Hitting and Commute Times in Large Random Neighborhood Graphs

Ulrike von Luxburg   LUXBURG@INFORMATIK.UNI-HAMBURG.DE
Department of Computer Science
University of Hamburg
Vogt-Koelln-Str. 30, 22527 Hamburg, Germany

Agnes Radl   AGNES.RADL@MATH.UNI-LEIPZIG.DE
Institute of Mathematics
University of Leipzig
Augustusplatz 10, 04109 Leipzig, Germany

Matthias Hein   HEIN@CS.UNI-SAARLAND.DE
Department of Mathematics and Computer Science
Saarland University
Campus E1 1, 66123 Saarbrücken, Germany

Editor: Gabor Lugosi

Abstract

In machine learning, a popular tool to analyze the structure of graphs is the hitting time and the commute distance (resistance distance). For two vertices u and v, the hitting time H_{uv} is the expected time it takes a random walk to travel from u to v. The commute distance is its symmetrized version $C_{uv} = H_{uv} + H_{vu}$. In our paper we study the behavior of hitting times and commute distances when the number n of vertices in the graph tends to infinity. We focus on random geometric graphs (ε-graphs, kNN graphs and Gaussian similarity graphs), but our results also extend to graphs with a given expected degree distribution or Erdős-Rényi graphs with planted partitions. We prove that in these graph families, the suitably rescaled hitting time H_{uv} converges to $1/d_v$ and the rescaled commute time to $1/d_u + 1/d_v$ where d_u and d_v denote the degrees of vertices u and v. In these cases, hitting and commute times do not provide information about the structure of the graph, and their use is discouraged in many machine learning applications.

Keywords: commute distance, resistance, random graph, k-nearest neighbor graph, spectral gap

1. Introduction

Given an undirected, weighted graph $G = (V, E)$ with n vertices, the commute distance between two vertices u and v is defined as the expected time it takes the natural random walk starting in vertex u to travel to vertex v and back to u. It is equivalent (up to a constant) to the resistance distance, which interprets the graph as an electrical network and defines the distance between vertices u and v as the effective resistance between these vertices. See below for exact definitions, for background reading see Doyle and Snell (1984); Klein and Randic (1993); Xiao and Gutman (2003); Fouss et al. (2006), Chapter 2 of Lyons and Peres (2010), Chapter 3 of Aldous and Fill (2001), or Section 9.4 of Levin et al. (2008).

The commute distance is very popular in many different fields of computer science and beyond. As examples consider the fields of graph embedding (Guattery, 1998; Saerens et al., 2004; Qiu and Hancock, 2006; Wittmann et al., 2009), graph sparsification (Spielman and Srivastava, 2008), social network analysis (Liben-Nowell and Kleinberg, 2003), proximity search (Sarkar et al., 2008), collaborative filtering (Fouss et al., 2006), clustering (Yen et al., 2005), semi-supervised learning (Zhou and Schölkopf, 2004), dimensionality reduction (Ham et al., 2004), image processing (Qiu and Hancock, 2005), graph labeling (Herbster and Pontil, 2006; Cesa-Bianchi et al., 2009), theoretical computer science (Aleliunas et al., 1979; Chandra et al., 1989; Avin and Ercal, 2007; Cooper and Frieze, 2003, 2005, 2007, 2011), and various applications in chemometrics and bioinformatics (Klein and Randic, 1993; Ivanciuc, 2000; Fowler, 2002; Roy, 2004; Guillot et al., 2009).

The commute distance has many nice properties, both from a theoretical and a practical point of view. It is a Euclidean distance function and can be computed in closed form. As opposed to the shortest path distance, it takes into account all paths between u and v, not just the shortest one. As a rule of thumb, the more paths connect u with v, the smaller their commute distance becomes. Hence it supposedly satisfies the following, highly desirable property:

Property ($\bigstar$): Vertices in the same "cluster" of the graph have a small commute distance, whereas vertices in different clusters of the graph have a large commute distance to each other.

Consequently, the commute distance is considered a convenient tool to encode the cluster structure of the graph.

In this paper we study how the commute distance behaves when the size of the graph increases. We focus on the case of random geometric graphs (k-nearest neighbor graphs, ε-graphs, and Gaussian similarity graphs). Denote by H_{uv} the expected hitting time and by C_{uv} the commute distance between two vertices u and v, by d_u the degree of vertex u, by $\mathrm{vol}(G)$ the volume of the graph. We show that as the number n of vertices tends to infinity, there exists a scaling term sc such that the hitting times and commute distances in random geometric graphs satisfy

$$ sc \cdot \left| \frac{1}{\mathrm{vol}(G)} H_{uv} - \frac{1}{d_v} \right| \longrightarrow 0 \qquad \text{and} \qquad sc \cdot \left| \frac{1}{\mathrm{vol}(G)} C_{uv} - \left(\frac{1}{d_u} + \frac{1}{d_v} \right) \right| \longrightarrow 0, $$

and at the same time $sc \cdot d_u$ and $sc \cdot d_v$ converge to positive constants (precise definitions, assumptions and statements below). Loosely speaking, the convergence result says that the rescaled commute distance can be approximated by the sum of the inverse rescaled degrees.

We present two different strategies to prove these results: one based on flow arguments on electrical networks, and another one based on spectral arguments. While the former approach leads to tighter bounds, the latter is more general. Our proofs heavily rely on prior work by a number of authors: Lovász (1993), who characterized hitting and commute times in terms of their spectral properties and was the first one to observe that the commute distance can be approximated by $1/d_u + 1/d_v$; Boyd et al. (2005), who provided bounds on the

spectral gap in random geometric graphs, and Avin and Ercal (2007), who already studied the growth rates of the commute distance in random geometric graphs. Our main technical contributions are to strengthen the bound provided by Lovász (1993), to extend the results by Boyd et al. (2005) and Avin and Ercal (2007) to more general types of geometric graphs such as k-nearest neighbor graphs with general domain and general density, and to develop the flow-based techniques for geometric graphs.

Loosely speaking, the convergence results say that whenever the graph is reasonably large, the degrees are not too small, and the bottleneck is not too extreme, then the commute distance between two vertices can be approximated by the sum of their inverse degrees. These results have the following important consequences for applications.

Negative implication: Hitting and commute times can be misleading. Our approximation result shows that the commute distance does not take into account any global properties of the data in large geometric graphs. It has been observed before that the commute distance sometimes behaves in an undesired way when high-degree vertices are involved (Liben-Nowell and Kleinberg, 2003; Brand, 2005), but our work now gives a complete theoretical description of this phenomenon: the commute distance just considers the local density (the degree of the vertex) at the two vertices, nothing else. The resulting large sample commute distance $dist(u, v) = 1/d_u + 1/d_v$ is completely meaningless as a distance on a graph. For example, all data points have the same nearest neighbor (namely, the vertex with the largest degree), the same second-nearest neighbor (the vertex with the second-largest degree), and so on. In particular, one of the main motivations to use the commute distance, Property (★), no longer holds when the graph becomes large enough. Even more disappointingly, computer simulations show that n does not even need to be very large before (★) breaks down. Often, n in the order of 1000 is already enough to make the commute distance close to its approximation expression. This effect is even stronger if the dimensionality of the underlying data space is large. Consequently, even on moderate-sized graphs, the use of the raw commute distance should be discouraged.

Positive implication: Efficient computation of approximate commute distances. In some applications the commute distance is not used as a distance function, but as a tool to encode the connectivity properties of a graph, for example in graph sparsification (Spielman and Srivastava, 2008) or when computing bounds on mixing or cover times (Aleliunas et al., 1979; Chandra et al., 1989; Avin and Ercal, 2007; Cooper and Frieze, 2011) or graph labeling (Herbster and Pontil, 2006; Cesa-Bianchi et al., 2009). To obtain the commute distance between all points in a graph one has to compute the pseudo-inverse of the graph Laplacian matrix, an operation of time complexity $O(n^3)$. This is prohibitive in large graphs. To circumvent the matrix inversion, several approximations of the commute distance have been suggested in the literature (Spielman and Srivastava, 2008; Sarkar and Moore, 2007; Brand, 2005). Our results lead to a much simpler and well-justified way of approximating the commute distance on large random geometric graphs.

After introducing general definitions and notation (Section 3), we present our main results in Section 4. This section is divided into two parts (flow-based part and spectral part). In

Section 5 we show in extensive simulations that our approximation results are relevant for many graphs used in machine learning. Relations to previous work is discussed in Section 2. All proofs are deferred to Sections 6 and 7. Parts of this work is built on our conference paper von Luxburg et al. (2010).

2. Related Work

The resistance distance became popular through the work of Doyle and Snell (1984) and Klein and Randic (1993), and the connection between commute and resistance distance was established by Chandra et al. (1989) and Tetali (1991). By now, resistance and commute distances are treated in many text books, for example Chapter IX of Bollobas (1998), Chapter 2 of Lyons and Peres (2010), Chapter 3 of Aldous and Fill (2001), or Section 9.4 of Levin et al. (2008). It is well known that the commute distance is related to the spectrum of the unnormalized and normalized graph Laplacian (Lovász, 1993; Xiao and Gutman, 2003). Bounds on resistance distances in terms of the eigengap and the term $1/d_u + 1/d_v$ have already been presented in Lovász (1993). We present an improved version of this bound that leads to our convergence results in the spectral approach.

Properties of random geometric graphs have been investigated thoroughly in the literature, see for example the monograph of Penrose (2003). Our work concerns the case where the graph connectivity parameter (ε or k, respectively) is so large that the graph is connected with high probability (see Penrose, 1997; Brito et al., 1997; Penrose, 1999, 2003; Xue and Kumar, 2004; Balister et al., 2005). We focus on this case because it is most relevant for machine learning. In applications such as clustering, people construct a neighborhood graph based on given similarity scores between objects, and they choose the graph connectivity parameter so large that the graph is well-connected.

Asymptotic growth rates of commute distances and the spectral gap have already been studied for a particular special case: ε-graphs on a sample from the uniform distribution on the unit cube or unit torus in $\mathbb{R}^d$ (Avin and Ercal, 2007; Boyd et al., 2005; Cooper and Frieze, 2011). However, the most interesting case for machine learning is the case of kNN graphs or Gaussian graphs (as they are the ones used in practice) on spaces with a non-uniform probability distribution (real data is never uniform). A priori, it is unclear whether the commute distances on such graphs behave as the ones on ε-graphs: there are many situations in which these types of graphs behave very different. For example their graph Laplacians converge to different limit objects (Hein et al., 2007). In our paper we now consider the general situation of commute distances in ε, kNN and Gaussian graphs on a sample from a non-uniform distribution on some subset of $\mathbb{R}^d$.

Our main techniques, the canonical path technique for bounding the spectral gap (Diaconis and Strook, 1991; Sinclair, 1992; Jerrum and Sinclair, 1988; Diaconis and Saloff-Coste, 1993) and the flow-based techniques for bounding the resistance, are text book knowledge (Section 13.5. of Levin et al., 2008; Sec. IX.2 of Bollobas, 1998). Our results on the spectral gap in random geometric graphs, Theorems 6 and 7, build on similar results in Boyd et al.

(2005); Avin and Ercal (2007); Cooper and Frieze (2011). These three papers consider the special case of an ε-graph on the unit cube / unit torus in $\mathbb{R}^d$, endowed with the uniform distribution. For this case, the authors also discuss cover times and mixing times, and Avin and Ercal (2007) also include bounds on the asymptotic growth rate of resistance distances. We extend these results to the case of ε, kNN and Gaussian graphs with general domain and general probability density. The focus in our paper is somewhat different from the related literature, because we care about the exact limit expressions rather than asymptotic growth rates.

3. General Setup, Definitions and Notation

We consider undirected graphs $G = (V, E)$ that are connected and not bipartite. By n we denote the number of vertices. The adjacency matrix is denoted by $W := (w_{ij})_{i,j=1,\ldots,n}$. In case the graph is weighted, this matrix is also called the weight matrix. All weights are assumed to be non-negative. The minimal and maximal weights in the graph are denoted by $w_{\min}$ and $w_{\max}$. By $d_i := \sum_{j=1}^n w_{ij}$ we denote the degree of vertex v_i. The diagonal matrix D with diagonal entries $d_1, \ldots, d_n$ is called the *degree matrix*, the minimal and maximal degrees are denoted $d_{\min}$ and $d_{\max}$. The *volume* of the graph is given as $\mathrm{vol}(G) = \sum_{j=1,\ldots,n} d_j$. The *unnormalized graph Laplacian* is given as $L := D - W$, the normalized one as $L_{\mathrm{sym}} = D^{-1/2} L D^{-1/2}$. Consider the natural random walk on G. Its transition matrix is given as $P = D^{-1} W$. It is well-known that λ is an eigenvalue of L_{sym} if and only if $1 - \lambda$ is an eigenvalue of P. By $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > -1$ we denote the eigenvalues of P. The quantity $1 - \max\{\lambda_2, |\lambda_n|\}$ is called the *spectral gap* of P.

The *hitting time* H_{uv} is defined as the expected time it takes a random walk starting in vertex u to travel to vertex v (where $H_{uu} = 0$ by definition). The *commute distance* (*commute time*) between u and v is defined as $C_{uv} := H_{uv} + H_{vu}$. Closely related to the commute distance is the *resistance distance*. Here one interprets the graph as an electrical network where the edges represent resistors. The conductance of a resistor is given by the corresponding edge weight. The resistance distance R_{uv} between two vertices u and v is defined as the effective resistance between u and v in the network. It is well known (Chandra et al., 1989) that the resistance distance coincides with the commute distance up to a constant: $C_{uv} = \mathrm{vol}(G) R_{uv}$. For background reading on resistance and commute distances see Doyle and Snell (1984); Klein and Randic (1993); Xiao and Gutman (2003); Fouss et al. (2006).

Recall that for a symmetric, non-invertible matrix A its Moore-Penrose inverse is defined as $A^\dagger := (A+U)^{-1} - U$ where U is the projection on the eigenspace corresponding to eigenvalue 0. It is well known that commute times can be expressed in terms of the Moore-Penrose inverse $L^\dagger$ of the unnormalized graph Laplacian (e.g., Klein and Randic, 1993; Xiao and Gutman, 2003; Fouss et al., 2006):

$$C_{ij} = \mathrm{vol}(G) \left\langle e_i - e_j, L^\dagger (e_i - e_j) \right\rangle,$$

where e_i is the i-th unit vector in $\mathbb{R}^n$. The following representations for commute and hitting times involving the pseudo-inverse $L_{\mathrm{sym}}^\dagger$ of the normalized graph Laplacian are direct consequences of Lovász, 1993:

Proposition 1 (Closed form expression for hitting and commute times) *Let G be a connected, undirected graph with n vertices. The hitting times H_{ij}, $i \neq j$, can be computed by*

$$H_{ij} = \text{vol}(G)\Big\langle \frac{1}{\sqrt{d_j}}e_j \ , \ L_{sym}^{\dagger}\Big(\frac{1}{\sqrt{d_j}}e_j - \frac{1}{\sqrt{d_i}}e_i\Big)\Big\rangle,$$

and the commute times satisfy

$$C_{ij} = \text{vol}(G)\Big\langle \frac{1}{\sqrt{d_i}}e_i - \frac{1}{\sqrt{d_j}}e_j \ , \ L_{sym}^{\dagger}\Big(\frac{1}{\sqrt{d_j}}e_j - \frac{1}{\sqrt{d_i}}e_i\Big)\Big\rangle.$$

Our main focus in this paper is the class of *geometric graphs*. For a *deterministic* geometric graph we consider a fixed set of points $X_1, \ldots, X_n \in \mathbb{R}^d$. These points form the vertices $v_1, \ldots, v_n$ of the graph. In the *ε-graph* we connect two points whenever their Euclidean distance is less than or equal to ε. In the undirected *symmetric k-nearest neighbor graph* we connect v_i to v_j if X_i is among the k nearest neighbors of X_j *or* vice versa. In the undirected *mutual k-nearest neighbor graph* we connect v_i to v_j if X_i is among the k nearest neighbors of X_j *and* vice versa. Note that by default, the terms ε- and kNN-graph refer to unweighted graphs in our paper. When we treat weighted graphs, we always make it explicit. For a general *similarity graph* we build a weight matrix between all points based on a similarity function $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, that is we define the weight matrix W with entries $w_{ij} = k(X_i, X_j)$ and consider the fully connected graph with weight matrix W. The most popular weight function in applications is the Gaussian similarity function $w_{ij} = \exp(-\|X_i - X_j\|^2/\sigma^2)$, where $\sigma > 0$ is a bandwidth parameter.

While these definitions make sense with any fixed set of vertices, we are most interested in the case of *random* geometric graphs. We assume that the underlying set of vertices $X_1, ..., X_n$ has been drawn i.i.d. according to some probability density p on $\mathbb{R}^d$. Once the vertices are known, the edges in the graphs are constructed as described above. In the random setting it is convenient to make regularity assumptions in order to be able to control quantities such as the minimal and maximal degrees.

Definition 2 (Valid region) *Let p be any density on $\mathbb{R}^d$. We call a connected subset $\mathcal{X} \subset \mathbb{R}^d$ a* valid region *if the following properties are satisfied:*

1. *The density on $\mathcal{X}$ is bounded away from 0 and infinity, that is for all $x \in \mathcal{X}$ we have that $0 < p_{\min} \leq p(x) \leq p_{\max} < \infty$ for some constants $p_{\min}, p_{\max}$.*

2. *$\mathcal{X}$ has "bottleneck" larger than some value $h > 0$: the set $\{x \in \mathcal{X} : dist(x, \partial\mathcal{X}) > h/2\}$ is connected (here $\partial\mathcal{X}$ denotes the topological boundary of $\mathcal{X}$).*

3. *The boundary of $\mathcal{X}$ is regular in the following sense. We assume that there exist positive constants $\alpha > 0$ and $\varepsilon_0 > 0$ such that if $\varepsilon < \varepsilon_0$, then for all points $x \in \partial\mathcal{X}$ we have $\text{vol}(B_\varepsilon(x) \cap \mathcal{X}) \geq \alpha\,\text{vol}(B_\varepsilon(x))$ (where vol denotes the Lebesgue volume). Essentially this condition just excludes the situation where the boundary has arbitrarily thin spikes.*

Sometimes we consider a valid region with respect to two points s, t. Here we additionally assume that s and t are interior points of $\mathcal{X}$.

