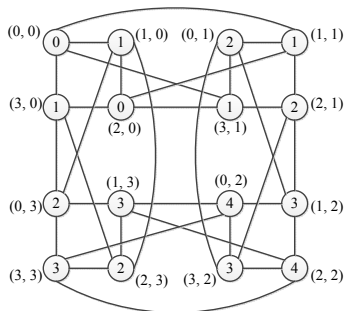


Fig. 7. An illustration of hypercube-based complex contagions (HCC).

Fig. 7 is an example of a 2-dimensional balanced hypercube. In a balanced hypercube, every node has another node matching it, i.e., these two nodes have the same adjacent nodes. As shown in Fig. 7, nodes $(0,0)$ and $(2,0)$ are a pair of matching nodes, which have the same adjacent nodes: $(1,0)$, $(3,0)$, $(1,1)$, and $(3,1)$.

An m -dimensional balanced hypercube has a matching pair property (*Property 1*): In an m -dimensional balanced hypercube, nodes can be partitioned into a set of matching pairs $v = (a_0, a_1, \dots, a_{m-1})$ and $v' = (a_0 + 2, a_1, \dots, a_{m-1})$. Therefore, in order to activate all nodes, it only needs one matching pair of nodes to be the initial activated nodes.

Another important property of balanced hypercube (*Property 2*) is: An m -dimensional balanced hypercube has 2^{2m} nodes, each of which has $2m$ adjacent nodes. Since, each node has $2m$ adjacent nodes, HCC is a fault-tolerant approach. HCC also does not have the bottleneck problem.

VIII. ANALYSIS

In this section, we will compare RLCC, GCC, TCC, CCC, and HCC from different network characteristics: number of links, diameter, node degree, and bisection width.

A. Number of Links

The number of links in the network measure the connectivity of the network. Normally, for a network model, the number of links is the fewer the better.

For RLCC, it is a 2-dimensional ring lattice structure. The number of links depends on the number of neighbors every node is connected to (K). In Fig. 2, K is equal to 4. The total number of links in RLCC is $\frac{nK}{2}$, where n is the total number of individuals in the network.

For GCC, it is an extension of 2-dimensional mesh with wrap structure. Each node has 8 local connected neighbors, e.g., in Fig. 3, N_1 has 8 local links to $N_2, N_3, N_4, N_5, N_6, N_7, N_8$, and N_9 , and 2 randomly selected long-range links, e.g., in Fig. 3, N_1 has 2 random long-range links to N_{10} . Therefore, the total number of links in GCC is $5n$.

For TCC, it is an extension of a binary tree. Here, we assume the height of the binary tree is h . For the number of links, we have the following conclusion:

For $h = 0$, there is only one node. So, there is no link.

For $h = 1$, there are three nodes, which are fully connected. Therefore, the total number of links is 3;

For $h = 2$, there are 7 nodes in the binary tree. The root node has only 2 connections (e.g., N_1 has 2 links: (N_1, N_2) and (N_1, N_3) in Fig. 5(a)). Each node in the second level has 6 connections (e.g., N_2 has 6 links: (N_2, N_1) , (N_2, N_3) , (N_2, N_4) , (N_2, N_5) , (N_2, N_6) , and (N_2, N_7) in Fig. 5(a)), while each node in the bottom level has only 3 connections. Therefore, the total number of links is $(1 \times 2 + 2 \times 6 + 4 \times 3)/2 = 13$.

For $h \geq 3$, there are $2^{h+1} - 1$ nodes in the binary tree. The root node has only 2 connections. Each node in second level has 6 connections, while each node in bottom level has only 3 connections. Each node between the second level and bottom level has 7 connections (e.g., N_4 has 7 links: (N_4, N_2) , (N_4, N_3) , (N_4, N_5) , (N_4, N_8) , (N_4, N_9) , (N_4, N_{10}) , and (N_4, N_{11}) in Fig. 5(a)). Therefore, the total number of links is

$$\begin{aligned} & (1 \times 2 + 2 \times 6 + 2^2 \times 7 + \dots + 2^{h-1} \times 7 + 2^h \times 3)/2 \quad (1) \\ & = 7 \times (1 + 2 + 2^2 + \dots + 2^{h-2}) + 2^{h-1} \times 3 \\ & = 10 \times 2^{h-1} - 7 = 10 \times 2^{\log_2 n - 1} - 7. \end{aligned}$$

For CCC, the number of links depends on the size of the k -clique. We can get the number of links as the equation: $\frac{(n-k) \times k + k \times (n-1)}{2} = (2n - k - 1) \times \frac{k}{2}$. The first term is the total number of links for outside clique nodes, and the second term is the total number of links for the nodes in the k -clique.

