The Complex Network Modeling Method and Its Application to the Knowledge System

Xi Yunjiang[1], **Dang Yanzhong**, **Wu Jiangning**
(Institute of Systems Engineering, Dalian University of Technology, Dalian, 116024, China)

ABSTRACT

In view of the system theory, the knowledge system is a system composed of different kinds of knowledge. Modeling knowledge system as a complex network is currently a hot topic in knowledge management research. However, it is very difficult to construct such a model because the elements and interrelations in knowledge system are often ambiguous and fuzzy. In this paper we aim to propose an approach to modeling the knowledge system by means of complex network theory and techniques. Firstly the knowledge system is divided into many knowledge fields, and each field can be divided into some sub-fields, and in the end the knowledge system can be divided into many small units that can be understood and limited in a small field and can be used to qualify the elements and interrelations of a knowledge system. Then the modeling method of knowledge points is proposed, in which knowledge naming and text mining methods are suggested, and then the interrelations between each pair of knowledge points are measured through comparison of their representation forms. The applications of the proposed network model are also presented in the paper together with a case study. The study shows that the proposed model can be applied in the representation, structure analysis, and classification of individual or organizational knowledge resources etc.

Keywords: knowledge network, complex network, knowledge system, knowledge management

1. Introduction

Many real systems can be described as complex networks with their elements being represented as nodes and interactions as edges [1-5]. Such networks can properly reflect and be used widely to analyze the features of many real systems. Frequently cited examples include the cell, a network of chemicals linked by chemical reactions, and the Internet, a network of routers and computers connected by physical links [3].

In view of the system theory, different knowledge that exists in human brain can form a system -knowledge system- with elements representing many different knowledge and interactions representing their interrelations.

However, it is very difficult to model the knowledge system as a complex network because of some special features of the system. For instance, it is very hard to qualify which knowledge is an element of a knowledge system, and what it is, let alone to qualify how many elements in it. The interrelations among different knowledge are also difficult to be identified. This means the knowledge system is an ambiguous and fuzzy system and can't be modeled and analyzed as a complex network directly in the usual way.

The goal of this paper is to propose a feasible method through which the elements and interrelations of knowledge system can be qualified and represented as nodes and edges, so that the knowledge system can be modeled and analyzed as a network, i.e. knowledge network. To qualify the elements, we divide different knowledge into a hierarchical structure which is composed of knowledge fields, sub-fields, knowledge units, and they are all called the knowledge points in the paper. A knowledge unit may refer to a very small part of knowledge that can be understood and limited in a small fields, and may merely relate to something, some facts or objects, some methods, tactics, skills and so on. And the interrelations between different knowledge points can be defined and qualified, therefore the knowledge network model can be constructed. The network formed in this way reflects the structure of knowledge system and can be applied in the representation, structure analysis, and classification of individual or organizational knowledge resources, which is significant in knowledge management.

2. Knowledge System and Its Features

Knowledge system is composed of many kinds of knowledge, which, in view of the system theory, are called 'elements', and may be reserved in documents, graphs, books, or in human brains, even in business processes, management regulations, products and services of an organization [6]. However, Knowledge system is quite different to physical or biological systems because of the following special features:

- The elements are ambiguous. The elements of knowledge system represent different kinds of knowledge. However, it is very difficult to identify an element in a knowledge system, let alone to specify "how many elements" in a system.
- The interrelations are ambiguous too. Interrelations of elements represent interrelations between different kinds of knowledge in the system. Owing

[1]Corresponding Author. Tel. 0086-0411-86927757. Email Address: xyj888@etang.com
Foundation Item：Project Supported by the National Science Foundation of China (No. 70271046)

to the ambiguity and fuzziness of knowledge, it's very difficult to judge whether there are interrelations between two knowledge points.

3. The Qualification of Elements and Interrelations of Knowledge System

To model a knowledge system, we need to qualify its elements and interrelations at first.

