ESAIM: PS 17 (2013) 370–418
DOI: 10.1051/ps/2012001

ESAIM: Probability and Statistics
www.esaim-ps.org

HOW THE RESULT OF GRAPH CLUSTERING METHODS DEPENDS ON THE CONSTRUCTION OF THE GRAPH

Markus Maier[1], Ulrike von Luxburg[1,2] and Matthias Hein[3]

Abstract. We study the scenario of graph-based clustering algorithms such as spectral clustering. Given a set of data points, one first has to construct a graph on the data points and then apply a graph clustering algorithm to find a suitable partition of the graph. Our main question is if and how the construction of the graph (choice of the graph, choice of parameters, choice of weights) influences the outcome of the final clustering result. To this end we study the convergence of cluster quality measures such as the normalized cut or the Cheeger cut on various kinds of random geometric graphs as the sample size tends to infinity. It turns out that the limit values of the same objective function are systematically different on different types of graphs. This implies that clustering results systematically depend on the graph and can be very different for different types of graph. We provide examples to illustrate the implications on spectral clustering.

Mathematics Subject Classification. 62G20, 05C80, 68Q87.

Received November 23, 2010. Revised December 21, 2011.

1. Introduction

Nowadays it is very popular to represent and analyze statistical data using random graph or network models. The vertices in such a graph correspond to data points, whereas edges in the graph indicate that the adjacent vertices are "similar" or "related" to each other. In this paper we consider the problem of data clustering in a random geometric graph setting. We are given a sample of points drawn from some underlying probability distribution on a metric space. The goal is to cluster the sample points into "meaningful groups". A standard procedure is to first transform the data to a neighborhood graph, for example a k-nearest neighbor graph. In a second step, the cluster structure is then extracted from the graph: clusters correspond to regions in the graph that are tightly connected within themselves and only sparsely connected to other clusters.

There already exist a couple of papers that study statistical properties of this procedure in a particular setting: when the true underlying clusters are defined to be the connected components of a density level set in the underlying space. In his setting, a test for detecting cluster structure and outliers is proposed in Brito *et al.* [3]. In Biau *et al.* [2] the authors build a neighborhood graph in such a way that its connected components

Keywords and phrases. Random geometric graph, clustering, graph cuts.

[1] Max Planck Institute for Intelligent Systems, Spemannstr. 38, 72076 Tübingen, Germany. mmaier@tuebingen.mpg.de

[2] Department of Computer Science, University of Hamburg, Vogt-Kölln-Str. 30, 22527 Hamburg, Germany. luxburg@informatik.uni-hamburg.de

[3] Faculty of Mathematics and Computer Science, Saarland University, Postfach 151150, 66041, Saarbrücken, Germany. hein@cs.uni-sb.de

Article published by EDP Sciences

converge to the underlying true clusters in the data. Maier *et al.* [8] compare the properties of different random graph models for identifying clusters of the density level sets.

While the definition of clusters as connected components of level sets is appealing from a theoretical point of view, the corresponding algorithms are often too simplistic and only moderately successful in practice. From a practical point of view, clustering methods based on graph partitioning algorithms are more robust. Clusters do not have to be perfectly disconnected in the graph, but are allowed to have a small number of connecting edges between them. Graph partitioning methods are widely used in practice. The most prominent algorithm in this class is spectral clustering, which optimizes the normalized cut (NCut) objective function (see below for exact definitions, and von Luxburg [13] for a tutorial on spectral clustering). It is already known under what circumstances spectral clustering is statistically consistent [14]. However, there is one important open question. When applying graph-based methods to given sets of data points, one obviously has to build a graph first, and there are several important choices to be made: the type of the graph (for example, k-nearest neighbor graph, the r-neighborhood graph or a Gaussian similarity graph), the connectivity parameter (k or r or σ, respectively) and the weights of the graph. Making such choices is not so difficult in the domain of supervised learning, where parameters can be set using cross-validation. However, it poses a serious problem in unsupervised learning. While different researchers use different heuristics and their "gut feeling" to set these parameters, neither systematic empirical studies have been conducted (for example, how sensitive the results are to the choice of graph parameters), nor do theoretical results exist which lead to well-justified heuristics.

In this paper we study the question if and how the results of graph-based clustering algorithms are affected by the graph type and the parameters that are chosen for the construction of the neighborhood graph. We focus on the case where the best clustering is defined as the partition that minimizes the normalized cut (Ncut) or the Cheeger cut.

Our theoretical setup is as follows. In a first step we ignore the problem of actually *finding* the optimal partition. Instead we fix some partition of the underlying space and consider it as the "true" partition. For any finite set of points drawn from the underlying space we consider the clustering of the points that is induced by this underlying partition. Then we study the convergence of the NCut value of this clustering as the sample size tends to infinity. We investigate this question on different kinds of neighborhood graphs. Our first main result is that depending on the type of graph, the clustering quality measure converges to different limit values. For example, depending on whether we use the kNN graph or the r-graph, the limit functional integrates over different powers of the density. From a statistical point of view, this is very surprising because in many other respects, the kNN graph and the r-graph behave very similar to each other. Just consider the related problem of density estimation. Here, both the k-nearest neighbor density estimate and the estimate based on the degrees in the r-graph converge to the same limit, namely the true underlying density (*cf.* Loftsgaarden and Quesenberry [7] for the consistency of the kNN density estimate). So it is far from obvious that the NCut values would converge to different limits.

In a second step we then relate these results to the setting where we optimize over all partitions to find the one that minimizes the NCut. We can show that the results from the first part can lead to the effect that the minimizer of NCut on the kNN graph is different from the minimizer of NCut on the r-graph or on the complete graph with Gaussian weights. This effect can also be studied in practical examples. First, we give examples of well-clustered distributions (mixtures of Gaussians) where the optimal limit cut on the kNN graph is different from the one on the r-neighborhood graph. The optimal limit cuts in these examples can be computed analytically. Next we can demonstrate that this effect can already been observed on finite samples from these distributions. Given a finite sample, running normalized spectral clustering to optimize Ncut leads to systematically different results on the kNN graph than on the r-graph. This shows that our results are not only of theoretical interest, but that they are highly relevant in practice.

In the following section we formally define the graph clustering quality measures and the neighborhood graph types we consider in this paper. Furthermore, we introduce the notation and technical assumptions for the rest of the paper. In Section 3 we present our main results on the convergence of NCut and the CheegerCut on different graphs. In Section 4 we show that our findings are not only of theoretical interest, but that they also influence concrete algorithms such as spectral clustering in practice. All proofs are deferred to Section 6. Note that a small part of the results of this paper has already been published in Maier *et al.* [9].

2. Definitions and assumptions

Given a directed graph $G = (V, E)$ with weights $w : E \to \mathbb{R}$ and a partition of the nodes V into $(U, V \setminus U)$ we define

$$\mathrm{cut}(U, V \setminus U) = \sum_{u \in U, v \in V \setminus U} (w(u,v) + w(v,u)),$$

and $\mathrm{vol}(U) = \sum_{u \in U, v \in V} w(u,v)$. If G is an undirected graph we replace the ordered pair (u,v) in the sums by the unordered pair $\{u,v\}$. Note that by doing so we count each edge twice in the undirected graph. This introduces a constant of two in the limits but it has the advantage that there is no need to distinguish in the formulation of our results between directed and undirected graphs.

Intuitively, the cut measures how strong the connection between the different clusters in the clustering is, whereas the volume of a subset of the nodes measures the "weight" of the subset in terms of the edges that originate in it. An ideal clustering would have a low cut and balanced clusters, that is clusters with similar volume. The graph clustering quality measures that we use in this paper, the normalized cut and the Cheeger cut, formalize this trade-off in slightly different ways: the normalized cut is defined by

$$\mathrm{NCut}(U, V \setminus U) = \mathrm{cut}(U, V \setminus U)\left(\frac{1}{\mathrm{vol}(U)} + \frac{1}{\mathrm{vol}(V \setminus U)}\right), \tag{2.1}$$

whereas the Cheeger cut is defined by

$$\mathrm{CheegerCut}(U, V \setminus U) = \frac{\mathrm{cut}(U, V \setminus U)}{\min\{\mathrm{vol}(U), \mathrm{vol}(V \setminus U)\}}\cdot \tag{2.2}$$

These definitions are useful for general weighted graphs and general partitions. As was said in the beginning we want to study the values of NCut and CheegerCut on neighborhood graphs on sample points in Euclidean space and for partitions of the nodes that are induced by a hyperplane S in $\mathbb{R}^d$. The two halfspaces belonging to S are denoted by H^+ and H^-. Having a neighborhood graph on the sample points $\{x_1, \ldots, x_n\}$, the partition of the nodes induced by S is $(\{x_1, \ldots, x_n\} \cap H^+, \{x_1, \ldots, x_n\} \cap H^-)$. In the rest of this paper for a given neighborhood graph G_n we set $\mathrm{cut}_n = \mathrm{cut}(\{x_1, \ldots, x_n\} \cap H^+, \{x_1, \ldots, x_n\} \cap H^-)$. Similarly, for $H = H^+$ or $H = H^-$ we set $\mathrm{vol}_n(H) = \mathrm{vol}(\{x_1, \ldots, x_n\} \cap H^+)$. Accordingly we define NCut_n and $\mathrm{CheegerCut}_n$.

In the following we introduce the different types of neighborhood graphs and weighting schemes that are considered in this paper. The graph types are:

- *the k-nearest neighbor* (kNN) *graphs*, where the idea is to connect each point to its k nearest neighbors. However, this yields a directed graph, since the k-nearest neighbor relationship is not symmetric. If we want to construct an undirected kNN graph we can choose between the mutual kNN graph, where there is an edge between two points if both points are among the k nearest neighbors of the other one, and the symmetric kNN graph, where there is an edge between two points if only one point is among the k nearest neighbors of the other one. In our proofs for the limit expressions it will become clear that these do not differ between the different types of kNN graphs. Therefore, we do not distinguish between them in the statement of the theorems, but rather speak of "the kNN graph";
- *the r-neighborhood graph*, where a radius r is fixed and two points are connected if their distance does not exceed the threshold radius r. Note that due to the symmetry of the distance we do not have to distinguish between directed and undirected graphs;

- *the complete weighted graph*, where there is an edge between each pair of distinct nodes (but no loops). Of course, in general we would not consider this graph a neighborhood graph. However, if the weight function is chosen in such a way that the weights of edges between nearby nodes are high and the weights between points far away from each other are almost negligible, then the behavior of this graph should be similar to that of a neighborhood graph. One such weight function is the Gaussian weight function, which we introduce below.

The weights that are used on neighborhood graphs usually depend on the distance of the end nodes of the edge and are non-increasing. That is, the weight $w(x_i, x_j)$ of an edge (x_i, x_j) is given by $w(x_i, x_j) = f(\mathrm{dist}(x_i, x_j))$ with a non-increasing weight function f. The weight functions we consider here are the *unit weight function* $f \equiv 1$, which results in the unweighted graph, and the *Gaussian weight function*

$$f(u) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{1}{2}\frac{u^2}{\sigma^2}\right)$$

with a parameter $\sigma > 0$ defining the bandwidth.

Of course, not every weighting scheme is suitable for every graph type. For example, as mentioned above, we would hardly consider the complete graph with unit weights a neighborhood graph. Therefore, we only consider the Gaussian weight function for this graph. On the other hand, for the kNN graph and the r-neighborhood graph with Gaussian weights there are two "mechanisms" that reduce the influence of far-away nodes: first the fact that far-away nodes are not connected to each other by an edge and second the decay of the weight function. In fact, it turns out that the limit expressions we study depend on the interplay between these two mechanisms. Clearly, the decay of the weight function is governed by the parameter σ. For the r-neighborhood graph the radius r limits the length of the edges. Asymptotically, given sequences $(\sigma_n)_{n\in\mathbb{N}}$ and $(r_n)_{n\in\mathbb{N}}$ of bandwidths and radii we distinguish between the following two cases:

- *the bandwidth σ_n is dominated by the radius r_n*, that is $\sigma_n/r_n \to 0$ for $n \to \infty$;
- *the radius r_n is dominated by the bandwidth σ_n*, that is $r_n/\sigma_n \to 0$ for $n \to \infty$.

For the kNN graph we cannot give a radius up to which points are connected by an edge, since this radius for each point is a random variable that depends on the positions of all the sample points. However, it is possible to show that for a point in a region of constant density p the k_n-nearest neighbor radius is concentrated around $\sqrt[d]{k_n/((n-1)\eta_d p)}$, where η_d denotes the volume of the unit ball in Euclidean space $\mathbb{R}^d$. This is plausible, considering that, by a standard result on density estimation, $k_n/((n-1)\eta_d \hat{r}^d)$, where $\hat{r}$ is the empirical k_n-nearest neighbor radius, is an estimate of the density at the point. That is, the kNN radius decays to zero with the rate $\sqrt[d]{k_n/n}$. In the following it is convenient to set for the kNN graph $r_n = \sqrt[d]{k_n/n}$, noting that this is not the k-nearest neighbor radius of any point but only its decay rate. Using this "radius" we distinguish between the same two cases of the ratio of r_n and σ_n as for the r-neighborhood graph.

For the sequences $(r_n)_{n\in\mathbb{N}}$ and $(\sigma_n)_{n\in\mathbb{N}}$ we always assume $r_n \to 0$, $\sigma_n \to 0$ and $nr_n \to \infty$, $n\sigma_n \to \infty$ for $n \to \infty$. Furthermore, for the parameter sequence $(k_n)_{n\in\mathbb{N}}$ of the kNN graph we always assume $k_n/n \to 0$, which corresponds to $r_n \to 0$, and $k_n/\log n \to \infty$.

In the rest of this paper we denote by $\mathcal{L}_d$ the Lebesgue measure in $\mathbb{R}^d$. Furthermore, let $B(x, r)$ denote the closed ball of radius r around x and $\eta_d = \mathcal{L}_d(B(0, 1))$, where we set $\eta_0 = 1$.

We make the following **general assumptions in the whole paper:**

- *the data points $x_1, \ldots, x_n$ are drawn independently from some density p on $\mathbb{R}^d$. The measure on $\mathbb{R}^d$ that is induced by p is denoted by μ; that means, for a measurable set $A \subseteq \mathbb{R}^d$ we set $\mu(A) = \int_A p(x)\,\mathrm{d}x$;*
- *the density p is bounded from below and above, that is $0 < p_{\min} \le p(x) \le p_{\max}$. In particular, it has compact support C;*
- *in the interior of C, the density p is twice differentiable and $\|\nabla p(x)\| \le p'_{\max}$ for a $p'_{\max} \in \mathbb{R}$ and all x in the interior of C;*

- *the cut hyperplane S splits the space $\mathbb{R}^d$ into two halfspaces H^+ and H^- (both including the hyperplane S) with positive probability masses, that is $\mu(H^+) > 0$, $\mu(H^-) > 0$. The normal of S pointing towards H^+ is denoted by n_S;*
- *if $d \geq 2$ the boundary ∂C is a compact, smooth $(d-1)$-dimensional surface with minimal curvature radius $\kappa > 0$, that is the absolute values of the principal curvatures are bounded by $1/\kappa$. We denote by n_x the normal to the surface ∂C at the point $x \in \partial C$. Furthermore, we can find constants $\gamma > 0$ and $r_\gamma > 0$ such that for all $r \leq r_\gamma$ we have $\mathcal{L}_d(B(x,r) \cap C) \geq \gamma \mathcal{L}_d(B(x,r))$ for all $x \in C$;*
- *if $d \geq 2$ we can find an angle $\alpha \in (0, \pi/2)$ such that $|\langle n_S, n_x \rangle| \leq \cos\alpha$ for all $x \in S \cap \partial C$. If $d = 1$ we assume that (the point) S is in the interior of C.*

The assumptions on the lower and upper bounds of the density are necessary to find lower and upper bounds on the k-nearest neighbor radii of points in C. The assumptions on differentiability are used to show concentration of the k-nearest neighbor radius for points in the "interior" of C, that is, points not within a boundary strip of C.

The assumptions on the boundary ∂C are necessary in order to bound the influence of points that are close to the boundary. The problem with these points is that the density is not approximately uniform inside small balls around them. Therefore, we cannot find a good estimate of their kNN radius and on their contribution to the cut and the volume. Furthermore, in the case of the r-neighborhood graph we cannot control the number of edges originating in these points. Under the assumptions above we can neglect these points.

The last assumption on the minimum angle between the normal of S and the normal of ∂C in $S \cap \partial C$ is used to ensure, that the intersection of the "boundary strip" of C, where we cannot control the kNN radii and the number of edges to other points, and S converges to zero sufficiently fast. Therefore, we can find a bound on the influence of the boundary points in our results and how fast this influence vanishes.

Appendix A contains a table of the notation used throughout the paper.

3. Main results: limits of the quality measures NCut and CheegerCut

As we can see in equations (2.1) and (2.2) the definitions of NCut and CheegerCut rely on the cut and the volume. Therefore, in order to study the convergence of NCut and CheegerCut it seems reasonable to study the convergence of the cut and the volume first. In Section 6 Corollaries 6.4–6.6 and Corollaries 6.9–6.11 state the convergence of the cut and the volume on the kNN graphs. Corollaries 6.13–6.16 state the convergence of the cut on the r-graph and the complete weighted graph, whereas Corollaries 6.18–6.21 state the convergence of the volume on the same graphs.

These corollaries show that there are scaling sequences $(s_n^{\mathrm{cut}})_{n\in\mathbb{N}}$ and $(s_n^{\mathrm{vol}})_{n\in\mathbb{N}}$ that depend on n, r_n and the graph type such that, under certain conditions, almost surely

$$\left(s_n^{\mathrm{cut}}\right)^{-1} \mathrm{cut}_n \to CutLim \qquad \text{and} \qquad \left(s_n^{\mathrm{vol}}\right)^{-1} \mathrm{vol}_n(H) \to VolLim(H)$$

for $n \to \infty$, where $CutLim \in \mathbb{R}_{\geq 0}$ and $VolLim(H^+), VolLim(H^-) \in \mathbb{R}_{>0}$ are constants depending only on the density p and the hyperplane S.

Having defined these limits we define, analogously to the definitions in Equations (2.1) and (2.2), the limits of NCut and CheegerCut as

$$NCutLim = \frac{CutLim}{VolLim\,(H^+)} + \frac{CutLim}{VolLim\,(H^-)} \tag{3.1}$$

and

$$CheegerCutLim = \frac{CutLim}{\min\left\{VolLim\,(H^+)\,, VolLim\,(H^-)\right\}}\cdot \tag{3.2}$$

In our following main theorems we show the conditions under which we have for $n \to \infty$ almost sure convergence of

$$\frac{s_n^{\mathrm{vol}}}{s_n^{\mathrm{cut}}}\,\mathrm{NCut}_n \to NCutLim \qquad \text{and} \qquad \frac{s_n^{\mathrm{vol}}}{s_n^{\mathrm{cut}}}\,\mathrm{CheegerCut}_n \to CheegerCutLim.$$

Furthermore, for the unweighted r-graph and kNN-graph and for the complete weighted graph with Gaussian weights we state the optimal convergence rates, where "optimal" means the best trade-off between our bounds for different quantities derived in Section 6. Note that we will not prove the following theorems here. Rather the Proof of Theorem 3.1 can be found in Section 6.2.4, whereas the Proofs of Theorems 3.2 and 3.3 can be found in Section 6.3.3.

Theorem 3.1 (NCut and CheegerCut on the kNN-graph)**.** *For a sequence $(k_n)_{n\in\mathbb{N}}$ with $k_n/n \to 0$ for $n \to \infty$ let G_n be the k_n-nearest neighbor graph on the sample $x_1, \ldots, x_n$. Set* $\mathrm{XCut} = \mathrm{NCut}$ *or* $\mathrm{XCut} = \mathrm{CheegerCut}$ *and let $XCutLim$ denote the corresponding limit as defined in equations* (3.1) *and* (3.2)*. Set*

$$\varDelta_n = \left|\frac{s_n^{\mathrm{vol}}}{s_n^{\mathrm{cut}}}\,\mathrm{XCut}_n - XCutLim\right|.$$

- *Let G_n be the unweighted* kNN *graph. If $k_n/\sqrt{n\log n} \to \infty$ in the case $d = 1$ and $k_n/\log n \to \infty$ in the case $d \geq 2$ we have $\varDelta_n \to 0$ for $n \to \infty$ almost surely. The optimal convergence rate is achieved for $k_n = k_0\sqrt[4]{n^3\log n}$ in the case $d = 1$ and $k_n = k_0 n^{2/(d+2)}(\log n)^{d/(d+2)}$ in the case $d \geq 2$. For this choice of k_n we have $\varDelta_n = O(\sqrt[d+4]{\log n/n})$ in the case $d = 1$ and $\varDelta_n = O(\sqrt[d+2]{\log n/n})$ for $d \geq 2$;*
- *let G_n be the* kNN*-graph with Gaussian weights and suppose $r_n \geq \sigma_n^\alpha$ for an $\alpha \in (0,1)$. Then we have almost sure convergence of $\varDelta_n \to 0$ for $n \to \infty$ if $k_n/\log n \to \infty$ and $n\sigma_n^{d+1}/\log n \to \infty$;*
- *let G_n be the* kNN*-graph with Gaussian weights and $r_n/\sigma_n \to 0$. Then we have almost sure convergence of $\varDelta_n \to 0$ for $n \to \infty$ if $k_n/\sqrt{n\log n} \to \infty$ in the case $d = 1$ and $k_n/\log n \to \infty$ in the case $d \geq 2$.*

Theorem 3.2 (NCut and CheegerCut on the r-graph)**.** *For a sequence $(r_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}_{>0}$ with $r_n \to 0$ for $n \to \infty$ let G_n be the r_n-neighborhood graph on the sample $x_1, \ldots, x_n$. Set* $\mathrm{XCut} = \mathrm{NCut}$ *or* $\mathrm{XCut} = \mathrm{CheegerCut}$ *and let $XCutLim$ denote the corresponding limit as defined in equations* (3.1) *and* (3.2)*. Set*

$$\varDelta_n = \left|\frac{s_n^{\mathrm{vol}}}{s_n^{\mathrm{cut}}}\,\mathrm{XCut}_n - XCutLim\right|.$$

- *Let G_n be unweighted. Then $\varDelta_n \to 0$ almost surely for $n \to \infty$ if $nr_n^{d+1}/\log n \to \infty$. The optimal convergence rate is achieved for $r_n = r_0\sqrt[d+3]{\log n/n}$ for a suitable constant $r_0 > 0$. For this choice of r_n we have $\varDelta_n = O(\sqrt[d+3]{\log n/n})$;*
- *let G_n be weighted with Gaussian weights with bandwidth $\sigma_n \to 0$ and $r_n/\sigma_n \to \infty$ for $n \to \infty$. Then $\varDelta_n \to 0$ almost surely for $n \to \infty$ if $n\sigma_n^{d+1}/\log n \to \infty$;*
- *let G_n be weighted with Gaussian weights with bandwidth $\sigma_n \to 0$ and $r_n/\sigma_n \to 0$ for $n \to \infty$. Then $\varDelta_n \to 0$ almost surely for $n \to \infty$ if $nr_n^{d+1}/\log n \to \infty$.*

The following theorem presents the limit results for NCut and CheegerCut on the complete weighted graph. One result that we need in the proof of this theorem is Corollary 6.14 on the convergence of the cut. Note that in Narayanan *et al.* [11] a similar cut convergence problem is studied for the case of the complete weighted graph, and the scaling sequence and the limit differ from ours. However, the reason is that in that paper the *weighted* cut is considered, which can be written as $f' L_{\mathrm{norm}} f$, where L_{norm} denotes the normalized graph Laplacian matrix and f is an n-dimensional vector with $f_i = 1$ if x_i is in one cluster and $f_i = 0$ if x_i is in the other cluster.

TABLE 1. The scaling sequences and limit expression for the cut and the volume in all the considered graph types. In the limit expression for the cut the integral denotes the $(d-1)$-dimensional surface integral along the hyperplane S, whereas in the limit expressions for the volume the integral denotes the Lebesgue integral over the halfspace $H = H^+$ or $H = H^-$.

The cut in the kNN-graph and the r-graph			
Weighting	s_n^{cut}	$CutLim$ kNN-graph	$CutLim$ r-graph
Unweighted	$n^2 r_n^{d+1}$	$\frac{2\eta_{d-1}}{(d+1)\eta_d^{1+1/d}} \int_S p^{1-1/d}(s)\ \mathrm{d}s$	$\frac{2\eta_{d-1}}{d+1} \int_S p^2(s)\ \mathrm{d}s$
Weighted $r_n/\sigma_n \to \infty$	$n^2\sigma_n$	$\frac{2}{\sqrt{2\pi}} \int_S p^2(s)\ \mathrm{d}s$	$\frac{2}{\sqrt{2\pi}} \int_S p^2(s)\ \mathrm{d}s$
Weighted $r_n/\sigma_n \to 0$	$\sigma_n^{-d} n^2 r_n^{d+1}$	$\frac{2\eta_{d-1}\eta_d^{-1-1/d}}{(d+1)(2\pi)^{d/2}} \int_S p^{1-1/d}(s)\ \mathrm{d}s$	$\frac{2\eta_{d-1}}{(d+1)(2\pi)^{d/2}} \int_S p^2(s)\ \mathrm{d}s$

The cut in the complete weighted graph		
Weighting	s_n^{cut}	$CutLim$ in complete weighted graph
Weighted	$n^2\sigma_n$	$\frac{2}{\sqrt{2\pi}} \int_S p^2(s)\ \mathrm{d}s$

The volume in the kNN-graph and the r-graph			
Weighting	s_n^{vol}	$VolLim(H)$ kNN-graph	$VolLim(H)$ r-graph
Unweighted	$n^2 r_n^d$	$\int_H p(x)\ \mathrm{d}x$	$\eta_d \int_H p^2(x)\ \mathrm{d}x$
Weighted, $r_n/\sigma_n \to \infty$	n^2	$\int_H p^2(x)\ \mathrm{d}x$	$\int_H p^2(x)\ \mathrm{d}x$
Weighted, $r_n/\sigma_n \to 0$	$\sigma_n^{-d} n^2 r_n^d$	$\frac{1}{(2\pi)^{d/2}} \int_H p(x)\ \mathrm{d}x$	$\frac{\eta_d}{(2\pi)^{d/2}} \int_H p^2(x)\ \mathrm{d}x$

The volume in the complete weighted graph		
Weighting	s_n^{vol}	$VolLim$ in complete weighted graph
Weighted	n^2	$\int_H p^2(x)\ \mathrm{d}x$

On the other hand, the standard cut, which we consider in this paper, can be written (up to a constant) as $f' L_{\text{unnorm}} f$, where L_{unnorm} denotes the unnormalized graph Laplacian matrix. (For the definitions of the graph Laplacian matrices and their relationship to the cut we refer the reader to von Luxburg [13]). Therefore, the two results do not contradict each other.

