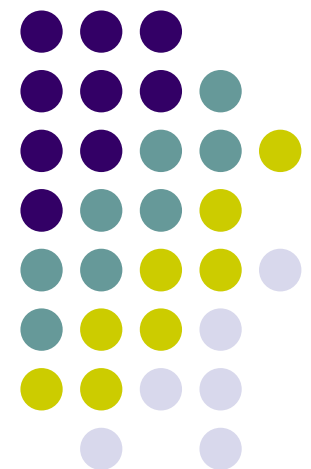


Ubiquitous and Mobile Computing CS 528: Characterizing Smartphone Usage Patterns from Millions of Android Users

Shaocheng Wang, Guojun Wu

*Computer Science Dept.
Worcester Polytechnic Institute (WPI)*





Introduction (cont.)

- Understanding user behaviors of using mobile apps is critical to the industry.
- There are some previous studies which are lack of data.
- They conduct a comprehensive study to find some user patterns.



Introduction

- How can we characterize the app popularity among millions of users?
- How do mobile users choose and manage apps?
- How do different apps perform in network activity?
- How do economic factors affect app selection and network usage?



Related Work

- Understanding User Behavior by Field Study
- Mining App Marketplace Data – ARMiner, WisCom
- Predicting Apps to Use



Data Set

- From: Wandoujia, an Android App Store which is the most popular in China.
- Size: 260,172 apps
- Subset:
 - App management data set: 8,112,145 users
 - Network usage data set: 2,031,007 users

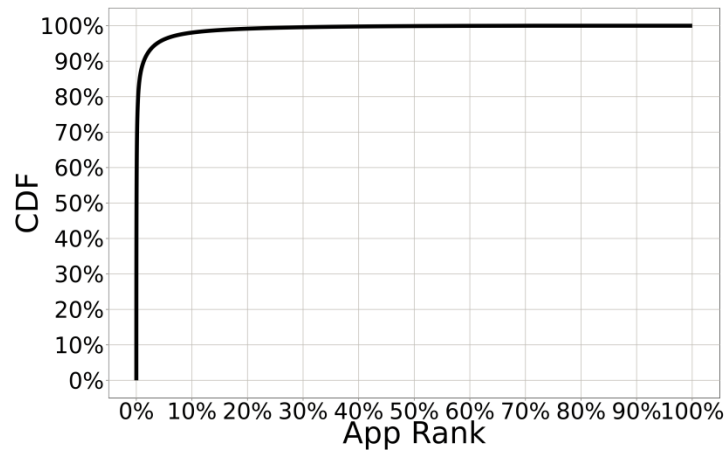
App Popularity



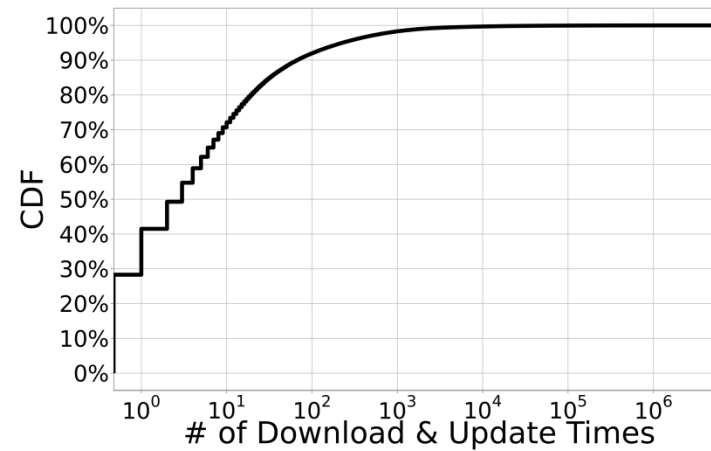
- Metrics:
 - the number of downloads of the app;
 - the number of unique devices that download the app;
 - the aggregated data traffic generated by the app;
 - the aggregated access time that users interact with the app.



