Università
degli Studi
di Padova

Dipartimento di Ingegneria dell'Informazione
Scuola di Dottorato di Ricerca in Ingegneria dell'Informazione
Indirizzo in Scienza e Tecnologia dell'Informazione
XXIV Ciclo — 2009–2011

Graph Models of Information Spreading in Wireless Networks

Direttore della Scuola: Ch.mo Prof. Matteo Bertocco

Coordinatore d'Indirizzo: Ch.mo Prof. Andrea Neviani

Supervisore: Ch.mo Prof. Geppino Pucci

Dottorando: Alberto Pettarin

Io stimo più il trovar un vero,
benché di cosa leggiera,
che 'l disputar lungamente
delle massime questioni
senza conseguir verità nissuna.

Galileo Galilei

Abstract

This thesis investigates the structural properties of graph models of wireless networks, where autonomous agents communicate using radios in order to accomplish a predefined task. Ad hoc, sensor, and vehicle networks are perhaps the most familiar examples. The goal of this thesis is the analytical characterization of information spreading in graph models of wireless networks, since this fundamental process is a primitive needed to accomplish more complex tasks.

The well-established graph-based approaches adopted when analyzing the behavior of "classical" distributed systems (e.g., P2P networks, computing clusters, etc.) fail to generalize to wireless networks, due to several causes, including the stricter physical constraints governing the operation of these systems (e.g., interference on the physical channel or scarce energy/computational resources) and the fact that the topology of the network might be unknown at design time or it might evolve over time.

This thesis shows how to tackle these problems by suitably defining and rigorously analyzing graph models and graph processes capturing the structure, evolution and operation of these networks. We present two reference scenarios.

In the first one we study a family of random graphs known as Bluetooth Topology, which closely model the connectivity of a network built by the device discovery phase of Bluetooth-like protocols, largely employed in wireless networks. Formally, the Bluetooth Topology generalizes the well-known Random Geometric Graph model, introducing a distributed pruning of the edge set. We investigate the expansion and the diameter of

these graphs, as they quantify the bandwidth and the latency of a wireless network. We give tight bounds on the expansion and, leveraging on these, we prove nearly-tight bounds on the diameter. Our results show that the Bluetooth Topology features the same global level of connectivity of the Random Geometric Graph but requires maintaining much fewer communication links.

Motivated by the recent and rapidly growing interest in mobile systems, in the second part of the thesis we turn our attention to the dynamics of information dissemination between agents performing random walks on a planar grid and communicating over short distances. This setting can also be employed to study phenomena like the spreading of a disease, where infections are the result of local interactions between agents. We prove that, for a sufficiently sparse system, the broadcast time of a message is independent from the transmission radius; indeed, we show that the broadcast time is dominated by the time needed for many agents to meet. Our findings nicely complements previous results that dealt with dense systems, where there is dependency from the transmission radius. Moreover, our analysis techniques extend to similar mobility-communication models, suggesting some interesting further research directions.

Sommario

Questa tesi studia le proprietà strutturali di alcuni modelli a grafo di reti di agenti autonomi che comunicano via radio per completare un prefissato compito. Reti ad hoc, di sensori e veicolari sono forse gli esempi più immediati. Lo scopo di questa tesi è caratterizzare la diffusione dell'informazione in questi modelli a grafo di reti wireless, considerata l'importanza di questo processo come primitiva fondamentale per realizzare protocolli più complessi.

Gli approcci basati su tecniche combinatorie adottati per l'analisi di sistemi distribuiti "classici", come le reti P2P o i cluster di calcolo, non possono essere estesi alle reti wireless, per varie ragioni: ad esempio a causa dei vincoli fisici che governano il funzionamento di questi sistemi (interferenza sul canale radio, scarse risorse energetiche/computazionali, ecc.) e per il fatto che la topologia della rete può essere ignota in fase di progettazione o può evolvere nel tempo.

Questa tesi suggerisce come sia possibile affrontare tali problemi tramite l'opportuna definizione e l'analisi rigorosa di modelli a grafo (o processi su grafi) che catturino l'evoluzione e il funzionamento delle reti wireless. Mostriamo come sia possibile applicare quest'approccio a due scenari di riferimento.

Innanzitutto studiamo una famiglia di grafi random nota come Bluetooth Topology, che ben rappresenta la connettività della rete creata dalla fase di device discovery in protocolli simili al Bluetooth, largamente utilizzati nelle reti wireless. Dal punto di vista formale, la Bluetooth Topology generalizza il ben noto modello Random Geometric Graph, introducendovi una selezione

distribuita degli archi. Studiamo l'espansione e il diametro di questi grafi, poiché quantificano la banda e la latenza della rete. Dimostriamo limiti stretti all'espansione e, sfruttando questa caratterizzazione, diamo dei limiti quasi stretti al diametro. I nostri risultati provano che la Bluetooth Topology presenta lo stesso livello globale di connettività del Random Geometric Graph, pur richiedendo molti meno link di comunicazione.

Motivati dal recente crescente interesse verso i sistemi mobili, nella seconda parte della tesi concentriamo la nostra attenzione sulle dinamiche di disseminazione dell'informazione tra agenti che effettuano random walk su una griglia planare e che comunicano su brevi distanze. Questo scenario può essere utilizzato per studiare fenomeni come la diffusione di malattie, dove le infezioni sono il risultato di interazioni locali tra gli agenti. Proviamo che, per un sistema sufficientemente sparso, il tempo di broadcast di un messaggio è indipendente dal raggio di trasmissione, dimostrando che esso è dominato dal tempo necessario affinché molti agenti si incontrino. I nostri risultati completano l'analisi, apparsa in lavori precedenti, di sistemi densi, dove viceversa vi è dipendenza del tempo di broadcast dal raggio di trasmissione. Inoltre le nostre tecniche di analisi possono essere estese a modelli di mobilità-comunicazione simili, suggerendo alcune interessanti linee di ulteriore ricerca.

Contents

Chapter 1

Introduction

Recent years witnessed an extraordinary evolution of technology, leading to the development of incredibly small electronic devices that have become familiar to everybody: laptops, portable music players, PDAs and smartphones, just to name a few. All these devices had a great impact on our lives, and we tend to consider them as “extensions” of our body and our cognitive functionality. Hence it is not surprising that a growing number of these devices have been equipped with radio technologies in order to let them have the capability of communicating with external services (e.g., mobile phones and WiFi-enabled laptops) or with other devices (e.g., a Bluetooth-enabled headset “connected” to a gaming console).

These aforementioned applications are just the tip of an iceberg: not-so-distant scenarios feature smart electronic appliances coordinating themselves to take care of our houses, intelligent cars that can predict and avoid safety hazards or traffic jams without relying on a fixed infrastructure, networks of sensors able to monitor sensitive areas or buildings, tracking of goods or raw materials in a manufacturing facility, and so on.

Many such applications heavily rely on the exchange of information among the nodes, which usually are small devices equipped with a cheap processor, few sensing probes, a radio antenna and a battery. A key characteristic of these networks is that their operation is decentralized and there is no single node coordinating the communication and hence managing the information flow inside the system. This approach offers several advantages, among the others, better fault tolerance, scalability, and deployability in a short time span.

Although many such systems have been deployed in the last few years, their development process has almost always been guided by an experimental and/or simulation approach. However, this approach is very time- and resource-consuming, and it is not very robust to changes in either the technology or the system specifications. Not surprisingly, many scientists have recently proposed abstract models for these networks, with the goal of establishing fundamental relationships between the model parameters and the resulting behavior of the system, independently from the particular technology adopted to build them.

The aim of this thesis is to investigate the structural properties of graph models of wireless networks at a foundational level. Specifically, the goal of this thesis is the analytical characterization of information spreading in wireless networks, since this basic process is a fundamental primitive needed to accomplish more complex tasks, like electing a leader, reaching consensus and computing a generic function of the nodes' internal states.

Unfortunately, the well-established graph-based approaches used to analyze the behavior of "classical" distributed systems (e.g., P2P networks, computing clusters, etc.) fail to generalize to wireless networks. This failure has several causes. From the operational point of view, these networks present stricter physical constraints than those encountered in wired networks, ranging from the interference originated by simultaneous accesses to the radio channel to the scarcity of energy resources. Hence, the need arises for interference- and energy-aware protocols, and more sophisticated techniques for their analysis. On a second level, the nodes in a wireless network are not usually endowed with large computing resources, and therefore the protocols they are supposed to run must be extremely simple and efficient. Finally, from a logical point of view, the topology of the network can be unknown at design time (e.g., sensor networks deployed by scattering its nodes in the environment) or it can even change over time (e.g., in mobile networks), and hence the precise pattern of interactions between nodes might be extremely hard to predict and exploit.

This thesis shows how to tackle these problems by suitably defining and rigorously analyzing graph models capturing the structure, evolution and operation of these networks. We present two reference scenarios. The first considers a family of random graphs representing the connectivity between devices which adopt a Bluetooth-like discovery protocol. The second scenario shows how messages spread in a system of mobile agents, under the constraint of short-range communications.

In the first part of the thesis, we study expansion and diameter of a family of subgraphs of the random geometric graph, known as Bluetooth Topology, which closely model the network built by the device discovery phase of Bluetooth-like protocols. By giving a characterization of the graph-theoretic concepts of expansion and diameter, we in fact quantify the bandwidth and latency characteristics of such networks. The main feature modeled by the Bluetooth Topology, denoted as $\mathcal{BT}\left(r(n), c(n)\right)$, is the small number $c(n)$ of links that each of the n devices (vertices) establishes with those located within its communication range $r(n)$. First, tight bounds are proved on the expansion of $\mathcal{BT}\left(r(n), c(n)\right)$ for the whole set of functions $r(n)$ and $c(n)$ for which connectivity has been established in previous works. Then, by leveraging on the expansion result, nearly-tight upper and lower bounds on the diameter of $\mathcal{BT}\left(r(n), c(n)\right)$ are derived. In particular, we show asymptotically tight bounds on the diameter when the communication range is near the minimum needed for connectivity, the typical scenario considered in practical applications.

Motivated by the recent and rapidly growing interest in mobile systems, in the second part of the thesis we turn our attention to the dynamics of information dissemination between agents moving independently on a planar domain. We formally model the problem by considering k mobile agents performing independent random walks on an n-node grid, which represents a discretization of the plane. At time 0, each agent is located at a random node of the grid and one agent has a rumor. The spread of the rumor is governed by a dynamic communication graph process $\{ G_t(r) \mid t \geq 0 \}$, where two agents are connected by an edge in $G_t(r)$ if and only if their distance at time t is within their transmission radius r. We study the broadcast time T_{B} of the system, which is the time it takes for all agents to know the rumor. We focus on the sparse case (below the percolation point $r_c \approx \sqrt{n/k}$) where, with high probability, no connected component in G_t has more than a logarithmic number of agents and the broadcast time is dominated by the time it takes for many independent random walks to meet one other. For a system below the percolation point, we give a tight characterization (up to logarithmic factors) of the broadcast time which quite surprisingly does not depend on the transmission radius, even when the latter is significantly larger than the mobility range in one step.

It is worth pointing out that our analysis techniques can also be applied to some related scenarios. For example, we can study the spreading of multiple messages or different interaction rules between agents, yielding to predator-prey models. Moreover, we can think of including communication barriers and mobility obstacles, further constraining the dynamics of the system to resemble an urban setting.

1.1 Thesis Overview

The remainder of this thesis is organized as follows.

Chapter 2 summarizes the peculiar characteristics of wireless networks and the challenges that they pose to the modeling and rigorous analysis of their behavior.

In Chapter 3 we consider the family of random graphs known as Bluetooth Topology. After illustrating the model definition and the results previously appeared in the literature, we proceed in characterizing its expansion and diameter, for those parameter ranges for which the graph is known to be connected. As a by-product of our analytic techniques, we derive a similar characterization of the properties of the Random Geometric Graph model, that is much simpler than the previous results. We also consider a dynamic version of the Bluetooth Topology, where the nodes can move over time but still rely on the neighbor selection protocol to establish communication links with other nodes.

Chapter 4 presents a model for mobile wireless networks, where moving agents perform random walks in the plane. We study the broadcast time of a message, that is, the time it takes for a rumor, originating at an agent in the system, to spread through the network. We give nearly-tight bounds (up to polylogarithmic factors) to the broadcast time in a sparse system, where the system density is below the percolation threshold. Unlike the previous results dealing with the dense case, we show that the broadcast time is asymptotically independent from the transmission radius. This result generalizes to multiple messages and to some other types of interaction between agents.

Finally, Chapter 5 contains some concluding observations and a summary of the contributions of this thesis, and it discusses some possible future research directions.

The main results of this thesis appeared in the open literature in [PPP09, PPP10, PPPU10, PPPU11]. Portion of this thesis is based on joint work with Andrea Pietracaprina, Geppino Pucci, and Eli Upfal.

Chapter 2

Wireless Networks

In this chapter we briefly survey the main characteristics, application fields and design challenges of wireless networks.

2.1 Historical Development

The first military projects about wireless networking date back to the 1970s and 1980s. The original intent was dealing with the lack of communication infrastructures on battlefields [LNT87, XHW03].

However, it was just since late 1990s that wireless (ad hoc) networks became a very active field of research. In fact, many circumstances contributed to the flourishing of new paradigms, models, protocols and eventually commercial products.

We can identify three main causes. At first, VLSI production processes favored the introduction of low-cost, portable devices capable of generating and processing data, such as laptops, PDAs, mobile phones, etc. Then, new emerging wireless technologies, such as BLUETOOTH, IEEE 802.11 and HYPERLAN 2 allowed the interconnection of such different units to exchange data and fostered the research in ultra-low power communications. Finally, MEMS-based devices and more efficient and manageable energy supplies made possible to develop more durable systems, with months or even years of operating autonomy.

2.2 Main Features

In this thesis we consider graph models for information spreading in wireless networks, defined as those distributed systems wherein autonomous devices can communicate via radio equipment [Per01].

The most evident property of these networks (sometimes also referred to as *Ad Hoc Networks*) consists in their ability to cope with the lack of regular, known infrastructures. For example, during rescue or emergency operations in a devastated area, nodes cannot be placed according to a scheme known *a priori*, since the orography of the area or the presence of obstacles might not permit it. In other contexts random placement occurs, for example in a smart dust, in intelligent fabrics or when sensor are literally thrown onto a geographic area. Additionally, the topology of the network can even vary during time, for example when devices are mobile (in this case, we speak about *Mobile Ad Hoc Networks* or MANETs).

These two features — unpredictable placements of nodes and varying topology — force algorithms and management policies to be "self-organizing", i.e. they should be able to modify duty-cycles, data routes, and energy management without a centralized supervisor. Hence, it is highly desirable that such a network could function even without possessing any information about the network topology and, more generally, absolute device location. In fact, especially in ad hoc and sensor networks, putting a GPS receiver or other localization units on a device could be neither practical nor affordable.

Each element of the network is supposed to operate autonomously, i.e. it performs its tasks quite "independently" from the presence of other devices in his neighborhood. Devices are usually equipped with some on-board processor that can perform local computation on the raw data, thus reducing the amount of network traffic by transmitting partially processed information. The latter is a crucial ability since, with state-of-the-art integrated circuits and over the entire system life, the energy required to receive and transmit data can be three or four orders of magnitude higher than the energy spent to perform local computation [PK00].

Sometimes data created at each network node has to be transmitted to gathering points, usually one or more *base stations* (also called *data sinks*) where it will be further processed or used. These endpoints can be inside the area where nodes are placed or in proximity of its borders and usually are directly connected to a wired backbone. This is the typical scenario for a sensor network that surveillance a certain environment, building or scientific experiment. Inter-node communication is performed through wireless links, but wired shortcuts can be issued to increase the overall system performances, giving birth to a Hybrid Sensor Network (see [SM05]).

More often, though, a wireless network is completely autonomous: data generated inside the network is used to coordinate the operation of the nodes of the network, in order to accomplish the task for which the system was designed. Such systems include vehicle or mobile ad hoc networks, where a device is attached to each vehicle and can communicate with other devices without relying upon a fixed infrastructure. The messages exchanged usually contain data about the vehicle (speed, distance covered since previous communication, etc.), and the environment (lighting conditions, presence of fog banks, etc.). These pieces of information can then be used to alert the driver about possible traffic jams or hazards on the route. In a not-so-distant future, we can imagine smart vehicles able to "drive themselves" by sensing the street conditions with cameras and low-range scanners for local decisions (e.g., avoiding a crossing pedestrian, maintaining the safety distance, etc.) and gathering traffic data from other vehicles for global decisions, like choosing the least crowded lane on a large avenue, taking into account temporary detours, or selecting the shortest route to reach the intended destination.

Radio technologies adopted in wireless networks include low-energy protocols such as BLUETOOTH, IEEE 802.11, HYPERLAN 2, and more recently IEEE 802.15 (including UWB and ZIGBEE) and IEEE 802.16 (WIMAX) protocols[1].

Typically, communication takes place with broadcast algorithms, such as *flooding* or *gossiping*, rather than adopting *address-based* routing policies [Per01, ASSC02].

Additionally, multi-hop strategies are preferred upon classical single-hop communication since they are expected to reduce the interference among nodes and with the environment (which is desirable in covert scenarios) and to consume less power. In fact, the transmitting power required by state-of-the-art antennas is proportional to the covered distance elevated to an exponent between two and four, depending on the directionality of the antenna, reflectivity of the terrain and the fading model for carrying medium [BC05].

Keeping the amount of network traffic low and adopting multi-hop transmission are not always sufficient to preserve battery life. Moreover, especially in ad hoc and sensor networks, the power source is irreplaceable and, once it becomes depleted, the single device becomes permanently unavailable. In order to prolong system life, management policies are power-aware and it is not uncommon that, unlike classical networks, Quality of Service (QoS) policies are not issued for the sake of energy efficiency.

Power consumption issues are not the unique concern in designing such systems. A careful identification of the task to be performed, usually the same for all the components, has to be pursued to produce a tailor-made project of final devices. Many constraints have to be satisfied, just to cite a few: minimize the form factor, reduce the manufacturing budget, reduce environmental interferences, enforce system robustness or security, guarantee the reconfigurability of the system.

Typical applications include:

- *Home/Business*: home automation, smart control of rooms and build-

[1] Specifications are available at: `http://www.bluetooth.com/Bluetooth/Technology/Building/Specifications/Default.htm`, `http://standards.ieee.org/getieee802/`, and `http://pda.etsi.org/`.

ings, industrial production processes [Int07], object/vehicle tracking [OS05];

- *Environmental*: fire, flood, and pollution detection [SM08], environmental protection [CAW06, APM07], precision agriculture [SCV+06];
- *Health*: telemonitoring, patients and personnel tracking inside a hospital [PLC05], drug administration;
- *Entertainment*: governance of visitors in a site [KW08], interactive museums, 3D reconstruction [WAK08];
- *Engineering*: infrastructure health monitoring [MZ04], pipe inspection [SNMT07], plant safety control;
- *Military*: accounting equipment and ammunition, battlefield surveillance, NBC attacks detection, damage assessment, intelligent clothing;
- *Surveillance and Control*: continuous sensing, event detection or identification and local control of actuators [ASSC02, CES04].

2.3 Analytic Models of Wireless Networks

Although many wireless networks have been deployed in the last few years, their development process has almost always been led by an experimental and/or simulation approach. In fact, nowadays, once the high-level functionalities of the system have been decided, network designers either build a small-scale prototype of the final system and test it or they build computer simulations with *ad hoc* tools and check that the behavior of the simulated system is aligned with the required one.

Unfortunately, both approaches are very time- and resource-consuming, and they also suffer from the serious drawback of being tightly coupled with the actual parameters of the technology adopted for the devices and radio protocols. Indeed, if the system specification is changed or updated, thus requiring re-engineering the network nodes and operation protocols, all the previous simulations and tests must be performed again. Clearly, this development cycle is not scalable in the long run. Moreover, little insight on the dynamics governing wireless networks is gained in this type of process.

Not surprisingly, several computer scientists have recently proposed mathematical models for wireless networks, with the goal of establishing fundamental relationships between the model parameters and the resulting behavior of the system, independently from the particular technology adopted to build them. The underlying assumption is that, despite the different technical details characterizing different electronic and radio technologies, some (relatively easily) recognizable interaction patterns dominate the behavior of such networks.

These efforts usually adopt a common approach. After recognizing these "abstract" interaction patterns, one usually want to define a suitable mathematical model capturing them, and then study its structure and behavior in order to understand the dynamics of the real-world system of interest.

Our thesis exploits this analytic approach to investigate the structural properties of graph models of wireless networks at a foundational level. Specifically, the goal of this thesis is the analytical characterization of information spreading in wireless networks. We chose to study information spreading processes because they are fundamental primitives needed to accomplish more complex tasks, like electing a leader, reaching consensus and computing a generic function of the nodes' internal states.

Unfortunately, modeling wireless networks is significantly harder than studying 'classical" distributed systems like P2P networks, computing clusters, desktop grids, etc. In fact, the well-established graph-based approaches used to analyze the behavior of these "classical" systems fail to generalize to wireless networks for several reasons.

From the operational point of view, wireless networks present stricter physical constraints than those encountered in wired networks, ranging from the interference originated by simultaneous accesses to the radio channel to the scarcity of energy resources. Hence, the need of interference- and energy-aware models and protocols, requiring more sophisticated techniques for their analysis.

On a second level, the nodes in a wireless network are not usually endowed with large computing resources, and therefore the protocols they are supposed to run must be extremely simple and efficient. Finally, from a logical point of view, the topology of the network can be unknown at design time: a typical example is a sensor network where nodes are deployed by scattering them in the environment. Moreover, the network topology can even change over time, as is the case for mobile networks, and hence the precise pattern of interactions between nodes might be extremely hard to predict and exploit.

This thesis shows how to tackle these problems by suitably defining and rigorously analyzing graph models capturing the structure, evolution and operation of wireless networks.

In the next chapters, we present two reference scenarios. The first one considers a family of random graphs representing the connectivity between devices adopting a Bluetooth-like discovery protocol, which is a common choice for ad hoc and sensor networks. The second scenario studies a sparse system of mobile agents that communicate over short distances; within this setting, we study how fast messages spread in the system.

Chapter 3

Bluetooth Topology

In this chapter we study the Bluetooth Topology, a family of graphs that captures the structure of wireless networks based on Bluetooth-like protocols. We present a characterization of its expansion and diameter for those parameter values for which connectivity has been established in previous works. The main technical results of this chapter show that the Bluetooth Topology guarantees roughly the same global level of connectivity of the Random Geometric Graph, despite being a possibly much sparser sub-graph of the latter.

The results presented in this chapter appeared in the literature in [PPP09, PPP10].

3.1 Introduction

Random graph models have been employed in the literature for the analytical characterization of topological properties of ad hoc wireless networks governed by a variety of network-formation protocols. One such case concerns networks based on the *Bluetooth* technology [WHC05, SZ06].

A Bluetooth network connects n devices, each endowed with a wireless transmitter/receiver able to communicate within a certain *visibility range*. The network is obtained by means of the following process: each device attempts to discover other devices contained within its visibility range and to establish reliable communication channels with them, in order to form a connected topology, called the *Bluetooth Topology*. Subsequently, a hierarchical organization is superimposed on this initial topology; Figure 3.1 illustrates the process. Since requiring each device to discover *all* of its neighbors is too time-consuming [BBMP04], the device discovery phase is terminated by a suitable time-out, hence only a limited number of neighbors are actually discovered.

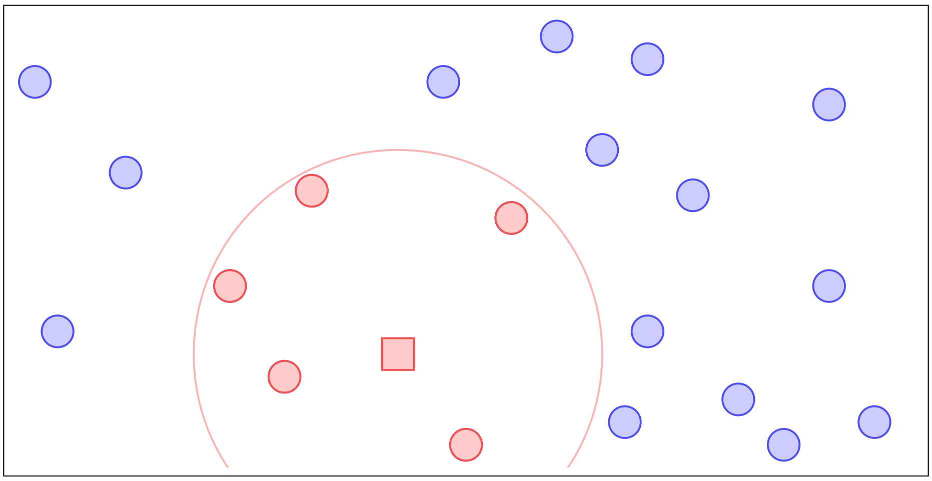
The following random graph model for the Bluetooth Topology has been proposed in [FMPP04] and subsequently generalized in [CNPP09]. The devices are represented by n nodes, whose coordinates are randomly chosen within the unit square $[0,1]^2$; each node selects $c(n)$ *neighbors* among all *visible nodes*, that is, among all nodes within the Euclidean distance $r(n)$, where $r(n)$ models the visibility range, which is assumed to be the same for all devices. The resulting graph, called $\mathcal{BT}(r(n), c(n))$, is the one where there is an undirected edge for each pair of neighbors. Note that such a graph is a subgraph of the well-known Random Geometric Graph [Pen03] in two dimensions. Experimental evidence shows that $\mathcal{BT}(r(n), c(n))$ is a good model for the Bluetooth Topology [FMPP04].

