Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2016) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

13th International Conference on Current Research Information Systems, CRIS2016, 9-11 June 2016, Scotland, UK

Capturing the collaboration intensity of research institutions using social network analysis

Stefan Schlattmann*

University of Muenster, Schlossplatz 2, 48149 Münster

Abstract

The relevance of collaborations among scientists is growing steadily in order to address complex research problems. On the basis of research institutions this trend of increased collaboration is being shaped through interdisciplinary research centres, institutional networks or centres of excellence. The identification and promotion of collaborative research within the own university has become a crucial part of strategic planning. Using CRIS-Data as enrichment for decision-making seems to be a practical and obvious approach since in this environment research information across all disciplines are getting consolidated explicitly. In the field of scientometrics examining research collaboration with methods taken from the social network analysis is widely accepted. Instead of analysing the macro-level of science, this paper deals with the potential benefits of using methods from the social network analysis in providing quantitative information about the intra-organizational collaboration for purposes of research management.

Keywords: quantitative measurement of CRIS Data: intra-organizational research collaboration; strategic planning; social network analysis

1. Introduction

In order to address complex research problems the relevance of global collaborations among scientists is growing steadily[1]. There are numerous reasons why collaboration is important in science. Foremost scientists seek to work with the most outstanding colleagues in their field [2]. Besides that interdisciplinary research approaches are

* Corresponding author. Tel.: 49 251 8321471
E-mail address: stefan.schlattmann@uni-muenster.de

substantial drivers of collaboration. On the basis of research institutions this trend of increased collaboration is being shaped through interdisciplinary research centres, institutional networks or centres of excellence [3]. This development can among other factors be traced to an ascending appreciation of interdisciplinary research by funding organizations [4]. In times in which third-party funds represent a remarkable percentage of a universities financial resources an in-depth knowledge of current as well as potential research focuses can be an valuable determinant to position in an increasingly evaluation-based competition with other universities. [5] Using CRIS-Data for such purposes seems to be a practical and obvious approach since in this environment research information across all disciplines are getting consolidated explicitly[6].
Foremost two strategic tasks were identified in which knowledge about the characteristics and the intensity of collaborations within a university can provide substantial information:

- Detection, analysis and visualization of (interdisciplinary) research communities as support for applications of third-party funds, focus areas or for establishing interdisciplinary research centres.
- Identification and strengthening of the university-wide research profile

Examining research collaboration with methods taken from the social network analysis is widely accepted [7]. Co-citation and co-authorship analyses are a growing sub-domain within the field of scientometrics by giving the study of science, communication in science and science policy a quantitative perspective, [8]. Differentiating from this global approach, this paper deals with the potential benefits of using methods from the social network analysis in providing quantitative information about the intra-organizational collaboration for purposes of research management and communication.
Considering this paper as introductory illustration of application scenarios for capturing the intra-organizational collaboration under usage of social network analysis, the research design as well as the findings is described on more general basis.

2. Collaboration analysis for strategic planning in research institutions

One central aspect for implementation and operation of a CRIS is to perform a university-wide harmonization and consolidation of the decentralized pools of research information [9]. The evolving centralized data pool can be used to fulfil several tasks in the field of research management. One major sub domain within the area of research management is the institutional strategic planning. In an environment where seeking of third-party funds represents a significant part of strategic planning, knowledge about the existing and potential research focuses of a university is crucial.[5] The ascending appreciation of interdisciplinary research by funding organizations demands explicit knowledge about present research communities within the institution. As illustrative example the Excellence initiative for German universities can be mentioned. Through the establishment of so called clusters of excellence an enhancement of scientific networking and cooperation was intended. Instead of relying completely on the personal perception of individual persons or groups in determining the potential topics and scientists to be organized in an excellence cluster, a data-oriented and evidence-based analysis under usage of CRIS data compromises extensive potentials to enrich such actions. Furthermore in order to establish a distinctive position in the national research landscape, universities perceive increasingly the need of characterizing their institution through the definition of explicit research focuses. While these processes are highly driven by university-political considerations, evidence-based decision-support can provide added value during the detection of research focuses as well as confirmation for existing focuses.[2] Finally not only at the institutional level insights of research collaboration can enrich strategic planning actions, but deans as well as other faculty staff involved in research management can take advantage of such exploratory approaches in developing future research orientations.

