社會網路分析
(Social Network Analysis)

Min-Yuh Day
戴敏育
Assistant Professor
專任助理教授

Dept. of Information Management, Tamkang University

http://mail.tku.edu.tw/myday/
2014-05-14
週次 (Week) 日期 (Date) 內容 (Subject/Topics)
1 103/02/19 商業智慧導論 (Introduction to Business Intelligence)
2 103/02/26 管理決策支援系統與商業智慧
   (Management Decision Support System and Business Intelligence)
3 103/03/05 企業績效管理 (Business Performance Management)
4 103/03/12 資料倉儲 (Data Warehousing)
5 103/03/19 商業智慧的資料探勘 (Data Mining for Business Intelligence)
6 103/03/26 商業智慧的資料探勘 (Data Mining for Business Intelligence)
7 103/04/02 教學行政觀摩日 (Off-campus study)
8 103/04/09 資料科學與巨量資料分析
   (Data Science and Big Data Analytics)
<table>
<thead>
<tr>
<th>週次</th>
<th>日期</th>
<th>內容（Subject/Topics）</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>103/04/16</td>
<td>期中報告 (Midterm Project Presentation)</td>
</tr>
<tr>
<td>10</td>
<td>103/04/23</td>
<td>期中考試週 (Midterm Exam)</td>
</tr>
<tr>
<td>11</td>
<td>103/04/30</td>
<td>文字探勘與網路探勘 (Text and Web Mining)</td>
</tr>
<tr>
<td>12</td>
<td>103/05/07</td>
<td>意見探勘與情感分析 (Opinion Mining and Sentiment Analysis)</td>
</tr>
<tr>
<td>13</td>
<td>103/05/14</td>
<td>社會網路分析 (Social Network Analysis)</td>
</tr>
<tr>
<td>14</td>
<td>103/05/21</td>
<td>期末報告 (Final Project Presentation)</td>
</tr>
<tr>
<td>15</td>
<td>103/05/28</td>
<td>畢業考試週 (Final Exam)</td>
</tr>
</tbody>
</table>
Outline

• Social Network Analysis (SNA)
  – Degree Centrality
  – Betweenness Centrality
  – Closeness Centrality

• Applications of SNA

Social Network Analysis

Source: http://www.fmsasg.com/SocialNetworkAnalysis/
Social Network Analysis

- A **social network** is a social structure of people, related (directly or indirectly) to each other through a common relation or interest
- **Social network analysis (SNA)** is the study of social networks to understand their structure and behavior

Source: (c) Jaideep Srivastava, srivasta@cs.umn.edu, Data Mining for Social Network Analysis
Social Network Analysis

• Using Social Network Analysis, you can get answers to questions like:
  – How highly connected is an entity within a network?
  – What is an entity's overall importance in a network?
  – How central is an entity within a network?
  – How does information flow within a network?

Source: http://www.fmsasq.com/SocialNetworkAnalysis/
Social Network Analysis

• Social network is the study of social entities (people in an organization, called **actors**), and their **interactions and relationships**.

• The interactions and relationships can be represented with a **network** or **graph**,
  – each **vertex** (or **node**) represents an actor and
  – each link represents a relationship.

• From the network, we can study the properties of its structure, and **the role, position and prestige** of each social actor.

• We can also find various kinds of sub-graphs, e.g., **communities** formed by groups of actors.

Social Network and the Web

• Social network analysis is useful for the Web because the Web is essentially a virtual society, and thus a virtual social network,
  – Each page: a social actor and
  – each hyperlink: a relationship.

• Many results from social network can be adapted and extended for use in the Web context.

• Two types of social network analysis,
  – Centrality
  – Prestige

closely related to hyperlink analysis and search on the Web

Degree

Source: https://www.youtube.com/watch?v=89mxOdwPfxA
Degree

A: 2  
B: 4  
C: 2  
D: 1  
E: 1

Source: https://www.youtube.com/watch?v=89mxOdwPfxA
Density

Edges (Links): 5
Total Possible Edges: 10
Density: 5/10 = 0.5

Source: https://www.youtube.com/watch?v=89mxOdwpfxA
Nodes (n): 10
Edges (Links): 13
Total Possible Edges: \((n \times (n-1)) / 2 = (10 \times 9) / 2 = 45\)
Density: \(13/45 = 0.29\)
Which Node is Most Important?
Centrality

• Important or prominent actors are those that are linked or involved with other actors extensively.

