# The f-Adjusted Graph Laplacian: a Diagonal Modification with a Geometric Interpretation

**Sven Kurras** KURRAS@INFORMATIK.UNI-HAMBURG.DE
**Ulrike von Luxburg** LUXBURG@INFORMATIK.UNI-HAMBURG.DE
Department of Computer Science, University of Hamburg, Germany

**Gilles Blanchard** GILLES.BLANCHARD@MATH.UNI-POTSDAM.DE
Department of Mathematics, University of Potsdam, Germany

## Abstract

Consider a neighborhood graph, for example a $k$-nearest neighbor graph, that is constructed on sample points drawn according to some density $p$. Our goal is to re-weight the graph's edges such that all cuts and volumes behave as if the graph was built on a different sample drawn from an alternative density $\overline{p}$. We introduce the $\mathbf{f}$-adjusted graph and prove that it provides the correct cuts and volumes as the sample size tends to infinity. From an algebraic perspective, we show that its normalized Laplacian, denoted as the $\mathbf{f}$-adjusted Laplacian, represents a natural family of diagonal perturbations of the original normalized Laplacian. Our technique allows to apply any cut and volume based algorithm to the $\mathbf{f}$-adjusted graph, for example spectral clustering, in order to study the given graph as if it were built on an unaccessible sample from a different density. We point out applications in sample bias correction, data uniformization, and multi-scale analysis of graphs.

## 1. Introduction

Assume that we are given a neighborhood graph $G$, say a $k$-nearest neighbor graph, based on a sample $x_1, ..., x_n$ of points drawn according to some probability density $p$. Many properties of $p$ reflect in properties of $G$. However, we are actually interested in another density $\overline{p}$ that differs from $p$ in some known aspects, but we cannot access a sample from $\overline{p}$. For example, $p$ may be biased in some way known to us. How could we correct for such a bias if we just have access to the given graph $G$, but not to the underlying sampling mechanism? In order to tackle this problem, our general strategy is to transform $G$ into another graph $\overline{G}$ that "reflects properties of $\overline{p}$", although the vertices still refer to the original sample points distributed according to $p$. We focus on volume and cut properties, as these are used in a large variety of machine learning algorithms. Our goal is that volumes and cuts of $\overline{G}$ reveal the same information on $\overline{p}$ as another graph would do that is directly built on a sample from $\overline{p}$. We provide a solution that provably achieves this goal in the limit of increasing sample size. We first define a vector $\mathbf{f}$ of vertex weights that reflects the continuous volume properties of $\overline{p}$. Then we "merge" these vertex weights into $G$ by re-weighting its edges in such a way that the modified graph $\overline{G}$ attains $\mathbf{f}$ as its degree vector. Any such "merging operation" implicitly sets all volumes in $\overline{G}$ corresponding to $\overline{p}$. The crux is to construct a merging operation that simultaneously ensures that the cut weights in $\overline{G}$ correspond to the continuous cut properties of $\overline{p}$. We provide such an operation by defining $\overline{G}$ as the "$\mathbf{f}$-adjusted graph of $G$".

The above delineates the *geometric interpretation* of $\mathbf{f}$-adjusting. It allows to think of transforming $G$ into $\overline{G}$ as of modifying the underlying density from $p$ to $\overline{p}$. This provides all the intuition that we need in order to choose $\mathbf{f}$ in a meaningful way and to interpret volumes and cut weights in the $\mathbf{f}$-adjusted graph. Although this motivation is purely geometric, it turns out that $\mathbf{f}$-adjusting also has appealing algebraic properties. We denote by the term "$\mathbf{f}$-adjusted Laplacian" the normalized Laplacian matrix of the $\mathbf{f}$-adjusted graph. The *algebraic interpretation* shows that all $\mathbf{f}$-adjusted Laplacians that can be obtained from $G$ represent a natural family of diagonal perturbations of the normalized Laplacian of $G$. Thus we can think of transforming $G$ into $\overline{G}$ as of a meaningful modification along the main diagonal of the original normalized Laplacian. The algebraic results do not require that $G$ is a geometric graph.

*Proceedings of the $31^{st}$ International Conference on Machine Learning*, Beijing, China, 2014. JMLR: W&CP volume 32. 

These interpretations allow us to apply $\mathbf{f}$-adjusting to a variety of problems from machine learning:

(a) Correcting for sampling bias: we obtain a method to correct for a sampling bias in graph-based algorithms.

(b) Spectral clustering: although $G$ is built on a fixed sample drawn from $p$, we can apply spectral clustering as if the graph were built on another sample drawn from density $\overline{p}$.

(c) Multi-scale analysis: the original density $p$ can be modified in various ways, for example by a convolution. Let $k_h$ denote some convolution kernel of bandwidth $h$. Estimates on $p * k_h$ can be extracted from the graph by averaging over neighborhoods. $\mathbf{f}$-adjusting can then be applied in order to merge these estimates into edge weights appropriately, so that the new volumes and cut weights represent $p * k_h$.

(d) Manifold learning and clustering: there is a complicated interplay between the geometric structure of the data and the sampling density. It can be helpful to separate these two aspects, for example if we are interested in a function $f$ of which we only know $f(x_i)$ with $x_i$ sampled according to some unknown sampling density on a manifold. We would like to remove the artifacts induced by the sampling density and construct a situation that looks like as if the sample points $x_i$ were distributed uniformly.

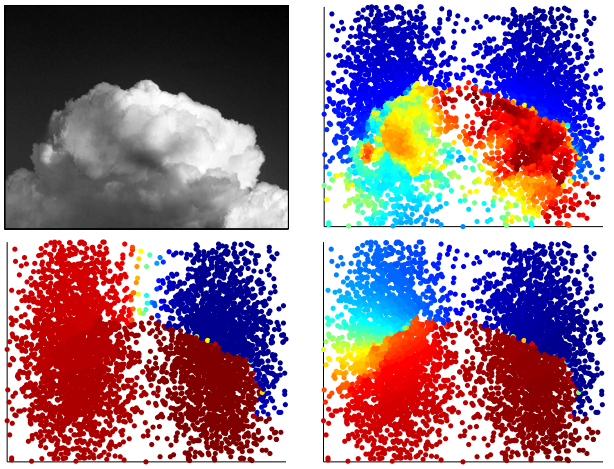

*Figure 1.* Top left: original image of a cloud, unknown to us. Top right: 5000 sample points drawn from two Gaussians, colors indicating intensity values. We can access data only at these positions. Bottom left: spectral image clustering applied to the non-uniform sample positions. We show the heat plot of the NCut score vector (eigenvector that solves the relaxed minimum normalized cut problem). The scores are biased to a wrong segmentation. Bottom right: heat plot of the NCut score vector of our density-corrected graph. It separates the cloud from its background because the spatial bias is removed.

## 2. Applications

Before we dive into the theoretical analysis, we motivate the reader by two example applications.

