

A Balanced Method for Budgeted Influence Maximization

Xinhui Xu*, Yong Zhang[†], Qingcheng Hu*, Chao Li[†], Chunxiao Xing[†]

Department of Compute Science and Technology

Tsinghua National Laboratory for Information Science and Technology

Research Institute of Information Technology, Tsinghua University, Beijing, China

* {xuxh13, hqc10}@mails.tsinghua.edu.cn

[†] {zhangyong05, li-chao, xingcx}@mail.tsinghua.edu.cn

Abstract—With the flourish of Web-based large Online Social Networks (OSNs), people on OSNs can easily yield influence on others. Finding how the influence spreads and maximizing influence spread within OSNs have been extensively studied. State-of-the-art researches suffer two defects: (a) need to acquire the topological structure of the network, which is impractical for the continuously changing networks in real life and thus can not balance very well between influence spread and running time; (b) assign the same cost for every node in OSN which cannot reflect the reality. To solve these problems we firstly propose PageRank Based Cost (PRBC) model to assess the cost of nodes in OSN according to their importance (influence); secondly we present Budgeted Random Maximal Degree Neighbor (BRMDN) algorithm by exploiting the scale free property. Results from extensive experiments show that BRMDN can well balance influence spread and running time.

Keywords-social network; cost model; influence maximization;

I. INTRODUCTION

A. Background and Motivations

The Web today is a growing universe of interlinked Web pages and Web applications, teeming with videos, photos, and interactive content. The Web-based large online social network(OSN)s such as Facebook, Twitter, WeChat etc. acquire great success for their interactive features like sharing, forwarding and discussing contents. According to eBizMBA, in February 2015, Facebook has estimated unique monthly visitors numbered 900 million¹. People spend a lot of time on these communication platforms making friends, sharing daily affairs, spreading interesting news, and expressing different opinions, which provides us with affluent real-life data to mine valuable information. Taking advantage of the popularity of OSNs, many researchers have studied diffusion phenomenon in OSNs, which includes the diffusion of news, ideas, innovations, and the adoption of new products [1]. These diffusion phenomenons are referred as influence diffusion or propagation in [1]. Influence maximization is an extensively investigated topic in influence diffusion [1] [2] [3] [4]. It tries to find a set of nodes in one OSN to maximal the influence spread over the OSN under certain diffusion model such as Independent Cascade (IC) proposed in [2].

However, the aforementioned research works suffer two defects: (a) need to acquire the topological structure of the network, which is impractical for the continuously changing networks in real life and thus can not balance very well between influence spread and running time; (b) assign the same cost for every nodes in OSN which cannot reflect the real situation in life. In reality, time can be a critical factor in some situations such as disease controlling, emergency evacuation. Furthermore, different nodes in one OSN should not be assigned the same cost. For example, in the domain of online advertisement, different service providers have different advertising prices [5].

Taking account of the above situations, also with the fact that most of the large OSNs are complex networks and have the scale free property [6], we firstly define a cost function to assess the cost of a given node in one OSN, then research budgeted influence maximization problem and propose our algorithm.

B. Our Contributions

In this paper, we propose a cost function to assess every node in terms of their influence and dedicate to solving the problem of budgeted influence maximization. Our contributions in this paper are summarized as follows.

- We propose PRBC model to assess nodes' costs according to their importances (influences) which are assessed by the nodes' PageRank value [7] and degrees in OSN.
- We exploit the scale-free property [6] [8] that most OSNs hold, and then propose BRMDN under PRBC model.
- Finally, we test the performance of BRMDN under two real datasets with extensive experiments, which proves the effectiveness and efficiency of the proposed algorithm.

The rest of this paper is organized as follows. Section II introduces the related work. Section III provides preliminaries for budgeted influence maximization. Section IV presents our algorithms and the theoretical analysis. Section V shows our experimental results. Finally we conclude the paper in section VI.

¹<http://www.ebizmba.com/articles/social-networking-websites>

II. RELATED WORK

A. Diffusion Models

In influence maximization process, we label every node with one status—activated or inactivated. We tag a node *activated* if it accepts the message or event we concern, *inactivated* otherwise. In a widely connected network, people may influence each other by publishing, sharing, re-directing messages or news etc. Therefore, influence spreading is to a large extent similar to information diffusion process. One pioneer information diffusion model is IC [2].

