Analyzing the research output presented at European Networked Knowledge Organization Systems workshops (2000-2015)

Fakhri Momeni and Philipp Mayr

GESIS - Leibniz Institute for the Social Sciences,
Unter Sachsenhausen 6-8
50667 Cologne, Germany
firstname.lastname@gesis.org

Abstract. In this paper we analyze a major part of the research output of the Networked Knowledge Organization Systems (NKOS) community in the period 2000 to 2015 from a network analytical perspective. We focus on the paper output presented at the European NKOS workshops in the last 15 years. Our open dataset, the "NKOS bibliography", includes 14 workshop agendas (ECDL 2000-2010, TPDL 2011-2015) and 4 special issues on NKOS (2001, 2004, 2006 and 2015) which cover 171 papers with 218 distinct authors in total. A focus of the analysis is the visualization of co-authorship networks in this interdisciplinary field. We used standard network analytic measures like degree and betweenness centrality to describe the co-authorship distribution in our NKOS dataset. We can see in our dataset that 15% (with degree=0) of authors had no co-authorship with others and 53% of them had a maximum of 3 cooperations with other authors. 32% had at least 4 co-authors for all of their papers. The NKOS co-author network in the "NKOS bibliography" is a typical co-authorship network with one relatively large component, many smaller components and many isolated co-authorships or triples.

Keywords: European NKOS workshops, Output analysis, Network analysis, Central authors, Collaboration

1 Introduction

The European NKOS network has held a long-running series of annual workshops at the European Conference on Digital Libraries (ECDL), latterly reformed as the International Conference on Theory and Practice of Digital Libraries (TPDL). Typically, recent advances of KOS have been reported at the NKOS workshops, e.g. including the Simple Knowledge Organization System (SKOS) W3C standard, the ISO 25964 thesauri standard, the CIDOC Conceptual Reference Model (CRM), Linked Data applications, KOS-based recommender systems, KOS mapping techniques, KOS registries and metadata, social tagging, user-centred issues, and many other topics. A comprehensive and well cited review article on KOS and NKOS topics was published in 2004 [11]. Special issues

on Networked Knowledge Organization Systems (NKOS) have been published in Journal of Digital Information in 2001 and 2004, in New Review of Hypermedia and Multimedia in 2006 and recently in International Journal of Digital Libraries in 2015 [6].

The motivation of this paper is to analyze the research output of the NKOS community. We are focusing on the informal part of this output, the paper presentations given at the past NKOS workshops (the first European NKOS workshop in 2000 to the 14th European NKOS workshop in 2015). The specialty of this research output is that these research papers typically are not published in journals or conference proceedings. These papers appear just as oral presentations at the workshop and are documented on the website.

To our knowledge nobody has done an analysis on this part of the NKOS research output before.

2 NKOS workshop bibliography

For our analysis we have compiled an open dataset the "NKOS bibliography"[1] which includes 14 workshop programs with all presented papers at ECDL 2000, ECDL 2003-2010 and TPDL 2011-2015 (see Table 1). We added papers from 4 special issues on NKOS which have been edited by members of the NKOS community in the same period (see Table 2).

Table 1. Overview of all NKOS workshop papers

venue	papers	authors
ECDL 2000	4	4
ECDL 2003	13	11
ECDL 2004	14	27
ECDL 2005	12	26
ECDL 2006	12	27
ECDL 2007	15	26
ECDL 2008	11	16
ECDL 2009	12	31
ECDL 2010	12	25
TPDL 2011	11	26
TPDL 2012	9	22
TPDL 2013	7	17
TPDL 2014	9	17
TPDL 2015	7	13

In a first step we have extracted all paper titles presented at the NKOS workshop websites. We excluded welcome and introduction presentations. We

[1] The NKOS workshop bibliography is maintained in the following github repository: https://github.com/PhilippMayr/NKOS-bibliography

added journal papers from the four mentioned special issues on NKOS. These journal papers are the only formal publications in our analysis. In the end we manually disambiguated author names of all papers.

