Spectral Graph Sparsification

Presenter: Derrick Blakely

June 28, 2019

University of Virginia

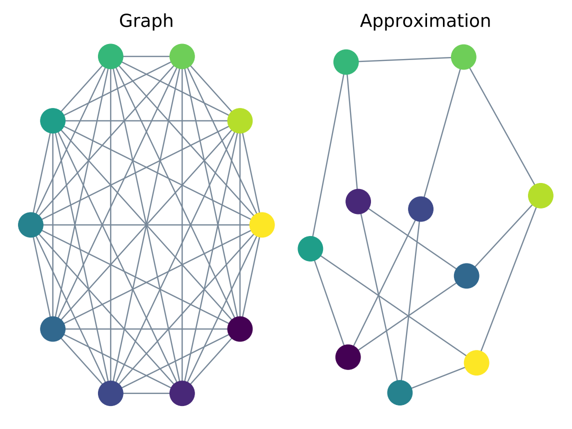
https://qdata.github.io/deep2Read/

Table of contents

Sparsification

Intro

Approximate any graph with a sparse graph:



Why?

- Compression
- Faster to compute with
- Less memory

Graph Approximation

Given: $G = (V, E, w)$ and $H = (V, F, z)$, we want:

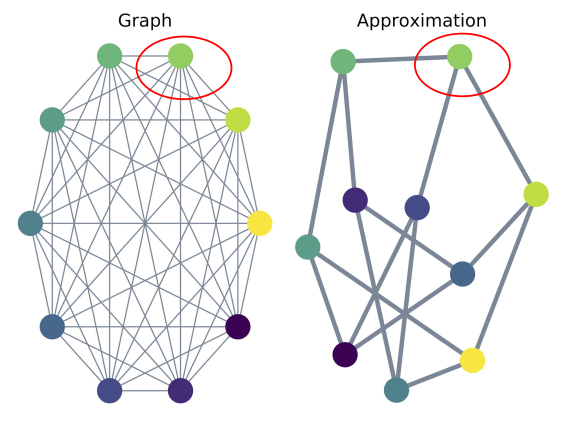
$$H \approx_\epsilon G$$

Properties we want to preserve:

- Cut sizes
- Communities/clusters
- Behavior of random walks

Cut Approximation

- $H \approx_\epsilon G$ if for every $S \subset V$, the sum of weights leaving S is the same in H and G



Cut Approximation Theorem [2]

Given a graph G with m edges and error parameter ϵ, we can find a graph H such that:

- H has $O(n \log n / \epsilon^2)$ edges
- The value of every cut in H is $(1 \pm \epsilon)$ the corresponding cut in G
- H can be constructed in $O(m \log^2 n)$ time if G is unweighted and $O(m \log^3 n)$ if G is weighted

Stronger Approximation: Spectral Sparsification

$H \approx_\kappa G$ if for some error parameter κ:

$$H \preccurlyeq G \preccurlyeq \kappa H$$

Where $H \preccurlyeq G$ if $\forall x : V \to \mathbb{R}$:

$$x^\top L_H x \leq x^\top L_G x$$

The Graph Laplacian

The ∇ Operator

- A pseudo-vector: $\nabla = [\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, ..., \frac{\partial}{\partial x_n}]$

The ∇ Operator

- A pseudo-vector: $\nabla = [\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, ..., \frac{\partial}{\partial x_n}]$
- Gradient: $\nabla f = [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}]$

The ∇ Operator

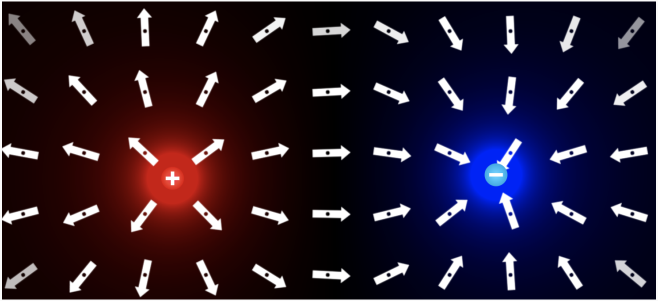
- A pseudo-vector: $\nabla = [\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, ..., \frac{\partial}{\partial x_n}]$
- Gradient: $\nabla f = [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}]$
- Divergence: $\nabla \cdot f = \nabla \cdot [f_1, f_2, ..., f_n] = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + ... + \frac{\partial f_n}{\partial x_n}$

