

Analyzing Scientific Structure of Digital Humanity

Hsin-Ning Su

Graduate Institute of Technology Management, National Chung Hsing University, Taichung, Taiwan

Abstract--The integration of humanity and digital technology has been becoming a new approach for social science research or even gradually forming a new discipline in academy. To understand currently developed knowledge structure of "Digital Humanity", this study utilized keyword analysis and social network analysis on investigating scientific papers published in the field of "Digital Humanity". The methodology used in this study is capable of creating three-dimensional "Research focus parallelship network" and "Keyword Co-occurrence Network", together with two-dimensional knowledge map. The networks and knowledge maps can be depicted differently by choosing different information as network actor, i.e. country, institute, paper and keyword, to reflect knowledge structures from macro, meso, to micro-levels. The quantitative exploration provides a way to visualize and assess the development of "Digital Humanity" to encourage the formation of this emerging research field.

I. INTRODUCTION

In this modern society, information communication technology contributes significantly to the development of knowledge and accelerate knowledge creation as well as its paradigm shift. The integration of digital technology and humanities study is advanced by the rapid-pace of ICT development. Humanities documents are digitized and analyzed to uncover knowledge that is hardly obtained by conventional methodologies. Therefore, the newly developed research fields named "Digital Humanities" is not only for creation of knowledge in a new field but also the creation of new method for knowledge development.

In order to catch up with the trend of the development of Digital Humanities research, Taiwan government has implemented national level of research projects in relation to Digital Humanities. For example, Digital Museum Project (1998~2002), National Digital Collection Project (2001), National Digital Archives Program (2002~2006), 2nd National Digital Archives Program (2007), and Taiwan e-Learning and Digital Archives Program (2008~2012). It is obvious that the focus before 2012 was to digitize historical documents, so digital-archive related programs were encouraged and sponsored by Taiwan government. However, the attention has been shifted to the use of digital methodologies to investigate digitized documents particularly after 2013. The use of digital tools on traditional social science may provide very significant contribution to knowledge development in human science and thus impact the concept of exploring humanities. Therefore, the Digital Humanity Projects was initialized in 2013 to analyze digital document to help humanities study.

Digital Humanities researches were promoted by many countries For example, Taiwan's Ministry of Science and

Technology initialized a Digital Humanities Program in 2013 to explore Digital Humanities researches in a comprehensive way. The program not only promote the development of Digital Humanities research community but also induce a question to be answered- What are the output of Digital Humanities researches worldwide. The performance evaluation of Digital Humanities research is a must when encouraging the development of this field. The output of Digital Humanities researches should be analyzed and the its developed knowledge structure should be understood before continuous resource allocation to the field of Digital Humanities.

A. Identification of Knowledge Structure

It can be observed in literature that knowledge structure has been investigated by bibliometric method based on keyword analysis. For instance, Ding et al. [1] investigated information retrieval research by the use of co-word analysis on papers from Science Citation Index and Social Science Citation Index from 1987 to 1997. ethics and dementia research is mapped by Baldwin et al. [2] also by keywords. Tian et al. [3] used SCI/SSCI database to measure scientific output of Geographic Information System. Different approaches are used to understand the structure of knowledge in a selected field [4][5][6][7].

A "structure" is a collection of inter-related components or services based on which knowledge can be developed. A well-defined structure can be depicted by a network which consists of (1) network actors and (2) network ties, which are corresponding to the "components" and "inter-relationship among components" in a well-defined structure. The application of network theory for understanding knowledge structure is therefore possible if both network actors and network ties can be well-defined. Our previous studies have conducted to understand knowledge structure of patented technology by combining network theory and patent citation analysis [8], and Research Focus Parallelship networks was proposed for understanding the structure of regional innovation system [7], electrical conducting polymer nanocomposite [9] as well as technology foresight [4].

B. Understanding the structure of Digital Humanity research

This paper aims to obtain structural analysis on Digital Humanity research[10][11]. The structure of Digital Humanities research is created and presented as networks consisting of network actors and network ties. Network actors which are depicted by Country, research institute, author and keyword are positioned in a relative manner to represent their roles in Digital Humanities researches

The created network structure can be used as an objective

evidence to allow decision makers to evaluate existing contest of global Digital Humanities and then decide what and how different research can be contributed to global Digital Humanities knowledge. The objective of this paper is to explore the knowledge structure of global Digital Humanities research to allows Digital Humanities researchers to understand how to position their research in global knowledge context.

