Information Dissemination Dynamics in Delay Tolerant Network: A Bipartite Network Approach

Sudipta Saha*
Indian Institute of Technology
Kharagpur-721302, India
sudipta.saha
@cse.iitkgp.ernet.in

Niloy Ganguly†
Indian Institute of Technology
Kharagpur-721302, India
niloy
@cse.iitkgp.ernet.in

Animesh Mukherjee‡
Indian Institute of Technology
Kharagpur-721302, India
animeshm
@cse.iitkgp.ernet.in

ABSTRACT

In this paper, we present a model of a delay tolerant network (DTN) and identify that this model can be suitably reformulated as a bipartite network and that the major predictions from the former are equivalent to that of the latter. In particular, we show that the *coverage* of the information dissemination process in the DTN matches accurately with the size of the largest component in the suitably thresholded one-mode projection of the corresponding bipartite network. In the process of this analysis, some of the important insights gained are that (a) arbitrarily increasing the number of agents participating in the dissemination process cannot increase the *coverage* once the system has reached the stationary state for a given buffer time (i.e., the time for which a message resides in the buffer of the places visited by the agents), (b) the *coverage* varies inversely with the square of the number of places in the system and directly with the square of the average social participation of the agents and (c) it is possible to design an optimal buffer time for a desired cost of *coverage*. To the best of our knowledge, this is the first such work that employs the rich theoretical backbone of bipartite networks as a "proxy" for the analysis of the otherwise intractable DTN dynamics thus allowing for various novel analytical estimates.

*PhD Student, Department of Computer Science & Engineering, Indian Institute of Technology, Kharagpur, India.

†Associate Professor, Department of Computer Science & Engineering, Indian Institute of Technology, Kharagpur, India.

‡Assistant Professor, Department of Computer Science & Engineering, Indian Institute of Technology, Kharagpur, India.

MobiOppŠ12, March 15–16, 2012, Zürich, Switzerland

Categories and Subject Descriptors

C.2.1 [**Computer-communication Networks**]: Network Architecture and Design—*Store and forward networks, Wireless communication*; G.2.2 [**Discrete Mathematics**]: Graph Theory—*Network Problems*

General Terms

Performance,Theory

Keywords

Delay Tolerant Network, DTN, Bipartite Network, Information Dissemination, Coverage

1. INTRODUCTION

The medium of communication among the devices in a DTN is wireless and therefore the connectivity among them is of short range [1] and inefficient. This creates a lack of certainty in direct communication among the devices. Therefore, the concept of indirect communication is introduced in DTN which involves storing a message temporarily in a certain 'throwbox' [26] or 'buffer' in different places so that other devices can pick up the message when the agents carrying the devices, visit these places. Consequently, it is reasonable to advocate that the performance of any search or information dissemination application, developed for such a network, is strongly influenced by the mobility pattern of these participating agents. The destination of the mobile agents in general are selected on a purely random basis as is the case for the random way point model [5] and various other similar models [4, 6]. However, in a realistic scenario, it becomes important to incorporate the social behavior of the agents (as in for humans) that has been recently introduced for instance *'Self Similar Least Action Walk'* based mobility model [15]. A very important component of such observation is that, the agents have a tendency to visit places depending on the attractiveness of those places. In other words, there is a preferential choice driving the mobility pattern of the agents. The preferential choice induces long and short range correlation among the walkers. This increases the complexity in the mobility pattern making it extremely difficult to use traditional mathematical techniques, like mean field theory, in calculating the *coverage*, i.e., the number of distinct nodes receiving the message in

[1]The radius of communication of the devices using technology such as bluetooth [7], is of the order of a few meters.

the steady-state. In fact, this is possibly one of the most important reasons why such an analysis is almost inexistent in the literature although there have been many works related to the design of dissemination algorithms directed to maximize coverage in both wired and wireless networks [2, 10, 17, 18].

In this paper, we identify that, the message dissemination process in DTN has a natural and one-to-one correspondence with a time varying bipartite network where one partition contains a fixed number of places and the other partition contains the agents whose number continuously grows starting from zero. Hence, we try to show that, without "re-inventing the wheel" (such as recasting the DTN scenario as an epidemic spreading process or otherwise), the existing rich theoretical backbone of the evolving bipartite network [8, 16, 19] can be exploited to analyze the otherwise intractable characteristics of the message dissemination process in DTN. In principle, we concentrate on the analysis of the coverage in terms of the largest component of the *one-mode projection* of the underlying bipartite network suitably thresholded by a time-varying threshold.

In the following, we first present a brief survey of the works already done to model different aspects of DTN. In section 3 we describe a realistic scenario of information dissemination in DTN and subsequently examine it under the lens of the analytical framework of the bipartite networks (section 4). In section 5, first we show that the time evolution of the fraction of nodes to which a message gets disseminated (i.e., coverage) in the DTN has a perfect overlap with the growth of the largest component size of the one-mode projection (explained later) of the bipartite network suitably thresholded by a time varying threshold. Next, we show the correlation of the parameters in the two domains. Finally, we provide a closed form expression for the largest component size of the thresholded one-mode projection which in turn gives the theoretical estimate for the coverage achievable in the DTN.

