Graphs in Machine Learning

Cut Graph Sparsifiers

Benczur-Karger Algorithm

Michal Valko

Inria & ENS Paris-Saclay, MVA

Partially based on material by: Rob Fergus, Nikhil Srivastava, Yiannis Koutis, Joshua Batson, Daniel Spielman
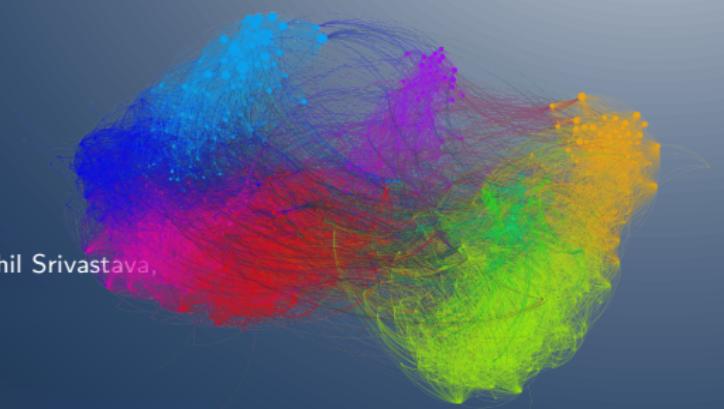
Cut Graph Sparsifiers

Define G and H are $(1 \pm \varepsilon)$-**cut similar** when $\forall S$

$$(1 - \varepsilon)\mathsf{cut}_H(S) \leq \mathsf{cut}_G(S) \leq (1 + \varepsilon)\mathsf{cut}_H(S)$$

Cut Graph Sparsifiers

Define G and H are $(1 \pm \varepsilon)$-**cut similar** when $\forall S$

$$(1 - \varepsilon)\mathsf{cut}_H(S) \leq \mathsf{cut}_G(S) \leq (1 + \varepsilon)\mathsf{cut}_H(S)$$

Why did they care?

Cut Graph Sparsifiers

Define G and H are $(1 \pm \varepsilon)$-**cut similar** when $\forall S$

$$(1 - \varepsilon)\mathsf{cut}_H(S) \leq \mathsf{cut}_G(S) \leq (1 + \varepsilon)\mathsf{cut}_H(S)$$

Why did they care? faster mincut/maxflow

Cut Graph Sparsifiers

Define G and H are $(1 \pm \varepsilon)$-**cut similar** when $\forall S$

$$(1 - \varepsilon)\mathsf{cut}_H(S) \leq \mathsf{cut}_G(S) \leq (1 + \varepsilon)\mathsf{cut}_H(S)$$

Why did they care? faster mincut/maxflow

Is this always possible?

Cut Graph Sparsifiers

Define G and H are $(1 \pm \varepsilon)$-**cut similar** when $\forall S$

$$(1 - \varepsilon)\mathsf{cut}_H(S) \leq \mathsf{cut}_G(S) \leq (1 + \varepsilon)\mathsf{cut}_H(S)$$

Why did they care? faster mincut/maxflow

Is this always possible? Benczúr and Karger (1996): Yes!

$\forall \varepsilon \; \exists \; (1 + \varepsilon)$-cut similar H with $\mathcal{O}(n \log n / \varepsilon^2)$ edges s.t. $E_H \subseteq E$ and computable in $\mathcal{O}(m \log^3 n + m \log n / \varepsilon^2)$ time n nodes, m edges

Cut Graph Sparsifiers

$G = K_n$

$H = d$-**regular** (random)

