Lecture 1: A local algorithm for low conductance clusters using the Lovasz-Simonovits Theorem

30/3, 31/3 and 6/4 2015

1.1 Introduction

1.1.1 Motivation

Clustering is the task of grouping vertices of a graph which are more tightly knit than others. In the age of the internet and large social networks, these networks tend to be prohibitively large to run algorithms of time complexity larger than $\tilde{O}(n)$ [1]. In some cases, like the internet for example, it would be impossible to access all of the nodes even once. This calls for more near-linear and local approaches.

1.1.2 Local Clustering

We say that the graph is a *local algorithm* if at each step it only examines the neighbors of the current vertex. Given a graph and a vertex v, we want to construct a local algorithm which examines T vertices to return a *good* cluster of size at least $T/2$ containing v, if it exists. This procedure can then be used to construct a near linear clustering algorithm.

Observe that if v is present in a good cluster, then a random walk starting from v would mostly be contained within the cluster. Thus the set of vertices which had the highest probability of reaching would be a good guess for the cluster. Hence our procedure would be to estimate the probability of reaching each vertex from v by performing a random walk starting from v, select k nodes which have the highest probability, and check if they form a good cluster. If not repeat till some predetermined limit is reached.

To show that the above algorithm works, we need to show some local bounds instead of global bounds involving λ_2 (the second largest eigenvalue of the adjacency matrix). Such bounds are given to us by the Lovász-Simonovits Theorem.

[1] $\tilde{O}(\cdot)$ hides constants and *polylog* factors

1.2 Preliminaries

Graphs and Random Walks

We are given a weighted graph $G = (V, E)$ with the adjacency matrix A and weight function a where edge (u, v) has weight $a_{u,v}$. We define the normalized adjacency matrix M for the graph G as

$$m_{u,v} = \frac{a_{u,v}}{\sum\limits_{w} a_{u,w}}$$

A *random walk* over G is defined as follows. If the position at time t is at vertex u, then $m_{u,v}$ is probability with which we will be at v at time $t+1$. Thus a random walk is a Markov process with the transition matrix defined by M.

For node $u \in V$ denote by d_u the total weight of edges going out of u, i.e. $d_u = \sum_{w|(u,w)\in E} a_{uw}$. Let D be the diagonal matrix $\mathrm{diag}(d_1, d_2 \ldots, d_n)$ consisting of weighted degrees of all nodes. Note that by the definition of M, the vector $m_{u\cdot} = \frac{a_{u\cdot}}{d_u}$, giving $M = D^{-1}A$. Thus given any matrix A with non-negative entries, we can obtain the corresponding transition matrix M by this normalization procedure. Also note that if p is a probability distribution on V, then pM denotes the probability distribution after a single step of the random walk.

Definition 1. *A matrix M is said to be diagonally dominant if for every row the magnitude of the diagonal entry is at least as much as the sum of absolute values of all other entries in the row, i.e.*

$$|m_{ii}| \geq \sum_{j \neq i} |m_{ij}| \quad \textit{for all } i$$

A random walk whose normalized adjacency matrix M is *diagonally dominant* is said to be a *lazy* random walk. Note that $m_{ii} \geq 1/2$ for each i and so at each step we remain at the current vertex with probability at least $\frac{1}{2}$. It is known that diagonally dominant matrices with non-negative entries are positive semi-definite [1] and eventually converge to a stationary distribution [2] given by

$$q(u) = \frac{d_u}{\sum_{v \in V} d_v}$$

$\mathbf{p_o}$ denotes the initial probability distribution and $\mathbf{p_t}$ similarly denotes the probability distributions after t steps of the random walk.

$$p_t = p_o M^t$$

Conductance of Sets

We call a clustering *good* if it has many more internal edges than external edges. This notion is formalized by the conductance of a set.

Definition 2. *For a set $S \subseteq V$ such that $\sum_{u \in S} d_u \leq \frac{1}{2} \sum_{v \in V} d_v$, the conductance of the set is*

$$\Phi(S) = \frac{\sum\limits_{u \in S, v \in V \setminus S} a_{uv}}{\sum\limits_{u \in S} d_u}$$

The conductance of the graph $\Phi(G) = \min\limits_{S \subset V} \Phi(S)$.

When the graph G being considered is obvious from context, $\Phi(G)$ is simply denoted by Φ. The conductance of the set S can alternatively be seen as the probability of a random walk on G leaving the set S in a single step, given that the initial probabilities over the vertices are the stationary distribution restricted to S.

