CS-774 Spectral Graph Theory Aug 07 and Aug 10, 2023

Lecture 2: Lovász-Simonovits Curve Technique

Lecturer: Akash Kumar *Scribe: Akash Kumar*

1 An overview: finding sparse cuts using random walks

In the last lecture, we saw that you can find a sparse cut in a graph by finding out the eigenvector $\boldsymbol{x}$ which is paired with the second smallest eigenvalue, λ_2 of the normalized Laplacian, $\overline{\boldsymbol{L}}$. In this lecture, let us denote the normalized Laplacian as $\boldsymbol{\mathcal{L}}$. This eigenvector based approach opens doors to what is possible and returns a cut which is guaranteed to have conductance within $O(\sqrt{\varphi(G)})$ factor of the optimum. Now, just like we do everything in life, we ask for more :-). The question that drives our discussion today is can you recover this low conductance cut really quickly? Of course, you need to spend time proportional to size of the set you are trying to find. So, as a loose benchmark, we ask: can you find an approximation $\widetilde{S}$ to the set S with smallest conductance which runs in time $O(|S|)$? As you can imagine, such a routine will be super handy if the set S is super small. Indeed, you expect real world networks to have small sized cuts. These sets with small conductance capture the real-life meaning of "a community". It is usually taken, in several disciplines, to refer to a set of vertices which is very "well-connected inside" and "loosely connected outside". First things first, let us make sure that the eigenvector based sweep approach cannot find this set in time proportional to only the size of this set. Indeed, you have to compute eigenvectors and even this requires you to write down the entire matrix $\boldsymbol{\mathcal{L}}$.

However, you might have heard of iterative methods to solve eigenvalue/eigenvector problems. Here, the entire idea is to find the eigenvector iteratively by doing power iteration on some appropriate matrix. Luckily for us, power iterations on $\boldsymbol{\mathcal{L}}$ bear a close resemblance to power iterations on the random walk matrix $\boldsymbol{M}$. Indeed, you might recall that these two matrices share the same eigenbasis for a (d-regular) graph and the eigenvalues of these two matrices also bear a linear relationship. So, it is fair game to expect power iterations on $\boldsymbol{M}$ to reveal information about results of running a power iteration on $\boldsymbol{\mathcal{L}}$. And power iterations on $\boldsymbol{M}$ can be simulated by just doing random walks! Does this mean we can expect a random walk based approach to be able to find sparse cuts in graphs? As you will see, in a truly remarkable feat, Spielman-Teng [ST12] made this high level plan work. Just one final note in this (super) high level story. Perhaps this is not that unexpected. Indeed, you might expect that if you truly have a low conductance cut (say community of people who play candy crush on the Facebook graph – this is a low conductance cut, you guys don't have too many friends) then a random walk which is started at a vertex within the community is unlikely to ever leave this community. And thus after doing enough of these walks and collecting all the end-points reached, intuitively you should be able to find a low conductance cut. Ok, *enough chitchat*, let us setup the stage to work out the relevant math.

2 The Lovász-Simonovits Theorem

Consider a connected graph $G' = (V, E')$. Define a (directed) graph $G'' = (V, E'')$ where G'' is obtained by taking an edge $(u, v) \in E'$ and making two copies of it (one in each direction). Finally, define a (lazy) graph $G = (V, E)$ where G is obtained by adding d_u loops at vertex u where d_u is the outdegree of u. Let $|E| = 2m$ where m denotes the number of self-loops in this graph. Recall from last lecture that the distribution $\boldsymbol{\pi}$ where $\boldsymbol{\pi}(v) = deg(v)/2|E|$ satisfies the detailed balance condition for G and thus $\boldsymbol{\pi}$ is the stationary distribution for random walks on G. Let $\boldsymbol{\pi}$ denote the "current distribution". For a vertex $u \in V$ and $v \in N(u)$, the probability mass sent by u along the edge (u, v) with respect to the current distribution $\boldsymbol{\pi}$ equals $\frac{\boldsymbol{\pi}(u)}{deg(u)}$. Since $\boldsymbol{\pi}(u) \propto d_u$ (where d_u is the degree of u in G' or the out-degree of u in G''), one notices that this distribution is uniform on all the edges of G. That is,

$$p(u, v) = \boldsymbol{\pi}(u)/deg_G(u) = \boldsymbol{\pi}(u)/2d_u = 1/2m$$

π yields uniform dist.

for all directed edges out of u in G.

