# Knowledge Organization and the 2015 UDC Seminar: An Editorial

Xin Cai, Melissa Davey Castillo, Musa Dauda Hassan,
Ann Graf, Hyoungjoo Park and Richard P. Smiraglia

Information Organization Research Group, School of Information Studies,
University of Wisconsin, Milwaukee, Northwest Quad Building B, 2025 E Newport Street,
Milwaukee, WI 53211, USA, smiragli@uwm.edu

Cai, Xin, Melissa Davey Castillo, Musa Dauda Hassan, Ann Graf, Hyoungjoo Park and Richard P. Smiraglia. 2016. "Knowledge Organization and the 2015 UDC Seminar: An Editorial." *Knowledge Organization* 43(6): 395-402. 3 references.

The 2015 biennial international Universal Decimal Classification Consortium (UDCC) seminar was held at The National Library of Portugal in Lisbon, Portugal, on 29-30 October 2015. The 2015 seminar was titled "Classification & Authority Control: Expanding Resource Discovery," and brought together 95 researchers and academics from 26 nations. This editorial follows in a sequence of earlier domain analyses of UDCC seminars with the purpose of establishing the extension and intension of the UDCC seminar, and of noting its overlap with the broader knowledge organization (KO) domain. For various reasons, this particular seminar was quite different from its intellectual predecessors—that is, it was quite different from earlier UDCC seminars, and it also was quite different from the core of the KO domain. The topic of "authority control" seems to have moved the audience away from classification as a scientific mode toward librarianship; few markers of traditional KO were present. Our informetric visualizations help to demonstrate this difference.

## 1.0 Some domain metrics

The conference consisted of twenty-six formal papers, of which twenty-four were full papers printed in the proceedings and two appeared as abstracts. Another seven contributions consisted of posters or short papers. The twenty-six formal papers had forty-eight authors, which demonstrates a significant increase in collaboration over previous UDCC seminars (mean 1.86 authors per paper; range 1-4). The authors of the papers listed fifteen national affiliations as shown in Figure 1; the largest cluster was from the United States. The national affiliations are consistent with the 2013 UDCC seminar (Beak et al. 2014, 192).

There were 279 citations in the twenty-six papers; the mean number of citations was 10.07 with a range from 8 to 25. This is a much lower mean by approximately half than the 2013 seminar (Beak et al. 2014, 191), which is consistent with a greater emphasis on applied reports than on research.

The year of cited works ranged from 1769 to 2015 and the mean age of cited work was 13.9 years indicating

