# Local Clustering in Hypergraphs
## (Mixing in irregular hypergraphs)

Stefano Huber

Department of Computer Science
EPFL

June 28, 2022

# Table of Contents

# Definitions

- A graph $G = (V, E)$ is a collection of $n$ vertices and $m$ edges.
- A hypergraph $H = (V, E)$ is a generalization of a graph to capture higher order relations between vertices.
- An interesting property of a graph is its conductance $\phi$:

**Conductance**

$$\phi(S) = \frac{E(S,\bar{S})}{\min(\text{vol}(S),\text{vol}(\bar{S}))}$$

It measures how well inter-connected the cut $S$ is, and well separated from the rest of the graph.

- In the Local Clustering problem, you want to find a cut $(S, \bar{S}) : S \subseteq V \wedge S \cup \bar{S} = V$ around a vertex $v \in S$ such that the conductance is at most a parameter $\phi$.
- The way to find good clusters is by studying mixing and leaking properties of random walks in a graph.

# Random Walks in graphs

Local clustering in graphs is often achieved by random walks. Here are some useful definitions:

- Random walks are *lazy* (w.p. $\frac{1}{2}$ you do not move)

**Transition probability matrix**

$M = \frac{1}{2}(I + AD^{-1})$

**Evolution of probability vector**

$\mathrm{p}_{t+1} = M\mathrm{p}_t$

- The Laplacian matrix $\mathcal{L} = I - AD^{-1}$ describes the rate of convergence of the probability vector wrt the time:

**Derivative of $p_t$ wrt time**

$\frac{d\mathrm{p}_t}{dt} = -\mathcal{L}\mathrm{p}_t$

# Random Walks in graphs

- Random walks converge to the stationary distribution $\pi(u) = \frac{d(u)}{\mathrm{vol}(G)}$

**Convergence to stationary distribution**

when $t \to \infty \implies \mathrm{p}_t \to \pi$

- Studying how fast the probability vector converges to stationary distribution is called *mixing*, and it is done with the Lovasz-Simonovits curve.

# Lovasz-Simonovits curve

- It is easy to define the Lovasz Simonovits curve using the definition of a sweep cut:

**Sweep Cut**

$S_j(\mathrm{p}_t) = \{u \in V : \frac{p_t(u)}{d(u)} \text{ is maximized} \wedge |S_j(\mathrm{p}_t)| = j\}$

- LS curve: for $k = \mathrm{vol}(S_j(\mathrm{p}_t))$ for some $j \in [0, n]$

**Lovasz-Simonovits curve**

$I_t(k) = \sum_{u \in S_j(\mathrm{p}_t)} p_t(u)$

- The LS curve is concave, and it decreases with time until it converges to a straight line from $(0, 0)$ to $(\mathrm{vol}(G), 1)$: for all $t$ and $k \in [0, \mathrm{vol}(G)]$

**Lovasz-Simonovits curve is decreasing**

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

# 1. Studying mixing with Lovasz-Simonovits curve

- It is possible to study the mixing time of a random walk by studying how fast the LS curve converges to a straight line. In particular, in between time steps, the LS-curve decreases piece-wise to be below a *large* chord depending on the conductance $\hat{\phi} = \min_{j\in[1,n],t} \phi(S_j(\mathrm{p}_t))$: $\forall k \in [0, \mathrm{vol}(G)]$, and $\hat{k} := \min(k, \mathrm{vol}(G) - k)$

**Rate of decrease of LS curve, depending on $\hat{\phi}$**

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

- This allows for exponentially fast convergence:

**Exponentially fast mixing wrt time**

$$I_t(k) - \pi(S_j(\mathrm{p}_t)) \leq \sqrt{\hat{k}} e^{-t\hat{\phi}^2}$$

EPFL

# 2. Leaking of random walks in a graph

- Leaking is the opposite of mixing: it provides a lower bound on the mixing time of a random walk, based on the optimal conductance of the graph.
- For graphs, leaking is very simple: when $p_0 = \psi_S$ for some $S \subseteq V$, then

