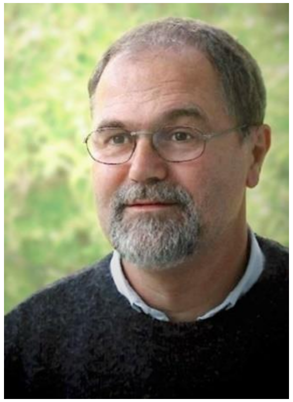
A Note On Spectral Clustering

Pavel Kolev and **Kurt Mehlhorn**



European Symposia on Algorithms'16

max planck institut informatik

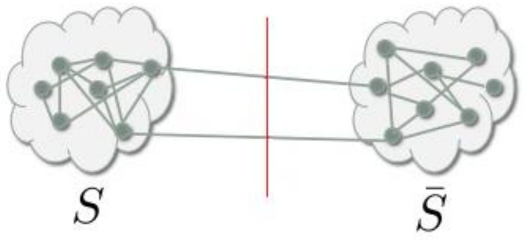
CLUSTER OF EXCELLENCE

Outline

- **Problem Formulation**
  - Algorithmic Tools
- Our Contribution
  - Structural Result
  - Algorithmic Result
    - Proof Overview
- Summary

k-way Partitioning

- **Def.** A **cluster** is a subset $S \subseteq V$ with small conductance



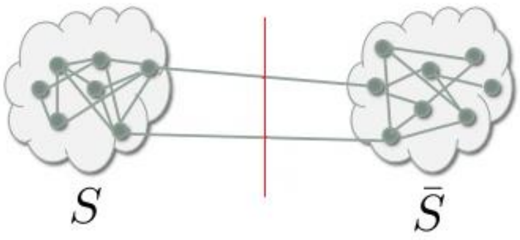
$\phi(S) = \frac{|E(S,\bar{S})|}{\mu(S)}$, where the volume $\mu(S) = \sum_{v \in S} \deg(v)$.

max planck institut informatik

CLUSTER OF EXCELLENCE

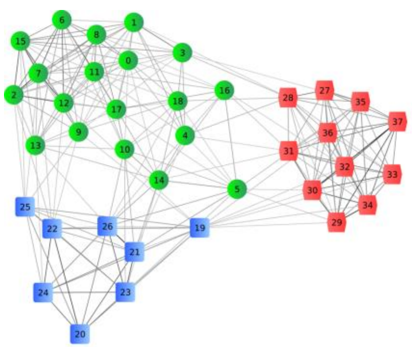
k-way Partitioning

- **Def.** A **cluster** is a subset $S \subseteq V$
  with small conductance



$\phi(S) = \frac{|E(S,\bar{S})|}{\mu(S)}$, where the volume $\mu(S) = \sum_{v \in S} \deg(v)$.

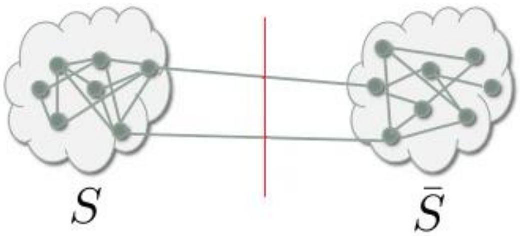
- **Def.** The order k conductance constant



$$\rho(k) = \min_{\text{partition } (P_1,\ldots,P_k)} \max_{i \in [1:k]} \phi(P_i)$$

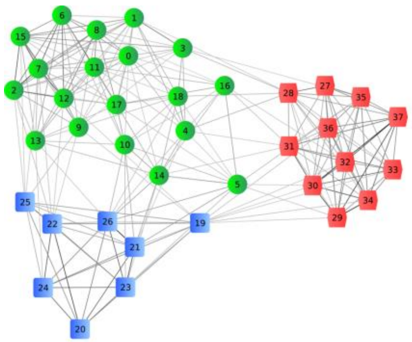
k-way Partitioning

- **Def.** A **cluster** is a subset $S \subseteq V$
  with small conductance



$$\phi(S) = \frac{|E(S,\bar{S})|}{\mu(S)},$$ where the volume $\mu(S) = \sum_{v \in S} \deg(v)$.

- **Def.** The order k conductance constant



$$\rho(k) = \min_{\text{partition } (P_1,\ldots,P_k)} \max_{i \in [1:k]} \phi(P_i)$$

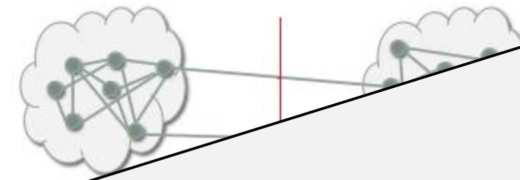
- **Goal:** Find an ***approximate*** k-way partition w.r.t $\rho(k)$.

max planck institut
informatik

CLUSTER OF EXCELLENCE

k-way Partitioning

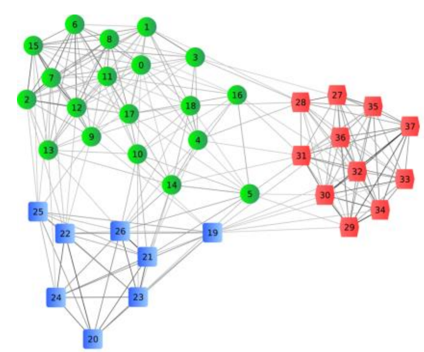
- **Def.** A **cluster** is a subset $S \subseteq V$
  with small conductance



$\phi(S) = \frac{|E(S,\bar{S})|}{\mu(S)}$, where the … $g(v)$.

- D… …ance constant

$$\min_{\text{partition } (P_1,\ldots,P_k)} \max_{i\in[1:k]} \phi(P_i)$$



Here, we analyze rigorously the Standard Spectral Clustering Paradigm.

- **Goal:** Find an ***approximate*** k-way partition w.r.t $\rho(k)$.

Standard Spectral Clustering Paradigm

Input: $G = (V, E)$, $3 \le k \ll n$ and $\epsilon \in (0,1)$.

Output: An *approximate* k-way partition of V.

Andrew Ng et al [NIPS'02]:

1. Computes an *approximate* Spectral Embedding $(F: V \mapsto R^k)$ using the Power Method.

2) Run a k-means clustering algorithm to compute an *approximate* k-way partition of $\{F(v)\}_{v \in V}$.

Outline

- Problem Formulation
  - **Algorithmic Tools**
- Our Contribution
  - Structural Result
  - Algorithmic Result
    - Proof Overview
- Summary

Spectral Graph Theory

- The normalized Laplacian matrix $\mathcal{L}$ has eigenvalues

$$0 = \lambda_1 \leq \cdots \leq \lambda_k \leq \lambda_{k+1} \leq \cdots \leq \lambda_n \leq 2.$$

- **Fact.** A graph has exactly k connected component iff

$$0 = \lambda_k < \lambda_{k+1}.$$

max planck institut informatik

CLUSTER OF EXCELLENCE

Spectral Graph Theory

- The normalized Laplacian matrix $\mathcal{L}$ has eigenvalues

$$0 = \lambda_1 \leq \cdots \leq \lambda_k \leq \lambda_{k+1} \leq \cdots \leq \lambda_n \leq 2.$$

