Effective Promotional Strategies Selection in Social Media: A Data-Driven Approach

Kun Kuang, Meng Jiang, Peng Cui, Hengliang Luo, Shiqiang Yang, *Senior Member, IEEE*

Abstract—Nowdays, many companies, organizations and individuals are using the function of sharing or retweeting information to promote their products, policies, and ideas on social media. While a growing body of research has focused on identifying the promoters from millions of users, the promoters themselves are seeking to know which strategy can improve promotional effectiveness, which is rarely studied in the literature. In this work, we investigate an open problem of effective promotional strategy selection via causal analysis which is challenging in identifying and quantifying promotional strategies as well as the selection bias when estimating the causal effect of promotional strategies from observational data. We study the promotional strategies not only on the content level (what to promote) but also on the context level (when and how to promote). To alleviate the issue of selection bias in observational studies, we propose a data-driven approach that is a Propensity Score Matching (PSM) based method, which helps to evaluate the causal effect of each promotional strategy and discover the set of effective strategies to predict the promotional effectiveness (i.e., the number of users infected by the promotion). We evaluate our proposed method on a real social dataset including 194 million users and 5 million promoted messages. Experimental results show that (1) the top-ranked strategies by our PSM based method significantly and consistently outperform the correlation based feature selection methods in predicting promotional effectiveness; (2) we conclude our observations from the real data with three interpretable and practical ideas for steering social media promotion.

Index Terms—Social media promotion, propensity score, promotional strategies, promotional effectiveness

1 Introduction

PROMOTION is everywhere and expensive. People always wonder how to use a smaller marketing budget for a smarter promotion. For example, the administration of the US President spent nearly $700 million dollars to promote the Obamacare[1]; the Starbucks spent $485 million dollars for media advertising in 2010-2014[2]. Governments and companies have realized the great value of the promotional function of social media that users can easily generate and share messages on a huge network. However, there is a lack of real data-driven approaches to support effective strategy-making processes to promote commercial product, disaster alert and public policy.

Fortunately, a rich line of work focus on identifying promoters in social media [33][32][25][21][20]. Thanks to these techniques that accurately distinguish the roles of accounts (promoters and ordinary users), we are able to observe various strategies from the promoters. For example, as "Strategy 1" shown in Figure 1, there are groups of promoters that believe *repeat promotion* can attract repeat customers. Thus, they post the same message (e.g., "Up to 30% Off Coupon Code") for many times. Another strategy, "Strategy 2", is to decorate the promoters' messages according to the profiles or interests of recipients (e.g., adding "a chance to win 30% Off for Prada and Gucci" when sending the clothing promotional messages to young ladies), called *personalized decoration*. They believe their personalized feature can significantly attract the target customers stronger than just sending the original messages. The question is, *which strategy is more effective, or what are the most effective strategies in social media promotions?* That is the effective promotional strategy selection problem.

It is nontrivial to address the problem, and there are two major challenges as follows. First, there is a lack of study on promotional strategies in social media, while we spot many cases in real data. Besides the aforementioned strategies "repeat promotion" and "personalized decoration", the promoters make their strategies from different factors including content, user, network and timing. Although the research literature of cascade prediction [12][10][37] and influence maximization [7][6][26][36] are proposed to analyze the patterns of information propagation or select the top influential users in social network. However, they ignore the behaviors of promoters and promotional strategies. Therefore, the effective promotional strategies selection problem has never been studied on the promoter side and from the data-driven angle. Specifically, the problem can be defined as follows: Given the social network and a group of promoters that a company can manipulate (i.e., using them to post any content at any time), which strategies will be effective for these promoters and achieve good effectiveness?

Another challenge is the issue of selection bias when evaluating the causal effect of a promotional strategy from observational data. Actually, in observational studies, the selection bias is a principal issue in the estimation of *treatment effect* [18] which here refers to the causal effect of a treatment variable (whether a promotion adopts a particular strategy) on an outcome variable (the number of users infected by this promotion) as shown in Figure 2. The selection bias issue is induced by that the treatments

- *Kun Kuang, Peng Cui and Shiqiang Yang are with the Department of Computer Science and Technology in Tsinghua University, Beijing 100084, China. Email: kkun2010@gmail.com, cuip@tsinghua.edu.cn, yangshq@tsinghua,edu.cn.*
- *Meng Jiang is with the Department of Computer Science and Engineering, University of Notre Dame, IN, USA. Email: mjiang2@nd.edu.*
- *Hengliang Luo is with the Data Computing Laboratory, Samsung R&D Institute of China - Beijing. Email: hl.luo@samsung.com.*

1. http://freebeacon.com/issues/govt-spent-700-million-promoting-obamacare/
2. http://www.statista.com/statistics/275195/starbucks-advertising-spending-in-the-us/

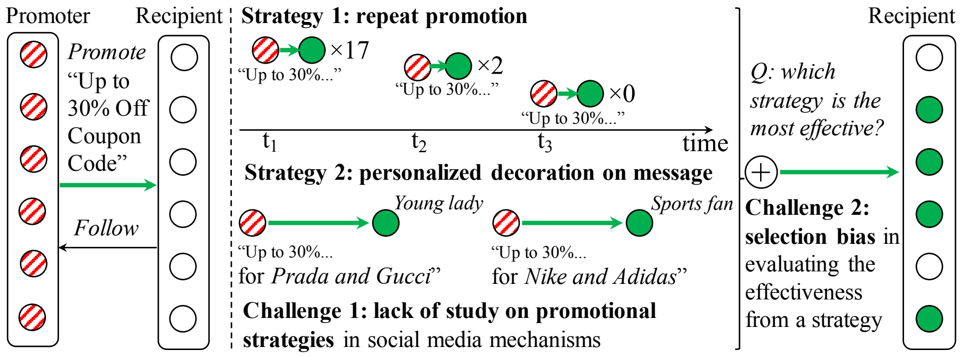


Fig. 1: *What are the promotional strategies, and which is the most effective?* There is a lack of study on social media promotional strategies, while we spot many in real data, for instance, repeat promotion and personalized decoration on message. However, it is difficult to answer the above questions due to the selection bias in evaluating their effectiveness with observational studies.

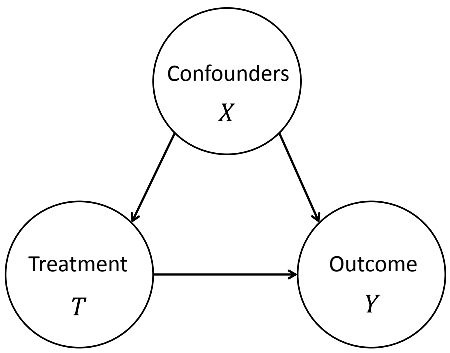


Fig. 2: *Framework of treatment effect estimation in observational studies.* The framework includes an Treatment (T), such as whether a promotion adopts a particular promotional strategy, an Outcome (Y), such as the number of users infected by the promotion, and a set of Confounders (X), which are causally related to both the treatment and outcome, such as the other features and strategies of the promotion. Since the strategy (treatment) is not randomly assigned to promotions in observational studies, the distributions of confounders may be different between promotions with and without the particular strategy and make selection bias on outcome. Hence, one have to reduce the selection bias induced by the confounders before estimating the treatment effect.

are not randomly assigned to promotions in observational studies, which makes the different distributions of confounders (other features and strategies) between the promotions with and without the particular strategy. Since the different distribution of confounders which may be associated with outcome variable, we can not distinguish the effect from treatment variable and confounders, leading to imprecise estimation of the treatment effect. Hence, to precisely estimate the treatment effect in observational studies, one have to reduce the selection bias by balancing the distribution of confounders among units (*i.e.*, promotions in this paper) with different treatment value.

In order to address these two challenges, in this paper, we provide an in-depth study of the social media promotional mechanism with Weibo (a Twitter-like social platform in China) data: we not only extract a large set of static features of promotion (e.g., degree, PageRank value) from the perspectives of root user (who generates the message first), root message content and promoter (who promotes the message to his/her followers), but also present as many as 17 promotional strategies that we observe in content level (e.g., personalized decoration) and context level (e.g., repeat promotion). To the best of our knowledge, this is the first time of a piece of research that investigates social media promotional strategies with rich complex behavioral data.

We propose a data-driven approach, Propensity Score Matching (PSM) based algorithm, to estimate the causal effect of each strategy and address the effective promotional strategy selection problem from causal angle. In particular, conditioned by the propensity score [43], the distribution of confounders will be similar between treated and untreated promotions, where a treated (untreated) promotion is a promotion that adopts (does not adopt) a given strategy. With PSM based method, we can successfully reduce the selection bias and precisely estimate the casual effect of all 17 promotional strategies, which helps to select the effective strategies. Furthermore, we summarize three tactics to steer social media promotions.

- *Personalized decoration*, as well as *early-stage message* (small $depth\text{-}in\text{-}path$, i.e., promote before the message has gone too deep in the network) and *proper timing* (big $user\text{-}active\text{-}time$, i.e., promote in the hours that users are active in social media) can significantly improve the promotional effectiveness.
- *Repeat promotion* does not equal to repeat customers. There is a trade-off: the more a promoter repeats the same content, the fewer adoptions he/she will harvest, though the total number of adopted promotions is monotonic nondecreasing as the number of promotions increasing.
- Popular and ordinary promoters have different effective strategies. Timing or contextual factors are more important for popular promoters, while the ordinary promoters should focus on generating good content.

It is worthwhile to highlight our contributions as follows.

- *Causal Angle:* We address the problem of effective promotional strategy selection from causal angle. With Weibo's real data, we identify and quantify a large set of promotional strategies from both context and content levels. Then, we select the most effective strategies by their causal effect on promotional effectiveness.
- *Selection bias reducing:* In order to overcome the serious issue of selection bias that regression-based methods and predictive models suffer from, we propose a data-driven approach to reduce the selection bias and estimate causal

effect of each promotional strategy from observational data.

- *Effectiveness and insights:* Experimental results demonstrate that the top-k strategies selected by our PSM based method are more consistently effective than baseline models (MRel and mRMR). And we summarize three insightful and practical points for steering social media promotions.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 gives the definition of our proposed problem of *effective promotional strategy selection in social media*. Promotional strategies are listed and discussed in Section 4. In Section 5, we introduce the details of our PSM based method. Section 6 shows experimental results, and finally, Section 7 concludes the paper.