App Popularity-Downloads



Percentage of Downloads against App Rank

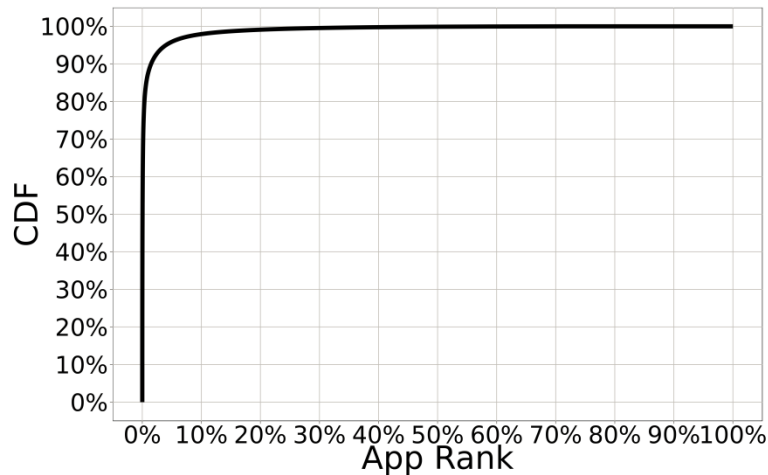


Downloads of an App

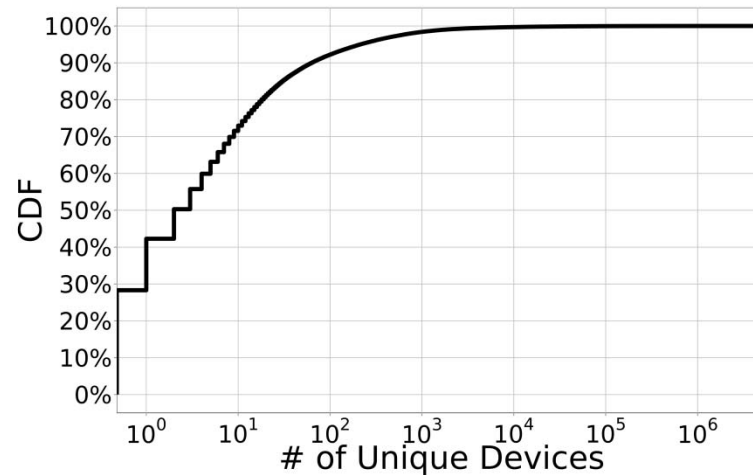


App Popularity-Unique Subscribers

- How many devices has downloaded a App



