Co-Authorship Networks Visualization System for Supporting Survey of Researchers' Future Activities

Takeshi Kurosawa, Yasufumi Takama
Graduate School of System Design, Tokyo Metropolitan University, Hino, Tokyo, Japan
Email: kurosawa@krectmt3.sd.tmu.ac.jp, ytakama@sd.tmu.ac.jp

Abstract— **This paper proposes a visualization system that supports users getting insight into future research activities from co-authorship networks. A bibliographic network such as a co-authorship network and a citation network is important information for researchers when doing a research survey. In particular, there are many requests on research survey that relate with researchers' future activities, such as identification of remarkable researchers including growing researchers and supervisors. Although a citation network has received many attentions from researchers, it is not suitable for such surveys because it reflects researchers' past activities. Since collaboration of researchers is essential for researchers' activities, co-authorship network is supposed to be suitable for predicting future activities. In order to get insights into future research activities by discriminating growing research areas from grown-up areas, the proposed visualization system provides the functions for identifying research areas as well as for identifying time variation of both network structure and keyword distribution. As a basis for getting insights into future research activities, this paper focuses on the task of discriminating growing researchers from supervisors. The effectiveness of the proposed system is evaluated through the detailed analysis of two participants' analyzing process of InfoVis 2004 Contest dataset. It is observed that different analyzing strategies are employed by even the same participant, when available support functions are different. The result indicates participants can successfully utilize the functions in their exploratory analysis process.**

Index Terms— **exploratory data analysis; interactive information visualization; temporal trend information; co-authorship networks; graph visualization;**

I. Introduction

This paper proposes a visualization system for co-authorship networks that has functionalities for supporting the prediction of future research activities. A bibliographic network is important information for researchers when doing a research survey. A bibliographic network is composed of several networks: a co-authorship network, a citation network, and a co-citation network. Today, there are a number of research bibliography sites on the web such as IEEE Xplore[1] and DBLP[2]. These sites provide detailed information about specific authors, papers and journals, which are useful for a research survey. Research surveys are sometimes conducted by researchers who are getting into their unfamiliar research field. In such a case, the purpose of the survey is to grasp an overview of the research field. However, as the structure of bibliographic networks is usually huge and complicated, it is difficult to grasp such an overview using simple interface of existing research bibliography sites. Therefore, information visualization techniques for bibliographic networks have been studied [1]–[6].

In the field of bibliographic network analysis, a citation network has received many attentions from researchers [1]. There also exist many visualization systems aiming at supporting user's research survey based on it [2]. A citation network represents researchers' past activities, from which we can identify important papers by finding the most frequently cited papers.

On the other hand, there are many requests on research survey that relate with researchers' future activities, such as the identification of researchers who will potentially write an interesting paper. As the direction of citation is from new paper to old one, it is difficult to predict such future activities from a citation network.

It is noted that collaboration of researchers is essential for these research surveys. That is, a research paper is the outcome of collaborative activities, and research areas are often emerged from collaboration among researchers working on different research topics. In that sense, a co-authorship network is suitable for predicting researchers' future activities, because it reflects the past and current status of collaboration.

This paper proposes a visualization system for co-authorship networks that supports users getting insight into future research activities. As for remarkable researchers to be identified for a research survey, this paper focuses on two types of researchers: growing researchers and supervisors. It is noted that this paper defines a supervisor as a researcher who has already achieved in his/her research fields. To identify growing researchers and supervisors, the proposed system provides two functions: the function for identifying research areas and that for identifying time variation of both network structure and keyword distribution. To identify research areas, the proposed system renders a researcher as a pie chart, which represents a ratio of keywords assigned to corresponding researcher's papers. By representing a node in the co-authorship network as a pie-chart, keyword distribution

[1] http://ieeexplore.ieee.org/
[2] http://www.informatik.uni-trier.de/~ley/db/

© 2012 ACADEMY PUBLISHER
doi:10.4304/jetwi.4.1.3-14

over the network is easily grasped by analysts.

Identifying time variation is composed of three sub functions. First, to determinate whether a researcher published papers in a period or not, the system uses animation. Second, the brightness and saturation of a segment in a pie chart represent the period when corresponding keyword is used. Finally, to determinate whether the researcher's activity is continuously or sporadically, a pie chart of a researcher can be bi-cylindrical style, in which inner and outer ring correspond to the earliest and the most recent year of his/her publications, respectively.

In order to evaluate the effectiveness of the system, the proposed system is applied to exploratory analysis of the co-authorship network extracted from the InfoVis 2004 Contest dataset. Two test participants used the system for identifying growing researchers and supervisors. Results are analyzed from the viewpoint of analyzing strategies they employed in performing tasks. Two types of systems, each of which provides different sets of support functions, are used in a user study. By comparing the results, how available functions affect participants' analyzing strategies is investigated.

The results show the participants who don't have background knowledge about InfoVis can identify growing researchers and supervisors using the system. It is observed that the proposed functions such as highlighting researchers who published in a certain period and visualizing a node as a pie chart of keywords are heavily used.

The content of the paper is extended from our previous works [7], [8]. Main difference from [7] is a result of comparing user behaviors with using two system having different functions. Expansion from [8] has been made according to the discussion at the conference, including more detailed description about the system and discussion about the result of user study.

The remainder of this paper is organized as follows: Section II discusses related works. In Section III we introduce the proposed visualization system. The effectiveness of the proposed system is shown with a user study, which is described in section IV and V.

II. Related works

A. Visual Analysis of Bibliographic Networks

Bibliographic information is important for researchers when doing a research survey. A bibliographic network is composed of three types of networks: a co-authorship network, a citation network, and a co-citation network. A co-authorship network represents the relationship between a researcher and his/her collaborators. A citation network represents reference relationship between papers. It is a directed network, in which a paper has a directed link to other paper by citing it. A co-citation network links two nodes (papers) when those appear together (co-cited) in at least one other paper.

A citation network/relationship is suitable for identifying the most cited papers and researchers, which are considered important papers and researchers. PaperLens provides "year by year top 10 cited papers/authors" view to identify the most cited papers and authors in each year. It can filter papers/authors by specifying research area [2].

On the other hand, a co-authorship network tends to be used to analyze the relationship between researchers [9]. Henry et al. identified collaboration patterns of researchers by visualizing a co-authorship network [3].

As such a network structure is complicated, information visualization techniques are usually employed for the analysis. In a node-link diagram which is the most often used representation of a network structure, a researcher or a paper is represented by a node and the relationship is represented by an edge. Ghoniem et al. [10] have shown the readability of node-link diagrams decreases for dense/large networks. To overcome that problem, various techniques have been proposed.

Node clustering/aggregations are commonly used to improve readability of a node-link diagram. Auber et al. [11] have proposed multiscale visualization. In this visualization, a network is divided into clusters, each of which corresponds to a small network and treated as a macro node, and the edges between clusters are clustered.

Ham et al. [12] have proposed an interactive visualization for node clusters to inspect inside of clusters. Holten [13] has proposed hierarchical edge bundles to improve the readability of global relationship trend. TreePlus [14] and Vizster [15] enabled exploratory analysis of local structure of a large network.

B. InfoVis 2004 Contest

The IEEE InfoVis 2004 Contest [16] provides a dataset which contains bibliographic information of InfoVis papers from 1995 to 2002 and those references. The aim of the contest was "to promote the development of benchmarks for information visualization and establish a forum to advance evaluation methods" [16]. In the contest, three first place winners [1], [4], [5], one student first place winner [6], and eight second winners were selected.

The tasks of the contest are defined as follows.

1) Create a static overview of the 10 years of the InfoVis
2) Characterize the research areas and their evolution
3) Where does a particular author/researcher fit within the research areas defined in task 2?
4) What, if any, are the relationships between two or more or all researchers?

