Knowledge Mapping of ITS Theses and Dissertations in Taiwan Using Social Network Analysis

Chi-Chung Tao
Department of Transportation Management
Tamkang University
Tamsui District, New Taipei City, Taiwan
cctao@mail.tku.edu.tw

Jennifer Yuh-Jen Wu
Institute of Transportation
Ministry of Transportation and Communications
Taipei, Taiwan
jenny@iot.gov.tw

Ruei-Jhih Jian
Department of Transportation Management
Tamkang University
Tamsui District, New Taipei City, Taiwan
fantasy10540@yahoo.com.tw

Abstract— **Theses and dissertations underpinning a university's academic performance may not really reflect the actual knowledge flows. It is the informal networks that have played a critical role in getting important work done in universities. In order to better understand the knowledge flow through these informal networks, knowledge maps can be developed to illustrate the actual knowledge flows. Social network analysis is a technique that can be applied in building knowledge maps and can help analyze the strengths and weaknesses of the networks effectively. This paper provides a case study to illustrate the application of social network analysis to develop ITS knowledge maps for 10 universities in Taiwan. Borrowing and adapting techniques from other disciplines, such as social network analysis, needs to be done to push the new frontiers of ITS knowledge management.**

Keywords-Intelligent Transportation Systems; Knowledge Mapping;Social Network Analysis

I. Introduction

Recently, because of the development in information technology and the setup of many high-volume and high quality data sets of scientific publications such as theses and dissertations, the study of scientific network has attracted considerable attention. To enhance the knowledge flows between people to stimulate innovative thinking, organizations should first conduct a knowledge audit and develop a knowledge map of the sources, sinks, and flows of knowledge in the organization.

Once the knowledge maps are created, the managers can then review the flows and determine if there is any potential inefficiency in knowledge exchange and implement changes to improve the overall networks. However, without a systematic way of analyzing the networks, one may find it difficult to determine the networks' strengths or weaknesses. It is therefore beneficial to apply social network analysis (SNA) in analyzing the knowledge flows depicted by the knowledge maps. A number of software tools, such as NetMiner, UCINET, and NetDraw, have been developed to aid the analysis and provide the visualization of the networks.

Since 1990 Intelligent Transportation Systems (ITS) has become one of frontiers of transportation science in Taiwan, numerous theses, dissertations and research reports increase rapidly. This paper aims at introducing the importance of the application of SNA to knowledge mapping through a case study of ITS theses and dissertations in Taiwan. It first provides a literature review on the recent research and application of knowledge mapping and SNA. It is followed by a discussion of the concepts of SNA. A case study based on National Digital Library of ITS theses and dissertation in Taiwan is then used to illustrate how SNA is applied to develop knowledge maps for the case study [1].

II. Literature Review

A. Knowledge Mapping

A key part of knowledge management (KM) is performing a knowledge audit to determine knowledge flows within an organization. Dodge and Kitchin [2] classified two types of knowledge maps- manual and automatic knowledge maps:

1) Manual knowledge maps: Concept map and mind map are drawings in which blocks represent concepts or things and connecting lines represent relationships [3]. However, the manual approach is not scalable to the processing of large amounts of information, because a manual knowledge map is not only limited in scope and timeliness, but it is also slow and cumbersome. In order to create a knowledge map that closely matches a mental model, an intelligent, automatic categorization algorithm must be employed.

2) Automatic knowledge maps: Automatic knowledge maps can be categorized into three categories based on their knowledge characteristics:

a) Numerical: Visualization of numbers was among the earliest map applications. When the numbers have physical correspondence, the maps are easily understood. For example, the Internet's statistics on inbound and outbound traffic and the domain host's density can be readily layered on top of the physical infrastructure maps to reveal the growth and demographics of the Internet over time.

b) Textual: Mapping textual knowledge sources is more difficult than mapping numerical knowledge sources because text has limited spatial meaning but strong abstract or conceptual relationships. One example of mapping textual knowledge is Cartia's NewsMap [4], which applies sophisticated and proprietary lexical algorithms to analyze the content of text documents and the relations between them to distill the key topics and form the topic island map.

c) Social: The third kind of knowledge maps visualizes human relationships in a community. Rich knowledge sources may include numerical, textual, and relationship links. Social visualization research represents human behavior graphically.

