Modeling Paying Behavior in Game Social Networks

Zhanpeng Fang[†], Xinyu Zhou[†], Jie Tang[†], Wei Shao[‡], A.C.M. Fong[♮], Longjun Sun[‡], Ying Ding[♯], Ling Zhou[⋆], and Jarder Luo[⋆]
[†]Department of Computer Science and Technology, Tsinghua University, China
[‡]Tencent Corporation, Shenzhen, China
[♮]School of Computing and Mathematical Sciences, Auckland University of Technology, New Zealand
[♯]Department of Information Science, Indiana University, USA
[⋆]Department of Sociology, Tsinghua University, China
{fzp13, zhouxy11}@mails.tsinghua.edu.cn, jietang@tsinghua.edu.cn, {abelyshao,stanleysun}@tencent.com

ABSTRACT

Online gaming is one of the largest industries on the Internet, generating tens of billions of dollars in revenues annually. One core problem in online game is to find and convert free users into paying customers, which is of great importance for the sustainable development of almost all online games. Although much research has been conducted, there are still several challenges that remain largely unsolved: What are the fundamental factors that trigger the users to pay? How does users' paying behavior influence each other in the game social network? How to design a prediction model to recognize those potential users who are likely to pay?

In this paper, employing two large online games as the basis, we study how a user becomes a new paying user in the games. In particular, we examine how users' paying behavior influences each other in the game social network. We study this problem from various sociological perspectives including strong/weak ties, social structural diversity and social influence. Based on the discovered patterns, we propose a learning framework to predict potential new payers. The framework can learn a model using features associated with users and then use the social relationships between users to refine the learned model.

We test the proposed framework using nearly 50 billion user activities from two real games. Our experiments show that the proposed framework significantly improves the prediction accuracy by up to 3-11% compared to several alternative methods. The study also unveils several intriguing social phenomena from the data. For example, influence indeed exists among users for the paying behavior. The likelihood of a user becoming a new paying user is 5 times higher than chance when he has 5 paying neighbors of strong tie. We have deployed the proposed algorithm into the game, and the *Lift_Ratio* has been improved up to 196% compared to the prior strategy.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Miscellaneous; H.2.8 [**Database Management**]: Data Mining

CIKM'14, November 03 - 07, 2014, Shanghai, China.

http://dx.doi.org/10.1145/2661829.2662024 .

General Terms

Algorithms, Experimentation

Keywords

Social game, Social networks, User behavior modeling

1. INTRODUCTION

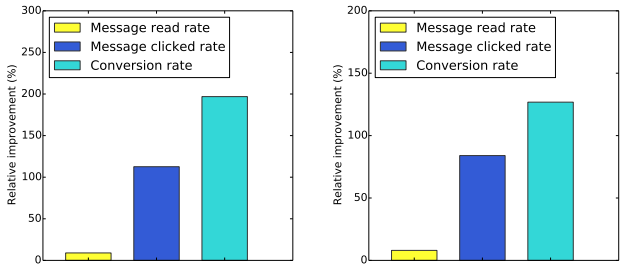
Online gaming is one of the largest industries on the Internet, generating tens of billions of dollars in revenues annually. According to Facebook' first quarter report in 2013, it has 250 million people playing games on Facebook monthly, and there are about 200 games right now on Facebook with more than 1 million active users each. About 12% of the company's revenue is directly from games. (This number does not even include the advertising shown alongside the games.) The situation is the same in China. Tencent is the largest Internet company and also the largest game service provider in China. It has more than 400 million gaming users. The payment revenue from games accounts for roughly 50% of Tencent's overall revenue.

In online game, one key problem is how to design strategies to keep players not just playing, but also paying. This is important for the sustainable development of the whole game industry. The problem has attracted attention of both academic and industry communities. For example, Yee [30] investigated how players differ from one another and how motivations of play relate to age, gender, usage patterns, and in-game behaviors. Recently, online games are becoming more and more *social*. Statistics show that 80% of Zynga's[1] revenue comes from Facebook users. Traditional research mainly focuses on analyzing users' personal attributes but ignores the social effects. Ducheneaut and Moore [11] studied the users' interaction patterns in the online game. Ducheneaut et al. [12] further used longitudinal data collected from the online game to examine play and grouping patterns. They found several interesting patterns that affect the formation and longevity of gaming communities. However, they do not consider users' paying behavior. In practice, users' play patterns and social activities in the game are strongly affected by their paying status.

In this paper, we try to systematically study users' paying behavior in online game. Precisely, we aim to understand what are the fundamental factors that trigger free users to pay. The problem is nontrivial and poses a set of unique challenges:

- **Sparsity:** the paying behavior is rather rare compared to traditional Internet applications such as recommendation and

[1]The largest game provider on Facebook.



(a) Online test 1 (b) Online test 2

Figure 1: Relative improvement in message read rate, message clicked rate and conversion rate compared to the prior strategy in (a) Online test 1 (with 0.8 million users), (b) Online test 2 (with 1 million users).

ranking. Zynga has more than 200 million monthly users; however only 3% of the users have purchased credits in the game.

- **Social effects:** social activity has already become one of the most important elements in designing online games. How is users' paying behavior influenced by friends and the social structure?

- **Predictive models:** in order to effectively identify potential paying users, it is important to develop methods that can combine users' attributes and the social effects.

By carefully studying users' paying behavior in large online game networks, we have discovered several intriguing social patterns. For example, strong influence on paying behavior indeed exists among users. The likelihood that a user who has 5 paying neighbors of strong tie becomes a new payer is 5 times higher than that of an average user. Meanwhile, when a user is aware that his friends have already paid a lot in the game, his willingness to pay will quickly cool down.

Based on the discoveries, we propose a learning framework referred to as local consistent factorization machines (LCFM) model. The framework can learn a model using features associated with users and then use the social relationships between users to locally regularize the learned model. By trading off the global optimum and the local regularization terms, the framework achieves a learning solution with local and global consistency. Our experiments show that the proposed framework significantly improves the prediction accuracy by up to 3-11% compared with several alternative methods. We have deployed the proposed algorithm into the games. Specifically, we use the proposed algorithm to predict who are the most likely to pay in the game. Then the system automatically sends a message to those users suggested by the algorithm. If a user responds and starts to pay in the game, we say the algorithm makes a correct prediction. We compare the proposed algorithm with the prior strategy used in the game (Cf. §6 for details). Figure 1 reports the relative improvement of the proposed algorithm against the prior strategy in two real online tests (with 800,000∼1,000,000 chosen users). It can be seen that the conversion rate (from free user to paying user) has been significantly improved up to 126-196% compared to the prior strategy.