IC is firstly proposed by D Kempe et al. in [2], and now becomes the most important model in influence maximization problem. Given OSN $G = (V, E)$, an initial seed set S , let $p(u, v)$ denote the probability of u influencing v and $p(u, v)$ is independent assigned for $\forall u, v \in V$. IC runs as follows: Let S_t denote the nodes activated in step (time) t , t_e represents the step when the activation process ends, initially we have $t = 0, S_t = S$. At step $t + 1$, every node $u \in S_t$ tries to activate its out-neighbors $v \in V \setminus \bigcup_{0 \leq i \leq t} S_i$ with an independent probability of $p(u, v)$. This procedure proceeds until no more nodes can be activated. It should be mentioned that each node can activate its out-neighbors one time and when a node is activated, it never fails. The final activated set can be calculated by $\bigcup_{0 \leq i \leq t_e} S_i$ which we denote as $\sigma_{IC}(S)$.

B. Influence Maximization

1) *Greedy Algorithms*: In [2], D Kempe et al. proposed a general greedy algorithm for influence maximization and proved that general greedy algorithm can approximate to the optimal solution by a factor of $1 - 1/e$, but it is extremely expensive to compute. CELF [9] improved general greedy algorithm by exploiting the property of submodular function. Results show that CELF achieves 700 times faster than general greedy algorithm. NewGreedy and MixGreedy in [10] presented two variants of greedy algorithm which aim at improving effectiveness and efficiency, yet still suffer high computation cost.

2) *Random Based Algorithm*: Random algorithm always contains a randomization procedure within itself, in the field of influence maximization, the most obvious and straightforward random based algorithm is referred as Random in [2] [10]. It runs as follows, it iterates for k times, every time it randomly selects a node which has not been added to the seed set. As we can see from Table I, Random runs fastest, but we will show later it performs the worst in terms of influence spread.

3) *Degree Based Heuristic Algorithms*: In OSNs, one node that has a larger number of in-neighbors most probably means that it is more important. It is referred as degree centrality [11]. DegreeHeuristic algorithm sorts all the nodes according to their degree and then chooses the Top-K nodes. In [10] SingleDiscount is a simple degree discount heuristic

Table I
TIME COMPLEXITY OF ALGORITHMS

Algorithms	Complexity
Random	$O(K)$
Degree Heuristic	$O(m)$
Degree Discount	$O(K * \log(n) + m)$
Single Discount	$O(K * \log(n) + m)$
New Greedy IC	$O(KRm)$
CELF Greedy	$O(KnRm/700+)$
General Greedy	$O(KnRm)$

where each neighbor of a newly selected seed discounts its degree by one, while DegreeDiscount is a more accurate degree discount heuristic algorithm. It excludes nodes that can possibly be influenced by nodes which have already been added into the seed set. Experiments show that SingleDiscount and DegreeDiscount have almost the same result in influence spreading as greedy algorithms, while achieving significant speedup in running time.

The time complexities of some aforementioned algorithms are shown in Table I, where K denotes # of nodes in the initial seed set, n represents # of vertices, m depicts # of edges in the given graph and R is # of rounds in the specific algorithm. These notations used in the following sections will have the same meaning if there is no explicit declaration.

C. Cost Models

Most state-of-the-art influence maximization algorithms such as [1] [2] [3] [4] [10] etc. take the constant cost (i.e. unit cost) model. That is, $\mathcal{CF}(s) = 1$ for all node $s \in V$, $\mathcal{CF}(\cdot)$ denotes cost function. In [9], J. Leskovec et al used # of posts for their case 1, they assigned a non-negative cost for case 2; [5] does not give their detailed cost model; while [12] randomly gives a value to nodes in OSN. To our best of knowledge, few non-constant cost models have been used for budgeted influence maximization.

III. PRELIMINARIES

A. Scale-free Networks

In the real world, there are numerous networks existing in form of complex network [6] such as biosphere, citation network, OSNs and so on. Many large OSNs (such as Facebook, Twitter, MySpace, Flickr [13]) share the property of scale-free [6]. A node in a network with degree k subjects to power-law distribution has probability $p(k) = ck^{-\gamma}$ [14]. When the power-law distribution exponent γ values are between 2 and 3, the network holds the property of scale-free.

