The $\mathbf{f}$-Adjusted Graph Laplacian: a Diagonal Modification with a Geometric Interpretation (Supplement)

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

This supplement consists of three sections. Section A provides further applications that were skipped from the paper. Section B provides all proofs for paper's Section 4 (Algebraic Interpretation), including several further details. Section C provides details for paper's Section 5 (Geometric Interpretation).

A. Further Applications

The first application studies a certain approach of biased random walks from our new geometric perspective. The second application deals with Semi-Supervised Learning.

A.1. Topologically biased random walks

$\mathbf{f}$-adjusting a graph changes most of the properties of the random walk on the graph. The stationary distribution is now obviously proportional to $\mathbf{f}$ rather than $\mathbf{d}$. Hitting times and commute times are affected as well. Zlatić et al. (2010) study what they call "vertex centered biased random walks": implicitly they modify the random walk matrix $P = D^{-1}W$ in a way that can be represented as $\tilde{P}_{\mathbf{f}} = \tilde{D}_{\mathbf{f}}^{-1}\tilde{W}_{\mathbf{f}}$ for $\tilde{W}_{\mathbf{f}}$ being the $\mathbf{f}$-scaled graph with $\mathbf{f} = \mathbf{d} \odot \exp(2\beta \cdot \mathbf{d}/\|\mathbf{d}\|_\infty)$, where $\odot$ denotes the Hadamard product and $\exp(\cdot)$ is applied entry-wise. They observe that their particular choice leads to much better clustering results. They conclude this from studying eigenvalues and eigenvectors of the symmetrization $\tilde{D}_{\mathbf{f}}^{-1/2}\tilde{W}_{\mathbf{f}}\tilde{D}_{\mathbf{f}}^{-1/2}$. Since these quantities correspond to the eigenvalues and eigenvectors of $\mathcal{L}(\tilde{W})$, their approach can be understood in terms of normalized cuts of the $\mathbf{f}$-scaled graph. Based on our analysis, we can now give an intuitive explanation of their approach: their new random walk aims at studying the modified density $\tilde{p}(x) = \exp(2\beta \cdot p(x)/p_{max}) \cdot p(x)^2$. Hence, they amplify high-density regions in space exponentially stronger than low-density regions, which drastically strengthens any density cluster structure. Of course this only works to a certain extent, because it runs into the same problem as studying the density p^r for large r: suppose that A and B denote disjoint geometric areas of two different clusters at slightly different "density plateaus", that is, there exist $a < b$ such that the level sets to every threshold $t \in [a, b]$ only show, say, A but not B. For any moderate bias/amplification ("small $\beta > 0$", or "small $r > 1$"), A and B are both emphasized over any much lower density areas between them. However, since A and B lie on different plateaus, they are affected in an increasingly different way for a larger amplification. If the amplification is too strong, then B is suppressed like other low-density areas, since A is favored too much over anything else.

Further, note that the exponential scaling in the modified random walk is applied to p^2 instead of p. Our technique suggests that applying the modification $\mathbf{f} = \exp(\beta \cdot \mathbf{d}/\|\mathbf{d}\|_\infty)$ is a more natural choice, since it applies the intended exponential influence directly on p. This implies to study the modified density $\tilde{p}(x) = \exp(\beta \cdot p(x)/p_{max}) \cdot p(x)$, without squaring. Finally, one could fix the influence of the deviation $\tilde{\mathbf{d}} \neq \mathbf{f}$ on the spectrum by considering $\mathcal{L}_{\mathbf{f}}(W)$ everywhere instead of $\mathcal{L}(\tilde{W}_{\mathbf{f}})$.

A.2. Semi-Supervised Learning (SSL)

A.2.1. Merging Vertex Weights into Edge Weights

Assume that we have additional knowledge given as vertex weights. For example in network traffic analysis one has to deal with both edge congestion and vertex congestion. Moreover, it has been shown by Montenegro (2003) and Kannan et al. (2006) that considering both edge bottlenecks and vertex bottlenecks simultaneously can drastically improve mixing time bounds for random walks. However, many algorithms focus only on edge similarities, so does label propagation. In order to make the vertex weights visible to the algorithm, we have to transform them into edge weights. The trivial strategy of introducing selfloops does not affect label propagation at all, since selfloops do not change the probabilities of the random trajectories. Our framework suggests to merge vertex weights into edge weights via $\mathbf{f}$-adjusting (where only the implied $\mathbf{f}$-scaling takes an influence on label propagation).

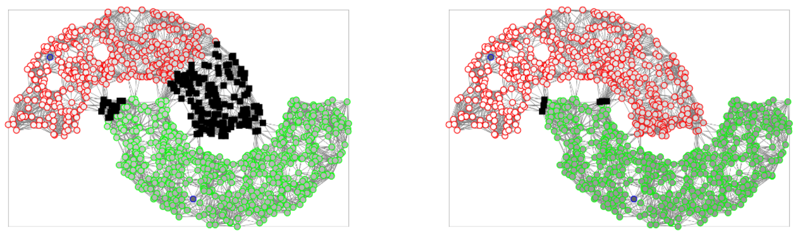

Figure 1. Label propagation on original graph (left) and $\mathbf{f}$-adjusted graph (right), showing Label Propagation on the two moons data set with one labeled point per class (shown in blue). Black squares mark misclassified vertices.

As an example, consider the two-moons dataset for an unweighted 10NN graph on 1000 vertices with one labeled vertex per cluster. Label propagation alone provides the rather bad results in Figure 1 (left). Now we add additional knowledge by computing a measure of local centrality for each vertex: denote by $N_3(i)$ the set of vertices that have shortest path distance exactly 3 to vertex i. Let φ_i be the sum of all pairwise shortest path distances between any two vertices in $N_3(i)$. Then φ_i^{-1} penalizes those vertices that lie close to sharp cluster boundaries. Incorporating these vertex weights by $\mathbf{f}$-adjusting with $f_i := \exp(-5 \cdot \phi_i)$ gives the almost perfect result for label propagation on the adjusted graph (Figure 1, right).

A.2.2. Label Propagation Limit Behavior

The soft labels computed by the label propagation algorithm in case of few labeled vertices tend to be flat, with sharp spikes at the labeled vertices. In this situation, a meaningful threshold is increasingly difficult to find (Nadler et al., 2009). Figure 2 (left) illustrates this problem for a Gaussian-weighted 50NN-graph on 2000 points (x_i, y_i, z_i) sampled from $\mathcal{N}(\mu, 1) \times \mathcal{N}(0, 1) \times \mathcal{N}(0, 0.1)$, where $\mu \in \{0, 4\}$ for the 1000 points of each class, respectively. For each class, a single vertex (denoted as ℓ_a resp. ℓ_b) near to μ is labeled.