Next, take a distribution $\boldsymbol{p}$ (different from $\boldsymbol{\pi}$) supported on vertices of G. This distribution also induces a natural distribution $\boldsymbol{\rho}$ on $E(G)$. For any edge e directed out of $u \in V$, we have the probability mass about to be dispatched along the edge e (with respect to $\boldsymbol{p}$) equals

$$\boldsymbol{\rho}(e) = \frac{\boldsymbol{p}(u)}{d_G(u)}.$$

Note that this value, the probability dispatched by u along any edge out of u, does not depend on the other end-point of this edge. This is nice, as we now have a distribution on edges coming from $\boldsymbol{p}$ and the other from $\boldsymbol{\pi}$. One natural approach to show convergence of random walks to the distribution $\boldsymbol{\pi}$ on vertices is to instead show convergence to the distribution on edges induced by $\boldsymbol{\pi}$. We would like to show convergence time bounds (and find sparse cuts in G) by using this framework of thinking about these induced edge distributions instead. We already discussed an approach to bounding the mixing time in the last lecture (which was expressed in terms of the gap $\lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M})$, where $\boldsymbol{M} = \boldsymbol{M}(G)$ denotes the random walk matrix on G). This approach tracked a single number: the ℓ_2 distance to the stationary at the current step. The key idea of Lovász and Simonovits was the suggestion that instead of tracking a single number, you could track an entire curve and use this information to bound the mixing time.

So, following Lovász-Simonovits, we now define a "greedy" potential function.

Definition 2.1. *Let $\boldsymbol{p}$ denote a distribution on $V(G)$. Denote the distribution induced on $E(G)$ as $\boldsymbol{\rho}$. Sort the edges in decreasing order of mass dispatched along the eges (with respect to $\boldsymbol{p}$) to get*

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

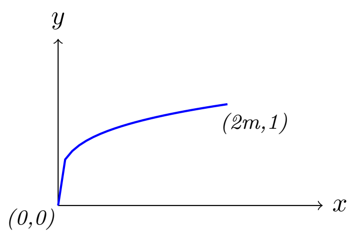
Then the discrete Lovász Simonovits potential, denoted $G_{\boldsymbol{p}}$ is a function defined as follows:

$$G_{\boldsymbol{p}}\{0, 1, 2, \cdots, 2m\} \to [0, 1] \text{ where } G_{\boldsymbol{p}}(x) = \sum_{i=1}^{x} \boldsymbol{\rho}(e_i).$$

We define the Lovász-Simonovits potential $g_{\boldsymbol{p}}$ as an extension of $G_{\boldsymbol{p}}$ to the domain $[0, 2m]$ by linearly interpolating $G_{\boldsymbol{p}}$ between two successive integer values $i, i+1$ where $i \in \{0, 1, 2, \cdots, 2m-1\}$. Often, we will also write $g_{\boldsymbol{p}}$ as $g_{\boldsymbol{\rho}}$.

We first make a few remarks to help digest this definition.

Remark 2.2. Note that the potential defined above is truly a greedy potential. After all, it "greedily aggregates" the x heaviest elements with respect to $\boldsymbol{\rho}$ and sums them up which results in $g_{\boldsymbol{p}}(x)$. One thing you might find annoying is the use of x to count number of edges [1]. Finally, you note that the curve $g_{\boldsymbol{p}}$ is increasing (or non-decreasing) and concave (why?). Here is a pictorial illustration.



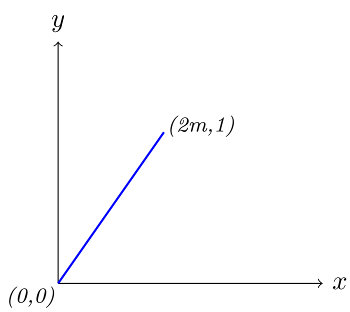
Note that $g_{\boldsymbol{p}}(0) = 0$ and $g_{\boldsymbol{p}}(2m) = 1$.

Next, I would like you to verify the following.

Problem 2.3. *Let $\boldsymbol{\pi}$ denote the stationary distribution of random walks on $G = (V, E)$ as defined above. Show that*

$$g_{\boldsymbol{\pi}}(x) = \frac{x}{2m}, \quad \forall 0 \leq x \leq 2m.$$

That is, it is given by the following picture.



