EXPLORATION OF TCM MASTERS KNOWLEDGE MINING*

Xijin TANG · Nan ZHANG · Zheng WANG

Abstract Traditional Chinese medicine (TCM) has a rich knowledge about human health and disease by its special way evolved along a very long history. As modern medicine is achieving much progress, arguments and disputes toward TCM never end. To avoid losing precious knowledge of living TCM masters, endeavors have been engaged to systematic collection of those knowledge of TCM masters, such as their growth experiences, effective practical cases toward diseases and typical therapeutic principles and methods. Knowledge mining methods have been expected to explore some useful or hidden patterns to unveil some mysteries of the TCM system. In the paper, some computerized methods are applied toward those collected materials about some living TCM masters in China mainland to show a different way of exposing essential ideas of those TCM masters by correspondence visualization which aims to help people understand TCM holistic views toward disease and body, and facilitate tacit knowledge transfer and sense-making of the essence of TCM. The work is one kind of qualitative meta-synthesis of TCM masters' knowledge.

Key words Idea map, knowledge mining, qualitative meta-synthesis, traditional Chinese medicine.

1 Introduction

Analysis is one of salient features of modern science. The analytical approach is the very foundation of modern medicine. Allied to the notion of analysis are the techniques of quantification and the idea of causality. Analysis is far less important to traditional Chinese medicine (TCM), which views human health and disease in terms of functional entities and disease-causing influence that are observed with the naked senses. "Its sophistication lies in its observation of correspondence between gross phenomena, and its organization of these observations through holistic systems of yin-yang and five phases"[1]. Qualitativity and holistic correspondence are two principal features in TCM whose basic concepts seem very simple while on the other hand create difficulties in applying them to practical situation. In TCM, the body is conceived as a system, rather than a machine. TCM diagnosis requires the identification of subtle variations of the working body and assessment of their significance in relation to each other. This is usually done by qualitative synthesis, rather than analytical reasoning. Ability to synthesize a host of subtle clues into a clear image of human's condition, to make differentiation

Xijin TANG · Nan ZHANG · Zheng WANG
Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, *China*. Email: xjtang@amss.ac.cn; zhangnan@amss.ac.cn; wzheng@amss.ac.cn.

of syndromes and then to give corresponding prescriptions is the mark of an experienced TCM doctor.

Due to complicated reasons, TCM is confronting difficulties in its own development in comparison to the modern medicine, which challenge current TCM education mode. Knowledge transmission of TCM meets problems, and even lots of precious knowledge of TCM masters are losing. Endeavors have been taken to save those masters' tacit experiences by systematically organizing to collect knowledge of those living TCM masters, such as their growth experiences, effective practical cases toward disease, typical therapeutic methods, principles, and prescriptions. On the other hand, a variety of information processing technologies have been applied to different facets of TCM research to explore some patterns or laws. Among those, data mining, text mining, and ontology are widely studied[2−9]. Those researches require lots of datasets or prerequisites for data mining. In this paper, instead of concerning IT applications to TCM research, the delivery of TCM masters' knowledge is firstly addressed as a knowledge conversion process where new insights may be acquired by TCM followers. This can also be achieved by qualitative meta-synthesis. To facilitate knowledge conversion and practice of qualitative meta-synthesis, some textual computing approaches are applied to exposing basic concepts or constructs of TCM masters' thoughts by visualizing possible correspondence, and are integrated into a computerized tool, TCM Master Miner, which may be applied to the meta-synthetic engineering of TCM knowledge conversion. Firstly, TCM knowledge transfer and meta-synthetic support are addressed.

2 TCM Knowledge Transfer and Meta-Synthetic Support

The ability of holistic correspondence and synthesis of determination of therapy by differentiation of syndromes can only be gained through a very long practice. Thus new TCM college graduates still require apprentice training after 5-year institutionalized learning to get familiar with at least 800 herbs and 120 standard traditional formulas of prescriptions. Usually junior medical doctors copy prescriptions (explicit knowledge) for their mentors during daily practice for rather a period of time to gain the ability of holistic observation. That is one of biggest differences in education between TCM and the modern medicine. Through learning and practice under guidance of the experienced TCM doctors, less experienced medical doctors may gradually sense the insights of their mentors' know-how by careful observation and practice based on their own institutional TCM knowledge, an indication of masters' tacit knowledge transfer to students' own knowledge, which could be regarded as a normal SECI (socialization, externalization, combination, and internalization) process of knowledge conversion proposed by Nonaka and his colleague[10]. Figure 1 shows the SECI model of TCM knowledge conversion. The cycle of TCM knowledge conversion lasts longer. Mass production of modern physicians is impractical to train genuine TCM doctors.
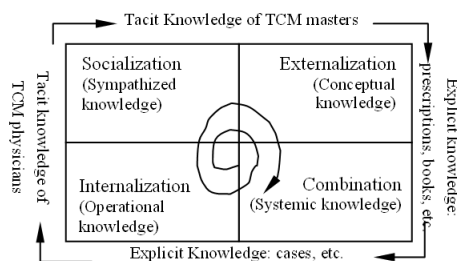

