# On Social Network Firewall Selection

Anisha Mazumder and Arunabha Sen
Computer Science and Engineering Program
School of Computing, Informatics and Decision System Engineering
Arizona State University
Tempe, Arizona 85287
Email: {anisha.mazumder,asen }@asu.edu

***Abstract*—It is common knowledge that the decision of an individual regarding adoption of a product or technology is, more often than not, heavily influenced by their friends and acquaintances. In real world, there are different competing products and innovations that try to garner as many loyal followers as possible. Over the past few years, there has been a significant interest in the research community to study social network problems with a flavor of competition. Such problems often focus on identification of a set of people in a given social network by the competing players in order to achieve some goal. In this paper, we introduce the *weighted Segregating Vertex Set (wSVS)* problem, in which we are given a weighted undirected graph with a subset of nodes identified as the seedset of the first player and the goal for the second player is to identify a subset of nodes (firewall) of minimum cumulative weight, such that the total weight of the nodes reachable by the first player is strictly less than the total weight of the nodes not reachable by the first player. Thus, the second player tries to contain the *reach* of the first player within the social network community. This problem is also relevant for containment of disease in epidemiology, containment of forest fire and several other domains. We prove that this problem is NP-complete and provide an optimal solution through the use of Mixed Integer Linear Programming. We also provide a heuristic solution for the wSVS problem and show its efficacy through detailed experimentation. Our heuristic solution delivers near optimal solution in lesser time compared to that needed to find the optimal solution.**

## I. Introduction

Over the past few years, particularly with the boom of the Online Social Networks (OSNs), there has been a considerable amount of research interest in the study of problems pertaining to OSNs. These include - problems related to Influence Maximization [1], [2], Influence Blocking Maximization[3], [4], Community Detection [5] and others.

In particular, such problems in a two-player setting has gained considerable interest in the recent past. The motivation for such interest is two fold. First, one's decision to adopt a new technology or a product is often influenced by one's friends and family. Second, in the real world, there are always more than one competing technology or product to choose from and the product manufacturers are often promoting their product so that the general population adopt their product instead of their competitor's. Each competing manufacturer (or player) in the market would like to win over as many loyal followers as possible but at the same time, she would also like to try to "contain" the spread of the other player's product. Besides, each individual, represented as a node in the social network, can be of different *utility* or *value* to the player and as such the player may assign a different weightage to each individual, taking into account an individual's age, social status, educational status, economic status and various other factors. Also, in order to win over a population, a company needs to incentivize individuals to subscribe to their products. Certainly, a company would like to spend as little of its resources in incentivization that will achieve its objective, which may be to capture a certain percentage of the market share or other similar goals. We can also think about the weightage of an individual from a company's perspective as the amount of incentivization needed by the individual to become a loyal customer of the company.

Motivated by such considerations, in this research, we consider the problem of *weighted Segregating Vertex Set problem (wSVS)*. In this problem, we are given a weighted undirected social network graph and we consider that there are two players - A and B who are competing against each other. Let us consider that player A has already selected a subset of the population, which we refer to as the *seedset* of player A. This means that player A has already won over the loyalty of these people and as a result these people are unavailable to the second player i.e., player B. Now, player B would like to *contain* the *possible* spread of player A's influence by selecting a *firewall* of nodes from the network. This firewall will ensure that the spread of player A's influence is limited to only a part of the network such that the total weight of the nodes beyond the reach of player A is more that the total weight of the nodes within the reach of player A. This will in turn mean that even if player A is able to *influence* all the people within the reach of her seedset, she would never be able to conquer half or more than half the social network - this will ensure that player B a reasonable chance to have a majority of the market share in this social network. However, from player B's perspective, she would like to construct this firewall with the least amount of investment for incentivization. Thus, the problem has a definite flavor of the *vertex separator problem* [6] that has been studied extensively in the Computer Science literature . The wSVS problem is formally defined later.

It may be noted that the wSVS problem is similar to an extent to the Influence Blocking Maximization (IBM) problem [3]. But, wSVS is significantly different from the IBM problem because in the IBM problem, it is not required that the influence of the first player be contained within less than half of the entire network. We can conceive many scenarios where the wSVS problem formulation can be used as a model of a



978-1-4673-8579-4/16/$31.00 ©2016 IEEE

real life situation. Consider two competing players, A and B, such as two house builders or two companies manufacturing some heavy products such as cars or expensive electronics. It is clear that an individual having made an expensive investments in such items at a certain point in time is extremely unlikely to invest again in recent future. If player A starts marketing while player B realizes that its product can only be ready in about half a year or so, player B would like to stop customers from buying from player A in the meantime. Such scenarios are exactly captured by the wSVS problem formulation.