**Leaking wrt optimal conductance $\phi^*$**

$p_t(S) \geq 1 - t\phi^*(S)$

- To conclude, it can also be proved that the volume of $S^g$ s.t. $\forall v \in S^g \subseteq S$ when $p_0 = \chi_v$ then $p_t(S) \geq 1 - t\phi^*(S)$ is large:

**Volume of $S^g$ is large**

$\mathrm{vol}(S^g) \geq \frac{1}{2}\mathrm{vol}(S)$

# Mixing+leaking = clustering algorithm

- Assumption: there is a set $S$ of large volume (but $\leq \frac{1}{2}\mathrm{vol}(G)$), s.t. for a parameter $\hat{\phi}$, its conductance is $\phi(S) \leq \frac{\hat{\phi}^2}{\log(n)} = \phi^*$
- You pick a vertex $v$ at random according to the stationary distribution (with good probability it falls in $S^g$) and evolve the probability vector from $\mathrm{p}_0 = \chi_v$ for $t = \frac{1}{4\phi^*}$ iterations.
- By leaking: $p_t(S) - \pi(S) \geq 1 - t\phi^* - \frac{1}{2} \geq \frac{1}{4}$
- in contrast, by mixing $p_t(S) - \pi(S) \leq \sqrt{\mathrm{vol}(S)}e^{-t\hat{\phi}^2}$
- This allows to conclude that $\hat{\phi} \leq \sqrt{\log(n)\phi^*}$ *not too far* from the original conductance.

# Local clustering algorithm for hypergraphs

- In order to find a proper clustering algorithm for hypergraphs, we need to generalize both mixing and leaking results to hypergraphs.
- Mixing:

**Known mixing result for $d$-regular hypergraphs**

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

- Leaking: no known leaking result as far as we know.
- In next sections: we provide a generalization to non $d$-regular hypergraphs and a short discussion on leaking.

# Why mixing is hard in hypergraphs?

- Proving mixing in hypergraphs is harder because it is not possible to define an equivalent discrete diffusion process as a random walk.
- This is because the Laplacian is a non-linear operator, for $B_e$ the convex hull of $\{\chi_v - \chi_u : u, v \in e\}$:

**Laplacian for hypergraphs**

$\mathcal{L}(\mathrm{x}) = \{\sum_{e \in E} w_e \mathrm{b}_e \mathrm{b}_e^T \mathrm{x} \mid \mathrm{b}_e = \arg\max_{\mathrm{b} \in B_e} \mathrm{x}^T \mathrm{b}\}$

- Such a laplacian operator is only good for describing $\frac{d\mathrm{p}_t}{dt} = -\mathcal{L}(D^{-1}\mathrm{p}_t)$ for infinitesimal time steps, hence the diffusion process is inherently continuous and harder to handle.
- Observation: notice that the laplacian is a singleton matrix when the vector $\frac{\mathrm{p}_t}{\mathrm{d}}$ has piece-wise unique entries. The problem is handling ties.

# Discrete diffusion process

- Idea: turn the hypergraph laplacian into a singleton matrix by solving ties *arbitrarily*.
- Question then is: is this simple resolution enough in order to obtain mixing?
- Quick answer: yes, it is possible to prove that solving ties easily is enough in order to build a discrete diffusion process with good mixing properties.

# Discrete diffusion process

- We collapse the hypergraph $H$ into a multigraph $G_t$ at every time step, by collapsing every hyperedge $e$ into $(v^t_{\max}(e), v^t_{\min}(e))$

**$v^t_{\max}(e)$**

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

**$v^t_{\min}(e)$**

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

ties are solved by vertex index (smallest first).

- Self loops are added in order to preserve the degree of every node.

# Discrete diffusion process

- Since $G_t$ is a graph, it is possible to evolve the probability vector $\mathrm{p}_{t+dt} = M_t \mathrm{p}_t$. Notice that the step size is $dt \leq \frac{1}{2}$
- How to prove good mixing properties? We need to make sure that the sweep cuts $S_j(\mathrm{p}_{t+dt})$ has the same conductance $\hat{\phi}$ in the collapsed graph $G_t$ and in the original hypergraph $H$.
- Since the graph $G_t$ was made using the probability vector $\mathrm{p}_t$, the conductance of the sweep cut $S_j(\mathrm{p}_{t+dt})$ is not ensured to be preserved in $G_t$.
- Idea is to build a support $\tilde{G}_t$ collapsed graph, such that the conductance of the sweep cut $S_j(\mathrm{p}_{t+dt})$ in $\tilde{G}_t$ is preserved.