In summary, the first place winner entries tended to use citation networks to identify research areas, their relationships, and/or evolution (task 2). Co-authorship networks tend to be used to identify collaboration relationships of researchers (task 3 and 4).

In task 2, Ke et al. [1] used burst analysis of keywords to identify the research areas and their evolution. They showed the results as tables. Lee et al. [5] clustered papers into research areas using their titles, references, and keywords. Based on the clusters shown as table, they showed the evolution of research areas. Wong et al. [4] identified discriminating research areas by co-occurrence of words appeared in titles and abstracts.

© 2012 ACADEMY PUBLISHER

Papers were placed based on thematic similarity. By filtering the papers by years, they showed the evolution of research areas. Ahmed et al. [6] showed that the evolution of research areas and their citation relationship can be identified by using a 3D "worm" representation in task 2. The research areas are identified by clustering papers by the word histogram of titles, abstracts, and keywords using SOM (Self Organizing Map).

In task 3, Ke et al. [1] analyzed the keyword usage of the researchers in non-visual way and showed researcher's interesting areas. In task 4, they used co-authorship networks to identify the collaboration relationship. In task 3 and 4, Lee et al. [5] analyzed a citation network in terms of research area. For each researcher, research areas of his/her papers as well as those citing his/her papers are identified. They also showed collaborators of a researcher using co-authorship relationship. Wong et al. [4] identified research areas of a researcher based on his/her publications. They also defined the thematic similarity between researchers. The results showed that the influence of a researcher can be identified from citation relationships by highlighting papers which cite his/her papers. Ahmed et al. [6] showed the hierarchical relationships of researchers with using a 3D network visualization of co-authorship networks. They classified the researchers in that visualization according to the number of collaboration relationships (degree): 20+ degree, 10-19 degree, and less than 10 degree. They assumed that researchers with high degree are senior researchers and those with low degree are students and younger researchers.

In summary, above-mentioned entries have identified research areas with the process of evolution. The collaboration of researchers has been also identified. However, each of those two findings has been obtained separately despite both are outcomes of the researchers' collaboration.

III. Visualization System for Co-Authorship Networks

A. Overview of the system

Fig. 1 shows the screenshot of the proposed system. The system consists of three parts; Network panel (Fig. 1(a)), Node Detail panel (Fig. 1(b)), and Operation panel (Fig. 1(c)).

Network panel shows researchers and their collaborations as a node-link diagram, in which the size of a node indicates the number of his/her collaborators. The thickness of an edge indicates the number of collaborations (i.e., collaborative papers).

Node Detail panel provides detailed information of the node selected in the Network panel. This panel shows an enlarged pie chart of the selected researcher and lists his/her publication list, collaborators, keywords, and publication years. For each item in the lists, more detailed information is provided by a tooltip.

Operation panel allows a user to change the node visualization according to the purpose of analysis. The operations include enabling/disabling the highlighting of nodes and so on.

B. Research area identification

To identify research areas, the system provides four functions as follows.

1) Listing all keywords which the selected researcher uses.
2) Rendering a researcher as a pie chart (i.e., a node) showing the ratio of his/her keywords usage.
3) Highlighting researchers who use at least one of keywords used by the selected researcher.
4) Aggregating the nodes of researchers if they use exactly the same keywords, for making it clear their collaboration is tighter than usual.

As for (1), a user can check keywords a researcher uses with the Node Detail panel. The Node Detail panel lists all keywords used by the selected researcher.

Regarding (2), to check distribution of the keywords, a keyword can be assigned a color (hue), which becomes a *selected* keyword, and the system renders its usage with a pie chart in the Network panel. A color can be assigned either manually or automatically. A user can manually assign a color to a keyword using the Node Detail panel.

A user can assign a color to a keyword automatically based on one of following indices as well. The colors are assigned from red to blue in descending order of the index value.

- TF (Term Frequency)
- TF/IDF (Inverse Document Frequency)
- Alphabetic

The TF is suitable for identifying frequently used keywords. Two types of TF indices are available. The TF weight w_k of a keyword k is the frequency of k among all papers in the co-authorship network. The w_{vk} of k for a researcher v is the frequency of k among all papers written by v.

The TF/IDF is suitable for identifying keywords used by specific researchers particularly. There are two types of indices as well. The TF/IDF weight of a keyword w'_k and w'_{vk} are given by following formulas, where N is the number of all papers over the network, N_k is the number of papers attached the keyword k, N_v is the number of papers written by v, and N_{vk} is the number of v's papers attached the keyword k.

$$w'_k = w_k \times IDF_k, \quad (1)$$

$$IDF_k = \log \frac{N+1}{N_k+1}, \quad (2)$$

$$w'_{vk} = w_{vk} \times IDF_{vk}, \quad (3)$$

$$IDF_{vk} = \log \frac{N_v+1}{N_{vk}+1}. \quad (4)$$

Alphabetic index sorts the keywords alphabetically and assigns color according to the order.

Segments in a pie chart are also ordered according to one of above-mentioned indices for each researcher v (i.e.,

© 2012 ACADEMY PUBLISHER

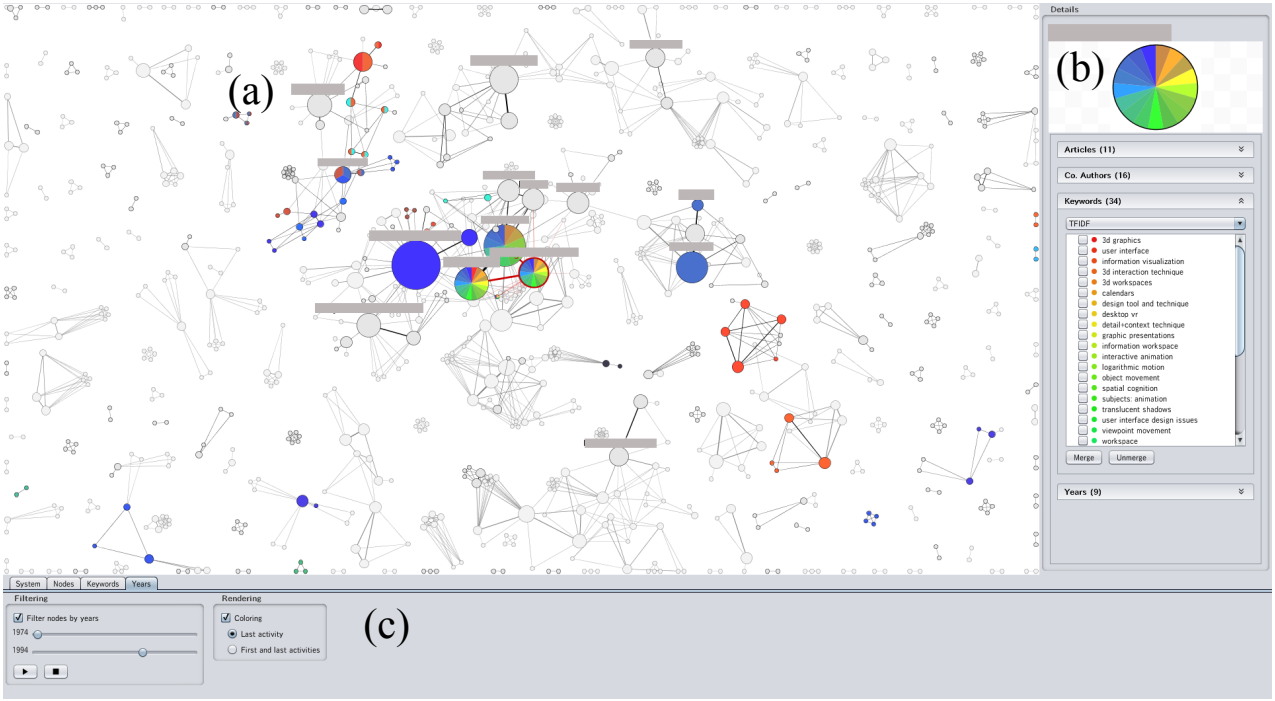

Figure 1. The overview of proposed system; (a) Network panel, (b) Node Detail panel, (c) Operation panel.

w_{vk}, w'_{vk} and alphabetic index). This means the positions of a segment in a pie chart can be different between researchers if their keywords usages are different.

