



HAL
open science

From Chinese Philosophy to Knowledge Discovery in Databases: A Case Study: Scientometric Analysis

Eric Boutin, Pei Liu

► **To cite this version:**

Eric Boutin, Pei Liu. From Chinese Philosophy to Knowledge Discovery in Databases: A Case Study: Scientometric Analysis. Encyclopedia of Database Technologies and Applications, Idea Group Reference, USA, pp.1-11, 2008. sic_00826989

HAL Id: sic_00826989

https://archivesic.ccsd.cnrs.fr/sic_00826989

Submitted on 28 May 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Chapter ??

From Chinese Philosophy to Knowledge Discovery in Databases: A Case Study: Scientometric Analysis

Pei Liu

Université du Sud Toulon Var, France

Eric Boutin

Université du Sud Toulon Var, France

INTRODUCTION

The field of scientometrics has been looking at the identification of co-authorship through network mapping. Research on this topic focuses on the cooperation of two authors who have published papers together. However, this paper is exploring the latent association of two authors. By ‘latent association’, we mean that the collaboration between two researchers has not yet occurred but might very likely take place in the future. In this paper, we will aim to find out a couple of authors who have never published together and who bear similar academic interests or study similar subjects. We will also show how the concepts of Yuan (Interdependent arising), Kong (Emptiness), Shi (Energy) and Guanxi (Relationship) in Chinese philosophy contribute to understand ‘latent associations’. These four Chinese concepts are

the theoretical basis of this paper. By explaining one by one what each concept is about we hope to tackle the two following questions: What do those four concepts exactly tell us? And how are they linked together? Finally, we will look at the empirical case study in scientometrics. We hope to show that this application of Chinese concepts can unravel latent associations between researchers in Database.

CONCEPTUAL MODEL

The Interdependent Arising (“Yuan” in Chinese 缘)

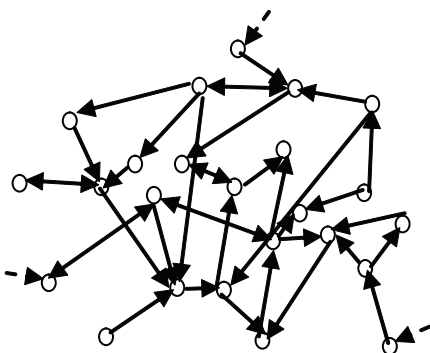
The Chinese believe that every phenomenon arise within the context of a mutually interdependent web of cause and effect as much in time as in

space. This concept is also the basis of Buddhism thought, as it is encompassed in the following classical formulation:

*“When this is, that is.
From the arising of this comes the arising of that.
When this isn’t, that isn’t.
From the cessation of this comes the cessation of that.”* (Samyutta Nikaya 雜阿含經)

This formulation illustrates the Asian ontology which is that everything appears only because of the arrival of its “arising dependent” (Yuan) and it exists because its “arising dependent” exists. They believe that everything is conditioned and reciprocally influenced (Lai 2003, Cai 1990, Wei 1982, Fo 1992, Duo 1996, 2006). There is nothing in the universe that can survive without its interconnection with other things. Everything depends on everything else. This interdependence forms an incredibly complex web of causation and result encompassing the three universes of foretime, present and future (Figure 1). Whatever we are in whatever form, whenever and wherever we are situated in the world, we are the elements of a connected web within a multidimensional causal nexus (of time and space). In other words, we are just the interconnected nodes of an intricate network receiving the arrows and then sending them to others.