• A person with extensive contacts (links) or communications with many other people in the organization is considered more important than a person with relatively fewer contacts.

• The links can also be called ties. A central actor is one involved in many ties.

Social Network Analysis (SNA)

• Degree Centrality
• Betweenness Centrality
• Closeness Centrality
Alice has the highest degree centrality, which means that she is quite active in the network. However, she is not necessarily the most powerful person because she is only directly connected within one degree to people in her clique—she has to go through Rafael to get to other cliques.

Source: http://www.fmsasq.com/SocialNetworkAnalysis/
Social Network Analysis: Degree Centrality

• Degree centrality is simply the number of direct relationships that an entity has.

• An entity with high degree centrality:
  – Is generally an active player in the network.
  – Is often a connector or hub in the network.
  – Is not necessarily the most connected entity in the network (an entity may have a large number of relationships, the majority of which point to low-level entities).
  – May be in an advantaged position in the network.
  – May have alternative avenues to satisfy organizational needs, and consequently may be less dependent on other individuals.
  – Can often be identified as third parties or deal makers.

Source: http://www.fmsasg.com/SocialNetworkAnalysis/
Social Network Analysis: Degree Centrality
Social Network Analysis: Degree Centrality

A: 2
B: 2
C: 5
D: 3
E: 3
F: 2
G: 4
H: 3
I: 1
J: 1
Rafael has the highest betweenness because he is between Alice and Aldo, who are between other entities. Alice and Aldo have a slightly lower betweenness because they are essentially only between their own cliques. Therefore, although Alice has a higher degree centrality, Rafael has more importance in the network in certain respects.

Social Network Analysis: Betweenness Centrality

• Betweenness centrality identifies an entity's position within a network in terms of its ability to make connections to other pairs or groups in a network.

• An entity with a high betweenness centrality generally:
  – Holds a favored or powerful position in the network.
  – Represents a single point of failure—take the single betweenness spanner out of a network and you sever ties between cliques.
  – Has a greater amount of influence over what happens in a network.

Source: http://www.fmsasq.com/SocialNetworkAnalysis/
Rafael has the highest closeness centrality because he can reach more entities through shorter paths. As such, Rafael's placement allows him to connect to entities in his own clique, and to entities that span cliques.
Social Network Analysis: Closeness Centrality

• Closeness centrality measures how quickly an entity can access more entities in a network.

• An entity with a high closeness centrality generally:
  – Has quick access to other entities in a network.
  – Has a short path to other entities.
  – Is close to other entities.
  – Has high visibility as to what is happening in the network.

Source: http://www.fmsasq.com/SocialNetworkAnalysis/
Social Network Analysis: Closeness Centrality

C: Closeness Centrality = \frac{15}{9} = 1.67
Social Network Analysis: Closeness Centrality

G: Closeness Centrality = 14/9 = 1.56
Social Network Analysis: Closeness Centrality

H: Closeness Centrality = 17/9 = 1.89
Social Network Analysis:
Closeness Centrality

G: Closeness Centrality = 14/9 = 1.56
C: Closeness Centrality = 15/9 = 1.67
H: Closeness Centrality = 17/9 = 1.89
Social Network Analysis: Eigenvalue

Alice and Rafael are closer to other highly close entities in the network. Bob and Frederica are also highly close, but to a lesser value.

Source: http://www.fmsasq.com/SocialNetworkAnalysis/
Social Network Analysis: Eigenvalue

• Eigenvalue measures how close an entity is to other highly close entities within a network. In other words, Eigenvalue identifies the most central entities in terms of the global or overall makeup of the network.

• A high Eigenvalue generally:
  – Indicates an actor that is more central to the main pattern of distances among all entities.
  – Is a reasonable measure of one aspect of centrality in terms of positional advantage.

Source: http://www.fmsasq.com/SocialNetworkAnalysis/
Hubs are entities that point to a relatively large number of authorities. They are essentially the mutually reinforcing analogues to authorities. Authorities point to high hubs. Hubs point to high authorities. You cannot have one without the other.

