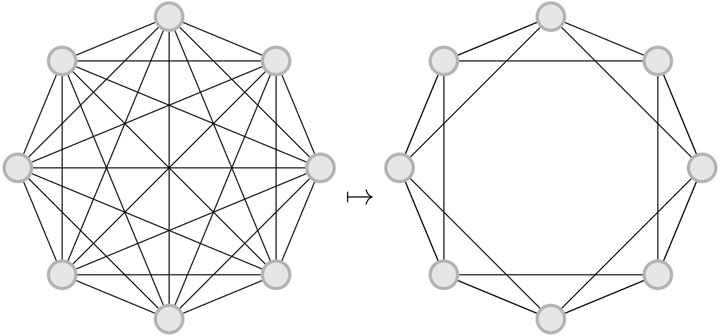
Graph sparsification

Bruno Ordozgoiti[1]

[1]Aalto University

Helsinki 2021



We write $A - B \succcurlyeq 0$ if $A - B$ is P.S.D.

We consider a graph $G = (V, E)$ with Laplacian $L_G = D_G - A_G$.

Definition: ϵ-approximation of a graph.

$H = (V, E')$ ϵ-approximates G if

$$(1 - \epsilon)L_G \preccurlyeq L_H \preccurlyeq (1 + \epsilon)L_G.$$

We write $A - B \succcurlyeq 0$ if $A - B$ is P.S.D.

We consider a graph $G = (V, E)$ with Laplacian $L_G = D_G - A_G$.

Definition: ϵ-approximation of a graph.

$H = (V, E')$ ϵ-approximates G if

$$(1-\epsilon)L_G \preccurlyeq L_H \preccurlyeq (1+\epsilon)L_G.$$

Implications:

- $(1-\epsilon)x^T L_G x \preccurlyeq x^T L_H x \preccurlyeq (1+\epsilon)x^T L_G x.$

We write $A - B \succcurlyeq 0$ if $A - B$ is P.S.D.

We consider a graph $G = (V, E)$ with Laplacian $L_G = D_G - A_G$.

Definition: ϵ-approximation of a graph.

$H = (V, E')$ ϵ-approximates G if

$$(1 - \epsilon)L_G \preccurlyeq L_H \preccurlyeq (1 + \epsilon)L_G.$$

Implications:

- $(1 - \epsilon)x^T L_G x \preccurlyeq x^T L_H x \preccurlyeq (1 + \epsilon)x^T L_G x.$
- Cuts: $(1 - \epsilon)1_S^T L_G 1_S \preccurlyeq 1_S^T L_H 1_S \preccurlyeq (1 + \epsilon)1_S^T L_G 1_S.$

First step: build H randomly to preserve G in expectation.

$$w_H(u,v) = \begin{cases} \frac{w_G(u,v)}{p_{u,v}} & \text{w. p. } p_{u,v} \\ 0 & \text{w. p. } 1 - p_{u,v}. \end{cases}$$

$\mathbb{E}\left[L_H\right] = \sum_{u,v \in E} p_{u,v} \frac{w_G(u,v)}{p_{u,v}} L_{u,v} = L_G$ where $L_{u,v}$ is the Laplacian of the graph with unique edge u, v.

First step: build H randomly to preserve G in expectation.

$$w_H(u,v) = \begin{cases} \frac{w_G(u,v)}{p_{u,v}} & \text{w. p. } p_{u,v} \\ 0 & \text{w. p. } 1 - p_{u,v}. \end{cases}$$

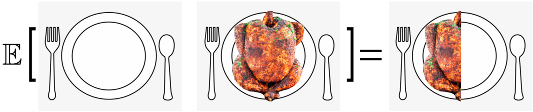
$\mathbb{E}[L_H] = \sum_{u,v \in E} p_{u,v} \frac{w_G(u,v)}{p_{u,v}} L_{u,v} = L_G$ where $L_{u,v}$ is the Laplacian of the graph with unique edge u, v.

First step: build H randomly to preserve G in expectation.

$$w_H(u, v) = \begin{cases} \frac{w_G(u,v)}{p_{u,v}} & \text{w. p. } p_{u,v} \\ 0 & \text{w. p. } 1 - p_{u,v}. \end{cases}$$

$\mathbb{E}\left[L_H\right] = \sum_{u,v \in E} p_{u,v} \frac{w_G(u,v)}{p_{u,v}} L_{u,v} = L_G$ where $L_{u,v}$ is the Laplacian of the graph with unique edge u, v.

We now need Chernoff bounds for the concentration of eigenvalues.



Matrix Chernoff bounds (Tropp, 2012)[1].