Our dataset covers 171 papers in total with a sum of 218 distinct author names. Table 1 provides an overview of all workshop papers[2]. We can see a slightly decreasing amount of papers presented at the workshops after ECDL has renamed to TPDL in 2011. In average 1.9 authors published a workshop paper.

Table 2. Overview of all NKOS special issue papers

venue	papers	authors
JODI 2001 [5]	5	8
JODI 2004 [9]	5	15
NREV 2006 [10]	6	11
IJDL 2015 [6]	7	20

Table 2 provides an overview of all papers in the special issues. We can see that a relative constant number of papers in the issues have an increasing number of authors. In average 2.3 authors published a special issue journal paper.

3 Analysis

In order to analyze the collaboration of the NKOS community we built a network of all authors at the workshops and special issues and computed the centrality of each author. For this purpose we utilized some standard centrality measures in Pajek[3]. The network is composed of pairs of author names. Each pair means that two authors cooperated for writing a paper. If we have n papers and the paper i has m_i authors, the number of pairs are

$$\sum_{i=1}^{n} \frac{m_i(m_i-1)}{2} \qquad if \quad m_i \geq 1 \tag{1}$$

These pairs built the network for our analysis in Pajek. To avoid repetition of pairs, we gave weight to pairs and this is equal to the number of cooperations of two authors in different papers. Two often used centrality measures of authors are degree and betweenness. Degree is the number of nodes that a focal node is connected to and measures the involvement of the node in the network [7]. In our authorship-network it specifies the sum of co-authors for all papers that each author has written. Betweenness assesses the degree to which a node lies

[2] See readme for details on the workshops under https://github.com/PhilippMayr/NKOS-bibliography/blob/master/readme.txt

[3] A program for analysis and visualization of very large networks (http://mrvar.fdv.uni-lj.si/pajek/)

on the shortest path between two other nodes and is able to funnel the flow in the network [7]. In the authorship-network the author with a high betweenness has a large influence on the transfer of information.

4 Results

Figure 1 demonstrates a general view of the network. In this view each author has at least one co-author. This network contains 31 components. From the network illustrated in this figure we selected the largest component that is represented in Figure 2. 68 authors (31% of all authors) are connected in this component. The NKOS co-author network in the "NKOS bibliography" is a typical co-authorship network with one relatively large component, many smaller components and many isolated co-authorships or triples.

