Theoretical Computer Science 412 (2011) 265–269

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

On positive influence dominating sets in social networks[☆]

Feng Wang[a,*], Hongwei Du[b], Erika Camacho[a], Kuai Xu[a], Wonjun Lee[c], Yan Shi[d], Shan Shan[d]

[a] Division of Mathematical and Natural Sciences, Arizona State University, Phoenix, AZ 85069, USA
[b] Department of Computer Science, Illinois Institute of Technology, Chicago, IL 60616, USA
[c] Division of Computer and Communication Engineering, Korea University, Seoul, Republic of Korea
[d] Department of Computer Science, University of Texas at Dallas, Richardson, TX 75080, USA

ARTICLE INFO

Keywords:
Dominating set
Positive influence dominating set
Social networks
APX-hard

ABSTRACT

In this paper, we investigate the positive influence dominating set (PIDS) which has applications in social networks. We prove that PIDS is APX-hard and propose a greedy algorithm with an approximation ratio of $H(\delta)$ where H is the harmonic function and δ is the maximum vertex degree of the graph representing a social network.

1. Introduction

Dominating set has been extensively studied and been adopted in many real-life applications. For example, it has been utilized in wireless networks to address media access, routing, power management, and topology control issues [1,7,8,10,11]. Recently, [5,9] studied the dominating set in social networks. [5] proposed a greedy approximation algorithm and proved that the algorithm gives a $1 + O(1)$ approximation with a small constant in $O(1)$ to the dominating set problem in a power-law graph. [9] introduced a variation of dominating set, called Positive Influence Dominating Set (PIDS), as follows: Given a graph $G = (V, E)$, a PIDS is a subset D of V such that any node v in V is dominated by at least $\lceil \frac{d(v)}{2} \rceil$ nodes (that is, v has at least $\lceil \frac{d(v)}{2} \rceil$ neighbors) in D where $d(v)$ is the degree of node v. Note that there are two requirements for PIDS: Firstly, every node not in D has at least half of its neighbors in D, secondly every node in D also has at least half of its neighbors in D.

Constructing a minimum PIDS of a social network which consists of individuals with a certain type of social problem (such as drinking, smoking and drug related issues) is helpful for the success of intervention programs. Intervention programs are important tools to help combat some of the social problems and consist of disseminated education and therapy via mail, Internet, or face-to-face interviews. In a social setting, people can have both positive and negative impacts on each other, and a person can take and move among different roles since they are affected by their peers. For example, within the context of drinking problem, a binge drinker can be converted to an abstainer through intervention program and have positive impact on his direct friends (called neighbors). However, he might turn back into a binge drinker and have negative impact on his neighbors if many of his friends are binge drinkers. Ideally, we want to educate all binge drinkers, since this will reduce the possibility of converted binge drinker being influenced by his binge drinker friends who are not chosen in the intervention program. On the other hand, due to the budget limitations, it is impossible to include all the binge drinkers in the intervention program. Therefore, how to choose a subset of individuals to be part of the intervention program so that the effect of the intervention program can spread through the whole group under consideration becomes an important research problem.

☆ This research was jointly supported by MEST, Korea under WCU (R33-2008-000-10044-0), and National Science Foundation of USA under grants CNS0831831, CNS0831579 and CCF0728851.

* Corresponding author. Tel.: +1 6025438150.
E-mail addresses: fwang25@asu.edu (F. Wang), david.duhongwei@gmail.com (H. Du), erika.camacho@asu.edu (E. Camacho), kxu01@asu.edu (K. Xu), wlee@korea.ac.kr (W. Lee), yxs055100@utdallas.edu (Y. Shi), sxs063100@utdallas.edu (S. Shan).

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2009.10.001

PIDS is a plausible solution since PIDS can guarantee that by selecting PIDS nodes to participate in the intervention program, each individual in the social network has more positive neighbors than negative ones thus to ensure that the intervention can result in a globally positive impact on the entire social network.

In this paper, we proved the complexity of PIDS problem in general graphs by constructing a L-reduction from vertex cover problem in cubic graphs to PIDS. Furthermore, we propose a greedy approximation algorithm and give theoretical analysis about its approximation ratio. The results show that PIDS is APX-hard and the greedy algorithm produces an approximation solution within a factor of $H(\delta)$ from optimal where H is the harmonic function and δ is the maximum vertex degree of input graph.

The rest of this paper is organized as follows. In Section 2, we analyze the complexity of PIDS problem. Section 3 presents a greedy approximation algorithm for PIDS and analyzes its approximation ratio. Section 4 concludes this paper and discusses the future work.