Cut Graph Sparsifiers

$G = K_n$
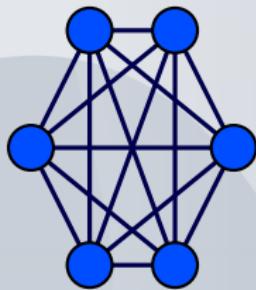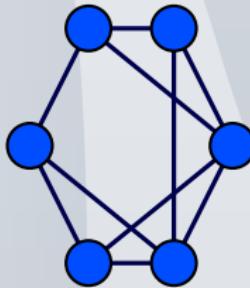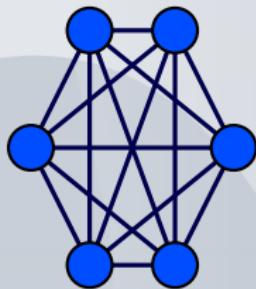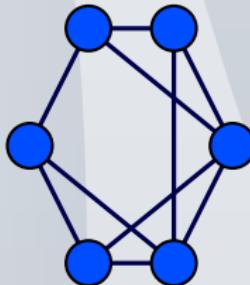
$H = d$-**regular** (random)

How many edges?

Cut Graph Sparsifiers

$G = K_n$
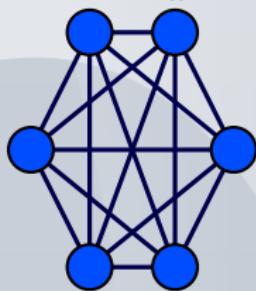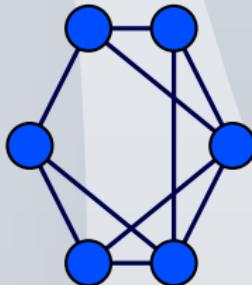
$H = d$-**regular** (random)

How many edges?

$|E_G| = \mathcal{O}(n^2)$

Cut Graph Sparsifiers

$G = K_n$
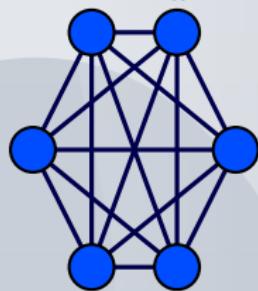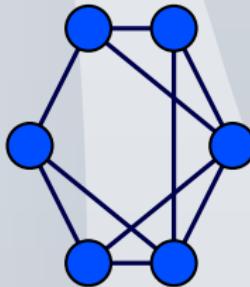
$H = d$-**regular** (random)

How many edges?

$|E_G| = \mathcal{O}(n^2)$

$|E_H| = \mathcal{O}(dn)$

Cut Graph Sparsifiers
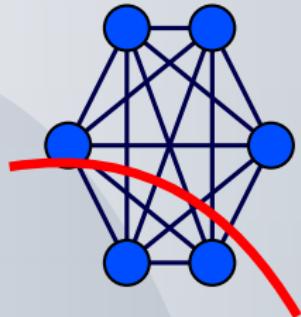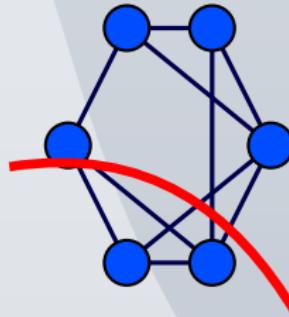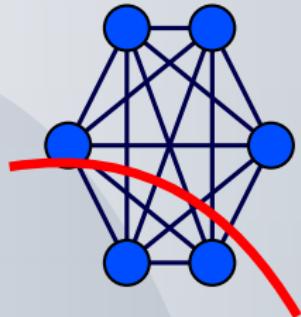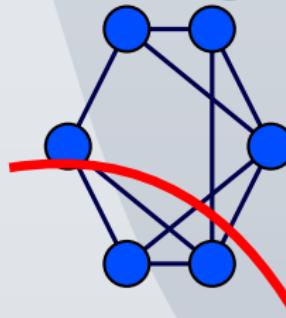
$G = K_n$

$H = d\text{-regular}$ (random)

Cut Graph Sparsifiers

$G = K_n$

$H = d\text{-regular}$ (random)

What are the cut weights for any S?