2 Related Work

In this section, we review four fields of related work, influence maximization, cascade prediction, causal discovery and propensity score matching, then point out the uniqueness of our work.

2.1 Influence Maximization

The influence maximization problem aims to select a small set of users in a social network that can maximize the influence of target messages under a certain influence cascade models. Domingos and Richarson first studied this problem [13][41], and Kempe *et al.* formulated it as an optimization problem and proved it is NP-hard [27]. Later, many researchers devoted themselves to find an efficient and scalable solution, such as, greedy algorithms [48][16] and heuristics algorithms [7][28][6][8]. However, in the previous methods, each user is assumed to have a binary state: adopt or not, which cannot fully and clearly describe the real cases (like recurrent adoptions) in social promotion. Moreover, their selected subset of users are often the top influential, while we study a new problem that given set of promoters who can be well manipulated, what promotional strategies can improve their promotional effectiveness. In this paper, we focus on effective promotional strategies selection problem for steering social media promotions.

2.2 Cascade Prediction

Recently, fruitful researches have been proposed to analyze information cascades especially in social media. Yang *et al.* [49] proposed a K-spectral centroid clustering algorithm which uses time series patterns to find information cascades in Twitter. Matsubara *et al.* [34] developed a unified model SPIKEM with seven parameters to explain cascade patterns. Cui *et al.* [12] selected nodes as sensors in social network to predict cascade outbreak in early stage. Yu *et al.* [50] uncovered and predicted the macro cascading process with micro behavioral dynamics. Myers *et al.* [37] predicted the bursty with social network structure dynamics. Cheng *et al.* [10] proposed to analyze if cascade growth is predictable and studied the problem that if a cascade will double the size in the future. Most of these works aim to find rules and patterns of information cascades and to predict cascade popularity in social network, ignoring the behaviors of promoters. The promotion phenomenon in social media is rarely investigated. Understanding and analyzing the behavioral patterns of the promoters and their promotional strategies is still an open and challenging problem. In this paper, we study the promotional strategies effect, and develop an effective promotional strategies selection algorithm to predict the promotional effectiveness.

2.3 Causal Discovery

Discovering causal relations is fundamental to reasoning and intelligence. The gold standard method for causal discovery is to perform active interventions (also called randomized experiments) [38]. However, interventions are often expensive, unethical, or impossible to realize in many situations. In all of these situations, there is a prime need to discovery and reason the causality from observational data. Over the last decade, the state-of-the-art in causal discovery has matured into a wide array of algorithms, including PC-type methods [45], time series methods [17][15], and identifiable methods [19][40][9][35][51][44]. [45] and [38] explored a formal causal semantics based on DAGs for causal discovery. [17] and [15] introduced the Granger causality and its extension for temporal or dynamic causal relations discovery from time series. [19], [35] and [40] involved the additive noise models for causal discovery, since these models are well understood, and [9] proposed a reproducing kernel Hilbert space embedinessbased method to solve a general causal discovery problem. [42] introduced the tool of proxy variables and projections for causal discovery. Casual discovery has proved to be a promising way to infer the causal direction from observations. But, in this paper, we focus on the causal effect estimation from observational data.

2.4 Propensity Score Matching

Evaluating treatment effect in observational studies often requires adjustment for selection bias in pre-treatment variables. In literature, Rosenbaum and Rubin [43] proposed a statistical framework based on propensity score adjustment. It is aiming to balance the difference distribution of confounders among group of units receiving different treatment. Such framework has been widely used in observational causal study, including matching, stratification, weighting and regression on propensity score [2][1][4][47][5][29]. Austin *et al.* [2] described these four propensity score methods. Sinan *et al.* [1] used propensity score matching to distinguish peer-to-peer influence from homophily in dynamic network. [4][47] evaluated the effect of online advertisement based on propensity score. [5] made propensity score matching on network structure. Kun *et al.* [29] proposed to separate the confounders from all variables by jointly optimize propensity score and regression bias. In this work, we introduce the propensity score matching based method to analyze the causal effect of promotional strategies in social media and select the most effective strategies to predict promotional effectiveness, which is a brand new problem to our research community.

Comparing to the preliminary version [30], this one comprises a new effective strategy selection method, experimental efforts and contributions. Key points of differences lie in the following aspects: based on the predefined problem of promotional strategy effect estimation, we move one step further to rank the promotional strategies by their absolute causal effect on promotional effectiveness and select the top-k effective strategies to predict the final promotional effectiveness. The extensive experimental results demonstrate that the top-k strategies ranked by our method are more consistently effective than two classical correlation-based feature selection methods (MRel [46] and mRMR [39]).

3 Problem Statement

Before we define the effective promotional strategy selection problem, we give the definitions of “promoter”, “promotion”, and

TABLE 1: Symbols and definitions.

Symbol	Definition
U_{pro}	The set of promoters
"\$"	Target message
p	Promotion
$c(p)$	Comment in promotion p
$U_{adp}(p)$	The set of ordinary users who adopt p
$PE(p)$	Effectiveness of promotion p
S_{static}	The set of static features
S_{pro}	The set of promotional strategies

"promotional effectiveness" ordinally. The symbols and definitions are given in Table 1.

Definition 1 (Promoter). A promoter $u_{pro} \in U_{pro}$ in social media (*e.g.*, Twitter) is one of a group of users that are manipulated by companies, organizations or individuals and operated to retweet target message (denoted by "\$") for monetary incentives or other purposes.

Among the fruitful research works [33][32][21][20] on identifying social promoters, we use a state-of-the-art effective and scalable algorithm called CROSSSPOT [25] to label every user as whether a promoter or not. CROSSSPOT evaluated the probability of a group of users' behavioral patterns in multi-faceted data, and searched for the most suspicious parts of behaviors (*e.g.*, retweeting, resharing) with a theoretical guarantee of accuracy.

Definition 2 (Promotion). Given a target message "\$" in social media (*e.g.*, Twitter), a promotion p is a retweet "\$+$c(p)$" generated by a promoter u_{pro}, where $c(p)$ is the comment with which u_{pro} decorates the target message "\$" when promoting.

The basic function of social media ensures that the promoter's followers will receive his/her promotion when they log in the platform in the near future. The promoter expects high promotional effectiveness, *i.e.*, his/her promotion will be adopted as many times as possible.

Definition 3 (Promotional effectiveness). The effectiveness of a promotion p is the number of the ordinary users who adopt the promotion (*e.g.*, retweeting/resharing the content to their followers/friends) in the future. In other words, the promotional effectiveness of promotion p, denoted by $PE(p)$, is the size of the set of the ordinary users $U_{adp}(p)$ who adopt the promotion p: $PE(p) = |U_{adp}(p)|$.

In order to improve the effectiveness, the promoters are seeking for a handbook to introduce effective strategies that have significant effect on promotional effectiveness. Thus, they would be able to practice (*i.e.*, make more effective strategy to attract their audiences) in the real world. Here we focus on the fundamental problem: how to define and select strategies which have significant effect on promotional effectiveness.

Problem 1 (Effective Promotional Strategy Selection in Social Media). **Given** a promotion p and multi-faceted information including the social network, the target message "\$" and comment $c(p)$, and **given** a set of static features S_{static} and a set of promotional strategies $S_{pro} = \{s_1, \ldots, s_m\}$, our task is to **evaluate** the causal effect of each strategy on $PE(p)$ and **select** a subset of significant effective strategies $S_{sigpro} \subset S_{pro}$ to predict the promotional effectiveness.

In Section 4, the static features and promotional strategies will be introduced in details. In practice, we evaluate the causal effect of each promotional strategy s_i by setting it as treatment T, other strategies $S_{pro} - \{s_i\}$ and static features S_{static} as confounders $\mathbf{X}$ and the promotional effectiveness $PE(.)$ as outcome Y with the framework as shown in Figure 2. The key challenge of causal effect estimation is the selection bias issue. Since in observational studies, the treatment is not assigned randomly, which make the distribution of confounders are distinct between treated and untreated promotions groups. We will introduce the propensity score matching based method to remove the selection bias when estimating causal effect of strategies, then rank the strategies with their estimated causal effect and select the top-k effective strategies to predict the promotional effectiveness.

4 FEATURES AND PROMOTIONAL STRATEGIES

In this section, we briefly list static features (see Table 2), and investigate promotional strategies (see Table 3) that drive the higher promotional effectiveness from two main dimensions: context and content.

4.1 Static features

Before we estimate the effect of strategies on the promotional effectiveness, we have to eliminate the selection bias induced by static features, which cannot be changed by anyone in the social networks. Table 2 lists the static features from three domains. First, the network structural characteristics of the promoters and their previous impact, including the number of followers/followees of the promoter u_{pro}, the PageRank value of u_{pro}, and the previous average promotional effectiveness of u_{pro}. Second, the characteristics of target message "\$", including the length of message, number of hashtags, mentions, emoticons, question marks, exclamation marks and URLs of message. Third, the characteristics of the root user u_{root} such as the number of followers/followees, PageRank value and the previous average promotional effectiveness.

4.2 Context-level Strategies

Now we proceed by describing context-level strategies mainly for answering when the promoters should act. The first natural factor contributing to improve the promotional effectiveness is the timing. We study the timing factors from many perspectives, for example, how long it has been since the root message was generated, which hour the promotion will be posted, or, how many messages online users usually post at this hour (while their activity often forms a periodic pattern), and the time interval between former and current promotion or current and the next promotion. We also study the depth of this promotion on the promoted message's information propagation path, *i.e.*, how many parent retweet nodes does this promotion node has in the path.

4.3 Content-level Strategies

The other natural factor is the content itself, *i.e.*, the comment $c(p)$ that the promoter decorates on the target message. We group the content-level strategies into two classes: (1) word-count based strategies and (2) topic-distribution based strategies. The first class of strategies are easy to compute, such as the length of comment, the number of hashtags, mentions, emoticons, question marks, exclamation marks and URLs. The second class relies on LDA topic models [3] that have been incorporated into many tasks [14][24]. We denote by $Pr(z|c(p))$ the probability distribution over topic $z \in Z$ assigned to the comment $c(p)$, where Z

TABLE 2: *Static features of a promotion:* it has a few facets that cannot be changed by strategy, including the promoter's popularity, the content of the target message, and the characteristics of the root user.

