Measuring and Improving the Readability of Network Visualizations

Cody Dunne
cdunne@cs.umd.edu

NIST ACMD Seminar – August 28, 2012
Network visualization is highly useful, but hard!

There are many ways to make it easier
Krist Wongsuphasawat wins the Interactive Category of the Information Is Beautiful Award. In a contest sponsored by The Guardian to show European debt crisis over time, Kris's visualization, cleverly called European Bubbles, wins the Interactive Category. The award committee described his visualization as "amazing, slick and full-featured interactive." Kris gives loads of room for exploratory fun, details and follow up. In short, a winner! Check out Kris's work at the European Bubbles website. On May 22nd – 23rd, 2012, HCIL 29th Symposium will offer a Masters degree in Human-Computer Interaction which includes internships, work with HCIL researchers, interdisciplinary classes, and more. Learn more about this exciting new program, or visit Maryland's iSchool website for application. Click here to register.

WHO WE ARE

The HCIL has a long, rich history of transforming the experience people have with new technologies. From understanding user needs, to developing and evaluating those technologies, the lab's faculty, staff, and students.

Social Network Strategies for Surviving the Zombie Apocalypse

Jennifer Golbeck
Some of my work…

NodeXL (Smith et al., 2009; Shneiderman & Dunne, 2012; etc...)

GraphTrail (Dunne et al., 2012)

Action Science Explorer (Dunne et al., 2012; Gove et al., 2011)

STICK (Shneiderman et al., 2011; Gove et al., 2011)

NetGrok (Blue et al., 2008)
Networks
### Edge List

<table>
<thead>
<tr>
<th>Vertex1</th>
<th>Vertex2</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>

### Adjacency Matrix

<table>
<thead>
<tr>
<th></th>
<th>Alice</th>
<th>Bob</th>
<th>Cathy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Bob</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cathy</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Who Uses Network Analysis

Sociology

Scientometrics

Politics

Urban Planning

Biology

Archaeology

WWW
What we have done: Open Tools

- **YASNAT**: NodeXL
- Data providers ("spigots")
- FOSS
NodeXL

Collect data, Excel analysis, statistics, visualization, layout algorithms, filtering, clustering, attribute mapping...
What we have done: *Open Data*

NodeXLGraphGallery.org
What we have done: Open Scholarship

- Webshop 2011: NSF, Google, Intel
- Webshop 2012: NSF, GRAND, Yahoo!, Google
- Other Workshops: ICWSM12, NetSci, HyperText12, Cape Town, Yeungnam, Italy, dg.o
What we have done: Open Scholarship
http://nodexl.codeplex.com
We still have a problem...
Node-Link Visualization is Hard
So what can we do?
Better Layouts...

Hachul & Jünger, 2006
Better Layouts...

Hachul & Jünger, 2006
Alternate visualizations...

Gove et al., 2011

Henry & Fekete, 2006

Dunne et al., 2012

Freire et al., 2010

Wattenberg, 2006
Plan of attack

Readability metrics
- Global/local
- Taxonomy/layout aids

Motif simplifications

Meta-layouts

Evaluations
- Readability metrics
- User studies
Readability Metrics
Why measure readability?

Lee et al., 2003
Measuring Readability

Simple rules or heuristics
Davidson & Harel, 1996

User performance
Huang et al., 2007

Global readability metrics
Purchase, 2002

Global Readability Metrics

• How understandable is the network drawing?
• Example: Journal may suggest
  • 0% node occlusion
  • <2% edge tunneling
  • <5% edge crossing
More Metrics!

- Global
  - Node overlap, drawing space used...
- Local for Nodes, Edges, & Groups
- Task-by-metric taxonomy

Dunne & Shneiderman, 2009, HCIL TR
E.g., Node Overlap

Global readability metric

\[ [0,1] \text{ where:} \]
\[ 0 = \text{Complete overlap} \]
\[ 1 = \text{No overlap} \]

\[
a = \text{area}\left(\bigcup_{j=1}^{n} \text{bounds}(n_j)\right)
\]

\[
a_{\text{max}} = \sum_{j=1}^{n} \text{area}(\text{bounds}(n_j))
\]

\[
\mathcal{N}_n = \frac{a}{a_{\text{max}}} - \text{max node area}
\]

Node readability metric

Ratio of node area that overlaps other nodes

\[
a_j = \text{area}\left(\bigcup_{k=1}^{n} \text{bounds}(n_j) \cap^* \text{bounds}(n_k)\right)
\]

\[
\mathcal{N}_{n_j} = 1 - \frac{a_j}{\text{area}(\text{bounds}(n_j))}
\]
Existing metrics

Our metrics

New
- Node overlap
- Edge tunnel
- Drawing space used
- Group overlap

Local
- Edge crossing
- Angular resolution
- Edge crossing angle
Assisted Manipulation

- Real-time ranking & coloring by metrics
- Snap-to-maxima

14 edge tunnels

0 edge tunnels

Images: Cody Dunne
Multi-Criteria Optimization

• User-defined energy function
  • Interactive view of task-by-metric taxonomy

• Simulated annealing
  • Metropolis et al., 1953; Kirkpatrick et al., 1983

• Searches layout space
  • Hill climbing

• Expensive...
Takeaway

• Raise awareness
• Localized identification of where improvement is needed
• Optimization recommendations for tasks
• Interactive, semi-automatic, and fully-automatic optimization
Motif Simplification
Observations

1: There are repeating patterns in networks (motifs)

2: Motifs often dominate the visualization

3: Motifs members can be functionally equivalent
Graph Summarization...