Cut Graph Sparsifiers
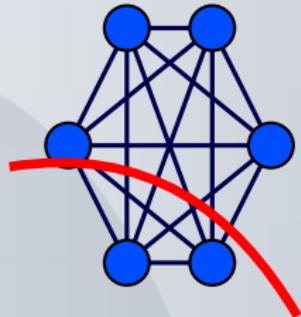
$G = K_n$ $H = d$-**regular** (random)
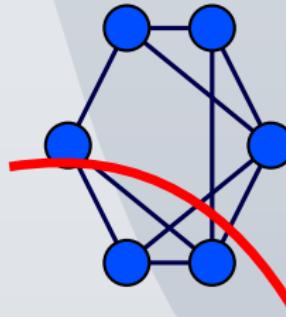
What are the cut weights for any S?

$$w_G(\delta S) = |S| \cdot |\overline{S}|$$

Cut Graph Sparsifiers
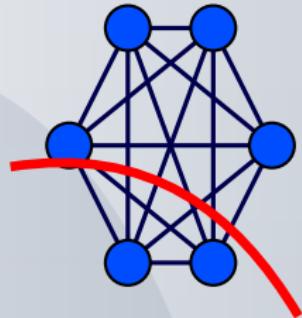
$G = K_n$ $H = d$-**regular** (random)
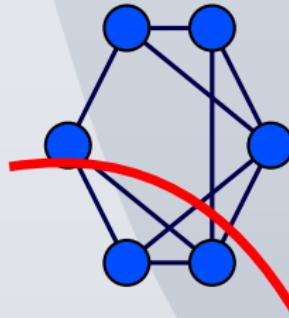
What are the cut weights for any S?

$$w_G(\delta S) = |S| \cdot |\overline{S}| \qquad\qquad w_H(\delta S) \approx \frac{d}{n} \cdot |S| \cdot |\overline{S}|$$

$$\forall S \subset V : \frac{w_G(\delta S)}{w_H(\delta S)} \approx \frac{n}{d}$$

Cut Graph Sparsifiers
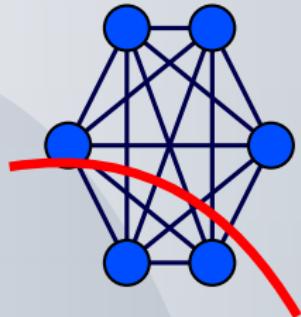
$G = K_n$ $H = d\text{-regular}$ (random)
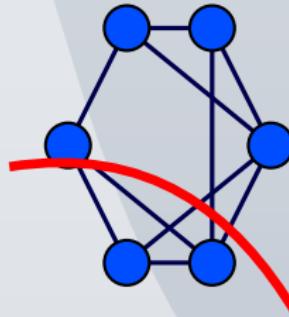
What are the cut weights for any S?

$$w_G(\delta S) = |S| \cdot |\overline{S}| \qquad w_H(\delta S) \approx \frac{d}{n} \cdot |S| \cdot |\overline{S}|$$

$$\forall S \subset V : \frac{w_G(\delta S)}{w_H(\delta S)} \approx \frac{n}{d}$$

Could be large :(

Cut Graph Sparsifiers
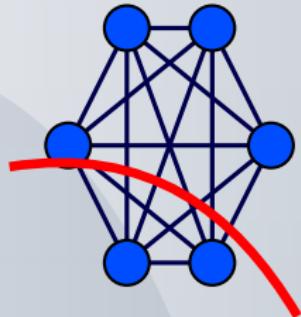
$G = K_n$ $H = d\text{-regular}$ (random)
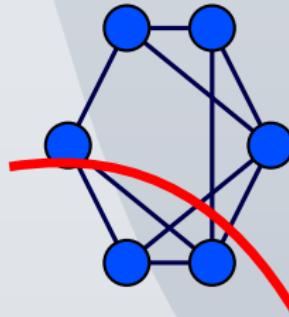
What are the cut weights for any S?

$$w_G(\delta S) = |S| \cdot |\overline{S}| \qquad\qquad w_H(\delta S) \approx \frac{d}{n} \cdot |S| \cdot |\overline{S}|$$

$$\forall S \subset V : \frac{w_G(\delta S)}{w_H(\delta S)} \approx \frac{n}{d}$$

Could be large :(What to do?

