Master Thesis

Stefano Huber

Local Clustering in irregular Hypergraphs.

Professor: Michael Kapralov
Supervisors: Kshiteej Sheth and Jakab Tardos

June 17, 2022

Abstract

Graph clustering is a widely studied graph problem. One of the main techniques used in order to develop such algorithms is the study of random walks: in particular, how fast random walks reach the stationary distribution in a cluster (mixing), and how much probability clusters lose during the random walk (leaking). Although this problem for graphs has been widely studied and understood, the generalization to hypergraphs (which are capable of representing higher than binary relations) remains to this point still open. In particular, it is hard to define a discrete Laplacian operator for hypergraphs, which makes discrete random walks inherently hard to study. Recently, a discrete diffusion process for regular hypergraphs with proper mixing guarantees has been described by Sheth et al. In this document, we are going to extend such result to irregular hypergraphs, and discuss whether the obtained guarantees are powerful enough in order to develop a conventional clustering algorithm. To conclude, we are briefly describing why the leaking result is in contrast hard to generalize to hypergraphs, and propose a possible direction for future works on the argument.

Contents

1 Introduction

An interesting graph problem is the one of clustering. It consists of finding a good set of vertices C which has many inner edges and it is sparsely connected with the rest of the graph. The way you measure the quality of the cluster, is by minimizing the *conductance* $\phi(C)$. A well-established method to find good clusters is using random walks: you must assume that there exists a cluster C^* with optimum (minimum) conductance $\phi(C^*) = \phi^*$. Then, if the random walk starts centered in a vertex $v_0 \in C^*$, after a few evolution steps you can take a sweep cut over the evolved probability vector (namely, a subset of vertices sorted by probability) and the resulting cluster C has a conductance $\phi(C) = \hat{\phi} \leq f(\phi^*)$, where f is some error function (usually the square root).
The usual way to prove such result is by dividing the problem of clustering into *mixing* and *leaking*: with the leaking result you can prove that after a few random step, the probability *leaking from* C is small, namely *the probability inside* $C \geq q(\phi^*)$ (where q is some function depending on the optimal conductance).
With the mixing result, you prove instead the opposite: namely, after a few iterations *the probability inside* C must be small $\leq r(\hat{\phi})$ (where r is some function depending on the output conductance of the algorithm).
This series of inequality finally allows you to relate the output conductance of the algorithm $\hat{\phi}$ with the optimal conductance of the graph ϕ^*.
Although this general approach is well studied in literature for standard graphs ([7], [1]), this fails to extend nicely to hypergraphs. In particular, it is hard to define a discrete random walk on a hypergraph, which makes both mixing and leaking results hard to prove. An interesting and encouraging approach is the one studied by Takai et al. [8]. But, although they could prove both a mixing and a leaking result for the personalized page rank vector, they use a continuous diffusion process (based on the hypergraph Laplacian explored by Chan et al. [2]) in order to compute the page rank vector (rather than a discrete one): this makes the algorithmic and numerical result somehow dependant on the approximation quality chosen (in their case, they use the Euler method in order to compute the solution of the heat equation), other than being theoretically more cumbersome to handle.
Recently, Sheth et al. managed to find a discrete process in order to perform a random walk in a d-regular hypergraph, and proved mixing using the well established Lovasz-Simonovits curve technique [5].
The goal of this project is to extend the mixing result found by Sheth et al. to irregular hypergraphs. In the remainder of the document, we are going to provide the proof for the extension of the mixing result to irregular hypergraphs in Section 3. Then we are going to discuss in Section 4 whether the result found is as powerful as the one for the d-regular case. In addition, in Section 5 we are going to prove with empirical evidence that the results obtained are in line with the theoretical guarantees. To conclude, we present in Section 6 a brief introduction of the leaking problem, which remains at the moment unsolved, and provide some insight for future works on the argument.

2 Background

2.1 Standard Graphs

A graph $G = (V, E, w)$ is a collection of vertices V, of edges $E \subseteq V \times V$, and a weight function $w(e) : E \to \mathbb{R}$ assigning a weight to every edge. We usually call $n := |V|$ and $m := |E|$. If the graph is unweighted, then we can assume that $w(e) = 1$, $\forall e \in E$. Edges are usually directed (namely, $(u, v) \neq (v, u)$). We define the degree of a node as $d(v) := \sum_{(v,u)\in E} w(v, u)$, the degree vector $\mathbf{d}$ as the vector of vertex degrees and the diagonal degree matrix $D := \mathrm{diag}(\mathbf{d})$. We define the volume of a subset of nodes $S \subseteq V$ as $\mathrm{vol}(S) := \sum_{u\in S} d(u)$. Conveniently, the volume of the graph is simply the volume of its entire vertex set: $\mathrm{vol}(G) := \mathrm{vol}(V)$. It is also convenient to define the adjacency matrix as $A := (w(u, v))_{(u,v)\in E}$. We always assume that the graphs analyzed are connected.
A lazy random walk is defined as the random walk that with probability $\frac{1}{2}$ stays still, and with probability $\frac{1}{2}$ moves from any vertex u along any of the adjacent edges w.p. $\frac{w(u,v)}{d(u)}$. It is easy to check that we can express the evolution of the random walk as

$$\mathbf{p}_{t+1} = \frac{1}{2}(I + AD^{-1})\mathbf{p}_t \tag{1}$$

We call $M = \frac{1}{2}(I + AD^{-1})$ the transition probability matrix.
It is a known result for Markov Chains that such random walks converge to the stationary distribution $\pi = \left(\frac{d(u)}{\mathrm{vol}(G)}\right)_{u\in V}$. When we are talking about mixing, we simply want to describe how fast the random walk converges to the stationary distribution: if we allow the notation $p(S) := \sum_{u\in S} p(u)$ for some $S \subseteq V$, then a good mixing result would be of the form $\forall S \subseteq V$, $|p(S) - \pi(S)| \leq f(t)$, namely the difference between the probability centered in any subset of nodes and the stationary distribution on the same subset of nodes must be some function f depending on the time (we are going to show that $f(t)$ is exponentially decreasing with respect to the time t).
To conclude the section, we are going to define the conductance ϕ: for any subset of nodes $S \subseteq V$,

$$\phi(S) := \frac{\delta(S, V \setminus S)}{\min(\mathrm{vol}(S), \mathrm{vol}(V \setminus S))} \tag{2}$$

with the cut weight δ defined as

$$\delta(S, V \setminus S) := \sum_{\substack{(u,v)\in E \text{ s.t.} \\ u\in S, v\in V\setminus S}} w(u, v) \tag{3}$$

It is easy to see why the conductance is a good measure of the quality of a cluster: when the conductance for a cut S is low, it means that only a little fraction of the edges crosses the cut and hence the cluster is well inter-connected but loosely connected with the rest of the graph. For ease of notation, we will define the bar function $\bar{S} := V \setminus S$, so that the conductance can be more easily rephrased $\phi(S) = \frac{\delta(S,\bar{S})}{\min(\mathrm{vol}(S),\mathrm{vol}(\bar{S})}$.

We can define the conductance for the entire graph $\phi(G) := \min_{S \subseteq V} \phi(S)$.
There are some special vectors for which standard notation is widely used in literature: $\mathbf{p}$ is usually a probability vector, namely a vector such that $\sum_i p(i) = 1$ and all entries $p(i) \geq 0$. χ_v is a special probability vector s.t.

$$\chi_v(u) = \begin{cases} 1, & \text{if } v = u \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

and ψ_S is another special vector, namely for some $S \subseteq V$

$$\psi_S(u) = \begin{cases} \frac{d(u)}{\text{vol}(S)}, & \text{if } u \in S \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

Notice that the vector ψ_S is nothing but the stationary distribution centered on the set of vertices S. The graph stationary distribution $\pi := \psi_V$.
To conclude, we will call the quantity $\hat{k} := \min(k, \text{vol}(G) - k)$.

2.2 Hypergraphs

A hypergraph $H = (V, E, w)$ is a generalization of a graph, where edges can be subsets of vertices, namely $E \subseteq 2^V$. We can then define the degree of a vertex u as $d(u) := \sum_{e: u \in e} w(e)$. The conductance formula does not change, but the cut weight δ does:

$$\delta(S, \bar{S}) := \sum_{e \in E: e \cap S \neq \emptyset \wedge e \cap \bar{S} \neq \emptyset} w(e) \tag{6}$$

A hypergraph is called d-regular if all edges have equal degree d. In contrast, it is called r-uniform if all hyperedges have the same size r. When the edges are unweighted (namely, they are all ones) and the hypergraph is d-regular, then it is easy to compute the volume $\text{vol}(H) = dn$. The same happens when the hypergraph is r-uniform: $\text{vol}(H) = rm$.

2.3 Mixing and Lovasz-Simonovits curve

A very useful approach in order to prove mixing is the Lovasz-Simonovits curve [5]. When dealing with the Lovasz-Simonovits curve, we always assume that the graph is undirected and we substitute any undirected edge $\{u, v\}$ with two directed edges $(u, v), (v, u)$. In order to define the curve, it is useful to define first a sweep cut.
Given a graph $G = (V, E)$ and a random walk probability vector $\mathbf{p}_t$, then you define a sweep cut $S_j(\mathbf{p}_t)$ as the set of j vertices u that maximize the ratio $\frac{p_t(u)}{d(u)}$. Notice that such ratio is nothing but the probability distributed on the outgoing edges of the vertex u.
In order to define the sweep cut, you first compute a permutation ρ of numbers in $[1, n]$, $\rho(1), \rho(2) ... \rho(n)$ s.t. $\frac{p_t(\rho(i))}{d(\rho(i))} \geq \frac{p_t(\rho(l)}{d(\rho(l))}$, $\forall l \geq i$. Notice that in case of a tie, you can solve it arbitrarily.
Then, you can simply define the sweep cut as the set of j nodes that maximize the probability on the outgoing edges:

$$S_j(\mathbf{p}_t) := \{\rho(i), \forall i \leq j\} \tag{7}$$

With this definition at hand, it is now easy to define the Lovasz-Simonovits curve: for any $j \in [1, n]$, you define $k_j = \sum_{i=1}^{j} d(\rho(i))$ as the volume of the sweep cut $S_j(\mathbf{p}_t)$ (notice that when j is clear from the context, we simply write k). Then the Lovasz-Simonovits curve $I_t : [0, \mathrm{vol}(G)] \to [0, 1]$ is such that

$$I_t(k) := \mathbf{p}_t(S_j(\mathbf{p}_t)) \tag{8}$$

We call the set of points $\{k : k = \sum_{i=1}^{j} d(\rho(i)) \text{ for some j}\}$ *hinge points*. The Lovasz-Simonovits curve is defined in between hinge points by linear interpolation: namely, for any k that is not a hinge point, we call k' the x-coordinate of the previous hinge point and k'' the x-coordinate of the next hinge point, and we can define the curve in k as:

$$I_t(k) := I_t(k') + \frac{k - k'}{k'' - k'}(I_t(k'') - I_t(k')) \tag{9}$$

It is interesting to notice a few things: first, the curve I is concave. In fact, the angular coefficient of the curve in any point k in between hinge points $\rho(j-1)$, $\rho(j)$ is simply $\frac{p_t(\rho(j))}{d(\rho(j))}$. Since we have sorted the vertices in such a way that this quantity is non-increasing, the angular coefficient is also non-increasing and hence the curve is concave. Another important aspect is that the curve in non-decreasing: in fact, between consecutive hinge points k_j and k_{j+1}, $I_t(k_{j+1}) = I_t(k_j) + p_t(\rho(j+1))$. Since the probability on any vertex is non-negative, and the curve is piece-wise linear, then it follows that it must be non-decreasing.
An important claim about the Lovasz-Simonovits curve is the following (for the proof, please see [6])

Lemma 2.1. *The Lovasz-Simonovits curve at time $t+1$ lies below the curve at time t. Namely, $\forall t$ and $\forall k \in [0, vol(H)]$*

$$I_{t+1}(k) \leq I_t(k) \tag{10}$$

So, we have seen that the Lovasz-Simonovits curve is a concave function that flattens out at every iteration step: we might wonder what happens when the probability $\mathbf{p}_t \to_{t\to\infty} \pi$: in this case the angular coefficient of the curve at any point in between hinge point j is $\frac{p_t(\rho(j))}{d(\rho(j))}$, and when $\mathbf{p}_t$ equals the stationary distribution, then this quantity is a constant:

$$\frac{p_t(\rho(j))}{d(\rho(j))} = \frac{\pi(\rho(j))}{d(\rho(j))} = \frac{\frac{d(\rho(j))}{\mathrm{vol}(G)}}{d(\rho(j)} = \frac{1}{\mathrm{vol}(G)} \tag{11}$$

Hence when the random walk converges to the stationary distribution, then the Lovasz-Simonovits curve flattens out to a straight line between $(0, 0)$ and $(\mathrm{vol}(G), 1)$.
Surprisingly, it turns out that studying how fast the Lovasz-Simonovits curve converges to a straight line is extremely convenient in order to understand how fast the random walk converges to the stationary distribution.
Hence, in order to study the mixing time of the random walk, it is enough to study the behaviour of the Lovasz-Simonovits curve. The first rule for proving mixing consists in understanding how much the Lovasz-Simonovits curve

decreases from one iteration to the other. The general case for standard lazy random walks says that any hinge point k decreases by:

$$I_{t+1}(k) \leq \frac{1}{2}(I_t(k - \phi\hat{k}) + I_t(k + \phi\hat{k})) \tag{12}$$

where ϕ is the conductance of the best sweep cut wrt the probability vector $\mathbf{p}_t$. This simply means that for every hinge point k, you can take a chord between the two symmetric points $\pm\phi\hat{k}$, and the next iteration curve will lie below such chord. Notice that since the curve is concave, then the chord lies below the hinge point. Next result for proving mixing consists in quantifying the distance between the stationary distribution and the Lovasz-Simonovits curve with respect to the time. The general rule is:

$$I_t(k) \leq \sqrt{\hat{k}} e^{-\frac{t\phi^2}{4}} + \frac{k}{\mathrm{vol}(G)} \tag{13}$$

It is easy now to see why mixing happens fast: the error that separates the Lovasz-Simonovits curve from the stationary distribution (a straight line with value $\frac{k}{\mathrm{vol}(G)}$) is exponentially decreasing with respect to the time. Hence, in order to achieve an error which is $\frac{1}{\mathrm{poly}(n)}$ for any point k, it is sufficient a number of iterations $\sqrt{\frac{\mathrm{vol}(G)}{2}} e^{-\frac{t\phi^2}{4}} \leq \frac{1}{\mathrm{poly}(n)} \implies t = O\left(\frac{\log(n)}{\phi^2}\right)$ (assuming that edge weights are mostly polynomial in n, which is the case for instance when the graph is unweighted).