Figure 1 SECI Model about TCM Knowledge Conversion

To enable effective knowledge conversion, ideas of computerized support are naturally adopted to TCM knowledge conversion process to help less experienced medical doctors or even non-professionals to understand those TCM masters' thoughts easier, i.e., to acquire the essential thoughts, especially the mechanism of qualitative correspondence in diagnosis and treatment. Such kind of supporting tools is expected to bring new threads for association and expand human's thinking space. If recognition of disease is an unstructured problem, the particular diagnosing way of TCM is a problem structuring process through which a clear image of the patient with a context of evidences of whole working body is acquired. And those computerized aids are expected to visualize the perspectives or structures of those TCM masters' diagnosis. Here we regard TCM masters belong to human expert system where human mind, humans' advantages in imagery thinking, experiences, and intuition, etc. are denoted as qualitative intelligence, as compared with the powerful computing capabilities of computers which are referred as quantitative intelligence. Then it is expected to take the advantages of both TCM masters in qualitative intelligence and machine system in quantitative intelligence to generate new TCM knowledge, detect hidden patterns, unveil mysteries and instill new and validated knowledge into TCM knowledge system, which may be denoted as a meta-synthetic TCM engineering for TCM knowledge creation and conversion based on the ideas of meta-synthesis system approach and its practicing platform - Hall of Workshop on Meta-Synthetic Engineering (HWMSE) proposed by Chinese system scientistXuesen Qian (Tsien Hsueshen)[11]. The diagnosing process of TCM is actually one kind of qualitative meta-synthesis, i.e., to find assumptions or hypotheses about problems (syndromes) for further actions (treatment). Then those computing aids are mainly oriented to qualitative meta-synthetic support for exposing structures along an unstructured problem-solving process.

Among various developed supporting tools, group argumentation environment (GAE) is specifically designed to support divergent group thinking and qualitative meta-synthesis by versatile ways, such as visualization of expert opinion structure, clustering of contributed opinions, various analysis about participation, etc., and it has been applied to various conference mining and complex problem solving[12−13]. However, few group activities such as conference exist in normal TCM practice. Even TCM people may prefer applying GAE, and it is inappropriate to explore TCM masters' thoughts by GAE which is originally dedicated to grasp the active interaction between participants during group discussing process for procedural rationality of problem-solving process. Then a TCM Master Miner is designed with improvements in analytical technologies.

3 TCM Master Miner for TCM Masters' Knowledge Mining

Current explorations fulfilled by TCM Master Miner are mainly based on those materials contributed by living TCM masters. One piece of a TCM master's thoughts can be expressed by a structure as

$$\langle\text{master's name, text of thoughts, keyword set}\rangle,$$

which indicates that a master expresses his thoughts by a text (one sentence) with a set of keywords. The keyword set are manually selected by domain people according to the related text in the context. Those keywords can indicate the origin of thoughts (eg. names of ancient TCM masters) and academic background of the master, principles of differentiation of syndromes, therapeutic approaches, and prescriptions.

3.1 Framework of TCM Master Miner

Figure 2 shows the functional framework of current TCM Master Miner. Based on simple

representation of thoughts, a variety of explorations toward those masters are provided in TCM Master Miner, such as

- Visualization of correspondence between masters and their academic thoughts by exploratory analysis;

- Clustering of masters' academic thoughts and key concept extraction based on correspondence analysis;

- Visualization of masters' knowledge structure by idea network;

- Comparisons between TCM masters, such as dominance, agreement, and discrepancy, etc.
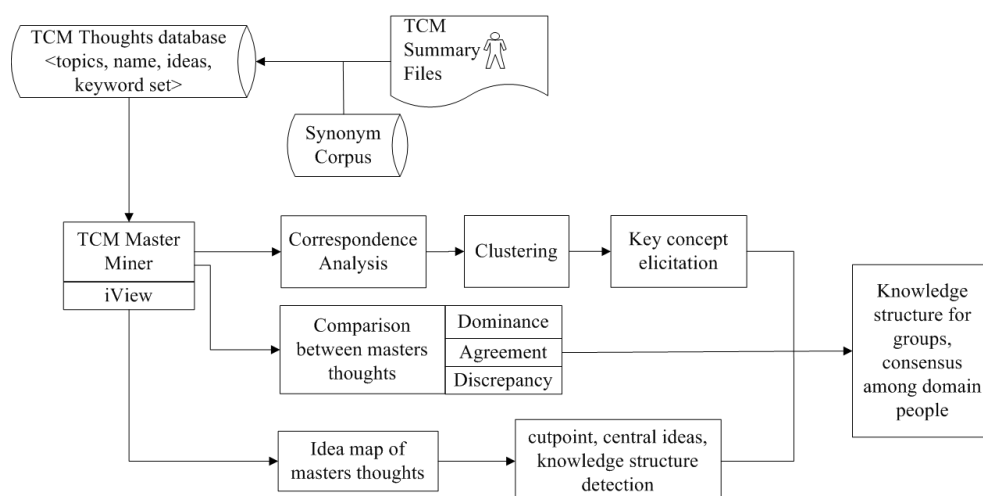

