A Game Theoretic Model for Influence Maximization

Yu Zhang
Dept. of Computer Science
Peking University
Beijing, China
yutazh@pku.edu.cn

Yan Zhang
Dept. of Machine Intelligence
Peking University
Beijing, China
zhy@cis.pku.edu.cn

APPENDIX

A. PROOF OF THEOREM 1

PROOF. Let us consider the seed minimization problem under Majority Vote model: minimize $|S_0|$ subject to $|S_n| = |V|$. This problem is the dual problem of influence maximization, aiming to find the smallest seed set to activate the whole graph. Our proof is divided into two parts. Firstly, we reduce the Vertex-Cover problem [2] to the seed minimization problem. Secondly, we reduce the seed minimization problem to our influence maximization problem.

In the first part, we consider three separate cases.

Case $\delta = \frac{1}{2}$**.** Given a graph $G = (V, E)$ and a constant k, we want to know whether G has a vertex cover of size at most k. We construct the following graph G'.

G' has $|V| + 3|E|$ vertices. $|V|$ vertices $\{u_{11}, u_{12}, ..., u_{1|V|}\}$ in G' correspond to $|V|$ vertices $\{v_1, v_2, ..., v_{|V|}\}$ in G respectively. $|E|$ vertices $\{u_{21}, u_{22}, ..., u_{2|E|}\}$ in G' correspond to $|E|$ edges $\{e_1, e_2, ..., e_{|E|}\}$ in G respectively. u_{1i} and u_{2j} has a link in G' if and only if v_i is an endpoint of e_j in G. There are another $2|E|$ "branching nodes" in G'. Each u_{1i} is linked with $deg(v_i)$ "branching nodes". We draw the brief structure of G' in Figure 7. Obviously, each u_{1i} has $2deg(v_i)$ neighbors in G', and each u_{2j} has 2 neighbors.

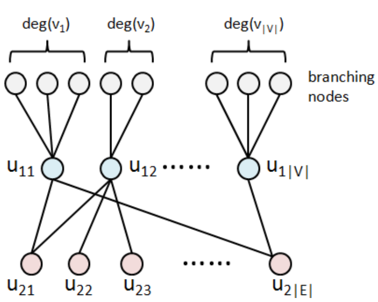

Figure 1: Structure of G' (Case $\delta = \frac{1}{2}$)

We claim that G has a vertex cover of size at most k if and only if G' has a seed set of size at most k to activate the whole graph.

If G has a vertex cover $\{v_{x_1}, v_{x_2}, ..., v_{x_l}\}$ ($l \le k$), we choose $S_0 = \{u_{1x_1}, u_{1x_2}, ..., u_{1x_l}\}$ as our seed set. According to the property of vertex cover, each u_{2j} has at least 1 active neighbor now, so all u_{2j} are activated after step 1. Then each u_{1i} has at least $deg(v_i)$ active neighbors, so all u_{1i} are activated after step 2. And all "branching nodes" will be activated after step 3. Therefore S_0 activates the whole graph.

If G does not have a vertex cover of size k, we assume that G' has a seed set S_0 of size k to get the contradiction.

(1) If there are any "branching nodes" w in S_0, it will be better if we choose a u_{1i} linked with w instead of w itself, because w will become active soon after u_{1i} is active.

(2) If there are any u_{2j} in S_0, it will be better if we choose a u_{1i} linked with u_{2j} instead of u_{2j} itself, because u_{2j} will become active soon after u_{1i} is active.

According to (1) and (2), we can assume $S_0 = \{u_{1x_1}, u_{1x_2}, ..., u_{1x_l}\}$, $l \le k$. Since $\{v_{x_1}, v_{x_2}, ..., v_{x_l}\}$ is not a vertex cover of G, there will be at least one u_{2j} remaining inactive after step 1. We suppose that u_{2j} is linked with u_{1s} and u_{1t} in G'. Obviously u_{1s} and u_{1t} are not in S_0, so they are inactive after step 1. Now we get the following "deadlock": If u_{2j} want to be activated, at least one of u_{1s} and u_{1t} should become active first. If u_{1s} or u_{1t} want to be activated, u_{2j} should become active first. So all of these 3 nodes will never become active. We get the contradiction and prove our claim.

Case $\delta \in (0, \frac{1}{2})$**.** We slightly modify the graph G' in Figure 7. Let each u_{1i} link with K_i "branching nodes" and each u_{2j} link with K_0 "branching nodes", where

$$\begin{aligned} K_0 &= \min\{K \in \mathbb{N} | 0 < \delta \le \frac{1}{K+2}\}, \\ K_i &= \min\{K \in \mathbb{N} | \frac{deg(u_i) - 1}{deg(u_i) + K} < \delta \le \frac{deg(u_i)}{deg(u_i) + K}\}. \end{aligned} \tag{1}$$

Note that the constraint of K_i is equivalent to

$$\frac{deg(u_i) - 1}{\delta} - deg(u_i) < K_i \le \frac{deg(u_i)}{\delta} - deg(u_i), \tag{2}$$

therefore the gap between the upper bound and lower bound is $\frac{1}{\delta} > 1$. So there is at least one integer satisfying the constraint.

Similar to the Case $\delta = \frac{1}{2}$, we can prove that G has a vertex cover of size at most k if and only if G' has a seed set of size at most k to activate the whole graph.

