

Advanced Computer Graphics
CS 525M: *Identifying diverse usage behaviors of smartphone apps*

Alec Mitnik

Computer Science Dept.
Worcester Polytechnic Institute (WPI)





Introduction

- The popularity of mobile devices is increasing.
- Apps are becoming more mainstream.
 - There are over 350k apps at the iOS AppStore with over 10 billion downloads.
- Companies are developing apps instead of just web-based services.
- We don't know nearly as much about app usage as web usage.



Related Work

- Existing studies of app usage have been isolated and small scale.
 - This project wishes to gather nation-wide data for location and time based variations.
- Other studies have used an app that relied on volunteer measurement.
 - This is too challenging, as many APIs don't enable measurement of other apps.



Methodology

- Collect anonymized network traces within a tier-1 cellular network in the U.S. for one week.
- Use HTTP headers and user agents to distinguish individual apps and locations.

top X apps	right (%)	wrong (%)	unknown (%)
10	8 (80%)	0 (0%)	2 (20%)
20	17 (85%)	1 (5%)	2 (10%)
50	46 (92%)	2 (4%)	2 (4%)
100	91 (91%)	4 (4%)	5 (5%)
200	176 (88%)	5 (3%)	19 (10%)
500	427 (85%)	14 (3%)	69 (14%)

Table 1: accuracy of using `User-Agent` to categorize apps (via manual comparison to app names in the app marketplace).



Methodology

- Record four main features for each app:
 - Traffic volume
 - Access time
 - Unique subscribers
 - Locations
- Use uniform random sampling to prevent traffic overflow.
- Only recognize apps involving network flows, but the interest of the study is just such apps anyway.



Results

- Recorded data for a total of about 600K individual devices and about 22K individual apps.
- When analyzing traffic volume, access time, and number of subscribers, many apps have very small values and do not provide enough data to analyze, so are excluded from detailed analysis.

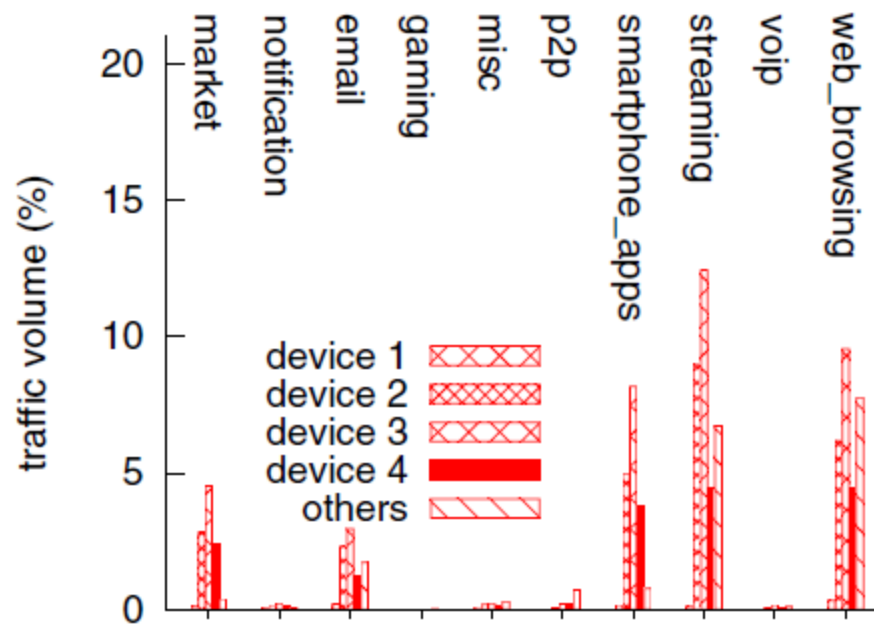
genre	books	business	education	entertainment	finance	games	healthcare	lifestyle	medical	music	navigation	news	photography	productivity	reference	social net.	sports	travel	utilities	weather	unknown
total apps	351	418	643	1827	283	3108	368	1298	205	1296	450	1126	475	527	515	721	787	590	1079	236	5865
popular apps	7	8	13	95	13	199	23	71	4	34	23	89	17	23	26	58	33	29	49	16	170

Table 2: distribution of the genre of apps.



Traffic Volume

- Note that app and web browsing traffic are comparable, and the significant market traffic.

