Performance Evaluation 160 (2023) 102343

Contents lists available at ScienceDirect

Performance Evaluation

journal homepage: www.elsevier.com/locate/peva

An analysis of probabilistic forwarding of coded packets on random geometric graphs

B.R. Vinay Kumar [a,*], Navin Kashyap [b], D. Yogeshwaran [c]

[a] INRIA, Sophia Antipolis, France
[b] Indian Institute of Science, Bengaluru, India
[c] Indian Statistical Institute, Bengaluru, India

ARTICLE INFO

Article history:
Received 19 July 2022
Received in revised form 12 February 2023
Accepted 15 February 2023
Available online 24 February 2023

Keywords:
Probabilistic broadcast
Percolation
Random geometric graph
Ad-hoc networks

ABSTRACT

We consider the problem of energy-efficient broadcasting on large ad-hoc networks. Ad-hoc networks are generally modelled using random geometric graphs (RGGs). Here, nodes are deployed uniformly in a square area around the origin, and any two nodes which are within Euclidean distance of 1 are assumed to be able to receive each other's broadcast. A source node at the origin encodes k data packets of information into n $(> k)$ coded packets and transmits them to all its one-hop neighbours. The encoding is such that, any node that receives at least k out of the n coded packets can retrieve the original k data packets. Every other node in the network follows a probabilistic forwarding protocol; upon reception of a previously unreceived packet, the node forwards it with probability p and does nothing with probability $1-p$. We are interested in the minimum forwarding probability which ensures that a large fraction of nodes can decode the information from the source. We deem this a *near-broadcast*. The performance metric of interest is the expected total number of transmissions at this minimum forwarding probability, where the expectation is over both the forwarding protocol as well as the realization of the RGG. In comparison to probabilistic forwarding with no coding, our treatment of the problem indicates that, with a judicious choice of n, it is possible to reduce the expected total number of transmissions while ensuring a near-broadcast.

1. Introduction

Ad-hoc networks are distributed networks with no centralized infrastructure. Applications involving the Internet of Things (IoT), such as healthcare, smart factories and homes, intelligent transport etc., have led to wide-spread presence of dense ad-hoc networks. Individual nodes in these networks are typically low-cost and energy-constrained, having limited computational ability and knowledge of the network topology.

Wireless ad-hoc networks are often modelled using random network models. In particular, random geometric graphs (RGGs) have been used in the literature to model spatially distributed networks (see e.g. [1,2]). These are generated by scattering (a Poisson number of) nodes in a finite area uniformly at random and connecting nodes within a pre-specified distance. The random distribution of nodes captures the variability in the deployment of the nodes of an ad-hoc network. The distance threshold conforms to the maximum range at which a transmission from a node, with maximum power, is received reliably. A more formal description of our network setting is provided in Section 3.

* Corresponding author.
E-mail addresses: vinay-kumar.bindiganavile-ramadas@inria.fr (B.R. Vinay Kumar), nkashyap@iisc.ac.in (N. Kashyap), d.yogesh@isibang.ac.in (D. Yogeshwaran).

https://doi.org/10.1016/j.peva.2023.102343
0166-5316/

Exchange of network-critical information for network control and routing happens primarily through broadcast mechanisms in these networks. A considerable number of broadcast mechanisms have been proposed in the literature. Naive algorithms such as flooding, although being light-weight and easy to implement, give rise to unnecessary transmissions and hence are not energy efficient. Flooding is also known to result in the 'broadcast-storm' problem (see [3]).

Probabilistic forwarding as a broadcast mechanism (see e.g., [4–6]) has been proposed in the literature as an alternative to flooding. Here, each node, on receiving a packet for the first time, either forwards it to all its one-hop neighbours with probability p or takes no action with probability $1-p$. While this mechanism reduces the number of transmissions, reception of a packet by a network node is not guaranteed.

To improve the chances of a network node receiving a packet and to handle packet drops, we introduce coding along with probabilistic forwarding. Let us suppose that the source possesses k_s message packets which need to be broadcast. These k_s message packets are first encoded into n coded packets such that, for some $k \geq k_s$, the reception of any k out of the n coded packets by a node, suffices to retrieve the original k_s message packets. Examples of codes with this property are Maximum Distance Separable (MDS) codes ($k = k_s$), fountain codes ($k = k_s(1+\epsilon)$ for some $\epsilon > 0$) etc. which are used in practice.

The source transmits the n coded packets to its one-hop neighbours and every other node in the network employs the probabilistic forwarding mechanism described above. Subsequent receptions of the same packet by a network node are neglected.

In this paper, we analyse the performance of the above algorithm on RGGs. In particular, we wish to find the minimum retransmission probability p for which the expected fraction of nodes receiving at least k out of the n coded packets is close to 1, which we deem a "near-broadcast". Here, it is to be clarified that the expectation is over both the realization of the RGG and the probabilistic forwarding protocol. This probability yields the minimum value for the expected total number of transmissions across all the network nodes needed for a near-broadcast. The expected total number of transmissions is taken to be a measure of the energy expenditure in the network.

To the best of our knowledge, we are the first to propose an algorithm that combines coding with a probabilistic forwarding based broadcast mechanism. (A survey of the related literature is provided in Section 2.) Specifically, the novelty in our proposed algorithm is to introduce redundancy in the form of coded packets into the probabilistic forwarding mechanism. The randomness brought about by the probabilistic forwarding algorithm can be compensated by the structural properties of the code we employ. This results in a simple, light-weight broadcast algorithm suitable for distributed implementation on ad-hoc networks.

In our previous work [7], we analysed the probabilistic forwarding mechanism described here on deterministic graphs such as trees and grids. It was found that introducing coded packets with probabilistic forwarding offered significant energy benefits in terms of the number of transmissions needed for a near-broadcast on well-connected graphs such as grids and other lattice structures. However, for d-regular trees, such energy savings were not observed. RGGs (in the super-critical regime) show similar behaviour as grids, i.e., for an intelligently chosen value of the number of coded packets, n, and the minimum forwarding probability, the energy expenditure in the network is considerably lesser for a near-broadcast, when compared to the scenario of probabilistic forwarding with no coding.

In this paper, we justify these observations using rigorous methods. Specifically, this work aims to build a mathematical framework to analyse the effect of introducing coding along with probabilistic forwarding for broadcasting on RGGs, which form an important class of models for ad-hoc networks. While some of the techniques used in our analysis are similar to those used for broadcasting on the grid in [7], we stress here that the additional complications due to the randomness of the underlying graph calls for the use of more sophisticated techniques. Ideas from continuum percolation, ergodic theory and Palm theory are employed to circumvent some of the technicalities encountered. These mathematical techniques could be of independent interest for related problems. Moreover, our method of analysis may also extend to more general broadcasting models and other point processes. Thus, we believe that our analysis of the proposed algorithm is in fact one of the more useful contributions of this paper.

The rest of the paper is organized as follows. Section 2 provides a literature overview of broadcast mechanisms for ad-hoc networks. In Section 3, we describe our network setup and formulate our problem. Section 4 provides the simulation results of the probabilistic forwarding algorithm on RGGs. In Section 5, we provide definitions and notations of RGGs on $\mathbb{R}^2$. Marked point processes (MPPs) are introduced to model probabilistic forwarding on the RGG. Section 6 relates probabilistic forwarding and marked point processes. Ergodic theorems on MPPs are used to obtain some key quantities. These will serve as the main ingredients in the analysis of the proposed algorithm on RGGs. Our main results are presented in Section 7, culminating in estimates for the minimum retransmission probability (see (20)) and the associated expected total number of transmissions (see (18)). Since the estimate for the minimum retransmission probability is difficult to compute explicitly, in Section 8, we provide a heuristic approach which is used to compare the expressions obtained theoretically with the simulation results. Section 9 discusses some aspects related to the assumptions and our results. In Section 10, some questions arising from this work are highlighted as possible future directions of research. The appendix contains technical results pertaining to the Palm expectations and the proof of one of our main theorems.

2. Related work

Algorithms for broadcast over ad-hoc networks have garnered considerable attention in the past. We refer the reader to [8–10] and the references therein for a review of the broad categories of algorithms employed for broadcasting. We further supplement this list with references relevant to our work here. The broadcast algorithm proposed and analysed in this paper is an amalgamation of probabilistic forwarding along with encoding of packets at the source. In the following, we highlight relevant literature from these two areas.

2.1. Coding based approaches

Network coding

Network coding has been used for efficient data dissemination in wireless networks in [9–14]. In [11], the authors propose random linear network coding (RLNC) for the multicast problem and give bounds on the probability that all the receivers are successful in obtaining the packets. The authors in [13] compare the number of transmissions in the RLNC based approach with that of store-and-forward approaches (which includes probabilistic forwarding) on a circular network topology. Network coding schemes are shown to be energy-efficient. Similar deductions are made via simulations in [14] for employing network coding in a medical sensor network. In [12], the authors provide transmission strategies for universal recovery and arrive at necessary and sufficient conditions on the number of transmissions required using network coding. However they assume complete knowledge of the network topology at every node.

Our work is closest in spirit to that in [9], where the authors propose a low-complexity distributed broadcast algorithm that improves upon the number of transmissions in flooding by a constant factor. Their approach based on network coding is well-suited for broadcasting on networks where individual nodes do not have any knowledge of the network topology, especially since, in their setting, all the nodes in the network have messages to broadcast. On the other hand, in our case, only a single source node has messages that need to be broadcast. This makes the two works incomparable.

Other coding schemes

Unlike network coding schemes, in our work, packets are encoded only at the source before transmission. The class of codes that we propose includes, among others, fountain codes, which have been used widely in broadcast mechanisms for ad-hoc networks. This is primarily because they form a convenient alternative to the ARQ (Automatic Repeat Request) protocol. In the ARQ scheme, an acknowledgement (ACK) needs to be sent every time a packet is received. By employing fountain codes, a node is required to send an ACK less frequently, thus saving on energy.

The authors in [15] employ fountain codes for broadcasting in vehicular networks. However, unlike our setting, all the nodes are in a star topology and receive transmissions from the source through erasure channels. In [16,17], the authors use Luby transform (LT) codes, a special case of fountain codes, which reduces the complexity of encoding and decoding at the network nodes. The LT encoding is done by randomly selecting d packets from n packets and doing an XOR of these packets to form a single encoded packet. The authors in [16] propose to employ LT codes in conjunction with transmission over a source-independent backbone network. They show via simulations that this approach not only reduces the number of transmissions required for flooding, but also reduces the packet delay. The variable d is an integer which is chosen according to a distribution. In [17], the authors propose a new distribution on d which further brings down the delay and the number of transmissions. However, both these approaches require the knowledge of a dominating set which is a subset of nodes of the network such that every node in the network is either in this set or adjacent to a node of this set. Finding a dominating set is computationally expensive. In [18], the authors construct novel codes called rateless online MDS (ROME) codes for wireless broadcasting. They are shown to have lesser coding redundancy and number of transmissions as compared to LT codes. However, they exploit feedback information from the receivers.

2.2. Probabilistic forwarding based approaches

Probabilistic forwarding mechanism, as described in Section 1, forms an energy-efficient alternative to the flooding mechanism. An excellent summary of the recent literature on probabilistic broadcast mechanisms is provided in [4, Chapter 3].

GOSSIP algorithm and variants

Probabilistic forwarding has also been referred to as the GOSSIP1(p) algorithm in [6]. The authors claim a 35% reduction in the transmission overhead as compared to flooding. Further, several variants of the probabilistic GOSSIP1(p) protocol are described, and heuristics and simulation results are provided for improving flooding and routing mechanisms in networks.

There have been numerous other works, for example see [19–23], which propose improvements on the GOSSIP protocol. In [20], the authors target a similar problem as ours: achieve a high degree of network coverage with limited number of transmissions. They even employ very similar analytical techniques based on continuum percolation to characterize two gossip algorithms: global gossip and distributed gossip. However, they assume some knowledge of the average degree of the random planar network at every node of the network. The authors in [24] propose a novel approach to combine tree-based and gossip protocols in order to achieve both low message complexity and high reliability. Hypergossiping has been proposed in [21] to overcome problems of connectivity in mobile ad-hoc networks. In [23], the authors propose the smart gossip protocol which aims to adaptively set the forwarding probability at each node by quantifying the "importance" of each node for achieving dissemination. However, all of these works evaluate the proposed algorithm using extensive simulations and lack sound analytical characterization.

Choice of forwarding probability

A significant portion of the literature on probabilistic forwarding dwells upon setting the forwarding probability based on different approaches, some of which are highlighted below:

- **Neighbour based approaches** [25–30]: In these schemes, the forwarding probability is decided based on the number of neighbours or the density of nodes in a region. The main rationale behind this approach is that, higher the density or the number of one-hop neighbours, lower the forwarding probability.
- **Area/distance based approaches** [31–34]: In area-based schemes, the forwarding probability is set based on an estimate of the additional area that will be covered by a node if it transmits. This additional area is estimated based on either the number of copies a node receives or the distance from the node whose transmission it receives.
- **Interference based schemes** [35,36]: In such schemes, nodes in the network choose a forwarding probability based on the signal strength with which they receive packets. Received signal strength is an indication of the channel quality, and hence nodes transmit with higher probability when the channel is good.

There are numerous other approaches which combine different methods to set the forwarding probability. The interested reader is referred to the survey paper [37] and Chapter 3 of [4].

While choosing the forwarding probability in a meaningful manner is also a motivation for our work, there are two main differences between the previously considered schemes and ours. Firstly, the previous schemes require some knowledge about the network topology, either in terms of the number of neighbours or distance from a nearest node etc., which we do not assume in our work. Secondly, and more importantly, most of these are simulation-based studies with no analytical backing. Our aim in this work is to provide a robust analytical framework to the algorithm we propose which can perhaps be extended to analyse some of these algorithms as well.

Other variants of probabilistic forwarding

The authors in [38] map randomized broadcast mechanisms to percolation on networks, which is the approach we take here as well. They, however, use directional antennas to reduce the transmission overhead and map it to a bond percolation problem. In [39], the authors propose Robust Probabilistic Flooding mechanism which takes into account the energy-harvesting nodes and the times they are active. The works in [40,41] consider broadcast problems on topologies similar to ours but a different mechanism. In [40], the authors model each edge of a tree as a binary symmetric channel and aim to recover the data present at the root of the tree using information from the nodes at level ℓ. Similar considerations are discussed on an infinite directed acyclic graph with the form of a $2D$ regular grid in [41].

3. Problem formulation

We begin by describing our setting for the specific case of random geometric graphs. This section introduces additional notation specific to RGGs as well.

3.1. Network setup

A random geometric graph is parametrized by the intensity λ and the distance threshold r. It suffices to study them by keeping one of the parameters fixed. In our treatment, we will fix the distance parameter r to be equal to 1, and study various properties as a function of the intensity, λ.

Construct a random geometric graph G_m with intensity λ and distance threshold $r = 1$ on $\Gamma_m := \left[\frac{-m}{2}, \frac{m}{2}\right]^2$ as follows:

- **Step 1:** Sample the number of points, N, from a Poisson distribution with mean $\lambda\nu(\Gamma_m)$. Here, $\nu(\cdot)$ is the Lebesgue measure on $\mathbb{R}^2$. Therefore, $N \sim \mathrm{Poi}(\lambda m^2)$.
- **Step 2:** Choose points $X_1, X_2, \ldots, X_N$ uniformly and independently from Γ_m. These form the points of a Poisson point process (see [42, Section 2.5]) Φ, and constitute the vertex set of G_m.
- **Step 3:** Place an edge between any two vertices which are within Euclidean distance $r = 1$ of each other.

To carry out probabilistic forwarding over G_m, we need to fix a source. For this, we will assume that there is a point at the origin $\mathbf{0} = (0, 0) \in \mathbb{R}^2$. More specifically, a graph $G_m^{\mathbf{0}}$ is created with the underlying point process $\Phi^{\mathbf{0}} \triangleq \Phi \cup \{\mathbf{0}\}$, as the vertex set and introducing additional edges from $\mathbf{0}$ to nodes which are within $B_1(\mathbf{0})$, to the edge set of G_m. Here, $B_1(\mathbf{0})$ (more generally, $B_1(\mathbf{v})$ for $\mathbf{v} \in \mathbb{R}^2$) is a closed Euclidean ball of radius 1 centred at $\mathbf{0}$ ($\mathbf{v}$).

The inclusion of an additional point at the origin $\mathbf{0}$ means that all the probabilistic computations need to be made with respect to the Palm probability given a point at the origin. We direct the reader to [43, Ch. 1.4] for an in-depth treatment of Palm theory. Heuristically, the Palm probability must be interpreted as the probability conditional on the event that the origin is a point of the point process. We denote the Palm probability by $\mathbb{P}^{\mathbf{0}}$ and the expectation with respect to it by $\mathbb{E}^{\mathbf{0}}$.

The origin here is a distinguished vertex. Broadcasts initiated from it can be received by the nodes which are present in the component of the origin only. Denote by $C_{\mathbf{0}} \equiv C_{\mathbf{0}}(G_m^{\mathbf{0}})$, the set of nodes in the component of the origin in $G_m^{\mathbf{0}}$. The component of the origin in $G_m^{\mathbf{0}}$ forms the underlying connected graph, which we denote by G.

3.2. Probabilistic forwarding on RGG

Equipped with the underlying network, G, we now describe the probabilistic forwarding algorithm on it. The source, $\mathbf{0}$, encodes k_s message packets into n coded packets and indexes them using integers from 1 to n. It then broadcasts each of these packets individually, which are received by all its one-hop neighbours. Every other node in the network, upon reception of a packet (say packet #j) uses the probabilistic forwarding mechanism: it broadcasts the packet with probability p and takes no action with probability $1-p$. A packet forwarded by a node (a single broadcast) is received by all its one-hop neighbours. Each packet is forwarded independently of other packets and other nodes. The node ignores all subsequent receptions of packet #j, irrespective of the decision it took at the time of first reception. Packet collisions and interference effects are neglected. See Section 9.2 for a discussion of this assumption.

We are interested in the following scenario. Let $R_{k,n}(G)$ be the number of nodes in $C_{\mathbf{0}}$ that receive at least k out of the n coded packets in G. We refer to these as *successful receivers*. We sometimes denote this by $R_{k,n}(G_m^{\mathbf{0}})$ to explicitly bring out the dependence on m. Given a $\delta > 0$, we are interested in the minimum forwarding probability p, such that the expected fraction of successful receivers is at least $1-\delta$. The expectation here is over the probabilistic forwarding protocol for a fixed realization of G. In reality, the proposed broadcasting algorithm of probabilistic forwarding with coded packets, should give a good performance for any realization of the underlying graph. In other words, we would want the expected fraction of successful receivers to be at least $1-\delta$, for every realization of G. However, in our formulation we relax this condition by asking for it only in an expected sense. More specifically, we define

$$p_{k,n,\delta} = \inf\left\{ p \;\middle|\; \mathbb{E}\left[\frac{R_{k,n}(G_m^{\mathbf{0}})}{|C_{\mathbf{0}}(G_m^{\mathbf{0}})|}\right] \geq 1-\delta \right\}, \tag{1}$$

where the expectation is over both the graph $G_m^{\mathbf{0}}$ as well as the probabilistic forwarding mechanism. Note that, from our construction, $R_{k,n}(G) = R_{k,n}(G_m^{\mathbf{0}}) \subseteq C_{\mathbf{0}}(G_m^{\mathbf{0}})$. The number of successful receivers is normalized by the total number of vertices in G, which is the same as the number of vertices within the component of the origin, $|C_{\mathbf{0}}(G_m^{\mathbf{0}})|$.