Theorem 1. *Given that $p_o(u) = \frac{d_u}{\sum\limits_{v \in S} d_v}$ for $u \in S$, and 0 elsewhere, Pr[leaving S in a single step$] = \Phi(S)$*

Proof. $Pr[\text{leaving S}] = \sum\limits_{u \in S, v \in V \setminus S} P_o(u) m_{uv} = \sum\limits_{u \in S, v \in V \setminus S} \frac{d_u}{\sum_{w \in S} d_w} \frac{a_{uv}}{d_u} = \Phi(S).$ □

Alternatively, the above can be restated as given the initial distribution is P_o, the set S *leaks* or *loses* a probability mass of $\Phi(S)$ in one step. Thus the conductance of the graph represents a bottleneck for random walks to converge to stationary distribution. Let $S*$ denote the set for which $\Phi(S*) = \Phi$. In each step of the random walk starting from P_o defined on $S*$, we lose atmost Φ. Thus to lose a probability mass of $1/4$, we need at least $\frac{1}{4\Phi}$.

Remark 1. *We had initially proved the same bound on the mixing time $t_{mix}(\frac{1}{4}) \geq \frac{1}{4\Phi}$, which is the time for which the distance of the probability from uniform to be $\frac{1}{4}$. This had been derived from the definitions of mixing time, conductance and markov chains (refer to Theorem 7.3 of [3] for the formal statement and proof).*

Lovász-Simonovits Theorem

We now have the required background to state the Lovász-Simonovits Theorem.

Theorem 2. *Let A be a non-negative diagonally dominant matrix corresponding to the adjacency matrix of a graph $G = (V, E)$. Let M be the transition matrix realizing the corresponding random walk. Let p_0 be any initial probability distribution on the vertices, and let $p_t = p_0 M^t$, for each positive integer t. Further, for each t, let π_t be the permutation such that*

$$\frac{p_t(\pi_t(1))}{d_{\pi_t(1)}} \geq \frac{p_t(\pi_t(2))}{d_{\pi_t(2)}} \geq \cdots \geq \frac{p_t(\pi_t(n))}{d_{\pi_t(n)}}$$

For some $T \geq 1$, let

$$\phi := \min_{0 \leq t \leq T} \min_{1 \leq k < n} \Phi(\{\pi_t(1), \pi_t(2), \ldots, \pi_t(k)\})$$

Then for each $W \subseteq V$,

$$|\sum_{w \in W} p_T(w) - q(w)| \leq \min\{\sqrt{x}, \sqrt{\sigma - x}\} \left(1 - \frac{\phi^2}{2}\right)^t,$$

where $x = \sum_{w \in W} d_w$, $\sigma = \sum_{v \in V} d_v$ and q denotes the steady-state distribution.

For some time t and k, the term $\Phi(\{\pi_t(1), \pi_t(2), \ldots, \pi_t(k)\})$ is called a *sweep-cut*. Thus ϕ is the smallest sweep-cut seem by the random walk in the T steps of the random walk. This captures the *local* nature of the bounds we want as opposed to the *global* bounds captured by Φ.

Remark 2. *We know that $\sum_{v \in V} |p_T(v) - q(v)| \leq \sqrt{n}(1 - \lambda_2)^t$. For a proof of this see Chapter 4.3 of [3]. Though the forms look similar, two crucial differences exist in the L-S theorem - the restriction to any arbitrary $W \subseteq V$ and that the summation is inside the modulus. The former makes the L-S theorem stronger, while the latter makes it weaker.*

We are going to prove the Lovász-Simonovits Theorem (Theorem 2) for the special case when the weights are binary, i.e. $a_{uv} \in \{0, 1\}$ for all $u \neq v$, and G is obtained from a simple (self-loop free) undirected graph by replacing each edge by two oppositely directed edges between the same nodes and adding the minimal number of self-loops to make A diagonally dominant. Clearly, this corresponds to the addition of one self-loop for each (directed) edge.[2] In this case, $|E| = 2m$, where m denotes the number of self-loops in the graph. We remark that the present arguments can be readily extended to establish the theorem in the general case.

Given a probability distribution p on V, we define an edge measure $\rho_p(e)$ as the probability that a single step random walk will cross the edge e starting from initial distribution p, i.e.

$$\rho_p(u, v) = \frac{p(u)}{d_u}$$

Note that $\rho_p(u, v)$ depends only on u and hence can be shortened to just $\rho_p(u)$. Note that for the stationary distribution $q(u) = \frac{d_u}{2m}$, we have $\rho_q(e) = \frac{1}{2m}$ for each edge $e \in E$.

We will mostly be working with the distributions $p_t, 0 \leq t \leq T$ in the statement of Theorem 2. Hence, for notational convenience we will often write $\rho_{p_t}(u, v)$ and $\rho_{p_t}(u)$ as simply $\rho_t(u, v)$ and $\rho_t(u)$ in the following. Also, we might drop the distribution p from the subscript when it is evident from context. The key insight of the proof is to study the variation with time of what we define below as the Lovász-Simonowitz curve.

