Social Network Analytics

Recap of Course
Social Media Analytics: Agenda/Learning Objectives

• What’s new here?
• Democratization of celebrity endorsements
• Content & connections
• Discoveries, insights & predictions from the wisdom of the crowd
• Who matters the most in social networks
• Network value of a customer
• Real time visibility into perceptions, events & phenomena
Network Structure Does Matter

• In addition to analyzing attention ....
• Who are most popular?
• Who are “best” connected?
• Who can help spread a message?
• Who can help connect diverse groups?
Insights & Predictions from Social Mentions

Well I'm a Benz fan first and foremost, but the Audi 3.0 CVT just offers more. I think both are excellent (sp) cars, however neither one is a low maintena- it has come to my attention that as of late MB has taken a hit in its reputatio- There is not much to debate on this subject. If you want a performance car give up nationality based generalizations a couple of years ago. I'm interested more towards performance than luxury. Also value is an imp cybers:... I can't fathom why you would want to go with slushbox, but that Sorry to disappoint with the slushbox, but if we were all the same life wou- The new Altima is great as my friends just bought one. However, dollar for c cybers:... Excitement and slushbox don't seem to be compatible. I can't ge “Excitement and slushbox don't seem to be compatible.” Agreed. Unfortuna- Riez, you have heard of performance sensitive steering? This varies the pe cybers:... There are two types. Those tied to performance output and tho- Yeah, I have that brochure and they don't mention the difference between cybers:... From what I've read, the better variable performance assist syst- So, if you read through most of this group, a lot of references went out to Ai. Have you driven them both? The S4 really moves. Even by Audi's own 0-60 I've tested the 3.0, not the S4. Hmmn... new S4 in '08, huh? Maybe I should Yeah, the current S4 has 250 hp and 258 lb-ft of performance. The perfo- Not so fast. There is merit to the "nationality based" debate. Let me explai

Brand & product associations

Predict sales rank, retention, spend, etc.

Find influencers

Metrics, metrics & more metrics

Copyright Anitesh Barua 2016

WHO WILL MAKE BUSINESS HAPPEN? SPARTANS WILL.
New Insights and Discoveries

• Can social media data be used to measure changes in customer perception of specific brand attributes?
• Can social media be used prior to the design stage to create a better campaign?
• New knowledge & insights about
  – Our brand
  – Our products & attributes
  – Competition
• What metrics can we use?
• Where can we apply these insights?
Brand Insights from Netzer et al. (2012)
Text Mining (from Netzer et al. 2012)

- Step 1. Downloading: The Web pages are downloaded from a given forum site in HTML format.
- Step 2. Cleaning: HTML tags and nontextual information such as images and commercials are cleaned from the downloaded files.
- Step 3. Information extraction: Terms for products and product attributes are extracted from the messages.
- Step 4. Chunking: The text is divided into informative units such as threads, messages, and sentences.
- Step 5. Identification of semantic relationships: Two forms of product comparisons are computed: First, we generate a semantic network of co-occurrences of product mentions in the forum. This analysis can provide an overview of the overall market structure. Second, we extract the relationship between products and terms and the nature and sentiment of the relationship.
Message downloaded from Edmunds.com

CarType: 2- Acura TL
MsgNumber: 2479
MsgTitle: r34
MsgAuthor: r34
MsgDate: Jun 24, 2004 (11:38 am)
MsgRepliesTo:
That's strange. I heard many people complaint[sic] about the Honda paint. I owned a 1995 Nissan Altima before and its paint was much better than my neighbour's Accord (1998+ model). I found the Altima interior was quiet [sic] good at that time (not as strange as today's).

Edmunds.com sedan forum

Honda Accords and Toyota Camrys are nice sedans, but hardly the best car on the road (for many people). It's just that they are very competent in their price range. So, a love fest of the best selling may not tell you what is "best".
## Attributes and co-mentions

<table>
<thead>
<tr>
<th></th>
<th>Toyota Camry</th>
<th>Volvo S40</th>
</tr>
</thead>
<tbody>
<tr>
<td># mentions</td>
<td>34,559</td>
<td>580</td>
</tr>
<tr>
<td># co-mentions with safety related words</td>
<td>379</td>
<td>51</td>
</tr>
<tr>
<td>Lift of make/model &amp; safety related words</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total # messages: 868,174 (assume one make/model appears only once in a message)  
Safety related messages: 4534
Co-occurrence/ Co-mentions