If there are keywords that correspond to the same topic, a user can manually group those into one keyword group. A keyword group can be assigned a color in the same way as a single keyword.

The system has two variations of the pie chart representation according to the types of analysis. The first one shows only selected keywords in a pie chart (type *K1*). This is used for checking the distribution of specific keywords over the network. The second one shows all keywords in a pie chart, in which non-selected keywords are rendered in achromatic color (type *K2*). This is used for checking the concordance rate of keywords between researchers.

As for (3), for identifying relations between researchers in terms of keywords, the system can highlight the researchers who use at least one of the keywords used by the selected researcher.

Regarding (4), if the researchers having the collaborative papers use exactly the same keywords, their collaboration is tighter than usual. To show it visually, the system aggregates nodes in the Network panel if the corresponding researchers use exactly the same keywords. This also improves the readability of network view. The left figure in Fig. 2 shows normal node placement. In the right figure in Fig. 2, researchers using exactly the same keywords are aggregated.

C. Time variation

The system handles two types of time variation; time variation of researcher's collaboration (i.e., publication activities) and that of keywords usage.

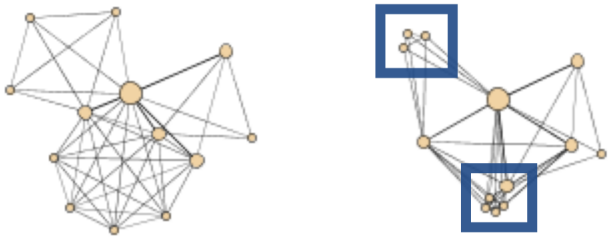

Figure 2. Nodes aggregating based on the keywords usage (left: without aggregation, right: with aggregation). In right figure aggregated nodes are enclosed by rectangles.

This paper divides the task of Identifying time variation of researchers' collaboration into three sub tasks:

- Identifying whether a researcher published papers in a certain period or not.
- Identifying whether a researcher published papers recently or not.
- Identifying whether a researcher published papers continuously or sporadically.

For the first task, a user can highlight the researchers who published papers in a certain period from y_s (year) to y_e (year) in the Network panel. The highlighted area can be varied from $[y_s, y_s]$ to $[y_s, y_e]$ with animation, so that the change of researchers' collaboration can be shown. The system also shows all the publication years of the selected researcher in the Node Detail panel.

For the second task, the brightness and saturation of a node represent the last period when a researcher published papers (type *T1*). Brighter node indicates the corresponding researcher published a paper more recently.

© 2012 ACADEMY PUBLISHER

Finally, to identify whether a researcher published papers continuously or sporadically, corresponding node can be represented with bi-cylindrical style (type *T2*). The inner ring represents the earliest year of his/her publication and outer ring represents the last year. In the left figure in Fig. 3, both of inner and outer rings of the node are dark, which indicates the corresponding researcher published papers early in the specified period only. In the center figure in Fig. 3, the inner ring of the node is dark, but the outer ring is bright. This indicates the corresponding researcher published papers continuously. In the right figure in Fig. 3, the both of inner and outer rings of the node are bright, which indicates corresponding researcher published papers late in the period only.

Figure 3. Visualization of time variation of researchers' collaboration (left: a researcher who published papers early in the period only, center: a researcher who published papers continuously, right: a researcher who published papers late in the period only)

Identification of time variation of keyword usage is done by combining {*K1*, *K2*} and {*T1*, *T2*} visualization as shown in Fig. 4. In the same way as type *T1*, the brightness and saturation of a segment in a pie chart represent the period when a certain keyword was used. When the bi-cylindrical style (*T2*) is used, the inner and the outer rings represent the earliest and the most recent year when the corresponding researcher used it, respectively. In Fig. 5(a), the inner ring of the keyword at top right segment is dark and outer ring is bright. This indicates that keyword has been used continuously. On the other hand, the rings of the keyword at middle right (Fig. 5(b)) are colored dark brown. This indicates that keyword was used early in the specified period only. The color of the keyword at the bottom right (Fig. 5(c)) is bright yellow, which indicates that keyword was used in late in the period only. In a pie chart representation, type

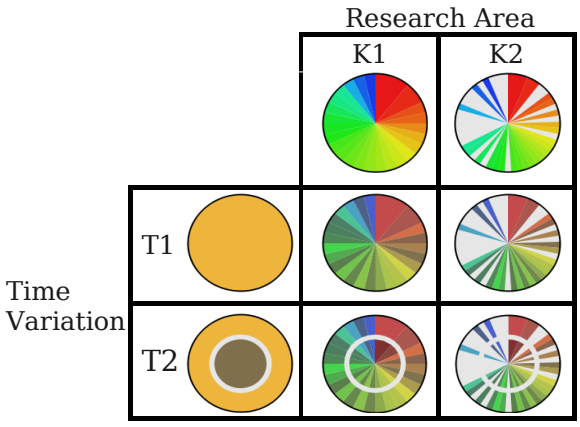

Figure 4. Variations of nodes visualization

K1 and *K2* as above-mentioned are available also in this case.

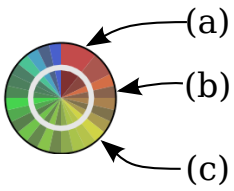

Figure 5. Visualization of time variation of keywords usage; (a) the keyword used continuously, (b) used in early period only, (c) used in later period only

IV. User Study

It has been shown in [17] that using the proposed functions in combination, authors could find the same insights found by first place winners in InfoVis 2004 Contest. The performed task corresponds to the analysis of past research activities from a co-authorship network.

On the contrary, this paper examines the effectiveness of the proposed system for supporting the analysis of future research activities, and its usability for users unfamiliar with target domain.

Although the comparing with other systems is useful for examining the effectiveness of the proposed system, detailed analysis, such as the contribution of each function, is difficult through the comparison between systems having considerably different interfaces and functions. Therefore, two types of systems are prepared: the system with full functions and that with limited functions. A "limited functions" indicates that test participants cannot use functions for assignment colors to keywords and its visualization, grouping keywords manually, and bi-cylindrical visualization of time variation of a researcher or a keyword. Other than those limitations are as the same as the system with full functions.

Participants are asked to identify researchers of following two types.

Growing researcher
: A researcher who is doing interesting work and would publish research papers actively in future from a period of a dataset.

Supervisor
: A researcher who did or is doing interesting research work but would not publish papers so actively in future from a period of a dataset, because s/he has already achieved in the period of a dataset.

The reason of adopting these kinds of abstract tasks is to let participants perform exploratory data analysis based on various assumptions. That is, when giving such an abstract purpose, participants are supposed to translate it into more concrete assumptions about the conditions the researchers to be identified should satisfy. The participants should perform analysis guided by such assumptions with the combination of functions provided by the system. It is also expected that a participant would

© 2012 ACADEMY PUBLISHER

make various assumptions until s/he obtains sufficient results. The purpose of the user study is to examine the relationship between assumptions and used functions.