# Discrete diffusion process

The proof of the mixing result goes like this:

- First, prove that the LS-curve $\tilde{I}_t$ is smaller than $I_t$.

**Lemma 1**

$\forall t,\ I_{t+dt}(k) \leq \tilde{I}_{t+dt}(k)$

- Then, take advantage of the known conductance of the sweep cuts in $\tilde{G}_t$ to prove

**Lemma 2**

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

- Which allows us to claim

**Lemma 3**

$I_t(k) - \pi(S_j(\mathrm{p}_t)) \leq \sqrt{k}e^{-t\hat{\phi}^2}$

# Definition of $\tilde{G}_t$

How can be sure that the conductance of the sweep cut $S_j(\mathrm{p}_{t+dt})$ is preserved in $\tilde{G}_t$?

- $\forall e \in E$, if $(v_{\min}^{t+dt}(e), v_{\max}^{t+dt}(e))$ crosses the cut and $(v_{\min}^{t}(e), v_{\max}^{t}(e))$ does not, then either $(v_{\min}^{t+dt}(e), v_{\min}^{t}(e))$ or $(v_{\max}^{t+dt}(e), v_{\max}^{t}(e))$ also crosses the cut.
- We can add to $G_t$ the new crossing collapsed edge, and remove enough self loops so that the degree of $\tilde{G}_t$ is the same as in $H$.
- Notice that in the graph $\tilde{G}_t$ the conductance of the sweep cut $S_j(\mathrm{p}_{t+dt})$ is $\phi(S_j(\mathrm{p}_{t+dt}))$, because we have ensured that every crossing hyperedge is collapsed into a crossing edge, and the degree has not changed.
- We can now define $\tilde{\mathrm{p}}_{t+dt} = \tilde{M}_t \mathrm{p}_t$, and accordingly $\tilde{I}_{t+dt}$

# Proof of Lemma 1

**Lemma 1**

$\forall k \in [0, \mathrm{vol}(H)], \forall t, I_{t+dt}(k) \leq \tilde{I}_{t+dt}(k)$

**Proof.**

The idea is that $\forall u \in S_j(\mathrm{p}_{t+dt})$, the new edges $(u, v)$ added to $\tilde{G}_t$ are such that $v$ has higher $\frac{p_t(v)}{d(v)}$ value than $\frac{p_t(u)}{d(u)}$. This ensures that the probability at time $t + dt$ on $u \in S_j(\mathrm{p}_{t+dt})$ is higher, and hence $\tilde{I}_{t+dt}(k) \geq I_{t+dt}(k)$. □

# Proof of Lemma 2

## Lemma 2

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

## Proof.

- In order to prove Lemma 2 you split edges (and self loops) in groups:
  - $W_1 = \{u, v : u, v \in \bar{S}_j(\mathrm{p}_{t+dt})\}$
  - $W_2 = \{u, v : u \in S_j(\mathrm{p}_{t+dt}) \text{ and } v \in S_j(\mathrm{p}_{t+dt})\}$
  - $W_3 = \{(u, u) : \mathrm{vol}(W_3) = dt \cdot \mathrm{vol}(S_j(\mathrm{p}_{t+dt}))\}$
  - $W_4 = \{(u, u) : \mathrm{vol}(W_4) = (1-2dt)\mathrm{vol}(S_j(\mathrm{p}_{t+dt}))\}$
- Then you prove the three bounds:
  - $\mathrm{p}_{t+dt}(W_1) \leq dt \cdot I_t(\mathrm{vol}(S_j(\mathrm{p}_{t+dt})) - \frac{\mathrm{vol}(W_2)}{dt})$
  - $\mathrm{p}_{t+dt}(W_2 \cup W_3) \leq dt \cdot I_t(\mathrm{vol}(S_j(\mathrm{p}_{t+dt})) + \frac{\mathrm{vol}(W_2)}{dt})$
  - $\mathrm{p}_{t+dt}(W_4) \leq (1-2dt)I_t(\mathrm{vol}(W_4))$
- and conclude by noticing that $I_t$ is concave, and $\frac{\mathrm{vol}(W_2)}{dt} \geq \hat{\phi}\hat{k}$.

