Graphical Modeling of Macro Behavioral Targeting in Social Networks

Yusheng Xie* Zhengzhang Chen Kunpeng Zhang Md. Mostofa Ali Patwary
Yu Cheng Haotian Liu Ankit Agrawal Alok Choudhary†

Abstract

We investigate a class of emerging online marketing challenges in social networks; macro behavioral targeting (MBT) is introduced as non-personalized broadcasting efforts to massive populations. We propose a new probabilistic graphical model for MBT. Further, a linear-time approximation method is proposed to circumvent an intractable parametric representation of user behaviors. We compare the proposed model with the existing state-of-the-art method on real datasets from social networks. Our model outperforms in all categories by comfortable margins.

1 Introduction

In the year of 2012, the total expense of US Internet display advertisement is around $12.7 billion and the number is expected to increase up to $28 billion by the end of 2017 [21]. Increased spending on online campaigns has turned the Internet, on both desktop and mobile devices, to one of the best media to place ads. In addition to conventional sponsored search [4] and sponsored story [23], Behavioral Targeting (BT) [3], which displays sponsored messages to targeted user segments by algorithmically learning users' search and browsing behaviors and interests, has become increasingly popular many data mining researchers and practitioners [1] [2] [20]. Many successful systems [3] and models have been proposed and are proven to be valuable in practice in the real world.

The very recent explosion of popular social networking websites such as Facebook and Twitter changes how people communicate on the Internet. Inevitably, this social change brings some subtle yet critical distinctions in BT on social networks. For example, on Facebook, brands make public posts to engage their fans/customers. Posting on Facebook or tweeting through Twitter is quite different from the traditional understanding of BT, but yet they target to a somewhat selected audience because only one's own Twitter followers (Facebook fans) have a chance to see one's tweets (Facebook posts). On the other hand, they are also different from traditional front-page-based advertising (e.g., MSN home page).

Based on the distinctions between traditional BT and BT in social networks, we introduce the term Macro Behavioral Targeting (MBT) to refer to BT in social networks. More formally, we define MBT as non-personalized broadcasting efforts that appeal to a massive targeted interactive population under competition from rivals for limited influence over the same population. Although MBT is different from traditional BT by non-personalized and broad targeting, such difference alone is not sufficient to motivate a separate study on MBT. After all, targeting at coarse granularity, like individual BT, is already well studied [2] [19]. For example, [2] looks at Yahoo! Homepage today's module as a means of broadcasting and macro targeting. We identify two other aspects, as discussed below, that make MBT a different problem from traditional BT and motivate this study.

First, the competition structure in MBT is different from traditional BT. MBT campaign starters (e.g., Coca-Cola, Pepsi) are competing for popularity on a particular platform (e.g., Facebook) with a lot of visibility of the fan's reaction to each brand's public campaigns. This scenario is different from traditional BT competing models. For example, consider two brands bidding on Google keywords for a particular demographic; there is no way for either of them to know how its opponent's campaigns are running, although their audiences overlap. However, if two brands run MBT campaigns through Facebook, most information about their campaigns (including how many comments / likes are received, what messages the efforts used, etc.) would just be public information. How to handle open competition becomes the key for MBT success.

In addition to its open competitiveness, viral effect is another trademark of MBT. Unlike traditional advertising environment, where interaction is limited to end user and advertiser without any inter-user interaction, MBT has a different story. Since in MBT the end users form an interconnected network, behavioral marketers could for the first time exploit inter-user interactions.

*Corresponding author: yxi389@eecs.northwestern.edu
†All authors are affiliated with EECS, Northwestern University, Evanston, IL 60208.

For example, if on Facebook Mike likes an MBT effort posted by a particular brand, the update "Mike likes this brand" would come up in the feed of Mike's friends. Even if the friends do not receive the original MBT effort and are not fans of the posting brand, they are still reached by this effort through Mike's influence like how virus infects.

Our contribution We identify and define a niche of Behavioral Targeting problems, called MBT, which differs from the traditional BT mechanism by its open competitiveness and viral marketing effect. A probabilistic graphical model is designed and implemented for the MBT problem that accounts for its aforementioned unique aspects. In the model, we construct an accurate parametrization for user behaviors, which goes beyond simple Poisson event modeling. In the experiments section, we compare the proposed model with a state-of-the-art method and show a significant improvement on real datasets from Facebook.

2 Related work

A close work to this paper is [16], where the authors try to predict the popularity of each brand post based on historical data. [16] represents brand posts as feature vectors. With enough historical posts as training data, an SVM classifier [13] is used to predict the success of each brand post. In fact, a similar implementation is introduced in our experiments for comparison to our proposed model in feedback prediction. Despite its simplification of important MBT's characteristics like viral effect, which has been well studied in earlier work like [7], the approach of [16] is simple and effective.

However, a critical problem in [16] is its absence of any user model [24]. In fact, the approach taken in [16] is advertiser-centric and is agnostic of the users. This one-sided simplification not only limits the generalizability, but also deviates from classical approaches to model BT problems [3] [1]. More generally, deciding on the detailed level at which the model should represent the users in MBT is a discretionary problem. Most BT models, which are built by the behavioral marketers themselves, with full access to user activity patterns, can allow complicated structure on user modeling as long as the computation is feasible. But MBT models often deal with public or incomplete data as implicit assumptions. The obvious way to handle this discretionary problem is to ignore the users, which is what [16] has done, at the cost of throwing away valuable user information. In response, an idea of "parameterize-and-approximate" is presented in the later part of this paper.