Organization. The rest of this paper is organized as follows: Section 2 introduces the dataset we use for our study; Section 3 presents indepth analysis on the data; Section 4 defines features used in the learning model; Section 5 describes our proposed model for the problem; Section 6 gives experimental results that validate the effectiveness and efficiency of the proposed framework; Section 7 presents related work; and finally Section 8 concludes the work.

Table 1: Statistics of the datasets.

Category	Type	QQSpeed	DNF
User	all users	7.60M	347K
	free users	$\sim 10^6$	$\sim 10^5$
	paying users	$\sim 10^6$	$\sim 10^5$
	new payers	$\sim 10^5$	$\sim 10^4$
Relationship	co-playing	134M	7.30M
Guild	guilds	600K	49.6K
	co-guild	66.7M	51.7M
Activity	activity types	58	64
	activity logs	44.7B	5.71B
Date span	from	2013.6.20	2013.4.1
	to	2013.8.20	2013.6.30

2. DATASET

We study the problem in two large online games: Dungeon & Fighter Online (DNF), the second largest online game in China, and QQSpeed, the fourth largest online game in China.

DNF is a game of melee combat between one user and an improbably large number of underpowered enemies. In 2013, the game attracted more than 400 million users from the world. Users in the game can fight against enemies as individuals, or form a group to fight together.

The DNF dataset comprises of all kinds of user activities from sampled users of the game during the period from Apr. 1st, 2013 to Jun. 30th, 2013. The user activities include subtraction of user money, change of user level, killing mob in games and etc. In summary, there are 5.71 billion activity logs from 347 thousand users.

QQSpeed is a racing game that users can take part in racing competitions to play against other users. The game is the largest online racing game in China and attracted more than 200 million users in 2013. Users in the game can race against other users as individuals, or form a group to race together. The game provider earns money by selling virtual items in the in-game shop, including different types of vehicles and accessories.

The QQSpeed dataset comprises of all kinds of user activities of sampled users in the game during the period from Jun. 20th, 2013 to Aug. 20th, 2013. There are totally 44.7 billion activity logs from 7.60 million users.

In addition, the two datasets also contain the paying logs of all the users. Based on the paying logs, we classify the users into three categories: free user, paying user and new payer, according to their paying behavior. Take QQSpeed for example. We categorize the users who have paying behavior before Jun. 20th, 2013 as paying users. For those users who do not have paying behavior before Jun. 20th, 2013, but paid during Jun. 20th, 2013 to Aug. 20th, 2013, we call them new payers. And the remaining users who do not have any paying behavior are considered as free users.

Besides, both games provide a *guild* system. A guild is similar to a community (group). Any player can request to join an existing guild or create a new guild and invite other players to join. The guild has different levels. Members of a high-level guild can gain additional experience in the game, and on the other hand, a guild can level up when many of its members are logged into the game.

In DNF and QQSpeed, we consider two types of relationships. If two users have played together, we deem the two users are

connected, which is considered to be the first type of relationship. Specifically, in DNF, we consider two users are connected if they joined in the same group to fight enemies together, and in QQSpeed, we consider two users are connected if they joined in the same competition to race together. We call this type of relationship as *co-playing* relationship. Besides, we also consider the strength of this relationship. If two users have played together for more than 5 times, we call the relationship as a strong tie, otherwise weak tie. (Cf. Figure 3(b) for more analysis for strong/weak ties.) The other type of relationship is derived from the guild system: if two users join in the same guild, then we consider a *co-guild relationship* between them. For a specific user v, we call those users $NB(v)$ who have a co-playing relationship (or a co-guild relationship) with this user as his/her *neighbors*, and those neighbors who have already paid in the game, i.e., $NB'(v) \subseteq NB(v)$, as *paying neighbors*.

Based on above methods, we construct a game network of 263 thousand users with 7.3 million co-playing relationships and 52 million co-guild relationships in DNF, and a game network of 3.54 million users with 134 million co-playing relationships and 66.7 million co-guild relationships in QQSpeed. Table 1 shows statistics of the datasets. For simplicity, we only consider the co-playing relationship in the following analysis and experiments, but the proposed model can be easily transferred to different types of relationship.

3. OBSERVATIONS

We first engage in some high-level investigation of how different factors influence users' paying behavior, since a major motivation of our work is to find out what the fundamental factors are that trigger users to pay in online game. In particular, we focus on the interplay of the following factors with the paying behavior in online game: **(1) Demographics:** how do users' demographic attributes affect their paying behavior? **(2) Social effects:** Do users who have the same relationships tend to have similar paying behavior? How do different social factors, e.g., strong/weak tie and social structural diversity, influence users' paying behavior? To which extent?

Due to limited space, we only present the analysis results in game QQSpeed, but similar results can be derived from the DNF dataset as well.

3.1 Demographics

We examine the correlation of the following demographic attributes against the paying behavior.

Gender: according to the gender of the user, we classify users into two groups - Male and Female.

Level: a number that represents the user's overall skill and experience. Improving level needs accomplishing tasks and participating in matches. By gaining a level, a user's abilities will enhance, and meanwhile he could play with users in higher levels. In QQSpeed, the user's level ranges from 1 to 200. We use a density-based discretization method [26] to classify users into three groups: Low Level, Middle Level, and High Level.

Login: represents the number of days that a user logins into the game during the two months. We evenly divide the game users into two groups - Login Bottom and Login Top to represent users who login less frequently and more frequently respectively.

Relation: an attribute that represents whether a user has at least one co-playing relationship with others.

Guild: an attribute that represents whether a user has joined a guild.

Centrality: an attribute that represents whether a user is a centrality in the network. In network analysis, centrality of a node measures its relative importance within a graph. We use the PageRank[23] algorithm to calculate the centrality scores of all the users and select the top 10% users[2] with the highest PageRank score as centrality users and the rest as ordinary user.

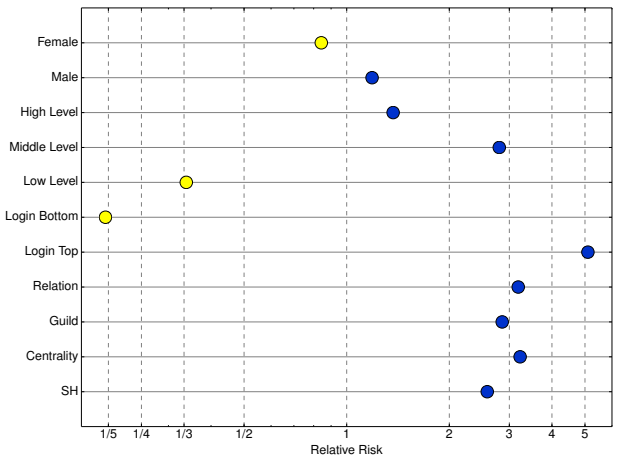


Figure 2: Correlation analysis of users' demographic attributes against the paying behavior.

