Visual Analysis of Classification Scheme

Veslava Osińska

Institute of Information Science and Book Studies
Nicolaus Copernicus University Toruń, Poland
Information Visualization (INFOVIS)

- Visualization is a process in which data, information and knowledge are transformed into a visual form exploiting people.
- Visualization is more than a method of computing. It is a process of transforming information into a visual form enabling the viewer to observe, browse, make sense, and understand the information.
- **Information visualization** concentrates on the use of computer-supported tools to explore large amount of abstract data.
- **INSIGHT…**
Three objectives of InfoVis

- Science Domains structure’s discovering
- Information Retrieval aid
- Data clusterization or classification
Infoviz - examples

Title words only from all abstracts accepted for the 2008 ONS Congress and 2008 Advanced Practice Nursing Conference
Tree maps
Visualization of English Wikipedia

by Todd Holloway, Miran Bozicevic, Katy Börner
Science mapping based on a combination of the bibliographic coupling of references and keyword vectors 7.2 million papers and over 16,000 separate journals, proceedings from a five year period: 2001-2005
Computing Classification System

Association for Computing Machinery (ACM) 1964, 1968, 1998

The 1998 ACM Computing Classification System

The ACM Computing Classification System [1998 Version]
Valid through 2009

A. General Literature
B. Hardware
C. Computer Systems Organization
D. Software
E. Data
F. Theory of Computation
G. Mathematics of Computing
H. Information Systems
I. Computing Methodologies
J. Computer Applications
K. Computing Milieux
Visualization principle: similar objects are located closer one another, distinct – far. How to define the similarity of tree’s objects: (sub)classes?
Nonlinear Similarity Metrics

- Metadata of CCS documents:
  - title
  - author
  - year
  - primary class symbol
  - additional classes symbols
  - keywords
  - general terms
  - subject descriptors

As closer thematically two subclasses the more common articles they include. Topic similarity between classes is proportional to the number of common documents.
Mapping Space

- continuos
- symmetric
- curved surface has more topological possibilities
- sphere is easy for navigation and retrieval processes
Similarity Matrix

- Documents metadata 2007 \( (N = 37 \, 543) \)
- Dimension: 353 classes and subclassess number
- Normalization
- Symetrization

\[
N = \sum_{i=1}^{353} \sum_{j=1}^{353} S_{ij}
\]

- \( S_{ij} = 1 \) \((i = j)\) and \( S_{ij} < 1 \) \((i \neq j)\)
### Similarity Matrix cont.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A.1</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>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A.2</td>
<td>0.000302</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>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A.m</td>
<td>0.0075</td>
<td>0.01882</td>
<td>0.05555</td>
<td>0.013616</td>
<td>0.0165</td>
<td>0.11639</td>
<td>0.125</td>
<td>0.0345</td>
<td>0.125</td>
<td>0.03333</td>
<td>0.19231</td>
<td>0.25</td>
<td>0.122</td>
<td></td>
</tr>
<tr>
<td>B.0</td>
<td>0.026586</td>
<td>0.00909</td>
<td>0.013616</td>
<td>0.0165</td>
<td>0.11639</td>
<td>0.125</td>
<td>0.0345</td>
<td>0.125</td>
<td>0.03333</td>
<td>0.19231</td>
<td>0.25</td>
<td>0.122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.1</td>
<td>0.000302</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B.1.0</td>
<td>0.000302</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B.1.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B.1.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D.1.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D.1.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D.1.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B.1.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B.1.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B.1.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
MDS plot

- $\Phi = 0.25$
- $r^2 = 0.47$
- $r_{ijk}^2 = x_i^2 + y_j^2 + z_k^2$

- Morse potential

$$E_s (r) = D_e \left[ \left( 1 - e^{-b(r-R)} \right)^2 - 1 \right]$$
The distribution \( r^2 \)

\[ r^2 = 0.47 \]
Attributes:

1. main class (color) 11
2. level 1,2,3 (luminosity)
3. population (size)
Documents Visualization principle

Classifications weights
Main: Additional

0.6:0.4
0.7:0.3
0.5:0.5
Cartographic projection of sphere surface
Evolution of CCS Scheme

- 1968
- 1978
- 1988
- 1998
- 2007
Evolution of CCS Scheme (2)

Longitudinal mapping

Hague

1968

1978

1988

1998

2007

2007

1998

1988

1978

1968
Evolution of CCS Scheme - conclusion

- **In 80th - systematic of 4 basic categories**: software, hardware, Computer Systems Organization and Information systems.

- **In 90th - clusterization**

  Classification evolves towards stronger adaptation CCS scheme in Digital Library.
Summary

- Metadata gathering
- Data extraction
- Data visualization
- Data analysis on the basis of:
  - cartographic projections;
  - fractal dimensions;
- Implementations:
  - in keywords mapping;
  - in CCS Scheme evaluation;
  - in CCS Scheme and CS domain evolution research
Conslusion

• **Innovation** - First visualization of:
  – classification and universe,
  – thematic categories of single domain,
  – cross-historical of CS,
  – use of own methods: sphere surface, cartographic projections, graphic filters, fractals and feedback mapping (one map verifies another).

• **Potencial of method:**
  – use for automatic generation and evaluation of classification schemes
  – analysis of inter- and multidisciplinarity
  – Forecasting Large Trends in Science
Classifications mapping

- Instead textual mapping -> topology transformation?
Thank you for attention