A much richer literature is available on traditional BT analysis and optimization. Several different estimating approaches are recently proposed with some common merits including proven real-world performance and scalability [3] [1] [20]. Linear Poisson regression is often favored [3] [20] for modeling behavioral events, given that the Poisson distribution is the intuitive statistical model for counting data: $p(y) = \lambda^y \exp(-\lambda)/y!$, where $\lambda = \mathbf{w}^T\mathbf{x}$. The Poisson parameter λ is determined by the inner product of a weight vector and a feature vector. [1] tries to solve the problem from a different angle by modeling users' interests as dynamic topics. The authors propose a time-varying user model based on Latent Semantic Indexing [14], Latent Dirichlet Allocation (LDA) [8] and Dynamic Topic Model (DTM) [9].

However, the modeling in both [3] [20] are less than ideal for MBT scenarios. For example, the single parametric Poisson model introduced in [3] makes assumption on homogeneous user clicks and therefore becomes inadequate in the MBT setting where heterogeneity is essential. The topic model in [1] is designed to be user-centric. If it were applied in MBT context, it would end up with an infeasible situation where the number of topics is orders of magnitude larger than the number of documents.

3 Methodology

In this section, we discuss the proposed methodology for solving the MBT problems presented above. We first motivate and describe a parametrization of user behaviors and the feature extraction of MBT elements in Section 3.1 and 3.2. Then a full parametrization of the MBT flow in terms of random variables is presented as a graphical model in Section 3.3. In Section 3.4, we implement the graphical model from Section 3.3 by fully specifying all necessary prior distributions. Finally, we present inference and query-answering on the proposed model in Section 3.5, where we further propose a linear-order-in-time approximation for user behavior parametrization to achieve computational feasibility.

3.1 Parametrizing user behavior Each user's feedback function is modeled by a parametric distribution with different parameters. The feedback function for the ith user X_i should follow the Gamma distribution $\Gamma(\beta_{1i}, \beta_{2i})$ since the Gamma distribution is often the default distribution for modeling wait time until the occurrence of an event [5] [6]. Note that by denoting X_i, each user is assumed to have the same distribution across efforts from different brands of different categories. In other words, this user parametrization is two-fold: 1) each X_i is across all topics and brands, which the user is a fan of; 2) all X_i's are of the same parametric family. Neither the shape parameter β_{1i} nor the rate parameter β_{2i} is fixed.

The choice on the prior distributions of the two parameters is a delicate matter [12]. This user parametrization only dictates that the wait time before a user's feedback follows a particular family of distributions; no constraint is put on the content of the feedback itself. This assumption is in many aspects an improvement over single parametric Poisson because it relaxes the homogeneous assumption [3], which is made implicitly by the Poisson model about the BT events.

Table 1: Feature vector for effort variables.

Feature	Notes (example)
brand_id	Pepsi
hour_of_day	3pm
day_of_week	Sunday
time_since_last	the number of hours passed since the last MBT effort of this brand was launched.
type_of_effort	photo, video, url, etc.
ask_to_like ask_to_comment ask_to_share	true if this effort appeals to fans for like/comment/share
is_a_question	true if contains a question in message
long_text	true if text is over 140 characters
[positive_list] [negative_list] [characteristic_list]	[List of Boolean values; true if word is in this effort message]

Table 2: Label vector for effort variables.

Dynamical Label	Volume Label
10min_d_percentage	10min_v_percentile
1hr_d_percentage	1hr_v_percentile
8hr_d_percentage	8hr_v_percentile
24hr_d_percentage	24hr_v_percentile
48hr_d_percentage	48hr_v_percentile
7day_d_percentage	7day_v_percentile

Table 3: Feature vector for individual feedback.

user_id	each user_id corresponds to a prior Gamma distribution according to the user parametrization.
time_since_last	the wait time in minutes since an effort is launched until user makes her feedback
feedback_sentiment	sentiment label for the feedback message.
[positive_list / negative_list]	[true if word is used in feedback text]

3.2 Extracting features and labels Efforts and feedback are two important elements in MBT. In this section, we describe how to represent these two elements as feature vectors.

Efforts: MBT efforts are represented as feature vectors, whose components are shown in Table 1. Most entries in Table 1 is explained with the aid of side notes, but positive_list, negative_list, and characteristic_list are worth more explanation. The positive/negative_list each contains a list of positive/negative adjectives. And each adjective is associated with a numeric score, a score that represents the word's sentiment and varies between between 5 and -5, with 5 being the most positive and -5 the most negative [10] [18] [17]. Then, for each word in the MBT effort message, the word is marked and its score is calculated if it is present in either list. After checking wall words from the effort message, a positive_score and a negative_score are generated by summing over the scores of all marked adjectives in both lists. The effort feature vector uses these two scores with other feature elements from Table 1. Similarly, it is done for characteristic_list, which contains a list of words or phrases that would help characterize an MBT effort. For example, words like "sale" and "promotion" are included because they can be effective indicators that MBT efforts including such words are related to brand's promotional events.

