Modelling 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[+]

[+]Tsinghua University

[#]Tencent Corporation

[*]Auckland University of Technology

[-]Indiana University

Billion Dollar Industry

- Facebook[1]
  - 250 million monthly players
  - 200 games with >1 million active players
  - 12% revenue
- Tencent[2] (Market Cap: ~150B $)
  - >400 million players
  - 50% revenue

[1] Facebook 2013 First Quarter Report
[2] Tencent 2013 Anual Report

Billion Dollar Industry

- Facebook[1]
  - 250 million monthly players

Not only keep players playing, but also make them pay.

- Tencent[2] (Market Cap: ~150B $)
  - >400 million players
  - 50% revenue

[1] Facebook 2013 First Quarter Report
[2] Tencent 2013 Anual Report

What we do

- Given users' data in online games, predict:

Free users -> Paying users

What we do

- Given users’ data in online games, predict:

Free users -> Paying users

- Our goal:
  - Fundamental factors
  - Social effect
  - Predictive model

Two games: DNF

- Dungeon & Fighter Online (DNF)
  - Fight enemies by individuals or groups
  - 400+ million users
  - 2nd largest online game in China.

Two games: QQ Speed

- QQ Speed
  - Car racing against other users
  - 200+ million users

Datasets

- 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

Datasets

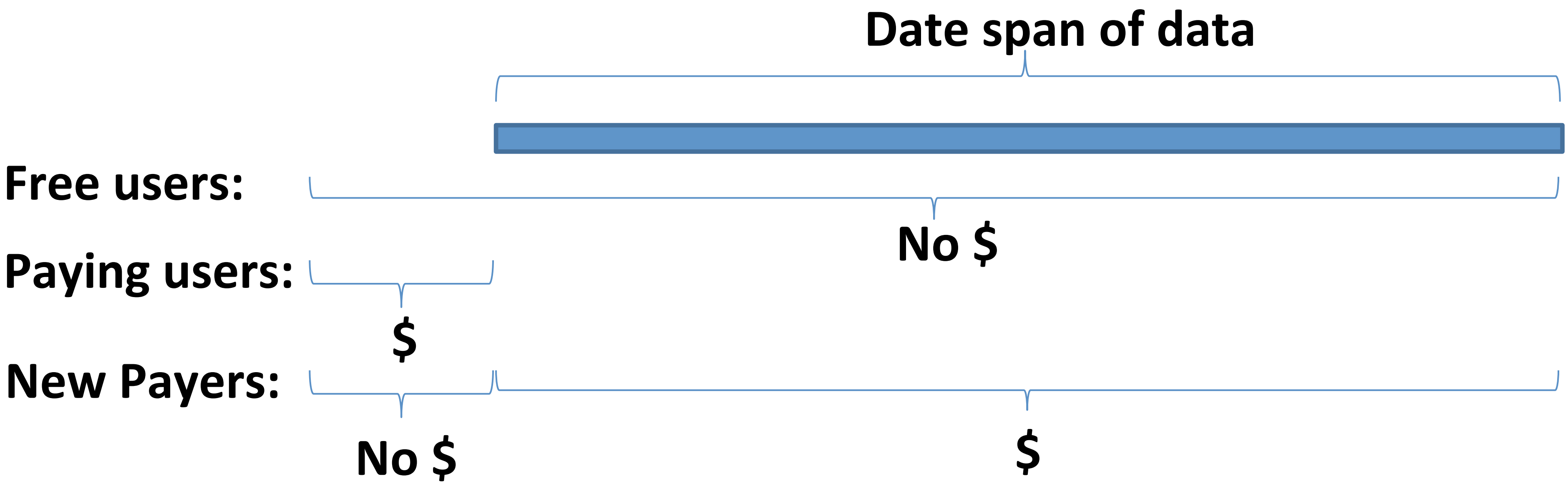
- Statistics of the datasets

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

Datasets

- 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	$\sim 10^5$	$\sim 10^4$



Datasets

- 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

Observation – Two Questions

- How do demographic attributes affect users' paying behavior?

- How do social factors influence users' paying behavior?

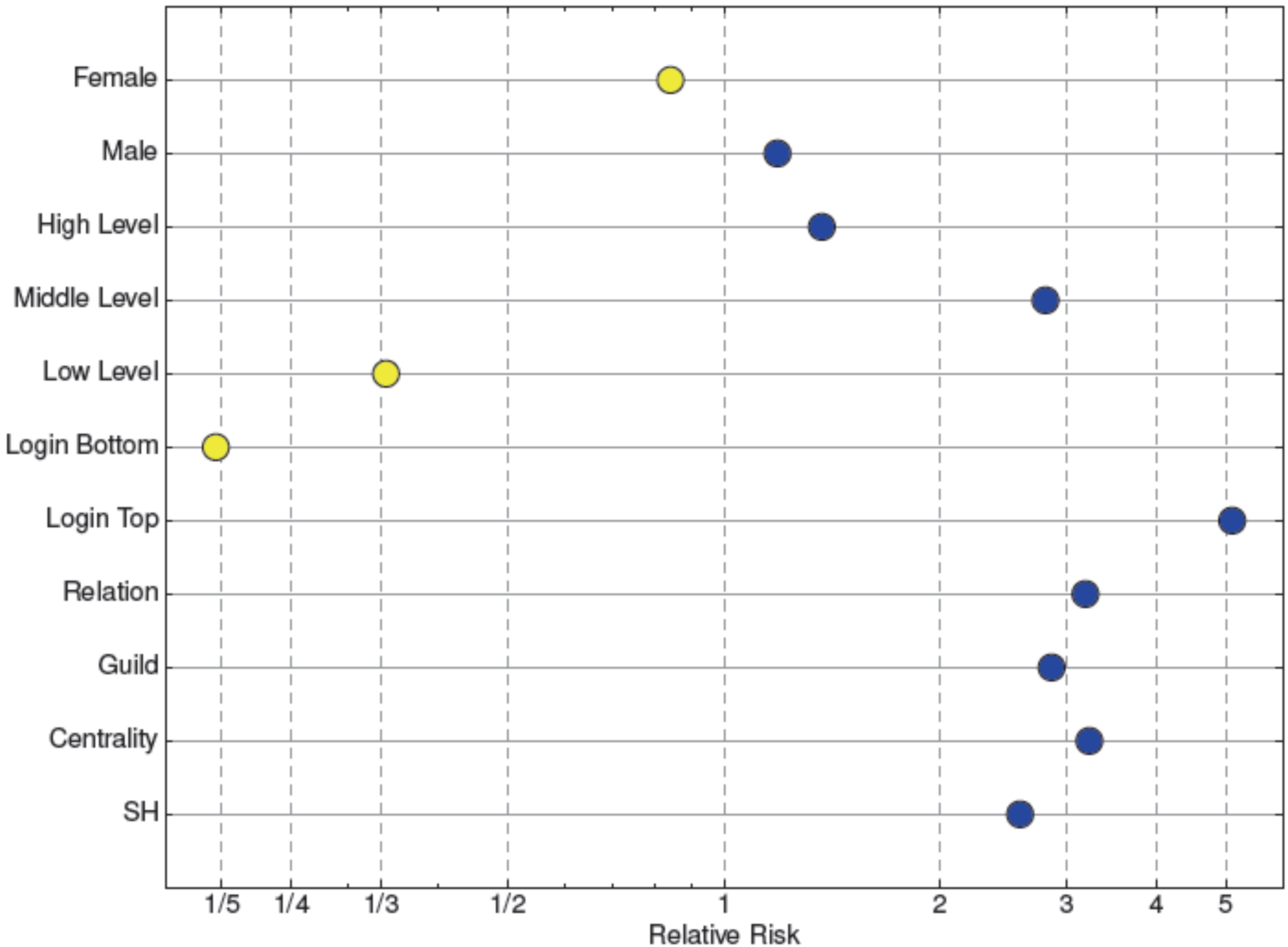
Observation - Demographics

- Relative risk for attribute i:

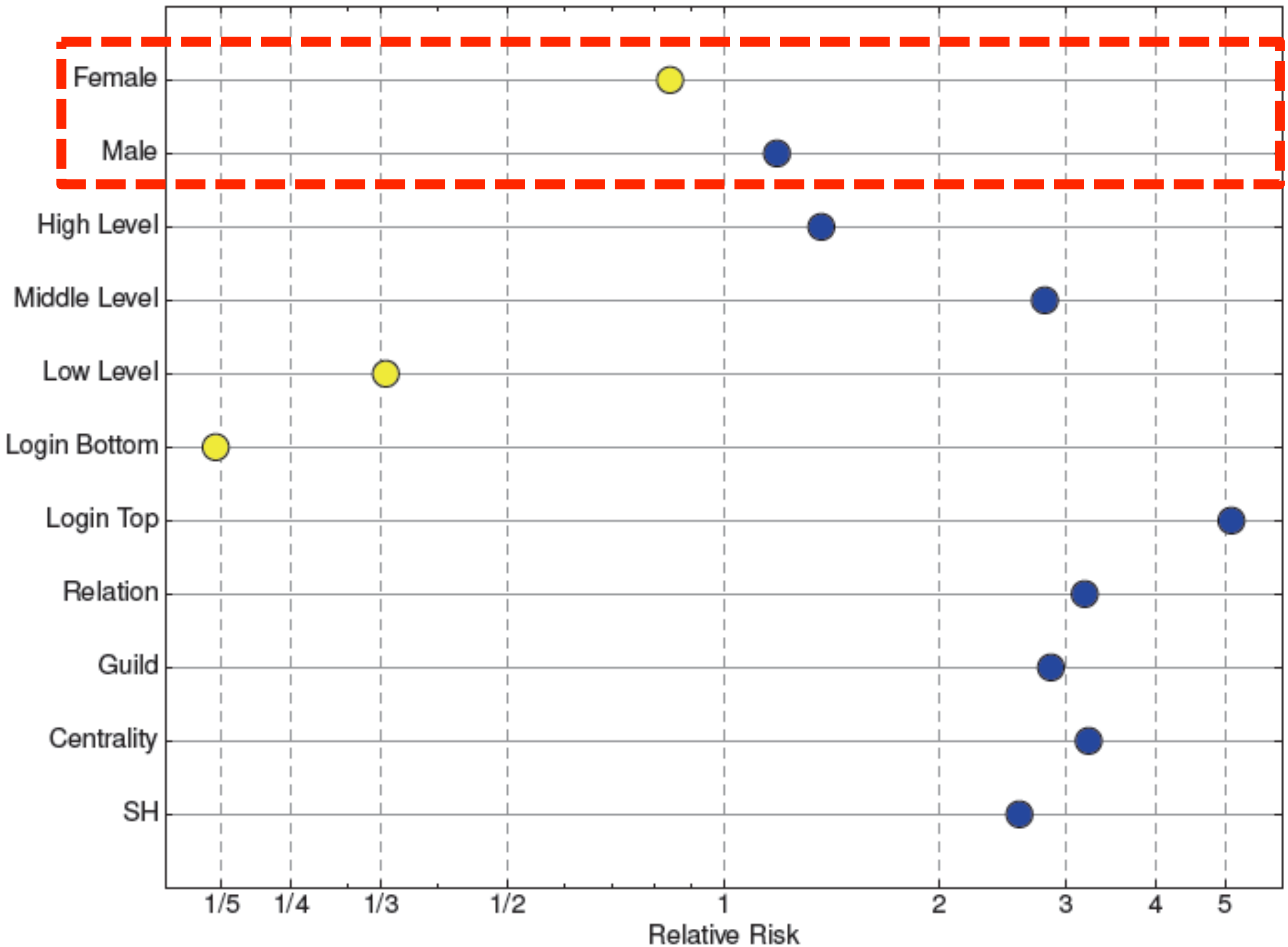
$$RR(i) = \frac{P(\text{new payer}|\text{has attribute } i)}{P(\text{new payer}|\text{does not have attribute } i)}$$

- RR(i) > 1: **more** likely to become paying users
- RR(i) < 1: **less** likely to become paying users