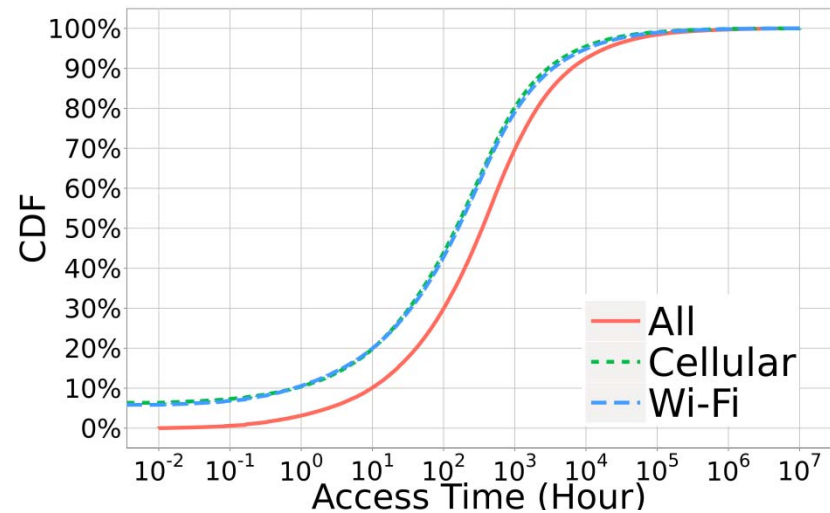
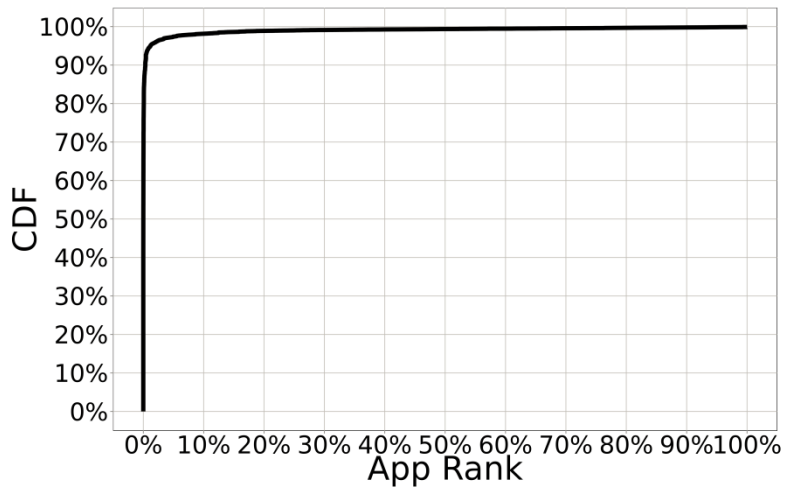
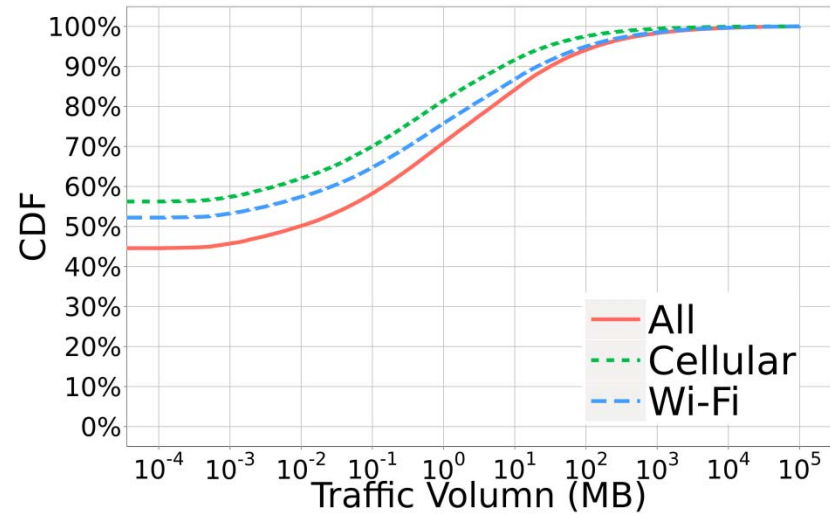
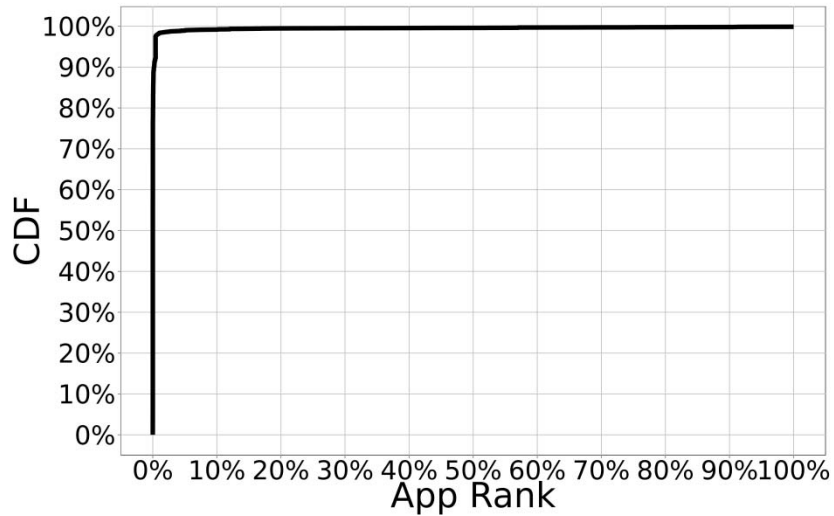
Percentage of Unique Subscribers against App Rank