In addition to feature vector, efforts also have a label vector. We are not only interested in the eventual steady state of an MBT effort's feedback, but also in the dynamics of how rapidly the feedback is accumulated and converges over time. To address these needs, a label vector for each effort is used to characterize its feedback dynamics, whose elements are presented in Table 2. A Dynamical Label for an MBT effort, say *1hr_d_percentage*, is simply the fraction of the amount of feedback received by this effort in 1 hour over the total amount of feedback this effort would ever receive in lifetime. On the other hand, a Volume Label, say *10min_v_percentile*, is simply the percentile rank of this effort among all efforts by a particular brand in terms of feedback received within 10 minutes after the release of each effort. 6 Volume Labels characterize how much feedback an MBT effort garners; and 6 Dynamic Labels characterize how rapidly it does so.

Feedback: The feature representation for feedback is shown in Table 3, which also provides explanatory notes for each of the feature elements.

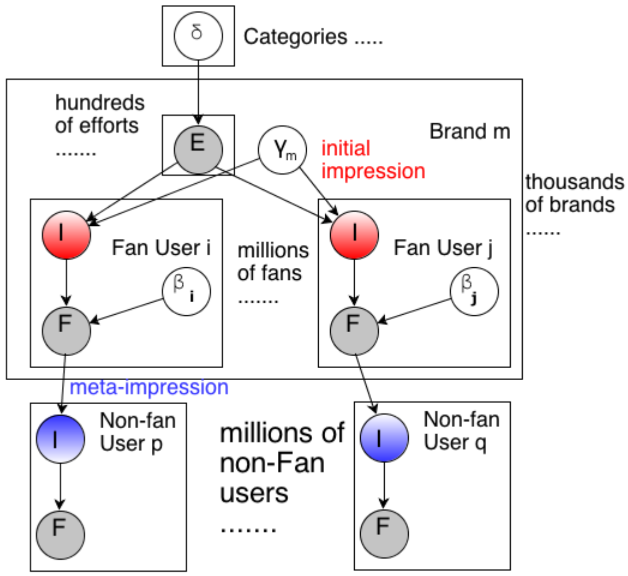

Figure 1: Graphical model for MBT. Directed arrows denote variable dependencies. Shadow nodes represent observable variables.

3.3 Parametrizing MBT flow After parametrizing the user behavior and extracting features and labels, we build a graphical model as shown in Figure 1. The model presented in Figure 1 tries to parametrize the conceptual flow of a complete MBT instance. When a brand starts an MBT instance to outreach, it needs to broadcast an effort to its fans or followers, which could be interactive messages, promotions, etc. Then the effort will appear to a certain number of its fans, which is called "initial impression" of this MBT effort. Having received the MBT effort, some fans would reply with his or her feedback as a "like" or a message, etc. When a user does so, her friends would automatically receive this fact and therefore see the MBT effort even if they are not originally fans of the brand. Such impressions are called "meta-impression" to differentiate from initial impression.

Effort and effort prior variables: Variable E in Figure 1 represents an MBT effort, whose representation is found in Table 1. Brands from different categories should have different effort priors, so an effort prior distribution δ is used to capture the inter-category distinction. In implementation, δ, given category, provides a prior distribution for each effort feature defined in Table 1. As a result, δ defines the difference among categories. For example, brands from retailer category would be more likely to include promotional messages in their MBT efforts than news / media category.

Impression and affinity prior variables: Variable I represents initial impression of each user. In implementation, I (each $I_{i,e}$) is a single probability of user i seeing an MBT effort e. Since potential impression from different brands varies a lot (e.g. Coca-cola vs. a local bakery), each brand's impression variable has a different affinity prior γ. In implementation, γ (each $\gamma_{i,m}$) provides a single probability of user i seeing any MBT effort from brand m as *initial impression* (red in Figure 1). γ is very sparse since each $\gamma_{i,m}$ is non-zero if and only if user i is a fan of Brand m. Both I and γ represent the probability of a user seeing an effort, but γ serves as a prior to I. On the other hand, I models the likelihood a user likes the contents of an effort, while γ models how likely a user likes a brand's effort in general.

User feedback and user prior variables: Variable F represents feedback from each user, which is modeled by a feature vector defined in Table 3. In implementation, F holds the probability distribution for each feedback feature defined in Table 3. Each user has different propensity to interact with MBT effort, depending on: how social she is, how much time she spends on the Internet, etc. Assigning each user a prior β, which is obtained from this user's past activities, can capture such inter-user differences. And in implementation, β defines a prior distribution on each feedback feature in Table 3. Finally, *meta-impression* is modeled through fan users' influence on non-fan users. The dependencies between fan users and non-fan users, on networks like Facebook, are essentially friends relationship links, which are usually not available in public datasets. As a workaround, uniform meta-impression is assumed. In other words, every non-fan user has equal probability of being "meta-impressed".