Source: http://www.fmsasg.com/SocialNetworkAnalysis/
Social Network Analysis: Hub and Authority

- Entities that many other entities point to are called Authorities. In Sentinel Visualizer, relationships are directional—they point from one entity to another.
- If an entity has a high number of relationships pointing to it, it has a high authority value, and generally:
  - Is a knowledge or organizational authority within a domain.
  - Acts as definitive source of information.

# Social Network Analysis

![Network Metrics Table](http://www.fmsasg.com/SocialNetworkAnalysis/)

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Degree</th>
<th>Betweenness</th>
<th>Closeness</th>
<th>Eigenvalue</th>
<th>Hub</th>
<th>Authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Osama bin Laden</td>
<td>Person</td>
<td>44</td>
<td>0.920432092358</td>
<td>1</td>
<td>0.0271</td>
<td>0</td>
<td>0.011</td>
</tr>
<tr>
<td>Abdallah Al-Halabi</td>
<td>Person</td>
<td>2</td>
<td>0</td>
<td>0.654867256637</td>
<td>0.0001</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Abu Mussab al-Zarqawi</td>
<td>Person</td>
<td>84</td>
<td>0.934887847326</td>
<td>0.869451697127</td>
<td>0.7028</td>
<td>0.6572</td>
<td>0.1076</td>
</tr>
<tr>
<td>Al Qaeda</td>
<td>Terrorist Organiz...</td>
<td>85</td>
<td>1</td>
<td>0.962427745664</td>
<td>0.0416</td>
<td>0.3341</td>
<td>0.0156</td>
</tr>
<tr>
<td>Ayman Al-Zawahiri</td>
<td>Person</td>
<td>14</td>
<td>0.045794908783</td>
<td>0.716129032258</td>
<td>0</td>
<td>0</td>
<td>0.0173</td>
</tr>
<tr>
<td>Enaam Arnaout</td>
<td>Person</td>
<td>4</td>
<td>0.031189325814</td>
<td>0.656804733727</td>
<td>0.0001</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Imad Eddin Barakat Yarba</td>
<td>Person</td>
<td>11</td>
<td>0.065045680938</td>
<td>0.704016913319</td>
<td>0.0015</td>
<td>0</td>
<td>0.0025</td>
</tr>
<tr>
<td>Khalid Shaikh Mohammed</td>
<td>Person</td>
<td>32</td>
<td>0.339916464724</td>
<td>0.866059817945</td>
<td>0.002</td>
<td>0</td>
<td>0.1528</td>
</tr>
<tr>
<td>Mohamed Atta</td>
<td>Person</td>
<td>61</td>
<td>0.666268740074</td>
<td>0.820197448334</td>
<td>0.0015</td>
<td>0</td>
<td>0.6316</td>
</tr>
</tbody>
</table>

Application of SNA

Social Network Analysis of Research Collaboration in Information Reuse and Integration

Example of SNA Data Source

IRO 2010: Las Vegas, NV, USA

- Proceedings of the IEEE International Conference on Information Reuse and Integration, IRI 2010, 4-6 August 2010, Las Vegas, Nevada, USA.
- IEEE Systems, Man, and Cybernetics Society 2010

- Reda Alhajj, James B. D. Joshi, Mei-Ling Shyu: Message from Program Co-Chairs. 1
- Stuart Harvey Rubin, Shu-Ching Chen: Forward. 1
- Lotfi A. Zadeh: Precisation of meaning - toward computation with natural language. 1-4
- Reda Alhajj, Shu-Ching Chen, Gongzhu Hu, James B. D. Joshi, Gordon K. Lee, Stuart Harvey Rubin, Mei-Ling Shyu, Lotfi A. Zadeh: Panel title: Critical need for funding of basic and applied research in large-scale computing. 1

Automation, Integration and Reuse across Various Apps

- László István Etesi, André Csillaghy, Lin-Ching Chang: A message-based interoperability framework with application to astrophysics. 1-6
- Awny Alnusair, Tian Zhao, Eric Bodden: Effective API navigation and reuse. 7-12
- Manabu Ohta, Ryohei Inoue, Atsuhiro Takasu: Empirical evaluation of active sampling for CRF-based analysis of pages. 13-18

Source: http://www.informatik.uni-trier.de/~ley/db/conf/iri/iri2010.html
Research Question

• RQ1: What are the scientific collaboration patterns in the IRI research community?

• RQ2: Who are the prominent researchers in the IRI community?

