

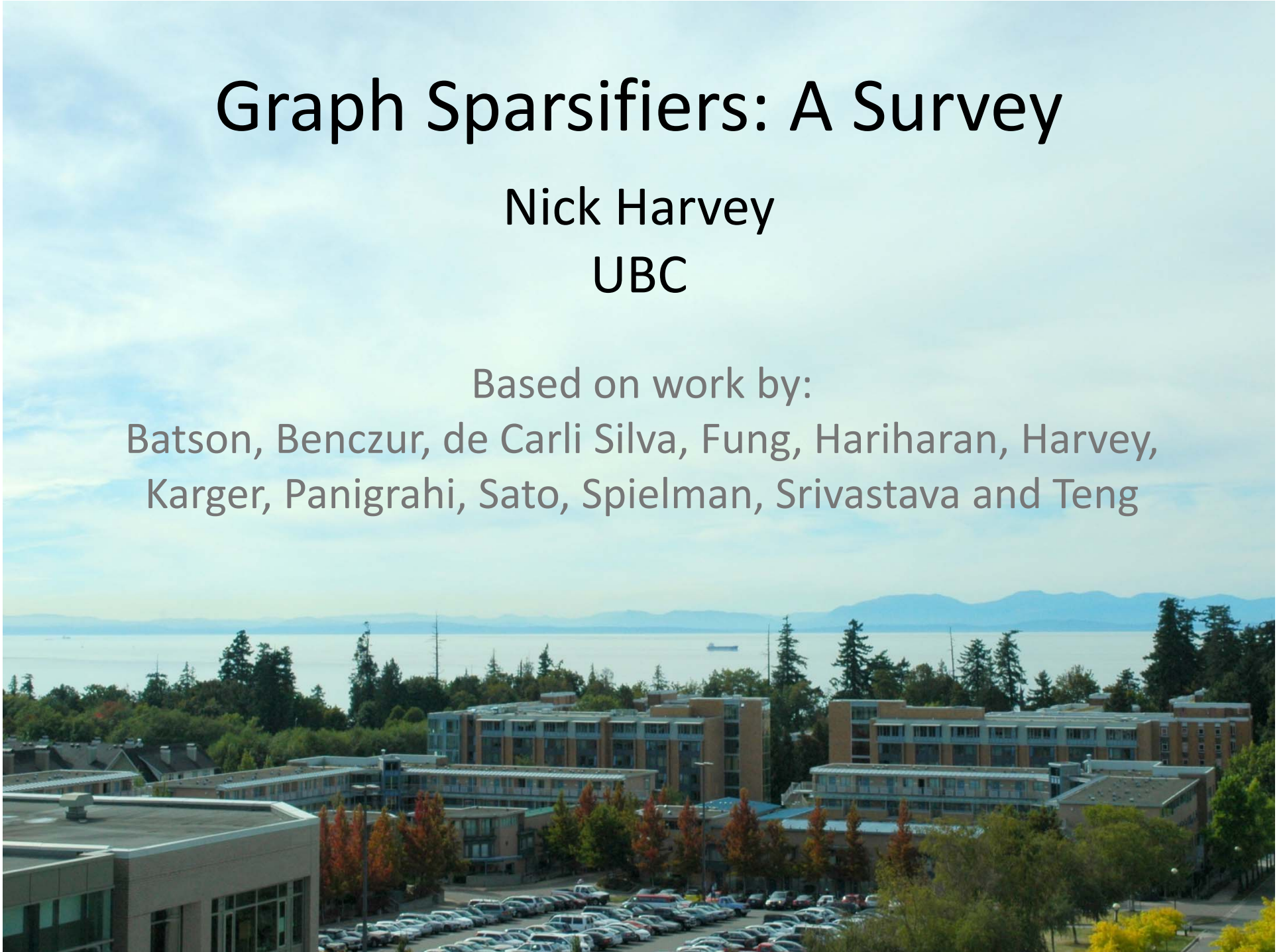
Graph Sparsifiers: A Survey

Nick Harvey

UBC

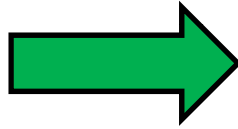
Based on work by:

Batson, Benczur, de Carli Silva, Fung, Hariharan, Harvey,
Karger, Panigrahi, Sato, Spielman, Srivastava and Teng

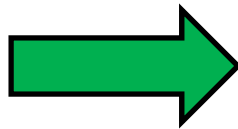


Approximating Dense Objects by Sparse Ones

- Floor joists

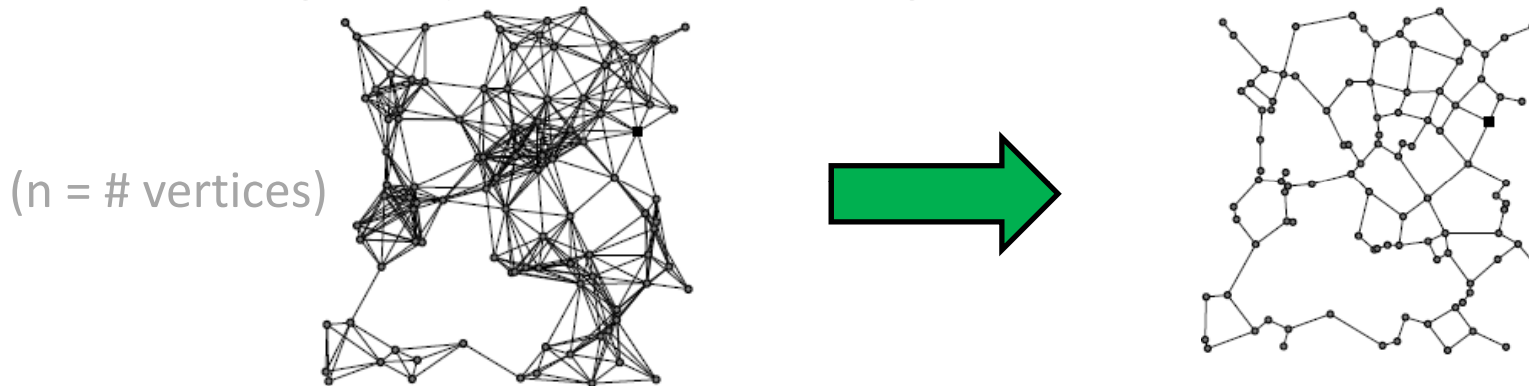


- Image compression

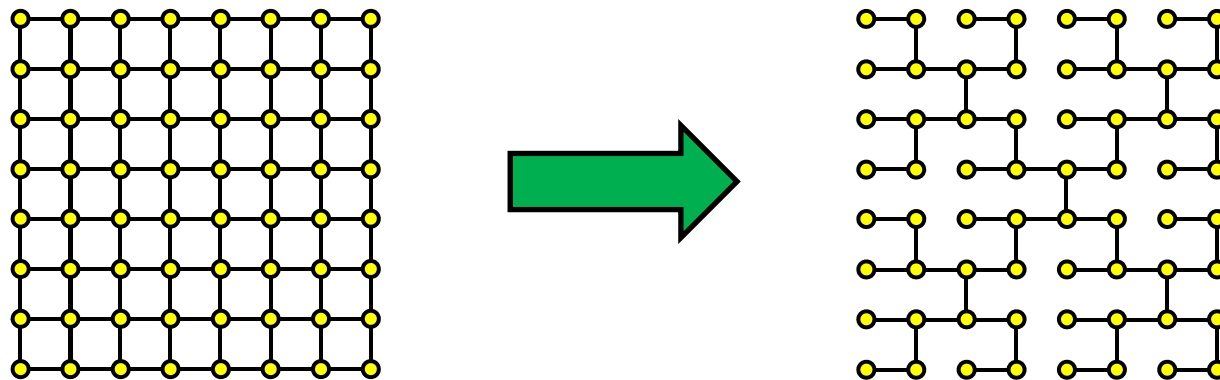


Approximating Dense Graphs by Sparse Ones

- **Spanners:** Approximate distances to within α using only $O(n^{1+2/\alpha})$ edges



