A. Omitted Proofs

A.1. Proof of Lemma 1

Proof. Denote the solutions to the two different objectives as follows: $\hat{S}_r = \arg\max_{S:|S|\le k} \min_{\mathbf{p}\in\mathcal{P}} \frac{f_{\mathbf{p}}(S)}{f_{\mathbf{p}}(S^*_{\mathbf{p}})}$ and $\hat{S}_v = \arg\max_{S:|S|\le k} \min_{\mathbf{p}\in\mathcal{P}} f_{\mathbf{p}}(S)$.

We will prove the lemma by contradiction. Specifically, let's assume that there exists a set of influence functions $\mathcal{P}$ for which $\min_{\mathbf{p}} f_{\mathbf{p}}(\hat{S}_r) < \frac{1}{\sqrt{n}} \min_{\mathbf{p}} f_{\mathbf{p}}(\hat{S}_v)$, i.e. for this $\mathcal{P}$, the solution for the robust ratio objective is suboptimal with respect to the total number of nodes influenced by a factor greater than $\sqrt{n}$.

To ease the notation let us denote with f_r the function that achieves

Hence it holds:

$$\begin{aligned}\sqrt{n}\cdot f_r(\hat{S}_r) &= \sqrt{n}\cdot \min_{\mathbf{p}} f_{\mathbf{p}}(\hat{S}_r)\\ &< \min_{\mathbf{p}} f_{\mathbf{p}}(\hat{S}_v) = f_v(\hat{S}_v) \le f_r(\hat{S}_v) \quad (3)\end{aligned}$$

where the last inequality is due to the minimality of f_v. Let us denote with f_m the function that has the minimum ratio for $\hat{S}_v$. That is, $\frac{f_m(\hat{S}_v)}{f_m(S^*_m)} = \min_{\mathbf{p}} \frac{f_{\mathbf{p}}(\hat{S}_v)}{f_{\mathbf{p}}(S^*_{\mathbf{p}})}$. Then,

$$\begin{aligned}\frac{1}{\sqrt{n}} &> \frac{f_r(\hat{S}_r)}{f_r(\hat{S}_v)} \ge \frac{f_r(\hat{S}_r)}{f_r(S^*_r)}\\ &= \max_S \min_{\mathbf{p}} \frac{f_{\mathbf{p}}(S)}{f_{\mathbf{p}}(S^*_{\mathbf{p}})} \ge \min_{\mathbf{p}} \frac{f_{\mathbf{p}}(\hat{S}_v)}{f_{\mathbf{p}}(S^*_{\mathbf{p}})} = \frac{f_m(\hat{S}_v)}{f_m(S^*_m)} \quad (4)\end{aligned}$$

where the last inequality holds due to the maximality of $\hat{S}_r$. Now we can prove a contradiction as follows:

$$\begin{aligned}f_m(S^*_m) &> \sqrt{n}\cdot f_m(\hat{S}_v) \ge \sqrt{n}\cdot f_v(\hat{S}_v)\\ &> n\cdot f_r(\hat{S}_r) \ge n\end{aligned}$$

The first inequality is due to (4), the second is due to the fact that $f_v(\hat{S}_v) = \arg\min_{\mathbf{p}} f_{\mathbf{p}}(\hat{S}_v)$, while the third is from (3). Finally, since $|\hat{S}_r| \ge 1$ the influence function is also at least 1 (at least all the nodes in $\hat{S}_r$ get influenced).

Now notice that the influence of any set of nodes cannot be more than n and as a result we have a contradiction. Thus, $\min_{\mathbf{p}} f_{\mathbf{p}}(\hat{S}_r) \ge \frac{1}{\sqrt{n}} \min_{\mathbf{p}} f_{\mathbf{p}}(\hat{S}_v)$.

The graph in Figure 2 shows that there exist a set $\mathcal{P}$ for which $\min_{\mathbf{p}} f_{\mathbf{p}}(\hat{S}_r) = \Omega\left(\frac{1}{\sqrt{n}}\right) \min_{\mathbf{p}} f_{\mathbf{p}}(\hat{S}_v)$ which concludes the proof.

□

A.2. Proof of Lemma 3

Consider a cycle on n nodes, connected with edges of diffusion probability $1-\lambda$, and an additional center node $v^\star$ that is connected to all the nodes of the cycle. Notice that the number of edges is $m = 2n$. To consider the Lipschitzness of the influence function on this graph we consider the change in the influence of $v^\star$ in case that the probabilities connecting it to the cycle are all λ and the case in which they are all $\epsilon = n\cdot\lambda$.

