

Improved Bounds on Information Dissemination by Manhattan Random Waypoint Model

Aria Rezaei
Stony Brook University
arezaei@cs.stonybrook.edu

Jeff M. Phillips
University of Utah
jeffp@cs.utah.edu

Jie Gao
Stony Brook University
jgao@cs.stonybrook.edu

Csaba D. Tóth
Cal State Univ. Northridge
csaba.toth@csun.edu

ABSTRACT

With the popularity of portable wireless devices it is important to model and predict how information or contagions spread by natural human mobility – for understanding the spreading of deadly infectious diseases and for improving delay tolerant communication schemes. Formally, we model this problem by considering M moving agents, where each agent initially carries a *distinct* bit of information. When two agents are at the same location or in close proximity to one another, they share all their information with each other. We would like to know the time it takes until all bits of information reach all agents, called the *flood time*, and how it depends on the way agents move, the size and shape of the network and the number of agents moving in the network.

We provide rigorous analysis for the Manhattan Random Waypoint model (which takes paths with minimum number of turns), a convenient model used previously to analyze mobile agents, and find that with high probability the flood time is bounded by $O(N \log M[(N/M) \log(NM)])$, where M agents move on an $N \times N$ grid. In addition to extensive simulations, we use a data set of taxi trajectories to show that our method can successfully predict flood times in both experimental settings and the real world.

CCS CONCEPTS

• **Information systems** → **Mobile information processing systems**; • **Computing methodologies** → **Mobile agents**; • **Networks** → **Mobile networks**;

KEYWORDS

Mobile Information Systems, Mobile Agents, Information Dissemination

ACM Reference Format:

Aria Rezaei, Jie Gao, Jeff M. Phillips, and Csaba D. Tóth. 2018. Improved Bounds on Information Dissemination by Manhattan Random Waypoint Model. In *26th ACM SIGSPATIAL International Conference on Advances in*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGSPATIAL '18, November 6–9, 2018, Seattle, WA, USA

© 2018 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.

ACM ISBN 978-1-4503-5889-7/18/11...\$15.00

<https://doi.org/10.1145/3274895.3274917>

Geographic Information Systems (SIGSPATIAL '18), November 6–9, 2018, Seattle, WA, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3274895.3274917>

1 INTRODUCTION

It has always been an interesting research topic to understand human mobility and how contagions spread via such motion. One of the motivations is to understand how infectious diseases spread by moving agents. Think of a strain of an infectious virus such as SARS or Ebola. When two individuals are at the same location at the same time, there is a possibility for one to spread that strain of virus to the other. Therefore the spread of contagions in a population is highly dependent on the density of the population and how individuals move around. In another example, human motion can be used for our benefit. It has become common that people carry wireless devices around. Short range low cost wireless communication can be established at the contact events to allow energy efficient information exchange. In this case the mobility model influences how long it takes for a piece of information from one node to reach all nodes in the network.

There have been two main approaches to study human mobility in the literature: data-driven methods versus theoretical analysis. In recent years wireless technology has made it possible to collect a large amount of mobility data through wireless devices. There has been a lot of work on finding exciting patterns in human movements and information spreading in real-world data [1, 19, 23]. It has been shown that human mobility is immensely complex. Accurate models can be built using historic traffic data to predict agents' locations and social ties [20, 42]. But these models each work only for a specific scenario. It is unclear how these models can help us analyze asymptotic behavior of moving agents or whether a model generalizes to a different geographical location, a different travel modality, or a different group of people. Furthermore, long-term mobility data can be identity revealing. Even with great efforts to anonymize the data and with removing big fractions of it, individuals are identifiable by their movement patterns. A seminal work revealed that 4 randomly selected points in an hourly location sequence of a person recorded over 15 months via cellphone antennas is enough to make that person identifiable among 1.5 million individuals [17]. As a result, mobility data sets are usually not published by companies due to concerns over user privacy, except for a few special cases of shared vehicles (taxis or shared bikes).

On the other hand, an extensive amount of work has been dedicated to mathematical models of mobility and their asymptotic

behaviors. Although these models cannot compete with the accuracy of data-driven models in the presence of enough historic mobility data, they have been used for their rigorous analysis and their ability to predict future events with provable certainty. Over the years theoretical models have evolved from simplistic models, inspired by known physical phenomena in real world, to more sophisticated ones, taking into account the complexity of human mobility. We briefly review these models and their analytical results below. For a comprehensive survey on these models refer to [8].

- **Random Walk:** Perhaps the most studied movement model. Inspired by the movement of floating particles in a liquid or gas, called *Brownian Motion* [22], in its simplest version an agent starts its movement in an arbitrary node in a given network. At each time step, the agent chooses one of the neighbors of its current node *uniformly at random* and moves to that neighbor. There are variations in which agents can rest in their current position for a period of time or use a non-uniform transition probability when choosing a neighbor.
- **Random Direction Model:** In this model, an agent chooses a random direction and possibly a random velocity, then moves in that direction until it collides with the boundary of the network. The agent then chooses a new direction and velocity and continues as before.
- **Random Way-point:** In this model, an agent, starting from an initial position in the network, chooses the next destination from *all* nodes in the network *uniformly at random*. Then, using one of the shortest paths, moves towards the destination and after reaching it finds a new destination with the same method. This has been widely used in modeling human motion and in many prior simulations for mobile networks [6, 7, 26, 29, 37].
- **Manhattan Random Way-point :** This is a special case of the Random Way-point model [6]. In a grid (or *torus*) networks, agents move to the destination with as few turns as possible. Thus they travel first horizontally and then vertically (or vice versa) to the destination. This model is inspired by the fact that in urban streets, turning can be time-consuming [13, 16].
- **Lévy Walk:** Studies on intelligent moving agents, especially humans, have revealed that the distance to the next destination chosen by such agents seems to follow a *fat-tailed* distribution [2, 25, 43]. A popular movement model in this category is called *Lévy Walk* [41]. This model is similar to Random Way-point, but instead of choosing a new destination uniformly at random, an agent chooses a node as its next destination with probability proportional to the inverse of its distance from the agent's current position, to some power $\alpha > 0$.

1.1 Our Contributions

In this paper we provide improved upper bounds on the rate of information dissemination when agents move according to the Manhattan Random Way-point model. Through a series of simulations we show that, combined with bounds on Random Walk, our bounds lead to a new conjecture on the time it takes for information to disseminate through a network when agents' movements follow

the Lévy Walk model, a challenging question that remains widely open. Finally, we report the result of a series of experiments we have designed, which show that our model is capable of predicting trends on experimental settings, as well as real-world data.