We remark that the $\mathcal{BT}(r(n), c(n))$ graph may be employed as a model for other real ad hoc network scenarios where nodes are constrained to maintain a small number of simultaneous connections, because of limited resources, both energetic and computational, or where establishing links to every visible node is too costly either in time or energy.

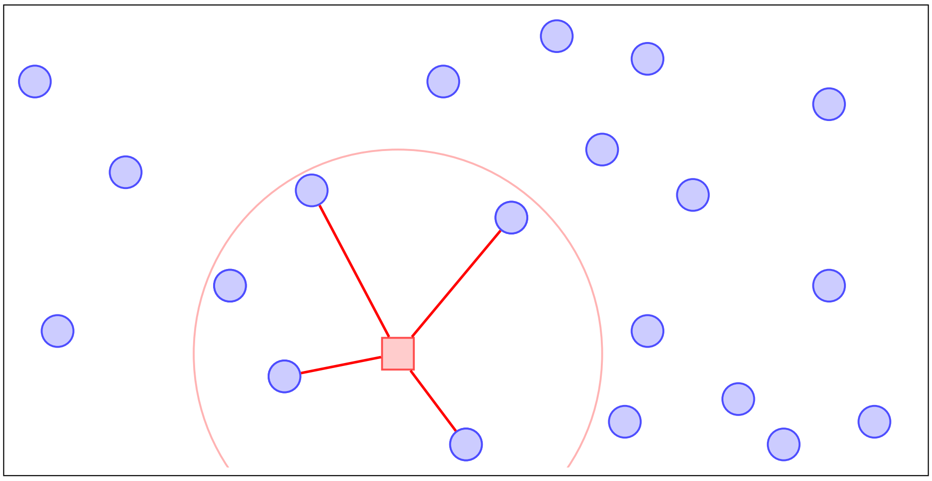
3.2 Previous Work

The topological properties of $\mathcal{BT}(r(n), c(n))$ have been investigated in a number of recent works.

In [PR04] the authors show that for any fixed constant $r > 0$ there exists a (large) constant c such that $\mathcal{BT}(r, c)$ is an expander with high probability. In [DJH$^+$05] it is proved that with high probability $\mathcal{BT}(r, c)$ is connected

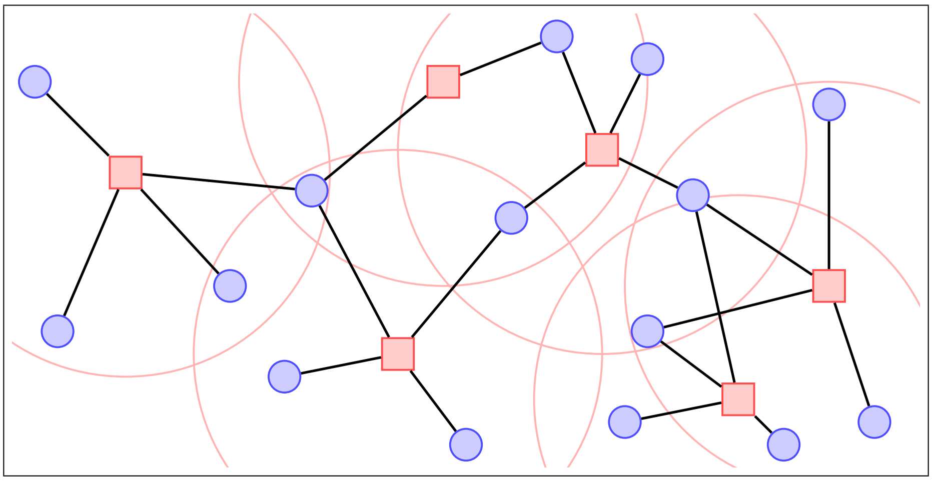


(a) A device, depicted as a square in the figure, attempts to discover its neighbors. The device discovery timeout limits the discovery range (circle in red).

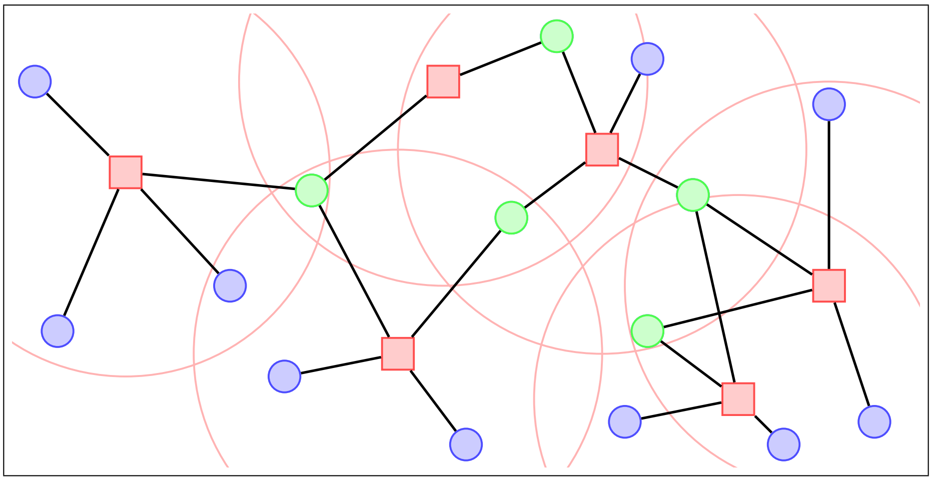


(b) The initiator (master node) establishes links with a subset of the discovered neighbors (slave nodes), forming a so-called Piconet.

Figure 3.1: Scatternet formation in a Bluetooth network.



(c) Slave nodes can belong to multiple Piconets.



(d) Slave nodes shared by multiple Piconets (green in the figure) behave like gateways between Piconets, giving birth to the global Scatternet.

Figure 3.1: Scatternet formation in a Bluetooth network.

for any fixed constant $r > 0$ and $c \geq 2$ whenever n becomes sufficiently large. Unfortunately, these results require a constant visibility range (where the domain is normalized to be the unit square) and this fact implies that every node can choose its neighbors among a constant fraction of all of the nodes in the system. Such an assumption becomes rapidly unfeasible as the number of devices grows large, and hence it might hinder the applicability of the above results to real-world networks.

To overcome the latter problem, a more general setting has been analyzed in [CNPP09], where it has been proved that $\mathcal{BT}\left(r(n), c(n)\right)$ stays connected, with high probability, also for vanishing values of $r(n)$ as $n \to \infty$, as long as each node selects a suitable number of neighbors. Precisely, if $r(n) = \Omega\left(\sqrt{\log n/n}\right)$, just allowing $c(n) = O\left(\log\left(1/r(n)\right)\right)$ neighbor selections per node ensures the connectivity of the resulting graph with high probability.

The above lower bound on $r(n)$ cannot be asymptotically improved: in fact, when $r(n) \leq \delta\sqrt{\log n/n}$, for some constant $0 < \delta < 1$, the visibility graph obtained connecting every node to *all* visible ones is disconnected with high probability [EJY05]. The latter graph is precisely the Random Geometric Graph $\mathcal{RGG}\left(r(n)\right)$ of [Pen03]. Since the edge-set of the Bluetooth Topology is a subset of the edge-set of the Random Geometric Graph (being the transmission radius the same for both models), $r(n) \geq \sqrt{\log n/n}$ is a necessary condition for the connectivity of $\mathcal{BT}\left(r(n), c(n)\right)$ as well.

A different approach to approximate the device discovery phase of Bluetooth networks has been proposed in [DHM$^+$07] under the name Blue Pleiades. This model is similar to the Bluetooth Topology, except for the neighbor selection protocol. In the Blue Pleiades model, agents choose to establish links with other devices in asynchronous rounds, where each attempt might fail if the polled node has already reached the maximum allowed degree $c^\star$. We note that also in this framework the transmission radius is a constant $0 < r < \sqrt{2}$, while the maximum degree of the agents is $c^\star = \Theta\left(1/r\right)$.

Very recently Broutin et al. [BDFL11] have studied the minimum number of neighbor choices needed to achieve a connected Bluetooth Topology when the transmission radius is $r(n) = \delta\sqrt{\log n/n}$, for some constant $\delta > 1$, and hence just a constant time greater than the minimum one needed to achieve connectivity. They show that $c(n) = \Theta\left(\sqrt{\log n/\log\log n}\right)$ choices yields a connected graph with high probability. Moreover, they prove that this threshold is sharp, that is, for a constant $\varepsilon \in (0, 2)$, the Bluetooth Topology with $c(n) = \sqrt{(2+\varepsilon)\log n/\log\log n}$ is connected with high probability, while for $c(n) = \sqrt{(2-\varepsilon)\log n/\log\log n}$ the resulting graph is not connected with high probability. We note that the authors of [BDFL11] do not characterize the optimal trade-off between $r(n)$ and $c(n)$ for which $\mathcal{BT}(r(n), c(n))$ is connected with high probability besides for the extremal case $r(n) = \Theta\left(\sqrt{\log n/n}\right)$. To the best of our knowledge, the latter problem is still open, and the best known result for non-constant and non-minimum transmission radii is the one by Crescenzi et al. [CNPP09]. Hence, we will adopt their characterization of $c(n)$ in function of $r(n)$ in our analysis of the expansion and diameter of the Bluetooth Topology.

3.3 Mathematical Model

In this section we formally define the Bluetooth Topology (see also Figure 3.2 for a pictorial description), illustrate the notation and recall some facts for later use.

Definition 3.3.1 (Bluetooth Topology)**.** *Given a positive integer n, a real-valued function $r(n) \to (0, \sqrt{2}]$ and a positive integer function $c(n)$, the Bluetooth Topology, denoted by $\mathcal{BT}(r(n), c(n))$, is the undirected random graph $G = (V_n, E_n)$, defined as follows.*

- *The node set V_n is a set of n points chosen uniformly and independently at random in $[0, 1]^2$.*
- *The edge set E_n is obtained through the following process: independently,*

each node selects a random subset of $c(n)$ neighbors among all nodes within distance $r(n)$ (all of them, if they are less than $c(n)$). An edge $\{u, v\} \in E_n$ exists if and only if u has selected v, or vice versa.

We say that two nodes *see each other* if they are within distance $r(n)$. In the next sections, we assume the following setting. Consider the standard tessellation of $[0, 1]^2$ into k^2 square *cells* of side $1/k$ where $k = \left\lceil \sqrt{5}/r(n) \right\rceil$. Consequently, any two nodes residing in the same or in two adjacent cells are at distance at most $r(n)$, hence they see each other. When the context is clear, with a slight abuse of notation, we identify a cell with the set of nodes residing therein. Figure 3.3 shows a graphical representation of the tessellation.

Recall that an event occurs *with high probability* if its probability is at least $1 - 1/\operatorname{poly}(n)$. Let $m = n/k^2 = \Theta\left(nr^2(n)\right)$ be the expected number of nodes residing in a cell. The following proposition will be used several times in the proofs of this chapter.

Proposition 3.3.2 ([CNPP09]). *Let $\alpha = 9/10$, $\beta = 11/10$. There exists a constant $\gamma_1 > 0$ such that for every $r(n) \geq \gamma_1 \sqrt{\log n/n}$ the following two events occur with high probability:*

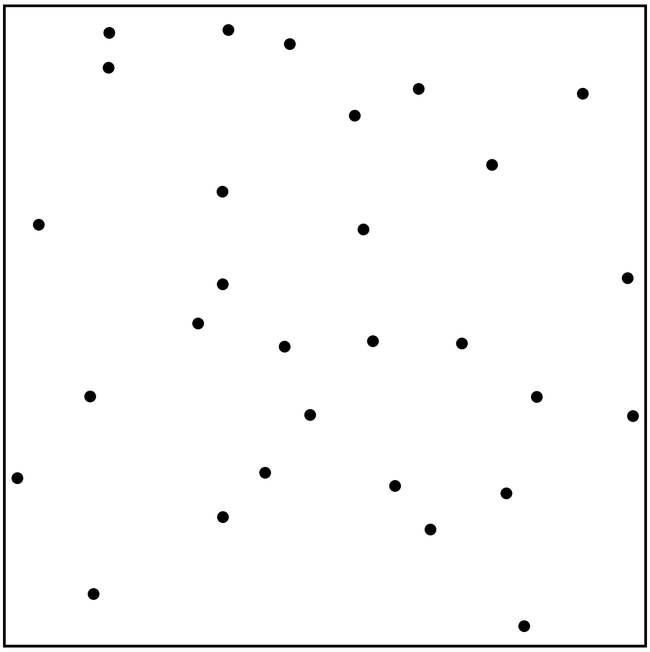
1. *every cell contains at least αm and at most βm nodes;*
2. *every node has at least $(\alpha/4)\pi n r^2(n)$ and at most $\beta \pi n r^2(n)$ nodes in its visibility range.*

Let $G = (V, E)$ be an undirected, connected graph. Below, we define the quantities at the core of our analysis.

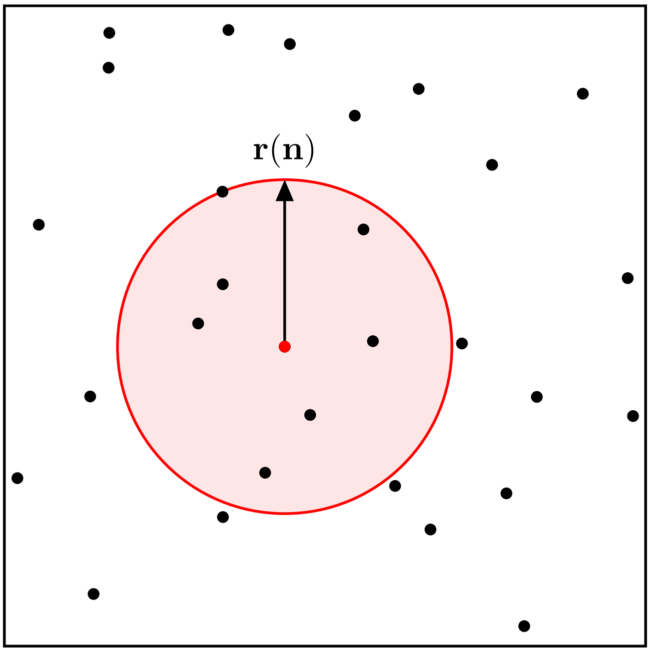
Definition 3.3.3 (Neighborhood). *Given a set of vertices $X \subseteq V$, its neighborhood is the set $\Gamma(X) = \{\, u \in V(G) : \exists e = \{u, v\} \in E(G), v \in X \,\}$.*

Definition 3.3.4 (Expansion). *The expansion of G is a function $\lambda(s)$, for $1 \leq s \leq |V|/2$, such that*

$$\lambda(s) = \min_{S \subseteq V : |S| = s} \frac{|\Gamma(S) - S|}{|S|}.$$

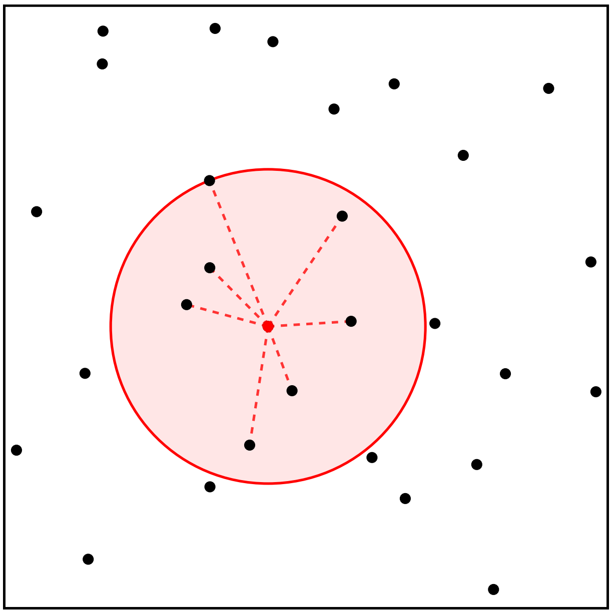


(a) In the unit square n nodes are spread independently and uniformly at random.

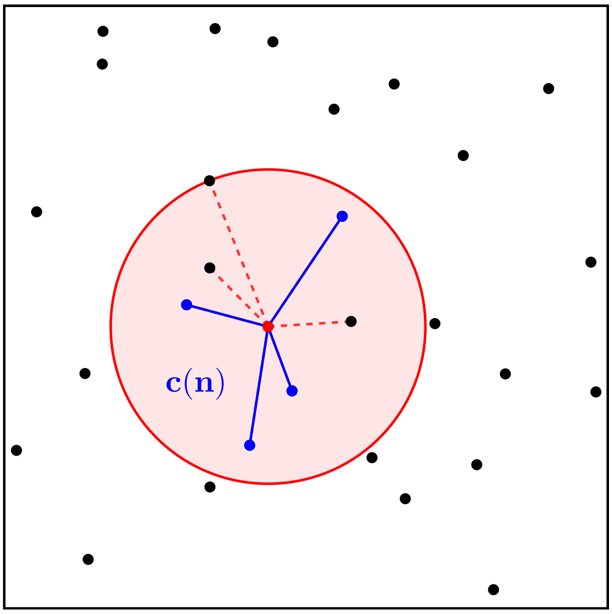


(b) Each node has a transmission range $r(n)$.

Figure 3.2: Bluetooth Topology model.

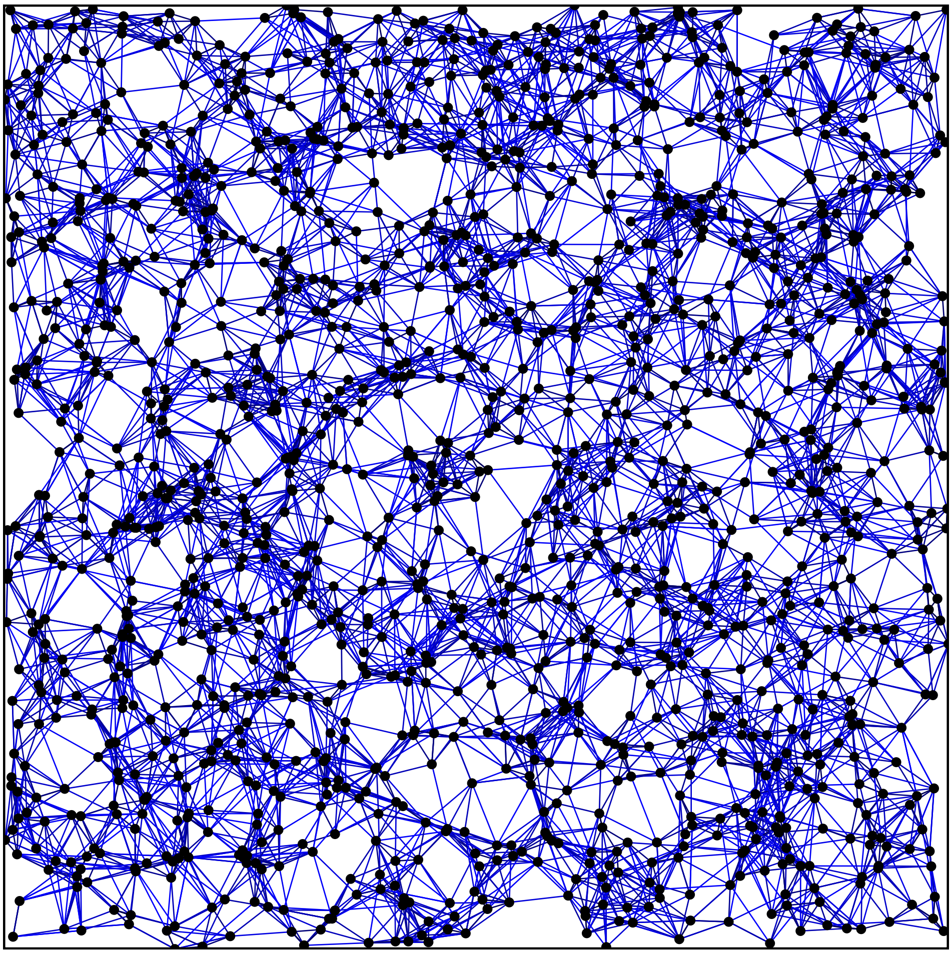


(c) Among all the nodes lying inside the disk of radius $r(n)$...



(d) ...each node selects uniformly at random $c(n)$ neighbors, establishing links with them.

Figure 3.2: Bluetooth Topology model.



(e) An instance of the Bluetooth Topology on $n = 1500$ nodes with transmission range $r(n) = 0.075$ and $c(n) = 5$ neighbor choices.

Figure 3.2: Bluetooth Topology model.

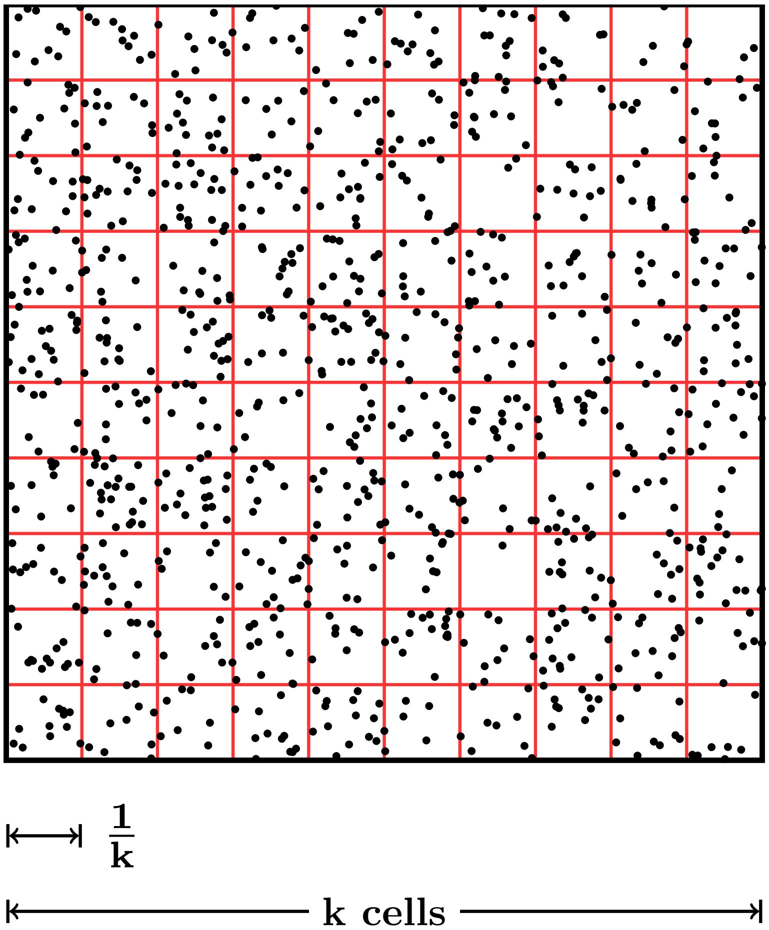


Figure 3.3: The tessellation of the unit square adopted in our analysis.

We remark that, in some works, the term "expansion" is used to refer to a "global" property of the graph, that is, the minimum value of the function $\lambda(s)$ taken over all subset sizes $1 \leq s \leq |V|/2$ (see [LR99]). In contrast, we offer a finer characterization of the expansion properties of $\mathcal{BT}(r(n), c(n))$ by proving explicit bounds on $\lambda(s)$ for all values of s.

Definition 3.3.5 (Diameter). *The diameter of a graph $G = (V, E)$, denoted as* $\mathrm{diam}(G)$*, is the maximum distance between any two nodes $u, v \in V$, where the distance between two nodes is the number of edges of a shortest path connecting them.*

Observe that, under any reasonable cost model for communication, the maximum latency to be expected of a point-to-point communication in a network is proportional to the diameter of its underlying topology.

In the next two sections we study the expansion and the diameter of $\mathcal{BT}(r(n), c(n))$ for those ranges of the parameters for which the connectivity is guaranteed by the results of [CNPP09], that is, $r(n) \geq \gamma_1 \sqrt{\log n/n}$ and $c(n) = \gamma_2 \log(1/r(n))$ for two suitable positive real constants γ_1 and γ_2.

