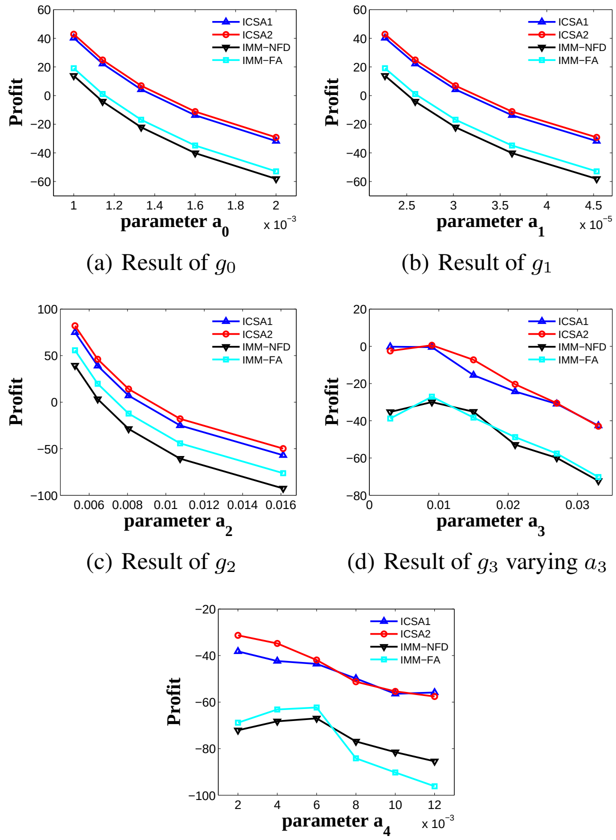

(a) Result of g_0 (b) Result of g_1

(c) Result of g_2 (d) Result of g_3 varying a_3

(e) Result of g_3 varying a_4

Figure 1: Profit with Different Functions: g_0 : $B_i = \frac{1}{a_0} f_i$, g_1: $B_i = \frac{1}{a_1} f_i^2$, g_2: $B_i = \frac{1}{a_2} f_i^{\frac{1}{2}}$ and g_3: $B_i = \frac{a_5}{1+e^{-(f_i - a_3)/a_4}}$.

1 Experimental Results on Epinions Dataset

We conduct experiments on the dataset of Epinions[1], which has 75K nodes and 508K directed edges. The results with the four functions with parameters varied are shown in Figure 1. In all the cases, we find the results are very similar to those of dataset Gowalla, which demonstrate the effectiveness of the proposed methods.

2 Proof of Theorems

2.1 Proof of Theorem 1

We prove the theorem by a reduction from 3-PARTITION [Garey and Johnson, 1986]. Given a set $X = \{x_1, ..., x_{3m}\}$ of positive integers whose sum is C, with $x_i \in (C/4m, C/2m), \forall i$, 3-PARTITION asks whether X can be partitioned into m disjoint 3-element subsets, such that the sum of elements in each partition is the same ($= C/m$). This problem is known as strongly NP-hard even if the integers x_i are bounded above by a polynomial in m.

Given an instance $\mathcal{I}$ of 3-PARTITION, we construct an instance $\mathcal{J}$ of HPM problem as follows. First, we set the number of marketers $K = m$ with $B_i = C/m, f_i = 1/m, \forall i$. Then, we construct a directed bipartite graph $G = (U \cup V, E)$: for each number x_i, G has one node $u_i \in U$ with $x_i - 1$ out-neighbors in V. Let spread probability on each edge and the threshold of each node all be 1. Let the cost of selecting each node as a seed be $C/3m$. This reduction will result in a total of C nodes in the instance $\mathcal{J}$. Since C is bounded by a polynomial in m, the reduction is achieved in polynomial time. Note after such reduction, any seed set S_i consists of node in U has an influence spread of $\sigma(S_i|\mathbf{S}) = \sum_{u_j \in S_i} \sigma(u_j|\mathbf{S}) = \sum_{u_j \in S_i} x_j$.

