EXPLORING THE SUBGRAPH DENSITY-SIZE TRADE-OFF VIA THE LOVASZ EXTENSION

Aritra Konar and Nicholas D. Sidiropoulos

University of Virginia, Charlottesville, USA

Introduction:

- **Densest-k-subgraph problem (DkS):** Given an undirected graph, find subgraph of size k with the max. number of induced edges
- **Formulation:** Let $\mathbf{A} \in \mathbb{R}_{+}^{n \times n}$ denote the (possibly weighted) adjacency matrix of an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

$$\boxed{\begin{aligned} &\max_{\mathbf{x} \in \{0,1\}^n} \left\{ f(\mathbf{x}) := \mathbf{x}^T \mathbf{A} \mathbf{x} \right\} \\ &\text{s.t.} \quad \mathbf{1}^T \mathbf{x} = k \end{aligned}}$$

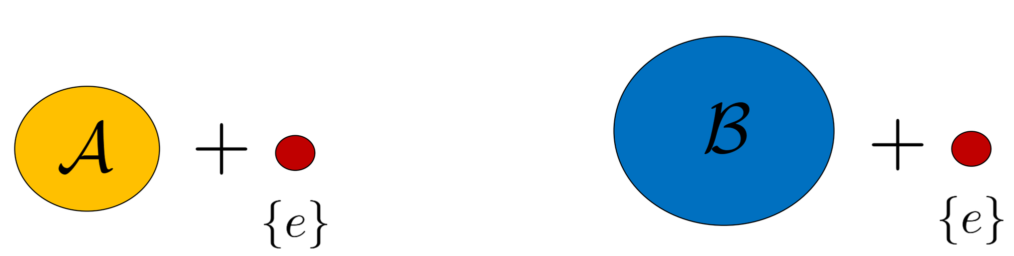
- **NP-hard, difficult to approximate**
- **Our contributions**: We propose a new convex relaxation for DkS based on the Lovasz extension of submodular functions and devise an ADMM algorithm for solving the problem at scale.

Prior Art:

- **State-of-the-art**: [Bhaskara *et al.*, 2010]. Provides $O(n^{1/4+\epsilon})$ approximation guarantee in time $n^{O(1/\epsilon)}$
- **Greedy Algorithm:** [Feige *et al.* 2001]. Provides $O(n/k)$ guarantee
- **Semidefinite relaxation:** [Srivastav-Wolf 1998, Feige-Langberg 2001, Bombina-Ames 2020]. Computationally expensive.
- **Low-rank Matrix approximation:** [Papailiopoulos *et al.* 2014]. Solve DkS with low (constant) rank approximation of adjacency matrix, provides data-dependent quality guarantees.

Submodularity:

- Given a set of elements $\mathcal{V} = [n] := \{1, \cdots, n\}$, consider a set-function $f : 2^{\mathcal{V}} \to \mathbb{R}$
- Submodular if for all $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V} \setminus \{e\}$,

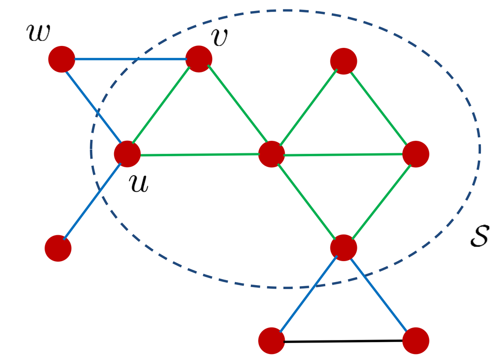


$$F(\mathcal{A} \cup \{e\}) - F(\mathcal{A}) \;\; \geq \;\; F(\mathcal{B} \cup \{e\}) - F(\mathcal{B})$$

i.e., diminishing returns

- **Link to convexity**: every submodular function possesses a continuous, convex extension (the Lovasz extension) [Lovasz 1983]. **Can be viewed as the tightest convex under-estimator of f (in a certain sense).**