3.4 Expansion

This section is devoted to the characterization of the expansion properties of $\mathcal{BT}(r(n), c(n))$.

First, we establish a lower bound on the expansion of this family of random graphs in Section 3.4.1. As an application of the latter result, in Section 3.4.2 we prove an upper bound on the flooding time of a message in a dynamic system closely related to the Bluetooth Topology. Then, Section 3.4.3 provides an upper bound on the expansion of $\mathcal{BT}(r(n), c(n))$, matching the above lower bound. Finally, Section 3.4.4 applies the analysis techniques developed for $\mathcal{BT}(r(n), c(n))$ to $\mathcal{RGG}(r(n))$, obtaining a finer characterization of its expansion than the ones previously known in the literature.

3.4.1 Lower Bound

The main result of this section is the following theorem.

Theorem 3.4.1. *Consider an instance of $\mathcal{BT}(r(n), c(n))$ with $r(n) \geq \gamma_1 \sqrt{\log n/n}$ and $c(n) = \gamma_2 \log(1/r(n))$, for two suitable positive constants γ_1 and γ_2. With high probability, for every integer s, $1 \leq s \leq n/2$, we have*

$$\lambda(s) = \begin{cases} \Omega\left(\min\{c(n), m/s\}\right) & \text{if} \quad s = O(m); \\ \Omega\left(\sqrt{m/s}\right) & \text{if} \quad s = \Omega(m). \end{cases}$$

The proof of Theorem 3.4.1 relies on three technical lemmas, which characterize the expansion of certain types of node subsets confined within a single cell. Consider a given subset of vertices S of size s. For any cell Q, we call the set $P = S \cap Q$ the *pocket* of S in Q.

Lemma 3.4.2 shows that a pocket is highly expanding either when the visibility range is $r(n) = O\left(n^{-\delta}\right)$ for some constant $\delta > 0$, or when we consider a sufficiently large pocket. Lemma 3.4.3 covers the case in which the visibility range is large but the pocket is of at most logarithmic size. Finally, for all radii, Lemma 3.4.4 assures that a pocket P containing a sufficiently large constant fraction of the nodes of its cell Q, expands roughly linearly into any adjacent cell Q' or in Q itself.

Lemma 3.4.2. *Let α' and ε' be two suitable positive constants, with $\alpha' \leq \min\{\varepsilon', 1/2\}$. Then, with high probability, for any cell Q and for every pocket $P \subseteq Q$ such that $c(n)|P| = \Omega(\log n)$ and $|P| \leq \alpha' m$, we have $|\Gamma(P) - P| \geq \varepsilon' \min\{c(n)|P|, m\}$.*

Proof. Fix a cell Q and a pocket P whose size $p = |P|$ satisfies the hypotheses of the lemma. We bound the probability that the entire neighborhood of P is contained in $P \cup T$, where T is a set of nodes not belonging to P with a certain (small) size t. For notational convenience, we abbreviate $c = c(n) = \gamma_2 \log(1/r(n))$ and introduce the following quantities:

- q is the number of nodes in Q;
- v is the total number of nodes visible by at least one node in Q;
- w is the minimum number of nodes visible by any node;
- w' is the maximum number of nodes visible by any node;
- z is the minimum number of nodes visible by all nodes in P.

Conditioning on the events of Proposition 3.3.2, we have that $q, v, w, w', z = \Theta(m)$.

Let $\mathcal{E}$ be the event that $|\Gamma(P) - P| \leq t$. We can bound the probability of the event $\mathcal{E}$:

$$\begin{aligned}
\Pr[\mathcal{E}] &\le \binom{q}{p}\binom{v}{t}\left(\frac{\binom{t+p}{c}}{\binom{w}{c}}\right)^{p}\left(\frac{\binom{w'-p}{c}}{\binom{w'}{c}}\right)^{z-(t+p)}\\
&\le \left(\frac{eq}{p}\right)^{p}\left(\frac{ev}{t}\right)^{t}\left(\frac{t+p}{w}\right)^{cp}\left(\frac{w'-p}{w'}\right)^{c(z-(t+p))}\\
&\le \left(\frac{eq}{p}\right)^{p}\left(\frac{ev}{t}\right)^{t}\left(\frac{t+p}{w}\right)^{cp} e^{-\frac{cp}{w'}(z-(t+p))}.
\end{aligned}$$

We distinguish between two cases, depending on the value of p.

Case 1: $1 \le p \le m/c$. Let $t = \varepsilon' cp$. We rewrite the bound on $\Pr[\mathcal{E}]$ as

$$\Pr[\mathcal{E}] \le \left(\left(\frac{eqc}{cp}\right)^{1/c}\left(\frac{ev}{\varepsilon' cp}\right)^{\varepsilon'}\left(\frac{\varepsilon' cp}{aw}\right)\right)^{cp},$$

where a is a positive constant, since $p = O(t)$ and $(z-(t+p))/w' = \Theta(1)$. By regrouping the factors, we obtain:

$$\Pr[\mathcal{E}] \le \left(\frac{c^{1/c}}{a\varepsilon'^{\varepsilon'}}\frac{(eq)^{1/c}(ev)^{\varepsilon'}}{w}(cp)^{1-\varepsilon'-1/c}\varepsilon'\right)^{cp} < \frac{1}{n^3},$$

where the last inequality holds for a sufficiently large γ_2 in $c = \gamma_2 \log(1/r(n))$, and for a sufficiently small ε', since $cp = \Omega(\log n)$. The claim follows by invoking the union bound over the $O(n)$ cells and the $O(n)$ choices of $p = |P|$.

Case 2: $m/c < p \le \alpha' m$. Note that in this case $cp > m$, whence we set $t = \varepsilon' m$. We rewrite the upper bound on $\Pr[\mathcal{E}]$ as

$$\begin{aligned}
\Pr[\mathcal{E}] &\le \left(\frac{eq}{p}\right)^{p}\left(\frac{ev}{\varepsilon' m}\right)^{\varepsilon' m}\left(\frac{\varepsilon' m + p}{aw}\right)^{cp}\\
&\le \left(\left(\frac{eq}{p}\right)^{1/c}\left(\frac{ev}{\varepsilon' m}\right)^{\varepsilon' m/(cp)}\left(\frac{\varepsilon' m + p}{aw}\right)\right)^{cp}.
\end{aligned}$$

The first and the second factor of the latter bound are bounded by a constant, for a suitable choice of c and ε'. By our choice of α', letting ε' be a sufficiently small value, we can make the product of the three factors at most a constant less than 1, so that $\Pr[\mathcal{E}] < 1/n^3$ since $cp = \Omega(\log n)$. The claim then follows by applying the union bound as done for Case 1. $\square$

Lemma 3.4.3. *Let $r(n) = \Omega\left(n^{-1/8}\right)$, and $c(n) \geq 3$. With high probability, for any cell Q and for every pocket $P \subseteq Q$, with $|P| < \log n$, we have $|\Gamma(P)| > \frac{1}{3}c(n)\,|P|$.*

Proof. Let $c = c(n) = \gamma_2 \log(1/r(n)) \geq 3$ and let $p = |P|$. Since $|\Gamma(P)| \geq c(n)$ for all pockets P, then the lemma is trivial when $p < 3$. Suppose now that $p \geq 3$, and fix a subset T of t possible neighbors of P such that $t = \frac{1}{3}cp \geq c$. There are at most $\binom{\beta m}{p}$ ways of choosing P and at most $\binom{24\beta m}{\frac{1}{3}cp}$ ways of choosing T. Since a node in Q can choose its neighbors from at least 3 cells, the probability that all cp neighbor choices from P are within nodes of T is at most

$$\left(\frac{\binom{t}{c}}{\binom{3\alpha m}{c}}\right)^p = \left(\frac{t!\,(3\alpha m - c)!}{(3\alpha m)!\,(t-c)!}\right)^p \leq \left(\frac{\frac{1}{3}cp}{3\alpha m}\right)^{cp}.$$

Define $\mathcal{E}$ to be the event that all of the sets $P \subseteq Q$, with $|P| < \log n$, choose a number of neighbors in T less than $\frac{1}{3}c\,|P|$. Then,

$$\begin{aligned}
\Pr[\mathcal{E}] &\leq \sum_{p=1}^{\log n - 1} \binom{\beta m}{p}\binom{24\beta m}{\frac{1}{3}cp}\left(\frac{\frac{1}{3}cp}{3\alpha m}\right)^{cp} \\
&\leq \sum_{p=1}^{\log n - 1} \left(\frac{e\beta m}{p}\right)^p \left(\frac{e24\beta m}{\frac{1}{3}cp}\right)^{\frac{1}{3}cp} \left(\frac{\frac{1}{3}cp}{3\alpha m}\right)^{cp} \\
&\leq \sum_{p=1}^{\log n - 1} \left(\frac{24e^2\beta^2 m^2}{\frac{1}{3}cp^2}\right)^{\frac{1}{3}cp} \left(\frac{\frac{1}{3}cp}{3\alpha m}\right)^{cp} \\
&\leq \sum_{p=1}^{\log n - 1} \left(\tau \frac{c^{2/3}p^{1/3}}{m^{1/3}}\right)^{cp} = O\left(\left(\frac{\log n}{m^{1/3}}\right)^{c}\right)
\end{aligned}$$

where $\tau = \left(72e^2\beta^2\right)^{1/3} / (9\alpha)$ is a positive constant. Since $r(n) = \Omega\left(n^{-1/8}\right)$, $\Pr[\mathcal{E}] < 1/n^3$ for a convenient choice of γ_2 in the definition of c. The result follows from the union bound over $k^2 = O\left(n^2\right)$ cells. □

Lemma 3.4.4. *Let α' be the constant defined in Lemma 3.4.2. With high probability, for any pair of cells Q and Q', with either $Q' = Q$ or Q' adjacent to Q, and for every pocket $P \subseteq Q$, with $|P| = \alpha' m$, we have $|\Gamma(P) \cap Q'| \geq (1/2 + \epsilon'')\,m$, for a suitable constant $\epsilon'' > 0$.*

Proof. Let $c = c(n) = \gamma_2 \log(1/r(n))$. To ease the argument, we suppose that nodes choose their neighbors by picking uniformly at random a node within distance $r(n)$ for $c(n)$ times. Clearly, this process is stochastically dominated by the actual one since in the latter a node cannot be chosen multiple times.

Consider a particular pair of adjacent cells Q and Q' (or $Q' = Q$) and a subset $P \subseteq Q$ of size $p = |P| = \alpha' m$. First we show that with high probability a constant fraction of the cp neighbor choices from nodes of P goes toward nodes of Q'. Then, conditioning on this event, we prove the lemma allowing a suitably large constant γ_2 in the definition of $c(n)$.

The probability q that a node $u \in P$ chooses a neighbor inside Q' is at least $q \geq \alpha m/(\beta\pi n r(n)^2)$, since there are at least αm nodes inside Q' and at most $\beta\pi n r(n)^2$ nodes within distance $r(n)$ from u. By our definition of $m = n/k^2$, we have that $q \geq \alpha/(5\pi\beta) = \Theta(1)$. Let L denote the number of neighbor choices from nodes of P which select a node inside Q'. Note that L does not count the number of distinct nodes of Q' reached from P but it is instead the number of edges from P to Q' (counted with repetitions). By the linearity of expectation, we have $\mathrm{E}[L] = qpc = \Omega(\log n)$ since $p = \alpha' m = \Omega(\log n)$. Applying the standard Chernoff bound A.1.3, we get

$$\Pr\left[L < \frac{1}{2}\mathrm{E}[L]\right] \leq e^{-\frac{1}{8}qpc}.$$

Since there are no more than $5n$ pairs of cells (Q, Q') to be accounted for, and at most $\binom{\beta m}{p} \leq \left(\frac{e\beta m}{\alpha' m}\right)^p$ ways of choosing a pocket P of size $p = \alpha' m$ inside Q, we can conclude that, for any such pair and any such pocket P,

$$\Pr\left[L < \frac{1}{2}qpc\right] \leq 5n\left(\frac{e\beta}{\alpha'}\right)^p e^{-\frac{1}{8}qpc} \leq \frac{1}{n},$$

where the last inequality holds by allowing a sufficiently large constant γ_2 in the definition of $c(n)$. The above inequality implies that, with high probability, $L \geq \sigma m$ for any constant $\sigma > 0$. In the following, we will make use of such a fact for a specific value σ to be determined by the analysis.

Since the neighbor choices are independent and uniform, we can model the neighbor selection process as an instance of the classical balls-and-bins problem, where L balls are thrown inside $b = |Q'|$ bins. Let Z_v be the indicator variable of the event "node $v \in Q'$ was not chosen by any node of P" (i.e., it is an empty bin) and let $Z = \sum_{v \in Q'} Z_v$ denote the total number of nodes of Q' which were not chosen by any node of P (i.e., the total number of empty bins). Conditioning on the event $L \geq \sigma m$, by the linearity of expectation we have

$$\mathrm{E}[Z] \leq b\left(1 - \frac{1}{b}\right)^{\sigma m} \leq b e^{-\frac{\sigma m}{b}} \leq \beta e^{-\frac{\sigma}{\beta}} m,$$

since $b \leq \beta m$. The Z_v variables are not independent but they satisfy the Lipschitz condition with bound 1. Therefore, we can apply the Azuma-Hoeffding concentration bound A.1.6, obtaining

$$\Pr\left[Z > 2\beta e^{-\frac{\sigma}{\beta}} m\right] \leq \Pr\left[Z > \mathrm{E}[Z] + \beta e^{-\frac{\sigma}{\beta}} m\right] \leq e^{-\frac{2\left(\beta e^{-\frac{\sigma}{\beta}} m\right)^2}{\sigma m}} \leq \frac{1}{n^2},$$

where the last inequality holds for any value of σ, provided that we choose a sufficiently large constant $\gamma_1 > 0$ in the definition of $r(n)$, since $m = \Theta\left(n r^2(n)\right)$.

Therefore, the number of distinct nodes of Q' reached by nodes of P is

$$\left|\Gamma(P) \cap Q'\right| \geq (\alpha m - 2\beta e^{-\frac{\sigma}{\beta}} m) = (1/2 + \epsilon'')m$$

by letting $\sigma = \beta \log \frac{2\beta}{\alpha - 1/2 - \epsilon''}$, and this can be achieved by selecting a suitably large constant γ_2 in the definition of $c(n)$. Invoking the union bound over $O(n)$ pairs of adjacent cells concludes the proof. □

We are now ready to prove the main result of this section.

Theorem 3.4.1. Throughout the proof, we condition on the events stated in Proposition 3.3.2 and in the three previous lemmas. We define $\bar{\varepsilon} = \min\{\varepsilon', \varepsilon''/2, 1/3\}$ where $\varepsilon', \varepsilon''$ are the constants appearing in the statements of Lemma 3.4.2 and Lemma 3.4.4, respectively, and let $\alpha' \leq \min\{\varepsilon', 1/2\}$ be a suitable small constant (thus, consistent with the constraint posed by the aforementioned lemmas). Consider an arbitrary set S of s vertices of $\mathcal{BT}(r(n), c(n))$, with $1 \leq s \leq n/2$. We classify the cells according to the size of the pockets of S that they contain: namely, a cell Q such that $Q \cap S \neq \emptyset$ is said to be *red* if it contains at least $\alpha' m$ nodes of S, and *blue* otherwise; please see Figure 3.4.

Two cases are possible: either a majority of nodes of S resides in red cells or a majority of nodes of S resides in blue cells.

In the first case, $N_R > s/2$ nodes of S belong to red cells. We further subdivide the red cells into two groups, depending on the number of nodes of S they contain. We say that a red cell is "dark red" if it contains at least $(1/2 + \varepsilon''/2)m$ nodes of S; otherwise we call it "light red".

Suppose that a majority of the N_R nodes resides inside dark red cells. There are at most $n/((1+\varepsilon'')m)$ dark red cells since $s \leq n/2$, and thus at least $\varepsilon'' n/((1+\varepsilon'')m)$ cells are not dark red. Hence, by well-known topological properties of two-dimensional meshes [BP86, DP09], there are $\Omega\left(\sqrt{s/m}\right)$ disjoint pairs of adjacent cells (Q, Q'), where Q is dark red and Q' is not. Consider one such pair: applying Lemma 3.4.4 by taking a pocket $P \subseteq Q \cap S$ of size $|P| = \alpha' m$, we have that

$$\left|\Gamma(Q) \cap Q' - S\right| \geq \left|\Gamma(P) \cap Q' - S\right| \geq \varepsilon'' m/2 \geq \bar{\varepsilon} m,$$

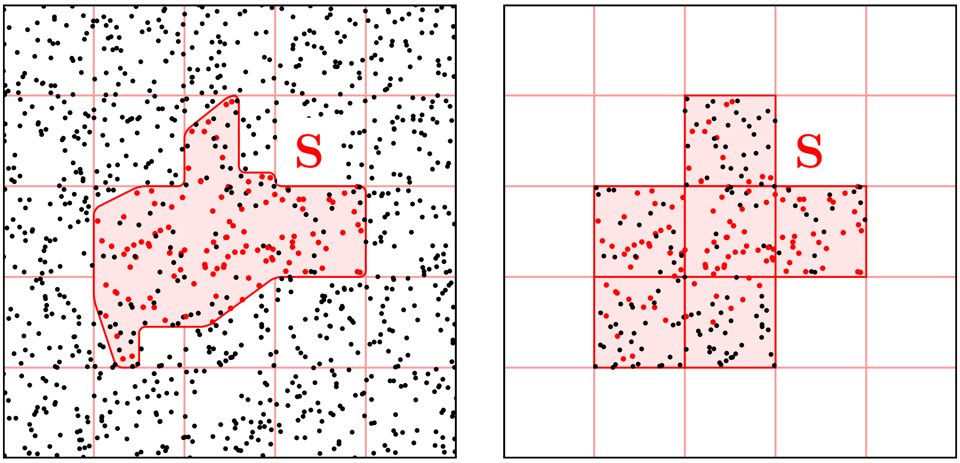
since, by definition, Q' contains less than $(1/2 + \varepsilon''/2)m$ nodes of S. Summing over the $\Omega\left(\sqrt{s/m}\right)$ disjoint pairs of cells (Q, Q'), we get $\lambda(s) = \Omega\left(\sqrt{m/s}\right)$.

The other subcase where a majority of the N_R nodes resides inside light red cells is easier to deal with since we can just consider the expansion of light red cells into themselves. Mimicking the application of Lemma 3.4.4 as in the previous subcase, we have that, for any light red cell Q,

$$|\Gamma(Q) \cap Q - S| \geq \varepsilon'' m/2 \geq \bar{\varepsilon} m.$$

Since there are at least $s/((1+\epsilon'')m)$ light red cells, we immediately obtain $\lambda(s) = \Omega(1) = \Omega\left(\sqrt{m/s}\right)$, which is the correct bound since $s = \Omega(m)$ in this case.

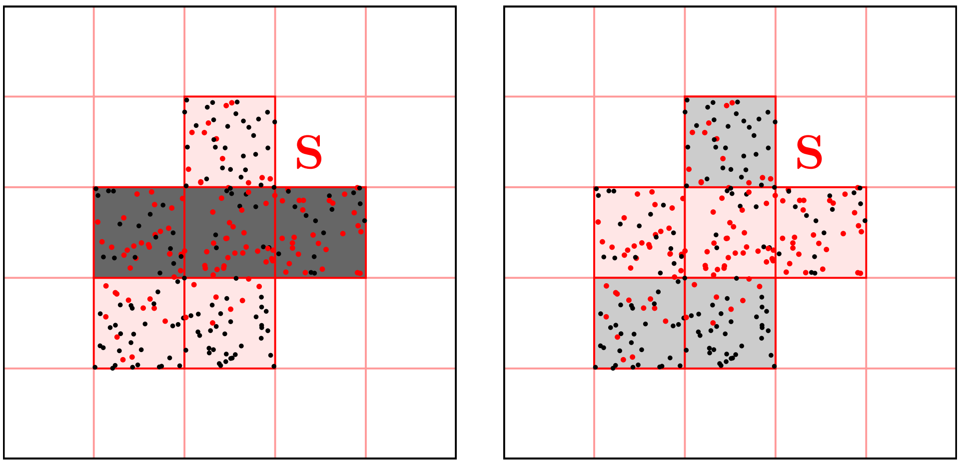
To analyze the second case, where at least $s/2$ nodes of S belong to blue cells, we resort to a proof strategy inspired by the one employed in [PR04].



(a) Consider an arbitrary subset of vertices S, red in figure.

(b) We consider only those cells containing at least one node of S.

Figure 3.4: Lower bound on the expansion.



(c) Red cells (dark gray in figure) contain at least $\alpha' m$ nodes of S.

(d) Blue cells (light gray in figure) contain less than $\alpha' m$ nodes of S.

Figure 3.4: Lower bound on the expansion.

Referring to the tessellation of $[0,1]^2$ into k^2 cells, let us index the cells as Q_{ij}, with $1 \leq i, j \leq k$. Define the *sector* $\mathcal{S}_{ij}$ of a cell Q_{ij} as

$$\mathcal{S}_{ij} = \bigcup_{\substack{\max\{i-6,1\} \leq x \leq \min\{i+6,k\} \\ \max\{j-6,1\} \leq y \leq \min\{j+6,k\}}} Q_{xy}.$$

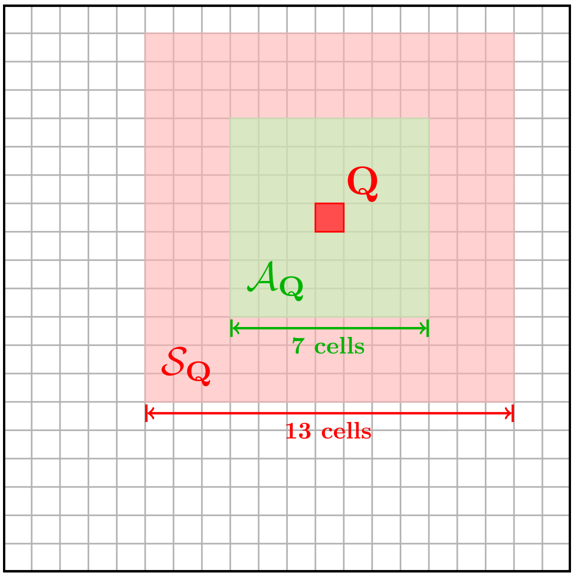
The *active area* $\mathcal{A}_{ij}$ of sector $\mathcal{S}_{ij}$ is defined as

$$\mathcal{A}_{ij} = \bigcup_{\substack{\max\{i-3,1\} \leq x \leq \min\{i+3,k\} \\ \max\{j-3,1\} \leq y \leq \min\{j+3,k\}}} Q_{xy}.$$

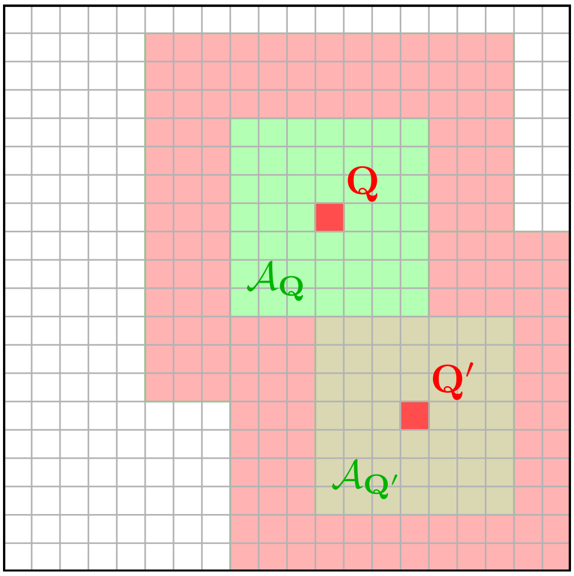
Cell Q_{ij} is called the *center* of both sector $\mathcal{S}_{ij}$ and its active area $\mathcal{A}_{ij}$. Note that the neighborhood of the pocket $P_{ij} = Q_{ij} \cap S$ is entirely contained in $\mathcal{A}_{ij}$ and that the definition of a sector ensures that given two sectors $\mathcal{S}_{ij}$ and $\mathcal{S}_{i'j'}$, with $Q_{i'j'} \cap \mathcal{S}_{ij} = \emptyset$, their active areas are non-overlapping. Figure 3.5 illustrates these concepts.

Let B be the set of at least $s/2$ nodes of S belonging to blue cells. To estimate the expansion of S, we first execute a greedy procedure, which selects a number of blue cells which are centers of non-overlapping active areas, and then obtain a lower bound on the expansion by adding up the contributions related to these selected cells. The selection of the centers is obtained via the following marking strategy. Initially all of the blue cells are unmarked. Then, iteratively, the center of the next active area is selected as the unmarked blue cell Q containing the largest pocket of S, and all of the unmarked cells of the sector centered at Q are marked. The procedure terminates as soon as every blue cell becomes marked. The procedure is described by the pseudo code in Algorithm CenterSelection, where sets I and U maintain, respectively, the indices of the selected centers and the indices of unmarked cells, and subroutine LargestPocket(U) returns the pair (i, j) corresponding to the unmarked cell with the largest pocket (ties broken arbitrarily). Figure 3.6 depicts an execution of the algorithm.