Suppose algorithm $\mathcal{A}$ can solve HPM problem optimally in polynomial time. Then we can use $\mathcal{A}$ to distinguish between YES- and NO-instances of 3-PARTITION as follows. Run $\mathcal{A}$ on $\mathcal{J}$ to yield an allocation $\mathbf{S} = (S_1, ..., S_m)$. Then $\mathcal{I}$ is a YES-instance of 3-PARTITION iff for all i, $\sigma(S_i|\mathbf{S}) = \sum_{u_j \in S_i} x_j = B_i = C \cdot f_i = \frac{C}{m}$.

($\Longrightarrow$): Suppose the above equation holds for all i. In this case, each S_i must consist of 3 nodes in U, whose influence spread sums to C/m. Then the allocation itself witness that $\mathcal{I}$ is a YES-instance. Suppose $|S_i| \neq 3$ for some i. Since $\forall i, x_i \in (C/4m, C/2m)$, then $\sigma(S_i|\mathbf{S}) = \sum_{u_j \in S_i} x_j \neq C/m$, which makes a contradictory.

($\Longleftarrow$): Suppose $S_1, ..., S_m$ are disjoint 3-element subsets that each sum to C/m. By using $(S_1, ..., S_m)$ as the allocation, we must obtain an optimal solution to HPM problem. It is trivial to see that removing any node from any set, or changing any two nodes with different influence spread between two sets will break the satisfactory of some f_i.

We just proved that HPM problem is NP-hard. To see hardness of approximation, suppose $\mathcal{B}$ is an algorithm that approximates HPM problem within a factor of α. That is, the profit achieved by algorithm $\mathcal{B}$ on any instance of HPM problem is $\geq \alpha \cdot OPT$, where OPT is the optimal (maximum) profit. See the above instance $\mathcal{J}$ of which the optimal profit is 0. On this instance, the profit achieved by algorithm $\mathcal{B}$ is $\geq \alpha \cdot 0 = 0$, i.e., algorithm $\mathcal{B}$ can solve HPM problem optimally in polynomial time, which is shown to be impossible unless $P = NP$.

2.2 Proof of Theorem 2

We consider an allocation $\mathbf{S} = (S_1, S_2, ..., S_K)$ yielded by the NFD algorithm. First, we have the following result.

Lemma 1. *If $\sigma(S_i|\mathbf{S}) \geq 2f_i$, then $|S_i| = 1$.*

Proof. Assume $\sigma(S_i|\mathbf{S}) \geq 2f_i$ and $|S_i| \geq 2$. Suppose the last allocated seed v_j to d_i makes $\sigma(S_i) > f_i$. Then we have $\sigma(S_i|\mathbf{S}) - \sigma(v_j) < f_i$ and $\sigma(S_i|\mathbf{S}) \geq f_i$. If $\sigma(S_i|\mathbf{S}) \geq 2f_i$, then it follows that $\sigma(v_j) > f_i$. Since NFD allocates candidate seeds in non-increasing order, for all $v_{j'} \in S_i \backslash \{j'\}$ it holds $\sigma(v_j) < f_i$, which makes a contradiction. Thus we have $|S_i| = 1$ if $\sigma(S_i|\mathbf{S}) \geq 2f_i$. □

Let i' be the smallest index with $\sigma(S_{i'}\mathbf{S}) \geq 2f_{i'}$ and $v_{j'}$ be the seed allocated to i'. Let $D' = D \backslash \{d_{i'}\}$ and $S' = S_{cand} \backslash \{v_{j'}\}$. Then removing $d_{i'}$ and $v_{j'}$ will lead the approximation ratio of NFD decrease.

Lemma 2. $\frac{NFD(D, S_{cand})}{OPT(D, S_{cand})} \geq \frac{NFD(D', S')}{OPT(D', S')}$.

[1] http://snap.stanford.edu/data/

Proof. Consider the smallest index i' with $\sigma(S_{i'}|\mathbf{S}) \geq 2f_{i'}$ and $v_{j'}$ be the seed allocated to $d_{i'}$. By the ordering, for every d_i with $i > i'$, if we allocate $v_{j'}$ to d_i, we have that $\sigma(v_{j'}) - f_i \geq \sigma(v_{j'}) - f_{i'}$. Because of this and because every d_i with $i \geq i'$ can be satisfied with only the candidate seed $v_{j'}$, we can assume an optimal algorithm would assign $v_{j'}$ to an advertiser d_i with i at most be i'.

