Diversified Social Influence Maximization

Fangshuang Tang, Qi Liu, Hengshu Zhu, Enhong Chen
School of Computer Science and Technology
University of Science and Technology of China
Hefei, P. R. China, 230027
{fstang,zhs}@mail.ustc.edu.cn, {qiliuql,cheneh}@ustc.edu.cn

Feida Zhu
School of Information Systems
Singapore Management University
Singapore, 178902
fdzhu@smu.edu.sg

Abstract—**For better viral marketing, there has been a lot of research on social influence maximization. However, the problem that who is influenced and how diverse the influenced population is, which is important in real-world marketing, has largely been neglected. To that end, in this paper, we propose to consider the magnitude of influence and the diversity of the influenced crowd simultaneously. Specifically, we formulate it as an optimization problem, i.e., *diversified social influence maximization*. First, we present a general framework for this problem, under which we construct a class of diversity measures to quantify the diversity of the influenced crowd. Meanwhile, we prove that a simple greedy algorithm guarantees to provide a near-optimal solution to the optimization problem. Furthermore, we relax the problem by focusing on the diversity of the nodes targeted for initial activation, and show how this relaxed form could be used to diversify the results of many heuristics, e.g., PageRank. Finally, we run extensive experiments on two real-world datasets, showing that our formulation is effective in generating diverse results.**

Keywords—*influence maximization; diversity; viral marketing*

Introduction

Viral marketing, as an important application, has attracted much attention from both industry and academia. Kempe *et al.* first formulate this as the influence maximization problem [1]. Namely, given a social graph $G = (V, E)$, where V is the set of nodes, and E is the set of edges, find a node set S ($S \subseteq V, |S| = K$) that could result in the maximum *influence* under certain influence models, such as Independent Cascade (IC) model and Linear Threshold (LT) model. S is the set of nodes targeted for initial activation, called seed set.

While large amounts of research have been devoted to the development of effective and efficient algorithms to maximize influence [2]–[4], there is little literature focused on how to measure the diversity of the influenced crowd. In real-world marketing, having a diverse target audience could bring many benefits. An obvious one is that this could reduce the risk of marketing campaigns, as the proverb goes: "Don't put all your eggs in one basket". Besides marketing, diversity in influence maximization could also benefit the recommendation systems based on social influence theories [5], since diversity has been found to be an important factor in recommender systems [6].

Seemingly, diversity is intrinsic to the traditional social influence maximization problem, since the optimal seed nodes are encouraged to lie far apart in the graph to minimize influence overlapping. However, this represents only structural characteristics of the graph but lacks semantic meanings. In this sense, It is not real diversity.

In this paper, we propose a general framework for the diversified social influence maximization problem, which is applicable to both IC model and LT model. To measure diversity, we assume that each node in the graph belongs to one or more semantic categories, and can be described by a distribution over these categories. Under this assumption, and inspired by the recent research on text summarization [7], we construct a class of diversity measures. We prove that our diversity measures and the objective function have the nice properties of submodularity and monotonicity, making it possible for a simple greedy algorithm to achieve a near-optimal solution.

Since people tend to associate with people similar to themselves [8], we further relax the problem by emphasizing the diversity of the seed set instead of the diversity of the influenced crowd. Intuitively, if the seed set is diverse, we could expect the activated node set is diverse, too. With this relaxation, we could easily extend our framework to several existing heuristics, such as PageRank [9] and degree centrality. To summarize, our contributions are as follows.

- We introduce diversity, for the first time, into the problem of influence maximization, with a general framework capable of accommodating different diversity measures and influence models.
- We construct a class of diversity measures to quantify the diversity of the influenced crowd, and prove that the diversity measures, as well as the objective functions based on them, are nondecreasing submodular, showing that a simple greedy algorithm could achieve a near-optimal solution.
- We provide a relaxation that only emphasizes the diversity of the seed set, and use it to incorporate diversity information into some heuristics, e.g., PageRank.