### 2.1. Removing density information

In some scenarios the sample distribution does *not* provide the structural information that we are interested in. For example in sea temperature recording or general wireless sensor networks one has to deal with measurement values $t_i$ at random spatial positions $x_i$. The goal is to study the measurement values without any bias introduced by their random positions. If the measurements were taken at uniform positions, then one could apply spectral clustering. But any non-uniform spatial distribution distorts all cut weights. There is no obvious mechanism to compensate for this influence. We demonstrate this effect in an image segmentation task. The classical spectral clustering for image data on a $d$-dimensional grid (Shi & Malik, 2000) constructs an $r$-graph of radius $r$ on the pixel positions. Edge weights are defined by the products $w_{ij} := \mathrm{sim}_s(x_i, x_j) \cdot \mathrm{sim}_v(t_i, t_j)$ of spatial similarities $\mathrm{sim}_s(x_i, x_j) := \exp(-0.5\|x_i - x_j\|^2/\sigma_s^2)$ and value similarities $\mathrm{sim}_v(t_i, t_j) := \exp(-0.5\|t_i - t_j\|^2/\sigma_v^2)$. The normalized cuts in the resulting graph provide useful clustering information on the intensity values. But this only holds because the spatial similarities of a grid have a uniform impact on the product in $w_{ij}$. It does not generalize to the case where the sample positions are distributed non-uniformly. In this case the spatial similarities distort the cut weights and hence the clustering result in an undesired way. Consider the image in Figure 1 (top left), and assume that its intensity values are not given at the positions of a grid, but only at sample positions drawn from an aggregation of two Gaussians (top right). In order to remove the spatial unbalancedness, we suggest to proceed as follows:

(1) Build the neighborhood graph $G$ on the sample points from only their spatial similarities $w_{ij} := \mathrm{sim}_s(x_i, x_j)$.

(2) With $\mathbf{d}$ the vector of vertex degrees, create the $\mathbf{d}^{-1}$-adjusted graph $\overline{G}$. It provides the edge weights $\overline{w}_{ij}$. We show below that these weights remove the density information from the original volumes and cut weights of $G$.

(3) Define the final edge weights by $\overline{w}_{ij} \cdot \mathrm{sim}_v(t_i, t_j)$.

(4) Apply spectral clustering to the resulting graph.

This approach removes the influence of the spatial distribution from the final edge weights. It solely considers intensity values. Figure 1 (bottom right) shows that this leads to the correct segmentation, whereas the biased segmentation provided by the original graph is not correct (bottom left).

### 2.2. Correcting for a sampling bias

Consider a data set in which some areas are known to be over- or underrepresented. For example, a poll among shop

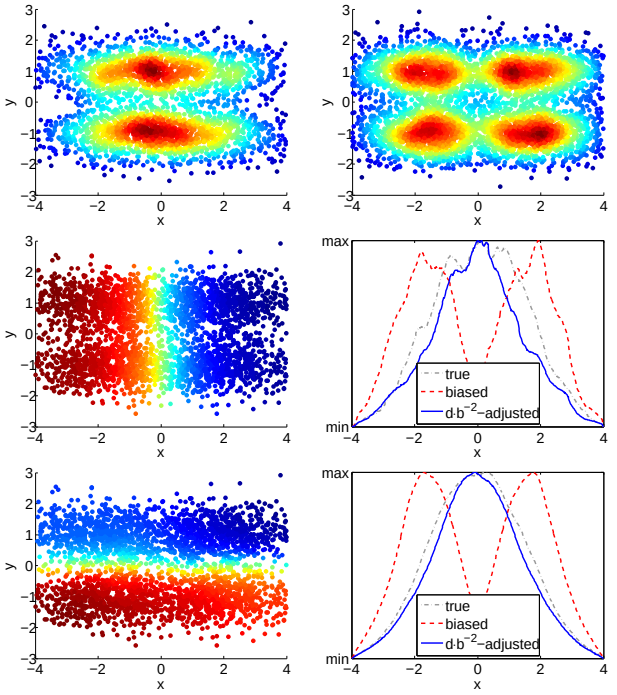


*Figure 2.* Top left: exemplary sample from true density $\overline{p}$ that we cannot access. Top right: 3000 sample points from biased density $p = b \cdot \overline{p}$ that underrepresents around $x \approx 0$. This is the sample we can access. Middle right: cut weights of vertical cuts at $x \in [-4, 4]$, for a single graph drawn from each the true and the biased density, and for the $\mathbf{db}^{-2}$-adjusted graph. Bottom right: same, averaged over 20 runs. Middle left: heat plot of the NCut score vector of the biased graph, giving the wrong vertical cut. Bottom left: heat plot of the NCut score vector of the $\mathbf{db}^{-2}$-adjusted graph. It identifies the correct clusters, since the bias is removed.

customers in the morning will particularly have the ages 20 to 30 being underrepresented. How can we correct for this in a graph-based learning scenario? Let $p = b \cdot \overline{p}$ denote the erroneous density that we can access under some bias $b : \mathbb{R}^d \to \mathbb{R}_{>0}$ and $\overline{p}$ the true density that we cannot access. Using our framework, one can use any estimate of $b$ in order to compensate for the bias. This can be achieved by the $\mathbf{f}$-adjusted graph for $\mathbf{f} = \mathbf{db}^{-2} := (d_i / b(x_i)^2)_i$. It provides volumes and cut weights just as if the sample points were drawn from $\overline{p}$ instead of $p$. Since $\mathbf{f}$-adjusting does not require any coordinates of the sample points, the bias $\mathbf{b}$ can also be given as external knowledge on the vertices.

Consider the example in Figure 2. It shows two Gaussians (top left) that are underrepresented around $x \approx 0$ due to bias $b(x, y) = \min\{1, (2 + 3|x|)/8\}$ (top right). The minimum normalized cut of a graph built on the biased sample is misdirected by this bias to the wrong vertical clusters (middle left). However, the $\mathbf{f}$-adjustment appropriately "repairs" the volumes and cut weights in the graph. Now the correct horizontal cut is revealed (bottom left).

## 3. The f-Adjusted Graph Laplacian

In this section we formally define the $\mathbf{f}$-adjusted graph and its Laplacian. $\mathbf{f}$-adjusting addresses two goals. The first goal is to re-weight the edges of a given graph in such a way that the new degree vector equals the vector $\mathbf{f}$. This implicitly sets the *volumes* of all vertex subsets according to $\mathbf{f}$. The second goal is to provide a geometric interpretation that simultaneously relates the new *cut weights* to $\mathbf{f}$. As shown in Section 5, $\mathbf{f}$-adjusting indeed fits both volumes and cut weights to $\mathbf{f}$ in a way that allows for a geometric interpretation: both quantities represent one and the same modified underlying density.

