

Discovering Research Collaboration Networks from Scientific Digital Libraries and Repositories

Mieczysław Muraszkiewicz
Henryk Rybiński
Piotr Szczepański

Institute of Computer Science, Warsaw University of Technology

Abstract

Purpose: The purpose of the study is to outline a practical model for discovering research collaboration networks on the basis of data and information stored in scientific digital libraries and repositories. The discovered relationships between researchers, projects, scientific institutions and other scientific entities are used for identifying collaboration networks of researchers and research institutions interested in or working on a given subject. Afterwards, such networks can be subject to various types of network analysis in order to get in-depth knowledge on the networks and their components.

Approach/methods: The method adopted in the study is twofold, that is: (i) it takes into consideration the way of discovering collaboration networks by means of simple tools that have been implemented within the Ω^{VR} system developed at Warsaw University of Technology; (ii) it develops an outline of a formal model of research collaboration networks that takes into account the specificity of scientific digital libraries and repositories and includes the network analysis techniques for discovering knowledge residing/hidden in the networks.

Results and conclusions: The outcome of the research is the outline of a formal model of research collaboration networks that includes: (i) a discovery mechanism for identifying thematically related scientists, projects, research institutions, and other scientific entities; and (ii) a set of network analysis methods for getting in-depth knowledge residing in the networks. The model is implementable and scalable in terms of functionality it offers and the network analysis techniques it includes. The model is founded on a solid ground, which is the Ω^{VR} system functionality to discover simple collaboration networks, and it is being used for enhancing the Ω^{VR} system.

Originality/value: The value of the research is the outline of a general research collaboration networks model that: (i) can help identify, build, and analyse research communities, and thereby increases the scope, value and impact of scientific endeavours on science and society; (ii) is used for enhancing the Ω^{VR} system.

Keywords

Research collaboration network. Model of research collaboration network. Network analysis. Knowledge discovery. Ω^{VR} system.

Received: 24 October 2015. Reviewed: 10 November 2015. Accepted: 12 November 2015.

1. Introduction

The relentless progress within the Internet universe including its mobile extension has not only been changing our perception of the world that surrounds us but also deeply affects the way we collaborate with human and institutional peers. Until recently we tended to look at the world through the systemic lens for the major notion to depict complex structures was the concept of *system*. This has changed and now the tool of choice to comprehend and explain what we are faced with and what challenges us, and/or what we are meant to do is the notion of *network*. More on this significant game-changing alterations and a transformative shift caused by networking is said in (Muraszkiewicz, 2004; 2013). Here let us note that retackling well-known issues by means of the networking conceptual apparatus, which have originally been approached and described by systemic methodologies, can bring innovative answers and resolutions. It seems that also in the area of information science that is still perceived by many of its researchers, scholars and practitioner mainly through the systemic lens a broader application of the networking approach, especially towards establishing various research and development, and business ecosystems could indubitably enhance the horizon of information science and thereby make it more relevant to and useful for scientists, scholars, and other beneficiaries of its actual and prospective offers.

The studies of real-life networks have also brought attention to the realm of scientists working in many different specialised fields that are important for the development of science itself and also for the development and growth of economy and social settings. Biologists are investigating the proteins interaction networks in order to find hidden patterns describing human organisms (Bork et al., 2003), sociologists are studying human interactions networks in hope to predict the development, needs and behaviours of modern societies and communities (Wasserman & Faust, 1994), computer scientists are analysing World Wide Web to optimise its mechanisms, make it robust against unsolicited attacks, and humanise its interfaces (Albert et al., 2000), economists are taking bank networks and financial systems under scrutiny in order to prevent future crises (Freixas et al., 2000), and so on, and so forth. This is why the slogan: “think network” is not just a rhetoric expression; it is a methodological and pragmatic beacon for organising activities of contemporary communities, in particular, the communities of scientists.

Generically a network consists of nodes (vertices) and connections (links) between them, along with a set of operations that can be executed on the nodes and links. Graphs are convenient and widely used tools to depict and model networks. Both nodes and links are important carriers of data and explicit or hidden information. Smart techniques for analysing data and information sitting in the networks allow us to discover concealed patterns and knowledge that reside in nodes and links. Obviously the greater the network and the more operations it performs the more data, information and eventually knowledge it contains and the greater is the utilitarian value for its owners and/or users.