We now solve a problem through which we confirm an interesting fact. The Lovász-Simonovits potential gives some kind of special treatment to the stationary distribution, $\boldsymbol{\pi}$.

Claim 2.4. *For any distribution $\boldsymbol{p}$ on $V(G)$, the curve $g_{\boldsymbol{\pi}}$ lies pointwise below $g_{\boldsymbol{p}}$.*

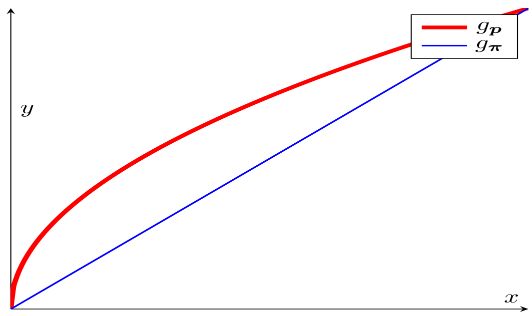
Proof of Claim 2.4. Indeed, this problem is fairly straightforward to show if you understand the definition of Lovász-Simonovits curves and that they have this defining feature of being constructed by this greedy aggregation. You just note the following

- All curves start at $(0, 0)$ and end at $(2m, 1)$.

[1] Get over it :-) It is not such a big thing

- You have $g_{\boldsymbol{p}}(1) > g_{\boldsymbol{\pi}}(1)$, that is the curve $g_{\boldsymbol{p}}$ starts out with a bigger slope.

Finally, you recall that all of these curves are concave. Thus, if the curve $g_{\boldsymbol{p}}$ were to ever dip below $g_{\boldsymbol{\pi}}$, it cannot reach all the way to $(2m, 1)$ without violating concavity. And this solves the problem above. The picture below illustrates this. $\square$



Alright, this is great. We now follow Lovász-Simonovits and ask: is it possible to argue convergence of random walks using this picture above? Namely, letting $\boldsymbol{p}_t^T = \boldsymbol{p}_{t-1}^T \boldsymbol{M}$ (where $\boldsymbol{M}$ denotes the random-walk matrix for G), is it conceivable that $g_{\boldsymbol{p}_t} \to g_{\boldsymbol{\pi}}$ as $t \to \infty$? As I already told you, the answer is yes. The rest of this lecture is focused on showing just that.

2.1 Convergence of Lovász-Simonovits curves

Let us start with a few notational simplications. First off, we will say LS instead of Lovász-Simonovits from now on. Additionally, we use another shorthand.

Notation 2.5. *Let $\boldsymbol{p}_0$ denote a point distribution of the form $\mathbf{1}_v$ for some $v \in V$. We let $\boldsymbol{p}_t^T = \boldsymbol{p}_{t-1}^T \boldsymbol{M}$ and we will denote the t-step LS curve as $g_t := g_{\boldsymbol{p}_t}$.*

We will prove Lemma 2.6 as a warmup towards showing this convergence of LS curves. It just asserts that the LS curves only fall downward with time (actually, the claim is *weaker*: it shows that the later LS curves don't rise above the former curves).

Lemma 2.6. *For every $t \geq 1$, for every $x \in [0, 2m]$, it holds that*

$$g_t(x) \leq g_{t-1}(x).$$

It would be helpful to make a definition before we proceed.

Definition 2.7. *Let $\boldsymbol{p}$ denote a distribution on $V(G)$ and let $\boldsymbol{\rho}$ denote the corresponding distribution induced on edges. Arrange edges in the order*

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

where all the out-edges incident on a vertex $u \in V$ occur consecutively in this sequence (of course, these out edges occur in an arbitrary order). An integer $0 \leq x \leq 2m$ is called a hinge point with respect to $\boldsymbol{\rho}$ (or $\boldsymbol{p}$) if the x heaviest edges under $\boldsymbol{\rho}$ correspond to all the out edges (including loops) incident on some subset S of vertices. If $|S| = i$, the integer x is called the ith hinge point on $g_{\boldsymbol{\rho}}$. We call this order on edges $e_1, e_2, \cdots e_{2m}$ a hinge respecting order.