With all necessary variables defined, our goal can be translated into the model's language. From the viewpoint of a human manager of MBT campaigns, E is the only variable under her control and she can only evaluate the performance of her MBT campaign by looking at the variables F. $\gamma_{i,m}$ encodes the consideration for open competitiveness, one of the key aspects that identifies MBT. So considering the competition from rival brands, the modeler's goal thus becomes to find the value E' for variable E, at which the probability of having maximum feedback is maximized.

$$E' = \arg\max_{E} \left(\Pr\left(F_{max}|E, I, \delta, \gamma, \beta\right)\right) \tag{3.1}$$

Exploiting the conditional independencies gives the following equivalent objective:

$$E' = \arg\max_{E} \left(\Pr\left(F_{max}|I, \beta\right) \Pr\left(I|E, \gamma\right) \Pr\left(E|\delta\right)\right) \tag{3.2}$$

3.4 Estimating prior knowledge Acquiring reliable prior information on β, γ, and δ is crucial to the final results since our model heavily relies on the three prior nodes in Figure 1.

Estimating β: Let X_i denotes user i's reaction time to give any of her feedback after an effort is made. Then the user parametrization gives $X_i \sim \Gamma(\beta_{1i}, \beta_{2i})$; X_i and X_j are also statistically independent for any $i \neq j$. Let $x_1, ..., x_{N_i}$ be user i's historical data, where N_i is the number of total MBT efforts this user has responded to and x_1 is this user's reaction time to effort 1. Then the parameters can be estimated by maximum likelihood estimation as below. Note that $\hat{\beta}_{1i}$ is the shape parameter, and $\hat{\beta}_{2i}$ is the scale parameter. $\hat{\beta}_{1i} = k$ such that: $\ln(k) - \Gamma'(k)/\Gamma'(k) = \ln\left(\sum_{i=1}^{N_i} x_i/N_i\right) - \sum_{i=1}^{N_i} \ln(x_i)/N_i$, and $\hat{\beta}_{2i} = \sum_{i=1}^{N_i} x_i/\hat{\beta}_{1i}$. The solution for $\hat{\beta}_{1i}$ is implicit in terms of k, so in practice the following popular approximation is used [15]: $\hat{\beta}_{1i} \approx 3 - s + \sqrt{(s-3)^2 + 24s}/12s$, where $s = \ln\left(\sum_{i=1}^{N_i} x_i/N_i\right) - \sum_{i=1}^{N_i} \ln(x_i)/N_i$ Thus, the prior β in our graphical model (see Figure 1) can be estimated.

Estimating γ: Now there is, for each i, $\hat{X}_i \sim \Gamma(\hat{\beta}_{1i}, \hat{\beta}_{2i})$. For each effort, let $\hat{X}_1, ..., \hat{X}_n$ denote the reaction time from n users to whom this effort reaches. n, the impression of an effort, is unknown and is often very difficult to estimate (therefore it is formulated as a latent variable in Figure 1). For example, on Facebook, n is determined by an algorithm called EdgeRank[11]; Given that, to the best of our knowledge, no previous studies on simulating EdgeRank algorithm results exist, n can be simply estimated by $\hat{n} = c|FAN_m|$ for brand m, where c is an unknown nuisance parameter and is assumed to be constant through all brands, and $|FAN_m|$ is the number of fans brand m has. In other words, for brand m and its fan user i and user j, $\gamma_{i,m} = \gamma_{j,m} \propto |FAN_m|$. This accounts for the prior γ for each brand from Figure 1 and allows us to estimate n with $\hat{n}$.

Estimating δ: This process is similar to that of estimating β, except that, unlike β, δ is not assumed to have a parametric distribution. Therefore, δ for each category is basically represented as a set of discrete histograms built from aggregated statistics, each of which corresponds to one of the features defined in Table 1.

So far, all prior nodes in Figure 1 are fully specified.

3.5 Model inference In MBT, one is often interested in knowing or even predicting how much feedback a particular effort can gather, and how quickly they gather. To answer such questions, we consider $\hat{X}_{(1)}, ..., \hat{X}_{(\hat{n})}$, the order statistics for $\hat{X}_1, ..., \hat{X}_{\hat{n}}$. Then for a fixed effort, the probability of having feedback from, say, at least y users in at most t seconds can be readily written as

(3.3)
$$\begin{aligned}
&\Pr(T \le t, Y \ge y) = \Pr(X_{(Y)} \le t, Y \ge y) \\
&= \sum_{i=y}^{\hat{n}} \Pr(X_{(Y)} \le t, Y = i) = \sum_{i=y}^{\hat{n}} \Pr(X_{(i)} \le t, X_{(i+1)} > t) \\
&= \sum_{i=y}^{\hat{n}} \Pr(X_{(i)} \le t) \Pr(X_{(i+1)} > t).
\end{aligned}$$

And similarly,

(3.4)
$$\Pr(T \le t, Y \le y) = \sum_{i=1}^{y} \Pr(X_{(i)} \le t) \Pr(X_{(i+1)} > t).$$

Using Equation 3.3, we can answer most typical questions one may ask in MBT, namely, how much feedback my campaign would gather on the social network in a certain time frame; while Equation 3.4 provides a way to estimate the entire joint density function of T and Y, which fulfills the effort prior δ from Figure 1.

