Streaming Robust Submodular Maximization: A Partitioned Thresholding Approach

Slobodan Mitrovic, Ilija Bogunovic, Ashkan Norouzi Fard, Jakub Tarnawski, Volkan Cevher

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Highlights

I. Submodular maximization in the **streaming setting** with **robustness** to **removals**.
II. Constant-factor guarantee for **any number** of **removals**.
III. Uses only **at most polylog** more space compared to the optimum.

Setup

Submodularity of a set function $f : 2^V \rightarrow \mathbb{R}_+$:

$$f(X \cup \{e\}) - f(x) \geq f(Y \cup \{e\}) - f(Y)$$

whenever $X \subseteq Y$ and $e \notin Y$.

Submodular maximization under cardinality constraint k:

$$OPT(X, k) := \underset{Y \subseteq X : |Y| = k}{\arg\max} f(Y)$$

The elements of V arrive in the **streaming** fashion in **arbitrary** order.

An arbitrary set $E \subseteq V$ is **removed** from V after the stream ends.

Robust monotone submodular maximization: find $S \subseteq V$ such that

1. $|S| \leq m$, and
2. $f(OPT(S\setminus E, k)) \geq c \cdot f(OPT(V\setminus E, k))$

for a parameter c and any E **not known a priori**.

Main result

- [MKK17]: Algorithm with c=0.499 and m= ***Õ(k \|E\|)***.
- [BMKK14]: If $E=\emptyset$, algorithm with c=0.499 and m= ***Õ(k)***.

Our result:
Algorithm with c=0.148 and m=***Õ(k + \|E\|)***.

Applications

Robust dominating set

Find a set of nodes S of size at most t that maximizes

$$|\mathcal{N}(OPT(S\setminus E, k)) \cup OPT(S\setminus E, k)|$$

$\mathcal{N}(A)$: denotes all the neighborhood nodes of nodes in A

Interactive personalized movie recommendation:

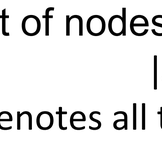
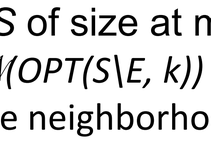
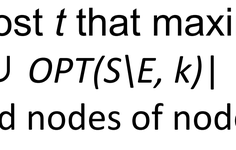
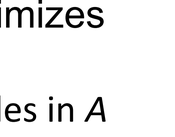
Given a feature vector u of a user and feature vectors v_z for a set of movies M, find a set of movies $S \subseteq M$ that maximizes

$$(1-\alpha) \sum_{z \in S} \langle u, v_z \rangle + \alpha \sum_{m \in M} \max_{z \in S} \langle v_m, v_z \rangle$$

The set of movies S is then used to provide, on the user's request, movies with specific properties (genre, year, not-seen, etc). The movies NOT falling in the category of the user's request are treated as the set E.

Setting

First phase: select S from the stream

$S = \{$  ,  ,  ,  $\}$

Second phase: given E, run `Greedy` to approximate $OPT(S\setminus E, k)$

$E = \{$  ,  ,  $\}$

$OPT(\{$  ,  ,  $\}, k)$

[first name].[last name]@epfl.ch

STAR-T Algorithm

Structure of the robust set ***S***:
- # of **partitions**: $\log k$
- partition i contains: $(k / 2^i)$ **buckets** with at most 2^i **items**

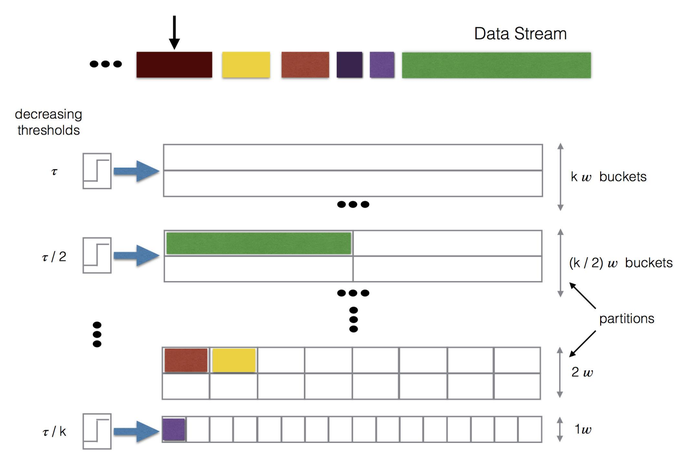
Streaming Algorithm to create ***S***:
- add the item e to the first non-full bucket B if

$$f(e|B) \geq \tau / 2^i,$$

i : partition of the bucket B
τ : threshold value that depends on $f(OPT(V\setminus E, k))$
($f(OPT(V\setminus E, k))$ is approximated by extending the techniques of [BMKK14])

