APPENDIX A
EVALUATION OF NODE SUMMARIZATION METHODS

We have validated SymNMF-based node summarization method by experimentally comparing with classical graph clustering approaches. Three alternative algorithms were implemented: graph partition by normalized cut, ratio association objectives [1], and k-means clustering [2] on the content of graph nodes. The experiment was carried out on AMiner [3] V8 citation database, and the citation influence graph of 18010 visualization-related papers are included. Details about the test data set can be found in Section 5.3.1. The content of each paper (node) is modeled by converting its title+abstract clip into a vector representation through the Word2Vec tool [4], which is then used in the k-means clustering. We consider two quality measures for the node summarization (clustering). One is the sum of all flow rates on the summarization, evaluating the static objective for influence graph summarization (IGS). The other measure is the silhouette value of clustering [5] where the node dissimilarity is defined by the Euclidean distance over the node's vector representations (e.g., the individual paper's content representation). This is to evaluate the content-based consistency by node clusterings.

The performance comparison on flow rate and silhouette value is shown in Figure 1 under the setting of 20 and 40 clusters respectively. In each sub-figure, the influence graphs are grouped into multiple bins by their initial graph size (#nodes), i.e., $(40,200]$[1], $(200,400]$, $(400,600]$, $\cdots$. In Figure 1(a), a dashed line is drawn to indicate the number of graphs in each bin. Most influence graphs have less than 3000 nodes (15859 graphs, 88.1% of all influence graphs). For each bin, the average flow rate (silhouette) of graphs is plotted on Y axis while the corresponding X axis value is set to the bin center, i.e., 120, 300, 500, $\cdots$. The result on flow rate (Figure 1(a)(b)) indicates that SymNMF achieves a much larger overall flow rate than classical graph clustering algorithms in all bins with varying graph sizes. The average flow rate by SymNMF is 42.4% larger than the best of alternatives (Ratio Association) when #clusters=20, and 47.7% larger when #clusters=40. This validates the theoretical result in [6] that SymNMF approximately maximizes the static IGS objective on overall flow rates. Meanwhile, on content-based clustering consistency, Figure 1(c)(d) indicate a divided performance. When the graph size is small ($\leq$1000 when #clusters=20, $\leq$1600 when #clusters=40), the silhouette value of SymNMF is comparable or even larger than k-means, and much better than graph topology based clustering algorithms. When the graph size is large, SymNMF performs worse than k-means and is comparable with ratio association based graph clustering algorithm. It is also noticed that when the graph size is large ($>$ 600 when #clusters=20, $>$ 1000 when #clusters=40), all tested algorithms converge to very small silhouette values ($<$0.05), which implies little content consistency in the resulting node clustering.

[1] We remove graphs smaller than 40 nodes because they do not need summarization and can not reach a cluster size of 40.

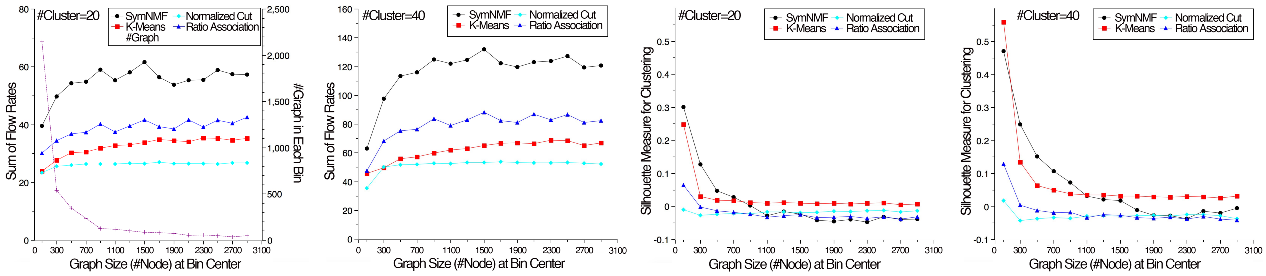
The experiment result on node summarization has three implications. First, the flow-based objective and the content-based measure form a trade-off. There is no algorithm that maximizes both measures in all conditions. Second, on small graphs, SymNMF obtains the best trade-off among all algorithms tested. It achieves flow rate maximization while has a comparable consistency performance with the content-based k-means algorithm. On large graphs, all algorithms under the current number of clusters (20 or 40) fail to detect a consistent node clustering. For these reasons, we select SymNMF as the node summarization method in Eiffel. Third, the number of clusters becomes an important parameter when the influence graph is large. With a higher number of clusters (20$\rightarrow$40), the consistency degradation for SymNMF becomes slower. This suggests that, if an appropriate number of clusters is chosen for the summarization, SymNMF can achieve the best trade-off between the flow-based heuristics and the content consistency.

