FSCI 2017 WT5

Identifying how scientific papers are shared and who is sharing them on Twitter

Stefanie Haustein & Juan Pablo Alperin
Who we are

Stefanie Haustein
Assistant Professor
School of Information Studies
University of Ottawa
stefanie.haustein@uottawa.ca
@stefhaustein

Juan Pablo Alperin
Assistant Professor
Canadian Institute for Studies in Publishing
Simon Fraser University
juan@alperin.ca
@juancommander

Scholarly Communications Lab
scholcommlab.ca

Welcome to the Scholarly Communications Lab

The ScholCommLab is an interdisciplinary team of researchers based in Vancouver and Ottawa, Canada, interested in all aspects of scholarly communication. We explore a wide range of questions using a combination of computational techniques (including applied statistics, machine learning, network analysis, and natural language processing), innovative methods (such as Twitter bot surveys), and traditional qualitative methods (such as interviews, surveys, and focus groups) to investigate how knowledge is produced, disseminated, and used.

We are associated with the Publishing Program and the Public Knowledge Project at Simon Fraser University and with the School of Information Studies at the University of Ottawa.
Day 1 – Wednesday

1. Introduction to altmetrics and Twitter in scholarly communication
2. Collecting Twitter data
3. Generating networks from Twitter data
Outline

Day 2 – Thursday

1. Introduction to social network analysis on Twitter
2. Limitations of data and methods
3. Visualizing and analyzing Twitter networks
4. Interpreting Twitter diffusion networks
5. Possibilities and future directions of altmetrics
1:30-2:20  Introduction to altmetrics and Twitter in scholarly communication
  •   Scholarly communication: From small elites to social media
  •   Research evaluation: From library management to altmetrics

2:20-2:30  Coffee break

2:30-3:20  Collecting Twitter data
  •   Tweets
  •   Twitter APIs
  •   Annotated code to collect Twitter data

3:20-3:30  Coffee break

3:30-4:10  Generating networks from Twitter data
  •   Annotated code to collect follower and friend networks
  •   Creation of edge and node lists
  •   Computation of network indicators
Twitter in Scholarly Communication
Invisible Colleges

Père Marin Mersenne  
(1588-1648)

Henry Oldenburg  
(1619-1677)

http://commons.wikimedia.org/wiki/File:Marin_mersenne.jpg#media/File:Marin_mersenne.jpg
Scientific Societies

L’Académie royale des sciences
22 December 1666

The Royal Society
28 November 1660
Scientific Journals

Le journal des scavans
5 January 1665


Philosophical Transactions
6 March 1665

Scientific Articles

Development from 17\textsuperscript{th} to 20\textsuperscript{th} century

• From natural philosophy to specialized disciplines
• From lengthy adhoc descriptions to IMRaD structure
• From personal to technical and codified writing style
• Development of:
  • Theory
  • Methodology
  • References
  • Gatekeeping
  • Style guides
Digital Revolution

- Improved access
- Acceleration
  - Collaboration
  - Peer review
  - Distribution of preprints

- Decreasing importance of scientific journal
  - Journal functions
  - Diversification of publication venues

Symbolic capital of journals unchanged

Submissions to arXiv

Open Access

Budapest Open Access Initiative

“immediate, free availability on the public internet, permitting any users to read, download, copy, distribute, print, search or link to the full text of these articles”

Budapest Open Access Initiative (2002)

- Gold and Green
- Libre and Gratis
- Hybrid
  - Elsevier: $500 to 5,000
  - Springer: $3,000
  - Wiley: $3,000

Freely available journal papers 2004 to 2011
Open Science

“opening up the research process by making all of its outcomes, and the way in which these outcomes were achieved, publicly available on the World Wide Web”

Kraker et al. (2011, p. 645)

- Open Data
- Open Source
- Open Methodology
- Open Access
- Open Peer Review

Social Media in Academia

- Increased use of social media
  - Informational use
  - Social use
- Uptake and use differs between:
  - Platforms
  - Disciplines
  - Individuals
- Concerns and risks

## Journals cited in the *Journal of the American Chemical Society* 1926

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ber.</em></td>
<td>78</td>
<td>30</td>
<td>67</td>
<td>115</td>
<td>79</td>
<td>64</td>
<td>60</td>
<td>56</td>
<td>53</td>
<td>44</td>
<td>33</td>
<td></td>
<td>686</td>
</tr>
<tr>
<td><em>J. Chem. Soc.</em></td>
<td>122</td>
<td>37</td>
<td>60</td>
<td>47</td>
<td>21</td>
<td>20</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>390</td>
</tr>
<tr>
<td><em>Ann.</em></td>
<td>278</td>
<td>26</td>
<td>8</td>
<td>37</td>
<td>33</td>
<td>23</td>
<td>22</td>
<td>21</td>
<td>19</td>
<td>18</td>
<td>13</td>
<td></td>
<td>191</td>
</tr>
<tr>
<td><em>Z. physik. Chem.</em></td>
<td>64</td>
<td>21</td>
<td>29</td>
<td>19</td>
<td>28</td>
<td>16</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>126</td>
</tr>
<tr>
<td><em>Compt. rend.</em></td>
<td>42</td>
<td>13</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>93</td>
</tr>
<tr>
<td><em>J. Phys. Chem.</em></td>
<td>18</td>
<td>4</td>
<td>13</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>93</td>
</tr>
<tr>
<td><em>Ann. Physik.</em></td>
<td>41</td>
<td>16</td>
<td>14</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
</tr>
<tr>
<td><em>J. Biol. Chem.</em></td>
<td>41</td>
<td>16</td>
<td>14</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
</tr>
<tr>
<td><em>Am. Chem. J.</em></td>
<td>21</td>
<td>11</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>6</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
</tr>
<tr>
<td><em>Z. anorg. Chem.</em></td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td><em>Ann. Chem.</em></td>
<td>5</td>
<td>0</td>
<td>6</td>
<td>9</td>
<td>7</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td><em>Bull. Soc. Chim.</em></td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td><em>Proc. Roy. Soc.</em></td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td><em>J. Ind. Eng. Chem.</em></td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td><em>Z. Phys.</em></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><em>Ann. Chem.</em></td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td><em>Phil. Mag.</em></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><em>Gez. chem. stoil.</em></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><em>Phys. Rev.</em></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><em>Physik. Zeit.</em></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><em>Z. Elektrochem.</em></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><em>Biochem. Z.</em></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><em>Rec. trav. chim.</em></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><em>Science</em></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><em>Trans. Far. Soc.</em></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><em>Proc. Nat'l Acad.</em></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><em>Nature</em></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

The abbreviations used above and in the tables to follow are those accepted by *Chemical Abstracts* and will be found in their list of periodicals abstracted, issued October 20, 1926.