2. RELATED WORKS

Routing/dissemination of information in DTN have been in focus for a long period of time. Many algorithms have been developed to solve these problems. For example, in the store-carry-forward paradigm based algorithms, the mobility of the agents is exploited to convey message packets. In these strategies, the devices carried by the agents temporarily buffer the data and forward it to other (appropriate) agents. Epidemic routing [11], spray and wait protocols [24] are some examples of these. Analysis and subsequent use of contact history among the agents have also been the focus of various works [3, 14, 25].

Modeling DTN through different analytical framework has also gained much interest in recent times. Epidemic modeling [20], ordinary differential equations [23], partial differential equations [1] or Markov models [21] have been successfully used to represent DTN. Performance of different routing strategies have been evaluated using these mathematical techniques. The primary objectives of these studies were to analyze the data delivery ratio and data delivery latency.

Augmenting the DTN with different types of stationary message storing devices, such as throwbox [26], has been the recent trend to enhance the communication opportunities in between the mobile devices. In [22], the authors show that, use of such relay devices effectively decreases the data delivery latency as well as increases the data delivery ratio. However, the existing modeling or analysis related works on DTN do not consider the presence of such message buffers. Recently, Gu et al [13] have addressed this issue in similar lines as of those presented here. They have proposed the use of message buffers as an instance of bio-inspired methods (e.g., pheromone or footprint). Using discrete Markov chain based modeling, they analyzed the importance of buffer time, i.e., the amount of time a message copy can stay in a message buffer, as well as the preferences of visiting different places by the mobile agents. They studied the impact of these two crucial system features on the latency and the message delivery ratio of the dissemination process in the network. In our work we specially emphasize on the inherent bipartite nature of the "buffer augmented DTN" and thus bring forward the fact that instead of starting from scratch, the existing theories of the bipartite network can be used (with necessary modification) to analyze the coverage problems related to DTN.

3. INFORMATION DISSEMINATION IN DTN

We consider a certain number of mobile agents (t) who participate in the information dissemination process and a certain number of common places (N) where the agents usually go. An agent is assumed to make μ number of visits to different places (hence, μ directly models the social participation of the agents). The place to be visited next by an agent, is chosen preferentially from the pool of places where the preference to be given on a place is directly proportional to the number of other agents who already visited the place. A sequential agent arrival pattern is assumed which implies that the next agent will join the system after the previous agent has visited all of the μ places. To make the process realistic, we introduce a concept of *time* which denotes the count of the agents who have joined the system and visited all the places they were supposed to visit. Within a single *time* unit an agent creates μ number of connections (τ varies from 1 to μ). To follow the store and forward paradigm, we assume that, each of the places as well as the agents has a buffer ('throwbox' [26]) where several pieces of information can be stored. Throughout this paper we assume that all the communications between the agents take place via these message buffers. Without any loss of generality, we consider here the dissemination of a single message. Due to the limited size, a message will be discarded from the buffer of a place, after a certain *time* duration b, termed as buffer time [2]. However, due to sequential agent arrival pattern, we assume that the agents can store a particular message for their full life time, after they have got the message. The information dissemination process along with the observable in the process are described below.

- *Initial Condition*: Initially, i.e., prior to the start of

[2]The buffer time happens to be a very crucial factor in the message dissemination process. A longer buffer time implies plenty of redundant message copies, higher CPU cycle as well as higher battery power consumption in the mobile devices. On the other hand small buffer time may imply insufficient node coverage. Therefore, arriving at an optimal buffer time is crucial for the system to maintain a balance between performance and load. Consequently, this factor plays as one of the most significant constraints in the system.

the dissemination process, the buffers of each place and device of the agents are assumed to have enough space to participate in the dissemination process.

- *Start*: The initiator of the dissemination process brings or creates a message in its buffer and sequentially visits μ number of places (preferentially) where the messages are dropped so that other agents can pick up the message and participate in the dissemination process.

- *Dissemination through places*: When some agent comes in a place where the initiator has already dropped the message, the message gets transferred to the buffer of the agent if the agent is not already containing the message.

- *Dissemination through other agents*: Once some agent picks up a message from some place, it also participates in the dissemination process. When such an agent visits a place where the message is not present, the message gets transferred to the buffer of the place. At the same time, the buffer timer of the place gets set to the value b which implies that the message can be stored in the place for b *time* units. This provides some chances to the place to convey the message to different agents.

- *Buffer timer manipulation*: After the end of each *time* unit, the buffer timers in all the places decrease by one. However, when an agent who has already got the message from some place, arrives at another place where the message is already there, we assume that a fresh copy of the message is brought to the place. We relate this event to the importance of the message which is being disseminated and therefore give advantage to the place by setting its buffer timer value back to the maximum value, i.e., b.

- *Observable*: Keeping the 'on the fly' [21] structure of DTN in mind, we abstract out the definition of coverage as number of different places a particular message can reach, under a given message dissemination scheme. Hence, in the process described above, we measure, the number of distinct places, the buffers of which contain the message at different *time* steps. We denote this quantity by G_d. This specific quantity is of interest from the perspective of dissemination because, the probability that any other mobile agent will receive the information while visiting a place, is directly proportional to the value of G_d.

