# Taxi Driving Behavior Analysis in Latent Vehicle-to-Vehicle Networks: A Social Influence Perspective

**Tong Xu[1], Hengshu Zhu[2], Xiangyu Zhao[1] , Qi Liu[1], Hao Zhong[3], Enhong Chen[1], Hui Xiong[3]**

[1] University of Science and Technology of China,

[2] Baidu Research, Big Data Lab

[3] Rugters, The State University of New Jersey

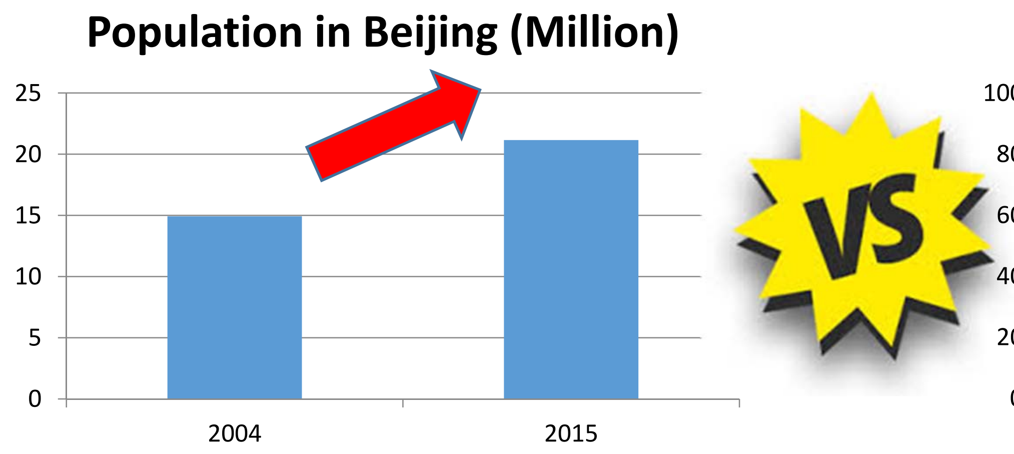
中国科学技术大学 University of Science and Technology of China

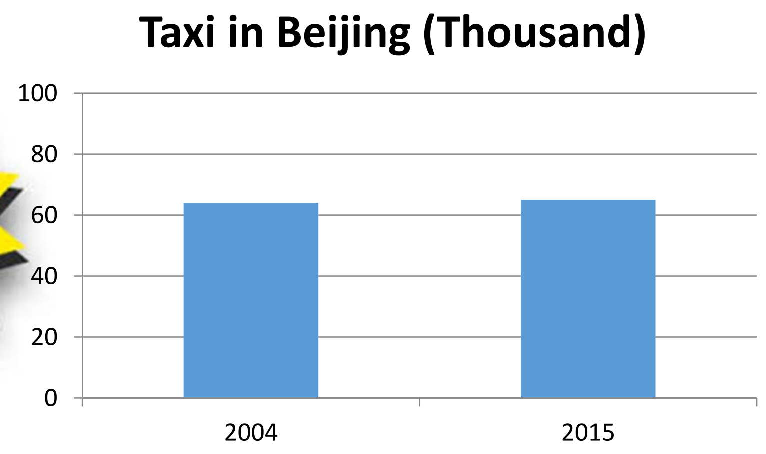
Baidu 百度

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

# Background

- **The dramatic expansion of urban areas and population, results in the urge demand of efficient taxi services.**
  - Limited cabs due to amount control.
  - Simply increasing the amount of cabs or drivers could hardly solve the dilemma.



