A ADDITIONAL PROOFS OF LEMMAS AND THEOREMS

THEOREM 3.1. Given a graph $G = (V, E)$, and two vertex sets $R, A \subseteq V$. In FDP, when

$$T > O\left(\frac{\Delta(G)(\mathrm{vol(R)} + |\mathrm{A}|\Delta(\mathrm{G})^2)}{\epsilon^2}\right),$$

the following convergence guarantee holds: $\|\boldsymbol{r} + \boldsymbol{\beta}\|_\infty - \rho_R^* \le \epsilon$, where vol(R) is the sum of degrees in G for all vertices in the set R, and $\Delta(G)$ denotes the maximum degree of graph G.

PROOF. In the t-th iteration, for simplify, we let $\boldsymbol{z}^{(t)} = \boldsymbol{r}^{(t)} + \boldsymbol{\beta}^{(t)}$, and $\boldsymbol{s}^{(t)} = \boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}$, we have:

$$\frac{\boldsymbol{z}^{(t)}}{t} = (1 - \frac{1}{t}) \cdot \frac{\boldsymbol{z}^{(t-1)}}{t-1} + \frac{1}{t} \cdot \boldsymbol{s}^{(t-1)}$$

Clearly, this update can be interpreted as a standard Frank-Wolfe step with a learning rate of $\gamma_t = \frac{1}{t}$. Now, consider the function $f(z) = (\|z\|)^2$, clearly, the $\min_{z \in C} f(z)$ is exactly the optimal solution of $\mathrm{QP}(G, A)$, where C denotes the set of all feasible vectors z for $\mathrm{QP}(G, A)$.

Then, we can introduce the following theorem to analyze the convergence rate of Algorithm 1.

THEOREM A.1 ([35]). *Let z^* denote the optimal solution of $\min_{z\in C} f(z)$. For each $t \ge 1$ in Algorithm 1, the iterates of $x^{(t)}$ satisfies:*

$$f(z^{(t)}) - f(z^*) \le \gamma_t \cdot \xi \cdot (1 + \delta),$$

where ξ is the curvature constant of $f(z)$, and δ is the error in the linear minimization step.

Now, we analyze the curvature constant ξ and the accuracy δ of the objective function:

$$\min f(z) = \sum_{u \in V} (r(u) + \beta(u))^2 = \|z\|^2$$

We can easily derive $\delta = 0$, and the upper bound of ξ can be calculated by the following theorem.

THEOREM A.2. *The curvature constant ξ of $f(z) = \sum_{u\in V}(r(u) + \beta(u))^2$ satisfies $\xi \le 8(\Delta(G) + 1) \cdot (|A|\Delta(G)^2 + \mathrm{vol(R)})$.*

PROOF. Denote C by the set of x of all feasible solutions for $\mathrm{QP}(G, A)$. By the definition of the curvature constant, we have

$$\xi = \sup_{\boldsymbol{x_1}, \boldsymbol{y} \in C, \gamma \in [0,1], \boldsymbol{x_2} = \boldsymbol{x_1} + \gamma(\boldsymbol{y} - \boldsymbol{x_1})} \frac{2}{\gamma^2}(f(\boldsymbol{x_2}) - f(\boldsymbol{x_1}) - \nabla f(\boldsymbol{x_1})^T(\boldsymbol{x_2} - \boldsymbol{x_1}))$$

The gradient of $f(x)$ is $\nabla f(x)_{u,v} = 2x_u$. Therefore,

$$\begin{aligned}
& f(\boldsymbol{x_2}) - f(\boldsymbol{x_1}) - \nabla f(\boldsymbol{x_1})^T(\boldsymbol{x_2} - \boldsymbol{x_1}) \\
=& \sum_{u \in V} (x_2(u)^2 - x_1(u)^2) - \sum_{u \in V} 2x_1(u)(x_2(u) - x_1(u)) \\
=& \sum_{u \in V} (x_1(u) - x_2(u))^2 \\
=& \|\boldsymbol{x_1} - \boldsymbol{x_2}\|^2 \\
=& \gamma^2 \|\boldsymbol{y} - \boldsymbol{x_1}\|^2 \\
\le& \gamma^2 C^2
\end{aligned}$$

where $\mathfrak{R}$ is the diameter of C:

$$\begin{aligned}
\mathfrak{R}^2 =& \sup_{\boldsymbol{x} \in C, \boldsymbol{x} = \boldsymbol{r} + \boldsymbol{\beta}} \sum_{u \in V} (r(u) + \beta(u))^2 \\
\le& \sup_{\boldsymbol{x} \in C, \boldsymbol{x} = \boldsymbol{r} + \boldsymbol{\beta}} \sum_{u \in V} \left(\sum_{v \in N(u)} \alpha_{u,v} + \beta(u) \right)^2 \\
=& \sup_{\boldsymbol{x} \in C, \boldsymbol{x} = \boldsymbol{r} + \boldsymbol{\beta}} \sum_{u \in V} (d(u) + 1) \left(\beta(u)^2 + \sum_{v \in N(u)} (\alpha_{u,v})^2 \right) \\
\le& (\Delta(G) + 1) \sup_{\boldsymbol{x} \in C, \boldsymbol{x} = \boldsymbol{r} + \boldsymbol{\beta}} \sum_{u \in V} \left(\beta(u)^2 + \sum_{v \in N(u)} (\alpha_{u,v})^2 \right) \\
=& (\Delta(G) + 1) \sup_{\boldsymbol{x} \in C, \boldsymbol{x} = \boldsymbol{r} + \boldsymbol{\beta}} \left(\sum_{u \in A} \beta(u)^2 + \sum_{u \in V} \sum_{v \in N(u)} \alpha_{u,v}^2 \right) \\
\le& (\Delta(G) + 1)(4|A|\Delta(G)^2 + 4\mathrm{vol(R)})
\end{aligned}$$