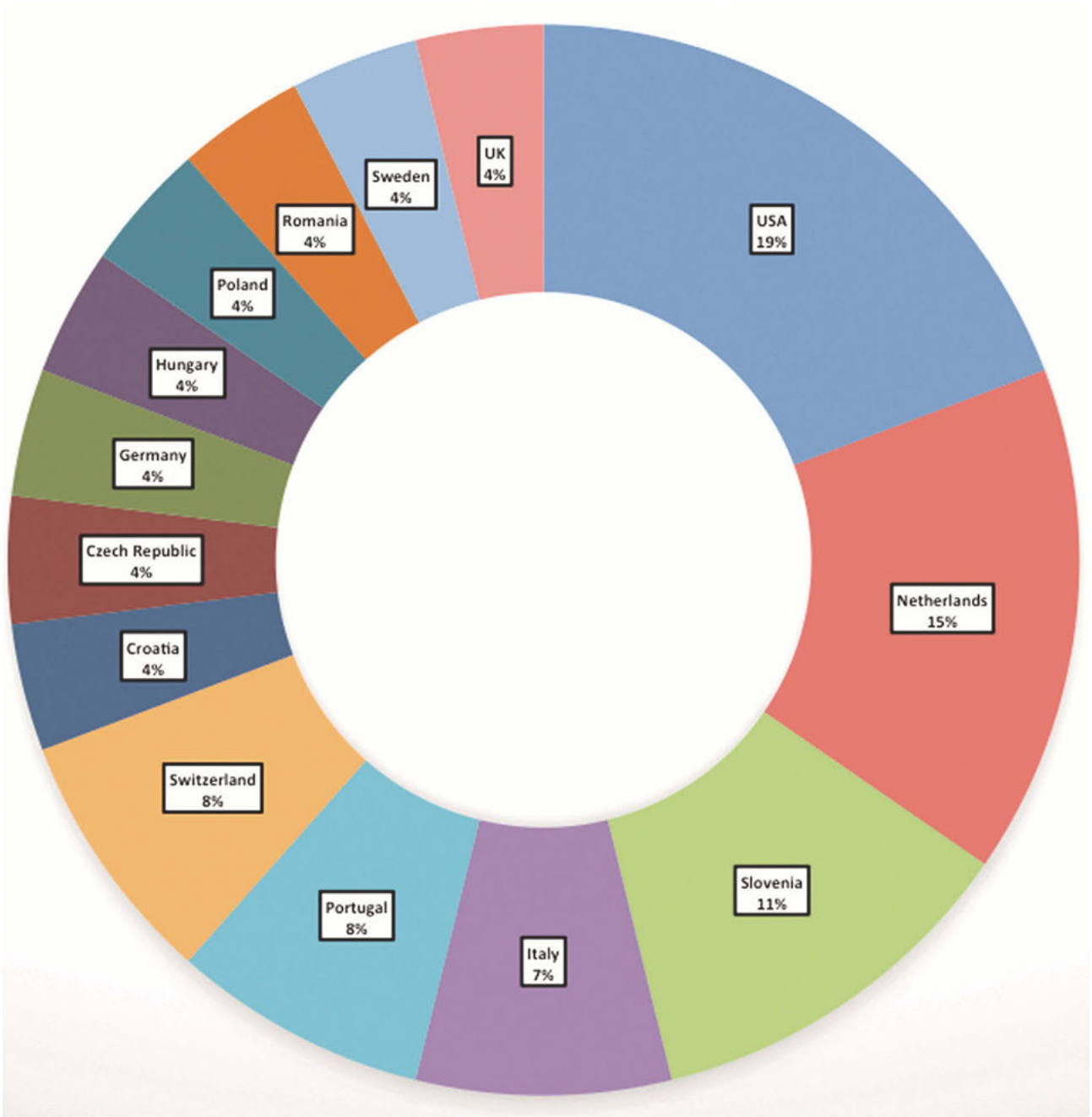



*Figure 1.* Countries of affiliation of contributing authors.

the majority of works cited were recent publications. The distribution of dates of publication of works cited is visualized in Figure 2. Seventy-six percent of the works cited were published in 2000 or later, which is comparable with the earlier seminar and shows the relative contemporaneity of the source works.

Ninety of the 279 citations were to journal articles (32%). Another 37 (13%) were to sources available only via the internet, 35 (12.5%) were to papers in conference proceedings, 48 (17%) to monographs and 3 (1%) to dissertations. The remaining 66 citations (23%) were to documents that could be considered unpublished but widely disseminated technical reports. This is quite a different distribution of sources than is typical in knowledge organization where sources usually are approximately evenly divided between journal articles and conference papers, on the one hand, and monographs on the other (Smiraglia 2013, 6). As well, the list of journals is quite different from the usual group associated with KO research (Table 1).

*Cataloging & Classification Quarterly*, *Knowledge Organization*, *Journal of Documentation*, and the *Journal of the Association for Information Science and Technology* (and its predecessors) are typically found in KO research publications. But the extension of the domain of the conference is clearly illustrated by both the top-ranked *Extensions & Corrections to the UDC* and the long list of publications with only one or two citations, which are divided among various journals of librarianship and periodicals with semantic web associations.

The titles of the references of the conference papers were entered into Provalis Research's WordStat™ software, which can be used to analyze the keywords and phrases in a corpus of text, and also shows terms that are