Let $\langle c_1, c_2, \ldots, c_w \rangle$ be the list of w centers picked by CenterSelection, where $c_t = (i_t, j_t)$ was chosen at the t-th iteration of the **while** loop. Let $p_t = |P_{c_t}|$, and let b_t be the number of nodes residing in unmarked blue

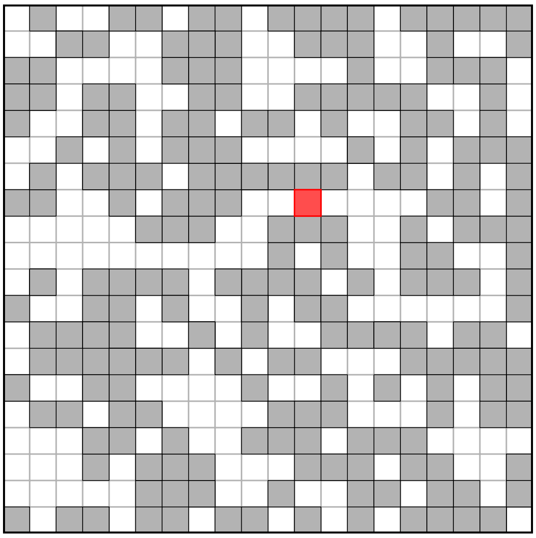


(a) Sector $\mathcal{S}_Q$ centered at cell Q, and its active area $\mathcal{A}_Q$.

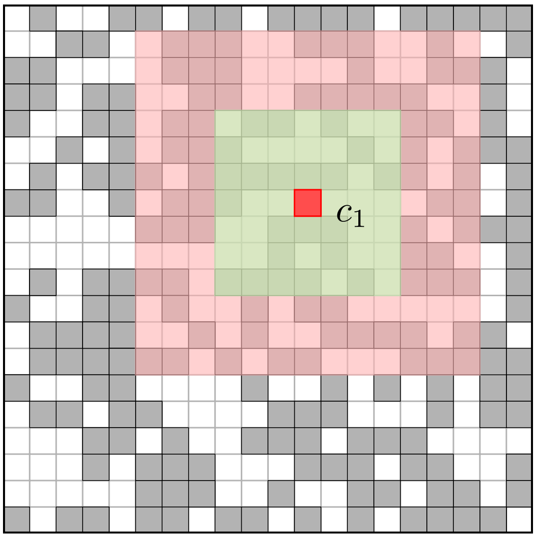


(b) The non-overlapping property of active areas belonging to different sectors.

Figure 3.5: Illustration of the concepts of sector and active area.

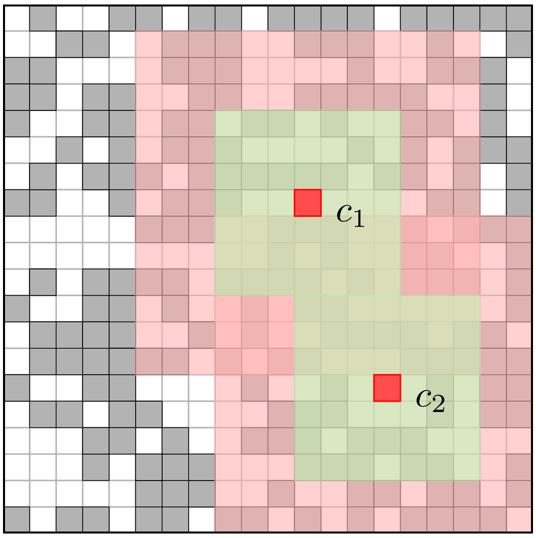


(a) Consider only the blue cells (light gray in figure) and select the cell Q (red in figure) containing the largest number of nodes of S to be the first center picked by Algorithm CenterSelection.

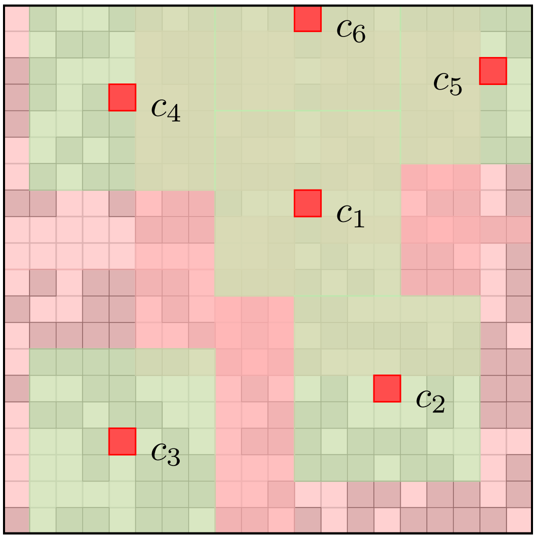


(b) All the unmarked blue cells in the sector of center c_1 are now marked.

Figure 3.6: The execution of Algorithm CenterSelection.



(c) The next center c_2 contains the largest number of nodes of S among all the unmarked blue cells. By the non-overlapping property, $\mathcal{A}_{c_2}$ does not overlap with $\mathcal{A}_{c_1}$.



(d) The greedy center selection procedure continues until not blue cell is left unmarked. Note that all the active areas do not overlap.

Figure 3.6: The execution of Algorithm CenterSelection.

Algorithm 1 CenterSelection

```
I ← ∅
U ← {(i,j) : Q_ij is a blue cell}
while U ≠ ∅ do
    (i,j) ← LargestPocket(U)
    I ← I ∪ (i,j)
    for all Q_xy ∈ S_ij do
        U ← U − {(x,y)}
    end for
end while
```

cells of $\mathcal{S}_{c_t}$ at the beginning of iteration t. Clearly, we have that $\sum_{t=1}^{w} b_t = |B|$ and, by the greedy choice of the centers, $b_t \leq 169p_t$.

In order to lower bound the expansion of S, we proceed as follows. For each t, with $1 \leq t \leq w$, we determine a suitable set of nodes $N_t \subseteq \Gamma(S)$, which belong to blue cells of the active area $\mathcal{A}_{c_t}$. We distinguish between two different cases. First, consider the case $s < \alpha' m$, which implies that no red cell exists. Let $N_t = \Gamma(P_{c_t}) - P_{c_t}$ and observe that by Lemmas 3.4.2 and 3.4.3, $|N_t| \geq \bar{\varepsilon} \min\{c(n)p_t, m\}$. Note that the N_t's are all disjoint, but the sum of their sizes does not immediately yield a lower bound on $|\Gamma(S) - S|$, since each set N_t may itself contain nodes of S, which have to be subtracted from the overall count. Specifically, the number of *external neighbors of S* (i.e., nodes of $\Gamma(S) - S$) accounted for by the N_t's is

$$\left(\sum_{t=1}^{w} |N_t|\right) - |B| = \sum_{t=1}^{w} (|N_t| - b_t) \geq \sum_{t=1}^{w} (|N_t| - 169p_t).$$

Since $p_t < \alpha' m$ and $|N_t| \geq \bar{\varepsilon} \min\{c(n)p_t, m\}$, then for a sufficiently large choice of γ_2 in $c(n) = \gamma_2 \log(1/r(n))$ and a sufficiently small value of α', we have that $|N_t| - 169p_t \geq \mu |N_t|$ for a certain constant $\mu > 0$. Hence,

$$\sum_{t=1}^{w} (|N_t| - 169p_t) = \Omega\left(\sum_{t=1}^{w} \bar{\varepsilon} \min\{c(n)p_t, m\}\right) = \Omega(\min\{c(n)s, m\}),$$

and the theorem follows.

Consider now the case $s \geq \alpha' m$. Observe that $\sum_{t=1}^{w} |N_t| = \Omega(|B|) = \Omega(s)$, and note that it is sufficient to show that the number of external neighbors of S is $\Omega\left(\sum_{t=1}^{w} |N_t|\right)$. Partition the index set $I = \{1, 2, \ldots, t\}$ into two disjoint subsets B_1 and B_2, such that $t \in B_1$ if $\mathcal{A}_{c_t}$ contains no red cells, and $t \in B_2$ otherwise. Suppose that $\sum_{t \in B_2} |N_t| \geq \tau \sum_{t \in B_1} |N_t|$, for a suitable positive constant τ which will be specified later. For each $t \in B_2$ the set N_t contains $(1/2 + \varepsilon'')m$ nodes, and at least $(1/2 + \varepsilon'' - \alpha')m$ of these nodes are external neighbors of S. Hence, the total number of external neighbors of S is at least

$$\sum_{t \in B_2} (1/2 + \varepsilon'' - \alpha')m = \frac{1/2 + \varepsilon'' - \alpha'}{1/2 + \varepsilon''} \sum_{t \in B_2} |N_t| \geq \frac{1/2 + \varepsilon'' - \alpha'}{1/2 + \varepsilon''} \frac{\tau}{1 + \tau} \sum_{t=1}^{w} |N_t|,$$

and the theorem follows. Finally, if $\sum_{t \in B_2} |N_t| < \tau \sum_{t \in B_1} |N_t|$, the number of external neighbors of S accounted for by the nodes in the N_t's is

$$\begin{aligned}
\left(\sum_{t=1}^{w} |N_t|\right) - |B| &= \sum_{t \in B_1} (|N_t| - 169 p_t) + \sum_{t \in B_2} (|N_t| - 169 p_t) \\
&\geq \sum_{t \in B_1} \mu |N_t| + \sum_{t \in B_2} ((1/2 + \varepsilon'')m - 169\alpha' m) \\
&> \sum_{t \in B_1} \mu |N_t| - \sum_{t \in B_1} \left(\frac{169\alpha'}{1/2 + \varepsilon''} - 1\right) \tau |N_t|.
\end{aligned}$$

By fixing τ such that $((169\alpha'/(1/2 + \varepsilon'')) - 1)\tau = \mu/2$, we get

$$\sum_{t \in B_1} \mu |N_t| - \sum_{t \in B_1} \left(\frac{169\alpha'}{1/2 + \varepsilon''} - 1\right) \tau |N_t| = \frac{\mu}{2} \sum_{t \in B_1} |N_t| = \Omega\left(\sum_{t=1}^{w} |N_t|\right),$$

and the theorem follows. □

3.4.2 Flooding Time of the Stationary Dynamic Bluetooth Topology

We now turn our attention to a dynamic version of the $\mathcal{BT}(r(n), c(n))$, where we allow both the positions of the nodes and the set of links to evolve over time. This framework has been adopted by Clementi *et al.* as a model for mobile agents in [CMPS09], where the primitive of interest is *flooding*, that is, the spreading of information from one agent to all the others. Building on the relation between graph expansion and flooding established in [CMPS09], in this section we study the flooding time of a system of mobile agents where the communication links are established through the neighbor selection protocol of the Bluetooth topology.

Suppose that we have n agents moving along the nodes of a square grid of side 1 and edge-length[1] $\epsilon > 0$. Time is discrete and, in a time step, each agent moves to a grid node chosen uniformly at random among the grid nodes within the Euclidean distance $0 < \rho \leq \sqrt{2}$ from its current position. The parameter ρ can be interpreted as the maximum velocity that any agent can achieve. We suppose that all the moves are synchronous and independent. After reaching its new position, each agent establishes communication links with $c(n)$ other agents, chosen uniformly at random among those within the Euclidean distance $r(n)$, or with all of them if they are less than $c(n)$. For $t \in \mathbb{N}$, let G_t be the graph induced by the positions of the agents and the links established at time t. We can formally describe the evolution of the system resulting from this stochastic process as the sequence of graphs $\mathcal{G}(n, \rho, r(n), c(n), \epsilon) = \{G_t : t \in \mathbb{N}\}$, which we call *Dynamic Bluetooth Topology*.

In the above dynamic scenario, we aim at upper bounding the flooding time of a message, that is, the minimum number of time steps required to inform all the agents in the system of a message originating from a source agent. When a link connecting an informed agent to an uninformed one is established, the latter becomes informed of the message. It is easily seen that $\mathcal{G}(n, \rho, r(n), c(n), \epsilon)$ constitutes a Markov chain, hence it is a *Markovian evolving graph* according to the Definition 2.1 of [CMPS09]. Moreover, when the positions of the nodes in G_0 are chosen according to the stationary

[1] The edge-length ϵ can be made arbitrarily small, and it is introduced only to guarantee the technical condition that the state space of the Markov chain describing the system is finite. In fact, ϵ only affects the constants, hence it does not appear in the results expressed in asymptotic notation.

distribution (in this case, $\mathcal{G}$ is referred to as *stationary Markovian evolving graph*), the flooding time can be bounded from above based on graph expansion, as established by the following proposition.

Proposition 3.4.5 ([CMPS09, Corollary 2.5]). *Let $\mathcal{M} = \{G_t : t \in \mathbb{N}\}$ be a stationary Markovian evolving graph. Assume a decreasing sequence $k_1 \geq k_2 \geq \ldots \geq k_{n/2}$ of positive real numbers exists such that, with probability at least $1 - 1/n^4$, G_t has expansion $\lambda(s) \geq k_s$, for every $s = 1, 2, \ldots, n/2$. Then the flooding time in $\mathcal{M}$ is with high probability*

$$O\left(\sum_{s=1}^{n/2} \frac{1}{s \log(1 + k_s)}\right).$$

In order to apply the above result to the Dynamic Bluetooth Topology $\mathcal{G}(n, \rho, r(n), c(n), \epsilon)$, we need to lower bound the expansion of each constituent graph G_t. Unfortunately, we cannot directly apply the result of the previous section, since the agents in G_t are not distributed uniformly in $[0,1]^2$, as it is the case for $\mathcal{BT}\,(r(n), c(n))$. However, the stationary distribution of the positions of the agents is *quasi-uniform*, meaning that, with respect to a suitably defined tessellation of the domain into non-overlapping cells of equal size, the number of agents in any two cells differs by at most a constant factor, which is sufficient to obtain a lower bound on the expansion. In what follows we consider a stationary Dynamic Bluetooth Topology $\mathcal{G}_{\mathrm{STAT}}(n, \rho, r(n), c(n), \epsilon)$, where the positions of the agents in G_0 are chosen randomly according to the stationary distribution, and tessellate the domain into k^2 cells where $k = \lceil \sqrt{5}/r(n) \rceil$, and $m = n/k^2$. The following proposition is analogous to Proposition 3.3.2.

Proposition 3.4.6 (Quasi-Uniformity). *Consider a stationary Dynamic Bluetooth Topology $\mathcal{G}_{\mathrm{STAT}}(n, \rho, r(n), c(n), \epsilon)$ with $r(n) \geq \gamma_1 \sqrt{\log n/n}$ and $c(n) = \gamma_2 \log(1/r(n))$, for two suitable positive constants γ_1 and γ_2 and arbitrary positive constants ρ and ϵ. Then, with probability $1 - 1/n^5$, in the stationary Dynamic Bluetooth Topology $\mathcal{G}_{\mathrm{STAT}}(n, \rho, r(n), c(n), \epsilon)$, the number of agents $N_Q(t)$ residing in cell Q at time t satisfies*

$$m/\mu \leq N_Q(t) \leq \mu m,$$

for each cell Q, $0 \leq t < n$, and for some constant $\mu > 0$.

Proof. Let $\pi(x)$ be the probability that an agent is located at grid node x in the stationary distribution. As noted in [CMPS09], for any two positions x and y, there exists a constant $\sigma > 0$, depending on ϵ and ρ, such that $1/\sigma \leq \pi(x)/\pi(y) \leq \sigma$. Let $N_Q(t)$ denote the number of agents residing inside cell Q at time t. Since the size of each cell is $1/k^2$ and the distribution of the agents' position is stationary, we have that $\mathrm{E}\left[N_Q(t)\right] \geq m/\sigma$ and $\mathrm{E}\left[N_Q(t)\right] \leq \sigma m$. Also, the agents move independently, so we can apply the Chernoff bound on $N_Q(t)$, obtaining that $\Pr\left[N_Q(t) \leq m/(2\sigma)\right] \leq 1/n^7$, and $\Pr\left[N_Q(t) \geq 3\sigma m/2\right] \leq 1/n^7$, for a sufficiently large constant γ_1. The proof follows by applying the union bound over $O(n)$ cells and the n time instants and setting $\mu = 2\sigma$. □

The quasi-uniform distribution of the agents described in Proposition 3.4.6 allows us to characterize the expansion of each snapshot G_t of the process, as stated in the next theorem.

Theorem 3.4.7. *Consider a stationary Dynamic Bluetooth Topology $\mathcal{G}_{\mathrm{STAT}}(n, \rho, r(n), c(n), \epsilon)$, with $r(n) \geq \gamma_1\sqrt{\log n/n}$ and $c(n) = \gamma_2 \log(1/r(n))$, for two suitable positive constants γ_1 and γ_2 and arbitrary positive constants ρ and ϵ. Then there exist positive constants $\delta_1, \delta_2, \delta_3, \alpha$ such that, with probability $1 - 1/n^4$, each graph $G_t \in \mathcal{G}_{\mathrm{STAT}}(n, \rho, r(n), c(n), \epsilon)$ has expansion*

$$
\begin{aligned}
\lambda(s) &\geq \delta_1 c(n) && \text{if} \quad 1 \leq s \leq m/c(n), \\
\lambda(s) &\geq \delta_2 m/s && \text{if} \quad m/c(n) < s \leq \alpha m, \\
\lambda(s) &\geq \delta_3\sqrt{m/s} && \text{if} \quad s > \alpha m,
\end{aligned}
$$

for all $0 \leq t < n$.

Proof. The proof follows exactly the same steps of the proof of Theorem 3.4.1 for $\mathcal{BT}\left(r(n), c(n)\right)$. In fact, it suffices to observe that the result of Proposition 3.4.6 enables us to prove lemmas analogous to Lemmas 3.4.2, 3.4.3 and

3.4.4, where the constants involved in the pocket expansion become suitable functions of μ. □

Theorem 3.4.8. *Consider a stationary Dynamic Bluetooth Topology $\mathcal{G}_{\mathrm{STAT}}(n, \rho, r(n), c(n), \epsilon)$, with $r(n) \geq \gamma_1 \sqrt{\log n / n}$ and $c(n) = \gamma_2 \log(1/r(n))$, for two suitable positive constants γ_1 and γ_2 and arbitrary positive constants ρ and ϵ. Then, with high probability, the flooding time in $\mathcal{G}_{\mathrm{STAT}}(n, \rho, r(n), c(n), \epsilon)$ is*

$$T_{\mathrm{F}} = O\left(\frac{1}{r(n)} + \log n\right).$$

Proof. By plugging in the lower bounds on the expansion stated in Theorem 3.4.7 in the formula given in Proposition 3.4.5, we have that

$$T_{\mathrm{F}} = O\left(\sum_{s=1}^{m/c(n)} \frac{1}{s\log(1+\delta_1 c(n))} + \sum_{s=m/c(n)+1}^{\alpha m} \frac{1}{s\log(1+\delta_2 m/s)} + \sum_{s=\alpha m+1}^{n/2} \frac{1}{s\log(1+\delta_3\sqrt{m/s})}\right).$$

We evaluate the three summations separately. The first summation easily yields

$$\sum_{s=1}^{m/c(n)} \frac{1}{s\log(1+\delta_1 c(n))} = \frac{1}{\log(1+\delta_1 c(n))} H\left(\frac{m}{c(n)}\right) = O(\log n).$$

For the second summation, since we can always presume that $\delta_2 \leq \alpha$, we obtain

$$\begin{aligned}\sum_{s=m/c(n)+1}^{\alpha m} \frac{1}{s\log(1+\delta_2 m/s)} &\leq 2\frac{\alpha}{\delta_2} \sum_{s=m/c(n)+1}^{\alpha m} \frac{1}{s\log(1+\delta_2 m/s)} \frac{\delta_2 m}{\alpha m + s} \\ &\leq 2\frac{\alpha}{\delta_2} \int_{m/c(n)}^{\alpha m} \frac{1}{x\log(1+\delta_2 m/x)} \frac{\delta_2 m}{\delta_2 m + x} \mathrm{d}x \\ &= O(\log\log c(n)).\end{aligned}$$

Since $c(n) = O(\log n)$, we have that the second summation is bounded by

$O(\log\log\log n)$. Finally, for the third summation we obtain

$$\begin{aligned}\sum_{s=\alpha m+1}^{n/2} \frac{1}{s\log(1+\delta_3\sqrt{m/s})} &\le \sum_{s=\alpha m+1}^{n/2} \frac{1+\delta_3/\sqrt{\alpha}}{\delta_3\sqrt{m}}\frac{1}{\sqrt{s}} \\ &\le 2\frac{1+\delta_3/\sqrt{\alpha}}{\delta_3}\frac{1}{\sqrt{m}}\int_{\alpha m}^{n/2}\frac{1}{\sqrt{x}}\mathrm{d}x \\ &= O\left(\frac{1}{r(n)}\right).\end{aligned}$$

Summing the three contributions concludes the proof of the theorem. □

3.4.3 Upper Bound

In this section we prove that the lower bound on the expansion of $\mathcal{BT}(r(n), c(n))$ established by Theorem 3.4.1 is asymptotically tight.

Theorem 3.4.9. *Consider an instance of $\mathcal{BT}(r(n), c(n))$ with $r(n) \ge \gamma_1\sqrt{\log n/n}$ and $c(n) = \gamma_2 \log(1/r(n))$, for two suitable positive constants γ_1 and γ_2. With high probability, for every integer s, $1 \le s \le n/2$, there exists a set of vertices S of size s whose expansion is*

$$\lambda(s) = \begin{cases} O(\min\{c(n), m/s\}) & \text{if} \quad s = O(m); \\ O\left(\sqrt{m/s}\right) & \text{if} \quad s = \Omega(m). \end{cases}$$

Proof. We fix the constant γ_1 so that Proposition 3.3.2 holds. If $s \le \alpha m$, we can choose any subset S of the nodes in a single corner cell, so that a total of at most $13\beta m$ nodes are visible from S. Hence, $\lambda(s) = O(m/s)$. Consider a list $\langle v_1, v_2, \dots, v_n \rangle$ of the vertices of V, sorted by non-decreasing degree. If we take $S = \{v_1, v_2, \dots, v_s\}$, then we are guaranteed that the sum of the degrees of all nodes in S is no greater than $2c(n)s$, or otherwise the sum of the degrees of the n nodes would exceed $2c(n)n$, which is impossible. Combining the two cases above proves the theorem for the case $s \le \alpha m$.

Consider now the case $s > \alpha m$ and choose a set S which occupies an approximately square area of $\Theta(s/m)$ cells in a corner of $[0,1]^2$. Since only the nodes in $O\left(\sqrt{s/m}\right)$ cells are visible from S, we have that $\lambda(s) = O\left(\sqrt{ms}/s\right) = O\left(\sqrt{m/s}\right)$, and the theorem follows. □

We remark that the tight bounds on the expansion of $\mathcal{BT}(r(n), c(n))$ provided by Theorems 3.4.1 and 3.4.9 extend the results in [PR04] from the case $r(n) = \Theta(1)$ to arbitrary values of $r(n)$ that guarantee the connectivity of the graph. Note also that if we consider the minimum expansion $\lambda = \min_{1 \le s \le n/2} \lambda(s)$, we obtain that $\lambda = \Theta(r(n))$ for the Bluetooth Topology.

3.4.4 Expansion of the Random Geometric Graph

The analysis performed in the previous section for $\mathcal{BT}(r(n), c(n))$ can also be applied to $\mathcal{RGG}(r(n))$ with simpler technical arguments, due to the absence of the neighbor selection procedure. Indeed, the following theorem establishes the asymptotic order of the expansion $\lambda(s)$ of a Random Geometric Graph, for all values of s.

Theorem 3.4.10. *Consider an instance of $\mathcal{RGG}(r(n))$ with $r(n) \ge \gamma_1 \sqrt{\log n / n}$, for a suitable constant γ_1. With high probability, for every integer s, $1 \le s \le n/2$, we have*

$$\lambda(s) = \begin{cases} \Theta(m/s) & \text{if} \quad s = O(m); \\ \Theta\left(\sqrt{m/s}\right) & \text{if} \quad s = \Omega(m). \end{cases}$$

Proof. We fix the constant γ_1 so that Proposition 3.3.2 holds. For the lower bound, consider a subset $S \subseteq V$, $|S| = s$, with $1 \le s \le n/2$. If $s \le \alpha m$, recall that Proposition 3.3.2 implies that any node $u \in S$ has at least $(\alpha/4)\pi n r^2(n)$ neighbors in $RGG(r(n))$. Since $(\alpha/4)\pi n r^2(n) - s = \Omega(m)$, we have that $\lambda(s) = \Omega(m/s)$. If $s > \alpha m$, similarly to the proof of Theorem 3.4.1 we say that a cell Q such that $Q \cap S \ne \emptyset$ is *red* if it contains at least $(3/4)\alpha m$ nodes of S, and *blue* otherwise. Two subcases are possible: either a majority of nodes of S resides in red cells or a majority of nodes of S resides in blue cells. In the first case, since $s \le n/2$, the number N_R of red cells satisfies $N_R \le (2/(3\alpha))n/m$ and therefore the number of non-red cells is at least $n/m - N_R = \Omega(n/m)$. Therefore, at least $\Omega\left(\sqrt{n/m}\right) = \Omega\left(\sqrt{s/m}\right)$ red cells are adjacent to a non-red cell [BP86], and each of these (disjoint) pairs contributes $\Omega(m)$ nodes to the expansion of S, yielding $\lambda(s) = \Omega\left(\sqrt{m/s}\right)$.