Research Evaluation

Information Retrieval

“It would not be excessive to demand that the thorough scholar check all papers that have cited or criticized such papers, if they could be located quickly. The citation index makes this check practicable.”

Garfield (1955, p. 108)

- Institute for Scientific Information
- Science Citation Index
- Source Author Index
- Citation Index

Research Evaluation

Citation Impact

• Part of hiring, promotion and funding decisions
• Dashboard tools
Research Evaluation

Altmetrics

• Information overload
  “We rely on filters to make sense of the scholarly literature, but the narrow, traditional filters are being swamped. However, the growth of new, online scholarly tools allows us to make new filters; these altmetrics reflect the broad, rapid impact of scholarship in this burgeoning ecosystem.”

  Priem et al. (2010)

• Criticism against current form of research evaluation
  • Alternative forms of research output
  • Alternative use and visibility of research output

Altmetrics

“study and use of scholarly impact measures based on activity in online tools and environments”

Priem (2014, p. 266)

“a good idea but a bad name”

Rousseau & Ye (2013, p. 3289)


Scholarly Metrics

adapted from: Björneborn & Ingwersen (2004, p. 1217)

“[S]cholarly metrics are thus defined as indicators based on recorded events of acts [...] related to scholarly documents [...] or scholarly agents [...].”

Haustein (2016, p. 348)
Scholarly Metrics

Acts related to research objects

viewing metadata
accessing content
storing research object

on a microblog platform
in a social network
in a comment
on a Q&A site
on a listserv
on a rating or voting platform
in podcasts and videos
in a presentation
in a blog post
in a Wikipedia article
in mainstream media and news
in a scientific document
in a policy document

theories, methods, or results
software code
datasets

collaboration

viewing homepage
emailing
downloading CV

on a microblog platform
in a social network
in a comment
on a Q&A site
on a listserv
on a rating or voting platform
in podcasts and videos
in a presentation
in a blog post
in a Wikipedia article
in mainstream media and news
in a scientific document
in a policy document

**Altmetrics**

**Coverage per platform**

<table>
<thead>
<tr>
<th>Platform</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>WoS 2012</td>
<td>N=1 339 279 (100%)</td>
</tr>
<tr>
<td>Mendeley</td>
<td>84,2 %</td>
</tr>
<tr>
<td>Twitter</td>
<td>21,5 %</td>
</tr>
<tr>
<td>Facebook</td>
<td>4,7 %</td>
</tr>
<tr>
<td>Blogs</td>
<td>1,9 %</td>
</tr>
<tr>
<td>Google+</td>
<td>0,8 %</td>
</tr>
<tr>
<td>News</td>
<td>0,7 %</td>
</tr>
</tbody>
</table>


Altmetrics

Coverage per discipline

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Mendeley Coverage</th>
<th>Twitter Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics &amp; Computer Science</td>
<td>76.4 %</td>
<td>7.5 %</td>
</tr>
<tr>
<td>Natural Sciences &amp; Engineering</td>
<td>83.7 %</td>
<td>12.9 %</td>
</tr>
<tr>
<td>Life &amp; Earth Sciences</td>
<td>91.4 %</td>
<td>21.6 %</td>
</tr>
<tr>
<td>Biomedical &amp; Health Sciences</td>
<td>86.5 %</td>
<td>31.7 %</td>
</tr>
<tr>
<td>Social Sciences &amp; Humanities</td>
<td>81.7 %</td>
<td>26.0 %</td>
</tr>
</tbody>
</table>


Altmetrics

Spearman correlations with citations

<table>
<thead>
<tr>
<th>Platform</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mendeley</td>
<td>0.386**</td>
</tr>
<tr>
<td>Twitter</td>
<td>0.194**</td>
</tr>
<tr>
<td>Facebook</td>
<td>0.097**</td>
</tr>
<tr>
<td>Blogs</td>
<td>0.124**</td>
</tr>
<tr>
<td>Google+</td>
<td>0.065**</td>
</tr>
<tr>
<td>News</td>
<td>0.083**</td>
</tr>
</tbody>
</table>


Scientific Papers on Twitter

Engagement with scientific papers on Twitter

- **A unification of RDE model and XCDM model**
  arxiv.org/abs/1212.5790

- **On the calculation of percentile-based bibliometric indicators**
  ow.ly/h8Kzp

- **Hysteretic response characteristics and dynamic phase transition via site dilution in the kinetic**

- **Burkert & Hartmann on star formation thresholds, should be good.**
  arxiv.org/abs/1212.4543 with nod to newly be-doctored @alunacentroid too!

- **Richard Ellis about possible detection of z=11.9 galaxy in Hubble data:**
  "While definitively real, we remain cautious of it's nature" #AAS221

- **Jan Hattenbach**
  @JanHattenbach
  Details
  Reply "La Herradura" Favorites More

Scientific Papers on Twitter

Journal accounts

[JASIST] The official scholarly journal of the Association for Information Science and Technology (@asist_org), published by Wiley.

[JASIST] In the 2/2017 @JASIST: Generating new indicators on universities by linking data in open platforms

Beyond university rankings? Generating new indi...
Scientific Papers on Twitter

Who is tweeting scientific papers?

• Bots?
  • No use, no impact?
  • Information diffusion?

• Researchers/institutions?
  • Scientific impact?
  • Scientific use?
  • Information diffusion?

• Students/teachers?
  • Educational impact?
  • Educational use?