In the spectral part of our paper, we always have to make a couple of assumptions that will be summarized by the term **general assumptions**. They are as follows: First we assume that $\mathcal{X} := \text{supp}(p)$ is a valid region according to Definition 2. Second, we assume that $\mathcal{X}$ does not contain any holes and does not become arbitrarily narrow: there exists a homeomorphism $h : \mathcal{X} \to [0,1]^d$ and constants $0 < L_{\min} < L_{\max} < \infty$ such that for all $x, y \in \mathcal{X}$ we have

$$L_{\min}\|x - y\| \ \leq \ \|h(x) - h(y)\| \ \leq \ L_{\max}\|x - y\|.$$

This condition restricts $\mathcal{X}$ to be topologically equivalent to the cube. In applications this is not a strong assumption, as the occurrence of "holes" with vanishing probability density is unrealistic due to the presence of noise in the data generating process. More generally we believe that our results can be generalized to other homeomorphism classes, but refrain from doing so as it would substantially increase the amount of technicalities.

In the following we denote the volume of the unit ball in $\mathbb{R}^d$ by η_d. For readability reasons, we are going to state our main results using constants $c_i > 0$. These constants are independent of n and the graph connectivity parameter (ε or k or h, respectively) but depend on the dimension, the geometry of $\mathcal{X}$, and p. The values of all constants are determined explicitly in the proofs. They are not the same in different propositions.

4. Main Results

Our paper comprises two different approaches. In the first approach we analyze the resistance distance by flow based arguments. This technique is somewhat restrictive in the sense that it only works for the resistance distance itself (not the hitting times) and we only apply it to random geometric graphs. The advantage is that in this setting we obtain good convergence conditions and rates. The second approach is based on spectral arguments and is more general. It works for various kinds of graphs and can treat hitting times as well. This comes at the price of slightly stronger assumptions and worse convergence rates.

4.1 Results Based on Flow Arguments

Theorem 3 (Commute distance on ε-graphs) *Let $\mathcal{X}$ be a valid region with bottleneck h and minimal density $p_{\min}$. For $\varepsilon \leq h$, consider an unweighted ε-graph built from the sequence $X_1, \ldots, X_n$ that has been drawn i.i.d. from the density p. Fix i and j. Assume that X_i and X_j have distance at least h from the boundary of $\mathcal{X}$, and that the distance between X_i and X_j is at least 8ε. Then there exist constants $c_1, \ldots, c_7 > 0$ (depending on the dimension and geometry of $\mathcal{X}$) such that with probability at least $1 - c_1 n \exp(-c_2 n \varepsilon^d) -$*

$c_3 \exp(-c_4 n\varepsilon^d)/\varepsilon^d$ *the commute distance on the ε-graph satisfies*

$$\left| \frac{n\varepsilon^d}{\text{vol}(G)} C_{ij} - \left(\frac{n\varepsilon^d}{d_i} + \frac{n\varepsilon^d}{d_j} \right) \right| \leq \begin{cases} c_5/n\varepsilon^d & \text{if } d > 3 \\ c_6 \cdot \log(1/\varepsilon)/n\varepsilon^3 & \text{if } d = 3 \\ c_7/n\varepsilon^3 & \text{if } d = 2 \end{cases}$$

The probability converges to 1 if $n \to \infty$ and $n\varepsilon^d/\log(n) \to \infty$. The right hand side of the deviation bound converges to 0 as $n \to \infty$, if

$$\begin{cases} n\varepsilon^d \to \infty & \text{if } d > 3 \\ n\varepsilon^3/\log(1/\varepsilon) \to \infty & \text{if } d = 3 \\ n\varepsilon^3 = n\varepsilon^{d+1} \to \infty & \text{if } d = 2. \end{cases}$$

Under these conditions, if the density p is continuous and if $\varepsilon \to 0$, then

$$\frac{n\varepsilon^d}{\text{vol}(G)} C_{ij} \longrightarrow \frac{1}{\eta_d p(X_i)} + \frac{1}{\eta_d p(X_j)} \qquad a.s.$$

Theorem 4 (Commute distance on kNN-graphs) *Let $\mathcal{X}$ be a valid region with bottleneck h and density bounds $p_{\min}$ and $p_{\max}$. Consider an unweighted kNN-graph (either symmetric or mutual) such that $(k/n)^{1/d}/(2p_{\max}) \leq h$, built from the sequence $X_1, \ldots, X_n$ that has been drawn i.i.d. from the density p.*

Fix i and j. Assume that X_i and X_j have distance at least h from the boundary of $\mathcal{X}$, and that the distance between X_i and X_j is at least $4(k/n)^{1/d}/p_{\max}$. Then there exist constants $c_1, \ldots, c_5 > 0$ such that with probability at least $1 - c_1 n \exp(-c_2 k)$ the commute distance on both the symmetric and the mutual kNN-graph satisfies

$$\left| \frac{k}{\text{vol}(G)} C_{ij} - \left(\frac{k}{d_i} + \frac{k}{d_j} \right) \right| \leq \begin{cases} c_4/k & \text{if } d > 3 \\ c_5 \cdot \log(n/k)/k & \text{if } d = 3 \\ c_6 n^{1/2}/k^{3/2} & \text{if } d = 2 \end{cases}$$

The probability converges to 1 if $n \to \infty$ and $k/\log(n) \to \infty$. In case $d > 3$, the right hand side of the deviation bound converges to 0 if $k \to \infty$ (and under slightly worse conditions in cases $d = 3$ and $d = 2$). Under these conditions, if the density p is continuous and if additionally $k/n \to 0$, then $\frac{k}{\text{vol}(G)} C_{ij} \to 2$ almost surely.

Let us make a couple of technical remarks about these theorems.

To achieve the convergence of the commute distance we have to rescale it appropriately (for example, in the ε-graph we scale by a factor of $n\varepsilon^d$). Our rescaling is exactly chosen such that the limit expressions are finite, positive values. Scaling by any other factor in terms of n, ε or k either leads to divergence or to convergence to zero.

In case $d > 3$, all convergence conditions on n and ε (or k, respectively) are the ones to be expected for random geometric graphs. They are satisfied as soon as the degrees grow faster than $\log(n)$. For degrees of order smaller than $\log(n)$, the graphs are not connected

anyway, see for example Penrose (1997, 1999); Xue and Kumar (2004); Balister et al. (2005). In dimensions 3 and 2, our rates are a bit weaker. For example, in dimension 2 we need $n\varepsilon^3 \to \infty$ instead of $n\varepsilon^2 \to \infty$. On the one hand we are not too surprised to get systematic differences between the lowest few dimensions. The same happens in many situations, just consider the example of Polya's theorem about the recurrence/ transience of random walks on grids. On the other hand, these differences might as well be an artifact of our proof methods (and we suspect so at least for the case $d = 3$; but even though we tried, we did not get rid of the log factor in this case). It is a matter of future work to clarify this.

The valid region $\mathcal{X}$ has been introduced for technical reasons. We need to operate in such a region in order to be able to control the behavior of the graph, e.g. the minimal and maximal degrees. The assumptions on $\mathcal{X}$ are the standard assumptions used regularly in the random geometric graph literature. In our setting, we have the freedom of choosing $\mathcal{X} \subset \mathbb{R}^d$ as we want. In order to obtain the tightest bounds one should aim for a valid $\mathcal{X}$ that has a wide bottleneck h and a high minimal density $p_{\min}$. In general this freedom of choosing $\mathcal{X}$ shows that if two points are in the same high-density region of the space, the convergence of the commute distance is fast, while it gets slower if the two points are in different regions of high density separated by a bottleneck.

We stated the theorems above for a fixed pair i, j. However, they also hold uniformly over all pairs i, j that satisfy the conditions in the theorem (with exactly the same statement). The reason is that the main probabilistic quantities that enter the proofs are bound on the minimal and maximal degrees, which of course hold uniformly.

4.2 Results Based on Spectral Arguments

The representation of the hitting and commute times in terms of the Moore-Penrose inverse of the normalized graph Laplacian (Proposition 1) can be used to derive the following key proposition that is the basis for all further results in this section.

Proposition 5 (Absolute and relative bounds in any fixed graph) *Let G be a finite, connected, undirected, possibly weighted graph that is not bipartite.*

1. For $i \neq j$

$$\left| \frac{1}{\mathrm{vol}(G)} H_{ij} - \frac{1}{d_j} \right| \leq 2 \left(\frac{1}{1 - \lambda_2} + 1 \right) \frac{w_{\max}}{d_{\min}^2}.$$

2. For $i \neq j$

$$\left| \frac{1}{\mathrm{vol}(G)} C_{ij} - \left(\frac{1}{d_i} + \frac{1}{d_j} \right) \right| \leq \frac{w_{\max}}{d_{\min}} \left(\frac{1}{1 - \lambda_2} + 2 \right) \left(\frac{1}{d_i} + \frac{1}{d_j} \right)$$

$$\leq 2 \left(\frac{1}{1 - \lambda_2} + 2 \right) \frac{w_{\max}}{d_{\min}^2}. \tag{1}$$

We would like to point out that even though the bound in Part 2 of the proposition is reminiscent to statements in the literature, it is much tighter. Consider the following formula from Lovász (1993)

$$\frac{1}{2}\left(\frac{1}{d_i}+\frac{1}{d_j}\right) \ \leq \ \frac{1}{\text{vol}(G)}C_{ij} \ \leq \ \frac{1}{1-\lambda_2}\left(\frac{1}{d_i}+\frac{1}{d_j}\right)$$

that can easily be rearranged to the following bound:

$$\left|\frac{1}{\text{vol}(G)}C_{ij}-\left(\frac{1}{d_i}+\frac{1}{d_j}\right)\right| \leq \frac{1}{1-\lambda_2}\frac{2}{d_{\min}}. \tag{2}$$

The major difference between our bound (1) and Lovasz' bound (2) is that while the latter has the term $d_{\min}$ in the denominator, our bound has the term $d_{\min}^2$ in the denominator. This makes all of a difference: in the graphs under considerations our bound converges to 0 whereas Lovasz' bound diverges. In particular, our convergence results are not a trivial consequence of Lovász (1993).

4.2.1 Application to Unweighted Random Geometric Graphs

In the following we are going to apply Proposition 5 to various random geometric graphs. Next to some standard results about the degrees and number of edges in random geometric graphs, the main ingredients are the following bounds on the spectral gap in random geometric graphs. These bounds are of independent interest because the spectral gap governs many important properties and processes on graphs.

Theorem 6 (Spectral gap of the ε-graph) *Suppose that the general assumptions hold. Then there exist constants $c_1, \ldots, c_6 > 0$ such that with probability at least $1-c_1 n \exp(-c_2 n\varepsilon^d)-c_3 \exp(-c_4 n\varepsilon^d)/\varepsilon^d$,*

$$1-\lambda_2 \ \geq \ c_5 \cdot \varepsilon^2 \qquad\qquad and \qquad\qquad 1-|\lambda_n| \ \geq \ c_6 \cdot \varepsilon^{d+1}/n.$$

If $n\varepsilon^d/\log n \to \infty$, then this probability converges to 1.

Theorem 7 (Spectral gap of the kNN-graph) *Suppose that the general assumptions hold. Then for both the symmetric and the mutual kNN-graph there exist constants $c_1, \ldots, c_4 > 0$ such that with probability at least $1 - c_1 n \exp(-c_2 k)$,*

$$1-\lambda_2 \ \geq \ c_3 \cdot (k/n)^{2/d} \qquad\qquad and \qquad\qquad 1-|\lambda_n| \ \geq \ c_4 \cdot k^{2/d}/n^{(d+2)/d}.$$

If $k/\log n \to \infty$, then the probability converges to 1.

The following theorems characterize the hitting and commute times for ε-and kNN-graphs. They are direct consequences of plugging the results about the spectral gap into Proposition 5. In the corollaries, the reader should keep in mind that the degrees also depend on n and ε (or k, respectively).

Corollary 8 (Hitting and commute times on ε-graphs) *Assume that the general assumptions hold. Consider an unweighted ε-graph built from the sequence $X_1, \ldots, X_n$ drawn i.i.d. from the density p. Then there exist constants $c_1, \ldots, c_5 > 0$ such that with probability at least $1 - c_1 n \exp(-c_2 n \varepsilon^d) - c_3 \exp(-c_4 n \varepsilon^d)/\varepsilon^d$, we have uniformly for all $i \neq j$ that*

$$\left| \frac{n \varepsilon^d}{\mathrm{vol}(G)} H_{ij} - \frac{n \varepsilon^d}{d_j} \right| \leq \frac{c_5}{n \varepsilon^{d+2}}.$$

If the density p is continuous and $n \to \infty, \varepsilon \to 0$ and $n \varepsilon^{d+2} \to \infty$, then

$$\frac{n \varepsilon^d}{\mathrm{vol}(G)} H_{ij} \longrightarrow \frac{1}{\eta_d \cdot p(X_j)} \quad \textit{almost surely.}$$

For the commute times, the analogous results hold due to $C_{ij} = H_{ij} + H_{ji}$.

Corollary 9 (Hitting and commute times on kNN-graphs) *Assume that the general assumptions hold. Consider an unweighted kNN-graph built from the sequence $X_1, \ldots, X_n$ drawn i.i.d. from the density p. Then for both the symmetric and mutual kNN-graph there exist constants $c_1, c_2, c_3 > 0$ such that with probability at least $1 - c_1 \cdot n \cdot \exp(-kc_2)$, we have uniformly for all $i \neq j$ that*

$$\left| \frac{k}{\mathrm{vol}(G)} H_{ij} - \frac{k}{d_j} \right| \leq c_3 \cdot \frac{n^{2/d}}{k^{1+2/d}}.$$

If the density p is continuous and $n \to \infty, k/n \to 0$ and $k(k/n)^{2/d} \to \infty$, then

$$\frac{k}{\mathrm{vol}(G)} H_{ij} \longrightarrow 1 \quad \textit{almost surely.}$$

For the commute times, the analogous results hold due to $C_{ij} = H_{ij} + H_{ji}$.

Note that the density shows up in the limit for the ε-graph, but not for the kNN graph. The explanation is that in the former, the density is encoded in the degrees of the graph, while in the latter it is only encoded in the k-nearest neighbor distance, but not the degrees themselves. As a rule of thumb, it is possible to convert the last two corollaries into each other by substituting ε by $(k/(np(x)\eta^d))^{1/d}$ or vice versa.

4.2.2 Application to Weighted Graphs

In several applications, ε-graphs or kNN graphs are endowed with edge weights. For example, in the field of machine learning it is common to use Gaussian weights $w_{ij} = \exp(-\|X_i - X_j\|^2/\sigma^2)$, where $\sigma > 0$ is a bandwidth parameter. We can use standard spectral results to prove approximation theorems in such cases.

Theorem 10 (Results on fully connected weighted graphs) *Consider a fixed, fully connected weighted graph with weight matrix W. Assume that its entries are upper and*

lower bounded by some constants $w_{\min}, w_{\max}$, that is $0 < w_{\min} \le w_{ij} \le w_{\max}$ for all i, j. Then, uniformly for all $i, j \in \{1, ..., n\}$, $i \ne j$,

$$\left| \frac{n}{\mathrm{vol}(G)} H_{ij} - \frac{n}{d_j} \right| \le 4n \left(\frac{w_{\max}}{w_{\min}} \right) \frac{w_{\max}}{d_{\min}^2} \le 4 \frac{w_{\max}^2}{w_{\min}^3} \frac{1}{n}.$$

For example, this result can be applied directly to a Gaussian similarity graph (for fixed bandwidth σ).

The next theorem treats the case of Gaussian similarity graphs with adapted bandwidth σ. The technique we use to prove this theorem is rather general. Using the Rayleigh principle, we reduce the case of the fully connected Gaussian graph to a truncated graph where edges beyond a certain length are removed. Bounds for this truncated graph, in turn, can be reduced to bounds of the unweighted ε-graph. With this technique it is possible to treat very general classes of graphs.

Theorem 11 (Results on Gaussian graphs with adapted bandwidth) *Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a compact set and p a continuous, strictly positive density on $\mathcal{X}$. Consider a fully connected, weighted similarity graph built from the points $X_1, \ldots, X_n$ drawn i.i.d. from density p. As weight function use the Gaussian similarity function $k_\sigma(x, y) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} \exp\left(-\frac{\|x-y\|^2}{2\sigma^2} \right)$. If the density p is continuous and $n \to \infty, \sigma \to 0$ and $n\sigma^{d+2}/\log(n) \to \infty$, then*

$$\frac{n}{\mathrm{vol}(G)} C_{ij} \longrightarrow \frac{1}{p(X_i)} + \frac{1}{p(X_j)} \quad \text{almost surely.}$$

Note that in this theorem, we introduced the scaling factor $1/\sigma^d$ already in the definition of the Gaussian similarity function to obtain the correct density estimate $p(X_j)$ in the limit. For this reason, the resistance results are rescaled with factor n instead of $n\sigma^d$.

4.2.3 Application to Random Graphs with Given Expected Degrees and Erdős-Rényi Graphs

Consider the general random graph model where the edge between vertices i and j is chosen independently with a certain probability p_{ij} that is allowed to depend on i and j. This model contains popular random graph models such as the Erdős-Rényi random graph, planted partition graphs, and random graphs with given expected degrees. For this class of random graphs, bounds on the spectral gap have been proved by Chung and Radcliffe (2011). These bounds can directly be applied to derive bounds on the resistance distances. It is not surprising to see that hitting times are meaningless, because these graphs are expander graphs and the random walk mixes fast. The model of Erdős-Rényi graphs with planted partitions is more interesting because it gives insight to the question how strongly clustered the graph can be before our results break down.