Figure 1 pictorially describes this information dissemination process. In the next section we describe the analysis of this process of dissemination using evolution of the bipartite network.

4. MODELING BY BIPARTITE NETWORK

In this section, we describe the modeling of the whole information dissemination dynamics in DTN as a bipartite network with one growing partition. We visualize the DTN dynamics as a bipartite network where one of the partitions corresponds to the places while the other corresponds to the agents. The number of places is fixed and finite ($=N$) while the number of agents grows over *time* and is modeled by

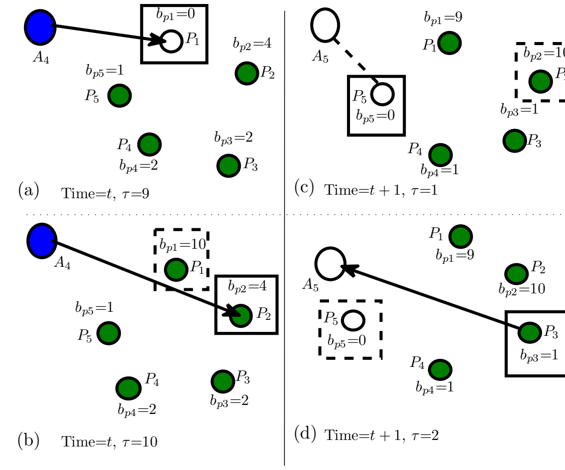

Figure 1: A schematic diagram showing the information dissemination process running in a DTN. There are 5 places: P_1 ... P_5 and 2 agents A_1 and A_2. The notation b_{pi} denotes the current value of the buffer timer of P_i, i.e., the rest *time* units for which the message can be stored in the buffer of place P_i. We assume here that the maximum buffer *time* is $10(=b)$ and the agents can travel $10(=\mu)$ places sequentially. The empty and filled up circles denote the absence and presence of message respectively. Arrows denote the direction of transfer of the message. Part (a) and (b) are snaps of the process at the beginning of creating the 9^{th} and 10^{th} connection at *time* t by agent A_4. In (a), the message gets transferred to P_1 from A_4. In (b), the buffer timer of P_2 gets set to 10. Part (c) and (d) are snaps at the beginning of creating 1^{st} and 2^{nd} connection by the agent A_5. In (c), no message transfer happens. In (d), the message gets transferred from P_3 to A_5. Most of the possible interactions are shown through this figure.

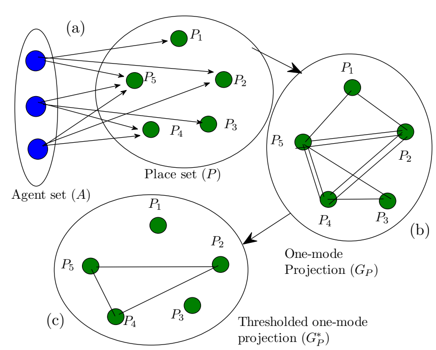

Figure 2: Schematic diagram of a possible scenario of the bipartite network corresponding to a DTN comprising five places, i.e., N=5 (P_1...P_5) and three agents, i.e., t=3 (A_1 ... A_3) with μ=3. The diagram shows a possible status after all the agents have joined the system. Part (a) shows the bipartite network, (b) is the one-mode projection and (c) is the thresholded one-mode projection for threshold value 2.

the parameter t. Each agent is allowed to make μ connections sequentially one by one, each time choosing a place in a preferential fashion (see Figure 2). Therefore, in both the bipartite network as well as the DTN domain, the parameters N, μ and t have the same significance. Table 1 summarizes the precise relationships between the parameters in the two domains.

4.1 One-mode projection in bipartite network

In the course of visiting from place to place, the underlying store and forward paradigm based algorithm, operating in the mobile devices of the agents, allows one agent to convey a message from one place to other. In order to capture this message flow, we take the one-mode projection of this bipartite network on the place set (one-mode projection, on the place set, is a place to place graph where two places are connected by an edge if there is one common agent who has visited/connected both of the places). In this projection, we assign weights to the edges where a particular weight denotes the number of parallel edges between the two places via same or different agents (if one agent has visited two places each twice, then there will be four parallel edges in the one-mode projection which captures the fact that there are actually four different possible communications between the two places). For a bipartite network G, we denote its one-mode projection on the place set by G_P (see Figure 2(b)).

4.2 Thresholding edge weight in the bipartite network

It can be intuitively understood that, the buffer time in the information dissemination process in a DTN, actually controls the flow of the message from one place node to the other. Therefore, the probability that a common visit will convey a message, is directly proportional to the buffer time b. Hence, for a given value of b, there is a minimum number of common visits required to effectively convey a message from one place to the other. To reflect this scenario corresponding to a buffer time b in DTN, we introduce a threshold edge weight c in the bipartite network. In particular, we prune those edges in the one-mode projection of the bipartite network whose edge weights fall below c. Hence, the rest graph contains only those edges which represent strong and stable inter-place communication and thereby accurately simulates the effect of b in DTN. For a certain bipartite network G we denote this thresholded one-mode projection on the place set, by G_P^* (see Figure 2(c)).

