Random Geometric Graphs: An Algorithmic Perspective

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Computer Science

by

Chen Avin

2006
The dissertation of Chen Avin is approved.

__________________________________________
Adnan Darwiche

__________________________________________
Mark Hansen

__________________________________________
Deborah Estrin, Committee Co-chair

__________________________________________
Judea Pearl, Committee Co-chair

University of California, Los Angeles
2006
To my family
## Table of Contents

1 Introduction ........................................... 1  

1.1 Random Graphs ................................... 1  
1.2 Random Geometric Graphs ....................... 2  
1.3 Questions of Interest and Overview of Results ............. 4  
1.3.1 Random Walks in Random Geometric Graphs .......... 4  
1.3.2 Restricted Delaunay Triangulation in Random Geometric Graphs .................... 7  
1.3.3 Random Distance Graphs ..................... 8  
1.4 Organization ................................... 11  

2 Random Walks on Random Geometric Graphs ................. 12  

2.1 Introduction ...................................... 12  
2.2 Preliminaries .................................... 14  
2.2.1 Markov chains and the Simple Random Walk .......... 14  
2.2.2 Mixing Time and the Spectral Gap \((1 - \lambda_1)\) .......... 15  
2.2.3 Cover Time, Partial Cover Time and Blanket Time ........ 17  
2.2.4 Conductance .................................. 21  
2.2.5 Bounding The Cover Time via Resistance .......... 22  
2.3 Geo-dense Geometric Graphs ...................... 27  
2.3.1 Geo-dense Random Geometric Graphs ............... 28  
2.4 The Mixing Time of Random Geometric Graphs .............. 31
4.2 Definitions and Statement of Results ......... 71
4.3 Proofs ............................................. 73
   4.3.1 Proof of Theorem 4.1 (Connectivity) .... 74
   4.3.2 Proof of Theorem 4.2 (Clustering) ......... 79
   4.3.3 Proof of Theorem 4.3 (Diameter) ......... 81
   4.3.4 Proof of Theorem 4.4 (Local Routing) .... 82
4.4 Notes and Related Work ..................... 83
4.5 Conclusions ..................................... 84

5 Experimental Results .......................... 86
   5.1 Application Example ....................... 86
   5.2 Efficiency of Random Walk ............... 88
      5.2.1 Biased Random Walk ................. 90
   5.3 Quality of Random Walk .................. 91
      5.3.1 Partial Cover Quality ............... 91
      5.3.2 Robustness to Dynamics ............... 92
      5.3.3 Load Balancing ....................... 95

References ........................................ 97
List of Figures

1.1 (A) \( D(n, g_1^1) \) is equivalent to \( G(n, r) \). (B) \( D(n, g_r^r) \) is equivalent to \( B(n, p) \). (C) typical \( D(n, g_\alpha^r) \) case for \( \pi r^2 < \alpha < 1, 0 < \beta < \pi r^2 \). 10

2.1 Unit flow for upper bound on the 2–dimension grid resistance ... 25

2.2 (A) Lower bound for the Conductance in \( G(n, r) \). (B) upper bound for the Conductance in \( G(n, r) \) ............................ 33

2.3 Approximating the Conductance in RGG .......................... 34

2.4 \( T(u, v) \) and the flow \( c \) between \( u \) and \( v \) in \( G(n, r) \) ................. 39

2.5 Lower bound for \( R_{uv} \) on the \( G(n, r) \) .......................... 41

3.1 Different Graphs over a set \( V \) of 50 random nodes in the unit square with \( r = 0.3 \) (A) \( G(V, r) \). (B) \( Vor(V) \). (C) \( Del(V) \). (D) The edges in \( Del(V) \) that are longer than \( r \) (E) \( LocalDel(G) \) where consistent edges are in dots and inconsistent edges are in solid lines. 57

3.2 A case where edges \( \{w, x\} \) and \( \{u, v\} \) are consistent and intersect in \( LocalDel(G) \) .............................. 61

3.3 A disk \( D \) that must be included in the area \( disk(u, v, w) \cap (disk(u) \cup disk(v)) \) ................................. 62

3.4 An example where inconsistent edge \( \{u, v\} \) exist next to the border of the unit square .......................... 63

3.5 Average number of messages in Algorithm 1 for different size random networks .......................... 66

4.1 Computing the conditional probability \( P(\{i, j\} \mid \{k, i\}, \{k, j\}) \) .......................... 79
4.2 an area that is proportional to $x^2$ when local routing from $i$ to $j$
with $x = d(i, j)$. ................................................. 82

5.1 An example of the temperature in an area with six random light
sources ................................................................. 87

5.2 Comparing the histogram founded by the 80% random walk on the
graph and the histogram of the real data from Figure 5.1 ........ 87

5.3 The progress of partial cover time as function of number of steps
normalized to $n$ for different graphs of size $n = 4096$ ....... 88

5.4 Partial Cover time in increasing size of random network with same
density ................................................................. 89

5.5 Partial Cover Time in random walks with increasing bias on ran-
dom network ......................................................... 90

5.6 Hole size as a function of the number of steps normalized to $n$ for
$G(4096, r)$ with different radii $r$ ............................. 91

5.7 The Partial Cover time required when the probability $p$ of each
node to fail is increasing. The result are for 4096 nodes networks . 93

5.8 An example of a 4096 random network with 4 disaster areas. We
can see the creation of bottlenecks ............................... 94

5.9 The Partial Cover time required when we increase the number of
disaster areas in the network ....................................... 94

5.10 Histogram of the expected number of visits to a node in a 80%
cover random walk ............................................... 96
<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\triangle uvw$</td>
<td>triangle of $u, v, w$</td>
<td>56</td>
</tr>
<tr>
<td>$B_G$</td>
<td>blanket time</td>
<td>20</td>
</tr>
<tr>
<td>$B(n, p)$</td>
<td>Bernoulli random graph</td>
<td>1</td>
</tr>
<tr>
<td>$C_G$</td>
<td>cover time of graph $G$</td>
<td>17</td>
</tr>
<tr>
<td>$C_G(c)$</td>
<td>partial cover time of fraction $c$</td>
<td>18</td>
</tr>
<tr>
<td>$C_{uv}$</td>
<td>commute time</td>
<td>18</td>
</tr>
<tr>
<td>$d(i, j)$</td>
<td>Euclidian distance between $i, j$</td>
<td>9</td>
</tr>
<tr>
<td>$\text{disk}_r(u)$</td>
<td>disk centered around $u$ with radius $r$</td>
<td>55</td>
</tr>
<tr>
<td>$\text{disk}(u, v)$</td>
<td>disk through $u, v$ with diameter $d(u, v)$</td>
<td>55</td>
</tr>
<tr>
<td>$\text{disk}(u, v, w)$</td>
<td>unique circumcircle over $u, v$ and $w$</td>
<td>55</td>
</tr>
<tr>
<td>$D_{el}(G)$</td>
<td>Delaunay triangulation of a geometric graph $G$</td>
<td>53</td>
</tr>
<tr>
<td>$D(n, g)$</td>
<td>random distance graph</td>
<td>71</td>
</tr>
<tr>
<td>$\Delta x(t)$</td>
<td>variation distance</td>
<td>15</td>
</tr>
<tr>
<td>$\delta(v)$</td>
<td>degree of $v$</td>
<td>14</td>
</tr>
<tr>
<td>$\delta_{avg}$</td>
<td>average degree in the graph</td>
<td>37</td>
</tr>
<tr>
<td>$E(G)$</td>
<td>electrical network of $G$</td>
<td>22</td>
</tr>
<tr>
<td>$H_{uv}$</td>
<td>hitting time</td>
<td>17</td>
</tr>
<tr>
<td>$H_{max}$</td>
<td>maximum hitting time</td>
<td>18</td>
</tr>
<tr>
<td>$G(n, r)$</td>
<td>random geometric graph</td>
<td>2</td>
</tr>
<tr>
<td>$G_k(n)$</td>
<td>$k$-fuzz of a grid of size $n$</td>
<td>44</td>
</tr>
<tr>
<td>$\text{lune}_r(i, j)$</td>
<td>the intersection of $\text{disk}_r(i)$ and $\text{disk}_r(j)$</td>
<td>73</td>
</tr>
<tr>
<td>$\lambda_i$</td>
<td>$i$'th eigenvalue</td>
<td>16</td>
</tr>
<tr>
<td>$\lambda_{max}$</td>
<td>second largest eigenvalue in absolute value</td>
<td>16</td>
</tr>
<tr>
<td>$\mathcal{M}$</td>
<td>Markov chain</td>
<td>14</td>
</tr>
</tbody>
</table>
\( N(u) \) \hspace{1cm} \text{set of neighbors of } u \text{ including } u \hspace{1cm} 55
\( N(u, v) \) \hspace{1cm} \text{ } \hspace{1cm} 55
\( p(n) \) \hspace{1cm} \text{edge probability} \hspace{1cm} 2
\( P(c) \) \hspace{1cm} \text{power of a flow } c \hspace{1cm} 24
\( \mathcal{P}(\rho, g) \) \hspace{1cm} \text{Poisson random graph} \hspace{1cm} 74
\( \Pi \) \hspace{1cm} \text{stationary distribution} \hspace{1cm} 14
\( \Phi \) \hspace{1cm} \text{conductance} \hspace{1cm} 21
\( r(n) \) \hspace{1cm} \text{radius} \hspace{1cm} 2
\( r_{\text{con}} \) \hspace{1cm} \text{critical radius for connectivity} \hspace{1cm} 3
\( R \) \hspace{1cm} \text{resistance} \hspace{1cm} 23
\( R_{uv} \) \hspace{1cm} \text{effective resistance between } u \text{ and } v \hspace{1cm} 23
\( RDG(G) \) \hspace{1cm} \text{Restricted Delaunay Graph of } G \hspace{1cm} 53
\( \tau_x(\epsilon) \) \hspace{1cm} \text{mixing time} \hspace{1cm} 16
\( U \) \hspace{1cm} \text{the unit disk} \hspace{1cm} 71
\( Vor(V) \) \hspace{1cm} \text{Voronoi diagram of a set of nodes } V \hspace{1cm} 55
I could not have reached the end of this long, challenging path without the support and help of many people. First, I would like to thank my advisor Judea Pearl for his support and for allowing me the freedom to pursue my own interest. Despite difficult times, he was always there when I needed him and I’m thankful for that. I would also like to thank my co-chair Deborah Estrin for introducing me to sensor networks and for her valuable feedback on my work. I thank the other member of my committee, Adnan Darwiche and Mark Hansen for their support and for interesting and enjoyable classes along the way.

Many friends at UCLA with whom I worked and discussed my research made it possible for me to complete this work. In particular, I would like to thank Gunes Ercal who is a co-author and a friend for life and Carlos Brito who put me on the right track and taught me how research is being done. Chapter 2 and 5 of this dissertation are based on joint work with Gunes and Carlos [AB04, AE05b, AE05a]. Thanks to other members in our windowless lab along the years: Blai Bonet, Mark Hopkins, Ilya Shpitser, and Shailesh Vaya, each has helped me in his own way along the road. The open door, good advice and friendship of Eli Gafni helped me to continue during my most difficult times and I am grateful for that. I would like to thanks Kaoru Mulvihill for being supportive and understanding, and for all her help. I would not have started this journey without the encouragements of Rachel Ben-Eliyahu and Ran Giladi and without the financial support of the Department of Communication System Engineering at Ben-Gurion University, Israel.

A special thought goes to Verra Morgan whom I met on my first day at UCLA, and who was ever-since a countless source of smiles, moral support and reminders to ”stay out of trouble”. To David, whom I also met during my first
days at UCLA and who now he is my best friend: thank you for being there whenever I need you.

Finally, I would like to thank my family who has always been the most important part of my life. To my late father Tzvi who never finished high school, but showed me the joy of learning and curiosity. To my late grandparents Yuda and Shlomit who inspired me with their knowledge and wisdom. You are always with me. To my mother Ilana who is always there to support me, in good and bad times and to my brothers and sisters Ayelet, Eran, Shira and Yagil for there unconditional love. Most of all there is my own "little" family: To my wonderful kids, Itamar and Maya, you are the source of my power. What I have learned from them, and in particular from Itamar, is priceless and beyond what I could ever imagine, and this is just the beginning. And last, my wife, my love, Yehudit who has stood by me all the way and makes all of this possible. Thank you.
Vita

1970 Born, Beer-Sheva, Israel.


2001–2003 M.S., Computer Science, University of California Los Angeles, Los Angeles, USA.

Publications


Avin, C., and Ercal, G. Bounds on the Mixing Time and Partial Cover of Ad-hoc


Abstract of the Dissertation

Random Geometric Graphs: An Algorithmic Perspective

by

Chen Avin

Doctor of Philosophy in Computer Science
University of California, Los Angeles, 2006

Professor Judea Pearl, Co-chair
Professor Deborah Estrin, Co-chair

A random geometric graph $G(n, r)$ is a graph resulting from placing $n$ points uniformly at random on the unit square (or on the unit disk) and connecting two points iff their Euclidean distance is at most the radius $r(n)$. Recently, this class of random graphs has gained relevance as a natural model for wireless ad-hoc and sensor networks. Investigating properties of these graphs can unearth properties of the real-life systems they model and allow for the design of efficient algorithms. In this work, we study properties of random geometric graphs motivated by challenges encounter in sensor networks applications.

Sensor networks are constructed from a large number of low-cost, low-power sensors equipped with wireless communication and limited processing capabilities. These devices are expected to be embedded densely into the environment and cooperate to achieve high level tasks. In many cases, the network created by these devices is subject to dramatic structural changes due to failures, node mobility and other factors. Motivated by the disadvantage entailed by topology driven algorithms (e.g. the need to maintain data structures and to execute
expensive recovery mechanisms), we investigate algorithms that require no knowledge of network topology, in particular algorithms based on random walks. The investigation we carry out in this dissertation attempts to assess the efficiency of random-walk-based algorithms. Surprisingly we show that despite their simplicity, random-walk-based algorithms can be competitive with optimal topology driven strategies for certain tasks. In particular, we analyze three properties of random walks on these graphs, the mixing time, the cover time and the partial cover time that are essential to determining the efficiency of this approach for sensor network tasks.

We also investigated another property of random geometric graphs which has implication for routing and topological control in sensor networks. The goal here is to construct a special subgraph, the Restricted Delaunay Graph, that permits efficient routing, based only on local information. We bound the number of messages needed for this task in these networks and presents a novel algorithm, based on the graph properties, that is more efficient than previous ones.

We offer a new extension of random geometric graphs called random distance graphs, to explain some interesting similarities between random geometric graphs and the familiar model of Bernoulli random graphs\footnote{A Bernoulli random graph (a.k.a. Erdős-Rényi graph) $B(n, p)$ is a random graph with $n$ nodes in which each edge is chosen independently at random with an edge probability $p(n)$.}. Interestingly, while, neither random geometric graphs nor Bernoulli random graphs are suitable to model social networks, a typical case of random distance graphs can captures important properties of social networks. These properties, known as ”Small World”, includes small average path length and high clustering have been the distinctive mark of many natural networks.
CHAPTER 1

Introduction

1.1 Random Graphs

The field of random graphs was established by the seminal work of Erdős and Rényi in 1959 "On Random Graphs" [ER59] and the subsequent series of papers they co-authored. Ever since, scientists have been using different models of random graphs to predict and understand the typical structure of the complex systems that pervade real-life. This approach has proved to be useful when systems are large, with partially unknown relationship, and no deterministic mechanism to explain the way those relationship arose. Random graphs have been used to model the network of social links, computer networks, the metabolic network of a cell, the electricity networks of power lines, business relations between companies, and the linking structure of the World Wide Web, to name a few [Gil61, WS98, NWS02, BCR03, Pen03].

Among the known models of random graphs, in this work we mostly consider two: i) *Bernoulli random graphs* which were introduced by Gilbert [Gil59] and later analyzed by Erdős and Rényi (a.k.a. Erdős-Rényi graphs) ii) *Random geometric graphs* (RGG) which are the main focus of our study [Gil61, Pen97, GK98, DPS98, Pen03].

i) A Bernoulli random graph $\mathcal{B}(n, p)$ is a random graph with $n$ nodes in which
each edge (out of the \( \binom{n}{2} \) possible edges) is chosen independently at random with an edge probability \( p(n) \).

ii) A random geometric graph \( G(n, r) \) is a graph resulting from placing \( n \) points uniformly at random on the unit square (or on the unit disk) and connecting two points iff their Euclidean distance is at most the radius \( r(n) \).

One of the main goals of studying random graphs is to elucidate the properties of a typical graph. That is, after setting the parameter of the model to a specific function of \( n \) (in our case setting \( r(n) \) or \( p(n) \) as functions of \( n \)), we let \( n \) go to infinity, and we look at the resulting graph. If a property \( Q \) exists with probability going to 1 as \( n \) goes to infinity (i.e. with high probability), we say that a typical graph has property \( Q \). In this context we can also speak of the evolution of the typical graph: the way the properties of a typical graph change as the typical graph evolves when we increase the order of the functions we set the parameter to. For example, we may watch if a property \( Q \) appears as we change the parameter \( p(n) \) in the Bernoulli graph from \( \frac{1}{n} \) to \( \frac{\log n}{n} \). When a random graph is used to mirror a real-life system, the properties of the typical graph, in turn, can be used to predict the expected behavior of the system.

1.2 Random Geometric Graphs

Although the origin of random geometric graphs can be traced back to the work of Gilbert in 1961 [Gil61], they were not theoretically analyzed until recent years. These graphs have traditionally been associated with areas such as statistical physics and hypothesis testing [Pen03], but have gained new relevance with the advent of wireless ad-hoc and sensor networks [EGH99, PK00]. In particular, Wireless Sensor Networks (WSN) are emerging as a new type of computing plat-
form that may revolutionize information gathering and processing [EGH99]. Sensor networks are constructed from a large number of low-cost, low-power sensors equipped with wireless communication and limited processing capabilities. These devices are expected to be embedded densely into the environment to create a network in which sensors can cooperate to achieve high level tasks. A wide range of applications have been offered for such systems in the last few years, ranging from environmental and habitat monitoring to disaster management and manufacturing process flow [CES04]. In these networks energy is mostly consumed by radio communication, so the number of messages being sent for a given task is considered the major efficiency metric.

In ad-hoc and sensor networks, interference grows with increased communication radius. So, for a property of interest $Q$ of random geometric graphs, one wants to find a tight upper bound on the smallest radius $r_Q(n)$, that will guarantee that $Q$ holds with high probability ($w.h.p.$). The radius $r_Q(n)$ is called the critical radius if $Q$ exhibits a sharp threshold (also known as phase transition), that is if the difference between the smallest radius for which the property holds with high probability and the largest radius for which the property holds with low probability goes to 0 as $n \to \infty$.

The critical radius for connectivity, $r_{con}$, has been of special interest, and it has been shown that if $r \geq r_{con} = \sqrt{\frac{\log n + \gamma_n}{\pi n}}$ then $G(n, r)$ is connected $w.h.p.$ as $n \to \infty$ iff $\gamma_n \to +\infty$ and disconnected $w.h.p.$ iff $\gamma_n \to -\infty$ [Pen97, GK98]. In ad-hoc and sensor networks we usually require that the network is connected, so information can be exchanged across the whole network. Therefore we will use the critical radius for connectivity, $r_{con}$, as a reference point to which we will compare other critical radii; naturally, it is always desirable that a critical radius be as close as possible to $r_{con}$ or, in other words, that the property appears at
the same evolution stage as connectivity.

1.3 Questions of Interest and Overview of Results

1.3.1 Random Walks in Random Geometric Graphs

Sensor networks have strict energy and memory constraints and, in many cases, are subject to dramatic structural changes created by failures, mobility and other factors [EGH99, PK00, Cer02]. Thus, topology driven algorithms are at a disadvantage for such networks as they need to maintain data structures (e.g. pointers to cluster heads, routing tables and spanning trees) and so have to handle critical points of failure (e.g. cluster heads, nodes close to the root in a spanning tree). Consequently, algorithms that require no knowledge of network topology are at an advantage, one such example is random-walks-based algorithms.

A Random Walk is the simple process of visiting the nodes of a graph $G$ in some sequential random order. The walk starts at some fixed node, and at each step it moves to a neighbor of the current node chosen randomly according to an arbitrary distribution. A simple random walk, which we consider here, is a walk where the next node is chosen uniformly at random from the set of neighbors. In a sensor networks context, a random walk will result when messages are sent at random from sensor to sensor. This process presents no critical points of failure; on the contrary, all the nodes are equally unimportant at all times$^1$.

Since random-walks-based algorithms display locality, simplicity, low-overhead and robustness to failures, questions regarding the properties of random walks on sensor networks have recently attracted interest [GRK04a, AE05a, BGP]. In par-

---

$^1$Assuming the probability of a node to fail during the short time it holds the message is negligible.
ticular, random walk techniques have been proposed for gossiping [KDG03], for information collection and query answering [SKH03, AB04] and even for routing [BE02, SB02].

At first glance, it would seem that a random walk is a simplified and highly inefficient strategy since it ignores the topology of the graph. The investigation carried-out in this dissertation attempt to asses the efficiency of random-walk-based algorithms viz-à-vis the optimal possible. Surprisingly we show that despite their simplicity, random-walk-based algorithms can be competitive in certain tasks.

To study the efficiency of the approach, two important properties of random walks on graphs need to be evaluated: mixing time [Sin92] and cover time [Ald83]. The mixing time $\tau_G$ of a graph $G$ is the time (measured by number of steps or in our case by the number of messages) taken by a simple random walk on $G$ to sample a node with likelihood equal to the steady state distribution of the Markov chain on $G$. For example, if $G$ is regular (i.e. all nodes have the same degree), this means sampling uniformly at random. $G$ is said to be rapid mixing if the mixing time is poly-logarithmic in the number of nodes. For an efficient sampling we would like our graph to be rapid mixing. The cover time $C_G$ of a graph $G$ is the expected time taken by a simple random walk to visit all nodes in $G$. This property is relevant to a wide range of algorithmic applications [KDG03, GMS04, WLB98, JS97, AB04], and methods of bounding the cover time of graphs have been thoroughly investigated [Mat88, Ald89, CRR89, BK89, Zuc90, AKL79]. Several bounds on the cover time of particular classes of graphs have been obtained with many positive results [CRR89, BK89, Jon98, JS00, CF03]. Let’s now pose our first two question of interest:
Question 1. What is the critical radius for random geometric graphs to be rapid mixing? More precisely, is there a critical radius $r_{\text{rapid}}$ that will guarantee with high probability that $\mathcal{G}(n, r)$ with $r \geq r_{\text{rapid}}$ is rapid mixing? If so, what is it?

