XuetangX

—The Platform for Chinese MOOC

Jie Tang

Tsinghua University

2012: The Year of MOOC

- **108** partners
- 633 courses
- **7.1 million** users

coursera
education for everyone

xuetangX.com 学堂在线

- **20+** courses
- **100,000** users
- **Chinese EDU association**

edX

- **50+** partners
- 160+ courses
- **2.1 million** users

網易公开课

- host >900 courses
- millions of users

- **~10** partners
- 40+ courses
- **1.6 million** users

UDACITY

……

Outline

- The platform—xuetangX.com
- Our courses
- Some interesting findings

Milestones

17 May	Tsinghua University joined edX
24 May	Tsinghua Center for Massive Online Education
06 June	edX releases OpenEdX framework
08 June	Tsinghua sets up XuetangX Team
10 Oct	Tsinghua releases XuetangX platform

Our Team

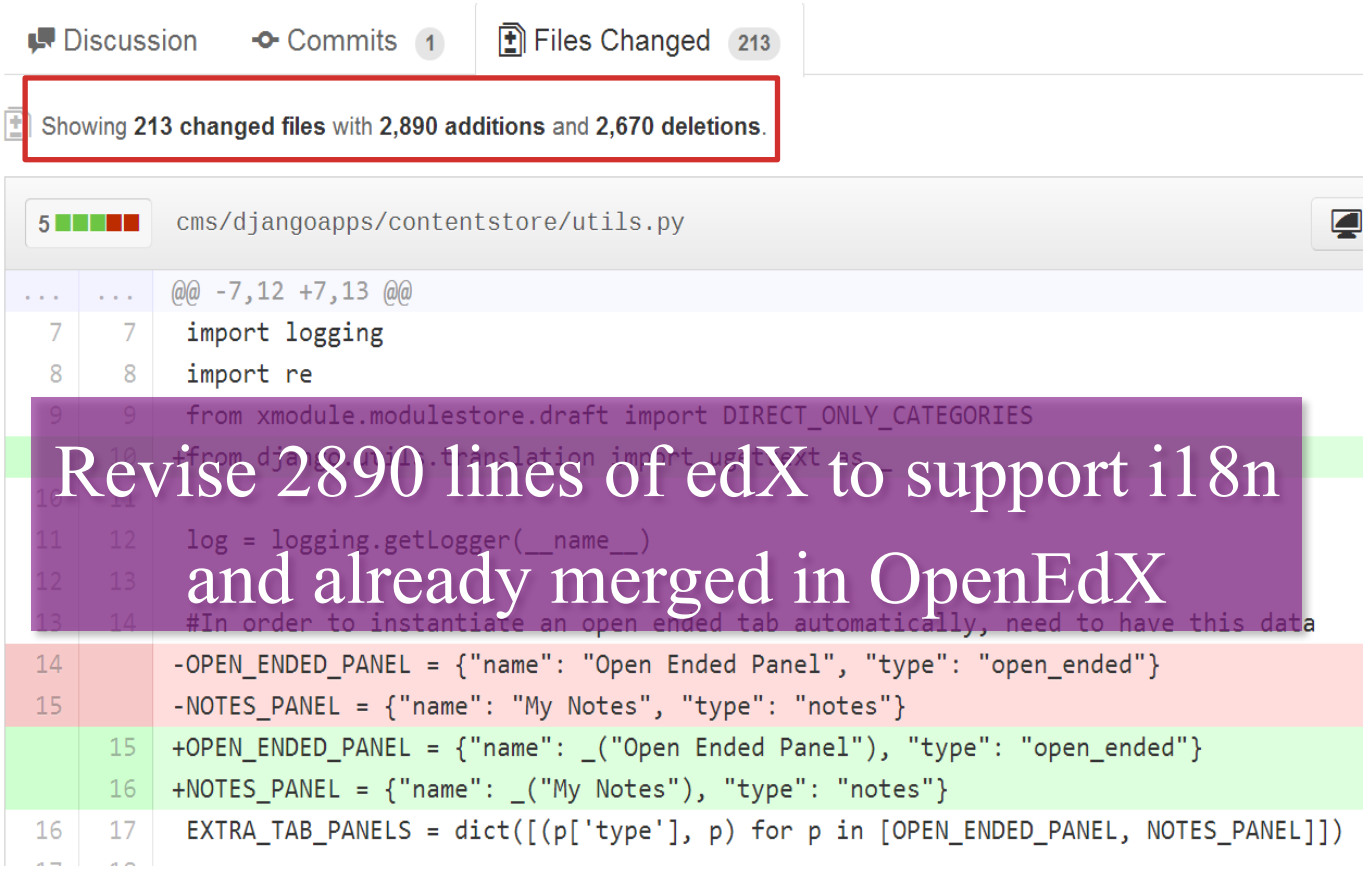
- Tsinghua Center for Massive Online Education
  - Jointly set up by Dept. Computer Science, Institute for Interdisciplinary Information Sciences, School of Social Sciences, and Institute of Education
- XuetangX Team
  - 30 members from different departments
  - Most from Dept. Computer Science
  - Some from Tsinghua Press and Dept. Electronic Engineering

Brief Introduction

- Develop based on OpenEdX
- XuetangX Team adds and modifies source codes of ~100,000 lines

Key Tech 1: i18n and l10n

- Help edX in i18n



Revise 2890 lines of edX to support i18n and already merged in OpenEdX

Key Tech 1: i18n and l10n

- Translate about 30 thousand English words into Chinese and finish l10n

The work speeds up i18n of edX. Currently. There are 30+ languages are performing translation for l10n. Chinese, French, Korean, Portuguese and Spanish have finished translation.