Related Work

Social Influence Maximization

Domingos and Richardson are the first to investigate the problem of marketing from an algorithmic perspective [10]. Kempe *et al.* first propose the influence maximization problem and formulate it as an optimization problem [1]. Namely, given a graph $G = (V, E)$,

$$\max_{S \subseteq V} \quad \sigma(S), \quad \text{s.t.} \quad |S| = K, \quad (1)$$

where $\sigma(S)$ is the *influence* of seed set S, i.e., the expected number of influenced nodes under certain influence propagation models, such as Independent Cascade (IC) model [1].

They further prove that the problem is NP-hard. Luckily, $\sigma(S)$ is nondecreasing submodular. Utilizing this property, Kempe *et al.* have proposed a greedy algorithm with a $(1 - 1/e)$-approximation to the optimal solution [1].

As the greedy algorithm is time-consuming, an important direction is to speedup computation [2]–[4], [11]. Meanwhile, there are some other directions of research, such as learning influence probabilities and proposing new models to better reflect the real-world influence propagation. For a comprehensive survey on influence diffusion, we recommend the book by Chen *et al.* [12].

Diversity

Diversity has been studied in many disciplines, such as sociology [13] and economics [14], with various measures proposed. In the field of data mining, diversity also has many applications, in search [15], recommendation [6], and text summarization [7], etc. To the best of our knowledge, diversity in influence maximization has remained unexplored, and this is the first work to introduce diversity into the problem of influence maximization.

DIVERSIFIED SOCIAL INFLUENCE MAXIMIZATION

The General Framework

Since the two goals of maximizing influence and diversity may be conflicting, we propose to set our optimization function as a linear combination of both. Specifically, given a graph $G = (V, E)$, the framework for diversified social influence maximization is as follows.

$$\max_{S} \quad F(S) = (1 - \gamma)\frac{\sigma(S)}{\overline{\sigma}} + \gamma\frac{D(\boldsymbol{\mu}^S)}{\overline{D}},$$
$$\text{s.t.} \quad S \subseteq V, \ |S| = K, \qquad (2)$$

where $\boldsymbol{\mu}^S$ is a vector of length $|V|$. The i-th entry μ_i^S represents the probability of node i being activated when the seed set is S. $\sigma(S)$ represents the influence of S, while $D(\boldsymbol{\mu}^S)$ represents the diversity of the influenced crowd when seed set is S. $\overline{\sigma}$ and $\overline{D}$ are normalization factors, commonly chosen as the maximum possible values of $\sigma(S)$ and $D(\boldsymbol{\mu}^S)$, respectively. $\gamma \in [0, 1]$ is a tunable parameter to specify the degree of diversity. When $\gamma = 0$, the formulation is equivalent to the original influence maximization problem; the larger γ is, the more diversity is emphasized. Furthermore, we have the following lemma.

Lemma 1. μ_j^S *is submodular with respect to* S. *Namely,* $\forall S \subseteq T, v \notin T, \mu_j^{S\cup\{v\}} - \mu_j^S \geq \mu_j^{T\cup\{v\}} - \mu_j^T.$

The proof follows the live-path argument by Kempe *et al* [1]. Details are omitted for brevity.

A Specific Formulation

To get a workable formulation of Equation (2), we just need to specify a diversity measure (i.e., $D(\boldsymbol{\mu}^S)$), since $\sigma(\cdot)$ is already well defined [1].

As already mentioned, many diversity measures rely on category information, such as the widely-used Shannon entropy and Gini index. Thus, we assume that each node belongs to

one or more categories, described as a distribution over these categories. Supposing there are C categories, and the category distribution of node i is denoted as $\mathbf{w}_i$, a vector of length C satisfying $\sum_{j=1}^{C} w_{ij} = 1$. We propose a class of crowd diversity measures for influence maximization as follows.

$$D(\boldsymbol{\mu}^S) = \sum_{i=1}^{C} f(\sum_{j=1}^{|V|} w_{ji}\mu_j^S), \qquad (3)$$

where $f : \mathbb{R} \to \mathbb{R}$ can be any nondecreasing concave function satisfying $f(0) = 0$ (to ensure $D(\boldsymbol{\mu}^{\emptyset}) = 0$). For example, we could choose $f(x) = \frac{x}{1+x}, x \geq 0$, $f(x) = ln(1 + x), x \geq 0$ or $f(x) = \sqrt{x}, x \geq 0$.

