A. Additional Approximation Results

A.1. The case of monotone submodular F

The main difficulty in proving Theorems 1 and 2 in Section 2.1 is that the ϵ-approximate oracle F is not monotone submodular. In the event that F is monotone submodular, existing results for SCSC (Wolsey, 1982; Wan et al., 2010; Soma & Yoshida, 2015) can be translated into results for SCSC with an ϵ-approximate oracle, as the following proposition shows.

Proposition 2. *Let $\mathcal{A}$ be a bicriteria approximation algorithm for SCSC that takes as input f, c, τ and has approximation ratio α, and feasibility guarantee $\tau - \delta$.*

Suppose that we have two instances of SCSC: instance $\mathcal{I}_1$ with f, c, τ and instance $\mathcal{I}_2$ with $F, c, \tau - \epsilon$ where F is an ϵ-approximation to f. Then if we run $\mathcal{A}$ for $\mathcal{I}_2$ and $\mathcal{A}$ returns the set A, it is guaranteed that $c(A) \leq \alpha c(A^*)$ and $f(A) \geq \tau - 2\epsilon - \delta$ where A^* is the optimal solution to instance $\mathcal{I}_1$.

Proof. A^* is a feasible solution to $\mathcal{I}_2$ since $F(A^*) \geq f(A^*) - \epsilon \geq \tau - \epsilon$. Therefore $\alpha c(A^*) \geq c(A)$. For the feasibility, we have that $f(A) \geq F(A) - \epsilon \geq \tau - 2\epsilon - \delta$. □

A.2. An alternative version of Theorem 1

The definitions and notation used in this section can be found in Section 1.3 of the paper.

The version of Theorem 1 in Section 2 requires that ϵ be small relative to the minimum marginal gain of an element added to the greedy set, μ, over the duration of Algorithm 1. In particular, $\mu > 4\epsilon c_{max}\rho/c_{min}$. Alternatively, we may ensure the approximation ratio of Theorem 1 if ϵ is sufficiently small by exiting Algorithm 1 if $F_\tau(A_i) - F_\tau(A_{i-1})$ falls below an input value, μ^*. The feasibility guarantee is weakened since Algorithm 1 does not necessarily run to completion, but not by much if ϵ is sufficiently small. In particular, we have the following alternative version of Theorem 1.

Theorem 1 (Alternative) Suppose we have an instance of SCSC with $n \geq 1$. Let F be a function that is ϵ-approximate to f.

Let $\mu^* > 0$ be given. Suppose Algorithm 1 is run with input F, c, and τ, but we exit the algorithm at the first iteration k such that $F_\tau(A_k) - F_\tau(A_{k-1}) \leq \mu^*$ and return $A = A_{k-1}$. Then $f(A) \geq \tau - n((c_{max}/c_{min})\mu^* + 2\epsilon)$ and if $\epsilon < \mu^*/(4c_{max}\rho/c_{min} + 2)$, then

$$c(A) \leq \frac{\rho}{1 - \frac{4\epsilon c_{max}\rho}{c_{min}(\mu^* - 2\epsilon)}} \left(\ln\left(\frac{\alpha}{\beta}\right) + 2 \right) c(A^*)$$

where A^* is the optimal solution to the instance of SCSC.

Proof of Theorem 1 (Alternative) Without loss of generality we re-define $f = \min\{f, \tau\}$ and $F = \min\{F, \tau\}$. This way, $f = f_\tau$ and $F = F_\tau$.

First, we prove the feasibility. If Algorithm 1 runs to completion, then the feasibility guarantee is clear since $f(A) \geq \tau - \epsilon$. Suppose that Algorithm 1 did not run to completion, but instead returned $A = A_{k-1}$ once $F(A_k) - F(A_{k-1}) \leq \mu^*$. Let x_k be the element that is selected in the kth iteration. Then for all x

$$\Delta F(A, x) \leq \frac{c(x)}{c(x_k)} \Delta F(A, x_k) \leq \frac{c_{max}}{c_{min}} \mu^*.$$

Therefore for all x $\Delta f(A, x) \leq \Delta F(A, x) + 2\epsilon \leq (c_{max}/c_{min})\mu^* + 2\epsilon$. By the submodularity of f we have that

$$\tau - f(A) \leq f(S) - f(A) \leq \sum_x \Delta f(A, x) \leq n \left(\frac{c_{max}}{c_{min}} \mu^* + 2\epsilon \right)$$

and so the feasibility guarantee is proven.