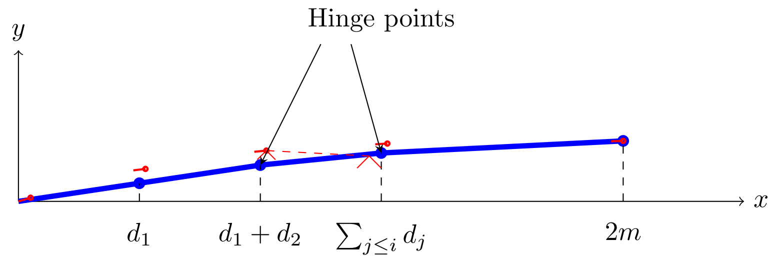


Figure 2.1: If g_{t-1} dominates g_t on all hinges of g_t then g_{t-1} cannot dip below g_t anywhere else without violating concavity.

As you can see, the hinges are merely a convenient way to organize the edges with respect to the distribution $\boldsymbol{\rho}$. Note that any curve $g_{\boldsymbol{\rho}}$ has exactly $|V(G)|$ hinges. We first argue that it suffices to prove the following claim.

Claim 2.8. *Fix $t \geq 1$ and let $I \subseteq [2m]$ denote the subset of hinges on g_t. Suppose*

$$\forall x \in I, \text{ we have } g_t(x) \leq g_{t-1}(x).$$

Then for all $0 \leq x \leq 2m$, we have $g_t(x) \leq g_{t-1}(x)$.

Proof of Lemma 2.6.[Assuming Claim 2.8] Convexity (rather concavity) argument. We will show that if g_t is below g_{t-1} on two successive hinges of g_t, then it must be below g_{t-1} at all the intermediate points. To this end, recall that both the LS curves, g_{t-1} and g_t start at $(0,0)$ and wind up at $(2m,1)$. Consider two consecutive hinges on g_{t-1}, say x_i and x_{i+1} which are the ith and the $(i+1)$th hinge points. Take $y \in [x_{i-1}, x_i]$. Suppose for contradiction sake that $g_t(y) > g_{t-1}(y)$. We know $g_t(x_i) \leq g_{t-1}(x_i)$ and $g_t(x_{i-1}) \leq g_{t-1}(x_{i-1})$. This means there is a hinge y' of g_{t-1} which lies between x_{i-1} and x_i as well. But this would mean g_{t-1} is not concave in the interval $[y', x_i]$. Here is a picture that assists the proof.

□

So, to finish off proving Lemma 2.6, it suffices to show Claim 2.8. We do that now.

Proof of Claim 2.8. Call the distribution on $E(G)$ with respect to distribution $\boldsymbol{p}_t$ on vertices $\boldsymbol{\rho}$ and suppose $\boldsymbol{\rho}(e_1) \geq \boldsymbol{\rho}(e_2) \geq \cdots \geq \boldsymbol{\rho}(e_{2m})$ where the edges are ordered in some hinge respecting order. Write $e_i = (u_i, v_i)$ where several edges might share the same u_i. Let x denote a hinge point on $\boldsymbol{p}_t$ corresponding to some subset S of vertices.

$$\begin{aligned} g_t(x) = \sum_{i=1}^{x} \boldsymbol{\rho}(e_i) = \sum_{i=1}^{x} \boldsymbol{\rho}(u_i, v_i) = \sum_{u \in S} \boldsymbol{p}_t(u) & \\ \overset{(\mathbf{1})}{=} \sum_{i=1}^{x} \boldsymbol{\rho}_{t-1}(\overleftarrow{e_i}) & \quad \overleftarrow{e_i} = (v_i, u_i) \text{ denotes the reverse edge to } e_i \\ \overset{(\mathbf{2})}{\leq} g_{t-1}(x) & \quad \text{by greedy property of LS curves} \end{aligned}$$

A little more detail on the above derivation. Here, (**1**) holds because the mass $\boldsymbol{p}_t(u)$ deposited on each $u \in S$ is precisely the mass which reached those vertices along the in-edges to u at the previous step. For (**2**), note that to evaluate g_{t-1} at x you must grab x heaviest edges under $\boldsymbol{\rho}_{t-1}$ and add up the probabilities on those edges. Here, we are grabbing *some* x edges which are reversals of edges incident on S which need not be the x heaviest edges with respect to $\boldsymbol{\rho}_{t-1}$ and therefore this step follows. □

2.2 A Brief Interlude

We pause the main story for a while where the plan is to go beyond the warmup done in Lemma 2.6 and show that the curve should fall significantly from one step to another if you know something about the graph (like the graph expands). Through this brief interlude, we will first solidfy our understanding of the LS potential function – which is now possible thanks to our little warmup proving Lemma 2.6. So, let us ask the following question.