- **Fact.** A graph has exactly k connected component iff

$$0 = \lambda_k < \lambda_{k+1}.$$

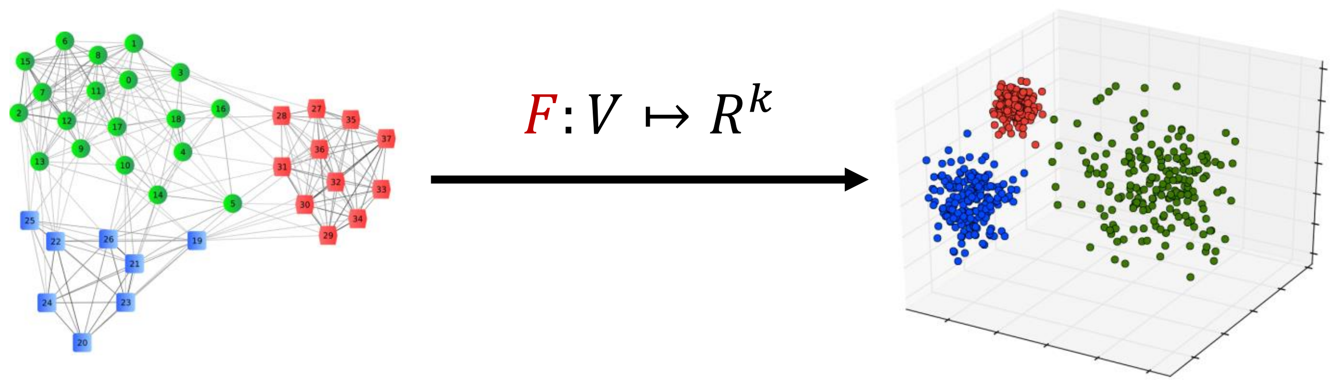
- **Trevisan et al. [STOC'12, SODA'14]** proved a robust version

$$\lambda_k/2 \leq \rho(k) \leq O(k^3)\sqrt{\lambda_k}.$$

($\rho(k)$ is **NP-hard** and λ_k is in **P**) → **approx**. scheme!

Exact Spectral Embedding

- $U_k = (v_1, v_2, \dots, v_k) \in R^{V \times k}$ - the bottom k eigenvectors of $\mathcal{L}$

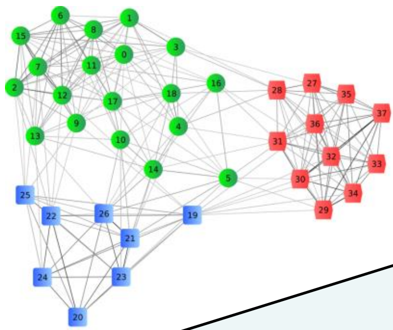


- Normalized Spectral Embedding:

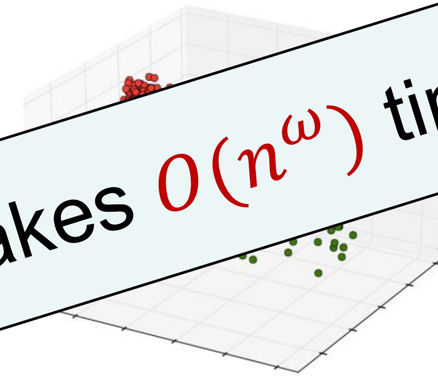
$$F(v) = \frac{1}{\sqrt{\deg(v)}} U_k(v,:), \text{ for every } v \in V.$$

Exact Spectral Embedding

- $U_k = (v_1, v_2, \dots, v_k) \in R^{V \times k}$ - the bottom k eigenvectors of $\mathcal{L}$



$F: V \mapsto R^k$



The exact computation of U_k takes $O(n^{\omega})$ time!

- zed Spectral Embedding:

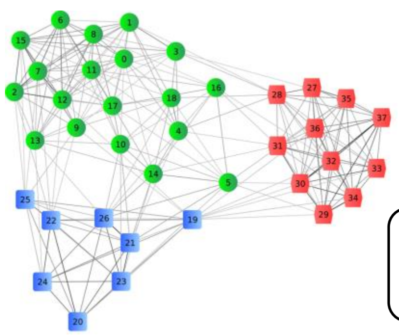
$$F(v) = \frac{1}{\sqrt{\deg(v)}} U_k(v,:), \text{ for every } v \in V.$$

max planck institut informatik

CLUSTER OF EXCELLENCE

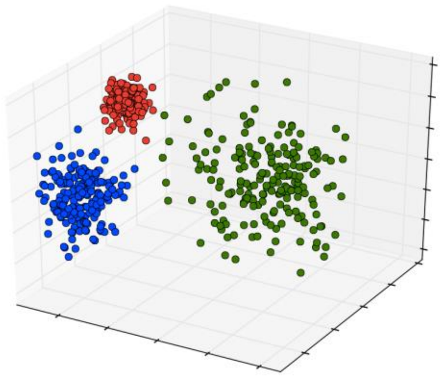
Approximate Spectral Embedding

- $\widetilde{U}_k \in R^{V \times k}$ approximation of the bottom k eigenvectors of $\mathcal{L}$



$\widetilde{F}: V \mapsto R^k$

Power Method

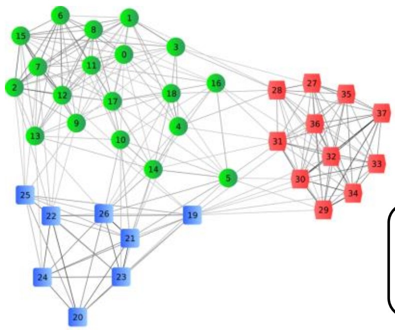


- Approximate Normalized Spectral Embedding:

$$\widetilde{F}(v) = \frac{1}{\sqrt{\deg(v)}} \widetilde{U}_k(v,:), \text{ for every } v \in V.$$

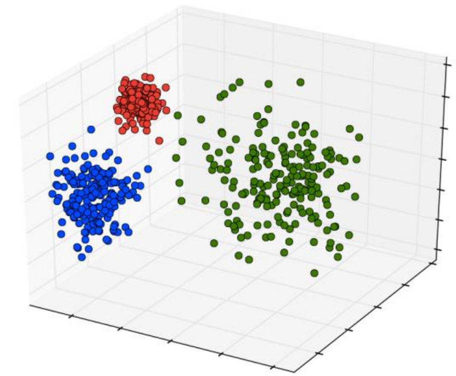
Approximate Spectral Embedding

- $\widetilde{U}_k \in R^{V \times k}$ approximation of the bottom k eigenvectors of $\mathcal{L}$



$\tilde{F}: V \mapsto R^k$

Power Method



- Approximate Normalized Spectral Embedding:

Point Sets:

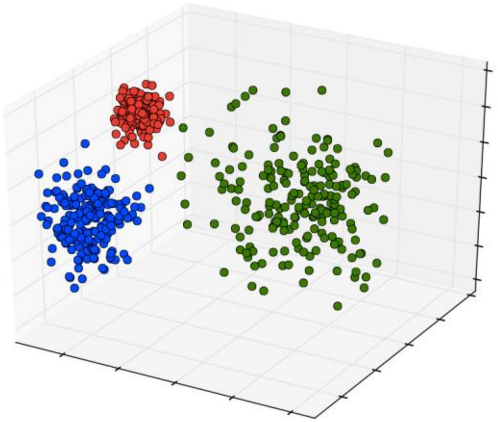
$$\widetilde{\mathcal{X}}_E = \{\deg(v) \text{ many copies of } \tilde{F}(v) \mid v \in V\}.$$

$$\widetilde{\mathcal{X}}_V = \{\tilde{F}(v) \mid v \in V\}.$$

max planck institut informatik

CLUSTER OF EXCELLENCE

k-means Clustering



Input: $\mathcal{X} = \{p_1, \ldots, p_n\}$ with $p_i \in R^k$.

Output: k-way partition of $\mathcal{X}$ such that

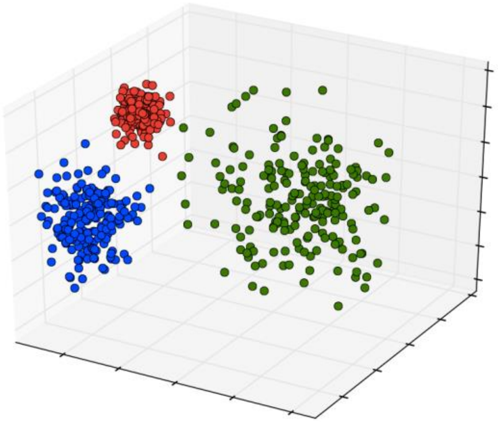
$$(A_1^\star, \ldots, A_k^\star) = \underset{\text{partition } (X_1,\ldots,X_k) \text{ of } \mathcal{X}}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{p \in X_i} \|p - c_i\|^2 ,$$

where c_i is the center of X_i.

max planck institut informatik

CLUSTER OF EXCELLENCE

k-means Clustering



Input: $\mathcal{X} = \{p_1, \dots, p_n\}$ with $p_i \in R^k$.