3.1. The Qualification of Elements

According to J. R. Anderson [7], a famous cognitive psychologist, there are two types of knowledge: declarative knowledge and procedural knowledge. The former is related to things, facts, or some objects and the latter is to approaches, skills, methods or strategies and so on. If we divide different knowledge into many small units, which may only refer to a very small part of knowledge that can be understood and limited in a small fields, and merely relate to something, some facts or objects, some methods, tactics, skills and so on, all the units may be distinguished from each other and can be qualified and identified according to the contents they are related.

The above mentioned method can be realized through the following process:

- To divide knowledge system into many knowledge fields;
- To divide each knowledge field into some small sub-fields;
- To re-divide each sub-fields into many smaller parts recursively until the sub-fields can't or needn't be divided. Then the knowledge system is divided into many small units of knowledge, which are called knowledge units in the paper.

The process can be shown as Fig. 1.

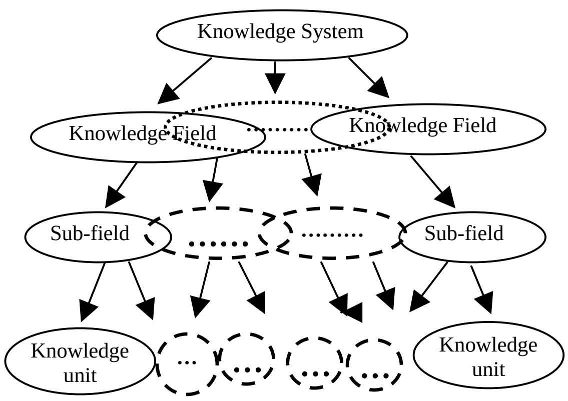

Fig 1 The Approach to Qualifying Elements of KS

Then the whole knowledge system can be viewed as a system composed of some knowledge fields, sub-fields, and many small units (shown as Fig. 2), which represents the hierarchical structure of knowledge system. To represent the hierarchical structure of different knowledge in the system, we mix the knowledge fields, sub-fields, and knowledge units together and view all of them as elements of knowledge system, and they are all called knowledge points hereafter in the paper.

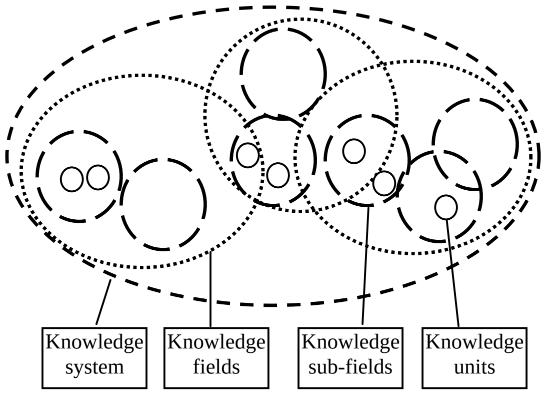

Fig. 2 The Composition of Knowledge System

Through this way, all the elements of a knowledge system can be qualified as knowledge points, and we can get two types of elements:

- Object-related elements: the knowledge points that may be related to different objects, things, facts and so on.
- Approach-related elements: the knowledge points that may be related to different approaches, skills, tactics and so on.

3.2. The Qualification of Interrelations

According to the contents that the knowledge points are related, i.e. the different things, facts, objects, approaches, skills, tactics and so on, the interrelations between different elements can also be qualified. Here the interrelation is defined as: the relationship between two knowledge points which exists when both elements are related to the same things, facts, objects, approaches, skills, tactics and so on. For instance, taking "Information Management" and "Information System" as two different knowledge points, both are related to the object "information", so we can say there is a relationship between knowledge point "information management" and "information system".