If an optimal algorithm decides to allocate $v_{j'}$ to d_i with $i < i'$, then it would not allocate all seeds of $v_1, ..., v_{j'-1}$ to $d_1, ..., d_{i'-1}$. This is because also NFD allocates the $v_1, ..., v_{j'-1}$ to $d_1, ..., d_{i'-1}$ and these are already satisfied, but $d_{i'}$ would not, in this case. Hence at least one of the seeds in $\{v_1, ..., v_{j'}\}$ would be allocated to d_i with $i \geq i'$, otherwise this allocation would not be optimal. With the same argumentation as above, for every such v_j, if it would be allocated to d_i with $i \geq i'$, we have $\sigma(v_j) - f_i \geq \sigma(v_j) - f_{i'}$. Because every such v_j satisfies every d_i with $i \geq i'$ singularly. We can thus assume v_j is allocated to i'. But then, since $\sigma(v_j) - f_{i'} \geq \sigma(v_{j'}) - f_{i'}$, we also can assume $v_{j'}$ is allocated by OPT to $d_{i'}$. Hence we have $\frac{NFD(D',S')}{OPT(D',S')} \leq \frac{NFD(D',S')+f_{i'}}{OPT(D',S')+f_{i'}} = \frac{NFD(D,S_{cand})}{OPT(D,S_{cand})}$, since $OPT(D',S') \geq NFD(D',S') > 0$. □

Iteratively applying Lemma 2, we can see all $\sigma(S_i|\mathbf{S}) > 2f_i$ are also the allocation of OPT. Then to derive a lower bound of the approximation ratio, we can assume w.l.o.g. that for all satisfied d_i, $\sigma(S_i|\mathbf{S}) \leq 2f_i$. We assume at least one d_i is satisfied, otherwise we obtain the optimal value of 1. Let k be the number of satisfied advertisers.

Lemma 3. *If NFD gives a solution that satisfies k marketers, then* $OPT(D, S_{cand})/NFD(D, S_{cand}) \leq 2 + 1/k$.

Proof. Let i' be the largest index of a d_i which is satisfied, i.e., $d_{i'}$. Let D_I be the set of all satisfied d_i. Then we have $NFD(D, S_{cand}) \geq \sum_{d_i \in D_I} f_i$ and $NFD(D, S_{cand}) \geq kf_{i'}$. See we have for every $d_i \in D_I$ that $\sigma(S_i|\mathbf{S}) \leq 2f_i$. Since $\sigma(S_{i'+1}|\mathbf{S}) = 0$, we have $\sum_{j=l}^{N} \sigma_{v_j} < f_{i'}$, otherwise NFD would have satisfied $d_{i'+1}$. It follows that $OPT \leq \sum_{j=1}^{N} \sigma(v_j) < \sum_{d_i \in I} 2f_i + f_{i'} \leq (2 + 1/k)NFD$. □

We can see if $k \geq 4$, the 4/9-approximation is guaranteed by Lemma 3. In the following, we discuss the cases with $k \leq 3$.

Case 1: $k = 1$**.** W.o.l.g., we assume d_1 is satisfied. Let t be the number of seeds allocated to advertiser d_1 which is satisfied. Observe that these seeds have in sum an influence of at most $\frac{t}{t-1}f_1$, which is by the ordering of seeds and the fact that NFD does not allocate seeds to already satisfied d. With the argument that the seed $v_{t+1}, ..., v_N$ could not satisfy any of $d_2, ..., d_K$ in NFD's solution, we have $\sum_{l=t+1}^{N} \sigma(v_l) < f_2 \leq f_1$. Altogether we can bound $OPT < \frac{t}{t-1}f_1 + f_1 =: f(t)$. As f is a monotone decreasing function, we only need to discuss whether $f(t) \geq 9/4f_1$ for $2 \leq t \leq 4$. If $t = 2$, the possible OPT is $f_1 + f_2 \leq 2f_1 < 9/4f_1$. Then we discuss the case of $t = 3$. If OPT wants to gain more than twice the profit NFD gains, it has to satisfy at least three d_i, which is because NFD satisfies the largest one in the instance. For ease of notation we relabel the d_i OPT satisfies as d_2, d_3 and d_4 in non-increasing order of f_i. It suffices to show that $f_2 + f_3 + f_4 \leq 9/4f_1$. For the case $t = 3$, we name the nodes allocated as v_1, v_2, v_3. We have $\sigma(v_1) + \sigma(v_2) + \sigma(v_3) \leq 3/2f_1$. In addition, $\sum_{l=t+1}^{N} \sigma(v_l) < f_4$. Assume $f_2 + f_3 + f_4 > 9/4f_1$. Then it must hold that $f_2 + f_3 > 3/2f_1$, which can not be satisfied by seeds v_1, v_2, v_3. Allocating new seeds to d_2 or d_3 will lead f_4 can not be satisfied by the remaining seeds. An analogous computation gives a contradiction for the case $t = 4$, too.