Main idea:



$$\text{vol}(\mathcal{S}) = \sum_{u \in \mathcal{S}} d_u \quad \text{where } d_u = \text{degree of vertex } u$$

$$= \underbrace{2e_2(\mathcal{S})}_{\text{\# edges with 2 endpoints in } \mathcal{S}} + \underbrace{e_1(\mathcal{S})}_{\text{\# edges with 1 endpoint in } \mathcal{S}}$$

- Re-arranging: $2e_2(\mathcal{S}) = \text{vol}(\mathcal{S}) - e_1(\mathcal{S}), \forall\, \mathcal{S} \subseteq \mathcal{V}$

Equivalently

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{d}^T \mathbf{x} - \sum_{(u,v) \in \mathcal{E}} |x_u - x_v|, \forall\, \mathbf{x} \in \{0,1\}^n$$

- Equivalent reformulation of DkS:

$$\boxed{\begin{aligned} &\max_{\mathbf{x} \in \{0,1\}^n} \; \mathbf{d}^T \mathbf{x} - \sum_{(u,v) \in \mathcal{E}} |x_u - x_v| \\ &\text{s.t.} \quad \mathbf{1}^T \mathbf{x} = k \end{aligned}}$$

- Relax discrete constraints to obtain convex problem

$$\boxed{\begin{aligned} &\min_{\mathbf{x} \in [0,1]^n} \; -\mathbf{d}^T \mathbf{x} + \sum_{(u,v) \in \mathcal{E}} |x_u - x_v| \\ &\text{s.t.} \quad \mathbf{1}^T \mathbf{x} = k \end{aligned}}$$

- **Main claims:**

(a) Cost function $-f$ of DkS is submodular

(b) Cost function of relaxed problem is the Lovasz extension of $-f$

Algorithm:

- Express relaxed problem as: $\min\limits_{\mathbf{B}^T \mathbf{x} = \mathbf{z}} \left\{ g(\mathbf{x}) + h(\mathbf{z}) \right\}$

where $\boxed{g(\mathbf{x}) = \begin{cases} -\mathbf{d}^T \mathbf{x}, & \mathbf{x} \in [0,1]^n, \mathbf{1}^T \mathbf{x} = k \\ +\infty, & \text{o.w.} \end{cases}}$ $\boxed{h(\mathbf{z}) = \sum_{e=1}^{m} |z_e| = \|\mathbf{z}\|_1}$

- Apply Linearized ADMM [Condat 2014]
- **L-ADMM updates:**

Require computing proximal operators of a convex function

$$\begin{aligned} \mathbf{x}^{k+1} &= \text{prox}_{g/\mu}((\mathbf{I}_n - \mu\rho\mathbf{L})\mathbf{x}^k - \mu\rho\mathbf{B}(\mathbf{z}^k - \mathbf{u}^k)) \\ \mathbf{z}^{k+1} &= \text{prox}_{\rho h}(\mathbf{B}^T \mathbf{x}^{k+1} + \mathbf{u}^k) \\ \mathbf{u}^{k+1} &= \mathbf{u}^k + \mathbf{B}^T \mathbf{x}^{k+1} - \mathbf{z}^{k+1} \end{aligned}$$

- **x-update:** Bisection search
- **z-update:** Soft-thresholding operator
- **Convergence rate:** Same as standard ADMM! $O(1/k)$

Experimental Setup:

- Use L-ADMM to solve L-relaxation; final solution obtained by
  (i) Projection onto discrete set
  (ii) Use Frank-Wolfe algorithm [Frank-Wolfe 1950] for last mile refinement