B. PageRank Algorithm & Proposed Cost Model

PageRank algorithm has been widely taken as a method for measuring the importance of web pages which was firstly proposed in [7]. With the web pages modeled as nodes, and hyperlinks between them represented as edges, the interlinked Web can be seen as a complicated graph. As

we have described in previous section, OSN is represented by $G = (V, E)$, so we can apply the PageRank algorithm to the OSN to estimate the importance of nodes. According to nodes' differences in importance we give them different cost. The cost model is shown as follows.

Definition 1. (*PageRank Based Cost*) Given an OSN $G = (V, E)$, a pre-defined increase factor δ and coefficient λ , for $\forall u \in V$, we define its PageRank Based Cost, $PRBC(u)$, as follows,

$$PRBC(u) = \frac{\lambda(PR(u) + \delta)\mathcal{D}(u)}{\mathcal{D}(v_{max})}, v_{max} = \operatorname{argmax}_{v \in \mathcal{N}(u) \cup \{u\}} \mathcal{D}(v)$$

where $PR(u)$ is the PageRank value of node u , $\mathcal{D}(u)$ is the degree of node u , $\mathcal{N}(u)$ is the neighbor set of node u .

C. Non-Constant Cost Influence Maximization

In reality, the nodes in OSN have different influences should be assigned different costs. Compared with normal unit cost influence maximization problem, Non-Constant Cost Influence Maximization (NCC-IM) has tighter constraints. We formally formulate it as follows,

Definition 2. (*NCC-IM*) Given OSN $G = (V, E)$, a constant K , a cost function $\mathcal{CF}(\cdot)$, and a budget limit \mathcal{B} , finding a set S with K nodes which subjects to the following constraints:

$$\sigma(S) = \operatorname{argmax}_{|S| \leq K \wedge S \subseteq V} \operatorname{Inf}(S), \sum_{u \in S} \mathcal{CF}(u) < \mathcal{B}$$

where $\sigma(S)$ denotes the expected nodes set finally activated by giving an initial set S under influence spread function Inf .

Noting that in NCC-IM problem, we have constraints that $|S| \leq K$, because that in our algorithm and other algorithms we implemented as baselines, we find that when $\sum_{u \in S} \mathcal{CF}(u) > \mathcal{B}$, all the algorithms should stop.

IV. BUDGETED INFLUENCE MAXIMIZATION ALGORITHM

Finding exact maximal influence spread is a NP-Hard problem and the improved greedy algorithm variants are expensive to compute. Random algorithms randomly choosing K nodes also perform arbitrarily bad. [8] [15] [16] find that in many OSNs vertices' connectivities follow scale-free property that most vertices are sparsely connected. However, a small number of vertices are densely connected. We design our own algorithm by combining randomly heuristic method with the properties of scale-free networks.

A. Algorithm Design

Enlightened by the random algorithm and the property of scale-free network, we propose Budgeted Random Maximal Degree Neighbor (BRMDN) algorithm. Firstly, BRMDN randomly selects a node u in the network; then it uses the algorithm MDN (Algorithm 1) to find the candidate node

which will be added to initial set S if it satisfies the budget constraints. we repeatedly run the above procedure to get the qualified S . We give BRMDN as Algorithm 2.

Algorithm 1 MDN: Find the candidate node

Input: $G = (V, E)$; node u ; exclusive set \mathcal{ES} ;

Output: Node v_{max} selected to be added to the seed set

```

1:  $v_{max} \leftarrow u$ ;  $v_{degree} \leftarrow \mathcal{D}(u)$ ;
2: for all  $nbr \in \mathcal{N}(u)$  do
3:   if  $nbr \notin \mathcal{ES} \wedge v_{degree} < \mathcal{D}(nbr)$  then
4:      $v_{degree} \leftarrow \mathcal{D}(nbr)$ ;
5:      $v_{max} \leftarrow nbr$ ;
6:   end if
7: end for
8: return  $v_{max}$ 

```

Algorithm 2 BRMDN: Compute the seed set

Input: $G = (V, E)$; seed set size K ; budget limit \mathcal{B} ; fault tolerance factor τ ; cost function \mathcal{CF} ;

