DRP Presentation: Spectral Graph Theory

Vivien Goyal

November 20, 2024

Outline

Expander Graphs

Expanders are sparse highly-connected graphs.

Constant degree (k regular, or close to that)

An expander graph on n vertices has $O(n)$ edges and diameter $O(log(n))$.

Sparsification is the problem of how can we remove edges from a graph while retaining high conductance - i.e, the process of turning non-expander graphs into expander graphs.

Spectral Graph Theory

A graph is an adjacency matrix with real-valued entries.

Undirected graphs have symmetric matrices, so spectral theorem applies and we have a basis of eigenvectors.

We can encode a graph as a matrix and then analyse the spectrum (eigenvals/eigenvecs)

Eigenvalues, Connected Components, and Random Walks

The multiplicity of the 0 eigenvalue measures the number of connected components.

$Lx = 0$ requires $x_i = x_j$ whenever i, j in the same component, and an orthogonal set of (nonzero) vectors which all satisfy this are the indicator vectors for the connected components.

Adjacency matrix (divided by degree) raised to powers gets you the random walk distribution.

So if $R = D^{-1}A$, $R^k_{i,j}$ is the probability that a k step walk from i ends at j.

Laplacians

D is the degree matrix of a graph: $D_{u,u} = \deg u$, and A is the adjacency matrix, with $A_{u,v} = w_{u,v}$

Quadratic Form of a matrix: $\langle x, Ax \rangle$ or $x^T Ax$.

$L = D - A$, so $L_{u,v} = -w_{u,v}$ if edge u, v exists and $L_{u,u} = \deg u$ along the diagonals. It encodes both the degree and adjacency matrices.

It can also be decomposed into a sum of simpler matrices $\sum_{u,v \in E} L_{u,v}$,

$\langle x, Lx \rangle = \sum_{u,v \in E} w_{u,v}(x_u - x_v)^2$, a useful result for later.

If we consider $\mathcal{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$, we notice that all of its entries are between -1 and 1. If G is d regular, $\mathcal{L} = \frac{1}{d} L$

Conductance

The *conductance* of a subgraph (precisely, a subset of vertices) S measures how connected S is to the rest of the graph.

$\Phi(S) = \frac{|E(S, V-S)|}{Vol(S)}$

The conductance of the graph is just $\Phi_G = \min_{S \subset V} \Phi(S)$, it measures the conductance of the most "island-y" subgraph (most disconnected from the rest of the graph).

If we take a subset of vertices S and write it as an indicator vector x_S, $\Phi(S) = \frac{\langle x_s, Lx_s \rangle}{\sum_{i \in V} \deg i x_s(i)^2} = \frac{\langle x_s, Lx_s \rangle}{d\langle x_s, x_s \rangle}$.

Numerator works out as a result of symmetry: x_u and x_v cancel out if both are in S, so only edges counted are those between S and $V - S$.

So Φ_G is the result of minimising this over all indicator vectors, i.e over $\{0, 1\}^n$.

Cheeger's Inequality

$\langle x, \mathcal{L}x\rangle = \langle x, \sum_i \lambda_i \langle \mathcal{L}x, e_i\rangle e_i\rangle = \sum_i \lambda_i \langle \mathcal{L}x, e_i\rangle^2 \leq \lambda_1 ||x||$. So quadratic forms are bounded by minimum and maximum eigenvalues.

$\mathcal{L}D^{\frac{1}{2}}\mathbf{1} = L\mathbf{1} = 0$ always holds, since each row sums to 0. So $D^{\frac{1}{2}}\mathbf{1}$ is always an eigenvector of $\mathcal{L}$ with eigenvalue 0.

$\mathcal{L}$ is also invariant over the orthogonal complement of span $(\mathbf{1})$, so if we restrict ourselves to $U = \mathbf{1}^{\perp}$, $\min_{u \in U} \frac{\langle x, \mathcal{L}x\rangle}{\langle x, x\rangle} = \lambda_2$.