The performance measure of interest, denoted by $\tau_{k,n,\delta}$, is the expected total number of transmissions across all nodes when the forwarding probability is set to $p_{k,n,\delta}$. Here, it should be clarified that whenever a node forwards a packet to all its one-hop neighbours, it is counted as a single broadcast transmission. Our aim is to determine, for a given k and δ, how $\tau_{k,n,\delta}$ varies with n, and the value of n at which it is minimized (if it is indeed minimized). To this end, it is necessary to first understand the behaviour of $p_{k,n,\delta}$ as a function of n. In subsequent sections, we will formulate the probabilistic forwarding mechanism as a marked point process and use results from ergodic theory to obtain the expected value of the number of successful receivers and the overall number of transmissions.

4. Simulation results

In this section, we provide simulation results of our algorithm on random geometric graphs. For simulations on other network topologies, we refer the reader to [7,44].

Simulations of the probabilistic forwarding mechanism with coded packets were performed on an RGG generated with $m = 101$ and intensity $\lambda = 4.5$ and 4. As stated before, the distance threshold parameter r was set to 1. The probabilistic forwarding mechanism was carried out with $k = 20$ packets and n varying from 20 to 40. The value of δ was set to 0.1. Twenty realizations of G were generated and 10 iterations of the probabilistic forwarding mechanism was carried out on each of the realizations. The fraction of successful receivers was averaged over each iteration and realization of the graph. This was used to find the minimum forwarding probability, $p_{k,n,\delta}$, required for a near-broadcast, which is plotted in Fig. 1(a). The bars indicate bounds on $p_{k,n,\delta}$ obtained using a 95% confidence interval around the computed empirical average for the number of successful receivers. The $p_{k,n,\delta}$ values so obtained were further used to find the expected total number of transmissions over the same realizations. The expected total number of transmissions $\tau_{k,n,\delta}$, normalized by λm^2, which is the average number of points within Γ_m, is shown in Fig. 1(b) along with a 95% confidence interval around the empirical average. This can be interpreted as the average number of transmissions per node in the graph.

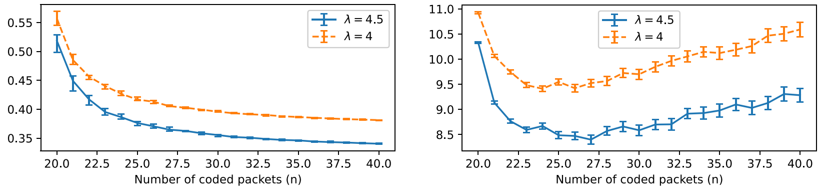
Notice that the expected number of transmissions decreases initially to a minimum and then increases. The decrease indicates the benefit of introducing coding along with probabilistic forwarding. The number of coded packets, n, and the probability, $p_{k,n,\delta}$, corresponding to the minimum point of Fig. 1(b) are the ideal parameters for operating the network to obtain maximum energy benefits.

Further, it can be observed from Fig. 1(a), that the minimum forwarding probability, $p_{k,n,\delta}$, decreases to 0 with n. This is formalized in the following lemma.

Lemma 4.1. *For fixed values of k and δ,*

(a) $p_{k,n,\delta}$ *is a non-increasing function of n.*
(b) $p_{k,n,\delta} \to 0$ *as $n \to \infty$.*

The proof is along similar lines as that for deterministic graphs expounded in [7]. However, unlike in deterministic graphs where the total number of nodes in the graph is a constant (N), here, the denominator in the expression for



(a) Minimum retransmission probability (b) Expected total no. of transmissions

Fig. 1. Simulations on a random geometric graph generated on Γ_{101} with intensity λ and distance threshold $r = 1$. Probabilistic forwarding done with $k = 20$ packets and $\delta = 0.1$. The error bars correspond to a 95% confidence interval computed around the empirical averages.

the fraction of successful receivers comprises of the nodes in the component of the origin (see (1)), which is a random quantity. Nevertheless, conditioning on the underlying point process, Φ, gives a deterministic graph, on which the result for deterministic graphs can be used.

Proof. (a) Denote a realization of the random geometric graph G by g. Let us define $\mathbb{E}_g$ to be the expectation over the probabilistic forwarding protocol when the underlying graph is g. Using the tower property of expectation, we obtain

$$\mathbb{E}\left[\frac{R_{k,n}}{N}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{R_{k,n}}{N} \,\middle|\, G\right]\right].$$

Conditioned on a realization g of G, N $(\equiv |C_{\mathbf{0}}(G_m^{\mathbf{0}})|)$ is fixed and it is true that $\mathbb{E}_g\left[\frac{R_{k,n}}{N}\right] \geq \mathbb{E}_g\left[\frac{R_{k,n-1}}{N}\right]$ due to a similar coupling argument as in [7, Lemma 1]. Therefore, we have that $p_{k,n,\delta}$ is a non-increasing function of n even when the underlying graph is random.

(b) For the second part, create $\lfloor \frac{n}{k} \rfloor$ non-overlapping (i.e., disjoint) groups of k packets each. For $i = 1, 2, \ldots, \lfloor \frac{n}{k} \rfloor$, let A_i be the event that the ith group of k coded packets is received by at least $(1 - \delta/2)N$ nodes. For a fixed realization g of the RGG, the randomness arises only because of the probabilistic forwarding mechanism. Since packets are forwarded independently of each other, and any two events A_i and A_j, for $i \neq j$, depend on disjoint sets of packets (for fixed g), they are independent conditioned on the RGG. In other words, conditional on $G = g$, the events A_i are independent and identically distributed (iid). Moreover, since we have a deterministic graph g, proceeding as in [7, Lemma 1], for all sufficiently large n we have that $\mathbb{E}_g\left[\frac{R_{k,n}}{N}\right] \geq 1 - \delta$ for any realization g of G. Therefore $\mathbb{E}\left[\frac{R_{k,n}}{N}\right]$ can be made arbitrarily close to 1 for sufficiently large n. This in turn means that $p_{k,n,\delta} \to 0$ as $n \to \infty$. □

5. Point process preliminaries

In this section, we introduce the tools required to characterize the performance of the probabilistic forwarding algorithm. The probabilistic forwarding mechanism on the RGG is modelled using marked point processes which are described here.

5.1. Random geometric graphs on $\mathbb{R}^2$

Our approach to analysing the probabilistic forwarding mechanism on G is to relate it to the probabilistic forwarding mechanism on a RGG generated on the whole $\mathbb{R}^2$ plane with the origin as the source. This means that the vertex set of the RGG is a Poisson point process, Φ, on $\mathbb{R}^2$. We refer the reader to [2] or [43] for the background needed on Poisson point processes. In particular, we use the procedure outlined in [43, Section 1.3] to construct the RGG on the whole $\mathbb{R}^2$ plane.

Create a tiling of the $\mathbb{R}^2$ plane with translations of Γ_m, i.e., $\Gamma_{i,j} := (im, jm) + \Gamma_m$ for $i, j \in \mathbb{Z}$. On each such translation, $\Gamma_{i,j}$, construct an independent copy of a Poisson point process with intensity λ as described in steps 1 and 2 of Section 3.1. The random geometric graph $(\mathcal{G})$ is constructed by connecting vertices which are within distance 1 of each other. We then say $\mathcal{G} \sim RGG(\lambda, 1)$.

It is known that the $RGG(\lambda, 1)$ model on $\mathbb{R}^2$ shows a phase transition phenomenon (see e.g. [45]). For $\lambda > \lambda_c$, the *critical intensity*, there exists a unique infinite cluster, $C \equiv C(\Phi)$, in the RGG almost surely. The value of λ_c is not exactly known, but simulation studies such as [46] indicate that $\lambda_c \approx 1.44$. The *percolation probability* $\theta(\lambda)$ is defined as the probability that the origin is present in the infinite cluster C, i.e., $\theta(\lambda) := \mathbb{P}^{\mathbf{0}}(\mathbf{0} \in C)$. We remark here that there is no known analytical expression for $\theta(\lambda)$ nor are there good approximations. Since we are interested in large networks, we will assume throughout our analysis that we operate in the super-critical region, i.e., $\lambda > \lambda_c$.

5.2. Marked point process

During the course of the probabilistic forwarding protocol on the RGG, each node decides independently whether to forward a particular packet with probability p. Marked point processes (MPPs) turn out to be a natural way to model such functions of an underlying point process.

Definition 5.1. Let $\Phi = \sum_i \varepsilon_{X_i}$[1] be a Poisson point process on $\mathbb{R}^2$. With each point X_i of Φ, associate a mark Z_i taking values in some measurable space $(\mathbb{K}, \mathcal{K})$ such that $\{Z_i\}_{i\in\mathbb{N}} \overset{iid}{\sim} \Pi(\cdot)$. Then, $\tilde{\Phi} = \sum_i \varepsilon_{(X_i,Z_i)}$ is called an *iid marked point process* on $\mathbb{R}^2 \times \mathbb{K}$ with *mark distribution* $\Pi(\cdot)$.

We now state an ergodic theorem for MPPs which is used to obtain some key results required in the analysis of the probabilistic forwarding protocol in Section 6.

5.3. Ergodic theorem

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space over which an iid marked point process $\tilde{\Phi} = \sum_i \varepsilon_{(X_i,Z_i)}$ is defined with mark distribution $\Pi(\cdot)$. Let $\theta_x : \Omega \to \Omega$, for $x \in \mathbb{R}^2$, be the operator which shifts each point of $\tilde{\Phi}$ by $-x$, i.e., $\theta_x\tilde{\Phi} = \sum_i \varepsilon_{(X_i-x,Z_i)}$ and let $(\mathbb{K}, \mathcal{K})$ be the measurable space of marks. Let $f : \mathbb{K} \times \Omega \to \mathbb{R}_+$ be a non-negative function of the MPP. Then, by the ergodic theorem for marked random measures (see [47, Theorem 8.4.4]), we have

$$\frac{1}{\nu(\Gamma_m)} \sum_{X_i\in\Gamma_m} f(Z_i, \theta_{X_i}(\omega)) \to \lambda \int_{\mathbb{K}} \mathbb{E}^{(\mathbf{0},z)} [f(z, \omega)]\, \Pi(dz) \qquad \mathbb{P}\text{-a.s.} \tag{2}$$

as $m \to \infty$, where $\mathbb{E}^{(\mathbf{0},z)}$ is the expectation with respect to the Palm probability $\mathbb{P}^{(\mathbf{0},z)}$ conditional on the mark, z. If $f(z, \omega) = f(\omega)$, then (2) reduces to

$$\frac{1}{\nu(\Gamma_m)} \sum_{X_i\in\Gamma_m} f(\theta_{X_i}(\omega)) \overset{m\to\infty}{\longrightarrow} \lambda\mathbb{E}^{\mathbf{0}} [f(\omega)] \qquad \mathbb{P}\text{-a.s.} \tag{3}$$

6. Probabilistic forwarding and MPPs

In this section, we formulate the probabilistic forwarding mechanism using the framework of marked point processes (MPPs). Ergodic theorems applied to MPPs are used to obtain the limiting values for the fraction of nodes in the infinite cluster (8) and in the infinite extended cluster (11). These help in obtaining estimates for $p_{k,n,\delta}$ and $\tau_{k,n,\delta}$ in Section 7. It should be noted here that all the graphs and point processes discussed in this section are defined on the whole $\mathbb{R}^2$ plane.

6.1. Single packet probabilistic forwarding

Consider the probabilistic forwarding of a single packet on $\mathcal{G} \sim RGG(\Phi, 1)$ defined on a Poisson point process (PPP) Φ of intensity λ on $\mathbb{R}^2$. Let $\mathcal{G}^{\mathbf{0}}$ be the graph created with the underlying point process being $\Phi^{\mathbf{0}} \triangleq \Phi \cup \{\mathbf{0}\}$ as the vertex set, and introducing additional edges from $\mathbf{0}$ to nodes which are within $B_1(\mathbf{0})$, to the edge set of $\mathcal{G}$. We assign a mark 1 to a node if it decides to transmit the packet and 0 otherwise. Thus, the mark space is $\mathbb{K} = \{0, 1\}$ and $\tilde{\Phi}$ is an iid MPP with a $Ber(p)$ mark distribution. Note that the origin, $\mathbf{0}$, has mark 1 since it always transmits the packet. Also, the subset of nodes which have mark 1 form a thinned point process of intensity λp, and the subset of vertices with mark 0 form a $\lambda(1-p)$–thinned process. Denote these by Φ^+ and Φ^- respectively, and the corresponding RGGs by $\mathcal{G}^+$ and $\mathcal{G}^-$. Notice that the set of vertices of Φ^+ which are in the same cluster as the origin are the vertices which receive the packet from the source and transmit it. Thus, the number of vertices in the cluster containing the origin in $\mathcal{G}^+$ (call this set of nodes $|C_{\mathbf{0}}^+|$), is the number of transmissions of the packet.

In addition to the nodes of the cluster containing the origin in $\mathcal{G}^+$, the nodes of $\mathcal{G}^-$ which are within distance 1 from them, also receive the packet. To account for them, we define for any cluster of nodes $S \subset \Phi^+$, the *boundary* of S as

$$\partial S = \{\mathbf{v} \in \Phi^- | B_1(\mathbf{v}) \cap S \neq \emptyset\},$$

and the *extended cluster* of S to be $S^{\text{ext}} = S \cup \partial S$. Then, the receivers are the nodes in $C_{\mathbf{0}}^{\text{ext}}$. We refer to this as the extended cluster of the origin.

Our interest is in large networks in which the origin is likely to be in the infinite cluster of $\mathcal{G}^{\mathbf{0}}$. Moreover, since we are interested in a large fraction of nodes in the network to be successful receivers, the extended cluster of the origin has to comprise of a significant number of nodes within Γ_m. In the limit of large m, this means that the extended cluster of the origin is the *infinite extended cluster* (IEC), C^{ext}, defined as the extended cluster of $C^+ := C(\Phi^+)$. This also means that the transmitters correspond to the nodes within Γ_m of the infinite cluster of Φ^+, C^+. Thus, in the thermodynamic limit,

[1] Here ε_x is the Dirac measure at x; for $A \subset \mathbb{R}^2$, $\varepsilon_x(A) = 1$ if $x \in A$ and $\varepsilon_x(A) = 0$ if $x \notin A$.

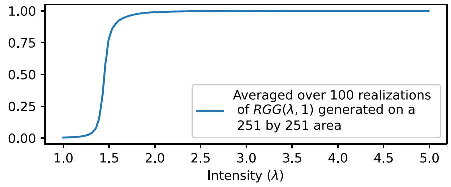


Fig. 2. Percolation probability $\theta(\lambda)$ vs. intensity λ.

the expected number of vertices in $C_{\mathbf{0}} \cap \Gamma_m$ (resp. $C_{\mathbf{0}}^{\text{ext}} \cap \Gamma_m$) is well-approximated by the expected number of vertices within Γ_m of the infinite cluster C^+ (resp., of the IEC C^{ext}) for large m. We use the ergodic theorem stated in Section 5.3 to obtain almost sure results for the fraction of nodes within Γ_m of the infinite cluster C^+ and the IEC C^{ext} in terms of the percolation probability $\theta(\lambda)$.

6.2. Application of the ergodic theorem

Specializing the statement in (2) to the probabilistic forwarding of a single packet where $\mathbb{K} = \{0, 1\}$ and the marks are independent, conditional on Φ, with distribution given by $\Pi(1) = 1 - \Pi(0) = p$, we obtain,

$$\frac{1}{\nu(\Gamma_m)} \sum_{X_i \in \Gamma_m} f(Z_i, \theta_{X_i}(\omega)) \xrightarrow{m\to\infty} \lambda p \mathbb{E}^{(\mathbf{0},1)}[f(1,\omega)] + \lambda(1-p)\, \mathbb{E}^{(\mathbf{0},0)}[f(0,\omega)] \qquad \mathbb{P}\text{-a.s.} \tag{4}$$

We will now use (3) and (4) to obtain key results which will be used to analyse the probabilistic forwarding of a single packet on $\mathbb{R}^2$. In particular, we substitute different functions f in (3) and (4) to obtain the following results:

- $f(z, \omega) = 1$. The ergodic theorem in (3) results in
$$\frac{\Phi(\Gamma_m)}{\nu(\Gamma_m)} \xrightarrow{m\to\infty} \lambda \qquad \mathbb{P}\text{-a.s.} \tag{5}$$
As a corollary, taking the reciprocals, we obtain
$$\frac{m^2}{\Phi(\Gamma_m)} \xrightarrow{m\to\infty} \frac{1}{\lambda} \qquad \mathbb{P}\text{-a.s.,} \tag{6}$$
which holds in our setting since $\lambda > \lambda_c$.
- $f(z, \omega) = z$. Substituting in (4), we see that the sum on the LHS counts the number of nodes which have mark 1 in Γ_m. Indeed, we obtain
$$\frac{\Phi^+(\Gamma_m)}{\nu(\Gamma_m)} \xrightarrow{m\to\infty} \lambda p \qquad \mathbb{P}\text{-a.s.} \tag{7}$$
- Let C be the unique infinite cluster in $\mathcal{G}$. Using the ergodic theorem in (3) with $f(z, \omega) = \mathbb{1}\{\mathbf{0} \in C\}$, we see that the sum on the LHS counts the number of vertices of Φ which are present in the infinite cluster. Then, we have that
$$\frac{|C \cap \Gamma_m|}{\nu(\Gamma_m)} \xrightarrow{m\to\infty} \lambda\, \theta(\lambda) \qquad \mathbb{P}\text{-a.s.} \tag{8}$$
Using the dominated convergence theorem (DCT) and (6), we also have that
$$\mathbb{E}\left[\frac{|C \cap \Gamma_m|}{\Phi(\Gamma_m)}\right] \xrightarrow{m\to\infty} \theta(\lambda). \tag{9}$$
This means that, for large m, the expected fraction of vertices of the infinite cluster within Γ_m is a good approximation for the percolation probability. We use this to obtain an empirical estimate of the percolation probability as follows. We generate 100 instantiations of the $RGG(\lambda, 1)$ model on Γ_{251}, for each value of λ between 1 and 5 (in steps of 0.01). The average number of vertices in the largest cluster within Γ_{251} is computed and taken as a proxy for the fraction of nodes of the infinite cluster. The graph obtained is shown in Fig. 2. We use the values from this plot in our numerical results. Similar plots are obtained in other works such as [48–50].
- Suppose $\lambda p > \lambda_c$, so that $\mathcal{G}^+$ operates in the super-critical region. Let C^+ be the unique infinite cluster in $\mathcal{G}^+$. Since Φ^+ is a thinned point process of intensity λp, we can use the result from (8) for the infinite cluster C^+ to obtain
$$\frac{|C^+ \cap \Gamma_m|}{\nu(\Gamma_m)} \xrightarrow{m\to\infty} \lambda p\, \theta(\lambda p) \qquad \mathbb{P}\text{-a.s.} \tag{10}$$