We formally define our problem as follows. Consider a set of M autonomous agents, each starting at time 0 with a unique bit of information, b_i , at a node selected uniformly at random in an $N \times N$ torus¹ denoted by G . Agents all follow the same movement model and they share information with each other when they meet during their move. Meeting is defined as being at the same location at the same time, where the location can be inside a node or on an edge between two nodes. Agents start their movement at the beginning of each time step in a synchronized fashion. We consider uniform speed for all agents² and transmission radius is practically 0 as agents have to be collocated in order to pass along information, for simplicity. However our findings can be extended for arbitrary constant transmission radius.

In the above setting, we are interested in finding bounds on the time it takes until *every* agent finds out about *every* piece of information. This value is called the *flood time* (T_F). An equally important statistic is the time it takes for all agents to learn a *specific* bit of information, called *broadcast time* and denoted by T_B . Clearly $T_B \leq T_F$. Using the *union bound*, any upper bound on T_B extends to T_F , too, if the probability of the bound occurring is *sufficiently high*³. Both T_F and T_B are important statistics in various applications. In the case of disease spreading, T_B corresponds to the time when an infection of *any* agent would have been passed to the entire population. In a delay-tolerant wireless mobile network, T_F corresponds to the time in the past, from which we can assume all information has been shared across the network; we can predicate the start of a new protocol based on assuming all agents are up to date after this delay. In mobile social networks, T_B corresponds to the time it takes a new piece of information to permeate society. Overall, both T_F and T_B are important statistics which capture information flow in a network, and will be the focus of our study.

Our contributions are as follows:

- We find a new upper bound for T_F and T_B that is tight for a wide range of settings. Specifically, when M agents move on an $N \times N$ grid with torus topology, we show $T_B \leq T_F = O(N \log M \lceil \frac{N}{M} \log(NM) \rceil)$. This bound improves upon recent upper bounds for topologies with more complex boundary conditions.
- We analyze the relation between Random Walk, Manhattan Random Way-point and Lévy Walk. Through simulations, we show empirically that the Lévy Walk model can be understood by carefully interpolating between the heavily studied Random Walk model and our new results on the Manhattan Random Way-point model.

¹A grid where nodes in the boundaries are each connected to their corresponding node in the opposite boundary.

²In some previous definitions of Random Way-point the agents choose their speed uniformly at random from a range. But this choice will lead to the average moving speed to be decreasing over time [45]. Also in reality vehicles/pedestrians often move with a fixed speed.

³If $\Pr\{T_B > T\} < 1/N^d$, for some $d > 0$, we know that $\Pr\{T_F > T\} < M/N^d < 1/N^{d'}$, when d is sufficiently large.

- We validate the theoretical bounds in a number of empirical studies using simulated scenarios, bike and taxi trajectory data sets.

2 BACKGROUND AND RELATED WORKS

Since the advent of social networks, researchers studied how information (e.g., rumors or viral videos) spreads through a network [21, 24, 30]. The rate of change in these networks is slow enough that they can be considered *static* throughout the course of information dissemination. This assumption simplifies the mathematical models tremendously. These studies do not fit *dynamic* mobile networks, due to the high rate of topology changes. The spreading behavior heavily depends on how agents move [32].

For analyzing the mobility models mentioned earlier, arguably the simplest model for a geographically spread network is a 2D grid. In some metropolitan settings the downtown area is a reasonable grid. Wrapping the grid around to a torus has been commonly adopted in prior papers that analyze information spreading in a mobile network. The benefit of the torus model is to get rid of the boundary effect, thus simplifying the analysis. The torus also removes the boundary effect in the Random Way-point model, which often caused unwanted artifacts in simulations [3, 11, 39].

For direct comparison, we review prior results in the same format as ours, where N and M are decoupled and there are no assumptions on their ratio, near-zero transmission radius ($R \approx 0$), and where each *move* consists of walking along a path from source node to destination node, rather than *jumping* to the destination instantaneously.

The Random Walk movement model has been extensively studied and tight bounds on many characteristics have been fully resolved [18, 31, 33]. One of the tightest bounds for Random Walk is given by Patterin et al. [38]. They found that even with a very small transmission radius, broadcast time does not depend on the relation between the mobility speed and the transmission radius. They prove that with high probability (w.h.p):

$$T_B = \tilde{O}\left(N \left\lceil \frac{N}{\sqrt{M}} \right\rceil\right). \quad (1)$$

Keeping track of a bit of information b , they first divide the initial grid into smaller cells and find the time by which an arbitrary cell is *infiltrated* by an agent carrying b . Then they show that this infiltrating agent will inform the majority of agents near this cell and spread the information *locally*. After the local spread is done within the cell, information *leaks* into adjacent cells and this whole process is repeated. Ultimately, every cell is infiltrated and every agent in the network finds out about b .

Clementi et al. [14] have proved bounds for T_F in Manhattan-like grids where agents move according to the Manhattan Random Way-point model. Their setting slightly differs from ours as they do not employ boundary loops, which results in two zones with very different traffic of agents. On the one hand, nodes in the *central zone* are visited by agents across all nodes in the grid with high probability. On the other hand, the periphery, called *suburb* areas, are starved of agents: the probability that an agent passes through these areas is significantly lower than that of the central zone. They prove that w.h.p. $T_F = \tilde{O}(N/R + N^3/vR^2M)$, where R is the transmission radius and v is the agents' speed. We can rewrite this

in our setting as:

$$T_F = \tilde{O}\left(N + \frac{N^3}{M}\right). \quad (2)$$

They found that the bulk of the Flooding Time is devoted to carrying the information to the “suburbs,” as it requires a flow of informed agents traveling from the central zone to the suburbs.

A different line of work focused on solving the problem in a general platform, oblivious to geometric considerations. Clementi et al. [15] have derived an upper bound for T_F , and the *mixing time*⁴ of the Markov chain corresponding to agent movements, as well as how independent the collisions between different pairs are, play a major role in the analysis of T_F . Applying Manhattan Random Way-point model to their general bound yields the following:

$$T_F = O\left(\frac{N}{v_{\max}} \left(\frac{N^2}{MR^2} + 1\right) \log^2 M\right),$$

where v_{\max} is the maximum speed of any agent. Rewritten in our setting, the bound is:

$$T_F = \tilde{O}\left(N \left\lceil \frac{N^2}{M} \right\rceil\right). \quad (3)$$

Since their method is very general, their bound is not competitive with the bounds on specific movement models and networks. Table 1 shows that our bound is a significant improvement over these bounds on different movement models.

Table 1: Recent bounds on Manhattan Random Way-point (MRWP), Random Way-point (RWP) and Random Walk (RW).

Authors	Model	Bound
Ours	MRWP	$\tilde{O}(N \lceil N/M \rceil)$
Clementi et al. 2010 [14]	MRWP	$\tilde{O}(N + N^3/M)$
Clementi et al. 2015 [15]	RWP	$\tilde{O}(N \lceil N^2/M \rceil)$
Pettarin et al. 2011 [38]	RW	$\tilde{O}(N \lceil N/\sqrt{M} \rceil)$

Rigorous analysis of the Lévy Walk model is much more challenging due to the strong spatiotemporal correlation [4, 34, 40]. The most relevant work is done by Wang et al. [44] where the authors have analyzed the distribution of the minimum time needed until a piece of information reaches a certain region. However, to the best of our knowledge, there are no bounds for T_F when the agents are moving according to a Lévy Walk.

