Workshop 1A: 
Data Collection & Network Analysis with 
@Netlytic & the iGraph R Package

Anatoliy Gruzd
gruzd@ryerson.ca
@gruzd

Associate Professor, Ted Rogers School of Management
Director, Social Media Lab
Ryerson University

#SMSociety15
Toronto, July 27, 2015
Outline

• Making Sense of Social Media Data
• Practice Part 1: Netlytic
• Practice Part 2: R + igraph
Growth of Social Media and Social Networks Data

Social Media have become an integral part of our daily lives!

Facebook: 1B+ users
Twitter: 500M+ users
How to Make Sense of Social Media Data?
How to Make Sense of Social Media Data?

Social Big Data -> Visualizations -> Understanding
(Development, Application & Validation)
How to Make Sense of Social Media Data?

Social Network Analysis (SNA)

- **Nodes** = People
- **Edges/Ties (lines)** = Relations/“Who retweeted/replied/mentioned whom”
Studying Online Social Networks

- Forum networks
- Blog networks
- Friends’ networks (Facebook, Twitter, Google+, etc…)
- Networks of like-minded people (YouTube, Flickr, etc…)
Advantages of Social Network Analysis

- Reduce the large quantity of data into a more concise representation
- Makes it much easier to understand what is going on in a group

Once the network is discovered, we can find out:
- How do people interact with each other,
- Who are the most/least active members of a group,
- Who is influential in a group,
- Who is susceptible to being influenced, etc…
2012 Olympics in London
#tarsand Twitter Community

Common approach for collecting social network data:

- Self-reported social network data may not be available/accurate
- Surveys or interviews

Problems with surveys or interviews

- Time-consuming
- Questions can be too sensitive
- Answers are subjective or incomplete
- Participant can forget people and interactions
- Different people perceive events and relationships differently
How Do We Collect Information About Social Networks?

• Common approach: surveys or interviews
• A sample question about students’ perceived social structures

Please indicate on a scale from [1] to [5],
YOUR FRIENDSHIP RELATIONSHIP WITH EACH STUDENT IN THE CLASS
[1] - don’t know this person
[2] - just another member of class
[3] - a slight friendship
[4] - a friend
[5] - a close friend

...

Source: C. Haythornthwaite, 1999
How Do We Collect Information About Online Social Networks?

**Goal:** Automated Networks Discovery

**Challenge:** Figuring out what content-based features of online interactions can help to uncover nodes and ties between group members.
Automated Discovery of Social Networks

Emails

• Nodes = People

• Ties = “Who talks to whom”

• Tie strength = The number of messages exchanged between individuals
Automated Discovery of Social Networks

“Many to Many” Communication
## Automated Discovery of Social Networks
### Approach 1: Chain Network (Reply-to)

<table>
<thead>
<tr>
<th>Source</th>
<th>Posting Header</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Method</strong></td>
<td>Connects a sender to the previous poster in the thread</td>
</tr>
<tr>
<td><strong>Discovered Tie(s)</strong></td>
<td><strong>Sam -&gt; Gabriel</strong></td>
</tr>
</tbody>
</table>

**Posting header**

FROM: Sam
PREVIOUS POSTER: Gabriel

...

...

...

Automated Discovery of Social Networks

Approach 1: Chain Network (Reply-to)

<table>
<thead>
<tr>
<th>Source</th>
<th>Posting Header</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>Connects a sender to the previous poster in the thread</td>
</tr>
<tr>
<td>Discovered Tie(s)</td>
<td>Sam -&gt; Gabriel</td>
</tr>
</tbody>
</table>

Posting header

FROM: Sam
PREVIOUS POSTER: Gabriel

Nick, Gina and Gabriel: I apologize for not backing this up with a good source, but I know from reading about this topic that … ”

Possible Missing Connections:

• Sam -> Nick
• Sam -> Gina
• Nick <-> Gina
FROM: Eva
REFERENCE CHAIN: Gabriel, Sam, Gina

“Gina, I owe you a cookie. This is exactly what I wanted to know. I was already planning on taking 402 next semester, and now I have something to look forward to!”

FROM: Fred

“I wonder if that could be why other libraries around the world have resisted changing – it's too much work, and as Dan pointed out, too expensive.”
Automated Discovery of Social Networks
Approach 2: Name Network

This approach looks for personal names in the content of the messages to identify social connections between group members.

FROM: Ann

“Steve and Natasha, I couldn't wait to see your site. I knew it was going to [be] awesome!”