Key Tech 2：Video Player

- Develop video player based on HTML 5 / Flash
  - Independent of YouTube. Support multiple sources
  - Support a unified interface to upload videos in CMS

Key Tech 3: Course Search

- Support search of video captions in Chinese/English
- Click search results to play the video from the expected point

Key Tech 4: WYSIWYG Editor of Math Formulas

- WYSIWYG editor of math formulas can facilitate teachers / students for math input
  - Support rich math symbols for input
  - Comfortable input experience in Web browsers suportted by HTML + CSS
  - Integrate with rich text editor

```
\begin{equation}
\int x^{\mu} dx = \frac{x^{\mu +1}}{\mu+1}
\end{equation}
```

$$\int x^{\mu} dx = \frac{x^{\mu+1}}{\mu+1}$$

LaTeX editor by EdX

WYSIWYG editor by XuetangX

Key Tech 5:
Auto Grading for Programming Homework

- Student can upload their source codes and get auto grading results

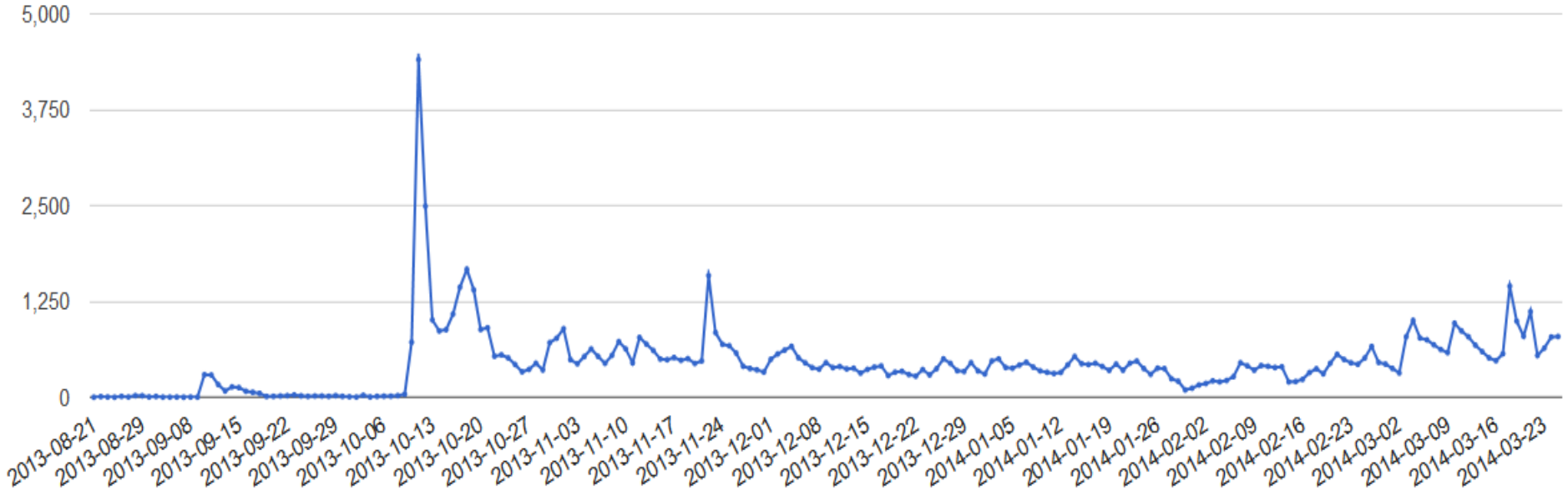
Summary

After several month operation, xuetangX has attracted ~100K users and the number is continuously increasing…

In Service

- Support ~20 Tsinghua MOOCs simultaneously with edX
  - Principles of Electric Circuits; History of Chinese Architecture; Data Structure; Historical Relic Treasures and Cultural China; Financial Analysis and Decision Making
- Partners' courses
  - **MIT:** Circuits and Electronics
  - **UC Berkeley:** Cloud Computing and Software Engineering
  - **Peking University:** Principles and Practice of Computer Aided Translation
- Support 2 Tsinghua SPOCs
  - C++ Programming by Prof. ZHENG Li for 93 students
  - Cloud Computing and Soft Engineering by Prof. XU Wei for 35 students