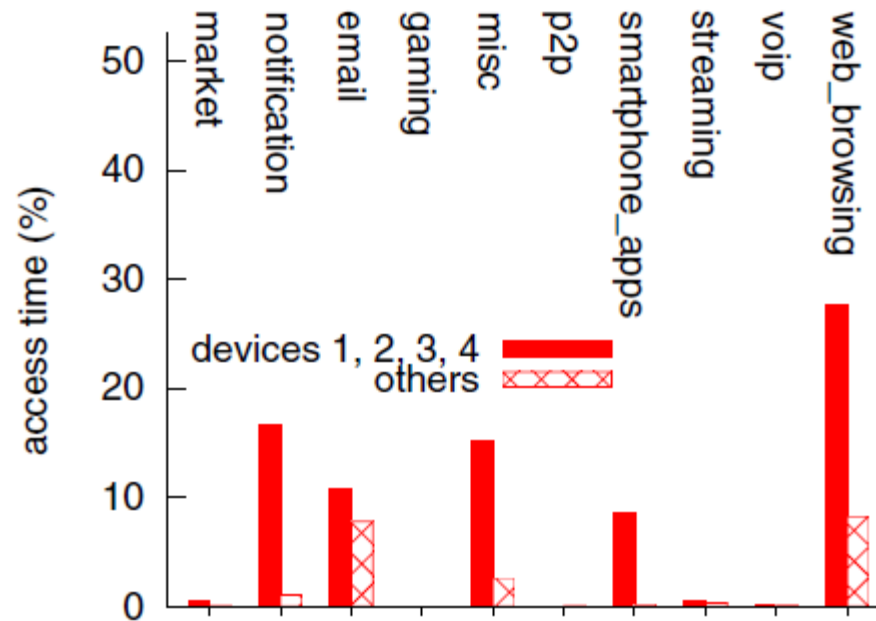


a. distribution of traffic volume



Access Time

- Gaming, p2p, and voip seem to not be commonly used on the captured devices.

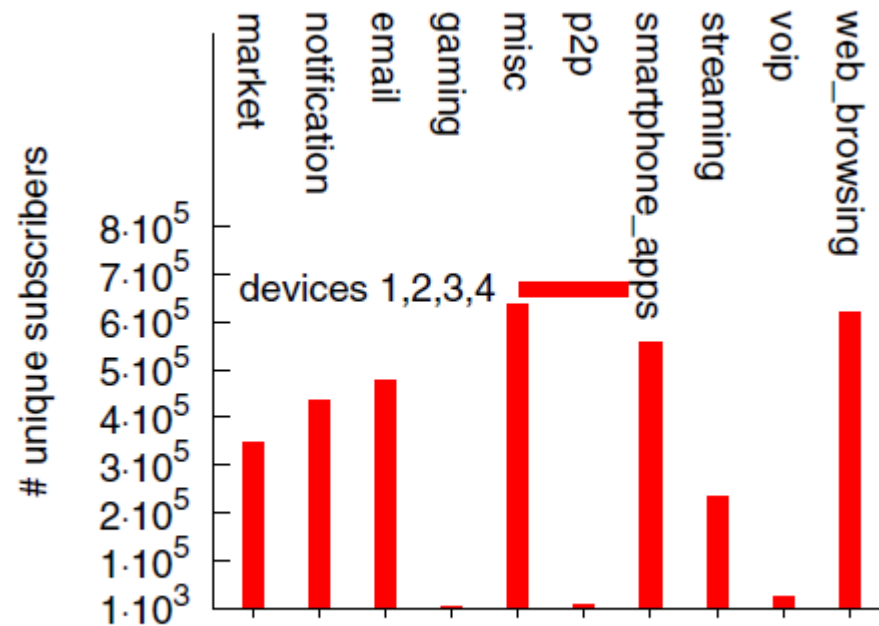


b. distribution of access time



Unique Subscribers

- The misc value reflects the total number of subscribers. Almost all use web browsing and apps.