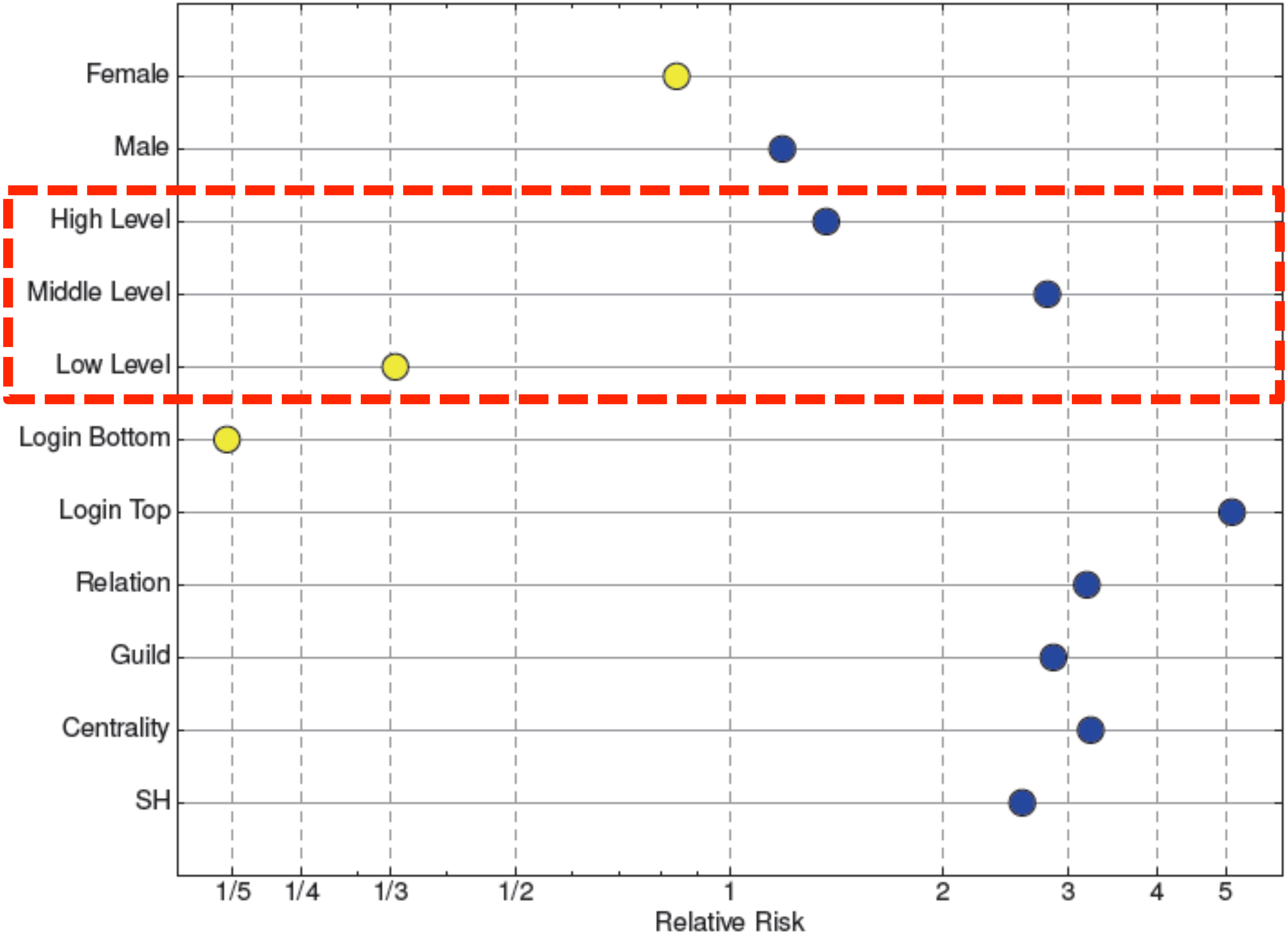
Observation - Demographics



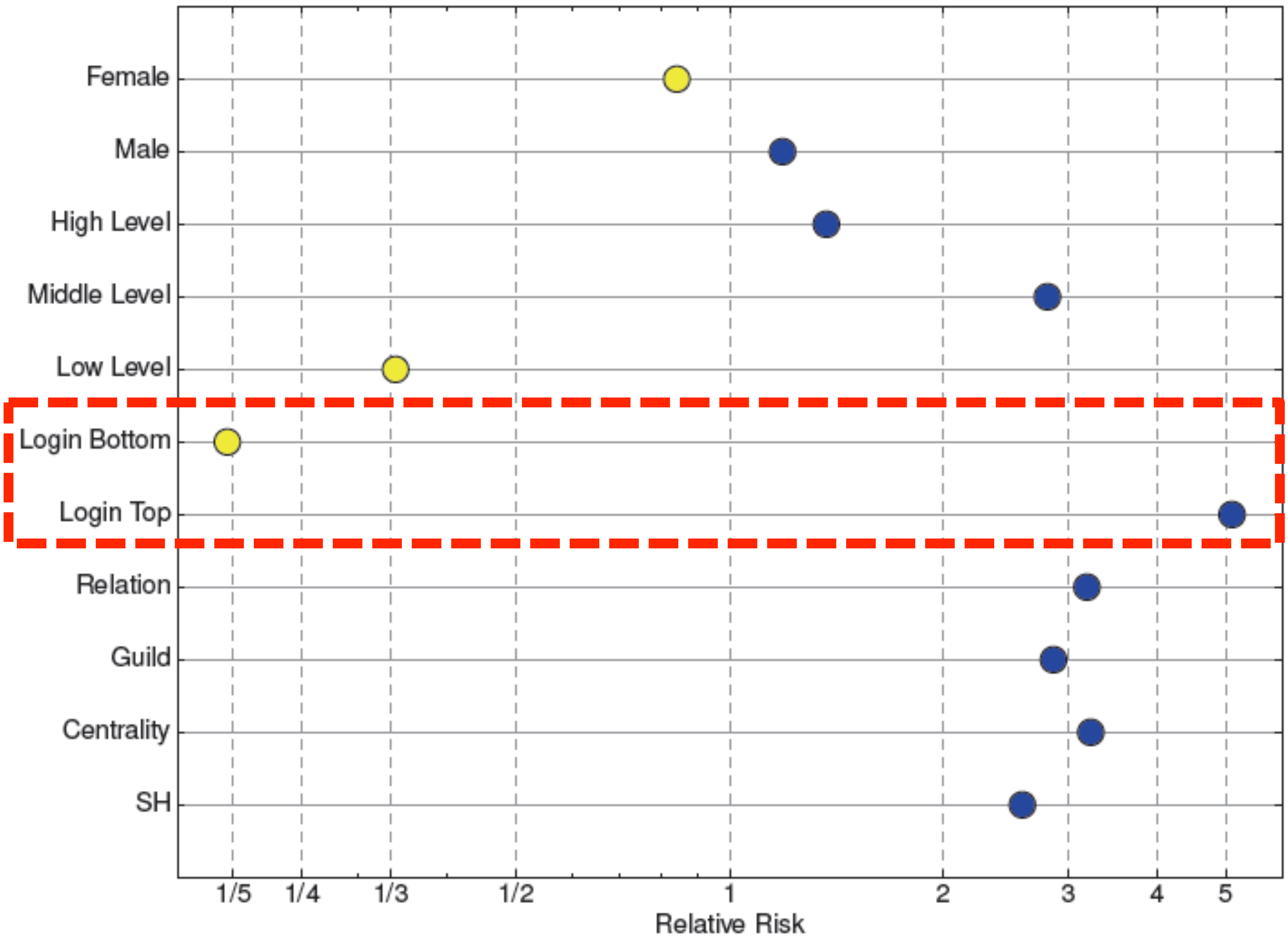
Observation - Demographics



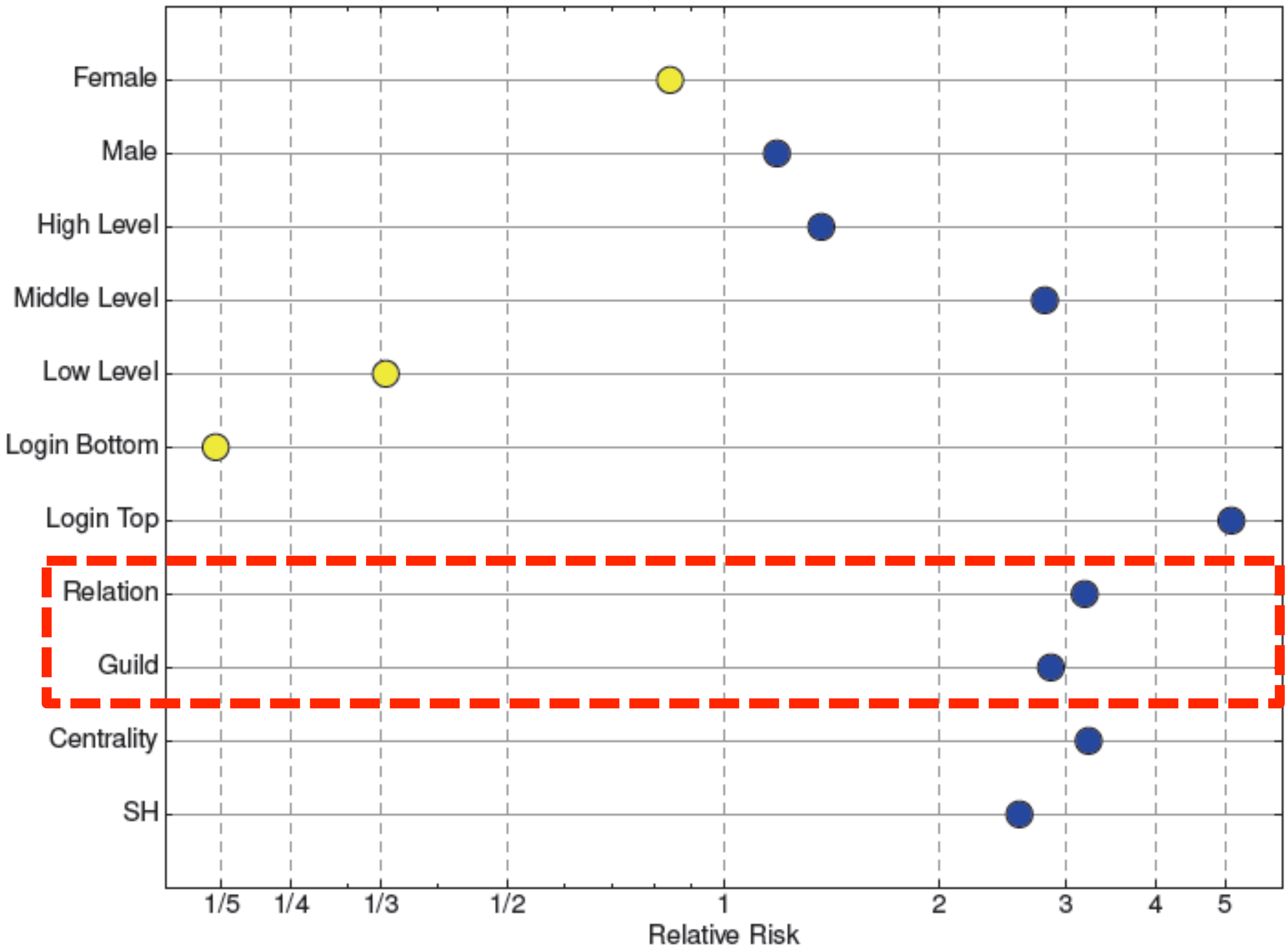
Observation - Demographics



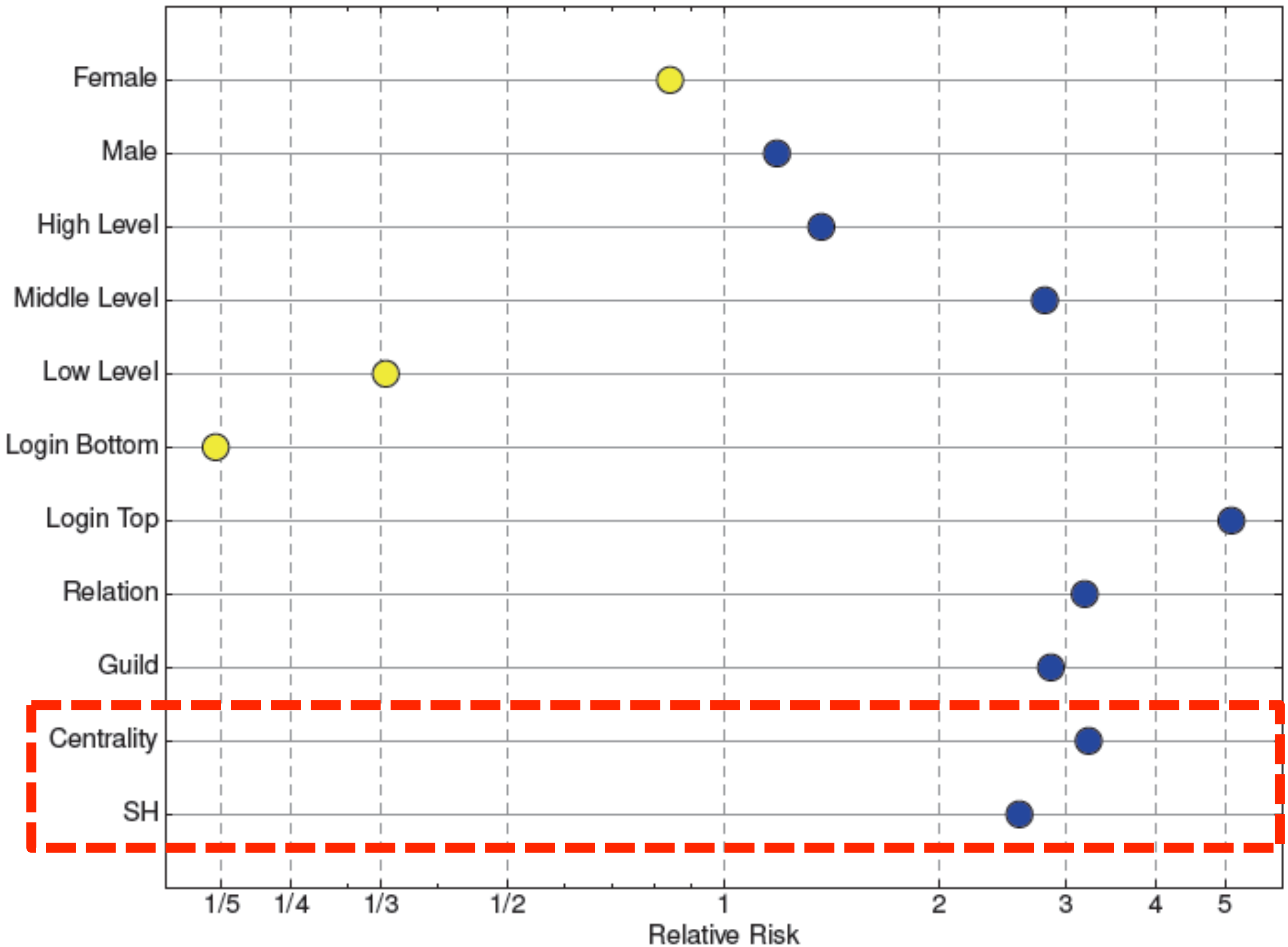
Observation - Demographics