• Do terms or words appear together by chance or due to association.
  – E.g., is Volvo more likely to be mentioned in conjunction with safety related words than Toyota?
  – Is Honda more likely to be associated with reliability than, say, Jaguar?
• Lift($A, B$) = $P(A \& B) / (P(A) \times P(B))$
• If $> (\leq 1)$ 1, association is stronger (weaker) than what is expected by chance
Message #1199 Civic vs. Corolla by mcmanus Jul 21, 2007 (4:05 pm). Yes DrFill, the Honda car model is sporty, reliable, and economical vs the Corolla that is just reliable and economical. Ironically its Toyota that is supplying 1.8L turbo ... Neon to his 16 year old brother. I drove it about 130 miles today. Boy does that put all this Civic vs. Corolla back in perspective! The Neon is very crudely designed and built, with no low ...

Product Co-occurrence Data

<table>
<thead>
<tr>
<th></th>
<th>Audi A6</th>
<th>Honda Civic</th>
<th>Toyota Corolla</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audi A6</td>
<td>---</td>
<td>252</td>
<td>101</td>
</tr>
<tr>
<td>Honda Civic</td>
<td>252</td>
<td>---</td>
<td>2762</td>
</tr>
<tr>
<td>Toyota Corolla</td>
<td>101</td>
<td>2762</td>
<td>---</td>
</tr>
</tbody>
</table>

Associative Network

\[ \text{lift}(A, B) = \frac{P(A, B)}{P(A) \times P(B)} \]
Figure 9  Problems Commonly Appearing with the Honda Civic, Nissan Sentra, and Toyota Corolla

- Sludge problem
- Fuel problem
- Starting problem
- Battery problem
- AC problem
- Honda Civic
- Oil consumption problem
- Radio problem
- Toyota Corolla
- Nissan Sentra
- Pump problem
- Engine hesitation problem
- Belt problem
- Power problem
- Suspension problem
- Performance problem
- Engine problem
Network of Car Sedan models

Figure 2: Spring-Embedded (Kamada and Kawai 1989) Network Graph of Sedan Car Models
Edmunds.com brand sentiment

<table>
<thead>
<tr>
<th>MODEL</th>
<th>SENTRA</th>
<th>COROLLA</th>
<th>CIVIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Commonalities</strong></td>
<td><strong>Differentiators</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Economy</td>
<td>Small-car</td>
<td>Subcompact</td>
</tr>
</tbody>
</table>

WHO WILL MAKE BUSINESS HAPPEN? SPARTANS WILL.
Can We Predict Switching Behavior?

Forum co-mentions

Actual brand switching
Data collection

Keywords: Python scripts to collect data from user forums, Twitter etc

Word frequency analysis

Find important words occurring with high frequencies. Also find words/names of companies to merge

Merge words

MDS scaling

Lift analysis

Extract relevant parts for sentiment analysis

Sentiment analysis

Draw insights and create actionable advice

**MDS**: Create dissimilarity matrix from lift matrix by taking dissimilarity = 1/lift

Pay attention to the range of lift values before creating the dissimilarity matrix.
Tools We Use to Collect & Analyze Social Mentions

- Python scripts to crawl sites & access the Twitter API (for example Web Scraper extension for Chrome, http://webscraper.io/)
  - Fetch social mentions
  - Analyze perceptions about brands, products and attributes
  - Rank users by “influence”, provide sentiments, reach, etc.

- Social network analysis tools: E.g., NodeXL, Gephi
- Sentiment analysis: E.g., SentiStrength, Weka, Lingpipe
- Text analytics and machine learning: LightSIDE, Weka, R (tm package)
- Natural Language Processing: Natural Language Toolkit (NLTK)
- Deep Learning: gensim, word2vec
Other network software tools

visualization and analysis:
- **UCInet** - user friendly social network visualization and analysis software (suitable for smaller networks)
- **Pajek** - network analysis and visualization; menu driven, suitable for large networks
  - **iGraph** - if you are familiar with R, you can use iGraph as a module to analyze or create large networks, or you can directly use the C functions
  - **Graph package for Matlab** - if Matlab is the environment you are most comfortable in
  - **SIENA** – for longitudinal analysis
  - **SNA package for R** - all sorts of analysis + heavy duty stats to boot
  - **NetworkX** - python based free package for analysis of large graphs

visualization only:
- **GraphViz** - open source network visualization software (can handle large/specialized networks)
- **TouchGraph** - need to quickly create an interactive visualization for the web

R libraries:
- degreenet, RSeina, PAFit, igraph, sna network, tnet, ergm, Bergm, hergm, latentnet and networksis, statnet

A useful reference: [INSNA long list of SNA packages](#)