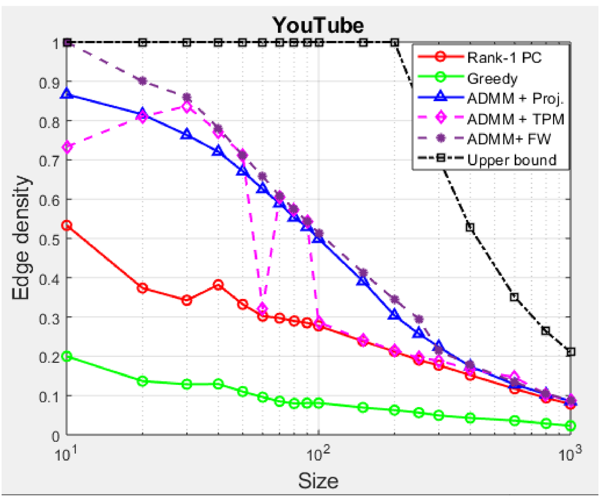
- **Baselines:**
  (i) Greedy algorithm [Feige *et al.* 2001]
  (ii) Truncated Power Method [Yuan-Zhang 2013]
  (iii) Rank-1 binary matrix principal component (larger ranks incur higher complexity) [Papailiopoulos *et al.* 2014].
  (iv) Edge-density upper-bound

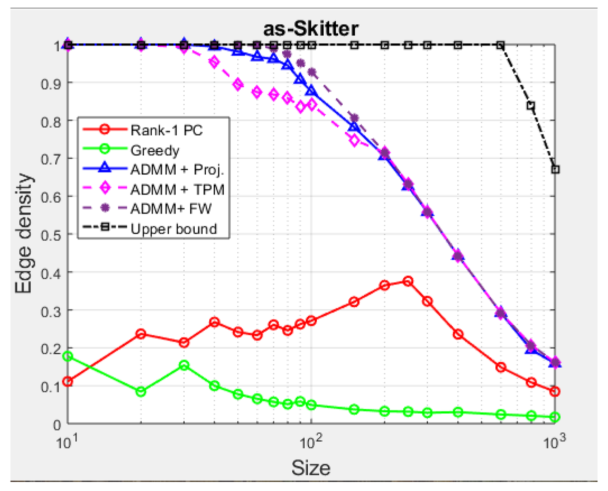
- **Performance metric:**
  Edge-density = (# of induced edges)/(# edges in k clique)

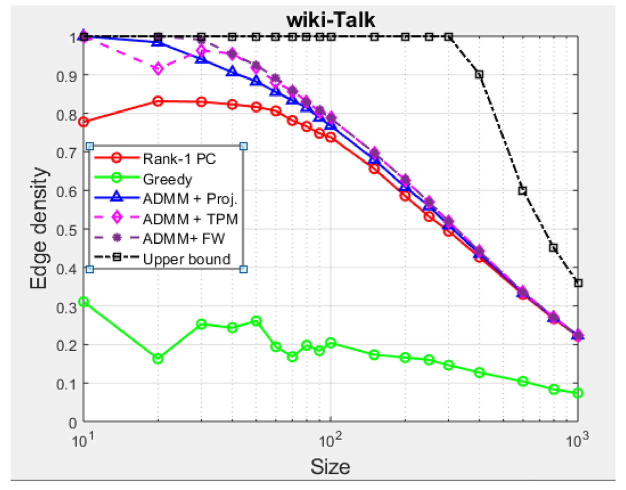
Results:

Datasets from SNAP [Krevl-Leskovec 2015]

Graph	n	m	Network Type
polBlog	1,224	16,714	Social
Facebook	4,039	88,234	Social
ppi-Human	21,557	342K	Biological
loc-Gowalla	196K	950K	Social
web-Google	875K	5.10M	Web
YouTube	1.1M	2.9M	Social
as-Skitter	1.7M	12M	Autonomous Systems
wiki-Talk	2.4M	5M	Communications







Our approach attains state-of-the-art performance

Conclusions:

- Proposed a new convex relaxation for DkS and a scalable ADMM algorithm for solving it.
- Attains state-of-the-art performance on real graphs.
- Future work: data-dependent performance guarantees, extension to higher-order motifs.