Output: k-way partition of $\mathcal{X}$ such that

$$(A_1^\star, \dots, A_k^\star) = \underset{\text{partition } (X_1,\dots,X_k) \text{ of } \mathcal{X}}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{p \in X_i} \|p - c_i\|^2 ,$$

where c_i is the center of X_i.

Def. The optimal k-means cost is

$$\Delta_k(\mathcal{X}) = \operatorname{cost}(A_1^\star, \dots, A_k^\star).$$

Outline

- Problem Formulation
  - Algorithmic Tools
- Our Contribution
  - **Structural Result**
  - Algorithmic Result
    - Proof Overview
- Summary

Structural Result

- **Peng et al. [COLT'15]**

  $\Upsilon := \lambda_{k+1}/\rho(k) \geq \Omega(k^3)$

$$\rho(k) = \max_{i \in [1:k]} \phi(P_i)$$

- **Our Result**

  $\Psi := \lambda_{k+1}/\rho_{\mathrm{avr}}(k) \geq \Omega(k^3)$

$$\rho_{\mathrm{avr}}(k) = \frac{1}{k}\sum_{i=1}^{k} \phi(P_i)$$

- $(P_1, \ldots, P_k)$ is an optimal k-way partition of G w.r.t. $\rho(k)$.

max planck institut informatik

CLUSTER OF EXCELLENCE

Structural Result

- **Peng et al. [COLT'15]**

  $\Upsilon := \lambda_{k+1}/\rho(k) \geq \Omega(k^3)$

$$\rho(k) = \max_{i\in[1:k]} \phi(P_i)$$

- **Our Result**

  $\Psi := \lambda_{k+1}/\rho_{\text{avr}}(k) \geq \Omega(k^3)$

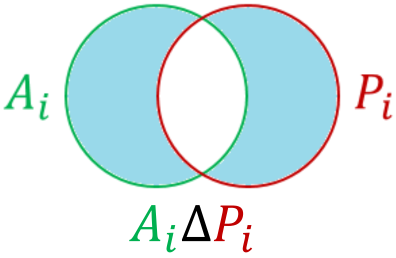
$$\rho_{\text{avr}}(k) = \frac{1}{k}\sum_{i=1}^{k} \phi(P_i)$$

- $(P_1, \dots, P_k)$ is an optimal k-way partition of G w.r.t. $\rho(k)$.
- $\text{cost}(A_1, \dots, A_k) \leq \gamma \cdot \Delta_k(\widetilde{\mathcal{X}}_E)$ for $\gamma \geq 1$.

max planck institut informatik

CLUSTER OF EXCELLENCE

Structural Result



- **Peng et al. [COLT'15]**

  If $\Upsilon := \lambda_{k+1}/\rho(k) \geq \Omega(k^3)$ then
  - $\mu(A_i \Delta P_i) \leq (\gamma/\Upsilon) \cdot \mu(P_i)$

$$\rho(k) = \max_{i \in [1:k]} \phi(P_i)$$

- **Our Result**

  If $\Psi := \lambda_{k+1}/\rho_{\mathrm{avr}}(k) \geq \Omega(k^3)$ then
  - $\mu(A_i \Delta P_i) \leq (\gamma/\Psi \boldsymbol{k}) \cdot \mu(P_i)$

$$\rho_{\mathrm{avr}}(k) = \frac{1}{k} \sum_{i=1}^{k} \phi(P_i)$$

- $(P_1, \ldots, P_k)$ is an optimal k-way partition of G w.r.t. $\rho(k)$.
- $\mathrm{cost}(A_1, \ldots, A_k) \leq \gamma \cdot \Delta_k(\widetilde{\mathcal{X}}_E)$ for $\gamma \geq 1$.

max planck institut
informatik

CLUSTER OF EXCELLENCE

Structural Result

- **Peng et al. [COLT'15]**

  If $\Upsilon := \lambda_{k+1}/\rho(k) \geq \Omega(k^3)$ then
  - $\mu(A_i \Delta P_i) \leq (\gamma/\Upsilon) \cdot \mu(P_i)$
  - $\phi(A_i) \leq (1 + \gamma/\Upsilon) \cdot \phi(P_i) + \gamma/\Upsilon$

$$\rho(k) = \max_{i \in [1:k]} \phi(P_i)$$

- **Our Result**

  If $\Psi := \lambda_{k+1}/\rho_{\mathrm{avr}}(k) \geq \Omega(k^3)$ then
  - $\mu(A_i \Delta P_i) \leq (\gamma/\Psi \boldsymbol{k}) \cdot \mu(P_i)$
  - $\phi(A_i) \leq (1 + \gamma/\Psi \boldsymbol{k}) \cdot \phi(P_i) + \gamma/\Psi \boldsymbol{k}$

$$\rho_{\mathrm{avr}}(k) = \frac{1}{k} \sum_{i=1}^{k} \phi(P_i)$$

- $(P_1, \ldots, P_k)$ is an optimal k-way partition of G w.r.t. $\rho(k)$.
- $\mathrm{cost}(A_1, \ldots, A_k) \leq \gamma \cdot \Delta_k(\widetilde{\mathcal{X}}_E)$ for $\gamma \geq 1$.

Structural Result

- **Peng et al. [COLT'15]**

  If $\Upsilon := \lambda_{k+1}/\rho(k) \geq \Omega(k^3)$ then
  - $\mu(A_i \Delta P_i) \leq (\gamma/\Upsilon) \cdot \mu(P_i)$
  - $\phi(A_i) \leq (1 + \gamma/\Upsilon) \cdot \phi(P_i) + \gamma/\Upsilon$

$$\rho(k) = \max_{i \in [1:k]} \phi(P_i)$$

- **Our Result**

  If $\Psi := \lambda_{k+1}/\rho_{\mathrm{avr}}(k) \geq \Omega(k^3)$ then
  - $\mu(A_i \Delta P_i) \leq (\gamma/\Psi \boldsymbol{k}) \cdot \mu(P_i)$
  - $\phi(A_i) \leq (1 + \gamma/\Psi \boldsymbol{k}) \cdot \phi(P_i) + \gamma/\Psi \boldsymbol{k}$

$$\rho_{\mathrm{avr}}(k) = \frac{1}{k} \sum_{i=1}^{k} \phi(P_i)$$

How to find such k-way partition $(A_1, \ldots, A_k)$?

Outline

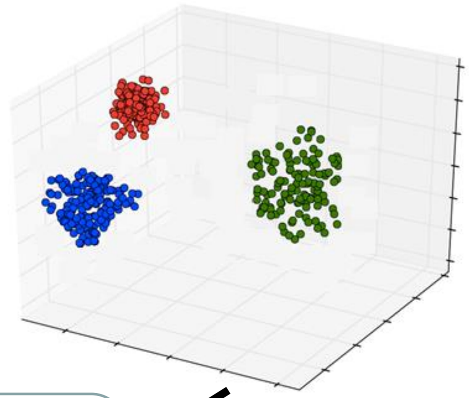
- Problem Formulation
  - Algorithmic Tools
- Our Contribution
  - Structural Result
  - **Algorithmic Result**
    - Proof Overview
- Summary

