2018 17th ACM/IEEE International Conference on Information Processing in Sensor Networks

On the Ability of Mobile Sensor Networks to Diffuse Information

Chen Gu
Google Inc.
USA
guc@google.com

Ian Downes
Twitter Inc.
USA

Omprakash Gnawali
University of Houston
USA
gnawali@cs.uh.edu

Leonidas Guibas
Stanford University
USA
guibas@cs.stanford.edu

ABSTRACT

We examine the ability of networks formed by mobile sensor nodes to diffuse information in the case when communication is only possible during opportunistic encounters. Our setting assumes that mobile nodes are continuously sensing the world and acquiring new information. We form an abstract model of this situation and show by theoretical analysis, simulation, and real mobility data that the diffusion of information in this setting cannot be as efficient as when we allow arbitrary contact patterns between the nodes with the same overall contact statistics. This establishes a fundamental asymptotic limitation on the information diffusion capacity of such opportunistic mobile sensor networks — the encounter patterns arising out of physical motions in a geometric space are not ideal for information diffusion.

CCS CONCEPTS

• **Theory of computation** → **Communication complexity**; **Computational geometry**;

KEYWORDS

mobile sensor network, information packet, capacity, combinatorial and geometric settings, line arrangement

ACM Reference Format:
Chen Gu, Ian Downes, Omprakash Gnawali, and Leonidas Guibas. 2018. On the Ability of Mobile Sensor Networks to Diffuse Information. In *Proceedings of The 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN '18).* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In this paper we study mobile sensor nodes, called agents, sharing information through opportunistic encounters with each other during their motion. Such nodes form an ad hoc mobile network in which communication is possible only when two agents are in sufficient proximity — no other global communication infrastructure is presumed available. Our goal is to understand what are the capabilities and limits of such networks to diffuse information or, more precisely, to quantify how much of the information sensed by each mobile node can ultimately be delivered to every other node. To make this fundamental question tractable, we make certain assumptions that make our model at the same time simple and realistic but also different from prior models used in estimating the capacity of mobile networks:

- Every mobile agent is a sensor and is continuously acquiring information about the world. Unlike gossip models where the information (secret) is known in advance to one or more agents [16], in our setting new information enters the system continuously, effectively guaranteeing that perfect diffusion is impossible in most cases.
- When two agents communicate, they may share any information observed directly by them, as well as information conveyed to them by other agents during earlier encounters. While such an assumption may be unrealistic in certain settings (e.g. continuous video acquisition), in many practical situations information can be encoded compactly enough for such complete exchanges to occur (e.g. for scalar measurements like temperature, etc.).
- Communication happens in discrete events between pairs of agents. Here we assume that only one pair of agents communicates at a time — this assumption is not fundamental, but is convenient for theoretical analysis. We will extend the model to broadcast communication in real trace analysis.

To make the problem more amenable to discrete analysis, we lump all the information collected by a mobile agent between two successive encounters with other agents into what we call an *information packet*. Agents always communicate in entire information packets and we use such packet counts to measure the ability of our mobile sensor networks to diffuse information.

Clearly it is the temporal pattern of encounter events between the mobile agents that gates the capability of the system to spread information around as well as the speed with which this happens. For our analysis we assume that, as the mobile agents move along their trajectories, a finite number of encounter events occur, and then the system reaches its final state and stops. Ideally, we would like every agent to know in the end all the information packets generated by all the other agents, but this is clearly infeasible — as information collected in the very recent past may have no opportunity to diffuse to far away agents when the system is near or reaches its final state. Nevertheless, it is this gold standard we would like to use to measure the capability of a mobile sensor network to diffuse information.

In our setting the physical motions of the mobile agents determine the encounter patterns. In general, higher mobile agent density and higher speeds will lead to more frequent encounter events and therefore the information diffuses more rapidly [5]. In [4], the authors present a nice scheme for fitting mobile ad hoc networks and disruption tolerant networks into a continuum, according to the density and speed criteria just mentioned. Their high-level classification, however, leaves open the question of how the actual patterns of encounters and communications affect the ability of the network to diffuse information, even if the first-order communication statistics are fixed (e.g. how many times each particular pair of nodes communicates). This is exactly the question we propose to initiate a study of in this paper: the connection and dependencies between (1) mobility patterns of the agents in a geometric sense, (2) encounter and communication events among the agents as enabled by the mobility, and (3) the capacity of the agent network to diffuse sensor information.

From a theoretical point of view it makes sense to first look at a simple and uniform communication setting: we assume that we have n mobile agents and that each of the $\binom{n}{2}$ distinct pairs of agents communicates exactly once during the course of the scenario we are interested in. Thus each agent has $n-1$ encounters and generates a total of n packets of sensor data. The ideal is that in the end each agent knows all n^2 information packets generated by all n agents, and therefore we hope for at most n^3 (information packet, agent) pair deliveries. Specifically, if S is the total number of successful deliveries, we look at S/n^3 as a measure of the *capacity* of the network to diffuse information. We examine a number of different scenario classes:

(1) **Combinatorial setting.** In these scenarios we allow the time ordering of the events to be an arbitrary permutation of $\binom{n}{2}$ pairwise communications, which means there is no geometric constraint (e.g. two agents may communicate using phone or internet, even if they are very far from each other). Since this is a huge space with much variability according to the specific pattern, we focus on an average case analysis, where we consider each of the $\binom{n}{2}!$ temporal event permutations as equally likely. We show that in the random combinatorial setting, the capacity asymptotically tends to 1 (as $n \to \infty$) and the variance is low — in other words, with high probability, all the information packets get delivered to all agents except for a vanishingly small fraction.

(2) **Geometric setting.** These are the true scenarios we are after analyzing. Here each mobile agent follows a path in the plane and encounters only occur when two agents are at the same point at the same time. We analyze a simple setting where each agent moves along an infinite straight line in the plane, and the motions of the agents are coordinated so as to guarantee that communications happen at all arrangement vertices. Again, we define an appropriate notion of a random arrangement and look at average case capacity, to factor out variability due to geometric reasons. We show that in the random geometric case, there is a hard asymptotic upper bound $\kappa < 1$ on the capacity. So no matter how large the network gets, some fraction of the information will not get through.