Besides theoretical work, there has been a lot of empirical analysis of how information can spread through opportunistic peer communication among mobile agents using simulations or empirical evaluations. Protocols for reducing communication cost (e.g., to avoid a message be delivered to a node multiple times) have been studied extensively [10, 28, 36]. Last, there has also been work on a model assuming that a supporting static wireless network is in place, which helps to cache and propagate these events [46]. This model is different from ours.

⁴Mixing time is the time needed for a Markov chain to reach its stationary distribution, starting from an arbitrary distribution.

3 BOUNDS ON FLOOD TIME

In this section, we present and prove our main theoretical results on the flooding time on torus networks. We start with a trivial lower bound.

THEOREM 3.1. *For M agents initially positioned at uniformly random nodes and moving with constant speed in an $N \times N$ torus, with constant probability, we have:*

$$T_F \geq T_B \geq \Omega(N). \quad (4)$$

PROOF. Let b_i be the bit of information initially carried by agent A_i , and A_j the farthest agent to A_i at time 0. The time it takes until b_i reaches A_j , denoted by $T(A_j, b_i)$, is a lower bound for both T_F and T_B . Since we are assuming uniformly random initial positions, with $1/2$ probability the distance between A_i and A_j is at least $N/2$. Also note that information can only travel as fast as agents can. This means that at each time step the distance between A_j and the closest copy of b_i is reduced by at most 2 units. As a result, with probability $1/2$, $T(A_j, b_i)$ is at least $N/4$, which completes the proof. \square

Now we move forward to our main theorem on the upper bound for Manhattan Random Way-point model.

THEOREM 3.2. *For M agents moving according to the Manhattan Random Way-point model with constant speed in an $N \times N$ torus, with high probability⁵, we have:*

$$T_B \leq T_F = O\left(N \log M \left\lceil \frac{N}{M} \log(NM) \right\rceil\right). \quad (5)$$

We present the proof in two parts. The first part is to analyze when and with what probability two agents have a collocation event so they can share information. This involves a few necessary conditions: time-wise the moves made by two agents with a collocation event need to overlap; geometrically their trajectories need to have an intersection; and thirdly, they arrive at the intersection at the same time. Next we need to analyze the global property: how information sharing enabled by collocation events leads to global dissemination. The next two subsections focus on these two parts of the proof respectively.

3.1 Bounding Collocation Probability

We consider agents moving non-stop following Manhattan Random Way-point model, and partition the mobility trace of each agent into *moves* between randomly chosen destinations. First, observe that any move by an agent following Manhattan Random Way-point model takes $\Omega(N)$ time with constant probability. We call the two straight parts of a movement, one horizontal and the other vertical, *segments*. Note that any move in this model can have at most two segments of different directions appearing in an arbitrary order.

We say two agents have a *connection* during their moves if they are at the same location at the same time. For that to happen, the moves of the two agents must at least overlap over time.

Definition 3.3. *Strongly overlapping moves* are moves made by two agents where the *time interval* of a segment of an agent's move is completely contained within the time interval of the other agent's whole move.

⁵High probability in our work means at least $1 - 1/N^d$ for some constant $d > 0$.

Note that this overlap only needs to happen in the time interval of two moves and no condition is imposed on their geometric locations.

LEMMA 3.4. *Every move M_i of an agent A_i strongly overlaps with at least one of the moves of another agent A_j , say M_j . The starting moment of the moves can be at most N time-steps apart and with constant probability, a segment of M_i will have a time duration overlap of $\Omega(N)$ with a segment of M_j .*

PROOF. First consider the move M_j^- of A_j that is still active when M_i starts. If M_j^- ends after the first segment of M_i , then M_j^- covers the first segment of M_i and the two moves strongly overlap (Figure 1 case (i)). If not, then consider the next move of A_j , M_j^+ . If M_j^+ ends after M_i ends, M_j^+ completely covers M_i 's second segment and the two moves strongly overlap (Figure 1 case (ii)). If not, then M_j^+ must be completely covered by M_i and the two moves strongly overlap (Figure 1 case (iii)). We have shown that a strong overlap occurs. Since the time duration of each move can be at most N , the starting point of M_i is at most N time-steps apart from the starting points of both M_j^- and M_j^+ .

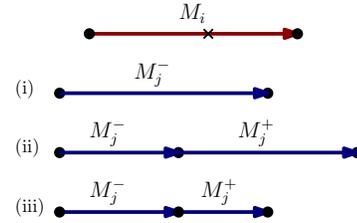


Figure 1: Any move of one agent strongly overlaps with at least one of the moves of another agent.

Without loss of generality, assume a segment S_j of M_j is covered completely by M_i . With probability $1/2$, the time duration of S_j (equal to its length due to the constant speed assumption) is at least $N/4$. Now take the segment of M_i with the longest time duration overlap with S_j , and denote it by S_i . This overlap should be at least $|S_j|/2$. As a result, with probability $1/2$, the time duration of this overlap between S_i and S_j is at least $N/8$, as required. \square

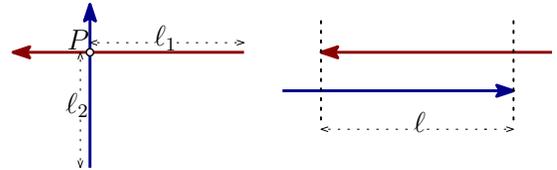


Figure 2: The segments are trimmed to duration of their overlap, ℓ . Left: The collocation event of two agents happens if the distance between the intersection point (P) and the starting point of both segments is equal ($\ell_1 = \ell_2$). Right: Connection happens if there is a non-zero overlap between two segments and there are 2ℓ such placements for the blue segment if the red segment is fixed.

LEMMA 3.5. *If two agents A_i and A_j have two segments S_i and S_j with a time interval overlap of ℓ , they meet with probability $\Theta(\ell/N^2)$.*

PROOF. We trim both segments to the duration of their overlap, making them of equal length ℓ . All cases of a connection between the two agents can be reduced to three main cases below by rotating the torus or swapping the agents:

- (1) S_i is horizontal and S_j vertical. Here, the two agents *connect*, if the two segments intersect geometrically at a point P and P is at equal distance from the starting points of S_i and S_j (see Figure 2, left). Since we assume that G is a torus, we can fix S_i 's position in our analysis. Out of all N^2 possible placements of S_j on G , there are ℓ placements that result in an intersection that meets the above condition. Hence the probability of a *connection* between A_i and A_j is ℓ/N^2 in this scenario.
- (2) Both S_i and S_j are horizontal and in opposite directions. In this case, any geometric intersection is enough for a connection to happen (see Figure 2, right). If both agents move in the same row, there would be 2ℓ placements of S_j once S_i 's position is fixed that results in an intersection. The probability of both agents moving in the same row is $1/N$, which makes the overall probability $2\ell/N^2$.
- (3) Both S_i and S_j are horizontal and in the same direction. In this case, the starting points of S_i and S_j have to be in the same exact node, which happens with probability $1/N^2$.