B. Social Network Analysis(SNA)

A common framework for SNA is the mathematical approach of graph theory [5]. A social network is a set of actors (or nodes) that may have relationships (or ties) with one another. It is common to refer to standard graphs shown in Figure 1 since the actors and group centrality greatly varies in those graphs. A quick look shows that in Fig.1(b), all nodes are equally interchangeable and hence should be equally central, in Fig.1(a) one node completely outranks the others (which are interchangeable), and in Fig.1(c) centrality decrease for peripheral nodes.

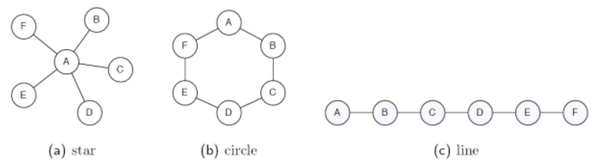

Figure 1. Network graphs illustrating centrality

There are two major concepts of describing network structures. These are centrality and substructures. Centrality is important to understand power, stratification, ranking, and inequality in social structures; whereas the idea of substructures (or subgroups) within a network is a powerful tool for understanding social structure and the embeddedness of individuals.

1) Centrality: Centrality is the measure of how close an individual is to the centre of the action in a network. The work of [6] yields to the use of the casual notation when it comes to actor centrality measures. C is a particular centrality measure, which will be a function of n_i, where subscript index i range from 1 to g. As there is different version of centrality, C will be subscripted with an index for the particular measure under study. Then centrality measure A of node i is $C_A(n_i)$. The group-level quantity is an index of centralization, and identify how variable or heterogenous the actors centralities are. The larger the group index is, the more likely there is one single very central actor in the network whereas other actors are considerably less central. Fig.1(a) is maximally central because one central actor has ties with all other actors. We define $C_A(n^*) = max_i C_A(n_i)$ as the largest value of index A among all actors in the network. The general group centralization index, called Freeman's index, is:

$$C_A = \frac{\sum_{i=1}^{g}[C_A(n^*) - C_A(n_i)]}{max\sum_{i=1}^{g}[C_A(n^*) - C_A(n_i)]} \tag{1}$$

a) Degree: Measuring the degrees of the actors in the network is the simplest and most straight-forward way to determine centrality. The degree is the number of an actor's direct ties. The actor-degree centrality index is defined as:

$$C_D(n_i) = d(n_i) = \sum_j x_{ij} = \sum_i x_{ji} \tag{2}$$

where $x_{ij} = 1$ if there is a link between nodes n_i and n_j .The measure is usually standardized with:

$$C'_D(n_i) = \frac{d(n_i)}{g-1} \tag{3}$$

According to the standard definition of group-level index, the group degree centrality index is then:

$$C_D = \frac{\sum_{i=1}^{g}[C_D(n^*) - C_D(n_i)]}{max\sum_{i=1}^{g}[C_D(n^*) - C_D(n_i)]} = \frac{\sum_{i=1}^{g}[C_D(n^*) - C_D(n_i)]}{(g-1)(g-2)} \tag{4}$$

It is 1 for star graph, and zero in a circle graph.

b) Closeness: Closeness focuses on how close an actor is to all the other actors in the network. Actor closeness index is a function of the geodesic distances. Define $d(n_i,n_j)$ be the number of links between n_i and n_j. Thus the actor closeness index is (5), the standard actor closeness index is (6) and the group closeness centralization index is (7):

$$C_C(n_i) = [\sum_{j=1,j\neq i}^{g} d(n_i,n_j)]^{-1} \tag{5}$$

$$C'_C(n_i) = \frac{g-1}{\sum_{j=1,j\neq i}^{g} d(n_i,n_j)} = (g-1)C_C(n_i) \tag{6}$$