Problem 2.9. *Let $\boldsymbol{p}_t$ denote the tth step distribution on $V(G)$ and $\boldsymbol{\rho}_t$ denote the corresponding distribution induced on $E(G)$ where*

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

denotes a hinge respecting order with respect to $\boldsymbol{\rho}$. Then can you interpret $g_t(x)$ as solution to some natural ILP?

It should not surprise you that there is indeed such an ILP. Indeed, the LS curve aggregates edges greedily by heaviness with respect to $\boldsymbol{\rho}$. Taking inspiration from the fractional knapsack problem, and writing c_{e_i} for ith component of $\boldsymbol{c}$, we get:

$$\begin{aligned} g_t(x) = \text{maximize } & \vec{\boldsymbol{c}} \cdot \vec{\boldsymbol{\rho}} \\ & \text{subject to} \quad 0 \leq c_{e_i} \leq 1 \;\; \forall 1 \leq i \leq 2m \\ & \sum_i c_{e_i} = x \end{aligned}$$

For those among you who are not yet comfortable with ILPs, let me present this result in a different/more elementary way.

Lemma 2.10. *For every $c_1, c_2, \cdots, c_{2m}$ such that $0 \leq c_i \leq 1$,*

$$\sum_{i=1}^{2m} c_i \left(\boldsymbol{\rho}(e_i)\right) \leq g_t(\sum_{i=1}^{2m} c_i).$$

You should prove this on your own. I will give a hint: recall that LS curves are built greedily and in some sense the right hand side is the greedy maximum you get for the left hand side. Okay, this finishes the brief interlude, let's get back to the main story.

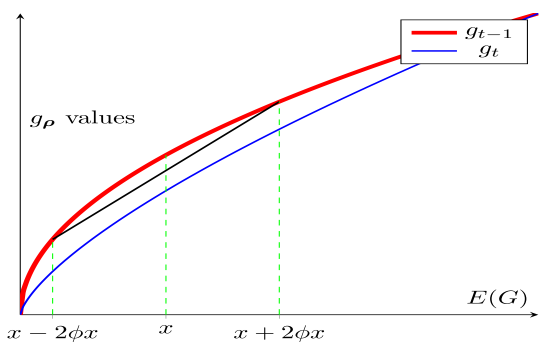
2.3 Resuming the LS convergence saga

There are two major results in this section – Theorem 2.11 and Theorem 2.12. Theorem 2.11 identifies a certain recurrence the LS potential functions satisfy which formalizes our intuition about the drops of the LS curve from one timestep to the next. Theorem 2.12 solves this recurrence to give a form more readily usable in applications. We start by stating Theorem 2.11.

Theorem 2.11. *Let $G = (V, E)$ be the graph defined before and suppose $\varphi(G) = \phi$. For every start distribution $\boldsymbol{p}_0$, for every $t \geq 1$, and every hinge point x of g_t, we have the following:*

- *If $x \leq m$, then*
$$g_t(x) \leq \frac{g_{t-1}(x - 2\phi x) + g_{t-1}(x + 2\phi x)}{2}.$$
- *If $x > m$, then*
$$g_t(x) \leq \frac{g_{t-1}(x - 2\phi(2m - x)) + g_{t-1}(x + 2\phi(2m - x))}{2}.$$

To understand what this theorem is saying, consider the following picture.



In this picture, we have two successive LS curves, g_{t-1} and g_t. What you notice is that the entire curve at time t drops significantly below the curve at time $t - 1$. And this drop is large if the conductance is large. Indeed, if the conductance is at least ϕ, you can draw chords of length $4\phi x$ below x such that at the point x, g_t is below this chord[2](!) This matches well with what we intended to show: if the conductance is large these curves not only fall, they fall *royally* and the entire sequence of curves rapidly heads towards the curve $g_{\boldsymbol{\pi}}$. One shortcoming of this theorem is that it is not fully transparent what the rate of convergence is. The theorem below teases out this rate and provides the convergence guarantee in a form which is mathematically more transparent.