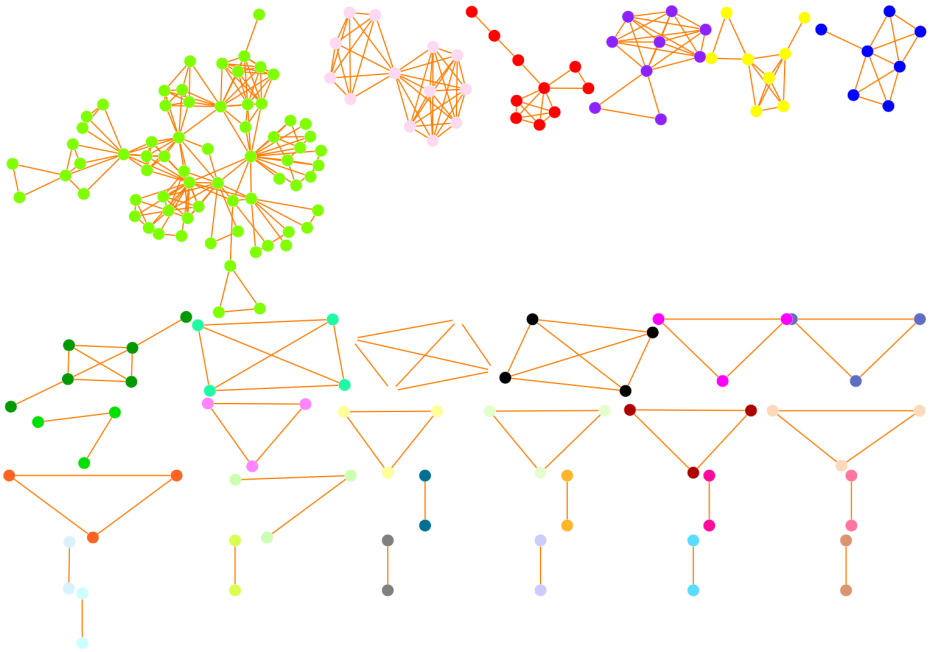

Fig. 1. Co-authorship network in the European NKOS community

To show the quantity of collaboration in the community we measured the degree centrality for each author. Figure 3 shows the percentage of authors with different degrees. In this figure we see that 15% (with degree=0) of authors had no co-authorship with others and 53% of them had a maximum of 3 cooperations with other authors. 32% had at least 4 co-authors for all their papers.

Figure 4 shows the authors with high degree (more than 8) in the network.