- Suppose that $\lambda p > \lambda_c$ and let C^{ext} denote the extended cluster of C^+, i.e. $C^{\text{ext}} = C^+ \cup \partial C^+$. Note that since C^+ is infinite, C^{ext} is also infinite. Hence, we refer to it as the *infinite extended cluster*, or IEC for short. Take $f(\omega) = \mathbb{1}(B_1(\mathbf{0}) \cap C(\Phi^+) \neq \emptyset)$. Observe that $\{X_i \in C^{\text{ext}}\} = \mathbb{1}(B_1(X_i) \cap C(\Phi^+) \neq \emptyset) = f(\theta_{X_i} w)$. So, using (3), we have that

$$\frac{1}{\nu(\Gamma_m)} \sum_{X_i \in \Gamma_m} \mathbb{1}\{X_i \in C^{\text{ext}}\} \xrightarrow{m\to\infty} \lambda \mathbb{P}(B_1(\mathbf{0}) \cap C(\Phi^+) \neq \emptyset) \qquad \mathbb{P}\text{-a.s.}$$

By definition, $\mathbb{P}(B_1(\mathbf{0}) \cap C(\Phi^+) \neq \emptyset) = \theta(\lambda p)$, the percolation probability of Φ^+. We then have,

$$\frac{|C^{\text{ext}} \cap \Gamma_m|}{\nu(\Gamma_m)} \xrightarrow{m\to\infty} \lambda\theta(\lambda p) \qquad \mathbb{P}\text{-a.s.} \tag{11}$$

Thus, it is natural to define, $\theta^{\text{ext}}(\lambda, p) := \mathbb{P}^{\mathbf{0}}(\mathbf{0} \in C^{\text{ext}}) = \theta(\lambda p)$.
Comparing RHS of (7) and (11) suggests an alternate viewpoint for the nodes that are present in the IEC. On the underlying point process Φ, define new iid marks $Z' \in \mathbb{K} = \{0, 1\}$ with $\text{Ber}(\theta^{\text{ext}}(\lambda, p))$ distribution. This means that a vertex is attributed mark 1, if it is in the IEC when probabilistic forwarding is carried out with forwarding probability p. Then, the fraction of nodes in the IEC when marks are Z corresponds to the fraction of nodes with mark 1 when marks are Z'. This interpretation will be useful in proposing a heuristic approach for probabilistic forwarding of multiple packets in Section 8.

6.3. Probabilistic forwarding of multiple packets

Consider now the probabilistic forwarding mechanism on n packets. Each node transmits a newly received packet with probability p independently of other packets. It is required to find the fraction of successful receivers, the nodes that receive at least k out of the n packets. From our discussion of probabilistic forwarding of a single packet (in Section 6.1), for large m, the number of nodes within Γ_m that receive a packet from the origin is well-approximated by the number of nodes in the IEC. In a similar way, the fraction of successful receivers within Γ_m can be well approximated by the fraction of nodes which are present in at least k out of the n IECs when probabilistic forwarding is done on the RGG, $\mathcal{G}^{\mathbf{0}}$. In this subsection, we will use the ergodic theorem and obtain explicit bounds on this fraction.

Equip each vertex of the point process Φ with mark $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n) \in \mathbb{K} = \{0, 1\}^n$. Here the jth co-ordinate of the mark represents transmission of the jth packet on Φ. More precisely, $Z_j(\cdot) \sim Ber(p)$ and, for two different vertices u and v, $\mathbf{Z}(X_u)$ and $\mathbf{Z}(X_v)$ are independent conditional on Φ. Therefore, it forms an iid marked point process. Define $C_{k,n}^{\text{ext}}$ to be the set of nodes which are present in at least k out of the n IECs. Taking $f(z, \omega) = \mathbf{1}\{\mathbf{0} \in C_{k,n}^{\text{ext}}\}$ in the statement of the ergodic theorem, we obtain

$$\frac{1}{\nu(\Gamma_m)} \sum_{X_i \in \Gamma_m} \mathbb{1}\{X_i \in C_{k,n}^{\text{ext}}\} \xrightarrow{m\to\infty} \lambda \; \mathbb{P}^{\mathbf{0}}(\mathbf{0} \in C_{k,n}^{\text{ext}}) \;\; \mathbb{P}\text{-a.s.}$$

Denote by $\theta_{k,n}^{\text{ext}}(\lambda, p) := \mathbb{P}^{\mathbf{0}}(\mathbf{0} \in C_{k,n}^{\text{ext}})$. Then the above statement reads as

$$\lim_{m\to\infty} \frac{|C_{k,n}^{\text{ext}} \cap \Gamma_m|}{\nu(\Gamma_m)} = \lambda \; \theta_{k,n}^{\text{ext}}(\lambda, p) \qquad \mathbb{P}\text{-a.s.} \tag{12}$$

7. Main results

In this section, we obtain expressions for $p_{k,n,\delta}$ and $\tau_{k,n,\delta}$ on the finite graph G based on the framework that has been developed in the previous section. Theorems 7.6 and 7.7 provide the limiting values of the expected fraction of transmitters and the expected fraction of successful receivers respectively, which are then used to obtain the estimates for $\tau_{k,n,\delta}$ (in (18)) and $p_{k,n,\delta}$ (in (20)). Prior to that, we first address some technical hurdles that arise while mapping the probabilistic forwarding mechanism on the finite graph to the MPP on $\mathbb{R}^2$.

While constructing $\mathcal{G}^{\mathbf{0}}$ (as described in Section 6.1), the graph corresponding to $\Gamma_{0,0}$ can be taken to be $G_m^{\mathbf{0}}$ (with additional edges from vertices in $\Gamma_{0,0}$ to those outside it). Alternately, $G_m^{\mathbf{0}}$ can be constructed by considering a restriction of $\mathcal{G} \sim RGG(\lambda, 1)$ to Γ_m and connecting the origin to nodes within $B_1(\mathbf{0})$. In essence, it is true that the distribution of nodes of $G_m^{\mathbf{0}}$ and $\mathcal{G}^{\mathbf{0}} \cap \Gamma_m$ is the same. Recall that the graph G on which the probabilistic forwarding mechanism is carried out, is the component of the origin in $G_m^{\mathbf{0}}$. In light of the correspondence between the vertices of $G_m^{\mathbf{0}}$ and $\mathcal{G}^{\mathbf{0}} \cap \Gamma_m$, the graph G should correspond to the graph induced on the nodes within Γ_m that are present in the cluster of the origin in $\mathcal{G}^{\mathbf{0}}$. However, these nodes also include those that are contained in the cluster of the origin through paths which go outside Γ_m but are not connected to the origin within Γ_m (see Fig. 3). We refer to these as, nodes in the cluster of the origin but *without a Γ_m-conduit* and denote them by $\widehat{C}_{\mathbf{0},m}$. The following theorem states that the number of nodes without Γ_m-conduits normalized by the area of Γ_m converges almost surely to 0.

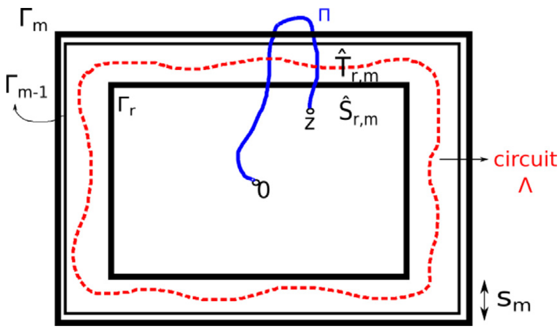


Fig. 3. Circuit in the annulus $\Gamma_{m-1} \setminus \Gamma_r$.

Theorem 7.1. *For $\lambda > \lambda_c$,*

$$\lim_{m\to\infty} \frac{|\widehat{C}_{\mathbf{0},m}|}{m^2} = 0 \qquad \mathbb{P}\text{-a.s.}$$

As a consequence, we have

$$\lim_{m\to\infty} \frac{|C_{\mathbf{0}}(G_m^{\mathbf{0}})|}{\lambda m^2} = \lim_{m\to\infty} \frac{|C_{\mathbf{0}}(\mathcal{G}^{\mathbf{0}}) \cap \Gamma_m|}{\lambda m^2} \qquad \mathbb{P}\text{-a.s.},$$

where $C_{\mathbf{0}}(\mathcal{G}^{\mathbf{0}})$ is the set of nodes in the cluster of the origin in $\mathcal{G}^{\mathbf{0}}$.

The latter part of the theorem is obtained by noting that $C_{\mathbf{0}}(\mathcal{G}^{\mathbf{0}}) \cap \Gamma_m = C_{\mathbf{0}}(G_m^{\mathbf{0}}) \cup \widehat{C}_{\mathbf{0},m}$ with $C_{\mathbf{0}}(G_m^{\mathbf{0}}) \cap \widehat{C}_{\mathbf{0},m} = \emptyset$. For the first part, we divide the nodes in $\widehat{C}_{\mathbf{0},m}$ into those which are present within a smaller concentric $r \times r$ area Γ_r, for $r < m$, and those in $\Gamma_m \setminus \Gamma_r$ (see Fig. 3). Denote these by

$$\widehat{S}_{r,m} = \widehat{C}_{\mathbf{0},m} \cap \Gamma_r \qquad \text{and} \qquad \widehat{T}_{r,m} = \widehat{C}_{\mathbf{0},m} \setminus \widehat{S}_{r,m}$$

respectively. In the following two lemmas, we show that for an appropriate value of r, the number of nodes in $\widehat{S}_{r,m}$ and $\widehat{T}_{r,m}$ normalized by m^2 converges to 0 almost surely.

Define $s_m = \frac{m-r}{2}$, the width of the annulus $\Gamma_m \setminus \Gamma_r$. Let us first look at the nodes in $\widehat{T}_{r,m}$. The following lemma states that the fraction of nodes of $\widehat{T}_{r,m}$ in a narrow annulus within Γ_m approaches 0 as $m \to \infty$.

Lemma 7.2. *For a sequence $s_m \to \infty$ with $\frac{s_m}{m} \to 0$ as $m \to \infty$, we have*

$$\lim_{m\to\infty} \frac{|\widehat{T}_{r,m}|}{m^2} = 0 \qquad \mathbb{P}\text{-a.s.}$$

Proof. The nodes in $\widehat{T}_{r,m}$ form a subset of the nodes of the underlying Poisson point process Φ which are within $\Gamma_m \setminus \Gamma_r$. Thus, we have,

$$|\widehat{T}_{r,m}| \le \Phi(\Gamma_m \setminus \Gamma_r) \qquad \mathbb{P}\text{-a.s.} \tag{13}$$

It suffices now to show that $\frac{\Phi(\Gamma_m \setminus \Gamma_r)}{m^2} \to 0$ as $m \to \infty$, which then proves the lemma. We proceed as follows:

$$\begin{aligned}\frac{\Phi(\Gamma_m \setminus \Gamma_r)}{m^2} &= \frac{\Phi(\Gamma_m \setminus \Gamma_r)}{m^2 - r^2} \cdot \frac{m^2 - r^2}{m^2} \\ &= \frac{\Phi(\Gamma_m \setminus \Gamma_r)}{m^2 - r^2} \cdot \left(\frac{4s_m}{m} - \frac{4s_m^2}{m^2}\right).\end{aligned} \tag{14}$$

Using the ergodic result in (5) with Γ_m replaced by $\Gamma_m \setminus \Gamma_r$, we obtain

$$\frac{\Phi(\Gamma_m \setminus \Gamma_r)}{m^2 - r^2} \to \lambda \qquad \mathbb{P}\text{-a.s.}$$

This is because the area of $\Gamma_m \setminus \Gamma_r$ is $m^2 - r^2$. Moreover, since the term within parenthesis in (14) converges to 0, from the condition in the statement of the lemma, we have that

$$\begin{aligned}\lim_{m\to\infty} \frac{|\widehat{T}_{r,m}|}{m^2} &\le \lim_{m\to\infty} \frac{\Phi(\Gamma_m \setminus \Gamma_r)}{m^2} \\ &= 0 \qquad \mathbb{P}\text{-a.s.} \quad \square\end{aligned}$$

We next address the nodes in $\widehat{S}_{r,m}$. These are nodes within Γ_r but without a Γ_m-conduit. We will show that $|\widehat{S}_{r,m}|$ converges to 0 almost surely using ideas from Russo–Seymour–Welsh (RSW) theory which is discussed in Appendix B. For this, let Ann_{s_m} denote the event of existence of a circuit in the annulus $\Gamma_{m-1} \setminus \Gamma_r$ as shown in Fig. 3. Notice that if $|\widehat{S}_{r,m}| > 0$, then there cannot be such a circuit. This is stated formally in the following lemma.

Lemma 7.3. *For $\lambda > \lambda_c$, let $\widehat{B}_{r,m}$ be the event that there exists at least one point of $\widehat{C}_{\mathbf{0},m}$ within Γ_r without a Γ_m-conduit i.e., $\widehat{B}_{r,m} = \{|\widehat{S}_{r,m}| > 0\}$. Then $\widehat{B}_{r,m} \subseteq Ann^c_{s_m}$.*

Proof. The proof proceeds by showing that the events $\widehat{B}_{r,m}$ and Ann_{s_m} cannot occur simultaneously. For this, suppose there is a circuit Λ within $\Gamma_{m-1} \setminus \Gamma_r$. Also, suppose that some point $z \in \Phi$ that lies within Γ_r is connected to the origin only via a path Π that leaves Γ_m. Then, Π must physically cross Λ at least twice as shown in Fig. 3. At any of the locations where such a crossing happens, consider the two adjacent points, x and y, of Φ that are on the path Π, but which fall on opposite sides of Λ. Note that, since Λ is at a distance of at least 1 from the boundary of Γ_m, both x and y are within Γ_m. Also consider the two adjacent points, u and v, of Φ that are on Λ, but which fall on opposite sides of Π. Now, x, u, y, v form a quadrilateral with diagonals xy and uv having length at most 1. Hence, at least one of the sides of this quadrilateral has length at most 1. This means that at least one of x and y is within distance 1 of either u or v (or both). Thus, at any crossing of Π and Λ, either Π and Λ intersect at some point of Φ, or Π is connected by an edge to the circuit Λ, and the connecting edge lies entirely within Γ_m. From this, one can construct a Γ_m-conduit between z and the origin. □

Corollary 7.4. *For $\lambda > \lambda_c$, there exists $s_m \ll m$ such that*

$$|\widehat{S}_{r,m}| \xrightarrow{m\to\infty} 0 \qquad \mathbb{P}\text{-a.s.}$$

Proof. Let $\epsilon > 0$. From the previous lemma and using Proposition B.3, we can write

$$\mathbb{P}(|\widehat{S}_{r,m}| > \epsilon) \le \mathbb{P}(\mathrm{Ann}^c_{s_m}) \le 8\left\lceil \frac{m}{s_m} \right\rceil \exp(-cs_m). \tag{15}$$

Taking $s_m = \frac{3\log m}{c}$ and summing over m, we obtain

$$\sum_m \mathbb{P}(|\widehat{S}_{r,m}| > \epsilon) \le \sum_m \frac{c'}{m^2 \log(m)} + \frac{c''}{m^3} < \infty. \tag{16}$$

Using the Borel–Cantelli lemma, this shows that $|\widehat{S}_{r,m}| \to 0$ as $m \to \infty$ almost surely. □

Proof of Theorem 7.1. The choice of $s_m = \frac{3\log m}{c}$ satisfies the condition of Lemma 7.2 as well. From Lemma 7.2 and Corollary 7.4, as m tends to infinity, we obtain

$$\frac{\widehat{C}_{\mathbf{0},m}}{m^2} = \frac{|\widehat{T}_{r,m}|}{m^2} + \frac{|\widehat{S}_{r,m}|}{m^2} \to 0 \qquad \mathbb{P}\text{-a.s.},$$

where $r = m - \frac{6\log m}{c}$. This proves the theorem. □

Continuing the discussion prior to Theorem 7.1, the fraction of nodes in the component of the origin that are not connected via Γ_m-conduits approaches 0 as $m \to \infty$ almost surely. The outcome of Theorem 7.1 is that in the asymptotic regime as $m \to \infty$, as long as we are interested in the fraction of nodes within the component of the origin, it does not matter whether these are connected to the origin via Γ_m-conduits or not. In other words, the fraction of nodes within G can be approximated by the fraction of nodes within Γ_m of the component of the origin in $\mathcal{G}^{\mathbf{0}}$ for a large m. To get a handle on the fraction of nodes within Γ_m of $C_{\mathbf{0}}(\mathcal{G}^{\mathbf{0}})$, we will need the following lemma.

Lemma 7.5. *Let $A = \{\mathbf{0} \in C(\mathcal{G}^{\mathbf{0}})\}$, where $C(\mathcal{G}^{\mathbf{0}})$ is the infinite cluster of $\mathcal{G}^{\mathbf{0}}$. For $\lambda > \lambda_c$, we then have*

$$\lim_{m\to\infty} \frac{|C_{\mathbf{0}}(\mathcal{G}^{\mathbf{0}}) \cap \Gamma_m|}{\lambda m^2} = \theta(\lambda)\mathbf{1}_A \qquad \mathbb{P}\text{-a.s.}$$

Proof. We can write

$$\frac{|C_{\mathbf{0}}(\mathcal{G}^{\mathbf{0}}) \cap \Gamma_m|}{\lambda m^2} = \frac{|C_{\mathbf{0}}(\mathcal{G}^{\mathbf{0}}) \cap \Gamma_m|}{\lambda m^2}\mathbf{1}_A + \frac{|C_{\mathbf{0}}(\mathcal{G}^{\mathbf{0}}) \cap \Gamma_m|}{\lambda m^2}\mathbf{1}_{A^c}.$$

Since A^c is the event that the origin is in some finite cluster, the number of nodes within $C_{\mathbf{0}}(\mathcal{G}^{\mathbf{0}})$ is finite. In the limit as $m \to \infty$, the latter term on the RHS above goes to 0. For the first term, notice that $A = \{C_{\mathbf{0}}(\mathcal{G}^{\mathbf{0}}) = C(\mathcal{G}^{\mathbf{0}})\}$. This gives

$$\frac{|C_{\mathbf{0}}(\mathcal{G}^{\mathbf{0}}) \cap \Gamma_m|}{\lambda m^2}\mathbf{1}_A = \frac{|C(\mathcal{G}^{\mathbf{0}}) \cap \Gamma_m|}{\lambda m^2}\mathbf{1}_A.$$

Further, from (A.1), we have that

$$\lim_{m\to\infty} \frac{|C(\mathcal{G}^{\mathbf{0}}) \cap \Gamma_m|}{\lambda m^2} = \lim_{m\to\infty} \frac{|C(\mathcal{G}) \cap \Gamma_m|}{\lambda m^2} \qquad \mathbb{P}\text{-a.s.}$$

Therefore, using (8) in the RHS of the above equation, we obtain that

$$\begin{aligned}\lim_{m\to\infty} \frac{|C_{\mathbf{0}}(\mathcal{G}^{\mathbf{0}}) \cap \Gamma_m|}{\lambda m^2}\mathbf{1}_A &= \lim_{m\to\infty} \frac{|C(\mathcal{G}^{\mathbf{0}}) \cap \Gamma_m|}{\lambda m^2}\mathbf{1}_A \\ &= \theta(\lambda)\mathbf{1}_A \qquad \mathbb{P}\text{-a.s.} \quad \square\end{aligned}$$

Note: It should be noted here that the statements in Theorem 7.1, Lemmas 7.3 and 7.5 and Corollary 7.4 hold $\mathbb{P}^{\mathbf{0}}$-a.s., since these are $\mathbb{P}$-a.s. statements made on the underlying graph $\mathcal{G}^{\mathbf{0}}$.

