Handling LP-Rounding for Hierarchical Clustering and Fitting Distances by Ultrametrics

Changyeol Lee

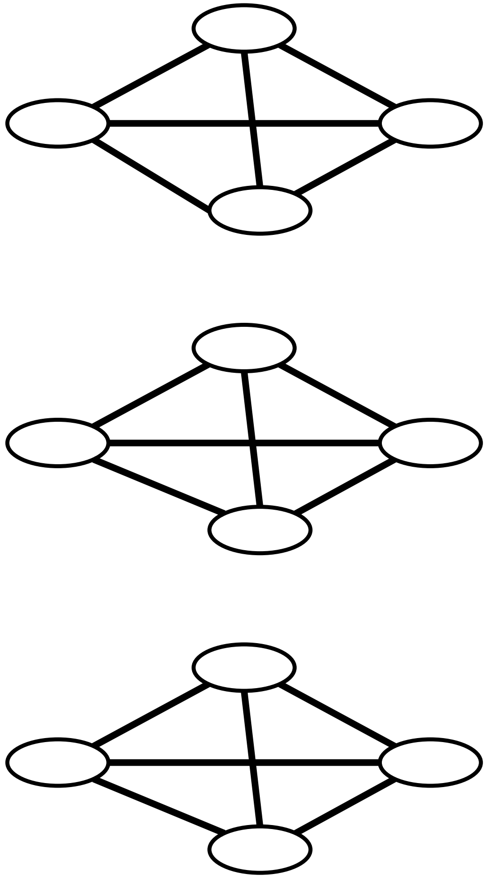
Yonsei University, South Korea

Joint work with

Hyung-Chan An Yonsei University, South Korea
Mong-Jen Kao National Yang-Ming Chiao-Tung University, Taiwan
Mu-Ting Lee National Yang-Ming Chiao-Tung University, Taiwan

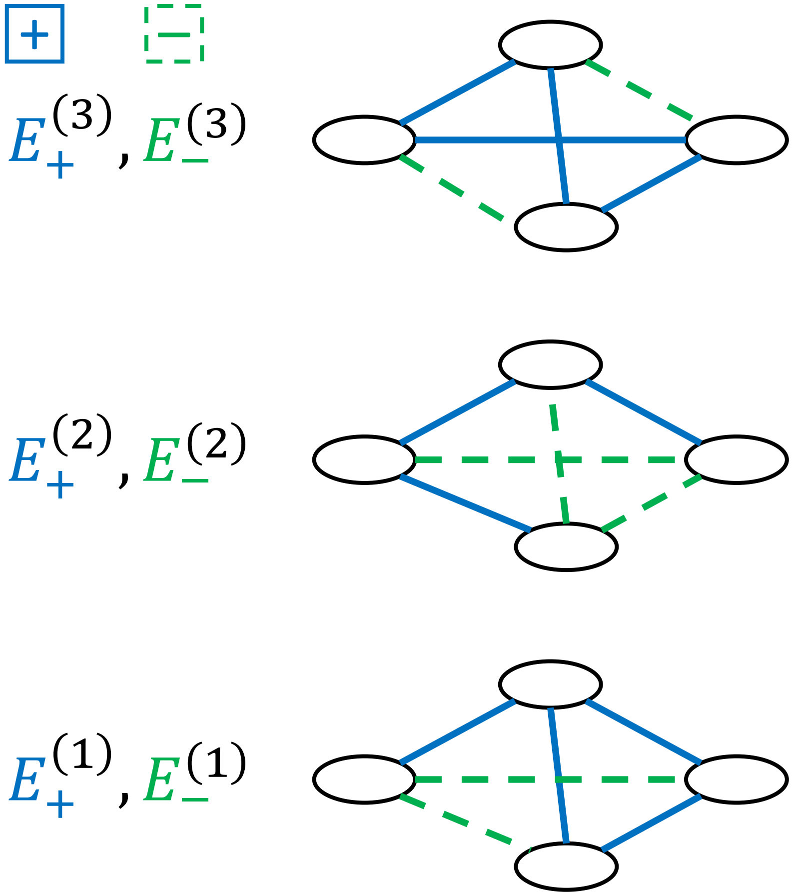
Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs (on the same vertex set)



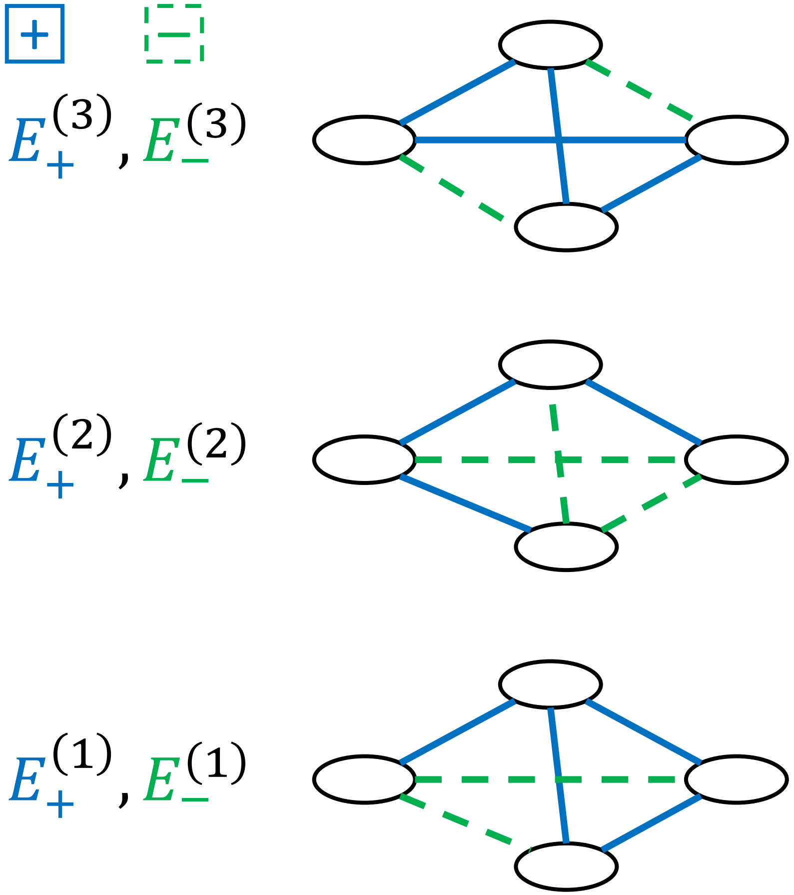
Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs (on the same vertex set)
  - each edge is labeled $\boxed{+}$ or $\boxed{-}$



Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs
- ℓ weights $\delta^{(1)},\ldots,\delta^{(\ell)} \geq 0$



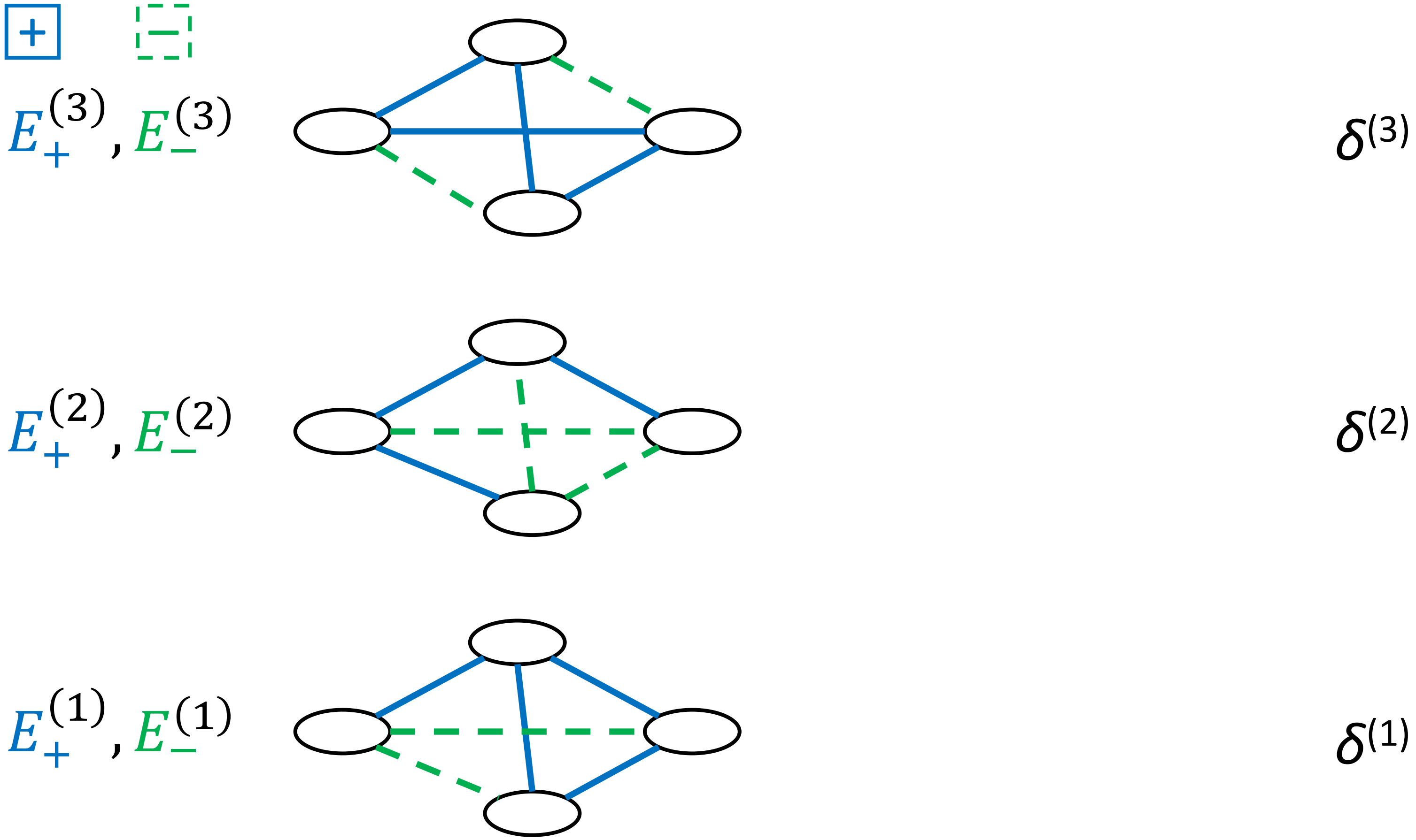
$E_+^{(3)}, E_-^{(3)}$ $\delta^{(3)}$

$E_+^{(2)}, E_-^{(2)}$ $\delta^{(2)}$

$E_+^{(1)}, E_-^{(1)}$ $\delta^{(1)}$

Hierarchical Correlation Clustering (HCC)

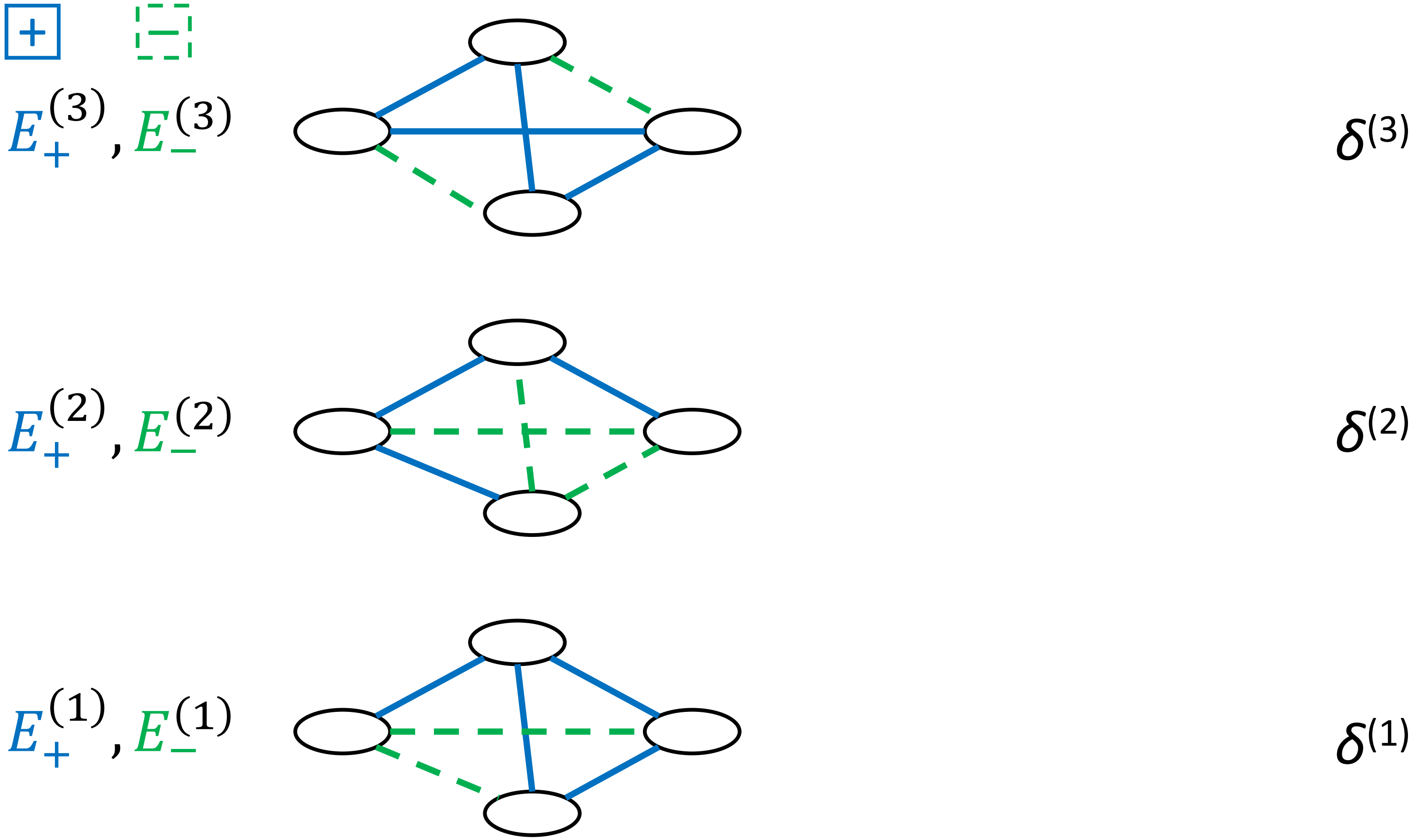
- ℓ *layers* of complete graphs
- find clusterings (partitions) $\mathcal{P}^{(1)}, \dots, \mathcal{P}^{(\ell)}$
- ℓ weights $\delta^{(1)}, \dots, \delta^{(\ell)} \geq 0$



Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs
- ℓ weights $\delta^{(1)},\ldots,\delta^{(\ell)} \geq 0$
- find clusterings (partitions) $\mathcal{P}^{(1)},\ldots,\mathcal{P}^{(\ell)}$

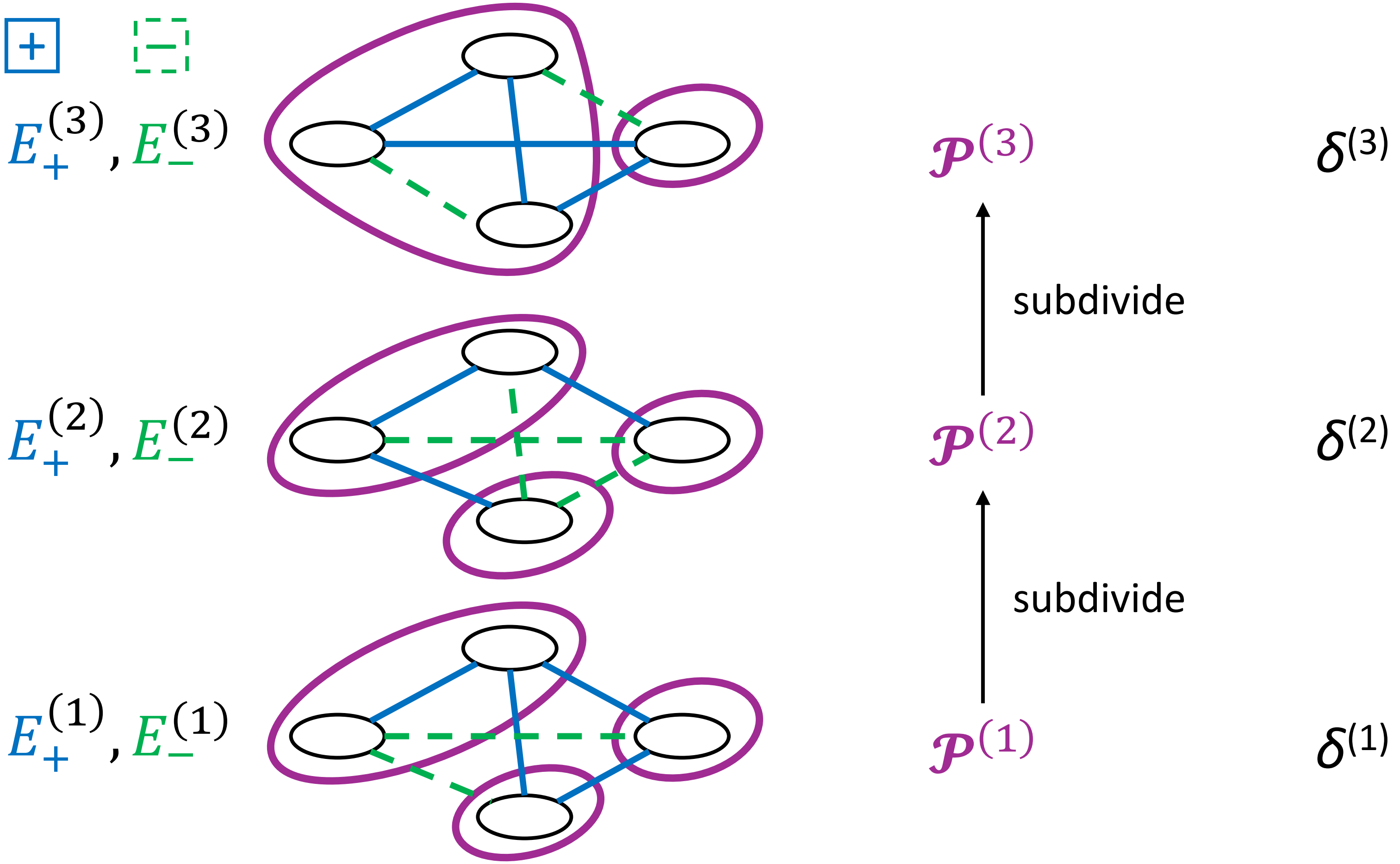
  s.t. $\mathcal{P}^{(t)}$ subdivides $\mathcal{P}^{(t+1)}$ for all $t < \ell$



Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs
- ℓ weights $\delta^{(1)},\ldots,\delta^{(\ell)} \geq 0$
- find clusterings (partitions) $\mathcal{P}^{(1)},\ldots,\mathcal{P}^{(\ell)}$

  s.t. $\mathcal{P}^{(t)}$ subdivides $\mathcal{P}^{(t+1)}$ for all $t < \ell$



$+$ $-$

$E_{+}^{(3)}, E_{-}^{(3)}$ $\mathcal{P}^{(3)}$ $\delta^{(3)}$

subdivide

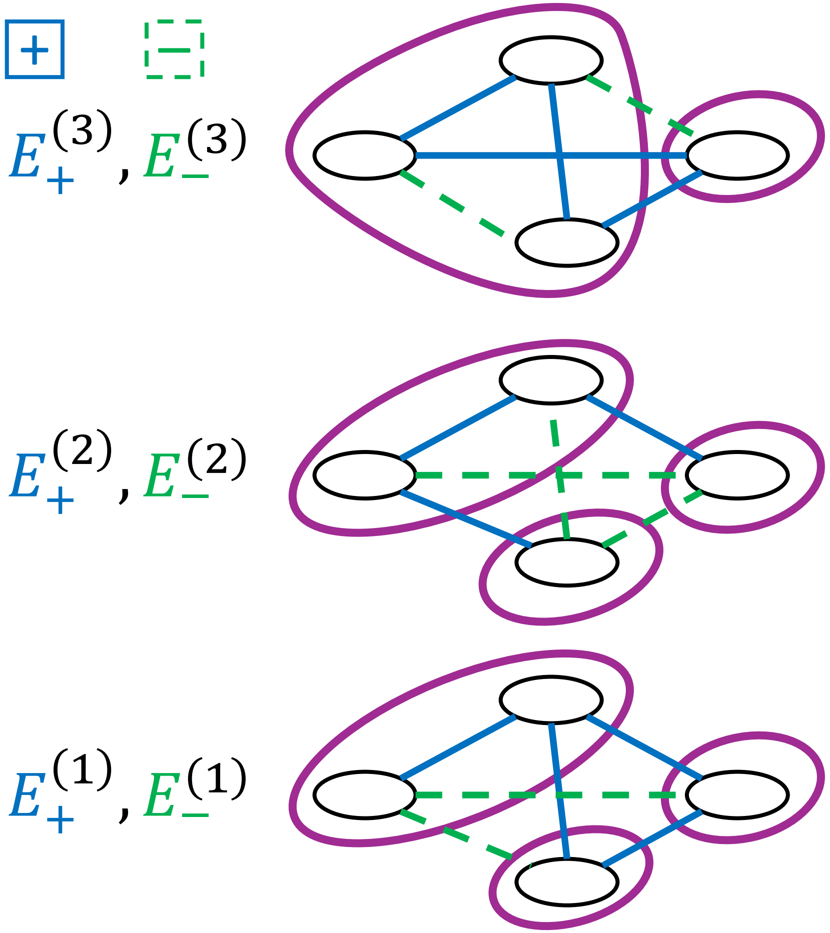
$E_{+}^{(2)}, E_{-}^{(2)}$ $\mathcal{P}^{(2)}$ $\delta^{(2)}$

subdivide

$E_{+}^{(1)}, E_{-}^{(1)}$ $\mathcal{P}^{(1)}$ $\delta^{(1)}$

Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs
- ℓ weights $\delta^{(1)}, \ldots, \delta^{(\ell)} \geq 0$
- find ***hierarchical*** clusterings $\mathcal{P}^{(1)}, \ldots, \mathcal{P}^{(\ell)}$



$\mathcal{P}^{(3)}$ $\delta^{(3)}$

↑ subdivide

$\mathcal{P}^{(2)}$ $\delta^{(2)}$

↑ subdivide

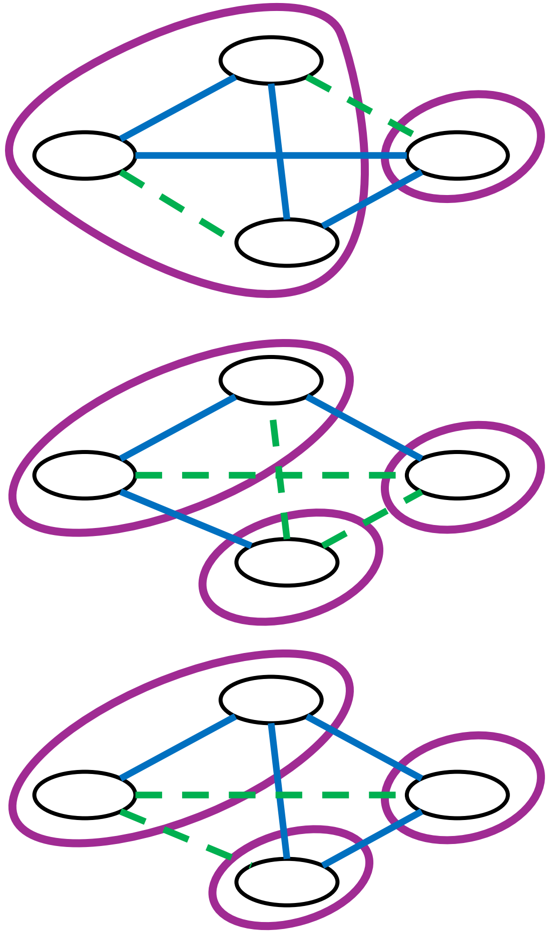
$\mathcal{P}^{(1)}$ $\delta^{(1)}$

Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs
- ℓ weights $\delta^{(1)}, \ldots, \delta^{(\ell)} \geq 0$

- find ***hierarchical*** clusterings $\mathcal{P}^{(1)}, \ldots, \mathcal{P}^{(\ell)}$

minimizing $\Sigma_t \, \delta^{(t)} \cdot$(#disagreements at layer *t*)



$\boxed{+}$ $\boxed{-}$

$E_+^{(3)}, E_-^{(3)}$

$\mathcal{P}^{(3)}$ $\delta^{(3)}$

subdivide

$E_+^{(2)}, E_-^{(2)}$

$\mathcal{P}^{(2)}$ $\delta^{(2)}$

subdivide

$E_+^{(1)}, E_-^{(1)}$

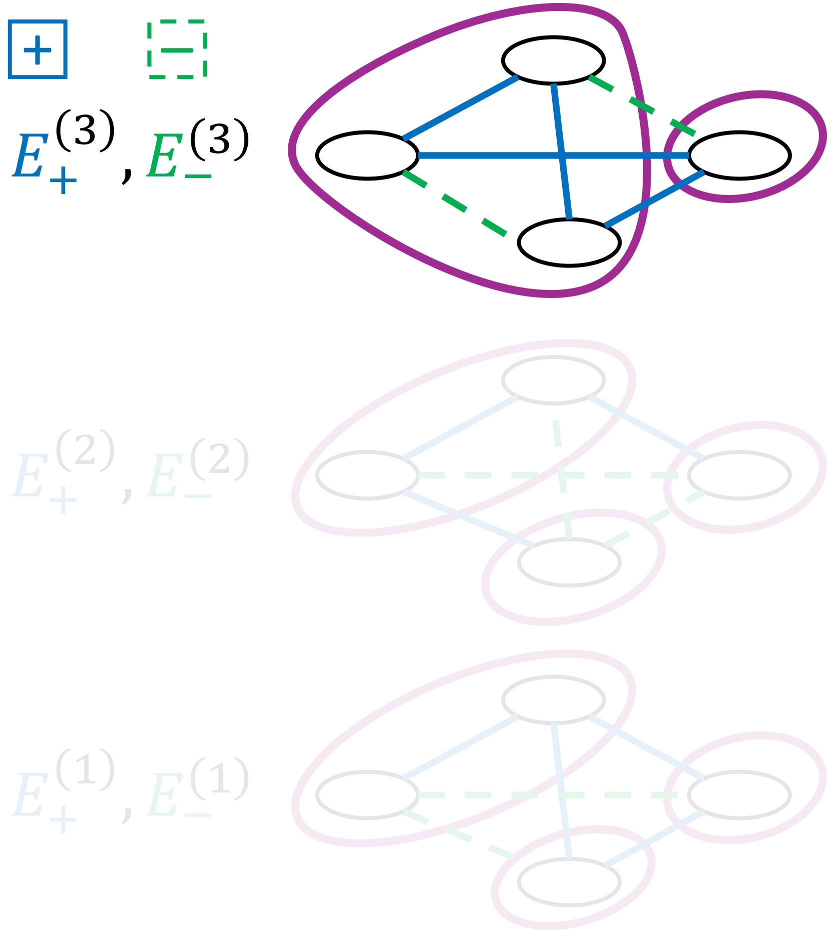
$\mathcal{P}^{(1)}$ $\delta^{(1)}$

Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs
- ℓ weights $\delta^{(1)}, \dots, \delta^{(\ell)} \geq 0$

- find ***hierarchical*** clusterings $\mathcal{P}^{(1)}, \dots, \mathcal{P}^{(\ell)}$

minimizing $\Sigma_t \, \delta^{(t)} \cdot (\#\text{disagreements at layer } t)$



$\delta^{(3)}$

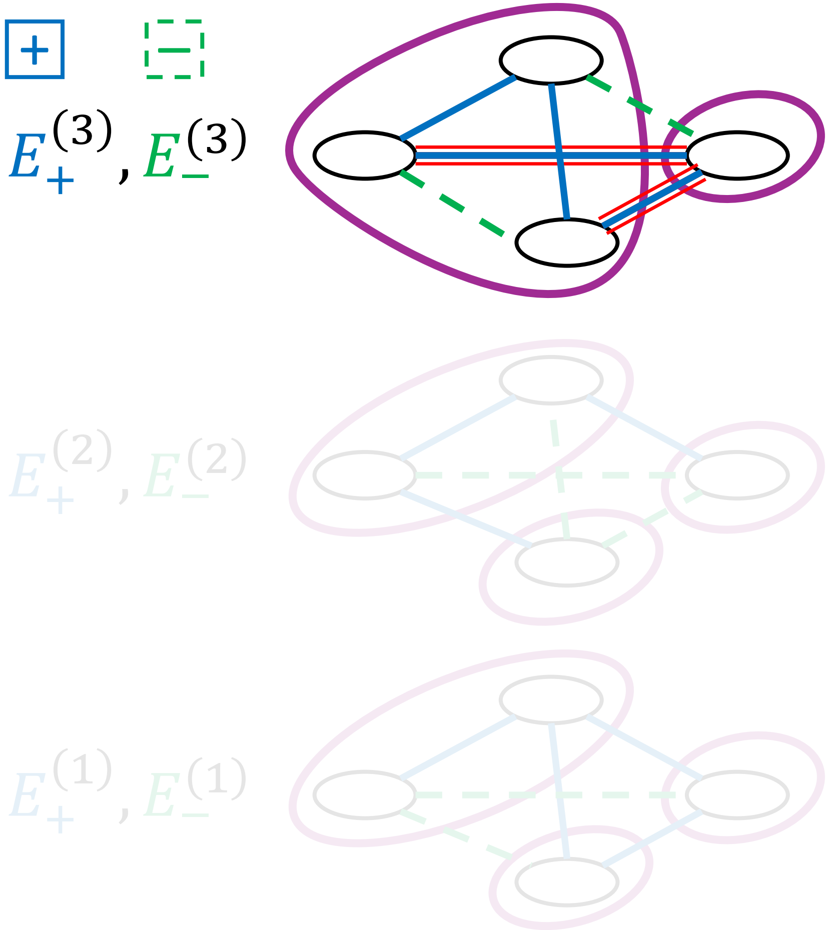
Definition. An edge is a *disagreement* if

- ⊞ edge and endpoints are separated
- ⊟ edge and endpoints are clustered

$\delta^{(1)}$

Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs
- ℓ weights $\delta^{(1)}, \dots, \delta^{(\ell)} \geq 0$



- find ***hierarchical*** clusterings $\mathcal{P}^{(1)}, \dots, \mathcal{P}^{(\ell)}$

minimizing $\Sigma_t \, \delta^{(t)} \cdot$(#disagreements at layer t)

$\delta^{(3)}$

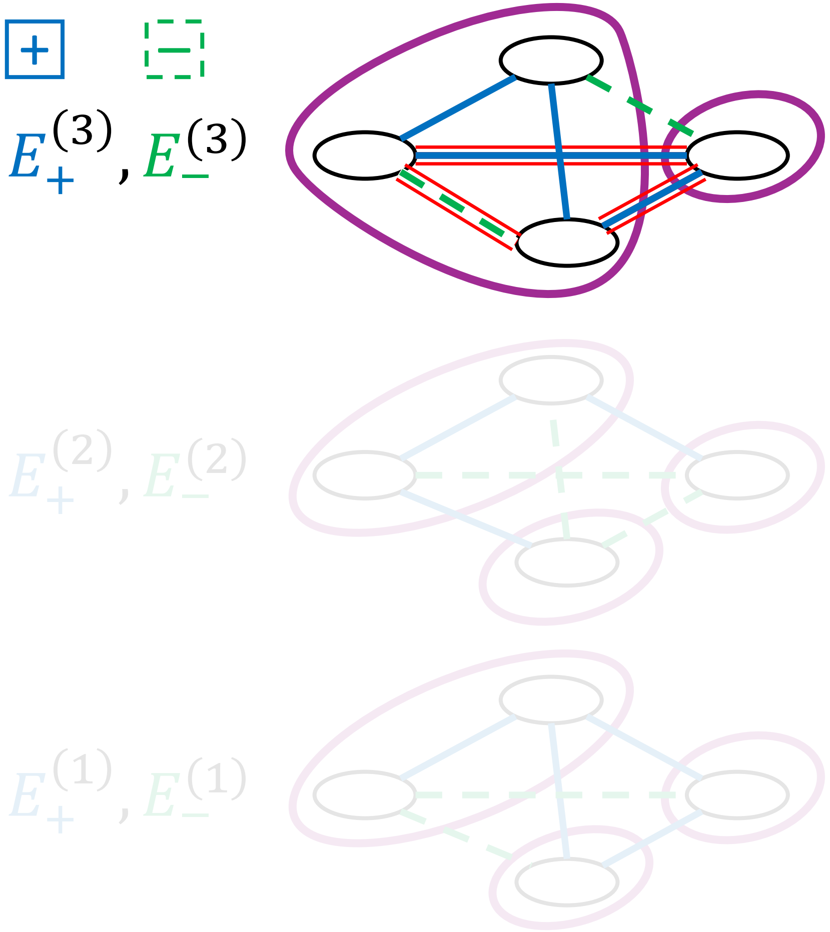
Definition. An edge is a *disagreement* if

- $\boxed{+}$ edge and endpoints are separated
- $\boxed{-}$ edge and endpoints are clustered

$\delta^{(1)}$

Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs
- ℓ weights $\delta^{(1)},\ldots,\delta^{(\ell)} \geq 0$



- find ***hierarchical*** clusterings $\mathcal{P}^{(1)}, \ldots, \mathcal{P}^{(\ell)}$

minimizing $\Sigma_t\, \delta^{(t)} \cdot$(#disagreements at layer t)

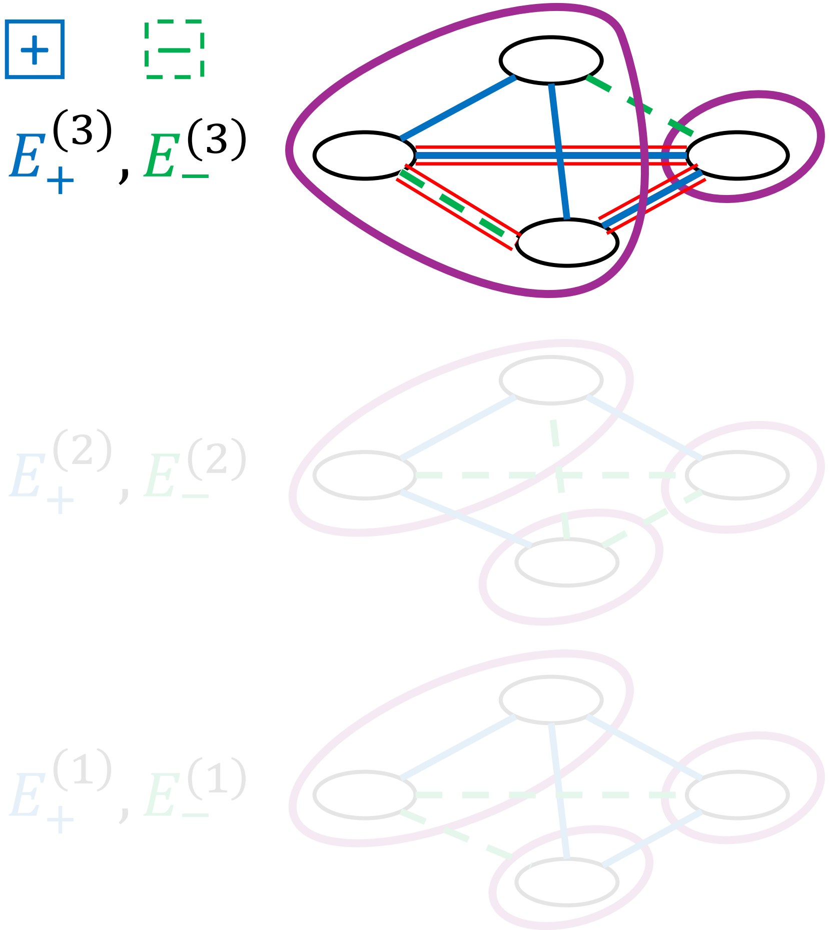
Definition. An edge is a *disagreement* if
- $\boxed{+}$ edge and endpoints are separated
- $\boxed{-}$ edge and endpoints are clustered

Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs
- ℓ weights $\delta^{(1)},\ldots,\delta^{(\ell)} \geq 0$

- find ***hierarchical*** clusterings $\mathcal{P}^{(1)}, \ldots, \mathcal{P}^{(\ell)}$

minimizing $\Sigma_t\, \delta^{(t)} \cdot$(#disagreements at layer t)