Actually Equation (3) can be seen as a generalized version of the one proposed by Lin and Bilmes [7]. We have chosen this diversity measure instead of the commonly-used Shannon entropy or Gini index, because this measure has the nice property of monotonicity and submodularity, shown in the following lemma.

Lemma 2. $D(\boldsymbol{\mu}^S)$ *in Equation (3) is nondecreasing submodular with respect to* S.

Proof: It is obvious that $D(\boldsymbol{\mu}^S)$ is nondecreasing with respect to S, since $\forall j, \mu_j^S \geq 0$ and $\forall i, j, w_{ij} \geq 0$. Now, we turn to prove its submodularity based on the following theorem [7]:

Theorem. *Given functions* $F : 2^V \to \mathbb{R}$ *and* $f : \mathbb{R} \to \mathbb{R}$, *the composition* $F' = f \circ F : 2^V \to \mathbb{R}(i.e., F'(S) = f(F(S)))$ *is nondecreasing submodular, if* f *is nondecreasing concave and* F *is nondecreasing submodular.*

Because $D(\boldsymbol{\mu}^S) = \sum_{i=1}^{C} f(\sum_{j=1}^{|V|} w_{ji}\mu_j^S)$, and f is nondecreasing concave while μ_j^S is nondecreasing submodular (Lemma 1), $D(\boldsymbol{\mu}^S)$ is also nondecreasing submodular with respect to S. ∎

Substitute Equation (3) into Equation (2), and we get a specific formulation for the diversified influence maximization problem as follows:

$$\max_{S} \quad F(S) = (1 - \gamma)\frac{\sigma(S)}{\overline{\sigma}} + \gamma\frac{\sum_{i=1}^{C} f(\sum_{j=1}^{|V|} w_{ji}\mu_j^S)}{\overline{D}},$$
$$\text{s.t.} \quad S \subseteq V, \ |S| = K. \qquad (4)$$

Since $\sigma(S)$ has been proved to be nondecreasing submodular [1], and a non-negative linear combination of submodular functions is also submodular, we have the following theorem.

Theorem 1. $F(S)$ *in Equation (4) is nondecreasing submodular with respect to* S.

Relaxation

According to the general framework, we have to compute μ_j^S, i.e., the probability of node j being activated when the seed set is S. However, the common way to compute μ_j^S is to run Monte Carlo simulations for sufficient times, which is time-consuming. This motivates us to search for a simpler form of diversity measure to enable efficient computation. Inspired by the homophily phenomenon observed in many social networks [16], we relax the problem from enforcing diversity on influenced crowd to just enforcing diversity on

seed set. The idea behind this relaxation is that if the seed set is diverse, then the resulting active node set would be diverse, too. The relaxed optimization problem is defined as follows.

$$\max_S \quad F_s(S) = (1-\gamma)\frac{\sigma(S)}{\overline{\sigma}} + \gamma\frac{D(S)}{\overline{D}},$$
$$\text{s.t.} \quad S \subseteq V, \ |S| = K, \qquad (5)$$

where $D(S)$ is the diversity of set S.

Specifically, we choose two variants for $D(S)$: $D(S) = \sum_{i=1}^{C} f(\sum_{j\in S} w_{ji} * 1)$ and $D(S) = \sum_{i=1}^{C} f(\sum_{j\in S} w_{ji}\sigma(\{j\}))$. We call the former *uniform* diversity measure, since each seed's contribution to the diversity is equal (i.e., 1). In contrast, the latter is called *weighted* diversity measure, as seed j's contribution to the diversity is weighted by its influence $\sigma(\{j\})$.

However, the computation of $\sigma(S)$ still needs Monte Carlo simulation. Nevertheless, this idea of relaxation could help incorporate diversity into some heuristics to get efficient and diversity-aware algorithms for the influence maximization problem. In the following, we take the degree centrality and PageRank heuristic as examples for illustration.

