Scalable Multi-Scale Visual Analytical Tools for Health Science

Katy Börner, Robert Light, Daniel Halsey

Cyberinfrastructure for Network Science Center
School of Informatics and Computing, Indiana University, USA

Regenstrief Institute, Indianapolis, IN
April 29, 2015

Language Communities of Twitter - Eric Fischer - 2012

Descriptive & Predictive Models

Find your way

Terra bytes of data

Find collaborators, friends

Identify trends
Descriptive Models

Multiple levels: Micro ... Macro


Different Levels of Abstraction/Analysis

Macro/Global Population Level

Meso/Local Group Level

Micro Individual Level
### Type of Analysis vs. Level of Analysis

<table>
<thead>
<tr>
<th></th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–100,000 records)</th>
<th>Macro/Global (100,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical Analysis/Profiling</strong></td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of USA, all of science.</td>
</tr>
<tr>
<td><strong>Temporal Analysis (When?)</strong></td>
<td>Funding portfolio of one individual</td>
<td>Mapping topic bursts in 20-years of PNAS</td>
<td>113 Years of Physics Research</td>
</tr>
<tr>
<td><strong>Geospatial Analysis (Where?)</strong></td>
<td>Career trajectory of one individual</td>
<td>Mapping a states intellectual landscape</td>
<td>PNAS publications</td>
</tr>
<tr>
<td><strong>Topical Analysis (What?)</strong></td>
<td>Base knowledge from which one grant draws.</td>
<td>Knowledge flows in Chemistry research</td>
<td>VxOrd/Topic maps of NIH funding</td>
</tr>
<tr>
<td><strong>Network Analysis (With Whom?)</strong></td>
<td>NSF Co-PI network of one individual</td>
<td>Co-author network</td>
<td>NIH’s core competency</td>
</tr>
</tbody>
</table>
Mapping the Evolution of Co-Authorship Networks
Ke, Visvanath & Börner. 2004. Won 1st prize at the IEEE InfoVis Contest.
Mapping the Evolution of Co-Authorship Networks

Ke, Viswanath & Börner. 2004. Won 1st prize at the IEEE InfoVis Contest.

Mapping Indiana’s Intellectual Space

Identify
- Pockets of innovation
- Pathways from ideas to products
- Interplay of industry and academia
Mapping Transdisciplinary Tobacco Use Research Centers Publications

Compare R01 investigator-based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

Supported by NIH/NCI Contract HHSN261200800812

Research Collaborations by the Chinese Academy of Sciences

Huang, Duhon, Hardy & Börner

[Diagram showing research collaborations]
Empowering Anyone to Visualize STI

Example: The Information Visualization MOOC

Course Schedule

Part 1: Theory and Hands-On
- **Session 1** – Workflow Design and Visualization Framework
- **Session 2** – “When:” Temporal Data
- **Session 3** – “Where:” Geospatial Data
- **Session 4** – “What:” Topical Data

Mid-Term
- **Session 5** – “With Whom:” Trees
- **Session 6** – “With Whom:” Networks
- **Session 7** – Dynamic Visualizations and Deployment

Final Exam

Part 2: Students work in teams on client projects.

Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project(30%).

Books Used in the IVMOOCC

**Teaches timely knowledge:**
Advanced algorithms, tools, and hands-on workflows.

**Teaches timeless knowledge:**
Visualization framework—exemplified using generic visualization examples and pioneering visualizations.
How to Classify Different Visualizations?

By
- User insight needs?
- User task types?
- Data to be visualized?
- Data transformation?
- Visualization technique?
- Visual mapping transformation?
- Interaction techniques?
- Or?
### Basic Task Types

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>selection</td>
<td>categorize</td>
<td>category</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>order</td>
<td>rank</td>
<td>ranking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distribution</td>
<td>distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>compare</td>
<td>nominal</td>
<td>differences</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time series</td>
<td>patterns</td>
<td>time</td>
<td>time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>quantity</td>
<td>part-to-whole</td>
<td>proportions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>association</td>
<td>correlate</td>
<td>correlation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Insight Need Types

- categorize/cluster
- order/rank/sort
- distributions
- part-to-whole
- trends (process and time)
- geographic
- compositions (also of text)
- correlations/relationships

