Supplementary material for the Densest SWAMP problem: subhypergraphs with arbitrary monotonic partial edge rewards

Vedangi Bengali[1] (✉), Nikolaj Tatti[2], Iiro Kumpulainen[2], Florian Adriaens[2], and Nate Veldt[1]

[1] Texas A&M University, College Station, TX, USA
{vedangibengali, nveldt}@tamu.edu
[2] University of Helsinki, HIIT, Helsinki, Finland
{nikolaj.tatti, iiro.kumpulainen, florian.adriaens}@helsinki.fi

A NP-hardness of $\textsc{swamp}(r)$

In this section we will prove the following result.

Theorem 1. *Assume r such that for some positive integers i and j we have $r(i)/i > r(1)$ and $2r(j) > r(j-1) + r(j+1)$. Then $\textsc{swamp}(r)$ is **NP**-hard.*

We can safely assume that i is the smallest index for which $r(i)/i > r(1)$.

Let us write $\delta(a) = r(a) - r(a-1)$. Define $\theta = r(i) - ir(1)$ and $\eta = \delta(j) - \delta(j+1)$.

Let $R = \max r$ and $\Delta = \max \delta$.

We will prove the hardness by a reduction from the independent set problem in 3-regular graphs. Assume we are given a 3-regular graph $H = (W, A)$ with n nodes and m edges.

We construct $G = (V, E)$ as follows. The nodes consists of the original nodes W and $\ell = k(m(j-1) + kbn(i-1))$ nodes S, where

$$b = \left\lceil \frac{4\Delta - r(1)}{\delta(i) - r(1)} \right\rceil \quad \text{and} \quad k = \left\lceil \frac{4R}{\eta} + \frac{6R}{\theta} \right\rceil.$$

Each subset of j nodes in S is connected with a hyperedge with a weight

$$\alpha = \beta \frac{\ell}{\binom{\ell}{i} r(i)} = \frac{\beta i}{\binom{\ell-1}{i-1} r(i)}, \quad \text{where} \quad \beta = k\frac{5\delta(j) + \delta(j+1)}{2} + kb\delta(i).$$

We will denote this set of hyperedges with E_1. For each $(u, v) \in A$ we connect u and v with $k(j-1)$ unique nodes in S with k hyperedges, each of size $j+1$. For each $u \in W$ we connect u with $kb(i-1)$ unique nodes in S with kb hyperedges, each of size i. We will denote this set of hyperedges with E_2, and the subset of the latter edges with E_2'.

Let us write

$$f_1(X) = \sum_{e \in E_1} r(|X \cap e|) \quad \text{and} \quad f_2(X) = \sum_{e \in E_2} r(|X \cap e|).$$

Let O be the optimal solution. Let $T = O \cap S$, and define $d = |T|$ and $c = |O \cap W|$.

We will prove the result as a sequence of lemmas.

Lemma 1 (Mediant inequality). *Assume 4 positive numbers a, b, c, and d. Then*

$$\frac{a+c}{b+d} \geq \frac{a}{b} \iff \frac{c}{d} \geq \frac{a}{b} \iff \frac{c}{d} \geq \frac{a+c}{b+d}.$$

Proof.

$$\frac{a+c}{b+d} \geq \frac{a}{b} \iff \frac{a+c}{b+d} - \frac{a}{b} \geq 0 \iff \frac{b(a+c)-(b+d)a}{b(b+d)} \geq 0$$
$$\iff ba + bc - ba - da \geq 0 \iff bc \geq da \iff \frac{c}{d} \geq \frac{a}{b}.$$

