Visualization Techniques for Analyzing and Sharing Relational Data

Cody Dunne
November 9, 2015
cdunne@us.ibm.com
ibm.biz/cogvislab
Visualization Techniques .... Relational Data

Agenda

Introduction  Node-Link  VoroGraph  GraphTrail  StoryFacets  Discussion
Cody Dunne, PhD – Cognitive Visualization RSM
Web: ibm.biz/codydunne
Email: cdunne@us.ibm.com

- Epidemiology/dynamic networks
- Layout readability
- Exploration provenance
- Network type overviews
- Group/set visualization
- Aggregation techniques
- Literature exploration
- News term occurrence
- Computer network traffic
The Data Problem
### Why Visualization?

**Anscombe’s quartet – Table**

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
</tr>
<tr>
<td>10.00</td>
<td>8.04</td>
<td>10.00</td>
<td>9.14</td>
<td>10.00</td>
</tr>
<tr>
<td>8.00</td>
<td>6.95</td>
<td>8.00</td>
<td>8.14</td>
<td>8.00</td>
</tr>
<tr>
<td>13.00</td>
<td>7.58</td>
<td>13.00</td>
<td>8.74</td>
<td>13.00</td>
</tr>
<tr>
<td>9.00</td>
<td>8.81</td>
<td>9.00</td>
<td>8.77</td>
<td>9.00</td>
</tr>
<tr>
<td>11.00</td>
<td>8.33</td>
<td>11.00</td>
<td>9.26</td>
<td>11.00</td>
</tr>
<tr>
<td>14.00</td>
<td>9.96</td>
<td>14.00</td>
<td>8.10</td>
<td>14.00</td>
</tr>
<tr>
<td>6.00</td>
<td>7.24</td>
<td>6.00</td>
<td>6.13</td>
<td>6.00</td>
</tr>
<tr>
<td>4.00</td>
<td>4.26</td>
<td>4.00</td>
<td>3.10</td>
<td>4.00</td>
</tr>
<tr>
<td>12.00</td>
<td>10.84</td>
<td>12.00</td>
<td>9.13</td>
<td>12.00</td>
</tr>
<tr>
<td>7.00</td>
<td>4.82</td>
<td>7.00</td>
<td>7.26</td>
<td>7.00</td>
</tr>
<tr>
<td>5.00</td>
<td>5.68</td>
<td>5.00</td>
<td>4.74</td>
<td>5.00</td>
</tr>
</tbody>
</table>
### Why Visualization?

**Anscombe’s quartet – Statistics & Visualization**

<table>
<thead>
<tr>
<th>Property in Each Case</th>
<th>Value</th>
<th>Equality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of x</td>
<td>9</td>
<td>Exact</td>
</tr>
<tr>
<td>Variance of x</td>
<td>11</td>
<td>Exact</td>
</tr>
<tr>
<td>Mean of y</td>
<td>7.50</td>
<td>2 decimal places</td>
</tr>
<tr>
<td>Variance of y</td>
<td>4.122 or 4.127</td>
<td>3 decimal places</td>
</tr>
<tr>
<td>Correlation between x &amp; y</td>
<td>0.816</td>
<td>3 decimal places</td>
</tr>
<tr>
<td>Linear regression line</td>
<td>$y = 3.00 + 0.500x$</td>
<td>2 &amp; 3 decimal places</td>
</tr>
</tbody>
</table>

![Graphs of Anscombe's quartet](image)
No catalogue of techniques can convey a willingness to look for what can be seen, whether or not anticipated. Yet this is at the heart of exploratory data analysis. ... the picture-examining eye is the best finder we have of the wholly unanticipated.

– Tukey, 1980
Design Choices
Charm vs. clarity

Charm and Direct Appeal

Clarity and Power
Design Choices
Charm vs. clarity

Nigel Holmes

Edward Tufte
# Data Analysis Process

<table>
<thead>
<tr>
<th>Kandel et al., 2012</th>
<th>Heer, 2013 (unpublished)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery</td>
<td>Acquisition</td>
</tr>
<tr>
<td>Wrangling</td>
<td>Cleaning</td>
</tr>
<tr>
<td></td>
<td>Integration</td>
</tr>
<tr>
<td>Profiling</td>
<td>Visualization</td>
</tr>
<tr>
<td>Modeling</td>
<td>Modeling</td>
</tr>
<tr>
<td>Reporting</td>
<td>Presentation</td>
</tr>
<tr>
<td></td>
<td>Dissemination</td>
</tr>
</tbody>
</table>
Data Analysis Process

Dissemination ← Acquisition

Presentation ↔ Cleaning

Modeling ↔ Integration

Visualization
Node-Link Network Visualizations
Node-Link Graph Visualization

General

Graph \approx Network
Node \approx Vertex \approx Entity
Edge \approx Link \approx Relationship \approx Tie

<table>
<thead>
<tr>
<th>Node 1</th>
<th>Node 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>Bob</td>
</tr>
<tr>
<td>Alice</td>
<td>Cathy</td>
</tr>
<tr>
<td>Cathy</td>
<td>Alice</td>
</tr>
</tbody>
</table>
## Node-Link Network Visualization