We also find that the separation between combinatorial setting and geometric line arrangement case is more generally true for arbitrary but "reasonable" geometric motions in the plane (e.g. for bounded degree algebraic motions, or piecewise linear motions with bounded number of waypoints). This implies the number of realizable communication patterns in such geometric settings is still a vanishingly small fraction of those possible in the combinatorial setting.

(3) **Realistic setting.** We examine GPS traces of real vehicles under a slightly relaxed communication model — we assume two vehicles can exchange information if they are within a fixed communication range [3]. We show that in the realistic setting the performance is close to the idealized geometric setting: for any fixed-size time window, the capacity is asymptotically bounded by some constant $\kappa < 1$, while κ increases for bigger time window.

2 RELATED WORK

- **Setting.** There is a vast body of prior work characterizing the limits of information delivery in wireless and mobile ad hoc networks [1, 8, 9, 13, 14]. These studies try to understand how fast and how efficiently information available at the beginning of time on some or all nodes can propagate to the rest of the network. We instead study the problem of information dissemination in a different setting: mobile ad hoc sensor networks. Given that the nodes are continuously acquiring new information through sensing, we try to understand how quickly and what fraction of the total information (in space and time) propagates to the rest of the network.
- **Routing algorithm.** We use a gossip protocol to study the information delivery limits in combinatorial networks. The gossip communication models, sometimes also called epidemic protocols, are widely used in social networks [16]. Some gossip protocols compute aggregates [12], while others exchange information without processing it [10]. We focus on the latter class of gossip protocols. Gossip algorithms can be used in static networks [6], intermittently connected networks [19], or mobile ad hoc networks [17]. These gossip protocols are one of many classes of protocols one could use to disseminate information: proactive and reactive protocols [11], data muling [7], and VANET routing [18]. We analyze a generalized version of such protocols in the context of mobile nodes that continuously generate new information to be shared with rest of the network.
- **Mobility model.** Most analysis of information delivery in mobile ad hoc networks uses some variation of the random waypoint model [2]. Random waypoint has properties that make it amenable to analysis, while it may not be the best model to use when trying to understand the limits of information delivery [20]. Our analysis in combinatorial networks allows us to understand upper bounds on information diffusion while our geometric analysis focuses on low-complexity geometric motions which more realistically model the motions of vehicles on roads, etc. Thus, we complement prior work by analyzing information delivery in new network and mobility models.
- **Information delivery bounds.** In gossip setting, nodes require $\Theta(n^2)$ message exchanges for network-wide information dissemination. With m mobile nodes, the average number of exchanges for convergence within ϵ of the true result drops to $\Theta(n^2 \log \epsilon^{-1}/m)$ [17]. These analyses describe the bounds on number of message exchanges, which can be a proxy for convergence time or information delivery latency. We ask a different question — given unlimited time for convergence or information delivery, what is the achievable bound on the fraction of nodes that will receive the information continuously generated by all the nodes in the mobile ad hoc network?

3 THEORETICAL UNDERPINNINGS

In this section, we introduce two communication models: combinatorial and geometric settings, and develop theoretical results for information diffusion in these networks.

3.1 Combinatorial setting

Suppose there are n nodes, each representing a mobile agent in the network. All nodes continuously acquire information and communicate everything they know when a communication event happens. In the combinatorial setting, we assume that each pair of nodes communicates exactly once. The ordering of these communications is random, so that each of the $\binom{n}{2}!$ possible orderings of the communications between the nodes is equally likely.

As described in Section 1, we lump all the information collected by a node between two successive encounters with other nodes into an information packet. Each node has $n-1$ encounters and generates a total of n packets of data. Note that the last information packet cannot be delivered to any other node because the data is collected after all encounters. Thus, more precisely, we only consider the first $n-1$ information packets for every node, and measure the diffusion of $n(n-1)$ information items in the network.

Furthermore, when two nodes i and j encounter, they share all information packets they have. Note that if the latest information packet node i generated just before this encounter can be delivered to some other node k, using subsequent inter-node communications, then the latest information packet node j generated just before this encounter can also be delivered to node k along the same path. Thus, we can further merge these two information packets into one item, denoted as *information packet* $\{i, j\}$ (unordered pair, see Figure 1). As a result, there are n nodes and $\binom{n}{2}$ information packets in total, and therefore in the end we can expect at most $n\binom{n}{2}$ information item deliveries. We then define the following notion to measure the ability of mobile sensor networks to diffuse information:

Definition 3.1. Given an ordering of all $\binom{n}{2}$ pairwise communications between n nodes, the information packet $\{i, j\}$ can reach node k if there exists a sequence of nodes $\{m_{-1} = i, m_0 = j, m_1, \ldots, m_{h-1}, m_h = k\}$, or $\{m_{-1} = j, m_0 = i, m_1, \ldots, m_{h-1}, m_h = k\}$ such that all successive pairs

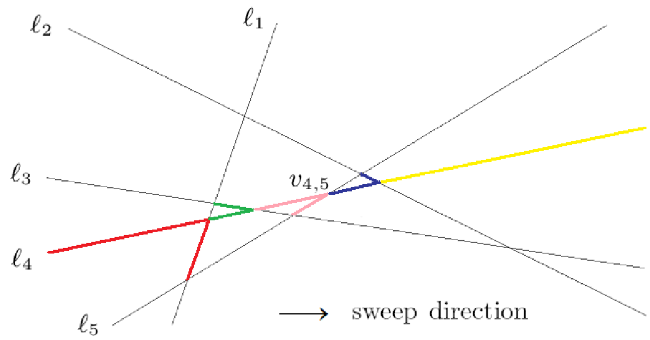

Figure 1: An arrangement of 5 lines. All information acquired is lumped into packets (pairs of segments with same color) at arrangement vertices. Information packet at $v_{4,5}$ can reach line ℓ_3 in 2 hops ($v_{4,5} \rightarrow v_{2,4} \rightarrow v_{2,3}$), but it cannot reach line ℓ_1.

$\{m_{-1}, m_0\}$, $\{m_0, m_1\}$, ..., $\{m_{h-1}, m_h\}$ appear as a subsequence in the ordering of all $\binom{n}{2}$ pairwise communications. If S is the total number of reachable $(\{i,j\},k)$ pairs $(1 \leq i < j \leq n, 1 \leq k \leq n)$, then the capacity is $\frac{S}{n\binom{n}{2}}$.