Similarly,

$$\frac{c}{d} \geq \frac{a+c}{b+d} \iff \frac{c}{d} - \frac{a+c}{b+d} \geq 0 \iff \frac{(b+d)c - d(a+c)}{b(b+d)} \geq 0$$
$$\iff bc + dc - da - dc \geq 0 \iff bc \geq da \iff \frac{c}{d} \geq \frac{a}{b},$$

proving the claim. □

Lemma 2. *Let O be the densest subgraph. Let $X \subsetneq O$. Then*

$$\frac{f(O) - f(O-X)}{|X|} \geq \Gamma(O).$$

Let Y be such that $Y \cap O = \emptyset$. Then

$$\frac{f(O \cup Y) - f(O)}{|Y|} \leq \Gamma(O).$$

Proof. For the first statement, apply Lemma 1 with $a = f(O-X)$, $b = |O-X|$, $c = f(O) - f(O-X)$, and $d = |X|$. By the optimality of O, we have

$$\frac{a+c}{b+d} = \frac{f(O)}{|O|} \geq \frac{f(O-X)}{|O-X|} = \frac{a}{b}.$$

Lemma 1 then implies

$$\frac{f(O) - f(O-X)}{|X|} = \frac{c}{d} \geq \frac{a+c}{b+d} = \Gamma(O).$$

For the second statement, apply Lemma 1 with $a = f(O \cup Y) - f(O)$, $b = |Y|$, $c = f(O)$, and $d = |O|$. By the optimality of O, we have

$$\frac{c}{d} = \frac{f(O)}{|O|} \geq \frac{f(O \cup Y)}{|O| + |Y|} = \frac{a+c}{b+d}.$$

Lemma 1 then implies

$$\Gamma(O) = \frac{f(O)}{|O|} = \frac{c}{d} \geq \frac{a}{b} = \frac{f(O \cup Y) - f(O)}{|Y|},$$

proving the claim. □

Lemma 3. *Let $v \in O \cap W$. Let $Z \subseteq S$ be the set of nodes connected to v by an edge in E_2'. There are at least k nodes shared by Z and O.*

Proof. Let $v \in O \cap W$. Lemma 2 implies

$$f(O) - f(O - v) \geq \Gamma(O) \geq \Gamma(S) \geq \beta \geq kb\delta(i).$$

Let x be the number of hyperedges $e \in E_2'$ for which $e \cap O = \{v\}$. Since there are $k(3 + b)$ edges adjacent to v we have

$$f(O) - f(O - v) \leq r(1)x + \Delta(k(3 + b) - x).$$

By definition of x, there are $kb - x$ edges in E_2' that contain at least one vertex in $Z \cap O$. Combining the two inequalities and solving for $kb - x$ leads to

$$kb - x \geq \frac{kb\delta(i) - k(3 + b)\Delta}{\Delta - r(1)} + bk = \frac{bk(\delta(i) - r_1) - 3k\Delta}{\Delta - r(1)} \geq k,$$

where the last inequality is due to the definition of b. The claim follows since the edges in E_2' are disjoint in S. □

Lemma 4. *$f_2(O) \leq R(3 + b)d$.*

Proof. Let Y be the edges in E_2 containing a vertex in $W \cap O$. Let Z be the edges in E_2 that contain a vertex in O but do not contain a vertex in $W \cap O$. Note that Y and Z are the only edges that contribute to $f_2(O)$. Let d_1 be the number of nodes in $O \cap S$ covered by Y, and let d_2 be the number of nodes in $O \cap S$ covered by Z. Note that $d = d_1 + d_2$.

Assume a vertex $v \in W \cap O$. There are $(3 + b)k$ hyperedges in Y attached to v. Lemma 3 implies that these edges contain k nodes in $O \cap S$. Since edges in Y are disjoint, $|Y| \leq (3 + b)d_1$.

Any edge in Z will have a vertex in $O \cap S$. Since these edges are disjoint in S, we have $|Z| \leq d_2$.

Consequently,

$$f_2(O) \leq R(|Y| + |Z|) \leq R((3 + b)d_1 + d_2) \leq R(3 + b)d,$$

proving the claim. □

Lemma 5. *Let $X \subseteq S$ be a non-empty set. Then*

$$f_1(X) \leq |X|r(1)\binom{\ell - 1}{i - 1} + |X|\theta\binom{|X| - 1}{i - 1}.$$

The inequality is tight if $X = S$.