Definition 3. *The Lovász-Simonowitz curve is a function $I : [0, 2m] \to [0, 1]$ associated with a distribution p on V. If ρ is the edge measure induced by p, we order the edges in $E = \{e_1, e_2, \ldots, e_{2m}\}$, as*

$$\rho(e_1) \geq \rho(e_2) \geq \cdots \geq \rho(e_{2m})$$

and define for each integer $k \in [0, 2m]$

$$I(k) = \sum_{i=1}^{k} \rho(e_i)$$

[2] More formally, starting with a simple undirected graph $G' = (V', E')$, we place two directed edges (u, v) and (v, u) in the place of each undirected edge (u, v). If $\hat{d_u}$ denotes the out-degree of a vertex u in this graph, we place exactly $\hat{d_u}$ self loops at u and denote the new degree of this vertex by $d_u = 2\hat{d_u}$. The adjacency matrix A of the final graph $G = (V, E)$ is given by $a_{u,u} = \hat{d_u}$, $a_{u,v} = 1$ iff $(u, v) \in E, u \neq v$ and 0 otherwise.

At the remaining points $x \in (k, k+1)$ (where $k \in \mathbf{Z} \cap [0, 2m]$), $I(x)$ is given by the straight line joining $(k, I(k))$ and $(k+1, I(k+1))$, i.e.

$$I(x) = (k+1-x)I(k) + (x-k)I(k+1)$$

Intuitively, $I_t(k)$ measures how much probability is transported over the k most utilized edges. Here we describe some of the properties of the L-S curve.

1. As the walk converges, I_t should eventually converge to a straight line.
2. $I_t(0) = 0$, $I_t(2m) = 1$. On convergence to stationary, $I_t(x) = \dfrac{x}{2m}$
3. The slope of I_t between k and $k+1$ is given by $I_t(k+1) - I_t(k) = \rho_t(e_{k+1})$.
4. Since all edges going out of a vertex u have the same $\rho_t(u)$, we can order the edges (according to π_t) such that the outgoing edges are consecutive.
5. I_t is non-decreasing and *concave* i.e. $I_t((x+y)/2) \geq (I_t(x) + I_t(y))/2$

By point 4 above, I_t can be viewed as $|V|$ piecewise linear components. We will refer to the end points of these components as the ***hinge points*** of the curve I_t.

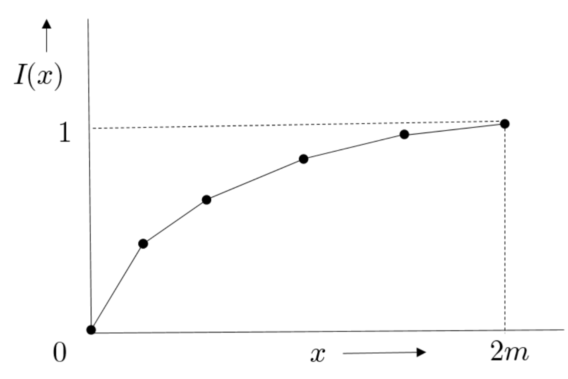

Figure 1: The Lovász Simonovitz Curve

1.3 Proof of the L-S Theorem

Lemma 1. *For any* $0 \le c_1, c_2, \ldots, c_j \le 1$, $1 \le j \le 2m$

$$\sum_{i=1}^{j} c_i \rho(e_i) \le I(\sum_{i=1}^{j} c_i)$$

Proof. This is immediate from the concavity and monotonicity of I. Let $c = \sum_{i=1}^{j} c_i$ denote the sum of weights. Intuitively, we can use the fact that the slope of I is non-negative and non-increasing and hence the weighted sum (with weights at most 1) of slopes in the first j intervals of the form $[k, k+1]$ (for integer k) can only increase if the weights are shifted left to the interval $[0, c]$. $\square$

Before proving the next lemma, note that ρ becomes the uniform (constant) measure at stationarity and hence I becomes a straight line joining $(0, 0)$ and $(2m, 1)$. Since, I_t is concave initially (and for any non-stationary distribution) with the same end points, it seems natural to conjecture that the curve decreases at each step till it finally becomes a straight line. This is indeed the case, as we show next.

Lemma 2. *For any* $0 < t \le T$ *and* k *a hinge point of* I_t,

$$I_t(k) \le I_{t-1}(k)$$

Proof. Let W denote the multiset of source vertices corresponding to the first k edges (E_t) according to permutation π_t, and W' be the corresponding set. Let $N(u) = \{v \in V \mid (u, v) \in E\}$ for each $u \in V$ denote the neighbors of node u. Since k corresponds to a hinge point of I_t, $\{(u, v) \mid u \in W', v \in N(u)\} = \{(u, v) \mid u \in W', (u, v) \in E\} = E_t$. Now

$$\begin{aligned}
I_t(k) &= \sum_{i=1}^{k} \rho_t(e_i) \text{ (where } e_i \text{ are ordered according to } \pi_t) \\
&= \sum_{u \in W} \rho_t(u) \\
&= \sum_{u \in W} \frac{p_t(u)}{d_u} \\
&= \sum_{u \in W'} p_t(u) \text{ (since } k \text{ is a hinge point of } I_t)
\end{aligned}$$