Before we proceed, we recall the definition of the minimum forwarding probability in (1):

$$p_{k,n,\delta} = \inf\left\{p \;\middle|\; \mathbb{E}\left[\frac{R_{k,n}(G_m^{\mathbf{0}})}{|C_{\mathbf{0}}(G_m^{\mathbf{0}})|}\right] \geq 1-\delta\right\},$$

where the expectation is over the graph as well as the probabilistic forwarding mechanism. Note that in our setting, the source, $\mathbf{0}$, always has mark 1 since it transmits all the n packets. To be more explicit, define $\mathbf{1} = (1, 1, \ldots, 1)$ to be the vector of all 1s of length n. We denote by $\mathbb{E}^{(\mathbf{0},\mathbf{1})}$ the expectation with respect to the Palm probability $\mathbb{P}^{\mathbf{0}}$ given a point at the origin, conditional on it having mark $\mathbf{Z}(\mathbf{0}) = \mathbf{1}$. In terms of this, the above equation translates to

$$p_{k,n,\delta} = \inf\left\{p \;\middle|\; \mathbb{E}^{(\mathbf{0},\mathbf{1})}\left[\frac{R_{k,n}(G_m)}{|C_{\mathbf{0}}(G_m)|}\right] \geq 1-\delta\right\}. \tag{17}$$

Next, since we are addressing a broadcast problem, it is necessary that a large fraction of nodes receive a packet. This, in turn necessitates that the fraction of nodes that transmit the packet is also large. With reference to the RGG on the whole plane, this means that the nodes in $\mathcal{G}^+$ need to have an infinite cluster. To allow for this, we make the following assumption.

Assumption 1. The forwarding probability p is such that $\lambda p > \lambda_c$.

Notice that the $p_{k,n,\delta}$ values obtained from simulations in Fig. 1 conform to this assumption. The assumption is discussed in slightly more detail in Section 9.1. We now obtain expressions for the minimum forwarding probability and the expected total number of transmissions based on these two assumptions.

7.1. Transmissions

Consider first the transmission of a single packet. Let $T(G_m)$ be the number of nodes of G_m that receive the packet from the source and transmit it and let $\mathcal{T}(\mathcal{G}) \cap \Gamma_m$ be the set of nodes within Γ_m that receive the packet from the source and transmit it when probabilistic forwarding is carried out on $\mathcal{G}$.[2] From our construction, it follows that $T(G_m)$ is stochastically dominated by $|\mathcal{T}(\mathcal{G}) \cap \Gamma_m|$ since there might be nodes which receive a packet from outside Γ_m and transmit it. However, it can be shown that,

$$\lim_{m\to\infty} \frac{\mathbb{E}^{(\mathbf{0},1)}\left[T(G_m)\right]}{m^2} = \lim_{m\to\infty} \frac{\mathbb{E}^{(\mathbf{0},1)}\left[|\mathcal{T}(\mathcal{G}) \cap \Gamma_m|\right]}{m^2}.$$

This is because the expected fraction of transmitting nodes with no Γ_m-conduits diminishes as $m \to \infty$. Thus, it suffices to evaluate $\lim_{m\to\infty} \frac{\mathbb{E}^{(\mathbf{0},1)}[|\mathcal{T}(\mathcal{G})\cap\Gamma_m|]}{m^2}$ to find the expected number of transmissions for a single packet.

In the jargon of marked point processes, $\mathcal{T}(\mathcal{G})$ is the set of vertices with mark $Z(\cdot) = 1$ that are in the cluster containing the origin. Note that the origin has mark 1, since it always transmits the packet. As the vertices with mark 1 form a thinned point process, Φ^+ of intensity λp, $\mathcal{T}(\mathcal{G})$ is the set of nodes in the cluster containing the origin in $\mathcal{G}^+$. In Section 6.1, we denoted this set by $C_{\mathbf{0}}^+$. From Assumption 1, the graph on Φ^+ is in the super-critical regime and thus possesses a unique infinite cluster, C^+. The following theorem provides the expected size of $C_{\mathbf{0}}^+ \cap \Gamma_m$. The proof proceeds by relating it to the expected size of $C^+ \cap \Gamma_m$ and using the ergodic result in (10).

Theorem 7.6. *For $\lambda p > \lambda_c$, we have*

$$\lim_{m\to\infty} \mathbb{E}^{(\mathbf{0},1)}\left[\frac{|C_{\mathbf{0}}^+ \cap \Gamma_m|}{\lambda m^2}\right] = p\,\theta(\lambda p)^2.$$

[2] It is implicit from the use of Palm probabilities that the origin is the source and probabilistic forwarding is formulated as an MPP as described in Section 6.1.

Proof. Denote by C^+, the unique infinite cluster of the thinned process Φ^+. Define the event $A^+ = \{\mathbf{0} \in C^+\} = \{B_1(\mathbf{0}) \cap C^+ \neq \emptyset\} \cap \{Z(\mathbf{0}) = 1\}$. Using Lemma 7.5 for the thinned process Φ^+ of intensity λp, we obtain

$$\frac{|C_{\mathbf{0}}^+ \cap \Gamma_m|}{\lambda p m^2} \stackrel{m\to\infty}{\longrightarrow} \theta(\lambda p)\mathbb{1}\{A^+\} \qquad \mathbb{P}\text{-a.s.}$$

From the note following Lemma 7.5 and using DCT, the expected values with respect to $\mathbb{P}^{\mathbf{0}}$ also converge giving,

$$\lim_{m\to\infty} \mathbb{E}^{\mathbf{0}}\left[\frac{|C_{\mathbf{0}}^+ \cap \Gamma_m|}{\lambda m^2}\right] = \theta(\lambda p)\mathbb{P}^{\mathbf{0}}(A^+) = p\,\theta(\lambda p)^2,$$

where the last equality uses the definition of A^+, $\mathbb{P}^{\mathbf{0}}(A^+) = \mathbb{P}^{\mathbf{0}}(B_1(\mathbf{0}) \cap C^+ \neq \emptyset)\,\mathbb{P}^{\mathbf{0}}(Z(\mathbf{0}) = 1) = p\theta(\lambda p)$, and we have also used that $\{B_1(\mathbf{0}) \cap C^+ \neq \emptyset\}$ and $\{Z(\mathbf{0}) = 1\}$ are independent events with respect to $\mathbb{P}^{\mathbf{0}}$. The proof is complete by noting that if $Z(\mathbf{0}) = 0$, then $C_{\mathbf{0}}^+ = \emptyset$ and so

$$\mathbb{E}^{\mathbf{0}}\left[\frac{|C_{\mathbf{0}}^+ \cap \Gamma_m|}{\lambda m^2}\right] = p\,\mathbb{E}^{(\mathbf{0},1)}\left[\frac{|C_{\mathbf{0}}^+ \cap \Gamma_m|}{\lambda m^2}\right]. \quad \square$$

Therefore, for large values of m, the expected number of transmissions, $\mathbb{E}^{\mathbf{0},1}[T(G_m)]$, can be approximated by

$$\mathbb{E}^{(\mathbf{0},1)}\left[|C_{\mathbf{0}}^+ \cap \Gamma_m|\right] \approx m^2\lambda p\;\theta(\lambda p)^2.$$

Consider now the transmission of multiple packets. The n coded packets are transmitted independently of each other. The expected total number of transmissions of all n packets would just be n times the expected transmissions of a single packet. Therefore, from Theorem 7.6, we then obtain

$$\tau_{k,n,\delta} \approx nm^2\lambda p_{k,n,\delta}\;\left(\theta(\lambda p_{k,n,\delta})\right)^2. \tag{18}$$

7.2. Minimum forwarding probability

In this section, we will obtain an expression for the minimum forwarding probability. Recall that this entails estimating $\mathbb{E}^{(\mathbf{0},1)}\left[\frac{R_{k,n}(G_m)}{|C_{\mathbf{0}}(G_m)|}\right]$, where $C_{\mathbf{0}}(G_m)$ is the set of nodes in the component of the origin in the underlying RGG on Γ_m and $R_{k,n}(G_m)$ are the number of nodes that receive at least k out of the n packets from the origin, which is the source. From Theorem 7.1, $C_{\mathbf{0}}(G_m)$ can be viewed as the set of nodes in the component of the origin in $\mathcal{G}^{\mathbf{0}}$ restricted to Γ_m but with only those nodes which are connected to the origin via Γ_m-conduits. $R_{k,n}(G_m)$ is the number of nodes among those in $C_{\mathbf{0}}(G_m)$, which are successful receivers. These arguments lets us think of the expectation $\mathbb{E}^{(\mathbf{0},1)}\left[\frac{R_{k,n}(G_m)}{|C_{\mathbf{0}}(G_m)|}\right]$, with respect to the RGG, $\mathcal{G}^{\mathbf{0}}$, instead of the finite RGG, $G_m^{\mathbf{0}}$.

Since we are interested in large networks, it is natural to assume that the origin is part of the infinite cluster of $\mathcal{G}^{\mathbf{0}}$. This means that the cluster of the origin in $G_m^{\mathbf{0}}$ connects to the infinite cluster in $\mathcal{G}^{\mathbf{0}}$ when $G_m^{\mathbf{0}}$ is embedded within it. In other words, the event $A = \{\mathbf{0} \in C(\mathcal{G}^{\mathbf{0}})\}$ occurs. The results of this section are made with this assumption, which is stated below explicitly. Additional justification for this is provided in Section 9.1.

Assumption 2. The origin is part of the infinite cluster of $\mathcal{G}^{\mathbf{0}}$.

From the discussion above and the assumption, our interest now is to estimate $\mathbb{E}_A^{(\mathbf{0},1)}\left[\frac{R_{k,n}(G_m)}{|C_{\mathbf{0}}(G_m)|}\right]$. The subscript A in the expectation $\mathbb{E}_A^{(\mathbf{0},1)}$ indicates conditional expectation given that the event A occurs. From Assumption 1, it is clear that such a conditioning can indeed be done, since $\mathbb{P}(A) = \theta(\lambda) > 0$.

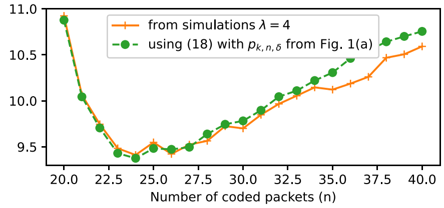
The following theorem gives the expected value of the fraction of successful receivers in the limit as $m \to \infty$ given the event A. Before we state the theorem, recall the formulation of probabilistic forwarding as a marked point process in Section 6. $C_{k,n}^{\text{ext}}$ was defined as the set of nodes which are present in at least k out of the n IECs and let $\theta_{k,n}^{\text{ext}} \equiv \theta_{k,n}^{\text{ext}}(\lambda, p) = \mathbb{P}^{\mathbf{0}}(\mathbf{0} \in C_{k,n}^{\text{ext}})$. Additionally, define $A_{[t]}^{\text{ext}}$ to be the event that the origin is present only in the IECs corresponding to the packets $1, 2, \ldots, t$.

Theorem 7.7. *For $\lambda p > \lambda_c$, we have*

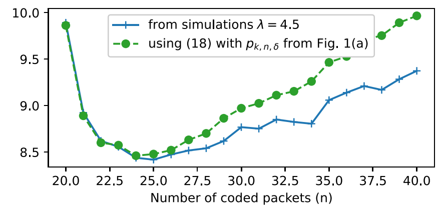
$$\lim_{m\to\infty} \mathbb{E}_A^{(\mathbf{0},1)}\left[\frac{R_{k,n}(G_m)}{|C_{\mathbf{0}}(G_m)|}\right] = \frac{1}{\theta(\lambda)^2}\sum_{t=k}^{n}\binom{n}{t}\theta_{k,t}^{ext}\;\mathbb{P}^{(\mathbf{0},1)}(A_{[t]}^{ext}).$$

The proof is on similar lines as that on the grid in [7]. It relies on carefully relating the fraction of successful receivers on G to the fraction of nodes present in at least k out of the n IECs corresponding to probabilistic forwarding on $\mathcal{G}^{\mathbf{0}}$. A step-by-step proof is given in Appendix C.

The following proposition is used to express $\mathbb{P}^{(\mathbf{0},1)}(A_{[t]}^{\text{ext}})$ in terms of $\theta_{k,n}^{\text{ext}}$.



(a) RGG(4, 1)



(b) RGG(4.5, 1)

Fig. 4. Comparison of the expected number of transmissions per node on Γ_{101} obtained using (18) with that obtained through simulations. Note that the $p_{k,n,\delta}$ value for each point on both the curves in each plot are from the simulations in Fig. 1(a).

Proposition 7.8.

$$\mathbb{P}^{(\mathbf{0},\mathbf{1})}\left(A_{[t]}^{ext}\right) = \begin{cases} \dfrac{\theta_{t,n}^{ext} - \theta_{t+1,n}^{ext}}{\binom{n}{t}} & 0 \le t \le n-1 \\ \theta_{n,n}^{ext} & t = n. \end{cases} \tag{19}$$

Proof. The second part follows directly from the definitions of $\theta_{n,n}^{\text{ext}}$ and the event $A_{[n]}^{\text{ext}}$. For the first part, define for $T \subseteq [n]$, A_T^{ext} to be the event that the origin is present in exactly the IECs indexed by T. Note that

$$\theta_{k,n}^{\text{ext}} = \mathbb{P}^{(\mathbf{0},\mathbf{1})}(\mathbf{0} \in C_{k,n}^{\text{ext}}) = \sum_{j=k}^{n} \sum_{\substack{T\subseteq[n] \\ |T|=j}} \mathbb{P}^{(\mathbf{0},\mathbf{1})}(A_T^{\text{ext}}).$$

Since the event A_T^{ext} depends only on the cardinality j (see Step 7 in Appendix C), we obtain

$$\theta_{k,n}^{\text{ext}} = \sum_{j=k}^{n} \binom{n}{j} \mathbb{P}^{(\mathbf{0},\mathbf{1})}(A_{[j]}^{\text{ext}}).$$

We then have that $\theta_{t,n}^{\text{ext}} - \theta_{t+1,n}^{\text{ext}} = \binom{n}{t}\mathbb{P}^{(\mathbf{0},\mathbf{1})}(A_{[t]}^{\text{ext}})$ for $0 \le t \le n-1$, which is the statement of the proposition. □

We remark here that the statement of Theorem 7.7 can be used to obtain an estimate for the expected fraction of successful receivers without the conditioning on the event A. We write

$$\mathbb{E}^{(\mathbf{0},\mathbf{1})}\left[\frac{R_{k,n}(G_m)}{|C_{\mathbf{0}}(G_m)|}\right] = \theta(\lambda)\, \mathbb{E}_A^{(\mathbf{0},\mathbf{1})}\left[\frac{R_{k,n}(G_m)}{|C_{\mathbf{0}}(G_m)|}\right] + (1-\theta(\lambda))\, \mathbb{E}_{A^C}^{(\mathbf{0},\mathbf{1})}\left[\frac{R_{k,n}(G_m)}{|C_{\mathbf{0}}(G_m)|}\right]$$

Notice from Fig. 2 that $\theta(\lambda)$ shows a phase transition phenomenon. For the intensities we are interested in, $\mathbb{P}(A^c) = 1-\theta(\lambda)$ is very small and the latter term in the above equation can be neglected. This also suggests that Assumption 2 is not a very strong requirement.

Consequently, for large m, using Theorem 7.7 and Proposition 7.8 in (17) yields an approximation for the minimum forwarding probability given by,

$$p_{k,n,\delta} \approx \inf\left\{p \middle| \sum_{t=k}^{n-1} \frac{\theta_{k,t}^{\text{ext}}(\theta_{t,n}^{\text{ext}} - \theta_{t+1,n}^{\text{ext}})}{\theta(\lambda)} + \frac{\theta_{k,n}^{\text{ext}}\theta_{n,n}^{\text{ext}}}{\theta(\lambda)} \ge 1-\delta\right\}. \tag{20}$$

7.3. Comparison with simulations

We have not been able to obtain exact expressions for the probability $\theta_{k,t}^{\text{ext}}(\lambda, p)$ in terms of the percolation probability $\theta(\lambda)$. However, in Section 9.4, we provide some bounds for it. We also develop an alternate heuristic approach, which provides comparable results for the minimum forwarding probability obtained through simulations, in Section 8.

Nevertheless, the approximation for the expected total number of transmissions, $\tau_{k,n,\delta}$ in (18) can be evaluated with the knowledge of the minimum forwarding probability. In Fig. 4, we show the plot of $\tau_{k,n,\delta}$ normalized by λm^2 with n in which we use $p_{k,n,\delta}$ values from Fig. 1(a)

It is observed that for $n \lesssim 26$, both the curves match pretty well. However, for $n > 26$ they diverge. This can be attributed to the fact that as n increases, $p_{k,n,\delta}$ decreases as in Fig. 1(a) and thus $\lambda p_{k,n,\delta} \searrow \lambda_c$. The estimate for the percolation probability, $\theta(\lambda)$, obtained via the ergodic result in (9) may not be accurate near the critical intensity, λ_c (which is itself not exactly known). In particular, Γ_{251} may not be large enough for the ergodic result in (9) to kick in, as we approach λ_c.

Nevertheless, this provides justification to our observation that the expected number of transmissions indeed decreases when we introduce coded packets along with probabilistic forwarding. This comes with a catch that the minimum forwarding probability for a near-broadcast behaves as in Fig. 1(a). In order to establish this, we provide a heuristic explanation in the next section.

8. Heuristic estimates for the minimum forwarding probability and the optimal number of coded packets

The minimum forwarding probability is expressed in terms of the probability $\theta_{k,n}$ in (20). This section provides heuristic estimates for the probability $\theta_{k,n}$, which, in turn, is used to obtain approximations for $p_{k,n,\delta}$. Additionally, a graphical procedure to obtain the optimal number of coded packets using these heuristic estimates is also outlined. For rigorous bounds on $\theta_{k,n}$, we refer the reader to Section 9.4.

