# Visualized Clustering of Ideas for Group Argumentation

Bin Luo, Xijin Tang
*Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China*
*{luobin, xjtang}@amss.ac.cn*

## Abstract

*This paper addresses visualized clustering methods that are embedded in CorMap and iView analysis of ideas towards the concerned topic. K-means clustering, automatic affinity diagram (KJ method) and self-organizing map are applied to CorMap analysis and graph clustering algorithm is applied to iView analysis are introduced. We report the visualized clustering results of workshops of a famous scientific forum, show the features of each clustering and compare their performance.*

*Keywords: visualized clustering, self-organization map (SOM), graph clustering, k-means clustering, affinity diagram*

## 1. Introduction

Group discussion produces ideas and opinions towards the concerned topic. Those ideas may not be well organized. There has been a growing interest in the analysis of group ideas by a variety of intelligent technologies, e.g. text clustering, text classification, etc. In those technologies to process human ideas, clustering is a major task. In this paper, we focus on exploring visualized clustering in order to get the main points and basic ideas of a group discussion.

CorMap and iView are two technologies which integrate different algorithms to process community ideas and have been developed for qualitative meta-synthesis by MSKS research group in AMSS, CAS [1, 2]. Three different types of clustering methods are applied in CorMap and iView analysis, *statistic-based analysis, artificial neural network and graph-based clustering analysis*. In this paper the workflow of CorMap and iView analysis is briefly introduced together with the applied visualized clustering methods including *k*-means clustering, automatic affinity diagram (KJ method), self-organizing map (SOM) and graph clustering that are integrated into both technologies. We apply those clustering methods to some dedicated workshops of one topic in a famous scientific forum in China. Analytical results of different methods are given in Session 3 together with a comparison of performance of those methods. Session 4 is the conclusion.

## 2. Different visualized clustering in CorMap and iView analysis

Both CorMap analysis and iView analysis aim to conduct exploratory analysis for textual data [1] [2]. The meta-data for both technologies is of a structure as <*topic, speaker, speech, keywords, time*>. The *keywords* are articulated as attributes of *speakers* and *speeches*. Either CorMap or iView analysis shows different perspectives toward the same data set based on different mechanisms with the same aim to acquire constructs of the problems from those textual data. Figure 1 shows the workflow and main functions of CorMap and iView analysis. Four different clustering methods belong to three different types are implemented in CorMap and iView analysis.

*K-means clustering* aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean [3]. *K*-means of keywords run on the spatial relations acquired in the correspondence analysis. A label is given to each cluster based on its centroid.

*Automatic affinity diagram.* Affinity diagram, also called the KJ method, is used to organize ideas based on their natural relationships. These diagrams may reveal new patterns and relationships between ideas, and then lead to more creative solutions [4]. Automatic affinity diagramming is conducted in CorMap analysis by dividing the 2D space into grids. The ideas spatially fall into the same grid are regarded as one cluster.