### 3.1. Graph notation

For any matrix $A = [a_{ij}]_{ij} \in \mathbb{R}^{n \times n}$ let $A \geq 0$ denote that all entries are non-negative. For vector $\mathbf{b} \in \mathbb{R}^n$ we refer by $\operatorname{diag}(\mathbf{b}) \in \operatorname{diag}(\mathbb{R}^n)$ to the corresponding diagonal matrix. For any $q \in \mathbb{R}$ we define $\mathbf{b}^q := (b_i^q)_i$ element-wise, as well as the product $\mathbf{bc} := (b_i c_i)_i$ of vectors $\mathbf{b}, \mathbf{c} \in \mathbb{R}^n$. We define the set of **graph matrices** as $\mathbb{W} := \{X \in \mathbb{R}^{n \times n}_{\geq 0} \mid X = X^T, X\mathbf{1} > 0\}$. We also study its generalization to the set of **weak graph matrices** $\mathbb{W}_{\ominus} := \{X \in \mathbb{W} + \operatorname{diag}(\mathbb{R}^n) \mid X\mathbf{1} > 0\}$, whose elements further allow for negative diagonal entries as long as all row sums remain positive. Obviously $\mathbb{W} \subset \mathbb{W}_{\ominus}$. To each $W \in \mathbb{W}_{\ominus}$ there corresponds an undirected weighted graph $\mathcal{G}(W) := (V, E, W)$ that has $W$ as its weighted adjacency matrix, and edge set $E := \{ij \mid w_{ij} \neq 0\}$. The degree of vertex $i \in V$ is $d_i := \sum_{j \in V} w_{ij} > 0$, leading to the all-positive degree vector $\mathbf{d} = W\mathbf{1}$. The volume of $S \subseteq V$ is defined as $\operatorname{vol}_G(S) := \sum_{i \in S} d_i$. Any partition $V = S \cup \overline{S}$ is a cut of weight $\operatorname{cut}_G(S, \overline{S}) := \sum_{i \in S, j \in \overline{S}} w_{ij}$. For the two frequently used vectors $\mathbf{f} = (f_i)_i$ and $\mathbf{d} = (d_i)_i$ we set $F := \operatorname{diag}(\mathbf{f})$ and $D := \operatorname{diag}(\mathbf{d})$.

### 3.2. Definition of the f-adjusted graph

From now on fix some $W \in \mathbb{W}$ and set $G := \mathcal{G}(W)$. There are two straightforward methods for modifying $G$'s edge weights (= entries in $W$) in order to fit all vertex degrees (= row/column sums of $W$) to some prescribed vector $\mathbf{f}$. $\mathbf{f}$-adjusting combines these two methods because they mutually compensate their respective drawbacks.

**$\mathbf{f}$-selflooping.** The naive strategy to adjust the vertex degrees of the graph $\mathcal{G}(W)$ to any prescribed vector $\mathbf{f} \in \mathbb{R}^n_{>0}$ is by modifying selfloops: fix all off-diagonal entries in $W$, and set its main diagonal such that the row sums equal $\mathbf{f}$:

$$W^{\circ}_{\mathbf{f}} := W - D + F \quad \in \mathbb{W}_{\ominus}.$$

This yields the vertex degrees $W^{\circ}_{\mathbf{f}}\mathbf{1} = \mathbf{f}$ exactly. We refer to $\mathcal{G}(W^{\circ}_{\mathbf{f}})$ as the $\mathbf{f}$-selflooped graph. Despite of its positive row sums, this approach can force some entries on the

main diagonal of $W^\circ_{\mathbf{f}}$ to take negative values. We can fix this by considering $c\mathbf{f}$-selflooping for $c > 0$. This yields vertex degrees $W^\circ_{c\mathbf{f}}\mathbf{1} = c\mathbf{f}$, and it is straightforward to see that the main diagonal of $\mathcal{W}^\circ_{c\mathbf{f}}$ is non-negative if and only if $c \geq \max_{i\in V}\{(d_i - w_{ii})/f_i\}$. The drawback is that selfloop weights only affect volumes but not the cut weights. In general there is no relation between volumes and cut weights in the $\mathbf{f}$-selflooped graph.

**f-scaling.** Another natural strategy to adapt the row sums of $W$ to any prescribed vector $\mathbf{f} \in \mathbb{R}^n_{>0}$ is by proportional scaling, that is by considering $FD^{-1}W$. This matrix yields the row sums $\mathbf{f}$ exactly, but is not symmetric in general. For that reason, we use a (multiplicative) symmetrization:

$$\tilde{W}_{\mathbf{f}} \;:=\; \sqrt{FD^{-1}}W\sqrt{FD^{-1}} \quad\in\; \mathbb{W}.$$

We refer to $\mathcal{G}(\tilde{W}_{\mathbf{f}})$ as the $\mathbf{f}$-scaled graph. The drawback of this approach is that its degree vector $\tilde{\mathbf{d}} = \tilde{W}_{\mathbf{f}}\mathbf{1}$ does *not* equal $\mathbf{f}$ exactly, since $\tilde{d}_i = \sum_{j\in V} w_{ij}\sqrt{(f_if_j)/(d_id_j)}$ differs from $f_i$ in general. In the following let $\tilde{D}_{\mathbf{f}} := \operatorname{diag}(\tilde{\mathbf{d}})$.

**f-adjusting.** For any $\mathbf{f} \in \mathbb{R}^n_{>0}$ and $c > 0$, we combine both approaches by first applying $\mathbf{f}$-scaling, followed by $c\mathbf{f}$-selflooping. This defines the weak graph matrix

$$\overline{W}_{\mathbf{f},c} \;:=\; \tilde{W}_{\mathbf{f}} - \tilde{D}_{\mathbf{f}} + cF \quad\in\; \mathbb{W}_{\ominus}.$$

We refer to $\mathcal{G}(\overline{W}_{\mathbf{f},c})$ as the $(\mathbf{f}, c)$**-adjusted graph** of $G$, and for the case $c = 1$ simply as the **$\mathbf{f}$-adjusted graph**. All vertex degrees are positive because $\overline{W}_{\mathbf{f},c}\mathbf{1} = c\mathbf{f} > 0$. The $(\mathbf{f}, c)$-adjusted graph has no negative selfloop if and only if $c \geq \max_{i\in V}\{(\tilde{d}_i - \tilde{w}_{ii})/f_i\} =: c^*$. In this case it holds that $\overline{W}_{\mathbf{f},c} \in \mathbb{W}$. Although this definition of $\mathbf{f}$-adjusting appears somewhat artificial at the first glance, it reveals appealing properties in its algebraic and geometric interpretation.