$\square$

# Proof of Lemma 3

## Lemma 4

$$I_t(k) - \pi(S_j(\mathrm{p}_t)) \leq \sqrt{\hat{k}} e^{-\frac{t\hat{\phi}^2}{4}}$$

## Proof.

First, we define support function

$$R_t(k) = \begin{cases} \sqrt{k} & t = 0 \\ (1 - 2dt)R_{t-dt}(k) + dt(R_{t-dt}(k - \hat{\phi}\hat{k}) + R_{t-dt}(k + \hat{\phi}\hat{k})) & t > 0 \end{cases}$$

And then prove $I_t(k) \leq R_t(k)$ by induction on $t$. Finally, it is possible to prove $R_t(k) \leq \sqrt{k} e^{-\frac{t\hat{\phi}^2}{4}}$ by induction on $t$, using Taylor expansion $\sqrt{1 - \hat{\phi}} + \sqrt{1 + \hat{\phi}} \leq (1 - \frac{\hat{\phi}^2}{8}) \leq e^{-\frac{\hat{\phi}^2}{8}}$ □

# Is this result as powerful as the one for graphs?

Here we state back to back the theorem for $d-$ *regular* hypergraphs and irregular hypergraphs:

## Mixing in $d$-regular hypergraphs

$$I_t(k) - \pi(S_j(\mathrm{p}_t)) \leq \sqrt{\frac{k}{d}} e^{-\frac{t\hat{\phi}^2}{4}}$$

## Mixing in irregular hypergraphs

$$I_t(k) - \pi(S_j(\mathrm{p}_t)) \leq \sqrt{k} e^{-\frac{t\hat{\phi}^2}{4}}$$

Interestingly, the $d$ factor has dropped, but this might be an issue: in fact, $k$ can be as large as $O(n2^n)$ for hypergraphs (whereas $\frac{k}{d}$ is only $O(n)$). So, is it enough to have mixing time $t = O\left(\frac{\log(n)}{\hat{\phi}^2}\right)$ also for irregular hypergraphs?

# Conductance is $O(1)$, but mixing is $O(n)$

**Lemma 5**

*There exists a multigraph s.t. the conductance is $\frac{1}{2}$ but the mixing time is $O(n)$.*

Here is an example:

- Nodes $[1, n]$ in a straight path.
- Edge $(i, i+1)$ has weight $w(i, i+1) = \sum_{j=0}^{i} w(i, i+1)$ so that weight doubles at every $i$.
- Total graph volume: $2^{n-1}$.
- Conductance is $\frac{1}{2}$.
- Mixing time is $O(n)$ for a large fraction of nodes.

Can there be a hypergraph which gets collapsed into such an unfavourable multigraph?

# Two arguments for not having this issue in hypergraphs

Let's discuss some arguments of why it is not possible to have a hypergraph that gets collapsed in such a way that the conductance is $O(1)$ but the mixing time is $O(n)$. This would be problematic because the approximation factor for the clustering algorithm would be $\hat{\phi} = O(\sqrt{\phi n})$ instead of $\hat{\phi} = O(\sqrt{\phi \log(n)})$

- argument 1, if the conductance is large, then the volume cannot be too large.

**Lemma 6**

*Given $\hat{r} = avg(|e|)_{e \in E}$ then $\phi(H) \leq \frac{3}{\hat{r}}$*

- If the starting vertex is chosen at random w.p. proportional to $\pi$, then it holds $I_t(k) \leq \sqrt{\frac{k}{d(v_0)}} e^{-\frac{t\hat{\phi}^2}{4}}$ (Simple corollary of Lemma 3).

# Proof of Lemma 6

## Lemma 6

$\phi(H) \leq \frac{3}{\hat{r}}$ with $\hat{r}$ the average edge size.