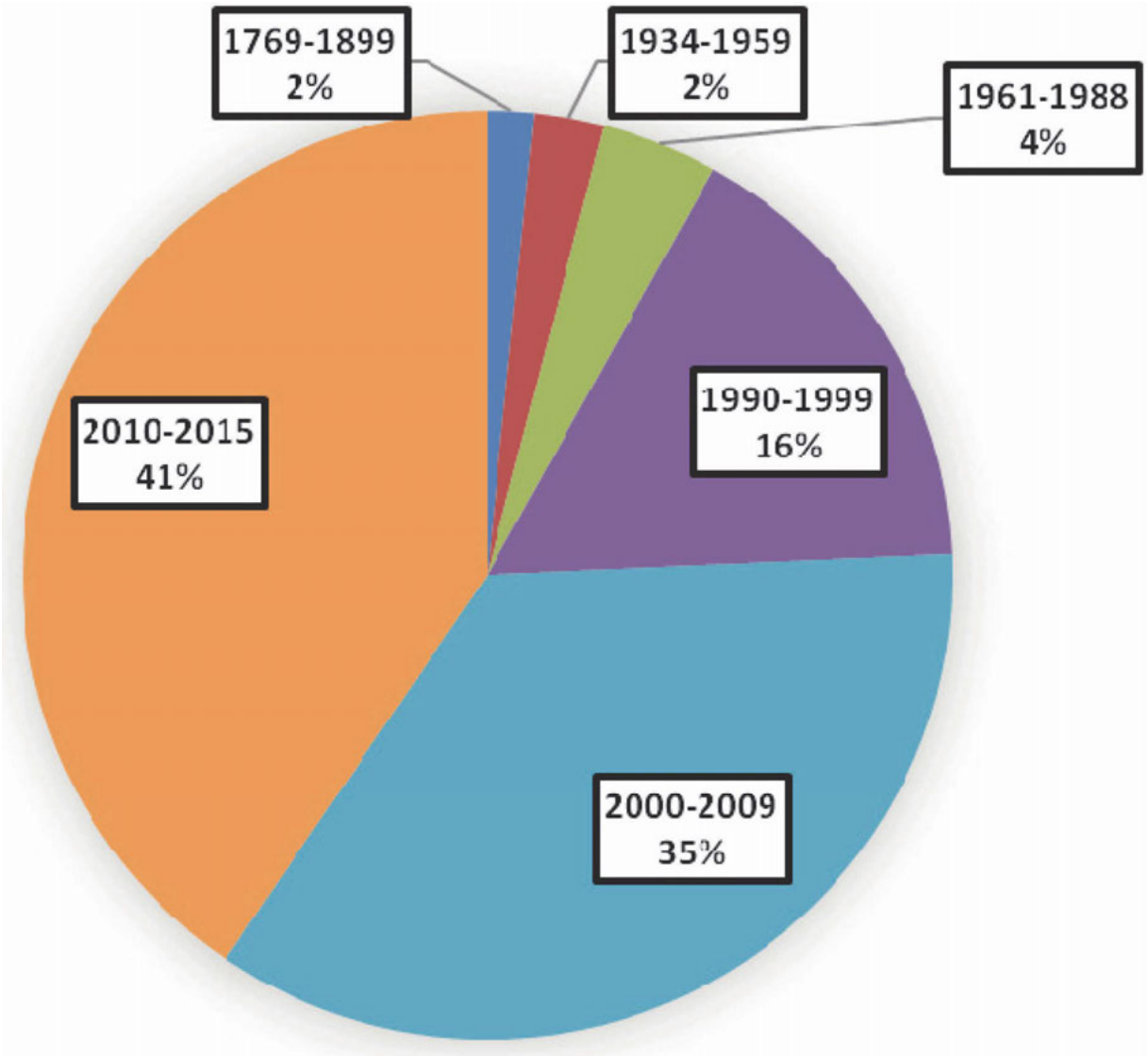



*Figure 2.* Years of works cited.

| Journal Title | Frequency |
|---|---|
| *Extensions & Corrections to the UDC* | 17 |
| *Cataloging & Classification Quarterly* | 12 |
| *Knowledge Organization* | 10 |
| *Journal of Documentation* | 8 |
| *Journal of the Association for Information Science and Technology* | 8 |
| *IBM Journal of Research and Development* | 2 |
| *JLIS.it: Italian Journal of Library, Archives and Information Science* | 2 |
| *Journal of Digital Information* | 2 |
| *Organizacija znanja* | 2 |
| *Advances in Neural Information Processing Systems* | 1 |
| *Aslib proceedings* | 1 |
| *Axiomathes* | 1 |
| *Bulletin of the Association for Information Science and Technology* | 1 |
| *Business Information Review* | 1 |
| *Computer Networks & ISDN Systems, 30, pp. 646-648.* | 1 |
| *D-Lib Magazine* | 1 |
| *Documentaliste-Sciences de l'Information* | 1 |
| *IEEE Software* | 1 |
| *IFLA Journal* | 1 |
| *Information research* | 1 |
| *Information Standards Quarterly* | 1 |
| *International Cataloguing and Bibliographic Control* | 1 |
| *International Classification* | 1 |
| *International Journal of Knowledge Content Development & Technology* | 1 |
| *Journal of Academic Librarianship* | 1 |
| *Journal of Information Retrieval* | 1 |
| *Journal of information Science* | 1 |
| *Library Hi Tech* | 1 |
| *Library Quarterly* | 1 |
| *Library Trends* | 1 |
| *Newsletter of the European Mathematical Society* | 1 |
| *NextSpace* | 1 |
| *Signum* | 1 |
| *SRELS Journal of Information Management* | 1 |
| *Journal of Machine Learning Research* | 1 |
| *Vjesnik bibliotekara Hrvatske* | 1 |

*Table 1.* Journals cited.

adjacent to keywords. Taken together a small core of themes can be isolated and plotted using multi-dimensional scaling (MDS). In this case, twenty of 1865 phrases occurred four times or more and these are shown in the MDS plot in Figure 3; the three- dimensional plot shows not only the core themes but also their relative proximity based on term co-occurrence in the text corpus.