2.4 Clustering algorithms for graphs

In this section, we are going to describe how clustering algorithms work in general graphs: first, let us assume that there is a cut S^* with optimum conductance ϕ^*, and volume $\leq \frac{1}{2}\mathrm{vol}(G)$. The idea (explored in [7] and [1]) then is that if you start the random walk from any vertex in $S^g \subseteq S^*$ s.t. $\mathrm{vol}(S^g) \geq \frac{1}{2}\mathrm{vol}(S^*)$, then you want to be able to find a good sweep cut S with conductance ϕ not too far from ϕ^*.
The general process is this: first, you want a leaking result in order to understand how many iterations it takes before a constant amount of probability leaks from S^*. In particular this leaking result is usually of the form: assuming that $\mathbf{p}_0 = \chi_v$ and $v \in S^g$

$$p_t(\bar{S^*}) \leq t\phi^* \tag{14}$$

Which means that after $t = \frac{1}{4\phi^*}$ we still have a probability $\geq \frac{3}{4}$ on the set S^*. This implies that

$$\left| \sum_{u \in S^*} (p_t(u) - \pi(u)) \right| \geq \frac{3}{4} - \frac{1}{2} = \frac{1}{4} \tag{15}$$

because we assumed the set S^* to have volume at most $\frac{1}{2}\mathrm{vol}(G)$.
At the same time, the mixing result ensures that, in contrast, the difference between the probability at time t and the stationary distribution should be

small, namely:

$$\left| \sum_{u \in S^*} (p_t(u) - \pi(u)) \right| \leq \sqrt{\text{vol}(S^*)} e^{-\frac{t\phi^2}{4}} \tag{16}$$

Where ϕ is the conductance of the best sweep cut on the probability vector $\mathbf{p}_t$. Combining the two results, and setting $t = \frac{1}{4\phi^*}$ we get that

$$\frac{1}{4} \leq \sqrt{\text{vol}(S^*)} e^{-\frac{t\phi^2}{4}} \implies \phi \leq O(\sqrt{\log(\text{vol}(S^*))\phi^*}) \tag{17}$$

namely, the conductance found with a sweep cut is not too far from the optimal conductance of the graph.
Notice that the leaking result only holds when the starting vertex v which determines $\mathbf{p}_0 = \chi_v$ must be such that $v \in S^g$. Hence, it is also of crucial importance to prove that the set S^g has a large volume: namely, in order to have a powerful algorithmic clustering primitive, we need to prove that $\text{vol}(S^g) \geq \frac{1}{2}\text{vol}(S^*)$. This means that when being allowed to pick the starting vertex at random according to the stationary distribution ψ_V, then the probability that we start from a vertex which is *good* (namely, which is in S^g) is high. This results in a probabilistic algorithm for clustering that the authors of [7] call `Nibble`.
Although the actual details of the algorithm `Nibble` are vastly more complicated and would need a specific section for a proper discussion (in particular, for the discussion of the running time and of the properties of the output set), notice that for the scope of our analysis it is enough to appreciate the importance of the leaking and of the mixing result, and how is it possible to combine them together to have an elementary primitive for a clustering algorithm.

2.5 Clustering algorithms for hypergraphs

For hypergraphs, clustering algorithms are more tricky: in fact, it is not possible to define an equivalent discrete diffusion process as a random walk in a graph. This is due to the fact that the Laplacian operator $\mathcal{L} = (I - AD^{-1})$, which describes how the probability mass varies from one iteration to the other $\frac{d\mathbf{p}_t}{dt} = -\mathcal{L}\mathbf{p}_t$, is not a linear operator. According to Li et al [4], the Laplacian operator for hypergraphs can be defined as

$$\mathcal{L}_H(\mathbf{x}) := \left\{ \sum_{e \in E} \mathbf{b}_e \mathbf{b}_e^T D^{-1} \mathbf{x} \mid \mathbf{b}_e \in \arg\max_{\mathbf{b} \in B_e} \mathbf{b}^T \mathbf{x} \right\} \tag{18}$$

with B_e the convex hull of $\{\chi_v - \chi_u \mid u, v \in e\}$.
The reason why this operator is somewhat hard to treat, is because there can be multiple choices of $\mathbf{b}_e$ when more than one vertex $u, v \in e$ have equal value $\frac{p_t(u)}{d(u)} = \frac{p_t(v)}{d(v)}$: when this happens, it is necessary to find the correct $\mathbf{b}_e$ which distributes the probability homogeneously among all vertices maximizing $\mathbf{b}_e^T \mathbf{x}$, also taking into account how the other edges behave. Instead, when the values $\frac{p_t(u)}{d(u)}$ are distinct for all $u \in V$, then vectors $\mathbf{b}_e$ are unique and it is easy to define the hypergraph laplacian.

Taking advantage of this operator, Takai et al [8] managed to define a personalized page rank vector $\mathbf{pr}_\alpha(\mathbf{s}) = \alpha\mathbf{s} + (1-\alpha)M\mathbf{pr}_\alpha(\mathbf{s})$ which they proved to be the solution $\mathbf{x}$ to

$$\mathbf{s} \in \left(I + \frac{1-\alpha}{2\alpha}\mathcal{L}_H\right)(\mathbf{x}) \tag{19}$$

Finally, using analogue techniques developed by Andersen et al. [1] , they could find good clusters by taking sweep cuts over the page rank vector.
The tricky problem about clustering in hypergraphs is then how to compute the personalized page rank vector: usually, you can compute the personalized page rank vector in a graph by simulating a random walk. But for hypergraphs, the page rank vector is the solution as $t \to \infty$ of the differential equation

$$\frac{d\mathbf{p}_t}{dt} \in \frac{2\alpha}{1+\alpha}(\mathbf{s} - \mathbf{p}_t) - \left(1 - \frac{2\alpha}{1+\alpha}\right)\mathcal{L}_H(\mathbf{p}_t) \tag{20}$$

Since the solution of such a differential equation cannot be computed exactly, the quality of the resulting clustering algorithm is only as good as the approximation factor chosen (in case of Takai et al. [8], they use the Euler Method).
With the issues arising from a continuous diffusion process in mind, in the next Section 3 we are going to describe a *discrete* diffusion process for hypergraphs with mixing time guarantees. The approach is an innovative idea originally developed by Sheth et al. for regular hypergraphs and yet to be published. The goal of this project is to extend it to irregular hypergraphs.

3 Mixing in non d-regular hypergraphs

In this section we are going to explore the mixing result found by Sheth et al. for d-regular hypergraphs and extend it to irregular hypergraphs: the authors have found a discrete process in order to evolve the probability vector $\mathbf{p}_t$, such that the Lovasz-Simonovits curve with respect to such a probability vector respects the mixing result:

Theorem 1. *Mixing result for d-regular hypergraphs: when I_t is the Lovasz-Simonovits curve computed using the probability vector $\mathbf{p}_t$, coming from the discrete diffusion process described in 3.1, then $\forall k \in [0, vol(H)]$, $t \in \mathbb{N}$*

$$I_t(k) \leq \sqrt{\min\left(\frac{k}{d}, \frac{vol(H)-k}{d}\right)}e^{-\frac{t}{4\phi^2}} + \frac{k}{vol(H)} \tag{21}$$

Before starting with the proof, it is important to point out that this process is of great relevance because it is a discrete one, unlike the processes known so far with bounded mixing properties in hypergraphs ([8], [2]) which are, instead, continuous. The benefit of a discrete process over a continuous one, apart from being of more immediate understanding, is that a discrete process can be simply implemented with an algorithm.
The underlying idea is as follows: at every iteration t, you collapse the hypergraph H into a multigraph G_t. On this multigraph, it is possible to perform a discrete step of size dt using the formula $\mathbf{p}_{t+dt} = ((1-dt)I + dtAD^{-1})\mathbf{p}_t$. The evolved probability vector respects the usual recursive Lovasz-Simonovits

upper bound $I_{t+dt}(k) \leq (1-2dt)I_t(k) + 2dt(\frac{1}{2}I_t(k-\phi\hat{k}) + \frac{1}{2}I_t(k+\phi\hat{k}))$ which is sufficient to achieve an exponentially fast (with respect to the time t) mixing: $I_t(k) \leq \sqrt{\hat{k}}e^{\phi^2 t} + \frac{k}{\mathrm{vol}(H)}$. In the following sections we are going to describe the details of the proof.

3.1 Preliminaries

Let $H = (V, E)$ be an hypergraph, and p_t be any probability probability vector. You can think of $\mathbf{p}_0$ as the probability vector centered in some vertex, namely $\mathbf{p}_0 = \chi_v$ for some $v \in V$. For every $e \in E$, let

$$v^t_{\max}(e) := \arg\max_{v \in e} \frac{p_t(v)}{d(v)} \tag{22}$$

$$v^t_{\min}(e) := \arg\min_{v \in e} \frac{p_t(v)}{d(v)} \tag{23}$$

where $d(v)$ is the degree of v. In case of a tie, it is important to solve it in such a way that $v^t_{\max}(e)$ is the vertex with the smallest index, and $v^t_{\min}(e)$ is the one with largest index. Notice the analogy with the hypergraph Laplacian described by Li et al in Section 2.5, in case the probability vector is made by distinct entries $\frac{p_t(u)}{d(u)}$ $\forall u \in V$. Then, the vector $\mathbf{b}_e$ has as non-zero entries exactly our $v^t_{\max}$ and $v^t_{\min}$. The discretization of our process w.r.t. the continuous one described in Section 2.5 comes precisely from the fact that we solve ties easily.
Let $G_t = (V, E_t)$ be the collapsed multigraph such that $\forall e \in E, (v^t_{\min}(e), v^t_{\max}(e)) \in E_t$. In addition, in order to preserve the vertex degree of the original hypergraph, we add enough self-loops to every $v \in V$. Notice that no vertex v can be in more than $d(v)$ edges of type $(v^t_{\min}(e), v^t_{\max}(e))$. This means that the degree $d_t(v)$ of every vertex in the collapsed graph is such that $d_t(v) \leq d(v)$. Hence, by adding the right amount of self loops $d(v) - d_t(v)$, we can always make sure that the degree of every vertex in the collapsed graph is equal to the degree of the same vertex in the original hypergraph. This also implies that $\mathrm{vol}(H) = \mathrm{vol}(G_t)$.
Let dt be a constant $\leq \frac{1}{2}$, the discrete length of the step. Let A_t be the symmetric adjacency matrix $A_t(u, v) = w(u, v)$ where $w(u, v)$ is the weight function, namely the number of equal edges (u, v) in the multigraph G_t. Let us define

$$M_t := (1 - dt)I + dtA_tD^{-1} \tag{24}$$

the random walk transition probability matrix, and finally let

$$\mathbf{p}_{t+dt} := M_t\mathbf{p}_t \tag{25}$$

be the evolved probability vector.
To conclude, we call the sweep cut $S_j(\mathbf{p}_t)$ the set of j vertices that maximize the quantity $\frac{p_t(v)}{d(v)}$. Once again, we are going to solve ties in such a way that vertices with smallest index come first.

3.2 Changing Edge Set

Let $S_j(\mathbf{p}_{t+dt})$ be a sweep cut with conductance ϕ with respect to the edge set E_{t+dt}. Because of how we solved ties in the choice of the sweep cut and in the choice of $v_{\max}^{t+dt}(e)$ and $v_{\min}^{t+dt}(e)$, it is possible to prove that also in the hypergraph H the conductance of the cut $(S_j(\mathbf{p}_{t+dt}), V \setminus S_j(\mathbf{p}_{t+dt}))$ is ϕ. In order to prove it, we will show this simple fact:

Fact 3.1.

$$e \in E \text{ crosses } S_j(\mathbf{p}_{t+dt}) \iff \text{the collapsed edge } (v_{max}^{t+dt}(e), v_{min}^{t+dt}(e)) \text{ crosses } S_j(\mathbf{p}_{t+dt}) \quad (26)$$

Notice that, as soon as the above fact is proved, then it is easy to see that the conductance of the cut $S_j(\mathbf{p}_{t+dt})$ is ϕ in both the hypergraph H and in the collapsed graph G_{t+dt}: in fact the volume of every vertex in H and in G_{t+dt} does not change by construction, and the volume of the set of crossing edges is equal, since we ensured with the fact above that there is a one-to-one correspondence between the crossing edges in H and in the collapsed graph G_{t+dt}.

Proof. We can prove the above fact by simply solving the two directions of the implication: in order to prove

$$\text{the collapsed edge } (v_{\max}^{t+dt}(e), v_{\min}^{t+dt}(e)) \text{ crosses } S_j(\mathbf{p}_{t+dt}) \implies e \in E \text{ crosses } S_j(\mathbf{p}_{t+dt}) \quad (27)$$

it is enough to notice that when the edge $(v_{\max}^{t+dt}(e), v_{\min}^{t+dt}(e))$ crosses the cut (namely, $v_{\max}^{t+dt}(e) \in S_j(\mathbf{p}_{t+dt})$ and $v_{\min}^{t+dt}(e) \in \bar{S}_j(\mathbf{p}_{t+dt})$), then since both $v_{\max}^{t+dt}(e)$ and $v_{\min}^{t+dt}(e) \in e$ by construction, then also e is cut because there are at least two nodes in opposite sides of the cut $S_j(\mathbf{p}_{t+dt})$.
For the inverse direction, we proceed with a different argument: assume ρ is the permutation of vertices that induces the sweep cut $S_j(\mathbf{p}_{t+dt})$. Then of all vertices $u \in e$, $v_{\max}^{t+dt}(e)$ is the one that appears first in ρ, and $v_{\min}^{t+dt}(e)$ is the one that appears last. Notice that this claim is enough to prove the inverse direction: when e is cut, then it means that $\exists u \in e : u \in S_j(\mathbf{p}_{t+dt})$ (hence, the position of u in the permutation ρ is smaller or equal than j), and also $\exists v \in e : v \in \bar{S}_j(\mathbf{p}_{t+dt})$ (hence the position of v in the permutation ρ is strictly larger than j). By the above claim, since $v_{\max}^{t+dt}(e)$ comes before u in ρ, and $v_{\min}^{t+dt}(e)$ comes after v in ρ, then also the collapsed edge is cut by the sweep cut $S_j(\mathbf{p}_{t+dt})$. It now remains to prove why $v_{\max}^{t+dt}(e)$ is the first vertex in e to appear in the permutation ρ (the argument for $v_{\min}^{t+dt}(e)$ to appear as the last is analogous): when values $\frac{p_{t+dt}(u)}{d(u)}$ are distinct for all $u \in e$, then $v_{\max}^{t+dt}(e)$ is the vertex $u \in e$ with highest $\frac{p_{t+dt}(u)}{d(u)}$ value by construction, hence it appears first (among the vertices in e) in the permutation ρ. The case which is more difficult to handle is in case there are multiple vertices $v, u \in e$ with equal $\frac{p_{t+dt}(u)}{d(u)} = \frac{p_{t+dt}(v)}{d(v)}$. But, since we have solved both ties in the construction of ρ and in the selection of $v_{\max}^{t+dt}(e)$ by letting first vertices with lower index, we are sure that indeed $v_{\max}^{t+dt}(e)$ is the first vertex in e to appear in ρ. A similar (but opposite) argument shows that $v_{\min}^{t+dt}(e)$ is the last vertex in e to appear in the permutation ρ. This concludes the proof.