**f-adjusted Laplacian.** For $W \in \mathbb{W}$, the unnormalized Laplacian matrix $L(W) := D - W$ and the normalized Laplacian matrix $\mathcal{L}(W) := I - \tilde{W}_{\mathbf{1}}$ are powerful and well-studied objects (see Chung, 1997). We define

$$\mathcal{L}_{\mathbf{f}}(W) \;:=\; \mathcal{L}(\overline{W}_{\mathbf{f},1})$$

for any $\mathbf{f} \in \mathbb{R}^n_{>0}$ as the **$\mathbf{f}$-adjusted Laplacian** of $\mathcal{G}(W)$.

## 4. Algebraic Interpretation

All diagonal modifications of the form $W + X \in \mathbb{W}_{\ominus}$ for $X \in \operatorname{diag}(\mathbb{R}^n)$ obviously represent all possible selfloop modifications of $W$, that is $\mathbf{f}$-selflooping for every $\mathbf{f} \in \mathbb{R}^n_{>0}$. In this section, we derive a similar result for the normalized Laplacian: all diagonal modifications of the form $\mathcal{L}(W) + X \in \mathcal{L}(\mathbb{W}_{\ominus})$ characterize all possible $\mathbf{f}$-adjustments of $W$. In this sense, $\mathbf{f}$-adjusting deals with diagonal modifications of the normalized Laplacian just in the same way as $\mathbf{f}$-selflooping deals with diagonal modifications of the adjacency matrix. This shows that $\mathbf{f}$-adjusting is a natural graph modification. All proofs and further details are provided in the supplement.

### 4.1. $(\mathbf{f}, c)$-adjusting is as powerful as $\mathbf{f}$-adjusting

The following lemma shows that it is sufficient to focus on $\mathbf{f}$-adjusting in order to study every $(\mathbf{f}, c)$-adjusting.

**Lemma 4.1 (Scaling Relation).** *For all $W \in \mathbb{W}$, $\mathbf{f} \in \mathbb{R}^n_{>0}$ and $c > 0$ it holds that*

$$\mathcal{L}(\overline{W}_{\mathbf{f},1}) = c \cdot \mathcal{L}(\overline{W}_{\mathbf{f},c}).$$

Consequently, every $\mathcal{L}(\overline{W}_{\mathbf{f},c})$ has the same eigenvectors as $\mathcal{L}(\overline{W}_{\mathbf{f},1})$, with all eigenvalues scaled by $c^{-1}$. In particular the order of the eigenvalues is preserved. This implies further that $\mathcal{L}(\overline{W}_{\mathbf{f},c})$ inherits many spectral properties from $\mathcal{L}(\overline{W}_{\mathbf{f},c^*})$ such as its positive semi-definiteness and that $\sqrt{\mathbf{f}}$ is an eigenvector to eigenvalue 0 whose multiplicity matches the number of connected components in $\mathcal{G}(W)$. For that reason, the spectral analysis of any $(\mathbf{f}, c)$-adjusted graph can be reduced to the $\mathbf{f}$-adjusted graph. In particular it does not matter whether or not it has negative selfloops.

### 4.2. f-adjustments are diagonally modified Laplacians

The next lemma shows that every $\mathbf{f}$-adjustment represents a meaningful modification along the main diagonal of $\mathcal{L}(W)$.

**Lemma 4.2 (Diagonally Modified Laplacian).** *For all $W \in \mathbb{W}$ and $\mathbf{f} \in \mathbb{R}^n_{>0}$ it holds that*

$$\mathcal{L}_{\mathbf{f}}(W) \;=\; \tilde{D}_{\mathbf{f}}F^{-1} - \tilde{W}_{\mathbf{1}}.$$

The main insight provided by Lemma 4.2 is that we can think of $\mathbf{f}$-adjusting as of replacing the identity matrix in $\mathcal{L}(W) = I - \tilde{W}_{\mathbf{1}}$ with the new matrix $\tilde{D}_{\mathbf{f}}F^{-1}$. Its diagonal entries reflect the relative deviation $\tilde{\mathbf{d}}/\mathbf{f}$ between the intended new degrees $\mathbf{f}$ and the degrees $\tilde{\mathbf{d}}$ that are obtained by $\mathbf{f}$-scaling without subsequent $\mathbf{f}$-selflooping. Note that $\mathbf{f}$-scaling alone does not reveal any clear relation to $\mathcal{L}(W)$, only its combination with the subsequent $\mathbf{f}$-selflooping collapses down to this "simple" algebraic form. We denote any modification along the main diagonal of a matrix as a **diagonal modification**. Whenever it exceeds just a tiny perturbation, it has a strong non-linear impact on the spectrum: the eigenvalues can only loosely and abstractly be bounded by Horn's inequalities (Bhatia, 2001), and nothing is known on the impact on the eigenvectors.

### 4.3. Diagonally modified Laplacians are f-adjustments

Lemma 4.2 shows that every $\mathbf{f}$-adjusting can be understood as a diagonal modification of the form $\mathcal{L}(W) + X = \mathcal{L}(A)$ for some $X \in \operatorname{diag}(\mathbb{R}^n)$ and $A \in \mathbb{W}_{\ominus}$. This rises the question whether also the converse is true: does *every* such diagonal modification imply that $A$ is an $\mathbf{f}$-adjustment of $W$? The following theorem gives a positive answer.

**Theorem 4.3 (Complete Characterization).** *For any $W \in \mathbb{W}$ with $\mathcal{L}(W) \neq 0$ consider all solutions $(X, A, c) \in \mathrm{diag}(\mathbb{R}^n) \times \mathbb{W}_\ominus \times \mathbb{R}$ of the equation*

$$\mathcal{L}(W) + X = c \cdot \mathcal{L}(A).$$

*For $c \leq 0$ no solution exists. For $c > 0$, all solutions are given by $A = \overline{W}_{\mathbf{f},c}$ and $X + I = \tilde{D}_{\mathbf{f}} F^{-1} = Z$ for any choice of $\mathbf{f} \in \mathbb{R}^n_{>0}$. For connected $\mathcal{G}(W)$, choosing $\mathbf{f}$ is equivalent to choosing any $Z \in \mathrm{diag}(\mathbb{R}^n_{>0})$ with spectral radius $\rho(Z^{-1}\tilde{W}_{\mathbf{1}}) = 1$. This determines $\sqrt{\mathbf{f}}$ uniquely (up to scaling) as the eigenvector corresponding to the simple eigenvalue 1 of the matrix $Z^{-1}\tilde{W}_{\mathbf{1}}$.*

Theorem 4.3 gives two further results for connected graphs:

(i) In order to transform $\mathcal{L}(W)$ by a diagonal modification into another normalized Laplacian matrix, we are free to orient the identity matrix in $\mathcal{L}(W) = I - \tilde{W}_{\mathbf{1}}$ toward any "direction" $Y \in \mathrm{diag}(\mathbb{R}^n_{>0})$. But at the same time this determines only a single possible "distance" $\mu := \rho(Y^{-1}\tilde{W}_{\mathbf{1}})$ for which indeed it holds that $\mu Y - \tilde{W}_{\mathbf{1}} \in \mathcal{L}(\mathbb{W}_\ominus)$, because only then $\rho((\mu Y)^{-1}\tilde{W}_{\mathbf{1}}) = 1$. Moreover, it follows that $\mu Y - \tilde{W}_{\mathbf{1}} = \mathcal{L}(\overline{W}_{\mathbf{f},1})$, where $\sqrt{\mathbf{f}}$ is the eigenvector of $(\mu Y)^{-1}\tilde{W}_{\mathbf{1}}$ corresponding to the simple eigenvalue 1.