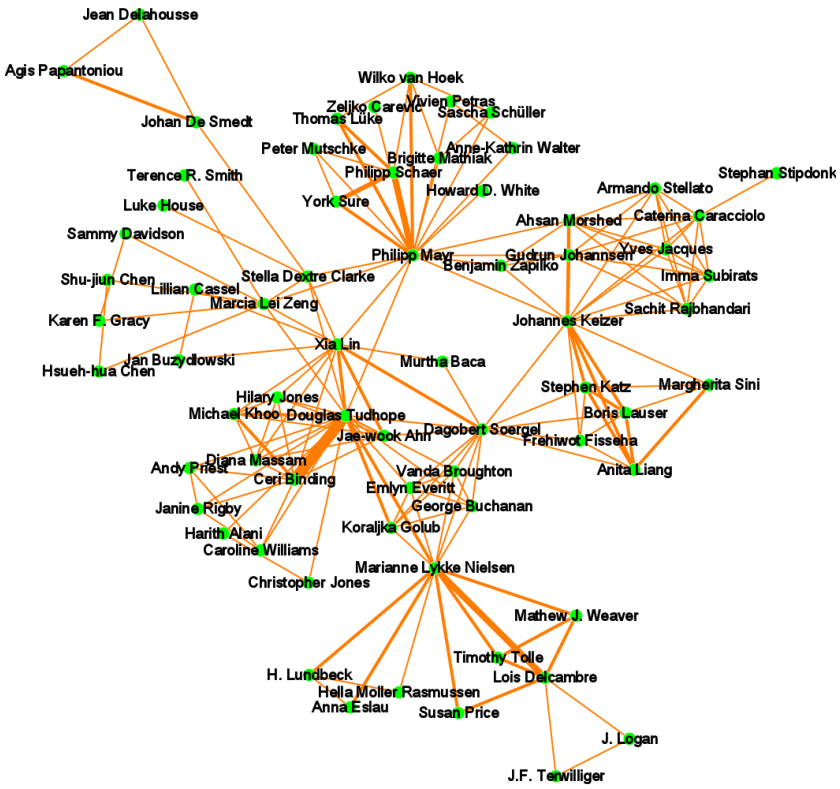

Fig. 2. Largest component in the NKOS co-authorship network

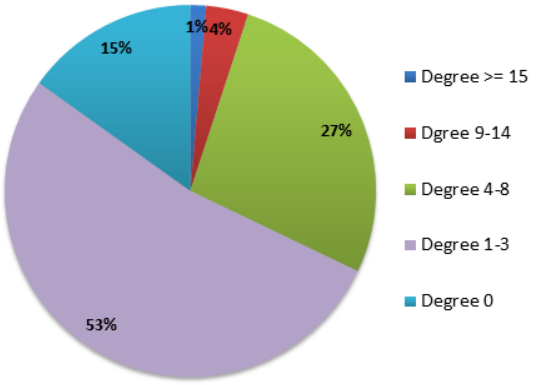

Fig. 3. Distribution of degree numbers of authors in the network

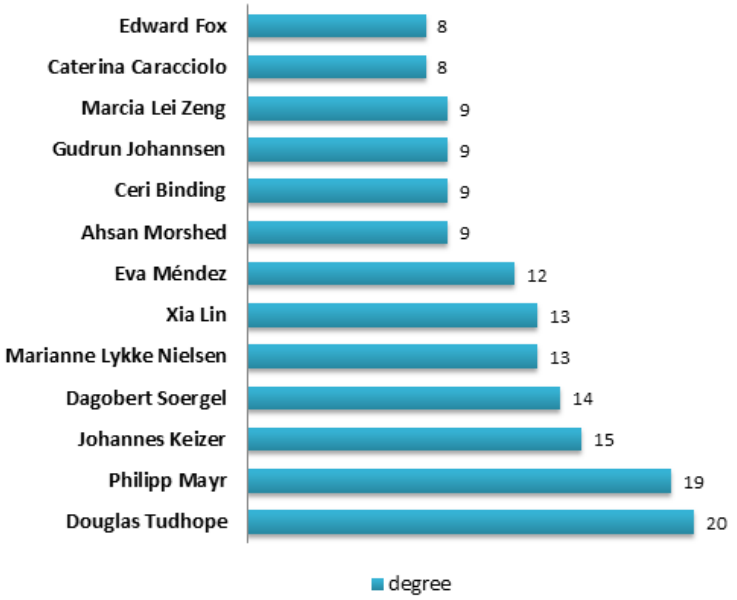

Fig. 4. Authors with degree more than 8

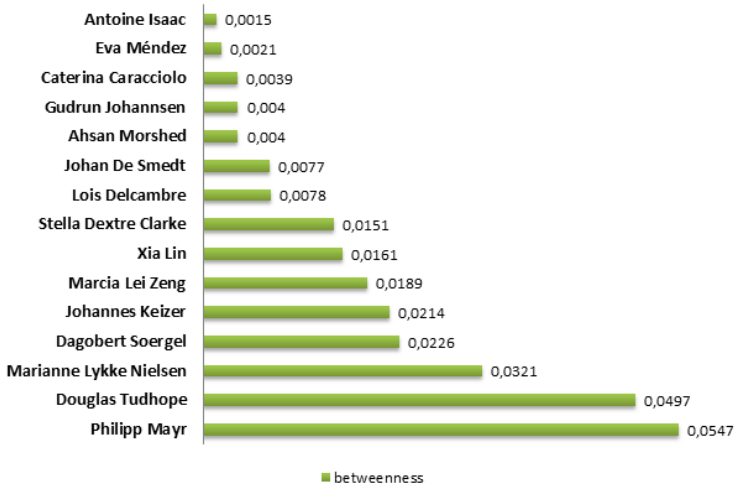

Fig. 5. Authors with betweenness more than 0.001

To detect the influence of authors on information exchange we calculated the betweenness centrality of authors. Figure 5 indicates the authors with high betweenness (more than 0.001). If we compare betweenness with degree we can see that the ranking of authors has changed. We observe that some authors have lower ranking in betweenness (despite their high cooperation with other authors) in comparison to others with lower degree. Comparing the largest component in Figure 2 with Figure 4 and Figure 5 we see that most authors with top centrality (top degree or top betweenness) are in this component. Nevertheless there are 3 authors ('Eva Mendez', 'Ceri Binding' and 'Antoine Isaac') who are not in the largest component, but are in the list of top authors for betweenness or degree centrality.

5 Conclusion

In this paper we have started to analyze the collaborative research of authors and their connectivity to each other in the special case of European NKOS workshop activities and four special issues on NKOS. The results show most active authors in this community who had an important role in the exchange of information exchange and in connecting researchers. We saw some details of the largest component in this network that covers one third of the authors of the whole network. Also we observed that most authors with highest centrality (degree and betweenness) are in this component. Our analyses show that NKOS workshops were successful in bringing researchers from different domains together.

