Collective Attention and the Dynamics of Group Deals

Mao Ye[*]
Social Computing Group
HP Labs
California, USA
mxy177@cse.psu.edu

Thomas Sandholm
Social Computing Group
HP Labs
California, USA
thomas.e.sandholm@hp.com

Chunyan Wang
Dept. of Applied Physics
Stanford University
California, USA
chunyan@stanford.edu

Christina Aperjis
Social Computing Group
HP Labs
California, USA
christina.aperjis@hp.com

Bernardo A. Huberman
Social Computing Group
HP Labs
California, USA
bernardo.huberman@hp.com

ABSTRACT

We present a study of the group purchasing behavior of daily deals in Groupon and LivingSocial and formulate a predictive dynamic model of collective attention for group buying behavior. Using large data sets from both Groupon and LivingSocial we show how the model is able to predict the success of group deals as a function of time. We find that Groupon deals are easier to predict accurately earlier in the deal lifecycle than LivingSocial deals due to the total number of deal purchases saturating quicker. One possible explanation for this is that the incentive to socially propagate a deal is based on an individual threshold in LivingSocial, whereas in Groupon it is based on a collective threshold which is reached very early. Furthermore, the personal benefit of propagating a deal is greater in LivingSocial.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavior Sciences; G.3 [**Mathematics of Computing**]: Probability and Statistics

General Terms

Economics, Theory, Algorithms

Keywords

group deals, collective attention, purchase dynamics

[*]Mao Ye is also a PhD student in the Department of Computer Science and Engineering, the Pennsylvania State University, Pennsylvania, USA

1. INTRODUCTION

Attracting the attention of potential customers in today's information rich social media is a challenge. As a result marketers have been forced to target customers in more sophisticated ways. Location-based (regional) and hyper-location-based (within eye-sight) targeting has turned out to be very effective in terms of improving conversion rates from views to purchases [11]. However, since people are unwilling to share their exact locations out of privacy concerns they need to be given some incentive to reveal their position. The most successful incentive employed to date is daily deals.[1] In spite of the success of this strategy it is not fully understood what makes it successful and what kind of social behavior the daily deals sites so effectively tap into and exploit. However, it is clear that deadlines and social propagation play important roles in addition to location-based targeting. The main question we are addressing in this work is how to describe the purchasing pattern more precisely in order to predict the future popularity of a deal.

We analyzed data from Groupon and LivingSocial, the current market leaders of daily deals in the US. Groupon promotes deals for different geographic markets, or cities, called divisions. In each division, there is typically one featured daily deal. A deal is a coupon for some product or service at a substantial discount off the regular price. Deals may be available for one or more days. Coupons are only redeemable if a certain minimum number of customers purchases the deal, and this number constitutes what Groupon calls a *tipping point*. Furthermore, sellers may set a maximum threshold size to limit the number of coupons that can be purchased. LivingSocial is similar to Groupon, except that there is no tipping point. The incentive that drives users to propagate deals is the following commitment made by LivingSocial: "Buy first, then share a special link with friends, if three friends buy, yours is free!".[2]

A closer examination of the mechanisms driving user behavior in group deals could provide useful guidance for merchants' local marketing campaigns as well as for deal providers. If the time a deal will run out of coupons or the total coupons sold at a given time can be predicted for a deal, then a deal marketing scheduler could maximize the profits from a port-

folio of live deals by trying to adjust the exposure so that the aggregate expected profit is maximized within some time horizon.

In this paper we study the evolution of collective attention measured as deal purchases. We base our analysis on data collected from Groupon over two months and from Living-Social over one month.

The contributions of this paper fall into two categories:

- **Structure of purchasing dynamics.** We present a stochastic model that analytically explains the observed purchasing behavior.

- **Prediction model for purchases.** We show how the model is able to predict the success of group deals as a function of time.

The paper is structured as follows. In Section 2, we discuss related work. In Section 3, we discuss the data sets and the collection strategies used in our study. Section 4 describes and empirically verifies our stochastic model. Then in Section 5 we use our model to predict purchase volume and benchmark it against some baselines. Section 6 concludes with possible applications of our work and future directions.

2. RELATED WORK

The related work comes from two broad areas, social purchasing behavior, and collective attention.

2.1 Social Purchasing Behavior

According to [8, 10], a buyer's social network strongly influences her purchasing behavior. In [10], Guo et al. analyze data from the e-commerce site Taobao[3] to understand how individuals' commercial transactions are embedded in their social graphs. In the study, they show that implicit information passing exists in the Taobao network, and that communication between buyers drives purchases. However, according to the study presented in [16] social factors may impose a different level of impact on the user purchase behavior for different e-commerce products.