II. DATA AND METHOD

A. Data

SSCI database is used to retrieve Digital Humanities related papers. Due to the interdisciplinary nature of Digital Humanities, it is very difficult to select a proper set of keywords to retrieve paper. This research used “Digital Humanities” as topic to retrieve 222 papers belong to this field on December 9, 2015. The keywords of the 222 paper are analyzed and found that “Digital Humanities” and “Digital History” are the two most general keywords that significant cover the field of Digital Humanities. The tow keywords, “Digital Humanities” and “Digital History” were therefore used a query strategy to obtained paper with title consisting any of of the two keywords. A total of 283 papers were obtained. However, some papers without author keywords or are not written in English are precluded. The rest papers are manually scanned to removed those which are not relevant to the field of Digital Humanities. The final corpus to be analyzed contains a total of 77 paper.

B. Knowledge mapping by network theory

1) The Research Focus Parallelship (RFP) network is

created for understanding how journal papers overlap each other through authors’ selection of keywords. Network actors with a minimum one same keyword are linked together Since journal papers can be aggregated into different levels of analyses, i.e. country, institute and author. The RFP network can be created as RFP-country network, RFP-institute network and RFP-author network, respectively [9][4].

2) Keyword Co-Occurrence (KCO) network is created for understanding how keywords are specified in each paper. Each set of author keywords in one paper are linked together because they are used to represent the research concept of one paper and the relation among these keywords should be emphasized [7][9][4].

C. Network centrality calculation

Network properties are calculated to understand how central each network is in a network. Brass and Burkhardt argued that a person with higher centrality is always the one with higher influence in an organization [12]. This study adopted three type of network centrality measurements proposed by Freeman: 1) Degree Centrality, 2) Between Centrality, and 3) Closeness Centrality [13].

III. RESULTS AND DISCUSSION

A. Descriptive analysis

No. of research papers along time horizon is provided in Fig. 1, it can be observed that Digital Humanities related papers was started in 1993 and reach its maximum in the year of 2012. A rapid increase after 2008 can be observed.

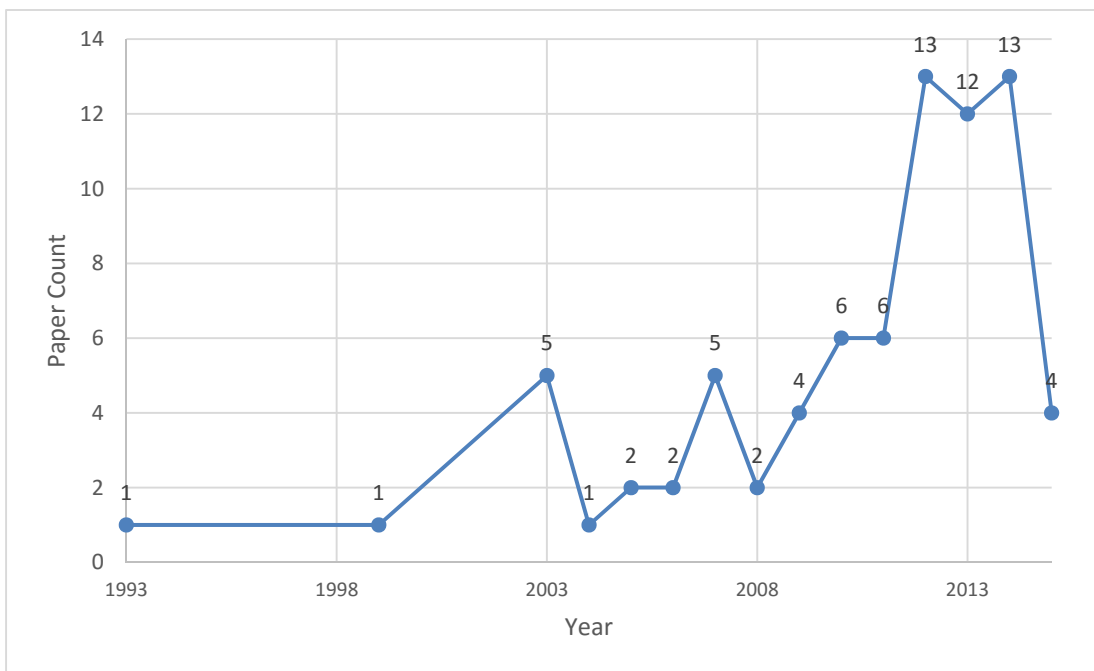


Fig. 1. Number of Digital Humanities Papers from 1993 to 1995.

2016 Proceedings of PICMET '16: Technology Management for Social Innovation

In TABLE 1, Journals containing these Digital Humanities paper are ranked by number of papers. The top three journals are HISTORICAL SOCIAL RESEARCH-HISTORISCHE SOZIALFORSCHUNG, ELECTRONIC LIBRARY, and PROFESIONAL DE LA INFORMACION. Most of journal contains only 1 or 2 Digital Humanities paper. This indicates the interdisciplinary nature of Digital Humanities that can not be covered by only several number of journals.