The influence of $v^\star$ in the first case is at most $n(1-(1-\lambda)^n)$. For sufficiently large n:

$$(1-\lambda)^n = \left(1 - \frac{n\lambda}{n}\right)^n \approx e^{-n\lambda} \approx 1 - n\lambda$$

So, the influence is at most $n^2\lambda = n\epsilon$. Once the probabilities on edges connecting it to the cycle increase from λ to ϵ its expected influence becomes at least $n(1-(1-\epsilon)^n)(1-\lambda)^n$. Using the same approximation as before for sufficient large n, we get that the influence of $v^\star$ is at least $n^2\epsilon(1-n\lambda) = n^2\epsilon(1-\epsilon) = n^2\epsilon - n^2\epsilon^2$.

Then, $|f_{\mathbf{p}}(v^\star) - f_{\mathbf{p'}}(v^\star)| = n^2\epsilon - n^2\epsilon^2 - n\epsilon$. After setting $\epsilon = \frac{1}{n}$, this bound becomes $n^2\epsilon - 2n\epsilon = (n\frac{m}{2} - 2n)\epsilon$. Thus, for small ϵ the Lipschitz constant is asymptotically achieved with n.

Notice that this example can be simplified if we use probabilities of 1 in the cycle, and 0 and ϵ in the the connections of $v^\star$. The reason why we avoided the values $0, 1$ is because for some generalized linear models, e.g. in the logistic or the probit model the values of the probabilities are strictly in $(0, 1)$ instead of $[0, 1]$.

A.3. Proof of Lemma 5

In Definition 2 we defined an ϵ-cover of $\mathcal{F}$ as a set $\mathcal{F}_\epsilon \subset F$ s.t. for any $f_\theta \in \mathcal{F}$ there exists a function $f_j \in \mathcal{F}_\epsilon$ such that: $|f_\theta(S) - f_j(S)| \le \epsilon$ for all $S \subseteq V$. Using this definition we can proceed as follows:

$$\forall S \subseteq V, \forall f_\theta \in \mathcal{F}, \exists f_j \in \mathcal{F}_\epsilon : |f_\theta(S) - f_j(S)| \le \epsilon$$

$$\Rightarrow -\min_{f_\theta\in\mathcal{F}} f_\theta(S) + f_j(S) \le \epsilon$$

$$\Rightarrow \min_{f_\theta\in\mathcal{F}} f_\theta(S) \ge \min_{f_i\in F} f_i(S) - \epsilon$$

Simultaneously it holds that:

$$\forall S \subseteq V,\ \min_{f_i\in\mathcal{F}_\epsilon} f_i(S) - \min_{f_\theta\in\mathcal{F}} f_\theta(S) \ge 0$$

since $\mathcal{F}_\epsilon \subset \mathcal{F}$. Let $S^* = \arg\max_{S:|S|\le k} \min_{f_\theta\in\mathcal{F}} f_\theta(x)$ and $S^*_\epsilon = \arg\max_{S:|S|\le k} \min_{f_i\in\mathcal{F}_\epsilon} f_i(x)$. Then it is:

$$\min_{f_i \in \mathcal{F}_\epsilon} f_i(S^*_\epsilon) - \min_{f_\theta \in \mathcal{F}} f_\theta(S^*) \geq$$
$$\min_{f_i \in \mathcal{F}_\epsilon} f_i(S^*) - \min_{f_\theta \in \mathcal{F}} f_\theta(S^*) \geq 0$$
$$\Rightarrow \max_{S:|S|\leq k} \min_{f_i \in \mathcal{F}_\epsilon} f_i(S) \geq \max_{S:|S|\leq k} \min_{f_\theta \in \mathcal{F}} f_\theta(S)$$

Hence, utilizing an algorithm that guarantees an $\hat{S} \subseteq V$ such that:

$$\min_{f_i \in \mathcal{F}_\epsilon} f_i(\hat{S}) \geq \alpha \cdot \max_{S:|S|\leq k} \min_{f_i \in \mathcal{F}_\epsilon} f_i(S)$$

we get that for the family $\mathcal{F}$ it holds:

$$\min_{f_\theta \in \mathcal{F}} f(\hat{S}) \geq \alpha \cdot \max_{S:|S|\leq k} \min_{f_\theta \in \mathcal{F}} f(S) - \epsilon.$$

□

B. Lower Bound

We build a similar reduction to the one in (He & Kempe, 2016), reducing from is GAP SET COVER.