The large blocks in the lower left corner represent the conference themes of authority control and linked data and emphasize the UDC, the *Dewey Decimal Classification* (*DDC*), and Library of Congress vocabularies. This could be called the conference domain's extension. In the upper right corner is what could be described as one aspect of the intension of the conference, representing authority data and the new conceptual model "Functional Requirements for Subject Authority Data" (IFLA 2010) also known by its initials FRSAD. The other aspect of the intension we see is the linkage between faceted classification and information retrieval, which though linked are relatively distant from each other; this aspect is also distant from the FRSAD aspect. This is a fairly good visualization of the core themes represented in the titles of the papers and it is quite different from what we observed in 2013, when classification and information retrieval hovered in the vicinity of the core themes of culture, collections and visualization.

Another way to visualize the extension and intension of a conference domain is to analyze author co-citation among the conference papers. In this instance, forty-three authors were cited twice or more. After removal of corporate bodies from the list, sixteen remained who had been cited three times or more. The most cited were Aïda Slavic (11), Rebecca Green (8) and Michael Buckland (8). Author co-citation demonstrates perceived thematic relationships among work by authors who are cited together, and thus can act as a methodological triangulation when compared to co-word analysis. An internal author co-citation matrix was prepared, and IBM-SPSS™ was used to generate an MDS plot. The same matrix was used with Gephi visualization software to generate a network map; both are shown in Figure 4.

We see three clusters in this visualization and they match more or less with the segments we saw in the phrase analysis in Figure 3. The cluster in the center represents the conference extension, the UDC, linked data, the semantic web and authority control, including Library of Congress vocabularies. On the right is a cluster related to FRSAD, and on the left is a cluster related to faceted classification and the *Dewey Decimal Classification*. The intension is represented to some extent by the classical KO domain, but the cluster on the right is not quite as typical of KO. Although Riesthuis and Francu are names found often in KO proceedings, Pika and Buxton are not. Jiri Pika is one co-author (with Milena Pika-Biolzi) of a paper reporting on the use of highly synthesized UDC strings together with authority-controlled natural language subject terms to permit complex multi-