TABLE 1 NUMBER OF PAPERS IN JOURNALS

| Journal | No. of papers | Percentage |
|--|---------------|------------|
| HISTORICAL SOCIAL RESEARCH-HISTORISCHE SOZIALFORSCHUNG | 10 | 12.99% |
| ELECTRONIC LIBRARY | 5 | 6.49% |
| PROFESIONAL DE LA INFORMACION | 5 | 6.49% |
| ORAL HISTORY REVIEW | 4 | 5.19% |
| MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY | 3 | 3.90% |
| PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS | 3 | 3.90% |
| ASLIB PROCEEDINGS | 2 | 2.60% |
| HISTORIA CRITICA | 2 | 2.60% |
| HISTORY OF THE FAMILY | 2 | 2.60% |
| JOURNAL OF DOCUMENTATION | 2 | 2.60% |
| Other | 39 | 50.65% |
| Total | 77 | 100.00% |

B. Knowledge Structure and Centrality Calculation

Keywords for each paper are retrieved to understand

fundamental elements of research projects. Due to the fact that different words can be used for describing the same concept, it is necessary to standardize words that used to express the same concept. A total of 432 keywords are standardized into 287 keywords. Subsequently, RFP-subject network, RFP-institute network, RFP-department network, Fig. 2-5, are constructed on the basis of keyword co-occurrence by choosing paper authors' background as network actor.

1) RFP-country network:

Keywords are grouped together by first author country. Any two countries that have same keywords are linked together. As shown in Fig. 2, a total of 28 network actors and 494 network ties are obtained. The size of actor is proportional to its degree centrality. It can be observed that Digital Humanities researches are almost equally distributed in different countries (actors' sizes are similar). This is very different from traditional knowledge development that is always dominated by the US. Table 2 shows centralities of country. Although the US is still ranked as No 1, but the centralities differences between the US and other countries is not significant. This leads the size of a large number of actors in Fig. 2 to be similar to that of the US. A balanced development of Digital Humanities worldwide can be observed.

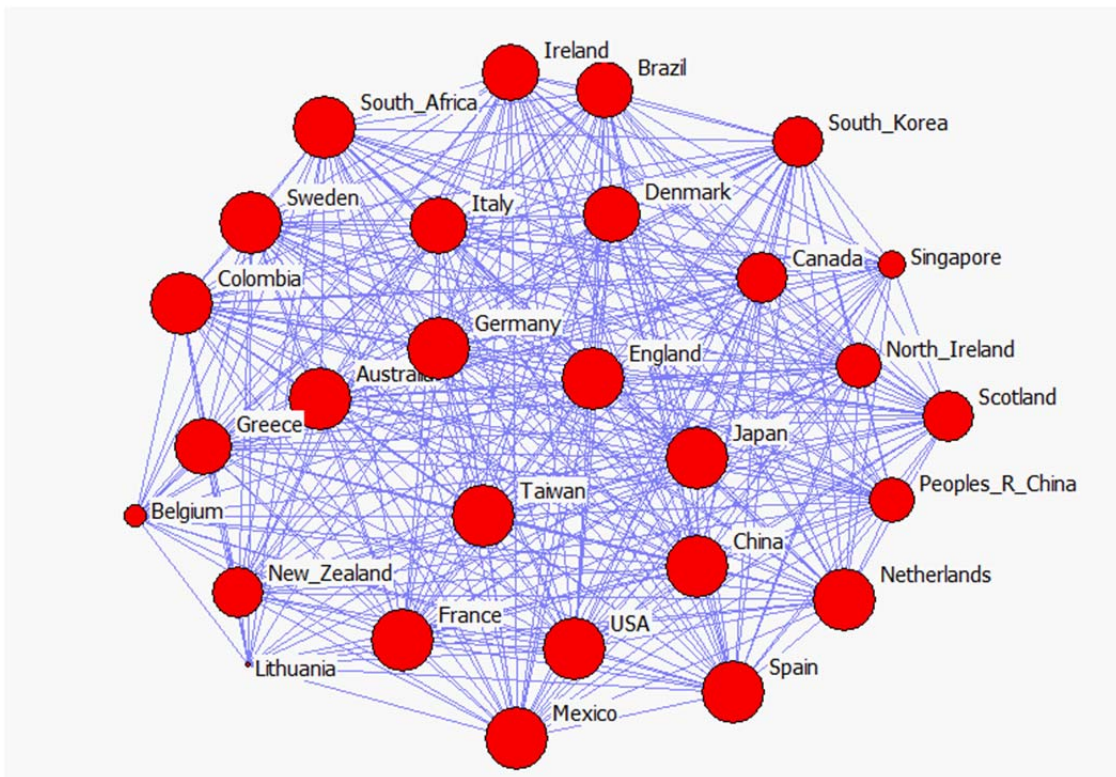


Fig. 2. RFP-country network.

