Finding Cheeger Cuts in Hypergraphs via Heat Equation

Masahiro Ikeda [1], Atsushi Miyauchi [2], Yuuki Takai [3], and Yuichi Yoshida[4]

Abstract

Cheeger's inequality states that a tightly connected subset can be extracted from a graph G using an eigenvector of the normalized Laplacian associated with G. More specifically, we can compute a subset with conductance $O(\sqrt{\phi_G})$, where ϕ_G is the minimum conductance of a set in G.

It has recently been shown that Cheeger's inequality can be extended to hypergraphs. However, as the normalized Laplacian of a hypergraph is no longer a matrix, we can only approximate to its eigenvectors; this causes a loss in the conductance of the obtained subset. To address this problem, we here consider the heat equation on hypergraphs, which is a differential equation exploiting the normalized Laplacian. We show that the heat equation has a unique solution and that we can extract a subset with conductance $\sqrt{\phi_G}$ from the solution. An analogous result also holds for directed graphs.

1 Introduction

The goal of spectral clustering of graphs is to extract tightly connected communities from a given weighted graph $G = (V, E, w)$, where $w\colon E \to \mathbb{R}_+$ is a weight function, using eigenvectors of matrices associated with G. One of the most fundamental results in this area is Cheeger's inequality, which relates the second-smallest eigenvalue of the normalized Laplacian of G and the conductance of G. Here, the *(random-walk) normalized Laplacian* of G is defined as $\mathcal{L}_G = I - A_G D_G^{-1}$, where $A_G \in \mathbb{R}^{V\times V}$ and $D_G \in \mathbb{R}^{V\times V}$ are the (weighted) adjacency matrix and the (weighted) degree matrix, respectively, of G. Further, D_G is a diagonal matrix with the (v, v)-th element for $v \in V$ being the (weighted) degree $d_G(v) := \sum_{e\in E|v\in e} w(e)$ of v. Note that all eigenvalues of $\mathcal{L}_G$ are non-negative and the smallest eigenvalue is always zero, as $\mathcal{L}_G D_G \mathbf{1} = \mathbf{0}$. The *conductance* of a set $\emptyset \subsetneq S \subsetneq V$ is defined as

$$\phi_G(S) := \frac{\sum_{e\in\partial_G(S)} w(e)}{\min\{\mathrm{vol}_G(S), \mathrm{vol}_G(V \setminus S)\}},$$

where $\partial_G(S)$ is the set of edges between S and $V \setminus S$, and $\mathrm{vol}_G(S) := \sum_{v\in S} d_G(v)$ is the *volume* of S. Intuitively, smaller $\phi_G(S)$ corresponds to more tightly connected S. The *conductance* of G is the minimum conductance of a set in G; that is, $\phi_G := \min_{\emptyset\subsetneq S\subsetneq V} \phi_G(S)$. Then, Cheeger's inequality [2, 3] states that

$$\frac{\lambda_G}{2} \leq \phi_G \leq \sqrt{2\lambda_G}, \tag{1}$$

where $\lambda_G \in \mathbb{R}_+$ is the second-smallest eigenvalue of $\mathcal{L}_G$. The second inequality of (1) is algorithmic in the sense that we can compute a set $\emptyset \subsetneq S \subsetneq V$ with conductance of at most $\sqrt{2\lambda_G} = O(\sqrt{\phi_G})$, which is called a *Cheeger cut*, in polynomial time from an eigenvector corresponding to λ_G. Moreover, Cheeger's inequality is tight in the sense that computing a set with conductance $o(\sqrt{\phi})$ is NP-hard [13], assuming the small set expansion hypothesis (SSEH) [12].

[1] Center for Advanced Intelligence Project, RIKEN, Tokyo, Japan/Department of Mathematics, Faculty of Science and Technology, Keio University 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522, Japan, E-mail: `masahiro.ikeda@riken.jp`

[2] Center for Advanced Intelligence Project, RIKEN, Tokyo, Japan, E-mail: `atsushi.miyauchi.hv@riken.jp`

[3] Center for Advanced Intelligence Project, RIKEN, Tokyo, Japan/Department of Mathematics, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522, Japan, E-mail: `yuuki.takai@riken.jp`

[4] National Institute of Informatics, Tokyo, Japan, E-mail: `yyoshida@nii.ac.jp`

Several attempts to extend Cheeger's inequality to hypergraphs have been made. To explain the known results, we first extend the concepts of conductance and the normalized Laplacian to hypergraphs. Let $G = (V, E, w)$ be a weighted hypergraph, where $w \colon E \to \mathbb{R}_+$ is a weight function. The *(weighted) degree* of a vertex $v \in V$ is $d_G(v) := \sum_{e \in E \mid v \in e} w(e)$. For a vertex set $\emptyset \subsetneq S \subsetneq V$, the *conductance* of S is

$$\phi_G(S) := \frac{\sum_{e \in \partial_G(S)} w(e)}{\min\{\mathrm{vol}_G(S), \mathrm{vol}_G(V \setminus S)\}},$$

where $\partial_G(S)$ is the set of hyperedges intersecting both S and $V \setminus S$, and $\mathrm{vol}_G(S)$ has the same definition as previously. The *conductance* of G is $\phi_G := \min_{\emptyset \subsetneq S \subsetneq V} \phi_G(S)$.

The *normalized Laplacian* $\mathcal{L}_G \colon \mathbb{R}^V \to 2^{\mathbb{R}^V}$ of a hypergraph G [4, 17] is multi-valued and no longer linear (see Section 2 for a detailed definition). In the simplest setting that the hypergraph G is unweighted and d-regular, that is, every vertex has degree d, and the elements of the given vector $\boldsymbol{x} \in \mathbb{R}^V$ are pairwise distinct, the $\mathcal{L}_G$ acts as follows: We create an undirected graph $G_{\boldsymbol{x}}$ on V from G by adding for each hyperedge $e \in E$ an undirected edge uv, where $u = \mathrm{argmin}_{w \in e} \boldsymbol{x}(w)$ and $v = \mathrm{argmax}_{w \in e} \boldsymbol{x}(w)$, then return $\mathcal{L}_{G_{\boldsymbol{x}}} \boldsymbol{x}$.

When $\mathcal{L}_G(\boldsymbol{v}) \ni \lambda \boldsymbol{v}$ holds for $\lambda \in \mathbb{R}$ and $\boldsymbol{v} \neq \boldsymbol{0}$, we can state that λ and $\boldsymbol{v}$ are an *eigenvalue* and an *eigenvector*, respectively, of $\mathcal{L}_G$. As with the graph case, all eigenvalues of $\mathcal{L}_G$ are non-negative and the first eigenvalue is zero as $\mathcal{L}_G(D\mathbf{1}) = \boldsymbol{0}$ holds. Moreover, the second-smallest eigenvalue $\lambda_G \in \mathbb{R}_+$ exists. Cheeger's inequality for hypergraphs [4, 17] states that

$$\frac{\lambda_G}{2} \leq \phi_G \leq 2\sqrt{\lambda_G}. \tag{2}$$

Again, the second inequality is algorithmic. If we can compute an eigenvector corresponding to λ_G, we can obtain a Cheeger cut; that is, a set $\emptyset \subsetneq S \subsetneq V$ with $\phi_G(S) = O(\sqrt{\phi_G})$, in polynomial time. Unlike the undirected graph case, however, only an $O(\log n)$-approximation algorithm is available for computing λ_G [17]. Further, this approximation ratio is tight under the SSEH [4]. Hence, the following question arises naturally: Can we compute a Cheeger cut without computing λ_G and applying Cheeger's inequality on the corresponding eigenvector?

To answer this question, we consider the following differential equation called the *heat equation* [4]:

$$\frac{d\boldsymbol{\rho}_t}{dt} \in -\mathcal{L}_G(\boldsymbol{\rho}_t) \quad \text{and} \quad \boldsymbol{\rho}_0 = \boldsymbol{s}, \tag{HE; $\boldsymbol{s}$}$$

where $\boldsymbol{s} \in \mathbb{R}^V$ is an initial vector. Intuitively, we gradually diffuse values (or *heat*) on vertices along hyperedges so that the maximum and minimum values in each hyperedge become closer. We can show that (HE; $\boldsymbol{s}$) always has a (unique) solution for $t \geq 0$[5] using the theory of monotone operators and evolution equations [11] (see Section 4 for details), and let $\boldsymbol{\rho}_t^{\boldsymbol{s}} \in \mathbb{R}^V$ be the vector at time $t \geq 0$. In particular, $\boldsymbol{\rho}_0^{\boldsymbol{s}} = \boldsymbol{s}$ holds. In addition, if $\sum_{v \in V} \boldsymbol{s}(v) = 1$, we can show that $\sum_{v \in V} \boldsymbol{\rho}_t^{\boldsymbol{s}}(v) = 1$ holds for any $t \geq 0$, and that $\boldsymbol{\rho}_t^{\boldsymbol{s}}$ converges to $\boldsymbol{\pi} \in \mathbb{R}^V$ when G is connected, where $\boldsymbol{\pi}(v) = d_G(v)/\mathrm{vol}(V)$ (see [4, Theorem 3.4]).

For a vector $\boldsymbol{x} \in \mathbb{R}^V$, let $\mathrm{sweep}(\boldsymbol{x})$ denote the set of all *sweep sets* with respect to $\boldsymbol{x}$; that is, sets of the form either $\{v \in V \mid \boldsymbol{x}(v) \geq \tau\}$ or $\{v \in V \mid \boldsymbol{x}(v) \leq \tau\}$, for some $\tau \in \mathbb{R}$. The following theorem can now be presented.

[5] Previous works [4, 17] only guaranteed that it has a solution for $0 \leq t \leq T$ for some $T \geq 0$.

Theorem 1. *Let $G = (V, E)$ be a hypergraph and $\emptyset \subsetneq S \subsetneq V$ be a set. Then, we have*

$$\phi_G(S) \geq \frac{\widetilde{\kappa}_t^2}{2} - \frac{2}{t} \log \frac{\mathrm{vol}(S)}{1 - \mathrm{vol}(S)/\mathrm{vol}(V)},$$

where $\widetilde{\kappa}_t = \min\{\phi_G(S) \mid v \in S, \xi \in [0, t], S \in \mathrm{sweep}(\boldsymbol{\rho}_\xi^{\boldsymbol{\pi}_v})\}$ and $\boldsymbol{\pi}_v \in \mathbb{R}^V$ is a vector for which $\boldsymbol{\pi}_v(v) = 1$ and $\boldsymbol{\pi}_v(u) = 0$ for $u \neq v$.

The proof of this theorem is given in Section 3. Theorem 1 means that, when t is sufficiently large, we can obtain a set $\emptyset \subsetneq S \subsetneq V$ such that $\phi_G(S) = O(\sqrt{\phi_G})$, thereby avoiding the problem of computing λ_G. Although we cannot solve the differential equation (HE; $\boldsymbol{s}$) exactly in polynomial time, we can efficiently simulate it by discretizing time using, e.g., the Euler method or the Runge-Kutta method. Indeed these methods have already been used in practice [16]. Alternatively, we can use difference approximation, developed in the theory of monotone operators and evolution equations [11], to obtain the following:

Theorem 2. *Let $G = (V, E, w)$ be a hypergraph and $v \in V$, and let $T \geq 1$ and $\lambda \in (0, 1)$. Then, we can compute (a concise representation) of a solution $\{\boldsymbol{\rho}_t^\lambda\}_{0 \leq t \leq T}$ such that $\|\boldsymbol{\rho}_t^{\boldsymbol{\pi}_v} - \boldsymbol{\rho}_t^\lambda\|_{D^{-1}} = O(\sqrt{\lambda T})$ for every $0 \leq t \leq T$, where $\|\boldsymbol{x}\|_{D^{-1}}^2 = \boldsymbol{x}^\top D^{-1} \boldsymbol{x}$, in time polynomial in $1/\lambda$, T, and $\sum_{e \in E} |e|$.*

1.1 Directed graphs

We briefly discuss directed graphs here, as we can show an analogue of Theorem 1 for such graphs with almost the same proof.