According to the definition of the balanced hypercube (Definition 1), for HCC, it is an m -dimensional balanced hypercube with 2^{2m} nodes. Each node has $2m$ number of links. Therefore, the total number of links in HCC is $m \times 2^{2m} = \frac{n}{2} \times \log_2 n$.

B. Diameter

The diameter of a network is the longest of all the calculated shortest paths in a network. In other words, once the shortest path length from every node to all other nodes is calculated, the diameter is the longest of all the calculated path lengths. The diameter is representative of the linear size of a network, which is an important measure of communication delay. Normally, the shorter the diameter, the lower the communication delay.

The RLCC and GCC are based on the small world model, in which, the diameter of the network is around 6 [18]. Six degrees of separation means that everyone is six or fewer steps away from any other person in the world, so that any two people can be connected in a maximum of six steps.

The diameter of the binary tree T is the largest of the following quantities: the diameter of T 's left subtree; the diameter of T 's right subtree; the longest path between leaves that goes through the root of T (this can be computed from the heights of the subtrees of T).

Theorem 1: In TCC, the diameter D of our proposed binary tree is as follows:

$$D = \begin{cases} 0, & h = 0 \\ 1, & h = 1 \\ 2(h-1), & \text{otherwise} \end{cases} \quad (2)$$

where h is the height of the binary tree.

Proof: As we can see from Fig. 5, when $h = 0$, there is only one node; therefore the diameter is 0. When $h = 1$, it

is a three-node fully connected graph, which has diameter 1. When the height is larger than 1, according to our proposed binary tree structure, the leaf nodes can connect their ‘cousin’ nodes through their ‘parents’, which do not need to go through the root. Therefore, the diameter is $2(h - 1)$ in this situation. ■

Since, there are $2^{h+1} - 1$ nodes in a binary tree with height h , we have

$$D = \begin{cases} 0, & n = 1 \\ 1, & n = 3 \\ 2(\log_2(n + 1) - 2), & \text{otherwise} \end{cases} \quad (3)$$

where n is the number of nodes.

In CCC, we can see clearly from Fig. 6, the distance between two nodes insider the clique is 1. The distance between two nodes outside the clique is 2. The distance from one node inside the clique to another node outside the clique is 1. Therefore, the diameter of CCC is 2.

Theorem 2: In HCC, The diameter D of m -dimensional balanced hypercubes is as follows:

$$D = \begin{cases} 2m, & m \text{ is even or } m = 1 \\ 2m - 1, & m \text{ is odd other than } 1. \end{cases} \quad (4)$$

The details of the proof can be found in [17]. The balanced hypercube has a smaller diameter than the traditional hypercube when m ($m > 1$) is odd.

In RLCC and GCC, weak ties play an important role in reducing the network diameter. Our proposed TCC, CCC, and HCC remove the long-range links, since the long-range links prevent the complex contagions. For fair comparison, if there are only local links in the small world models, in RLCC, the diameter of the ring lattice is $n/2$, which is much larger than the diameters in TCC and HCC in large scale OMSNs. Since, the diameter of 2-dimensional mesh with wrap is $2 \lfloor \frac{\sqrt{n}}{2} \rfloor$, for GCC, the diameter of the grid is $\lceil \sqrt{n} \rceil - 1$ [11].

C. Node Degree

The degree of a node in a network is the number of connections or edges the node has to other nodes.