Based on the 4 possible directions of each segment, there are 16 possible cases for S_i and S_j , all of which happens with equal probability and can be reduced to one of the 3 cases above. The overall probability of a connection between two agents can be bounded as:

$$\frac{\ell}{N^2} \leq \Pr\{\text{Connect}(A_i, A_j, \ell)\} \leq 32 \frac{\ell}{N^2}. \quad (6)$$

□

An immediate consequence of Lemmas 3.5 and 3.4 is:

LEMMA 3.6. *Two agents with strongly overlapping moves have a connection with probability $\Theta(1/N)$.*

PROOF. For the probability of two strongly overlapping agents connecting we have:

$$\Pr\{\text{Connect}(A_i, A_j)\} = \sum_{x=1}^{N/2} P(\ell = x) \Pr\{\text{Connect}(A_i, A_j, x)\}, \quad (7)$$

where $P(\ell = x)$ is the probability that the time duration of the overlap between segments of A_i and A_j is x . From Lemma 3.4 we know that $P(\ell \geq N/8) \geq 1/2$. As a result we can rewrite (7) as:

$$\Pr\{\text{Connect}(A_i, A_j)\} \geq \frac{1}{2} \frac{N/8}{N^2} \geq \frac{1}{16N}.$$

The upper bound in (6) yields:

$$\Pr\{\text{Connect}(A_i, A_j)\} \leq \max_x \Pr\{\text{Connect}(A_i, A_j, x)\} \leq \frac{16}{N}.$$

□

3.2 Bounding the Flood Time

Given the probability of two agents sharing information during their moves, we are now ready to argue for how information propagates to the entire network. For simplicity, we find an upper bound for the broadcast time T_B (one specific message reaching everyone) and extend it to the flood time T_F (all messages reaching everyone). There are two issues that we need to address. First, the positions of an agent are temporally correlated—but fortunately, if sufficiently far apart in time, the positions of agents are independent which will help to simplify our analysis. Second, we need to track agents who have been informed and who have not, and analyze how information spreads from the informed ones to the uninformed ones.

LEMMA 3.7. *The locations of two agents at $2N$ steps apart are independent of each other.*

PROOF. Let the sequence of destinations chosen by an agent A be (X_0, X_1, \dots, X_k) . Observe that regardless of what X_i is, every node in the torus (including X_i) has the same probability of being X_{i+2} . Also, nodes visited between two consecutive destinations are only dependent on those two destinations. Let P_t and P_{t+2N} be the position of A at time t and $t + 2N$. Since the agents move along the shortest path towards their destination, a move is at most N steps long. As a result, at least two destinations will be visited between time t and $t + 2N$ by A . Take the destination immediately after P_t , X_i , and immediately before P_{t+2N} , X_j . As noted above, P_t is only dependent on X_{i-1} and X_i , while P_{t+2N} is only dependent on X_j and X_{j+1} , where X_i and X_j are distinct positions. This makes P_t and P_{t+2N} completely independent. □

As mentioned earlier, we track the spread of a single bit of information b among the agents. We first divide the time span of the whole process into windows of size $6N$ time steps, called *cycles*. The k^{th} cycle starts at time $6kN$ and ends at time $6(k+1)N$. Each agent would visit at least 4 destinations between time $(6k+1)N$ and $(6k+5)N$, which yield at least 3 moves independent of other cycles in an agent's trajectory (note the $2N$ margin between selected moves in each cycle, which guarantees independence). According to Lemma 3.4, we know that a move by an agent will strongly overlap with a move by another agent and the starting point of the two moves are at most N time steps apart. As a result, for each pair of agents and each cycle, we can find two strongly overlapping moves independent of other cycles, which according to Lemma 3.6 have a c/N chance of connection, where c is a constant. This essentially makes collocation events between a fixed pair of agents in different cycles i.i.d.

We now divide the whole process into consecutive *epochs*. Each epoch consists of s_i cycles, and starts when we have a set I_i of agents who know about b (referred to as *informed* agents) and a set U_i of agents who do not (referred to as *uninformed* agents). An epoch ends when the number of informed agents doubles, or the number of uninformed agents drops to zero. For the broadcast time of b , T_B , we can write:

$$T_B \leq 6N \sum_{i=1}^{\log M} s_i. \quad (8)$$

We now find the number of cycles needed for that w.h.p. each agent in I_i is paired with a *distinct* agent in U_i during the i^{th} epoch.

By artificially forcing informed agents to find distinct partners, we only slow down the process of information spread, and an upper bound found in this manner is valid as an upper bound for the main problem. The reason we require distinct partners for each informed agent is to ensure that connections between different pairs are independent.

The probability of an informed agent $A \in I_i$ connecting to *any* $A' \in U_i$ is as follows (arrow shows the direction of information exchange):

$$\begin{aligned} \Pr\{\nexists A' \in U_i : A \rightarrow A'\} &= \left(1 - \frac{c}{N}\right)^{s_i |U_i|}, \\ \Pr\{\exists A' \in U_i : A \rightarrow A'\} &= 1 - \left(1 - \frac{c}{N}\right)^{s_i |U_i|}. \end{aligned} \quad (9)$$

Let there be an arbitrary order for agents in I_i and one for agents in U_i . The first informed agent can match to any of the $|U_i|$ uninformed agents. After the first matching is done, there will be $|U_i| - 1$ potential matches for the second informed agent and so on. Assuming that there are q pairs at the end of this epoch ($q \leq \min(|I_i|, |U_i|)$) and using Equation (9), the probability of this happening (i.e., having q pairs of matched informed/uninformed agents) is:

$$\begin{aligned} \Pr\{q\text{-matching}\} &= \prod_{i=0}^{q-1} \left(1 - \left(1 - \frac{c}{N}\right)^{s_i(M-q-i)}\right) \\ &\geq \left(1 - \left(1 - \frac{c}{N}\right)^{s_i(M/2)}\right)^{q/2} \\ &\geq \left(1 - \exp\left(\frac{-s_i c(M/2)}{N}\right)\right)^{q/2} \\ &\geq 1 - q \exp\left(\frac{-s_i c(M/2)}{N}\right). \end{aligned}$$

For the above to happen w.h.p., for $d > 1$, we need:

$$\begin{aligned} 1 - q \exp\left(\frac{-s_i c(M/2)}{N}\right) &\geq 1 - \frac{1}{N^d} \\ \exp\left(\frac{s_i c(M/2)}{N}\right) &\geq q N^d \\ s_i &\geq \frac{2N}{cM} (\log q + d \log N) \\ s_i &> \frac{2dN}{cM} \log(NM). \end{aligned}$$

The last step is due to the fact that $q < M$. To make sure that s_i is a non-zero integer and we have at least one cycle, we set:

$$s_i = \left\lceil \frac{2dN}{cM} \log(NM) \right\rceil. \quad (10)$$

Substituting (10) into (8), we have:

$$T_B \leq 6N \sum_{i=1}^{\log M} \left\lceil \frac{2dN}{cM} \log(NM) \right\rceil \quad (11)$$

$$= O\left(N \log M \left\lceil \frac{N}{M} \log(NM) \right\rceil\right). \quad (12)$$

Since our bound for T_B works for arbitrarily high probability ($1 - 1/N^d$, for a constant $d > 0$), it extends to T_F using the Union Bound. This completes the proof of Theorem 3.2.