The ∇ Operator

- A pseudo-vector: $\nabla = [\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, ..., \frac{\partial}{\partial x_n}]$
- Gradient: $\nabla f = [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}]$
- Divergence: $\nabla \cdot f = \nabla \cdot [f_1, f_2, ..., f_n] = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + ... + \frac{\partial f_n}{\partial x_n}$
- Laplacian: $\nabla \cdot \nabla f = \nabla^2 f = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + ... + \frac{\partial^2 f}{\partial x_n^2}$

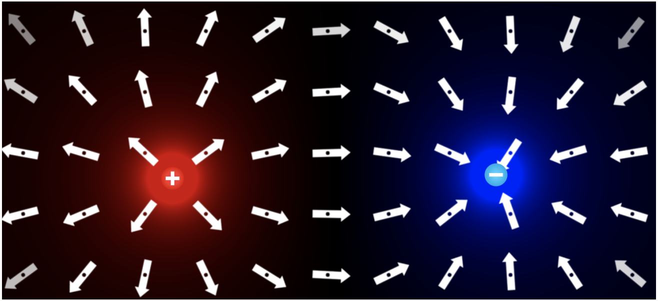
Physics Explanation

- Let $V : \mathbb{R}^3 \to \mathbb{R}$ be the electric potential



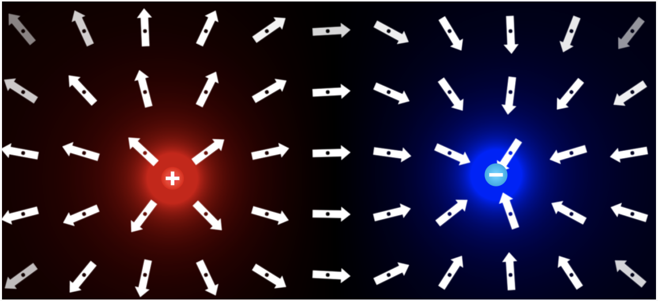
Physics Explanation

- Let $V : \mathbb{R}^3 \to \mathbb{R}$ be the electric potential
- $E = -\nabla V : \mathbb{R}^3 \to \mathbb{R}^3$ is the electric field



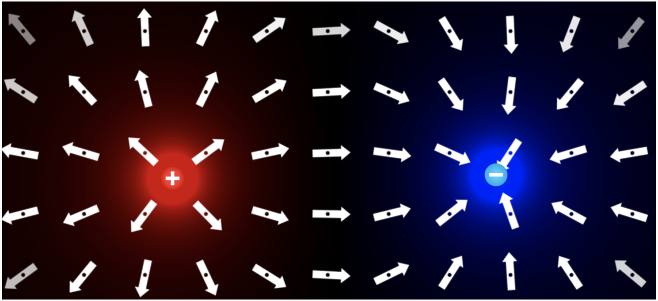
Physics Explanation

- Let $V : \mathbb{R}^3 \to \mathbb{R}$ be the electric potential
- $E = -\nabla V : \mathbb{R}^3 \to \mathbb{R}^3$ is the electric field
- $div(E) = \nabla \cdot E : \mathbb{R}^3 \to \mathbb{R}$ is divergence of E