For a directed graph $G = (V, E, w)$, the *degree* of a vertex $v \in V$ is $d_G(v) = \sum_{e \in E | v \in e} w(e)$ and the *volume* of a set $S \subseteq V$ is $\mathrm{vol}_G(S) = \sum_{v \in S} d_G(v)$. Note that we do not distinguish out-going and in-coming edges when calculating degrees. Then, the *conductance* of a set $\emptyset \subsetneq S \subsetneq V$ is defined as

$$\phi_G(S) := \frac{\min\{\sum_{e \in \partial_G^+(S)} w(e), \sum_{e \in \partial_G^-(S)} w(e)\}}{\min\{\mathrm{vol}_G(S), \mathrm{vol}_G(V \setminus S)\}},$$

where $\partial_G^+(S)$ and $\partial_G^-(S)$ are the sets of edges leaving and entering S, respectively. Then, the *conductance* of G is $\phi_G := \min_{\emptyset \subsetneq S \subsetneq V} \phi_G(S)$. Note that $\phi_G = 0$ when G is a directed acyclic graph.

Yoshida [16] introduced the notion of a Laplacian for directed graphs and derived Cheeger's inequality for such graphs, which relates ϕ_G and the second-smallest eigenvalue λ_G of the normalized Laplacian of G. As with the hypergraph case, computing λ_G is problematic, and we can apply an analogue of Theorem 1 to obtain a set of small conductance without computing λ_G. In this paper, we focus on hypergraphs for simplicity of exposition.

1.2 Proof sketch

Previously, Chung [5] presented an analogue of Theorem 1 for graphs. Here, we review the proof of that analogue, because the proof of Theorem 1 presented in Section 3 is based on that proof.

For the graph case, we consider the following differential equation:

$$\frac{d\boldsymbol{\rho}_t}{dt} = -\mathcal{L}_G \boldsymbol{\rho}_t \quad \text{and} \quad \boldsymbol{\rho}_0 = \boldsymbol{s}.$$

This differential equation has a unique solution $\boldsymbol{\rho}_t^{\boldsymbol{s}} = \exp(-t\mathcal{L}_G)\boldsymbol{s}$. We define a function $f^{\boldsymbol{s}} \colon \mathbb{R}_+ \to \mathbb{R}$ as

$$f^{\boldsymbol{s}}(t) = \|\boldsymbol{\rho}_{t/2}^{\boldsymbol{s}} - \boldsymbol{\pi}\|_{D^{-1}}^2.$$

When G is connected, $\boldsymbol{\rho}_t^{\boldsymbol{s}}$ converges to $\boldsymbol{\pi}$ irrespective of $\boldsymbol{s}$; hence, $f^{\boldsymbol{s}}$ measures the difference between $\boldsymbol{\rho}_{t/2}^{\boldsymbol{s}}$ and its unique stationary distribution $\boldsymbol{\pi}$. For a set $S \subseteq V$, we define $\boldsymbol{\pi}_S \in \mathbb{R}^V$ as $\boldsymbol{\pi}_S(v) = d(v)/\mathrm{vol}(S)$ if $v \in S$ and $\boldsymbol{\pi}_S(v) = 0$ otherwise. Then, we can show that

$$\exp(-O(\phi(S)t)) \leq f^{\boldsymbol{\pi}_S}(t) \leq \exp\left(-\Omega\left((\kappa_t^{\boldsymbol{\pi}_S})^2 t\right)\right), \tag{3}$$

for every $S \subseteq V$, where $\kappa_t^{\boldsymbol{\pi}_S}$ is the minimum conductance of a sweep set with respect to the vector $(\boldsymbol{\rho}_t^{\boldsymbol{\pi}_S}(v)/d(v))_{v \in V}$. From the closed solution of $\boldsymbol{\rho}_t^{\boldsymbol{s}}$, we observe that $\boldsymbol{\rho}_{t/2}^{\boldsymbol{\pi}_S} = \sum_{v \in S} \frac{d(v)}{\mathrm{vol}(S)} \boldsymbol{\rho}_{t/2}^{\boldsymbol{\pi}_v}$. Then, we have

$$\exp(-O(\phi(S)t)) \leq f^{\boldsymbol{\pi}_S}(t) = \|\boldsymbol{\rho}_{t/2}^{\boldsymbol{\pi}_S} - \boldsymbol{\pi}\|_{D^{-1}}^2 \leq \left(\sum_{v \in S} \frac{d(v)}{\mathrm{vol}(S)} \|\boldsymbol{\rho}_{t/2}^{\boldsymbol{\pi}_v} - \boldsymbol{\pi}\|_{D^{-1}}\right)^2 \quad \text{(by triangle inequality)}$$

$$\leq \max_{v \in S} \|\boldsymbol{\rho}_{t/2}^{\boldsymbol{\pi}_v} - \boldsymbol{\pi}\|_{D^{-1}}^2 = \max_{v \in S} f^{\boldsymbol{\pi}_v}(t) \leq \max_{v \in S} \exp\left(-\Omega\left((\kappa_t^{\boldsymbol{\pi}_v})^2 t\right)\right).$$

Taking the logarithm yields the desired result.

The main obstacle to extending the above argument to hypergraphs is that $\boldsymbol{\rho}_t$ does not have a closed-form solution as $\mathcal{L}_G$ is no longer a linear operator. To overcome this obstacle, we observe that the sequence $t_0 = 0 < t_1 < t_2 < \cdots$ with $\lim_{i \to \infty} t_i = \infty$ exists, such that $\mathcal{L}_G$ acts as a linear operator $\mathcal{L}_i$ in each interval $[t_i, t_{i+1})$. Here, $\mathcal{L}_i$ is the normalized Laplacian of a graph constructed from the hypergraph G and the vector $\boldsymbol{\rho}_{t_i}$. Then, we can show a counterpart of (3) for each $f_i^{\boldsymbol{s}} \colon \mathbb{R}_+ \to \mathbb{R}$ defined as $f_i^{\boldsymbol{s}}(\Delta) = \|\boldsymbol{\rho}_{t_i + \Delta/2}^{\boldsymbol{s}} - \boldsymbol{\pi}\|$, which is sufficient for our analysis.

Another obstacle is that the triangle inequality applied in the above argument is not a priori true, because $\boldsymbol{\rho}_{t/2}^{\boldsymbol{\pi}_S}$ may not generally be equivalent to $\sum_{v \in S} \frac{d(v)}{\mathrm{vol}(S)} \boldsymbol{\rho}_{t/2}^{\boldsymbol{\pi}_v}$ for the hypergraph case. To derive the triangle inequality, we exploit the theory of maximal monotone operators and evolution equations [11] and borrow the concept of difference approximation of the solution.

1.3 Related work

As noted above, an analogue of Theorem 1 for graphs has been presented by Chung [5]. However, as the normalized Laplacian $\mathcal{L}_G = I - A_G D_G^{-1}$ is a matrix for the graph case, that analysis is simpler than that presented herein. Kloster and Gleich [9] have presented a deterministic algorithm that approximately simulates the heat equation for graphs. Hence, they extracted a tightly connected subset by considering a local part of the graph only.

The concept of the Laplacian for hypergraphs has been implicitly employed in semi-supervised learning on hypergraphs in the form $\boldsymbol{x}^\top L_G(\boldsymbol{x})$, where $\mathcal{L}_G(\boldsymbol{x}) = L_G(D_G^{-1}\boldsymbol{x})$ [8, 18]. This concept was then formally presented by Chan et al. [4] at a later time. Subsequently, the Laplacian concept was further generalized to handle submodular transformations [10, 17]; this development encompasses Laplacians for graphs, hypergraphs [4], and directed graphs [16].

Finally, we note that another type of Laplacian for hypergraphs, which essentially replaces each hyperedge with a clique, has been used in the literature [1, 14]. We stress that that Laplacian differs from the Laplacian for hypergraphs studied in this work.

1.4 Organization

The remainder of this paper is organized as follows. In Section 2, we introduce the basic concepts used throughout this paper. In Section 3, we prove Theorem 1. We show that (HE; $\boldsymbol{s}$) has a unique solution in Section 4. In Section 5, we prove the triangle inequality discussed in Section 1.2. A proof of Theorem 2 is given in Section 6.

2 Preliminaries

For a vector $\boldsymbol{x} \in \mathbb{R}^V$ and a set $S \subseteq V$, let $\boldsymbol{x}(S) = \sum_{v \in S} \boldsymbol{x}(v)$. For a vector $\boldsymbol{x} \in \mathbb{R}^V$ and a positive semidefinite matrix $A \in \mathbb{R}^{V \times V}$, we define $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_A = \boldsymbol{x}^\top A \boldsymbol{y}$ and $\|\boldsymbol{x}\|_A = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle_A} = \sqrt{\boldsymbol{x}^\top A \boldsymbol{x}}$.

Let $G = (V, E, w)$ be a hypergraph. Hereafter, we omit the subscript G from symbols such as A_G, L_G, d_G, and ϕ_G when it is clear from context. For $S \subseteq V$, let $\mathbf{1}_S \in \mathbb{R}^V$ denote the characteristic vector of S; that is, $\mathbf{1}_S(v) = 1$ if $v \in S$ and $\mathbf{1}_S(v) = 0$ otherwise. When $S = V$ or $S = \{v\}$, we simply write $\mathbf{1}$ and $\mathbf{1}_v$, respectively. Further, for $S \subseteq V$, we define a vector $\boldsymbol{\pi}_S \in \mathbb{R}^V$ as $\boldsymbol{\pi}_S(v) = \frac{d_G(v)}{\mathrm{vol}_G(S)}$ if $v \in S$ and $\boldsymbol{\pi}_S(v) = 0$ otherwise. When $S = V$ or $S = \{v\}$, we simply write $\boldsymbol{\pi}$ and $\boldsymbol{\pi}_v$, respectively. For a vector $\boldsymbol{\rho} \in \mathbb{R}^V$, we write $\boldsymbol{\rho}/d_G$ to denote a vector with $(\boldsymbol{\rho}/d_G)(v) = \boldsymbol{\rho}(v)/d_G(v)$ for each $v \in V$.

2.1 Normalized Laplacian for hypergraphs

Here, we formally define the (random-walk) normalized Laplacian for hypergraphs. Let $G = (V, E, w)$ be a hypergraph. For each hyperedge $e \in E$, we define a polytope $B_e = \mathrm{conv}(\{\mathbf{1}_u - \mathbf{1}_v \mid u, v \in e\})$, where $\mathrm{conv}(S)$ denotes the convex hull of $S \subseteq \mathbb{R}^V$. Then, the *Laplacian* $L \colon \mathbb{R}^V \to \mathbb{R}^V$ of G is defined as

$$L_G(\boldsymbol{x}) = \left\{ \sum_{e \in E} w(e) \boldsymbol{b}_e \boldsymbol{b}_e^\top \boldsymbol{x} \mid \boldsymbol{b}_e \in \operatorname*{argmax}_{\boldsymbol{b} \in B_e} \boldsymbol{b}^\top \boldsymbol{x} \right\}, \tag{4}$$

and the *normalized Laplacian* is defined as $\mathcal{L}_G \colon \boldsymbol{x} \mapsto L_G(D_G^{-1}\boldsymbol{x})$.

We can write $L_G(\boldsymbol{x})$ more explicitly, as follows. For each $e \in E$, let $S_e = \mathrm{argmax}_{v \in e} \boldsymbol{x}(v)$ and $I_e = \mathrm{argmin}_{v \in e} \boldsymbol{x}(v)$. Let $E' = \{uv \mid e \in E, u \in S_e, v \in I_e\} \cup \{vv \mid v \in V\}$. Then, we arbitrarily define a function $w'_e \colon E' \to \mathbb{R}_+$ such that $w'_e(uv) > 0$ only if $u \in S_e$ and $v \in I_e$, and we have $\sum_{u \in S_e, v \in I_e} w'_e(uv) = w(e)$. Then, we construct a graph $G' = (V, E', w')$, where $w'(uv) = \sum_{e \in E \mid u \in S_e, v \in I_e} w'_e(uv)$ for each $uv \in E'$ and $w'(vv) = d_G(v) - \sum_{e \in E' \mid v \in e} w'(e)$ for each $v \in V$. Note that $d_G(v) = d_{G'}(v)$ for every $v \in V$. Let $\mathcal{G}(G, \boldsymbol{x})$ be the set of graphs constructed in this manner. Hence, we have $L_G(\boldsymbol{x}) = \{L_{G'}\boldsymbol{x} \mid G' \in \mathcal{G}(G, \boldsymbol{x})\}$.