We caution that the node summarization that maximizes overall flow rate (e.g., SymNMF) and the content-based node clustering (e.g., k-means) might not be appropriately compared. They serve different objectives, influence flow maximization and content consistency. This work focused on the visualization of influence patterns, for which the flow-based objective should be prioritized and the SymNMF method should be selected.

APPENDIX B
EVALUATION OF EDGE SUMMARIZATION METHODS

We conducted an experiment to evaluate the edge summarization algorithms presented in Section 4.2.2, i.e., connected top-n flow graph, MWST, and Maximal Padded MWST. The experiment used the same citation influence graphs with the experiment of node summarization algorithms (Appendix A). These graphs are first summarized by SymNMF using a cluster number of 20. Next, each alternative edge summarization algorithm is applied to the IGS generated by SymNMF. In Section 4.2.2, two objectives are proposed for edge summarization: 1) maximize the overall flow rate; 2) reduce visual clutter and minimize edge crossings in the final display. Note that all the three edge summarization algorithms produce visualizations with little visual cluster, which satisfy the first objective. MWST and maximal padded MWST guarantee a tree structure for visualization, so that there will be no edge crossing. The connected top-n flow graph, although does not guarantee a free of edge crossings, will also introduce little visual clutter because there is only $n-1$ remaining edges in the summarization. Therefore, in this experiment we mainly consider the first objective, i.e., the maximization of overall flow rate in the summarization.

Figure 2 shows the performance of three edge summarization algorithms in achieving the first objective. For each citation influence graph, we computed the overall flow rate after the edge summarization. The percentage of this rate preserved from that of the IGS generated by SymNMF is displayed in the Y axis of Figure 2, averaged across all the graphs in a same bin. The X axis in Figure 2 indicates the



(a) Flow rate (#Cluster=20) (b) Flow rate (#Cluster=40) (c) Silhouette (#Cluster=20) (d) Silhouette (#Cluster=40)

Fig. 1: Performance of node summarization algorithms on AMiner V8 citation influence graphs. Only papers on the visualization topic are considered to reduce the experiment data size.

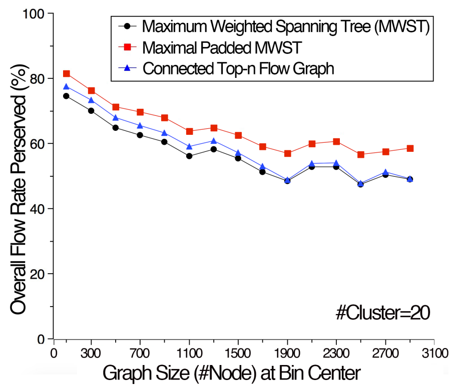


Fig. 2: Edge summarization performance measured by the percentage of flow rate preserved from the IGS after node summarization (larger is better). The experiment setting is the same with Figure 1.

average graph size in a bin. We found that maximal padded MWST preserves a higher percentage of overall flow rate in edge summarization compared with MWST and connected top-n flow graph, for graphs of any size. Connected top-n flow graph is better than MWST when the graph size is small ($\leq$2000) and is similar to MWST for large graphs. Based on these experimental results, we choose maximal padded MWST as the default edge summarization algorithm in Eiffel.

Appendix C
Eiffel System Prototype

We implemented a web-based, functional Eiffel prototype on citation influence analysis, available at http://118.190.210.193/eiffel/. It now indexes research papers on the visualization topic from CiteSeerX and AMiner data sets. In the back end, these data sets are first imported and stored in MongoDB. By offline processing, for each indexed paper, we build a maximal influence graph and its evolutionary IGS summarization according to Section 3, and finally save them in JSON files. In the front end, influence graph summarizations are retrieved online upon search request, and then visualized by the Eiffel flow map implemented with D3 package [7].

From user's perspective, the system starts from a search interface. S/he can type in interested keywords and access a list of related papers, or select a related venue (e.g., TVCG) on the search page and get returned with a list top papers in that venue. Each list view follows a Google Scholar style design, i.e., the cited/citing paper list and the corresponding venue can be further expanded. The main Eiffel citation influence graph view will show up when one paper in the list is selected. An additional feature is, for these papers included in both citation data sets (AMiner and CiteSeerX), the influence graph summarized with each data set can be displayed side-by-side for comparison. The evolutionary visualization will then be animated in a synchronized manner. Screenshots and more details about the prototype can also be found in the supplemental video.