□

Although we have just ensured that the conductance of $S_j(\mathbf{p}_{t+dt})$ is the same in the hypergraph H and in the collapsed graph G_{t+dt}, it might be possible that the conductance of $S_j(\mathbf{p}_{t+dt})$ on the edge set E_t is less than ϕ: it might happen, in fact, that for a crossing hyperedge $e \in E$, the collapsed edge $(v_{\min}^{t+dt}(e), v_{\max}^{t+dt}(e))$ crosses the bipartition but the edge $(v_{\min}^{t}(e), v_{\max}^{t}(e))$ does not. When this happens, then the following must hold:

- first, it must be true that $v_{\max}^{t+dt}(e) \in S_j(\mathbf{p}_{t+dt})$ and $v_{\min}^{t+dt}(e) \notin S_j(\mathbf{p}_{t+dt})$, because $S_j(\mathbf{p}_{t+dt})$ contains the vertices v that maximize the quantity $\frac{p_{t+dt}(v)}{d(v)}$.
- Second, since $(v_{\min}^{t}(e), v_{\max}^{t}(e))$ does not cross the bipartition, either $v_{\min}^{t+dt}(e)$, $v_{\min}^{t}(e)$ and $v_{\max}^{t}(e)$ are outside the set $S_j(\mathbf{p}_{t+dt})$ (which means that the edge $(v_{\max}^{t}(e), v_{\max}^{t+dt}(e))$ cuts the bipartition), or $v_{\max}^{t+dt}(e), v_{\min}^{t}(e)$ and $v_{\max}^{t}(e)$ are inside $S_j(\mathbf{p}_{t+dt})$, and then $(v_{\min}^{t}(e), v_{\min}^{t+dt}(e))$ cuts the bipartition.

So, we can always add to E_t another collapsed edge $(v_{\max}^{t+dt}(e), v_{\max}^{t}(e))$ or $(v_{\min}^{t+dt}(e), v_{\min}^{t}(e))$ so that the conductance of the cut $S_j(\mathbf{p}_{t+dt})$ on the edge set E_t has conductance ϕ. In order to preserve the conductance, though, we cannot alter the volume of the graph: it is enough to remove enough self loops from the vertices where we have added edges. Let us call the new edge set $\tilde{E}_t$: now, we can see that E_t, E_{t+dt} and $\tilde{E}_t$ have the same degree on every vertex, and the number of edges cutting the bipartition $S_j(\mathbf{p}_{t+dt})$ is equal in both edge sets $\tilde{E}_t$ and E_{t+dt}.

Observation 3.2. *It might happen that, when adding an edge of type $(v_{min}^{t}(e), v_{min}^{t+dt}(e))$ to $\tilde{E}_t$ (or, analogously $(v_{max}^{t}(e), v_{max}^{t+dt}(e))$), and we need to remove one self loop from both $v_{min}^{t+dt}(e)$ and from $v_{min}^{t}(e)$, there are not enough self loops to remove. For example, the vertex $u = v_{min}^{t}(e)$ is the one that minimizes the quantity $\frac{p_t(v)}{d(v)}$ $\forall v \in V$. Hence, for every hyperedge e s.t. $u \in e$, the edge $(u, v_{max}^{t}(e))$ will be added to u. This implies that u has already degree $d(u)$, without adding any self loop. Hence, if for some other $e' \in E$ we need to add the edge $(u, v_{min}^{t+dt}(e'))$, we would not have any self loop to remove from u, so that the graph volume remains unchanged. In order to solve this possible event, it is simply enough to double the degree of every node by adding self loops: this ensures that the number of self loops in every $v \in V$ is always sufficiently large. At the same time, the conductance only changes by a constant factor: since we have doubled the volume without adding any crossing edge (self loops never cross any cut), the conductance is half the original conductance, which is perfectly fine for the scope of our analysis.*

We call $\tilde{\mathbf{p}}_{t+dt}$ the evolution of $\mathbf{p}_t$ on $\tilde{E}_t$. It is possible now to claim the following lemma:

Lemma 3.3. *$\tilde{p}_{t+dt}(v) \geq p_{t+dt}(v)$ $\forall v \in S_j(\mathbf{p}_{t+dt})$, and $\tilde{p}_{t+dt}(u) \leq p_{t+dt}(u)$ $\forall u \notin S_j(\mathbf{p}_{t+dt})$.*

Proof. Let's first prove that when $v \in S_j(\mathbf{p}_{t+dt}) \implies \tilde{p}_{t+dt}(v) \geq p_{t+dt}(v)$. First, notice that the v-th coordinate of $\mathbf{p}_{t+dt} = M_t\mathbf{p}_t$ (and equivalently for $\tilde{p}_{t+dt} = \tilde{M}_t\mathbf{p}_t$) is only affected by the direct neighbors of v (the v-th row of

the adjacency matrix A_t is zero in all entries that are not direct neighbors of v). Hence, when the neighborhoods of v $\text{neigh}_{E_t}(v) = \text{neigh}_{\tilde{E}_t}(v)$, then also the value $p_{t+dt}(v) = \tilde{p}_{t+dt}(v)$. When the neighborhood of v changes, instead, it must hold that for some $e \in E$ that contains v, the edge $(v_{\min}^{t+dt}(e), v_{\max}^{t+dt}(e))$ crosses the cut, and $(v_{\min}^{t}(e), v_{\max}^{t}(e))$ does not. This further implies that it has been added to $\tilde{E}_t$ either the edge $(v_{\min}^{t}(e), v_{\min}^{t+dt}(e))$ or the edge $(v_{\max}^{t}(e), v_{\max}^{t+dt}(e))$. Let us now try to understand what vertex can v be: since we assumed that $v \in S_j(\mathbf{p}_{t+dt})$, then certainly $v \neq v_{\min}^{t+dt}(e)$; in fact, we assumed that $(v_{\min}^{t+dt}(e), v_{\max}^{t+dt}(e))$ crosses the cut, and by how we select the vertices in $S_j(\mathbf{p}_{t+dt})$, $v_{\max}^{t+dt}$ must be the vertex in the bipartition (because by definition it has higher $\frac{p_{t+dt}(v)}{d(v)}$ value). At the same time, $v \neq v_{\max}^{t}(e)$ because the added edge $(v_{\max}^{t}(e), v_{\max}^{t+dt}(e))$ would need to cross the bipartition, but both vertices by assumption would need to belong to $S_j(\mathbf{p}_{t+dt})$ and hence the edge could not be a crossing edge. Hence, the vertex v can be either $v = v_{\min}^{t}(e)$ (in which case, it would get connected with $v_{\min}^{t+dt}(e)$) or $v = v_{\max}^{t+dt}(e)$ (and it would get connected to $v_{\max}^{t}(e)$). In both cases, v gets connected in $\tilde{E}_t$ with a vertex u' that has a higher value than the self loop, namely $\frac{p_t(u')}{d(u')} \geq \frac{p_t(v)}{d(v)}$. We can now say that the value $\tilde{p}_{t+dt}(v) - p_{t+dt}(v) \geq 0$:

$$\tilde{p}_{t+dt}(v) - p_{t+dt}(v) = (\tilde{M}_t \mathbf{p}_t)(v) - (M_t \mathbf{p}_t)(v) \tag{28}$$

$$= (1 - dt)p_t(v) + dt \sum_{u \in \text{neigh}_{\tilde{E}_t}(v)} \frac{p_t(u)}{d(u)} -$$

$$(1 - dt)p_t(v) - dt \sum_{u \in \text{neigh}_{E_t}(v)} \frac{p_t(u)}{d(u)} \tag{29}$$

$$\geq \frac{p_t(u')}{d(u')} - \frac{p_t(v)}{d(v)} \tag{30}$$

$$\geq 0 \tag{31}$$

Where the $\geq$ between equations 29 and 30 comes from the fact that there might be more than one edge difference in the two neighborhoods (all of which have higher $\frac{p_t}{d}$ value than the self loops that get removed in $\tilde{E}_t$). This finally imples that $\tilde{p}_{t+dt}(v) \geq p_{t+dt}(v)$ when $v \in S_j(\mathbf{p}_{t+dt})$. The proof for the case $v \notin S_j(\mathbf{p}_{t+dt})$ is analogous: v can only be $v_{\min}^{t+dt}(e)$ or $v_{\max}^{t}(e)$, and gets connected to a vertex u' which has lower $\frac{p_t(u')}{d(u')} \leq \frac{p_t(v)}{d(v)}$ value, while losing a self loop. Hence, the value of $\tilde{p}_{t+dt}(v) \leq p_{t+dt}(v)$ when $v \notin S_j(\mathbf{p}_{t+dt})$. □

Lemma 3.3 has a nice corollary:

Corollary 2. *the sweep cut* $S_j(\mathbf{p}_{t+dt}) = S_j(\tilde{\mathbf{p}}_{t+dt})$

Proof. If $v \in S_j(\mathbf{p}_{t+dt})$, then it means that there are $n-j$ vertices $u \in \bar{S}_j(\mathbf{p}_{t+dt})$ with smaller $\frac{p_{t+dt}(u)}{d(u)} \leq \frac{p_{t+dt}(v)}{d(v)}$ value. Due to lemma 3.3, the value of $\frac{\tilde{p}_{t+dt}(v)}{d(v)} \geq \frac{p_{t+dt}(v)}{d(v)}$, and at the same time for all other entries $n-j$ vertices $u \notin S_j(\mathbf{p}_{t+dt})$, $\frac{\tilde{p}_{t+dt}(u)}{d(u)} \leq \frac{p_{t+dt}(u)}{d(u)}$. This clearly means that there are still at least $n-j$ vertices u such that $\frac{\tilde{p}_{t+dt}(v)}{d(v)} \geq \frac{\tilde{p}_{t+dt}(u)}{d(u)}$, and hence $v \in S_j(\tilde{\mathbf{p}}_{t+dt})$. □

With these two claims at hand, it is now possible to study the rate of convergence of the Lovasz-Simonovits curve I_t, I_{t+dt} and $\tilde{I}_{t+dt}$ that are respectively computed using $\mathbf{p}_t$, $\mathbf{p}_{t+dt}$ and $\tilde{\mathbf{p}}_{t+dt}$. First, we can see that if k is the number of edges incident on the cut $S_j(\mathbf{p}_{t+dt})$ in the edge set E_{t+dt}, then it must be the case that k is also the number of edges incident on $S_j(\mathbf{p}_{t+dt})$ on both edge sets E_t and $\tilde{E}_{t+dt}$: in fact, we ensured adding and removing carefully self loops that the degree of every vertex remains unchanged in all three edge sets.
With these observations at hand, we are going to claim all meaningful lemmas in order to prove that the mixing happens in logarithmic time. Proofs will be presented further in the section.

Lemma 3.4. *When $k = vol(S_j(\mathbf{p}_{t+dt}))$ then $I_{t+dt}(k) \leq \tilde{I}_{t+dt}(k)$*

Proof.

$$\begin{aligned} I_{t+dt}(k) &= p_{t+dt}(S_j(\mathbf{p}_{t+dt})) && \text{by definition of } I_{t+dt} && (32)\\ &\leq \tilde{p}_{t+dt}(S_j(\mathbf{p}_{t+dt})) && \text{Lemma 3.3} && (33)\\ &= \tilde{p}_{t+dt}(S_j(\tilde{\mathbf{p}}_{t+dt})) && \text{Corollary 2} && (34)\\ &= \tilde{I}_{t+dt}(k) && \text{definition of } \tilde{I}_{t+dt} && (35) \end{aligned}$$

□

This simply implies that if we are able to find an upper bound on the curve $\tilde{I}_{t+dt}(k)$ (for which, conveniently, we know the conductance ϕ of the sweep cut $S_j(\mathbf{p}_{t+dt})$ on the underlying collapsed graph $\tilde{E}_t$), then such upper bound also holds for the curve I_{t+dt}.
Next lemma is key to the Lovasz-Simonovits mixing result: in particular, it says that the curve $\tilde{I}_{t+dt}$ decreases point-wise with respect to the curve I_t to be below a large chord:

Lemma 3.5. *When the conductance of the sweep cut $S_j(\mathbf{p}_{t+dt})$ in the graph $\tilde{G}_t$ is ϕ (or larger), then:*

$$\tilde{I}_{t+dt}(k) \leq (1-2dt)I_t(k) + 2dt\left(\frac{1}{2}I_t(k-\phi\hat{k}) + \frac{1}{2}I_t(k+\phi\hat{k})\right) \tag{36}$$

Where $\hat{k} := \min(k, vol(H) - k)$

An immediate corollary of Lemma 3.5 and Lemma 3.4 is:

Corollary 3.

$$I_{t+dt}(k) \leq (1-2dt)I_t(k) + 2dt\left(\frac{1}{2}I_t(k-\phi\hat{k}) + \frac{1}{2}I_t(k+\phi\hat{k})\right) \tag{37}$$

The consequence of the Corollary 3 is the final mixing theorem:

Theorem 4. *If H is a hypergraph with conductance at least ϕ, then $\forall k \in [0, vol(H)]$ and $\forall t \geq 0$*

$$I_t(k) \leq \sqrt{\hat{k}}e^{-\phi^2 t} + \frac{k}{vol(H)} \tag{38}$$

With Theorem 4 at hand, it is a simple corollary to check that the mixing time is indeed exponentially fast in t:

Corollary 5. *In order to mix in hypergraph H with conductance of sweep cuts $\geq \phi$ (namely, make sure that $|\sum_{v\in S}(p_t(v)-\pi(v))| \leq \frac{1}{poly(n)}$, $\forall S \subseteq V$), we need a time $t = O\left(\frac{\log(vol(S))}{\phi^2}\right)$.*

In the next subsection, we are going to present the proof of the Lovasz-Simonovits mixing result (namely, Lemma 3.5 and Theorem 4) for non d-regular hypergraphs.