This is a pretty tight bound. Consider a semi-dense scenario where $M \approx N$, our bound becomes $\tilde{O}(N)$, which nearly meets the trivial lower bound of Equation (4).

Compared to previously mentioned bounds for the Random Way-point model in (2) and (3), our bound is stronger. For the same movement model, although under slightly different network assumptions, Clementi et al. found the bound of $\tilde{O}(N + N^3/M)$ [14], which is mainly due to the choice of not having a torus as they intended to study the impact of rarely visited areas on the total information spread time. Our bound also improves that of [15] by a huge margin. This can be due to the fact that their method is a general framework to find an upper bound for T_F . Our version of the Manhattan Random Way-point model assumes that agents complete their move in one coordinate then start moving in another. As suggested in [15], this assumption increases the probability of connection between two agents, which in turn leads to a better upper bound for T_F and T_B . Further, the Manhattan Random Way-point model is a better fit for mobility in urban areas it implicitly incorporates the cost of turning during movement.

Depending on the application, one can think of various extensions to our model, such as an arbitrary transmission radius or random waiting time between two consecutive moves of an agent. In this case, an approach similar to ours can be adopted to find a bound for T_F if these three components are available: (1) suitably sized independent time windows (called *cycle* here), (2) guarantee of a long enough time interval overlap between segments of moves by two agents and (3) probability of connection between those two segments.

4 EXPERIMENTS

In this section, through a series of experiments we test the accuracy of our discovered bound. First, we test our model where agents are moving in a torus-like grid, following Manhattan Random Way-point model. Next, using bike sharing records in 3 major cities, we create synthetic trajectories in real-world road networks, and compare simulated behavior of flood time against our model. Finally, using GPS traces of taxis in a major city, we verify our model against a real-world case of information dissemination via mobile agents.

4.1 Simulated Movements in a Grid

We simulate the movement of agents in a torus-like grid following a Manhattan Random Way-point model, and compare the average flood times, \overline{T}_F , against our bound. First, we fix N to 100 and find \overline{T}_F for $M = \{25, 50, \dots, 1000\}$ by averaging T_F over 100 realizations. To fit the resulting values to our bound, we use function $f(M) = \frac{c_1}{M} \log(c_2 M) \log(c_3 M)$, where each c_i is a positive constant, accounting for fixed N and constants in our asymptotic analysis. The results of the simulation along with fitted values are shown in Figure 3, middle. Our bound has accurately captured the changes in \overline{T}_F as the coefficient of determination, R^2 , is equal to 0.9894. Next, using a similar procedure, we fix M to 100 and find \overline{T}_F for $N = \{25, 50, \dots, 500\}$. We fit the values to the function $f(N) = c_1 N^2 (\log N + c_2)$, where again each c_i is a positive constant. The simulation results and the fitted values are depicted in Figure 3. As expected, we showed good performance here too by yielding an R^2 of 0.9894 (equality between the two R^2 values is coincidental).

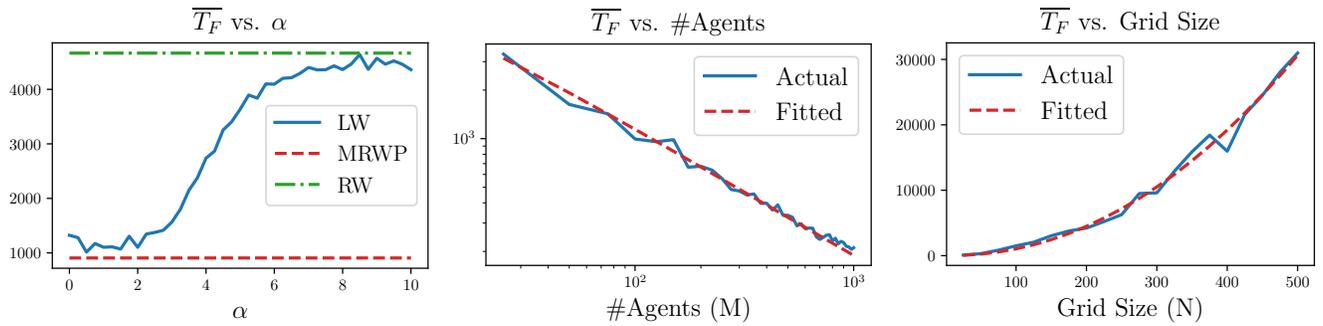


Figure 3: (left) The sigmoid-like behavior of Lévy Walk, sandwiched by Random Walk and Manhattan Random Way-point. (middle, right) Our bound accurately captures the changes in \overline{T}_F as M and N are tweaked while other parameters are fixed.