Promoter (u_{pro})	Target message ("\$")	Root user (u_{root})
num-of-followers-of-u_{pro}	length-of-message-"\$"	if-u_{root}-is-promoter
num-of-followees-of-u_{pro}	num-hashtags-of-message-"\$" ("#XXX")	num-of-followers-of-u_{root}
ratio-of-female-followers-of-u_{pro}	num-mentions-of-message-"\$" ("@XXX")	num-of-followees-of-u_{root}
PageRank-value-of-u_{pro}	num-emoticons-of-message-"\$" (":D")	PageRank-value-of-u_{root}
average-PE-of-u_{pro}	num-question-marks-of-message-"\$" ("...?")	average-PE-of-u_{root}
	num-exclamation-marks-of-message-"\$" ("...!")	
	num-URLs-of-message-"\$" ("http:...")	

TABLE 3: *Promotional strategies of a promotion:* we present both context-level and content-level strategies. Practitioners can easily compute the values after reading the descriptions.

	Strategy	Description
Context	$depth\text{-}in\text{-}path$	Depth of the promotion p in the propagation path (*i.e.*, #parent-retweets)
	$num\text{-}of\text{-}repeat$	Number of repeat: the promoter u_{pro} may repeat retweeting the content "\$"
	$user\text{-}active\text{-}time$	Users' activeness in the *hour* of the promotion p (*i.e.*, periodic pattern)
	$time\text{-}after\text{-}the\text{-}root$	Time interval between the root (target) message "\$" and the promotion p
	$interval\text{-}after\text{-}the\text{-}former$	Time interval between the former promotion and the current one
	$interval\text{-}before\text{-}the\text{-}next$	Time interval between the current promotion and the next one
Content	$length\text{-}of\text{-}comment$	Length of promotional comment $c(p)$
	$num\text{-}of\text{-}hashtags$	Number of hashtags ("#XXX") in promotional comment $c(p)$
	$num\text{-}of\text{-}mentions$	Number of mentioned users ("@XXX") in promotional comment $c(p)$
	$num\text{-}of\text{-}emoticons$	Number of emoticons (":D") in promotional comment $c(p)$
	$num\text{-}of\text{-}question\text{-}marks$	Number of question marks ("...?...?") in promotional comment $c(p)$
	$num\text{-}of\text{-}exclamation\text{-}marks$	Number of exclamation marks ("...!...!") in promotional comment $c(p)$
	$num\text{-}of\text{-}URLs$	Number of URLs ("http:...") in promotional comment $c(p)$
	$topic\text{-}popularity$	Popularity of the topics in the comment $c(p)$ (see Eq. 1)
	$topic\text{-}diversity$	Diversity of all the topics of the comment $c(p)$ (see Eq. 3)
	$topic\text{-}novelty$	Difference between topics of the comment $c(p)$ and target message "\$" (see Eq. 4)
	$topic\text{-}interest$	Similarity between the comment $c(p)$ and recipient's interest (see Eq. 5)

is the set of all the 100 topics. Now we define the following topic-distribution based strategies, including $topic\text{-}popularity$, $topic\text{-}diversity$, $topic\text{-}novelty$, and $topic\text{-}interest$.

The topic-level popularity describes how popular the topics in a given promotional comment of promotion are:

$$topic\text{-}popularity(p) = \sum_{z \in Z} Pr(z|c(p)) \cdot popularity(z), \quad (1)$$

where $popularity(z)$ is the popularity of topic z in social media, which is defined as follow:

$$popularity(z) = \sum_{p \in P} Pr(z|c(p)) \cdot PE(p), \quad (2)$$

where P is the all promotions set in our training dataset.

The topic-level diversity describes how much the topics in the comment of the promotion differ. We define it as the Shannon entropy of its topic distribution:

$$topic\text{-}diversity(p) = \sum_{z \in Z} -Pr(z|c(p)) \cdot log(Pr(z|c(p))). \quad (3)$$

The topic-level novelty has been adopted to evaluate paper quality [14]. It was measured by the difference between a particular paper and other related papers. Here we define it as the KL divergence [31] between the topic distributions of the comment of promotion and the target message:

$$topic\text{-}novelty(p) = \sum_{z \in Z} Pr(z|c(p)) ln \frac{Pr(z|c(p))}{Pr(z|\$)}. \quad (4)$$

The topic-level interest describes the similarity between the comment of promotion $c(p)$ and the recipient's interesting of promoter u_{pro}:

$$topic\text{-}interest(p) = \sum_{z \in Z} Pr(z|c(p)) \cdot recipient\text{-}interest(u_{pro}, z), \quad (5)$$

where $recipient\text{-}interest(u_{pro}, z)$ is the recipient's interest of promoter u_{pro} on topic z, which is defined as follow:

$$recipient\text{-}interest(u_{pro}, z) = \sum_{p \in P_{u_{pro}}} Pr(z|c(p)) \cdot PE(p), \quad (6)$$

where $P_{u_{pro}}$ is a set of previous promotions by promoter u_{pro}.

5 Effective Promotional Strategy Selection with Propensity Score Matching

In this section, we present our Propensity Score Matching (PSM) based algorithm to reduce the selection bias and estimate causal effect of promotional strategies from observational data.

As described in Section 3, we evaluate the effect of each promotional strategy s_i by setting it as treatment T, other strategies $S_{pro} - \{s_i\}$ and static features S_{static} as confounders $\mathbf{X}$ and the promotional effectiveness $PE(.)$ as outcome Y. Then, for each promotion p in our problem, we observe a treatment $T_p = t$, a outcome $Y_p(t) = PE(p)$ and a vector of other strategies and features X_p. In this paper, we only consider the binary treatment, that is $t \in \{0, 1\}$. We define the promotions which adopts the strategy, that is $T = 1$, as treated promotions and the others with $T = 0$ as untreated promotions.

To evaluate the causal effect of a given strategy (treatment T) on the outcome Y, we have to remove the selection bias induced by $\mathbf{X}$. And there are two standard assumptions usually made for unbiased evaluating the treatment effect in observational studies.

Assumption 1: stable unit treatment value [11]. The distribution of potential outcome for one unit is assumed to be unaffected by the particular assignment of treatment of another unit given the confounders.

Assumption 2: strong ignorability of treatment assignment [43]. The distribution of treatment is independent of the potential outcome given the confounders. Formally, $T \perp (Y(0), Y(1))|\mathbf{X}$.

When estimating the treatment effect, the primary interest is the distribution of $Pr(Y(t)|\mathbf{X})$ for each $t \in \{0, 1\}$ when fixing $\mathbf{X}$, or its average over the population $Pr(Y(t))$. Due to the fact that we observed only one potential outcome $Y(T = t)$ for each unit, therefore, in order to obtain $Pr(Y(t))$, we have to condition on the observed treatment assignment and confounders [23]. Under assumption 2, we have

$$\begin{aligned} Pr(Y(t)|do(T=t),\mathbf{X}) &= \tfrac{Pr(do(T=t)|Y(t),\mathbf{X})Pr(Y(t)|\mathbf{X})}{Pr(do(T=t)|\mathbf{X})} \\ &= Pr(Y(t)|\mathbf{X}), \end{aligned} \tag{7}$$

hence,

$$Pr(Y(t)) = \textstyle\int_{\mathbf{X}} Pr(Y(t)|do(T=t),\mathbf{X})Pr(\mathbf{X})d\mathbf{X}, \tag{8}$$

where $Pr(Y(t)|do(T = t), \mathbf{X})$ is the conditional distribution of $Y(t)$ by setting T to t and giving $\mathbf{X}$, and $Pr(\mathbf{X})$ is the distribution of $\mathbf{X}$.

In principle, we can model $Pr(Y(t)|do(T = t), \mathbf{X})$ directly, but the result will be strongly biased if the relation between T and $\mathbf{X}$ is omitted or misspecified [22]. Matching and subclassification according to $\mathbf{X}$ can avoid the bias when the confounders $\mathbf{X}$ is in low dimensions. However, as the number of the dimensions of $\mathbf{X}$ increasing, existing methods become computationally infeasible.

To address the high dimensional issue of confounders $\mathbf{X}$, we employ the balancing score, denoted by $b(\mathbf{X})$, to summarize the required information to balance the distribution of $\mathbf{X}$. The balancing score was proposed in [43] for treatment effect estimation with binary treatment. And it had been proved in [43] that the treatment assignment is strongly ignorable when giving the balancing score. Formally, $T \perp (Y(0), Y(1))|b(\mathbf{X})$. The propensity score, denoted by $e(\mathbf{X})$, is the most commonly used balancing score in treatment effect estimation with observational data, which is defined as the conditional probability of be treated when giving the confounders, that is $e(\mathbf{X}) = Pr(T = 1|\mathbf{X})$.

We have the propensity score which is an balancing score as follows:

$$Pr(do(T=t)|Y(t), e(\mathbf{X})) = Pr(do(T=t)|e(\mathbf{X})). \tag{9}$$

Hence we obtain $p(Y(t))$ as

$$Pr(Y(t)) = \int_{e(\mathbf{X})} Pr(Y(t)|do(T=t), e(\mathbf{X}))Pr(e(\mathbf{X}))de(\mathbf{X}). \tag{10}$$

We approximate the integral in Eq. (10) by propensity score matching based algorithm, which matches promotions with the similar propensity score $e(\mathbf{X})$ between treated (T=1) and untreated (T=0) groups, then estimates the average treatment effect $E(Y(1) - Y(0))$ under the matched treated and untreated promotion groups. The propensity score matching algorithm is summarized in Algorithm 1.

Algorithm 1 (Propensity Score Matching Based Algorithm)

Input: the outcome Y_i, the treatment T_i, and the confounders X_i of units indexed by $i = 1, 2, \cdots, N$,
Output: the estimated average treatment effect $E(Y(1) - Y(0))$.
Step 1: find a propensity score $e(X_i)$ for each unit such that the treatment $T_i \perp X_i|e(X_i)$.
Step 2: match the promotions with similar propensity score $e(\mathbf{X})$ between treated and untreated groups.
Step 3: calculate the average outcome $Y(t)$ of promotion within each matched group.
Step 4: estimate the average treatment effect $E(Y(1) - Y(0))$ by comparing with the average value of the outcomes between the matched treated and untreated groups.