• Authors/institutions?
  • Information diffusion?
  • Self promotion?

• Publishers/editors?
  • Information diffusion?
  • Self promotion?

• The “general public”?
  • Public engagement?
  • Societal impact?
Scientific Papers on Twitter

Who is tweeting scientific papers?

- Analyzing Twitter bios
- Coding Twitter users
- Twitter bot surveys
- Analysis of Twitter diffusion networks
Scientific Papers on Twitter

Analyzing Twitter bios
Altmetric.com classification

Demographic breakdown

<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
<th>As %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Members of the public</td>
<td>141</td>
<td>71%</td>
</tr>
<tr>
<td>Practitioners (doctors, other healthcare pros)</td>
<td>26</td>
<td>13%</td>
</tr>
<tr>
<td>Scientists</td>
<td>21</td>
<td>11%</td>
</tr>
<tr>
<td>Science communicators (journalists, bloggers, editors)</td>
<td>11</td>
<td>6%</td>
</tr>
</tbody>
</table>

*based on Altmetric.com data 06/2015

https://www.altmetric.com/details/2088143
Scientific Papers on Twitter

Analyzing Twitter bios

Node size
number of accounts associated with term

Node color
cluster affiliation

1. personal

2. topics and collectives

3. academic

Coding Twitter users

- Among a sample of 2,000 accounts tweeting papers, 34% of individuals identified as having PhD
  
  (Tsou, Bowman, Ghazinejad, & Sugimoto, 2010)

- Random sample of 800 of 89,768 English accounts tweeting 2012 WoS articles:

Is the Twitter handle maintained by an:

- Unable to tell: 93 (12%)
- Organization: 165 (21%)
- Individual: 542 (68%)

Does the Twitter account appear:

- Unable to tell: 42%
- Completely automated: 8%
- Partially automated: 5%
- Not automated: 45%

n=800

Scientific Papers on Twitter

Does the individual indicate that they are:

- **Student**:
  - 7%
  - PhD student
  - PhD candidate
  - Med student
  - Grad student
  ...  

- **Researcher**:
  - 22%
  - Scientist
  - Researcher
  - Professor
  - Postdoc
  ...  

- **Professional**:
  - 47%
  - Coach
  - Doctor
  - MD
  - Photographer
  ...  

- **Science communicator**:
  - 13%
  - Writer
  - Author
  - Journalist
  - Blogger
  ...  

- **Unable to tell**:
  - 29%
  - Wife
  - Likes [...]
  - Runner
  - Conservative
  ...  

n=542
## Scientific Papers on Twitter

### Twitter bot surveys

- Of 286 users linking to SciELO articles, 24% employed at university, 23% students *(Alperin, 2015)*

- Optimization of response rates: *(Alperin et al., 2017)*

<table>
<thead>
<tr>
<th>#</th>
<th>Question types</th>
<th>N</th>
<th>Response rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Open</td>
<td>119</td>
<td>16.8 %</td>
</tr>
<tr>
<td>2</td>
<td>No egoistic appeal</td>
<td>112</td>
<td>12.5 %</td>
</tr>
<tr>
<td>3</td>
<td>Yes/No</td>
<td>116</td>
<td>17.2 %</td>
</tr>
<tr>
<td>4</td>
<td>No context</td>
<td>111</td>
<td>9.9 %</td>
</tr>
<tr>
<td>5</td>
<td>Multiple choice</td>
<td>106</td>
<td>22.6 %</td>
</tr>
<tr>
<td>6</td>
<td>No context</td>
<td>112</td>
<td>21.4 %</td>
</tr>
<tr>
<td>7</td>
<td>Open</td>
<td>110</td>
<td>28.2 %</td>
</tr>
<tr>
<td>8</td>
<td>No context</td>
<td>109</td>
<td>22.0 %</td>
</tr>
<tr>
<td>9</td>
<td>Yes/No</td>
<td>114</td>
<td>29.0 %</td>
</tr>
<tr>
<td>10</td>
<td>No context</td>
<td>105</td>
<td>24.8 %</td>
</tr>
<tr>
<td>11</td>
<td>Multiple choice</td>
<td>110</td>
<td>40.0 %</td>
</tr>
<tr>
<td>12</td>
<td>No context</td>
<td>107</td>
<td>35.5 %</td>
</tr>
</tbody>
</table>

1,331 (Alperin et al., 2017)
Scientific Papers on Twitter

Analysis of Twitter networks

• Social network analysis (SNA) of Twitter follower networks
• Distinguishing diffusion patterns
• Combination of SNA with NLP and Twitter bot surveys

➤ Differentiation of diffusion, use and impact
  • Scientific
  • Educational
  • Societal
Twitter Data Collection
Twitter Data Collection

Twitter APIs

• Streaming API
  • Push-based global stream of live Twitter data
  • Bandwidths:
    • 100% (“firehose”) – paid access
    • 10% sample (“gardenhose”) – occasionally granted
    • 1% sample (“spritzer”) – free for all users

• Representational State Transfer (REST) API
  • Pull-based access to Twitter data
  • Restricted through rate limits

## Twitter Data Collection

### REST API rate limits

*Per user (access token) and 15-minute window*

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Resource family</th>
<th>Requests / window (user auth)</th>
<th>Requests / window (app auth)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GET account/verify_credentials</td>
<td>application</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>GET application/rate_limit_status</td>
<td>application</td>
<td>180</td>
<td>180</td>
</tr>
<tr>
<td>GET favorites/list</td>
<td>favorites</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>GET followers/ids</td>
<td>followers</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>GET followers/list</td>
<td>followers</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>GET friends/ids</td>
<td>friends</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>GET friends/list</td>
<td>friends</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>GET friendships/show</td>
<td>friendships</td>
<td>180</td>
<td>15</td>
</tr>
<tr>
<td>GET geo/id/:place_id</td>
<td>geo</td>
<td>75</td>
<td>0</td>
</tr>
</tbody>
</table>

[https://dev.twitter.com/rest/public/rate-limits](https://dev.twitter.com/rest/public/rate-limits)
Twitter Data Collection