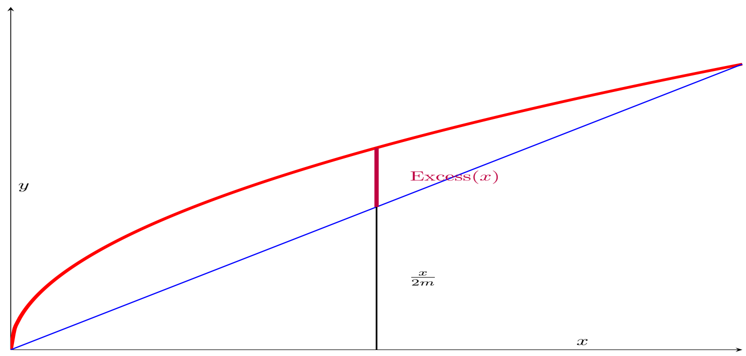
Theorem 2.12. *Let $G = (V, E)$ be a graph as defined above. Suppose $\varphi(G) \geq \phi$. Then, for every distribution $\boldsymbol{p}_0$, every $t \geq 1$ and every $x \in [0, 2m]$, we have*

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

[2]For technical reasons, we show this only for hinge points on g_t. However, this is enough to show convergence as you will see in the proof of Theorem 2.12

We know that any distribution pointwise dominates the uniform distribution on the edges. In particular, this means the LS value of a distribution $\boldsymbol{\rho}$ on the edges at a point $0 \leq x \leq 2m$ is always at least $x/2m$ (which is the value of LS curve for uniform distribution at x). Theorem 2.12 asserts that if the conductance is large, this value is only slightly bigger (additively) and it is within an at most $Excess(x) \leq \sqrt{x}\left(1 - \frac{\phi^2}{2}\right)^t$ of $x/2m$ at time t provided the graph is an ϕ-expander. Let us have another picture to appreciate what this theorem asserts.

i.e. at $t\uparrow$, we eventually encounter low conductance cut causing large fall in g.



We delegate the proof of these last two theorems to the following two subsubsections.

2.3.1 Proof of Theorem 2.11

Proof. This proof is what we call a "mass reshuffling argument." I will only prove item 1. Item 2 follows by an analogous argument. Let x denote a hinge point of g_t. Then just like we did in proof of Lemma 2.6 (actually, proof of Claim 2.8), we have

$$\begin{aligned}
g_t(x) &= \sum_{i=1}^{x} \boldsymbol{\rho}(e_i) \\
&= \sum_{i=1}^{x} \boldsymbol{\rho}(u_i, v_i) \\
&= \sum_{u \in S} \boldsymbol{p}_t(u) \\
&= \sum_{i=1}^{x} \boldsymbol{\rho}_{t-1}(\overleftarrow{e_i}) \quad \overleftarrow{e_i} \text{ denotes the reverse edge to } e_i
\end{aligned} \tag{2.1}$$

We just skipped the last step where we (loosely) upper bounded the last expression in the above chain of equalities as $g_{t-1}(x)$. This time around, we know the graph is an ϕ-expander and we would like to exploit this fact to get a better bound. Let S denote the set of vertices corresponding to which x is the hinge point on g_t. That is, $\sum_{u \in S} deg_G(u) = 2d_u = x$. Let

$$E_x = \{(u_i, v_i)\}_{u_i \in S} \text{ i.e., } \textit{set of out-edges or loops incident on some vertex in } S$$

and W is obtained by reversing them all. That is,

$$W = \{(v_i, u_i) \colon (u_i, v_i) \in E_x\}.$$

We divide this set of *reversed edges* W into two groups:

$$\begin{aligned} W_1 &= \{(v_i, u_i) \colon u_i, v_i \in S, u_i \neq v_i\} && \text{non loop edges fully in S} \\ W_2 &= \{(v_i, u_i) \colon u_i \in S, v_i \notin S\} \bigcup \{(u_i, u_i) \colon u_i \in S\} && \text{edges coming into S plus all loops} \end{aligned} \tag{2.2}$$

At this stage, you might want to try the following: write $\sum_{(u,v)\in W} \boldsymbol{\rho}_{t-1}(u,v) = \sum_{(u,v)\in W_1} \boldsymbol{\rho}_{t-1}(u,v) + \sum_{(u,v)\in W_2} \boldsymbol{\rho}_{t-1}(u,v)$. Note that $|W_2| \geq x/2 + \phi x$ and thus, $|W_1| \leq x/2 - \phi x$. By the greedy property of LS curves, you get

$$\sum_{(u,v)\in W_1} \boldsymbol{\rho}_{t-1}(u,v) \leq g_{t-1}(|W_1|) \leq g_{t-1}(x/2 - \phi x).$$