Case 2: $k = 2$ **and** $k = 3$**.** For $k = 2$, suppose there are two seeds allocated to each d. We know that in this case $NFD = f_1 + f_2, \sigma(v_1) + \sigma(v_2) + \sigma(v_3) + \sigma(v_4) \geq f_1 + f_2$ and the remaining seeds can not satisfy any remaining d_i. Thus $OPT \leq 2(f_1 + f_2) = 2NFD$ where the equation holds when $f_1 = f_2 = f_3 = f_4$ and all the seeds can satisfy d_1, d_2, d_3, d_4. If d_i for $i = 1, 2, 3, 4$ are satisfied, the remaining seeds, however, cannot satisfy any other d_i, and this is the maximum possible OPT can be obtained. The same result holds when $k = 3$. If at least one d_i is allocated with at least three seeds. Firstly consider the case $k = 3$. Let w.l.o.g. d_1, d_2 and d_3 are satisfied by NFD. Suppose there are t seeds allocated to d_1, d_2, d_3. We have $\sigma(S_i|\mathbf{S}) \leq 3/2f_i$ and it is clear this term is smaller, if there are more than three seeds. Then we can bound $\sum_{l=1}^{t} \sigma(v_l) \leq 2f_1 + 2f_2 + 3/2f_3$ (Assume d_3 is allocated 3 seeds.). Further, $\sum_{l=t+1}^{N} \sigma(v_l) < f_4 \leq f_3$. Thus $OPT < 2f_1 + 2f_2 + 5/2f_3$. Then $\frac{NFD}{OPT} > \frac{f_1+f_2+f_3}{2f_1+2f_2+5/2f_3} > 4/9$. For $k = 2$ we deduce analogously $OPT \leq 2f_1 + 3/2f_2 + f_2 = 2f_1 + 5/2f_2$ and also yield $NFD/OPT \leq 9/4$. It is easy to see that if there is another d_i with 3 seeds or on one d_i there are more than 3 seeds, the bound is even better. Hence the claim follows.

2.3 Proof of Theorem 3

We use reduction from the PARTITION problem. Recall that for this we are given a set of items $P = \{1, ..., m\}$, where item j has integral size s_j. Our goal is to find an index set $L \in P$, such that $s(L) = s(P \backslash L)$, i.e., the items from P are partitioned in two sets of equal size.

Let P be a PARTITION instance and we refer to the sizes of the items as s_j. We now set an instance of Allocation Problem. Let candidate seed set $S_{cand} = \{v_1, v_2, ..., v_{m+K-2}\}$ with $\sigma(v_j) = 2Ks_j$ for $j = 1, ..., m$ and $\sigma(v_j) = 1$ for $j = m+1, ..., m+K-2$, i.e., the items of the Partition instance are scaled by a factor of $2K$. As $K \leq m$ this is done in polynomial time.

Let $s := \sum_{j=1}^{m} \sigma(v_j)$ and set requirements $f_1 = s/2N$, $f_2 = s/2N$, where we assume $s/2$ is integral, which is due to the integral s_j (and thus integral $\sigma(v_j)$). Further we set $f_3 = ... = f_K = 1/N$. Now we see that the solution of Allocation Problem has a value of $(s + K - 2)\gamma/N$, if the Partition problem has a solution. Meanwhile, if the Partition problem has no solution, the value of the solution to Allocation problem is at most $(s/2 + K - 2)\gamma/N$.