Three hundred years ago it was possible that a single man could work on almost all disciplines of a contemporary world. Since the time of Leibniz, who was called “The Last Man Who Knew Everything” (Hockney, 2012), the way we define and practice science changed considerably. The research of a modern scientist is usually limited to a very specialised discipline. The number of such specialised fields is growing rapidly, which makes that people from different branches of science can hardly collaborate. Additionally, each year the

number of published scientific papers, books and reports is rapidly growing, which causes that even by means of advanced search engines it is hard to find out relevant worth-reading articles. The solution to this problem may lay in a smart combination of: digitalised data, researchers' collaboration networks and advanced network analysis and retrieval tools.

In this paper we focus our attention on one particular kind of networks, namely collaboration networks linking up scientists, scientific endeavours, and other scientific entities. These are the networks in which scientists or scientific institutions are nodes, and their jointed publications, research projects and/or research artefacts define the links between nodes. The strength of such connections depends on the number of the jointed items. We argue that the growing digitalised libraries and repositories containing myriads of scientific papers, books, reports, computer programmes, sets of experiment data, and other material of scientific and academic value together with advanced networks analysis tools could be employed for creating a facility that would allow researchers and science managers to identify and/or establish research collaboration networks. Such networks could, on the one hand, increase the overall efficiency of scientific research, and on the other hand, mitigate the negative effects of the phenomenon of excessive science specialisation (Radosevic & Yoruk, 2014) that narrows down scientific research and limits its horizons. The paper presents an outline of a formal model of research collaboration networks, and then it mentions some assorted network analysis techniques and advises on how they could support the process of discovering actual and potential collaboration networks (Chapter 3). The work on developing the model has been inspired by two facts, which are: (i) the Ω^{WR} system (Muraskiewicz et al., 2014) that is an advanced digital academic repository developed and implemented at Warsaw University of Technology (Chapter 2); and (ii) the research carried out at the Institute of Computer Science of Warsaw University of Technology on discovering collaboration links in scientific digital libraries and repositories (Chapter 3).

2. Ω^{WR} – A Testbed for Developing Research Networks Discovery Mechanism

As part of the research grant *Establishment of a general, open hosting and communications repository platform for network knowledge resources for science, education and open knowledge society*¹, SP/I/1/77065/10, funded by the Polish National Centre of Research and Development, which aimed at initiating the process of setting up a forward-looking country-wide scientific information infrastructure offering rich functionality and best available technical facilities, a team of the Institute of Computer Science (Faculty of Electronics and Information Technology, Warsaw University of Technology) designed and implemented an advanced scientific information and knowledge management system dubbed Ω^{WR} . Important part of the system is an academic repository whose scope reaches beyond the functionality of classic repositories since in addition to standard repository functionality