(ii) We get deeper insights into the structure of the deviation error $\tilde{\mathbf{d}}/\mathbf{f}$ under $\mathbf{f}$-scaling: given any vector of relative deviations $\boldsymbol{\xi} \in \mathbb{R}^n_{>0}$, there exists a *unique* (up to scaling) vector $\mathbf{f} \in \mathbb{R}^n_{>0}$ that satisfies $\tilde{\mathbf{d}}/\mathbf{f} = \alpha\boldsymbol{\xi}$ for a unique $\alpha > 0$. In particular this implies that no vector $\mathbf{f} \neq \mathbf{d}$ can be obtained by $\mathbf{f}$-scaling *exactly* as the new degree vector $\tilde{\mathbf{d}}$, not even up to scaling, because the zero-deviation case $\boldsymbol{\xi} = \mathbf{1}$ is already reached for $\mathbf{f} = \mathbf{d}$.

These two results generalize to unconnected graphs by applying them individually to each connected component. In particular the uniqueness is affected in the way that $\mathbf{f}$ may be scaled individually within each connected component.

### 4.4. Application: $\mathbf{f}$-adjusted spectral clustering

We are going to show that the spectral clustering technique generalizes to weak graph matrices. This allows to apply it to any $\mathbf{f}$-adjusted graph. Spectral clustering of a connected graph $G = (V, E, W)$ relies on a relaxation for solving the NP-hard problem of minimizing the normalized cut value

$$NCut(S, \overline{S}) = \mathrm{cut}_G(S, \overline{S}) \left(\mathrm{vol}_G(S)^{-1} + \mathrm{vol}_G(\overline{S})^{-1}\right)$$

over all $S \subseteq V$ and $\overline{S} := V \backslash S$. Instead of the optimal characteristic vector, the relaxation determines a real-valued score vector $\mathbf{s} := \sqrt{D^{-1}}\mathbf{v}_2$, where $\mathbf{v}_2$ denotes the second smallest eigenvector of $\mathcal{L}(W)$. Thresholding is used to define $S$ from $\mathbf{s}$. We refer to $\mathbf{s}$ as the **NCut score vector**. Variants use multiple smallest eigenvectors $\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_\ell$ for embedding vertex $i \in V$ in $\mathbb{R}^{\ell-1}$ at the $i$'th coordinate entries of $\mathbf{v}_2, \ldots, \mathbf{v}_\ell$, and then apply for example $k$-means on the embedded points in order to identify $k$ clusters.

Since the $(\mathbf{f}, c^*)$-adjusted graph has no negative selfloops, we can apply any spectral clustering technique to $\mathcal{G}(\overline{W}_{\mathbf{f},c^*})$ as usual. Lemma 4.1 shows that $\mathcal{L}(\overline{W}_{\mathbf{f},c^*})$ and $\mathcal{L}_{\mathbf{f}}(W)$ share the same eigenvectors in the same order. Thus, the NCut score vector does not change by considering $\mathcal{L}_{\mathbf{f}}(W)$ instead of $\mathcal{L}(\overline{W}_{\mathbf{f},c^*})$. Consequently, in all variants of spectral clustering we can simply replace $\mathcal{L}(W)$ by $\mathcal{L}_{\mathbf{f}}(W)$ for any $\mathbf{f} \in \mathbb{R}^n_{>0}$ in order to study the normalized cuts of the graphs $\mathcal{G}(\overline{W}_{\mathbf{f},c})$ for all $c > 0$ simultaneously, even though they contain negative selfloops for $c < c^*$.

## 5. Geometric Interpretation

Assume that we are given a graph $G$ plus the meta-information that it is some neighborhood graph built on an unknown sample from an unknown density. Our goal is to infer from $G$ about structural properties of the underlying density. The additional meta-information allows to interpret certain quantities in $G$, such as volumes and cut weights, in a meaningful way.

In the following we show that our *modification* of $G$ still keeps the geometric interpretation of volumes and cut weights in the *modified* graph $\overline{G}$. Thus we can think of modifying edge weights as of modifying the underlying density, although the unknown sample points remain fixed at their original positions. These insights provide answers to the following questions: how to interpret clustering results for $\overline{G}$? How to choose $\mathbf{f}$ in a meaningful way? Can we use $\mathbf{f}$ to "merge" external information into the graph in order to extract hidden information from $G$?

### 5.1. Geometric graphs

In a random geometric graph, each vertex $i \in V$ is identified with a sample point $x_i \in \mathbb{R}^d$ drawn i.i.d. according to some continuous probability density $p : \mathcal{X} \to \mathbb{R}$ on a compact domain $\mathcal{X} \subseteq \mathbb{R}^d$. The edge set and weights are given from the construction of a **neighborhood graph** on these sample points. Prominent choices are:

(i) the **Gaussian graph**: the complete graph with Gaussian weights $w_{ij} = \exp(-\frac{1}{2}\|x_i - x_j\|^2/\sigma^2))$

(ii) the **unweighted $r$-graph**, which has an edge $ij$ of unit weight if and only if $\|x_i - x_j\| \leq r$

(iii) the Gaussian **weighted $k$NN graph**, which has an edge $ij$ of Gaussian weight if and only if $x_j$ is among the nearest $k$ neighbors of $x_i$ or vice versa.

The results in this section refer to the limit case $n \to \infty$. We consider the following **convergence conditions**: for the Gaussian graph we require that $n \to \infty$, $\sigma \to 0$, $n\sigma^{d+1}/\log n \to \infty$. For the $r$-graph that $n \to \infty$, $r \to 0$, $nr^{d+1}/\log n \to \infty$. For the Gaussian weighted $k$NN graph that $n \to \infty$, $k \to \infty$, $k/\log n \to \infty$, $n\sigma^{d+1} \to \infty$, $(k/n)^{1/d} \geq \sigma^\alpha$ for some $\alpha \in ]0, 1[$.

We require that the following regularity conditions are satisfied: $\mathcal{X}$ has a smooth boundary with bounded curvature. $p$ is twice differentiable with bounded gradient, and bounded away from zero, that is $p(x) \geq p_{min} > 0$ for $x \in \mathcal{X}$.

### 5.2. Different views on graphs

We introduce a distinction between three different "views" on the graph. These views help us to interpret graph modifications in terms of modifications applied to the underlying probability distribution.