c. distribution of subscribers

Results for Smartphone Apps Only

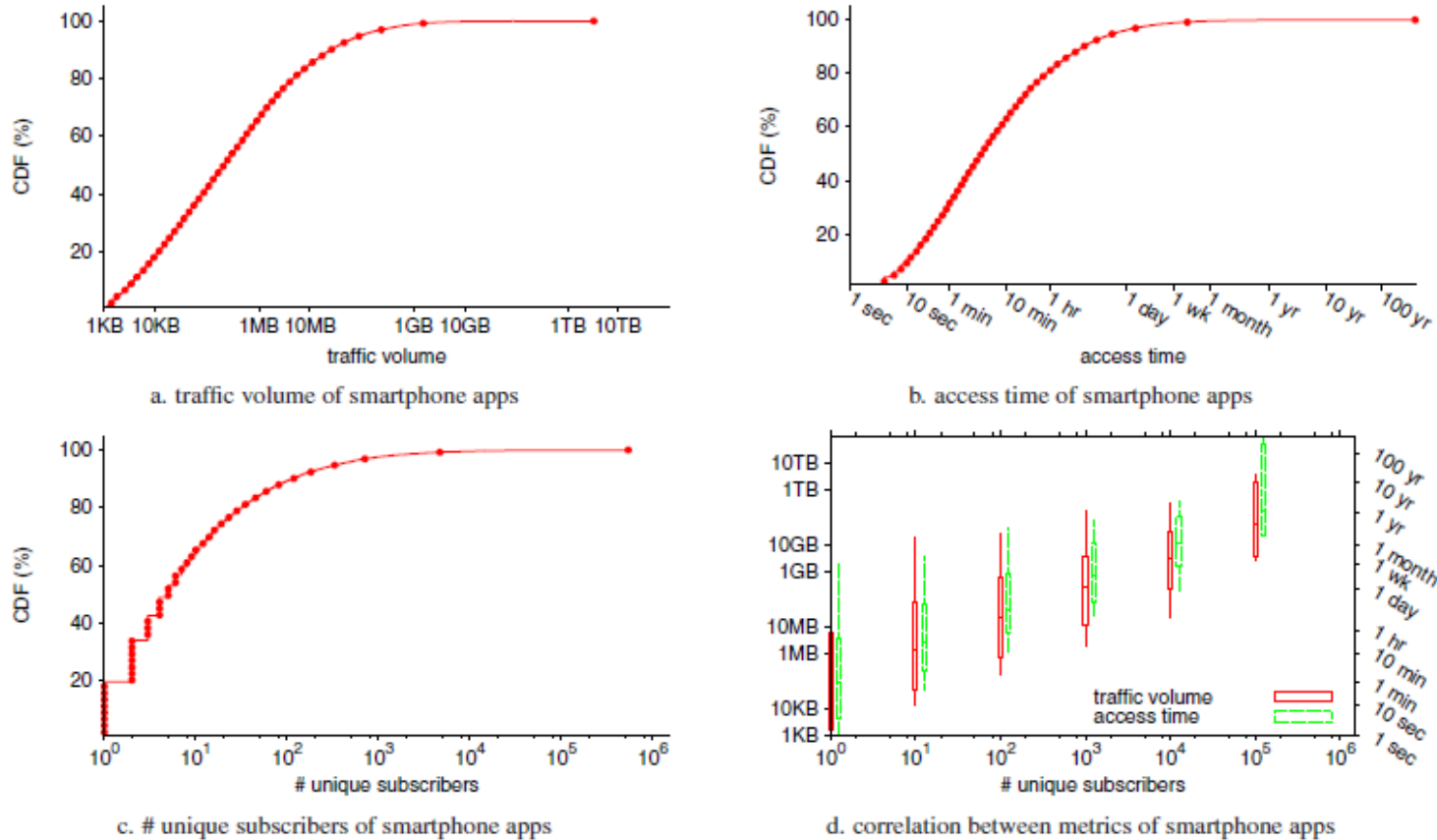
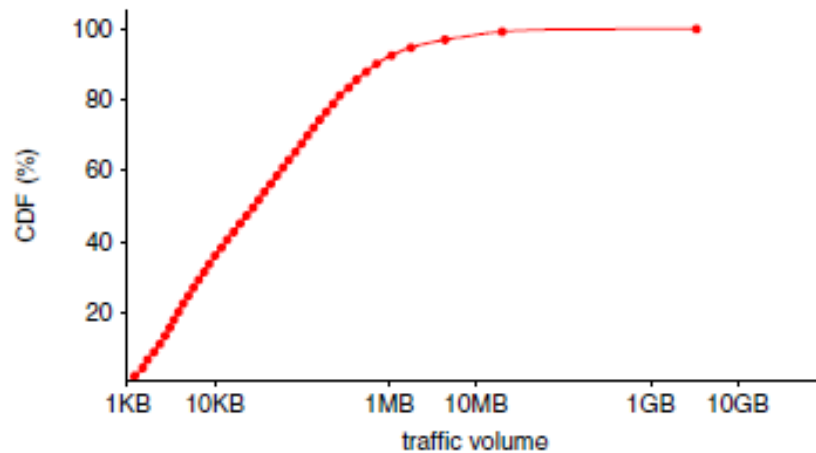


Figure 2: (a)-(c) CDF of volume, access time, and users, with one data point per smartphone app, aggregating users together in one week.