Proof. Assume $v \in X$. Then

$$\begin{aligned} f_1(X) &= \sum_{e \in E_1} r(|X \cap e|) \\ &= |X| \sum_{e \in E_1; v \in e} \frac{r(|X \cap e|)}{|X \cap e|} \\ &\leq |X| \sum_{e \in E_1; v \in e} r(1) + \sum_{e \in E_1; v \in e \subseteq X} \theta \\ &= |X| r(1) \binom{\ell - 1}{i - 1} + |X| \theta \binom{|X| - 1}{i - 1}, \end{aligned}$$

where the inequality is due to the fact that $r(a)/a \leq r(1)$ for any $a < i$ and the definition of θ. Note that the inequality is tight if $X = S$. □

Lemma 6. *$S \subseteq O$.*

Proof. Assume otherwise. Let $O' = O \cup S$. Note that Lemma 3 guarantees that T is not empty.

To prove the claim, we need to show that $\Gamma(O') > \Gamma(O)$. We will do this by bounding the individual terms. We start with f_1, that can be bounded with

$$\begin{aligned} \frac{f_1(O')}{|O'|} - \frac{f_1(O)}{|O|} &= \frac{f_1(S)}{\ell + c} - \frac{f_1(T)}{d + c} \\ &\geq \frac{f_1(S)}{\ell + c} - \frac{f_1(T)}{d + \frac{d}{\ell} c} && (d \leq \ell) \\ &= \frac{\ell}{\ell + c} \left(\frac{f_1(S)}{\ell} - \frac{f_1(T)}{d} \right) \\ &\geq \frac{1}{2} \left(\frac{f_1(S)}{\ell} - \frac{f_1(T)}{d} \right) && (\ell \geq m \geq c) \\ &\geq \frac{1}{2} \theta \left(\binom{\ell - 1}{i - 1} - \binom{d - 1}{i - 1} \right). && \text{(Lemma 5)} \end{aligned}$$

Consequently,

$$\frac{\alpha f_1(O')}{|O'|} - \frac{\alpha f_1(O)}{|O|} \geq \frac{\theta \beta i}{2r(i)} \left(1 - \binom{d - 1}{i - 1} \Big/ \binom{\ell - 1}{i - 1} \right) \geq \frac{\theta i}{2r(i)} \beta \frac{\ell - d}{\ell - 1}, \qquad (1)$$

where the second inequality is due to the fact that

$$\binom{d - 1}{i - 1} \Big/ \binom{\ell - 1}{i - 1} \leq \frac{d - 1}{\ell - 1}.$$

Next we bound f_2 terms with

$$\begin{aligned}
\frac{f_2(O')}{|O'|} - \frac{f_2(O)}{|O|} &\geq \frac{f_2(O)}{|O'|} - \frac{f_2(O)}{|O|} \\
&= \frac{f_2(O)}{d}\left(\frac{d}{\ell + c} - \frac{d}{d + c}\right) \\
&\geq R(3+b)\left(\frac{d}{\ell + c} - \frac{d}{d + c}\right) \qquad \text{(Lemma 4)} \\
&= R(3+b)d\frac{d - \ell}{(\ell + c)(d + c)} \\
&\geq R(3+b)d\frac{d - \ell}{d\ell} \\
&= R(3+b)\frac{d - \ell}{\ell}.
\end{aligned} \tag{2}$$

Note that the latter two inequalities rely on the fact that $d \leq \ell$.

To match the coefficients in the bounds, we can use the definition of k and β and show that

$$\beta \geq kb\delta(i) > kbr(1) \geq kb\frac{r(i)}{i} \geq \frac{6bRr(i)}{\theta i} > \frac{2(3+b)Rr(i)}{\theta i}. \tag{3}$$

Finally, we have

$$\begin{aligned}
\Gamma(O') - \Gamma(O) &= \frac{\alpha f_1(O')}{|O'|} - \frac{\alpha f_1(O)}{|O|} + \frac{f_2(O')}{|O'|} - \frac{f_2(O)}{|O|} \\
&\geq \frac{\theta i}{2r(i)}\beta\frac{\ell - d}{\ell - 1} + R(3+b)\frac{d - \ell}{\ell} \qquad \text{(Eqs. 1–2)} \\
&> \left(\frac{\theta i}{2r(i)}\beta - R(3+b)\right)\frac{\ell - d}{\ell} \\
&> 0, \qquad \text{(Eq. 3)}
\end{aligned}$$

which contradicts the optimality of O. □

Lemma 7. *Let $(u, v) \in A$. If $u \in O$, then $v \notin O$.*

Proof. Let a be the number of neighbors of u in H that are in O. The fact that $S \subseteq O$ and Lemma 2 imply

$$\beta \leq \Gamma(S) \leq \Gamma(O) \leq f(O) - f(O - u) = kb\delta(i) + ak\delta(j+1) + (3-a)k\delta(j).$$