So $\lambda_2 = \min_{u \in U} \frac{\langle x, Lx\rangle}{d\langle x, x\rangle}$

Both of these are optimisation problems with the same objective, and λ_2 is a relaxed version of Φ_G (we optimise over all vectors, not just indicators).

Cheeger's inequality formalises this similarity, stating that $2\Phi_G \geq \lambda_2 \geq \frac{\Phi_G^2}{2}$.

Cheeger's: Easy Direction

Let $S \subseteq V$ be a set of vertices. Define our 'indicator' vector x_S to have $x_s(i) = \frac{1}{|S|}$ if $i \in S$, else $x_s(i) = \frac{-1}{|V-S|}$

$\lambda_2 \leq \frac{\langle x_s, Lx_S \rangle}{d\langle x_S, x_S \rangle} \leq 2\phi_G$.

Random Walk

Suppose we have a d regular graph. The random walk matrix is $\frac{1}{d}A$, and so $(A/d)^n$ gives the distribution of a random walk after n steps.

$\mathcal{L} = \frac{1}{d}(I - A) = \frac{1}{d}(L) = \frac{I}{d} - \frac{A}{d}$.

Suppose $A/d = U^{-1}RU$. Then if λ_{n-1} is small, it goes to 0 as we take powers, leaving $\lambda_n = 1$ so random walk is uniform.

The eigenvalues of $\mathcal{L}$ are related to negatives of the eigenvalues of A/d. So if λ_2 of $\mathcal{L}$ is large, λ_{n-1} of A/d is small and so the random walk is uniform. We can see why that makes it a good expander.

So if λ_2 is high, it means it is 'easy' to walk from any vertex to any other vertex in a fairly short number of steps, and that all vertices are (roughly) equally easy to get to - which matches up with our intuitions about good sparsifiers.

Paper and Results

Spectral Sparsification of Graphs by Daniel A. Spielman and Shang-Hua Teng

Spectral Approximation: Quadratic forms are bounded by $\frac{1}{\sigma}\langle x, L_{\tilde{G}}x\rangle \leq \langle x, L_G x\rangle \leq \sigma\langle x, L_{\tilde{G}}x\rangle$.

For every G, we can find a sparsifier $\tilde{G}$ with (n/ϵ^2) edges that is a $(1+\epsilon)$ spectral approximation of G.

Their algorithm takes (m) time, where $m = |E|$.

Preserving Expected Quadratic Forms

If we select edge with probability $p_{u,v}$, and assign it weight $1/p_{u,v}$ in our sparsified graph, $\mathbb{E}[\langle x, Lx\rangle] = \mathbb{E}[\langle x, \tilde{L}x\rangle]$.

We can see this from decomposing $L = \sum_{u,v\in E} L_{u,v}$ so $\mathbb{E}[\tilde{L}] = \sum_{u,v\in E} p_{u,v} L_{u,v}$: Dividing by $p_{u,v}$ undoes the effect of the probability of dropping it.