According to the contents that the knowledge points are related, the interrelations in knowledge system can be classified into two types: object-based interrelation and approach-based interrelation (shown in Table 1):

Tab.1 Types of knowledge interrelations

Knowledge Interrelation Types	Characteristics	Examples
Object-based Interrelation	related to same objects, facts, or things and so on.	Business-Process-Reengineer VS Process-Management
Approach-based Interrelation	related to same approaches, skills, tactics, strategies and so on	Genetic-Algorism VS Immune-Genetic-Algorism

Tab.2 Naming Methods for Knowledge Points

Naming Methods	Characteristics	Types of Knowledge Points	Examples
Object-based Naming	Objects, things, or facts that a knowledge point is related to are represented in name.	declarative knowledge	Information system, Workflow
Approach-based Naming	Approaches, skills, tactics that a knowledge point is related to are represented in name.	procedural knowledge	Genetic Algorism, Immune Genetic Algorism
Combined Naming	Both objects and approaches that a knowledge point is related to are represented in name.	declarative knowledge or procedural knowledge	Business Process Reengineering, Project Management

4. The Network Model and the Modelling Method for Knowledge System

4.1. The Network Model of Knowledge System

Based on the above analyses, the knowledge system is composed of knowledge points and the corresponding interrelations.

Let $K = \{k_1, k_2, \cdots, k_n\}$ denote the set of knowledge points, and $S = \{(k_i, k_j)\}$, $i, j = 1,2,\cdots,n$, denote the set of interrelations, then the knowledge system can be represented as: $KS = (K, S)$.

Its network model, i.e. knowledge network, in which nodes represent knowledge points and edges represent their interrelations, can be represented as:

$$G = (K, E) \qquad (1)$$

where $K = \{k_1, k_2, \cdots, k_n\}$ represents the set of nodes, $E = \{(k_i, k_j)\}$, $i, j = 1,2,\cdots,n$, represents the set of edges. Its adjacent matrix is:

$$R = \{(r_{ij})\}, \quad i, j = 1,2,\cdots,n,$$

where r_{ij} represents the weight of interrelation from nodes k_i to k_j.

4.2. The Network Modelling Method for Knowledge System

As shown in Eq. (1), the model of knowledge network is composed of knowledge points and their interrelations. So the modelling method for knowledge system includes two parts: the modelling method for knowledge points and their interrelations.

4.2.1. The Modelling Method for Knowledge Points

(1) The Representation Method for Knowledge Points

Knowledge point represents a part of knowledge. So we can conclude that its representation form means the name of knowledge point, which can be written in natural language of human beings, and usually shown as a few words or phrases, i.e.:

$$k_i = w_{i1}\_w_{i2}\_\cdots\_w_{im} \qquad (2)$$

where w_{im} denotes a word in nature language, and all the words stringed sequentially form a phrase, denoting the name of a knowledge point.

In fact, the representation form shown in Eq. (2) can be seen the name of a knowledge point in natural language. According to the objects or approaches that knowledge points are related to, there are three naming methods for knowledge points (shown in Table 2).

(2) The Acquisition Method for the Names of Knowledge Points

The naming methods can help us represent knowledge points as words or phrases in human language, which can be acquired through text-mining or directly naming by experts. To tacit knowledge, which is mainly reserved in human brains, the names can be acquired through directly naming; while to most explicit knowledge, which is mainly reserved in documents, the names can be acquired through text mining. Since much of expert knowledge can be mined from their published papers, here we mainly discuss how to acquire the names of knowledge points through text-mining.

Let $P = \{P_1, P_2, \cdots, P_m\}$ donate the set of documents of the published papers, and $P_i \in P$ can be represented as:

$$P_i = \{(k_{i1}, q(k_{i1}))\}$$

where k_{i1} denotes the name of a knowledge point, $q(k_{i1})$ donates the frequency that k_{i1} emerges in document P_i, then the set of knowledge points can be represented as:

$$K = \{(k_i, q_i)\} \tag{3}$$

where q_i is the weight of k_i, and can be calculated by:

$$q_i = \sum q(k_i) \tag{4}$$

Eq. (4) means the total weight of q_i equaling to the accumulation of frequencies that k_i emerges in each document.

4.2.2. The Modelling Method for Interrelations of Knowledge Points

The interrelation between two knowledge points means they are related to the same objects or approaches. Because of the ambiguity and complexity of knowledge, the works to qualify the interrelations have usually to be done by field experts. Here we present a method which may simplify the process to value the interrelations approximately.