Now the definition of β results in

$$2.5 \times \delta(j) + 0.5 \times \delta(j+1) \leq a\delta(j+1) + (3-a)\delta(j).$$

Since $\delta(j) > \delta(j+1)$, this is only possible when $a = 0$. □

Proof (Proof of Theorem 1). Lemma 7 shows that $O \cap W$ is an independent set. We prove the claim by showing that $O \cap W$ is the maximum independent set.

Since $S \subseteq O$, we can rewrite $f_2(O)$ as

$$\begin{aligned} f_2(O) &= ckbr(i) + 3ckr(j) + (n-c)kbr(i-1) + (m-3c)kr(j-1) \\ &= ckb\delta(i) + 3ck\delta(j) + \omega, \end{aligned}$$

where $\omega = nkbr(i-1) + mkr(j-1)$.

The bound

$$\begin{aligned} \frac{\alpha\binom{\ell}{i}r(i) + \omega}{\ell} &= \beta + \frac{\omega}{\ell} \\ &\le \beta + R \\ &\le \beta + k\eta/4 \\ &= k\frac{11\delta(j) + \delta(j+1)}{4} + kb\delta(i) \\ &\le kb\delta(i) + 3k\delta(j) \end{aligned}$$

together with Lemma 1 imply that

$$\Gamma(O) = \frac{\alpha\binom{\ell}{i}r(i) + \omega + ckb\delta(i) + 3ck\delta(j)}{\ell + c}$$

is maximized when c is maximized. □

B Projection-based Method Approximation Guarantee

Proposition 1. *Let $r\colon [0,k] \to \mathbb{R}^+$ be a monotonic reward function satisfying $r(0) = 0$. There exists a nonnegative monotonic convex function $\hat{r}$ satisfying $\hat{r}(i) \le r(i) \le k \cdot \hat{r}(i)$ for every $i \in \{1, 2, \ldots, k\}$.*

Proof. To prove the result, we explicitly describe a procedure for finding the optimal $\hat{r}$. Although not the most efficient in terms of runtime, the procedure is conceptually simple and allows us to prove the desired bound.

At step 0, we initialize the procedure by setting $t_0 = 0$ and $\hat{r}(0) = 0$. At the beginning of step i, we assume we have already defined $\hat{r}$ on the interval $[0, t_{i-1}]$ for some positive integer t_{i-1}. If $t_{i-1} < k$, we then find a new point t_i as

$$t_i = \min_{j \in \{t_{i-1}+1, \ldots, k\}} \frac{r(j) - r(t_{i-1})}{j - t_{i-1}}. \tag{4}$$

In other words, we find the line connecting $(t_{i-1}, r(t_{i-1}))$ to another point $(j, r(j))$ with the smallest possible slope. We then define $\hat{r}$ on the interval $[t_{i-1}, t_i]$ to be the line segment from $(t_{i-1}, r(t_{i-1}))$ to $(t_i, r(t_i))$. This continues until an iteration m where $t_m = k$.

The proof follows by showing three properties for this procedure:

(a) $\hat{r}(j) \leq r(j)$ for $j \in \{0, 1, \dots k\}$,
(b) $\hat{r}$ is convex, and
(c) $r(j) \leq k \cdot \hat{r}(j)$ for $j \in \{0, 1, \dots k\}$.