$$C_C = \frac{\sum_{i=1}^{g}[C'_C(n^*) - C'_C(n_i)]}{[(g-2)(g-1)/(2g-3)]} \tag{7}$$

c) Betweenness: Betweenness is the extent to which a particular actor lies between the various other actors in the network. A fundamental assumption is that all geodesics (i.e. shortest paths) are equally likely to be used. The actor betweenness index is (8) and the standard actor betweenness index is (9):

$$C_B(n_i) = \sum_{n_s \neq n_t \neq n_i} \frac{g_{st}(i)}{g_{st}} \tag{8}$$

$$C'_B(n_i) = \frac{C_B(n_i)}{(G-1)(G-2)} \tag{9}$$

with g_{st} number of geodesics linking node *s* and node *t*, g_{st} (n_i) if number of geodesics linking the two actors that contains actor *i*, and *G* being the total number of actors *G*. Actor betweenness index is an unbounded number whereas standardized actor betweenness index range in *[0,1]* with *1* when node is maximally central (e.g. middle node in a star graph) and *0* if it on no geodesics (e.g. edge node in star graph). In order to measure the centrality on the whole graph, the group betweenness centralization index is thus defined as:

$$C'_B(n_i) = \frac{C_B(n_i)}{(G-1)(G-2)} \tag{10}$$

with $C_B(n^*)$ largest realized actor betweenness index for the set of actors. Group betweenness index allow to compare different networks with respect to the heterogeneity of the betweenness centrality of the members of different networks. Maximum value, unity, is reached for a star graph. Minimum value, zero, occurs when all actors have the same actor betweenness index (e.g. a circle graph).

2) Substructures: There are basically two approaches: bottom-up and top-down [7,8].

a) The bottom-up approach is to think of networks as building up out of the combining of the simplest relations among three actors. A clique is a subset of actors in which every possible pair of actors is directly connected by a relation and the clique is not contained within any other clique. Cliques are used to identify how larger structures are compounded from smaller ones. Variations of cliques are *n*-cliques, *n*-clans, *k*-plexes, and *k*-cores, which have more relaxed definitions than cliques.

b) The top-down approach is to start from the whole network, instead of the dyads, and identify substructures as parts that are locally denser than the others as a whole. In a sense, this approach is looking for holes or vulnerabilities in the overall network. The top-down approach includes components, blocks and cutpoints, and Lambda sets. Components are subsets of a graph that are connected within, but disconnected between sub-graphs. Within a component, all actors are connected through paths, but no paths run to points outside the component. Isolates within graphs are also regarded as components. The pattern of components of a graph – their number and size – is taken as an indication of the opportunities and obstacles to communication or transfer of resources in the associated network .

C. SNA Tools

SNA will not be performed efficiently without the use of an appropriate tool. SNA tools provide functionalities for network data analysis, relieving the burden of intensive computation for the researchers. SNA tools also provide 2-D and 3-D visualization of the network data so that the researchers can have a better understanding of the network structure by viewing it. Currently, there are a number of SNA tools available in the market. Some of the common ones are NetMiner, UCINET, NetDraw, MAGE and Pajek [9, 10]. The website of the International Network for Social Network Analysis (INSNA) (http://www.sfu.ca/~insna/) is an excellent source for accessing these tools.

III. THE CASE STUDY

Total of 547 theses and dissertations classified into in the 10 ITS application fields gathered from 1991 to 2010 for the survey are analyzed to determine knowledge flow within 10 universities in Taiwan (shown in Table I).

TABLE I. THESES DISTRIBUTION OF 10 UNIVERSITIES IN TAIWAN

Cluster / University	ATMS	ATIS	APTS	ETC /EPS	CVOS	AVCSS	VIPS	EMS	Others	Sub-total
NCKU	48	10	7	3	5	15	5	6	0	99
NTU	32	7	15	10	6	10	4	4	0	88
TKU	29	17	15	4	5	5	3	6	0	84
NCTU-Taipei	15	12	4	5	17	2	7	8	1	71
NCTU-HsinChu	33	6	10	1	10	1	5	6	0	72
FCU	25	2	13	2	11	1	0	4	0	58
CHU	3	0	3	1	4	11	9	0	0	31
NCPU	4	0	8	3	4	0	1	0	0	20
NKFUST	0	0	0	1	11	0	0	2	1	15
NCYU	3	0	0	1	0	0	1	0	0	5
KNU	0	0	0	2	2	0	0	0	0	4
Total	192	54	75	33	75	45	35	36	2	547

The curve showing the trend of ITS theses and dissertations is depicted in Figure 2. It demonstrates that the amount of theses and dissertations increases rapidly since 2000 because of numerous ITS deployment projects sponsored by central and local governments in Taiwan.