<table>
<thead>
<tr>
<th>Method</th>
<th>Connect the sender to people mentioned in the message</th>
<th>Connect people whose names co-occur in the same message(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovered Tie(s)</td>
<td>Ann -&gt; Steve, Ann -&gt; Natasha</td>
<td>Steve &lt;-&gt; Natasha</td>
</tr>
</tbody>
</table>
Comparing Chain vs Name Networks
Example: Youtube comments

Chain Network (less connections)

Name Network (more connections)
Automated Discovery of Social Networks
Approach 2: Name Network

• Main Communicative Functions of Personal Names (Leech, 1999)
  – getting attention and identifying addressee
  – maintaining and reinforcing social relationships

• Names are “one of the few textual carriers of identity” in discussions on the web (Doherty, 2004)

• Their use is crucial for the creation and maintenance of a sense of community (Ubon, 2005)
Automated Discovery of Social Networks

Name Network Method: Challenges

Kurt Cobain, a lead singer for the rock band Nirvana

chris is not a group member

Santa Monica Public Library

John Dewey, philosopher & educator

mark up language

Solution:
- Name alias resolution
How to Make Sense of Social Media Data?

Example: Twitter Networks

- Nodes = People
- Ties = “Who retweeted/replied/mentioned whom”
- Tie strength = The number of retweets, replies or mentions
Automated Discovery of Social Networks
Twitter Data Example

Report by @JoeProf on how consumers dislike #datamining but feel helpless. HT
@VMosco nytimes.com/2015/06/05/tec…

<table>
<thead>
<tr>
<th>Chain Network ties</th>
<th>Name Network ties</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>@Cheeflo -&gt; @JoeProf</td>
</tr>
<tr>
<td></td>
<td>@Cheeflo -&gt; @VMosco</td>
</tr>
</tbody>
</table>
Automated Discovery of Social Networks
Twitter Data Example

Anatoliy Gruzd
@gruzd

@sidneyeve thank you for helping to spread the word about the position!

<table>
<thead>
<tr>
<th>Chain Network ties</th>
<th>Name Network ties</th>
</tr>
</thead>
<tbody>
<tr>
<td>@gruzd -&gt; @sidneyeve</td>
<td>@gruzd -&gt; @sidneyeve</td>
</tr>
</tbody>
</table>
Comparing Chain vs Name Networks
Example: Twitter data - #SMSociety15 hashtag

Chain Network
10 nodes, 19 ties

Name Network
105 nodes, 152 ties
Netlytic.org
cloud-based research infrastructure for automated text analysis & discovery
of social networks from social big data
Tutorial: Analyzing #SMSociety15 on Twitter

https://netlytic.org/home/?p=10676

Step 1: Connect your Twitter account to Netlytic
# Social Media Research Toolkit

maintained by the Social Media Lab

[http://socialmedialab.ca/?page_id=7801](http://socialmedialab.ca/?page_id=7801)

<table>
<thead>
<tr>
<th>Tool</th>
<th>Platforms</th>
<th>Data collection</th>
<th>Network discovery</th>
<th>Text analysis</th>
<th>Export</th>
<th>Visualization</th>
<th>Published works</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foller.me</td>
<td>🌐</td>
<td>Direct collection from Twitter</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>Journal articles</td>
</tr>
<tr>
<td>NameGenWeb</td>
<td>🌐</td>
<td>Direct collection from Facebook profile</td>
<td>✓</td>
<td>✗</td>
<td>.GraphML</td>
<td>✓</td>
<td>Journal articles</td>
</tr>
<tr>
<td>Netlytic</td>
<td>🌐</td>
<td>RSS, Google Docs, Twitter, YouTube ID, .TXT, .CSV</td>
<td>✓</td>
<td>✓</td>
<td>Pajek, UCINET, .CSV, .PNG</td>
<td>✓</td>
<td>Journal articles</td>
</tr>
<tr>
<td>NodeXL</td>
<td>🌐</td>
<td>GraphML, Pajek, UCINet Social Network Importer</td>
<td>✓</td>
<td>✓</td>
<td>GraphML, Pajek, UCINet, and matrix formats.JPG, .PNG, .EPS, etc.</td>
<td>✓</td>
<td>Books Journal articles</td>
</tr>
<tr>
<td>Panalytics</td>
<td>🌐</td>
<td>Direct connection to Pinterest or Google Analytics</td>
<td>✗</td>
<td>✗</td>
<td>.TSV</td>
<td>✗</td>
<td></td>
</tr>
</tbody>
</table>