Have you ever wondered...

- Where people who attend a football game on Sunday go after the game is over
- Where people who come to a mall on Saturday afternoon were in the morning
- What are the neighborhoods most at risk if a given individual is found to be infected with a new strand of the flu virus
- How the construction of a new highway in a given neighborhood might impact people’s activities
- What people in a given area of town are most likely to be doing on a given day and at a given time
- How you might want to modulate public bus schedules to accommodate shifting traffic demands over time
Taken together, these mental maps represent the collective wisdom of a city's inhabitants. They also provide unique insights into the social fabric of a city.

Could social media help us?

The Livehoods Project

18 million foursquare check-ins that people shared publicly on their Twitter feeds

(patent pending)

"J. Cranshaw, R. Schwartz, J. Hong, N. Sadeh, "The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City", in Proc. of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM-12), Dublin, Ireland, June 2012 – best paper award

www.livehoods.org
info@livehoods.org
We can aggregate these patterns to compute relationships between check-in venues.

These relationships can then be used to identify natural borders in the urban landscape.

We call the discovered clusters “Livehoods” reflecting their dynamic character.
Whenever I was living down on 15th Street [LH7] I had to worry about drunk people following me home, but on 23rd [LH8] I need to worry about people trying to mug you... so it's different. It's not something I had anticipated, but there is a distinct difference between the two areas of the South Side.

"Whenever I was living down on 15th Street [LH7] I had to worry about drunk people following me home, but on 23rd [LH8] I need to worry about people trying to mug you... so it's different. It's not something I had anticipated, but there is a distinct difference between the two areas of the South Side."

There is this interesting mix of people there [LH6] I don't see walking around the neighborhood. I think they are coming to the Giant Eagle from lower income neighborhoods... I always assumed they came from up the hill.

"They came from up the hill..."

"from an urban standpoint it is a lot tighter on the western part once you get west of 17th or 18th [LH7]."
Beyond Municipal Boundaries

Some Limitations
Foursquare data is skewed
...but the methodology can be applied to other data sets

Typology of a City
- Can we identify canonical city neighborhoods – across different cities?
  - Shopping areas
  - University neighborhoods
  - Entertainment districts

Topic Modeling Applied to Venues

Table 2. Here we highlight 10 topics discovered by posterior inference on our model. The 15 most probable categories from each topic are shown. Topic names are supplied by us to capture what we felt each collection of venues represents.

(patent pending)

Examples: College Neighborhoods

Or Shopping Areas in Lower Manhattan
SoHo, NoLIta, Union Square (and Madison Square), Harold Square, and the Meat Packing District

Imagine the Possibilities
Urban Planning
Transportation
Real Estate
Personal Recommendations
Public Health
Travel & Tourism
Public Safety
Marketing
The WisCom Project

13 million mobile app reviews
171,000 Android Apps

*Why People Hate Your App: Making Sense of User Feedback in a Mobile App Store*,
Bin Fu, Jialiu Lin, Lei Li, Jason I. Hong,
Christos Faloutsos, Norman Sadeh – to appear KDD 2013

Life Before the App Store

- Mobile Apps have been around for a dozen years...
- ...but, prior to the App store, few people:
  - Knew how to find and download apps
  - Were willing to give their credit card details to companies they had never heard of
- …at best 2 guys in a garage…at worst a scheme to get your credit card credentials

The App Store is about addressing fundamental trust and usability issues

Lots of Bad Apps

- From malware

Could App stores leverage user comments to automatically take down bad apps and help developers fix their apps?

...and more...