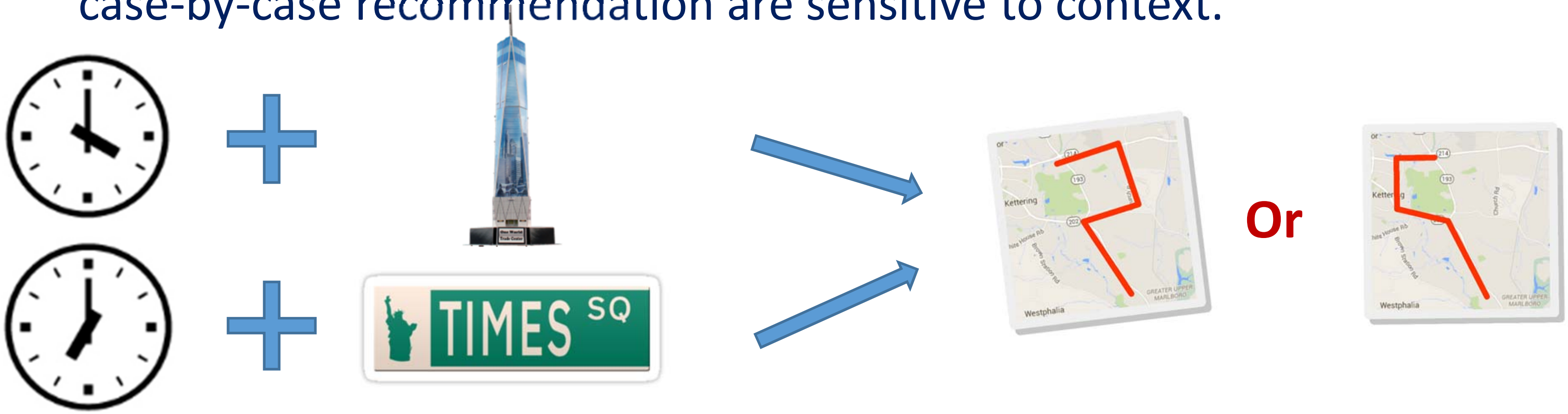
# Prior Arts

- **Thanks to the rapid development of wireless sensor technologies in mobile environments, the abundant real-time trajectories could be promptly collected.**
- **Corresponding, intelligent services can be enabled to support cab drivers or passengers.**
  - Fastest driving routes [Zheng, et. al, Ubicomp'2011]
  - Sequence of pick-up points [Qu, et. al, KDD'2014]
  - Customers within the shortest driving distance [Ge, et. al, KDD'2010]
  - Etc…



# Limitation

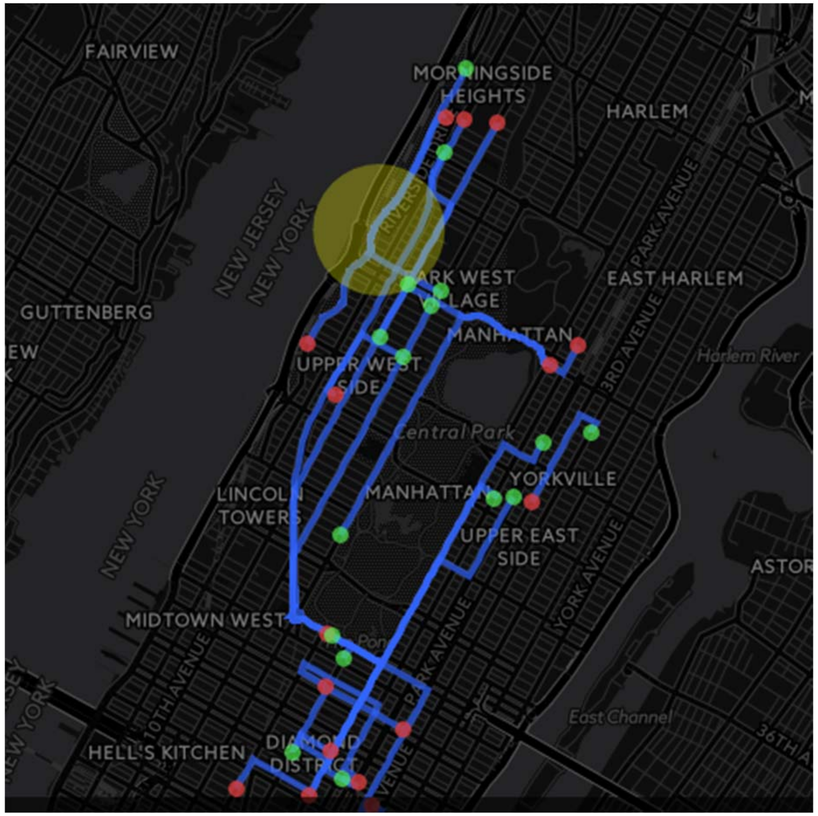
- **Although the above works can effectively enhance taxi business, they may still suffer some defects.**
  - Frequent update is required which raises heavy burden of system, as the case-by-case recommendation are sensitive to context.




