

# Analytics and Visualization of Citation Network Applying Topic-based Clustering

Rina Nakazawa · Takayuki Itoh ·  
Takafumi Saito

Received: date / Accepted: date

**Abstract** Survey of papers is not an easy task for novice researchers because it may happen that they miss appropriate keywords for their survey. It often takes a long time for young researchers to find research papers even when they use famous search engines like Google Scholar. In addition, they may not be familiar with understanding positions of papers in their research fields smoothly. To resolve this problem, many researchers have studied citation network visualization techniques for surveying papers. However, it is still often difficult to observe the complicated relations across multiple research fields or traverse the entire relations in their interest. Additional clues as well as a citation network are therefore important for survey of papers. In this paper, we proposed a visualization technique for citation networks applying a topic-based paper clustering. Our technique categorizes papers applying LDA (Latent Dirichlet Allocation), and constructs clustered networks consisting of the papers. We applied the technique to three datasets. The results of our visualization technique demonstrated that the proposed technique could contribute to help users to understand the positions of papers in the research fields. We conducted subjective evaluation comparing with time-oriented technique, and demonstrated that our technique was more helpful for novice researchers like students to find papers.

**Keywords** Visualization · Citation network · Edge bundling · Reference · Topic-based clustering

## 1 Introduction

Finding research papers is a very important task to understand trends of research fields and find related papers. Researchers use text-based portal Web

---

Rina Nakazawa · Takayuki Itoh  
2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610 Japan  
E-mail: leena,itot @itolab.is.ocha.ac.jp

sites such as Google Scholar [22], ACM Digital Library [21], and IEEE Explore Digital Library [23]. Researchers look up for the references of papers they have read. However, it is not always easy for novice researchers to survey papers they want to read and instantly understand positions of the papers in the research fields with their search results. Moreover, young researchers may miss papers in case they do not find the appropriate keywords, or in case papers they really want to survey straddle multiple research fields. We define “keywords” in this paper as the terms which consist of a topic not terms which the authors annotate. A topic includes multiple keywords.

There have been many studies on visualization of citation networks, including Mackinlay et al. [12] and Small et al. [17], which aimed to alleviate these difficulties. However, we suppose still there are many open problems on visualization of citation networks. For example, researchers continuously trigger for new fusions of multiple fields, and therefore they need to organize and understand the relations of papers that cover multiple research fields. Another problem is while surveying papers in unfamiliar research fields. Papers in the unfamiliar research fields sometimes do not include the terms well-used in a research field which the users are familiar with. Conversely, terms may have very different meanings depending on the research field. In such cases, we find that the papers are not what we expected after we read them. Understanding the positions of the papers in the research fields is important for researchers to identify whether the papers are related to the topics which they want to survey.

To organize these open problems, we define the requirements in visualization of citation networks for survey of papers as follows:

- R1: Find much-cited papers which include user-specified topics.
- R2: Find papers whose contents are similar to the focused papers, such as the papers which do not contain the user-specified keywords but belong to the user-interested topics.
- R3: Find the contents of papers which have citation relations between the papers using user-specified keywords or belonging to user-interested topics.
- R4: Find tightly related pairs of topics.

We propose a visualization technique that satisfies these requirements by implementing the following solutions.

- S1: Categorize papers that have similar topics to the same group. (Section 3.1)
- S2: Place papers that belong to the category in a circular region. (Section 3.2)
- S3: Place citing and cited papers closer by applying a force-directed layout algorithm. (Section 3.2)
- S4: Summarize citation relations by applying an edge bundling method. (Section 3.2)

Solutions S1 and S2 satisfy R1 and R2 because users can easily find papers belonging to user-interested topics by looking at the particular circular regions. Solution S3 satisfies R3 because users can follow the citation relations between closely placed papers. Solution S4 satisfies R4 because bundled edges effectively represent tight relations between pairs of topics.

We applied the proposed technique to the datasets describing the citations of papers published in ACM SIGGRAPH, IEEE Transactions on Visualization and Computer Graphics, and IEEE Computer Graphics and Applications. This paper introduces the visualization results with these datasets and discusses the effectiveness of our technique.

Our technique applies a general purpose graph layout technique, against many studies on visualization of citation relations apply time-oriented visualization design. We aimed to represent the positions of user-interested papers in a set of topics, and relations between pairs of topics. We therefore applied our own algorithm on clustered graph layout [9, ?], not time-oriented visualization. The paper presents our subjective evaluation comparing with an existing time-ordered technique. We asked the questionnaires to the 21 graduated students as an example of novice researchers. We have presented the result of our technique [15], we introduce a technique to apply the dataset of multiple conferences in addition.

Our contributions of this paper are as follows:

- Proposed a technique to visualize citation network based on their topics.
- Demonstrated that our technique could be helpful to grasp the positions of papers in research fields with the visualization results.
- Compared with the time-ordered visualization technique.

Our technique would help novice researchers to understand the differences between the tendencies of similar research fields.