As support for decision-making in general most CRIS vendors have integrated reporting platforms into their solutions. Such solutions offer the creation of flexible and aggregated reports and visualizations. Although the level of

[2] The German Rectors' Conference established a national research map consisting all german universities including their institutional research focuses. In this setting the acceptance of a research focus requires a cooperation involving 25 professors. For more detail see: http://www.forschungslandkarte.de/profilbildende-forschung-an-universitaeten/kartensuche.html

aggregation of the existing data can be handled variably, uncovering implicit patterns is hardly possible. At this point more sophisticated approaches are needed. However, the social network analysis offers various methods to capture structural aspects of entities being connected through social relations.

3. Design

The social network analysis constitutes the methodical framework of this paper as mentioned before. The structural components of a network - nodes which are connected by edges - are in this setting the organizational units (nodes) connected with one another through one or more collaborative research projects (edges). More formalized the collaboration network is a weighted and undirected graph - the direction of an edge does not matter but capturing the number of shared project is an essential condition. (For basic information about social network analysis, see e.g. [10]. In an exemplary Application all organizational units from German university were taken regardless of the hierarchical level which ranges from the University itself to academic chairs.

As quantifiable measure for a research collaboration all third-party funded projects are considered instead of the well examined practice of taking co-authored publications. From a research-theoretic perspective this is reasoned at first sight through the more intensive and longer term co-work of research projects. Thus a stricter notion of collaboration was chosen [11]. Another possible bias which can be neglected are the quite divergent discipline-specific publication characteristics which range from research communities with a strong emphasis on monographs to communities in which conference papers are the preferred publication format [12].

A framework describing the potential dimensions of analysis was developed (see Figure 1). The basic idea is to point out the capabilities for obtaining the desired output network. This can be accomplished through variation in the selection and aggregation of data taken from the CRIS as well as the selection of an appropriate method from the social network analysis.

Having the collaboration network as output variable for all problem definitions, the relevant data for a specific research question can be chosen under usage of the dimensions *organization*, *project* and *network metrics*. In case of the dimensions project and organization, a variation in the selection of the relevant nodes and edges can be performed. Reasonable selection criteria for organizations are for instance the hierarchical level as well as the associated discipline of an organization. A particular mapping of relevant projects can be performed based on temporal aspects and project-based attributes like external funders or research areas. This dynamic mode of network construction consisting of various possible node-edge combinations offers on the one hand great flexibility in addressing different research questions. On the other hand a crucial network-theoretic issue the network boundary definition is taken into account. Since many structural aspects of interest can already be affected by the presence or absence small numbers of irrelevant relationships in key locations of the network [13].

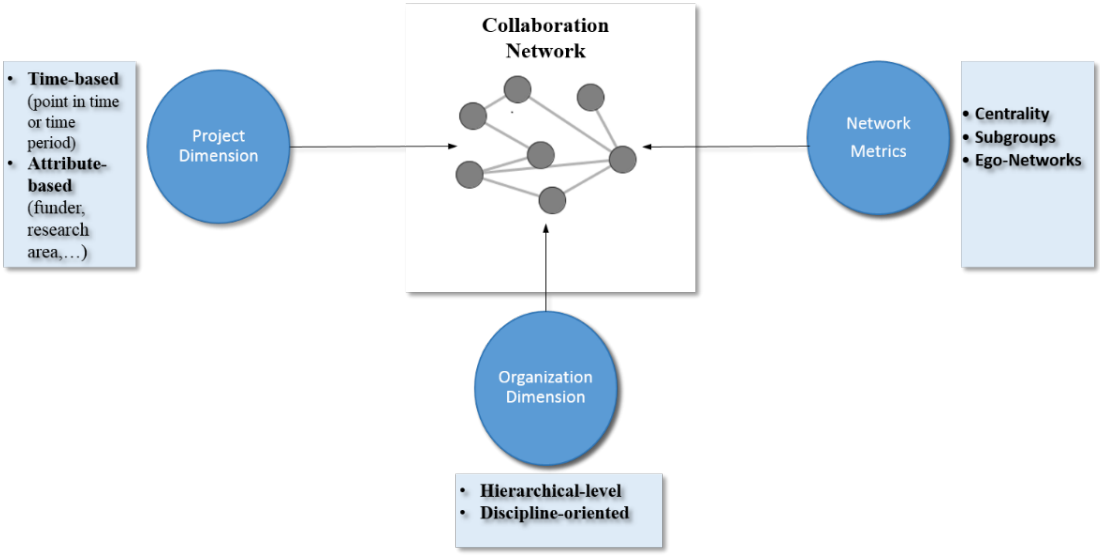

Figure 1Analytical dimensions of collaboration analysis

Depending on the particular question different methods taken from the social network analysis can be suitable. The execution of community detection algorithms can provide valuable insights for a discovery of potential research areas by identifying subgroups based on specific network properties (A commonly used algorithm is the Girvan-Newman Algorithm, see [14]. For other problem definitions which for instance have a strong emphasis on identifying the best connected scientific organizations within a university the concept of network-centrality is capable (e.g. see [15]). Commonly used centrality measures are among others the *degree*, *betweenness* or *eigenvector*. [3] Finally with one specific organization at the centre, the construction of ego networks allows the analysis on a micro-level consisting all relations to other organizations [10].