Observation - Demographics



Observation - Demographics

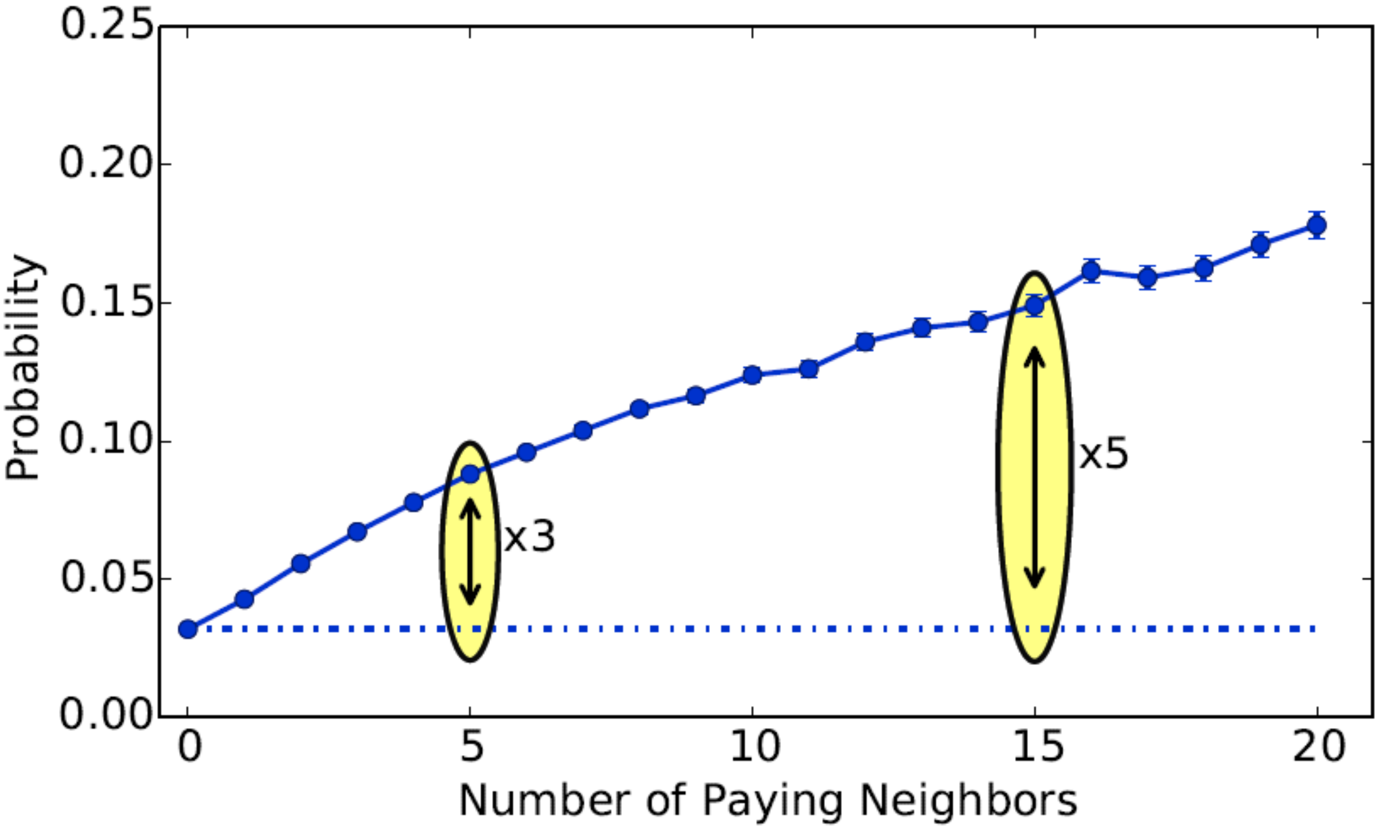


Observation – Social Effects

- Social network construction
  - Co-playing network

- Social relationship
  - Social influence
  - Strong/Weak tie
  - Status
- Structural diversity

