MatLink: Enhanced Matrix Visualization for Analyzing Social Networks

Nathalie Henry and Jean-Daniel Fekete

Presented by:
Chang-Han Jong and Andreea Olea
Social Networks

- **Actors** (people, subgroups, organizations, collectivities)
- **Relationships** (friendship, interactions, communications, transactions, movement, kinship)

**Categories:**
- **Almost trees** (e.g. genealogy, STD transmission patterns)
- **Almost complete graphs** (e.g. trade between countries or companies)
- **Small world networks**
  - locally dense, high clustering coefficient, small average shortest path, power-law degree distribution
Social Network Representations

1. Node Link
2. Matrix
Node Link Representation

- Great for visualizing “almost trees” networks
- Better at path finding than matrix representations

- Link overlapping - interferes with neighborhood finding and link counting
- Link length - interferes with neighborhood finding
- Hard to find node by name when # nodes is large
- Nodes not laid out in a predictive order
- Poor at visualizing almost complete graphs and dense networks (links starting or ending at a node vs. links crossing the node)
Matrix Representation

- Can display dense networks without edge overlap
- Orderable
- Easier to count links

- Display area quadratic in the # of vertices
- Poor for path-finding tasks
Node Link vs. Matrix Study

- Study by Ghoniem, M., Fekete, J.D. and Castagliola, P.

- **Task 1: Approx. Node Count**
  - Node Link Representation - reduced readability as size/density of graph increases

- **Task 2: Approx. Edge Count**
  - Statistically insignificant difference between representations

- **Task 3: Most Connected Node**
  - Correctness slightly better on matrix representations

- **Task 4: Find Node by Label**
  - Readability of Node Link diagram decreases with the increase in size of the graph
  - Matrix not affected by link desity
Node Link vs. Matrix Study

- Study by Ghoniem, M., Fekete, J.D. and Castagliola, P.

- **Task 5: Find a Link between Two Nodes**
  - Node Link - larger graphs take longer
  - Matrix - doesn’t depend on size/density
  - Node Link much faster for small graphs, Matrix much faster for large/dense graphs

- **Task 6: Find Common Neighbor**
  - Node Link slower for larger graphs
  - Matrix doesn’t depend on size/density

- **Task 7: Find Path**
  - Both deteriorate with size
  - Node Link performs much better
NL, MAT and MatLink

(a) Node-Link (NL)  (b) Matrix (MAT)  (c) MatLink
MatLink

- Hybrid Approach
  - links overlaid on the borders of a matrix
  - dynamic topological feedback as the pointer moves
- Outperforming
  - Significant advantages for most tasks
  - Especially *path-related* tasks
- (Nodes x Nodes) 2D matrix
  - Need non-random order
  - Modified TSP algorithm for find ordering
  - Minimize but not guarantee line-crossing
- Once click
  - Red lines: selected nodes/paths
  - Green line: a possible shortest path
Video Demo...

http://www.lri.fr/~nhenry/matlink/matlink.mov
EVALUATE MAT, NL AND MATLINK
EXPERIMENTS

- 3 visualizations
  - NL, MAT, MatLink

- 6 social networks
  - Real social networks from Pajek, UCINET, Complex Networks, Graph Drawing contests and InfoVis 2004 contests

- 5 Tasks
  - Common neighbor
  - Shortest path
  - Most connected
  - Articulation point
  - Largest clique
DATA SETS

- **Measures**
  - Vertex number \((S)\)
  - Edge number \((E)\)
  - Density \((D)\)
  - Cluster coefficient \((CC)\)

- **Not intend to control**
  - Average path length
  - Degree distribution

- **Excludes by Ghoniem et al.**
  - Very small graphs with <24 nodes
  - Small and very sparse (NL certainly will do good)
  - Very dense (MAT certainly will do good)

- **Minimize extraneous interpretation issues**
  - Assigning different labels when presenting the same graph twice
  - Produced a set of randomly filtered graphs for each instance
  - Selected the filtered graphs with \(D/CC\) similar to the original
  - Relaxed the \(D\) constraint when shrinking large graphs (difficult to keep low \(D\) with high \(CC\))
## OVERVIEW OF DATA SETS

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Vertex number</th>
<th>Edge Number</th>
<th>Density</th>
<th>Clustering Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infovis</td>
<td>47</td>
<td>114</td>
<td>0.23</td>
<td>0.83</td>
</tr>
<tr>
<td>Dolphins</td>
<td>47</td>
<td>202</td>
<td>0.30</td>
<td>0.42</td>
</tr>
<tr>
<td>Fraternity</td>
<td>47</td>
<td>294</td>
<td>0.36</td>
<td>0.72</td>
</tr>
<tr>
<td>Genealogy</td>
<td>94</td>
<td>192</td>
<td>0.15</td>
<td>0.59</td>
</tr>
<tr>
<td>Collaboration</td>
<td>94</td>
<td>313</td>
<td>0.19</td>
<td>0.90</td>
</tr>
<tr>
<td>USAirports</td>
<td>94</td>
<td>990</td>
<td>0.33</td>
<td>0.84</td>
</tr>
</tbody>
</table>
EXPERIMENT SETUP

- **Fair Game**
  - Let MAT/NL also have dynamic highlighting

- **Only three interactions**
  - Mousing over an element
  - Clicking on an element
  - Dragging on several elements using the left mouse button

- **Only vertices are clickable**

- **NL**
  - Uses LinLog algorithm to draw

- **MAT**
  - Augmented TSP to order matrices
SUBJECTS

- Univ. of Sydney
  - 18 students/researchers in Graph Drawing Research Group
- Univ. of Paris-Sud
  - 18 students/researchers in HCI
PROCEDURE

1. Brief interview
   1. Tutorial sheet

2. Practice for few minutes (40 min in average)
   1. Ends if all questions are correctly answered

3. Answering questions (<40 min)
   1. Rule: answer correctly as rapidly as possible.
   2. Skip is allowed
   3. 90 questions: 5 tasks performed on 3 visualization of 6 graphs
   4. 60 seconds limit for each question
   5. Divided Sessions with rest

4. Questionnaire
   1. Use of each representation
   2. Rating by task and preference

5. Debriefing
   1. Answering remaining questions
   2. Collecting final comments
DATA COLLECTED

- Primary measure
  - Answer correctness (0-3)

- Secondary measure
  - Time

- All trials are reported, even those with errors

- Cautious about high error rate and fast wrong answer
RESULTS

- Four way ANOVA (Analysis of Variances) on Visualization, Task, |Nodes|, |Edges|
- Tukey’s HSD Post Hoc test (TT)

Fig. 2: Overall average Score and Time by visualization. Average number of error by task and visualization.
ALL TASKS

- ANOVA reveals significant for Vis on Score and Time

Fig. 2: Overall average Score and Time by visualization. Average number of error by task and visualization.
CONCLUSION

- Users were significantly more accurate with MatLink than with MAT for most tasks

- Accurate for path-based tasks:
  - NL > MAT mostly
  - MatLink > NL in shortPath task

- ....more comparison

- Users had fewer errors with NL on dense graphs than sparse ones
  - Explanation: highlighting of the shortest paths
Critiques

- Window game ‘Minesweeper’
- Mirror in undirected Graphs may be better utilized.
  - The cells can be coded in color/texture for more information
  - Automatically Traverse by specified algorithms
  - Generated NL for selected subgraph or community might be useful
  - Local dense is solved. But how to deal with large scale networks?