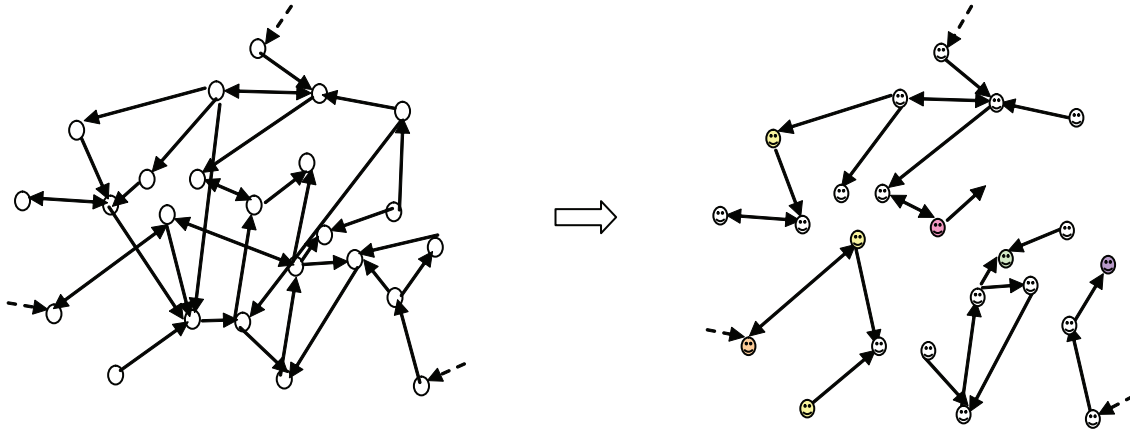
Figure 1. Complex web of causation and result of all things



Yuan (缘) → Guanxi (in Chinese 关系)

The belief of this arising dependent (缘) has deeply influenced the Chinese people. They often conclude that something is inexplicable as a result of the arising dependent, known in Chinese as “Yuan”. Two persons can know each other as a result of the Yuan. Someone who knows many people will be complimented as “you ren Yuan”, means “have persons’ Yuan”. Owing to the Yuan’s power on which the existence of everything in the world depends, the person involved in many relationships with others can become stronger. If the nodes of the Figure 1 are replaced by people, the web of causation and result of all things is translated into the web of a person’s relationships. On the other hand, as the “Yuan” produces outcomes, the relationship between persons is transformed from cause->effect into relationship->outcome. Consequently a person who knows a lot of people has lots of “ren yuan”. What is more, this person is more powerful when connected to other persons. He can produce many outcomes and advantages. These outcomes and advantages can also be exchanged for other favours in order to achieve certain purposes, like for example, business activities within a network of informal and interpersonal relationships (Lovette et al., 1999) called “Guanxi”. Although, it is not the point to discuss in this paper the theory of Guanxi, it is not difficult to understand why Chinese people and even other East Asian people unconsciously share this same Guanxi philosophy. Guanxi is an intricate and pervasive relational network consisting of mutual obligations, assurances and understandings (Park and Luo, 2001). The source of Guanxi philosophy has its roots in the web of all arising dependent. In that respect, Guanxi can be considered as a subset of the web of Yuan (Figure 2). The whole web and its subset (Guanxi) are also mutually interdependent.

Figure 2. Guanxi network



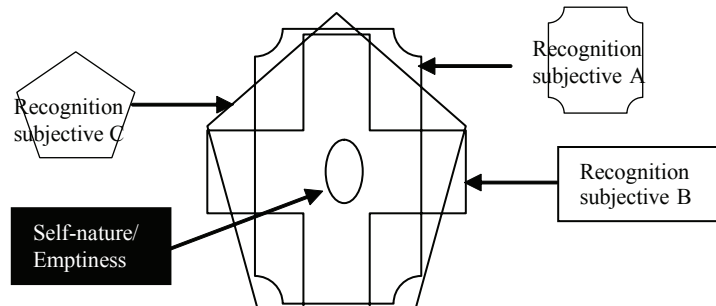
Yuan → Kong (Emptiness, in Chinese空)

All things are inter-conditioned and exist in the universe in a way of impermanence (in Chinese called Wuchang, 无常). As a result, the appearance of phenomena stems from an interdependent effect of the main cause and of countless intrinsic and extrinsic factors. If this arising interdependent stops existing, the phenomena disappear. So there are no permanent phenomena. None truly exists. Everything is therefore asomatous and empty (Nagarjuna, Mūlamadhamaka kârikas). This concept of the ‘emptiness’ means that things do not have a natural form like what we can see (Wu 1996). An object or a person we see is only an identity shaped by our imagination and recognition (Wu 1999).