In RLCC, the average degree is equal to a predefined parameter K , which determines the number of neighbors every node is connected to. In Fig. 2, the node degree is 4. In GCC, the average node degree is 10 (or 9), which is the summation of its 8 nearest neighbors with Manhattan distance 2 or smaller and 2 different long-range linked nodes (or 2 long-range links connected to the same node, as shown in Fig 3).

Theorem 3: The node degree of TCC is smaller than or equal to 7.

Proof: In TCC, the degree for the top root is 2. For the bottom leaves, the node degree is 3, which is one link to their ‘brother’, one link to their ‘parent’, and another link to their ‘uncle’. For the remaining nodes, the node degree is 7, which is the summation of the links to 4 of its next generation (e.g., for node N_4 : (N_4, N_8) , (N_4, N_9) , (N_4, N_{10}) , and (N_4, N_{11}) in Fig. 5(a)), one link to its same generation (e.g., for node N_4 :

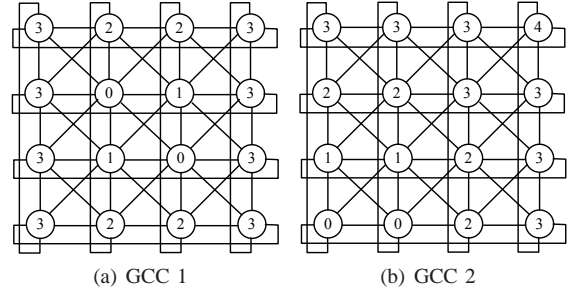


Fig. 8. Comparison of 2-Complex Contagions Spreading Speed: the number in the node is the activated step. 0 means the initial activated node, while 1 means the node activated in step 1.

(N_4, N_5) in Fig. 5(a)), and two links to its elder generation (e.g., for node N_4 : (N_4, N_2) and (N_4, N_3) in Fig. 5(a)). ■

In CCC, the degree of the nodes which are outside the k -clique is k . However, the degree of the nodes which belong to the k -clique is $n - 1$. Therefore, CCC is not a load balance architecture. In HCC, the degree of any node in an m -dimensional balanced hypercube is equal to $2m$.

D. Bisection Width

Definition 2: Bisection width is the minimum number of arcs that must be removed to partition the network into two equal halves.

In other words, bisection width is the minimum number of communication links that can be removed to break it into two equally-sized disconnected networks. Normally, the bisection width is the larger the better.

For RLCC, the bisection width is according to the number of neighbors (K). The bisection width of RLCC follows the following theorem:

Theorem 4: The bisection width of RLCC with K neighbors is smaller than or equal to $2K$.

Proof: For complex contagions, K must be larger than 1. For $K = 2$, RLCC will be a simple 1-dimensional ring. Therefore, in order to partition this ring, it only needs to break 2 links. In this situation, the bisection width is 2. When K is equal to 3, RLCC will be a 2-dimensional ring. The bisection width is 4, which needs to break the links in 2 dimensions. When K is 4, there are two more links which cross the 2 dimensions, which must be broken in order to split into two halves. The bisection width is 6 in this situation. When K is larger than 4, the additional links to connect the 2-dimensional nodes in long distance also need to break. Thus, in order to split the graph into two halves, it needs to break all links on the boundary, which is $2K$. Therefore, The bisection width of RLCC with K neighbors is smaller than or equal to $2K$. ■

For GCC, in order to partition the network into two equal halves, we need to remove \sqrt{n} links. Therefore, the bisection width in GCC is \sqrt{n} [11]. For TCC, we only need to break the binary tree into its left subtree and right subtree to split the graph into two halves. From Fig. 5, we can see clearly that we only need to break the intermediate links to partition the tree structure. Therefore, the bisection width is 1 for $n \leq 3$ or 5 for $n > 3$. For CCC, the bisection width is according to the size of the clique (k). The bisection width is $(k(n - k + 1))/2$, which splits the k -clique into halves, and other nodes also into halves.