4. Findings

The data analysis framework which was introduced in the research design section offers plenty of possible analysis scenarios.

As exemplary case a community detection algorithm from [14][4] was executed. The possible value of this method is to identify coherent research communities. Applied to the collaboration network of the considered university the algorithm produced 38 communities, where the predominant number of communities consists only of few members. However there are some greater communities. An evaluation regarding the informative value as well as the representation of actual circumstances was conducted through a mapping of the produced communities with known collaborative hubs. The so called clusters of excellence were classified as known greater research

[3] An in-depth overview of centrality measures can be found at: [16].

[4] Entitled as *edge betweenness community* the used algorithm is well-implemented in several statistical programming languages.

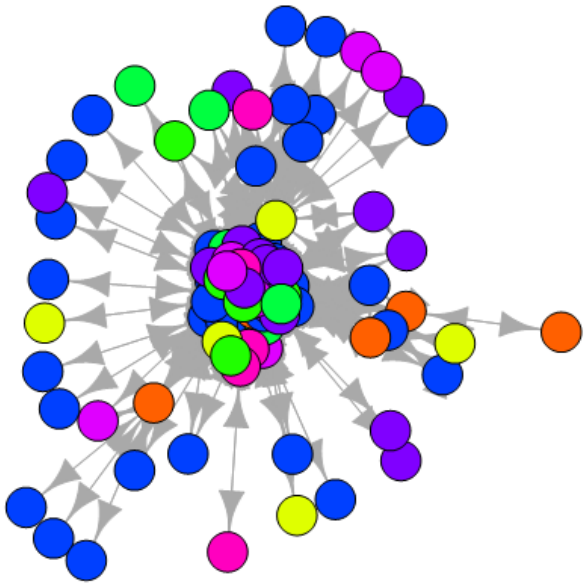

Figure 2 Sub-network based on Girvan-Newman algorithm

communities.[5] Having those actual structures in mind, the community detection algorithm produced sub-groups which show a large accordance. In Figure 2 a sub graph consisting nearly all involved organizational units of one of the university's clusters of excellence is visualized. Especially at the centre of the network a really dense sub-structure is noticeable from which an intensive collaboration can be concluded. The visualization of the dense network core was emphasized using a layout algorithm which arranges adjacent nodes close together.
Thus only by analysing the structure of the constructed network patterns of collaboration can be detected. Furthermore a first impression of the interdisciplinary nature of the observed community can be observed since the faculty affiliation of each node is highlighted in a specific colour. In order to pick up one of the earlier discussed use cases – identification of existing (interdisciplinary) research communities - again, the community detection algorithm could function as a starting point of analysis. The resulting communities could get further examined under usage of appropriate centrality measures. Possible criteria of distinction of resulting communities are the frequency of collaboration (number of collaborative projects) or the overall degree of connectivity (size of cohesive subgroup).

Besides the presented detection of research communities the methodical spectrum of the social network analysis can be used to pursue several other questions regarding the collaboration at the institution. Considering the analytical dimensions (see Figure 1) variation in the research design can be performed in many ways. For instance the collaboration relevance – does intra-organizational collaboration matter for scientists at all - can be analyzed on the whole institution or for a more dedicated view on the basis of certain faculties. Another additional perspective could be a comparison of the collaborative and interdisciplinary nature with regard to the most important funders of third-party projects.

[5] Clusters of excellence are part of the excellence initiative which is a large-scale promotional program initiated from the Federal Ministry of Education and Research and the German Research Foundation

5. Conclusion and Outlook

In this study based on research information taken from an institutional CRIS a collaboration network was constructed. Furthermore a framework consisting analytical dimensions for collaboration analyses was introduced. In this framework perspectives, aggregations and network metrics for possible problem definitions are arranged. To outline the informative value of such approaches a widely accepted community detection algorithm was executed. As a result known patterns of university-wide collaboration were identified only by analysing the relation between the involved actors.