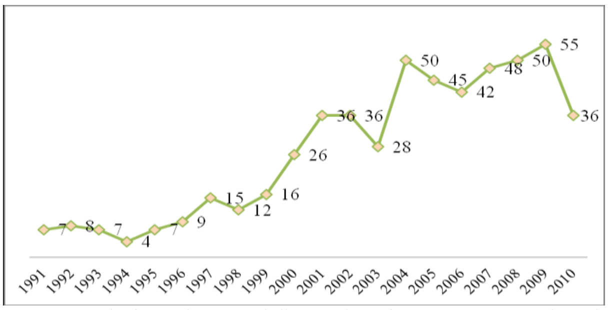

Figure 2. Trend of ITS theses and dissertations from 1991 to 2010 in Taiwan

The analysis considered two kinds of detailed knowledge flows: among the universities and among the individual professors. The SNA tool, UCINET, is used to assist in the analysis. Table II shows that National Taiwan University (NTU), National Chiao Tung University (NCTU)-Taipei, NCTU-Hsinchu, National Cheng Kung University (NCKU) and Tamkang University (TKU) are main contributors of ITS theses and dissertations for 9 application fields because of higher values of centrality of degree.

TABLE II. CENTRALITY OF ITS THESES AND DISSERTATIONS OF TAIWAN'S UNIVERSITIES

University	Centrality of Degree	Centrality of Betweenness
NTU	0.40	0.01
NCTU-Taipei	0.40	0.01
NCTU-HsinChu	0.40	0.01
NCKU	0.40	0.01
TKU	0.40	0.01
FCU	0.35	0.008
CHU	0.30	0.005

Figure 3 shows the social network graphs drawn by UCINET. The dicker lines represent bigger amount of theses and dissertations in different application fields. Many theses and dissertations concerning ATMS (Advanced Traffic Management Systems) are produced prominently by each university. However, each university has its own focus. For example, NCTU-Taipei is interested in researching CVO (Commercial Vehicle Operation)；NTU and NCKU pay more attention to AVCSS (Advanced Vehicle Control Safety Systems); NTU, NCTU-Hsinchu and TKU aim at emphasizing ATIS (Advanced Traveler Information Systems) and APTS (Advanced Public Transportation Systems); Feng Chia University (FCU) is absorbed in ATMS and Chunghwa University (CHU) stresses particularly on APTS and VIPS (Vulnerable Individual Protection System).

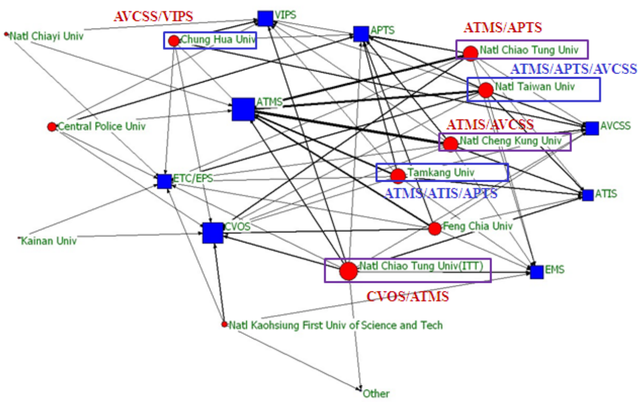

Figure 3. ITS Knowledge mapping of social network graphs of Taiwan's universities

All research topics corresponding to 9 application fields for each university are classified according to key words and context correlations. For each professor in 10 universities, his/her academic publications concerned research topics are also analyzed in the same way. Taking TKU as an example, table III shows the distribution of research topics in ITS application fields at Tamkang University. The high frequencies of traffic management and route guidance reveal TKU's specific research interests in ATMS and ATIS.