Further, since G is assumed to have been obtained from an undirected graph, $u \in N(v)$ if and only if $v \in N(u)$. This gives

$$\begin{aligned} I_t(k) &= \sum_{u \in W', u \in N(v)} p_{t-1}(v) m_{vu} \\ &= \sum_{u \in W', v \in N(u)} \frac{p_{t-1}(v)}{d_v} \\ &= \sum_{(u,v) \in E_t} \rho_{t-1}(v) \end{aligned}$$

Now we use Lemma 1 about the LS-curve with $j = 2m$, $c_i = 1$ if $e_i \in E_t$ and $c_i = 0$ otherwise. Here $\sum_{i=1}^{j} c_i = |E_t| = k$. This gives

$$\begin{aligned} I_t(k) &= \sum_i c_i \rho_{t-1}(e_i) \\ &\leq I_{t-1}(k) \end{aligned}$$

as desired. $\square$

In the lecture, it was further conjectured that the above proof can be extended to all $k \in [0, 2m]$. With some additional effort, it seems possible to extend the above proof to this case. For sake of continuity, the proof of this fact has been relegated to the appendix. Thus, the curve I_t decreases at all points as t increases, losing its concavity at each step until it finally becomes a straight line at steady-state. We now attempt to get a stronger result, quantifying the decrease with t using the conductance $\Phi(G) = \Phi$. It should then be possible to use this to bound the convergence rate.

Theorem 3. *For every distribution p_0 on V, $0 < t \leq T$ we have the following. If $x \in [0, m]$, then*

$$I_t(x) \leq \frac{1}{2} \left(I_{t-1}(x - 2\phi x) + I_{t-1}(x + 2\phi x) \right)$$

Similarly for each $x \in [m, 2m]$,

$$I_t(x) \leq \frac{1}{2} \left(I_{t-1}(x - 2\phi(2m - x)) + I_{t-1}(x + 2\phi(2m - x)) \right)$$

Proof. It is not a priori clear why the terms on the right hand side of the above inequalities are even well-defined. We will show this later in the proof. Assuming this is indeed the case, the theorem implies that I_t lies below a chord of I_{t-1}. From Figure 2 it is further evident that if ϕ increases, then the chord gets lowered, and hence the convergence rate increases.

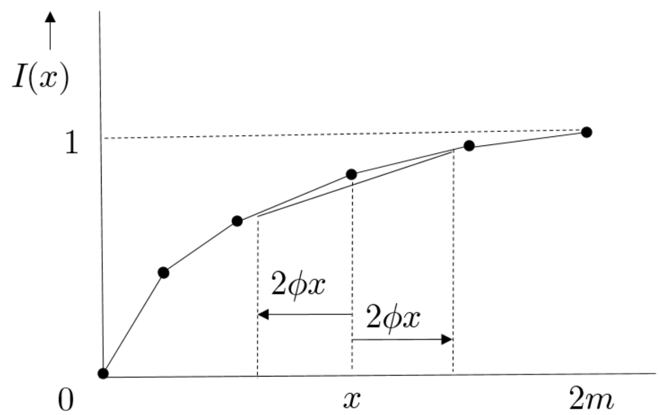

Figure 2: Dependence of convergence on ϕ

We will prove the theorem only for hinge points k. As in Lemma 2 on the monotonic decrease in the LS-curve, this case constitutes the heart of the proof and with some additional care it is possible to extend the result to any $x \in [0, 2m]$.

Let $k = x$. Also let $W = \{u_1, u_2, \ldots, u_k\}$ be the source vertices corresponding to the edges $e_1, e_2, \ldots, e_k$ sorted according to $\pi(t)$, and $v_1, v_2, \ldots, v_k$ be the corresponding destination vertices. As noted in the proof of Lemma 2,

$$I_t(x) = \sum_{i=1}^{k} \rho_t(u_i, v_i) = \sum_{i=1}^{k} \rho_{t-1}(v_i, u_i) \tag{1}$$

We note that the sum on the right is on set of the reversed edges $E' = \{(v_i, u_i) \mid 1 \le i \le k\}$. Clearly, $|E'| = k = x$. To handle both the cases of the theorem simultaneously, we replace W by its complement in the following if $x > m$ and define $y = m - |m - x|$.