Even though the primary setting of the wSVS problem is in the domain of social networks, it may be noted that in abstraction, the problem can be easily transported to the domains of damage control or epidemic control. Consider, for example, that a part of a population has become infected by a contagious disease. A primary prevention method by health service officials could be to try to vaccinate individuals in such a way that the vaccinated individuals form a firewall that ensures that at least a majority of the population would be safe-guarded against the epidemic. Similarly, in a distributed data storage network, if a part of the network becomes compromised, it may be prudent to build such a firewall so that more than half of the network is protected. Although it might appear that these scenarios are completely different from the social network scenarios described earlier, in abstraction the underlying problem is the same.

The rest of the paper is organized as follows - in section II, we discuss related works for the wSVS problem; in section III, we formally define the wSVS problem and provide a formal proof of hardness of the wSVS problem; in section IV, we provide an optimal solution to the wSVS problem using Mixed Integer Linear Program formulation and a heuristic solution to solve the wSVS problem in polynomial time. In section V, we demonstrate the efficacy of our heuristic solution through experimentation on three families of network namely Barabasi-Albert graph, Erdos-Renyi graph and Watts-Strogatz graph, besides a real world Facebook dataset. Our heuristic computes near optimal solution in lesser time compared to that needed by the optimal solution. Finally, section VI concludes our paper.

## II. Related Works

The research community in recent times has seen a heightened level of interest in *social computing* or *social network* problems. In the Influence Maximization (IM) problem [1], [7], given a network, a player wants to select the minimum number of nodes that she must incentivize so as to obtain the maximum number of loyalists in the network. The natural generalization of this problem is to extend the setting to a multi-player scenario [2], where there might be multiple players trying to capture a given market. A number of different models of propagation of influence through a social network has been proposed - these include probabilistic influence propagation [1] as well as deterministic influence propagation [8]. [9] is a variation of the IM problem in a two player setting where the first player has already selected a seedset and the goal of the second player is to select a subset (of minimum cardinality) of nodes from the remaining population such that after the influence from the seedset of both the players propagate, the expected number of nodes influenced by second player is strictly greater than that by the first player.

Influence Blocking Maximization (IBM) is the problem where the second player attempts to stall the influence propagation of the first player (under a budget constraint) to as high an extent as possible through strategic selection of a seedset that could initiate influence propagation of its own. [3] proposes a solution technique for the IBM problem under competitive linear threshold (CLT) model. [10], [11] are works on variations of the IBM problem from a game theoretic perspective.

Another line of research involves the propagation of negative influence, contagious diseases and so on. [12] studies the problem of minimizing the influence of negative information. In [13], given a graph where a node has been marked to be the 'source', the goal is to obtain a cut minimizing the number of nodes on the partition containing the source, such that the capacity of the cut does not exceed a pre-determined budget.

Although all these studies delve into different aspects of social network problems and there are numerous studies on the Set Partition Problem [14],[15], to the best of our knowledge none of them focus on problems related to the wSVS problem where there is a strict constraint that the weighted sum of the nodes reachable from the first player is to be restricted to less than half of the total weighted sum of all the nodes in the network. Here, by reachability, we generalize to any model of influence propagation. This means that irrespective of the model of propagation considered, the total weighted influence of the first player must be restricted to less than half of the total weight of the entire network.

## III. Problem Formulation

In this section, we provide a formal statement of the weighted Segregating Vertex Set (wSVS) problem. The wSVS problem is defined as follows:
Given a weighted undirected graph $G$ and a subset $R \subset V(G)$, find a least weighted vertex separator $C$ ($C \subset V(G) \setminus R$) such that $C$ divides the graph G into two components (say, $P$ and $Q$), where (i) $R \subseteq P$ and (ii) $weight(P) < weight(Q) + weight(C)$.
It may be noted that $R$ in this formulation represents the seedset of the first player (referred to as $seedset_A$) in discussion earlier. For a subset of nodes $V' \subset V(G)$, we define $weight(V') = \sum_{v \in V'} w_v$, where $w_v$ is the weight of the node $v$. For the decision version of the wSVS problem, the question is as follows: Is there a vertex separator $C$ of weight at most $B$, such that

$$weight(P) < weight(Q) + weight(C)\text{....... (i)}$$

We assume that $\sum_{v \in seedset_A} w_v < \sum_{v \in V(G) \setminus seedset_A} w_v$ i.e., if the second player selects all the remaining nodes after selection of the nodes by the first player, constraint (i) will certainly be satisfied.