REST API endpoints

• POST
• DELETE
• GET
  • users
    api.get_user(screen_name)
  • statuses
    api.get_status(id)
  • followers
    api.followers(user.id)
  • friends
    api.friends(user.id)

https://dev.twitter.com/rest/reference  https://dev.twitter.com/overview/api/tweets
Twitter Data Collection

Jupyter Notebook

- Accessing Twitter API
- Collect tweet and tweet metadata using Tweepy Python library
- Collect user information

https://twitter-research.scholcommlab.ca:8888
# Twitter Data Collection

## Jupyter Notebook

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>Name</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>#</td>
</tr>
<tr>
<td>1</td>
<td>Haustein, Stefanie</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Troia, Lily Cristina</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>Agate, Nicky</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>Peter, Katharin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>Grishagin, Ivan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>Elfenbein, Timothy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>Kunze, John</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
</tbody>
</table>
Twitter Networks
Twitter networks

Twitter entities

• Tweet
  • Type of tweet
    • Original tweet
    • Retweet
    • @reply
    • Private message
  • Tweet content
    • Hashtags
    • URL
    • Media
  • Tweet metadata
    • Timestamp
    • Location

• User
  • Follower/friend relationship
  • User information
Twitter Networks

Users

Twitter follower/friends information

Jonathan D G Jones
@jonathandgjones

I study plant disease resistance to solve crop disease 34 happy years making GM plants. Pic with wonderful multicultural team @ Xmas meal. Incorrigible Brexmooner

© Sainsbury Lab, Norwich, UK
tsl.ac.uk/groups/jones-g...
Joined May 2010

Followers 3,195

Following 444
Twitter networks

Twitter follower network

Twitter user

Follower/friend relationship
Twitter networks

Twitter follower network

Twitter user
- Screen name
- Sign-up date
- Twitter bio
- Location
- Number of followers
- Number of friends
- ...

Follower/friend relationship
- A is friend of B / B is follower of A
- A is follower of B / B is friend of A
Twitter Networks

Jupyter Notebook

• Collect follower information
• Create Twitter follower/friend network
• Visualize network
• Compute network indicators

https://twitter-research.scholcommlab.ca:8888
Twitter Networks

Jupyter Notebook

To set your working directory in cell m), please use the following participant numbers:

```python
# m) set your working directory cell

# Put your number here. This will avoid you overwriting each other's work.
student_number = 1

datadir = 'data/%s' % student_number
```

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Haustein, Stefanie</td>
</tr>
<tr>
<td>2</td>
<td>Troia, Lily Cristina</td>
</tr>
<tr>
<td>3</td>
<td>Agate, Nicky</td>
</tr>
<tr>
<td>4</td>
<td>Peter, Katharin</td>
</tr>
<tr>
<td>5</td>
<td>Grishagin, Ivan</td>
</tr>
<tr>
<td>6</td>
<td>Elfenbein, Timothy</td>
</tr>
<tr>
<td>7</td>
<td>Kunze, John</td>
</tr>
<tr>
<td>8</td>
<td>Berry, Alyssa</td>
</tr>
<tr>
<td>9</td>
<td>Neylon, Cameron</td>
</tr>
<tr>
<td>10</td>
<td>Chuen, Lorraine</td>
</tr>
<tr>
<td>11</td>
<td>Suiter, Amy</td>
</tr>
<tr>
<td>12</td>
<td>Ricaurte, Paola</td>
</tr>
<tr>
<td>13</td>
<td>Participant 13</td>
</tr>
<tr>
<td>14</td>
<td>Participant 14</td>
</tr>
</tbody>
</table>

Jupyter Notebook Cell m)
Twitter Networks

Wednesday

density: 0.19
num_nodes: 13
num_edges: 30
diameter: 4
in_degree_mean: 2.31
out_degree_mean: 2.31
degree_mean: 4.62
biggest_wcc_num_nodes: 9
biggest_wcc_num_nodes_p: 69.23
biggest_wcc_density: 0.42
biggest_wcc_infomap_modularity: 0.00

Thursday

density: 0.28
num_nodes: 13
num_edges: 43
diameter: 4
in_degree_mean: 3.31
out_degree_mean: 3.31
degree_mean: 6.62
biggest_wcc_num_nodes: 11
biggest_wcc_num_nodes_p: 84.62
biggest_wcc_density: 0.39
biggest_wcc_infomap_modularity: 0.00
FSCI 2017 WT5

Identifying how scientific papers are shared and who is sharing them on Twitter

Stefanie Haustein & Juan Pablo Alperin
WT5 Day 2

1:30-2:20  Introduction to social network analysis
   - Theoretical framework, basic concepts and development of social network analysis
   - Network representations, indicators and visualizations

2:20-2:30  Limitations of data and methods

2:30-2:40  *Coffee break*

2:40-4:15  Applied social network analysis: Scientific papers on Twitter
   - Visualizing Twitter networks with Gephi
   - Computing social network indicators with Gephi
   - Interpreting Twitter diffusion networks

4:15-4:30  Possibilities and future directions of altmetrics
Introduction to Social Network Analysis
Introduction to SNA

What is Social Network Analysis (SNA)?

• Analysis of connections between entities
  • Identifying structures of groups and clusters
  • Identifying positions of entities

• Entities/actors = nodes
• Connections/relationships = edges or arcs

Introduction to SNA

Friendship choices among fourth graders

Introduction to SNA

Women’s attendance at social events

Introduction to SNA

UCSD Map of Science

Introduction to SNA

Networks in bibliometrics

Citation network of publications

Co-authorship network of authors or organizations

Co-citation network of publications, authors or journals

Bibliographic coupling network of publications, authors or journals

Co-occurrence network of terms

Introduction to SNA

Facebook friendship networks
Introduction to SNA

Twitter follower network

http://allthingsgraphed.com/2014/11/02/twitter-friends-network/
Introduction to SNA

Why SNA?