TABLE I. CHARACTERISTICS OF COMPLEX CONTAGIONS MODELS WITH n NODES

Model	Number of links	Diameter	Node degree	Bisection width
RLCC	$\frac{nK}{2}$	$\frac{n}{2}$	K	$2K$
GCC	$5n$	$\lceil \sqrt{n} \rceil - 1$	10 or 9	\sqrt{n}
TCC	$10 \cdot 2^{\log_2 n - 1} - 7$	$2(\log_2(n+1) - 2)$	7	5
CCC	$(2n - k - 1) \frac{k}{2}$	2	k or $n - 1$	$\frac{k(n - k + 1)}{2}$
HCC	$\frac{n}{2} \log_2 n$	$\log_2 n$ or $\log_2 n - 1$	$\log_2 n$	$\frac{n}{2}$

Theorem 5: The bisection width of TCC is $n/2$.

Proof: For HCC, according to the matching pair property (*Property 1*), the nodes in the balanced hypercube can be partitioned into a set of matching pairs. Thus, in order to split the structure into two halves, we need to break the links between the matching pairs, which are $n/2$ links [17]. Therefore, the bisection width for HCC is $n/2$. ■

E. Discussion

We conclude our analysis results in Table I. Compared with RLCC, our proposed TCC reduces the diameter; at the same time, it does not increase the node degree and bisection width. In other words, it increases the spread speed for complex contagions, while it guarantees the structure to be congestion-free. CCC has a higher bisection width and lower diameter compared with RLCC and GCC. However, it increases the number of links and node degree dramatically, which means it is not suitable for sparse networks. HCC is a more balanced model, which has a higher bisection width and smaller diameter, while the number of links and node degree are not big. HCC is also a congestion-free model, which is fault tolerant and load balancing.

To evaluate the efficiency of the complex contagions, we need to calculate the spreading speed. Here, we compare the 2-complex contagions spreading speed in different models in a 16 nodes network (in the TCC model, there is only 15 nodes). From Figs. 2, 5, 7, and 8, we can see that RLCC is the most inefficient model. Even for a small network, RLCC needs 7 steps to activate all nodes. For GCC and TCC, the initial activated nodes are very important, especially for TCC. Figs. 5(a) and 5(b) indicate that if we choose higher level nodes as the initial activated nodes, the spreading speed is much faster. HCC is a load balance model. If we choose a pair of matching nodes as the initial activated nodes, the spreading speed can be controlled, and is very fast. For CCC, the spread speed is the fastest among all approaches. If the initial activated nodes belongs to the k -clique, they only need 1 step to spread the information to all other nodes. At the same time, CCC is also suitable for a -complex contagions, while the number of required activated neighbors (a) is larger than 2.

IX. SIMULATION

In this section, we compare the performance of our proposed three complex contagions models (TCC, CCC, and HCC) with two state-of-the-art models (RLCC and GCC).

(1) RLCC: for RLCC, it is a ring lattice structure, while each node has 4 nearest neighbors. It has three random rewired ties connecting to long distance nodes, which are the weak ties, as shown in Fig. 2. (2) GCC: the grid-based complex contagions model follows the rules we presented in Section III-C. In our simulation, we set α to 2. (3) TCC: we create a complex binary tree with extra links, as discussed

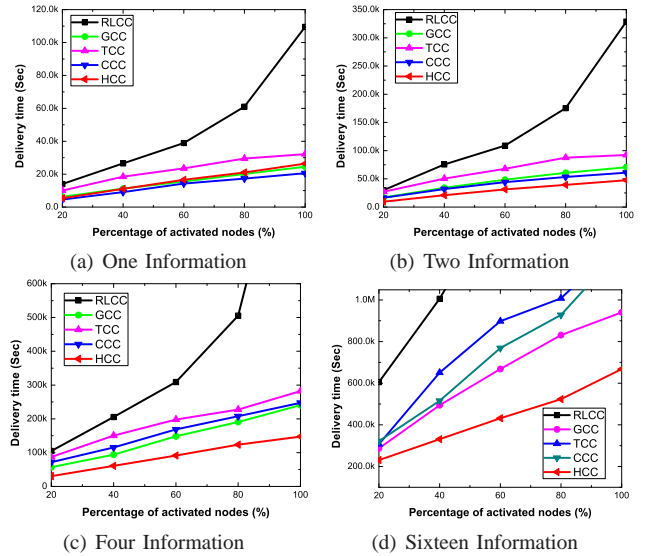


Fig. 9. Comparison of the delivery time for complex contagions.

in Section V. (4) CCC: in the simulation, we create a 3-clique based star structure. (5) HCC: the balanced hypercube generation rule follows the Definition 1.