In the 1^{st} step of algorithm 1, we estimate the propensity score $e(\mathbf{X})$ using logistic regression model. Hence,

$$e(\mathbf{X}) = p(T=1|\mathbf{X}) = \frac{1}{1+e^{-(\alpha+\beta\mathbf{X})}}, \tag{11}$$

where α and β are the parameters to learn.

In the 2^{nd} step of Algorithm 1, we match the promotions into 2 groups (treated group where $T = 1$ and untreated group where $T = 0$) by employing the nearest neighbor matching method. Specifically, for each treated promotion p whose $T = 1$, we choose an untreated promotion q whose $T = 0$ to match with the constraint that $|e(X_p) - e(X_q)|$ is minimum. And we set a threshold ϵ to filtrate the pairs of p and q that satisfy the condition: $|e(X_p) - e(X_q)| < \epsilon$. After that, we can obtain the matched promotions set $P_{matched}$, including the matched treated and untreated promotions. In this step, we reduce the selection bias in observational data by units matching with propensity score.

Next, in the 3^{th} step of algorithm 1, we calculate the average outcome of treated group $T = 1$ and untreated group $T = 0$. And in the 4^{th} step, we estimated the average treatment effect (ATE) as:

$$ATE = \frac{\sum_{p \in P_{matched}, T(p)=1} PE(p)}{\sum_{q \in P_{matched}, T(q)=0} PE(q)} - 1. \tag{12}$$

The propensity score matching based algorithm helps us to reduce the selection bias and evaluate the treatment (promotional strategy) effect more accurately. Then we rank the promotional strategies by their estimated causal effect and select the top-k effective strategies to predict the promotional effectiveness.

6 EXPERIMENTS

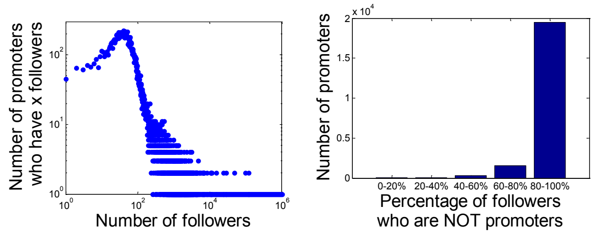
In this section, we first introduce our dataset and then provide comprehensive analysis for the effect of promotional strategies after reducing the selection bias by the Propensity Score Matching (PSM) method, and finally, we demonstrate the high accuracy and efficiency of selected strategies by their estimated effect in promotional effectiveness prediction with a series of experiments.

6.1 Dataset

We crawled a large dataset of both user and tweet information during Nov. 9^{th}, 2011 to Dec. 22^{th}, 2011, from Tencent Weibo, a Twitter-style social platform in China. For the user information, we have a social graph of nearly *200 million* users; for the

TABLE 4: *Data statistics:* we identified 21K promoters from 194M users, and collected over 4M messages that were generated after the promoters posted 814K retweets.

Description	Value
Number of users	193,998,829
Number of promoters	21,378
Number of target messages	13,314
Number of promotions	814,824
Number of adopted promotions	4,213,545



(a) The median value of #followers is 71. (b) Promoters' followers are often NOT promoters.

Fig. 3: *Distributions of promoters' followers:* (a) Few promoters have over 1,000 followers. (b) Only 8.95% of the promoters have more than 20% of followers who are promoters.

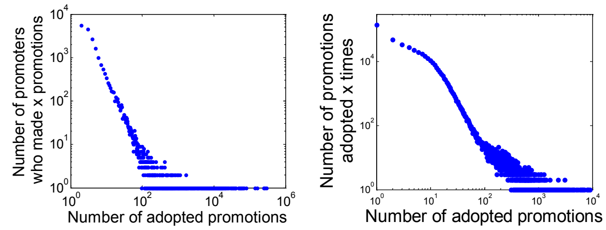
tweet information, we have retweeting paths (*i.e.*, parent-to-child retweeting relationships) consisted of 13,314 target messages and over *4 million* retweets as well as their content including comments and timestamps. We adopt a recently-published, fast suspicious behavior detection algorithm CrossSpot [25] that can identify fraudsters (*e.g.*, promoters in their examples) with multi-dimensional analysis; its suspiciousness score can automatically filter inactive promoters in the social dataset. The data statistics can be found in Table 4.

In Figure 3, we examine the distributions of promoters' followers. Figure 3a plots the frequency of promoters that have a specific number of followers in log-log scale, from which we observe that (i) few promoters have over 1,000 followers, and (ii) the majority of promoters have 20 to 120 followers with the median value 71. Figure 3b plots the number of promoters versus the percentages of their followers who are NOT promoters. It shows that only 8.95% of the promoters have more than 1/5 followers who are also promoters. Figure 4 shows the distributions of adopted promotions. From figure 4a we can see a power-law distribution between the number of adopted promotions (*i.e.*, promotional effectiveness) and the number of promoters whose promotions result in that many adoptions, and from Figure 4b we can also see power-law distributions between the promotional effectiveness and the number of promotions that have received that many adoptions.

6.2 Experiment Settings

Task description: As we have described in the former section, our task is to evaluate the effect of each strategy on promotional effectiveness and select a subset of significant effective strategies to predict the promotional effectiveness.

Evaluation metrics: We adopt two classical metrics to evaluate the prediction accuracy: (1) Root Mean Square Logarithm Error (RMSLE) and (2) Mean Absolute Logarithm Error (MALE). The metrics are generated based on the standard RMSE and MAE



(a) Power-law: most promoters are NOT successful. (b) Power-law: most promotions are NOT effective.

Fig. 4: *Distributions of adopted promotions:* we spot two power-law distributions, one is (a) the relationship between the number of adopted promotions and the number of promoters who have achieved this value; the other is (b) the relationship between the number of adopted promotions and the number of promotions that have received that many adoptions.

definitions, and they have been used in [50]. The power-law distribution of promotional effectiveness in Figure 4b indicates that it does not make sense to calculate the error with the value of effect, and thus, we use the logarithm of the value. Formally, the RMSLE metric is defined as

$$RMSLE = \sqrt{\frac{1}{|P|}\sum_{p_i \in P}(\log(PE(p_i)+1) - \log(\widehat{PE}(p_i)+1))^2}, \tag{13}$$

where p_i denotes the i-th promotion in the set of promotions P, $PE(p_i)$ is the observed promotional effectiveness of promotion p_i, and $\widehat{PE}(p_i)$ is predicted promotional effectiveness of p_i. Similarly, the MALE metric is defined as

$$MALE = \frac{1}{|P|}\sum_{p_i \in P}|\log(PE(p_i)+1) - \log(\widehat{PE}(p_i)+1)|. \tag{14}$$

Note that a smaller RMSLE or MALE, indicates a more accurate result given by a more effective algorithm.

Implementation and parameter settings: We implemented our Propensity Score Matching (PSM) method with MATLAB. All experiments are performed on a single machine with Intel Xeon CPU at 2.40GHz and 32GB RAM. In matching step of algorithm 1, we set $\epsilon = 0.05$ as default threshold parameter for the nearest neighbor matching. To conduct all the experiments, we randomly selected 80% of promotion cases for training and used the remaining 20% for testing.

Baseline methods: In order to compare the performance on promotional effectiveness prediction with the effective strategies selected by our PSM algorithm, we implemented the following two algorithms of feature selection as baselines.

- MRel (maximum relevance)[46]: It measures the importance of promotional strategies by evaluating the mutual information between a feature vector of the strategies and a score vector of the promotional effect.
- mRMR (minimum-redundancy maximum-relevancy)[39]: It maximizes the relevance between promotional strategies and the promotional effectiveness, while minimizes the redundancy between each pair of strategies.

After selecting the top-k strategies with baselines and our PSM algorithm, we fairly apply linear regression on them for promotional effectiveness prediction.

TABLE 5: *Difference in mean of confounders with and without PSM by setting $user\text{-}active\text{-}time$ as the treatment.* With PSM, we make the difference in mean of most confounders between treated and untreated objects become more closer. We alleviate the selection bias and balance the distribution of confounder in observational studies with PSM based method.

Confounders	Without PSM		With PSM	
	Mean of Treated Objects	Mean of Untreated Objects	Mean of Treated Objects	Mean of Untreated Objects
num-of-followers-of-u_{root}	11.724	11.396	11.528	11.521
length-of-message-“$”	308.720	316.843	306.514	306.806
average-PE-of-u_{root}	29.607	24.779	30.337	25.556
num-of-followers-of-u_{pro}	9.823	10.4	10.07	10.127
ratio-of-female-followers-of-u_{pro}	0.486	0.467	0.474	0.472
$depth\text{-}in\text{-}path$	1.194	1.196	1.194	1.201
$num\text{-}of\text{-}repeat$	17.385	25.76	23.058	23.76
$interval\text{-}after\text{-}the\text{-}former$	10917.36	10262.016	8959.105	10272.235
$interval\text{-}before\text{-}the\text{-}next$	10966.023	10225.86	2222159.543	2201697.916
$time\text{-}after\text{-}the\text{-}root$	100317.575	93663.06	97861.786	95831.294
$length\text{-}of\text{-}comment$	33.608	44.218	38.054	38.669
$num\text{-}of\text{-}hashtags$	0.041	0.037	0.037	0.033
$num\text{-}of\text{-}mentions$	0.053	0.054	0.049	0.051
$num\text{-}of\text{-}URLs$	0.001	0.001	0.001	0.001
$num\text{-}of\text{-}mentions$	0.076	0.06	0.062	0.062
$num\text{-}of\text{-}question\text{-}marks$	0.094	0.105	0.109	0.104
$num\text{-}of\text{-}exclamation\text{-}marks$	0.391	0.541	0.461	0.463
$topic\text{-}popularity$	33954.501	45130.029	39129.365	39989.802
$topic\text{-}novelty$	1.807	2.163	1.705	1.771
$topic\text{-}diversity$	2.158	2.804	2.455	2.505
$topic\text{-}interest$	1124.136	1608.045	1255.169	1309.918

6.3 Experimental Results

In this section, we first evaluate the effect of each strategy on promotional effectiveness with our PSM algorithm. Then we conduct extensive experiments to demonstrate the effectiveness of our algorithm in strategy selection for promotional effectiveness prediction, comparing with the state-of-the-art methods.