The approximation ratio follows by using the same argument as in Theorem 1 where μ is replaced by $\mu^* - 2\epsilon$ since

$$\mu = \min\{f(A_i) - f(A_{i-1}) : i \leq k\} \geq \min\{F(A_i) - F(A_{i-1}) : i \leq k\} - 2\epsilon > \mu^* - 2\epsilon.$$

□

B. Proof of Theorem 1

The definitions and notation used in this section can be found in Section 1.3 of the paper.

Theorem 1 Suppose we have an instance of SCSC. Let F be a function that is ϵ-approximate to f.

Suppose we run Algorithm 1 with input F, c, and τ. Then $f(A) \geq \tau - \epsilon$. And if $\mu > 4\epsilon c_{max}\rho/c_{min}$,

$$c(A) \leq \frac{\rho}{1 - \frac{4\epsilon c_{max}\rho}{c_{min}\mu}} \left(\ln\left(\frac{\alpha}{\beta}\right) + 2 \right) c(A^*)$$

where A^* is an optimal solution to the instance of SCSC.

Proof of Theorem 1 The feasibility guarantee is clear from the stopping condition on the greedy algorithm: $f(A) \geq F(A) - \epsilon \geq \tau - \epsilon$.

We now prove the upper bound on $c(A)$ if $\mu > 4\epsilon c_{max}\rho/c_{min}$. Without loss of generality we re-define $f = \min\{f, \tau\}$ and $F = \min\{F, \tau\}$. This way, $f = f_\tau$ and $F = F_\tau$. Notice that this does not change that F is an ϵ-approximation of f since the error is absolute.

Let $x_1, ..., x_k$ be the elements of A in the order that they were chosen by the greedy algorithm. If $k = 0$, then $c(A) = 0$ and the approximation ratio is clear. For the rest of the proof, we assume that $k \geq 1$.

We define a sequence of elements $\tilde{x}_1, ..., \tilde{x}_k$ where

$$\tilde{x}_i = \text{argmax}_{x \in S \setminus A_{i-1}} \frac{\Delta f(A_{i-1}, x)}{c(x)}.$$

$\tilde{x}_i$ has the most cost-effective marginal gain of being added to A_{i-1} according to f, while x_i has the most cost-effective marginal gain of being added to A_{i-1} according to F. Note that the same element can appear multiple times in the sequence $\tilde{x}_1, ..., \tilde{x}_k$. In addition, we have a lower bound on $\Delta f(A_{i-1}, \tilde{x}_i)$:

$$\Delta f(A_{i-1}, \tilde{x}_i) \geq \frac{c(\tilde{x}_i)}{c(x_i)} \Delta f(A_{i-1}, x_i) \geq \frac{c_{min}}{c_{max}} \mu. \tag{1}$$

Our argument to bound $c(A)$ will follow the following three steps: **(a)** We bound $c(A)$ in terms of the costs of the elements $\tilde{x}_1, ..., \tilde{x}_k$. **(b)** We charge the elements of A^* with the costs of the elements $\tilde{x}_1, ..., \tilde{x}_k$, and bound $c(A)$ in terms of the total charge on all elements in A^*. **(c)** We bound the total charge on the elements of A^* in terms of $c(A^*)$.

(a) First, we bound $c(A)$ in terms of the costs of the elements $\tilde{x}_1, ..., \tilde{x}_k$. At iteration i of Algorithm 1, the most cost-effective element to add to the set A_{i-1} according to F is x_i. Using the fact that F is ϵ-approximate to f, we can bound how much more cost-effective $\tilde{x}_i$ is compared to x_i according to f as follows:

$$\frac{\Delta f(A_{i-1}, x_i) + 2\epsilon}{c(x_i)} \geq \frac{\Delta F(A_{i-1}, x_i)}{c(x_i)} \geq \frac{\Delta F(A_{i-1}, \tilde{x}_i)}{c(\tilde{x}_i)} \geq \frac{\Delta f(A_{i-1}, \tilde{x}_i) - 2\epsilon}{c(\tilde{x}_i)}$$

which implies that

$$\frac{\Delta f(A_{i-1}, x_i)}{c(x_i)} + \frac{2\epsilon(c(x_i) + c(\tilde{x}_i))}{c(x_i)c(\tilde{x}_i)} \geq \frac{\Delta f(A_{i-1}, \tilde{x}_i)}{c(\tilde{x}_i)}. \tag{2}$$