User enrolment in the past months



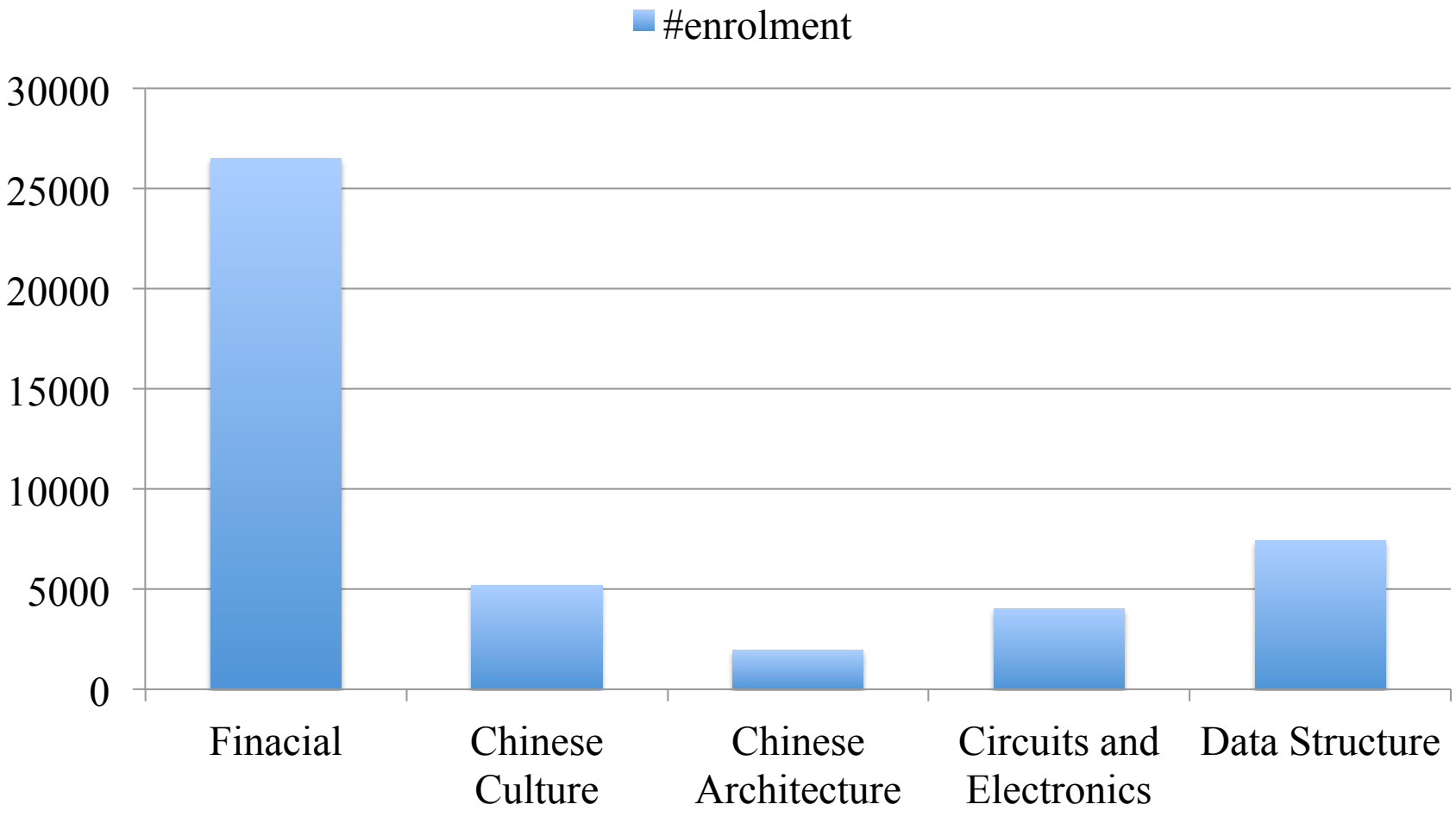
Outline

- The platform—xuetangX.com
- Our courses
- Some interesting findings

Courses

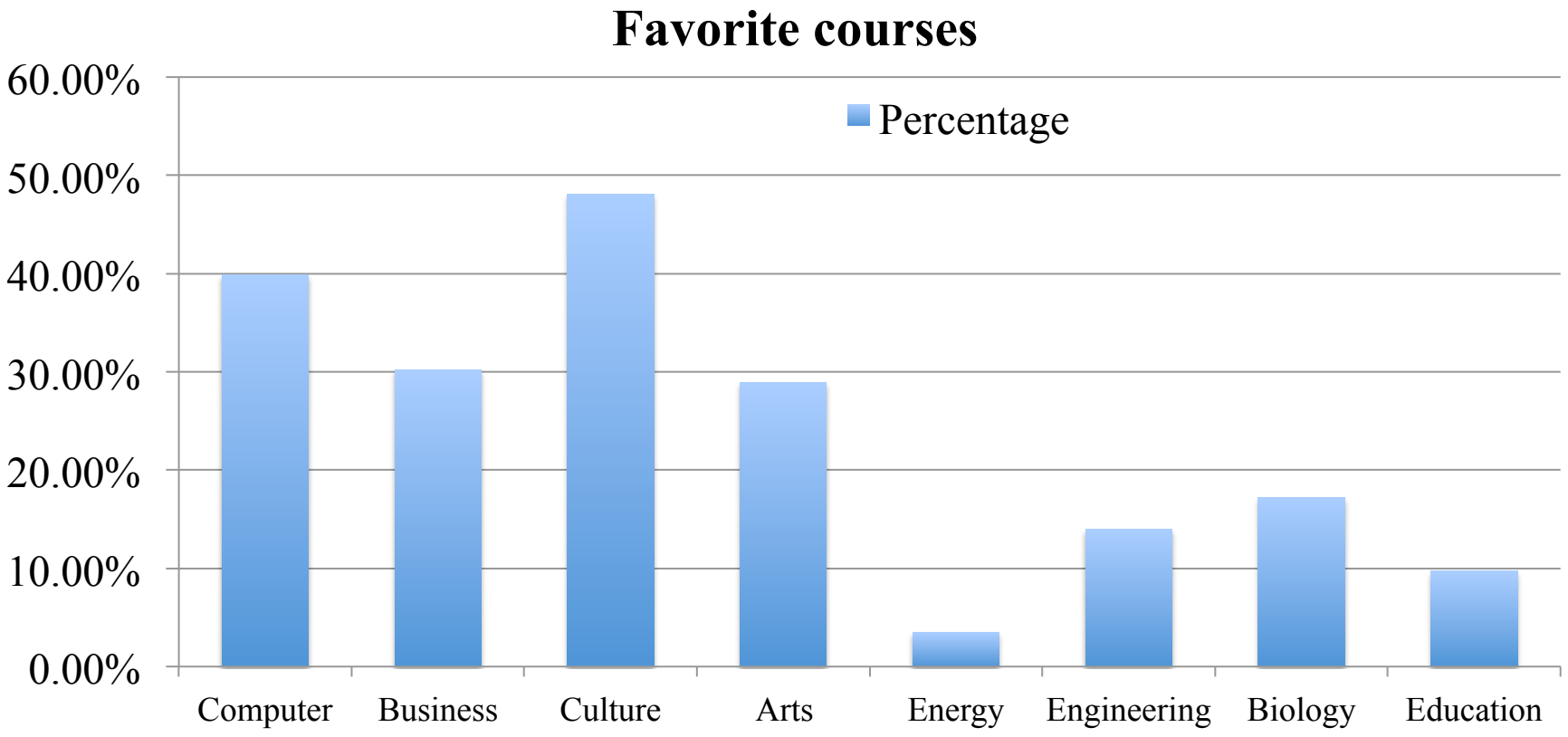
- ~20 courses offered by 4 partners (MIT, UC Berkeley, Peking U, Tsinghua)
  - Data Structure
  - Principles and Practice of Computer Aided Translation
  - Financial Analysis and Decision Making
  - Historical Relic Treasures and Cultural China
  - Cloud Computing and Soft Engineering
  - Circuits and Electronics
  - Principles of Electric Circuits
  - Introduction to Psychology
  - Combinatorics
  - Football Science
  - …
- 7 Tsinghua courses also on edX
  - History of Chinese Architecture; Data Structure; Historical Relic Treasures and Cultural China; Financial Analysis and Decision Making