6.3.1 Strategies effect discovery

Before we present our strategies effect analysis, we show strong evidences that we reduce the selection bias by our PSM algorithm.

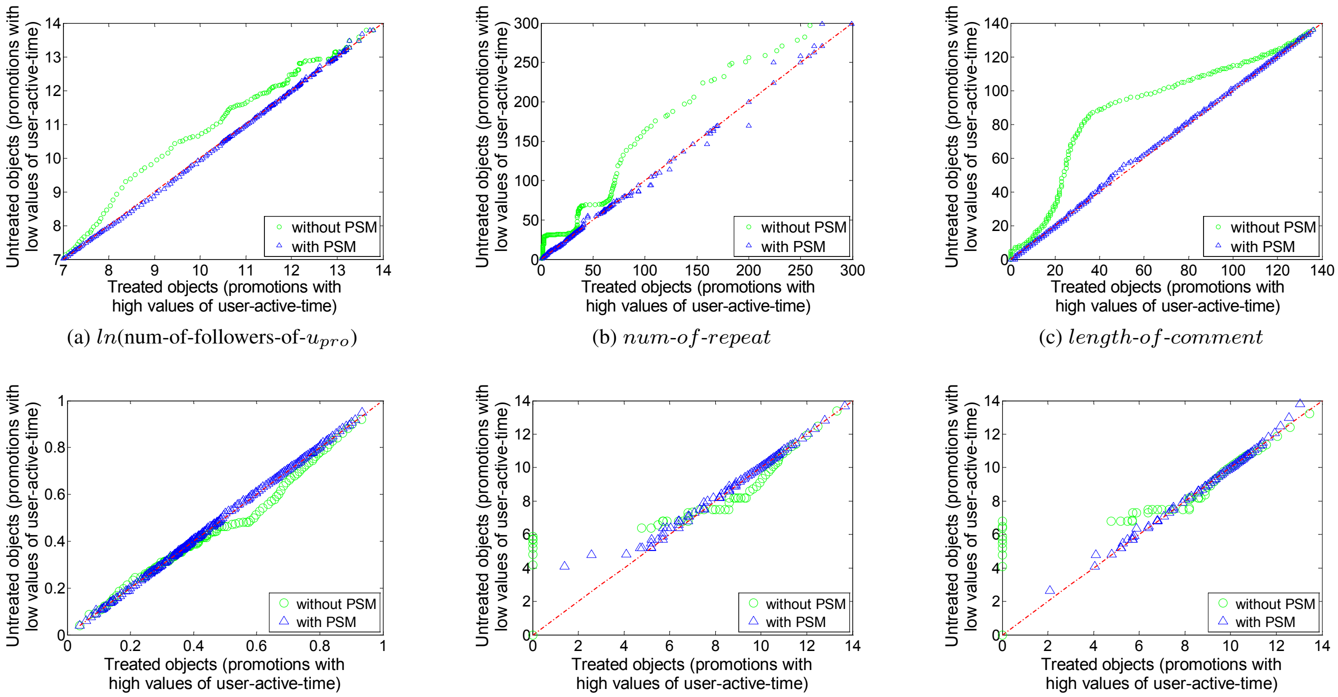
Selection bias reduction. Given a specific strategy as treatment, we examine the data distribution between the treated objects (*i.e.*, promotions) and the untreated objects that have been matched based on the nearest neighbors of propensity score. Hopefully, the distributions of confounders between the matched treated and untreated objects should be similar. The more similar indicates the less selection bias. Quantile-quantile plot (Q-Q plot) provides a standard visualization to examine the distributions. We expect that the treated and untreated objects can have a perfect matching (dots are closely aligned with $y = x$ in Q-Q plot) for every confounder. For example, when we choose $user\text{-}active\text{-}time$ as the treatment, Figure 5 shows Q-Q plots of six confounders: num-of-followers-of-u_{pro}, $num\text{-}of\text{-}repeat$, $length\text{-}of\text{-}comment$, the-ratio-of-female-followers-of-u_{pro}, $interval\text{-}after\text{-}the\text{-}former$, and $interval\text{-}before\text{-}the\text{-}next$. A dot represents a matching of a treated object and an untreated one with the same quantile. We observe that the green circle-dots (original dataset without PSM) deviate the red dashed line $y = x$, but the blue triangle-dots (with PSM) are closely aligned with $y = x$, which indicates that the distributions of confounders are very similar between the matched treated and untreated objects after selection bias reducing with our PSM algorithm.

And in Table 5, we demonstrate the difference in mean of confounders with and without PSM by setting $user\text{-}active\text{-}time$ as the treatment. With PSM, we make the difference in mean of most confounders between treated and untreated objects become more closer. It demonstrate that we can eliminate the selection bias and balance the distribution of confounder in observational studies with PSM based method.

Therefore, we can better estimate the effect of promotional strategies by reducing selection bias with our PSM method.

Strategies effect analysis. We present the results of our effect analysis of promotional strategies. For different levels of the number of the promoters' followers and different promotional strategies, we discuss the polarity (positive or negative), degree of strategies effect and its significance level, as shown in Table 6. A positive (negative) value of the effect indicates that a bigger (smaller) value of the strategy will achieve better effectiveness. We include the SEM in parentheses. Moreover, we conduct a paired t-test on the dataset: a smaller p-value indicates higher significance of the strategy. We divide the p-values into four levels: $p < 0.001$(***), $0.001 \leq p < 0.01$(**), $0.01 \leq p < 0.05$ (*), and $p \geq 0.05$ (*NO* star). The strategies that are *not* significant in *any* type of promotions are omitted for space, such as $topic\text{-}novelty$ and $topic\text{-}diversity$. From the information-rich table, we have the following observations.

Observation 1. Three significant, stable strategies. We find that three strategies $topic\text{-}interest$ (1.316 in average, positive), $user\text{-}active\text{-}time$ (0.313 in average, positive), and $depth\text{-}in\text{-}path$ (-0.209 in average, negative) have strong and robust effects on the promotional effectiveness. First, promotions that are generated when the users are active in the social media can be very effective. For example, adopted promotions often appear at 12 a.m (after lunch) and 8 p.m (after dinner) when people get entertained by the Internet. Therefore, strategy $user\text{-}active\text{-}time$ has strong positive effect on the promotional effectiveness. Second, given an target message, if the promoter decorates it with well-designed comments that match the recipient's personal interest, it is more probable to be adopted by him/her. For example, when a group of promoters are trying to promote “*50% off Discount Men's Shoes including Nike, Adidas...*” to *a teenage NBA fan*, if the promoter decorates the message with “*Basketball shoes! Nike Hyperdunk! RT ... 50% off Discount ...*”, the fan must be more interested in it than the original message.



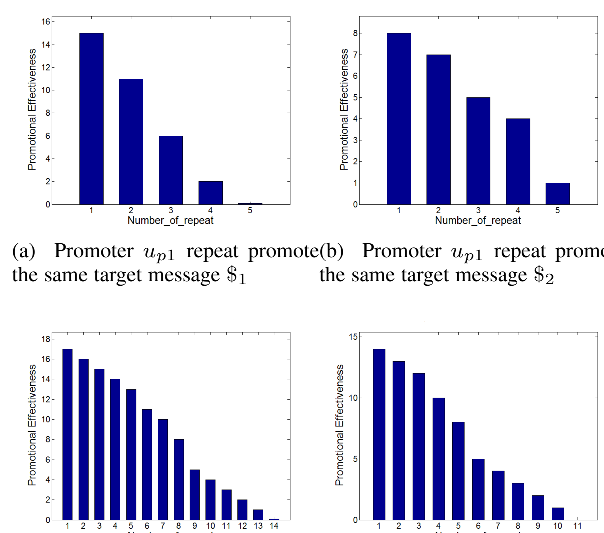
(a) ln(num-of-followers-of-u_{pro}) (b) $num\text{-}of\text{-}repeat$ (c) $length\text{-}of\text{-}comment$

(d) ratio-of-female-followers-of-u_{pro} (e) $ln(interval\text{-}after\text{-}the\text{-}former)$ (f) $ln(interval\text{-}before\text{-}the\text{-}next)$

Fig. 5: *Demonstration of selection bias reducing of six covariates with setting $user\text{-}active\text{-}time$ as the treatment.* The Q-Q plot which is closely aligned with $y = x$ (red dashed line) indicates that the quantile distributions between treated and untreated objects are similar. The distributions between treated objects and untreated objects without PSM (green circle-dots) shows the existing of selection bias in confounders. We remove selection bias with PSM (blue triangle-dots) to ensure the reliability.

So $topic\text{-}interest$ or personalized decoration can work as such an effective strategy in social promotion. Third, in a propagation path, the grandchild promotion retweet (*i.e.*, the retweet of the target message's retweet) often has fewer adoptions than the child promotion retweet (*i.e.*, the retweet of the target message). Thus, we find that more $depth\text{-}in\text{-}path$ indicates weaker promotional effectiveness. The potential reason is the recipients of the grandchild may have received the same message from the child and its siblings.

Observation 2. Two critical trade-offs in the context-level strategies. We present two trade-offs for the practitioners who want to promote their target message in online social media. One is the trade-off between the value of $num\text{-}of\text{-}repeat$ and the negative influence of its growth on a specific promotional effectiveness. As we have introduced in Figure 1, the more a promoter repeats the same promoted content, the fewer adoptions he/she will harvest. However, the total number of adopted promotions is monotonic nondecreasing with the number of promotions increasing. The promoter may hope to get as many as adoptions as possible but should stop promoting when its benefit becomes zero. The other is the trade-off between the negative effect of $interval\text{-}after\text{-}the\text{-}former$ and the positive effect of $interval\text{-}before\text{-}the\text{-}next$ that a promoter has to deal with when he/she decides the posting time. From Table 6 we know that the shorter the time $interval\text{-}after\text{-}the\text{-}former$ promotion is, the more adopted message this promotion will get, because the current promotion is still at the early stage of the process. Therefore, a promoter hopes to *frequently* promote. However, he/she also hopes to *infrequently* promote the message, since the longer the time $interval\text{-}before\text{-}the\text{-}next$ promotion is, the more adopted message this promotion will get, because the current promotion could sufficiently disseminated before the next promotion appears.