SH: an attribute that represents whether a user is a structural hole spanner [7] in the network. In sociology, a structural hole spanner represents the user who connects different communities [2]. We use the HIS algorithm [22] to estimate the structural hole score (with the following formula).

$$I(v, C_i) = \max_{\substack{e_{uv} \in E, \\ S \subseteq \mathbf{C} \wedge C_i \in S}} \{I(v, C_i), \alpha_i I(u, C_i) + \beta_S H(u, S)\}$$

$$H(v, S) = \min_{C_i \in S} \{I(v, C_i)\}$$

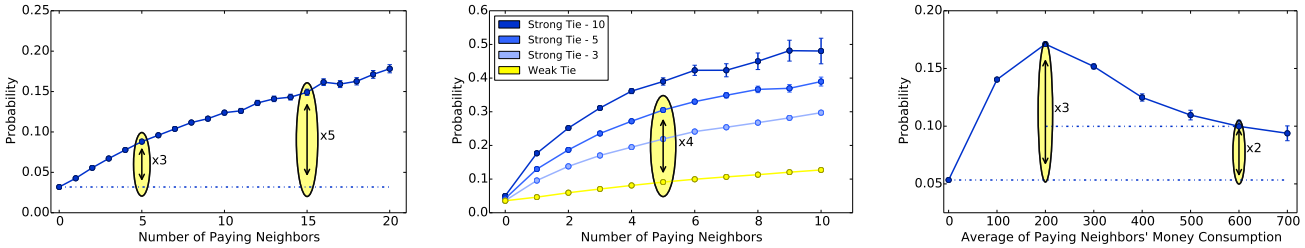
where $\mathbf{C} = \{C_1, \cdots, C_l\}$ denotes l communities in the network, $I(v, C_i) \in [0, 1]$ be the importance of v in community C_i and $H(v, S) \in [0, 1]$ as the structural hole score of v in S, i.e., the likelihood of v spanning structural holes across all communities in $S, S \subseteq \mathbf{C}$. The communities in the relationship network are detected by using Clauset-Newman-Moore algorithm [8].

Results. We use relative risk to study the effect of different attributes on user's paying behavior. Relative risk is an effect measure widely used in statistics and other fields. The relative risk of new payer associated with an attribute i can be calculated as follows:

$$RR(i) = \frac{P(\text{new payer}|\text{has attribute } i)}{P(\text{new payer}|\text{does not have attribute } i)}$$

For a specific attribute, a larger relative risk indicates that users with this attribute are more likely to become new payers. The results are given in Figure 2. We can see that the probability of a male user becoming a new payer is 1.2 times as high as that of a female user, which suggests that male users are easier to pay than female users. Moreover, users in the middle level group have higher probability to become a new payer than the others. This sounds reasonable because low level users might be unfamiliar with the game, and high level users might already play very well in the game, thus have

[2]Statistics have shown that 12% of users form core groups to actively play together[13].



(a) Social Influence (b) Strong/Weak tie (c) Paying status

Figure 3: Convert probability conditioned on paying neighbors. Y-axis: probability that a free user converts to a paying user; X-axis: (a) the number of paying neighbors, (b) the number of paying neighbors of different types of relationships, (c) the total amount of money paid by the paying neighbors.

no incentive to pay in the game. For login frequency, the probability of becoming a new payer for the Login Top group is five times higher than that of the Login Bottom group. For the attributes of relation and guild, the users who have relationship with others are 3 times higher in probability of becoming a new payer than users who do not have any relationship with others, and the users who have joined guilds are almost 3 times higher in probability of becoming a new payer than the users who do not join any guilds. For centrality and structural hole, centrality users are 300% higher in probability of becoming a new payer than other users while structural hole spanning users are 250% more likely to become a new payer than other users. Both centrality and structural hole attributes have positive correlation with users' paying behavior.

3.2 Social Effects

We investigate a number of representative sociology theories and quantitatively analyze the correlations between the users' paying behavior and these fundamental social concepts. Particularly, we focus on social relationships and social structures.

3.2.1 Social Relationship

Social Influence. The principle of social influence [16, 28] suggests that users tend to change their behavior so as to match to their friends' behaviors. Influence is an important factor that governs the dynamics of social networks. In the game network, we try to examine whether users' paying behavior will be influenced by their paying neighbors. In particular, we examine the probability of a free user becoming a new payer when he has different types of relationships with other paying users.

Figure 3(a) shows the probability that a free user becomes a new payer, conditioned on the number of paying neighbors he has in the game network. Clearly paying money is an epidemic behavior. Simply speaking, when a free user played with 5 paying users, the likelihood that he will also pay increases to 3 times higher than that he do not play with any paying users. Another trend can be learned from Figure 3(a) is that the conversion likelihood (from a free user to a paying user) continues to increase when the user has more paying neighbors. Is it really true? Will different types of relationships have different effects?

Strong/Weak Tie. We further study whether strong tie and weak tie have different influence on users' paying behavior. Strong ties are connections with people who you are close to and associate regularly with, while weak ties are more distant connections. Strong/Weak Tie [15] is one of the most basic principles in social network theories. Specifically, we classify the constructed social relationships into strong ties and weak ties by the number of times that two users played together in the game. If two users played together for more than 5 times, we call the relationship as a strong tie, otherwise a weak tie. Figure 3(b) shows some interesting results: strong ties indeed have strong influence on the paying behavior, while the influence of weak ties is relatively weak. Still take the free user with five paying neighbors for example. If the relationship is strong tie, the probability that the user converts to a new payer increases to 30%, almost 4 times higher than the case when the relationship is weak tie, and 5 times higher than that of an average user. Figure 3(b) also shows the effect comparisons of different definitions for the strong tie by varying the co-playing threshold. In our prediction, we tried different values for the threshold and finally chose 5 by cross-validation.

Status. Another interesting analysis is how the total amount of money paid by the paying neighbors would influence the user's paying behavior. We calculate the total amount of money (Chinese yuan) consumption denoted as consumption degree in the two month period for each paying user in our dataset. Then we figure out the probability of a free user becoming a new payer, conditioned on the average of paying neighbors' consumption degree, which is round to multiples of 100 for convenience in statistics. For users who do not have paying neighbors, the average of their paying neighbors' consumption degree is set to zero. We show the results in Figure 3(c), when the average of paying neighbors' consumption is 200, the probability is 300% higher than that of zero. However, as the average of paying neighbors' consumption increases, the probability of becoming new payer drops down. This is possibly because such users could receive more gift props from their "rich friends" than other users.