#Enrolment on edX

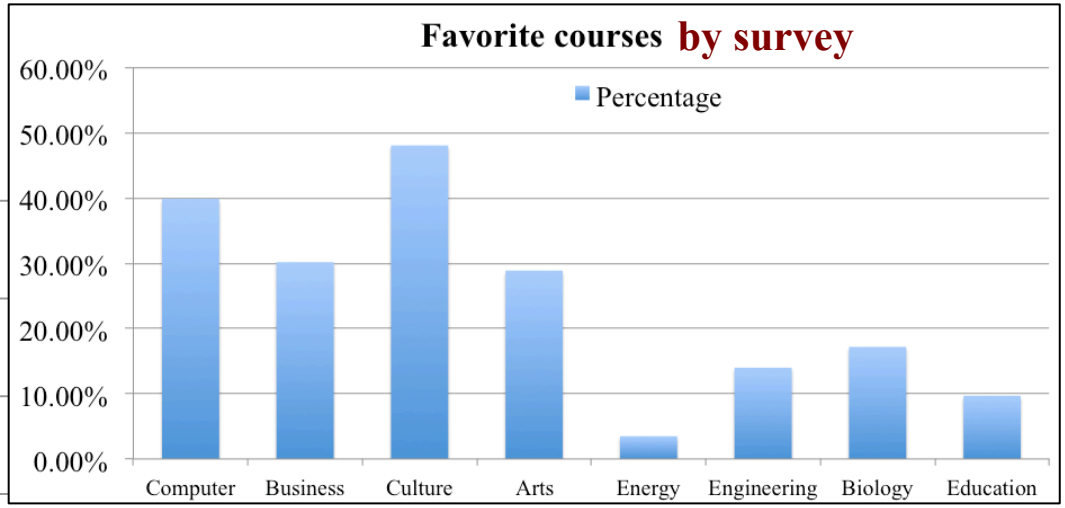


Enrolment on XuetangX

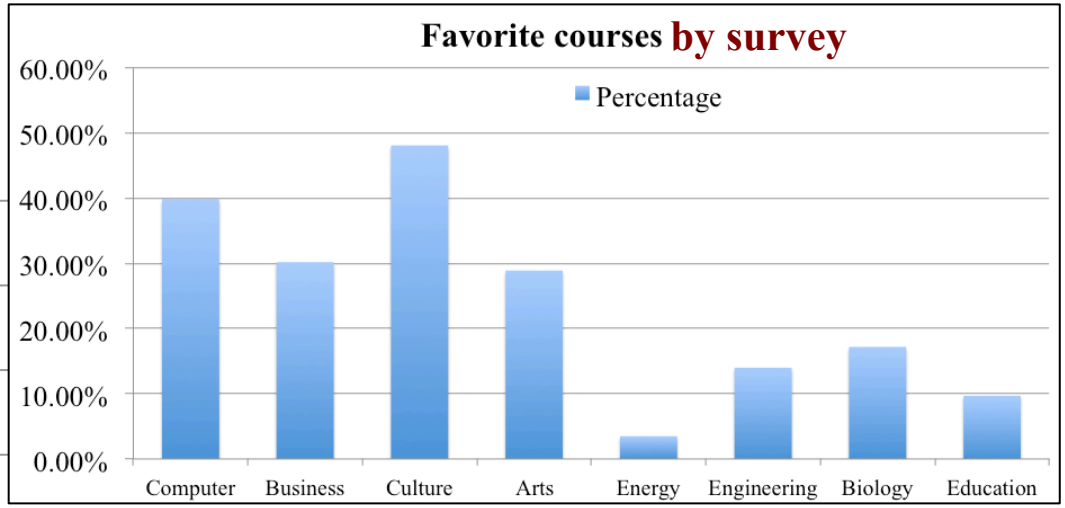
A survey conducted on 6,116 randomly chosen Internet users before we released xuetangX…



Enrolment on XuetangX



Enrolment on XuetangX



Two SPOC courses in Tsinghua

- Two Tsinghua SPOCs
  - C++ Programming for 93 students
  - Cloud Computing and Soft Engineering for 35 students
- Deploying multiple SPOC systems in China
  - Wenzhou University
  - Jiangxi Finance Institute
  - …

Outline

- The platform—xuetangX.com
- Our courses
- Some interesting findings

Rich tracking logs of student behaviors

- The huge amount of data available in MOOC offers a unique opportunity for understanding student behavior
- Such logs include: watch video, homework, forum, etc.