We can understand Laplacian for hypergraphs in terms of submodular functions. Let $F_e \colon 2^V \to \{0, 1\}$ be the cut function associated with a hyperedge $e \in E$; that is, $F_e(S) = 1$ if and only if $S \cap e \neq \emptyset$ and $(V \setminus S) \cap e \neq \emptyset$. It is known that F_e is *submodular*; that is, $F_e(S) + F_e(T) \geq F_e(S \cap T) + F_e(S \cup T)$ holds for every $S, T \subseteq V$. Then, B_e is the base polytope of F_e and $\boldsymbol{b}_e$ in (4) is chosen so that $\boldsymbol{b}_e^\top \boldsymbol{x} = f_e(\boldsymbol{x})$, where $f_e \colon \mathbb{R}^V \to \mathbb{R}$ is the *Lovász extension* of F_e. See [7] for detailed definitions of these concepts.

When $G = (V, E, w)$ is a graph, its *Laplacian* $L_G \in \mathbb{R}^{V \times V}$ and the *(random-walk) normalized Laplacian* $\mathcal{L}_G \in \mathbb{R}^{V \times V}$ are defined as $D_G - A_G$ and $I_G - A_G D_G^{-1}$, respectively. Indeed, this coincides with (4) when we regard G as a hypergraph with each hyperedge having size two.

2.2 Heat equation

Let us briefly review some facts regarding the heat equation (HE; $\boldsymbol{s}$). We say that $\{\boldsymbol{\rho}_t\}_{t\geq 0}$ is a *solution* of (HE; $\boldsymbol{s}$) if $\boldsymbol{\rho}_t$ is absolutely continuous with respect to t (hence, $\boldsymbol{\rho}_t$ is differentiable at almost all t) and $\boldsymbol{\rho}_0 = \boldsymbol{s}$, and satisfies $\frac{d}{dt}\boldsymbol{\rho}_t \in -\mathcal{L}_G(\boldsymbol{\rho}_t)$ for almost all $t \geq 0$. As discussed in Section 4 below, the heat equation (HE; $\boldsymbol{s}$) always has a unique solution. In addition, when G is connected, $\boldsymbol{\rho}_t$ converges to $\boldsymbol{\pi}$ as $t \to \infty$ for any $\boldsymbol{s} \in \mathbb{R}^V$ with $\sum_{v\in V} \boldsymbol{s}(v) = 1$, as mentioned previously.

Suppose that we begin the heat equation on a hypergraph $G = (V, E, w)$ with an initial vector $\boldsymbol{s} \in \mathbb{R}^V$. Then, there is a time sequence $t_0 = 0 < t_1 < t_2 < \cdots$ such that a weighted graph $G_i \in \mathcal{G}(G, \boldsymbol{\rho}_{t_i})$ exists for each $i \in \mathbb{Z}_+$, such that the heat equation on the interval $[t_i, t_{i+1})$ satisfies

$$\frac{d\boldsymbol{\rho}_t}{dt} = -\mathcal{L}_i \boldsymbol{\rho}_t,$$

where $\mathcal{L}_i$ is the normalized Laplacian associated with G_i. Hence, we can write the solution $\boldsymbol{\rho}_{i,\Delta} := \boldsymbol{\rho}_{t_i+\Delta}$ for $\Delta \in [0, t_{i+1} - t_i)$ as

$$\boldsymbol{\rho}_{i,\Delta} := H_{i,\Delta}\boldsymbol{\rho}_{t_i}, \quad \text{where} \quad H_{i,\Delta} := e^{-\Delta\mathcal{L}_i} = \sum_{n=0}^{\infty} \frac{(-\Delta)^n \mathcal{L}_i^n}{n!}. \tag{5}$$

For $t \in [t_i, t_{i+1})$, it is easy to see that

$$\boldsymbol{\rho}_t = \boldsymbol{\rho}_{i,t-t_i} = H_{i,t-t_i}\boldsymbol{\rho}_{t_i} = H_{i,t-t_i} H_{i-1,t_i-t_{i-1}} \cdots H_{1,t_2-t_1} H_{0,t_1} \boldsymbol{s}.$$

Although $\boldsymbol{\rho}_{i,\Delta}$ was originally defined for $\Delta \in [0, t_{i+1} - t_i)$, we can extend it to any $\Delta \geq 0$ using (5). Note that, when we wish to stress the initial vector, we write $\boldsymbol{\rho}_t^{\boldsymbol{s}}$, $\boldsymbol{\rho}_{i,\Delta}^{\boldsymbol{s}}$, etc.

3 Proof of Theorem 1

In this section, we prove Theorem 1.

Consider the heat equation (HE; $\boldsymbol{s}$). For each $i \in \mathbb{Z}_+$, we define a function $f_i \colon \mathbb{R}_+ \to \mathbb{R}$ as

$$f_i(\Delta) := \boldsymbol{\rho}_{i,0}^\top D^{-1} \left(\boldsymbol{\rho}_{i,\Delta} - \boldsymbol{\pi}\right).$$

Note that, what we wish to stress the initial vector $\boldsymbol{s} \in \mathbb{R}^V$, we write $f_i^{\boldsymbol{s}}$. As the following proposition implies, the value of $f_i(\Delta)$ indicates the difference between $\boldsymbol{\rho}_{i,\Delta/2}$ and the stationary distribution $\boldsymbol{\pi}$.

Proposition 1. *For any initial vector $\boldsymbol{s} \in \mathbb{R}^V$, $i \in \mathbb{Z}_+$, and $\Delta \geq 0$, we have*

$$f_i(\Delta) = \|\boldsymbol{\rho}_{i,\Delta/2} - \boldsymbol{\pi}\|_{D^{-1}}^2 = \sum_{w\in V} \left(\frac{\boldsymbol{\rho}_{i,\Delta/2}(w)}{d(w)} - \frac{1}{\mathrm{vol}(V)}\right)^2 d(w) \geq 0.$$

Proof.

$$\boldsymbol{\rho}_{i,0}^\top D^{-1}(\boldsymbol{\rho}_{i,\Delta} - \boldsymbol{\pi}) = \|D^{-1/2}(H_{i,\Delta/2} - \boldsymbol{\pi}\mathbf{1}^\top)\boldsymbol{\rho}_{i,0}\|^2 = \|D^{-1/2}(\boldsymbol{\rho}_{i,\Delta/2} - \boldsymbol{\pi})\|^2.$$

The second equality is obtained through a direct calculation. □

Theorem 1 is obtained by bounding $f_i(\Delta)$ from above and below. To obtain an upper bound, we define

$$\kappa_{i,\Delta} = \min\left\{\phi_{G_i}(S) \mid 0 \leq \xi \leq \Delta,\, S \in \mathrm{sweep}\left(\frac{\boldsymbol{\rho}_{i,\xi}}{d}\right)\right\} \quad (i \in \mathbb{Z}_+, \Delta \geq 0),$$
$$\kappa_i = \kappa_{i,t_{i+1}-t_i} \quad (i \in \mathbb{Z}_+),$$
$$\kappa_t = \min\left\{\min_{j=0,\dots,i-1} \kappa_j,\, \kappa_{i,\Delta}\right\}, \text{where } i \in \mathbb{Z}_+ \text{ is such that } t \in [t_i, t_{i+1}) \text{ and } \Delta = t - t_i \quad (t \geq 0).$$

Again, when we wish to stress the initial vector $\boldsymbol{s} \in \mathbb{R}^V$, we write $\kappa^{\boldsymbol{s}}_{i,\Delta}$, etc. In the following lemma, we present an upper bound on $f_i(\Delta)$ when the initial vector $\boldsymbol{s}$ is $\boldsymbol{\pi}_S$ for some set $S \subseteq V$.

Lemma 1. *Consider the heat equation* (HE; $\boldsymbol{\pi}_S$) *for a set $S \subseteq V$. For $t \geq 0$, let $i \in \mathbb{Z}_+$ be such that $t \in [t_i, t_{i+1})$ and let $\Delta = t - t_i$ and $t' = t_i + \Delta/2$. Then, we have*

$$f_i(\Delta) \leq \frac{1-\boldsymbol{\pi}(S)}{\mathrm{vol}(S)} \exp\left(-\kappa_{t'}^2 t'\right).$$

Next, we consider a lower bound on $f_i(\Delta)$, when the initial vector $\boldsymbol{s}$ is $\boldsymbol{\pi}_S$ for some set $S \subseteq V$. For simplicity, we write $\phi_0(S)$ to denote $\phi_{G_0}(S)$.

Lemma 2. *Consider the heat equation* (HE; $\boldsymbol{\pi}_S$) *for a set $S \subseteq V$. For $t \geq 0$, let $i \in \mathbb{Z}_+$ be such that $t \in [t_i, t_{i+1})$ and let $\Delta = t - t_i$ and $t' = t_i + \Delta/2$. Then, we have*

$$f_i(\Delta) \geq \frac{1-\boldsymbol{\pi}(S)}{\mathrm{vol}(S)} \exp\left(-\frac{2\phi_0(S)}{1-\boldsymbol{\pi}(S)} t'\right).$$

The following lemma is useful to relate the heat equation solutions to the different initial vectors.

Lemma 3. *Let $\boldsymbol{s}_1, \dots, \boldsymbol{s}_m \in \mathbb{R}^V$ be vectors and let $\boldsymbol{s} = \sum_{i=1}^m \boldsymbol{s}_i$. Then, we have*

$$\|\boldsymbol{\rho}_t^{\boldsymbol{s}}\|_{D^{-1}} \leq \sum_{i=1}^m \|\boldsymbol{\rho}_t^{\boldsymbol{s}_i}\|_{D^{-1}}.$$

We provide proofs of Lemmas 1, 2, and 3 in Sections 3.2, 3.3, and 5, respectively.

Based on these lemmas, we show the following:

Theorem 3. *Let $G = (V, E, w)$ be a hypergraph, and $S \subseteq V$ and $t \geq 0$. Then, we have*

$$\frac{\phi_0(S)}{1-\boldsymbol{\pi}(S)} \geq \frac{\kappa_{t,S}^2}{2} - \frac{2}{t} \log \frac{\mathrm{vol}(S)}{1-\boldsymbol{\pi}(S)},$$

where $\kappa_{t,S} = \min_{v \in S} \kappa_t^{\boldsymbol{\pi}_v}$.