The analytical dimensions are a good baseline for a systematization of limitations and resulting future research. Regarding the dimensions of projects and organizations, the data structure and the level of detail in a CRIS has to be evaluated more detailed. A potential area of improvement is the accountability of organizational units and if applicable of projects to its specific research discipline. Especially for capturing the interdisciplinary nature of collaborative research a well-defined classification of fields of research is needed.[6] The application of network metrics needs to be amplified in many ways. This could lead to a further examination towards more precise problem definitions including an explicit selection of appropriate network metrics and further methods of interest.
Even on a more general level basic components of the network construction could get further questioned. At first place the intention of this paper is to evaluate the possible uses of social network analysis in highlighting collaboration pattern within a university based on data taken from a CRIS. The selection of common third-party funded research projects in pursuing this problem definition was undertaken among other factors due to university specific characteristics which to a certain extent have to be reconsidered within a more generic setting.

References

[1] C. Wagner, L. Staheli, R. Silberglitt, and A. Wong, “Linking effectively: Learning lessons from successful collaboration in science and technology,” *Rand*, no. April, 2002.
[2] Royal Society, *Knowledge , networks and nations Global scientific collaboration in the 21st century*, vol. 03/11, no. RS Policy document 03/11. 2011, p. 114.
[3] T. Heinze, P. Shapira, J. D. Rogers, and J. M. Senker, “Organizational and institutional influences on creativity in scientific research,” *Res. Policy*, vol. 38, no. 4, pp. 610–623, 2009.
[4] U. Schimank, “‘New public management’ and the academic profession: Reflections on the German situation,” *Minerva*, vol. 43, no. 4, pp. 361–376, 2005.
[5] J. Gläser, S. Lange, G. Laudel, and U. Schimank, “Evaluationsbasierte Forschungsfinanzierung und ihre Folgen,” in *Wissensproduktion und Wissenstransfer*, U. Mayntz, Renate;Neidhardt, Friedhelm; Weingart, Peter; Wengenroth, Ed. transcript, 2008, pp. 145–169.
[6] EuroCRIS, “Why does one need a CRIS?,” *The Research Process and how a CRIS can support it.*, 2016. [Online]. Available: http://www.eurocris.org/why-does-one-need-cris.
[7] H. Hou, H. Kretschmer, and Z. Liu, “The structure of scientific collaboration networks in Scientometrics,” *Scientometrics*, vol. 75, no. 2, pp. 189–202, 2008.
[8] L. Leydesdorff and S. Milojevic, “Scientometrics,” *CoRR*, vol. abs/1208.4566, 2012.
[9] S. Herwig and T. Höllrigl, “‘ All roads lead to Rome ’: Establishing Best Practices for the Implementation and Introduction of a CRIS : Insights and Experiences from a CRIS Project at the University of Münster Research Information Systems as an Instrument for the Dissemination of Re,” *E-Infrastructures Res. Innov. Link. Inf. Syst. to Improv. Sci. Knowl. Prod. Proc. 11th Int. Conf. Curr. Res. Inf. Syst.*, pp. 93–102, 2012.
[10] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.
[11] J. S. Katz and B. R. Martin, “What is research collaboration?,” *Res. Policy*, vol. 26, no. 1, pp. 1–18, 1997.
[12] S. Hornbostel, B. Klingsporn, and B. von Ins, “Messung von Forschungsleistungen – eine Vermessenheit?,” *Publ. unterschiedlichen wissenschaftlichen Disziplinen Beiträge zur Beurteilung von Forschungsleistungen*, vol. 49, no. 0, pp. 14–21, 2009.
[13] C. T. Butts, “Social network analysis: A methodological introduction,” *Asian J. Soc. Psychol.*, vol. 11, no. 1, pp. 13–41, 2008.
[14] M. Girvan, M. Girvan, M. E. J. Newman, and M. E. J. Newman, “Community structure in social and biological networks,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 12, pp. 7821–7826, 2002.
[15] M. E. J. Newman, “Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality,” *Phys. Rev. E*, vol. 64, no. 1, p. 016132, 2001.
[16] U. Brandes and T. Erlebach, *Network Analysis: Methodological Foundations (Lecture Notes in Computer Science)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.

[6] The classification of fields of research is a general issue in research management. Applied to the German environment, a set of recommendations was compiled in [17]

[17] Institut für Forschungsinformation und Qualitätssicherung (ifQ); Fraunhofer-Institut für Angewandte Informationstechnik, “Interdisziplinarität im Forschungsberichtswesen Empfehlungen zur Abbildung,” 2015.