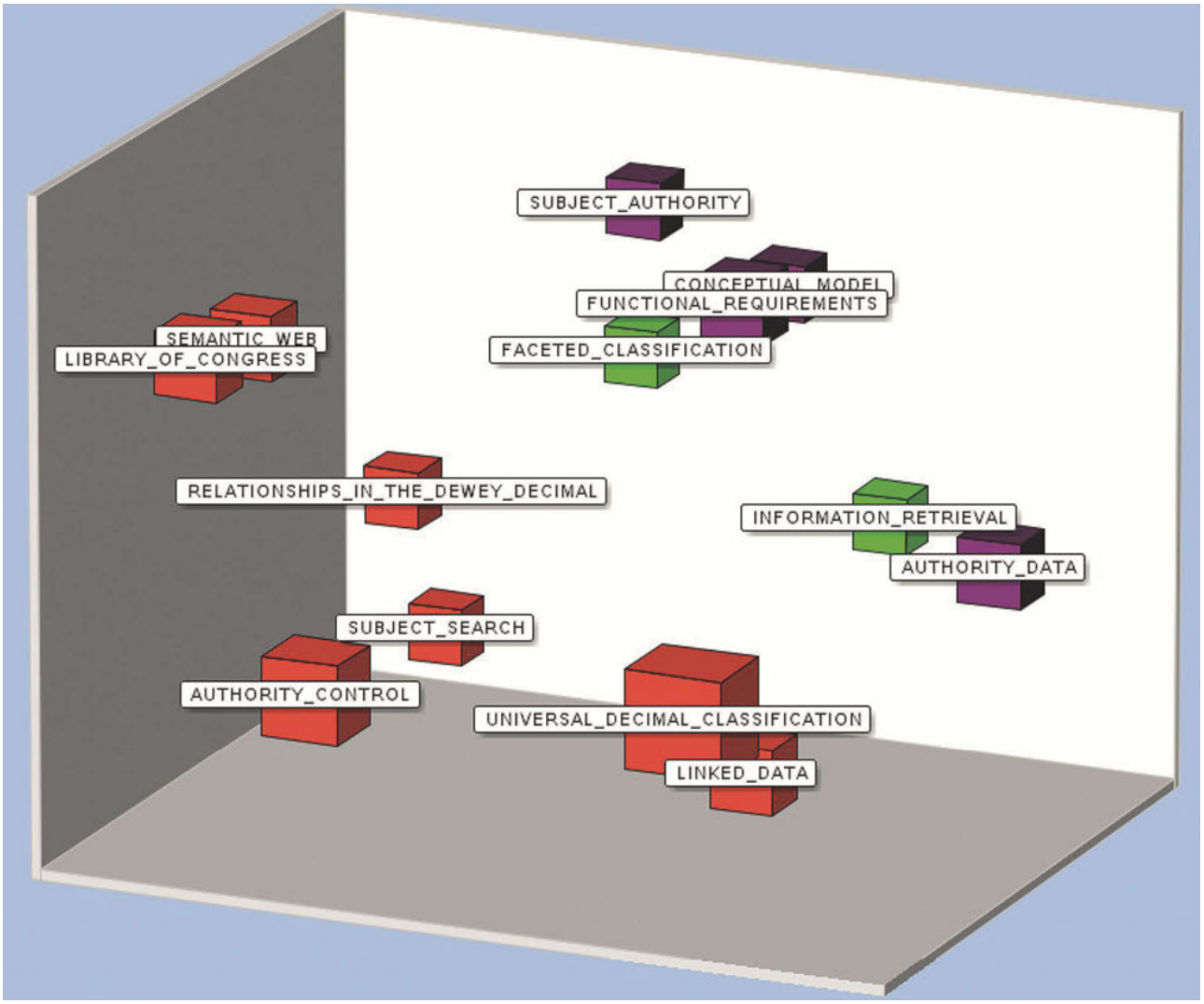



*Figure 3*. MDS plot of core themes (stress = .22 $R^2$ = .90).

lingual searching at the ETH Bibliothek at the Swiss Federal Institute of Technology in Zurich. They have cited two 1990 papers by A.B. Buxton on computer searching of UDC numbers. Although these authors are cited only a few times, they are co-cited with Francu and Slavic at about the same rate as Riesthuis. Essentially this cluster on the right represents the segment of the conference's intension that is new to KO because it is unique to this conference—authority control and classification as approaches to the integration of FRSAD. The network visualization emphasizes the coherence of the network of perceptions of the intension as well as the anchoring roles of Slavic, Green and Tillett.

Finally, because journals represent perceived scientific (or at least, arguably, authoritative) source material as a domain evolves, the perceptions of citing authors about relationships among the sources on which they rely can reveal another dimension of relativity in the core ontological positions within a domain. Journal co-citation is a means of visualizing this dimension. For this analysis we created a matrix of the co-citation of the nine journals cited more than once (see Table 1 above). An MDS plot generated using IBM-SPSS™ helps to visualize the perceptions among citing authors. The same matrix was used with Gephi visualization software to generate a network map; both are shown in Figure 5.

The MDS plot in Figure 5 provides further methodological triangulation in support of the results already revealed. That is, there is one cluster of tightly related journals anchored by the paired *Cataloging & Classification Quarterly* and *Knowledge Organization* together with the paired *Journal of Documentation* and *JLISItalian*. These are journals representing the citing authors' perceptions about both theoretical and applied research in KO and KO Systems (KOSs). The most highly cited periodical in the conference is the *Extensions & Corrections to the UDC*, which is not a research journal *per se* but is the authoritative updating mechanism used to extend the UDC and