3.3 Lovasz-Simonovits for small time-steps in multigraphs

Let G be a multigraph with conductance of the sweep cut $S_j(\mathbf{p}_{t+dt})$ of size k at least ϕ (it is, indeed, the graph $\tilde{G}_t = (V, \tilde{E}_t)$), and define its transition probability matrix $M = (1-dt)I + dtAD^{-1}$. Finally, we evolve the probability vector with the rule $\mathbf{p}_{t+dt} = M\mathbf{p}_t$.
Notice that having such transition probability matrix is equivalent to assigning to each outgoing edge (we call these edges *non trivial*) from any vertex v a weight $dt\frac{p_t(v)}{d(v)}$ (so that, a dt fraction of the probability $p_t(v)$ remains on such outgoing edges), and the remaining $(1-dt)$ fraction of the probability divided into $\frac{1-dt}{dt}d(v)$ additional self loops (in this way, every self loops has the same weight as the non-trivial edges: $\frac{1-dt}{dt}d(v)\cdot dt\frac{p_t(v)}{d(v)} = (1-dt)p_t(v)$). The key point is that for every edge going out of any vertex v, both self loops and non-trivial edges have the same weight $dt\frac{p_t(v)}{d(v)}$. Notice that the edge weight is different when the starting vertex changes.
We can now prove the convergence claim:

Lemma 3.6. *When the multigraph G has expansion of the sweep cut $S_j(\mathbf{p}_{t+dt}) \geq \phi$, and $vol(S_j(\mathbf{p}_{t+dt})) = k$, then*

$$I_{t+dt}(k) \leq (1-2dt)I_t(k) + 2dt\left(\frac{1}{2}I_t(k-\phi\hat{k}) + \frac{1}{2}I_t(k+\phi\hat{k})\right) \tag{39}$$

Notice that, although the naming convention can be misleading, Lemma 3.6 is equivalent to Lemma 3.5 rather than Corollary 3: in fact, by construction, only the graph $\tilde{G}_t$ has conductance on the sweep cut $S_j(\mathbf{p}_{t+dt}) \geq \phi$.

Proof. To begin with, we can see that

$$\begin{aligned} I_{t+dt}(k) &= \sum_{u\in S_j(\mathbf{p}_{t+dt})} p_{t+dt}(u) && (40)\\ &= \sum_{(u,v)\,:\,u\in S_j(\mathbf{p}_{t+dt})} p_{t+dt}(u,v) && (41)\\ &= \sum_{(v,u):u\in S_j(\mathbf{p}_{t+dt})} p_t(v,u) \qquad \text{the reversed edges} && (42) \end{aligned}$$

Now, let us call the set of reversed edges W: we can further partition it into four sets:

$$\begin{aligned}
W_1 &:= \{(v,u) : u,v \in S_j(\mathbf{p}_{t+dt}\} \\
W_2 &:= \{(v,u) : v \notin S_j(\mathbf{p}_{t+dt}), u \in S_j(\mathbf{p}_{t+dt})\} \\
W_3 &:= \{(v,v) : v \in S_j(\mathbf{p}_{t+dt}) \wedge \mathrm{vol}(W_3) = dt \cdot k\} \\
W_4 &:= \{(v,v) : v \in S_j(\mathbf{p}_{t+dt}) \wedge \mathrm{vol}(W_4) = (1-2dt)k\}
\end{aligned}$$

Notice that self loops are partitioned in two groups: out of the edges incident to $S_j(\mathbf{p}_{t+dt})$ (with total volume k), we know that $(1-dt)k$ are self loops, and the remaining $dt \cdot k$ are "non-trivial" edges (there can be also self loops among the non-trivial edges, in this case they belong to W_1). Then, for every node $v \in S_j(\mathbf{p}_{t+dt})$ there are $d(v)$ non trivial edges, and $\frac{1-dt}{dt}d(v)$ self loops. So, for every node we put $\frac{1-2dt}{dt}d(v)$ self loops in W_4, and the remaining $d(v)$ self loops in W_3 (notice that the sum is $\frac{1-2dt}{dt}d(v) + d(v) = \frac{1-dt}{dt}d(v)$, namely the total number of self loops going out of v). Notice that here the assumption $dt \leq \frac{1}{2}$ comes into play: otherwise, the quantity $1-2dt$ would be negative.
Out of the total volume k, then, we have that $k(dt)$ resides in non-trivial edges $(W_1 \cup W_2)$, $k(dt)$ of it is in self loops W_3 and the remaining $(1-2dt)k$ is in self loops in W_4. Also notice that all edges going out of the same vertex u have the same weight $\frac{p_t(u)}{d(u)}dt$.
In the remaining of the proof, we are going to prove that:

$$p_t(W_4) \leq (1-2dt)I_t(k) \tag{43}$$

$$p_t(W_1) \leq dtI_t\left(k - \frac{\mathrm{vol}(W_2)}{dt}\right) \tag{44}$$

$$p_t(W_2 \cup W_3) \leq dtI_t\left(k + \frac{\mathrm{vol}(W_2)}{dt}\right) \tag{45}$$

To understand why proving Equations 43, 44 and 45 is enough to conclude the theorem, it is sufficient to notice that according to Equation 42 we get that

$$I_{t+dt}(k) = \sum_{i=1}^{k} p_t(v_i,u_i) \tag{46}$$

$$= p_t(W_1) + p_t(W_2 \cup W_3) + p_t(W_4) \tag{47}$$

$$\leq dtI_t\left(k - \frac{\mathrm{vol}(W_2)}{dt}\right) + dtI_t\left(k + \frac{\mathrm{vol}(W_2)}{dt}\right) + (1-2dt)I_t(k) \tag{48}$$

Finally, due to the fact that $\mathrm{vol}(W_2)$ is nothing but the volume of the edges cutting the sweep cut $S_j(\mathbf{p}_{t+dt})$, we know by the hypothesis about the conductance of the sweep cut being $\geq \phi$ that among the kdt volume of non-trivial edges, a fraction of the volume $\geq \phi\hat{k}dt$ is made of crossing edges, where $\hat{k} := \min(k, \mathrm{vol}(G)-k)$: hence $\frac{\mathrm{vol}(W_2)}{dt} \geq \hat{k}\phi$. To conclude, we take advantage of the concavity of the curve I_t: in fact $I_t(x+y) + I_t(x-y) \leq I_t(x+z) + I_t(x-z)$ when $z \leq y$. This observation finally yields:

$$I_{t+dt}(k) \leq dtI_t\left(k - \frac{\mathrm{vol}(W_2)}{dt}\right) + dtI_t\left(k + \frac{\mathrm{vol}(W_2)}{dt}\right) + (1-2dt)I_t(k) \quad (49)$$

$$\leq dtI_t\left(k - \phi\hat{k}\right) + dtI_t\left(k + \phi\hat{k}\right) + (1-2dt)I_t(k) \quad (50)$$

It is left to prove the three bounds in Equations 43, 44 and 45. We start with Equation 43: we know that by construction, every vertex has a fraction $(1-2dt)$ of its weight in self loops that belong to W_4:

$$p_t(W_4) = \sum_{v \in S_j(\mathbf{p}_{t+dt}} p_t(v)(1-2dt) \quad (51)$$

$$= (1-2dt) \sum_{v \in S_j(\mathbf{p}_{t+dt}} p_t(v) \quad (52)$$

$$= (1-2dt)I_t(k) \qquad \mathrm{vol}(S_j(\mathbf{p}_{t+dt})) = k \quad (53)$$

We now prove Equation 44: $p_t(W_1) \leq dtI_t\left(k - \frac{\mathrm{vol}(W_2)}{dt}\right)$: first, notice that $\mathrm{vol}(W_1) + \mathrm{vol}(W_2) = dt \cdot k$. Moreover, for every node v there are at most $d(v)$ edges belonging to W_1. Since the number of self loops in every node is $\frac{1-dt}{dt}d(v)$, this means that for every edge in W_1 there are at least $\frac{1-dt}{dt}$ self loops starting from the same vertex (and hence with the same weight $\frac{p_t(v)dt}{d(v)}$). Calling W_1^{charge} this set of self loops, then we know that $p_t(W_1^{\mathrm{charge}}) = \frac{1-dt}{dt}p_t(W_1)$ (because for every edge in W_1 we have $\frac{1-dt}{dt}$ many self loops with the same weight, which further means that

$$p_t(W_1 \cup W_1^{\mathrm{charge}}) = p_t(W_1) + \frac{1-dt}{dt}p_t(W_1) \quad (54)$$

$$= \frac{1}{dt}p_t(W_1) \quad (55)$$

Which implies that

$$p_t(W_1) = dtp_t(W_1 \cup W_1^{\mathrm{charge}}) \quad (56)$$

$$\leq dtI_t(\mathrm{vol}(W_1) + \mathrm{vol}(W_1^{\mathrm{charge}})) \quad (57)$$

$$= dtI_t(dt \cdot k - \mathrm{vol}(W_2) + \frac{1-dt}{dt}(dt \cdot k - \mathrm{vol}(W_2))) \quad (58)$$

$$= dtI_t\left(k - \frac{\mathrm{vol}(W_2)}{dt}\right) \quad (59)$$

To conclude, let us prove the last Equation 45: we use the same trick, and try to find some duplicate edges with equal weights as the ones in $W_2 \cup W_3$. In particular, we know for sure that every node $v \in S_j(\mathbf{p}_{t+dt})$ has a fraction of edges dt as self loops in W_3, and a fraction of self loops $(1-2dt)$ in W_4 with the same weight. This means that for every self loop in W_3 there are $\frac{1-2dt}{dt}$ self loops in W_4 with equal weight (in fact, the volume of W_3 times $\frac{1-2dt}{dt}$ is $k \cdot dt\frac{1-2dt}{dt} = k(1-2dt)$ which is the volume of W_4). In addition, since the volume of W_3 is dtk, which is the same as the volume of $W_1 \cup W_2$, then we know for sure that every self loop in W_3 also has at least one corresponding

"non-trivial" edge with equal weight. We add both self loops and non-trivial edges to W_3^{charge}, which has weight $|W_3|\left(\frac{1-2dt}{dt}+1\right) = |W_3|\left(\frac{1-dt}{dt}\right)$. Regarding W_2, instead, since all edges of W_2 start outside $S_j(\mathbf{p}_{t+dt})$, then we know for sure that for every edge in W_2 there is a number of self loops as large as $\frac{1-dt}{dt}$, which certainly is completely disjoint from the self loops in W_3^{charge} due to the fact that all self loops in W_3^{charge} are instead inside $S_j(\mathbf{p}_{t+dt})$. Let us call this set of self loops W_2^{charge}. This leads us to the result:

$$\text{vol}(W_3 \cup W_3^{\text{charge}}) = \text{vol}(W_3)\left(1+\frac{1-dt}{dt}\right) = \frac{1}{dt}\text{vol}(W_3) \tag{60}$$

$$\text{vol}(W_2 \cup W_2^{\text{charge}}) = \text{vol}(W_2)\left(1+\frac{1-dt}{dt}\right) = \frac{1}{dt}\text{vol}(W_2) \tag{61}$$

And hence

$$\begin{aligned} p_t(W_2 \cup W_3) &= \frac{1}{dt}p_t(W_2 \cup W_2^{\text{charge}} \cup W_3 \cup W_3^{\text{charge}}) && (62)\\ &\leq \frac{1}{dt}I_t(\text{vol}(W_2 \cup W_2^{\text{charge}}) + \text{vol}(W_3 \cup W_3^{\text{charge}})) && (63)\\ &= \frac{1}{dt}I_t\left(\frac{1}{dt}\text{vol}(W_2) + \frac{1}{dt}\text{vol}(W_3)\right) && (64)\\ &= \frac{1}{dt}I_t\left(\frac{1}{dt}\text{vol}(W_2) + k\right) && (65)\\ & && (66) \end{aligned}$$

Where the last inequality follows from the fact that $\text{vol}(W_3) = dt \cdot k$. □

With this recursive upper bound at hand, we are able to prove the mixing result:

Lemma 3.7. *Let's define the function $R_t(k)$ as follows:*

$$R_0(k) = \min(\sqrt{k}, \sqrt{vol(G)-k}) \tag{67}$$

$$R_{t+dt}(k) = 2dt\left(\frac{1}{2}R_t(k-\phi\hat{k}) + \frac{1}{2}R_t(k+\phi\hat{k})\right) + (1-2dt)R_t(k) \tag{68}$$

Then $\forall k$

$$I_t(k) \leq R_t(k) \tag{69}$$

and

$$R_t(k) \leq \sqrt{\hat{k}}e^{-\frac{t\phi^2}{4}} + \frac{k}{vol(G)} \tag{70}$$

This final lemma is what is needed to conclude mixing in logarithmic time for non d-regular hypergraphs. Notice that, in the case of d-regular hypergraphs, we can re-define $R_0(k) = \min\left(\sqrt{\frac{k}{d}}, \sqrt{\frac{\text{vol}(H)-k}{d}}\right)$ in order to achieve the desired result as in Theorem 1. For irregular hypergraphs, when the starting probability vector is χ_v for some $v \in V$, then we can also use a similar bound using instead of d the quantity $d(v)$, as it will be described later in Section 4.2.2.
Let us now explain the proof.