Several studies have been conducted to understand various aspects of Groupon. In [1], Arahbshai examined the business model of Groupon. In [7], Dholakia conducted a survey-based study on Groupon, in order to understand how businesses fare when running group promotions. Employee satisfaction, rather than features of the promotion or its effect, was found to be the factor that correlates most strongly with the profit gained from a promotion. Effectiveness in reaching new customers and the percentage of Groupon users who bought more than the deal's value during the visit were important factors for the small merchants when considering whether to run another promotion. In [9], Grabchak et al. study the problem of selecting Groupon style chunked reward advertisements. To address the problem, they devise several adaptive greedy algorithms in a stochastic Knapsack framework.

The papers most related to our work are [4, 5], where data on the purchase history of Groupon, and LivingSocial deals were analyzed. One key outcome of [4] is the preliminary evidence that Groupon is behaving strategically to optimize

deal offerings, giving customers "soft" incentives (e.g., deal scheduling and duration, deal featuring, and limited inventory) to make a purchase. In [5], the authors study the effect of social influence by external sites such as Facebook and Yelp. They conclude that these sites can help promote deals but they do not have any effect in terms of the long term reputation of the merchants. Our work differs from these studies by focusing on modeling the deal purchasing dynamics over time, as opposed to using a linear regression model of deal attributes. Furthermore, in our work the effects of social propagation are inferred from the purchasing information itself (as opposed to using external data sources such as Facebook or Yelp), which simplifies deployment.

2.2 Collective Attention

In [14, 13, 15], Lerman et al. propose to use a stochastic model to describe the social dynamics of web users, with Digg as a case study. The stochastic model focuses on describing the aggregated (by average quantities) behavior of the system, including average rate at which users contribute new stories and vote on existing stories. With the devised stochastic model, popularity of a Digg story can be predicted shortly after it was submitted (or with 10 to 20 votes). Studies in [12, 3, 6] have found that early diffusion of information within a community could be a good predictor of how far it will spread.

Recent studies of collective attention on social media sites such as Twitter, Digg and YouTube [18, 17, 2] have clarified the interplay between popularity and novelty of user generated content. The allocation of attention across items was found to be universally log-normal, as a result of a multiplicative process that can be explained by an information propagation mechanism inherent in all these sites. While the specific time scales over which novelty decays differ between different systems depending on their typical type of content, the functional form of the decay is consistent and thus future popularity is predictable.

3. DATASETS

We collected data from Groupon's socially promoted and local daily deal websites in the US. We also collected data from LivingSocial to verify that our models could be applied more generally across group deal sites.

Groupon provides a convenient API,[4] which allows us to obtain more detailed information about the deals. By the end of April 2011, Groupon's business covered about 120 cities in the US.[5] We monitored all Groupon deals offered in 60 different randomly selected cities during the period between April 4th and June 16th, 2011. In total we collected the entire purchase traces of 4349 deals.

In LivingSocial, there is no API available for us to periodically obtain information about deals, so we developed a crawler to visit the web pages of deals. After crawling for one month, we collected traces from over 900 deals.

To give a flavor of the type of data we use, we next examine the features of Groupon deals in more detail. A similar examination for LivingSocial is outside the scope of this work. However, we will later see that the models inferred from these observations apply to LivingSocial as well.

---

[3]Taobao is a Chinese Consumer Market place, and also the world's largest e-commerce website, http://www.taobao.com.

[4]http://www.groupon.com/pages/api

[5]Statistics obtained from Groupon API.

Description	coefficient	standard error	t-value	p-value
log(Tipping Point)	0.7316	0.029	25.2276	6.5792×10^{-125}(***)
Featured position	0.7004	0.0463	15.1189	2.0166×10^{-49}(***)
Duration	0.0062	4.8862×10^{-4}	12.6412	1.6054×10^{-35}(***)
Is limited	-2.6105×10^{-4}	2.0969×10^{-5}	-12.4494	1.5597×10^{-34}(***)
Discount	-0.0011	1.6681×10^{-4}	-6.3744	2.1908×10^{-10}(***)
Sunday	0.0061	0.0022	2.7358	0.0063 (***)
Nightlife	0.3208	0.1515	2.1180	0.0343 (*)
Health&Fitness	0.6429	0.0849	7.5722	5.1827×10^{-14}(***)
Travel	-0.1789	0.0782	-2.2874	0.0223 (*)
Automotive	-0.3289	0.1366	-2.4074	0.0161 (*)
Atlanta	-2.0460	0.9373	-2.1829	0.0291 (*)
Albuquerque	-1.8548	0.9365	-1.9806	0.0478 (*)
Austin	-2.4329	0.9516	-2.5567	0.0106 (*)
Abbotsford	-2.1012	0.9392	-2.2371	0.0254 (*)
Barrie	-2.2454	0.9496	-2.3646	0.0181 (*)
...	...	...	...	...