Unique Subscribers of an App

- They did find some apps with high downloads and low unique users

App Popularity-Network Traffic



App Popularity-Network Traffic

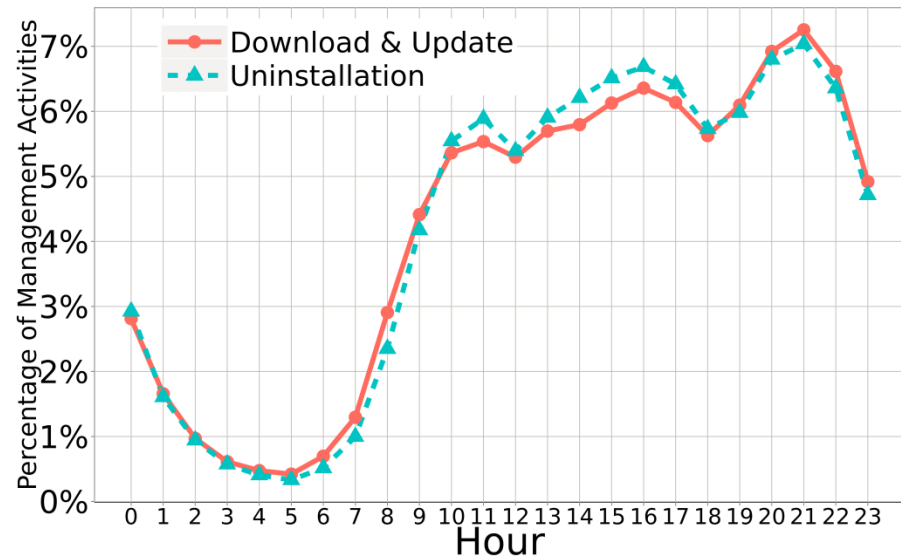


- The data traffic comes from both foreground and background.
- The access time are only from foreground.
- 97% apps consume less than 100 MB traffic volume in one month, and about 95% apps are used less than 100 hours.

APP MANAGEMENT PATTERNS- Duration



- About 32% downloading and updating activities are performed during 7:00 pm to 11:00 pm
- Midnight: still have 7%