TABLE III. TKU'S DISTRIBUTION OF RESEARCH TOPICS

Application Field	Research Topic	Amount
ATMS	Traffic management	20
	Traffic signal control	2
	Automatic vehicle classification	3
	Automatic vehicle recognition	1
	Incident management	3
ATIS	Pre-trip information	2
	Route guidance	15
APTS	Dynamic bus information system	5
	Intelligent railway system	2
	Public transit operation	2
	Smart bus	2
	Bus signal priority	1
	Demand responsive bus	2
	High occupancy bus	1
ETC/EPS	Electronic toll collection	3
	IC smart card	1
CVO	Fleet management	3
	Taxipooling	2
AVCSS	Driving simulation	1
	Driver assistant system	3
	Advanced vehicle safety	1
VIPS	Travel assistance for elderly people	3
EMS	Emergency incident warning	1
	Emergency fleet management	1
	Rescue vehicles dispatching	2
	Traffic management after disaster damage	2

Table IV demonstrates centrality of degree and betweenness of each professor at TKU. The value of centrality of degree more than 0.20 represents that more than 6 theses and dissertations supervised by regarding professor cover at least 4 ITS application fields.

TABLE IV. CENTRALITY OF ITS THESES AND DISSERTATIONS OF TKU'S PROFESSORS

Name	Centrality of Degree	Centrality of Betweenness
CHANG, Sheng-Hsiung	0.20	0.007
CHEN, Dun-Ji	0.25	0.043
CHIU, Hsien-Ming	0.15	0.011
FAN, Chun-Hai	0.15	0.015
HSU, Chao-Che	0.05	0.000
HU, Shou-Ren	0.10	0.002
LIU, Shih-Sien	0.20	0.013
LUO, Shiaw-Shyan	0.25	0.029
SHIH, Fon-Yu	0.05	0.000
TAO, Chi-Chung	0.20	0.016
TONG, Chee-Chung	0.20	0.012

Considering the dick lines connecting different ITS application fields, it is found that Tong devoted himself to researching ATIS, especially in route guidance and driver information assistance. Although Luo, Fan and Liu are interested in ATMS, their specialized subjects still differ each other. Liu puts effort on dynamic traffic assignment and travel time estimation; Luo's studies cover traffic control, smart bus and emergency management; while Fan focuses on automatic vehicle classification and image processing. Chang pays more attention to APTS, especially the subjects of demand responsive transportation and dynamic bus information system. Chen's interests include ITS business models and industrial development; Tao's publications cover ATMS, ATIS, APTS and ETC, particularly in the field of intelligent railway systems.

Figure 4 shows TKU's social network graphs based on individual professors.

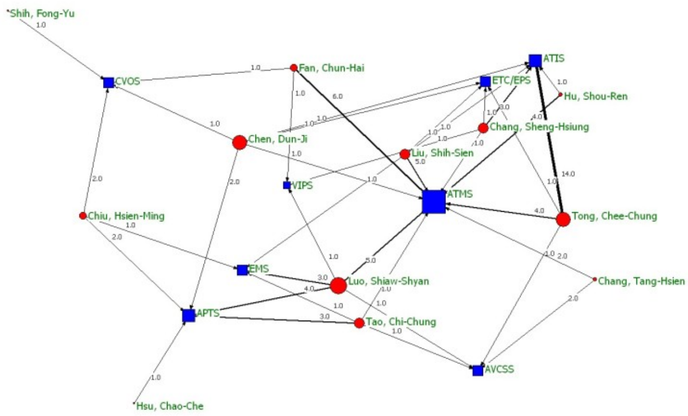

Figure 4. ITS Knowledge mapping of social network graphs of TKU's professors

In summary, Taiwan's main contributors of ITS theses and dissertations are 6 universities including NTU, NCTU, NCKU, TKU, FCU and CHU with the help of knowledge mapping by SNA. The most popular application field of ITS theses and dissertations is ATMS. From knowledge maps of each university, detailed research topics can be classified such as incident management, traffic signal control and automatic vehicle detection for ATMS. And specialized interests of each professor can also be identified by using SNA.