The degree centrality method simply selects K nodes with largest degrees, while the PageRank heuristic selects K nodes with largest PageRank values. These heuristics could also be regarded as maximizing certain submodular objectives. For example, degree centrality could be seen as finding a set S of size K to maximize $\sum_{i\in S} deg(i)$, where $deg(i)$ is the degree of node i. Hence, we could easily extend this relaxed form (Equation (5)) to incorporate diversity into degree centrality to get a nondecreasing submodular objective:

$$\max_S \quad F_d(S) = (1-\gamma)\frac{\sum_{i\in S} deg(i)}{\overline{deg}} + \gamma\frac{D(S)}{\overline{D}},$$
$$\text{s.t.} \quad S \subseteq V, \ |S| = K. \qquad (6)$$

Similarly, we also have two choices—the uniform diversity measure: $D(S) = \sum_{i=1}^{C} f(\sum_{j\in S} w_{ji} * 1)$ and the weighted one: $D(S) = \sum_{i=1}^{C} f(\sum_{j\in S} w_{ji} deg(j))$.

By simply substituting the degrees of nodes with PageRank values, we could get the formulation of the diversified PageRank. Details are omitted for brevity.

Algorithm

Since the function $F(S)$s (e.g., $F_s(S)$, $F_d(S)$) defined in this paper are all nondecreasing submodular [1], a simple greedy algorithm could achieve a $(1 - 1/e)$-approximation of the optimal solution. This process is summarized in Algorithm 1.

Algorithm 1 Greedy(G, K, $F(\cdot)$)

Input: $G(V, E), K, F(\cdot)$
Output: S
  $S \Leftarrow \emptyset$
  while $|S| < K$ **do**
    $v^* \Leftarrow \arg\max_{v\in V\setminus S} F(S\cup\{v\}) - F(S)$
    $S \Leftarrow S\cup\{v^*\}$
  return S
---

[1] Proof is omitted for brevity.

EXPERIMENTS

Experimental Data

We use two datasets collected from MovieLens[2] (a movie-recommending website) and Yahoo! Answers (an online Q&A website) [17], [18], which are both publicly available. Diversified influence maximization performed on these two datasets could help tasks such as recommending a set of movies with diverse flavors to users and finding a team of experts (answerers) with a comprehensive scope of knowledge.
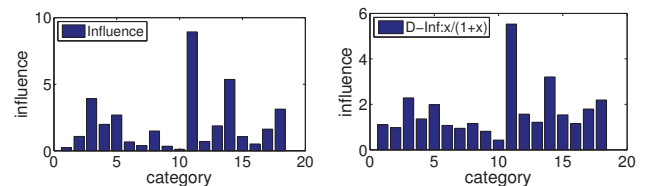
In the Movie graph, each node represents a movie. If a user rates (watches) movie A before movie B, we insert a directed edge $(A \to B)$ into the graph. After that, we substitute parallel edges between each pair of nodes with a single edge with weight being the number of the parallel edges. Then we filter out those edges with weight less than a certain threshold (set to 10 in our experiments), which are likely to be noise. Then we reweight the edges to ensure the weights of in-edges of each node sum up to a certain value between 0 and 1 (set to 0.1 in our experiments), so we could regard these weights as influence probabilities. So an edge $(A \to B)$ with weight p intuitively means that the probability of a user watching movie B after movie A is p.

In the Yahoo! graph, we select top-30 categories with maximum number of answers. The nodes represent users, including askers and answerers. The categories a node belongs to are just the categories of the questions he/she asked. If user A answers a question asked by user B, we insert a directed edge $(A \to B)$ into the graph, and we reweight the edges in a similar way as in the movie graph.

Due to the lack of prior knowledge, we assume that the category distribution of each node is uniform over labeled categories. For a movie, categories are genres of the movie; for a user on Yahoo! Answers, the categories are just the categories of questions the user participated in. Some basic statistics about these graphs are illustrated in Table I.

TABLE I: statistics of graph data

Graphs	#Nodes	#Edges	Density	#Categories
Movie	1,112	275,675	0.22	19
Yahoo!	41,042	193,618	1.15e-4	30

(a) Original Result   (b) Diversified Result

Fig. 2: Influence distribution comparison between *Influence* and *D-Inf* on the Movie graph, where $f(x) = \frac{x}{1+x}$, $\gamma = 1$, and the seed set size is 15.