We know that our dataset has some severe limitations. First of all we have included just paper presentations. Editing and organizing activities at the workshops have not been covered in our dataset. This leads to artifacts; e.g. Traugott Koch,[4] a long-term organizer of the NKOS workshops and editor of the early JoDI special issues on NKOS, is not included in our dataset and the network.

Second, we have not included the activities of the NKOS community in the United States of America. The website at Kent State[5] would be a great starting point to look up more research activities of the US NKOS community.

Third, we have not included bibliometric data to complete our analysis. This is because most of the NKOS workshop activities (presentations) are not formally cited or even mentioned in scientific papers. In difference to the workshop output, the few journal papers in the special issues on NKOS are cited. Some works (e.g. [1, 3, 4, 2, 8]) are cited well in the literature.

6 Future work

We are planning to extend the analysis of the NKOS network. In this way we first plan to complement the dataset with other NKOS research output. We

[4] Traugott Koch was an central protagonist and networker of the European NKOS community. He retired and left the NKOS community in 2012.

[5] http://nkos.slis.kent.edu/

also plan to analyze the development of topics in the titles and abstracts of the presentations and papers. Combining network analytic measures with bibliometric analysis (e.g. co-citations, bibliographic coupling) would complement our preliminary observations. We invite people to contribute to the data set.

7 Acknowledgment

We thank our colleague Julia Achenbach (GESIS) who helped us to compile and correct the dataset. We thank Peter Mutschke (GESIS) who helped us with the interpretation of the NKOS co-authorship network.

References

1. Cranefield, S.: Networked knowledge representation and exchange using uml and rdf. Journal of Digital Information (2001), https://journals.tdl.org/jodi/index.php/jodi/article/view/30
2. Dagobert Soergel, Boris Lauser, A.L.F.F.J.K., Katz, S.: Reengineering thesauri for new applications: the agrovoc example. Journal of Digital Information (2004), https://journals.tdl.org/jodi/index.php/jodi/article/view/112
3. Doerr, M.: Semantic problems of thesaurus mapping. Journal of Digital Information (2001), https://journals.tdl.org/jodi/index.php/jodi/article/view/31
4. Douglas Tudhope, H.A., Jones, C.: Augmenting thesaurus relationships: possibilities for retrieval. Journal of Digital Information (2001), https://journals.tdl.org/jodi/index.php/jodi/article/view/181/160
5. Hill, L., Koch, T.: Networked Knowledge Organization Systems: introduction to a special issue. Journal of Digital Information 1(8) (2001), https://journals.tdl.org/jodi/index.php/jodi/article/view/32/33
6. Mayr, P., Tudhope, D., Clarke, S.D., Zeng, M.L., Lin, X.: Recent applications of Knowledge Organization Systems: introduction to a special issue. International Journal on Digital Libraries 17(1), 1–4 (2016), http://link.springer.com/10.1007/s00799-015-0167-x
7. Opsahl, T., Agneessens, F., Skvoretz, J.: Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks 32(3), 245–251 (2010), http://dx.doi.org/10.1016/j.socnet.2010.03.006
8. Trant, J., with the participants in the steve.museum project: Exploring the potential for social tagging and folksonomy in art museums: Proof of concept. New Review of Hypermedia and Multimedia (2006), http://www.tandfonline.com/doi/abs/10.1080/13614560600802940
9. Tudhope, D., Koch, T.: New Applications of Knowledge Organization Systems: introduction to a special issue. Journal of Digital Information 4(4) (2004), https://journals.tdl.org/jodi/index.php/jodi/article/view/109/108
10. Tudhope, D., Lykke Nielsen, M.: Introduction to Knowledge Organization Systems and Services. New Review of Hypermedia and Multimedia 12(1), 3–9 (2006)
11. Zeng, M.L., Chan, L.M.: Trends and Issues in Establishing Interoperability Among Knowledge Organization Systems. Journal of the American Society for Information Science and Technology 55(3), 377–395 (2004)