Biased Respondent Group Selection under Limited Budget for Minority Opinion Survey *

Donghyun Kim[1], Wei Wang[2], Matthew Tetteh[1],
Jun Liang[4], Soyoon Park[3], Wonjun Lee[3]**

[1] Dept. of Math. and Physics, North Carolina Central Univ., Durham, NC, USA.
donghyun.kim@nccu.edu, mtetteh@eagles.nccu.edu

[2] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China.
wang_weiw@163.com

[3] Dept. of Computer Science, University of Texas at Dallas, Richardson, TX, USA.
jxl130131@utdallas.edu

[4] Dept. of Computer Science and Engineering, Korea University, Seoul, South Korea.
{sogehao, wlee}@korea.ac.kr

Abstract. This paper discusses a new approach to use the information from a special social network with high homophily to select a survey respondent group under a limited budget such that the result of the survey is biased to the minority opinions. This approach has a wide range of potential applications, e.g. collecting complaints from the customers of a new product while most of them are satisfied. We formally define the problem of computing such group with better utilization as the p-biased-representative selection problem (p-BRSP). This problem has two separate objectives and is difficult to deal with. Thus, we also propose a new unified-objective which is a function of the two optimization objectives. Most importantly, we introduce two polynomial time heuristic algorithms for the problem, where each of which has an approximation ratio with respect to each of the objectives.

1 Introduction

Recently, the value of the information from online resources such as online social networks are getting more recognized and thus lots of research efforts are made

* This work was supported in part by US National Science Foundation (NSF) CREST No. HRD-1345219. This research was jointly supported by National Natural Science Foundation of China under grants 11471005.

** Corresponding Author

to maximize its utilization [1–3]. Following the trend, online survey is also being recognized as a critical tool to make a wide range of significant marketing and political decisions. Due to the reason, a huge amount of investment is being made for various researches on online survey [4]. There are several motives that promote online survey [5]. Most of all, it costs much less and produces results much faster than its counterpart. How to select a meaningful survey respondent group has been a tough but critical question to deal with to make a traditional off-line survey method more effective and reliable, and this is still true for online survey. Here, the definition of "meaningful" can differ based on the purpose of the survey. In most cases, a survey aims to learn the general opinions from the public of interest by sampling, and thus it is significant to elect a group of unbiased respondents among the public using a proper method, e.g. random sampling.

Previously, Kim et. al. [6] introduced a new strategy to elect a survey respondent group to perform efficient biased survey and collect more minority opinions with the assistance from an artificial social network graph constructed by them. They argued that in opposition to the widely accepted belief, biased survey could be useful, and discussed an example to support their claim. In the example, they pointed out that when the majority of the people, who purchased a new smartphone, are very satisfied, the opinions of unsatisfied users of the new smartphone, who can be classified as minority opinion holders, could provide useful information to the new smartphone's product quality manager, who is more interested in complaints. Based on this observation, Kim et. al. introduced a new strategy to select a respondent group more suitable for such survey in the sense that from which the diversified voices from unsatisfied users (minority opinions) can be heard more loudly. Most importantly, their strategy only requires the expected similarity of the opinions between each pair of users on the issue to construct the biased respondent group. Kim et. al.'s approach is rather localized and consumes less resources, and thus is more practical in big data environment compared to its alternative straightforward approach which analyzes the sentiment of each individual on the issue first, and then identifies the minority opinions as this certainly requires global analysis.

To achieve the goal, they first compute a new social network graph G in which each node represents a member of the society, and there is an edge between two nodes only if the opinions of the two people, who are represented by the nodes, are similar enough on the subject of interest. In the literature, such social network graph, in which two nodes are neighboring only if they are sharing close opinion, is told to have high level of homophily [7]. Once such a graph is constructed, the algorithm attempts to compute a smaller size inverse k-core dominating set D of G, which is a subset of the nodes in $G = (V, E)$ such that all nodes in V is either in D or neighboring to a node in D (domination property), and the degree of each node in D in the induced graph by D in G is at most k (inverse k-core property). Note that the dominating property on D is necessary to ensure that D has a representation over all of the members in the society, and the inverse k-core property is enforced to make sure to obtain more minority opinions with greater diversity. Most of all, in the simulation, the authors have shown that their approach is in fact effective.