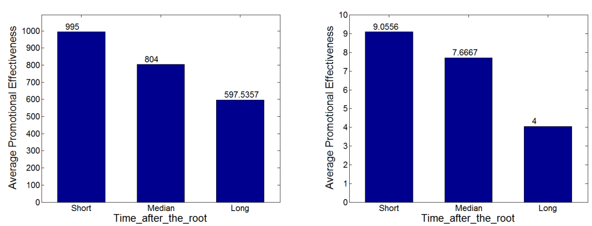
(a) Promoter u_{p1} repeat promote the same target message $\$_1$ (b) Promoter u_{p1} repeat promote the same target message $\$_2$

(c) Promoter u_{p2} repeat promote the same target message $\$_3$ (d) Promoter u_{p2} repeat promote the same target message $\$_4$

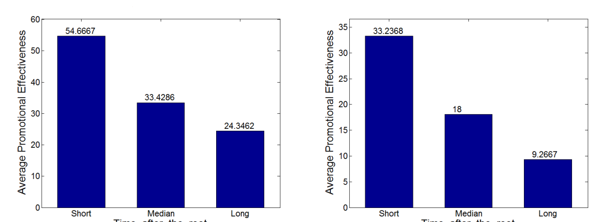
Fig. 6: *Case studies on strategy $num\text{-}of\text{-}repeats$:* The more a promoter repeats the same target message, the fewer adoptions he/she will harvest.

TABLE 6: *The effect of strategies on promotional effectiveness:* A positive (negative) value of the effect means that a higher (smaller) value of the strategy will achieve better effectiveness with standard error of the mean (SEM) in parentheses. In a paired t-test, a smaller p-value indicates high significance of the strategy: ***: $p < 0.001$, **: $0.001 \leq p < 0.01$, *: $0.01 \leq p < 0.05$, NO star: $p \geq 0.05$. Non-significant strategies are omitted for space.

	Num. followers of promoter Pct. such promotions in data	[0, 100] 36%	(100, 1,000] 14%	(1,000, 10,000] 8%	(10,000, 100,000] 11%	(100,000, +∞) 31%
Context	$^{(-)}depth\text{-}in\text{-}path$	-0.163 *** (0.018)	-0.131 *** (0.035)	-0.662 *** (0.122)	-0.175 *** (0.034)	0.086 (0.192)
	$^{(-)}num\text{-}of\text{-}repeat$	0.068 (0.047)	0.613 (0.335)	n/a	0.728 (0.839)	-0.525 *** (0.034)
	$^{(+)}user\text{-}active\text{-}time$	0.158 *** (0.010)	0.171 *** (0.031)	0.695 *** (0.138)	0.123 *** (0.012)	0.418 *** (0.052)
	$^{(-)}time\text{-}after\text{-}the\text{-}root$	0.043 (0.017)	n/a	n/a	0.010 (0.023)	-0.263 *** (0.075)
	$^{(-)}interval\text{-}after\text{-}the\text{-}former$	-0.066 (0.101)	0.068 (0.105)	n/a	-0.029 (0.141)	-0.336 *** (0.075)
	$^{(+)}interval\text{-}before\text{-}the\text{-}next$	n/a	n/a	n/a	n/a	0.558 *** (0.079)
Content	$^{(+)}length\text{-}of\text{-}comment$	0.188 *** (0.035)	0.274 *** (0.063)	6.749 *** (0.668)	-0.040 (0.023)	-0.122 (0.092)
	$^{(+)}num\text{-}of\text{-}hashtags$ "#XXX"	0.766 * (0.360)	-0.121 (0.131)	-0.685 (0.579)	-0.096 (0.082)	-0.216 (0.237)
	$^{(+)}num\text{-}of\text{-}mentions$ "@XXX"	0.171 * (0.083)	-0.184 (0.146)	-0.439 (0.385)	-0.494 (0.385)	-0.208 (0.322)
	$^{(+)}num\text{-}of\text{-}emoticons$ ":D"	0.101 ** (0.037)	0.016 (0.071)	-0.198 (0.223)	-0.008 (0.027)	0.478 *** (0.141)
	$^{(?)}num\text{-}of\text{-}questions$ "...?"	0.567 * (0.279)	0.539 (0.453)	0.874 (0.954)	-0.089 *** (0.026)	-0.246 * (0.097)
	$^{(+)}topic\text{-}interest$	1.062 *** (0.118)	1.154 *** (0.235)	3.251 *** (0.506)	0.199 *** (0.052)	0.914 *** (0.161)

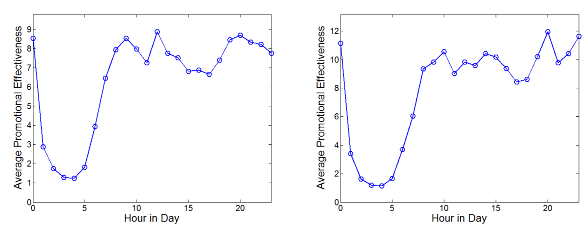


(a) Promoter u_{p1}: Average PE v.s. $time\text{-}after\text{-}the\text{-}root$ (b) Promoter u_{p2}: Average PE v.s. $time\text{-}after\text{-}the\text{-}root$

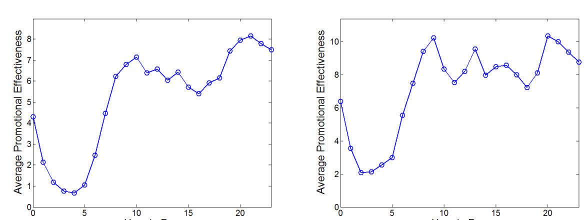


(c) Promoter u_{p3}: Average PE v.s. $time\text{-}after\text{-}the\text{-}root$ (d) Promoter u_{p4}: Average PE v.s. $time\text{-}after\text{-}the\text{-}root$

Fig. 7: *Case studies on strategy $time\text{-}after\text{-}the\text{-}root$ for popular promoters:* For the promoters who have more than 100,000 followers, their promotional effectiveness significantly affected by the context level strategy $time\text{-}after\text{-}the\text{-}root$.



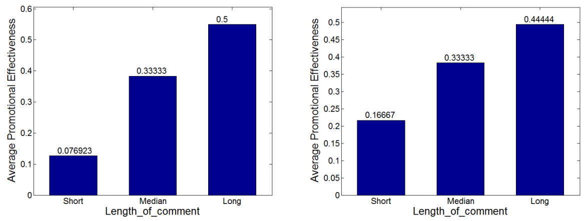
(a) Average PE varies as hour in day for promoter u_{p1} (b) Average PE varies as hour in day for promoter u_{p2}



(c) Average PE varies as hour in day for promoter u_{p3} (d) Average PE varies as hour in day for promoter u_{p4}

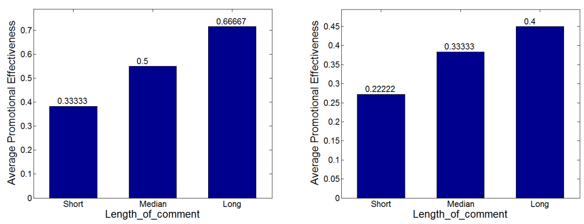
Fig. 8: *Case studies on strategy $user\text{-}active\text{-}time$ for popular promoters:* For the promoters who have more than 100,000 followers, their promotional effectiveness significantly affected by the context level strategy $user\text{-}active\text{-}time$.

Observation 3. Different promoters should focus on different promotional strategies. Specifically, the context-level strategies are significant for popular promoters, while ordinary promoters should focus on the content-level strategies. Table 6 shows that for the promoters who have more than 100,000 followers, the context-level strategies including $interval\text{-}before\text{-}the\text{-}next$ (0.558), $num\text{-}of\text{-}repeat$ (-0.525), $user\text{-}active\text{-}time$ (0.418), $interval\text{-}after\text{-}the\text{-}former$ (-0.336), and $time\text{-}after\text{-}the\text{-}root$ (-0.263) have significant effect on promotional effectiveness. However, if a promoter is not that popular, for example, if he/she has not more than 100 followers, the promoter must focus on content-level instead of context-level strategies. More appropriate decorations will be more appreciated by the recipients, such as using hashtags (0.766) to explicitly



(a) Promoter u_{p1}: Average PE v.s. $length\text{-}of\text{-}comment$

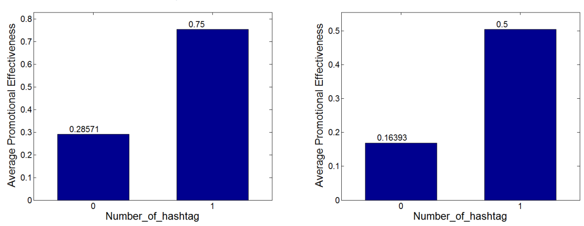
(b) Promoter u_{p2}: Average PE v.s. $length\text{-}of\text{-}comment$



(c) Promoter u_{p3}: Average PE v.s. $length\text{-}of\text{-}comment$

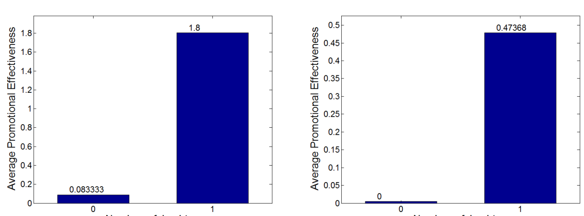
(d) Promoter u_{p4}: Average PE v.s. $length\text{-}of\text{-}comment$

Fig. 9: *Case studies on strategy $length\text{-}of\text{-}comment$ for ordinary promoters:* For the promoters who have less than 100 followers, their promotional effective increased with the length of comment.



(a) Promoter u_{p1}: Average PE v.s. $num\text{-}of\text{-}hashtag$

(b) Promoter u_{p2}: Average PE v.s. $num\text{-}of\text{-}hashtag$



(c) Promoter u_{p3}: Average PE v.s. $num\text{-}of\text{-}hashtag$

(d) Promoter u_{p4}: Average PE v.s. $num\text{-}of\text{-}hashtag$

Fig. 10: *Case studies on strategy $num\text{-}of\text{-}hashtag$ with ordinary promoters:* With the help of hashtag, the promoter who have less than 100 followers can improve their promotional effectiveness.

represent its topic, using question marks (0.567) to inspire users to respond, using longer comments (0.188) to decorate with interesting content, using mentions (0.171) to notify some users (who are even not his/her followers), and using emoticons (0.101) to make the message look sentimental, will be more appreciated by the recipients.