Methodology

• Developed a simple **web focused crawler** program to download literature information about all IRI papers published between **2003 and 2010** from IEEE Xplore and DBLP.
  - 767 paper
  - 1599 distinct author

• Developed a program to convert the list of coauthors into the **format of a network file** which can be readable by social network analysis software.

• **UCINet** and **Pajek** were used in this study for the social network analysis.

Top10 prolific authors (IRI 2003-2010)

1. Stuart Harvey Rubin
2. Taghi M. Khoshgoftaar
3. Shu-Ching Chen
4. Mei-Ling Shyu
5. Mohamed E. Fayad
6. Reda Alhajj
7. Du Zhang
8. Wen-Lian Hsu
9. Jason Van Hulse
10. Min-Yuh Day

Data Analysis and Discussion

• Closeness Centrality
  – Collaborated widely

• Betweenness Centrality
  – Collaborated diversely

• Degree Centrality
  – Collaborated frequently

• Visualization of Social Network Analysis
  – Insight into the structural characteristics of research collaboration networks

### Top 20 authors with the highest closeness scores

<table>
<thead>
<tr>
<th>Rank</th>
<th>ID</th>
<th>Closeness</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>0.024675</td>
<td>Shu-Ching Chen</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.022830</td>
<td>Stuart Harvey Rubin</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.022207</td>
<td>Mei-Ling Shyu</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0.020013</td>
<td>Reda Alhajj</td>
</tr>
<tr>
<td>5</td>
<td>61</td>
<td>0.019700</td>
<td>Na Zhao</td>
</tr>
<tr>
<td>6</td>
<td>260</td>
<td>0.018936</td>
<td>Min Chen</td>
</tr>
<tr>
<td>7</td>
<td>151</td>
<td>0.018230</td>
<td>Gordon K. Lee</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>0.017962</td>
<td>Chengcui Zhang</td>
</tr>
<tr>
<td>9</td>
<td>1043</td>
<td>0.017962</td>
<td>Isai Michel Lombera</td>
</tr>
<tr>
<td>10</td>
<td>1027</td>
<td>0.017962</td>
<td>Michael Armella</td>
</tr>
<tr>
<td>11</td>
<td>443</td>
<td>0.017448</td>
<td>James B. Law</td>
</tr>
<tr>
<td>12</td>
<td>157</td>
<td>0.017082</td>
<td>Keqi Zhang</td>
</tr>
<tr>
<td>13</td>
<td>253</td>
<td>0.016731</td>
<td>Shahid Hamid</td>
</tr>
<tr>
<td>14</td>
<td>1038</td>
<td>0.016618</td>
<td>Walter Z. Tang</td>
</tr>
<tr>
<td>15</td>
<td>959</td>
<td>0.016285</td>
<td>Chengjun Zhan</td>
</tr>
<tr>
<td>16</td>
<td>957</td>
<td>0.016285</td>
<td>Lin Luo</td>
</tr>
<tr>
<td>17</td>
<td>956</td>
<td>0.016285</td>
<td>Guo Chen</td>
</tr>
<tr>
<td>18</td>
<td>955</td>
<td>0.016285</td>
<td>Xin Huang</td>
</tr>
<tr>
<td>19</td>
<td>943</td>
<td>0.016285</td>
<td>Sneh Gulati</td>
</tr>
<tr>
<td>20</td>
<td>960</td>
<td>0.016071</td>
<td>Sheng-Tun Li</td>
</tr>
</tbody>
</table>

