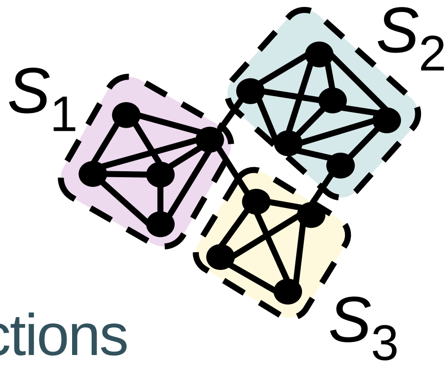
Minimum Average Cost Clustering

***Kiyohito Nagano (University of Tokyo)**
Yoshinobu Kawahara (Osaka University)
Satoru Iwata (Kyoto University)

For clustering problems with submodular objective functions, we introduce the minimum average cost criterion

The proposed algorithm

- does not require # of clusters in advance
- computes an optimal # of clusters and an optimal partition in polynomial time
- uses the theory of intersecting submodular functions



Keywords
clustering, submodular functions, combinatorial optimization

- $V = \{1, \ldots, n\}$ is a finite set of data points
- A function f defined on $2^V = \{S : S \subseteq V\}$ is **submodular** if

$$f(S) + f(T) \geq f(S \cup T) + f(S \cap T), \ \forall S, T \subseteq V$$

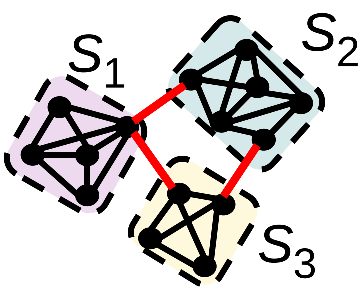
  - A generalization of cut functions, entropy functions, etc.

Clustering problem with submodular objective function

[Narasimhan-Jojic-Bilmes, NIPS 2005]

Given set V, integer $k (\leq n)$, and submodular function f,
find a k-partition of V, $\{S_1, \ldots, S_k\}$ that minimizes $\sum_{i=1}^{k} f(S_i)$

optimal k-clustering



optimal 3-clustering

In the case of a network, $\Sigma_i f(S_i) = 2 \times \#$(red edge)
(V is a set of nodes, and f is a cut function)

■ Optimal *k*-clustering problem [Narasimhan *et al.*, NIPS 2005]

min $\sum_{i=1}^{k} f(S_i)$
s. t. $\{S_1, \dots, S_k\}$ is a k-partition of V

- k (# of clusters) should be computed via some method
- NP-hard

■ Minimum Average Cost (MAC) clustering [This work]

min $\sum_{S \in \mathcal{P}} f(S) / (|\mathcal{P}| - \beta)$
s. t. $\mathcal{P}$ is a partition of V
$|\mathcal{P}| > \beta$

where $0 \leq \beta < n$

β-MAC clustering

averaged objective function

If $\beta = 1$, then the objective function is a natural average cost of $\mathcal{P}$

- $k = |\mathcal{P}|$ and a partition $\mathcal{P}$ are determined at the same time
- Solvable in poly time
- competitive with other methods

If β is small, $\mathcal{P}$ is coarse. If β is big, $\mathcal{P}$ is fine.

Theory of intersecting submodular functions

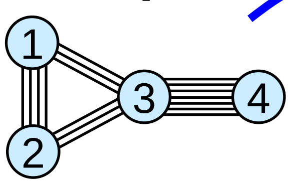
⇨ **Theorem** [This work]**.** There is an algorithm that computes all the β-MAC clusterings in polynomial time in total

Observation. Suppose that a partition $\mathcal{P}$ is a β-MAC clustering for some β, and let $k = |\mathcal{P}|$. Then, $\mathcal{P}$ is a k-optimal clustering

⇨ The information about MAC clusterings gives a portion of the information about optimal k-clusterings

(Remember that an optimal k-clustering problem is NP-hard)

Example



Compute all β-MAC clusterings

$0 \le \beta < 1$	$1 \le \beta < 11/7$	$11/7 \le \beta < 4$
{{1,2,3,4}}	{{1},{2},{3,4}}	{{1},{2},{3},{4}}
opt 1-clustering	opt 3-clustering	opt 4-clustering

In this case, our algorithm computes optimal k-clusterings for k = 1,3, & 4

For more information, please visit Poster T26