VIGOR: INTERACTIVE VISUAL EXPLORATION OF GRAPH QUERY RESULTS

Author: Robert Pienta, Fred Hohman, Alex Endert, Acar Tamersoy, Kevin Roundy, Chris Gates, Shamkant Navathe, Duen Horng Chau

PRESENTER: JIAHONG CHEN
INSTRUCTOR: PROF. TAMARA MUNZNER

Nov, 20th, 2017
How can we extract useful information from large scale network?
BACKGROUND

• Graph querying: locate entities with specific relationships among them
  • financial transaction networks
    • flag “near cliques” formed among company insiders
  • money-laundering
  • online auctions
    • uncover fraudsters and their accomplices
• Bioinformatics
• Social network analysis
BACKGROUND

- Few work focused on developing visualization system to help understand graph structure and rich data.
  - underlying data from the nodes
  - structure of each subgraph result
  - large number of results
  - potential overlap in node and edges among

https://vimeo.com/237670479
DATA TO VIS AND DERIVED RESULTS

- DBLP Dataset.
- DBLP is a computer science bibliography website.
- Co-authorship network of DBLP’s computer science bibliography data, focusing on the data mining and information visualization communities
  - 59,655 authors; 48,677 papers; 7,236 sessions
  - 417 proceedings; 21 conferences; 1,634,742 relations
- Derived results
  - A novel interactive visual analytics system, for exploring and making sense of query results

<table>
<thead>
<tr>
<th>VAD Idiom</th>
<th>VIGOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>What: Data</td>
<td>Network data with vertex and edges</td>
</tr>
<tr>
<td>What: Derived</td>
<td>Subgraph and feature clusters</td>
</tr>
<tr>
<td>Why: Tasks</td>
<td>Find subgraph according to query results and cluster features</td>
</tr>
<tr>
<td>Scale</td>
<td>Millions of relations and tens of thousands of co-authors</td>
</tr>
</tbody>
</table>
ILLUSTRATIVE USAGE SCENARIO

Exemplar View
• The analyst starts with only the structure of the graph query, then incrementally adds node value constraints to narrow in on specific results
• Choose conference by name
• Narrows down the network by choosing mutual authors.

<table>
<thead>
<tr>
<th>VAD Idiom</th>
<th>VIGOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>How: Encode</td>
<td>Use lines to show connected relationships; colors for different nodes</td>
</tr>
<tr>
<td>How: Reduce</td>
<td>Item filtering</td>
</tr>
</tbody>
</table>
ILLUSTRATIVE USAGE SCENARIO

Fusion Graph
- After adding Exemplar View filters, induced subgraph of all the combined results from the original query will be generated in Fusion Graph.
- Shixia Liu’s papers and co-authors who have published papers together at VAST and KDD.

VAD Idiom | VIGOR
---|---
How: Manipulate | Reorder, realign, hovering highlight
ILLUSTRATIVE USAGE SCENARIO
ILLUSTRATIVE USAGE SCENARIO

Subgraph Embedding
- Query: an author who has published two papers with a co-author, where the papers were published to VAST and another conference will return 2550 results.
- Subgraph Embedding view provides an overview of all results by clustering

<table>
<thead>
<tr>
<th>VAD Idiom</th>
<th>VIGOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>How: Facet</td>
<td>Linked highlighting</td>
</tr>
<tr>
<td>How: Encode</td>
<td>colors for different clusters</td>
</tr>
</tbody>
</table>
ILLUSTRATIVE USAGE SCENARIO

Feature Explorer
• Compare two clusters in the Feature Explorer
• Color: same as the cluster color
• X-axis: # Papers/ # co-authors/publication year/ # authors
• Y-axis: number of papers
• The bar charts show the top-k most common values,

<table>
<thead>
<tr>
<th>VAD Idiom</th>
<th>VIGOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>How: Encode</td>
<td>colors for different clusters</td>
</tr>
</tbody>
</table>
ILLUSTRATIVE USAGE SCENARIO
METHODOLOGY & ARCHITECTURE