**Discrete graph view.** This is the well-known setting of the discrete graph with $\mathrm{cut}_G$ and $\mathrm{vol}_G$ as introduced in Section 3.1. Specifically, we do not have access to the information about the underlying space. This is the standard situation for algorithms like spectral clustering or Isomap. In particular, in the graph view we cannot evaluate index sets of the form $\{i \mid x_i \in A\}$ where $A$ is a subset of $\mathbb{R}^d$.

**Continuous space view.** Here we define volumes and cuts with respect to the underlying density $p$. Specifically, we define the **$p$-volume** of a measurable $A \subseteq \mathcal{X}$ as $\mathrm{vol}_p(A) := p(A) = \int_A p(x)\,dx$. Given a hyperplane $H$, we interpret it as a cut of the underlying space in the two half-spaces denoted by $H^+$ and $H^-$. We denote its corresponding **$p$-weight** as $\mathrm{cut}_p(H) := \int_H p(x)\,dx$. In this work we focus on cuts induced by hyperplanes. We believe that all stated results can be generalized to any other cut surfaces that are sufficiently regular.

**Interspace view.** This view "mediates" between the discrete and continuous world. Its objects are not accessible to algorithms on graphs, but they serve as a theoretical construction to express the relationships between discrete and continuous cuts and volumes. In this view, we assume that we can evaluate index sets of the form $V(A) := \{i \mid x_i \in A\} \subseteq V$ where $A$ is a subset of $\mathbb{R}^d$. For any vector $\mathbf{d}$ and any sample drawn from $p$, we refer to $\mathrm{vol}_{\mathbf{d}}(A) := \mathrm{vol}_G(V(A)) = \sum_{x_i \in A} d_i$ as the **interspace volume** of $A$ with respect to vector $\mathbf{d}$. Intuitively, it sums over the degrees as provided in the discrete graph view, while the sum is indexed by the sample points as provided by the continuous space view. Moreover, let $\mathrm{cut}_W(H) := \mathrm{cut}_G(V(H^-), V(H^+)) = \sum_{x_i \in H^-, x_j \in H^+} w_{ij}$ denote the **interspace cut weight** of the cut $H$ with respect to weight matrix $W$. Again, it sums over the edge weights as in the graph view, while the summands are indicated by the sample points in the space view.

### 5.3. Geometric problem statement

We already stated the problem informally as modifying a given neighborhood graph from density $p$ such that its volumes and cuts "looks like" those of a neighborhood graph from an alternative density $\overline{p}$. We are now ready to state this requirement formally by using the interspace view concept.

**Geometric Graph Adjustment Problem** *Given a neighborhood graph $G = (V, E, W)$ built on a sample from density $p : \mathcal{X} \to \mathbb{R}$, and a second density $\overline{p} : \mathcal{X} \to \mathbb{R}$. How can we define a modified graph $\overline{G} = (V, \overline{E}, \overline{W})$ such that, in the limit case, $\mathrm{vol}_{\overline{\mathbf{d}}}(A) \approx \mathrm{vol}_{\overline{p}}(A)$ for all measurable sets $A \subseteq \mathcal{X}$ and $\mathrm{cut}_{\overline{W}}(H) \approx \mathrm{cut}_{\overline{p}}(H)$ for all cuts $H$?*

The vertex set and the corresponding sample points remain fixed throughout the modification. All that is allowed to change are edges and edge weights. As the term *modification* indicates, $\overline{G}$ should stay "similar" to $G$. This can be quantified in different ways. Here, we focus on the constraint that the edge sets $\overline{E}$ and $E$ coincide up to selfloops.

It is well known that $|V(A)|$ is proportional to $p(A)$ in expectation and that the degree $d_i$ of vertex $i$ can serve, up to global scaling, as a consistent estimate of the underlying density $p(x_i)$ at sample point $x_i$. This motivates the following obvious approach to fit $\overline{G}$'s interspace volumes to $\overline{p}$-volumes: determine $f_i$ proportional to $\overline{p}(x_i)$, and use $\mathbf{f}$-selfflooping to let the modified graph $\overline{G}$ attain $\mathbf{f}$ as its degree vector. However it is not obvious how to define $\overline{G}$ such that further the interspace cut weights in $\overline{G}$ correspond to $\overline{p}$-weights. We will see in the following that this is provided by $\mathbf{f}$-scaling. Since $\mathbf{f}$-adjusting combines both approaches, it provides both goals simultaneously. Moreover, if we want to avoid negative selfloops in $\overline{G}$, then we can apply $(\mathbf{f}, c^*)$-adjusting, which keeps all cut weights and volume proportions intact; it just scales all volumes by $c^*$.

### 5.4. Convergence of f-adjusted volumes

Recall that the interspace volume of any $A \subseteq \mathbb{R}^d$ is defined as $\mathrm{vol}_{\mathbf{d}}(A) = \sum_{x_i \in A} d_i$. This expression already shows how we can intuitively think of discrete vertex degrees as a weight distribution on the underlying space: the positions of the sample points are distributed according to density $p$, and each sample point is additionally weighted by $d_i$. Since the degree $d_i$ can serve as a density estimate for $p(x_i)$, this weighting brings in an additional factor $p$. Therefore we expect that the interspace volume $\mathrm{vol}_{\mathbf{d}}(A)$ behaves like the $p^2$-volume $\int_A p^2(x)dx$ in the limit case. Moreover, if we consider any graph modification that changes the original degree vector $\mathbf{d}$ into $\mathbf{f}$, then this affects the above expression by replacing each individual $d_i$ by $f_i$, while the index set $\{i \mid x_i \in A\}$ remains the same. Thus, if the new vertex degrees $f_i$ are given as $f_i := f(x_i)$ for some continuous function $f : \mathcal{X} \to \mathbb{R}_{>0}$ then we expect that the interspace volume $\mathrm{vol}_{\mathbf{f}}(A)$ behaves like the $fp$-volume $\int_A f(x)p(x)dx$ in the limit case. Indeed this intuition is correct. The following proposition shows that the interspace volume and the continuous $fp$-volume are proportionally related as $n \to \infty$.

**Proposition 5.1. (Interspace Volumes)** *Let $G$ be a geometric graph based on $n$ vertices drawn according to $p$. Denote its degree vector by $\mathbf{d}$ and let $f : \mathcal{X} \to \mathbb{R}_{>0}$ be a continuous function. Define the vector $\mathbf{f} := (f(x_i))_i$, and let $\overline{G}$ be any graph modification of $G$ that attains the degrees $\overline{\mathbf{d}} = \mathbf{f}$. Then, under the convergence conditions mentioned above, for any measurable $A \subset \mathbb{R}^d$, $C \cdot \mathrm{vol}_{\overline{\mathbf{d}}}(A) \to \mathrm{vol}_{f \cdot p}(A)$ almost surely as $n \to \infty$, where $C$ is a scaling constant that depends on $n$, $d$.*

This result shows that we can modify the graph such that the interspace volume of the new graph corresponds to any density $\overline{p}$ we would like, as long as $\overline{p}$ is absolutely continuous with respect to $p$ (that is, sets $A$ with $p(A) = 0$ also have $\overline{p}(A) = 0$). In Section 5.6 and Section 2 we outline a number of important consequences and applications.