978-0-7695-4191-4/10 $26.00 © 2010 IEEE
DOI 10.1109/WI-IAT.2010.280

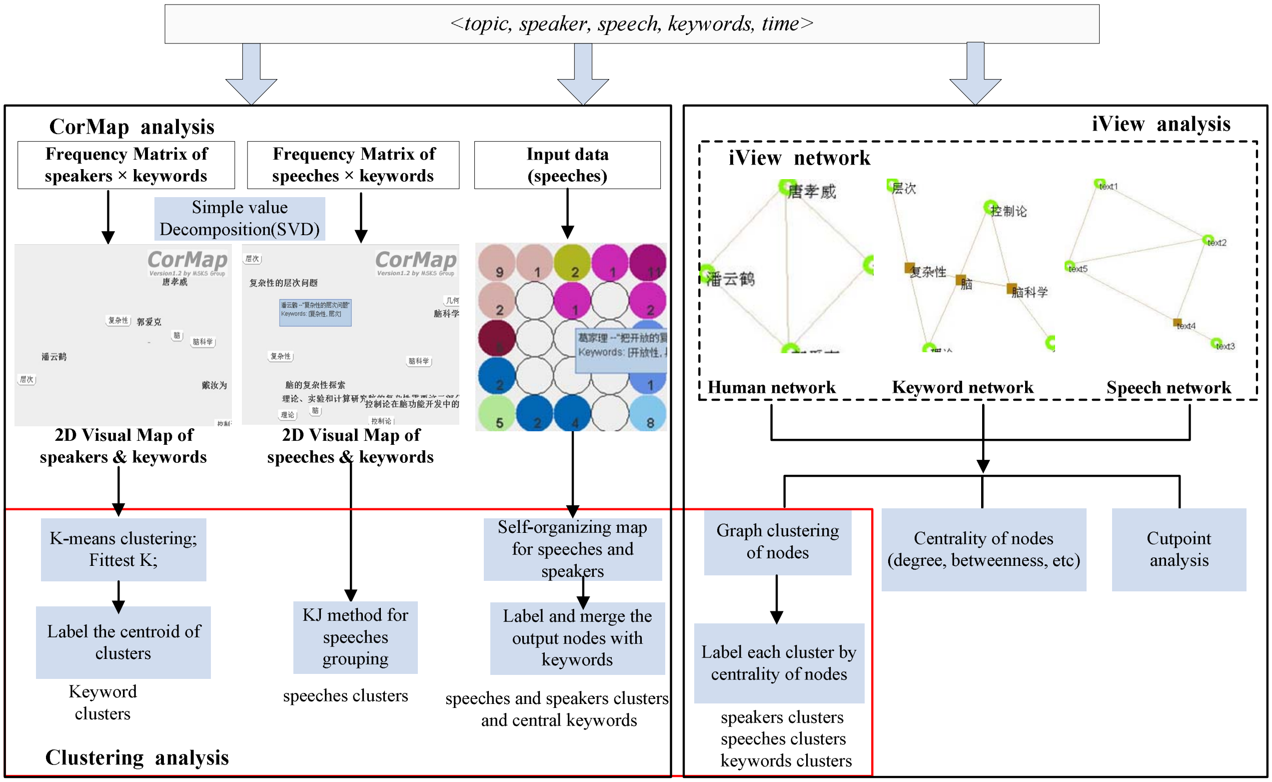


**Figure 1 Workflow and main functions of CorMap and iView analysis**

*Self-organizing map (SOM)* is a type of artificial neural network that is trained using unsupervised learning to visualize low-dimensional views of high-dimensional data. SOM forms a semantic map where similar samples are mapped close together and dissimilar apart [5]. Each unit in the lattice is called a neuron, and adjacent neurons are connected to each other, which gives the clear topology of how the network fits itself to the input space. Input data sharing common characteristics will activate adjacent areas on the map. A speech can be regarded as an input vector of the keyword.

*Graph clustering* is the task of grouping the nodes of the graph into clusters taking into consideration the edge structure of the graph in such a way that there should be many edges within each cluster and relatively few between the clusters [6]. Network structure detection method toward iView network will help to find the clusters of speech or keyword through the network structure.

## 3. Visualized clustering of XSSC workshops

Xianshan Science Conference (XSSC) is a famous science forum which denotes a series of small-scale workshops which bring together a group of scientists working at the research frontier and enable them to discuss all aspects of the most recent advances in the field in depth and to stimulate new directions for research in China. Then each workshop could be regarded as a group discussing process toward the concerned scientific problems [7, 8].

### 3.1 Dataset and preprocessing

One plenary speech in XSSC workshop can be represented as the meta-data used in CorMap and iView analysis. For example, an item

*<Complex systems and complexity research, Yu Jingyuan, Complex Giant System Engineering, {open, giant system, complexity, qualitative, quantitative}, 1998-12-22>* indicates that scientist *Yu Jingyuan* gave a speech titled *Complex Giant System Engineering* about the topic *complex systems and complexity research* on Dec 22, 1998. *Open, giant system, complexity, qualitative, quantitative* are 5 representative keywords of that speech. Here keywords of each speech are selected by the analysts.

A same data set analyzed by Ref. [7, 8] toward complexity-related XSSC workshops is applied with minor modification towards keywords of speeches in this study. For example, keyword like "*systems science*" in some speeches is amended to "*systems* *engineering*" according to the context of the speeches at the workshop. Then the data set includes 23 speakers, 92 keywords and 61 speeches. Next the analytical clustering results by the four different methods embedded in either CorMap or iView are given.