Let the *optimal cover time* be a cover time of the same order as that associated with the complete graph (which is known to be the best case [Fei95a]).

Question 2. What is the critical radius above which random geometric graphs have optimal cover time? More precisely, is there a critical radius $r_{\text{opt}}$ that will guarantee with high probability that $\mathcal{G}(n, r)$ with $r \geq r_{\text{opt}}$ has optimal cover-time? If so, what is it?

We show in this dissertation that such thresholds do indeed exist and, surprisingly, that the threshold for optimal cover time occurs at a radius $r_{\text{opt}} = \Theta(r_{\text{con}})^2$. On the other hand, $r_{\text{rapid}} = \omega(r_{\text{con}})$ (i.e. an order larger than $r_{\text{con}}$), and, in particular, the radius required for rapid mixing is $r_{\text{rapid}} = \Theta(1/\text{poly}(\log n))$. Moreover, we also prove that both the *partial cover time* [AB04] - the expected time taken by a random walk to visit a constant fraction of the nodes, and the *blanket time* [WZ96] - the expected time taken by a random walk to visit all nodes with stationary frequencies, are optimal for random geometric graphs. This results imply the efficiency of random walk approaches in sensor networks. Optimal cover time implies low number of messages and the short radius implies low power transmission.

The techniques we use to prove these results rely on two theoretical tools: geo-density and network flow. First, we show that random geometric graphs are *geo-dense* (to be define formally in the text), namely, geometric graphs that have $\frac{1}{r} = \Theta(\sqrt{\frac{n}{\log n}})$.

\[ \frac{1}{r} = \Theta(\sqrt{\frac{n}{\log n}}) \]
two desirable properties: almost uniform node distribution across the unit square and regularity of the node degree (i.e the maximum and minimum degree are of the same order.) Second, we use flow-based arguments to prove our theorems. To answer Question 1 we use a flow to bound the conductance $\Phi$ of the graph [SJ89] which, in turn, bounds the spectral gap and the mixing time. For Question 2 we use a flow to bound the resistance $R$ of the graph [DS84] which, in turn, bounds the cover time.

1.3.2 Restricted Delaunay Triangulation in Random Geometric Graphs

In this section we focus on another property of random geometric graphs which mostly has implication to routing. Our goal is to find a special subgraph, the Restricted Delaunay Graph [GGH01, LCW02], that permits an efficient routing based only on local information. The tasks of topological control and routing in sensor networks have been studied extensively and, in particular, lead to the advent of geo-routing [BMS99, KK00]. In geo-routing the assumption is that each node knows its own location (i.e. its coordinates) and the location of the destination to which it wants to deliver a message. The goal then is to find an efficient route from source to destination using only the local information available at each node and limited memory. A restricted Delaunay graph of a geometric graph is a planar sub-graph which is also a spanner: the shortest path between any pair of points is at most a constant factor longer than the shortest path in the original graph. It has been shown that using this sub-graph an efficient geo-routing can be obtained [GGH01, LCW02]. The question then is whether such sub-graph can be computed efficiently in random geometric graphs. More formally:

**Question 3.** For $r = \Theta(r_{\text{con}})$ can we, efficiently and locally, construct a Re-
We show that, while for general graphs this construction requires $\Theta(n)$ messages, $O(\sqrt{n \log n})$ messages suffices in $\mathcal{G}(n, r)$. We further present a novel algorithm that achieves this bound. Our algorithm exhibits two unique features which results in reduced message count. First, it requires only one round of communication and second, only ”problematic” nodes need to send messages.

### 1.3.3 Random Distance Graphs

The critical parameter for connectivity also played a major role in the study of Bernoulli random graphs $\mathcal{B}(n, p)$ (recall that $p$ is the probability with which each edge is chosen independently at random). In what may seem surprising, it has been shown that, for connectivity, both graphs have closely related critical thresholds for the radius and the edge probability. In particular if $\pi r^2 = \frac{\log n + \gamma_n}{n}$ then $\mathcal{G}(n, r)$ is connected w.h.p.iff $\gamma_n \to +\infty$ and disconnected w.h.p.iff $\gamma_n \to -\infty$ [Pen97, GK98] and likewise, if $p = \frac{\log n + \gamma_n}{n}$ then $\mathcal{B}(n, p)$ is connected w.h.p.iff $\gamma_n \to +\infty$ and disconnected w.h.p.iff $\gamma_n \to -\infty$ [ER59, Bol85]. This gives rise to the natural question:

**Question 4.** How can one explain the similarities in the connectivity threshold between these two graphs? Is there a general model that captures both $\mathcal{G}(n, r)$ and $\mathcal{B}(n, p)$ and exhibits the same connectivity threshold?

Note that these two graphs have quite diverse characteristics: in $\mathcal{B}(n, p)$ nodes appear to be only ”place holders” for random edges selection, while in $\mathcal{G}(n, r)$ nodes are assigned some properties (i.e. coordinates $x$ and $y$), on which the existence of edges depends. This leads to major differences in structure. For example, in the connectivity regime, $\mathcal{B}(n, p)$ has small diameter, and no small
cliques while $G(n,r)$ has large diameter and many small cliques.

Another important distinction between these graphs is manifested in the conditional probability that an edge $\{i,j\}$ exists given the existence of edges $\{k,i\}$ and $\{k,j\}$ (i.e. $P(\{i,j\} | \{i,k\}, \{j,k\})$). In $B(n,p)$ all three events are independent so $P(\{i,j\} | \{i,k\}, \{j,k\}) = P(\{i,j\})$, while in $G(n,r)$ the three events are not independent and $P(\{i,j\} | \{i,k\}, \{j,k\}) \gg P(\{i,j\})$.

This lack of independence, also called "locality", is one of the two properties that define Small World networks [WS98]. For example, in social networks two individuals that have a common friend are more likely to be friends with each others than two individuals chosen at random from the population; this intuition is confirmed by a wide range of real life data [WS98, NWS02]. The other property of small world networks is the existence of a small average path length between any two nodes and is commonly referred to as "six degrees of separation". Interestingly, in the connectivity regime $G(n,r)$ displays locality but not small path length, and $B(n,p)$ displays small path length but not locality, so neither graphs are adequate to model social networks. The following question arises:

**Question 5.** Does a general class of random graph exist, one that contains both $G(n,r)$ and $B(n,p)$ as special pathological cases, but whose typical case is adequate to model social networks?

To answer these questions we define a new class of random graphs, Random Distance Graphs. A random distance graph, $D(n,g)$, results from the following process: First, place $n$ nodes uniformly at random on the unit disk. Second, for each pair of nodes $i,j$ with distance $d(i,j)$ place an edge between $i$ and $j$ independently from all other edges with probability $p_{ij} = g(d(i,j))$.

Intuitively, we add edges between nodes as a function of their distance and therefore the underlying structure of the graph depends on the connection func-
tion $g$. Here, we primarily consider a specific type of function, a step function, $g^\alpha_r$, with parameters $r$ and $\alpha$. This function creates edges with probability $\alpha = \alpha(n)$ for nodes at distance less or equal to a radius $r = r(n)$ (short edges), and with probability $\beta = \beta(n)$ for nodes at distance larger than $r$ (long edges) where we choose $\beta(\alpha)$ as a function of $\alpha$.

We define $\beta(\alpha)$ in such a way that $\mathcal{D}(n, g^\alpha_r)$ captures $\mathcal{G}(n, r)$ and $\mathcal{B}(n, p)$ in two extremes: when $\alpha = 1$, $\beta = 0$, and we have $\mathcal{D}(n, g^1_r) \equiv \mathcal{G}(n, r)$, a random geometric graph. (where $\equiv$ stands for having the same generating process.) On the other extreme, when $\alpha = \pi r^2$, $\beta = \alpha$ and edges do not depend anymore on the distance and every edge is chosen independently with probability $p = \alpha = \beta$, so $\mathcal{D}(n, g^{\pi r^2}_r) \equiv \mathcal{B}(n, p = \pi r^2)$ which is the original random Bernoulli graph (See Fig 1.1 for clarifications).

Using this model we are able to answer Question 4 and prove that for $\pi r^2 = \frac{\log n + \gamma n}{n}$ and $\pi r^2 \leq \alpha \leq 1$, $\mathcal{D}(n, g^\alpha_r)$ is connected w.h.p. iff $\gamma_n \to \infty$ and is disconnected w.h.p. iff $\gamma_n \to -\infty$.

Regarding Question 5, we prove that in the evolution stage of connectivity and for a wide range of values for $\alpha$, $\mathcal{D}(n, g^\alpha_r)$ exhibits both the desired properties: it has small diameter and locality, which makes it a small world graph.
In addition, for a different connection function we present a distance graph that displays the following properties \(w.h.p.\): It is connected, has an average degree of \(\Theta(\log n)\), exhibits locality, has a small diameter and enables short local routing\(^3\).

1.4 Organization

Each Chapter is organized to allow independent reading as much as possible, it starts with a short introduction and statement of the results, follows with sections for preliminaries, results and proofs, notes and related work and ends with conclusions. We provide a list of notation at the beginning of the dissertation for more convenient reading. The dissertation is organized as follows: Chapter 2 deals with random walks on random geometric graphs and proves the theorems regarding the cover time and mixing time on these graphs. In Chapter 3 we present an algorithm to compute the restricted Delaunay triangulation, prove its correctness and bound the number of messages. We introduce random distance graphs in Chapter 4. We show that these graphs capture both random geometric graphs and Bernoulli graphs and prove Theorems regarding the properties of these graphs for different parameters settings. Chapter 5 present simulation results from early phase of this work. It gives an application example for random-walk-based algorithm and empirical results for its efficiency and robustness.

\(^3\)A graph has the property of short local routing only if there exist an algorithm that achieves short routing using only local information. (\(D(n, g^\alpha)\) doesn’t have short local routing.)


CHAPTER 2

Random Walks on Random Geometric Graphs

2.1 Introduction

In this chapter we study the existence of critical radii for properties of optimal cover time and rapid mixing. In particular, we study the existence of a radius $r_{\text{opt}}$ that will guarantee with high probability that $G(n, r)$ with $r \geq r_{\text{opt}}$ has *optimal cover-time* and a radius $r_{\text{rapid}}$ that will guarantee with high probability that $G(n, r)$ with $r \geq r_{\text{rapid}}$ is rapid mixing. Optimal cover time is cover time of $\Theta(n \log n)$ [Fei95a], the same order as the complete graph. We show that such thresholds do in did exist, and, surprisingly, the threshold for optimal cover time occurs at a radius $r_{\text{opt}} = \Theta(r_{\text{con}})$. On the other hand, $r_{\text{rapid}} = w(r_{\text{con}})$, and, in particular, the radius required for rapid mixing is $r_{\text{rapid}} = \Theta(1/\text{poly}(\log n))$.

The main contribution of this chapter is in giving new tight theoretical bounds on the cover time and sharp threshold width associated with cover time for random geometric graphs. Our main result can be formalized as follows:

**Theorem 2.1** (Cover Time of RGG). For $c > 1$, if $r^2 \geq \frac{c8\log n}{n}$, then w.h.p.\footnote{Event $E_n$ occurs with high probability if probability $P(E_n)$ is such that $\lim_{n \to \infty} P(E_n) = 1$.} $G(n, r)$ has cover time $\Theta(n \log n)$. If $r^2 \leq \frac{\log n}{\pi n}$, then $G(n, r)$ has infinite cover time with positive probability (bounded away from zero).

Our result has important implications for applications. Corollaries to our
result is that both the partial cover time [AB04] - the expected time taken by a random walk to visit a constant fraction of the nodes, and the blanket time [WZ96] - the expected time taken by a random walk to visit all nodes with frequencies according to the stationary distribution, are optimal for random geometric graphs. This demonstrates both the efficiency and quality of random walk approaches and certain token-management schemes for some ad-hoc and sensor networks [DSW02, KDG03, AB04].

The second major contribution is bounding the mixing-time and spectral gap of random geometric graphs:

**Theorem 2.2 (Mixing Time of RGG).** Radius $r = \Omega(1/\text{poly}(\log n))$ is w.h.p. necessary and sufficient for $G(n, r)$ to be rapidly mixing.

Note that the bounds on the cover-time in Theorem 2.1 improve upon bounds on the cover time obtainable via Theorem 2.2 as cover time can be bounded by the spectral gap [BK89]. In particular, the spectral gap method and Theorem 2.2 only guarantees optimal cover time of $G(n, r)$ for $r = \Theta(1)$.

The techniques we use to prove our results rely on two main features. First, we show that random geometric graphs are *geo-dense* (to be define formally later), that is geometric graphs that have desirable properties of uniform node distribution across the unit square and regularity on the node degree. In particular, in geo-dense graphs every *bin* larger than a certain size has the number of nodes inside it proportional to its size. Second, we use flow-based arguments to prove our theorems. In both cases, bins are the building blocks in the flow constructions, and we use the fact that for certain size bins all the nodes inside it form a clique. In the proof of Theorem 2.1 we use a flow to bound the resistance $R$ of the graph [DS84] which in turn bounds the cover time. In the proof of Theorem 2.2 we use a flow to bound the conductance $\Phi$ of the graph [SJ89], which in turn
bounds the spectral gap and the mixing time.

2.2 Preliminaries

2.2.1 Markov chains and the Simple Random Walk

The probabilistic rules by which a random walk operates is defined by the corresponding Markov chain. Let $\mathcal{M}$ be a Markov chain over state space $\Omega$ and probability transition matrix $P$ (i.e $P(x, y)$ is the probability to move from $x$ at time $t$ to $y$ at time $t+1$). In such terms, the stationary distribution of $\mathcal{M}$, if such exists, is then defined as the unique probability vector $\Pi$ such that

$$\Pi P = \Pi$$

A primary motivation in considering a random walk approach as opposed to a deterministic protocol is the simplicity and locality of computation. So, if the random walk is currently at state $q \in \Omega$, then the simplest probabilistic rule by which to choose the next state is simply to choose a state uniformly at random from among the set of states accessible from $q$. We call this set the neighbors of $q$. We call the Markov chain $\mathcal{M} = (\Omega, P)$ corresponding to such a random walk the simple random walk. Note that we may just as well define such $\mathcal{M}$ by its underlying graph $G = (V, E)$ where $V = \Omega$ and states are represented by nodes. For such $G$, for any node $v \in V$, let $\delta(v)$ denote the degree of $v$, that is the number of neighbors of $v$ in $G$ and let $P(v, u) = \frac{1}{\delta(v)}$ for $(v, u) \in E$ and 0 otherwise. It is well known that the simple random walk $\mathcal{M} = (\Omega, P)$ over an undirected connected graph $G = (V, E)$ has a stationary distribution $\Pi$ such that, for any node $q \in V$ [Lov96],

$$\Pi(q) = \frac{\delta(q)}{2m}$$ (2.1)
where \( m = |E| \). Further, when the underlying graph \( G \) is regular, that is when there is \( d \) such that for all \( q \) in \( \mathcal{M} \), \( \delta(q) = d \), the stationary distribution is the uniform distribution \([\text{Lov}96]\)

\[
\Pi(q) = \frac{d}{2m} = \frac{1}{n} \quad \forall q \in \Omega
\]

where \( n = |\Omega| = |V| \). It is also easy to confirm that the Markov chain corresponding to the simple random walk is reversible, that it satisfies the detailed balance condition with respect to \( \Pi \)

\[
Q(v, u) = \Pi(v)P(v, u) = \Pi(u)P(u, v) = Q(u, v) \quad \forall v, u \in V
\]

and in particular

\[
Q(u, v) = \frac{1}{2m} \quad \forall \{u, v\} \in E
\]

If \( P \) is also aperiodic (i.e, \( G \) is non-bipartite, which we assume true in our case\(^2\)) then the chain is ergodic and the distribution of the states at time \( t \) approaches \( \Pi \) as \( t \to \infty \), regardless of the starting state.

### 2.2.2 Mixing Time and the Spectral Gap \((1 - \lambda_1)\)

The efficiency with which a random walk of \( \mathcal{M} \) may be used to sample over state space \( \Omega \) with respect to the stationary distribution \( \Pi \) is precisely given by the rate at which the distribution of the states at time \( t \) converges to \( \Pi \) as \( t \to \infty \). In order to speak of convergence of probabilities, one must have a notion of distance over time. Let \( x \) be the state at time \( t = 0 \) and denote by \( P^t(x, \cdot) \) the distribution of the states at time \( t \). The variation distance at time \( t \) with respect to the initial state \( x \) is defined to be \([\text{Sin}92]\)

\[
\Delta_x(t) = \max_{S \subseteq \Omega} |P^t(x, S) - \Pi(S)| = \frac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \Pi(y)|
\]

\(^2\)One odd length cycle is sufficient to guarantee that \( G \) is non-bipartite.
The rate of convergence to stationary may be measured by the *mixing time*, the function \[ \tau_x(\epsilon) = \min\{t \mid \Delta_x(t') \leq \epsilon, \forall t' \geq t\} \]
which intuitively is the minimum number of steps \( t \) required, starting from node \( x \), to guarantee that for any node \( y \) the probability of being at \( y \) after \( t \) or more steps is at most \( \epsilon \) away from the probability of being at \( y \) under the stationary distribution (i.e. \( \Pi(y) \)). A chain \( \mathcal{M} \) is considered *rapidly mixing* iff \( \tau_x(\epsilon) \) is \( O(\text{poly}(\log(n/\epsilon))) \). For \( \mathcal{M} \) to be used for efficient sampling (according to its stationary distribution), we want \( \mathcal{M} \) to be rapidly mixing.

As the stationary distribution \( \Pi \) is defined to be such that \( \Pi P = \Pi \), it corresponds to the eigenvalue 1 = \( \lambda_0 \) of \( P \). Let the rest of the eigenvalues of \( P \) in decreasing order be: \( 1 = \lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{n-1} \geq -1 \). Since \( \mathcal{M} \) is ergodic \( \lambda_{n-1} > -1 \), and it is well known that the rate of convergence to \( \Pi \) is governed by the second largest eigenvalue in absolute value \( \lambda_{\text{max}} = \max\{\lambda_1, |\lambda_{n-1}|\} \), and in particular by the *spectral gap* \( 1 - \lambda_{\text{max}} \) [Sin92]:

**Proposition 2.3.** For an ergodic Markov chain, the quantity \( \tau_x(\epsilon) \) satisfies

\[
(i) \quad \tau_x(\epsilon) \leq (1 - \lambda_{\text{max}})^{-1}(\ln \Pi(x)^{-1} + \ln \epsilon^{-1})
\]

\[
(ii) \quad \max_{x \in \Omega} \tau_x(\epsilon) \geq \frac{1}{2} \lambda_{\text{max}}(1 - \lambda_{\text{max}})^{-1} \ln(2\epsilon)^{-1}
\]

As we want the starting state of a random walk to be arbitrary, the statement above implies that a large spectral gap \( 1 - \lambda_{\text{max}} \) is both a necessary and sufficient condition for rapid mixing. In practice the smallest eigenvalue, \( \lambda_{n-1} \), is not important since by simply adding self-loop probabilities of \( \frac{1}{2} \) ("staying" probability) at each node, we create a new chain that has the same stationary distribution, and its eigenvalues, \( \{\lambda'_i\} \), are similarly ordered and satisfy \( \lambda'_{n-1} > 0 \).
and $\lambda'_{\text{max}} = \lambda'_1 = \frac{1}{2}(1 + \lambda_1)$ [SJ89]. This shows that it is sufficient to bound $\lambda_1$ to prove rapid mixing.

### 2.2.3 Cover Time, Partial Cover Time and Blanket Time

#### 2.2.3.1 Cover Time

The *cover time* $C_G$ of a graph $G$ is the expected time taken by a simple random walk on $G$ to visit all nodes in $G$. The *cover time* of graphs and methods of bounding it have been extensively investigated [Mat88, Ald89, CRR89, BK89, Zuc90, AKL79]. Several bounds on the cover times for particular classes of graphs have been obtained with many positive results [CRR89, BK89, Jon98, JS00, CF03]. Results for the cover time of specific graphs vary from the *optimal cover time* of $\Theta(n \log(n))$ associate with the complete graph to the worst case of $\Theta(n^3)$ associate with the lollipop graph [Fei95b, Fei95a]. The known best cases correspond to dense, highly connected graphs such as the complete graph, $d$-regular graphs with $d > \frac{n}{2}$, and the hypercube which is the $d$-dimensional mesh with $d = \log(n)$. As one may find surprising, any $d$-dimensional mesh for $d \geq 3$ has *optimal cover time*; note that these graphs are regular with only constant degree, 6 in the case of the 3-dimensional mesh. However, the case of dimension $d = 2$ was considered to be the hardest one to analyze [AF99] and its cover time has been shown to be $\Theta(n \log^2(n))$ (i.e. not optimal). When connectivity decreases and bottlenecks exist in the graph, the cover time increases. Two examples are the line graph which has cover time $\Theta(n^2)$ and the lollipop graph mentioned above.

Let $G(V, E)$ be the graph on which the simple random walk is performed and let $n = |V|$ and $m = |E|$. For $v \in V$ let $C_v$ be the expected number of steps for the simple random walk starting at $v$ to visit all the nodes in $G$, and the cover time of $G$ is $C_G = \max_v C_v$. Let $H_{uv}$ be the *hitting time*, the expected time for
a random walk starting at $u$ to arrive to $v$ for the first time, and let $C_{uv}$ be the *commute time*, the expected time for a random walk starting at $u$ to first arrive at $v$ and then return to $u$. Let $H_{max}$ be the maximum $H_{uv}$ over all ordered pairs of nodes and let $H_{min}$ to be defined similarly. The following theorem provides bounds on the cover time $C_G$ in term of $H_{max}$ and $H_{min}$.

**Theorem 2.4** (Matthews’ theorem [Mat88]). *For any graph $G$,*

$$H_{min} \cdot H_n \leq C_G \leq H_{max} \cdot H_n$$

*where $H_k = \ln(k) + \Theta(1)$ is the $k$-th harmonic number.*

Notice that this bound is not always tight, since in the line, for instance, we have $C_G = H_{max}$.