Algorithmic Result

- **Peng et al. [COLT'15]**

  $\Upsilon := \lambda_{k+1}/\rho(k) \geq \Omega(\boldsymbol{k^5})$

Concentration



more restrictive by

$\Omega(k^2)$-factor

Heat Kernel and

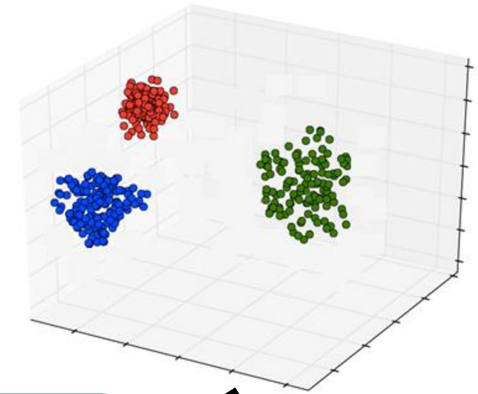
Local Sensitive Hashing

Algorithmic Result

- **Peng et al. [COLT'15]**

  $\Upsilon := \lambda_{k+1}/\rho(k) \geq \Omega(\boldsymbol{k^5})$

Concentration



more restrictive by

$\Omega(k^2)$-factor

Heat Kernel and

Local Sensitive Hashing

- **Our Result**

  $\Psi := \lambda_{k+1}/\rho_{\mathrm{avr}}(k) \geq \Omega(k^3)$

  and $\Delta_k(\mathcal{X}_V) \geq n^{-O(1)}$

Approx. Spectral Embedding

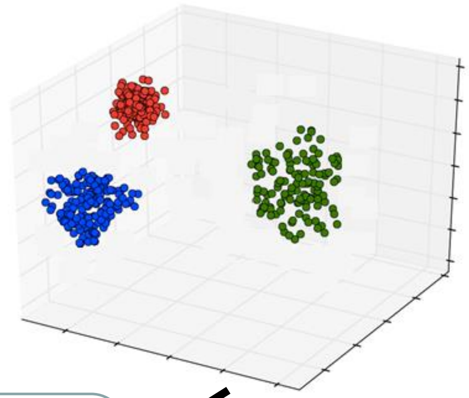
and k-means Clustering

Algorithmic Result

- **Peng et al. [COLT'15]**

  $\Upsilon := \lambda_{k+1}/\rho(k) \geq \Omega(\boldsymbol{k^5})$

Concentration →



more restrictive by $\Omega(k^2)$-factor

Heat Kernel and Local Sensitive Hashing

- **Our Result**

  $\Psi := \lambda_{k+1}/\rho_{\text{avr}}(k) \geq \Omega(k^3)$

  and $\Delta_k(\mathcal{X}_V) \geq n^{-O(1)}$

Approx. Spectral Embedding and k-means Clustering

This is the 1st rigorous ***algorithmic*** analysis of the Standard Spectral Clustering Paradigm!

max planck institut informatik

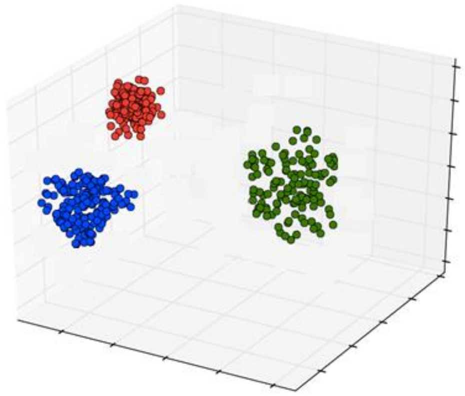
M2CI CLUSTER OF EXCELLENCE

Algorithmic Result

- **Peng et al. [COLT'15]**

  $\Upsilon := \lambda_{k+1}/\rho(k) \geq \Omega(\boldsymbol{k^5})$

  Concentration →



constant = 10^5

Heat Kernel and Local Sensitive Hashing

- **Our Result**

  $\Psi := \lambda_{k+1}/\rho_{\text{avr}}(k) \geq \Omega(k^3)$

  and $\Delta_k(\mathcal{X}_V) \geq n^{-O(1)}$

Approx. Spectral Embedding

and k-means Clustering

constant = $10^7/\epsilon_0$

$\epsilon_0 = 6/10^7$ is **Ostrovsky et al's [FOCS'13]**

k-means alg. constant (is not optimized!)

Algorithmic Result

- **Peng et al. [COLT'15]**

  If $\Upsilon := \lambda_{k+1}/\rho(k) \geq \Omega(\boldsymbol{k}^{\mathbf{5}})$ then
  - $\mu(A_i \Delta P_i) \leq \left(\frac{\log^2 k}{k^2}/\Upsilon\right) \cdot \mu(P_i)$
  - $\phi(A_i) \leq \left(1 + \frac{\log^2 k}{k^2}/\Upsilon\right) \cdot \phi(P_i) + \frac{\log^2 k}{k^2}/\Upsilon$

Heat Kernel and

Local Sensitive Hashing

- **Our Result**

  If $\Psi := \lambda_{k+1}/\rho_{\mathrm{avr}}(k) \geq \Omega(k^3)$

  and $\Delta_k(\mathcal{X}_V) \geq n^{-O(1)}$ then
  - $\mu(A_i \Delta P_i) \leq (1/\Psi \boldsymbol{k}) \cdot \mu(P_i)$
  - $\phi(A_i) \leq (1 + 1/\Psi \boldsymbol{k}) \cdot \phi(P_i) + (1/\Psi \boldsymbol{k})$

Approx. Spectral Embedding

and k-means Clustering

Algorithmic Result

- **Peng et al. [COLT'15]**

  If $\Upsilon \coloneqq \lambda_{k+1}/\rho(k) \geq \Omega(\boldsymbol{k^5})$ then

  - $\mu(A_i \Delta P_i) \leq \left(\frac{\log^2 k}{k^2}/\Upsilon\right) \cdot \mu(P_i)$
  - $\phi(A_i) \leq \left(1 + \frac{\log^2 k}{k^2}/\Upsilon\right) \cdot \phi(P_i) + \frac{\log^2 k}{k^2}/\Upsilon$

Heat Kernel and

Local Sensitive Hashing

Runtime: $O(m\log^c n)$

- **Our Result**

  If $\Psi \coloneqq \lambda_{k+1}/\rho_{\mathrm{avr}}(k) \geq \Omega(k^3)$

  and $\Delta_k(\mathcal{X}_V) \geq n^{-O(1)}$ then

  - $\mu(A_i \Delta P_i) \leq (1/\Psi\boldsymbol{k}) \cdot \mu(P_i)$
  - $\phi(A_i) \leq (1 + 1/\Psi\boldsymbol{k}) \cdot \phi(P_i) + (1/\Psi\boldsymbol{k})$

Approx. Spectral Embedding

and k-means Clustering

Runtime: $\mathrm{O}\left(m\left(k^2 + \frac{\ln n}{\lambda_{k+1}}\right)\right)$