In the marked point process formulation, probabilistic forwarding of multiple packets was modelled using marks given by $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ with $Z_i \sim Ber(p)$ conditional on the underlying point process Φ. We refer to this as the *original model*. Motivated by the alternate interpretation for the nodes in the IEC expounded at the end of Section 6.2, in this section we provide a heuristic approach for evaluating the minimum forwarding probability.

As before, let $\theta^{\text{ext}}(\lambda, p)$ denote the probability that the origin is in the IEC for a single packet transmission. Associate a new mark $\mathbf{Z}' = (Z'_1, Z'_2, \ldots, Z'_n) \in \mathbb{K} = \{0, 1\}^n$ to each vertex of Φ. The ith co-ordinate of $\mathbf{Z}'$ corresponds to probabilistic forwarding of the ith packet. The mark $\mathbf{Z}'$ is chosen such that each of the i co-ordinates is either 1 with probability $\theta^{\text{ext}}(\lambda, p)\,(= \theta(\lambda p))$ or 0 with the remaining probability, independent of the others. Similar to the viewpoint for the single packet transmission, our idea is to use Z'_i as a proxy for a vertex to be present in the IEC in probabilistic forwarding of the i−th packet. We refer to this as the *mean-field model.*

There are two key differences between the two models defined here. Firstly, in the original model, presence of a node in the IEC is not independent of other nodes being present in the IEC. Whereas, in the mean-field model, $Z'_i(\mathbf{u})$ and $Z'_i(\mathbf{v})$ are chosen to be independent $Ber(\theta(\lambda p))$ random variables for two distinct vertices $\mathbf{u}$ and $\mathbf{v}$. Since Z'_i is interpreted as an indicator whether a vertex is present in the ith IEC, this independence is enforced, conditional on Φ. Secondly, in the original model, presence of a particular node in IECs corresponding to two different packets, are not independent. They are independent conditional on Φ but not otherwise. In the mean-field model, since $\mathbf{Z}'_i(\mathbf{v})$ and $\mathbf{Z}'_j(\mathbf{v})$ are taken to be iid, this dependence is over-looked.

To analyse the mean-field model, let us use the ergodic theorem (2) with

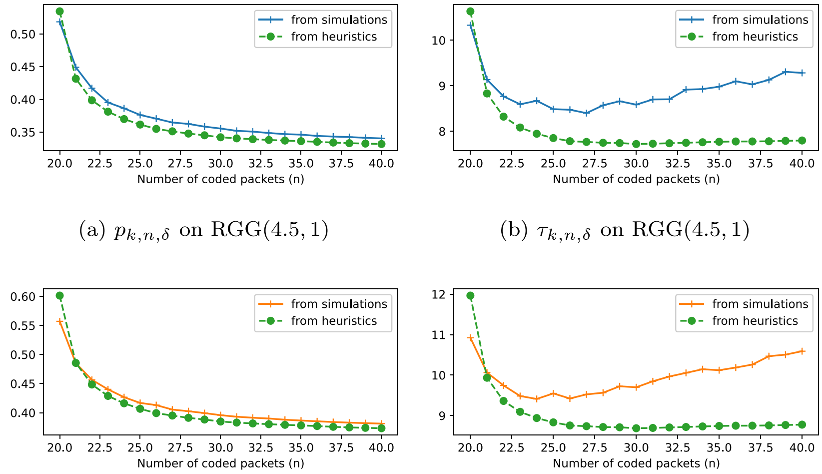
$$f(\mathbf{z}', \omega) = \sum_{j=k}^{n} \sum_{\substack{T \subseteq [n] \\ |T|=j}} \prod_{i \in T} z'_i \prod_{i \notin T} (1 - z'_i).$$

The inner summation is 1 only if a node has mark 1 in exactly the co-ordinates indexed by T (which has cardinality j). Since the outer sum goes over all $j \geq k$, the value of the function is 1 for a vertex which has mark 1, in at least k out of the n co-ordinates. From our interpretation of $\mathbf{Z}'$, the value of the function, f, for a vertex is equal to 1 if it is present in at least k out of the n IECs of the original model. Define $C'_{k,n}$ to be the set of nodes which have mark $Z'_i(\cdot) = 1$ in at least k out of the n packet transmissions in the mean-field model. Here, $C'_{k,n}$ acts as a proxy for $C^{\text{ext}}_{k,n}$. Since $f(\mathbf{Z}'(\mathbf{v}), \omega) = 1$ if $\mathbf{v} \in C'_{k,n}$, we can apply Theorem (2), to obtain for $\mathbb{P}$ almost surely

$$\begin{aligned}
&\frac{1}{\nu(\Gamma_m)} \sum_{X_i \in \Gamma_m} \mathbb{1}\{X_i \in C'_{k,n}\} \\
&\quad \overset{m\to\infty}{\longrightarrow} \lambda \sum_{\mathbf{z}' \in \{0,1\}^n} \mathbb{P}(\mathbf{Z}' = \mathbf{z}') \mathbb{E}^{(\mathbf{0},\mathbf{z}')} \left[\sum_{j=k}^{n} \sum_{\substack{T \subseteq [n] \\ |T|=j}} \prod_{i \in T} \mathbf{Z}'_i \prod_{i \notin T} (1 - \mathbf{Z}'_i) \right] \\
&\quad = \lambda \sum_{j=k}^{n} \sum_{\substack{T \subseteq [n] \\ |T|=j}} \sum_{\mathbf{z} \in \{0,1\}^n} \mathbb{P}(\mathbf{Z}' = \mathbf{z}') \times \prod_{i \in T} \mathbf{z}'_i \prod_{i \notin T} (1 - \mathbf{z}'_i).
\end{aligned}$$

For a fixed j and a set T with $|T| = j$, there is exactly one $\mathbf{z}'$ such that $\prod_{i \in T} \mathbf{z}'_i \prod_{i \notin T} (1 - \mathbf{z}'_i) = 1$ and the probability of such a $\mathbf{z}'$ is given by $\mathbb{P}(\mathbf{Z}' = \mathbf{z}') = \theta^{\text{ext}}(\lambda, p)^j \times (1 - \theta^{\text{ext}}(\lambda, p))^{n-j}$. Thus, the expression above reduces to

$$\begin{aligned}
\frac{|C'_{k,n} \cap \Gamma_m|}{\nu(\Gamma_m)} &\overset{m\to\infty}{\longrightarrow} \lambda \sum_{j=k}^{n} \sum_{\substack{T \subseteq [n] \\ |T|=j}} \theta^{\text{ext}}(\lambda, p)^j (1 - \theta^{\text{ext}}(\lambda, p))^{n-j} \\
&= \lambda \sum_{j=k}^{n} \binom{n}{j} \theta^{\text{ext}}(\lambda, p)^j (1 - \theta^{\text{ext}}(\lambda, p))^{n-j} \qquad \mathbb{P}\text{-a.s.}
\end{aligned}$$



(a) $p_{k,n,\delta}$ on RGG(4.5, 1) (b) $\tau_{k,n,\delta}$ on RGG(4.5, 1)

(c) $p_{k,n,\delta}$ on RGG(4, 1) (d) $\tau_{k,n,\delta}$ on RGG(4, 1)

Fig. 5. Comparison of simulation results with results obtained via (21) and (18) on $RGG(4.5, 1)$ (top) and $RGG(4, 1)$ (bottom) on Γ_{101} with $k = 20$ packets and $\delta = 0.1$.

Define

$$\theta'_{k,n} \equiv \theta'_{k,n}(\lambda, p) = \sum_{j=k}^{n} \binom{n}{j} \theta(\lambda p)^j (1 - \theta(\lambda p))^{n-j}.$$

From our interpretation of $C'_{k,n}$ as representing $C^{\text{ext}}_{k,n}$ of the original model, we use $\theta'_{k,n}$ instead of $\theta^{\text{ext}}_{k,n}$ in (20), and after a series of manipulations, the minimum forwarding probability obtained via this heuristic approach, $p'_{k,n,\delta}$, would be the minimum probability p such that

$$\frac{1}{\theta(\lambda)} \sum_{t=k}^{n} \sum_{j=k}^{t} \binom{n}{t} \binom{t}{j} \theta(\lambda p)^{t+j} (1 - \theta(\lambda p))^{n-j} \geq 1 - \delta.$$

This expression is similar to the expression that was obtained for the case of a grid in [7]. Using [7, Prop. VI.11], we then have

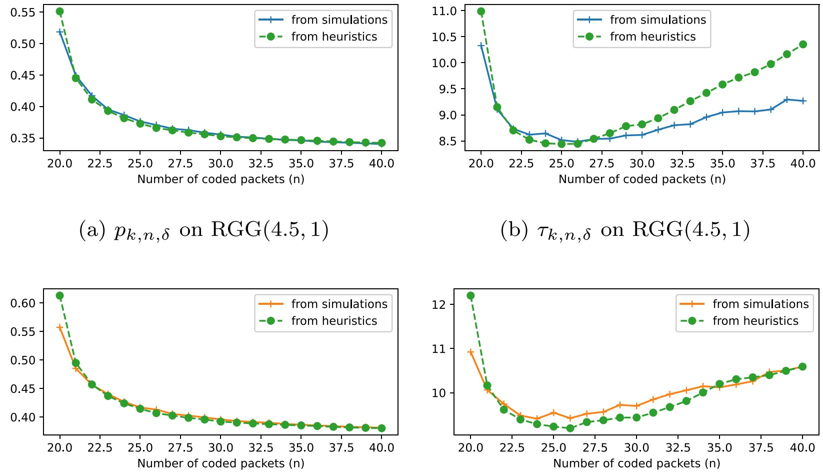
$$p'_{k,n,\delta} = \inf \left\{ p \;\middle|\; \frac{\mathbb{P}(Y \geq k)}{\theta(\lambda)} \geq 1 - \delta \right\} \tag{21}$$

where $Y \sim Bin(n, (\theta(\lambda p))^2)$.

The $p'_{k,n,\delta}$ values obtained using this expression is compared alongside the simulation results in Figs. 5(a) ($\lambda = 4.5$) and 5(c) ($\lambda = 4$). The expected total number of transmissions obtained via (18) is plotted in Figs. 5(b) and 5(d) respectively. The simulation setup is the same as described in Section 4.

It is observed that the curve for the minimum forwarding probability obtained via our analysis tracks the simulation curve pretty well. However, the curve for the expected total number of transmissions deviates from the simulation results substantially for larger values of n. This can be attributed to the drastic change in $\theta(\lambda)$ around the critical intensity λ_c. Even though there seems to be a minor difference in the forwarding probability of the original and the mean-field model, the behaviour of the percolation probability around λ_c creates a huge divide between the two transmission plots in Fig. 5(b). This behaviour is similar to what was obtained on the grid in [7]. Nevertheless, note that the $\tau_{k,n,\delta}$ curve initially decreases to a minimum and then gradually increases with n (albeit very slowly). This shows that probabilistic forwarding with coding is indeed beneficial on RGGs in terms of the number of transmissions required for a near-broadcast.

From a practical viewpoint, this heuristic can be used to obtain the optimal values of the minimum forwarding probability and the number of coded packets required for a near-broadcast. The expression for $p'_{k,n,\delta}$ in (21) provides an easy and relatively accurate way (as seen in Figs. 5(a) and 5(c)) to graph the curve for the optimal forwarding probability for given values of k and δ. This could then be used to choose the value of n for encoding the k data packets so as to have a near-broadcast with the least number of transmissions. A heuristic way to go about this would be to use the smallest n corresponding to the point on the $p'_{k,n,\delta}$ curve where the (discrete) gradient between successive points does not vary



(a) $p_{k,n,\delta}$ on RGG(4.5, 1)   (b) $\tau_{k,n,\delta}$ on RGG(4.5, 1)

(c) $p_{k,n,\delta}$ on RGG(4, 1)   (d) $\tau_{k,n,\delta}$ on RGG(4, 1)

Fig. 6. Comparison of simulation results with heuristics shown in Fig. 5 but with additional correction for the probabilities. The probabilities in Figs. 5(a) and 5(c) were scaled by 1.0189 and 1.032 to obtain the curves corresponding to the heuristics in (a) and (c). Eq. (18) was then used to generate the green curves in (b) and (d). (For interpretation of the references to colour in this figure caption, the reader is referred to the webversion of this article.)

significantly.[3] In Fig. 5(a), this corresponds to the point $n \approx 26$. This point is expected to give the best result for the following two reasons:

- When operating with a lesser value of n (< 26), it is possible to introduce more coded packets so that the effect of the decrease in $p'_{k,n,\delta}$ from Fig. 5(a) overshadows the increase in the number of transmissions brought about by the introduction of the extra packets.
- With additional coded packets ($n > 26$), the number of transmissions increases owing to the forwarding probability not decreasing significantly.

In this way, a judicious choice can be made for n and the forwarding probability $p_{k,n,\delta}$ at which to operate the scheme.

We now provide an alternate methodology to deduce the optimal number of coded packets motivated from our numerical results. It is evident from Fig. 4 that the deviation of the two curves in Figs. 5(b) and 5(d) in predominantly due to the minor difference between the corresponding minimum probability curves in Figs. 5(a) and 5(c). With this insight, we scale the curves obtained via the heuristics in Fig. 5(a) and 5(c) by a factor equal to the ratio of the minimum forwarding probabilities from the simulations and heuristics averaged over all n from 20 to 40. The plots accounting for this correction are provided in Fig. 6 . It can be seen that the corrected curves track the simulations well, and can additionally recover the optimal number of coded packets.

9. Discussion

9.1. A note on our assumptions

In this subsection, we provide some justifications for the assumptions made in our analysis. Our interest in this paper is to broadcast information on large networks. A basic requirement for this is that a large number of nodes in the network must be reachable from the origin. In the sub-critical regime, i.e. $\lambda < \lambda_c \approx 1.44$, the clusters are finite and small. To model large ad-hoc networks, we need the graph to be connected on a large area Γ_m. This necessitates λ to be in the super-critical regime and the component of the origin within Γ_m to be large. In the limit as $m \to \infty$, this requires that the origin be present in the infinite cluster of the underlying RGG, thus justifying Assumption 2.

Further, notice that for a near-broadcast, we need the expected fraction of successful receivers to be close to 1, i.e., $\mathbb{E}^{\mathbf{0}}\left[\frac{|\mathcal{R}_{k,n}(\mathcal{G}^{\mathbf{0}})\cap\Gamma_m|}{\lambda\theta(\lambda)m^2}\right] \geq 1-\delta$ for some small $\delta > 0$ (The denominator here is the expected number of nodes within Γ_m of the infinite cluster C.). If we would like this to hold for sufficiently large m, then the forwarding probability must be such that $\mathcal{R}_{k,n}(\mathcal{G}^{\mathbf{0}})$ has infinite cardinality. This implies that p must be such that there is an IEC during probabilistic forwarding on $\mathcal{G}^{\mathbf{0}}$. Now, since existence of an IEC implies existence of an infinite cluster, the p value must ensure presence of an infinite cluster. Thus $\lambda p > \lambda_c$. This justifies Assumption 1.

[3] This is referred to as the "elbow method" in the clustering literature.

It can also be seen from the simulation results in Fig. 1 that $\tau_{k,n,\delta}$ is minimized when the forwarding probability is such that $\lambda p_{k,n,\delta} > \lambda_c$ or $p_{k,n,\delta} > 0.32$. Further, results obtained from our heuristic approach in Fig. 5 also suggest that the expected total number of transmissions is indeed minimized when operating in the super-critical regime.

9.2. Communication aspects

In this subsection, we consider some of the issues involved in the practical implementation of the proposed probabilistic forwarding algorithm with coded packets. With numerous packets traversing the network, packet collisions are bound to happen. These interference effects need to be handled. Moreover, the channel between adjacent nodes could be error prone resulting in a transmission being lost. Such channel outages need to be addressed as well.

When there are multiple packets in the network, interference effects can be avoided by separating the transmissions either in the frequency domain or in the time domain.

- In the frequency domain, a possible solution is for nodes to transmit on orthogonal sub-carriers of an Orthogonal Frequency Division Multiplexing (OFDM) signal. Alternately, each packet could be transmitted on a different orthogonal sub-carrier. The latter scheme, however, limits the number of packets that can be transmitted concurrently.
- In the time domain, a scheduling algorithm has to be implemented to avoid concurrent transmissions which might interfere. A schedule must not have a pair of nodes that are within two hops from each other in the same slot.

From a graph-theoretic perspective, both these solutions can be viewed as vertex-colouring problems on the underlying graph $G = (V, E)$. The problem of *broadcast scheduling* captures this from a graph-colouring setting and has a vast literature (see e.g., [51,52]). Once a vertex-colouring is obtained, it can be used to implement the above strategies by associating a colour with a particular sub-carrier frequency band or a particular time slot as required. These assignments could be done during the time of network deployment with an associated one-time cost.

When links between adjacent nodes are not ideal noiseless links, the broadcast information received could be distorted. Knowledge of the channel state information (CSI) could be used to further regulate the forwarding probability at every node to overcome channel outages. Each link could be modelled to be either in a 'good' or a 'bad' state based on CSI statistics. The problem then reduces to carrying out the probabilistic forwarding mechanism on a random subgraph of the original network. This involves modelling the process as a joint site-bond percolation on the underlying graph. The techniques necessary to analyse this scenario are more sophisticated than those used in this work, and form a possible future research direction.

9.3. Optimization framework

Our treatment of the problem has shown that on RGGs in the connectivity regime, there exists an optimal value of the number of coded packets n^* and a corresponding forwarding probability p^* that minimizes the expected total number of transmissions while ensuring that the fraction of successful receivers is close to 1. However, our analysis does not provide a way to obtain these optimal values. In the following, we formally state the relevant optimization problem and comment on the same.