Figure 2 Framework of TCM Master Miner (v1.0)

Next, the mechanisms of two feature functions are explained briefly. Those functions may be referred as textual computing approaches toward visualization of human ideas.

3.2 Visualization of Correspondence Between Masters and Their Academic Thoughts by Exploratory Analysis

This function mainly adopts correspondence analysis which provides a method of factoring categorical variables and displaying them in a property space which maps their association in 2 or more dimensions. This method has been widely used in many disciplines[14].

Given the records of TCM masters' thoughts, a frequency matrix $F = (f_{ij})$ between masters and their academic thoughts can be acquired, where f_{ij} denotes the frequency of keyword j referred by the master i, $i = 1, 2, \cdots, m$; $j = 1, 2, \cdots, n$. The keywords are articulated as attributes of the masters.

Let the grand total of F be t and the correspondence matrix, or matrix of relative frequencies, be P so that the (i, j)-th cell entry is $p_{ij} = \frac{a_{ij}}{t}$. Define the i-th row marginal proportion by $p_{i\cdot} = \sum_{j=1}^{n} p_{ij}$ and the j-th column marginal proportion as $p_{\cdot j} = \sum_{i=1}^{m} p_{ij}$. The row marginal proportions are called row masses and the column marginal proportions are called column masses. Consider $\alpha_{ij} = \frac{p_{ij}}{p_{i\cdot} p_{\cdot j}}$ be the independence coefficient of p_{ij}. Then the chi-squared

statistics can be expressed by $\chi^2 = t \sum_{i=1}^{m} \sum_{j=1}^{n} p_i. p_{.j}(\alpha_{ij} - 1)^2$, and its degree of freedom is $(m-1)(n-1)$.

Correspondence analysis considers χ^2/t – referred to as the total inertia of the contingency table – to describe the level of association, or dependence, between two categorical variables. By decomposing the total inertia, we can identify important sources of information that help describe this association. The most common type of decomposition used in correspondence analysis is singular value decomposition (SVD).

In TCM Master Miner, the SVD is performed on a matrix A after several transformations of matrix F. The matrix A is decomposed into components by $A = U \Sigma V^{\mathrm{T}}$, where U is an m-by-m unitary matrix containing a set of orthonormal vectors called row singular vectors, Σ is m-by-n matrix where the diagonal elements $\Sigma_{i,i}$ are nonnegative singular values sorted in descending order, and V^{T} denotes the conjugate transpose of V, an n-by-n unitary matrix containing a set of orthonormal vectors called column singular vectors.

And the total inertia can be written in terms of the singular values such that $\frac{\chi^2}{t} = \sum_{i=1}^{M^*} \lambda_i^2$, where $M^* = \min(m, n) - 1$ and λ_i is the ith positive singular value. The keywords are mapped into 2-dimension space using the largest two singular values, so the 2D space can account for $\frac{(\lambda_1^2 + \lambda_2^2)}{(\chi^2/t)}$ of total variation.

Given the spatial relations acquired by correspondence analysis, a variety of clustering approaches such as k-means clustering can then be applied to ideas clustering and concept extraction for qualitative meta-synthesis. Instead fully depended on human's judgment of clustering numbers, a method that determines the true number of clusters is applied[15]. This method is based on distortion, a measure of within cluster dispersion, which is from the field of rate distortion theory in the information theory. Let X be a p-dimensional random variable with a mixture distribution of G components, each with covariance Γ; let $c_1, c_2, \cdots, c_K$ be a set of candidate cluster centers; c_X be the one closest to X. The minimum achievable distortion associated with fitting K centers to the data is

$$d_K = \frac{1}{p} \min_{c_1, c_2, \cdots, c_K} \mathrm{E}[(X - c_X)^{\mathrm{T}} \Gamma^{-1} (X - c_X)],$$

which is simply the average Mahalanobis distance, per dimension, between X and c_X. In practice, the covariance matrix Γ is rarely known, and then Γ is ignored, and Eculidian distance rather than Mahalanobis distance is used.