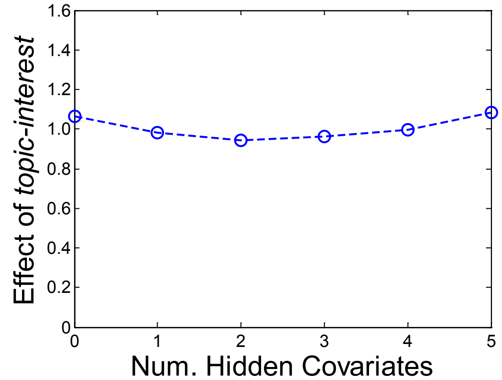


Fig. 11: The effect of strategy $topic\text{-}interest$ is stable when we randomly hide 0 to 5 covariates as unobserved.

6.3.2 Case studies on Promotional Strategies

In this section, we give some real promotion examples on five strategies, including $num\text{-}of\text{-}repeats$, $length\text{-}of\text{-}comment$, $num\text{-}of\text{-}hashtag$, $time\text{-}after\text{-}the\text{-}root$, and $user\text{-}active\text{-}time$.

From observation 2, we know that the more a promoter repeats the same target message, the fewer adoptions he/she will harvest. Figure 6 shows the real examples on strategy $num\text{-}of\text{-}repeats$. For example, in figure 6a, the promoter u_{p1} promoted the same target message $\$_1$ *five* time, but the promotional effectiveness of each promotion monotone decreased from 15 to 0.

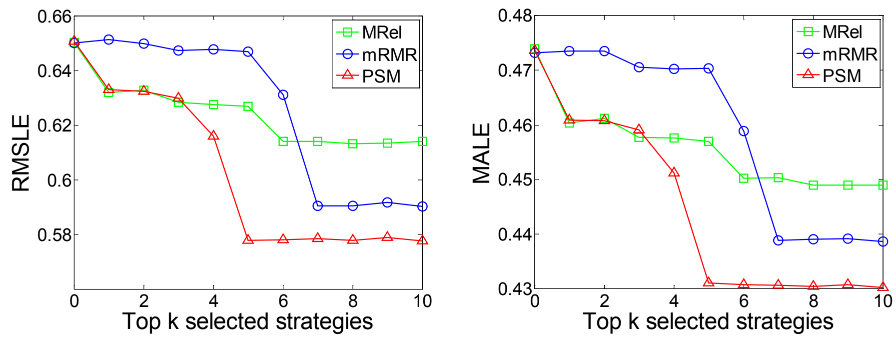
From observation 3, we find that different promoters should focus on different promotional strategies. Specifically, the popular promoters should focus on context level strategies, while the ordinary promoters should focus on content level strategies.

Figure 7 and 8 demonstrate the context level strategies, $time\text{-}after\text{-}the\text{-}root$, and $user\text{-}active\text{-}time$, are important for popular promoters with real examples. For instance, in figure 7a, the average promotional effectiveness of popular promoter u_{p1} decreased from *995* to *804* and then to *597.5* with the strategy $time\text{-}after\text{-}the\text{-}root$ changed from *short* to *median* then to *long*; and in figure 8a, for the popular promoter u_{p1}, his/her average promotional effectiveness at 12 a.m (after lunch) and 8 p.m (after dinner) when people get entertained by Internet are higher than 3 a,m or 4 a.m when people are sleeping.

Figure 9 and 10 demonstrate the content level strategies, $length\text{-}of\text{-}comment$ and $num\text{-}of\text{-}hashtag$, are important for ordinary promoter with real examples. For instance, in figure 9a, the average promotional effectiveness of ordinary promoter u_{p1} increased from 0.07 to 0.33 and then to 0.5 with the length of comment changed from *short* to *median* and then to *long*; and in figure 10a, for the ordinary promoter u_{p1}, his/she average promotional effectiveness with hashtag (0.75) are higher than without hashtag (0.29).

6.3.3 Unobserved covariates testing.

Here we test the robustness of our propensity score matching based method for unobserved covariates. Specifically, we randomly hide some covariates as unobserved. We assume that the unobserved covariates are not such relevant with treatment, since the majority of relevant covariates would be considered when we crawled the observational dataset. Figure 11 shows that the treatment effect of $topic\text{-}interest$ strategy estimated by our PSM method is stable



(a) RMSLE (b) MALE

Fig. 12: *Our PSM based method outperforms the baselines when selecting the top k effective strategies.* When $k = 0$, all the prediction algorithms use only static features; with the k value increasing, the RMSLE and MALE of PSM decrease much faster than bselines.

TABLE 7: The top ranked strategies based on static features by PSM and baselines.

Rank	PSM	mRMR	MRel
1	$topic\text{-}interest$	$depth\text{-}in\text{-}path$	$topic\text{-}interest$
2	$num\text{-}of\text{-}emoticons$	$topic\text{-}popularity$	$length\text{-}of\text{-}comment$
3	$interval\text{-}before\text{-}the\text{-}next$	$time\text{-}after\text{-}the\text{-}root$	$topic\text{-}popularity$
4	$num\text{-}of\text{-}repeat$	$num\text{-}of\text{-}URLs$	$topic\text{-}diversity$
5	$user\text{-}active\text{-}time$	$num\text{-}of\text{-}question\text{-}marks$	$topic\text{-}novelty$
6	$time\text{-}after\text{-}the\text{-}root$	$topic\text{-}interest$	$num\text{-}of\text{-}repeat$
7	$interval\text{-}after\text{-}the\text{-}former$	$user\text{-}active\text{-}time$	$interval\text{-}before\text{-}the\text{-}next$
8	$num\text{-}of\text{-}hashtags$	$num\text{-}of\text{-}exclamation\text{-}marks$	$time\text{-}after\text{-}the\text{-}root$
9	$num\text{-}of\text{-}question\text{-}marks$	$num\text{-}of\text{-}hashtags$	$interval\text{-}after\text{-}the\text{-}former$
10	$num\text{-}of\text{-}exclamation\text{-}marks$	$topic\text{-}novelty$	$depth\text{-}in\text{-}path$

when we randomly hide 0 to 5 covariates, which are not such relevant with $topic\text{-}interest$, as unobserved.

6.3.4 Effective strategy ranking and promotional effectiveness prediction

Based on the static features set, we use both baseline methods and our PSM method to rank effective promotional strategies, predict the promotional effectiveness with static features and the top k ranked effective strategies, and report the RMSLE and MALE. With our PSM, we rank all promotional strategies by their estimated effect on promotional effectiveness. Table 7 demonstrates the top ranked strategies based on static features by PSM and baselines. Figure 12 shows that our PSM method can reach a smaller error more faster than the baselines MRel and mRMR. We observe that when $k = 5$, the RMSLE value of PSM (0.578) is much smaller than those values of MRel (0.647) and mRMR (0.627), and the same with the MALE value.

Besides the winning in effective strategy ranking, we provide further analysis of experimental results in Figure 12 to emphasize the difference between PSM method and correlation-based methods. The reasons are listed below.

- **High correlation with effectiveness does not equal to huge error reduction when the strategy is also correlated with static features.** In Figure 12, the 2^{nd} selected strategy in MRel is the $length\text{-}of\text{-}comment$, which has high correlation coefficient (0.54) with promotional effectiveness, but has no error reduction on RMSLE. Since it has higher correlation coefficient (0.78) with num-of-followers-of-u_{pro}, a pre-selected static feature of promotion. The strategies effect analysis with our PSM method is able to select strategies that have higher directly effect on promotional effectiveness by controlling the other features, including the static features.
- **Correlation-based methods cannot reduce much error when meeting selection bias and unbalance.** In Figure 12, the strategy $num\text{-}of\text{-}URLs$ was ranked in the 4^{th} by mRMR method, but has no error reduction on RMSLE because of the selection bias and unbalance in our dataset: 99% promotions in our data are without URLs and mRMR method reduces redundancy between strategies by linear function, which can not completely remove the redundancy and selection bias. Our PSM method is able to eliminate the selection bias and balance data via strategies effect analysis with a data sampling step by propensity score matching. After the data sampling, the percentage of promotions with and without URLs are fifty-fifty, and the distributions of other covariate strategies are similar between the treated and untreated promotions. Thus, our PSM method concludes that the strategy $num\text{-}of\text{-}URLs$ has little effect on promotional effectiveness, and ranks it out of top-10.

In summation, it is crucial for strategies effect analysis to *reduce the selection bias* (proved in Figure 5). The correlation-based methods (*e.g.*, MRel and mRMR) could not handle the issue. But our PSM method is able to *capture the strategies effect more accuracy* by selection bias reducing in observational studies.

6.3.5 Efficiency testing

We randomly sampled different portions of our dataset, and plot Figure 13 to show the running time cost of our PSM method. The time cost has linear increase with the size of our sampled data. The total time cost of our whole dataset is less than 10,000 seconds, indicating that we need only *12 milliseconds* to learn and predict a single promotion.

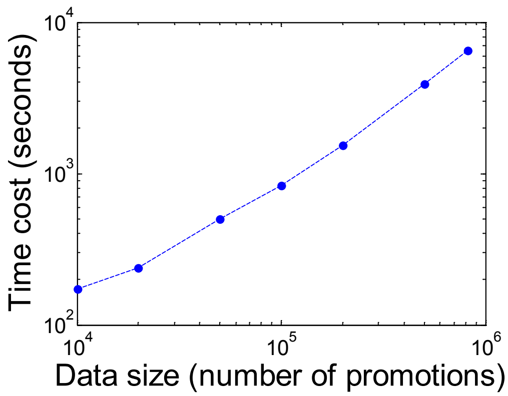


Fig. 13: *PSM method is scalable to the data size:* The running time cost is linear to the number of randomly-sampled promotions.