In a SET COVER instance we have a universe elements $U = \{u_1, u_2, \ldots, u_\ell\}$ and a collection of subsets of U, $\mathcal{T} = \{T_1, T_2, \ldots, T_M\}$, where $T_i \subseteq U$ for all $i \in [M]$. The goal is to find a cover $C \subseteq \mathcal{T}$ such that $\cup_{T\in C} T = U$ and the size of C is minimized. In the decision version of the problem we are also given an integer k and we are asked whether the optimal solution has value $|C| \leq k$ or $|C| > k$. The GAP SET COVER is a slightly stronger problem that asks whether there is a solution C such that $|C| \leq k$ or $|C| > (1-\delta)\log Nk$, for any $\delta \in (0,1)$. We will assume that $k \leq \min\{M, \ell\}$, otherwise we can always find the optimal solution by simply picking all the elements of $\mathcal{T}$ or at least one set per element of U that contains it (assuming that a set cover exists, such a set always exists as well). Both problems are NP-hard as proved in (Karp, 1972) and (Dinur & Steurer, 2014).

For any given instance of GAP SET COVER we construct an instance of robust influence maximization (RIM) by constructing a graph on n nodes, and ℓ different influence functions. The goal is to maximize the influence with respect to the worse influence function. Each influence function is associated with a different set of diffusion probabilities. We will prove that if we can find a seed set that is a better than $\frac{1}{n^{1-\epsilon}}$-approximation to the maximin solution of this RIM problem, then we can solve gap set cover.

We construct the following bipartite graph with vertex set $V = A \cup B$. The set A contains exactly M nodes, one node a_T for each T, $i \in [M]$. The set B contains m nodes (m to be fixed later in the proof) for each element $u \in U$: $\{b_{u1}, b_{u2}, \ldots, b_{um}\}$, so $m\ell$ nodes in total. The total number of nodes in the graph is $n = M + m\ell$.

We create the edges of the graph according to the the set cover solution C. For every $T \in C$ we add the directed edges from a_T to $\{b_{u1}, b_{u2}, \ldots, b_{um}\}$ for all $u \in T$. That is $m|T|$ edges per element $T \in C$.

Each influence function induces different probabilities on the edges. We have ℓ functions. For the u^{th} function, set the probability of the edges $\{(a_T, b_{u1}), (a_T, b_{u2}), \ldots, (a_T, b_{um})\}$ for which $u \in T$ to $1-\lambda$ and the probability of the rest of the edges to λ.

There are two cases: $|C| \leq k$ and $|C| > (1-\delta)\ln \ell k$. Let us focus on the case where $|C| \leq k$ first. One can easily see that if we choose the a_Ts for which $T \in C$ as seeds, we can achieve expected diffusion of at least $|C| + (1-\lambda)m$ on each of the ℓ influence functions due to the fact that every $j \in [\ell]$, u_j is covered by the solution C, i.e. there exists T such that $u_j \in T$. Thus, in this case $\max_S \min_{j\in[\ell]} f_j(S) \geq |C| + (1-\lambda)m$.

In the second case, there is no cover of size at most $(1-\delta)\ln \ell k$. However, we are allowed to choose at most $(1-\delta)\ln \ell k$ as seeds. Hence, for any choice of seeds there is definitely an element $u_j \in U$ that is not covered. As a result, for the j^{th} influence function the expected number of influenced nodes is at most $(1-\delta)\ln \ell k \cdot (1 + m(\ell-1)\lambda)$. That is because each node is connected to at most $m(\ell-1)\lambda$ other nodes (since it is definitely not connected to u_j) and there are no high probability edges that are triggered in this function. As a result $\max_S \min_{j\in[\ell]} f_j(S) \leq (1-\delta)\ln \ell k \cdot (1 + m(\ell-1)\lambda)$.

We want to be able to distinguish between the first and the second case, i.e. the case where $\max_S \min_{j\in[\ell]} f_j(S) \geq |C| + (1-\lambda)m$ and when $\max_S \min_{j\in[\ell]} f_j(S) \leq (1-\delta)\ln \ell k \cdot (1+m(\ell-1)\lambda)$. To this end, we consider the ratio: $\frac{(1-\delta)\ln \ell k(1+m(\ell-1)\lambda)}{|C|+(1-\lambda)m}$ which we want to prove that is less than $\frac{1}{n^{1-\epsilon}}$ for any $\epsilon > 0$. Remember that for the number of nodes in the graph it holds $n = M + m\ell$.

Hence, if $\frac{|C|+(1-\lambda)m}{n^{1-\epsilon}} > (1-\delta)\ln \ell k(1 + m(\ell-1)\lambda)$ we will be able to separate the two cases.