Output: Seed set S with $|S| \leq K \wedge \sum_{s \in S} \mathcal{CF}(s) < \mathcal{B}$

```

1:  $S \leftarrow \Phi$ ;  $i \leftarrow 0$ ;  $totalcost \leftarrow 0$ 
2: while  $i < K$  do
3:   Randomly select a node  $u \in V \setminus S$ 
4:   Select  $u_{max} \leftarrow MDN(G, u, S)$ 
5:   if  $\mathcal{CF}(u_{max}) + totalcost < \mathcal{B} + \tau$  then
6:      $S \leftarrow S \cup \{u_{max}\}$ 
7:      $totalcost \leftarrow totalcost + \mathcal{CF}(u_{max})$ 
8:      $i \leftarrow i + 1$ 
9:   if  $totalcost > \mathcal{B}$  then
10:     $i \leftarrow K + 1$ 
11:   end if
12: end if
13: end while
14: return  $S$ 

```

Noting that through MDN procedure in BRMDN, we only need to know the local topological structure around node u , which can greatly boost computing efficiency.

B. Feasibility of BRMDN

For network $G = (V, E)$ which subjects to the power-law distribution, a node with degree k has probability $p(k) = ck^{-\gamma}$. Let k_{max} be # of the maximal degree and k_{min} be the minimal one, so we have:

$$\int_{k_{min}}^{+\infty} p(k)dk = \frac{1}{n}, \int_{k_{min}}^{+\infty} p(k)dk = 1 \quad (1)$$

By solving (1), we get, $k_{max} = k_{min}n^{\frac{1}{\gamma-1}}$. For arbitrary edge which starts from node u , let p_{Top-K} denote the probability of connecting a node with degree great than or equal to Top-K (such node is also known as one Hub of the

network). We can get

$$p_{Top-K} = \int_{k_{Top-K}}^{k_{max}} p(m)dm = \frac{k_{max}^{2-\gamma} - k_{Top-K}^{2-\gamma}}{k_{max}^{2-\gamma} - k_{min}^{2-\gamma}} \quad (2)$$

If the seed set size is K , the probability to get at least one hub node is $p_{hub} = 1 - (1 - p_{Top-K})^K - \epsilon$. Where ϵ denotes the budget \mathcal{B} affects the p_{hub} , actually if $\mathcal{B} > K\zeta$ then $\epsilon = 0$, where $\zeta = \max_{u \in V} \{\mathcal{C}\mathcal{F}(u)\}$ is the highest cost in the OSN. If K is bigger enough (say $K = 30$) then we can have $(1 - p_{Top-K})^K \rightarrow 0$, and finally $p_{hub} \approx 1 - \epsilon$. Here ϵ is a random factor which is determined by random process in BRMDN. But we can alleviate the random disturbance by iterating BRMDN many times to decrease constraint ϵ to a small number and so p_{hub} is close to 1. A high value of p_{hub} indicates that BRMDN with a large possibility can have at least one hub node in initial set S .

C. Time Complexity

According to Algorithm 2, after a node is randomly chosen, we have to traverse its all neighbors. Let \bar{k} denote the average degree in a scale-free network, we have:

$$\bar{k} = \sum_1^n kp(k) = \sum_1^n kck^{-\gamma} = c \sum_1^n \frac{1}{k^{\gamma-1}}, p(k) = ck^{-\gamma} \quad (3)$$

Lemma 1. Let $G = (V, E)$ be a network that holds scale-free property, if there are no self-loops or multiple links between two nodes in G , there does not exist a G with $1 < \gamma < 2$.

Proof: From section IV-B, we already have $k_{max} = k_{min}n^{\frac{1}{\gamma-1}}$. Now assume that there exists a network that subjects to power-law distribution and has $1 < \gamma < 2$, then we can get $0 < \gamma - 1 < 1$, $n^{\frac{1}{\gamma-1}} > n$, and $k_{max} = k_{min}n^{\frac{1}{\gamma-1}} > n$, which means that a node with the largest degree is even bigger than the number of the total nodes in network G . It contradicts the fact that n is the total # of nodes of network G . ■

According to lemma 1, $\gamma > 2$. Then

$$\bar{k} = c \sum_1^n \frac{1}{k^{\gamma-1}} \leq c \sum_1^n \frac{1}{k} = c \ln(n), n \rightarrow +\infty \quad (4)$$

If the size of the initial seed set is K , the time complexity of Algorithm 2 can be computed by $K\bar{k}$. By substituting \bar{k} with equation (4), we get $O(K \log(n))$.