Proof. Let $i \in \mathbb{Z}_+$ such that $t \in [t_i^{\boldsymbol{\pi}_S}, t_{i+1}^{\boldsymbol{\pi}_S})$, $\Delta = t - t_i^{\boldsymbol{\pi}_S}$, and $t' = t_i^{\boldsymbol{\pi}_S} + \Delta/2$. By Lemma 2, we have

$$\frac{1-\boldsymbol{\pi}(S)}{\mathrm{vol}(S)} \exp\left(-\frac{2\phi_0(S)}{1-\boldsymbol{\pi}(S)} t\right) p \leq f_i^{\boldsymbol{\pi}_S}(\Delta) = \|\boldsymbol{\rho}_{i,\Delta/2}^{\boldsymbol{\pi}_S} - \boldsymbol{\pi}\|_{D^{-1}}^2 = \|\boldsymbol{\rho}_{t'}^{\boldsymbol{\pi}_S} - \boldsymbol{\pi}\|_{D^{-1}}^2.$$

By [4, Lemma 4.12], we have

$$-\mathcal{L}_G(\boldsymbol{\rho}_t^{\boldsymbol{\pi}_S} - \boldsymbol{\pi}) = -\mathcal{L}_G(\boldsymbol{\rho}_t^{\boldsymbol{\pi}_S}) \ni \frac{d}{dt}\boldsymbol{\rho}_t^{\boldsymbol{\pi}_S} = \frac{d}{dt}(\boldsymbol{\rho}_t^{\boldsymbol{\pi}_S} - \boldsymbol{\pi}).$$

Hence $\boldsymbol{\rho}_t^{\boldsymbol{\pi}_S} - \boldsymbol{\pi}$ is the solution to the heat equation $(\mathrm{HE}; \boldsymbol{\pi}_S - \boldsymbol{\pi})$. Similarly, $\boldsymbol{\rho}_{t, \frac{d(v)}{\mathrm{vol}(S)}\boldsymbol{\pi}_v} - \frac{d(v)}{\mathrm{vol}(S)}\boldsymbol{\pi}$ is also the solution to the heat equation $(\mathrm{HE}; \frac{d(v)}{\mathrm{vol}(S)}(\boldsymbol{\pi}_v - \boldsymbol{\pi}))$. Then, Lemma 3 implies that

$$\|\boldsymbol{\rho}_{t'}^{\boldsymbol{\pi}_S} - \boldsymbol{\pi}\|_{D^{-1}} = \left\|\boldsymbol{\rho}_{t'}^{\boldsymbol{\pi}_S} - \sum_{v \in S} \frac{d(v)}{\mathrm{vol}(S)}\boldsymbol{\pi}\right\|_{D^{-1}} \leq \sum_{v \in S} \left\|\boldsymbol{\rho}_{t'}^{\frac{d(v)}{\mathrm{vol}(S)}\boldsymbol{\pi}_v} - \frac{d(v)}{\mathrm{vol}(S)}\boldsymbol{\pi}\right\|_{D^{-1}}$$
$$= \sum_{v \in S} \frac{d(v)}{\mathrm{vol}(S)} \|\boldsymbol{\rho}_{t'}^{\boldsymbol{\pi}_v} - \boldsymbol{\pi}\|_{D^{-1}} \leq \|\boldsymbol{\rho}_{t'}^{\boldsymbol{\pi}_{v^*}} - \boldsymbol{\pi}\|_{D^{-1}},$$

where the last equality follows from $\mathcal{L}_G(a\boldsymbol{x}) = a\mathcal{L}_G(\boldsymbol{x})$ for any scalar $a \in \mathbb{R}_+$ and $\boldsymbol{x} \in \mathbb{R}^V$, and $v^* \in S$ is a maximizer of $\|\boldsymbol{\rho}_{t'}^{\boldsymbol{\pi}_v} - \boldsymbol{\pi}\|_{D^{-1}}$. Let i^* be such that $t' \in [t_{i^*}^{\boldsymbol{\pi}_{v^*}}, t_{i^*+1}^{\boldsymbol{\pi}_{v^*}})$, and $\Delta'/2 = t' - t_{i^*}^{\boldsymbol{\pi}_{v^*}}$. Then, Lemma 1 instantiated with $\boldsymbol{\pi}_{v^*}$ implies

$$\|\boldsymbol{\rho}_{t'}^{\boldsymbol{\pi}_{v^*}} - \boldsymbol{\pi}\|_{D^{-1}}^2 = f_{i^*}^{\boldsymbol{\pi}_{v^*}}(\Delta') \leq \left(\frac{1}{d(v^*)} - \frac{1}{\mathrm{vol}(V)}\right) \exp\left(-\left(\kappa_{t'}^{\boldsymbol{\pi}_{v^*}}\right)^2 t'\right) \leq \exp\left(-\left(\kappa_{t'}^{\boldsymbol{\pi}_{v^*}}\right)^2 t'\right).$$

To summarize, we have obtained the following inequality:

$$\frac{1 - \boldsymbol{\pi}(S)}{\mathrm{vol}(S)} \exp\left(-\frac{2\phi_0(S)}{1 - \boldsymbol{\pi}(S)} t'\right) \leq \exp\left(-\left(\kappa_{t'}^{\boldsymbol{\pi}_{v^*}}\right)^2 t'\right).$$

Hence, by taking the logarithm, we have

$$\frac{\phi_0(S)}{1 - \boldsymbol{\pi}(S)} \geq \frac{(\kappa_{t',S})^2}{2} - \frac{1}{t'} \log \frac{\mathrm{vol}(S)}{1 - \boldsymbol{\pi}(S)}.$$

Note that we have $t' \geq t_i^{\boldsymbol{\pi}_S} + \Delta/2 \geq (t_i^{\boldsymbol{\pi}_S} + \Delta)/2 \geq t/2$; hence, $-1/t' \geq -2/t$ and $\kappa_{t',S} \leq \kappa_{t,S}$. Thus, the claim holds. $\square$

The following lemma relates the conductance of a set in a hypergraph G and that in a graph in $\mathcal{G}(G, \boldsymbol{x})$.

Lemma 4. *Let $G = (V, E, w)$ be a hypergraph and $\boldsymbol{x} \in \mathbb{R}^V$ be a vector. For any $\emptyset \subsetneq S \subsetneq V$, we have $\phi_{G'}(S) \leq \phi_G(S)$ for every $G' = (V, E', w') \in \mathcal{G}(G, \boldsymbol{x})$. When S is a sweep set with respect to $\boldsymbol{x}$, the equality is attained.*

Proof. We have

$$\sum_{uv \in \partial_{G'}(S)} w'(uv) = \sum_{uv \in \partial_{G'}(S)} \sum_{e \in E} w'_e(uv) = \sum_{e \in E} \sum_{uv \in \partial_{G'}(S)} w'_e(uv) = \sum_{e \in \partial_G(S)} \sum_{uv \in \partial_{G'}(S)} w'_e(uv)$$
$$\leq \sum_{e \in \partial_G(S)} \sum_{u \in S_e, v \in I_e} w'_e(uv) = \sum_{e \in \partial_G(S)} w(e).$$

Thus, we have $\phi_{G'}(S) \leq \phi_G(S)$ as $\mathrm{vol}_{G'}(S) = \mathrm{vol}_G(S)$. In addition, the equality holds when S is a sweep set with respect to $\boldsymbol{x}$, because $\sum_{uv \in \partial_{G'}(S)} w'_e(uv) = \sum_{u \in S_e, v \in I_e} w'_e(uv)$ holds for every hyperedge $e \in E$. $\square$

Proof of Theorem 1. By Lemma 4, we have $\phi_0(S) \leq \phi_G(S)$ and $\kappa_{t,S} = \widetilde{\kappa}_t$. Hence, Theorem 3 implies Theorem 1. $\square$

3.1 Useful lemmas

In this section, we derive several inequalities on f_i that will be useful later. Note that the proofs are deferred to Section A. We define $\mathcal{R}_i \colon \mathbb{R}^V \to \mathbb{R}$ as

$$\mathcal{R}_i(\boldsymbol{x}) = \frac{\boldsymbol{x}^\top L_{G_i} \boldsymbol{x}}{\|\boldsymbol{x}\|_D} = \frac{\sum_{uv \in E_i} (\boldsymbol{x}(u) - \boldsymbol{x}(v))^2 w_i(uv)}{\sum_{v \in V} \boldsymbol{x}(v)^2 d(v)}. \tag{6}$$

Lemma 5. *For any $i \in \mathbb{Z}_+$, we have*

$$\frac{d}{d\Delta} \log f_i(\Delta) = \frac{\boldsymbol{\rho}_{i,0}^\top D^{-1} \frac{d}{d\Delta} \boldsymbol{\rho}_{i,\Delta}}{\boldsymbol{\rho}_{i,0}^\top D^{-1} (\boldsymbol{\rho}_{i,\Delta} - \boldsymbol{\pi})} = -\mathcal{R}_i \left(\frac{\boldsymbol{\rho}_{i,\Delta/2}}{d} - \frac{1}{\mathrm{vol}(V)} \right).$$

Lemma 6. *We have*

$$-\frac{d}{d\Delta} \log f_0(\Delta)|_{\Delta=0} = \frac{\phi_0(S)}{1 - \boldsymbol{\pi}(S)}.$$

Lemma 7. *For any $i \in \mathbb{Z}_+$, we have*

$$\frac{d^2}{d\Delta^2} \log f_i(\Delta) \geq 0.$$

3.2 Proof of Lemma 1

We first derive a lower bound on the log derivative of $f_i(\Delta)$.

Lemma 8. *For any $i \in \mathbb{Z}_+$ and $\Delta \geq 0$, we have*

$$-\frac{d}{d\Delta} \log f_i(\Delta) \geq \frac{\kappa_{i,\Delta/2}^2}{2}.$$

Proof. By Lemma 5, we have

$$-\frac{d}{d\Delta} \log f_i(\Delta) = \mathcal{R}_i \left(\frac{\boldsymbol{\rho}_{i,\Delta/2}}{d} - \frac{1}{\mathrm{vol}(V)} \right).$$

Then, by applying Cheeger's inequality on the vector $\boldsymbol{\rho}_{i,\Delta/2}/d$, we obtain

$$\max_{c \in \mathbb{R}} \mathcal{R}_i \left(\frac{\boldsymbol{\rho}_{i,\Delta/2}}{d} - c \right) \geq \frac{\kappa_{i,\Delta/2}^2}{2}.$$

Hence, it suffices to show that the left hand side (LHS) attains the maximum value when $c = 1/\mathrm{vol}(V)$. Let $\varphi \colon \mathbb{R} \to \mathbb{R}$ be the denominator of the LHS (recall (6)) as a function of c. Then,

$$\varphi'(c) = -2 \sum_{v \in V} \left(\frac{\boldsymbol{\rho}_{i,\Delta/2}(v)}{d(v)} - c \right) d(v).$$

Hence $\varphi'(c) = 0$ yields

$$\sum_{v \in V} \boldsymbol{\rho}_{i,\Delta/2}(v) - \left(\sum_{v \in V} d(v) \right) c = 0,$$

which implies $c = 1/\mathrm{vol}(V)$ attains the minimum of φ. $\square$

Proof of Lemma 1. By Lemma 8, we have

$$\log f_i(\Delta) - \log f_i(0) \leq -\frac{\kappa_{i,\Delta/2}^2}{2}\Delta,$$
$$\log f_j(2(t_{j+1} - t_j)) - \log f_j(0) \leq -\kappa_j^2(t_{j+1} - t_j) \quad (j = 0, \dots, i-1).$$

Hence, we have

$$\begin{aligned}
&f_i(\Delta) \leq f_i(0) \exp\left(-\frac{\kappa_{i,\Delta/2}^2}{2}\Delta\right) = \|\boldsymbol{\rho}_{i,0} - \boldsymbol{\pi}\|_{D^{-1}}^2 \exp\left(-\frac{\kappa_{i,\Delta/2}^2}{2}\Delta\right) \\
&= \|\boldsymbol{\rho}_{i-1,t_i-t_{i-1}} - \boldsymbol{\pi}\|_{D^{-1}}^2 \exp\left(-\frac{\kappa_{i,\Delta/2}^2}{2}\Delta\right) = f_{i-1}(2(t_i - t_{i-1})) \exp\left(-\frac{\kappa_{i,\Delta/2}^2}{2}\Delta\right) \\
&\leq f_{i-1}(0) \exp\left(-\frac{\kappa_{i,\Delta/2}^2}{2}\Delta - \kappa_{i-1}^2(t_i - t_{i-1})\right) \leq \cdots \\
&\leq f_0(0) \exp\left(-\frac{\kappa_{i,\Delta/2}^2}{2}\Delta - \sum_{j=0}^{i-1}\kappa_j^2(t_{j+1} - t_j)\right) = \|\boldsymbol{\pi}_S - \boldsymbol{\pi}\|_{D^{-1}}^2 \exp\left(-\frac{\kappa_{i,\Delta/2}^2}{2}\Delta - \sum_{j=0}^{i-1}\kappa_j^2(t_{j+1} - t_j)\right) \\
&\leq \frac{1 - \boldsymbol{\pi}(S)}{\mathrm{vol}(S)} \exp\left(-\kappa_{t'}^2 t'\right).
\end{aligned}$$