According to the naming methods in Table 2, knowledge points may be named by some words or phrases that represent the objects or methods they are related. So we can conclude that:

Rule 1: There are interrelations between knowledge points k_i and k_j if there are identical words except some stop words in their names.

Here stop words refer to art words and preposition words, such as "a, the, by, of, through, with, for", and so on.

Let r_{ij} denote the interrelation from knowledge points k_i to k_j, and $l(k_i)$ denote the number of words except stop words in the representation form of k_i, $l(k_i, k_j)$ denote the number of identical words that emerge in the representation forms of both k_i and k_j except stop words, then the interrelations between k_i and k_j can be calculated as:

$$\left.\begin{aligned} r_{ij} &= \frac{l(k_i, k_j)}{l(k_i)} \\ r_{ji} &= \frac{l(k_i, k_j)}{l(k_j)} \end{aligned}\right\} \tag{5}$$

Through Eq. (5), the adjacent matrix can be acquired if all interrelations are calculated, and the network of knowledge system can then be constructed.

5. The Analysis of Knowledge Network Together with a Case Study

The above proposed modelling method can help us model a knowledge system as a complex network, i.e. knowledge network. Hereafter we'll analyze its features by a case study.

The case is on an expert who has domain knowledge in some research fields. Although he may have comprehensive knowledge in many fields, only those in his research fields are cared which is called domain knowledge and mainly reserved in his published papers. The domain knowledge of the expert can be viewed as a knowledge system, and be modelled as a network through the approach proposed in this paper. According to "Full text database of Chinese journals", the given expert has published total 24 papers that can be retrieved from the database by Dec.2004. Then we can acquire the model of knowledge points from the papers according to Eq. (3), which can be done through text mining and be represented as a set of many phrases and their weights. To simplify the process, we use key words of the papers as the names of knowledge points directly, and their weights are assigned as 1 in each paper, then their total weights can be calculated according to Eq. (4). According to Eq. (5), the adjacent matrix of the network can also be acquired, and the network of the expert domain knowledge can be constructed (shown as Fig. 3).

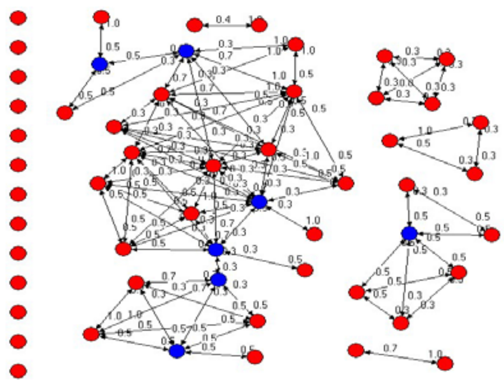

Fig.3 Knowledge network for a domain expert (In the paper all network figures are created in Ucnet6.71 [8])

As shown in the case study, there are some features of knowledge network which can be concluded from the modelling process:

- Knowledge network is a directed and asymmetric network. In the adjacent matrix, r_{ij} doesn't always equal to r_{ji}, which means the interrelation from k_i to k_j is not equal to that from k_j to k_i, so the network is directed and asymmetric.

- The edges have weights and the points have attributes. The former represents to what degree one knowledge point is similar to another; the latter represents how important that a knowledge point is in a knowledge system.
- There are many isolates and sub-groups in a knowledge network, and the largest sub-group is called component [9]. A sub-group is composed of many points that are linked together, which implies that there are direct or indirect interrelations among the nodes.
- There are many cliques in a sub-group. The definition of a clique is: a sub set of points, in which all pairs of points are directly connected by lines and the clique is not contained in any other clique [10].

Rule 2: All nodes that have common knowledge in a knowledge network may form a clique.

Here common knowledge means the identical words except stop words in all the names of the points in a clique, and is called “knowledge core” in the paper hereafter.

Rule 2 can be used to digest knowledge cores in a clique, and all knowledge cores can be mined in a knowledge network.

There may be some sub-cliques in a clique, and it can be defined as: A sub-clique is a sub set of nodes of a clique, in which a node links to any other node of the sub-clique with higher weight than it links to any nodes outside the sub-clique (but still in the clique), and the sub-clique isn’t contained in any other sub-clique.