Although from Equation 3.4 the distribution function of $\Pr(T, Y)$ can be recovered, $\Pr(T, Y)$ may not be computationally feasible. It is possible to solve Equations 3.3 and 3.4 but due to the large $\hat{n}$ and the non-iid nature of $\hat{X}_1, ..., \hat{X}_{\hat{n}}$, the analytical expression for $\Pr(T, Y)$ would be intractable in length. So instead, $\Pr(T, Y)$ must be rapidly approximated. Since each $\hat{X}_{(i)}$ is independent and has different parameters, the density function for $\hat{X}_{(i)}$ according to the definition of order statistics can be expressed analytically [12]:

(3.5)
$$\hat{X}_{(i)} \sim \sum_{\sigma \in \hat{\Sigma}} \left\{ \prod_{j=1}^{i-1} \left[F_{\sigma(j)}(t)\right] \times f_{\sigma(j)}(t) \times \prod_{k=i+1}^{\hat{n}} \left[1 - F_{\sigma(k)}(t)\right] \right\},$$

where $\hat{\Sigma}$ is the set of all permutations of $\hat{X}_1, ..., \hat{X}_{\hat{n}}$, $\sigma(j)$ denotes the jth random variable in a particular permutation σ, $F_{\sigma(j)}$ is the cumulative distribution function for the random variable identified by $\sigma(j)$, and similarly $f_{\sigma(j)}$ is the density function for the random variable identified by $\sigma(i)$. In our case, the density functions and distribution functions are all Gamma functions with different parameters. Equation 3.5 certainly provides a mechanism to finding the order statistics and therefore completes our quest. But Equation 3.5 yields intractable calculations due to the size of permutation set, which is irreducible because of the heterogeneity in $\hat{X}_{(1)}, ..., \hat{X}_{(\hat{n})}$. However, we empirically find

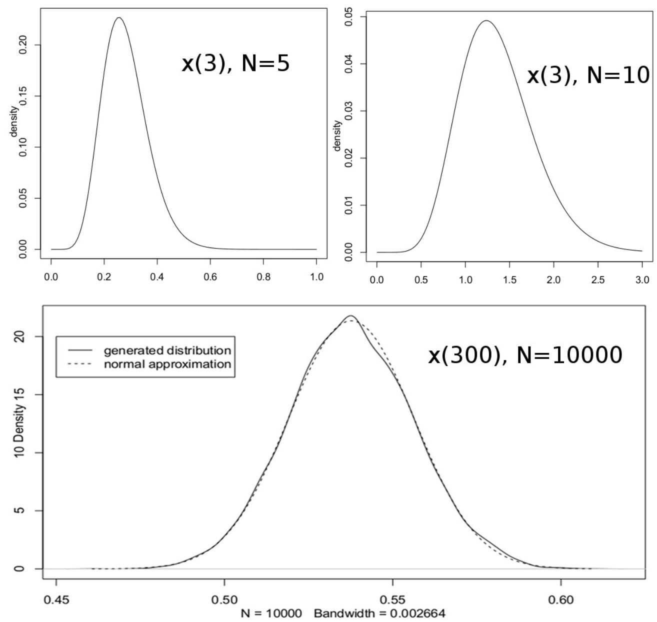

Figure 2: Order statistic densities empirically converge to normal approximation at large N.

out that for large $\hat{n}$, $\hat{X}_{(i)}$ can be conveniently approximated. Figure 2 illustrates this idea of approximation. Figure 2 first plots the distribution of $X_{(3)}$, with each $X_{(i)}, i = 1, 2, 3, ..., N$, being Gamma, at both $N = 5$ and $N = 10$; then we plot $X_{(300)}$ when $N = 10,000$, which is compared with its Normal approximation in dashed curve. Figure 2 suggests that order statistics of Gamma distributions at large N values can be approximated by corresponding Normal densities. In other words, the estimator $\hat{X}_{(i)}$ is approximated by $\tilde{X}_{(i)}$, where

$$\tilde{X}_{(i)} \sim \mathcal{N}\left(\mathbf{E}\left[\hat{X}_{(i)}\right], s^2\right), \text{ where} \tag{3.6}$$

$s^2 = \sum_{j=1}^{\hat{n}}(\mathbf{E}[\hat{X}_{(i)}] - \mathbf{E}_K[\mathbf{E}[\hat{X}_{(j)}]|K = j])/(\hat{n} - 1) = \sum_{j=1}^{\hat{n}}(\mathbf{E}[\hat{X}_i] - \mathbf{E}_K[\mathbf{E}[\hat{X}_j]|K = j])/(\hat{n}-1)$. Both $\mathbf{E}[\hat{X}_{(i)}]$ and s^2 are straightforward to compute. This approximation reduces the computational cost from $O(n!)$ to $O(n)$.

4 Experiments

In this section, we evaluate our MBT model on two important applications: predicting the performance of MBT efforts and discovering effective MBT contents. Facebook datasets are used in our experiments, and goodness-of-fit tests are performed to validate our model.

Table 4: Facebook dataset basic statistics.