Consider the case that the candidate seeds $v_1, ..., v_m$ are assigned to satisfy f_1 and f_2 and the candidate seeds $v_{m+1}, ..., v_{m+K-2}$ are assigned to satisfy $f_3, ..., f_K$. In the

PARTITION instance P, for every index set L we have the property that $s(L) \neq s(P\backslash L)$, i.e., the two parts differ by at least 1. Hence in the instance for Allocation problem, which uses the sizes from the scaled instance P we have for every allocation of candidates $v_1, v_2, ..., v_m$ to satisfy f_1 and f_2, $\sigma(S_1)$ and $\sigma(S_2)$ must differ by at least $2K$. Let w.l.o.g. be $\sigma(S_1) > \sigma(S_2)$, then we have $\sigma(S_1) - \sigma(S_2) \geq 2K$, and thus $\sigma(S_2) \leq s/2 - K$. Consequently, even, if all candidate seeds $v_{m+1}, ..., v_{m+K-2}$ are allocated to satisfy f_2, we have $\sigma(S_2) \leq s/2 - K + (K-2) < s/2 < Nf_2$. We can see f_2 is not satisfied. And if the candidate seeds $v_1, ..., v_m$ are allocated arbitrarily, the value of the solution may only decrease and we have shown that the value of a solution on the given instance is at most $(s/2 + K - 2)\gamma/N$, if the Partition problem has no solution.

Hence in case an algorithm has approximation ratio larger than $\rho = \frac{1}{2} + \frac{K-2}{2s+2K-4}$, it can distinguish the cases and solve the Partition problem.

2.4 Proof of Theorem 4

First, Algorithm 3 can ensure a $1/4$-approximation for the CAMB instance (D_{NS}, S_{cand}^{NS}), where D_{NS}, S_{cand}^{NS} are the set of d_i satisfied non-singularly and the corresponding allocated seed set in the return of $OPT(D, S_{cand})$. Here we first show that all seeds are used after running NFD-AE.

Lemma 4. *For any instance (D', S_{cand}^{NS}), where $D_{NS} \subseteq D'$ the algorithm can allocate all the candidate seeds.*

Proof. Suppose there is a seed v_j unallocated after running NFD-AE. Then all d_i with $f_i \geq \sigma(v_j)$ must be satisfied and all the seeds v allocated to such f_i must have $\sigma(v) \geq \sigma(v_j)$. The remaining unsatisfied f_i all have $f_i < \sigma(v_j)$. We see there exists an allocation that uses up all the seeds and each f_i is satisfied non-singularly. In this allocation, v_j must be allocated to a d_i where $f_i > \sigma(v_j)$. But now v_j is not used while d_i is also satisfied. Our algorithm allocates nodes with decreasing σ thus v_j cannot be replaced by multiple nodes with smaller σ. This means a node v' with $\sigma(v') \geq \sigma(v_j)$ is allocated to d_i instead of v_j. In turn, the place of v' is also replaced by a node with larger σ, otherwise we can allocate v_j. Iterate the above process until the node with largest σ is used. Thus there must exist an unsatisfied d_j with $f_j \geq \sigma(v_j)$, which v_j can be allocated and hence the contradictory follows. □

Basing on Lemma 4, we see NFD-AE algorithm is guaranteed to use up all candidate seeds for instance (D', S_{cand}^{NS}) with $D_{NS} \subseteq D'$, which brings conveniens to analyze the approximation ratio.

Lemma 5. *For instance (D', S_{cand}^{NS}) where $D_{NS} \subseteq D'$, NFD-AE ensures a 1/4-approximation.*

Proof. First, after the for loop in Algorithm 3, we have already satisfy some d_i, denote as set D_{sa}. Meanwhile, we denote D_{usa} as the set of d_i that allocated some seeds but not satisfied. We have that the additional size $f_c^+ = \sum_{i \in D_{sa}} (\sum_{v \in S_i} \sigma(v) - f_i)$, which is smaller than $f_c = \sum_{i \in D_{sa}} f_i$. Actually, if D_{usa} is empty, we already ensure a 1/2-approximation. This is because we allocate by decreasing of e and the possible budget can be earned is at most twice of current result since $\sigma(S_i) < 2f_i$.