Table 1: **Multivariate linear regression of log(total number of purchases).** $N = 3876$, R-square $= 0.5952$, adjusted R-square $= 0.5857$. **Note that, due to space limitation, we only show the result with p-value smaller than 5% for the launching day, category and division study.**

3.1 Groupon Deal Characteristics

At the time of our study the Groupon website presented the following relevant deal information: description, discount, time of launch, tipping point (purchases required for a deal to actually be sold), and the maximum number of sales of the deal. Additionally, users could monitor the current number of purchases[6] and whether the deal has tipped or sold out. We monitored the number of purchases and the position of each deal in 20-minute time intervals. A surprisingly large portion (10%) of all deals exhibited dramatic non-monotonically increasing behavior, e.g. a decrease of 10 purchases between subsequent intervals. This may indicate that something was wrong with the deal, e.g. false marketing due to an inflated list price, and customers who initially purchased the deal requested a refund (an option Groupon supports and markets). Due to the unknown user behavior behind these deal actions we exclude these deals from our study. Hence, 3876 deals were left to analyze. In our dataset, 270 deals (out of 3876) had not reached their tipping point when they expired. In the following, these deals are called *failed* deals; and deals that are turned on successfully are called *tipped* deals.

3.2 Attributes of Groupon Deals

Here we present some statistics about attributes of the deals in our Groupon dataset, including retail price, discount, deals needed to tip (tipping point), time that tipping occurs (tipping time), lifetime of a deal and total number of purchases.

Groupon deals have different retail prices and discounts. The mean value of retail price is $44 and the mode value is $10. We observe that most of the discounts range from 50% to 60%, and the mode value is 50%. Based on these statistics, we see that the product and services deals provided on the Groupon website are not expensive most of the time, and the discounts are usually very big.

In Groupon, deals may have different tipping points and successful deals may also have different tipping times even when they have the same tipping points. The average number of tipping points or units needed to tip is 22 (mode value is around 10) and the expected tipping time is about 10.5 hours (the mode is around 7 hours). Most of the time, deals in Groupon were tipped within one day.

Note that the lifetime of a deal in Groupon is usually set to 1 day, 2 days, 3 days or 4 days. The average number of purchases of a deal is 373. A deal may be specified with a limited available quantity. So these numbers are mixtures of different factors, such as the quality of a deal itself, the quantity available etc.

3.3 Factors Impacting Purchases on Groupon

As we are ultimately interested in modeling purchase dynamics of deals, we first need to understand what factors impact purchases. We now consider how the attributes discussed in the previous section affect the total number of purchases of a deal. It is worth noting that if the Groupon commission is known,[7] this number also gives a good estimate of the merchant's revenue from a deal.

We regress the logarithm of the total number of purchases for each deal against: the logarithm of the tipping point, whether the deal was listed in featured position (1) or not (0), the duration, whether the deal inventory is limited (1) or not (0), the discount, the retail price, the launch day, the category, and the city. The parameter estimates, their standard errors, t-values and p-values are listed in Table 1. Due to space limitations, only attributes with significance level (p-value) smaller than 5% are shown in the table. Among those attributes, we find that tipping point and featured position are the two most significant factors that can help predict the number of purchases. Surprisingly, tipping point seems to have better predicting power than featured position (i.e., the t-value is much larger for the tipping point factor than for the featured position factor). In the next

---

[6]The current number of purchases has since our study been removed and replaced with an obfuscated threshold to make it harder to make predictions.

[7]reportedly 50% in [1]

section, we show how the tipping time can be generalized as an inflection point in the purchase dynamics of group deals.

4. PURCHASE DYNAMICS

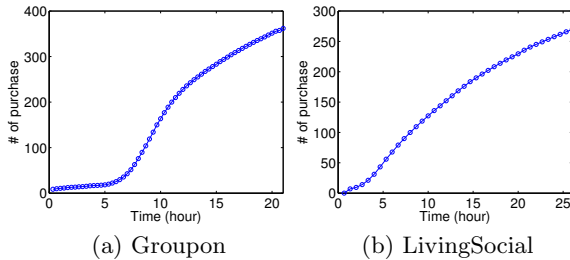In this section, we propose a model of the purchase dynamics of group deals.

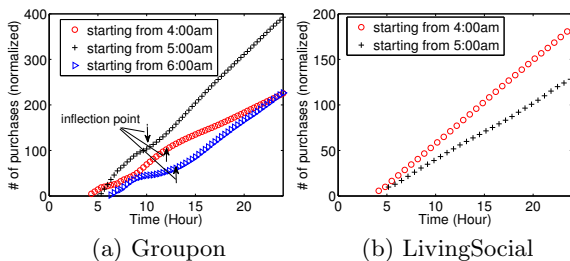**Figure 1: Purchase growth of deals**

Figure 2: Normalized Purchase growth of deals

We average the number of purchases of deals for each time step in both Groupon and LivingSocial. As shown in Figure 1, deals in LivingSocial grow faster than Groupon in the first few hours. A possible reason is due to the different incentive that LivingSocial is using to promote deals. LivingSocial users who want to get free deals may disseminate deal information more eagerly.