The contact table is generated with one contact per second, while the probability of choosing the links is equal. Since, we compare the 2-complex contagions, there are two initial activated nodes, which have at least one common neighbor.

The simulation is compared in the following categories: (1) *One message 2-complex contagions*: there is only one message, which needs two activated nodes to trigger the adoption, spreading by the initial activated nodes. (2) *Multiple messages 2-complex contagions*: there are multiple messages spreading at the same time by different initial pairs of activated nodes. Due to the limited contact duration and bandwidth, at each contact, we assume that only one message can be spread. When one node keeps more than one message, it randomly chooses one message from its buffer to propagate. In our simulation, we will compare the spreading speed in different stages: 20%, 40%, 60%, 80%, and 100% nodes activated. We generate a 1,024-node network with 1,048,576 contacts. The contact nodes are uniformly selected from the pairs of linked nodes in these models.

A. Simulation Results

(1) One message 2-complex contagions: in Fig. 9(a), we compare the delivery time for one message 2-complex contagions in these five schemes. We can see that it takes the longest time to activate the nodes in RLCC, especially at the final stage from 80% activated nodes to 100% activated nodes. Since the three random rewired weak ties may not form a ‘broad’ bridge, the complex contagions cannot propagate to long distance neighbors in this situation. CCC creates many more links; therefore, it has the best performance in terms of delivery time. GCC and HCC have a similar performance, while GCC has shorter latency to finish the whole complex contagions process than HCC.

(2) Multiple messages 2-complex contagions: in Figs. 9(b), 9(c) and 9(d), we conduct three comparisons for multiple messages 2-complex contagions: two, four, sixteen messages,

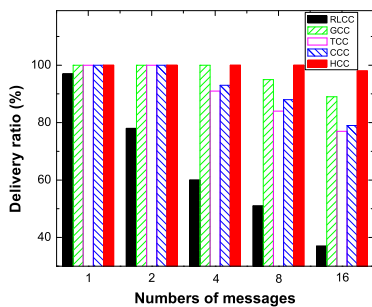


Fig. 10. Comparison of delivery ratios of different numbers of messages.

respectively. Since the spreading speed of CCC highly relies on the k -clique, when the number of spreading messages increases, the performance of CCC decreases. Therefore, initially, there are many traffics on the k -clique, which will increase the complex contagions time. We find that HCC has the shortest delivery time, especially when the amount of messages increases. In HCC, every pair of matching nodes has multiple common neighbors. Therefore, multiple messages can be spread in parallel paths, which can achieve efficient complex contagions. When there are four messages in the network, RLCC cannot activate all nodes for these messages. When the number of messages increases to 16, RLCC can only spread all the messages to about 40% of nodes, while TCC and CCC also cannot activate all other opinion-free nodes.

(3) Comparison of the delivery ratios: we compare the delivery ratios in different number of spreading messages. From Fig. 10, we can see that when the number of messages increases, the delivery ratios for RLCC decreases dramatically. When there are 16 messages, RLCC can only achieve about 40% delivery ratio. The delivery ratios of HCC are almost 100% in all situations. GCC, TCC, and CCC cannot achieve high delivery ratio when the amount of messages increases.

B. Summary of Simulation

In the simulation, we compare the spreading speed of different 2-complex contagions schemes in different numbers of messages environments. Since CCC has many more links, it has the best performance in terms of delivery time in single message 2-complex contagions. However, when the amount of messages increases, CCC has a congestion problem because of the high dependence of the k -clique. Although HCC has a greater delivery time than CCC and GCC in a single message environment, it has the best performance for multiple messages complex contagions. This is because that in HCC, the matching pair of nodes has multiple adjacent neighbors, which can solve the congestion problem. Since the random rewired long-range links cannot reform a bridge for complex contagions, RLCC has the largest delivery time and lowest delivery ratio among all models. Because of the bottleneck problem, TCC has a larger delivery time than GCC, CCC, and HCC.