Item	Number
Users	76,215
Courses	9
Logs	~7.5M
Date span	2013/09/28-2014/01/15

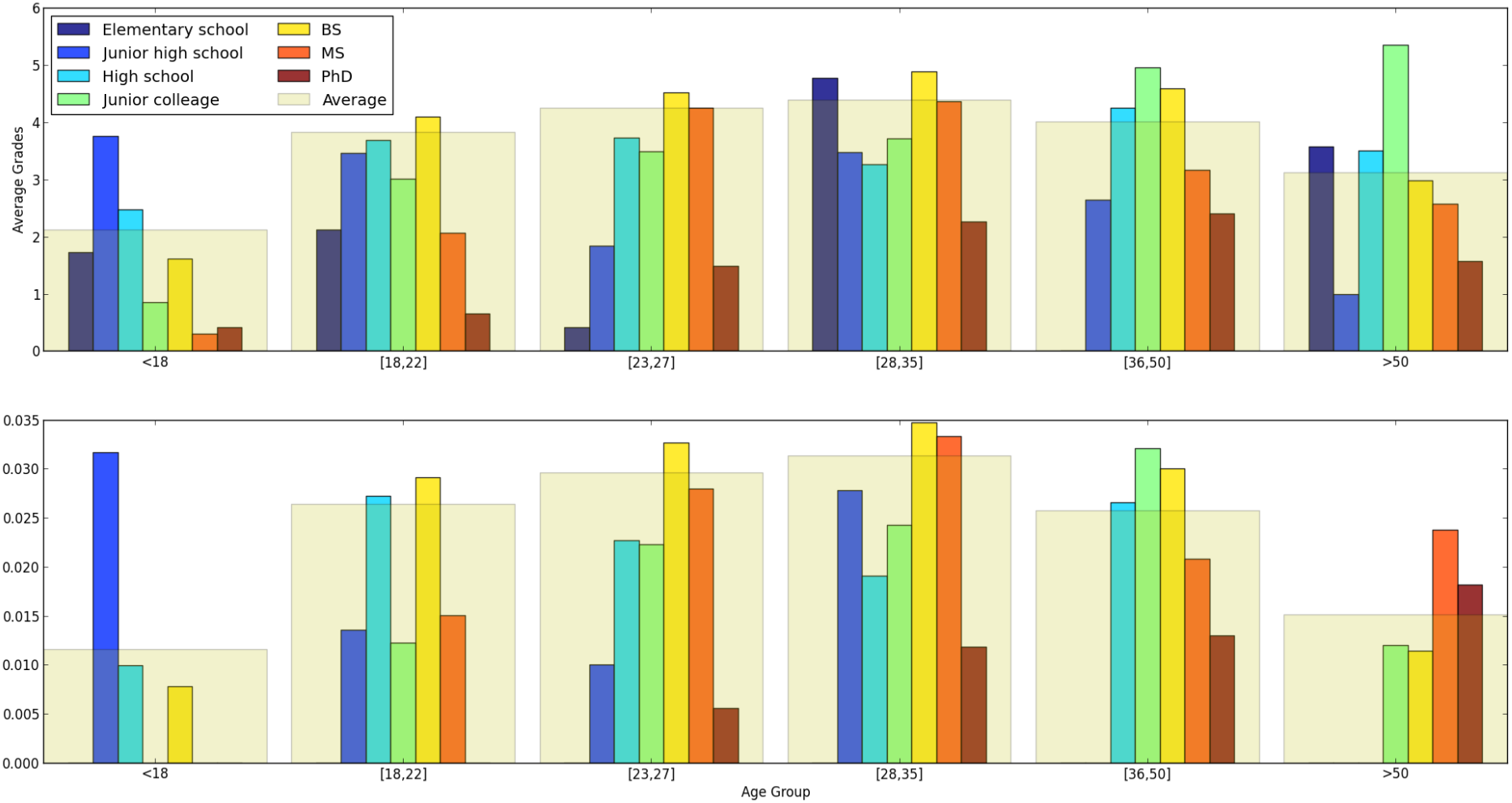
Recent Research

- Personalized service based on education data analysis
- Learning behavior analysis
- Peer assessment based on social networks
- Intelligent course QA systems
- Auto grading for student essays
- Virtual labs
- Mobile apps
- …