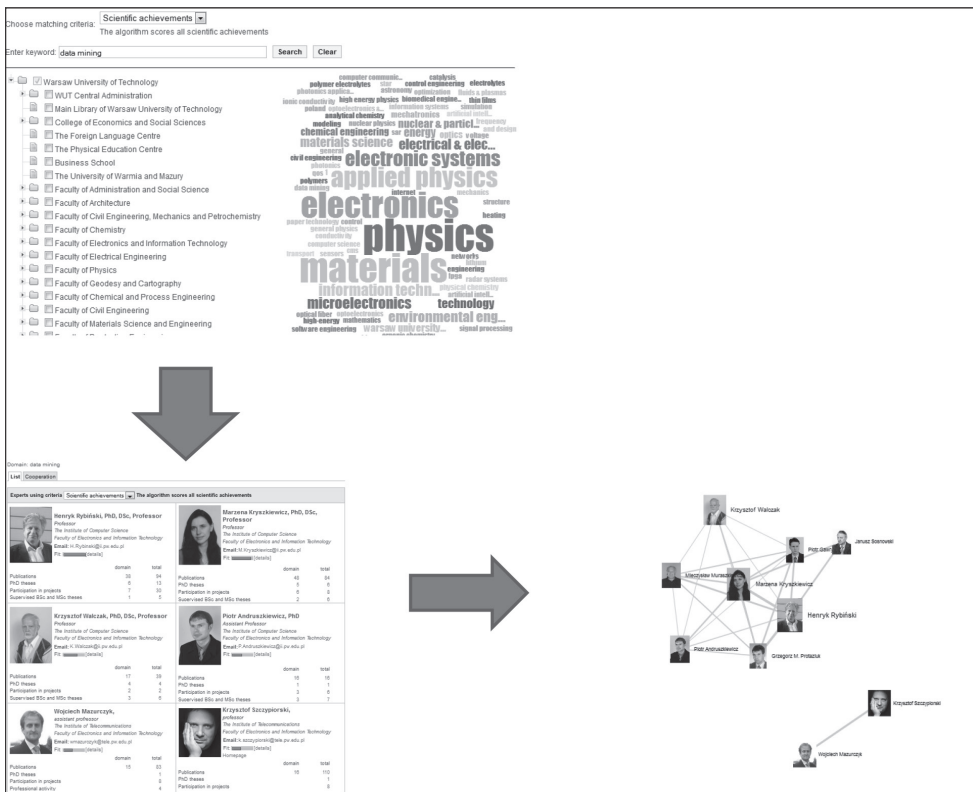
¹ *Utworzenie uniwersalnej, otwartej, repozytoryjnej platformy hostingowej i komunikacyjnej dla sieciowych zasobów wiedzy dla nauki, edukacji i otwartego społeczeństwa wiedzy*, SP/I/1/77065/10, the grant of the NCBiR; duration: 16 August 2010 through 24 June 2014, executed by a consortium of 16 leading Polish scientific institutions and libraries.

it: (i) acquires and stores not only local publications and documents (papers, textbooks, Ph.D. theses, M.Sc. theses, project reports, conference presentations, etc.), but also information on institutions, projects, people, patents, conferences, etc.; and (ii) provides a rich set of research-centric analytical tools, especially useful for managing research tasks by project leaders and for generating various statistics and reports necessary for the university administration and ministerial authorities, and also of value for researchers, students and external users (Muraszkiewicz et al., 2014).

What is important from this paper point of view is that the Ω^{VR} system focuses in particular on the presentation and promotion of researchers as well as on university units, and informal research teams. To this end, advanced algorithms have been devised and implemented for:

- tagging the expertise of a given researcher and visualizing it by a cloud of tags;
- discovering experts in a given domain, based on the research achievements registered in the knowledge base;
- finding the networks of cooperating researchers.

Figure 1. Cloud of tags and experts search



A sequence of steps for discovering a community of researchers employing these algorithms is shown in Fig. 1. The system Ω^{VR} is a pilot tool for identifying research communities and learning about their activities. Now we have started developing a model for discovering

collaboration networks, in particular within the Ω^{VR} system, along with recommendations on the methods for getting it implemented, which work we outline in the following Chapter.

3. An Outline of the Model for Discovering Research Collaboration Networks

The natural universe where we can pursue the process of identifying/discovering research collaboration networks is the realm of scientific digital libraries and academic repositories. A digital library or repository, generally speaking, is the collection of literature and multimedia objects, possibly of different kinds and formats, and their descriptions along with metadata, all of them stored electronically; the collection is endowed with a set of direct interaction interfaces, e.g. search engines, and a toolset of application programme interfaces (API) allowing users to define and execute tasks on the stored assets. Nowadays, there exist a number of valuable scientific libraries storing a large amount of information; for instance the ACM Digital Library on computer science, <http://librarians.acm.org/digital-library>, includes:

- 407,367 full-text articles;
- 2.0+ million pages of full-text articles;
- 18,000+ new full-text articles added each year;
- 44+ high impact journals with 2–3 new journals being launched each year;
- 275+ conference proceedings titles added each year;
- 2,000+ proceedings volumes;
- 8 magazines (including the flagship Communications of the ACM, the most heavily cited publication in the field of computing according to Thomson-Reuters);
- 37 technical newsletters from ACM's Special Interest Groups (SIGs);
- 6,500+ video files;
- 594 audio files.