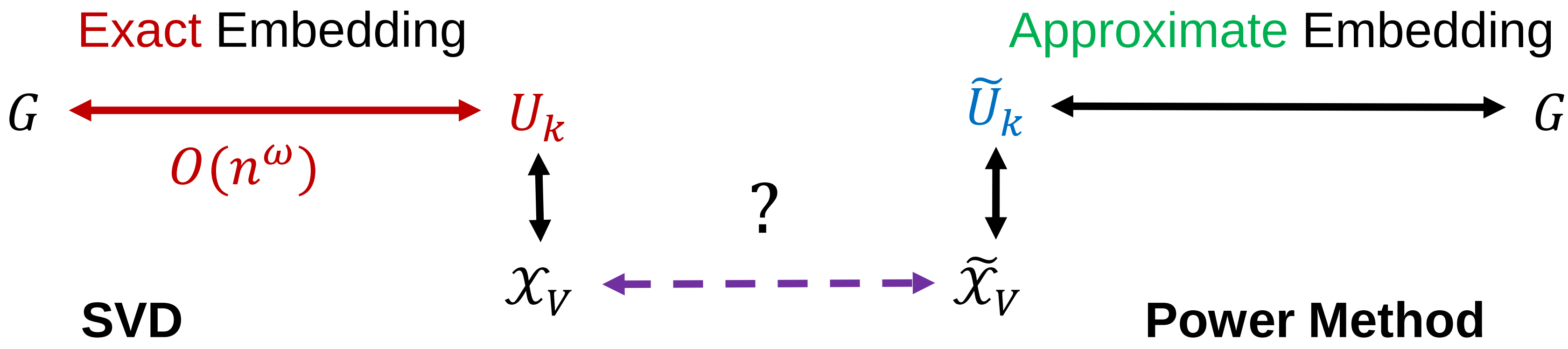
max planck institut informatik

CLUSTER OF EXCELLENCE

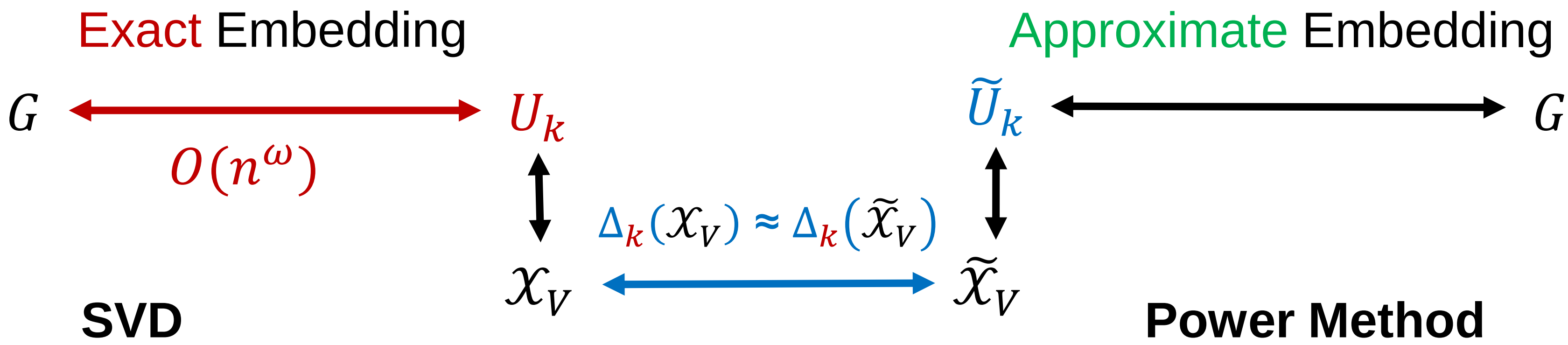
Outline

- Problem Formulation
  - Algorithmic Tools
- Our Contribution
  - Structural Result
  - Algorithmic Result
    - **Proof Overview**
- Summary

Proof Overview



Proof Overview



Boutsidos et al [ICML'15] Let $(A_1, \ldots, A_k)$ be a partition such that

$$\mathrm{cost}(A_1, \ldots, A_k) \leq (1+\gamma)\Delta_k(\tilde{\mathcal{X}}_V)$$

then

$$\mathrm{cost}(A_1, \ldots, A_k) \leq (1+4\epsilon)(1+\gamma)\Delta_k(\mathcal{X}_V) + 4\epsilon^2.$$

max planck institut informatik

CLUSTER OF EXCELLENCE

Proof Overview

Exact Embedding

$G \longleftrightarrow U_k$: $O(n^\omega)$

SVD

$U_k \updownarrow \mathcal{X}_V$

$\Delta_k(\mathcal{X}_V) \approx \Delta_k(\widetilde{\mathcal{X}}_V)$

$\mathcal{X}_V \longleftrightarrow \widetilde{\mathcal{X}}_V$

No fast P.M.
No fast k-means

Approximate Embedding

$\widetilde{U}_k \longleftrightarrow G$: $O(mkp + nk^2)$

$\widetilde{U}_k \updownarrow \widetilde{\mathcal{X}}_V$

Power Method

$$p = f\left(n, \log\frac{1}{\epsilon}, \lambda_k, \lambda_{k+1}\right)$$

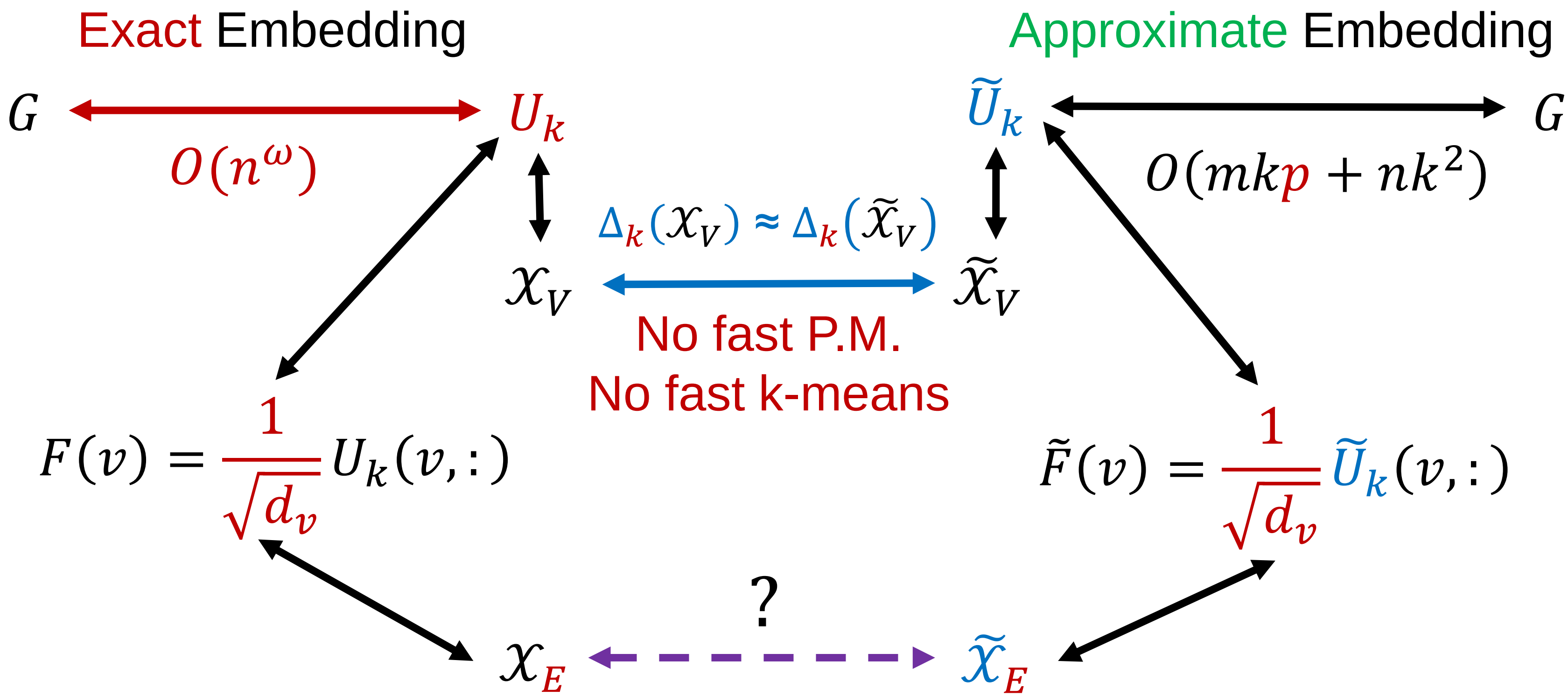
Boutsidos et al [ICML'15] Let $(A_1, \ldots, A_k)$ be a partition such that

$$\text{cost}(A_1, \ldots, A_k) \le (1+\gamma)\Delta_k(\widetilde{\mathcal{X}}_V)$$

then

$$\text{cost}(A_1, \ldots, A_k) \le (1+4\epsilon)(1+\gamma)\Delta_k(\mathcal{X}_V) + 4\epsilon^2.$$