Theorem 3.3 (NCut and CheegerCut on the complete weighted graph)**.** *Let G_n be the complete weighted graph with Gaussian weights and bandwidth σ_n on the sample points $x_1, \ldots, x_n$. Set* XCut = NCut *or* XCut = CheegerCut *and let $XCutLim$ denote the corresponding limit as defined in equations* (3.1) *and* (3.2)*. Set*

$$\Delta_n = \left| \frac{s_n^{\text{vol}}}{s_n^{\text{cut}}} \operatorname{XCut}_n - XCutLim \right|.$$

Under the conditions $\sigma_n \to 0$ and $n\sigma_n^{d+1}/\log n \to \infty$ we have almost surely $\Delta_n \to 0$ for $n \to \infty$. The optimal convergence rate is achieved setting $\sigma_n = \sigma_0 \sqrt[d+3]{\log n / n}$ with a suitable $\sigma_0 > 0$. For this choice of σ_n the convergence rate is in $O((\log n/n)^{\alpha/(d+3)})$ for any $\alpha \in (0,1)$.

Let us decrypt these results and for simplicity focus on the cut value. When we compare the limits of the cut (*cf.* Tab. 1) it is striking that, depending on the graph type and the weighting scheme, there are two substantially different limits: the limit $\int_S p^2(s)\ \mathrm{d}s$ for the unweighted r-neighborhood graph, and the limit $\int_S p^{1-1/d}(s)\ \mathrm{d}s$ for the unweighted k-nearest neighbor graph.

The limit of the cut for the complete weighted graph with Gaussian weights is the same as the limit for the unweighted r-neighborhood graph. There is a simple reason for that: on both graph types the weight of an edge only depends on the distance between its end points, no matter where the points are. This is in contrast to the kNN-graph, where the radius up to which a point is connected strongly depends on its location: if a point is in a region of high density there will be many other points close by, which means that the radius is small. On the other hand, this radius is large for points in low-density regions. Furthermore, the Gaussian weights decline very rapidly with the distance, depending on the parameter σ. That is, σ plays a similar role as the radius r for the r-neighborhood graph.

The two types of r-neighborhood graphs with Gaussian weights have the same limit as the unweighted r-neighborhood graph and the complete weighted graph with Gaussian weights. When we compare the scaling sequences s_n^{cut} it turns out that in the case $r_n/\sigma_n \to \infty$ this sequence is the same as for the complete weighted graph, whereas in the case $r_n/\sigma_n \to 0$ we have $s_n^{\mathrm{cut}} = n^2 r_n^{d+1}/\sigma_n^d$, which is the same sequence as for the unweighted r-graph corrected by a factor of σ_n^{-d}. In fact, these effects are easy to explain: if $r_n/\sigma_n \to \infty$ then the edges which we have to remove from the complete weighted graph in order to obtain the r_n-neighborhood graph have a very small weight and their contribution to the value of the cut can be neglected. Therefore this graph behaves like the complete weighted graph with Gaussian weights. On the other hand, if $r_n/\sigma_n \to 0$ then all the edges that remain in the r_n-neighborhood graph have approximately the same weight, namely the maximum of the Gaussian weight function, which is linear in σ_n^{-d}.

Similar effects can be observed for the k-nearest neighbor graphs. The limits of the unweighted graph and the graph with Gaussian weight and $r_n/\sigma_n \to 0$ are identical (up to constants) and the scaling sequence has to correct for the maximum of the Gaussian weight function. However, the limit for the kNN-graph with Gaussian weights and $r_n/\sigma_n \to \infty$ is different: in fact, we have the same limit expression as for the complete weighted graph with Gaussian weights. The reason for this is the following: since r_n is large compared to σ_n at some point all the k-nearest neighbor radii of the sample points are very large. Therefore, all the edges that are in the complete weighted graph but not in the kNN graph have very low weights and thus the limit of this graph behaves like the limit of the complete weighted graph with Gaussian weights.

Finally, we would like to discuss the difference between the two limit expressions, where as examples for the graphs we use only the unweighted r-neighborhood graph and the unweighted kNN-graph. Of course, the results can be carried over to the other graph types. For the cut we have the limits $\int_S p^{1-1/d}(s)\ \mathrm{d}s$ and $\int_S p^2(s)\ \mathrm{d}s$. In dimension 1 the difference between these expressions is most pronounced: the limit for the kNN graph does not depend on the density p at all, whereas in the limit for the r-graph the exponent of p is 2, independent of the dimension. Generally, the limit for the r-graph seems to be more sensitive to the absolute value of the density. This can also be seen for the volume: the limit expression for the kNN graph is $\int_H p(x)\ \mathrm{d}x$, which does not depend on the absolute value of the density at all, but only on the probability mass in the halfspace H. This is different for the unweighted r-neighborhood graph with the limit expression $\int_H p^2(x)\ \mathrm{d}x$.

4. Examples where different limits of Ncut lead to different optimal cuts

In Theorems 3.1–3.3 we have proved that the limit expressions for NCut and CheegerCut are different for different kinds of neighborhood graphs. In fact, apart from constants there are two limit expressions: that of the unweighted kNN-graph, where the exponent of the density p in the limit integral for the cut is $1-1/d$ and for the volume is 1, and that of the unweighted r-neighborhood graph, where the exponent in the limit of the cut is 2 and in the limit of the vol is 1. Therefore, we consider here only the unweighted kNN-graph and the unweighted r-neighborhood graph.

In this section we show that the difference between the limit expressions is more than a mathematical subtlety without practical relevance: if we select an optimal cut based on the limit criterion for the kNN graph we can obtain a different result than if we use the limit criterion based on the r-neighborhood graph.

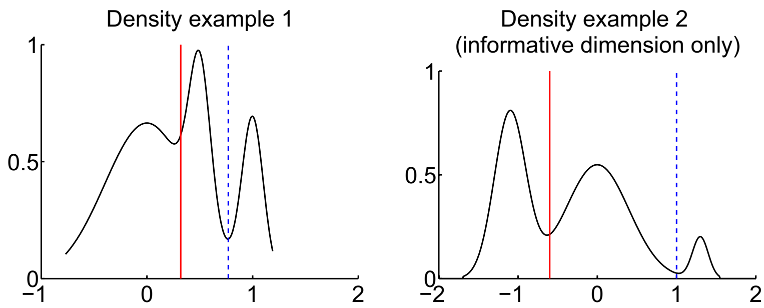

FIGURE 1. Densities in the examples. In the two-dimensional case, we plot the informative dimension (marginal over the other dimensions) only. The dashed blue vertical line depicts the optimal limit cut of the r-graph, the solid red vertical line the optimal limit cut of the kNN graph.

Consider Gaussian mixture distributions in one (Example 1) and in two dimensions (Example 2) of the form $\sum_{i=1}^{3} \alpha_i N([\mu_i, 0, \ldots, 0], \sigma_i I)$ which are set to zero where they are below a threshold θ and properly rescaled. The specific parameters in one and two dimensions are

Dim	μ_1	μ_2	μ_3	σ_1	σ_2	σ_3	α_1	α_2	α_3	θ
1	0	0.5	1	0.4	0.1	0.1	0.66	0.17	0.17	0.1
2	-1.1	0	1.3	0.2	0.4	0.1	0.4	0.55	0.05	0.01

Plots of the densities of Examples 1 and 2 can be seen in Figure 1. We first investigate the theoretic limit cut values, for hyperplanes which cut perpendicular to the first dimension (which is the "informative" dimension of the data). For the chosen densities, the limit NCut expressions from Theorems 3.1 and 3.2 can be computed analytically and optimized over the chosen hyperplanes. The solid red line in Figure 1 indicates the position of the minimal value for the kNN-graph case, whereas the dashed blue line indicates the the position of the minimal value for the r-graph case.

Up to now we only compared the limits of different graphs with each other, but the question is, whether the effects of these limits can be observed even for finite sample sizes. In order to investigate this question we applied normalized spectral clustering (*cf.* von Luxburg [13]) to sample data sets of $n = 2000$ points from the mixture distribution above. We used the unweighted r-graph and the unweighted symmetric k-nearest neighbor graph. We tried a range of reasonable values for the parameters k and r and the results we obtained were stable over a range of parameters. Here we present the results for the 30- (for $d = 1$) and the 150-nearest neighbor graphs (for $d = 2$) and the r-graphs with corresponding parameter r, that is r was set to be the mean 30- and 150-nearest neighbor radius. Different clusterings are compared using the minimal matching distance:

$$d_{MM}(\text{Clust}_1, \text{Clust}_2) = \min_{\pi} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\text{Clust}_1(x_i) \neq \pi(\text{Clust}_2(x_i))}$$

where the minimum is taken over all permutations π of the labels. In the case of two clusters, this distance corresponds to the 0-1-loss as used in classification: a minimal matching distance of 0.35, say, means that 35% of the data points lie in different clusters. In our spectral clustering experiment, we could observe that the clusterings obtained by spectral clustering are usually very close to the theoretically optimal hyperplane splits predicted by theory (the minimal matching distances to the optimal hyperplane splits were always in the order of 0.03 or smaller). As predicted by theory, the two types of graph give different cuts in the data. An illustration of this phenomenon for the case of dimension 2 can be found in Figure 2. To give a quantitative evaluation of this phenomenon, we computed the mean minimal matching distances between clusterings obtained by the

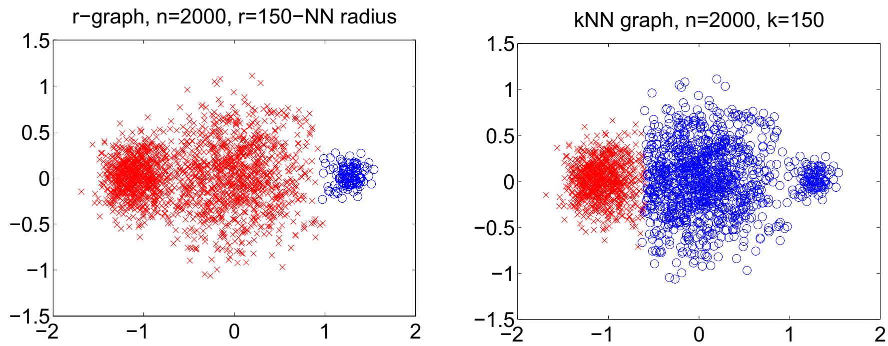

FIGURE 2. Results of spectral clustering in two dimensions, for the unweighted r-graph (left) and the unweighted kNN graph (right).

same type of graph over the different samples (denoted d_{kNN} and d_r), and the mean difference $d_{\mathrm{kNN}\,-r}$ between the clusterings obtained by different graph types:

Example	d_{kNN}	d_r	$d_{\mathrm{kNN}\,-r}$
1 dim	0.0005 ± 0.0006	0.0003 ± 0.0004	0.346 ± 0.063
2 dim	0.005 ± 0.0023	0.001 ± 0.001	0.49 ± 0.01

We can see that for the same graph, the clustering results are very stable (differences in the order of 10^{-3}) whereas the differences between the kNN graph and the r-neighborhood graph are substantial (0.35 and 0.49, respectively). This difference is exactly the one induced by assigning the middle mode of the density to different clusters, which is the effect predicted by theory.

It is tempting to conjecture that in Examples 1 and 2 the two different limit solutions and their impact on spectral clustering might arise due to the fact that the number of Gaussians and the number of clusters we are looking for do not coincide. Yet the following Example 3 shows that this is not the case: for a density in one dimension as above but with only two Gaussians with parameters

μ_1	μ_2	σ_1	σ_2	α_1	α_2	θ
0.2	0.4	0.05	0.03	0.8	0.2	0.1

the same effects can be observed. The density is depicted in the left plot of Figure 3.

In this example we draw a sample of 2000 points from this density and compute the spectral clustering of the points, once with the unweighted kNN-graph and once with the unweighted r-graph. In one dimension we can compute the place of the boundary between two clusters, that is the middle between the rightmost point of the left cluster and the leftmost point of the right cluster. We did this for 100 iterations and plotted histograms of the location of the cluster boundary. In the middle and the right plot of Figure 3 we see that these coincide with the optimal cut predicted by theory.

5. OUTLOOK

In this paper we have investigated the influence of the graph construction on the graph-based clustering measures normalized cut and Cheeger cut. We have seen that depending on the type of graph and the weights, the clustering quality measures converge to different limit results.

This means that ultimately, the question about the "best NCut" or "best Cheeger cut" clustering, given infinite amount of data, has different answers, depending on which underlying graph we use. This observation opens Pandora's box on clustering criteria: the "meaning" of a clustering criterion does not only depend on the exact definition of the criterion itself, but also on how the graph on the finite sample is constructed. This means that one graph clustering quality measure is not just "one well-defined criterion" on the underlying space, but

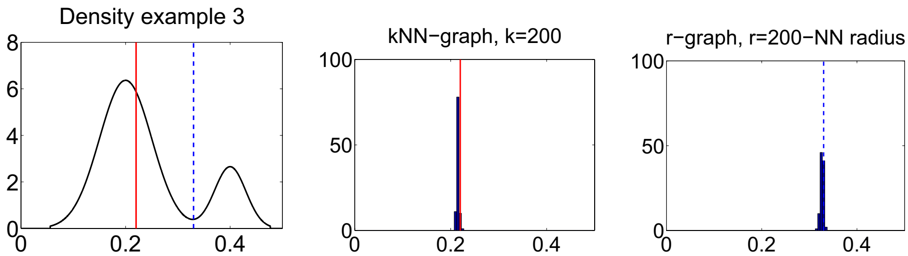

FIGURE 3. The Example 3 with the sum of two Gaussians, that is two modes of the density. In the left figure the density with the optimal limit cut of the r-graph (dashed blue vertical line) and the optimal limit cut of the kNN graph (the solid red vertical line) is depicted. The two figures on the right show the histograms of the cluster boundary over 100 iterations for the unweighted r-neighborhood and kNN-graphs.

it corresponds to a whole bunch of criteria, which differ depending on the underlying graph. More sloppy: a clustering quality measure applied to one neighborhood graph does something different in terms of partitions of the underlying space than the same quality measure applied to a different neighborhood graph. This shows that these criteria cannot be studied isolated from the graph they are applied to.

From a theoretical side, there are several directions in which our work can be improved. In this paper we only consider partitions of Euclidean space that are defined by hyperplanes. This restriction is made in order to keep the proofs reasonably simple. However, we are confident that similar results could be proven for arbitrary smooth surfaces.

Another extension would be to obtain uniform convergence results. Here one has to take care that one uses a suitably restricted class of candidate surfaces S (note that uniform convergence results over the set of all partitions of $\mathbb{R}^d$ are impossible, *cf.* Bubeck and von Luxburg [4]). This result would be especially useful, if there existed a practically applicable algorithm to compute the optimal surface out of the set of all candidate surfaces.

For practice, it will be important to study how the different limit results influence clustering results. So far, we do not have much intuition about when the different limit expressions lead to different optimal solutions, and when these solutions will show up in practice. The examples we provided above already show that different graphs indeed can lead to systematically different clusterings in practice. Gaining more understanding of this effect will be an important direction of research if one wants to understand the nature of different graph clustering quality measures.

6. Proofs

In many of the proofs that are to follow in this section a lot of technique is involved in order to come to terms with problems that arise due to effects at the boundary of our support C and to the non-uniformity of the density p. However, if these technicalities are ignored, the basic ideas of the proofs are simple to explain and they are similar for the different types of neighborhood graphs. In Section 6.1 we discuss these ideas without the technical overhead and define some quantities that are necessary for the formulation of our results.

In Section 6.2 we present the results for the k-nearest neighbor graph and in Section 6.3 we present those for the r-graph and the complete weighted graph. Each of these sections consists of three parts: the first is devoted to the cut, the second is devoted to the volume, and in the third we proof the main theorem for the considered graphs using the results for the cut and the volume.

The sections on the convergence of the cut and the volume always follow the same scheme: first, a proposition concerning the convergence of the cut or the volume for general monotonically decreasing weight functions is given. Using this general proposition the results for the specific weight functions we consider in this paper follow as corollaries.

Since the basic ideas of our proofs are the same for all the different graphs, it is not worth repeating the same steps for all the graphs. Therefore, we decided to give detailed proofs for the k-nearest neighbor graph, which is the most difficult case. The r-neighborhood graph and the complete weighted graph can be treated together and we mainly discuss the differences to the proof for the kNN graph.

The limits of the cut and the volume for general weight function are expressed in terms of certain integrals of the weight function over "caps" and "balls", which are explained later. For a specific weight function these integrals have to be evaluated. This is done in the lemmas in Section 6.4. Furthermore, this section contains a technical lemma that helps us to control boundary effects.

6.1. Basic ideas

In this section we present the ideas of our convergence proofs non-formally. We focus here on NCut, but all the ideas can easily be carried over to the Cheeger cut.

First step: Decompose NCut_n *into* cut_n *and* vol_n

For sequences a_n, b_n that converge to the limits $a > 0$ and $b > 0$, the convergence of $a_n - b_n$, a_n/b_n can be expressend in terms of the convergence speed of a_n and b_n. Therefore, under our general assumptions, there exist constants c_1, c_2, c_3, which may depend on the limit values of the cut and the volume, such that for sufficiently large n

$$\left|\frac{s_n^{\mathrm{vol}}}{s_n^{\mathrm{cut}}}\left(\frac{\mathrm{cut}_n}{\mathrm{vol}_n(H^+)}+\frac{\mathrm{cut}_n}{\mathrm{vol}_n(H^-)}\right)-\frac{CutLim}{VolLim(H^+)}+\frac{CutLim}{VolLim(H^-)}\right|$$
$$\leq c_1\underbrace{\left|\frac{\mathrm{cut}_n}{s_n^{\mathrm{cut}}}-CutLim\right|}_{\text{cut term}}+c_2\underbrace{\left|\frac{\mathrm{vol}_n(H^+)}{s_n^{\mathrm{vol}}}-VolLim(H^+)\right|}_{\text{volume-term}}+c_3\underbrace{\left|\frac{\mathrm{vol}_n(H^-)}{s_n^{\mathrm{vol}}}-VolLim(H^-)\right|}_{\text{volume-term}}.$$

This decomposition is used in order to proof the main theorems, Theorems 3.1–3.3; the goal of the following steps is to find bounds on the terms on the right hand side of this equation.

Second step: Bias/variance decomposition of cut *and volume terms*

In order to show the convergence of the cut-term we do a bias/variance decomposition

$$\left|\frac{\mathrm{cut}_n}{s_n^{\mathrm{cut}}}-CutLim\right|\leq\underbrace{\left|\frac{\mathrm{cut}_n}{s_n^{\mathrm{cut}}}-\mathbb{E}\left(\frac{\mathrm{cut}_n}{s_n^{\mathrm{cut}}}\right)\right|}_{\text{variance term}}+\underbrace{\left|\mathbb{E}\left(\frac{\mathrm{cut}_n}{s_n^{\mathrm{cut}}}\right)-CutLim\right|}_{\text{bias term}}$$

and show the convergence to zero of these terms separately. Clearly, the same decomposition can be done for the volume terms. In the following we call these terms the "bias term of the cut" and the "variance term of the cut" and similarly for the volume.

In Propositions 6.1 and 6.12 bounds on the bias term and the variance term of the cut are shown for the k-nearest neighbor graph and the r-graph, respectively, for rather general weight functions. Similarly in Propositions 6.7 and 6.17 for the bias and the variance term of the volume. The following steps in this section show the ideas that are used in the proofs of these propositions.

Third step: Use concentration of measure inequalities for the variance term

Bounding the deviation of a random variable from its expectation is a well-studied problem in statistics and there are a couple of so-called concentration of measure inequalities that bound the probability of a large deviation from the mean. In this paper we use McDiarmid's inequality for the kNN graphs and a concentration of measure result for U-statistics by Hoeffding for the r-neighborhood graph and the complete weighted graph. The reason for this is that each of the graph types has its particular advantages and disadvantages when it comes to the prerequisites for the concentration inequalities: the advantage of the kNN graph is that we can bound the degree of a node linearly in the parameter k, whereas for the r-neighborhood graph we can bound the degree only by the trivial bound $(n-1)$ and for the complete graph this bound is even attained. Therefore, using the same proof as for the kNN-graph is suboptimal for the latter two graphs. On the other hand, in these graphs the connectivity between points is not random given their position and it is always symmetric. This allows us to use a U-statistics argument, which cannot be applied to the kNN-graph, since the connectivity there may be unsymmetric (at least for the directed one) and the connectivity between each two points depends on all the sample points.

Note that these results are of a probabilistic nature, that is we obtain results of the form

$$\Pr\left(\left|\frac{\mathrm{cut}_n}{s_n^{\mathrm{cut}}} - \mathbb{E}\left(\frac{\mathrm{cut}_n}{s_n^{\mathrm{cut}}}\right)\right| > \varepsilon\right) \le p_n,$$

for a sequence (p_n) of non-negative real numbers. If for all $\varepsilon > 0$ the sum $\sum_{i=1}^{\infty} p_i$ is finite, then we have almost sure convergence of the variance term to zero by the Borel-Cantelli lemma.

Formal proofs of probabilistic bounds on the variance terms can be found in the proofs of Propositions 6.1 and 6.12 for the cut, and in the proofs of Propositions 6.7 and 6.17 for the volume.

Fourth step: Bias of the cut *term*

While all steps so far were pretty much standard, this part is the technically most challenging part of our convergence proof. We have to prove the convergence of $\mathbb{E}(\mathrm{cut}_n / s_n^{\mathrm{cut}})$ to $CutLim$ (and similarly for the volume). Omitting all technical difficulties like boundary effects and the variability of the density, the basic ideas can be described in a rather simple manner. The formal proofs can be found in the proofs of Proposition 6.1 for the k-nearest neighbor graph, and in the proof of Proposition 6.12 for the r-neighborhood and the complete graph.

The first idea is to break the cut down into the contributions of each single edge. We define a random variable W_{ij} that attains the weight of the edge between x_i and x_j, if these points are connected in the graph and on different sides of the hyperplane S, and zero otherwise. By the linearity of the expectation and the fact that the points are sampled i.i.d.

$$\mathbb{E}\left(\mathrm{cut}_n\right) = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \ne i}}^{n} \mathbb{E} W_{ij} = n(n-1)\mathbb{E} W_{12}.$$

Now we fix the positions of the points $x_1 = x$ and $x_2 = y$. In this case W_{ij} can attain only two values: $f_n(\mathrm{dist}(x,y))$ if the points are connected and on different sides of S, and zero otherwise. We first consider the r-neighborhood graph with parameter r_n, since here the existence of an edge between two points is determined by their distance, and is not random as in the kNN graph. Two points are connected if their distance is not greater than r_n and thus $W_{ij} = 0$ if $\mathrm{dist}(x,y) > r_n$. Furthermore, $W_{ij} = 0$ if x and y are on the same side of S. That is, for a point $x \in H^+$ we have

$$\mathbb{E}(W_{12} \mid x_1 = x, x_2 = y) = \begin{cases} f_n(\mathrm{dist}(x,y)) & \text{if } y \text{ is in the cap } B(x,r_n) \cap H^- \\ 0 & \text{otherwise.} \end{cases}$$

By integrating over $\mathbb{R}^d$ we obtain

$$\mathbb{E}(W_{12} \mid x_1 = x) = \int_{B(x,r_n)\cap H^-} f_n(\mathrm{dist}(x,y)) p(y) \; \mathrm{d}y$$

and denote the integral on the right hand side in the following by $g(x)$.

Integrating the conditional expectation over all possible positions of the point x in $\mathbb{R}^d$ gives

$$\mathbb{E}\left(W_{12}\right) = \int_{\mathbb{R}^d} g(x)\; p(x) \; \mathrm{d}x = \int_{H^+} g(x)\; p(x)\; \mathrm{d}x + \int_{H^-} g(x)\; p(x)\; \mathrm{d}x.$$

We only consider the integral over the halfspace H^+ here, since the other integral can be treated analogously. The important idea in the evaluation of this integral is the following: instead of integrating over H^+, we initially

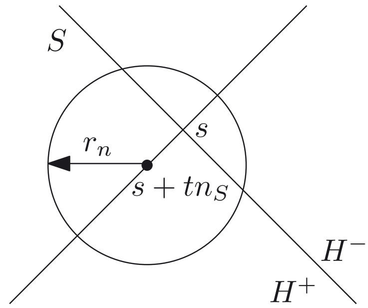

FIGURE 4. Integration along the normal line through s. Obviously, for $t \geq r_n$ the intersection $B(s + tn_S, r_n) \cap H^-$ is empty and therefore $g(s + tn_S) = 0$. For $0 \leq t < r_n$ the points in the cap are close to s and therefore the density in the cap is approximately $p(s)$.

integrate over the hyperplane S and then, at each point $s \in S$, along the normal line through s, that is the line $s + tn_S$ for all $t \in \mathbb{R}_{\geq 0}$. This leads to

$$\int_{H^+} g(x)\ p(x)\ \mathrm{d}x = \int_S \int_0^\infty g(s + tn_S)\ p(s + tn_S)\ \mathrm{d}t\ \mathrm{d}s.$$

This integration is illustrated in Figure 4. It has two advantages: first, if x is far enough from S (that is, $\mathrm{dist}(x, s) > r_n$ for all $s \in S$), then $g(x) = 0$ and the corresponding terms in the integral vanish. Second, if x is close to $s \in S$ and the radius r_n is small, then the density on the ball $B(x, r_n)$ can be considered approximately uniform, that is we assume $p(y) = p(s)$ for all $y \in B(x, r_n)$. Thus,

$$\begin{aligned}
\int_0^\infty g(s + tn_S)\ p(s + tn_S)\ \mathrm{d}t &= \int_0^{r_n} g(s + tn_S)\ p(s + tn_S)\ \mathrm{d}t \\
&= p(s) \int_0^{r_n} g(s + tn_S)\ \mathrm{d}t = p^2(s) \int_0^{r_n} \int_{B(x,r_n)\cap H^-} f_n(\mathrm{dist}(x, y))\ \mathrm{d}y\ \mathrm{d}t \\
&= \eta_{d-1} \int_0^{r_n} u^d f_n(u)\ \mathrm{d}u\ p^2(s)
\end{aligned}$$

where the last step follows with Lemma 6.3.