### 2.2.3.2 Partial Cover Time

The Partial Cover Time (PCT) of graph $G$ is define formally as follow: For $0 \leq c \leq 1$, let $C_G(c)$ be the expected time taken by a simple random walk on $G$ to visit $\lfloor cn \rfloor$ of the nodes of $G$. In the next Lemma we prove that as long as we want to cover a constant fraction of the graph we can reduce Matthew’s bound by an order of $\log(n)$ so that the bounds becomes linear in $H_{max}$. Intuitively, this means that, on sufficiently large graphs, almost all the time used by a walk to cover the entire graph is spent trying to reach the last $\log(n)$ nodes.

**Lemma 2.5** (Partial Cover Lemma). *For any graph $G$, and $0 \leq c \leq \frac{n-1}{n}$*

$$C_G(c) < 2 \cdot H_{max} \cdot \left\lceil \log_2\left(\frac{1}{1-c}\right) \right\rceil = O(H_{max})$$

This implies the following interesting results: for graphs in which $H_{max} = n$, PCT becomes linear in $n$ and we consider it to be *optimal partial cover*; known
graph of this type are the complete graph, the Star, the hypercube, and the 3-
dimensional mesh. Covering \( x \) percent of the nodes in the 3-dimensional mesh,
for example, is linear in \( n \) while the cover time is \( O(n \log(n)) \). On the other
hand, for the 2-dimension grid, the maximum hitting time is \( O(n \log(n)) \) [Zuc90]
so PCT becomes \( O(n \log(n)) \).

**Proof of Partial Cover Lemma.** For simplicity of exposition, we prove
the result for special cases where \( c \) is of the form \( c_l = \frac{2^l-1}{2^l}, l \geq 0 \). The general case
will follow directly by taking the minimum \( l \) such that \( \frac{2^l-1}{2^l} \geq c \). Let \( n = 2^l k + 1 \)
and \( c_l = \frac{2^l-1}{2^l} \), and the proof will be by induction on \( l \). Let \( c_l = \frac{2^l-1}{2^l} \). For a fixed
random walk, let \( \gamma_l \) be the minimum number of steps to visit more than \( c_l \cdot n \) of
the nodes. Let \( \alpha_v \) be the time (step number) when node \( v \) is visited for the first
time and let \( S_l = \{ v \in V \mid \alpha_v \geq \gamma_l \} \) be the set of all nodes visited at time \( \gamma_l \) or
later. Note that \( |S_l| = n - \lfloor c_l n \rfloor = 2^l - l k + 1 \) and that \( S_0 = V \) is the set of all
nodes.

**Base Cases:** The \( l = 0 \) case is trivially true. For \( l = 1 \) and \( c_1 = \frac{1}{2} \), the proof
is stated in Lemma 2.8 [Lov96] that proves that visiting more than half of the
nodes take less then \( 2H_{\text{max}} \) steps. We will follow this proof here:

Let \( k' = 2^{l-1} k \) and so \( n = 2k' + 1 \) is odd. The time \( \gamma_1 \) when we reach more
than half of the nodes is the \( (k' + 1) \)-st largest of the \( \alpha_v \). Hence

\[
\sum_{v \in S_0} \alpha_v \geq (k' + 1) \gamma_1
\]

taking the expectation on both sides and noting that \( E(\alpha_v) \leq H_{\text{max}} \)

\[
C_G(c_1) = E(\gamma_1) \leq \frac{1}{k' + 1} \sum_{v \in S_0} E(\alpha_v) \\
\leq \frac{2k' + 1}{k' + 1} H_{\text{max}} \\
< 2H_{\text{max}} \log_2(\frac{1}{1-c_1}) = 2H_{\text{max}}
\]

19
Induction Step: Assume true for $L - 1 \geq l \geq 0$. We will prove true for $l + 1$.

Let $\alpha'_v$ be the number of steps until node $v$ is first visited after $\gamma_l$ steps and $k' = 2^{L-l}k$.

$$\sum_{v \in S_l} \alpha_v = \sum_{v \in S_l} (\gamma_l + \alpha'_v) \geq k'\gamma_l + (k' + 1)\gamma_{l+1}$$

Taking the expectation, the l.h.s is:

$$E \left[ \sum_{v \in S_l} (\gamma_l + \alpha'_v) \right] = \sum_{v \in S_l} E[\gamma_l + \alpha'_v]$$

$$= (2k' + 1)C_G(c_l) + \sum_{v \in S_l} E[\alpha'_v]$$

The r.h.s will be:

$$E [k'\gamma_l + (k' + 1)\gamma_{l+1}] = k'C_G(c_l) + (k' + 1)C_G(c_{l+1})$$

Putting it together:

$$k'C_G(c_l) + (k' + 1)C_G(c_{l+1}) \leq (2k' + 1)C_G(c_l) + \sum_{v \in S_l} E[\alpha'_v]$$

Using the induction assumption and $E(\alpha'_v) \leq H_{\max}$ we get:

$$C_G(c_{l+1}) \leq 2H_{\max} \log_2\left(\frac{1}{1 - c_l}\right) + \frac{2k' + 1}{k' + 1}H_{\max}$$

$$< 2H_{\max} \log_2\left(\frac{1}{1 - c_l}\right) + 2H_{\max} \log_2(2)$$

$$= 2H_{\max} \log_2\left(\frac{1}{1 - c_{l+1}}\right)$$

$\square$

2.2.3.3 Blanket Time

The blanket time[KKL02] $B_G$ of a graph $G$ is the expected time for a random walk to visit every node of $G$ within a constant factor of the number predict by
the stationary distribution. Clearly $B_G \geq C_G$, surprisingly it was conjectured in [WZ96] that $B_G = O(C_G)$. The conjecture was proven for the case where the cover time $C_G$ and the maximum hitting time $H_{\text{max}}$ differ by a logarithmic factor. When $B_G$ is of the same order as $C_G$ and the cover time is optimal cover time (i.e $O(n \log n)$) we say that the blanket time is optimal blanket time.

2.2.4 Conductance

A well-known method for bounding $\lambda_1$ to prove rapid mixing when the underlying graph has a geometric interpretation is a Conductance argument [JS97]. This is the method we shall use, as random geometric graphs have a strong geometric interpretation. Intuitively, one would expect that when the graph that underlies the Markov chain $\mathcal{M}$ doesn’t have bottlenecks, the lower the probability of getting stuck in any particular set of states, and thus the more rapidly mixing $\mathcal{M}$ is. The property of “no bottlenecks” is formalized in a continuous manner with the notion of conductance.

The conductance of a reversible Markov chain $\mathcal{M}$ is defined by [SJ89]

$$\Phi = \Phi(\mathcal{M}) = \min_{S \subset \Omega, \pi(S) \leq 1/2} \frac{Q(S, \overline{S})}{\Pi(S)}$$

where $\overline{S} = \Omega - S$, $\Pi(S)$ is the probability density of $S$ under the stationary distribution $\Pi$, and $Q(S, \overline{S})$ is the sum of $Q(v, u)$ over all $(v, u) \in S \times \overline{S}$.

In graph-theoretic terms, the conductance of $\mathcal{M}$ is the minimum over all subsets $S \subset \Omega$ of the ratio of the weighted flow across the cut $\text{Cut}(S, \overline{S})$ to the weighted capacity of $S$. The higher the conductance of $\mathcal{M}$, there are fewer bottlenecks in $\mathcal{M}$, and the more rapidly mixing $\mathcal{M}$ is. This intuition is confirmed by the following theorem:

**Theorem 2.6** ([Sin92]). *The second eigenvalue $\lambda_1$ of a reversible Markov chain*
\( \mathbb{M} \) satisfies

\[ 1 - 2\Phi \leq \lambda_1 \leq 1 - \frac{\Phi^2}{2} \]

The above Theorem along with Proposition 2.3 yield the following powerful corollary bounding the mixing time via conductance:

**Corollary 2.7** ([JS97]). Let \( \mathbb{M} \) be a finite, reversible, ergodic Markov chain with loop probabilities \( P(x, x) \geq \frac{1}{2} \) for all states \( x \). Let \( \Phi \) be the conductance of \( \mathbb{M} \). Then, for any initial state \( x \), the mixing time of \( \mathbb{M} \) satisfies

\[ \tau_x(\epsilon) \leq 2\Phi^{-2}(\ln \Pi(x)^{-1} + \ln \epsilon^{-1}) \]

Regarding cover time, the best cases are usually also correspond to dense, highly connected graphs. When connectivity decreases and bottlenecks exist in the graph, the cover time increases, therefore, intuitively, one would anticipate a relationship between the spectral gap \((1 - \lambda_1)\) and small cover time. In confirmation of this intuition, a bound for the cover time for regular graphs \( G \) that is based on the spectral gap \((1 - \lambda_1)\) is as follows:

**Theorem 2.8** ([BK89, AF99]). For regular graph \( G = (V, E) \) with \( n = |V| \) and second largest eigenvalue \( \lambda_1 \) the cover time of \( G \) is bounded as follows:

\[ C_G = O(n \log n/(1 - \lambda_1)) \]

In the next subsection we present, yet another way to bound the Cover Time; via resistance.

### 2.2.5 Bounding The Cover Time via Resistance

For a graph \( G = (V, E) \) with \( |V| = n, |E| = m \), the electrical network \( \mathcal{E}(G) \) is obtained by replacing each edge \( e \in E \) with a 1 Ohm resistor, and this is the
network we analyze when we speak of the resistance properties of $G$. For $u, v \in V$ let $R_{uv}$ be the effective resistance between $u$ and $v$: the voltage induced between $u$ and $v$ by passing a current flow of one ampere between them. Let $R$ be the electrical resistance of $G$: the maximum effective resistance between any pair of nodes [DS84].

Chandra et al. [CRR89] proved the following equality for the commute time $C_{uv}$ in terms of the effective resistance $R_{uv}$:

**Theorem 2.9 ([CRR89]).** For any two vertices $u$ and $v$ in $G$ the commute time

$$C_{uv} = 2mR_{uv}$$

Using this direct relation between resistance and random walks and Matthews’ theorem (Theorem 2.4) [Mat88] the authors introduced the following bound on the cover time $C_G$ for any graph $G$ with $n$ nodes and $m$ edges, where $R$ is the the electrical resistance of the graph [CRR89]:

$$mR \leq C_G \leq O(mR \log n) \quad (2.2)$$

We can also bound $H_{\text{max}}$, the maximum hitting time by $R$. Since $H_{uv} \leq C_{uv}$ it we have:

$$H_{\text{max}} = \max_{u,v \in V} H_{uv} \leq \max_{u,v \in V} C_{uv} = \max_{u,v \in V} 2mR_{uv} = 2mR$$

Lemma 2.5 bounds the partial cover time by $H_{\text{max}}$, so combining: for a constant $0 \leq c < 1$

$$C_G(c) \leq O(mR) \quad (2.3)$$

Thus, by bounding the resistance $R$ we may obtain tight bounds on the cover time $C_G$ through (2.2) and on the partial cover time through (2.3).
A powerful method used to bound resistance is by bounding the power of a current flow in the network. The following definitions and propositions from the literature \cite{CRR89, DS84, Syn51} help to formalize this method.

**Definition 2.1** (Power of a flow). Given an electrical network $(V, E, \rho)$, with resistance $\rho(e)$ for each edge $e \in E$, a flow $c$ from a source $u$ to a sink $v$ is a function from $V \times V$ to $\mathbb{R}$, having the property that $c(x, y) = 0$ unless $\{x, y\} \in E$, and $c$ is anti-symmetric, i.e., $c(x, y) = -c(y, x)$. The net flow out of a node will be denoted $c(x) = \sum_{y \in V} c(x, y)$ and $c(x) = 0$ if $x \neq u, v$. The flow along an edge $e$ is $c(e) = |c(u, v)|$. The power $P(c)$ in a flow is $P(c) = \sum_{e \in E} \rho(e)c^2(e)$. A flow is a current flow if it satisfies Kirchoff's voltage law, i.e., for any directed cycle $x_0, x_1, \ldots, x_{k-1}, x_0$, $\sum_{i=0}^{k-1} c(x_i, x_{i+1 \text{ mod } k}) \cdot \rho(x_i, x_{i+1 \text{ mod } k}) = 0$.

**Proposition 2.10.** [Thomson Principle \cite{DS84, Syn51}] For any electrical network $(V, E, \rho)$ and flow $c$ with only one source $u$, one sink $v$, and $c(u) = -c(v) = 1$ (i.e a unit flow), we have $R_{uv} \leq P(c)$, with equality when the flow is a current flow.

Finally,

**Proposition 2.11.** [Rayleigh’s Short/Cut Principle \cite{DS84}] Resistance is never raised by lowering the resistance on an edge, e.g. by ”shorting” two nodes together, and is never lowered by raising the resistance on an edge, e.g. by ”cutting” it. Similarly, resistance is never lowered by ”cutting” a node, leaving each incident edge attached to only one of the two ”halves” of the node.

In the following subsection we give an example that demonstrates these ideas. We consider a finite 2-dimensional grid, not the torus that was analyzed in the work mentioned above. To the best of our knowledge, all previous work on meshes has only considered the torus.
Let $G$ be a $\sqrt{n} \times \sqrt{n}$ 2-dimensional grid of $n$ nodes, and for convenience denote $s = \sqrt{n}$. Let $u_0$ be the point at the origin, $(0,0)$, and $u_n$ the point at $(\sqrt{n}, \sqrt{n})$.

Claim 2.12. The resistance between $u_0$ and $u_n$ in $G$ is $\Theta(\log n)$

Proof. The proof is similar to that for the mesh [CRR89] (6.3, 8.7), and we state parts of it in a way that is generalizable to other proofs.

The upper bound is found using Thomson Principle and by constructing a unit flow $c$ from $u_0$ to $u_n$, where $u_0$ is the only source and $u_n$ is the only sink. Let $V_l$ denote the set of nodes at Manhattan distance (i.e. $L_1$ distance) $l$ from the origin $u_0$. Note that $|V_l| = l + 1$ for $0 \leq l \leq s - 1$ and $|V_l| = l_u + 1$ for $s \leq l \leq 2s - 1$ where $l_u$ is the distance from $u_n$ (i.e. $l_u = 2s - 2 - l$). Let $E_l$ denote the set of edges between $V_l$ and $V_{l+1}$. Note that, for any $l$, the set $V_l$ can be viewed as a contour of equi-potential nodes. So, we refer to $V_l$ as layer $l$, and the flow that we construct can be seen as progressing uniformly through the $E_l$'s. The first layer $V_0$ is only $u_0$, which after the layers grows linearly up to layer $V_{s-1}$ and then shrink back, until layer $V_{2s-1}$ which is only $u_n$.

The **outgoing** flow from a point $(x, y) \in V_l$ where $0 \leq l \leq s - 2$ is $\frac{x+1}{(l+2)(l+1)}$.
to point \((x + 1, y)\) and \(\frac{y+1}{(l+2)(l+1)}\) to point \((x, y + 1)\) (see Fig. 2.1 (A)). For points \((x, y)\) ∈ \(V_l\) where \(s \leq l \leq 2s - 1\) the incoming flow from \((x-1, y)\) is \(\frac{s-x+1}{(2s-l+2)(2s-l+1)}\) and \(\frac{s-y+1}{(2s-l+2)(2s-l+1)}\) from the point \((x, y - 1)\) (see Fig. 2.1 (B)). Clearly, the flow out of \(u_0\) is 1 and the flow into \(u_n\) is 1. Moreover, it is easy to check that the net flow at each point \((x, y) \neq u_0, u_n\) is 0.

Now by Thomson Principle we have \(R_{u_0u_n} \leq P(c) = \sum_{e \in E} r(e)c^2(e)\). Note that when \(0 \leq l \leq s - 2\), \(|E_l| = 2(l + 1)\), and \(\forall e \in E_l, c(e) \leq \frac{1}{1+l}\). When \(s \leq l \leq 2s - 1\), \(|E_{l-1}| = 2(l_u + 1)\), and \(\forall e \in E_{l-1}, c(e) \leq (1/(l_u + 1))\). So,

\[
R_{u_0u_n} \leq P(c) = \sum_{e \in E} \rho(e)c^2(e) \\
= \sum_{l=1}^{s-2} \sum_{e \in E_l} O(1/l^2) + \sum_{l=s}^{2s-2} \sum_{e \in E_{l-1}} O(1/l_u^2)) \\
= \sum_{l=0}^{s-2} O(l)O(1/l^2) + \sum_{k=0}^{s-2} O(k)O(1/k^2)) \\
= 2 \sum_{l=0}^{s-2} O(1/l) \\
= O(\log n)
\]

The lower bound is by the “Short/Cut” Principle. For any \(l \geq 0\), shorting all vertices in \(V_l\) to each other we get the following:

\[
R_{u_0u_n} \geq \frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{2(s-1)} + \frac{1}{2(s-1)} + \frac{1}{2} + \ldots + \frac{1}{2} \\
= 2 \sum_{l=1}^{s-1} \frac{1}{2(s-1)} \\
= \Omega(\log n)
\]

Noting that in the 2 dimensional grid case \(m \approx 2n\) and using (2.2) we get the following bound on the cover time \(C = O(n \log^2 n)\).

26
2.3 Geo-dense Geometric Graphs

Our analytical results for random geometric graph are based on the “nice” properties that those graphs have. These properties include the uniformity of nodes distribution and the regularity of node degree. In the following section we will define this using the notion of *geo-dense* graph, that is, geometric graph (random or deterministic) with almost uniform node density across the unit square. In *geo-dense* graphs there are no large areas that fail to contain a sufficient number of nodes. We show that random geometric graphs are *geo-dense* and for radius $r_{\text{reg}} = \Theta(r_{\text{con}})$ all nodes have the same order degree $w.h.p$. To do so we introduce the notion of *bins*, equal size areas that partition the unit square. These bins will be our building blocks for future proofs as well.

A **geometric graph** is a graph $G(n, r) = (V, E)$ with $n = |V|$ such that the nodes of $V$ are embedded into the unit square with the property that $e = (u, v) \in E$ if and only if $d(u, v) \leq r$ (where $d(u, v)$ is the Euclidean distance between points $u$ and $v$).

**Definition 2.2.** Let $G(n, r(n))$ be a geometric graph. For a constant $\mu \geq 1$ we say that such a graph is $\mu$-**geo-dense** if every square bin of size $A \geq r^2/\mu$ (in the unit square) has $\Theta(nA)$ nodes.

We can claim the following on geo-dense graphs:

**Lemma 2.13.** Let $G(n, r)$ be a 2-geo-dense geometric graph with $V$ the set of nodes and $E$ the set of edges. Let $\delta(v)$ denote the degree (i.e number of neighbors) of $v \in V$. Then: (i) $\forall v \in V \ \delta(v) = \Theta(nr^2)$ and (ii) $m = |E| = \Theta(n^2r^2)$.

**Proof.** (i). First note that the geo-dense property guarantees that if we divide the unit square into square bins of size $\frac{r}{\sqrt{2}} \times \frac{r}{\sqrt{2}}$ each, then the number
of nodes in every bin will be $\Theta(nr^2)$. Since, for every such bin, the set of nodes in the bin forms a clique\(^3\), and every node $v \in V$ is in some bin, we have that $\delta(v) = \Omega(nr^2), \forall v \in V$. Similarly, when we divide the area into bins of size $r \times r$ every node may be connected to the nodes of at most nine bins (that is its own bin and the bordering bins), and we have that $\delta(v) = O(nr^2), \forall v \in V$. (ii) follow directly from (i). \hfill \Box

### 2.3.1 Geo-dense Random Geometric Graphs

To prove the geo-dense property for $G(n, r)$ we utilize the following lemma which seems to be folklore [WZ96] although we include a proof since we have not found a reference including a proof of the minimum condition.

**Lemma 2.14 (Balls in Bins).** For a constant $c > 1$, if one throws $n \geq cB \log B$ balls uniformly at random into $B$ bins, then w.h.p. both the minimum and the maximum number of balls in any bin is $\Theta(\frac{n}{B})$.

**Proof.** Let $n = cB \log B$ and note that when $n \to \infty$ then $B \to \infty$. The upper bound (maximum) of $\Theta(\frac{n}{B})$ on the number of balls in any bin is given in [MR95]. The following is proof of the lower bound: Fix a bin, say the first bin, and let $X_1$ denote the size of the first bin. Consider $Pr[X_1 = \frac{\log B}{c_1}]$ for some constant $c_1$.\(^4\)

\[
Pr[X_1 = \frac{\log B}{c_1}] = \left(\frac{cB \log B}{\log B/c_1}\right) \left(\frac{1}{B}\right)^{\log B/c_1} \left(1 - \frac{1}{B}\right)^{cB \log B - \log B/c_1} \\
\leq (ec_1)^{\frac{\log B}{c_1}} e^{\frac{\log B}{c_1} - cB \log B} \\
= B^{\frac{1}{c_1}} (c_1 e)^{\frac{\log B}{c_1} - B^{\frac{1}{c_1} - c}} \\
= B^{\frac{\log c_1 + \log e}{c_1}} + \frac{1}{c_1} + \frac{1}{B^c_1 - c}
\]

---

\(^3\)The maximum distance between any two nodes in the bin is less that $r$.

\(^4\)by using $(1 - \frac{1}{n})^r \leq e^{-r/n}$, $(\frac{n}{k})^k \leq (\frac{nc_1}{k})^k$ and $c_1 = e^{\log(c_1)}$.