This is not quite what we want. We would like to "pull the one-half" out of this expression[3]. In particular, we would like to show the following:

$$\begin{aligned} \sum_{(v,u)\in W_1} \boldsymbol{\rho}_{t-1}(v,u) &\leq 1/2 \cdot g_{t-1}(x - 2\phi x) \\ \sum_{(v,u)\in W_2} \boldsymbol{\rho}_{t-1}(v,u) &\leq 1/2 \cdot g_{t-1}(x + 2\phi x) \end{aligned} \tag{2.3}$$

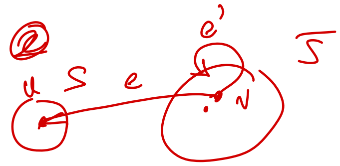
As you will see, our final proof ends up diverging from this plan. Let us press on, we will see where we diverge. Towards showing the first bound above, following Lovàsz and Simonovits, let us introduce a new set of edges. To this end, take an edge $e = (v,u) \in W_1$ and take a unique self loop e' at v and throw them both in a new set, W_1'. Note $|W_1'| = 2|W_1|$ and that $\boldsymbol{\rho}_{t-1}(e) = \boldsymbol{\rho}_{t-1}(e')$. In particular, this means $\sum_{e\in W_1'} \boldsymbol{\rho}_{t-1}(e) = 2\sum_{e\in W_1} \boldsymbol{\rho}_{t-1}(e)$. By the greedy rule again,

$$\sum_{e\in W_1} \boldsymbol{\rho}_{t-1}(e) = \frac{1}{2} \cdot \sum_{e\in W_1'} \boldsymbol{\rho}_{t-1}(e') \leq \frac{1}{2} \cdot g_{t-1}(|W_1'|)$$

Note that one can pull out the one-half factor which gives us the desired bound on the first term as the last term above satisfies $g_{t-1}(|W_1'|)) \leq g_{t-1}(x - 2\phi x)$[4]. Lovàsz and Simonovits resist this temptation *(and this is where we diverge from the plan)* as the same argument cannot be done with W_2 as we only know $|W_2| \geq x/2 + \phi x$ and we cannot use monotonicity for showing such a bound. Instead, this will follow from a clever mass shuffling together with concavity arguments.

[3]When life gives you lemons, make lemonade

[4]by monotonicity of LS curves and recalling $|W_1'| \leq x - 2\phi x$



$|w_2'| = 2|w_1|$

We again define a new set W_2' using the set W_2. Recall, W_2 contains loops and edges $e \in E(\overline{S}, S)$. As before, we take a non-loop edge $e = (v, u) \in W_2$ and we pair it with a unique loop e' at v. Next, we take a loop $e = (u, u) \in W_2$ and we pair it with a unique edge e' directed out of u[5]. We add all these e, e' pairs in W_2'. By an entirely similar argument as before, we get

↳ $\#e = \#$loops by construction.

$$\sum_{(v,u) \in W_2} \boldsymbol{\rho}_{t-1}(v, u) \leq \frac{1}{2} \cdot g_{t-1}(|W_2'|).$$

Thus, $g_t(x)$ lies below the "midpoint of a longer chord"[6] on g_{t-1} – namely the chord connecting

$$p_1 = (|W_1'|, g_{t-1}(|W_1'|)) \text{ and } p_2 = (|W_2'|, g_{t-1}(|W_2'|)) .$$

And therefore, $g_t(x)$ will also lie below the "mid-point of a shorter chord" which connects

$$q_1 = (x - 2\phi x, g_{t-1}(x - 2\phi x)) \text{ and } q_2 = (x + 2\phi x, g_{t-1}(x + 2\phi x)) .$$

□

For those of you who want it, here is Claim 2.13 for completeness. Note that this claim is not standalone and it will only help if you read up proof of Theorem 2.11 above.