Since this integral of the weight function f_n over the "caps" plays such an important role in the derivation of our results we introduce a special notation for it: for a radius $r \in \mathbb{R}_{\geq 0}$ and $q = 1, 2$ we define

$$F_C^{(q)}(r) = \eta_{d-1} \int_0^r u^d f_n^q(u)\ \mathrm{d}u.$$

Although these integrals also depend on n we do not make this dependence explicit. In fact, the parameter r is replaced by the radius r_n in the case of the r-neighborhood graph or by a different graph parameter depending on n for the other neighborhood graphs. Therefore the dependence of $F_C^{(q)}(r_n)$ on n will be understood. Note that we allow the notation $F_C^{(q)}(\infty)$, if the indefinite integral exists. The integral $F_C^{(q)}$ for $q = 2$ is needed for the following reason: for the U-statistics bound on the variance term we do not only have to compute the expectation of W_{ij}, but also their variance. But the variance can in turn be bounded by the expectation of W_{ij}^2, which is expressed in terms of $F_C^{(2)}(r_n)$.

In the r-neighborhood graph points are only connected within a certain radius r_n, which means that to compute $\mathbb{E}(W_{12} \mid x_1 = x)$ we only have to integrate over the ball $B(x, r_n)$, since all other points cannot be connected to $x_1 = x$. This is clearly different for the complete graph, where every point is connected to every other point. The idea is to fix a radius r_n in such a way as to make sure that the contribution of edges to points outside $B(x, r_n)$ can be neglected, because their weight is small. Since $W_{12} = f_n(\mathrm{dist}(x_1, x_2))$ if the points are on different sides of S we have for $x \in H^+$

$$\begin{aligned}\mathbb{E}(W_{12} \mid x_1 = x) &= \int_{B(x,r_n)\cap H^-} f_n(\mathrm{dist}(x,y))\ p(y)\ \mathrm{d}y + \int_{B(x,r_n)^c\cap H^-} f_n(\mathrm{dist}(x,y))\ p(y)\ \mathrm{d}y \\ &\le g(x) + p_{\max} \int_{B(x,r_n)^c} f_n(\mathrm{dist}(x,y))\ \mathrm{d}y.\end{aligned}$$

For the Gaussian weight function the integral converges to zero very quickly, if $r_n/\sigma_n \to \infty$ for $n \to \infty$. Thus we can treat the complete graph almost as the r-neighborhood graph.

For the k-nearest neighbor graph the connectedness of points depends on their k-nearest neighbor radii that is, the distance of the point to its k-th nearest neighbor, which is itself a random variable. However, one can show that with very high probability the k-nearest neighbor radius of a point in a region with uniform density p is concentrated around $(k_n/((n-1)\eta_d p)^{1/d}$. Since we assume that $k_n/n \to 0$ for $n \to \infty$ the expected kNN radius converges to zero. Thus the density in balls with this radius is close to uniform and the estimate becomes more accurate. Upper and lower bounds on the k-nearest neighbor radius that hold with high probability are given in Lemma 6.2. The idea is to perform the integration above for both, the lower bound on the kNN radius and the upper bound on the kNN radius. Then it is shown that these integrals converge to the same limit.

Fifth step: Bias of the volume terms

The bias of the volume term, which is dealt with in in Propositions 6.7 for the k-nearest neighbor graph and in Proposition 6.17 for the r-neighborhood and the complete graph, can be treated similarly to that of the cut term. We define $W_{ij} = f_n(\mathrm{dist}(x_i, x_j)$ if x_i and x_j are connected in the graph and $W_{ij} = 0$ otherwise. Note that we do not need the condition that the points have to be on different sides of the hyperplane S as for the cut. Then, for a point $x \in C$ if we assume that the density is uniform within distance r_n around x

$$\begin{aligned}\mathbb{E}(W_{12} \mid x_1 = x) &= \int_{B(x,r_n)} f_n(\mathrm{dist}(x,y))p(y)\ \mathrm{d}y = p(x)\int_{B(x,r_n)} f_n(\mathrm{dist}(x,y))\ \mathrm{d}y \\ &= d\eta_d \int_0^{r_n} u^{d-1} f_n(u)\ \mathrm{d}u\ p(x),\end{aligned}$$

where the last integral transform follows with Lemma 6.8. Integrating over $\mathbb{R}^d$ we obtain

$$\mathbb{E}(W_{12}) = \int_{\mathbb{R}^d} \mathbb{E}(W_{12} \mid x_1 = x)p(x)\ \mathrm{d}x = d\eta_d \int_0^{r_n} u^{d-1} f_n(u)\ \mathrm{d}u \int_{\mathbb{R}^d} p^2(x)\ \mathrm{d}x.$$

Since the integral over the balls is so important in the formulation of our general results we often call it the "ball integral" and introduce the notation

$$F_B^{(q)}(r) = d\eta_d \int_0^r u^{d-1} f_n(u)\ \mathrm{d}u$$

for a radius $r > 0$ and $q = 1, 2$. The remarks that were made on the "cap integral" $F_C(r)$ above also apply to the "ball integral" $F_B(r)$.

Sixth step: Plugging in the weight functions

Having derived results on the bias term of the cut and volume for general weight functions, we can now plug in the specific weight functions in which we are interested in this paper. This boils down to the evaluation of the "cap" and "ball" integrals $F_C(r_n)$ and $F_B(r_n)$ for these weight functions. For the unit weight function the integrals can be computed exactly (Lem. 6.22), whereas for the Gaussian weight function we study the asymptotic behavior of the "cap" and "ball" integral in the cases $r_n/\sigma_n \to 0$ (Lem. 6.23) and $r_n/\sigma_n \to \infty$ for $n \to \infty$ (Lem. 6.24).

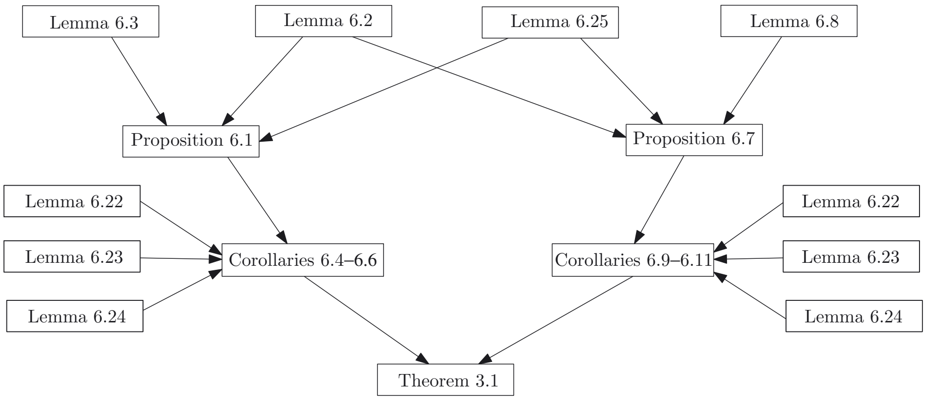

FIGURE 5. The structure of the proofs in this section. Proposition 6.1 and 6.7 state bounds for general weight functions on the bias and the variance term of the cut and the volume, respectively. Lemma 6.2 shows the concentration of the kNN radii, Lemma 6.25 is needed to bound the influence of points close to the boundary. Lemma 6.3 and 6.8 perform the integration of the weight function over "caps" and "balls". In Lemmas 6.22–6.24 the general "ball" and "cap" integrals are evaluated for the specific weight functions we use. Using these results, Corollaries 6.4–6.6 dealing with the cut and Corollaries 6.9–6.11 dealing with the volume are proved. Finally, in Theorem 3.1 the convergence of NCut and CheegerCut are analyzed using the result of these corollaries.

6.2. Proofs for the k-nearest neighbor graph

As we have already mentioned we will give the proofs of our general propositions in detail here and then discuss in Section 6.3 how they have to be adapted to the r-neighborhood graph and the complete weighted graph. This means, that Lemmas 6.3 and 6.8 that are necessary for the proof of the general propositions can be found in this section, although they are also needed for the r-graph and the complete graph with Gaussian weights.

This section consists of four subsections: in Section 6.2.1 we define some quantities that help us to deal with the fact that the connectivity between two points is random even if we know their distance. These quantities will play an important role in the succeeding sections. Section 6.2.2 presents the results for the cut term, whereas Section 6.2.3 presents the results for the volume term. Finally, these results are used to proof Theorem 3.1, the main theorem for the k-nearest neighbor graph in Section 6.2.4.

In the subsections on the cut-term and the volume term we always present the proposition for general weight functions first. Then the lemmas follow that are used in the proof of the proposition. Finally, we show corollaries that apply these general results to the specific weight functions we consider in this paper. An overview of the proof structure is given in Figure 5.

6.2.1. k-nearest neighbor radii

As we have explained in Section 6.1 the basic ideas of our convergence proofs are similar for all the graphs. However, there is one major technical difficulty for the k-nearest neighbor graph: the existence of an edge between two points depends on all the other sample points and it is random, even if we know the distance between the points. However, each sample point x_i is connected to its k nearest neighbors, that means to all points with a distance not greater than that of the k-th nearest neighbor. This distance is called the k-nearest neighbor radius of point x_i. Unfortunately, given a sample point we do not know this radius without looking at all the other points. The idea to overcome this difficulty is the following: given the position of a sample point we give lower and upper bounds on the kNN radius that depend on the density around the point and show that with high probability the true radius is between these bounds. Then we can replace the integration over balls of a fixed radius with the integration over balls with the lower and upper bound on the kNN radius in the proof for the bias term and then show that these integrals converge towards each other. Furthermore, under our assumptions the radius of all the points can be bounded from above, which helps to bound the influence of far-away points.

In this section we define formally the bounds on the k-nearest neighbor radii, since these will be used in the statement of the general proposition. In Lemma 6.2 we state the bounds on the probabilities that the true kNN radius is between our bounds for the cases we need in the proofs.

We first introduce the upper bound $r_n^{\max}$ on the maximum k-nearest neighbor radius of a point not depending on its position. Second, we use that given a point x (far enough) in the interior of C the conditional kNN radius of a sample point at x is highly concentrated around a radius $r_n(x)$. Formally, we define

$$r_n^{\max} = \sqrt[d]{\frac{4}{\gamma p_{\min}\eta_d}\frac{k_n}{n-1}}, \qquad \text{and} \qquad r_n(x) = \sqrt[d]{\frac{k_n}{(n-1)p(x)\eta_d}} \qquad \text{for all } x \in C.$$

As to the concentration we state sequences of lower and upper bounds, $r_n^-(x)$ and $r_n^+(x)$ that converge to $r_n(x)$ such that for all $x \in C$ that are not in a small boundary strip the probability that a point in x is connected to a point in y becomes small if the distance between x and y exceeds $r_n^+(x)$ and becomes large if the distance is smaller than $r_n^-(x)$.

Clearly, the accuracy of the bounds depends on how much the density can vary around x. Setting $\xi_n = 2p'_{\max}r_n^{\max}/p_{\min}$ the density in the ball of radius $2r_n^{\max}$ around x can vary between $(1-\xi_n)p(x)$ and $(1+\xi_n)p(x)$. Furthermore, we have to "blow up" or shrink the radii a bit in order to be sure that the true kNN radius is between them. To this end we introduce a sequence $(\delta_n)_{n\in\mathbb{N}}$ with $\delta_n \to 0$ and $\delta_n k_n \to \infty$ for $n \to \infty$. Then we can define

$$r_n^-(x) = \sqrt[d]{(1-2\xi_n)(1-\delta_n)}r_n(x) \quad \text{and} \quad r_n^+(x) = \sqrt[d]{(1+2\xi_n)(1+\delta_n)}r_n(x).$$

Note that ξ_n converges to zero, since $r_n^{\max}$ converges to zero as $\sqrt[d]{k_n/n}$. The sequence δ_n is chosen such that it converges to zero reasonably fast, but that with high probability $r_n^+(x)$ and $r_n^-(x)$ are bounds on the kNN radius of a point at x.

In order to quantify the probability of connections, which we seek to bound, we define the function $c : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ by

$$c(x,y) = \begin{cases} \Pr(C_{12} \mid x_1 = x, x_2 = y) & \text{if } x \in C \text{ and } y \in C \\ 0 & \text{otherwise,} \end{cases}$$

where C_{12} denotes the event that there is an edge between the sample points x_1 and x_2 in the (directed or undirected) k-nearest neighbor graph.

6.2.2. The cut *term in the* kNN *graph*

Proposition 6.1. *Let G_n be the directed, symmetric or mutual k-nearest neighbor graph with a monotonically decreasing weight function f_n. Set $\delta_n = \sqrt{(8\delta_0 \log n)/k_n}$ for some $\delta_0 \geq 2$ in the definition of $r_n^-(x)$. Then we have for the bias term*

$$\begin{aligned}\left|\mathbb{E}\left(\frac{\mathrm{cut}_n}{n(n-1)}\right) - 2\int_{S\cap C} p^2(s)F_C^{(1)}\left(r_n(s)\right)\, \mathrm{d}s\right| &= O\left(F_C^{(1)}\left(r_n^{\max}\right)\sqrt[d]{\frac{k_n}{n}}\right)\\
&+ O\left(\min\left\{n^{-\delta_0}f_n\left(\inf_{x\in C} r_n(x)\right), F_B^{(1)}(\infty) - F_B^{(1)}\left(\inf_{x\in C} r_n(x)\right)\right\}\right)\\
&+ O\left(\min\left\{\left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}}\right) f_n\left(\inf_{x\in C} r_n^-(x)\right)\left(\frac{k_n}{n}\right)^{1+1/d}, F_C^{(1)}(\infty) - F_C^{(1)}(\inf_{x\in C} r_n^-(x))\right\}\right).\end{aligned}$$

Furthermore, we have for the variance term for a suitable constant $\tilde{C}$

$$\Pr\left(\left|\mathrm{cut}_n - \mathbb{E}\left(\mathrm{cut}_n^{(i)}\right)\right| > \varepsilon\right) \leq 2\exp\left(-\frac{\tilde{C}\varepsilon^2}{nk_n^2 f_n^2(0)}\right).$$

Proof. We define for $i, j \in \{1, \ldots, n\}$, $i \neq j$ the random variable W_{ij} as

$$W_{ij} = \begin{cases} f_n(\mathrm{dist}(x_i, x_j) & \text{if } x_i \in H^+, x_j \in H^- \text{ and } (x_i, x_j) \text{ edge in } G_n \\ 0 & \text{otherwise.} \end{cases}$$

For both, a directed and an undirected graph we have

$$\mathrm{cut}_n = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} W_{ij},$$

and by the linearity of expectation and the fact that the points are independent and identically distributed, we have

$$\mathbb{E}\left(\frac{\mathrm{cut}_n}{n(n-1)}\right) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbb{E}(W_{ij}) = \frac{1}{n(n-1)} n(n-1)\mathbb{E}(W_{12}) = \mathbb{E}(W_{12}).$$

In the convergence proof for the variance term of the cut for the r-neighborhood graph in Proposition 6.12 we need a bound on $\mathbb{E}(W_{12}^2)$. Since this can be derived similarly to $\mathbb{E}(W_{12})$ we state the following for $\mathbb{E}(W_{12}^q)$ for $q = 1, 2$.

We define C_{12} to be the event that the sample points x_1 and x_2 are connected in the graph. Conditioning on the location of the points $x_1 \in C$ and $x_2 \in C$ we obtain $W_{12} = 0$ if x_1 and x_2 on the same side of the hyperplane$S$, otherwise

$$W_{12} = \begin{cases} f_n(\mathrm{dist}(x_1, x_2)) & \text{if } C_{12} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, if $x_1 \in C$ and $x_2 \in C$ are on different sides of S

$$\mathbb{E}\left(W_{12}^q \mid x_1 = x, x_2 = y\right) = f_n^q(\mathrm{dist}(x, y)) \Pr\left(C_{12} \mid x_1 = x, x_2 = y\right).$$

With $c(x, y)$ as above we have

$$\begin{aligned} \mathbb{E}(W_{12}^q) &= \int_C \int_C \mathbb{E}(W_{12}^q \mid x_1 = x, x_2 = y) p(y)\, \mathrm{d}y\; p(x)\, \mathrm{d}x \\ &= \int_{H^+ \cap C} \int_{H^- \cap C} f_n^q(\mathrm{dist}(x, y)) \Pr\left(C_{12} \mid x_1 = x, x_2 = y\right) p(y)\, \mathrm{d}y\; p(x)\, \mathrm{d}x \\ &\quad + \int_{H^- \cap C} \int_{H^+ \cap C} f_n^q(\mathrm{dist}(x, y)) \Pr\left(C_{12} \mid x_1 = x, x_2 = y\right) p(y)\, \mathrm{d}y\; p(x)\, \mathrm{d}x \\ &= \int_{H^+} \int_{H^-} f_n^q(\mathrm{dist}(x, y)) c(x, y) p(y)\, \mathrm{d}y\; p(x)\, \mathrm{d}x \\ &\quad + \int_{H^-} \int_{H^+} f_n^q(\mathrm{dist}(x, y)) c(x, y) p(y)\, \mathrm{d}y\; p(x)\, \mathrm{d}x. \end{aligned}$$

Setting

$$g(x) = \begin{cases} \int_{H^-} f_n^q(\operatorname{dist}(x,y))c(x,y)p(y)\ \mathrm{d}y & \text{if } x \in H^+ \\ \int_{H^+} f_n^q(\operatorname{dist}(x,y))c(x,y)p(y)\ \mathrm{d}y & \text{if } x \in H^- \end{cases}$$

we obtain

$$\mathbb{E}(W_{12}^q) = \int_{\mathbb{R}^d} g(x)p(x)\ \mathrm{d}x = \int_{H^+} g(x)p(x)\ \mathrm{d}x + \int_{H^-} g(x)p(x)\ \mathrm{d}x.$$

We only deal with the first integral here, the second can be computed analogously. By a simple transformation of the coordinate system we can write this integral as an integral along the hyperplane S, and for each points s in S we integrate over the normal line through s. In the following we find lower and upper bounds on the integral

$$\int_S \int_0^\infty g(s+tn_S)p\left(s+tn_S\right) \mathrm{d}t\mathrm{d}s = \int_S h_n(s)\mathrm{d}s,$$

where we have set

$$h_n(s) = \int_0^\infty g\left(s+tn_S\right) p\left(s+tn_S\right)\ \mathrm{d}t.$$

We set $\mathcal{I}_n = \{x \in C \mid \operatorname{dist}(x, \partial C) \geq 2r_n^{\max}\}$ and use the following decomposition of the integral

$$\left| \int_S h_n(s)\ \mathrm{d}s - \int_S p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s \right| \leq \left| \int_S h_n(s)\ \mathrm{d}s - \int_{S\cap\mathcal{I}_n} h_n(s)\ \mathrm{d}s \right| \tag{6.1}$$

$$+ \left| \int_{S\cap\mathcal{I}_n} h_n(s)\ \mathrm{d}s - \int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s \right| \tag{6.2}$$

$$+ \left| \int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s - \int_{S\cap C} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s \right|. \tag{6.3}$$

We first give a bound on the right hand side of equation (6.1). Setting $\mathcal{R}_n = \{x \in \mathbb{R}^d \mid \operatorname{dist}(x, \partial C) < 2r_n^{\max}\}$ and $\mathcal{A}_n = \mathbb{R}^d \setminus (\mathcal{I}_n \cup \mathcal{R}_n)$, we have (considering that the integrand is positive and $S \cap \mathcal{I}_n \subseteq S$)

$$\left| \int_S h_n(s)\ \mathrm{d}s - \int_{S\cap\mathcal{I}_n} h_n(s)\ \mathrm{d}s \right| = \int_{S\cap\mathcal{R}_n} h_n(s)\ \mathrm{d}s + \int_{S\cap\mathcal{A}_n} h_n(s)\ \mathrm{d}s,$$

that is, we have to derive upper bounds on the two integrals on the right hand side.

First let $s \in S \cap \mathcal{A}_n$, that is $s \notin C$ and $\operatorname{dist}(s, C) \geq 2r_n^{\max}$. Consequently $p(s+tn_S) = 0$ for $t < 2r_n^{\max}$. On the other hand, if $t \geq 2r_n^{\max}$ we have $\operatorname{dist}(s+tn_S, y) \geq 2r_n^{\max}$ for all $y \in H^-$. Setting $c_n = 2\exp(-k_n/8)$ we have with Lemma 6.2 $c(s+tn_S, y) \leq c_n$ for all $y \in H^-$. Hence

$$\begin{aligned} g(s+tn_S) \leq & \int_{B(s+tn_S, r_n^{\max})\cap H^-} f_n^q(\operatorname{dist}(s+tn_S, y))c(s+tn_S, y)p(y)\ \mathrm{d}y \\ & + \int_{B(s+tn_S, r_n^{\max})^c\cap H^-} f_n^q(\operatorname{dist}(s+tn_S, y))c(s+tn_S, y)p(y)\ \mathrm{d}y \\ & \leq f_n^q\left(r_n^{\max}\right) \int_{H^-} c(s+tn_S, y)p(y)\ \mathrm{d}y \leq c_n f_n^q\left(r_n^{\max}\right), \end{aligned}$$

since $B(s+tn_S, r_n^{\max}) \cap H^- = \emptyset$ for $t > r_n^{\max}$ and f_n is monotonically decreasing. Therefore, for all $s \in S \cap \mathcal{A}_n$

$$\begin{aligned} h_n(s) &= \int_0^\infty g\left(s+tn_S\right) p\left(s+tn_S\right)\ \mathrm{d}t \le \int_{2r_n^{\max}}^\infty g(s+tn_S) p\left(s+tn_S\right)\ \mathrm{d}t \\ &\le c_n f_n^q\left(r_n^{\max}\right) \int_0^\infty p\left(s+tn_S\right)\ \mathrm{d}t, \end{aligned}$$

and thus

$$\begin{aligned} \int_{S\cap\mathcal{A}_n} h_n(s)\ \mathrm{d}s &\le \int_{S\cap\mathcal{A}_n} c_n f_n^q\left(r_n^{\max}\right) \int_0^\infty p\left(s+tn_S\right)\ \mathrm{d}t\ \mathrm{d}s \\ &\le c_n f_n^q\left(r_n^{\max}\right) \int_S \int_0^\infty p\left(s+tn_S\right)\ \mathrm{d}t\ \mathrm{d}s \le c_n f_n^q\left(r_n^{\max}\right). \end{aligned}$$

Now let $s \in S \cap \mathcal{R}_n$. Then

$$\begin{aligned} g(s+tn_S) &= \int_{H^-} f_n^q(\operatorname{dist}(s+tn_S, y)) c(s+tn_S, y) p(y)\ \mathrm{d}y \\ &\le \int_{B(s+tn_S, r_n^{\max})\cap H^-} f_n^q(\operatorname{dist}(s+tn_S, y)) c(s+tn_S, y) p(y)\ \mathrm{d}y \\ &\quad + \int_{B(s+tn_S, r_n^{\max})^c\cap H^-} f_n^q(\operatorname{dist}(s+tn_S, y)) c(s+tn_S, y) p(y)\ \mathrm{d}y \\ &\le p_{\max} \int_{B(s+tn_S, r_n^{\max})\cap H^-} f_n^q(\operatorname{dist}(s+tn_S, y))\ \mathrm{d}y + c_n f_n^q\left(r_n^{\max}\right). \end{aligned}$$

Considering that $B(s+tn_S, r_n^{\max}) \cap H^- = \emptyset$ for $t > r_n^{\max}$ and therefore the first integral vanishes in this case, we have for all $s \in S \cap \mathcal{R}_n$

$$\begin{aligned} h_n(s) &= \int_0^\infty g\left(s+tn_S\right) p\left(s+tn_S\right)\ \mathrm{d}t \\ &\le \int_0^{r_n^{\max}} p_{\max} \int_{B(s+tn_S, r_n^{\max})\cap H^-} f_n^q(\operatorname{dist}(s+tn_S, y))\ \mathrm{d}y\ p\left(s+tn_S\right)\ \mathrm{d}t \\ &\quad + c_n f_n^q\left(r_n^{\max}\right) \int_0^\infty p\left(s+tn_S\right)\ \mathrm{d}t \\ &\le p_{\max}^2 \int_0^{r_n^{\max}} \int_{B(s+tn_S, r_n^{\max})\cap H^-} f_n^q(\operatorname{dist}(s+tn_S, y))\ \mathrm{d}y\ \mathrm{d}t \\ &\quad + c_n f_n^q\left(r_n^{\max}\right) \int_0^\infty p\left(s+tn_S\right)\ \mathrm{d}t \\ &\le p_{\max}^2 F_C^{(q)}\left(r_n^{\max}\right) + c_n f_n^q\left(r_n^{\max}\right) \int_0^\infty p\left(s+tn_S\right)\ \mathrm{d}t, \end{aligned}$$

and thus

$$\begin{aligned} \int_{S\cap\mathcal{R}_n} h_n(s)\ \mathrm{d}s &\le \int_{S\cap\mathcal{R}_n} p_{\max}^2 F_C^{(q)}\left(r_n^{\max}\right) + c_n f_n^q\left(r_n^{\max}\right) \int_0^\infty p\left(s+tn_S\right)\ \mathrm{d}t\ \mathrm{d}s \\ &\le p_{\max}^2 F_C^{(q)}\left(r_n^{\max}\right) \mathcal{L}_{d-1}\left(S\cap\mathcal{R}_n\right) + c_n f_n^q\left(r_n^{\max}\right). \end{aligned}$$

For some weight functions, for example the Gaussian, it is preferable to use that for all $x \in \mathbb{R}^d$ and all radii r

$$\begin{aligned}\int_{B(x,r)^c \cap H^-} f_n^q(\operatorname{dist}(x,y)) c(x,y) p(y) \, \mathrm{d}y &\le p_{\max} \int_{B(x,r)^c} f_n^q(\operatorname{dist}(x,y)) \, \mathrm{d}y \\ &= p_{\max} \left(\int_{\mathbb{R}^d} f_n^q(\operatorname{dist}(x,y)) \, \mathrm{d}y - \int_{B(x,r)} f_n^q(\operatorname{dist}(x,y)) \, \mathrm{d}y \right) \\ &= p_{\max} \left(F_B^{(q)}(\infty) - F_B^{(q)}(r) \right).\end{aligned}$$

We have according to Lemma 6.25 $\mathcal{L}_{d-1}(S \cap \mathcal{R}_n) = O(r_n^{\max})$. Consequently, using $r_n^{\max} = O(\sqrt[d]{k_n/n})$ and plugging in c_n

$$\begin{aligned}&\left| \int_S h_n(s) \, \mathrm{d}s - \int_{S \cap \mathcal{I}_n} h_n(s) \, \mathrm{d}s \right| \\ &\quad = O\left(F_C^{(q)}(r_n^{\max}) \sqrt[d]{\frac{k_n}{n}} + \min\left\{ \exp(-k_n/8) f_n^q\left(\inf_{x \in C} r_n(x) \right), \left(F_B^{(q)}(\infty) - F_B^{(q)}(r_n^{\max}) \right) \right\} \right).\end{aligned}$$

Now we consider the term in equation (6.2). In the following, note that with $\xi_n = 2p'_{\max} r_n^{\max}/p_{\min}$ we have for all $x \in C$ with $B(x, 2r_n^{\max}) \subseteq C$ and $y \in B(x, 2r_n^{\max})$

$$(1 - \xi_n) p(x) \le p(y) \le (1 + \xi_n) p(x).$$

We assume that n is sufficiently large such that $\xi_n < 1/2$.