## 2 Related Work

This section introduces existing visualization techniques for topics or keywords of text corpus and science literature. There have been many visualization techniques for topics in text corpus [11], [18]. Some of these techniques focus on the topics of science literature. Lee et al. presented PaperLens [10] as a visualization technique that firstly applies the mixture distribution model to the titles and keywords, then estimates their topics, and finally shows papers by topics and publication years. Shahaf et al. [16] introduced their technique which visualizes the relationships among terms looking like metro maps. As another examples of the analyses for topics of publications, Henry et al. [8] visualized networks of well-used keywords and co-authors, and analyzed features of multiple conferences. CiteRivers [7] visualized the trend of the topics mentioned in the papers and the number of papers for each conference or journal by citation flow. Users can easily understand the trend of the topic they focus on and which conference or journal is related to it. These techniques support to understand the transition and trend of the research topics in a certain conference or a journal. However, these techniques just visualize topics mentioned in the papers; they do not support citation relations. It is difficult for these existing techniques to navigate to the related papers that are not directly related to the user-interested topics but cited by papers of the user-interested

topics. Therefore, they do not satisfy the requirements mentioned in Section 1.

Many researchers have been studied visualization of citation networks. Citation network topology [2], [3] is one of these approaches to visualize citation patterns. Brandes et al. [2] presented a visualization technique for citation networks with topographic maps that places the hub papers cited by many papers higher than the other papers. It also arranges the papers that have similar citation pattern closer. That enables us to easily find the hub papers and the groups of papers that have similar citation patterns. These techniques assist users to easily find well-cited papers and understand citation patterns. However, it is difficult to find papers by only applying the citation topology when users do not know appropriate papers for a clue to track the citation relations.

Many researchers have also studied time-ordered visualization techniques [13], [19], [20] for citation network. CiteNetExplorer[20] applies transitive reduction of citation network and put them in chronological order. It assigns colors of nodes which denote publications to the attributes like successor and predecessor. Citeology [13] orders papers based on the numbers of their citations with respect to each year, and places them from the center of a display space. It can visualize up to eight hops of the citations. This study represents structures of citation networks by placing nodes corresponding to papers in the time-series order. When a citation network has complicated relations across multiple research fields, it causes serious edge crossing and cluttering which bring bad impact on readability. Visualization results with heavy cluttering prevent the users from finding the positions of papers, while the users want to understand the positions of the interested papers in the research fields. New papers always cite the older papers, so we do not need to apply time-ordered visualization design for citation network when we show direction of the citations. CiteVis[19] visualizes citation relations by highlighting the citing nodes and the cited nodes when a user clicks a node. This technique reduces a visual clutter of edges and enables users to understand the citation relations of the clicked paper and the trend of the number of presented papers in a conference.

One of our goals is to help novice researchers to survey papers that they want to read and understand the positions of the papers in the research fields with search results instantly. To achieve this goal, we think both of the citation relations and the topics of the papers are important as described in the previous section. Therefore, we propose a visualization technique that concerned both of the citation relations and the topics of the papers. Still small number of studies have addressed both of them. Dunne et al. [5] proposed an integrated visualization of citation network and paper summary description. Users can simultaneously look at the citations, ranking based on the citation count, and summary description of papers in the cluster generated by graph clustering based on citation structure. This representation has a bottleneck that it may require larger display spaces. Also, the network visualization shows only papers extracted by the keyword-based search. Therefore, users may miss papers

if these papers do not use the user-specified keywords or they are not cited by papers using such keywords.

Though these novel visualization techniques have been presented, it is not still always easy to find important papers by using such techniques. One of the reasons is that these existing techniques often require users to manually specify the papers whose citations they want to figure out. It often happens that novice researchers do not know all the appropriate keywords, and therefore it is not easy for them to determine which papers they should read. The second reason is that it may happen to miss the papers which do not have citation relations but have similar contents when we only focus on citation relations to find papers. The third reason is that many recent new research fields have triggered as fusions of multiple research fields. Researchers need to organizationally understand the relations of papers that cover such multiple fields along their fusion. However, there seems no visualization techniques addressing this problem.

### 3 Proposed Technique

This section describes the processing flow of the presented technique. We treat the papers as nodes, and citations as directed edges of a network. The technique classifies the papers based on their contents to construct a hierarchical network. The technique then applies our hierarchical network layout technique with an edge bundling algorithm. Our implementation also provides rendering and interaction techniques.