Furthermore, there is an inflection point in the purchase dynamics for both Groupon and LivingSocial deals (after around 7 and 4 hours in Figure 1(a) and (b), respectively), after which the number of purchases grows faster; whereas the number of purchases grows relatively slowly and steadily before the inflection point. Note that after the inflection point, the number of purchases grows dramatically for about 11.6 and 14.8 hours in Groupon and LivingSocial, respectively, after which the purchase rate drops.

One may argue that this inflection point could be caused by time-of-day seasonality given that all deals are local for a region belonging to a single time zone. For example, most people do not buy deals at night, but early in the morning when they wake up. Hence, we normalize the number of purchases by removing the seasonal impact to examine whether the inflection point is caused by the time the deal is launched, as shown in Figure 2. In Groupon, 95% of the deals are launched before 7:00am and 50% of these are launched between 4:00am and 6:00am. Hence, we cluster deals in three groups, those that launch around 4:00am, 5:00am, and 6:00am respectively. As shown in Figure 2(a), normalized purchase growth of deals clearly has two-stage

growth, which is divided by an inflection point. Before the inflection point, it shows non-linear growth; while after the inflection point, it obeys linear growth. In LivingSocial, deals are launched during 4:00am~6:00am, like Groupon. Interestingly, in Figure 2(b), we find the inflection point in the purchase growth of LivingSocial deals disappears after the normalization. In addition, deals launched from the same time (e.g., from 4:00 am) exhibit different purchase dynamics behavior in Groupon and LivingSocial, e.g., in Figure 2, the purchase dynamics of Groupon deals still exhibit an inflection point, while there is none in LivingSocial deals. These observations suggest that: (1) the consistent launch times may cause the two-stage purchase growth in LivingSocial; but (2) the inflection point cannot solely be attributed to the time the deal is launched in Groupon: the tipping-point mechanism may also play a role here.

4.1 Purchase Dynamics Before Inflection

The purchase dynamics before the inflection point in Groupon follow simple linear growth and can be modeled by a random discovery process with independently and identically distributed increments in the number of purchases. The time before inflection is too short to gain any predictive insights.

4.2 Purchase Dynamics After Inflection

We now focus on purchase dynamics after a deal tips. For expositional clarity consider the time of inflection as time 0, and denote by N_t the number of purchases t time units after the inflection point. Thus, N_0 denotes the number of purchases of a deal at the inflection point time.

We start with the multiplicative process given by the following equation:

$$N_{t+1} - N_t = r(t)X_t N_t \qquad (1)$$

According to (1), after the inflection point, the increase in the number of purchases $(N_{t+\Delta t} - N_t)$ is proportional to the number of people that have purchased the deal up to time t. Intuitively, a fraction of the people that already purchased the deal will notify some of their friends about it, and a fraction of these friends will purchase the deal. These fractions are represented by the positive random variable X_t. We assume that $\{X_t\}$ are independent and identically distributed random variables. Since X_t is assumed to be positive, N_t can only increase over time. This growth in time is eventually curtailed by a decay in novelty, which is parameterized by the factor $r(t)$. As we discuss later, $r(t)$ is decreasing in t. This notation of social propagation is borrowed from and motivated in more depth in [18]. This process, called "growth with random multiplicative noise", describes the dynamics of users' attention to web content.

According to Equation (1), the number of purchases at time T (that is, T time units after the inflection point) is given by

$$N_T = \prod_{t=1}^{T} (1 + r(t)X_t)N_0 \qquad (2)$$

Note that the realization of X_t will in general be different in different time periods; however all random variables X_t follow the same distribution. When X_t is small (which is the case for small time steps), we have the following approximate

solution for N_T:

$$N_T \approx \left(\prod_{t=1}^{T} e^{r(t)X_t} \right) N_0 = e^{\sum_{t=1}^{T} r(t)X_t} N_0. \qquad (3)$$

Taking the logarithm on both sides, we get

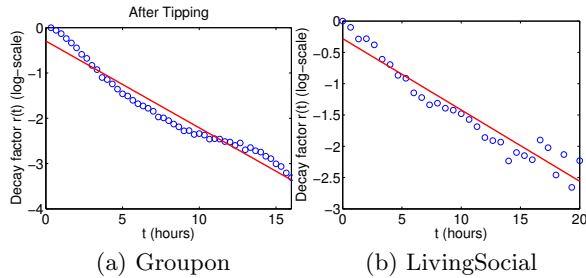$$\log N_T - \log N_0 \approx \sum_{t=1}^{T} r(t)X_t \qquad (4)$$