Obviously, the definition of sub-clique is similar to that of clique. The Fig.4 shows a clique and sub-clique in our case study.

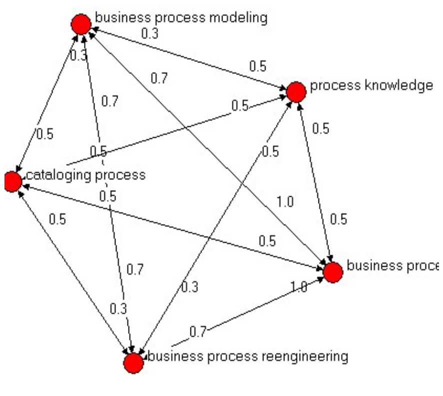

Fig.4 Cliques and sub-cliques in a knowledge network

In Fig.4, there are 5 knowledge points: process knowledge, business process, business process reengineering, business process modelling, cataloging process, each of which are represented as k_1, k_2, k_3, k_4, k_5 respectively, and the adjacent matrix of the clique is:

$$\begin{array}{c} \\ k_1 \\ k_2 \\ k_3 \\ k_4 \\ k_5 \end{array} \begin{array}{c} \begin{array}{ccccc} k_1 & k_2 & k_3 & k_4 & k_5 \end{array} \\ \begin{bmatrix} 0 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0 & 1 & 1 & 0.5 \\ 0.33 & 0.67 & 0 & 0.67 & 0.33 \\ 0.33 & 0.67 & 0.67 & 0 & 0.33 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0 \end{bmatrix} \end{array}$$

Rule 3: In an adjacent matrix of a clique, if all values of row i except $r_{ii} = 0$ don’t equal to each other, there must exist a sub-clique in the clique at least and k_i must be a member of the sub-clique.

Rule 3 can be used to discover sub-cliques and their members in a clique.

In a sub-clique, there is common knowledge that all nodes are related, and the common knowledge is composed of more identical words than the knowledge core of the clique. Hereafter we call it Knowledge Sub-Core in the paper.

Rule 4: In a clique, knowledge sub-cores, if exist, must be related to knowledge core, and the weight of interrelation from knowledge core to sub-core is 1.

6. The Application of Knowledge Network

Knowledge network model can be applied to knowledge management for individuals and organizations including representation and structural analysis of knowledge resource, knowledge classification and so on.

(1) Knowledge representation: as shown in Fig.1, the model can represent knowledge resources of an individual or organization. The attributes of knowledge

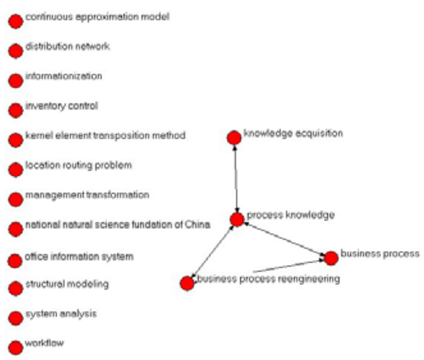

Fig.5 Knowledge Network (q_i >1)

points and weights of edges can be used to show knowledge network in different scales, as shown in Figures 5 and 6, which represent a same knowledge network as that shown in Fig.3, but are resized by attributes of nodes(Fig.5, q_i >1) or weights of edges (Fig.6, r_{ij} >=0. 4).

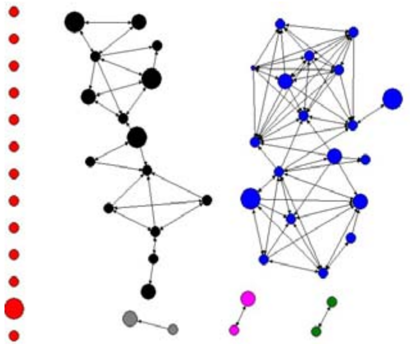

Fig.6 Knowledge Network (r_{ij} >=0.4)

(2) Structural analysis of knowledge resource: Knowledge network is composed of many isolates and sub-groups, including components, cliques, sub-cliques, which indicate some structural features of knowledge