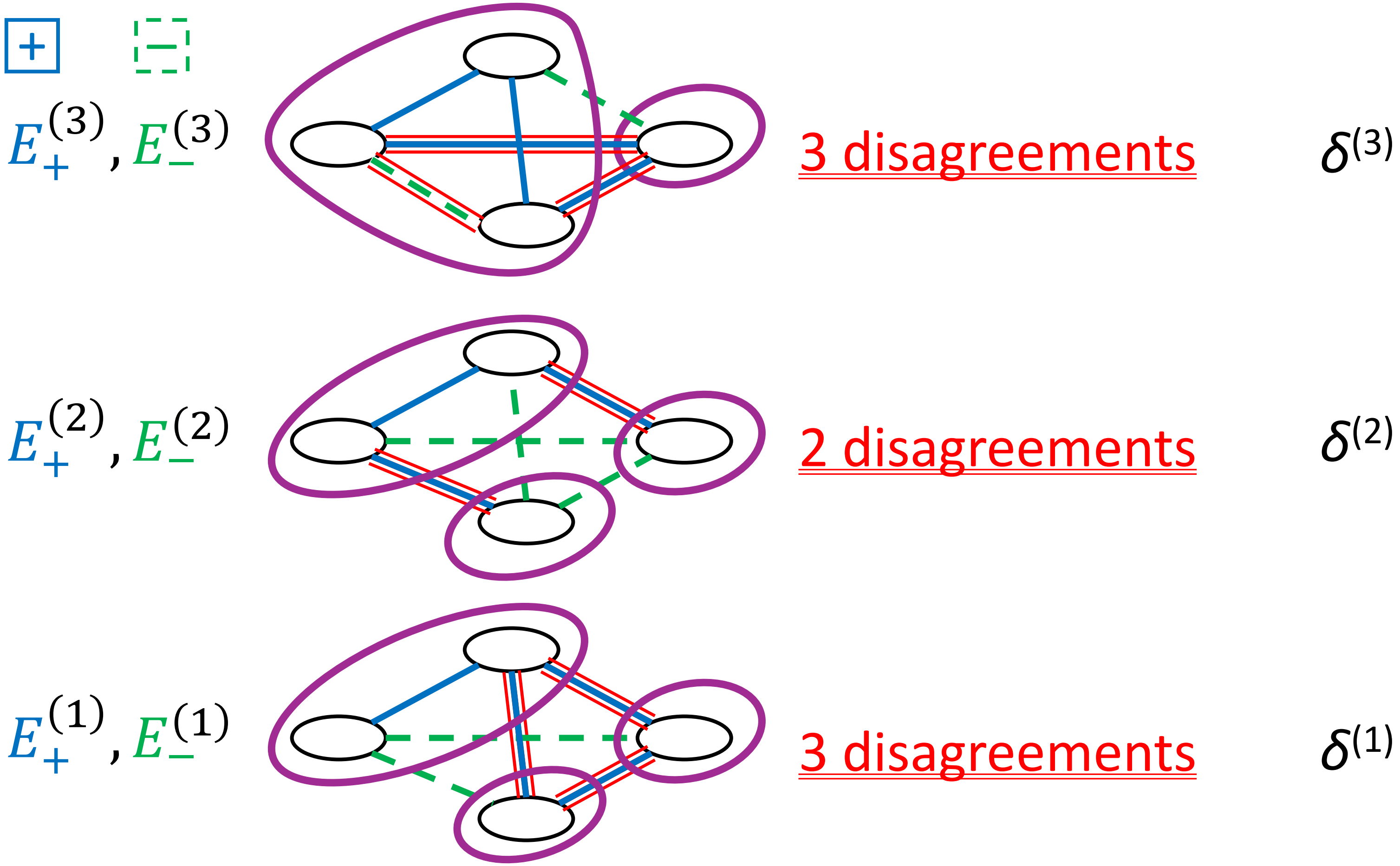
3 disagreements (at layer 3)

Definition. An edge is a *disagreement* if
- [+] edge and endpoints are separated
- [−] edge and endpoints are clustered

Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs
- ℓ weights $\delta^{(1)},\ldots,\delta^{(\ell)} \geq 0$
- find ***hierarchical*** clusterings $\mathcal{P}^{(1)},\ldots,\mathcal{P}^{(\ell)}$

minimizing $\Sigma_t\,\delta^{(t)}\cdot$(#disagreements at layer t)

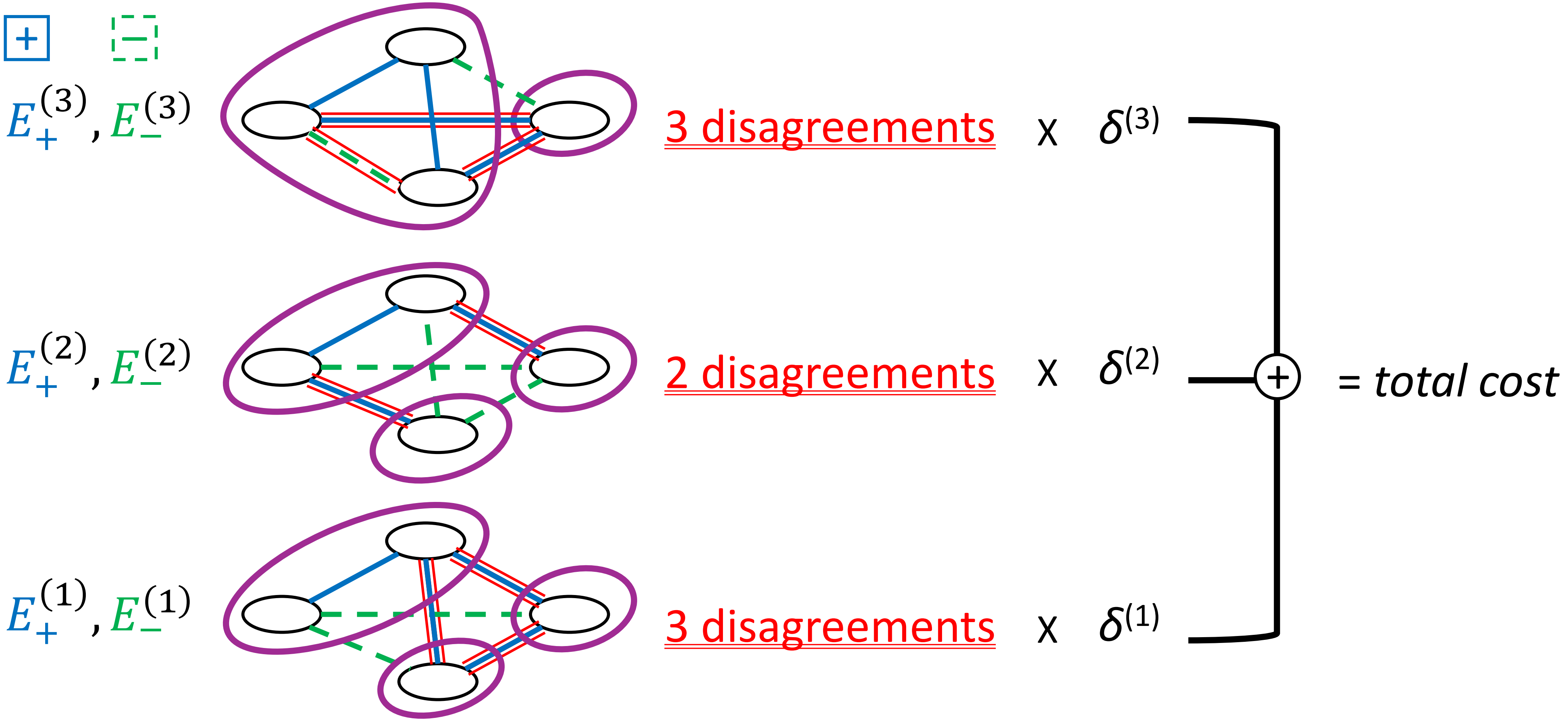


Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs
- ℓ weights $\delta^{(1)},\ldots,\delta^{(\ell)} \geq 0$

- find ***hierarchical*** clusterings $\mathcal{P}^{(1)},\ldots,\mathcal{P}^{(\ell)}$

minimizing $\Sigma_t\, \delta^{(t)} \cdot$(#disagreements at layer t)



Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs
- ℓ weights $\delta^{(1)}$,..., $\delta^{(\ell)} \geq 0$

- find ***hierarchical*** clusterings $\boldsymbol{\mathcal{P}}^{(1)}, \ldots, \boldsymbol{\mathcal{P}}^{(\ell)}$

minimizing $\Sigma_t \, \delta^{(t)} \cdot$(#disagreements at layer t)

- HCC generalizes Correlation Clustering

Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs
- ℓ weights $\delta^{(1)},\ldots,\delta^{(\ell)} \geq 0$

- find ***hierarchical*** clusterings $\boldsymbol{\mathcal{P}}^{(1)}, \ldots, \boldsymbol{\mathcal{P}}^{(\ell)}$

minimizing $\Sigma_t \delta^{(t)} \cdot$(#disagreements at layer t)

- HCC generalizes Correlation Clustering
- **HCC generalizes L_1 *Ultrametric Fitting*!** [Harb, Kannan, and McGregor, 2005]

Hierarchical Correlation Clustering (HCC)

- ℓ *layers* of complete graphs
- ℓ weights $\delta^{(1)}, \ldots, \delta^{(\ell)} \geq 0$

- find ***hierarchical*** clusterings $\mathcal{P}^{(1)}, \ldots, \mathcal{P}^{(\ell)}$

minimizing $\Sigma_t \, \delta^{(t)} \cdot$(#disagreements at layer t)

- HCC generalizes Correlation Clustering
- **HCC generalizes L_1 *Ultrametric Fitting*!** [Harb, Kannan, and McGregor, 2005]
  - Harb, Kannan, and McGregor (2005)
  - Ailon and Charikar (2005)
  - Cohen-Addad, Das, Kipouridis, Parotsidis, and Thorup (2021)

Ultrametric Fitting

- **Ultrametric Fitting**

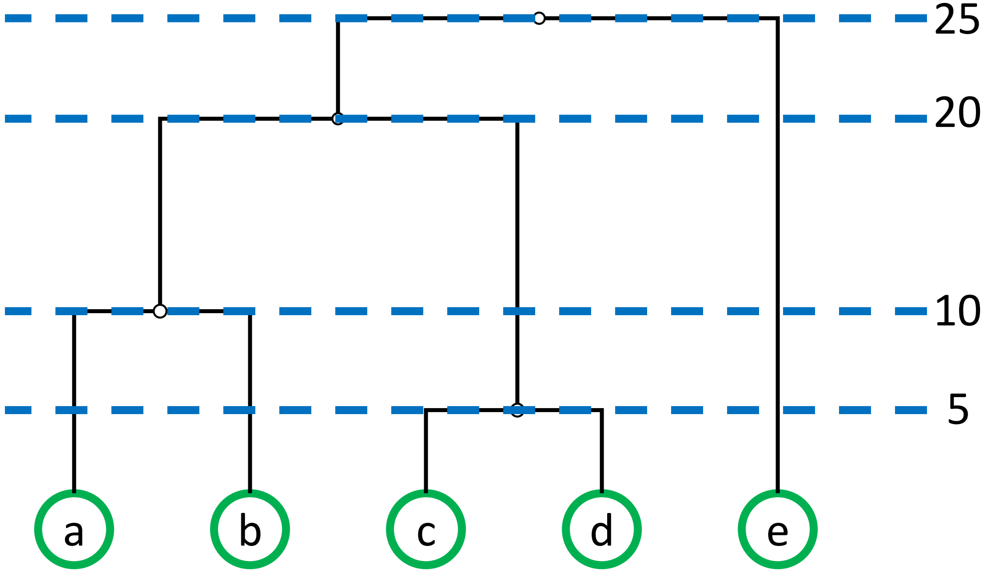
  Given a distance function *D*, find an ultrametric *T* that "*best fits*" *D*

Ultrametric Fitting

- **Ultrametric Fitting**

Given a distance function *D*, find an ultrametric *T* that "*best fits*" *D*

all root–leaf paths have equal length



example of an unltrametric

Ultrametric Fitting

- **Ultrametric Fitting**

  Given a distance function *D*, find an ultrametric *T* that “*best fits*” *D*

- Extensively studied since 1960s [Sneath and Sokal, 1962], [Cavalli-Sforza and Edwards, 1967], [Farris, 1972], [Agarwala, Bafna, Farach, Paterson, and Thorup, 1996]
  - Numerical taxonomy, phylogeny reconstruction

Ultrametric Fitting

- **Ultrametric Fitting**

  Given a distance function *D*, find an ultrametric *T* that "*best fits*" *D*

- Extensively studied since 1960s [Sneath and Sokal, 1962], [Cavalli-Sforza and Edwards, 1967], [Farris, 1972], [Agarwala, Bafna, Farach, Paterson, and Thorup, 1996]
  - Numerical taxonomy, phylogeny reconstruction

- **L_1 Ultrametric Fitting:** best fit ≡ minimize $\|D - T\|_1$ —— $\sum_{ij} |D(i,j) - T(i,j)|$
- **L_0 Ultrametric Fitting:** best fit ≡ minimize $\|D - T\|_0$ —— $\sum_{ij} \mathbb{1}[D(i,j) \neq T(i,j)]$

Ultrametric Fitting

- **Ultrametric Fitting**

  Given a distance function *D*, find an ultrametric *T* that "*best fits*" *D*

- Extensively studied since 1960s [Sneath and Sokal, 1962], [Cavalli-Sforza and Edwards, 1967], [Farris, 1972], [Agarwala, Bafna, Farach, Paterson, and Thorup, 1996]
  - Numerical taxonomy, phylogeny reconstruction

(special case of HCC)

- ***L_1* Ultrametric Fitting:** best fit ≡ minimize $\|D - T\|_1$ —— $\sum_{ij}|D(i,j) - T(i,j)|$
- ***L_0* Ultrametric Fitting:** best fit ≡ minimize $\|D - T\|_0$ —— $\sum_{ij} \mathbb{1}[D(i,j) \neq T(i,j)]$

Previous results on L_1 Ultrametric Fitting/HCC

k ≔ #distinct distances in the input distance function D

- APX-hard
- min{n, $O(k \log n)$}-approximation for L_1 Ultrametric Fitting
  - Harb, Kannan, and McGregor (APPROX 2005)
- min{k+2, $O(\log n \log\log n)$}-approximation for L_1 Ultrametric Fitting
  - Ailon and Charikar (FOCS 2005; SIAM J. Comput. 2011)
- first $O(1)$-approximation for HCC
  - Cohen-Addad, Das, Kipouridis, Parotsidis, and Thorup (FOCS 2021; J. ACM 2024)

Our results on HCC

k ≔ #distinct distances in the input distance function D

- APX-hard
- min{n, $O(k \log n)$}-approximation for L_1 Ultrametric Fitting
  - Harb, Kannan, and McGregor (APPROX 2005)
- min{k+2, $O(\log n \log\log n)$}-approximation for L_1 Ultrametric Fitting
  - Ailon and Charikar (FOCS 2005; SIAM J. Comput. 2011)
- first $O(1)$-approximation for HCC
  - Cohen-Addad, Das, Kipouridis, Parotsidis, and Thorup (FOCS 2021; J. ACM 2024)
- **Our result: 25.7846-approximation for HCC**

Previous results on L_0 Ultrametric Fitting

- APX-hard
- first $O(1)$-approximation
    - Cohen-Addad, Fan, Lee, and Mesmay (FOCS 2022, SIAM J. Comput. 2025)
- 5-approximation
    - Charikar and Gao (SODA 2024)

Our results on L_0 Ultrametric Fitting

- APX-hard
- first $O(1)$-approximation
    - Cohen-Addad, Fan, Lee, and Mesmay (FOCS 2022, SIAM J. Comput. 2025)
- 5-approximation
    - Charikar and Gao (SODA 2024)
- **Our result: (simple) 5-approximation**

Our Algorithm for HCC

Standard LP

[Ailon and Charikar, 2005]
[Cohen-Addad, Das, Kipouridis, Parotsidis, and Thorup, 2021]

- “Distance” variable $x_{ij}^{(t)}$ for all *ij* ∈ *E*, *t* ∈ [*ℓ*]
  - $x_{ij}^{(t)} = 1 \;\Rightarrow$ *i* and *j* are separated at layer *t*
  - $x_{ij}^{(t)} = 0 \;\Rightarrow$ *i* and *j* are clustered at layer *t*

Standard LP

[Ailon and Charikar, 2005]
[Cohen-Addad, Das, Kipouridis, Parotsidis, and Thorup, 2021]

- “Distance” variable $x_{ij}^{(t)}$ for all *ij* ∈ *E*, *t* ∈ [ℓ]
  - $x_{ij}^{(t)} = 1$ ⇒ *i* and *j* are separated at layer *t* ($\boxed{+}$ edge is a disagreement)
  - $x_{ij}^{(t)} = 0$ ⇒ *i* and *j* are clustered at layer *t* ($\boxed{-}$ edge is a disagreement)

Standard LP

[Ailon and Charikar, 2005]
[Cohen-Addad, Das, Kipouridis, Parotsidis, and Thorup, 2021]

- “Distance” variable $x_{ij}^{(t)}$ for all *ij* ∈ *E*, *t* ∈ [*ℓ*]
  - $x_{ij}^{(t)} = 1$ ⇒ *i* and *j* are separated at layer *t* (⊞ edge is a disagreement)
  - $x_{ij}^{(t)} = 0$ ⇒ *i* and *j* are clustered at layer *t* (⊟ edge is a disagreement)
- LP relaxation

$$\sum_{ij \in E_{+}^{(t)}} x_{ij}^{(t)} + \sum_{ij \in E_{-}^{(t)}} \left(1 - x_{ij}^{(t)}\right)$$

(fractional) number of disagreements (at layer *t*)

Standard LP

[Ailon and Charikar, 2005]
[Cohen-Addad, Das, Kipouridis, Parotsidis, and Thorup, 2021]

- “Distance” variable $x_{ij}^{(t)}$ for all *ij* ∈ *E*, *t* ∈ [*ℓ*]
    - $x_{ij}^{(t)} = 1$ ⇒ *i* and *j* are separated at layer *t* (⊞ edge is a disagreement)
    - $x_{ij}^{(t)} = 0$ ⇒ *i* and *j* are clustered at layer *t* (⊟ edge is a disagreement)
- LP relaxation
    - minimize $\sum_{t\in[\ell]} \delta^{(t)} \left(\sum_{ij\in E_+^{(t)}} x_{ij}^{(t)} + \sum_{ij\in E_-^{(t)}} \left(1 - x_{ij}^{(t)}\right)\right)$

Standard LP

[Ailon and Charikar, 2005]
[Cohen-Addad, Das, Kipouridis, Parotsidis, and Thorup, 2021]

- "Distance" variable $x_{ij}^{(t)}$ for all *ij* ∈ *E*, *t* ∈ [ℓ]
  - $x_{ij}^{(t)} = 1$ ⇒ *i* and *j* are separated at layer *t* (⊞ edge is a disagreement)
  - $x_{ij}^{(t)} = 0$ ⇒ *i* and *j* are clustered at layer *t* (⊟ edge is a disagreement)
- LP relaxation
  - minimize $\sum_{t\in[\ell]} \delta^{(t)} \left(\sum_{ij\in E_{+}^{(t)}} x_{ij}^{(t)} + \sum_{ij\in E_{-}^{(t)}} \left(1 - x_{ij}^{(t)}\right)\right)$
  - subject to $\boldsymbol{x}$ satisfies the triangle inequalities for each layer