Meanwhile, online survey may not be completely free-of-cost, even though it is usually much cheaper than the traditional off-line survey approaches. For instance, a recent study conducted by Singer and Ye shows that a reasonable compensation can certainly improve the response rate of the survey [8]. However, we notice the approach proposed by Kim et. al. does not provide any explicit way to control the cost of their approach (the size of the group returned by their algorithm), and this can be a very critical issue in order to make the approach more practical. Motivated by this observation, in this paper, we introduce two new approaches to perform effective biased survey using homophily rich graph under limited budget. The main contributions of this paper can be summarized as follows.

(a) For the first time in the literature, we discuss the motivation for selecting a survey respondent group to better capture more diversified minority opinions under limited budget. We formulate the problem of our interest as a new optimization problem two independent objectives. We also discuss how the two objectives can be combined into one objective function.

(b) We introduce two polynomial time algorithms to solve the proposed problem. The first algorithm is based on Kim et. al. [6]'s approximation algorithm for the minimum inverse k-core dominating set problem, and can be considered as its generalization. This algorithm has a proven approximation ratio with respect to Objective 1. The second one is a simpler greedy algorithm and has the best possible approximation ratio with respect to Objective 2.

The rest of this paper is organized as follows. Section 2 discusses some preliminaries. We introduce the two new approaches for the problem of our interested in Section 3. Section 4 concludes this paper and presents future works.

2 Preliminaries

2.1 Notations and Definitions

In this paper, $G = (V, E)$ represents a social network graph with a node set $V = V(G)$ and an edge set $E = E(G)$. We assume the relationship between each pair of members is symmetric, which is true in homophily high social network graph, and thus the edges in E are bidirectional. Also, we use n to denote the number of nodes in V, i.e. $n = |V|$. For any subset $D \subseteq V$, $G[D]$ is a subgraph of G induced by D. For a pair of nodes $u, v \in V(G)$, $Hopdist(u, v)$ is the hop distance between u and v over the shortest path between them in G. Given a node v in G, $deg(v, G)$ is the degree of v in G. For any $V' \subseteq V$ in $G = (V, E)$, $deg(V', G)$ is $\max_{v \in V'} deg(v, G)$. Also, $deg(G)$ is $\max_{v \in V} deg(v, G)$. $dia(G)$ is the diameter of G, which is the length of the longest shortest path between any pair of nodes in the graph G. For each node $v \in V$, $N_{v,V}(G)$ is the set of nodes in V neighboring to v in G. In other words, the nodes in $N_{v,V}(G)$ are the 1-hop neighbors of v in G. Similarly, $N^d_{v,V}(G)$ is the set of nodes in V, which are at most d-hops far from v in G. Note that we will use $N_{v,V}(G)$ and $N^1_{v,V}(G)$ interchangeably. Given a graph G, a subset $D \subseteq V$ is a *dominating set (DS)* of G if for each node $u \in V \setminus D$, $\exists v \in D$ such that $(v, u) \in E$. In general, a subset $D \subseteq V$ is a *d-hop dominating set (d-DS)* of G if for each node $u \in V \setminus D$, $\exists v \in D$ such that $Hopdist(v, u) \leq d$. In graph theory, the *minimum dominating set problem (MDSP)* is to find a minimum size DS in a given G. Also, the goal of the *minimum d-hop dominating set problem (MdDSP)* is to find a minimum size d-DS in G. Given a graph G, a subset $D \subseteq V$, and a positive integer k such that $0 \leq k \leq \Delta$, where Δ is the degree of G, D is an *inverse k-core* in G if for each $v \in D$, $|N_{v,D}(G)| \leq k$. Generally speaking, D is an *inverse* (k, d)*-core* in G if for each $v \in D$, $|N^d_{v,D}(G)| \leq k$. Given $\langle G, k \rangle$, the minimum inverse k-core dominating set problem is to find a minimum size inverse k-core dominating set of G. Similarly, given $\langle G, k, d \rangle$, the minimum inverse (k, d)-core dominating set problem is to find a minimum size inverse k-core d-hop dominating set of G.