4.3 Time-varying threshold

The weights of the edges in G_P vary with time for continuous arrival of the agents and their connection patterns. Therefore, in order to bring the concept of temporal stability of the edge weights between places of the bipartite network (always imposed by b in DTN over the entire time evolution of the system), we calculate the threshold c as a function of *time* t, i.e., number of agents who already joined the system (as we consider the arrival rate as one agent per time step). In this work, we assume that $c \propto t$ and hence $c = v \times t$. We take this constant of proportionality v as the characteristic parameter equivalent to a buffer time b in DTN.

Table 1: Relationship between the parameters in DTN and the bipartite network

Type	DTN	Bipartite Network	Remarks
Parameters	Agents(t)	Agent partition (t)	Growing
	Places(N)	Place partition (N)	Fixed and finite
	Number of place an agent visits (μ)	Number of connections an agent creates with different places (μ)	Constant (can be taken from some specified distribution also)
	Buffer time (b)	Threshold varying with t (v)	(See subsection 4.3)
Observable	Number of places where the message could reach under the dissemination process (G_d)	Size of the largest component of the thresholded one-mode projection (G_b)	These quantities should match

5. RESULTS

The coverage in DTN for a certain value of b is conceptually the same as the size of the largest component (denoted by G_b) in the one-mode projection of the bipartite network suitably thresholded (keeping other parameters same in the two domains). We simulate the time evolution for both G_d as well as G_b for many different parameter combinations. We find that these quantities become independent of t (become stable) after a certain number of agents have joined the system and their time evolutions also match accurately with each other for different value pairs of v and b. Figure 3 shows few such sample cases for different combinations of the other two parameters N and μ.

Thus, it can be understood that, to analyze the coverage, i.e., number of distinct sites where a message can be spread by means of a specific information dissemination scheme in DTN, the very first step to be carried out is to estimate the parameters of the bipartite network setup that can work as a perfect "proxy" for the DTN setup. Therefore, this parameter estimation process is our next focus.

5.1 Estimation of parameters of bipartite network from DTN

In order to successfully establish a correlation between the two domains we need to find a value of the time varying threshold v in bipartite network functionally equivalent to the value of the parameter b in DTN. In other words, the relationship between these two parameters creates a bridge between the two domains. To understand this equivalence we start from the ground state in the DTN, where there is no buffer i.e. b=0. In that case the information will not pass on to a single other node. Therefore, the coverage will be zero. This situation is captured in the bipartite network by using a sufficiently high value of v which prunes all the parallel edges in the one-mode projection of the bipartite network resulting in a set of isolated nodes only. Similarly, in the opposite case,

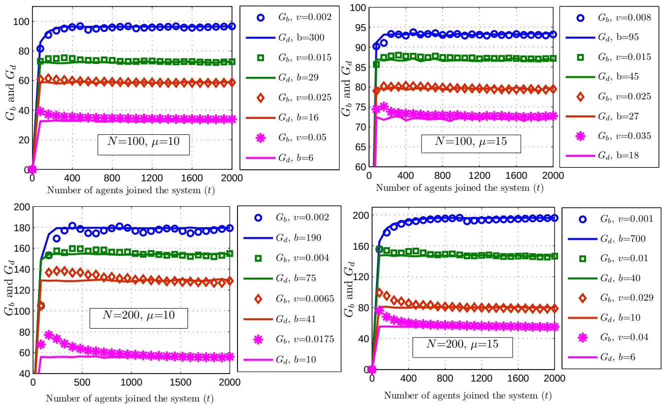

Figure 3: Comparison of the size of the largest component (G_b) in the thresholded one-mode projection of the bipartite network and the number of places in DTN where the information, being disseminated, is found i.e. *coverage* (G_d). The bipartite network is formed corresponding to a DTN set up consisting of (N=) 100 common places and (t=) 2000 mobile agents each of which creates (μ=) 10 connections with the places i.e. visits 10 places sequentially. The four parts of the figure shows the results for four different combinations of N and μ.

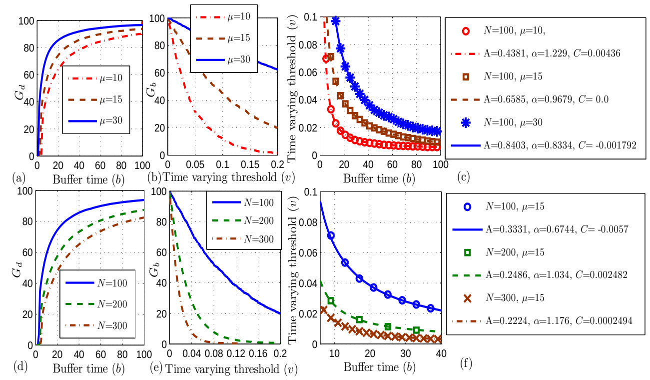

Figure 4: Plots in (a) and (b) show the relationship of G_d and G_b with the buffer time b and the time varying threshold v respectively for different values of μ while the value of N is set at 100. Plots in (d) and (e) show the same relationship for different values of N while the value of μ is set at 15. Plots in (c) and (f) show the relationship between b and v for different values of μ and N along with the parameters A, C and α for the best fitting instance (99% confidence level) of the equation 1. In all these cases the values of G_b and G_d have been collected after enough number of users have joined the system and the system has stabilized.