• Efficient information transfer
  • Preattentive perception:
    • Size
    • Colors
    • Form
    • Spatial position

https://medium.com/@jonmyers/design-seeing-without-thinking-783d018bb82f#.emeowtdei
Introduction to SNA

Preattentive processing

Color:

https://www.csc2.ncsu.edu/faculty/healey/PP/

https://www.youtube.com/watch?v=wnvoZxe95bo
Introduction to SNA

Preattentive processing

Shape:

https://www.csc2.ncsu.edu/faculty/healey/PP/

https://www.youtube.com/watch?v=wnvoZxe95bo
Introduction to SNA

Preattentive processing absent

Color and shape:

https://www.csc2.ncsu.edu/faculty/healey/PP/

https://www.youtube.com/watch?v=wnvoZxe95bo
Different representations of social networks

• Network graphs (sociograms)
  “The proper placement of every individual and of all interrelations of individuals can be shown on a sociogram. It is at present the only available scheme which makes structural analysis of a community possible.”
  Moreno, 1953, p. 96

• Matrices
• Edge lists

Introduction to SNA

Different representations: Graph
### Introduction to SNA

**Different representations: Matrix**

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>A8</th>
<th>A9</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>A6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>A9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Introduction to SNA

Different representations: Edge list

A1 A3
A3 A2
A3 A4
A2 A4
A4 A5
A5 A6
A5 A9
A6 A7
A7 A8
A7 A9
A8 A9
Different types of networks

Introduction to SNA

Complete network

Ego network

1-mode

2-mode
Introduction to SNA

Network elements

Dyad

*Subgraph containing two nodes*

Triad

*Subgraph containing three nodes*

Isolate

*Single node without ties*
Introduction to SNA

Network elements

Component

*Subgraph in which every node can reach every other*

Cluster

*Subgraph of densely connected nodes*

Clique

*Subgraph in which all nodes are connected*
Introduction to SNA

Network elements

Component
Subgraph in which every node can reach every other

Cluster
Subgraph of densely connected nodes

Clique
Subgraph in which all nodes are connected
Introduction to SNA

Network elements

Component
Subgraph in which every node can reach every other

Cluster
Subgraph of densely connected nodes

Clique
Subgraph in which all nodes are connected
Introduction to SNA

Network elements

Component

*Subgraph in which every node can reach every other*

Cluster

*Subgraph of densely connected nodes*

Clique

*Subgraph in which all nodes are connected*
Introduction to SNA

Network elements

Component
Subgraph in which every node can reach every other

Cluster
Subgraph of densely connected nodes

Clique
Subgraph in which all nodes are connected
Introduction to SNA

Analyzing and comparing networks

- **Visual interpretation of network structures**
  - Core vs. periphery
  - Components, clusters and cliques
- **Network indicators**
  - Node properties
  - Edge properties
  - Network properties
Analyzing and comparing networks

- **Node properties**
  - Central, peripheral nodes or brokers

- **Edge properties**
  - Directed or undirected
  - Valued or binary
  - Reciprocal or unidirectional
  - Multiplex or uniplex

- **Network properties**
  - Densely or sparsely connected
  - Modularity
SNA Indicators: Nodes

Degree centrality (undirected networks)

Number of direct links

Central nodes have many direct links

\[ C_D(n_i) = d(n_i) \]

Example:
\[ C_D(A7) = 3 \]

d = number of direct links
SNA Indicators: Nodes

In-degree (directed networks)

*Number of direct incoming arcs*

- A node with many incoming links is an authority

Out-degree (directed networks)

*Number of direct outgoing arcs*

- A node with many outgoing arcs is a hub
SNA Indicators: Nodes

Closeness centrality

*Sum of the reciprocal of the shortest paths to all nodes*

- Central nodes are close to all other nodes in the network

\[
C_C(n_i) = \sum_{\substack{j=1 \\ i \neq j}}^{n} \frac{1}{d(n_i, n_j)}
\]

\[d(n_i, n_j) = \text{shortest geodesic distance between nodes } n_i \text{ and } n_j\]
SNA Indicators: Nodes

Shortest path \((\text{geodesic distance})\)

*Shortest path length between two nodes*

Path length: number of edges between two nodes
SNA Indicators: Nodes

Closeness centrality

\[ C_C(n_i) = \sum_{j=1}^{n} \frac{1}{d(n_i, n_j)} \quad \text{for} \quad i \neq j \]

\[ C_C(A5) = \frac{1}{3} + \frac{1}{2} + \frac{1}{2} + \frac{1}{1} + \frac{1}{1} + \frac{1}{2} + \frac{1}{2} + \frac{1}{1} \]

\[ = 5.33 \]

Shortest paths from A5 to A1, A2, A3, A4, A6, A7, A8 and A9
SNA Indicators: Nodes

Betweenness centrality

*Sum of the ratio of shortest paths between two nodes through a third node and all shortest paths between two nodes*

Central nodes are positioned on the shortest paths of many nodes of the network

\[ C_B(n_i) = \sum_{j,k \neq i} \frac{g_{jik}}{g_{jk}} \]

\( g_{jik} = \) shortest paths between \( j \) and \( k \) through \( i \)

\( g_{jk} = \) shortest paths between \( j \) and \( k \)
SNA Indicators: Nodes

Betweenness centrality

$$C_B(n_i) = \sum \frac{g_{jik}}{g_{jk}}$$

$$C_B(A5) = \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} = 16.5$$

Shortest paths from A6 to A9
SNA Indicators: Nodes

Eigenvector centrality

Iterative multiplication of adjacency matrix by eigenvector

Central nodes are linked to by many other central nodes in the network

\[
\begin{pmatrix}
0 & 1 & 1 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 \\
\end{pmatrix}
\begin{pmatrix}
3 \\
2 \\
3 \\
3 \\
1 \\
\end{pmatrix}
\]
SNA Indicators: Nodes