That is, $\xi \le 8(|A|\Delta(G)^3 + \mathrm{vol(R)}\Delta(\mathrm{G}))$ □

Now, at the T-th iteration, for the objective function $f(z)$, we have achieved the following bound:

$$\begin{aligned}
f(z^{(T)}) - f(z^*) \le& \gamma_t \cdot \xi \\
\le& \frac{1}{T} 8(|A|\Delta(G)^3 + \mathrm{vol(R)}\Delta(\mathrm{G}))
\end{aligned}$$

To ensure $f(z^{(T)}) - f(z^*) \le \epsilon$, we need to show:

$$\frac{8(|A|\Delta(G)^3 + \mathrm{vol(R)}\Delta(\mathrm{G}))}{T} \le \epsilon^2$$

In other words,

$$T \ge \frac{8(|A|\Delta(G)^3 + \mathrm{vol(R)}\Delta(\mathrm{G}))}{\epsilon^2}$$

Absorbing constant factors into the asymptotic notation, this yields the final complexity bound:

$$T = O\left(\frac{|A|\Delta(G)^3 + \mathrm{vol(R)}\Delta(\mathrm{G})}{\epsilon^2} \right)$$

This completes the proof. □

THEOREM 4.1. Given a graph $G = (V, E)$ and two vertex sets $R, A \subseteq V$, we have $\mathrm{OPT}(\mathrm{QP}(G, A)) = \mathrm{OPT}(\mathrm{QP}'(G, A))$.

PROOF. Let

$$g(\mathbf{r}, \boldsymbol{\beta}) = \sum_{u \in V} \big(r(u) + \beta(u)\big)^2, \; z(\mathbf{r}) = |A| \Big(\max_{v \in V} r(v) \Big)^2 + \sum_{u \in V \setminus A} r(u)^2,$$

denote the objective functions of $\mathrm{QP}(G, A)$ and $\mathrm{QP}'(G, A)$, respectively. We prove the theorem in two steps.

(i) $\mathtt{OPT}(\mathrm{QP}(G,A)) \geq \mathtt{OPT}(\mathrm{QP}'(G,A))$. The program $\mathrm{QP}'(G,A)$ can be considered as a relaxation of $\mathrm{QP}(G,A)$ that removes only the constraints related to $\boldsymbol{\beta}$. Therefore, any feasible triple $(\mathbf{r},\boldsymbol{\beta},\boldsymbol{\alpha})$ of $\mathrm{QP}(G,A)$ implies that the pair $(\mathbf{r},\boldsymbol{\alpha})$ is a feasible solution to $\mathrm{QP}'(G,A)$. Moreover,

$$\begin{aligned} g(\mathbf{r},\boldsymbol{\beta}) &= \sum_{u\in V}\big(r(u)+\beta(u)\big)^2 \\ &\geq |A|\cdot\Big(\max_{v\in V}(r(v)+\beta(v))\Big)^2+\sum_{u\in V\setminus A}\big(r(u)+\beta(u)\big)^2 \\ &\geq z(\mathbf{r}). \end{aligned}$$

(ii) $\mathtt{OPT}(\mathrm{QP}(G,A)) \leq \mathtt{OPT}(\mathrm{QP}'(\mathrm{G},\mathrm{A}))$. Let $(\mathbf{r}^*,\boldsymbol{\alpha}^*)$ be an optimal pair for $\mathrm{QP}'(G,A)$ and set $\tau=\max_{v\in V} r^*(v)$. Define

$$\beta^*(u)=\begin{cases}\tau-r^*(u), & u\in A,\\ 0, & u\in V\setminus A.\end{cases}$$

Then $r^*(u)+\beta^*(u)=\tau$ for all $u\in A$ and $r^*(v)+\beta^*(v)\leq\tau$ for $v\notin A$; hence $(\mathbf{r}^*,\boldsymbol{\beta}^*,\boldsymbol{\alpha}^*)$ is a feasible solution for $\mathrm{QP}(G,A)$. Its objective value is

$$\begin{aligned} g(\mathbf{r}^*,\boldsymbol{\beta}^*) &= \sum_{u\in V}\big(r^*(u)+\beta^*(u)\big)^2 \\ &= |A|\,\tau^2+\sum_{u\in V\setminus A} r^*(u)^2 \;=\; z(\mathbf{r}^*) \;\leq\; \mathtt{OPT}(\mathrm{QP}'(G,A)). \end{aligned}$$

Combining (i) and (ii), we obtain $\mathtt{OPT}(\mathrm{QP}(G,A))=\mathtt{OPT}(\mathrm{QP}'(G,A))$, as claimed. □