Theorem 12 (Chung and Radcliffe, 2011) *Let G be a random graph where edges between vertices i and j are put independently with probabilities p_{ij}. Consider the normalized Laplacian L_{sym}, and define the expected normalized Laplacian as the matrix $\overline{L_{sym}} :=$*

$I - \overline{D^{-1/2}\overline{A}D^{-1/2}}$ where $\overline{A}_{ij} = E(A_{ij}) = p_{ij}$ and $\overline{D} = E(D)$. Let $\overline{d}_{\min}$ be the minimal expected degree. Denote the eigenvalues of L_{sym} by μ, the ones of $\overline{L_{sym}}$ by $\overline{\mu}$. Choose $\varepsilon > 0$. Then there exists a constant $k = k(\varepsilon)$ such that if $\overline{d}_{\min} > k \log(n)$, then with probability at least $1 - \varepsilon$,

$$\forall j = 1, ..., n : \quad |\mu_j - \overline{\mu_j}| \leq 2\sqrt{\frac{3 \log(4n/\varepsilon)}{\overline{d}_{\min}}}.$$

Random graphs with given expected degrees. For a graph of n vertices we have n parameters $\overline{d}_1, ..., \overline{d}_n > 0$. For each pair of vertices v_i and v_j, we independently place an edge between these two vertices with probability $\overline{d}_i \overline{d}_j / \sum_{k=1}^{n} \overline{d}_k$. It is easy to see that in this model, vertex v_i has expected degree $\overline{d}_i$ (cf. Section 5.3. in Chung and Lu, 2006 for background reading).

Corollary 13 (Random graphs with given expected degrees) *Consider any sequence of random graphs with expected degrees such that $\overline{d}_{\min} = \omega(\log n)$. Then the commute distances satisfy for all $i \neq j$,*

$$\left|\frac{1}{\text{vol}(G)}C_{ij} - \left(\frac{1}{d_i} + \frac{1}{d_j}\right)\right| \Bigg/ \left(\frac{1}{d_i} + \frac{1}{d_j}\right) = O\left(\frac{1}{\log(2n)}\right) \longrightarrow 0, \text{ almost surely.}$$

Planted partition graphs. Assume that the n vertices are split into two "clusters" of equal size. We put an edge between two vertices u and v with probability p_{within} if they are in the same cluster and with probability $p_{between} < p_{within}$ if they are in different clusters. For simplicity we allow self-loops.

Corollary 14 (Random graph with planted partitions) *Consider an Erdős-Rényi graph with planted bisection. Assume that $p_{within} = \omega(\log(n)/n)$ and $p_{between}$ such that $np_{between} \to \infty$ (arbitrarily slow). Then, for all vertices u, v in the graph*

$$\left|\frac{1}{n} \cdot H_{ij} - 1\right| = O\left(\frac{1}{n \, p_{between}}\right) \to 0 \quad \text{in probability.}$$

This result is a nice example to show that even though there is a strong cluster structure in the graph, hitting times and commute distances cannot see this cluster structure any more, once the graph gets too large. Note that the corollary even holds if $np_{between}$ grows much slower than np_{within}. That is, the larger our graph, the more pronounced is the cluster structure. Nevertheless, the commute distance converges to a trivial result. On the other hand, we also see that the speed of convergence is $O(np_{between})$, that is, if $p_{between} = g(n)/n$ with a slowly growing function g, then convergence can be slow. We might need very large graphs before the degeneracy of the commute time will be visible.

5. Experiments

In this section we examine the convergence behavior of the commute distance in practice. We ran a large number of simulations, both on artificial and real world data sets, in order to evaluate whether the rescaled commute distance in a graph is close to its predicted limit expression $1/d_u + 1/d_v$ or not. We conducted simulations for artificial graphs (random geometric graphs, planted partition graphs, preferential attachment graphs) and real world data sets of various types and sizes (social networks, biological networks, traffic networks; up to 1.6 million vertices and 22 million edges). The general setup is as follows. Given a graph, we compute the pairwise resistance distance R_{ij} between all points. The relative deviation between resistance distance and predicted result is then given by

$$RelDev(i,j) := \frac{|R_{ij} - 1/d_i - 1/d_j|}{R_{ij}}.$$

Note that we report relative rather than absolute deviations because this is more meaningful if R_{ij} is strongly fluctuating on the graph. It also allows to compare the behavior of different graphs with each other. In all figures, we then report the maximum, mean and median relative deviations. In small and moderate sized graphs, these operations are taken over all pairs of points. In some of the larger graphs, computing all pairwise distances is prohibitive. In these cases, we compute mean, median and maximum based on a random subsample of vertices (see below for details).

The bottom line of our experiments is that in nearly all graphs, the deviations are small. In particular, this also holds for many moderate sized graphs, even though our results are statements about $n \to \infty$. This shows that the limit results for the commute distance are indeed relevant for practice.

5.1 Random Geometric Graphs

We start with the class of random geometric graphs, which is very important for machine learning. We use a mixture of two Gaussian distributions on $\mathbb{R}^d$. The first two dimensions contain a two-dimensional mixture of two Gaussians with varying separation (centers $(-sep/2, 0)$ and $(+sep/2, 0)$, covariance matrix $0.2 \cdot Id$, mixing weights 0.5 for both Gaussians). The remaining $d - 2$ dimensions contain Gaussian noise with variance 0.2 as well. From this distribution we draw n sample points. Based on this sample, we either compute the unweighted symmetric kNN graph, the unweighted ε-graphs or the Gaussian similarity graph. In order to be able to compare the results between these three types of graphs we match the parameters of the different graphs: given some value k for the kNN-graph we choose the values of ε for the ε-graph and σ for the Gaussian graph as the maximal k-nearest neighbor distance in the data set.

Figure 1 shows the results of the simulations. We can see that the deviations decrease fast with the sample size. In particular, already for small sample sizes reported, the maximal deviations get very small. The more clustered the data is (separation is larger), the larger the deviations get. This is the case as the deviation bound scales inversely with the spectral gap, which gets larger the more clustered the data is. The deviations also decrease with
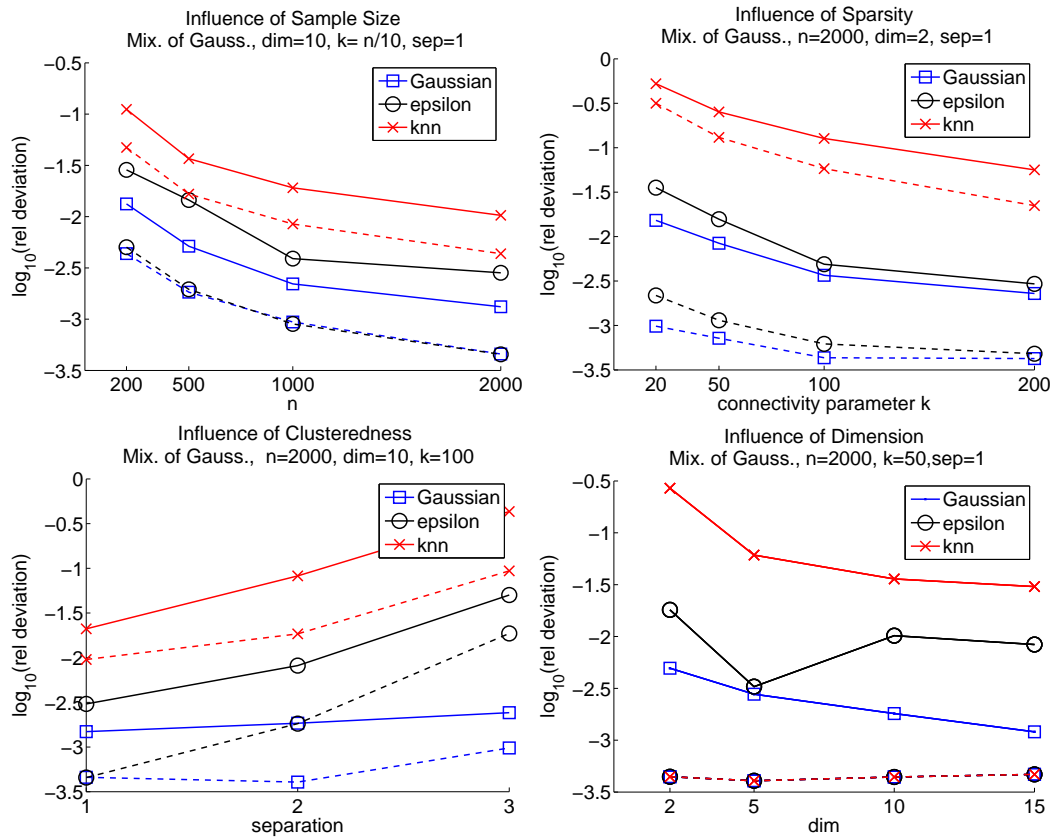
Figure 1: Deviations in random geometric graphs. Solid lines show the maximum relative deviation, dashed lines the mean relative deviation. See text for more details.

increasing dimension, as predicted by our bounds. The intuitive explanation is that in higher dimensions, geometric graphs mix faster as there exist more "shortcuts" between the two sides of the point cloud. For a similar reason, the deviation bounds also decrease with increasing connectivity in the graph. All in all, the deviations are very small even though our sample sizes in these experiments are modest.

5.2 Planted Partition Graphs

Next we consider graphs according to the planted partition model. We modeled a graph with n vertices and two equal sized clusters with connectivity parameters p_{within} and $p_{between}$. As the results in Figure 2 show, the deviation decreases rapidly when n increases, and it decreases when the cluster structure becomes less pronounced.
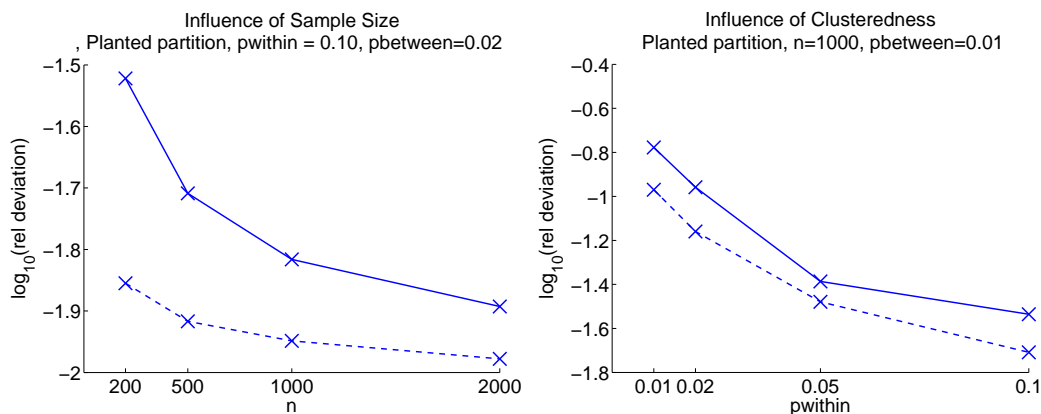
Figure 2: Deviations in planted partition graphs. Solid lines show the maximum relative deviation, dashed lines the mean relative deviation. See text for more details.

5.3 Preferential Attachment Graphs

An important class of random graphs is the preferential attachment model (Barabási and Albert, 1999), because it can be used to model graphs with power law behavior. Our current convergence proofs cannot be carried over to preferential attachment graphs: the minimal degree in preferential attachment graphs is constant, so our proofs break down. However, our simulation results show that approximating commute distances by the limit expression $1/d_u + 1/d_v$ gives accurate approximations as well. This finding indicates that our convergence results seem to hold even more generally than our theoretical findings suggest.

In our simulation we generated preferential attachment graphs according to the following standard procedure: Starting with a graph that consists of two vertices connected by an edge, in each time step we add a new vertex to the graph. The new vertex is connected by a fixed number of edges to existing vertices (this number is called NumLinks in the figures below). The target vertices of these edges are chosen randomly among all existing vertices, where the probability to connect to a particular vertex is proportional to its degree. All edges in the graph are undirected. As an example, we show the adjacency matrix and the degree histogram for such a graph in Figure 3. The next plots in this figure shows the relative deviations in such a preferential attachment graph, plotted against the sparsity of the graph (number of outlinks). Overall we can see that the deviations are very small, they are on the same scale as the ones in all the other simulations above. The mean deviations simply decrease as the connectivity of the graph increases. The maximum deviations show an effect that is different from what we have seen in the other graphs: while it decreases in the sparse regime, it starts to increase again when the graph becomes denser. However, investigating this effect more closely reveals that it is just generated by the three vertices in the graph which have the largest degrees. If we exclude these three vertices when computing the maximum deviation, then the maximum deviation decreases as gracefully as the mean deviation. All in all we can say that in the sparse regime, the commute distance can be well
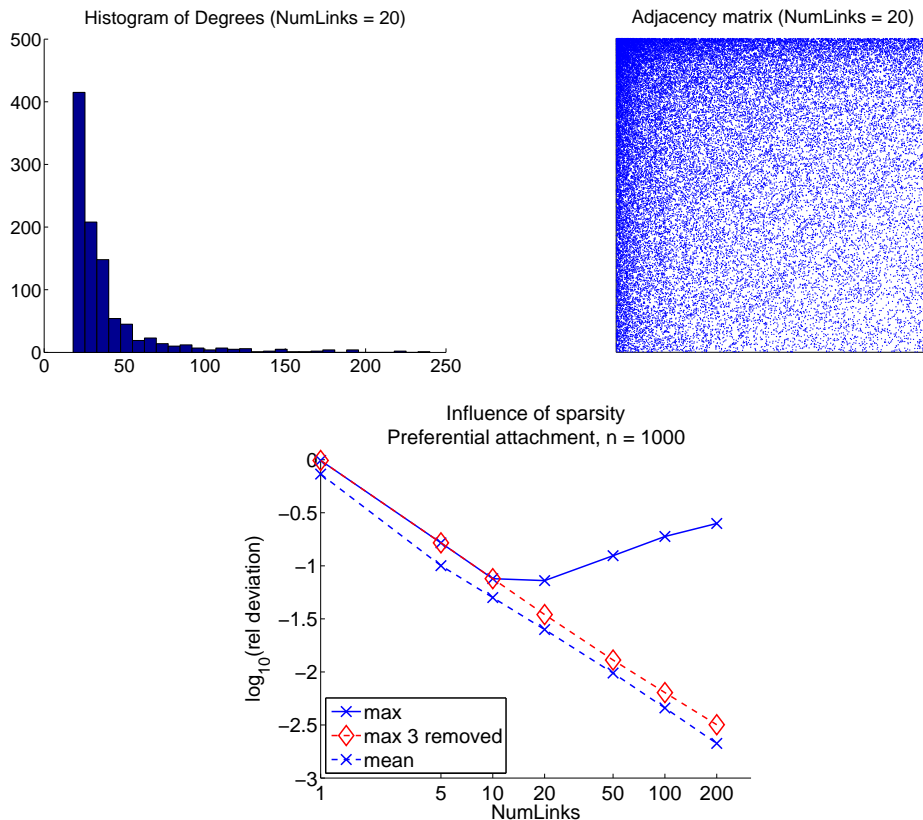
Figure 3: Relative deviations for a preferential attachment graph. First two figures: illustration of such a graph in terms of histogram of degrees and heat plot of the adjacency matrix. Third figure: Relative deviations. As before the solid line shows the maximum relative deviation and the dashed line the mean relative deviation. The line "max 3 removed" refers to the maximum deviation when we exclude the three highest-degree vertices from computing the maximum. See text for details.

approximated by the expression $1/d_u + 1/d_v$. In the dense regime, the same is true unless u or v is among the very top degree vertices.

5.4 Real World Data

Now we consider the deviations for a couple of real world data sets. We start with similarity graphs on real data, as they are often used in machine learning. As example we use the full USPS data set of handwritten digits (9298 points in 256 dimensions) and consider the different forms of similarity graphs (kNN, ε, Gaussian) with varying connectivity parameter. In Figure 4 we can see that overall, the relative errors are pretty small.
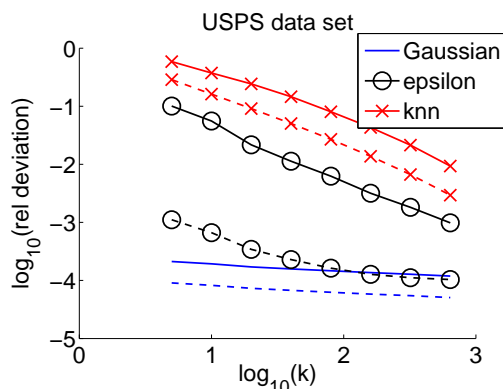
Figure 4: Relative deviations on different similarity graphs on the USPS data. As before the solid line shows the maximum relative deviation and the dashed line the mean relative deviation. See text for details.

Name	Type of Network	Number Vertices	Number Edges	mean rel dev	median rel dev	max rel dev
mousegene[2]	gene network	42923	28921954	0.12	0.01	0.70
jazz[1]	social network	198	5484	0.09	0.06	0.51
soc-Slashdot0902[2]	social network	82168	1008460	0.10	0.06	0.33
cit-HepPh[2]	citation graph	34401	841568	0.09	0.07	0.25
soc-sign-epinions[2]	social network	119130	1408534	0.14	0.09	0.53
pdb1HYS[2]	protein databank	36417	4308348	0.12	0.10	0.28
as-Skitter[2]	internet topology	1694616	22188418	0.16	0.12	0.54
celegans[1]	neural network	453	4065	0.18	0.12	0.77
email[1]	email traffic	1133	10903	0.15	0.13	0.77
Amazon0505[2]	co-purchase	410236	4878874	0.28	0.24	0.80
CoPapersCiteseer[2]	coauthor graph	434102	32073440	0.33	0.30	0.76
PGP[1]	user network	10680	48632	0.54	0.53	0.97

Figure 5: Relative Error of the approximation in real world networks. Downloaded from: 1, `http://deim.urv.cat/~aarenas/data/welcome.htm` and 2, `http://www.cise.ufl.edu/research/sparse/matrices/`. See text for details.