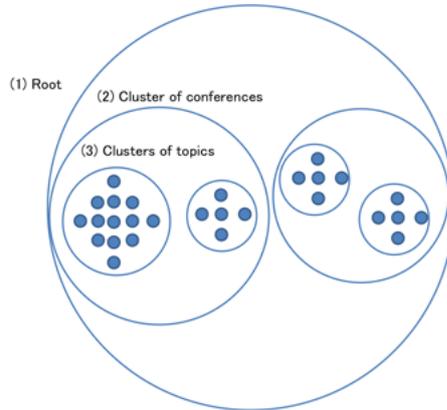
#### 3.1 Clustering Papers

The proposed technique applies LDA (Latent Dirichlet Allocation) [1] to categorize papers based on the contents of papers. LDA is a generative topic model which allows a document to include various topics. It is generally used because it can avoid overfitting the data. As a paper can include multiple topics, we think LDA is appropriate for our purpose. It could solve the problem to categorize papers that straddle multiple research fields. The technique applies LDA to the set of all the paper abstracts to estimate topics and calculate the topic distribution for each abstract. LDA needs to be given the number of topics, so we determine the number heuristically. We regard these topics as research fields and categorize all papers based on them. The technique supposes a paper is related to the particular topic, if a value of the topic distribution is larger than the threshold. We removed unnecessary words from the abstract as a preprocessing to improve the quality of clustering results. The removed words included non-important words such as prepositions, or too frequently used terms such as “propose” and “technique.” Then, we presumed the contents of the topic from 20 words whose probability is highest in the topic. Our clustering allows a paper to belong to multiple topics. A cluster in this paper

can include multiple topics. For example, there is a cluster which including only topic A whereas papers about topic A and topic B belong to another cluster.

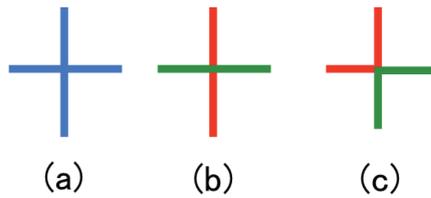
### 3.2 Network layout

Next, our technique arranges nodes applying a hybrid force-directed and space-filling graph drawing algorithm [9] to calculate the positions of nodes corresponding to the individual papers. We define a dataset of papers has a hierarchical structure of conferences and topics. In this paper, conferences are upper-level clusters and topic clusters are lower-level ones to show the trend of topics and citation relationships in conferences.



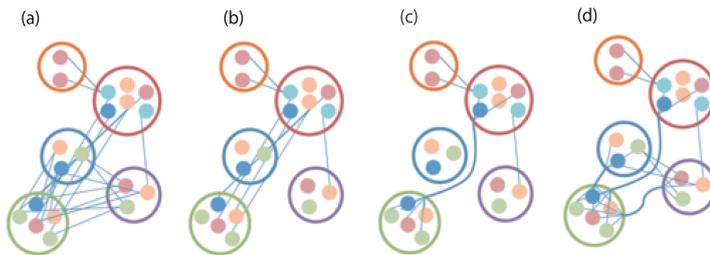
**Fig. 1** Space-filling hierarchy layout for our system. The algorithm calculates positions of nodes or clusters from (3) the lowest level clusters to (1) the top-level of the dataset.

We calculate positions of papers under topic clusters shown as (3) in Figure 1, position of topic clusters under conference clusters shown as (2) in Figure 1, and position of conference clusters under the root node using Itoh’s algorithm. The technique displays the nodes supposing that their sizes are proportional to the number of citations. The force-directed algorithm enables to place papers that belong to the same research category closely, and also, papers that have citation relations closely. Then, the space-filling algorithm enables to avoid the node cluttering and improve the display space utilization. After the above process, the technique summarizes the edges corresponding to citations by applying an edge bundling algorithm. Our implementation of the edge bundling enables users to adjust the threshold controlling whether it bundles the edges or not. We have already implemented the edge bundling algorithm in our previous work [14]; however, it had a problem that straight bundles with which summarizes a lot of edges may lead to misinterpretations of Gestalt principle as Figure 2 shows when the bundles avoid nodes and bend at a right angle.



**Fig. 2** A misconception of Gestalt principle (a) The appearance of two bundles. (b) It usually looks like two bundles crossing. (c) The two bundles actually bent at a right angle.

To prevent the misconceptions, we place nodes in circular and bundle the citation edges with Catmull-Rom spline curves (Figure 3). Our technique firstly calculates the shapes of all bundle paths so that they do not overlap the node clusters. According to the threshold the user sets, the technique determines whether the number of the edges of one cluster with the others is larger than the threshold like Figure 3-(b). Then, it bundles the edges only when the number of edges between two clusters is larger than the threshold. (Figure 3-(c)). The technique applies this process to all pairs of the node clusters (Figure 3-(d)).



**Fig. 3** The processes of edge bundling: (a) Calculate the shapes of all bundle paths. (b) Count the number of edges between two clusters. (c) Bundle the edges only when the number of edges between two clusters is larger than the threshold. (d) Apply the processes (b)(c) to all pairs of the node clusters.

### 3.3 Color scaling for network rendering

Since citation networks have directionality so-called “cited” and “citing,” our technique draws the cited side of the edges in bright pink, and the citing side of the edges in dark pink, to represent the directionality of the edges. When a user click a node without edge bundling technique, the system highlight the links of the clicked node with blue or green color. We chose the colors based on the following requirements, so the other colors are also appropriate if they satisfy them.