    $\boldsymbol{x}$ is monotone w.r.t. layers

    $\boldsymbol{x} \in [\mathbf{0}, \mathbf{1}]$

Notations

- $\widetilde{\boldsymbol{x}}$: an optimal LP solution
- $\tilde{x}_{ij}^{(t)}$: the *distance* of *ij* at layer *t*

Useful Lemma

- $\tilde{\boldsymbol{x}}$: an optimal LP solution
- $\tilde{x}_{ij}^{(t)}$: the *distance* of *ij* at layer *t*

- **Lemma**. #(⊟ edges at layer *t* with dist < 1)

Useful Lemma

- $\tilde{\boldsymbol{x}}$: an optimal LP solution
- $\tilde{x}_{ij}^{(t)}$: the *distance* of *ij* at layer *t*

- **Lemma**. $\Sigma_t \delta^{(t)} \cdot$ #(⊟ edges at layer *t* with dist < 1)

Useful Lemma

- $\tilde{\boldsymbol{x}}$: an optimal LP solution
- $\tilde{x}_{ij}^{(t)}$: the *distance* of *ij* at layer *t*

- **Lemma**. $\Sigma_t \, \delta^{(t)} \cdot$ #(⊟ edges at layer *t* with dist < 1) ≤ $\mathrm{OPT}_{\mathrm{LP}}$

Useful Lemma

- $\tilde{\boldsymbol{x}}$: an optimal LP solution
- $\tilde{x}_{ij}^{(t)}$: the *distance* of *ij* at layer *t*

- **Lemma**. $\Sigma_t\, \delta^{(t)} \cdot$ #(⊟ edges at layer *t* with dist < 1) ≤ $\mathrm{OPT}_{\mathrm{LP}}$

"Disregarding ⊟ edges with distance < 1 in each layer

only incurs an additive factor of 1"

Useful Lemma

- $\tilde{\boldsymbol{x}}$: an optimal LP solution
- $\tilde{x}_{ij}^{(t)}$: the *distance* of *ij* at layer *t*

- **Lemma**. $\Sigma_t \delta^{(t)} \cdot$ #(⊟ edges at layer *t* with dist < 1) ≤ OPT_{LP}
  - Proved using complementary slackness & weak duality

“Disregarding ⊟ edges with distance < 1 in each layer

only incurs an additive factor of 1”

Algorithm Overview

- Construct a *pre-clustering* $\boldsymbol{Q}^{(t)}$ (partition) for each layer *t*

$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Algorithm Overview

(partition)

- Construct a *pre-clustering* $\boldsymbol{\mathcal{Q}}^{(t)}$ for each layer *t*

```
(Fix layer $t$)
$\boldsymbol{\mathcal{Q}} \leftarrow \{V\}$
while there is $Q \in \boldsymbol{\mathcal{Q}}$ whose diameter is large,
    split $Q$ into two so that #(edges separated & dist < 1) is small
```

$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Algorithm Overview

(partition)

- Construct a *pre-clustering* $\mathcal{Q}^{(t)}$ for each layer *t*

```
(Fix layer $t$)
$\mathcal{Q} \leftarrow \{V\}$
while there is $Q \in \mathcal{Q}$ whose diameter is large,
    split $Q$ into two so that #(edges separated & dist < 1) is small
```

- **small-diameter property**: every *pre-cluster* has a small diameter
- **few-separated-edges property**: #(edges separated by $\mathcal{Q}$ & dist < 1) is small

Algorithm Overview

- Construct a *pre-clustering* $\boldsymbol{\mathcal{Q}}^{(t)}$ for each layer *t*
- Construct hierarchical clusterings bottom-up:

$^*\boldsymbol{\mathcal{P}}^{(0)} \coloneqq \{\ \{u\}: u \in V\ \}$

(at layer *t*) construct the *clustering* $\boldsymbol{\mathcal{P}}^{(t)}$ by only merging clusters in $\boldsymbol{\mathcal{P}}^{(t-1)}$

Algorithm Overview

- Construct a *pre-clustering* $\boldsymbol{\mathcal{Q}}^{(t)}$ for each layer *t*
- Construct hierarchical clusterings bottom-up:

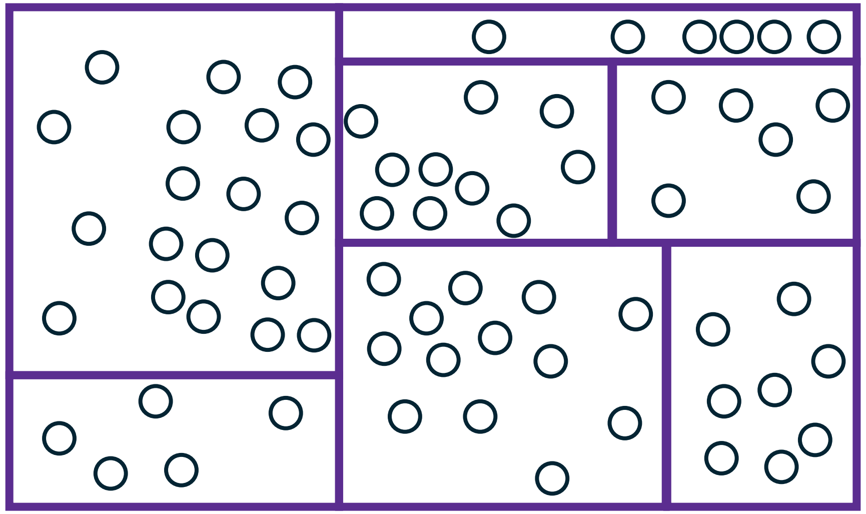
$^*\boldsymbol{\mathcal{P}}^{(0)} := \{ \{u\}: u \in V \}$

(at layer *t*) construct the *clustering* $\boldsymbol{\mathcal{P}}^{(t)}$ by only merging clusters in $\boldsymbol{\mathcal{P}}^{(t-1)}$

Hierarchical constraints satisfied!

Algorithm Overview

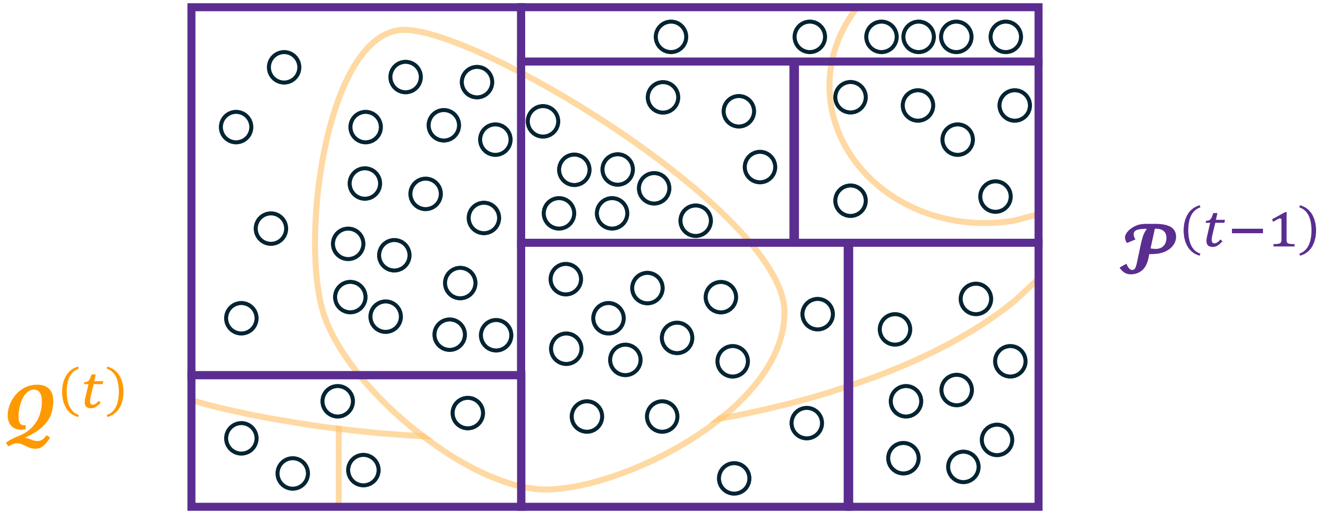
(Fix layer t)



$\mathcal{P}^{(t-1)}$

Algorithm Overview

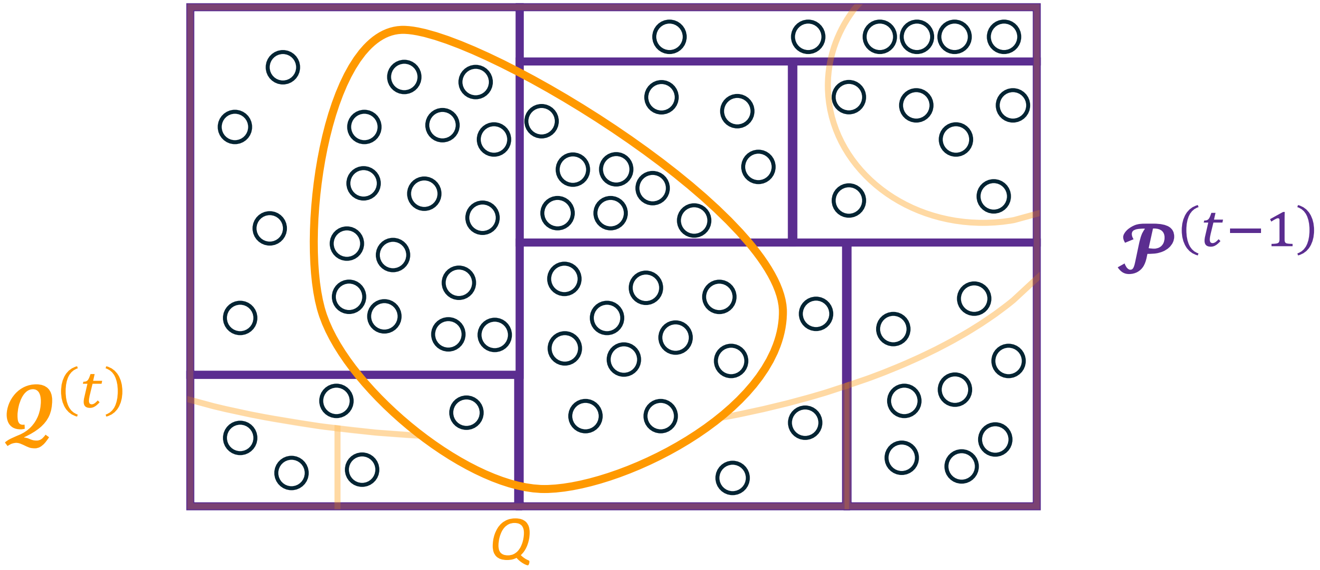
(Fix layer t)



Algorithm Overview

(Fix layer t)

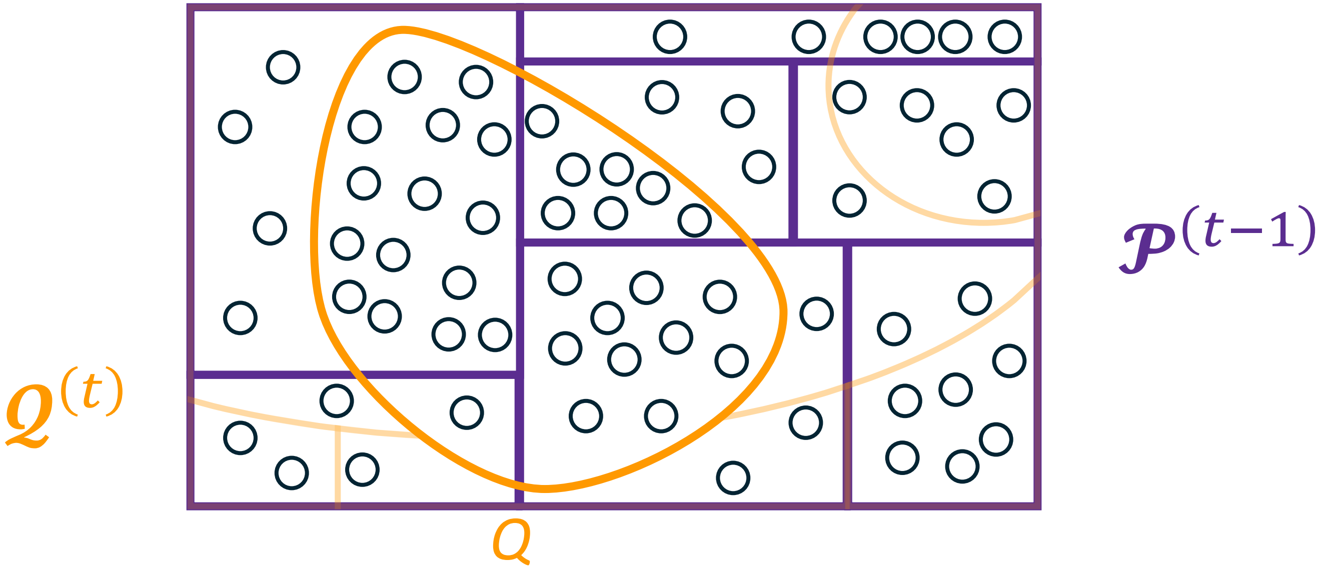
```
for each pre-cluster $Q$ at layer $t$,
```



Algorithm Overview

(Fix layer *t*)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
```



Algorithm Overview

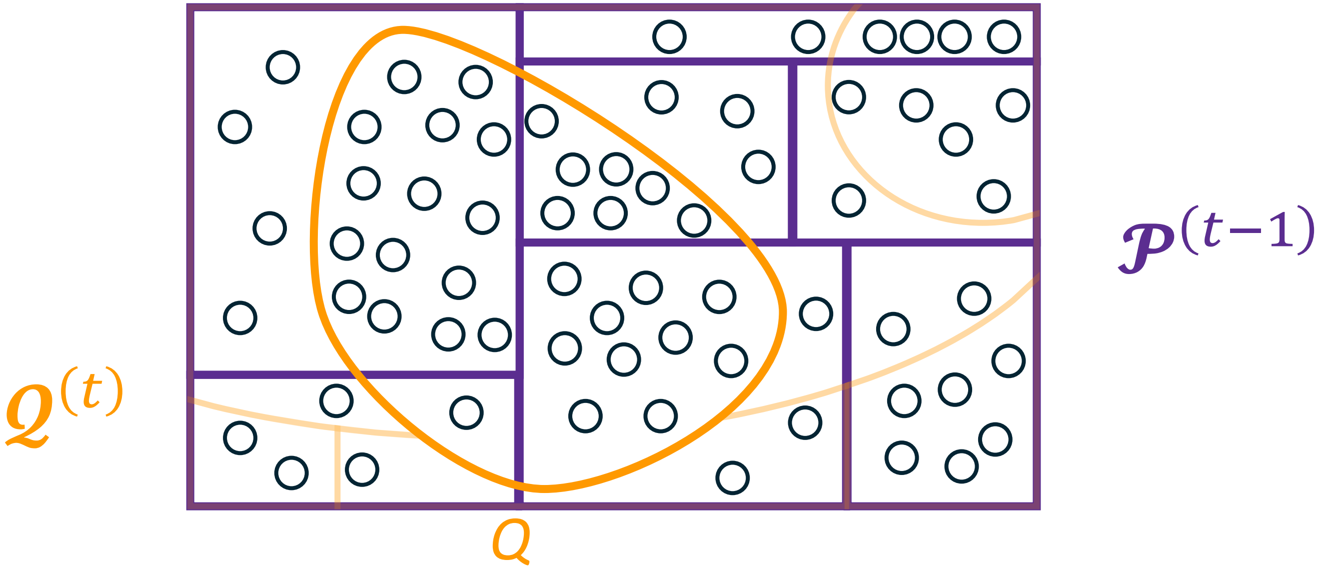
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster Q at layer t,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster Q
```



(roughly)

merging condition:

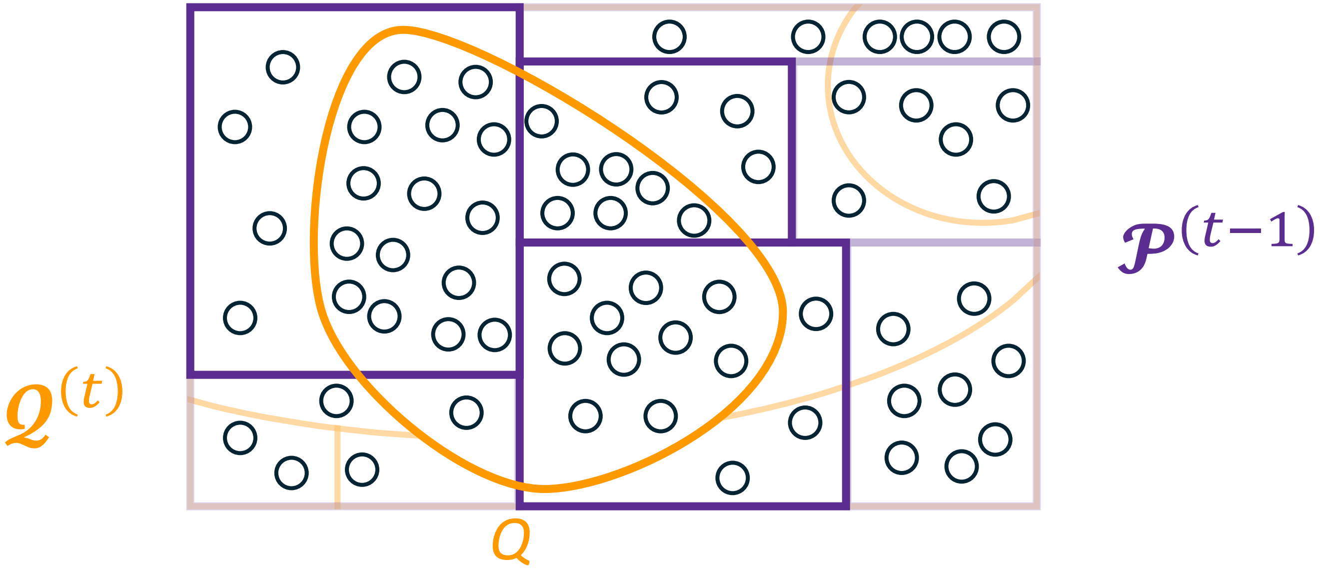
most points in P' are in $P' \cap Q$

Algorithm Overview

(Fix layer t)

for each pre-cluster Q at layer t,

find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a *merging condition* with pre-cluster Q