  - Except for the recommendation task, cabs should be distributed by extra scheme for keeping regional balances.
  - Only algorithm try to teach people what to do!

# Motivation

- **Individual factors of taxi drivers, i.e., social sharing or learning between drivers have been largely ignored.**




- **Cab drivers usually hold their own driving behavior patterns.**
  - **E.g., the drivers in left figure tend to drive around the central park.**
  - Designed by Chris Whong (nyctaxi.herokuapp.com).

# Motivation (cont.)

- **Indeed, even without intellectual services, cab drivers could also rely on themselves to well fit the taxi business.**
  - For the experienced drivers, they could effectively summarize the knowledge to meet circumstances.
  - For the rest inexperienced ones, luckily, driving experience sharing within drivers may help them.

# Motivating Example

| Author | Topic |
|---|---|
| **fschase** (4 Posts) 3/6/201 | New driver's need help... |

| | |
|---|---|
| **ksktaxi** (5 Posts) 3/7/2012 7:53:20 PM | Re: New driver's need help... |

Hi my friend

I'm not a old driver but I will share my tactic for new drivers.I'm driving about 9 months as a day driver,first I was start like what you mentioned by using gps so now I don't need to use it anymore.And I will tell what I'm doing everyday and it brings approximately $130-$180 everyday.
I work 5 to 5 pm.

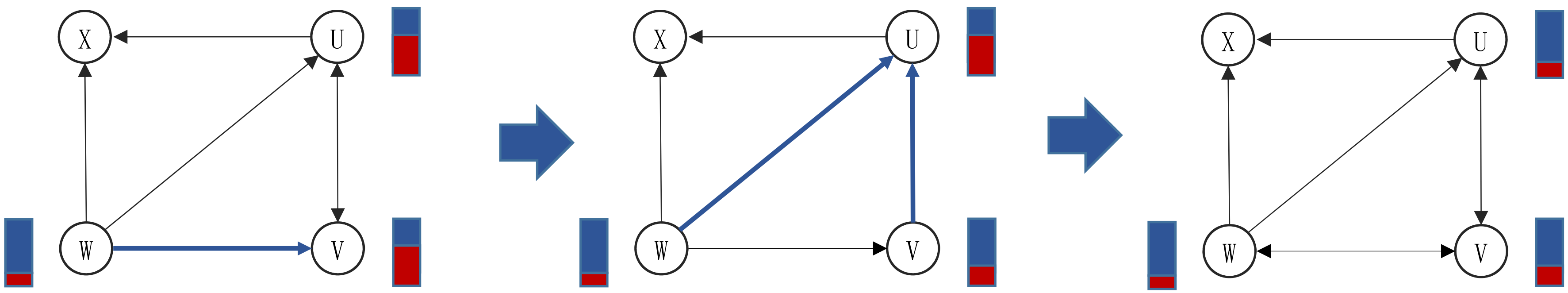
I get the car downtown like 5a.m and start to work.My first though is find to someone who already wanna go home from meat pack, village or midtown bars.Usually drunk people or who are employees of bars.It goes like 5.45 am.And then I drive to Port Authority.There are two lanes over there I usually take outer one because people prefer who go shorty like 48-6 ave 52-lex. etc...I drop and come back P.A quickly. It keeps going until 7 or 7.30 it depends.After 7.30 traffic starts I go up to uptown usually east side.Take someone and drop in midtown and go back empty but check it out hen you go up, if you get someone use the advantage of it. I keep it until 10 am.There are heavy traffic and less business at this time so its time to out of city.Im looking for people who have bag heavy bags most.It means airport,if you lucky it is jfk, if you less lucky lga, if you so lucky newark.But if you have a bad like it could be penn station or short distance.It is all about luck nothing to do.I dont like to stay in front of the hotel both it is wasting time and doormans take your money.But if you wanna stay i advise shorten on 53&7 or Hilton 53&6.Because in Sheraton there is short line and airport line.In hilton doormans never talk to you they just whistle and whoever on the first he goes.It looks fair for me.Other hotels has tipping and limo friend system I dont advise for new guys.