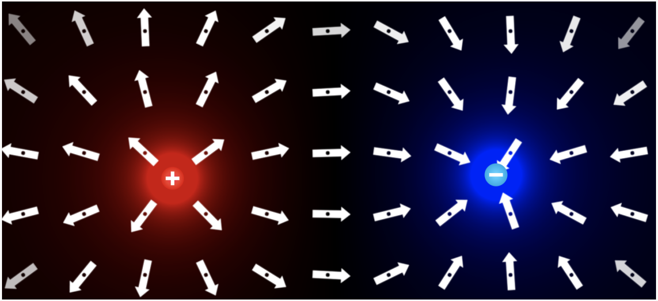
Physics Explanation

- Let $V : \mathbb{R}^3 \to \mathbb{R}$ be the electric potential
- $E = -\nabla V : \mathbb{R}^3 \to \mathbb{R}^3$ is the electric field
- $div(E) = \nabla \cdot E : \mathbb{R}^3 \to \mathbb{R}$ is divergence of E
- $Laplacian(V) = div(E) = \nabla \cdot \nabla V = \nabla^2 V$



Interpreting the Laplacian

- $Laplacian(V) = div(E) = \nabla \cdot E = \nabla \cdot \nabla V = \nabla^2 V$
- Extent to which a point behaves like a positive voltage source
- Net flux density through a volume at a point
- Second derivative of V: smoothness of V over space

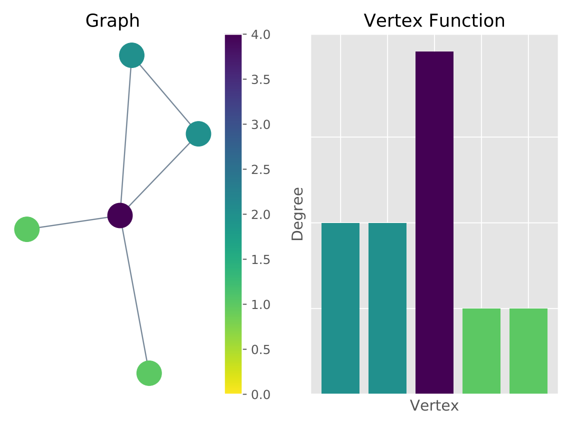


What is the Graph Laplacian?

How do we define $\nabla \cdot \nabla f = \nabla^2 f$ for graphs? Three questions:

1. What does f mean for graphs?
2. What does the gradient ∇f mean for graphs?
3. What does the Laplacian $\nabla \cdot \nabla f$ mean for graphs?

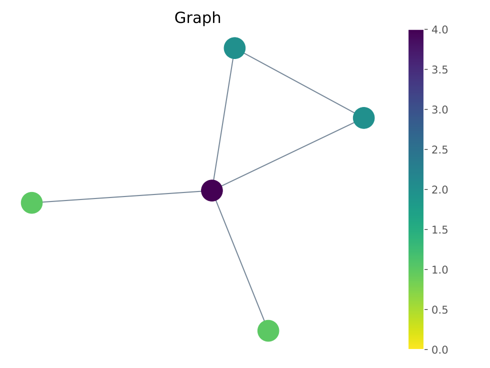
- $f : V \to \mathbb{R}$
- For example: the degree of each node
- In other words: degree of a node is like its potential



2. Gradient of Degree Function

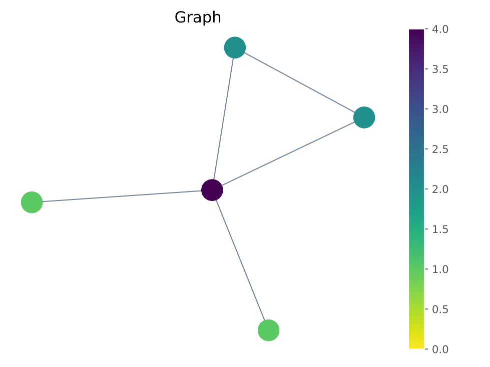
- Need an analog to $\nabla = [\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}]$
- Incidence matrix K: each node gets a row and each edge gets a column
- If outgoing edge, $K_{n,e} = -1$
- If incoming edge, $K_{n,e} = 1$
- Neither: $K_{n,e} = 0$