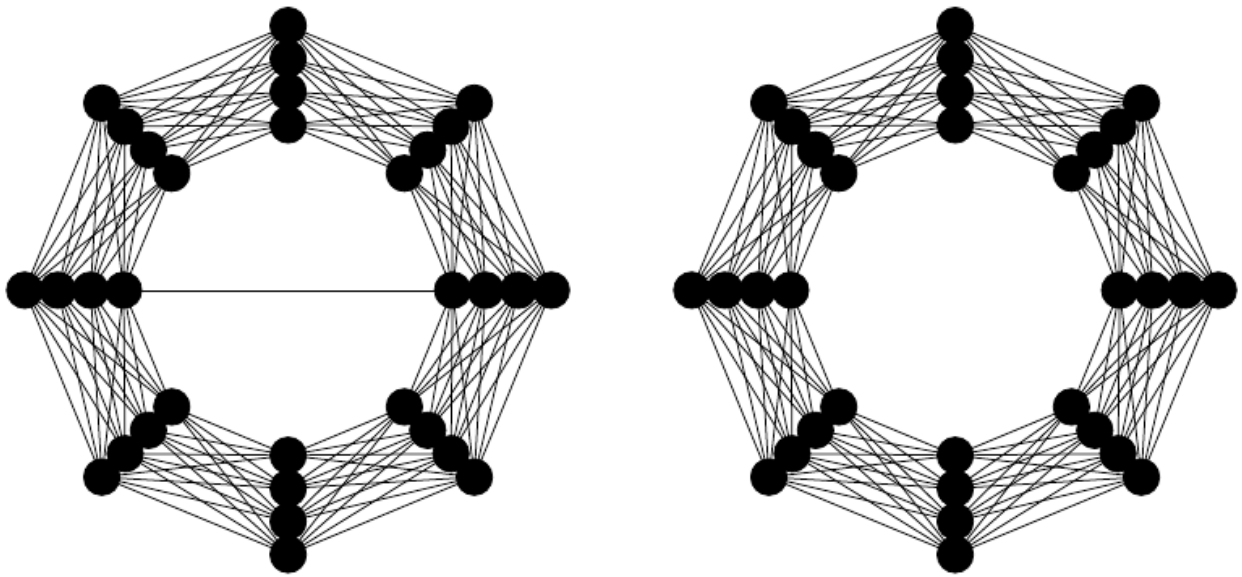
Problem is now assigning $p_{u,v}$ in a way that preserves extremal quadratic forms, at least with high probability.

Spectral vs Cut Sparsifiers

Intuitively, a good sparsifier should preserve 'important' edges, such as if there is one edge which connects two otherwise-distant vertices.

Spectral Sparsifiers are sensitive to that distance concept, while other kinds of sparsifiers like cut sparsifiers aren't.

One way we translate that intuition is by making the probability of $\{i, j\}$ being included proportional to $\frac{1}{\min(\deg i, \deg j)}$. In fact, this is actually one of the methods the main sampling algorithm uses.



Overview of Algorithm

$p_{i,j} = \min(1, \frac{\gamma}{\min(d_i, d_j)})$.

γ is defined in terms of our tolerances (of quality of sparsifier and probability of success), and also inversely proportional to λ_2^2. Intuition - highly connected graphs can survive more 'aggressive' sparsification.

Show that if $||D^{-\frac{1}{2}}(L - \tilde{L})D^{-\frac{1}{2}}||$ is small and λ_2 is large, $\tilde{G}$ is a good spectral approximation for G.

Show that this sparsifier gives us $||D^{-\frac{1}{2}}(\tilde{D} - D)D^{-\frac{1}{2}}||$ and $||D^{-\frac{1}{2}}(\tilde{A} - A)D^{-\frac{1}{2}}||$ small with high probability

So by triangle inequality $||D^{-\frac{1}{2}}(L - \tilde{L})D^{-\frac{1}{2}}||$ is small with high probability.

Small Laplacian Difference $\Rightarrow$ Good Spectral Approximation (Lemma 6.2)

If Laplacian difference $||D^{-\frac{1}{2}}(L - \tilde{L})D^{-\frac{1}{2}}|| \leq \epsilon$ and $\lambda_2(D^{-\frac{1}{2}}LD^{-\frac{1}{2}}) \geq \lambda$, $\tilde{G}$ is a $\sigma = \frac{\lambda}{\lambda - \epsilon}$ spectral approximation.