Define $W_1 = \{(v_i, u_i) \mid (u_i, v_i) \in W, v_i \neq u_i\}$ as the edges in E' internal to the set W and $W_2 = E' \setminus W_1$ as the union of external edges entering W and self-loops of W.[3] Note that the cut defined by W is a *sweep cut* as defined earlier. Thus, of the x edges starting in W, $\frac{x}{2}$ are self loops and at least ϕy leave W. Thus,

$$|W_1| \leq x - \frac{x}{2} - \phi y = \frac{x}{2} - \phi y$$

and

$$|W_2| = x - |W_2| \geq \frac{x}{2} - \phi y$$

For each edge $e = (v, u) \in W_1$ we associate a unique self-loop e' at node v. We note that $\rho(e) = \rho(e') = \rho(v)$. Let W_1' denote the set of edges in W_1 along with these associated self-loops. Now $\sum_{e \in W_1'} \rho_{t-1}(e) = 2 \sum_{e \in W_1} \rho_{t-1}(e)$ by the previous observation. Applying Lemma 1 again with $j = 2m$ and weights $c_i = 1$ if $e_i \in W_1'$, $c_i = 0$ otherwise, we get

$$\begin{aligned} \sum_{e \in W_1} \rho_{t-1}(e) &= \frac{1}{2} \sum_{e \in W_1'} \rho_{t-1}(e) \\ &\leq \frac{1}{2} I_{t-1}(|W_1'|) \end{aligned}$$

Note that $|W_1'| = 2|W_1| \leq x - 2\phi x$ and I_{t-1} is monotonic, and hence $\frac{1}{2} I_{t-1}(|W_1'|) \leq \frac{1}{2} I_{t-1}(x - 2\phi y)$, giving us part of the right hand side of the desired inequality. But this uses an upper bound on $|W_1|$, and we cannot adapt the above argument for the remainder of the inequality as we only have a lower bound on the size of W_2. We instead use concavity of I in a clever way in the following.

Like the previous argument, we augment W_2 to W_2' by adding edges outside W_2 with corresponding sources. As before we use a bijection between self-loops and non-self-loops (sharing a common source) to get an edge $e' \notin W_2$ for each $e \in W_2$. Explicitly, for each external edge $e = (v, u)$ entering W we add the corresponding self-loop e' (outside of W) and for each self-loop $e \in W_2$ (inside of W) we add the corresponding non-self-loop edge with the same source (starting in W). Thus, W_2' has $2|W_2|$ distinct edges and reasoning as above we get $\sum_{e \in W_2} \rho_{t-1}(e) \leq \frac{1}{2} I_{t-1}(|W_2'|)$. Putting the two inequalities together and using (1), we get

$$I_t(x) \leq \frac{1}{2} I_{t-1}(|W_1'|) + \frac{1}{2} I_{t-1}(|W_2'|)$$

[3] This follows from the hinge point assumption on k and that G was constructed from an undirected graph. Thus E' consists of reversals of exactly the set of edges with source in W.

Also, we have

$$|W_1'| = 2|W_1| \leq x - 2\phi y \leq x + 2\phi y \leq 2|W_2| = |W_2'|$$

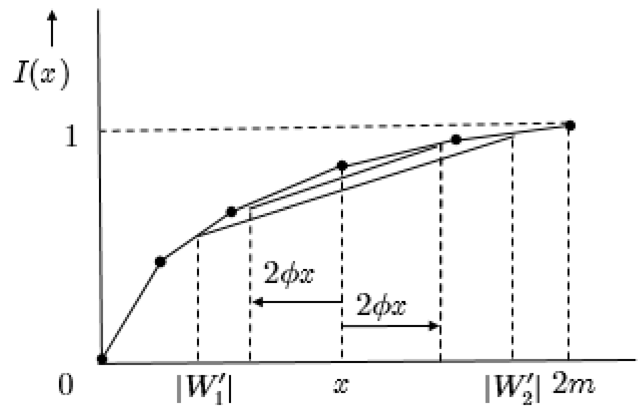

Figure 3: Using concavity of I to prove Theorem 3

Since $0 \leq |W_1'|$ and $|W_2'| \leq 2m$, the above implies that the inequalities in the theorem are well-defined. Also, by concavity of I_{t-1}, the chord joining $(|W_1'|, I_{t-1}(|W_1'|))$ and $(|W_2'|, I_{t-1}(|W_2'|))$ lies below the chord joining $(|x - 2\phi y|, I_{t-1}(x-2\phi y))$ and $(x+2\phi y, I_{t-1}(x+2\phi y))$ (see Figure 3). In particular, this is true for the respective mid-points which coincide at x,[4] and hence

$$I_t(x) \leq \frac{1}{2}I_{t-1}(|W_1'|) + \frac{1}{2}I_{t-1}(|W_2'|) \leq \frac{1}{2}I_{t-1}(x - 2\phi y) + \frac{1}{2}I_{t-1}(x + 2\phi y)$$

□

We now proceed to establish the Lovász-Simonovits Theorem (Theorem 2). To this end, we define a family of functions $\{R_t\}_{t\geq 0}$ inductively as follows

$$R_t(x) = \begin{cases} \min\{\sqrt{x}, \sqrt{2m-x}\} + \frac{x}{2m} & \text{if } t = 0,\ x \in [0, 2m] \\ \frac{1}{2}\left(R_{t-1}(x - 2\phi x) + R_{t-1}(x + 2\phi x)\right) & \text{if } t > 0,\ x \in [0, m] \\ \frac{1}{2}\left(R_{t-1}(x - 2\phi(2m - x)) + R_{t-1}(x + 2\phi(2m - x))\right) & \text{if } t > 0,\ x \in [m, 2m] \end{cases}$$

We now make the following observation.