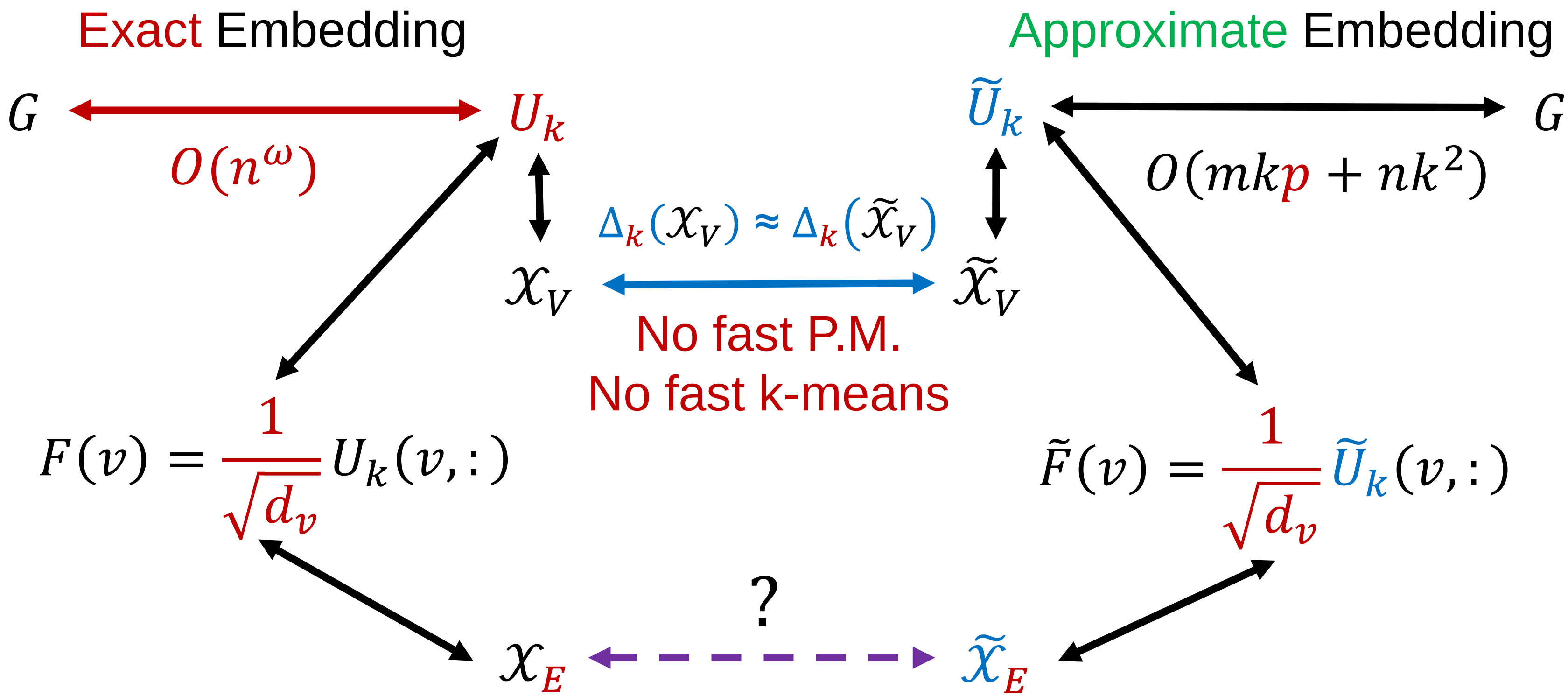
Proof Overview



max planck institut informatik

CLUSTER OF EXCELLENCE

Proof Overview

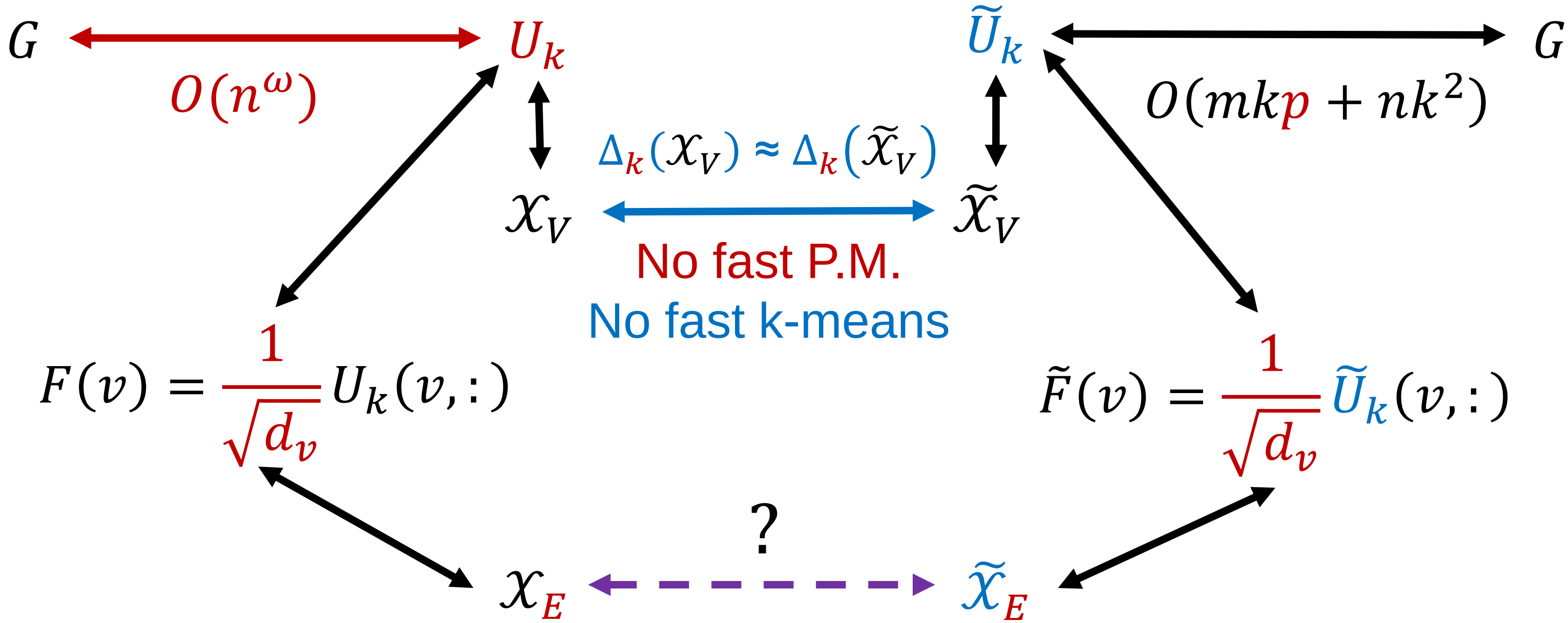


Questions:

1. Find an efficient k-means clustering algorithm for $\widetilde{\mathcal{X}}_E$?
2. Extend **Boutsidos et al's [ICML'15]** analysis?