In the final inequality, we used the following fact:

$$\|\boldsymbol{\pi}_S - \boldsymbol{\pi}\|_{D^{-1}}^2 = \boldsymbol{\pi}_S^\top D^{-1}(\boldsymbol{\pi}_S - \boldsymbol{\pi}) = \frac{1}{\mathrm{vol}(S)} - \frac{1}{\mathrm{vol}(V)} = \frac{1 - \boldsymbol{\pi}(S)}{\mathrm{vol}(S)}. \qquad \square$$

3.3 Proof of Lemma 2

Proof of Lemma 2. By Lemma 7, the function

$$F_j(\Delta) = -\frac{d}{d\Delta}\log f_j(\Delta)$$

is non-increasing in Δ for every $j \in \mathbb{Z}_+$. By Lemma 5, we have $F_j(2(t_{j+1} - t_j)) = F_{j+1}(0)$ for any $j \in \mathbb{Z}_+$. Hence, $F_0(0)$ attains the maximum of $F_j(\Delta)$ over $j \in \mathbb{Z}_+$ and $\Delta \geq 0$. By Lemma 6, we have

$$-\log f_i(\Delta) + \log f_i(0) \leq \frac{\phi_0(S)}{1 - \boldsymbol{\pi}(S)}\Delta.$$

Hence, we have

$$\begin{aligned}
f_i(\Delta) &\geq f_i(0) \exp\left(-\frac{\phi_0(S)}{1 - \boldsymbol{\pi}(S)}\Delta\right) = \|\boldsymbol{\rho}_{i,0} - \boldsymbol{\pi}\|_{D^{-1}}^2 \exp\left(-\frac{\phi_0(S)}{1 - \boldsymbol{\pi}(S)}\Delta\right) \\
&= \|\boldsymbol{\rho}_{i-1,t_i-t_{i-1}} - \boldsymbol{\pi}\|_{D^{-1}}^2 \exp\left(-\frac{\phi_0(S)}{1 - \boldsymbol{\pi}(S)}\Delta\right) = f_{i-1}(2(t_i - t_{i-1})) \exp\left(-\frac{\phi_0(S)}{1 - \boldsymbol{\pi}(S)}\Delta\right) \geq \cdots \\
&\geq f_0(0) \exp\left(-\frac{\phi_0(S)}{1 - \boldsymbol{\pi}(S)}\Delta - \frac{2\phi_0(S)}{1 - \boldsymbol{\pi}(S)}t_i\right) = \frac{1 - \boldsymbol{\pi}(S)}{\mathrm{vol}(S)} \exp\left(-\frac{2\phi_0(S)}{1 - \boldsymbol{\pi}(S)}t'\right).
\end{aligned}$$

In the last inequality, we used the relation $f_0(0) = \|\boldsymbol{\pi}_S - \boldsymbol{\pi}\|_{D^{-1}}^2 = (1 - \boldsymbol{\pi}(S))/\mathrm{vol}(S)$. $\square$

4 Existence of Solution

In this section, we show the existence of a solution to the heat equation (HE; $\boldsymbol{s}$) using the theory of monotone operators. We refer the interested reader to the books by Miyadera [11] and Showalter [15] for a detailed description of this topic.

We begin by introducing some definitions. Let $X = (X, \langle\cdot,\cdot\rangle)$ be a Hilbert space, $\|\cdot\|$ be the norm defined from the inner product, and $A\colon X \to 2^X$ be a multi-valued operator on X. Let $R(A) \subseteq X$ be the *range* of A. We often identify A with the *graph* of A; that is, $\{(x, y) \mid x \in X, y \in A(x)\} \subseteq X \times X$.

Definition 1. *An operator $A\colon X \to 2^X$ is* monotone *(or* accretive*) if, for any $x, x' \in X$ and $y \in A(x), y' \in A(x')$, we have*

$$\langle y - y', x - x'\rangle \geq 0.$$

When $-A$ is monotone, A is called dissipative.

Definition 2. *A monotone operator $A\colon X \to 2^X$ is* maximal *if A is maximal as a graph of the monotone operator on X; i.e., if there is a monotone operator $B\colon X \to 2^X$ with $A(x) \subseteq B(x)$ for any $x \in X$. Then we have $A = B$.*

To show that the heat equation (HE; $\boldsymbol{s}$) has a solution, by the theory of monotone operators, it is sufficient to show that $\mathcal{L}_G\colon \mathbb{R}^V \to 2^{\mathbb{R}^V}$ is a maximal monotone operator. In our case, the Hilbert space is $X = \mathbb{R}^V$ equipped with the inner product $\langle\cdot,\cdot\rangle_{D^{-1}}$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^V$.

Lemma 9. *The operator $\mathcal{L}_G$ is monotone.*

Proof. For any $\boldsymbol{x} \in \mathbb{R}^V$ and $\boldsymbol{y} \in \mathcal{L}_G(\boldsymbol{x})$, we can write

$$\boldsymbol{y} = BWB^\top D^{-1}\boldsymbol{x} = \sum_{e\in E} w(e)\boldsymbol{b}_e\boldsymbol{b}_e^\top\widetilde{\boldsymbol{x}},$$

where $\widetilde{\boldsymbol{x}} = D^{-1}\boldsymbol{x}$. Further, $W \in \mathbb{R}^{E\times E}$ is a diagonal matrix with the (e, e)-th entry being $w(e)$. $B = (\boldsymbol{b}_e)_{e\in E}$ is a matrix with column vectors $\boldsymbol{b}_e \in \mathbb{R}^V$, for which

$$\boldsymbol{b}_e \in \operatorname*{argmax}_{\boldsymbol{b}\in B_e}\langle\boldsymbol{b}, \widetilde{\boldsymbol{x}}\rangle.$$

We use this to show monotonicity. For $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^V$ and $\boldsymbol{y}_1 \in \mathcal{L}_G(\boldsymbol{x}_1), \boldsymbol{y}_2 \in \mathcal{L}_G(\boldsymbol{x}_2)$, we have

$$\boldsymbol{y}_1 = B_1WB_1^\top\widetilde{\boldsymbol{x}}_1, \quad \boldsymbol{y}_2 = B_2WB_2^\top\widetilde{\boldsymbol{x}}_2.$$

Then, we have

$$\begin{aligned}\langle\boldsymbol{y}_1 - \boldsymbol{y}_2, \boldsymbol{x}_1 - \boldsymbol{x}_2\rangle_{D^{-1}} &= \langle\boldsymbol{y}_1, \boldsymbol{x}_1\rangle_{D^{-1}} + \langle\boldsymbol{y}_2, \boldsymbol{x}_2\rangle_{D^{-1}} - \langle\boldsymbol{y}_2, \boldsymbol{x}_1\rangle_{D^{-1}} - \langle\boldsymbol{y}_1, \boldsymbol{x}_2\rangle_{D^{-1}} \\ &= \|B_1^\top\widetilde{\boldsymbol{x}}_1\|_W^2 + \|B_2^\top\widetilde{\boldsymbol{x}}_2\|_W^2 - \widetilde{\boldsymbol{x}}_2^\top B_2WB_2^\top\widetilde{\boldsymbol{x}}_1 - \widetilde{\boldsymbol{x}}_1^\top B_1WB_1^\top\widetilde{\boldsymbol{x}}_2 \\ &\geq \|B_1^\top\widetilde{\boldsymbol{x}}_1\|_W^2 + \|B_2^\top\widetilde{\boldsymbol{x}}_2\|_W^2 - \widetilde{\boldsymbol{x}}_2^\top B_2WB_1^\top\widetilde{\boldsymbol{x}}_1 - \widetilde{\boldsymbol{x}}_1^\top B_1WB_2^\top\widetilde{\boldsymbol{x}}_2 \\ &= \|B_1^\top\widetilde{\boldsymbol{x}}_1 - B_2^\top\widetilde{\boldsymbol{x}}_2\|_W^2 \geq 0.\end{aligned}$$

□

Lemma 10. *The operator $\mathcal{L}_G$ is maximal.*

Proof. By [15, IV.1. Proposition 1.6], it is sufficient to show that $R(I + \mathcal{L}_G) = \mathbb{R}^V$. This condition means that, for any $\boldsymbol{b} \in \mathbb{R}^V$, the equation $\boldsymbol{x} + \mathcal{L}_G(\boldsymbol{x}) \ni \boldsymbol{b}$ has a solution $\boldsymbol{x}$ in $\mathbb{R}^V$. In a previous work [6, §.3.1], an equivalent condition of the existence of the solution to $\mathcal{L}_G(\boldsymbol{x}) \ni \boldsymbol{b}$ was given. By a similar argument, we can give an equivalent condition for $\boldsymbol{x}+\mathcal{L}_G(\boldsymbol{x}) \ni \boldsymbol{b}$ and show the existence of the solution to $\boldsymbol{x}+\mathcal{L}_G(\boldsymbol{x}) \ni \boldsymbol{b}$. ☐

Corollary 1. *The heat equation* (HE; $\boldsymbol{s}$) *has a unique solution.*

Proof. This proof is an immediate consequence of Lemmas 9 and 10. See [15, IV, Proposition 3.1] for details. ☐

5 Triangle Inequality

In this section, we prove Lemma 3. Our proof consists of two steps:

1. Demonstration that the unique solution to the heat equation (HE; $\boldsymbol{s}$) given by Corollary 1 is integral.
2. Proof of Lemma 3 using a difference approximation of the integral solution.

Here, we state that $\{\boldsymbol{\rho}_t\}_{t\geq 0}$ is an *integral solution (of type* 0*)* to the heat equation (HE; $\boldsymbol{s}$) if $\boldsymbol{\rho}_t$ satisfies the following conditions (c.f., [11, Definition 5.1]):

1. $\boldsymbol{\rho}_0 = \boldsymbol{s}$;
2. $\boldsymbol{\rho}_t$ is continuous with respect to t;
3. for any $t, t' \geq 0$ with $t < t'$, $\boldsymbol{x} \in \mathbb{R}^V$, and $\boldsymbol{y} \in -\mathcal{L}_G(\boldsymbol{x})$, we have

$$\|\boldsymbol{\rho}_{t'} - \boldsymbol{x}\|_{D^{-1}}^2 - \|\boldsymbol{\rho}_t - \boldsymbol{x}\|_{D^{-1}}^2 \leq 2\int_t^{t'} \langle \boldsymbol{y}, \boldsymbol{\rho}_\xi - \boldsymbol{x}\rangle \, d\xi.$$

Lemma 11. *The heat equation* (HE; $\boldsymbol{s}$) *has a unique integral solution of type* 0*.*

Proof. By [11, Theorem 5.10], it is sufficient to check that $-\mathcal{L}_G$ is dissipative of type 0. For any $\boldsymbol{x} \in \mathbb{R}^V$, we have

$$\liminf_{\lambda \to 0+} \lambda^{-1} d(R(I + \lambda\mathcal{L}_G), \boldsymbol{x}) = 0,$$

where $d(R(I + \lambda\mathcal{L}_G), \boldsymbol{x})$ is the distance between the set $R(I + \lambda\mathcal{L}_G)$ and $\boldsymbol{x}$. In our case, $\mathcal{L}_G$ is monotone by Lemma 9 and satisfies $R(I + \lambda\mathcal{L}_G) \simeq \mathbb{R}^V$ for any $\lambda > 0$ by the proof of Lemma 10 and [15, IV.1. Lemma 1.3]. Hence, for any $\boldsymbol{x} \in \mathbb{R}^V$, we have $d(R(I + \lambda\mathcal{L}_G), \boldsymbol{x}) = 0$. ☐

Furthermore, we can show the following proposition:

Proposition 2. *The solution* $\boldsymbol{\rho}_t$ *of the heat equation* (HE; $\boldsymbol{s}$) *given by Corollary 1 is its (unique) integral solution.*