Methods for Comparison

To evaluate the effectiveness of our formulation in promoting diversity in the influence maximization problem, we select several benchmark methods as follows.

- **Influence** is the basic greedy algorithm to maximize social influence (i.e., Equation (1)).

[2] http://movielens.umn.edu/, data available at http://grouplens.org/datasets/

(a) Diversity Gain (Movie)   (b) Diversity Gain (Yahoo!)   (c) Influence Loss (Movie)   (d) Influence Loss (Yahoo!)
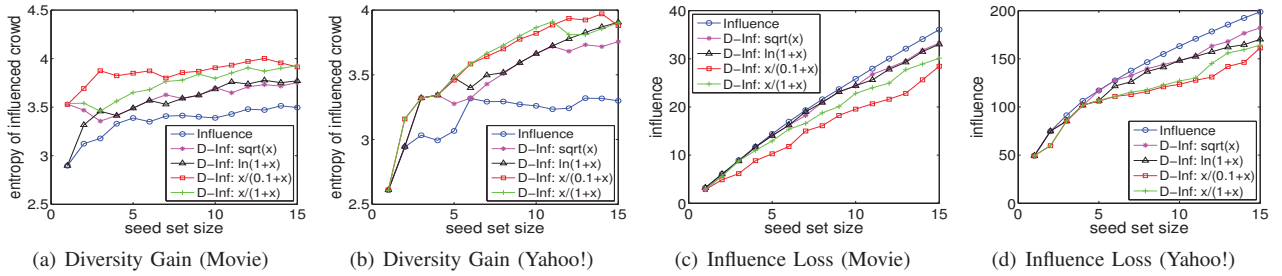
Fig. 1: The performance of diversification in social influence.

- **PageRank** is the greedy algorithm to maximize the sum of PageRank values.
- **Degree** is the greedy algorithm to maximize the sum of Degrees (i.e., degree centrality).

The methods that we have developed in the previous section are summarized as follows.

- **D-Inf** is the greedy algorithm to maximize the Diversified Influence (i.e., Equation (4)).
- **Seed-DU(W)** is the greedy algorithm to maximize the Diversified influence of Seed set with Uniform (Weighted) diversity measures (i.e., Equation (5)).
- **Deg-DU(W)** is the greedy algorithm to maximize the sum of Degrees with Diversification using Uniform (Weighted) diversity measures (i.e., Equation (6)).
- **PR-DU(W)** is the greedy algorithm to maximize the sum of PageRank values with Diversification, using Uniform (Weighted) diversity measures.

Although our formulations for these methods can be used in either IC or LT model without modification, we have chosen only the IC model, which is more widely used, for our experimental evaluation.

Evaluation Metrics

We leverage the widely-used Shannon entropy[3] to evaluate the diversity of influenced crowd. After obtaining the seed set S using each method, we run Monte Carlo simulations under IC model for 20,000 times to approximate $\sigma(S)$ and $\boldsymbol{\mu}^S$. The Shannon entropy of the influenced crowd is defined as $Shannon\ entropy = \sum_{i=1}^{C} -p_i \log_2 p_i$, where $p_i = \frac{\sum_{j=1}^{|V|} w_{ji}\mu_j}{\sum_{j=1}^{|V|} \mu_j}$, and p_i can be interpreted as the proportion of influence distributed to the i-th category.

Effect of Diversification

To show the effect of diversification of social influence on different seedset sizes, we fix the value of γ to 1 (which corresponds to the maximum possible diversity that could be achieved under our formulation). Fig. 1 shows the overall performance of *D-Inf*. Specifically, the results in Fig. 1 (a) and (b) clearly validate that *D-Inf* can improve the diversity of influenced crowd compared with traditional influence maximization. Furthermore, from Fig. 1 (c) and (d) we can observe that the overall influence of the seed set selected by *D-Inf* is often less than that of the traditional method, though the

---

[3]Note that we did not use it to measure diversity in the objective function, because the objective would be hard to optimize otherwise.

influence loss is relatively low. This is reasonable, since *D-Inf* does not directly maximize the overall influence of seed set but tries to strike a balance between influence and diversity. Fig. 2 shows the influence distribution over different categories on the movie graph with respect to different methods, which indicates that *D-Inf* leads to a seed set with more balanced distribution of influence in different categories.