[4]Note that $\frac{|W_1'|+|W_2'|}{2} = |W_1| + |W_2| = x$, $\frac{x-2\phi y+x+2\phi y}{2} = 2$

Lemma 3. *$I_t(x) \le R_t(x)$ for any $0 \le t \le T$ and $x \in [0, 2m]$.*

Proof. We do an induction on t. For $t = 0$, we have the following three cases

1. $1 < x < 2m - 1$: $I_0(x) \le 1 \le R_0(x)$.
2. $x \in [0, 1]$: $I_0(x) \le x \le \sqrt{x} \le R_0(x)$.
3. $x \in [2m-1, 2m]$: Note that $R_0(x) = \sqrt{2m - x} + \frac{x}{2m}$. Let $y = 2m - x \in [0, 1]$. Then we have

$$\begin{aligned} R_0(x) &= \sqrt{y} + \frac{2m - y}{y} \\ &= 1 + \sqrt{y} - \frac{y}{2m} \\ &\ge 1 + \sqrt{y} - y \\ &\ge 1 \ge I_0(x) \end{aligned}$$

For $t > 0$ we use the inductive hypothesis together with Theorem 3 to conclude that

$$\begin{aligned} I_t(x) &\le \frac{1}{2}\left(I_{t-1}(x - 2\phi y) + I_{t-1}(x + 2\phi y)\right) \\ &\le \frac{1}{2}\left(R_{t-1}(x - 2\phi y) + R_{t-1}(x + 2\phi y)\right) = R_t(x) \end{aligned}$$

where $y = m - |m - x|$. $\square$

To finish the proof, it suffices to now show that

$$R_t(x) \le \min\{\sqrt{x}, \sqrt{2m - x}\}\left(1 - \frac{\phi^2}{2}\right)^t + \frac{x}{2m}$$

For $t = 0$ this holds with equality. For $t > 0, x \in [0, m]$

$$\begin{aligned}
R_t(x) &= \frac{1}{2}\left(R_{t-1}(x - 2\phi x) + R_{t-1}(x + 2\phi x)\right) \\
&\leq \frac{1}{2}\Bigg(\min\{\sqrt{x - 2\phi x}, \sqrt{2m - (x - 2\phi x)}\}\left(1 - \frac{\phi^2}{2}\right)^{t-1} + \\
&\quad \min\{\sqrt{x + 2\phi x}, \sqrt{2m - (x + 2\phi x)}\}\left(1 - \frac{\phi^2}{2}\right)^{t-1}\Bigg) + \frac{x}{2m} \\
&\leq \frac{1}{2}\left(\sqrt{x - 2\phi x}\left(1 - \frac{\phi^2}{2}\right)^{t-1} + \sqrt{x + 2\phi x}\left(1 - \frac{\phi^2}{2}\right)^{t-1}\right) + \frac{x}{2m} \\
&= \frac{1}{2}\left(1 - \frac{\phi^2}{2}\right)^{t-1}\sqrt{x}\left(\sqrt{1 - 2\phi} + \sqrt{1 + 2\phi}\right) + \frac{x}{2m} \\
&\leq \frac{1}{2}\left(1 - \frac{\phi^2}{2}\right)^{t-1}\sqrt{x}\left(\left(1 - \phi - \frac{4\phi^2}{8}\right) + \left(1 + \phi - \frac{4\phi^2}{8}\right)\right) + \frac{x}{2m} \\
&= \frac{1}{2}\left(1 - \frac{\phi^2}{2}\right)^{t-1}\sqrt{x}.2\left(1 - \frac{\phi^2}{2}\right) + \frac{x}{2m} \\
&= \left(1 - \frac{\phi^2}{2}\right)^{t}\sqrt{x} + \frac{x}{2m} \\
&= \min\{\sqrt{x}, \sqrt{2m - x}\}\left(1 - \frac{\phi^2}{2}\right)^{t} + \frac{x}{2m}
\end{aligned}$$

where we have used the binomial expansion identity to obtain the last inequality and $x \in [0, m]$ for the last equality. The case $t > 0, x \in [m, 2m]$ is similar,