THEOREM 4.3. The Lipschitz constant L_f of $\nabla f(\alpha)$, is bounded by $2\cdot(|A|+1)\cdot\sqrt{\Delta(\mathrm{G})\cdot\mathrm{vol}(\mathrm{R})}$.

$$L_f^2=\|\nabla f(\alpha)\|^2=\sum_{(u,v)\in E}\sum_{(x,y)\in E}\left(\frac{\partial\left(\nabla f(\alpha)_{u,v}\right)}{\partial\alpha_{x,y}}\right)^2$$

We have:

$$\frac{\partial\left(\nabla f(\alpha)_{u,v}\right)}{\partial\alpha_{x,y}}=\frac{\partial\left(2\cdot(\mathbb{1}[u\notin A]+g_{u,v})\cdot r(u)\right)}{\partial\alpha_{x,y}}\leq\begin{cases}2\cdot(|A|+1), & u=x\\ 0, & u\neq x\end{cases}$$

Therefore,

$$\begin{aligned} L_f^2 &\leq \sum_{u\in V}\sum_{(u,v)\in E}\sum_{(u,y)\in E}(2\cdot(|A|+1))^2 \\ &\leq \sum_{u\in V}\sum_{(u,v)\in E}4(|A|+1)^2\cdot d(u) \\ &\leq 4(|A|+1)^2\Delta(G)\sum_{u\in V}\sum_{(u,v)\in E}1 \\ &\leq 4(|A|+1)^2\Delta(G)\cdot\mathrm{vol}(R) \end{aligned}$$

Therefore, $L_f\leq 2(|A|+1)\sqrt{\Delta(G)\cdot\mathrm{vol}(R)}$.

THEOREM 4.4. Given a graph $G=(V,E)$, and two vertex sets $R,A\subseteq V$. In PGD-ADS, when

$$T>O\left(\frac{\sqrt{|A|}\cdot\mathrm{vol}(R)^{\frac{1}{4}}\cdot\Delta(G)^{\frac{3}{4}}\cdot\sqrt{(\mathrm{vol}(R)+|A|\Delta(G))}}{\epsilon}\right),$$

the following convergence guarantee holds: $\|\boldsymbol{\lambda}\|_\infty-\rho_R^*\leq\epsilon$.

PROOF. We first provide our proof roadmap: (1) we prove that if $f(\boldsymbol{\alpha})-f(\boldsymbol{\alpha}^*)\leq\mu$, and we have $\|\boldsymbol{\lambda}\|_\infty-\|\boldsymbol{\lambda}^*\|_\infty\leq\sqrt{\mu}$, and (2) Let $\mu:=\frac{2\cdot\|x^{(0)}-x^*\|^2}{\eta\cdot T^2}=\epsilon^2$. Then, by Theorem 4.2, we establish the relationship between the target accuracy ϵ and the number of iterations T, thus completing the proof.

Firstly, $f(\boldsymbol{\alpha})-f(\boldsymbol{\alpha}^*)$ can be rewritten as:

$$\|\boldsymbol{\lambda}\|^2-\|\boldsymbol{\lambda}^*\|^2.$$

For each $u\in V$, let $\delta(u)=\lambda(u)-\lambda^*(u)$, we have:

$$\begin{aligned} \|\boldsymbol{\lambda}\|^2-\|\boldsymbol{\lambda}^*\|^2 &= \sum_{u\in V}\lambda(u)^2-\lambda^*(u)^2 \\ &= \sum_{u\in V}(2\lambda^*(u)\cdot\delta(u)+\delta(u)^2)\leq\mu. \end{aligned}$$

Afterwards, we aim to prove that:

$$\sum_{u\in V}(2\lambda^*(u)\cdot\delta(u))\geq 0,$$

thus $\sum_{u\in V}\delta(u)^2\leq\mu$, in other words, $(\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^*\|)^2\leq\mu$.

To achieve this goal, we need to establish two results: (1) show that $\lambda^*(u)$ is identical for all $u\in V$; (2) prove that $\sum_{u\in V}\delta(u)\geq 0$.

(1) For the first item, by Fujishige's theorem [27], when the feasible region is the base contrapolymatroid of a supermodular function, the optimal solution to the ℓ_2-norm minimization assigns identical values to all coordinates. A set function $g:2^{\mathcal{U}}\to\mathbb{R}$ is said to be *supermodular* if for any subsets $X,Y\subseteq\mathcal{U}$, it satisfies: $g(X)+g(Y)\leq g(X\cup Y)+g(X\cap Y)$.

Let $g:2^{\mathcal{U}}\to\mathbb{R}^+$ be a supermodular function, and $\vec{x}\in\mathbb{R}^{\mathcal{U}}$ be a nonnegative vector. The *base contrapolymatroid* defined by g is:

$$B_g=\left\{\vec{x}\in\mathbb{R}^{|\mathcal{U}|}\,\middle|\,\vec{x}\geq 0,\ \vec{x}(S)\geq g(S),\ \forall S\subseteq\mathcal{U},\ x(\mathcal{U})=g(\mathcal{U})\right\},$$

where $\forall S\subseteq\mathcal{U},\vec{x}(S)=\sum_{u\in S}x(u)$. Intuitively, the base contrapolymatroid B_g can be viewed as the feasible region for a $|\mathcal{U}|$-dimensional vector under a set of lower bounds and a single linear equality constraint.