---

28
Since we want for this probability to be \( \frac{1}{B^{1+\epsilon'}} \) for \( \epsilon' > 0 \) we need
\[
c - \left( \frac{\log c_1 + \log c}{c_1} + \frac{1}{c_1} + \frac{1}{Bc_1} \right) > 1
\]
Let \( c = 1 + \epsilon \) where \( \epsilon > 0 \) can be arbitrary small constant and so we need \( c_1 \) s.t
\[
1 + \epsilon - \left( \frac{\log c_1 + \log(1 + \epsilon)}{c_1} + \frac{1}{c_1} + \frac{1}{Bc_1} \right) > 1
\]
(2.4)
The following \( c_1 \) will satisfy (2.4)
\[
\frac{1}{c_1 - 1} + \frac{1}{B(c_1 - 1)} + \frac{\log c_1}{c_1 - 1} < \epsilon
\]
(2.5)
So it is clear that for any constant \( \epsilon > 0 \) small as we want, there exists a constant \( c_1 \) that will satisfy (2.5), let that constant be \( c^* \). Note easily that \( \Pr[X_1 = \frac{\log B}{c^*}] \geq \Pr[X_1 = \frac{\log B}{c^*} - Q] \) for any \( 0 \leq Q \leq \frac{\log B}{c^*} \). Therefore, we have that for large enough \( B \)
\[
\Pr[X_1 \leq \frac{\log B}{c^*}] \leq \left( \frac{\log B}{c^*} \right) \Pr[X_1 = \frac{\log B}{c^*}]
\]
\[
\leq \left( \frac{\log B}{c^*} \right) \frac{1}{B^{1+\epsilon'}}
\]
Finally to get the lower bound (minimum) for all bins, we use that the probability of the union of events is no more than their sum. Letting \( U \) denote the event that some bin has less than \( \frac{\log B}{c^*} \) balls:
\[
\Pr[U] \leq \sum_{i=1}^{B} \Pr[X_i \leq \frac{\log B}{c^*}]
\]
\[
= \sum_{i=1}^{B} \Pr[X_1 \leq \frac{\log B}{c^*}]
\]
\[
= B \frac{\log B}{B^{1+\epsilon'}}
\]
\[
= \frac{\log B}{c^*} = o(1)
\]
Therefore, with high probability every bin has at least \( \frac{\log B}{c'} = \Theta(\frac{n}{B}) \) balls. Now, clearly, choosing \( n > cB \log B \) can only increase the probability that every bin has at least \( \frac{\log B}{c'} = \Theta(\frac{n}{B}) \) balls. So, we are done. \( \square \)

Following Balls in Bins Lemma we can now make the claim about the geo-density of \( G(n, r(n)) \) precise:

**Lemma 2.15** (Geo-density of \( G(n, r) \)). For constants \( c > 1 \) and \( \mu \geq 1 \), if \( r^2 = \frac{cn \log n}{\mu} \) then w.h.p. \( G(n, r) \) is \( \mu \)-geo-dense, that is, any bin area of size \( r^2/\mu \) in \( G(n, r) \) has \( \Theta(\log n) \) nodes w.h.p.

**Proof.** Let an area of \( r^2/\mu \) be a bin. If we divide the unit square into such equal size bins we have \( B = \frac{n}{c \log n} \) bins. For the result to follow we check that Lemma 2.14 holds by showing that \( n \geq cB \log B \) for some constant \( c' > 1 \):

\[
B \log B = \frac{n}{c \log n} \log(\frac{n}{c \log n})
\]
\[
= \frac{n}{c \log n} (\log(n) - \log(c \log n))
\]
\[
= \frac{n}{c} - (\frac{n}{c \log n}) (\log(c \log n))
\]
\[
\leq n/c
\]

\( \square \)

Now combining the results of Lemmas 2.13 and 2.15 we can also claim the following about \( G(n, r(n)) \):

**Corollary 2.16.** For \( c > 1 \), if \( r^2 \geq \frac{c2 \log n}{\mu} \), then w.h.p. \( G(n, r) = (V, E) \) we have \( \delta(v) = \Theta(nr^2) \) \( \forall v \in V \), and \( m = |E| = \Theta(n^2 r^2) \).

Recall that the critical radius for connectivity \( r_{con} \) is s.t \( \pi r_{con}^2 = \frac{\log n}{n} \). We have just showed that for \( r_{reg} = \Theta(r_{con}) \) w.h.p. \( G(n, r_{reg}) \) will have the nice properties mentioned above. Note however, that even though \( G(n, r_{reg}) \) is geo-dense in our
terms, it is not a dense graph in graph theoretic terms (i.e a graph with $\Theta(n^2)$ edges), but is a sparse graph with expected number of $\Theta(n \log n)$ edges.

## 2.4 The Mixing Time of Random Geometric Graphs

In this section we demonstrate that for sufficiently large $n$, the conductance $\Phi$ of $\mathcal{G}(n, r)$ is $\Phi(\mathcal{G}(n, r)) = \Theta(r)$ with high probability, and we give a useful continuous approximation to $\Phi$ in Section 2.4.2. Based on the conductance results, we show that for $\mathcal{G}(n, r)$ to be rapidly mixing, radius at least $r_{\text{rapid}} = \Theta(1/\text{poly}(\log n))$ is necessary and sufficient.

### 2.4.1 Bounding the Conductance of $\mathcal{G}(n, r)$

The main result of this section is as follows:

**Theorem 2.17 (Conductance of RGG).** For $c > 1$, if $r^2 \geq \frac{c^4 \log n}{n}$, then w.h.p.

$$\Phi(\mathcal{G}(n, r)) = \Theta(r)$$

From Theorem 2.17, Theorem 2.6 and Corollary 2.7 we obtain these bounds:

**Corollary 2.18.** For $c > 1$, if $r^2 \geq \frac{c^4 \log n}{n}$, then w.h.p. the mixing time of $\mathcal{G}(n, r)$ is as follows:

1. $\tau_x(\epsilon) = O(r^{-2}(\ln n + \ln \epsilon^{-1}))$

2. $1 - \lambda_1 = \Omega(r^2)$ and $1 - \lambda_1 = O(r)$

Together with Proposition 2.3 (ii) we also obtain the necessary condition:

31
**Theorem 2.2.** Radius \( r = \Omega(1/poly(\log n)) \) is w.h.p. necessary and sufficient for \( G(n,r) \) to be rapidly mixing.

Now we may begin the proof of the main result of this section:

**Proof of Theorem 2.17.** Let \( \text{Cut}(S, \bar{S}) \) denote the cut size between \( S \) and \( \bar{S} \) in \( G(n,r) \) (the total number of edges crossing from \( S \) to \( \bar{S} \)). From Lemma 2.15 and corollary 2.16 w.h.p. \( G(n,r) \) is 4-geo-dense and "almost regular". We can observe that the minimum conductance is when we divide the area into two halves \( S \) and \( \bar{S} \) with \( \Pi(S) \approx \Pi(\bar{S}) \approx \frac{1}{2} \) and such that the length of the boundary between \( S \) and \( \bar{S} \) is minimized. Similarly to the the grid case the separation satisfying this is with a separating line \( l \) parallel to one of the axis. Let \( \text{Cut}_\Phi(S, \bar{S}) \) be the above cut, the one that minimizes \( \Phi(G(n,r)) \). Next we bound \( \text{Cut}_\Phi(S, \bar{S}) \) and show it to be \( \Theta(r^3n^2) \).

For the lower bound of \( \text{Cut}_\Phi(S, \bar{S}) \), partition the area into bins of size \( \frac{r}{2\sqrt{2}} \times \frac{r}{\sqrt{2}} \) as in Figure 2.2 (A). By the 4-geo-dense property w.h.p. the number of nodes in any bin is \( \Theta(nr^2) \). Notice that the set of nodes in any two horizontally adjacent bins (such as \( B_0 \) and \( B_1 \) in Fig 2.2 (A)) forms a clique. Therefore, to lower bound \( \text{Cut}_\Phi(S, \bar{S}) \), we are only considering the crossing edges within each separate such clique along the dividing line \( l \). Since there are at least \( \frac{\sqrt{2}}{r} \) cliques along the dividing line \( l \), and for each bin on the left side of \( l \) we have \( \Omega(n^2r^4) \) such edges crossing to the right of \( l \), we obtain the desired lower bound \( \text{Cut}_\Phi(S, \bar{S}) = \Omega(r^3n^2) \).

For the upper bound partition the area into bins of size \( r \times r \) as in Figure 2.2 (B). Note that for each edge \((u, v)\) crossing \( l \), \( v \) must be in some left bin \( B_0 \) adjacent to \( l \), and so \( u \) must be in one of three possible bins \( B_1, B_2, B_3 \) that are on the right of \( l \) and touching \( B_0 \) as shown in the picture. To upper bound \( \text{Cut}_\Phi(S, \bar{S}) \), we consider the maximum number of crossing edges from any \( r \times r \).
Figure 2.2: (A) Lower bound for the Conductance in $\mathcal{G}(n, r)$. (B) upper bound for the Conductance in $\mathcal{G}(n, r)$

sized bin $B_0$ in $S$ to three $r \times r$ sized bins $B_1, B_2$ and $B_3$ in $\bar{S}$. As there are $\frac{1}{r}$ such bins as $B_0$, and from the 4-geo-dense property, w.h.p. the number of nodes in any bin is $\Theta(nr^2)$, we get the desired upper bound as follows:

$$Cut_{\Phi}(S, \bar{S}) = O\left(\frac{1}{r} \cdot nr^2 \cdot 3nr^2\right) = O(r^3n^2)$$

So, combining the upper and lower bounds, we have that w.h.p.,

$$Cut_{\Phi}(S, \bar{S}) = \Theta(r^3n^2)$$

And, thus, by corollary 2.16, equation (2.1), and the definition of $P(x, y)$ we
complete the proof:

\[
\Phi(G(n, r)) = \min_{S \subseteq \Omega, 0 < \Pi(S) \leq 1/2} \frac{Q(S, \overline{S})}{\Pi(S)}
\]

\[
= \min_{S \subseteq \Omega, 0 < \Pi(S) \leq 1/2} \frac{1}{\Pi(S)} \sum_{x \in S, y \in \overline{S}} \Pi(x)P(x, y)
\]

\[
= \min_{S \subseteq \Omega, 0 < \Pi(S) \leq 1/2} \Theta(2) \sum_{x \in S} \Theta\left(\frac{1}{n}\right) \Theta\left(\frac{1}{nr^2}\right)
\]

\[
= \Theta(2) \frac{1}{\Theta(r^2 n^2)}
\]

\[
= \Theta(r^3 n^2) / \Theta(r^2 n^2)
\]

\[
= \Theta(r)
\]

\[\square\]

### 2.4.2 Continuous Approximation of Conductance

Follow Fig. 2.3 Let \( l \) be the dividing line. A point \( p \) in \( S \) that is at distance \( x < r \) from \( l \) neighbors the nodes in the gray area \( A \) in the Figure. The size of
$A$ is given by $\frac{1}{2}r^2(\theta - \sin \theta)$. (Observe that $\theta = \arccos(\frac{x}{r})$ and $A$ is a function of $x$.) So $p$ has an expected number of $nA$ edges crossing to $\bar{S}$. Taking the integral over all the points in distance $0 \leq x \leq r$ and assuming that there are $n\Delta x$ nodes in the area $1 \cdot \Delta x$ we get that the expected number of edges crossing from $S$ to $\bar{S}$ is (ignoring the effect of the borders)\(^5\)

$$E[\text{Cut}(S, \bar{S})] \leq \int_0^r \frac{1}{2}r^2 n \text{And}x$$

$$= \int_0^r \frac{1}{2}r^2 n [\arccos(\frac{x}{r}) - \sin(\arccos(\frac{x}{r}))] \text{nd}x$$

$$= \frac{1}{2}r^2 n^2 \left[ \frac{-2r\sqrt{1 - \frac{x^2}{r^2}}}{-2r + 2\frac{3}{r}x \arccos(\frac{x}{r})} \right]_0^r$$

$$= \frac{1}{2}r^2 n^2 (0 - (-2r + \frac{2}{3}r))$$

$$= \frac{2}{3}r^3 n^2$$

To approximate the conductance we use the expected node degree as $\pi r^2 n$ and consider the border effect as we take the integral over the area $(1 - r) \cdot \Delta x$

\(^5\)Note that as $n \to \infty$ and $r \to 0$ the above bound tightens and approaches equality.
(assuming $r << 1$

\[
\Phi(G(n, r)) = \frac{Q(S, \bar{S})}{\Pi(S)} = 2 \sum_{x \in S, y \in \bar{S}} \Pi(x)P(x, y)
\]

\[
= 2 \sum_{x \in S, y \in \bar{S}} \frac{1}{n} \frac{1}{\pi r^2 n}
\]

\[
\approx 2 \frac{2}{3} \frac{n^2 r^2 (1 - r)}{\pi r^2 n^2} \frac{1}{\pi r^2 n^2}
\]

\[
= \frac{4}{3\pi} r(1 - r)
\]

(2.6)

2.5 The Cover Time of Random Geometric Graphs

It follows from the previous section, Corollary 2.18 and Theorem 2.8 that we can bound the cover time of $G(n, r)$ as follows:

**Corollary 2.19.** For $c > 1$, if $r^2 \geq \frac{c4\log n}{n}$, then w.h.p.

\[
C_{G(n, r)} = O(r^{-2}n \log n)
\]

If this bound on cover time were tight, then the only way to achieve optimal cover time for random geometric graphs would be by choosing a radius $r$ that is constant irrespective of the network size $n$. Recalling that our definition of $G(n, r)$ is normalized to a unit area, this would mean that only broadcast networks of constant hop diameter may have optimal cover and partial cover. Even the minimum radius required for rapid mixing, which is $r_{\text{rapid}} = \Theta(1/\text{poly}(\log n))$, is several orders lower than such a radius. However, fortunately, the bounds given by Theorem 2.8 and correspondingly Corollary 2.19 are not tight. The method that we used to improve upon these results and to derive Theorem 2.1 is by bounding the electrical resistance of $G(n, r)$, which bounds the cover time.
by (2.2). In turn, we bound the resistance $R$ of $G(n, r)$ by bounding the power of a unit flow as permitted by Thomson’s Principle. For any pair of points $u$ and $v$, we construct a flow $c$ in such a manner that the power of the flow satisfies $P(c) = O\left(\frac{n}{m}\right) = O\left(\frac{1}{\delta_{\text{avg}}}\right)$ where $\delta_{\text{avg}}$ denotes the average degree of a node in $G(n, r)$. Since $R \leq P(c)$ the above flow together with (2.2) establish to be sufficient for $G(n, r)$ to have optimal cover time.

To construct a flow from $u$ to $v$, we partition the nodes into contour layers based on distance from $u$ and expanding outward until the midpoint between $u$ and $v$, then from the midpoint line onward contracting towards $v$ in a mirror fashion. The idea of using contour layers that expand with distance from a point is similar to the layering ideas used by Chandra et al. [CRR89] for meshes and originally by Doyle and Snell [DS84] for infinite grids. Layers in our case can be visualized as slices of an isosceles right triangle along the hypotenuse that connects $u$ and $v$. The flow can thus be thought of as moving through consecutive layers, with the total flow on the edges connecting consecutive layers being 1. Just as the variance of a probability function is minimized for the uniform distribution, we minimize the power by allocating flow almost uniformly along the set of edges used between layer $l$ and layer $l + 1$.

The construction of the above flow is based, as before, on the geo-dense property of random geometric graphs. To construct the flow we define bins as equal size areas that partition the unit square. These bins are used to construct our layered flow: Nodes in neighboring bins are in the same clique, and only edges between neighboring bins contribute to the flow.
2.5.1 The Cover Time and Resistance of Geometric Graphs

Before proving Theorem 2.1 about random geometric graphs we are going to prove a more general theorem about geometric graphs.

Theorem 2.20. A geometric graph $G(n, r)$ that is 8-geo-dense and has $r = \Theta\left(\frac{\log n}{n}\right)$ has optimal cover time of $\Theta(n \log n)$, optimal partial cover time of $\Theta(n)$, and optimal blanket time of $\Theta(n \log n)$.

Let $G(n, r)$ be a geometric graph that is 8-geo-dense. We will prove Theorem 2.20 using Eq. (2.2) that bounds the cover time as $C_G \leq mR \log n$ and by bounding the resistance between any two points $u, v$ in $G(n, r)$. If we’ll show that the resistance $R$ of $G(n, r)$ is $O\left(\frac{n}{m}\right)$ then it follows that $C_G \leq O(n \log n)$ and we done. Since we showed that $m = |E| = \Theta(n^2 r^2)$, all we need is to prove that $R = O\left(\frac{n}{m}\right) = O\left(\frac{1}{nr^2}\right)$.

Theorem 2.21. The resistance $R_{uv}$ between $u, v \in V$ is $\Theta\left(\frac{1}{nr^2} + \frac{\log(d(u,v)/r)}{n^2 r^4}\right)$.

**Proof.** The proof of the upper bound will be by bounding the power of a unit flow $c$ that we construct between $u$ and $v$.

Let $T(u, v)$ be an isosceles right triangle such that the line $(u, v)$ is the hypotenuse. It is clear that such a triangle which lies inside the unit square must exist. We divide our flow $c$ into two disjoint flows $c_1$ and $c_2$ where $c_1$ carries a unit flow from $u$ up to the line perpendicular to the midpoint of $d(u, v)$ in increasing layer size, and $c_2$ forwards the flow in decreasing layer size up to $v$ which is the only sink. By symmetry we can talk only about $c_1$ since the construction of $c_2$ mirrors that of $c_1$ and $P(c) = P(c_1 + c_2) = 2P(c_1)$ since the flows are disjoint. To construct the flow in $c_1$ we divide the line $(u, \text{midpoint}(u, v))$ into $d(u, v)\sqrt{2}/r$ segments of size $r/\sqrt{8}$, and number them from 0 to $d(u, v)\sqrt{2}/r - 1$ (see Fig 2.4).
Figure 2.4: $T(u, v)$ and the flow $c$ between $u$ and $v$ in $G(n, r)$ (A).

Let $S_l$ be the largest rectangle of width $r/\sqrt{8}$ included in the intersection of the area perpendicular to the $l^{th}$ segment and $T(u, v)$. $S_l$ will define the $l^{th}$ layer in our flow. Note that the area of $S_l$ is $lr^2/8$ and contains $l$ squares of area $r^2/8$, each of them containing $\Theta(nr^2)$ nodes by the geo-dense property.

Let $V_l \subseteq V$ be the set of nodes in layer $l$. $V_0 = u$, and for $l > 0$ a node $v$ is in layer $l$ if and only if it is located inside $S_l$. It follows that $|V_l| = \Theta(nr^2 l)$. Edges in our flow are only among edges $e = (x, y)$ s.t. $x \in V_l$ and $y \in V_{l+1}$, and all other edges have zero flow. In particular, the set of edges $E_l$ that carries flow from layer $l$ to layer $l + 1$ in $c_1$ is defined as follows: for the case $l = 0$, $E_0$ contains all the edges from $u$ to nodes in $V_1$, noting that $|E_0| = |V_1| = \Theta(nr^2)$ since $u \cup V_1$ is a clique (i.e. the maximum $d(u, x), x \in V_1$ is $r$). This allows us to make the flow uniform such that each node in $V_1$ has incoming flow of $1/|V_1|$ and for each edge $e \in E_0$, $c_1(e) = 1/|E_0|$. For $l > 0$ (see again Fig. 2.4 (A)) we divide $S_l$ into $l$ equal squares $A_1, A_2, \ldots, A_l$ each of size $r^2/8$. Let $V_{A_i}$ be the set of

---

6Assume for simplicity the expression divides nicely, if not, the proof holds by adding one more segment that will end at the midpoint and overlap with the previous segment.
nodes contained in the area $A_i$. We then divide $S_{i+1}$ into $l$ equal sized rectangles $B_1, B_2 \ldots B_l$ and define $V_{B_i}$ similarly, with $B_i$ touching $A_i$ for each $i$.

Now let $E_l = \{(x, y) \mid x \in V_{A_i} \text{ and } y \in V_{B_i}\}$. Note again that since, for each $i$, the maximum $d(x, y)$ between nodes in $A_i$ and nodes in $B_i$ is $r$ (see Fig. 2.4 (B)), $V_{A_i} \cup V_{B_i}$ is a clique (as the worst case distance occurs between the first two layers). So, the number of edges crossing from $A_i$ to $B_i$ is $|V_{A_i}||V_{B_i}| = \Theta(n^2 r^4)$ by geo-dense property. The clique construction allows us to easily maintain the uniformity of the flow such that into each node in $V_{B_i}$ the total flow is $1/l|V_{B_i}|$, and each edge carries a flow of $\Theta(1/n^2 r^4 l) = \Theta(1/E_l)$. All other edges have no flow. Now we compute the power of $c$:

$$R_{uv} \leq \sum_{e \in c} c(e)^2 = \sum_{e \in c_1} c_1(e)^2 + \sum_{e \in c_2} c_2(e)^2 =$$

$$= 2 \sum_{l=0}^{\lceil \sqrt{2d(u,v)/r} \rceil} \sum_{e \in E_l} c_1(e)^2 + 2 \sum_{l=1}^{\lceil \sqrt{2d(u,v)/r} \rceil} \frac{1}{l} \sum_{e \in E_l} 1$$

$$= 2O\left(\frac{1}{nr^2}\right) + 2O\left(\frac{1}{n^2r^4}\right) \sqrt{2d(u,v)/r} \sum_{l=1}^{\lceil \sqrt{2d(u,v)/r} \rceil} \frac{1}{l}$$

$$= O\left(\frac{1}{nr^2} + \frac{\log(d(u,v)/r)}{n^2r^4}\right)$$

To prove the lower bound we again follow in the spirit of [DS84] and use the "Short/Cut" Principle. We partition the graph into $\lceil d(u, v)/r \rceil + 1$ partitions by drawing $\lceil d(u, v)/r \rceil$ squares perpendicular to the line $(u, v)$, where the first partition $P_0$ is only $u$ itself and the $l^{th}$ partition $P_l$ is the area of the $l^{th}$ square excluding the $(l-1)^{th}$ square area. The last partition contains all the nodes outside the last square including $v$ (see Fig 2.5 (A)). We are shorting all vertices in the same partition (see Fig. 2.5 (B), and following the reasoning of the upper bound, let $m_l$ be the number of edges between partition $l$ and $l+1$. $m_0$ is $\Theta(nr^2)$
Figure 2.5: Lower bound for $R_{av}$ on the $G(n, r)$

and for $l > 0$, $m_l = \Theta(n^2r^4l)$, so

$$R_{av} \geq \sum_{l=0}^{\lfloor d(u,v)/r \rfloor} \frac{1}{m_l}$$

$$= \Omega\left(\frac{1}{nr^2}\right) + \sum_{l=1}^{\lfloor d(u,v)/r \rfloor} \Omega\left(\frac{1}{n^2r^4l}\right)$$

$$= \Omega\left(\frac{1}{nr^2} + \frac{\log(d(u,v)/r)}{n^2r^4}\right)$$

Corollary 2.22. The resistance $R$ of $G(n, r)$ is $\Theta\left(\frac{1}{nr^2} + \frac{\log(\sqrt{2}/r)}{n^2r^4}\right)$.

This follows directly from the fact that $\max_{u,v} d(u,v) \leq \sqrt{2}$. Now we can prove Theorem 2.20.