X. CONCLUSION

Information dissemination is a key application for opportunistic mobile social networks (OMSNs). Most of the recent research is focused on simple contagions, in which the information dissemination only requires one activated source to infect the opinion-free nodes. However, for some behaviors, beliefs, or preferences, the willingness to participate may

require independent affirmation or reinforcement from multiple sources, which called complex contagions.

In this paper, we study the complex contagions problem in OMSNs. We propose three novel hierarchical complex contagions schemes: tree-based, clique-based, and hypercube-based complex contagions. The tree-based scheme is a complete binary tree model. The clique-based scheme is a star structure with a k -clique in the center. The hypercube-based scheme is based on a balanced hypercube model. We compare the characteristics of these five models in formal analysis and simulation. The results show that CCC has the best performance in terms of delivery time when there is only one message. When the number of messages increases, HCC has the smallest delivery delay among these five models. In this paper, we consider the 2-complex contagions in undirected social networks. In the future work, we will extend our research in a -complex contagions in directed networks.

REFERENCES

- [1] T. Ning, Z. Yang, X. Xie, and H. Wu, "Incentive-aware data dissemination in delay-tolerant mobile networks," in *Proc. of 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, 2011, pp. 539–547.
- [2] W. Gao and G. Cao, "User-centric data dissemination in disruption tolerant networks," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, 2011, pp. 3119–3127.
- [3] A.-K. Pietiläinen and C. Diot, "Dissemination in opportunistic social networks: the role of temporal communities," in *Proc. of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, 2012, pp. 165–174.
- [4] D. Centola and M. Macy, "Complex Contagions and the Weakness of Long Ties," *American Journal of Sociology*, vol. 113, no. 3, pp. 702–734, Nov. 2007.
- [5] G. Ghasemiesfeh, R. Ebrahimi, and J. Gao, "Complex contagion and the weakness of long ties in social networks: revisited," in *Proc. of the 14th ACM Conference on Electronic Commerce (EC)*, 2013, pp. 507–524.
- [6] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440–442, June 1998.
- [7] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.
- [8] M. Cebrian and L. M. Vaquero, "The rich club phenomenon in the classroom," *Nature Scientific Reports*, vol. 3, no. 1174, 2013.
- [9] J. Wu and Y. Wang, "Social feature-based multi-path routing in delay tolerant networks," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, 2012, pp. 1368–1376.
- [10] M. Newman and D. Watts, "Scaling and percolation in the small-world network model," *Phys. Rev. E*, vol. 60, no. 6, pp. 7332–7342, 1999.
- [11] J. Kleinberg, "The small-world phenomenon: an algorithmic perspective," in *Proc. of the 32nd Annual ACM Symposium on Theory of Computing (STOC)*, 2000, pp. 163–170.
- [12] J. M. Kleinberg, "Small-world phenomena and the dynamics of information," in *Proc. in Neural Information Processing Systems (NIPS)*, 2001, pp. 431–438.
- [13] M. Granovetter, "The Strength of Weak Ties," *The American Journal of Sociology*, vol. 78, pp. 1360–1380, 1973.
- [14] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, "CRAWDAD trace cambridge/haggle/imote/infocom2006 (v. 2009-05-29)," 2009.
- [15] A. P. N. Eagle and D. Lazer, "Inferring social network structure using mobile phone data," in *Proc. of the National Academy of Sciences*, vol. 106(36), 2009, pp. 15,274–15,278.
- [16] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, "CRAWDAD trace cambridge/haggle/imote/cambridge (v. 2006-01-31)," Jan. 2006.
- [17] J. Wu and K. Huang, "The balanced hypercube: A cube-based system for fault-tolerant applications," *IEEE Trans. Comput.*, vol. 46, no. 4, pp. 484–490, Apr. 1997.
- [18] F. W. Takes and W. A. Kosters, "Determining the diameter of small world networks," in *Proc. of the 20th ACM International Conference on Information and Knowledge Management (CIKM)*, 2011, pp. 1191–1196.