### 5.5. Convergence of f-adjusted cut weights

In order to study the cut weights after $(\mathbf{f}, c)$-adjusting, we can solely focus on $\mathbf{f}$-scaling, since it provides exactly the same cut weights. Again we can derive the intuition from the interspace view: let $\mathrm{cut}_W(H) = \sum_{x_i \in H^-, x_j \in H^+} w_{ij}$ denote the interspace cut weight according to any hyperplane $H$. $\mathbf{f}$-scaling replaces each edge weight $w_{ij}$ in this sum by the new weight $\tilde{w}_{ij} := w_{ij} \cdot \sqrt{(f_i f_j)/(d_i d_j)}$. Assume that the original interspace cut weight $\mathrm{cut}_W(H)$ behaves like the $p^2$-weight $\int_H p^2(x)dx$ in the limit. Then we expect due to $\sqrt{(f_i f_j)/(d_i d_j)} \sim f p^{-1}$ that the modified interspace cut weight $\mathrm{cut}_{\tilde{W}}(H)$ behaves like the $fp$-weight $\int_H f(x)p(x)dx$. The following proposition makes this intuition explicit, by relating the interspace cut weights after $\mathbf{f}$-scaling to the continuous $fp$-weights.

**Proposition 5.2. (Interspace Cuts)** *Let $G$ be a geometric graph based on $n$ vertices drawn according to $p$. Denote its degree vector by $\mathbf{d}$ and let $f : \mathcal{X} \to \mathbb{R}_{>0}$ be a continuous function that is twice differentiable and has bounded gradient. Define the vector $\mathbf{f} := (f(x_i))_i$, and let $\tilde{G}$ be the corresponding $\mathbf{f}$-scaled graph with weight matrix $\tilde{W}_{\mathbf{f}}$. Consider a hyperplane $H$ in $\mathbb{R}^d$. Then, under the convergence conditions mentioned above, $C \cdot \mathrm{cut}_{\tilde{W}_{\mathbf{f}}}(H) \to \mathrm{cut}_{f \cdot p}(H)$ almost surely as $n \to \infty$, where $C$ is a scaling constant that depends on $n, d$.*

The proofs of both propositions in this section are based on the arguments in Maier et al. (2009), who study volumes and cut weights in neighborhood graphs.

Note that the results on the interspace cut weights are perfectly aligned with the results on the interspace volumes in the sense that both share the same integrand $\int f(x)p(x)dx$. Since $\mathbf{f}$-adjusting provides the same cut weights as $\mathbf{f}$-scaling, and further degree vector $\mathbf{f}$, we get that the interspace volumes and interspace cuts of the $\mathbf{f}$-adjusted graph both converge to the continuous $fp$-volumes and $fp$-cuts, respectively (up to scaling). Further, $(\mathbf{f}, c)$-adjusting simply puts another global scaling factor $c$ on all volumes.

### 5.6. Interesting consequences

The intuitive interpretation of the geometric results is as follows: the continuous volumes and cuts change under $\mathbf{f}$-adjusting from $\int p^2$ to $\int fp$. This implies a number of interesting special cases:

(1) In the original graph (case $\mathbf{f} = \mathbf{d}$), both interspace volumes and interspace cuts correspond to the continuous quantities of the density $p^2$. This shows that the original notion of volume and cut in the graph has an artifact in the geometric setting in the sense that it does not correspond to volumes according to the original density, but to the squared density. See below for an illustration.

(2) $\mathbf{f}$-adjusting to uniform vertex degrees $\mathbf{f} = \mathbf{1}$ corresponds to the continuous quantities of the original density $p$. This choice removes the artifact of (1).

(3) $\mathbf{f}$-adjusting to inverse degrees $\mathbf{f} = \mathbf{d}^{-1}$ corresponds to a continuous density that is uniform. This can be used to remove density information altogether and make the underlying volumes behave "uniformly". The density removal application in Section 2.1 relies on this strategy.

(4) More generally, all three above cases are modifications of the form $f(x) = g(p(x))$ for some $g : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$. In practice, we can replace the $p(x)$ inside $g$ by any density estimate (such as the degrees in the original graph). If, additionally, $g(\alpha x) = const(\alpha)g(x)$ for all $\alpha > 0$, then any global scaling factor of the density estimate simply translates in a global scaling factor of the continuous quantities. This approach describes the target density $\overline{p}$ implicitly, relative to $p$. See the supplement for an application to the biased random walks studied by Zlatić et al. (2010), and for further details on the implicit definition of $\overline{p}$ from $p$.

(5) For any twice differentiable function $f : \mathcal{X} \to \mathbb{R}_{>0}$, define the new degree vector $\mathbf{f} = (f(x_i))_i$. In this setting, interspace cuts and volumes correspond to the quantities as provided by the density $f \cdot p$. This case is particularly interesting for spatial corrections of biased sample data, as in Section 2.2, or if sample coordinates are known.

**Anomaly in Volumes and Cuts** As mentioned above in (1), cuts and volumes in the original graph represent integrals over the *squared* density $p^2$. To illustrate this anomaly, we consider in Figure 3 the density $p : [0,1]^2 \to \mathbb{R}$, $p(x, y) = 2x$. Let $G = (V, E, W)$ denote the 200-NN graph built on 5000 sample points drawn from $p$ with Gaussian weighted edges ($\sigma = 0.03$).

*Volume anomaly*: assume that we want to partition $V$ into $L \cup R$ by splitting the sample points at some $x \in [0, 1]$ into vertices to its left $L$ and to its right $R$. The split point should be such that $L$ and $R$ cover the same probability

mass of the underlying density $p$. The correct approach is then to choose $L$ and $R$ such that $|L| \approx |R|$, which defines $x_2$ in Figure 3. In contrast to that, constraining to $\mathrm{vol}_G(L) \approx \mathrm{vol}_G(R)$ expands low density regions, since the split point $x_3$ is now implicitly set according to the squared density. Constraining to $\sum_{i\in L} d_i^{-1} \approx \sum_{j\in R} d_j^{-1}$ gives $x_1$. This lets $L$ and $R$ cover the same uniform amount of $p$'s support $\{x \in \mathbb{R}^d \mid p(x) > 0\}$, hiding any other density information. Note that the analytically expected value of $x_i$ is $0.5^{1/i}$, which gives roughly $(0.5, 0.71, 0.79)$.