Category	Electronics	Food beverages	Product service	Cars
Brands	34	346	614	91
Efforts	19.4K	169.4K	424.6K	46.2K
Feedback	1.49M	13.8M	21.0M	2.24M
Total fans	2.45M	5.42M	28.3M	13.3M
Active fans (%)	353K 14.4%	410K 7.51%	1.02M 3.61%	721K 5.40%
Feedback _per_fan	4.20 (0.605)	33.7 (2.53)	20.5 (0.741)	3.12 (0.168)
avg_wait (minutes)	176.0	106.1	234.3	222.9
avg_ sentiment	0.284	0.309	0.298	0.521
feedback_ half_life (minutes)	41.6	47.0	72.1	71.9

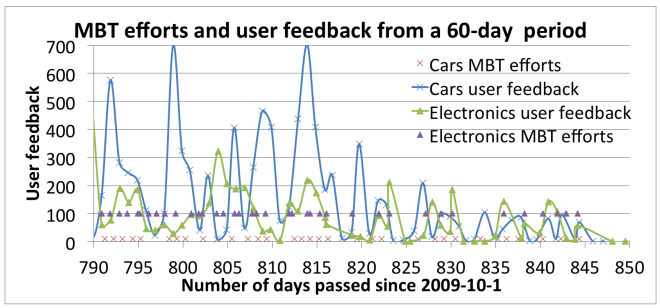

Figure 3: Effort vs feedback.

4.1 Datasets and model validation. Table 4 summarizes some statistical properties of our datasets. The datasets include over 1000 Facebook brand walls from four categories with different volumes of interactions with fans. In Table 4, "Feedback_per_fan" measures the average amount of feedback by an active fan while the number in parentheses is the average amount of feedback by a fan, active or inactive; "avg_wait" is the average number of minutes for a user to respond to an MBT effort with his/her feedback; "avg_sentiment" is the percentage number, averaged over all efforts, of feedback, whose sentiment is identified as positive, among all feedback; "feedback_half_life" is the amount of time needed for an effort, averaged over all efforts, to garner 50% of all feedback that it would eventually collect. Figure 3 substantially reveals the dynamics in MBT and what it reveals corroborate our model. Fan activities are strongly tied to MBT efforts in a particular, repeatable, even mechanical kind of dynamics. For example, the

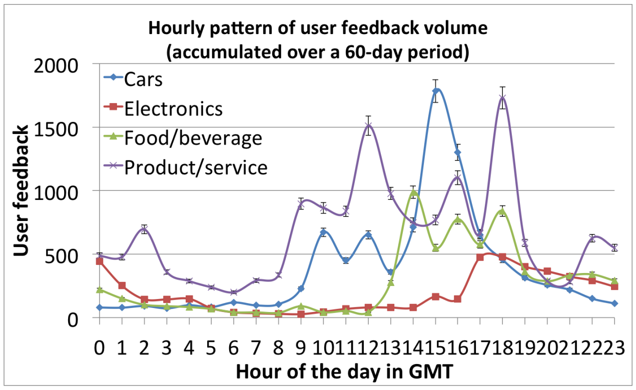

Figure 4: Hourly feedback distribution.

red crosses (Cars MBT efforts) and the responding blue curve (Cars feedback) between day840 and day845 confirm the outreach-and-feedback model. In addition, the five red crosses around day815 confirms the conjecture about the linear additivity of the appeal of consecutive MBT efforts, which is part of our proposed model.

Furthermore, the Day-Of-Week (DOW) pattern in daily feedback volume change, which is often considered essential in coarse-granularity targeting, is not clearly shown in Figure 3. In previous works on coarse-granularity targeting like [2], the DOW pattern clearly indicates repeatedly and significantly more active user feedback on, say, Friday and Sunday, over several weeks; but a similar pattern is not found in our datasets. Likewise, neither the Hour-Of-Day (HOD) pattern is clear from Figure 4. The argument here is not against the existence of DOW/HOD pattern. On the contrary, they are quite important for us (after all, both patterns are included in the feature vectors for effort in Table 1). With the goodness-of-fit tests on the feedback distributions presented in Figure 3 and Figure 4, we claim that the DOW/HOD influence is expressive through the placements of MBT effort and is not directly onto the user feedback. In other words, DOW / HOD influence and user feedback are conditionally independent once given the MBT efforts, to which all feedback is tied to. The details of goodness-of-fit tests performed are presented in the supplementary file. [1]

4.2 Predicting MBT effort performance Here, we compare our model to previously state-of-the-art results [16] on predicting the performance of MBT efforts.

Baseline approach: The baseline method chosen is based on what is presented in [16]. Our implementation basically employs a classifier that considers effort features defined in Table 1 and classifies on the effort labels defined in Table 3. Each of the 12 labels from Table 3 are trained and classified independently into one of the 5 categories: "very less" (10% of sample maximum) / "less" (30% of sample maximum) / "mediocre" (50% of sample maximum) / "high" (70% of sample maximum)/ "every high" (90% of sample maximum) amount of attention received. Then the numerical amount of feedback, based on which category the effort is classified into, is taken as the median of the amount of feedback of the efforts that are classified into this category. In order to obtain the 12 labels on the training data, MBT feedback timestamps are taken into consideration. But the feedback authorship,feedback sentiment information are ignored in this approach. The baseline classifier, in addition to lifetime feedback prediction, can also produce dynamic prediction of feedback popularity: simply predict on incremental time intervals (e.g., initial hour, initial 3 hours, initial 8 hours, etc.).