$\Delta f(A_{i-1}, x_i) \geq \mu > 0$ by assumption, and $\Delta f(A_{i-1}, \tilde{x}_i) \geq (c_{min}/c_{max})\mu > 0$ by Equation (1). Therefore we can re-arrange Equation (2) to be

$$\frac{c(\tilde{x}_i)}{\Delta f(A_{i-1}, \tilde{x}_i)} + \alpha_i \geq \frac{c(x_i)}{\Delta f(A_{i-1}, x_i)}, \text{ where } \alpha_i = \frac{2\epsilon(c(x_i) + c(\tilde{x}_i))}{\Delta f(A_{i-1}, \tilde{x}_i)\Delta f(A_{i-1}, x_i)}. \tag{3}$$

Inequality (3) and the submodularity of c imply that

$$\begin{aligned} c(A) &\leq \sum_{i=1}^{k} c(x_i) = \sum_{i=1}^{k} \Delta f(A_{i-1}, x_i) \frac{c(x_i)}{\Delta f(A_{i-1}, x_i)} \\ &\leq \sum_{i=1}^{k} \Delta f(A_{i-1}, x_i) \left(\frac{c(\tilde{x}_i)}{\Delta f(A_{i-1}, \tilde{x}_i)} + \alpha_i \right) \\ &= \sum_{i=1}^{k} \Delta f(A_{i-1}, x_i) \frac{c(\tilde{x}_i)}{\Delta f(A_{i-1}, \tilde{x}_i)} + \sum_{i=1}^{k} \frac{2\epsilon(c(\tilde{x}_i) + c(x_i))}{\Delta f(A_{i-1}, \tilde{x}_i)}. \end{aligned} \tag{4}$$

We now bound the second term on the right side of Equation (4) by

$$\begin{aligned} \sum_{i=1}^{k} \frac{2\epsilon(c(\tilde{x}_i) + c(x_i))}{\Delta f(A_{i-1}, \tilde{x}_i)} &= \sum_{i=1}^{k} \frac{2\epsilon c(\tilde{x}_i)}{\Delta f(A_{i-1}, \tilde{x}_i)} + \sum_{i=1}^{k} \frac{2\epsilon c(x_i)}{\Delta f(A_{i-1}, \tilde{x}_i)} \\ &\leq \sum_{i=1}^{k} \frac{2\epsilon c(x_i)}{\Delta f(A_{i-1}, x_i)} + \sum_{i=1}^{k} \frac{2\epsilon c(x_i)}{\Delta f(A_{i-1}, \tilde{x}_i)} \\ &\leq \frac{4\epsilon c_{max}}{c_{min}\mu} \sum_{i=1}^{k} c(x_i) \\ &\leq \frac{4\epsilon c_{max}\rho}{c_{min}\mu} c(A). \end{aligned} \tag{5}$$

The second to last inequality in Equation (5) follows from the fact that $\Delta f(A_{i-1}, x_i) \geq \mu \geq (c_{min}/c_{max})\mu$, and that by Equation (1) $\Delta f(A_{i-1}, \tilde{x}_i) \geq (c_{min}/c_{max})\mu$. The last inequality in Equation (5) uses the definition of the curvature ρ of c.

Combining Equations (4) and (5) gives us the following bound on $c(A)$ in terms of the costs of the elements $\tilde{x}_1, ..., \tilde{x}_k$:

$$\left(1 - \frac{4\epsilon c_{max}\rho}{c_{min}\mu}\right) c(A) \leq \sum_{i=1}^{k} \frac{\Delta f(A_{i-1}, x_i)}{\Delta f(A_{i-1}, \tilde{x}_i)} c(\tilde{x}_i). \tag{6}$$