Proof Overview

Exact Embedding Approximate Embedding



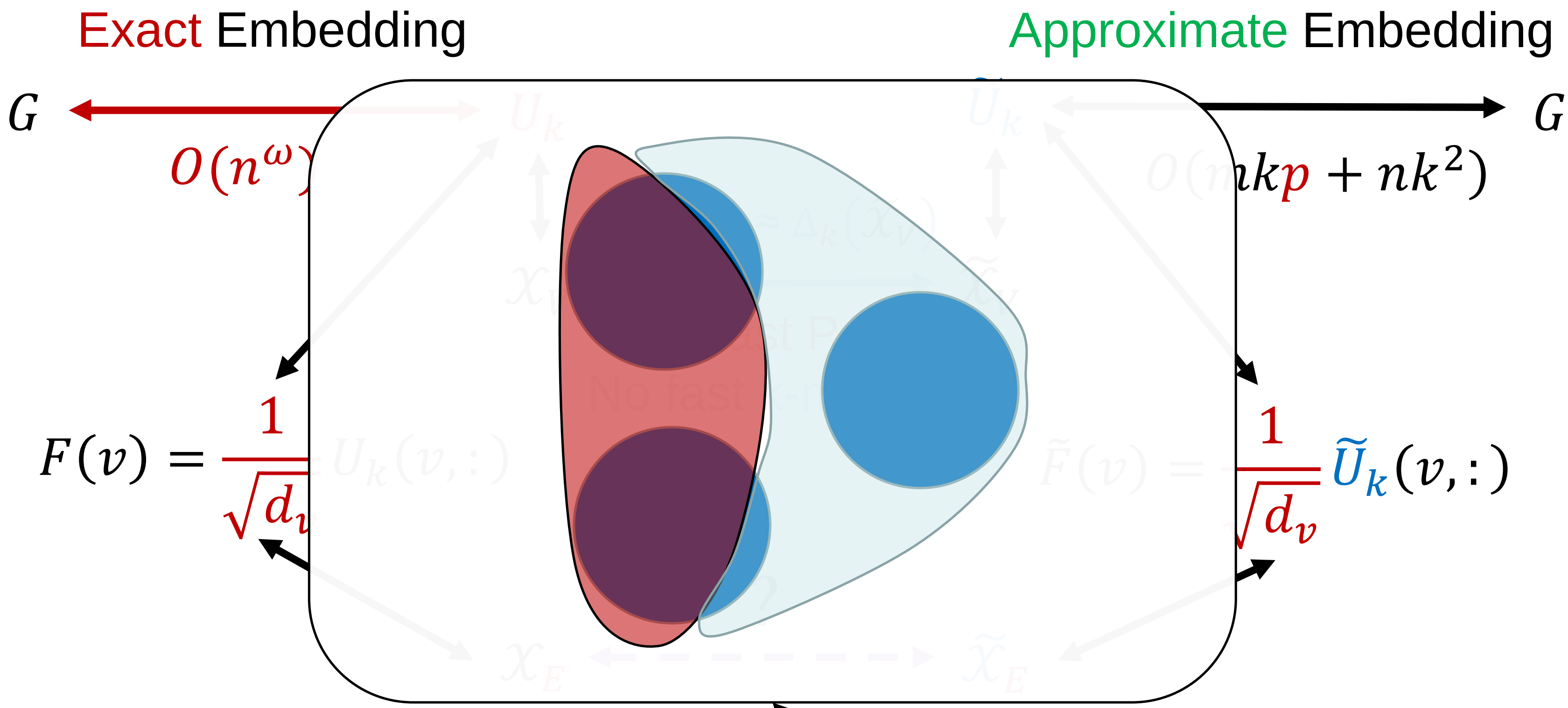
Ostrovsky et al's [FOCS'13] gave an *approximate* k-means algorithm with fast runtime $O(mk^2)$, but requires $\Delta_k(\mathcal{X}) \leq \epsilon_0^2 \cdot \Delta_{k-1}(\mathcal{X})$

where $\epsilon_0 = 6/10^7$.

max planck institut informatik

CLUSTER OF EXCELLENCE

Proof Overview

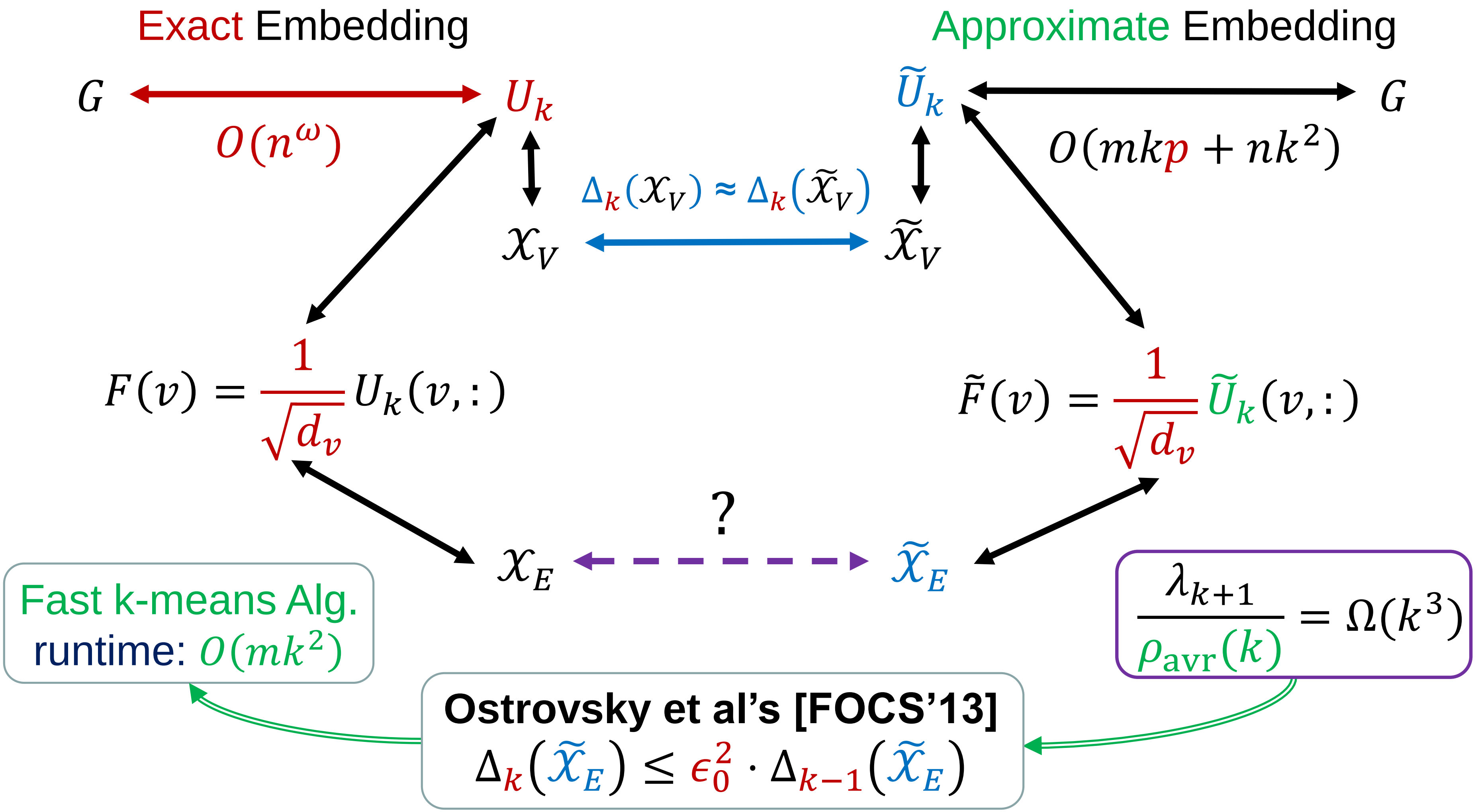


Ostrovsky et al's [FOCS'13] gave an *approximate* k-means algorithm with fast runtime $O(mk^2)$, but requires $\Delta_k(\mathcal{X}) \leq \epsilon_0^2 \cdot \Delta_{k-1}(\mathcal{X})$ where $\epsilon_0 = 6/10^7$.