3.2.2 Structural Diversity

Besides peer relationships, we also study how the structure of one's personal social circle influences his paying behavior. In particular, we leverage the idea of structural diversity [29], which suggests that different inner structures of a user's neighbors have different effects on the user's behavior. We test whether the inner structure of one's paying friends will influence a free user to become a new payer. Regarding the inner structure, we count the number of connected components. We give an example in Figure 4, the centric node is the free user we are going to study. The surrounding nodes are neighbors where blue nodes indicating paying users. We can see that, B and C are connected as a component; A and D are separated as two independent components. This structure corresponds to the second one from the right in the structural

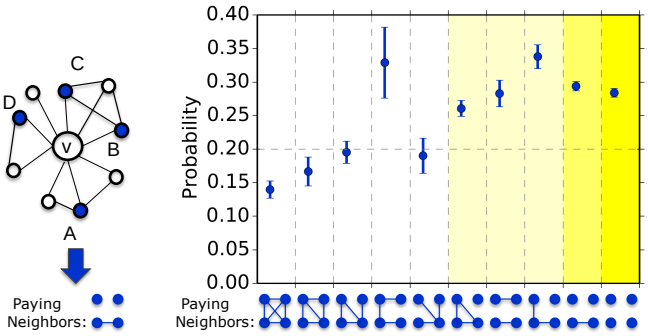


Figure 4: Convert probability conditioned on the inner structure of the paying neighbors. Y-axis: probability that a free user converts to a paying user; X-axis: the inner structure of paying neighbors by taking 4 paying neighbors as the example.

spectrum analysis. We have several intriguing discoveries from the structural spectrum analysis. First, the conversion probability (from a free user to a new payer) generally increases with the number of disconnected components. For example, when a user has a number of, e.g. 4, paying neighbors, if all the neighbors do not know each other (the last one in structural spectrum analysis), then the probability that the user will be influenced to become a new payer is almost twice higher than the case when the four neighbors know each other (the first one in structural spectrum analysis). This is somehow counterintuitive, but explainable to some extent. For an example, the first structure of the user's paying neighbors in Figure 4 indicates that all the four neighbors know each other, which implies that all the users are from the same circle. In this case, the centric user may have a feeling that he can leverage the circle to acquire some advantage (e.g., borrow an equipment from friends of the circle in the game). On the other hand, the last structure in Figure 4 implies that the four users may be from totally different circles. In this case, the centric user may feel that becoming a paying user is very common in the game, because his friends from different circles already paid money. The second discovery is the exception case when the paying neighbors are weakly connected, say A is connected to B, B to C, and C to D (the third one). The influence on the centric user's paying behavior is very strong, but also with a high variance.

4. FEATURE REPRESENTATION

Based on the analysis in § 3, we extract features from user attributes and social relationships. In addition, we extract features from users' in-game behavior logs. These features are also useful because they can capture behavioral patterns of users, say, whether the user likes to purchase props in the game, which can complement the features extracted by attributes and relationships. To summarize, the features used by our approach can be grouped into the following three categories:

User attribute features: this category of features are derived from user attributes based on the analysis in § 3.1, including:

- **Gender:** the binary feature represents the gender of the user.
- **Level:** the integer-valued feature represents the level of the users.
- **Login time:** the real-valued feature represents the length of time the user logins to the game.
- **Relation:** the binary feature represents whether the user has at least one co-playing relationship with others.
- **Guild:** the binary feature represents whether the user joins a guild.
- **Centrality:** the binary feature represents whether the user is a centrality.
- **SH:** the binary feature represents whether the user is a structural hole spanner in the game network.

Social effect features: social effect features refer to those features which model the social effects analyzed in Section 3.2, including:

- **Paying neighbor count:** the integer-valued feature represents the number of paying neighbors.
- **Strong tie count:** the integer-valued feature represents the number of paying neighbors with strong tie.
- **Weak tie count:** the integer-valued feature represents the number of weak tie paying neighbor with weak tie.
- **Average neighbor status:** the real-valued feature represents the average amount of money paid by the paying neighbors.
- **Structure diversity:** two integer-valued features respectively represent the number of paying neighbors and the number of connected components formed by paying neighbors.

In-game behavior features: this kind of features are extracted from users' in-game behavior logs, including a list of statistical summary features such as the number of purchased items and the sum of virtual money consumption. We also extract some domain specific features such as the maximum value of the user's missing items and the number of specific competitions the user participates.

5. MODEL FRAMEWORK

Factorization models have been proposed and successfully applied to recommendation and prediction tasks [1, 17, 18, 25]. However, the traditional factorization model does not model the social network information. We propose a local consistent factorization machines (LCFM) model that incorporates the social network information into the factorization model.

To present the model precisely, we introduce some necessary notations. Let $G = (V, E, W, X)$ be a social network, where V represents the set of all users, E represents the set of all relationships, and $e_{i,j} \in E$ represents a relationship between node v_i and node v_j. Each relationship $e_{i,j} \in E$ is associated with a weight $W_{i,j} \in W$, which represents the strength of the relationship. X is the set of feature vectors of all users. Based on the features defined in Section 4, each user v_i has a feature vector, denoted as $\mathbf{x}_i \in X$; the j^{th} entry in vector $\mathbf{x}_i$ is denoted as $x_{i,j}$; d represents the length of the feature vector. And $y_i \in [0, 1]$ indicates the paying potential of user v_i.

Next, we will first briefly introduce a generic factorization model, the factorization machines model, and then describe the proposed LCFM model.

5.1 Factorization Machines (FM) Model

Factorization Machines works with real valued feature vector and can leverage the interactions between variables using factorized parameters. This allows us to learn more complex interactions between variables. For example, FM can learn that female users of low level in the game have higher paying potential, whereas male users of high level in game have lower paying potential.

For feature vector $\mathbf{x}_i$, the prediction is computed by:

$$\hat{y}(\mathbf{x}_i) = w_0 + \sum_{j=1}^{d} w_j x_{i,j} + \sum_{j=1}^{d-1} \sum_{j'=j+1}^{d} x_{i,j} x_{i,j'} \langle \mathbf{p}_j, \mathbf{p}_{j'} \rangle \quad (1)$$

where $\mathbf{p}_j, \mathbf{p}_{j'}$ are two k-dimensional latent vectors and $\langle \mathbf{p}_j, \mathbf{p}_{j'} \rangle$ models the interactions between variables $x_{i,j}, x_{i,j'}$ with the dot product of the two latent vectors:

$$\langle \mathbf{p}_j, \mathbf{p}_{j'} \rangle = \sum_{l=1}^{k} p_{j,l} p_{j',l} \tag{2}$$