7 Conclusions

In this paper, we involved causal analysis to addressed a novel real-world problem that how to make strategy for high promotional effectiveness in social media from causal angle. We introduced a series of promotional strategies in both context and content levels, and presented their causal effect analysis after reducing selection bias by propensity score matching (PSM) based method in observational data. The results provided comprehensive suggestions to the practitioners (promoters) to operate (*i.e.*, when and how to promote the messages) for steering social media promotions. We conducted extensive experiments on a large social platform of over 194 million users, and demonstrated that our PSM method could find the significant effective strategies and outperform the state-of-the-art methods in promotional effectiveness prediction. Moreover, we provided three insights of making promotional strategy: (1) three significant, stable strategies, (2) two critical trade-offs, and (3) different strategies for promoters with different popularity. Our in-depth study may inspire the future of more productive promotions for products and public policies.

8 Acknowledgement

This work was supported in part by the National Program on Key Basic Research Project under Grant 2015CB352300 and in part by the National Natural Science Foundation of China under Grant 61210008, Grant U1611461, Grant 61521002, and Grant 61531006.

References

[1] S. Aral, L. Muchnik, and A. Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *PNAS*, 106(51):21544–21549, 2009.

[2] P. C. Austin. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3):399–424, 2011.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *JMLR*, 3:993–1022, 2003.

[4] D. Chan, R. Ge, O. Gershony, T. Hesterberg, and D. Lambert. Evaluating online ad campaigns in a pipeline: causal models at scale. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 7–16. ACM, 2010.

[5] M. Charkhabi. Adjustments to propensity score matching for network structures. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, pages 628–633. IEEE, 2014.

[6] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD*, pages 1029–1038, 2010.

[7] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *KDD*, pages 199–208, 2009.

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

[9] Z. Chen, K. Zhang, L. Chan, and B. Schölkopf. Causal discovery via reproducing kernel hilbert space embeddings. *Neural computation*, 26(7):1484–1517, 2014.

[10] J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec. Can cascades be predicted? In *WWW*, pages 925–936, 2014.

[11] D. R. Cox. Planning of experiments. 1958.

[12] P. Cui, S. Jin, L. Yu, F. Wang, W. Zhu, and S. Yang. Cascading outbreak prediction in networks: a data-driven approach. In *KDD*, pages 901–909, 2013.

[13] P. Domingos and M. Richardson. Mining the network value of customers. In *KDD*, pages 57–66, 2001.

[14] Y. Dong, R. A. Johnson, and N. V. Chawla. Will this paper increase your h-index?: Scientific impact prediction. In *ACM WSDM*, pages 149–158, 2015.

[15] M. Gong, K. Zhang, B. Schoelkopf, D. Tao, and P. Geiger. Discovering temporal causal relations from subsampled data. In *International Conference on Machine Learning*, pages 1898–1906, 2015.

[16] A. Goyal, W. Lu, and L. V. Lakshmanan. Celf++: optimizing the greedy algorithm for influence maximization in social networks. In *WWW*, pages 47–48, 2011.

[17] C. W. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.

[18] D. N. Hill, R. Moakler, A. E. Hubbard, V. Tsemekhman, F. Provost, and K. Tsemekhman. Measuring causal impact of online actions via natural experiments: Application to display advertising. In *Proceedings of the 21th KDD International Conference on Knowledge Discovery and Data Mining*, pages 1839–1847. ACM, 2015.

[19] P. O. Hoyer, D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf. Nonlinear causal discovery with additive noise models. In *Advances in neural information processing systems*, pages 689–696, 2009.

[20] X. Hu, J. Tang, and H. Liu. Online social spammer detection. In *AAAI*, pages 59–65, 2014.

[21] X. Hu, J. Tang, Y. Zhang, and H. Liu. Social spammer detection in microblogging. In *IJCAI*, volume 13, pages 2633–2639. Citeseer, 2013.

[22] K. Imai and D. A. Van Dyk. Causal inference with general treatment regimes. *Journal of the American Statistical Association*, 99(467), 2004.

[23] K. Imai and D. A. Van Dyk. Causal inference with general treatment regimes. *Journal of the American Statistical Association*, 2012.

[24] M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, and S. Yang. Social contextual recommendation. In *ACM CIKM*, pages 45–54, 2012.

[25] M. Jiang, B. Hooi, A. Beutel, S. Yang, P. Cui, and C. Faloutsos. A general suspiciousness metric for dense blocks in multimodal data. In *IEEE ICDM*, 2015.

[26] K. Jung, W. Heo, and W. Chen. Irie: Scalable and robust influence maximization in social networks. In *2012 IEEE 12th International Conference on Data Mining*, pages 918–923. IEEE, 2012.

[27] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD*, pages 137–146, 2003.

[28] M. Kimura and K. Saito. Tractable models for information diffusion in social networks. In *PKDD*, pages 259–271. 2006.

[29] K. Kuang, P. Cui, B. Li, M. Jiang, S. Yang, and F. Wang. Treatment effect estimation with data-driven variable decomposition. In *AAAI*, pages 140–146, 2017.

[30] K. Kuang, M. Jiang, P. Cui, and S. Yang. Steering social media promotions with effective strategies. In *IEEE ICDM*, 2016.

[31] S. Kullback and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, pages 79–86, 1951.

[32] H. Li, A. Mukherjee, B. Liu, R. Kornfield, and S. Emery. Detecting campaign promoters on twitter using markov random fields. In *IEEE ICDM*, pages 290–299, 2014.

[33] C.-T. Lu, H.-H. Shuai, and P. S. Yu. Identifying your customers in social networks. In *ACM CIKM*, pages 391–400, 2014.

[34] Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos. Rise and fall patterns of information diffusion: model and implications. In *KDD*, pages 6–14, 2012.

[35] J. M. Mooij, D. Janzing, T. Heskes, and B. Schölkopf. On causal discovery with cyclic additive noise models. In *Advances in neural information processing systems*, pages 639–647, 2011.

[36] F. Morone and H. A. Makse. Influence maximization in complex networks through optimal percolation. *Nature*, 2015.

[37] S. A. Myers and J. Leskovec. The bursty dynamics of the twitter information network. In *WWW*, pages 913–924, 2014.

[38] J. Pearl. Causality: models, reasoning and inference. In *Cambridge University Press, 2nd edition*, 2009.

[39] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE TPAMI*, 27(8):1226–1238, 2005.

[40] J. Peters, J. M. Mooij, D. Janzing, B. Schölkopf, et al. Causal discovery with continuous additive noise models. *Journal of Machine Learning Research*, 15(1):2009–2053, 2014.

[41] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *KDD*, pages 61–70, 2002.

[42] M. Rojas-Carulla, M. Baroni, and D. Lopez-Paz. Causal discovery using proxy variables. *arXiv preprint arXiv:1702.07306*, 2017.

[43] P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[44] S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(Oct):2003–2030, 2006.

[45] P. Spirtes, C. N. Glymour, and R. Scheines. *Causation, prediction, and search*. MIT press, 2000.

[46] K. Torkkola. Feature extraction by non parametric mutual information maximization. *JMLR*, 3:1415–1438, 2003.

[47] P. Wang, Y. Liu, M. Meytlis, H.-Y. Tsao, J. Yang, and P. Huang. An efficient framework for online advertising effectiveness measurement and comparison. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 163–172. ACM, 2014.

[48] Y. Wang, G. Cong, G. Song, and K. Xie. Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In *KDD*, pages 1039–1048, 2010.

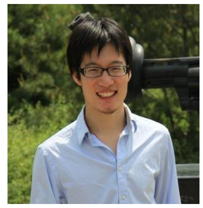
[49] J. Yang and J. Leskovec. Patterns of temporal variation in online media. In *ACM WSDM*, pages 177–186, 2011.

[50] L. Yu, P. Cui, F. Wang, C. Song, and S. Yang. From micro to macro: Uncovering and predicting information cascading process with behavioral dynamics. *arXiv preprint arXiv:1505.07193*, 2015.

[51] K. Zhang and A. Hyvärinen. On the identifiability of the post-nonlinear causal model. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, pages 647–655. AUAI Press, 2009.

Kun Kuang received the B.E. degree from the Department of Computer Science and technology of Beijing Institute of Technology in 2014. He is a third-year Ph.D. candidate in the Department of Computer Science and Technology of Tsinghua University. His main research interests including data mining, high dimensional inference and data driven causal model.

Meng Jiang received his B.E. degree and Ph.D. degree in 2010 and 2015 at the Department of Computer Science and Technology in Tsinghua University. He is now a Postdoctoral Research Associate at University of Illinois at Urbana-Champaign. His main research interests include data mining, behavior analysis and social network analysis. He visited Carnegie Mellon University from 2012 to 2013. He has over 15 published papers on social recommendation and suspicious behavior detection in top conferences/journals of the relevant field such as IEEE TKDE, SIGKDD, TKDD and AAAI. He got the best paper finalist in ACM SIGKDD 2014



Peng Cui received his Ph.D. degree in computer science in 2010 from Tsinghua University and he is an Assistant Professor at Tsinghua. He has vast research interests in data mining, multimedia processing, and social network analysis. Until now, he has published more than 60 papers in conferences such as SIGIR, AAAI, ICDM, etc. and journals such as IEEE TMM, IEEE TIP, DMKD, etc. He won 5 best paper awards in recent 4 years, including ICDM2015 Best Student Paper Award, ICME 2014 Best Paper Award, etc. In 2015, he was awarded as ACM China Rising Star. Now his research is sponsored by National Science Foundation of China, Samsung, Tencent, etc. He also serves as Guest Editor, Co-Chair, PC member, and Reviewer of several high-level international conferences, workshops, and journals

Hengliang Luo received his Ph.D. degree in Pattern Recognition in 2005 from Beijing Institute of Technology and Join Samsung since 2006. Now he is leading a big data analytic group at Samsung Beijing R&D Centre, which focus on heterogeneous interaction data mining between user and devices, includes media, activity, structure etc.

Shiqiang Yang g received his B.E. and M.E. degrees from the Department of Computer Science and Technology of Tsinghua University in 1977 and 1983, respectively. He is now a Professor at Tsinghua University. His research interests include multimedia technology and systems, video compression and streaming, content-based retrieval for multimedia information, multimedia content security, and digital right management. He has published more than 100 papers and MPEG standard proposals. Professor Yang has organized many conferences as program Chair or TPC member including PCM05, PCM06 Workshop On ACM Multimedia05, MMM06, ICME06, MMSP05, ASWC06, etc.