Figure 3: Process of novelty decay

The decay factor $r(t)$ is estimated according to Equation (4) as follows:

$$r(t) = \frac{\mathbb{E}(\log N_t) - \mathbb{E}(\log N_{t-1})}{\mathbb{E}(\log N_1) - \mathbb{E}(\log N_0)} \qquad (5)$$

where we normalize $r(1)$ to 1.[8] This calculation is again borrowed from and evaluated in more detail in [18].
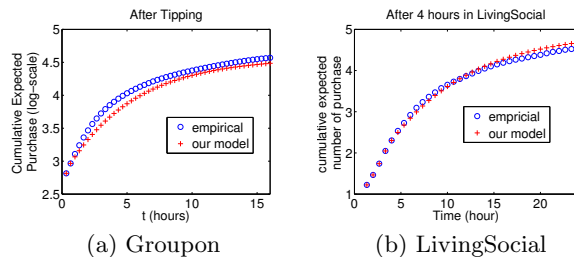

Figure 4: Empirical verification of our model

In Figure 3, we plot the novelty decay $r(t)$ for the first 16 and 20 hours after the inflection point in Groupon and LivingSocial, respectively, as estimated from our dataset. Note that tipping time is usually around 8 hours, so we focus on the time duration of 16 hours after tipping in Groupon. We observe that $r(t)$ decreases over time. Moreover, Figure 3 suggests that the novelty decay is exponential. In particular,

$$r(t) \approx \exp(at + b), \qquad (6)$$

where in Groupon $a = -0.21$ and $b = -2$, and the R^2 value for this fit is 0.88; and in LivingSocial $a = -0.11$ and $b = -0.28$ and R^2 value for this fit is 0.92.

Next, we are interested in evaluating how well our model helps explain the purchase growth after a deal has turned on. With both a, b estimated, we can use our results to explain the growth of purchases. In Figure 4, we demonstrate

---

[8]Note that since we normalize $r(1)$ to 1, Equation (4) implies that $\mathbb{E}X_t = \mathbb{E}(\log N_1) - \mathbb{E}(\log N_0)$.

the potential predictive power of our model by empirically verifying the growth of purchases of deals after they have tipped. For the model fitting in Figure 4, the R^2 value is 0.94 and 0.99 in Groupon and LivingSocial, respectively.

5. PURCHASE PREDICTION

In this section, we discuss how to use our models to predict the number of purchases of deals at a given time. Purchase prediction is important for both group deal websites and local merchants. Accurate forecasts may help group deal websites design more optimized deal scheduling and promotion strategies and aid local merchants in allocating resources more efficiently.

We now discuss methods which make predictions based on h hours of previous observations.

5.1 Predictors

5.1.1 Baselines

The first simple baseline algorithm (denoted as `baseline1`) is to treat the current number of purchases as the future number of purchases, and hence it guarantees less than 100% relative error, given that the number is increasing and always positive.

Another baseline algorithm (denoted as `baseline2`) is to assume a linear relationship between the current number of purchases and the future number of purchases. Suppose we know the number of purchases N_{t_1} at time t_1, and aim to predict the number of purchases N_{t_2} at time t_2, where $t_1 < t_2$. Then we assume that

$$N_{t_2} = \alpha N_{t_1} + \beta \qquad (7)$$

where α and β are model parameters that can be learned from training data.