For any $s \in S \cap \mathcal{I}_n$ and any $t \ge 0$ we have

$$\begin{aligned}g(s + t n_S) &= \int_{H^-} f_n^q(\operatorname{dist}(s + t n_S, y)) c(s + t n_S, y) p(y) \, \mathrm{d}y \\ &\ge \int_{B(s + t n_S, r_n^-(s)) \cap H^-} f_n^q(\operatorname{dist}(s + t n_S, y)) c(s + t n_S, y) p(y) \, \mathrm{d}y.\end{aligned}$$

If $t > r_n^-(s)$ we use the trivial bound $g(s + t n_S) \ge 0$. Otherwise we have with Lemma 6.2 for all $y \in B(s + t n_S, r_n^-(s)) \cap H^-$ that $c(s + t n_S, y) \ge 1 - a_n$ with $a_n = 6 \exp\left(-\delta_n^2 k_n/3\right)$. Using, furthermore, the bound $p(y) \ge (1 - \xi_n) p(s)$ we obtain

$$\begin{aligned}g(s + t n_S) &\ge \int_{B(s + t n_S, r_n^-(s)) \cap H^-} f_n^q(\operatorname{dist}(s + t n_S, y))(1 - a_n)(1 - \xi_n) p(s) \, \mathrm{d}y \\ &= (1 - a_n)(1 - \xi_n) p(s) \int_{B(s + t n_S, r_n^-(s)) \cap H^-} f_n^q(\operatorname{dist}(s + t n_S, y)) \, \mathrm{d}y.\end{aligned}$$

That is, we obtain for $s \in \mathcal{I}_n$

$$\begin{aligned}h_n(s) &= \int_0^\infty g(s + t n_S) \, p(s + t n_S) \, \mathrm{d}t \ge \int_0^{r_n^-(s)} g(s + t n_S) \, p(s + t n_S) \, \mathrm{d}t \\ &\ge (1 - \xi_n) p(s) \int_0^{r_n^-(s)} g(s + t n_S) \, \mathrm{d}t \\ &\ge (1 - a_n)(1 - \xi_n)^2 p^2(s) \int_0^{r_n^-(s)} \int_{B(s + t n_S, r_n^-(s)) \cap H^-} f_n^q(\operatorname{dist}(s + t n_S, y)) \, \mathrm{d}y \, \mathrm{d}t \\ &\ge (1 - a_n)(1 - \xi_n)^2 p^2(s) F_C^{(q)}\left(r_n^-(s)\right),\end{aligned}$$

where in the last inequality we have applied Lemma 6.3.

Therefore

$$\begin{aligned}\int_{S\cap\mathcal{I}_n} h_n(s)\ \mathrm{d}s \geq& (1-a_n)(1-\xi_n)^2 \int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n^-(s)\right)\ \mathrm{d}s\\ \geq& (1-a_n)(1-\xi_n)^2 \int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s\\ &- \int_{S\cap\mathcal{I}_n} p^2(s)\left(F_C^{(q)}\left(r_n(s)\right) - F_C^{(q)}\left(r_n^-(s)\right)\right)\ \mathrm{d}s\\ \geq& \int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s - (a_n+\xi_n)\int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s\\ &- p_{\max}^2 \int_{S\cap\mathcal{I}_n} \left(F_C^{(q)}\left(r_n(s)\right) - F_C^{(q)}\left(r_n^-(s)\right)\right)\ \mathrm{d}s,\end{aligned}$$

and thus

$$\begin{aligned}\int_{S\cap\mathcal{I}_n} h_n(s)\ \mathrm{d}s - \int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s \geq& -(a_n+\xi_n)\int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s\\ &- p_{\max}^2 \mathcal{L}_{d-1}(S\cap C) \sup_{s\in S\cap\mathcal{I}_n}\left(F_C^{(q)}\left(r_n^+(s)\right) - F_C^{(q)}\left(r_n(s)\right)\right).\end{aligned} \tag{6.4}$$

Now, we want to find an upper bound on $g(s+tn_S)$ for $s\in S\cap\mathcal{I}_n$, that is $B(s, 2r_n^{\max})\subseteq C$. We use the following decomposition

$$\begin{aligned}g(s+tn_S) =& \int_{H^-} f_n^q(\mathrm{dist}(s+tn_S,y))c(s+tn_S,y)p(y)\ \mathrm{d}y\\ \leq& \int_{B(s+tn_S,r_n^+(s))\cap H^-} f_n^q(\mathrm{dist}(s+tn_S,y))c(s+tn_S,y)p(y)\ \mathrm{d}y\\ &+ \int_{B(s+tn_S,r_n^+(s))^c\cap H^-} f_n^q(\mathrm{dist}(s+tn_S,y))c(s+tn_S,y)p(y)\ \mathrm{d}y.\end{aligned}$$

We use in the first term the trivial bound $c(s+tn_S,y)\leq 1$ and in the second term the monotonicity of f_n and the bound $b_n = 6\exp(-\delta_n^2 k_n/4)$ on the probability of connectedness when the distance is greater than $r_n^+(s)$ from Lemma 6.2 to obtain

$$\begin{aligned}g(s+tn_S) \leq& \int_{B(s+tn_S,r_n^+(s))\cap H^-} f_n^q(\mathrm{dist}(s+tn_S,y))p(y)\ \mathrm{d}y + b_n f_n^q\left(r_n^+(s)\right)\int_{B(s+tn_S,r_n^+(s))^c\cap H^-} p(y)\ \mathrm{d}y\\ \leq& \int_{B(s+tn_S,r_n^+(s))\cap H^-} f_n^q(\mathrm{dist}(s+tn_S,y))p(y)\ \mathrm{d}y + b_n f_n^q\left(r_n^+(s)\right).\end{aligned}$$

Using a bound on the density in the balls $B(s+tn_S, r_n^+(s))$ we obtain

$$g(s+tn_S) \leq (1+\xi_n)p(s)\int_{B(s+tn_S,r_n^+(s))\cap H^-} f_n^q(\mathrm{dist}(s+tn_S,y))\ \mathrm{d}y + b_n f_n\left(r_n^+(s)\right),$$

and observe that $g(s+tn_S)\leq b_n f_n^q(r_n^+(s))$ if $t > r_n^+(s)$ since in this case $B(s+tn_S, r_n^+(s))\cap H^- = \emptyset$.

That is,

$$\begin{aligned}
h_n(s) &= \int_0^\infty g\left(s+tn_S\right) p\left(s+tn_S\right)\ \mathrm{d}t \\
&\le \int_0^{r_n^+(s)} (1+\xi_n) p(s) \int_{B(s+tn_S, r_n^+(s))\cap H^-} f_n^q(\mathrm{dist}(s+tn_S, y))\ \mathrm{d}y\ p\left(s+tn_S\right)\ \mathrm{d}t \\
&\quad + \int_0^\infty b_n f_n^q\left(r_n^+(s)\right) p\left(s+tn_S\right)\ \mathrm{d}t \\
&\le (1+\xi_n)^2 p^2(s) \int_0^{r_n^+(s)} \int_{B(s+tn_S, r_n^+(s))\cap H^-} f_n^q(\mathrm{dist}(s+tn_S, y))\ \mathrm{d}y\ \mathrm{d}t \\
&\quad + b_n f_n^q\left(r_n^+(s)\right) \int_0^\infty p\left(s+tn_S\right)\ \mathrm{d}t \\
&= (1+\xi_n)^2 p^2(s) F_C^{(q)}\left(r_n^+(s)\right) + b_n f_n^q\left(r_n^+(s)\right) \int_0^\infty p\left(s+tn_S\right)\ \mathrm{d}t.
\end{aligned}$$

Therefore, considering that $\xi_n < 1/2$

$$\begin{aligned}
\int_{S\cap\mathcal{I}_n} h_n(s)\ \mathrm{d}s &\le (1+\xi_n)^2 \int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n^+(s)\right)\ \mathrm{d}s + b_n \int_{S\cap\mathcal{I}_n} f_n^q\left(r_n^+(s)\right) \int_0^\infty p\left(s+tn_S\right)\ \mathrm{d}t\ \mathrm{d}s \\
&\le (1+3\xi_n) \int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s + 3\int_{S\cap\mathcal{I}_n} p^2(s)\left(F_C^{(q)}\left(r_n^+(s)\right) - F_C^{(q)}\left(r_n(s)\right)\right)\ \mathrm{d}s \\
&\quad + b_n f_n^q\left(\inf_{s\in S\cap C} r_n^+(s)\right).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\int_{S\cap\mathcal{I}_n} h_n(s)\ \mathrm{d}s - \int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s &\le 3p_{\max}^2 \sup_{s\in S\cap\mathcal{I}_n}\left(F_C^{(q)}\left(r_n^+(s)\right) - F_C^{(q)}\left(r_n(s)\right)\right)\mathcal{L}_{d-1}(S\cap C) \\
&\quad + 3\xi_n \int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s + b_n f_n^q\left(\inf_{s\in S\cap C} r_n^+(s)\right). \quad (6.5)
\end{aligned}$$

Similarly to the remark above, using the boundedness of the density p, we can replace $b_n f_n^q\left(\inf_{s\in S\cap C} r_n^+(s)\right)$ by

$$p_{\max}\left(F_B^{(q)}(\infty) - F_B^{(q)}(\inf_{s\in S\cap C} r_n(s))\right),$$

which gives a better bound for some weight functions, especially the Gaussian.

Combining Equations (6.4) and (6.5), using the monotonicity of $F_C^{(q)}$ and f we obtain

$$\begin{aligned}
\left|\int_{S\cap\mathcal{I}_n} h_n(s)\ \mathrm{d}s - \int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s\right| &= O\left(\sup_{s\in S\cap\mathcal{I}_n}\left(F_C^{(q)}\left(r_n^+(s)\right) - F_C^{(q)}\left(r_n^-(s)\right)\right)\right) \\
&\quad + O\left((a_n+\xi_n) F_C^{(q)}\left(r_n^{\max}\right) + \min\left\{b_n f_n^q\left(\inf_{x\in C} r_n(s)\right), F_B^{(q)}(\infty) - F_B^{(q)}(\inf_{x\in C} r_n(x))\right\}\right).
\end{aligned}$$

We still have to bound the first term. For some weight functions, especially the Gaussian, we have

$$\sup_{s\in S\cap\mathcal{I}_n}\left(F_C^{(q)}\left(r_n^+(s)\right) - F_C^{(q)}\left(r_n^-(s)\right)\right) \le F_C^{(q)}\left(\infty\right) - F_C^{(q)}\left(\inf_{x\in C} r_n^-(x)\right).$$

For the other weight functions we use

$$\begin{aligned}
F_C^{(q)}\left(r_n^+(s)\right) - F_C^{(q)}\left(r_n^-(s)\right) &= \int_0^{r_n^+(s)} u^d f_n^q(u)\ \mathrm{d}u - \int_0^{r_n^-(s)} u^d f_n^q(u)\ \mathrm{d}u \\
&\le f_n^q\left(r_n^-(s)\right) \int_{r_n^-(s)}^{r_n^+(s)} u^d\ \mathrm{d}u = \frac{1}{d+1} f_n^q\left(r_n^-(s)\right)\left(\left(r_n^+(s)\right)^{d+1} - \left(r_n^-(s)\right)^{d+1}\right) \\
&= \frac{1}{d+1} f_n^q\left(r_n^-(s)\right) r_n^{d+1}(s)\left(\left(\frac{r_n^+(s)}{r_n(s)}\right)^{d+1} - \left(\frac{r_n^-(s)}{r_n(s)}\right)^{d+1}\right).
\end{aligned}$$

Since, with $\xi_n < 1/2$ and $\delta_n < 1$,

$$\begin{aligned}
\left(\frac{r_n^+(s)}{r_n(s)}\right)^{d+1} &= \left(\frac{(1+2\xi_n)(1+2\delta_n)k_n}{(n-1)p(s)\eta_d}\frac{(n-1)p(s)\eta_d}{k_n}\right)^{1+1/d} \\
&= \left((1+2\xi_n)(1+2\delta_n)\right)^{1+1/d} \le 1 + 54\xi_n + 8\delta_n
\end{aligned}$$

and a similar bound holds for the other quotient we have

$$F_C^{(q)}\left(r_n^+(s)\right) - F_C^{(q)}\left(r_n^-(s)\right) = O\left((\xi_n+\delta_n) f_n^q\left(\inf_{x\in C} r_n^-(x)\right)\left(r_n^{\max}\right)^{d+1}\right).$$

With our choice of δ_n we have, considering that $\delta_0 \ge 2$,

$$a_n = 6\exp\left(-\delta_n^2 k_n/3\right) = 6\exp\left(-(8\delta_0\log n)/3\right) \le 6\exp\left(-5\log n\right) = 6/n^5,$$

that is, for n sufficiently large such that $6/n^5 \le \xi_n$, considering that $\xi_n = O(\sqrt[d]{k_n/n})$ and plugging in b_n we have

$$\begin{aligned}
&\left|\int_{S\cap\mathcal{I}_n} h_n(s)\ \mathrm{d}s - \int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s\right| \\
&\quad = O\left(\min\left\{\left(\sqrt[d]{\frac{k_n}{n}} + \delta_n\right) f_n^q\left(\inf_{x\in C} r_n^-(x)\right)\left(r_n^{\max}\right)^{d+1}, F_C^{(q)}(\infty) - F_C^{(q)}\left(\inf_{x\in C} r_n^-(x)\right)\right\}\right) \\
&\quad\quad + O\left(\sqrt[d]{\frac{k_n}{n}} F_C^{(q)}\left(r_n^{\max}\right) + \min\left\{\exp\left(-\delta_n^2\frac{k_n}{4}\right) f_n^q\left(\inf_{x\in C} r_n(s)\right), F_B^{(q)}(\infty) - F_B^{(q)}\left(\inf_{x\in C} r_n(x)\right)\right\}\right).
\end{aligned}$$

Finally, we bound the term in equation (6.3). Setting $\mathcal{R}_n' = C \setminus \mathcal{I}_n$ we have

$$\begin{aligned}
\left|\int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s - \int_{S\cap C} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s\right| &= \int_{S\cap\mathcal{R}_n'} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s \\
&\le p_{\max}^2 F_C^{(q)}\left(\max_{x\in C} r_n(x)\right)\mathcal{L}_{d-1}\left(S\cap\mathcal{R}_n'\right) \\
&\le p_{\max}^2 F_C^{(q)}\left(\max_{x\in C} r_n(x)\right)\mathcal{L}_{d-1}\left(S\cap\mathcal{R}_n\right).
\end{aligned}$$

Using Lemma 6.25 we have $\mathcal{L}_{d-1}\left(S\cap\mathcal{R}_n\right) = O(r_n^{\max})$, and thus

$$\left|\int_{S\cap\mathcal{I}_n} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s - \int_{S\cap C} p^2(s) F_C^{(q)}\left(r_n(s)\right)\ \mathrm{d}s\right| = O\left(F_C^{(q)}\left(\max_{x\in C} r_n(x)\right)\sqrt[d]{\frac{k_n}{n}}\right).$$

Deriving the same bounds for the other halfspace and collecting the three bounds we obtain the result, considering that $k_n/8 \geq \delta_n^2 k_n/8$, $\delta_n^2 k_n/4 \geq \delta_n^2 k_n/8$ and $r_n^{\max} \geq \max_{x\in C} r_n(x)$ due to the monotonicity of $F_C^{(1)}$.

Finally, we discuss the choice of δ_n. With this choice of δ_n we have $\exp\left(-\delta_n^2 k_n/8\right) = n^{-\delta_0}$. Note that this is the fastest convergence rate of δ_n for which the exponential term converges polynomially in $1/n$, which we will need in the proof of the following corollaries. In all the other terms above δ_n has to be chosen as small as possible, so this is the best convergence rate for δ_n. Note further that for this choice of δ_n we require $k_n/\log n \to \infty$, since δ_n has to converge to zero.

Now we proof the bound for the variance term. According to Corollary 3.2.3 from Miller *et al.* [10] the maximum degree of the symmetric k_n-nearest neighbor graph is bounded by $(\tau_d + 1)k_n$, where τ_d denotes the kissing number in dimension d, that is, the maximum number of unit hypershpheres that touch another unit hypersphere without any intersections.

Thus, removing a point from the graph and inserting it in a different place the number of (undirected) edges in the cut can change by at most $2(\tau_d + 1)$. Since we count undirected edges twice we obtain for all types of k-nearest neighbor graphs

$$\left|\mathrm{cut}_n - \mathrm{cut}_n^{(i)}\right| \leq 4(\tau_d + 1)k_n f_n(0),$$

where $\mathrm{cut}_n^{(i)}$ denotes the value of the cut in a graph where exactly one point has been moved to a different place. Thus by McDiarmid's inequality for a suitable constant $\tilde{C} > 0$

$$\Pr\left(\left|\mathrm{cut}_n - \mathbb{E}\left(\mathrm{cut}_n^{(i)}\right)\right| > \varepsilon\right) \leq 2\exp\left(-\frac{2\varepsilon^2}{n\left(4(\tau_d+1)k_n f_n(0)\right)^2}\right) = 2\exp\left(-\frac{\tilde{C}\varepsilon^2}{n k_n^2 f_n^2(0)}\right). \qquad \square$$

The following lemma states bounds on $c(x, y)$, that is the probability of edges between points at x and y, in the cases that we need in the convergence proofs for the cut and the volume.

Lemma 6.2 (kNN radii)**.** *Let G_n be the directed, mutual or symmetric k_n-nearest neighbor graph. Let k_n/n be sufficiently small such that $r_n^{\max} \leq r_\gamma$. Then, if $x, y \in \mathbb{R}^d$ and $\mathrm{dist}(x, y) \geq r_n^{\max}$ we have $c(x, y) \leq 2\exp(-k_n/8)$.*

Set $\xi_n = 2p'_{\max} r_n^{\max}/p_{\min}$ and define $\mathcal{I}_n = \{s \in C \mid B(s, 2r_n^{\max}) \subseteq C\}$. Let n be sufficiently large such that $\xi_n < 1/2$ and let $\delta_n \in (0, 1)$ with $\delta_n \to 0$ for $n \to \infty$ and $k_n\delta_n > 1$ for sufficiently large n.

Let $x = s + tn_S$ with $s \in \mathcal{I}_n \cap S$. If $t \in \mathbb{R}_{\geq 0}$ and $y \in H^-$ or $t \in \mathbb{R}_{\leq 0}$ and $y \in H^+$, and, furthermore, $\mathrm{dist}(x, y) \geq r_n^+(s)$ then $c(x, y) \leq 6\exp\left(-\delta_n^2 k_n/4\right)$. The same holds for $x \in \mathcal{I}_n$ and $y \in C$ with $\mathrm{dist}(x, y) \geq r_n^+(x)$.

Let $x = s + tn_S$ with $t \in [0, r_n^-(s)]$ and $y \in H^-$ or $t \in [-r_n^-(s), 0]$ and $y \in H^+$. If $\mathrm{dist}(x, y) \leq r_n^-(s)$ then $c(x, y) \geq 1 - 6\exp\left(-\delta_n^2 k_n/3\right)$. The same holds for $x \in \mathcal{I}_n$ and $y \in C$ with $\mathrm{dist}(x, y) \leq r_n^-(x)$.

Proof. We first show bounds on the probability of connectedness for the directed k-nearest neighbor graph. These are used in the second part of this proof in order to show bounds for the undirected graph as well. Let D_{ij} denote the event that there exists an edge between x_i and x_j in the directed k-nearest neighbor graph.

First we show the statement concerning the maximal k-nearest neighbor radius. For any $x \in C$ we have, plugging in the definition of $r_n^{\max}$, and using the assumptions that the density p is bounded from below on C and that for balls of a sufficiently small radius around points in C at least a proportion of γ of the volume of the ball is within C,

$$\begin{aligned}\mu(B(x, r_n^{\max}) &= \mu\left(B\left(x, \sqrt[d]{\frac{4}{\gamma p_{\min}\eta_d}\frac{k_n}{n-1}}\right)\right) \geq p_{\min}\mathcal{L}_d\left(B\left(x, \sqrt[d]{\frac{4}{\gamma p_{\min}\eta_d}\frac{k_n}{n-1}}\right) \cap C\right)\\ &\geq p_{\min}\gamma\mathcal{L}_d\left(B\left(x, \sqrt[d]{\frac{4}{\gamma p_{\min}\eta_d}\frac{k_n}{n-1}}\right)\right) = p_{\min}\gamma\frac{4}{\gamma p_{\min}\eta_d}\frac{k_n}{n-1}\eta_d = \frac{4k_n}{n-1}.\end{aligned}$$

Now suppose we fix x_1 and x_2 with $\mathrm{dist}(x_1, x_2) \geq r_n^{\max}$. If U denotes the random variable that counts the number of points $x_3, \ldots, x_n$ in $B(x_1, r_n^{\max})$ we have $U \sim Bin(n-2, \mu(B(x_1, r_n^{\max})))$. Setting $V \sim Bin(n-2, 4k_n/(n-1))$, we certainly have $0 < k_n/(n-2) < 4k_n/(n-1)$ for $n \geq 3$ and thus we obtain with a tail bound for the binomial distribution from Srivastav and Stangier [12], which was first proved in Angluin and Valiant [1],

$$\Pr(D_{12}) = \Pr(U < k_n) \leq \Pr(V < k_n) \leq \exp\left(-\frac{1}{2}\frac{\left((n-2)\frac{4k_n}{n-1} - k_n\right)^2}{(n-2)\frac{4k_n}{n-1}}\right) \leq \exp\left(-\frac{k_n}{8}\right).$$

In the following we show the statements concerning the upper bound $r_n^+(s)$ on the k-nearest neighbor radii of points in regions of relatively homogeneous density. The proof for the lower bound $r_n^-(s)$ is similar and is therefore omitted. Note, however, that the technical condition $\delta_n k_n > 1$ is needed for this case.

First we show how we can bound the density in the balls $B(s, 2r_n^{\max})$: for any $z \in B(s, 2r_n^{\max})$ we have by Taylor's theorem, using the assumptions on differentiability of p,

$$p(s) - 2p'_{\max} r_n^{\max} \leq p(y) \leq p(s) + 2p'_{\max} r_n^{\max},$$

and thus, with $\xi_n = 2p'_{\max} r_n^{\max}/p_{\min}$,

$$(1-\xi_n)p(s) \leq p(y) \leq (1+\xi_n)p(s).$$

These bounds are used below to bound the probability mass of balls within $B(s, 2r_n^{\max})$.

Now, we bound the probability mass in $B(x, \mathrm{dist}(x, y))$ and $B(y, \mathrm{dist}(x, y))$ from below, when $\mathrm{dist}(x, y) \geq r_n^+(s)$. We first observe that, for $\xi_n < 1/2$, $\delta_n < 1$, and using the lower bound on p,

$$r_n^+(s) = \sqrt[d]{\frac{(1+2\xi_n)(1+\delta_n)k_n}{(n-1)p(s)\eta_d}} \leq \sqrt[d]{\frac{4k_n}{(n-1)\gamma p_{\min}\eta_d}} = r_n^{\max}.$$

Suppose $t = \mathrm{dist}(x, s) \leq r_n^+(s)$. Then

$$\mu\left(B(x, \mathrm{dist}(x, y))\right) \geq \mu\left(B(x, r_n^+(s))\right)$$

with $B(x, r_n^+(s)) \subseteq B(s, 2r_n^{\max})$. If $t = \mathrm{dist}(x, s) > r_n^+(s)$ we know that $\mathrm{dist}(x, y) > \mathrm{dist}(x, s)$, since x and y are on different sides of the hyperplane S. We set $x' = s + r_n^+(s) n_S$, that is the point on the line connecting s and x with distance $r_n^+(s)$ from s. Then, by construction, $B(x', r_n^+(s)) \subseteq B(s, 2r_n^{\max})$ and $B(x', r_n^+(s)) \subseteq B(x, \mathrm{dist}(x, s))$. Thus

$$\mu\left(B(x, \mathrm{dist}(x, y))\right) \geq \mu\left(B(x, \mathrm{dist}(x, s))\right) \geq \mu\left(B(x', r_n^+(s))\right).$$

Now we consider balls around the other point y. First, suppose $\mathrm{dist}(y, s) = r_n^+(s)$. Then

$$\mu\left(B(y, \mathrm{dist}(x, y))\right) \geq \mu\left(B(y, r_n^+(s))\right)$$

with $B(y, r_n^+(s)) \subseteq B(s, 2r_n^{\max})$.