2. Complexity

Theorem 1. *PIDS is APX-hard.*

Proof. It can be found in [4] that the vertex cover problem in cubic graph (let us call it VC-cubic) is APX-complete. A cubic graph is a graph with every vertex's degree of exactly three. Given a cubic graph, the VC-cubic is to find a minimum vertex cover, i.e., a minimum-cardinality subset C of vertices such that every edge has an endpoint in C. To prove our theorem, we construct a L-reduction from VC-cubic to PIDS.

Consider a cubic graph $G = (V, E)$ as in Fig. 1(a) as an instance of VC-cubic. First, we construct a bipartite graph $B = (V \cup U_E, F)$ as in Fig. 1(b). Each vertex $u_e \in U_E$ is corresponding to an edge $e \in E$. An edge (v, u_e) exists in F if and only if $v \in e$, i.e., v is an endpoint of e.

Now, for each vertex $u_e \in U_E$, we attach a path (u_e, a_e, b_e, c_e) and for each vertex $v \in V$, we attach three paths (v, p_v, q_v, r_v), (v, p'_v, q'_v, r'_v) and (v, p''_v, q''_v, r''_v). The construction is illustrated in Fig. 1(c). The obtained graph is denoted by G'. As can be seen from the illustration, every node in U_e has degree of 3, every node in V has degree of 6 in the constructed graph G'.

Next, we show a claim.

Claim 2. *G has a vertex cover of size at most k if and only if G' has a positive influence dominating set of size at most $k + 9n$ where $n = |V|$.*

Proof. First, suppose G has a vertex cover C of size k. Let $D = C \cup \{a_e, b_e \mid e \in E\} \cup \{p_v, q_v, p'_v, q'_v, p''_v, q''_v \mid v \in V\}$. We can easily verify that D is a positive influence dominating set of G'. Note that in a cubic graph G, $|E| = 3|V|/2$. Therefore, $|D| = |C| + 2 * |E| + 6 * |V| = |C| + 9|V| = k + 9n$.

Conversely, suppose G' has a positive influence dominating set D of size k', we show that G has a vertex cover of size $k' - 9n$. This will complete the proof of the claim.

Note that D must contain b_e for every $e \in E$ in order to make c_e have a neighbor in D. Next, we note that either a_e or c_e in D in order to make b_e have a neighbor in D. Since we can always replace c_e by a_e, which does not increase the size of D, we may assume $c_e \notin D$ and $a_e \in D$ without loss of generality. Similarly, we may assume $p_v, q_v, p'_v, q'_v, p''_v, q''_v \in D$ and $r_v, r'_v, r''_v \notin D$ for every $v \in V$. Under these assumptions, every v has at least half of its neighbors (p_v, p'_v, p''_v) in D. Furthermore, a_e which has degree of 2 has one neighbor b_e in D. These imply that every neighbor of u_e has at least half of its neighbors in D even if $u_e \notin D$. Therefore, we may assume without loss of generality that none of u_e for $e \in E$ belongs to D.

Now we know that $D - V$ consists of only $a_e, b_e, p_v, q_v, p'_v, q'_v, p''_v, q''_v$ hence we have $|D \cap V| = k' - 2*|E| - 6*|V| = k' - 9n$. Note that each u_e for $e \in E$ has degree exactly three in G'. Moreover, it already has one neighbor a_e in D. So, u_e must have at least one neighbor in $D \cap V$. This means that $D \cap V$ is a vertex cover of G. □

An immediate consequence from this claim is that G has a minimum vertex cover of size $opt_{VC-cubic}(G)$ if and only if G' has a minimum positive influence dominating set of size $opt_{PIDS}(G') = opt_{VC-cubic}(G) + 9n$. Since each vertex has degree of three in G, we have $n/2 = |E|/3 \leq opt_{VC-cubic}(G)$. Plugging $n = (opt_{PIDS}(G') - opt_{VC-cubic}(G))/9$ into the inequality, we have