The computing process is as follows:

1) Run the k-means algorithm for different numbers of clusters, K, and calculate the corresponding distortion, d_k;

2) Select a transformation power, $Y > 0$;

3) Calculate the "jumps" in transformed distortion, $J_K = d_K^{-Y} - d_{K-1}^{-Y}$;

4) Estimate the number of clusters in the dataset by $K^* = \arg\max_K J_K$, the value of K associated with the largest jump.

Such kind of analysis can be applied to any combination of available TCM masters, and may help to "drill down" into those masters' thoughts to detect some possible or emerging academic schools among those masters which may have never been realized or admitted before. If applied to an individual master, exploratory analysis may reveal personal thinking structure.

The goal of such visualization is for holistic thinking and knowledge delivery. For TCM students or nonprofessionals, it is easier to grasp the fundamental concepts of TCM from the visualized map of masters' academic thoughts, find interesting or strange ideas for further studies.

It should be indicated that the main goal of applying correspondence map in TCM Master Miner is to provide a basic or possible association between masters and their academic thoughts represented by keywords. As correspondence analysis is only a method for exploratory analysis, the visualized association is not a confirmatory one, and 2 dimensions may not visualize more than 75% of the association between masters and keywords. It is necessary to do further analysis instead of directly drawing conclusions from the visualized relevance. Such kind of association is actually as a catalyst for shared understanding and wider thinking. Interesting or strange ideas toward the relevance, especially those isolated ideas far away from the majority or seemingly unreasonable associations may lead to some in-depth investigation for curiosity, even those associations are not confirmatory.

3.3 Idea Map by Keyword Network

The clustering of the thoughts of the concerned masters by spatial correspondence provides perspectives of those masters, which is easier for novices to understand the major ideas of those masters. TCM Master Miner provides another way to detect structures of academic thoughts. Each textual record of the masters' thought has a set of keywords, which actually explain the basic constructs or ideas applied to the specific problem by the masters. Then a keyword graph $G_l = (K_l, E_l)$ of the lth record of the thoughts can be constructed where the vertex refers to a keyword $k_i \in K_l$ (K_l is the keyword set of the lth record), and if both keywords k_i and k_j belong to one record, then an edge exists between two vertices $e_{ij} = (k_i, k_j), i \neq j$, $e_{ij} \in E_l$ (E_l is the edge set). Each vertex is connected with all others at one keyword graph for one record of text. The aggregation of all keyword graphs of one master or a group of selected masters brings forward a topological keyword network, $G = (K, E)$, $K = \cup K_l = \{k_1^l, k_2^l, \cdots, k_m^l\}$, $E = \cup E_l = \cup \{e_{ij}\}, i, j = 1, 2, \cdots, m; i \neq j$. This map is a weighted undirected network where the weight of edge refers to the frequency of co-occurrence of keywords among all texts contributed by the master(s) and is referred as an idea map of the concerned master(s). Given such a network, more senses may be acquired by a variety of network analysis in detecting some features of the idea map, such as cutpoints, centrality of keywords, clustering of keywords, etc.[16] which may help expose different perspectives of the master's knowledge scope. For example, a cutpoint (articulation point) of a graph is a vertex whose removal increases the number of connected components[17]; then the cutpoint keyword in the idea network may reveal the important concept for holistic correspondence. So does the centrality analysis of the keyword vertex. With the keyword clustering by community structure detection, it may be of more senses about the major topics from those keyword clusters than only counting individual keywords with highest frequency.

Next some trials are given.

4 Practical Analysis of Some Masters' Thoughts Using TCM Master Miner

A national key technologies R&D program for TCM research was initiated to build a platform* and collect those renowned TCM masters' materials in China since 2005. It is a time-consuming process to upload personalized academic summaries at a standard format. With only very limited finished materials contributed by active TCM masters, original design of a variety of data mining could not be carried out. A different attempt is to try to analyze those small-size datasets of academic thoughts of a few masters by using TCM Master Miner (v1.0).

---

*http://tcm.wfcms.org/

4.1 Data Preprocessing

Before applying TCM Master Miner, data preprocessing is undertaken.

1) Select representative texts from each TCM master's contributing file, which is fulfilled by domain people;

2) Convert each selected text into the structure: <master's name, text of thoughts, keyword set> where each keyword denotes only one idea, syndrome, disease, diagnosis, or treating principle;

3) Put all structured records of the concerned master(s) together into a data base file;

4) Converge keywords by replacing some keywords with the available synonyms based on a synonym corpus of TCM masters' thoughts. The corpus is not a comprehensive one but growing with increasing TCM masters' materials. For example, TCM masters prefer to cite an ancient book; while sometimes they refer to its author. Then if the keyword is the book's name, it is replaced by the author's name.