Algorithm Overview

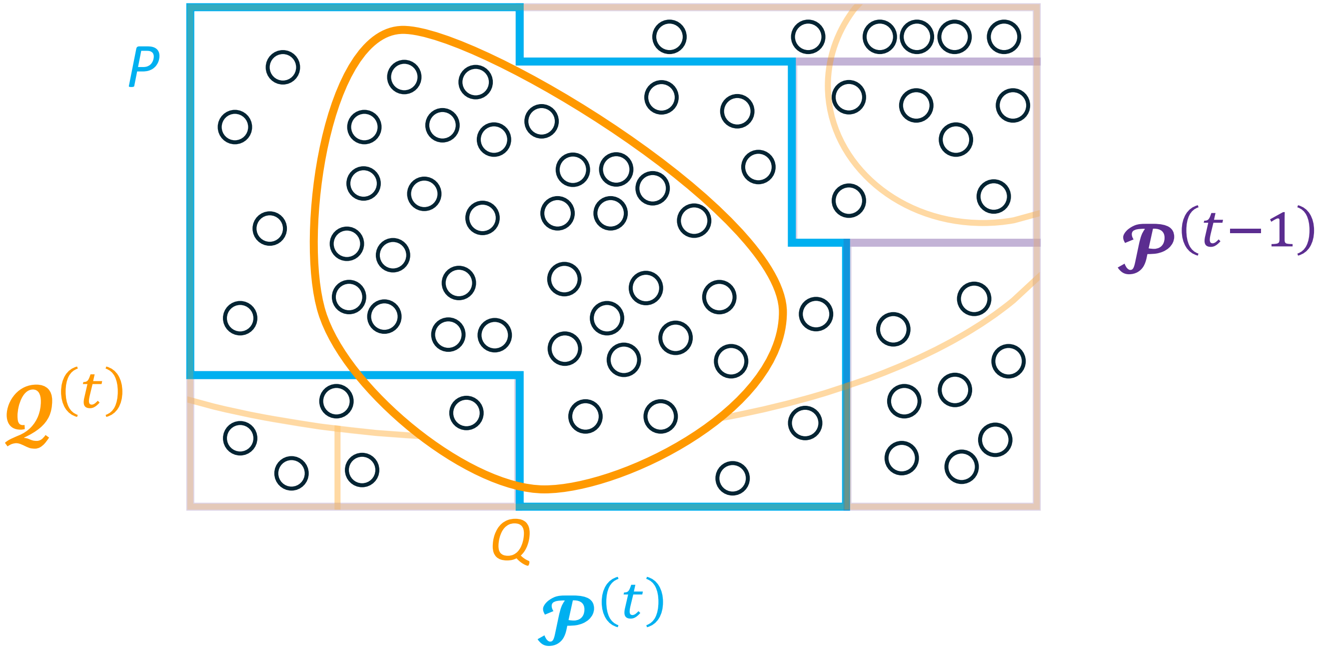
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
```



Algorithm Overview

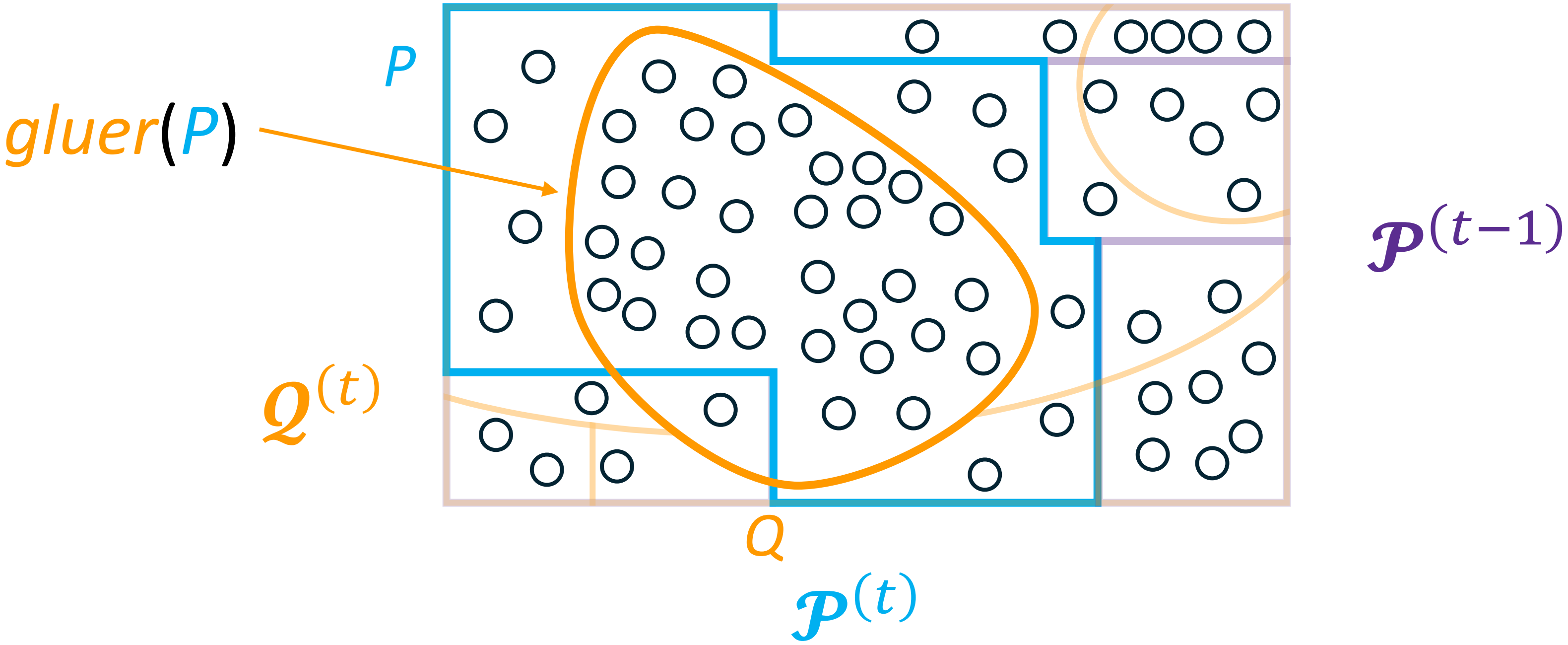
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
```



Algorithm Overview

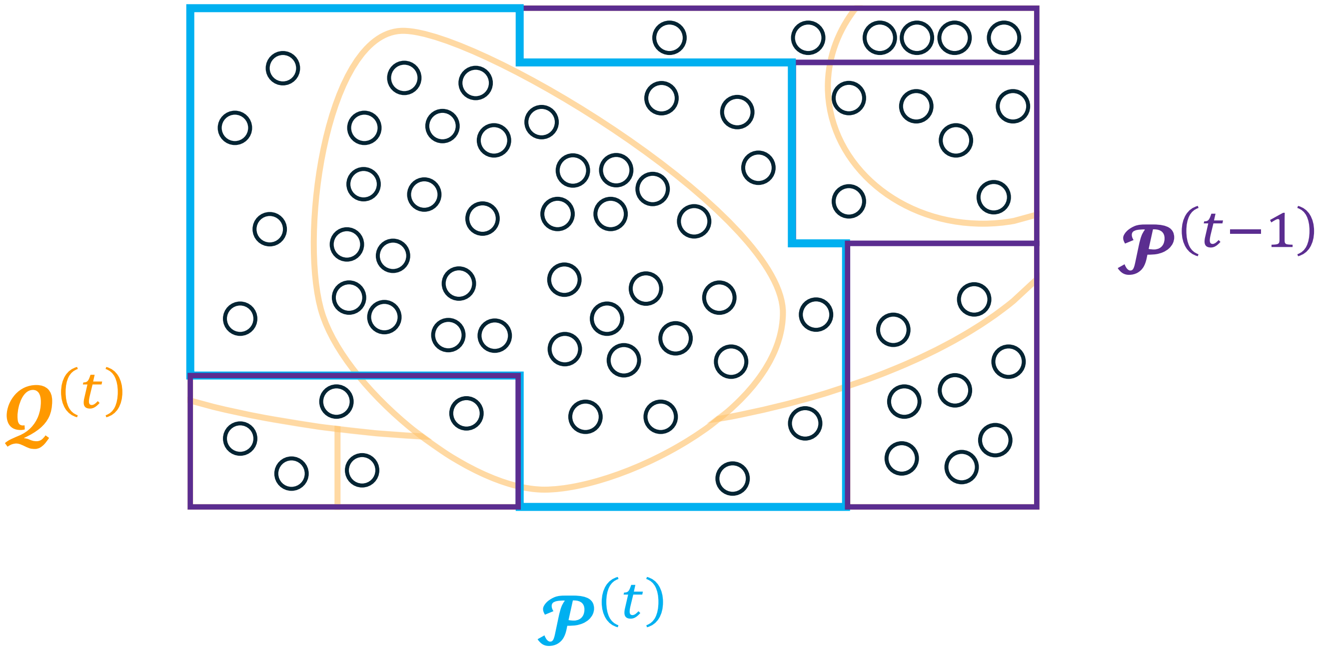
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
```



Algorithm Overview

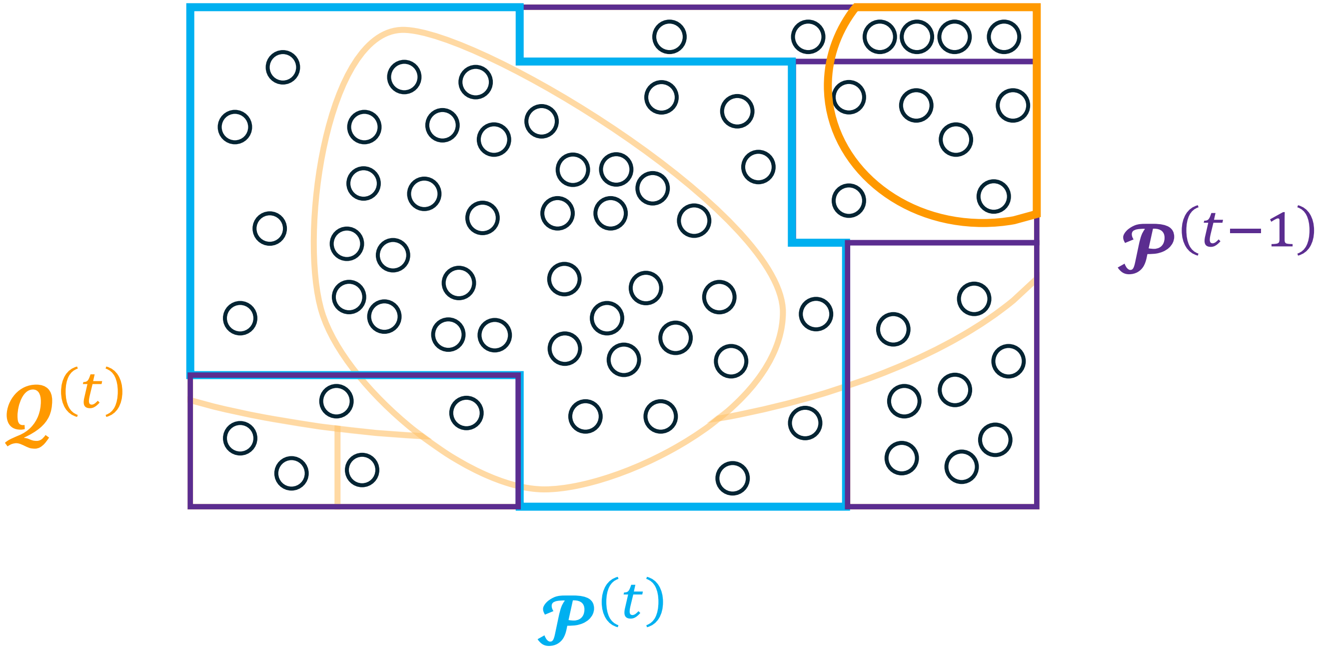
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
```



Algorithm Overview

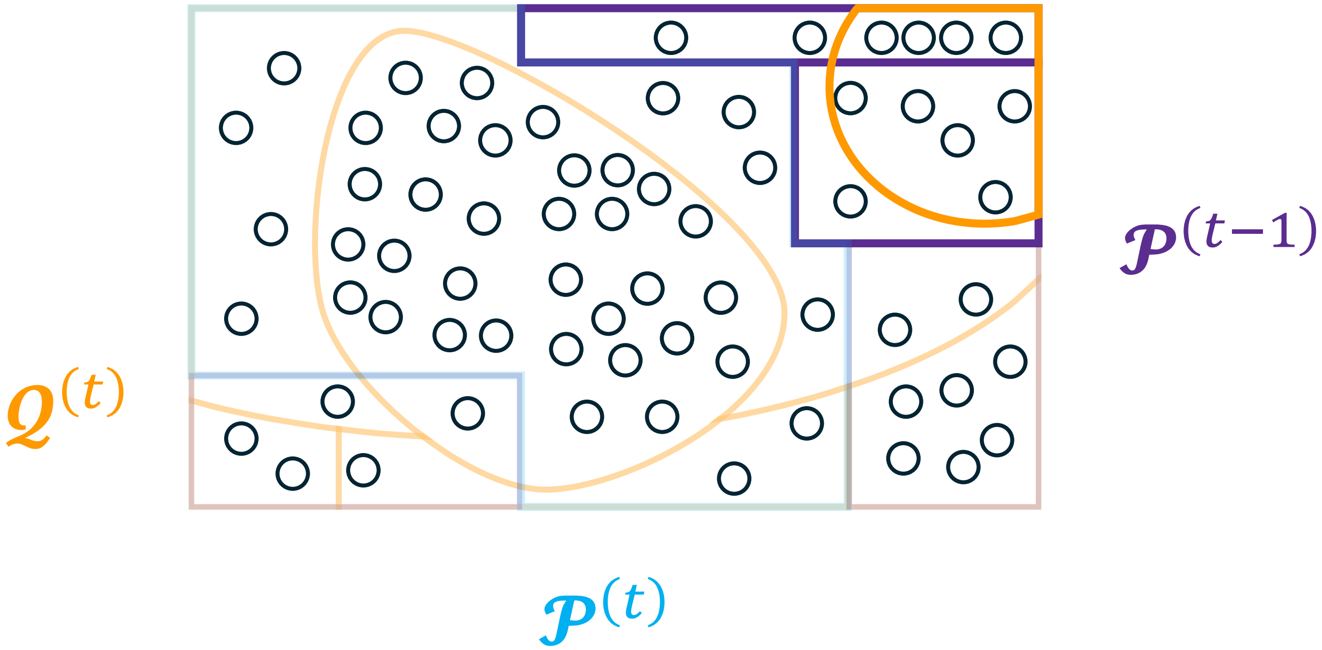
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
```



Algorithm Overview

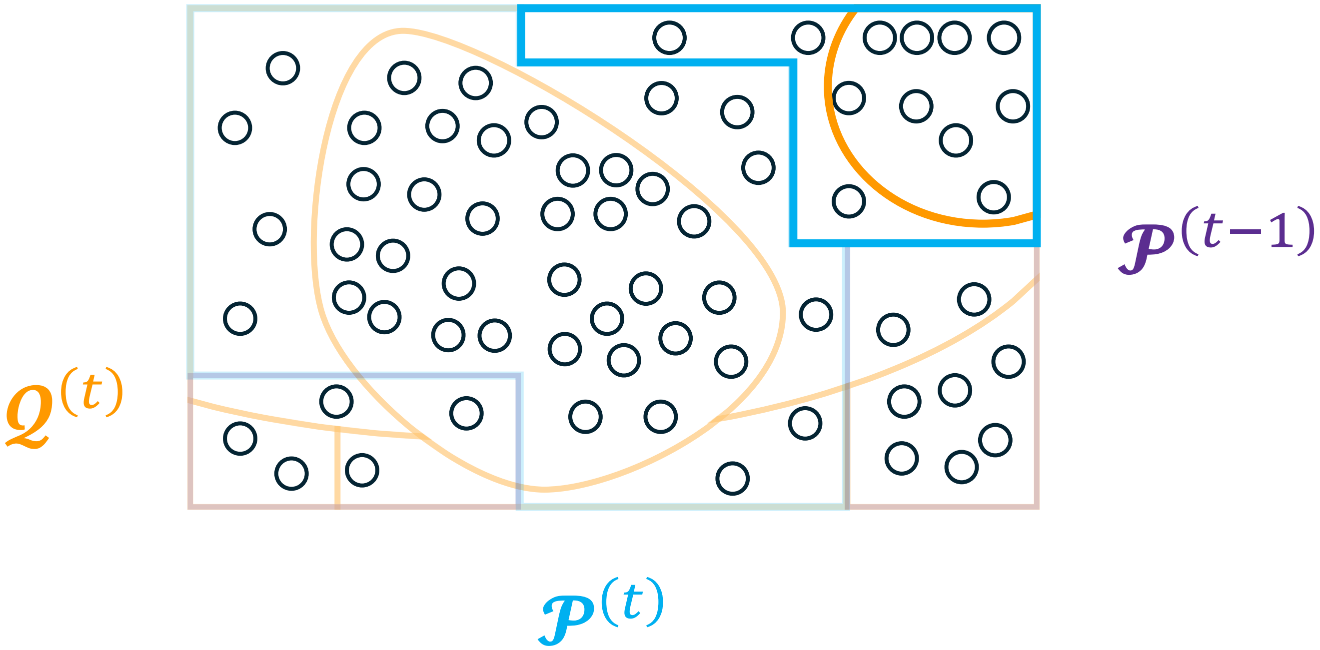
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
```