Given this, Eq.(1) can be also rewritten as:

$$\hat{y}(\mathbf{x}_i) = w_0 + \sum_{j=1}^{d} w_j x_{i,j} + \frac{1}{2} \sum_{l=1}^{k} \left[\left(\sum_{j=1}^{d} p_{j,l} x_{i,j} \right)^2 - \sum_{j=1}^{d} p_{j,l}^2 x_{i,j}^2 \right] \tag{3}$$

where $\Theta = \{w_0, w_1, ..., w_d, p_{1,1}, ..., p_{d,k}\}$ are the model parameters. The model has a closed model equation, and can be learned efficiently by the gradient descent method, e.g. stochastic gradient descent (SGD), based on a variety of loss functions, like square loss, hinge loss, etc. Here, we use square loss as loss function and optimize the model parameters by applying L-2 regularization on latent vector parameters to overcome the overfitting problem. The objective function is defined as follows:

$$\mathcal{O}(\Theta) = \sum_{v_i \in V} \big(\hat{y}(\mathbf{x}_i) - y_i\big)^2 + \lambda \sum_{i=1}^{d} \|\mathbf{p}_i\|^2 \tag{4}$$

where λ is a parameter that controls the regularization value. We adopt stochastic gradient descent method to solve the objective function. The partial derivative of $\hat{y}(\mathbf{x}_i)$ with respect to the model parameters can be written as:

$$\frac{\partial \hat{y}(\mathbf{x}_i)}{\partial \theta} = \begin{cases} 1, & \text{if } \theta \text{ is } w_0 \\ x_{i,j}, & \text{if } \theta \text{ is } w_j \\ x_{i,j} \sum_{f=1}^{d} p_{f,l} x_{i,f} - p_{j,l} x_{i,j}^2, & \text{if } \theta \text{ is } p_{j,l} \end{cases} \tag{5}$$

Then the parameters are updated by moving in the opposite direction of the gradient, yielding:

$$\begin{aligned} w_0 &\leftarrow w_0 - \eta \cdot 2(\hat{y}(\mathbf{x_i}) - y_i) \frac{\partial \hat{y}(\mathbf{x_i})}{\partial w_0} \\ w_j &\leftarrow w_j - \eta \cdot 2(\hat{y}(\mathbf{x_i}) - y_i) \frac{\partial \hat{y}(\mathbf{x_i})}{\partial w_j} \\ p_{j,l} &\leftarrow p_{j,l} - \eta \cdot \big(2(\hat{y}(\mathbf{x_i}) - y_i) \frac{\partial \hat{y}(\mathbf{x_i})}{\partial p_{j,l}} + 2\lambda p_{j,l}\big) \end{aligned} \tag{6}$$

where $\eta \in R^+$ is the learning rate for gradient descent. Given a training dataset, we iteratively update each parameter according to the gradient until convergence or the maximum number of iterations is reached. Then we can obtain the training model parameters $\Theta = \{w_0, w_1, ..., w_d, p_{1,1}, ..., p_{d,k}\}$.

5.2 Local Consistent FM Model

Since the standard FM model cannot utilize the network information between unlabeled users, we propose a Local Consistent FM (LCFM) model that incorporates the network information by *local consistency*. The general idea is that we assume neighborhood nodes in the network should be similar with each other, and the tendency depends on the strength of the relationship between them. Formally, based on the strength weight $W_{i,j}$, we define a *consistency degree*:

Input: Training network G_1, test network G_2, balance parameters (λ, μ), iteration numbers T_1 and T_2;
Output: estimated paying potentials $(\hat{y}_1, ..., \hat{y}_{|V_2|})$;

Initialize model parameters $\Theta \leftarrow \mathbf{0}$;
$V' \leftarrow$ Under-sampling training users $v \in V_1$;
$L \leftarrow$ a list of random shuffle $v \in V'$;
for $t = 1$ **to** T_1 **do**
  foreach $v_i \in L$ **do**
    Calculate the paying potential by Eq.(3): $\tilde{y}_i \leftarrow \hat{y}(\mathbf{x}_i)$;
    Calculate the gradient of all parameters by Eq.(5), and update parameters:
    $w_0 \leftarrow w_0 - \eta \cdot 2(\tilde{y}_i - y_i) \frac{\partial}{\partial w_0} \hat{y}(\mathbf{x_i})$;
    for $j \in \{1, ..., d\} \wedge x_{i,j} \neq 0$ **do**
      $w_j \leftarrow w_j - \eta \cdot 2(\tilde{y}_i - y_i) \frac{\partial}{\partial w_j} \hat{y}(\mathbf{x_i})$;
      for $l \in \{1, ..., k\}$ **do**
        $p_{j,l} \leftarrow p_{j,l} - \eta \cdot \big(2(\tilde{y}_i - y_i) \frac{\partial}{\partial p_{j,l}} \hat{y}(\mathbf{x_i}) + 2\lambda p_{j,l}\big)$;
      end
    end
  end
end
Initialize paying potentials of test users by Eq.(3):
for $v_i \in V_2$ **do**
  $\hat{y}_i \leftarrow \hat{y}(\mathbf{x}_i)$;
end
Propagate the paying potential scores to neighborhood:
for $t = 1$ **to** T_2 **do**
  foreach $v_i \in V_2$ **do**
    Update $\hat{y}_i$ according to Eq.(9);
  end
end

Algorithm 1: Learning and inference by LCFM

$$c_{i,j} = \frac{\log(1 + W_{i,j})}{\sum_{v_{j'} \in NB(v_i)} \log(1 + W_{i,j'})} \tag{7}$$

where $NB(v_i)$ is the set of nodes that has a relationship with node v_i, and we use a logarithm function to smooth the link strength.

Now, the problem becomes how to combine the local consistency degree into the FM model. Our strategy is to use a regularization framework to define the local consistency factor as a regularization term. In this way, we can rewrite the objective function as follows:

$$\begin{aligned} \mathcal{O}(\Theta) = \sum_{v_i \in V} \big(\hat{y}(\mathbf{x}_i) - y_i\big)^2 + \lambda \sum_{i=1}^{d} \|\mathbf{p}_i\|^2 \\ + \mu \sum_{v_i \in V} \sum_{v_j \in NB(v_i)} c_{i,j} \big(\hat{y}(\mathbf{x}_i) - \hat{y}(\mathbf{x}_j)\big)^2 \end{aligned} \tag{8}$$

where the third term of the right-hand side in the objective function is the local consistency constraint, which means linked users should have similar paying behaviors; $\mu > 0$ is a tunable parameter to balance the effect of the local consistency factor in the objective function.

Model Learning. The new objective function still has a closed model and can be solved by the SGD method. However the complexity of each iteration increases to $O(|E|kd)$, which is infeasible in most cases. Therefore, we propose a two-step approach to separately handle the FM terms and the local consistency term in the objective function.