THEOREM 3.2. *In the combinatorial setting, the capacity is $1 - O(\log^2 n/n)$ with high probability.*

PROOF. We partition the sequence of all $\binom{n}{2}$ pairwise communications into groups of size $s = \lceil n \log n \rceil$. For any node i $(1 \leq i \leq n)$, the probability that i does not appear in one group is

$$\begin{aligned}
\frac{\binom{\binom{n-1}{2}}{s}}{\binom{\binom{n}{2}}{s}} &= \prod_{i=0}^{s-1} \frac{\binom{n-1}{2}-i}{\binom{n}{2}-i} \\
&< \prod_{i=0}^{s-1} \frac{\binom{n-1}{2}}{\binom{n}{2}} \\
&= \left(1 - \frac{2}{n}\right)^s \\
&\leq \left(1 - \frac{2}{n}\right)^{n \log n} \\
&< e^{-2\log n} \\
&= \frac{1}{n^2}
\end{aligned}$$

The probability that all n nodes appear in one group is at least

$$\begin{aligned}
& P\left(\bigcap_{i=1}^{n} \{\text{node } i \text{ appears in the group}\}\right) \\
=\; & 1 - P\left(\bigcup_{i=1}^{n} \{\text{node } i \text{ does not appear in the group}\}\right) \\
\geq\; & 1 - \sum_{i=1}^{n} P\left(\text{node } i \text{ does not appear in the group}\right) \\
>\; & 1 - \sum_{i=1}^{n} \frac{1}{n^2} \\
=\; & 1 - \frac{1}{n}
\end{aligned}$$

which means every node has been touched with high probability. Thus, each group can be considered as one round in the gossip algorithm, where every node communicates to some random partner [16].

In the original form of gossip algorithm, a secret can be diffused to all n nodes in $O(\log n)$ rounds with high probability. So, each information packet $\{i,j\}$ from communication pairs before the last $O(\log n)$ groups can be delivered to all nodes in the end with probability at least

$$\begin{aligned}
& P\left(\bigcap_{i=1}^{O(\log n)} \{i\text{ -th to last group is a gossip round}\}\right) \\
=\; & 1 - P\left(\bigcup_{i=1}^{O(\log n)} \{i\text{ -th to last group is not a gossip round}\}\right) \\
\geq\; & 1 - \sum_{i=1}^{O(\log n)} P\left(i\text{ -th to last group is not a gossip round}\right) \\
>\; & 1 - \sum_{i=1}^{O(\log n)} \frac{1}{n} \\
=\; & 1 - O\left(\frac{\log n}{n}\right)
\end{aligned}$$

that all the last $O(\log n)$ groups are gossip rounds. Therefore, the capacity is at least

$$\frac{n\left(\binom{n}{2} - sO(\log n)\right)}{n\binom{n}{2}} = 1 - O\left(\frac{\log^2 n}{n}\right)$$

with high probability. □

COROLLARY 3.3. *In the combinatorial setting, the variance of capacity is $O(\log^2 n/n)$.*

PROOF. Let X be the capacity in a random combinatorial network. Since $X \leq 1$, we have

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}(X^2) - (\mathrm{E}(X))^2 \\
&\leq 1 - \left(1 - O\left(\frac{\log^2 n}{n}\right)\right)^2 \\
&= O\left(\frac{\log^2 n}{n}\right)
\end{aligned}$$

□

We next look at a more refined measure of how the information packets get to their destinations: *minimum hop count*, which is defined as the minimum number of times the information packet has to be transferred from one node to another along any path to the destination (e.g. if the shortest sequence of pairwise communications from $\{i, j\}$ to k is $\{i, j\} \to \{j, p\} \to \{p, q\} \to \{q, k\}$, then the minimum hop count for $(\{i, j\}, k)$ is 3).

THEOREM 3.4. *In the combinatorial setting, 2/3 of the information packet deliveries need at most 1 hop.*

PROOF. Given an information packet $\{i, j\}$ and a node k, $\{i, j\}$ can reach k in 0 hops if and only if $k = i$ or $k = j$. So, the number of reachable $(\{i, j\}, k)$ pairs with 0 hops is $2\binom{n}{2}$. Suppose $\{i, j\}$ can reach k in 1 hop where i, j, k are all distinct, then either node i takes the information packet $\{i, j\}$ directly to the encounter at $\{i, k\}$, or node j takes the information packet $\{i, j\}$ directly to the encounter at $\{j, k\}$. So, we can ignore all other $n-3$ nodes and only consider the combinatorial network for nodes i, j and k. Note that for any ordering of pairwise communications $\{i, j\}$, $\{j, k\}$ and $\{k, i\}$, exactly 2 pairs from $(\{i, j\}, k)$, $(\{j, k\}, i)$ and $(\{k, i\}, j)$ are reachable: only the information packet from the last communication pair cannot reach the third node. Thus, the number of reachable $(\{i, j\}, k)$ pairs with 1 hop is $2\binom{n}{3}$. Therefore,

$$\frac{2\binom{n}{2} + 2\binom{n}{3}}{n\binom{n}{2}} > \frac{2}{3}$$

of the information packet deliveries need at most 1 hop. □

3.2 Geometric setting

In the geometric setting, since the ordering of all pairwise communications is constrained by the physical motions of the nodes, only certain communication patterns are possible. In this section, we consider a simple line arrangement model and analyze information diffusion.

We assume that each node moves along an infinite straight line in the plane, and all n lines $L = \{\ell_1, \ell_2, \ldots, \ell_n\}$ form an arrangement in general position. Two nodes only encounter each other if they are at the same point at the same time. To make the setting as conducive to diffusion as possible, we coordinate the motions of all nodes by sweeping the arrangement from left to right with a vertical line in the plane. During the sweeping procedure, all nodes move according to the intersections of their lines and the sweep line. These coordinated motions guarantee that encounters occur at all $\binom{n}{2}$ arrangement vertices, and thus each pair of nodes communicates exactly once.

By analogy with the combinatorial setting, we lump the information acquired by all nodes into information packets $\{i, j\}$ at arrangement vertices $v_{i,j} = \ell_i \cap \ell_j$. Geometrically, we say *a vertex $v_{i,j}$ can reach a line ℓ_k* if there exists an x-monotone path (directed from left to right) from $v_{i,j}$ to some vertex $v_{k,h}$ $(1 \leq h \leq n)$ on line ℓ_k. This means the information packet $\{i, j\}$ can be delivered to node k, using subsequent inter-node communications. The *minimum hop count* for a reachable $(v_{i,j}, \ell_k)$ pair is equal to the number of times the packet changes lines along the min-link path from $v_{i,j}$ to ℓ_k.