## Proof.

- Take random cuts $W$ of size $\mathrm{vol}(W) \in [\frac{1}{3}\mathrm{vol}(H), \frac{2}{3}\mathrm{vol}(H)]$.
- compute $\mathbb{E}[\phi(W) \mid \mathrm{vol}(W) \in [\frac{1}{3}, \frac{2}{3}]\mathrm{vol}(H)] = \frac{3}{\hat{r}}$
- Conclude that $\mathbb{P}(\{\phi(W) \leq \mathbb{E}[\cdot] \mid \mathrm{vol}(W) \in [\frac{1}{3}, \frac{2}{3}]\mathrm{vol}(H)\}) > 0$
- This implies that $\phi(W) \leq \frac{3}{\hat{r}}$.

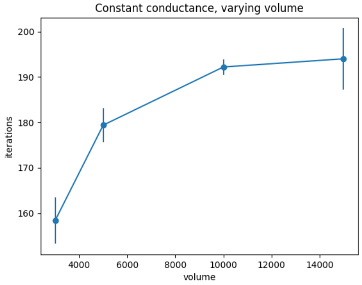
$\square$

# Experiments

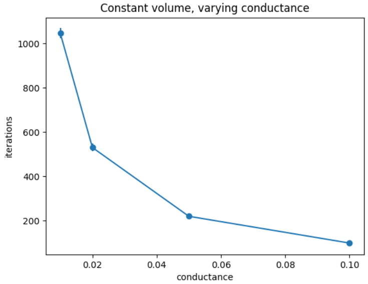
We performed some experiments to show empirically our mixing result.

- First, we want to prove that the mixing time grows logarithmically with the volume, and decreases quadratically with the conductance, namely $t = O\left(\frac{\log(n)}{\hat{\phi}^2}\right)$.
- We know that for $r$-uniform hypergraphs there is a theoretical upper-bound for the continuous diffusion process $t = O\left(\frac{\log(n)}{\hat{\phi}^2 r}\right)$. We want to see if such upper bound also holds for our simplified discrete diffusion process.

# Experiment 1

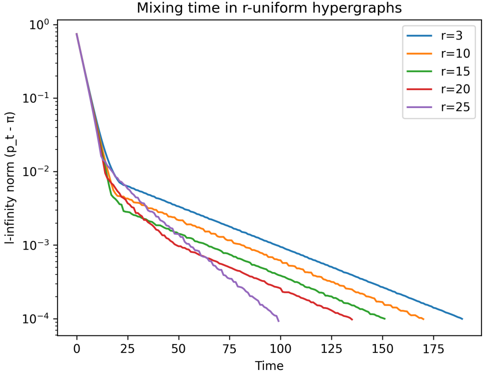



When the conductance is constant and the volume changes, the mixing time is *logarithmic* wrt the volume.




Instead, when the volume is constant and the conductance changes, then the mixing time decreases *quadratically* fast.

# Experiment 2




Indeed, the larger the $r$ the smaller the mixing time, suggesting that also with our discrete diffusion process when the hypergraph is $r$-uniform it might hold $t = O\left(\frac{\log(n)}{r\hat{\phi}^2}\right)$.

# Discussion about leaking

Leaking is indeed still an open problem in hypergraphs. In particular, it is indeed possible to prove that

**Leaking starting from the stationary distribution on a set $S$**

$\chi_S^T(\prod_{t' \leq t}(D_S M_{t'}))\psi_S \geq 1 - t\phi(S)$

But, in contrast, it is not possible to prove it when starting from a single vertex:

**Leaking starting from centered distribution in one vertex**

When $S^g = \{v \in S : \chi_S^T(\prod_{t' \leq t}(D_S M_{t'}))\chi_v \geq 1 - \phi(S)t\}$ then $\mathrm{vol}(S^g) \ngeq \frac{1}{2}\mathrm{vol}(H)$

# Hypothesis for leaking

Mildly speaking, when averaging (starting vertex $\psi_S$) we count every crossing hyperedge once. Instead, when starting from a single vertex $\chi_v$, we count every crossing edge up to $\approx r$ times (in case the graph is $r$-uniform). This introduces an additional $r$ factor in the leaking $p_t(S) \geq 1 - r\phi t$, but still it is not possible to prove that this claim holds for $t > dt$

# The end.

k Thank you!

# References I

Dirac, P. (1964).
*Lectures on Quantum Mechanics*.
New York, Belfer Graduate School of Science, Yeshiva University.