Bilu-Linial Stability

Grégoire Fournier

MSCS, MCS 549

March 3, 2025

Motivation

- Informally, an instance of a problem is Bilu-Linial stable if the optimal solution does not change when the instance is perturbed.
- Such instances yields more robust solutions than worst cases in many problems, and are closer to real word application.

Plan

1. Introduction, definition
2. Presentation of some results on Bilu Linial stability
3. *k*-center, α-approximation and stability
4. A clustering algorithm for α-resilient instances

Definition

Some studied applications for a perturbation α are:

- Graph partitioning, $G = (V, E, w)$ and $G' = (V, E, w)$ st. $\forall e \in E, w(e) \leq w'(e) \leq \alpha w(e)$.
- Clustering $J = (V, d)$ and $J' = (V, d')$ st. $\forall u, v \in V, d(u, v) \leq d'(u, v) \leq \alpha d(u, v)$.

An instance is α-stable if its every α perturbation does not change the optimal solution.

When studying LPs-SDPs, an instance is α-weak Bilu-Linial stable for an (α, ε) perturbation resilience.

Known Results

problem	main results	reference
Max Cut & 2-correlation clustering	$O(\sqrt{\log n} \log\log n)$ (incl. weakly stable instances) SDP gap and hardness result	Makarychev et al. (2014b)
Min Multiway Cut	4, (incl. weakly stable instances)	Makarychev et al. (2014b)
Max k-Cut	hardness for ∞-stable instances	Makarychev et al. (2014b)
sym./asym. k-center	2	Balcan et al. (2015)
	hardness for $(2-\varepsilon)$-pert. resil.	
s.c.b. objective	$1+\sqrt{2}$	Balcan and Liang (2016)
	$(2+\sqrt{3}, \varepsilon)$ for k-median	
	2, assuming cluster verifiability	Balcan et al. (2015)
s.c.b., Steiner points	$2+\sqrt{3}$	Awasthi et al. (2012)
min-sum objective	$O(\rho)$ and $(O(\rho), \varepsilon)$, where ρ is the ratio between the sizes of the largest and smallest clusters	Balcan and Liang (2016)
TSP	1.8	Mihalák et al. (2011)

Figure: Known results for Bilu-Linial stability [3], α-stable/perturbation resilient instances; it shows (α, ε) for (α, ε) -perturbation/resilient instances

Metric perturbation resilience for k-Center [1]

Motivation: Fast algorithm to solve 2-metric perturbation resilient instances of k-center.

Definition (k-Center)

Given vertices V, a metric d, in V define k-centers, $c_1, ...c_k$ which induce a clustering $C_1, ...C_k$ on V based on the d-nearest centers. Hence:

$$C_i = \{u : \forall j \neq i, d(u, c_i) \leq d(u, c_j)\}$$

$$cost = max_i max_{u \in C_i} d(u, c_i)$$

Metric perturbation resilience for k-Center [1]

Note that given a set of centers or a clustering, it is possible to efficiently find the corresponding clustering or the corresponding optimal set of centers.

Theorem (Stability and approximation for k-center)

Every α-approximation algorithm finds the optimal solutions of α-metric perturbation resilient instances.

Metric perturbation resilience for k-Center [1]

Proof

The goal is to show $C'_1, ..C'_k$ is an optimal clustering for a distance d' at the same cost. Then $C'_1, ..C'_k$ is an optimal clustering for d', by stability it is equal to $C_1, ..C_k$.

Consider $(C_i)_i$ the optimal clustering solution with cost r^* and $(C'_i)_i$ the approximation. By definition of an approximation $\forall i, \forall u \in C'_i, d(u, c'_i) \leq \alpha r^*$. Now define d':

$$\forall u, v \in V, d'(u,v) = \begin{cases} d(u,v)/\alpha & \text{if } d(u,v) \geq \alpha r^* \\ r^* & \text{if } d(u,v) \in [r^*, \alpha r^*] \\ d(u,v) & \text{if } d(u,v) \leq r^* \end{cases}$$

Metric perturbation resilience for k-Center [1]

So that d' defines a distance, it must satisfy the triangle inequality, so define f:

$$f(x) = \begin{cases} 1/\alpha & \text{if } x \geq \alpha r^* \\ r^*/x & \text{if } x \in [r^*, \alpha r^*] \\ 1 & \text{if } x \leq r^* \end{cases}$$

Metric perturbation resilience for k-Center [1]