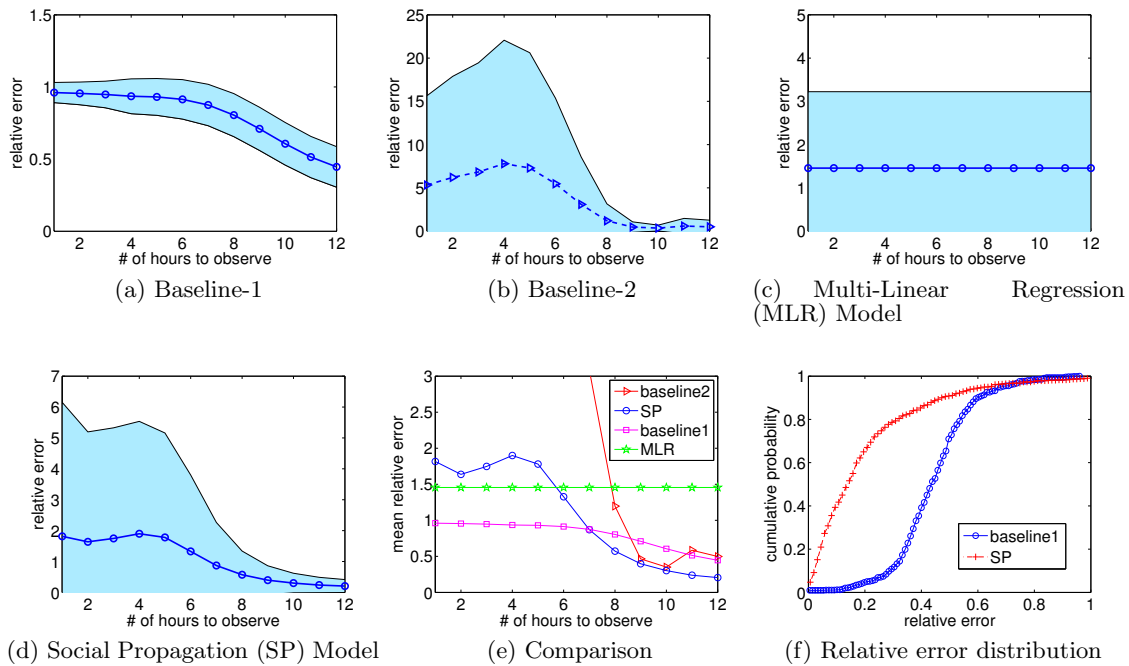#### 5.1.2 Social Propagation Model

As seen in Figure 4, the growth in sales after tipping in Groupon is described well by a multiplicative process. What follows from the model is that to obtain the popularity for the next time step we multiply the current popularity by a small, random amount. More specifically, let t_1 and t_2 denote two different time steps and $t_1 < t_2$. Then, we have

$$\log N_{t_2} \approx \log(N_{t_1}) + \sum_{t=t_1}^{t_2} r(t)X_t \qquad (8)$$

according to Equation (4)

Here, we introduce the process used to model and predict the future number of purchases of a deal. To help determine whether the number of purchases early on is a predictor of later number of purchases we plot the number of purchases at the reference time $t_1 = 8$ hours vs. the number of purchases at the end of a day (i.e., $t_2 = 24$ hours) in both Groupon and LivingSocial (see Figure 5). We logarithmically rescaled the horizontal and vertical axes in the figure to show the number of purchases for different deals, which span four orders of magnitude.

Figure 5 shows that there is a strong correlation between the earlier observations of the number of purchases of a deal and the later observations. So we can determine the linear regression coefficients between t_1 and t_2 on a given training dataset, and then use the estimated coefficients to extrapolate on the test dataset.

(a) Baseline-1   (b) Baseline-2   (c)   Multi-Linear   Regression (MLR) Model

(d) Social Propagation (SP) Model   (e) Comparison   (f) Relative error distribution

Figure 6: Performance comparison of prediction of the number of purchases after one day in Groupon. In (a)-(e), lines denote the average relative error, and shaded regions cover the areas of one-standard error.

(a) Groupon   (b) LivingSocial
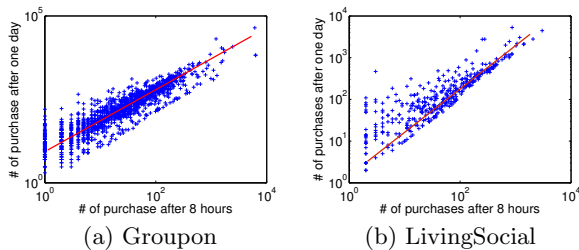
Figure 5: Number of purchases after 8 hours vs. number of purchases after one day (log-log-scale). The bold line is the linear fit to the data.

Note that there is a limitation to this approach. As we discussed before, the multiplicative process which gives rise to Equation (3) governs the dynamics only after tipping — and not before. So this approach may not perform well for the very early observations. Nevertheless, it is applicable to both Groupon and LivingSocial since the multiplicative process is the main process during the life cycle of a deal for both services.

5.2   Evaluation

In this subsection, we conduct an experimental study to evaluate the proposed prediction algorithms. As discussed before, the important task is to be able to predict how successful a deal will be. Since there are many deals with a lifetime of one day we evaluate the performance of different algorithms by how accurately they can predict the number of purchases of a deal after one day. Here, we use relative error, i.e., $\frac{|\text{real purchases - predicted purchases}|}{\text{real purchases}}$, as the performance metric to measure accuracy.

5.2.1   Experiments with Groupon Deals

First, we conduct experiments on the Groupon dataset by randomly splitting it into halves, where one half is used for training and another half is for testing.

In Figure 6, we find `baseline1` shows the best performance among all the testing algorithms with less than 7-hours of observations. After 7 hours of observations, our proposed social propagation model (denoted `SP`) shows the best performance. Note that a deal which attracts more than hundred purchases within the first hour after launching (6 deals in total in the experiment) is treated differently by applying `baseline1`, as these deals are extremely popular and don't follow the general multiplicative process. The justification for applying `baseline1` is that, these deals are so appealing that local merchants usually place quantity limits.

As we observed before, deals in Groupon are usually tipped after about 7 hours. Before tipping, the purchase dynamics is governed by random discovery instead of the multiplicative process, thus the social propagation model fails to achieve good performance. However, we find that there is an inflection point which occurs at about 7 hours. After 7 hours of observations, the social propagation model exhibits relatively good performance, and it performs much better with more hours of observation. In Figure 6 (f), relative error distributions of `baseline1` and `SP` with 12 hours of observations are examined. We find that the relative error is less than 50% for over 90% of deals when using `SP`, and there are about 70% of deals achieving less than 20% relative error when applying `SP`.