Furthermore, we consider a number of network data sets that are available online, see Figure 5 for a complete list. Some of these networks are directed, but we use their undirected versions. In cases the graphs were not connected, we ran the analysis on the largest connected component. On the smaller data sets, we computed all pairwise commute distances and used all these values to compute mean, median and maximum relative deviations. On the larger data sets we just draw 20 vertices at random, then compute all pairwise commute distances between these 20 vertices, and finally evaluate mean, median and maximum

based on this subset. Figure 5 shows the mean, median and maximum relative commute distances in all these networks. We can see that in many of these data sets, the deviations are reasonable small. Even though they are not as small as in the artificial data sets, they are small enough to acknowledge that our approximation results tend to hold in real world graphs.

6. Proofs for the Flow-Based Approach

For notational convenience, in this section we work with the resistance distance $R_{uv} = C_{uv}/\operatorname{vol}(G)$ instead of the commute distance C_{uv}, then we do not have to carry the factor $1/\operatorname{vol}(G)$ everywhere.

6.1 Lower Bound

It is easy to prove that the resistance distance between two points is lower bounded by the sum of the inverse degrees.

Proposition 15 (Lower bound) *Let G be a weighted, undirected, connected graph and consider two vertices s and t, $s \neq t$. Assume that G remains connected if we remove s and t. Then the effective resistance between s and t is bounded by*

$$R_{st} \geq \frac{Q_{st}}{1 + w_{st}Q_{st}},$$

where $Q_{st} = 1/(d_s - w_{st}) + 1/(d_t - w_{st})$. Note that if s and t are not connected by a direct edge (that is, $w_{st} = 0$), then the right hand side simplifies to $1/d_s + 1/d_t$.

Proof. The proof is based on Rayleigh's monotonicity principle that states that increasing edge weights (conductances) in the graph can never increase the effective resistance between two vertices (cf. Corollary 7 in Section IX.2 of Bollobas, 1998). Given our original graph G, we build a new graph G' by setting the weight of all edges to infinity, except the edges that are adjacent to s or t (setting the weight of an edge to infinity means that this edge has infinite conductance and no resistance any more). This can also be interpreted as taking all vertices except s and t and merging them to one super-node a. Now our graph G' consists of three vertices s, a, t with several parallel edges from s to a, several parallel edges from a to t, and potentially the original edge between s and t (if it existed in G). Exploiting the laws in electrical networks (resistances add along edges in series, conductances add along edges in parallel; see Section 2.3 in Lyons and Peres (2010) for detailed instructions and examples) leads to the desired result. ∎

6.2 Upper Bound

This is the part that requires the hard work. Our proof is based on a theorem that shows how the resistance between two points in the graph can be computed in terms of flows on the graph. The following result is taken from Corollary 6 in Section IX.2 of Bollobas (1998).

Theorem 16 (Resistance in terms of flows, cf. Bollobas, 1998) *Let $G = (V, E)$ be a weighted graph with edge weights w_e ($e \in E$). The effective resistance R_{st} between two fixed vertices s and t can be expressed as*

$$R_{st} = \inf \left\{ \sum_{e \in E} \frac{u_e^2}{w_e} \;\middle|\; u = (u_e)_{e \in E} \text{ unit flow from } s \text{ to } t \right\}.$$

Note that evaluating the formula in the above theorem for any fixed flow leads to an upper bound on the effective resistance. The key to obtaining a tight bound is to distribute the flow as widely and uniformly over the graph as possible.

For the case of geometric graphs we are going to use a grid on the underlying space to construct an efficient flow between two vertices. Let $X_1, ..., X_n$ be a fixed set of points in $\mathbb{R}^d$ and consider a geometric graph G with vertices $X_1, ..., X_n$. Fix any two of them, say $s := X_1$ and $t := X_2$. Let $\mathcal{X} \subset \mathbb{R}^d$ be a connected set that contains both s and t. Consider a regular grid with grid width g on $\mathcal{X}$. We say that grid cells are neighbors of each other if they touch each other in at least one edge.

Definition 17 (Valid grid) *We call the grid* valid *if the following properties are satisfied:*

1. *The grid width is not too small: Each cell of the grid contains at least one of the points $X_1, ..., X_n$.*

2. *The grid width g is not too large: Points in the same or neighboring cells of the grid are always connected in the graph G.*

3. *Relation between grid width and geometry of $\mathcal{X}$: Define the bottleneck h of the region $\mathcal{X}$ as the largest u such that the set $\{x \in \mathcal{X} \mid dist(x, \partial \mathcal{X}) > u/2\}$ is connected.*

   We require that $\sqrt{d}\, g \leq h$ (a cube of side length g should fit in the bottleneck).

Under the assumption that a valid grid exists, we can prove the following general proposition that gives an upper bound on the resistance distance between vertices in a fixed geometric graph. Note that proving the existence of the valid grid will be an important part in the proofs of Theorems 3 and 4.

Proposition 18 (Resistance on a fixed geometric graph) *Consider a fixed set of points $X_1, ..., X_n$ in some connected region $\mathcal{X} \subset \mathbb{R}^d$ and a geometric graph on $X_1, ..., X_n$. Assume that $\mathcal{X}$ has bottleneck not smaller than h (where the bottleneck is defined as in the definition of a valid grid). Denote $s = X_1$ and $t = X_2$. Assume that s and t can be connected by a straight line that stays inside $\mathcal{X}$ and has distance at least $h/2$ to $\partial \mathcal{X}$. Denote the distance between s and t by $d(s, t)$. Let g be the width of a valid grid on $\mathcal{X}$ and assume that $d(s, t) > 4\sqrt{d}\, g$. By $N_{\min}$ denote the minimal number of points in each grid cell, and define a as*

$$a \;:=\; \left\lfloor h/(2g\sqrt{d-1}) \right\rfloor.$$

Assume that points that are connected in the graph are at most Q grid cells apart from each other (for example, two points in the two grey cells in Figure 6b are 5 cells apart from each other). Then the effective resistance between s and t can be bounded as follows:

$$\text{In case } d > 3: \quad R_{st} \leq \frac{1}{d_s} + \frac{1}{d_t} + \left(\frac{1}{d_s} + \frac{1}{d_t}\right)\frac{1}{N_{\min}} + \frac{1}{N_{\min}^2}\left(6 + \frac{d(s,t)}{g(2a+1)^3} + 2Q\right)$$

$$\text{In case } d = 3: \quad R_{st} \leq \frac{1}{d_s} + \frac{1}{d_t} + \left(\frac{1}{d_s} + \frac{1}{d_t}\right)\frac{1}{N_{\min}}$$
$$+ \frac{1}{N_{\min}^2}\left(4\log(a) + 8 + \frac{d(s,t)}{g(2a+1)^2} + 2Q\right)$$

$$\text{In case } d = 2: \quad R_{st} \leq \frac{1}{d_s} + \frac{1}{d_t} + \left(\frac{1}{d_s} + \frac{1}{d_t}\right)\frac{1}{N_{\min}} + \frac{1}{N_{\min}^2}\left(4a + 2 + \frac{d(s,t)}{g(2a+1)} + 2Q\right)$$

Proof. The general idea of the proof is to construct a flow from s to t with the help of the underlying grid. On a high level, the construction of the proof is not so difficult, but the details are lengthy and a bit tedious. The rest of this section is devoted to it.

Construction of the flow — overview. Without loss of generality we assume that there exists a straight line connecting s and t which is along the first dimension of the space. By $C(s)$ we denote the grid cell in which s sits.

Step 0. We start a unit flow in vertex s.

Step 1. We make a step to all neighbors $\text{Neigh}(s)$ of s and distribute the flow uniformly over all edges. That is, we traverse d_s edges and send flow $1/d_s$ over each edge (see Figure 6a). This is the crucial step which, ultimately, leads to the desired limit result.

Step 2. Some of the flow now sits inside $C(s)$, but some of it might sit outside of $C(s)$. In this step, we bring back all flow to $C(s)$ in order to control it later on (see Figure 6b).

Step 3. We now distribute the flow from $C(s)$ to a larger region, namely to a hypercube $H(s)$ of side length h that is perpendicular to the linear path from s to t and centered at $C(s)$ (see the hypercubes in Figure 6c). This can be achieved in several substeps that will be defined below.

Step 4. We now traverse from $H(s)$ to an analogous hypercube $H(t)$ located at t using parallel paths, see Figure 6c.

Step 5. From the hypercube $H(t)$ we send the flow to the neighborhood $\text{Neigh}(t)$ (this is the "reverse" of steps 2 and 3).

Step 6. From $\text{Neigh}(t)$ we finally send the flow to the destination t ("reverse" of step 1).

Details of the flow construction and computation of the resistance between s and t in the general case $d > 3$. We now describe the individual steps and their contribution to the
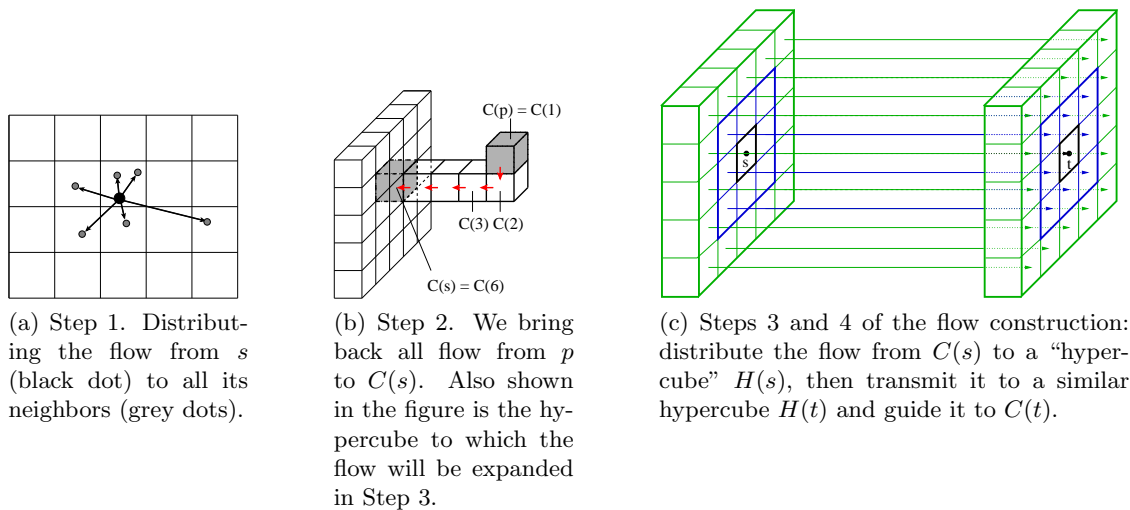
(a) Step 1. Distributing the flow from s (black dot) to all its neighbors (grey dots).

(b) Step 2. We bring back all flow from p to $C(s)$. Also shown in the figure is the hypercube to which the flow will be expanded in Step 3.

(c) Steps 3 and 4 of the flow construction: distribute the flow from $C(s)$ to a "hypercube" $H(s)$, then transmit it to a similar hypercube $H(t)$ and guide it to $C(t)$.

Figure 6: The flow construction — overview.

(a) Definition of layers.

(b) Before Step 3A starts, all flow is uniformly distributed in Layer $i-1$ (dark area).

(c) Step 3A then distributes the flow from Layer $i-1$ to the adjacent cells in Layer i

(d) After Step 3A: all flow is in Layer i, but not yet uniformly distributed

(e) Step 3B redistributes the flow in Layer i.

(f) After Step 3B, the flow is uniformly distributed in Layer i.

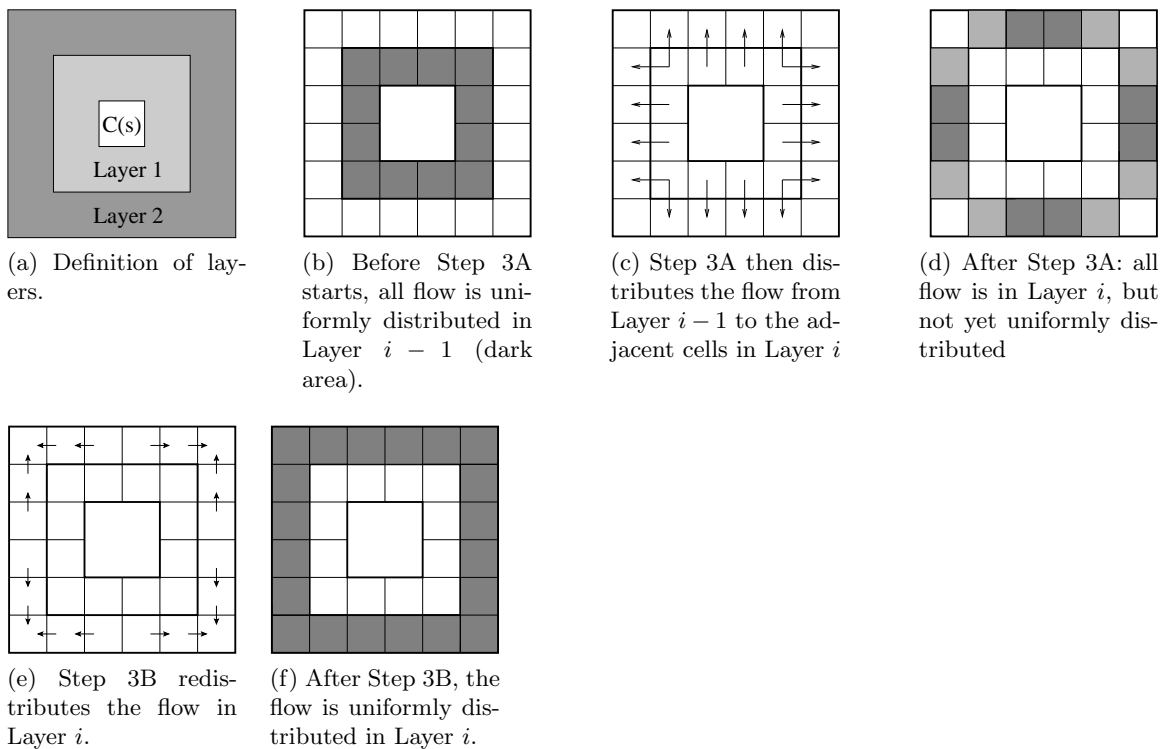Figure 7: Details of Step 3 between Layers $i-1$ and i. The first row corresponds to the expansion phase, the second row to the redistribution phase. The figure is shown for the case of $d = 3$.

bound on the resistance. We start with the general case $d > 3$. We will discuss the special cases $d = 2$ and $d = 3$ below.

In the computations below, by the "contribution of a step" we mean the part of the sum in Theorem 16 that goes over the edges considered in the current step.

Step 1: We start with a unit flow at s that we send over all d_s adjacent edges. This leads to flow $1/d_s$ over d_s edges. According to the formula in Theorem 16 this contributes

$$r_1 = d_s \cdot \frac{1}{d_s^2} = \frac{1}{d_s}$$

to the overall resistance R_{st}.

Step 2: After Step 1, the flow sits on all neighbors of s, and these neighbors are not necessarily all contained in $C(s)$. To proceed we want to re-concentrate all flow in $C(s)$. For each neighbor p of s, we thus carry the flow along a Hamming path of cells from p back to $C(s)$, see Figure 6b for an illustration.

To compute an upper bound for Step 2 we exploit that each neighbor p of s has to traverse at most Q cells to reach $C(s)$ (recall the definition of Q from the proposition). Let us fix p. After Step 1, we have flow of size $1/d_s$ in p. We now move this flow from p to all points in the neighboring cell $C(2)$ (cf. Figure 6b). For this we can use at least $N_{\min}$ edges. Thus we send flow of size $1/d_s$ over $N_{\min}$ edges, that is each edge receives flow $1/(d_s N_{\min})$. Summing the flow from $C(p)$ to $C(2)$, for all points p, gives

$$d_s N_{\min} \left(\frac{1}{d_s N_{\min}} \right)^2 = \frac{1}{d_s N_{\min}},$$

Then we transport the flow from $C(2)$ along to $C(s)$. Between each two cells on the way we can use $N_{\min}^2$ edges. Note, however, that we need to take into account that some of these edges might be used several times (for different points p). In the worst case, $C(2)$ is the same for all points p, in which case we send the whole unit flow over these edges. This amounts to flow of size $1/(N_{\min}^2)$ over $(Q-1)N_{\min}^2$ edges, that is a contribution of

$$\frac{Q-1}{N_{\min}^2}.$$

Altogether we obtain

$$r_2 \leq \frac{1}{d_s N_{\min}} + \frac{Q}{N_{\min}^2}.$$

Step 3: At the beginning of this step, the complete unit flow resides in the cube $C(s)$. We now want to distribute this flow to a "hypercube" of three dimensions (no matter what d

is, as long as $d > 3$) that is perpendicular to the line that connects s and t (see Figure 6c, where the case of $d = 3$ and a 2-dimensional "hypercube" are shown). To distribute the flow to this cube we divide it into layers (see Figure 7a). Layer 0 consists of the cell $C(s)$ itself, the first layer consists of all cells adjacent to $C(s)$, and so on. Each side of Layer i consists of

$$l_i = (2i + 1)$$

cells. For the 3-dimensional cube, the number z_i of grid cells in Layer i, $i \geq 1$, is given as

$$z_i = \underbrace{6 \cdot (2i - 1)^2}_{\text{interior cells of the faces}} + \underbrace{12 \cdot (2i - 1)}_{\text{cells along the edges (excluding corners)}} + \underbrace{8}_{\text{corner cells}} = 24i^2 + 2.$$

All in all we consider

$$a = \left\lfloor h/(2g\sqrt{d-1}) \right\rfloor \leq \left\lfloor h/(2(g-1)\sqrt{d-1}) \right\rfloor$$

layers, so that the final layer has diameter just a bit smaller than the bottleneck h. We now distribute the flow stepwise through all layers, starting with unit flow in Layer 0. To send the flow from Layer $i - 1$ to Layer i we use two phases, see Figure 7 for details. In the "expansion phase" 3A(i) we transmit the flow from Layer $i - 1$ to all adjacent cells in Layer i. In the "redistribution phase" 3B(i) we then redistribute the flow in Layer i to achieve that it is uniformly distributed in Layer i. In all phases, the aim is to use as many edges as possible.