Top 20 authors with the highest *betweenness* scores

<table>
<thead>
<tr>
<th>Rank</th>
<th>ID</th>
<th>Betweenness</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.000752</td>
<td>Stuart Harvey Rubin</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.000741</td>
<td>Shu-Ching Chen</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.000406</td>
<td>Taghi M. Khoshgoftaar</td>
</tr>
<tr>
<td>4</td>
<td>66</td>
<td>0.000385</td>
<td>Xingquan Zhu</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.000376</td>
<td>Mei-Ling Shyu</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0.000296</td>
<td>Reda Alhajj</td>
</tr>
<tr>
<td>7</td>
<td>65</td>
<td>0.000256</td>
<td>Xindong Wu</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>0.000194</td>
<td>Chengcui Zhang</td>
</tr>
<tr>
<td>9</td>
<td>39</td>
<td>0.000185</td>
<td>Wei Dai</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>0.000107</td>
<td>Narayan C. Debnath</td>
</tr>
<tr>
<td>11</td>
<td>31</td>
<td>0.000094</td>
<td>Qianhui Althea Liang</td>
</tr>
<tr>
<td>12</td>
<td>151</td>
<td>0.000094</td>
<td>Gordon K. Lee</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>0.000085</td>
<td>Du Zhang</td>
</tr>
<tr>
<td>14</td>
<td>30</td>
<td>0.000072</td>
<td>Baowen Xu</td>
</tr>
<tr>
<td>15</td>
<td>41</td>
<td>0.000067</td>
<td>Hongji Yang</td>
</tr>
<tr>
<td>16</td>
<td>270</td>
<td>0.000060</td>
<td>Zhiwei Xu</td>
</tr>
<tr>
<td>17</td>
<td>5</td>
<td>0.000043</td>
<td>Mohamed E. Fayad</td>
</tr>
<tr>
<td>18</td>
<td>110</td>
<td>0.000042</td>
<td>Abhijit S. Pandya</td>
</tr>
<tr>
<td>19</td>
<td>106</td>
<td>0.000042</td>
<td>Sam Hsu</td>
</tr>
<tr>
<td>20</td>
<td>8</td>
<td>0.000042</td>
<td>Wen-Lian Hsu</td>
</tr>
</tbody>
</table>

**Top 20 authors with the highest degree scores**

<table>
<thead>
<tr>
<th>Rank</th>
<th>ID</th>
<th>Degree</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>0.035044</td>
<td>Shu-Ching Chen</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.034418</td>
<td>Stuart Harvey Rubin</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.030663</td>
<td>Taghi M. Khoshgoftaar</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0.028786</td>
<td>Reda Alhajj</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>0.028786</td>
<td>Wen-Lian Hsu</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>0.024406</td>
<td>Min-Yuh Day</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0.022528</td>
<td>Mei-Ling Shyu</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>0.021277</td>
<td>Richard Tzong-Han Tsai</td>
</tr>
<tr>
<td>9</td>
<td>14</td>
<td>0.017522</td>
<td>Eduardo Santana de Almeida</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
<td>0.017522</td>
<td>Roumen Kountchev</td>
</tr>
<tr>
<td>11</td>
<td>40</td>
<td>0.016896</td>
<td>Hong-Jie Dai</td>
</tr>
<tr>
<td>12</td>
<td>15</td>
<td>0.015645</td>
<td>Narayan C. Debnath</td>
</tr>
<tr>
<td>13</td>
<td>9</td>
<td>0.015019</td>
<td>Jason Van Hulse</td>
</tr>
<tr>
<td>14</td>
<td>25</td>
<td>0.013767</td>
<td>Roumiana Kountcheva</td>
</tr>
<tr>
<td>15</td>
<td>28</td>
<td>0.013141</td>
<td>Silvio Romero de Lemos Meira</td>
</tr>
<tr>
<td>16</td>
<td>24</td>
<td>0.013141</td>
<td>Vladimir Todorov</td>
</tr>
<tr>
<td>17</td>
<td>23</td>
<td>0.013141</td>
<td>Mariofanna G. Milanova</td>
</tr>
<tr>
<td>18</td>
<td>5</td>
<td>0.013141</td>
<td>Mohamed E. Fayad</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>0.012516</td>
<td>Chengcui Zhang</td>
</tr>
<tr>
<td>20</td>
<td>18</td>
<td>0.011890</td>
<td>Waleed W. Smari</td>
</tr>
</tbody>
</table>

Visualization of IRI (IEEE IRI 2003-2010)
co-authorship network (global view)

Visualization of Social Network Analysis

Visualization of Social Network Analysis

Visualization of Social Network Analysis

Summary

• Social Network Analysis (SNA)
  – Degree Centrality
  – Betweenness Centrality
  – Closeness Centrality

• Applications of SNA
References

  http://www.cs.uic.edu/~liub/WebMiningBook.html

• Jennifer Golbeck (2013), Analyzing the Social Web, Morgan Kaufmann.
  http://analyzingthesocialweb.com/course-materials.shtml

• Sentinel Visualizer, http://www.fmsasg.com/SocialNetworkAnalysis/