Two participants took part in the study. Both of them are male graduated students of system design major and aged early 20s. It should be noted we selected test participants who don't have background knowledge about Information Visualization. We think support of research survey by users unfamiliar with target domain is important, because research surveys are inevitable for researchers / companies getting into new domains. In particular, the importance of such surveys is growing for companies trying to adapt to rapidly changing business environment.

A participant first did the task by using the system with limited functions. Before doing the task, participants were lectured about the usage of *T1* for node visualization and keyword list in Node Detail panel without explanation about how to assign color for keyword and group keywords manually. After finishing the task, they are lectured about remaining functions: *T2*, *K1* and *K2* and the manual color assignment to keywords / keyword groups. After the lecture, they did the same task again, with using full functions.

After each task, we had interviews with participants about assumptions and reasons of their analysis procedures.

We use the co-authorship networks extracted from InfoVis 2004 dataset in the experiments. The InfoVis 2004 Contest [16] dataset contains bibliographic information of InfoVis papers from 1995 to 2002 and those references. As some entries of the contest have pointed out that there are duplications or errors in the dataset [1], [6], we cleaned up the dataset based on [1]. The extracted network contains 969 researchers (nodes), 1736 collaborations (edges) and 1777 keywords. We treated the publishing date by year. Published period is from 1974 to 2004, in which there are no missing years.

V. Results

A. System with Limited Functions

The analyses of two participants were based on almost the same assumptions:

- A growing researcher has a certain number of papers and also published papers in recent period (2000-2004).
- A supervisor has a certain number of papers but doesn't published papers in recent period (2000-2004).

They used combination of following steps to ensure the assumptions.

- Identify the number of papers which a researcher published by his/her node size
- Highlight researchers who published papers in a certain period

In particular, if corresponding node is relatively large and highlighted in recently period, a researcher is considered as growing researcher. If a node is relatively large but not highlighted, corresponding researcher is considered as supervisor.

They also employed different types of clues to ensure above-mentioned assumptions. Participant A focused on groups of researchers. He looked the co-authorship network as "Map of Researcher Groups," which consists of nodes with various sizes (Fig. 6). By changing time period for highlighting nodes, he identified "Rise and Fall of Research Groups." He considered relatively large nodes highlighted in recent period (2002-2004) as growing researchers and those unhighlighted relatively large nodes as supervisors.

Participant B used *T1* to visualize time variation of node in addition to the function of highlighting researchers. First he highlighted recent 5 years (2000-2004). In this situation, researchers who published papers during 2003 and 2004 become blight nodes (enclosed by red solid rectangle) and those who published during 2000 and 2001 become dark nodes (enclosed by blue dashed rectangles) in *T1* visualization (Fig. 7). From this result he considered blight nodes as growing researchers and dark nodes as supervisors.

It can be said that there is a relationship between their analyzing strategies and the functions provided by the system. That is, as they could not focus on keyword usage, they had to ensure their assumptions only from information about publications.

B. System with Full Functions

Also in this case, both of participants used almost the same schemes to complete sub-tasks. To identify growing researchers, they used following schemes:

- (Step1) Identify a researcher who published papers in recent 5 years (2000-2004)
- (Step2) Assign colors to keywords according to participant's assumption
- (Step3) Identify keywords usage over the network

For identification of a researcher who published papers in recent period, both participants used the function of highlighting researchers.

On the other hand, in the second step, they selected keywords according to different assumptions. The assumption of participant A is following:

- If a researcher is a growing researcher, some of his/her keywords are used by only his/her collaborators in both recent and early period.

Participant A used the function of automatically assigning colors to selected keywords. As he assigned colors based on TF/IDF value (w'_{vk}), in which the colors are assigned from red (high TF/IDF value) to blue (low TF/IDF value), unique keywords used by few other researchers are colored red or yellow (Fig. 8). Therefore he could easily check the usage of such unique keywords over the entire network in recent and early periods. If the red or yellow keywords are concentrated to his/her collaborators in both periods (Fig. 8), he considered the researcher is a growing researcher. In Fig. 8, a researcher enclosed by

© 2012 ACADEMY PUBLISHER

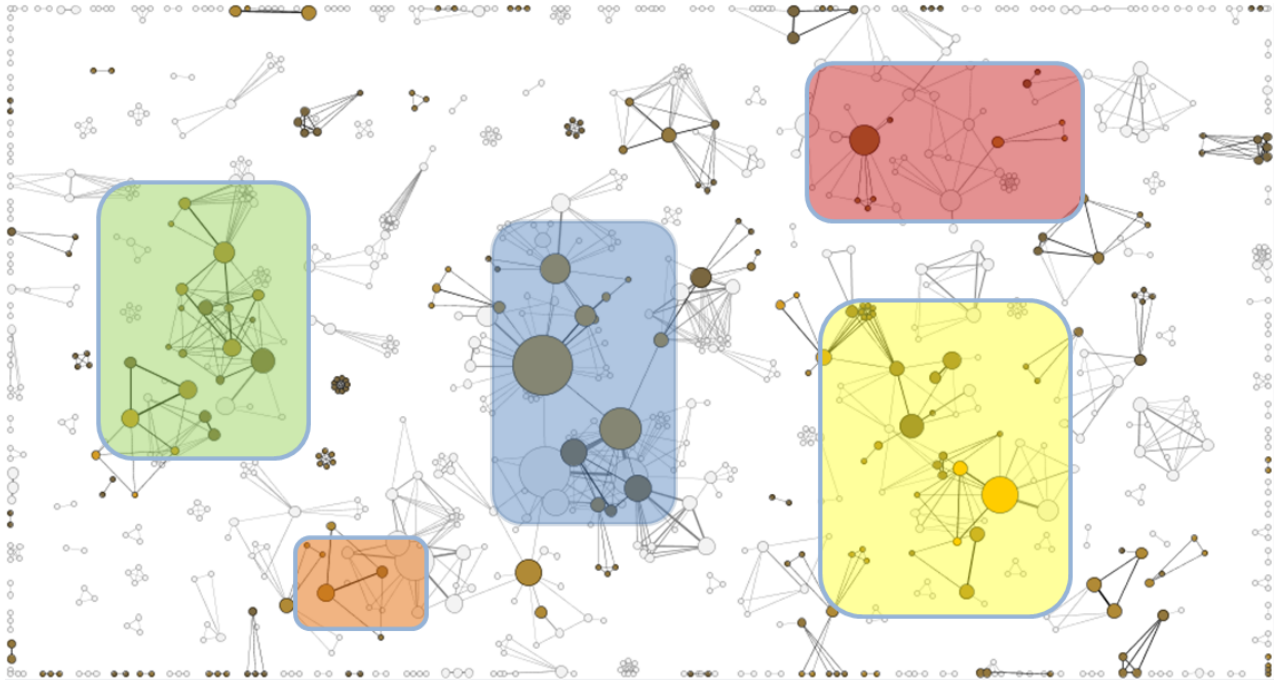

Figure 6. "Map of Researcher Groups" in recent period (2000-2004). Researcher groups are enclosed by rectangles.