As before, let G be the component of the origin in an RGG generated on Γ_m of intensity λ. Denote by N the total number of nodes in G and let T_i be the number of transmissions of packet i, for $i \in [n]$. Recall that $R_{k,n}$ was the number of successful receivers within G. For fixed k and δ, our interest is to find

$$\begin{aligned} (n^*, p^*) \;=\; & \arg\min_{(n,p)} \quad \mathbb{E}\left[\sum_{i=1}^{n} T_i\right] \\ & \text{subject to} \quad \mathbb{E}\left[\frac{R_{k,n}}{N}\right] \geq 1-\delta, \\ & \qquad\qquad\quad n \in \mathbb{N},\; p \in [0,1]. \end{aligned} \tag{22}$$

In this work, we have analysed the above problem when the size of the area on which the RGG is deployed, Γ_m, goes to infinity. We obtain analytical expressions for $\mathbb{E}[T_i]$ and $\mathbb{E}\left[\frac{R_{k,n}}{N}\right]$ in terms of the percolation probability of the RGG as $m \to \infty$. To be more precise, denote the successful receivers by $R_{k,n}(G_m^{\mathbf{0}})$. Using a coupling argument as in the proof of Lemma 4.1(a), we have that $\mathbb{E}\left[\frac{R_{k,n+1}(G_m^{\mathbf{0}})}{N}\right] \geq \mathbb{E}\left[\frac{R_{k,n}(G_m^{\mathbf{0}})}{N}\right]$. Taking the limit as $m \to \infty$, we obtain

$$\lim_{m\to\infty} \mathbb{E}\left[\frac{R_{k,n+1}(G_m^{\mathbf{0}})}{N}\right] \geq \lim_{m\to\infty} \mathbb{E}\left[\frac{R_{k,n}(G_m^{\mathbf{0}})}{N}\right]. \tag{23}$$

Following similar arguments as in obtaining (20) from Theorem 7.7, the limit on the RHS above is given by $\sum_{t=k}^{n-1}\frac{\theta_{k,t}^{\text{ext}}(\theta_{t,n}^{\text{ext}}-\theta_{t+1,n}^{\text{ext}})}{\theta(\lambda)}+\frac{\theta_{k,n}^{\text{ext}}\theta_{n,n}^{\text{ext}}}{\theta(\lambda)}$ and the limit on the LHS is the same expression with n replaced by $n+1$. If we now define for fixed n, k and δ,

$$\mathcal{P}_n \triangleq \left\{p \;\middle|\; \sum_{t=k}^{n-1}\frac{\theta_{k,t}^{\text{ext}}(\theta_{t,n}^{\text{ext}}-\theta_{t+1,n}^{\text{ext}})}{\theta(\lambda)}+\frac{\theta_{k,n}^{\text{ext}}\theta_{n,n}^{\text{ext}}}{\theta(\lambda)} \geq 1-\delta\right\} \text{ and } p(n)\triangleq \inf\mathcal{P}_n, \tag{24}$$

(23) shows that $\{\mathcal{P}_n, n\geq 0\}$ is a non-decreasing sequence of sets and therefore $p(n)$ decreases monotonically as $n\to\infty$. Using this along with the observation that the expression for the expected total number of transmissions in (18) is an increasing function of the forwarding probability, we obtain that,

$$n^* \;=\; \arg\min_{n\geq k}\; n\lambda p(n)\;(\theta(\lambda p(n)))^2 \quad \text{and} \quad p^* \;=\; p(n^*) \quad \text{as } m\to\infty, \tag{25}$$

In the following, we discuss two variations of the above problem:

- **Fixed forwarding probability $\mathbf{p}=\mathbf{p_0}$** : If $p_0 > p_{k,k,\delta}$, then introducing coded packets does not provide any benefit in terms of the expected total number of transmissions. However if $p_0 < p_{k,k,\delta}$, since $\{\mathcal{P}_n, n\geq k\}$ forms an increasing sequence of sets, there exists a minimum value of $n=n_0$ after which $p_0\in\mathcal{P}_n$ for all $n\geq n_0$. This n_0 is therefore the optimal number of coded packets for the fixed probability p_0.
- **Fixed number of coded packets $\mathbf{n}=\mathbf{n_0}\;\geq\;\mathbf{k}$**: Here again, owing to the sequence $p(n)$ decreasing to 0 from Lemma 4.1, there exists a unique probability $p_0=p(n_0)$ given by (20).

In both these scenarios, obtaining a closed form expression for the unknown n_0 or p_0 would require an analytical characterization of the probabilities $\theta_{k,n}^{\text{ext}}$ and $\theta(\lambda)$ appearing in (24). This remains true even if the optimal values are provided, i.e., even if either $p_0=p^*$ or $n_0=n^*$ is given. Some bounds on $\theta_{k,n}^{\text{ext}}$ are provided in the next subsection and obtaining expressions for $\theta(\lambda)$ is an open problem (even on deterministic graphs such as grids). Nevertheless, the heuristic approach presented in Section 8 can be used to obtain the optimal quantities in both the above scenarios, or to solve the optimization problem in (22) directly. A further discussion on this appears in Section 10.

9.4. Bounds on $\theta_{k,n}^{ext}$

We give two lower bounds for $\theta_{k,n}^{\text{ext}}(\lambda,p)$. The probability $\theta_{k,n}^{\text{ext}}(\lambda,p)$ can be expressed in terms of the events A_T^{ext} as follows.

$$\theta_{k,n}^{\text{ext}}(\lambda,p)=\mathbb{P}^{\mathbf{0}}\left(\bigcup_{|T|\geq k}A_T^{\text{ext}}\right)=\sum_{|T|\geq k}\mathbb{P}^{\mathbf{0}}(A_T^{\text{ext}})$$

A simple lower bound for $\theta_{k,n}^{\text{ext}}(\lambda,p)$ can be obtained by taking the term corresponding to $T=[n]$ in the above summation.

$$\begin{aligned}\theta_{k,n}^{\text{ext}}(\lambda,p)\geq \mathbb{P}^{\mathbf{0}}(A_{[n]}^{\text{ext}}) &= \mathbb{P}^{\mathbf{0}}\left(\bigcap_{i=1}^{n}\{\mathbf{0}\in C_i^{\text{ext}}\}\right)\\ &\overset{(a)}{\geq}\prod_{i=1}^{n}\mathbb{P}^{\mathbf{0}}\left(\mathbf{0}\in C_i^{\text{ext}}\right)\\ &=\mathbb{P}^{\mathbf{0}}\left(\mathbf{0}\in C_i^{\text{ext}}\right)^n\end{aligned}$$

Here, the inequality in (a) is via the FKG inequality since the events $\{\mathbf{0}\in C_i^{\text{ext}}\}$ are increasing events. This gives

$$\theta_{k,n}^{\text{ext}}(\lambda,p)\geq\theta(\lambda p)^n. \tag{26}$$

Note that this, along with Assumption 1, suffices to ensure that our analysis yields non-trivial results for all values of k and n.

We now provide a second bound. For this, recall the iid marked point process Φ equipped with the mark structure $\mathbf{Z}$. Define a new marked point process Φ_T with the underlying point process Φ and marks $Z_T=\prod_{i\in T}Z_i\prod_{j\notin T}(1-Z_j)$. The points with mark 1 in Φ_T, form a thinned version of Φ where each vertex is retained with probability $\mathbb{P}(Z_T=1|\Phi)=\mathbb{P}(Z_i=\mathbf{1}\{i\in T\}\;,\;\; i\in[n]|\Phi)=p^{|T|}(1-p)^{n-|T|}$. Thus Φ_T is an iid marked point process with $Ber(p^{|T|}(1-p)^{n-|T|})$ marks.

Let $C^{\text{ext}}(\Phi_T)$ denote the IEC of Φ_T. Notice that

$$\bigcup_{|T|\geq k}\{\mathbf{0}\in C^{\text{ext}}(\Phi_T)\}\subseteq\{\mathbf{0}\in C_{k,n}^{\text{ext}}\}.$$

The probability of the event in the LHS above can be found as

$$\mathbb{P}^{\mathbf{0}}\left(\bigcup_{|T|\geq k}\{\mathbf{0}\in C^{\text{ext}}(\Phi_T)\}\right) = 1-\mathbb{P}^{\mathbf{0}}\left(\bigcap_{|T|\geq k}\{\mathbf{0}\notin C^{\text{ext}}(\Phi_T)\}\right)$$
$$= 1-\prod_{j=k}^{n}\left(1-\theta^{\text{ext}}(\lambda, p^j(1-p)^{n-j})\right)^{\binom{n}{j}}$$

Therefore, the probability $\theta^{\text{ext}}_{k,n}(\lambda,p)$ can be bounded as

$$\theta^{\text{ext}}_{k,n}(\lambda,p)\geq 1-\prod_{j=k}^{n}\left(1-\theta(\lambda p^j(1-p)^{n-j})\right)^{\binom{n}{j}} \tag{27}$$

10. Future work

In this section, we collect some of the questions arising in this work that could lead to possible future research directions. These encompass problems arising from our analysis, numerical experiments, algorithm variants and practical implementation.

1. Probabilistic forwarding of n packets on the RGG gave rise to the term $\theta^{\text{ext}}_{k,n}=\mathbb{P}^{\mathbf{0}}(\mathbf{0}\in C^{\text{ext}}_{k,n})$ in the expression for the expected fraction of successful receivers. While some bounds were obtained for this in Section 9.4, an analytical expression for $\theta^{\text{ext}}_{k,n}$ in terms of $\theta^{\text{ext}}(\lambda,p)$ (which was the probability that the origin belongs to the IEC for site percolation on the RGG), would be useful in obtaining better estimates of $p_{k,n,\delta}$ and $\tau_{k,n,\delta}$. Perhaps, a simpler problem is to find the probability $\mathbb{P}^{\mathbf{0}}(\mathbf{0}\in C^{+}_{k,n})$. In terms of the marked point process formulation, for a point process $\Phi^{\mathbf{0}}$ with independent marks $\mathbf{Z}=(Z_1,Z_2,\ldots,Z_n)$ where $Z_i\sim Ber(p)$, this is the probability that the origin is present in at least k out of the n infinite clusters. Each Z_i corresponds to a site percolation process on the underlying realization of the RGG. Conditional on the underlying RGG (or equivalently, Φ), the events corresponding to the presence of the origin in the infinite cluster of the ith and the jth percolation processes are independent. However, this is not true unconditionally. Intuitively, it is expected that the presence of the origin in the ith infinite cluster makes it more likely for it to be present in the jth infinite cluster as well. A mathematically rigorous understanding of this phenomenon is necessary.
2. Concerning the optimization framework developed in Section 9.3, it was indicated that the heuristics presented in Section 8 can be employed to obtain the optimal values of the number of coded packets and the forwarding probability. This assumed that the estimates for the probabilities $\theta^{\text{ext}}_{k,n}$ and $\theta(\lambda)$ obtained numerically using the ergodic theorems (9) and (12) well-approximated the actual values in the limit as $m\to\infty$. Naturally, a second-moment characterization of the fraction of successful receivers and the expected total number of transmissions in the asymptotic regime will provide better indication of the validity of these estimates. Alternately, one could consider solving the optimization problem stated in (22) for a fixed m using other methodologies. We believe the techniques required for these approaches are more sophisticated and span an interesting future direction for this line of work.
3. The assumption that the random geometric graph operates in the super-critical region is inherent in our analysis. In fact, as discussed in Section 9.1, most of our results require $\lambda p>\lambda_c$. However, as shown in Lemma 4.1(b), the forwarding probability diminishes to 0 as $n\to\infty$. Thus, for large n, the thinned RGG of intensity λp consisting of only the transmitters operates in the sub-critical regime ($\lambda p<\lambda_c$). A comprehensive study of the probabilistic forwarding mechanism in the sub-critical regime will help provide an overall understanding of the problem. In particular, this might provide further insight into the deviation of our heuristics from the observed simulations in Figs. 5(b) and 5(d).
4. The probabilistic forwarding mechanism with coded packets is a completely decentralized and distributed algorithm. This makes it amenable to be deployed on mobile ad-hoc networks (MANETs) or vehicular networks (VANETs) where the individual nodes are moving. Moreover, simulation studies indicate that mobility improves connectivity in such networks (see e.g., [53–56]). Additionally, other metrics of performance can be incorporated in this scenario such as delay [57–59], age of information (AoI) [60,61], percolation and connection times [62] etc. Thus, an interesting future direction is to investigate the performance of the probabilistic forwarding mechanism with coded packets on MANETs considering these metrics as well.
5. As an extension of the techniques presented here, one could consider each communication link between nodes to be noisy. Then, even though a node might forward a packet with probability p, it will be received only by a subset of its neighbours depending on the packet drop probability, q, induced by the noisy channel. This can be modelled as simultaneous bond and site percolation on the underlying graph, a process that does not seem to have received much attention on random graphs.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: B. R. Vinay Kumar reports financial support was provided by Centre for Networked Intelligence (a Cisco CSR initiative), Indian Institute of Science. Navin Kashyap reports financial support was provided by Swarnajayanti Fellowship, Dept. of Science & Technology, Govt. of India. D. Yogeshwaran reports financial support was provided by SERB-MATRICS grant MTR-2020-000470. This work was presented in part at the 19th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt 2021), which was held virtually during Oct 18-21, 2021.

Data availability

No data was used for the research described in the article.

Acknowledgements

B. R. Vinay Kumar was supported in part by a fellowship from the Centre for Networked Intelligence (a Cisco CSR initiative) of the Indian Institute of Science, and in part by the RISE Academie-1 post-doctoral fellowship of the Université Côte d'Azur, France.

N. Kashyap was supported in part by a Swarnajayanti Fellowship awarded by the Department of Science & Technology, Government of India.

D. Yogeshwaran was supported in part by CPDA, Indian Statistical Institute, and in part by a SERB-MATRICS grant MTR/2020/000470.

This work was presented in part at the 19th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt 2021), which was held virtually during Oct 18–21, 2021.

Appendix A. Palm expectations of infinite cluster densities

In this section, we prove three main propositions which are used in the analysis of the probabilistic forwarding protocol. Let $\mathcal{G} \sim RGG(\lambda, 1)$ be a random geometric graph on $\mathbb{R}^2$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The underlying Poisson point process, Φ, is of intensity λ. The intensity λ is such that we operate in the super-critical region, i.e., $\lambda > \lambda_c$. Let $C \equiv C(\Phi)$ be the unique infinite cluster in $\mathcal{G}$. Let $\Phi^{\mathbf{0}} = \Phi \cup \{\mathbf{0}\}$ denote the Palm version of Φ and let $C(\Phi^{\mathbf{0}})$ be the infinite cluster in it. Denote by $\mathbb{P}^{\mathbf{0}}$, the Palm probability of the origin and $\mathbb{E}^{\mathbf{0}}$, the expectation with respect to $\mathbb{P}^{\mathbf{0}}$. We now show that the limiting fraction of vertices in C within Γ_m remains the same with respect to both $\mathbb{E}$ and $\mathbb{E}^{\mathbf{0}}$.

Proposition A.1.

$$\lim_{m\to\infty} \mathbb{E}^{\mathbf{0}}\left[\frac{|C \cap \Gamma_m|}{m^2}\right] = \lim_{m\to\infty} \mathbb{E}\left[\frac{|C \cap \Gamma_m|}{m^2}\right]$$

Proof. Let $C_1, C_2, \dots, C_K$ be finite components in $\mathcal{G}$ which intersect the ball of radius 1 centred at the origin, i.e., $C_i \cap B_1(\mathbf{0}) \neq \emptyset, \quad \forall i \in \{1, 2, \dots, K\}$. Since vertices from distinct finite components C_i and C_j, should be at least at a distance of 1 from each other, the number of such components is bounded. In particular, K is a random variable with $K \le 7$ *a.s.* The infinite clusters in the $RGG(\Phi^{\mathbf{0}}, 1)$ and $RGG(\Phi, 1)$ models can be related in the following way:

$$C(\Phi^{\mathbf{0}}) = \begin{cases} C(\Phi) \cup C_1 \cup \cdots \cup C_K \cup \{\mathbf{0}\} & \text{if } C \cap B_1(\mathbf{0}) \neq \emptyset \\ C(\Phi) & \text{if } C \cap B_1(\mathbf{0}) = \emptyset \end{cases}$$

Using this, we can write

$$\frac{|C(\Phi^{\mathbf{0}}) \cap \Gamma_m|}{m^2} = \frac{|C(\Phi) \cap \Gamma_m|}{m^2} + \sum_{i=1}^{K} \frac{|C_i \cap \Gamma_m|}{m^2} \mathbb{1}\{C \cap B_1(\mathbf{0}) \neq \emptyset\}$$

Since $K \le 7$ *a.s.* and $|C_i| < \infty$ for all $i = 1, 2, \dots, K$, we have

$$\sum_{i=1}^{K} \frac{|C_i \cap \Gamma_m|}{m^2} \xrightarrow{m\to\infty} 0 \quad \mathbb{P}\text{-a.s.}.$$

Thus, we deduce that

$$\lim_{m\to\infty} \frac{|C(\Phi^{\mathbf{0}}) \cap \Gamma_m|}{m^2} = \lim_{m\to\infty} \frac{|C(\Phi) \cap \Gamma_m|}{m^2} \qquad \mathbb{P}\text{-a.s.} \tag{A.1}$$

Since the random variables involved are bounded by 1, applying the dominated convergence theorem (DCT) gives the desired result. □

Corollary A.2.

$$\lim_{m\to\infty} \mathbb{E}^{\mathbf{0}}\left[\frac{|C\cap \Gamma_m|}{m^2}\right] = \lambda\theta(\lambda)$$

Proof. This directly follows from the previous proposition and (8). □

Next, consider the formulation of the marked point process described in Section 6. Let $C^{\mathrm{ext}} \equiv C^{\mathrm{ext}}(\Phi)$ be the infinite extended cluster (IEC). We now show an analogue of the previous proposition for C^{ext}.

Proposition A.3.

$$\lim_{m\to\infty} \mathbb{E}^{\mathbf{0}}\left[\frac{|C^{ext}\cap \Gamma_m|}{m^2}\right] = \lim_{m\to\infty} \mathbb{E}\left[\frac{|C^{ext}\cap \Gamma_m|}{m^2}\right]$$

Proof. The proof is along the same lines as that in Proposition A.1. Let $C_1, C_2, \ldots, C_K$ be finite components in $\mathcal{G}^+$ which intersect the ball of radius 1 centred at the origin, i.e., $C_i \cap B_1(\mathbf{0}) \neq \emptyset, \quad \forall i \in \{1, 2, \ldots, K\}$. Here again $K \leq 7$ *a.s.* Now, suppose that $C^+ \cap B_1(\mathbf{0}) \neq \emptyset$, then regardless of the mark of the origin, it is true that $C^{\mathrm{ext}}(\Phi^{\mathbf{0}}) \subseteq C^{\mathrm{ext}}(\Phi) \cup C_1^{\mathrm{ext}} \cup \cdots \cup C_K^{\mathrm{ext}}$ (with equality being true when the origin has mark 1). If on the other hand $C^+ \cap B_1(\mathbf{0}) = \emptyset$, then $C^{\mathrm{ext}}(\Phi^{\mathbf{0}}) = C^{\mathrm{ext}}(\Phi)$. Using this, we can write

$$\frac{|C^{\mathrm{ext}}(\Phi^{\mathbf{0}})\cap \Gamma_m|}{m^2} \leq \frac{|C^{\mathrm{ext}}(\Phi)\cap \Gamma_m|}{m^2} + \sum_{i=1}^{K}\frac{|C_i^{\mathrm{ext}}\cap \Gamma_m|}{m^2}\,\mathbb{1}\{C^+\cap B_1(\mathbf{0})\neq\emptyset\}.$$

Note that, if C_i is a finite cluster, then so is C_i^{ext} and hence the summation on the RHS above tends to 0 as $m \to \infty$. Since we trivially have that

$$\frac{|C^{\mathrm{ext}}(\Phi)\cap \Gamma_m|}{m^2} \leq \frac{|C^{\mathrm{ext}}(\Phi^{\mathbf{0}})\cap \Gamma_m|}{m^2},$$

in the limit of large m, the fraction $\frac{|C^{\mathrm{ext}}(\Phi^{\mathbf{0}})\cap \Gamma_m|}{m^2}$ is sandwiched between the two limits yielding

$$\lim_{m\to\infty}\frac{|C^{\mathrm{ext}}(\Phi^{\mathbf{0}})\cap \Gamma_m|}{m^2} = \lim_{m\to\infty}\frac{|C^{\mathrm{ext}}(\Phi)\cap \Gamma_m|}{m^2} \qquad \mathbb{P}\text{-a.s.}$$

Using DCT gives the statement of the proposition. □

A similar argument extends to $C_{k,n}^{\mathrm{ext}}$ as well, which is stated in the following proposition.