- **Low-stretch trees:** Approximate most distances to within $O(\log n)$ using only $n-1$ edges



Overview

- Definitions
 - Cut & Spectral Sparsifiers
 - Applications
- Cut Sparsifiers
- Spectral Sparsifiers
 - A random sampling construction
 - Derandomization

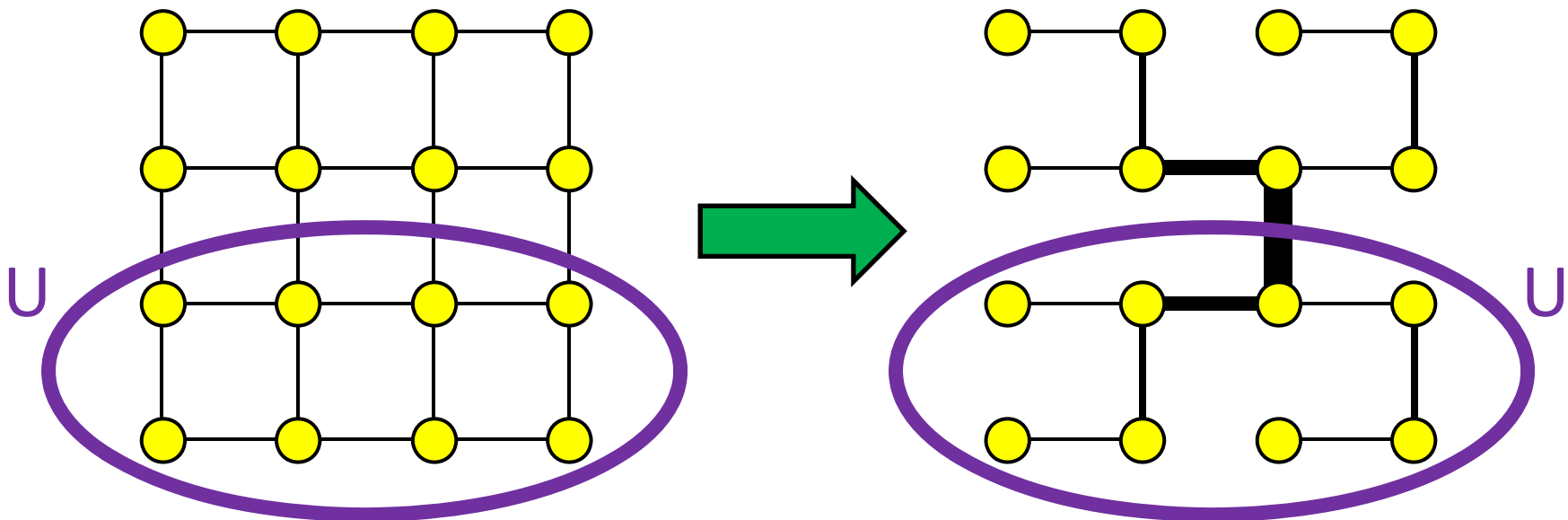
Cut Sparsifiers (Karger '94)

- **Input:** An undirected graph $G=(V,E)$ with weights $u : E \rightarrow \mathbb{R}_+$
- **Output:** A subgraph $H=(V,F)$ of G with weights $w : F \rightarrow \mathbb{R}_+$ such that $|F|$ is small and

$$w(\delta_H(U)) = (1 \pm \epsilon) u(\delta_G(U)) \quad \forall U \subseteq V$$

weight of edges between U and $V \setminus U$ in H

weight of edges between U and $V \setminus U$ in G



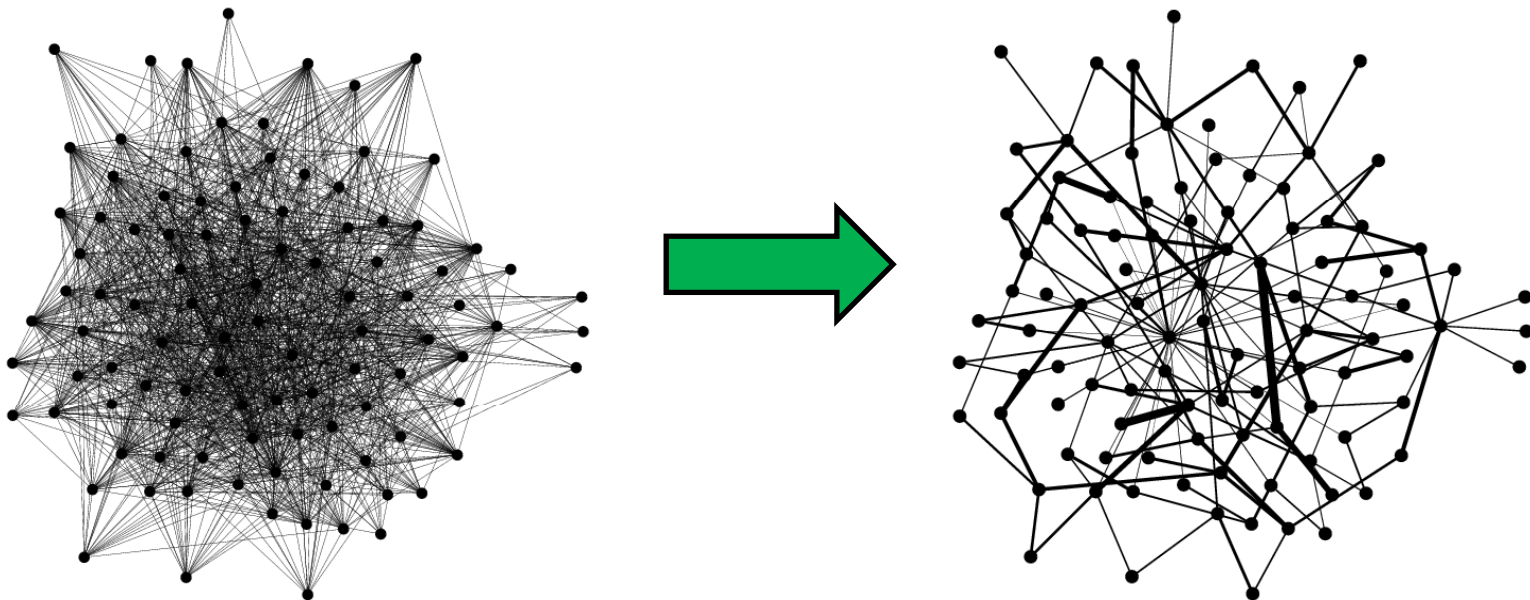
Cut Sparsifiers (Karger '94)

- **Input:** An undirected graph $G=(V,E)$ with weights $u : E \rightarrow \mathbb{R}_+$
- **Output:** A subgraph $H=(V,F)$ of G with weights $w : F \rightarrow \mathbb{R}_+$ such that $|F|$ is small and