To define a random line, we take a point (a, b) randomly sampled from region $(a^-, a^+) \times (b^-, b^+)$, and then dualize this point to the line $y = ax + b$. A random arrangement of n lines is obtained by repeating this i.i.d. process n times. Here $a^- < a^+$ and $b^- < b^+$ are arbitrary real numbers. Note that since we sweep the arrangement from left to right, a and b can also be considered as the speed and position (at time $x = 0$) of the node, which are randomly sampled from their respective ranges.

THEOREM 3.5. *In the geometric setting, the average capacity is bounded by $\kappa \leq 5/6$.*

PROOF. For any $a^- < s < a^+$, let

$$R_s = \{(x, y) \mid x > 0,\ b^- + a^- x < y < b^- + sx\}$$

We first claim that a vertex v in region R_s cannot reach a line ℓ with a slope higher than s (see Figure 2). Otherwise, there must exist another line ℓ' below v with a slope higher than the slope of ℓ, which takes the information packet at vertex v to line ℓ. However, such a line ℓ' must have a y-intercept less than b^-, which is out of the range of b.

We next compute the probability $P(s)$ that v appears in R_s. Let

$$v = \left(\frac{b_1 - b_2}{a_2 - a_1}, \frac{a_2 b_1 - a_1 b_2}{a_2 - a_1}\right)$$

be the intersection of two random lines $y = a_1 x + b_1$ and $y = a_2 x + b_2$. Assuming $a_1 < a_2$, from $x > 0$ and $y < b^- + sx$, we have

$$b_1 > b_2 \text{ and } a_1 > \frac{a_2(b_1 - b^-) + s(b_2 - b_1)}{b_2 - b^-}$$

Since $a_2 > a_1 > \frac{a_2(b_1 - b^-) + s(b_2 - b_1)}{b_2 - b^-}$, we also have $a_2 < s$. So, all random variables need to satisfy:

$$\begin{cases} b^- < b_2 < b^+ \\ b_2 < b_1 < b^+ \\ a^- < a_2 < s \\ \max\left(a^-, \frac{a_2(b_1 - b^-) + s(b_2 - b_1)}{b_2 - b^-}\right) < a_1 < a_2 \end{cases}$$

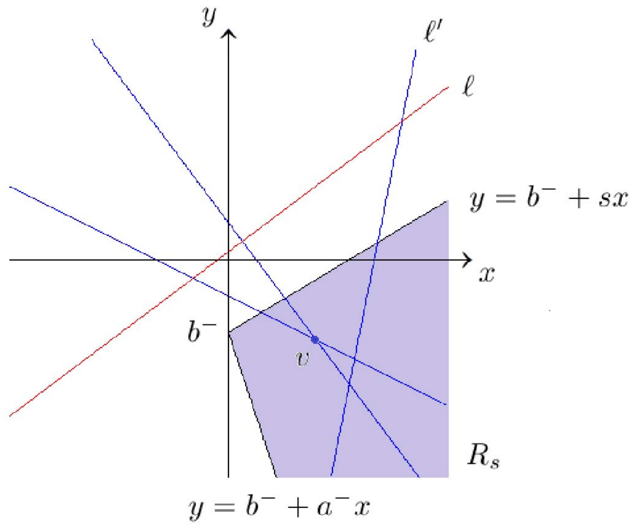

Figure 2: Vertex v in region R_s cannot reach line ℓ with a slope higher than s.

$$
\begin{aligned}
P(s) &= P(s \mid a_1 < a_2) \; + \; P(s \mid a_1 > a_2) \\
&= 2\; P(s \mid a_1 < a_2) \\
&= \tfrac{2}{(a^+ - a^-)^2 (b^+ - b^-)^2} \int_{b^-}^{b^+} \int_{b_2}^{b^+} \int_{a^-}^{s} \\
&\quad \left(a_2 - \max\left(a^-, \tfrac{a_2(b_1 - b^-) + s(b_2 - b_1)}{b_2 - b^-}\right)\right) \mathrm{d}a_2 \, \mathrm{d}b_1 \, \mathrm{d}b_2 \\
&= \tfrac{2}{(a^+ - a^-)^2 (b^+ - b^-)^2} \int_{b^-}^{b^+} \int_{b_2}^{b^+} \\
&\quad \Bigg(\int_{a^-}^{\frac{a^-(b_2 - b^-) + s(b_1 - b_2)}{b_1 - b^-}} \left(a_2 - a^-\right) \mathrm{d}a_2 \; + \\
&\quad \int_{\frac{a^-(b_2 - b^-) + s(b_1 - b_2)}{b_1 - b^-}}^{s} \left(a_2 - \tfrac{a_2(b_1 - b^-) + s(b_2 - b_1)}{b_2 - b^-}\right) \mathrm{d}a_2 \Bigg) \\
&\quad \mathrm{d}b_1 \, \mathrm{d}b_2 \\
&= \tfrac{2}{(a^+ - a^-)^2 (b^+ - b^-)^2} \int_{b^-}^{b^+} \int_{b_2}^{b^+} \tfrac{(s - a^-)^2 (b_1 - b_2)}{2(b_1 - b^-)} \, \mathrm{d}b_1 \, \mathrm{d}b_2 \\
&= \tfrac{(s - a^-)^2}{(a^+ - a^-)^2 (b^+ - b^-)^2} \int_{b^-}^{b^+} \big(b^+ - b_2 - \\
&\quad (b_2 - b^-)(\ln(b^+ - b^-) - \ln(b_2 - b^-))\big) \, \mathrm{d}b_2 \\
&= \tfrac{(s - a^-)^2}{4(a^+ - a^-)^2}
\end{aligned}
$$

Let $P'(s) = \mathrm{d}P(s)/\mathrm{d}s$ be the probability density at slope s. Since the probability that line ℓ has a slope higher than s is $\frac{a^+ - s}{a^+ - a^-}$, the probability for such a non-reachable (v, ℓ) pair is

$$\int_{a^-}^{a^+} P'(s) \; \frac{a^+ - s}{a^+ - a^-} \, \mathrm{d}s = \int_{a^-}^{a^+} \frac{(s - a^-)(a^+ - s)}{2(a^+ - a^-)^3} \, \mathrm{d}s = \frac{1}{12}$$

Symmetrically, a vertex in region

$$\left\{(x, y) \mid x > 0, \; b^+ + sx < y < b^+ + a^+ x\right\}$$

cannot reach a line ℓ with a slope lower than s. The probability for such a non-reachable (v, ℓ) pair is also $1/12$. Therefore,

$$\kappa \leq 1 - 2 \times 1/12 = 5/6$$

$\square$

Note that the asymptotic bound $\kappa < 1$ does not hold for every arrangement of lines. However, the probability for such set of constructions in the geometric setting is only vanishingly small.