Cut Graph Sparsifiers
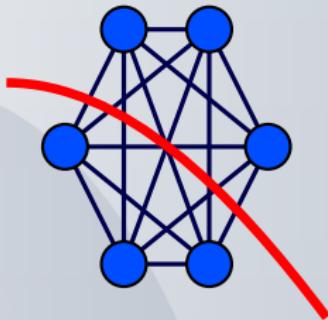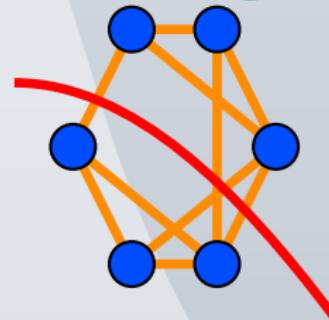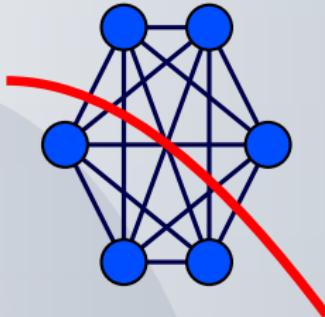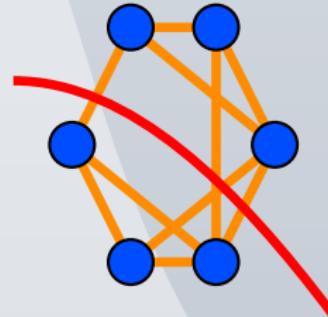
$G = K_n$

$H = d\text{-}\textbf{regular}$ (random)

Cut Graph Sparsifiers

$G = K_n$ $H = d\text{-}\mathbf{regular}$ (random)

What are the cut weights for any S?

Cut Graph Sparsifiers

$G = K_n$

$H = d$-**regular** (random)

What are the cut weights for any S?

$$w_G(\delta S) = |S| \cdot |\overline{S}|$$

Cut Graph Sparsifiers

$G = K_n$
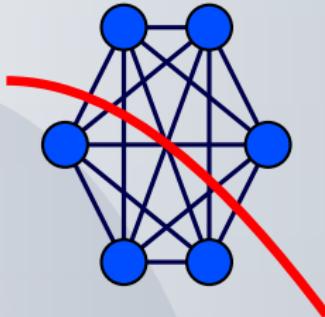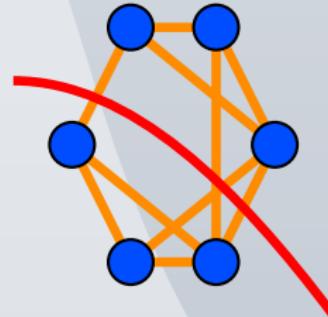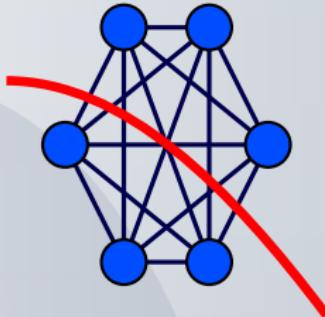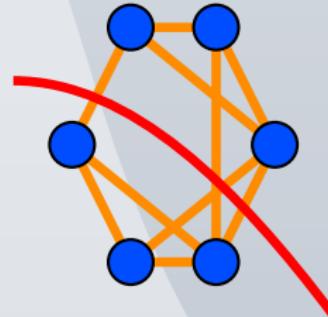
$H = d\text{-regular}$ (random)

What are the cut weights for any S?

$$w_G(\delta S) = |S| \cdot |\overline{S}| \qquad w_H(\delta S) \approx \frac{d}{n} \cdot \frac{n}{d} \cdot |S| \cdot |\overline{S}|$$

$$\forall S \subset V : \frac{w_G(\delta S)}{w_H(\delta S)} \approx 1$$

Cut Graph Sparsifiers
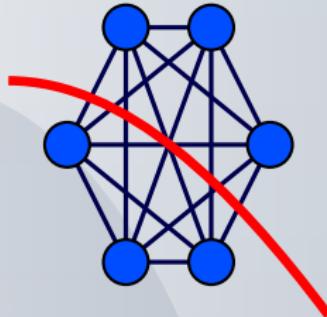
$G = K_n$ $H = d\text{-regular}$ (random)
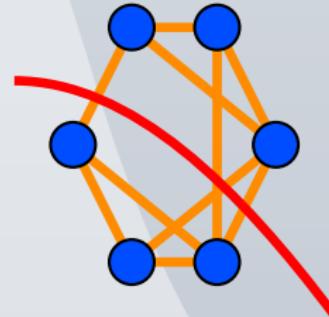
What are the cut weights for any S?