2.2 Formal Definition of Problem

In this paper, we are interested in selecting a survey respondent group whose size is p, which is a positive constant determined by the available budget, such that (a) more members with minority opinions are selected (biased to minority opinions), and (b) the group can well-represent the overall minority opinions (well-representation of diversified minority opinion). In the following, we explain the desirable properties of the group to be elected for our purpose and their implications in terms of graph theory.

Property 1: higher bias to minority opinion holders. Previously, Kim et. al. [6] introduced a way to construct a homophily high social network graph, in which there exists an edge between two nodes only if the opinions of the members represented by the two nodes are similar enough. They also found that in a homophily high social network graph, a node with lower node degree tends to be a minority opinion holder. This implies that a group with size p possibly includes more minority opinion holders (and thus the group is more biased) when the average degree of the selected nodes (or their total node degree) in the given social network graph is lower. In this paper, we will assume a homophily high social network graph G as an input of our algorithms and thus, prefer to have a node subset V' with size p such that $\sum_{v \in V'} deg(v, G)$ becomes as small as possible as an output of our algorithm.

Property 2: better representation of minority opinion holders. In a homophily high social network graph G, a pair of nodes are connected in G only if their expected opinions on the subject of interest are similar enough. In the literature, the minimum size dominating set problem is widely used to select an efficient representative group. For instance, in [6], Kim et. al. were looking for a minimum size dominating set of G with certain properties to elect a group of survey respondents which can represent the rest.

Unfortunately, there are two issues to extend this approach to the problem of our interest directly. First, depending on the input graph G, a dominating set (or 1-hop dominating set) with the enforced size constraint p may not exist. Second, we may ignore the majority opinion holders in the process of selecting the representatives for minority opinion holders in contrast to the fact that a dominating set implicitly does not ignore them.

One way to address the first concern is to relax the 1-hop domination constraint and allow a representative of a node to be multiple hops far from the node. In this way, the size of the dominating set can be reduced. However, as the hop distance between two nodes in G generally implies the degree of difference on the opinions between them, and thus as the hop distance grows, the effect of the representation becomes smaller. Consequently, it would be more desirable to find a subset of nodes, V' from V with size p such that the maximum hop distance from a minority opinion holder to its nearest node in V' becomes minimized. To address the second concern, the concept of nodes with "minority opinion holders" should be more clearly defined. Based on [6], a node with lower degree has a better chance to be a minority opinion holder. Therefore, we may attempt to identify those minority opinion holders by computing the degree of each node in G (by following Property 1) and consider those nodes with smaller node degree as minority opinion holders, where the concept of "smaller" is dependent on the context and can be specified by the survey organizer.

Property 3: greater diversification of minority opinion holders. In practice, there can be a number of different minority opinions, and thus minority opinions are quite diversified. Therefore, it is important to construct a size p representative group in a way that more diversified minority opinions can be collected. Now, we have the following remark.

Remark 1. We argue that Property 3 is already included in Property 2. For instance, given a connected graph G with one huge complete subgraph (majority opinion holders) and two non-adjacent smaller size complete subgraphs (minorities), if we select two representatives from the same smaller size complete subgraph, the total hop distance discussed in Property 2 will be greater compared to the case in which one representative is selected from each smaller size complete subgraph.

Based on Properties 1, 2, 3, and Remark 1, we formally define our problem of interest.

Definition 1 (p-BRSP). *Given a homophily high social network graph $G = (V, E)$, a subset $S \subset V$, which is the group of nodes in V whose node degree is no greater than a threshold level (and therefore are suspected as nodes representing minority opinion holders), a positive integer $p \leq |V| = n$, the p-biased-*

representative selection problem (p-BRSP) is to find a subset $V' \subseteq V$ whose size is p from G such that

(a) ***Objective 1****: the total node degree of V' is minimum, or equivalently $\sum_{v \in V'} deg(v, G)$ is minimum, and*

(b) ***Objective 2****: the maximum hop distance between a node in V to its nearest node in V', or equivalently $\max_{u \in S \setminus V'} \arg\min_{v \in V'} Hopdist(u, v)$ is minimum.*

Meanwhile, it is uncertain that which of the requirements is more significant. As a result, we redefine the problem such that its objective is to minimize

$$\alpha \times \frac{\sum_{v \in V'} deg(v, G)}{\sum_{v \in W} deg(v, G)} + (1 - \alpha) \times \frac{\max_{S \in V \setminus V'} \arg\min_{v \in V'} Hopdist(u, v)}{dia(G)}, \quad (1)$$

from some $0 \leq \alpha \leq 1$, which is determined by the operator of the survey, where W is the set of the first p nodes in G with largest node degree. During the rest of paper, we discuss how to quality solutions of p-BRSP with the objective function in Eq. (1).