If D_{usa} is not empty. We can see the final solution is the maximum one of S' and S'' in the algorithm. In S', only d_l is satisfied since we assume allocating all seeds can satisfy any one of d_i. In S'', we can see all $d_i \in R$ with $i = 1, 2, 3, ..., l-1$ are satisfied and d_l is unsatisfied. This is because in Algorithm, d_i is unsatisfied if no seeds v with $\sigma(v) \leq f_i$ is unallocated and the subsequent allocated seeds all can satisfy d_i singularly. This also mean $d_i \leq d_j$ with $i < j$. So d_l is unsatisfied.

Let OPT be the optimal earned budget and $NFD\text{-}AE$ be the budget earned by the return of the algorithm. Then if $B_l \geq 1/2OPT$, $NFD\text{-}AE \geq S' \geq 1/2OPT$. If $B_l \leq 1/2OPT$. We have $NFD\text{-}AE \geq S'' \geq 1/2(OPT - B_l) \geq 1/4OPT$. □

To deduce the approximation on the instance (D, S_{cand}), we further show the monotonicity of $NFD\text{-}AE$ algorithm.

Lemma 6. *For the instances (D, S_{cand}) and (D, S_{cand}^{NS}) where $S_{cand}^{NS} \subseteq S_{cand}$, $NFD\text{-}AE(D, S_{cand}^{NS}) \leq NFD\text{-}AE(D, S_{cand})$.*

Proof. To prove the lemma, it is sufficient to show for an instance D, S, $NFD\text{-}AE(D, S) \leq NFD\text{-}AE(D, S \cup \{v\})$. Suppose v is allocated to d_i. There are three cases.

Case 1: d_i is not satisfied in (D, S) but satisfied by allocating v without removing any already allocated seeds. In this case, the earned profit is increased.

Case 2: d_i is not satisfied by (D, S) but satisfied by allocating v with removing some already allocated seeds. In this case, first, the earned profit is increased. Second, suppose we remove one seed v', it is actually still the case for adding seeds to instance $(D, S\backslash v' \cup \{v\})$.

Case 3: d_i is satisfied by (D, S) and also satisfied by allocating v with removing some already allocated seeds. In this case, first, the earned profit is unchanged. Second, suppose we remove one seed v', it is actually still the case for adding seeds to instance $(D, S\backslash v' \cup \{v\})$. □

Now we can show the approximation of FA algorithm on instance (D, S_{cand}). Let $D_S, D_{NS}, S_{cand}^S, S_{cand}^{NS}$ be the singularly (non-singularly) satisfied set and corresponding allocated seed set of the optimal solution. Then we have $OPT(D, S_{cand}) = OPT(D_S, S_{cand}^S) + OPT(D_{NS}, S_{cand}^{NS})$.

If $OPT(D_{NS}, S_{cand}^{NS}) < 4/5OPT(D, S_{cand})$, then $KM(D, S_{cand}) \geq KM(D_S, S_{cand}^S) = OPT(D_S, S_{cand}^S) > 1/5OPT(D, S_{cand})$, which is achieved by solution $\mathbf{S}_1$.

Else if $OPT(D_{NS}, S_{cand}^{NS}) \geq 4/5OPT(D, S_{cand})$, then by Lemma 5 and 6, we have $NFD\text{-}AE(D, S_{cand}) \geq NFD\text{-}AE(D_{NS}, S_{cand}^{NS}) \geq 1/4OPT(D_{NS}, S_{cand}^{NS}) \geq 1/5OPT(D, S_{cand})$, which is achieved by solution $\mathbf{S}_2$.

Therefore the theorem is proved.

2.5 Proof of Theorem 5

Lemma 7. *Given $\theta \geq \frac{(2+\lambda)\cdot\log(2N)}{f_m\lambda^2}$, then for any set S with $\sigma(S) \geq N \cdot f_m$, with at least $1 - 1/n$ probability, we have $\big|F_R(S) - \sigma(S)\big| < \lambda\sigma(S)$.*