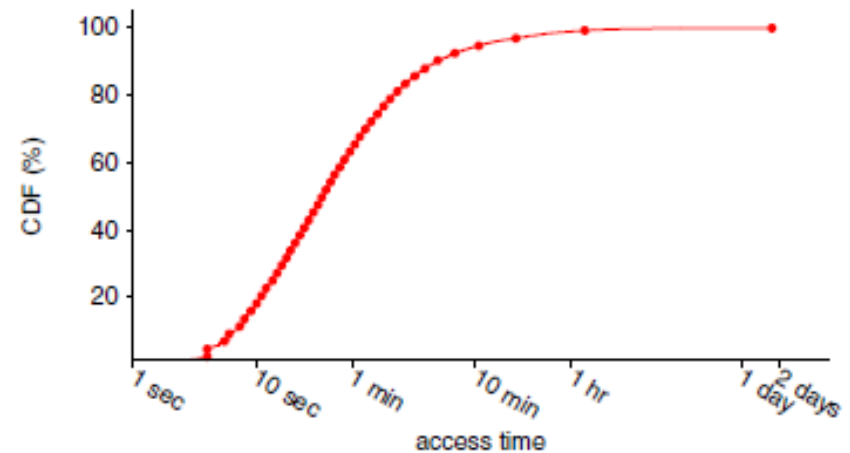


Results Normalized by Subscribers

- A few big values, but most are very small. Must filter out small values for proper analysis.



a. traffic volume of smartphone apps per subscriber



b. access time of smartphone apps per subscriber

Figure 3: metrics of smartphone apps (ii) – averaged based on users.



Filtering Results

- Over 90% of total traffic and access time is contained within the 1000 most-subscribed apps.

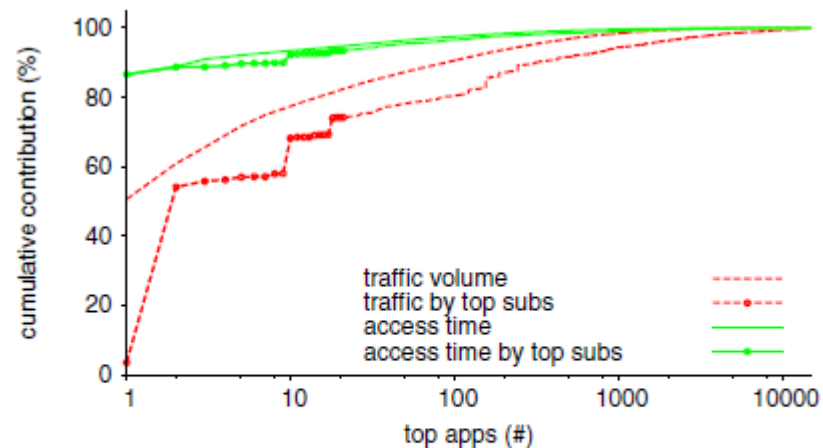


Figure 4: is the number of unique subscribers a good metric for filtering? Contributions of the top X apps to total volume and access time respectively



Results for Location

- 20% of popular apps are local, such as radio or news.
 - Amounts to 2% of total traffic.