If $\mathrm{dist}(y, s) > r_n^+(s)$ we set $y' = s + (y-s)/\|y-s\|$, that is the point on the line connecting s and y with distance $r_n^+(s)$ from s. Then, by construction, $B(y', r_n^+(s)) \subseteq B(s, 2r_n^{\max})$ and $B(y', r_n^+(s)) \subseteq B(y, \mathrm{dist}(y, s))$. Since x and y are on different sides of S we have $\mathrm{dist}(y, s) \leq \mathrm{dist}(y, x)$. Therefore

$$\mu\left(B(y, \mathrm{dist}(y, x))\right) \geq \mu\left(B(y, \mathrm{dist}(y, s))\right) \geq \mu\left(B(y', r_n^+(s))\right).$$

We show how to bound $\mu(B(x, r_n^+(s)))$. The same bound can be shown for the probability mass in $B(x', r_n^+(s))$, $B(y, r_n^+(s))$ and $B(y', r_n^+(s))$, since all of these balls lie in $B(s, 2r_n^{\max})$. We have, since $\xi_n < 1/2$,

$$\begin{aligned}\mu\left(B(x, r_n^+(s))\right) &\geq (1-\xi_n)p(s)\eta_d \left(r_n^+(s)\right)^d = (1-\xi_n)p(s)\eta_d \frac{(1+2\xi_n)(1+\delta_n)k_n}{(n-1)p(s)\eta_d} \\ &= (1-\xi_n)(1+2\xi_n)(1+\delta_n)\frac{k_n}{n-1} \geq (1+\delta_n)\frac{k_n}{n-1}.\end{aligned}$$

Let $U_x^+ \sim Bin\left(n-2, \mu\left(B(x, r_n^+(s))\right)\right)$ and $V_x^+ \sim Bin\left(n-2, (1+\delta_n)k_n/(n-1)\right)$. Then, we have for $(n-2)\delta_n > 1$

$$0 \leq \frac{k_n}{n-2} = \left(1 + \frac{1}{n-2}\right)\frac{k_n}{n-1} < (1+\delta_n)\frac{k_n}{n-1}$$

and thus, by the tail bound from Angluin and Valiant [1],

$$\Pr(D_{12}) = \Pr\left(U_x^+ < k\right) \leq \Pr\left(V_x^+ < k\right) \leq \exp\left(-\frac{1}{2}\frac{\left((n-2)(1+\delta_n)\frac{k_n}{n-1} - k_n\right)^2}{(n-2)(1+\delta_n)\frac{k_n}{n-1}}\right).$$

We have

$$\begin{aligned}\left((n-2)(1+\delta_n)\frac{k_n}{n-1} - k_n\right)^2 &= \left(\left(1 - \frac{1}{n-1}\right)(1+\delta_n)k_n - k_n\right)^2 \\ &= \left(\delta_n k_n - \frac{1+\delta_n}{n-1}k_n\right)^2 \geq \delta_n^2 k_n^2 - 2\delta_n(1+\delta_n)\frac{k_n}{n-1}k_n \geq \delta_n^2 k_n^2 - 4\delta_n k_n\end{aligned}$$

and

$$(n-2)(1+\delta_n)\frac{k_n}{n-1} = \left(1 - \frac{1}{n-1}\right)(1+\delta_n)k_n \leq 2k_n,$$

and thus, using $\delta_n < 1$,

$$\Pr(D_{12}) \leq \exp\left(-\frac{\delta_n^2 k_n^2 - 4\delta_n k_n}{4k_n}\right) \leq \exp\left(-\frac{\delta_n^2 k_n}{4} + \delta_n\right) \leq 3\exp\left(-\frac{\delta_n^2 k_n}{4}\right).$$

This analysis can be carried over to the case $t > r_n^+(s)$ and the same bound holds.

The same bound holds also for $\Pr(D_{21})$, since the same bounds for the probability mass in the balls $B(y, r_n^+(s))$ and $B(y', r_n^+(s))$ hold.

In the final step of the proof we use the results derived so far to show the results for the undirected k-nearest neighbor graphs. For the mutual kNN graph we have by definition $\Pr(C_{12}) = \Pr(C_{21}) = \Pr(D_{12} \cap D_{21})$. Thus, clearly, $\Pr(C_{12}) \leq \Pr(D_{12})$ and

$$\begin{aligned}\Pr(C_{12}) &= \Pr(D_{12} \cap D_{21}) = 1 - \Pr(D_{12}^c \cup D_{21}^c) \geq 1 - \Pr(D_{12}^c) - \Pr(D_{21}^c) \\ &= 1 - (1 - \Pr(D_{12})) - (1 - \Pr(D_{21})) = \Pr(D_{12}) + \Pr(D_{21}) - 1.\end{aligned}$$

This implies

$$\Pr(D_{12} \mid x_1 = x, x_2 = y) + \Pr(D_{21} \mid x_1 = x, x_2 = y) - 1 \leq \Pr(C_{12} \mid x_1 = x, x_2 = y) \leq \Pr(D_{12} \mid x_1 = x, x_2 = y).$$

For the symmetric kNN graph we have $\Pr(C_{12}) = \Pr(C_{21}) = \Pr(D_{12} \cup D_{21})$, which implies $\Pr(C_{12}) \geq \Pr(D_{12})$ and by a union bound $\Pr(C_{12}) \leq \Pr(D_{12}) + \Pr(D_{21})$. Therefore

$$\Pr(D_{12} \mid x_1 = x, x_2 = y) \leq \Pr(C_{12} \mid x_1 = x, x_2 = y) \leq \Pr(D_{12} \mid x_1 = x, x_2 = y) + \Pr(D_{21} \mid x_1 = x, x_2 = y).$$

Thus, using the worse out of the two possible bounds we obtain for both undirected kNN graph types

$$\begin{aligned}\Pr(D_{12} \mid x_1 = x, x_2 = y) + \Pr(D_{21} \mid x_1 = x, x_2 = y) - 1 \le \Pr(C_{12} \mid x_1 = x, x_2 = y)&\\ \le \Pr(D_{12} \mid x_1 = x, x_2 = y) + \Pr(D_{21} \mid x_1 = x, x_2 = y).&\end{aligned}$$

Plugging in the results for $\Pr(D_{12})$ and $\Pr(D_{21})$ in the cases studied above, we obtain the result. □

Lemma 6.3 (integral over caps)**.** *Let the general assumptions hold and let $f : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ be a monotonically decreasing function and $s \in S$. Then we have for any $R \in \mathbb{R}_{>0}$*

$$\int_0^R \int_{B(s+tn_S,R)\cap H^-} f(\mathrm{dist}(s + tn_S, y))\ \mathrm{d}y\ \mathrm{d}t = \eta_{d-1} \int_{u=0}^R u^d f(u)\ \mathrm{d}u$$

and

$$\int_{-R}^0 \int_{B(s+tn_S,R)\cap H^-} f(\mathrm{dist}(s + tn_S, y))\ \mathrm{d}y\ \mathrm{d}t = \eta_{d-1} \int_{u=0}^R u^d f(u)\ \mathrm{d}u.$$

Proof. By a translation and rotation of our coordinate system in $\mathbb{R}^d$ such that $s + tn_S$ is the origin and $-n_S$ the first coordinate axis we obtain for $t \ge 0$

$$\begin{aligned}\int_{B(s+tn_S,R)\cap H^-} f(\mathrm{dist}(s + tn_S, y))\ \mathrm{d}y\ &= \int_{B(0,R)\cap\{z_1\ge t\}} f(\mathrm{dist}(0, z))\ \mathrm{d}z\\ &= \int_{z_1=t}^R \int_{\{z_2^2+\ldots+z_d^2\le R^2-z_1^2\}} f(\mathrm{dist}(0, z))\ \mathrm{d}z_d \ldots\ \mathrm{d}z_2\ \mathrm{d}z_1\\ &= \int_{z_1=t}^R \int_{\{z_2^2+\ldots+z_d^2\le R^2-z_1^2\}} f\left(\sqrt{z_1^2 + \ldots + z_d^2}\right)\ \mathrm{d}z_d \ldots\ \mathrm{d}z_2\ \mathrm{d}z_1\\ &= \int_{z_1=t}^R A(z_1)\ \mathrm{d}z_1,\end{aligned}$$

where we have set

$$A(r) = \int_{\{z_2^2+\ldots+z_d^2\le R^2-r^2\}} f\left(\sqrt{r^2 + z_2^2 + \ldots + z_d^2}\right) \mathrm{d}z_d \ldots \mathrm{d}z_2.$$

Thus,

$$\begin{aligned}\int_{t=0}^R \int_{B(s+tn_S,R)\cap H^-} f(\mathrm{dist}(s + tn_S, y))\ \mathrm{d}y\ \mathrm{d}t &= \int_{t=0}^R \int_{r=t}^R A(r)\ \mathrm{d}r\ \mathrm{d}t\\ &= \int_{r=0}^R \int_{t=0}^r A(r)\ \mathrm{d}t\ \mathrm{d}r = \int_{r=0}^R A(r) \int_{t=0}^r\ \mathrm{d}t\ \mathrm{d}r = \int_{r=0}^R rA(r)\ \mathrm{d}r.\end{aligned}$$

Similarly, by the same translation and a rotation such that n_S is the first coordinate axis we obtain for $t < 0$

$$\begin{aligned}\int_{B(s+tn_S,R)\cap H^+} f(\mathrm{dist}(s + tn_S, y))\ \mathrm{d}y &= \int_{B(0,R)\cap\{z_1\ge -t\}} f(\mathrm{dist}(0, z))\ \mathrm{d}z\\ &= \int_{z_1=-t}^R A(z_1)\ \mathrm{d}z_1,\end{aligned}$$

that is,

$$\int_{-R}^{0}\int_{B(s+tn_S,R)\cap H^-} f(\mathrm{dist}(s+tn_S,y))\ \mathrm{d}y\ \mathrm{d}t = \int_{t=-R}^{0}\int_{r=-t}^{R} A(r)\ \mathrm{d}r\ \mathrm{d}t$$
$$= \int_{r=0}^{R}\int_{t=-r}^{0} A(r)\ \mathrm{d}t\ \mathrm{d}r = \int_{r=0}^{R} A(r)\int_{t=-r}^{0}\ \mathrm{d}t\ \mathrm{d}r = \int_{r=0}^{R} rA(r)\ \mathrm{d}r.$$

Therefore, both the integrals we want to compute are equal to $\int_{r=0}^{R} rA(r)\ \mathrm{d}r$ which we will treat in the following. First we are going to compute the $(d-1)$-dimensional integral $A(r)$. Setting $\tilde{f}_r(s) = f(\sqrt{r^2+s^2})$ we can write $A(r)$ as the following integral in $\mathbb{R}^{d-1}$:

$$A(r) = \int_{\{x_1^2+\ldots+x_{d-1}^2 \le R^2-r^2\}} f\left(\sqrt{r^2+x_1^2+\ldots+x_{d-1}^2}\right) \mathrm{d}x_{d-1}\ldots\mathrm{d}x_1$$
$$= \int_{\|x\|\le\sqrt{R^2-r^2}} \tilde{f}_r(\|x\|)\ \mathrm{d}x = \int_0^{\sqrt{R^2-r^2}} (d-1)\eta_{d-1}s^{d-2}\tilde{f}_r(s)\ \mathrm{d}s$$
$$= (d-1)\eta_{d-1}\int_0^{\sqrt{R^2-r^2}} s^{d-2} f\left(\sqrt{r^2+s^2}\right)\ \mathrm{d}s.$$

Plugging in this expression for $A(r)$ we obtain

$$\int_{r=0}^{R} rA(r)\ \mathrm{d}r = (d-1)\eta_{d-1}\int_{r=0}^{R}\int_{s=0}^{\sqrt{R^2-r^2}} rs^{d-2} f\left(\sqrt{r^2+s^2}\right)\ \mathrm{d}s\ \mathrm{d}r.$$

Substituting with polar coordinates $(r,s) = (u\cos\theta, u\sin\theta)$ with $u\in[0,R]$ and $\theta\in[0,\pi/2]$, we have

$$\int_{r=0}^{R}\int_{s=0}^{\sqrt{R^2-r^2}} rs^{d-2} f\left(\sqrt{r^2+s^2}\right)\ \mathrm{d}s\ \mathrm{d}r = \int_{u=0}^{R}\int_{\theta=0}^{\pi/2} u\cos\theta u^{d-2}\sin^{d-2}\theta f(u)u\ \mathrm{d}\theta\ \mathrm{d}u$$
$$= \int_{u=0}^{R} u^d f(u)\int_{\theta=0}^{\pi/2}\cos\theta\sin^{d-2}\theta\ \mathrm{d}\theta\ \mathrm{d}u$$
$$= \int_{u=0}^{R} u^d f(u)\left[\frac{1}{d-1}\sin^{d-1}\theta\right]_{\theta=0}^{\pi/2}\ \mathrm{d}u = \frac{1}{d-1}\int_{u=0}^{R} u^d f(u)\ \mathrm{d}u.$$

Combining the last two equations we obtain

$$\int_{r=0}^{R} rA(r)\ \mathrm{d}r = \eta_{d-1}\int_{u=0}^{R} u^d f(u)\ \mathrm{d}u.$$

Note that the integral exists due to the monotonicity of f and the compactness of the interval $[0,R]$. $\square$

Corollary 6.4 (unweighted kNN-graph)**.** *Let G_n be the unweighted k-nearest neighbor graph and let f_n be the unit weight function. Then*

$$\left|\frac{1}{nk_n}\sqrt[d]{\frac{n}{k_n}}\,\mathrm{cut}_n - \frac{2\eta_{d-1}}{(d+1)\eta_d^{1+1/d}}\int_S p^{1-1/d}(s)\ \mathrm{d}s\right| = O\left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}}\right)$$

and, for a suitable constant $\tilde{C}>0$

$$\Pr\left(\left|\frac{1}{nk_n}\sqrt[d]{\frac{n}{k_n}}\,\mathrm{cut}_n - \mathbb{E}\left(\frac{1}{nk_n}\sqrt[d]{\frac{n}{k_n}}\,\mathrm{cut}_n\right)\right| > \varepsilon\right) \le 2\exp\left(-\tilde{C}\varepsilon^2 n^{1-2/d}k_n^{2/d}\right).$$

Proof. With Lemma 6.22 we have for any $s \in S \cap C$, plugging in the definition of $r_n(s)$,

$$F_C^{(1)}(r_n(s)) = \frac{\eta_{d-1}}{d+1}\left(\frac{k_n}{(n-1)p(s)\eta_d}\right)^{1+1/d} = \frac{\eta_{d-1}}{(d+1)\eta_d^{1+1/d}}\left(\frac{k_n}{n-1}\right)^{1+1/d} p^{-1-1/d}(s).$$

Therefore,

$$2\int_{S\cap C} p^2(s)F_C^{(1)}(r_n(s))\ \mathrm{d}s = 2\int_{S\cap C} p^2(s)\frac{\eta_{d-1}}{(d+1)\eta_d^{1+1/d}}\left(\frac{k_n}{n-1}\right)^{1+1/d} p^{-1-1/d}(s)\ \mathrm{d}s$$
$$= \left(\frac{k_n}{n-1}\right)^{1+1/d}\frac{2\eta_{d-1}}{(d+1)\eta_d^{1+1/d}}\int_S p^{1-1/d}(s)\ \mathrm{d}s.$$

Multiplying this term with the factor $(k_n/(n-1))^{-1-1/d}$ we obtain a constant limit. We now multiply the inequality for the bias term in Proposition 6.1 with this factor and deal with the error terms.

For the first term we derive an upper bound on $F_C^{(1)}(r_n^{\max})$ similarly to above and obtain

$$\left(\frac{k_n}{n-1}\right)^{-1-1/d} F_C^{(1)}(r_n^{\max})\sqrt[d]{\frac{k_n}{n}} = O\left(\sqrt[d]{\frac{k_n}{n}}\right).$$

For the second error term we have with $\delta_0 = 3$ and $f_n \equiv 1$

$$\left(\frac{k_n}{n-1}\right)^{-1-1/d} n^{-\delta_0} f_n\left(\inf_{x\in C} r_n(x)\right) \le n^2 n^{-3} = O\left(n^{-1}\right).$$

For the last error term we have

$$\left(\frac{k_n}{n-1}\right)^{-1-1/d}\left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}}\right) f_n\left(\inf_{x\in C} r_n^-(x)\right)\left(\frac{k_n}{n}\right)^{1+1/d} = O\left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}}\right).$$

Thus, considering that $n^{-1} \le \sqrt[d]{k_n/n}$, we obtain

$$\left|\frac{1}{nk_n}\sqrt[d]{\frac{n-1}{k_n}}\,\mathrm{cut}_n - \frac{2\eta_{d-1}}{(d+1)\eta_d^{1+1/d}}\int_S p^{1-1/d}(s)\ \mathrm{d}s\right| = \left(\frac{n-1}{k_n}\right)^{1+1/d}\left|\frac{\mathrm{cut}_n}{n(n-1)} - 2\int_S p^2(s)F_C^{(1)}(r_n(s))\ \mathrm{d}s\right|$$
$$= O\left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}}\right).$$

For the *variance term* we have with Proposition 6.1 and $f_n(0) = 1$

$$\Pr\left(\left|\frac{1}{nk_n}\sqrt[d]{\frac{n-1}{k_n}}\,\mathrm{cut}_n - \mathbb{E}\left(\frac{1}{nk_n}\sqrt[d]{\frac{n-1}{k_n}}\,\mathrm{cut}_n\right)\right| > \varepsilon\right) = \Pr\left(\left|\mathrm{cut}_n - \mathbb{E}\left(\mathrm{cut}_n\right)\right| > nk_n\sqrt[d]{\frac{k_n}{n-1}}\varepsilon\right)$$
$$\le 2\exp\left(-\tilde{C}\frac{\varepsilon^2 n^2 k_n^2 (k_n/(n-1))^{2/d}}{nk_n^2 f_n^2(0)}\right) \le 2\exp\left(-\tilde{C}\varepsilon^2 n^{1-2/d} k_n^{2/d}\right).$$

Since $1/n = O(\sqrt[d]{k_n/n})$ we can change $\sqrt[d]{(n-1)/k_n}$ in the scaling factor to $\sqrt[d]{n/k_n}$ without changing the convergence rate. $\square$

Corollary 6.5 (Gaussian weights and $1/\sigma_n(k_n/n)^{1/d} \to 0$)**.** *Let G_n be the k-nearest neighbor graph with Gaussian weight function and let $1/\sigma_n(k_n/n)^{1/d} \to 0$. Then*

$$\left|\mathbb{E}\left(\frac{\sigma_n^d}{nk_n}\sqrt[d]{\frac{n}{k_n}}\,\mathrm{cut}_n\right) - \frac{2\eta_{d-1}\eta_d^{-1-1/d}}{(d+1)(2\pi)^{d/2}}\int_S p^{1-1/d}(s)\ \mathrm{d}s\right| = O\left(\left(\frac{1}{\sigma_n}\sqrt[d]{\frac{k_n}{n}}\right)^2 + \sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}}\right)$$

and, for a suitable constant $\tilde{C} > 0$

$$\Pr\left(\left|\frac{1}{nk_n}\sqrt[d]{\frac{n}{k_n}}\,\mathrm{cut}_n - \mathbb{E}\left(\frac{1}{nk_n}\sqrt[d]{\frac{n}{k_n}}\,\mathrm{cut}_n\right)\right| > \varepsilon\right) \le 2\exp\left(-\tilde{C}\varepsilon^2 n^{1-2/d}k_n^{2/d}\right).$$

Proof. According to Lemma 6.23 we have for all $s \in S \cap C$

$$\left|\frac{\sigma_n^{qd}}{r_n^{d+1}(s)}F_C^{(q)}(r_n(s)) - \frac{\eta_{d-1}}{(d+1)(2\pi)^{qd/2}}\right| \le 2\left(\frac{r_n(s)}{\sigma_n}\right)^2.$$

Plugging in $r_n(s) = \sqrt[d]{k_n/((n-1)\eta_d p(s))}$ we obtain

$$\left|\sigma_n^{qd}\left(\frac{n-1}{k_n}\right)^{1+1/d}(\eta_d p(s))^{1+1/d}\,F_C^{(q)}(r_n(s)) - \frac{\eta_{d-1}}{(d+1)(2\pi)^{qd/2}}\right| \le 2\left(\frac{1}{\sigma_n}\sqrt[d]{\frac{k_n}{(n-1)\eta_d p(s)}}\right)^2$$

and therefore

$$\left|\sigma_n^{qd}\left(\frac{n-1}{k_n}\right)^{1+1/d}F_C^{(q)}(r_n(s)) - \frac{\eta_{d-1}\eta_d^{-1-1/d}}{(d+1)(2\pi)^{qd/2}}p(s)^{-1-1/d}\right|$$
$$\le 2(\eta_d p(s))^{-1-1/d}\left(\frac{1}{\sigma_n}\sqrt[d]{\frac{k_n}{(n-1)\eta_d p(s)}}\right)^2 \le \tilde{C}_1\left(\frac{k_n}{\sigma_n^d n}\right)^{2/d}$$

for a suitable constant $\tilde{C}_1 > 0$. Therefore

$$\begin{aligned}
&\left|\sigma_n^d\left(\frac{n-1}{k_n}\right)^{1+1/d} 2\int_{S\cap C} p^2(s)F_C^{(1)}(r_n(s))\ \mathrm{d}s - \frac{2\eta_{d-1}\eta_d^{-1-1/d}}{(2\pi)^{d/2}(d+1)}\int_S p^{1-1/d}(s)\ \mathrm{d}s\right| \\
&= \left|\sigma_n^d\left(\frac{n-1}{k_n}\right)^{1+1/d} 2\int_{S\cap C} p^2(s)F_C^{(1)}(r_n(s))\ \mathrm{d}s - 2\int_S p^2(s)\frac{\eta_{d-1}\eta_d^{-1-1/d}}{(2\pi)^{d/2}(d+1)}p^{-1-1/d}(s)\ \mathrm{d}s\right| \\
&\le 2\int_{S\cap C} p^2(s)\left|\sigma_n^d\left(\frac{n-1}{k_n}\right)^{1+1/d}F_C^{(1)}(r_n(s)) - \frac{\eta_{d-1}\eta_d^{-1-1/d}}{(2\pi)^{d/2}(d+1)}p^{-1-1/d}(s)\right|\ \mathrm{d}s \\
&\le 2\int_{S\cap C} p^2(s)\tilde{C}_1\left(\frac{k_n}{n\sigma_n^d}\right)^{2/d}\ \mathrm{d}s = 2\tilde{C}_1\left(\frac{k_n}{n\sigma_n^d}\right)^{2/d}p_{\max}^2\mathcal{L}_{d-1}(S\cap C).
\end{aligned}$$

Now, we consider the error terms of Proposition 6.1. For the first one we have, using that $F_C^{(1)}(r_n^{\max}) = O((r_n^{\max})^{d+1}/\sigma_n^d)$ and, furthermore, $r_n^{\max} = O(\sqrt[d]{k_n/(n-1)})$

$$\sigma_n^d\left(\frac{n-1}{k_n}\right)^{1+1/d}F_C^{(1)}(r_n^{\max})\sqrt[d]{\frac{k_n}{n}} = O\left(\sigma_n^d\left(\frac{n-1}{k_n}\right)^{1+1/d}\sigma_n^{-d}\left(\frac{k_n}{n-1}\right)^{1+1/d}\sqrt[d]{\frac{k_n}{n}}\right) = O\left(\sqrt[d]{\frac{k_n}{n}}\right).$$

For the second error term we have with $\delta_0 = 4$

$$\sigma_n^d \left(\frac{n-1}{k_n}\right)^{1+1/d} n^{-\delta_0} f_n \left(\inf_{x\in C} r_n(x)\right) \le \sigma_n^d n^2 n^{-4} \frac{1}{(2\pi)^{d/2}\sigma_n^d} = O\left(n^{-2}\right).$$

For the third error term we have with $f_n(0) = O(\sigma_n^{-d})$ and the monotonicity of f_n

$$\sigma_n^d \left(\frac{n-1}{k_n}\right)^{1+1/d} \left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}}\right) f_n \left(\inf_{x\in C} r_n^-(x)\right) \left(\frac{k_n}{n}\right)^{1+1/d} = O\left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}}\right).$$

For the *variance term* we have with Proposition 6.1 and $f_n(0) = (2\pi)^{-d/2}\sigma_n^{-d}$ for a suitable constant $\tilde{C}' > 0$

$$\begin{aligned}\Pr\left(\left|\frac{\sigma_n^d}{nk_n}\sqrt[d]{\frac{n-1}{k_n}}\operatorname{cut}_n - \mathbb{E}\left(\frac{\sigma_n^d}{nk_n}\sqrt[d]{\frac{n-1}{k_n}}\operatorname{cut}_n\right)\right| > \varepsilon\right) &= \Pr\left(\left|\operatorname{cut}_n - \mathbb{E}\left(\operatorname{cut}_n\right)\right| > \frac{nk_n}{\sigma_n^d}\sqrt[d]{\frac{k_n}{n-1}}\varepsilon\right)\\
&\le 2\exp\left(-\tilde{C}'\frac{\varepsilon^2 n^2 k_n^2 \sigma_n^{-2d}(k_n/(n-1))^{2/d}}{nk_n^2 f_n^2(0)}\right) \le 2\exp\left(-\tilde{C}\varepsilon^2 n^{1-2/d} k_n^{2/d}\right),\end{aligned}$$

where we have set $\tilde{C} = (2\pi)^d \tilde{C}'$.

Since $1/n = O(\sqrt[d]{k_n/n})$ we can change $\sqrt[d]{(n-1)/k_n}$ in the scaling factor to $1/(nk_n)\sqrt[d]{n/k_n}$ without changing the convergence rate. □

Corollary 6.6 (Gaussian weights and $\sigma_n(k_n/n)^{-1/d} \to 0$)**.** *We consider the* kNN *graph with Gaussian weight function. Let $\sigma_n(k_n/n)^{-1/d} \to 0$ and $n\sigma_n^{d+1} \to \infty$ for $n \to \infty$. Then there exists a constant $\tilde{C} > 0$ such that*

$$\left|\mathbb{E}\left(\frac{1}{n^2\sigma_n}\operatorname{cut}_n\right) - \frac{2}{\sqrt{2\pi}}\int_S p^2(s)\,\mathrm{d}s\right| = O\left(\sqrt[d]{\frac{k_n}{n}} + \frac{1}{\sigma_n}\exp\left(-\tilde{C}\left(\frac{1}{\sigma_n}\sqrt[d]{\frac{k_n}{n}}\right)^2\right)\right).$$

Furthermore, suppose $\sqrt[d]{k_n/n} \ge \sigma_n^\alpha$ for an $\alpha \in (0,1)$ and n sufficiently large. Then there exist non-negative random variables $D_n^{(1)}, D_n^{(2)}$ such that

$$\left|\frac{\operatorname{cut}_n}{n^2\sigma_n} - \mathbb{E}\left(\frac{\operatorname{cut}_n}{n^2\sigma_n}\right)\right| = O(\sigma_n) + D_n^{(1)} + D_n^{(2)},$$

with $\Pr(D_n^{(1)} > \varepsilon) \le 2\exp(\tilde{C}_2 n\sigma_n^{d+1}\varepsilon^2)$ for a constant $\tilde{C}_2 > 0$, and $\Pr(D_n^{(2)} > \sigma_n) \le 1/n^3$.

Proof. With Lemma 6.24 we have for for $\sqrt[d]{k_n/n}/\sigma_n$ sufficiently large

$$\begin{aligned}\left|\frac{2}{\sigma_n}\int_{S\cap C} p^2(s) F_C^{(1)}(r_n(s))\,\mathrm{d}s - \frac{2}{\sqrt{2\pi}}\int_S p^2(s)\,\mathrm{d}s\right| &\le 2\int_{S\cap C} p^2(s)\left|\frac{1}{\sigma_n}F_C^{(1)}(r_n(s)) - \frac{1}{\sqrt{2\pi}}\right|\,\mathrm{d}s\\
&= O\left(\exp\left(-\frac{1}{4(p_{\max}\eta_d)^{2/d}}\left(\frac{1}{\sigma_n}\sqrt[d]{\frac{k_n}{n}}\right)^2\right)\right),\end{aligned}$$

where we use that p and $\mathcal{L}_{d-1}(S\cap C)$ are bounded.