Theorem 3.6. *There exists arrangements of lines with capacity $1 - O(1/n)$.*

Proof. Given an arbitrary arrangement of $n-2$ lines, we add at the very right (after all line intersections) a collector line with highest slope a^+, followed by a distributor line with lowest slope a^-. As a result, all intersections between the first $n-2$ lines can reach all n lines by following the collector, and then the distributor. Therefore, the capacity is at least

$$\frac{n\binom{n-2}{2}}{n\binom{n}{2}} \geq 1 - \frac{4}{n} = 1 - O\left(\frac{1}{n}\right)$$

Finally, it is easy to normalize the y-intercept of all lines to the sampling range by mapping $y = a_i x + b_i$ to

$$y = a_i x + \frac{b_i - \min b_i}{\max b_i - \min b_i}(b^+ - b^-) + b^-$$

$\square$

Corollary 3.7. *The fraction of combinatorial patterns that can be realized in the geometric setting is $O(\log^2 n/n)$.*

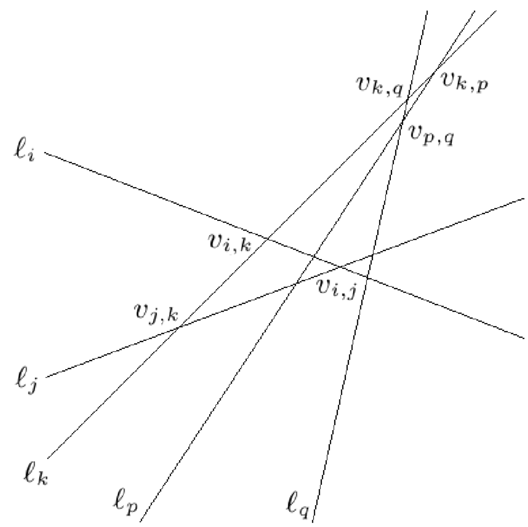

Figure 3: In the geometric setting, the min-link path takes at most 2 hops.

PROOF. Let X be the capacity in a random combinatorial network, and p be the fraction of combinatorial patterns realizable in the geometric setting. Since

$$\mathrm{E}(X) \leq \kappa \times p + 1 \times (1-p)$$

we have

$$p \leq \frac{1-\mathrm{E}(X)}{1-\kappa} = O\left(\frac{\log^2 n}{n}\right)$$

□

THEOREM 3.8. *In the line arrangement model, if vertex $v_{i,j}$ can reach line ℓ_k, then there exists a path from $v_{i,j}$ to ℓ_k with at most 2 hops.*

PROOF. Suppose on the contrary ℓ_k is reachable from $v_{i,j}$, and their min-link path takes at least 3 hops (see Figure 3). First, ℓ_k cannot intersect ℓ_i or ℓ_j to the right of $v_{i,j}$, otherwise it only takes 1 hop from $v_{i,j}$ to ℓ_k.

Assuming the slope of ℓ_k is positive, consider the min-link path from $v_{i,j}$ to ℓ_k: it must reach some vertex $v_{k,p}$ on ℓ_k from another line ℓ_p. The slope of ℓ_p must be higher than ℓ_k, otherwise we would reach ℓ_k before ℓ_p. Also, ℓ_p cannot intersect ℓ_i or ℓ_j to the right of $v_{i,j}$, otherwise it only takes 2 hops from $v_{i,j}$ to ℓ_k. Therefore, ℓ_p must intersect ℓ_i between $v_{i,k}$ and $v_{i,j}$, and also intersect ℓ_j between $v_{j,k}$ and $v_{i,j}$.

Similarly, to reach vertex $v_{k,p}$ from line ℓ_p, the path must first reach some vertex $v_{p,q}$ on line ℓ_p from another line ℓ_q. The slope of ℓ_q must be higher than ℓ_p, otherwise we would reach ℓ_p before ℓ_q. However, in this case, we can travel along ℓ_q directly to reach ℓ_k at vertex $v_{k,q}$ (without using ℓ_p), which gives a shorter path — a contradiction. □

4 EXPERIMENTAL VALIDATION

In this section, we validate results in theoretical settings, extend the model beyond linear motions, and examine the performance in realistic setting.

4.1 Algorithm

We first present an algorithm to compute the network capacity and minimum hop counts for reachable $(v_{i,j}, \ell_k)$ pairs. The input here is a sequence of pairwise communications, so the algorithm works for both combinatorial and geometric settings.

For network capacity, we need to find all reachable $(v_{i,j}, \ell_k)$ pairs. Let $S(v_{i,j}) = \{\ell_k \mid (v_{i,j}, \ell_k) \text{ is reachable}\}$. We first find the next encounter $v_{i,p}$ after $v_{i,j}$ for node i, and also the next encounter $v_{q,j}$ after $v_{i,j}$ for node j. Then, we can compute $S(v_{i,j})$ recursively by $S(v_{i,j}) = \{\ell_i, \ell_j\} \cup S(v_{i,p}) \cup S(v_{q,j})$. By using dynamic programming, we can find all reachable $(v_{i,j}, \ell_k)$ pairs in $O(n^3)$ time.