Expansion phase 3A(i). We can lower bound the number of edges between Layer $i - 1$ and Layer i by $z_{i-1}N_{\min}^2$: each of the z_{i-1} cells in Layer $i - 1$ is adjacent to at least one of the cells in Layer i, and each cell contains at least $N_{\min}$ points. Consequently, we can upper bound the contribution of the edges in the expansion phase 3A(i) to the resistance by

$$r_{3A(i)} \leq z_{i-1}N_{\min}^2 \cdot \left(\frac{1}{z_{i-1}N_{\min}^2} \right)^2 = \frac{1}{z_{i-1}N_{\min}^2}.$$

Redistribution phase 3B(i). We make a crude upper bound for the redistribution phase. In this phase we have to move some part of the flow from each cell to its neighboring cells. For simplicity we bound this by assuming that for each cell, we had to move *all* its flow to neighboring cells. By a similar argument as for Step 3A(i), the contribution of the redistribution step can be bounded by

$$r_{3B(i)} \leq z_i N_{\min}^2 \cdot \left(\frac{1}{z_i N_{\min}^2} \right)^2 = \frac{1}{z_i N_{\min}^2}.$$

All of Step 3. All in all we have a layers. Thus the overall contribution of Step 3 to the resistance can be bounded by

$$r_3 = \sum_{i=1}^{a} r_{3A(i)} + r_{3B(i)} \leq \frac{2}{N_{\min}^2} \sum_{i=1}^{a} \frac{1}{z_{i-1}} \leq \frac{2}{N_{\min}^2} \left(1 + \frac{1}{24} \sum_{i=1}^{a-1} \frac{1}{i^2} \right) \leq \frac{3}{N_{\min}^2}.$$

To see the last inequality, note that the sum $\sum_{i=1}^{a-1} 1/i^2$ is a partial sum of the over-harmonic series that converges to a constant smaller than 2.

Step 4: Now we transfer all flow in "parallel cell paths" from $H(s)$ to $H(t)$. We have $(2a+1)^3$ parallel rows of cells going from $H(s)$ to $H(t)$, each of them contains $d(s,t)/g$ cells. Thus all in all we traverse $(2a+1)^3 N_{\min}^2 d(s,t)/g$ edges, and each edge carries flow $1/((2a+1)^3 N_{\min}^2)$. Thus step 4 contributes

$$r_4 \leq (2a+1)^3 N_{\min}^2 \frac{d(s,t)}{g} \cdot \left(\frac{1}{(2a+1)^3 N_{\min}^2} \right)^2 = \frac{d(s,t)}{g(2a+1)^3 N_{\min}^2}.$$

Step 5 is completely analogous to steps 2 and 3, with the analogous contribution $r_5 = \frac{1}{d_t N_{\min}} + \frac{Q}{N_{\min}^2} + r_3$.

Step 6 is completely analogous to step 1 with overall contribution of $r_6 = 1/d_t$.

Summing up the general case $d > 3$. All these contributions lead to the following overall bound on the resistance in case $d > 3$:

$$R_{st} \leq \frac{1}{d_s} + \frac{1}{d_t} + \left(\frac{1}{d_s} + \frac{1}{d_t} \right) \frac{1}{N_{\min}} + \frac{1}{N_{\min}^2} \left(6 + \frac{d(s,t)}{g(2a+1)^3} + 2Q \right)$$

with a and Q as defined in Proposition 18. This is the result stated in the proposition for case $d > 3$.

Note that as spelled out above, the proof works whenever the dimension of the space satisfies $d > 3$. In particular, note that even if d is large, we only use a 3-dimensional "hypercube" in Step 3. It is sufficient to give the rate we need, and carrying out the construction for higher-dimensional hypercube (in particular Step 3B) is a pain that we wanted to avoid.

The special case $d = 3$. In this case, everything works similar to above, except that we we only use a 2-dimensional "hypercube" (this is what we always show in the figures). The only place in the proof where this really makes a difference is in Step 3. The number z_i of grid cells in Layer i is given as $z_i = 8i$. Consequently, instead of obtaining an over-harmonic sum in r_3 we obtain a harmonic sum. Using the well-known fact that $\sum_{i=1}^{a} 1/i \leq \log(a) + 1$ we obtain

$$r_3 \leq \frac{2}{N_{\min}^2} \left(1 + \frac{1}{8} \sum_{i=1}^{a-1} \frac{1}{i}\right) \leq \frac{2}{N_{\min}^2} \left(2 + \log(a)\right).$$

In Step 4 we just have to replace the terms $(2a+1)^3$ by $(2a+1)^2$. This leads to the result in Proposition 18.

The special case $d = 2$. Here our "hypercube" only consists of a "pillar" of $2a+1$ cells. The fundamental difference to higher dimensions is that in Step 3, the flow does not have so much "space" to be distributed. Essentially, we have to distribute all unit flow through a "pillar", which results in contributions

$$r_3 \leq \frac{2a+1}{N_{\min}^2}$$
$$r_4 \leq \frac{d(s,t)}{g} \frac{1}{(2a+1)N_{\min}^2}.$$

This concludes the proof of Proposition 18. ■

Let us make a couple of technical remarks about this proof. For the ease of presentation we simplified the proof in a couple of respects.

Strictly speaking, we do not need to distribute the whole unit flow to the outmost Layer a. The reason is that in each layer, a fraction of the flow already "branches off" in direction of t. We simply ignore this leaving flow when bounding the flow in Step 3, our construction leads to an upper bound. It is not difficult to take the outbound flow into account, but it does not change the order of magnitude of the final result. So for the ease of presentation we drop this additional complication and stick to our rough upper bound.

When we consider Steps 2 and 3 together, it turns out that we might have introduced some loops in the flow. To construct a proper flow, we can simply remove these loops. This would then just reduce the contribution of Steps 2 and 3, so that our current estimate is an overestimation of the whole resistance.

The proof as it is spelled out above considers the case where s and t are connected by a straight line. It can be generalized to the case where they are connected by a piecewise linear path. This does not change the result by more than constants, but adds some technicality at the corners of the paths.

The construction of the flow only works if the bottleneck of $\mathcal{X}$ is not smaller than the diameter of one grid cell, if s and t are at least a couple of grid cells apart from each other, and if s and t are not too close to the boundary of $\mathcal{X}$. We took care of these conditions in Part 3 of the definition of a valid grid.

6.3 Proof of the Theorems 3 and 4

First of all, note that by Rayleigh's principle (cf. Corollary 7 in Section IX.2 of Bollobas, 1998) the effective resistance between vertices cannot decrease if we delete edges from the graph. Given a sample from the underlying density p, a random geometric graph based on this sample, and some valid region $\mathcal{X}$, we first delete all points that are not in $\mathcal{X}$. Then we consider the remaining geometric graph. The effective resistances on this graph are upper bounds on the resistances of the original graph. Then we conclude the proofs with the following arguments:

Proof of Theorem 3. The lower bound on the deviation follows immediately from Proposition 15. The upper bound is a consequence of Proposition 18 and well known properties of random geometric graphs (summarized in the appendix). In particular, note that we can choose the grid width $g := \varepsilon/(2\sqrt{d-1})$ to obtain a valid grid. The quantity $N_{\min}$ can be bounded as stated in Proposition 28 and is of order $n\varepsilon^d$, the degrees behave as described in Proposition 29 and are also of order $n\varepsilon^d$ (we use $\delta = 1/2$ in these results for simplicity). The quantity a in Proposition 18 is of the order $1/\varepsilon$, and Q can be bounded by $Q = \varepsilon/g$ and by the choice of g is indeed a constant. Plugging all these results together leads to the final statement of the theorem. ∎

Proof of Theorem 4. This proof is analogous to the ε-graph. As grid width g we choose $g = R_{k,\min}/(2\sqrt{d-1})$ where $R_{k,\min}$ is the minimal k-nearest neighbor distance (note that this works for both the symmetric and the mutual kNN-graph). Exploiting Propositions 28 and 30 we can see that $R_{k,\min}$ and $R_{k,\max}$ are of order $(k/n)^{1/d}$, the degrees and $N_{\min}$ are of order k, a is of the order $(n/k)^{1/d}$ and Q a constant. Now the statements of the theorem follow from Proposition 18. ∎

7. Proofs for the Spectral Approach

In this section we present the proofs of the results obtained by spectral arguments.

7.1 Proof of Propositions 1 and 5

First we prove the general formulas to compute and approximate the hitting times.

Proof of Proposition 1. For the hitting time formula, let $u_1, \ldots, u_n$ be an orthonormal set of eigenvectors of the matrix $D^{-1/2}WD^{-1/2}$ corresponding to the eigenvalues $\mu_1, \ldots, \mu_n$. Let u_{ij} denote the j-th entry of u_i. According to Lovász (1993) the hitting time is given by

$$H_{ij} = \mathrm{vol}(G) \sum_{k=2}^{n} \frac{1}{1-\mu_k} \left(\frac{u_{kj}^2}{d_j} - \frac{u_{ki}u_{kj}}{\sqrt{d_i d_j}} \right).$$

A straightforward calculation using the spectral representation of L_{sym} yields

$$H_{ij} = \mathrm{vol}(G) \left\langle \tfrac{1}{\sqrt{d_j}}e_j, \sum_{k=2}^{n} \frac{1}{1-\mu_k} \left\langle u_k, \tfrac{1}{\sqrt{d_j}}e_j - \tfrac{1}{\sqrt{d_i}}e_i \right\rangle u_k \right\rangle$$

$$= \text{vol}(G) \left\langle \frac{1}{\sqrt{d_j}} e_j, \, L_{\text{sym}}^\dagger \left(\frac{1}{\sqrt{d_j}} e_j - \frac{1}{\sqrt{d_i}} e_i \right) \right\rangle.$$

The result for the commute time follows from the one for the hitting times. ∎

In order to prove Proposition 5 we first state a small lemma. For convenience, we set $A = D^{-1/2} W D^{-1/2}$ and $u_i = e_i / \sqrt{d_i}$. Furthermore, we are going to denote the projection on the eigenspace of the j-the eigenvalue λ_j of A by P_j.

Lemma 19 (Pseudo-inverse $L_{\text{sym}}^\dagger$) *The pseudo-inverse of the symmetric Laplacian satisfies*

$$L_{sym}^\dagger = I - P_1 + M,$$

where I denotes the identity matrix and M is given as follows:

$$M \;=\; \sum_{k=1}^\infty (A - P_1)^k \;=\; \sum_{r=2}^n \frac{\lambda_r}{1 - \lambda_r} P_r. \tag{3}$$

Furthermore, for all $u, v \in \mathbb{R}^n$ we have

$$|\langle u, Mv \rangle| \;\leq\; \frac{1}{1 - \lambda_2} \cdot \|(A - P_1)u\| \cdot \|(A - P_1)v\| + |\langle u, \, (A - P_1)v \rangle|. \tag{4}$$

Proof. The projection onto the null space of L_{sym} is given by $P_1 = \sqrt{d}\sqrt{d}^T / \sum_{i=1} d_i$ where $\sqrt{d} = (\sqrt{d_1}, \dots, \sqrt{d_n})^T$. As the graph is not bipartite, $\lambda_n > -1$. Thus the pseudoinverse of L_{sym} can be computed as

$$L_{\text{sym}}^\dagger = (I - A)^\dagger = (I - A + P_1)^{-1} - P_1 = \sum_{k=0}^\infty (A - P_1)^k - P_1.$$

Thus

$$M := \sum_{k=1}^\infty (A - P_1)^k \;=\; \sum_{k=0}^\infty (A - P_1)^k (A - P_1)$$

$$= \left(\sum_{k=0}^\infty \sum_{r=2}^n \lambda_r^k P_r\right)\left(\sum_{r=2}^n \lambda_r P_r\right) \;=\; \left(\sum_{r=2}^n \frac{1}{1 - \lambda_r} P_r\right)\left(\sum_{r=2}^n \lambda_r P_r\right)$$

$$= \sum_{r=2}^n \frac{\lambda_r}{1 - \lambda_r} P_r$$

which proves Equation (3). By a little detour, we can also see

$$M \;=\; \sum_{k=0}^\infty (A - P_1)^k (A - P_1)^2 + (A - P_1) \;=\; \left(\sum_{r=2}^n \frac{1}{1 - \lambda_r} P_r\right)(A - P_1)^2 + (A - P_1).$$

Exploiting that $(A - P_1)$ commutes with all P_r gives

$$\langle u, Mv \rangle = \left\langle (A - P_1)u \ , \ \left(\sum_{r=2}^{n} \frac{1}{1 - \lambda_r} P_r \right)(A - P_1)v \right\rangle + \langle u, (A - P_1)v \rangle .$$

Applying the Cauchy-Schwarz inequality and the fact $\| \sum_{r=2}^{n} \frac{1}{1-\lambda_r} P_r \|_2 = 1/(1 - \lambda_2)$ leads to the desired statement. ∎

Proof of Proposition 5. This proposition now follows easily from the Lemma above. Observe that

$$\langle u_i, A u_j \rangle = \frac{w_{ij}}{d_i d_j} \leq \frac{w_{\max}}{d_{\min}^2}$$

$$\|A u_i\|^2 = \sum_{k=1}^{n} \frac{w_{ik}^2}{d_i^2 d_k} \leq \frac{w_{\max}}{d_{\min} d_i^2} \sum_{k} w_{ik} = \frac{w_{\max}}{d_{\min}} \frac{1}{d_i} \leq \frac{w_{\max}}{d_{\min}^2}$$

$$\|A(u_i - u_j)\|^2 \leq \frac{w_{\max}}{d_{\min}} \left(\frac{1}{d_i} + \frac{1}{d_j} \right) \leq \frac{2 w_{\max}}{d_{\min}^2}.$$

Exploiting that $P_1(u_i - u_j) = 0$ we get for the hitting time

$$\left| \frac{1}{\mathrm{vol}(G)} H_{ij} - \frac{1}{d_j} \right| = |\langle u_j, M(u_j - u_i) \rangle|$$

$$\leq \frac{1}{1 - \lambda_2} \|A u_j\| \cdot \|A(u_j - u_i)\| + |\langle u_j, A(u_j - u_i) \rangle|$$

$$\leq \frac{1}{1 - \lambda_2} \frac{w_{\max}}{d_{\min}} \left(\frac{1}{\sqrt{d_j}} \sqrt{\frac{1}{d_i} + \frac{1}{d_j}} \right) + \frac{w_{ij}}{d_i d_j} + \frac{w_{jj}}{d_j^2}$$

$$\leq 2 \frac{w_{\max}}{d_{\min}^2} \left(\frac{1}{1 - \lambda_2} + 1 \right).$$

For the commute time, we note that

$$\left| \frac{1}{\mathrm{vol}(G)} C_{ij} - \left(\frac{1}{d_i} + \frac{1}{d_j} \right) \right| = |\langle u_i - u_j, M(u_i - u_j) \rangle|$$

$$\leq \frac{1}{1 - \lambda_2} \|A(u_i - u_j)\|^2 + |\langle u_i - u_j, A(u_i - u_j) \rangle|$$

$$\leq \frac{w_{\max}}{d_{\min}} \left(\frac{1}{1 - \lambda_2} + 2 \right) \left(\frac{1}{d_i} + \frac{1}{d_j} \right).$$

∎

We would like to point out that the key to achieving this bound is not to give in to the temptation to manipulate Eq. (3) directly, but to bound Eq. (4). The reason is that we can compute terms of the form $\langle u_i, A u_j \rangle$ and related terms explicitly, whereas we do not have any explicit formulas for the eigenvalues and eigenvectors in (3).

7.2 The Spectral Gap in Random Geometric Graphs

As we have seen above, a key ingredient in the approximation result for hitting times and commute distances is the spectral gap. In this section we show how the spectral gap can be lower bounded for random geometric graphs. We first consider the case of a fixed geometric graph. From this general result we then derive the results for the special cases of the ε-graph and the kNN-graphs. All graphs considered in this section are unweighted and undirected. We follow the strategy in Boyd et al. (2005) where the spectral gap is bounded by means of the Poincaré inequality (see Diaconis and Stroock, 1991, for a general introduction to this technique; see Cooper and Frieze, 2011, for a related approach in simpler settings). The outline of this technique is as follows: for each pair (X, Y) of vertices in the graph we need to select a path γ_{XY} in the graph that connects these two vertices. In our case, this selection is made in a random manner. Then we need to consider all edges in the graph and investigate how many of the paths γ_{XY}, on average, traverse this edge. We need to control the maximum of this "load" over all edges. The higher this load is, the more pronounced is the bottleneck in the graph, and the smaller the spectral gap is. Formally, the spectral gap is related to the maximum average load b as follows.

Proposition 20 (Spectral gap, Diaconis and Stroock, 1991) *Consider a finite, connected, undirected, unweighted graph that is not bipartite. For each pair of vertices $X \neq Y$ let P_{XY} be a probability distribution over all paths that connect X and Y and have uneven length. Let $(\gamma_{XY})_{X,Y}$ be a family of paths independently drawn from the respective P_{XY}. Define $b := \max_{\{e\ edge\}} \mathbb{E}|\{\gamma_{XY} \mid e \in \gamma_{XY}\}|$. Denote by $|\gamma_{\max}|$ the maximum path length (where the length of the path is the number of edges in the path). Then the spectral gap in the graph is bounded as follows:*

$$1 - \lambda_2 \geq \frac{\mathrm{vol}(G)}{d_{\max}^2 |\gamma_{\max}| b} \qquad and \qquad 1 - |\lambda_n| \geq \frac{2}{d_{\max} |\gamma_{\max}| b}.$$

For deterministic sets Γ, this proposition has been derived as Corollary 1 and 2 in Diaconis and Stroock (1991). The adaptation for random selection of paths is straightforward, see Boyd et al. (2005).