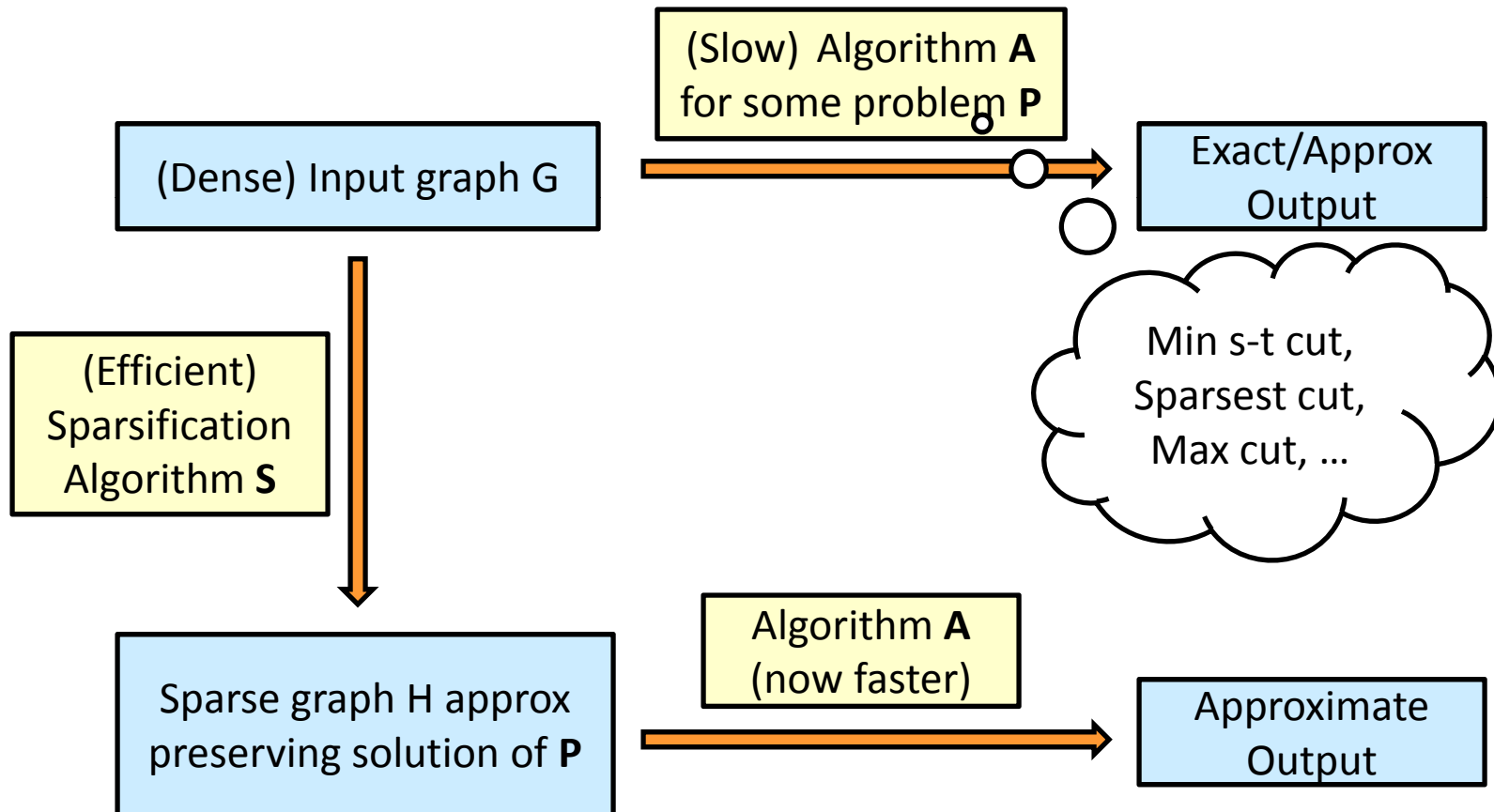
$$w(\delta_H(U)) = (1 \pm \epsilon) u(\delta_G(U)) \quad \forall U \subseteq V$$

weight of edges between U and $V \setminus U$ in H

weight of edges between U and $V \setminus U$ in G



Generic Application of Cut Sparsifiers



Relation to Expander Graphs

- Graph H on V is an **expander** if, for some constant c ,

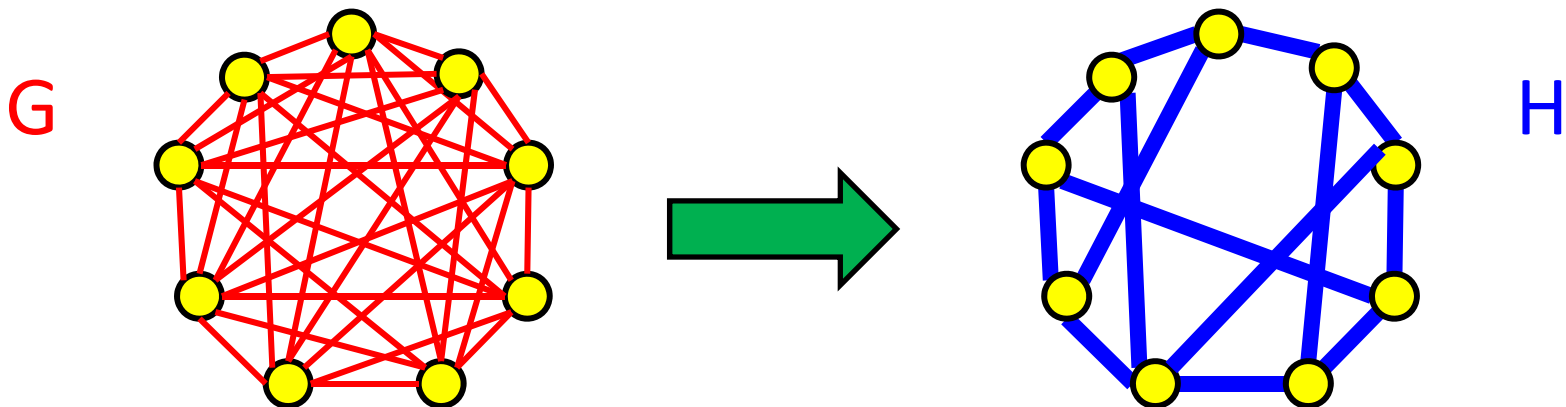
$$|\delta_H(U)| \geq c |U| \quad \forall U \subseteq V, |U| \leq n/2$$

- Let G be the complete graph on V .

If we give all edges of H weight $w=n$, then

$$w(\delta_H(U)) \geq c n |U| \approx c |\delta_G(U)| \quad \forall U \subseteq V, |U| \leq n/2$$

- Expanders are similar to sparsifiers of complete graph

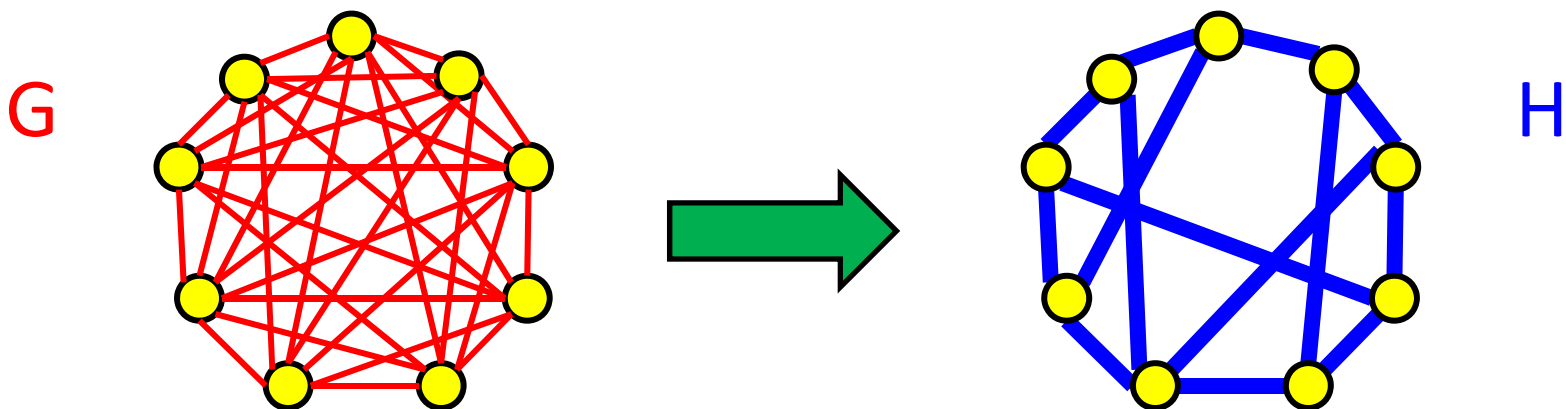


Relation to Expander Graphs

- **Simple Random Construction:**

Erdos-Renyi graph G_{np} is an expander if $p = \Theta(\log(n)/n)$, with high probability. This gives an expander with $\Theta(n \log n)$ edges with high probability.