2016 Proceedings of PICMET '16: Technology Management for Social Innovation

TABLE 2 TOP 15 CENTRALITIES COUNTRIES

| Ranking | Degree Centrality | Betweenness Centrality | Closeness Centrality |
|---------|-------------------|------------------------|----------------------|
| 1 | USA | USA | USA |
| 2 | Spain | Spain | Spain |
| 3 | Germany | Germany | Germany |
| 4 | Netherlands | Netherlands | Netherlands |
| 5 | Japan | Japan | Japan |
| 6 | Mexico | Mexico | Mexico |
| 7 | France | France | France |
| 8 | Taiwan | Taiwan | Taiwan |
| 9 | China | China | China |
| 10 | England | England | England |
| 11 | Australia | Australia | Australia |
| 12 | South Africa | South Africa | South Africa |
| 13 | Colombia | Colombia | Colombia |
| 14 | Sweden | Sweden | Sweden |
| 15 | Ireland | Greece | Greece |

2) RFP-institute network:

Keywords are grouped together by research institutes, any two institutes that have same keywords are linked together. A total of 74 network actors and 2790 network ties are obtained in Fig. 3. The large number of network ties means a certain ratio of research similarity can be anticipated in these 74 institutes. This suggest all these 75 institutes are not going solo but are co-evolving in the field of Digital Humanities. It is found in Table 3 that Centro de Ciencias

Humanas y Sociales, Universidad Nacional Autonoma de Mexico, National Chengchi University, New York Univ Shanghai, and University of Pretoria are the top five centralities institute. All these institutes mainly focus on social science studies. Digital Humanities integrates science and social science, but these top centralities institutes suggest that Digital Humanities are investigated mainly by social science based institutes.

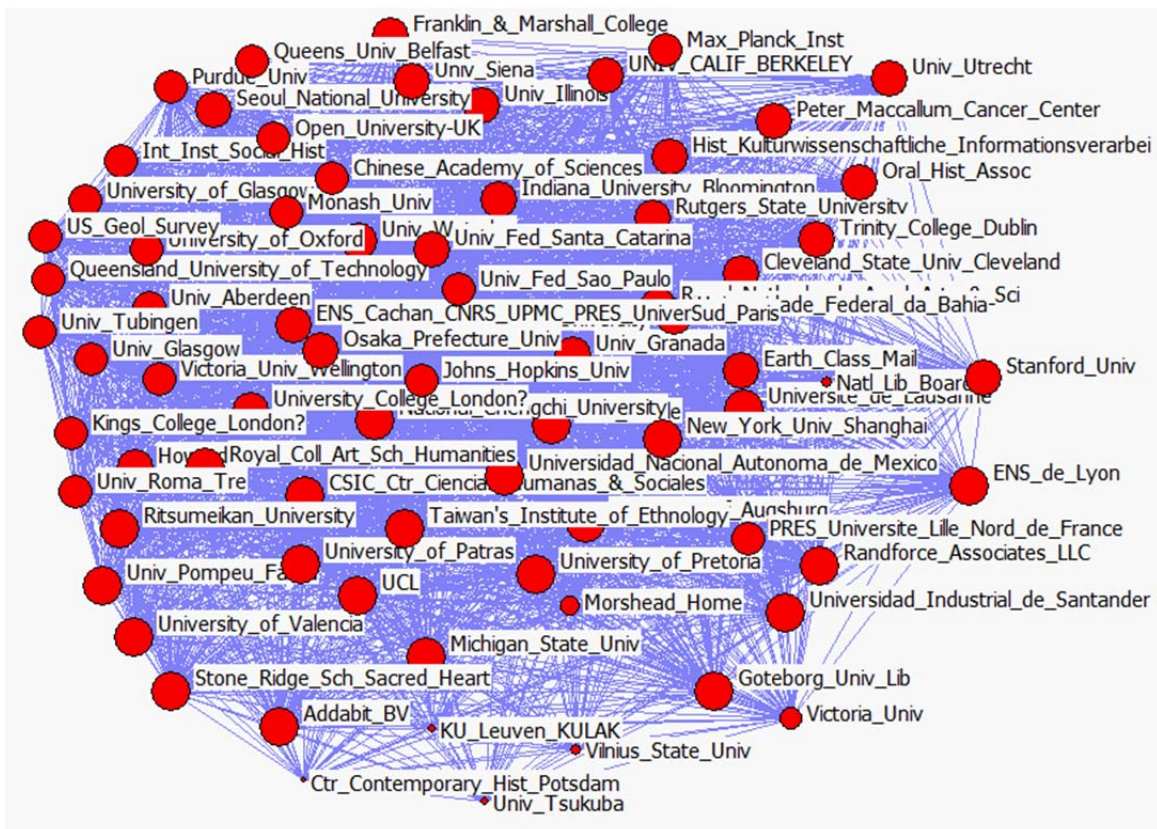


Fig. 3. RFP-institute network.