For d regular graphs, we have $\frac{1}{d}||(L - \tilde{L})|| \leq \epsilon$, and $\lambda_2(L) = d \cdot \lambda_2(\mathcal{L}) \geq d \cdot \lambda$

We can split x up into u parallel to $\mathbf{1}$ and v orthogonal to it and use the fact $\langle v, Lv \rangle \geq d \cdot \lambda$ (since quadforms are bounded by eigenvalues)

In the regular case, a good intuition is that if the spectrum of $L - \tilde{L}$ has small eigenvalues ($\leq \epsilon$), then their quadratic forms behave similarly.

As we increase λ our bound gets tighter, reflecting that highly-connected graphs are easier to sparsify well.

$||D^{-\frac{1}{2}}(A - \tilde{A})D^{-\frac{1}{2}}||$ is Small (Lemma 6.3-4)

Since the norm is just the highest eigenvalue, it instead bounds the trace (product of eigenvalues) of $\Delta = D^{-1}(\tilde{A} - A)$.

Idea from earlier: Random walk matrix encodes probability of walking from v_0 to v_k. Still applies here in some form .

For a given walk $v_0, v_1, \cdots v_k$, its probability is nonzero iff all of the necessary edges are included in A, so each significant sequence is analogous to a walk $v_0, \cdots, v_k$ on A.

$\mathbb{E}[\Delta_{v_i,v_j}] = 0$ from the definition. Δ_{v_i,v_j} is independent of all others except Δ_{v_j,v_i}. So we can split $\mathbb{E}[\Pi_{i=1}^{k}\Delta_{v_{i-1},v_i}]$ into independent pairs $E[\Delta_{v_i,v_{i+1}}\Delta_{v_{i+1},v_i}]$, so a walk has nonzero contribution only if each edge appears at least twice (i.e $\Delta_{v_i,v_{i+1}}\Delta_{v_{i+1},v_i} \neq 0$).

They use an ingenious method to bound the number of such walks.

$||D^{-\frac{1}{2}}(D - \tilde{D})D^{-\frac{1}{2}}||$ is Probably Small (Lemma 6.7)

The probability that $||D^{-\frac{1}{2}}(D - \tilde{D})D^{-\frac{1}{2}}|| \geq \epsilon$ is proportional to $e^{-\epsilon^2}$.

The norm of a diagonal matrix is just it's largest entry, and $D^{-1}(\tilde{D} - D)_{i,i} = 1 - \frac{\tilde{d}_i}{d_i}$.

$\mathbb{E}[\tilde{d}_i] = d_i$, and $\tilde{d}_i$ decomposes into a sum of independent d_i indicator variables.

So we can apply the Chernoff bound to the probability that $\tilde{d}_i - d_i > \epsilon d_i$ for a given i and then use union bound to show that the probability of this occuring for any i is small.

High-Conductance Subgraphs Exist (Theorem 7.1)

We can always find a reasonably large subgraph with fairly high conductance.

From that fact, it follows that we can partition a graph into high-conductance components, by repeatedly extracting these subgraphs.

Let S be the largest set with $\Phi_B(S) \leq \phi$, for some $B \subseteq V$. If S is small with $Vol(S) = aVol(B)$ with $a \leq 1/3$, then $\Phi_{B-S} \geq \phi(\frac{1-3a}{1-a})$

Consider the minimal-conductance subgraph R of the graph $B - S$. Suppose it has conductance $\Phi_{B-S}(R) < \phi(\frac{1-3a}{1-a})$.

Let $T = R \cup S$ so $vol(T) > vol(S)$.

Either $\phi_B(T) \leq \phi_B(S)$ or $\phi_B(B - T) \leq \phi_B(S)$, contradiction.

Sparsifying Arbitrary Graphs (Section 8-10)

These deal with approximating this optimal partition, and using that to sparsify arbitrary weighted and unweighted graphs.

It's mainly extending the ideas from sections 6 and 7 in routine ways.

In general, if we union high-conductance subgraphs, the resulting subgraph has (with high probability) higher conductance.

So the algorithm they use calls a *Partition* algorithm to find high-conductance subgraphs, and unions the resulting subgraphs together to improve the conductance.