DBLP Coauthor Network of InfoVis Conference Program Committee

Rongjian Lan  |  111 93 7264  |  rongjian.lan@gmail.com  |  Nov.8 2012

Data Description:

The data used in this analysis is from DBLP computer science bibliography website using DBLP XML Requests[1]. The data is generated in this way: 1) find 36 famous researchers in the field of Visualization as anchors; 2) find their direct coauthors and their number of collaborations to form a graph; 3) get all the author’s 1-author publications. The attribute for nodes (authors) is their number of 1-author publications and the attribute for edges (author-author collaboration) is the number of collaborations.

Author list (all served as InfoVis program committee):


Headlines:
1. The Important Person May Not Have A Big Name

I use the force-directly layout to show the overall network in the top big image, which shows that authors working in the InfoVis field are tightly connected. The red nodes represent the author who served as PC (program committee) in InfoVis Conference, and the black nodes represent the other collaborators. The size of the nodes corresponds to the number of first-author publications of the author. The width of the edges corresponds to the number of collaborations (publishing a paper) between the two authors.
We can easily perceive that several big nodes (like Ben Shneiderman, Daniel A. Keim, George G. Robertson and John T. Stasko) serve as the anchor author who are connected most of the authors in the graph. They should have high value of betweenness centrality. To clear the graph, I filter the graph by betweenness centrality to the point where for the first time all authors are anchors, as shown in the first image at the bottom. We can see that even though the whole graph is well connected, the important nodes themselves are not all connected. Helen C. Purchase and Pat Hanrahan are not directly connected with the other anchors in the graph. Also, John T. Stasko acts as a 5-connector in the graph which shows his effort on collaborating with other famous researchers in the field.

Then I reverse the filter constraint to show the first non-anchor author who turns out to be Catherine Plaisant. She acts as a 4-connector in the filtered graph and her betweenness centrality is higher than most of the other anchors, which means she collaborates actively and broadly with the searchers in this field. To make the current graph a connected graph, I further reverse the filter constraint and get two non-anchor authors, Maneesh Agrawala and Terry Winograd, who serve as bridges connecting Pat Hanrahan to the main part of the graph.
2. Collaboration Pattern Differs Among Groups
The above image shows a connection-based clustered graph of the original network. Each rectangle shows a cluster of authors in force-directed layout and the edges between groups are hidden for clarity. The clustering reveals groups of authors who collaborate intensely with each other. They usually are with the same research lab or with different labs frequently collaborating with each other. As in the image, some clusters feature nodes connecting around one big node, while other clusters shows nodes of small cliques connected by some important bridge nodes. This shows that people in some research labs tend to collaborates within the lab while other labs are collaborating with each other a lot.

To simplify the graph, I filter the nodes by the number of first-author publications of equal to or bigger than 10. This leaves the authors who are well focused in this field and may be well-experienced researchers. We can see that some clusters, even big in size, have a few well-focused researchers, while other clusters have a lot of experienced researchers collaborating with each other intensely.
3. To Become Famous and Successful, Work Hard And Socialize Hard

The last image show a coordinate-based layout of the network with X-Axis representing the degree of the node and Y-Axis representing the number of first-author publications of the nodes (authors). As we can see, most of the program committee serving researchers (red nodes) have a balanced number of publications and number of collaborations with different people.

Critiques:

1. The selection of edge or nodes is not working well. Sometimes it's hard to select what I want.
2. Node label size is better to be proportional to the node's size to show the importance of the node.
3. The graph is scaled as the panel is scaled which changes the visual distance of nodes, confusing user's understanding of the data.
4. The color mapping for categorical field is limited. It’s automatically mapped and I have no way to assign color to different categories.
5. The analysis process including the current graph and configuration cannot be saved which make it hard to do continual analysis.
6. The scale slider controls both nodes and edges while scaling edge and node separately can make the graph cleaner.
7. No way to export graph shot as SpotFire image exporting feature.
8. No progress report on the graph layout process.
9. No revert function provided.
10. The limit on the node size (100) make it hard to emphasize important nodes in very large data set.
11. Edge filter doesn't work on the combined edge between group.
12. The system doesn’t provide filtering on categorical column, or user-defined columns.