V. EXPERIMENTS

Considering different networks with different topological structures, we choose two datasets which subject to power-law distribution and have different γ values (Table II). We will show that BRMDN performs very well in terms of both influence spread and running time.

Table II
STATISTICS OF TWO REAL NETWORKS

Dataset	n	m	\bar{k}	k_{max}	γ
Blogs	3982	6803	3.42	189	2.453
Facebook	4039	88234	43.69	1045	2.509

A. Experimental Setup

The two real scale-free networks listed in Table II are summarized as follows:

- **Blogs [17].** It contains about 4K nodes and 6K edges. Obviously this network is sparsely connected, and $\frac{\#edges}{\#nodes} = 1.70$.
- **Facebook [18].** It is just a small part of users of Facebook. This network has 4K nodes but with 88K edges, which is different from Blogs greatly in its high density of connection, and $\frac{\#edges}{\#nodes} = 21.84$.

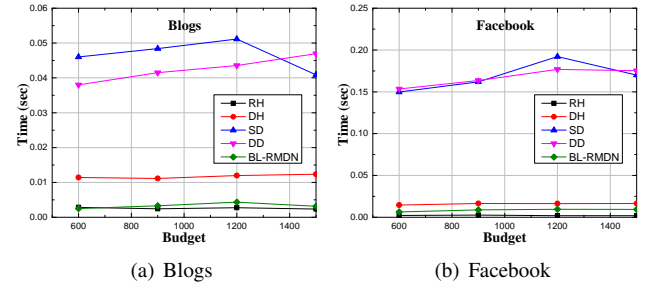


Figure 1. Running time under datasets Blogs and Facebook with $K = 30$

We apply IC model to BRMDN and compare the results with some state-of-the-art heuristic based influence maximization algorithms. We list those algorithms as follows,

RandomHeuristic(RH): It is a baseline for heuristic approaches. It simply selects K random vertices in the graph, which is also evaluated in [2] and [10].

DegreeHeuristic(DH): It intuitively selects K vertices which have the largest degrees in the given graph.

SingleDiscount(SD): A simple degree discount heuristic where each neighbor of a newly selected seed decreases its degree by one, proposed in [10].

DegreeDiscount(DD): Compared to SingleDiscount, it is a refined heuristic method [10].

The greedy algorithms perform very well in terms of influence spread but are intolerable slow for large networks. For example in our experiments, the general greedy algorithm spends 20.78 hours running dataset Blogs with $R = 5$ (normally we have $R = 1000$), while one degree based heuristic algorithm spends about 0.05 seconds, which almost is 1,500,000 times faster. Therefore we do not compare with them in this paper. We gain all the results on a server computer with 24 cores of Intel(R) Xeon(R) CPU E5-2640 2.50GHz and 128G Memory.

For the IC model with relatively large propagation probability p , the influence spread is not very sensitive to different

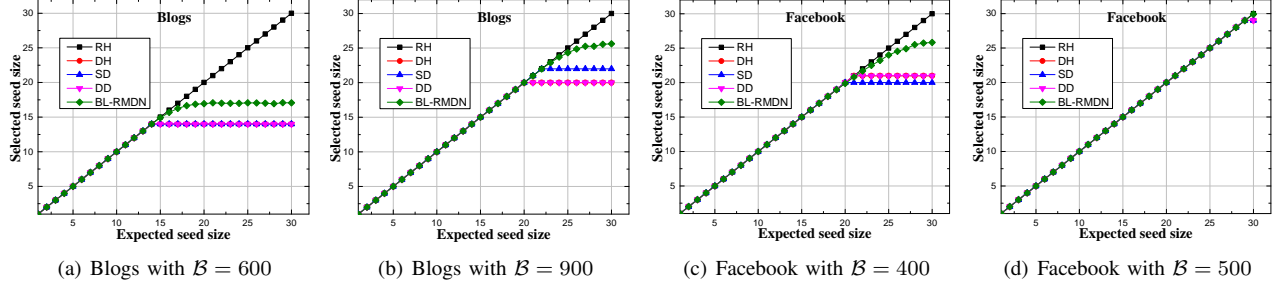


Figure 2. Performance under IC model for dataset Blogs and Facebook with different budget B