Proof. It is easy to check by induction that $I_t(k) \leq R_t(k)$, $\forall k \in [0, \mathrm{vol}(G)]$. To see the base case, $R_0(k) \geq 1$ when $k \geq 1$, whereas $I_t(k) \leq 1$ $\forall k, t$ by its own definition. The only thing to check is when $k \leq 1$: in such case, we know that the angular coefficient of $I_t(k)$ cannot be larger than 1 (when the probability vector p_0 is concentrated in a vertex with degree 1), and the square root function lies above $y = x$ in the interval $[0, 1]$. For the inductive case, it is enough to see that the definition of R makes it trivial to achieve the desired upper bound:

$$I_{t+dt}(k) \leq (1 - 2dt) I_t(k) + dt \left(\frac{1}{2} I_t(k - \phi\hat{k}) + \frac{1}{2} I_t(k + \phi\hat{k}) \right) \qquad \text{Lemma 3.6} \tag{71}$$

$$\leq (1 - 2dt) R_t(k) + dt \left(\frac{1}{2} R_t(k - \phi\hat{k}) + \frac{1}{2} R_t(k + \phi\hat{k}) \right) \quad \text{induction} \tag{72}$$

$$= R_{t+dt}(k) \tag{73}$$

In order to prove that $R_t(k) \leq \sqrt{\hat{k}} e^{-\frac{\phi^2 t}{4}} + \frac{k}{\mathrm{vol}(G)}$, we proceed once again by induction: when $t = 0$, then $R_0(k) = \sqrt{\hat{k}} \leq \sqrt{\hat{k}} e^0 + \frac{k}{\mathrm{vol}(G)}$. For the inductive case:

$$R_{t+dt}(k) = 2dt \left(\frac{1}{2} R_t(k - \phi\hat{k}) + \frac{1}{2} R_t(k + \phi\hat{k}) \right) + (1 - 2dt) R_t(k) \tag{74}$$

$$\leq 2dt \left(\frac{1}{2} \left[\sqrt{g(k - \hat{k}\phi)} e^{-\frac{t\phi^2}{4}} \right] + \frac{1}{2} \left[\sqrt{g(k + \phi\hat{k})} e^{-\frac{\phi^2 t}{4}} \right] \right) + \frac{2dt \cdot k}{\mathrm{vol}(G)} + (1 - 2dt) \left[\sqrt{\hat{k}} e^{-\phi^2 t} + \frac{k}{\mathrm{vol}(G)} \right] \tag{75}$$

Where the function $g()$ is nothing but the hat function: $g(x) := \min(x, \mathrm{vol}(G) - x)$. A short analysis allows us to conclude that

$$\sqrt{g(k - \phi\hat{k})} + \sqrt{g(k + \phi\hat{k})} \leq \sqrt{\hat{k}} \left(\sqrt{1 - \phi} + \sqrt{1 + \phi} \right) \tag{76}$$

We proceed to prove the inequality 76 by cases:

- $k + \phi\hat{k} \leq \frac{1}{2}\mathrm{vol}(G)$: in this case, $\hat{k} = k$ and we have that

$$\sqrt{g(k - \phi\hat{k})} + \sqrt{g(k + \phi\hat{k})} = \sqrt{\hat{k} - \phi\hat{k}} + \sqrt{\hat{k} + \phi\hat{k}} \tag{77}$$

- $k \leq \frac{1}{2}\mathrm{vol}(G) \wedge k + \hat{k}\phi \geq \frac{1}{2}\mathrm{vol}(G)$: Also here it holds that $k = \hat{k}$, and

$$\sqrt{g(k - \phi\hat{k})} + \sqrt{g(k + \phi\hat{k})} = \sqrt{\hat{k} - \phi\hat{k}} + \sqrt{g(\hat{k} + \phi\hat{k})} \tag{78}$$

$$= \sqrt{\hat{k} - \phi\hat{k}} + \sqrt{\mathrm{vol}(G) - (\hat{k} + \phi\hat{k})} \tag{79}$$

$$\leq \sqrt{\hat{k} - \phi\hat{k}} + \sqrt{\hat{k} + \phi\hat{k}} \tag{80}$$

.

- $k \geq \frac{1}{2}\text{vol}(G) \wedge k - \phi\hat{k} \leq \frac{1}{2}\text{vol}(G)$: $\hat{k} = \text{vol}(G) - k$ and it must hold that:

$$\sqrt{g(k-\phi\hat{k})} + \sqrt{g(k+\phi\hat{k})} = \sqrt{k - \hat{k}\phi} + \sqrt{\text{vol}(G) - k - \hat{k}\phi} \tag{81}$$

$$= \sqrt{k - \hat{k}\phi} + \sqrt{\hat{k} - \phi\hat{k}} \tag{82}$$

$$\leq \sqrt{\hat{k} + \phi\hat{k}} + \sqrt{\hat{k} - \phi\hat{k}} \tag{83}$$

  In fact $k - \hat{k}\phi \leq \frac{1}{2}\text{vol}(G)$ and $\hat{k} + \phi\hat{k} = (\text{vol}(G) - k) + \phi\hat{k} = \text{vol}(G) - (k - \phi\hat{k}) \geq \frac{1}{2}\text{vol}(G)$.

- $k \geq \frac{1}{2}\text{vol}(G) \wedge k - \phi\hat{k} \geq \frac{1}{2}\text{vol}(G)$: then $\hat{k} = \text{vol}(G) - k$ and

$$\sqrt{g(k-\phi\hat{k})} + \sqrt{g(k+\phi\hat{k})} = \sqrt{\text{vol}(G) - (k - \phi\hat{k})} + \sqrt{\text{vol}(G) - (k + \phi\hat{k})} \tag{84}$$

$$= \sqrt{\hat{k} + \hat{k}\phi} + \sqrt{\hat{k} - \hat{k}\phi}. \tag{85}$$

With the claim $\sqrt{g(k-\phi\hat{k})} + \sqrt{g(k+\phi\hat{k})} \leq \sqrt{\hat{k}}(\sqrt{1-\phi} + \sqrt{1+\phi})$ at hand, we can continue the inequality 75 (also taking advantage of the Taylor expansion $\frac{1}{2}(\sqrt{1-\phi} + \sqrt{1+\phi}) \leq \left(1 - \frac{\phi^2}{8}\right)$) which finally yields:

$$R_{t+dt}(k) \leq 2dt \left(\frac{1}{2} \left[\sqrt{g(k - \hat{k}\phi)} e^{-\frac{t\phi^2}{4}} \right] + \frac{1}{2} \left[\sqrt{g(k + \phi\hat{k})} e^{-\frac{\phi^2 t}{4}} \right] \right) + \frac{2dt \cdot k}{\text{vol}(G)} +$$
$$(1 - 2dt) \left[\sqrt{\hat{k}} e^{-\phi^2 t} + \frac{k}{\text{vol}(G)} \right] \tag{86}$$

$$\leq 2dt \left(\sqrt{\hat{k}} \left(\frac{1}{2}\sqrt{1-\phi} + \frac{1}{2}\sqrt{1+\phi} \right) \right) e^{\frac{-t\phi^2}{4}} +$$
$$(1 - 2dt)\sqrt{\hat{k}} e^{-\frac{\phi^2 t}{4}} + \frac{k}{\text{vol}(G)} \tag{87}$$

$$\leq 2dt \left(\sqrt{\hat{k}} \left(1 - \frac{\phi^2}{8} \right) \right) e^{-\frac{t\phi^2}{4}} + (1 - 2dt)\sqrt{\hat{k}} e^{-\frac{\phi^2 t}{4}} + \frac{k}{\text{vol}(G)} \tag{88}$$

$$= \sqrt{\hat{k}} e^{-\frac{t\phi^2}{4}} \left(1 - \frac{dt\phi^2}{4} \right) + \frac{k}{\text{vol}(G)} \tag{89}$$

$$\leq \sqrt{\hat{k}} e^{-\frac{\phi^2(t+dt)}{4}} + \frac{k}{\text{vol}(G)} \tag{90}$$

As desired. □

4 Discussion of mixing result

4.1 Is this result enough for logarithmic mixing time?

In this section, we are going to discuss the mixing result found for non d-regular hypergraphs (Theorem 4) and try to understand if it is enough to achieve a mixing time $O\left(\frac{\log(n)}{\phi^2}\right)$ as in the case of regular hypergraphs.

4.1.1 Difference with d-regular mixing result

To begin with, we are going to state back to back the two mixing results: first the one for non d-regular hypergraphs:

$$I_t(k) \leq \sqrt{\hat{k}} e^{-\frac{\phi^2 t}{4}} + \frac{k}{\mathrm{vol}(G)} \tag{91}$$

And second, the one for d-regular hypergraphs:

$$I_t(k) \leq \sqrt{\frac{\hat{k}}{d}} e^{-\frac{\phi^2 t}{4}} + \frac{k}{\mathrm{vol}(G)} \tag{92}$$

Recall that $\hat{k}$ can be as large as $\frac{1}{2}\mathrm{vol}(H)$. The only clear difference among the two results is the d factor: the d regularity factor, in fact, is not a thing for non d-regular hypergraphs, and hence it is perfectly reasonable to think that it somehow disappears in the generalized version of the theorem. Moreover, in the Lovasz-Simonovits mixing result for standard non d-regular graphs, the convergence result is perfectly equivalent to our generalization to non d-regular hypergraphs. What we may ask is: does this difference affect the quality of the result?
If we limit our analysis to standard graphs, we can see that the bound $I_t(k) \leq \sqrt{\hat{k}} e^{-\frac{\phi^2 t}{4}} + \frac{k}{\mathrm{vol}(G)}$ is perfectly fine for mixing time in $O\left(\frac{\log(n)}{\phi^2}\right)$: in fact, the volume of G (and hence $\hat{k}$) can be only as large as $\binom{n}{2} \leq n^2$. So, if we want to limit the error of the Lovasz-Simonovits curve to any number $O\left(\frac{1}{\mathrm{poly}(n)}\right)$, it is sufficient to have a number of iterations:

$$\sqrt{\hat{k}} e^{\frac{-t\phi^2}{4}} \leq \sqrt{n^2} e^{\frac{-t\phi^2}{4}} \leq \mathrm{poly}\left(\frac{1}{n}\right) \implies e^{-\frac{t\phi^2}{4}} \leq \frac{1}{n^2} \tag{93}$$

$$\implies t = O\left(\frac{\log(n)}{\phi^2}\right) \tag{94}$$

But if we turn our attention to hypergraphs, then the volume of the hypergraph (and also of the collapsed multigraph) can be as large as $n2^{n-1}$ (when all hyperedges are present, every node appears in exactly 2^{n-1} hyperedges). This yields a mixing time of:

$$\sqrt{\hat{k}} e^{\frac{-t\phi^2}{4}} \leq \sqrt{n2^n} e^{\frac{-t\phi^2}{4}} \leq \mathrm{poly}\left(\frac{1}{n}\right) \implies e^{-\frac{t\phi^2}{4}} \leq \frac{1}{n\sqrt{n2^n}} \tag{95}$$

$$\implies t = O\left(\frac{\log(n2^n) + \log(n)}{\phi^2}\right) = O\left(\frac{n}{\phi^2}\right) \tag{96}$$

which is now way worse than the result found for graphs.
Notice that, instead, the mixing result for d-regular hypergraphs has the same logarithmic complexity as the standard graph mixing result: in fact, regardless of the total volume of the graph (which, recall, can be as large as $O(2^n)$), the

following relationship binds the volume of the graph and the degree in any d-regular hypergraph:

$$\text{vol}(H) = nd \tag{97}$$

This claim ensures that the convergence result of Theorem 1

$$I_t(k) \leq \sqrt{\frac{\hat{k}}{d}} e^{-\frac{t\phi^2}{4}} + \frac{k}{\text{vol}(G)} \tag{98}$$

is enough to have $O\left(\frac{\log(n)}{\phi^2}\right)$ mixing time: in fact,
$\sqrt{\frac{\hat{k}}{d}} \leq \sqrt{\frac{\text{vol}(G)}{d}} \leq \sqrt{n}$, so that with $t = O\left(\frac{\log(n)}{\phi^2}\right)$ you can mix with an error up to $\frac{1}{\text{poly}(n)}$.
This little difference in the mixing result has some undesired and unexpected consequences: for standard graphs, the main parameter that affects the mixing time is the conductance. In fact, when the conductance is large (like a constant), then you can expect to mix very fast in $t = O(\log(n))$. Whenever instead the conductance is low (like a path graph with conductance $O\left(\frac{1}{n}\right)$), then you cannot hope to mix in a time which is smaller than n (it is, indeed, $t = O(\log(n)n^2)$). Although this principle applies for d-regular hypergraphs as well, at least by looking at Theorem 4 it is not possible to conclude the same also for irregular hypergraphs: what the theorem suggests is that it might happen that a hypergraph has very large conductance ($\phi = O(1)$), but still a mixing time $O(n)$ because the volume of the hypergraph is extremely large.
The main problem about having such a high mixing time for a clustering algorithm, is that the guarantee on the output conductance inevitably gets worse: in fact, if having a mixing time $O\left(\frac{\log(n)}{\phi^2}\right)$ ensures that the output conductance is at most $\leq \sqrt{\log(n)\phi^*}$ (with ϕ^* the optimal conductance), then having a mixing time $O\left(\frac{n}{\phi^2}\right)$ allows just a poor $\leq \sqrt{n\phi^*}$ approximation, which is extremely worse than the approximation for graphs and regular hypergraphs when n gets large.
In the following sections we are going to discover whether this is actually the case: can there be a hypergraph with high conductance and large volume, such that even if the conductance is a large constant, the mixing time is still $O(n)$?
We start with a simple observation: notice that Theorem 4 manages to find a bound on the mixing time using the collapsed multigraphs G_t's. Hence, in order to prove that it is impossible to mix in time $O(n)$ when the conductance is a large constant, it might be tempting to prove such claim in multigraphs (which are easier to handle than hypergraphs).
In fact, if we can say that no such multigraph with high conductance and exponential volume exists, then it is also not possible to collapse a hypergraph into a multigraph with such undesirable characteristics.
Unfortunately, we will show in next Section 4.1.2 that such a multigraph with high conductance and exponential volume indeed exists.