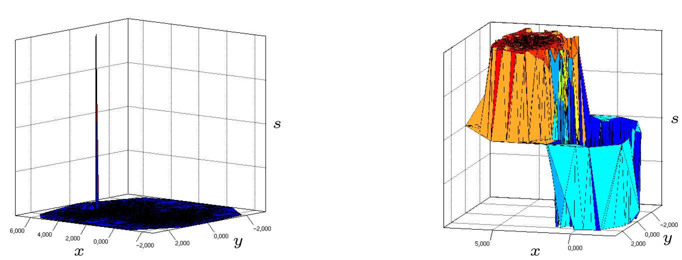

Figure 2. Soft threshold values s of label propagation on original graph (left) and $\mathbf{f}$-adjusted graph (right). Both graphs are built on the same sample drawn from two Gaussians, where each Gaussian is labeled by a single label point near to its center.

Figure 2 (right) shows the soft labels for the $\mathbf{f}$-adjusted graph with $f_i := d_i \cdot \exp(-4 \cdot \min\{\mathrm{hd}(i, \ell_a), \mathrm{hd}(i, \ell_b)\})$, wherein $\mathrm{hd}(x, y)$ denotes the hop-distance, that is the smallest number of edges on an xy-path in G. This adjustment can be interpreted as extending the areas of attraction from single points (the labeled ones) to larger sets (neighborhoods around labeled points). This biases the label propagation algorithm towards any labeled vertex, providing a sharp separation in the soft labels. Instead of hd, any other distance measure can be used, for example some quantity that is derived from a coarser cluster structure.

B. Algebraic Interpretation

First we summarize the notation. Then we derive some general properties of weak graph matrices. Finally, we put some effort into proving the main result of the algebraic interpretation, Theorem 4.3.

B.1. Notation

By the term **graph matrix** we refer to any element of $\mathbb{W} := \{X \in \mathbb{R}^{n\times n}_{\geq 0} \mid X = X^T, X\mathbf{1} > 0\}$. We generalize this to $\mathbb{W}_{\ominus} := \{X \in \mathbb{W} + \operatorname{diag}(\mathbb{R}^n) \mid X\mathbf{1} > 0\} = \{X \in \mathbb{R}^{n\times n}_{\geq 0} + \operatorname{diag}(\mathbb{R}^n) \mid X = X^T, X\mathbf{1} > 0\}$, the set of **weak graph matrices**, whose elements further allow for negative diagonal entries as long as all row sums remain positive. Obviously $\mathbb{W} \subset \mathbb{W}_{\ominus}$.

Every weak graph matrix $W \in \mathbb{W}_{\ominus}$ corresponds to an undirected weighted graph $\mathcal{G}(W) := (V, E, W)$ that has W as its weighted adjacency matrix, and edge set $E := \{ij \mid w_{ij} \neq 0\}$. With $\mathbf{d} := W\mathbf{1}$ and $D := \operatorname{diag}(\mathbf{d})$, we define the unnormalized Laplacian $L(W) = [l_{ij}] := D - W$ and the normalized Laplacian $\mathcal{L}(W) = [\ell_{ij}] := \sqrt{D^{-1}}L(W)\sqrt{D^{-1}} = I - \sqrt{D^{-1}}W\sqrt{D^{-1}}$. If all off-diagonal entries of W are zero, then all vertices in $\mathcal{G}(W)$ are isolated (each with a positive selfloop attached). We refer to this special case as a **trivial graph**. A graph is trivial if and only if $L(W) = \mathcal{L}(W) = 0$. It is non-trivial if and only if $n > 1$ and $w_{ij} \neq 0$ for some $i \neq j$.

For $\mathbf{f} \in \mathbb{R}^n_{>0}$ and $c > 0$ we define the following (weak) graph matrices

for $W \in \mathbb{W}$	:	$\mathbf{f}$-scaled	$\tilde{W}_{\mathbf{f}} := \sqrt{FD^{-1}}W\sqrt{FD^{-1}}$	$\in \mathbb{W}$	where $F := \operatorname{diag}(\mathbf{f})$
for $W \in \mathbb{W}_{\ominus}$	:	$\mathbf{f}$-selflooped	$W^{\circ}_{\mathbf{f}} := W - D + F$	$\in \mathbb{W}_{\ominus}$	
for $W \in \mathbb{W}$	:	$(\mathbf{f}, c)$-adjusted	$\overline{W}_{\mathbf{f},c} := \tilde{W}_{\mathbf{f}} - \tilde{D}_{\mathbf{f}} + cF$	$\in \mathbb{W}_{\ominus}$	where $\tilde{D}_{\mathbf{f}} := \operatorname{diag}(\tilde{W}_{\mathbf{f}}\mathbf{1})$.

We consider the matrix decorations $\tilde{}$, $^\circ$ and $\overline{}$ as operators on the matrix W, with the indices $\mathbf{f}$ and c as additional parameter values. For example, for any matrix $A \in \mathbb{W}$ and any vector $\mathbf{x} \in \mathbb{R}^n_{>0}$ we get that $A^{\circ}_{\mathbf{x}} := A - \operatorname{diag}(A\mathbf{1}) + \operatorname{diag}(\mathbf{x})$. Note that $(\mathbf{f}, c)$-adjusting is equal to $\mathbf{f}$-scaling followed by $c\mathbf{f}$-selflooping, that is $\overline{W}_{\mathbf{f},c} = (\tilde{W}_{\mathbf{f}})^{\circ}_{c\mathbf{f}}$.

For $W \in \mathbb{W}$, we define the **f-adjusted Laplacian** $\mathcal{L}_{\mathbf{f}}(W)$ as the normalized Laplacian of the $\mathbf{f}$-adjusted graph,

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

We study some of its properties in the following. In particular, Lemma 4.2 shows that $\mathcal{L}_{\mathbf{f}}(W)$ is a diagonal modification of $\mathcal{L}(W)$, that is $\mathcal{L}_{\mathbf{f}}(W) = \mathcal{L}(W) + X$ for some $X \in \operatorname{diag}(\mathbb{R}^n)$.

B.2. General properties of weak graph matrices

$\mathbb{W}$ is closed under $\mathbf{f}$-scaling: $W \in \mathbb{W} \Rightarrow \tilde{W}_{\mathbf{f}} \in \mathbb{W}$. However, $\mathbb{W}_{\ominus}$ is *not* closed under $\mathbf{f}$-scaling, since for example $W = \begin{pmatrix} -5 & 6 \\ 6 & 3 \end{pmatrix} \in \mathbb{W}_{\ominus}$ would give that $\tilde{W}_{\mathbf{1}} = \begin{pmatrix} -5 & 2 \\ 2 & 1/3 \end{pmatrix} \notin \mathbb{W}_{\ominus}$. For that reason, we consider $\mathbf{f}$-scaling and $\mathbf{f}$-adjusting only applied to non-weak graph matrices.