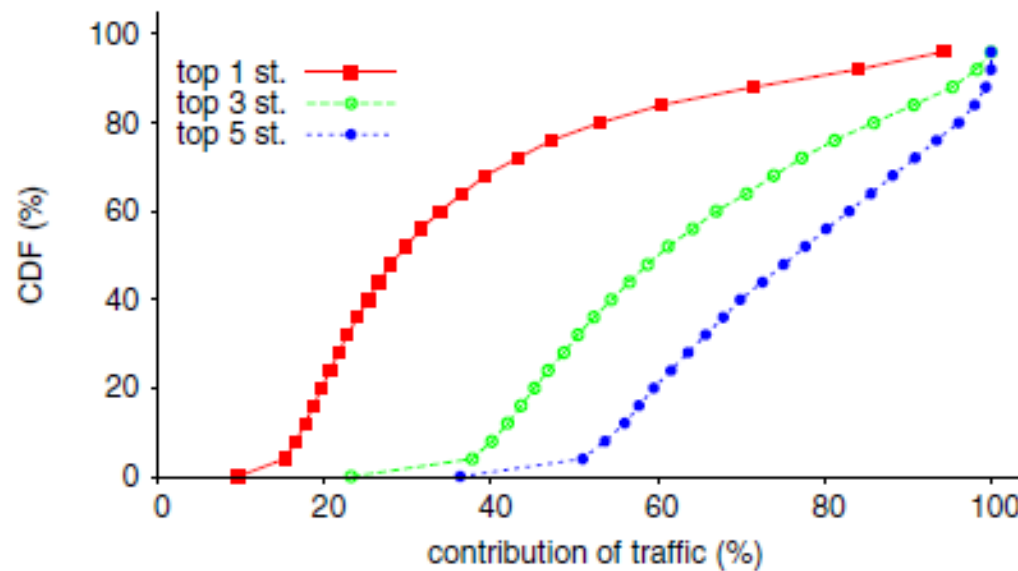


Figure 6: contribution of volume from top X states.