4.1.2 An example of a multigraph with high conductance and O(n) mixing time.

In this section, we are going to provide an example of a multigraph that has high constant conductance $\frac{1}{2}$, but still has a mixing time $\geq n$ for some special starting probability vectors $\mathbf{p}_0$. Here is the example: assume we have n vertices like in a path, but allowing repetitions of the edges. In particular, v_1 (the leftmost node) has degree 1 and is connected to v_2. In turns v_2 has an outgoing edge with the same weight as all the edges coming before it (in this case, only 1). v_3 then has one incoming edge from v_2, and one outgoing edge going to v_4 with weight 2 (the sum of the edge weights coming before it).
In particular, the following rule applies in order to describe the weight between two consecutive vertices $a(i, i+1)$:

$$a(i, i+1) = \sum_{j=1}^{i-1} a(j, j+1) \tag{99}$$

namely, the weight of the edge going from i to $i+1$ is the sum of all weights of the edges coming before i (including the weight of the edge $(i-1, i)$). I.e the weights of the edges are: 1, 1, 2, 4, 8, 16 ...
We will address this special graph as the *path graph with exponential edge weights*.
We are going to prove this fact regarding the path graph with exponential edge weights:

Claim 4.1. *If G is a path graph with exponential edge weights, then the conductance is constant ($\frac{1}{2}$), but the mixing time is $O(n)$.*

In order to bound the mixing time, we can see that the total time to reach v_n, when starting from vertex v_1, is n. Moreover, the last vertex has a stationary probability which is approximately $\frac{1}{4}$ (in fact, it is incident to half the edges in G, and the total volume is $2|E|$). So, if we start in v_1 with probability $p_0(1) = 1$, it is impossible to carry a constant amount of probability to a vertex which is at distance n, in less then n steps. At the same time, we will show that the total volume is indeed 2^n-ish and that the conductance is a constant, $\frac{1}{2}$.
Let's first prove that the volume of the graph is 2^n: in particular, the volume of the graph is simply twice the sum of the weights of the edges:

$$\begin{aligned}
\mathrm{vol}(G) &= 2 \sum_{i=1}^{n-1} a(i, i+1) && (100)\\
&= 2 \left(1 + \sum_{i=2}^{n-1} 2^{i-2} \right) && (101)\\
&= 2 \left(1 + \sum_{i=0}^{n-3} 2^{i} \right) && (102)\\
&= 2 \left(1 + \frac{1 - 2^{n-2}}{1-2} \right) && (103)\\
&= 2^{n-1} && (104)
\end{aligned}$$

It is now easy to see that the conductance of every sweep cut $S_k = v_1, v_2, ..., v_k$ is $\frac{1}{2}$. The reason why is because the edge (v_k, v_{k+1}), which is the only edge that is crossing the cut, has weight exactly equal to $\sum_{i=1}^{k-1} a(i, i+1)$ by construction, which means that the volume of $S_k = \sum_{i=1}^{k} a(i, i+1) = \sum_{i=1}^{k-1} a(i, i+1) + a(k, k+1) = 2a(k, k+1)$, hence only twice the volume of the cut. Now it remains to prove that no other cut can have a conductance which is lower than $\frac{1}{2}$.

Lemma 4.2. *In the path graph G with increasingly exponential edge weights (as described in the above paragraph), no cut has a conductance lower than the sweep cuts.*

The above lemma has the following direct corollary:

Corollary 6. *The path graph G has conductance $\frac{1}{2}$.*

Proof. We will proceed by contradiction. In particular, assume that the cut is made of $k > 1$ edges, like $e_{\pi_1}, e_{\pi_2}..., e_{\pi_k}$ and such that e_{π_i} is an edge of the form (v_j, v_{j+1}) for some j. Assume that π_k is the edge with highest index j^*. Then, the following scenario can happen: if $j^* = n-1$ (namely, we are cutting the last edge of weight 2^{n-3}), then clearly the smallest bipartition must be smaller than half the volume of the entire graph, namely $\min(vol(S), \text{vol}(V \setminus S)) \leq 2^{n-2}$, and hence the conductance must be $\geq \frac{1}{2}$. Notice that the last edge has been treated carefully because it might have been a bit tricky: the last node does not have the highest degree. The highest degree node is, in fact, the penultimate one. When, instead, j^* is not $n-1$, then clearly the largest bipartition is the one on the right (because it contains the last two nodes that have volume $> \frac{1}{2}\text{vol}(G)$). Hence, given this edge cut, the highest conductance value we can achieve for this situation is by not partitioning further the left bipartition (because it would increase the number of edges crossing, and reduce the value of the volume of the left bipartition). But we have already proved that the conductance of every cut of one edge is $\frac{1}{2}$, hence we are done. □

This example, unfortunately, is saying that we cannot hope to mix in logarithmic time in any multigraph with high conductance. Hence, if we collapse our input hypergraph into a multigraph with the characteristics of the path graph just described, we might have a slow mixing time, even though the conductance is high.
What is left to be proved, is than that it is actually impossible to collapse a hypergraph into a multigraph with very high conductance and high volume: in the next Section we are actually going to prove that there is no hypergraph with high conductance (constant) and high volume (exponential). Hence, this is going to allow us to conclude that the mixing time also for irregular hypergraphs is mainly determined by the conductance, as it is the case for general graphs.

4.2 Some solutions

In this section we are going to discuss when the mixing result found for irregular hypergraphs can still be useful: in particular, in Section 4.2.1 we are going to prove that the mixing theorem found for irregular hypergraphs is actually almost equivalent to the one for graphs, when the conductance is high $\gg \frac{1}{n}$: in fact,

we will show that there cannot be any hypergraph with high conductance and exponential volume, which means that as long as the conductance is high $\gg \frac{1}{n}$, then we are certain to mix fast $O\left(\frac{\log(n)}{\phi^3}\right)$ even in irregular hypergraphs.
Moreover, in Section 4.2.2 we show that the mixing time is tightly related to the degree of the starting vertex: hence, when we can choose the starting vertex at random according to the stationary distribution, as in the case for most probabilistic clustering algorithms ([7]), then we can expect to have a fast mixing time $\frac{\log(n)}{\phi^2}$ rather than the mixing time showed by Theorem 4.

4.2.1 High conductance implies low volume

In this section we are going to prove that if a hypergraph has high conductance, then it must have low volume ($\ll 2^n$): assume that the graph has high conductance $\phi \gg \frac{1}{n}$. Then, we are going to discuss that the average size of the edge cannot be too large: in particular, if we call $\hat{r}$ the average size of the hyperedges, then the following bound must hold: $\hat{r} \leq \frac{3}{\phi}$. Notice that it is the same bound as in the case of r-uniform hypergraphs, but with $\hat{r} := \frac{1}{m}\sum_{i=1}^{m} |e_i|$, namely the average hyperedge size. This allows us to conclude that the volume of the hypergraph cannot be larger than a poly factor of n which gets handled by the logarithmic mixing time in Theorem 4.
We will prove the following lemma (very similar to the r-uniform hypergraph case)

Lemma 4.3. *If the hypergraph H (with no self loops) has high conductance $\phi \gg poly\left(\frac{1}{n}\right)$, the mixing time is low: $t = O\left(\frac{\log(n)}{\phi^3}\right)$.*

We will prove this with the following argument: first, we will say that the average hyperedge size and the conductance are related.

Lemma 4.4. *If the hypergraph (with no self loops) has conductance ϕ, then the average size of the hyperedge $\hat{r} := \frac{1}{m}\sum_{i=1}^{m} |e_i|$ is s.t. $\hat{r} \leq \frac{3}{\phi}$.*

With this lemma at hand, we can say that when $\phi \gg \frac{1}{n}$ then

Lemma 4.5. *The hypergraph with average hyperedge size $\hat{r}$ cannot have volume larger than $2\hat{r}^2 n^{\hat{r}}$.*

These facts combined allow us to conclude Lemma 4.3: in fact when knowing that the volume of the hypergraph is $\leq 2\hat{r}^2 n^{\hat{r}} \leq 2\frac{9}{\phi^2} n^{\frac{3}{\phi}}$, and since the mixing time according to Lemma 3.7 is

$$t \leq \frac{\log(\mathrm{vol}(H))}{\phi^2} \tag{105}$$

$$\leq \frac{\log(2\frac{9}{\phi^2} n^{\frac{3}{\phi}})}{\phi^2} \tag{106}$$

$$\leq O\left(\frac{\log(n)}{\phi^3} + \frac{2\log\left(\frac{1}{\phi}\right)}{\phi^2}\right) \qquad \frac{1}{\phi} \ll n \tag{107}$$

$$\leq O\left(\frac{\log(n)}{\phi^3}\right) \tag{108}$$

Now it only needs to be proved Lemma 4.4 and Lemma 4.5. We start with Lemma 4.4

Proof. The proof of Lemma 4.4 is similar to the proof for the r-uniform case: in particular using the probabilistic method, we are going to claim that

$$\mathbb{P}\left(\left\{\phi(S) \leq \frac{3}{\hat{r}}\right\}\right) \geq \tag{109}$$

$$\mathbb{P}\left(\left\{\phi(S) \leq \frac{3}{\hat{r}}\right\} \mid \left\{\mathrm{vol}(S) \in \left[\frac{\mathrm{vol}(H)}{3}, \frac{2\mathrm{vol}(H)}{3}\right]\right\}\right) \mathbb{P}\left(\left\{\mathrm{vol}(S) \in \left[\frac{\mathrm{vol}(H)}{3}, \frac{2\mathrm{vol}(H)}{3}\right]\right\}\right) > 0 \tag{110}$$

We are going to say that $\mathbb{P}\left(\left\{\mathrm{vol}(S) \in \left[\frac{\mathrm{vol}(H)}{3}, \frac{2\mathrm{vol}(H)}{3}\right]\right\}\right) > 0$ with this easy argument: assume that there is a node with degree $> \frac{2}{3}\mathrm{vol}(H)$. Then it would be impossible to have a set S with the desired volume: we are claiming that such vertex with very high degree cannot exist. In fact, if $d(v_i) > \frac{2}{3}\mathrm{vol}(H) = \frac{2}{3}m\hat{r} \geq m$ (assuming that $\hat{r} \geq 2$ since all hyperedges should have at least two nodes in them, if we do not allow self-loops). But, no node can have degree strictly larger than the number of edges. If we assume that there is a node with degree $\in \left[\frac{\mathrm{vol}(H)}{3}, \frac{2\mathrm{vol}(H)}{3}\right]$, then we are done (we can achieve a set S with the desired volume by simply adding the single node to it). To conclude if all nodes have degree $\leq \frac{1}{3}\mathrm{vol}(G)$, then it means that by adding nodes one by one to S, we cannot overshoot from $\frac{1}{3}\mathrm{vol}(H)$ to $\frac{2}{3}\mathrm{vol}(H)$.
Once we have proven that, it is easy to prove that

$$\mathbb{P}\left(\left\{\phi(S) \leq \frac{3}{\hat{r}}\right\} \middle| \left\{\mathrm{vol}(S) \in \left[\frac{\mathrm{vol}(H)}{3}, \frac{2\mathrm{vol}(H)}{3}\right]\right\}\right) > 0 \tag{111}$$

In fact first we can prove that

$$\mathbb{E}\left[\phi(S) \middle| \left\{\mathrm{vol}(S) \in \left[\frac{1}{3}\mathrm{vol}(H), \frac{2}{3}\mathrm{vol}(H)\right]\right\}\right] \leq \frac{3}{\hat{r}} \tag{112}$$

To see why this is true, it is enough to notice that

$$\mathbb{E}\left[\frac{\delta(S, V \setminus S)}{\min(\mathrm{vol}(S), \mathrm{vol}(V \setminus S)} \middle| \left\{\mathrm{vol}(H) \in \left[\frac{1}{3}\mathrm{vol}(H), \frac{2}{3}\mathrm{vol}(H)\right]\right\}\right] \leq \tag{113}$$

$$\frac{\sum_{i=1}^{m} \mathbb{P}\left(\{e_i \text{ is cut}\} \middle| \left\{\mathrm{vol}(S) \in \left[\frac{1}{3}\mathrm{vol}(H), \frac{2}{3}\mathrm{vol}(H)\right]\right\}\right)}{\frac{1}{3}\mathrm{vol}(H)} \leq \tag{114}$$

$$\frac{m}{\frac{1}{3}\hat{r}m} = \frac{3}{\hat{r}} \tag{115}$$

Hence, using the fact that

$$\mathbb{P}\left(\left\{\phi(S) \leq \frac{3}{\hat{r}}\right\} \middle| \left\{\mathrm{vol}(S) \in \left\{\frac{1}{3}\mathrm{vol}(H), \frac{2}{3}\mathrm{vol}(H)\right\}\right\}\right) \geq \tag{116}$$

$$\mathbb{P}\Bigg(\Bigg\{\phi(S) \leq \mathbb{E}\left[\phi(S)\middle|\left\{\mathrm{vol}(S) \in \left[\frac{1}{3}\mathrm{vol}(H), \frac{2}{3}\mathrm{vol}(H)\right]\right\}\right]\Bigg\} \mid \left\{\mathrm{vol}(S) \in \left[\frac{1}{3}\mathrm{vol}(H), \frac{2}{3}\mathrm{vol}(H)\right]\right\}\Bigg) > 0 \quad (117)$$

Which concludes the proof. □

Now, we prove Lemma 4.5:

Proof. In order to prove that when the average $\hat{r}$ is small, then the volume of the graph is not larger than $2\hat{r}^2 n^{\hat{r}}$ it is enough to notice the following fact:

$$\mathrm{vol}(H) = \hat{r}m \quad (118)$$

In order to estimate how many edges we can have when the average edge size must be $\hat{r}$, we notice the following: in order to maximize m, we can take all edges of size $l \leq \hat{r}$, and for every edge of size l we can add for free another edge of size $q = 2\hat{r} - l$ without altering the average edge size $\hat{r}$. (Notice that it is not possible to create more edges than with this technique, without altering the average edge size).
Now, it is enough to count how many edges of size $l \leq \hat{r}$ we can create:

$$\begin{aligned}
\mathrm{vol}(H) &\leq \hat{r}m && (119)\\
&\leq 2\hat{r}\,|\{e \in E \text{ s.t. } |e| \leq \hat{r}\}| && (120)\\
&= 2\hat{r}\sum_{l=1}^{\hat{r}} \binom{n}{l} && (121)\\
&\leq 2\hat{r}^2 n^{\hat{r}} && (122)
\end{aligned}$$

□

With this fact we have just been able to prove that the mixing time for irregular hypergraphs is $O\left(\frac{\log(n)}{\phi^3}\right)$. Although this mixing time is not as good as the one for graphs and regular hypergraphs, notice that whenever the conductance of the graph is $\gg \frac{1}{n}$ (for example, a small constant $O(1)$) than the result found is almost equivalent to Theorem 1, and more importantly ensures an approximation factor for a clustering algorithm based on this mixing result which would be logarithmic in the number of nodes rather than linear.