If I get Jfk , i wait on taxi hold, depends on how many cabs are there.Usually I get off like 12or 12.30 but depends.After that the city is busy.I advise to you downtown like tribeca, soho.

But afternoon you should avoid canal st. 36.st and 48th 52nd st.Too much traffic, 30 and 31st also.best way to cross town is 38th to east 39th to west.if you go all the way west or east like 1st ave to 9 ave. i advise 42nd st. and for to the west 34th, but do not use 34th to east bound.I have many things to say but I couldnt remember at this time.I can write soon.Good luck you all

# Assumption

- **Social interaction (learning) may influence driving behavior.**



# Problem Definition

- **Given the target group of taxi drivers, and for each driver, corresponding behavior patterns during each time period (timestamp as 1,2,...,T) are summarized.**
- **The problems could be defined as follows:**
  - Mining latent social connections within drivers.
  - Predicting future behavior pattern of drivers.

# Preliminaries

- **Driving behaviors are summarized as vectors in time series, where**
- $\mathbf{s_i^t}$ : Driving behavior pattern vector of driver $u_i$ in time t.
- $s_{i,k}^{t}$ : Ratio of *k*-th driving behavior pattern, of driver $u_i$ in time t.
- $p_{i,k}^{t}$ : Corresponding social influence on *k*-th driving behavior pattern.
- **Further, to describe the time-variation of driving behaviors, we have**
- $\mathbf{R}_i^t$ : The set of patterns whose frequency increased in time t.
- $\mathbf{D}_i^t$ : The set of patterns whose frequency decreased in time t.

# Partial Ranking

- To be specific, different social influence may cause increasing / decreasing on different driving patterns.
- Correspondingly, the increasing / decreasing ratio indicates partial ordering of social influence.
- For $\forall \langle r, d \rangle_{i,t}$ , where $r \in \mathbf{R}_i^t$ and $d \in \mathbf{D}_i^t$ , we have $p_{i,d}^t < p_{i,r}^t$
- In which, social influence could be estimated by the basic Independent Cascade (IC) Model (replaceable)

$$p_{i,k}^t = 1 - \prod_{j \in N_i} (1 - w_{ji} \cdot \delta_{i,j,k}^{t-1}),$$

# Loss Function

- **Finally, to optimize the partial ranking, we formulate the loss function of pairwise ranking problem as follows**