For minimum hop counts, we need to record additional information on the path directions. Given a vertex $v_{i,j}$ and a path ℓ_k, we define $f_k^{(1)}(i,j)$ as the minimum hop count from $v_{i,j}$ to ℓ_k with path direction along ℓ_i, and $f_k^{(2)}(i,j)$ as the minimum hop count from $v_{i,j}$ to ℓ_k with path direction along ℓ_j. Then, we can compute $f_k^{(1)}(i,j)$ and $f_k^{(2)}(i,j)$ recursively as follows:

$$f_k^{(1)}(i,j) = \begin{cases} 0 & \text{, if } i = k \\ 1 & \text{, if } j = k \\ f_k^{(2)}(j,i) & \text{, if } i > j \\ \min\left(f_k^{(1)}(i,p), 1 + f_k^{(2)}(q,j)\right) & \text{, otherwise} \end{cases}$$

$$f_k^{(2)}(i,j) = \begin{cases} 0 & \text{, if } j = k \\ 1 & \text{, if } i = k \\ f_k^{(1)}(j,i) & \text{, if } i > j \\ \min\left(f_k^{(2)}(q,j), 1 + f_k^{(1)}(i,p)\right) & \text{, otherwise} \end{cases}$$

Finally, the minimum hop count from $v_{i,j}$ to ℓ_k is $f_k(i,j) = \min\{f_k^{(1)}(i,j), f_k^{(2)}(i,j)\}$. Therefore, we can compute all values for $f_k(i,j)$ in $O(n^3)$ time.

4.2 Theoretical settings

We next verify theoretical results in Section 3. We generate random networks in different settings, and use the above algorithm to measure information diffusion. The experiment results are shown in Figure 4(a) and Table 1.

For network capacity, in the combinatorial setting, as we increase the network size n, the capacity converges to 1 (Theorem 3.2). In contrast, in the geometric setting, the capacity is bounded by $\kappa = 0.822 < 5/6$ (Theorem 3.5). Thus there is a clear gap between the two settings.

For minimum hop counts, in both settings, 2/3 of the information packet deliveries need only 1 hop (Theorem 3.4). In the geometric setting, the minimum hop count is at most 2 (Theorem 3.8). In the combinatorial setting, the min-link path can take 3 or more hops, while 2 hops are enough for most connectivity.

4.3 Beyond linear motions

The Definition 3.1 of network capacity can be easily generalized to the case when we allow each pair of nodes to encounter zero or multiple times. Given n node trajectories $\{\ell_1, \ell_2, \ldots, \ell_n\}$ with m pairwise intersections $\{v_1, v_2, \ldots, v_m\}$, we define the capacity as S/nm, where S is the total number of reachable (v, ℓ) pairs. In this setting, the capacity of combinatorial networks (sequence of random pairwise communications) in Theorem 3.2 becomes $1 - O(n\log^2 n/m)$. The notions in the algorithm in Section 4.1 can also be slightly modified to distinguish multiple encounters "$v_{i,j}$" for each pair of nodes.

Now we extend the line arrangement model to more general geometric motions. In Figure 4(b), we examine mobile nodes following algebraic motions, where each node moves according to a polynomial curve

$$y = \sum_{i=0}^{d} a_i x^i$$

with some bounded degree d and random coefficients $\{a_i\}$. The intersections between curves are computed using polynomial roots from companion matrix eigenvalues. Figure 4(c) considers an alternative geometric model based on piecewise linear motions, where each node trajectory is defined as an x-monotone path with k waypoints. More precisely, each path starts with a line with random slope and interception, and then the slope is randomly resampled after traveling along the line for a random length. This i.i.d. resampling process repeats $k-1$ times, forming a trace with k piecewise linear segments.

In both scenarios, we see that there exists some asymptotic upper bound $\kappa < 1$ on the capacity for every geometric network, while the constant κ increases as the degree d or number of waypoints k gets larger. Thus the gap we have established between combinatorial setting and geometric line arrangement case still holds true for these more general geometric motions. This also shows that the number of realizable communication patterns in such geometric settings is still a vanishingly small fraction of those possible in the combinatorial setting.

The proof of Theorem 3.5 can also be generalized to the piecewise linear model. In Figure 2, for any vertex v in region R_s, it cannot reach a path in which each of the k line segments has a slope higher than s. Therefore, the capacity

$$\kappa \leq 1 - 2\int_{a^-}^{a^+} P_k'(s) \left(\frac{a^+ - s}{a^+ - a^-}\right)^k \mathrm{d}s$$

where $P_k(s)$ is the probability that v appears in R_s with model complexity k. This explains why for any constant k we would expect a bounded capacity $\kappa < 1$, while as $k \to \infty$, $\left(\frac{a^+ - s}{a^+ - a^-}\right)^k \to 0$ and thus $\kappa \to 1$.

4.4 Realistic setting

In this section, we test information diffusion on a real mobility dataset from CRAWDAD [15]. It contains GPS coordinates of taxis collected in San Francisco Bay Area.

4.4.1 Broadcast communication. In real mobility traces, it is possible for multiple nodes to communicate simultaneously (e.g. group meeting) — in the geometric line arrangement model this corresponds to the degenerate case of concurrent lines. Consider nodes a, b, and c encounter at the same time, if we convert this event into a sequence of pairwise communications $\{ab, bc, ca\}$ (i.e. slightly perturb the lines), then both (ab, c) and (bc, a) are reachable, but (ca, b) is not. To make these three pairs equivalent, we can add another copy of ab in the end (so the new sequence becomes $\{ab, bc, ca, ab\}$), then (ca, b) is also reachable in 1 hop. In general, if there is a group of k nodes that communicates simultaneously at time t, we first create $\binom{k}{2}$ "real" vertices (send information packets before time t), and then add another copy of $\binom{k}{2}$