if we employ a sufficiently high value of b, eventually almost all the nodes in the DTN will receive the message (after sufficiently large number of agents have joined the system with a sufficiently high value of μ). This scenario can be mimicked by a very small value of v (≈ 0) in the bipartite network which does not prune any of the edges from the one-mode projection. Hence, an inverse relationship among these two parameters can be observed. It can also be realized that the relationship between v and b, is not independent of N and μ. We extensively simulate the relationship between v and b for different values of N and μ. Figure 4 shows few sample results and the nature of the increase in the G_d and decrease in the G_b with the increase in b and v respectively for few combinations of N and μ. Using the commonality of G_d and G_b, we infer values of v for given values of b. We find that all these relationships between v and b (for different combinations of N and μ) fit the following equation-

$$v = Ab^{-\alpha} + C \tag{1}$$

where A and C are certain constants. Figure 4 also shows the values of the parameters A, C and α for few different combinations of N and μ. Through extensive simulation we find that the value of the exponent α is directly proportional to a non-linear combination of N and μ that we plan to explore further as a part of our future work.

Next we focus on the theoretical analysis of the size of the largest component in the thresholded one-mode projection of the bipartite network for a given value of v. We describe this process in the following subsections.

5.2 Component formation in bipartite network

In real life, the selection process for the next place to be visited by the mobile agents, incorporates a little randomness, rather than being fully preferential. However, for simplicity purpose, in this work, we have assumed a pure preferential selection model. Due to this reason, the place to place graph generated after application of threshold on the one-mode projection of the bipartite network on the place set, exhibits a special property described below. We denote this property by $\mathbb{P}$.

$\mathbb{P}$: *After any number of agents have joined, the thresholded one-mode projection of the bipartite network on the place set, consists of a single connected component while the rest of the places that are not part of the largest component are degenerate, i.e., have degree zero.*

The full proof of this property (as empirically observed by us for the first time in Figure 5) is out of the scope of this paper. However, the basic intuition behind this scenario can be sketched as follows. Due to preferential attachment, the nodes of the connected component necessarily have high degree in the bipartite network. Hence, any new agent would almost surely make some of the connections with these nodes. Conversely, none of the new agents will make all its connections only with the isolated nodes. Hence, isolated nodes will either get absorbed in the giant component or stay as single entity.

In order to test the above hypothesis, we simulate the evolution process of the bipartite network for various combinations of the parameters N, μ and v for large values of t. For pure preferential attachment process, we always find that the sum of the two quantities : *size of the largest component*, i.e., G_b and *the number of components* (denoted by C_b), is equal to N+1. Plots of Figure 5 show the evolution of G_b and C_b for two different values of v in four different combinations of N and μ. It is clear from the plots that the relationship G_b+C_b=N+1, is maintained through-

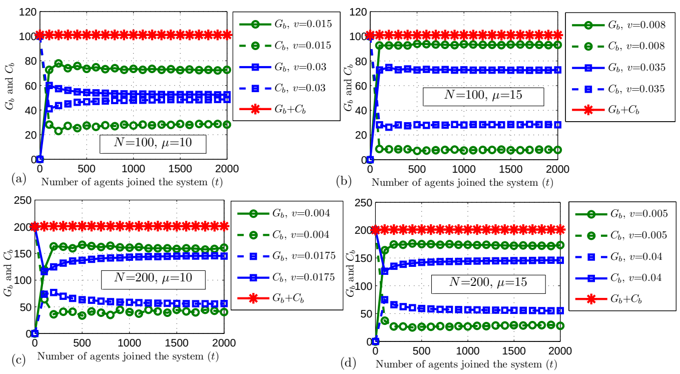

Figure 5: Comparison of the number of components (C_b) and largest component size (G_b) in four different combinations of N and μ. For each such combination, the relationship is shown for two sample values of v throughout the evolution of the bipartite network as 2000(=t) agents join the system sequentially. Each plot also shows the sum of the two quantities (i.e. C_b+G_b) in all these cases.

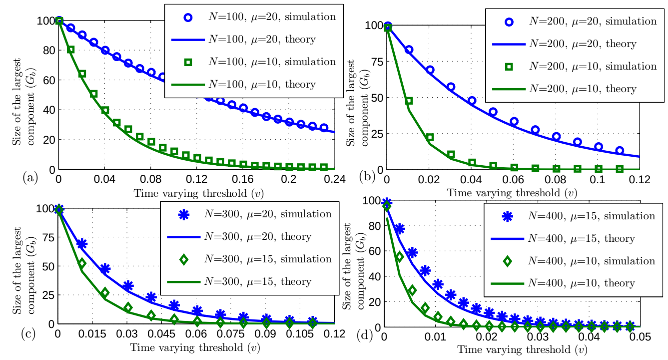

Figure 6: Comparison of the size of the largest component in the bipartite network obtained from equation 3 and the same obtained from simulation of the evolution process in the bipartite network under different combinations of N and μ.

out the whole evolution of the bipartite network for all v which implies that the property $\mathbb{P}$ is satisfied throughout the process. This result would help us to calculate the size of the largest component.