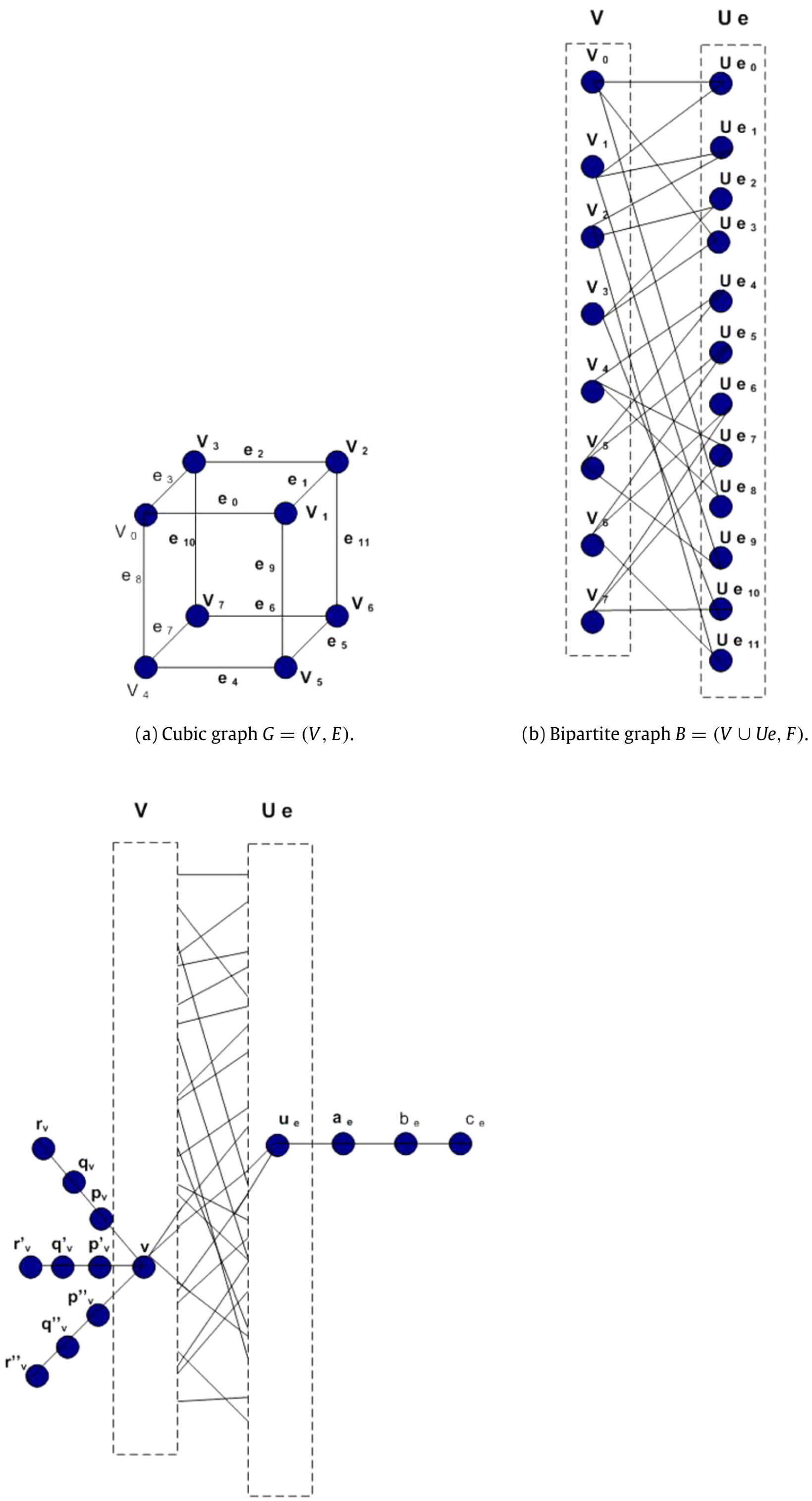
$$opt_{PIDS}(G') \leq 19 \cdot opt_{VC-cubic}(G).$$

Moreover, from the proof of Claim 2, we see that if G' has a positive influence dominating set D, then we can construct, in polynomial time, a vertex cover $C = D \cap V$ of G with size $|D| - 9n$. Therefore,

$$||C| - opt_{VC-cubic}(G)| = ||D| - opt_{PIDS}(G')|.$$

This means that VC-cubic is L-reducible to PIDS. □

APX-hardness of PIDS means that if $NP \neq P$, then PIDS has no PTAS (polynomial-time approximation scheme). If PIDS has a polynomial-time constant-approximation, then we can conclude that PIDS is in APX and hence is APX-complete. However, so far, we are unable to find a constant-approximation. Instead, we would like to present a polynomial-time $H(\delta)$-approximation for PIDS where δ is the maximum vertex degree of the input graph and $H(\delta)$ is the harmonic function, i.e., $H(\delta) = \sum_{i=1}^{\delta} \frac{1}{i} \leq 1 + \ln \delta$.



(a) Cubic graph $G = (V, E)$.

(b) Bipartite graph $B = (V \cup Ue, F)$.

(c) Construction of G'.

Fig. 1. Illustration of the *L*-reduction.

3. Approximation

In this section, we describe a greedy approximation algorithm and analyze its approximation ratio. To do so, we first define a function f as follows.