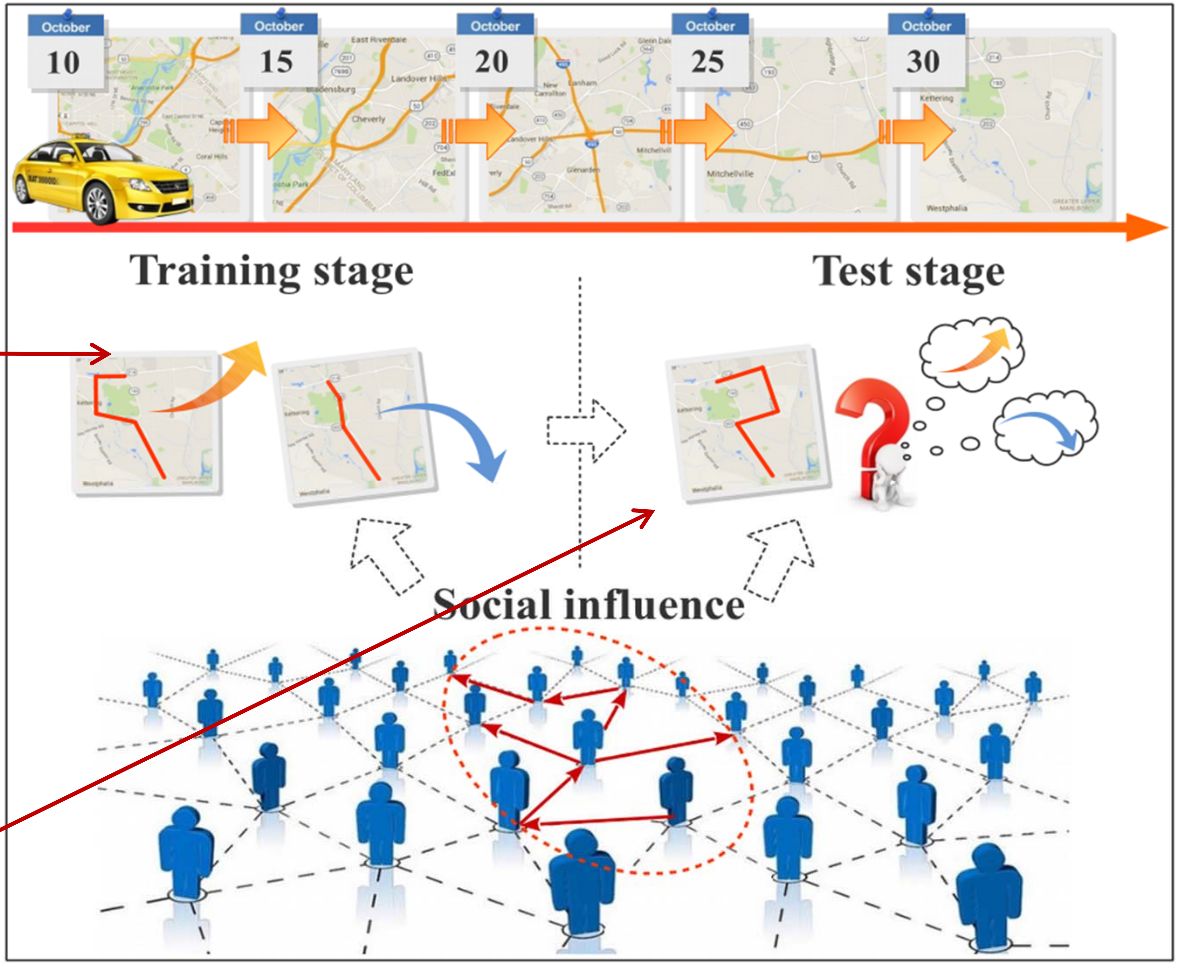
$$min_w \mathcal{F}(w) = \sum_{i,t} \sum_{r \in \mathbf{R}_i^t, d \in \mathbf{D}_i^t} h(p_{i,d}^t - p_{i,r}^t)$$

- **Here h() indicates punish function of ranking error (for non-negative term).**

# Framework

**Training Stage**
**Latent social connections** will be revealed and **weighted** via optimizing the loss function.

**Test Stage**
**Social influence** on driving pattern will be estimated to predict **future driving behaviors**.



# Experimental Setup

- **Prediction Task.**
  - Predicting future driving behavior pattern.
  - Experiments on NYC taxi data during 2013.

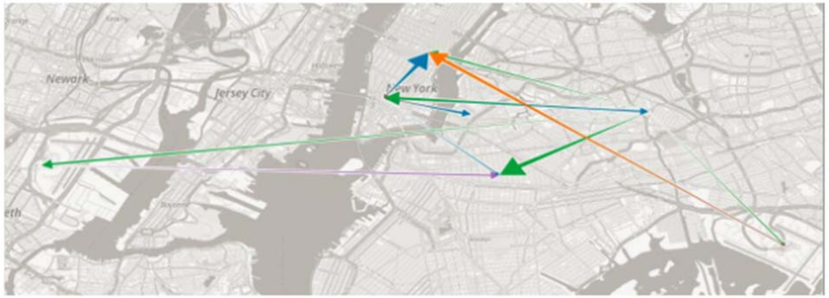
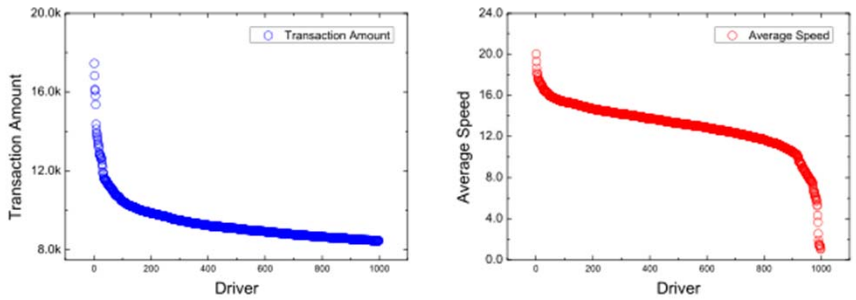