TABLE 3. TOP 15 CENTRALITIES INSTITUTES

| Ranking | Degree Centrality | Betweenness Centrality | Closeness Centrality |
|---------|---|---|---|
| 1 | CSIC Ctr Ciencias Humanas & Sociales | CSIC Ctr Ciencias Humanas & Sociales | CSIC Ctr Ciencias Humanas & Sociales |
| 2 | Universidad Nacional Autonoma de Mexico | Universidad Nacional Autonoma de Mexico | Universidad Nacional Autonoma de Mexico |
| 3 | National Chengchi University | National Chengchi University | National Chengchi University |
| 4 | New York Univ Shanghai | New York Univ Shanghai | New York Univ Shanghai |
| 5 | University of Pretoria | University of Pretoria | University of Pretoria |
| 6 | UCL | UCL | UCL |
| 7 | Michigan State Univ | Michigan State Univ | Michigan State Univ |
| 8 | Universidad Industrial de Santander | Universidad Industrial de Santander | Universidad Industrial de Santander |
| 9 | Randforce Associates LLC | Randforce Associates LLC | Randforce Associates LLC |
| 10 | University of Augsburg | University of Augsburg | University of Augsburg |
| 11 | Goteborg Univ Lib | Goteborg Univ Lib | Goteborg Univ Lib |
| 12 | Taiwan's Institute of Ethnology | Taiwan's Institute of Ethnology | Taiwan's Institute of Ethnology |
| 13 | Univ Pompeu Fabra | Univ Pompeu Fabra | Univ Pompeu Fabra |
| 14 | Univ Seville | Universite de Lausanne | Universite de Lausanne |
| 15 | Universite de Lausanne | Royal Coll Art Sch Humanities | Royal Coll Art Sch Humanities |

3) RFP-Author network

Keywords are grouped together by author, any two authors that have same keywords are linked together. A total of 77 network actors and 3102 network ties are obtained in Fig. 4. Similar to Fig. 3, the large number of network ties suggest all these 77 authors are co-evolving in the field of Digital Humanities. It is found in Table 4 that 1) Priani,

Saiso Ernesto, 2) Pun, R., 3) Rodriguez-Yunta L., 4) Galina, Russell Isabel, 5) Chen, Chih Ming are the top centralities researchers. The researches they conducts have the highest generality in the field of Digital Humanities and might be able to be regarded as fundamental Digital Humanities researches.