But aren't there much better expanders?



Spectral Sparsifiers (Spielman-Teng '04)

- **Input:** An undirected graph $G=(V,E)$ with weights $u : E \rightarrow \mathbb{R}_+$
- **Def:** The **Laplacian** is the matrix L_G such that
$$x^T L_G x = \sum_{st \in E} u_{st} (x_s - x_t)^2 \quad \forall x \in \mathbb{R}^V.$$
- L_G is positive semidefinite since this is ≥ 0 .
- **Example:** Electrical Networks
 - View edge st as resistor of resistance $1/u_{st}$.
 - Impose voltage x_v at every vertex v .
 - Ohm's Power Law: $P = V^2/R$.
 - Power consumed on edge st is $u_{st} (x_s - x_t)^2$.
 - Total power consumed is $x^T L_G x$.

Spectral Sparsifiers (Spielman-Teng '04)

- **Input:** An undirected graph $G=(V,E)$ with weights $u : E \rightarrow \mathbb{R}_+$
- **Def:** The **Laplacian** is the matrix L_G such that

$$x^T L_G x = \sum_{st \in E} u_{st} (x_s - x_t)^2 \quad \forall x \in \mathbb{R}^V.$$
- **Output:** A subgraph $H=(V,F)$ of G with weights $w : F \rightarrow \mathbb{R}$ such that $|F|$ is small and

$$x^T L_H x = (1 \pm \epsilon) x^T L_G x \quad \forall x \in \mathbb{R}^V$$

$\Downarrow$ Restrict to $\{0,1\}$ -vectors

$$w(\delta_H(U)) = (1 \pm \epsilon) u(\delta_G(U)) \quad \forall U \subseteq V$$

**Spectral
Sparsifier**



**Cut
Sparsifier**

Cut vs Spectral Sparsifiers

- Number of Constraints:

- Cut: $w(\delta_H(U)) = (1 \pm \epsilon) u(\delta_G(U)) \quad \forall U \subseteq V$ (2^n constraints)

- Spectral: $x^T L_H x = (1 \pm \epsilon) x^T L_G x \quad \forall x \in \mathbb{R}^V$ (∞ constraints)

- Spectral constraints are SDP feasibility constraints:

$$(1-\epsilon) x^T L_G x \leq x^T L_H x \leq (1+\epsilon) x^T L_G x \quad \forall x \in \mathbb{R}^V$$

$$\Leftrightarrow (1-\epsilon) L_G \preceq L_H \preceq (1+\epsilon) L_G$$

Here $X \preceq Y$ means $Y-X$ is positive semidefinite

- Spectral constraints are actually easier to handle

- Checking “Is H is a spectral sparsifier of G?” is in P

- Checking “Is H is a cut sparsifier of G?” is non-uniform sparsest cut, so NP-hard

Application of Spectral Sparsifiers

- Suppose you wanted to solve $L_G x = b$
- Since $(1-\epsilon) L_G \preceq L_H \preceq (1+\epsilon) L_G$
we have $(1-\epsilon) L_G^{-1} b \leq L_H^{-1} b \leq (1+\epsilon) L_G^{-1} b$
- **Hope:** Solving $L_H^{-1} b$ is easier since H is sparse
- **Theorem:** [Spielman-Teng '04, Koutis-Miller-Peng '10]
An approximate solution to $L_G x = b$ can be found
in $O(m \log n (\log \log n)^2)$ time. ($m = \#$ edges of G)

Results on Sparsifiers

Cut Sparsifiers

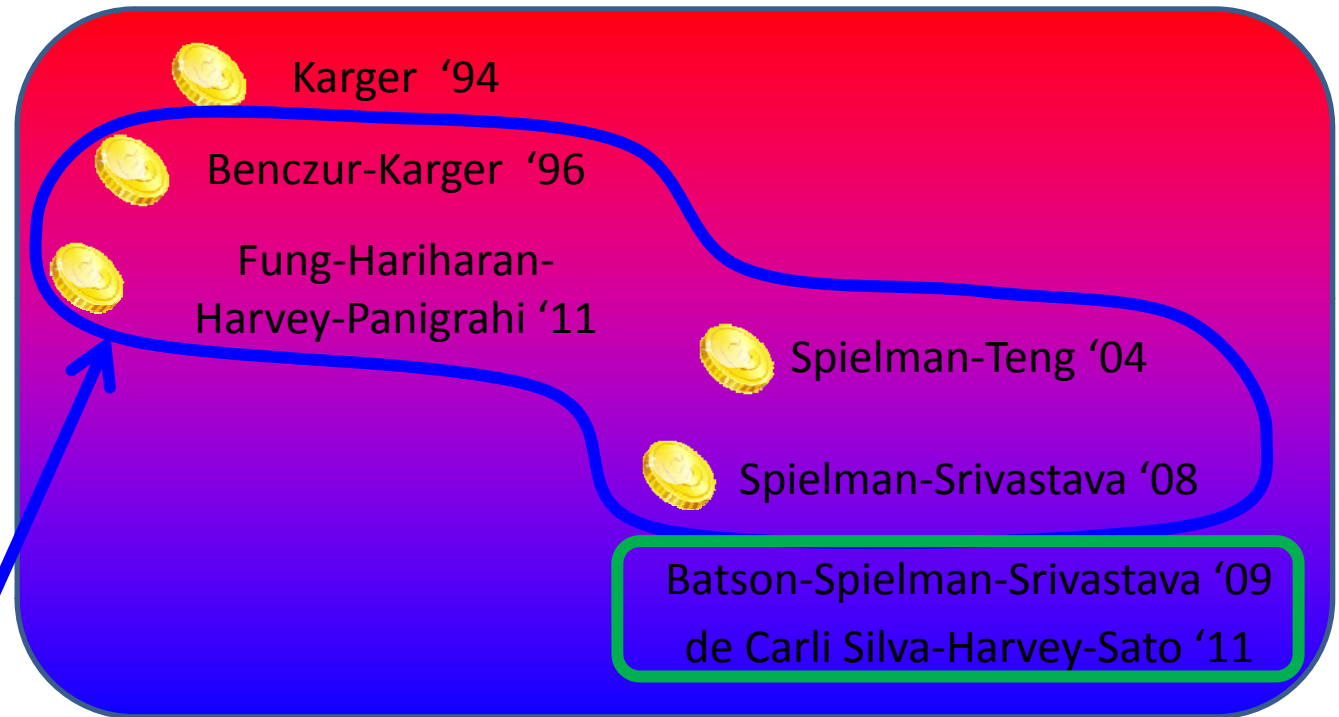
Spectral Sparsifiers

Combinatorial

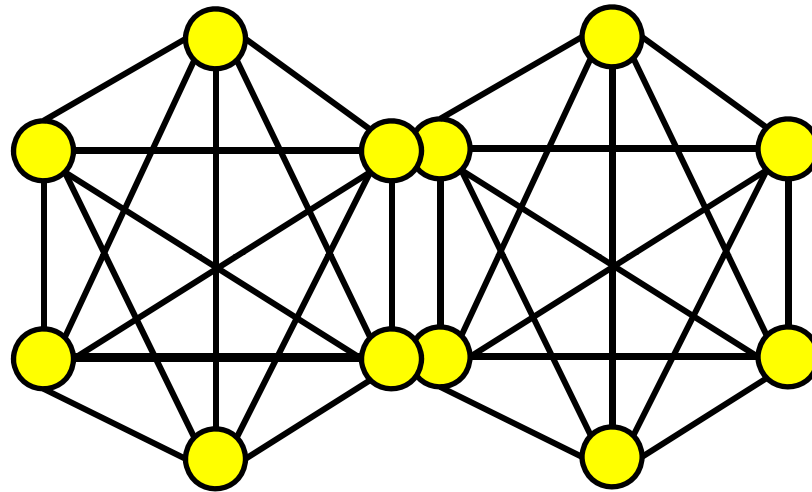
Linear Algebraic

Construct sparsifiers with $n \log^{O(1)} n / \epsilon^2$ edges,
in nearly linear time