$\mathbb{W}_{\ominus}$ is closed under $\mathbf{f}$-selflooping, but $\mathbb{W}$ is *not*: for example $W = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix} \in \mathbb{W}$ gives that $W^{\circ}_{\mathbf{1}} = \begin{pmatrix} -1 & 2 \\ 2 & -1 \end{pmatrix} \notin \mathbb{W}$. However, we achieve some sense of closedness of $\mathbb{W}$ when restricting to $c\mathbf{f}$-selflooping for $c \geq c^*_{W,\mathbf{f}}$ as defined in the next proposition.

Proposition B.1. *Let* $W \in \mathbb{W}_{\ominus}$, $\mathbf{f} \in \mathbb{R}^n_{>0}$ *and* $c > 0$. *Define* $c^*_{W,\mathbf{f}} := \max_{i\in V}\{(d_i - w_{ii})/f_i\} \geq 0$. *Then it holds that* $W^{\circ}_{c\mathbf{f}} \in \mathbb{W}$ *if and only if* $c \geq c^*_{W,\mathbf{f}}$.

Proof. From $W^{\circ}_{c\mathbf{f}} = W - D + cF$ we get that $W^{\circ}_{c\mathbf{f}}$ has no negative selfloop if and only if $w_{ii} - d_i + cf_i \geq 0$ for all $i \in V$. In the case $c \geq c^*_{W,\mathbf{f}}$ we get that $w_{ii} - d_i + cf_i \geq w_{ii} - d_i + \max_{i\in V}\{(d_i - w_{ii})/f_i\} \cdot f_i = \max_{i\in V}\{(d_i - w_{ii})\} - (d_i - w_{ii}) \geq 0$ for all $i \in V$. For $c < c^*_{W,\mathbf{f}}$, choose $k \in \{i \in V \mid (d_i - w_{ii})/f_i = c^*_{W,\mathbf{f}}\}$, and get that $w_{kk} - d_k + cf_k < w_{kk} - d_k + c^*_{W,\mathbf{f}} \cdot f_k = w_{kk} - d_k + (d_k - w_{kk})/f_k \cdot f_k = 0$. ∎

Remark 1. It holds that $c^*_{W,\mathbf{f}} = 0$ if and only if W is trivial (because then $d_i = w_{ii}$ for all $i \in V$).

Remark 2. For non-trivial $W \in \mathbb{W}_\ominus$ and any $\mathbf{f} \in \mathbb{R}^n_{>0}$ we get the non-trivial partition:

$$\{W^\circ_{c\mathbf{f}} \mid c > 0\} \quad = \quad \underbrace{\{W^\circ_{c\mathbf{f}} \mid 0 < c < c^*_{W,\mathbf{f}}\}}_{\subseteq \mathbb{W}_\ominus \setminus \mathbb{W}} \,\dot\cup\, \underbrace{\{W^\circ_{c\mathbf{f}} \mid c \geq c^*_{W,\mathbf{f}}\}}_{\subseteq \mathbb{W}}.$$

Fact B.2. *For every $W \in \mathbb{W}_\ominus$ it holds that $W^\circ_{c\mathbf{d}} \in \mathbb{W}$ for all positive $c \geq c^*_{W,\mathbf{d}}$.*

Proof. For trivial W this is obvious, and for non-trivial W this follows from Remark 2. □

Fact B.3. *For every $W \in \mathbb{W}$, $\mathbf{f} \in \mathbb{R}^n_{>0}$ and $c, a > 0$ it holds that $\tilde{W}_{a\mathbf{f}} = a \cdot \tilde{W}_\mathbf{f}$ and that $\overline{W}_{a\mathbf{f},c} = a \cdot \overline{W}_{\mathbf{f},c}$.*

Proof. By definition we get that $\tilde{W}_{a\mathbf{f}} = \sqrt{aFD^{-1}}W\sqrt{aFD^{-1}} = a\sqrt{FD^{-1}}W\sqrt{FD^{-1}} = a \cdot \tilde{W}_\mathbf{f}$, which further implies that $\tilde{D}_{a\mathbf{f}} = a \cdot \tilde{D}_\mathbf{f}$. It follows that $\overline{W}_{a\mathbf{f},c} = \tilde{W}_{a\mathbf{f}} - \tilde{D}_{a\mathbf{f}} + caF = a \cdot (\tilde{W}_\mathbf{f} - \tilde{D}_\mathbf{f} + cF) = a \cdot \overline{W}_{\mathbf{f},c}$. □

Let us summarize some easy facts on the main diagonal of $L(W) = [l_{ij}]$ and $\mathcal{L}(W) = [\ell_{ij}]$ without proofs:

Fact B.4. *For $W \in \mathbb{W}$ it holds for all $i \in V$ that:*

(i) $0 \leq w_{ii} \leq d_i, \quad 0 \leq w_{ii}/d_i \leq 1$

(ii) $l_{ii} = d_i - w_{ii} \in [0, d_i]$ *with* $l_{ii} = 0$ *if and only if* i *is an isolated vertex.*

(iii) $\ell_{ii} = 1 - w_{ii}/d_i \in [0, 1]$ *with* $\ell_{ii} = 0$ *if and only if* i *is an isolated vertex.*

These results generalize to weak graph matrices (now $w_{ii} < 0$ is possible) as follows:

Fact B.5. *For $W \in \mathbb{W}_\ominus$ it holds for all $i \in V$ that:*

(i) $-\infty < w_{ii} \leq d_i, \quad -\infty < w_{ii}/d_i \leq 1$

(ii) $l_{ii} = d_i - w_{ii} \in [0, \infty)$ *with* $l_{ii} = 0$ *if and only if* i *is an isolated vertex, and* $l_{ii} > d_i$ *if and only if* i *has a negative selfloop.*

(iii) $\ell_{ii} = 1 - w_{ii}/d_i \in [0, \infty)$ *with* $\ell_{ii} = 0$ *if and only if* i *is an isolated vertex, and* $\ell_{ii} > 1$ *if and only if* i *has a negative selfloop.*

Trivially it holds that $W = W^\circ_\mathbf{d}$. The next lemma shows how W and $W^\circ_{c\mathbf{d}}$ are related for all $c > 0$.