(b) Next, we charge the elements of A^* with the costs of the elements $\tilde{x}_1, ..., \tilde{x}_k$, and bound $c(A)$ in terms of the total charge on all elements in A^*. By this we mean that we give each $y \in A^*$ a portion of the total cost of the elements $\tilde{x}_1, ..., \tilde{x}_k$. In particular, we give each $y \in A^*$ a charge of $w(y)$, defined by

$$w(y) = \sum_{i=1}^{k} (\pi_i(y) - \pi_{i+1}(y))\omega_i, \text{ where } \omega_i = \frac{c(\tilde{x}_i)}{\Delta f(A_{i-1}, \tilde{x}_i)}, \text{ and } \pi_i(y) = \begin{cases} \Delta f(A_{i-1}, y) & i \in \{1, ..., k\} \\ \Delta f(A, y) & i = k+1 \end{cases}.$$

Recall that $\Delta f(A_{i-1}, \tilde{x}_i) > 0$ for all i by Equation (1), and so we can define ω_i as above. Wan et al. charged the elements of A^* with the cost of elements $x_1, ..., x_k$ analogously to the above. We charge with the cost of elements $\tilde{x}_1, ..., \tilde{x}_k$ because they exhibit diminishing cost-effectiveness, i.e. $\omega_i - \omega_{i-1} \geq 0$ for all $i \in \{1, ..., k\}$, which is needed to proceed with the argument. Because we choose $x_1, ..., x_k$ with F, which is not monotone submodular, $x_1, ..., x_k$ do not exhibit diminishing cost-effectiveness even if we replace f with F in the definition of $w(y)$ above. We now follow an argument analogous to Wan et al. but with the elements $\tilde{x}_1, ..., \tilde{x}_k$ in order to prove Equation (9).

Consider any $y \in A^*$. We can re-write $w(y)$ as

$$\begin{aligned}
w(y) &= \sum_{i=1}^{k} (\pi_i(y) - \pi_{i+1}(y))\omega_i \\
&= \sum_{i=1}^{k} \pi_i(y)\omega_i - \sum_{i=1}^{k} \pi_{i+1}(y)\omega_i \\
&= \pi_1(y)\omega_1 + \sum_{i=2}^{k} \pi_i(y)\omega_i - \sum_{i=2}^{k} \pi_i(y)\omega_{i-1} - \pi_{k+1}(y)\omega_k \\
&= \pi_1(y)\omega_1 - \Delta f(A, y)\omega_k + \sum_{i=2}^{k} (\omega_i - \omega_{i-1})\pi_i(y).
\end{aligned}$$

Summing over $y \in A^*$, we have

$$\sum_{y \in A^*} w(y) = \omega_1 \sum_{y \in A^*} \pi_1(y) + \sum_{i=2}^{k} (\omega_i - \omega_{i-1}) \sum_{y \in A^*} \pi_i(y) - \sum_{y \in A^*} \Delta f(A, y)\omega_k. \tag{7}$$

On the other hand, starting with Equation (6), we see that

$$\begin{aligned}
\left(1 - \frac{4\epsilon c_{max}\rho}{\mu c_{min}}\right) c(A) &\leq \sum_{i=1}^{k} \frac{\Delta f(A_{i-1}, x_i)}{\Delta f(A_{i-1}, \tilde{x}_i)} c(\tilde{x}_i) \\
&= \Delta f(A_{k-1}, x_k)\omega_k + \sum_{i=1}^{k-1} \Delta f(A_{i-1}, x_i)\omega_i \\
&= \Delta f(A_{k-1}, x_k)\omega_k + \sum_{i=1}^{k-1} (\sum_{j=i}^{k} \Delta f(A_{j-1}, x_j) - \sum_{j=i+1}^{k} \Delta f(A_{j-1}, x_j))\omega_i \\
&= \sum_{i=1}^{k} \sum_{j=i}^{k} \Delta f(A_{j-1}, x_j)\omega_i - \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \Delta f(A_{j-1}, x_j)\omega_i \\
&= \omega_1 \sum_{j=1}^{k} \Delta f(A_{j-1}, x_j) + \sum_{i=2}^{k} \sum_{j=i}^{k} \Delta f(A_{j-1}, x_j)\omega_i - \sum_{i=2}^{k} \sum_{j=i}^{k} \Delta f(A_{j-1}, x_j)\omega_{i-1} \\
&= \omega_1 \sum_{j=1}^{k} \Delta f(A_{j-1}, x_j) + \sum_{i=2}^{k} (\omega_i - \omega_{i-1}) \sum_{j=i}^{k} \Delta f(A_{j-1}, x_j).
\end{aligned} \tag{8}$$