APP MANAGEMENT PATTERNS-

App Selection

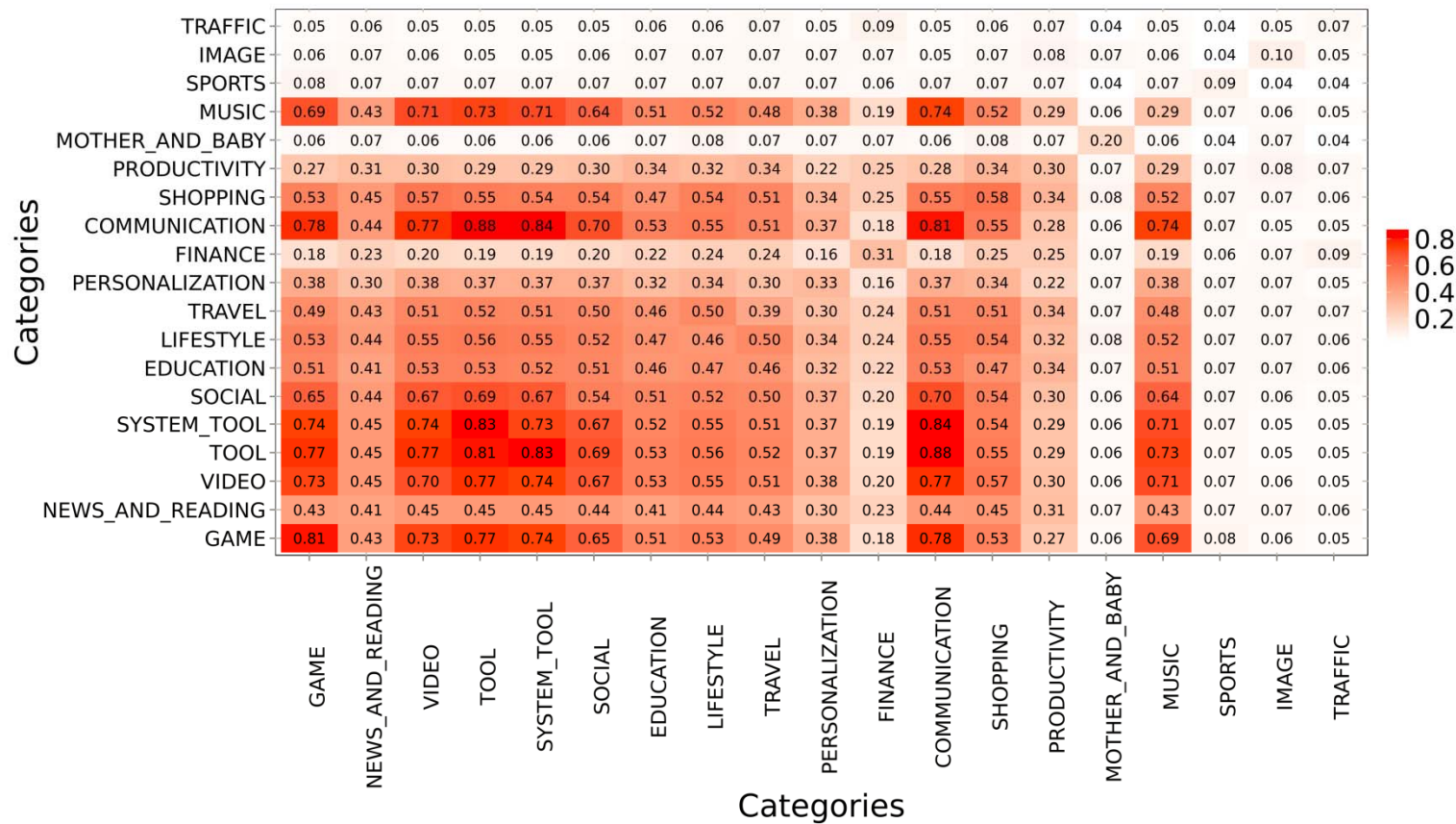


- Jaccard Coefficient: $\frac{\text{number of two things appear together}}{\text{number of at least one appears}}$.
- Apps from same vendor could be installed together.

APP MANAGEMENT PATTERNS- App Selection



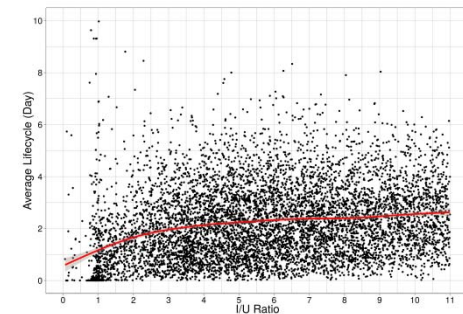
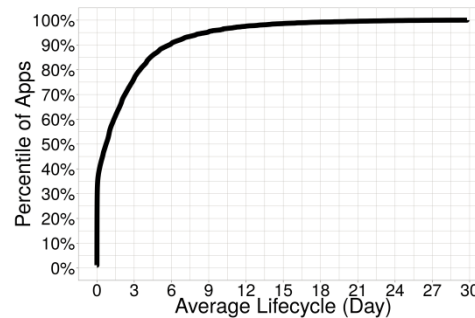
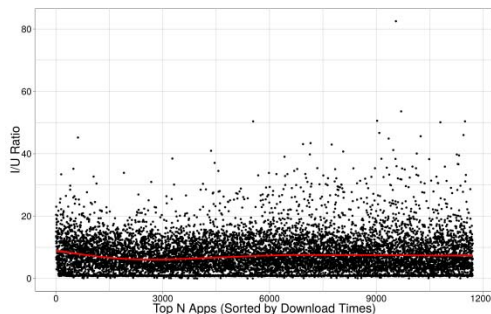
- The co-occurrence of two categories



APP MANAGEMENT PATTERNS- Uninstallation



- I/U ratio: Number of Installation/Number of Uninstallation.
- Most of Apps can't survive more than one day.



NETWORK ACTIVITY PATTERNS



- Distinguish daily data traffic and access time
 - from Wi-Fi and cellular network, respectively
 - from foreground and background, respectively
- Findings:
 - know which apps are network-intensive, and result in more data traffic and battery life



Data Traffic Patterns

- Aggregate apps by their categories and summarize the total traffic consumption from Wi-Fi and cellular, respectively

Table 1: Chosen Top Apps by Category.