This absolute subjective identification is based on a kind of incorporeal inherent existence. Just like the figure 3, different people see diverse things from the same object, and the object himself does not have an exact shape either. It is a composition of natural factors, like the water and the earth in an inter-conditioned background.

These natural factors which are united by Yuan build up the Self of this object. They become invisible when the Yuan flies away. So the Kong implicates an identification of the self-nature of things. It offers us the opportunity to apprehend and predict the evolution of things. If we keep a vision of the “emptiness”, we can find out the self-nature of things and the change it undergoes. The development and the movement of things are in our control. While cultivating this vision of

Figure 3. Different recognition based on the same ‘self-nature (emptiness)’



emptiness, we can release our thinking which is often limited and pre-defined by the appearance of phenomena to search in the background the cause of things. So it is important to keep at all time a view of emptiness for every phenomenon.

Kong → Shi (Force and Energy, in Chinese 勢)

A metaphor of water is often used by Chinese people. Although small brooks are weak and easy to divert it is strongly believed that they should not be ignored. A powerful large river always starts from hundreds of weak brooks which later become bigger and bigger. “The rush of a torrent which will even roll stones along in its course (Art of the war by Suntsu)”. It’s the result of the ‘Shi’ (勢) which in Chinese means the energy and force. This implies that we should use the momentum of the water in movement which possesses ‘Shi’. The vigour of the torrent will make us more powerful to achieve our final goal. This approach has twofold. One is to study and analyse clearly the environment in order to position ourselves in the right place as early as possible and the other is to be pushed by the force of Shi in the right direction. In other words, a vision of emptiness (Kong) needs to be inside everything in order to control every opportunity of evolution.

Only if we can clearly analyze the entire situation by using the vision of emptiness (Kong) in the context of a cause-effect web of Yuan, we will be able to situate ourselves in a small streamlake. This streamlake has the potential to become a big river in the future when we need it. In contrast if we situate ourselves but not in a river, which for now seems to possess a good Shi, this river might however become droughty tomorrow. That is why it is always worthwhile to choose carefully since the emptiness is a latent force for great changes, such as the Western attitude to the study of weak signals.

Shi and Guanxi (关系)

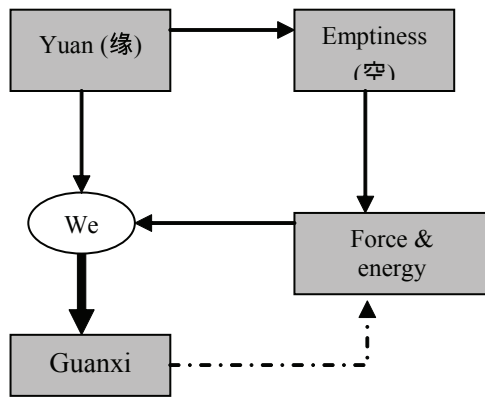
Guanxi (Part 2 Yuan->Guanxi) signifies in China the value of a social capital accumulated by a person. (Warren et al., 2004). Somebody can be hired just because he has a good Guanxi. (Aufrecht and Bun, 1995). A good Guanxi in China equals a good social situation. Guanxi is absolutely necessary to get through social life in China.

Chinese people like to be in “Guanxi Wang”, which consists of cultivating carefully a web of relationships and to only use these Guanxi when really needed. This Guanxi philosophy is herewith a good way to balance the power of the Shi. This is because if we consider each tie of the Guanxi network as a small brook, the day when someone has many ties in his web, he will become as powerful as the great river. He will as such be able to get rid of any problem easily. As a consequence, it is viewed logical and efficient for a Chinese person to make use of relationships as a resource and to mobilise one’s Guanxi in order to create a favourable Shi and to realize one’s own desire.