First assume $\min\{M, \ell\} = \ell$:

$$\frac{|C| + (1-\lambda)m}{(M+m\ell)^{1-\epsilon}} \geq (1-\lambda)\frac{1+m}{((1+m)M)^{1-\epsilon}}$$
$$\geq (1-\lambda)\frac{m^\epsilon}{M^{1-\epsilon}}$$

Now for $m = M^{3/\epsilon}, \lambda = 1/m$ we have that as M grows

$$\left(1 - \frac{1}{m}\right)\frac{m^\epsilon}{M^{1-\epsilon}} > (1-\delta)M^{2+\epsilon} >$$
$$(1-\delta)M^2 \ln M$$

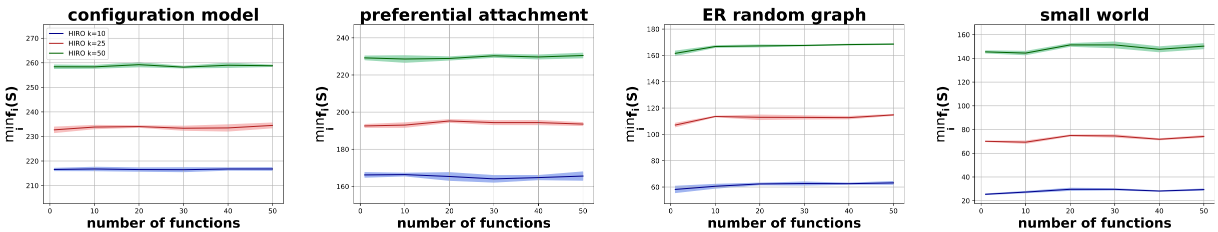

Figure 6. Number of functions needed to cover the Hyperparameter's space.

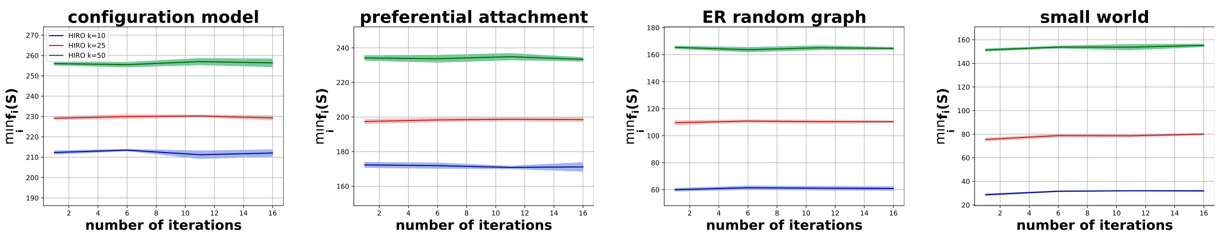

Figure 7. Speed of convergence of HIRO for multiple seed set sizes.

which is asymptotically larger than $(1-\delta)\ln \ell k(1+m(\ell-1)\lambda)$. Choosing $m=\ell^{3/\epsilon}$ solves the other case. Hence setting $m=(\max\{M,\ell\})^{3/\epsilon}$ and $\lambda=1/m$ completes the reduction. Hence, if we have a better than $\frac{1}{n^{1-\epsilon}}$-approximation algorithm for robust influence maximization then we can solve gap set cover.

C. Omitted Details from Experiments

Synthetic Graphs: As we discussed in Section 6 different graph models yield graphs with different topological properties. The ones we selected for our experiments are the following:

- *Small-World network:* In this model most nodes are not neighbors of one another, but the path from each node to another is short. Specifically we use the Watts-Strogatz model that is known for its high clustering coefficient and small diameter properties. Each node is connected to 5 nodes in the ring topology and the probability of rewiring an edge is $1/n$ where n is the number of nodes in the graph. We work with graphs of sizes 100-250.
- *Preferential Attachment (Barabási-Albert):* The degree distribution of this model is a power law and hence captures interesting properties of the real-world social networks. We took 2 initial vertices and added 2 edges at each step, using the preferential attachment model, until we reached 100-250 vertices.
- *Configuration model:* The configuration model allows us to construct a graph with a given degree distribution. We chose 100-250 vertices and a power-law degree distribution with parameter $\alpha = 2$.
- *Erdös-Rényi:* We used the celebrated $G(n,m)$ model to create a graph with 100-250 vertices and edges with probability $p = 3/n$. $G(n,m)$ does not capture some of the properties of real social networks, however it is a very impactful model with variety of applications in several areas of science.