Construct sparsifiers with $O(n/\epsilon^2)$ edges,
in $\text{poly}(n)$ time

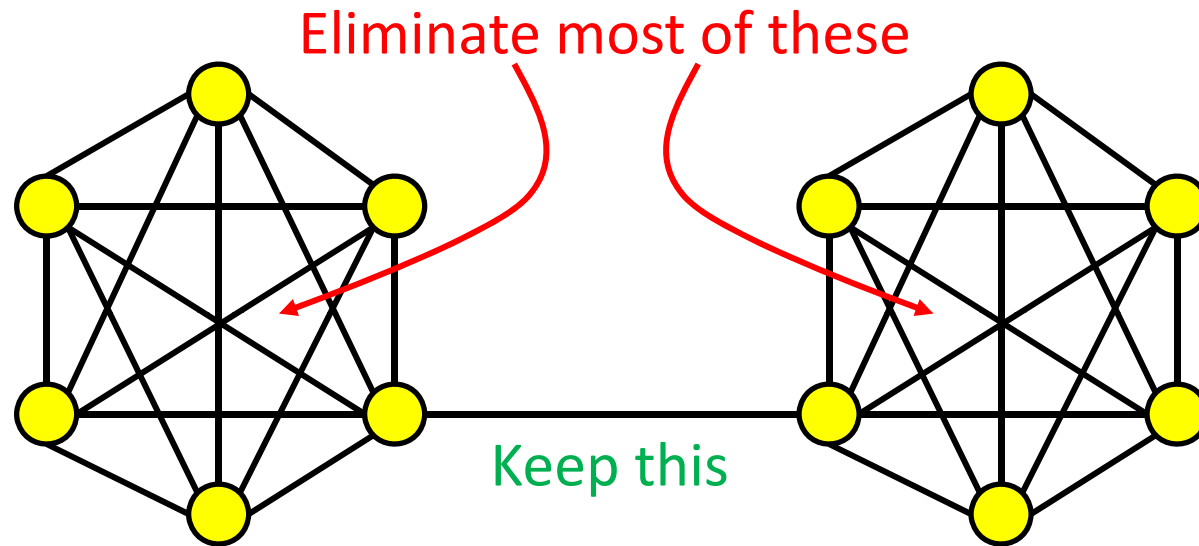


Sparsifiers by Random Sampling



- The complete graph is easy!
Random sampling gives an expander (ie. sparsifier)
with $O(n \log n)$ edges.

Sparsifiers by Random Sampling



- Can't sample edges with same probability!
- Idea [BK'96]
Sample low-connectivity edges with high probability,
and high-connectivity edges with low probability

Non-uniform sampling algorithm [BK'96]

- **Input:** Graph $G=(V,E)$, weights $u : E \rightarrow \mathbb{R}_+$
- **Output:** A subgraph $H=(V,F)$ with weights $w : F \rightarrow \mathbb{R}_+$

Choose parameter ρ
Compute probabilities $\{p_e : e \in E\}$
For $i=1$ to ρ
 For each edge $e \in E$
 With probability p_e ,
 Add e to F
 Increase w_e by $u_e/(\rho p_e)$

Can we do **this** so that the **cut values** are tightly concentrated and $E[|F|] = n \log^{O(1)} n$?

- **Note:** $E[|F|] \leq \rho \cdot \sum_e p_e$
- **Note:** $E[w_e] = u_e \quad \forall e \in E$
 $\Rightarrow$ For every $U \subseteq V$, $E[w(\delta_H(U))] = u(\delta_G(U))$

Benczur-Karger '96

- **Input:** Graph $G=(V,E)$, weights $u : E \rightarrow \mathbb{R}_+$
- **Output:** A subgraph $H=(V,F)$ with weights $w : F \rightarrow \mathbb{R}_+$

Choose parameter ρ
Compute probabilities $\{ p_e : e \in E \}$
For $i=1$ to ρ

But what is “strength”?
Can’t we use “connectivity”?

Add e to F
Increase w_e by $u_e/(\rho p_e)$

Can we do **this** so that the **cut values** are tightly concentrated and $E[|F|] = n \log^{O(1)} n$?

Can **approximate** all values in $m \log^{O(1)} n$ time.

- Set $\rho = O(\log n / \epsilon^2)$.
- Let $p_e = 1/\text{“strength”}$ of edge e .
- Cuts are preserved to within $(1 \pm \epsilon)$ and $E[|F|] = O(n \log n / \epsilon^2)$

Fung-Hariharan-Harvey-Panigrahi '11

- **Input:** Graph $G=(V,E)$, weights $u : E \rightarrow \mathbb{R}_+$
- **Output:** A subgraph $H=(V,F)$ with weights $w : F \rightarrow \mathbb{R}_+$

Choose parameter ρ
Compute probabilities $\{p_e : e \in E\}$
For $i=1$ to ρ
 For each edge $e \in E$
 With probability p_e ,
 Add e to F
 Increase w_e by $u_e/(\rho p_e)$

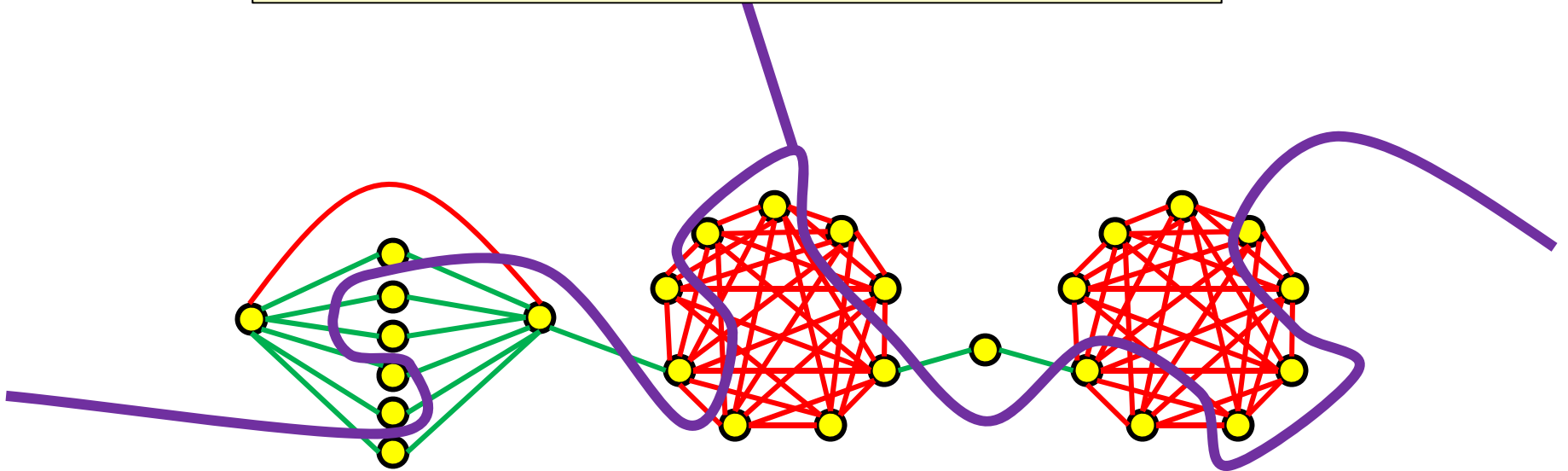
Can we do **this** so that the **cut values** are tightly concentrated and $E[|F|] = n \log^{O(1)} n$?