*Why People Hate Your App: Making Sense of User Feedback in a Mobile App Store*,
Bin Fu, Jialiu Lin, Lei Li, Jason I. Hong,
Christos Faloutsos, Norman Sadeh – to appear KDD 2013
Distinct Features of App Reviews (in contrast to other product reviews)

- Shorter in length
- More typos, slang, ill-structured sentences
- Reviews of the same app can refer to different versions

Motivations for Analyzing User Reviews

- **App store**: Manual inspection of hundreds of thousands of apps is impractical. Maintaining user trust & prevent manipulation
- **End-user**: Develop summary of what others think about an app
- **App developer**: Help identify & diagnose problems associated with user complaints
- Analyze overall market & spot trends
  - What are those issues that matter most to users in different app categories

WisCom: A Multi-level Review Summarization System

- **Per Review Analysis**: Sentiment analysis on each review, identify inconsistent reviews
- **Per App Analysis**: Topic modeling on aggregated reviews, detect major releases and identify major sources of complaints
- **Market Analysis**: analyze market segments & spot trends, incl. user preferences. Can also help develop/refine app development guidelines

Dataset: 13 million Android App Reviews

- Crawled Google Play – Looked at 171,000 Android apps (Nov. 2012)
- Info about the identity of the reviewer is not available
Apps by Categories

Dataset

- The number of comments follows a heavy tailed distribution.
- Average number of comments = 101.90, STD = 460.25, median=8

Dataset

PER REVIEW ANALYSIS

- Avg rating is 3.9 and 54% of apps are rated 5 stars.
- Used 3 as threshold between "positive review" and "negative review".

Preprocessing

- Sample 8% of the comments (1 million reviews)
- Remove HTML tags and delete non-English reviews
- Segment comments into words
- Remove rarely used words.
- Yield 19K distinct words, from 988K comments.
Linear Regression Model

\[ L(W) = \sum_{i=1}^{m} (Y_i - (w_0 + \sum_{j=1}^{n} X_{ij} w_j))^2 + \gamma P(W) \]

Where \( m = 988K, n = 19K \), \( Y \): reported ratings
\( X_{ij} \) is frequency of \( j \)-th word in \( i \)-th comment
\( W_j \) is the weight of \( j \)-th word in vocabulary

\[ P(W) = \sum_{i=1}^{n} \left( (1 - \alpha) w_i^2 + \alpha |w_i| \right) \]

Normal Tikhonov regularization to reduce overfitting,
L1 norm, penalty of complexity of model – forces a number of zero weights for words that don’t matter

Most Positive and Negative Words

<table>
<thead>
<tr>
<th>Word</th>
<th>Weight</th>
<th>Freq</th>
<th>Word</th>
<th>Weight</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>awesome</td>
<td>0.67</td>
<td>4893</td>
<td>sucks</td>
<td>-1.24</td>
<td>13178</td>
</tr>
<tr>
<td>excellent</td>
<td>0.67</td>
<td>31971</td>
<td>lame</td>
<td>-1.16</td>
<td>2701</td>
</tr>
<tr>
<td>awesome</td>
<td>0.63</td>
<td>63257</td>
<td>rubbish</td>
<td>-1.12</td>
<td>2127</td>
</tr>
<tr>
<td>fault</td>
<td>0.61</td>
<td>1027</td>
<td>worthless</td>
<td>-1.03</td>
<td>1628</td>
</tr>
<tr>
<td>sweet</td>
<td>0.60</td>
<td>3572</td>
<td>poor</td>
<td>-1.02</td>
<td>6307</td>
</tr>
<tr>
<td>superb</td>
<td>0.58</td>
<td>3694</td>
<td>boring</td>
<td>-0.99</td>
<td>3529</td>
</tr>
<tr>
<td>brilliant</td>
<td>0.58</td>
<td>6384</td>
<td>useless</td>
<td>-0.98</td>
<td>8075</td>
</tr>
<tr>
<td>yay</td>
<td>0.57</td>
<td>1134</td>
<td>horrible</td>
<td>-0.96</td>
<td>4428</td>
</tr>
<tr>
<td>greatest</td>
<td>0.56</td>
<td>1148</td>
<td>crap</td>
<td>-0.95</td>
<td>7515</td>
</tr>
<tr>
<td>amazing</td>
<td>0.56</td>
<td>18753</td>
<td>garbage</td>
<td>-0.93</td>
<td>2217</td>
</tr>
</tbody>
</table>