5.3 Calculation of the size of the largest component

In this subsection, we use the existing theory of bipartite network to derive the expected coverage in the described information dissemination process in DTN. We denote the ground state of the bipartite network G as G_0 where set A is empty and set P contains N places. One agent joins the set A per time step and creates μ connections with some elements in the set P. We consider here only the fully preferential attachment process. Using the basics of Polya Urn scheme and the de Finetti theorem [9], it can be shown that the probability that a new agent will create a connection with a place i in P (let us denote this probability as θ_i) of G, is marginally Beta distributed with the parameters b_i and b_0 where b_i is the initial degree of node i in G_0 and b_0 is the sum of the degrees of the other nodes in P of G_0. The work in [12], assumes that all the nodes in set P have same initial degree 1, i.e., b_i=1, $\forall i \in P$ which implies that θ_i, $\forall i \in P$, are identically Beta distributed with parameters (1, N-1). It has been shown in [12] that the expected number of edges at large time between node i and j of set P is $(\mu^2-\mu)\theta_i\theta_j$. From this it has been shown that at large time, the probability that a place i with attractiveness θ_i will be connected to some other place in the thresholded one-mode projection of G, i.e., G^*_P (conversely the probability that a place i with attractiveness θ_i have more than $c = v \times t$ parallel edges with some other node in the one-mode projection, i.e., G_P) is equal to $\left(1-\frac{v}{(\mu^2-\mu)\theta_i}\right)^{(N-1)}$. Finally, deriving the expected number of such connections of node i and using the Beta distribution of the attractiveness of the places of set P, [12] develops the cumulative degree distribution of the nodes of the set P in G^*_P. Equation 2 shows this cumulative degree distribution which effectively gives the probability that a randomly selected node in G^*_P has degree greater or equal to k at large time t.

$$F_k(t) = \left(1-\frac{v}{(\mu^2-\mu)x}\right)^{(N-1)} \tag{2}$$

where v is the time varying threshold and x=$1-\left(\frac{k}{N-1}\right)^{\frac{1}{N-1}}$.

We use this result for calculating the size of the largest component as follows. As we consider only fully preferential model of attachment, the property $\mathbb{P}$ holds true throughout the evolution process. Hence, the fraction of nodes which form the largest component are the nodes that have degree 1 or higher. This fraction could be obtained by putting k=1 in equation 2. We multiply this probability with N to get the number of nodes in the largest component which reads as follows.

$$G_b = N \times \left[1-\left(\frac{\sqrt[N-1]{(N-1)}}{\sqrt[N-1]{(N-1)}-1}\right)\times\left(\frac{v}{\mu^2-\mu}\right)\right]^{N-1} \tag{3}$$

We simulate the evolution of the bipartite network for various combinations of N and μ and measure the size of the largest component. Figure 6 shows the match of theory and the simulation results for eight such different combinations.

For large value of N, the ratio $\left(\frac{\sqrt[N-1]{(N-1)}}{\sqrt[N-1]{(N-1)}-1}\right)$ is almost equal to 1. The value v is generally less than 1 and $(\mu^2-\mu)$ is comparatively a large value ($\gg 1$). Hence, ignoring the higher order terms in binomial expansion in equation 3, we get the following simplified form of G_b.

$$G_b = N - \frac{N(N-1)}{\mu(\mu-1)} \times v \tag{4}$$

5.4 Calculation of the coverage in DTN

Putting the exact expression for v in terms of b in equation 4 we get the following formula which provides a very close estimation of the coverage in DTN.

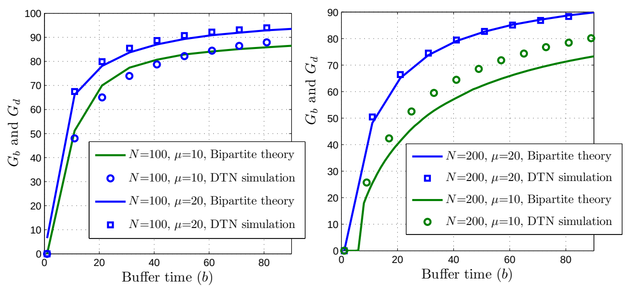

Figure 7: Comparison of the size of the largest component, i.e., G_b in the bipartite network obtained from equation 3 and the *coverage* values obtained by simulating the information dissemination process in DTN for various buffer times under four different combinations of N and μ.