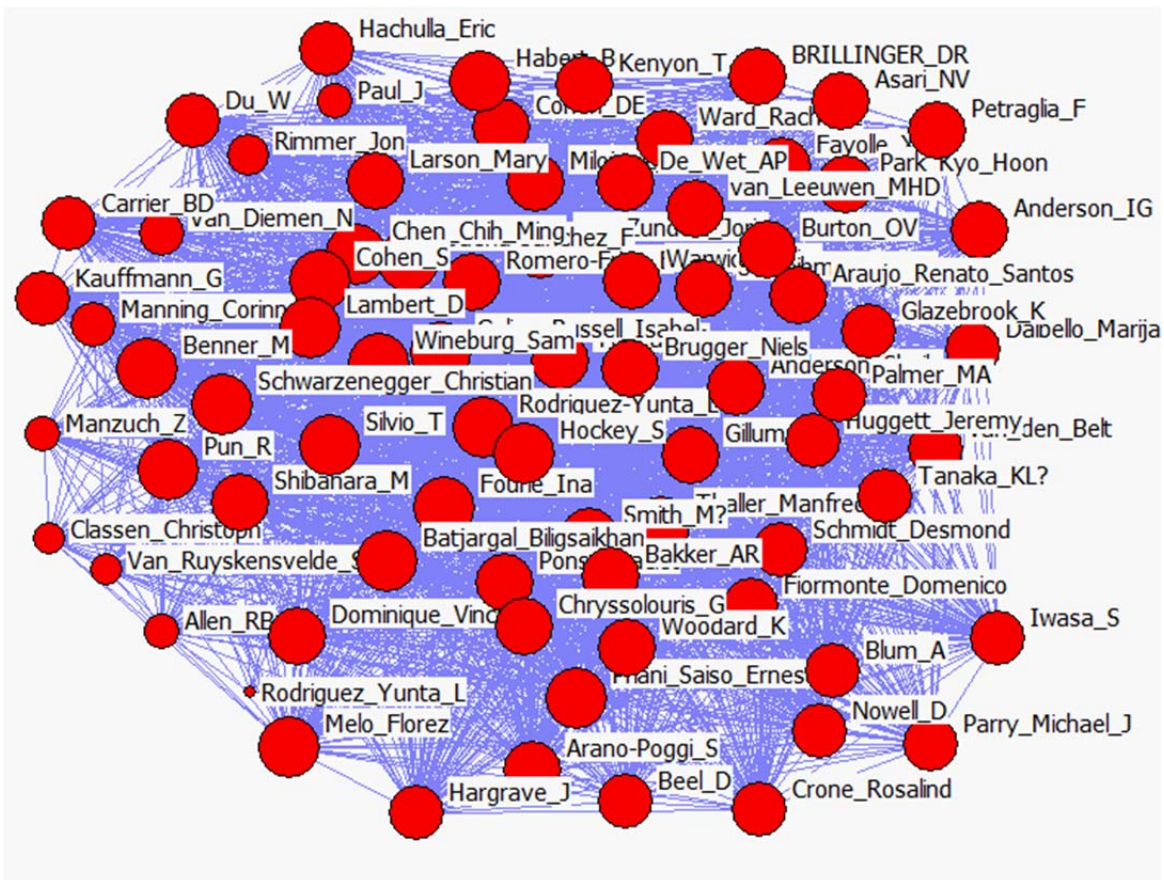


Fig. 4. RFP- author network.

TABLE 4 TOP 15 CENTRALITIES AUTHORS

| Ranking | Degree Centrality | Betweenness Centrality | Closeness Centrality |
|---------|--------------------------|------------------------|--------------------------|
| 1 | Priani Saiso Ernesto | Priani Saiso Ernesto | Priani Saiso Ernesto |
| 2 | Pun R | Pun R | Pun R |
| 3 | Rodriguez-Yunta L | Wineburg Sam | Rodriguez-Yunta L |
| 4 | Galina Russell Isabel | Fayolle X | Galina Russell Isabel |
| 5 | Chen Chih Ming | Shibahara M | Chen Chih Ming |
| 6 | Fourie Ina | Larson Mary | Fourie Ina |
| 7 | Hockey S | Blum A | Hockey S |
| 8 | Cohen S | Nowell D | Cohen S |
| 9 | Melo Florez | Rodriguez-Yunta L | Melo Florez |
| 10 | Lambert D | Galina Russell Isabel | Lambert D |
| 11 | Schwarzenegger Christian | Chen Chih Ming | Schwarzenegger Christian |
| 12 | Benner M | Fourie Ina | Benner M |
| 13 | Silvio T | Hockey S | Silvio T |
| 14 | Baena-Sanchez F | Cohen S | Wineburg Sam |
| 15 | Batjargal Biligsaikhan | Melo Florez | Batjargal Biligsaikhan |

4) KCO network:

Keywords are grouped together by research department, any two institutes that have same keywords are linked together. A total of 287 network actors and 1243 network ties are obtained, Fig. 5. The 287 networks can be categorized into diverse discipline. The phenomenon indicates that Digital Humanities is a cross-discipline research field which encourages various types of investigations in different discipline. TABLE 5 lists top 15 high centralities departments. The top departments with highest centralities are digital humanities, information, digital libraries, university, institutional, history, archives, digital images, etc. It can be observed that these top centralities keywords are

mainly related to digital archives related. This indicates digital archives is the most fundamental issue in Digital Humanities. The work of Digital Humanities first requires document to be digitized or digital archives to be available. By examining the top centralities keywords, it can also be found that the digitization was first introduced in the history and library, and this is why the two terms “digital library” and “digital history” are on the top centrality keyword list. However, it can be anticipated that digitization of other fields is also an inevitable process to allow digitized documents to be analyze. Therefore, new information can be unveiled by digital technologies and new knowledge can thus be developed.