Clearly, $\forall S\subseteq V$, let $f(S)=2|E(S)|-\sum_{u\in S\setminus R}|N(u,G[S])|$, f is a supermodular function, so that the constraints of $\mathrm{QP}'(G,A)$ can be compactly written as:

$$\lambda(S)\geq f(S),\quad\forall S\subseteq V,\quad\lambda(V)=f(V),$$

where $\lambda(S)=\sum_{u\in S}\lambda(u)$.

The *base contrapolymatroid* of $f(S)$ is:

$$B_f=\left\{\boldsymbol{\lambda}\in\mathbb{R}^n\,\middle|\,\boldsymbol{\lambda}\geq 0,\ \lambda(S)\geq f(S),\ \forall S\subseteq V\right\},$$

where $\lambda(V)=f(V)$.

Based on the B_f, and vector $\boldsymbol{\lambda}$ we can rewrite the $\mathrm{QP}'(G,A)$:

$$\mathrm{QP}'(G,A):\qquad \min_{\boldsymbol{\lambda}\in B_f}\sum_{u\in V}\lambda(u)^2.$$

Thus, by Fujishige's theorem [27], the optimal solution of $\mathrm{QP}'(G,A)$ is the lexicographically minimal base $\boldsymbol{\lambda}^*\in B_f$. That is, it follows that $\lambda^*(u)$ is identical for all $u\in V$.

(2) As $\boldsymbol{\lambda}\in B_f$ and $\lambda(V)=f(V)$, we have $\sum_{u\in V}\delta(u)=\lambda(V)-\lambda^*(V)\geq 0$.

By combining (1) and (2), we include $(\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^*\|)^2\leq\mu$. Let μ=ϵ^2, we have $\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^*\|\leq\epsilon$, on the other hands, we also have:

$$\|\boldsymbol{\lambda}\|_\infty-\|\boldsymbol{\lambda}^*\|_\infty\leq\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^*\|_\infty\leq\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^*\|\leq\epsilon.$$

In other words, if we have $f(\boldsymbol{\alpha}) - f(\boldsymbol{\alpha}^*) \le \epsilon^2$, we can finish the proof. Based on the Theorem 4.2, we have:

$$f(\boldsymbol{\alpha}) - f(\boldsymbol{\alpha}^*) \le \frac{2 \cdot \|\boldsymbol{\lambda}^{(0)} - \boldsymbol{\lambda}^*\|^2}{\eta \cdot T^2}$$

The gap of $\|\boldsymbol{\lambda}^{(0)} - \boldsymbol{\lambda}^*\|^2$ is graduated by the following:
As $\|\boldsymbol{\lambda}^{(0)} - \boldsymbol{\lambda}^*\| \le \|\boldsymbol{\lambda}^{(0)}\|$, we consider:

$$\|\boldsymbol{\lambda}^{(0)}\|^2 = \sum_{u \in V} \lambda(u)^2 = |A| \cdot (\max_{v \in V} r(v))^2 + \sum_{u \in V} \left(\sum_{(u,v) \in E \wedge (u,v) \notin A} \alpha_{u,v} \right)^2$$

$$\le |A| \cdot 4 \cdot \Delta(G)^2 + \sum_{u \in V} d(u,G) \left(\sum_{(u,v) \in E} \alpha_{u,v}^2 \right)$$

$$= |A| \cdot 4 \cdot \Delta(G)^2 + \Delta(G) \cdot \left(\sum_{(u,v) \in E \wedge (u<v)} \alpha_{u,v}^2 + \alpha_{v,u}^2 \right)$$

$$\le |A| \cdot 4 \cdot \Delta(G)^2 + \Delta(G) \cdot \left(\sum_{(u,v) \in E} w_{u,v}^2 \right)$$

$$\le 4 \cdot |A| \cdot \Delta(G)^2 + 4\Delta(G) \cdot \mathrm{vol}(R)$$

Therefore, $\|\boldsymbol{\lambda}^{(0)} - \boldsymbol{\lambda}^*\| \le 2\sqrt{|A| \cdot \Delta(G)^2 + \mathrm{vol}(R) \cdot \Delta(G)}$.
To ensure $f(\boldsymbol{\alpha}) - f(\boldsymbol{\alpha}^*) \le \epsilon^2$, we need to show:

$$\frac{2 \cdot (4 \cdot |A| \cdot \Delta(G)^2 + 4 \cdot \mathrm{vol}(R) \cdot \Delta(G))}{\frac{1}{2 \cdot (|A|+1) \cdot \sqrt{\Delta(\mathrm{G}) \cdot \mathrm{vol}(\mathrm{R})}} \cdot T^2} \le \epsilon^2$$

In other words,

$$T^2 \ge \frac{16 \cdot (\mathrm{vol}(R) \cdot \Delta(G) + |A| \cdot \Delta(G)^2) \cdot (|A|+1) \cdot \sqrt{\Delta(\mathrm{G}) \cdot \mathrm{vol}(\mathrm{R})}}{\epsilon^2}$$