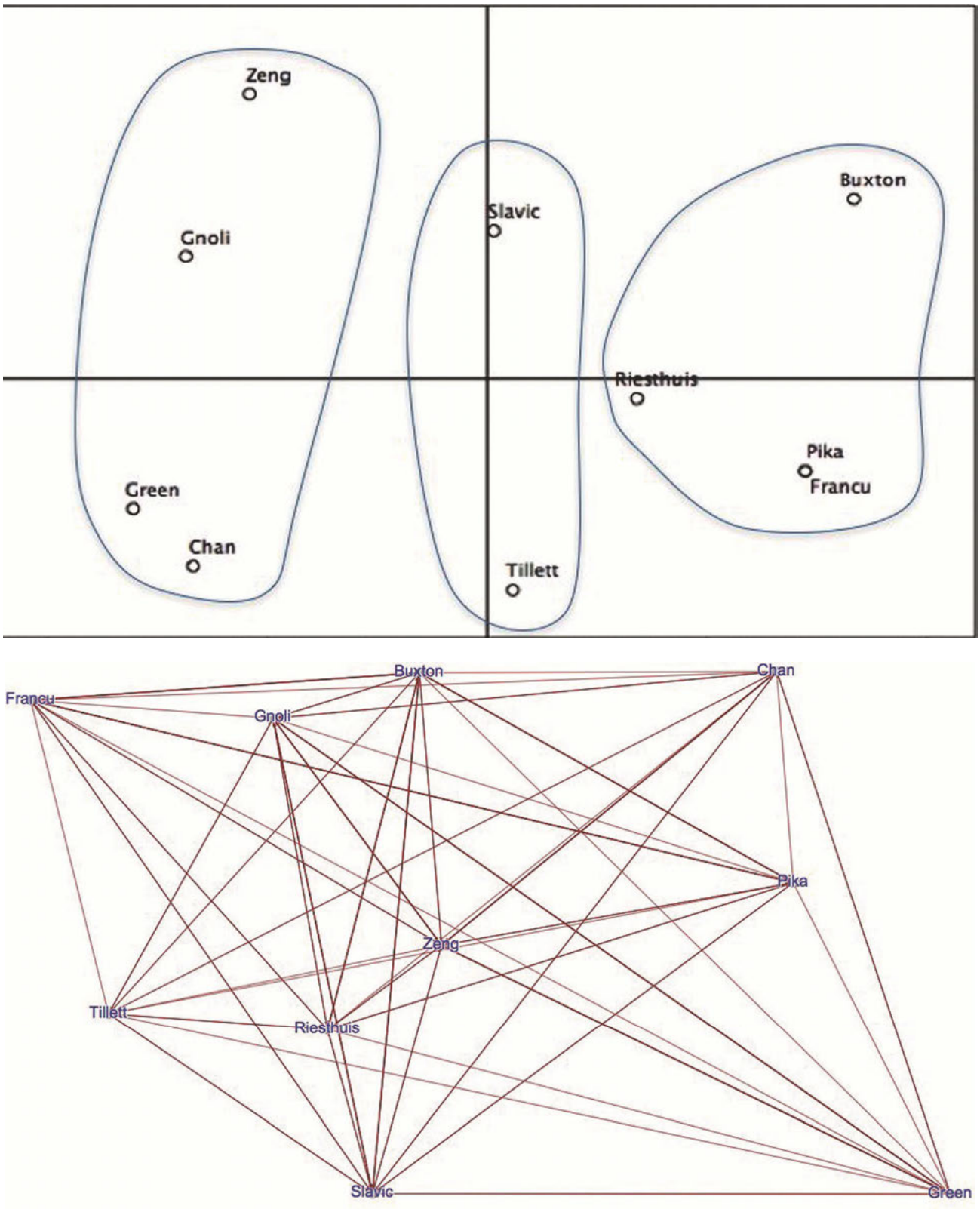