max planck institut informatik

CLUSTER OF EXCELLENCE

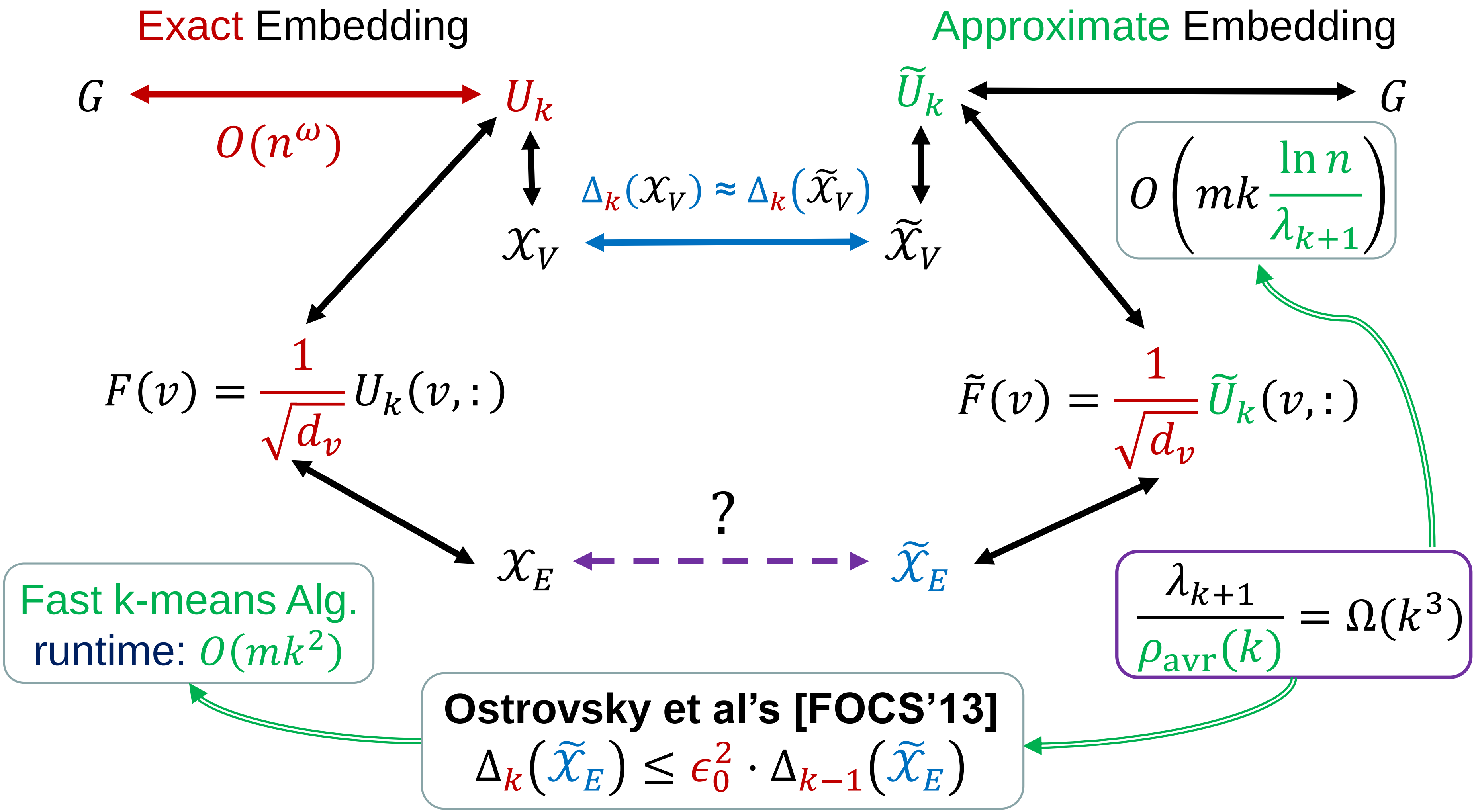
Proof Overview



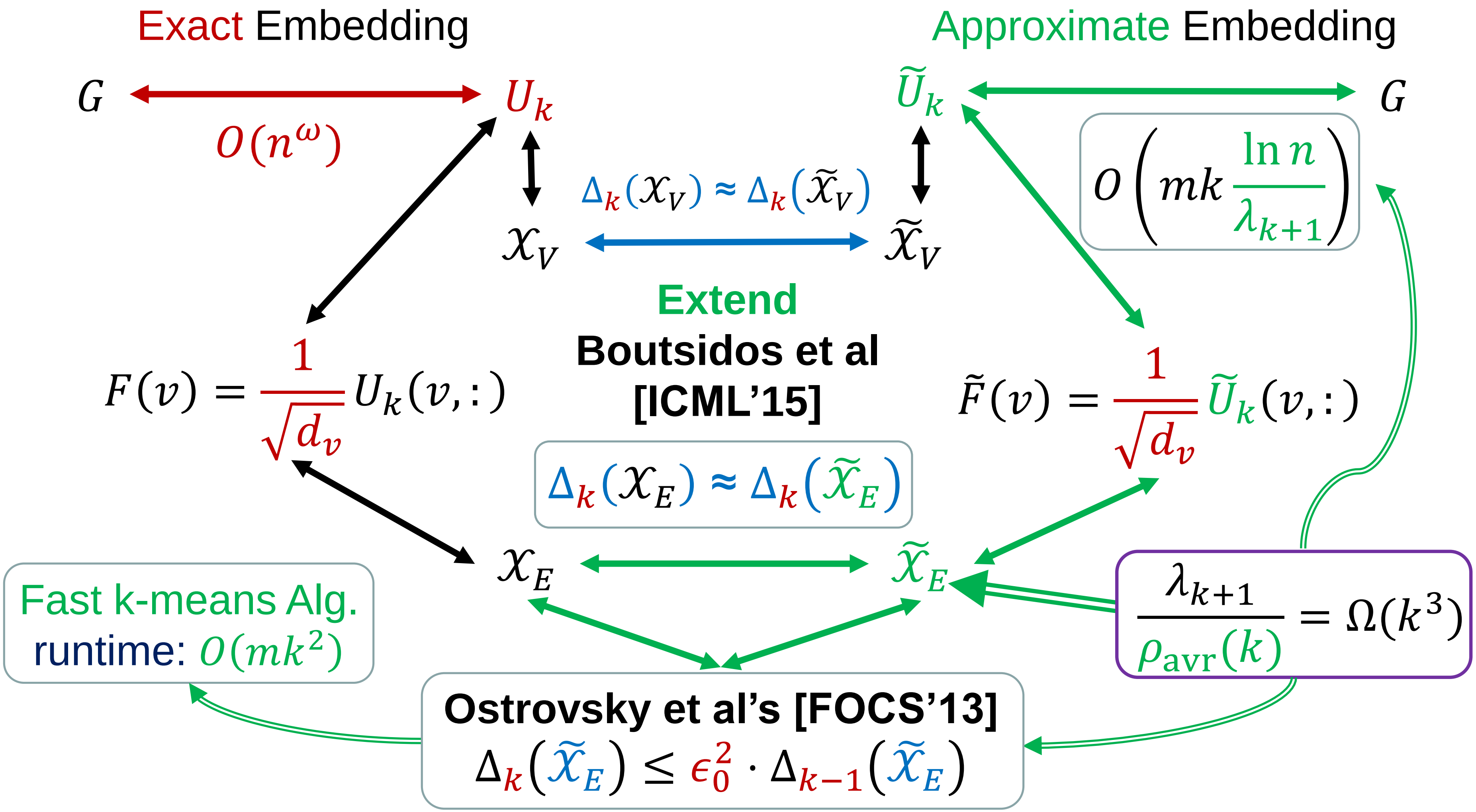
max planck institut
informatik

CLUSTER OF EXCELLENCE

Proof Overview



Proof Sketch (Overview)



max planck institut
informatik

CLUSTER OF EXCELLENCE

Outline

- Problem Formulation
  - Algorithmic Tools
- Our Contribution
  - Structural Result
  - Algorithmic Result
    - Proof Overview
- **Summary**

max planck institut informatik

CLUSTER OF EXCELLENCE

Summary

- *We proved rigorously that*

  the Standard Spectral Clustering Paradigm

  efficiently computes a k-way partition

  under asymptotically less restrictive gap assumption.

Open Problems

- *Show that the **SSCP** has a **good** behavior on **small** graphs.*

  Our approach **fails** due to **large** constants in $\Psi \geq \Omega(k^3)$:

  - $10^7/\epsilon_0$ - **Ostrovsky et al.** (is not optimized)

$$\Delta_k(\mathcal{X}) \leq \epsilon_0^2 \cdot \Delta_{k-1}(\mathcal{X}), \text{ where } \epsilon_0 = 6/10^7.$$

Open Problems

- *Show that the **SSCP** has a **good** behavior on **small** graphs.*

  Our approach **fails** due to **large** constants in $\Psi \geq \Omega(k^3)$:

  – $10^7/\epsilon_0$ - **Ostrovsky et al.** (is not optimized)

  $$\Delta_k(\mathcal{X}) \leq \epsilon_0^2 \cdot \Delta_{k-1}(\mathcal{X}), \text{ where } \epsilon_0 = 6/10^7.$$

- Can we obtain a **multiplicative** **conductance** guarantee:

  $$\phi(A_i) \leq (1 + \gamma/\Psi \boldsymbol{k}) \cdot \phi(P_i) + \gamma/\Psi \boldsymbol{k}.$$

Thank you!

max planck institut
informatik

CLUSTER OF EXCELLENCE