Consider a graph $G = (V, E)$ as an instance of PIDS. For any vertex subset $A \subseteq V$, $n_A(v)$ denotes the number of neighbors of v in A. For a vertex v with degree $deg(v)$ in G, we denote $h(v) = \lceil deg(v)/2 \rceil$. Now, define

$$f(A) = \sum_{v \in V} \min(h(v), n_A(v)).$$

The following lemma states important properties of f.

Lemma 3. (1) $f(\emptyset) = 0$.
(2) $f(A) = \sum_{v \in V} h(v)$ *if and only if A is a positive influence dominating set.*
(3) *If* $f(A) < \sum_{v \in V} h(v)$, *then there exists a vertex* $u \in V - A$ *such that* $f(A \cup \{u\}) > f(A)$.

Proof. Note that $n_\emptyset(v) = 0$ for all $v \in V$. Therefore (1) holds.

For (2), we note that $f(A) = \sum_{v \in V} h(v)$ if and only if $h(v) \le n_A(v)$ for every $v \in V$ if and only if A is a positive influence dominating set.

To see (3), note that $f(A) < \sum_{v \in V} h(v)$ implies the existence of $v \in V$ such that $h(v) > n_A(v)$. Let u be a neighbor of v which is not in A, then $f(A \cup \{u\}) > f(A)$. □

The above lemma suggests a greedy algorithm as follows.

Greedy Algorithm
$A \leftarrow \emptyset$;
while $f(A) < \sum_{v \in V} h(v)$ **do**
  choose $u \in V - A$ to maximize $f(A \cup \{u\})$
  and set $A \leftarrow A \cup \{u\}$;
output A.

To establish the performance ratio of this greedy algorithm, we will employ a well-known theorem [12] as follows.

Theorem 4. *Suppose f is a monotone increasing, submodular integer function with* $f(\emptyset) = 0$. *Then the above Greedy Algorithm produces an approximation solution within a factor of* $H(\gamma)$ *from optimal, where* $\gamma = \max_{v \in V} f(\{v\})$ *and H is the harmonic function, i.e.,* $H(\gamma) = \sum_{i=1}^{\gamma} \frac{1}{i}$.

To verify our current f satisfying all conditions of this theorem, let us first recall the definitions of monotone increasing and submodular properties of f.

(a) f is monotone increasing if $A \subset B$ implies $f(A) \le f(B)$.
(b) f is submodular if for any two subsets A and B,

$$f(A) + f(B) \ge f(A \cup B) + f(A \cap B).$$

Lemma 5. *f is monotone increasing.*

Proof. Suppose $A \subset B$. Then $n_A(v) \le n_B(v)$ for all $v \in V$. Hence

$$\begin{aligned} f(A) &= \sum_{v \in V} \min(h(v), n_A(v)) \\ &\le \sum_{v \in V} \min(h(v), n_B(v)) \\ &= f(B) \quad \square \end{aligned}$$

Lemma 6. *f is submodular.*

Proof. It has been known [2,3] that f is submodular if and only if for $u \notin B$, $A \subset B$ implies $\Delta_u f(A) \ge \Delta_u f(B)$ where $\Delta_u f(A) = f(A \cup \{u\}) - f(A)$.

Now, we have

$$\begin{aligned} \Delta_u f(A) &= \sum_{v \in V} [\min(h(v), n_{A \cup \{u\}}(v)) - \min(h(v), n_A(v))] \\ \Delta_u f(B) &= \sum_{v \in V} [\min(h(v), n_{B \cup \{u\}}(v)) - \min(h(v), n_B(v))]. \end{aligned}$$

For $u \notin V_B$ and $A \subset B$, we have

$$n_A(v) \le n_B(v) \text{ and } n_{A \cup \{u\}}(v) \le n_{B \cup \{u\}}(v).$$

Case 1. $n_{A\cup\{u\}}(v) \leq h(v)$. In this case,

$$\begin{aligned}
\Delta_u f(A) &= \min(h(v), n_{A\cup\{u\}}(v)) - \min(h(v), n_A(v)) \\
&= n_{A\cup\{u\}}(v) - n_A(v) \\
&= n_{\{u\}}(v) \\
&= n_{B\cup\{u\}}(v) - n_B(v) \\
&\geq \min(h(v), n_{B\cup\{u\}}(v)) - \min(h(v), n_B(v)) \\
&= \Delta_u f(B).
\end{aligned}$$