Then $d' = (f \circ d)d$. Note that f is non increasing, $xf(x)$ non decreasing. Now for u, v, w, since $xf(x)$ non decreasing $f(d(u,w)) \geq min(f(d(u,v)), f(d(v,w)))$ and then:

$$d'(u,v) + d'(v,w) = ((f \circ d)d)(u,v) + ((f \circ d)d)(v,w)$$

$$d'(u,v) + d'(v,w) \geq (f \circ d)(u,w)(d(u,v) + d(v,w))$$

So by triangular inequality for d:

$$d'(u,v) + d'(v,w) \geq (f \circ d)(u,w)d(u,w) \geq d'(u,w)$$

Metric perturbation resilience for k-Center [1]

d' induces an α-perturbation as:

$$\forall u, v \in V, \frac{d'(u,v)}{d(u,v)} = f(d(u,v)) \in [1/\alpha, 1]$$

So $C_1, ..., C_k$ still the optimal clustering for d' by α resilience.
Denote $c_1", ..., c_k"$ the optimal centers for d' of $C_1, ..C_k$.

$$\text{Define } r(C_i) = min_{c \in C_i} max_{u \in C_i} d(u, c)$$

Fix i st $r(C_i) = r^*$. In particular:

$$\exists u, c \in C_i \text{ st } d(u,c) \geq r(C_i) = r^* \text{ and then } d(u, c_i") \geq r^*$$

The cost of $C_1, ...C_k$ for d' is at least r^*, cannot be more than r^* since $d' \leq d$ on V.
So the cost of the clustering $C_1, ..C_k$ for d' is r^*.

Metric perturbation resilience for k-Center [1]

The cost of $C'_1, ..C'_k$ is also r^*:
Denote $c'_1, ..., c'_k$ the optimal set of centers for d', then:

$$\forall i, \forall u \in C'_i, d(u, c'_i) \leq \alpha r^* \text{ and so } d'(u, c'_i) \leq r^*$$

Therefore $C'_1, ..C'_k$ is an optimal clustering for d', by stability it is equal to $C_1, ..C_k$.

Theorem (Stability and approximation for k-center)

Every α-approximation algorithm finds the optimal solutions of α-metric perturbation resilient instances.

A clustering algorithm for α-resilient instances

Definition (Separable center-based objectives)

A clustering problem has a center-based objective if:

- Given $S \subset V$ and a distance d_S on S, it is possible to find the optimal center or set of optimal centers (subset of S).
- The set of centers does not change when multiplying all distances in S by α.
- If $C_1, ..., C_k$ is an optimal clustering of V. Then $\forall i, \forall p \in C_i \ d(p, c_i) < d(p, c_j)$.

A clustering algorithm for α-resilient instances

Definition (Separable center-based objectives)

It is separable if:

- The cost of the clustering is either the maximum or sum of the cluster scores.
- The score $score(S, d|S)$ of each cluster S depends only on $(S, d|S$, and can be computed in poly time.

k-center, k-means have separable center-based objectives.

A clustering algorithm for α-resilient instances in $O(n^3)$ [2]

Theorem (The C_is appear in $\mathcal{T}$)

Consider a cluster C_i in the optimal clustering.

- *Let C be a cluster/node in the decomposition tree. Then:*
$$C \subset C_i, \; C_i \subset C \;\; or \;\; C \cap C_i = \emptyset$$
- *C_i appears in the decomposition tree $\mathcal{T}$.*

A clustering algorithm for α-resilient instances in $O(n^3)$

- Assign a cluster to each vertex. For $n-1$ steps, fusion the two nearest clusters. Assign a binary decomposition tree $\mathcal{T}$ to this process.
- Using a dynamic program algorithm, identify the best k $(C_i)_i$ in $\mathcal{T}$.

The distance used is called the closure distance

A clustering algorithm for α-resilient instances in $O(n^3)$

Definition (Closure distance D_S)

$$\forall A_1, A_2 \subset V,$$

$D_S(A_1, A_2)$ is the minimum r st. $A_1 \cup A_2$ has an r-central point

$x \in A$ is r-central for A if:

- $A \subset B(x, r)$
- if $d(p, q) \leq d(p, x) \leq r$ then $d(q, x) \leq r$

References and Questions

[1] Maria-Florina Balcan, Nika Haghtalab, and Colin White. Symmetric and asymmetric \$k\$-center clustering under stability. *CoRR*, abs/1505.03924, 2015.

[2] Maria-Florina Balcan and Yingyu Liang. Clustering under perturbation resilience. *CoRR*, abs/1112.0826, 2011.

[3] K. Makarychev. Y. Makarychev. *Bilu–Linial Stability*.