Examples of Negative Words that Imply Problems

<table>
<thead>
<tr>
<th>Word</th>
<th>Weight</th>
<th>Freq</th>
<th>Word</th>
<th>Weight</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>bloatware</td>
<td>-0.89</td>
<td>1778</td>
<td>slow</td>
<td>-0.44</td>
<td>9939</td>
</tr>
<tr>
<td>misleading</td>
<td>-0.84</td>
<td>465</td>
<td>confusing</td>
<td>-0.38</td>
<td>1300</td>
</tr>
<tr>
<td>crashes</td>
<td>-0.71</td>
<td>9081</td>
<td>expensive</td>
<td>-0.26</td>
<td>1538</td>
</tr>
<tr>
<td>spam</td>
<td>-0.62</td>
<td>1601</td>
<td>permission</td>
<td>-0.18</td>
<td>1409</td>
</tr>
<tr>
<td>freezes</td>
<td>-0.54</td>
<td>3960</td>
<td>privacy</td>
<td>-0.10</td>
<td>962</td>
</tr>
</tbody>
</table>

Some negative words are particularly informative and hint to the source of a user’s complaint

Detect Inconsistent Reviews

- 0.9% of the reviews are inconsistent with their ratings.
- Might be attributed to carelessness or could be sign of intentional manipulation
- Inconsistent reviews were removed for later analysis
PER APP ANALYSIS

Can we monitor apps and, based on reviews, identify and help diagnose problems associated with complaints?

Topic Analysis

- Preprocessing:
  - Only on negative reviews, and negative words → avg. 71 characters, median length is 47 characters.
  - Removed inconsistent comments
  - Only negative comments associated with 1-star and 2-star ratings
  - Removed words with non-negative weights
  - Concatenate comments from the same app as documents (otherwise too short)
  - Remove apps with too few comments: This left us with >52K apps to analyze

- Latent Dirichlet Allocation model (LDA)
  - Identify 10 topics

Top 10 Root Causes

<table>
<thead>
<tr>
<th>Words</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>boring</td>
<td>18.2%</td>
</tr>
<tr>
<td>close</td>
<td>17.6%</td>
</tr>
<tr>
<td>location</td>
<td>17.6%</td>
</tr>
<tr>
<td>galaxy</td>
<td>17.5%</td>
</tr>
<tr>
<td>battery</td>
<td>17.1%</td>
</tr>
<tr>
<td>error</td>
<td>16.7%</td>
</tr>
<tr>
<td>money</td>
<td>16.0%</td>
</tr>
<tr>
<td>want</td>
<td>15.0%</td>
</tr>
<tr>
<td>picture</td>
<td>14.0%</td>
</tr>
<tr>
<td>sound</td>
<td>13.7%</td>
</tr>
<tr>
<td>notification</td>
<td>12.9%</td>
</tr>
</tbody>
</table>

Dynamic View: Life Story of App (I)
Market-Level Analysis

- Do users care about the same issues and/or exhibit the same level of tolerance for different app categories?
- Possible guidelines for developers