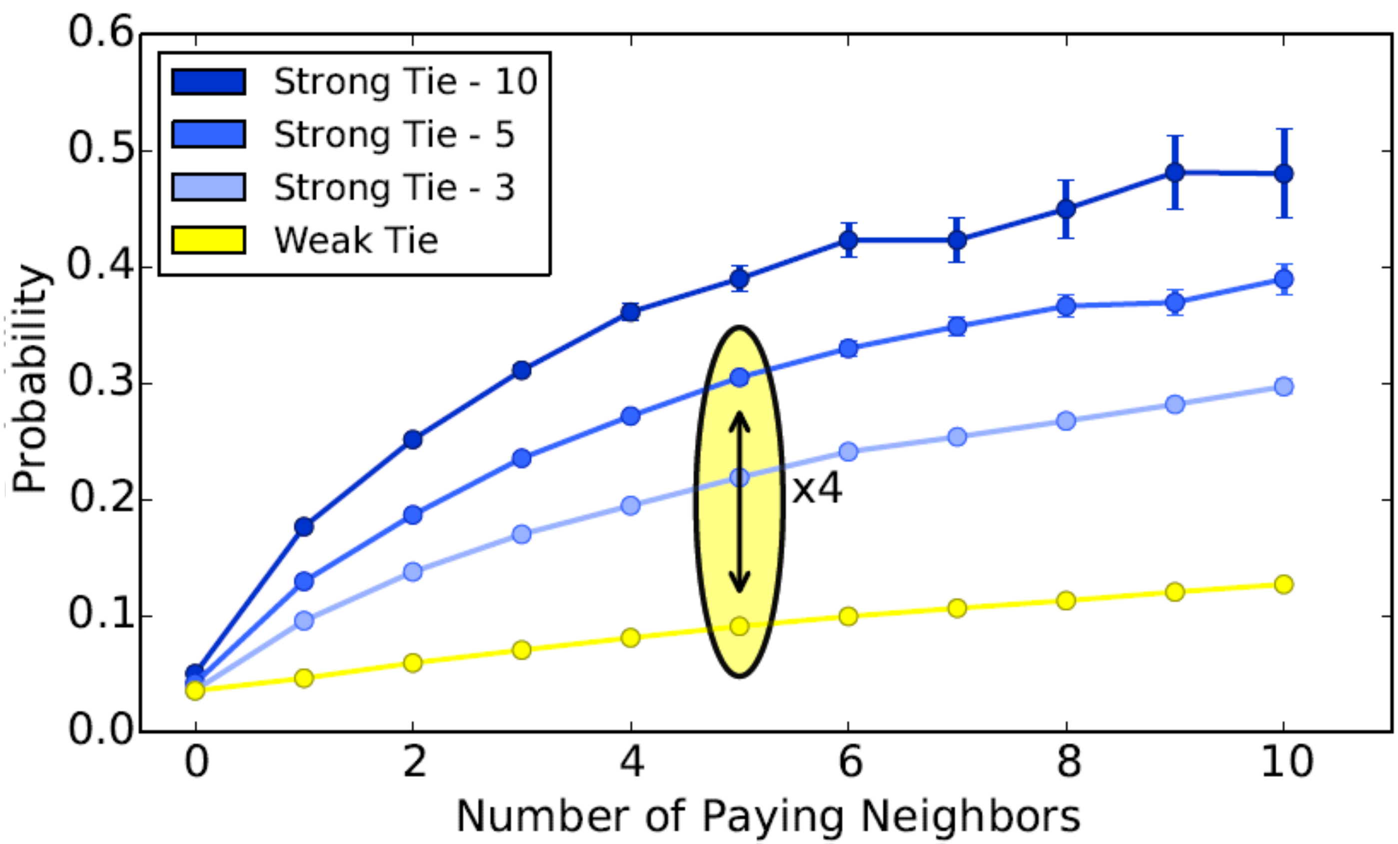
Social Relationship – Social Influence



Y: probability that a free user converts to a new payer

More paying neighbors ➡ Higher conversion probability

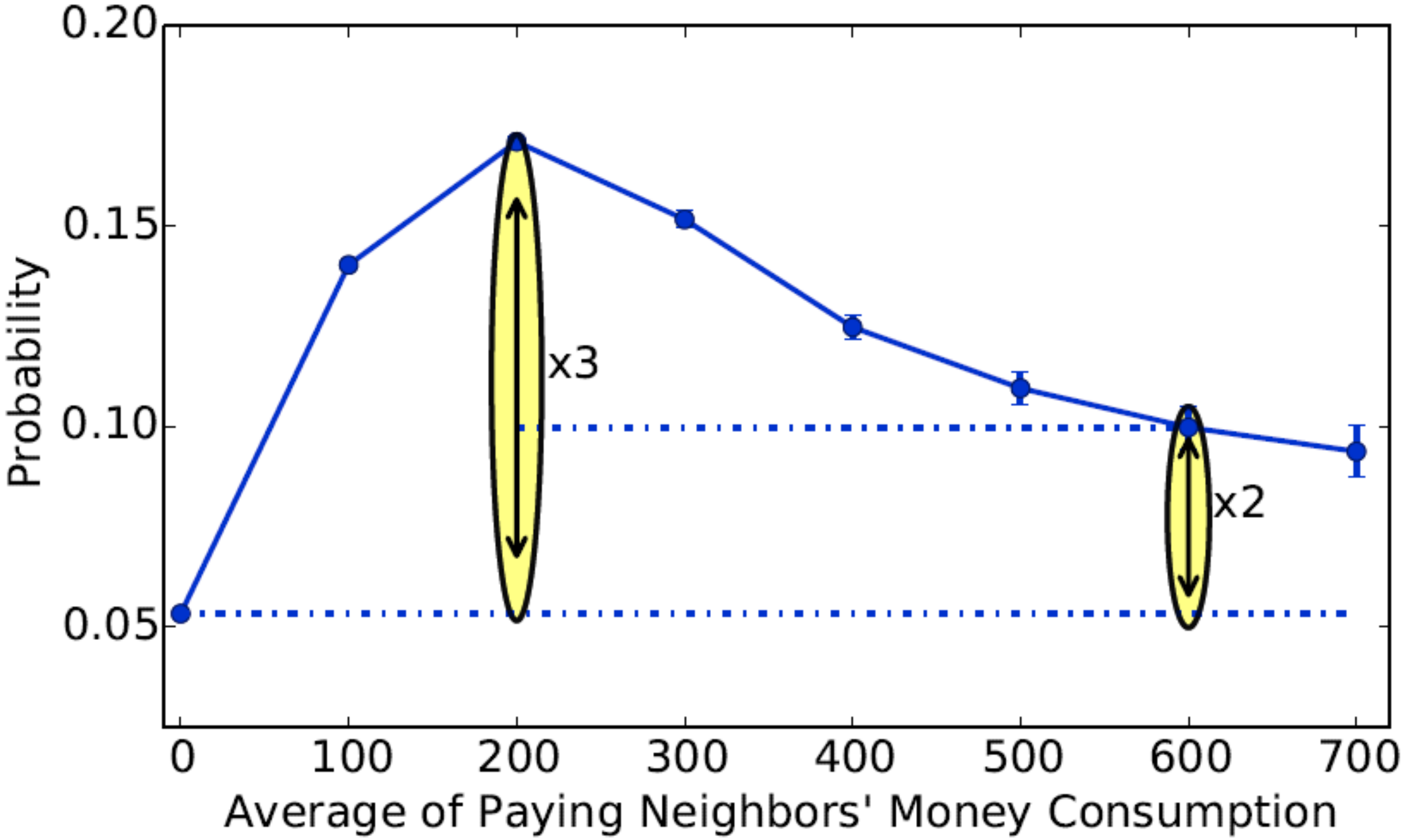
Social Relationship – Strong/Weak Tie



Y: probability that a free user converts to a new payer

Strong ties ➡ Strong influence

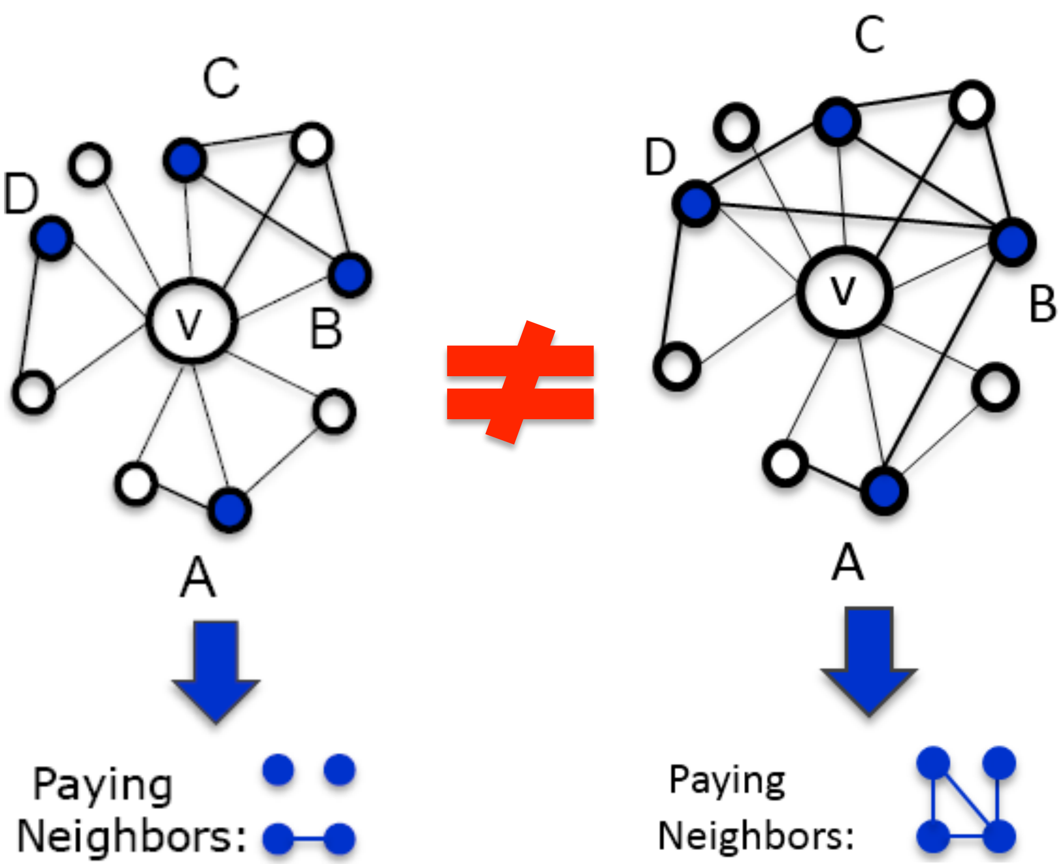
Social Relationship – User Status



Y: probability that a free user converts to a new payer

Neighbors’ money consumption increases ➡ Conversion probability follows a unimodal shape