Since the results of *Seed-DU(W)*, *Deg-DU(W)*, and *PR-DU(W)* show similar trends, we omit them due to space limit.

Trade-off between Diversity and Influence

To further study the flexibility of *D-Inf* in balancing diversity and influence, we evaluate its performance by tuning parameter γ. Fig. 3 shows the flexibility test on the Movie graph, where we choose two representative diversity functions with $f(x) = sqrt(x)$ and $f(x) = \frac{x}{1+x}$. From the results, we can observe that, with the increase of parameter γ, diversity increases while influence decreases, which indicates that γ is effective in controlling the balance between influence and diversity. Moreover, we can also observe that the uniform measures have higher tunability in terms of controlling diversity and influence than weighted measures (note that the solid lines, except the red one representing *D-Inf*, span larger ranges with regard to Y-axis than those dash lines).

Efficiency Comparison

We implement Influence, *D-Inf* and *Seed-DU(W)* by applying CELF optimization [2] to the greedy algorithm, and PR-DU(W) and Deg-DU(W) using simple greedy algorithm without CELF optimization (which is already quite efficient). To compare the efficiency of different methods, we choose $f(x) = ln(1+x)$ and fix the seed set size to 15 as an example. Results are shown in table II. Note that *Influence* is not listed in the column of "Methods", since it is actually the special case of *D-Inf* given $\gamma = 0$.

From the table, we have the following observations: 1) The diversified versions ($\gamma \neq 0$) of *D-Inf* and *Seed-DU(W)* are much more time-consuming than original ones ($\gamma = 0$). The reason is that in the greedy algorithm with CELF optimization, we maintain an upperbound of influence contribution for each node. In the original version ($\gamma = 0$), adding a node to the seed set will cause very few nodes (mainly the neighbors) to update their upperbound. However, in the diversified versions, adding a node to the seed set will also cause the nodes belonging to the same category (or having overlapping categories) to update their upperbounds. So, the diversified versions often have more updates and need more computation; 2) The diversified PR-DU(W) and Deg-DU(W) remain efficient and scalable despite the overhead of computing diversity values.

(a) $f(x) = sqrt(x)$
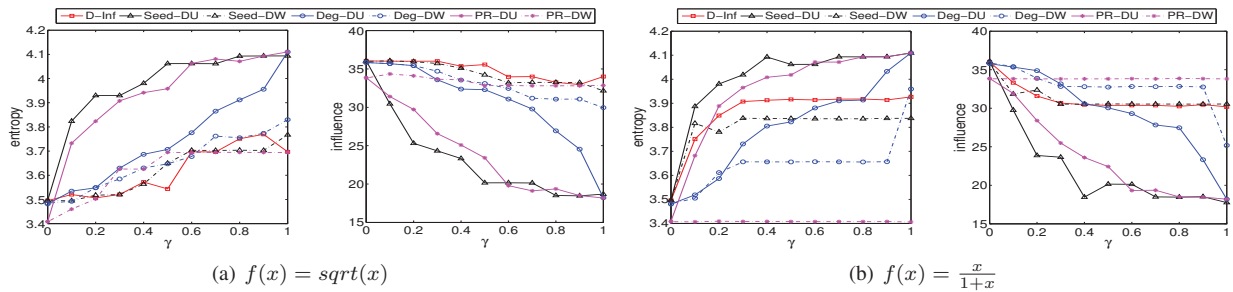
(b) $f(x) = \frac{x}{1+x}$

Fig. 3: Trade-off between diversity and influence by tuning γ on the Movie Graph for different diversity functions.

TABLE II: Efficiency comparison (in seconds).