Lemma B.6. *For all $W \in \mathbb{W}_\ominus$ it holds that $\mathcal{L}(W) = c \cdot \mathcal{L}(W^\circ_{c\mathbf{d}})$ for all $c > 0$.*

Proof. $\mathcal{L}(W^\circ_{c\mathbf{d}}) = \sqrt{(cD)^{-1}}(cD - (W - D + cD))\sqrt{(cD)^{-1}} = c^{-1}\sqrt{D^{-1}}(D - W)\sqrt{D^{-1}} = c^{-1}\mathcal{L}(W)$. □

This easy yet powerful lemma gives that all normalized Laplacians in $\{\mathcal{L}(W^\circ_{c\mathbf{d}}) \mid c > 0\}$ share the same eigenvectors with their corresponding eigenvalues simply scaled according to c. This implies for example that positive semi-definiteness of $L(W)$ and $\mathcal{L}(W)$ generalizes to non-weak graph matrices, as the next lemma shows.

Lemma B.7. *For $W \in \mathbb{W}_\ominus$ it holds that $\mathcal{L}(W)$ is positive semi-definite. Further, 0 is an eigenvalue of $\mathcal{L}(W)$ with multiplicity equal to the number of connected components in $\mathcal{G}(W)$, and $\sqrt{\mathbf{d}}$ a corresponding eigenvector.*

Proof. Adding/Removing selfloops does not affect the number of connected components. Thus, all results follow directly from Lemma B.6 and Fact B.2, since there exists $c > 0$ with $W^\circ_{c\mathbf{d}} \in \mathbb{W}$ and $\mathcal{L}(W) = c \cdot \mathcal{L}(W^\circ_{c\mathbf{d}})$. All stated properties are well-known for non-weak graph matrices. □

It is obvious that $A = cW$ implies that $\mathcal{L}(A) = \mathcal{L}(W)$. The following lemma shows that for connected graph matrices furthermore equivalence holds.

Lemma B.8. *Let $n > 1$. For connected $\mathcal{G}(W) \in \mathcal{G}(\mathbb{W}_\ominus)$ and any $A \in \mathbb{W}_\ominus$ it holds that $\mathcal{L}(W) = \mathcal{L}(A)$ if and only if $A = c \cdot W$ for some $c > 0$.*

Proof. "$\Leftarrow$". $A = cW$ for $c > 0$ yields $\mathcal{L}(A) = I - \sqrt{(cD)^{-1}}cW\sqrt{(cD)^{-1}} = I - \sqrt{D^{-1}}W\sqrt{D^{-1}} = \mathcal{L}(W)$. "$\Rightarrow$". Let $[\ell_{ij}] := \mathcal{L}(W) = \mathcal{L}(A) =: [\ell^A_{ij}]$. Further let $\mathbf{t} := A\mathbf{1} > 0$. First, we show that $\mathbf{t} = \alpha\mathbf{d}$ for some $\alpha > 0$. By Lemma B.7 we get that $\sqrt{\mathbf{d}}$ is the unique (up to scaling) eigenvector of $\mathcal{L}(W)$ to the eigenvalue 0, similarly $\sqrt{\mathbf{t}}$ for $\mathcal{L}(A)$. Hence, if $\mathbf{t} \neq \alpha\mathbf{d}$ for all $\alpha \in \mathbb{R}$, then $\mathcal{L}(W) \neq \mathcal{L}(A)$, since they differ in their first eigenspace. Otherwise we may choose a positive α and get that $t_i = \alpha \cdot d_i$ for all $i \in V$.

For every $i \neq j$ we get from $\ell_{ij} = w_{ij}/\sqrt{d_i d_j} = a_{ij}/\sqrt{t_i t_j} = \ell^A_{ij}$ that $a_{ij} = w_{ij} \cdot \sqrt{t_i t_j (d_i d_j)^{-1}} = \alpha \cdot w_{ij}$. For $i = j$ we get from $\ell_{ij} = (d_i - w_{ii})/d_i = (t_i - a_{ii})/t_i = \ell^A_{ij}$ that $d_i - w_{ii} = d_i - a_{ii}/\alpha$, hence $a_{ii} = \alpha \cdot w_{ii}$. □

Remark 3. With Fact B.3 this lemma also gives that $\mathcal{L}_\mathbf{f}(W) = \mathcal{L}_{\alpha\mathbf{f}}(W)$ for all $\alpha > 0$. Further, for connected $\mathcal{G}(W)$, even the opposite holds: $\mathcal{L}_\mathbf{f}(W) = \mathcal{L}_\mathbf{g}(W)$ if and only if $\mathbf{f} = \alpha\mathbf{g}$ for some $\alpha \in \mathbb{R}_{>0}$.

B.3. f-adjustments are diagonally modified Laplacians

The following two lemmas are stated in the paper. Their proofs are straightforward.

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

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

Proof. This can be seen from $\overline{W}_{\mathbf{f},1} = (\tilde{W}_\mathbf{f})^\circ_{c\mathbf{f}}$ and applying Lemma B.6 to $\tilde{W}_\mathbf{f}$, or directly as follows:

$$\begin{aligned} c \cdot \mathcal{L}(\overline{W}_{\mathbf{f},c}) &= c \cdot \sqrt{(cF)^{-1}}(cF - (\tilde{W}_\mathbf{f} - \tilde{D}_\mathbf{f} + cF))\sqrt{(cF)^{-1}} \\ &= \sqrt{F^{-1}}(F - (\tilde{W}_\mathbf{f} - \tilde{D}_\mathbf{f} + F))\sqrt{F^{-1}} \\ &= \mathcal{L}(\overline{W}_{\mathbf{f},1}) \end{aligned}$$

□

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

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

Proof. Right from the definitions we get that:

$$\begin{aligned}\mathcal{L}_{\mathbf{f}}(W) &= I - \sqrt{F^{-1}}(\tilde{W} - \tilde{D}_{\mathbf{f}} + F)\sqrt{F^{-1}} \\ &= I - \sqrt{F^{-1}}\sqrt{FD^{-1}}W\sqrt{FD^{-1}}\sqrt{F^{-1}} + \tilde{D}_{\mathbf{f}}F^{-1} - I \\ &= \tilde{D}_{\mathbf{f}}F^{-1} - \sqrt{D^{-1}}W\sqrt{D^{-1}}\end{aligned}$$

□

Remark 4. Recall that $\mathcal{L}(W) = I - \tilde{W}_{\mathbf{1}}$, hence

$$\mathcal{L}_{\mathbf{f}}(W) \;=\; \mathcal{L}(W) - \underbrace{(I - \tilde{D}_{\mathbf{f}}F^{-1})}_{\in\, \mathrm{diag}(\mathbb{R}^n)} \;=:\; \mathcal{L}(W) - \mathrm{diag}(\mathbf{h})$$

for $\mathbf{h} := (I - \sqrt{F^{-1}D^{-1}}W\sqrt{FD^{-1}})\mathbf{1}$, where $\mathbf{h} = (h_i)_i$ can be expressed element-wise as:

$$h_i = 1 - \sum_{j=1}^{n} w_{ij}\frac{\sqrt{f_j/f_i}}{\sqrt{d_i d_j}}.$$

Remark 5. The trivial relation $\mathcal{L}_{\mathbf{d}}(W) = \mathcal{L}(W)$ is attained for $\mathbf{h} = (I - D^{-1}W)\mathbf{1} = \mathbf{0}$, which is noteworthy, since $D^{-1}W$ is the random walk transition matrix.