With 8 TCM masters' materials, some testing is undertaken here to show basic features of TCM Master Miner in exposing different perspectives about those masters' thoughts, and help to experience the holistic view in TCM thinking.

4.2 Exploratory Analysis

Figure 3 shows a global correspondence structure of 8 selected TCM masters. It is easy to find that at the center of the map lies the keyword 内经 (actually denotes the famous TCM book Yellow Emperor's Inner Canon) which is surrounded by names of some famous ancient TCM masters (keywords). This reveals a basic fact that those living TCM masters mainly get basic ideas from the Inner Canon written in the beginning of the first millennium and from other ancient masters. Moreover, the specialty of some masters could also be speculated, for example both 郑新 and 胡建华 are experts on stomach and spleen disease according to their surrounding keywords. Here 4 experts, 郑新 (in the central area), 胡建华 (left above 郑新), 张珍玉 (close to the top) and 何任 (close to left border), are selected and their group structure is shown in Figure 4.

The absolute location of each expert is changed in Figure 4 while the relative location of each expert still maintains, which may infer a somewhat stable joint knowledge structure of these 4 experts. Further observation indicates that these 4 experts all treat stomach and spleen disease. Moreover, it could be noticed that those keywords at the central area of Figure 5 are 行气, 益气 and 理气, all related with 气 (qi, the dynamic product of the orchestration of muscle action, or the invisible but observable force that carries food downward or upward in the digestive tract), which also exposes the treatment principles applied by those TCM masters to stomach disease. With simple materials, basic principles of those TCM experts' thoughts are easily acquired.

With the spatial relations as shown in Figure 3, a centroid-based k-means clustering of keywords is undertaken. Then the algorithm mentioned in Section 3.2 for right cluster numbers is applied. As $k = 5$, we get the largest jump in transformed distoration $J_5 = 13.126$. The second largest jump is $J_6 = 4.627$. Then five clusters are supposed to be an appropriate clustering result as shown in Figure 5.

The keyword (whose label is of bigger size of fonts) which is closest to the centroid of the affiliated cluster could be regarded as the label of the cluster. For example, the keyword 解表 in Cluster No. 2 of 103 keywords located in the central area of Figure 5 is denoted as the representative of that cluster. Such a way actually reflects one kind of concept extraction. Analysts can check details of that cluster and assign a more appropriate label.
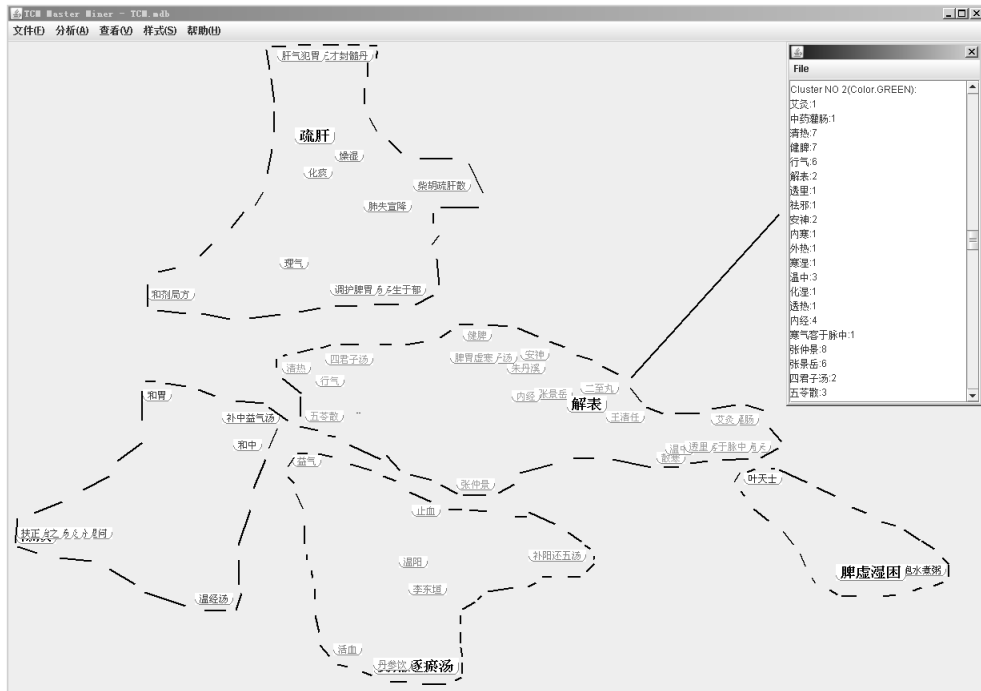
Figure 3 Visualization of 8 TCM masters' thought structure

Figure 4 Visualization of the selected 4 TCM masters' thought structure

Figure 5 Clustering of keywords of the 8 TCM masters' thought

4.3 Masters' Knowledge Structure

The idea maps of both individual master and a combination of masters can be generated. From available 14 records about one selected TCM master 张珍玉 ("Zhang Zhenyu") in the TCM master database, 57 keywords are extracted. Figure 6 is the keyword network of "Zhang Zhenyu" generated by TCM Master iView.