## 3.2 Clustering of keywords

Figure 2 shows some keyword clustering results of the XSSC workshops as group discussion. By performing a series of transformation and singular value decomposition (SVD), *k*-means clustering results based on the location of author and keyword are shows in Figure 2 (a) (*k*=7, which is the fittest number of clusters, and users are allowed to select different *k* from 1 to 20). The keywords are shown in boxed labels and the names of speakers are in bold black font. 7 clusters are found and different clusters are visualized in different colors. Keywords "心智 *human mind*", "意识 *consciousness*", "决策竞争 *competition of decision-making*", "多尺度 *multi-scale*", "方法论 *methodology*", "自然控制论 *natural cybernetics"* and "经济系统 *economic system"* are acquired as labels of the 7 clusters. The label is assigned as the keyword that is closest to the centroid of each cluster. Regarding the corresponding authors with those keyword clusters, author grouping can be acquired simultaneously. Figure 2 (b) shows visualized clustering result toward keyword network (8 clusters). The labels of each cluster are *"综合集成 meta-synthesis", "定性 qualitative", "复杂巨系统 complex giant system", "多尺度 multi-scale", "自然控制论 natural cybernetics", "研讨厅 Hall for Workshop of Metasynthetic Engineering - HWMSE", "系统科学 systems science" and "意识 consciousness*. The label of each cluster is the highest betweenness keyword of the cluster. The cutpoints of the keyword network are *"开放 open", "控制论 cybernetics" , "开放性 openness", "系统工程 systems engineering " and "脑 brain"* which may be regarded as the major points discussed in the *"complex systems and complexity research"* workshops. More interactive functions are provided by CorMap and iView technologies for a variety of analysis to explore structures of those scientific discussions.

We find that keywords "多尺度 *multi-scale*", "意识 *consciousness*", "自然控制论 *natural cybernetics*" are the same words selected by *k*-means clustering and graph clustering. Table 1 shows the keyword list of each cluster has the same labels by different clustering. The cluster labeled "意识 consciousness" by *k*-means clustering contains 5 words, "意识 consciousness", "神经网络 neural network", "记忆 memory", "认知 cognition", and "脑 brain". The cluster given by keyword network clustering contains 3 words, "意识 consciousness", "神经网络 neural network" and "记忆 memory". The cluster labeled as "自然控制论 *natural cybernetics*" contains keyword "控制论 *cybernetics*" that is not in the keyword list of the cluster with the same label by graph clustering. Keyword "非线性 *nonlinearity*" is contained in the keyword list of Cluster "多尺度 *multi-scale*" but not in the same cluster by graph clustering. Even by clustering with different mechanisms, some keywords still exist in the same cluster, some may belong to others. Different clustering of ideas shows different visions of the main points of the workshops.

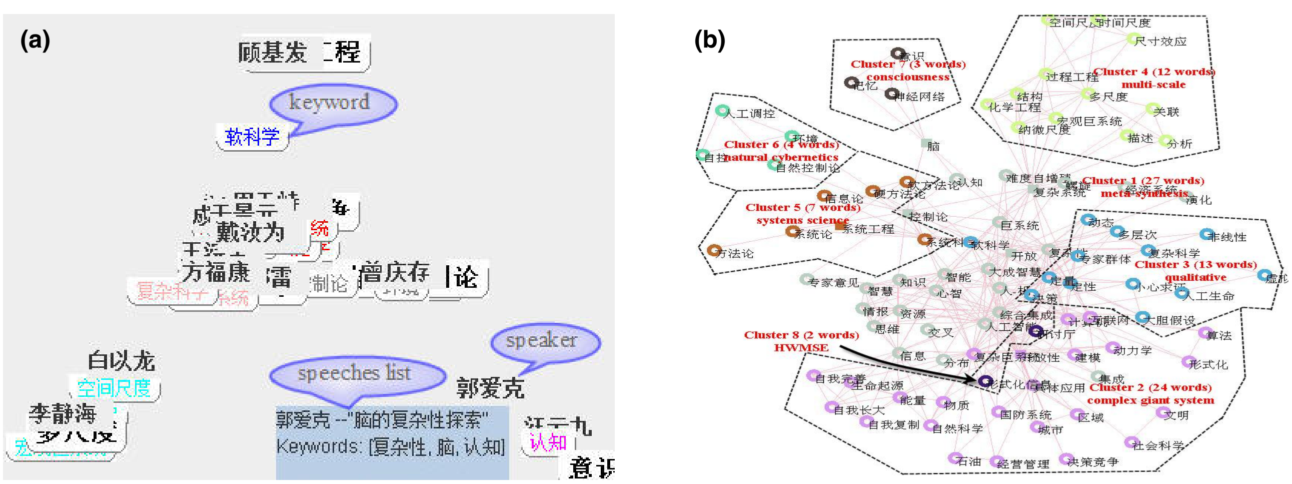


**Figure 2 Keyword clustering of XSSC complex related workshops (a) *K*-means clustering result of keywords (*K*-=7); (b) graph clustering of keyword network (8 clusters, Q=0.513)**