Appendix D
User Experiment on Eiffel Summarization

We conducted a controlled user experiment to compare the user performance in understanding the Eiffel summarization result and using a Google Scholar (GS) like interface. We apply two influence graph data in the formal test session of the experiment. The first data is a large-sized influence graph of the paper: "Graph Visualization and Navigation in Information Visualization: A Survey" in the AMinerV8 data set, with 29324 nodes (Graph I). Figure 3(a)(b) give the Eiffel visualization (EI) and the GS-like interface (GI) of this graph respectively. The second data is a medium-sized influence graph of the paper: "Planet-Sized Batched Dynamic Adaptive Meshes (P-BDAM)" in the AMinerV8 data set, with 1080 nodes (Graph II). Figure 4(a)(b) give the Eiffel visualization (EII) and the GS-like interface (GII) of this graph respectively. An example of the study document for one participant is provided in Appendix E.

Appendix E
User Study Document on Eiffel Summarization

Note: This is a simplified user study procedure translated from its full Chinese version for description purpose only. As the subjects are all Chinese students, we prepared all the user study documents in Chinese to reduce the learning cost. In the full document, we also include the detailed description of both visualization interfaces (i.e., Eiffel and Google Scholar) in the training phase. Four versions of the document were prepared, i.e., EI-GII, EII-GI, GI-EII, GII-EI, to counterbalance the learning and interaction effects. The below document is for EI-GII. (E=Eiffel, G=Google Scholar, I=Data I, II=Data II)

1. Training Session (E)

Eiffel on a sample paper: http://118.190.210.193/eiffel/graph.html?aminerV8_id=945237&citeseerx_id=5699756&selected=aminerV8&source=aminerV8_citeseerx_&r=269487022

Task: Write down THREE topic streams influenced most by the selected paper, both directly and indirectly (using 2∼3 keywords in sequence separated by comma for each topic)

2. Formal Test Session (EI)

Eiffel on paper "Graph Visualization and Navigation in Information Visualization: A Survey": http://118.190.210.193/eiffel/graph.html?aminerV8_id=288226&citeseerx_id=4592929&selected=aminerV8&source=aminerV8_citeseerx_&r=344554744

Task: Write down THREE topic streams influenced most by the selected paper, both directly and indirectly (using 2∼3 keywords in sequence separated by comma for each topic)

Subjective Questions: Rate your degree of agreement with the below statements (give a score of 0∼6: 0=strong disagree, 3=neutral, 6=strong agree) Q1. This visualization tool helps me to complete the above tasks. Q2. I like to use this visualization tool.

3. Training Session (G)

Google Scholar on a sample paper: http://118.190.210.193/eiffel/citation.html?id=945237&source=aminerV8&title=VAST%202007%20Contest%20-%20Blue%20Iguanodon&action=s&r=544635488

Task: Write down THREE topic streams influenced most by the selected paper, both directly and indirectly (using 2∼3 keywords in sequence separated by comma for each topic)

4. Formal Test Session (GII)

Google scholar on paper "Planet-Sized Batched Dynamic Adaptive Meshes (P-BDAM)": http://118.190.210.193/eiffel/citation.html?id=665758&source=aminerV8&title=Planet-Sized%20Batched%20Dynamic%20Adaptive%20Meshes%20(P-BDAM)&action=s&r=105316987

Task: Write down THREE topic streams influenced most by the selected paper, both directly and indirectly (using 2∼3 keywords in sequence separated by comma for each topic)

Subjective questions: Rate your degree of agreement with the below statements (give a score of 0∼6: 0=strong disagree, 3=neutral, 6=strong agree) Q1. This visualization tool helps me to complete the above tasks. Q2. I like to use this visualization tool.

Appendix F
Expert Feedback

We invited a senior professor on computer science (with 200+ refereed scientific publications) to a short expert study. First, we demonstrated the Eiffel prototype to him with the case study of the Jigsaw paper as an example (Section 6.3.1). Next, the expert was allowed to interact with the prototype to better understand Eiffel techniques. He was not required to complete any tasks because we currently do not support the influence visualization in his research area. Finally, the expert provided his feedback in an interactive, unstructured discussion session. We prepared and asked a few questions, including "Do you have any requirement regarding the citation influence analysis of a paper? What are those?", etc., but for most of the time, he talked about his impression on Eiffel and suggestions of its improvement, without sticking to our questions.