The idea map of "Zhang Zhenyu" has two components and 5 cutpoints existing in the biggest component, 柴胡疏肝散, 清热, 疏肝, 燥湿 and 通降 (from bottom right to up left) which connect 5 clusters detected with the highest measure of modularity ($Q = 0.676$) by Newman's fast algorithm[18]. The centrality of vertex is measured, too. For degree centrality of vertex, the top 5 vertexes are 通降 (24.00), 疏肝 (23.00), 理气 (18.00), 燥湿 (14.00) and 清热 (10.00). For centrality of betweeness, 疏肝 (600.33), 理气 (436.17), 通降 (419.83), 燥湿 (266.67), 清热 (192.00) and 柴胡疏肝散 (100.00) rank the top 6 vertexes, and betweennesses of all other vextexes are all zeros. Among those keywords, only 理气 is not the cutpoint. Except the prescription 柴胡疏肝散, all other keywords belong to treatment principles. By synthesis of these analyses, we may say those 6 keywords denote the constructs of knowledge system of the master "Zhang Zhenyu". We concentrate more on attaining some qualitative concepts instead of paying too much attention to the exact values of those centrality analysis of keywords.

Figure 7 is the collective keyword network of the selected 4 TCM masters whose knowledge correspondence is as shown in Figure 4. There are 35 clusters detected by Newman's fast algorithm ($Q = 0.766$), where the biggest component contains 6 clusters. 理气, 疏肝 and 益气 are cutpoints and of large values in centrality of betweenness, which may reflect their principal roles among those 4 masters' knowledge scope. Also they locate in the central area of Figure 4. Thus together with Figure 4, more senses could be acquired about the major treating principles
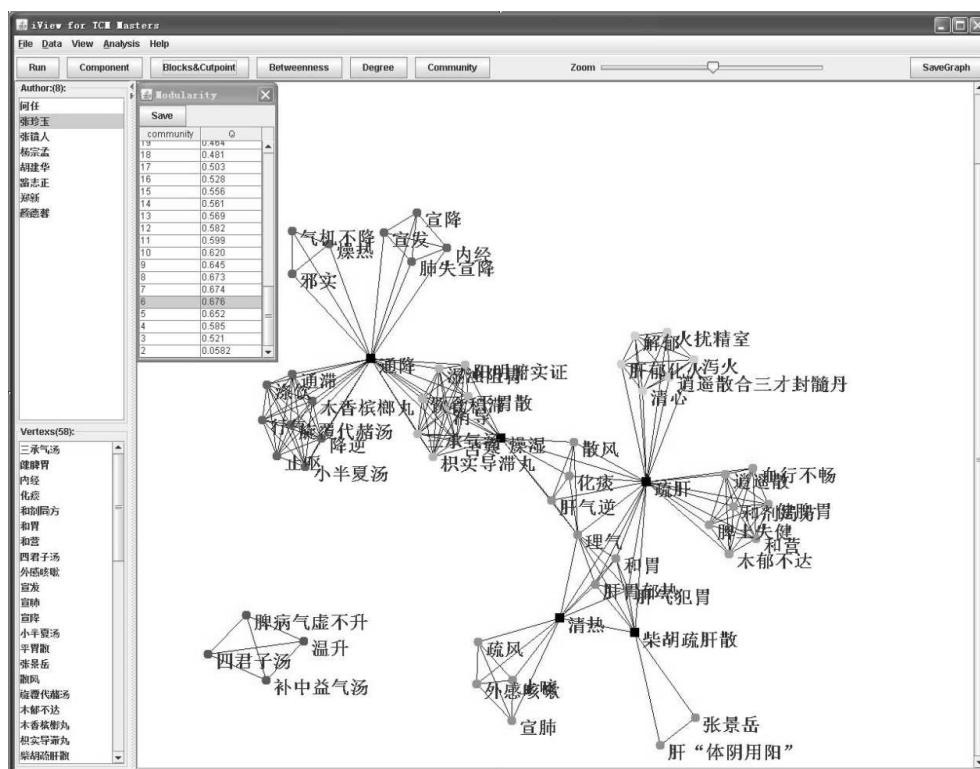
applied during long practice performed by the 4 masters.