B.4. Diagonally modified Laplacians are f-adjustments

We are now going to prove two lemmas that finally lead to our main result.

Lemma B.9 (Characterization of f-adjusted Laplacians). *For connected $\mathcal{G}(W) \in \mathcal{G}(\mathbb{W})$ let*

$$\Lambda := \{\mathcal{L}_{\mathbf{f}}(W) \mid \mathbf{f} \in \mathbb{R}^n_{>0}\}$$

denote the **Laplacian orbit of W under f-adjusting**. *Further let*

$$\Lambda' := \{Z - \tilde{W}_{\mathbf{1}} \mid Z \in \mathrm{diag}(\mathbb{R}^n_{>0}),\ \rho(Z^{-1}\tilde{W}_{\mathbf{1}}) = 1\}$$

for $\rho(\cdot)$ the Perron root of its argument. Then it holds that $\Lambda = \Lambda'$ with the relation $Z = \tilde{D}_{\mathbf{f}}F^{-1}$, wherein $\sqrt{\mathbf{f}}$ is the unique (up to scaling) right eigenvector of $Z^{-1}\tilde{W}_{\mathbf{1}}$ to eigenvalue 1.

Proof. It is well-known that $\mathcal{G}(W)$ is connected if and only if W is irreducible. This implies, for any choice of $Z \in \mathrm{diag}(\mathbb{R}^n_{>0})$, that $Z^{-1}\tilde{W}_{\mathbf{1}}$ is irreducible (and non-negative), too, since it has the same non-zero-pattern as W. This allows to apply various aspects of the Perron-Frobenius-Theorem (PFT), see for example Stańczak et al. (2006) for an overview.

$\Lambda \subseteq \Lambda'$: fix any $\mathcal{L}_{\mathbf{f}}(W) \in \Lambda$ for some $\mathbf{f} > 0$ and set $Z := \tilde{D}_{\mathbf{f}}F^{-1} \in \mathrm{diag}(\mathbb{R}^n_{>0})$. We prove that $\rho(Z^{-1}\tilde{W}_{\mathbf{1}}) = 1$ by finding an all-positive eigenvector $\mathbf{x}$ to the following eigenvalue problem:

$$Z^{-1}\sqrt{D^{-1}}W\sqrt{D^{-1}}\mathbf{x} = \mathbf{x}. \qquad (\star)$$

We propose that $\sqrt{\mathbf{f}}$ is such an eigenvector. Plugging $\tilde{D}_{\mathbf{f}} = \operatorname{diag}(\sqrt{FD^{-1}}W\sqrt{FD^{-1}}\mathbf{1})$ into $\tilde{D}_{\mathbf{f}}F^{-1} = Z$ element-wise, gives with $Z = \operatorname{diag}(z_1, \ldots, z_n)$ that

$$\begin{aligned}
& \tilde{D}_{\mathbf{f}}F^{-1} &=& \; Z \\
\Leftrightarrow \quad & \sum_{j=1}^{n} \frac{w_{ij}}{\sqrt{d_i d_j}} \cdot \sqrt{f_j} &=& \; z_i \cdot \sqrt{f_i} \quad \text{for all } i = 1, \ldots, n \\
\Leftrightarrow \quad & \sqrt{D^{-1}}W\sqrt{D^{-1}}\sqrt{\mathbf{f}} &=& \; Z\sqrt{\mathbf{f}} \\
\Leftrightarrow \quad & Z^{-1}\sqrt{D^{-1}}W\sqrt{D^{-1}}\sqrt{\mathbf{f}} &=& \; \sqrt{\mathbf{f}}
\end{aligned} \tag{$\star\star$}$$

Thus $\sqrt{\mathbf{f}}$ is indeed a solution to ($\star$), hence an all-positive (right) eigenvector of $Z^{-1}\tilde{W}_{\mathbf{1}}$ to the (existing) eigenvalue 1. By PFT, there is exactly one eigenvalue providing all-positive eigenvectors. Further it determines the spectral radius and is simple. Thus, we have that $\rho(Z^{-1}\tilde{W}_{\mathbf{1}}) = 1$ and that the corresponding left and right eigenvectors are unique (up to scaling). With Lemma 4.2 we get that $Z = \tilde{D}_{\mathbf{f}}F^{-1}$ is a valid choice for Z to represent $\mathcal{L}_{\mathbf{f}}(W)$ as an element in Λ'.

$\Lambda' \subseteq \Lambda$: fix any $Z - \tilde{W}_{\mathbf{1}} \in \Lambda'$ for some $Z \in \operatorname{diag}(\mathbb{R}^n_{>0})$ with $\rho(Z^{-1}\tilde{W}_{\mathbf{1}}) = 1$. By irreducibility of $Z^{-1}\tilde{W}_{\mathbf{1}}$, there exists by PFT a unique (up to scaling) all-positive solution $\hat{\mathbf{x}}$ of ($\star$). By defining $\mathbf{f} := \hat{\mathbf{x}}^2$ we get that $\sqrt{\mathbf{f}}$ is a solution to the eigenvalue problem ($\star$). From ($\star\star$) we see that this is equivalent to $Z = \tilde{D}_{\mathbf{f}}F^{-1}$. Thus, we get that $Z - \tilde{W}_{\mathbf{1}} = \tilde{D}_{\mathbf{f}}F^{-1} - \tilde{W}_{\mathbf{1}} = \mathcal{L}_{\mathbf{f}}(W) \in \Lambda$ for this unique (up to scaling) choice of $\mathbf{f}$. □

Remark 6. This lemma generalizes to unconnected graphs as follows: there is a permutation P of rows and columns such that $W' = PWP^T$ has block diagonal form, wherein each block-submatrix is irreducible. Thus, all above arguments can be carried out on each block individually, by restricting $\mathbf{f}$ to just the entries belonging to that block (= connected component). This relaxes the uniqueness in the way, that now $\mathbf{f}$ may be scaled by an individual scaling factor chosen independently for each connected component.