The key to tight bounds based on Proposition 20 is a clever choice of the paths. We need to make sure that we distribute the paths as "uniformly" as possible over the whole graph. This is relatively easy to achieve in the special situation where $\mathcal{X}$ is a torus with uniform distribution (as studied in Boyd et al., 2005; Cooper and Frieze, 2011) because of symmetry arguments and the absence of boundary effects. However, in our setting with general $\mathcal{X}$ and p we have to invest quite some work.

7.2.1 Fixed Geometric Graph on the Unit Cube in $\mathbb{R}^d$

We first treat the special case of a fixed geometric graph with vertices in the unit cube $[0, 1]^d$ in $\mathbb{R}^d$. Consider a grid on the cube with grid width g. For now we assume that the grid cells are so small that points in neighboring cells are always connected in the geometric graph, and so large that each cell contains a minimal number of data points. We will specify the exact value of g later. In the following, cells of the grid are identified with their center points.
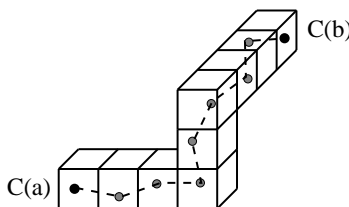
Figure 8: Canonical path between a and b. We first consider a "Hamming path of cells" between a and b. In all intermediate cells, we randomly pick a point.

Construction of the paths. Assume we want to construct a path between two vertices a and b that correspond to the points $a = (a_1, \ldots, a_d)$, $b = (b_1, \ldots, b_d) \in [0, 1]^d$. Let $C(a)$ and $C(b)$ denote the grid cells containing a and b, denote the centers of these cells by $c(a) = (c(a)_1, \ldots, c(a)_d)$ and $c(b) = (c(b)_1, \ldots, c(b)_d)$. We first construct a deterministic "cell path" between the cells $C(a)$ and $C(b)$ (see Figure 8). This path simply follows a Hamming path: starting at cell $C(a)$ we change the first coordinate until we have reached $c(b)_1$. For example, if $c(a)_1 < c(b)_1$ we traverse the cells

$$\big(c(a)_1, c(a)_2, \ldots, c(a)_d\big) \rightsquigarrow \big(c(a)_1 + g, c(a)_2, \ldots, c(a)_d\big) \rightsquigarrow \ldots \rightsquigarrow \big(c(b)_1, c(a)_2, \ldots, c(a)_d\big).$$

Then we move along the second coordinate from $c(a)_2$ until we have reached $c(b)_2$, that is we traverse the cells $(c(b)_1, *, c(a)_3, \ldots, c(a)_d)$. And so on. This gives a deterministic way of traversing adjacent cells from $C(a)$ to $C(b)$. Now we transform this deterministic "cell path" to a random path on the graph. In the special cases where a and b are in the same cell or in neighboring cells, we directly connect a and b by an edge. In the general case, we select one data point uniformly at random in each of the interior cells on the cell path. Then we connect the selected points to form a path. Note that we can always force the paths to have uneven lengths by adding one more point somewhere in between.

Proposition 21 (Path construction is valid) *Assume that (1) Each cell of the grid contains at least one data point. (2) Data points in the same and in neighboring cells are always connected in the graph. Then the graph is connected, and the paths constructed above are paths in the graph.*

Proof. Obvious, by construction of the paths. ∎

In order to apply Proposition 20 we now need to compute the maximal average load of all paths.

Proposition 22 (Maximum average load for fixed graph on cube) *Consider a geometric graph on $[0, 1]^d$ and the grid of width g on $[0, 1]^d$. Denote by $N_{\min}$ and $N_{\max}$ the minimal and maximal number of points per grid cell. Construct a random set of paths as described above.*

1. *Let C be any fixed cell in the grid. Then there exist at most d/g^{d+1} pairs of cells (A, B) such that cell paths starting in cell A and ending in cell B pass through C.*

1781

2. If the path construction is valid, then the maximal average load is upper bounded by

$$b \le 1 + \left(\frac{N_{\max}^2}{N_{\min}^2} + 2\frac{N_{\max}}{N_{\min}} \right) \frac{d}{g^{d+1}}.$$

Proof. *Part 1.* We identify cells with their centers. Consider two different grid cells A and B with centers a and b. By construction, the Hamming path between A and B has the corners

$$a = (a_1, a_2, a_3, \ldots, a_d) \rightsquigarrow (b_1, a_2, a_3, \ldots, a_d) \rightsquigarrow (b_1, b_2, a_3, \ldots, a_d)$$
$$\rightsquigarrow \ldots \rightsquigarrow (b_1, b_2, b_3, \ldots, b_{d-1}, a_d) \rightsquigarrow (b_1, b_2, b_3, \ldots, b_{d-1}, b_d) = b.$$

All cells on the path have the form $(b_1, b_2, \ldots, b_{l-1}, *, a_{l+1}, \ldots, a_d)$ where $*$ can take any value between a_l and b_l. A path can only pass through the fixed cell with center c if there exists some $l \in \{1, \ldots, d\}$ such that

$$(c_1, \ldots, c_d) = (b_1, b_2, \ldots, b_{l-1}, *, a_{l+1}, \ldots, a_d).$$

That is, there exists some $l \in \{1, \ldots, d\}$ such that

$\quad$ *(I)* $b_i = c_i$ for all $i = 1, \ldots, l-1$ $\qquad$ and $\qquad$ *(II)* $a_i = c_i$ for all $i = l+1, \ldots, d.$

For the given grid size g there are $1/g$ different cell centers per dimension. For fixed l there thus exist $1/g^{d-l+1}$ cell centers that satisfy *(I)* and $1/g^l$ cell centers that satisfy *(II)*. So all in all there are $1/g^{d+1}$ pairs of cells A and B such that both *(I)* and *(II)* are satisfied for a fixed value of l. Adding up the possibilities for all choices of $l \in \{1, \ldots, d\}$ leads to the factor d.

Part 2. Fix an edge e in the graph and consider its two adjacent vertices v_1 and v_2. If v_1 and v_2 are in two different cells that are not neighbors to each other, then by construction none of the paths traverses the edge. If they are in the same cell, by construction at most one of the paths can traverse this edge, namely the one directly connecting the two points. The interesting case is the one where v_1 and v_2 lie in two neighboring grid cells C and $\tilde{C}$. If both cells are "interior" cells of the path, then by construction each edge connecting the two cells has equal probability of being selected. As there are at least $N_{\min}$ points in each cell, there are at least $N_{\min}^2$ different edges between these cells. Thus each of the edges between the cells is selected with probability at most $1/N_{\min}^2$. We know by Part 1 that there are at most d/g^{d+1} pairs of start/end cells. As each cell contains at most $N_{\max}$ points, this leads to $N_{\max}^2 d/g^{d+1}$ different paths passing through C. This is also an upper bound on the number of paths passing through both C and $\tilde{C}$. Thus, the number of paths using each edge is at most $dN_{\max}^2/(g^{d+1}N_{\min}^2)$.

If at least one of the cells is the start cell of the path, then the corresponding vertex, say v_1, is the start point of the path. If v_2 is an intermediate point, then it is selected with probability at most $1/N_{\min}$ (the case where v_2 is an end point has already been treated at the beginning). Similarly to the last case, there are at most $N_{\max} d/g^{d+1}$ paths that start in v_1 and pass through $\tilde{C}$. This leads to an average load of $dN_{\max}/(g^{d+1}N_{\min})$ on edge e. The same holds with the roles of v_1 and v_2 exchanged, leading to a factor 2.

The overall average load is now the sum of the average loads in the different cases. $\blacksquare$

7.2.2 FIXED GEOMETRIC GRAPH ON A DOMAIN $\mathcal{X}$ THAT IS HOMEOMORPHIC TO A CUBE

Now assume that $\mathcal{X} \subset \mathbb{R}^d$ is a compact subset that is homeomorphic to the cube $[0,1]^d$ in the following sense: we assume that there exists a homeomorphism $h : \mathcal{X} \to [0,1]^d$ and constants $0 < L_{\min} < L_{\max} < \infty$ such that for all $x, y \in \mathcal{X}$ we have

$$L_{\min}\|x - y\| \ \leq \ \|h(x) - h(y)\| \ \leq \ L_{\max}\|x - y\|. \tag{5}$$

The general idea is now as follows. Assume we are given a geometric graph on $X_1, \ldots, X_n \in \mathcal{X}$. In order to construct the paths we first map the points in the cube using h. Then we construct the paths on $h(X_1), \ldots, h(X_n) \in [0,1]^d$ as in the last section. Finally, we map the path back to $\mathcal{X}$.

Proposition 23 (Maximum average load for fixed graph on general domain)
Let G be a geometric graph based on $X_1, \ldots, X_n \in \mathcal{X}$. Assume that there exists some $\tilde{g} > 0$ such that points of distance smaller than $\tilde{g}$ are always connected in the graph. Consider a mapping $h : \mathcal{X} \to [0,1]^d$ as in Equation (5) and a grid of width g on $[0,1]^d$. Let $(C_i)_i$ be the cells of the g-grid on $[0,1]^d$, denote their centers by c_i. Let B_i and B_i' be balls in $\mathcal{X}$ with radius $r = g/(2L_{\max})$ and $R = \sqrt{d}\, g/L_{\min}$ centered at $h^{-1}(c_i)$.

1. *These balls satisfy $B_i \subseteq h^{-1}(C_i) \subseteq B_i'$.*

2. *Denote by $\tilde{N}_{\min}$ the minimal number of points in B_i and $\tilde{N}_{\max}$ the maximal number of points in B_i'. Construct paths between the points $h(X_i) \in [0,1]^d$ as described in the previous subsection. If $\tilde{N}_{\min} \geq 1$ and $g \leq L_{\min}\tilde{g}/\sqrt{d+3}$, then these paths are valid.*

3. *In this case, the maximal average load can be upper bounded by*

$$1 + \left(\frac{\tilde{N}_{\max}^2}{\tilde{N}_{\min}^2} + 2\frac{\tilde{N}_{\max}}{\tilde{N}_{\min}} \right) \frac{d}{(\tilde{g}L_{\min}/\sqrt{d+3})^{d+1}}.$$

Proof. *Part (1).* To see the first inclusion, consider any point $x \in B_i$. By the Lipschitz continuity and the definition of B_i we get

$$\|h(x) - c_i\| \leq L_{max}\|x - h^{-1}(c_i)\| \leq L_{max} \cdot r = g/2.$$

Hence, $h(x) \in C_i$. To show the other inclusion let $x \in h^{-1}(C_i)$. Then by the Lipschitz continuity and the definition of C_i we get

$$\|x - h^{-1}(c_i)\| \leq \|h(x) - c_i\|/L_{\min} \leq \sqrt{d}g/L_{\min} = R,$$

hence $x \in B_i'$.
Part (2). By the definition of $\tilde{N}_{\min}$ and Part (1) it is clear that each cell of the grid contains at least one point. Consider two points $X_i, X_j \in \mathcal{X}$ such that $h(X_i)$ and $h(X_j)$ are in neighboring cells of the g-grid. Then $\|h(X_i) - h(X_j)\| \leq g\sqrt{d+3}$. By the properties of h,

$$\|h^{-1}(X_i) - h^{-1}(X_j)\| \leq \frac{1}{L_{\min}}\|X_i - X_j\| \leq \frac{1}{L_{\min}}\sqrt{d+3}\, g \ \leq \ \tilde{g}.$$

Thus, by the definition of $\tilde{g}$ the points X_i and X_j are connected in G.

Part (3). Follows directly from Proposition 22 and Parts (1) and (2). ∎

7.2.3 Spectral Gap for the ε-Graph

Now we are going to apply Proposition 23 to ε-graphs. We will use the general results on ε-graphs summarized in the appendix.

Proposition 24 (Maximal average load for ε-graph) *Assume that $\mathcal{X}$ is homeomorphic to the cube with a mapping h as described in Equation (5). Then there exist constants $c_1, c_2, c_3 > 0$ such that with probability at least $1 - c_1 \exp(-c_2 n \varepsilon^d)/\varepsilon^d$, the maximum average load is upper bounded by c_3/ε^{d+1}. If $n\varepsilon^d/\log n \to \infty$, then this probability tends to 1 as $n \to \infty$.*

Proof. The proof is based on Proposition 23. By construction we know that points with distance at most $\tilde{g} = \varepsilon$ are always connected in the ε-graph. By Part 2 of Proposition 23, to ensure that points in neighboring grid cells are always connected in the graph we thus need to choose the grid width $g = \varepsilon \cdot L_{\min}/\sqrt{d+3}$. The radius r defined in Proposition 23 is then given as

$$r = \frac{g}{2L_{\max}} = \varepsilon \cdot \frac{L_{\min}}{2\sqrt{d+3}L_{\max}}.$$

The probability mass of the balls B_i is thus bounded by

$$b_{\min} \geq r^d \eta_d p_{\min}\alpha = \varepsilon^d \cdot \left(\frac{L_{\min}}{L_{\max}}\right)^d \frac{\eta_d}{2^d(d+3)^{d/2}} p_{\min}\alpha =: \varepsilon^d \cdot c_{\min}$$

(recall that the constant α takes care of boundary points, see Definition 2 of the valid region).
We have

$$K = 1/g^d = \sqrt{d+3}/L_{\min}^d \cdot (1/\varepsilon^d) =: \kappa \cdot (1/\varepsilon^d)$$

grid cells and thus the same number of balls B_i. We can now apply Proposition 28 (with $\delta := 1/2$) to deduce the bound for the quantity $\tilde{N}_{\min}$ used in Proposition 23:

$$P\left(\tilde{N}_{\min} \leq n\varepsilon^d c_{\min}/2\right) \leq \frac{\kappa}{\varepsilon^d} \exp(-n\varepsilon^d c_{\min}/12).$$

Analogously, for $\tilde{N}_{\max}$ we have $R = \varepsilon > \varepsilon\sqrt{d}/\sqrt{d+3}$ and $b_{\max} = R^d \eta_d p_{\max} = \varepsilon^d \eta_d p_{\max} := \varepsilon^d \cdot c_{\max}$. With $\delta = 0.5$ we then obtain

$$P\left(\tilde{N}_{\max} \geq n\varepsilon^d c_{\max}3/2\right) \leq \frac{\kappa}{\varepsilon^d} \exp(-n\varepsilon^d c_{\max}/12).$$

Plugging these values into Proposition 23 leads to the result. ∎

Proof of Theorem 6. With probability at least $1 - c_1 n \exp(-c_2 n\varepsilon^d)$, both the minimal and maximal degrees in the graph are of the order $\Theta(n\varepsilon^d)$ (cf. Proposition 29), and the volume of G is of order $\Theta(n^2\varepsilon^d)$. To compute the maximal number $|\gamma_{\max}|$ of edges in each of the paths constructed above, observe that each path can traverse at most $d \cdot 1/g = (d\sqrt{d+3}/L_{\min}) \cdot (1/\varepsilon)$ cubes, and a path contains just one edge per cube. Thus $|\gamma_{\max}|$ is of the order $\Theta(1/\varepsilon)$. In Proposition 24 we have seen that with probability at least $c_4 \exp(-c_5 n\varepsilon^d)/\varepsilon^d$ the maximum average load b is of the order $\Omega(1/\varepsilon^{d+1})$. Plugging all these quantities in Proposition 20 leads to the result. ∎

7.2.4 SPECTRAL GAP FOR THE kNN-GRAPH

As in the case of the flow proofs, the techniques in the case of the kNN-graphs are identical to the ones for the ε-graph, we just have to replace the deterministic radius ε by the minimal kNN-radius. As before we exploit that if two sample points have distance less than $R_{k,\min}$ from each other, then they are always connected both in the symmetric and mutual kNN-graph.

Proposition 25 (Maximal average load in the kNN-graph) *Under the general assumptions, with probability at least $1 - c_1 \cdot n \cdot \exp(-c_2 k)$ the maximal average load in both the symmetric and mutual kNN-graph is bounded from above by $c_3 (n/k)^{1+1/d}$. If $k/\log n \to \infty$, then this probability converges to 1.*

Proof. This proof is analogous to the one of Proposition 24, the role of ε is now taken over by $R_{k,\min}$. ∎

Proof of Theorem 7. With probabilities at least $1 - n \exp(-c_1 k)$ the following statements hold: the minimal and maximal degree are of order $\Theta(k)$, thus the number of edges in the graph is of order $\Theta(nk)$. Analogously to the proof for the ε-graph, the maximal path length $|\gamma_{\max}|$ is of the order $1/R_{k,\min} = \Theta((k/n)^{1/d})$. The maximal average load is of the order $O((n/k)^{d+1/d})$. Plugging all these quantities in Proposition 20 leads to the result. ∎

7.3 Proofs of Corollaries 8 and 9

Now we collected all ingredients to finally present the following proofs.

Proof of Corollary 8. This is a direct consequence of the results on the minimal degree (Proposition 29) and the spectral gap (Theorem 6). Plugging these results into Proposition 5 leads to the first result. The last statement in the theorem follows by a standard density estimation argument, as the degree of a vertex in the ε-graph is a consistent density estimator (see Proposition 29). ∎

Proof of Corollary 9. Follows similarly as Theorem 8 by applying Proposition 5. The results on the minimal degree and the spectral gap can be found in Proposition 30 and Theorem 7. The last statement follows from the convergence of the degrees, see Proposition 30. ∎

7.4 Weighted Graphs

For weighted graphs, we use the following results from the literature.