**Proof of Theorem 2.20.** Remember that $m = \Theta(n^2r^2)$, so all we need is $R = O(n/m) = O(1/nr^2)$ and then the cover time bound will follow by (2.2), the partial cover time bound will follow from (2.3), and the blanket time will follow from [WZ96] and the log $n$ order difference between the cover time and maximum.
hitting time. In order to have \( R = \Theta(\frac{1}{n^{r^2}}) \) we want that \( \frac{\log(\sqrt{2}/r)}{n^{r^2}} = O(\frac{1}{n^{r^2}}) \), which means \( \frac{\log(1/r)}{n^{r^2}} \leq \alpha \) for some constant \( \alpha \). Taking \( r^2 = \frac{c \log n}{n} \), for a constant \( \beta \), we get \( \frac{\log(n/\beta \log n)}{\beta 2 \log n} = \frac{1}{2\beta} - \frac{\log(\beta \log n)}{2 \beta \log n} \leq \frac{1}{2\beta} \). \qed

\section{Cover Time and Resistance of \( G(n, r) \)}

After Proving Theorem \ref{thm:degree}, in order to prove Theorem \ref{thm:cover-time} all we need to show is that for \( c > 1 \), \( r^2 = \frac{c \log n}{n} \) is sufficient to guarantee with high probability that \( G(n, r) \) is \( 8 \)-geo-dense. Note however that the second part of the theorem follows directly from [GK98] since if \( G(n, r) \) is disconnected with positive probability bounded away from zero when \( r^2 \leq \frac{\log n}{\pi n} \), then it has infinite cover time with at least the same probability.

Now combining the results of Lemmas \ref{lem:degree} and \ref{lem:cover-time} we can prove Theorem \ref{thm:cover-time}.

\begin{theorem}
For \( c > 1 \), if \( r^2 \geq \frac{c \log n}{n} \), then w.h.p. \( G(n, r) \) has cover time \( \Theta(n \log n) \). If \( r^2 \leq \frac{\log n}{\pi n} \), then \( G(n, r) \) has infinite cover time with positive probability (bounded away from zero).
\end{theorem}

\begin{proof}
Clearly from Lemma \ref{lem:cover-time} for \( c > 1 \), \( r^2 = \frac{c \log n}{n} \) satisfies the \( 8 \)-geo-dense property w.h.p., and since \( r^2 \) is also \( \Theta(\frac{\log n}{n}) \) the result follows from Theorem \ref{thm:degree}.
\end{proof}

Similarly we can state the following:

\begin{corollary}
For \( c > 1 \), if \( r^2 \geq \frac{c \log n}{n} \), then w.h.p. \( G(n, r) \) has optimal partial cover time \( \Theta(n) \) and optimal blanket time \( \Theta(n \log n) \).
\end{corollary}
2.5.3 The Threshold Width of Optimal Cover Time

Recently Goel, Rai, and Krishnamachari [GRK04b] have shown that any monotone property of random geometric graphs has a sharp threshold and bounded the corresponding threshold width. A monoton property is a graph property that is preserved under edge addition, for example connectivity. A threshold width associated with a monotone property $Q$ in $G(n,r)$ is define as the difference $\delta(n,\epsilon) = r_{Q:\text{high}} - r_{Q:\text{low}}$ where $r_{Q:\text{high}}$ is the smallest radius for which property $Q$ holds with probability at least $1 - \epsilon$, and $r_{Q:\text{low}}$ is the largest radius for which property $Q$ holds with probability at most $\epsilon$, for any $0 \leq \epsilon \leq \frac{1}{2}$. When $\delta(n,\epsilon) = o(1)$ $Q$ said to have a \textit{sharp threshold}. Goel \textit{et al.} proved that any monotone property in the two dimensional random geometric graph\footnote{they proved bounds for all dimensions, but in the current work we only concern with the result for $G(n,r)$ in 2 dimensions, $[0, 1]^2$.} has a threshold width of $O(r_{\text{con}} \log^\frac{1}{4} n)$. While for general graphs optimality of cover time is not a monotonic property (see next subsection for a proof), it follows clearly from our bound that optimality of cover time is monotone for random geometric graphs and has a threshold width of $O(r_{\text{con}})$. Interestingly this bound is an order lower than the bounds on the threshold width obtained by Goel \textit{et al.}, but actually supports their \textit{conjectured} threshold width for monotone properties.

2.5.4 Optimal Cover Time is not Monotone

An immediate and well-known corollary to Rayleigh’s Short/Cut Principle is that the Resistance $R$ of a graph is monotone, as adding new edges can not increase the resistance $R$. On the other hand, it is also well-known that, in general, cover time is not a monotone property of graphs. As a simple demonstrative example we can take the line of $n$ nodes which has cover time of $O(n^2)$, and by adding
edges we can create the lollipop graph which is known to have cover time of \( O(n^3) \), and if we keep adding edges we will get the complete graph which has optimal cover time, \( O(n \log n) \) [CRR89]. One can wonder if this is still the case if the graph \( G \) already has cover time of \( O(n \log n) \). In other words, can we create, by adding more edges, a graph \( G' \) which has cover time of \( \omega(n \log n) \)?

**Lemma 2.24.** Cover time of \( O(n \log n) \) is not a monotone property of graphs.

**Proof.** The proof will be by counter example and by the lower bound for cover time given in (2.2). Let \( G \) be the 3-dimentional grid of \( n \) nodes. It is known that \( G \) has cover time of \( C_G = O(n \log n) \) [CRR89]. We construct a graph \( G' \) be adding \( O(n^2) \) edges to \( G \) is such a way that the resistance of the graph will not change: Let \( u_0 \) be the node at \((0,0,0)\) and \( u_n \) the node at \((\sqrt[3]{n}, \sqrt[3]{n}, \sqrt[3]{n})\). Make all the points with \( L_1 \) distance of at most \( \sqrt{n} \) from \( u_0 \) a clique. The number of nodes in this clique is \( \approx n/2 \), and so the number of edges in this clique is \( \approx n^2/8 \), making the total number of edges in \( G' \) \( m = \Theta(n^2) \). Since the minimum degree in \( G' \) is the same as in \( G \), namely degree of 3 at \( u_n \), the resistance \( R \) of \( G' \) \( \geq 1/3 \), and by Theorem 2.2 \( mR \leq C_G \) so we get \( C_{G'} = \Omega(n^2) \). \( \square \)

### 2.5.5 Cover Time and Resistance of Deterministic Geometric Graphs

As an example of other applications of our results consider the following: for an integer \( k \), let the \textbf{\( k \)-fuzz} [DS84] of a graph \( G \) be the graph \( G_k \) obtained from \( G \) by adding an edge \( xy \) if \( x \) is at most \( k \) hops away from \( y \) in \( G \). In particular, let \( G_1(n) \) denote the 2-dimensional grid of \( n \) nodes, and let \( G_k(n) \) be the \( k \)-fuzz of \( G_1(n) \). It is known that the cover time of \( G_1(n) \) is \( \Theta(n \log^2 n) \); and so we ask what is the minimum \( k \) s.t. \( G_k(n) \) has an optimal cover time of \( \Theta(n \log n) \)?

**Definition 2.3.** Let \( T = T(n, r(n)) \), where \( n \) is s.t. \( \sqrt{n} \in \mathbb{Z} \) denote the class of
geometric graphs, where the nodes of each instance of $T(n, r)$ are placed on the unit square exactly as the 2-dimensional grid of $n$ nodes. In other words, there is exactly one node at each position $(\frac{i}{\sqrt{n}}, \frac{j}{\sqrt{n}})$ $0 \leq i, j \leq \sqrt{n}, i, j \in \mathbb{Z}$.

Note the following:

**Corollary 2.25.** For $G_k(n)$, a $k$–fuzz of the 2-dimensional grid

1. $G_1(n) = T(n, \frac{1}{\sqrt{n}})$ (i.e the 2-dimensional grid of $n$ nodes).

2. $G_k(n)$ is a super-graph of $T(n, \frac{k}{\sqrt{2n}})$

3. $G_k(n)$ is a sub-graph of $T(n, \frac{k}{\sqrt{n}})$

**Claim 2.26.** For a constant $k$ the resistance of $T(n, \frac{k}{\sqrt{n}})$ is $\Theta(k^{-4} \log n)$.

**Proof.** It is clear that $T(n, \frac{k}{\sqrt{n}})$ satisfies the 8-geo-dense property,\(^8\) so the result follows directly from Theorem 2.21:

\[
R = \Theta\left(\frac{1}{nr^2} + \frac{\log(\sqrt{2}/r)}{n^2 r^4}\right)
= \Theta(k^{-2} + k^{-4} \log(\sqrt{2n}/k))
= \Theta(k^{-4} \log n)
\]

\[\square\]

**Theorem 2.27.** For any constant $k$, the cover time of $G_k$ is $\Theta(k^{-2} n \log^2 n)$.

**Proof.** The upper bound follows directly from Corollary 2.25, Claim 2.26, and equation (2.2). To prove the lower bound note that the resistance $R_{uv}$ is $\Theta(k^{-4} \log(\frac{d(u,v)\sqrt{n}}{k}))$. Letting $d'(u,v) = d(u,v)\sqrt{n}$ denote the non-normalized distance (hop distance), we have that $R_{uv} = \Omega(k^{-4} \log(d'(u,v)))$ where $1 \leq \ldots$

---

\(^8\)for large enough $k$
\( d'(u, v) \leq \sqrt{2n} \). Now we can use the method of Zuckerman [Zuc90] (specifically in Lemma 2 and Theorem 4). And, by noting that it is known that the commute time \( C_{uv} = 2mR_{uv} \), we have that hitting time \( (E_uT_v \text{ in Zuc90 notation}) \) is \( \Omega(k^{-2}n \log(d'(u, v))) \). Then the result follows directly from the proof of Theorem 4 in [Zuc90].

Thus, we have the solution to our question:

**Corollary 2.28.** \( G_K(n) \) has Cover Time of \( \Theta(n \log n) \) if \( k = \gamma_n \) and \( \lim_{n \to \infty} \frac{\log n}{\gamma_n^2} \leq c \) for some constant \( c \).

### 2.6 Notes and Related Work

There is a vast body of literature on cover times, mixing time and on geometric graphs, and to attempt to summarize all of the relevant work would not do it justice. We have already mentioned some of the related results previously, however, here we would like to highlight the related literature that has been most influential to our result, namely that of Chandra et al. [CRR89] and Doyle and Snell [DS84].

The work of Doyle and Snell [DS84] is a seminal work regarding the connection between random walks and electrical resistance. In particular, they proved that while the infinite 2-dimensional grid has infinite resistance, for any \( d \geq 3 \) the resistance of the \( d \)-dimensional grid is bounded from above, and these results were established to be sufficient in re-proving Pólya’s beautiful result that a random walk on the infinite 2-dimensional grid is recurrent whereas a random walk on the infinite \( d \)-dimensional grid for any \( d \geq 3 \) is transient. In obtaining this result, essentially thors bounded the power of a unit current flow from the origin out to infinity and found that the power diverges for the 2-dimensional case.
and converges for every dimension greater than two. The authors used a layering argument, namely partitioning nodes into disjoint contour layers based on their distance from the origin, and the rate of growth of consecutive layers can be seen as the crucial factor yielding the difference between the properties of the different dimensions. Later, Chandra et al. [CRR89] proved the tight relation between commute time and resistance, and used that relationship to extend Doyle and Snell’s result by bounding the cover time of the finite $d$-dimensional mesh by computing the power and resistance via an expanding contour layers argument. Together with the tight lower bound of Zuckerman [Zuc90], they showed that the 2-dimensional torus has cover time of $\Theta(n \log^2 n)$, and for $d \geq 3$ the $d$-dimensional torus has an optimal cover time of $\Theta(n \log n)$.

While this chapter deals with random geometric graphs there are striking similarities between $G(n, r)$ and the Bernoulli graphs $B(n, p)$ which we discuss in more details in Chapter 4. For example, the critical parameters for connectivity and sharp thresholds for monotone properties [GRK04b, Bol85]. Regarding cover time, Jonasson [Jon98] and Cooper and Frieze [CF03] gave tight bounds on the cover time and an interesting aspect of our result is that we add another similarity and both classes of graphs have optimal cover time around the same threshold for connectivity. Yet, despite the similarities between $G(n, r)$ and $B(n, p)$, Bernoulli graphs are not appropriate models for connectivity in wireless networks since edges are introduced independent of the distance between nodes. In wireless networks the event of edges existing between $i$ and $j$ and between $j$ and $k$ is not independent of the event of an edge existing between $k$ and $i$. There are other notable differences between $G(n, r)$ and $B(n, p)$ as well. For example, the proof techniques for the above results for $G(n, r)$ are very different than the proof techniques for the respective results for $B(n, p)$. Interestingly, whereas the proof of [CF03] for optimality of cover time in Bernoulli graphs of $\Theta(\log n)$ average degree
depends on the property that Bernoulli graphs do not have small cliques (and, in particular that small cycles are sufficiently far apart), in the case of random geometric graphs the existence of many small cliques uniformly distributed over the unit square like bins, in other words geo-denseness, is essential in our analysis.

Geo-denseness is also essential in our method of bounding the conductance of $G(n,r)$ to bound the mixing time. Previous work on the use of conductance to bound the mixing time of graphs has been primarily geared towards approximations for hard counting problems and has utilized large, sophisticated constructions of Markov Chains [SJ89].

Another recent result with a bin-based analysis technique for random geometric graphs is that of Muthukrishnan and Pandurangan [MP05]. However, as their technique uses large overlapping bins where the overlap is explicitly stated to be essential and there is no direct utilization of cliques.

As mention before there had been other work that proposed random walk techniques for different networks settings. Servetto and Barrenechea [SB02] offer to use a constrained random walk to perform routing in a $n \times n$ grid network where each node knows his location in the grid. The network is dynamic since nodes are switching between active and in-active states at random times. Using the location knowledge they show that constrained (constrained since the message is only forward in the destination direction) random walk can achieve reliable and load balanced routing from $(0,0)$ to $(n,n)$. Like any random walk, their protocol is memory-less and does not require any path discovery or repair computation due to the failures.

Braginsky and Estrin [BE02] take another random walk oriented approach and offer a rumor routing scheme for query answering. The idea is based on the fact that two lines in a bounded rectangle are to intersect with high probability.
In their scheme every event in the network issues a few random walks that leave a trace of the event in the network. When the sink sends a query, yet another random walk, it intersects with one of the event traces with high probability. Then the query can route to the event source and get the detailed answer which will save the cost of flooding the query to the network. This scheme is suitable when nodes generate events, but it will not apply if we, for example, are trying to find the average value of the network. It is also not clear, how robust it is in a dynamic network, since it would have to maintain event and routing tables.

The most similar solution to ours is ACQUIRE proposed by Sadagopan et al. [SKH03] This system uses random walks to answer one-shot, non-aggregate, complex, replicate data queries. The system works by issuing a query as a random walk with parameter $d$, at each node the walk retrieves data from all nodes that are less than $d$-hops away neighbors and continue by moving randomly to a node at distance $d$. The walk continues until it finds the node with the answer and then sent it back on the same way it arrives. The authors compare ACQUIRE to a flooding based system and to an expanding ring search. Assuming a uniform distribution of data values over the nodes in the network and given that the walk doesn’t return to the same node twice, the authors analytically and experimentally show that ACQUIRE can be more efficient than the other two methods.

Gkantsidis et al. [GMS04] have proposed random walk techniques for peer-to-peer networks based on that random walks are especially efficient when the underlying topology of the network is an expander [MR95]. As such, they have advocated expander construction in peer-to-peer network design and the use of random walks to perform search.

Kempe et al. [KDG03] proposed and analyzed parallel random walk tech-
niques for gossiping in networks. They were concerned with disseminating global information throughout the network to be stored in every node (all-to-all), for example the task of computing the average temperature in the network and storing it at each node in the network. Motivated by this work, and independently to our earlier version [AE05a], Boyd et al. [BGP] and Rai [Rai04] proposed the idea of using conductance to bound the mixing time of broadcast networks and obtained a similar result to Theorem 2.2.

2.7 Conclusions

In this Chapter we have shown that for a two dimensional random geometric graph $G(n, r)$, if the radius $r_{\text{opt}}$ is chosen just on the order of guaranteeing asymptotic connectivity (i.e the same evolution stage of connectivity) then $G(n, r)$ has optimal cover time of $\Theta(n \log n)$ for any $r \geq r_{\text{opt}}$. Noting that $G(n, r_{\text{opt}})$ still has a long diameter of $\Theta(\frac{1}{r_{\text{opt}}}) = \Theta(\sqrt{\frac{n}{\log n}})$, it is not surprising that it is not rapid mixing, a property which we have shown requires a radius of at least $r_{\text{rapid}} = \Theta(1/\text{poly}(\log n))$. Intuitively, this gap seems to indicate that although the partial cover is optimal, that is linear, the distribution of the uncovered nodes after the partial cover may be such that contiguous uncovered geometric regions may remain (see simulation results in chapter 5).

Our proof techniques can be generalized to the $d$-dimensional random geometric graph $G^d(n, r)$, yielding that for any given dimension $d$, $r_{\text{opt}} = \Theta(r_{\text{con}})$ with correspondingly optimal cover time. However, both grow exponentially with $d$ which seems to be a consequence of a separation between average degree and minimum degree for higher dimensions rather than just an artifact of our method. Nevertheless, the case of dimension $d = 2$ is considered to be the hardest one [AF99]. This can intuitively be seen from the mesh results. The case for $d = 1$
(i.e the cycle) is easy to analyze. For $d > 2$ the cover time of the $d$-dimensional mesh is optimal \cite{CRR89}, and we can show that for any $k$ the cover time of the $k$-fuzz is also optimal. On the other hand, as we show in Section 2.5.5, the cover time of the $k$–fuzz in 2 dimensions (i.e. $G_k(n)$) for constant $k$ is not optimal making this the most interesting case.
CHAPTER 3

Efficient Restricted Delaunay Triangulation in Random Geometric Graphs

3.1 Introduction

Sensor Networks introduce new design challenges; the strict energy and memory constraints of the sensors and the large scale of the network require the use of distributed, localized algorithms which minimize memory and energy use \cite{EGH99}. Since energy is mostly consumed by radio communication, the number of messages being sent by a given algorithm is considered the efficiency metric. Naturally, these restrictions and the theoretical model of random geometric graphs have led to a variety of analytical work aimed at investigating different properties of such networks \cite{GK98, GRK04b, MP05, AE05b}.

The tasks of topological control and routing in sensor networks have been studied extensively and, in particular, lead to the advent of \textit{geo-routing} (geographical routing) \cite{BMS99, KK00}. In geo-routing the assumption is that each node knows its own location (i.e. its coordinates) and the location of the destination to which it wants to deliver a message (via Global Positioning System (GPS) at each node or some other mechanism). The goal then is to find an efficient route from source to destination using only the local information\footnote{The location of your immediate neighbors.} available
at each node and a limited amount of memory.

Most of the early work on this issue, beginning with the proposals of Bose et al [BMS99] and Karp and Kung [KK00], has been based on greedy forwarding combined with face routing over a planar sub-graph of the network. That is, the message is always forwarded to the neighbor closest to the destination, and if such a neighbor does not exist, recovery from the local minima is obtained using a route along the current face of the planar sub-graph. Although this method guarantees delivery, the efficiency of the method depends on the properties of the planar sub-graph. Ideally, our sub-graph should be sparse and locally constructed, but at the same time a spanner: the shortest path between any pair of points is at most a constant factor longer than the shortest path in the original graph. The sparseness and locality reduce energy and memory consumption in the construction phase while the spanner property allows efficient routing.

Several candidate graphs for goe-routing in wireless networks have been offered recently in the literature. Let $G$ be a geometric graph (i.e the nodes are embedded in the plane). The Relative Neighborhood graph, $RNG(G)$ [Tou80], and the Gabriel Graph, $GG(G)$ [GS69], are both planar and can be efficiently constructed locally, but they are not good spanners, even in random geometric graphs [BDE02]. Another well-known planar graph is the Delaunay triangulation whose known to be a spanner of the complete graph [Che86, DFS90]. Unfortunately the Delaunay triangulation of a geometric graph $G$, $Del(G)$, cannot be constructed locally and may contain long edges; In other words $Del(G)$ is not necessarily a sub-graph of $G$. To overcome this problem several authors proposed the Restricted Delaunay Graph, $RDG(G)$. This is a planar sub-graph of $G$ that contains all the edges of $Del(G)$ that are also in $G$, and was proved to be a spanner of $G$ [GGH01, LCW02]. Note, that by definition $RDG(G)$ is
not unique and different methods have been suggested to construct such graphs [GGH01, LCW02, WL03, AR04].

In the context of random geometric graphs, as before, we are interested in the relation between the range of communication $r$, the number of nodes $n$ in the graph and some desired property $Q$ (e.g. connectivity). In ad-hoc and sensor networks interference grows with increased communication radius. So one wants to find a tight upper bound on the smallest radius $r_Q$, that will guarantee that $Q$ holds $w.h.p.$\footnote{Recall that event $E_n$ occurs with high probability if probability $P(E_n)$ is such that $\lim_{n \to \infty} P(E_n) = 1.$}. As before we use the critical radius for connectivity, $r_{\text{con}}$ as a reference point and recall that if $\pi r^2 \geq \pi r_{\text{con}}^2 = \frac{\log n + \gamma_n}{n}$ then $\mathcal{G}(n, r)$ is connected $w.h.p.$ iff $\gamma_n \to +\infty$ [Pen97, GK98].

It is well known that the maximum edge length of the Delaunay triangulation of $\mathcal{G}(n, r)$ in the unit square, and in particular on the convex hull, is $w(r_{\text{con}})$. Recently, a similar result has been proved also for the unit disk [KLS04]. Therefore it is clear that when $r = O(r_{\text{con}})$ the Delaunay triangulation cannot be computed locally (i.e. with information obtained only from nodes that are a constant number of hops away).

In this chapter we show that if $r = O(r_{\text{con}})$, namely on the order that guarantees connectivity, then $w.h.p.$, we can efficiently and locally construct a Restricted Delaunay Graph $\text{RDG}(G)$. We show that, while for general graphs this construction requires $\Theta(n)$ messages, an order of $O(\sqrt{n \log n})$ messages suffices in $\mathcal{G}(n, r)$. We further present a novel algorithm that achieves this bound. Our algorithm exhibits two unique features which results in reduce message count. First, it requires only one round of communication and second, only ”problematic” nodes need to send messages. Our results are stated for geometric graphs that have some nice properties, but are not necessarily random nor in a specific bounded
area (i.e. square, disk). Later we show that random geometric graphs in the unit square (or unit disk) have these nice properties with high probability and the results follow.