Proof. Let $\boldsymbol{x} \in \mathbb{R}^V$ and $\boldsymbol{y} \in \mathcal{L}_G(\boldsymbol{x})$. Then, for almost all $t \geq 0$, we have

$$\frac{d}{dt}\|\boldsymbol{\rho}_t - \boldsymbol{x}\|_{D^{-1}}^2 = 2\left\langle \frac{d}{dt}\boldsymbol{\rho}_t, \boldsymbol{\rho}_t - \boldsymbol{x}\right\rangle_{D^{-1}}.$$

Hence, we have

$$\left\langle \frac{d}{dt}\boldsymbol{\rho}_t, \boldsymbol{\rho}_t - \boldsymbol{x} \right\rangle_{D^{-1}} = \left\langle \frac{d}{dt}\boldsymbol{\rho}_t - \boldsymbol{y} + \boldsymbol{y}, \boldsymbol{\rho}_t - \boldsymbol{x} \right\rangle_{D^{-1}}$$
$$= \left\langle \frac{d}{dt}\boldsymbol{\rho}_t - \boldsymbol{y}, \boldsymbol{\rho}_t - \boldsymbol{x} \right\rangle_{D^{-1}} + \langle \boldsymbol{y}, \boldsymbol{\rho}_t - \boldsymbol{x}\rangle_{D^{-1}} \leq \langle \boldsymbol{y}, \boldsymbol{\rho}_t - \boldsymbol{x}\rangle_{D^{-1}},$$

where the last inequality follows from the monotonicity of $\mathcal{L}_G$ and by $\frac{d}{dt}\boldsymbol{\rho}_t \in -\mathcal{L}_G(\boldsymbol{\rho}_t)$. Taking the integral, we obtain

$$\|\boldsymbol{\rho}_t - \boldsymbol{x}\|_{D^{-1}}^2 - \|\boldsymbol{\rho}_{t'} - \boldsymbol{x}\|_{D^{-1}}^2 \leq 2\int_t^{t'} \langle \boldsymbol{y}, \boldsymbol{\rho}_\xi - \boldsymbol{x}\rangle_{D^{-1}}\, d\xi$$

for any $t, t' \geq 0$ with $t \leq t'$. □

Lemma 12 (Difference approximation)**.** *For any $\lambda > 0$, there exists a triple of sequences $\{t_k^\lambda\}, \{\boldsymbol{x}_k^\lambda\}$, and $\{\boldsymbol{y}_k^\lambda\}$ satisfying the following conditions, where $\boldsymbol{x}_k^\lambda \in \mathbb{R}^V$ and $\boldsymbol{y}_k^\lambda \in -\mathcal{L}_G(\boldsymbol{x}_k^\lambda)$:*

1. $0 = t_0^\lambda < t_1^\lambda < \cdots < t_k^\lambda < \cdots$ *with* $\lim_{k\to\infty} t_k^\lambda = \infty$,
2. $t_k^\lambda - t_{k-1}^\lambda < \lambda \quad (k = 1, 2, \dots)$,
3. $\|\boldsymbol{x}_k^\lambda - \boldsymbol{x}_{k-1}^\lambda - (t_k^\lambda - t_{k-1}^\lambda)\boldsymbol{y}_k^\lambda\|_{D^{-1}} < \lambda(t_k^\lambda - t_{k-1}^\lambda) \quad (k = 1, 2, \dots)$,

where $\boldsymbol{x}_0^\lambda = \boldsymbol{s}$. We define $\{\boldsymbol{\rho}_t^\lambda\}_{t\geq 0}$ as

$$\boldsymbol{\rho}_t^\lambda = \begin{cases} \boldsymbol{x} & \text{if } t = t_0^\lambda = 0, \\ \boldsymbol{x}_k^\lambda & \text{if } t \in (t_{k-1}^\lambda, t_k^\lambda]. \end{cases}$$

Then, the function $\boldsymbol{\rho}_t = \lim_{\lambda\to 0+} \boldsymbol{\rho}_t^\lambda$ is the unique integral solution of (HE; $\boldsymbol{s}$).

Proof. The existence of the sequences $\{t_k^\lambda\}, \{\boldsymbol{x}_k^\lambda\}, \{\boldsymbol{y}_k^\lambda\}$ follows from [11, Lemma 5.11]. The latter claim on $\boldsymbol{\rho}_t$ is in the proof of [11, Theorem 5.10]. □

By the uniqueness of an integral solution, our solution can be approximated using $\boldsymbol{\rho}_t^\lambda$. Next we prove Lemma 3.

Proof of Lemma 3. For simplicity, we only check the case of $m = 2$; i.e., we prove

$$\left\|\boldsymbol{\rho}_t^{\boldsymbol{a}+\boldsymbol{b}}\right\|_{D^{-1}} \leq \|\boldsymbol{\rho}_t^{\boldsymbol{a}}\|_{D^{-1}} + \|\boldsymbol{\rho}_t^{\boldsymbol{b}}\|_{D^{-1}}.$$

Note that the extension to general m is straightforward and is performed by applying this inequality $m-1$ times.

For $\lambda > 0$, let $\{t_k^\lambda(\boldsymbol{c})\}, \{\boldsymbol{x}_k^\lambda(\boldsymbol{c})\}$, $\{\boldsymbol{y}_k^\lambda(\boldsymbol{c})\}$, and $\{\boldsymbol{\rho}_t^\lambda(\boldsymbol{c})\}$ be the sequences corresponding to $\boldsymbol{c} = \boldsymbol{a}, \boldsymbol{b}, (\boldsymbol{a}+\boldsymbol{b})/2$ given in Lemma 12. From condition 3 of Lemma 12 and the triangle inequality, we have

$$\|\boldsymbol{x}_{k-1}^\lambda(\boldsymbol{c})\|_{D^{-1}} - \|(t_k^\lambda(\boldsymbol{c}) - t_{k-1}^\lambda(\boldsymbol{c}))\boldsymbol{y}_k^\lambda(\boldsymbol{c})\|_{D^{-1}} - \lambda(t_k^\lambda(\boldsymbol{c}) - t_{k-1}^\lambda(\boldsymbol{c})) < \|\boldsymbol{x}_k^\lambda\|_{D^{-1}}$$
$$< \|\boldsymbol{x}_{k-1}^\lambda(\boldsymbol{c})\|_{D^{-1}} + \|(t_k^\lambda(\boldsymbol{c}) - t_{k-1}^\lambda(\boldsymbol{c}))\boldsymbol{y}_k^\lambda(\boldsymbol{c})\|_{D^{-1}} + \lambda(t_k^\lambda(\boldsymbol{c}) - t_{k-1}^\lambda(\boldsymbol{c})).$$

By using this inequality and induction on k, we prove that

$$\|\boldsymbol{\rho}_t^\lambda(\boldsymbol{a})\|_{D^{-1}} + \|\boldsymbol{\rho}_t^\lambda(\boldsymbol{b})\|_{D^{-1}} - \|\boldsymbol{\rho}_t^\lambda(\boldsymbol{a}+\boldsymbol{b})\|_{D^{-1}} > F_k(\lambda) \tag{7}$$

holds for $t \in (t_{k-1}^\lambda, t_k^\lambda]$, for some function $F_k(\lambda)$ with $\lim_{\lambda\to 0+} F_k(\lambda) = 0$.

For $k = 0$, $t = t_0^\lambda = 0$. Then, by the usual triangle inequality, we have

$$\|\boldsymbol{\rho}_0^\lambda(\boldsymbol{a})\|_{D^{-1}} + \|\boldsymbol{\rho}_0^\lambda(\boldsymbol{b})\|_{D^{-1}} - \|\boldsymbol{\rho}_0^\lambda(\boldsymbol{a}+\boldsymbol{b})\|_{D^{-1}} = \|\boldsymbol{a}\|_{D^{-1}} + \|\boldsymbol{b}\|_{D^{-1}} - \|\boldsymbol{a}+\boldsymbol{b}\|_{D^{-1}} \geq 0.$$

We assume that inequality (7) holds for k; i.e., for $t \in (t_{k-1}^\lambda, t_k^\lambda]$, the inequality

$$\|\boldsymbol{\rho}_t^\lambda(\boldsymbol{a})\|_{D^{-1}} + \|\boldsymbol{\rho}_t^\lambda(\boldsymbol{b})\|_{D^{-1}} - \|\boldsymbol{\rho}_t^\lambda(\boldsymbol{a}+\boldsymbol{b})\|_{D^{-1}} > F_k(\lambda)$$

holds for some function $F_k(\lambda)$ with $\lim_{\lambda\to 0+} F_k(\lambda) = 0$. By using the above inequality, we have

$$\begin{aligned}
&\|\boldsymbol{x}_{k+1}^\lambda(\boldsymbol{a})\|_{D^{-1}} + \|\boldsymbol{x}_{k+1}^\lambda(\boldsymbol{b})\|_{D^{-1}} - \|\boldsymbol{x}_{k+1}^\lambda(\boldsymbol{a}+\boldsymbol{b})\|_{D^{-1}} > \\
&(\|\boldsymbol{x}_k^\lambda(\boldsymbol{a})\|_{D^{-1}} - \|(t_{k+1}^\lambda(\boldsymbol{a}) - t_k^\lambda(\boldsymbol{a}))\boldsymbol{y}_{k+1}^\lambda(\boldsymbol{a})\|_{D^{-1}} - \lambda(t_{k+1}^\lambda(\boldsymbol{a}) - t_k^\lambda(\boldsymbol{a})) \\
&+ \|\boldsymbol{x}_k^\lambda(\boldsymbol{b})\|_{D^{-1}} - \|(t_{k+1}^\lambda(\boldsymbol{b}) - t_k^\lambda(\boldsymbol{b}))\boldsymbol{y}_{k+1}^\lambda(\boldsymbol{b})\|_{D^{-1}} - \lambda(t_{k+1}^\lambda(\boldsymbol{b}) - t_k^\lambda(\boldsymbol{b})))- \\
&(\|\boldsymbol{x}_k^\lambda(\boldsymbol{a}+\boldsymbol{b})\|_{D^{-1}} - \|(t_{k+1}^\lambda(\boldsymbol{a}+\boldsymbol{b}) - t_k^\lambda(\boldsymbol{a}+\boldsymbol{b}))\boldsymbol{y}_{k+1}^\lambda(\boldsymbol{a}+\boldsymbol{b})\|_{D^{-1}} - \lambda(t_{k+1}^\lambda(\boldsymbol{a}+\boldsymbol{b}) - t_k^\lambda(\boldsymbol{a}+\boldsymbol{b}))) \\
&> F_k(\lambda) + G_{k+1}(\lambda).
\end{aligned}$$

The last inequality follows from the assumption of induction and by defining $G_{k+1}(\lambda)$ as

$$\begin{aligned}
G_{k+1}(\lambda) = &-\|(t_{k+1}^\lambda(\boldsymbol{a}) - t_k^\lambda(\boldsymbol{a}))\boldsymbol{y}_{k+1}^\lambda(\boldsymbol{a})\|_{D^{-1}} - \lambda(t_{k+1}^\lambda(\boldsymbol{a}) - t_k^\lambda(\boldsymbol{a})) \\
&- \|(t_{k+1}^\lambda(\boldsymbol{b}) - t_k^\lambda(\boldsymbol{b}))\boldsymbol{y}_{k+1}^\lambda(\boldsymbol{b})\|_{D^{-1}} - \lambda(t_{k+1}^\lambda(\boldsymbol{b}) - t_k^\lambda(\boldsymbol{b})) \\
&+ \|(t_{k+1}^\lambda(\boldsymbol{a}+\boldsymbol{b}) - t_k^\lambda(\boldsymbol{a}+\boldsymbol{b}))\boldsymbol{y}_{k+1}^\lambda(\boldsymbol{a}+\boldsymbol{b})\|_{D^{-1}} + \lambda(t_{k+1}^\lambda(\boldsymbol{a}+\boldsymbol{b}) - t_k^\lambda(\boldsymbol{a}+\boldsymbol{b})).
\end{aligned}$$

We set $F_{k+1}(\lambda) = F_k(\lambda) + G_{k+1}(\lambda)$. Then, $F_{k+1}(\lambda)$ is a finite sum of terms that go to 0 as $\lambda \to 0+$. Thus, $\lim_{\lambda\to 0+} F_{k+1}(\lambda) = 0$ holds.