On the other hand, if a majority of nodes of S resides in blue cells, there are at least $s/((3/4)\alpha m)$ blue cells, and each contributes at least $(\alpha/4)m$ nodes to the expansion of S, since any node residing in that cell connects to every other node residing in the same cell. In this second case, we get $\lambda(s) \geq 1/3 = \Omega\left(\sqrt{m/s}\right)$ since $s > \alpha m$.

In order to complete the proof, we derive a matching upper bound. If $s \leq \alpha m$, we can pick the set S entirely contained in a single corner cell. Since the number of nodes visible from S is bounded by $13\beta m$, we have that $\lambda(s) = O\left(m/s\right)$. On the other hand, if $s > \alpha m$, consider the "densest" set S as in the proof of Theorem 3.4.9. We immediately conclude that the neighborhood of S has at most $O\left(m\sqrt{s/m}\right)$ nodes, and thus $\lambda(s) = O\left(\sqrt{m/s}\right)$. □

Quite surprisingly, Theorems 3.4.1, 3.4.9, and 3.4.10 imply that the expansion $\lambda(s)$ of $\mathcal{BT}\left(r(n), c(n)\right)$ is, within a constant factor, equal to the expansion of $\mathcal{RGG}\left(r(n)\right)$, as soon as we consider a set of $s = \Omega\left(m/c(n)\right)$ vertices, although $\mathcal{BT}\left(r(n), c(n)\right)$ is a (possibly very sparse) subgraph of $\mathcal{RGG}\left(r(n)\right)$.

3.5 Diameter

In this section we provide upper and lower bounds on the diameter of $\mathcal{BT}\left(r(n), c(n)\right)$ by leveraging on the expansion result of Section 3.4. Specifically, the upper bound relies on the following lemma, which relates diameter and expansion.

Lemma 3.5.1. *Given a connected undirected graph $G = (V, E)$ with n nodes and expansion $\lambda(s)$, for $1 \leq s \leq n/2$, consider the following recurrence:*

$$\begin{aligned} N_0 &= 1 \\ N_i &= \left(1 + \lambda(N_{i-1})\right) N_{i-1}. \end{aligned} \tag{3.1}$$

Define $i^\star$ as the smallest index such that $N_{i^\star} > n/2$. Then, $\mathrm{diam}(G) \leq 2i^\star$.

Proof. Let $d = \mathrm{diam}(G)$ and let u and v be two nodes at distance d in G. Consider a breadth-first tree rooted at u. For $0 \le i \le d$, let W_i denote the set of nodes at level i in the tree, and $Y_i = \bigcup_{\ell=0}^{i} W_\ell$. Note that the expansion properties of G imply that $|Y_i| \ge N_i$. Define now $j^\star$ as the smallest index such that $|Y_{j^\star}| > n/2$, which implies that $j^\star \le i^\star$. Also, w.l.o.g., we can assume that $j^\star \ge \lceil d/2 \rceil$, or otherwise we repeat the argument considering the breadth-first tree rooted at v. Indeed, since u and v are at distance d, one of the two breadth-first trees must reach at most $n/2$ nodes up to level $\lceil d/2 \rceil - 1$, or there would be a path shorter than d connecting u and v. The lemma follows. □

Theorem 3.5.2. *Consider an instance of $\mathcal{BT}(r(n), c(n))$ with $r(n) \ge \gamma_1 \sqrt{\log n / n}$ and $c(n) = \gamma_2 \log(1/r(n))$, for two suitable positive constants γ_1 and γ_2. With high probability,*

$$\mathrm{diam}(\mathcal{BT}(r(n), c(n))) = O\left(\frac{1}{r(n)} + \log n\right).$$

Proof. We apply Lemma 3.5.1 by estimating the value $i^\star$ for the graph $\mathcal{BT}(r(n), c(n))$, conditioning on the fact that the expansion of $\mathcal{BT}(r(n), c(n))$ is $\lambda(s) = \Omega(\min\{c(n), m/s\})$ for $s = O(m)$, and $\lambda(s) = \Omega\left(\sqrt{m/s}\right)$ for $s = \Omega(m)$, which happens with high probability (see Theorem 3.4.1).

In order to account for these two different expansion regimes, we proceed as follows. Let $K(j) = \min\left\{i : N_i \ge 2^j\right\}$, so that $i^\star = K(\log n - 1)$ and let j_1 be such that $2^{j_1} = \Theta(m)$. Since $\lambda(N_i) = \Omega(1)$ for $0 \le i < K(j_1)$, it follows that $K(j_1) = O(\log n)$. Observe that for $i > K(j_1)$, there exists a constant σ such that $\lambda(N_i) \ge \sigma\sqrt{m/N_i}$. As a consequence, for $j > j_1$ and for every

$\ell \geq K(j-1)$ we have:

$$\begin{aligned} N_\ell &\geq N_{K(j-1)} \prod_{s=K(j-1)}^{\ell-1} \left(1 + \frac{\sigma\sqrt{m}}{\sqrt{N_s}}\right) \\ &\geq N_{K(j-1)} \left(1 + \frac{\sigma\sqrt{m}}{\sqrt{N_{\ell-1}}}\right)^{\ell-K(j-1)} \\ &\geq 2^{j-1} \left(1 + \frac{\sigma\sqrt{m}}{2^{j/2}}\right)^{\ell-K(j-1)}. \end{aligned}$$

Since $K(j)$ is defined as the smallest index for which $N_{K(j)} \geq 2^j$, from the above inequalities it follows that $K(j) \leq \min\{\ell : (1+\sigma\sqrt{m}/2^{j/2})^{\ell-K(j-1)} \geq 2\}$, hence $K(j) - K(j-1) = O\left(2^{j/2}/(r(n)\sqrt{n})\right)$. Therefore,

$$\begin{aligned} i^\star = K(\log n - 1) &= \sum_{j=1}^{\log n - 1} (K(j) - K(j-1)) \\ &= \sum_{j=1}^{j_1} (K(j) - K(j-1)) + \sum_{j=j_1+1}^{\log n - 1} (K(j) - K(j-1)) \\ &= O(\log n) + O\left(\frac{1}{r(n)}\right), \end{aligned}$$

and the theorem follows from Lemma 3.5.1. □

We now show that Theorem 3.5.2 gives a tight estimate for the diameter of $\mathcal{BT}(r(n), c(n))$ when $r(n) = O(1/\log n)$.

Theorem 3.5.3. *Consider an instance of $\mathcal{BT}(r(n), c(n))$ with $r(n) \geq \gamma_1\sqrt{\log n/n}$ and $c(n) \geq \gamma_2 \log(1/r(n))$, for two suitable positive constants γ_1 and γ_2. With high probability,*

$$\mathrm{diam}(\mathcal{BT}(r(n), c(n))) = \Omega\left(\frac{1}{r(n)}\right).$$

Proof. Consider the natural tessellation introduced in Section 3.3. By Proposition 3.3.2, with high probability the top leftmost cell and the bottom rightmost cell contain at least one node each, hence the Euclidean distance between these two nodes is $\Theta(1)$. Therefore, any path in $\mathcal{BT}(r(n), c(n))$ connecting them must contain at least $\Omega(1/r(n))$ nodes. □

The above lower bound can be improved for large visibility radii $r(n) = \Omega(\log\log n / \log n)$, yielding a lower bound almost matching the $O(\log n)$ upper bound given by Theorem 3.5.2.

Theorem 3.5.4. *Consider an instance of $\mathcal{BT}(r(n), c(n))$ with $\gamma_1 \log\log n / \log n \le r(n) \le \sqrt{2}$ and $c(n) \ge \gamma_2 \log(1/r(n))$, for two suitable positive constants γ_1 and γ_2. With high probability,*

$$\mathrm{diam}(\mathcal{BT}(r(n), c(n))) = \Omega\left(\frac{\log n}{\log\log n}\right).$$

Proof. We show that, with high probability, each node of the graph has degree bounded by $\Delta = c(n) + 2\log_2 n$ and therefore the diameter of $\mathcal{BT}(r(n), c(n))$ cannot be smaller than the diameter of a tree with arity Δ.

Consider an arbitrary node $u \in \mathcal{BT}(r(n), c(n))$ and denote by $\deg(u)$ its degree. By the definition of Bluetooth Topology, the number $\deg_{\mathrm{OUT}}(u)$ of neighbors chosen by u satisfies (deterministically) $\deg_{\mathrm{OUT}}(u) \le c(n)$. For each node $v \ne u$ within distance $r(n)$ from u, let X_v be a 0–1 random variable, taking value 1 iff node v selects u as its neighbor. Observe that the number $\deg_{\mathrm{IN}}(u)$ of nodes which selected u as their neighbor can be written as

$$\deg_{\mathrm{IN}}(u) = \sum_v X_v,$$

where the summation ranges over all nodes within distance $r(n)$ from u. It is straightforward to see that $\Pr[X_v = 1] = c(n)/(N_v - 1)$, where N_v denotes the number of nodes in the visibility disk of v. By Proposition 3.3.2, we have that

$$\mathrm{E}\left[\deg_{\mathrm{IN}}(u)\right] \le \frac{4\beta}{\alpha} c(n) = O(\log\log n).$$

Let $t = 2\log_2 n$. For a sufficiently large n, $t \ge 6\mathrm{E}\left[\deg_{\mathrm{IN}}(u)\right]$. Since the neighbor choices performed by different nodes are independent, we can apply the Chernoff bound A.1.3 to $\deg_{\mathrm{IN}}$ to obtain that $\Pr\left[\deg_{\mathrm{IN}} \ge t\right] \le 2^{-t} = 1/n^2$. Applying the union bound over the n nodes yields that with probability at least $1 - 1/n$ all the nodes have in-degree at most t and hence their degree is at most $\Delta = c(n) + t = O(\log n)$. The theorem follows by observing that $\mathrm{diam}(\mathcal{BT}(r(n), c(n))) \ge \log_\Delta n = \Omega\left(\frac{\log n}{\log\log n}\right)$. □

We conclude this section by noticing that the $\mathcal{BT}\left(r(n), c(n)\right)$ exhibits, for not too large visibility radii, the same asymptotic diameter of the denser Random Geometric Graph with the same parameter $r(n)$. Indeed, [EJY05] proves that $\mathrm{diam}(\mathcal{RGG}\left(r(n)\right)) = O\left(1/r(n)\right)$, while Theorem 3.5.2 yields $\mathrm{diam}(\mathcal{BT}\left(r(n), c(n)\right)) = O\left(1/r(n)\right)$ for $r(n) = O\left(1/\log n\right)$.

Chapter 4

Spreading Information among Random Walkers

Recently mobile networks have attracted a lot of interest, both in the network and in the theory communities. Many theoretical models have been proposed, with the explicit goal of characterizing the spreading of information in such mobile networks. In this chapter we study the broadcast time of a message among agents performing random walks in the plane. We consider a sparse system below the percolation threshold, thus complementing the analysis carried out by previous works that dealt with dense scenarios. We show that in the sparse case the time needed to broadcast a message is asymptotically independent from the transmission radius of the moving agents, and we show how our analysis extends to other similar models of communication/interaction between agents.

The results presented in this chapter appeared in [PPPU10, PPPU11].

4.1 Introduction

The emergence of mobile computing devices has added a new intriguing component, *mobility*, to the study of distributed systems. In fully mobile systems, such as wireless mobile ad-hoc networks (MANETs), information is generated, transmitted and consumed within the mobile nodes, and communication is carried out without the support of static structures such as cell towers. These systems have been implemented in vehicular networks and sensor networks attached to soldiers on a battlefield or animals in a nature reserve [OW09, Ger05, JOW+02, Sto02]. Characterizing the power and limitations of mobile networks requires new models and analytical tools that address the unique properties of these systems [GT02, CPS09], which include:

- *Small transmission radius*: the transmission range of individual agents is restricted by limitations on energy consumption and interference from other agents;
- *Planarity*: agents reside, move and transmit on (subsets of) a plane. Low diameter graphs that are often used to model static communication networks are not useful here;
- *Dynamic communication graphs*: communication channels between agents are changing dynamically as mobile agents move in and out of the transmission radius of other agents;
- *Relative speeds*: transmission speed is significantly faster than the physical movement of the agents. A message can execute several hops before the network is altered by motion.

In this chapter we study the dynamics of information dissemination between agents moving independently on a plane. We consider a system of k mobile agents performing independent random walks on an n-node grid, starting at time 0 in a uniform distribution over the grid nodes. We focus on the fundamental communication primitive of broadcasting a rumor originating at one arbitrary agent to all other agents in the system. We characterize the *broadcast time* T_{B} of the system, which is the time it takes for all agents to receive the rumor.

We model the spreading of information in a mobile system by a dynamic communication graph process $\{G_t(r) \mid t \geq 0\}$, where the nodes of $G_t(r)$

are the mobile agents, and two agents are connected by an edge iff their distance at time t is within their transmission radius r. We are interested in *sparse systems* in which the transmission radius is below the percolation point $r_c \approx \sqrt{n/k}$ (i.e., the minimum radius which guarantees that $G_t(r_c)$ features a giant connected component), and where, with high probability, no connected component of $G_t(r_c)$ has more than a logarithmic number of agents [Pen03, PSSS11]. The broadcast time in sparse systems is dominated by the time it takes for many independent random walks to meet one another. Incorporating the fact that radio transmission is much faster than the motion of the agents, we assume that the rumor can travel throughout a connected component of G_t within one step, before the graph is altered by the motion.

Our main result is quite surprising: we show that below the percolation point the broadcast time does not depend on the transmission radius. We prove that $T_{\mathrm{B}} = \tilde{\Theta}\left(n/\sqrt{k}\right)$ for any r below r_c, giving a tight characterization up to logarithmic factors[1]. Our bound holds both when the transmission radius is significantly larger than the mobility range (i.e., the distance an agent can travel in one step), and when, in contrast to previous work [CMPS09, CPS09], the transmission radius as well as the the mobility range are very small. Our work complements a recent result by Peres et al. [PSSS11] who proved an upper bound polylogarithmic in k for the broadcast time in a system of k mobile agents which follow independent Brownian motions in $\mathbb{R}^d$, with transmission radius above the percolation point.

Our analysis techniques are applicable to a number of interesting related problems such as covering the grid with many random walks and bounding the extinction time in random predator-prey systems.

[1]The tilde notation hides polylogarithmic factors, e.g. $\tilde{O}\left(f(n)\right) = O\left(f(n)\log^c n\right)$ for some constant c.

4.2 Previous Work

Information dissemination has been extensively studied in the literature under a variety of scenarios and objectives. Here we restrict our attention to the results more directly related to our contribution.

A prolific line of research has addressed broadcasting and gossiping in static graphs, where the nodes of the graph represent active entities which exchange messages along incident edges according to specific protocols (e.g., *push*, *pull*, *push-pull*). The most recent results in this area relate the performance of the protocols to expansion properties of the underlying topology, with particular attention to the case of social networks, where broadcasting is often referred to as *rumor spreading* [CLP10]. (For a relatively recent, comprehensive survey on this subject, see [HKP$^+$05].) Unfortunately, mobile networks do not feature properties similar to those of social networks, mostly because of the physical limitations of both the movement and the radio transmission processes. Indeed, as noted in [Kle07], the short range of communication attainable by low-power antennas enforces the same local dynamics typical of disease epidemics [Dur99], which require physical proximity to propagate. Indeed, the analysis of opportunistic networks, where nodes relay messages as they come close to one another, employs models from the study of human mobility [CHC$^+$07, CFL08].

In the theory community there has been growing interest in modeling and analyzing information dissemination in dynamic scenarios, where a number of agents move either in a continuous space or along the nodes of some underlying graph and exchange information when their positions satisfy a specified proximity constraint. In [CMPS09, CPS09] the authors study the time it takes to broadcast information from one of k mobile agents to all others. The agents move on a square grid of n nodes and in each time step an agent can (a) exchange information with all agents at distance at most r from it, and (b) move to any random node at distance at most ρ from its current position. The results in these papers only apply to a very dense scenario where the number of agents is linear in the number of grid nodes (i.e., $k = \Theta(n)$). They show that the broadcast time is $\Theta\left(\sqrt{n}/r\right)$ w.h.p., when $\rho = O(r)$ and $r = \Omega\left(\sqrt{\log n}\right)$ [CMPS09], and it is $O\left((\sqrt{n}/\rho) + \log n\right)$ w.h.p., when $\rho = \Omega\left(\max\{r, \sqrt{\log n}\}\right)$ [CPS09]. These results crucially rely on $r + \rho = \Omega\left(\sqrt{\log n}\right)$, which implies that the range of agents' communications or movements at each step defines a connected graph.

In more realistic scenarios, like the one considered in the next section, the number of agents is decoupled from the number of locations (i.e., the graph nodes) and a smoother dynamics is enforced by limiting agents to move only between neighboring nodes. A reasonable model consists of a set of multiple, simple random walks on a graph, one for each agent, with communication between two agents occurring when they are nodes whose distance is at most $r \geq 0$. One variant of this setting is the so-called *Frog Model*, where initially one of k agents is active (i.e., is performing a random walk), while the remaining agents do not move. Whenever an active agent hits an inactive one, the latter is activated and starts its own random walk. This model was mostly studied in the infinite grid focusing on the asymptotic (in time) shape of the set of vertices containing all active agents [AMP02, KS03].

A model similar to our scenario is often employed to represent the spreading of computer viruses in networks, and the spreading time is also referred to as *infection time*. Kesten and Sidoravicius [KS05] characterized the rate at which an infection spreads among particles performing continuous-time random walks with the same jump rate. In [DNS06], the authors provide a general bound on the average infection time when k agents (one of them initially affected by the virus) move in an n-node graph. For general graphs, this bound is $O\left(t^* \log k\right)$, where t^* denotes the maximum average meeting time of two random walks on the graph, and the maximum is taken over all pairs of starting locations of the random walks. Also, in the paper tighter bounds are provided for the complete graph and for expanders. Observe that the $O(t^* \log k)$ bound specializes to $O(n \log n \log k)$ for the n-node grid by applying the known bound on t^* of [AF03]. A tight bound of $\Theta((n \log n \log k)/k)$ on the infection time on the grid is claimed in [WKK08], based on a rather informal argument where some unwarranted independence assumptions are made. Our results show that this latter bound is incorrect.

Recent work by Peres et al. [PSSS11] studies a process in which agents follow independent Brownian motions in $\mathbb{R}^d$. They investigate several properties of the system, such as detection, coverage and percolation times, and characterize them as functions of the spatial density of the agents, which is assumed to be greater than the percolation point. Leveraging on these results, they show that the broadcast time of a message is polylogarithmic in the number of agents, under the assumption that a message spreads within a connected component of the communication graph instantaneously, before the graph is altered by agents' motion.

4.3 Mathematical Model

We study the dynamics of information exchange among a set A of k agents performing independent random walks on an n-node 2-dimensional square grid $\mathcal{G}_n$, which is commonly adopted as a discrete model for the domain where agents wander. We assume that $n \geq 2k$, since sparse scenarios are the most interesting from the point of view of applications; however, our analysis can be easily extended to denser scenarios. We suppose that the agents are initially placed uniformly and independently at random on the grid nodes. The distance between two grid nodes u and v, denoted by $||u - v||$, is defined to be the Manhattan distance. Time is discrete and agent moves are synchronized. At each step, an agent residing on a node v with n_v neighbors ($n_v = 2, 3, 4$), moves to any such neighbor with probability $1/5$ and stays on v with probability $1 - n_v/5$, as depicted in Figure 4.1. With these probabilities it is easy to see that at any time step the agents are placed uniformly and independently at random on the grid nodes. The following two lemmas contain important properties of random walks on $\mathcal{G}_n$, which will be employed for deriving our results. In what follows, all logarithms are taken to the base two.

Lemma 4.3.1. *Consider a random walk on $\mathcal{G}_n$, starting at time $t = 0$ at node v_0. There exists a positive constant c_1 such that for any node $v \neq v_0$, the probability $p(v, v_0)$ that v is visited within $(||v - v_0||)^2$ steps is*

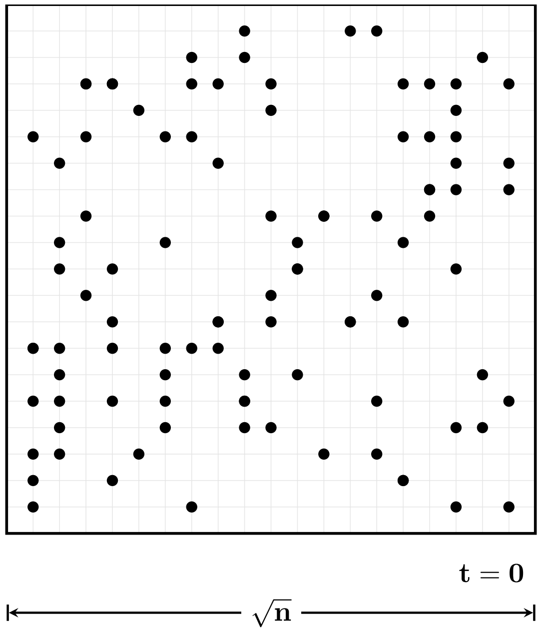
$$p(v, v_0) \geq \frac{c_1}{\max\{1, \log(||v - v_0||)\}}.$$

Proof. The Lemma is proved in [AMP02, Theorem 2.2] for the infinite grid $\mathbb{Z}^2$. By the "Reflection Principle" [Fel68, Page 72], for each walk in $\mathbb{Z}^2$ that started in $\mathcal{G}_n$, crossed a boundary and then crossed the boundary back to $\mathcal{G}_n$, there is a walk with the same probability that does not cross the boundary and visits all the nodes in $\mathcal{G}_n$ that were visited by the first walk. Thus, restricting the walks to $\mathcal{G}_n$ can only change the bound by a constant factor. □

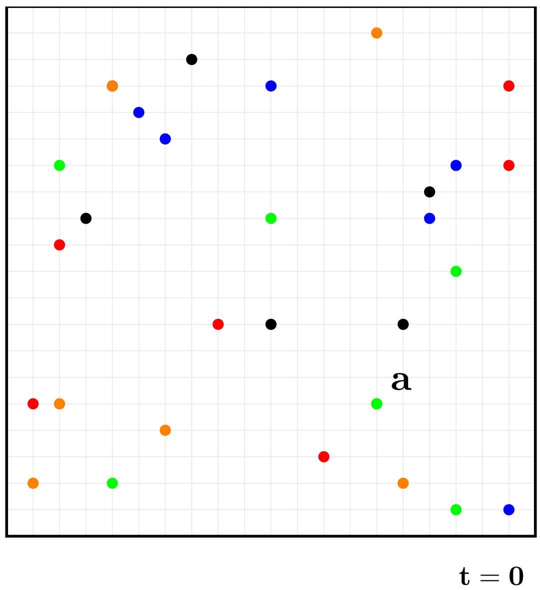
Lemma 4.3.2. *Consider the first ℓ steps of a random walk in $\mathcal{G}_n$ which was at node v_0 at time 0.*

1. *The probability that at any given step $1 \leq i \leq \ell$ the random walk is at distance at least $\geq \lambda\sqrt{\ell}$ from v_0 is at most $2e^{-\lambda^2/2}$.*
2. *There is a constant c_2 such that, with probability greater than $1/2$, by time ℓ the walk has visited at least $c_2\ell/\log \ell$ distinct nodes in $\mathcal{G}_n$.*

Proof. We observe that the distance from v_0 in each coordinate defines a martingale with bounded difference 1. Then, the first property follows from the Azuma-Hoeffding Inequality A.1.6. As for the second property, let R_ℓ be the set of nodes reached by the walk in ℓ steps. By Lemma 4.3.1, $\mathrm{E}[R_\ell] = \Omega(\ell/\log \ell)$ (even when v_0 is near a boundary), while $\mathrm{Var}[R_\ell] = \Theta\left(\ell^2/\log^4 \ell\right)$ (see [Tor86]). The result follows by applying Chebyshev's inequality. □



(a) At time $t = 0$, n agents are spread independently and uniformly at random over the $\sqrt{n} \times \sqrt{n}$ planar grid.