In the first step, we optimize the FM terms in training data. As the paying behavior is sparse in games, the training data might be

rather unbalanced. Hence, we firstly conduct under-sampling [19] on the training data. Then we use SGD to solve Eq.(4), the objective function without local consistency assumption, to obtain the estimated model parameters Θ^*.

In the second step, we optimize the local consistency terms in test data. For each test node v_i, we define $\hat{y}_i$ as its estimated paying potential. With the estimated model parameters Θ^*, we set the initial value of $\hat{y}_i$ as $\hat{y}(\mathbf{x}_i)$ according to Eq.(3). To optimize the local consistency term, we use a propagation strategy to iteratively update the value of $\hat{y}_i$. In each iteration, for every $\hat{y}_i$, if $NB(v_i)$ is not empty, $\hat{y}_i$ is updated by paying potential scores of its neighbors as:

$$\hat{y}_i = (1 - \gamma\mu)\hat{y}_i + \gamma \cdot \mu \sum_{v_j \in NB(v_i)} c_{i,j}\hat{y}_j \tag{9}$$

where $\gamma \in [0, 1]$ is a parameter to control the propagation rate. The learning algorithm is summarized in Algorithm 1.

Complexity analysis. The time complexity of the standard FM model is in $O(kd)$. The complexity of the proposed model consists of two parts. The complexity of the first step is in $O(|V|T_1kd)$. And for the second step of the algorithm, the time complexity of each iteration is $O(|E|)$, and the total time complexity is $O(|E|T_2)$. Finally the total complexity is $O(|V|T_1kd + |E|T_2)$, which is much faster than directly applying SGD to solve Eq. 8, of which the time complexity is $O(|E|T_1kd)$.

6. EXPERIMENTS

In this section, we first evaluate our approach in the offline datasets described in Section 2 and compare the performance results with different approaches. We further deploy the algorithm in the real online game system and report the online test results to further evaluate the performance of our method.

6.1 Experimental Setup

Prediction Setting. We use the two datasets described in Section 2 in our experiments. Our task is to predict whether a free user will become a new payer, given the user's network information and activity logs. We split the datasets into training and test sets by time. For the QQSpeed dataset, the training set is comprised of data ranged from Jun. 20th, 2013 to Jul. 19th, 2013, and the test set is comprised of data from Jul. 20th, 2013 to Aug. 20th, 2013. For the DNF dataset, the training set is comprised of data from Apr. 1st, 2013 to May. 31th, 2013, and the test set is comprised of data from Jun. 1st, 2013 to Jun. 30th, 2013.

Comparison Methods. We compare our model with several widely used recommendation models:

Factorization Machines (FM)[25]: it uses the same features as LCFM to train factorization machines model and then employs the model to predict users' labels in the test set.

Logistic Regression (LRC)[20]: it trains a logistic regression classification model and then predicts users' labels in the test set.

SVM [14]: it trains a SVM regression model and then employs the regression model to predict whether a free user will become a new payer in the test data. For SVM, we employ LIBLINEAR [14].

Random Forest (RF)[5]: it trains a random forest model with features associated with each user. Random forest is an ensemble method. It builds a library of decision trees by random sampling instances from the train data. Each decision tree is grown by randomly selecting the features to split data upon. For random forest and the following Gradient Boosted Decision Tree (GBDT) model, we employ scikit-learn[3].

Gradient Boosted Decision Tree (GBDT)[10]: it trains a gradient boosted decision tree model with features associated with each user. GBDT is an ensemble method. It constructs an additive regression model, utilizing decision trees as the weak learner.

As for the tunable parameters k, λ, and μ in the LCFM model, we find the best configuration through cross validation in the offline datasets (i.e., $k = 10$, $\lambda = 0.1$, and $\mu = 0.1$).

Table 2: Prediction performance of different methods on the datasets.(%)

Data	Method	AUC	Recall	Precision	F1-score
QQSpeed	FM	73.61	33.16	13.62	19.31
	LRC	73.17	30.75	14.00	19.24
	SVM	72.78	32.72	14.13	19.74
	RF	73.57	33.36	13.52	19.25
	GBDT	73.64	28.88	14.44	19.25
	LCFM	**74.90**	**33.67**	**14.72**	**20.49**
DNF	FM	77.56	34.76	24.51	28.75
	LRC	77.03	34.78	24.25	28.57
	SVM	76.48	32.53	25.31	28.47
	RF	77.11	30.91	24.00	27.02
	GBDT	77.73	34.10	25.05	28.88
	LCFM	**78.32**	**35.66**	**25.71**	**29.88**

Evaluation Measures. We evaluate the performance of different approaches in terms of Precision (Prec.), Recall (Rec.), F1-Measure (F1), and Area Under Curve (AUC) [6].

All algorithms are implemented in C++ and Python, and the experiments are performed on an x64 machine with E5-4650 2.70GHz Intel Xeon CPU (with 64 cores) and 128GB RAM. The operation system is Ubuntu 12.04. The proposed algorithm has tractable running times on networks of 7,600,000 size/order of magnitude and requires 1 to 3 minutes for training and prediction.

6.2 Offline Performance

Results of Different Methods. Table 2 lists the results of different methods in the offline datasets. We can see that the proposed LCFM model clearly outperforms the baseline methods (+3-11% in terms of F1, +1-3% in terms of AUC). The advantages of the LCFM model lie in the following aspects. First, LCFM captures the interactions between variables, therefore it can estimate more accurately than models like LRC and SVM which can only model effect of single variable. Second, LCFM models interactions between variables by dot products between latent vectors, which smooths the effect of strong variables and avoids overfitting. Third, LCFM leverages the network information by the local consistency assumption, thus can better fit the data. We also conducted significant test and all the p-value < 0.01, which confirms that our method significantly outperforms the baselines. In addition, we can see that the prediction result in the DNF dataset is better than that in the QQSpeed dataset, which might be due to the reason that new payer is more sparse in the QQSpeed dataset.

Feature Contribution Analysis. In the proposed method, we consider three different categories of features: user attribute features (A), social effect features (S) and in-game behavior features

[3]http://scikit-learn.org

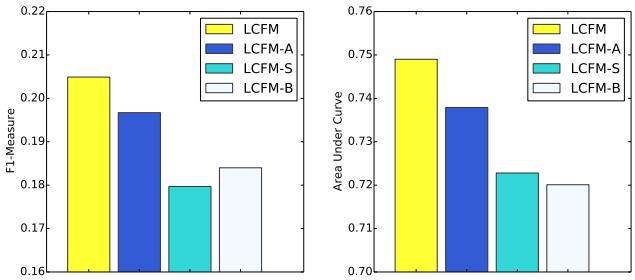


Figure 5: Feature contribution analysis based on the QQSpeed dataset

(B). Here we examine the contribution of different categories of features in LCFM model based on the QQSpeed dataset. Specifically, first we use all three categories of features to train a LCFM model, denoted as LCFM. Then we respectively remove one of the three categories of features, denoted as LCFM-A, LCFM-S, and LCFM-B. We train and evaluate the prediction performance of the different versions of the LCFM model. In Figure 5, we can observe clear drop on the performance when ignoring each category of features. This indicates that our method works well by combining different categories of features and each category of features in our method contributes improvement in the performance.