Figure 6 One TCM masters' thought map via keyword network (cutpoint: black square vertex)

5 Concluding Remarks

With long history, traditional Chinese medicine confronts a lot of difficulties, such as the dissemination of its thoughts with its salient features in holistic thinking toward the working body and determination of treatment based on differentiation of syndromes. Many current TCM doctors even do not know how to apply the fundamental TCM approaches in their practice due to inappropriate modes of TCM education and practical training. This paper addresses conceptual modeling of the TCM knowledge conversion and proposes a computerized tool, TCM Master Miner, which aims to facilitate TCM knowledge conversion and qualitative metasynthesis during structuring process of masters' thought.

With integration of various methods, such as correspondence analysis, graph theory, and complex networks analysis, TCM Master Miner provides

- Perspective analysis of TCM masters' thoughts, which help people to acquire TCM basic scenario about the working body of human beings easier;

- Exploratory detection of possible academic schools of current TCM masters;

- Extraction of essential TCM masters' ideas;

- Awareness of unknown correspondence between different masters, between syndromes, diagnosis and treatment, etc.
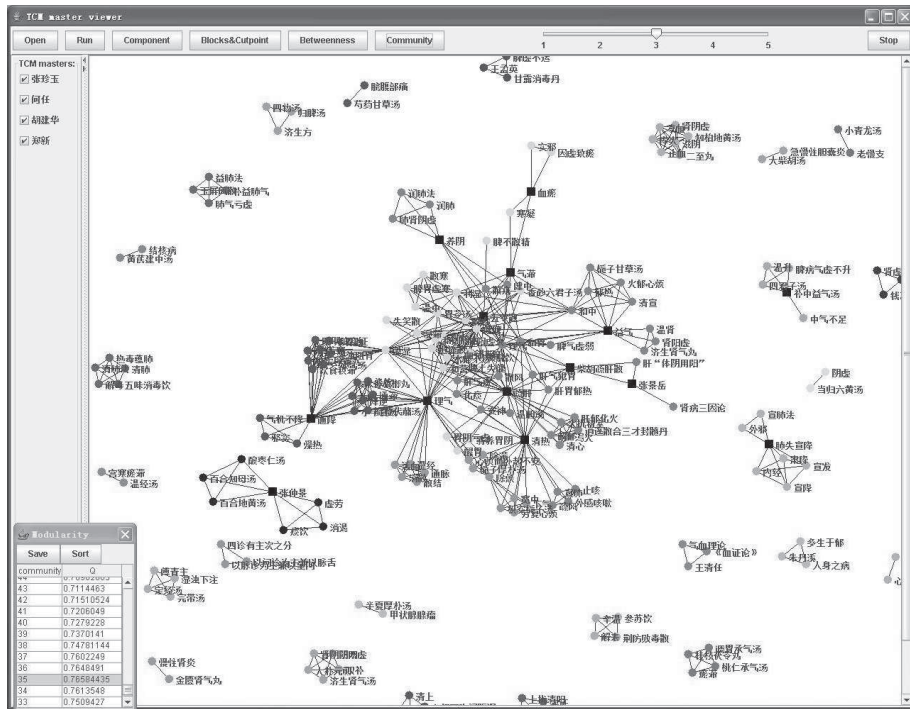
Figure 7 Four TCM masters' thought map via keyword network (cutpoint: black square vertex)

Many disputes are held towards the possibility of survival of TCM, while some western philosophers and system scientists even engage in disseminating TCM thoughts[19], which may bring stimuli to unveil mysteries of TCM.

TCM masters belong to TCM expert system and a variety of IT supports is regarded as machine system for quantitative computing and analysis. Both systems contribute knowledge to the TCM knowledge system which is being increasing and validating. Then these three systems (human expert system, machine system and knowledge system) construct a meta-synthetic system of TCM masters for TCM knowledge production. TCM Master Miner belongs to machine system and undertakes somewhat knowledge mining by exposing hidden structure of those TCM masters' thought and characteristics of basic TCM thinking, which may even reflect basic situations of current TCM diagnosis and treatment, and then help understand the situation of TCM in China in a right way.

Our current work is still at a very initial stage in both research and practice. Here we only show a very basic analysis using TCM Master Miner in exploring living TCM famous experts' thoughts. Lots of further work will be under exploration, such as expert group detection by considering the working location of TCM masters, consideration of more semantic meanings of keywords, such as origins or background of the thoughts, concerned patterns or syndromes, treating methods and principles or prescriptions in correspondence analysis, etc. It is also worth adopting or comparing with others' ideas[20]. Exploring a new linkage may be more difficult to be validated at TCM field. Besides, with more TCM masters' materials provided, more analysis will be undertaken for verification and validation of TCM Master Miner in practice.