Can **approximate** all values in $O(m + n \log n)$ time

- Set $\rho = O(\log^2 n / \epsilon^2)$.
- Let $p_{st} = 1/(\text{min cut separating } s \text{ and } t)$
- Cuts are preserved to within $(1 \pm \epsilon)$ and $E[|F|] = O(n \log^2 n / \epsilon^2)$

Overview of Analysis

Most cuts hit a huge number of edges
 $\Rightarrow$ extremely concentrated
 $\Rightarrow$ whp, most cuts are close to their mean



Overview of Analysis

Hits only one red edge
 $\Rightarrow$ poorly concentrated

Hits many red edges
 $\Rightarrow$ reasonably concentrated

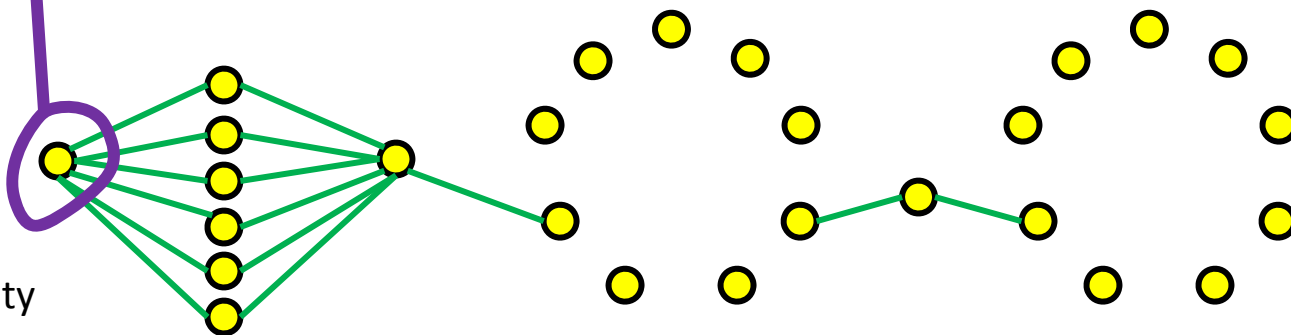
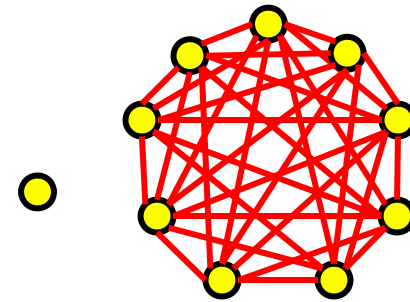
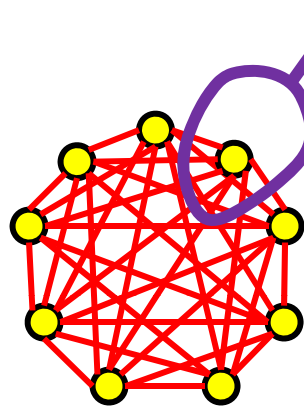
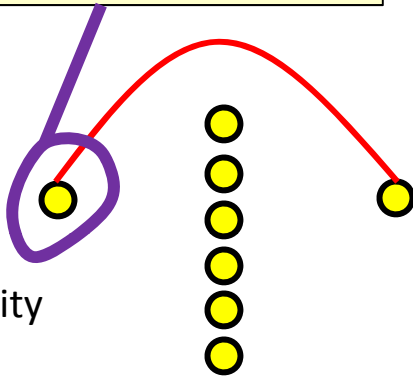
High connectivity

Low sampling probability

The same cut also hits many green edges
 $\Rightarrow$ highly concentrated

Low connectivity

High sampling probability



Summary for Cut Sparsifiers

- Do non-uniform sampling of edges, with probabilities based on “connectivity”
- Decomposes graph into “connectivity classes” and argue concentration of all cuts
- BK’96 used “strength” not “connectivity”
- Can get sparsifiers with $O(n \log n / \epsilon^2)$ edges
 - Optimal for any independent sampling algorithm

Spectral Sparsification

- **Input:** Graph $G=(V,E)$, weights $u : E \rightarrow \mathbb{R}_+$

- **Recall:** $x^T L_G x = \sum_{st \in E} \underbrace{u_{st} (x_s - x_t)^2}_{\text{Call this } x^T L_{st} x}$

- **Goal:** Find weights $w : E \rightarrow \mathbb{R}_+$ such that most w_e are zero, and

$$(1-\epsilon) x^T L_G x \leq \sum_{e \in E} w_e x^T L_e x \leq (1+\epsilon) x^T L_G x \quad \forall x \in \mathbb{R}^V$$

$$\Leftrightarrow (1-\epsilon) L_G \preceq \sum_{e \in E} w_e L_e \preceq (1+\epsilon) L_G$$

- **General Problem:** Given matrices L_e satisfying $\sum_e L_e = L_G$, find coefficients w_e , mostly zero, such that

$$(1-\epsilon) L_G \preceq \sum_e w_e L_e \preceq (1+\epsilon) L_G$$

The General Problem:

Sparsifying Sums of PSD Matrices

- **General Problem:** Given PSD matrices L_e s.t. $\sum_e L_e = L$, find coefficients w_e , mostly zero, such that
$$(1-\epsilon) L \preceq \sum_e w_e L_e \preceq (1+\epsilon) L$$
- **Theorem:** [Ahlswede-Winter '02]
Random sampling gives w with $O(n \log n / \epsilon^2)$ non-zeros.
- **Theorem:** [de Carli Silva-Harvey-Sato '11],
building on [Batson-Spielman-Srivastava '09]
Deterministic alg gives w with $O(n / \epsilon^2)$ non-zeros.
 - Cut & spectral sparsifiers with $O(n / \epsilon^2)$ edges [BSS'09]
 - Sparsifiers with more properties and $O(n / \epsilon^2)$ edges [dHS'11]

Vector Case

- **Vector problem:** Given vectors $\mathbf{v}_e \in [0,1]^n$ s.t. $\sum_e \mathbf{v}_e = \mathbf{v}$, find coefficients w_e , mostly zero, such that

$$\left\| \sum_e w_e \mathbf{v}_e - \mathbf{v} \right\|_{\infty} \leq \epsilon$$

- **Theorem** [Althofer '94, Lipton-Young '94]:
There is a w with $O(\log n/\epsilon^2)$ non-zeros.

- **Proof:** Random sampling & Hoeffding inequality. ■

$\Rightarrow$ ϵ -approx equilibria with $O(\log n/\epsilon^2)$ support in zero-sum games

- **Multiplicative version:** There is a w with $O(n \log n/\epsilon^2)$ non-zeros such that

$$(1-\epsilon) \mathbf{v} \leq \sum_e w_e \mathbf{v}_e \leq (1+\epsilon) \mathbf{v}$$

Concentration Inequalities

- **Theorem:** [Chernoff '52, Hoeffding '63]

Let $Y_1, \dots, Y_k$ be i.i.d. **random non-negative real numbers** s.t. $E[Y_i] = Z$ and $Y_i \leq uZ$. Then

$$\Pr \left[(1 - \epsilon)Z \leq \sum_{i=1}^k \frac{Y_i}{k} \leq (1 + \epsilon)Z \right] \geq 1 - 2 \exp(-\Omega(\epsilon^2 k/u))$$

- **Theorem:** [Ahlsvede-Winter '02]

Let $Y_1, \dots, Y_k$ be i.i.d. **random PSD nxn matrices** s.t. $E[Y_i] = Z$ and $Y_i \preceq uZ$. Then

$$\Pr \left[(1 - \epsilon)Z \preceq \sum_{i=1}^k \frac{Y_i}{k} \preceq (1 + \epsilon)Z \right] \geq 1 - 2n \exp(-\Omega(\epsilon^2 k/u))$$

 The only difference

“Balls & Bins” Example

- **Problem:** Throw k balls into n bins. Want $(\text{max load}) / (\text{min load}) \leq 1+\epsilon$. How big should k be?
- **AW Theorem:** Let $Y_1, \dots, Y_k$ be i.i.d. random PSD matrices such that $E[Y_i] = Z$ and $Y_i \preceq uZ$. Then

$$\Pr \left[(1 - \epsilon)Z \preceq \sum_{i=1}^k \frac{Y_i}{k} \preceq (1 + \epsilon)Z \right] \geq 1 - 2n \exp(-\Omega(\epsilon^2 k/u))$$

- **Solution:** Let Y_i be all zeros, except for a single n in a random diagonal entry.

$$Y_i = \begin{pmatrix} \ddots & & \\ & n & \\ & & \ddots \end{pmatrix}$$

Then $E[Y_i] = I$, and $Y_i \preceq nI$. Set $k = \Theta(n \log n / \epsilon^2)$.