Social Effects Analysis. We further present an in-depth analysis on how different social effects affect the performance of new payer prediction. Table 3 shows the prediction performance of the proposed LCFM by considering different social effects in the QQSpeed dataset. We firstly analyze each individual social effect. In Part A of Table 3, Attribute&Behavior stands for the LCFM method by considering only user attribute features and in-game behavior features. +Social influence indicates that we add the social influence based features into the Attribute&Behavior method. +Strong/Weak tie (or +Status or +Structural diversity) indicates we further add more social based features. By incorporating each type of social effect features, we observe clear improvement compared to the Attribute&Behavior method. We can also see that the strong/weak tie features and the structural diversity features are the most useful social effect features in predicting new paying users.

Next, we analyze the combination effects of the social based features. In Part B of Table 3, "All Features" stands for the LCFM method by considering all three categories of features. -Social influence (or -Strong/Weak tie or -Status or -Structural diversity) respectively indicates that we remove individual social based features from the All Features method. It is interesting to see that the performance indeed drops but not significantly when only removing one type of the social feature. This implies that different types of social features have strong inter-correlation with each other. Another interesting phenomenon is that status seems to be a "weak" social feature to improve the prediction performance (Part A); however, without it (Part B) the model results in a most performance decrease compared to other social features.

6.3 Online Performance

We have deployed a new payer prediction system based on the proposed algorithm and applied it to several famous games such as Dungeon & Fighter Online (DNF) and QQSpeed. Basically, for each game, with users' activity logs, the system trains a LCFM model periodically on data collected from all users, and then uses the trained model to estimate the paying potential of free users.

Table 3: Social effect analysis based on the QQSpeed dataset.(%)

	Features used	AUC	Rec.	Prec.	F1
A	Attribute&Behavior	72.28	32.30	12.45	17.97
	+Social influence	74.65	33.29	14.46	20.17(+2.20%)
	+Strong/Weak tie	74.75	33.31	14.67	20.37(+2.40%)
	+Status	74.08	32.39	13.88	19.43(+1.46%)
	+Structural diversity	74.75	32.39	14.86	20.38(+2.41%)
B	All Features	74.90	33.67	14.72	20.49
	-Social influence	74.88	33.73	14.67	20.45(-0.04%)
	-Strong/Weak tie	74.88	33.19	14.80	20.48(-0.01%)
	-Status	74.77	33.15	14.66	20.33(-0.16%)
	-Structural diversity	74.89	32.90	14.84	20.45(-0.04%)

Figure 6: A screenshot of the promotion activity for online test

Finally, the game operator uses the prediction results for market decision making. To validate its effectiveness, we conducted two online tests in the game QQSpeed based on the deployed system, and compared our method with the prior strategy used in the game. The prior strategy suggests users mainly according to their activities. We use a new metric *Lift_Ratio* to evaluate our method , which can evaluate different methods at the macro-level. It is defined as

$$Lift\_Ratio = \frac{CR - CR_{prior}}{CR_{prior}} \tag{10}$$

where CR is the new payer conversion rate of a specific method and CR_{prior} is the new payer conversion rate of the prior strategy.

For each online test, we use different methods to select a *test group* of users and a *control group* of users. A user may belong to more than one group, e.g. both the test group and the control group. Then, we send invitation messages to all the selected users. A user will receive only one message even if the user belongs to more than one group. The content of the messages is to invite the user to attend a promotion activity[4] in which the user only needs to pay some amounts of money, then the user can participate in a lottery draw to get props in QQSpeed. Figure 6 shows a screenshot of the promotion activity's web page.

In online test 1, we focus on validating the effectiveness of the proposed method in online scenario, while in online test 2, we emphatically analyze the contribution of social factor.

[4]http://speed.qq.com/act/a20131220djcj/index.htm

Table 4: Results of the two online tests.

	Online Test 1 2013.12.27 - 2014.1.3		Online Test 2 2014.1.24 - 2014.1.27		
Group name	test group	control group	test group	control group	prior group
Group size	600K	200K	400K	400K	200K
#Message read	345K	106K	229K	215K	106K
Message read rate	57.50%	53.00%	57.25%	53.75%	53.00%
#Message clicked	47584	7466	23325	20922	6299
Message clicked rate	7.93%	3.73%	5.83%	5.23%	3.15%
Lift_Ratio	196.87%	0%	126.81%	73.40%	0%

Online Test 1. We select all the free users who have logined during Nov. 20th, 2013 to Dec. 20th, 2013 to form a candidate user set. Then we use the LCFM model to calculate the paying potential score of each user in the candidate user set. We select the top 600,000 users to form the *test group* and use the prior strategy to select 200,000 users to form the *control group*. We send the invitation tips on Dec. 27th, 2013, and collect results from Dec. 27th, 2013 to Jan. 3rd, 2014.

The test results are showed in Table 4. We can see that our method significantly outperforms the baseline (p-value < 0.01) and brings 196% relative improvement in the conversion rate compared to the prior selection method, which validates the effectiveness of our method in online scenario.

Online Test 2. We evaluate the contribution of social factor in the online scenario in this test using the similar method with online test 1. The candidate user set is comprised of free users who have logined during Dec. 20th, 2013 to Jan. 20th, 2014. We select three groups of users from the candidate user set. The first group is the *test group*. It contains high paying potential users recommended by the proposed LCFM model. The second group is the *control group*, which contains high score users predicted by our model by removing social effect related features. The third group is prior group, which contains users selected by the prior strategy. The invitation tips are sent on Jan. 24th, 2014, and the results are collected from Jan. 24th, 2014 to Jan. 27th, 2014.

Table 4 shows the results of online test 2. As we can see, Lift_Ratio in the test group is 172% higher than that in the control group, and the differences between the algorithms are statistically significant (p-value < 0.01), which demonstrates that social factor plays an important role in modeling and predicting users' paying behavior. Besides, we can see that the Lift_Ratio of the test group in online test 2 is lower than that in online test 1. We investigate this issue by calculating the overlap of the test groups in the two online tests. And we find that 17% of the users in the test group of online test 2 have been sent invitation tips in the first online test. Therefore, the performance decrease in online test 2 might be due to the reason that most of the users who have high predicted score and are interested in the promotion activity have already been covered in online test 1. Hence, the test group of online test 2 contains a part of the users who are not interested in the promotion activity in online test 1. Therefore the Lift_Ratio measure is lower in online test 2.