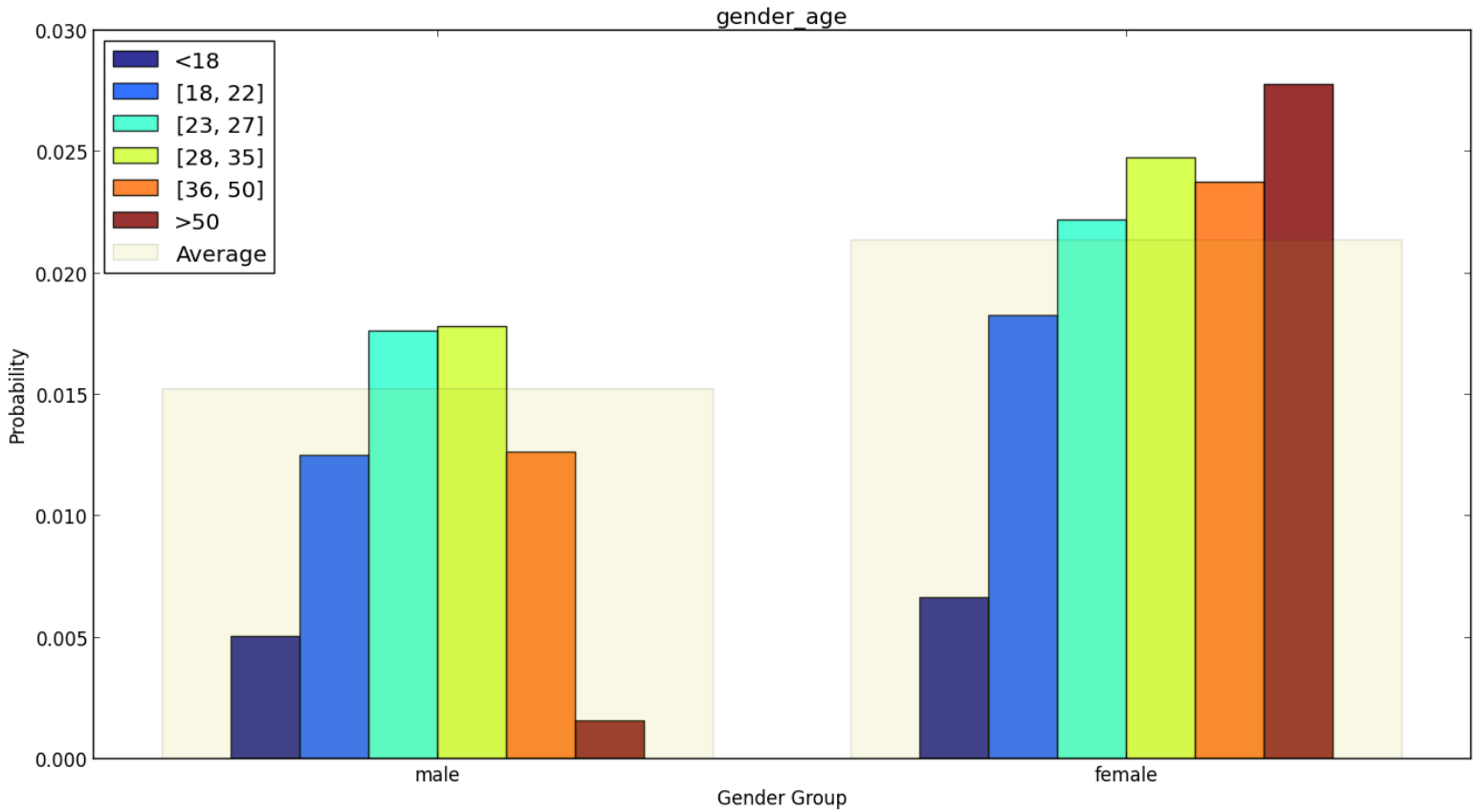
One particular question

- One fact: 76,215 users and only 3%-6% received the certificates
- An interesting question is: Who finally received the certificates?