Absorbing constant factors into the asymptotic notation, this yields the final complexity bound:

$$T = O\left(\frac{\sqrt{|A|} \cdot \mathrm{vol}(R)^{\frac{1}{4}} \cdot \Delta(G)^{\frac{3}{4}} \cdot \sqrt{(\mathrm{vol}(R) + |A|\Delta(G))}}{\epsilon} \right)$$

This completes the proof. ◻

Theorem 5.3 Given a graph $G = (V, E)$, two vertex sets R and A, and the $\lceil \rho_R^* \rceil$-$r$core, $\mathcal{M}^*$, the ADS $\mathcal{D} = G[S^*]$ must be contained within $G[V(\mathcal{M}^*) \cup A]$, where ρ_R^* is the density of $\mathcal{D}$.

Proof. We prove this by contradiction. Suppose that the ADS $\mathcal{D}$=$G[S^*]$ is not fully contained in $G[V(\mathcal{H}^*) \cup A]$. Then there exists a vertex $u \in S^* \wedge u \notin A$ such that:

$$d_R(u, G[S^*]) < \rho_R^*.$$

Consider removing u from S^*, and let $S' = S^* \setminus \{u\}$. Since $\mathcal{D}$ is an ADS, it holds that $\rho_R(G[S']) \le \rho_R^*$. On the other hand, we can compute $\rho_R(G[S'])$ as:

$$\rho_R(G[S']) = \frac{2|E(S')|}{|S'|} = \frac{2|E(S^*)| - 2d_R(u, G[S^*])}{|S^*| - 1},$$

and since $d_R(u, G[S^*])$ counts the number of R-neighbors of u in $G[S^*]$, we further have:

$$2|E(S')| = \rho_R^* \cdot |S^*| - d_R(u, G[S^*]).$$

Therefore,

$$\rho_R(G[S']) = \frac{\rho_R^* \cdot |S^*| - d_R(u, G[S^*])}{|S^*| - 1}.$$

Since $\rho_R(G[S']) \le \rho_R^*$, we obtain:

$$\frac{\rho_R^* \cdot |S^*| - d_R(u, G[S^*])}{|S^*| - 1} \le \rho_R^*$$

$$\rho_R^* \cdot |S^*| - d_R(u, G[S^*]) \le \rho_R^* \cdot (|S^*| - 1)$$

$$d_R(u, G[S^*]) \ge \rho_R^*.$$

This contradicts the assumption that $d_R(u, G[S^*]) < \rho_R^*$. Hence, we have $\forall u \in S^* \setminus A$, $d_R(u, G[S^*]) \ge \rho_R^*$. By the definition of the k-$r$core, the theorem holds. ◻

Lemma 5.5. Given a graph G=(V, E), and its subgraph S, we have: $\sum_{v \in V(S)} d_R(v, S) = 2 \times (2|E(S)| - \sum_{u \in V(S) \setminus R} |N(u, S)|)$.

Proof. To prove this lemma, we first have:

$$2|E(S)| - \sum_{u \in V(S) \setminus R} |N(u, S)|$$

$$= \sum_{u \in V(S)} d(u, S) - \sum_{u \in V(S) \wedge u \notin R} |N(u, S)|$$

$$= \sum_{u \in V(S) \wedge u \in R} d(u, S) + \sum_{u \in V(S) \wedge u \notin R} d(u,]) - \sum_{u \in V(S) \wedge u \notin R} |N(u, S)|$$

$$= \sum_{u \in V(S) \wedge u \in R} d(u, S).$$

On the other hand, we have:

$$\sum_{u \in V(S)} d_R(u, S)$$

$$= \sum_{u \in V(S) \wedge u \in R} d_R(u, S) + \sum_{u \in V(S) \wedge u \notin R} d_R(u, S)$$

$$= \sum_{u \in V(S) \wedge u \in R} d(u, S) + d(u, G[V(S) \cap R]) + \sum_{u \in V(S) \wedge u \notin R} d(u, G[V(S) \cap R])$$

$$= \sum_{u \in V(S) \wedge u \in R} d(u, S) + \sum_{u \in V(S)} d(u, G[V(S) \cap R])$$

$$= \sum_{u \in V(S) \wedge u \in R} d(u, S) + \sum_{(u,v) \in E(S)} \mathbb{1}[u \in R]$$

$$= \sum_{u \in V(S) \wedge u \in R} d(u, S) + \sum_{v \in V(S) \wedge v \in R} d(v, S)$$

Thus, the above lemma holds. ◻

Lemma 5.6. Given a graph $G = (V, E)$ and two vertex sets $R, A \subseteq V$, if there exists a k-$r$core $\mathcal{M}$ in G, then the NR-density of the ADS in G is at least $\frac{k^2}{2(k+|A|)}$.