As a conclusion, for any $\lambda > 0$ and $t \in \mathbb{R}_+$, if $t \in (t_{k-1}^\lambda, t_k^\lambda]$, we have

$$\|\boldsymbol{\rho}_t^\lambda(\boldsymbol{a})\|_{D^{-1}} + \|\boldsymbol{\rho}_t^\lambda(\boldsymbol{b})\|_{D^{-1}} - \|\boldsymbol{\rho}_t^\lambda(\boldsymbol{a}+\boldsymbol{b})\|_{D^{-1}} > F_k(\lambda).$$

Thus, by taking limit $\lambda \to 0+$ we have the inequality:

$$\|\boldsymbol{\rho}_t^{\boldsymbol{a}}\|_{D^{-1}} + \|\boldsymbol{\rho}_t^{\boldsymbol{b}}\|_{D^{-1}} - \|\boldsymbol{\rho}_t^{\boldsymbol{a}+\boldsymbol{b}}\|_{D^{-1}} \geq 0.$$

□

6 Computation and Error Analysis of Difference Approximation

In this section, we prove Theorem 2. In what follows, we fix a hypergraph $G = (V, E, w)$, $v \in V$, $T \geq 1$, and $\lambda \in (0, 1)$.

We first review the construction of difference approximation $\boldsymbol{\rho}_t^\lambda$ given in [11, Section 5.3]. By the condition (5.27) in [11] and the maximality of $\mathcal{L}_G$, for any $\boldsymbol{x} \in \mathbb{R}^V$, there is a real number μ satisfying the following conditions:

$$\begin{cases} 0 < \mu \leq \lambda, \\ \boldsymbol{x}_\mu \in \mathbb{R}^V,\ \boldsymbol{y}_\mu \in -\mathcal{L}_G(\boldsymbol{x}_\mu), \\ \|\boldsymbol{x}_\mu - \boldsymbol{x} - \mu\boldsymbol{y}_\mu\|_{D^{-1}} < \mu\lambda. \end{cases} \tag{8}$$

We define $\mu(\boldsymbol{x})$ as the least upper bound on μ satisfying (8). We consider an initial vector $\boldsymbol{x}_0 \in \mathbb{R}^V$. Then, there is $h_1 \in \mathbb{R}$ such that $\mu(\boldsymbol{x}_0)/2 < h_1 \leq \lambda$ and there are $\boldsymbol{x}_1 \in \mathbb{R}^V$ and $\boldsymbol{y}_1 \in -\mathcal{L}_G(\boldsymbol{x}_1)$ satisfying $\|\boldsymbol{x}_1 - \boldsymbol{x}_0 - h_1\boldsymbol{y}_1\|_{D^{-1}} < h_1\lambda$. By repeating this argument, we can take sequences $\{h_k\}$, $\{\boldsymbol{x}_k\}$, and $\{\boldsymbol{y}_k\}$ for $k = 1, 2, \ldots$ satisfying the following conditions:

1. $\mu(\boldsymbol{x}_{k-1})/2 < h_k \leq \lambda$,
2. $\|\boldsymbol{x}_k - \boldsymbol{x}_{k-1} - h_k \boldsymbol{y}_k\|_{D^{-1}} < h_k \lambda$.

Let $t_k = \sum_{j=1}^{k} h_k$. Then, it is easy to show that $\{t_k\}$, $\{\boldsymbol{x}_k\}$, and $\{\boldsymbol{y}_k\}$ satisfies the conditions for $\{t_k^\lambda\}$, $\{\boldsymbol{x}_k^\lambda\}$, and $\{\boldsymbol{y}_k^\lambda\}$ in Lemma 12. Then, the function $\boldsymbol{\rho}_t^\lambda$ was defined by

$$\boldsymbol{\rho}_t^\lambda = \begin{cases} \boldsymbol{x}_0 & \text{if } t = 0, \\ \boldsymbol{x}_k^\lambda & \text{if } t \in (t_k^\lambda, t_{k+1}^\lambda] \cap (0, T]. \end{cases} \tag{9}$$

Theorem 2 follows from Lemmas 13 and 14 below.

Lemma 13. *We can compute (a concise representation) of $\{\boldsymbol{\rho}_t^\lambda\}_{0 \leq t \leq T}$ for every $0 \leq t \leq T$ in time polynomial in $1/\lambda$, T, and $\sum_{e \in E} |e|$.*

Proof. From the construction of $\boldsymbol{\rho}_t^\lambda$, it suffices to compute $\boldsymbol{x}_k^\lambda$ until $t_k \geq T$.

Note that we can obtain $\boldsymbol{x}_k^\lambda$ from $\boldsymbol{x}_{k-1}^\lambda$ by solving the equation

$$\boldsymbol{x} - \boldsymbol{x}_{k-1}^\lambda \in -\lambda \mathcal{L}_G(\boldsymbol{x}), \tag{10}$$

because, then, we can set $h_k = \lambda$ and $\boldsymbol{x}_k^\lambda$ to be the obtained solution.

Let $\widetilde{\boldsymbol{x}} = D^{-1}\boldsymbol{x}$ for any $\boldsymbol{x} \in \mathbb{R}^V$. Then, solving (10) is equivalent to solving

$$D\widetilde{\boldsymbol{x}} - D\widetilde{\boldsymbol{x}}_{k-1}^\lambda \in -\lambda L_G(\widetilde{\boldsymbol{x}}). \tag{11}$$

By an argument similar to [6, Section 3.1], solving (11) is equivalent to computing the following proximal operator

$$\operatorname{prox}(\widetilde{\boldsymbol{x}}_{k-1}^\lambda) := \operatorname*{argmin}_{\widetilde{\boldsymbol{x}} \in \mathbb{R}^V} \left(\frac{\lambda}{2} \sum_{e \in E} w(e) f_e(\widetilde{\boldsymbol{x}})^2 + \frac{1}{2} \|\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}_{k-1}^\lambda\|_D^2 \right), \tag{12}$$

which can be computed in time polynomial in $\sum_{e \in E} |\mathcal{V}_e|$, where $\mathcal{V}_e$ is the set of extreme points of B_e [6, Theorem D.1 (i)]. As $\mathcal{V}_e \leq |e|^2$, we can compute $\boldsymbol{x}_k^\lambda = D \operatorname{prox}(\widetilde{\boldsymbol{x}}_{k-1}^\lambda)$ in time polynomial in $\sum_{e \in E} |e|$.

As $h_k = \lambda$, we need to compute $\boldsymbol{x}_k^\lambda$ for $k \leq \lceil T/k \rceil$. Hence, the total time complexity is polynomial in $1/\lambda$, T, and $\sum_{e \in E} |e|$. □

Lemma 14. *We have*

$$\|\boldsymbol{\rho}_t^\lambda - \boldsymbol{\rho}_t^{\boldsymbol{\pi}_v}\|_{D^{-1}} = O(\sqrt{\lambda T}).$$

Proof. Let $|||\mathcal{L}_G(\boldsymbol{x})||| = \inf\{\|\boldsymbol{y}\|_{D^{-1}} \mid \boldsymbol{y} \in \mathcal{L}_G(\boldsymbol{x})\}$. We set $N_\lambda \in \mathbb{Z}_+$ as $t_{N_\lambda}^\lambda < T \leq t_{N_\lambda}^\lambda$, $|\Delta_\lambda| = \max\{t_k^\lambda - t_{k-1}^\lambda; k = 1, 2, \ldots, N_\lambda\}$ and $\mathcal{E}_\lambda = \sum_{k=1}^{N_\lambda} \|\boldsymbol{\mathcal{E}}_k^\lambda\|_{D^{-1}} (t_k^\lambda - t_{k-1}^\lambda)$, where $\boldsymbol{\mathcal{E}}_k^\lambda$ is defined as

$$\boldsymbol{\mathcal{E}}_k^\lambda = \frac{\boldsymbol{x}_k^\lambda - \boldsymbol{x}_{k-1}^\lambda}{t_k^\lambda - t_{k-1}^\lambda} - \boldsymbol{y}_k^\lambda \quad (k = 1, 2, \ldots).$$

Then, by the equation (5.20) of [11] instantiated with $\omega_0 = 0$, $t = s$, $x_p = x$, we have

$$\|\boldsymbol{\rho}_t^\lambda - \boldsymbol{\rho}_t^\mu\|_{D^{-1}} \leq \mathcal{E}_\lambda + \mathcal{E}_\mu + \Big((|\Delta_\lambda| + |\Delta_\mu|)^2 + |\Delta_\lambda|(t + |\Delta_\lambda|) + |\Delta_\mu|(t + |\Delta_\mu|) \Big)^{\frac{1}{2}} \times |||\mathcal{L}_G(\boldsymbol{\pi}_v)|||$$

for $t \in [0, T]$ and $\mu > 0$. The condition (3) in Lemma 12 implies $\|\boldsymbol{\mathcal{E}}_k^\lambda\|_{D^{-1}} < \lambda$. Hence, $\mathcal{E}_\lambda < \lambda t_{N_\lambda}^\lambda < \lambda(T + \lambda)$ as $t_{N_\lambda - 1}^\lambda < T \leq t_{N_\lambda}^\lambda$.

Therefore by taking limit $\mu \to 0+$, we have

$$\|\boldsymbol{\rho}_t^\lambda - \boldsymbol{\rho}_t^{\boldsymbol{\pi}_v}\|_{D^{-1}} < \lambda(T + \lambda) + \sqrt{\lambda^2 + \lambda(t + \lambda)} |||\mathcal{L}_G(\boldsymbol{\pi}_v)||| = O(\sqrt{\lambda T}).$$ □

References

[1] S. Agarwal, K. Branson, and S. Belongie. Higher order learning with graphs. In *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, pages 17–24, 2006.

[2] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[3] N. Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.

[4] T.-H. H. Chan, A. Louis, Z. G. Tang, and C. Zhang. Spectral properties of hypergraph Laplacian and approximation algorithms. *Journal of the ACM*, 65(3):15–48, 2018.

[5] F. Chung. The heat kernel as the pagerank of a graph. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50):19735–19740, 2007.

[6] K. Fujii, T. Soma, and Y. Yoshida. Polynomial-time algorithms for submodular Laplacian systems. *arXiv preprint, arXiv:1803.10923*, 2018.

[7] S. Fujishige. *Submodular Functions and Optimization*. Elsevier, 2005.

[8] M. Hein, S. Setzer, L. Jost, and S. S. Rangapuram. The total variation on hypergraphs — Learning on hypergraphs revisited. In *Proceedings of the 27th Annual Conference on Neural Information Processing Systems (NIPS)*, pages 2427–2435, 2013.

[9] K. Kloster and D. F. Gleich. Heat kernel based community detection. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1386–1395, 2014.

[10] P. Li and O. Milenkovic. Submodular hypergraphs: p-Laplacians, Cheeger inequalities and spectral clustering. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pages 3014–3023, 2018.

[11] I. Miyadera. *Nonlinear Semigroups*, volume 109. American Mathematical Soc., 1992.

[12] P. Raghavendra and D. Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd ACM Annual Symposium on Theory of Computing (STOC)*, pages 755–764, 2010.

[13] P. Raghavendra and D. Steurer. Reductions between expansion problems. In *Proceedings of the 27th IEEE Annual Conference on Computational Complexity (CCC)*, pages 64–73, 2012.

[14] B. Schölkopf, J. Platt, and T. Hofmann. Learning with hypergraphs: Clustering, classification, and embedding. In *Proceedings of the 19th Annual Conference on Neural Information Processing Systems (NIPS)*, pages 1601–1608, 2006.

[15] R. E. Showalter. *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, volume 49. American Mathematical Soc., 2013.

[16] Y. Yoshida. Nonlinear Laplacian for digraphs and its applications to network analysis. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM)*, pages 483–492, 2016.

[17] Y. Yoshida. Cheeger inequalities for submodular transformations. *arXiv preprint, arXiv:1708.08781*, 2017.