Brief Summary

The relationship of Yuan, Kong, Shi and Guanxi are illustrated in the Figure 4. Everything in the world comes with the arrival of Yuan and disappears when Yuan vanish. This makes us conclude that everything is empty and temporal. The Kong (emptiness) implies that everything in the world does not truly exist. Therefore the importance of a vision free from what we see through our eyes to find out about the self-nature is necessary. It is useful to cultivate this valuable vision to determine and create the Shi because the Shi is powerful. We can also achieve the target more easily if we control it. The world we live in can be seen as an immense causal web. A web which ties people’s relationship is called “Guanxi Wang” in Chinese. It’s a type of Chinese working platform: people first react by mobilising their

Figure 4. Causal relations of four concepts



web to position themselves in a good Shi and to pursue their goals.

CREATION OF SHI FOR SCIENTIFIC RESEARCH

The development of our conceptual model brings up the following question “Can we stimulate the development of scientific research?” The affirmative response to this question has led us to another question “Can we re-establish intelligently our ‘Guanxi Wang’ in order to scientifically enrich ourselves and to ameliorate the production of publications?” According to the perspective of the mentioned above Chinese philosophy, the idea of a Guanxi and Shi arises spontaneously. In the second part of the paper, I would like to describe the identification of Latent Association using the scientometrics method.

In the field of scientometrics, many publications have focused on analyzing the output of scientific research (Aksnes & Taxt, 2004; Weingart, 2005). In particularly in bibliometrics, most scientists studied the exploration of quotation as well as co-publication. They consider these elements as impact factors of scientific evaluation (Moed, 2005; Yi H. & Ao X. et al. 2008; Ulrich S. & Torben S. 2008). But the research which pays

attention to the identification of co-authorship through network mapping is normally focusing on the cooperation of two authors who have published papers together. This co-publication relationship represents the actual Guanxi among researchers and shape Shi’s existence. In line with our above conclusion, the creation of the new Guanxi and powerful latent Shi in scientific research could facilitate the development of research at a high speed thanks to the force of Shi. So we endeavour to identify the potential of Guanxi between scholars and at the same time to create a new Shi for the scientific research through scientometrics analysis based on the points just raised.

We first define the latent Guanxi of two authors, which refers to the collaboration of researchers that has not yet occurred but might very likely take place in the future.

The concepts of Yuan and Kong teach us a constructive view of the cognitive process that is useful to understand the Yuan (the origin of things) and to find out about the self-nature of things. Co-authorship comes from the fact that the two authors have written an article together. In such association, it must be noted that they also share the same title, keywords and bibliographies. In other words, because scientists in the same field of research are interested in similar subjects, there is a possibility of co-publication. So the identification criteria for latent co-authorship might be, for example:

- Two researchers publish articles using the same keywords
- Two researchers publish articles which are associating the same people
- Two researchers publish articles which have similar references in bibliography.

Methodology

The methodology of this empirical study is technically based on data mining and knowledge discovery. Knowledge Discovery is normally used in

bibliographical databases. The study will employ this method to carry out a comprehensive analysis of network research through studying the latent Guanxi by looking at similar keywords. Hence we will build a network of latent keywords by relating two authors which have not published together, but who have at least published separately one article which is at least described by one identical keyword.

Suppose two articles for which we have the information of authors and keywords. The authors and keywords are expressed as letters and numbers (Table 1).

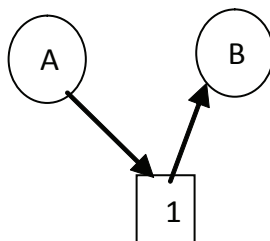
The keyword “1” has been used by author A and B, so if we follow the transitive reasoning we get the Figure 5 because A is described by keyword 1 and the keyword is found in the article of B (Figure 5).

From the example, there are present in six associations: AB, AC, BC, CD, BD and AD. Among these identified associations, we can from now on filter to retain the latent associations. AC, AD, DC, BD are not the latent associations because

Table 1. Demo corpus

Article α : Authors: A, C, D Keywords: 1, 2, 3 Article β : Authors: B, D Keywords: 1, 5, 7
--

Figure 5. Transitive logic



they have published the same paper. The list of the latent associations is therefore: AB and BC.