Proof. Consider the subgraph $G[V(\mathcal{M}) \cup A]$, it's NR-density satisfies:

$$\begin{aligned}\rho_R(G[V(\mathcal{M}) \cup A]) =& \frac{\sum_{u \in V(\mathcal{M}) \cup A} d_R(u, G[V(\mathcal{M}) \cup A])}{2 \cdot |V(\mathcal{M}) \cup A|} \\ \geq& \frac{\sum_{u \in V(\mathcal{M})} d_R(u, \mathcal{M})}{2 \cdot (|V(\mathcal{M})| + |A|)} \\ \geq& \frac{k \cdot |V(\mathcal{M})|}{2 \cdot (|V(\mathcal{M})| + |A|)}.\end{aligned}$$

Consider the function $f(x) = \frac{k \cdot x}{x+|A|}$. Its derivative is given by

$$f'(x) = \frac{k \cdot |A|}{(x + |A|)^2} > 0,$$

which indicates that $f(x)$ is strictly increasing with respect to x.

Now, let $x = |V(\mathcal{M})|$. Since $|V(\mathcal{M})| \geq k$, we have:

$$\rho_R\big(G[V(\mathcal{M}) \cup A]\big) \geq \frac{k^2}{2 \cdot (k + |A|)}.$$

Therefore, $\frac{k^2}{2 \cdot (k+|A|)}$ is a lower bound on the NR-density of G. □

Theorem 6.1. Given a graph $G = (V, E)$, and two vertex sets $R, A \subseteq V$. In PASTA, when

$$T > O\left(\frac{\sqrt{|A|} \cdot \Delta(G) \cdot \sqrt{(\mathrm{vol}(R) + |A|\Delta(G))}}{\epsilon}\right),$$

the algorithm is guaranteed to return a $(1 + \epsilon)$-approximation solution.

The result in Theorem 6.1 is straightforward, since PASTA is built upon PGD-ADS, and thus inherits its convergence properties. Specifically, the convergence rate of our method is significantly better than that of FDP, as shown in the following theorem:

THEOREM 6.2. To find a $(1 + \epsilon)$-approximation solution, PASTA reduces the overall time complexity at least by $O\left(\frac{\mathrm{vol}(R)^{\frac{3}{4}} \cdot \Delta(G)^{\frac{5}{4}}}{\epsilon}\right)$ over FDP, where all variables keep the same meanings as Theorem 4.4

Proof. To obtain a $(1 + \epsilon)$-approximation solution, the baseline method FDP and PASTA take $O\left(\frac{(\Delta(G) \cdot (\mathrm{vol}(R) + |A|\Delta(G)^2)) \cdot \mathrm{vol}(R)}{\epsilon^2}\right)$, and $O\left(\frac{\sqrt{|A|} \cdot \mathrm{vol}(R)^{\frac{1}{4}} \cdot \Delta(G)^{\frac{3}{4}} \cdot \sqrt{(\mathrm{vol}(R) + |A|\Delta(G))}}{\epsilon}\right)$ time theoretically, respectively. Compared to FDP, our method achieves a reduction of:

$$O\left(\frac{\mathrm{vol}(R)^{\frac{3}{4}} \Delta(G)^{\frac{1}{4}} (\mathrm{vol}(R) + |A|\Delta(G)^2)}{\epsilon \cdot \sqrt{|A|}\sqrt{\mathrm{vol}(R) + |A|\Delta(G)}}\right),$$

which can be further reduced to $O(\frac{\mathrm{vol}(R)\Delta(G)}{\epsilon})$. This implies that the function $g(x)$ (where $x = |A|$) should be defined as:

$$g(x) = \frac{\mathrm{vol}(R)^{\frac{3}{4}} \Delta(G)^{\frac{1}{4}} (\mathrm{vol}(R) + x\Delta(G)^2)}{\sqrt{x}\sqrt{\mathrm{vol}(R) + x\Delta(G)}}$$

To find the minimum of $g(x)$, we can equivalently minimize $g(x)^2$ as the square root is a monotonic function. We can rewrite $g(x)^2$ (ignoring the constant $\Delta(G)$ for minimization purposes):

$$\frac{g(x)^2}{\mathrm{vol}(R)^{\frac{3}{2}} \Delta(G)^{\frac{1}{2}}} = \frac{(\mathrm{vol}(R) + x\Delta(G)^2)^2}{x(\mathrm{vol(R)} + x\Delta(G))}$$

Let $h(x) = \frac{(\mathrm{vol}(R)+x\Delta(G)^2)^2}{x(\mathrm{vol(R)}+x\Delta(G))}$. Let $a = \mathrm{vol}(R), b = \Delta(G)$, we can rewrite $h(x)$ as $h(x) = \frac{(a+b^2x)^2}{x(a+bx)}$ Now, we compute the derivative of $h(x)$ with respect to x:

$$h'(x) = \frac{d}{dx}\left(\frac{(a + b^2x)^2}{x(a + bx)}\right) = \frac{a(a + b^2x)}{x^2(a + bx)^2}((b^2 - 2b)x - a)$$

Given that $a, b > 0$ and $x \in [0, |R|]$, there are two cases regarding whether $0 < b \leq 2$ or $b > 2$.