4.2.2 We can choose the starting vertex wrt the stationary distribution

Another possible solution takes advantage of this fact: it is easy to see that in the proof of Lemma 3.7 we can easily substitute the definition of $R_0(k) := \sqrt{\frac{\hat{k}}{d(v_0)}}$ with $d(v_0)$ being the degree of the vertex v_0 s.t. the initial probability vector is χ_{v_0}. This yields an improved upper bound of $I_t(k) \leq \sqrt{\frac{\hat{k}}{d(v_0)}} e^{-\frac{t\phi^2}{4}} + \frac{k}{\mathrm{vol}(H)}$
We might ask whether we are now able to conclude a mixing time which is logarithmic in the number of nodes n: the answer to that is that it clearly depends on how large is $d(v_0)$. In fact if $\mathrm{vol}(G) = 2^n$ and $d(v_0) = O(1)$, then

we cannot hope to achieve a better mixing time than $t = O(n)$. But if, instead, $d(v_0) = \text{poly}\left(\frac{1}{n}\right)\text{vol}(G)$, then $\sqrt{\frac{\tilde{k}}{\tilde{d}}} = \text{poly}(n)$ and the mixing time becomes once again logarithmic in n.
So the idea is this: if we can pick the starting vertex with probability $\sim \pi = \left(\frac{d(i)}{\text{vol}(G)}\right)_{i=1}^{n}$, then we can estimate the probability that we pick a bad (with very little degree) vertex as:

$$\mathbb{P}\left(\left\{\text{starting vertex has degree} \ll \frac{\text{vol}(G)}{\text{poly}(n)}\right\}\right) \leq n \cdot \frac{\text{poly}(n)}{\text{vol}(G)} \tag{123}$$

$$\xrightarrow[(\text{vol}(G))\to 2^n]{} 0 \tag{124}$$

This is saying that when we are allowed to select the starting vertex with probability proportional to the stationary distribution, then we have a probability close to 1 to select a *good* starting vertex, namely a vertex with high volume which ensures that $\frac{\text{vol}(G)}{\tilde{d}} = O(\text{poly}(n))$, and yields a mixing time $O\left(\frac{\log(n)}{\phi^2}\right)$. Since in the vast majority of probabilistic clustering algorithms, the starting vertex is chosen at random w.p. $\sim \pi$, then we can conclude that our result Theorem 4 is still powerful enough to provide good mixing guarantees with high probability.

5 Experiments

In this section we are going to show with empirical evidence that, indeed, the mixing result found makes sense. In particular, we want to address the following questions:

- To begin with, it is interesting to understand if mixing happens within the time limit set by Theorem 4. In particular, if the result is correct, the mixing time for the Lovasz-Simonovits curve should behave like $O\left(\frac{\log(\text{vol}(H))}{\phi^2}\right)$. Hence, by providing the algorithm a family of hypergraphs with constant volume and different conductance or vice versa, we should observe a mixing time logarithmic with respect to a change in the volume, and inversely quadratic when the conductance changes.

- In addition, we want to study empirically the following behaviour: in r-uniform hypergraphs, the theoretical upper bound according to [2] is dependent on r: in particular $t = O\left(\frac{\log(\text{vol}(H))}{\gamma_2}\right) \leq O\left(\frac{\log(\text{vol}(H))}{r\phi^2}\right)$ (Thanks to the Hypergraph Cheeger's Inequality theorem in [3]). Notice that, in contrast, the parameter r disappears in our discussion (it is possible to see that by assuming that the hypergraph is r-uniform, the proof of Theorem 4 cannot leverage such additional information). What we want to understand is whether the additional r factor is important also in our analysis and it has been somehow overseen, or if the method is simply not powerful enough to explain a mixing time also dependent on r.

We start our discussion with the analysis of a family of hypergraphs with constant volume and varying conductance, and vice versa.

In order to reproduce the experiments, you can use the code available on GitHub [1]

5.1 Generate a hypergraph with desired conductance

Algorithm 1: Create hypergraph with desired conductance and volume

Requires: $n \geq 0, \text{vol}(H), 0 < \phi < 1$
Returns: $H : \phi(H) = \phi$
$A := [1, ..., \frac{n}{2}]$
$B := [\frac{n}{2} + 1, ..., n]$
$E := \emptyset$
vol_so_far := 0
while $|E| < vol(H)\frac{\phi}{2}$ **do**
  $r = \text{random}(2, 1/\phi)$
  $e = \{v : |e| = r, e \cap A \neq \emptyset \wedge e \cap B \neq \emptyset\}$
  $E = E \cup e$
end
vol_bip = vol(A)
while *vol(A) < vol(H)/2* **do**
  $r = \text{random}(2, 1/\phi)$
  $e = \{v : |e| = r, e \cap A \neq \emptyset \wedge e \cap B = \emptyset\}$
  $E = E \cup e$
end
vol_bip = vol(B)
while *vol(B) < vol(H)/2* **do**
  $r = \text{random}(2, 1/\phi)$
  e = $\{v : |e| = r, e \cap A = \emptyset \wedge e \cap B \neq \emptyset\}$
  $E = E \cup e$
end
Return H

In order to generate a hypergraph with the desired volume and conductance, the approach described by the pseudocode in Algorithm **??** has been used.
We explain a bit more into details how and why the following approach works: first, we create the optimum cut A and B consisting of half nodes each. Then, we add enough crossing hyperedges so that the final conductance of the cut (A, B) is indeed the desired ϕ. Then, we finally conclude by simply adding internal edges to the two partitions A and B. Since adding random hyperedges inside the two partitions makes them two expanders (with high conductance), then we can still expect that the minimum cut conductance is indeed (A, B). Although a proper proof bounding the probability of having a cut with better conductance than (A, B) could not be found, empirical evidence show that by taking random cuts in the graph, we are not able to find a cut with better conductance than ϕ. Notice that in order to build an r-uniform hypergraph, it is simply enough to substitute the random choice of the r parameter with the r uniformity factor.

[1] https://github.com/steber97/spectral_local_clustering

Table 1: Family of four graphs with constant volume, and variable conductance.

	Volume	Conductance 1	Conductance 2	Conductance 3	Conductance 4
H_v	10000	0.1	0.05	0.02	0.01

Table 2: Family of four graphs with constant conductance, and variable volume.

	Conductance	Volume 1	Volume 2	Volume 3	Volume 4
H_c	0.05	3000	5000	10000	15000

5.2 Const volume vs const conductance

The first experiment is the one showing that, indeed, the mixing time for the discrete process described in Section 3 respects the rule $t = O\left(\frac{\log(\mathrm{vol}(H))}{\phi^2}\right)$. We are going to check that it is indeed the case by simply showing these two simple experiments: first, we are going to build two families $\tilde{H}_v$ and $\tilde{H}_c$ with constant volume and conductance (hence the subscript). Notice that when the volume is constant, the conductance changes and vice versa. The two families of hypergraphs built are as described in Tables 1 and 2. Every family of hypergraph is made of four instances, and every instance has a constant factor (either the volume or the conductance) and the other variable term.
The experiments are held with this simple algorithm:

Algorithm 2: Discrete Random Walk in Hypergraph

```
p_0 = χ_{v_0}
t = 0
while not mixed do
    G_t = (V, ∅)
    for e ∈ E do
        e' = collapse e
        Add e' to G_t
    M_t = (1 − dt)I + dt A_t D^{−1}
    p_{t+dt} = M_t p_t
    t = t + dt
```

Notice that G_t is the collapsed graph at time t, where with the term *collapsed* we mean substituting every hyperedge with a simple edge between $(v_t^{\max}, v_t^{\min})$ plus a good number of self loops as defined in Section 3.1. The error bars are due to different repetitions of the algorithm, with different random starting points. The random walk is iterated until we have mixed, namely until the probability on the cut $S = [1, \frac{n}{2}]$ are less then $\frac{1}{n}$ far from the stationary distribution.
Then, we simply report how many iterations are necessary before the random walk converges to the stationary distribution, using the discrete process described.
What we would expect, is that the time to converge grows logarithmic w.r.t. the volume, and it decreases quadratically when the conductance increases.
Below, we report in two plots the evolution of the time w.r.t. the change in volume (when the conductance is constant) and the evolution of time when the conductance changes and the volume remains constant.

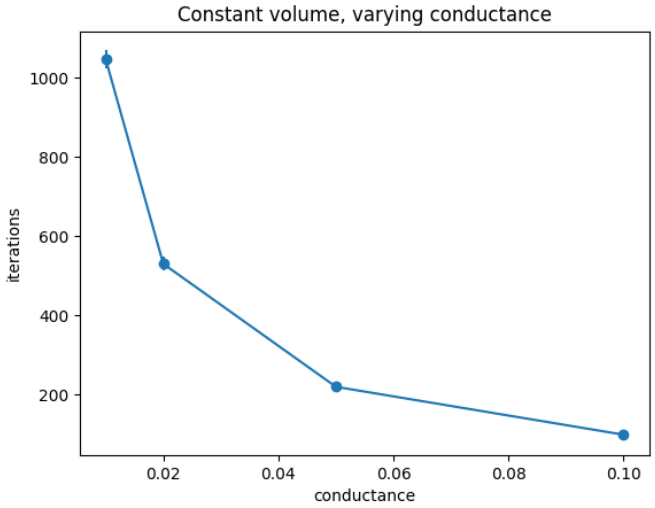

Figure 1: Constant volume, varying conductance: the mixing time decreases as a quadratic like curve wrt the conductance.

As it is possible to see in Figure 2, when the volume grows and the conductance remains constant, then we can observe a logarithmic growth in the mixing time (although it is not exactly a logarithm, it is interesting to see that the curve is concave and resembles a logarithm). At the same time, we can also see that in Figure 1 it is presented the mixing time when the volume is constant and the conductance changes: it is possible to see that the increase of the conductance correspond to a decrease in the mixing time. The curve resembles a quadratic curve $\frac{1}{x^2}$, which is our initial claim since the convergence time depends on $\frac{1}{\phi^2}$ when the volume is a constant.

5.3 R-uniform family of hypergraphs

In this experiment, we want to study the mixing time of the discrete process described in Section 3 applied to r-uniform hypergraphs. In fact, although our proof does not improve the mixing time when there is the additional assumption of r-uniformity, we know from [2] that actually a better upper bound exists for r-uniform hypergraphs: $t = O\left(\frac{\log(\mathrm{vol}(H))}{r\phi^2}\right)$.
Since the process described is discrete, it has the additional advantage of being easily simulable. Hence, we can hope to gain some insights on whether we can hope to improve the bound found for r -uniform hypergraphs: in particular, we will create a family of r-uniform hypergraphs for different values of r, keeping constant both the volume and the conductance. Then, we are going to estimate the mixing time and see if the variable r is inversely proportional to the mixing time.
After having generated some r-uniform hypergraph families with constant volume and conductance (with the same procedure as described in Section 5.1), we evolve the probability vector with the discrete procedure presented in Section

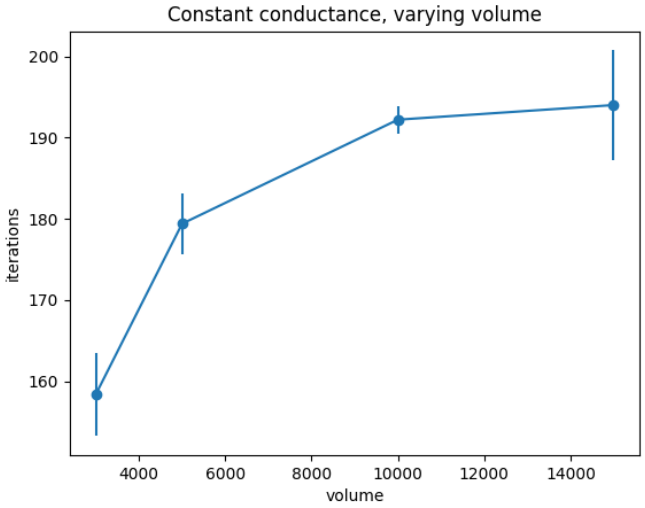

Figure 2: Constant conductance, varying volume: the mixing time grows with a logarithmic like curve wrt the volume.

3.1 and observe the mixing time w.r.t. the r-uniformity factor.
You can observe the mixing time in Figure 3.
The plot shows the l-infinity norm of the vector $\|\mathbf{p}_t - \pi\|_\infty$ for different t (on the x-axis), on a logarithmic scale. Indeed, it looks like the higher the r-uniformity factor, the faster the mixing time is, hence suggesting that stronger upper bounds for r-uniform hypergraphs can be achieved using our discrete diffusion method. This is encouraging because it might suggest that with a better analysis we might be able to prove theoretically a better mixing bound for our discrete process.

6 Conclusion and further work

In the previous sections, we have been exploring the mixing properties of the discrete process that describes a random walk in a hypergraph. Recalling what was said in Section 2.4 about the principles of clustering algorithms, mixing is only half of the required results in order to achieve a proper clustering algorithm: in particular, we also need a leaking result. Unfortunately, proving a proper leaking result for hypergraph is yet (as far as we are aware) an open question. In the following sections we are going to give an intuition of why solving the leaking problem is so hard, and propose some directions for future research in this area.

6.1 Missing leaking result for hypergraphs

The goal of this section is to be able to prove an equivalent result to the leaking result in Spielman et al. [7] (Lemma 2.7 on the volume of the set S^g). Notice that the proofs are re-adaptations of the proofs contained in the above mentioned

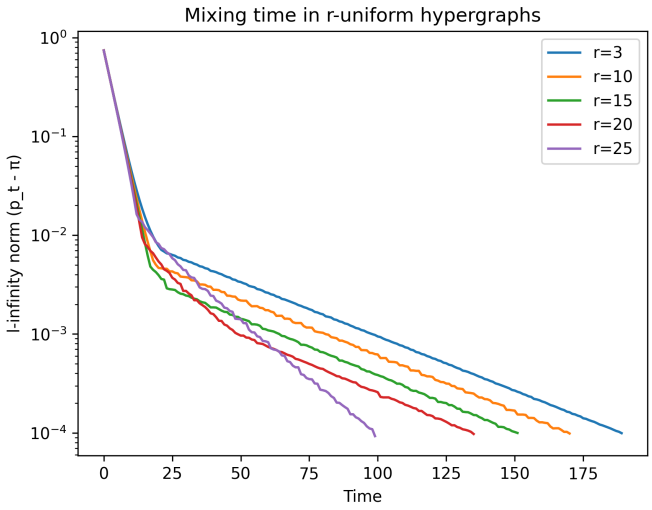

Figure 3: Constant conductance, constant volume, varying r-uniformity factor. On the y-axis, the logarithm of the $\|\mathbf{p}_t - \pi\|_\infty$, on the x-axis the time t

.

paper, and for brevity they have been omitted. As it will be clear later on in the proof, proving an equivalent statement is hard and could not be done, though we hope to shed some light on the argument, and provide some intuition for a possible solution.
The final lemma that we would be able to prove is:

Lemma 6.1. *$\forall S \subset V$ s.t. $vol(S) \leq \frac{1}{2}vol(H), \exists S^g \subset S$ s.t. $vol(S^g) \geq \frac{1}{2}vol(S)$ and such that $\forall v \in S^g, \forall t \geq 0$ and with $\mathbf{p}_0 = \chi_v$*

$$p_t(S) \geq \mathbb{P}(\{\textit{random walk never leaves } S\}) \geq 1 - \frac{t\phi_H(S)}{2} \tag{125}$$

Where with ϕ_H we mean the conductance of the cut S in the hypergraph H.