Experimental Validation

From the data base of archiveSIC¹, we construct a corpus of 30 articles of bibliometrics and scientometrics. The Table 2 shows the first 5 articles of the corpus.

From these data, we create a table of 49 real co-authorship associations by respecting the criterion of two authors who have at least published one paper together. We show a part of associations in Table 3.

Since the original reference doesn't include the keywords of the articles so we enter into each article to find the keywords. Then according to the methodology mentioned above, we can obtain any possible association which link authors who share at least one common keyword. As these associations also compose the existing Guanxi of the co-publication due to the fact that a same article has exactly the same keywords, we removed real associations and listed in Table 4 21 latent Guanxi. The column of frequency is the synthetical value of common keywords and the frequency of each common keyword. It implies the possibility to transpose latent-reality association. Thus the Shi contains the energy of the torrent that can even roll the stones.

The network analysis (Wasserman & Faust 1994) provides a methodology of representation and characterization of the interactions between objects. “*Networks are the language of our times.*” (McCarthy et al. 2004) We therefore represent the interactions of real Guanxi in Figure 6. In Figure 7, we represent the interactions of all relationships including the latent association obtained by network analysis. Each node symbolizes an author, and the size represents the number of articles for each author. A bigger size denotes that more papers have been published in the Bibliometrics field of ArchiveSIC database. For the reason of publication, the different relationships are dis-

Table 2. The first 5 articles of archiveSIC in bibliometrics

 <p><u><i>Les nouvelles formes d'évaluation scientifique : quelles évolutions en sciences, technique et médecine ?</i></u> Durand-Barthez, M. [sic_00260459 – version 1]</p>
 <p><u><i>Citations des revues de paléontologie : bilan 2006 et 2007</i></u> Néraudeau, D., Dubigeon, I., Le Gall, A.-H. [sic_00254730 – version 1]</p>
 <p><u><i>Citations des revues de paléontologie : bilan 2005</i></u> Néraudeau, D., Le Gall, A.-H. [sic_00166236 – version 1]</p>
 <p><u><i>Citations des revues de paléontologie : bilan 2004</i></u> Néraudeau, D., Le Gall, A.-H. [sic_00166233 – version 1]</p>
 <p><u><i>Citations des revues de paléontologie : bilan 2003</i></u> Néraudeau, D., Le Gall, A.-H. [sic_00166231 – version 1]</p>

Table 3. Several co-authorship associations

Author	Author
Alain-Hervé_Le_Gall	Didier_Néraudeau
Bador_P.	Boukacem_C.
Bador_P.	Lafouge_T.
Bador_P.	Prost_H.
Bador_P.	SchÖpfel_J.
Beigbeder_M.	Lafouge_T.

tinguished by thinner and thicker ties. Thinner ties mean real associations and the thicker ones signify latent ones. The thickest ties represent a higher frequency of latent associations.

Table 4. Latent association

Latent association		
auteur1	auteur2	Frequency
Alain_Hervé_Le_Gall	Boukacem_C.	3
Alain_Hervé_Le_Gall	Lafouge_T.	3
Alain_Hervé_Le_Gall	LE_COADIC_Y_F.	3
Alain_Hervé_Le_Gall	Salaün_J._M.	3
Boukacem_C.	Didier_Néraudeau	3
Didier_Néraudeau	Lafouge_T.	3
Didier_Néraudeau	LE_COADIC_Y_F.	3
Didier_Néraudeau	Salaün_J._M.	3
Prime_C.	Smolczewska_A.	3
Lafouge_T.	LE_COADIC_Y_F.	2
Michel_C.	Prime_C.	2
Beigbeder_M.	Michel_C.	1
Beigbeder_M.	Smolczewska_A.	1
Boukacem_C.	Epron_B.	1
Boukacem_C.	LE_COADIC_Y_F.	1
Boukacem_C.	Pouchot_S.	1
Boukacem_C.	Prime_C.	1
Ibekwe_SanJuan_F.	Michel_C.	1
LE_COADIC_Y_F.	Salaün_J._M.	1
Michel_C.	SanJuan_E.	1
Michel_C.	Smolczewska_A.	1