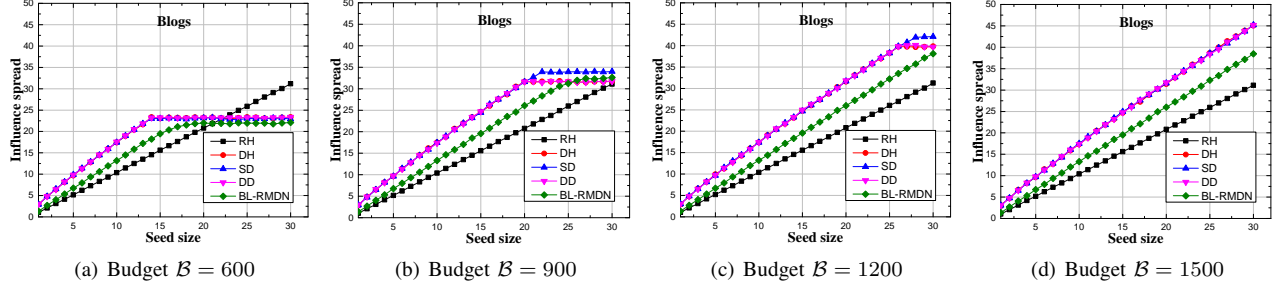


Figure 3. Performance under IC model for dataset Blogs with different budget B

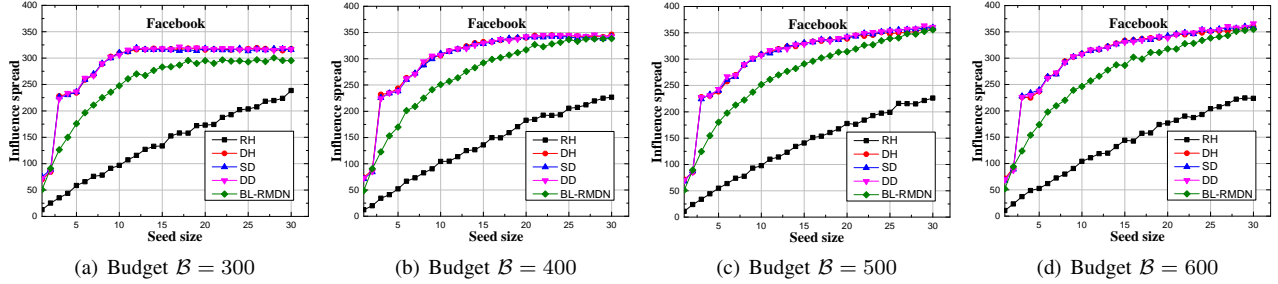


Figure 4. Performance under IC model for dataset Facebook with different budget B

algorithms. When simulating propagation process with IC model, we set transmitting probability to a relative small number 0.01. To correctly evaluate our method with a reasonable precision, we set our iteration number $R = 1000$. We use PageRank Based Cost Model to evaluate the cost of every node in all the aforementioned algorithms, and we set $\lambda = 100, \delta = 0.5$.

In scale-free networks, the most nodes are sparsely connected, while the hub nodes are densely connected. In the logarithmic coordinate system, a network having scale-free property is a straight-line with gradient valued between -2 and -3. We use the method proposed in [19] [20] to estimate γ , the result is listed in Table II.

B. Experimental Results

The results show that under IC model, BRMDN performs almost as good as DegreeDiscount according their influence spread, while achieves great speedup with respect to running time.

Influence spread: Considering influence spread in Figure 3 and Figure 4, we can see that DegreeDiscount, DegreeHeuristic, SingleDiscount achieve the best; RandomHeuristic, without any surprise, performs the worst; BRMDN approximately approaches to the best algorithms by different ratios with respect to different datasets. In Figure 3, with the increasing of seed set size K , BRMDN gets the result more and more close to DegreeDiscount. Let ∇_{fig} denotes the ratio between the influence spread of BRMDN and DD in Figure fig , then from Figure 3 we have $\nabla_{3(a)} = 95\%$, $\nabla_{3(b)} = 96\%$, $\nabla_{3(c)} = 91\%$, $\nabla_{3(d)} = 85\%$. While in Figure 4, we can see that when the seed set size $K < 15$, BRMDN gets a result approximate to DegreeDiscount with a ratio less than 90%. But when K grows near to 30, BRMDN achieves almost as good as DegreeDiscount. Finally we have $\nabla_{4(a)} = 93\%$, $\nabla_{4(b)} = 99\%$, $\nabla_{4(c)} = 99\%$, $\nabla_{4(d)} = 97\%$. From Figure 3 and Figure 4, BRMDN performs very well in terms of influence spread.