- Use the color except pink to compare the edges of the clicked node and the other edges

- Use two different colors to compare the nodes which are clicked at the first time and the next time

We can also draw arrows or assign different hues to the each side of the edges for the representation of directionality of the edges. However, these representations are not always adequate for large-scale networks and networks in which there are many hubs. When we represent the direction by arrows, heavy cluttering may happen around hub nodes or dense regions, which would degrade the readability. Besides, we assign hues to the nodes, and our technique controls brightness to represent the edge direction. As Figure 4 shows, we draw nodes with the color scale corresponding to the publication years.

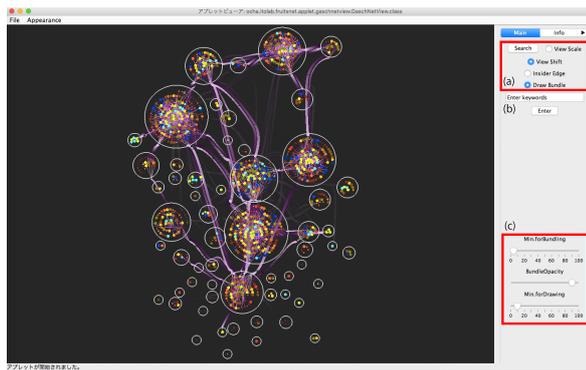


**Fig. 4** Color scaling (a): The node color, (b): The edge color.

### 3.4 User Interface

Figure 5 is a snapshot of the user interface we implemented. The left side of the window features the drawing space, while the right side features two tabs. One of the tabs features various GUI widgets. Users can scale and shift the view, switch the edge bundling mode, and set its threshold, by using the GUI widgets shown in Figure 5 (a)(c). When a user clicks a node corresponding to a particular paper, the technique displays the details of the paper such as the digital object identifier (DOI), title, authors, year, and abstract, on the panel featured by the other tab. At the same time, it highlights the edges of the clicked node, and those of the nodes that are connected to the clicked node. This edge highlight function is applicable to two nodes together, and this enables to compare the citations of each paper.

By the way, it is not always easy for the novice researchers to find the paper that they should read first, just by observing the citation networks. Such users can filter papers on the display by selecting a research category or entering a keyword. When the user enters a keyword in the text input widget shown in Figure 5 (b), the technique displays only the papers whose titles include the keyword. Also, When selecting a research category that the user is interested in, the node cluster that has only the research category is magnified in the center of the display in Figure 6 and Figure 7. It is useful to firstly overview, and then narrow down the focus cluster by selecting a category or entering a keyword when users want to survey whole contents of the conference or research fields. Users can track bundles of the focus cluster, and then move to focus on other clusters. In case that users want to look into respective papers, they can also narrow down the focus paper in the same procedure. If users click a paper node, its citation edges are highlighted. Users can follow



**Fig. 5** User Interface (a) scale and shift the view, switch the edge bundling mode (b) enter the keyword to display only the papers whose titles include it (c) set its threshold

these edges and trace them. VIGOR[6] is an integrated visualization technique to show the query, the result of joining all query matches, subgraphs of the search result with clustering, and summarization of each cluster’s feature distributions. We think visualizing the words in the user-focused cluster by tag cloud or visualizing its feature distribution will also enable novice researchers to understand papers and topics for their survey.

## 4 Results

We implemented the proposed technique with Java Development Kit (JDK) 1.6.0. In this section, we show some results visualizing citation networks.

### 4.1 An example of a conference proceeding

We applied a citation network dataset consisting of 1072 full papers published in the SIGGRAPH conferences during 1990 to 1994, and during 2000 to 2010, provided by the ACM Digital Library [21]. We extracted the title, publication year, abstract, references, and authors from html files of the papers. We did not apply the paper information during 1995 to 1999, because we could not extract the abstracts from ACM Digital Library.

#### 4.1.1 Example of hardware

Suppose that a user surveys for research papers on “hardware and GPU.” Figure 6 is an example when the user selected the “hardware and GPU” category. We could observe that the cluster in the center contained papers categorized only to “hardware and GPU” had dense relationships between the “physical simulation”, “lighting”, and “shape modeling” categories. We also found that the cited bundles of the “hardware and GPU” cluster are thicker than the

citing ones, which means many papers in these research fields “physical simulation,” “lighting,” and “shape modeling” refer to the papers in the “hardware and GPU” cluster, and the researches in these fields have often evolved based on the researches in the “hardware and GPU” category. Among these relationships, especially, the relation between the “hardware and GPU” and “lighting” clusters clearly shows the above fact. Therefore, we expect that the “hardware and GPU” cluster could give a clue to the research team that develops hardware systems when they want to know which research fields their products are well applied.

#### 4.1.2 Example of lighting and CG algorithm

Next, we supposed that a user searched for papers related to lighting. Figure 7 shows an example of visualization under this supposition. The cluster A is a group that categorized into “lighting & CG (Computer Graphics) algorithm.” We found the nodes in this cluster were colored in light blue or yellow-green, where the colors depicted that the papers corresponding to the nodes were published in 1994 and 2000. Although this cluster is small, the problems in this research field were addressed once in 1994 and discussed again in 2000. The papers in the cluster A are as follows:

- A fast shadow algorithm for area light sources using back projection (in 1994)
- The irradiance Jacobian for partially occluded polyhedral sources (in 1994)
- A clustering algorithm for radiosity in complex environments (in 1994)
- Illuminating micro geometry based on precomputed visibility (in 2000)
- Efficient image-based methods for rendering soft shadows (in 2000)
- Conservative volumetric visibility with occluder fusion (in 2000)

We especially suppose the above papers published in 2000 might have triggered the invention of PRT (Precomputed Radiance Transfer).