3.2 Preliminaries

We consider a wireless ad hoc network (or sensor network) over a set $V$ of $n$ nodes distributed in the unit square, where each node can communicate with all the nodes in its transmission range, that is a disk of radius $r$ centered at the node. The resulting is a geometric graph $G = G(V, r)$ with $V$ the set of nodes and $E = \{\{u, v\} \mid u, v \in V \wedge d(u, v) \leq r\}$ the set of edges. This is similar to the Unit Disk Graph $UDG(V)$ [CC91] in which the set of nodes $V \in \mathbb{R}^2$ and the radius is assumed to be one unit, but in our case we are interested in a network in a bounded area and in the relation between the number of nodes $n$ and the transmission range $r$ as a function of $n$.

Let $N(u)$ denote the neighbors of $u$ including $u$ and $N(u, v)$ the set of the common neighbors of $u$ and $v$, i.e. $N(u, v) = N(v, u) = N(u) \cap N(v)$. Throughout the paper we use three disk definitions: let $\text{disk}_r(u)$ be the disk centered around $u$ with radius $r$ (with $r$ omitted when the context is clear), $\text{disk}(u, v)$ be the disk through $u, v$ with diameter $d(u, v)$ and $\text{disk}(u, v, w)$ be the unique circumcircle over $u, v$ and $w$.

Next we present more graphs over the set of nodes $V$, note that in some cases the graphs are derived directly from $V$, others are a function of $G$ (i.e. $r$ is needed to compute them). The Voronoi diagram of a set of nodes (or sites) $V$ in the space, $\text{Vor}(V)$, is the partition of the space into cells $V_u, u \in V$ such that all the points inside $V_u$ are closer to $u$ than to any other node in $V$. The Delaunay
triangulation, \( \text{Del}(V) \), is the dual graph of \( \text{Vor}(V) \): an edge \( \{u, v\} \) is in \( \text{Del}(V) \) if and only if \( V_u \) and \( V_v \) share a common boundary. It is well known that \( \text{Del}(V) \) is a spanner of the complete graph \( K_n \) [Che86, DFS90], which means that the shortest path between any two points on \( \text{Del}(V) \) is at most \( t \) times the shortest path on \( K_n \), where \( t \) is a positive constant called the stretch factor. In the case of \( \text{Del}(V) \) and the complete graph \( t \approx 5.08 \). A useful property of the Delaunay triangulation is that a triangle \( \triangle uvw \in \text{Del}(V) \) if and only if \( \text{disk}(u, v, w) \) is empty: there is no other node from \( V \) in it\(^3\) [BKO97].

Let \( \text{UDel}(G) \) be the sub-graph of \( \text{Del}(V) \) that contains only the short edges of \( \text{Del}(V) \), that is the edges that are shorter than \( r \); therefore \( \text{UDel}(G) = \text{Del}(V) \cap G \) and is also a sub-graph of \( G \) [LCW02, GGH01].

**Definition 3.1** ([GGH01]). A Restricted Delaunay Graph \( \text{RDG}(G) \) is a planar graph such that

\[
\text{UDel}(G) \subseteq \text{RDG}(G) \subseteq G
\]

Let \( T(u) \) be the set of edges in \( \text{Del}(N(u)) \) (i.e the Delaunay triangulation of the nodes in \( N(u) \)) and similarly \( T(u, v) = \text{Del}(N(u, v)) \). Note that there may be edges in \( T(u) \) and \( T(u, v) \) which are not presented in \( \text{Del}(G) \).

**Definition 3.2.** \( \text{LocalDel}(G) \) is the graph resulting from computing \( T(u) \) at each node: an edge \( u, v \) is in \( \text{LocalDel}(G) \) if and only if there exist \( u \) s.t \( \{u, v\} \in T(u) \).

Figure 3.1 illustrates the graphs discussed above for a set \( V \) of 50 random points in the unit square.

\(^3\)For simplicity we assume that no four points in \( V \) are co-circular.
Figure 3.1: Different Graphs over a set $V$ of 50 random nodes in the unit square with $r = 0.3$ (A) $G(V, r)$. (B) $Vor(V)$. (C) $Del(V)$. (D) The edges in $Del(V)$ that are longer than $r$ (E) $LocalDel(G)$ where consistent edges are in dots and inconsistent edges are in solid lines.

57
3.3 Computing $RDG(\mathcal{G}(n, r))$

We offer an efficient algorithm to construct $RDG(\mathcal{G}(n, r))$. There are several advantages to our algorithm: (i) it assumes the transmission range is on the same order as the necessary range required for connectivity. (ii) there is only one round of communication, and most importantly (iii) the number of messages it sends is adaptive. Our algorithm is based on $LocalDel(G)$ which is known not to be a planar graph. There has been previous proposal to solve this problem but all of them required a constant number of messages per node. In our algorithm only messages that are needed to eliminate problematic edges are sent, enabling us to reduce the number of messages from $O(n)$ to $O(\sqrt{n \log n})$. Before describing the algorithm we define the following:

**Definition 3.3.** An edge $\{u, v\}$ is **local-inconsistent** at $u$ if $\{u, v\} \notin T(u)$ and $\{u, v\} \in T(u, v)$ and a **local-consistent** otherwise.

**Claim 3.1.** For each local-inconsistent edge $\{u, v\} \in T(u, v)$ at $u$ there is a triangle $\triangle uwv \in T(u)$ which is the **proof** that $\{u, v\}$ is local-inconsistent, i.e. $(w, x)$ intersect with $\{u, v\}$.

This is clearly true since otherwise $\{u, v\}$ will be in $T(u)$. Now we can introduce our algorithm:

The main results of this work are the following theorems about the correctness and number of messages in Algorithm 1:

**Theorem 3.2.** For $r \geq \sqrt{\frac{32}{3}r_{\text{con}}}$ Algorithm 1 computes $RDG(\mathcal{G}(n, r))$ w.h.p.

**Theorem 3.3.** For $r \geq \sqrt{\frac{32}{3}r_{\text{con}}}$ the number of messages in Algorithm 1 is $O(\sqrt{n \log n})$ and the number of bits is $O(\sqrt{n}(\log n)^{3/2})$ w.h.p.
Algorithm 1. local RDG(G) construction at node $u$

1. Compute $T(u)$ and for each neighbor $v \in N(u)$ compute $T(u,v)$.

2. Keep all edges $\{u, v\} \in T(u)$.

3. If there are local-inconsistent edges, broadcast proofs for all of them.

4. Remove edge $\{u, v\}$ if received a proof of its inconsistency.

To prove this theorems we next establish a few helping lemmas.

3.4 Properties of $LocalDel(G)$

The $LocalDel(G)$ graph can be constructed locally without exchanging messages assuming that each node knows the locations of all its neighbors. Here we assume that this information is obtained by each node using some other mechanism that is shared with other applications. We therefore do not count the messages required in this process as part of our algorithm (otherwise $\Omega(n)$ messages are necessary for any nontrivial task) and only consider algorithm-specific messages. Next we prove some properties of $LocalDel(G)$ that are based on the following proposition:

**Proposition 3.4.** Let $V' \subseteq V$. $(u, v \in V' \land \{u, v\} \in Del(V)) \Rightarrow \{u, v\} \in Del(V')$.

This clearly follows from the fact that if two Voronoi cells $V_u$ and $V_v$ share a boundary in $Vor(V)$ they must share a boundary in $Vor(V')$, since removing nodes cannot decrease their boundary.

**Definition 3.4.** An edge $\{u, v\}$ in $LocalDel(G)$ is **consistent** if $\{u, v\} \in T(u)$.
and \( \{u, v\} \in T(v) \), and **inconsistent** otherwise.

**Lemma 3.5.** If \( \{u, v\} \in U\text{Del}(G) \), then \( \{u, v\} \) is a consistent edge in \( \text{LocalDel}(G) \).

**Proof.** This follows directly from Proposition 3.4. Since \( \{u, v\} \in U\text{Del}(V) \) we have \( \{u, v\} \in \text{Del}(V) \), \( u \in N(v) \) and \( v \in N(u) \). Setting \( V' = N(u) \) we get \( \{u, v\} \in T(u) = \text{Del}(V') \) and similarly for \( V' = N(v) \) we get \( \{v, u\} \in T(v) = \text{Del}(V') \). \( \square \)

It is clear from Lemma 3.5 that \( U\text{Del}(G) \subseteq \text{LocalDel}(G) \), however, it is still not \( \text{RDG}(G) \) since it may be not a planar. There are two type of edges in \( \text{LocalDel}(G) \): consistent and inconsistent, and both may cross other edges. First we take care of the inconsistent edges.

**Lemma 3.6.** An edge \( \{u, v\} \in \text{LocalDel}(G) \) is inconsistent if and only if \( \{u, v\} \) is local inconsistent at \( u \) or \( v \).

**Proof.**  \( \Rightarrow \): Assume edge \( \{u, v\} \in \text{LocalDel}(G) \) is inconsistent. Recall that an edge \( \{u, v\} \) is in \( \text{LocalDel}(G) \) iff it is in \( T(u) \) or \( T(v) \). But if it is also inconsistent it cannot be in both. w.l.o.g let \( \{u, v\} \in T(u) \) and \( \{v, u\} \notin T(v) \). By Proposition 3.4 \( \{u, v\} \in T(u, v) = T(v, u) \) so \( \{v, u\} \) must be local-inconsistent at \( v \). \( \Leftarrow \): By Proposition 3.4 if an edge is consistent it must be local consistent at \( u \) and \( v \). \( \square \)

Next we bound the number of **proofs** each node can have for its inconsistent edges.

**Lemma 3.7.** A node can have at most 6 proofs for all its local-inconsistent edges in \( \text{LocalDel}(G) \).

**Proof.** A triangle \( \triangle uwx \in T(u) \) with \( \angle wux \leq \pi/3 \) cannot be a proof for a local-inconsistent edge \( \{u, v\} \) since \( v \) must then be neighbor with \( w \) and \( x \) and \( \text{Del}(N(u)) \) and \( \text{Del}(N(u, v)) \) agree on \( \{u, v\} \) and \( \{w, x\} \). \( \square \)
3.4.1 Well-distributed Geometric Graphs

From now on we turn to a more specific type of geometric graphs. First let's define them formally:

**Definition 3.5.** A geometric graph $G$ is **well-distributed** if every convex area of size at least $\frac{3\pi}{32}r^2$ (in the unit square) has at least one node in it.

In these graphs the nodes are distributed ”nicely” across the unit square and in particular do not contains large ”holes”: empty convex regions with area larger than $\frac{3\pi}{32}r^2$.

**Lemma 3.8.** If $G$, is a well-distributed geometric graph, then consistent edges do not intersect in $\text{LocalDel}(G)$.

**Proof.** Assume $\{u, v\}$ and $\{w, x\}$ are two consistent edges that intersect in $\text{LocalDel}(G)$, by Proposition 3.4 we can remove all nodes but $u, v, w$ and $x$ from the graph and consider only these two edges that must still exist and intersect. From Lemma 4.1 in [GGH01] it must be the case that one of the four nodes is
a neighbor of all other three, w.l.o.g let it be $w$. If any of the other three nodes sees (i.e. neighbor of) all the four nodes it must be the case that either $u$ or $v$ sees all four of them, w.l.o.g let it be $u$. But in this case since $w$ and $u$ see all four nodes $T(w) = T(u)$ and either $\{w, x\}$ and $\{u, v\}$ do not intersect or at least one of them is inconsistent which leads to contradiction. So assume $w$ is the only node that sees all the other four. Note that since $w$ selected $\{w, x\}$ as an edge while having information on the four nodes, $\{u, v\}$ is a the non-Delaunay edge among the two. Observe that $x$ must be outside $\text{disk}(u) \cup \text{disk}(v)$ (otherwise $u$ or $v$ see the four nodes), so it must be the case that $d(w, x) \geq \frac{\sqrt{3}}{2}r$, since $d(u, v)$ is at most $r$ and the edges intersect by assumption, see Fig. 3.2. Note also that since $u$ and $v$ choose $\{u, v\}$ as an edge in $\text{LocalDel}(G)$ we must have that the circumcircle $\text{disk}(u, v, w) \cap (\text{disk}(u) \cup \text{disk}(v))$ is empty. In particular, this imply that the disk $D$ of diameter $\frac{\sqrt{3}}{2}r$ which is tangent to the midpoint between $u$ and $v$ is empty. (see the gray disk in Fig 3.3). Since $w, x, u, v$ are all in the unit
Figure 3.4: An example where inconsistent edge \( \{u, v\} \) exist next to the border of the unit square

Now since \( G \) is well-distributed and the area of half of \( D \) is \( \frac{3\pi}{32} r^2 \) there is at least one node in that half, contradicting the consistency of \( \{u, v\} \) in \( \text{LocalDel}(G) \). \( \square \)

This lemma stands behind the core of our algorithm; For well distributed graph \( G \) all one needs to do to compute \( \text{RDG}(G) \) is to remove all inconsistent edges. Note, however, that it is still the case that even for well-distributed \( G \), there may be inconsistent edges in \( \text{LocalDel}(G) \). As Fig. 3.4 illustrates an edge \( \{u, v\} \) can be inconsistent at \( v \) since the area \( \text{disk}(v) \cap \text{disk}(u, v, w) \) (the gray area in the figure) can become arbitrary small next to the boundaries of the unit square. (For a similar reason it can be shown that long Delaunay edges can exist in \( \text{Del}(G) \), and in particular on the convex hull of \( V \)). Before formally proving the correctness of Algorithm 1 we need to show that random geometric graphs are well-distributed. We will do so by utilize a coupon collector argument.

**Lemma 3.9.** If \( r \geq \sqrt{\frac{32}{3}} r_{\text{con}} \) then w.h.p. \( \mathcal{G}(n, r) \) is well-distributed.
Proof. Recall that \( r_{\text{con}}^2 = \frac{\log n + \gamma_n}{\pi n} \) and \( \gamma_n \to \infty \). Partition the unit square into convex bins of size \( \frac{3\pi}{32} r^2 \) so the number of bins \( B \) is

\[
B = \frac{32}{3\pi r^2} = \frac{32}{3\pi} \frac{3\pi n}{32(\log n + \gamma_n)} = \frac{n}{\log n + \gamma_n}
\]

Now it is a known result [MR95] that if one throws balls, uniformly at random into \( B \) bins, the expected number of balls needed to fill every bin with at least one ball is \( B \log B \). If we want the result \( w.h.p. \) then one need to throw \( B \log B + \gamma_n B \) balls. To conclude the proof we need to show that \( n \geq B \log B + \gamma_n B \).

\[
B \log B + \gamma_n B = \frac{n}{\log n + \gamma_n} \log \left( \frac{n}{\log n + \gamma_n} \right) + \gamma_n \frac{n}{\log n + \gamma_n}
\]

\[
= n \left( \frac{\log n}{\log n + \gamma_n} - \frac{\log(\log n + \gamma_n)}{\log n + \gamma_n} + \frac{\gamma_n}{\log n + \gamma_n} \right)
\]

\[
= n \left( 1 - \frac{\log(\log n + \gamma_n)}{\log n + \gamma_n} \right)
\]

\[
\leq n
\]

Now we can proceed and prove Algorithm 1 correctness.

Proof of Theorem 3.2. From the last lemma for \( r \geq \sqrt{\frac{32}{3}} r_{\text{con}} \) \( w.h.p. G(n, r) \) in is well-distributed. Steps 1 and 2 compute a sub-graph of \( G \). In step 4 the algorithm removes all local-inconsistent edges which by Lemma 3.6 is equal to removing all inconsistent edges. After step 4 that resulting graph contain only consistent edges, and so by Lemma 3.5 is a super graph of \( UDel(G) \). Lemma 3.8 guarantees that the graph is also a planar (since consistent edges do not intersect in well-distributed graphs) so we get a \( RDG(G) \). \( \square \)

3.4.2 Bounding the number of messages

Let \( I = \left[ \frac{\sqrt{3}}{4} r, 1 - \frac{\sqrt{3}}{4} r \right]^2 \) be the inner square centered in the unit square where each side of I is at distance \( \frac{\sqrt{3}}{4} r \) from the side of the unit square. For a well distributed
Lemma 3.10. If $u \in I$ and $\{u, v\} \in \text{LocalDel}(G)$ then $d(u, v) < \frac{\sqrt{3}}{4} r$. 

**Proof.** Let $\{u, v\} \in \text{LocalDel}(G)$ and $u \in I$. Assume $d(u, v) \geq \frac{\sqrt{3}}{4} r$, then the area of each half of $\text{disk}(u, v)$ is at least $\frac{3\pi}{32} r^2$ and at least this area of each half is inside the unit square. Since $G$ is well distributed each half contains at least one node, so $\{u, v\}$ can’t be an edge in $\text{LocalDel}(G)$. Contradiction. \(\square\)

Lemma 3.11. If $\{u, v\} \in \text{LocalDel}(G)$ and $u, v \in I$ then the edge $\{u, v\}$ is consistent.

**Proof.** Let $\{u, v\} \in \text{LocalDel}(G)$ and $u, v \in I$. Assume $\{u, v\}$ is inconsistent and w.l.o.g assume it is local-inconsistent at $u$. Then $u$ must have a proof for the inconsistency, let it be $\triangle uwx \in T(u)$. Now $w$ or $x$ must be in $I$ so from previous lemma $d(w, x) < \frac{\sqrt{3}}{4} r$. Since $d(u, v)$ is also less than $\frac{\sqrt{3}}{4} r$ both $x, w$ are in $N(u, v)$ and $\triangle uwx$ cannot be a proof. Contradiction. \(\square\)

Now we can also prove the upper bound on the number of messages.

**Proof of Theorem 3.3.** As before $G(n, r)$ is well-distributed w.h.p. There is only one step of communication and messages are sent only from nodes with local-inconsistent edges. From Lemma 3.11 only edges $\{\{u, v\} \mid u, v \notin I\}$ can be inconsistent. The result follows since there are $\Theta(rn) = \Theta(\sqrt{n} \log n)$ nodes outside $I$ and each sends at most 6 proofs. \(\square\)

Using bins we can show that there are $\Theta(\sqrt{n/\log n})$ bins outside $I$, each with $O(\log n)$ nodes. We conjecture that only constant number of inconsistent edges exist between neighboring bins (note that each bin is a clique so a bin cannot contain inconsistent edge) and state the following:

**Conjecture 3.12.** For $r \geq \sqrt{\frac{32}{3} r_{\text{con}}}$ the number of messages in Algorithm 1 is $O(\sqrt{\frac{n}{\log n}})$. 

65
Figure 3.5 shows early results of the number of messages in Algorithm 1 for different sizes of random networks, ranging from 100 to 3200 nodes. We choose $r$ to be the same as in the theorems and plot the average number of messages for 10 different runs. These results are compared with the plot of $n/\log n$.

3.5 Notes and Related Work

**Unit Disk Graphs:** The *Gabriel Graph, GG(G)*, [GS69] is a graph where there is an edge \( \{u, v\} \) if and only if there is no other node in \( \text{disk}(u, v) \). Bose *et al* [BMS99] offered a distributed local algorithm to construct the Gabriel Graph over a wireless network and then used face routing to guarantee message delivery. Later, Bose and Morin [BM99] considered different face routing methods in triangulation and in particular in the Delaunay triangulation. In [KK00], Karp and Kung independently proposed GPSR, a memory-less routing algorithm that combines greedy forwarding and local minimum recovery that is also based on
face routing over the Gabriel Graph. Subsequently, later work aimed at finding better planar graphs that can be constructed locally.

Gao et al [GGH01] proposed the use of a Restricted Delaunay Graph $RDG(G)$, a graph that contains all the short edges of the Delaunay Graph and is also planar. The authors proved that $RDG(G)$ is Euclidean spanner of the Unit Disk Graph $G$ and gave an algorithm to construct it that, in general, can be inefficient with $O(n^2)$ messages. Similarly, Li et al [LCW02] proved that the $UDel(G)$ is a spanner of the Unit Disk Graph $G$ and offered a local algorithm to build a planar supergraph of $UDel(G)^4$ called $PLDel(G)$ in $\Theta(n)$ messages and $\Theta(n \log n)$ bits. They presented yet another graph, $LDel^{(k)}(G)$, a local Delaunay triangulation where the circumcircle of $u, v, w$ does not contain any node which is $k$ hops away from $u, v$ or $w$. The authors proved that $LDel^{(k)}(G)$, $k \geq 1$ is a supergraph of $UDel(G)$ and a sub-graph of $LDel^{(k+1)}(G)$ and therefore a spanner. In addition they showed that for $k = 1$ $LDel^{(k)}(G)$ is not planar, but for $k > 1$ it is. Recently Wang and Li [WL03] showed how to bound the maximum degree of such graphs, since $PLDel(G)$ or in general $UDel(G)$ are not bounded degree graphs. Arajo and Rodrigues [AR04] reduced the number of steps in [LCW02] but their algorithm still has the same order of messages, $\Theta(n)$. All of the above algorithms are non adaptive: in some cases they send unnecessary messages. Essentially, they require each node $u$ to broadcast all the triangles in $T(u)$ with a $\angle wux \geq \pi/3$. Since the total number of such triangles (faces) in the above graphs is linear, all the algorithms require a linear number of messages.

Random Geometric Graphs: Bose et al [BDE02] proved (among other results) that the Gabriel Graph is not a spanner of the Unit Disk Graph and that, in the worst case, its stretch factor is $\Theta(\sqrt{n})$. Moreover, they proved that for

\footnote{$RDG(G)$ in [GGH01] notation.}
random geometric graphs in the unit square the stretch factor of the Gabriel Graph is \( w.h.p. \Theta(\sqrt{\log n / \log \log n}) \), which proves its inefficiency for face routing in random networks. In [KLS04], Kozma et al bounded the longest edge of \( \text{Del}(G) \) in a random geometric graph in the unit disk. They showed that, due to boundary effects, the longest edge is of order \( O(\sqrt{n / \log n}) \), an order larger than \( r_{\text{con}} \), and left open the question of an algorithm for the case where \( r = O(r_{\text{con}}) \).

Bern et al [BEY91] proved that the Delaunay triangulation of a uniform set of points does not have bounded degree, and that the maximum degree grows like \( \Theta(\log n / \log \log n) \). In particular, they showed that this does not happen next to the boundary. Since higher degree leads to greater load imbalance, one wants to have a constant degree planar graph; in our case, the algorithm we offer does not solve this problem.