Navlakha et al., 2008
Our approach: Motif Simplification to reduce visualization complexity
Motif Simplification

Fan Motif

2-Parallel Motif
Glyph Design Guidelines

• Representative: topology, count & attributes
• Easily distinguishable
• Easily comparable
• Allow overlaps
Glyph Design: Fan
Glyph Design: Parallel
Glyph Design: General D-Parallel
Cliques too!
Interactivity

Fan motif: 133 leaf vertices with head vertex “Theory”
## Quantifying Effectiveness

<table>
<thead>
<tr>
<th>Metric (before ⇒ after)</th>
<th>Lostpedia</th>
<th>VOSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes</td>
<td>513 ⇒ 25</td>
<td>3958 ⇒ 559</td>
</tr>
<tr>
<td>Number of edges</td>
<td>586 ⇒ 40</td>
<td>4380 ⇒ 765</td>
</tr>
<tr>
<td>Graph density</td>
<td>0.00446</td>
<td>0.00056</td>
</tr>
<tr>
<td>Fan motifs</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>2-parallel motifs</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>Fan sizes</td>
<td>7–247</td>
<td>17–852</td>
</tr>
<tr>
<td>2-parallel sizes</td>
<td>7–28</td>
<td>2–50</td>
</tr>
<tr>
<td>Node-node overlap</td>
<td>0.981 ⇒ 0.983</td>
<td>0.709 ⇒ 0.971</td>
</tr>
</tbody>
</table>
User Impressions

“I’m overwhelmed, ... this is like one of those vision tests at the eye doctor”

“Now I can see the central pages...[and] pairwise connections”
Discussion

Motif simplification effective for

• Reducing complexity
• Understanding larger or hidden relationships

However

• Frequent motifs may not be covered
• Glyph design has tradeoffs

Details & algs. in Dunne & Shneiderman, HCIL TR 2012
Meta-Layouts
Meta-Layouts

- Layout using groupings
- Attributes
- Topology
- Manual

Rodrigues et al., 2011
Pennsylvania Innovation

Grouped using Clauset et al. (2004)

Laid out with Group-in-a-Box (Rodrigues et al., 2011)

Data:
Scott Dempwolf

Image:
Cody Dunne
Pennsylvania Innovation

Largest groups and no inter-group edges

Data: Scott Dempwolf
Image: Cody Dunne
Clustered Twitter Network

Source: Rodrigues et al., 2011
Group-in-a-Box Layout

Source: Rodrigues et al., 2011
Current Meta-Layouts

• Poorly show ties (Rodrigues et al., 2011)
  • Long ties
  • Group arrangement
  • Aggregate relationships

OR

• Poorly show nodes & groups (Noack, 2003)
  • Require much more space
  • Harder to see groups
Upcoming Meta-Layouts

- Donut layout
  - See topology better, slight space increase
  - Extension of Tu & Shen, 2007

- Group-aware force-directed layout
  - See topology well, at the cost of space

- Space-filling force-directed layout
  - Balances space and topology
  - Extension of Wood & Dykes, 2008
Meta-layouts Good For...

- Large or high density networks
- Highlighting hidden relationships
- Recursive nesting
Better Node-Link Visualizations

Readability metrics
- Global/local
- Taxonomy/layout aids

Motif simplifications

Meta-layouts

Evaluations
- Readability metrics
- User studies
We still have some concerns...

Networks can be
- Large & complex
- Multivariate
- Heterogeneous

Analysis can take
- Many sessions
- Many users
Beyond Node-Link Visualizations
GraphTrail

• Aggregation
• Drag-and-drop interactions
• Integrated exploration history

Dunne et al., CHI 2012
GraphTrail Demo
Design–Aggregating Charts

Authors
- Allison Druin
- Ben Shneiderman
- Bonnie John Brad Myers
- Elizabeth Mynatt
- Elliot Soloway
- Hiroshi Ishii
- I. MacKenzie
- Jakob Nielsen
- James Landay
- John Carroll
- Jonathan Grudin
- Mary Rosson
- Michael Muller
- Shumin Zhai
- Stephen Rosebrough
- Stuart Card
- William Buxton

Papers
- Anthropomorphism
- Applications
- Cognitive Factors in Design
- CSCW
- End User Programming
- InfoVis
- Lab Reports
- Multimodal UI
- Miscellaneous
- Web
- User Centered Design
- User Modeling
- Target Acquisition
- Tangible UI
Design–Drag & Drop Interaction
Design–Pivoting & Derived Attributes

Author
- Name
- Affiliation
- Paper.Topic*

Paper
- Topic
- Year

Proceeding
- Country
- Date
Design–Visual Exploration History
Lab Study

• Can GraphTrail make the same findings as other tools?
• Can new users make findings?
• Can new users understand the exploration history?
Lab Study

• Can GraphTrail make the same findings as other tools?