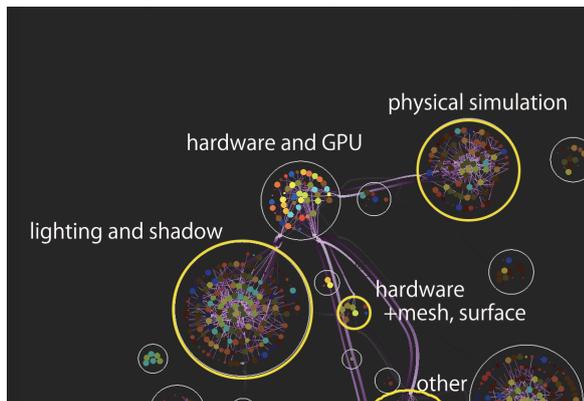


Fig. 6 Example of hardware

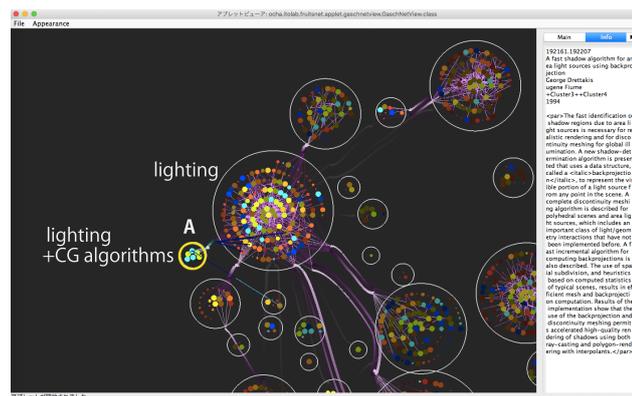


Fig. 7 Example of lighting and CG algorithm

#### 4.1.3 Example with a keyword

Figure 8 (a) shows an example that a user entered the keyword “skin”. When the user did not apply the edge bundling and clicked the two orange nodes, many edges are drawn as shown in Figure 8 (b). The technique highlights the edges connected to the clicked nodes and the citations of the cited and citing nodes. Figure 8 (a) demonstrates that we can classify the research papers whose titles contain the term “skin” into two research fields. Therefore, we clicked two orange nodes, one in a larger cluster, and the other categorized in the different cluster far from the first one. As a result, we could grasp the two streams containing each of the clicked nodes, because all the displayed nodes in Figure 8 (b) connect with either blue or green edges. We listed all the titles and figures (see Figure 9 and Figure 10 ) of the papers classified into these two groups. The papers connected with green edges as follows:

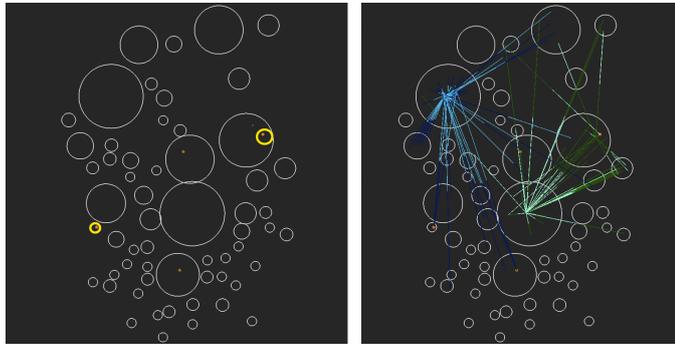
- (a) Continuous capture of skin deformation [P. Sand et. al., 03] (in 2003)
- (b) Building efficient, accurate character skins from examples [A. Mohr et. al., 03] (in 2003)
- (c) Capturing and animating skin deformation in human motion [S. II. Park et. al., 06] (in 2006)
- (d) Data-driven modeling of skin and muscle deformation [S. II. Park et. al., 08] (in 2008)

We listed the papers that belong to the blue stream.

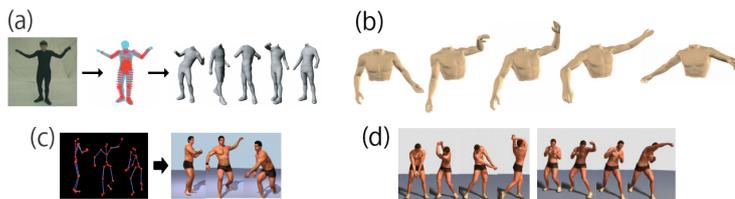
- (e) Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin [N. Tsumura et. al., 03] (in 2003)
- (f) Analysis of human faces using a measurement-based skin reflectance model [T. Weyrich et. al., 06] (in 2006)