### 3.6 Conclusions

We offer a novel local algorithm to construct planar spanner in random wireless networks. Previous algorithms for computing restricted Delaunay graphs send a message for each triangle in the Restricted Delaunay Graph, and in particular by the node with the largest angle. On the contrary, avoids sending unnecessary messages far from the boundary and thus reduces the total number of messages from \( \Theta(n) \) to \( O(\sqrt{n \log n}) \). Moreover, our results are stated in terms of well-distributed graphs, deterministic or random, and thus can be applied to more general graphs than the ones discussed here.
CHAPTER 4

Random Distance Graphs

4.1 Introduction

In this chapter we define a new class of random graphs, an extension of random geometric graphs called random distance graphs. This class can capture, as extreme cases, both random geometric graphs and Bernoulli random graphs and helped explain some of the similarities between them. Interestingly, neither $\mathcal{G}(n, r)$ nor $\mathcal{B}(n, p)$ are suitable to model social networks. On the contrary, we show that a typical case of random distance graph is adequate to model social networks and, in particular, is a ”Small World” graph, capturing both high clustering and small average path length. As opposed to previous Small World models that rely on deterministic sub-structures to grantee connectivity, random distance graphs offer a completely randomized model with a proven connectivity threshold.

The critical parameter for connectivity, both for $\mathcal{G}(n, r)$ and $\mathcal{B}(n, p)$, has been of special interest. In what may seem surprising, it has been shown that both graphs have closely related critical thresholds for the radius and the edge probability. In particular if $\pi r^2 = p = \frac{\log n + \gamma_n}{n}$ then both $\mathcal{G}(n, r)$ and $\mathcal{B}(n, p)$ are connected with probability going to one as $n \to +\infty$ iff $\gamma_n \to +\infty$ [Pen97, GK98, ER59]. This gives rise to the natural question of how to explain the similarities in the connectivity threshold between these two graphs? Recently, more results on random geometric graphs have been coupled to known results on Bernoulli graphs,
revealing some interesting similarities around the critical connectivity thresholds, for example both graphs have sharp threshold for all monotone properties [GRK04b] and both graphs have optimal cover time near the connectivity threshold [CF03, AE05b] (and Chapter 2 here).

Note, however, that these two graphs have quite different characteristics: in $\mathcal{B}(n, p)$ nodes appear to be only ”place holders” for random edges selection, while in $\mathcal{G}(n, r)$ nodes are assigned some properties (i.e. coordinates $x$ and $y$), on which the existence of edges depends. This leads to major differences in structure. For example, under the connectivity regime, $\mathcal{B}(n, p)$ has small diameter, and no small cliques (and, in particular, small cycles are sufficiently far apart), while $\mathcal{G}(n, r)$ has large diameter and many small cliques. Another important distinction between these graphs concerns the probability that given the existence of edges $\{k, i\}$ and $\{k, j\}$ there is also an edge $\{i, j\}$. (i.e. the conditional probability $P(\{i, j\} \mid \{k, i\}, \{k, i\})$.) In $\mathcal{B}(n, p)$ these events are independent so $P(\{i, j\} \mid \{i, k\}, \{j, k\}) = P(\{i, j\})$, while in $\mathcal{G}(n, r)$ the two events are not independent and $P(\{i, j\} \mid \{i, k\}, \{j, k\}) \gg P(\{i, j\})$.

This lack of independence, also called ”locality”, is one of two important properties of social networks, and in particular, Small World graphs [WS98]. The locality property is measured by the clustering coefficient [WS98]: a number between 0 and 1 that reflects the fraction of a vertex’s neighbors which are neighbors themselves. This reflects the observation that in social networks people that have a common friend are more likely to be friends with each others than two people chosen at random from the population; this intuition is confirmed by a wide range of real life data [WS98, NWS02]. The other major property of Small World graphs is the existence of a small average path length between nodes. (and from an algorithmic perspective, the ability to find such a path locally [Kle00].)
Interestingly, in the evolution stage of connectivity $G(n, r)$ displays locality but not small path length, and $B(n, p)$ displays small path length but not locality, so neither graphs are adequate to model social networks.

This motivates our investigation of whether a more general class of random graph exist, one that contains both $G(n, r)$ and $B(n, p)$ as special pathological cases, but whose typical case, in the evolution stage of connectivity, is adequate to model social networks. That is a class of graphs that present both locality and small average path length.

4.2 Definitions and Statement of Results

Let $U$ be the unit area disk centered at the origin.

**Definition 4.1** (Random Distance Graph). For $n$ nodes and a connection function $g : [0, \frac{2}{\sqrt{\pi}}] \rightarrow [0, 1]$ let $\mathcal{D}(n, g)$ be a random distance graph resulting from the following process. First place $n$ nodes uniformly at random in the unit disk $U$. Second, for each pair of nodes $i, j$ with distance $d(i, j)$ place an edge between $i$ and $j$ independently from all other edges with probability $p_{ij} = g(d(i, j))$.

Intuitively, we add edges between nodes as a function of their distance and therefore the underlying structure of the graph depends on the connection function $g$. Here we primarily consider a specific type of function, a step function $g_{r, \alpha}$, with parameters $r$ and $\alpha$. This function creates edges with probability $\alpha = \alpha(n)$ for nodes at distance less or equal to a radius $r = r(n)$ (short edges), and with probability $\beta = \beta(n)$ for nodes at distance larger than $r$ (long edges). Moreover we choose $\beta(\alpha)$ as a function of $\alpha$, and in order to keep the average degree of the graph invariant with respect to $\alpha$, we require:

$$ (1 - \alpha)\pi r^2 = \beta(1 - \pi r^2) \quad (4.1) $$
Formally, for $\pi r^2 \leq \alpha \leq 1$ the function $g^\alpha_r$ is defined as follow:

$$g^\alpha_r(d) = \begin{cases} 
\alpha & \text{if } d \leq r, \\
\beta = \frac{(1-\alpha)\pi r^2}{1-\pi r^2} & \text{if } d > r.
\end{cases}$$ (4.2)

note that this guarantees that the integral of $g^\alpha_r$ over $\mathcal{U}$, relative to the origin, is also invariant with respect to $\alpha$:

$$\int_{\mathcal{U}} g^\alpha_r(x)dx = \int_0^r 2\pi x\alpha dx + \int_r^{\frac{1}{2}} 2\pi x\beta dx = \pi r^2$$

For $\pi r^2 \leq \alpha \leq 1$, $\mathcal{D}(n, g^\alpha_r)$ is an extension of random geometric graphs that can capture both $\mathcal{B}(n, p)$ and $\mathcal{G}(n, r)$. On one hand when $\alpha = 1$ we have $\mathcal{D}(n, g^1_r) \equiv \mathcal{G}(n, r)$, a random geometric graph. (where $\equiv$ stand for the same generating process.) On the other hand when $\alpha = \pi r^2$ from (4.1) we get that $\beta = \pi r^2$.

Since $\alpha = \beta$ edges do not depend any more on the distance and every edge is chosen independently with probability $p = \alpha$ so $\mathcal{D}(n, g^\pi r^2) \equiv \mathcal{B}(n, p = \pi r^2)$ which is the original random Bernoulli graph. Note that both these graphs, as well as $\mathcal{D}(n, g^\alpha_r)$ for any $\pi r^2 \leq \alpha \leq 1$, have the same average degree, $\delta_{\text{avg}} = (n - 1)\pi r^2$ and the same expected number of edges $\binom{n}{2}\pi r^2$.

Our first result extends the previous result on the sharp threshold for connectivity to $\mathcal{D}(n, g^\alpha_r)$:

**Theorem 4.1** (Connectivity). Let $\pi r^2 = \frac{\log n + \gamma_n}{n}$. Then for $\pi r^2 \leq \alpha \leq 1$, $\mathcal{D}(n, g^\alpha_r)$ is connected w.h.p. iff $\gamma_n \to \infty$ and is disconnected w.h.p. iff $\gamma_n \to -\infty$.

Next we can prove the following the about clustering and the diameter of $\mathcal{D}(n, g^\alpha_r)$:

**Theorem 4.2** (Clustering). Let $\pi r^2 = \frac{\log n + \gamma_n}{n}$ and $\gamma_n \to \infty$. For $\pi r^2 \leq \alpha \leq 1$ the cluster coefficient of $\mathcal{D}(n, g^\alpha_r)$ represented as $C = P(\{i, j\} \mid \{k, i\}, \{k, j\})$ is w.h.p.$C = \alpha \ast 0.5865 + o(1)$.
Theorem 4.3 (Diameter). Let $\pi r^2 = \frac{\log n + \gamma_n}{n}$ and $\gamma_n \to \infty$. For a constant $\epsilon > 0$ and $\pi r^2 \leq \alpha \leq 1 - \epsilon$ the diameter of $D(n, g^\alpha_r)$ is w.h.p. $\Theta\left(\frac{\log n}{\log \log n}\right)$.

Another important perspective of Small World graphs mentioned earlier is the algorithmic one [Kle00]. From this perspective, motivated by the original Milgram experiment [Mil67], social network not only have a short average path length, but enable for such a path to be found locally in a distributed manner, namely by local routing. Local routing is a mechanism for which a message is sent from a source to a destination using only local information available at each node. The destination’s location is known (for example in the message header), and each node can only forward a message to one of its immediate neighbors based on their locations, thus we can route the message in a distributed way without global knowledge. We are interested in a graph that permits short local routing. Namely, the expected route length (number of steps) from source to destination is on the same order as the graph diameter. Finally we can show the following:

Theorem 4.4 (Local Routing). There exist a random distance graph $D(n, g)$ that has the following properties with high probability: It is connected, has average degree $\Theta(\log n)$, high clustering, small diameter and short local routing.

4.3 Proofs

Recall that $\mathcal{U}$ is the unit area disk at the origin. Let $\bar{\mathcal{U}}$ be the disk of radius $\frac{1}{\pi} - r$ and let $\Delta \mathcal{U}$ be $\mathcal{U} \setminus \bar{\mathcal{U}}$. Let $\text{disk}_r(i)$ be the disk of radius $r$ centered at $X_i$, and $\text{lune}_r(i, j)$ be the intersection of $\text{disk}_r(i)$ and $\text{disk}_r(j)$. (we omit $r$ where there is no confusion.) If, for nodes uniformly distributed in $\mathcal{U}$, it is given that a node $i \in \text{disk}(j)$ then $i$ is uniformly distributed in $\text{disk}(j)$ and if it is given that
i \notin \text{disk}(j) \text{ then } i \text{ is uniformly distributed in } U \setminus \text{disk}(j).

4.3.1 Proof of Theorem 4.1 (Connectivity)

Proof. We make use of results from \textit{continuum percolation}, in particular a model that was first introduced by Gilbert [Gil61] and later analyzed rigorously by Penrose [Pen91]. Here we concern ourselves only with the two dimensional case; Let $g(x), x \in \mathbb{R}^2$, be a measurable function taking values in $[0, 1]$ such that

\begin{equation}
 g(x) = g(|x|), \quad x \in \mathbb{R}^2 \tag{4.3}
\end{equation}

\begin{equation}
 0 < \int_{\mathbb{R}^2} g(x) dx < \infty \tag{4.4}
\end{equation}

Let $\mathcal{P}$ be a homogeneous Poisson process on $\mathbb{R}^2$ with rate $\rho$: the expected number of point in any region is equal to the area of the region multiplied by $\rho$. Let $\{X_1, X_2, X_3 \ldots \}$ be the set of points placed by $\mathcal{P}$ in $\mathbb{R}^2$. In addition, a point $X_0 = 0$ is added and when considering this point at the origin it is assumed to be an "arbitrary point of the Poisson process" [Pen91].

Let $\mathcal{P}(\rho, g)$ denote the following random graph: Given a generalization of $\mathcal{P} \cup X_0$ we connect every two points $\{X_i, X_j\} i \neq j$ with probability $g(d(X_i, X_j))$, independently of any other pair of points. The connected components of $\mathcal{P}(\rho, g)$ are called \textit{clusters}, and let $C(0)$ be the "cluster at the origin", the set of points that have a path to $X_0$ in $\mathcal{P}(\rho, g)$. Let $\text{card}(C(0))$ be the cardinality of $C(0)$. For a given $g(x), \mathcal{P}(\rho, g)$ and an integer $k$, let $q_k(\rho)$ denote the probability that $C(0)$ has $k$ points, i.e. the probability that $\text{card}(C(0)) = k$

We make use of the following results:

I. \textbf{Probability of Isolation} (Theorem 3 [Pen91]): If $g$ enclose zero (essentially $g$ is symmetric, $g$ has bounded support, and $g$ is bounded away from zero in some open neighborhood of the surface, see
all the functions considered in this paper encloses zero) then for large $\rho$ the probability of the origin to be in a size one cluster and the probability of the origin to be in any finite size cluster are the same:

$$
\frac{P_\rho(\text{card}(c(0)) < \infty)}{P_\rho(\text{card}(C(0)) = 1)} = \frac{1}{q_1(\rho)} \sum_{k=1}^{\infty} q_k(\rho) \rightarrow 1 \text{ as } \rho \rightarrow \infty \quad (4.5)
$$

In other words, for large $\rho$, w.h.p. the point at the origin is either isolated or connected to an infinite cluster.

II. **Uniqueness Theorem** (Theorem 6.3 [MR96]): w.h.p. there is at most one infinite cluster in $\mathcal{P}(\rho, g)$.

Putting these results together we get that in order to bound the probability of $\mathcal{P}(\rho, g)$ to be connected (for an appropriate $g$), it is sufficient to bound $q_1(\rho)$, the probability that the point at the origin is isolated, i.e. $\text{card}(C(0)) = 1$. In [Pen91] it has been shown that

$$
q_1(\rho) = P_\rho[\text{card}(C(0)) = 1] = \exp\left(-\rho \int_{\mathbb{R}^d} g(x) dx\right) \quad (4.6)
$$

If we let $g(x) = g_r^\alpha$ and $\rho = n$, then we can bound the probability of isolation at the origin:

$$
\int_{U} g_r^\alpha(x) dx = \alpha(\pi r^2) + \beta(1 - \pi r^2)
= \alpha(\pi r^2) + \frac{(1 - \alpha)\pi r^2}{1 - \pi r^2}(1 - \pi r^2)
= \pi r^2
= \frac{\log n + \gamma_n}{n}
$$
So from (4.6) the probability of the origin to be isolated is:

\[ q_1(n) = \exp \left( -n \int_{\mathbb{R}^2} g(x) dx \right) \]

\[ = \exp \left( -n \frac{\log n + \gamma_n}{n} \right) \]

\[ = \exp(- \log n - \gamma_n) \]

\[ = \frac{1}{n} e^{-\gamma_n} \]

Let \( E^1(G) \) be the expected number of order 1 components in \( \mathcal{U} \) for random graph \( G \) and \( P^1(G) \) the probability that there is at least one order 1 component in \( \mathcal{U} \) for \( G \). Since we are still (not for long) considering \( \mathbb{R}^2 \) any other point in \( \mathcal{U} \) has the same probably to be isolated and we have

\[ P^1(\mathcal{P}(n, g)) \leq E^1(\mathcal{P}(n, g)) = n q_1(n) = e^{-\gamma_n} \quad (4.7) \]

When \( \lim_{n \to \infty} \gamma_n = \infty \) then \( P^1(\mathcal{P}(n, g)) \to 0 \) and the graph is connected with high probability since every point is in the unique infinite cluster. The problem is that this result relays on the fact that \( \mathcal{P}(n, g) \) is defined over \( \mathbb{R}^2 \) and each point in \( \mathcal{U} \) is isomorphic to the origin. In our case we are concerned with a process that throws points only inside the unit disk, and the problem of the border’s effect arise.

For a connection function \( g(x) \) let \( \mathcal{P}^\mu(n, g) \) be the graph resulting from the Poisson process with density \( n \) on the unit disk \( \mathcal{U} \) and \( \mathcal{D}(n, g) \), as before, the graph resulting from uniformly distributing \( n \) points in \( \mathcal{U} \). Let \( I \) be the indicator function and for a random graph \( \mathcal{D}(j, g) \) with \( j \) nodes let \( \mathcal{D}(j, g)_i \) stand for the event "\( X_i \) is isolated in \( \mathcal{D}(j, g) \)" (and similarly we have \( \mathcal{G}(j, r)_j \) and \( \mathcal{B}(j, p)_j \)). By
definition of the Poisson process and expectation we have:

\[
P^1(\mathcal{P}^{\text{id}}(n, g)) = \sum_{j=1}^{\infty} P^1(\mathcal{D}(j, g)) e^{-n \frac{n^j}{j!}}
\]

\[
\leq \sum_{j=1}^{\infty} E^1(\mathcal{D}(j, g)) e^{-n \frac{n^j}{j!}}
\]

\[
= \sum_{j=1}^{\infty} E\left[\sum_{i=1}^{j} I_{\mathcal{D}(j,g)_i}\right] e^{-n \frac{n^j}{j!}}
\]

\[
= \sum_{j=1}^{\infty} jP(\mathcal{D}(j, g)_j)e^{-n \frac{n^j}{j!}}
\]

In [GK98] Gupta and Kumar proved that for \( \pi r^2 = \frac{\log n + \gamma n}{n} \)

\[
P^1(\mathcal{P}^{\text{id}}(n, g^1_r)) \leq \sum_{j=1}^{\infty} j P(\mathcal{D}(j, g^1_r)_j)e^{-n \frac{n^j}{j!}} \leq e^{-\gamma n} \quad (4.8)
\]

More over they bound the probability \( P_{\text{diss}} \) that \( \mathcal{D}(n, g) \) is disconnected\(^1\). For \( \epsilon > 0 \)

\[
P_{\text{diss}}(\mathcal{D}(n, g)) \leq 2(1 + 6\epsilon) \left[ P^1(\mathcal{P}^{\text{id}}(n, g)) + \sum_{j=1}^{\infty} j P(\mathcal{D}(j, g^1_r)_j)e^{-n \frac{n^j}{j!}} \right] \quad (4.9)
\]

In particular for \( \pi r^2 = \frac{\log n + \gamma n}{n} \) and \( \lim_{n \to \infty} \gamma n = c \), using (4.8) they proved:

\[
\limsup_{n \to \infty} P_{\text{diss}}(\mathcal{G}(n, r)) = \limsup_{n \to \infty} P_{\text{diss}}(\mathcal{D}(n, g^1_r)) \leq 4e^{-c}
\]

To prove our result we want to show that for \( \pi r^2 \leq \alpha \leq 1 \) and \( 1 \leq j \)

\[
\limsup_{n \to \infty} P_{\text{diss}}(\mathcal{D}(n, g^0_r)) \leq 4e^{-c}
\]

This will follow from (4.8) and (4.9) and by showing that for \( \pi r^2 \leq \alpha \leq 1 \)

\[
P(\mathcal{D}(j, g^1_r)_j) \geq P(\mathcal{D}(j, g^0_r)_j) \geq P(\mathcal{D}(j, g^{\pi r^2}_r)_j) \quad (4.10)
\]

\(^1\)They prove it for \( \mathcal{G}(n, r) \), but the result follows for the general case.
meaning that, for any \( \alpha \) in the range, the probability that a node is isolated in \( \mathcal{D}(n, g^\alpha) \) is smaller than it is in random geometric graph \( \mathcal{G}(n, r) \) and larger than in random Bernoulli graph \( \mathcal{B}(n, p = \pi r^2) \).

There are two cases: (i) \( X_j \in \bar{U} \) and (ii) \( X_j \in \Delta U \). For the first case we have that for \( \mathcal{G}(j, r) \) and \( \mathcal{B}(j, \pi r^2) \) the probability that \( X_j \) is isolated is \((1 - \pi r^2)^{j-1}\).

For \( \mathcal{D}(j, g^\alpha) \) we have:

\[
P(\mathcal{D}(j, g^\alpha), j) = ((1 - \alpha)\pi r^2 + (1 - \beta)(1 - \pi r^2))^{j-1} 
= ((1 - \alpha)\pi r^2 + (1 - \frac{(1 - \alpha)\pi r^2}{1 - \pi r^2})(1 - \pi r^2))^{j-1} 
= ((1 - \alpha)\pi r^2 + (1 - \pi r^2) - (1 - \alpha)\pi r^2)^{j-1} 
= (1 - \pi r^2)^{j-1}
\]

For (ii) \( X_j \) is less than \( r \) away from the border of \( U \) and let \( A = disk_r(X_j) \cap U < \pi r^2 \). Then for \( \pi r^2 < \alpha < 1 \) we have

\[
P(\mathcal{D}(j, g^\alpha), j) = ((1 - \alpha)A + (1 - \beta)(1 - A))^{j-1} 
= ((1 - \alpha)A + (1 - A) - \frac{(1 - \alpha)\pi r^2}{1 - \pi r^2}(1 - A))^{j-1}
\]

When \( \alpha = 1 \) we get the case for \( \mathcal{G}(j, r) \), \( P(\mathcal{G}(j, r), j) = (1 - A)^{j-1} \) and when \( \alpha = \pi r^2 \) we get the case for \( \mathcal{B}(j, \pi r^2) \), \( P(\mathcal{B}(j, \pi r^2), j) = (1 - \pi r^2)^{j-1} \).