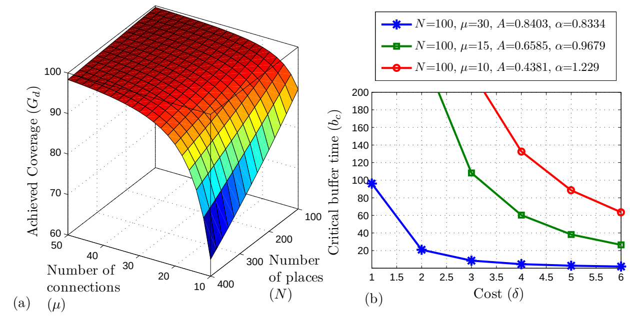

Figure 8: Part (a) shows the plot of the percentage of the number of places covered (i.e., G_d) as a function of the total number of places (N) and the number of places an agent visits (μ). Part (b) shows the plot of the critical buffer time value (b_c) constrained by different cost values (δ).

$$G_d(= G_b) = N - \frac{N(N-1)}{\mu(\mu-1)} \times (Ab^{-\alpha} + C) \qquad (5)$$

To test the accuracy, we simulate the dissemination process in DTN, for several buffer times (for different combinations of N and μ) for which we already know the relationship with their equivalent time varying threshold. Using that relationship we calculate the equivalent values of v and also calculate the size of the largest component in the thresholded one-mode projection of the equivalent bipartite network from equation 3. We find that these theoretical results and the coverage obtained by simulation of information dissemination in DTN, match quite accurately thus pointing to the fact that the empirically derived relationship between v and b is quite appropriate. Figure 7 shows four such cases for different combinations of N and μ.

5.5 Insights obtained from the analysis

We arrive at several significant insights from the closed form expression of coverage derived in terms of N, μ and b (equation 5). Most importantly, it has been found that, for a given value of N, μ and b, the number of nodes covered in the dissemination process does not grow unboundedly with the increase in the number of agents (t) joining the system; rather, after a certain value of t, the total number of place nodes covered, gets stabilized and is limited by the buffer time b (see Figure 3).

Further, to visualize the significance of equation 5, we present a three-dimensional plot (Figure 8(a)) showing the interaction of G_d, μ and N. It is clear from the plot that G_d bears an inverse relationship with N and a direct relationship with μ. Precisely, a closer look at the equation 5 reveals that indeed G_d is inversely proportional to N^2 and directly to μ^2.

As a final remark, we note that the rate of growth of G_d slows down with an increasing value of b (see Figure 7) and particularly after a critical value (say b_c) G_d is almost stable. From the perspective of a design engineer, the value of b_c could be crucial. Naturally the cost associated with dissemination process (buffer space, CPU consumption, bandwidth etc.), should bear a connection with b_c. We assume that this cost is a ratio of the coverage achieved to the overhead incurred per unit increase of b and denote it by δ. Therefore, the rate at which the cost increases should not overshoot the rate at which G_d increases and so we have the following relationship.

$$\frac{dG_d}{db} > \delta \qquad (6)$$

Evaluating the derivative of G_d with respect to b and using relationship 6, we get the following expression for the critical buffer time (b_c).

$$b_c = \sqrt[\alpha+1]{\frac{\alpha AN(N-1)}{\delta\mu(\mu-1)}} \qquad (7)$$

In Figure 8(b), we plot the values of b_c for different values of δ under few different combinations of N and μ. The value of δ can be chosen freely according to the design requirement. However as a proof of concept we inspect the values of delta between 1 and 6 to investigate the nature of the equation 7. The values of b_c, for say N=100 and μ=10, effectively mean that, to satisfy different values of δ, the employed values of b should be below the corresponding curve in Figure 8(b). It is seen that, with increasing cost, increasing the buffer time rapidly becomes uneconomical.

6. CONCLUSION

In this paper, we have identified a novel way of looking at the problem of estimating the coverage of the information dissemination process in delay tolerant networks. In particular, we found that the DTN system has an underlying "bipartite mechanism" which can be used as a "proxy" for analytically estimating the coverage. We have shown that the complexity of computing this quantity for DTN can be reduced to the problem of inferring a suitable value of v (bipartite domain) from a given value of b (DTN domain). In addition, we also observed that (i) arbitrarily increasing t does not amount to an arbitrary increase in the coverage while constrained by a specific value of b, (ii) the coverage achievable is inversely proportional to N^2 and directly proportional to μ^2 and (iii) it is possible to design an optimal value of b for a desired cost of coverage.

Some of the limitations of the current approach are that we assumed a sequential arrival of agents and a fully preferential choice of their movements. It is quite straightforward to relax both of these assumptions by respectively allowing for overlapping life span of the agents in both the domains and introducing a randomness parameter in the model that can control the preference factor of the agents. Preliminary experiments on both of these issues (to be reported elsewhere), show that the major trends are (almost) equivalent to what has been presented in this paper. Furthermore, in this work we mainly focus on the coverage achieved after the system has stabilized/saturated. However, the coverage achieved within a given period of time is also of high importance. Many works have been already done to analyze this in general peer-to-peer networks [17, 18]. We plan to incorporate this factor as a part of the future extension of this work.

7. ACKNOWLEDGMENTS

This work was funded by a project under *Department of Information and Technology* (DIT), Govt. of India. The authors are grateful to Saptarshi Ghosh for his comments and suggestions that helped in improving the quality of the paper.

8. REFERENCES

[1] J. Y. L. B. A. Chaintreau and N. Ristanovit'c. The age of gossip: Spatial mean-field regime. In *Proceedings of ACM Sigmetrics*, 2009.