In order to find a link between Equations (7) and (8), we first notice that for any $i \in \{1, ..., k\}$

$$\sum_{j=i}^{k} \Delta f(A_{j-1}, x_j) = f(A_k) - f(A_{i-1}) \leq \sum_{y \in A^*} \Delta f(A_{i-1}, y) = \sum_{y \in A^*} \pi_i(y).$$

In addition, for any $i \in \{1, ..., k\}$

$$\begin{aligned}
\omega_i - \omega_{i-1} &= \frac{c(\tilde{x}_i)}{\Delta f(A_{i-1}, \tilde{x}_i)} - \frac{c(\tilde{x}_{i-1})}{\Delta f(A_{i-2}, \tilde{x}_{i-1})} \\
&\geq \frac{c(\tilde{x}_i)}{\Delta f(A_{i-2}, \tilde{x}_i)} - \frac{c(\tilde{x}_{i-1})}{\Delta f(A_{i-2}, \tilde{x}_{i-1})} \\
&\geq \frac{c(\tilde{x}_{i-1})}{\Delta f(A_{i-2}, \tilde{x}_{i-1})} - \frac{c(\tilde{x}_{i-1})}{\Delta f(A_{i-2}, \tilde{x}_{i-1})} = 0.
\end{aligned}$$

We may therefore link Equations (7) and (8) to see that

$$\begin{aligned}\left(1-\frac{4\epsilon c_{max}\rho}{\mu c_{min}}\right)c(A) &\leq \omega_1\sum_{j=1}^{k}\Delta f(A_{j-1},x_j)+\sum_{i=2}^{k}(\omega_i-\omega_{i-1})\sum_{j=i}^{k}\Delta f(A_{j-1},x_j)\\ &\leq \sum_{y\in A^*}w(y)+\sum_{y\in A^*}\Delta f(A,y)\omega_k. \end{aligned} \tag{9}$$

Consider $y\in A^*$. $f(A)$ is not necessarily τ since the stopping condition for Algorithm 1 is only that $F(A)\geq\tau$. In this case, if $\Delta f(A,y)\neq 0$ for $y\in A^*$ (which implies that $y\notin A_{k-1}$) then by the submodularity of f

$$\Delta f(A,y)\omega_k=\Delta f(A,y)\frac{c(\tilde{x}_k)}{\Delta f(A_{k-1},\tilde{x}_k)}\leq\Delta f(A,y)\frac{c(y)}{\Delta f(A_{k-1},y)}\leq\Delta f(A,y)\frac{c(y)}{\Delta f(A,y)}=c(y).$$

Therefore we can bound $c(A)$ in terms of the total charge on all elements in A^*:

$$\left(1-\frac{4\epsilon c_{max}\rho}{\mu c_{min}}\right)c(A)\leq\sum_{y\in A^*}w(y)+\sum_{y\in A^*}c(y)\leq\sum_{y\in A^*}w(y)+\rho c(A^*). \tag{10}$$

(c) Finally, we bound the total charge on the elements of A^* in terms of $c(A^*)$.

We first define a value ℓ_y for every $y\in A^*$. For each $y\in A^*$, if $\pi_1(y)=0$ we set $\ell_y=0$, otherwise ℓ_y is the value in $\{1,...,k\}$ such that if $i\in\{1,...,\ell_y\}$ then $\pi_i(y)>0$, and if $i\in\{\ell_y+1,...,k\}$ then $\pi_i(y)=0$. Such an ℓ_y can be set since f is submodular and monotonic. Then

$$\begin{aligned} w(y) &= \sum_{i=1}^{\ell_y}(\pi_i(y)-\pi_{i+1}(y))\omega_i\\ &\leq \sum_{i=1}^{\ell_y}(\pi_i(y)-\pi_{i+1}(y))\frac{c(y)}{\pi_i(y)}\\ &\leq c(y)\left(\left(\sum_{i=1}^{\ell_y-1}\frac{\pi_i(y)-\pi_{i+1}(y)}{\pi_i(y)}\right)+1\right)\\ &\leq c(y)\left(\sum_{i=1}^{\ell_y-1}(\ln(\pi_i(y))-\ln(\pi_{i+1}(y)))+1\right)\\ &= c(y)\left(\ln(\pi_1(y))-\ln(\pi_{\ell_y}(y))+1\right)\\ &\leq c(y)\left(\ln\left(\frac{\alpha}{\beta}\right)+1\right). \end{aligned} \tag{11}$$