Eigenvector centrality

Iterative multiplication of adjacency matrix by eigenvector

Central nodes are linked to by many other central nodes in the network

\[
\begin{pmatrix}
0 & 1 & 1 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
3 \\
2 \\
3 \\
3 \\
1
\end{pmatrix}
= 
\begin{pmatrix}
0 \times 3 + 1 \times 2 + 1 \times 3 + 1 \times 3 + 0 \times 1 \\
1 \times 3 + 0 \times 2 + 1 \times 3 + 0 \times 3 + 0 \times 1 \\
1 \times 3 + 1 \times 2 + 0 \times 3 + 1 \times 3 + 0 \times 1 \\
1 \times 3 + 0 \times 2 + 1 \times 3 + 0 \times 3 + 1 \times 1 \\
0 \times 3 + 0 \times 2 + 0 \times 3 + 1 \times 3 + 0 \times 1
\end{pmatrix}
= 
\begin{pmatrix}
8 \\
6 \\
8 \\
7 \\
3
\end{pmatrix}
\]
Eigenvector centrality

*Iterative multiplication of adjacency matrix by eigenvector*

- Central nodes are linked to by many other central nodes in the network

\[
\begin{pmatrix}
0 & 1 & 1 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
8 \\
6 \\
8 \\
7 \\
3
\end{pmatrix}
\]
SNA Indicators: Nodes

Eigenvector centrality

Iterative multiplication of adjacency matrix by eigenvector

Central nodes are linked to by many other central nodes in the network

\[
\begin{pmatrix}
0 & 1 & 1 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 \\
\end{pmatrix}
\begin{pmatrix}
8 \\
6 \\
8 \\
7 \\
3 \\
\end{pmatrix}
=
\begin{pmatrix}
21 \\
16 \\
21 \\
19 \\
7 \\
\end{pmatrix}
\]
### SNA Indicators: Nodes

**Node centrality**

<table>
<thead>
<tr>
<th>Node</th>
<th>Degree</th>
<th>Closeness</th>
<th>Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>1</td>
<td>3.2</td>
<td>0.0</td>
</tr>
<tr>
<td>A2</td>
<td>2</td>
<td>4.2</td>
<td>0.0</td>
</tr>
<tr>
<td>A3</td>
<td>2</td>
<td>4.7</td>
<td>7.0</td>
</tr>
<tr>
<td>A4</td>
<td>3</td>
<td>5.1</td>
<td>15.0</td>
</tr>
<tr>
<td>A5</td>
<td>3</td>
<td>5.3</td>
<td>16.5</td>
</tr>
<tr>
<td>A6</td>
<td>2</td>
<td>4.4</td>
<td>2.5</td>
</tr>
<tr>
<td>A7</td>
<td>3</td>
<td>4.5</td>
<td>1.5</td>
</tr>
<tr>
<td>A8</td>
<td>2</td>
<td>4.0</td>
<td>0.0</td>
</tr>
<tr>
<td>A9</td>
<td>3</td>
<td>4.9</td>
<td>7.5</td>
</tr>
</tbody>
</table>
SNA Indicators: Nodes

Node centrality

<table>
<thead>
<tr>
<th></th>
<th>Degree</th>
<th>Closeness</th>
<th>Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>1</td>
<td>3.2</td>
<td>0.0</td>
</tr>
<tr>
<td>A2</td>
<td>2</td>
<td>4.2</td>
<td>0.0</td>
</tr>
<tr>
<td>A3</td>
<td>2</td>
<td>4.7</td>
<td>7.0</td>
</tr>
<tr>
<td>A4</td>
<td>3</td>
<td>5.1</td>
<td>15.0</td>
</tr>
<tr>
<td>A5</td>
<td>3</td>
<td>5.3</td>
<td>16.5</td>
</tr>
<tr>
<td>A6</td>
<td>2</td>
<td>4.4</td>
<td>2.5</td>
</tr>
<tr>
<td>A7</td>
<td>3</td>
<td>4.5</td>
<td>1.5</td>
</tr>
<tr>
<td>A8</td>
<td>2</td>
<td>4.0</td>
<td>0.0</td>
</tr>
<tr>
<td>A9</td>
<td>3</td>
<td>4.9</td>
<td>7.5</td>
</tr>
</tbody>
</table>
SNA Indicators: Networks

Number of nodes

Average distance

*Mean of shortest paths between all nodes*

Diameter

*Longest shortest path in the network*
Density (undirected network)

Actual edges divided by potential edges

- In a dense network, most nodes are connected

\[
D_{\text{undirected}} = \frac{2l}{n(n-1)}
\]

- \( l \) = number of actual edges
- \( n \) = number of nodes
SNA Indicator: Networks

Density (undirected network)

Actual edges: 11

Potential edges: 36

Number of potential edges in undirected networks: \( n(n-1)/2 \)
Density (undirected network)

*Actual edges divided by potential edges*

\[
D_{\text{undirected}} = \frac{2l}{n(n - 1)}
\]

\[
= \frac{2 \times 9}{9 \times (9 - 1)}
\]

\[
= \frac{18}{72}
\]

\[
= 0.25
\]
SNA Indicators: Networks

Density (directed network)

*Actual arcs divided by potential edges*

- In a dense network, most nodes are connected

\[
D_{\text{directed}} = \frac{l}{n(n-1)}
\]

- \( l = \) number of actual arcs
- \( n = \) number of nodes

Number of potential edges in directed networks: \( n(n-1) \)
Clustering and community detection

Partitioning a network into communities based on their relative density

- Highly modular networks are made up of many small, densely connected subgroups

Louvain modularity

Limitations of Data and Methods
Limitations

• Missing tweets without link to paper
• Replies
• Informal citations
• Second-order events
• Follower network based on time of data collection
• Diffusion pattern limited to Twitter network
Applied SNA: Scientific Papers on Twitter
Applied SNA

Research questions

• How are scientific papers diffused on Twitter?
• To what extent do communities form around scientific articles?

➢ Who discusses scientific papers on Twitter?
➢ What does Twitter visibility of scientific papers primarily reflect?