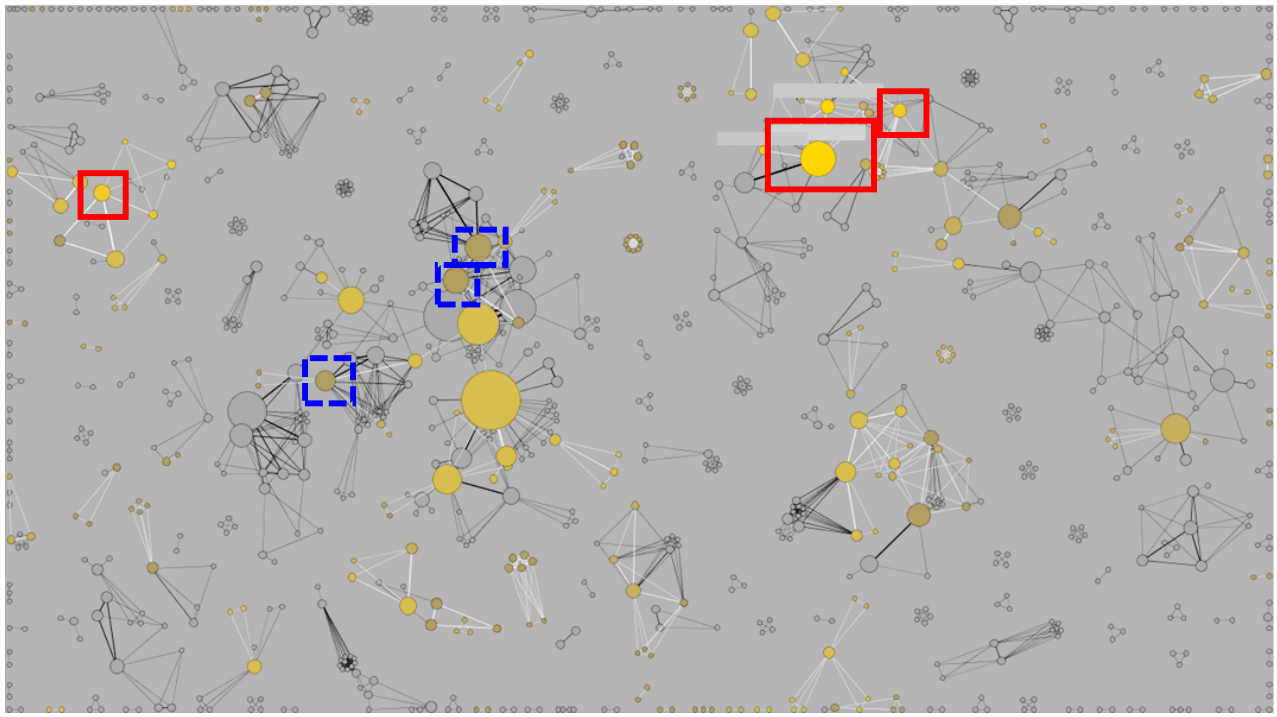

Figure 7. The distribution of last publishing year in recent period (2000-2004).

© 2012 ACADEMY PUBLISHER

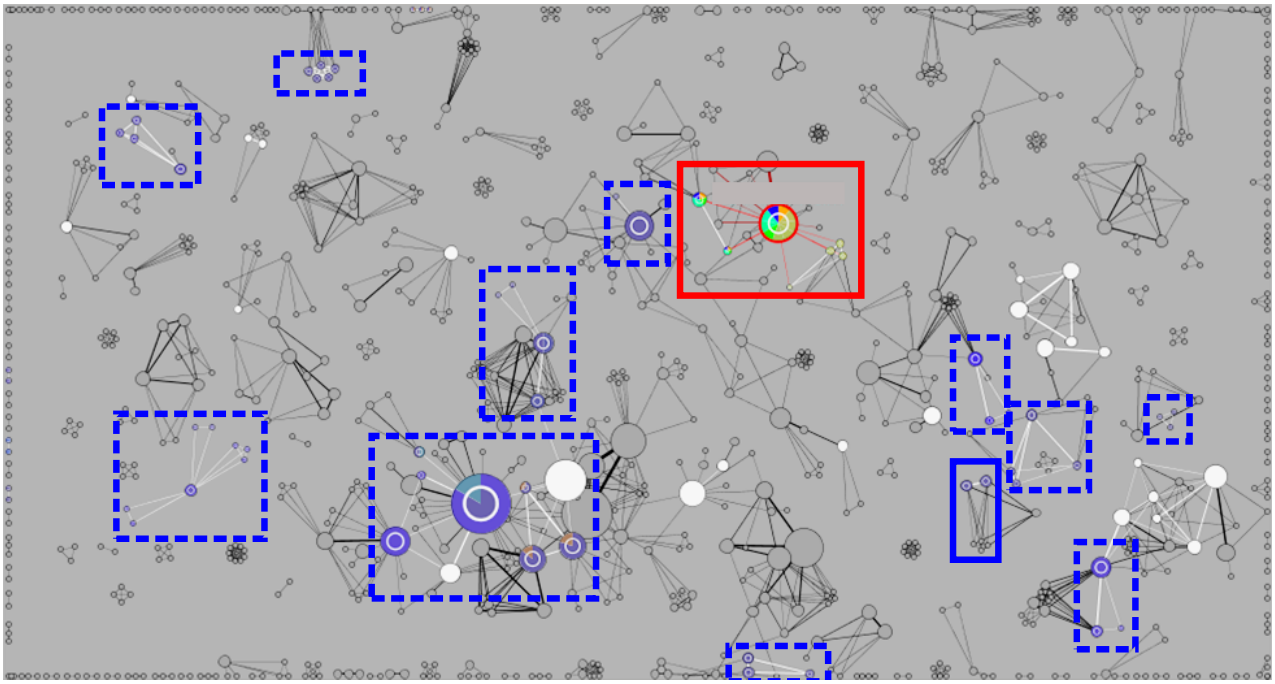

Figure 8. The usage of keywords used by a growing researcher in recent period (2000-2004) (analyzed by participant A)

.

a red solid rectangle is considered a growing researcher. The researchers enclosed by blue dashed rectangles use keywords used by the selected researcher. However they use only non-unique (green or blue) keywords.

Fig. 9 shows the change of his keyword usage from early period to recent period, which satisfies the assumption of participant A.

On the other hand, participant B focused on researcher group, based on the following assumption:

- A growing researcher shares "hot" keywords with several researcher groups in recent period.

In this case, a "hot" keyword means a recently used keyword. He assigned colors to all keywords in the dataset based on TF/IDF value (w'_k), then checked whether a researcher group shares keywords with other groups or not. To identify that, he used the function of highlighting researchers who share at least one of keywords used by the selected researcher. Fig. 10 shows the usage of keywords used by a growing researcher in recent period (2000-2004). If the selected researcher and other several groups use the same keywords in recent period, he considered the selected researcher as a growing researcher. In Fig. 10, a researcher group enclosed by a red rectangle contains a growing researcher. The groups enclosed by green or yellow rectangles use at least one of keywords used by the growing researcher.

To identify supervisors, both of participants used similar schemes as used to identify growing researchers.

- (Step1) Identify a researcher who published papers in early period but rarely published papers in recent period (2000-2004)
- (Step2) Assign colors to keywords according to a participant's assumption
- (Step3) Identify keywords usage over the network

In Step2 and 3, they employed the different assumptions each other. The assumption of participant A is following:

- If a researcher is a supervisor, his/her keywords with high TF/IDF value (w'_{vk}) are used by researchers other than his/her collaborators in recent and early periods.

As opposite to identification of growing researchers, if unique (red or yellow) keywords are shared without collaborations in early and recent periods, he considered the researcher as a supervisor. In Fig. 11, a supervisor is enclosed by a red solid rectangle. The researchers enclosed by blue dashed rectangles use keywords used by the selected researcher. They use unique (red or yellow) keywords without collaborations.

The assumption of participant B is following:

- If a researcher is a supervisor, keywords which were used by him/her in past period (1974-1999) are used by several other researchers over the network in recent period (2000-2004)

As similar to participant A, participant B assigned colors to keywords based on TF/IDF value (w'_{vk}), in which unique keywords used by few other researchers are colored red or yellow. With this visualization, he could easily identify the keywords usage over the network. Interestingly there were two types of supervisors in his results; researchers who used red or yellow keywords (Fig. 12) and those without red or yellow keywords (Fig. 13). In Fig. 12 and 13, a supervisor is enclosed by a red solid rectangle. The researchers enclosed by blue dashed rectangles use keywords used by the supervisor. In Fig.