**Table 1 Keyword list of the same labeled clusters**

| Keyword list of each cluster | *K*-means clustering | Keyword network clustering |
|---|---|---|
| **Cluster labeled** "多尺度 *multi-scale*" | 多尺度，关联，时间尺度，空间尺度，纳微尺度，分析，描述，尺寸效应，过程工程，化学工程，宏观巨系统 非线性 | 多尺度，关联，时间尺度，空间尺度，纳微尺度，分析，描述，尺寸效应，过程工程，化学工程，宏观巨系统， |
| **Cluster labeled** "意识 *consciousness*" | 意识，神经网络，记忆 认知，脑 | 意识，神经网络，记忆 |
| **Cluster labeled** "自然控制论 *natural cybernetics*" | 自然控制论，人工调控，自控，环境 控制论 | 自然控制论，人工调控，自控，环境 |

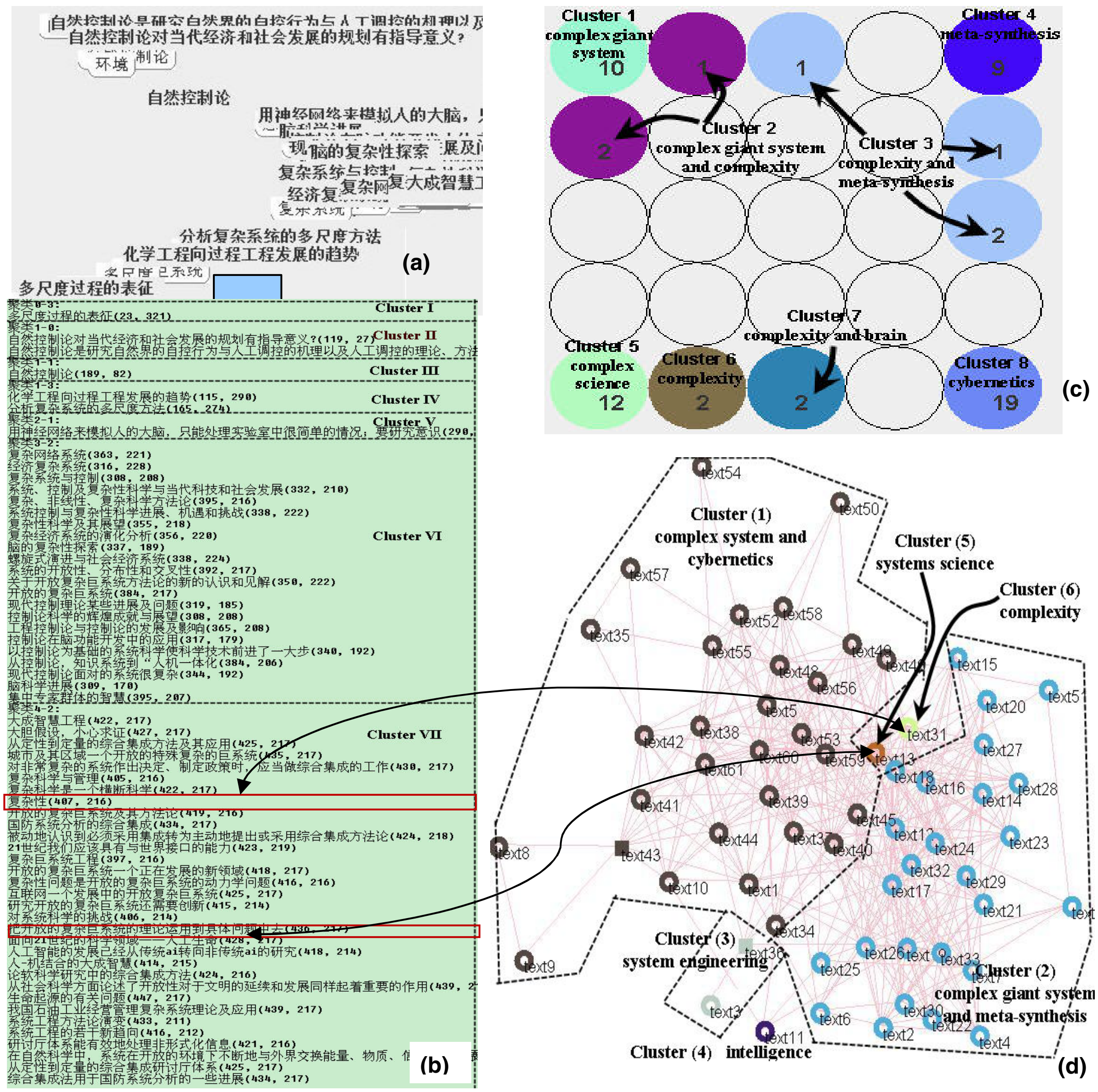


**Figure 3 Some speech clusters of XSSC workshops (a) a CorMap of speech and keyword; (b) automatic affinity diagramming of speeches (5×5 grids); (c) SOM clustering of speeches (training times=10000, learning rate=0.095, radius=4); (d) graph clustering of speech network(6 clusters)**

### 3.3 Clustering of speeches

Speech grouping results of "*complex systems and complexity research*" topic in the XSSC are shown in Figure 3. Based on the SVD results of the speeches and keywords frequency matrix, the correspondence structure of each speeches and keywords are shown in the Figure 3 (a). Automatic affinity diagramming of speeches is displayed in Figure 3 (b). 7 affinity groups are found under input 5×5 grids. Cluster VI and VII are the two biggest groups. Figure 3(c) is the result of SOM clustering under the input of initial learning rate=0.095, initial radius=4, 5×5 grid and training number=10000. 8 clusters of speeches (in different color) are generated by SOM clustering. The number in the colored circle shows the number of the speeches that are mapped into the node. Here those clusters are tagged as *"complex giant system"*(1)*, "complex giant system and complexity"* (2)*, "complexity and meta-synthesis"*(3)*, "meta-synthesis"*(4)*, "complex science"*(5)*, "complexity"* (6)*, "complexity and brain"*(7) *and "cybernetics"*(8) by human analysts. Both Cluster 1 & 2 may be combined into one cluster further accordingly.