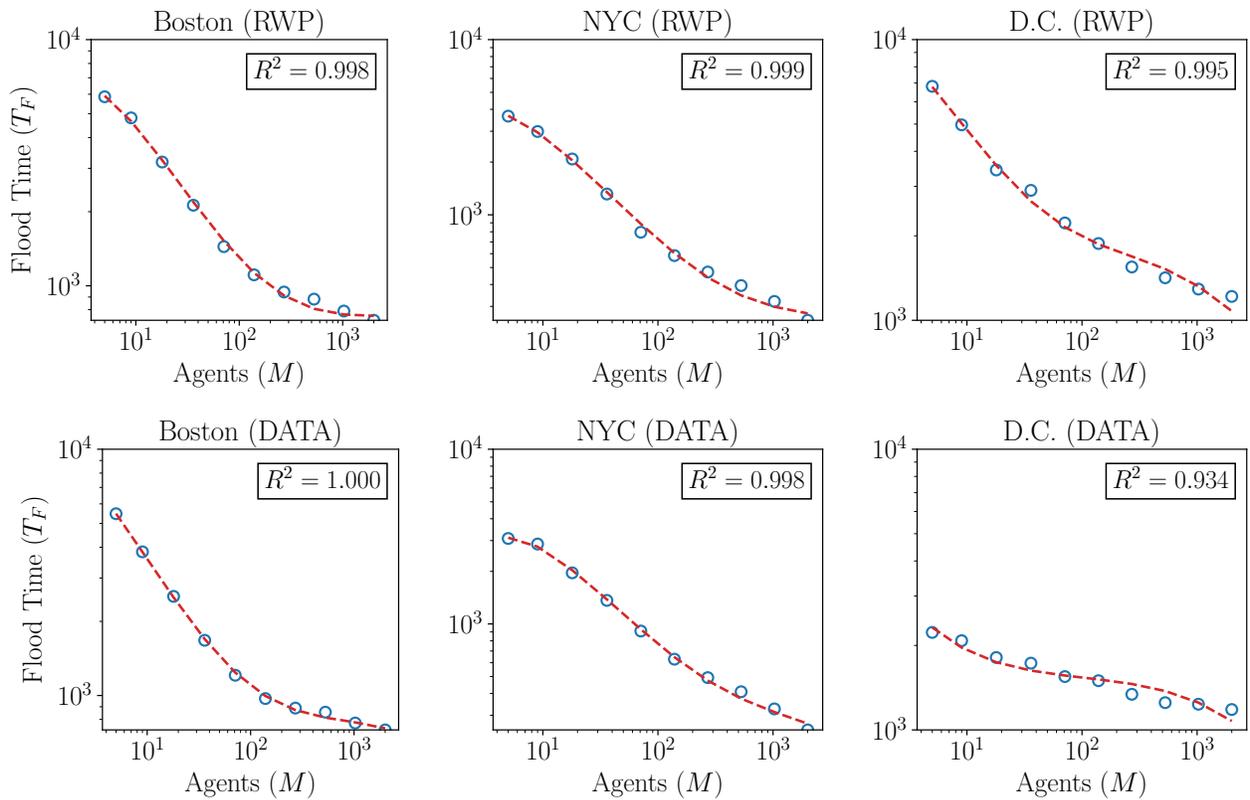


Figure 4: Actual (empty circles) and fitted (dotted lines) \overline{T}_F for two movement policies in 3 cities.

4.2 Simulated Movements in Real Networks

To test our model against non-grid networks, we use the bike rental records of 3 major US cities [9, 12, 27]. Each city has a unique road network and a set of fixed stations, $\{S_1, S_2, \dots, S_k\}$, which are used as the set of possible destinations each agent can choose from. The goal here is to test our model against a setting beyond the grid network and uniformly random selection of destinations. The data sets include an origin and a destination station for each

trip made. Using these records, we can estimate the probability of choosing a destination S_j given that an agent is currently positioned in station S_i , called the *transition* probability between S_i and S_j and denoted by $P(S_i, S_j)$. We can also calculate the probability of initiating a trajectory from any given station, from here on called initiation probability and denoted by $P(S_i)$. To gradually move our tests away from the theoretical settings, we use the following movement models throughout our simulations:

Table 2: Fitting score, R^2 , for all movement policies in all cities.

City	RWP	DATA
Boston	0.998	1.000
New York City	0.999	0.998
Washington, D.C.	0.995	0.934

- (1) Similar to Section 4.1, we select each station in the sequence of stations visited by an agent uniformly at random. This is equivalent to the Random Way-point model and denoted by *RWP* in this experiment.
- (2) Next, we use the calculated $P(S_i)$ and $P(S_i, S_j)$ values to build synthetic trajectories. We call this model *DATA* throughout this experiment.

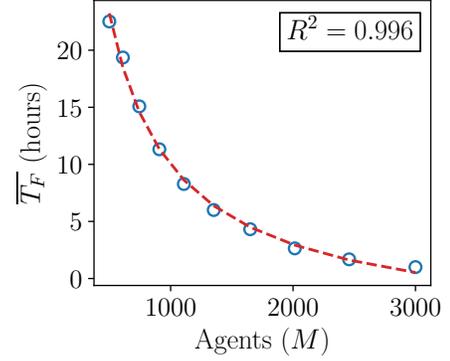
In each experiment, after selecting a sequence of stations visited by each agent, we find the shortest paths between consecutive stations using Routino [5] and OpenStreetMap [35] extracts. Two agents will *connect*, if at any time t they are closer than 100 meters from each other. Finally, for each of these experiments, we iterate over 10 values of M between 5 and 2000, and report the average flood time (\overline{T}_F) by aggregating over 25 realizations.

Using a function similar to Section 4.1, we can fit the simulation results to our bound. Figure 4 shows the actual and fitted values for the 3 cities, along with the R^2 value of the fitting. In these simulations, our bound closely approximates the flood time in the simulations, even when the movement policy used is data-specific rather than the Random Way-point model. The R^2 values are compared in Table 2. Across all settings, we achieve > 0.93 (5 of them > 0.99), which shows the flexibility of our model to variations of network and movement policy. Note that here the network was a real-world road network, and far from a torus.

There can be many different factors contributing to the flood time in networks as complicated as urban maps, which are beyond the scope of this study. Here, we tried to explore the limits of our model’s prediction capabilities by tweaking the settings of experiment in a controlled manner. Further investigation in the effects of structural properties of road networks, and different distributions of frequent origins and destinations on the flood time is needed to fully understand the process of information dissemination by human mobility in real road networks.

4.3 Real-World Data

Next, we try to fit our model to real-world GPS traces. Ideally, one may want to experiment on personal trajectories, as the behavior of a *single* moving agent is best understood by looking at individuals’ mobility traces. However, due to the sensitivity of such data, large and high-quality data sets containing personal mobility traces are extremely rare. As a substitute, we can study the mobility of shared vehicles, as we did in the previous section. Here, we study GPS traces of taxi cabs in the city of Shenzhen in China [19]. Over the course of 24 hours, the location of 9386 taxis are sampled every 1.01 minutes. We set the *transmission radius* to 100 meters and, for simplicity, assume that connections can happen only on sampled points in time. To generate different numbers of moving agents (M), we have to subsample from the set of all taxis. Since these

**Figure 5: Actual (empty circles) and fitted (dotted lines) \overline{T}_F for Shenzhen.**

trajectories are fixed, we cannot extend them in the event of having no information flood. Hence, we filter out those taxis that meet less than 100 distinct taxis during the whole 24 hours, and 3905 taxis will remain. We iterate over 10 different values of M between 500 and 3000, each time finding the flood time in hours. We average the results of 25 realizations for each M and report it. The results are shown in Figure 5. We have followed the same procedure to fit the simulation values to our bound. The resulting *fitted* line is drawn in Figure 5, achieving an R^2 value of 0.996. This shows that our model is capable of predicting flood times for real-world scenarios to some degree. It is worth noting that the real-world experiments did not have significant fluctuations in the flood time value and, similar to controlled experiments in the two sections before, shows a smooth behavior, even with only hundreds of moving agents in some cases.