(b) Agents perform independent random walks on the grid. In each of the following subfigures, the position of each agent at the previous time step is marked by a shadow.

Figure 4.1: The random walkers mobility model.

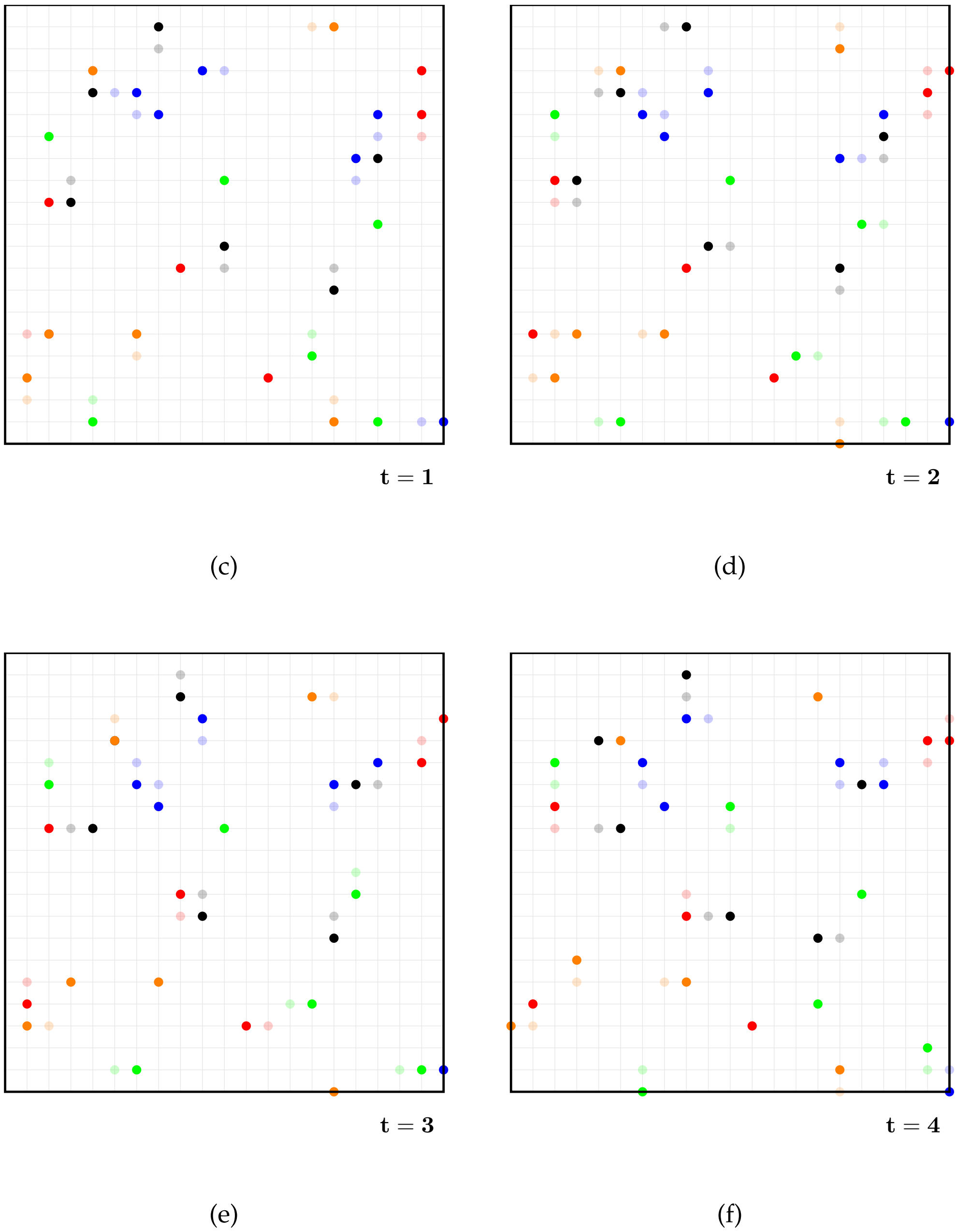


Figure 4.1: The random walkers mobility model.

Let M be a set of messages, which will be referred to as *rumors* henceforth, such that for each $m \in M$ there is (at least) one agent *informed* of m at time $t = 0$. W.l.o.g., we can assume that the number of distinct rumors is at most equal to the number of agent. We denote by $M_a(t)$ the set of rumors that agent $a \in A$ is informed of at time t, for any $t \geq 0$; possibly, $M_a(0) = \emptyset$. We assume that each agent is equipped with a *transmission radius* $r \in \mathbb{N}$, representing the maximum distance at which the agent can send information in a single time step.

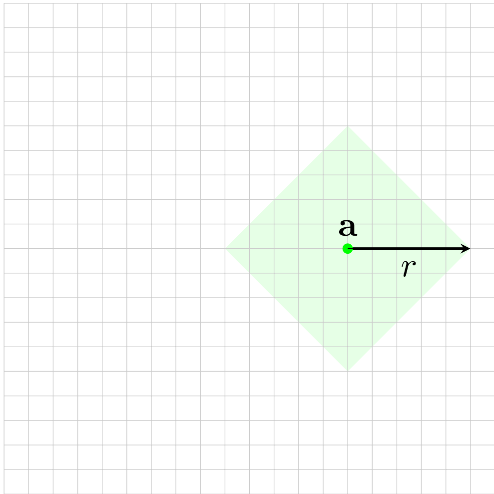
The spread of rumors can be represented by a dynamic communication graph process $\{G_t(r) \mid t \geq 0\}$, where $G_t(r)$, the *visibility graph at time t*, is a graph with vertex set A and such that there is an edge between two vertices iff the corresponding agents are within distance r at time t. For a graphical representation, see Figure 4.2.

Following a common assumption justified by the physical reality that the speed of radio transmission is much faster than the motion of the agents [PSSS11], we suppose that rumors can travel throughout a connected component of $G_t(r)$ before the graph is altered by the motion. We assume that within the same connected component agents exchange all rumors they are informed of. Formally, let C be a connected component of $G_t(r)$: for all $a \in C$, $M_a(t) = \bigcup_{a' \in C} M_{a'}(t-1)$. Note that the sets $M_a(t)$ can only grow over time, that is, agents do not "forget" rumors.

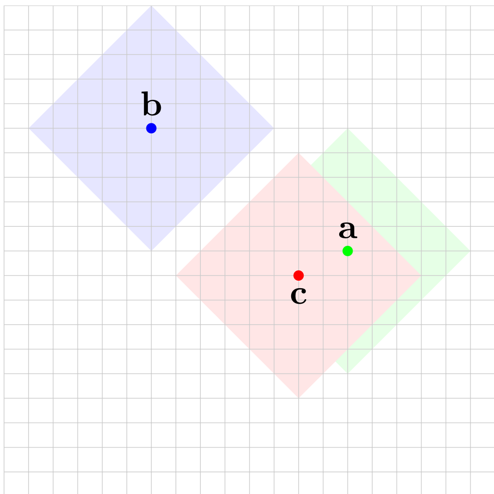
We are interested in studying the following quantities:

Definition 4.3.3 (Broadcast Time, Gossip Time)**.** *The broadcast time T_{B}^m of a rumor $m \in M$ is the first time at which every agent is informed of m, that is, for all $t \geq T_{\mathrm{B}}^m$ and $a \in A$, $m \in M_a(t)$. The gossip time T_{G} of the system is the first time at which every agent is informed of every rumor, that is, for any $t \geq T_{\mathrm{G}}$ and $a \in A$, $M_a(t) = M$.*

Note that both T_{B}^m and T_{G} depend on the transmission radius r, but we will omit this dependence to simplify the notation. We will also write T_{B} instead of T_{B}^m when the message m is clearly identified by the context.

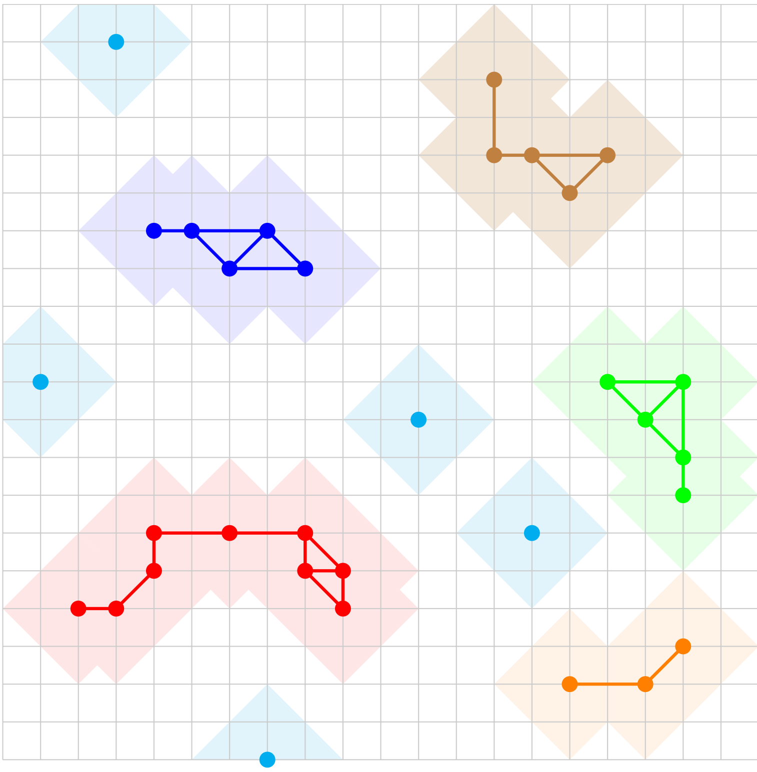


(a) Each agent is endowed with a transmission radius r. The visible grid locations are those contained in the shaded square.



(b) Agent a cannot communicate with agent b, since their distance is greater than r. On the other hand, a and c can exchange information, since they are within distance r.

Figure 4.2: The random walkers communication model.



$G_t(r)$

(c) A snapshot of the visibility graph $G_t(r)$. Each connected component is called "island", and nodes in two different islands are at distance at least r.

Figure 4.2: The random walkers communication model.

4.4 Broadcast Time in a Sparse System

In this section we give bounds to the broadcast time T_B of a rumor when the transmission radius is below the percolation point $r_c \approx \sqrt{n/k}$, that is, when all the connected components of $G_t(r)$ comprise at most a logarithmic number of agents. In this regime, we show that quite surprisingly T_B does not depend on the transmission radius, the reason being that the broadcast time is dominated by the time it takes for many independent random walks to intersect one another. In Subsection 4.4.1 we prove an upper bound on the broadcast time T_B in the extreme case $r = 0$, that is, when agents can exchange information only when they meet on a grid node. The same upper bound clearly holds for any other $r > 0$. Then, in Subsection 4.4.2 we show that the upper bound is tight, within logarithmic factors, for all values of the transmission radius below the percolation point. We also argue that the bounds on T_B easily extend to gossip time T_G.

4.4.1 Upper Bound on the Broadcast Time

The main technical ingredient of the analysis carried out in this section is the following lower bound on the probability that two random walks $\bar{a}, \bar{b}$ on the grid meet within a given time interval and not too far from their starting positions, which is a result of independent interest.

Lemma 4.4.1. *Consider two independent simple random walks on the grid* $\bar{a} = \langle a_0, a_1, \ldots \rangle$, *and* $\bar{b} = \langle b_0, b_1, \ldots \rangle$, *where* a_t *and* b_t *denote the locations of the walks at time* $t \geq 0$. *Let* $d = ||a_0 - b_0|| \geq 1$ *and define* D *to be the set of nodes at distance at most* d *from both* a_0 *and* b_0. *For* $T = d^2$, *there exists a constant* $c_3 > 0$ *such that*

$$\begin{aligned} P_{\bar{a},\bar{b}}(T) &\triangleq \Pr\left[\exists t \leq T \text{ such that } a_t = b_t \in D\right] \\ &\geq c_3 / \max\{1, \log d\}. \end{aligned}$$

Proof. The case $d = 1$ is immediate. Consider now the case $d > 1$. Let $P_t(w, x)$ denote the probability that a walk that started at node w at time 0 is at node x at time t, and let $R(w,u,D,s)$ be the expected number of times that two walks which started at nodes w and u at time 0 meet at nodes of D during the time interval $[0,s]$, then

$$R(w,u,D,s) = \sum_{t=0}^{s} \sum_{x \in D} P_t(w,x) P_t(u,x).$$

Let $\tau(a,b)$ be the first meeting time of the walks $\bar{a}$ and $\bar{b}$ at a node of D. Then

$$\begin{aligned} R(a_0,b_0,D,T) &= \sum_{t=0}^{T} \Pr\left[\tau(a,b)=t\right] R(a_t,a_t,D,T-t) \\ &\leq P_{\bar{a},\bar{b}}(T) \max_x R(x,x,D,T). \end{aligned}$$

Thus,

$$P_{\bar{a},\bar{b}}(T) \geq \frac{R(a_0,b_0,D,T)}{\max_x R(x,x,D,T)}.$$

It is easy to verify that $|D| \geq d^2/4$. Applying Theorem 1.2.1 in [Law91] we have:

$$\begin{aligned} R(a_0,b_0,D,T) &\geq \sum_{t=0}^{T} \sum_{x \in D} P_t(a_0,x) P_t(b_0,x) \\ &\geq \sum_{t=\frac{T}{2}+1}^{T} \sum_{x \in D} 4 \left(\frac{1}{\pi t}\right)^2 e^{-\frac{||x-a_0||^2+||x-b_0||^2}{t}}. \end{aligned}$$

By bounding $||x-a_0||^2$ and $||x-b_0||^2$ from above with T in the formula, easy calculations show that $R(a_0,b_0,D,T) = \Omega(1)$. Similarly, using the fact that there are no more than $4i$ nodes at distance exactly i from x, we have:

$$\begin{aligned} \max_x R(x,x,D,T) &\leq 1 + \sum_{t=1}^{T} \sum_{i=1}^{t} 4i\, 4 \left(\frac{1}{\pi t}\right)^2 2e^{-\frac{i^2}{t}} \\ &\leq 1 + \left(\frac{4}{\pi}\right)^2 \sum_{t=1}^{T} \frac{1}{t^2} \left(\left(\sum_{i=1}^{\sqrt{t}} i \right) + \left(\sum_{i=1+\sqrt{t}}^{t} i e^{-i^2/t} \right) \right) \\ &\leq 1 + \left(\frac{4}{\pi}\right)^2 \sum_{t=1}^{T} \frac{1}{t^2} \left(\frac{t}{2} + \left(\sum_{i=1+\sqrt{t}}^{t} i^2 e^{-i^2/t} \right) \right) \\ &\leq 1 + \left(\frac{4}{\pi}\right)^2 \sum_{t=1}^{T} \frac{1}{t^2} \left(\frac{t}{2} + \frac{e}{(e-1)^2} t \right) \\ &= O\left(\log T\right). \end{aligned}$$

We conclude that there is a constant $c_3 > 0$ such that

$$P_{\bar{a},\bar{b}}(T) \geq c_3 / \log d.$$

□

Observe that considering the difference random walk $\bar{a} - \bar{b} = \langle a_0 - b_0, a_1 - b_1, \ldots \rangle$ and computing the probability that it hits the origin in the prescribed number of steps does not provide any information about the place where the meeting occurs, hence it is not immediate to derive the above result through that approach.

The remainder of this section is devoted to proving the following upper bound on the broadcast time of a single rumor m in the case $r = 0$. We assume that $M_a(0) = \{m\}$ for some $a \in A$, and $M_{a'}(0) = \emptyset$ for any other $a' \neq a$.

Theorem 4.4.2. *Let $r = 0$. For any $k \geq 2$, with probability at least $1 - 1/n^2$,*

$$T_{\mathrm{B}} = \tilde{O}\left(\frac{n}{\sqrt{k}}\right).$$

We observe that since the diameter of $\mathcal{G}_n$ is $2\sqrt{n} - 2$, we can use Lemma 4.4.1 to show that with probability at least $1 - 1/n^2$, at time $8n \log^2 n$ an agent has met all other agents walking in $\mathcal{G}_n$. Thus, the theorem trivially holds for $k = O(\mathrm{polylog}(n))$.

From now on we concentrate on the case $k = \Omega\left(\log^3 n\right)$. We tessellate $\mathcal{G}_n$ into *cells* of side $\ell \triangleq \sqrt{14 n \log^3 n / (c_3 k)}$, where c_3 is defined in Lemma 4.4.1. (See Figure 4.3.)

We say that a cell Q is *reached* at time t_Q if t_Q is the first time when a node of the cell hosts an agent informed of the rumor, and we call this first visitor the *explorer* of Q. We first show that, after a suitably chosen number $T_1 = O\left(\ell^2 \log^4 n\right)$ of steps past t_Q, there is a large number of informed agents within distance $O\left(\ell \log^{5/2} n\right)$ from Q. Furthermore, we show that while the rumor spreads to cells adjacent to Q, at any time $t \geq t_Q + T_1$ a large number of informed agents are at locations close to Q. Figure 4.4

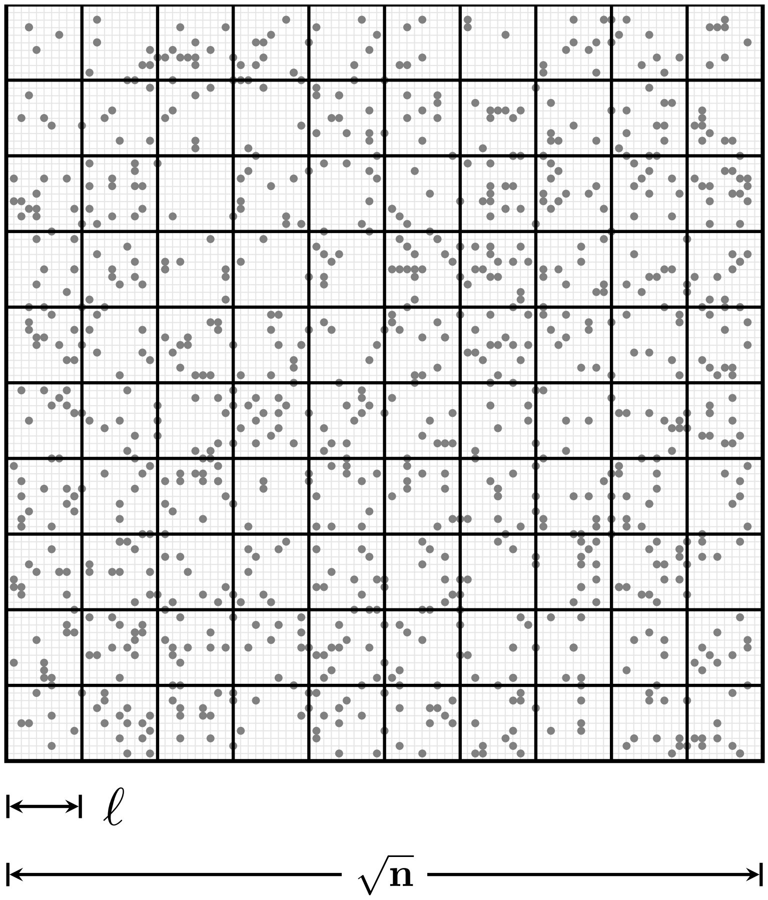


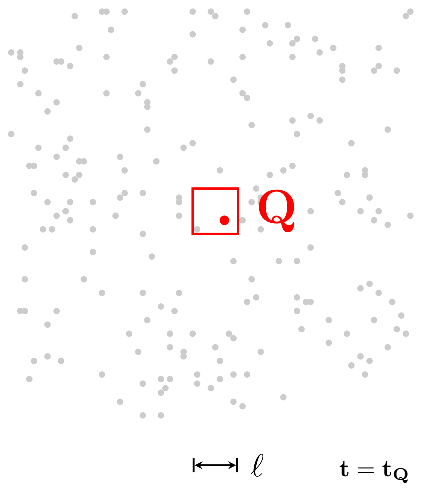
Figure 4.3: The tessellation of the planar grid adopted in our analysis.

offers some intuition about this process. These facts will imply that the exploration process proceeds smoothly and that all agents are informed of the rumor shortly after all cells are reached.

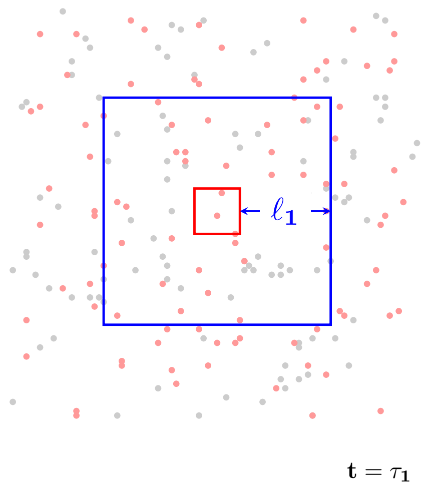
The above argument is made rigorous in the following sequence of lemmas.

Lemma 4.4.3. *Consider an arbitrary $\ell \times \ell$ cell Q of the tessellation. Let $T_1 = 16\beta\gamma\ell^2 \log^4 n$ and $c_4 = 8\sqrt{5\beta\gamma}$, where $\beta = 7/(2c_1)$ and $\gamma = 18/c_3$. By time $\tau_1 = t_Q + T_1$, at least $4\beta \log^2 n$ agents are informed and are at distance at most $2(1 + c_4 \log^{5/2} n)\ell$ from Q, with probability $1 - 1/n^8$, for sufficiently large n.*

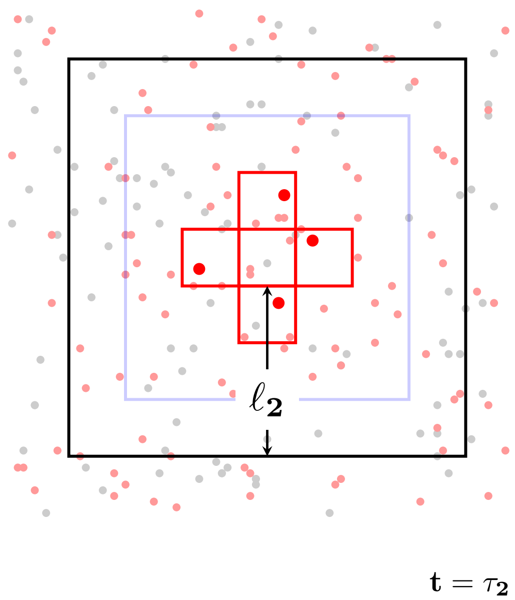
Proof. Since at any given time the agents are at random and independent locations, by the Chernoff bound we have that the following *density condition* holds with probability at least $1 - 1/n^9$, for sufficiently large n: for any cell



(a) Consider cell Q at time t_Q, that is, when it contains an informed agent for the first time.



(b) After T_1 time steps, many newly informed agents will be at distance at most ℓ_1 from Q.



(c) After additional T_2 steps, each cell adjacent to Q has been reached by the message, and many informed agents will stay at distance at most ℓ_2 from Q.

Figure 4.4: Intuition for the analysis of the upper bound on the Broadcast Time.

Q' and any time instant $t \in [0, n \log^4 n]$, the number of agents residing in cell Q' at time t is at least $(7 \log^3 n)/c_3$. In the rest of the proof, we assume that the density condition holds.

First, we prove that, by time τ_1, there are at least $4\beta \log^2 n$ informed agents in the system. We assume that at every time step $t \in [t_Q, \tau_1]$ there is always an uninformed agent in the same cell where the explorer resides (otherwise the sought property follows immediately by the density condition). For $1 \leq i \leq 4\beta \log^2 n$, let $t_i \geq t_Q$ be the time at which the explorer of Q informs the i-th agent. For notational convenience, we let $t_0 = t_Q$. To upper bound t_i, for $i > 0$, we consider a sequence of $\gamma \log^2 n$ consecutive, non-overlapping time intervals of length $4\ell^2$ beginning from time t_{i-1}. By the previous assumption, at the beginning of each interval the cell where the explorer resides contains an uninformed agent a. Hence, by Lemma 4.4.1, the probability that the explorer fails to meet an uninformed agent during all of these intervals is

$$\begin{aligned} \Pr\left[t_i > t_{i-1} + 4\gamma \ell^2 \log^2 n\right] &\leq (1 - c_3/\log(2\ell))^{\gamma \log^2 n} \\ &\leq 1/n^9, \end{aligned}$$

where the last inequality holds for sufficiently large n by our choice of γ. By iterating the argument for every i, we conclude that with probability at least $1 - 4\beta \log^2 n/n^9$, there are at least $4\beta \log^2 n$ informed agents at time τ_1. Let I denote the set of informed agents identified through the above argument, and observe that each agent of I was in the cell containing the explorer at some time step $t \in [t_Q, \tau_1]$.