We have shown that $\mathbf{f}$-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 $A \in \mathbb{W}_{\ominus}$. So far the question is left open if further the converse is true: does *every* such diagonal modification imply that A is an $\mathbf{f}$-adjustment of W? We now answer this in the affirmative.

Lemma B.10 (Characterization of Diagonally Modified Laplacians). *Let $W \in \mathbb{W}$. Then $\mathcal{L}(W) + X = \mathcal{L}(A)$ holds true for $X \in \operatorname{diag}(\mathbb{R}^n)$ and $A \in \mathbb{W}_{\ominus}$ if and only if A is an $\mathbf{f}$-adjustment of W for some $\mathbf{f} \in \mathbb{R}^n_{>0}$. Formally,*

$$\Lambda = \{\mathcal{L}(W) + X \mid X \in \operatorname{diag}(\mathbb{R}^n),\ \mathcal{L}(W) + X \in \mathcal{L}(\mathbb{W}_{\ominus})\}$$

with Λ the Laplacian orbit of W under $\mathbf{f}$-adjusting.

Proof. "$\subseteq$". This direction is clear by Lemma 4.2.

"$\supseteq$". With $Y := X + I$ we have that $\mathcal{L}(A) = Y - \sqrt{D^{-1}}W\sqrt{D^{-1}} =: [\ell^A_{ij}]$ for some $A \in \mathbb{W}_{\ominus}$. First we want to show that this implies that $Y = \operatorname{diag}(y_1, \ldots, y_n)$ has an all-positive diagonal. Fix any $i \in V$. From $A \in \mathbb{W}_{\ominus}$ and Fact B.5 it follows that $\ell^A_{ii} \geq 0$, and from $W \in \mathbb{W}$ that $w_{ii} \geq 0$. Thus, we get from $\ell^A_{ii} = y_i - w_{ii}/d_i \geq 0$ that $y_i \geq w_{ii}/d_i \geq 0$. Thus y_i is non-negative. Now assume that $y_i = 0$. This implies that $w_{ii} = 0$, hence $\ell^A_{ii} = 0$, so i must be an isolated vertex in $\mathcal{G}(A)$. However, in $\mathcal{G}(W)$ vertex i cannot be isolated, since $w_{ii} = 0$ implies by the positive degree constraint that $w_{ij} > 0$ for some $j \neq i$. For such j it holds that $-a_{ij}/\sqrt{d^A_i d^A_j} = \ell^A_{ij} = \ell_{ij} = -w_{ij}/\sqrt{d_i d_j} < 0$, hence that $a_{ij} > 0$ in contradiction to i being an isolated vertex in A. Therefore, $y_i > 0$ for all $i \in V$.

In the following we assume w.l.o.g. that W is connected, since all arguments can be applied to each connected component individually, independent of all other components.

Lemma B.9 gives that for any all-positive diagonal matrix Y (in particular as chosen above) there exists some $\alpha > 0$ and some $\mathbf{f}$-adjustment $\overline{W}_{\mathbf{f}} =: M_\alpha \in \mathbb{W}_{\ominus}$ for some $\mathbf{f} \in \mathbb{R}^n_{>0}$ such that $\mathcal{L}(M_\alpha) = \alpha Y - \sqrt{D^{-1}}W\sqrt{D^{-1}} \in \Lambda$. We now show that for no $\beta \neq \alpha$ any $M_\beta \in \mathbb{W}_\ominus$ with $\mathcal{L}(M_\beta) = \beta Y - \sqrt{D^{-1}} W \sqrt{D^{-1}}$ exists. This finally implies that M_1 itself is the $\mathbf{f}$-adjustment of W, which gives that $\mathcal{L}(A) = \mathcal{L}(M_1) \in \Lambda$.

Therefore, fix any $M_\beta \in \mathbb{W}_\ominus$ with $\mathcal{L}(M_\beta) = \beta Y - \sqrt{D^{-1}} W \sqrt{D^{-1}}$ for some $\beta \in \mathbb{R}$. Setting $\epsilon := \beta - \alpha$ gives that $\mathcal{L}(M_\beta) = \beta Y - \sqrt{D^{-1}} W \sqrt{D^{-1}} = \mathcal{L}(M_\alpha) + \epsilon Y$. Now assume that $\epsilon > 0$. By Lemma B.7, $\mathcal{L}(M_\beta)$ is positive semi-definite with $v := \sqrt{M_\beta \mathbf{1}}$ an all-positive eigenvector to the eigenvalue 0. We get the contradiction $0 = v^T \mathcal{L}(M_\beta) v = v^T \mathcal{L}(M_\alpha) v + \epsilon v^T Y v > 0$, because $v^T \mathcal{L}(M_\alpha) v \geq 0$ by positive semi-definiteness of $\mathcal{L}(M_\alpha)$, and $\epsilon v^T Y v > 0$ by all-positivity of v and Y. Now assume that $\epsilon < 0$. Let $w := \sqrt{M_\alpha \mathbf{1}}$ denote the all-positive eigenvector of $\mathcal{L}(M_\alpha)$ to the eigenvalue 0. We get the contradiction $0 \leq w^T \mathcal{L}(M_\beta) w = w^T \mathcal{L}(M_\alpha) w + \epsilon w^T Y w < 0$, because $w^T \mathcal{L}(M_\alpha) w = 0$, and $\epsilon w^T Y w < 0$ by all-positivity of w and Y, and the first inequality due to positive semi-definiteness of $\mathcal{L}(M_\beta)$. Thus, for no $\epsilon \neq 0$ any graph matrix of this form exists. □

Now we have all ingredients to prove our main result on the algebraic interpretation of $\mathbf{f}$-adjusting:

Theorem B.11 (Complete Characterization for Connected Graphs). *For $n > 1$ and connected $\mathcal{G}(W) \in \mathcal{G}(\mathbb{W})$ consider all solutions $(X, A, c) \in \operatorname{diag}(\mathbb{R}^n) \times \mathbb{W}_\ominus \times \mathbb{R}$ 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 $\mathbf{f} \in \mathbb{R}^n_{>0}$. This is equivalent to choosing any $Z \in \operatorname{diag}(\mathbb{R}^n_{>0})$ with Perron root $\rho(Z^{-1}\tilde{W}_\mathbf{1}) = 1$, which determines $\sqrt{\mathbf{f}}$ as the unique (up to scaling) right Perron eigenvector.

Proof. Let $\mathcal{L}(W) =: [\ell_{ij}]$ and $\mathcal{L}(A) =: [\ell^A_{ij}]$. Since W is non-trivial, there exist $i \neq j$ with $w_{ij} > 0$, hence $\ell_{ij} < 0$. The case $c < 0$ would imply that $\ell^A_{ij} > 0$, which is impossible for all $A \in \mathbb{W}_\ominus$. The case $c = 0$ would imply that $\mathcal{L}(W) + X = 0$, hence that all off-diagonal elements in W are zero, in contradiction to being non-trivial. Thus, no solutions for $c \leq 0$ exist.