CONCLUSION: INTERPRETATION OF RESULTS AND FUTURE TRENDS

The real research of Guanxi is dispersed in the graphic. Latent ties connect the largest research group with several other researches. Communications and Guanxi could be stabilised with these ties. The core axe of research is more balanceable than horizontal cooperation, if it creates some cooperating opportunities to build new Guanxi

Figure 6. Real association

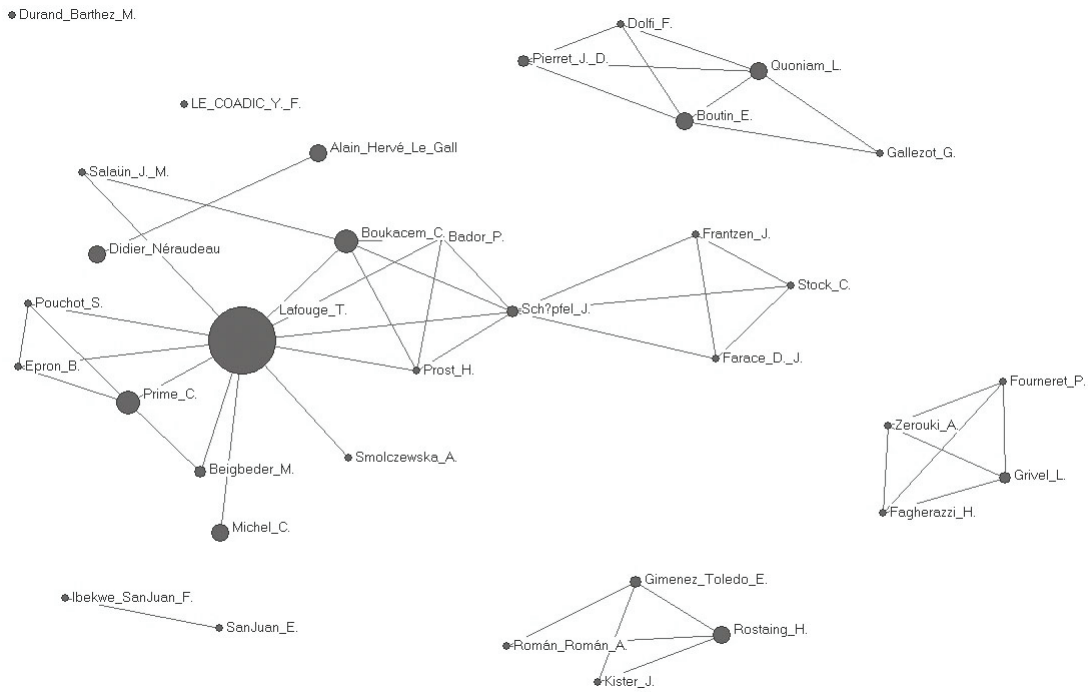
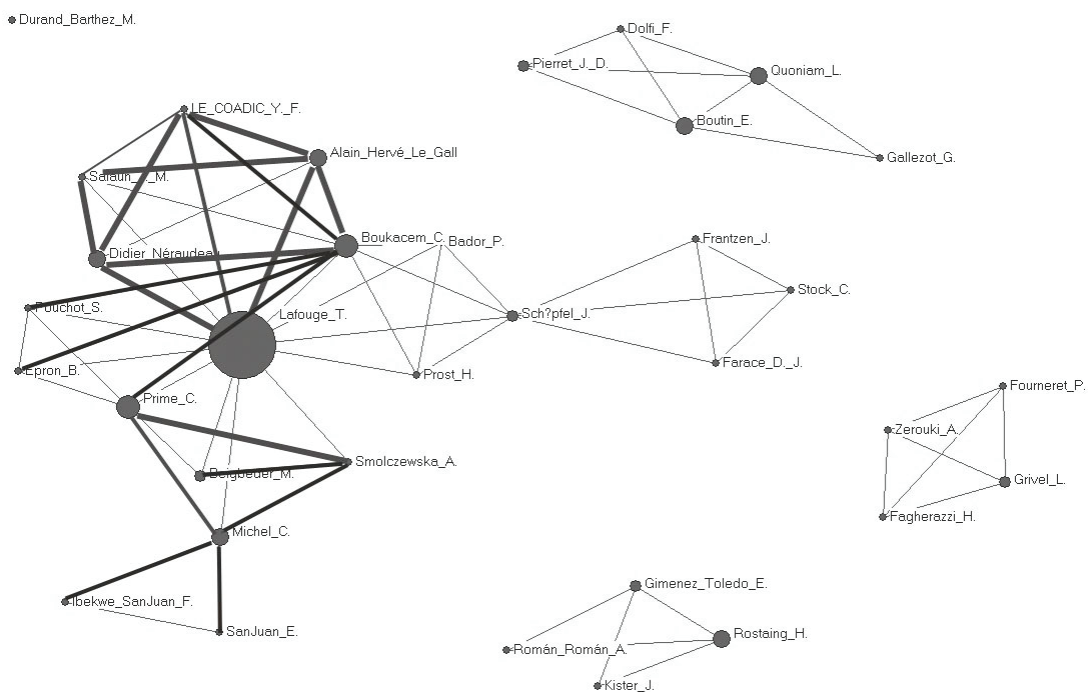