Incidence Matrix K



$$K = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{1}$$

Gradient of Degree Function



$$\nabla f = K^\top f = K^\top \begin{bmatrix} 2 \\ 2 \\ 4 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 2 \\ -3 \\ -3 \end{bmatrix} \tag{2}$$

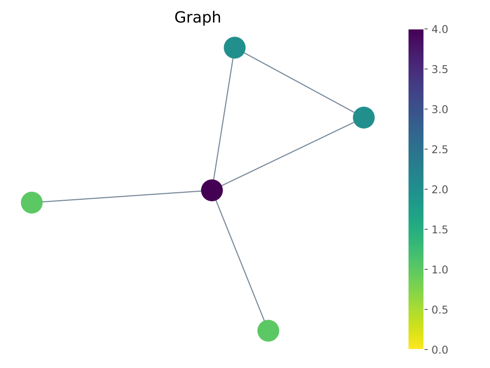
3. Laplacian Operator for Graphs

$$\nabla \cdot \nabla = KK^\top = L = \begin{bmatrix} 2 & -1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ -1 & -1 & 4 & -1 & -1 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{bmatrix} \tag{3}$$

$$L = D - A \tag{4}$$

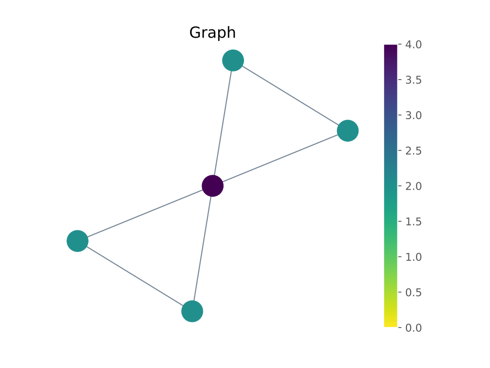
(where D is the degree matrix and A is the adjacency matrix)

Graph Divergence



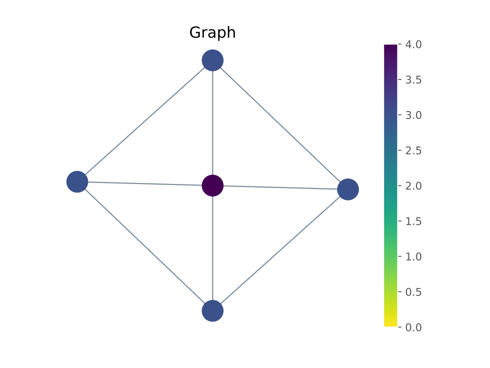
$$\nabla^2 f = KK^\top f = Lf = \begin{bmatrix} -2 \\ -2 \\ 10 \\ -3 \\ -3 \end{bmatrix} \tag{5}$$

Divergence and Smoothness



$$\nabla^2 f = KK^\top f = Lf = \begin{bmatrix} -2 \\ -2 \\ 8 \\ -2 \\ -2 \end{bmatrix} \tag{6}$$

Divergence and Smoothness



$$\nabla^2 f = KK^\top f = Lf = \begin{bmatrix} -1 \\ -1 \\ 4 \\ -1 \\ -1 \end{bmatrix} \tag{7}$$

Summary

Physics	Graphs
Potential: V	Vertex function: f
∇	K: incidence matrix
∇V: field	$K^\top f$: graph gradient
∇^2: Laplacian	$KK^\top$ Graph Laplacian
$\nabla^2 V$: Laplacian of V	$KK^\top f$: Graph Laplacian of f

Back to Sparsification

Stronger Approximation: Spectral Sparsification

$H \approx_\epsilon G$ if for some error parameter κ:

$$H \preccurlyeq G \preccurlyeq \kappa H$$

Where $H \preccurlyeq G$ if $\forall x : V \rightarrow \mathbb{R}$:

$$x^\top L_H x \leq x^\top L_G x$$

Laplacian Quadratic Form

Sum of square differences across the edges:

$$x^\top L_G x = \sum_{(u,v)\in E} w(u,v)(x(u) - x(v))^2$$

Spectral Approximation Theorem [1]

Given G and error parameter ϵ, we can find an approximation H such that:

- $H \preccurlyeq G$
- H has $O(n \log n/\epsilon^2)$ edges
- H can be found in time $\tilde{O}(m/\epsilon^2)$

Why is spectral approx. stronger than cut approx.?