© 2012 ACADEMY PUBLISHER

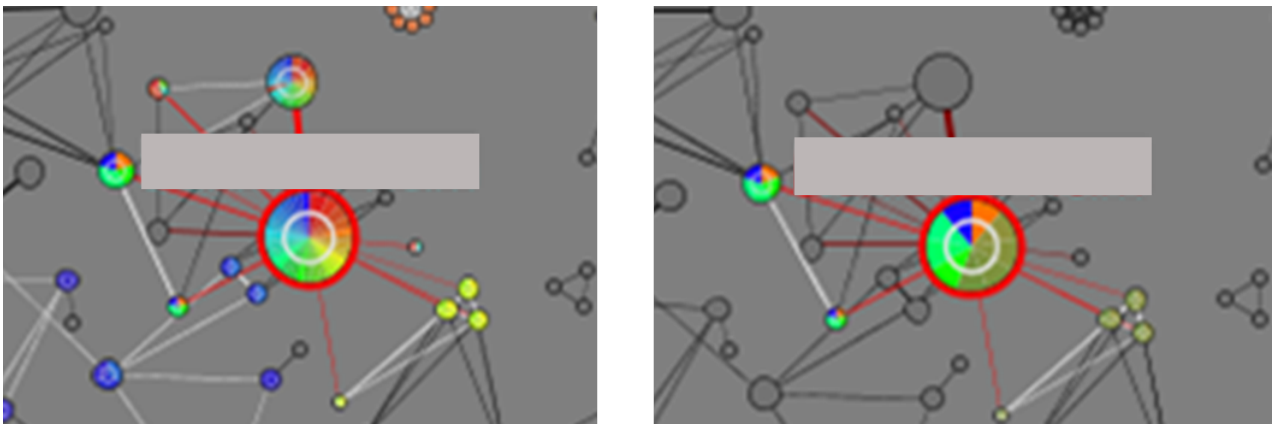

Figure 9. Change of keyword usage: (a) early period (1974-1999), (b) recent period (2000-2004)

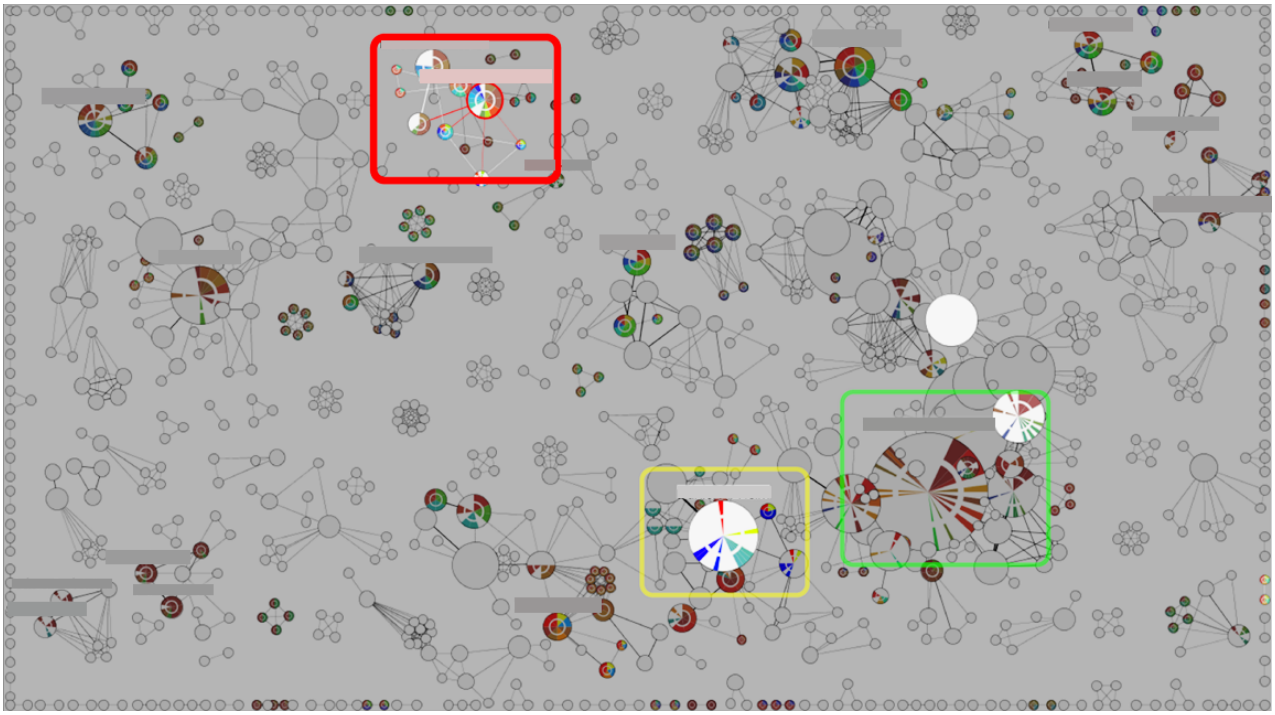

Figure 10. The usage of keywords used by a growing researcher in recent period (2000-2004) (analyzed by participant B)

12, non-unique (green or blue) keywords used by the supervisor are shared by others. In Fig. 13, unique (red or yellow) keywords are shared by those who have no collaboration with the supervisor.

It is not surprising that assumptions of participants are different, as the tasks are abstract and can be performed based on various criteria. It should be noted that both of participants consider keywords usage in a co-authorship networks is important information, in spite of having different assumptions each other. That is, the function is not single-purpose, by which analysis with various assumptions is possible.

Furthermore, the comparison of employed strategies between using the system with limited functions and with full functions shows that the available functions affect their analyzing strategies. By using functions for visualizing keyword usage, they had assumptions regarding research topics.

It is also observed that bi-cylindrical visualization for time variation of node (type *T2*) was not used in their strategies. One possible reason is that it could increase visual complexity. Because there are many nodes and edges in Network panel, participants preferred to the visualization functions which reduce (highlighting) or do not increase (type *T1*) visual complexity. The functions of manual color assignment to keyword and manually grouping keyword were not used as well. Although the functions are useful for analyzing time variation of a few keywords [10], it is supposed that manual operations are not suitable for a large number of keywords.

It is noted that no predetermined correct answers (ground truth) is given in the user study, because our aim