$$\begin{aligned}
R_t(x) &= \frac{1}{2}\left(R_{t-1}(x - 2\phi(2m - x)) + R_{t-1}(x + 2\phi(2m - x))\right) \\
&\leq \frac{1}{2}\Bigg(\min\{\sqrt{x - 2\phi(2m - x)}, \sqrt{2m - (x - 2\phi(2m - x))}\}\left(1 - \frac{\phi^2}{2}\right)^{t-1} + \\
&\quad \min\{\sqrt{x + 2\phi(2m - x)}, \sqrt{2m - (x + 2\phi(2m - x))}\}\left(1 - \frac{\phi^2}{2}\right)^{t-1}\Bigg) + \frac{x}{2m}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{2}\left(\sqrt{(2m-x)+2\phi(2m-x)}\left(1-\frac{\phi^2}{2}\right)^{t-1}+\right.\\
&\left.\sqrt{(2m-x)-2\phi(2m-x)}\left(1-\frac{\phi^2}{2}\right)^{t-1}\right)+\frac{x}{2m}\\
&=\frac{1}{2}\left(1-\frac{\phi^2}{2}\right)^{t-1}\sqrt{2m-x}\left(\sqrt{1+2\phi}+\sqrt{1-2\phi}\right)+\frac{x}{2m}\\
&\leq \frac{1}{2}\left(1-\frac{\phi^2}{2}\right)^{t-1}\sqrt{2m-x}\left(\left(1+\phi-\frac{4\phi^2}{8}\right)+\left(1-\phi-\frac{4\phi^2}{8}\right)\right)+\frac{x}{2m}\\
&=\frac{1}{2}\left(1-\frac{\phi^2}{2}\right)^{t-1}\sqrt{2m-x}.2\left(1-\frac{\phi^2}{2}\right)+\frac{x}{2m}\\
&=\left(1-\frac{\phi^2}{2}\right)^{t}\sqrt{2m-x}+\frac{x}{2m}\\
&=\min\{\sqrt{x},\sqrt{2m-x}\}\left(1-\frac{\phi^2}{2}\right)^{t}+\frac{x}{2m}
\end{aligned}
$$

This together with Lemma 3, which bounds the LS-curve from above by R_t, establishes the proof of Theorem 2.

We can extend the proof of the Lovász-Simonovits Theorem to obtain the following corollary.

Corollary 1. *With the notation and assumptions as in Theorem 2, we have*

$$|\sum_{w\in W} p_T(w)-q(w)| \leq \min\{\sqrt{x},\sqrt{\sigma-x}\}\left(1-\frac{\Phi(W)^2}{2}\right)^{t},$$

(See [4] for proof.)

This corollary can be used to design local clustering algorithms for graphs whose running time is nearly linear in the size of the cluster returned. [5]

1.4 Local Clustering Algorithm

We use Theorem 2 for clustering as follows. We will always start with a probability distribution p_0 that has all its weight concentrated on one vertex. If there is a cut S in the graph of conductance less than ϕ^o which is an input to our algorithm, and that vertex is chosen at random from S, Theorem

1 shows that, after $\frac{1}{4\phi^o}$ steps, most of the probability mass will still be in S. This means that the distance from stationary distribution will be large, and ϕ will have to be small by Theorem 2. ϕ can be computed by simply computing the sweep cuts after time T and checking the best of these cuts.

The problem with this approach is that computing all the probabilities will be too slow. In particular after a constant number of steps we have too many nonzero values. One solution proposed by Lovász and Simonovits is to simply zero out the smallest probabilities and prove that it doesnt hurt much.

Local Clustering in Linear Time

Let $\mu(S) = \sum_{u \in S} d_u$ for any $S \subseteq V$. Given a vertex v of G and parameters $0 < \phi < 1$ indicating bound on the conductance cluster and $b \in \mathbf{Z}^+$ governing the size of the cluster returned, we have the following theorem.

Constant	*Value*
l	$\lceil \log_2(\mu(V)/2) \rceil$
t_1	$\lceil \frac{2}{\phi^2} \ln \left(200(l+2)\sqrt{(\mu(V)/2)} \right) \rceil$
t_h	ht_1
t_{last}	$(l+1)t_1$
$f_1(\phi)$	$\dfrac{1}{280(l+2)t_{last}}$

Table 1: Table of constants

Theorem 4. *There exists an algorithm* `Nibble`(G, v, ϕ, b) *which runs in time* $O(2^b(\log^6 m)/\phi^4)$ *the output cluster* $C \subseteq V$ *satisfies (see Table 1 for definitions of* l *and* $f_1(\phi)$*)*

1. *If* $C \neq \emptyset$, $\Phi(C) \leq \phi$ *and* $\mu(C) \leq 5/6\mu(V)$.

2. *Let* $S \subseteq V$ *be such that* $\mu(S) \leq 2/3\mu(V)$ *and* $\Phi(S) \leq f_1(\phi)$. *Then* $\exists S' \subseteq S$ *with* $\mu(S') \geq 1/2\mu(S)$ *such that for any* $v \in S'$ *if* $C =$ `Nibble`$(G, v, \phi, b) \neq \emptyset$, *then* $\mu(C \cap S) \geq 2^{b-1}$

3. S' *can be partitioned into* $S'_0, S'_1, \ldots, S'_l$ *such that if* $v \in S'_b$ *then* $C =$ `Nibble`$(G, v, \phi, b) \neq \emptyset$.