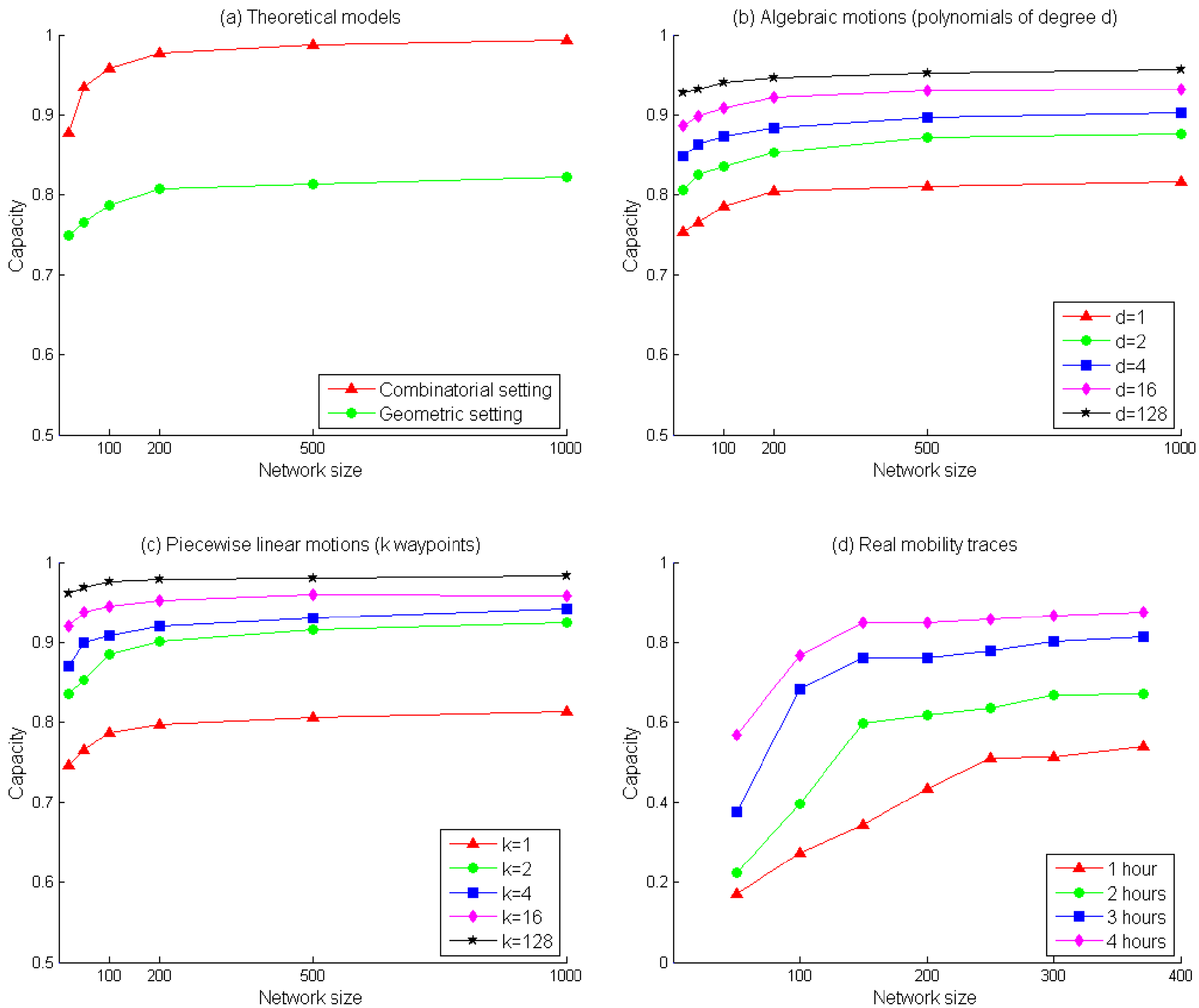

Figure 4: Network capacity.

"virtual" vertices (receive information packets after time t). Note that the delivery of all information packets outside this group is not affected by this construction. Therefore, we can convert broadcast communication model to pairwise communication model, and use the generalized algorithm in Section 4.3 to analyze information diffusion. Here we only count information packet deliveries from "real" vertices to node trajectories. The "virtual" vertices are only used for routing in the network.

4.4.2 Real world mobility. The dataset contains taxi traces collected in a 4-hour time window (8 am – 12 noon, Sunday morning) in San Francisco, California. We assume that two taxis can communicate if they are within 50 meters of each other. There are 371 taxis with 10428 communication events during that period.

In the realistic setting, we may compare the size of the time window to the degree of algebraic curves or number of waypoints in the geometric setting, which bounds how many times each pair of nodes can encounter during that period. So, for any fixed-size time window, we expect the capacity to be asymptotically bounded by some constant $\kappa < 1$. Figure 4(d) shows that, in each time window, the

Table 1: Distribution of minimum hop counts in combinatorial/geometric settings (1000 nodes) and realistic setting (371 taxis).

Setting \ Capacity \ Hop	0	1	2	3	4	≥ 5	Total
Combinatorial setting	0.002	0.665	0.306	0.015	0.003	0.002	0.993
Geometric setting	0.002	0.665	0.155				0.822
Algebraic motions ($d = 128$)	0.002	0.862	0.089	0.002	10^{-4}	10^{-5}	0.955
Piecewise linear motions ($k = 128$)	0.002	0.876	0.107	10^{-4}	10^{-5}	10^{-5}	0.985
Realistic setting (4 hours)	0.005	0.174	0.561	0.107	0.020	0.009	0.876

capacity roughly converges after there are more than 200 taxis in the network. We also see that the constant κ increases for bigger time window, as node trajectories become more complex and lead to more encounters in a longer time period (similar to curves of higher degree or polylines with more waypoints).

Finally, in Table 1 we see that 56.1% of the information packet deliveries require 2 hops. Since in this 4-hour period, there are only 9018 unique pairs of taxis encountered, rather than $\binom{371}{2} = 68625$ pairs, so we would expect longer delivery paths (in terms of hop count).

4.5 Optimistic capacity

As we have seen previously, random combinatorial contact patterns do not have realizations via reasonable geometric motions. In the realistic setting, assuming we allow some taxis to move under "unreasonable" geometric motions, what happens to the capacity in this optimistic scenario?

In the real world, it is a common phenomenon that a large fraction of encounters only happens in a small number of places (e.g. airport). So, we propose the following model to optimize diffusion via planned mobility. We partition the 2D map into grids of size L, and select the top N grids where encounters most likely occur. For each selected grid, we assume that there exists a *super taxi* which can instantly appear and exchange information during each real taxi-taxi encounter in that grid. But we do not create new encounter events between the super taxi and real taxis. The super taxi has a role analogous to the collector/distributor lines in Theorem 3.6, within its local grid area.

Table 2 shows that, with small grid size L, it is hard to increase the capacity as communications are still constrained by more reasonable physical motions. The capacity is higher with larger grids, because two taxis far apart can exchange information through a super taxi, as long as they communicate to some other taxis within the same local grid. When $L = 10^6$, there is only one global super taxi in the network, and the capacity reaches a maximum $\kappa = 0.949$. But of course it is impractical to imagine a physical mechanism to diffuse information instantly over such a large geographical area. This further confirms our gap result that separates combinatorial and geometric contact patterns.