Proof. Let p be the probability that $S \cap R \neq \emptyset$ for a random RR set R. Then we have $p = \mathbb{E}[F_R(S)/N] = \sigma(S)/N$. Let $\delta = \lambda\sigma(S)/Np$. Then by Chernoff bounds, we have

$$\begin{aligned}
&\Pr\left[\big|F_R(S) - \sigma(S)\big| \geq \lambda\sigma(S)\right] \\
&= \Pr\left[\big|\theta \cdot F_R(S)/N - p\theta\big| \geq \frac{\lambda\sigma(S)}{Np}\theta p\right] \\
&< 2\exp\left(-\frac{\delta^2}{2+\delta}\cdot p\theta\right) \leq 2\exp\left(-\frac{\lambda^2 f_m}{2+\lambda}\theta\right) = \frac{1}{N}.
\end{aligned}$$

Therefore, the lemma is proved. □

Now we have $(1-\lambda)\sigma(S) \leq F_R(S) \leq (1+\lambda)\sigma(S)$. It is easy to see that $\sigma(S) \geq F_R(S)(1+\lambda) \geq N \cdot f_m$. Thus we mainly focus on proving the bound of $C(S)$.

Let S_i be the set selected after i^{th} iteration, $\eta_i = Nf_m - \sigma(S_i)$ and $\hat{\eta_i} = (1+\lambda)Nf_m - F_R(S_i)$. Denote S° as the optimal solution and c° as the cost of S^*.

First, if $\forall x \in V\backslash S_i$, $\frac{\sigma(S_i \cup \{x\}) - \sigma(S_i)}{c(x)} < \frac{\eta_i}{c^*}$, we have

$$\begin{aligned}
\sigma(S^* \cup S_i) &\leq \sigma(S_i) + \sum_{x \in S^*\backslash S_i} (\sigma(S_i \cup \{x\}) - \sigma(S_i)) \\
&\qquad\qquad \text{(by submodularity of } f\text{)} \\
&< \sigma(S_i) + \frac{\eta^i}{c^*} \cdot \sum_{x \in S^*\backslash S_i} c(x) < Nf_m
\end{aligned}$$

which is a contradiction since the left hand side is no less than the optimal influence spread. Thus there must exist a node $x \in V\backslash S_i^j$ satisfying $\frac{\sigma(S_i \cup \{x\}) - \sigma(S_i)}{c(x)} \geq \frac{\eta_i}{c^*}$.

Let v_i, c_i be the node added in iteration i and its cost. Let the total number of iterations be ℓ. Consider S_ℓ with $F_R(S_\ell) \geq (1+\lambda)(N \cdot f_m - \beta)$ and $F_R(S_{\ell-1}) \leq (1+\lambda)(N \cdot f_m - \beta)$. For $1 \leq i \leq \ell - 1$, we have

$$\begin{aligned}
\hat{\eta_{i-1}} - \hat{\eta_i} &= F_R(S_i) - F_R(S_{i-1}) \\
&\geq (1-\lambda)\sigma(S_i) - (1+\lambda)\sigma(S_{i-1}) \\
&= \sigma(S_i) - \sigma(S_{i-1}) - \lambda(\sigma(S_i) + \sigma(S_{i-1})) \\
&\geq \frac{c_i\eta_i}{c^*} - \frac{2\lambda(1+\lambda)}{1-\lambda}Nf_m \\
&\geq \frac{c_i\hat{\eta_{i-1}}}{(1-\lambda)c^*} - \frac{2\lambda Nf_m}{1-\lambda}\left(\frac{c_i}{2c^*} + \lambda + 1\right),
\end{aligned}$$

that is

$$\hat{\eta}_i \leq \left(1 - \frac{c_i}{(1-\lambda)c^*}\right)\hat{\eta}_{i-1} + \frac{2\lambda Nf_m}{1-\lambda}\left(\frac{c^+}{2c^*} + \lambda + 1\right).$$