When $0 < b \leq 2$, $h'(x) < 0$ when $x \in [0, |R|]$. Therefore, the minimum value of $h(x)$ (thus $g(x)$) occurs at $x = |R|$. Note that $|R| < \mathrm{vol}(R) = a$, That is, we have $g(x) \geq g(|R|) \geq g(a) = \frac{a^{\frac{3}{4}} b^{\frac{1}{4}} (a+b^2 \cdot a)}{\sqrt{a(a+b \cdot a)}} = \frac{a^{\frac{3}{4}} b^{\frac{1}{4}} (1+b^2)}{\sqrt{1+b}} = \frac{\mathrm{vol}(R)^{\frac{3}{4}} \Delta(G)^{\frac{1}{4}} (1+\Delta(G)^2)}{\sqrt{1+\Delta(G)}} = O(\mathrm{vol}(R)^{\frac{3}{4}} \Delta(G)^{\frac{7}{4}})$.

When $b > 2$, we have

$$h'(x) \begin{cases} < 0, & x \in [0, \frac{a}{b(b-2)}) \\ > 0, & x \in (\frac{a}{b(b-2)}, |R|] \end{cases}$$

Therefore, the minimum value of $h(x)$ (thus $g(x)$) occurs at $x = \frac{a}{b(b-2)}$. That is, we have $g(x) \geq g(\frac{a}{b(b-2)}) = \frac{a^{\frac{3}{4}} b^{\frac{1}{4}} (a+b^2 \frac{a}{b(b-2)})}{\sqrt{\frac{a}{b(b-2)} (a+b \cdot \frac{a}{b(b-2)})}} = 2a^{\frac{3}{4}} b^{\frac{1}{4}} \sqrt{b(b-1)} = 2(\mathrm{vol}(R)^{\frac{3}{4}} \Delta(G)^{\frac{1}{4}} \sqrt{\Delta(G)(\Delta(G) - 1)}) = O(\mathrm{vol}(R)^{\frac{3}{4}} \Delta(G)^{\frac{5}{4}})$

In summary, we conclude that our algorithm, PASTA, reduces the number of iterations by at least $O\left(\frac{\mathrm{vol}(R)^{\frac{3}{4}} \cdot \Delta(G)^{\frac{5}{4}}}{\epsilon}\right)$ compared to FDP, when a $(1 + \epsilon)$-approximate solution is required. □

B MORE DISCUSSIONS

B.1 Review of FISTA

We first briefly review the well-known PGD-based first-order optimization algorithm, fast iterative shrinkage-thresholding algorithm (FISTA) [9].

Algorithm 5: FISTA [9]

input : functions f and h, learning rate η, and iterations T
output: $x^{(T)}$
1 initialize $x^{(0)}$, s.t. $h(x^{(0)}) \leftarrow 0$;
2 $y^{(0)} \leftarrow x^{(0)}$;
3 **foreach** $t = 1, 2, \cdots, T$ **do**
4 $\quad x^{(t)} \leftarrow \mathrm{prox}_h(y^{(t)} - \eta \cdot \nabla f(y^{(t)}))$;
5 $\quad y^{(t)} \leftarrow x^{(t)} + \frac{t-1}{t+2}(x^{(t)} - x^{(t-1)})$;
6 **return** $x^{(T)}$;

As shown in Algorithm 5, the FISTA proceeds in an iterative manner (lines 3–5). In each iteration, it computes the gradient of the smooth function f, followed by a proximal step w.r.t. the non-smooth function h via gradient projection (line 4). FISTA is particularly efficient when the projection step can be computed at low cost.

B.2 The naive Fran-Wolfe-based ADS algorithm

Indeed, the basic Frank-Wolfe algorithm can be adapted to solve the ADS problem by replacing the above algorithm with Frank-Wolfe-ADS in Algorithm 1.

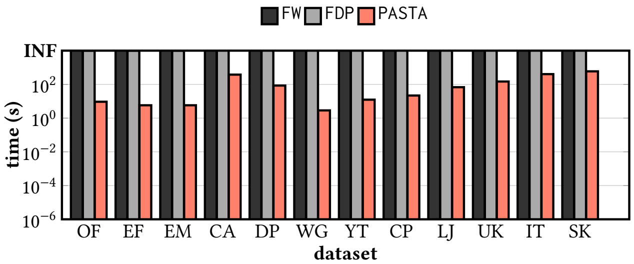

Figure 16: Efficiency of algorithms on ADS ($\epsilon = 0.01$).

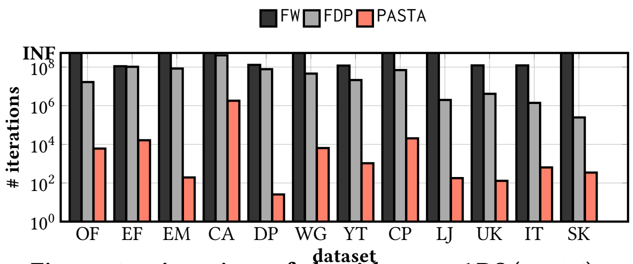

Figure 17: # iterations of algorithms on ADS ($\epsilon = 0.5$).

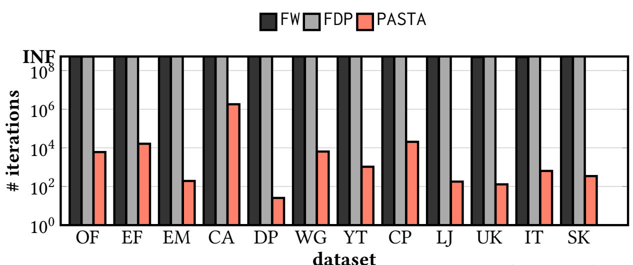

Figure 18: # iterations of algorithms on ADS ($\epsilon = 0.05$).