Results for Location of “National” Apps, by Genre

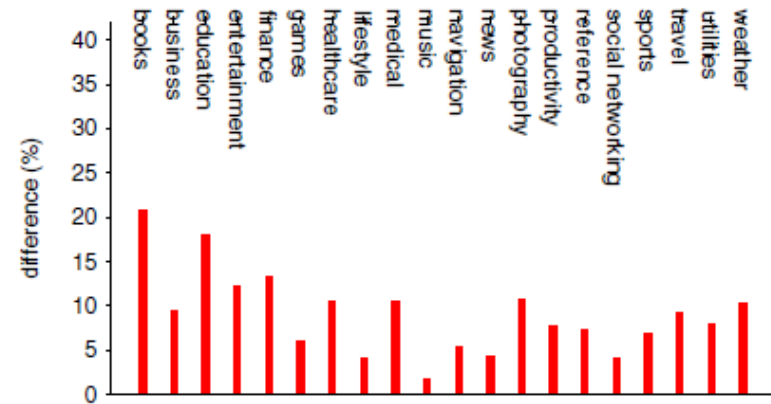


Figure 9: difference in the geographic usage of different app genres

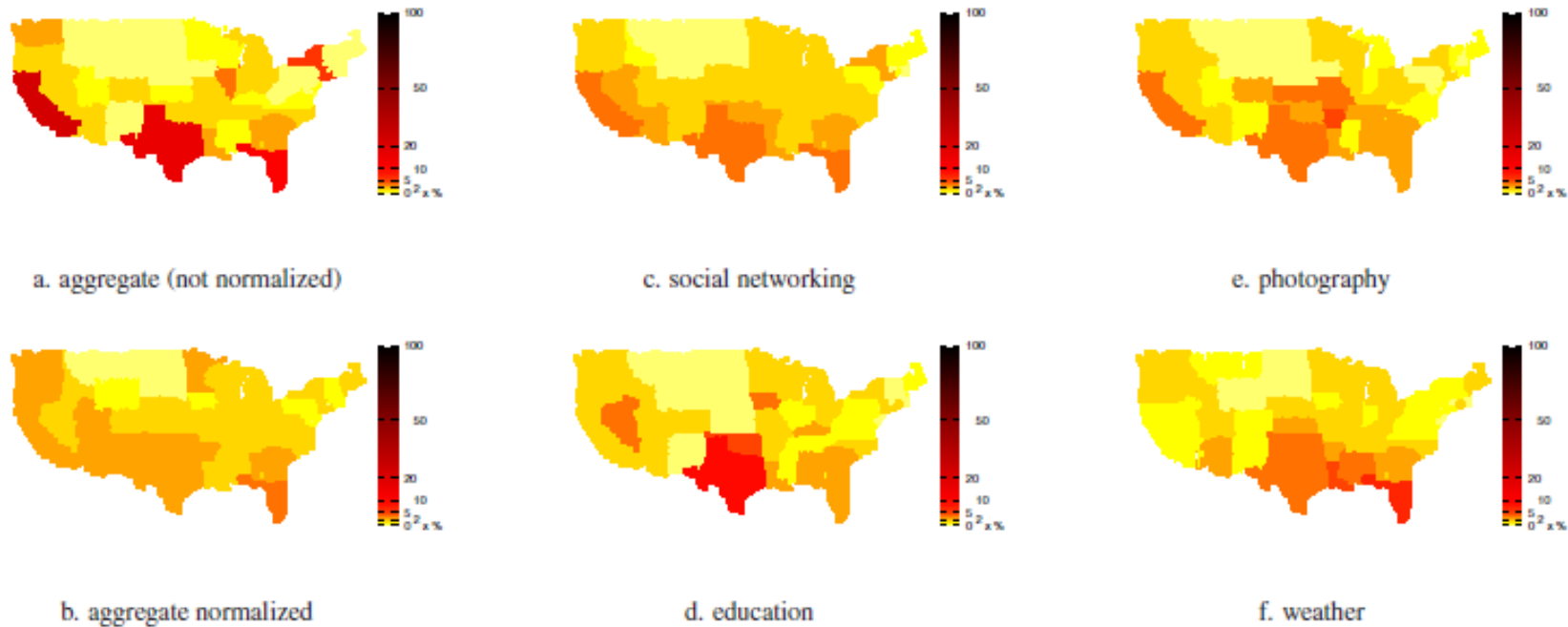


Figure 8: distribution of the geographic usage of apps in different genres.



Results for Mobility

- About 10% of apps access the network more than two sectors.
- Most mobile apps are social networks or games.

genre	books	business	finance	games	healthcare	lifestyle	music	news	productivity	reference	social net.	sports	travel	unknown
# apps	2	1	4	19	3	5	6	1	3	1	31	2	1	18

Table 5: genres of high travel-area smartphone apps.

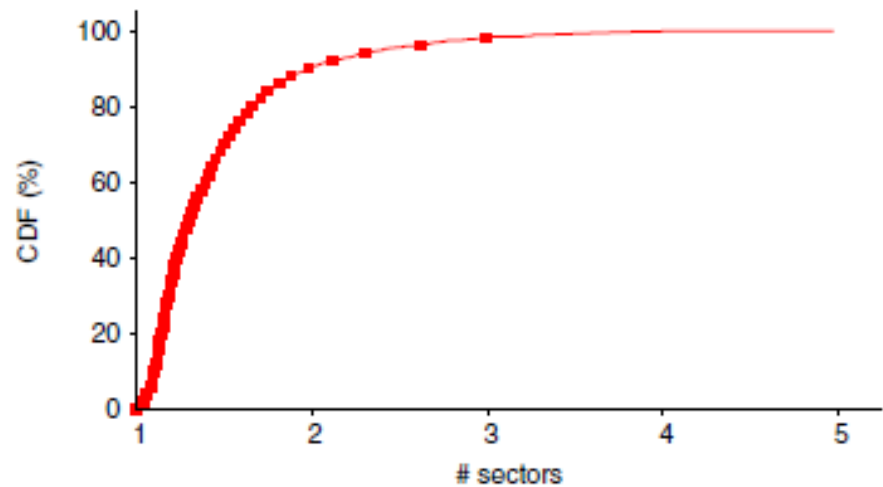


Figure 11: travel-area of apps.



Results for Correlation

- A JSC of 0.05 for two apps with 2000 subscribers each means 100 subscribers use both.
- Popular apps share more subscribers, naturally.

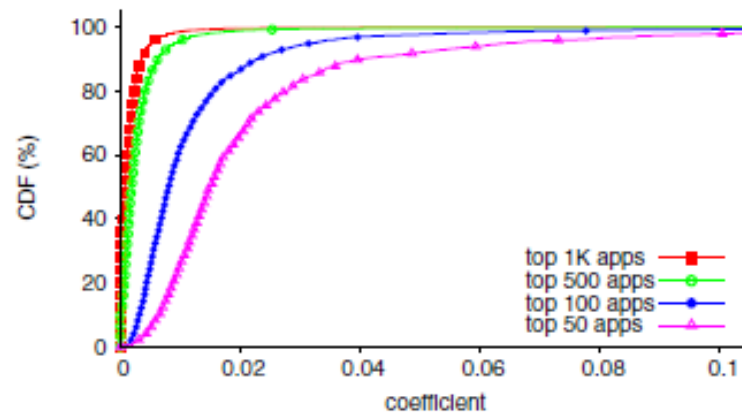


Figure 12: distribution of the Jaccard Similarity Coefficient of the popular apps.