Theorem

Theorem 32.3.1. Let $X_1, \ldots, X_m$ be independent random n-dimensional symmetric positive semidefinite matrices so that $\|X_i\| \leq R$ almost surely. Let $X = \sum_i X_i$ and let μ_{max} and μ_{min} be the maximum and minimum eigenvalues of

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X_i].$$

Then

$$Pr\left[\lambda_{min}\left(\sum_i X_i\right) \leq (1-\epsilon)\mu_{min}\right] \leq n\left(\frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}}\right)^{\frac{\mu_{min}}{R}} \leq ne^{\frac{-\epsilon^2\mu_{min}}{2R}},$$

$$Pr\left[\lambda_{max}\left(\sum_i X_i\right) \geq (1+\epsilon)\mu_{max}\right] \leq n\left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^{\frac{\mu_{max}}{R}} \leq ne^{\frac{-\epsilon^2\mu_{max}}{3R}}$$

[1] Tropp, Joel A. "User-friendly tail bounds for sums of random matrices." Foundations of computational mathematics 12.4 (2012): 389-434.

Matrix Chernoff bounds (Tropp, 2012)[1].

Theorem

Theorem 32.3.1. Let $X_1, \ldots, X_m$ be independent random n-dimensional symmetric positive semidefinite matrices so that $\|X_i\| \leq R$ almost surely. Let $X = \sum_i X_i$ and let μ_{max} and μ_{min} be the maximum and minimum eigenvalues of

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X_i] \,.$$

Then

$$Pr\left[\lambda_{min}\left(\sum_i X_i\right) \leq (1-\epsilon)\mu_{min}\right] \leq n\left(\frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}}\right)^{\frac{\mu_{min}}{R}} \leq ne^{\frac{-\epsilon^2\mu_{min}}{2R}},$$

$$Pr\left[\lambda_{max}\left(\sum_i X_i\right) \geq (1+\epsilon)\mu_{max}\right] \leq n\left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^{\frac{\mu_{max}}{R}} \leq ne^{\frac{-\epsilon^2\mu_{max}}{3R}}$$

[1] Tropp, Joel A. "User-friendly tail bounds for sums of random matrices." Foundations of computational mathematics 12.4 (2012): 389-434.

In a nutshell, if $\|X_i\| \leq R$, we fail with probability at most $ne^{\frac{-\epsilon^2 \mu_{min}}{2R}}$ (similar for μ_{max}).

Recall: $L_H = \sum_{u,v \in E} \frac{w_G(u,v)}{p_{u,v}} L_{u,v}$.

Thus, we just need to choose $p_{u,v}$ carefully, so that $\left\| \frac{w_G(u,v)}{p_{u,v}} L_{u,v} \right\| \leq R$.

But first, let's take care of a little annoyance: $ne^{\frac{-\epsilon^2 \mu_{min}}{2R}}$

For PD matrices A, B,

$$A \preccurlyeq B(1+\epsilon) \Leftrightarrow B^{-1/2}AB^{-1/2} \preccurlyeq (1+\epsilon)I.$$

For PD matrices A, B,

$$A \preccurlyeq B(1+\epsilon) \Leftrightarrow B^{-1/2}AB^{-1/2} \preccurlyeq (1+\epsilon)I.$$

For singular matrices we can use the pseudoinverse:

$$L_H \preccurlyeq (1+\epsilon)L_G \Leftrightarrow L_G^{+/2}L_H L_G^{+/2} \preccurlyeq (1+\epsilon)L_G^{+/2}L_G L_G^{+/2},$$

where $L_G^{+/2}L_G^{+/2} = L_G^{+}$.

For PD matrices A, B,

$$A \preccurlyeq B(1+\epsilon) \Leftrightarrow B^{-1/2} A B^{-1/2} \preccurlyeq (1+\epsilon) I.$$

For singular matrices we can use the pseudoinverse:

$$L_H \preccurlyeq (1+\epsilon) L_G \Leftrightarrow L_G^{+/2} L_H L_G^{+/2} \preccurlyeq (1+\epsilon) L_G^{+/2} L_G L_G^{+/2},$$

where $L_G^{+/2} L_G^{+/2} = L_G^{+}$. Now,

- $\Pi = L_G^{+/2} L_G L_G^{+/2}$ is a projection. Indeed, $\Pi\Pi = L_G^{+/2} L_G L_G^{+/2} L_G^{+/2} L_G L_G^{+/2} = L_G^{+/2} L_G L_G^{+/2} = \Pi$. So $\mu_{min} = \mu_{max} = 1$.
- And $\mathbb{E}\left[L_G^{+/2} L_H L_G^{+/2}\right] = \Pi$.