App Category	Apps	Users (10 ⁶ devices)	Downloads (10 ⁶ times)	Traffic (GB)	Access- Time (10 ⁷ hours)	C- Traffic	C- Time	W- Traffic	W- Time
GAME	1,227	3.87	15.15	13,669.71	0.38	2.98%	5.19%	0.76%	6.39%
NEWS_AND_READING	274	1.17	1.97	13,143.17	0.23	3.11%	2.91%	0.72%	3.95%
VIDEO	238	2.86	6.52	1,196,978.79	0.38	28.41%	1.42%	81.08%	10.54%
TOOL	227	3.84	9.43	77,329.87	0.68	15.63%	10.79%	4.40%	9.46%
SYSTEM_TOOL	217	3.37	7.54	34,012.16	0.25	3.05%	3.37%	2.17%	4.24%
SOCIAL	188	2.18	4.01	35,926.76	0.35	8.96%	4.77%	1.94%	5.66%
EDUCATION	172	1.68	2.98	13,893.55	0.34	1.46%	5.35%	0.87%	4.71%
LIFESTYLE	156	1.68	2.85	2,388.59	0.07	0.72%	1.00%	0.12%	1.06%
TRAVEL	111	1.62	2.75	8,182.24	0.03	0.78%	0.53%	0.52%	0.25%
PERSONALIZATION	104	1.49	3.68	7,426.38	0.86	0.85%	12.03%	0.46%	13.67%
FINANCE	99	0.32	0.50	382.60	0.02	0.13%	0.24%	0.02%	0.26%
COMMUNICATION	85	4.09	8.45	54,394.71	2.85	24.74%	49.01%	2.26%	35.26%
SHOPPING	78	1.57	3.00	21,808.51	0.07	3.16%	0.65%	1.32%	1.60%
PRODUCTIVITY	75	0.76	1.17	2,712.50	0.01	0.18%	0.17%	0.18%	0.26%
MOTHER_AND_BABY	48	0.10	0.15	525.72	0.01	0.07%	0.04%	0.03%	0.12%
MUSIC	43	2.33	3.39	49,540.12	0.17	5.66%	2.47%	3.08%	2.49%
SPORTS	27	0.31	0.36	61.40	0.00	0.02%	0.05%	0.00%	0.04%
IMAGE	23	0.14	0.17	801.64	0.00	0.06%	0.01%	0.05%	0.03%
TRAFFIC	14	0.10	0.12	78.10	0.00	0.02%	0.03%	0.00%	0.01%

The users, downloads, traffic, and access time are all computed by aggregating the data of each app in the category

The percentile of *W*-Traffic (*C*-Traffic) and *W*-Time (*C*-Time) refer to the data traffic and foreground access time over Wi-Fi (*W*) and cellular (*C*) network, respectively.



Data Traffic of Wi-Fi and Cellular

- Wi-Fi accounts for more than 60% in total traffic.

Table 2: Network Summary by App Category

App Category	<i>C</i> -Traffic (B)	<i>W</i> -Traffic (B)	<i>C</i> -Traffic (F)	<i>W</i> -Traffic (F)	<i>C</i> -Time (B)	<i>W</i> -Time (B)	<i>C</i> -Time (F)	<i>W</i> -Time (F)
VIDEO	0.81%	45.13%	1.28%	52.78%	42.62%	56.66%	0.10%	0.63%
TOOL	8.16%	39.13%	9.56%	43.14%	48.57%	50.42%	0.57%	0.43%
COMMUNICATION	12.42%	15.90%	27.48%	44.20%	48.01%	46.85%	3.15%	1.99%
MUSIC	4.35%	35.19%	5.67%	54.80%	49.23%	50.09%	0.36%	0.32%
SOCIAL	7.26%	20.65%	14.63%	57.47%	48.43%	50.41%	0.57%	0.59%
SYSTEM_TOOL	5.07%	51.57%	2.80%	40.55%	50.02%	49.48%	0.23%	0.26%
SHOPPING	3.29%	17.09%	9.42%	70.21%	43.34%	56.42%	0.08%	0.17%
EDUCATION	3.76%	39.38%	5.46%	51.40%	45.57%	52.83%	0.90%	0.69%
GAME	10.34%	43.11%	8.80%	37.74%	48.13%	51.34%	0.26%	0.28%
NEWS_AND_READING	5.91%	24.64%	14.83%	54.62%	43.43%	55.25%	0.60%	0.71%

W and *C* refer to Wi-Fi and Cellular, respectively.
B refers to background and *F* refers to foreground.