Acknowledgment

basic data and help in segmentation of TCM terms in this study. We also thank Mr. Zhengwen ZHANG who developed a new version of iView for better performance of analysis.

References

[1] N. Wiseman and K. Boss, *Introduction to Glossary of Chinese Medical Terms and Acupuncture Points*, Library of Congress number 89–2982, Paradigm, Brookline, 1995.

[2] Z. G. Xiang, A 3-stage voting algorithm for mining optimal ingredient pattern of traditional Chinese medicine, *Journal of Software*, 2003, **14**(11): 1882–1890.

[3] Y. N. Sun, S. Y. Ning, M. Y. Lu, and Y. C. Lu, OLAP and data mining technology in decision supporting system for Chinese traditional medical diagnosis, *Computer Engineering* (in Chinese), 2006, **32**(9): 251–252, 255.

[4] H. Y. Liu, Y. F. Cao, and L. N. Qin, Knowledge acquisition method of traditional Chinese medical expert by case base on ontology, *Computers System and Applications* (in Chinese), 2005, (3): 80–83.

[5] J. F. Yan and W. F. Zhu, Apply rough sets theory in TCM syndrome factor diagnosis research, *Chinese Journal of Basic Medicine in Traditional Chinese Medicine* (in Chinese), 2006, **12**(2): 90–93.

[6] Y. H. Zhu and W. F. Zhu, Syndrome differentiation system of traditional Chinese medicine based on Bayesian network, *Journal of Hunan University* (*Natural Sciences Part*) (in Chinese), 2006, **33**(4): 123–125.

[7] C. G. Cao, Extracting and sharing medical knowledge, *Journal of Computer Science and Technology*, 2002, **17**(3): 295–303.

[8] C. Li, C. J. Tang, J. Peng, J. J. Hu, L. M. Zeng, X. X. Yin, Y. G. Jiang, and J. Liu, TCMiner: A high performance data mining system for multi-dimensional data analysis of traditional Chinese medicine prescriptions, in *Proc. ER Workshops* 2004 (ed. by S. Wang), LNCS 3289, Springer-Verlag, Berlin Heidelberg, 2004, 246–257.

[9] Z. H. Wu, X. Z. Zhou, B. Y. Liu, and J. L. Chen, Text mining for finding functional community of related genes using tcm knowledge, in *Proc. PKDD* 2004 (ed. by J. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi), LNAI 3202, Springer-Verlag, Berlin Heidelberg, 2004, 459–470.

[10] I. Nonaka and Y. Takeuchi, *Knowledge Creating Company*, Oxford University Press, New York, 1995.

[11] X. S. Qian, *Establishing Systematology* (in Chinese), Shanxi Science and Technology Press, Taiyuan, 2001.

[12] X. J. Tang, Towards meta-synthetic support to unstructured problem solving, in *Proc. of the 4th International Conference on Systems Science and Systems Engineering* (ed. by G. Y. Chen et al.), Global-Link, Hong Kong, 2003, 203–209.

[13] X. J. Tang and Y. J. Liu, Computerized support for meta-synthesis as perspective development for complex problem solving, in *Creativity and Innovation in Decision Making and Decision Support* (*Proceedings of IFIP WG* 8.3 *International Conference, CIDMDS'* 2006, ed. by F. Adam et al.), Decision Support Press, London, 2006, **1**: 432–448.

[14] E. J. Beh, Simple Correspondence Analysis: a bibliographic review, *International Statistical Review*, 2004, **72**(2): 257–284.

[15] C. A. Sugar and G. M. James, Finding the number of clusters in a dataset: an information-theoretic approach, *Journal of the American Statistical Association*, 2003, **98**(463): 750–763.

[16] R. A. Hanneman and M. Riddle, *Introduction to Social Network Methods*, University of California, Riverside, 2005. URL: http://faculty.ucr.edu/hanneman/nettext/.

[17] F. Harary, *Graph Theory*, Addison-Wesley, Reading, 1969.

[18] M. E. J. Newman, Fast algorithm for detecting community structure in networks, *Phys. Rev. E*, 2004, **69**(6): 066133.

[19] F. G. Wallner, *What Practitioners of TCM should Know*, Peter Lang Pub Inc, Berlin, 2006.

[20] Y. Matsuo, Y. Ohsawa, and M. Ishizuka, Keyworld: extracting keywords from a document as a small world, in *Proc. DS* 2001 (ed. by K. P. Jantke and A. Shinohara), LNAI 2226, Springer-Verlag, Berlin Heidelberg, 2001, 271–281.