As we could understand from these pictures of the papers, our technique demonstrated that researches of SIGGRAPH related to “skin” could be divided into two groups, based on their topics and citations. One of the topics is

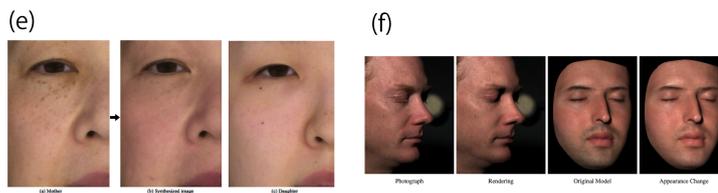
related to human animation generation using motion capture systems, and the other discusses generation or analysis of human face skins. This result demonstrates that the technique enables the novice researchers, who study computer graphics and want to read papers related to skin, to understand that there are two research fields related to skin and to choose which field they should survey.



**Fig. 8** (a) Result with a keyword “skin,” (b) Result when an user click two nodes. The edge bundling is not applied



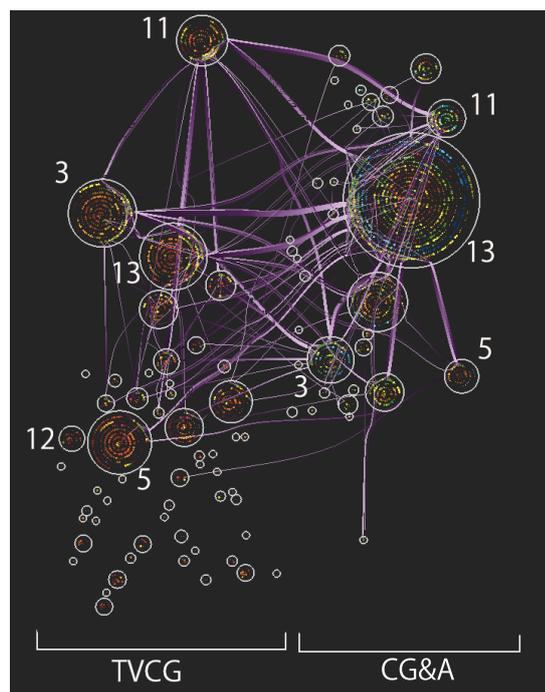
**Fig. 9** Pictures in papers of the green stream. (a) [P. Sand et. al., 03] (b) [A. Mohr et. al., 03] (c) [S. II. Park et. al., 06] (d) [S. II. Park et. al., 08]



**Fig. 10** Pictures in papers of the blue stream. (e) [N. Tsumura et. al., 03] (f) [T. Weyrich et. al., 06]

## 4.2 Examples of journals

As an example of multiple journals, we applied a citation network dataset consisting of 3604 full papers published in IEEE Transactions on Visualization and Computer Graphics (TVCG) and IEEE Computer Graphics and Applications (CG&A) from 1981 to 2015 provided by the IEEE Explore Digital Library [23]. We removed papers whose abstract cannot be provided in the library from the dataset. Figure 11 shows a visualization result of our technique. The papers of TVCG are placed on the left side of Figure 11 and the papers of CG&A are placed on the right side of Figure 11. This result shows that there are many citation relationships between TVCG and CG&A.



**Fig. 11** Example of IEEE Transactions on Visualization and Computer Graphics (TVCG) and IEEE Computer Graphics and Applications (CG&A). The numbers correspond to the itemization of the topics.

The following is a list of topics of the papers published by TVCG and CG&A.

1. VR and AR
2. simulation
3. geometry and modeling
4. animation and motion
5. interactive system

6. lighting and rendering
7. volume rendering
8. software and environment
9. art
10. color and projection
11. GPU and hardware
12. graph visualization
13. others

As the trend of topics in TVCG and CG&A, papers which include only topic 12 are published in TVCG, but not much published in CG&A. The papers in cluster 12 are colored in brown and red. They do not cite papers in CG&A. This explains that CG&A would not be helpful for users who are interested in "graph visualization." Then, we introduce another feature between TVCG and CG&A. Figure 11 shows that papers in the clusters which include only topic 3 or 11 have dense relationships between the two journals. They are related to "geometry and modeling" and "GPU and hardware." The papers of these two topics cite each other. This result would give a clue to find papers about the topics as the visualization result of SIGGRAPH dataset shown in the previous section. On the other hand, papers in the clusters which include only topic 5 do not cite each other much between the two journals. However, the papers in these clusters have some common relationships to papers in the largest cluster of CG&A. The papers in the clusters 13 are not categorized in any of the above 12 topics. In this way, our proposed technique can help to observe the trend of papers published in multiple journals.

## 5 Evaluation

### 5.1 Preliminary questionnaires

There have been a lot of citation visualization techniques as we mentioned in Section 2. As against our technique applies a general purpose graph layout technique, typical existing techniques places nodes corresponding to the papers in time-ordered. We assume the time-ordered layout policy is not mandatory, since it is sufficient for many users to recognize each of the visualized papers are old or new. For example, we often just want to know whether the paper is the oldest one as the roots in the research field, or the newest one. To prove our hypothesis, we conducted the subjective evaluation to compare our technique and the time-oriented visualization technique. Before the evaluation, we had a questionnaire to define what we carefully observe while surveying papers. We asked three questions to ten graduated students majoring computer science.