Since $F_R(S_{\ell-1}) \leq (1+\lambda)(Nf_m - \beta)$. We have $\hat{\eta}_{\ell-1} > (1+\lambda)\beta$. Since $\forall z, 1 + z \leq e^z$, we have

$$\begin{aligned}
&(1+\lambda)\beta < \hat{\eta}_{\ell-1} \\
&\leq \exp\left(-\frac{c_{\ell-1}}{(1-\lambda)c^*}\right)\hat{\eta}_{\ell-2} + \frac{2\lambda f_m}{1-\lambda}\left(\frac{c^+}{2c^*} + \lambda + 1\right) \\
&\leq \exp\left(-\frac{1}{(1-\lambda)c^*}\sum\nolimits_{j=1}^{\ell-1} c_j\right)(1+\lambda)Nf_m \\
&\quad + \frac{1 - (1 - \frac{c^-}{(1-\lambda)c^*})^{\ell-1}}{1 - (1 - \frac{c^-}{(1-\lambda)c^*})}\frac{2\lambda Nf_m}{1-\lambda}\left(\frac{c^+}{2c^*} + \lambda + 1\right),
\end{aligned}$$

that is

$$\begin{aligned}
&\exp\left(-\frac{1}{(1-\lambda)c^*}\sum\nolimits_{j=1}^{\ell-1} c_i\right) \\
&\geq \frac{(1+\lambda)\beta - \frac{2c^*\lambda Nf_m}{c^-}(\frac{c^+}{2c^*} + \lambda + 1)}{(1+\lambda)Nf_m} \\
&\geq \frac{\beta}{Nf_m} - \frac{\lambda}{c^-}\left(2c^* + \frac{c^+}{(1+\lambda)}\right)
\end{aligned}$$

Thus, let $\beta = Nf_m(\frac{\lambda}{c^-}(2c^* + \frac{c^+}{(1+\lambda)}) + \alpha')$, we have

$$c(S_\ell) \leq c(S_{\ell-1}) + c^+ \leq c^*(1-\lambda)\ln\alpha'.$$

If $F_R(S_\ell) \geq (1+\lambda)Nf_m$ and letting $\alpha' = \alpha$, we have done. Otherwise, $F_R(S_\ell) < (1+\lambda)Nf_m$ and we have

$$\begin{aligned}
F_R(S_{\ell+1}) - F_R(S_\ell) &\geq (1-\lambda)\sigma(S_{\ell+1}) - (1+\lambda)\sigma(S_\ell) \\
&\geq \sigma(S_{\ell+1}) - \sigma(S_\ell) - \lambda(\sigma(S_{\ell+1}) + \sigma(S_\ell)) \\
&\geq \frac{N}{c_m}\left(1 - \frac{1+\lambda}{1-\lambda}f_m\right) - \lambda N\left(1 + \frac{1+\lambda}{1-\lambda}f_m\right) \\
&\geq (1+\lambda)\left(\frac{N}{1+\lambda}\left(\frac{1}{c_m} - \lambda\right) - \frac{f_m}{1-\lambda}\left(\frac{N}{c_m} + \lambda N\right)\right)
\end{aligned}$$

We should ensure $\frac{N}{1+\lambda}(\frac{1}{c_m} - \lambda) - \frac{f_m}{1-\lambda}(\frac{N}{c_m} + \lambda N) \geq \beta$ to make $F_R(S_{\ell+1}) \geq (1+\lambda)Nf_m$, which results from

$$\alpha' \leq \frac{1/c_m - \lambda}{f_m(1+\lambda)} - \frac{1/c_m + \lambda}{1-\lambda} - \frac{2c_m\lambda}{c^-} - \frac{c^+\lambda}{c^-(1+\lambda)}$$

We set $y = g(\lambda)$ where $g(\lambda)$ is the right side of the above inequality. When $\lambda = 0$, we have $g(0) = (\frac{1}{f_m} - 1)\frac{1}{c_m} > 0$ and

$$g'(0) = -\left(\frac{1}{f_m} + 1\right)\left(\frac{1}{c_m} + 1\right) - \frac{1}{c^-}(2c_m + c^+) < 0.$$

$$g''(0) = \left(\frac{2}{f_m} - 1\right)\left(\frac{2}{c_m} - 1\right) + \frac{2c^+}{c^-} - 1 > 0.$$

Let $\lambda^* \leq \frac{1-f_m}{(1+c_m)(1+f_m) + \frac{2c_m + c^+}{c^-}c_m f_m}$. We can see $(\lambda^*, 0)$ is the intersection point between tangent line of $g(\lambda)$ at $\lambda = 0$ and $y = 0$. Thus, $g(\lambda) > 0$ with the given λ and by setting $\alpha' = \alpha$, we have done.

References

[Garey and Johnson, 1986] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1986.