As shown in Figure 3(d), Text 36 and Text 43 are the cutpoints of the speech network; 6 clusters are detected in the speech network (keyword-sharing speech network), actually a component including 2 biggest clusters which are labeled as *"complex system and cybernetics" and "meta-synthesis and complex giant system"* respectively by iView analysis. The other 4 clusters only contain few speeches does not mean those ideas are not important. Cluster 5 includes Text 13 presented by *Song Jian* and Cluster 6 includes Text 31 by *Qian Xuesen*. Those speeches may represent novel or special ideas which are not understood by the majority. Text 12 proposed by *Qian Xuesen* on complex giant system is the biggest betweenness node in the network. Here the analytical results are very good since those important or special speeches are actually given by influential systems scientists in China.

### 3.4 A brief comparison of ideas clustering

The effectiveness of different clustering methods should be considered. We find that visualization results of SVD & *k*-means clustering may show the keywords and speakers' correspondence structure, but may hide some keywords and speakers on common display screen. Some selected keywords of *k*-means clustering may show the novel ideas of the concerned topic, such as *consciousness*, *multi-scale and wisdom*. The cluster labels given by keyword network clustering may display the main point of those speeches among "*complex systems and complexity research*" workshops, such as *complex systems, complex giant system and meta-synthesis*. The results of keyword network structure may help us to find the connection of speeches.

Table 2 shows different results for speech grouping. Three different methods are applied in this dataset. Labels given by machine are speeches by *graph clustering and SOM*. And the *automatic affinity diagrams* cannot give labels of cluster automatically. However, analysts can make further review and summary based on their understanding and experience. Automatic affinity diagram is simple and cannot show the important or central idea of the argumentation topic. Results of SOM provide a different discuss perspective of cluster labels, for example, in Figure 3(c), the Cluster 2 and Cluster 1 are about the idea of *"complex giant system"*, and speeches in Cluster 2 discuss the *"complex giant system"* in different perspective. However, SOM clustering results are not always stable when the dataset is small scale because results are strongly correlated to the initial parameters.

