Extracting data from MySpace

-an example of large scale analysis of information about individuals

Mike Thelwall
Statistical Cybermetrics Research Group
University of Wolverhampton, UK
MySpace members data

1. a random sample of 15,043 members
2. a systematic sample of 7,627 members who joined on July 3, 2006
3. 403 members from July 3, 2006
   - excluded: music sites, ex-members
   - excluded (normally): members with 0 or 1 friends
   - all information on home pages was automatically downloaded and harvested by SocSciBot 4 -> Excel

Can access profiles of 1/3 of members with public profiles.
days since last access - members use MySpace once or frequently

all Members >>
days since last access - members use MySpace once or frequently

July 3 members
MySpace “age” profile: average “age” is 21
younger members have more MySpace friends
gender factors

- Female users more likely to be “here for” friendship and male users more likely to be “here for” dating (but only a minority)
- Males and females both preferred to have more female friends and top 8 friends
- Females preferred a greater proportion of female Top 8 friends

(403 data set)
who swears most?

- for US MySpace home pages:
  - male = more likely to contain strong swearing

- for UK MySpace home pages:
  - male = more likely to contain moderate swearing
  - no difference in strong swearing - possibly more strong swearing in female home pages in the younger age groups

- apparent reversal in gendered strong swearing in the UK for young people

July 3, 2006 members, extended collection >>
Percentage of profiles containing swearing

<table>
<thead>
<tr>
<th></th>
<th>moderate</th>
<th>strong</th>
<th>very strong</th>
<th>sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>US males 16-19</strong></td>
<td>10%</td>
<td>47%</td>
<td>2%</td>
<td>1,530</td>
</tr>
<tr>
<td><strong>US females 16-19</strong></td>
<td>11%</td>
<td>38%</td>
<td>2%</td>
<td>1,287</td>
</tr>
<tr>
<td><strong>UK males 16-19</strong></td>
<td>33%</td>
<td>33%</td>
<td>8%</td>
<td>171</td>
</tr>
<tr>
<td><strong>UK females 16-19</strong></td>
<td>18%</td>
<td>38%</td>
<td>3%</td>
<td>130</td>
</tr>
</tbody>
</table>

(Typical sample size 20-148 for non-web swearing research)
emphatic adverb/adjective OR adverbial booster
OR premodifying intensifying negative adjective
(36% of swearing)

- and we r guna go to town again n make a ryt fuckin nyt of it again lol
- see look i'm fucking commenting u back
- lol and stop fucking tickleing me!!
- Thanks for the party last night it was fucking good and you are great hosts.
- That 50's rock and roll weekender was fucking mint!
- Fuckin my space, my arse
- 1/2 d ppl cudnt even speak fuckin english!
- yeah so me and sarah broke up and everythings fucking shit
personal insult referring to defined entity (28% of swearing)

- tehe i am sorry.. i m such a sleep deprived twat alot of the time! lol
- Maxy is the soundest cunt in the world!!!!
- 3rd? i thought i was your main man number one? Fucker
- write bak cunt xxx
- You Godless bastard!
- You evil cunt! Haha
- CHEEKY LITTLE CUNT !
- lucky fuck
think am gonna get him an album or summet fuck nows
got another copy of the reaction CD (will had fucked the last one lol)
qu'est ce que fuck?
what the fuck pubehead whos pete and why is this necicery mate
Heh long story.. cant be fucked to explain :D
geography of MySpace Friends

Tobias Escher’s 2007 presentation...

Tobias is also helping to develop a Perl module `WWW::Myspace` for automatic extraction of information
emotion in Friend comments

- The majority of comments contain positive emotion (including formal expressions, such as “Love, Mike” or “mike x”)
- A minority (20%) contain negative emotion
- & People who give positive emotion tend to get it back

<table>
<thead>
<tr>
<th>Emotion</th>
<th>+ve</th>
<th>-ve</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (none)</td>
<td>34%</td>
<td>80%</td>
</tr>
<tr>
<td>2</td>
<td>28%</td>
<td>6%</td>
</tr>
<tr>
<td>3</td>
<td>35%</td>
<td>11%</td>
</tr>
<tr>
<td>4</td>
<td>3%</td>
<td>2%</td>
</tr>
<tr>
<td>5 (strong)</td>
<td>0%</td>
<td>1%</td>
</tr>
</tbody>
</table>

Emotion strength in 819 random comments
emotion in Friend comments

Positive emotion mainly used by females and mainly directed at females

No gender difference in negative emotions

<table>
<thead>
<tr>
<th></th>
<th>From female</th>
<th>From male</th>
</tr>
</thead>
<tbody>
<tr>
<td>To female</td>
<td>2.4 (+)</td>
<td>2.0 (+)</td>
</tr>
<tr>
<td></td>
<td>1.3 (-)</td>
<td>1.3 (-)</td>
</tr>
<tr>
<td>To male</td>
<td>2.2 (+)</td>
<td>1.7 (+)</td>
</tr>
<tr>
<td></td>
<td>1.3 (-)</td>
<td>1.5 (-)</td>
</tr>
</tbody>
</table>

Average emotion strength in 819 random comments
gendered MySpace comments?

<table>
<thead>
<tr>
<th></th>
<th>Female sender</th>
<th>Male sender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female recipient</td>
<td>i love you!!!!!!</td>
<td>YOU HAVE A PRETTY COOL PAGE....:)</td>
</tr>
<tr>
<td>Male recipient</td>
<td>i can't believe u would say goodbye to the ducky--he's so cute :( lol</td>
<td>you wish lol i'll be there tuesday if you will</td>
</tr>
</tbody>
</table>
conclusions

- There are significant gender differences in social network site communication.
- Language is very informal but this does not mean that it is wrong in context.
- Social network sites are apparently a valuable source of social support.


All available online – Google Thelwall