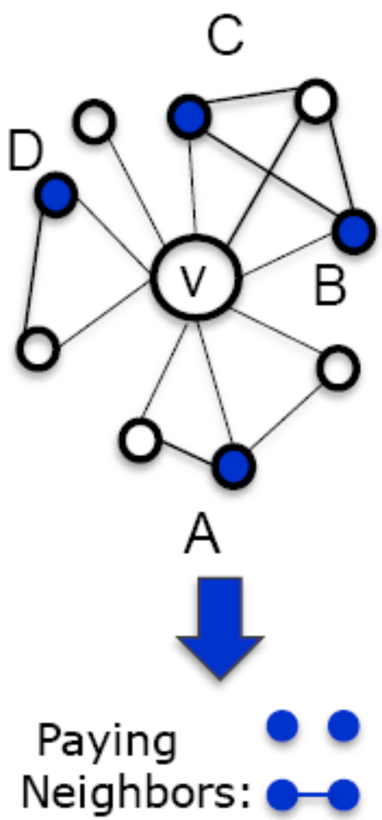
Structure Diversity

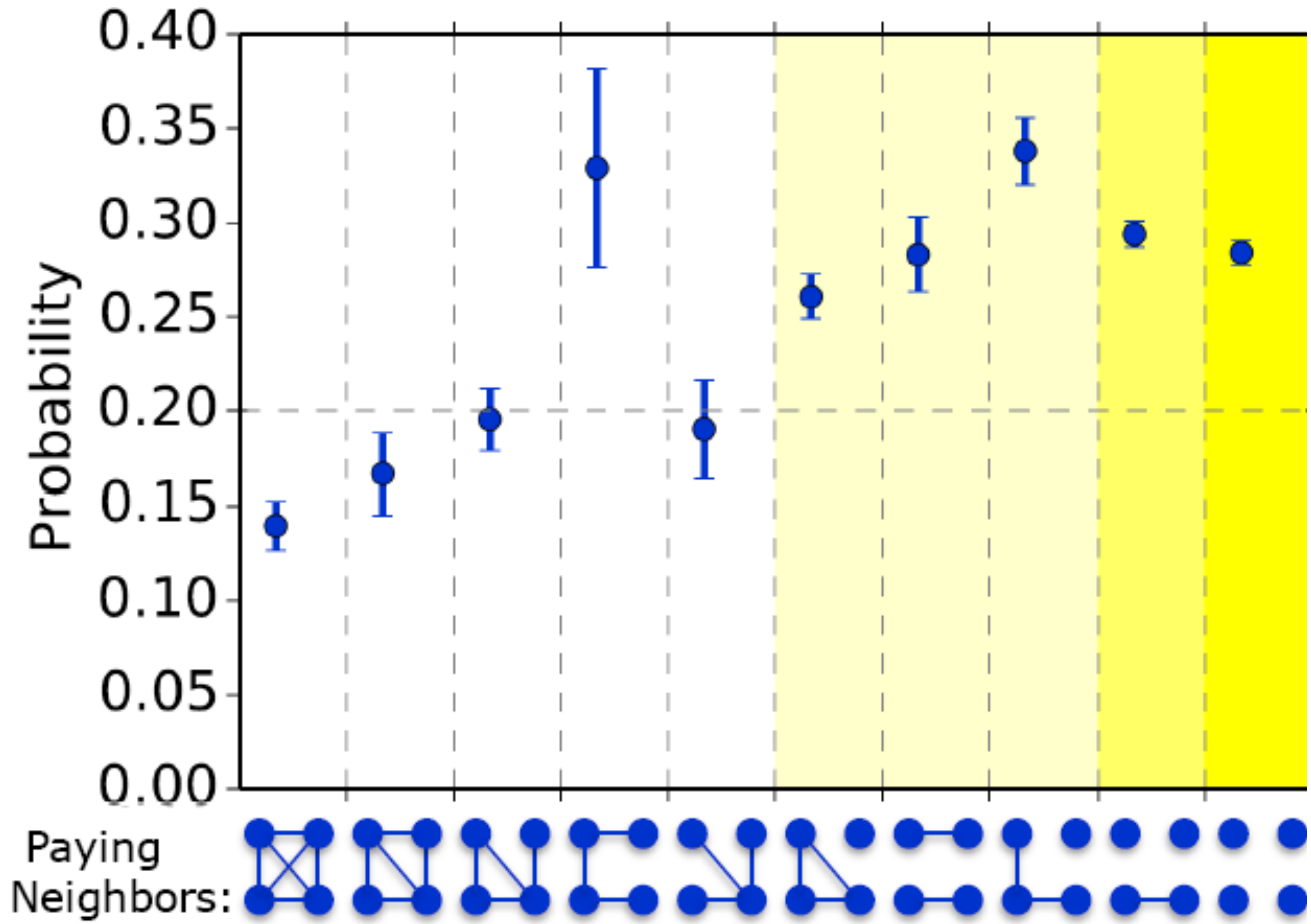


Different structures of a user's neighbors have different effects on the user's behavior[1]

[1] Ugander, J., Backstrom, L., Marlow, C., & Kleinberg, J. Structural diversity in social contagion. In PNSA'12.

Structure Diversity





Extracted Features

- User attributes features
- Social effect features
- In-game behavior features
  - #purchased items
  - sum of virtual money consumption
  - etc.

Model Framework - Notations

- $G=(V,E,W,\boldsymbol{X})$ be a social network.
- $W_{i,j} \in W$: weight on edge $e_{i,j} \in E$
- $\boldsymbol{x}_i \in X$: feature vector for user v_i
- $y_i \in Y$: paying potential for user v_i

Input:
$G = (V, E, W, \mathbf{X})$

Output:
Y

Factorization Machines

- The prediction for feature vector $x↓i$:

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

- Model parameters:

$$\Theta = \{w_0, w_1, ..., w_d, p_{1,1}, ..., p_{d,k}\}$$

Factorization Machines

- The prediction for feature vector x_i:

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

- Model parameters:

$$\Theta = \{w_0, w_1, ..., w_d, p_{1,1}, ..., p_{d,k}\}$$

- It can be rewritten as:

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

Factorization Machine (cont')

- Objective function:

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

- Solve by Stochastic Gradient Descent (SGD)

$$\frac{\partial \hat{y}(\mathbf{x})}{\partial \theta} = \begin{cases} 1, & \text{if } \theta \text{ is } w_0 \\ x_i, & \text{if } \theta \text{ is } w_i \\ x_i \sum_{j=1}^{d} p_{j,l} x_j - p_{j,l} x_i^2, & \text{if } \theta \text{ is } p_{i,l} \end{cases}$$

Local Consistent FM Model

- Consistency degree between two nodes:

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

- Incorporate the local consistency factor by a regularization term:

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

Model Learning – Two-step approach

- First step
  - Optimize the FM terms in training data by SGD.

.

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

Model Learning – Two-step approach

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

$$\boxed{+ \mu \sum_{v_i \in V} \sum_{v_j \in NB(v_i)} c_{i,j} \big(\hat{y}(\mathbf{x}_i) - \hat{y}(\mathbf{x}_j)\big)^2}$$

- Second step
  - Optimize the local consistency terms by local propagation.

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

Where $\gamma \in [0, 1]$ is a parameter to control the propagation rate.