$$\frac{\text{joint}(a,b)}{\text{union}(a,b)}$$

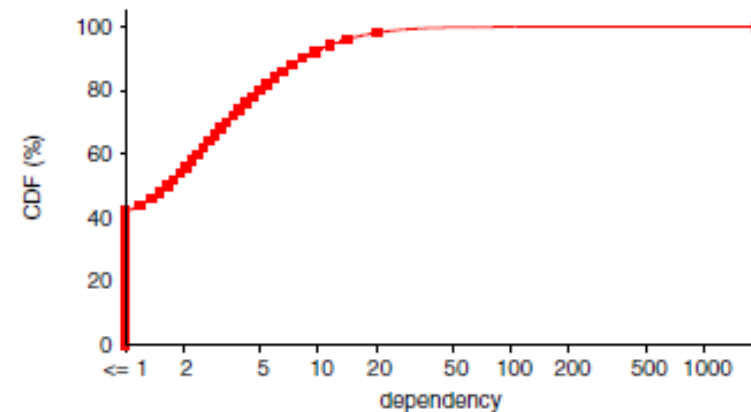
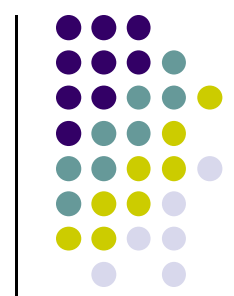
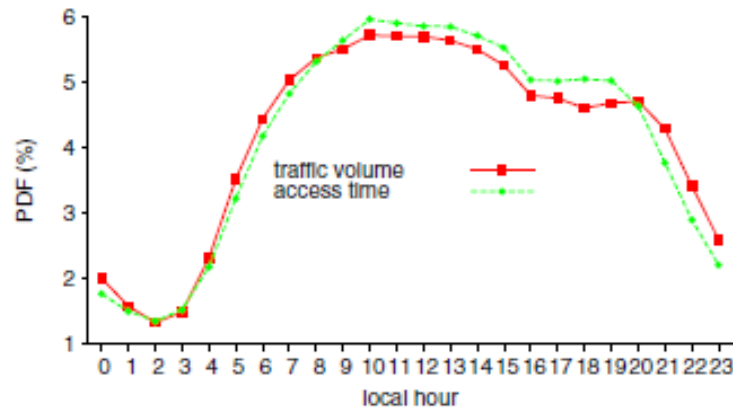


Figure 13: distribution of the dependency between popular apps.

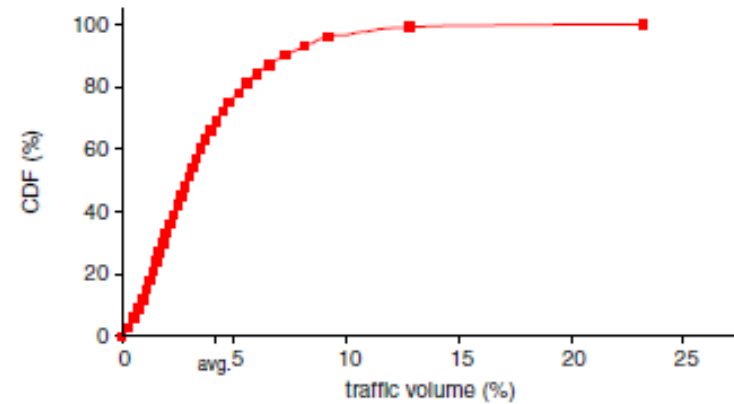
$$\frac{\text{Pr}[ab]}{\text{Pr}[a][b]}$$



Results for Temporal Patterns



a. diurnal patterns of traffic volume & access time



b. traffic contribution by late night (1:00 – 3:59 AM)

Figure 14: diurnal patterns.

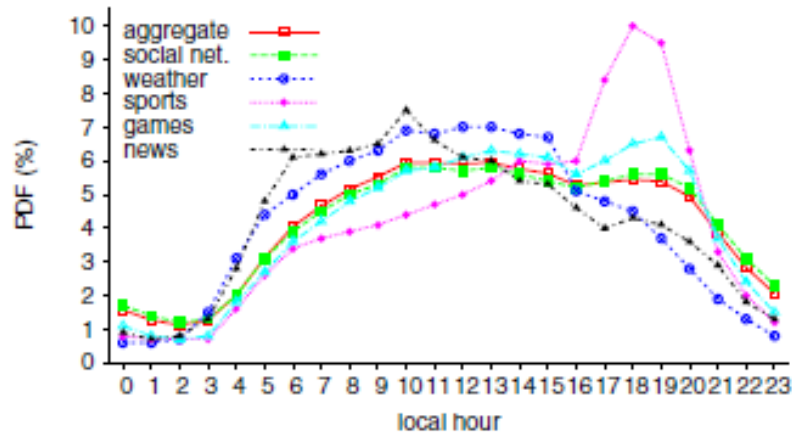


Figure 15: diurnal patterns across different genres.

