Flanders Research Information Space as a tool to monitor interdisciplinary research in Flanders

Bram Vancraeynest, Hoang-Son Pham, Hanne Poelmans, Sadia Vancauwenbergh
ECOOM-UHasselt, DSI-UHasselt Belgium
{Bram.Vancraeynest, Hoangson.Pham, Hanne.Poelmans, Sadia.Vancauwenbergh}@uhasselt.be

I. FRIS research portal

Flanders Research Information Space (FRIS)[1] is a current research information system (CRIS) governed by the Department Economy, Sciences and Innovation (EWI) of the Flemish government. The FRIS research portal discloses information on (partially) publicly funded research (e.g. researchers, research institutions, projects and publications) from over 40 data providers (research performing and research funding organisations) in Flanders.

The main goal of FRIS is to accurately reflect the state of research in the Flanders at any given time. To obtain this, all FRIS data providers make use of the CERIF-XML (Common European Research Information Format)[2] data format to push information from their institutional CRIS system to the FRIS platform and incrementally, in real-time, propagate data changes to FRIS.

The FRIS platform is used by the Flemish government for reports, analysis and statistics in the context of policy making and for monitoring trends in research and innovation. In addition, it is designed to unite researchers, encourage interdisciplinary research, stimulate networking among researchers and enable the identification of experts in certain disciplines.

II. FRIS as tool for research monitoring

This study investigates how an interorganisational CRIS, such as the FRIS portal, can be used to monitor trends and evolutions in the research landscape of a certain region. Moreover, as stimulating and measuring interdisciplinarity in research is one of the points of interest of the Flemish government, we aimed to develop an approach to measure interdisciplinarity in research projects based on the information available on the FRIS portal.

The FRIS portal is a rich source of information to monitor evolutions in research and innovation in Flanders. It not only contains constantly up-to-date information on a variety of research objects (i.e. researchers, research institutions, projects and publications) and from a wide range of stakeholders in the research landscape, it also labels (by means of semantically enriched classification systems) every research object with descriptive information on the scientific discipline ("Vlaamse Onderzoeksdiscipline Standaard", abbreviated VODS, in Dutch [1]) or funding source (financial source classification) it relates to.

III. Identifying interdisciplinary research projects in Flanders

Interdisciplinary research (IDR) is a mode of research that integrates information, data, techniques, tools, perspectives, concepts, and theories from two or more scientific disciplines [2]. IDR is essential to deal with boundary-spanning problems and to encourage the development of emerging research fields. IDR has recently become a hot topic in universities and research funding organisations, at (sub)national, European, and international levels [3], [4].

From a conceptual viewpoint, there are two main approaches to study interdisciplinarity:

- organisational approach - on the basis of the authors' and researchers' professional skills and/or affiliation;
- cognitive approach - on the basis of information flows between disciplines and similarity of textual content.

Organisational approach

The organisational approach assumes that a diverse team of researchers or a diverse team of research organisations that collaborate on a project indicates a high probability of IDR. Here we evaluate the interdisciplinarity of a project based on how diverse (in terms of disciplines) the researchers or organisations that participate in the project are.

To calculate the diversity of researchers (similar for calculating diversity of organisations), we define a researcher as a subset of disciplines of the total set of the disciplines. Noticeably, in this study, the disciplines of a researcher are not only based on the information declared in the researcher entity, but also inferred from different sources such as the disciplines of the organisations where they worked, the disciplines of the projects that they worked on, the disciplines of the publications that they published, the disciplines of the co-authors of the projects and the disciplines of the co-authors of the publications. By combining information from different resources, the disciplines of researchers can be determined more precisely.

To calculate the diversity of researchers, we apply the Rao-Stirling diversity index [5] which is calculated as follows:

$$DR(p) = \sum_{i,j=1}^{n} d_M(i,j) f_i f_j, \tag{1}$$

[1] https://researchportal.be/en
[2] https://eurocris.org/eurocris_archive/cerifsupport.org/category/cerif/cerif-xml/index.html

TABLE I
EXAMPLE OF PROBABILITY DISTRIBUTIONS OF DISCIPLINES OVER PROJECTS

	dis_1	dis_2	dis_3	...	dis_k
pro_1	0.9	0.1	0.0	...	...
pro_2	0.5	0.5	0.0	...	...
pro_3	0.1	0.4	0.5	...	...
...					
pro_n	0.6	0.1	0.3	...	...

where $d_M(i, j)$ is the distance between vectors represented researcher i and j; f_i and f_j are frequencies of researcher i and j, respectively.