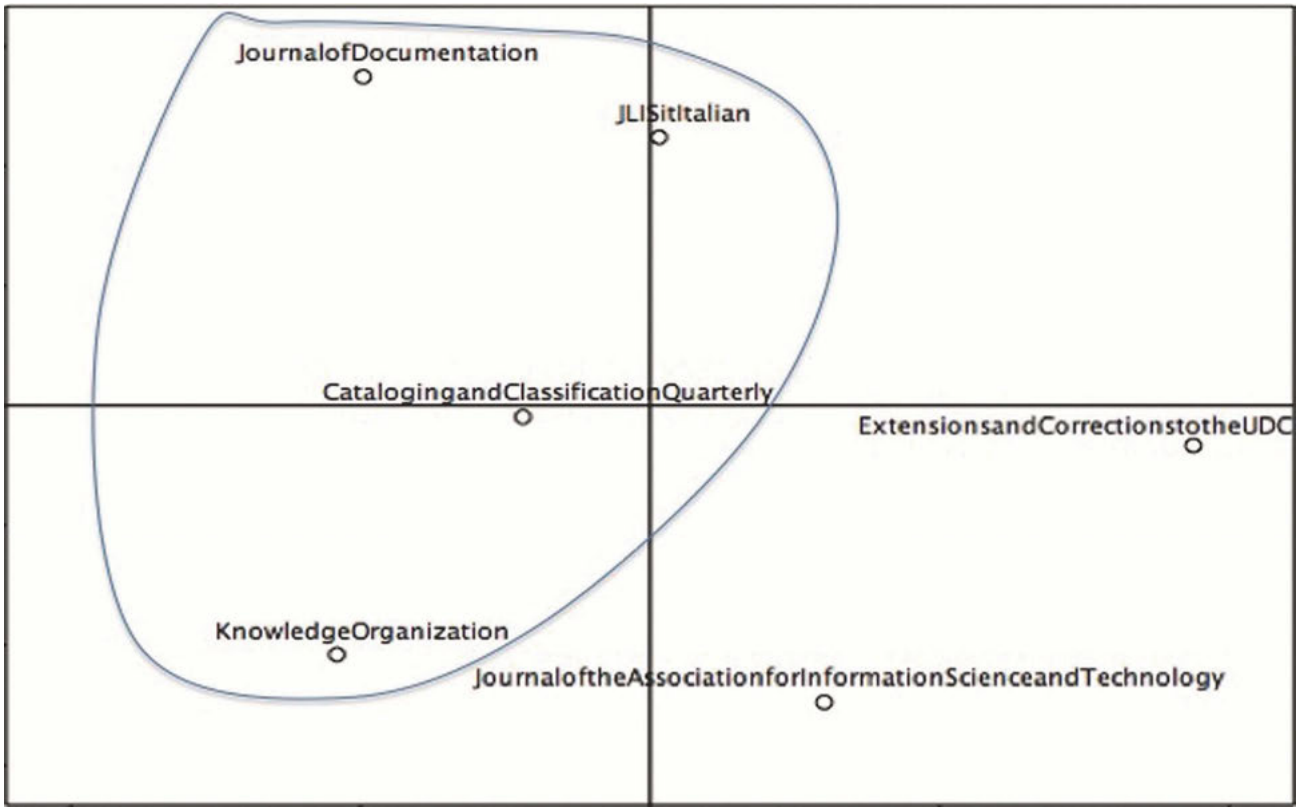