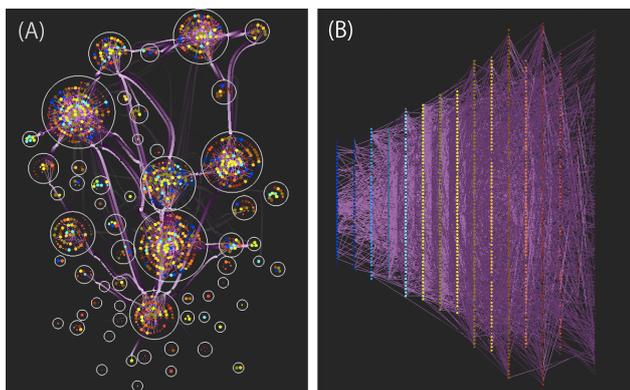
1. What do you want to know when you search for papers?
2. What technique do you want for surveying papers well?
3. What do you want to know if you look into the citation network visualization in a particular conference for twenty years?

Regarding the question 1, a half of the students answered that they would like to know whether the papers are similar to their researches. In other words, it is important to define criteria of similarity of research topics and papers. Other answers are regarding citations and research topics or fields of papers. These answers suggests the usefulness of visualizing topic-based structures of papers and citations. We also suppose the structures of topics and citations can be used to determine the similarity among papers. Several students answered they wanted to know the differences (e.g. advantages and disadvantages) among the techniques presented in the papers. We would like to solve this issue as a future work because both our technique and the existing techniques cannot represent the concrete contents only as the visualization results.

Regarding the question 2, more than half of the students mentioned that word-based smart search techniques are important for paper survey processes, including synonym recommendation and search refinement. This result proves that novice researchers including graduated students had troubles while selecting keywords to search for papers. Regarding the question 3, we roughly divide the answers into three categories, “the transition of research fields”, “the citation relations”, and “both research fields and citation relations, or what they reveal in combination”. It demonstrates the demands to understand both research fields and citations.

## 5.2 Evaluation: comparison with time-oriented visual representation

According to the result of the questionnaire, we asked 21 graduate students majoring computer science to compare our technique shown in Figure 12 (A) with the time-oriented citation visualization shown in Figure 12 (B), and evaluate which visualization is proper to know the contents below. We implemented the time-oriented technique mimicking Citeology.



**Fig. 12** (A) Our technique (B) Time-oriented technique

We asked participants to answer the questions as 5-level scores, where 1 represents a strong agreement with A, and 5 represents a strong agreement with B. The following are the contents for which we asked to participants which visualization is more suitable:

1. The transition of papers amount published in the conference every year.
2. The main topic in the conference.
3. The trend of a research topic by year.
4. The research fields that seem to have a strong relationship with a field you focus on.
5. Much-cited papers on a certain topic.
6. The latest paper on a certain topic.
7. The content trends of papers citing the paper you read (or clicked).
8. Papers whose contents are similar to the paper you read (or clicked).
9. Papers that had a great influence on the paper you read (or clicked).

Figure 13 shows the evaluation result. The x-axis denotes the sequential number of questions, and the y-axis denotes the quantity of responses. Our technique was evaluated as more beneficial in the questions 2, 4, 5, 7, and 8, while the time-oriented visualization B was evaluated as more effective in the questions 1, 6, and 9. The questions 4, 5, 7, 8 correspond to the requirements R1-4. Therefore, the result of this evaluation demonstrated that our technique satisfy the requirements.

Although we expected the time-oriented technique B has an advantage on the questions 3 and 9, the figure demonstrates their rates varied widely. The result denotes our technique are also effective for the questions 3 and 9. Especially, the rate of the questions 9 resulted in the variation because we did not need to know the publication year strictly to distinguish papers that had the great influence. This result proves that we do not need to assign the publication year to the x-axis of the display space.