Property (a) follows from the fact that at each iteration, we chose the line segment with the smallest possible slope. To prove Property (b), consider the points $\{t_{i-1}, t_i, t_{i+1}\}$ identified in three consecutive iterations of the procedure. Using the minimality property of t_i in Eq. (4), we have

$$\begin{aligned}\frac{r(t_i) - r(t_{i-1})}{t_i - t_{i-1}} \leq \frac{r(t_{i+1}) - r(t_{i-1})}{t_{i+1} - t_{i-1}} &= \frac{[r(t_{i+1}) - r(t_i)] + [r(t_i) - r(t_{i-1})]}{[t_{i+1} - t_i] + [t_i - t_{i-1}]} \\ &= \alpha \frac{r(t_{i+1}) - r(t_i)}{t_{i+1} - t_i} + (1 - \alpha) \frac{r(t_i) - r(t_{i-1})}{t_i - t_{i-1}},\end{aligned}$$

where the last equation must hold for some $\alpha \in [0, 1]$. Rearranging gives

$$\frac{r(t_i) - r(t_{i-1})}{t_i - t_{i-1}} \leq \frac{r(t_{i+1}) - r(t_i)}{t_{i+1} - t_i}.$$

In other words, the slopes from iteration i to iteration $i + 1$ are nondecreasing. Since $\hat{r}$ is a piecewise linear function with nondecreasing slopes, it is convex.

Finally, to show Property (c), it is sufficient to show that the bound holds for each individual line segment. On interval $[t_i, t_{i+1}]$, $\hat{r}$ is defined by the line

$$\hat{r}(x) = \frac{r(t_{i+1}) - r(t_i)}{t_{i+1} - t_i}(x - t_i) + r(t_i).$$

For $j \in \{t_i + 1, \cdots, t_{i+1} - 1\}$, we have

$$\begin{aligned}\frac{r(j)}{\hat{r}(j)} &= \frac{r(j)(t_{i+1} - t_i)}{(r(t_{i+1}) - r(t_i))(j - t_i) + r(t_i)(t_{i+1} - t_i)} \\ &= \frac{r(j)(t_{i+1} - t_i)}{r(t_{i+1})(j - t_i) + r(t_i)(t_{i+1} - j)} \\ &\leq \frac{r(j)}{r(t_{i+1})} \cdot \frac{(t_{i+1} - t_i)}{(j - t_i)} \leq k.\end{aligned}$$

The last inequality follows from the monotonicity of r and the fact that $j - t_i \geq 1$ and $t_{i+1} - t_i \leq k$. □

C Approximation Guarantee for CARD-SWAMP

Theorem 2. *Assume that we can α-approximate SWAMP, then Algorithm 2 yields $\frac{\alpha}{\alpha+1}$ approximation for CARD-SWAMP. Consequently, using Algorithm 1 together with Algorithm 2 yields $\frac{1}{k+1}$ approximation.*

Proof. Let us write $\beta = \alpha/(1 + \alpha)$.

Let O be the optimal solution, and let S_i, H_i, S' and (final) S be as used in Algorithm 2.

Let Γ_i be the density function in H_i, and let T_i be the subgraph optimizing the density Γ_i.

Define $W_i = \bigcup_{j<i} S_j$ be the nodes removed from H to obtain H_i. Let c be the index of the last S_i added, and write $W = W_c$.

We split the proof in two cases:

Assume that $f(W) \leq \beta f(O)$. We have immediately

$$\begin{aligned}\Gamma_i(T_i) &\geq \Gamma_i(O \setminus W_i)\\ &= \frac{f(O \cup W_i) - f(W_i)}{|O \setminus W_i|} \geq \frac{f(O) - f(W)}{|O \setminus W_i|} \geq \frac{1}{\alpha + 1}\frac{f(O)}{|O \setminus W_i|} \geq \frac{\Gamma(O)}{\alpha + 1}.\end{aligned}$$

Consequently, $\Gamma_i(S_i) \geq \beta\Gamma(O)$ and

$$\Gamma(S) = \frac{\sum_i f(S_i \cup W_i) - f(W_i)}{\sum_i |S_i|} = \frac{\sum_i |S_i|\Gamma_i(S_i)}{\sum_i |S_i|} \geq \beta\frac{\sum_i |S_i|\Gamma(O)}{\sum_i |S_i|} = \beta\Gamma(O),$$

proving the first case.

Assume now that $f(W) \geq \beta f(O)$. Then

$$\Gamma(S') = \frac{f(S')}{\ell} \geq \frac{f(W)}{\ell} \geq \beta\frac{f(O)}{\ell} \geq \beta\Gamma(O),$$

proving the second case. $\square$