- Extract Features - Calculate the topological- and node-features.
- Vectorize - Merge the common features into per-result vectors.
- Aggregate & Normalize into Signature - Reduce the large input vectors into uniform signatures.
- Reduce & Cluster - Reduce the signatures using dimensionality reduction.
METHODOLOGY & ARCHITECTURE (CONT’D)

• **Extract Features.**
  - **Structural features**
    - Subgraph neighborhood and egonet information
    - An egonet of a node, \( i \), is (a) the neighbor nodes of \( i \), (b) the edges to these neighbors and (c) all the edges among neighbors.
  - Node degree – number of neighbors
    - \( d_i = |N(i)|, N(i) \) is the neighboring nodes of node \( i \)
  - Egonet edges - a unweighted graph, simply counting the number of edges
    - \( E(ego(i)) = \sum_{j \in N(i)} (\sum_{e_{jk} \in E(j)} \delta_{ik}) \)
    - \( \delta_{ik} = \begin{cases} 1, & \text{if } k \in N(i) \\ 0, & \text{if } k \notin N(i) \end{cases} \)
  - Egonet neighboring nodes - the number of neighbor nodes of neighbor nodes
    - \( |N(ego(i))| = | \cup_{j \in N(i)} N(j) | \)
  - Clustering coefficient – ratio of closed loop subgraph and total number of edges
    - \( c_i = \frac{2|e_{jk} \in E(i): j,k \in N(i)|}{|N(i)|\cdot(|N(i)|-1)} \)
METHODOLOGY & ARCHITECTURE (CONT’D)

- Vectorize
  - Nodes feature
    - Author name
    - Number of co-authors
    - Number of conference
  - Merge common feature
METHODOLOGY & ARCHITECTURE (CONT’D)

• Aggregate & Normalize
  • For each feature, statistic characteristics are extracted: mean, variance, skewness, and kurtosis
  • Generate feature at same length: \( 4 \cdot (|f_s| + |f_t|) \)

• Reduce & Cluster
  • Dimensionality reduction reduces the feature dimension to 2D, which helps to vis.

<table>
<thead>
<tr>
<th>VAD Idiom</th>
<th>VIGOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>How: Encode</td>
<td>Attribute aggregation</td>
</tr>
</tbody>
</table>
EVALUATION

• User Study
  • 12 participants from computing related majors.
    • 7 female, 5 male
    • age 21 to 31
  • Paid $10 for 70 minutes test.
  • Dataset: DBLP co-authorship network
• Real World Application: Discovering Cybersecurity Blindspots
USER STUDY

- **Task 1:** Find the count of ICDM conference papers by Daniel Keim.
- **Task 2:** From the last two years of KDD publications, find and list the authors who are on more than one paper with “entity” in the name.
- **Task 3:** Find the number of distinct groups of researchers that Tobias Shreck is in from INFOVIS publications.
- **Task 4:** Among coauthors of at least two papers together at INFOVIS and KDD, who has the most publications.
USER STUDY

- Quantitative Results
  - Tasks: find out the software affect by executing four tasks and exam the average task time, and average # of errors.

- Observations and Subjective Results
  - Participants rate various aspects comparing both systems
CONTRIBUTIONS OF VIGOR

• Novel visual analytics system, VIGOR
• Exploring and making sense of graph querying results
• Exemplar-based interactive exploration
  • bottom-up: how many similar values are matched to each query-node
  • top-down: how a particular node value filters the results from the whole structure
• Novel result summarization through feature-aware subgraph result embedding and clustering.
  • VIGOR provides a top-down, high-level overview
  • Clustering node-feature and structural result similarity
• An integrated system fusing multiple coordinated views
  • Brushable linked views among Exemplar View, Subgraph Embedding View, and the Fusion Graph
CRITIQUE

• The number of people for user study might not enough and they are all professional users.
• Query sentence is hard to generate for non-professionals.
• The co-authorship is limited to one-hop
Thank you!