© 2012 ACADEMY PUBLISHER

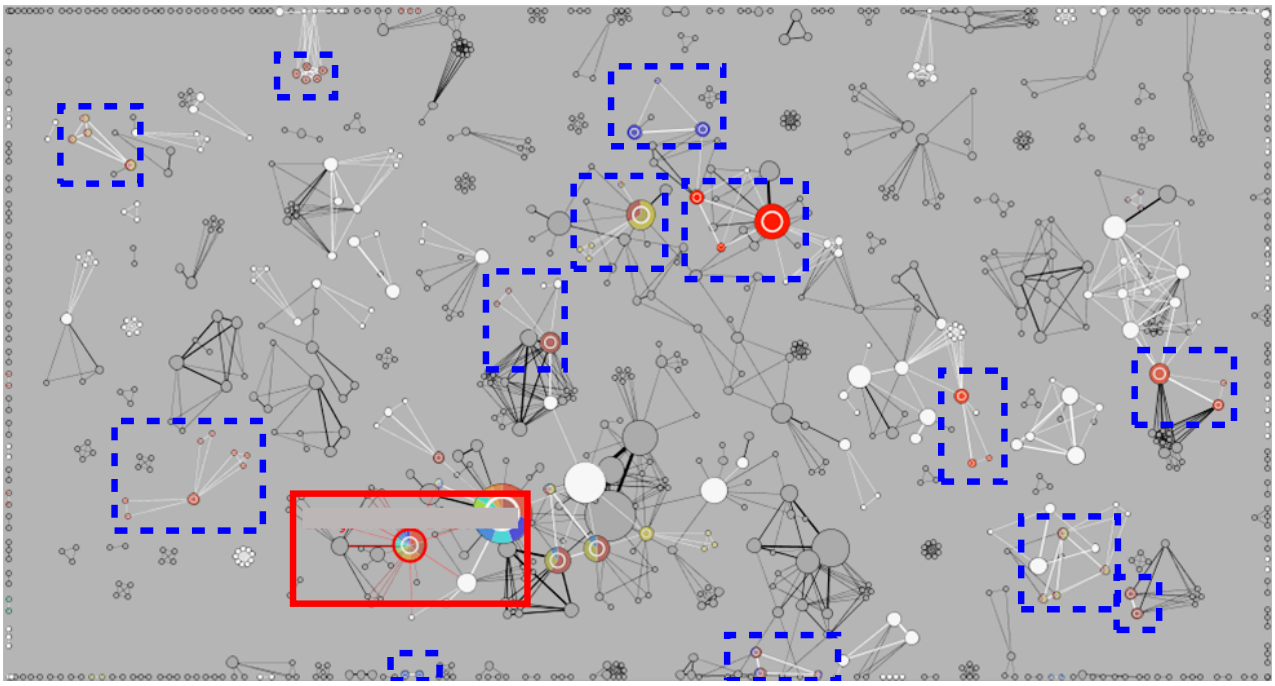

Figure 11. The usage of keywords used by a supervisor in recent period (2000-2004) (analyzed by participant A).

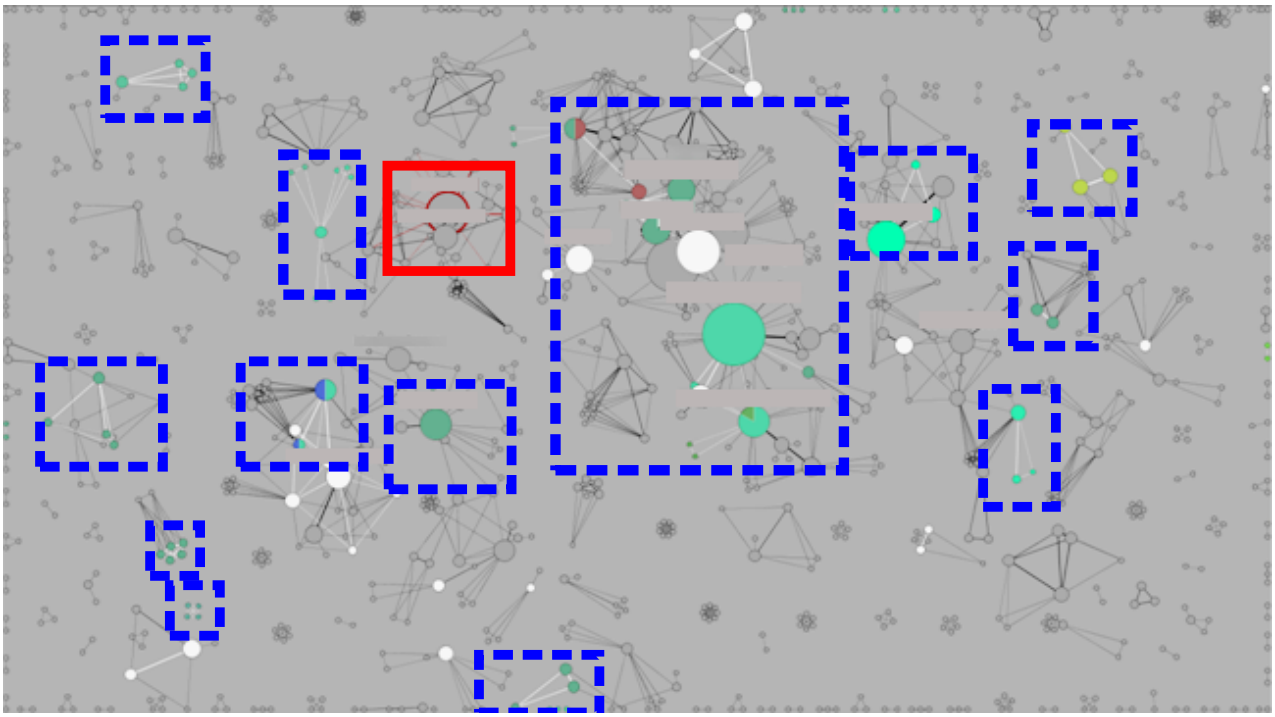

Figure 12. The usage of non-unique (green / blue) keywords used by a supervisor in recent period (2000-2004) (analyzed by participant B).

© 2012 ACADEMY PUBLISHER

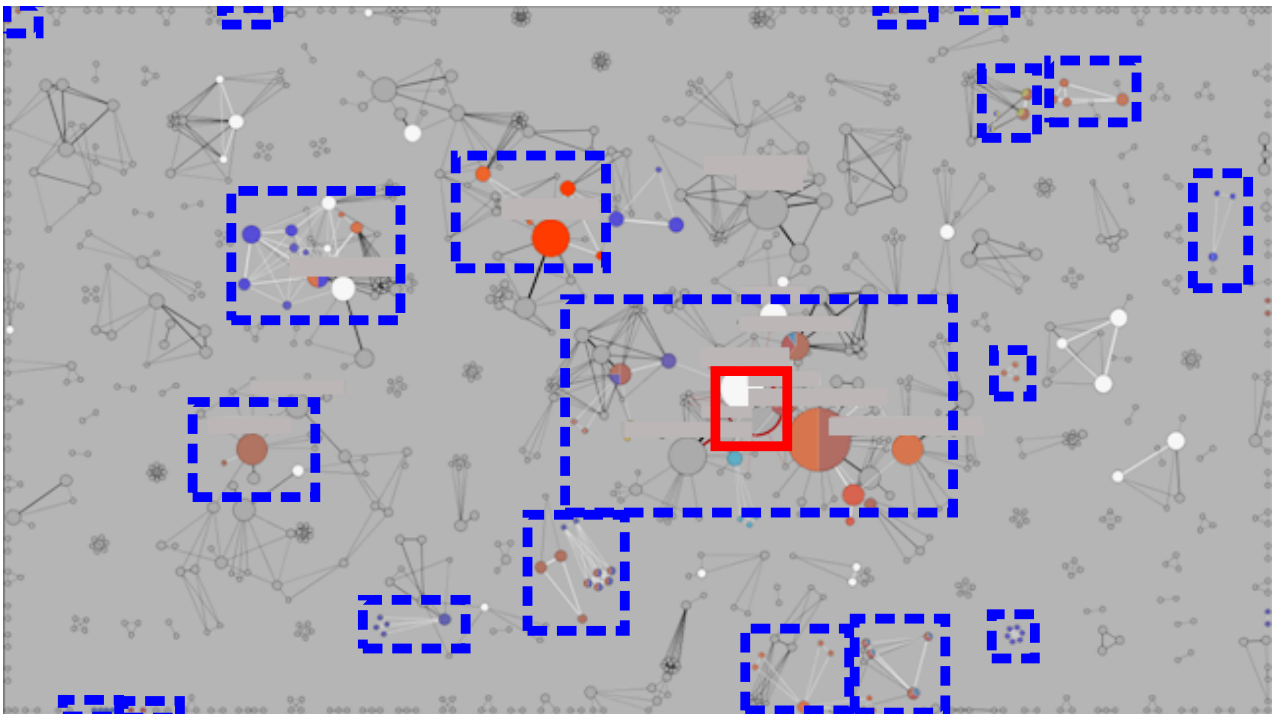

Figure 13. The usage of unique (red / yellow) keywords used by a supervisor in recent period (2000-2004) (analyzed by participant B).

is not to support accurate analysis but to encourage users to perform exploratory data analysis, based on their own assumptions. However, it is also important to confirm that users can analyze data reasonably. Therefore, instead of using ground truth, we asked the participants to conduct an additional task. After finishing all tasks, participants investigated the number of research papers from 2004 to 2007 published by their identified researchers using DBLP. They found that most of growing researchers they identified published 10 or more papers per year and most of supervisors published less than 5 papers per year in that period. These facts show their predictions using the system are reasonable.