**Table 1: Data Set Description**

| | Data Statistic |
|---|---|
| Number of Taxis | 14,144 |
| Number of Drivers | 43,191 |
| Average Num. of Transactions | 3,928.86 |
| Average Num. of Passengers | 1.68 |
| Average Trip Time | 15.05 min |
| Average Trip Distance | 8.86 miles |
| Average Trip Fare | $15.39 |




(a) (b)

**Figure 1: The distribution of transaction amount and average driving speed of taxi drivers.**

# Experimental Setup (cont.)

- **Metrics**
  - Precision / Recall (for binary classification)
  - NDCG / MAP@k (for ranking performance).
- **Benchmarks**
  - Personalized Average
  - Overall Popularity
  - Vector Auto-regression

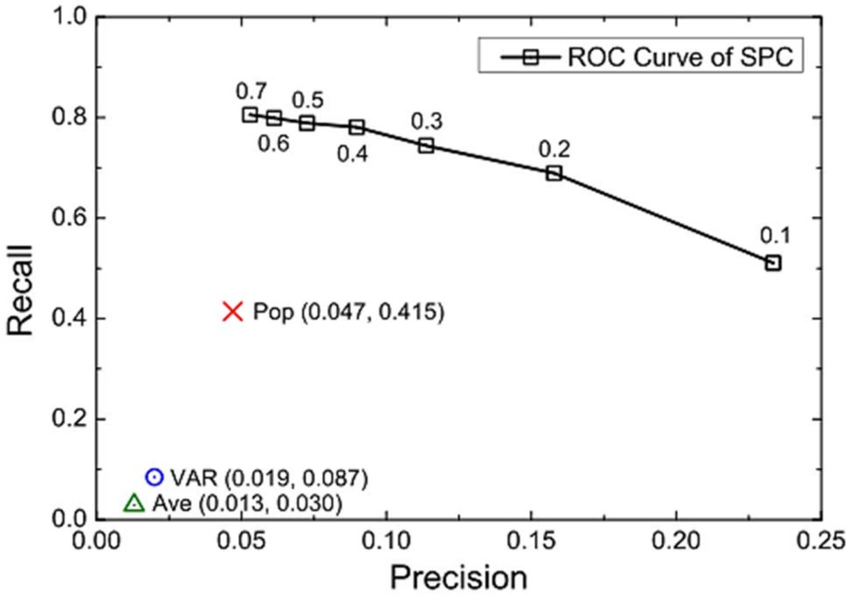



**Table 4. Overall performance of each approach.**

| | SPC | Ave | Pop | VAR | | SPC | Ave | Pop | VAR |
|---|---|---|---|---|---|---|---|---|---|
| NDCG | 0.3502 | 0.1603 | 0.2211 | **0.3619** | MAP@10 | **0.2128** | 0.0254 | 0.1042 | 0.2018 |
| Improvement (%) | - | +118.46 | +58.39 | -3.23 | Improvement (%) | - | +737.79 | +104.22 | +5.45 |
| P-Value | - | 0.000 | 0.000 | 0.755 | P-Value | - | 0.000 | 0.000 | 0.472 |
| Precision | **0.1579** | 0.0134 | 0.0474 | 0.0192 | Recall | **0.6892** | 0.0298 | 0.4151 | 0.0875 |
| Improvement (%) | - | +1078.35 | +233.12 | +722.39 | Improvement (%) | - | +2212.75 | +66.03 | +687.66 |
| P-Value | - | 0.000 | 0.000 | 0.000 | P-Value | - | 0.000 | 0.000 | 0.000 |