Now we bound the error terms from Proposition 6.1 of the other difference

$$\left|\mathbb{E}\left(\frac{1}{n(n-1)\sigma_n}\operatorname{cut}_n\right) - \frac{2}{\sigma_n}\int_{S\cap C} p^2(s) F_C^{(1)}(r_n(s))\,\mathrm{d}s\right|.$$

For the first one we observe that with Lemma 6.24 we have $F_C^{(1)}(r_n^{\max}) = O(\sigma_n)$ and therefore $\sigma_n^{-1} F_C^{(1)}(r_n^{\max}) \sqrt[d]{k_n/n} = O(\sqrt[d]{k_n/n})$.

For the second one we have with Lemma 6.24

$$\frac{1}{\sigma_n}\left(F_B^{(1)}(\infty) - F_B^{(1)}(\inf_{x\in C} r_n(x))\right) = O\left(\frac{1}{\sigma_n}\exp\left(-\frac{1}{4(p_{\max}\eta_d)^{2/d}}\left(\frac{1}{\sigma_n}\sqrt[d]{\frac{k_n}{n}}\right)^2\right)\right).$$

For the third error term we observe that if n is sufficiently large such that $\delta_n \le 1/2$ and $\xi_n \le 1/4$ then for all $x \in C$,

$$r_n^-(x) = \sqrt[d]{\frac{(1-2\xi_n)(1-\delta_n)k_n}{(n-1)p(x)\eta_d}} \ge \sqrt[d]{\frac{k_n}{4p_{\max}\eta_d n}}.$$

Then we have with Lemma 6.24

$$\frac{1}{\sigma_n}\left(F_C^{(1)}(\infty) - F_C^{(1)}(\inf_{x\in C} r_n^-(x))\right) = O\left(\exp\left(-\frac{1}{4(4p_{\max}\eta_d)^{2/d}}\left(\frac{1}{\sigma_n}\sqrt[d]{\frac{k_n}{n}}\right)^2\right)\right).$$

Now we proof the bound for the variance term. Unfortunately, the bound in Proposition 6.1 based on McDiarmid's inequality does not give good results. Therefore we proof a bound on the variance term directly. We set $\overline{\mathrm{cut}_n}$ to be the cut_n in the complete graph with Gaussian weights on the sample and we set $\mathrm{cut}_n^{\mathrm{miss}}$ to be sum of the weights of the edges that are in the cut but not in the kNN graph. Then $\mathrm{cut}_n = \overline{\mathrm{cut}_n} - \mathrm{cut}_n^{\mathrm{miss}}$ and we have

$$\begin{aligned}\left|\frac{\mathrm{cut}_n}{n(n-1)\sigma_n} - \mathbb{E}\left(\frac{\mathrm{cut}_n}{n(n-1)\sigma_n}\right)\right| &= \left|\frac{\overline{\mathrm{cut}_n}}{n(n-1)\sigma_n} - \mathbb{E}\left(\frac{\overline{\mathrm{cut}_n}}{n(n-1)\sigma_n}\right) - \left(\frac{\mathrm{cut}_n^{\mathrm{miss}}}{n(n-1)\sigma_n} - \mathbb{E}\left(\frac{\mathrm{cut}_n^{\mathrm{miss}}}{n(n-1)\sigma_n}\right)\right)\right| \\ &\le \left|\frac{\overline{\mathrm{cut}_n}}{n(n-1)\sigma_n} - \mathbb{E}\left(\frac{\overline{\mathrm{cut}_n}}{n(n-1)\sigma_n}\right)\right| + \frac{\mathrm{cut}_n^{\mathrm{miss}}}{n(n-1)\sigma_n} + \mathbb{E}\left(\frac{\mathrm{cut}_n^{\mathrm{miss}}}{n(n-1)\sigma_n}\right).\end{aligned}$$

The first deviation term is dealt with in Corollary 6.14.

We denote with $\mathcal{D}$ the event that the k-nearest neighbor radius of all the points is greater than $r_n^{\min} = \sqrt[d]{k_n/(2p_{\max}\eta_d(n-1))}$. One can show similarly to the proof of Lemma 6.2 that $\Pr(\mathcal{D}^c) \le \exp(\log n - k_n/8)$ and thus $\Pr(\mathcal{D}^c) \le 1/n^3$ for sufficiently large n, since $k_n/\log n \to \infty$. If $\mathcal{D}$ holds, all the edges in $\mathrm{cut}_n^{\mathrm{miss}}$ must have weight lower than $f_n(r_n^{\min})$, whereas if $\mathcal{D}^c$ holds the maximum edge weight is $f_n(0)$. There are $n(n-1)$ possible edges and thus

$$\begin{aligned}\mathbb{E}\left(\frac{\mathrm{cut}_n^{\mathrm{miss}}}{n(n-1)\sigma_n}\right) &\le \frac{1}{n(n-1)\sigma_n} n(n-1) f_n(0)\Pr(\mathcal{D}^c) + \frac{1}{n(n-1)\sigma_n} n(n-1) f_n(r_n^{\min})\Pr(\mathcal{D}) \\ &= O\left(\frac{1}{\sigma_n^{d+1}}\frac{1}{n^3} + \frac{1}{\sigma_n^{d+1}}\exp\left(-\frac{(r_n^{\min})^2}{2\sigma_n^2}\right)\right) = O\left(\frac{1}{n^2} + \frac{1}{\sigma_n^{d+1}}\exp\left(-\frac{(r_n^{\min})^2}{2\sigma_n^2}\right)\right),\end{aligned}$$

since $n\sigma_n^{d+1} \to \infty$ for $n \to \infty$.

Under the condition $\sqrt[d]{k_n/n} \ge \sigma_n^\alpha$ with $\alpha \in (0,1)$ we have for sufficiently large n and a suitable constant $\tilde{C}_1$

$$\frac{1}{\sigma_n^{d+1}}\exp\left(-\frac{(r_n^{\min})^2}{2\sigma_n^2}\right) \le \frac{1}{\sigma_n^{d+1}}\exp\left(-\tilde{C}_1\sigma_n^{2(\alpha-1)}\right) \le \sigma_n,$$

where we use that the exponential term converges to zero faster than any power of σ_n.

For the other term we clearly have for n sufficiently large

$$\Pr\left(\frac{\mathrm{cut}_n^{\mathrm{miss}}}{n(n-1)\sigma_n} > \sigma_n\right) \le \Pr\left(\frac{\mathrm{cut}_n^{\mathrm{miss}}}{n(n-1)\sigma_n} > \frac{1}{\sigma_n^{d+1}}\exp\left(-\frac{(r_n^{\min})^2}{2\sigma_n^2}\right)\right) \le \Pr(\mathcal{D}^c) \le \frac{1}{n^3}.$$

Clearly, we can replace $n(n-1)$ in the scaling factor by n^2 without changing the convergence rate. □

6.2.3. The volume term of the kNN *graph*

Proposition 6.7. *Let G_n be the k-nearest neighbor graph with a monotonically decreasing weight function f_n and let $H = H^+$ or $H = H^-$. Then*

$$\begin{aligned}&\left|\mathbb{E}\left(\frac{\mathrm{vol}_n(H)}{n(n-1)}\right) - \int_{H\cap C} F_B^{(1)}\left(r_n(x)\right) p^2(x)\,\mathrm{d}x\right| \\ &\quad= O\left(\sqrt[d]{\frac{k_n}{n}} F_B^{(1)}\left(r_n^{\max}\right)\right) + O\left(\min\left\{f_n^q\left(\inf_{x\in C} r_n(x)\right) n^{-\delta_0}, F_B^{(1)}(\infty) - F_B^{(1)}(\inf_{x\in C} r_n(x))\right\}\right) \\ &\qquad + O\left(\min\left\{f_n^q\left(\inf_{x\in C} r_n^-(x)\right)\left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}}\right)\frac{k_n}{n}, F_B^{(1)}(\infty) - F_B^{(1)}\left(\inf_{x\in C} r_n^-(x)\right)\right\}\right).\end{aligned}$$

where we set $\delta_n = \sqrt{(4\delta_0 \log n)/k_n}$ for a $\delta_0 \ge 2$ in the definition of $r_n^-(x)$.

For the variance term we have for a suitable constant $\tilde{C} > 0$

$$\Pr\left(\left|\mathrm{vol}_n(H) - \mathbb{E}\left(\mathrm{vol}_n(H)\right)\right| > \varepsilon\right) \le 2\exp\left(-\tilde{C}\frac{\varepsilon^2}{nk_n^2 f_n^2(0)}\right).$$

Proof. Similarly to the proof of for the cut we define for $i, j \in \{1,\dots,n\}$, $i \ne j$ the random variable W_{ij} as

$$W_{ij} = \begin{cases} f_n(\mathrm{dist}(x_i, x_j) & \text{if } x_i \in H \text{ and } (x_i, x_j) \text{ edge in } G_n \\ 0 & \text{otherwise}\end{cases}$$

and then have $\mathbb{E}\left(\mathrm{vol}_n(H)\right) = n(n-1)\mathbb{E}(W_{12})$. With a function $c(x,y)$ that indicates the probability of connectedness we obtain

$$\mathbb{E}(W_{12}^q) = \int_{H\cap C}\int_C f_n^q(\mathrm{dist}(x,y))c(x,y)p(y)\,\mathrm{d}y\; p(x)\,\mathrm{d}x.$$

Setting $\mathcal{R}_n = \{y \in H\cap C \mid \mathrm{dist}(y, \partial(H\cap C)) \le 2r_n^{\max}\}$ and $\mathcal{I}_n = (H\cap C)\setminus\mathcal{R}_n$ we can decompose the outer integral into integrals over $\mathcal{R}_n$ and $\mathcal{I}_n$.

First suppose $x \in \mathcal{R}_n$ and let c_n denote a bound on the probability that points in distance at least $r_n^{\max}$ are connected. Then, using $c_n \le 2\exp(-k_n/8)$ and Lemma 6.8,

$$\begin{aligned}\int_C f_n^q(\mathrm{dist}(x,y))c(x,y)p(y)\,\mathrm{d}y &\le p_{\max}\int_{B(x,r_n^{\max})\cap C} f_n^q(\mathrm{dist}(x,y))\,\mathrm{d}y + f_n^q\left(r_n^{\max}\right)c_n\int_C p(y)\,\mathrm{d}y \\ &\le p_{\max} d\eta_d\int_0^{r_n^{\max}} u^{d-1} f_n^q(u)\,\mathrm{d}u + 2f_n^q\left(r_n^{\max}\right)\exp\left(-k_n/8\right) \\ &= p_{\max}F_B^{(q)}\left(r_n^{\max}\right) + 2f_n^q\left(r_n^{\max}\right)\exp\left(-k_n/8\right).\end{aligned}$$

As was explained in the proof for the cut we can replace the term $2f_n^q\left(r_n^{\max}\right)\exp\left(-k_n/8\right)$ by the term

$$p_{\max}\left(F_B^{(q)}(\infty) - F_B^{(q)}(r_n^{\max})\right),$$

which is better suited, for example for the Gaussian.

Therefore, using that according to Lemma 6.25 the volume of $\mathcal{R}_n$ is in $O(r_n^{\max})$,

$$\int_{\mathcal{R}_n}\int_C f_n^q(\operatorname{dist}(x,y))c(x,y)p(y)\ \mathrm{d}y\ \mathrm{d}x = O\left(\sqrt[d]{\frac{k_n}{n}}F_B^{(q)}\left(r_n^{\max}\right)\right) + O\left(\min\left\{\sqrt[d]{\frac{k_n}{n}}\left(F_B^{(q)}(\infty) - F_B^{(q)}(r_n^{\max})\right),\right.\right.$$
$$\left.\left.\sqrt[d]{\frac{k_n}{n}} f_n^q\left(r_n^{\max}\right)\exp\left(-k_n/8\right)\right\}\right).$$

For $x \in \mathcal{I}_n$ we introduce as in the proof for the cut radii $r_n^-(x) \le r_n^{\max}$ and $r_n^+(x) \le r_n^{\max}$ that depend on δ_n and ξ_n defined there. These radii approximate the true kNN radius. For a lower bound we obtain

$$\int_C f_n^q(\operatorname{dist}(x,y))c(x,y)p(y)\ \mathrm{d}y \ge F_B^{(q)}\left(r_n(x)\right)p(x) - p_{\max}\left(F_B^{(q)}\left(r_n(x)\right) - F_B^{(q)}\left(r_n^-(x)\right)\right)$$
$$- \left(\xi_n + 6\exp\left(-\delta_n^2 k_n/3\right)\right)p_{\max}F_B^{(q)}\left(r_n^{\max}\right).$$

For some weight functions, especially the Gaussian, we can use

$$F_B^{(q)}\left(r_n(x)\right) - F_B^{(q)}\left(r_n^-(x)\right) \le F_B^{(q)}\left(\infty\right) - F_B^{(q)}\left(\inf_{x\in C} r_n^-(x)\right),$$

whereas for other ones it is better to use

$$F_B^{(q)}\left(r_n(x)\right) - F_B^{(q)}\left(r_n^-(x)\right) = d\eta_d\int_{r_n^-(x)}^{r_n(x)} u^{d-1}f_n^q(u)\ \mathrm{d}u$$
$$\le \eta_d f_n^q\left(\inf_{x\in C} r_n^-(x)\right)\left(\xi_n+\delta_n\right)\left(r_n^{\max}\right)^d.$$

Similarly we obtain an upper bound, with an additional term $f_n^q\left(\inf_{x\in C} r_n(x)\right)\exp\left(-\delta_n^2 k_n/4\right)$ or $p_{\max}(F_B^{(q)}(\infty) - F_B^{(q)}(\inf_{x\in C} r_n(x)))$ bounding the influence of points that are further away than $r_n^+(x)$. Combining the bounds we obtain

$$\left|\int_{\mathcal{I}_n}\int_C f_n^q(\operatorname{dist}(x,y))c(x,y)p(y)\ \mathrm{d}y - \int_{\mathcal{I}_n} F_B^{(q)}\left(r_n(x)\right)p^2(x)\ \mathrm{d}x\right|$$
$$= O\left((\xi_n + \exp\left(-\delta_n^2 k_n/3\right))F_B^{(q)}\left(r_n^{\max}\right)\right)$$
$$+ O\left(\min\left\{f_n^q\left(\inf_{x\in C} r_n^-(x)\right)\left(\xi_n+\delta_n\right)\left(r_n^{\max}\right)^d, F_B^{(q)}\left(\infty\right) - F_B^{(q)}\left(\inf_{x\in C} r_n^-(x)\right)\right\}\right)$$
$$+ O\left(\min\left\{f_n^q\left(\inf_{x\in C} r_n(x)\right)\exp\left(-\delta_n^2 k_n/4\right), F_B^{(q)}(\infty) - F_B^{(q)}\left(\inf_{x\in C} r_n(x)\right)\right\}\right).$$

Setting $\delta_n = \sqrt{(4\delta_0 \log n)/k_n}$ we obtain $\exp\left(-\delta_n^2 k_n/3\right) \le n^{-\delta_0}$ and the same for $\exp\left(-\delta_n^d k_n/4\right)$. Clearly, for $\delta_0 \ge 2$ we have $n^{-\delta_0} \le \xi_n$ and $n^{-\delta_0} \le (\xi_n r_n^{\max})^d$. Thus, with $\xi_n = O(r_n^{\max}) = O(\sqrt[d]{k_n/n})$,

$$\left|\int_{\mathcal{I}_n}\int_C f_n^q(\operatorname{dist}(x,y))c(x,y)p(y)\ \mathrm{d}y - \int_{\mathcal{I}_n} F_B^{(q)}\left(r_n(x)\right)p^2(x)\ \mathrm{d}x\right| = O\left(\sqrt[d]{\frac{k_n}{n}}F_B^{(q)}\left(r_n^{\max}\right)\right)$$
$$+ O\left(\min\left\{f_n^q\left(\inf_{x\in C} r_n^-(x)\right)\left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}}\right)\frac{k_n}{n}, F_B^{(q)}\left(\infty\right) - F_B^{(q)}\left(\inf_{x\in C} r_n^-(x)\right)\right\}\right)$$
$$+ O\left(\min\left\{f_n^q\left(\inf_{x\in C} r_n(x)\right)n^{-\delta_0}, F_B^{(q)}(\infty) - F_B^{(q)}\left(\inf_{x\in C} r_n(x)\right)\right\}\right).$$

Finally, by finding an upper bound on the integrand and the volume of $(H \cap C) \setminus \mathcal{I}_n$ we obtain

$$\left| \int_{\mathcal{I}_n} F_B^{(q)}\left(r_n(x)\right) p(x) \,\mathrm{d}x - \int_{H \cap C} F_B^{(q)}\left(r_n(x)\right) p^2(x) \,\mathrm{d}x \right| = O\left(\sqrt[d]{\frac{k_n}{n}} F_B^{(q)}\left(r_n^{\max}\right) \right).$$

Combining all the bounds above we obtain the result for the bias term. The bound for the variance term can be obtained with McDiarmid's inequality similarly to the proof for the cut in Proposition 6.1. □

The following lemma is necessary for the proof of the general theorem for both, the r-graph and the kNN-graph. It is an elementary lemma and therefore stated without proof.

Lemma 6.8 (integration over balls)**.** *Let $f_n : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a monotonically decreasing function and $x \in \mathbb{R}^d$. Then we have for any $R \in \mathbb{R}_{>0}$*

$$\int_{B(x,R)} f(\mathrm{dist}(x,y)) \,\mathrm{d}y = d\eta_d \int_0^R u^{d-1} f(u) \,\mathrm{d}u.$$

Corollary 6.9 (unweighted kNN-graph)**.** *Let G_n be the unweighted* kNN *graph with weight function $f_n \equiv 1$ and let $H = H^+$ or $H = H^-$. Then we have for the bias term*

$$\left| \frac{\mathrm{vol}_n(H)}{nk_n} - \int_H p(x) \,\mathrm{d}x \right| = O\left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}} \right),$$

and for the variance term for a suitable constant $\tilde{C}$

$$\Pr\left(\left| \frac{\mathrm{vol}_n(H)}{nk_n} - \mathbb{E}\left(\frac{\mathrm{vol}_n(H)}{nk_n} \right) \right| > \varepsilon \right) \leq 2\exp\left(-\tilde{C} n \varepsilon^2 \right).$$

Proof. With Lemma 6.22 we have, plugging in the definition of $r_n(x)$,

$$\int_{H \cap C} F_B^{(1)}(r_n(x)) p^2(x) \,\mathrm{d}x = \int_{H \cap C} \eta_d \frac{k_n}{(n-1)\eta_d p(x)} p^2(x) \,\mathrm{d}x = \frac{k_n}{n-1} \int_H p(x) \,\mathrm{d}x.$$

Therefore by multiplying the expression in Proposition 6.7 with $(n-1)/k_n$ we obtain for any $\delta_0 \geq 2$

$$\left| \frac{\mathrm{vol}_n(H)}{nk_n} - \int_H p(x) \,\mathrm{d}x \right| \leq O\left(\frac{n-1}{k_n} \sqrt[d]{\frac{k_n}{n}} F_B^{(1)}\left(r_n^{\max}\right) \right) + O\left(\frac{n-1}{k_n} f_n\left(\inf_{x \in C} r_n^-(x) \right) n^{-\delta_0} \right)$$
$$+ O\left(\frac{n-1}{k_n} \frac{k_n}{n} \left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}} \right) f_n\left(\inf_{x \in C} r_n^-(x) \right) \right).$$

Using $F_B^{(1)}(r_n^{\max}) \sim (n-1)/k_n$ and $f_n \equiv 1$ we obtain

$$\left| \frac{\mathrm{vol}_n(H)}{nk_n} - \int_H p(x) \,\mathrm{d}x \right| = O\left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}} \right).$$

For the variance term we use the bound in Proposition 6.7 and plug in $f_n(0) = 1$. □

Corollary 6.10 (Gaussian weights and $(k_n/n)^{1/d}/\sigma_n \to 0$)**.** *Consider the* kNN *graph with Gaussian weights and $(k_n/n)^{1/d}/\sigma_n \to 0$. Let $H = H^+$ or $H = H^-$. Then we have for the bias term*

$$\left| \frac{\sigma_n^d}{nk_n} \mathrm{vol}_n(H) - \frac{1}{(2\pi)^{d/2}} \int_H p(x) \,\mathrm{d}x \right| = O\left(\left(\frac{1}{\sigma_n} \sqrt[d]{\frac{k_n}{n}} \right)^2 + \sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}} \right)$$

and for the variance term, for a suitable constant $\tilde{C} > 0$,

$$\Pr\left(\left|\frac{\sigma_n^d}{nk_n}\mathrm{vol}_n(H) - \mathbb{E}\left(\frac{\sigma_n^d}{nk_n}\mathrm{vol}_n(H)\right)\right| > \varepsilon\right) \le 2\exp\left(-\tilde{C}n\varepsilon^2\right).$$

Proof. According to Lemma 6.23 we have for all $x \in C$

$$\left|\frac{\sigma_n^{qd}}{r_n^d(x)}F_B^{(q)}(r_n(x)) - \frac{\eta_d}{(2\pi)^{qd/2}}\right| \le 3\left(\frac{r_n(x)}{\sigma_n}\right)^2.$$

Plugging in $r_n(x) = \sqrt[d]{k_n/((n-1)\eta_d p(x))}$ and dividing by $\eta_d p(x)$ we obtain for points in the support of p

$$\left|\sigma_n^{qd}\left(\frac{n-1}{k_n}\right)F_B^{(q)}(r_n(x)) - \frac{1}{(2\pi)^{qd/2}p(x)}\right| = O\left(\left(\frac{k_n}{\sigma_n^d n}\right)^{2/d}\right).$$

Therefore, using the boundedness of p

$$\left|\sigma_n^d\left(\frac{n-1}{k_n}\right)\int_{H\cap C} p^2(x)F_B^{(1)}(r_n(x))\,\mathrm{d}x - \frac{1}{(2\pi)^{d/2}}\int_H p(x)\,\mathrm{d}x\right| = O\left(\left(\frac{k_n}{n\sigma_n^d}\right)^{2/d}\right).$$

Now, we consider the error terms from Proposition 6.7 of the other difference

$$\left|\frac{\sigma_n^d}{nk_n}\mathrm{vol}_n(H) - \sigma_n^d\left(\frac{n-1}{k_n}\right)\int_{H\cap C} p^2(x)F_B^{(1)}(r_n(x))\,\mathrm{d}x\right|.$$

As we have seen above $\sigma_n^d(n-1)/k_n F_B^{(1)}(r_n^{\max})$ can be bounded by a constant. Thus we have for the first term

$$\sigma_n^d\left(\frac{n-1}{k_n}\right)\sqrt[d]{\frac{k_n}{n}}F_B^{(1)}\left(r_n^{\max}\right) = O\left(\sqrt[d]{\frac{k_n}{n}}\right).$$

For the second term we have for n sufficiently large and setting $\delta_0 = 3$

$$\sigma_n^d\left(\frac{n-1}{k_n}\right)f_n\left(\inf_{x\in C} r_n^-(x)\right)n^{-\delta_0} \le \sigma_n^d\left(\frac{n-1}{k_n}\right)f_n(0)\,n^{-\delta_0} \le \left(\frac{n-1}{k_n}\right)n^{-\delta_0} \le n^{-2}.$$

For the third term we have

$$\begin{aligned}\sigma_n^d\left(\frac{n-1}{k_n}\right)\frac{k_n}{n}\left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}}\right)f_n\left(\inf_{x\in C} r_n^-(x)\right) &\le \left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}}\right)\sigma_n^d f_n(0)\\ &= \left(\sqrt[d]{\frac{k_n}{n}} + \sqrt{\frac{\log n}{k_n}}\right)\frac{1}{(2\pi)^{d/2}}.\end{aligned}$$

For *the variance term* we have for a suitable constant $\tilde{C}' > 0$

$$\begin{aligned}\Pr\left(\left|\frac{\sigma_n^d}{nk_n}\mathrm{vol}_n(H) - \mathbb{E}\left(\frac{\sigma_n^d}{nk_n}\mathrm{vol}_n(H)\right)\right| > \varepsilon\right) &= \Pr\left(\left|\mathrm{vol}_n(H) - \mathbb{E}\left(\mathrm{vol}_n(H)\right)\right| > nk_n\sigma_n^{-d}\varepsilon\right)\\ \le 2\exp\left(-\tilde{C}'\frac{n^2k_n^2\sigma_n^{-2d}\varepsilon^2}{nk_n^2 f_n^2(0)}\right) &\le 2\exp\left(-\tilde{C}'\frac{n\sigma_n^{-2d}\varepsilon^2}{\frac{1}{(2\pi)^d}\sigma_n^{-2d}}\right) = 2\exp\left(-\tilde{C}n\varepsilon^2\right),\end{aligned}$$

where we have set $\tilde{C} = (2\pi)^d\tilde{C}'$. $\square$

Corollary 6.11 (Gaussian weights and $(k_n/n)^{1/d}/\sigma_n \to \infty$)**.** *Let G_n be the* kNN *graph with Gaussian weights. Then for the bias term for a constant $\tilde{C}_1 > 0$*

$$\left|\mathbb{E}\left(\frac{\mathrm{vol}_n(H)}{n^2}\right) - \int_H p^2(x)\ \mathrm{d}x\right| = O\left(\sqrt[d]{\frac{k_n}{n}} + \exp\left(-\tilde{C}_1\left(\frac{1}{\sigma_n}\sqrt[d]{\frac{k_n}{n}}\right)^2\right)\right).$$

Let, furthermore, $\sqrt[d]{k_n/n} \geq \sigma_n^\alpha$ for an $\alpha \in (0,1)$ and n sufficiently large. Then there exist non-negative random variables $D_n^{(1)}, D_n^{(2)}$ such that

$$\left|\frac{\mathrm{vol}_n(H)}{n^2} - \mathbb{E}\left(\frac{\mathrm{vol}_n(H)}{n^2}\right)\right| = O(\sigma_n) + D_n^{(1)} + D_n^{(2)},$$

with $\Pr(D_n^{(1)} > \varepsilon) \leq 2\exp(\tilde{C}_2 n\sigma_n^{d+1}\varepsilon^2)$ for a constant $\tilde{C}_2 > 0$, and $\Pr(D_n^{(2)} > \sigma_n) \leq 1/n^3$.