We conclude the section with a presentation of the algorithm `Nibble`(G, v, ϕ, b). For the definitions of various constants in Theorem 1 and Algorithm 1, see Table 1.

Algorithm 1 `Nibble`(G, v, ϕ, b)

$\epsilon \leftarrow \dfrac{1}{1800(l+2)t_{last}2^b}$
$q_0 \leftarrow \chi_v$ (where $\chi_v(u) = \delta_{uv}$)
$r_0 \leftarrow [q_0]_\epsilon$ ($[.]_\epsilon$ denotes truncation of components below ϵ)
for $t = 1$ to t_{last} **do**
  $q_t \leftarrow r_{t-1}M$
  $r_t \leftarrow [q_t]_\epsilon$
  $S_j^t \leftarrow j^{\text{th}}$ sweep set[5] according to q_t
  if $\exists j$ such that **then**
    $\Phi(S_j^t) \leq \phi$,
    $2^b \leq \mu(S_j^t) \leq 5/6\mu(V)$ and
    $I_x(2^b) \geq \frac{1}{140(l+2)2^b}$ (where $I_x = \frac{\partial I_t(x)}{x}$)
    Then $C \leftarrow S_j^t$
  else
    $C \leftarrow \emptyset$
Return C

1.5 Appendix

Lemma 4. *For any $0 < t \leq T$ and $k \in [0, 2m]$,*

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

Proof. Note that since I is a linear extrapolation of its values at integral points, it suffices to establish the lemma for all integral $k \in [0, 2m]$. Clearly, $I_t(k) = I_{t-1}(k) = 0$. So we assume $k > 0$ in the following. As before, let W denote the multiset of source vertices corresponding to the first k edges (E_t) according to permutation π_t, W' be the corresponding set of vertices and $N(u) = \{v \in V \mid (u, v) \in E\}$ for each $u \in V$. Unlike Lemma 2, in this case the vertex $u^* = \text{argmin}_{u \in W'} \rho(u)$ may not have all its outgoing edges in E_t. Also, let $W'' = W' \setminus \{u^*\}$, $E^* = \{(u^*, v) \mid (u^*, v) \in E\}$ and $d^* = |E^* \cap E_t|$.

Now

$$
\begin{aligned}
I_t(k) &= \sum_{i=1}^{k} \rho_t(e_i) \text{ (where } e_i \text{ are ordered according to } \pi_t) \\
&= \sum_{u \in W} \rho_t(u) \\
&= \sum_{u \in W} \frac{p_t(u)}{d_u} \\
&= \sum_{u \in W''} p_t(u) + p_t(u^*)\frac{d^*}{d_{u^*}}
\end{aligned}
$$

Further, since G is assumed to have been obtained from an undirected graph, $u \in N(v)$ if and only if $v \in N(u)$. This gives

$$
\begin{aligned}
I_t(k) &= \sum_{u \in W'', u \in N(v)} p_{t-1}(v) m_{vu} + \sum_{u^* \in N(v)} p_{t-1}(v) m_{vu^*} \frac{d^*}{d_{u^*}} \\
&= \sum_{u \in W'', v \in N(u)} \frac{p_{t-1}(v)}{d_v} + \sum_{v \in N(u^*)} \frac{p_{t-1}(v)}{d_v} \frac{d^*}{d_{u^*}} \\
&= \sum_{(u,v) \in E_t \setminus E^*} \rho_{t-1}(v) + \sum_{(u,v) \in E^*} \frac{d^*}{d_{u^*}} \rho_{t-1}(v)
\end{aligned}
$$

Now we use Lemma 1 with $j = 2m$, and

$$
c_i(e) = \begin{cases} 1 & \text{if } e \in E_t \setminus E^* \\ \frac{d^*}{d_{u^*}} & \text{if } e \in E^* \\ 0 & \text{otherwise} \end{cases}
$$

Note that $\sum\limits_{i=1}^{j} c_i = |E_t \setminus E^*| + |E^*| \frac{d^*}{d_{u^*}} = (k - d^*) + d^* = k$. This gives

$$
\begin{aligned}
I_t(k) &= \sum_i c_i \rho_{t-1}(e_i) \\
&\le I_{t-1}(k)
\end{aligned}
$$

as desired. $\square$

References

[1] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

[2] K. R. James. Convergence of matrix iterations subject to diagonal dominance. *SIAM Journal on Numerical Analysis*, 10(3):478–484, 1973.

[3] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009.

[4] L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. *Random structures & algorithms*, 4(4):359–412, 1993.

[5] D. A. Spielman and S.-H. Teng. A local clustering algorithm for massive graphs and its application to nearly-linear time graph partitioning. *arXiv preprint arXiv:0809.3232*, 2008.