Outstanding Complaints in Each Category

<table>
<thead>
<tr>
<th>Category</th>
<th>1st Complaints</th>
<th>2nd Complaints</th>
<th>3rd Complaints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arcade &amp; Action</td>
<td>Attractiveness (60%)</td>
<td>Stability (18%)</td>
<td>Cost (11%)</td>
</tr>
<tr>
<td>Brain &amp; Puzzle</td>
<td>Attractiveness (49%)</td>
<td>Stability (18%)</td>
<td>Cost (8%)</td>
</tr>
<tr>
<td>Cards &amp; Casino</td>
<td>Attractiveness (41%)</td>
<td>Cost (25%)</td>
<td>Stability (10%)</td>
</tr>
<tr>
<td>Racing</td>
<td>Attractiveness (61%)</td>
<td>Stability (14%)</td>
<td>Cost (11%)</td>
</tr>
<tr>
<td>Sports Games</td>
<td>Attractiveness (65%)</td>
<td>Stability (15%)</td>
<td>Cost (10%)</td>
</tr>
<tr>
<td>Casual</td>
<td>Attractiveness (54%)</td>
<td>Stability (17%)</td>
<td>Cost (8%)</td>
</tr>
<tr>
<td>Books &amp; Reference</td>
<td>Accuracy (26%)</td>
<td>Phone (13%)</td>
<td>Connection (43%)</td>
</tr>
<tr>
<td>Business</td>
<td>Connection (31%)</td>
<td>Accuracy (22%)</td>
<td>Cost (15%)</td>
</tr>
<tr>
<td>Comics</td>
<td>Attractiveness (29%)</td>
<td>Picture (17%)</td>
<td>Connection (46%)</td>
</tr>
<tr>
<td>Communication</td>
<td>Phone (33%)</td>
<td>Connection (38%)</td>
<td>Compatibility (13%)</td>
</tr>
<tr>
<td>Education</td>
<td>Attractiveness (17%)</td>
<td>Accuracy (13%)</td>
<td>Phone (43%)</td>
</tr>
<tr>
<td>Entertainment</td>
<td>Attractiveness (26%)</td>
<td>Media (16%)</td>
<td>Stability (11%)</td>
</tr>
<tr>
<td>Finance</td>
<td>Connection (43%)</td>
<td>Accuracy (25%)</td>
<td>Cost (9%)</td>
</tr>
<tr>
<td>Health &amp; Fitness</td>
<td>Accuracy (39%)</td>
<td>Stability (14%)</td>
<td>Attractiveness (10%)</td>
</tr>
<tr>
<td>Libraries &amp; Demo</td>
<td>Attractiveness (21%)</td>
<td>Compatibility (16%)</td>
<td>Phone (15%)</td>
</tr>
<tr>
<td>Lifestyle</td>
<td>Accuracy (26%)</td>
<td>Stability (12%)</td>
<td>Connection (12%)</td>
</tr>
<tr>
<td>Media &amp; Video</td>
<td>Media (28%)</td>
<td>Picture (15%)</td>
<td>Stability (12%)</td>
</tr>
<tr>
<td>Medical</td>
<td>Accuracy (36%)</td>
<td>Cost (19%)</td>
<td>Connection (12%)</td>
</tr>
<tr>
<td>Music &amp; Audio</td>
<td>Media (32%)</td>
<td>Stability (17%)</td>
<td>Attractiveness (11%)</td>
</tr>
<tr>
<td>News &amp; Magazines</td>
<td>Connection (40%)</td>
<td>Stability (19%)</td>
<td>Accuracy (12%)</td>
</tr>
<tr>
<td>Personalization</td>
<td>Picture (29%)</td>
<td>Compatibility (19%)</td>
<td>Spam (13%)</td>
</tr>
<tr>
<td>Photography</td>
<td>Picture (61%)</td>
<td>Stability (10%)</td>
<td>Cost (6%)</td>
</tr>
<tr>
<td>Productivity</td>
<td>Accuracy (31%)</td>
<td>Compatibility (16%)</td>
<td>Connection (15%)</td>
</tr>
<tr>
<td>Shopping</td>
<td>Accuracy (38%)</td>
<td>Connection (16%)</td>
<td>Stability (12%)</td>
</tr>
<tr>
<td>Social</td>
<td>Connection (34%)</td>
<td>Phone (13%)</td>
<td>Stability (12%)</td>
</tr>
<tr>
<td>Sports</td>
<td>Connection (25%)</td>
<td>Accuracy (21%)</td>
<td>Stability (16%)</td>
</tr>
<tr>
<td>Tools</td>
<td>Compatibility (29%)</td>
<td>Accuracy (16%)</td>
<td>Phone (15%)</td>
</tr>
</tbody>
</table>

Summary

- WisCom provides analysis at 3 different levels
- WisCom can
  - Detect inconsistent reviews,
  - Identify root causes of complaints
  - Provide historical view of apps
  - Help identify market trends and develop guidelines

Users’ Reception of Free and Paid Apps

100 free/paid apps/games with the most reviews:
- For paid apps: cost dominates
- For paid games: other factors matter too
Concluding Remarks

- Lots of complex decisions can benefit from crowd sourcing
- Combining ML and crowd sourcing can help us:
  - Understand complex questions, automate processes, and more
  - Make crowd sourcing more efficient
- This presentation was intended to illustrate some of the research in this area at the Mobile Commerce Lab.