*Figure 4*. MDS plot (stress = .07 $R^2$ = .96) and network visualization of internal author co-citation.



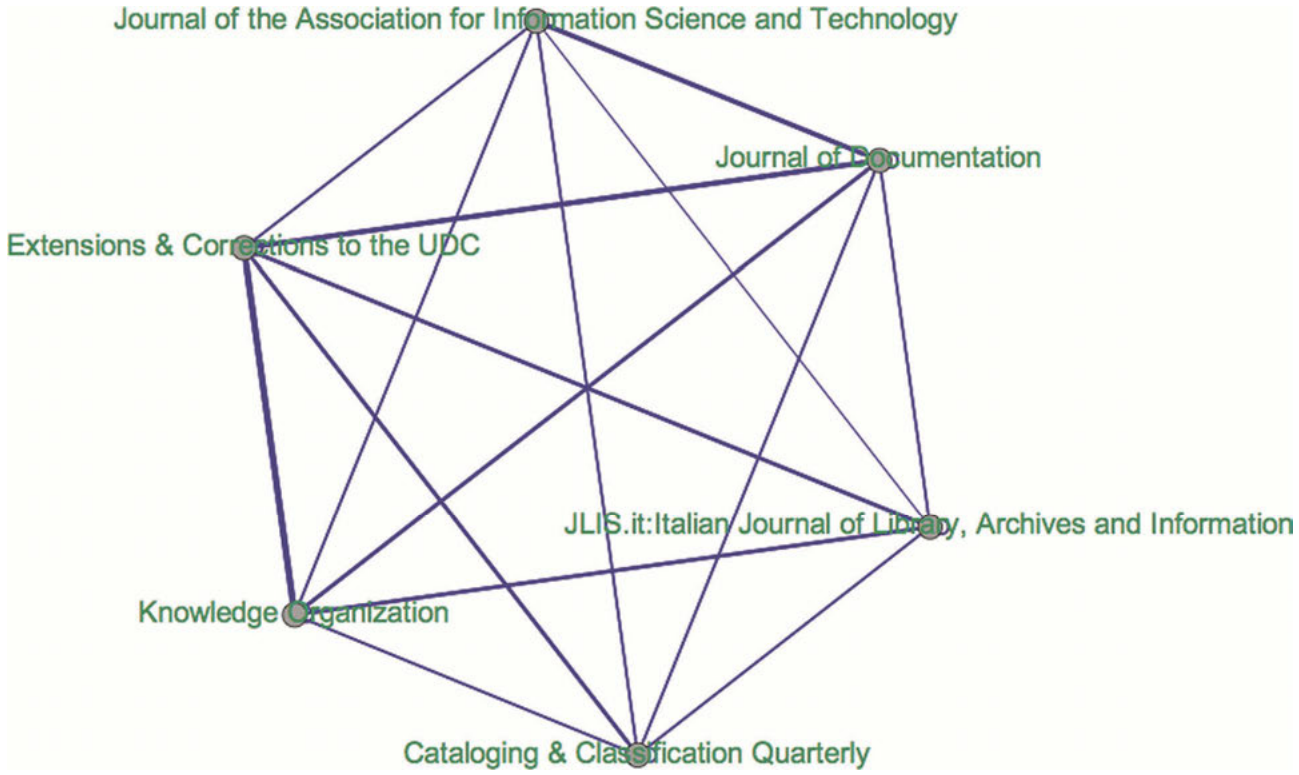




*Figure 5.* MDS plot and network visualization of internal journal co-citation (stress = .009 $R^2$ = .99).

which often reports local applied extensions of the UDC. It is plotted on the map at some distance from the *Journal of the Association for Information Science and Technology* (and its predecessors) representing, in this case, information retrieval and semantic web technologies. The latter two publications are only loosely associated with the core cluster suggesting three nodes quite similar to those observed earlier. The network visualization emphasizes the centrality of *Extensions & Corrections*, *Knowledge Organization* and *Journal of Documentation*, anchoring this conference domain's extension to the KO core.

## 2.0 Some concluding thoughts

The focus of the UDC Seminar is, of course, the evolution and growth of the use of the UDC as it expands from a bibliographic classification into a catalyst for KO on a semantic web. What we see in the phrase analysis as well as the internal author co-citation analysis is the unique contributions of this conference to the KO domain on that front. The core extension of this conference tightly reflects the conference theme of authority control, linked data and the UDC. Meanwhile, the intension is stretched by, on the one hand, a focus on the conceptual model of subject representation and retrieval contained in FRSAD, and on the other, a theoretical approach to faceted classification and information retrieval. From the brief journal co-citation analysis we also gain understanding of the key coordination in this conference between theoretical research and applied KOS development. This latter line between theory and practice is more sharply drawn than has been the case in general in the KO domain.

What is equally interesting is the absence of the traditional KO core author set represented by Hjørland, Smiraglia, López-Huertas and Tennis (Smiraglia 2015, 609). Citing authors seem to have relied on the authority of the leading contributors Aïda Slavic (Editor in Chief of the UDC), Barbara Tillett (retired chief of the Library of Congress' Cataloging Policy and Support Office and director of the library's Integrated Library System program) and Rebecca Green (Senior Editor of the *Dewey Decimal Classification*) to represent the conference theme, which emerges here as the domain extension. Around this core hover two other clusters representing FRSAD (Riesthuis, Francu) as applied to the use of authority-controlled subject vocabulary together with the UDC (Pika and Buxton) and the use of faceted classification for information retrieval (Gnoli, Zeng and Chan). To some extent, then, this conference has relied on quite specific poles within the KO domain rather than on the typical theoretical extension of the domain, which is why we encounter such a unique distribution of cited authors. In this, the UDCC has succeeded with its 2015 international seminar in constructively stretching the intension of the KO domain as it simultaneously extended the influence of the UDC for semantic web applications.

## References

Beak, Jihee, Inkyung Choi, Sukwon Lee, Hyoungjoo Park, Laura Ridenour and Richard P. Smiraglia. 2014. "Knowledge Organization and the 2013 UDC Seminar: An Editorial." *Knowledge Organization* 41: 191-4.

IFLA Working Group on the Functional Requirements for Authority Records. 2010. "Functional Requirements for Subject Authority Data (FRSAD): A Conceptual Model." Unpublished report. http://www.ifla.org/node/5849

Smiraglia, Richard P. 2013. "The Epistemological Dimension of Knowledge Organization." *IRIS Revista de Informação, Memória e Tecnologia* 2, no. 1. http://www.repositorios.ufpe.br/revistas/index.php/IRIS/article/view/498

Smiraglia, Richard P. 2015. "Domain Analysis of Domain Analysis for Knowledge Organization: Observations on an Emergent Methodological Cluster." *Knowledge Organization* 42:602-11.