Evaluation methodology: Figure 5 compares the performance of the proposed model and the baseline approach. Figure 5 shows that, in all four categories, our prediction outperforms the previously proposed state-of-the-art approach presented in [16] by various margins; our prediction curves maintain better accuracies, in terms of feedback volume, over a longer period of time and can more precisely model the dynamics over time as well. The continuous prediction from our model is quantized into the same granularity as the baseline prediction and their regression performance is compared in Table 5. In most cases, the predictive R^2 of our model is significantly better than that of the state-of-the-art approach. The advantage of our method mainly comes from direct specification of the dependencies in the MBT dynamics (captured in our graphical model) and a wider consideration of MBT features, which include high level semantic sentiments.

4.3 Discovering effective MBT contents In addition to the ability to predict the performance of MBT efforts, it is equally interesting and more practical if the model produces more effective MBT efforts by automatically optimizing contents. Textual message is probably the foremost and the most natural type of contents that can be optimized. Thanks to our general modeling, solving the objective proposed in Equation 2.3 can optimize the contents. Table 6 summarizes some of the most interesting findings. For each of the four categories, Table 6 records the phrases that are most correlated to the popularity of its mentioning MBT efforts, using the correlation measure defined below.

[1] http://cucis.ece.northwestern.edu/publications/pdf/sdm2013_mbt_supp.pdf

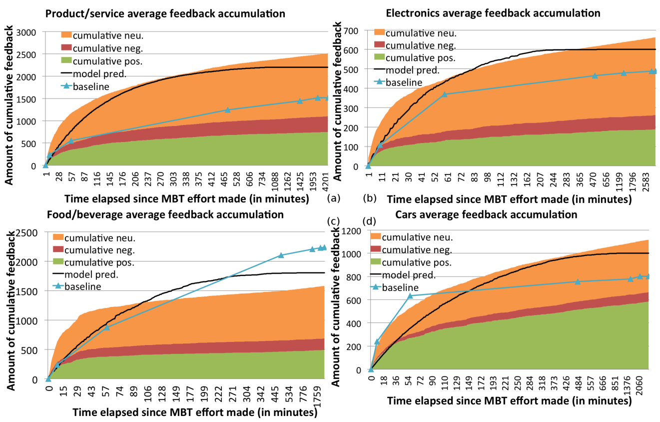

Figure 5: Cumulative feedback distribution and model / baseline predictions.

Table 5: Model / baseline prediction performance.

Category	Period	Our Model R^2	Baseline R^2
Product/service	10min	0.7624	0.3542
	1hr	0.7614	0.3356
	8hr	0.7629	0.3440
	24hr	0.7583	0.3522
	48hr	0.7549	0.3495
	7day	0.7632	0.3508
Electronics	10min	0.8189	0.5416
	1hr	0.8133	0.5290
	8hr	0.8041	0.5088
	24hr	0.7867	0.4726
	48hr	0.7882	0.4862
	7day	0.7643	0.4987
Food/beverage	10min	0.7428	0.6168
	1hr	0.7363	0.6583
	8hr	0.7355	0.7144
	24hr	0.7371	0.7199
	48hr	0.7165	0.6861
	7day	0.7309	0.6671
Cars	10min	0.7537	0.4182
	1hr	0.7648	0.4116
	8hr	0.7522	0.4353
	24hr	0.7679	0.4544
	48hr	0.7911	0.4381
	7day	0.8498	0.5161

Correlation Let O be the set of all MBT efforts in a category. Suppose there are m MBT efforts mentioning a phrase p and let M be the subset of O that contains all efforts containing the phrase p. Now Let N be the subset of O that contains the most popular efforts in terms of feedback they received. The pre-normalized correlation is simply $|N \cap M|/|N|$. Then the correlations are renormalized by setting the largest pre-normalized correlation to 1 and other correlations will be inflated accordingly. The phrases are not discovered by sorting them in descending order based on correlation; instead, they are obtained by solving Equation 2.3. Correlation is used to quantify the results in a solid and visual way.

A few remarks about our findings:

- N-gram phrases with $N = 1, 2, 3$, instead of simple 1-gram, are considered in experiment.
- Non-English entries suggest the multilingual capability of our model.
- Some terms (e.g. *favorite*, *free*, *share*, etc.) appear at the top across multiple categories while other terms (e.g. Product/service's *movie*, Cars' *sport*, etc.) are more domain-specific.

With an incremental amount of work, similar correlation results can be extracted on, not only text features, but also other dimensions including HOD, DOW, multimedia features, etc. [22].

5 Conclusion and future work

In this paper, an emerging class of MBT online marketing challenges in social networks is introduced. MBT differs from traditional targeting mechanism by two of its distinctions: open competitiveness and viral marketing effect. Unlike most previous related works, a two-parametric Gamma assumption is used to model user behavior. A linear-time approximation is devised and applied to the user behavior distributions, whose order statistics would otherwise be exponentially complex in computing time. A probabilistic model, based on the parametric assumption of user behaviors, is proposed to incorporate the two aforementioned distinctions of MBT. Experimental results show advantages of the proposed model over the state-of-the-art approach in predicting MBT behaviors. An interesting direction to further explore would be incorporating network-based model and weaker assumptions for the viral effect of MBT influence.