[2] N. Alon, C. Avin, M. Koucky, G. Kozma, Z. Lotker, and M. R. Tuttle. Many random walks are faster than one. In *Proceedings of SPAA*, pages 119–128, 2008.

[3] A. Balasubramanian, B. Levine, and A. Venkataramani. Dtn routing as a resource allocation problem. In *SIGCOMM Computer Communication Review*, 2007.

[4] C. Bettstetter. Mobility modeling in wireless networks: categorization, smooth movement, and border effects. *Sigmobile Mob. Comput. Commun. Rev.*, 5:55–66, July 2001.

[5] J. Broch, D. A. Maltz, D. B. Johnson, Y. C. Hu, and J. Jetcheva. A performance comparison of multi-hop wireless ad hoc network routing protocols. In *Proceedings of MobiCom*, pages 85–97, 1998.

[6] T. Camp, J. Boleng, and V. Davies. A survey of mobility models for ad hoc network research. *WCMC: Special Issue*, 2:483–502, 2002.

[7] O. S. Chau, P. Hui, and V. O. K. Li. An architecture enabling bluetooth$^{\mathrm{tm}}$/jini$^{\mathrm{tm}}$ interoperability. In *Proceedings of PIMRC*, pages 3013–3018, 2004.

[8] M. Choudhury, N. Ganguly, A. Maiti, A. Mukherjee, L. Brusch, A. Deutsch, and F. Peruani. Modeling discrete combinatorial systems as alphabetic bipartite networks: Theory and applications. *Phys. Rev. E*, 81:036103, 2010.

[9] P. Diaconis and D. Freedman. Finite exchangeable sequences. *Annals of Probability*, 8(4):744–764, 1980.

[10] V. V. Dimakopoulos and E. Pitoura. On the performance of flooding-based resource discovery. *IEEE Trans. Parallel Distrib. Syst.*, 17(11):1242–1252, 2006.

[11] W. Gao, Q. Li, B. Zhao, and G. Cao. Multicasting in delay tolerant networks: A social network perspective. In *Proceedings of MobiHoc*, 2009.

[12] S. Ghosh, A. Srivastava, T. Krueger, and N. Ganguly. Degree distribution of one mode projection of evolving alpha bipartite networks. *Available at http://cse.iitkgp .ac.in/resgrp/cnerg/Files/BipartiteNetwork.pdf.*

[13] B. Gu, X. Hong, and P. Wang. Analysis for bio-inspired thrown-box assisted message dissemination in delay tolerant networks. *Telecommun Syst, Springer*, 2011.

[14] R. Jathar and A. Gupta. Probabilistic routing using contact sequencing in delay tolerant networks. In *Proceedings of COMSNETS*, 2010.

[15] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong. Slaw: A new mobility model for human walks. In *Proceedings of INFOCOM*, pages 855–863, 2009.

[16] A. Mukherjee, M. Choudhury, and N. Ganguly. Understanding how both the partitions of a bipartite network affect its one-mode projection. *Physica A*, 390(20):3602 – 3607, 2011.

[17] S. Nandi, L. Brusch, A. Deutsch, and N. Ganguly. Coverage-maximization in networks under resource constraints. *Phys. Rev. E*, 81(6):061124, Jun 2010.

[18] K. Oikonomou, D. Kogias, and I. Stavrakakis. A study of information dissemination under multiple random walkers and replication mechanisms. In *Proceedings of MobiOpp*, pages 118–125, 2010.

[19] F. Peruani, M. Choudhury, A. Mukherjee, and N. Ganguly. Emergence of a non-scaling degree distribution in bipartite networks: a numerical and analytical study. *Euro. Phys. Lett.*, 79:28001, 2007.

[20] F. Peruani, A. Maiti, S. Sadhu, H. Chatt'e, R. R. Choudhury, and N. Ganguly. Modeling broadcasting using omnidirectional and directional antenna in delay tolerant networks as an epidemic dynamics. *IEEE-JSAC*, 28(4):524, 2010.

[21] A. Picu and T. Spyropoulos. Distributed stochastic optimization in opportunistic networks: The case of optimal relay selection. In *Proceedings of CHANTS*, 2010.

[22] S. Shahbazi, S. Karunasekera, and A. Harwood. Improving performance in delay/disruption tolerant networks through passive relay points. In *Wireless Network*, 2011.

[23] T. Small and Z. J. Haas. The shared wireless infostation model: a new ad hoc networking paradigm (or where there is a whale, there is a way). In *Proceedings of MobiHoc*, 2003.

[24] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Spray and focus: Efficient mobility-assisted routing for heterogeneous and correlated mobility. In *Proceedings of IEEE PerCom workshops*, 2007.

[25] Q. Yuan, I. Cardei, and J. Wu. Predict and relay: an efficient routing in disruption-tolerant networks. In *Proceedings of MobiHoc*, 2009.

[26] W. Zhao, Y. Chen, M. Ammar, M. Corner, B. Levine, and E. Zegura. Capacity enhancement using throwboxes in dtns. In *Proceedings of MASS*, pages 31 –40, 2006.