VI. Conclusions

This paper proposes the visualization system for co-authorship networks. To identify growing researchers and supervisors, the system provides support functions for identifying research areas as well as for examining its time variation.

This paper examines whether the proposed system is able to support users predicting future research activities of their unfamiliar domain. In the user study, test participants unfamiliar with InfoVis performed exploratory analysis of the co-authorship network to identify growing researchers and supervisors. The results show that they could perform the tasks even though they had no background knowledge about InfoVis. During performing tasks, it was observed that they employed different analyzing strategies according to available functions. Among the functions the system provided, the function for highlighting researchers who published papers in a specified period is heavily used to identify researchers who recently published papers. The functions for assigning colors to keywords and visualizing node as pie chart of keywords are also frequently used to identify types of researchers. By combining these functions, they could perform the tasks successfully. As one of future works, it is expected that automated assistance of frequently used assumptions would improve the effectiveness of the system.

In this paper, the result of user study by two participants is shown, as it is important for the study of user interface to analyze specific users' behaviors in detail. On the other hand, an experiment with many participants is also important, which should be conducted as one of future studies. The result of this paper will contribute to an experimental design.

Future works also include the introduction of other information than co-authorship networks. Participants suggested the order of authors in a research paper can be useful to determinate whether a researcher is young researcher or a supervisor, as young researcher tends to be first author and supervisor tends to be the second or later. It was also suggested that collaboration of researchers is clearly appeared when they received research grants. As their relationship is stronger than usual, analyzing grant programs could provide interesting information.

Acknowledgment

We would like to thank Mr. Ippei Ozawa and Mr. Kenji Takamiya for their participations of our study.

This work was partially supported by a grant from National Institute of Informatics (NII) and Japan-Taiwan Joint Research Program by Interchange Association, Japan.

© 2012 ACADEMY PUBLISHER

References

[1] W. Ke, K. Borner, and L. Viswanath, "Major Information Visualization Authors, Papers and Topics in the ACM Library," in *Proc. IEEE Symp. Information Visualization*, 2004, p. 216.

[2] B. Lee, M. Czerwinski, G. Robertson, and B. B. Bederson, "Understanding Research Trends in Conferences using PaperLens," in *Extended Abstracts on Human Factors in Computing Systems*, 2005, pp. 1969–1972.

[3] N. Henry, J.-D. Fekete, and M. J. McGuffin, "NodeTrix: a Hybrid Visualization of Social Networks," *IEEE Trans. Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1302–1309, 2007.

[4] P. C. Wong, B. Hetzler, C. Posse, M. Whiting, S. Havre, N. Cramer, A. Shah, M. Singhal, A. Turner, and J. Thomas, "IN-SPIRE InfoVis 2004 Contest Entry," in *Proc. IEEE Symp. Information Visualization*, 2004, p. 216.

[5] B. Lee, M. Czerwinski, G. Robertson, and B. B. Bederson, "Understanding Eight Years of InfoVis Conferences Using PaperLens," in *Proc. IEEE Symp. Information Visualization*, 2004, p. 216.

[6] A. Ahmed, T. Dwyer, C. Murray, L. Song, and Y. X. Wu, "WilmaScope Graph Visualisation," in *Proc. IEEE Symp. Information Visualization*, 2004, p. 216.

[7] T. Kurosawa and Y. Takama, "Visualization-based Support of Hypothesis Verification for Research Survey with Co-Authorship Networks," in *Proc. Int'l Workshop on Intelligent Web Interaction*, 2011, pp. 134–137.

[8] ——, "Predicting Researchers' Future Activities using Visualization System for Co-Authorship Networks," in *Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence*, 2011, pp. 332–339.

[9] M. E. J. Newman, "The structure of scientific collaboration networks," *Proc. National Academy of Sciences of the United States of America*, vol. 98, no. 2, pp. 404–409, 2001.

[10] M. Ghoniem, J.-D. Fekete, and P. Castagliola, "A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations," in *IEEE Symp. Information Visualization*, 2004, pp. 17–24.

[11] D. Auber, Y. Chiricota, F. Jourdan, and G. Melançon, "Multiscale Visualization of Small World Networks," in *Proc. IEEE Symp. Information Visualization*, 2003, pp. 75–81.

[12] F. v. Ham and J. J. v. Wijk, "Interactive Visualization of Small World Graphs," in *Proc. IEEE Symp. Information Visualization*, 2004, pp. 199–206.

[13] D. Holten, "Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data," *IEEE Trans. Visualization and Computer Graphics*, vol. 12, no. 6, pp. 741–748, 2006.

[14] B. Lee, C. S. Parr, C. Plaisant, B. B. Bederson, V. D. Veksler, W. D. Gray, and C. Kotfila, "TreePlus: Interactive Exploration of Networks with Enhanced Tree Layouts," *IEEE Trans. Visualization and Computer Graphics*, vol. 12, no. 6, pp. 1414–1426, 2006.

[15] J. Heer and D. Boyd, "Vizster: Visualizing Online Social Networks," in *Proc. IEEE Symp. Information Visualization*, 2005, p. 5.

[16] J.-D. Fekete, G. Grinstein, and C. Plaisant, "
IEEE InfoVis 2004 Contest, The History of InfoVis,"
http://www.cs.umd.edu/hcil/iv04contest, 2004.

[17] T. Kurosawa and Y. Takama, "Visualization System for Co-authorship Networks to Get Insight into Future Research Activities," in *Proc. Joint Int'l Conf. Soft Computing and Intelligent Systems and Int'l Advanced Intelligent Systems*, 2010, pp. 339–344.

Takeshi Kurosawa, Japan, 1987. He received the B.E. in Information Communication Technology from Tokyo Metropolitan University Tokyo, Tokyo Japan in 2010. 2010–present, he is a master degree student in Graduate School of System Design, Tokyo Metropolitan University, Japan.

Mr. Kurosawa is a student member of IEEE.

Yasufumi Takama, Japan, 1971. Dr. Eng, University of Tokyo, Tokyo Japan, 1999.

He was a JSPS (Japan Society for the Promotion of Science) Research Fellow from 1997 to 1999. From 1999 to 2002 he was a Research Associate at Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology in Japan. From 2002 to 2005, he was an Associate Professor at Department of Electronic Systems and Engineering, Tokyo Metropolitan Institute of Technology, Tokyo, Japan. Since 2005, he has been an Associate Professor at Faculty of System Design, Tokyo Metropolitan University, Tokyo, Japan. He also participated in PREST (Precursory Research for Embryonic Science and Technology), JST (Japan Science and Technology Corporation) from 2000 to 2003. His current research interest includes Web intelligence, information visualization, and intelligent interaction.

Dr. Takama is a member of IEEE, JSAI (Japanese Society of Artificial Intelligence), IPSJ (Information Processing Society of Japan), IEICE (Institute of Electronics, Information and Communication Engineers), and SOFT (Japan Society for Fuzzy Theory and Intelligent Informatics).

© 2012 ACADEMY PUBLISHER