3 Two Polynomial Time Algorithms for p-BRSP

In this paper, we introduce two new heuristic algorithms for p-BRSP along with some interesting theoretical analysis.

3.1 First Approach: Greedy-MI(k, d)CDSA

Previously, Kim et. al. introduced Greedy-MIkCDSA, a simple greedy strategy for the minimum inverse k core dominating set problem (MIkCDSP). Given an MIkCDSP instance $\langle G, k \rangle$, Greedy-MIkCDSA first prepares an empty set D, which will eventually include the output, an inverse k-core dominating set (IkCDS) of G. For each node $v_i \in V$, the algorithm creates a counter n_i which is initialized to 0. The counter will be used to track the number of neighbors of v_i in D. Depending on the counter, the algorithm creates a partition of the nodes in V, $X_0, X_1, \cdots$, where X_j is the subset of nodes in V whose counter is j. This means that initially X_0 is equal to V and each of the rest is empty. Clearly, the number of the subsets is bounded by n. Then, the algorithm iteratively picks a node v_i from $V \setminus \left((\bigcup_{j \geq k} X_j) \bigcup D \bigcup Q \right)$, i.e. v_i is a node which is

Algorithm 1 Greedy-MI(k,d)CDSA (G,S,p)

for $d = 1$ to $dia(G)$ **do**
  for $k = 0$ to $deg(G)$ **do**
    Prepare an empty set $D_{(d,k)}$, i.e. $D_{(d,k)} \leftarrow \emptyset$.
    For each $v_i \in V$, prepare a counter n_i which is initialized to 0, i.e. $n_i \leftarrow 0$.
    Suppose $X_j = \{v_i | v_i \in V \text{ and } n_i = j\}$.
    while $X_0 \bigcap S \neq \emptyset$ **do**
      Find $v_i \in V \setminus \left((\bigcup_{j \geq k} X_j) \bigcup D_{(d,k)} \right)$ so that $|N^w_{v_i, X_0 \bigcap S}(G)|$ is maximized, where $1 \leq w \leq d$. A tie can be broken by selecting a node with smaller node degree.
      Set $D_{(d,k)} \leftarrow D_{(d,k)} \cup \{v_i\}$.
      for each node $v_j \in N^w_{v_i,V}(G)$ with any $1 \leq w \leq d$ **do**
        $n_j \leftarrow n_j + 1$.
      end for
    end while
  end for
end for
Output the $D_{(d,k)}$ whose size is p and which minimizes the objective function in Eq. (1).

(a) with a counter n_i whose value is less than k (i.e. has less than k neighbors in DS),
(b) not selected as a DS node yet, and
(c) without any neighboring node w_l which is in D (otherwise v_i is already dominated) and, at the same time, in X_j for some $j \geq k$ (otherwise adding v_i to D will violate the k-inverse-core property),

such that the number of neighbors of v_i in X_0 is the maximum, where $Q = \{w_1, \cdots, w_q\}$ such that $w_l \in Q$ has at least one neighbor in $(\bigcup_{j \geq k} X_j)$. Any tie can be broken arbitrarily. The algorithm eventually terminates when all nodes in V is either in D or dominated by some node in D while maintaining $G[D]$ as an inverse k-core.

The main idea of Greedy-MIkCDSA is still applicable to p-BRSP. However, due to the size constraint p, we need to employ d-hop dominating set instead of 1-hop dominating set. To extend the main idea of Greedy-MIkCDSA which utilizes 1-hop dominating set to d-hop dominating set, there are several challenges to deal with at the same time. That is, we must find a subset V' of nodes with size p such that V' can d-hop dominates all nodes in S. At the same time, we need to adjust k properly, otherwise, there might be no feasible solution. Note that there can be more than one $\langle V', d, k\rangle$ computed in this way, and we need to find the one which can minimize the objective of p-BRSP in Eq. (1).