Differentiating between:
• Scholarly diffusion/use/impact
• Societal use/impact
"The more #sugar you eat, the more [your brain] wants you to eat" - Lewis Cantley (@WeillCornell) talks to @BMCBiology buff.ly/1fht7EK
10:10 AM - 31 Jan 2014

Fascinating interview with Lou Cantley about why fructose is worse for you than glucose; no more corn syrup for me biomedcentral.com/1741-7007/12/8
4:08 AM - 7 Feb 2014

Fructose, obesity and cancer. Phenomenal interview with Lew Cantley. biomedcentral.com/1741-7007/12/8
11:32 PM - 5 Feb 2014

https://www.altmetric.com/details/2088143
they should tell them to cut back on sugar

bmcbiol.biomedcentral.com/articles/10.11 ...
Cancer, metabolism, fructose, artificial sweeteners, and going cold turkey on sugar
Overview of attention for article published in BMC Biology, January 2014

So far, Altmetric has seen 222 tweets from 199 users, with an upper bound of 887,097 followers.

Showing items 1–100

Jonathan D G Jones
@jonashandgjones
they should tell them to cut back on sugar https://t.co/rDZikK67Er https://t.co/zjA5rSTKE9
28 May 2016

Philip Hopkins
@hopbios
Cancer, metabolism, fructose, artificial sweeteners, and going cold turkey on

YUUGA Kemistri
@YUUGAKemistri
RT @sarahwilson : Fructose doesn’t supply any energy. Nor to muscle; it only gets stored as fat. http://t.co/Vv
07 Mar 2016

Jan Vyjdak
@janovyjdak

Geographical breakdown

<table>
<thead>
<tr>
<th>Country</th>
<th>Count</th>
<th>As %</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>35</td>
<td>18%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>32</td>
<td>16%</td>
</tr>
<tr>
<td>Australia</td>
<td>16</td>
<td>8%</td>
</tr>
<tr>
<td>Canada</td>
<td>16</td>
<td>8%</td>
</tr>
<tr>
<td>New Zealand</td>
<td>8</td>
<td>4%</td>
</tr>
<tr>
<td>Spain</td>
<td>4</td>
<td>2%</td>
</tr>
</tbody>
</table>

Demographic breakdown

<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
<th>As %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Members of the public</td>
<td>143</td>
<td>71%</td>
</tr>
<tr>
<td>Practitioners (doctors, other healthcare professionals)</td>
<td>26</td>
<td>13%</td>
</tr>
<tr>
<td>Scientists</td>
<td>21</td>
<td>11%</td>
</tr>
<tr>
<td>Science communicators (journalists, bloggers, editors)</td>
<td>11</td>
<td>6%</td>
</tr>
</tbody>
</table>
Applied SNA

Dataset

Scientific journal articles mentioned on Twitter
Applied SNA

Dataset

Case study: *BMC Biology* and *BMC Evolutionary Biology*

- Identification of highly 2014 articles via Altmetric.com

≥50 tweets

- **BMC Biology**
  - 108 tweeted articles

- **BMC Evolutionary Biology**
  - 228 tweeted articles
# Applied SNA

## Dataset

### Case study: *BMC Biology* and *BMC Evolutionary Biology*

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>Tweets</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biol1</td>
<td>Cancer, metabolism, fructose, artificial sweeteners, and going cold turkey on sugar</td>
<td>223</td>
<td>203</td>
</tr>
<tr>
<td>Biol2</td>
<td>Models in biology: accurate descriptions of our pathetic thinking</td>
<td>199</td>
<td>157</td>
</tr>
<tr>
<td>Biol3</td>
<td>Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity</td>
<td>54</td>
<td>44</td>
</tr>
<tr>
<td>Biol4</td>
<td>Reported Drosophila courtship song rhythms are artifacts of data analysis</td>
<td>55</td>
<td>50</td>
</tr>
<tr>
<td>Biol5</td>
<td>The Earth Microbiome project: successes and aspirations</td>
<td>88</td>
<td>70</td>
</tr>
<tr>
<td>Biol6</td>
<td>An inside-out origin for the eukaryotic cell.</td>
<td>204</td>
<td>176</td>
</tr>
<tr>
<td>Biol7</td>
<td>Reagent and laboratory contamination can critically impact sequence-based microbiome analyses</td>
<td>235</td>
<td>191</td>
</tr>
<tr>
<td>Evol1</td>
<td>Revealing the maternal demographic history of Panthera leo using ancient DNA and a spatially explicit genealogical analysis.</td>
<td>62</td>
<td>58</td>
</tr>
<tr>
<td>Evol2</td>
<td>The mysterious Spotted Green Pigeon and its relation to the Dodo and its kindred</td>
<td>56</td>
<td>53</td>
</tr>
<tr>
<td>Evol3</td>
<td>Ingestion of radioactively contaminated diets for two generations in the pale grass blue butterfly.</td>
<td>225</td>
<td>206</td>
</tr>
<tr>
<td>Evol4</td>
<td>Gradual and contingent evolutionary emergence of leaf mimicry in butterfly wing patterns</td>
<td>189</td>
<td>181</td>
</tr>
</tbody>
</table>
Applied SNA

Selected indicators

• Twitter statistics
  • Number of Twitter users
  • Percentage of retweets

• Network
  • Density
  • Mean shortest diffusion path
  • Modularity

• Nodes
  • In-degree
  • Out-degree

• Tweet half-life
• Tweet lifespan

• Percentage of nodes in largest component

• Centrality indicators
• Tweet order
Cantley (2014): Cancer, metabolism, fructose, artificial sweeteners, and going cold turkey on sugar
doi: 10.1186/1741-7007-12-8
Cantley (2014): Cancer, metabolism, fructose, artificial sweeteners, and going cold turkey on sugar

doi: 10.1186/1741-7007-12-8

Tweets: 223
Users: 203
RTs: 64.6%
Half-life: 56.3
Lifespan: 385.6

Hashtags (60)
#diet (15)
#fructose (15)
#sugar (8)
#bmcbiology (7)
#cancer (2)
Cantley (2014): Cancer, metabolism, fructose, artificial sweeteners, and going cold turkey on sugar
doi: 10.1186/1741-7007-12-8