Graphs	Methods	γ			
		0	0.3	0.6	1
Movie	D-Inf	24	1,199	1,085	1,058
	Seed-DU	20	1,628	1,044	1,080
	Seed-DW	21	737	1,415	1,299
	PR-DU	0.56	0.57	0.57	0.59
	PR-DW	0.56	0.58	0.57	0.57
	Deg-DU	1e-4	9.4e-3	9.0e-3	9.9e-3
	Deg-DW	1e-4	8.7e-3	8.8e-3	8.8e-3
Yahoo!	D-Inf	8	2,369	3,123	4,865
	Seed-DU	7	9,147	6,213	6,527
	Seed-DW	7	745	1,014	1,167
	PR-DU	6.99	7.73	7.50	7.65
	PR-DW	6.99	7.57	7.60	7.68
	Deg-DU	2.3e-3	0.41	0.41	0.41
	Deg-DW	2.2e-3	0.40	0.40	0.40

CONCLUSION AND FUTURE WORK

In this paper, we formulated the diversified social influence problem as an optimization problem. Based on this formulation, we used a class of diversity measures to quantify the diversity of the influenced crowd. Meanwhile, we proved that the objective function is nondecreasing submodular, and provided a greedy solution. Moreover, we proposed a relaxation that only enforces diversity on the seed set. With this relaxation, we integrated diversity information into some widely-used heuristics, i.e., PageRank and degree centrality. Extensive experimental results showed that the proposed objective functions prefer seed sets that are capable of influencing a more diverse crowd, demonstrating the effectiveness of our formulation.

In the future, we will attempt to design more effective and efficient algorithms for the problem of maximizing diversified influence. Hopefully, we will also explore the opposite direction: the targeted influence maximization problem, which would be of interest in targeted marketing.

ACKNOWLEDGMENT

REFERENCES

[1] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *ACM SIGKDD*. ACM, 2003, pp. 137–146.

[2] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *ACM SIGKDD*. ACM, 2007, pp. 420–429.

[3] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *ACM SIGKDD*. ACM, 2010, pp. 1029–1038.

[4] K. Jung, W. Heo, and W. Chen, "Irie: Scalable and robust influence maximization in social networks," in *IEEE 12th International Conference on Data Mining (ICDM)*. IEEE, 2012, pp. 918–923.

[5] Q. Liu, B. Xiang, E. Chen, Y. Ge, H. Xiong, T. Bao, and Y. Zheng, "Influential seed items recommendation," in *ACM conference on Recommender systems*. ACM, 2012, pp. 245–248.

[6] G. Adomavicius and Y. Kwon, "Optimization-based approaches for maximizing aggregate recommendation diversity," *INFORMS Journal on Computing*, 2014.

[7] H. Lin and J. Bilmes, "A class of submodular functions for document summarization." in *ACL*, 2011, pp. 510–520.

[8] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual review of sociology*, pp. 415–444, 2001.

[9] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." 1999.

[10] P. Domingos and M. Richardson, "Mining the network value of customers," in *ACM SIGKDD*. ACM, 2001, pp. 57–66.

[11] Y. Yang, E. Chen, Q. Liu, B. Xiang, T. Xu, and S. A. Shad, "On approximation of real-world influence spread," in *Machine Learning and Knowledge Discovery in Databases*. Springer, 2012, pp. 548–564.

[12] W. Chen, L. V. Lakshmanan, and C. Castillo, "Information and influence propagation in social networks," *Synthesis Lectures on Data Management*, vol. 5, no. 4, pp. 1–177, 2013.

[13] S. Lieberson, "Measuring population diversity," *American Sociological Review*, pp. 850–862, 1969.

[14] A. Stirling, "On the economics and analysis of diversity," *SPRU, Electronic Working Papers Series, Paper*, vol. 28, 1998.

[15] R. L. Santos, C. Macdonald, and I. Ounis, "Exploiting query reformulations for web search result diversification," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 881–890.

[16] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010, pp. 261–270.

[17] Y. Liu, J. Bian, and E. Agichtein, "Predicting information seeker satisfaction in community question answering," in *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2008, pp. 483–490.

[18] H. Zhu, E. Chen, H. Xiong, H. Cao, and J. Tian, "Ranking user authority with relevant knowledge categories for expert finding," *World Wide Web*, pp. 1–27, 2013.