Data Traffic of Foreground and Background

- The background traffic means the app is still connecting to network even users do not interact with it

Table 2: Network Summary by App Category

App Category	<i>C</i> -Traffic (B)	<i>W</i> -Traffic (B)	<i>C</i> -Traffic (F)	<i>W</i> -Traffic (F)	<i>C</i> -Time (B)	<i>W</i> -Time (B)	<i>C</i> -Time (F)	<i>W</i> -Time (F)
VIDEO	0.81%	45.13%	1.28%	52.78%	42.62%	56.66%	0.10%	0.63%
TOOL	8.16%	39.13%	9.56%	43.14%	48.57%	50.42%	0.57%	0.43%
COMMUNICATION	12.42%	15.90%	27.48%	44.20%	48.01%	46.85%	3.15%	1.99%
MUSIC	4.35%	35.19%	5.67%	54.80%	49.23%	50.09%	0.36%	0.32%
SOCIAL	7.26%	20.65%	14.63%	57.47%	48.43%	50.41%	0.57%	0.59%
SYSTEM_TOOL	5.07%	51.57%	2.80%	40.55%	50.02%	49.48%	0.23%	0.26%
SHOPPING	3.29%	17.09%	9.42%	70.21%	43.34%	56.42%	0.08%	0.17%
EDUCATION	3.76%	39.38%	5.46%	51.40%	45.57%	52.83%	0.90%	0.69%
GAME	10.34%	43.11%	8.80%	37.74%	48.13%	51.34%	0.26%	0.28%
NEWS_AND_READING	5.91%	24.64%	14.83%	54.62%	43.43%	55.25%	0.60%	0.71%

W and *C* refer to Wi-Fi and Cellular, respectively.
B refers to background and *F* refers to foreground.



Access Time Patterns

- The foreground access time of an app indicates how long a user interacts with it.
- The background access time indicates how long an app connects to network when users do not interact with it.



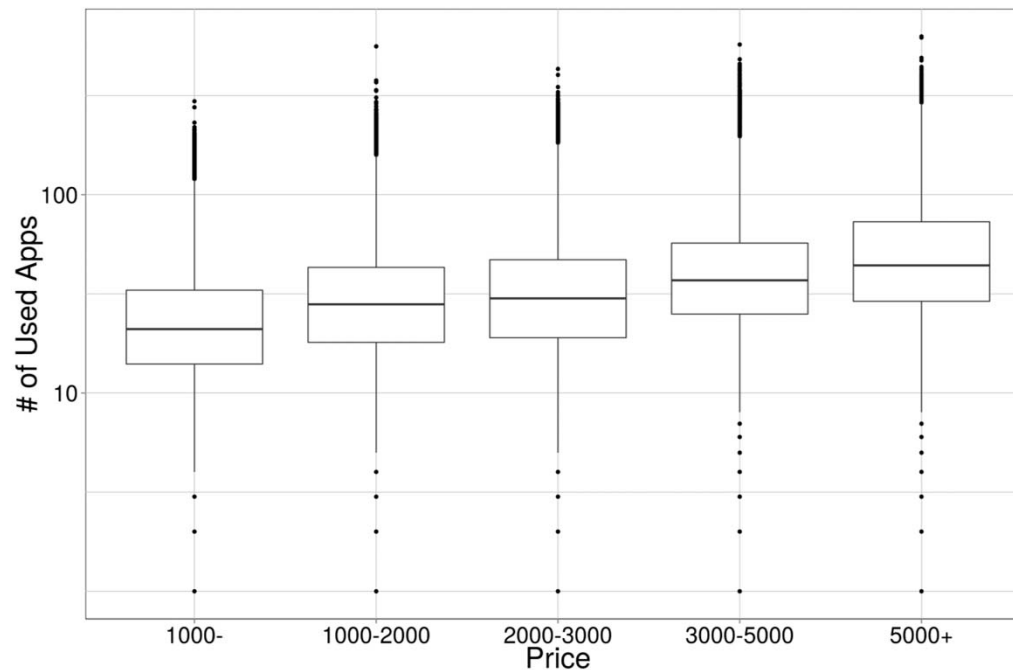
Price-Sensitive Patterns

- Classify users by the prices of their devices
- Device Model Clustering
 - Cluster users by their device models from the Networked App Set, that covers about 2 million users
 - 96% device models have less than 500 users
 - Take a set of “popular” device models
 - Categorize 327 device models according to their on-market sale prices.