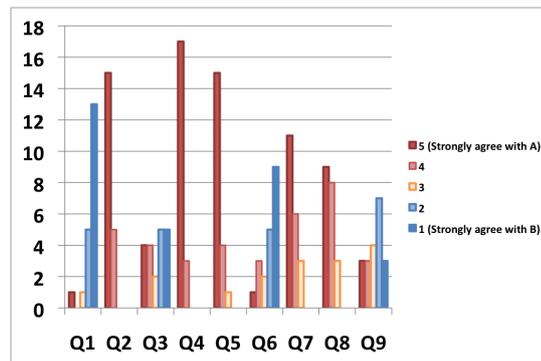


Fig. 13 Result of the evaluation

## 6 Conclusions

We presented a visualization technique of citation networks for survey of research papers, and discussed the visualization results. Our technique applies topic-based paper clustering to construct hierarchical network. It then applies a hybrid force-directed and space-filling network layout algorithm, and an edge bundling technique with Catmull-Rom spline curve. Our GUI design realizes the requirement R1, as a function of topic selection filtering. We applied datasets of publication of ACM SIGGRAPH, IEEE Transactions on Computer Graphics and Visualization, and IEEE Computer Graphics and Applications. These results showed our technique could help to understand the positions of papers in research fields and find papers even when users do not know all appropriate keywords. Also, our technique could show the trends of topics and citation relations in a particular conference or journal. The case of visualization with a keyword “skin” demonstrated our technique satisfies the requirements R2 and R3 and the case of hardware topic of ACM SIGGRAPH showed for the requirement R4. This paper also introduced the results of the user evaluation comparing with a time-oriented visualization technique. The result demonstrated that our technique was more helpful for novice researchers like students to find papers. For future work, we think combining paper information and author information like PivotPaths [4] would be helpful for survey of papers.

## References

1. D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation.” *Journal of Machine Learning Research*, 3, pp. 993-1022, 2003.
2. U. Brandes, T. Willhalm, “Visualization of bibliographic networks with a reshaped landscape metaphor.” In *Proceedings of the symposium on Data Visualisation 2002*, pp.159-164, 2002.
3. C. Chen, “CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature.” *Journal of the American Society for Information Science and Technology*, 57(3), 359-377, 2006.
4. M. Dörk, N. H. Riche, G. Ramos, and S. Dumais, “PivotPaths: Strolling through Faceted Information Spaces.” *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2709-2718, 2012.
5. C. Dunne, B. Shneiderman, R. Gove, J. Klavans, B. Dorr, “Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization.” *Journal of the American Society for Information Science and Technology*, 63(12), pp. 2351-2369, 2012.
6. R. Pienta, F. Hohman, A. Endert, A. Tamersoy, K. Roundy, C. Gates, and D. H. Chau, “VIGOR: Interactive Visual Exploration of Graph Query Results.” *IEEE Transactions on Visualization and Computer Graphics*, 2017.
7. F. Heimerl, Q. Han, S. Koch, and T. Ertl, “CiteRivers: visual analytics of citation patterns.” *IEEE transactions on visualization and computer graphics*, 22(1), pp. 190-199, 2016.
8. N. Henry, H. Goodell, N. Elmqvist, J. D. Fekete, “20 years of four HCI conferences: A visual exploration.” *International Journal of Human-Computer Interaction*, 23(3), pp. 239-285, 2007.
9. T. Itoh, C. Muelder, K. Ma, J. Sese, “A Hybrid space-filling and force-directed layout method for visualizing multiple-category graphs.” *IEEE Pacific Visualization Symposium*, pp. 121-128, 2009 ☒.

10. B. Lee, M. Czerwinski, G. Robertson, B.B. Bederson, "Understanding research trends in conferences using PaperLens." CHI'05 extended abstracts on Human factors in computing systems, pp. 1969-1972, 2005.
11. S. Liu, M. X. Zhou, S.Pan, Y.Song, W. Qian, W. Cai, and X. Lian, "TIARA: Interactive, Topic-Based Visual Text Summarization and Analysis." ACM Trans. Intell. Syst. Technol. 3, 2, Article 25 (February 2012), 28 pages, 2012.
12. J. D. Mackinlay, R. Rao, S. K. Card, "An organic user interface for searching citation links." the SIGCHI conference on Human factors in computing systems, pp. 67-73, 1995.
13. J. Matejka, T. Grossman, and G. Fitzmaurice, "Citeology: visualizing paper genealogy." CHI'12 Extended Abstracts on Human Factors in Computing Systems, pp. 181-190, 2012.
14. R. Nakazawa, T. Itoh, J. Sese, A. Terada, "Integrated visualization of gene network and ontology applying a hierarchical graph visualization technique." 16th International Conference on Information Visualization (iV), pp. 81-86, 2012.
15. R. Nakazawa, T. Itoh, and T. Saito, "A visualization of research papers based on the topics and citation network." In 19th International Conference on Information Visualisation (iV), pp. 283-289, 2015.
16. D. Shahaf, C. Guestrin, E. Horvitz, "Metro maps of science." In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1122-1130, 2012.
17. H. Small, "Visualizing science by citation mapping." Journal of the American society for Information Science, 50(9), pp. 799-813, 1999.
18. J. Stasko, G. Carsten, and L. Zhicheng, "Jigsaw: supporting investigative analysis through interactive visualization." Information Visualization 7, 2 (April 2008), pp. 118-132, 2008.
19. J. Stasko, J. Choo, Y. Han, M. Hu, H. Pileggi, R. Sadana, and C. D. Stolper, "Citevis: Exploring conference paper citation data visually." Posters of IEEE InfoVis, 2013.
20. N. J. Van Eck, L. Waltman, "CitNetExplorer: A new software and tool for analyzing and visualizing citation networks." Journal of Informetrics, 8(4), pp. 802-823, 2014.
21. ACM Digital Library, <http://dl.acm.org/>
22. Google Scholar, <https://scholar.google.co.jp/>
23. IEEE Xplore Digital Library, <http://ieeexplore.ieee.org/>