**Table 2 A comparison of different clustering of speeches in XSSC workshops**

| Clustering methods | Clusters # | Cluster labels by human analysis | F Measure | Entropy | Sum of the square errors (SSE) |
|---|---|---|---|---|---|
| **Automatic affinity diagram** | 7 | *complex system, meta-synthesis, multi-scale, system engineering and consciousness* | 0.370 | 0.367 | 58.88 |
| **SOM** | 8 | *complex giant system, meta-synthesis, complex system, complexity and cybernetics* | 0.487 | 0.363 | 18.07 |
| **Idea network clustering** | 6 | *complex system and cybernetics, system engineering, systems science, complex giant system and meta-synthesis, intelligence and complexity* | 0.361 | 0.396 | -- |

(Note: automatic affinity diagram (5×5 grid), SOM (learning rate=0.095, initial radius=4, 5×5 grid and training time=10000))

Generally, F-Measure, entropy and Sum of the Square Errors (SSE) are widely used measures of clustering results of the quality of clustering [9]. Based on clustering by a human expert (16 clusters are detected), F Measure and Entropy measure of SOM, automatic affinity diagram and graph clustering are given in Table 2. SOM performs best (with the highest F Measure and the lowest Measure). Clustering is to summarize those pieces of ideas, then the role of human is more important. At this point, quantitative measures of those methods may not so important especially visualized clustering of ideas is just a kind of exploratory analysis. The explanation of clustering results is more interesting when depends on human's understanding. On the other hand, SSE may serve as a kind of indicator of convergence of the group discussion.

## 4. Conclusion

Currently, both CorMap and iView analysis provide visualized analytical results about the ideas proposed by group discussion. Clustering in the CorMap and iView analysis may give different perspectives of idea and help participants to understand and capture the main or interesting points during the discussion process. In this paper, we address 4 different kinds of clustering that are applied to either CorMap or iView analysis, i.e. *k*-means clustering, automatic affinity diagram, SOM and graph clustering, and show their practical study toward a scientific workshop in acquiring the main points of that scientific discussion and a system vision of the concerned topics.

The effectiveness of different clustering needs to be considered. As the scale of the dataset and initial parameters of clustering methods will affect the results, quantitative measures of clustering is not always necessary for idea clustering of a practical group discussion. Clustering results are reviewed and analyzed based on humans' understanding and experience. Detection of structures of ideas is more interesting. The analytical results of the case study shows interesting information detected from such a scientific think tank. Those results can be pushed to more scientists and interested people together with the summary proposed by workshop's secretary to stimulate further thinking and wider discussions.

As blog, micro blog, and social networking have become popular, those technologies such as CorMap and iView may be embedded into those Web 2.0 tools to help acquire information both efficiently and effectively.

## Acknowledgement

This work is supported by the National Basic Research Program of China (973 Program) under Grant No. 2010CB731405.

## References

[1] X. J. Tang, "Approach to Detection of Community's Consensus and Interest", in *Proceedings of APWeb'2008 Workshops (Y. Ishikawa et al. eds.), LNCS 4977*, Springer-Verlag , Berlin Heidelberg, 2008.

[2] X. J. Tang, "Qualitative Meta-synthesis Techniques for Analysis of Public Opinions for in-depth Study", *Complex 2009, Part II, LNICST 5*, Springer-Verlag, Berlin Heidelberg, 2009.

[3] Han J. W. and M. Kamber, *Data Mining: Concepts and Techniques (2nd edition)*, Morgan Kaufmann, San Fransisco, 2006.

[4] Wikipedia, "Affinity Diagrams", *http://en.wikipedia.org-/wiki/Affinity_diagram.* (Retrieved June 20, 2010)

[5] T. Kohonen, *Self-Organizing Maps (3rd edition)*, Springer-Verlag, Secaucus, 2001.

[6] S. E. Schaeffer, "Graph Clustering", C*omputer Science Review,* vol 1, no.1, 2007, pp27-64.

[7] X. J. Tang and Y. J. Liu, "Computerized Support for Qualitative Meta-synthesis as Perspective Development for Complex Problem Solving", in *Creativity and Innovation in Decision Making and Decision Support*. Decision Support Press, London, 2006, pp432-448.

[8] X. J. Tang, Y. J. Liu and W. Zhang, "Augmented Analytical Exploitation of a Scientific Forum", Studies in Computational Intelligence, Springer, Berlin, 2008.

[9] Tan P.N., M. Steinbach, V. Kumar, *Introduction to Data Mining*. Pearson Addison-Wesley, Massachusetts, 2006.