Time Complexity

- Our approach: $O(|V|T_1kd + |E|T_2)$
- Directly apply SGD: $O(|E|T_1kd)$

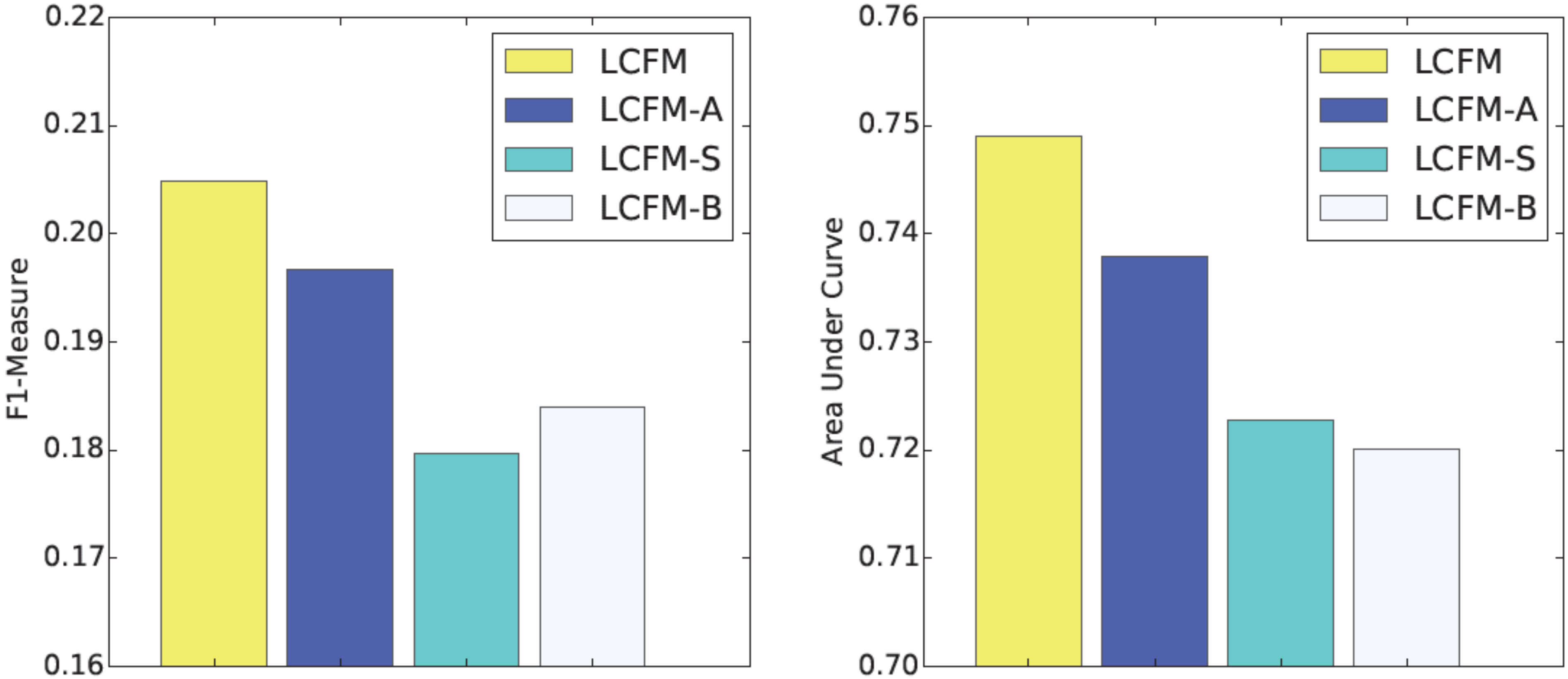
Experimental Setup

- Prediction setting
  - Predict whether a free user will become a new payer
  - Split the datasets into training and test sets by time
- Evaluation measures
  - Precision (Prec.)
  - Recall (Rec.)
  - F1-Measure (F1)
  - Area under Curve (AUC)

Results of Different Methods

Data	Method	AUC	Rec.	Prec.	F1
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**

Feature Contribution



LCFM-A: stands for removing attribute features

LCFM-S: stands for removing social effect features

LCFM-B: stands for removing in-game behavior features

Social Effect Contribution

	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%)

Social Effect Contribution

	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%)

Online Test

- Test setting
  - Two groups: *test group* and *control group*.
  - Send messages to invite the user to attend a promotion activity.

Online Test

- Test setting
  - Two groups: *test group* and *control group*.
  - Send messages to invite the user to attend a promotion activity.
- Evaluation metric:

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

where CR means the new payer **c**onverting **r**ate.

Prior strategy: suggests users mainly by their activities.

Online Test

- Test setting
  - Two groups: *test group* and *control group*.
  - Send messages to invite the user to attend a promotion activity.
- Evaluation metric:

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

where CR means the new payer converting rate.

Prior strategy: suggests users mainly by their activities.

Online Test Results

- Online test 1
  - Test the effectiveness of our approach in online scenario.
  - Test group: LCFM
  - Control group: Prior strategy

	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 Results

- Online test 2
  - Test the contribution of social factors in online scenario.
  - Test group: LCFM
  - Control group: LCFM – Social effect features

	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%

Conclusion

- Discovered strong social influence on users' paying behavior in the game network.
- Proposed a LCFM model that incorporates network information into FM model.
- Confirmed the effectiveness of our approach by online test results.

Thank you!

Online gaming is one of the largest industries on the Internet…

- Facebook
  - 250 million users play games monthly
  - 200 games with more than 1 million active users

Not only keep players playing, but also make them pay.

- Tencent (Market Cap: ~150B $)
  - More than 400 million gaming users
  - 50% of Tencent's overall revenue is from online games

Online games are become more and more *social*…

- Social activity has already become one of the most important elements in designing online games.
  - Statistics show that 80% of Zynga’s revenue comes from Facebook users.

What we do

- Given users' data in online games, predict:

Free users -> Paying users

- Precisely, we aim to answer:
  - What are the fundamental factors that trigger free users to play?
  - 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?

Online games are become more and more *social…*

- Statistics show that 80% of Zynga's revenue comes from Facebook users.

Online games are become more and more *social…*

- Statistics show that 80% of Zynga's revenue comes from Facebook users.

How to model paying behavior in game social networks?

Online games are become more and more *social*…

- Statistics show that 80% of Zynga's revenue comes from Facebook users.

How to model paying behavior in game social networks?

Free users -> Paying users

Challenges

- Sparsity
  - Only 3% of the users in Zynga have purchased credits in the game.
- Social effect
  - How is users' paying behavior influenced by friends and the social structure?
- Predictive models
  - How to develop methods that can effectively identify potential paying users?

Related Work

- Attribute analysis
  - Motivation for play [Yee CPB'06]
  - Gender swapping [Lou WWW'13]
- Social Analysis
  - Interaction patterns [Ducheneaut CSCW'04]
  - Grouping patterns [Ducheneaut CHI'06]
  - Group stability [Patil WWW'13]
  - Types of interaction networks [Son PloS one 12]

Two games: QQ Speed

- QQ Speed
  - A racing game that users can take part in competitions to play against other users.
  - 200+ million users.
  - Users can race against other users by individuals or forma a group to race together.

Datasets

- Statistics of the datasets

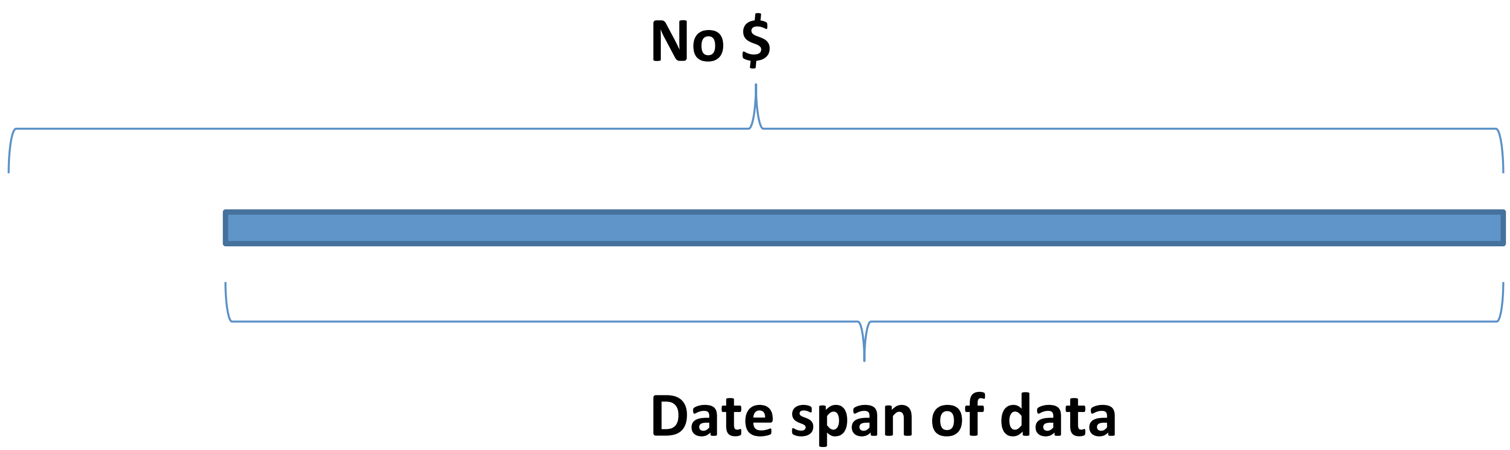
Category	Type	QQSpeed	DNF
User	all users	7.60M	347K
	free users	$\sim 10^6$	$\sim 10^5$
	paying users	$\sim 10^6$	$\sim 10^5$
	new payers	$\sim 10^5$	$\sim 10^4$
Relationship	co-playing	134M	7.30M
Guild	guilds	600K	49.6K
	co-guild	66.7M	51.7M
Activity	activity types	58	64
	activity logs	44.7B	5.71B
Date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