Figure 7. Real network (thinner tie) and latent network (thicker tie)



with the upper author. We clearly see from the figure that there is a potential to create new energy and that the research process could advance more smoothly by enlarging the Shi of the research group's origin.

The approach helps to identify the potential relationships within researchers of the same field who haven't yet published papers together. It contributes not only to the identification of latent research associations, but also it inspires researchers to network for potential cooperation across institutions, countries and field of studies. The potential for relationship brings the Shi to the field of scientific research. Shi as it is developed in latent association can become "the momentum of a round stone rolling down a mountain thousands of feet in height. There is so much on the subject of Shi." (Art of the war by Suntsu)

In our research, the technical method is based on the author and the keywords' information. It could also be studied by basing it on the author and a series of quotations or even on a combination of multi-information including author, keywords, year of publication and citations. Because the year of publication announces the age and the historical context of papers, and citations reflect the way of thinking and the appreciation of relative achievements of authors, since it is considered like a prize awarded for past research (Merton, 1979) and it could be used as a tool to convince others (Latour & Wolgar, 1979). More refined criteria also allow to reinforce the pertinence of identified latent association as well as a direct access to a mighty Shi.

EPILOGUE

We have taken a somewhat unconventional approach to conclude this article. In the first part of the paper, we have adopted a theoretical approach to try to present a new concept inspired by Chinese philosophy. And in the second part,

we came back to the fields of scientometrics to analyse the network and verify our conceptual model. Specialised in scientometric study and influenced deeply by Chinese philosophy, we recognize the intellectual connection between these two completely separate concepts. We also consider necessary to share this original thinking with other researchers by presenting this theoretical and empirical research. Since Chinese philosophy is usually talked about by sociologists and is most unlikely acquainted to the researchers in the field of scientometrics, we had to write many texts to describe this conceptual model. The experimentation is based on the title, author and keyword information. The method could also use information such as references, years of publications and other descriptive of the authors, like university or country as this would help to find out and create Shi across institutions and countries.

REFERENCES

- Aksnes, D., & Taxt, R. (2004). Peer review and bibliometrics indicators: A comparative study at a Norwegian university. *Research Evaluation*, 13(1), 33-41.
- Aufrecht, S. E., & Bun, L. S. (1995). Reform with Chinese Characteristics: The Context of Chinese Civil Service Reform. *Public Administration Review*, 55, 175-183.
- Cai, H. M. (1990). The standpoint of the primitive buddhism (原始佛教的缘起观). *Dharmaghosa (The Voice of Dharma)*, 1990(05), 25-27.
- Duo, S. (1996). Some characters of Buddhism different from the other religion. *Tibetan Folklore*, 1996(03), 52-64.
- Duo, S. (2006). Several Key Features of the Difference between Buddhism and Other Religions. *N.W. Ethno-National Studies*, 2006(02).