Claim 2.13. *Let x, W_1, W_2, W_1', W_2' be as defined in the above argument. Then*

$$|W_1'| + |W_2'| = x.$$

Proof of Claim 2.13. Recall that

$$\begin{aligned} |W_1'| = 2|W_1| &= 2 \cdot \# \text{ of non-loop edges internal to S} \\ &= \# \text{ of non-loop edges internal to S} + \# \text{ loops internal to S} \\ &= x - |E(S, \overline{S})| \end{aligned}$$

Since each non-loop edge has a reverse copy, $|E(S, \overline{S})| = |E(\overline{S}, S)|$ in our directed graph G. Next up, note that

$$\begin{aligned} |W_2'| = 2|W_2| &= 2 \cdot \# \text{ loops internal to S} + 2 \cdot |E(\overline{S}, S)| \\ &= \# \text{ loops internal to S} + \# \text{ non-loop edges internal to S} + 2 \cdot |E(\overline{S}, S)| \\ &= x - |E(S, \overline{S})| + 2 \cdot |E(\overline{S}, S)| \\ &= x + |E(S, \overline{S})| \end{aligned}$$

where recall the last step follows because $|E(S, \overline{S})| = |E(\overline{S}, S)|$. And therefore, it follows that x is the average of $|W_1'|$ and $|W_2'|$. □

[5] why is this pairing possible?

[6] If you are worried why x is the midpoint of this chord, try proving it yourselves. The proof is given in Claim 2.13 anyway

2.3.2 Proof of Theorem 2.12

We introduce a new curve,

$$R_0(x) = \min(\sqrt{x}, \sqrt{2m - x}) + x/2m.$$

It is easy to check that $\forall x \in [0, 2m], g_0(x) \le R_0(x)$. We use define a new curve R_t where R_t's are obtained by using the same recurrence that LS curves obey. In particular, we define

$$R_t(x) = \begin{cases} \frac{1}{2}\left(R_{t-1}(x - 2\phi x) + R_{t-1}(x + 2\phi x)\right) & \text{if } x \le m \\ \frac{1}{2}\left(R_{t-1}(x - 2\phi(2m - x)) + R_{t-1}(x + 2\phi(2m - x))\right) & \text{Otherwise} \end{cases}$$

We will now show the following:

1. First, we show
$$\forall x, R_t(x) \le \min(\sqrt{x}, \sqrt{2m - x}) + x/2m.$$
2. Second, we show that $g_t(x) \le R_t(x)$ holds at all hinges of g_t.
3. Last, note that the above item implies that $g_t(x) \le R_t(x)$ holds for all x. Indeed, both the functions are concave and increasing[7]. Indeed, R_t cannot dip below g_t between two hinges of g_t without violating concavity.

Thus, it suffices to show the first two items above. Let us start with the second item. We will show this only for hinge x of g_t where $x \le m$. We have,

$$\begin{aligned} g_t(x) &\le 1/2 \cdot (g_{t-1}(x - 2\phi x) + g_{t-1}(x + 2\phi x)) \\ &\le 1/2 \cdot (R_{t-1}(x - 2\phi x) + R_{t-1}(x + 2\phi x)) \\ &= R_t(x) \end{aligned}$$

Great. Now, let us show item 1 from our plan above. We again show this for $x \le m$ and leave the other part as an exercise. The right hand side is:

#ask why?

$$\begin{aligned} RHS &= \frac{1}{2}\left(R_{t-1}(x - 2\phi x) + R_{t-1}(x + 2\phi x)\right) \\ &\le \frac{1}{2} \cdot \left(1 - \frac{\phi^2}{2}\right)^{t-1} \cdot \Big(\min(\sqrt{x - 2\phi x}, \sqrt{2m - (x - 2\phi x)}) + \min(\sqrt{x + 2\phi x}, \sqrt{2m - (x + 2\phi x)})\Big) + x/2m \\ &\le \frac{1}{2}\left(\sqrt{x - 2\phi x} + \sqrt{x + 2\phi x}\right)\left(1 - \frac{\phi^2}{2}\right)^{t-1} + x/2m \\ &= \frac{1}{2}\left(1 - \phi^2/2\right)^{t-1} \cdot \sqrt{x}\left(\sqrt{1 - 2\phi} + \sqrt{1 + 2\phi}\right) + x/2m \\ &\le \sqrt{x}\left(1 - \phi^2/2\right)^t + x/2m \end{aligned}$$

taylor

[7] Q: Why are R_t's concave? This is a good exercise you should think about

You should verify the last step using Taylor expansions which tells you $1/2(\sqrt{1-2\phi}+\sqrt{1+2\phi}) \leq 1-\phi^2/2$ which holds for all $\phi \in (0, 1/2)$.

References

[ST12] Dan Spielman and Shanghua Teng, "A Local Clustering Algorithm for Massive Graphs and its Applications to Nearly-Linear Time Graph Partitioning," *SICOMP*, 2012