**Theorem III.1.** *The wSVS problem is NP-complete.*

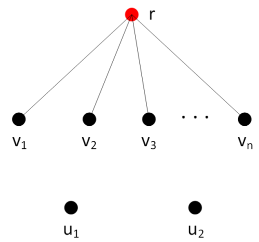



Fig. 1. Construction for hardness proof of wSVS problem

*Proof:* Given an instance of the wSVS problem and a solution to the problem (i.e., the vertex separator $C$), it is easy to verify in polynomial time whether $C$ indeed provides a feasible solution. Accordingly, wSVS is in NP. We prove that the wSVS problem is NP-complete by reducing the Set Partition Problem, a well-known NP complete problem, to it. *Set Partition Problem:* An instance of the Set Partition problem is made up of a set of integer numbers $S$, and it asks the following question: Is there a partition of the set $S$ into two subsets $A$ and $\bar{A}$ ($\bar{A} = S \setminus A$), such that the sum of the integers in the two sets $A$ and $\bar{A}$ are equal (i.e., $\sum_{x \in A} x = \sum_{x \in \bar{A}} x$). Given an instance of the set partition problem, we construct an instance of the wSVS problem as shown in Fig. 1:

Let us enumerate the numbers in $S$ as $s_1, s_2, \ldots, s_n$. For each number $s_i \in S$, we create a node $v_i$ of weight $w_{v_i} = value\ of\ s_i$. For e.g., if the enumeration of $S$ is as follows $5, 3, 6, \ldots$, then $w_{v_1} = 5, w_{v_2} = 3, w_{v_3} = 6, \ldots$. For notational purpose, let us denote $V = \{v_1, v_2, \ldots, v_n\}$. We add a single red node $r$ (forming the seedset of the first player) and add undirected edges $(r, v_i), 1 \le i \le n$. Also, we add two disjoint nodes $u_1$ and $u_2$, where $w_r = w_{u_1} = w_{u_2} = 1$. Let, $T = \sum_{i \in S} i$ i.e., sum of all the elements of $S$. Let $B = \frac{T}{2}$. We next provide both directions of the NP-hardness proof.

**If case**: If there is a partition of $S$ into $A$ and $\bar{A}$, then $C = A$. So, $weight(C) = B = \frac{T}{2}$ and $P = \{r\} \cup \bar{A}$ and $Q = \{u_1, u_2\}$ and thus condition (i) is satisfied.

**Only if case**: If there is a yes answer for the wSVS problem, let $C$ be the corresponding vertex separator where $weight(C) \le \frac{T}{2}$ and condition (i) is satisfied.

Now, if $weight(C) < \frac{T}{2}$, then $weight(V \setminus C) > \frac{T}{2}$. Since, all the weights are integers, we can say that

$$weight(V \setminus C) - weight(C) \ge 1 = weight(\{u_1, u_2\}) - w_r$$
$$\implies weight(V \setminus C) + w_r \ge weight(C) + weight(\{u_1, u_2\})$$

which violates condition (i). This implies that $weight(C) = \frac{T}{2} = weight(V \setminus C)$ and a partition of $S$ exists. ■

## IV. Solutions for the wSVS problem

In this section, we provide optimal and heuristic solutions for the wSVS problem.

### A. Optimal solution

We provide an optimal solution for the wSVS problem using Mixed Integer Linear Program (MILP) formulation. Given a graph $G$, we define the following variables:

For each node $v \in V(G)$:
$X_v = 1$, if node $v$ belongs to $P$ and 0 otherwise,
$Y_v = 1$, if node $v$ belongs to $Q$ and 0 otherwise.
Let $V(G)$ and $E(G)$ denote the vertex set and edge set of $G$ respectively. The MILP can now be written as:

$$min \sum_{v \in V} w_v \times (1 - X_v - Y_v)$$

$$X_u + Y_v \le 1 \quad \forall (u, v) \in E(G) \tag{1}$$

$$X_v + Y_v \le 1 \quad \forall v \in V(G) \tag{2}$$

$$X_v = 1 \quad \forall v \in R \tag{3}$$

$$\sum_v (w_v \times X_v) < \sum_v (w_v \times Y_v) + \sum_v (w_v \times (1 - X_v - Y_v)) \tag{4}$$

$$X_v \in \{0, 1\}; Y_v \ge 0;$$

The objective function implies that we want to minimize the total weight of the nodes in the separator $C$ (or equivalently maximize the total weight of the nodes which are as assigned to components $P$ and $Q$). Constraint (1) implies that there can be no edge between $P$ and $Q$. Constraint (2) implies that $P \cap Q = \emptyset$. Constraint (3) implies that $R \subseteq P$ or equivalently $seedset_A \subseteq P$. Finally, constraint (4) implies that $\sum_{v \in P} w_v < \sum_{v \in Q} w_v + \sum_{v \in C} w_v$.

### B. Heuristic Solution

Since, solving MILP can be NP hard, we provide a heuristic solution by solving the Linear Program (LP) with relaxed integrality constraints of the MILP formulation given in section IV-A and then using the output of the LP in order to obtain the final firewall.

---

**Algorithm 1:** Heuristic algorithm for solving wSVS problem

---

Solve the relaxed linear program formulation for the wSVS problem
Initialize $C$ as the empty-set
**while** total weight of nodes reachable from $seedset_A$ is greater than or equal to total weight of nodes unreachable from $seedset_A$ **do**
    Add node $v$ to $C$ where $v$ has the highest value of $1 - X_v - Y_v$ among all $v' \in V(G) \setminus C$; breaking ties randomly
**end while**
Return $C$

---

*1) Description:* The output of the relaxed LP formulation for the MILP formulation given in section IV-A gives fractional values (referred to as *lp* values by us) to the nodes of the input social network graph $G$. The MILP formulation very evidently has no properties such as half integrality or total-unimodularity. So, our heuristic solution performs rounding of the *lp* values for the nodes. Since, the heuristic will select some nodes from $V(G) \setminus seedset_A$ for the final output set

$C$, constraints $1-3$ of the MILP formulation are automatically satisfied. Besides, we are explicitly checking whether constraint (4) is satisfied in the condition of the **while** loop of line 3 and so constraint (4) is also always satisfied. The heuristic selects node in non-increasing order of the values $(1-X_v-Y_v), \forall v \in V(G)$ assigned by the solution of the relaxed LP formulation, breaking ties randomly. The intuition behind this is that higher the value $(1-X_v-Y_v)$ for a node $v$, the greater fraction of the node $v$ has been used by the linear program solution in its final solution. Since, we can not use a fraction of a node as the final solution of wSVS problem, the heuristic includes the entire node in its solution set $C$. The efficacy of this simple algorithm is proven empirically through our experimentation provided in section V.

*2) Time Complexity:* Step 1 of solving the relaxed LP formulation of the MILP formulation given in section IV-A takes polynomial time. The condition for the **while** loop of steps $3-5$ can be computed through a graph traversal algorithm such as depth-first search or breadth-first search which takes $O(|V(G)|+|E(G)|)$ time. We can sort and store the $1-X_v-Y_v$ values as a pre-computation step for efficient computation of step 4. A standard sorting algorithm on $O(n)$ nodes takes $O(n\ log\ n)$ time. The **while** loop can be executed for a maximum of $V(G)-|seedset_A|$ which is $O(n)$. Hence, Algorithm 1 runs in polynomial time in $|V(G)|$.