Algorithm 1 is the formal definition of this modified algorithm, namely Greedy-MI(k, d)CDSA. The core idea of our modification is that we vary d and k (Line 1 and Line 2 of Algorithm 1) and compute an inverse k-core d dominating set whose size is p. From Line 3 to Line 12, the strategy of Kim et. al's algorithm for the minimum inverse k-core dominating set problem is used to compute a smaller size inverse (k, d) core dominating set $D_{(d,k)}$ of G in a greedy manner. Note that $D_{(d,k)}$ intends to d-hop dominate the nodes in S which is a subset of V rather than the whole nodes in V. In Line 15, the algorithm returns the best $D_{(d,k)}$, which minimizes the objective function in Eq. (1) among all feasible ones.

Theorem 1. *Algorithm 1 is correct.*

Proof. This algorithm always returns a feasible solution and therefore correct as with $d = dia(G)$ and $k = deg(G)$, we can always find a single node in G which is an inverse k-core d dominating set of G.

Note that the potential function of Algorithm 1 is submodular and thus the algorithm has an approximation factor of $O(\log \delta)$ with respect to Objective 1, where *delta* is the maximum degree of the input graph. The proof of this claim is very similar to the proof of Theorem 2 in [10]. The only difference is that now we are considering d-hop domination instead of 1-hop the ratio becomes $O(\log \delta^d) = O(d \log \delta) = O(\log \delta)$.

3.2 Second Approach: Simple-p-RSPA

Now, we propose a simpler greedy algorithm for p-BRSP. Before discussing our new strategy, we first introduce a related problem, namely the p-center problem with degree constraint (p-CDC), and propose a 2-approximation algorithm for it, where 2 is the best possible. Then, this algorithm is used to design our second strategy, the simple-p-RSP algorithm (Simple-p-RSPA).

Definition 2 (p-CDC). *Given a graph $G = (V, E)$, a positive integer p, a subset $S \subset V$ representing the group of people with minority opinion, and a degree constraint W, the p-center problem with degree constraint (p-CDC) is to find a subset of nodes D satisfying (a) $|D| \leq p$, and (b) $\sum_{v \in D} deg(v, G) \leq W$*

such that the furthest distance from a node in S to its nearest node in D becomes minimum.

Now, we present a 2-approximation algorithm for the p-CDC problem, which is best possible unless $P = NP$. The main idea of the algorithm and corresponding analysis are motivated by Hochbaum's algorithm for the p-center problem [9]. However, p-CDC has a additional degree-sum constraints, and the objective function is also different from that of p-center (note we try to minimize the maximum hop-distance of a node in S (instead of V) to its nearest center in D). So the method in [9] cannot be applied directly here.

Let $\Gamma = (V, E')$ be a complete weighted graph constructed from G, in which the edge weight, $cost(e)$ of $e = (u, v)$, is the length of shortest-path between nodes u and v, and the node weight of v is the degree of v in G for every $v \in V$. Now, we order the edges of Γ in the following way: $cost(e_1) \leq cost(e_2) \leq \cdots \leq cost(e_m)$; where $m = \binom{n}{2}$ is the number of edges in the complete graph Γ.

Let $G_i = (V, E_i)$ with $E_i = \{e_1, e_2, \cdots e_i\} = \{e \mid cost(e) \leq cost(e_i)\}$. Let H_i be the subgraph of G_i induced by the 1-hop neighbors of S, together with S, i.e., $H_i = G_i[\cup_{v \in S} N_{v,V}(G_i) \cup S]$. Let H_i^2 be the square of H_i, i.e., H_i^2 is a graph obtained from H_i such that two nodes are adjacent in H_i^2 if and only if the hop-distance between them is no more than two in H_i. The algorithm is as follows:

(a) Step 1. Compute $H_1^2, H_2^2, \cdots, H_m^2$ and $(G_1[S])^2, (G_2[S])^2, \cdots, (G_m[S])^2$.
(b) Step 2. For each $i = 1, 2, \cdots, m$, compute a Maximal Independent Set (MIS) M_i of $(G_i[S])^2$ (which is also an independent set of H_i^2) as follows: At each time, choose a node $v \in S$ with the lightest node weight, then remove all the neighbors of v together with v in H_i^2, i.e., $N_{v,V}(H_i^2)$, from $(G_i[S])^2$; in the remaining graph, repeat the same process until there is no node left in $G_i^2[S]$.
(c) Step 3. Choose the smallest index i (say j) such that $|M_i| \leq p$ and $w(M_i) = \sum_{v \in M_i} deg(v) \leq W$.
(d) Step 4. Output $S = M_j$ as the centers.

Now, we show the algorithm describe above is a 2-approximation for p-CDC.

Lemma 1. *M_i dominates S in graph $(G_i[S])^2$ for every i.*

Proof. Note M_i is a maximal independent set of $(G_i[S])^2$, it is also a dominating set of $(G_i[S])^2$. Since if there is one node, say v in S, which is not dominated by

M_i, then $M_i \cup \{v\}$ is also an independent set; contradicts the fact that M_i is a maximal indecent set.

Lemma 2. *Let D_i^* be a subset of V with minimum size which dominates S in graph G_i, then $|D_i^*| \geq |M_i|$.*

Proof. Since M_i is an independent set in H_i^2, the hop distance of any two nodes $u, v \in M_i$ is at least three in H_i. Thus all the stars $S(u) = \{v \in V \mid (u, v) \in E(H_i)\}$ centered at $u \in M_i$ are pairwise disjoint each other. For each star $S(u)$, at least one vertex has to be selected into D_i^* in order to dominate S. Therefore, we have $|D_i^*| \geq |M_i|$.

Lemma 3. *Let WD_i^* be a subset of V with minimum total weight which dominates S, then $w(WD_i^*) \geq w(M_i)$.*

Proof. The proof is similar to that of Lemma 3. Now the key point is that by the construction of M_i, for each star $S(u)$ $(u \in M_i)$, we have $w(u) \leq w(v)$ for any $v \in N_{G_i}(v) = \{v|(u, v) \in E(H_i)\}$. Note $S(u) \cap S(v) = \emptyset$ for $u, v \in M_i$ and $u \neq v$. Thus, at least one node in each $S(u)$ $(u \in M_i)$ has to be selected into WD_i^*. Note u is the lightest node in $S(u)$. It follows that $w(WD_i^*) \geq w(M_i)$.

Theorem 2. *Above algorithm is a 2-approximation for p-CDC.*

Proof. Let i^* be the smallest index such that there exists a subset D_{i^*} of G_{i^*} that dominates S such that $|D_{i^*}| \leq p$ and $w(D_{i^*}) \leq W$. Then we have $OPT = cost(e_{i^*})$, where OPT is the optimal value of the p-CDC problem. By our algorithm, for each i $(i = 1, 2, \cdots, j-1)$, we have either $|M_i| > p$ or $w(M_i) > W$. It follows from Lemma 2 and Lemma 3 that either $|D_i^*| \geq |M_i| > p$ or $w(WD_i^*) \geq w(M_i) > W$. Thus we have $i^* > i$ for $i = 1, 2, \cdots, j-1$, i.e., $j \leq i^*$ and $cost(e_j) \leq OPT$. Since M_i is a maximal independent set of $(G_i[S])^2$, it also a dominating set of $(G_i[S])^2$. So in $(G_i[S])^2$, the stars centered at each $u \in M_i$ span all the nodes in $S = V((G_i[S])^2)$. Let v be any node in a star centered at some $u \in M_i$. Then v is at most two hops away from u in $G_i[S]$. By triangle inequality, $cost(e) \leq 2cost(e_j)$ for any edge $e = (u, v)$ in the star. Note $cost(e_j) \leq OPT$. We have $cost(e) \leq 2OPT$. This completes the proof.