Algorithm Overview

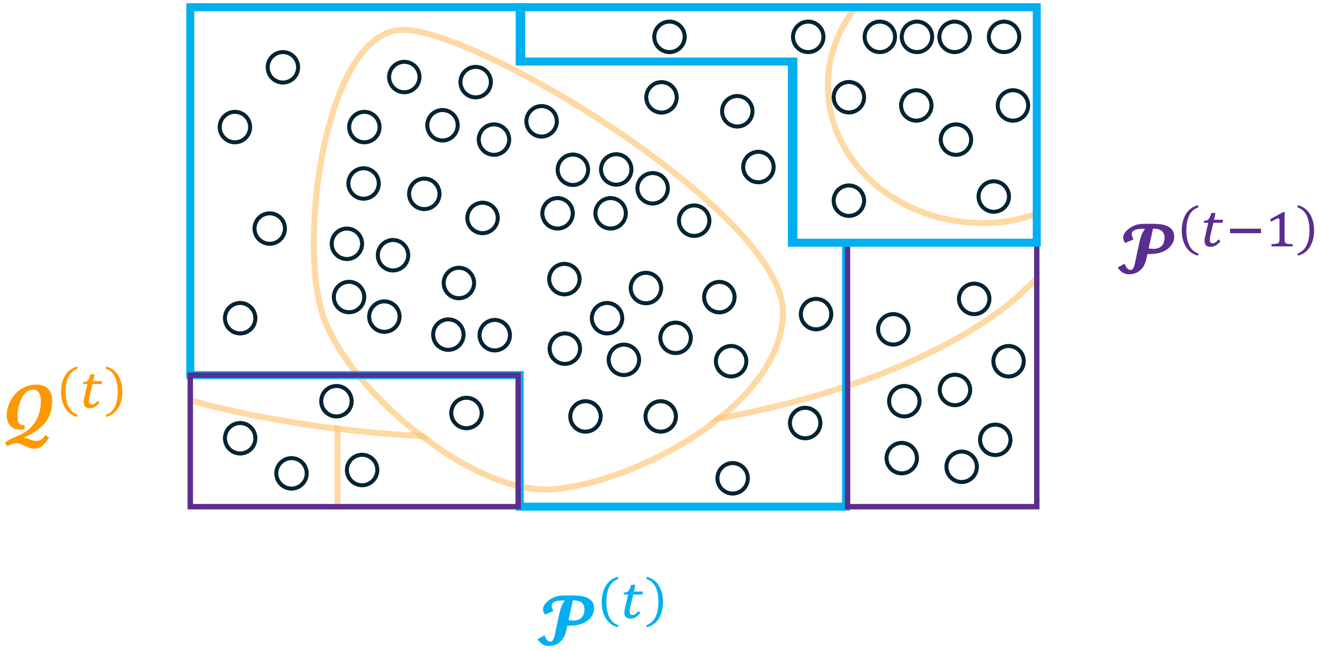
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
```



Algorithm Overview

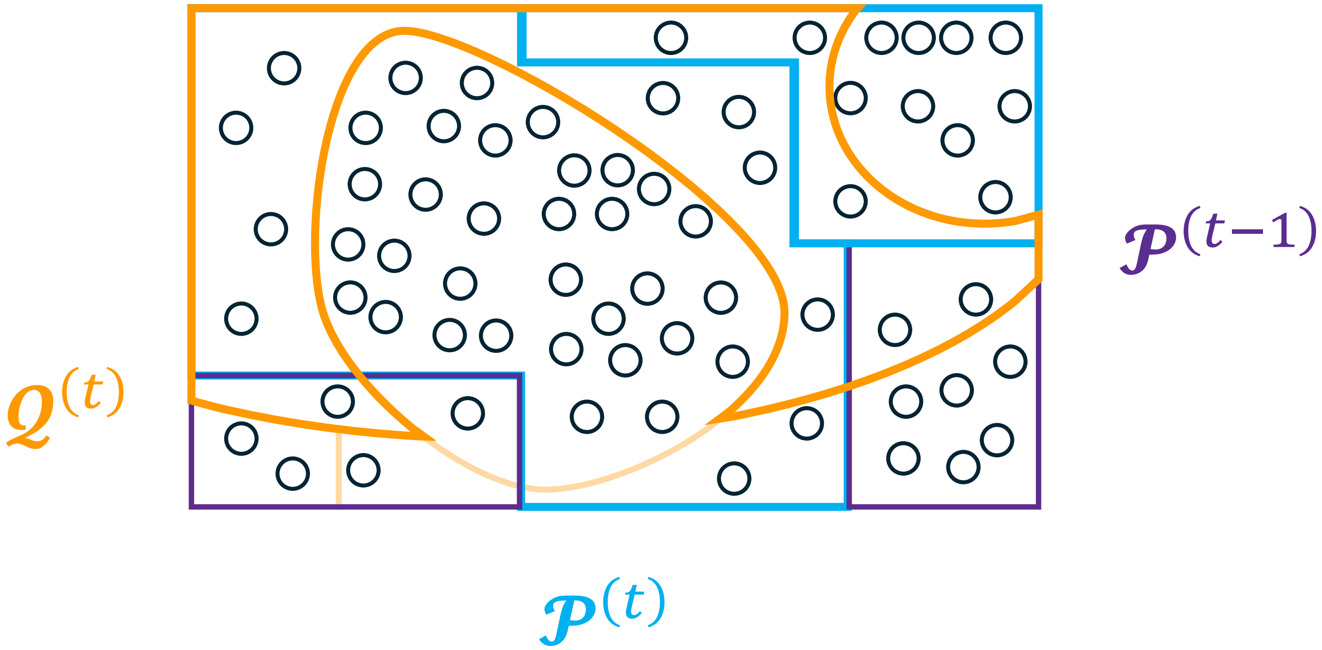
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a *merging condition* with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
```



Algorithm Overview

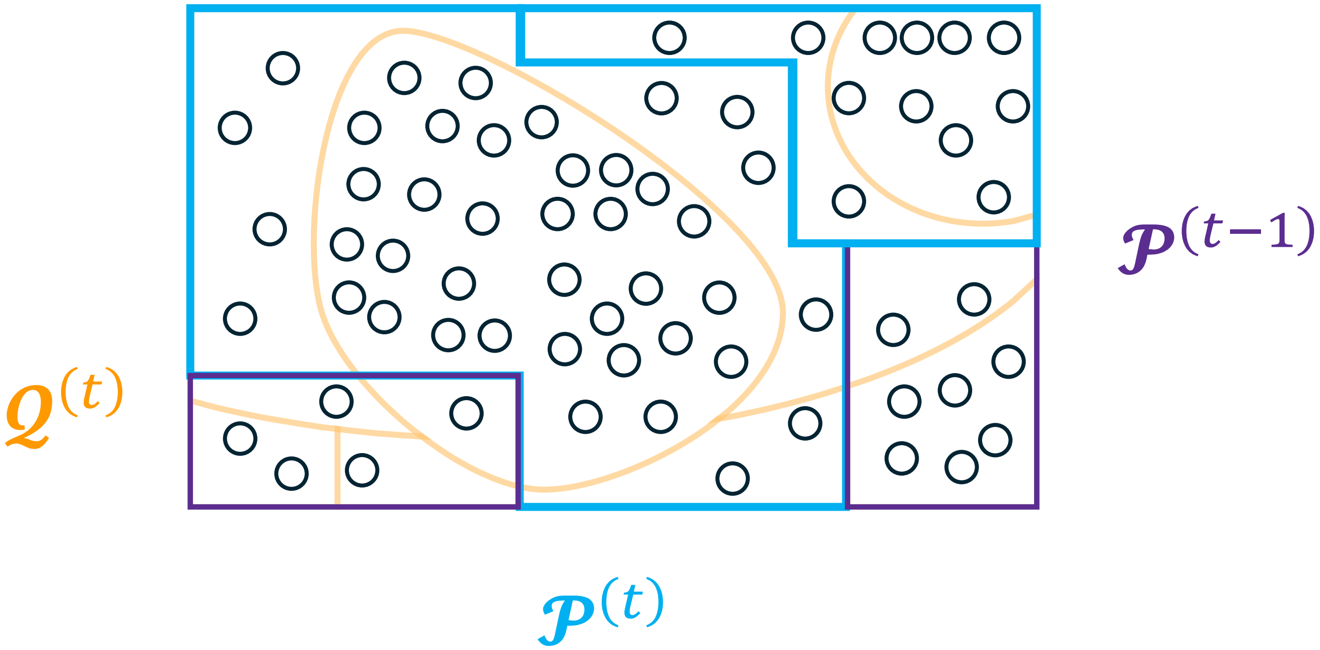
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
```



Algorithm Overview

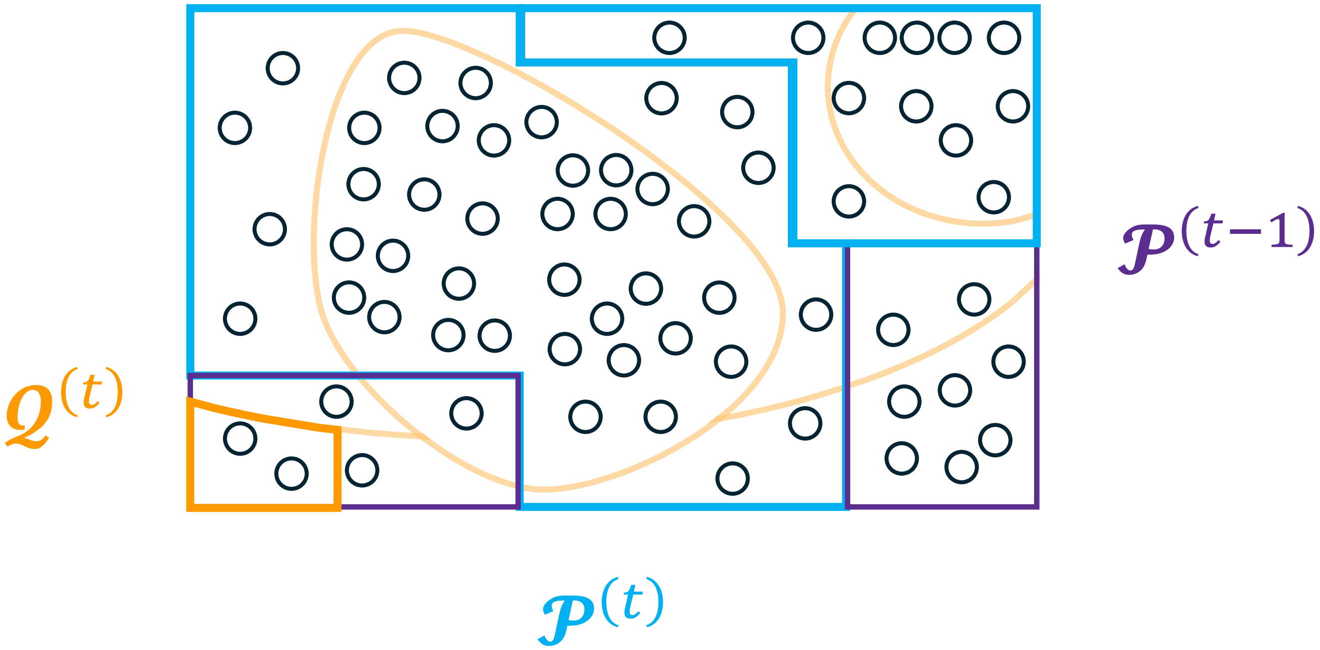
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
```



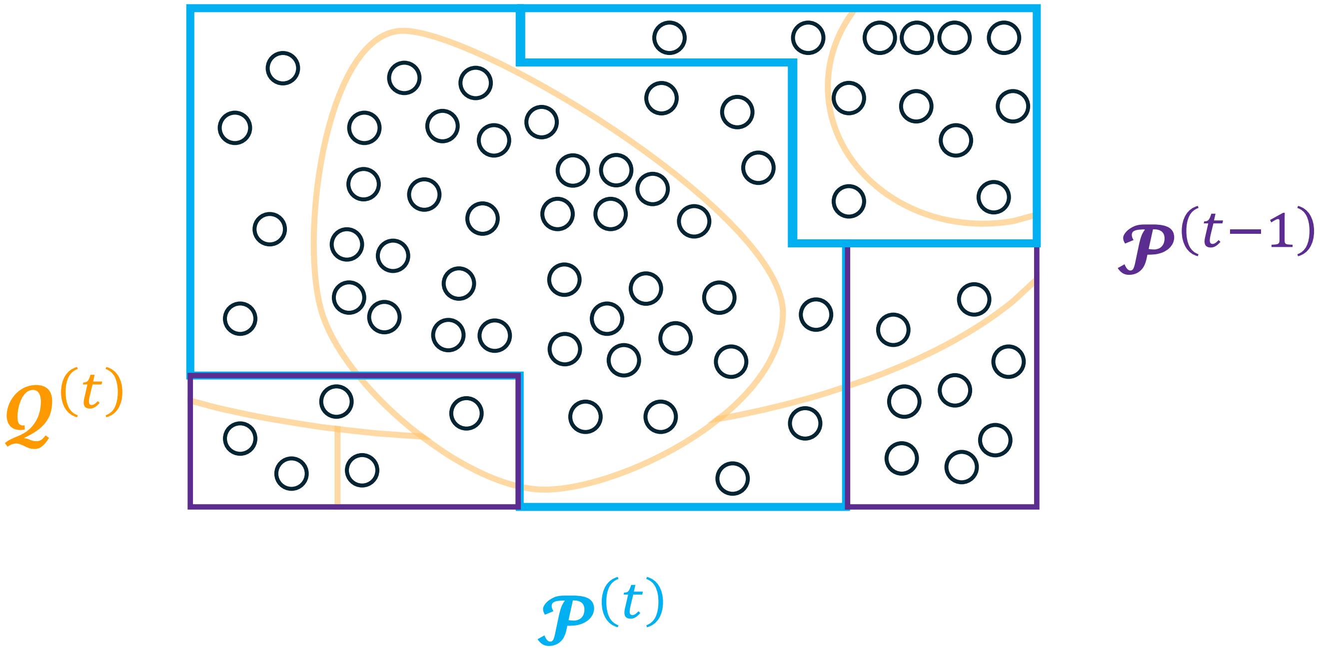
Algorithm Overview

(roughly)

merging condition:
most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
```



Algorithm Overview

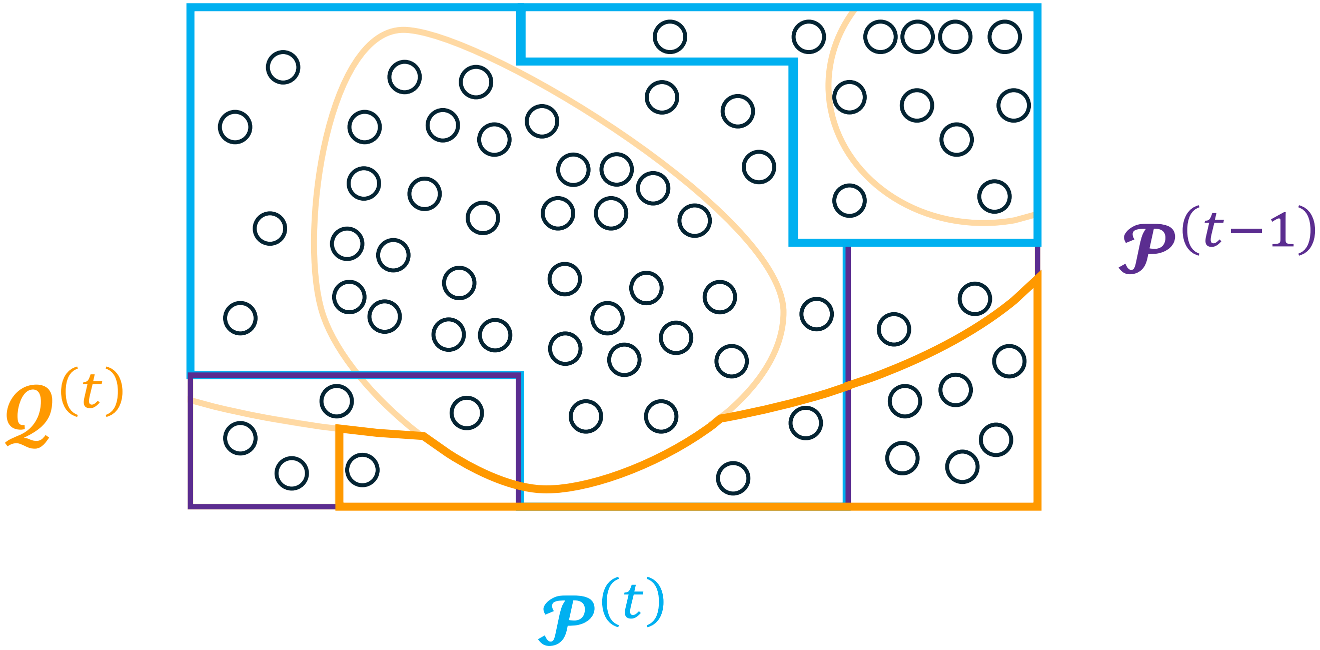
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
```