$$w_G(\delta S) = |S| \cdot |\overline{S}| \qquad w_H(\delta S) \approx \frac{d}{n} \cdot \frac{n}{d} \cdot |S| \cdot |\overline{S}|$$

$$\forall S \subset V : \frac{w_G(\delta S)}{w_H(\delta S)} \approx 1$$

Benczúr & Karger: Can find such H quickly for any G!

Graph Sparsification: What is good sparse?

Recall if $\mathbf{f} \in \{0,1\}^n$ represents S then $\mathbf{f}^\mathsf{T} \mathbf{L}_G \mathbf{f} =$

Graph Sparsification: What is good sparse?

Recall if $\mathbf{f} \in \{0,1\}^n$ represents S then $\mathbf{f}^\intercal \mathbf{L}_G \mathbf{f} = \text{cut}_G(S)$

Graph Sparsification: What is good sparse?

Recall if $\mathbf{f} \in \{0,1\}^n$ represents S then $\mathbf{f}^\intercal \mathbf{L}_G \mathbf{f} = \text{cut}_G(S)$

$$(1-\varepsilon)\text{cut}_H(S) \leq \text{cut}_G(S) \leq (1+\varepsilon)\text{cut}_H(S)$$

Graph Sparsification: What is good sparse?

Recall if $\mathbf{f} \in \{0,1\}^n$ represents S then $\mathbf{f}^\mathsf{T} \mathbf{L}_G \mathbf{f} = \text{cut}_G(S)$

$$(1-\varepsilon)\text{cut}_H(S) \leq \text{cut}_G(S) \leq (1+\varepsilon)\text{cut}_H(S)$$

becomes

$$(1-\varepsilon)\mathbf{f}^\mathsf{T} \mathbf{L}_H \mathbf{f} \leq \mathbf{f}^\mathsf{T} \mathbf{L}_G \mathbf{f} \leq (1+\varepsilon)\mathbf{f}^\mathsf{T} \mathbf{L}_H \mathbf{f}$$

Graph Sparsification: What is good sparse?

Recall if $\mathbf{f} \in \{0,1\}^n$ represents S then $\mathbf{f}^\mathsf{T}\mathbf{L}_G\mathbf{f} = \text{cut}_G(S)$

$$(1-\varepsilon)\text{cut}_H(S) \leq \text{cut}_G(S) \leq (1+\varepsilon)\text{cut}_H(S)$$

becomes

$$(1-\varepsilon)\mathbf{f}^\mathsf{T}\mathbf{L}_H\mathbf{f} \leq \mathbf{f}^\mathsf{T}\mathbf{L}_G\mathbf{f} \leq (1+\varepsilon)\mathbf{f}^\mathsf{T}\mathbf{L}_H\mathbf{f}$$

If we ask this only for $\mathbf{f} \in \{0,1\}^n \to (1+\varepsilon)$-cut similar _{combinatorial}
Benczúr & Karger (1996)

Graph Sparsification: What is good sparse?