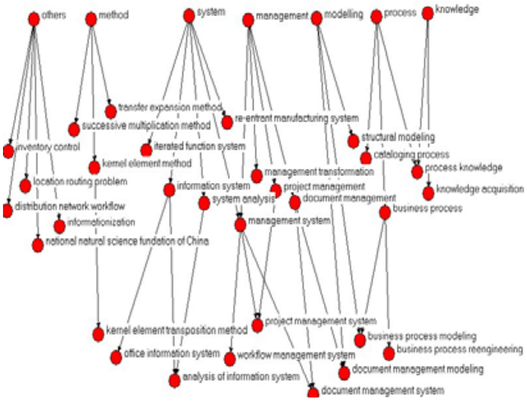

Fig.7 Hierarchy model of knowledge resource of an expert(r_{ij} =1)

Tab.3 Classification of knowledge resources for an expert ($q(KC) > 4, q(KSC) > 1$)

	Knowledge Core	Knowledge Sub-Core	Knowledge Points		
1	system	information system management system system analysis	iterated function system analysis of information system document management system analysis of information system	re-entrant manufacturing system office information system project management system	workflow management system
2	process	business process	cataloging process business process reengineering	process knowledge business process modeling	
3	management	document management management system project management	management transformation document management system document management system project management system	document management modeling project management system	workflow management system
4	knowledge		knowledge acquisition	process knowledge	
5	modeling		business process modeling	document management modeling	
6	method	kernel element method	successive multiplication method kernel element transposition method	transfer expansion method	
7	others		distribution network inventory control	workflow location routing problem	informationization national natural science foundation of China

resource. The hierarchal structure of individual or organizational knowledge resource can also be modelled as a network diagram (shown in Fig. 7, only edges with weights equalling to 1 are displayed), in which knowledge cores and sub-cores are added to the set of knowledge points and are treated as nodes of knowledge network.

(3) Classification of knowledge resource: The model can also be applied in knowledge classification. Through network analysis, cliques and sub-cliques can be obtained along with knowledge cores and sub-cores, which can be used to classify knowledge points into different classes according to cliques or sub-cliques they belong to (As shown in Table3, the case is same as that in Fig.3).

(4) The model can also be used to analyze the evolving and robust features of the knowledge system, which may be discussed in future studies.

7. Conclusion

In summary, the modelling approach to knowledge network is proposed in this paper, which can model a knowledge system as a complex network. The model can be applied to represent, analyze the structure, and classify knowledge resources that belong to an individual or organization, as is proved successfully in the case study.

Further work may focus on the evolving and robust features of knowledge system, which can also be modelled and analyzed on the basis of knowledge network model.

REFERENCES

[1]. D. J.Watts, S. H. Strogatz, Collective dynamics of small-world networks, Nature,393 (1998) 440-442.

[2]. A. L. Barab´asi, R. Albert, Emergence of scaling in random networks, Science, Science, 286, (1999) 509-512.

[3]. R. Albert and A.-L. Barab´asi, Statistical mechanics of complex networks, Rev. Mod. Phys., 74, (2002) 47-97.

[4]. S. N. Dorogovtsev and J. F. F. Mendes, Evolution of networks, Adv. Phys., 51, (2002) 1079-1187.

[5]. M. E. J. Newmann, The structure and function of complex networks, SIAM Rev., 45, (2003) 167-256.

[6]. Wang Zhongtuo. Knowledge System Engineering[M]. Beijing: Science Press, 2003. 21-21.

[7]. Anderson, J. R. Cognitive psychology and its implications, 2nd ed., New York: Freeman, 1985.

[8]. Borgatti S P, Everett M G, Freeman L C. Ucinet for Windows: Software for Social Network Analysis[M]. Harvard MA: Analytic Technologies, 2002.

[9]. John Scott. Social Network Analysis[M]. London: Sage Publications Inc, 2000. 100-101.

[10]. John Scott. Social Network Analysis[M]. London: Sage Publications Inc, 2000. 115-122.