Proposition A.4.

$$\lim_{m\to\infty} \mathbb{E}^{\mathbf{0}}\left[\frac{|C_{k,n}^{ext}\cap \Gamma_m|}{m^2}\right] = \lim_{m\to\infty} \mathbb{E}\left[\frac{|C_{k,n}^{ext}\cap \Gamma_m|}{m^2}\right]$$

Proof. Firstly, note that

$$\frac{|C_{k,n}^{\mathrm{ext}}(\Phi^{\mathbf{0}})\cap \Gamma_m|}{m^2} \geq \frac{|C_{k,n}^{\mathrm{ext}}(\Phi)\cap \Gamma_m|}{m^2}. \tag{A.2}$$

The nodes in $C_{k,n}^{\mathrm{ext}}(\Phi^{\mathbf{0}})$ can be related to those in $C_{k,n}^{\mathrm{ext}}(\Phi)$ in the following way. Let $C_1^+, C_2^+, \ldots, C_n^+$ denote the infinite clusters corresponding to each of the n packets and let $C_{i,1}, C_{i,2}, \ldots, C_{i,K_i}$ denote the finite clusters corresponding to the $i-$th packet which intersect the ball of radius 1 at the origin. Here again, $K_i \leq 7$ *a.s.* for all i. Proceeding with similar reasoning as that of Proposition A.3, we can obtain

$$\frac{|C_{k,n}^{\mathrm{ext}}(\Phi^{\mathbf{0}})\cap \Gamma_m|}{m^2} \leq \frac{|C_{k,n}^{\mathrm{ext}}(\Phi)\cap \Gamma_m|}{m^2} + \sum_{\substack{i\in[n] \\ C_i^+\cap B_1(\mathbf{0})\neq\emptyset}}\sum_{j=1}^{K_i}\frac{|C_{i,j}^{\mathrm{ext}}\cap \Gamma_m|}{m^2} \tag{A.3}$$

The summation on the RHS is a finite sum with at most $7n$ terms with each term consisting of fraction of nodes in some finite cluster. By taking limits as $m \to \infty$, this fraction vanishes. Therefore the fraction $\frac{|C_{k,n}^{\mathrm{ext}}(\Phi^{\mathbf{0}})\cap \Gamma_m|}{m^2}$ is sandwiched between the two limits in (A.2) and (A.3) yielding

$$\lim_{m\to\infty}\frac{|C_{k,n}^{\mathrm{ext}}(\Phi^{\mathbf{0}})\cap \Gamma_m|}{m^2} = \lim_{m\to\infty}\frac{|C_{k,n}^{\mathrm{ext}}(\Phi)\cap \Gamma_m|}{m^2} \qquad \mathbb{P}\text{-a.s.}$$

Using DCT gives the statement of the proposition. □

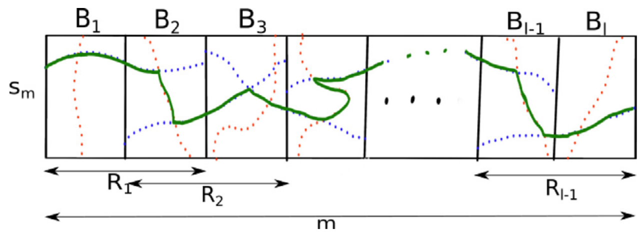


Fig. B.7. Left–right crossing in the $s_m \times m$ rectangular box B through the events LR_j and TB_i.

Appendix B. Estimates on some crossing probabilities

Let Φ be a homogeneous Poisson point process of intensity $\lambda > \lambda_c$ on the whole $\mathbb{R}^2$ plane. On a box $B_{a,b} = [0, b]\times[0, a]$, a *left–right crossing* of $B_{a,b}$ is defined as a sequence of vertices $\{X_i, i = 1, 2, \ldots, s\}$, such that[4] $\|X_i - X_{i-1}\| \le 1$ for $i = 2, 3, \ldots, s$ and $\|X_1 - x\| \le 1$ and $\|X_s - y\| \le 1$ for some $x \in \{0\} \times [0, a]$ and $y \in \{b\} \times [0, a]$. A *top-bottom crossing* is defined similarly but with $x \in [0, b] \times \{0\}$ and $y \in [0, b] \times \{a\}$. If $a = b$, we simply denote the square box by B_a.

Define $LR(a)$ to be the event that there is a left right crossing in a rectangular box $R_a = B_{a,2a} = [0, 2a] \times [0, a]$. The probability of $LR(a)$ in the super-critical region is exponentially close to 1 as formalized in [45, Lemma 10.5]. We reproduce the same here.

Lemma B.1. *For $\lambda > \lambda_c$, there exists $c > 0$ and $a_1 > 0$ such that $1 - \mathbb{P}(LR(a)) \le \exp(-ca)$ for all $a \ge a_1$.*

We will use this lemma to obtain the probability of a left–right crossing in a $s_m \times m$ rectangular box, where $s_m \ll m$. Let CR be the event that there is a left–right crossing of the box $B = [0, m] \times [0, s_m]$. We then have the following proposition.

Proposition B.2. *For $\lambda > \lambda_c$, there exists $c > 0$ and $a_1 > 0$ such that $1 - \mathbb{P}(CR) \le 2\left\lceil \frac{m}{s_m} \right\rceil \exp(-cs_m)$ for all $s_m \ge a_1$.*

Proof. Denote $\ell = \left\lceil \frac{m}{s_m} \right\rceil$. Let $B_i = [(i-1)s_m, is_m]\times[0, s_m]$ for $i \in \{1, 2, \ldots, \ell\}$ and let $R_j = B_j \cup B_{j+1}$ for $j \in \{1, 2, \ldots, \ell-1\}$ (see Fig. B.7). Define LR_j to be the event that there is a left–right crossing in R_j and let TB_i be the event that there is a top-bottom crossing of B_i. Notice that

$$CR \supseteq \bigcap_{i=1}^{\ell} TB_i \cap \bigcap_{j=1}^{\ell-1} LR_j,$$

which gives

$$\mathbb{P}(CR^c) \le \sum_{i=1}^{\ell} \mathbb{P}(TB_i^c) \; + \; \sum_{j=1}^{\ell-1} \mathbb{P}(LR_j^c).$$

The probability of there being no left–right crossings in the rectangles R_j, for $j \in \{1, 2, \ldots, \ell - 1\}$, are identical (due to translation invariance) and hence the latter term in the above expression can be replaced by $(\ell - 1)\mathbb{P}(LR_1^c)$. For the first term, note that absence of a top-bottom crossing of B_i implies that there is no top-bottom crossing in the rectangle $R_i' = [(i-1)s_m, is_m] \times [0, 2s_m]$. But a top-bottom crossing in R_i' is the same as a left–right crossing in R_1 (say), since the underlying homogeneous Poisson point process Φ is isotropic. This gives

$$\mathbb{P}(CR^c) \le (2\ell - 1)P(LR_1^c),$$

which from Lemma B.1 gives the statement of the proposition. □

Next, we apply Proposition B.2 to the four rectangles surrounding Γ_r as depicted in Fig. B.8 . Let CR_d for $d \in \{n, s, e, w\}$ be the event denoting the existence of crossings inside the four rectangles and let Ann_{s_m} be the event that there is a circuit in the annulus $\Gamma_{m-1} \setminus \Gamma_r$ as shown in Fig. B.8. Since the presence of crossings in the four rectangles ensures the occurrence of Ann_{s_m}, we obtain

$$\mathbb{P}(\text{Ann}_{s_m}^c) \le \mathbb{P}\left(\bigcup_d CR_d^c\right),$$

[4] Here $\|\cdot\|$ is the L^2 norm.

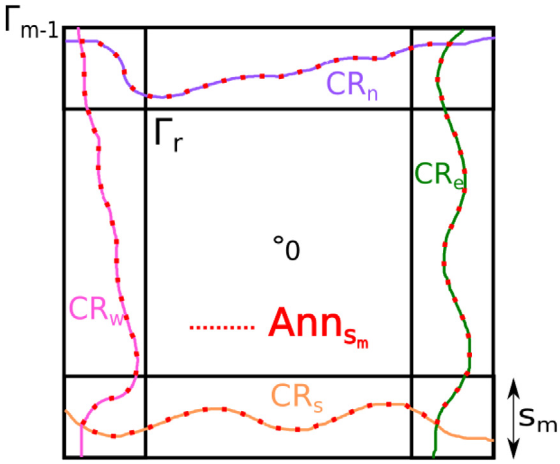


Fig. B.8. Circuit formed by the four left–right crossings $LR_d, d \in \{n, s, e, w\}$.

$$\leq \sum_d \mathbb{P}(CR_d^c),$$
$$\leq 8\left\lceil \frac{m}{s_m} \right\rceil \exp(-cs_m).$$

We state this formally in the following proposition.

Proposition B.3. *For $\lambda > \lambda_c$, there exists $c > 0$ and $a_1 > 0$ such that $1 - \mathbb{P}(Ann_{s_m}) \leq 8\left\lceil \frac{m}{s_m} \right\rceil \exp(-cs_m)$ for all $s_m \geq a_1$.*

Remark. Note that the statement of the above proposition holds even with respect to the Palm probability $\mathbb{P}^{\mathbf{0}}$. This is because introducing a point at the origin does not affect the event Ann_{s_m}, and hence $\mathbb{P}^{\mathbf{0}}(\mathrm{Ann}_{s_m}) = \mathbb{P}(\mathrm{Ann}_{s_m})$.

Appendix C. Proof of Theorem 7.7

Theorem C.1 (*Restatement of Theorem 7.7*)**.** *For $\lambda p > \lambda_c$, we have*

$$\lim_{m\to\infty} \mathbb{E}_A^{(\mathbf{0},\mathbf{1})}\left[\frac{R_{k,n}(G_m)}{|C_{\mathbf{0}}(G_m)|}\right] = \frac{1}{\theta(\lambda)^2}\sum_{t=k}^{n}\binom{n}{t}\theta_{k,t}^{ext}\ \mathbb{P}^{(\mathbf{0},\mathbf{1})}(A_{[t]}^{ext}).$$

Proof. Step 1: We first evaluate

$$\lim_{m\to\infty} \mathbb{E}^{(\mathbf{0},\mathbf{1})}\left[\frac{R_{k,n}(G_m)}{|C_{\mathbf{0}}(G_m)|}\mathbf{1}_A\right] = \lim_{m\to\infty} \mathbb{E}^{(\mathbf{0},\mathbf{1})}\left[\frac{R_{k,n}(G_m)\mathbf{1}_A}{|C_{\mathbf{0}}(G_m)|\mathbf{1}_A}\right]$$

and then divide it by $\mathbb{P}(A) = \mathbb{P}(\mathbf{0} \in C(\mathcal{G}^0)) = \theta(\lambda)$ to obtain the required conditional expectation. We take the convention that $\frac{0}{0} = 0$. Note that Assumption 1 ensures that $\theta(\lambda) > 0$.

Step 2: Specializing the statement of Theorem 7.1 on the event A, we obtain

$$\lim_{m\to\infty} \frac{|C_{\mathbf{0}}(G_m^{\mathbf{0}})|}{\lambda m^2}\mathbf{1}_A = \lim_{m\to\infty}\frac{|C_{\mathbf{0}}(\mathcal{G}^{\mathbf{0}}) \cap \Gamma_m|}{\lambda m^2}\mathbf{1}_A \qquad \mathbb{P}\text{-a.s.}$$

Notice that on the event A, $C_{\mathbf{0}}(\mathcal{G}^{\mathbf{0}}) = C(\mathcal{G}^{\mathbf{0}})$. Using (A.1), (8) and the note following Lemma 7.5, we have for $\lambda > \lambda_c$

$$\lim_{m\to\infty} \frac{|C_{\mathbf{0}}(G_m)|}{\lambda m^2}\mathbf{1}_A = \lim_{m\to\infty}\frac{|C(\mathcal{G}) \cap \Gamma_m|}{\lambda m^2}\mathbf{1}_A$$
$$= \theta(\lambda)\mathbf{1}_A \qquad \mathbb{P}^{\mathbf{0}}\text{-a.s.}$$

Conditional on the mark of the origin $\mathbf{Z}(\mathbf{0}) = \mathbf{1}$, we have

$$\lim_{m\to\infty} \frac{|C_{\mathbf{0}}(G_m)|}{\lambda m^2}\mathbf{1}_A = \theta(\lambda)\mathbf{1}_A \qquad \mathbb{P}^{(\mathbf{0},\mathbf{1})}\text{-a.s.}$$

Step 3: Let $\mathcal{R}_{k,n}(\mathcal{G})$ be the set of nodes that receive at least k out of the n packets from the origin when probabilistic forwarding is carried out on $\mathcal{G}$. Using arguments similar to those of Theorem 7.1 for nodes without Γ_m-conduits, we have that

$$\lim_{m\to\infty} \mathbb{E}^{(\mathbf{0},\mathbf{1})}\left[\frac{R_{k,n}(G_m)}{\lambda m^2}\mathbf{1}_A\right] = \lim_{m\to\infty} \mathbb{E}^{(\mathbf{0},\mathbf{1})}\left[\frac{|\mathcal{R}_{k,n}(\mathcal{G})\cap\Gamma_m|}{\lambda m^2}\mathbf{1}_A\right].$$

Step 4: For $T \subseteq [n]$, let A_T^{ext} be the event that the origin is present in exactly the IECs indexed by T. Conditioning on the event A_T^{ext}, we obtain

$$\mathbb{E}^{(\mathbf{0},\mathbf{1})}\left[\frac{|\mathcal{R}_{k,n}(\mathcal{G})\cap\Gamma_m|}{\lambda m^2}\mathbf{1}_A\right] = \sum_{t=0}^{n}\sum_{\substack{T\subseteq[n]\\|T|=t}} \mathbb{E}^{(\mathbf{0},\mathbf{1})}\left[\frac{|\mathcal{R}_{k,n}(\mathcal{G})\cap\Gamma_m|}{\lambda m^2}\mathbf{1}_A\middle|A_T^{\text{ext}}\right]\mathbb{P}^{(\mathbf{0},\mathbf{1})}(A_T^{\text{ext}}). \tag{C.1}$$

If $|T| < k$, then the nodes of $\mathcal{R}_{k,n}(\mathcal{G})$ within Γ_m must reside in finite clusters whose fraction vanishes in the limit of large m. If $|T| \geq k$, then it is only the nodes which are within at least k IECs among those packet transmissions which are indexed by T, that contribute towards the expectation. Denote such nodes by $\mathcal{R}_{k,T}$. The remaining nodes of $\mathcal{R}_{k,n}(\mathcal{G})$ within Γ_m, must be in at least one finite cluster and hence their fraction vanishes in the limit. Additionally, given A_T^{ext} for $|T| > 0$, the $\mathbf{0}$ must be present in the infinite cluster of the underlying graph i.e., $\mathbf{1}_A = 1$. Putting all these together, we obtain

$$\lim_{m\to\infty} \mathbb{E}^{(\mathbf{0},\mathbf{1})}\left[\frac{|\mathcal{R}_{k,n}(\mathcal{G})\cap\Gamma_m|}{\lambda m^2}\mathbf{1}_A\right] = \lim_{m\to\infty}\sum_{t=k}^{n}\sum_{\substack{T\subseteq[n]\\|T|=t}} \mathbb{E}^{(\mathbf{0},\mathbf{1})}\left[\frac{|\mathcal{R}_{k,T}\cap\Gamma_m|}{\lambda m^2}\middle|A_T^{\text{ext}}\right]\mathbb{P}^{(\mathbf{0},\mathbf{1})}(A_T^{\text{ext}}). \tag{C.2}$$

Step 5: Define $\mathbf{O}$ to be the event that the origin has mark 1 in all the n packet transmissions. The expectation on the RHS in the above equation can be written as

$$\mathbb{E}^{(\mathbf{0},\mathbf{1})}\left[\frac{|\mathcal{R}_{k,T}\cap\Gamma_m|}{\lambda m^2}\middle|A_T^{\text{ext}}\right] = \mathbb{E}^{\mathbf{0}}\left[\frac{|\mathcal{R}_{k,T}\cap\Gamma_m|}{\lambda m^2}\middle|A_T^{\text{ext}}\cap\mathbf{O}\right].$$

$\mathcal{R}_{k,T}$ is independent of the packet transmissions which are not in T. The event $\mathbf{O}$ can be thus restricted to only those indices in T. However, the conditioning event $A_T^{\text{ext}}\cap\mathbf{O}$ is then the event that $\mathbf{0}$ is in the infinite cluster C^+ in the packet transmissions indexed by T. Call this event A_T^+. We then have

$$\mathbb{E}^{(\mathbf{0},\mathbf{1})}\left[\frac{|\mathcal{R}_{k,T}\cap\Gamma_m|}{\lambda m^2}\middle|A_T^{\text{ext}}\right] = \mathbb{E}^{\mathbf{0}}\left[\frac{|\mathcal{R}_{k,T}\cap\Gamma_m|}{\lambda m^2}\middle|A_T^{+}\right] \tag{C.3}$$

Step 6: Conditional on the event A_T^+, the set $\mathcal{R}_{k,T}$ has the same distribution as the set $C_{k,|T|}^{\text{ext}}$, which was defined in Section 6.3. This gives

$$\mathbb{E}^{\mathbf{0}}\left[\frac{|\mathcal{R}_{k,T}\cap\Gamma_m|}{\lambda m^2}\middle|A_T^{+}\right] = \mathbb{E}^{\mathbf{0}}\left[\frac{|C_{k,|T|}^{\text{ext}}\cap\Gamma_m|}{\lambda m^2}\right].$$

From Proposition A.4, by taking limits as $m\to\infty$, the expectation with respect to the Palm probability, $\mathbb{E}^{\mathbf{0}}$, can be written in terms of the expectation $\mathbb{E}$, yielding

$$\lim_{m\to\infty}\mathbb{E}^{\mathbf{0}}\left[\frac{|\mathcal{R}_{k,T}\cap\Gamma_m|}{\lambda m^2}\middle|A_T^{+}\right] = \lim_{m\to\infty}\mathbb{E}\left[\frac{|C_{k,|T|}^{\text{ext}}\cap\Gamma_m|}{\lambda m^2}\right] \tag{C.4}$$