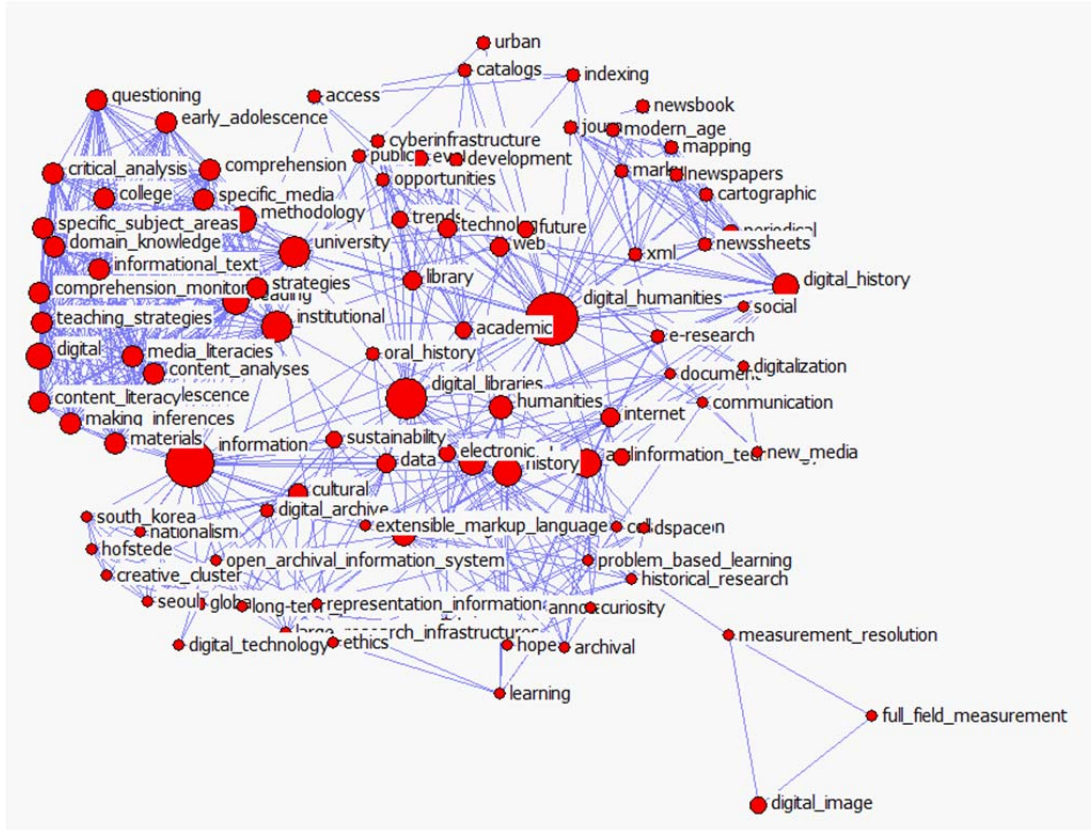


Fig. 5. KCO network (Degree centrality > 7).

TABLE 4 TOP 15 CENTRALITIES KEYWORDS

| Ranking | Degree Centrality | Betweenness Centrality | Closeness Centrality |
|---------|--------------------|------------------------|----------------------|
| 1 | digital humanities | digital humanities | digital humanities |
| 2 | information | information | information |
| 3 | digital libraries | research | digital libraries |
| 4 | university | measurement resolution | history |
| 5 | institutional | digital libraries | research |
| 6 | history | history | humanities |
| 7 | archives | digital image | university |
| 8 | digital history | urban | internet |
| 9 | reading | archives | academic |
| 10 | methodology | digital history | library |
| 11 | research | humanities | institutional |
| 12 | digital | periodical | digital history |
| 13 | knowledge | institutional | knowledge |
| 14 | humanities | cultural | technology |
| 15 | adolescence | information technology | archives |

The above results indicate that Digital Humanities is a cross-discipline research field which encourages diverse types of investigations in different discipline. It is so far in the stage of applying IT technology to conventional media, i.e. digitizing historical documents. Although different ways of digital archiving methodologies [14] have been proposed by researchers in their own field or region. It is expected that essential methodologies should be proposed to standardize the way to digitize or preserve historical documents in order to facilitate communications between different researcher and global collaboration in the field of digital humanities. The next stage should be expected to be the analysis of digitized document by proper techniques to uncover hidden implications in large volume of document. The recent fast-paced development of Big Data research [15][16] will satisfy the need of proper techniques for analyzing large set of data to discover important message or implication that are difficult to be retrieved from conventional method. However, the keyword “Big Data” was not found in the keyword list of this research because Big data is still under development or it is still in its early stage of analyzing numerical data. However, it is clear that the effort contributed by Big data will have positive impact to the field of Digital Humanities in the near future.

IV. CONCLUSION

This paper utilized keyword based method to map scientific structure of Digital Humanities research. It can be observed that Digital Humanities is a growing research field in modern society. The obtained knowledge structure can be served as evidence for understanding how to position each paper in the global Digital Humanity research map. The study analyzes 77 scientific journal papers obtained from SSCI database to create three-dimensional FRP-network and calculate network centralities to obtain quantitative analyses on country, institute, author and keyword networks. The investigation of the three different actors- country, institute, author and keyword networks represents observation on the overview of global Digital Humanities research in macro, meso, and micro levels, respectively. Policy implications can

be summarized as 1) Digital Humanity research has been first introduced in digital archive digital library as well as digital library and will be continuously explored in other fields; 2) the development of Digital Humanities is relatively well-distributed in different countries which indicates similar degree of awareness of Digital Humanities in different, 3) The complex network ties and small network centrality difference indicate those network actors are co-evolving in a balanced way. 4) the maturity of Digital Humanities has a lot to do with how efficient new knowledge can be uncovered by digital technologies.