Datasets

- 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	$\sim 10^5$	$\sim 10^4$

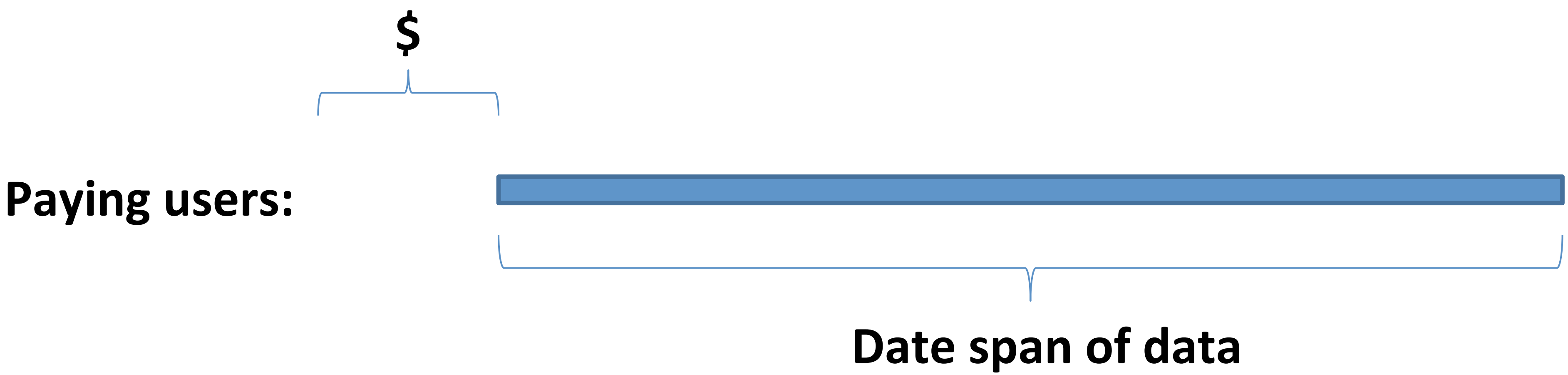
Free users:



Datasets

- Statistics of the datasets

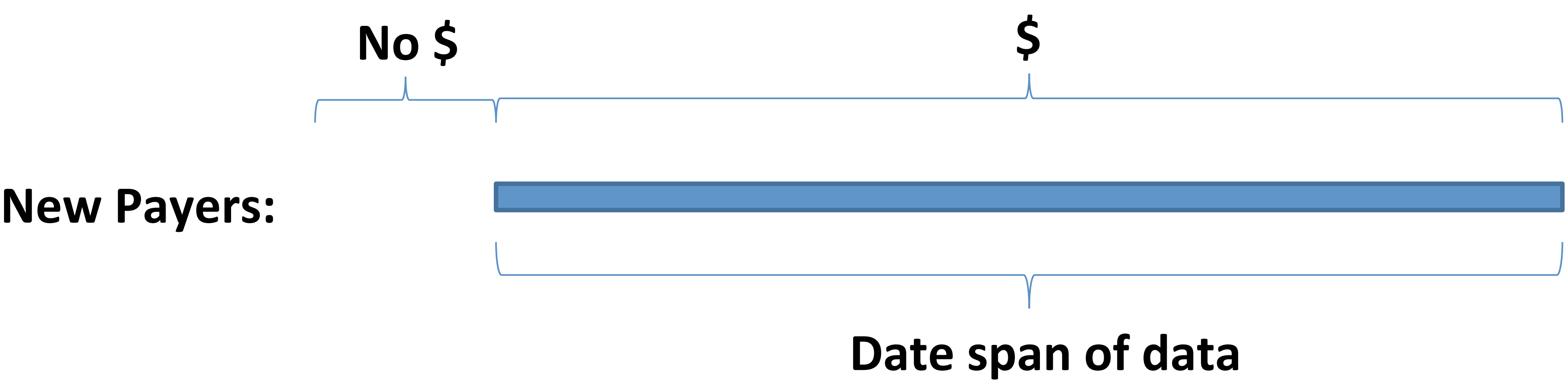
Category	Type	QQSpeed	DNF
User	all users	7.60M	347K
	free users	$\sim 10^6$	$\sim 10^5$
	paying users	$\sim 10^6$	$\sim 10^5$
	new payers	$\sim 10^5$	$\sim 10^4$



Datasets

- 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	$\sim 10^5$	$\sim 10^4$



Social Relationship – Social Influence

- **Social influence**: users tend to change their behavior so as to match to their friends' behavior.

- Examine the probability of a free user becoming a new payer, conditioned on the number of paying neighbors in the game network.

Social Relationship – Strong/Weak Tie

- Strong tie: connections with people who you are close to and associate regularly with.
- Weak tie: more distant connections.

- Classify the relationships into strong/weak ties by the number of times that two users played together in the game.

Social Relationship – User Status

- User status: the total amount of money consumption in the two month period.

- Examine the probability of a free user becoming a new payer, conditioned on the average of paying neighbors' status.

Factorization Machines

- The prediction for feature vector x_i :

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

MF

$w_{jj'}$

- Model parameters:

$$\Theta = \{w_0, w_1, \ldots, w_d, p_{1,1}, \ldots, p_{d,k}\}$$

Model Framework - Notations

- Let $G=(V,E,W,X)$ be a social network.
- $e_{i,j} \in E$ represents a relationship between node v_i and node v_j.
- Each $e_{i,j} \in E$ is associated with a weight $W_{i,j} \in W$.
- Each user v_i has a feature vector $\boldsymbol{x}_i \in X$; the j^th^ entry in $\boldsymbol{x}_i$ is $x_{i,j}$.
- d represents the length of the feature vector.
- $y_i \in [0,1]$ indicates the paying potential.

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 Θ ← 0;
V′ ← Under-sampling training users v ∈ V_1;
L ← a list of random shuffle v ∈ V′;
for t = 1 to T_1 do
    foreach v_i ∈ L do
        Calculate the paying potential by Eq.(3):
        ỹ_i ← ŷ(x_i);
        Calculate the gradient of all parameters by Eq.(5),
        and update parameters:
        w_0 ← w_0 − η · 2(ỹ_i − y_i) ∂/∂w_0 ŷ(x_i);
        for j ∈ {1, ..., d} ∧ x_{i,j} ≠ 0 do
            w_j ← w_j − η · 2(ỹ_i − y_i) ∂/∂w_j ŷ(x_i);
            for l ∈ {1, ..., k} do
                p_{j,l} ← p_{j,l} − η · (2(ỹ_i − y_i) ∂/∂p_{j,l} ŷ(x_i) + 2λ p_{j,l});
            end
        end
    end
end
Initialize paying potentials of test users by Eq.(3):
for v_i ∈ V_2 do
    ŷ_i ← ŷ(x_i);
end
Propagate the paying potential scores to neighborhood:
for t = 1 to T_2 do
    foreach v_i ∈ V_2 do
        Update ŷ_i according to Eq.(8);
    end
end
```

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**
        $p_{j,l} \leftarrow p_{j,l} - \eta \cdot \left(2(\tilde{y}_i - y_i)\frac{\partial}{\partial p_{j,l}}\hat{y}(\mathbf{x_i}) + 2\lambda p_{j,l}\right)$;
      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.(8);
  end
end

Online Test

- Test setting
  - Two groups: *test group* and *control group*.
  - Send messages to invite the user to attend a promotion activity.