The third to last inequality follows since

$$\frac{\pi_i(y)-\pi_{i+1}(y)}{\pi_i(y)}=\int_{\pi_{i+1}(y)}^{\pi_i(y)}\frac{1}{\pi_i(y)}dx\leq\int_{\pi_{i+1}(y)}^{\pi_i(y)}\frac{1}{x}dx=\ln(\pi_i(y))-\ln(\pi_{i+1}(y)).$$

We sum inequality (11) over all $y\in A^*$ to get that

$$\sum_{y\in A^*}w(y)\leq\left(\ln\left(\frac{\alpha}{\beta}\right)+1\right)\sum_{y\in A^*}c(y)\leq\rho\left(\ln\left(\frac{\alpha}{\beta}\right)+1\right)c(A^*). \tag{12}$$

Finally, we combine inequality (10) and inequality (1) to see that

$$\left(1-\frac{4\epsilon c_{max}\rho}{c_{min}\mu}\right)c(A)\leq\sum_{y\in A^*}w(y)+\rho c(A^*)\leq\rho\left(\ln\left(\frac{\alpha}{\beta}\right)+1\right)c(A^*)+\rho c(A^*).$$

If $\mu>(4\epsilon c_{max}\rho)/(c_{min})$, this completes the proof of the approximation guarantee in the theorem statement. □

C. Proof of Theorem 2

The definitions and notation used in this section can be found in Section 2.1 of the paper.

Theorem 2 Suppose we have an instance of SCSC. Let F be a function that is ϵ-approximate to f.

Suppose we run Algorithm 1 with input F, c, and τ. Then $f(A) \geq \tau - \epsilon$. And if $\mu > 4\epsilon c_{max}\rho/c_{min}$, then for any $\gamma \in (0, 1 - 4\epsilon c_{max}\rho/c_{min}\mu)$,

$$c(A) \leq \frac{\rho}{1 - \frac{4\epsilon c_{max}\rho}{c_{min}\mu} - \gamma}\left(\ln\left(\frac{n\alpha\rho}{\gamma\mu}\right) + 2\right)c(A^*)$$

where A^* is an optimal solution to the instance of SCSC.

Proof of Theorem 2 The argument for the proof of Theorem 2 is the same as Theorem 1, except for part **(c)** which is what we present here. In particular, we have gotten to the point of the proof of Theorem 1 where we have proven that

$$\left(1 - \frac{4\epsilon c_{max}\rho}{c_{min}\mu}\right)c(A) \leq \sum_{y \in A^*} w(y) + \rho c(A^*). \tag{1}$$

Let $\lambda > 0$. We first define a value m_y for every $y \in A^*$. For each $y \in A^*$, if $\pi_1(y) \leq \lambda$ we set $m_y = 0$, otherwise m_y is the value in $\{1, ..., k\}$ such that if $i \in \{1, ..., m_y\}$ then $\pi_i(y) > \lambda$, and if $i \in \{m_y + 1, ..., k\}$ then $\pi_i(y) \leq \lambda$. Such an m_y can be set since f is submodular and monotonic. Then

$$w(y) = \sum_{i=1}^{m_y}(\pi_i(y) - \pi_{i+1}(y))\omega_i + \sum_{i=m_y+1}^{k}(\pi_i(y) - \pi_{i+1}(y))\omega_i. \tag{2}$$

A similar analysis as in the proof of Theorem 1 can be used to show that

$$\sum_{i=1}^{m_y}(\pi_i(y) - \pi_{i+1}(y))\omega_i \leq c(y)\left(\ln\left(\frac{\alpha}{\lambda}\right) + 1\right). \tag{3}$$

The remaining part of inequality (2) can be bounded by

$$\begin{aligned}
\sum_{i=m_y+1}^{k}(\pi_i(y) - \pi_{i+1}(y))\omega_i &\leq \sum_{i=m_y+1}^{k}(\pi_i(y) - \pi_{i+1}(y))\frac{c(x_i)}{\Delta f(A_{i-1}, x_i)} \\
&\leq \sum_{i=m_y+1}^{k}\pi_i(y)\frac{c(x_i)}{\Delta f(A_{i-1}, x_i)} \\
&< \sum_{i=m_y+1}^{k}\lambda\frac{c(x_i)}{\mu} \\
&\leq \frac{\lambda\rho}{\mu}c(A).
\end{aligned} \tag{4}$$