Thus, we sample matrices X as follows:

$$w_X(u,v) = \begin{cases} \frac{w_G(u,v)}{p_{u,v}} L_G^{+/2} L_{u,v} L_G^{+/2} & \text{w. p. } p_{u,v} \\ 0 & \text{w. p. } 1 - p_{u,v}. \end{cases}$$

so that $\sum_{(u,v)\in E} X_{u,v} = L_G^{+/2} L_H L_G^{+/2}$.

Since we want $\|X_{u,v}\| \leq R$, we choose $p_{u,v} = \frac{1}{R} w_G(u,v) \|L_G^{+/2} L_{u,v} L_G^{+/2}\|$.

We now fail with probability at most $ne^{\frac{-\epsilon^2 \mu_{\min}}{2R}} = ne^{\frac{-\epsilon^2}{2R}}$ (similar for μ_{max}).

How may edges do we pick?

Conveniently, $\|L_G^{+/2} L_{u,v} L_G^{+/2}\| = (\delta_u - \delta_v)^T L_G^{+}(\delta_u - \delta_v) = R_{eff}(u, v)$, which is the effective resistance between u and v.

How may edges do we pick?

Conveniently, $\|L_G^{+/2} L_{u,v} L_G^{+/2}\| = (\delta_u - \delta_v)^T L_G^{+}(\delta_u - \delta_v) = R_{eff}(u, v)$, which is the effective resistance between u and v.

Interpretations of $w_G(u, v) R_{eff}(u, v)$:

- probability that (u, v) appears in a random spanning tree,
- leverage score of the edge (u, v) in the incidence matrix.

So $\mathbb{E}[|E_H|] = \sum_{(u,v)\in E} p_{u,v} = \sum_{(u,v)\in E} \frac{1}{R} w_G(u, v) R_{eff}(u, v) =$

How may edges do we pick?

Conveniently, $\|L_G^{+/2} L_{u,v} L_G^{+/2}\| = (\delta_u - \delta_v)^T L_G^+ (\delta_u - \delta_v) = R_{eff}(u, v)$, which is the effective resistance between u and v.

Interpretations of $w_G(u, v) R_{eff}(u, v)$:

- probability that (u, v) appears in a random spanning tree,
- leverage score of the edge (u, v) in the incidence matrix.

So $\mathbb{E}[|E_H|] = \sum_{(u,v)\in E} p_{u,v} = \sum_{(u,v)\in E} \frac{1}{R} w_G(u, v) R_{eff}(u, v) = \frac{n-1}{R}$.

Thus, we sample about $\frac{n}{R}$ edges in expectation, and our probability failure is at most $ne^{\frac{-\epsilon^2}{2R}}$.

How to choose R?

Thus, we sample about $\frac{n}{R}$ edges in expectation, and our probability failure is at most $ne^{\frac{-\epsilon^2}{2R}}$.

How to choose R?

$R = \frac{\epsilon^2}{3.5 \log n}$.

Thus, we sample about $\frac{n}{R}$ edges in expectation, and our probability failure is at most $ne^{\frac{-\epsilon^2}{2R}}$.

How to choose R?

$R = \frac{\epsilon^2}{3.5 \log n}$.

We have found a graph $H = (V, E')$ with $\mathbb{E}\left[|E'|\right] = \mathcal{O}\left(\frac{n \log n}{\epsilon^2}\right)$ satisfying

$$
\begin{aligned}
Pr\left[\lambda_{min}\left(L_G^{+/2} L_H L_G^{+/2}\right) \leq (1-\epsilon)\mu_{min}\right] &\leq ne^{-\epsilon^2/(2R)} \leq ne^{-(3.5/2)\log n} = n^{-3/4}, \\
Pr\left[\lambda_{max}\left(L_G^{+/2} L_H L_G^{+/2}\right) \geq (1+\epsilon)\mu_{max}\right] &\leq ne^{-\epsilon^2/(3R)} \leq ne^{-(3.5/3)\log n} = n^{-1/6}.
\end{aligned}
$$

There are linear-sized $\left(O\left(\frac{n}{\epsilon^2}\right)\right)$ sparsifiers! (Batson et al., 2012)[2]

[2] Batson, Joshua, Daniel A. Spielman, and Nikhil Srivastava. "Twice-ramanujan sparsifiers." SIAM Journal on Computing 41.6 (2012): 1704-1721.

There are linear-sized $\left(O\left(\frac{n}{\epsilon^2}\right)\right)$ sparsifiers! (Batson et al., 2012)[2]

Will this work with adjacency matrices?

[2] Batson, Joshua, Daniel A. Spielman, and Nikhil Srivastava. "Twice-ramanujan sparsifiers." SIAM Journal on Computing 41.6 (2012): 1704-1721.