Proposition 26 (Spectral gap in weighted graphs) *1. For any row-stochastic matrix P,*

$$\lambda_2 \leq \frac{1}{2} \max_{i,j} \sum_{k=1}^{n} \left| \frac{w_{ik}}{d_i} - \frac{w_{jk}}{d_j} \right| \leq 1 - n \min_{i,j} \frac{w_{ij}}{d_i} \leq 1 - \frac{w_{\min}}{w_{\max}}.$$

2. Consider a weighted graph G with edge weights $0 < w_{\min} \leq w_{ij} \leq w_{\max}$ and denote its second eigenvalue by $\lambda_{2,weighted}$. Consider the corresponding unweighted graph where

all edge weights are replaced by 1, and denote its second eigenvalue by $\lambda_{2,unweighted}$. Then we have

$$(1 - \lambda_{2,unweighted}) \cdot \frac{w_{\min}}{w_{\max}} \quad \le \quad (1 - \lambda_{2,weighted}) \quad \le \quad (1 - \lambda_{2,unweighted}) \cdot \frac{w_{\max}}{w_{\min}}.$$

Proof.

1. This bound was obtained by Zenger (1972), see also Section 2.5 of Seneta (2006) for a discussion. Note that the second inequality is far from being tight. But in our application, both bounds lead to similar results.

2. This statement follows directly from the well-known representation of the second eigenvalue μ_2 of the normalized graph Laplacian L_{sym} (see Sec. 1.2 in Chung, 1997),

$$\mu_2 = \inf_{f \in \mathbb{R}^n} \frac{\sum_{i,j=1}^n w_{ij}(f_i - f_j)^2}{\min_{c \in \mathbb{R}} \sum_{i=1}^n d_i (f_i - c)^2}.$$

   Note that the eigenvalue μ_2 of the normalized Laplacian and the eigenvalue λ_2 of the random walk matrix P are in relation $1 - \lambda_2 = \mu_2$. ■

Proof of Theorem 10. Follows directly from plugging in the first part of Proposition 26 in Proposition 5. ■

Proof of Theorem 11. We split

$$\left| nR_{ij} - \frac{1}{p(X_i)} - \frac{1}{p(X_j)} \right| \le \left| nR_{ij} - \frac{n}{d_i} - \frac{n}{d_j} \right| + \left| \frac{n}{d_i} + \frac{n}{d_j} - \frac{1}{p(X_i)} - \frac{1}{p(X_j)} \right|.$$

Under the given assumption, the second term on the right hand side converges to 0 a.s. by a standard kernel density estimation argument (e.g., Section 9 of Devroye and Lugosi, 2001). The main work is the first term on the right hand side. We treat upper and lower bounds of $R_{ij} - 1/d_i - 1/d_j$ separately.

To get a lower bound, recall that by Proposition 15 we have

$$R_{ij} \ge \frac{Q_{ij}}{1 + w_{ij}Q_{ij}},$$

where $Q_{ij} = 1/(d_i - w_{ij}) + 1/(d_j - w_{ij})$ and w_{ij} is the weight of the edge between i and j. Observe that under the given conditions, for any two fixed points X_i, X_j the Gaussian edge weight w_{ij} converges to 0 (as $\sigma \to 0$). Thus

$$n\left(R_{ij} - \frac{1}{d_i} - \frac{1}{d_j}\right) \ge n\left(\frac{Q_{ij}}{1 + w_{ij}Q_{ij}} - \frac{1}{d_i} - \frac{1}{d_j}\right) \to 0 \ \ a.s.$$

To treat the upper bound, we define the ε-truncated Gauss graph G^ε as the graph with edge weights

$$w_{ij}^\varepsilon := \begin{cases} w_{ij} & \text{if } \|X_i - X_j\| \le \varepsilon, \\ 0 & \text{else.} \end{cases}$$

Let $d_i^\varepsilon = \sum_{j=1}^n w_{ij}^\varepsilon$. Because of $w_{ij}^\varepsilon \leq w_{ij}$ and Rayleigh's principle, we have $R_{ij} \leq R_{ij}^\varepsilon$, where R^ε denotes the resistance of the ε-truncated Gauss graph. Obviously,

$$
nR_{ij} - \left(\frac{n}{d_i} + \frac{n}{d_j}\right) \leq \left| nR_{ij}^\varepsilon - \left(\frac{n}{d_i} + \frac{n}{d_j}\right) \right|
$$

$$
\leq \underbrace{\left| nR_{ij}^\varepsilon - \left(\frac{n}{d_i^\varepsilon} + \frac{n}{d_j^\varepsilon}\right) \right|}_{(*)} + \underbrace{\left| \left(\frac{n}{d_i^\varepsilon} + \frac{n}{d_j^\varepsilon}\right) - \left(\frac{n}{d_i} + \frac{n}{d_j}\right) \right|}_{(**)}.
$$

To bound term $(**)$ we show that the degrees in the truncated graph converge to the ones in the non-truncated graph. To see this, note that

$$
\mathbb{E}\left(\frac{d_i^\varepsilon}{n} \,\Big|\, X_i\right) = \frac{1}{(2\pi)^{\frac{d}{2}}} \frac{1}{\sigma^d} \int_{B(X_i,\varepsilon)} e^{\frac{-\|X_i-y\|^2}{2\sigma^2}} p(y)\, dy
$$

$$
= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{B(0,\frac{\varepsilon}{\sigma})} e^{-\frac{\|z\|^2}{2}} p(X_i + \sigma z)\, dz
$$

$$
= \mathbb{E}\left(\frac{d_i}{n} \,\Big|\, X_i\right) - \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d \setminus B(0,\frac{\varepsilon}{\sigma})} e^{-\frac{\|z\|^2}{2}} p(X_i + \sigma z)\, dz.
$$

Exploiting that

$$
\frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d \setminus B(0,\frac{\varepsilon}{\sigma})} e^{-\frac{\|z\|^2}{2}} \leq \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{\varepsilon^2}{4\sigma^2}} \int_{\mathbb{R}^d} e^{-\frac{\|z\|^2}{4}}
$$

$$
\leq 2^{\frac{d}{2}} e^{-\frac{\varepsilon^2}{4\sigma^2}} = 2^{\frac{d}{2}} \frac{1}{(n\varepsilon^{d+2})^{\frac{1}{4}}}
$$

we obtain the convergence of the expectations: under the assumptions on n and σ from the theorem,

$$
\left| \mathbb{E}\left(\frac{d_i^\varepsilon}{n} \,\Big|\, X_i\right) - \mathbb{E}\left(\frac{d_i}{n} \,\Big|\, X_i\right) \right| \to 0.
$$

Now, a probabilistic bound for term $(**)$ can be obtained by standard concentration arguments.

We now bound term $(*)$. Up to now, our argument holds for arbitrary ε. For this last step, we now require that ε satisfies $\sigma^2 = \omega(\varepsilon^2 / \log(n\varepsilon^{d+2}))$. Note that for this choice of ε, the truncated Gaussian graph "converges" to the non-truncated graph, as we truncate less and less weight.

Denote by $\lambda^{\varepsilon,\text{weighted}}$ the eigenvalues of the ε-truncated Gauss graph, and by $w_{\min}^\varepsilon$, $w_{\max}^\varepsilon$ its minimal and maximal edge weights. Also consider the graph G'' that is the unweighted version of the ε-truncated Gauss graph G^ε. Note that G'' coincides with the standard ε-graph. We denote its eigenvalues by $\lambda^{\varepsilon,\text{unweighted}}$. By applying Proposition 5, Corollary 8

and Proposition 26 we get

$$\left| nR_{ij}^{\varepsilon} - \left(\frac{n}{d_i^{\varepsilon}} + \frac{n}{d_j^{\varepsilon}} \right) \right| \leq \frac{w_{\max}^{\varepsilon}}{d_{\min}^{\varepsilon}} \left(\frac{1}{1 - \lambda_2^{\varepsilon,\text{weighted}}} + 2 \right) \left(\frac{n}{d_i^{\varepsilon}} + \frac{n}{d_j^{\varepsilon}} \right)$$

$$\leq \frac{w_{\max}^{\varepsilon}}{d_{\min}^{\varepsilon}} \left(\frac{w_{\max}^{\varepsilon}}{w_{\min}^{\varepsilon}} \frac{1}{1 - \lambda_2^{\varepsilon,\text{unweighted}}} + 2 \right) \left(\frac{n}{d_i^{\varepsilon}} + \frac{n}{d_j^{\varepsilon}} \right) \qquad (6)$$

where the first inequality holds with probability at least $1 - c_1 n \exp(-c_2 n \sigma^d) - c_3 \exp(-c_4 n \sigma^d)/\sigma^d$. By $(**)$ we already know that the last factor in (6) converges to a constant:

$$\left(\frac{n}{d_i^{\varepsilon}} + \frac{n}{d_j^{\varepsilon}} \right) \to 1/p(X_i) + 1/p(X_j).$$

For the other factors of Term (6) we use the following quantities:

$$w_{\min}^{\varepsilon} \geq \frac{1}{\sigma^d} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$$

$$w_{\max}^{\varepsilon} \leq \frac{1}{\sigma^d}$$

$$d_{\min}^{\varepsilon} \geq n\varepsilon^d w_{\min}^{\varepsilon}$$

$$1 - \lambda_2^{\varepsilon,\text{unweighted}} \geq \varepsilon^2$$

Plugging these quantities in (6) and exploiting $\sigma^2 = \omega(\varepsilon^2/\log(n\varepsilon^{d+2}))$ leads to the desired convergence of $(*)$ to 0. ∎

Proof of Corollary 13. We use the result from Theorem 4 in Chung and Radcliffe (2011) which states that under the assumption that the minimal expected degree $\overline{d_{\min}}$ satisfies $\overline{d_{\min}}/\log(n) \to \infty$, then with probability at least $1 - 1/n$ the spectral gap is bounded by a term of the order $O(\log(2n)/\overline{d_{\min}})$. Plugging this in Part (2) of Proposition 5 shows that with high probability,

$$\frac{\left| \frac{1}{\text{vol}(G)} C_{ij} - \frac{1}{d_i} - \frac{1}{d_j} \right|}{\frac{1}{d_i} + \frac{1}{d_j}} s \leq \left(\frac{\overline{d_{\min}}}{\log(2n)} + 2 \right) \frac{1}{d_{\min}} = O\left(\frac{1}{\log(2n)} \right).$$

∎

Proof of Corollary 14. The expected degree of each vertex is $n(p_{\text{within}} + p_{\text{between}})/2$, the expected volume of the graph is $n^2(p_{\text{within}} + p_{\text{between}})/2$. The matrix $\overline{A}$ has the form $\begin{pmatrix} pJ & qJ \\ qJ & pJ \end{pmatrix}$ where J is the $(n/2 \times n/2)$-matrix of all ones and $p = p_{\text{within}}$ and $q = p_{\text{between}}$. The expected degree of all vertices is $n(p+q)/2$. Hence, $\overline{D^{-1/2}} \overline{A} \overline{D^{-1/2}} = \frac{2}{n(p+q)} \cdot \overline{A}$. This matrix has rank 2, its largest eigenvalue is 1 (with eigenvector the constant 1 vector), the other eigenvalue is $(p-q)/(p+q)$ with eigenvector $(1, ..., 1, -1, ..., -1)$. Hence, the spectral gap in this model is $2q/(p+q)$.

Under the assumption that $p = \omega(\log(n)/n)$, the deviations in Theorem 12 converge to 0. Plugging the spectral gap in our bound in Proposition 5 shows that with high probability,

$$\frac{n(p_{\text{within}} + p_{\text{between}})}{2} \cdot \left| \frac{1}{\text{vol}(G)} H_{ij} - \frac{1}{d_i} \right| \leq \frac{4}{np_{\text{between}}} + \frac{4}{n(p_{\text{within}} + p_{\text{between}})}$$
$$= O\left(\frac{1}{np_{\text{between}}} \right).$$

$\blacksquare$

8. Discussion

We have presented different strategies to prove that in many large graphs the commute distance can be approximated by $1/d_i + 1/d_j$. Both our approaches tell a similar story. Our result holds as soon as there are "enough disjoint paths" between i and j, compared to the size of the graph, and the minimal degree is "large enough" compared to n. Most relevant for machine learning, our results hold for all kinds of random geometric graphs (ε-graphs, kNN graphs, Gaussian similarity graphs). Here, the limit distance function $dist(i, j) = 1/d_i + 1/d_j$ is meaningless: It just considers the local density (the degree) at the two vertices, but does not take into account any global property such as the cluster structure of the graph. As the speed of convergence is very fast (for example, of the order $1/n$ in the case of Gaussian similarity graphs), the use of the raw commute distance should be discouraged even on moderate sized graphs.

An important point to note is that the results on the degeneracy of the hitting and commute times are not due to pathologies such as a "misconstruction" of the graphs or "wrong scaling constants". For example, in the random geometric graph setting the graph Laplacian can be proved to converge to the Laplace-Beltrami operator on the underlying space under similar assumptions as the ones above (Hein et al., 2007). But even though the Laplacian itself converges to a meaningful limit, the resistance distance, which is computed based on point evaluations of the inverse of this Laplacian, does not converge to a useful limit.

There are two important classes of graphs that are not covered in our approach. In power law graphs as well as in grid-like graphs, the minimal degree is constant, thus our results do not lead to tight bounds. The resistance distances on grid-like graphs has been studied in some particular cases. For example, Cserti (2000) and Wu (2004) prove explicit formulas for the resistance on regular one-and two-dimensional grids, and Benjamini and Rossignol (2008) characterize the variance of the resistance on random Bernoulli grids.

Beyond the commute distance, there exists a large variety of distance functions on graphs, and it is an interesting question to study their convergence behavior as $n \to \infty$. In particular, several authors constructed parametric families of distances that include both the shortest path distance and commute distance as special cases (e.g., Yen et al., 2008, Chebotarev, 2011, Alamgir and von Luxburg, 2011). For such families, the limit behavior is particularly interesting. On the one hand, it is well-known that shortest path distances

on random geometric graphs converge to the Euclidean or geodesic distances in the underlying space (Bernstein et al., 2000). On the other hand, as we have just seen in this paper, the commute distance converges to a meaningless distance function. Hence it is an interesting problem to characterize those distances that are degenerate. For the family of p-resistance distances this question has been solved in Alamgir and von Luxburg (2011). If $p > p^*$ for some critical threshold p^*, we are in a similar regime as the commute distance and the p-resistance is degenerate. For $p < p^*$ the commute distance is in a similar regime as the shortest path distance and is not degenerate. We believe that a solution to this question would be particularly interesting for the family of logarithmic forest distances by Chebotarev (2011). This family has many nice properties and, in our experience, tends to perform nicely in practice. So what are the parameters α for which we can guarantee that the distance is not degenerate as $n \to \infty$?

Acknowledgements

Appendix A. General Properties of Random Geometric Graphs

In this appendix we collect some basic results on random geometric graphs. These results are well-known, cf. Penrose (2003), but we did not find any reference where the material is presented in the way we need it (often the results are used implicitly or are tailored towards particular applications).

In the following, assume that $\mathcal{X} := \text{supp}(p)$ is a valid region according to Definition 1. Recall the definition of the boundary constant α in the valid region.

A convenient tool for dealing with random geometric graphs is the following well-known concentration inequality for binomial random variables with small p (originally found by Chernoff, 1952 and Hoeffding, 1963, we use the version as published in Angluin and Valiant, 1977).

Proposition 27 (Concentration inequalities) *Let N be a $Bin(n, p)$-distributed random variable. Then, for all $\delta \in\,]0, 1]$,*

$$
\begin{aligned}
P\Big(N \;\leq\; (1-\delta)np\Big) &\;\leq\; \exp(-\frac{1}{3}\delta^2 np) \\
P\Big(N \;\geq\; (1+\delta)np\Big) &\;\leq\; \exp(-\frac{1}{3}\delta^2 np).
\end{aligned}
$$

We will see below that computing expected, minimum and maximum degrees in random geometric graphs always boils down to counting the number of data points in certain balls

in the space. The following proposition is a straightforward application of the concentration inequality above and serves as "template" for all later proofs.

Proposition 28 (Counting sample points) *Consider a sample $X_1, \ldots, X_n$ drawn i.i.d. according to density p on $\mathcal{X}$. Let $B_1, \ldots, B_K$ be a fixed collection of subsets of $\mathcal{X}$ (the B_i do not need to be disjoint). Denote by $b_{\min} := \min_{i=1,\ldots,K} \int_{B_i} p(x)dx$ the minimal probability mass of the sets B_i (similarly by $b_{\max}$ the maximal probability mass), and by $N_{\min}$ and $N_{\max}$ the minimal (resp. maximal) number of sample points in the sets B_i. Then for all $\delta \in]0,1]$*

$$P\Big(N_{\max} \geq (1+\delta)nb_{\max}\Big) \;\leq\; K \cdot \exp(-\delta^2 nb_{\max}/3)$$
$$P\Big(N_{\min} \leq (1-\delta)nb_{\min}\Big) \;\leq\; K \cdot \exp(-\delta^2 nb_{\min}/3).$$

Proof. This is a straightforward application of Proposition 27 using the union bound. ∎

When working with ε-graphs or kNN-graphs, we often need to know the degrees of the vertices. As a rule of thumb, the expected degree of a vertex in the ε-graph is of the order $\Theta(n\varepsilon^d)$, the expected degree of a vertex in both the symmetric and mutual kNN-graph is of the order $\Theta(k)$. The expected kNN-distance is of the order $\Theta((k/n)^{1/d})$. Provided the graph is "sufficiently connected", all these rules of thumb also apply to the minimal and maximal values of these quantities. The following propositions make these rules of thumb explicit.

Proposition 29 (Degrees in the ε-graph) *Consider an ε-graph on a valid region $\mathcal{X} \subset \mathbb{R}^d$.*

1. *Then, for all $\delta \in]0,1]$, the minimal and maximal degrees in the ε-graph satisfy*

$$P\Big(d_{\max} \geq (1+\delta)n\varepsilon^d p_{\max}\eta_d\Big) \;\leq\; n \cdot \exp(-\delta^2 n\varepsilon^d p_{\max}\eta_d/3)$$
$$P\Big(d_{\min} \leq (1-\delta)n\varepsilon^d p_{\min}\eta_d\alpha\Big) \;\leq\; n \cdot \exp(-\delta^2 n\varepsilon^d p_{\min}\eta_d\alpha/3).$$

   In particular, if $n\varepsilon^d/\log n \to \infty$, then these probabilities converge to 0 as $n \to \infty$.