Algorithm Overview

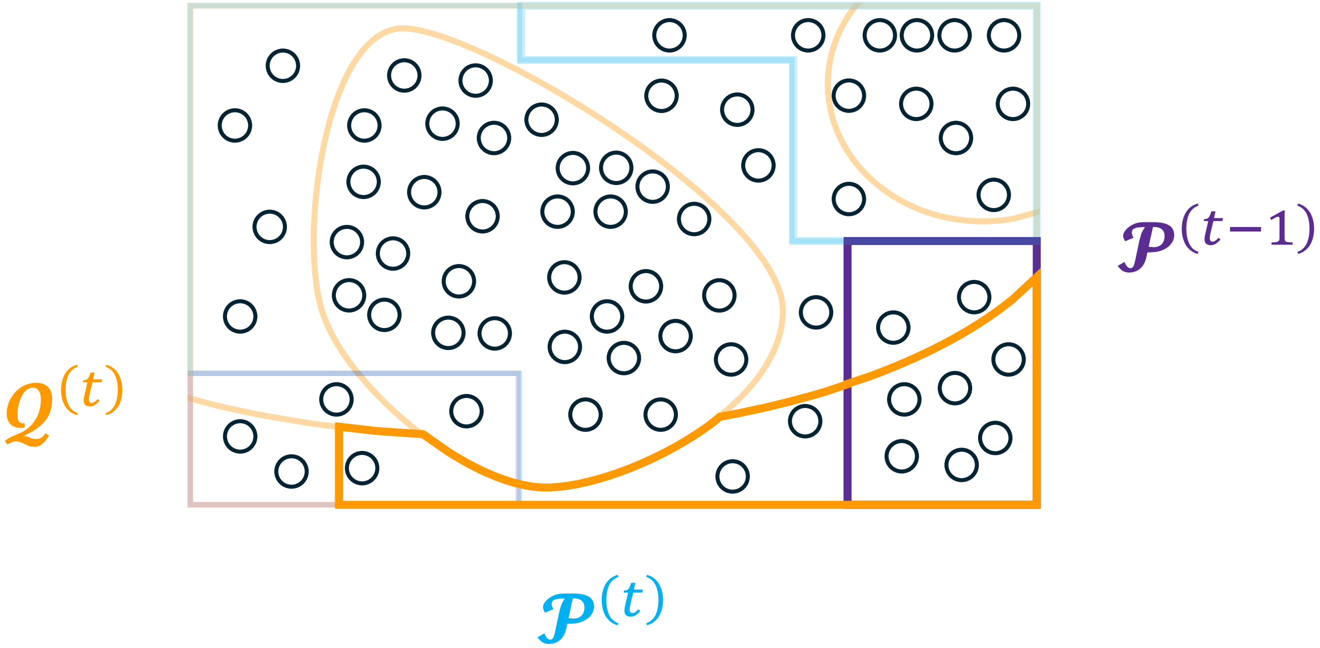
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer *t*)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
```



Algorithm Overview

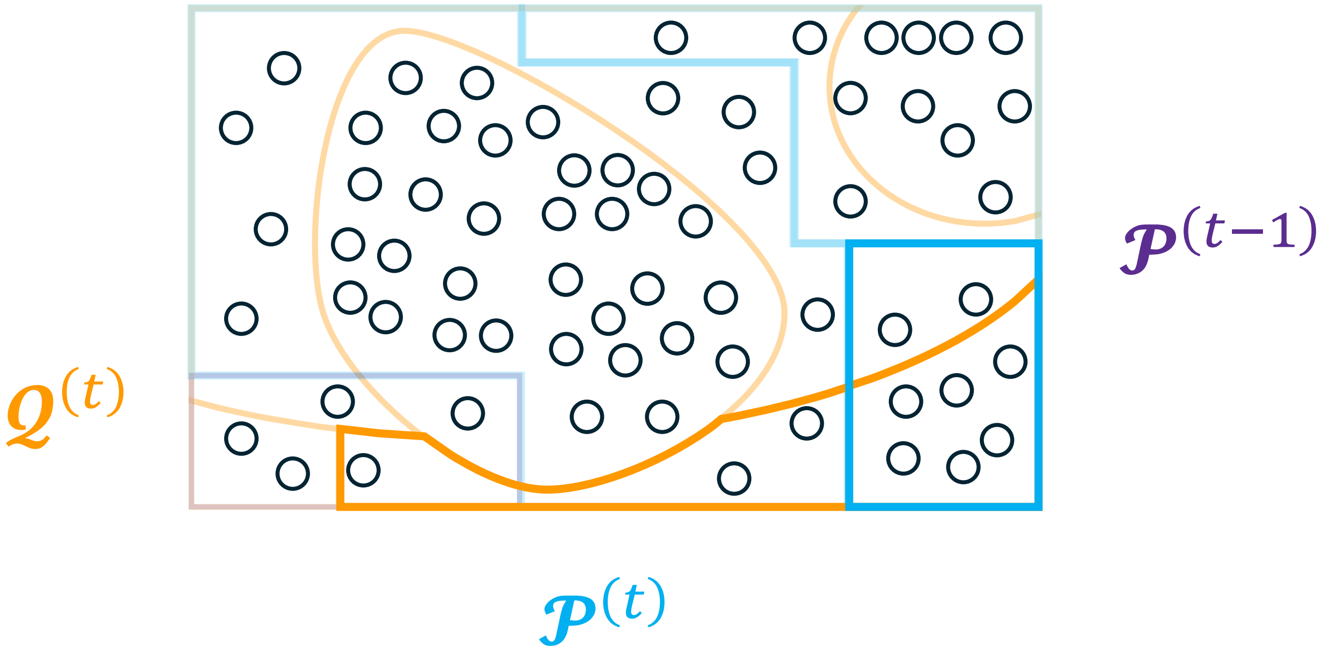
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
```



Algorithm Overview

(roughly)

merging condition:

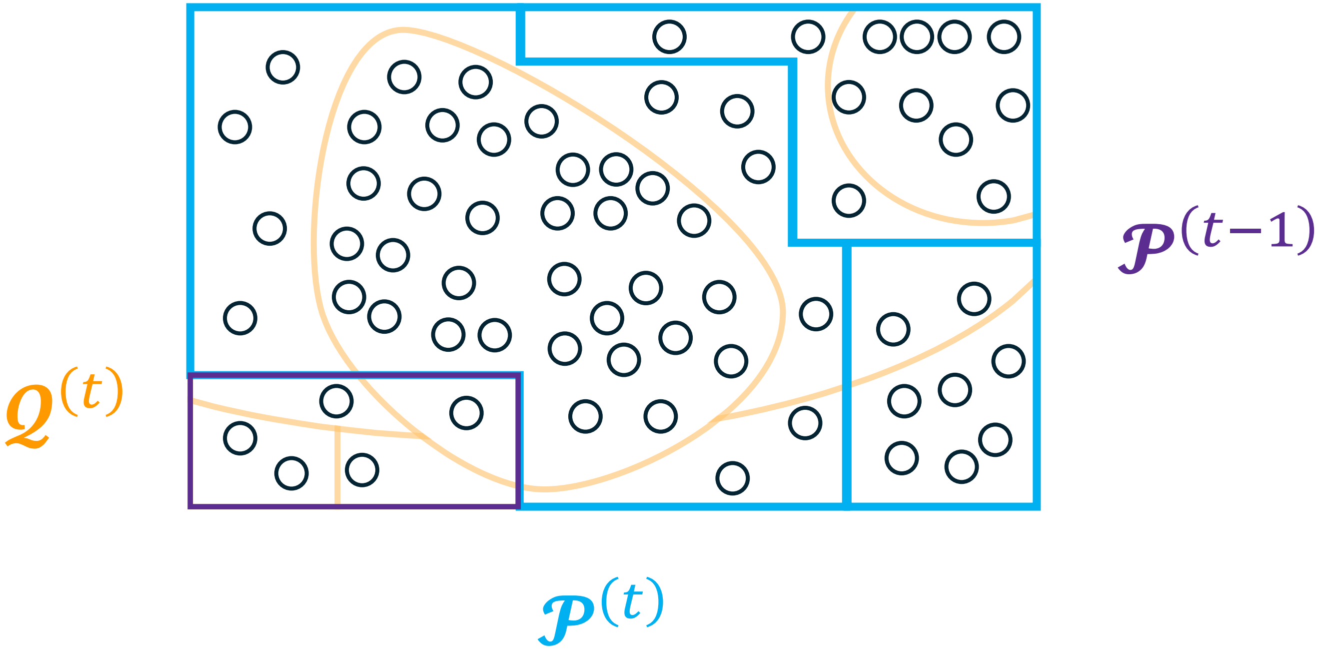
most points in P' are in $P' \cap Q$

(Fix layer t)

for each pre-cluster Q at layer t,

find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a *merging condition* with pre-cluster Q

let P be the merged cluster; add P to $\mathcal{P}^{(t)}$



Algorithm Overview

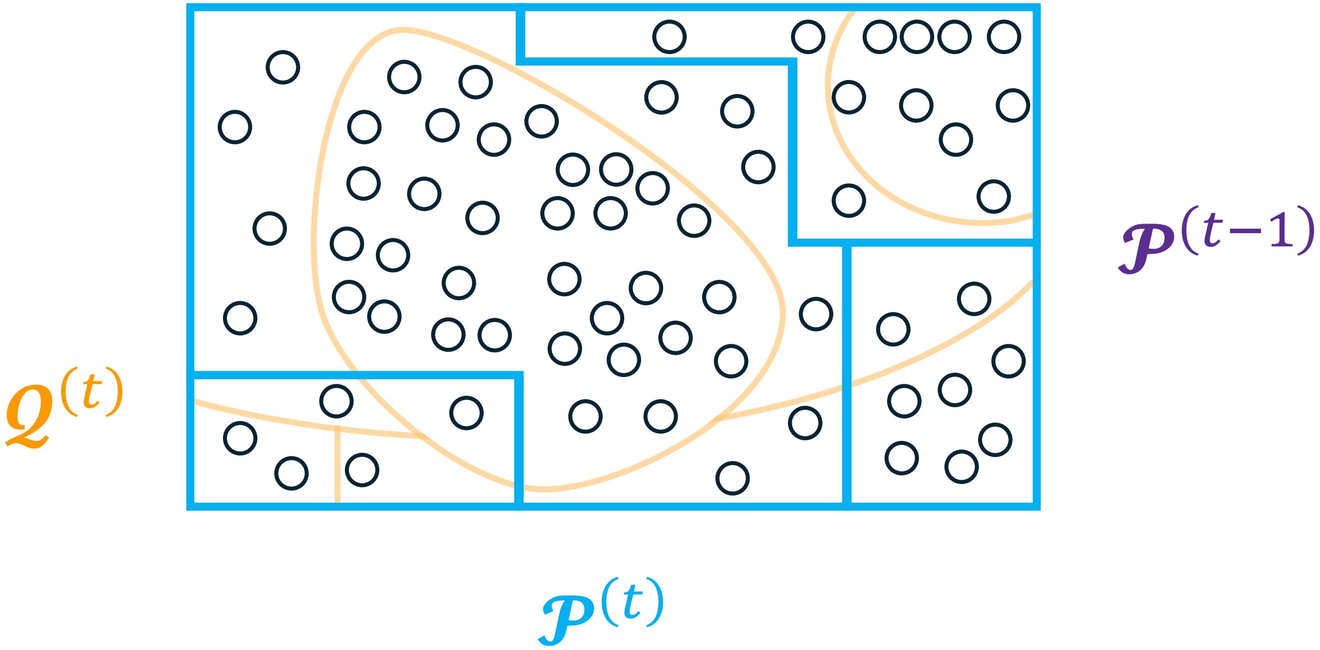
(roughly)

merging condition:

most points in P' are in $P' \cap Q$

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
add all remaining unmerged clusters $P' \in \mathcal{P}^{(t-1)}$ to $\mathcal{P}^{(t)}$
```



(roughly)

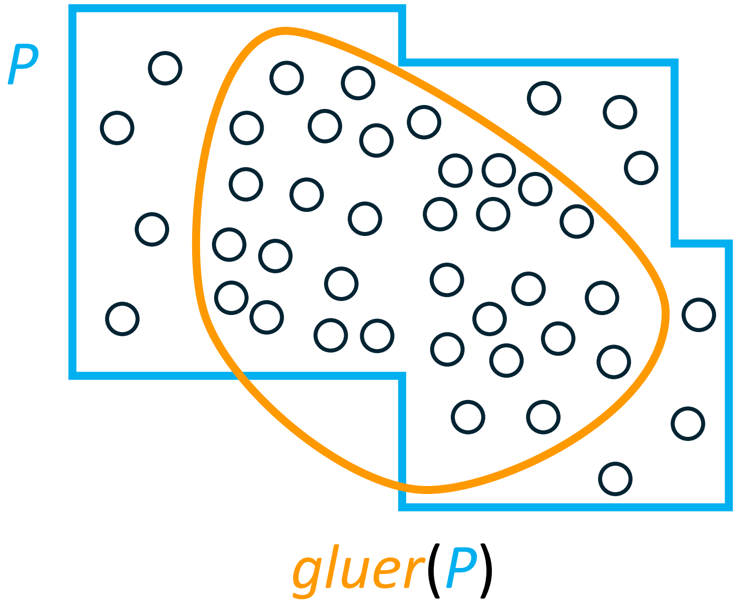
merging condition:

most points in P' are in $P' \cap Q$

Algorithm Overview

(Fix layer t)