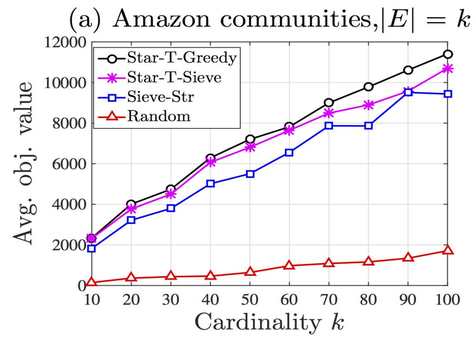
- if such bucket does not exist discard the item.

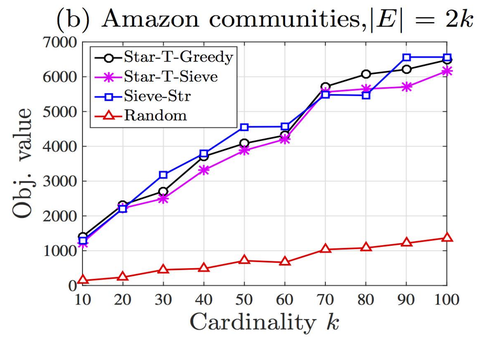
Query Stage:
- for a given ***E*** return ***k*** elements greedily selected from ***S \ E***

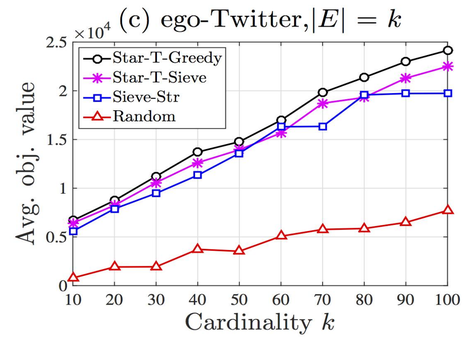


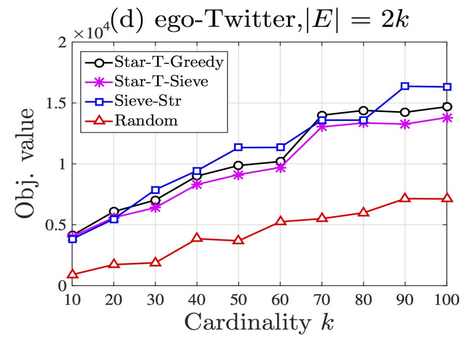
Numerical results

Robust dominating set:

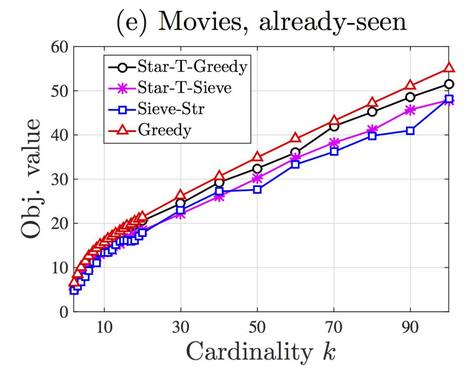


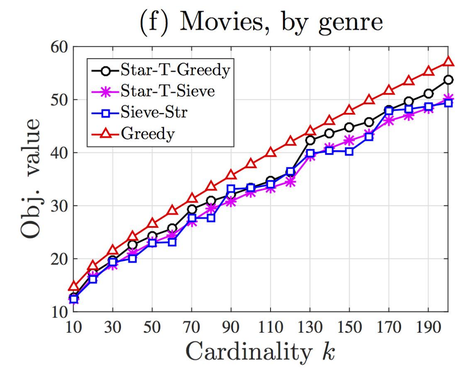






Interactive personalized movie recommendation:





References

[MKK17] B. Mirzasoleiman, A. Karbasi, A. Krause, *Deletion-Robust Submodular Maximization: Data Summarization with "the Right to be Forgotten"*, ICML 2017

[BMKK14] A. Badanidiyuru, B. Mirzasoleiman, A. Karbasi, A.Krause, *Streaming submodular maximization: Massive data summarization on the fly*, SIGKDD 2014