The development of a knowledge-based economy relies significantly on information technology and Humanity development can never be ignored in this society. Digital technology plays a more and more significant role in the development of Humanities and is therefore recommended to be monitored on a regular basis. This study employs methodology based on network theory and keyword analysis. The method used in this paper demonstrates an evidence-based approach for managing resources, knowledge mapping and visualization as well as performance evaluation. The same methodology can be used in other research fields, too. The major research limitation is that it is difficult to define precisely the range of Digital Humanities. The query strategy used in this study is not perfect and should be evaluated in the near future. Digital Humanity has been evolving from Digital Archive, Digital History, Digital Library to other interdisciplinary field. It can be expected that digital tool will be implemented into all humanities studies and thus conventional humanities researches will all become digital humanities. Therefore, the analysis of Digital Humanities is only a snapshot. The desirable longitudinal analysis on the development of Digital Humanities is necessary but is extremely difficult because the definition of Digital Humanities is changing rapidly over time.

REFERENCES

[1] Y. Ding, G. G. Chowdhury, and S. Foo, “Bibliometric cartography of information retrieval research by using co-word analysis,” *Inf. Process. Manag.*, vol. 37, no. 6, pp. 817–842, 2001.

2016 Proceedings of PICMET '16: Technology Management for Social Innovation

- [2] C. Baldwin, J. Hughes, T. Hope, R. Jacoby, and S. Ziebland, "Ethics and dementia: mapping the literature by bibliometric analysis," *Int. J. Geriatr. Psychiatry*, vol. 18, no. 1, pp. 41–54, 2003.
- [3] Y. Tian, C. Wen, and S. Hong, "Global scientific production on GIS research by bibliometric analysis from 1997 to 2006," *J. Informetr.*, vol. 2, no. 1, pp. 65–74, 2008.
- [4] H.-N. Su and P.-C. Lee, "Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in Technology Foresight," *Scientometrics*, vol. 85, no. 1, pp. 65–79, Jun. 2010.
- [5] Z. J. Acs, L. Anselin, and A. Varga, "Patents and innovation counts as measures of regional production of new knowledge," *Res. Policy*, vol. 31, no. 7, pp. 1069–1085, 2002.
- [6] Y.-T. Park and M.-S. Kim, "A Taxonomy of Industries Based on Knowledge Flow Structure," *Technol. Anal. Strateg. Manag.*, vol. 11, no. 4, pp. 541–549, 1999.
- [7] P. C. Lee and H. N. Su, "Investigating the structure of regional innovation system research through keyword co-occurrence and social network analysis," *Innov. Manag. Policy Pract.*, 2010.
- [8] P. C. Lee, H. N. Su, and F. S. Wu, "Quantitative mapping of patented technology—The case of electrical conducting polymer nanocomposite," *Technol. Forecast. Soc. Change*, vol. 77, no. 3, pp. 466–478, 2009.
- [9] P. C. Lee and H. N. Su, "Quantitative mapping of scientific research—The case of electrical conducting polymer nanocomposite," *Technol. Forecast. Soc. Change*, 2010.
- [10] L. McLeod, J. Broady-Preston, and R. Atkins, "Accessing Humanities Research in a Digital Environment," *Procedia - Soc. Behav. Sci.*, vol. 147, pp. 140–146, Aug. 2014.
- [11] C. A. Sula, "Digital Humanities and Libraries: A Conceptual Model," *J. Libr. Adm.*, vol. 53, no. 1, pp. 10–26, Jan. 2013.
- [12] D. J. Brass and M. E. Burkhardt, "Centrality and power in organizations," *Netw. Organ. Struct. Form Action*, vol. 191, p. 215, 1992.
- [13] L. C. Freeman, "Centrality in social networks conceptual clarification," *Soc. Netw.*, vol. 1, no. 3, pp. 215–239, 1978.
- [14] A. Sabharwal, "3 - Digital history, archives, and curating digital cultural heritage," in *Digital Curation in the Digital Humanities*, Chandos Publishing, 2015, pp. 49–67.
- [15] U. Iqbal, C.-K. Hsu, P. A. (Alex) Nguyen, D. L. Clinciu, R. Lu, S. Syed-Abdul, H.-C. Yang, Y.-C. Wang, C.-Y. Huang, C.-W. Huang, Y.-C. Chang, M.-H. Hsu, W.-S. Jian, and Y.-C. (Jack) Li, "Cancer-disease associations: A visualization and animation through medical big data," *Comput. Methods Programs Biomed.*, vol. 127, pp. 44–51, Apr. 2016.
- [16] W. S. Lee, E. J. Han, and S. Y. Sohn, "Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents," *Technol. Forecast. Soc. Change*, vol. 100, pp. 317–329, Nov. 2015.