Summing Equation (2) over all $y \in A^*$ and applying the upper bounds in Equations (3) and (4) gives us that

$$\sum_{y \in A^*} w(y) \leq \rho\left(\ln\left(\frac{\alpha}{\lambda}\right) + 1\right)c(A^*) + \frac{\lambda\rho n}{\mu}c(A) \tag{5}$$

By combining Equations (5) and (1), we have that

$$\left(1 - \frac{4\epsilon c_{max}\rho}{c_{min}\mu} - \frac{\lambda\rho n}{\mu}\right)c(A) \leq \rho\left(\ln\left(\frac{\alpha}{\lambda}\right) + 2\right)c(A^*).$$

If we set $\lambda = (\gamma\mu)/(n\rho)$ we have the approximation ratio in the theorem statement. $\square$

D. Application and Experiments

Influence Threshold (IT) (Non-Simulation) In contrast to the version of IT in Section 3 of the paper, we may define IT to directly use the influence model as follows.

Let $G = (V, E)$ be a social network where nodes V represents users, edges E represent social connections, and $\mathcal{D}$ is a probability distribution over subsets of E that represents the probability of a set of edges being "alive". Activation of users in the social network starts from an initial seed set and then propagates across alive edges. For $X \subseteq S$, $f(X)$ is the expected number of reachable nodes in V from X when a set of alive edges is sampled from $\mathcal{D}$. $c : 2^V \to \mathbb{R}_{\geq 0}$ is a monotone submodular function that gives the cost of seeding a set of users. The Influence Threshold problem (IT) is to find a seed set A such that $c(A)$ is minimized and $f(A) \geq \tau$.

f as Average Reachability One popular choice for the distribution $\mathcal{D}$ is the Independent Cascade (IC) model (Kempe et al., 2003). In the IC model, every edge has a probability assigned to it $w_E : E \to [0, 1]$. Each edge e is independently alive with probability $w_E(e)$. However, computing the expected influence f under the IC model is $\#P$-hard and therefore impractical to compute (Chen et al., 2010).

As an alternative to working directly with the influence model Kempe et al. proposed a simulation-based approach to approximating it: Random samples from $\mathcal{D}$ are drawn, and for every $X \subseteq S$ $f(X)$ is approximated by the average number of reachable nodes from X over the samples. The resulting approximation of f is unbiased, converges to the expected value, and is monotone submodular. This simulation-based approach is also advantageous since it can be used for more complex models than IC or for instances generated from traces (Cohen et al., 2014).

Because the average reachability is monotone submodular, it is easy to translate approximation results for IT where f is replaced by the average reachability into approximation results for IT (see Proposition 2 of Appendix A).

Additional Experimental Setup Details We use two real social networks: the Facebook ego network (Leskovec & Mcauley, 2012) with $n = 4039$, and the ArXiV General Relativity collaboration network with $n = 5242$ (Leskovec et al., 2007), which we refer to as GrQc.

Influence propagation follows the Independent Cascade (IC) model (Kempe et al., 2003). The probabilities assigned to the edges $w_E : E \to [0, 1]$ follow the weighted cascade model (Kempe et al., 2003). In the weighted cascade model, an edge that goes from $u \in V$ to $v \in V$ is assigned probability $\frac{q}{d_v}$ where d_v is the number of incoming edges to node v and $q \in (0, 1]$. For facebook $q = 0.5$, and for GrQc $q = 0.8$.

The average reachability oracle, f, is over $N = 25000$ random realizations of the influence graph. The approximate average reachability oracle of Cohen et al., F, is computed over these realizations with various oracle errors ϵ and the greedy algorithm is run using these oracles. For comparison, we also run the greedy algorithm with f.