[18] C. Zhang, S. Hu, Z. G. Tang, and T.-H. H. Chan. Re-revisiting learning on hypergraphs: Confidence interval and subgradient method. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pages 4026–4034, 2017.

A Proofs of Section 3.1

Lemma 15. *For any $i \in \mathbb{Z}_+$ and $\Delta \geq 0$, we have*

$$\boldsymbol{\rho}_{i,0}^\top D^{-1}\frac{d\boldsymbol{\rho}_{i,\Delta}}{d\Delta} = -(D^{-1}\boldsymbol{\rho}_{i,\Delta/2})^\top (D - A_i)(D^{-1}\boldsymbol{\rho}_{i,\Delta/2}) = -\sum_{uv\in E_i}\left(\frac{\boldsymbol{\rho}_{i,\Delta/2}(u)}{d(u)} - \frac{\boldsymbol{\rho}_{i,\Delta/2}(v)}{d(v)}\right)^2 w_i(uv) \leq 0.$$

Proof. We have

$$\begin{aligned}
\boldsymbol{\rho}_{i,0}^\top D^{-1}\frac{d\boldsymbol{\rho}_{i,\Delta}}{d\Delta} &= -\boldsymbol{\rho}_{i,0}^\top D^{-1}H_{i,\Delta}\mathcal{L}_i\boldsymbol{\rho}_{i,0} = -\boldsymbol{\rho}_{i,0}^\top D^{-1}H_{i,\Delta/2}H_{i,\Delta/2}\mathcal{L}_i\boldsymbol{\rho}_{i,0} && (\text{by } H_{i,\Delta} = H_{i,\Delta/2}H_{i,\Delta/2})\\
&= -\boldsymbol{\rho}_{i,0}^\top (H_{i,\Delta/2})^\top D^{-1}(D - A_i)D^{-1}H_{i,\Delta/2}\boldsymbol{\rho}_{i,0} && (\text{by } DH_{i,\Delta/2} = (H_{i,\Delta/2})^\top D)\\
&= -(D^{-1}\boldsymbol{\rho}_{i,\Delta/2})^\top (D - A_i)(D^{-1}\boldsymbol{\rho}_{i,\Delta/2}).
\end{aligned}$$

The second equality in the statement is obtained through a direct calculation. ☐

Proof of Lemma 5. The first equality is obtained through direct calculation and the second equality follows from Proposition 1 and Lemma 15. ☐

Proof of Lemma 6. By Lemma 5, we have

$$\text{LHS} = \frac{\sum_{uv\in E_i}\left(\frac{\boldsymbol{\pi}_S(u)}{d(u)} - \frac{\boldsymbol{\pi}_S(v)}{d(v)}\right)^2 w_0(uv)}{\sum_w \left(\frac{\mathbf{1}_S(w)}{d(w)} - \frac{1}{\mathrm{vol}(V)}\right)^2 d(w)} = \frac{\frac{1}{\mathrm{vol}(S)}\phi_0(S)}{\sum_w \left(\frac{\mathbf{1}_S(w)}{\mathrm{vol}(S)} - \frac{1}{\mathrm{vol}(V)}\right)^2 d(w)} = \frac{\frac{1}{\mathrm{vol}(S)}\phi_0(S)}{\frac{1}{\mathrm{vol}(S)}(1 - \boldsymbol{\pi}(S))} = \frac{\phi_0(S)}{1-\boldsymbol{\pi}(S)},$$

based on the fact that

$$\begin{aligned}
\sum_{v\in V}\left(\frac{\mathbf{1}_S(v)}{\mathrm{vol}(S)} - \frac{1}{\mathrm{vol}(V)}\right)^2 d(v) &= \left(\frac{1}{\mathrm{vol}(S)} - \frac{1}{\mathrm{vol}(V)}\right)^2 \sum_{v\in S} d(v) + \frac{1}{\mathrm{vol}(V)^2}\sum_{v\notin S} d(v)\\
&= \left(\frac{1}{\mathrm{vol}(S)} - \frac{1}{\mathrm{vol}(V)}\right)^2 \mathrm{vol}(S) + \frac{\mathrm{vol}(V) - \mathrm{vol}(S)}{\mathrm{vol}(V)^2}\\
&= \frac{(\mathrm{vol}(V) - \mathrm{vol}(S))^2}{\mathrm{vol}(S)\mathrm{vol}(V)^2} + \frac{\mathrm{vol}(V) - \mathrm{vol}(S)}{\mathrm{vol}(V)^2} = \frac{\mathrm{vol}(V) - \mathrm{vol}(S)}{\mathrm{vol}(V)\mathrm{vol}(S)} = \frac{1}{\mathrm{vol}(S)}(1 - \boldsymbol{\pi}(S)).
\end{aligned}$$
☐

Proof of Lemma 7. By Lemma 5, we have

$$\begin{aligned}
&-\frac{d^2}{d\Delta^2}(-\log(\boldsymbol{\rho}_{i,0}^\top D^{-1}(\boldsymbol{\rho}_{i,\Delta} - \boldsymbol{\pi}))) = \frac{d}{d\Delta}\left(-\frac{\boldsymbol{\rho}_{i,0}^\top D^{-1}\frac{d}{d\Delta}\boldsymbol{\rho}_{i,\Delta}}{\boldsymbol{\rho}_{i,0}^\top D^{-1}(\boldsymbol{\rho}_{i,\Delta} - \boldsymbol{\pi})}\right) = \frac{d}{d\Delta}\left(\frac{\boldsymbol{\rho}_{i,0}^\top D^{-1}\mathcal{L}_i\boldsymbol{\rho}_{i,\Delta}}{\boldsymbol{\rho}_{i,0}^\top D^{-1}(\boldsymbol{\rho}_{i,\Delta} - \boldsymbol{\pi})}\right)\\
&= \frac{\boldsymbol{\rho}_{i,0}^\top D^{-1}(\boldsymbol{\rho}_{i,\Delta} - \boldsymbol{\pi})\boldsymbol{\rho}_{i,0}^\top D^{-1}\mathcal{L}_i^2\boldsymbol{\rho}_{i,\Delta} - (\boldsymbol{\rho}_{i,0}^\top D^{-1}\mathcal{L}_i\boldsymbol{\rho}_{i,\Delta})^2}{(\boldsymbol{\rho}_{i,0}^\top D^{-1}(\boldsymbol{\rho}_{i,\Delta} - \boldsymbol{\pi}))^2}.
\end{aligned}$$

It is sufficient to check the positivity of the numerator. Note that the numerator can be written as

$$(\boldsymbol{\rho}_{i,0}^\top D^{-1}(\boldsymbol{\rho}_{i,\Delta} - \boldsymbol{\pi}))(\boldsymbol{\rho}_{i,0}^\top D^{-1}\mathcal{L}_i^2\boldsymbol{\rho}_{i,\Delta}) - (\boldsymbol{\rho}_{i,0}^\top D^{-1}\mathcal{L}_i\boldsymbol{\rho}_{i,\Delta})^2. \tag{13}$$

The first factor of the first term of (13) is

$$\boldsymbol{\rho}_{i,0}^\top D^{-1}(\boldsymbol{\rho}_{i,\Delta} - \boldsymbol{\pi}) = \|D^{-1/2}(\boldsymbol{\rho}_{i,\Delta/2} - \boldsymbol{\pi})\|^2$$

by Proposition 1. The second factor of the first term of (13) is

$$\begin{aligned}
&\boldsymbol{\rho}_{i,0}^\top D^{-1}\mathcal{L}_i^2\boldsymbol{\rho}_{i,\Delta} = \boldsymbol{\rho}_{i,0}^\top D^{-1}(I - A_iD^{-1})^2 H_{i,\Delta}\boldsymbol{\rho}_{i,0} = \boldsymbol{\rho}_{i,0}^\top D^{-1} D(H_{i,\Delta/2})^\top D^{-1}(I - A_iD^{-1})^2 H_{i,\Delta/2}\boldsymbol{\rho}_{i,0} \\
&= \boldsymbol{\rho}_{i,0}^\top (H_{i,\Delta/2})^\top D^{-1}(D - A_i)D^{-1}(D - A_i)D^{-1}H_{i,\Delta/2}\boldsymbol{\rho}_{i,0} \\
&= \|D^{-1/2}(D - A_i)D^{-1}H_{i,\Delta/2}\boldsymbol{\rho}_{i,0}\|^2 = \|D^{-1/2}(I - A_iD^{-1})\boldsymbol{\rho}_{i,\Delta/2}\|^2 = \|D^{-1/2}\mathcal{L}_i\boldsymbol{\rho}_{i,\Delta/2}\|^2.
\end{aligned}$$

The second term of (13) is

$$\begin{aligned}
\boldsymbol{\rho}_{i,0}^\top D^{-1}\mathcal{L}_i\boldsymbol{\rho}_{i,\Delta} &= \boldsymbol{\rho}_{i,0}^\top D^{-1}(I - A_iD^{-1})H_{i,\Delta}\boldsymbol{\rho}_{i,0} = \boldsymbol{\rho}_{i,0}^\top (H_{i,\Delta/2})^\top D^{-1}(I - A_iD^{-1})H_{i,\Delta/2}\boldsymbol{\rho}_{i,0} \\
&= \boldsymbol{\rho}_{i,\Delta/2}^\top D^{-1}(I - A_iD^{-1})\boldsymbol{\rho}_{i,\Delta/2} = \boldsymbol{\rho}_{i,\Delta/2}^\top D^{-1}\mathcal{L}_i\boldsymbol{\rho}_{i,\Delta/2}.
\end{aligned} \tag{14}$$

We can rephrase (14) as the inner product of the vectors $D^{-1/2}\mathcal{L}_i\boldsymbol{\rho}_{i,\Delta/2}$ and $D^{-1/2}(\boldsymbol{\rho}_{i,\Delta/2} - \boldsymbol{\pi})$, as follows:

$$\begin{aligned}
&(D^{-1/2}\mathcal{L}_i\boldsymbol{\rho}_{i,\Delta/2})^\top D^{-1/2}(\boldsymbol{\rho}_{i,\Delta/2} - \boldsymbol{\pi}) = \boldsymbol{\rho}_{i,\Delta/2}^\top \mathcal{L}_i^\top D^{-1/2}D^{-1/2}(\boldsymbol{\rho}_{i,\Delta/2} - \boldsymbol{\pi}) \\
&= \boldsymbol{\rho}_{i,\Delta/2}^\top \mathcal{L}_i^\top D^{-1}\boldsymbol{\rho}_{i,\Delta/2} - \boldsymbol{\rho}_{i,\Delta/2}^\top \mathcal{L}_i^\top D^{-1}\boldsymbol{\pi} = \boldsymbol{\rho}_{i,\Delta/2}^\top \mathcal{L}_i^\top D^{-1}\boldsymbol{\rho}_{i,\Delta/2},
\end{aligned}$$

where the last equality follows from

$$\boldsymbol{\rho}_{i,\Delta/2}^\top \mathcal{L}_i^\top D^{-1}\boldsymbol{\pi} = \boldsymbol{\rho}_{i,\Delta/2}^\top \mathcal{L}_i^\top \frac{1}{\mathrm{vol}(V)}\mathbf{1}$$

and $\mathcal{L}_i^\top\mathbf{1} = D^{-1}(D - A_i)\mathbf{1} = D^{-1}\mathbf{0} = \mathbf{0}$.

Hence, we have

$$(13) = \|D^{-1/2}\mathcal{L}_i\boldsymbol{\rho}_{i,\Delta/2}\|^2 \cdot \|D^{-1/2}(\boldsymbol{\rho}_{i,\Delta/2} - \boldsymbol{\pi})\|^2 - \left((D^{-1/2}\mathcal{L}_i\boldsymbol{\rho}_{i,\Delta/2})^\top D^{-1/2}(\boldsymbol{\rho}_{i,\Delta/2} - \boldsymbol{\pi})\right)^2 \geq 0,$$

where the last inequality follows from the Cauchy-Schwarz inequality. □