Running time: From Figure 1, we can see that for every

algorithm, the running time is almost a horizontal line which indicates that budget \mathcal{B} has little affect on running time. From Figure 1, we can see that RH runs fastest, second comes BRMDN, DH is the third, SD and DD are the slowest. Specifically, from Figure 1(a) BRMDN is almost 14 times faster than DD, and from Figure 1(b) BRMDN is almost 19 times faster than DD. For datasets Blogs and Facebook, they have relative small number of nodes, so DH's running time is approximate to BRMDN. But from Table I, we can see that when OSN has a large number of nodes (n), BRMDN will show its great speedup.

Noting that in Figure 3(a), when $K > 23$, RH outperforms other algorithms. From Figure 2(a), we find that when budget \mathcal{B} is very limited, other algorithms find a small number of nodes with high cost, while RH can find a large number of nodes with low costs, and finally RH can achieve a better influence spread. From Figure 4, when budget $\mathcal{B} > 400$, all the algorithms can not improve their performance when K is limited to 30. From Figure 2(d), we can see that it is because the nodes selected by all the algorithms are close to 30, and most influential nodes have been already included.

VI. CONCLUSION

Influence maximization is important for activities like products promotion, information transmission, emergency evacuation etc.. It has been extensively studied to find K nodes within the given budget in OSN to achieve maximal spread under constant-cost (i.e. unit-cost) model. In this paper, we firstly propose PRBC to assess nodes' cost according their importance (influence); secondly research budgeted influence maximization under PRBC and propose BRMDN algorithm; finally run extensive experiments to test the performance of BRMDN. The experimental results show: (1) budget \mathcal{B} has little affects on running time when a specific K is given; (2) BRMDN has almost the same influence spread with SD and DD, but achieves great speedup. Therefore, BRMDN balances well between running time and influence spread.

VII. ACKNOWLEDGMENT

This work was supported by National Basic Research Program of China (973 Program) No.2011CB302302, the National High-tech R&D Program of China under Grant No. SS2015AA020102, Tsinghua University Initiative Scientific Research Program.

REFERENCES

- [1] 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.
- [2] 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.
- [3] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, and Y. Yuan, "Influence maximization in social networks when negative opinions may emerge and propagate." in *SDM*, vol. 11. SIAM, 2011, pp. 379–390.
- [4] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 1029–1038.
- [5] S. Han, F. Zhuang, Q. He, and Z. Shi, "Balanced seed selection for budgeted influence maximization in social networks," in *Advances in Knowledge Discovery and Data Mining*. Springer, 2014, pp. 65–77.
- [6] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics Reports-review Section of Physics Letters*, vol. 424, pp. 175–308, 2006.
- [7] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." 1999.
- [8] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [9] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007, pp. 420–429.
- [10] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 199–208.
- [11] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *Journal of Mathematical Sociology*, vol. 2, no. 1, pp. 113–120, 1972.
- [12] H. Nguyen and R. Zheng, "On budgeted influence maximization in social networks," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 6, pp. 1084–1094, 2013.
- [13] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 2007, pp. 29–42.
- [14] R. Cohen and S. Havlin, "Scale-free networks are ultrasmall," *Physical review letters*, vol. 90, no. 5, p. 058701, 2003.
- [15] L. A. Adamic and B. A. Huberman, "Power-law distribution of the world wide web," *Science*, vol. 287, no. 5461, pp. 2115–2115, 2000.
- [16] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [17] Q. Hu, Y. Gao, P. Ma, Y. Yin, Y. Zhang, and C. Xing, "A new approach to identify influential spreaders in complex networks," in *Web-Age Information Management*. Springer, 2013, pp. 99–104.
- [18] J. Leskovec and J. J. McAuley, "Learning to discover social circles in ego networks," in *Advances in neural information processing systems*, 2012, pp. 539–547.
- [19] A.-L. Barabási, R. Albert, and H. Jeong, "Mean-field theory for scale-free random networks," *Physica A: Statistical Mechanics and its Applications*, vol. 272, no. 1, pp. 173–187, 1999.
- [20] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.