## V. Experimental Results and Discussions

In this section we present results of our experimentations to prove the effectiveness of our simple heuristic algorithm. For this, on one hand, we consider three families of graphs namely - Barabasi-Albert graph [16], Erdos-Renyi graph [17] and Watts-Strogatz graph [18] on 100 nodes and different parameters relevant for the particular type of graph. And on the other hand, we conisder an ego-Facebook real world dataset (freely available for download from https://snap.stanford.edu/data/) consisting of $4039$ nodes and $88,234$ edges. For generating data for the three graph families, we have used the NetworkX python library and because these are random graphs, we have experimented with $500$ instances for each set of parameters. Barabasi-Albert graphs are random scale-free networks generated using a preferential attachment algorithm. This means that a graph of $n$ nodes is constructed through the process of attaching new nodes each with a specified number of edges that are preferentially attached to existing nodes with high degree. The reason we have chosen Barabasi-Albert graph as one of the families of graphs for our experiment is that such scale-free networks are frequently observed in different social networks. For Barabasi-Albert graphs, in the context of wSVS problem, the parameters that we have experimented with are - (i) *m* i.e., the number of edges to attach from a new node to existing nodes and it has been varied from $10$ to $50$ in steps of $10$ as well as (ii) the size of seedset of player A and it has been varied from $5$ to $25$ in steps of $5$. Erdos-Renyi graphs are a family of random graphs. We have considered Erdos-Renyi graphs as baseline graph family. The parameters of Erdos-Renyi graphs, in the context of wSVS problem, that we have experimented with are - (i) *p* i.e., the probability that each edge exists and it has been given values of $0.1, 0.25, 0.5, 0.75$ as well as (ii) the size of the seedset of player A and it has been varied from $5$ to $25$ in steps of $5$. Finally, Watts-Strogatz graphs is a family of graphs with *small-world* properties, which include high clustering properties and short average path lengths. Such networks are also seen in social networks. For the Watts-Strogatz graphs, the parameters that we have experimented with in the context of the wSVS problem are - (i) *k* i.e., each node is connected to *k* nearest neighbors in ring topology and it has been given values of $4, 10, 16$, (ii) *p* i.e., the probability of rewiring each edge and it has been given values of $0.25, 0.5, 0.75$, and (iii) size of the seedset of player A and it has been varied from $5$ to $25$ in steps of $5$. For the Facebook dataset, we have used the data as is and have considered the entire network graph with all the nodes and all the edges from all the egonets combined. For all the datasets, the degree of each node in the graph is assigned as its weight. With these datasets in hand, we have computed the optimal solution for the wSVS problem by solving the MILP formulation as provided in section IV-A by using CPLEX Optimization Studio 12.5. The heuristic solution is implemented in Java and run for these datasets on a Windows 7 Intel core i7 laptop. The results are plotted in Fig. 2, where for each graph, we plot along x-axis the percentage of total number of nodes selected by the first player as her seedset - for e.g., at $x = 15$, we plot the results when the first player has selected $15\%$ of the total number of nodes as her seedset. And along y-axis, we plot the ratio of the weight of the seedset selected by the heuristic solution to the weight of the seedset selected by the optimal solution. In all of our experiments, the heuristic has obtained a solution value within a factor of 2 of the optimal solution value but in lesser time compared to that required to compute the optimal solution.

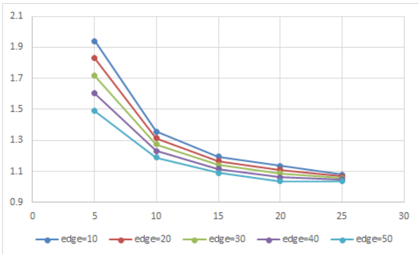
## VI. Conclusion

We have introduced a new problem involving a two-player game where, given a social network represented as a weighted undirected graph, the goal is to select a subset of nodes of minimum weight for the second player from the graph excluding a subset of nodes marked as seedset of the first player, such that the total weight of the nodes reachable by the first player's seedset is less than half the total weight of the nodes of the graph. We prove that the problem is NP-complete and provide optimal and heuristic solutions for the same.

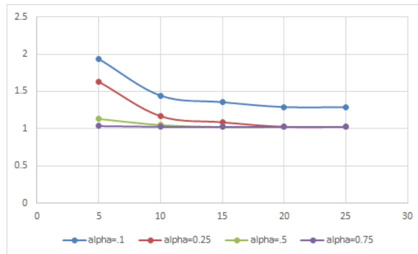
## References


[1] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 137–146.

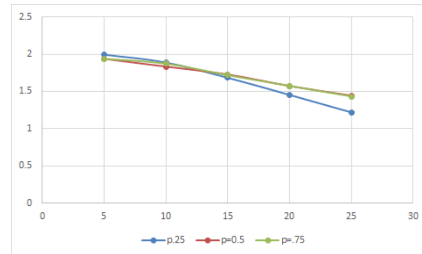
[2] T. Carnes, C. Nagarajan, S. M. Wild, and A. Van Zuylen, "Maximizing influence in a competitive social network: a follower's perspective," in *Proceedings of the ninth*