5 TOWARDS BOUNDING THE LÉVY WALK

Compared to other mobility models, the Lévy Walk is far less studied. Formally, in a Lévy Walk, given a constant $\alpha > 0$, an agent positioned at its i^{th} destination, $X = X_i$, chooses node Y as its next destination, X_{i+1} , with the following probability:

$$\Pr\{X_{i+1} = Y \mid X_i = X\} = \frac{1}{Z \|Y - X\|^\alpha}, \quad (13)$$

where Z is the normalizing factor and $\|X - Y\|$ is the distance between nodes X and Y , such as *Manhattan Distance* in a grid, *Euclidean Distance* in the 2D plane or *Graph Shortest Path Distance* in any given network. Figure 6 compares a Random, a Lévy and a Random Way-point walker, simulated for 250 steps. Notice that a Random Way-point walker tends to take big steps and cover a vast area in the grid, while a Random walker is concentrated to a small area around its initial position. A Lévy walker shows a mixture of the two behaviors. It roams around in a small area most of the time, but occasionally makes a long move to a different region in the grid.

To compare how information propagates in the three movement models, we have to first observe that Random Walk (or *Brownian motion*) and Random Way-point can be thought of as two extreme

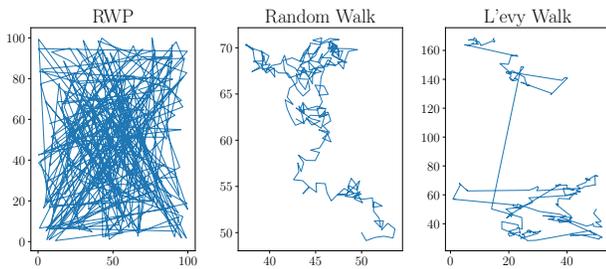


Figure 6: A Random Way-point walker (left), a random walker (middle), and a Lévy walker ($\alpha = 2$) after 250 steps.

ends of the spectrum of all possible Lévy Walks. In (13), setting α to 0 (and applying the corresponding Z value) yields a constant probability regardless of the distance between X and Y , similar to Random Way-point. On the other hand, given any time limit T , we can make α high enough so that w.h.p. no agent selects a destination more than one unit distance away at any time $t \leq T$, effectively forcing them to follow Random Walk.

Figure 3, left, shows the simulated results for the average flood time, denoted by \overline{T}_F , of a system where agents are moving by Random Walk, Manhattan Random Way-point or Lévy Walk model with different values of α . Both N and M are set to 100, α goes from 0 to 10 in 0.1 increments, and each point is created by aggregating the results of 100 realizations. Additionally, with our newly discovered bound for \overline{T}_F in Random Walk and Manhattan Random Way-point have gotten very close. We now have a reason to believe that any future bound for Lévy Walk should be close to either of the bounds for these two movement models. And since their bounds are close, it is worth investigating whether or not a careful interpolation of the bounds for Random Walk and Manhattan Random Way-point is a good predictor of how a Lévy Walker moves in a network.

6 CONCLUSION

Thanks to ever-present portable devices, there has been a growing interest in a better understanding of mobile networks (also called *vehicular* networks), where autonomous agents move independently and are capable of carrying and transmitting information. We studied the case of M agents moving in an $N \times N$ torus.

We made a new improvement to the flood time bound for Manhattan Random Way-point model, $T_F = \tilde{O}(N \lceil N/M \rceil)$, that is tight for a wide range of problem settings. To the best of our knowledge, this bound is stronger than all previous bounds found for this movement model. Through extensive experiments, we showed that our bound can accurately predict flood time for a wide variety of simulated and real-world settings.

Lastly, given the shrinking difference between the bounds for Random Walk and Random Way-point, and the fact that Lévy Walk behaves in between the former two movement models, it is now worth investigating whether a careful interpolation of Random Walk and Random Way-point can describe Lévy Walk accurately enough. Finding theoretical bounds for Lévy Walk can be a valuable future work that further expands our knowledge of the relation

between these three movement models and ultimately of human mobility.

ACKNOWLEDGEMENTS

A. Rezaei and J. Gao acknowledge support through NSF CCF-1535900, CNS-1618391, and DMS-1737812. J. Phillips acknowledges support by NSF CCF-1350888, ACI-1443046, CNS-1514520, and CNS-1564287. Research by C.D. Tóth was supported in part by NSF CCF-1422311 and CCF-1423615. The experiments were conducted with equipment purchased through NSF CISE Research Infrastructure Grant No. 1405641. The authors thank Dagstuhl Seminar 15111 on Computational Geometry during which some of the ideas were developed.

REFERENCES

- [1] Paolo Bajardi, Chiara Poletto, Jose J Ramasco, Michele Tizzoni, Vittoria Colizza, and Alessandro Vespignani. 2011. Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PLoS one* 6, 1 (2011), e16591.
- [2] Albert-lászló Barabási. 2005. The origin of bursts and heavy tails in human dynamics. *Nature* 435, 7039 (2005), 207–211.
- [3] Christian Bettstetter. 2001. Smooth is Better than Sharp: A Random Mobility Model for Simulation of Wireless Networks. *citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.23.3460* (2001).
- [4] B Birand, M Zafer, G Zussman, and K W Lee. 2011. Dynamic Graph Properties of Mobile Networks under Levy Walk Mobility. In *2011 IEEE Eighth International Conference on Mobile Ad-Hoc and Sensor Systems*. 292–301.
- [5] Andrew M. Bishop. 2017. Router for OpenStreetMap Data. <http://www.routino.org/>. (2017).
- [6] J Y Le Boudec and M Vojnovic. 2006. The Random Trip Model: Stability, Stationary Regime, and Perfect Simulation. *IEEE/ACM Trans. Netw.* 14, 6 (Dec. 2006), 1153–1166.
- [7] Josh Broch, David A Maltz, David B Johnson, Yih-Chun Hu, and Jorjeta Jetcheva. 1998. A Performance Comparison of Multi-hop Wireless Ad Hoc Network Routing Protocols. In *Proceedings of the 4th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom '98)*. ACM, New York, NY, USA, 85–97.
- [8] Tracy Camp, Jeff Boleng, and Vanessa Davies. 2002. A survey of mobility models for ad hoc network research. *Wireless communications and mobile computing* 2, 5 (2002), 483–502.
- [9] Capital Bikeshare. 2017. Washington DC Trip Histories. <https://www.capitalbikeshare.com/system-data>. (2017).
- [10] Maurice Chu, Horst Haussecker, and Feng Zhao. 2002. Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks. *The International Journal of High Performance Computing Applications* 16, 3 (2002), 293–313.
- [11] Tommy Chu and Ioanis Nikolaidis. 2002. On the Artifacts of Random Waypoint Simulations. In *International Conference on Internet Computing*.
- [12] Citi Bike. 2017. New York City Trip Histories. <https://www.citibikenyc.com/system-data>. (2017).
- [13] Andrea Clementi, Angelo Monti, and Riccardo Silvestri. 2011. Modelling mobility: A discrete revolution. *Ad Hoc Networks* 9, 6 (Aug. 2011), 998–1014.
- [14] Andrea Clementi, Angelo Monti, and Riccardo Silvestri. 2013. Fast flooding over Manhattan. *Distributed computing* 26, 1 (2013), 25–38.
- [15] Andrea Clementi, Riccardo Silvestri, and Luca Trevisan. 2015. Information spreading in dynamic graphs. *Distributed Computing* 28, 1 (2015), 55–73.
- [16] Pilu Crescenzi, Miriam Di Ianni, Andrea Marino, Gianluca Rossi, and Paola Vocca. 2009. Spatial Node Distribution of Manhattan Path Based Random Waypoint Mobility Models with Applications. In *Structural Information and Communication Complexity (Lecture Notes in Computer Science)*. Springer, Berlin, Heidelberg, 154–166.
- [17] Yves-Alexandre De Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. 2013. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports* 3 (2013), 1376.
- [18] Tassos Dimitriou, Sotiris Nikolettas, and Paul Spirakis. 2006. The infection time of graphs. *Discrete Applied Mathematics* 154, 18 (2006), 2577–2589.
- [19] Jiaxin Ding, Jie Gao, and Hui Xiong. 2015. Understanding and modelling information dissemination patterns in vehicle-to-vehicle networks. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 41.
- [20] Nathan Eagle, Alex (Sandy) Pentland, and David Lazer. 2009. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences* 106, 36 (2009), 15274–15278. <https://doi.org/10.1073/pnas.0900282106> arXiv:<http://www.pnas.org/content/106/36/15274.full.pdf+html>