Tweets of the #Win09 Workshop

<table>
<thead>
<tr>
<th></th>
<th>User 1</th>
<th>User 2</th>
<th></th>
<th>User 1</th>
<th>User 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20andlife</td>
<td>barrywellman</td>
<td>15</td>
<td>danevans87</td>
<td>informor</td>
</tr>
<tr>
<td>2</td>
<td>20andlife</td>
<td>BrianDavidson</td>
<td>16</td>
<td>danevans87</td>
<td>NetSciWestPoint</td>
</tr>
<tr>
<td>3</td>
<td>barrywellman</td>
<td>elizabethmdaly</td>
<td>17</td>
<td>danielequercia</td>
<td>BrianDavidson</td>
</tr>
<tr>
<td>4</td>
<td>barrywellman</td>
<td>informor</td>
<td>18</td>
<td>danielequercia</td>
<td>drewconway</td>
</tr>
<tr>
<td>5</td>
<td>BrianDavidson</td>
<td>hcraygliangjie</td>
<td>19</td>
<td>danielequercia</td>
<td>ipeirotis</td>
</tr>
<tr>
<td>6</td>
<td>BrianDavidson</td>
<td>informor</td>
<td>20</td>
<td>danielequercia</td>
<td>johnflurry</td>
</tr>
<tr>
<td>7</td>
<td>BrianDavidson</td>
<td>NetSciWestPoint</td>
<td>21</td>
<td>danielequercia</td>
<td>loyan</td>
</tr>
<tr>
<td>8</td>
<td>byaber</td>
<td>barrywellman</td>
<td>22</td>
<td>danielequercia</td>
<td>loyan</td>
</tr>
<tr>
<td>9</td>
<td>byaber</td>
<td>danielequercia</td>
<td>23</td>
<td>danielequercia</td>
<td>mcscharf</td>
</tr>
<tr>
<td>10</td>
<td>byaber</td>
<td>mcscharf</td>
<td>24</td>
<td>danielequercia</td>
<td>NetSciWestPoint</td>
</tr>
<tr>
<td>11</td>
<td>chrisnordyke</td>
<td>RebeccaBadger</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>12</td>
<td>danevans87</td>
<td>barrywellman</td>
<td>106</td>
<td>sechrest</td>
<td>Japportreport</td>
</tr>
<tr>
<td>13</td>
<td>danevans87</td>
<td>BrianDavidson</td>
<td>107</td>
<td>sechrest</td>
<td>loyan</td>
</tr>
<tr>
<td>14</td>
<td>danevans87</td>
<td>drewconway</td>
<td>108</td>
<td>sechrest</td>
<td>RebeccaBadger</td>
</tr>
</tbody>
</table>
Node-Link Network Visualization
Tweets of the #Win09 Workshop
Node-Link Network Visualization

Who uses network analysis?

- Sociology
- Scientometrics
- Politics
- Urban Planning
- Biology
- Archaeology
- WWW
Network Visualization in IBM

Services exposing network data (many more coming)
Alternate Network Visualizations

Dunne et al., 2012

Gove et al., 2011

Blue et al., 2008

Henry & Fekete, 2006

Freire et al., 2010

Wattenberg, 2006
GLEAM Epidemic Model
Population basins, local commuting, and global flights

- Population layer: census areas
- Short range mobility layer: commuting
- Long range mobility layer: air travel
Choropleth Maps
Gastner and Newman 2012 Election (State)

http://www-personal.umich.edu/~mejn/election/2012/
Area Morphing
Gastner and Newman 2012 Election (State)

http://www-personal.umich.edu/~mejn/election/2012/
Area Morphing
Gastner and Newman 2012 Election (County)

http://www-personal.umich.edu/~mejn/election/2012/
Contiguous Density Equalizing Cartograms
Gastner and Newman 2012 Election (County)

http://www-personal.umich.edu/~mejn/election/2012/
Voronoi Tessellation

Western Europe
Centroidal Voronoi Tessellation – Animated!

Western Europe
VoroGraph
Animated Transitions
**New York Influenza Scenario:** Consider a new strain of influenza starting in New York City in mid-February. In a dense population, the disease could quickly reach pandemic proportions. Indeed, the millions of commuters and visitors could carry the virus home with them. As with H1N1, the authorities would have to face both the local and global spreading of the disease.
VoroGraph
Non-Contiguous Edge Coding
VoroGraph
Non-Contiguous Edge Coding
VoroGraph
Force-Directed Group-in-a-Box
• Equal-population hexagons discretize the space for **countability**
• Easier **attribute comparison** with color/size coding
• Hexagons make clear it is an **artificial representation**
• Enforces a degree of **generalization**
• **Contiguous relationship** display

GraphTrail
Networks can be
- Large & complex
- Multivariate
- Heterogeneous

Analysis can take
- Many sessions
- Many users
• Aggregation
• Drag-and-drop interactions
• Integrated exploration history
GraphTrail
Aggregating charts
GraphTrail
Drag and drop interaction
GraphTrail
Provenance/history visualization

Municipality = Écija OR Marchena

Terra Sigillata
Black Gloss Ware | Iberian /
GraphTrail
Provenance & chart parameterization

Tag Cloud
Dig Sites
Municipality = Écija OR Marchena
Grouped by Ceramic Component

Provenance & chart parameterization
GraphTrail
Pivoting & Derived Attributes

A1: author of P1
A2
A3
P1: part of P2

cites

P2

Author
- Name
- Affiliation
- Paper.Topic*

Paper
- Topic
- Year

Proceeding
- Country
- Date
GraphTrail can make the same findings as other tools
  – And more!