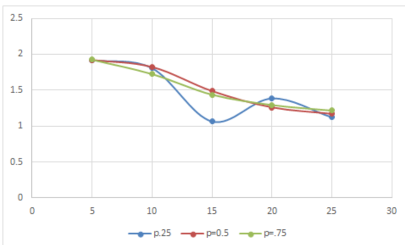

(a) Results for Barabasi-Albert Network



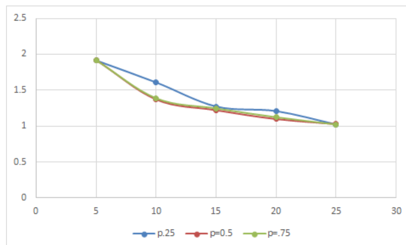

(b) Results for Erdos-Renyi Network



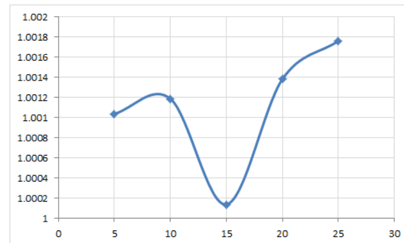

(c) Results for Watts-Strogatz graph for $k = 4$




(d) Results for Watts-Strogatz graph for $k = 10$




(e) Results for Watts-Strogatz graph for $k = 16$



(f) Results for Facebook data

Fig. 2. Experimental results for Barabasi-Albert Network, Erdos-Renyi Network, Watts-Strogatz Network and Facebook graph for different parameters

*international conference on Electronic commerce*. ACM, 2007, pp. 351–360.

[3] X. He, G. Song, W. Chen, and Q. Jiang, “Influence blocking maximization in social networks under the competitive linear threshold model.” in *SDM*. SIAM, 2012, pp. 463–474.

[4] G. Tuli, C. J. Kuhlman, M. V. Marathe, S. Ravi, and D. J. Rosenkrantz, “Blocking complex contagions using community structure,” 2012.

[5] J. Leskovec, K. J. Lang, and M. Mahoney, “Empirical comparison of algorithms for network community detection,” in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 631–640.

[6] M. D. Biha and M.-J. Meurs, “An exact algorithm for solving the vertex separator problem,” *Journal of Global Optimization*, vol. 49, no. 3, pp. 425–434, 2011.

[7] M. Kimura, K. Saito, and R. Nakano, “Extracting influential nodes for information diffusion on a social network,” in *AAAI*, vol. 7, 2007, pp. 1371–1376.

[8] P. Shakarian and D. Paulo, “Large social networks can be targeted for viral marketing with small seed sets,” in *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*. IEEE Computer Society, 2012, pp. 1–8.

[9] S. Shirazipourazad, B. Bogard, H. Vachhani, A. Sen, and P. Horn, “Influence propagation in adversarial setting: how to defeat competition with least amount of investment,” in *Proceedings of the 21st ACM international conference on Information and knowledge management*. ACM, 2012, pp. 585–594.

[10] J. Tsai, T. H. Nguyen, and M. Tambe, “Security games for controlling contagion.” in *AAAI*, 2012.

[11] M. Jain, D. Korzhyk, O. Vaněk, V. Conitzer, M. Pěchouček, and M. Tambe, “A double oracle algorithm for zero-sum security games on graphs,” in *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, 2011, pp. 327–334.

[12] S. Wang, X. Zhao, Y. Chen, Z. Li, K. Zhang, and J. Xia, “Negative influence minimizing by blocking nodes in social networks,” in *Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence*, 2013.

[13] A. Hayrapetyan, D. Kempe, M. Pál, and Z. Svitkina, “Unbalanced graph cuts,” in *Algorithms–ESA 2005*. Springer, 2005, pp. 191–202.

[14] M. Conforti, M. Di Summa, F. Eisenbrand, and L. A. Wolsey, “Network formulations of mixed-integer programs,” *Mathematics of Operations Research*, vol. 34, no. 1, pp. 194–209, 2009.

[15] M. Padberg and G. Rinaldi, “A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems,” *SIAM review*, vol. 33, no. 1, pp. 60–100, 1991.

[16] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[17] P. Erd6s and A. Rényi, “On the evolution of random graphs,” *Publ. Math. Inst. Hungar. Acad. Sci*, vol. 5, pp. 17–61, 1960.

[18] D. J. Watts and S. H. Strogatz, “Collective dynamics of small-worldnetworks,” *nature*, vol. 393, no. 6684, pp. 440–442, 1998.