Whp, every diagonal entry of $\sum_i Y_i / k$ is in $[1-\epsilon, 1+\epsilon]$.

Solving the General Problem

- **General Problem:** Given PSD matrices L_e s.t. $\sum_e L_e = L$, find coefficients w_e , mostly zero, such that

$$(1-\epsilon) L \preceq \sum_e w_e L_e \preceq (1+\epsilon) L$$

- **AW Theorem:** Let $Y_1, \dots, Y_k$ be i.i.d. random PSD matrices such that $E[Y_i] = Z$ and $Y_i \preceq uZ$. Then

$$\Pr \left[(1 - \epsilon)Z \preceq \sum_{i=1}^k \frac{Y_i}{k} \preceq (1 + \epsilon)Z \right] \geq 1 - 2n \exp(-\Omega(\epsilon^2 k/u))$$

- To solve General Problem with $O(n \log n / \epsilon^2)$ non-zeros
 - Repeat $k := \Theta(n \log n / \epsilon^2)$ times
 - Pick an edge e with probability $p_e := \text{Tr}(L_e L^{-1}) / n$
 - Increment w_e by $1/k \cdot p_e$

Derandomization

- **Vector problem:** Given vectors $\mathbf{v}_e \in [0,1]^n$ s.t. $\sum_e \mathbf{v}_e = \mathbf{v}$, find coefficients w_e , mostly zero, such that
$$\left\| \sum_e w_e \mathbf{v}_e - \mathbf{v} \right\|_{\infty} \leq \epsilon$$
- **Theorem** [Young '94]: The multiplicative weights method deterministically gives w with $O(\log n/\epsilon^2)$ non-zeros
 - Or, use pessimistic estimators on the Hoeffding proof
- **General Problem:** Given PSD matrices L_e s.t. $\sum_e L_e = L$, find coefficients w_e , mostly zero, such that
$$(1-\epsilon) L \preceq \sum_e w_e L_e \preceq (1+\epsilon) L$$
- **Theorem** [de Carli Silva-Harvey-Sato '11]:
The matrix multiplicative weights method (Arora-Kale '07) deterministically gives w with $O(n \log n/\epsilon^2)$ non-zeros
 - Or, use matrix pessimistic estimators (Wigderson-Xiao '06)

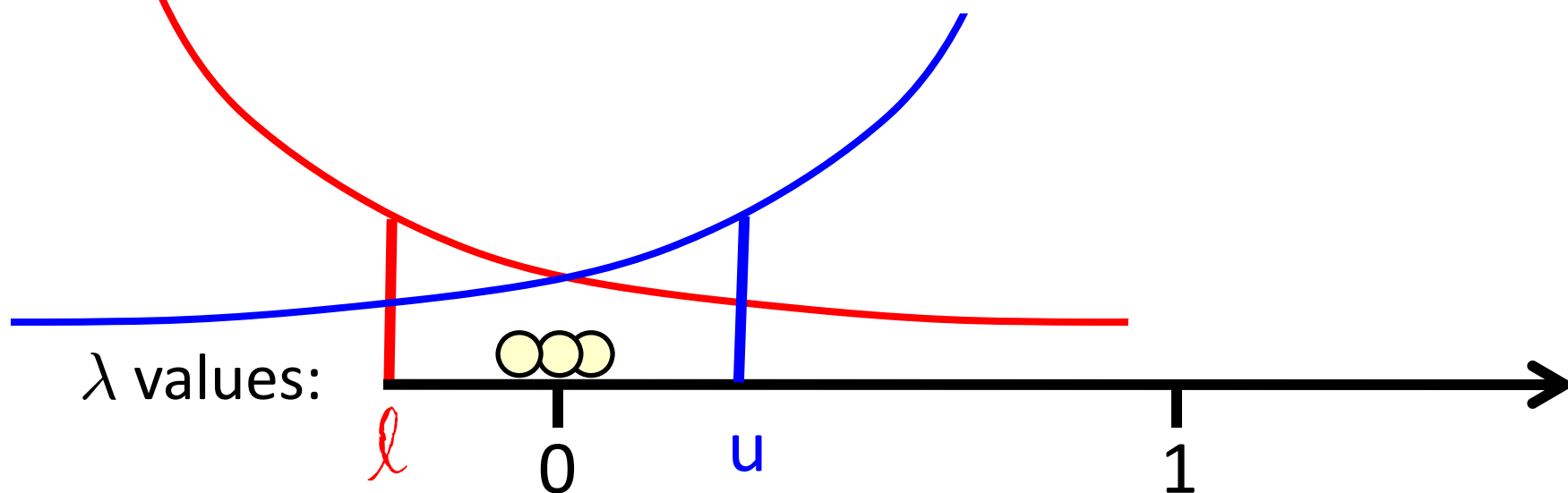
MWUM for “Balls & Bins”

- Let λ_i = load in bin i . Initially $\lambda=0$. Want: $1 \leq \lambda_i$ and $\lambda_i \leq 1$.
- Introduce penalty functions “ $\exp(\ell - \lambda_i)$ ” and “ $\exp(\lambda_i - u)$ ”
- Find a bin λ_i to throw a ball into such that, increasing ℓ by δ_ℓ and u by δ_u , the penalties don't grow.

$$\sum_i \exp(\ell + \delta_\ell - \lambda_i') \leq \sum_i \exp(\ell - \lambda_i)$$