Proof. With Lemma 6.24 we have for n sufficiently large such that $r_n(x)/\sigma_n$ sufficiently large uniformly over all $x \in C$

$$\begin{aligned}\left|\int_{H\cap C} F_B^{(1)}(r_n(x))p^2(x)\ \mathrm{d}x - \int_H p^2(x)\ \mathrm{d}x\right| &\leq \int_{H\cap C}\left|F_B^{(1)}(r_n(x)) - 1\right| p^2(x)\ \mathrm{d}x \\ &= O\left(\exp\left(-\frac{1}{4(p_{\max}\eta_d)^{2/d}}\frac{1}{\sigma_n^2}\left(\frac{k_n}{n}\right)^{2/d}\right)\right).\end{aligned}$$

Now we bound the error terms from Proposition 6.7 of the other difference

$$\left|\mathbb{E}\left(\frac{1}{n(n-1)}\mathrm{vol}_n(H)\right) - \int_{H\cap C} p^2(x)F_C^{(1)}(r_n(x))\ \mathrm{d}x\right|.$$

For the first error term we use that according to Lemma 6.24 $F_B^{(1)}(r_n^{\max})$ is bounded by one for n sufficiently large. Therefore $\sqrt[d]{k_n/n}F_B^{(1)}(r_n^{\max}) = O(\sqrt[d]{k_n/n})$.

For the second and third error term we observe that if n is sufficiently large such that $\delta_n \leq 1/2$ and $\xi_n \leq 1/4$ then

$$\inf_{x\in C} r_n(x) \geq \inf_{x\in C} r_n^-(x) = \inf_{x\in C}\sqrt[d]{\frac{(1-2\xi_n)(1-\delta_n)k_n}{(n-1)p(x)\eta_d}} \geq \sqrt[d]{\frac{k_n}{4p_{\max}\eta_d n}},$$

and therefore, for both, the second and the third error term,

$$F_B^{(1)}(\infty) - F_B^{(1)}(\inf_{x\in C} r_n(x)) = O\left(\exp\left(-\frac{1}{4(4p_{\max}\eta_d)^{2/d}}\left(\frac{1}{\sigma_n}\sqrt[d]{\frac{k_n}{n}}\right)^2\right)\right).$$

The proof of the bound for the variance term is identical to the corresponding part in the proof of Corollary 6.6. Therefore, we do not repeat it here.

Clearly, we can replace $n(n-1)$ in the scaling factor by n^2 without changing the convergence rate. □

6.2.4. The main theorem for the kNN *graph*

Proof of Theorem 3.1. As discussed in Section 6.1 we can study the convergence of the bias and variance terms of the cut and the volume separately.

For the *unweighted graph* we have with Corollary 6.4 that under the condition $k_n/\log n \to \infty$ the bias term for the cut is in $O(\sqrt[d]{k_n/n} + \sqrt{\log n/k_n})$. For some $\varepsilon > 0$ the probability that the variance term exceeds ε is bounded by $2\exp(-\tilde{C}\varepsilon^2 n^{1-2/d}k_n^{2/d})$ for a suitable constant $\tilde{C}$. Clearly, the bias term converges to zero under the condition $k_n/\log n \to \infty$. For the almost sure convergence of the variance term we need the stricter condition in dimension $d = 1$. The convergence of the volume-term follows with Corollary 6.9, since the requirements for this convergence are weaker. In the case $d \geq 2$ we obtain the optimal rates by equating the two bounds of the bias term and checking that the variance term converges as well at this rate. In the case $d = 1$ the optimal rate is determined by the variance term.

For the kNN-*graph with Gaussian weights and* $r_n/\sigma_n \to \infty$ we need the stronger condition $r_n \geq \sigma_n^\alpha$ for an $\alpha \in (0, 1)$ in order to show convergence of both, the bias term and the variance term. Under this condition we have according to Corollaries 6.6 and 6.11 that the bias term of both, the cut and the volume, is in $O(r_n)$, since the exponential term converges as σ_n.

Furthermore, the almost sure convergence of the variance term can be shown with the Borel−Cantelli lemma if $n\sigma_n^{d+1}/\log n \to \infty$ for $n \to \infty$.

For the kNN-*graph with Gaussian weights and* $r_n/\sigma_n \to 0$ according to Corollary 6.5 the bias term of the cut is in $O(r_n + (r_n/\sigma_n)^2 + \sqrt{\log n/k_n})$. The probability that the variance term of the cut exceeds an $\varepsilon > 0$ is bounded by $2\exp(-\tilde{C}n^{1-2/d}k_n^{2/d})$ for a suitable constant $\tilde{C}$, which is the same expression as in the unweighted case. Therefore, we have almost sure convergence of the cut-term to zero under the same conditions as for the unweighted kNN graph.

From Corollary 6.10 we can see that the convergence conditions for the volume are less strict than that of the cut. □

6.3. The r-graph and the complete weighted graph

This section consists of three parts: in the first one the convergence of the bias and variance term of the cut is studied, whereas in the second part that convergence is studied for the volume. Combining these results we can proof the main theorems on the convergence of NCut and CheegerCut for the r-graph and the complete weighted graph.

Section 6.3.1 and Section 6.3.2 are built up similarly: first, a proposition for a general weight function is given. The results are stated in terms of the "cap" and "ball" integrals and some properties of the weight function. Then four corollaries follow, where the general result is applied to the complete weighted graph with Gaussian weight function and to the r-graph with the specific weight functions we consider in this paper.

Some words on the proofs: the results on the bias terms for general weight functions can be shown analogously to the corresponding results for the kNN graph. Since the connectivity in these graphs given the position of two points is not random they are even simpler. Furthermore, all the error terms in the result for the kNN graph that are due to the uncertainty in the connectivity radius can be dropped for the r-graph and the complete weighted graph. Therefore, in the proof of the bias term of the cut we only discuss the adaptations that are made to the proof of the kNN graph.

As explained in Section 6.1 the situation is different for the variance term, where the convergence proof for the kNN-graph would lead to suboptimal results when carried over to the other two graphs. For this reason we give a different proof for the convergence of the variance term in the proof of the general result for the cut. It can be easily carried over to the volume and thus we omit it there.

As to the corollaries we only proof two of them: that for the complete weighted graph and that for the r-graph with Gaussian weights and $r_n/\sigma_n \to 0$ for $n \to \infty$. The proof of the corollary for the unweighted graph is very simple, that of the corollary for the r-graph with Gaussian weights and $\sigma_n/r_n \to 0$ is identical to the proof for the complete weighted graph where we can ignore one term.

The proofs in Section 6.3.2 are completely omitted: the general result on the bias term can be proved analogously to that for the kNN graph, if the adaptations that are discussed in the proof for the bias term of the cut are made. The general result on the variance term of the volume is proved analogously to that on the variance term of the cut. The proofs of the corollaries also work analogously to the corresponding proofs for the cut.

The proofs of the main theorems in Section 6.3.3 collect the bounds of the corollaries and identify the conditions that have to hold for the convergence of NCut and CheegerCut.

6.3.1. The cut *term in the r-graph and the complete weighted graph*

Proposition 6.12 (the cut in the r-neighborhood and the complete weighted graph)**.** *Let $(r_n)_{n\in\mathbb{N}}$ be a sequence that fulfills the conditions on parameter sequences of the r-neighborhood graph. Let G_n denote the r-neighborhood graph with parameter r_n or the complete weighted graph on $x_1,\dots,x_n$ with a monotonically decreasing weight function $f_n : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. We set*

$$1_c = \begin{cases} 1 & \text{if } G_n \text{ is the complete weighted graph} \\ 0 & \text{if } G_n \text{ is the } r_n\text{-neighborhood graph.} \end{cases}$$

Then for the bias term

$$\left| \mathbb{E}\left(\frac{\mathrm{cut}_n}{n(n-1)F_C^{(1)}(r_n)} \right) - 2\int_S p^2(s)\ \mathrm{d}s \right| = O\left(r_n + \frac{F_B^{(1)}(\infty) - F_B^{(1)}(r_n)}{F_C^{(1)}(r_n)} 1_c \right).$$

Furthermore, there are constants $\tilde{C}_1, \tilde{C}_2$ such that for the variance term

$$\Pr\left(\left| \frac{\mathrm{cut}_n}{n(n-1)F_C^{(1)}(r_n)} - \mathbb{E}\left(\frac{\mathrm{cut}_n}{n(n-1)F_C^{(1)}(r_n)} \right) \right| \geq \varepsilon \right)$$
$$\leq 2\exp\left(- \frac{n\left(F_C^{(1)}(r_n)\right)^2 \varepsilon^2}{\tilde{C}_1 F_C^{(2)}(r_n) + \tilde{C}_2 (F_B^{(2)}(\infty) - F_B^{(2)}(r_n)) 1_c + 2\varepsilon F_C^{(1)}(r_n)\ f_n(0)} \right).$$

Proof. As was said in the introduction we do not give the detailed proof of this proposition here, since it is similar to the proof of the corresponding proposition for the kNN-graph but simpler: the radius r_n is the same everywhere, that is we can set $r_n^{\max} = r_n^+(s) + = r_n^-(s) = r_n$ for all $s \in S$. Furthermore, the connectivity is not random, that is we can set $a_n = b_n = c_n = 0$ for the r-neighborhood graph, whereas we set $a_n = 0$,$b_n = 1$ and $c_n = 1$ for the complete weighted graph. We obtain

$$\left| \mathbb{E}\left(W_{12}^q\right) - 2F_C^{(q)}(r_n)\ \int_S p^2(s)\ \mathrm{d}s \right| = O\left(F_C^{(q)}(r_n) r_n + \left(F_B^{(q)}(\infty) - F_B^{(q)}(r_n) \right) 1_c \right),$$

and thus the result for the bias term immediately.

In order to bound the variance term we use a U-statistics argument. We have

$$\frac{\mathrm{cut}_n}{n(n-1)F_C^{(1)}(r_n)} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n \frac{1}{F_C^{(1)}(r_n)} W_{ij}.$$

For the upper bound on the properly rescaled variable W_{ij} clearly

$$\frac{1}{F_C^{(1)}(r_n)} W_{ij} \leq \frac{1}{F_C^{(1)}(r_n)} f_n(0)$$

and for the variance

$$\mathrm{Var}\left(\frac{1}{F_C^{(1)}(r_n)} W_{ij} \right) = \mathbb{E}\left(\left(\frac{1}{F_C^{(1)}(r_n)} W_{ij} \right)^2 \right) - \left(\mathbb{E}\left(\frac{1}{F_C^{(1)}(r_n)} W_{ij} \right) \right)^2 \leq \left(\frac{1}{F_C^{(1)}(r_n)} \right)^2 \mathbb{E}\left(W_{ij}^2\right).$$

With a Bernstein-type concentration inequality for U-statistics from Hoeffding [6] we obtain

$$\Pr\left(\left|\frac{\mathrm{cut}_n}{n(n-1)F_C^{(1)}(r_n)} - \mathbb{E}\left(\frac{\mathrm{cut}_n}{n(n-1)F_C^{(1)}(r_n)}\right)\right| \geq \varepsilon\right) \leq 2\exp\left(-\frac{\lfloor n/2\rfloor\varepsilon^2}{2\left(\frac{1}{F_C^{(1)}(r_n)}\right)^2\mathbb{E}\left(W_{ij}^2\right)+\frac{2}{3}\frac{1}{F_C^{(1)}(r_n)}\varepsilon f_n(0)}\right)$$
$$\leq 2\exp\left(-\frac{n\varepsilon^2\left(F_C^{(1)}(r_n)\right)^2}{6\mathbb{E}\left(W_{ij}^2\right)+2\varepsilon F_C^{(1)}(r_n)\, f_n(0)}\right)$$

where we have used $\lfloor n/2\rfloor \geq n/3$ for $n \geq 2$

Clearly, for $r_n \to 0$ we can find constants (depending on p and S) $\tilde{C}_1$ and $\tilde{C}_2$ such that for n sufficiently large $6\mathbb{E}(W_{ij}^2) \leq \tilde{C}_1 F_C^{(2)}(r_n) + \tilde{C}_2(F_B^{(2)}(\infty) - F_B^{(2)}(r_n))1_c$. □

The following corollary can be proved by plugging in the results of Lemma 6.22 into the bounds of Proposition 6.12. We do not give the details here.

Corollary 6.13 (unweighted r-graph)**.** *For the r-neighborhood graph and the weight function $f_n = 1$ we obtain*

$$\left|\mathbb{E}\left(\frac{\mathrm{cut}_n}{n^2 r_n^{d+1}}\right) - \frac{2\eta_{d-1}}{d+1}\int_S p^2(s)\ \mathrm{d}s\right| = O(r_n),$$

and, for a suitable constant $\tilde{C} > 0$,

$$\Pr\left(\left|\frac{\mathrm{cut}_n}{n^2 r_n^{d+1}} - \mathbb{E}\left(\frac{\mathrm{cut}_n}{n^2 r_n^{d+1}}\right)\right| \geq \varepsilon\right) \leq 2\exp\left(-\tilde{C} n r_n^{d+1}\varepsilon^2\right).$$

Corollary 6.14 (complete weighted graph)**.** *Consider the complete weighted graph G_n with Gaussian weight function. Then we have for the bias term for any $\alpha \in (0,1)$*

$$\left|\mathbb{E}\left(\frac{\mathrm{cut}_n}{n^2\sigma_n}\right) - \frac{2}{\sqrt{2\pi}}\int_S p^2(s)\ \mathrm{d}s\right| = O\left(\sigma_n^\alpha\right).$$

For the variance term we can find a constant $\tilde{C} > 0$ such that for n sufficiently large

$$\Pr\left(\left|\frac{\mathrm{cut}_n}{n^2\sigma_n} - \mathbb{E}\left(\frac{\mathrm{cut}_n}{n^2\sigma_n}\right)\right| \geq \varepsilon\right) \leq 2\exp\left(-\tilde{C} n\sigma_n^{d+1}\varepsilon^2\right).$$

Proof. Let r_n be a sequence with $r_n \to 0$ and $r_n/\sigma_n \to \infty$ for $n \to \infty$. We use the bound from Proposition 6.12 and the fact that $F_C^{(1)}(r_n)/\sigma_n$ can be bounded by a constant due to Lemma 6.24 to obtain

$$\left|\mathbb{E}\left(\frac{\mathrm{cut}_n}{n(n-1)\sigma_n}\right) - 2\frac{F_C^{(1)}(r_n)}{\sigma_n}\int_S p^2(s)\ \mathrm{d}s\right| = O\left(r_n + \frac{F_B^{(1)}(\infty) - F_B^{(1)}(r_n)}{\sigma_n}\right)$$
$$= O\left(r_n + \frac{1}{\sigma_n}\exp\left(-\frac{r_n^2}{4\sigma_n^2}\right)\right).$$

On the other hand, using Lemma 6.24, the boundedness of p and $\mathcal{L}_{d-1}(S\cap C)$, we have for r_n/σ_n sufficiently large

$$\left|2\frac{F_C^{(1)}(r_n)}{\sigma_n}\int_S p^2(s)\ \mathrm{d}s - \frac{2}{\sqrt{2\pi}}\int_S p^2(s)\ \mathrm{d}s\right| \leq \left|\frac{F_C^{(1)}(r_n)}{\sigma_n} - \frac{1}{\sqrt{2\pi}}\right| 2\int_S p^2(s)\ \mathrm{d}s = O\left(\exp\left(-\frac{r_n^2}{4\sigma_n^2}\right)\right).$$

Combining these two bounds und using $\log \sigma_n \leq 0$ for n sufficiently large we obtain

$$\left| \mathbb{E}\left(\frac{\mathrm{cut}_n}{n(n-1)\sigma_n} \right) - \frac{2}{\sqrt{2\pi}} \int_S p^2(s)\ \mathrm{d}s \right| = O\left(r_n + \exp\left(-\frac{r_n^2}{4\sigma_n^2} \right) \right).$$

Setting $r_n = \sigma_n^\alpha$ we have to show that the exponential term converges as fast. We have

$$\sigma_n^{-\alpha} \exp\left(-\frac{r_n^2}{4\sigma_n^2} \right) = \sigma_n^{-\alpha} \exp\left(-\frac{1}{4}\sigma_n^{2\alpha-2} \right) = \left(\sigma_n^{2\alpha-2}\right)^{\frac{-\alpha}{2\alpha-2}} \exp\left(-\frac{1}{4}\sigma_n^{2\alpha-2} \right) \to 0$$

for $n \to \infty$, since $x^r \exp(-x) \to 0$ for $x \to \infty$ and all $r \in \mathbb{R}$.

For the variance term we have with Proposition 6.12 and for constants $\tilde{C}_1, \tilde{C}_2$

$$\begin{aligned}
&\Pr\left(\left| \frac{\mathrm{cut}_n}{n(n-1)\sigma_n} - \mathbb{E}\left(\frac{\mathrm{cut}_n}{n(n-1)\sigma_n} \right) \right| \geq \varepsilon \right) \\
&\qquad = \Pr\left(\left| \frac{\mathrm{cut}_n}{n(n-1)F_C^{(1)}(r_n)} - \mathbb{E}\left(\frac{\mathrm{cut}_n}{n(n-1)F_C^{(1)}(r_n)} \right) \right| \geq \frac{\sigma_n}{F_C^{(1)}(r_n)}\varepsilon \right) \\
&\qquad \leq 2\exp\left(-\frac{n\sigma_n^2\varepsilon^2}{\tilde{C}_1 F_C^{(2)}(r_n) + \tilde{C}_2(F_B^{(2)}(\infty) - F_B^{(2)}(r_n)) + 2\varepsilon F_C^{(1)}(r_n)\ f_n(0)} \right).
\end{aligned}$$

With Lemma 6.24 we have for r_n/σ_n sufficiently large $F_C^{(2)}(r_n) = O(\sigma_n^{1-d})$, and

$$F_B^{(2)}(\infty) - F_B^{(2)}(r_n) = O\left(\sigma_n^{-d} \exp\left(-\frac{r_n^2}{4\sigma_n^2} \right) \right) = O\left(\sigma_n^{1-d}\right),$$

if we choose $r_n = \sigma_n^\alpha$ for $\alpha \in (0,1)$ similarly to above.

For the last term in the denominator we have $F_C^{(1)}(r_n)\ f_n(0) = O\left(\sigma_n \sigma_n^{-d}\right) = O\left(\sigma_n^{1-d}\right)$. Therefore, we can find a constant $\tilde{C}_3 > 0$ such that

$$\Pr\left(\left| \frac{\mathrm{cut}_n}{n(n-1)\sigma_n} - \mathbb{E}\left(\frac{\mathrm{cut}_n}{n(n-1)\sigma_n} \right) \right| \geq \varepsilon \right) \leq 2\exp\left(-\tilde{C}_3 \frac{n\sigma_n^2\varepsilon^2}{\sigma_n^{1-d}} \right) = 2\exp\left(-\tilde{C}_3 n \sigma_n^{d+1}\varepsilon^2 \right).$$

Since we assume that $n\sigma_n \to \infty$ for $n \to \infty$ we can replace $n(n-1)$ in the scaling factor by n^2. □

We do not state the proof of the following corollary, since it is similar to the proof of the last one. The difference is, that we do not have to consider the 1_c-terms, which are zero in the case of the r-graph.

Corollary 6.15 (r-graph with Gaussian weights and $\sigma_n/r_n \to 0$)**.** *Let G_n be the r-graph with Gaussian weight function and let $\sigma_n/r_n \to 0$ for $n \to \infty$. Then we have for the bias term*

$$\left| \mathbb{E}\left(\frac{\mathrm{cut}_n}{n^2\sigma_n} \right) - \frac{2}{\sqrt{2\pi}} \int_S p^2(s)\ \mathrm{d}s \right| = O\left(r_n + \exp\left(-\frac{r_n^2}{4\sigma_n^2} \right) \right).$$

For the variance term we can find a constant $\tilde{C}_2 > 0$ such that

$$\Pr\left(\left| \frac{\mathrm{cut}_n}{n^2\sigma_n} - \mathbb{E}\left(\frac{\mathrm{cut}_n}{n^2\sigma_n} \right) \right| \geq \varepsilon \right) \leq 2\exp\left(-\tilde{C}_2\ n\ \sigma_n^{d+1}\varepsilon^2 \right)$$

Corollary 6.16 (r-graph with Gaussian weights and $r_n/\sigma_n \to 0$)**.** *Consider the r-neighborhood graph with Gaussian weight function and let $r_n/\sigma_n \to 0$ for $n \to \infty$. Then we can find a constant $\tilde{C} > 0$ such that*

$$\left| \mathbb{E}\left(\frac{\sigma_n^d}{r_n^{d+1}} \frac{\mathrm{cut}_n}{n^2} \right) - \frac{2\eta_{d-1}}{(d+1)(2\pi)^{d/2}} \int_S p^2(s)\ \mathrm{d}s \right| = O\left(r_n + \frac{r_n^2}{\sigma_n^2} \right)$$

and

$$\Pr\left(\left|\frac{\sigma_n^d}{r_n^{d+1}}\frac{\mathrm{cut}_n}{n^2} - \mathbb{E}\left(\frac{\sigma_n^d}{r_n^{d+1}}\frac{\mathrm{cut}_n}{n^2}\right)\right| \geq \varepsilon\right) \leq 2\exp\left(-\tilde{C}n\varepsilon^2 r_n^{d+1}\right).$$

Proof. Multiplying the bound in Proposition 6.12 with $\sigma_n^d F_C^{(1)}(r_n)/r_n^{d+1}$, which can be bounded by a constant according to Lemma 6.23, and using $1_c = 0$ we obtain

$$\left|\mathbb{E}\left(\frac{\sigma_n^d F_C^{(1)}(r_n)}{r_n^{d+1}}\frac{\mathrm{cut}_n}{n(n-1)}\right) - 2\frac{\sigma_n^d F_C^{(1)}(r_n)}{r_n^{d+1}}\int_S p^2(s)\ \mathrm{d}s\right| = O\left(r_n\right).$$

On the other hand, by the boundedness of p and $\mathcal{L}_{d-1}(S\cap C)$, and with Lemma 6.23

$$\left|2\frac{\sigma_n^d F_C^{(1)}(r_n)}{r_n^{d+1}}\int_S p^2(s)\ \mathrm{d}s - \frac{2\eta_{d-1}}{(d+1)(2\pi)^{d/2}}\int_S p^2(s)\ \mathrm{d}s\right| = O\left(\frac{r_n^2}{\sigma_n^2}\right).$$

Combining these two bounds we obtain the result for the bias term.

For the variance term we have with Proposition 6.12 and for a constant $\tilde{C}_1$

$$\begin{aligned}
&\Pr\left(\left|\frac{\sigma_n^d}{r_n^{d+1}}\frac{\mathrm{cut}_n}{n(n-1)} - \mathbb{E}\left(\frac{\sigma_n^d}{r_n^{d+1}}\frac{\mathrm{cut}_n}{n(n-1)}\right)\right| \geq \varepsilon\right)\\
&\qquad = \Pr\left(\left|\frac{\mathrm{cut}_n}{n(n-1)F_C^{(1)}(r_n)} - \mathbb{E}\left(\frac{\mathrm{cut}_n}{n(n-1)F_C^{(1)}(r_n)}\right)\right| \geq \frac{r_n^{d+1}}{\sigma_n^d F_C^{(1)}(r_n)}\varepsilon\right)\\
&\qquad \leq 2\exp\left(-\frac{n\left(r_n^{d+1}/\sigma_n^d\right)^2\varepsilon^2}{\tilde{C}_1 F_C^{(2)}(r_n)\ + 2\varepsilon F_C^{(1)}(r_n)\ f_n(0)}\right).
\end{aligned}$$

With Lemma 6.23 we obtain $F_C^{(2)}(r_n) = O(r_n^{d+1}/\sigma_n^{2d})$ for sufficiently large n. With the same proposition and plugging in $f_n(0)$ we obtain $F_C^{(1)}(r_n)f_n(0) = O(r_n^{d+1}/\sigma_n^{2d})$. Plugging in these results above we obtain the bound for the variance term.

Since we always assume that $nr_n \to \infty$ for $n \to \infty$ we can replace $n(n-1)$ in the scaling factor by n^2. □

6.3.2. The volume term in the r-graph and the complete weighted graph

The following results are stated without proof: Proposition 6.17 can be proved analogously to Proposition 6.7 if the remarks on the difference between the kNN-graph and r-neighborhood graph in the proof of Proposition 6.12 are considered. The corollaries can be shown similarly to the corresponding corollaries in the previous section.