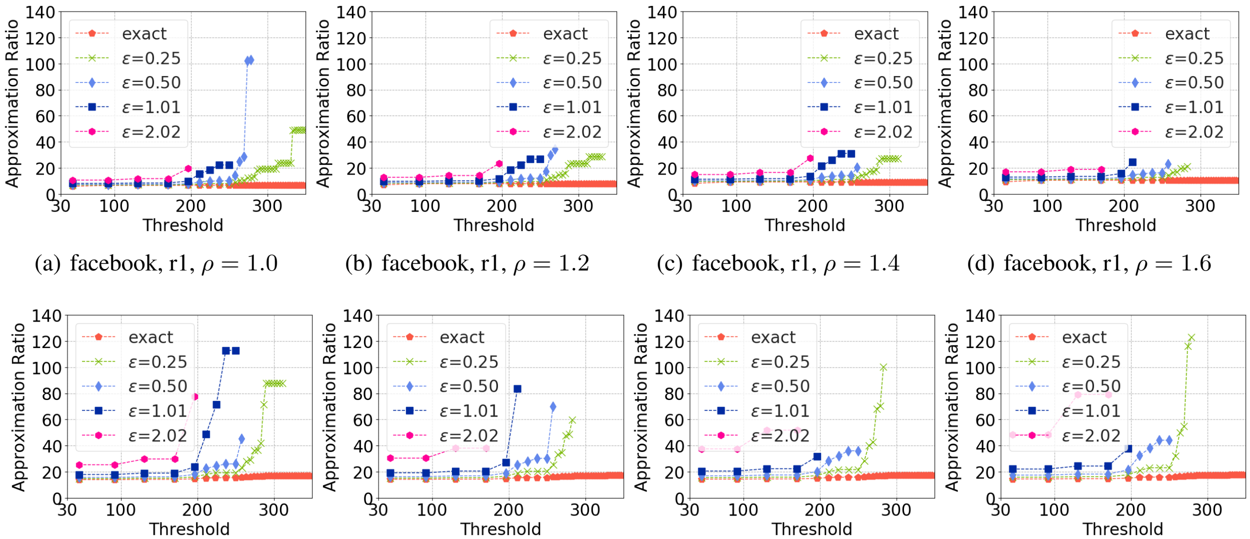
For the cost function $c : 2^V \to \mathbb{R}_{\geq 0}$, we choose a cost c_v for each $v \in V$ by sampling from a normal distribution with mean 1 and standard deviation 0.1, and then define $c(X) = \sum_{x \in X} c_x$.

To select the parameter γ when computing the ratio of Theorem 2, we discretize the domain of γ, compute the ratio on each of the points, and select the γ that gives the smallest approximation ratio.

Additional Experimental Results The experimental results presented in Section 3 of the paper consider a modular cost function. However, the approximation results presented in Theorems 1 and 2 are for general submodular cost functions. In this section, we present additional experimental results analogous to those in Section 3 of the paper but with non-modular cost function.

Recall from Section 1.3 that a measure of how far a function is from being is modular is measured by its curvature, $\rho \in [1, \infty)$. Our approximation ratios in Theorems 1 and 2 depend on ρ: the smaller ρ, and hence the closer to being modular the function is, the better the approximation ratio.

Notice that the greedy algorithm, presented in Section 1.3, chooses elements only according to singleton costs. Therefore, the experimental results in Section 3 of the paper can easily be extended to cost functions that are not modular by simply computing the ratio for different values of cost curvature, ρ; higher values of ρ would require smaller epsilon to get the same ratio.



(a) facebook, r1, $\rho = 1.0$
(b) facebook, r1, $\rho = 1.2$
(c) facebook, r1, $\rho = 1.4$
(d) facebook, r1, $\rho = 1.6$

(e) facebook, r2, $\rho = 1.0$
(f) facebook, r2, $\rho = 1.2$
(g) facebook, r2, $\rho = 1.4$
(h) facebook, r2, $\rho = 1.6$

Figure 4. The approximation ratios of Theorem 1 (r1) and an upper bound on that of Theorem 2 (r2) at thresholds indicated by the markers.

The approximation ratio of Theorem 1 on the Facebook dataset is plotted in Figures 4(a) to 4(d) for varying curvature ρ, and that of Theorem 2 is plotted in Figures 4(e) to 4(h). As ρ increases, the approximation ratios are subtly greater. The greater that ρ is, the greater μ must be relative to ϵ in order to have the approximation ratios of Theorems 1 and 2. This results in the ratios of Theorems 1 and 2 not being guaranteed at lower thresholds for larger ρ.