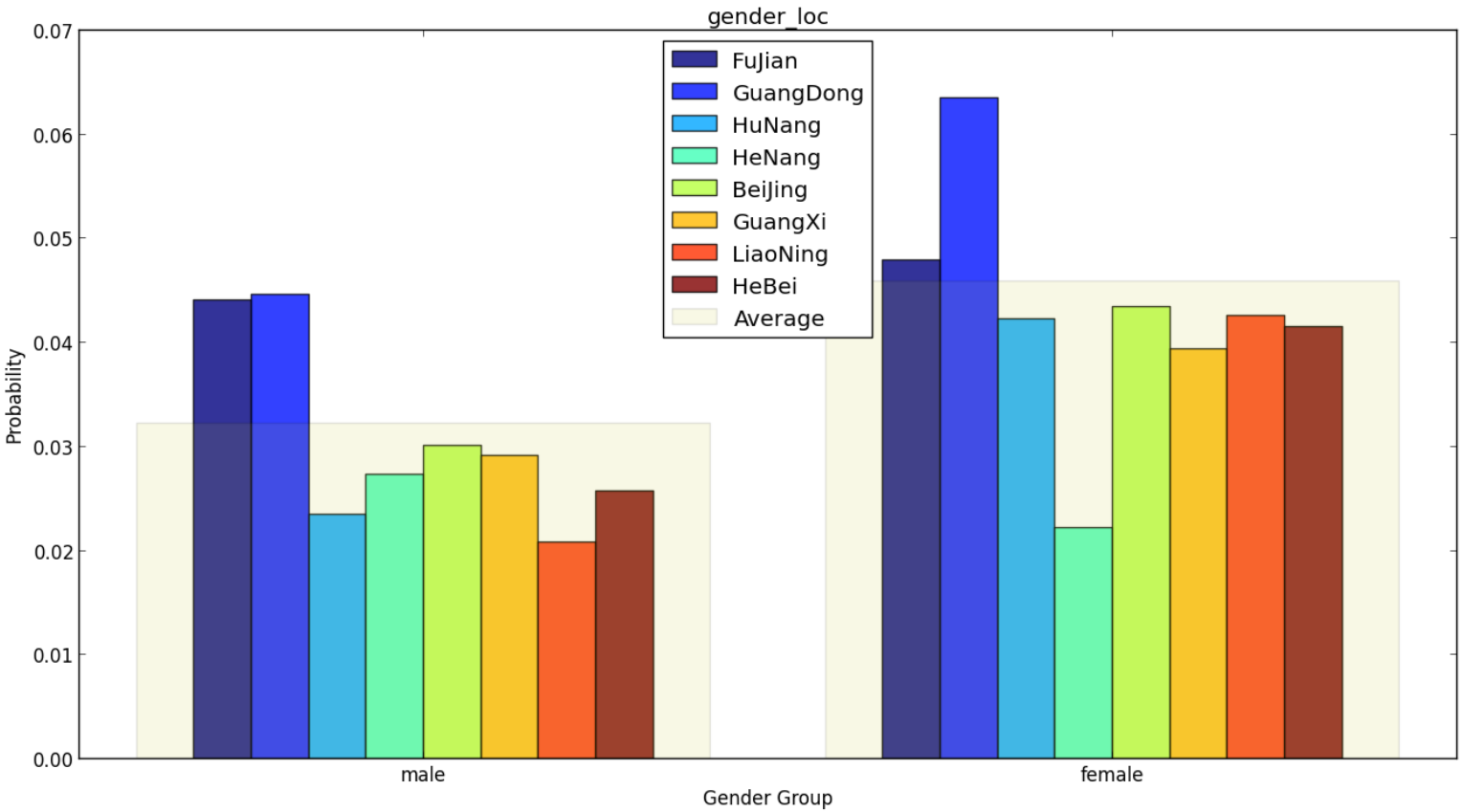
Age+Education vs. Certificate



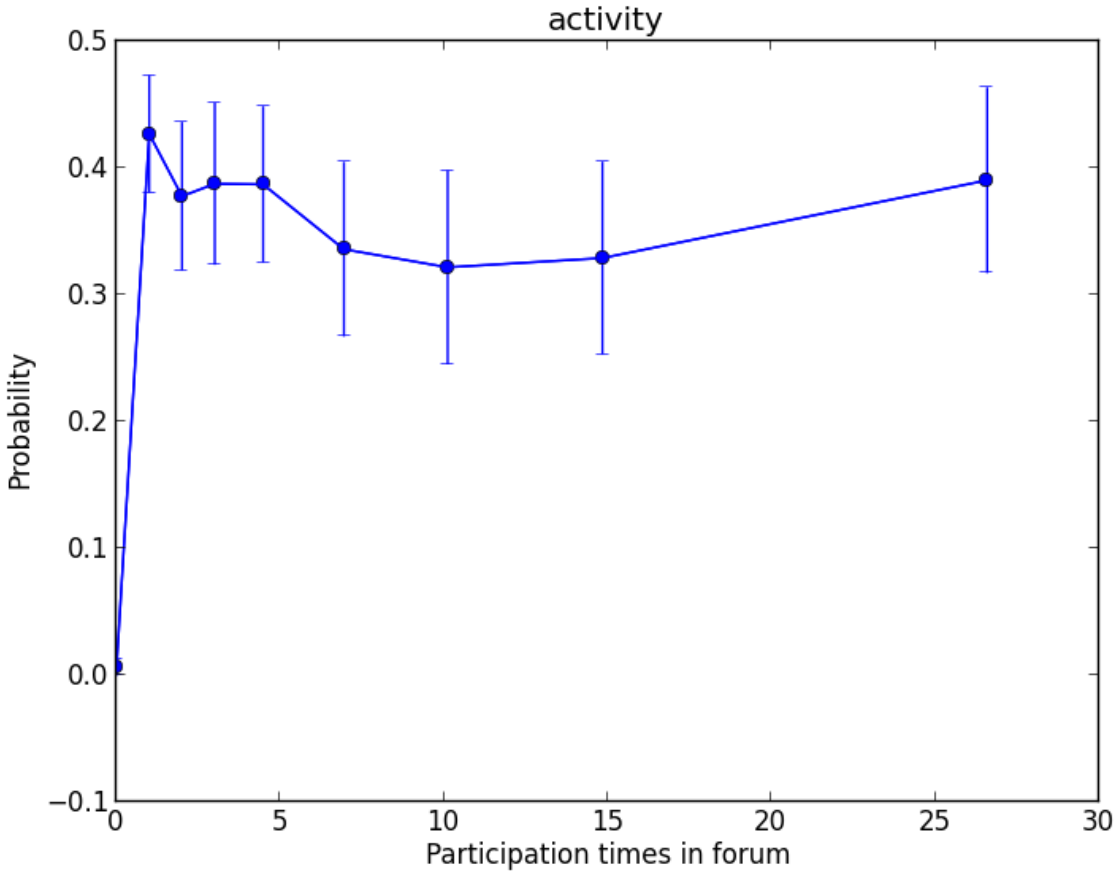
Gender+Age vs. Certificate



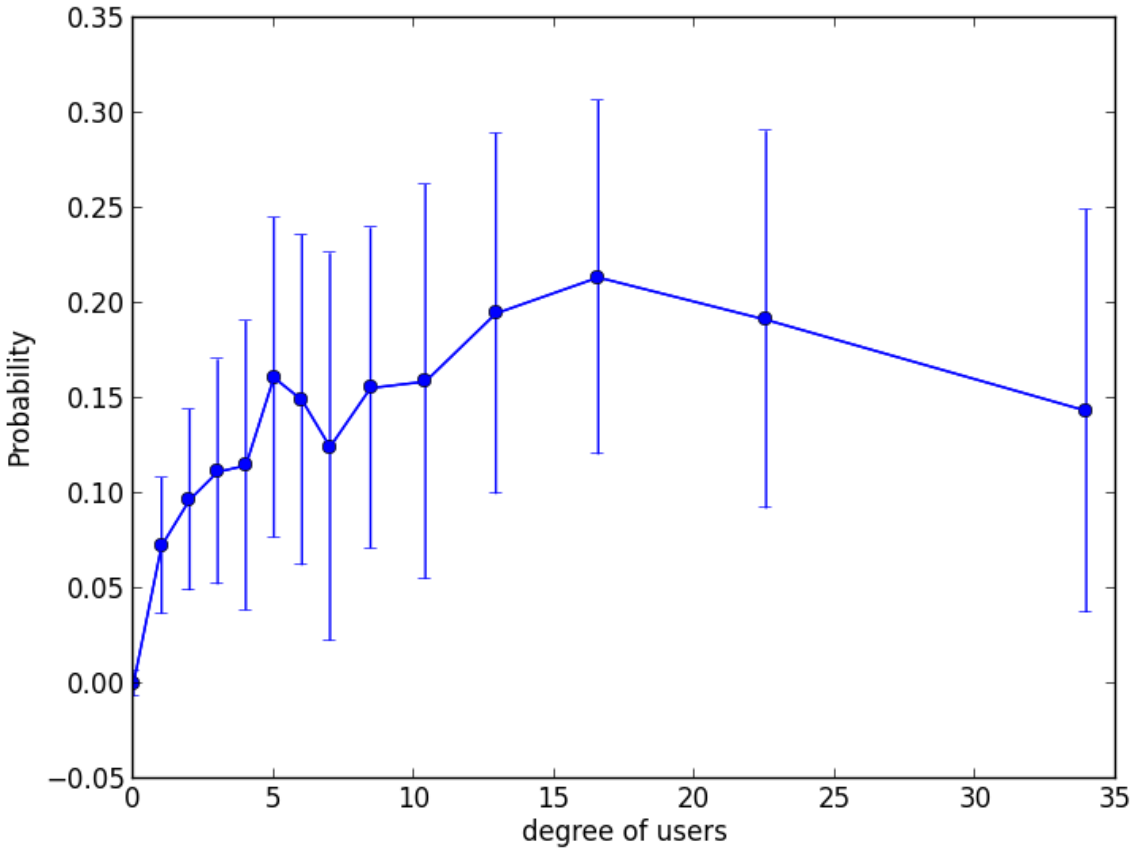
Gender+Location vs. Certificate



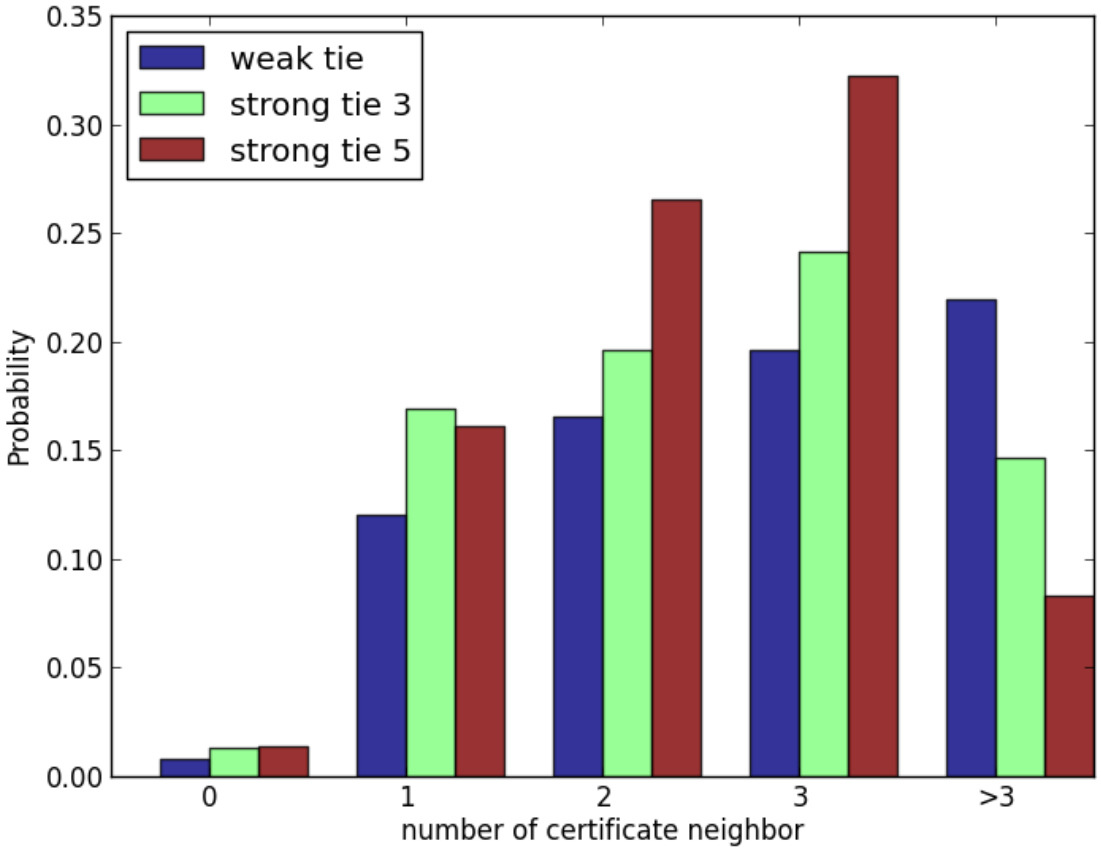
Forum vs. Certificate



#friends vs. Certificate



Influence vs. Certificate



Can we predict who will/could receive the certificate

- Given behavior log data by all users in the MOOC system,

- Predict whether a user will finally graduate and receive the certificate of a specific course.

Factorization Machines

- The prediction of feature vector $\mathbf{x}_i$:

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

- and the corresponding objective function:

$$O = \sum_{x_i} (y(\mathbf{x}_i) - y_i)^2 + \lambda \sum_{i=1}^{d} \left\| \mathbf{p}_i \right\|^2$$

Preliminary Results

Method	Features	AUC	Precision	Recall	F1
Factorization Machines	Demographics	90.80	5.91	45.24	9.89
	+ Social Influence	96.28	27.90	51.89	36.53
SVM	Demographics	84.36	5.54	42.31	9.81
	+ Social influence	96.49	25.90	50.85	34.27

* SVM is a state-of-the-art algorithm for classification/prediction. We use it as the baseline method in our experiments.

Outline

- The platform—xuetangX.com
- Our courses
- Some interesting findings
- Summaries

Summaries

- We have developed and deployed one of the largest MOOC platforms in China
  - attracted hundreds of thousands of users
- Interesting findings
  - “Financial analysis” is the most popular course
  - Most users are browsers, and 3-6% are certificate earners
  - There is a strong influence between users’ learning behavior

Future Research Directions

- Personalized learning path recommendation
- Learning pattern discovery
- Peer assessment based on social networks
- Intelligent course QA systems
- …

Thanks!

http://xuetangx.com
http://keg.cs.tsinghua.edu.cn/jietang

Age+Education vs. Certificate