Recall if $\mathbf{f} \in \{0,1\}^n$ represents S then $\mathbf{f}^\mathsf{T} \mathbf{L}_G \mathbf{f} = \text{cut}_G(S)$

$$(1-\varepsilon)\text{cut}_H(S) \leq \text{cut}_G(S) \leq (1+\varepsilon)\text{cut}_H(S)$$

becomes

$$(1-\varepsilon)\mathbf{f}^\mathsf{T} \mathbf{L}_H \mathbf{f} \leq \mathbf{f}^\mathsf{T} \mathbf{L}_G \mathbf{f} \leq (1+\varepsilon)\mathbf{f}^\mathsf{T} \mathbf{L}_H \mathbf{f}$$

If we ask this only for $\mathbf{f} \in \{0,1\}^n \to (1+\varepsilon)$-cut similar combinatorial

Benczúr & Karger (1996)

If we ask this for all $\mathbf{f} \in \mathbb{R}^n \to$

Graph Sparsification: What is good sparse?

Recall if $\mathbf{f} \in \{0, 1\}^n$ represents S then $\mathbf{f}^\mathsf{T} \mathbf{L}_G \mathbf{f} = \text{cut}_G(S)$

$$(1 - \varepsilon)\text{cut}_H(S) \leq \text{cut}_G(S) \leq (1 + \varepsilon)\text{cut}_H(S)$$

becomes

$$(1 - \varepsilon)\mathbf{f}^\mathsf{T} \mathbf{L}_H \mathbf{f} \leq \mathbf{f}^\mathsf{T} \mathbf{L}_G \mathbf{f} \leq (1 + \varepsilon)\mathbf{f}^\mathsf{T} \mathbf{L}_H \mathbf{f}$$

If we ask this only for $\mathbf{f} \in \{0, 1\}^n \to (1 + \varepsilon)$-cut similar combinatorial
Benczúr & Karger (1996)

If we ask this for all $\mathbf{f} \in \mathbb{R}^n \to (1 + \varepsilon)$-spectrally similar
Spielman & Teng (2004)

Graph Sparsification: What is **good** sparse?

Recall if $\mathbf{f} \in \{0,1\}^n$ represents S then $\mathbf{f}^\intercal \mathbf{L}_G \mathbf{f} = \text{cut}_G(S)$

$$(1-\varepsilon)\text{cut}_H(S) \leq \text{cut}_G(S) \leq (1+\varepsilon)\text{cut}_H(S)$$

becomes

$$(1-\varepsilon)\mathbf{f}^\intercal \mathbf{L}_H \mathbf{f} \leq \mathbf{f}^\intercal \mathbf{L}_G \mathbf{f} \leq (1+\varepsilon)\mathbf{f}^\intercal \mathbf{L}_H \mathbf{f}$$

If we ask this only for $\mathbf{f} \in \{0,1\}^n \rightarrow (1+\varepsilon)$-cut similar <small>combinatorial</small>
<small>Benczúr & Karger (1996)</small>
If we ask this for all $\mathbf{f} \in \mathbb{R}^n \rightarrow (1+\varepsilon)$-**spectrally similar**
<small>Spielman & Teng (2004)</small>

Spectral sparsifiers are stronger!

Graph Sparsification: What is good sparse?

Recall if $\mathbf{f} \in \{0,1\}^n$ represents S then $\mathbf{f}^\mathsf{T} \mathbf{L}_G \mathbf{f} = \text{cut}_G(S)$

$$(1-\varepsilon)\text{cut}_H(S) \leq \text{cut}_G(S) \leq (1+\varepsilon)\text{cut}_H(S)$$

becomes

$$(1-\varepsilon)\mathbf{f}^\mathsf{T}\mathbf{L}_H\mathbf{f} \leq \mathbf{f}^\mathsf{T}\mathbf{L}_G\mathbf{f} \leq (1+\varepsilon)\mathbf{f}^\mathsf{T}\mathbf{L}_H\mathbf{f}$$

If we ask this only for $\mathbf{f} \in \{0,1\}^n \to (1+\varepsilon)$-cut similar combinatorial
Benczúr & Karger (1996)
If we ask this for all $\mathbf{f} \in \mathbb{R}^n \to (1+\varepsilon)$-**spectrally similar**
Spielman & Teng (2004)

Spectral sparsifiers are stronger!

but checking for spectral similarity is easier

Michal Valko

michal.valko@inria.fr

Inria & ENS Paris-Saclay, MVA

https://misovalko.github.io/mva-ml-graphs.html