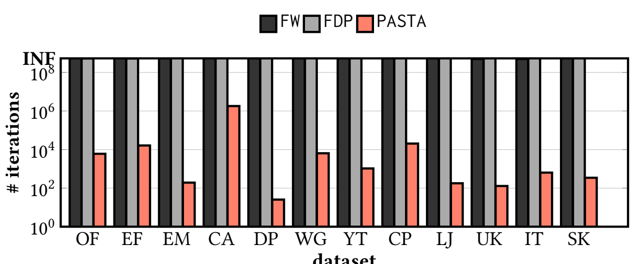

Figure 19: # iterations of algorithms on ADS ($\epsilon = 0.01$).

Algorithm 6: `Naive-Frank-Wolfe`

input : Graph $G=(V,E)$, two vertex sets R, A and an integer T
output: An approximate ADS $\mathcal{D}$

foreach $(u,v) \in E$ **do** $\alpha_{u,v}^{(0)} \leftarrow \frac{1}{2}$; $\alpha_{v,u}^{(0)} \leftarrow \frac{1}{2}$;
foreach $u \in V$ **do** $r^{(0)}(u) = \sum_{v \in N(u,G)} \alpha_{u,v}^{(0)}$;
foreach $t \leftarrow 1, 2, 3, \cdots, T$ **do**
  $\eta_t \leftarrow \frac{2}{t+2}$;
  for $(u,v) \in E$ **do**
    $\widehat{\alpha}_{u,v} \leftarrow \alpha_{u,v}^{(t)} - \eta \cdot \nabla f(\alpha)_{u,v}$;
    $\widehat{\alpha}_{v,u} \leftarrow \alpha_{v,u}^{(t)} - \eta \cdot \nabla f(\alpha)_{v,u}$;
  $\alpha^{(t)} \leftarrow (1-\gamma_t) \cdot \alpha^{(t-1)} + \gamma_t \cdot \widehat{\alpha}$;
  foreach $u \in V$ **do** $r^{(t)}(u) = \sum_{v \in N(u,G)} \alpha_{u,v}^{(t)}$;
// Extract the ADS
sort all $v \in V$ by weight $r^{(T)}(v)$;
$s^* \leftarrow \arg\max_{1 \le i \le n} \rho_R(V_i)$, where V_i is the top-i vertices in V;
return $G[V_{s^*}]$;

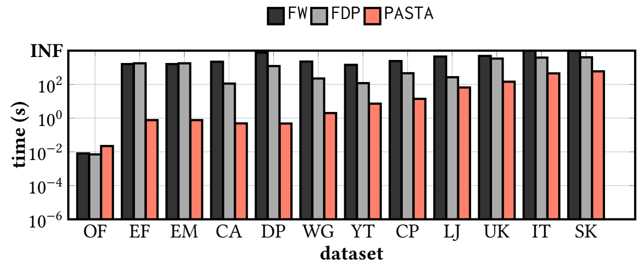

Figure 14: Efficiency of algorithms on ADS ($\epsilon = 0.5$).

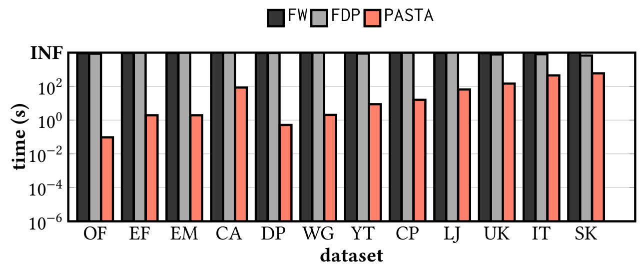

Figure 15: Efficiency of algorithms on ADS ($\epsilon = 0.05$).

C ADDITIONAL EXPERIMENTS

In this section, we present the additional experimental results:

• **1. Efficiency of approximation algorithms.** We evaluate the efficiency of three approximations on all datasets with ϵ from 0.5 to 0.01. As shown in Figures 14 ∼ 16, we can see that: (1) Our algorithm PASTA is up to four orders of magnitude faster than the baseline methods, since it has a better theoretical guarantee and can find the ADS over the very small graph. For example, on the OF dataset, PASTA only takes 10s to find the 1.01-approximation solution, while both FDP and Frank-Wolfe take more than three days. (2) On more than half of the datasets, neither Frank-Wolfe nor FDP can find solutions that meet the accuracy requirement within three days, which is consistent with our theoretical analysis.

• **2. The number of iterations.** In this experiment, we report the number of iterations that each algorithm needed to find the solutions with various accuracy requirements, i.e., ϵ from 0.5 to 0.01 in Figures 17 ∼ 19. We make the following observations: (1) PASTA requires significantly fewer iterations to obtain approximate solutions compared to its competitors. For example, on the DP dataset, Frank-Wolfe, FDP, and PASTA require around 129M, 77M, and 41 iterations, respectively. (2) (2) As the accuracy requirement increases (i.e., ϵ decreases from large to small), the number of iterations required by Frank-Wolfe and FDP grows significantly, whereas our method PASTA remains stable.