Step 7: Using (12) with n replaced by $|T| = t$ and employing DCT, we obtain

$$\lim_{m\to\infty}\mathbb{E}\left[\frac{|C_{k,|T|}^{\text{ext}}\cap\Gamma_m|}{\lambda m^2}\right] = \theta_{k,t}^{\text{ext}}(\lambda,p) \tag{C.5}$$

Step 8: Clubbing the expressions from (C.3), (C.4) and (C.5) into (C.2), and using C.1, we obtain

$$\lim_{m\to\infty}\mathbb{E}^{(\mathbf{0},\mathbf{1})}\left[\frac{R_{k,n}(G_m)}{\lambda m^2}\mathbf{1}_A\right] = \sum_{t=k}^{n}\sum_{\substack{T\subseteq[n]\\|T|=t}} \theta_{k,t}^{\text{ext}}\,\mathbb{P}^{(\mathbf{0},\mathbf{1})}(A_T^{\text{ext}}).$$

Step 9: The event A_T^{ext} can be expressed as

$$A_T^{\text{ext}} = \bigcap_{i \in T} \{\mathbf{0} \in C_i^{\text{ext}}\} \bigcap_{j \notin T} \{\mathbf{0} \notin C_j^{\text{ext}}\}.$$

Here, $C_1^{\text{ext}}, C_2^{\text{ext}}, \ldots, C_n^{\text{ext}}$ denote the IECs corresponding to the n packet transmissions. Since $\{\mathbf{0} \in C_i^{\text{ext}}\} = \{B_1(\mathbf{0}) \cap C_i^+ \neq \emptyset\}$, the event A_T^{ext} does not depend on the specific mark of $\mathbf{0}$. Furthermore, the event A_T^{ext} does not depend on the specific choice of the set T, but just on the cardinality $|T|$. This is because a relabelling of the packets does not alter the probability of A_T^{ext}. For a particular value of $|T| = t$, define

$$A_{[t]}^{\text{ext}} = \bigcap_{i=1}^{t} \{\mathbf{0} \in C_i^{\text{ext}}\} \bigcap_{j=t+1}^{n} \{\mathbf{0} \notin C_j^{\text{ext}}\}.$$

Notice now that the terms within the summation in Step 7, $\theta_{k,t}^{\text{ext}}\ \mathbb{P}^{(\mathbf{0},\mathbf{1})}(A_T^{\text{ext}})$ are identical for different T with the same cardinality. Therefore,

$$\lim_{m \to \infty} \mathbb{E}^{(\mathbf{0},\mathbf{1})} \left[\frac{R_{k,n}(G_m)}{\lambda m^2} \mathbf{1}_A \right] = \sum_{t=k}^{n} \binom{n}{t} \theta_{k,t}^{\text{ext}}\ \mathbb{P}^{(\mathbf{0},\mathbf{1})}(A_T^{\text{ext}}).$$

Step 10: Putting together the results from Step 2 and Step 9 and dividing by $\theta(\lambda)$ gives the statement of the theorem. □

References

[1] R. Vaze, Random Wireless Networks, Cambridge Univ. Press, 2015.
[2] M. Franceschetti, R. Meester, Random Networks for Communication: From Statistical Physics To Information Systems, Cambridge Univ. Press, 2008.
[3] Y.-C. Tseng, S.-Y. Ni, Y.-S. Chen, J.-P. Sheu, The broadcast storm problem in a mobile ad hoc network, Wirel. Netw. 8 (2/3) (2002) 153–167.
[4] F. Forero Rodríguez, Geometric Aspects and Random Delays in Probabilistic Broadcasting for Wireless Ad Hoc Networks (Ph.D. thesis), Univ. Estadual de Campinas, Brazil, and Univ. de Los Andes, Colombia, 2019.
[5] Y. Sasson, D. Cavin, A. Schiper, Probabilistic broadcast for flooding in wireless mobile ad hoc networks, in: Proc. WCNC 2003, Vol. 2, 2003, pp. 1124–1130.
[6] Z.J. Haas, J.Y. Halpern, L. Li, Gossip-based ad hoc routing, IEEE/ACM Trans. Netw. 14 (3) (2006) 479–491.
[7] B.R. Vinay Kumar, N. Kashyap, Probabilistic forwarding of coded packets on networks, IEEE/ACM Trans. Netw. 29 (1) (2021) 234–247, http://dx.doi.org/10.1109/TNET.2020.3031467.
[8] B. Williams, T. Camp, Comparison of broadcasting techniques for mobile ad hoc networks, in: Proc. 3rd ACM Int. Symp. Mobile Ad Hoc Networking & Computing, 2002, pp. 194–205.
[9] C. Fragouli, J. Widmer, J.-Y. Le Boudec, Efficient broadcasting using network coding, IEEE/ACM Trans. Netw. 16 (2) (2008) 450–463.
[10] S. Wang, G. Tan, Y. Liu, H. Jiang, T. He, Coding opportunity aware backbone metrics for broadcast in wireless networks, IEEE Trans. Parallel Distrib. Syst. 25 (8) (2013) 1999–2009.
[11] T. Ho, M. Médard, R. Koetter, D.R. Karger, M. Effros, J. Shi, B. Leong, A random linear network coding approach to multicast, IEEE Trans. Inform. Theory 52 (10) (2006) 4413–4430.
[12] T.A. Courtade, R.D. Wesel, Efficient universal recovery in broadcast networks, in: 2010 48th Annual Allerton Conference on Communication, Control, and Computing, Allerton, 2010, pp. 1542–1549, http://dx.doi.org/10.1109/ALLERTON.2010.5707096.
[13] D. Platz, D.H. Woldegebreal, H. Karl, Random network coding in wireless sensor networks: Energy efficiency via cross-layer approach, in: 2008 IEEE 10th International Symposium on Spread Spectrum Techniques and Applications, 2008, pp. 654–660.
[14] X. Wang, R. Hu, L. Wang, D. Gao, Y. Su, B. Yang, An efficient network-wide reliable broadcast protocol for medical sensor networks, in: Industrial IoT Technologies and Applications: 4th EAI International Conference, Industrial IoT 2020, Virtual Event, December 11, 2020, Springer International Publishing, 2021, pp. 34–44.
[15] M.A. Graham, A. Ganesh, R.J. Piechocki, Fountain coding enabled data dissemination for connected and automated vehicles, in: 2019 IEEE 89th Vehicular Technology Conference, VTC2019-Spring, 2019, pp. 1–5, http://dx.doi.org/10.1109/VTCSpring.2019.8746545.
[16] K. Al Agha, N. Kadi, I. Stojmenovic, Fountain codes with XOR of encoded packets for broadcasting and source independent backbone in multi-hop networks using network coding, in: 2009 IEEE 69th Vehicular Technology Conference, VTC2009-Spring, 2009, pp. 1–5.
[17] M.S. Zaman, G. RamaMurthy, A new degree distribution for LT codes for broadcasting in ad-hoc network using network coding, in: 2009 1st UK-India International Workshop on Cognitive Wireless Systems, UKIWCWS, 2009, pp. 1–5, http://dx.doi.org/10.1109/UKIWCWS.2009.5749389.
[18] N. He, Y. Xu, J. Cao, Z. Li, H. Chen, Y. Ren, ROME: Rateless online MDS code for wireless data broadcasting, in: 2010 IEEE Global Telecommunications Conference, GLOBECOM, 2010, pp. 1–5, http://dx.doi.org/10.1109/GLOCOM.2010.5683211.
[19] A. Gaba, S. Voulgaris, K. Iwanicki, M. van Steen, Revisiting gossip-based ad-hoc routing, in: 2012 21st International Conference on Computer Communications and Networks, ICCCN, 2012, pp. 1–6, http://dx.doi.org/10.1109/ICCCN.2012.6289312.
[20] I.S. Lysiuk, Z.J. Haas, Controlled gossiping in ad hoc networks, in: 2010 IEEE Wireless Communication and Networking Conference, 2010, pp. 1–6, http://dx.doi.org/10.1109/WCNC.2010.5506346.
[21] A. Khelil, P.J. Marrón, C. Becker, K. Rothermel, Hypergossiping: A generalized broadcast strategy for mobile ad hoc networks, Ad Hoc Netw. 5 (5) (2007) 531–546.
[22] B. Bako, E. Schoch, F. Kargl, M. Weber, Optimized position based gossiping in VANETs, in: 2008 IEEE 68th Vehicular Technology Conference, 2008, pp. 1–5, http://dx.doi.org/10.1109/VETECF.2008.449.
[23] P. Kyasanur, R.R. Choudhury, I. Gupta, Smart gossip: An adaptive gossip-based broadcasting service for sensor networks, in: 2006 IEEE International Conference on Mobile Ad Hoc and Sensor Systems, 2006, pp. 91–100, http://dx.doi.org/10.1109/MOBHOC.2006.278671.
[24] J. Leitao, J. Pereira, L. Rodrigues, Epidemic broadcast trees, in: 2007 IEEE 26th International Symposium on Reliable Distributed Systems, SRDS, 2007, pp. 301–310, http://dx.doi.org/10.1109/SRDS.2007.27.
[25] J. Cartigny, D. Simplot, Border node retransmission based probabilistic broadcast protocols in ad-hoc networks, in: Proc. of the 36th Annual Hawaii International Conference on System Sciences, 2003, 2003, http://dx.doi.org/10.1109/HICSS.2003.1174853, 10 pp.–.

[26] V. Drabkin, R. Friedman, G. Kliot, M. Segal, RAPID: Reliable probabilistic dissemination in wireless ad-hoc networks, in: 2007 IEEE 26th International Symposium on Reliable Distributed Systems, SRDS, 2007, pp. 13–22, http://dx.doi.org/10.1109/SRDS.2007.9.
[27] V. Drabkin, R. Friedman, G. Kliot, M. Segal, On reliable dissemination in wireless ad hoc networks, IEEE Trans. Dependable Secure Comput. 8 (6) (2010) 866–882.
[28] A.M. Hanashi, A. Siddique, I. Awan, M. Woodward, Performance evaluation of dynamic probabilistic flooding under different mobility models in MANETs, in: 2007 International Conference on Parallel and Distributed Systems, 2007, pp. 1–6, http://dx.doi.org/10.1109/ICPADS.2007.4447829.
[29] X.M. Zhang, E.B. Wang, J.J. Xia, D.K. Sung, A neighbor coverage-based probabilistic rebroadcast for reducing routing overhead in mobile ad hoc networks, IEEE Trans. Mob. Comput. 12 (3) (2012) 424–433.
[30] Q. Zhang, D. Agrawal, Analysis of leveled probabilistic routing in mobile ad hoc networks, in: 2004 IEEE International Conference on Communications (IEEE Cat. No.04CH37577), Vol. 7, 2004, pp. 3896–3900, http://dx.doi.org/10.1109/ICC.2004.1313282.
[31] H. Ling, D. Mossé, T. Znati, Coverage-based probabilistic forwarding in ad hoc routing, in: Proc. of the 14th International Conference on Computer Communications and Networks,, ICCCN, 2005, pp. 13–18.
[32] A. Mohammed, M. Ould-Khaoua, L.M. Mackenzie, J.-D. Abdulai, Dynamic probabilistic counter-based broadcasting in mobile ad hoc networks, in: 2009 2nd International Conference on Adaptive Science & Technology, ICAST, 2009, pp. 120–127.
[33] N. Wisitpongphan, O. Tonguz, Scalable broadcast strategies for ad hoc routing protocols, in: 2006 1st International Symposium on Wireless Pervasive Computing, 2006, p. 6, http://dx.doi.org/10.1109/ISWPC.2006.1613593.
[34] L. Zhou, NPPB: A broadcast scheme in dense VANETs, Inform. Technol. J. 9 (2) (2010) 247–256.
[35] H.Y. Adarbah, S. Ahmad, B. Arafeh, A. Duffy, Efficient broadcasting for route discovery in mobile ad-hoc networks, in: 2015 International Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS, 2015, pp. 1–7.
[36] P. Shete, R. Awale, Channel quality based adaptive gossip flooding mechanism for AODV, in: 2015 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, WiOpt, 2015, pp. 553–559.
[37] D.G. Reina, S. Toral, P. Johnson, F. Barrero, A survey on probabilistic broadcast schemes for wireless ad hoc networks, Ad Hoc Netw. 25 (2015) 263–292.
[38] C.-C. Shen, Z. Huang, C. Jaikaeo, Directional broadcast for mobile ad hoc networks with percolation theory, IEEE Trans. Mob. Comput. (4) (2006) 317–332.
[39] G. Koufoudakis, K. Oikonomou, G. Tsoumanis, Adapting probabilistic flooding in energy harvesting wireless sensor networks, J. Sens. Actuator Netw. 7 (3) (2018) URL https://www.mdpi.com/2224-2708/7/3/39.
[40] W. Evans, C. Kenyon, Y. Peres, L.J. Schulman, Broadcasting on trees and the ising model, Ann. Appl. Probab. (2000) 410–433.
[41] A. Makur, E. Mossel, Y. Polyanskiy, Broadcasting on two-dimensional regular grids, 2020, arXiv Preprint, URL https://arXiv:2010.01390.
[42] S.N. Chiu, D. Stoyan, W.S. Kendall, J. Mecke, Stochastic geometry and its applications, John Wiley & Sons, 2013.
[43] F. Baccelli, B. Błaszczyszyn, Stochastic Geometry and Wireless Networks, Volumes I and II, Now Publishers Inc, 2010.
[44] B.R. Vinay Kumar, N. Kashyap, R. Antony, The effect of introducing redundancy in a probabilistic forwarding protocol, in: 2018 Twenty Fourth National Conference on Communications, NCC, 2018, pp. 1–6, http://dx.doi.org/10.1109/NCC.2018.8599936.
[45] M. Penrose, Random Geometric Graphs, Oxford Univ. Press, 2003.
[46] J. Quintanilla, S. Torquato, R.M. Ziff, Efficient measurement of the percolation threshold for fully penetrable discs, J. Phys. A: Math. Gen. 33 (42) (2000) L399.
[47] F. Baccelli, B. Błaszczyszyn, M.K. Karray, Random Measures, Point Processes, and Stochastic Geometry, 2020, preprint, URL https://web.ma.utexas.edu/simons/wp-content/uploads/2020/02/BBK.pdf.
[48] J. Dall, M. Christensen, Random geometric graphs, Phys. Rev. E 66 (1) (2002) 016121.
[49] W.H.J. Debany, M.H. Linderman, Power Law Models For Connectivity of Random Geometric Graphs, AFRL/RI - Information Directorate, 2022, URL https://apps.dtic.mil/sti/citations/AD1187085.
[50] P. Villegas, T. Gili, A. Gabrielli, G. Caldarelli, Characterizing spatial point processes by percolation transitions, J. Stat. Mech. Theory Exp. 2022 (7) (2022) 073202.
[51] S. Ramanathan, E.L. Lloyd, Scheduling algorithms for multihop radio networks, IEEE/ACM Trans. Netw. 1 (2) (1993) 166–177.
[52] T.-T. Nguyen, T. Kim, T. Kim, A distributed TDMA scheduling algorithm using topological ordering for wireless sensor networks, IEEE Access 8 (2020) 145316–145331.
[53] S. Durrani, X. Zhou, A. Chandra, Effect of vehicle mobility on connectivity of vehicular ad hoc networks, in: 2010 IEEE 72nd Vehicular Technology Conference-Fall, 2010, pp. 1–5.
[54] Z. Khan, P. Fan, S. Fang, On the connectivity of vehicular ad hoc network under various mobility scenarios, IEEE Access 5 (2017) 22559–22565.
[55] C. Chen, X. Du, Q. Pei, Y. Jin, Connectivity analysis for free-flow traffic in VANETs: A statistical approach, Int. J. Distrib. Sens. Netw. 9 (10) (2013) 598946.
[56] A. Pramanik, B. Choudhury, T.S. Choudhury, W. Arif, J. Mehedi, Simulative study of random waypoint mobility model for mobile ad hoc networks, in: 2015 Global Conference on Communication Technologies, GCCT, 2015, pp. 112–116.
[57] E. Altman, F. De Pellegrini, Forward correction and fountain codes in delay-tolerant networks, IEEE/ACM Trans. Netw. 19 (1) (2010) 1–13.
[58] E. Altman, F. De Pellegrini, L. Sassatelli, Dynamic control of coding in delay tolerant networks, in: 2010 Proceedings IEEE INFOCOM, IEEE, 2010, pp. 1–5.
[59] L. Rashidi, D. Towsley, A. Mohseni-Kabir, A. Movaghar, On the performance analysis of epidemic routing in non-sparse delay tolerant networks, IEEE Trans. Mob. Comput. (2022).
[60] C. Rao, E. Modiano, Age of broadcast and collection in spatially distributed wireless networks, 2022, arXiv preprint arXiv:2212.05094.
[61] M. Wang, Y. Dong, Broadcast age of information in CSMA/CA based wireless networks, in: 2019 15th International Wireless Communications & Mobile Computing Conference, IWCMC, IEEE, 2019, pp. 1102–1107.
[62] C. Hirsch, B. Jahnel, E. Cali, Percolation and connection times in multi-scale dynamic networks, Stochastic Process. Appl. 151 (2022) 490–518.

B.R. Vinay Kumar received a B.E.(Hons.) in Electrical and Electronics Engineering and an M.Sc.(Hons.) in Mathematics from the Birla Institute of Technology and Sciences, Pilani, India. He obtained his Ph.D. from the Department of Electrical Communication Engineering at the Indian Institute of Science, Bengaluru. He is currently a post-doctoral researcher with the NEO team at INRIA Sophia Antipolis, France. His research interests lie in areas of stochastic geometry, community detection on graphs, percolation theory and random graphs.

Navin Kashyap received a B.Tech. degree in Electrical Engineering from the Indian Institute of Technology, Bombay, in 1995, an M.S. in Mathematics and a Ph.D. in Electrical Engineering from the University of Michigan, Ann Arbor, in 2001. After a two-year postdoctoral stint at the University of California, San Diego, he joined the faculty of the Department of Mathematics and Statistics at Queen's University, Kingston, Ontario, Canada, where he remained till December 2010. Thereafter, he has been with the Department of Electrical Communication Engineering at the Indian Institute of

Science, where he is currently a Professor. His research interests lie primarily in the application of combinatorial, algebraic, and probabilistic methods in information and coding theory.

D. Yogeshwaran's bachelors degree was at Sri Sathya Sai Institute of Higher Learning, Puttaparthi. He received M.S. in Mathematics from the Indian Institute of Science, Bangalore. He completed Ph.D. in Applied Probability from University of Pierre and Marie Curie (University of Paris 6). The Ph.D. thesis was focused on ordering of point processes and its applications. He was a post-doctoral researcher at the Faculty of Electrical Engineering, Technion, Haifa, Israel. Currently, he is a faculty at the Theoretical Statistics and Mathematics Unit at Indian Statistical Institute, Bangalore. His research interests are in stochastic geometry and random topology.