7. RELATED WORK

Online games have billions of users worldwide. Analyzing and mining the big data from online games becomes an important topic for understanding users' behaviors. We review related literature from three aspects: attribute analysis, social analysis, and game application.

Attribute Analysis. This line of research mainly focuses on studying users' demographic attributes and their effects on users' behavior in the game. For example, Yee [30] investigated how players differ from one another and how motivations of play relate to age, gender, usage patterns, and in-game behaviors. Lou et al. [21] investigated an interesting problem of "gender swapping", i.e., users choose avatars of gender opposite to their natural ones. However, all the studies ignore the social effects.

Social Analysis. Recently, more and more researchers start to analyze users' interaction in the game network. Ducheneaut and Moore [11] studied the users' interaction patterns in the online game. Son et al. [27] also provided an interesting study on an MMORPGs game. They focused on studying the interplay between distinct types of user interaction networks in the virtual world. Some other research also studies the group patterns of game users. For example, Patil et al. [24] investigated the problem of group dynamics in game social network. They developed a predictive model to predict the dynamic change of group in an online role-playing game (World of Warcraft). Ducheneaut et al. [12] used longitudinal data collected from the online game to examine play and grouping patterns. They found several interesting patterns that affect the formation and longevity gaming communities. Tang et al. [28] studied how to quantify social influence in large social networks. However, all of the aforementioned works do not consider the users' paying behavior. In practice, users' play patterns and social activities in the game are strongly affected by their paying status.

Game Application. The game-related approaches have also been used in many other fields such as web search and image recognition. For example, Bennett et al. [4] studied users' collaborative behavior in the labeling game, where users get reward if they provide consensus ranking labels of images for given queries. Arase et al. [3] developed a game based approach to attract users to label geographical relevance of web images. Daniel et al. [9] aimed to collect a crowd of users by leveraging games as a tool, and their focus is to develop an underlying data management framework.

8. CONCLUSION

In this paper, we study an interesting problem of predicting potential new paying users in online game networks. By investigating several social patterns and their effects on users' paying behavior, we found strong social influence on users' paying behavior in the game network. Based on the discoveries, we develop a local consistent factorization machines(LCFM) model by incorporating the network information into the factorization model. The model can effectively recognize potential paying users and significantly out-

performs alternative methods. We have deployed the model in online games and the online test results further confirm the effectiveness of the proposed model.

Using social network techniques to study online gaming data is an intriguing research direction. As for the future work, it would be interesting to connect users' virtual network in the game with their physical daily life to study how online affects offline. It would be also interesting to connect the study with other social theories to further investigate dynamic changes of the network structure in online games. As for the model, it is also important to develop an interactive learning mechanism so that the model can directly incorporate users' feedback.

ACKNOWLEDGMENTS

The work is supported by the National High-tech R&D Program (No. 2014AA015103), National Basic Research Program of China (No. 2014CB340500), Natural Science Foundation of China (No. 61222212), and a research fund of Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology.

9. REFERENCES

[1] D. Agarwal and B.-C. Chen. Regression-based latent factor models. In *KDD'09*, pages 19–28, 2009.

[2] C. C. Aggarwal. *An introduction to social network data analytics*. Springer, 2011.

[3] Y. Arase, X. Xie, M. Duan, T. Hara, and S. Nishio. A game based approach to assign geographical relevance to web images. In *WWW'09*, pages 811–820, 2009.

[4] P. N. Bennett, D. M. Chickering, and A. Mityagin. Learning consensus opinion: mining data from a labeling game. In *WWW'09*, pages 121–130, 2009.

[5] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[6] C. Buckley and E. M. Voorhees. Retrieval evaluation with incomplete information. In *SIGIR'2004*, pages 25–32, 2004.

[7] R. S. Burt. *Structural holes: The social structure of competition*. Harvard University Press, 2009.

[8] A. Clauset, M. E. Newman, and C. Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.

[9] D. Deutch, O. Greenshpan, B. Kostenko, and T. Milo. Declarative platform for data sourcing games. In *WWW'12*, pages 779–788, 2012.

[10] F. Diaz, D. Metzler, and S. Amer-Yahia. Relevance and ranking in online dating systems. In *SIGIR'10*, pages 66–73, 2010.

[11] N. Ducheneaut and R. J. Moore. The social side of gaming: a study of interaction patterns in a massively multiplayer online game. In *CSCW'04*, pages 360–369, 2004.

[12] N. Ducheneaut, N. Yee, E. Nickell, and R. J. Moore. Alone together?: exploring the social dynamics of massively multiplayer online games. In *CHI'06*, pages 407–416, 2006.

[13] N. Ducheneaut, N. Yee, E. Nickell, and R. J. Moore. Alone together?: exploring the social dynamics of massively multiplayer online games. In *CHI'06*, pages 407–416, 2006.

[14] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874, 2008.

[15] M. S. Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973.

[16] H. C. Kelman. Compliance, identification, and internalization: Three processes of attitude change. *Journal of Conflict Resolution*, 2(1):51–60, 1958.

[17] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *KDD'08*, pages 426–434, 2008.

[18] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[19] M. Kubat, S. Matwin, et al. Addressing the curse of imbalanced training sets: one-sided selection. In *ICML*, volume 97, pages 179–186, 1997.

[20] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting positive and negative links in online social networks. In *WWW'10*, pages 641–650, 2010.

[21] J.-K. Lou, K. Park, M. Cha, J. Park, C.-L. Lei, and K.-T. Chen. Gender swapping and user behaviors in online social games. In *WWW'13*, pages 827–836, 2013.

[22] T. Lou and J. Tang. Mining structural hole spanners through information diffusion in social networks. In *WWW'13*, pages 825–836, 2013.

[23] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: bringing order to the web. Technical report, Stanford InfoLab, 1999.

[24] A. Patil, J. Liu, and J. Gao. Predicting group stability in online social networks. In *WWW'13*, pages 1021–1030, 2013.

[25] S. Rendle. Factorization machines with libfm. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(3):57, 2012.

[26] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. *Data Mining and Knowledge Discovery*, 2(2):169–194, 1998.

[27] S. Son, A. R. Kang, H.-c. Kim, T. Kwon, J. Park, and H. K. Kim. Analysis of context dependence in social interaction networks of a massively multiplayer online role-playing game. *PloS one*, 7(4):e33918, 2012.

[28] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *KDD'09*, pages 807–816, 2009.

[29] J. Ugander, L. Backstrom, C. Marlow, and J. Kleinberg. Structural diversity in social contagion. *PNAS*, 109(16):5962–5966, 2012.

[30] N. Yee. Motivations for play in online games. *CyberPsychology & Behavior*, 9(6):772–775, 2006.