This case study attempts to illustrate the application of the SNA tool to knowledge mapping. Through the use of visualization and analysis reports of the SNA tool, potential issues not easily discovered by other means are identified. It is important to note, however, that further investigation into these issues should be performed before any changes to the network are done in order to confirm whether those are real issues. For instance, it is indicated in the analysis that the knowledge of the 'Information Management System (IMS)' application field is not clustered notably. A further analysis might find out that all subject matter knowledge of IMS can be documented as explicit knowledge for reference from other clusters and that the knowledge mapping will be more effective.

There are a couple of limitations to this case study:

1) Almost all theses and dissertations are collected from national digital library, however, they are searched and classified by the manual process because of lacking automatic Chinese extraction tools. Errors or Misunderstanding might occur while screening ITS theses and dissertations by key words and context correlations.

2) Not only transportation science, but also other disciplines such as information and communication sciences should be included to provide an overall ITS knowledge map . Due to limited manpower and budgets the ITS knowledge mapping in this study is principally valid for road transportation experts.

3) Visualizing knowledge domains is a very promising field of research and its commercial potential is great. Knowledge-poor approaches refer to those essentially relying on statistical and probabilistically methods to solve problems in text summarization, whereas knowledge-rich approaches make use of external sourcesof rules and heuristics in the summarization process [11]. Author co-citation analysis (ACA) provides a capable basis for modeling and visualizing knowledge structures reflected in scientific literatures [12]. The particular intellectual bonds between scientists based on ACA are needed to be explored in the future.

IV. Conclusions

In this case study, the state of the art of knowledge mapping of ITS theses and dissertations in Taiwan is highlighted by using SNA. Six universities including NTU, NCTU, NCKU, TKU, FCU and CHU are main contributors to ITS theses and dissertations in Taiwan. Their knowledge mappings in form of social network graphs drawn by UCINET are also provided to emphasize the key research topics of ITS application fields and its profound connections among universities, especially among individual professors. Some of the most influential research topics and professors have been discussed. Limitations of this case study are also presented for the follow-up research.

In conclusion, ITS knowledge mapping is a challenging but ultimately rewarding route to capture the essence of a scientific paradigm. Potential ITS applications will be valuable to multiple domains. These are tough challenges, but the potential benefits are tremendous and profound.

V. References

[1] National Digital Library of Theses and Dissertations in Taiwan, http://ndltd.ncl.edu.tw, 2012.

[2] M. Dodge and R. Kitchin, Atlas of Cyberspace, Addison-Wesley,2001.

[3] J.D. Novak, D.B. Gowin, Learning How to Learn, Cambridge Univ. Press, New York, 1984.

[4] M. Dodge, NewsMaps: topographic mapping of information, Available at: http://mappa.mundi.net/maps/maps_015/ 2000.

[5] J. Scott, Social Network Analysis: A Handbook, Sage Publications Inc, Newbury Park, CA, 1991.

[6] R. Hanneman, Introduction to Social Network Methods, retrieved from http://www.faculty.ucr.edu/~hanneman/, 2001.

[7] S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications, Cambridge University Press, New York, NY., 1997.

[8] K. Chan and J. Liebowitz, "The synergy of social network analysis and knowledge mapping: a case study " in Int. J. Management and Decision Making, Vol. 7, No. 1, 2006, pp. 19–35.

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

[10] J. Scott, Social Network Analysis, Sage Publications, 2007.

[11] U. Habn and I. Mani, "The Challenges of Automatic Summarization", in IEEE Computer, 33 (11), 2010, pp. 29-36.

[12] C. M. Chen, "Visualising Semantic Spaces and Author Co-citation Networks in Digital Libraries", in Information Processing and Management, 35 (2), 2010, pp. 401-420.