- [21] David Easley and Jon Kleinberg. 2010. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press.
- [22] Richard P Feynman, Robert B Leighton, and Matthew Sands. 2013. *The Feynman Lectures on Physics, Desktop Edition Volume I*. Vol. 1. Basic books.
- [23] Enrique Frias-Martinez, Graham Williamson, and Vanessa Frias-Martinez. 2011. An agent-based model of epidemic spread using human mobility and social network information. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*. IEEE, 57–64.
- [24] Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause. 2010. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1019–1028.
- [25] Marta C González, César A Hidalgo, and Albert-László Barabási. 2008. Understanding individual human mobility patterns. *Nature* 453, 7196 (2008), 779–782.
- [26] Yih-Chun Hu and David B Johnson. 2000. Caching Strategies in On-demand Routing Protocols for Wireless Ad Hoc Networks. In *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MobiCom '00)*. ACM, New York, NY, USA, 231–242.
- [27] Hubway. 2017. Boston Trip Histories. <https://www.thehubway.com/system-data>. (2017).
- [28] Chalermek Intanagonwiwat, Ramesh Govindan, and Deborah Estrin. 2000. Directed diffusion: A scalable and robust communication paradigm for sensor networks. In *Proceedings of the 6th annual international conference on Mobile computing and networking*. ACM, 56–67.
- [29] David B Johnson and David A Maltz. 1996. Dynamic Source Routing in Ad Hoc Wireless Networks. In *Mobile Computing*. Springer, Boston, MA, 153–181.
- [30] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 137–146.
- [31] Harry Kesten and Vladas Sidoravicius. 2005. The spread of a rumor or infection in a moving population. *Annals of Probability* (2005), 2402–2462.
- [32] Jon Kleinberg. 2007. Computing: The wireless epidemic. *Nature* 449, 7160 (2007), 287–288.
- [33] Gregoynr Lawler. 2013. *Intersections of random walks*. Springer Science & Business Media.
- [34] K Lee, Y Kim, S Chong, I Rhee, and Y Yi. 2011. Delay-capacity tradeoffs for mobile networks with Lévy walks and Lévy flights. In *2011 Proceedings IEEE INFOCOM*. 3128–3136.
- [35] OpenStreetMap contributors. 2017. Planet dump retrieved from <https://planet.osm.org>. <https://www.openstreetmap.org>. (2017).
- [36] Wei Peng and Xi-Cheng Lu. 2000. On the reduction of broadcast redundancy in mobile ad hoc networks. In *Proceedings of the 1st ACM international symposium on Mobile ad hoc networking & computing*. IEEE Press, 129–130.
- [37] C E Perkins, E M Royer, S R Das, and M K Marina. 2001. Performance comparison of two on-demand routing protocols for ad hoc networks. *IEEE Pers. Commun.* 8, 1 (Feb. 2001), 16–28.
- [38] Alberto Pettarin, Andrea Pietracaprina, Geppino Pucci, and Eli Upfal. 2011. Tight bounds on information dissemination in sparse mobile networks. In *Proceedings of the 30th annual ACM SIGACT-SIGOPS symposium on Principles of distributed computing*. ACM, 355–362.
- [39] E M Royer, P M Melliar-Smith, and L E Moser. 2001. An analysis of the optimum node density for ad hoc mobile networks. In *ICC 2001. IEEE International Conference on Communications. Conference Record (Cat. No.01CH37240)*, Vol. 3. 857–861 vol.3.
- [40] K Shinki, M Nishida, and N Hayashibara. 2017. Message Dissemination Using Lévy Flight on Unit Disk Graphs. In *2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*. 355–362.
- [41] Michael F Shlesinger, Joseph Klafter, and Gert Zumofen. 1999. Above, below and beyond Brownian motion. *American Journal of Physics* 67, 12 (1999), 1253–1259.
- [42] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010. Limits of Predictability in Human Mobility. *Science* 327, 5968 (2010), 1018–1021. <https://doi.org/10.1126/science.1177170>
- [43] G M Viswanathan, F Bartumeus, Sergey V. Buldyrev, J Catalan, U L Fulco, Shlomo Havlin, M G E da Luz, M L Lyra, E P Raposo, and H Eugene Stanley. 2002. Lévy flight random searches in biological phenomena. *Physica A: Statistical Mechanics and its Applications* 314, 1 (Nov. 2002), 208–213.
- [44] S Wang, X Wang, X Cheng, J Huang, and R Bie. 2014. The Tempo-Spatial Information Dissemination Properties of Mobile Opportunistic Networks with Levy Mobility. In *2014 IEEE 34th International Conference on Distributed Computing Systems*. 124–133.
- [45] J Yoon, M Liu, and B Noble. 2003. Random waypoint considered harmful. In *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No.03CH37428)*, Vol. 2. 1312–1321 vol.2.
- [46] Dengpan Zhou and Jie Gao. 2009. Opportunistic Processing and Query of Motion Trajectories in Wireless Sensor Networks. In *Proceedings of the 28th Annual IEEE Conference on Computer Communications (INFOCOM'09)*. 1197–1205.