$$\sum_i \exp(\lambda_i' - (u + \delta_u)) \leq \sum_i \exp(\lambda_i - u)$$

- Careful analysis shows $O(n \log n / \epsilon^2)$ balls is enough



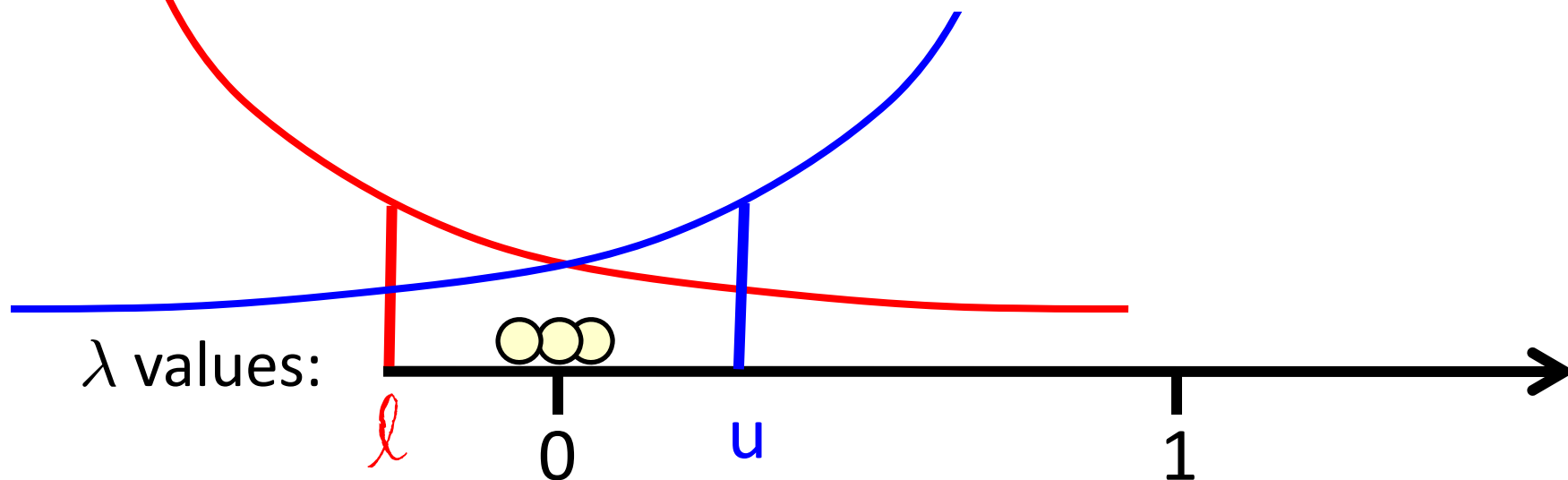
MMWUM for General Problem

- Let $A=0$ and λ its eigenvalues. Want: $1 \leq \lambda_i$ and $\lambda_i \leq 1$.
- Use penalty functions $\text{Tr exp}(\ell I - A)$ and $\text{Tr exp}(A - uI)$
- Find a matrix L_e such that adding αL_e to A , increasing ℓ by δ_ℓ and u by δ_u , the penalties don't grow.

$$\text{Tr exp}((\ell + \delta_\ell)I - (A + \alpha L_e)) \leq \text{Tr exp}(\ell I - A)$$

$$\text{Tr exp}((A + \alpha L_e) - (u + \delta_u)I) \leq \text{Tr exp}(A - uI)$$

- Careful analysis shows $O(n \log n / \epsilon^2)$ matrices is enough



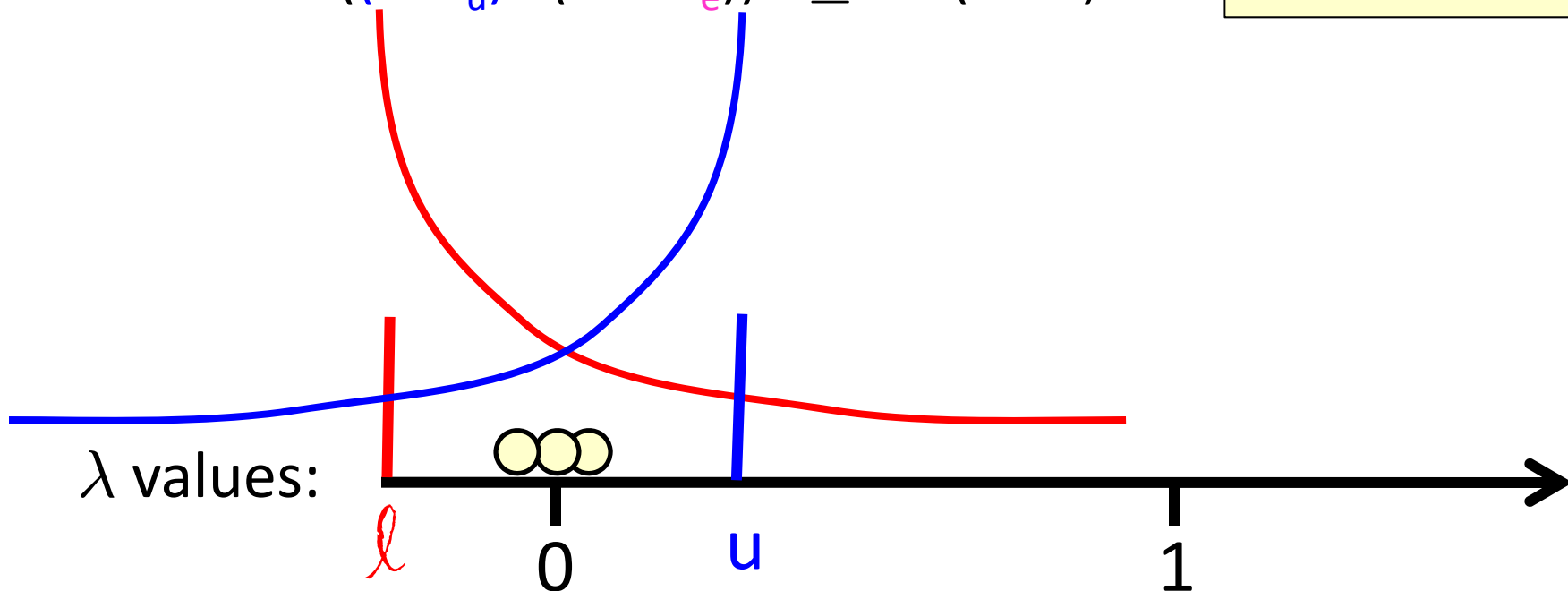
Beating Sampling & MMWUM

- To get a better bound, try changing the penalty functions to be steeper!
- Use penalty functions $\text{Tr} (A - \ell I)^{-1}$ and $\text{Tr} (u I - A)^{-1}$
- Find a matrix L_e such that adding αL_e to A , increasing ℓ by δ_ℓ and u by δ_u , the penalties don't grow.

$$\text{Tr} ((A + \alpha L_e) - (\ell + \delta_\ell) I)^{-1} \leq \text{Tr} (A - \ell I)^{-1}$$

$$\text{Tr} ((u + \delta_u) I - (A + \alpha L_e))^{-1} \leq \text{Tr} (u I - A)^{-1}$$

All eigenvalues
stay within $[\ell, u]$



Beating Sampling & MMWUM

- To get a better bound, try changing the penalty functions to be steeper!
- Use penalty functions $\text{Tr} (A - \ell I)^{-1}$ and $\text{Tr} (uI - A)^{-1}$
- Find a matrix L_e such that adding αL_e to A , increasing ℓ by δ_ℓ and u by δ_u , the penalties don't grow.

$$\text{Tr} ((A + \alpha L_e) - (\ell + \delta_\ell) I)^{-1} \leq \text{Tr} (A - \ell I)^{-1}$$

$$\text{Tr} ((u + \delta_u) I - (A + \alpha L_e))^{-1} \leq \text{Tr} (uI - A)^{-1}$$

- **General Problem:** Given PSD matrices L_e s.t. $\sum_e L_e = L$, find coefficients w_e , mostly zero, such that

$$(1 - \epsilon) L \preceq \sum_e w_e L_e \preceq (1 + \epsilon) L$$

- **Theorem:** [Batson-Spielman-Srivastava '09] in rank-1 case, [de Carli Silva-Harvey-Sato '11] for general case
This gives a solution w with $O(n/\epsilon^2)$ non-zeros.

Applications

- **Theorem:** [de Carli Silva-Harvey-Sato '11]

Given PSD matrices L_e s.t. $\sum_e L_e = L$, there is an algorithm to find w with $O(n/\epsilon^2)$ non-zeros such that

$$(1-\epsilon) L \preceq \sum_e w_e L_e \preceq (1+\epsilon) L$$

- **Application 1:** Spectral Sparsifiers with Costs

Given costs on edges of G , can find sparsifier H whose cost is at most $(1+\epsilon)$ the cost of G .

- **Application 2:** Sparse SDP Solutions

$\min \{ c^T y : \sum_i y_i A_i \succeq B, y \geq 0 \}$ where A_i 's and B are PSD
has nearly optimal solution with $O(n/\epsilon^2)$ non-zeros.

Open Questions

- Sparsifiers for directed graphs
- More constructions of sparsifiers with $O(n/\epsilon^2)$ edges. Perhaps randomized?
- Iterative construction of expander graphs
- More control of the weights w_e
- A combinatorial proof of spectral sparsifiers
- More applications of our general theorem