5 CONCLUSION

In this paper we have initiated an investigation of the capability of mobile sensor nodes to diffuse information based on opportunistic encounters. Our main objective has been to understand how contact patterns with the same overall statistics differ in the ability to diffuse information. We have established a gap between arbitrary combinatorial patterns and those that arise out of physical motions in a geometric space.

In the theory section, we have proved fundamental bounds on the network capacity for random combinatorial patterns and geometric line arrangement setting. These are two ideal cases and in the experiment section, we have validated on more practical motion models and also built a chain that connects two theoretical models to explain how geometry affects information diffusion along this chain: i.e. geometric lines (Section 3.2, simplest model) $\rightarrow$ algebraic / piecewise linear motions (Section 4.3, more realistic theory models) $\rightarrow$ real 2D GPS traces (Section 4.4) $\rightarrow$ optimistic assisted diffusion (Section 4.5, adding "unreasonable" motions without geometric constraint) $\rightarrow$ random combinatorial patterns (Section 3.1, completely no geometric constraint).

In conclusion, we list a few issues that need further work:

- In the geometric settings, we showed $\kappa < 1$ for the straight lines and polylines models, where convex regions in the plane of positive area get some non-vanishing fraction of the arrangement vertices and a non-vanishing slope interval captures some fraction of the slopes of the random lines. We hope to have similar proofs for algebraic curves of higher degree, as well as motions in higher dimensions.
- During opportunistic encounters, we allow nodes to exchange all information they have. In the worst case, two nodes can exchange $\Theta(n^2)$ packets at a time (consider

Table 2: Optimistic capacity in realistic setting (371 taxis, 4 hours).

Grid size (L in meters)	200	200	200	200	200	500	500	500	500	500
Super taxis (N)	1	10	50	100	500	1	5	10	50	100
Capacity	0.876	0.877	0.880	0.881	0.881	0.885	0.890	0.892	0.894	0.895
Grid size (L in meters)	10^3	10^3	10^3	10^3	10^4	10^4	10^4	10^5	10^5	10^6
Super taxis (N)	1	5	10	50	1	5	10	1	2	1
Capacity	0.890	0.895	0.897	0.897	0.916	0.928	0.928	0.947	0.948	0.949

a line to intersect with other $n-1$ lines after all their $\binom{n-1}{2}$ intersections). However, in the combinatorial setting it is sufficient for each node to keep only the latest $O(n\log^2 n)$ packets, which is almost linear. On the other hand, limiting the number of packets exchange would only degrade performance in the geometric setting, and thus increase the gap. The problem with sublinear memory size is still open to investigation.

ACKNOWLEDGMENTS

The authors wish to acknowledge the support of NSF grant CCF-1514305 and a gift from the Google corporation.

REFERENCES

[1] H. Baumann, P. Crescenzi, and P. Fraigniaud. 2009. Parsimonious flooding in dynamic graphs. In *Symposium on Principles of Distributed Computing*.

[2] J. Broch, D. Maltz, D. Johnson, Y.-C. Hu, and J. Jetcheva. 1998. A performance comparison of multi-hop wireless ad hoc network routing protocols. In *International Conference on Mobile Computing and Networking*.

[3] N. Cenerario, T. Delot, and S. Ilarri. 2008. Dissemination of information in inter-vehicle ad hoc networks. In *Intelligent Vehicles Symposium*.

[4] Y. Chen, V. Borrel, M. Ammar, and E. Zegura. 2011. A framework for characterizing the wireless and mobile network continuum. *Computer Communication Review* (2011).

[5] A. Clementi, F. Pasquale, and R. Silvestri. 2009. Manets: high mobility can make up for low transmission power. In *International Colloquium on Automata, Languages and Programming*.

[6] G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall, and W. Vogels. 2007. Dynamo: amazon's highly available key-value store. In *Symposium on Operating Systems Principles*.

[7] M. Dunbabin, P. Corke, I. Vasilescu, and D. Rus. 2006. Data muling over underwater wireless sensor networks using an autonomous underwater vehicle. In *International Conference on Robotics and Automation*.

[8] C. Dutta, G. Pandurangan, R. Rajaraman, Z. Sun, and E. Viola. 2013. On the Complexity of Information Spreading in Dynamic Networks. In *Symposium on Discrete Algorithms*.

[9] M. Grossglauser and D. Tse. 2002. Mobility increases the capacity of ad-hoc wireless networks. *Transactions on Networking* (2002).

[10] Z. Haas, J. Halpern, and L. Li. 2006. Gossip-based ad hoc routing. *Transactions on Networking* (2006).

[11] C. Intanagonwiwat, R. Govindan, and D. Estrin. 2000. Directed diffusion: a scalable and robust communication paradigm for sensor networks. In *International Conference on Mobile Computing and Networking*.

[12] D. Kempe, A. Dobra, and J. Gehrke. 2003. Gossip-based computation of aggregate information. In *Foundations of Computer Science*.

[13] Y. Peres, A. Sinclair, P. Sousi, and A. Stauffer. 2011. Mobile geometric graphs: detection, coverage and percolation. In *Symposium on Discrete Algorithms*.

[14] A. Pettarin, A. Pietracaprina, G. Pucci, and E. Upfal. 2011. Tight bounds on information dissemination in sparse mobile networks. In *Symposium on Principles of Distributed Computing*.

[15] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser. 2009. CRAWDAD dataset epfl/mobility (v. 2009-02-24). (2009).

[16] B. Pittel. 1987. On spreading a rumor. *Journal on Applied Mathematics* (1987).

[17] A. Sarwate and A. Dimakis. 2009. The impact of mobility on gossip algorithms. In *International Conference on Computer Communications*.

[18] A. Skordylis and N. Trigoni. 2011. Efficient data propagation in traffic-monitoring vehicular networks. *Transactions on Intelligent Transportation Systems* (2011).

[19] T. Spyropoulos, K. Psounis, and C. Raghavendra. 2008. Efficient routing in intermittently connected mobile networks: the single-copy case. *Transactions on Networking* (2008).

[20] J. Yoon, M. Liu, and B. Noble. 2003. Random waypoint considered harmful. In *International Conference on Computer Communications*.