The feedback can be summarized as the followings, most of which we have incorporated into the future plan of Eiffel. First, the expert found the visualization of indirect influence to be the most useful feature of Eiffel, whereas the previous systems such as Google Scholar lists the direct influence only. He proceeded to suggest us to visualize the influence of scientific giants and their work which had big indirect influence to the research communities. Some of the names he mentioned include Albert Einstein, Alan Turing, etc. Second, he thought Eiffel could help to identify the original innovation related to a research work. To better serve this goal, we would need to visualize influence graphs constructed by a backward search along the influence links. Third, he pointed out another potential of Eiffel to be the visualization of development of a discipline or a topic. The multi-source evolutionary influence graphs are one way for completing this task, but we must resolve the deficiency of Eiffel in visualizing very large influence graphs.

References

[1] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 888–905, 1997.

[2] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proceedings of Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, 1967.

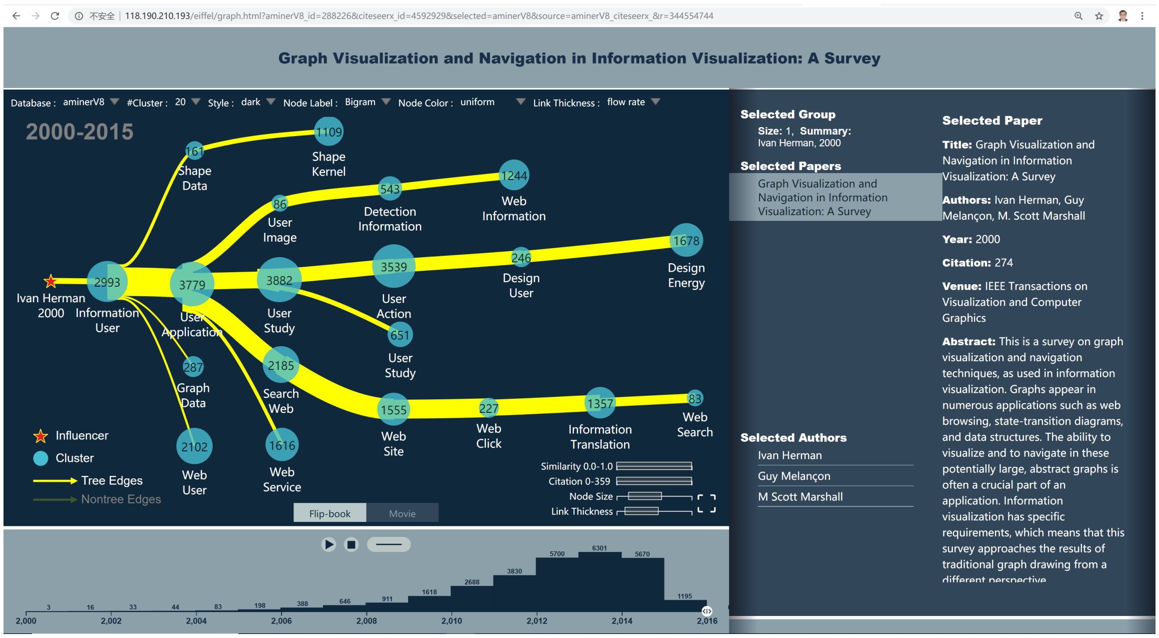
[3] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "Arnetminer: Extraction and mining of academic social networks," in *KDD'08*, 2008, pp. 990–998.

[4] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *NIPS'13*, 2013, pp. 3111–3119.

[5] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[6] L. Shi, H. Tong, J. Tang, and C. Lin, "VEGAS: Visual influEnce GrAph Summarization on Citation Networks," *IEEE TKDE*, vol. 27, no. 12, pp. 3417–3431, 2015.