Now consider the case $c = 1$, that is any solution of the form $\mathcal{L}(W) + X = \mathcal{L}(A)$. We get from Lemma B.10 that every such solution corresponds to fixing some $\mathbf{f} \in \mathbb{R}^n_{>0}$ and setting $A := \overline{W}_{\mathbf{f},1}$. This implies by Lemma 4.2 that $X + I = \tilde{D}_\mathbf{f} F^{-1}$. It remains to show that this is equivalent to choosing $Z \in \operatorname{diag}(\mathbb{R}^n_{>0})$ with the desired properties. For fixed $\mathbf{f}$, we get from Lemma B.9 that $\tilde{D}_\mathbf{f} F^{-1} = X + I = Z$ for some $Z \in \operatorname{diag}(\mathbb{R}^n_{>0})$ with Perron eigenvalue $\rho(Z^{-1}\tilde{W}_\mathbf{1}) = 1$ and $\sqrt{\mathbf{f}}$ the corresponding right Perron eigenvector. The other way round, Lemma B.9 gives that choosing any $Z \in \operatorname{diag}(\mathbb{R}^n_{>0})$ with $\rho(Z^{-1}\tilde{W}_\mathbf{1}) = 1$ implies by setting $X := Z - I$ that $\mathcal{L}(W) + X = \mathcal{L}(A)$ for $A = \overline{W}_{\mathbf{f},1}$, and further that $Z = \tilde{D}_\mathbf{f} F^{-1}$ with $\sqrt{\mathbf{f}}$ being determined as the unique (up to scaling) right Perron eigenvector.

Now consider the case $c > 0$, that is $\mathcal{L}(W) + X = c \cdot \mathcal{L}(A)$ for $X \in \operatorname{diag}(\mathbb{R}^n)$ and $A \in \mathbb{W}_\ominus$. From $c \cdot \mathcal{L}(A) \overset{B.6}{=} \mathcal{L}(A^\circ_{c^{-1}A\mathbf{1}})$ we get that $\mathcal{L}(W) + X$ equals the normalized Laplacian of a weak graph matrix. Thus, $\mathcal{L}(W) + X \overset{B.10}{=} \mathcal{L}(\overline{W}_{\mathbf{g},1})$ for some $\mathbf{g} \in \mathbb{R}^n_{>0}$. This gives that $\mathcal{L}(A) = c^{-1}\mathcal{L}(\overline{W}_{\mathbf{g},1}) \overset{4.1}{=} \mathcal{L}(\overline{W}_{\mathbf{g},c})$. Thus we have that $A \overset{B.8}{=} \alpha \cdot \overline{W}_{\mathbf{g},c} \overset{B.3}{=} \overline{W}_{\alpha\mathbf{g},c}$ for some $\alpha > 0$. So A is the $(\mathbf{f}, c)$-adjustment of W for $\mathbf{f} := \alpha\mathbf{g}$. It follows as before that $X + I = \tilde{D}_\mathbf{g} \operatorname{diag}(\mathbf{g})^{-1} = \tilde{D}_\mathbf{f} F^{-1} = Z$ for some $Z \in \operatorname{diag}(\mathbb{R}^n_{>0})$ with Perron eigenvalue $\rho(Z^{-1}\tilde{W}_\mathbf{1}) = 1$ and $\sqrt{\mathbf{g}}$ a corresponding right Perron eigenvector as well as $\sqrt{\mathbf{f}} = \sqrt{\alpha\mathbf{g}}$ another one, unique up to scaling.

For the other way round, choose any $\mathbf{f} \in \mathbb{R}^n_{>0}$ (or equivalently any $Z \in \operatorname{diag}(\mathbb{R}^n_{>0})$). Setting $A := \overline{W}_{\mathbf{f},c}$ implies that $\mathcal{L}(\overline{W}_{\mathbf{f},1}) \overset{4.1}{=} c \cdot \mathcal{L}(A) = \mathcal{L}(W) + X$, hence by Lemma 4.2 that $X + I = \tilde{D}_\mathbf{f} F^{-1}$. □

Remark 7. For arbitrary $c > 0$, the theorem implies that $\mathcal{L}(W) + X = c \cdot \mathcal{L}(A)$ is a solution if and only if $c \cdot \mathcal{L}(A) = c \cdot \mathcal{L}(\overline{W}_{\mathbf{f},c}) = \mathcal{L}(\overline{W}_{\mathbf{f},1}) = \mathcal{L}_\mathbf{f}(W)$ for the specific vector $\mathbf{f}$, that is if and only if $\mathcal{L}(W) + X \in \mathcal{L}(\mathbb{W}_\ominus)$.

Remark 8. From the special case $c = 1$ we get that all solutions $(X, A) \in \operatorname{diag}(\mathbb{R}^n) \times \mathbb{W}_{\ominus}$ of the equation $\mathcal{L}(W) + X = \mathcal{L}(A)$ are given by $\mathcal{L}(W) + X = \mathcal{L}_{\mathbf{f}}(W)$ for every $\mathbf{f} \in \mathbb{R}^n_{>0}$. Furthermore, $\mathbf{f}$ and X are related to each other by $\rho((X + I)^{-1}\tilde{W}_{\mathbf{1}}) = 1$ with $X + I \in \operatorname{diag}(\mathbb{R}^n_{>0})$ and $\sqrt{\mathbf{f}}$ the corresponding eigenvector.

In our paper we state Theorem B.11 slightly different as follows:

Theorem 4.3 (Complete Characterization). *For any* $W \in \mathbb{W}$ *with* $\mathcal{L}(W) \neq 0$ *consider all solutions* $(X, A, c) \in \operatorname{diag}(\mathbb{R}^n) \times \mathbb{W}_{\ominus} \times \mathbb{R}$ *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* $Z \in \operatorname{diag}(\mathbb{R}^n_{>0})$ *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* $Z^{-1}\tilde{W}_{\mathbf{1}}$*.*

Remark 9. All diagonal modifications of W with $W + X \in \mathbb{W}_{\ominus}$ obviously correspond to $\mathbf{f}$-selflooping, that is modifying selfloops in any possible way. Theorem 4.3 shows a similar result for the normalized Laplacian, namely that all diagonal modifications of $\mathcal{L}(W)$ with $\mathcal{L}(W) + X \in \mathcal{L}(\mathbb{W}_{\ominus})$ correspond to $\mathbf{f}$-adjusting for every $\mathbf{f} \in \mathbb{R}^n_{>0}$. This shows that $\mathbf{f}$-adjusting is a very natural graph modification.