Case 2. $n_{A\cup\{u\}}(v) > h(v)$. In this case, $n_A(v) \geq h(v)$. Hence, $n_{B\cup\{u\}}(v) > h(v)$ and $n_B(v) \geq h(v)$. It follows that

$$\begin{aligned}
\Delta_u f(A) &= \min(h(v), n_{A\cup\{u\}}(v)) - \min(h(v), n_A(v)) \\
&= 0 \\
&= \min(h(v), n_{B\cup\{u\}}(v)) - \min(h(v), n_B(v)) \\
&= \Delta_u f(B). \quad \square
\end{aligned}$$

Theorem 7. *The Greedy Algorithm for PIDS produces an approximation solution within a factor of $H(\delta)$ from optimal where δ is the maximum vertex degree of input graph.*

Proof. Note that $\gamma = \max_{v\in V} f(\{v\}) = \delta$. Therefore, this theorem follows immediately from Lemmas 3, 5 and 6 and Theorem 4. $\square$

The positive influence dominating set problem can be considered as a special case of the Set Multicover problem in [13] via the following mapping: Let the vertex set V in G as the universe, each vertex in V as an element, for each v, the collection contains the subsets of set S_v which contains all its neighbors. Then define the coverage requirement of each element v to be the ceiling of $deg(v)/2$. The $H(\delta)$ approximation greedy algorithm follows directly. However, The proof in [13] is based on LP approximation and dual fitting techniques which are rather complicated, while we solve the problem in the framework of submodular functions.

4. Conclusion and future work

In this paper, we investigated the positive influence dominating set (PIDS) which has applications in social networks. we proved the complexity of PIDS problem in general graphs by constructing an L-reduction from VC-cubic to PIDS. Furthermore, we propose a greedy approximation algorithm and give theoretical analysis about its approximation ratio. [6] studied the hardness of optimization in power-law graph and found that dominating set is theoretically an easier problem in a power-law graph than in a general graph. We are interested in studying PIDS in a power-law graph since most social networks follow the power-law.

References

[1] Jeremy Blum, Min Ding, Andrew Thaeler, Xiuzhen Cheng, Connected dominating set in sensor networks and MANETs, in: D.-Z. Du, P. Pardalos (Eds.), Handbook of Combinatorial Optimization, 2004, pp. 329–369.
[2] D.-Z. Du, Ker-I Ko, X.-D. Hu, Design and analysis of approximation algorithms, manuscript.
[3] Ding-Zhu Du, R.L. Graham, P.M. Pardalos, P.-J. Wan, W. Wu, W. Zhao, Analysis of greedy approximations with nonsubmodular potential functions, in: Proceedings of the 19th Annual ACM–SIAM Symposium on Dicrete Algorithms, SODA, 2008, pp. 167–175.
[4] D.-Z. Du, Ker-I Ko, Theory of Computational Complexity, John Wiley, 2000.
[5] Eubank Stephen, Kumar V.S. Anil, Madhav V. Marathe, Srinivasan Aravind, Wang Nan, Structural and algorithmic aspects of massive social networks, in: Proceedings of the 15th Annual ACM–SIAM symposium on Discrete Algorithms, SODA, 2004, pp. 718–727.
[6] Alessandro Ferrante, Gopal Pandurangan, Kihong Park, On the hardness of optimization in power-law graphs, Theoretical Computer Science 393 (2008) 220–230.
[7] X. Gao, Y. Wang, X. Li, W. Wu, Analysis on theoretical bonds for approximating dominating set problems, Discrete Mathematics, Algorithms and Applications 1 (1) (2009) 71–84.
[8] D. Li, L. Liu, H. Yang, Minimum connected r-hop k-dominating set in wireless networks, Discrete Mathematics, Algorithms and Applications 1 (1) (2009) 45–57.
[9] Feng Wang, Erika Camacho, Kuai Xu, Positive influence dominating set in online social networks, Lecture Notes in Computer Science 5573 (2009) 313–321.
[10] W. Wang, D. Kim, N. Sohaee, C. Ma, W. Wu, A PTAS for minimum d-hop underwater sink placement problem in 2-D underwater sensor networks, Discrete Mathematics, Algorithms and Applications 1 (2) (2009) 283–289.
[11] W. Wang, D. Kim, J. Willson, B. Thuraisingham, W. Wu, A better approximation for minimum average routing path clustering problem in 2-D underwater sensor networks, Discrete Mathematics, Algorithms and Applications 1 (2) (2009) 175–191.
[12] L.A. Wolsey, An analysis of the greedy algorithm for the submodular set covering problem, Combinatorica 4 (2) (1982) 385–393.
[13] Sridhar Rajagopalan, Vijay V. Vazirani, Primal-dual RNC approximation algorithms for set cover and covering integer program, SIAM Journal on Computing 28 (2) (1999) 525–540.