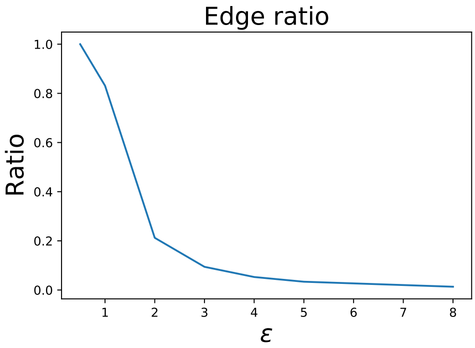
There are linear-sized $\left(O\left(\frac{n}{\epsilon^2}\right)\right)$ sparsifiers! (Batson et al., 2012)[2]

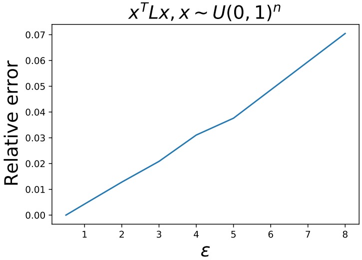
Will this work with adjacency matrices?

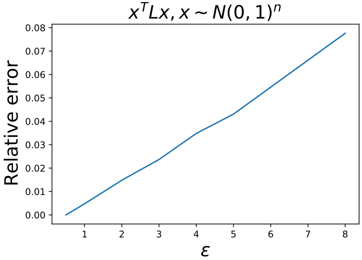
Will this work with signed graphs?

[2] Batson, Joshua, Daniel A. Spielman, and Nikhil Srivastava. "Twice-ramanujan sparsifiers." SIAM Journal on Computing 41.6 (2012): 1704-1721.

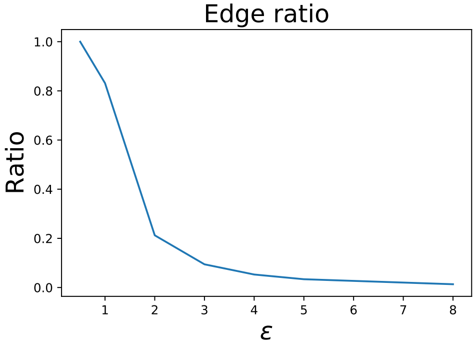
$G = (V, E)$. $|V| = 1000$. Density: 0.05. Planted subgraph with $|V'| = 100$, density: 0.75.

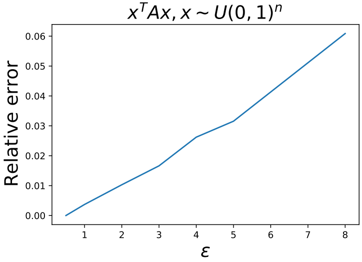


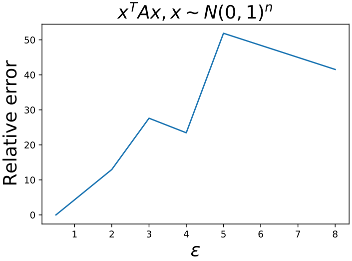




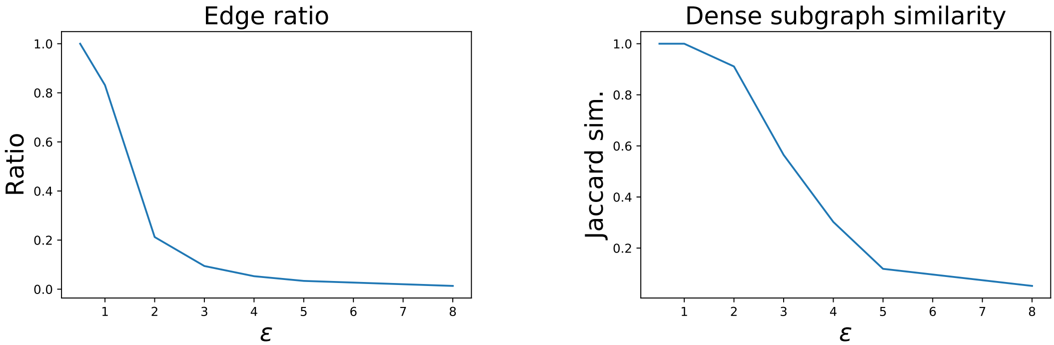
$G = (V, E)$. $|V| = 1000$. Density: 0.05. Planted subgraph with $|V'| = 100$, density: 0.75.







$G = (V, E)$. $|V| = 1000$. Density: 0.05. Planted subgraph with $|V'| = 100$, density: 0.75.



Thanks!

References I

Batson, J., Spielman, D. A., and Srivastava, N. (2012). Twice-ramanujan sparsifiers. *SIAM Journal on Computing*, 41(6):1704–1721.

Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434.