C. Geometric Interpretation

In this section we provide more details on the convergence of the modified interspace quantities to their corresponding modified continuous quantities. Both our propositions are already motivated in the paper in terms of the sum expressions that define the interspace volumes and cut weights, respectively.

The proofs heavily rely on the arguments in Maier et al. (2009), where the convergence of interspace volumes and cut weights to their corresponding continuous quantities is proven for different types of neighborhood graphs, for a fixed density $p : \mathcal{X} \to \mathbb{R}_{>0}$. They show that volumes and cut weights refer to the continuous quantities $\int p^2$ in the limit. They study *different types* of neighborhood graphs for a *single density* p. In contrast to that, we relate two neighborhood graphs of the *same type*, but according to *different densities* p and $\tilde{p} = fp$ to each other.

The limit quantities $\int p^2$ combine two different effects, each rising one factor of p. These two effects are:

(i) one factor p comes from the sampling mechanism, which distributes the sample points according to p

(ii) another factor p refers to the weighted degrees in the original graph, which serve as a density estimate on p

This gets particularly apparent in the interspace view. In light of this we ask the question: "*Can we modify these effects by modifying edge weights in the given graph ?*"

The first effect (i) cannot be modified by us, since the unknown sample is fixed. Therefore, the positions of the sample points are always determined by a sampling mechanism according to p. However, we observe that the second effect (ii) is *not* required to refer to degrees that estimate p. It can be chosen to represent any other function f on the underlying space (as long as it satisfies the same technical assumptions that are made on p). Since we have access to the graph, we can change its degrees by changing its edge weights plus adding selfloops in any way that we want. This gives us the opportunity to influence the second effect freely.

There is no geometric interpretation of volumes and cuts under *arbitrary* modifications applied to the graph. But if we modify the degrees and the cut weights by $\mathbf{f}$-adjusting, and if we further assume that $\mathbf{f}$ is determined by any suitable continuous function f, then we can interpret the resulting volumes and cuts in $\overline{G}$ in terms of $\int fp$, compared to their original interpretation as $\int p^2$ in G.

Technically, we achieve this by sneaking in a new term corresponding to f everywhere along the proofs in Maier et al. (2009). This allows for "changing" the limit quantities $\int p^2$ into any $\int \overline{p} = \int fp$, for free to choose f. The full proofs would cover at least 10 pages, and mainly deal with technical considerations on boundary effects. For that reason, we decided to just sketch the general proof strategies here. The interested reader is referred to Maier et al. (2009).

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 all 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.*

Proof Sketch. The general line of argument is to decompose the deviations in bias and variance term. The convergence of the bias term is straightforward to see, the convergence of the variance term can be proved by concentration arguments. □

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.*

Proof Sketch. The variance part can be solved by concentration arguments. For the bias term, we need to count the edges in the graph that cross the hyperplane H. The main ingredient therefore is that we can control the distance of connected points in the graph, with high probability. By concentration arguments we know that with high probability, each point x is connected to all points within a certain distance r_x, and is not connected to points exceeding a certain distance R_x. For counting edges, we then need to compute the probability mass of the intersection of those balls with the given hyperplane H. □

Remark 10. We believe that all results generalize from hyperplanes to any other cut surfaces that are sufficiently regular.

Remark 11. A nice consequence of our modification is that we can *implicitly* define f *relatively* to p, just by defining $\mathbf{f}$ as a modification applied to $\mathbf{d}$ in the graph. For example, and as argued below, $\mathbf{f} := \mathbf{d}^r$ refers implicitly to defining the underlying function $f := \beta \cdot p^{r+1}$ for some global scaling factor β that only depends on r, the sample size n and the intrinsic dimension d. This approach is only limited by the fact that the degrees $\mathbf{d}$ do not estimate p exactly, since they provide estimates that are *proportional* to p. Precisely, $d_i = \alpha_{n,d} \cdot p(x_i)$ for some global scaling factor $\alpha_{n,d}$ that depends on n and d. Usually n is known, but as long as we do not know d, we have to make sure that the unknown estimation factor $\alpha_{n,d}$ passes our modification just as another global scaling factor, without introducing some distortion. For example, in the case $\mathbf{f} = \mathbf{d}^r$ we get that the implicitly defined function f equals $f = \alpha_{n,d}^{r+1} \cdot p^{r+1} = const \cdot p^{r+1}$ as intended. However, if we define $\mathbf{f} := a\mathbf{d}^2 + b\mathbf{d} + c\mathbf{1}$ for $a, b, c \in \mathbb{R}_{>0}$, then the implicitly defined function f would *not* refer to $f = ap^2 + bp + c$, but to $f = a\alpha_{n,d}^2 p^2 + b\alpha_{n,d} p + c$ instead. Hence the unknown estimation factor $\alpha_{n,d}$ introduces a distortion that implies that f cannot be represented as $f = \gamma(ap^2 + bp + c)$ for any scaling factor γ.

Note that $\mathbf{f} = \mathbf{d}^r$ is not the only valid choice of implicitly defining f in terms of p. See for example Section A.1.

Moreover, if d is given as prior information, or from some estimate on the intrinsic dimension, then we can determine $\alpha_{n,d}$ and use it to define f implicitly from p in much more various ways. For example, if we want to study $ap^2 + bp + c$, then defining the new degrees by $\mathbf{f} := a\mathbf{d}^2 + b\alpha_{n,d}\mathbf{d} + c\alpha_{n,d}^2\mathbf{1}$ would indeed represent $f = a\alpha_{n,d}^2 p^2 + b\alpha_{n,d}^2 p + \alpha_{n,d}^2 c = const \cdot (ap^2 + bp + c)$ as intended. Hence any good estimate on the intrinsic dimension allows for more powerful ways of how to define f from p implicitly, in order to study volumes and cuts according to $\int fp$.

References

Kannan, R., Lovász, L., and Montenegro, R. Blocking Conductance and Mixing in Random Walks. *Combinatorics, Probability and Computing*, 15(4):541–570, 2006.

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

Montenegro, R. Faster Mixing for Conductance Methods. Technical report, 2003.

Nadler, B., Srebro, N., and Zhou, X. Statistical analysis of semi-supervised learning: The limit of infinite unlabelled data. In *Neural Information Processing Systems (NIPS)*, 2009.

Stańczak, S., Wiczanowski, M., and Boche, H. Chapter 1: On the Perron Root of Irreducible Matrices. In *Resource Allocation in Wireless Networks*, volume 4000 of *Lecture Notes in Computer Science*. 2006.

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