The distance between researchers are based on the distances of their respective disciplines. For instance, two researchers having similar disciplines are considered as less different (i.e. small distance) than researchers who have different disciplines (i.e. large distance).

Cognitive approach

The organisational approach is promising; however, it is not enough to fully capture IDR. For example, a diverse team of researchers does not guarantee that the project is interdisciplinary. To more efficiently measure IDR in projects, we need to incorporate the content of the project in the analysis. That is, the terms or keywords embedded in the content of the project can give a rich indication on the topics of the project. For example, if the used keywords in a project relate to different disciplines, the project will probably incorporate multiple disciplines and can potentially be interdisciplinary.

To discover the topics of a project, we used supervised topic modelling [6]. Supervised topic modelling works as follows: Given a set of documents (i.e., a set of projects' content) and a set of K topics (i.e., disciplines), the topic model can learn and discover the probability of the distribution of topics in a set of documents. Table I shows an example of output of supervised topic modeling.

Based on the output of the topic model and the probability distribution of topics over a document, we can apply the Rao-Striling diversity index to calculate the diversity of topics related to each document. This diversity score is used to measure the interdisciplinarity of project. A higher diversity score indicates a higher possibility of IDR.

Organisational and cognitive framework

To efficiently measure IDR in projects, we combine the organisational and the cognitive approach together. In particular, we combine the diversity of researchers (DR) and the diversity of topics (DT) to measure IDR of projects. In this framework, a high diversity of researchers and a high diversity of topics indicate potential IDR.

We evaluate the proposed framework on 2283 projects available on the FRIS portal. Figure 1 shows the result of this approach. The projects with a high diversity of researchers (DR) and a high diversity of topics (DT) (blue dots in red rectangle) can be considered as potentially IDR.

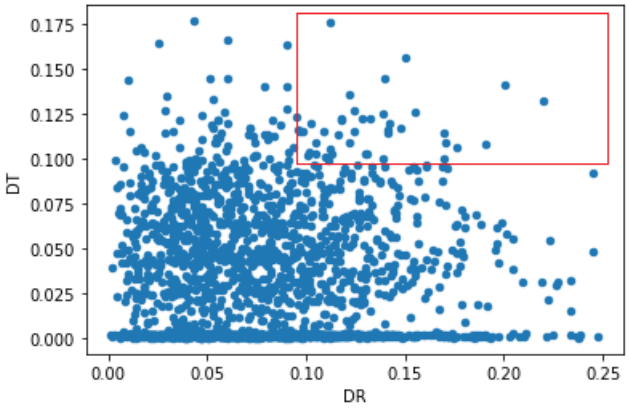

Fig. 1. Diversity of researchers (DR) and diversity of topics (DT) of 2283 projects

IV. CONCLUSION

In this study we used the data of an interorganisational CRIS to identify interdisciplinary research projects in Flanders. Because of the amount and the richness of the data available on the FRIS portal, we were able to combine two approaches to approximate interdisciplinarity, i.e. an organisational approach, based on the degree of diversity of the team of researchers involved in a project, and a cognitive approach, based on the diversity of topics used in the content of the project. With these two approaches we were able to identify a set of research projects that are potentially interdisciplinary in nature.

With the proposed mechanism, we can monitor how interdisciplinary research is situated in Flanders and how it evolves over time. Additionally, it can be further developed into a visualisation tool for the FRIS portal to visually mark those projects that are identified as IDR. Despite the fact that this study only focused on the FRIS portal, the proposed methods can easily be applied to other (inter)institutional CRIS systems.

REFERENCES

[1] S. Vancauwenbergh and H. Poelmans, "The creation of the flemish research discipline list, an important step forward in harmonising research information," *Procedia Computer Science*, vol. 146, pp. 265–278, 2019.

[2] National-Science-Foundation, "What is interdisciplinary research?" 2005, https://www.nsf.gov/od/oia/additional_resources/interdisciplinary_research/definition.jsp.

[3] D. Wernli and F. Darbellay, "Interdisciplinarity and the 21st century research-intensive university," 2016, source at www.leru.org.

[4] J. Allmendinger, "Quests for interdisciplinarity: a challenge for the era and horizon 2020." 2015, https://ec.europa.eu/.

[5] A. Stirling, "A general framework for analysing diversity in science, technology and society," *Journal of The Royal Society Interface*, vol. 4, no. 15, pp. 707–719, 2007.

[6] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning, "Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*. Singapore: Association for Computational Linguistics, Aug. 2009, pp. 248–256. [Online]. Available: https://aclanthology.org/D09-1026