### Data Scale Types

- nominal
- ordinal
- interval
- ratio

### Visualization Types

- table
- chart
- map
- network layout

### Graphic Symbol Types

- geometric symbols
- line
- area
- surface
- volume
- linguistic symbols
- text
- numerals
- punctuation marks
- pictorial symbols
- images
- icons
- statistical glyphs

### Graphic Variable Types

- spatial
- position
- form
- color
- optics
- motion

### Interaction Types

- overview
- zoom
- search and locate
- filter
- details-on-demand
- history
- extract
- link and brush
- projection
- distortion
Visualization Types (Reference Systems)

1. **Charts**: No reference system—e.g., Wordle.com, pie charts

2. **Tables**: Categorical axes that can be selected, reordered; cells can be color coded and might contain proportional symbols. Special kind of graph.

3. **Graphs**: Quantitative or qualitative (categorical) axes. Timelines, bar graphs, scatter plots.

4. **Geospatial maps**: Use latitude and longitude reference system. World or city maps.

5. **Network layouts**: Node position might depend on node attributes or node similarity. **Trees**: hierarchies, taxonomies, genealogies. **Networks**: social networks, migration flows.

---

**Types**

<table>
<thead>
<tr>
<th>Insight Need Types page 26</th>
<th>Data Scale Types page 28</th>
<th>Visualization Types page 30</th>
<th>Graphic Symbol Types page 32</th>
<th>Graphic Variable Types page 34</th>
<th>Interaction Types page 26</th>
</tr>
</thead>
<tbody>
<tr>
<td>category/cluster</td>
<td>nominal</td>
<td>geometric symbols</td>
<td>spatial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>order/rank/sort</td>
<td>ordinal</td>
<td>point</td>
<td>position</td>
<td></td>
<td></td>
</tr>
<tr>
<td>distributions</td>
<td>interval</td>
<td>line</td>
<td>form</td>
<td></td>
<td></td>
</tr>
<tr>
<td>also outliers, gaps</td>
<td>ratio</td>
<td>area</td>
<td>color</td>
<td></td>
<td></td>
</tr>
<tr>
<td>comparisons</td>
<td></td>
<td>surface</td>
<td>optics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>trends</td>
<td></td>
<td>volume</td>
<td>motion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(process and time)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>geospatial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>compositions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(also of text)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>correlations/relationships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

*See page 25*
### Graphic Variable Types Versus Graphic Symbol Types

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Point</th>
<th>Line</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image" alt="Point" /></td>
<td><img src="image" alt="Line" /></td>
<td><img src="image" alt="Area" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Size</th>
<th><img src="image" alt="Size" /> NA (Not Applicable)</th>
<th><img src="image" alt="Size" /></th>
<th><img src="image" alt="Size" /></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Shape</th>
<th><img src="image" alt="Shape" /></th>
<th><img src="image" alt="Shape" /></th>
<th><img src="image" alt="Shape" /></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Rotation</th>
<th><img src="image" alt="Rotation" /> NA</th>
<th><img src="image" alt="Rotation" /></th>
<th><img src="image" alt="Rotation" /></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Contour</th>
<th><img src="image" alt="Contour" /> NA</th>
<th><img src="image" alt="Contour" /></th>
<th><img src="image" alt="Contour" /></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Angle</th>
<th><img src="image" alt="Angle" /> NA</th>
<th><img src="image" alt="Angle" /></th>
<th><img src="image" alt="Angle" /></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Closure</th>
<th><img src="image" alt="Closure" /> NA</th>
<th><img src="image" alt="Closure" /></th>
<th><img src="image" alt="Closure" /></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Value</th>
<th><img src="image" alt="Value" /> NA</th>
<th><img src="image" alt="Value" /></th>
<th><img src="image" alt="Value" /></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Hue</th>
<th><img src="image" alt="Hue" /> NA</th>
<th><img src="image" alt="Hue" /></th>
<th><img src="image" alt="Hue" /></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Saturation</th>
<th><img src="image" alt="Saturation" /> NA</th>
<th><img src="image" alt="Saturation" /></th>
<th><img src="image" alt="Saturation" /></th>
</tr>
</thead>
</table>

---

See page 36

---

See pages 36-39
Sci2 Tool – OSGi/CIShell-based Macroscope
Download for free at http://sci2.cns.iu.edu

Use
• **Menu** to read data, run algorithms.
• **Console** to see work log, references to seminal works.
• **Data Manager** to select, view, save loaded, simulated, or derived datasets.
• **Scheduler** to see status of algorithm execution.