  Yes, and more!

  “Find all papers written by all Japanese authors”
Lab Study–Exploration

Can new users make findings?

“What can you tell me about HCI research in Georgia?”
Lab Study–History

Can new users understand the exploration history?

Yes, and often motivation too!
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
Field Study–Current Practice
Field Study–GraphTrail Analyses
Field Study–Sharing

“the way I structured it makes sense to me.”
Field Study–Sharing

<table>
<thead>
<tr>
<th>Sites</th>
<th>Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Municipality</strong></td>
<td><strong>Municipality</strong></td>
</tr>
<tr>
<td>Écija</td>
<td>Écija OR Marchena</td>
</tr>
<tr>
<td>Marchena</td>
<td>Marchena</td>
</tr>
<tr>
<td>Alcalá de Guadaira</td>
<td>Alcalá de Guadaira</td>
</tr>
<tr>
<td>Córdoba</td>
<td>Córdoba</td>
</tr>
<tr>
<td>Osuna</td>
<td>Osuna</td>
</tr>
<tr>
<td>Fuentes de Andalucia</td>
<td>Fuentes de Andalucia</td>
</tr>
<tr>
<td>Estepa</td>
<td>Estepa</td>
</tr>
<tr>
<td>Lora del Río</td>
<td>Lora del Río</td>
</tr>
<tr>
<td>Utrera</td>
<td>Utrera</td>
</tr>
<tr>
<td>Coria</td>
<td>Coria</td>
</tr>
<tr>
<td>Márden de la Frontera</td>
<td>Márden de la Frontera</td>
</tr>
<tr>
<td>Alcalá del Río</td>
<td>Alcalá del Río</td>
</tr>
<tr>
<td>Arco de la Frontera</td>
<td>Arco de la Frontera</td>
</tr>
<tr>
<td>Alcalá de la Ria</td>
<td>Alcalá de la Ria</td>
</tr>
<tr>
<td>Alba de la Ria</td>
<td>Alba de la Ria</td>
</tr>
<tr>
<td>Alcazar</td>
<td>Alcazar</td>
</tr>
<tr>
<td>Cabeza de San Juan</td>
<td>Cabeza de San Juan</td>
</tr>
<tr>
<td>Cos Harnosa</td>
<td>Cos Harnosa</td>
</tr>
<tr>
<td>Espada</td>
<td>Espada</td>
</tr>
<tr>
<td>Gómez</td>
<td>Gómez</td>
</tr>
<tr>
<td>Lebrija</td>
<td>Lebrija</td>
</tr>
<tr>
<td>Montemellano</td>
<td>Montemellano</td>
</tr>
<tr>
<td>Palma del Río</td>
<td>Palma del Río</td>
</tr>
<tr>
<td>Peñafiel</td>
<td>Peñafiel</td>
</tr>
<tr>
<td>San Lucar de Barrameda</td>
<td>San Lucar de Barrameda</td>
</tr>
<tr>
<td>San Roque</td>
<td>San Roque</td>
</tr>
<tr>
<td>Santa María</td>
<td>Santa María</td>
</tr>
<tr>
<td>San Juan de Amarguesa</td>
<td>San Juan de Amarguesa</td>
</tr>
<tr>
<td>Vilaflor</td>
<td>Vilaflor</td>
</tr>
<tr>
<td>Villaverde del Río y Minas</td>
<td>Villaverde del Río y Minas</td>
</tr>
<tr>
<td>Alcalá de los Rios</td>
<td>Alcalá de los Rios</td>
</tr>
<tr>
<td>Zafra</td>
<td>Zafra</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sites</th>
<th>Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Terra Sigillata</strong></td>
<td><strong>Terra Sigillata</strong></td>
</tr>
<tr>
<td>Black Gloss Ware</td>
<td>Black Gloss Ware</td>
</tr>
<tr>
<td>Iberian</td>
<td>Iberian</td>
</tr>
</tbody>
</table>

**Municipality = Écija OR Marchena**
Field Study–Sharing

Tag Cloud

Dig Sites

Municipality = Écija OR Marchena

Grouped by Ceramic Component
Discussion

• Exposing exploration history without extra effort
• Scalability

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
GraphTrail Overview

- 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.
Take Away Messages

Create effective node-link visualizations in NodeXL:

- **Readability metrics** to guide improvements
- **Motif simplification** to reduce complexity
- **Meta-layouts** to more clearly show ties and groups

Explore heterogeneous networks with GraphTrail:

- **Drag-and-drop exploration** of attribute aggregates
- **Remember & share** exploration history

Cody Dunne
cdunne@cs.umd.edu