Case $\delta \in (\frac{1}{2}, 1)$**.** We slightly modify the graph G' in Figure 7. Let each u_{1i} link with K_i "branching nodes" and

WWW '17 April 3–7, 2017, Perth, Australia

DOI: 10.475/123_4

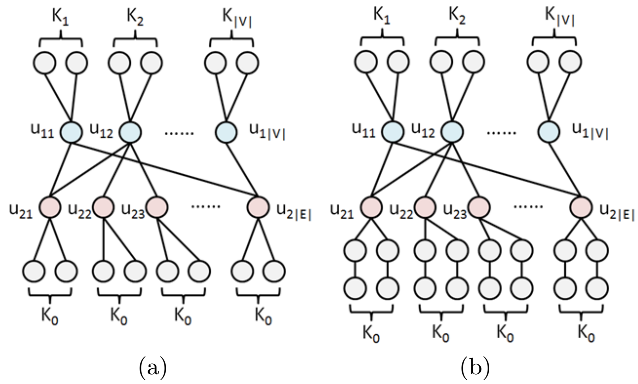

Figure 2: Structure of G'. (a) Case $\delta \in (0, \frac{1}{2})$. (b) Case $\delta \in (\frac{1}{2}, 1)$.

each u_{2j} link with K_0 "2-node bands", where

$$K_0 = \min\{K \in \mathbb{N} | \frac{K}{K+2} < \delta \leq \frac{K+1}{K+2}\},$$
$$K_i = \min\{K \in \mathbb{N} | \frac{deg(u_i) - 1}{deg(u_i) + K} < \delta \leq \frac{deg(u_i)}{deg(u_i) + K}\}. \quad (3)$$

Note that the constraint of K_0 is equivalent to

$$\frac{2\delta - 1}{1 - \delta} < K_i \leq \frac{2\delta}{1 - \delta}, \quad (4)$$

therefore the gap between the upper bound and lower bound is $\frac{1}{1-\delta} > 1$. So there is at least one integer satisfying the constraint. The existence of K_i is the same with Case $\delta \in (0, \frac{1}{2})$.

Similar to the Case $\delta = \frac{1}{2}$, we can prove that G has a vertex cover of size at most k if and only if G' has a seed set of size at most $k + K_0|E|$ to activate the whole graph (since each "2-node band" needs an active seed node).

We draw the brief structures of G' for Case $\delta \in (0, \frac{1}{2})$ and Case $\delta \in (\frac{1}{2}, 1)$ in Figure 8.

It is easy to see that the construction of G' can be done in polynomial time. So we finish our reduction in the first part. Note that G' in all three cases is a bipartite graph.

The second part is rather easy. We want to minimize $|S_0|$ subject to $|S_n| = |V|$. For $i = 1, 2, ..., |V|$, we solve the influence maximization problem $\max_{|S_0| \leq i} |S_n|$. And the answer to the seed minimization problem is the smallest i which makes $\max_{|S_0| \leq i} |S_n| = |V|$. □

B. PROOF OF THEOREM 3

Proof. Our proof is similar to the one in [1] except some modifications. In this proof, we assume that the graph is directed. We reduce this problem from the problem of counting simple paths in a directed graph. Given a directed graph $G = (V, E)$, counting the total number of simple paths in G is #P-hard [3]. Let $n = |V|$ and $D = \max_{v \in V} deg_{in}(v)$. From G, we construct $n+1$ graphs $G_1, G_2, ..., G_{n+1}$. To get G_i $(1 \leq i \leq n+1)$, we first add $D + i - deg_{in}(v)$ "branching nodes" linking to node v for all $v \in V$. And then we add a node s linking to all nodes in V. Thus each node in G_i has $D + i + 1$ in-links except "branching nodes" and s.

According to our assumption, the weight on each edge in G_i is $w_i = \frac{1}{D+i+1}$. Let $S_0 = \{s\}$ and $\mathcal{P}$ denote the set of all simple paths starting from s in G_i. (Note that $\mathcal{P}$ is identical in all G_i because "branching nodes" are unreachable from s.) According to [1], we have

$$\sigma_{G_i}(S_0) = \sum_{\pi \in \mathcal{P}} \prod_{e \in \pi} w_i, \quad (1 \leq i \leq n+1), \quad (5)$$

where $\sigma_{G_i}(S_0)$ means $\sigma(S_0)$ in G_i. Let B_j be the set of simple paths of length j in $\mathcal{P}$ $(0 \leq j \leq n)$. We have

$$\sigma_{G_i}(S_0) = \sum_{j=0}^{n} \sum_{\pi \in B_j} \prod_{e \in \pi} w_i = \sum_{j=0}^{n} \sum_{\pi \in B_j} w_i^j = \sum_{j=0}^{n} w_i^j |B_j|. \quad (6)$$

We want to solve these $n+1$ linear equations with $n+1$ variables $|B_0|, |B_1|, ..., |B_n|$. Since the coefficient matrix is a Vandermonde matrix, $(|B_0|, |B_1|, ..., |B_n|)$ is unique and easy to compute.

Finally, we notice that for each $j = 1, 2, ..., n$, there is a one-to-one correspondence between paths in B_j and simple paths of length $j - 1$ in G. Therefore, $\sum_{j=1}^{n} |B_j|$ is the total number of simple paths in G. We complete our reduction. □

C. REFERENCES

[1] W. Chen, Y. Yuan, and L. Zhang. Scalable influence maximization in social networks under the linear threshold model. In *ICDM'10*, pages 88–97. IEEE, 2010.

[2] R. M. Karp. Reducibility among combinatorial problems. In *Proceedings of a symposium on the Complexity of Computer Computations*, pages 85–103. Plenum Press, New York, 1972.

[3] L. G. Valiant. The complexity of enumeration and reliability problems. *SIAM Journal on Computing*, 8:410–421, 1979.