6 Acknowledgements

This work is supported in part by NSF award numbers CCF-0833131, CNS-0830927, IIS- 0905205, CCF-0938000, CCF-1029166, and OCI-1144061, and in part by DOE grants DE-FG02-08ER25848, DESC0001283, DE-SC0005309, DESC0005340, and DESC0007456.

References

Table 6: Phrases most correlated to effort popularity.

Category	Top Phrase	Correlation
Product/service	favorite	1.000
	friends	0.878
	app, apps	0.798
	win	0.704
	movie	0.680
	play	0.671
	share	0.642
	choice	0.579
	win	0.501
Electronics	free	1.000
	app, apps	0.827
	forever	0.749
	win	0.688
	share	0.612
	future	0.589
	amazing	0.571
	facebook	0.550
	삼성 (Samsung)	0.549
Food/beverage	free	1.000
	win	0.911
	like, likes	0.903
	find out	0.797
	cerveza (beer)	0.784
	morning	0.774
	yum, yummy	0.765
	cookies	0.763
	sour cream	0.754
Cars	favorite	1.000
	share	0.891
	sport	0.860
	dream	0.806
	summer	0.762
	bring back	0.599
	check	0.559
	leather	0.548
	heute (today)	0.463

[1] A. Ahmed, Y. Low, M. Aly, V. Josifovski, and A. J. Smola, *Scalable distributed inference of dynamic user interests for behavioral targeting*, KDD '11. ACM, 2011

[2] D. Agarwal, B.-C. Chen, and P. Elango. *Spatio-Temporal Models for Estimating Click-through Rate*, WWW '09. ACM, 2009

[3] Y. Chen, D. Pavlov, and J. F. Canny, *Large-scale behavioral targeting*, KDD '09. ACM, 2009

[4] T. Li, N. Liu, J. Yan, G. Wang, F. Bai, and Z. Chen, *A Markov Chain Model for Integrating Behavioral Targeting into Contextual Advertising*, ADKDD '09. 2009

[5] M. Uncles, A. Ehrenberg, and K. Hammond, *Patterns of Buyer Behavior: Regularities, Models, and Extensions*, Marketing Science, Vol. 14, No. 3, Part 2 of 2: Special Issue on Empirical Generalizations in Marketing (1995), pp. G71-G78. 1995

[6] W. W. Moe, and P. S. Fader, *Capturing evolving visit behavior in clickstream data*, J. Interactive Mark. 2004

[7] D. Kempe, J. Kleinberg, and E. Tardos, *Maximizing the spread of influence through a social network*, KDD '03. ACM, 2003

[8] D. M. Blei, A. Y. Ng, and M. I. Jordan, *Latent Dirichlet allocation*, J. Mach. Learn. Res. 3 (March 2003), 993-1022

[9] D. M. Blei and J. D. Lafferty, *Dynamic topic models*, ICML '06. 2006

[10] K. Dave, S. Lawrence, and D. M. Pennock, *Review sentiment scoring via a parse-and-paraphrase paradigm*, WWW '03. 2003

[11] http://edgerank.net/

[12] G. Casella and R. L. Berger, *Statistical Inference*, 2001

[13] T. Joachims, *Training Linear SVMs in Linear Time*, KDD '06. 2006

[14] T. Hofmann, *Probabilistic Latent Semantic Indexing*, SIGIR '99. 1999

[15] S. C. Choi, and R. Wette, *Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias*, Technometrics , Vol. 11, No. 4 (Nov., 1969), pp. 683-690

[16] H. Lakkaraju, and J. Ajmera, *Attention Prediction on Social Media Brand Pages*, CIKM '11. 2011

[17] X. Hu, L. Tang, J. Tang, and H. Liu *Exploiting Social Relations for Sentiment Analysis in Microblogging*, WSDM '13. 2013

[18] K. Denecke, *Using sentiwordnet from multilingual sentiment analysis*, ICDE '08. 2008

[19] K. Lerman, and T. Hogg, *Using a model of social dynamics to predict popularity of news*, WWW '10. 2010

[20] J. Canny, S. Zhong, S. Gaffiney, C. Brower, P. Berkhin, and G. H. John, *Granular data for behavioral targeting*, U.S. Patent Application 20090006363

[21] http://techcrunch.com/2012/10/09/forrester-us-online-display-ad-spend-to-hit-12-7b-in-2012-rich-media-video-leading-the-charge/

[22] X. Hu, N. Sun, C. Zhang, and T.-S. Chua, *Exploiting internal and external semantics for the clustering of short texts using world knowledge*, CIKM '09. 2009

[23] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang, and Z. Chen *How much can Behavioral Targeting Help Online Advertising?*, WWW '09. 2009

[24] Y. Xie, Y. Gao, J. Gou, Y. Cheng, D. Honbo, K. Zhang, A. Agrawal, and A. Choudhary, *Probabilistic Marco Behavioral Targeting*, CIKM DUBMMSM workshop. October 2012