All workflows are recorded into a log file (see /sci2/logs/...), and can be re-run for easy replication. If errors occur, they are saved in a error log to ease bug reporting.

All algorithms are documented online; workflows are given in Sci2 Manual at http://sci2.wiki.cns.iu.edu
Needs-Driven Workflow Design

Stakeholders

Types and levels of analysis determine data, algorithms & parameters, and deployment

Data

READ

ANALYZE

VISUALIZE

DEPLOY

Visually encode data

Overlay data

Select visualiz. type

Validation

Interpretation

Types and levels of analysis determine data, algorithms & parameters, and deployment

Data

READ

ANALYZE

VISUALIZE

DEPLOY

Visually encode data

Overlay data

Select visualiz. type

Validation

Interpretation

Needs-Driven Workflow Design
Load **One** File and Run **Many** Analyses and Visualizations

<table>
<thead>
<tr>
<th>Times Cited</th>
<th>Publication Year</th>
<th>City of Publisher</th>
<th>Country</th>
<th>Journal Title (Full)</th>
<th>Title</th>
<th>Subject Category</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>2011</td>
<td>NEW YORK</td>
<td>USA</td>
<td>COMMUNICATIONS OF THE ACM</td>
<td>Plug-and-Play Macroscopes</td>
<td>Computer Science</td>
<td>Borner, K</td>
</tr>
<tr>
<td>18</td>
<td>2010</td>
<td>MALDEN</td>
<td>USA</td>
<td>CTS-CLINICAL AND TRANSLATIONAL SCIENCE</td>
<td>Advancing the Science of Research &amp; Team Science</td>
<td>Experimental Medicine</td>
<td>Falk-Krzesinski, HJ; Borner, K; Contractor, N; Fiore, SM; Hall, KL; Keyton, J; Spring, B; Stokols, DJ; Trochim, W; Uzzi, B</td>
</tr>
<tr>
<td>13</td>
<td>2010</td>
<td>WASHINGTON</td>
<td>USA</td>
<td>SCIENCE TRANSLATIONAL MEDICINE</td>
<td>A Multi-Level Systems Perspective for the Science of Team Science</td>
<td>Cell Biology Research &amp; Experimental Medicine</td>
<td>Borner, K; Contractor, N; Falk-Krzesinski, HJ; Fiore, SM; Hall, KL; Keyton, J; Spring, B; Stokols, DJ; Trochim, W; Uzzi, B</td>
</tr>
</tbody>
</table>

---

Co-author and many other bi-modal networks.
Existing Interfaces for Health-related Data

9) NIH RePORTER: Visual Interface to Biomedical Funding Data in U.S.
10) CShell/Sci2 World and Science Visualizations of NIH RePORTER Data
11) NIH RePORTER: NIH Map
12) BBSRC: Visual Interface to Biomedical Funding Data in UK
13) IAI Multidimensional Analysis
14) Scraawl: Twitter Analysis
15) Illuminated Diagram: Searchable World and Science Maps

NIH RePORTER: Visual Interface to Biomedical Funding Data in US

http://projectreporter.nih.gov/
CIShell/Sci2 World and Science Visualizations of NIH RePORTER Data

Sci2 Desktop
NIH RePORTER: NIH Map

http://nihmaps.org/

BBSRC: Visual Interface to Biomedical Funding Data in UK

http://www.bbsrc.ac.uk/
BBSRC: Temporal animation

IAI Multidimensional Analysis
Scraawl: Twitter Analysis

https://www.scraawl.com/

Illuminated Diagram: Searchable World and Science Maps

http://cns.iu.edu/interactive_displays.html
Illuminated Diagram: Search detail

References


All papers, maps, tools, talks, press are linked from [http://cns.iu.edu](http://cns.iu.edu)

These slides will soon be at [http://cns.iu.edu/docs/presentations](http://cns.iu.edu/docs/presentations)

CNS Facebook: [http://www.facebook.com/cnscenter](http://www.facebook.com/cnscenter)

Mapping Science Exhibit Facebook: [http://www.facebook.com/mappingscience](http://www.facebook.com/mappingscience)