Apps Used Among Groups

- The higher price a device model has, the more apps are used on this device model





Network Activity Among Groups

- The usage of Wi-Fi also tends to take a positive correlation with device model prices

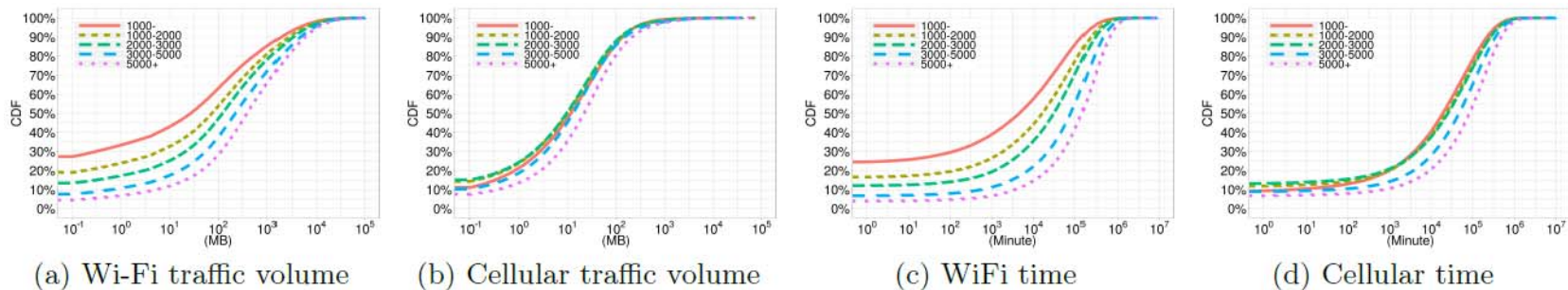
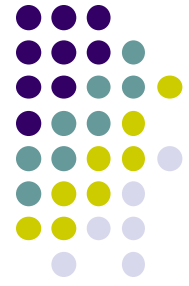


Figure 12: Network Activity Distribution among User Groups



User Interest on Similar Apps

- News Reader:
 - Netease, Sohu, Phoenix, Tencent, and Sina
- Browser:
 - Chrome, FireFox, Opera, Maxthon, UCWeb, 360Safe, Baidu, and Sogou

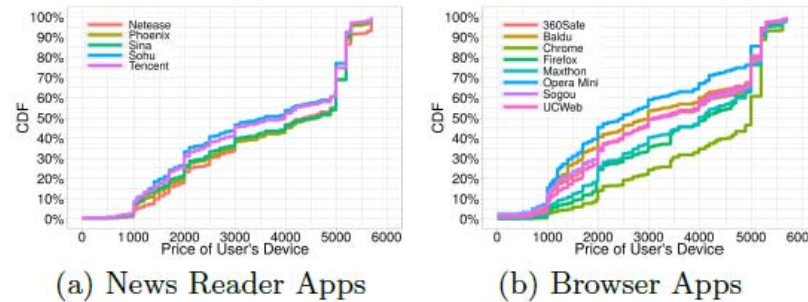


Figure 13: Similar App Preferences among User Groups



Limitations

- The dataset is collected from a single app marketplace in China
- Users under study are mainly Chinese, the region differences should be considered
- The analysis is done on one month of usage data

Conclusion

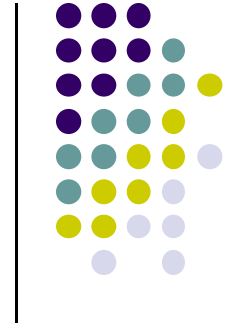


In this paper, the authors conducted a systematic descriptive analysis of a large collection of mobile app usage behaviors. Their findings in usage patterns provide valuable implications for different stakeholders in mobile app industry and research community.



References

- Huoran Li et al., **“Characterizing Smartphone Usage Patterns from Millions of Android Users”**
- N. Chen et al., **AR-miner: mining informative reviews for developers from mobile app marketplace. In Proc. of ICSE, pages 767–778, 2014**
- <https://www.wandoujia.com/>



Thank you!