Next, we discuss how to use the 2-approximation algorithm for the p-CDC problem to construct Simple-p-RSPA. There is a major challenge to apply the 2-approximation algorithm for the p-CDC problem to our problem of interest, since we are looking for a subset of nodes with size exactly p. If we enforce this, then the algorithm may not produce a feasible solution with insufficient W. To address this concern, it is necessary for us to find a valid W. To this purpose, we may set W to be $W_i = \sum_{v \in V_i} deg(v_i, G)$, where V_i is the subset of the first i nodes with largest node degree, for each $i = n, n-1, \cdots, 1$ and apply the modified 2-approximation algorithm for the p-CDC problem. Finally, we choose the one out of all feasible outputs such that the objective of p-BRSP in Eq. (1) is minimized. Note that this final result still has the approximation factor of 2 with respect to Objective 2.

4 Concluding Remarks

This paper introduces a new application of the information which can be extracted from social network information. The main focus of this paper is to use the information for biased survey so that more amount of minority opinions can be heard. We formalize the problem of our interest as a new optimization problem with two separate objectives. Then, we propose two heuristic algorithms for the problem, each of which has an approximation factor with respect to each of the objectives. As a future work, we will conduct simulations to evaluate the performance of the proposed algorithms. We are also interested in using real data to see if our approach is in fact effective. We also plan to use apply approach to identify the users with less satisfaction and compensate them so that the negative reputation of a new product can be suppressed. We believe this can compensate the existing approaches which focus on how to compensate users to spread positive reputation [11, 12].

References

1. E. Anifantis, E. Stai, V. Karyotis, and S. Papavassiliou, “Exploiting Social Features for Improving Cognitive Radio Infrastructures and Social Services via Combined MRF and Back Pressure Cross-layer Resource Allocation,” *Computational Social Networks*, vol. 1, issue 4, 2014.
2. C. Ai, W. Zhong, M. Yan, and F. Gu, “A Partner-matching Framework for Social Activity Communities,” *Computational Social Networks*, vol. 1, issue 5, 2014.
3. M. Ventresca and D. Aleman, “Efficiently Identifying Critical Nodes in Large Complex Networks,” *Computational Social Networks*, vol. 2, issue 6, 2015.

4. V. Vehovar and K.L. Manfreda, "Overview: Online Surveys," *The SAGE Handbook of Online Research Methods*, London: SAGE (edited by N.G. Fielding, R.M. Lee, and G. Blank), pp. 177-194, 2008.
5. B. Duffy, K. Smith, G. Terhanian, and J. Bremer, J, "Comparing Data from Online and Face-to-face Surveys," *International Journal of Market Research*, vol. 47, no. 6, pp. 615-639, 2005.
6. D. Kim, J. Zhong, M. Lee, D. Li, Y. Li, and A.O. Tokuta, "Efficient Respondents Selection for Biased Survey using Homophily-high Social Network Graph," submitted to *Optimization Letters (OPTL)*, under 3rd review.
7. H. Bisgin, N. Agarwal, and X. Xu, "A Study of Homophily on Social Media," *World Wide Web*, vol. 15, issue 2, pp. 213-232, 2012.
8. E. Singer and C. Ye, "The Use and Effects of Incentives in Surveys," *The Annals of the American Academy of Political and Social Science*, vol. 645, issue 1, pp. 112-141, 2013.
9. D.S. Hochbaum and D.B. Shmoys, "A Best Possible Heuristic for the k-Center Problem", *Mathematics of Operations Research*, vol. 10, issue 2, pp. 180184, 1985.
10. D. Kim, J. Zhong, M. Lee, D. Li, and A.O. Tokuta, "Efficient Respondents Selection for Biased Survey using Online Social Networks," *Proceedings of the 2nd Workshop on Computational Social Networks (CSoNet 2014)* in conjunction with COCOON 2014, August 4-5, 2014, Atlanta, USA.
11. Z. Lu, L. Fan, W. Wu, B. Thuraisingham, and K. Yang, "Efficient Influence Spread Estimation for Influence Maximization under the Linear Threshold Model," *Computational Social Networks*, vol. 1, issue 2, 2014.
12. H. Kim, K. Beznosov, and E. Yoneki, "A Study on the Influential Neighbors to Maximize Information Diffusion in Online Social Networks," *Computational Social Network*, vol. 2, issue 3, 2015.