Proposition 6.17. *Let G_n be the r_n-neighborhood graph or the complete weighted graph with a weight function f_n and set 1_c as in Proposition 6.12. Then*

$$\left|\mathbb{E}\left(\frac{\mathrm{vol}_n(H)}{n(n-1)F_B^{(1)}(r_n)}\right) - \int_H p^2(x)\ \mathrm{d}x\right| \leq O\left(r_n + \frac{F_B^{(1)}(\infty) - F_B^{(1)}(r_n)}{F_B^{(1)}(r_n)}1_c\right).$$

For the variance term we have

$$\begin{aligned}
&\Pr\left(\left|\frac{\mathrm{vol}_n(H)}{n(n-1)F_B^{(1)}(r_n)} - \mathbb{E}\left(\frac{\mathrm{vol}_n(H)}{n(n-1)F_B^{(1)}(r_n)}\right)\right| \geq \varepsilon\right)\\
&\qquad \leq 2\exp\left(-\frac{n\varepsilon^2\left(F_B^{(1)}(r_n)\right)^2}{\tilde{C}_1 F_B^{(2)}(r_n) + \tilde{C}_2 1_c(F_B^{(2)}(\infty) - F_B^{(2)}(r_n)) + 2\varepsilon f_n(0) F_B^{(1)}(r_n)}\right).
\end{aligned}$$

Corollary 6.18 (unweighted graph)**.** *For $f_n \equiv 1$ and the r_n-neighborhood graph we have*

$$\left| \mathbb{E}\left(\frac{\mathrm{vol}_n(H)}{n^2 r_n^d} \right) - \eta_d \int_{H \cap C} p^2(x) \ \mathrm{d}x \right| \leq O(r_n)$$

and, for a constant $\tilde{C} > 0$,

$$\Pr\left(\left| \frac{\mathrm{vol}_n(H)}{n^2 r_n^d} - \mathbb{E}\left(\frac{\mathrm{vol}_n(H)}{n^2 r_n^d} \right) \right| \geq \varepsilon \right) \leq 2 \exp\left(-\tilde{C} n \varepsilon^2 r_n^d \right).$$

Corollary 6.19 (complete weighted graph with Gaussian weights)**.** *Consider the complete weighted graph with the Gaussian weight function and a parameter sequence $\sigma_n \to 0$. Then we have for any $\alpha \in (0,1)$*

$$\left| \mathbb{E}\left(\frac{\mathrm{vol}_n(H)}{n^2} \right) - \int_{H} p^2(x) \ \mathrm{d}x \right| = O\left(\sigma_n^{\alpha} \right).$$

Furthermore there is a constant $\tilde{C}' > 0$ such that

$$\Pr\left(\left| \frac{\mathrm{vol}_n(H)}{n^2} - \mathbb{E}\left(\frac{\mathrm{vol}_n(H)}{n^2} \right) \right| \geq \varepsilon \right) \leq \exp\left(-\tilde{C}' n \varepsilon^2 \sigma_n^d \right).$$

Corollary 6.20 (r-graph with Gaussian weights and $\sigma_n / r_n \to 0$)**.** *Let G_n be the r-neighborhood graph with Gaussian weights and let $\sigma_n / r_n \to 0$ for $n \to \infty$. Then we have for the bias term for sufficiently large n*

$$\left| \mathbb{E}\left(\frac{\mathrm{vol}_n(H)}{n^2} \right) - \int_{H} p^2(x) \ \mathrm{d}x \right| = O\left(r_n + \exp\left(-\frac{1}{4} \frac{r_n^2}{\sigma_n^2} \right) \right),$$

and for the variance term for a suitable constant $\tilde{C}' > 0$

$$\Pr\left(\left| \frac{\mathrm{vol}_n(H)}{n^2} - \mathbb{E}\left(\frac{\mathrm{vol}_n(H)}{n^2} \right) \right| \geq \varepsilon \right) \leq \exp\left(-\tilde{C}' n \varepsilon^2 \sigma_n^d \right).$$

Corollary 6.21 (r-graph with Gaussian weights and $r_n / \sigma_n \to 0$)**.** *Let G_n be the r-neighborhood graph with Gaussian weights and let $r_n / \sigma_n \to 0$ for $n \to \infty$. Then we have for the bias term for sufficiently large n*

$$\left| \mathbb{E}\left(\frac{\sigma_n^d}{n^2 r_n^d} \mathrm{vol}_n(H) \right) - \frac{\eta_d}{(2\pi)^{d/2}} \int_{H} p^2(x) \ \mathrm{d}x \right| = O\left(r_n + \left(\frac{r_n}{\sigma_n} \right)^2 \right),$$

and for the variance term for a suitable constant $\tilde{C} > 0$

$$\Pr\left(\left| \frac{\sigma_n^d}{n^2 r_n^d} \mathrm{vol}_n(H) - \mathbb{E}\left(\frac{\sigma_n^d}{n^2 r_n^d} \mathrm{vol}_n(H) \right) \right| > \varepsilon \right) \leq 2 \exp\left(-\tilde{C} n \varepsilon^2 r_n^d \right).$$

6.3.3. The main theorems for the r-graph and the complete weighted graph

Proof of Theorem 3.2. As discussed in Section 6.1 we can study the convergence of the bias and variance terms of the cut and the volume separately.

For the *unweighted r-graph* we have with Corollary 6.13 that the bias term of the cut is in $O(r_n)$ and that for $\varepsilon > 0$ we can find a constant $\tilde{C}$ such that the probability that the variance term of the cut exceeds ε is bounded by $2\exp(-\tilde{C} n \sigma_n^{d+1} \varepsilon^2)$. Thus the cut-term converges almost surely to zero for $r_n \to 0$ and $n r_n^{d+1} / \log n \to \infty$. It follows from Corollary 6.18 that under these conditions the vol-term also converges to zero. The best convergence rate for the cut-term is $\sqrt[d+3]{\log n / n}$, which is achieved setting $r_n \sim \sqrt[d+3]{\log n / n}$. Setting r_n in this way the convergence rate of the vol-term is also $\sqrt[d+3]{\log n / n}$.

For the r-graph with *Gaussian weights and* $r_n/\sigma_n \to \infty$ we have with Corollaries 6.15 and 6.20 that the bias term of both, the cut and the volume, is in $O(r_n + \exp(-1/4(r_n/\sigma_n)^2))$. Furthermore, we can find a constant $\tilde{C} > 0$ such that the probability that the variance term of the cut exceeds an $\varepsilon > 0$ is bounded by $2\exp(-\tilde{C}n\sigma_n^{d+1}\varepsilon^2)$. Similarly, the variance term of the volume would converge almost surely for $n\sigma_n^d/\log n \to \infty$. This implies almost sure convergence of Δ_n to zero under the condition $n\sigma_n^{d+1}/\log n \to \infty$ for $n \to \infty$.

For the r-graph with *Gaussian weights and* $r_n/\sigma_n \to 0$ we have with Corollary 6.16 a rate of $O(r_n + (r_n/\sigma_n)^2)$ for the bias term of the cut. Furthermore, the probability that the variance term exceeds an $\varepsilon > 0$ is bounded by $2\exp(-\tilde{C}n\varepsilon^2 r_n^{d+1})$ with a constant $\tilde{C}$. Therefore, the cut-term almost surely converges to zero under the conditions $r_n \to 0$ and $nr_n^{d+1}/\log n \to \infty$. Under these conditions with Corollary 6.21 the volume-term also converges to zero. □

Proof of Theorem 3.3. As discussed in Section 6.1 we can study the convergence of the bias and variance terms of the cut and the volume separately.

With Corollaries 6.14 and 6.19 we have that the bias term of both, the cut and the volume is in $O(\sigma_n^\alpha)$ for any $\alpha \in (0,1)$. Furthermore, the probability that the variance term of the cut exceeds an $\varepsilon > 0$ is bounded by $2\exp(-\tilde{C}\ n\ \sigma_n^{d+1}\varepsilon^2)$ with a suitable constant $\tilde{C}$. For the variance term of the volume the exponent in this bound is only d. Consequently, we have almost sure convergence to zero under the condition $n\sigma_n^{d+1}/\log n \to \infty$.

For any fixed $\alpha \in (0,1)$ the optimal convergence rate is achieved setting $\sigma_n = ((\log n)/n)^{1/(d+1+2\alpha)}$. Since the variance term has to converge for any $\alpha \in (0,1)$ we choose $\sigma_n = ((\log n)/n)^{1/(d+3)}$ and achieve a convergence rate of σ_n^α for any $\alpha \in (0,1)$. □

6.4. The integrals $F_C^{(q)}(r)$ and the size of the boundary strips

Lemma 6.22 (unit weights)**.** *Let $f_n \equiv 1$ be the unit weight function. Then for any $r > 0$*

$$F_C^{(1)}(r) = F_C^{(2)}(r) = \frac{\eta_{d-1}}{d+1} r^{d+1}$$

and

$$F_B^{(1)}(r) = F_B^{(2)}(r) = \eta_d r^d.$$

Lemma 6.23 (Gaussian weights and $r_n/\sigma_n \to 0$)**.** *Let f_n denote the Gaussian weight function with parameter σ_n and let $r_n > 0$. Then we have for $q = 1,2$ for the cap integral*

$$\left| \frac{\sigma_n^{qd}}{r_n^{d+1}} F_C^{(q)}(r_n) - \frac{\eta_{d-1}}{(d+1)(2\pi)^{qd/2}} \right| \le 2 \left(\frac{r_n}{\sigma_n} \right)^2.$$

For the ball integral $F_B^{(q)}(r_n)$ we have

$$\left| \frac{\sigma_n^{qd}}{r_n^{d}} F_B^{(q)}(r_n) - \frac{\eta_{d}}{(2\pi)^{qd/2}} \right| \le 3 \left(\frac{r_n}{\sigma_n} \right)^2.$$

Proof. For the "ball integral" we have (with the substitution $v = u/r_n$)

$$\begin{aligned} F_B^{(q)} &= d\eta_d \int_0^{r_n} u^{d-1} f_n^q(u)\ \mathrm{d}u = d\eta_d \int_0^{r_n} u^{d-1} \frac{1}{(2\pi)^{qd/2}\sigma_n^{qd}} \exp\left(-\frac{q}{2}\frac{u^2}{\sigma_n^2}\right)\ \mathrm{d}u \\ &= \frac{d\eta_d}{(2\pi)^{qd/2}\sigma_n^{qd}} \int_0^1 (vr_n)^{d-1} \exp\left(-\frac{q}{2}\frac{v^2 r_n^2}{\sigma_n^2}\right) r_n\ \mathrm{d}v \\ &= \frac{d\eta_d r_n^d}{(2\pi)^{qd/2}\sigma_n^{qd}} \int_0^1 v^{d-1} \exp\left(-\frac{qv^2}{2}\frac{r_n^2}{\sigma_n^2}\right)\ \mathrm{d}v. \end{aligned}$$

Clearly,

$$\int_0^1 v^{d-1} \exp\left(-\frac{qv^2}{2}\frac{r_n^2}{\sigma_n^2}\right) \, \mathrm{d}v \le \int_0^1 v^{d-1} \, \mathrm{d}v = \frac{1}{d}$$

and, on the other hand

$$\int_0^1 v^{d-1} \exp\left(-\frac{qv^2}{2}\frac{r_n^2}{\sigma_n^2}\right) \, \mathrm{d}v \ge \exp\left(-\frac{q}{2}\frac{r_n^2}{\sigma_n^2}\right) \int_0^1 v^{d-1} \, \mathrm{d}v \ge \left(1 - \frac{q}{2}\frac{r_n^2}{\sigma_n^2}\right)\frac{1}{d} \ge \left(1 - \frac{r_n^2}{\sigma_n^2}\right)\frac{1}{d}.$$

Therefore,

$$\left(1 - \frac{r_n^2}{\sigma_n^2}\right)\frac{\eta_d}{(2\pi)^{qd/2}} \le \frac{\sigma_n^{qd}}{r_n^d} F_B^{(q)}(r_n) \le \frac{\eta_d}{(2\pi)^{qd/2}}.$$

Using $\eta_d/\sqrt{2\pi} \le 3$ we obtain the result for the ball integral. The result for the cap integral is shown similarly. □

Lemma 6.24 (Gaussian weights and $\sigma_n/r_n \to 0$)**.** *Let f_n denote the Gaussian weight function with a parameter σ_n and let $r_n/\sigma_n \ge 4d$. Then we have $F_C^{(1)}(\infty) = \sigma_n/\sqrt{2\pi}$ and*

$$\left|\frac{1}{\sigma_n}F_C^{(1)}(r_n) - \frac{1}{\sqrt{2\pi}}\right| = O\left(\exp\left(-\frac{1}{4}\left(\frac{r_n}{\sigma_n}\right)^2\right)\right).$$

Furthermore, $F_C^{(2)}(\infty) = O(\sigma_n^{1-d})$ and $F_C^{(2)}(\infty) - F_C^{(2)}(r_n) = O(\sigma_n^{1-d}\exp\left(-(r_n/\sigma_n)^2/4\right))$.

For the ball integral we have under the same conditions $F_B^{(1)}(\infty) = 1$

$$\left|F_B^{(1)}(r_n) - 1\right| = O\left(\exp\left(-\frac{1}{4}\left(\frac{r_n}{\sigma_n}\right)^2\right)\right).$$

Furthermore, $F_B^{(2)}(\infty) = O(\sigma_n^{-d})$ and $F_B^{(2)}(\infty) - F_B^{(2)}(r_n) = O(\sigma_n^{-d}\exp\left(-(r_n/\sigma_n)^2/4\right))$.

Proof. We have with an integral table, for example in Harris and Stocker [5], for $q = 1, 2$

$$\begin{aligned}
\int_0^\infty x^d f_n^q(x) \, \mathrm{d}x &= \int_0^\infty x^d \frac{1}{(2\pi)^{qd/2}\sigma_n^{qd}} \exp\left(-\frac{q}{2\sigma_n^2}x^2\right) \, \mathrm{d}x \\
&= \frac{1}{(2\pi)^{qd/2}\sigma_n^{qd}} \int_0^\infty x^d \exp\left(-\frac{q}{2\sigma_n^2}x^2\right) \, \mathrm{d}x = \frac{1}{(2\pi)^{qd/2}\sigma_n^{qd}} \frac{\Gamma(\frac{d+1}{2})}{2\left(\frac{q}{2\sigma_n^2}\right)^{(d+1)/2}} \\
&= \frac{\Gamma(\frac{d+1}{2})2^{(d-1)/2}}{(2\pi)^{qd/2}q^{(d+1)/2}}\sigma_n^{(1-q)d+1}.
\end{aligned}$$

This implies for all $r_n > 0$

$$F_C^{(2)}(r_n) \le F_C^{(2)}(\infty) = \eta_{d-1}\int_0^\infty x^d f_n^q(x) \, \mathrm{d}x = O(\sigma_n^{1-d}).$$

For $q = 1$ we have

$$F_C^{(1)}(\infty) = \eta_{d-1}\int_0^\infty x^d f_n(x) \, \mathrm{d}x = \frac{\pi^{(d-1)/2}}{\Gamma(\frac{d+1}{2})}\frac{\Gamma(\frac{d+1}{2})2^{(d-1)/2}}{(2\pi)^{d/2}}\sigma_n = \frac{\sigma_n}{\sqrt{2\pi}}.$$

We now bound the error we make, when the integral does not run to ∞ but to r_n. We have

$$\begin{aligned}\int_{r_n}^{\infty} x^d f_n^q(x)\ \mathrm{d}x &= \frac{1}{(2\pi)^{qd/2}} \int_{r_n}^{\infty} x^d \frac{1}{\sigma_n^{qd}} \exp\left(-\frac{q}{2}\frac{x^2}{\sigma^2}\right) \\ &= \frac{1}{(2\pi)^{qd/2}} \int_{r_n/\sigma_n}^{\infty} (u\sigma_n)^d \frac{1}{\sigma_n^{qd}} \exp\left(-\frac{q}{2}u^2\right)\sigma_n\ \mathrm{d}u = \frac{\sigma_n^{(1-q)d+1}}{(2\pi)^{qd/2}} \int_{r_n/\sigma_n}^{\infty} u^d \exp\left(-\frac{q}{2}u^2\right)\ \mathrm{d}u,\end{aligned}$$

where we applied the substitution $u = x/\sigma_n$.

We have for $r_n/\sigma_n \geq 4d$

$$\begin{aligned}\int_{r_n/\sigma_n}^{\infty} u^d \exp\left(-\frac{q}{2}u^2\right)\ \mathrm{d}u &\leq \int_{r_n/\sigma_n}^{\infty} \exp\left(d\log u - \frac{1}{2}u^2\right)\ \mathrm{d}u \leq \int_{r_n/\sigma_n}^{\infty} \exp\left(du - \frac{1}{2}u^2\right)\ \mathrm{d}u \\ &\leq \int_{r_n/\sigma_n}^{\infty} \exp\left(-\frac{1}{4}u^2\right)\ \mathrm{d}u \leq \int_{r_n/\sigma_n}^{\infty} \frac{u}{2}\exp\left(-\frac{1}{4}u^2\right)\ \mathrm{d}u = \exp\left(-\frac{1}{4}\left(\frac{r_n}{\sigma_n}\right)^2\right).\end{aligned}$$

For the ball integral we have with the substitution $v = u/\sigma_n$

$$\begin{aligned}F_B^{(q)}(\infty) &= d\eta_d \int_0^{\infty} u^{d-1} \frac{1}{(2\pi)^{qd/2}\sigma_n^{qd}} \exp\left(-\frac{q}{2}\frac{u^2}{\sigma^2}\right)\ \mathrm{d}u \\ &= d\eta_d \int_0^{\infty} (\sigma_n v)^{d-1} \frac{1}{(2\pi)^{qd/2}\sigma_n^{qd}} \exp\left(-\frac{q}{2}v^2\right)\sigma_n\ \mathrm{d}v = \frac{d\eta_d}{(2\pi)^{qd/2}}\sigma_n^{(1-q)d} \int_0^{\infty} v^{d-1} \exp\left(-\frac{q}{2}v^2\right)\ \mathrm{d}v \\ &= \frac{d\eta_d}{(2\pi)^{qd/2}}\sigma_n^{(1-q)d} \frac{\Gamma(d/2)}{2(q/2)^{d/2}} = \sigma_n^{(1-q)d} \frac{d}{2^{qd/2}\pi^{qd/2}} \frac{\pi^{d/2}}{\Gamma(d/2+1)} \frac{\Gamma(d/2)}{2(q/2)^{d/2}} \\ &= \sigma_n^{(1-q)d}\pi^{(1-q)d/2}2^{(1-q)d/2},\end{aligned}$$

since $\Gamma(d/2+1) = d/2\Gamma(d/2)$.

We have, again with the substitution $v = u/\sigma_n$ and for $r_n/\sigma_n \geq 1$

$$\begin{aligned}d\eta_d \int_{r_n}^{\infty} u^{d-1} \frac{1}{(2\pi)^{qd/2}\sigma_n^{qd}} \exp\left(-\frac{q}{2}\frac{u^2}{\sigma^2}\right)\ \mathrm{d}u &= d\eta_d \int_{r_n/\sigma_n}^{\infty} (\sigma_v)^{d-1} \frac{1}{(2\pi)^{qd/2}\sigma_n^{qd}} \exp\left(-\frac{q}{2}v^2\right)\sigma_n\ \mathrm{d}v \\ &= \frac{d\eta_d}{(2\pi)^{qd/2}}\sigma_n^{(1-q)d} \int_{r_n/\sigma_n}^{\infty} v^{d-1} \exp\left(-\frac{q}{2}v^2\right)\ \mathrm{d}v \\ &\leq \frac{d\eta_d}{(2\pi)^{qd/2}}\sigma_n^{(1-q)d} \int_{r_n/\sigma_n}^{\infty} v^{d} \exp\left(-\frac{q}{2}v^2\right)\ \mathrm{d}v.\end{aligned}$$

We have for $r_n/\sigma_n \geq 4d$

$$\begin{aligned}\int_{r_n/\sigma_n}^{\infty} u^d \exp\left(-\frac{q}{2}u^2\right)\ \mathrm{d}u &\leq \int_{r_n/\sigma_n}^{\infty} \exp\left(d\log u - \frac{1}{2}u^2\right)\ \mathrm{d}u \leq \int_{r_n/\sigma_n}^{\infty} \exp\left(du - \frac{1}{2}u^2\right)\ \mathrm{d}u \\ &\leq \int_{r_n/\sigma_n}^{\infty} \exp\left(-\frac{1}{4}u^2\right)\ \mathrm{d}u \leq \int_{r_n/\sigma_n}^{\infty} \frac{u}{2}\exp\left(-\frac{1}{4}u^2\right)\ \mathrm{d}u = \exp\left(-\frac{1}{4}\left(\frac{r_n}{\sigma_n}\right)^2\right). \quad \square\end{aligned}$$

The following lemma is necessary to bound the influence of points close to the boundary on the cut and the volume. The first statement is used for the cut, whereas the second statement is used for the volume.

Lemma 6.25. *Let the general assumptions hold and let $(r_n)_{n\in\mathbb{N}}$ be a sequence with $r_n \to 0$ for $n \to \infty$. Define $\mathcal{R}_n = \{x \in \mathbb{R}^d \mid \operatorname{dist}(x, \partial C) \leq 2r_n\}$. Then $\mathcal{L}_{d-1}(S \cap \mathcal{R}_n) = O(r_n)$.*

For $H = H^+$ or $H = H^-$ define $\bar{\mathcal{R}}_n = \{x \in H \cap C \mid \operatorname{dist}(x, \partial(H \cap C)) \leq 2r_n\}$. Then $\mathcal{L}_d(\bar{\mathcal{R}}_n) = O(r_n)$.

Appendix. Table of notation

The following table contains an overview of the most important notation used througout the paper:

$x_1, \ldots, x_n$	sample points in $\mathbb{R}^d$
n	sample size
d	dimension of the space $\mathbb{R}^d$
$p(x)$	density, the points are sampled from
C	compact support of the density p
$p_{\min}$, $p_{\max}$	minimum and maximum value of the density p on C
$p'_{\max}$	supremum of the norm of the gradient $\|\nabla p(x)\|$ in the interior of C
μ	measure induced by the density p, that is, $\mu(A) = \int_A p(x) \, \mathrm{d}x$
∂C	boundary of the set C
κ	minimal curvature radius of ∂C
n_x	normal to the surface ∂C at the point $x \in \partial C$
γ, r_γ	for balls of radius $r \leq r_\gamma$ around points in C at least γ of the total volume of the ball is within C
S	hyperplane in $\mathbb{R}^d$ that defines the cuts we consider in the neighborhood graphs
H^+, H^-	halfspaces of $\mathbb{R}^d$ defined by S
n_S	normal of S pointing towards H^+
α	minimum angle between n_S and n_x for all $x \in S \cap \partial C$
$\langle x_1, x_2 \rangle$	Euclidean dot product of $x_1, x_2 \in \mathbb{R}^d$
$\|x\|$	Euclidean norm of $x \in \mathbb{R}^d$, *i.e.* $\|x\| = \sqrt{\langle x, x \rangle}$
$\mathrm{dist}(x, y)$	distance between x and y
$\mathcal{L}$	the Lebesgue volume
$\mathcal{L}_{d-1}$	the $(d-1)$-dimensional Lebesgue measure in a $(d-1)$-dimensional affine subspace or the $(d-1)$-dimensional area of a $(d-1)$-dimensional surface
$\mathcal{L}_{d-2}$	the $(d-2)$-dimensional area of a $(d-2)$-dimensional surface
$B(x, r)$	the closed ball of radius r around $x \in \mathbb{R}^d$, that is, $B(x, r) = \{y \in \mathbb{R}^d \mid \mathrm{dist}(x, y) \leq r\}$
η_d	volume of the d-dimensional unit ball in the Euclidean metric, that is, $\eta_d = \mathcal{L}_d(B(0, 1))$
τ_d	kissing number in dimension d
$\Pr(A)$	probability of the event A
$\mathbb{E}(U)$	expectation of the random variable U
$\mathrm{Var}(U)$	variance of the random variable U
$Bin(n, p)$	discrete density of the binomial distribution with parameters n and p
$\xrightarrow{a.s.}$	almost sure convergence
$f = O(g)$	f is bounded above by g asymptotically up to a constant factor
$\nabla f(x)$	gradient of f at x
$\frac{\partial f(x)}{\partial x_i}$	partial derivative of the function f in the direction x_i
k	neighborhood parameter of the k-nearest neighbor graph
r	neighborhood size of the r-neigborhood graph
σ	bandwidth of the Gaussian weight function
1_c	1 for complete graph, 0 otherwise
$\mathrm{cut}(C, V \setminus C)$	cut size of the cut defined by $(C, V \setminus C)$ in the graph $G(V, E)$ with vertice set V and edge set E
$\mathrm{vol}(C)$	volume of $C \subseteq V$ in the graph $G(V, E)$

$\mathrm{NCut}(C, V \setminus C)$	the normalized cut measure for the partition $(C, V \setminus C)$ in the graph $G(V, E)$
cut_n	cut in neighborhood graph on n points defined by S
$\mathrm{vol}_n(H)$	volume of sample points in the halfspace H in neighborhood graph on n points
NCut_n	normalized cut in neighborhood graph on n points
$r_n^{\max}$	maximum (with a high probability) k-nearest neighbor radius
$r_n(x)$	expected k-nearest neighbor radius in point x
r_n^+, $r_n^-(x)$	sequences that converge to $r_n(x)$ from above and below
ξ_n	variation of the density in balls of radius $2r_n^{\max}$
δ_n	sequence determining the convergence of r_n^+ and $r_n^-(x)$ to $r_n(x)$
C_{ij}	event that there is an edge between x_i and x_j in the undirected neighborhood graph
D_{ij}	event that there is an edge between x_i and x_j in the directed neighborhood graph
$c(x, y)$	probabilty of an edge between a point in x and a point in y
$(s_n^{\mathrm{cut}})_{n\in\mathbb{N}}$, $(s_n^{\mathrm{vol}})_{n\in\mathbb{N}}$	scaling sequences for the cut and the volume
W_{ij}	random variable for the weight of an edge between x_i and x_j
$F_B^{(q)}(r)$	integral over balls $d\eta_d \int_0^r u^{d-1} f_n^q(u)\,\mathrm{d}u$
$F_C^{(q)}(r)$	integral over caps $\eta_{d-1} \int_0^r u^d f_n^q(u)\,\mathrm{d}u$
$CutLim$	limit of the cut induced by S on the neighborhood graph
$VolLim(H)$	limit of the volume of the halfspace H.

References

[1] D. Angluin and L. Valiant, Fast probabilistic algorithms for Hamiltonian circuits. *J. Comput. Syst. Sci.* **18** (1979) 155–193.

[2] G. Biau, B. Cadre and B. Pelletier, A graph-based estimator of the number of clusters. *ESAIM: PS* **11** (2007) 272–280.

[3] M. Brito, E. Chavez, A. Quiroz and J. Yukich, Connectivity of the mutual k-nearest-neighbor graph in clustering and outlier detection. *Stat. Probab. Lett.* **35** (1997) 33–42.

[4] S. Bubeck and U. von Luxburg, Nearest neighbor clustering: a baseline method for consistent clustering with arbitrary objective functions. *J. Mach. Learn. Res.* **10** (2009) 657–698.

[5] J.W. Harris and H. Stocker, *Handbook of Mathematics and Computational Science*. Springer (1998).

[6] W. Hoeffding, Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **58** (1963) 13–30.

[7] D.O. Loftsgaarden and C.P. Quesenberry, A nonparametric estimate of a multivariate density function. *Ann. Math. Stat.* **36** (1965) 1049–1051.

[8] M. Maier, M. Hein and U. von Luxburg, Optimal construction of k-nearest neighbor graphs for identifying noisy clusters. *Theoret. Comput. Sci.* **410** (2009) 1749–1764.

[9] M. Maier, U. von Luxburg and M. Hein, Influence of graph construction on graph-based clustering measures, in *Advances in Neural Information Processing Systems,* vol. 21, edited by D. Koller, D. Schuurmans, Y. Bengio and L. Bottou. MIT Press (2009) 1025–1032.

[10] G. Miller, S. Teng, W. Thurston and S. Vavasis, Separators for sphere-packings and nearest neighbor graphs. *J. ACM* **44** (1997) 1–29.

[11] H. Narayanan, M. Belkin and P. Niyogi, On the relation between low density separation, spectral clustering and graph cuts, in *Advances in Neural Information Processing Systems,* vol. 19, edited by B. Schölkopf, J. Platt and T. Hoffman. MIT Press (2007) 1025–1032.

[12] A. Srivastav and P. Stangier, Algorithmic Chernoff-Hoeffding inequalities in integer programming. *Random Struct. Algorithms* **8** (1996) 27–58.

[13] U. von Luxburg, A tutorial on spectral clustering. *Stat. Comput.* **17** (2007) 395–416.

[14] U. von Luxburg, M. Belkin and O. Bousquet, Consistency of spectral clustering. *Ann. Stat.* **36** (2008) 555–586.