New users can make findings

New users understand the exploration history
  – And usually motivation!
Field Study With Archaeologists

“How were Iron-Age communities integrated into the political and economic structure of the Roman Empire?”

“How were urban social hierarchies within the Roman provinces structured and articulated?”

Brughmans, T., et al., 2011
GraphTrail
Field study – current practice
GraphTrail
Field study – analyses
1. Number of nodes, edges, types

<table>
<thead>
<tr>
<th></th>
<th>Nodes</th>
<th>Types</th>
<th>Edges</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHI</td>
<td>10K</td>
<td>3</td>
<td>20K</td>
<td>3+</td>
</tr>
<tr>
<td>Archaeology</td>
<td>13K</td>
<td>24</td>
<td>20K</td>
<td>35</td>
</tr>
</tbody>
</table>

2. Number of charts

20 – 30 per session
A system for exploring large multivariate, heterogeneous networks using aggregation by node and edge attributes,

A method for capturing a user’s exploration history and integrating it directly into the workspace, and

A longitudinal field study and a qualitative lab study that prove the utility of these approaches.


StoryFacets
Individual exploration

Analyst A

Data filtering → Analysis → Report
StoryFacets
Real-world exploratory analysis

Data filtering → Analysis → Report

Analyst A

Analyst B
What happens if different filtering is done?

C-Suite
I need an overview. But I also want to understand the analysis process.

Sales
Can you explain step C in more detail?
I am your database.
StoryFacets
Star Wars data
STORYFACETS
Generating Multiple Representations of Exploratory Data Analysis for Communication
StoryFacets
Trail facet – Star Wars trilogy ships & pilots

Filtering

Pivoting

Attribute exploration

Branching
StoryFacets
Dashboard facet – Star Wars male vs. female

Male
4 times larger than female
5 times more starships
Appeared at all series

(a)

Starships

(b)

Female
Drives fastest spaceship
Only appear from new trilogy

(d)

Starships

(e)

Starships

(f)

max_atmosphering_speed

Starships

(g)

max_atmosphering_speed
The only hermaphrodite in Star Wars is Jabba Desilijic Tiure.
StoryFacets

Story (slideshow) facet – Star Wars character height

Among the residents select who is tallest.
StoryFacets
Linked back to trail facet

(a) Trail Facet

(b) Dashboards or Info Graphics Facet

(c) Story Facet
• 19 casual participants in a public setting
  – Novice users were able to select different communication format based on the target audience and nature of the content
• Three visualization professional expert reviewers
• Exploratory data analysis is much more than the initial exploration session
• Unified platform to support exploration, collaboration, discussion, and presentation
• Study with casual participants and expert reviewers reveal future directions

Discussion

Introduction
Node-Link
VoroGraph
GraphTrail
StoryFacets
Discussion
Graph Drawing 2017
Hosted in September by IBM
Cambridge, MA, USA

Cody Dunne, IBM Watson
cdunne@us.ibm.com

T. Alan Keahey, IBM Watson
alankeahey@us.ibm.com
iCity: Urban Informatics for Sustainable Metropolitan Growth

Version 2, November 2015

A project funded by:
The Ontario Research Fund – Research Excellence Round 7

Principal Investigator:
Prof. Eric J. Miller, Research Director
University of Toronto Transportation Research Institute

The research team

University of Toronto
Eric Miller
Baher Abdullai
Steve Listerbrook
Judy Farveden
Mark Fox
Mark Merat
Paul Hoss
Matt Roorda
Amir Shafiee
Matter Serafini
Evd Slezk

OCAD University
Sara Diamond
Greg Van Alstyne
Isabel Morales

University of Waterloo
Jeff Cassellio
University of Regina
Mohamed El Darabieb
IBM
Cody Dune
Magda Mourad
Samir Sahm

Our partners

Cellnet
City of Toronto
IBM Canada
Maximum City
Region of Waterloo
Waterfront Toronto
Ontario Ministry of Transportation
Ontario Ministry of Municipal Affairs & Housing
VARDEC
GraphTrail & StoryFacets
Blocks to build on

• Additional visualizations
• Context aware comments
• Exploration hints for new paths
• Streaming/temporal data, intelligent updates, and resurgent relevancy
• Linked chart parameterization for comparisons with auto chain layout and compression
• Advanced modeling and analytics (IBM Catalyst, IBM Watson)
• User management & security