*Cut weight anomaly*: consider vertical cuts at all positions $x \in [0.1, 0.9]$. Fitting a cubic polynomial $ax^3+bx^2+cx+d$ to the cut weights of the original graph gives, averaged over 10 runs, the coefficients $(-0.06, 1.30, -0.04, 0.01)$, which clearly shows the quadratic behavior. For the $\mathbf{1}$-adjusted graph we get $(-0.08, -0.08, 1.13, -0.11)$, which is basically linear as expected, since it corresponds to $p$ without squaring. Finally, the $\mathbf{d}^2$-adjusted graph provides $(1.21, 0.16, -0.06, 0.01)$ which has a significant cubic term as expected, since its cuts integrate over $p^3$.

Whenever we are interested in $p$ (which is usually the case), we must *not* consider the original graph, but the $\mathbf{1}$-adjusted graph, as it provides the desired volumes and cuts. In particular, spectral clustering of a sample drawn from $p$ approximates the normalized cut of $p^2$, not of $p$. We can remove this artifact by replacing the typically used normalized Laplacian $\mathcal{L}(W)$ with the $\mathbf{1}$-adjusted Laplacian $\mathcal{L}_{\mathbf{1}}(W)$. But note that squaring $p$ can be *beneficial* for the clustering result, since it emphasizes high density clusters. However, we are no longer restricted to squaring: we can approximate the normalized cut of the modified density $p^r$ for any $r \in \mathbb{R}$ simply by taking the $\mathbf{d}^{r-1}$-adjusted Laplacian for spectral clustering.

## 6. Further Related Work

Another approximative strategy to change the degree vector of a graph to some vector $\mathbf{f}$ is iterative matrix scaling. Pukelsheim (2014) presents recent results on Iterative Proportional Fitting, which scales row and column sums alternately to $\mathbf{f}$ until convergence. Knight et al. (2014) study a multiplicative symmetrization for $\mathbf{f} = \mathbf{1}$. Iterative matrix scaling allows for a statistical interpretation, as it converges to the relative entropy closest solution among all non-negative matrices that provide the degree vector $\mathbf{f}$ (Csiszar, 1975). However, in contrast to $\mathbf{f}$-adjusting, no geometric interpretation of the resulting cut weights is known.

Bapat et al. (2001) study diagonal modifications of the unnormalized Laplacian. Note that $L(W) + X \notin L(\mathbb{W}_\ominus)$ for all $X \neq 0$, hence no diagonal modification of the unnormalized Laplacian represents the unnormalized Laplacian of any other graph. Nevertheless, diagonal modifications of $L(W)$ can provide useful meta-information on $G$. Wu et al. (2012) consider this diagonal modification for $X \geq 0$, and show how to interpret $(L(W)+X)^{-1}X$ as meaningful random walk absorption probabilities on $G$.

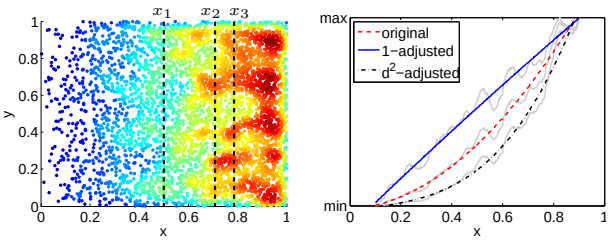


*Figure 3.* Left: 5000 sample points with heat colors indicating the degrees in the weighted $k$NN-graph. Dotted lines mark from left to right the estimated split points $(x_1, x_2, x_3) \approx (0.51, 0.7, 0.77)$ as described in text. Right: polynomials fitted to the cut weights of a single instance of the original graph $G$, its $\mathbf{1}$-adjustment and its $\mathbf{d}^2$-adjustment, for vertical cuts at $x \in [0.1, 0.9]$.

## 7. Discussion

We introduce $\mathbf{f}$-adjusting as a transformation of a given graph $G$ into another graph $\overline{G}$. We show that $\mathbf{f}$-adjusting corresponds to a natural diagonal modification of the normalized Laplacian that further allows for a clear geometric interpretation in terms of a density shift.

The *algebraic interpretation* shows that $\mathbf{f}$-adjusting represents all diagonal modifications of $\mathcal{L}(W)$ of the form $\mathcal{L}(W) + X \in \mathcal{L}(\mathbb{W}_\ominus)$. This is the normalized Laplacian's pendant to the fact that $\mathbf{f}$-selflooping represents all diagonal modifications of $W$ of the form $W + X \in \mathbb{W}_\ominus$. Thus, $\mathbf{f}$-adjusting acts on the normalized Laplacian in the same way as $\mathbf{f}$-selflooping acts on the adjacency matrix.

For the *geometric interpretation* we introduce an explicit distinction between three different views on a graph. In particular the interspace view is a helpful tool, since it serves as a bridge between the discrete graph world and the continuous world of density functions beneath. In terms of these views we express how $\mathbf{f}$-adjusting represents a modification applied to the underlying density $p$. As a result, volumes and cuts in $\overline{G}$ refer to well-specified continuous quantities. This allows to apply any volume and cut based algorithm to the graph as if it were drawn according to another distribution.

More experiments and details can be found in the supplementary material.

**Acknowledgements** This research is supported by the German Research Foundation via the Research Unit 1735 "*Structural Inference in Statistics: Adaptation and Efficiency*" and grant LU1718/1-1.

## References


Bapat, R. B., Kirkland, S. J., and Pati, S. The perturbed Laplacian matrix of a graph. *Linear and Multilinear Algebra*, 49(3):219–242, 2001.

Bhatia, R. Linear Algebra to Quantum Cohomology: The Story of Alfred Horn's Inequalities. *The American Mathematical Monthly*, 108(4):289–318, 2001.

Chung, F. R. K. *Spectral Graph Theory*. American Mathematical Society, 1997.

Csiszar, I. $I$-Divergence Geometry of Probability Distributions and Minimization Problems. *The Annals of Probability*, 3(1):146–158, 1975.

Knight, P., Ruiz, D., and Uçar, B. A Symmetry Preserving Algorithm for Matrix Scaling (to appear). *SIAM Journal on Matrix Analysis and Applications*, 2014.

Maier, M., von Luxburg, U., and Hein, M. Influence of graph construction on graph-based clustering measures. In *Neural Information Processing Systems (NIPS)*, 2009.

Pukelsheim, F. Biproportional scaling of matrices and the iterative proportional fitting procedure. *Annals of Operations Research*, 215(1):269–283, 2014.

Shi, J. and Malik, J. Normalized Cuts and Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8): 888–905, 2000.

Wu, X., Li, Z., So, A., Wright, J., and Chang, S. Learning with Partially Absorbing Random Walks. In *Neural Information Processing Systems (NIPS)*, 2012.

Zlatić, V., Gabrielli, A., and Caldarelli, G. Topologically biased random walk and community finding in networks. *Physical Review E*, 82(6):066109, 2010.