In the experiment, we incorporated all the attributes of the deals into the multi-linear regression (denoted as `MLR`) model, including the tipping point. Tipping points can be considered as the observation of the number of purchases at around 6-8 hours. Therefore, as shown in Figure 6(e), the

[Deal Title: The Magnetic Field - Asheville] $12 for Two Tickets to a Theater Performance (Up to $28 Value)			
Algorithms	Real purchases	Predicted purchases	Relative error
`baseline-1` 12-hour observation	251	93	0.63
`baseline-2` 12-hour observation	251	482	0.92
`MLR`	251	51	0.80
SP with 12-hour observation	251	355	0.42
[Deal Title: Lime Leaf Thai Cuisine - Hendersonville] $10 for $20 Worth of Thai Fusion Cuisine			
Algorithms	Real purchases	Predicted purchases	Relative error
`baseline-1` 12-hour observation	384	169	0.56
`baseline-2` 12-hour observation	384	714	0.86
`MLR`	384	1,452	2.783
SP with 12-hour observation	384	463	0.21

Table 2: Example prediction results for Groupon deals.

(a) Baseline-1   (b) Baseline-2   (c) Social Propagation Model   (d) Comparison   (e) Relative error distribution
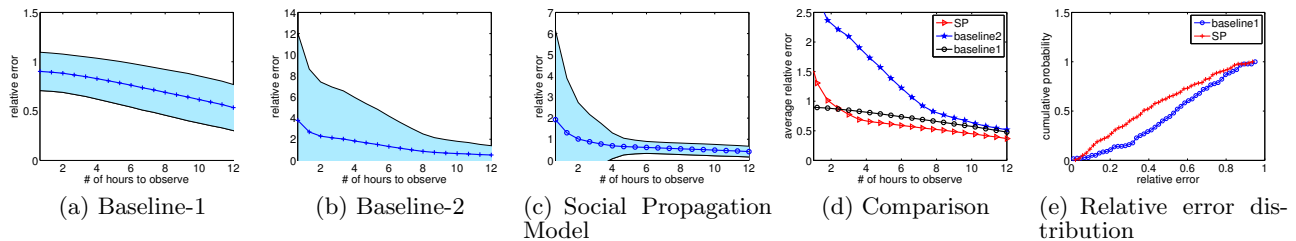
Figure 7: Performance comparison of prediction of the number of purchases after one day in LivingSocial.In (a)-(d), lines denote the average relative error, and shaded regions cover the areas of one-standard error.

multi-linear regression model achieves a comparable performance with our model within an observation period of 6 hours. To exemplify the prediction accuracy, we show the results from a few Groupon deals in Table 2.

As a refinement for Groupon deals, we perform `baseline1` if the deal has not tipped; otherwise, we apply the social propagation (`SP`) model.

5.2.2 Experiments with LivingSocial Deals

We conducted similar experiments on the LivingSocial dataset. As shown in Figure 7, our social propagation model (`SP`) always outperforms `baseline2` and beats `baseline1` with more than 2-hours of observations. Because of the limitations of the crawling technique, we do not have information about which deal is the featured one in a given city; and there is no tipping point in LivingSocial, which prevents the multi-linear regression model from generating good predictions. However, the social propagation model shows very good performance in LivingSocial. In particular, we examine the distribution of relative errors for predictions based on `SP` and `baseline1` with 12-hours of observations in LivingSocial. As shown in Figure 7(e), we find that there are about 65% of deals with less than 50% relative error; and `SP` always outperforms `baseline1`.

Similarly, we show prediction results from some LivingSocial deals in Table 3. As shown in Table 3, the social propagation model exhibits better prediction performance than both baselines, in terms of relative error.

Finally, our design for purchase prediction of Groupon deals is that we perform `baseline1` with less than 3 hours of observations; otherwise, we apply the social propagation (`SP`) model. Note that due to different mechanisms in Groupon and LivingSoical, inflection points are placed

at very different times (i.e., 6-8 hours in Groupon, and 2-4 hours in LivingSocial). Therefore, `SP` can be applied earlier in LivingSocial than in Groupon. However, as shown in Figure 6(e) and Figure 7(d), the relative error measured on the test set decreases rapidly for Groupon, while for LivingSocial the prediction converges more slowly to the actual value. After 17 hours, the expected relative error obtained when estimating one-day purchases of a deal by using `SP` is about 20%, while the same relative error is attained 13 hours after a Groupon deal is launched. This is due to the fact that novelty decay is faster in Groupon than in LivingSocial, i.e. it takes 7 hours in Groupon to reach the saturating point; while it takes about 14 hours in LivingSocial to reach the saturating point in Figure 4. So it is easier to predict the one-day purchases of Groupon deals with fewer hours of observations (after tipping). One possible explanation of this is that the tipping point incentive mechanism for propagating deals in Groupon disappears after the tipping point has been reached. In LivingSocial, on the other hand, the incentive to propagate a deal is always present for at least some users and furthermore the individual gain of propagating is greater. It is worth noting, that in addition to these incentives provided in Group and LivingSocial, people also propagate deals to friends in order to enjoy the corresponding activities with them.