- Fo, R. (1992). Dependent origination and nature origination. *Dharmaghosa (The Voice of Dharma)*, 1992(04), 7-15.
- Lai, Y. (2003). The interdependent arising is the fundamental theory of Buddhism. *Social Science Front*, 2003(5), 50-52.
- Latour, B., & Wolgar, S. (1979). *Laboratory Life: The Social Construction of Scientific Facts*. Sage Publishing, Los Angeles.
- Lovett, S., & Simmons, L. C., & Kali, R. (1999). Guanxi Versus the Market: Ethics and Efficiency. *Journal of International Business Studies*, 30(2), 231-248.
- McCarthy, H., & Miller, P., & Skidmore, P. (Eds.) (2004). *Network logic: Who governs in an interconnected world?* London: Demos. Individual essays available at <http://www.demos.co.uk/catalogue/networks/>.
- Merton, R. (1979). Foreword. In E. Garfield (Ed.), *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*, vii-xi. John Wiley and Sons, New York.
- Moed, H. F. (2005). Citation Analysis in Research Evaluation. *Information Science and Knowledge Management*. Springer, Dordrecht.
- Park, S. H., & Luo, Y. (2001). Guanxi and Organizational Dynamics: Organizational Networking in China Firms. *Strategic Management Journal*, 22(5), 455-477.
- Ulrich, S., & Torben, S. (2008). Are international co-publications an indicator for quality of scientific research? *Scientometrics*, 74(3), 361-377.
- Warren, D. E., Dunfee, T. W., & Li, N. (2004). Social Exchange in China: The Double-Edged Sword of Guanxi. *Journal of Business Ethics*, 55, 355-372.
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge, England, and New York: Cambridge University Press.
- Wei, A. (1982). The dependent origination-the core of Buddhism. *Dharmaghosa (The Voice of Dharma)*, 1982(01), 31-33.
- Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics*, 62, 117-131.
- Wu, G. (1996). Commentary to the dependent arising and emptiness. *Wuhan University Journal (Philosophy & Social Science Edition)*, 1996(02), 35-38.
- Wu, X. (1999). The reconstruction of dependent arising and emptiness contributed by the mind-only theory. *Fudan Journal (Social Sciences)*, 1999(03), 67-88.
- Yi, H., Ao, X., & Ho, Y. (2008). Use of citation per publication as an indicator to evaluate pentachlorophenol research. *Scientometrics*, 75, 67-80.

KEY TERMS

Guanxi: Chinese concept, a kind of interpersonal relationship, from a type point of view, Guanxi is considered as a standard of personal social evaluation.

Kong (Emptiness): Buddhism philosophy. It means everything in the world is empty and it implies many possibilities. By finding out the self-nature of things, the Kong vision enhance a clearly observation and a profound thinking that are often blinded by the appearance.

Knowledge Discovery: A kind of nontrivial study on extracting implicit and potential, useful information and then transform them accurately into intelligent information.

Latent Association: From a type point of view, latent associations refers to the co-work of researchers that has not occurred until present but will be very possible to take place in the future.

Scientometrics: It means a science of all applications of mathematical and statistical methods which are often developed to measure and evaluate the scientific publications.

Shi (Energy): A kind of momentum contained in the surrounding environment that could be borrowed, transferred or even created to help people achieve his target.

Yuan: It means all phenomena arise in the context of a mutually independent web of cause and effect as much in time as in space. Yuan could be considered as the reason of everything.

ENDNOTE

- ¹ archiveSIC: an open archive of Science of Information and Communication in France
<http://archivesic.ccsd.cnrs.fr/>