- A spectral approximation is also a cut approximation
- The converse is not always true

Why is it better?

- H will inherit a bunch of properties of G
- The eigenvectors and eigenvalues will be similar
- Can use H to obtain approximate solutions to linear systems of G
- If $x \in \mathbb{R}^n$, we can use convex solvers
- If $x \in \{0, 1\}^n$, we can relax it to $x \in \mathbb{R}^n$ to obtain approximate solutions

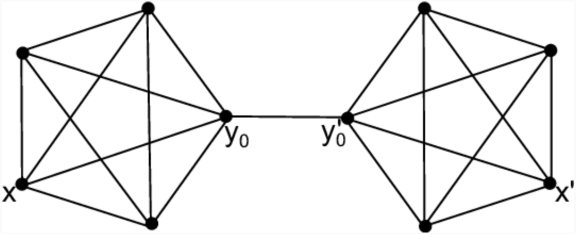
How to Obtain H

Basic Idea: Random Sampling

- Choose edge e with probability p_e
- Take k independent samples
- Add e to H with weight $1/kp_e$

Why not Uniform Sampling?

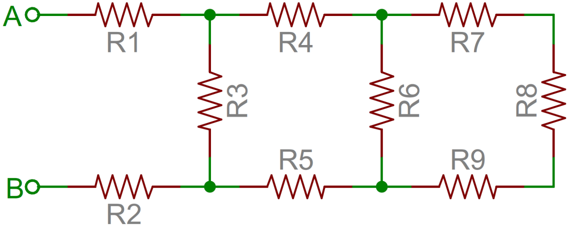
Don't want to disconnect the graph. E.g.,:



Instead: bias probabilities based on the "effective resistance" of the edges

Sparsification using Effective Resistance

Resistor Networks



Circuit Basics

Ohm's law: voltage drop across a resistor is

$$V = IR$$

Power dissipation across a resistor:

$$P = V^2/R$$

Quadratic Form and Power Dissipation

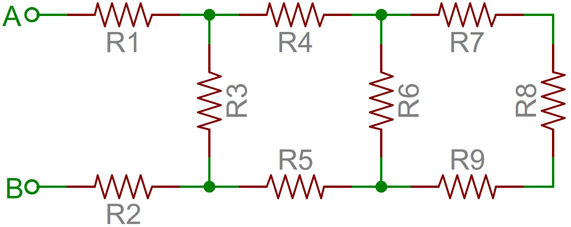
- If we interpret x as voltages and E as conductances, then

$$x^\top L_G x$$

  gives the power dissipation of the graph.
- If $H \approx_\kappa G$, then H and G have approximate "electrical equivalence."

Effective Resistance

$R_{eff}(e)$: power dissipation when a unit of current is sent across the ends of e



$$R_{eff}(e) = ||L_G^{-1/2} b_e|| = b_e^\top L_G^{-1} b_e$$

where $b_e \in 1, -1, 0^n$, $b_e(u) = 1$, $b_e(v) = -1$, 0 everywhere else

Sparsification with Effective Resistance [3]

- Choose edge e with probability $p_e \propto R_{\mathit{eff}}(e)$
- Take k independent samples
- Add e to H with weight $1/kp_e$

J. Batson, D. A. Spielman, N. Srivastava, and S.-H. Teng.
Spectral sparsification of graphs: theory and algorithms.
Communications of the ACM, 56(8):87–94, 2013.

A. Benczur and D. R. Karger.
Randomized approximation schemes for cuts and flows in capacitated graphs.
arXiv preprint cs/0207078, 2002.

D. A. Spielman and N. Srivastava.
Graph sparsification by effective resistances.
SIAM Journal on Computing, 40(6):1913–1926, 2011.