2. *If $n \to \infty, \varepsilon \to 0$ and $n\varepsilon^d/\log n \to \infty$, and the density p is continuous, then for each interior point $X_i \in \mathcal{X}$ the degree is a consistent density estimate: $d_i/(n\varepsilon^d\eta_d) \longrightarrow p(X_i)$ a.s.*

Proof. *Part 1* follows by applying Proposition 28 to the balls of radius ε centered at the data points. Note that for the bound on $d_{\min}$, we need to take into account boundary effects as only a part of the ε-ball around a boundary point is contained in $\mathcal{X}$. This is where the constant α comes in (recall the definition of α from the definition of a valid region). *Part 2* is a standard density estimation argument: the expected degree of X_i is the expected number of points in the ε-ball around X_i. For ε small enough, the ε-ball around X_i is completely contained in $\mathcal{X}$ and the density is approximately constant on this ball because we assumed the density to be continuous. The expected number of points is approximately $n\varepsilon^d\eta_d p(X_i)$ where η_d denotes the volume of a d-dimensional unit ball. The

result now follows from Part 1. ∎

Recall the definitions of the k-nearest neighbor radii: $R_k(x)$ denotes the distance of x to its k-nearest neighbor among the X_i, and the maximum and minimum values are denoted $R_{k,\max} := \max_{i=1,\dots,n} R_k(X_i)$ and $R_{k,\min} := \max_{i=1,\dots,n} R_k(X_i)$. Also recall the definition of the boundary constant α from the definition of a valid region.

Proposition 30 (Degrees in the kNN-graph) *Consider a valid region $\mathcal{X} \subset \mathbb{R}^d$.*

1. *Define the constants $a = 1/(2p_{\max}\eta_d)^{1/d}$ and $\tilde{a} := 2/(p_{\min}\eta_d\alpha)^{1/d}$. Then*

$$P\left(R_{k,\min} \leq a\left(\frac{k}{n}\right)^{1/d}\right) \leq n\exp(-k/3)$$

$$P\left(R_{k,\max} \geq \tilde{a}\left(\frac{k}{n}\right)^{1/d}\right) \leq n\exp(-k/12).$$

   If $n \to \infty$ and $k/\log n \to \infty$, then these probabilities converge to 0.

2. *Moreover, with probability at least $1 - n\exp(-c_4 k)$ the minimal and maximal degree in both the symmetric and mutual kNN-graph are of the order $\Theta(k)$ (the constants differ).*

3. *If the density is continuous, $n \to \infty$, $k/\log n \to \infty$ and additionally $k/n \to 0$, then in both the symmetric and the mutual kNN-graph, the degree of any fixed vertex v_i in the interior of $\mathcal{X}$ satisfies $k/d_i \to 1$ a.s.*

Proof. *Part 1.* Define the constant $a = 1/(2p_{\max}\eta_d)^{1/d}$ and the radius $r := a\,(k/n)^{1/d}$, fix a sample point x, and denote by $\mu(x)$ the probability mass of the ball around x with radius r. Set $\mu_{\max} := r^d\eta_d p_{\max} \geq \max_{x\in\mathcal{X}}\mu(x)$. Note that $\mu_{\max} < 1$. Observe that $R_k(x) \leq r$ if and only if there are at least k data points in the ball of radius r around x. Let $M \sim Bin(n,\mu)$ and $V \sim Bin(n,\mu_{\max})$. Note that by the choices of a and r we have $E(V) = k/2$. All this leads to

$$P\left(R_k(x) \leq r\right) \leq P\left(M \geq k\right) \leq P\left(V \geq k\right) = P\left(V \geq 2E(V)\right).$$

Applying the concentration inequality of Proposition 27 (with $\delta := 1$)) and using a union bound leads to the following result for the minimal kNN-radius:

$$P\left(R_{k,\min} \leq a\left(\frac{k}{n}\right)^{1/d}\right) \leq P\left(\exists i \,:\, R_k(X_i) \leq a\left(\frac{k}{n}\right)^{1/d}\right)$$

$$\leq n\max_{i=1,\dots,n}\left(R_k(X_i) \leq r\right)$$

$$\leq n\exp(-k/3).$$

By a similar approach we can prove the analogous statement for the maximal kNN-radius. Note that for the bound on $R_{k,\max}$ we additionally need to take into account boundary

effects: at the boundary of $\mathcal{X}$, only a part of the ball around a point is contained in $\mathcal{X}$, which affects the value of $\mu_{\min}$. We thus define $\tilde{a} := 2/(p_{\min}\eta_d\alpha)^{1/d}$, $r := \tilde{a}(k/n)^{1/d}$, $\mu_{\min} := r^d\eta_d p_{\min}\alpha$ where $\alpha \in]0,1]$ is the constant defined in the valid region. With $V = Bin(n, \mu_{\min})$ with $EV = 2k$ we continue similarly to above and get (using $\delta = 1/2$)

$$P\left(R_{k,\max} \geq \tilde{a}\left(\frac{k}{n}\right)^{1/d}\right) \quad \leq \quad n\exp(-k/12).$$

Part 2. In the directed kNN-graph, the degree of each vertex is exactly k. Thus, in the mutual kNN-graph, the maximum degree over all vertices is upper bounded by k, in the symmetric kNN-graph the minimum degree over all vertices is lower bounded by k.

For the symmetric graph, observe that the maximal degree in the graph is bounded by the maximal number of points in the balls of radius $R_{k,\max}$ centered at the data points. We know that with high probability, a ball of radius $R_{k,\max}$ contains of the order $\Theta(nR_{k,\max}^d)$ points. Using Part 1 we know that with high probability, $R_{k,\max}$ is of the order $(k/n)^{1/d}$. Thus the maximal degree in the symmetric kNN-graph is of the order $\Theta(k)$, with high probability.

In the mutual graph, observe that the minimal degree in the graph is bounded by the minimal number of points in the balls of radius $R_{k,\min}$ centered at the data points. Then the statement follows analogously to the last one.

Part 3, proof sketch. Consider a fixed point x in the interior of $\mathcal{X}$. We know that both in the symmetric and mutual kNN-graph, two points cannot be connected if their distance is larger than $R_{k,\max}$. As we know that $R_{k,\max}$ is of the order $(k/n)^{1/d}$, under the growth conditions on n and k this radius becomes arbitrarily small. Thus, because of the continuity of the density, if n is large enough we can assume that the density in the ball $B(x, R_{k,\max})$ of radius $R_{k,\max}$ around x is approximately constant. Thus, all points $y \in B(x, R_{k,\max})$ have approximately the same k-nearest neighbor radius $R := (k/(n \cdot p(x)\eta_d))^{1/d}$. Moreover, by concentration arguments it is easy to see that the actual kNN-radii only deviate by a factor $1 \pm \delta$ from their expected values.

Then, with high probability, all points inside of $B(x, R(1 - \delta))$ are among the k nearest neighbors of x, and all k nearest neighbors of x are inside $B(x, R(1+\delta))$. On the other hand, with high probability x is among the k nearest neighbors of all points $y \in B(x, R(1 - \delta))$, and not among the k nearest neighbors of any point outside of $B(x, R(1 + \delta))$. Hence, in the mutual kNN-graph, with high probability x is connected exactly to all points $y \in B(x, R(1 - \delta))$. In the symmetric kNN-graph, x might additionally be connected to the points in $B(x, R(1+\delta)) \setminus B(x, R(1-\delta))$. By construction, with high probability the number of sample points in these balls is $(1+\delta)k$ and $(1-\delta)k$. Driving δ to 0 leads to the result. ∎

References

M. Alamgir and U. von Luxburg. Phase transition in the familiy of p-resistances. In *Neural Information Processing Systems (NIPS)*, 2011.

D. Aldous and J. Fill. Reversible markov chains and random walks on graphs. *Preprint, available at http://www.stat.berkeley.edu/users/aldous/RWG/book.html*, 2001.

R. Aleliunas, R. Karp, R. Lipton, L. Lovász, and C. Rackoff. Random walks, universal traversal sequences, and the complexity of maze problems. In *20th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 218–223, 1979.

D. Angluin and L.G. Valiant. Fast probabilistic algorithms for Hamiltonian circuits and matchings. In *Symposium on the Theory of Computing (STOC)*, 1977.

C. Avin and G. Ercal. On the cover time and mixing time of random geometric graphs. *Theor. Comput. Sci*, 380(1-2):2–22, 2007.

P. Balister, B. Bollobás, A. Sarkar, and M. Walters. Connectivity of random k-nearest-neighbour graphs. *Advances in Applied Probability*, 37(1):1–24, 2005.

A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439): 509–512, 1999.

I. Benjamini and R. Rossignol. Submean variance bound for effective resistance of random electric networks. *Communications in Mathematical Physics*, 280(2):445–462, 2008.

M. Bernstein, V. De Silva, J. Langford, and J. Tenenbaum. Graph approximations to geodesics on embedded manifolds. Technical report, Department of Psychology, Stanford University, 2000.

B. Bollobas. *Modern Graph Theory*. Springer, 1998.

S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Mixing times for random walks on geometric random graphs. In C. Demetrescu, R. Sedgewick, and R.Tamassia, editors, *Proceedings of the Second Workshop on Analytic Algorithmics and Combinatorics, ANALCO*, pages 240–249, 2005.

M. Brand. A random walks perspective on maximizing satisfaction and profit. In *SIAM International Conference on Data Mining, April 21-23 (SDM)*, 2005.

M. Brito, E. Chavez, A. Quiroz, and J. Yukich. Connectivity of the mutual k-nearest-neighbor graph in clustering and outlier detection. *Statistics and Probability Letters*, 35: 33 – 42, 1997.

N. Cesa-Bianchi, C. Gentile, and F. Vitale. Fast and optimal prediction on a labeled tree. In *Conference on Learning Theory (COLT)*, 2009.

A. Chandra, P. Raghavan, W. Ruzzo, R. Smolensky, and P. Tiwari. The electrical resistance of a graph captures its commute and cover times. In *Proceedings of the 21st Annual Symposium on Theory of Computing (STOC)*, pages 574–586, 1989.

P. Chebotarev. A class of graph-geodetic distances generalizing the shortest path and the resistance distances. *Discrete Applied Mathematics*, 159(295 – 302), 2011.

H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 1952.

F. Chung. *Spectral graph theory*, volume 92 of the CBMS Regional Conference Series in Mathematics. Conference Board of the Mathematical Sciences, Washington, 1997.

F. Chung and L. Lu. *Complex Graphs and Networks*. American Mathematical Society, Providence, 2006.

F. Chung and M. Radcliffe. On the spectra of general random graphs. *Electronic Journal of Combinatorics*, 18(P215):1, 2011.

C. Cooper and A. Frieze. The cover time of sparse random graphs. In *Proceedings of the fourteenth Annual Symposium on Discrete Algorithms (SODA)*, pages 140–147, 2003.

C. Cooper and A. Frieze. The cover time of random regular graphs. *SIAM Journal on Discrete Mathematics*, 18(4):728–740, 2005.

C. Cooper and A. Frieze. The cover time of the preferential attachment graph. *J. Comb. Theory, Ser. B*, 97(2):269–290, 2007.

C. Cooper and A. Frieze. The cover time of random geometric graphs. *Random Structures & Algorithms*, 38(3):324–349, 2011.

J. Cserti. Application of the lattice Green's function for calculating the resistance of an infinite network of resistors. *American Journal of Physics*, 68:896, 2000.

L. Devroye and G. Lugosi. *Combinatorial Methods in Density Estimation*. Springer, New York, 2001.

P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *The Annals of Applied Probability*, 3(3):696–730, 1993.

P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *The Annals of Applied Probability*, pages 36–61, 1991.

P. G. Doyle and J. L. Snell. *Random walks and electric networks*, volume 22 of *Carus Mathematical Monographs*. Mathematical Association of America, Washington, DC, 1984.

F. Fouss, A. Pirotte, J.-M. Renders, and M. Saerens. A novel way of computing dissimilarities between nodes of a graph, with application to collaborative filtering and subspace projection of the graph nodes. Technical Report IAG WP 06/08, Université catholique de Louvain, 2006.

P.W. Fowler. Resistance distances in fullerene graphs. *Croatica Chemica Acta*, 75(2):401–408, 2002.

S. Guattery. Graph embeddings, symmetric real matrices, and generalized inverses. Technical report, Institute for Computer Applications in Science and Engineering, NASA Langley Research Center, 1998.

G. Guillot, R. Leblois, A. Coulon, and A.C. Frantz. Statistical methods in spatial genetics. *Molecular Ecology*, 18(23):4734–4756, 2009.

J. Ham, D. D. Lee, S. Mika, and B. Schölkopf. A kernel view of the dimensionality reduction of manifolds. In *International Conference on Machine Learning (ICML)*, pages 369–376. ACM, 2004.

M. Hein, J.-Y. Audibert, and U. von Luxburg. Graph Laplacians and their convergence on random neighborhood graphs. *Journal of Machine Learning Research*, 8:1325 – 1370, 2007.

M. Herbster and M. Pontil. Prediction on a graph with a perceptron. In Bernhard Schölkopf, John C. Platt, and Thomas Hoffman, editors, *Neural Information Processing Systems (NIPS)*, pages 577–584, 2006.

W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13 – 30, 1963.

O. Ivanciuc. QSAR and QSPR molecular descriptors computed from the resistance distance and electrical conductance matrices. *ACH Models in Chemistry*, 137(5/6):607–632, 2000.

M. Jerrum and A. Sinclair. Conductance and the rapid mixing property for markov chains: the approximation of permanent resolved. In *Proceedings of the twentieth annual ACM symposium on Theory of computing (STOC)*, pages 235–244, 1988.

D. Klein and M. Randic. Resistance distance. *Journal of Mathematical Chemistry*, 12: 81 – 95, 1993.

D. Levin, Y. Peres, and E. Wilmer. *Markov Chains and Mixing Times.* Americal Mathematical Society, 2008.

D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In *International Conference on Information and Knowledge Management (CIKM)*, pages 556–559, 2003.

L. Lovász. Random walks on graphs: a survey. In *Combinatorics, Paul Erdös is eighty*, Bolyai Soc. Math. Stud., pages 353 – 397. János Bolyai Math. Soc., Budapest, 1993.

R. Lyons and Y. Peres. Probability on trees and networks. *Book in preparation, available online on the webpage of Yuval Peres*, 2010.

M. Penrose. The longest edge of the random minimal spanning tree. *The Annals of Applied Probability*, pages 340–361, 1997.

M. Penrose. A strong law for the longest edge of the minimal spanning tree. *Ann. of Prob.*, 27(1):246 – 260, 1999.

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

H. Qiu and E. Hancock. Graph embedding using commute time. *Structural, Syntactic, and Statistical Pattern Recognition*, pages 441–449, 2006.

H. Qiu and E.R. Hancock. Image segmentation using commute times. In *Proceedings of the 16th British Machine Vision Conference (BMVC)*, pages 929–938, 2005.

K. Roy. Topological descriptors in drug design and modeling studies. *Molecular Diversity*, 8(4):321–323, 2004. ISSN 1381-1991.

M. Saerens, F. Fouss, L. Yen, and P. Dupont. The principal components analysis of a graph, and its relationships to spectral clustering. In *Proceedings of the 15th European Conference on Machine Learning (ECML)*, pages 371 – 383, 2004.

P. Sarkar and A. Moore. A tractable approach to finding closest truncated-commute-time neighbors in large graphs. In *23rd Conference on Uncertainty in Artificial Intelligence(UAI)*, 2007.

P. Sarkar, A. Moore, and A. Prakash. Fast incremental proximity search in large graphs. In W. Cohen, A. McCallum, and S. Roweis, editors, *Proceedings of the Twenty-Fifth International Conference of Machine Learning (ICML)*, pages 896–903, 2008.

E. Seneta. *Non-negative Matrices and Markov Chains*. Springer, 2006.

A. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1(04):351–370, 1992.

D. Spielman and N. Srivastava. Graph sparsification by effective resistances. In R. Ladner and C. Dwork, editors, *Proceedings of the 40th Annual Symposium on Theory of Computing (STOC)*, pages 563–568, 2008.

P. Tetali. Random walks and the effective resistance of networks. *Journal of Theoretical Probability*, 4(1):101–109, 1991.

U. von Luxburg, A. Radl, and M. Hein. Getting lost in space: Large sample analysis of the commute distance. In *Neural Information Processing Systems (NIPS)*, 2010.

D. Wittmann, D. Schmidl, F. Blöchl, and F. Theis. Reconstruction of graphs based on random walks. *Theoretical Computer Science*, 410(38-40):3826–3838, 2009.

F. Y. Wu. Theory of resistor networks: The two-point resistance. *Journal of Physics A: Mathematical and General*, 37:6653–6673, 2004.

W. Xiao and I. Gutman. Resistance distance and Laplacian spectrum. *Theoretical Chemistry Accounts*, 110:284 – 298, 2003.

F. Xue and P. Kumar. The number of neighbors needed for connectivity of wireless networks. *Wireless networks*, 10(2):169–181, 2004.

L. Yen, D. Vanvyve, F. Wouters, F. Fouss, M. Verleysen, and M. Saerens. Clustering using a random walk based distance measure. In *Proceedings of the 13th Symposium on Artificial Neural Networks (ESANN)*, pages 317–324, 2005.

L. Yen, M. Saerens, A. Mantrach, and M. Shimbo. A family of dissimilarity measures between nodes generalizing both the shortest-path and the commute-time distances. In *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–793, 2008.

C. Zenger. A comparison of some bounds for the nontrivial eigenvalues of stochastic matrices. *Numer. Math.*, 9:209 – 22, 1972.

D. Zhou and B. Schölkopf. Learning from labeled and unlabeled data using random walks. In *Pattern Recognition, Proceedings of the 26th DAGM Symposium*, pages 237 – 244, 2004.