6. CONCLUSIONS

In this paper, we presented a study of the group purchasing behavior of daily deals in Groupon and LivingSocial and introduced a predictive dynamic model of collective attention for group buying behavior. Using large data sets from both Groupon and LivingSocial we showed how the model was able to predict the popularity of group deals as a func-

[Deal Title: Coastal Contacts] $60 to Spend on Prescription Eyeglasses (Now $19)			
Model	Real purchases	Predicted purchases	Relative error
`baseline1` with 12-hour observation	129	32	0.75
`baseline2` with 12-hour observation	129	245	0.90
SP with 12-hour observation	129	110	0.14

[Deal Title: Dawgs!] $10 (Pay $5) or $20 (Pay $10) to Spend on Food and Drink			
Model	Real purchases	Predicted purchases	Relative error
`baseline1` with 12-hour observation	75	28	0.63
`baseline2` with 12-hour observation	75	147	0.96
SP with 12-hour observation	75	110	0.47

Table 3: Example prediction results for LivingSocial deals.

tion of time. Our main finding is that the different incentive mechanisms in Groupon and LivingSocial lead to different propagation behavior, which in turn leads to differences in predictability. However, the basic stochastic processes as well as the distributional parameters of growth and decay are strikingly similar. Given that Groupon no longer provides detailed statistics of purchases over time, the models presented here will not be as accurate for an external observer of Groupon purchases anymore, just for LivingSocial. However, both deal site owners and merchants would benefit from analyzing the early stream of purchases using the models presented here. Our work also gives some insights into how different incentive mechanisms can affect the longevity of propagation momentum. These insights could be exploited in local marketing campaigns where viral and social dissemination of offers is desirable.

7. REFERENCES

[1] A. Arabshahi. Undressing Groupon: An analysis of the Groupon business model, December 2010. Available at http://www.ahmadalia.com.

[2] S. Asur, B. A. Huberman, G. Szabó, and C. Wang. Trends in social media : Persistence and decay. *CoRR*, abs/1102.1402, 2011.

[3] E. Bakshy, B. Karrer, and L. A. Adamic. Social influence and the diffusion of user-created content. In *ACM Conference on Electronic Commerce*, pages 325–334, 2009.

[4] J. W. Byers, M. Mitzenmacher, M. Potamias, and G. Zervas. A month in the life of Groupon. *CoRR*, abs/1105.0903, 2011.

[5] J. W. Byers, M. Mitzenmacher, M. Potamias, and G. Zervas. Daily deals: Prediction, social diffusion, and reputation ramifications. *CoRR*, abs/1105.0903, 2011.

[6] R. Colbaugh and K. Glass. Early warning analysis for social diffusion events. In *ISI*, pages 37–42, 2010.

[7] U. M. Dholakia. How effective are Groupon promotions for businesses? September 28, 2010. Available at http://ssrn.com/abstract=1696327.

[8] P. DiMaggio and H. Louch. Socially embedded consumer transactions: For what kinds of purchases do people most often use networks? *American Sociological Review*, 63:619–637, 1998.

[9] M. Grabchak, N. L. Bhamidipati, R. Bhatt, and D. Garg. Adaptive policies for selecting Groupon style chunked reward ads in a stochastic knapsack framework. In *Proceedings of 20th International World Wide Web Conference (WWW)*, 2011.

[10] S. Guo, M. Wang, and J. Leskovec. The role of social networks in online shopping: Information passing, price of trust, and consumer choice. In *ACM Conference on Electronic Commerce (EC)*, 2011.

[11] JiWire. JiWire Mobile Audience Insights Report Q1 2011, 2011. Available at http://www.jiwire.com/insights.

[12] K. Lerman and A. Galstyan. Analysis of social voting patterns on Digg. In *ACM SIGCOMM Workshop on Online Social Networks*, 2008.

[13] K. Lerman and R. Ghosh. Information contagion: An empirical study of spread of news on Digg and Twitter social networks. In *Proceedings of 4th International Conference on Weblogs and Social Media (ICWSM)*, May 2010.

[14] K. Lerman and T. Hogg. Using a model of social dynamics to predict popularity of news. In *Proceedings of 19th International World Wide Web Conference (WWW)*, 2010.

[15] K. Lerman and T. Hogg. Using stochastic models to describe and predict social dynamics of web users. *to appear in ACM Transactions on Intelligent Systems and Technology*, 2011.

[16] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web*, 1(1), 2007.

[17] G. Szabó and B. A. Huberman. Predicting the popularity of online content. *Commun. ACM*, 53(8):80–88, 2010.

[18] F. Wu and B. A. Huberman. Novelty and collective attention. *PNAS*, 104(45), 2007.