If we were able to prove such a leaking result, notice that we would be able to build a probabilistic clustering algorithm for hypergraphs using the strategy described in Section 2.4.
We start with this simple proposition:

Proposition 6.2. *$\forall t \geq 0$, Let G_t be the collapsed graph at time t as described in Section 3.1. Then $\forall S \subset V$ s.t. $vol(S) \leq \frac{1}{2}vol(G)$*

$$\phi_H(S) \geq \phi_{G_t}(S) \tag{126}$$

Proof. In order to prove this proposition, we can easily notice that by construction of G_t it is clear that vol(S) is the same in both the hypergraph H and the collapsed graph G_t. So, it is enough to prove that $\forall S \subset V$, $E_H(S, \bar{S}) \geq E_{G_t}(S, \bar{S})$. We can prove the previous claim by simply saying that it holds that $\forall e \in E_{G_t}, e \in E_{G_t}(S, \bar{S}) \implies e' \in E_H(S, \bar{S})$, with e' being the hyperedge that when collapsed created the edge e. This is clearly true by

construction. So, since the number of crossing edges in the hypergraph H is certainly larger or equal than the crossing edges in the collapsed graph G_t, it also holds that the conductance is higher. □

Here we need to introduce some notation: we define $M_t^{\mathbf{s}}$ the transition probability matrix, obtained at time t with the collapsing procedure described in Section 3.1, when the starting probability distribution $\mathbf{p}_0 = \mathbf{s}$.
Moreover, let us define the matrix D_S as the diagonal matrix with entries $D(u,u) = 1$ if $u \in S$ and zero otherwise.
In addition, we need to take into account that the transition probability matrix differs between our discrete process (described in Section 3.1) and Spielman et al.'s ([7], Section 2.2): in particular, for standard graphs the transition probability matrix for a random walk at time t is the t power matrix M^t (where M is the transition probability matrix $M := \frac{1}{2}(I + AD^{-1})$). Instead, in our process, the transition probability matrix differs for different times t' since at every iteration we collapse the hypergraph into a different $G_{t'}$ with a transition probability matrix $M_{t'}^{\mathbf{s}}$ (of course, it also depends on the starting probability distribution). Hence, the transition probability matrix for our process at time t can be described as the product of the transition probability matrices at time $t' \leq t$:

$$\prod_{\substack{0 \leq t' \leq t: \\ t' = k \cdot dt \\ \text{for some } k \in \mathbb{N}}} M_{t'}^{\mathbf{s}} \tag{127}$$

To begin with, we claim that it is possible to prove an analogous claim as the Escaping Mass proposition:

Proposition 6.3. *Escaping Mass:* $\forall t \geq 0, \forall S \subseteq V$ *s.t.* $vol(S) \leq \frac{1}{2} vol(H)$

$$\mathbf{1}^T \Bigg(\prod_{\substack{0 \leq t' \leq t: \\ t' = k \cdot dt \\ for\ some\ k \in \mathbb{N}}} (D_S M_{t'}^{\psi_S}) \Bigg) \psi_S \geq 1 - \frac{t \phi_H(S)}{2} \tag{128}$$

The proof is omitted, but it is a simple re-adaptation of the proofs in Propositions 2.2, 2.4, 2.5 in [7] for our discrete process, with the additional Lemma 6.2.
Notice that, although this result resembles our leaking result goal, it is not equivalent: in fact, this leaking result only works when the starting probability vector is the stationary distribution over some set S. In order to understand why this is not enough to build an algorithm for clustering, notice that a clustering algorithm starts the random walk from a probability distribution χ_v, for some v picked at random. Hence, it is essential to prove that there is a large set of vertices $S^g \subset S$ such that when starting the random walk from χ_v for any $v \in S^g$, the probability leaking out of S is small.
So, if we define S^g to be

$$S^g := \left\{ v \in S : \chi_S^T \Bigg(\prod_{\substack{0 \leq t' \leq t: \\ t' = k \cdot dt \\ \text{for some } k \in \mathbb{N}}} M_{t'}^{\chi_v} \Bigg) \chi_v \geq 1 - t \phi_H(S) \right\} \tag{129}$$

then it is enough to prove that the volume of S^g is large: if this is the case, in fact, then by picking vertices at random we have a high chance of picking one which is *good*, namely which falls in S^g and hence has good leaking properties. As we will see, the generalization of the leaking result for hypergraphs fails here: in particular, it is not possible to prove the following lemma:

Lemma 6.4. *$vol(S^g) \geq \frac{1}{2} vol(S)$*

The idea behind this lemma is that when picking vertices at random according to the stationary distribution, the probability that the Escaping Mass Proposition 6.3 holds also when the starting probability distribution is of the form χ_v, is a constant. Unlike previous propositions, we will show that this lemma is hard to generalize for hypergraphs.
To begin with, we define a new subset of S:

$$S' := \left\{ v \in S : \mathbf{1}^T \Bigg(\prod_{\substack{0 \leq t' \leq t: \\ t' = k \cdot dt \\ \text{for some } k \in \mathbb{N}}} (D_S M_{t'}^{\chi_v}) \Bigg) \chi_v \geq 1 - t\phi_H(S) \right\} \tag{130}$$

Namely, the set of vertices $v \in S$ such that the probability of never escaping from S when starting from v at any step $t' \leq t$ is larger than $1 - \phi_H(S)t$.
It is easy to see that $S' \subseteq S^g$: in fact the condition to belong to S^g is looser than the condition to belong to S'.

$$\chi_S^T \Bigg(\prod_{\substack{0 \leq t' \leq t: \\ t' = k \cdot dt \\ \text{for some } k \in \mathbb{N}}} M_{t'}^{\chi_v} \Bigg) \chi_v \geq \chi_S^T \Bigg(\prod_{\substack{0 \leq t' \leq t: \\ t' = k \cdot dt \\ \text{for some } k \in \mathbb{N}}} (D_S M_{t'}^{\chi_v}) \Bigg) \chi_v \tag{131}$$

Due to Proposition 2.4 in [7]
So, if we are able to prove that $\text{vol}(S') \geq \frac{1}{2}\text{vol}(S)$, then it must also hold that $\text{vol}(S^g) \geq \frac{1}{2}\text{vol}(S)$.
Now it comes the tricky part which is hard to generalize: in order to prove the final claim, for general graphs we take advantage of this fact: when the transition probability matrix M does not change between iterations, it is possible to re-write the outcome of Claim 6.3 as

$$\begin{aligned}
\frac{1}{2} t\phi_H(S) &\geq 1 - \mathbf{1}^T (D_S M)^t \psi_S && (132)\\
&= \sum_{u \in S} \frac{d(u)}{\text{vol}(S)} \left(1 - \mathbf{1}^T (D_S M)^t \chi_v \right) && (133)\\
&\geq \sum_{u \in \bar{S}'} \frac{d(u)}{\text{vol}(S)} (1 - \mathbf{1}^T (D_S M)^t \chi_v)) && (134)\\
&\geq \frac{\text{vol}(\bar{S}')}{\text{vol}(S)} t\phi(S) && (135)\\
\implies & \frac{\text{vol}(S')}{\text{vol}(S)} \geq \frac{1}{2} && (136)
\end{aligned}$$

However, when the transition probability matrix changes according to the starting distribution (as it is the case for the discrete process described in 3.1), then the Equality in between Equations 132 and 133 does not hold:

$$1 - \mathbf{1}^T \Bigg(\prod_{\substack{0 \leq t' \leq t: \\ t' = k \cdot dt \\ \text{for some } k \in \mathbb{N}}} (D_S M_{t'}^{\psi_S}) \Bigg) \psi_S \neq \sum_{u \in S} \frac{d(u)}{\mathrm{vol}(S)} \Bigg(1 - \mathbf{1}^T \Bigg(\prod_{\substack{0 \leq t' \leq t: \\ t' = k \cdot dt \\ \text{for some } k \in \mathbb{N}}} (D_S M_{t'}^{\chi_u}) \Bigg) \chi_u \Bigg) \tag{137}$$

and, instead, there is a $\leq$ relation between them, which is the opposite of what we need. To see why this is the case, let's try to quantify the two sides of the inequality above, at the very first time step of the random walk: on the left hand side, we have that

$$1 - \mathbf{1}^T \left(D_S M_0^{\psi_S} \right) \psi_S = dt \sum_{e \in E(S, \bar{S})} \frac{1}{\mathrm{vol}(S)} \tag{138}$$

This is because whenever there is a crossing hyperedge $e \in E$, then the collapsed edge in the graph G_0 will be of the form $(u, v) : u \in S \wedge v \in \bar{S}$ by the way we have chosen the collapsing strategy (in fact, vertices in S have a non-zero probability, vertices in $\bar{S}$ have a zero probability when the initial probability distribution is ψ_S), and the probability flowing on such edge is $\frac{p(u)}{d(u)} = \frac{\psi_S(u)}{d(u)} = \frac{1}{\mathrm{vol}(S)}$. The dt factor is due to the length of the step.
For the right hand side, instead, we have that

$$\sum_{u \in S} \frac{d(u)}{\mathrm{vol}(S)} \left(1 - \mathbf{1}^T (D_S M_{t'}^{\chi_u}) \chi_u \right) = \tag{139}$$

assuming all crossing hyperedges $e \in E(S, \bar{S}) : u \in e$ are collapsed into $(u, v) : v \in \bar{S}$

$$dt \sum_{u \in S} \frac{d(u)}{\mathrm{vol}(S)} |\{e \in E(S, \bar{S}) : u \in e\}| \frac{\chi_u(u)}{d(u)} = \tag{140}$$

$$dt \sum_{e \in E(S, \bar{S})} \sum_{u \in e : u \in S} \frac{1}{\mathrm{vol}(S)} \tag{141}$$

which is definitely larger than the left hand side quantity.
This inequality, though, might suggest a reasonable way to find a proper lower bound: for instance, when the graph is r-uniform, then the right hand side is at most r times larger than the left hand side (since every hyperedge can be counted up to r times).
Hence, though this claim is true for the first iteration:

Claim 6.5. *in an r-uniform hypergraph*

$$1 \;-\; \mathbf{1}^T(D_S M_0^{\psi_S})\psi_S \quad \geq \quad \frac{1}{r}\sum_{u\in S}\frac{d(u)}{vol(S)}\left(1 \;-\; \mathbf{1}^T(D_S M_0^{\chi_u})\chi_u\right) \tag{142}$$

Additional empirical evidence show that this fact is also true for larger $t' > 0$ though, unfortunately, a proper proof of this conjecture could not be found. Hence, assuming the conjecture

Conjecture 6.6. *In r-uniform hypergraphs,*

$$\frac{1}{2}t\phi_H(S) \geq 1 - \mathbf{1}^T\left(\prod_{\substack{0\leq t'\leq t:\\ t'=k\cdot dt\\ for\ some\ k\in\mathbb{N}}}(D_S M_{t'}^{\psi_S})\right)\psi_S \tag{143}$$

$$\geq \frac{1}{r}\sum_{u\in S}\frac{d(u)}{vol(S)}\left(1 - \mathbf{1}^T\left(\prod_{\substack{0\leq t'\leq t:\\ t'=k\cdot dt\\ for\ some\ k\in\mathbb{N}}}(D_S M_{t'}^{\chi_u})\right)\chi_u\right) \tag{144}$$

where r is the r-uniformity factor of the hypergraph, or in case of a non uniform hypergraph, $r := \max_{e\in E}|e|$.
An extremely interesting future result would be to find proper theoretical evidence that it is actually the case. In case it was true, in fact, we would be able to conclude that the leaking result for r-uniform hypergraphs is *only* r times weaker than in general graphs. To see this, it is enough to re-define $S^g := \{u \in S : \chi_S^T \prod_{t'\leq t}(M_{t'})\chi_u \leq rt\phi_H(S)\}$. This ensures that $\text{vol}(S^g) \geq \frac{1}{2}\text{vol}(S)$. The reason why the leaking result is r times weaker than in standard graphs, is because when picking a node $v \in S^g$, the probability that leaks out of S is r times larger than what would leak out of S in a graph. At the same time, the fact that instead the mixing result is theoretically r-times stronger ($t = O(\frac{\log(\text{vol}(H))}{r\phi^2})$ due to [2], [3]) would allow us to find clustering algorithms for r-uniform hypergraphs with equivalent guarantees than the ones for graphs.

References

[1] R. Andersen, F. Chung, and K. Lang. Local graph partitioning using pagerank vectors. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 475–486, 2006.

[2] T. H. Chan, A. Louis, Z. G. Tang, and C. Zhang. Spectral properties of hypergraph laplacian and approximation algorithms. *CoRR*, abs/1605.01483, 2016.

[3] M. Kapralov, R. Krauthgamer, J. Tardos, and Y. Yoshida. Towards Tight Bounds for Spectral Sparsification of Hypergraphs. *arXiv*, Nov. 2020.

[4] P. Li and O. Milenkovic. Submodular hypergraphs: p-laplacians, cheeger inequalities and spectral clustering. *CoRR*, abs/1803.03833, 2018.

[5] L. M. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Struct. Algorithms*, 4:359–412, 1993.

[6] D. Spielman. The lovasz - simonovits approach to random walks, 2018.

[7] D. A. Spielman and S. Teng. A local clustering algorithm for massive graphs and its application to nearly-linear time graph partitioning. *CoRR*, abs/0809.3232, 2008.

[8] Y. Takai, A. Miyauchi, M. Ikeda, and Y. Yoshida. ACM, aug 2020.