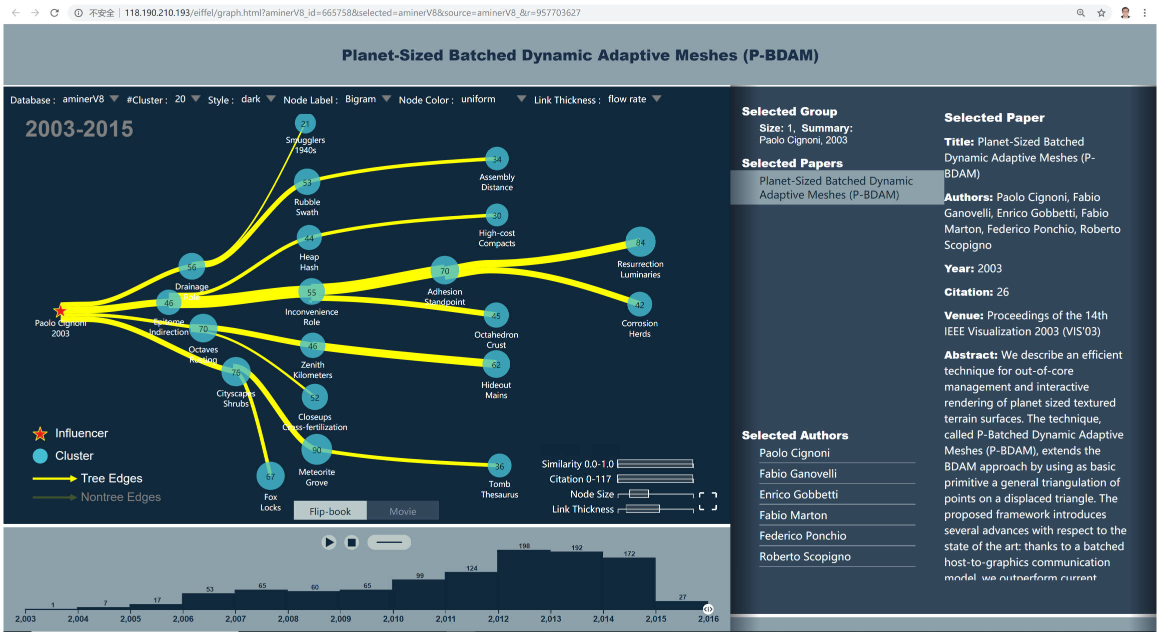
[7] M. Bostock, V. Ogievetsky, and J. Heer, "D3: data-driven documents," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2301–2309, 2011.



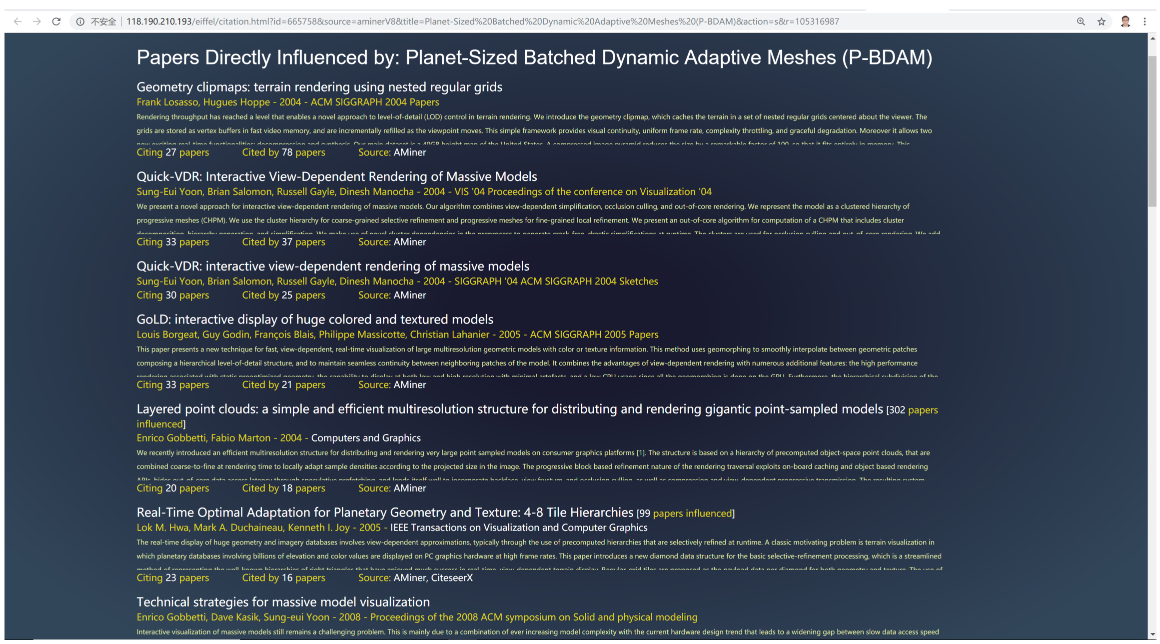
(a)

(b)

Fig. 3: Visualization of the influence graph of “Graph Visualization and Navigation in Information Visualization: A Survey” by: (a) Eiffel (EI); (b) GS (GI).



(a)



(b)

Fig. 4: Visualization of the influence graph of "Planet-Sized Batched Dynamic Adaptive Meshes (P-BDAM)" by: (a) Eiffel (EII); (b) GS (GII).