```
for each pre-cluster $Q$ at layer $t$,
    find all $P' \in \mathcal{P}^{(t-1)}$ that satisfy a merging condition with pre-cluster $Q$
    let $P$ be the merged cluster; add $P$ to $\mathcal{P}^{(t)}$
add all remaining unmerged clusters $P' \in \mathcal{P}^{(t-1)}$ to $\mathcal{P}^{(t)}$
```



- **concentration property**:

most points in P are in $P \cap$ *gluer*(P)

Analysis

Analysis Overview

(Fix layer *t*)

- #(disagreements)

$$\leq O(1)\cdot\left(\underbrace{\sum_{ij\in E_+^{(t)}} x_{ij}^{(t)} + \sum_{ij\in E_-^{(t)}} (1 - x_{ij}^{(t)})}_{\text{LP value at layer } t} \right)$$

Analysis Overview

(Fix layer *t*)

- #(disagreements)

$$\leq O(1)\cdot\left(\underbrace{\sum_{ij\in E_+^{(t)}} x_{ij}^{(t)} + \sum_{ij\in E_-^{(t)}} (1 - x_{ij}^{(t)})}_{\boxed{\text{LP value at layer } t}} \right)$$

Objective: $\sum_{t\in[\ell]} \delta^{(t)} \underline{\left(\sum_{ij\in E_+^{(t)}} x_{ij}^{(t)} + \sum_{ij\in E_-^{(t)}} \left(1 - x_{ij}^{(t)}\right)\right)}$

$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

(Fix layer *t*)

- #(disagreements disregarding – edges with dist < 1)

(Useful Lemma)

$$\leq O(1)\cdot\left(\sum_{ij\in E_+^{(t)}} x_{ij}^{(t)} + \sum_{ij\in E_-^{(t)}} (1 - x_{ij}^{(t)}) \right)$$

$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

(Fix layer *t*)

- #(disagreements disregarding – edges with dist < 1) **(Useful Lemma)**

  ≤ *O*(1)·#(edges separated by $\boldsymbol{Q}$ & dist < 1)

  ≤ *O*(1)·($\sum_{ij \in E_+^{(t)}} x_{ij}^{(t)} + \sum_{ij \in E_-^{(t)}} (1 - x_{ij}^{(t)})$) **(few-separated-edges property)**

$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

(Fix layer *t*)

Want:

- #(disagreements disregarding – edges with dist < 1) **(Useful Lemma)**

  ≤ *O*(1)·#(edges separated by $\mathcal{Q}$ & dist < 1)

  ≤ *O*(1)·($\sum_{ij \in E_+^{(t)}} x_{ij}^{(t)} + \sum_{ij \in E_-^{(t)}} (1 - x_{ij}^{(t)})$) **(few-separated-edges property)**

$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

(Fix layer *t*)

- #(disagreements disregarding – edges with dist < 1)
  - ①: #(– edges clustered in $\mathcal{P}$ & dist = 1)
  - ②: #(+ edges separated by $\mathcal{P}$)

Want:

① + ② ≤ $O(1)$·#(edges separated by $\mathcal{Q}$ & dist < 1)

$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

(Fix layer *t*)

- #(disagreements disregarding – edges with dist < 1)
  - ①: #(– edges clustered in $\mathcal{P}$ & dist = 1)
  - ②: #(+ edges separated by $\mathcal{P}$)

① ≤ $O(1)$·#(edges separated by $\mathcal{Q}$ & dist < 1 & clustered in $\mathcal{P}$)

+) ② ≤ $O(1)$·#(edges separated by $\mathcal{Q}$ & dist < 1 & separated by $\mathcal{P}$)

Want:

$$\boxed{① + ② \le O(1)\cdot\#(\text{edges separated by } \mathcal{Q} \text{ \& dist} < 1)}$$

$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

(Fix layer *t*)

- #(disagreements disregarding – edges with dist < 1)
  - ①: #(– edges clustered in $\mathcal{P}$ & dist = 1)
  - ②: #(+ edges separated by $\mathcal{P}$)

Our focus: ① ≤ *O*(1)·#(edges separated by $\mathcal{Q}$ & dist < 1 & clustered in $\mathcal{P}$)

+) ② ≤ *O*(1)·#(edges separated by $\mathcal{Q}$ & dist < 1 & separated by $\mathcal{P}$)

Want:

① + ② ≤ *O*(1)·#(edges separated by $\mathcal{Q}$ & dist < 1)

$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview (Fix layer *t*)

Our focus: #(– edges clustered in $\mathcal{P}$ & dist = 1)

≤ *O*(1)·#(edges clustered in $\mathcal{P}$ & separated by $\mathcal{Q}$ & dist < 1)

$\tilde{x}_e^{(t)}$: *distance* of e (at layer t)

Analysis Overview (Fix layer t, $P \in \mathcal{P}$)

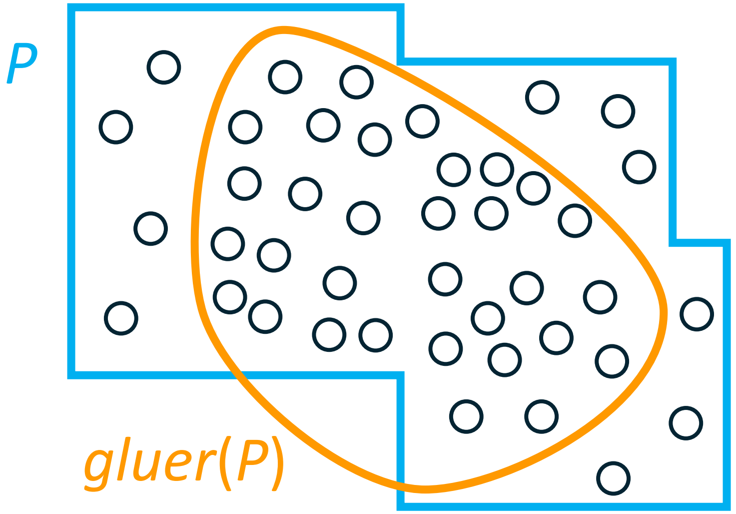
Our focus: #(– edges clustered in P & dist = 1)

≤ $O(1)$·#(edges clustered in P & separated by $\mathcal{Q}$ & dist < 1)

$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer t)

Analysis Overview

#(– edges clustered in *P*   & dist = 1)



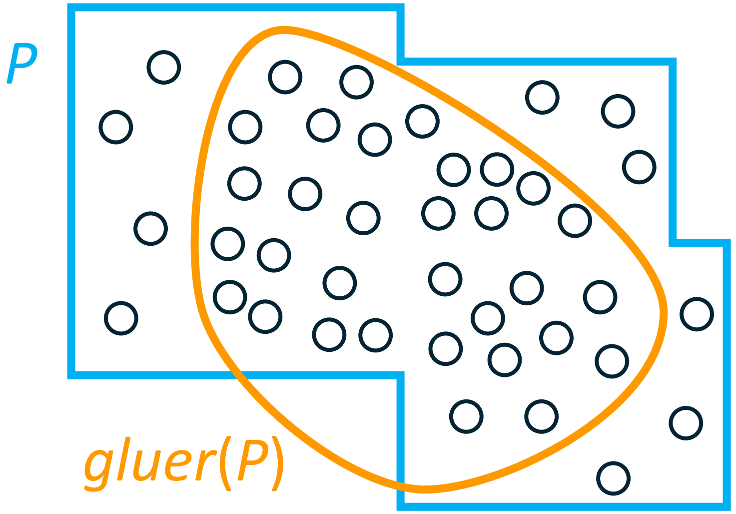
$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

#(– edges clustered in *P*, separated by *gluer*(*P*) & dist = 1)

⊕ #(– edges clustered in *P* \ *gluer*(*P*) & dist = 1)

#(– edges clustered in *P* ∩ *gluer*(*P*) & dist = 1)



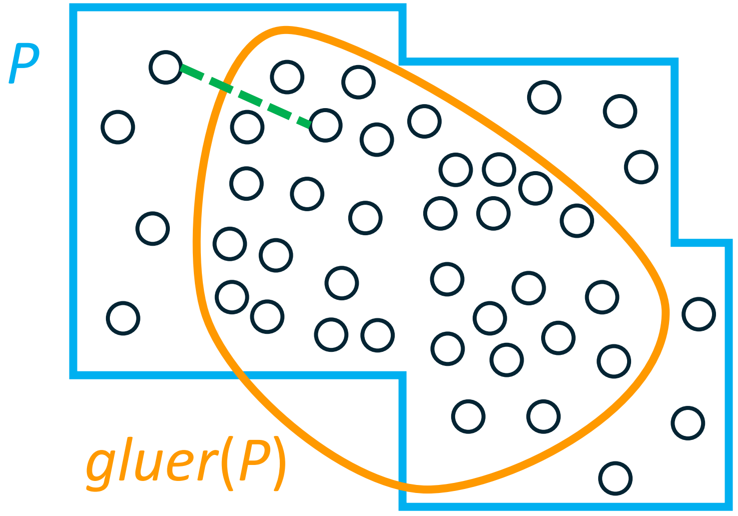
$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

#(– edges clustered in *P*, separated by *gluer*(*P*) & dist = 1)

⊕ #(– edges clustered in *P* \ *gluer*(*P*) & dist = 1)

#(– edges clustered in *P* ∩ *gluer*(*P*) & dist = 1)



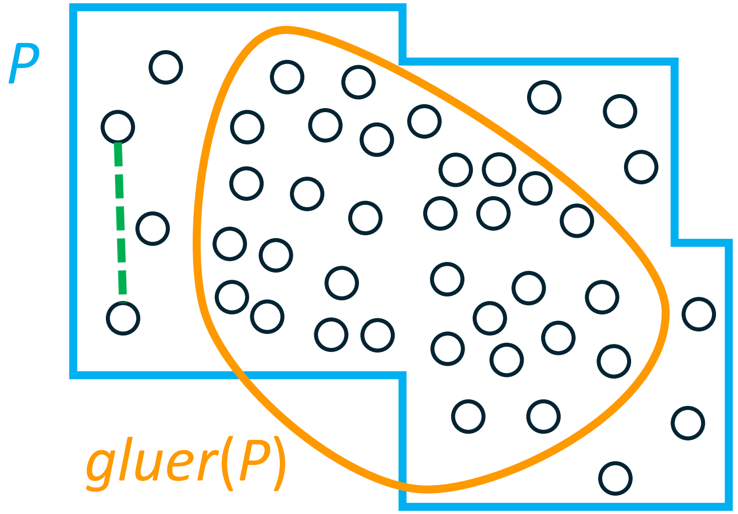
$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

#(– edges clustered in *P*, separated by *gluer*(*P*) & dist = 1)

⊕ #(– edges clustered in *P* \ *gluer*(*P*) & dist = 1)

#(– edges clustered in *P* ∩ *gluer*(*P*) & dist = 1)



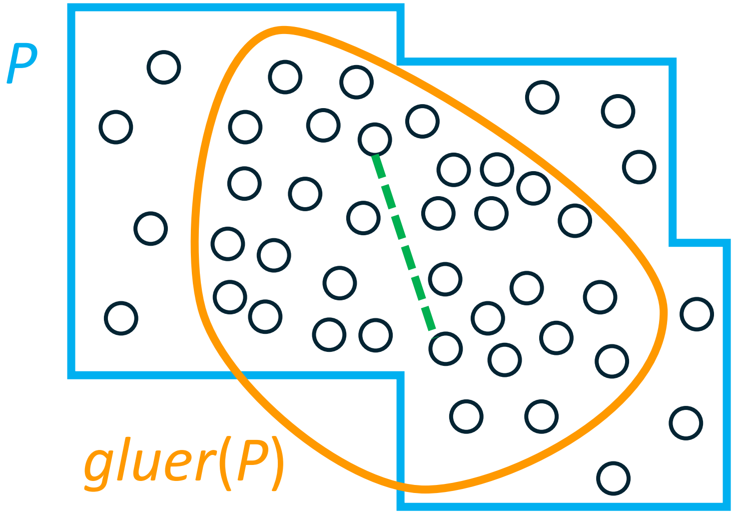
$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

#(– edges clustered in *P*, separated by *gluer*(*P*) & dist = 1)

⊕ #(– edges clustered in *P* \ *gluer*(*P*) & dist = 1)

#(– edges clustered in *P* ∩ *gluer*(*P*) & dist = 1)



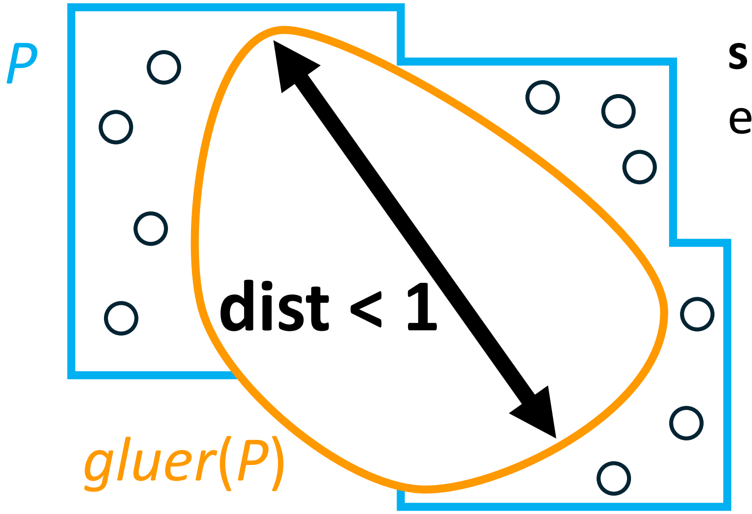
$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

#(– edges clustered in *P*, separated by *gluer*(*P*) & dist = 1)

⊕ #(– edges clustered in *P* \ *gluer*(*P*) & dist = 1)

#(– edges clustered in *P* ∩ *gluer*(*P*) & dist = 1) --- impossible



small-diameter property:
every *pre-cluster* has a small diameter

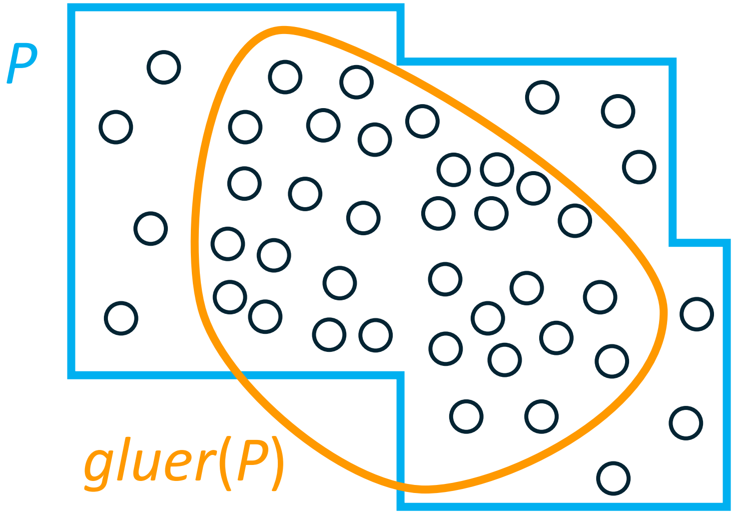
$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

#(– edges clustered in *P*, separated by *gluer*(*P*) & dist = 1)

⊕

#(– edges clustered in *P* \ *gluer*(*P*) & dist = 1)



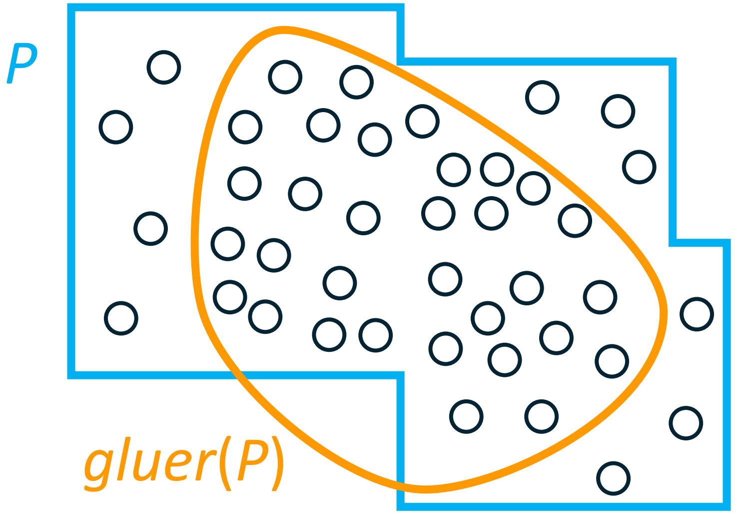
$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

⊕

#(~~−~~edges clustered in *P*, separated by *gluer*(*P*) & ~~dist = 1~~)

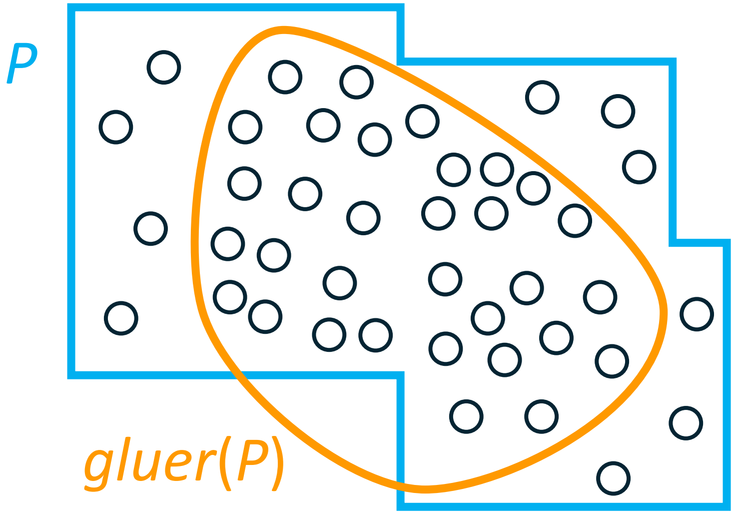
#(~~−~~edges clustered in *P* \ *gluer*(*P*) & ~~dist = 1~~)



$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

$$\leq \oplus \begin{cases} \#(\text{edges clustered in } P\text{, separated by } gluer(P)) \\ \#(\text{edges clustered in } P \setminus gluer(P) \qquad\qquad\qquad) \end{cases}$$



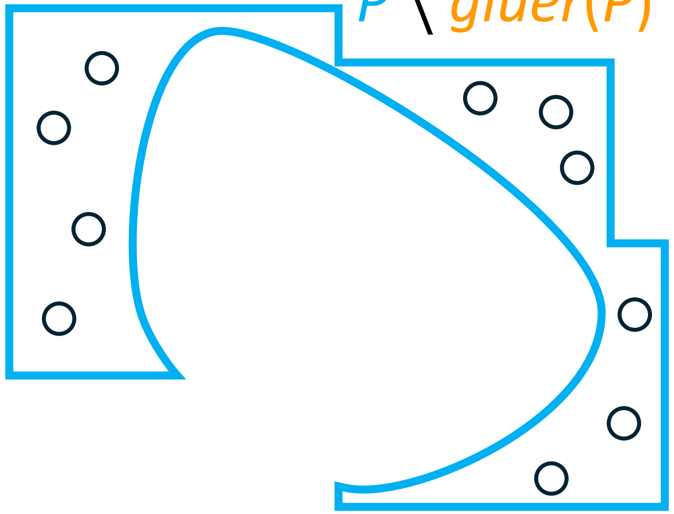
$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

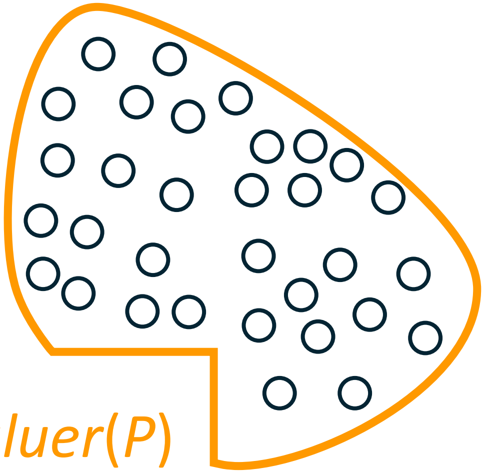
Analysis Overview

≤ ⊕ #(edges clustered in *P*, separated by *gluer*(*P*))

#(edges clustered in *P* \ *gluer*(*P*))

P \ *gluer*(*P*)





P ∩ *gluer*(*P*)

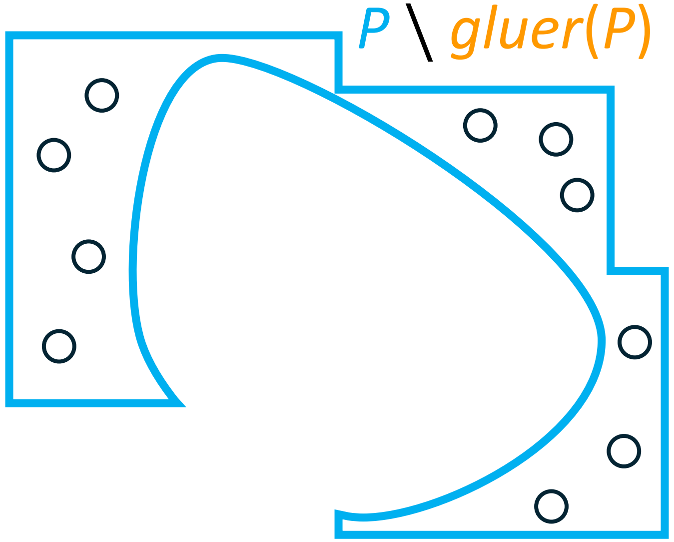
concentration property:
most points in *P* are in *P* ∩ *gluer*(*P*)

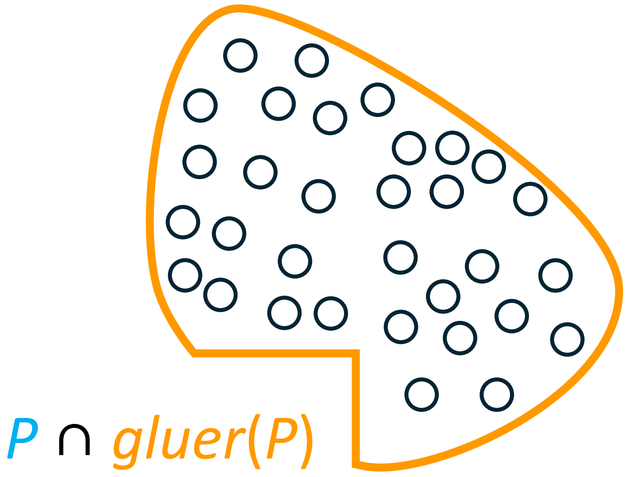
$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

$\leq$ $\oplus$ {
#(edges clustered in *P*, separated by *gluer*(*P*))
#(edges clustered in *P* \ *gluer*(*P*)) --- negligible

$\approx$ #(edges clustered in *P*, separated by *gluer*(*P*))





concentration property:
most points in *P* are in *P* ∩ *gluer*(*P*)

$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

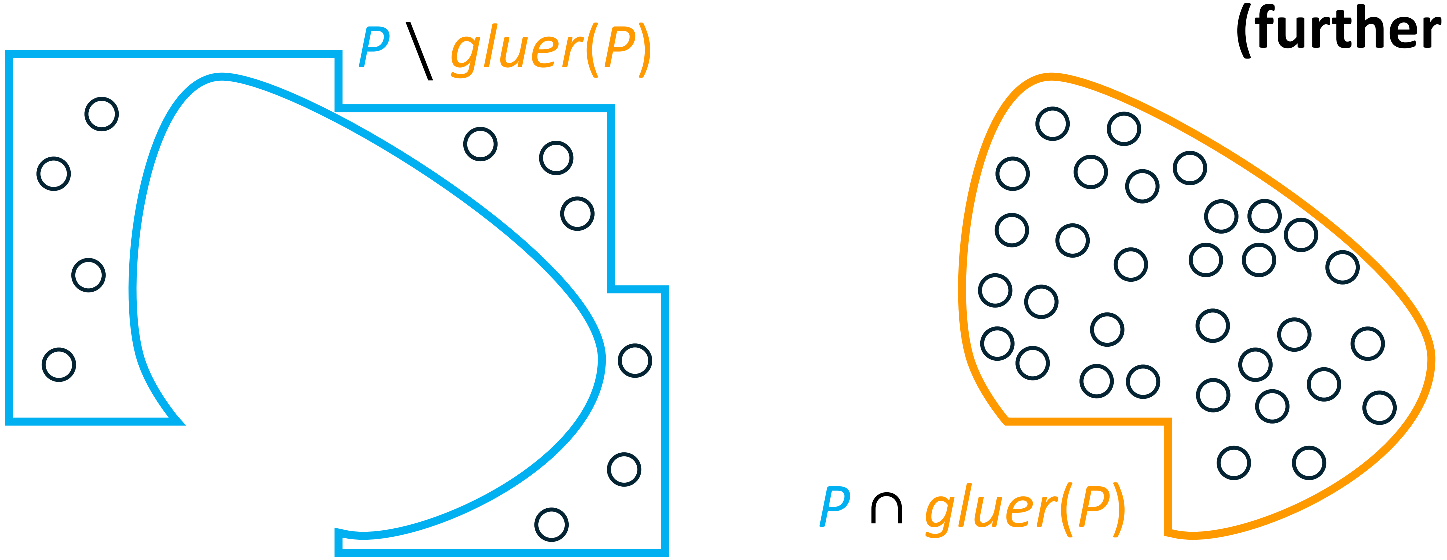
Analysis Overview

$\leq \oplus \begin{cases} \text{\#(edges clustered in } P\text{, separated by } gluer(P)\text{)} \\ \text{\#(edges clustered in } P \setminus gluer(P)\text{)} \quad \text{--- negligible} \end{cases}$

$\approx$ #(edges clustered in *P*, separated by *gluer*(*P*))

$\approx$ #(edges clustered in *P*, separated by *gluer*(*P*) & dist < 1)

(further exploiting concentration property)



$\tilde{x}_e^{(t)}$: *distance* of *e* (at layer *t*)

Analysis Overview

≤ ⊕ { #(edges clustered in *P*, separated by *gluer*(*P*))

#(edges clustered in *P* \ *gluer*(*P*)) --- negligible

≈ #(edges clustered in *P*, separated by *gluer*(*P*))

≈ #(edges clustered in *P*, separated by *gluer*(*P*) & dist < 1)

∴ #(– edges clustered in *P* & dist = 1)

≤ *O*(1)·#(edges clustered in *P*, separated by ***Q*** & dist < 1)

Conclusion

- **25.7846-approximation for HCC**
- Main ingredients
  - Useful lemma
  - Cut properties (of pre-clusterings)

- 5-approximation for L_0 Ultrametric Fitting
  - with the same ingredients

- Extension to other hierarchical clustering problems?

Thank You