To conclude the proof of the lemma, we note that, by Lemma 4.3.2, the probability that the explorer, during the interval $[t_Q, \tau_1]$, reaches a grid node at distance greater than $(c_4 \log^{5/2} n)\ell$ from its position at time t_Q is bounded by $2T_1/n^{10}$. Consider an arbitrary agent $a \in I$. As observed above, there must have been a time instant $\bar{t} \in [t_Q, \tau_1]$ when a and the explorer were in the same cell, hence at distance at most $(2 + c_4 \log^{5/2} n)\ell$ from Q. From time $\bar{t}$ until time τ_1 the random walk of agent a proceeds independently of the random walk of the explorer. By applying again Lemma 4.3.2, we can conclude that the probability that one of the agents of I is at distance greater than $2(1 + c_4 \log^{5/2} n)\ell$ from Q at time τ_1 is at most $8\beta \log^2 n/n^9$. By adding up the upper bounds to the probabilities that the event stated in the lemma does not hold, we get $1/n^9 + 4\beta \log^2 n/n^9 + 2T_1/n^{10} + 8\beta \log^2 n/n^9$, which is less than $1/n^8$ for sufficiently large n. □

Lemma 4.4.4. *Consider an arbitrary $\ell \times \ell$ cell Q of the tessellation. Let T_1, τ_1, c_4 and β be defined as in Lemma 4.4.3, and let $T_2 = (2(2 + c_4 \log^{5/2} n)\ell)^2$, $\tau_2 = \tau_1 + T_2$, and $c_5 = (4\sqrt{\log 16})c_4$. Then, the following two properties hold with probability at least $1 - 1/n^6$ for n sufficiently large:*

1. *For Q and for each of its adjacent cells, there exists a time t, with $\tau_1 \le t \le \tau_2$, at which there is an informed agent in the cell;*

2. *At any time t, with $\tau_1 \le t \le \tau_2 + T_1$, there are at least $\beta \log^2 n$ informed agents at distance at most $(2 + (2c_4 + c_5) \log^{5/2} n)\ell$ from Q.*

Proof. We condition on the event stated in Lemma 4.4.3, which occurs with probability $1 - 1/n^8$. Hence, assume that by time τ_1 there are at least $4\beta \log^2 n$ informed agents at distance at most $d_4 \triangleq 2(1 + c_4 \log^{5/2} n)\ell$ from Q. Consider the center node v of Q (resp., Q' adjacent to Q), so that at τ_1 there are at least $4\beta \log^2 n$ informed agents at distance at most $d_4 + 2\ell$ from v. By Lemma 4.3.1 the probability that v is not touched by an informed agent between τ_1 and τ_2 is at most $(1 - (c_1/\log(d_4 + 2\ell)))^{4\beta \log^2 n}$, which is less than $1/n^7$, for sufficiently large n, by our choice of β. Thus, Point 1 follows.

As for Point 2, consider an informed agent a which, at time τ_1, is at a node x at distance at most d_4 from Q. Fix a time $t \in [\tau_1, \tau_2 + T_1]$. By Lemma 4.3.2 the probability that at time t agent a is at distance greater than $(c_5 \log^{5/2} n)\ell$ from x is at most $1/2$. Hence, at time t the average number of informed agents at distance at most $d_4 + (c_5 \log^{5/2} n)\ell$ from Q is at least $2\beta \log^2 n$. Since agents move independently, Point 2 follows by applying the Chernoff bound to bound the probability that at time t there are less than $\beta \log^2 n$ informed agents at distance at most $d_4 + (c_5 \log^{5/2} n)\ell$ from Q, and by applying the union bound over all time steps of the interval $[\tau_1, \tau_2 + T_1]$. □

We are now ready to prove the main theorem of this section:

of Theorem 4.4.2. As observed at the beginning of the section, we can limit ourselves to the case $k = \Omega\left(\log^3 n\right)$. Consider the tessellation of $\mathcal{G}_n$ into $\ell \times \ell$ cells defined before, and focus on a cell Q reached for the first time at t_Q. By Lemma 4.4.4, we know that with probability at least $1 - 1/n^6$, in each time step $t \in [\tau_1, \tau_2 + T_1]$ there are at least $\beta \log^2 n$ informed agents at distance at most $d_5 \triangleq (2 + (2c_4 + c_5) \log^{5/2} n)\ell$ from Q and there exists a time $t' \in [\tau_1, \tau_2]$ such that an informed agent is again inside Q. By applying again the lemma, we can conclude that, with probability at least $(1 - 1/n^6)^2$, at any time step $t'' \in [t' + T_1, t' + 2T_1 + T_2]$ there are at least $\beta \log^2 n$ informed agents at distance at most d_5 from Q. Note that the two time intervals $[\tau_1, \tau_2 + T_1]$ and $[t' + T_1, t' + 2T_1 + T_2]$ overlap and the latter one ends at least T_1 time steps later. Thus, by applying the lemma $n \log^4 n$ times, we ensure that, with probability at least $(1 - 1/n^6)^{n \log^4 n} \geq 1 - \log^4 n/n^5$, from time τ_1 until the end of the broadcast, there are always at least $\beta \log^2 n$ informed agents at distance at most d_5 from Q.

Lemma 4.4.4 shows that each of the neighboring cells of Q is reached within time $\tau_2 = t_Q + T_1 + T_2$ with probability $1 - 1/n^6$. Therefore, all cells are reached within time $T^* = (2\sqrt{n}/\ell)(T_1 + T_2)$ with probability at least $1 - 1/n^5$. Hence, by applying a union bound over all cells, we can conclude that with probability at least $(1 - 1/n^5)(1 - \log^4 n/n^4) \geq 1 - 1/n^3$ there are at least $\beta \log^2 n$ informed agents at distance at most d_5 from each cell of the tessellation, from time $T^* + T_1$ until the end of the broadcast.

Consider now an agent a which at time $T^* + T_1$ is uninformed and resides in a certain cell Q. By an argument similar to the one used to prove Lemma 4.4.3, we can prove that a meets at least one of the informed agents around Q within $O\left(\ell^2 \log^5 n\right)$ time steps with probability at least $1 - 1/n^6$.

A union bound over all uninformed agents completes the proof. □

Observe that the broadcast time is a non-increasing function of the transmission radius. Therefore, the upper bound developed for the case $r = 0$ holds for any $r > 0$, as stated in the following corollary.

Corollary 4.4.5. *For any $k \geq 2$ and $r > 0$, $T_{\mathrm{B}} = \tilde{O}\left(n/\sqrt{k}\right)$ with probability at least $1 - 1/n^2$.*

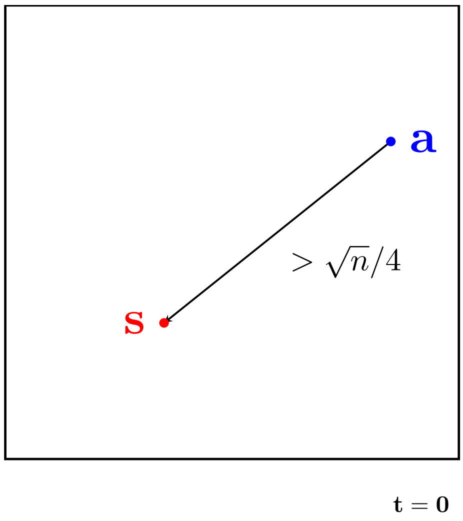
As another immediate corollary of the above theorem, we can prove that the gossiping of multiple distinct rumors completes within the same time bound, with high probability.

Corollary 4.4.6. *For any $k \geq 2$ and $r > 0$, $T_{\mathrm{G}} = \tilde{O}\left(n/\sqrt{k}\right)$ with probability at least $1 - 1/n$.*

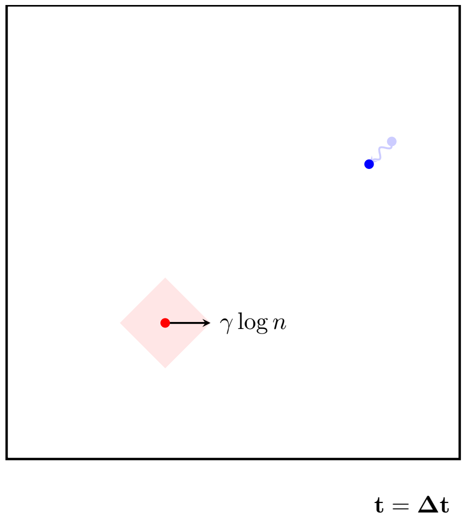
4.4.2 Lower Bound on the Broadcast Time

In this section we prove that the result of Corollary 4.4.5 is indeed tight, up to logarithmic factors, for any value r of the transmission radius below the percolation point. Note that this result is also a lower bound on T_{G} if there are multiple rumors in the system. First observe that with probability at least $1 - 2^{-(k-1)}$, there exists an agent placed at distance at least $\sqrt{n}/2$ from the source of m. Without loss of generality, we assume that the x-coordinates of the positions occupied by such an agent and the source agent differ by at least $\sqrt{n}/4$ and that the latter is at the left of the former. (The other cases can be dealt with through an identical argument.) In the proof, we cannot solely rely on a distance-based argument since we need to take into account the presence of "many" agents which may act as relay to deliver the rumor. Figure 4.5 presents graphically the intuition behind the proof of this lower bound.

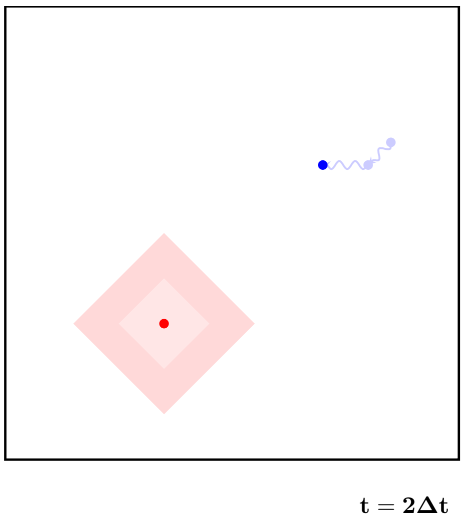
We define the *informed area* $\mathcal{I}(t)$ at time t as the set of grid nodes visited by any informed agent up to time t, and let $x(t)$ to be the rightmost grid



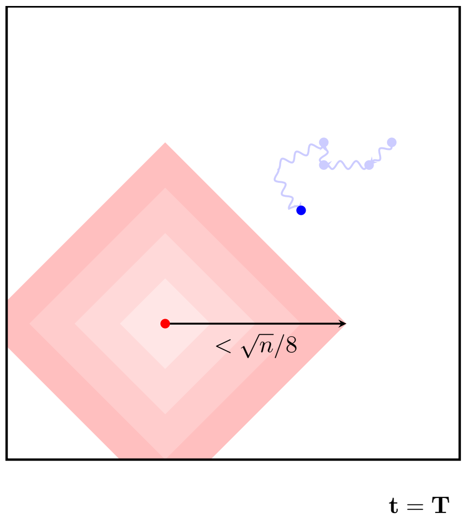
(a) Initially, with high probability there exists an agent a whose distance from the message source s is at least $\sqrt{n}/4$.



(b) In Δt time steps the message cannot reach a point further than $\gamma \log n$ from the initial position of the source.



(c) We look at the area reached by the message every Δt time steps. Meanwhile, the blue agent performs its own random walk.



(d) The blue agent cannot be informed at time T, since it could not reach the informed area by time T.

Figure 4.5: Intuition for the analysis of the lower bound on the Broadcast Time.

node in $\mathcal{I}(t)$. We will show that there is a sufficiently large value T such that, at time T, there is at least one uniformed agent right of $x(T)$.

We need the following definition:

Definition 4.4.7 (Island)**.** *Let A be the set of agents. For any $\gamma > 0$, let $G_t(\gamma)$ be the graph with vertex set A and such that there is an edge between two vertices iff the corresponding agents are within distance γ at time t. Then any connected component of $G_t(\gamma)$ is called an island of parameter γ at time t.*

Next, we prove an upper bound on the size of the islands.

Lemma 4.4.8. *Let $\gamma = \sqrt{n/(4e^6 k)}$. Then, the probability that there exists an island of parameter γ with more than $\log n$ agents at any time t, with $0 \leq t \leq 8n \log^2 n$, is at most $1/n^2$.*

Proof. Since at any given time the agents are uniformly distributed in $\mathcal{G}_n$, the probability that a given agent is within distance γ of another given agent at time t_0 is bounded by $4\gamma^2/n$. Fix a time t_0 and let $\mathcal{B}_w(t_0)$ denote the event that there exists an island with at least $w > \log n$ elements at time t_0. Then, recalling that w^{w-2} is the number of unrooted trees over w labeled nodes, we have that

$$\begin{aligned} \Pr\left[\mathcal{B}_w(t_0)\right] &\leq \binom{k}{w} w^{w-2} \left(\frac{4\gamma^2}{n}\right)^{w-1} \\ &\leq \left(\frac{ek}{w}\right)^{w} w^{w-2} \left(\frac{4\gamma^2}{n}\right)^{w-1} . \end{aligned}$$

Using definition of γ and the bound $w \geq 1 + \log n$ and $k \leq n$, we have

$$\Pr\left[\mathcal{B}_w(t_0)\right] \leq \frac{ek}{w^2} e^{-5(w-1)} \leq \frac{en}{w^2}\frac{1}{n^5} \leq \frac{1}{n^4},$$

for a sufficiently large n. Applying the union bound over $O\left(n \log^2 n\right)$ time steps concludes the proof. □

Next we show that, with high probability, for values of r below percolation, the informed area cannot expand to the right too fast.

Lemma 4.4.9. *Suppose* $r \leq \sqrt{n/(64e^6k)}$. *Let* $\gamma = \sqrt{n/(4e^6k)}$ *and let* t_0 *and* $t_1 = t_0 + \gamma^2/(144 \log n)$ *be two time steps. Then, with probability* $1 - 2/n^2$,

$$||x(t_1) - x(t_0)|| \leq \gamma \log n.$$

Proof. By Lemma 4.3.2, with probability $1 - 2/n^3$ an agent cannot cover a distance of more than $(\gamma - r)/2$ in $\gamma^2/(144 \log n)$ time steps. Thus with probability $1 - 1/n^2$, any two agents belonging to distinct islands of $G_{t_0}(\gamma)$ cannot come within distance r of each other in the interval $[t_0, t_1]$. Therefore, in that time interval, the rumor can propagate exclusively among agents belonging to those islands of $G_{t_0}(\gamma)$ containing at least one informed agent. By Lemma 4.4.8 we conclude that with probability $1 - 2/n^2$, in the interval $[t_0, t_1]$, $x(t)$ can move to the right of at most $\gamma(\log n - 1) + (\gamma - r)/2 < \gamma \log n$ positions. □

Finally, we can prove the main theorem of the section:

Theorem 4.4.10. *Let* $k \geq 2$ *and suppose that* $r \leq \sqrt{n/(64e^6k)}$. *Then, with probability* $1 - (2^{-(k-1)} + 1/n + 2/n^2)$,

$$T_{\mathrm{B}} = \Omega\left(\frac{n}{\sqrt{k}\log^2 n}\right).$$

Proof. As mentioned before, with probability at least $1 - 2^{-(k-1)}$ there exists an agent a placed at distance at least $\sqrt{n}/2$ from the source of the rumor; we may assume that their x-coordinates differ by at least $\sqrt{n}/4$ and that the uninformed agent is to the right of the source agent. Let $T = n/(2304e^3\sqrt{k}\log^2 n)$ and $\gamma = \sqrt{n/(4e^6k)}$. By Lemma 4.4.9, with probability $1 - 1/n$ the frontier cannot move right in T steps more than $(\gamma \log n/2)T/(\gamma^2/(144 \log n)) < \sqrt{n}/8$. By Lemma 4.3.2, with probability $1 - 2/n^2$, agent a cannot move left more than $2\sqrt{T \log n} < \sqrt{n}/8$, so that agent a cannot be informed by time T. Hence, the broadcast time is at least $T_{\mathrm{B}} > T = \Omega\left(n/(\sqrt{k}\log^2 n)\right)$ with probability at least $1 - (2^{-(k-1)} + 1/n + 2/n^2)$. □

4.5 Related Models

In this work we took a step toward a better understanding of the dynamics of information spreading in mobile networks. We proved a tight bound (up to logarithmic factors) on the broadcast of a rumor in a mobile network where agents perform independent random walks on a grid and the transmission radius defines a system below the percolation point. Our results complement the work of Peres et al. [PSSS11], who studied the behavior of a similar system above the percolation point. A similar bound holds for the gossip problem in this model, where at time 0 each agent has a distinct rumor and all agents need to receive all rumors.

Observe that by modeling the mobility process of each agent as a random walk on the grid, we are implicitly normalizing both the domain side and the transmission radius with respect to the mobility range, that is, the maximum distance that each agent can travel in unit time. This approach is reasonable, since in real-world mobile systems the mobility range is clearly much smaller than both the transmission radius and the physical diameter of the area spanned by the system. Clearly, our results are readily ported to the more general scenario featuring an arbitrary mobility range (like the one considered in [CMPS09, CPS09]) where we still consider the n-node grid populated by k agents with transmission range r but where, at each step, an agent can jump uniformly at random to a grid node within distance R from its current position. In this scenario, we can prove an upper bound of $\Theta\left(n/(R\sqrt{k})\right)$ to the broadcast time, still independent from the transmission radius r, under the assumption that both R and r are below the new percolation threshold $\Theta\left((\sqrt{n/k})/R\right)$. To show this result, the main idea is to consider a "coarsening" of the n-node grid into a new grid where each node represents a cell of side $\Theta(R)$. Then, we can apply the proof strategy we developed in Section 4.4.1, by noticing that each agent performs a random walk on this "coarse" grid, once we perform a suitable time scaling.

Our analysis techniques are applicable to some interesting related problems. For example, similar bounds on the broadcast time T_B can be obtained for the Frog Model [AMP02], where only informed agents move and uninformed agents remain at their initial positions. In particular, we can show that the broadcast time in the Frog Model is upper bounded by $T_B = \tilde{O}\left(n/\sqrt{k}\right)$. The argument is similar to the proof of Theorem 4.4.2, where Lemma 4.4.1 is replaced with Lemma 4.3.1 and the analysis of the initial phase of the information dissemination process is carried out by using Point 2 of Lemma 4.3.2. Also, a closer look at Theorem 4.4.10 reveals that the same argument employed in our dynamic model to bound T_B from below applies to the Frog Model. Thus, we have tight bounds, up to logarithmic factors, in this latter model as well.

Another measure of interest in systems of mobile agents is the *coverage time* T_C, that is, the first time at which every grid node has been visited at least once by an informed agent [PSSS11]. While in the Frog Model the broadcast time is obviously upper bounded by the coverage time, this relation is not so obvious in our dynamic model, since the coverage of the grid nodes does not imply that all agents have been informed of the rumor. Nevertheless, one can verify that, in our model, $T_C \approx T_B = \tilde{O}\left(n/\sqrt{k}\right)$. Indeed, by Point 2 of Lemma 4.4.4 and by Lemma 4.3.1, after $O\left(\ell^2\right)$ steps from the first time at which an informed agent reached a given cell, all the nodes of that cell have been visited by some informed agent. Hence, by the cell-by-cell spreading process devised in the proof of Theorem 4.4.2, we can conclude that the coverage time is bounded by $\tilde{O}\left(n/\sqrt{k}\right)$. (In fact, the same tight relation between T_C and T_B can be proved in the Frog Model.)

Another by-product of our techniques is a high-probability upper bound $O\left((n \log^2 n)/k + n \log n\right)$ on the *cover time* of k independent random walks on the n-grid (i.e., the time until each grid node has been touched by at least one such walk), improving on the previous results of [AAK$^+$08, ES09] which provide the same bound only for the expected value. Finally, in a closely related scenario, namely a random *predator-prey system* where

$k = \Omega(\log n)$ predators are to catch moving preys on an n-node grid by performing independent random walks [CFR09], our techniques yield a high-probability upper bound $O\left((n \log^2 n)/k\right)$ on the extinction time of the preys.

Chapter 5

Conclusions

This final chapter is devoted to a summary of the main contributions of this thesis and a discussion of some future research directions.

5.1 Summary

In this thesis we highlighted the need for a rigorous analytic approach to the study of information dissemination in wireless networks. We contributed several results on different network models, both static and dynamic. These results are interesting *per se* from the mathematical point of view, but they also have some important implications for real-world applications.

In Chapter 2 we described the main characteristics of wireless networks, with special attention to ad hoc, sensor and mobile networks. Then, we highlighted the essential problem with their state-of-the-art development process, which is nowadays substantially left to prototypes and numerical simulations. Hence, the necessity of a deeper, analytic understanding of their behavior, since it enables a better design approach, by identifying the key issues of those networks and quantifying their impact on the performances of the whole system. In particular, we suggested that information dissemination plays a crucial role in these distributed networks, where usually nodes are anonymous (not directly addressable because they do not hold a "global" identifier), there is no supporting infrastructure to rely upon and the system topology might change over time.

In Chapter 3 we studied a static random graph model, the Bluetooth Topology, known to capture the device discovery phase of Bluetooth-like networks. Our main result is a tight characterization of the expansion properties of the Bluetooth Topology. Since expansion is essentially a measure of bandwidth, being able to provide a quantitative estimate of this property is useful for the design and analysis of routing strategies [LR99]. Our result is valid for the entire set of visibility ranges $r(n)$ and number of neighbor choices $c(n)$ which are known to produce a connected graph, as opposed to the results of [PR04] and of [BDFL11] which hold only for the extreme cases $r(n) = \Theta(1)$ and $r(n) = \Theta\left(\sqrt{\log n/n}\right)$, respectively.

By leveraging on the expansion properties, we also derived nearly tight bounds on the diameter of the same topology, which is again an important measure for routing, related to the latency of the network. Our bounds are tight for a large spectrum of visibility ranges (i.e., $r(n) = O(1/\log n)$), which includes "small ranges", that is, those which are most interesting for the large scale deployment of the technology. For the larger ranges $r(n) = \Omega(1/\log n)$ we provided a more sophisticated lower bound which matches the upper bound up to a $\log\log n$ factor.

A somewhat surprising consequence of our results is that $\mathcal{BT}(r(n), c(n))$ exhibits roughly the same expansion as the Random Geometric Graph $\mathcal{RGG}(r(n))$, which is a much denser supergraph of $\mathcal{BT}(r(n), c(n))$. Also, the diameters of the two graphs differ by at most a logarithmic additive term. These are important considerations for real ad hoc networks, especially for what concerns routing capabilities, since they imply that $\mathcal{BT}(r(n), c(n))$ features similar bandwidth and latency characteristics of $\mathcal{RGG}(r(n))$ at only a fraction of the costs.

In Chapter 4 we considered a system of mobile agents, performing random walks on a planar grid, focusing on the dynamics of information spreading.

We proved a tight bound (up to logarithmic factors) on the broadcast time of a rumor in a sparse network, that is, when the transmission radius of the agents defines a system below the percolation point. Our results complement the work of Peres et al. [PSSS11], which studied the behavior of a similar system above the percolation point.

Our proofs show that in a sparse system the broadcast time of a message is asymptotically independent from the transmission radius of the agents. Indeed, we showed that the time needed to spread the message is dominated by the time needed for many agents to meet. This result is quite surprising and it marks a sharp difference between a sparse and a dense system, since we know from previous works that in a dense system the broadcast time depends on the transmission radius.

Our analysis techniques also apply to the gossip problem in the Random Walk Model, where at time 0 each agent has a distinct rumor and all agents need to receive all rumors. Moreover, our framework allows us to study several related scenarios with different mobility patterns or different rules of interaction between agents.

5.2 Further Research

Given the broad scope of the subject of this thesis, many problems in the field necessarily remain unsolved and many future research directions remain open. Here we discuss the most interesting ones and those that are most relevant to our contribution.

With reference to the Bluetooth Topology, we recall that it is still an open problem to establish, for *every* given visibility range $r(n) = \omega\left(\sqrt{\log n / n}\right)$, the *minimum* number $c(n)$ of neighbor choices which yield connectivity and to assess the corresponding expansion and diameter properties.

Moreover, there is a small asymptotic gap in our characterization of the diameter of $\mathcal{BT}(r(n), c(n))$, for $r(n) = \omega(1/\log n)$ and $r(n) = o(\log\log n/\log n)$, which stems from the proof of Lemma 3.5.1. We think that this gap might be removed by a tighter analysis, taking explicitly into account the process generating $\mathcal{BT}(r(n), c(n))$.

Furthermore, given the rather regular structure of $\mathcal{BT}(r(n), c(n))$, one can hope to derive sharp bounds on its expansion and diameter. In particular, one might want to improve the results presented in this thesis by obtaining sharper bounds that hold, with high probability, within a $(1 + o(1))$-factor from the respective "true" values.

Finally, from a more practical point of view, one might want to devise specific algorithms for communication tasks performed on a Bluetooth-like network. In particular, the expansion characterization given in Chapter 3 might help in designing algorithms which are efficient with respect to their completion time and the number of exchanged messages, and hence energy-efficient.