# Performance

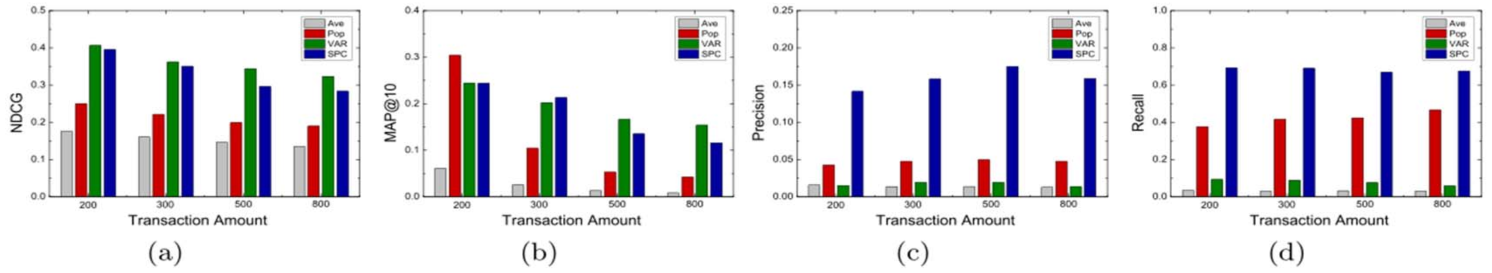



**Figure 7: The verification on robustness with different set of patterns in terms of different metrics, a) NDCG, b) MAP@10, c) Precision, d) Recall.**

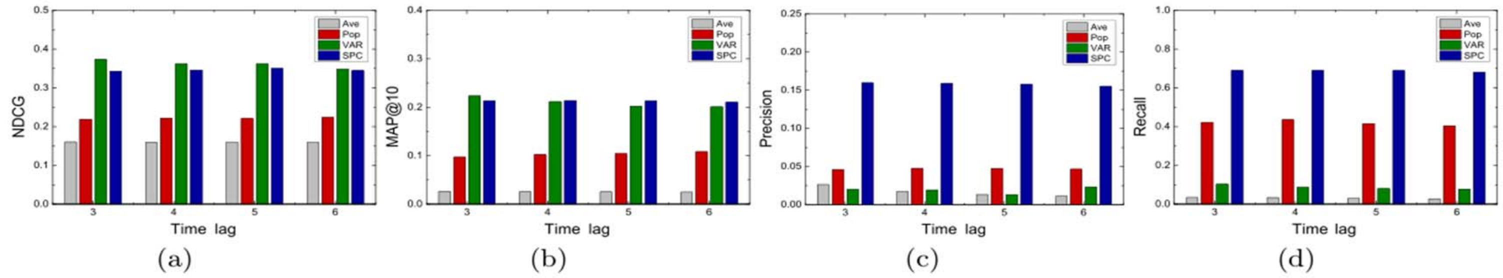



**Figure 8: The verification on robustness with different time lag in terms of different metrics, a) NDCG, b) MAP@10, c) Precision, d) Recall.**

- Without extra information, our framework outperforms the baselines dramatically, only introducing social learning scheme.

# Discussion

- **Social learning: Quantity v.s. Quality**
  - Drivers prefer "better tutor" to "more tutors“.
  - Top drivers tend to keep changing behaviors, while the majority usually cling to old patterns.
- **Skill sharing: Efficiency & Effectiveness**
  - Patterns with long-distance trip could hardly spread (“students” are not willing to learn)
  - Patterns with high benefits never spread beyond small groups (“tutors” are not willing to share)

1 Introduction

2 Motivation

3 Framework & Formulation

4 Experiments & Discussion

5 Conclusion

# Conclusion

- **In this paper, we investigated the latent vehicle-to-vehicle network with simulating the social-driven behaviors of taxi drivers.**
  - A social-driven two-stage framework is propose to better explain drivers' future behaviors.
  - Driving behavior prediction has been transferred as partial ranking with social influence.
- **Future works: Social-oriented Intelligent Service for Taxi**
  - "Tutor" or pattern recommendation: "People's Choice".
  - Cross validation between real social interaction and estimated social connection.

# Thanks!

Email: tongxu@ustc.edu.cn