To confirm that (4.10) hold also for case (ii) it is sufficient to show that

\[
\frac{\partial}{\partial \alpha} P(\mathcal{D}(j, g^\alpha), j) \geq 0,
\]

and in particular it is enough to show that

\[
\frac{\partial}{\partial \alpha} \left( (1 - \alpha)A + (1 - A) - \frac{(1 - \alpha)\pi r^2}{1 - \pi r^2}(1 - A) \right) = \frac{\pi r^2}{1 - \pi r^2}(1 - A) - A 
\geq \pi r^2 - A \geq 0
\]

The second part of the theorem also follow from (4.10). Since it is known that for \( \mathcal{B}(n, \pi r^2) \) and \( \gamma_n \to -\infty \), \( \mathcal{B}(n, \pi r^2) \) is disconnected w.h.p., and in particular
that there is an isolated node, from (4.10) the same will hold for \( D(n, g^n_\alpha) \) and \( \alpha \geq \pi r^2 \).

\[ \square \]

### 4.3.2 Proof of Theorem 4.2 (Clustering)

**Proof.** When computing the conditional probability \( P(\{i, j\} \mid \{k, i\}, \{k, j\}) \) we ignore the border effect (nodes in \( \Delta U \)) since the fraction of these nodes is vanishing and their contribution to the final result goes to zero. We can consider three disjoint cases: (i) \( i, j \in \text{disk}(k) \). (ii) one of \( i, j \) is in \( \text{disk}(k) \) and the other is not. (iii) \( i, j \notin \text{disk}(k) \). First let’s define the following helping probabilities:

\[
\begin{align*}
p^* &= P(j \in \text{disk}(i) \mid j, i \in \text{disk}(k)) = P(j \in \text{lune}(k, i) \mid j, i \in \text{disk}(k)) \\
p^{**} &= P(j \in \text{disk}(i) \setminus \text{lune}(k, i) \mid j \notin \text{disk}(k) \land i \in \text{disk}(k))
\end{align*}
\]

Let \( i \in \text{disk}(k) \) and let \( y = 2x \) be the distance between \( k \) and \( i \). Then \( \text{lune}(k, i) \) is equal to twice the half lune (see Fig. 4.1):

\[
\text{lune}(k, i) = r^2 \left( 2 \arccos\left(\frac{x}{r}\right) - \sin\left(2 \arccos\left(\frac{x}{r}\right)\right) \right) \tag{4.11}
\]
taking the integral over \( ydy = 2x^2dx \) we get:

\[ p^* = \int_0^{r/2} \frac{2\pi x^2}{\pi r^2} \cdot 2 \arccos\left(\frac{x}{r}\right) - \sin(2 \arccos\left(\frac{x}{r}\right)) \, dx \]

\[ = \frac{8}{\pi r^2} \int_0^{r/2} x^2 \arccos\left(\frac{x}{r}\right) - \sin(2 \arccos\left(\frac{x}{r}\right)) \, dx \]

\[ = \frac{8}{\pi r^2} \left( -\frac{x(r^2 + 2x^2)}{2} \arccos\left(\frac{x}{r}\right) + \frac{4r}{3} \arccos\left(\frac{x}{r}\right) + r^3 \arcsin\left(\frac{x}{r}\right) \right) \bigg|_0^{r/2} \]

\[ = \frac{2}{\pi r^3} \left( -\frac{r}{2}(r^2 + \frac{r^2}{2}) \frac{\sqrt{3}}{2} + r^3 \frac{\pi}{3} + r^3 \frac{\pi}{6} - 0 \right) \]

\[ = \frac{2}{\pi} \left( -\frac{\sqrt{33}}{8} + \frac{\pi}{3} + \frac{\pi}{6} \right) \]

\[ \approx 0.5865 \]

and

\[ p^{**} = \int_0^{r/2} \frac{2\pi x^2}{\pi r^2} \left( \frac{\pi r^2 - r^2 (2 \arccos\left(\frac{x}{r}\right) - \sin(2 \arccos\left(\frac{x}{r}\right)))}{1 - \pi r^2} \right) \, dx \]

\[ = \frac{\pi x^2}{1 - \pi r^2} - \frac{8}{1 - \pi r^2} \left( -\frac{x(r^2 + 2x^2)}{2} \arccos\left(\frac{x}{r}\right) + \frac{4r}{3} \arccos\left(\frac{x}{r}\right) + r^3 \arcsin\left(\frac{x}{r}\right) \right) \bigg|_0^{r/2} \]

\[ = \frac{2\pi r^2}{4(1 - \pi r^2)} - \frac{2r^2}{1 - \pi r^2} \left( \frac{\sqrt{33}}{8} + \frac{\pi}{3} + \frac{\pi}{6} \right) \]

\[ = o(1) \]

where the last step is a result of \( \pi r^2 = o(1) \). Now we can calculate the clustering coefficient as a function of \( \alpha \) (note that \( \beta = o(1) \) and \( \pi r^2 = o(1) \)):

\[ P(\{i, j\} \mid \{k, i\}, \{k, j\}) = P(\{i, j\} \mid \{k, i\}, \{k, j\}, (i)) + P(\{i, j\} \mid \{k, i\}, \{k, j\}, (ii)) \]

\[ + P(\{i, j\} \mid \{k, i\}, \{k, j\}, (iii)) \]

\[ = (\alpha p^* + \beta (1 - p^*)) + (\alpha p^{**} + \beta (1 - p^{**})) + (\alpha \pi r^2 + \beta (1 - \pi r^2)) \]

\[ = \alpha p^* + o(1) \]

\[ \approx \alpha \cdot 0.5865 + o(1) \]
4.3.3 Proof of Theorem 4.3 (Diameter)

Proof. It is well known that the diameter of a connected $\mathcal{B}(n, p)$ is $\frac{\log n}{\log np}$ and that the unique giant component of order $n$ emerge when $p > \frac{1}{n}$ [Pal85]. Since in our case $\alpha < 1 - \epsilon$ is bounded away from 1 we get that $\beta \geq \frac{\epsilon\pi r^2}{1-\pi r^2} \geq \frac{\epsilon' \log n}{n}$ for a constant $\frac{\epsilon}{1-\pi r^2} \geq \epsilon' > 0$. The graph $\mathcal{D}(n, g^\alpha_r)$ can be thought of as being built in two phases: in the first phase we give each node its location and create $\mathcal{B}(n, p)$ with $p = \beta$, and in the second phase we add the rest of the short edges with the appropriate probability. (i.e. $\frac{\alpha-\beta}{1-\beta}$.) For the above $p = \beta$ we can say the following about $\mathcal{B}(n, p)$ and its giant component [Pal85]: Let $I(n)$ be the set of nodes not in the giant component, then the expected size of $I(n)$ is $E(|I(n)|) = \frac{x(c)}{2c} n$ where $x(c)$ is the fraction of nodes not in the giant component, and:

$$x(c) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (2ce^{-2c})^k$$

where $c > 1/2$ is defined by $p = 2c/n$. In our case $c \geq \epsilon' \log n \rightarrow \infty$ so $x(c)/2c \rightarrow 0$. In particular

$$E(|I(n)|) = \frac{x(c)}{2c} n = O(n^{1-2\epsilon'})$$

From [CL01] we have that for $\log n > np \rightarrow \infty$ the diameter of the giant component of $\mathcal{B}(n, p)$ is $(1 + o(1))\frac{\log n}{\log np}$. To prove the theorem we need to show that there is no path longer than $\frac{\log n}{\log np}$ in $\mathcal{D}(n, g^\alpha_r)$ from nodes not in the giant component of $\mathcal{B}(n, \beta)$. Since we already proved that $\mathcal{D}(n, g^\alpha_r)$ is connected the theorem follows. A necessary condition to have a path longer than $\frac{\log n}{\log np}$ is to have $\frac{\log n}{\log np}$ nodes from $I(n)$ in an area of size less than or equal to $\left(\frac{\log n}{\log np}\right)\pi r^2 = \frac{\log^2 n}{n \log np}$. For $i \in I(n)$ let $S_i$ be the number of nodes from $I(n)$ that are in an area of $\frac{\log^2 n}{n \log np}$ which $i$ belongs to. Let $S = \max\{S_i \mid i \in I(n)\}$. Now since:

$$E(S_i) = E(|I(n)|) \frac{\log^2 n}{n \log np} = O\left(\frac{\log^2 n}{n^{2\epsilon'} \log np}\right) \rightarrow 0$$

(4.12)

81
Figure 4.2: an area that is proportional to $x^2$ when local routing from $i$ to $j$ with $x = d(i, j)$.

it follows that $P(S > \frac{\log n}{\log np}) \to 0$ and so the diameter of $D(n, g^0)$ is of the same order as the diameter of the giant component of $B(n, p = \beta)$.

4.3.4 Proof of Theorem 4.4 (Local Routing)

Proof. The proof follows the same idea of Kleinberg [Kle00], we give a connection function with the following guarantees (in expectation): after a constant number of steps the distance to the destination is reduced by a factor of two. We then modify the function to make sure the graph is also connected with average degree of $\Theta(\log n)$. Let $g'(x) = (x\sqrt{n\pi} + 1)^{-2}$ for $x \in [0, \frac{2}{\pi}]$ and 0 otherwise, so it is defined up to the maximum distance on $U$. Observe that for any two points $i, j \in U$ where $x = d(i, j)$, there is an area $A$ of at least $\pi(\frac{x}{2})^2/6$ s.t $A \subset U$ and all the points in $A$ are at most distance $x/2$ from $j$ and at most distance $x$ from $i$. The expected number of nodes (for large enough $A$) is $\Theta(nA)$. Consider a message at $i$ on its way to $j$, the probability that $i$ will have a long range neighbor $^2A$ is large since for $x < r \log n$ we can route to destination, without using long edges, in $\Theta(\log n)$ steps.
in $A$ is at least

$$n \frac{\pi \left( \frac{x}{2} \right)^2 / 6}{(x / \sqrt{n \pi} + 1)^2} = \frac{\pi n x^2}{24(\pi n x^2 + 2x \sqrt{\pi n} + 1)}$$

$$= \frac{1}{24} - o(1)$$

To guarantee that the graph is connected and that if a step at $i$ fails (there is no long edge to a node in $A$) there is a close neighbor $k$ with the same order of distance to $j$, we can just compose $g'(x)$ with $g_1$ s.t $g(x) = \max\{g'(x), g_1(x)\}$. Now the expected number of steps to reduce the distance to $j$ by a factor of two is constant, and the total number of steps to reach the destination is $\Theta(\log n)$. In this case the average degree of $D(n, g)$ is bounded by:

$$n \int_{\mathbb{R}} g(x) \leq \int_{\mathbb{R}} g'(x) + \int_{\mathbb{R}} g_1^1(x)$$

$$= n \left( \frac{1}{n + 2n^{1.5}} + \frac{\log(1 + 2\sqrt{n})}{n} - \frac{1}{n} \right) + n \frac{\log n + \gamma_n}{n}$$

$$= \Theta(\log n)$$

The high clustering is guaranteed by $\alpha = 1$ and $D(n, g_1^1)$, it can be decreased by taking $\alpha = 1 - \epsilon$ for a small enough $\epsilon$.

### 4.4 Notes and Related Work

This work was influenced by a large variety of recent work on random geometric graphs, and in particular work that exposed the similarities to the Bernoulli model, in connectivity [Pen97, GK98], monotone properties [GRK04b] and the cover time of random walk [AE05b]. The origin of distance graphs goes back to the work of Gilbert [Gil61] and later to the random connection model used in continuum percolation by Penrose and others [Pen03, MPS97]. Those models where concern with a Poisson process on the entire plane while the random distance graph is define on the unit disk and with a uniform point distribution. We
are, however, not aware of previous work that explores the connection between $B(n, p)$ and $G(n, r)$ by showing that the integral over the connection function $g(x)$ is the same for both graphs. In [GK98] Gupta and Kumar conjectured that if $\pi r^2 p(n) = \frac{\log n + \gamma n}{n}$ the same results on connectivity holds, but this seems to be harder case since it increases the border effect. (their proof for connectivity of $G(n, r)$, as well as ours, depends on the negligibility of the effect of nodes that are less than $r$ away from the border.)

In [Kle00] Kleinberg first propose the algorithmic perspective of the Milligram experiment and offered a grid-based distance graph that support local routing. In his model, as well as in the original Small World model of Watts and Strogatz [WS98], the graph starts from a deterministic connected graphs (i.e ring or grid) and random edges are only later introduced or rewired. In Watts and Strogatz model the initial graph is a connected ring of clusters, when edges are rewired the graph become a Small World, at the extreme enough edges are rewired to result in a Bernoulli graph. Random Intersection Graphs (RIG) [Sin95] are another model of random graphs motivated by social structure, in particular by a collaboration networks such as authors-papers or actors-movies, but it does not have any geometric flavor. Under certain parameters values this model is also identical to $B(n, p)$ [FSS00] and under others differs significantly. Unfortunately, across its parameters range it has only two asymptotic clustering coefficient values, either $o(1)$ or $1-o(1)$.

4.5 Conclusions

We offer the perspective that the similarity of the connectivity threshold results for $G(n, r)$ and $B(n, p)$ is a consequence of the integral over the connection function when the nodes are uniformly distributed on the unit disk. In both cases the
integral is $\frac{\log n}{n}$ and it determines the probability of isolated nodes which, in turn, govern the connectivity threshold by a result from continuum percolation. Following this view, we introduce a class of random distance graphs, $D(n, g^c_r)$, with a connection function that has the same integral as $G(n, r)$ and $B(n, p)$. This guarantees that the connectivity threshold is identical for $G(n, r)$ and $B(n, p)$. In addition, we show that for a wide range of parameters this class behaves as a Small World graph, contrary to $G(n, r)$ and $B(n, p)$. As opposed to previous Small World models, we propose a completely random model which seems to be more suitable for real-life situations.

We conjecture that a similar connectivity result can be obtained for more general functions than presented here and we state the following:

**Conjecture 4.5.** Let $c > 1$ be a constant and let $g(x)$ be a function $[0, \frac{2}{\sqrt{\pi}}] \to [0, 1]$ s.t $g(x)$ encloses zero and $\int_U g(x) = \frac{c \log n}{n}$. Then $D(n, g)$ is connected with high probability.

In future work we would like to prove more properties of random distance graphs like the sharp threshold of monotone properties, the cover time and the mixing time and to offer a random scale-free graph that is based on distance graphs.
CHAPTER 5

Experimental Results

In the following chapter, we study in more detail some of the properties of simple random walks and PCT, using simulations. We performed experiments on grids, Hypercube and random geometric graphs $G(n, r(n))$. When not explicitly mentioned we used $n = 4096$ and $R = 0.04$. For each experiment, we took the average results of 100 runs.

5.1 Application Example

Until now we only mentioned briefly applications or tasks that can be done using random walks, but it is clear that many types of queries can be answered using this method, for example; finding the min, max or mean of the data, calculating statistics of the network, finding sensors with specific thresholds and so on. We can even use a SQL like language to describe these queries as offered in [MFH02, BGS00].

To illustrate this ability we choose as an example a simple query, finding the histogram of the data in the network. The procedure is as described earlier, the token is moving in the network, each time it arrives at an unvisited node it updates its histogram (we note that in order to distinguish between visited and unvisited nodes, each node should keep a flag if it has been visited by this walk). After seeing 80% of the nodes the token reports its histogram back.
Figure 5.1: An example of the temperature in an area with six random light sources

Figure 5.2: Comparing the histogram founded by the 80% random walk on the graph and the histogram of the real data from Figure 5.1
Figure 5.3: The progress of partial cover time as function of number of steps normalized to $n$ for different graphs of size $n = 4096$

We wanted to make reasonable assumptions about the data observed by the sensors in order for this task not to be trivial. For example, assuming uniform distribution over the data did not seem reasonable, so we created a model of light sources that will produce a non-uniform temperature distribution over the network, Figure 5.1 is an example of such a distribution. Figure 5.2 compares the histogram obtained from the real data and the one collected by the random walk with 80% cover. As we can see from this example, the histogram is accurate with expected error as low as 0.37%.

### 5.2 Efficiency of Random Walk

Fig. 5.3 represents the number of steps normalized to $N$ (on log scale) as a function of the progress of the partial cover. The Figure presents three different well known graphs: 2−dimensional grid, 3−dimensional grid, and Hypercube, and three different random $\mathcal{G}(4096, r)$ with $r = 0.03, 0.05176, 0.1$ which closely
achieve the same partial cover as the respective known graphs. This exemplified how increasing $r$ change the graph structure drastically. Although both the 3—dimension grid and the Hypercube have optimal cover time the constant in the $O$ notation are different.

From Fig. 5.3, one may directly observe the sharp increase in the number of steps for every graph as the partial cover approaches the full cover. This confirms the non-negligible gap proved in section 2.2.3.2 between the order of the cover time and the order of the PCT for $G(n, r(n))$, which stated that most of the time is spent on the last nodes, further justifying consideration of the partial cover.

Figure 5.4 validates our theoretical result that Partial Cover in sensor networks is scalable. The graph shows the increase in the Partial Cover normalized to $n$ in increasing random network size with constant density. While increasing the network 16 times from 1024 nodes to 16384 the expected number of steps to cover 80% increases from $2.92n$ to $3.37n$. Although there is a small increase, our analytical results states that for large enough $n$ this would remain constant.
5.2.1 Biased Random Walk

Can we improve on the previous result? What if we can direct our random walk toward unvisited nodes? Biased random walk gives priority to unvisited neighbors instead of choosing uniformly at random. We define a bias parameter $0 \leq \text{bias} \leq 1$ and select our next node according to the following rule: Let $d$ be the number of neighbors of the current node, and let $d_u$ be the number of unvisited neighbors. Then: (i) A visited neighbor is selected with probability $(1 - \text{bias})/d$. (ii) An unvisited neighbor is selected with probability $(1 - \text{bias})/d + \text{bias}/d_u$. If all neighbors are already visited and $\text{bias} = 1$, a neighbor selected uniformly at random is returned. When each node knows whether it has been visited then we can execute this protocol without knowledge of neighbors, again using the broadcast nature of the communication.

We realize that $\text{bias} = 1$ cannot be always maintained due to errors, but as we can see in Figure 5.5 the substantial improvement in the number of steps required is obtained even with a small bias. Notice also that covering 80% with bias greater

Figure 5.5: Partial Cover Time in random walks with increasing bias on random network
than 0.8 requires less than $n$ steps. A future improvement can be to have “super bias” which take the neighborhood into consideration the neighborhood. A node will decrease its priority even if it hears the token passing by in his neighborhood. We can do this easily, again, because of the broadcast channel.

5.3 Quality of Random Walk

5.3.1 Partial Cover Quality

When the mixing time is better one would expect that the quality of the partial cover will improve, meaning that the random walk will not leave large contiguous areas in the network uncovered. To make this measurement more precise let $\text{min}(v)$ be the minimum distance from $v$ to a visited node in a (partial) random walk. We define the hole size of a random walk as the maximum of $\text{min}(v)$ over all the nodes in the graph. Note that the hole size is decreasing as the random walk proceeds and more nodes are visited; after cover time it is 0.
Fig. 5.6 presents the decrease in the hole size as a function of number of steps of random walk for $G(n, r(n))$ with increasing $r$. The figure shows that the rate of improvement in the quality is strongly dependent upon $r$, similarly to the mixing rate. Note that each walk was sampled at 10%, 20%, and up to 100% cover where the hole size is 0. An interesting point to discuss is that this experiment also validates the fact that graphs with different spectral gap (and mixing time) can have the same cover time (such as the 3D grid and the hypercube). For example the graphs with $r > 0.06$ seems to have very similar cover time but nevertheless very different partial cover quality.

5.3.2 Robustness to Dynamics

Dynamics in sensor networks are due to many different factors. Nodes can fail, turn off their radio in a duty cycle, or move. Wireless communication obstacles can disconnect links and more. How robust is our process to these dynamics? We know that as long as the network is strongly connected with no bottlenecks, the random walk will do well. Actually we do not care about the whole network, since we only need this requirement in the area of the token. The nice thing is that this condition is, in many cases, orthogonal to the network dynamic. A high duty cycle rate, for example, can still leave the network strongly connected at any time, so the random walk is robust to that. No fault tolerance or recovery mechanism is in needed. As long as the token is alive, there is nothing to recover from.

To illustrate this point we choose the following model of failures. Each node can fail independently with probability $p$ during the run (or nodes switching on-off in a duty cycle). The failure can occur at any time during the walk, so the probability that a node will fail exactly when it has the token is negligible.
Figure 5.7: The Partial Cover time required when the probability $p$ of each node to fail is increasing. The result are for 4096 nodes networks simulated the worst case for the random walk, where all the failures happen at once before the walk starts. It is worth mentioning that for other methods all-fail-at-once may not be the worst case. For example for a spanning tree it will be the best case, while nodes turning on-off during the data collection will be worst. Figure 5.7 shows the result for this model for $p = \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$. The vertical axis is normalized to the number of active nodes in each run. We see that the walks in this case continue to perform well (better than the grid) even at high rate of failure.

What about dependent, or correlated failures? Our second model tries to answer that by creating random areas of “disaster” where at the center of each there is a source and the probability of failures decreases exponentially from the source with parameter $\alpha$ (up to $r$ hops). Figure 5.8 shows an example of a 4096 nodes random network with 4 holes. As we increase the number of holes, we create more bottlenecks. Figure 5.9 supports that the random walk is highly robust to such failures. We can see that the effect of this type of failure is mostly
Figure 5.8: An example of a 4096 random network with 4 disaster areas. We can see the creation of bottlenecks.

Figure 5.9: The Partial Cover time required when we increase the number of disaster areas in the network.
on the covering of the last nodes and not on the Partial Cover.

But a token can still be lost, so what do we do then? First, to solve the problem of unreliable communication which can destroy our token, we suggest using a reliable protocol. The cost of that will not be too high, assuming there is a probability $p$ of each transmission to fail. The expected number of transmission to pass the token reliably (including ACK) will be $\frac{2-p}{(1-p)^2}$. For $p = 0.2$ for example we get 2.8125 transmissions. Second, if we still lost it, we will just issue another one. Since the probability for that is so small, it will not affect the overall performance of a few tokens traveling across the network at the same time.

5.3.3 Load Balancing

Random walk is an uncontrolled process. It may be the case that the walk will go to a neighbor and return back to the same node after one step. Even if we forbid this move, we cannot prevent it from happening in small cycles. This leads to the intuition that there will be nodes that are much more visited than others, so consuming more energy. However, for very long random walks this is not necessarily true. Since this process is a Markov Chain, it is known that the stationary distribution $\pi$ of it is $\pi = \{\pi_1, \ldots, \pi_n\}$ where $\pi_i = \frac{d_i}{2m}$, $d_i$ is the number of neighbors of $i$ and $m$ is the number of edges in the network [AKL79]. So if the graph is regular the stationary distribution will be uniform. This means that after a long enough time the probability of the token to be at any node is the same. From this, it follows that the frequency of returning to a node is the same for all nodes, which gives us load balancing.

As been shown, sensor networks are “almost regular” graphs with same order degree for all nodes by a constant, but, in our case we are issuing “short” random walks not long ones. There is no guarantee that in any given walk there will
Figure 5.10: Histogram of the expected number of visits to a node in a 80% cover random walk

not be nodes that are much more visited than others. Although this is true, our simulation results show that if we issue many (i.e 100) such “short” walks, we will still get very close to this property.

To measure the load balance we looked at the expected number of visits to each node in a 80% cover random walk. Figure 5.10 present the histogram of these values in a random walk with 13100 steps. The mean of the expected number of visits for each node is 3.19 which is approximately 13100/4096 and the standard deviation is 0.81606. This is showing that only small part of the network will be visited much less (more), and will use less (more) energy. For this histogram more than 95% of the nodes are in $mean(x) \pm std(x)$.


**References**


[Che86] P Chew. “There is a planar graph almost as good as the complete graph.” In SCG ’86: Proceedings of the second annual symposium on Computational geometry, pp. 169–177, New York, NY, USA, 1986. ACM Press.