<table>
<thead>
<tr>
<th></th>
<th>Altmetric.com (June 2016)</th>
<th>Twitter REST API (November 2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweets:</td>
<td>223</td>
<td>216</td>
</tr>
<tr>
<td>Users:</td>
<td>203</td>
<td>196</td>
</tr>
<tr>
<td>RTs:</td>
<td>64.6%</td>
<td></td>
</tr>
<tr>
<td>Half-life:</td>
<td>56.3</td>
<td></td>
</tr>
<tr>
<td>Lifespan:</td>
<td>385.6</td>
<td></td>
</tr>
<tr>
<td>Nodes in largest component:</td>
<td>81.6%</td>
<td></td>
</tr>
<tr>
<td>Mean shortest diffusion path:</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>Density:</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Diameter:</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Modularity (infomap):</td>
<td>0.57</td>
<td></td>
</tr>
</tbody>
</table>
Cantley (2014): Cancer, metabolism, fructose, artificial sweeteners, and going cold turkey on sugar

doi: 10.1186/1741-7007-12-8

Users/nodes: 196
Edges: 436
Nodes in largest component: 81.6%
Mean shortest diffusion path: 4.9
Density: 0.01
Diameter: 11
Modularity (infomap): 0.57

Components: [160, 4, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]

In-degree
Number of followers in network
1st tweet last tweet
Cantley (2014): Cancer, metabolism, fructose, artificial sweeteners, and going cold turkey on sugar
doi: 10.1186/1741-7007-12-8

@kamkmk Consultant specialising in digital comms. Recovering Londoner. Ex-charity bod. Also Oxford grad, hill walker & avid tea drinker. All views are my very own.

@MetabolismConf BioMed Central is hosting its second Metabolism, Diet and Disease conference (#MDD2014) in Washington DC from 28-30 May 2014.

@FlavourJournal A peer-reviewed, open access journal that publishes research on how we perceive food and drink, from all relevant disciplines. Most tweets from @drbenjohnson

@drbenjohnson Head of Communities & Engagement at Nature Research, former virologist & also LibDem Councillor in Southwark covering Bermondsey & London Bridge. Personal views

@BMCBiology BMC Biology publishes research and methodology articles of special importance and broad interest, plus reviews and editorial content, in any area of biology
Cantley (2014): Cancer, metabolism, fructose, artificial sweeteners, and going cold turkey on sugar

@kamkmk "The more #sugar you eat, the more [your brain] wants you to eat" - Lewis Cantley (@WeillCornell) talks to @BMCBiology http://t.co/olM3Ol1tWg

@MetabolismConf The problem with #fructose - pioneering metabolism and cancer researcher Lew Cantley talks to @BMCBiology #diet http://t.co/cTexGUGfDe

@FlavourJournal RT @MetabolismConf: The problem with #fructose - pioneering metabolism and cancer researcher Lew Cantley talks to @BMCBiology #diet http://â€”

@drbenjohnson RT @MetabolismConf: The problem with #fructose - pioneering metabolism and cancer researcher Lew Cantley talks to @BMCBiology #diet http://â€”

@BMCBiology .@MetabolismConf committee member Lew Cantley talks to us about cancer, metabolism and why fructose is bad for you http://t.co/oyPH3EB6PI

In-degree

Number of followers in network

1st tweet last tweet
Cantley (2014): Cancer, metabolism, fructose, artificial sweeteners, and going cold turkey on sugar
doi: 10.1186/1741-7007-12-8

@tmotola Founder #RezilirHealth located in Hollywood,Fl. Passionate about #health & #healthcare. #BrainHealth #healthyaging #Lifecoach #antibullying @rezilir
116,297 followers

@Mark_Sisson The real Mark Sisson. Former elite runner-triathlete, author of The Primal Blueprint, blogger at http://t.co/wQvbq5il.
112,047 followers

@ProfTimNoakes Lore of Running, Challenging Beliefs, Waterlogged, Real Meal Revolution, Raising Superheroes author. Emeritus Professor, runner, low carbohydrate diet proponent
79,259 followers

@_sarahwilson_ Journalist + NYT bestselling author. Bike rider. Compulsive hiker. Food waste campaigner. #simplicious #foodwastegirl #IQS
46,573 followers
Cantley (2014): Cancer, metabolism, fructose, artificial sweeteners, and going cold turkey on sugar
doi: 10.1186/1741-7007-12-8

@tmotola RT @Mark_Sisson: A fascinating interview with a cancer researcher about sugar, fructose, and cancer. http://t.co/td6aCekv5r
116,297 followers

@Mark_Sisson A fascinating interview with a cancer researcher about sugar, fructose, and cancer. http://t.co/td6aCekv5r
112,047 followers

@ProfTimNoakes RT @PeterAttiaMD: Must read by Lew Cantley: http://t.co/e3b7lqCPo2
79,259 followers

@sarahwilson Fructose doesn’t supply any energy to your brain at all.. Nor to muscle; it only gets stored as fat. http://t.co/VuR7YeJka6
46,573 followers
Applied SNA

Gunawardena (2014): Models in biology: 'accurate descriptions of our pathetic thinking

doi: 10.1186/1741-7007-12-29

Users/nodes: 153
Edges: 947
Nodes in largest component: 90.9%
Mean shortest diffusion path: 3.9
Density: 0.04
Diameter: 8
Modularity (infomap): 0.25

Half-life: 6.0
Lifespan: 390.8
RTs: 59.0%
Hashtags: 129

Components: [139, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]

#bmcbiology
#genomics
#bioinformatics
#datascience
#mathchat
Applied SNA

Gephi

• Visualize Twitter diffusion network
• Compute network indicators
• Analyze and interpret Twitter diffusion network

Gephi 0.9.1:  https://gephi.org/users/download/
Possibilities and Future Directions of Altmetrics
Future of Altmetrics

• Differentiation of various types of users and uses
  • Across and within underlying platforms

• Enhanced metrics
  • Normalization for age and field differences
  • Network indicators instead of raw counts
  ➢ Validation of indicators!

• Starting point for qualitative analysis
  • Filter/pointer for interviews and focus groups
WT5 Feedback

- Did you learn something new?
- Can you apply this to your work?
- What was missing?

Thank you for providing feedback in the FSCI course evaluation survey: https://goo.gl/forms/YIoVrQql9OLATbnu1 (via Slack)
Thank you for participating!