category	# apps	description
entertainment	20	small games, video channels, <i>etc.</i>
radio	28	music radio channels, news radio channels, <i>etc.</i>
healthcare	12	sleep aid utilities, <i>etc.</i>
books	6	bible, references, <i>etc.</i>

Table 7: description of late night apps.



Results for Devices

- More advanced devices consume more traffic.
- Power users likely upgrade to latest devices.

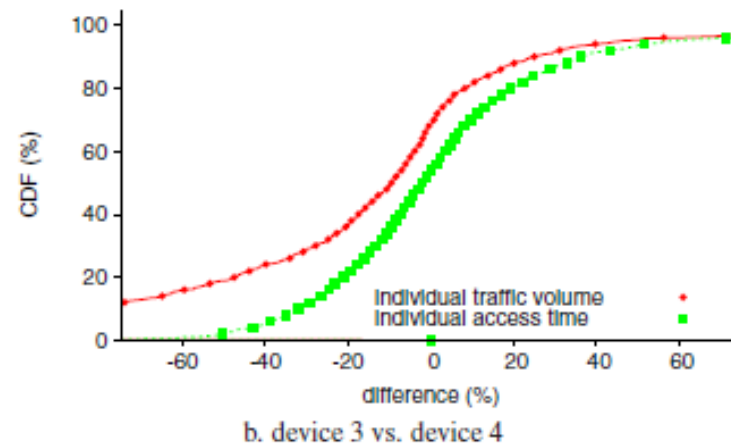
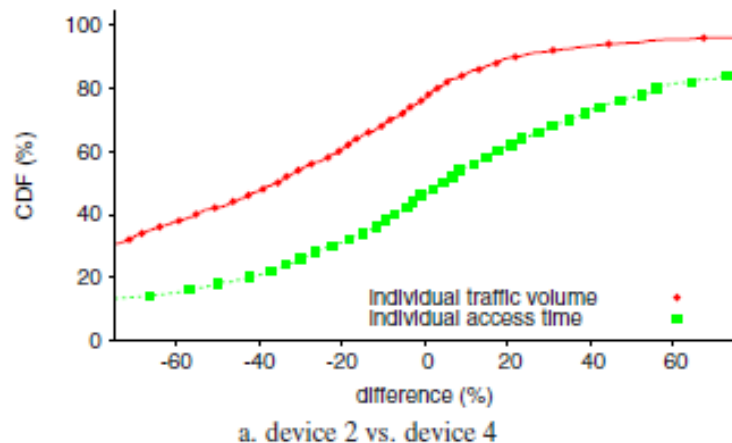


Figure 16: impact of devices used.



Conclusions

- Findings show many opportunities for optimization (such as moving content to local servers) and profiling (for recommending apps).
- Some apps are often used together, and some types of apps have alternatives that are interchangeable.
- There are trends in time of use (news in morning, sports in evening).
- There are trends in use while stationary or mobile.
- Results are mostly intuitive...

Thoughts



- Very broad and thorough analysis, but bland results.
- Is one week really long enough?
 - Weather app usage during hurricane season
- Why not name specific companies?
 - Tier-1 cellular network
 - Personalized Internet radio app
 - Social utility connecting people app (Facebook?)
 - 6 out of 7 devices use it, according to the data
- Graphs should use different colors instead of or in addition to different patterns.



References

- ***Identifying diverse usage behaviors of smartphone apps***
Qiang Xu, Jeffrey Erman, Alexandre Gerber, Zhuoqing Mao, Jeffrey Pang, Shobha Venkataramanin in Proc IMC 2011
<http://delivery.acm.org/10.1145/2070000/2068847/p329-xu.pdf?ip=130.215.29.166&acc=ACTIVE%20SERVICE&CFID=160083051&CFTOKEN=61884003&acm=1357578921fd49d3071b1b7accda3adef7c2eeb94c>