Nowadays mobile networks are attracting much interest, from both the network research community and the theory community. Recently our results have been generalized to higher dimensions by Lam et al. [LLM$^+$12]. They prove that the behavior of the system in $d \geq 3$ dimensions is qualitatively different than on the plane. In particular, they show a phase transition, as the ratio between the volume of the space and the number of agents varies, that does not occur in two dimensions.

Starting from the Random Walk Model described in Chapter 4, a first natural extension might be considering more complex mobility patterns. We might want to replace the grid with a generic graph, for example modeling the streets in a city. In this case, we might look for a characterization of the broadcast time in terms of the graph properties, for example the expansion of the underlying graph. Progress in this direction has been recently achieved by Clementi et al. [CST11].

We might also consider a different evolution process for the position of an agent, including more realistic features like group mobility, attraction toward popular places, and so on. The final goal is the formulation of sound analytical mobility models representing the dynamics of people traveling on road or subway networks, whereas, up to now, these types of systems have been studied by physicists, transportation scientists and engineers only by means of empirical or simulation techniques.

Another enhancement of the model, which is very interesting for practical purposes, consists in adopting a more realistic communication model. For example, instead of using a distance-based visibility graph, we can adopt a SINR-like communication model. Such a model takes explicitly into account the local effect of interference and hence might be more appealing when studying real-world networks. From the theoretical point of view, adopting such a communication model it might lead to slightly different dynamics at a system scale due to the effect of many unsuccessful communication attempts inside a very crowded area, leading to interesting tradeoffs between agent density and speed of information spreading.

As a first step toward this goal, we are studying a dynamic version of the "Line-of-Sight" model of Frieze et al. [FKRD09]. In this scenario agents are disseminated uniformly at random on a square grid. An agent can communicate only with those agents lying in the same row or column, up to a certain distance ℓ. This model aims at capturing the significant signal decay due to concrete buildings in a Manhattan-like environment, which forces communication to take place only when a direct line-of-sight between two agents exists. In a work in progress [PU12], we consider the dynamic model obtained by allowing agents to perform random walks on the grid, and we characterize the broadcast time of a message, in function of the agent density and the transmission range ℓ.

Another dimension of mobile networks that is certainly worth exploring is the communication complexity of generic distributed computations among moving agents. We might want to characterize the tradeoff between the speed of information spreading and the number of messages exchanged by the agents. One possible approach is the "parsimonious" one outlined in [BCF11, CS11], where agents might decide to stop communicating a certain message, for example because they have reached a pre-defined retransmission threshold. The intriguing aspect of these protocols lies in the analysis leading to the equilibrium threshold that keeps the number of transmission low while maintaining the message "alive" and "spreading" in the system.

Finally, we might investigate caching policies in mobile networks for space-consuming items, like songs or videos, being transferred directly from a user to another. In our results we assumed that each agent could store a number of messages linear in the number of nodes in the system, so that it can act as a relay for any message. A natural question to ask is about the existence of optimal caching policies when the local buffer of each agent is bounded. This scenario is particularly interesting since provides better understanding of social phenomena like the spreading of viral contents on social networks.

Appendix A

Analytic Tools

A.1 Probability Concentration Inequalities

For the sake of completeness, we report here some standard results, often referred through the text of this thesis, concerning the concentration of certain probability distributions.

We adopt the terminology and notation of Motwani and Raghavan [MR95], Mitzenmacher and Upfal [MU03], and Dubhashi and Panconesi [DP09]. The interested reader might consult one of the above books to find the full proofs and lots of applications of the following theorems.

Theorem A.1.1 (Markov's Inequality). *Let X be a random variable that assumes only nonnegative values. Then, for all $a > 0$,*

$$\Pr\left[X \geq a\right] \leq \frac{\mathrm{E}\left[X\right]}{a}.$$

Theorem A.1.2 (Chebyshev's Inequality). *For any $a > 0$,*

$$\Pr\left[\left|X - \mathrm{E}\left[X\right]\right| \geq a\right] \leq \frac{\mathrm{Var}\left[X\right]}{a^2}.$$

Theorem A.1.3 (Chernoff Bound for Poisson Trials). *Let $X_1, \ldots, X_n$ be independent 0-1 random variables such that $\Pr\left[X_i = 1\right] = p_i$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathrm{E}\left[X\right]$. Then the following Chernoff bounds hold:*

1. *for $0 < \delta \leq 1$,*

$$\Pr\left[X \geq (1+\delta)\mu\right] \leq e^{-\mu\delta^2/3};$$

2. *for* $0 < \delta < 1$,

$$\Pr\left[X \leq (1-\delta)\mu\right] \leq e^{-\mu\delta^2/2};$$

3. *for* $R \geq 6\mu$,

$$\Pr\left[X \geq R\right] \leq 2^{-R}.$$

Corollary A.1.4. *With the notation of Theorem A.1.3, for* $0 < \delta < 1$,

$$\Pr\left[|X-\mu| \geq \delta\mu\right] \leq 2e^{-\mu\delta^2/3}.$$

Generally, when analyzing complex systems, the random variables involved are not mutually independent. However, if their dependencies can be "bound", we might apply the following generalization of the Chernoff Bound (Theorem A.1.3), usually known as Azuma-Hoeffding Inequality (Theorem A.1.6).

We introduce the definition of martingale first.

Definition A.1.5 (Martingale)**.** *A sequence of random variables* $Z_0, Z_1, \ldots$ *is a martingale when:*

- $\mathrm{E}\left[|Z_n|\right] < +\infty$, *and*

- $\mathrm{E}\left[Z_{n+1} \mid Z_0, \ldots, Z_n\right] = Z_n$.

Now we can state the aforementioned concentration result.

Theorem A.1.6 (Azuma-Hoeffding Inequality)**.** *Let* $X_0, \ldots, X_n$ *be a martingale such that*

$$|X_k - X_{k-1}| \leq c_k.$$

For $t \geq 0$, *define* $C_t = \sum_{k=1}^{t} c_k^2$. *Then, for all* $t \geq 0$ *and any* $\lambda > 0$,

$$\Pr\left[|X_t - X_0| \geq \lambda\right] \leq 2e^{-\lambda^2/2C_t}.$$

A.2 Markov Chains

In this section we collect several definitions and facts concerning Markov chains that are necessary to precisely state several results in this thesis. Two classical references on the subject are the online notes by Aldous and Fill [AF03] and the recent book by Levin, Peres and Wilmer [LPW08]. A useful introduction to finite Markov chains and some related results is presented in Mitzenmacher and Upfal's textbook [MU03], from which we borrow the notation for this section.

In what follows we consider discrete-time, finite-state, and time-homogeneous Markov chains, that is, Markov chains that evolve in discrete steps, have a finite state space, and where the probability of any state transition does not change over time. Given a Markov chain **P**, we denote by $P^{\ell}_{i,j}$ the probability that the chain moves from state i to state j in ℓ steps.

The structure of a Markov chain might be characterized by inspecting its states. We recall a couple of definitions.

Definition A.2.1 (Irreducibility)**.** *A Markov chain is irreducible if all the states belong to one communicating class, that is, for any two states i, j there a positive integer ℓ such that $P^{\ell}_{i,j} > 0$.*

Consider the graph representation of a Markov chain **P**, that is, the directed graph having a node for each chain state and an edge from state i to state j iff $P_{i,j} > 0$. The above definition says that the chain is irreducible iff its graph representation is strongly connected.

Note that even a irreducible chain might present an unstable behavior, due to its particular periodic structure.

Definition A.2.2 (Periodicity)**.** *A state j in a discrete time Markov chain is periodic if there exists an integer $\Delta > 1$ such that*

$$\Pr\left[X_{t+s} = j \mid X_t = j\right] = 0$$

unless s is divisible by Δ. A discrete time Markov chain is periodic if any state in the chain is periodic. A state or a chain that is not periodic is called aperiodic.

However, note that a sufficient condition for the aperiodicity of an irreducible Markov chain is the existence of a state i such that $P_{i,i} > 0$. With respect to the graph representation of the chain, this means that state i presents a self-loop.

For the graph models described in this thesis we often provide analytic results that hold when the system has reached a steady state. This concept is formalized by the following definition.

Definition A.2.3 (Stationary distribution)**.** *A stationary distribution of a Markov chain* $\mathbf{P}$ *is a probability distribution* $\bar{\pi}$ *such that*

$$\bar{\pi} = \bar{\pi}\mathbf{P}.$$

Interestingly, a Markov chain with a "well behaved" structure has a unique stationary distribution, independent from the starting state of the chain. The classical result is the following.

Theorem A.2.4. *Any finite, irreducible, and aperiodic Markov chain has the following properties:*

- *the chain has a unique stationary distribution*

$$\bar{\pi} = (\pi_1, \pi_1, \ldots, \pi_n)\,;$$

- *for all states* $1 \leq i, j \leq n$*, the limit*

$$\lim_{t \to +\infty} P^t_{j,i}$$

  exists and it is independent of j;

- *the limit probability for the chain to be in state i satisfy*

$$\pi_i = \lim_{t \to +\infty} P^t_{j,i} = \frac{1}{h_{i,i}},$$

  where $h_{i,i}$ *is the expected number of steps for the chain, starting at state i, to return to the same state i.*

Note that once the hypotheses of Theorem A.2.4 are met, one can compute the stationary distribution of a Markov chain $\mathbf{P}$ by solving the system of equations $\bar{\pi} = \bar{\pi}\mathbf{P}$, together with the normalization constraint $\sum_i \pi_i = 1$.

Bibliography

[AAK$^+$08] N. Alon, C. Avin, M. Koucký, G. Kozma, Z. Lotker, and M. R. Tuttle. Many random walks are faster than one. In *Proceedings of the 20th Annual ACM Symposium on Parallelism in Algorithms and Architectures*, pages 119–128, 2008.

[AF03] D. Aldous and J. Fill. Reversible markov chains and random walks on graphs. http://stat-www.berkeley.edu/users/aldous/RWG/book.html, 2003.

[AMP02] O. S. M. Alves, F. P. Machado, and S. Y. Popov. The shape theorem for the frog model. *The Annals of Applied Probability*, 12(2):533–546, 2002.

[APM07] I. F. Akyildiz, D. Pompili, and T. Melodia. State of the art in protocol research for underwater acoustic sensor networks. *SIGMOBILE Mobile Computing Communication Review*, 11(4):11–22, 2007.

[ASSC02] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer Networks*, 38(4):393–422, 2002.

[BBMP04] S. Basagni, R. Bruno, G. Mambrini, and C. Petrioli. Comparative performance evaluation of scatternet formation protocols for networks of Bluetooth devices. *Wireless Networks*, 10(2):197–213, 2004.

[BC05] N. Benvenuto and G. Cherubini. *Algorithms for Communication Systems and their Applications*. Wiley, 2005.

[BCF11] H. Baumann, P. Crescenzi, and P. Fraigniaud. Parsimonious flooding in dynamic graphs. *Distributed Computing*, 24(1):31–44, 2011.

[BDFL11] N. Broutin, L. Devroye, N. Fraiman, and G. Lugosi. Connectivity threshold of Bluetooth graphs. In *arXiv:1103.0351v1*, 2011.

[BP86] G. Bilardi and F.P. Preparata. Area-time lower-bound techniques with applications to sorting. *Algorithmica*, 1(1):65–91, 1986.

[CAW06] C.-E. Chen, A. M. Ali, and H. Wang. Design and testing of robust acoustic arrays for localization and enhancement of several bird sources. In *Proceedings of the 5th international conference on Information processing in sensor networks*, pages 268–275, New York, NY, USA, 2006. ACM.

[CES04] D. Culler, D. Estrin, and M. Srivastava. Guest editors' introduction: Overview of sensor networks. *Computer*, 37(8):41–49, 2004.

[CFL08] A. Chaintreau, P. Fraigniaud, and E. Lebhar. Opportunistic spatial gossip over mobile social networks. In *Proceedings of the 1st ACM Workshop on Online Social Networks*, pages 73–78, 2008.

[CFR09] C. Cooper, A. Frieze, and T. Radzik. Multiple random walks in random regular graphs. *SIAM Journal of Discrete Mathematics*, 23(4):1738–1761, 2009.

[CHC+07] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing*, 6(6):606–620, 2007.

[CLP10] F. Chierichetti, S. Lattanzi, and A. Panconesi. Almost tight bounds for rumour spreading with conductance. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, pages 399–408, 2010.

[CMPS09] A. E. F. Clementi, A. Monti, F. Pasquale, and R. Silvestri. Information spreading in stationary markovian evolving graphs. In *Proceedings of the 23rd International Parallel and Distributed Processing Symposium*, 2009.

[CNPP09] P. Crescenzi, C. Nocentini, A. Pietracaprina, and G. Pucci. On the connectivity of Bluetooth-based ad hoc networks. *Concurrency and Computation: Practice and Experience*, 21(7):875–887, 2009.

[CPS09] A. E. F. Clementi, F. Pasquale, and R. Silvestri. MANETS: High mobility can make up for low transmission power. In *Proceedings of the 36th International Colloquium on Automata, Languages and Programming*, pages 387–398, 2009.

[CS11] A. E. F. Clementi and R. Silvestri. Parsimonious flooding in geometric random-walks - (extended abstract). In *Proceedings of the 25th International Symposium on Distributed Computing*, pages 298–310, 2011.

[CST11] A. E. F. Clementi, R. Silvestri, and L. Trevisan. Information spreading in dynamic graphs. *arXiv*, abs/1111.0583, 2011.

[DHM+07] D. P. Dubhashi, O. Häggström, G. Mambrini, A. Panconesi, and C. Petrioli. Blue pleiades, a new solution for device discovery and scatternet formation in multi-hop bluetooth networks. *Wireless Networks*, 13(1):107–125, 2007.

[DJH+05] D. P. Dubhashi, C. Johansson, O. Häggström, A. Panconesi, and M. Sozio. Irrigating ad hoc networks in constant time. In *Proceedings of the 17th ACM Symposium on Parallelism in Algorithms and Architectures*, pages 106–115, 2005.

[DNS06] T. Dimitriou, S. Nikoletseas, and P. Spirakis. The infection time of graphs. *Discrete Applied Mathematics*, 154(18):2577–2589, 2006.

[DP09] D. P. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.

[Dur99] R. Durrett. Stochastic spatial models. *SIAM Review*, 41:677–718, 1999.

[EJY05] R. Ellis, X. Jia, and C. Yan. On random points in the unit disk. *Random Structures and Algorithms*, 29(1):14–25, 2005.

[ES09] R. Elsässer and T. Sauerwald. Tight bounds for the cover time of multiple random walks. In *Proceedings of the 36th International Colloquium on Automata, Languages and Programming*, pages 415–426, 2009.

[Fel68] W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. I*. Wiley, 3 edition, 1968.

[FKRD09] A. M. Frieze, J. M. Kleinberg, R. Ravi, and W. Debany. Line-of-sight networks. *Combinatorics, Probability and Computing*, 18(1-2):145–163, 2009.

[FMPP04] F. Ferraguto, G. Mambrini, A. Panconesi, and C. Petrioli. A new approach to device discovery and scatternet formation in Bluetooth networks. In *Proceedings of the 18th International Parallel and Distributed Processing Symposium*, 2004.

[Ger05] M. Gerla. From battlefields to urban grids: New research challenges in ad hoc wireless networks. *Pervasive and Mobile Computing*, 1(1):77–93, 2005.

[GT02] M. Grossglauser and D. N. C. Tse. Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Transactions on Networking*, 10(4):477–486, 2002.

[HKP+05] J. Hromkovic, R. Klasing, A. Pelc, P. Ruzicka, and W. Unger. *Dissemination of Information in Communication Networks*. Springer, Berlin, 2005.

[Int07] Intel Corp., http://www.intel.com/research/exploratory/motes.htm. *Intel Mote research project*, 2007.

[JOW+02] P. Juang, H. Oki, Y. Wang, M. Martonosi, L.-S. Peh, and D. Rubenstein. Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with zebranet. In *Proceedings of the 10th International Conference on Architectural Support for Programming Languages and Operating Systems*, pages 96–107, 2002.

[Kle07] J. Kleinberg. The wireless epidemic. *Nature*, 449(7160):287–288, 2007.

[KS03] H. Kesten and V. Sidoravicius. A shape theorem for the spread of an infection. arXiv:math/0312511v1, 2003.

[KS05] H. Kesten and V. Sidoravicius. The spread of a rumor or infection in a moving population. *The Annals of Probability*, 33(6):2402–2462, 2005.

[KW08] L. Klingbeil and T. Wark. A wireless sensor network for real-time indoor localisation and motion monitoring. In *Proceedings of the 2008 International Conference on Information Processing in Sensor Networks*, pages 39–50, Washington, DC, USA, 2008. IEEE Computer Society.

[Law91] G. F. Lawler. *Intersections of random walks*. Birkhäuser, Boston, 1991.

[LLM+12] H. Lam, Z. Liu, M. Mitzenmacher, X. Sun, and Y. Wang. Information dissemination via random walks in d-dimensional space. In *Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms*, 2012.

[LNT87] M. Leiner, D. L. Nielson, and F. A. Tobagi, editors. *Special Issue on Packet Radio Networks*, volume 75. IEEE, Jan 1987.

[LPW08] D. A. Levin, Y. Peres, and E. L. Wilmer. *Randomized algorithms*. American Mathematical Society, 2008.

[LR99] T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM*, 46(6):787–832, 1999.

[MR95] R. Motwani and P. Raghavan. *Randomized algorithms*. Cambridge University Press, 1995.

[MU03] M. Mitzenmacher and E. Upfal. *Probability and Computing. Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2003.

[MZ04] V. Mehta and M. El Zarki. A bluetooth based sensor network for civil infrastructure health monitoring. *Wireless Networks*, 10(4):401–412, 2004.

[OS05] S. Oh and S. Sastry. Tracking on a graph. In *Proceedings of the 4th international symposium on Information processing in sensor networks*, page 26, Piscataway, NJ, USA, 2005. IEEE Press.

[OW09] S. Olariu and M. C. Weigle. *Vehicular Networks: From Theory to Practice*. Chapman and Hall/CRC, 2009.

[Pen03] M. Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.

[Per01] C. E. Perkins. *Ad hoc networking: an introduction*, chapter 1, pages 1–28. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001.

[PK00] G. J. Pottie and W. J. Kaiser. Wireless integrated network sensors. *Commununications of the ACM*, 43(5):51–58, 2000.

[PLC05] C. Park, J. Liu, and P. H. Chou. Eco: an ultra-compact low-power wireless sensor node for real-time motion monitoring. In *Proceedings of the 4th international symposium on Information processing in sensor networks*, page 54, Piscataway, NJ, USA, 2005. IEEE Press.

[PPP09] A. Pettarin, A. Pietracaprina, and G. Pucci. On the Expansion and Diameter of Bluetooth-Like Topologies. In *Proceedings of the 17th European Symposium on Algorithms (ESA'09)*, volume LNCS 5757, pages 528–539, 2009.

[PPP10] A. Pettarin, A. Pietracaprina, and G. Pucci. On the Expansion and Diameter of Bluetooth-Like Topologies. *To appear in Theory of Computing Systems*, 2010.

[PPPU10] A. Pettarin, A. Pietracaprina, G. Pucci, and E. Upfal. Infectious Random Walks. In *arXiv:1007.1604*, 2010.

[PPPU11] A. Pettarin, A. Pietracaprina, G. Pucci, and E. Upfal. Tight Bounds on Information Dissemination in Sparse Mobile Networks. In *Proceedings of the 30th Symposium on Principles of Distributed Computing (PODC'11)*, pages 355–362, 2011.

[PR04] A. Panconesi and J. Radhakrishnan. Expansion properties of (secure) wireless networks. In *Proceedings of the 16th ACM Symposium on Parallelism in Algorithms and Architectures*, pages 281–285, 2004.

[PSSS11] Y. Peres, A. Sinclair, P. Sousi, and A. Stauffer. Mobile geometric graphs: Detection, coverage and percolation. In *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms*, 2011.

[PU12] A. Pettarin and E. Upfal. Dynamic Line-of-Sight Networks. Work in progress, 2012.

[SCV$^+$06] P. Sikka, P. Corke, P. Valencia, C. Crossman, D. Swain, and G. Bishop-Hurley. Wireless adhoc sensor and actuator networks on the farm. In *Proceedings of the fifth international conference on Information processing in sensor networks*, pages 492–499, New York, NY, USA, 2006. ACM.

[SM05] G. Sharma and R. Mazumdar. Hybrid sensor networks: a small world. In *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*, pages 366–377, New York, NY, USA, 2005. ACM.

[SM08] A. Steed and R. Milton. Using tracked mobile sensors to make maps of environmental effects. *Personal Ubiquitous Computing*, 12(4):331–342, 2008.

[SNMT07] I. Stoianov, L. Nachman, S. Madden, and T. Tokmouline. PIPENET: a wireless sensor network for pipeline monitoring. In *Proceedings of the 6th international conference on Information processing in sensor networks*, pages 264–273, New York, NY, USA, 2007. ACM.

[Sto02] I. Stojmenovic. *Handbook of Wireless Networks and Mobile Computing*. Wiley, 2002.

[SZ06] I. Stojmenovic and N. Zaguia. Bluetooth scatternet formation in ad hoc wireless networks. *In Misic, J., Misic, V., eds.: Performance Modeling and Analysis of Bluetooth Networks, Auerbach Publications*, pages 147–171, 2006.

[Tor86] D. C. Torney. Variance of the range of a random walk. *Journal of Statistical Physics*, 44(1):49–66, 1986.

[WAK08] C. Wu, H. Aghajan, and R. Kleihorst. Real-time human posture reconstruction in wireless smart camera networks. In *Proceedings of the 2008 International Conference on Information Processing in Sensor Networks*, pages 321–331, Washington, DC, USA, 2008. IEEE Computer Society.

[WHC05] R. Whitaker, L. Hodge, and I. Chlamtac. Bluetooth scatternet formation: a survey. *Ad Hoc Networks*, 3:403–450, 2005.

[WKK08] Y. Wang, S. Kapadia, and B. Krishnamachari. Infection spread in wireless networks with random and adversarial node mobilities. In *Proceedings of the SIGMOBILE Workshop on Mobility Models*, pages 17–24, 2008.

[XHW03] B. Xu, S. Hischke, and B. Walke. The role of ad hoc networking in future wireless communications. *Proceedings of the International Conference on Communication Technology*, 2:1353–1358, April 2003.

Acknowledgments

I want to express my gratitude to Andrea and Geppino who are among the best mentors a graduate student might hope to have. I learned from you how being an open-minded person and a disciplined researcher feels like. More importantly, I learned that one can achieve any professional accomplishment while remaining a patient, humble, and generous person. Special appreciation goes to Geppino, who had to deal with all the bureaucratic aspects of my doctoral studies, being my "official" supervisor.

I cannot forget to thank Eli, who hosted me for three semesters at the Computer Science Department of Brown University, where I felt the rather unique privilege of working on cutting-edge problems in an inspiring environment. I want to thank you for your invaluable insights and teachings, and for the courtesy shown in delivering them.

I acknowledge the support of Fondazione "Ing. Aldo Gini" of Padova, which granted me two fellowships to study and perform research abroad. Moreover, I want to thank Alessandra, who has always been kind and helpful when I had to deal with the formal requirements of the Doctoral School.

I am in debt with all my co-authors, because they shared with me the excitement of several intellectual endeavors. This recognition extends to all the colleagues of the High Performance Computing Group in Padova and of the Theory Group in Providence. Thanks for setting up a friendly working environment where I learned a lot from each of you, probably more than you think!

My friends, both in Italy and in the USA, deserve a special place in my

heart. I am not going to list your names, as they are simply too many and I do not want to undergo the risk of forgetting someone. You helped me when I was struggling in difficult times, and you celebrated with me for my accomplishments. We had several interesting discussions on very profound questions, and we laugh together on light matters. I enjoyed the time I spent with you during my Ph.D., and I apologize for it being too short. I felt as you always were well aware of my moods and needs, and I owe you deep gratitude for your understanding.

My family (at large!) greatly supported me during these years, constantly providing me with material assistance, encouragement, and love. You listened patiently to my rants and complaints, and then you gave me invaluable pieces of advice. You forced me to keep my feet on the ground, reminding me of the important goals to pursue and of the essential values in life. This thesis would not have been written without you!

Padova, January 25, 2012 *A. P.*

Colophon

This thesis was typeset using $\LaTeX 2_\varepsilon$ on GNU/Linux Debian machines, mainly a Lenovo ThinkPad X120e laptop running Debian Wheezy.
The text editors used in this project are Vim and Kile. Figures have been produced using the PGF/TikZ package for LaTeX.
The main font for the text body is URW Palladio, a free clone of the famous Palatino designed by Herman Zapf.

The author can be contacted via email at `pettarin@dei.unipd.it`.

Padova, January 25, 2012