Given a historical perspective it should be noted that the issue of discovering networks in scientific digital libraries was pursued by M. E. J. Newman who in his pioneering work extracted a collaboration network of high-energy theory physics (Newman, 2001) by using data drawn from a number of databases, including the Los Alamos e-Print Archive (physics) and NCSTRL (computer science); the network contained 8,361 scientists and 5,751 relations among them. A similar job was performed by J. Leskovec, J. Kleinberg and C. Faloutsos who extracted from the arXiv digital library a network containing 18,772 astrophysics and 19,8110 connections between them (Leskovec et al., 2007). Noteworthy, none of the aforementioned exercises offered analytical tools for further analysis of the discovered networks.

Let's now outline our model of research collaboration networks. We start with the definition of a weighted multilayer network. Mathematically speaking, the *weighted 2-layer network* is a tuple $N = (V, E_1, E_2, w_1, w_2)$ where V is the set of nodes, E_i – set of edges, and $w_i : E_i \rightarrow \mathbb{R}$ is a weighted function that assigns to each edge some real number. Each edge is a pair of nodes: $e = (v_1, v_2); v_1, v_2 \in V$. The intuition standing behind multilayer networks is that these mathematical entities represent networks with different types of connections. In this paper these types represent two ways of how individual researchers or scientific

institutions can collaborate: (i) through joint publications; and (ii) through joint projects. Now, we can define more formally the basic entities used to build collaboration networks:

- *Author* is an individual researcher defined in our model as a ordered pair: $A = (Name, I)$, where *Name* is a combination of first, second and middle name of the author, and *I* stands for *Institution* and is his affiliation.
- *Institution* can be associated with the university, with a department, or some other scientific institution. More formally, we will define it as an triple: $= (Name, City, Country)$.
- *Project* will be the $(n+1)$ -tuple: $PR = (Number, I_1, I_2, \dots, I_n)$, where *Number* is an unique id, and I_k an institution, where the project is being developed.
- *Publication* is the most important entity, which delivers the information about collaboration between researchers. It is also the central notion of the most of repositories. We will define it as a $(n+n+3)$ -tuple: $P = (Title, Venue, Year, A_1, A_2, \dots, A_n, PR_1, \dots, PR_n)$, where *Venu* stands for the title of the journal, or the name of a conference proceedings, A_k is an author, and PR_k is a project cofounding work done by A_k on this article.

The above definitions contain the essential information about basic entities only. They can be easily extended and thus making the definitions and the model itself scalable, but for the sake of clarity we will keep the minimum necessary to build fully informative collaboration network. Now, let us focus on the basic retrieval procedures. The basic data mining operations on each publication are as follow:

- Function *auth*: $\mathcal{P} \rightarrow 2^{\mathcal{A}}$ is a function from the set of publications \mathcal{P} to the power set of authors.
- Function *inst*: $\mathcal{A} \rightarrow \mathcal{I}$ applied to author retrieves its affiliation, and with a little abuse of notation will write for some $P \in \mathcal{P}$: $inst: (P) = inst \circ auth(P)$ as a function $\rightarrow 2^{\mathcal{I}}$, where \circ stands for the composition of two functions.
- Function *pro*: $\mathcal{P} \rightarrow 2^{\mathcal{P}\mathcal{R}}$ is a function from the set of publications to the power set of projects. This function is overloaded and can also have a form: $pro: \mathcal{A} \times \mathcal{P} \rightarrow \mathcal{P}\mathcal{R}$, informally, it retrieves the information which project sponsors given scientists in the publication.

The above operators allow us to extract two types of collaboration networks from the available databases.

A research information store can be seen as the set of items, e.g. publications. Thus we have: $R = \{P_1, P_2, \dots, P_n\}$, where $|R| = n$ is the size of the store, e.g. in Scopus $n > 33,000,000$. Let us denote $\mathcal{R} = \{R_1, R_2, \dots, R_n\}$ the set of all stores. The first type is a network between individual researchers. Thus, the first network $N^A = (V^A, E_{\mathcal{P}}^A, E_{\mathcal{P}\mathcal{R}}^A, w_{\mathcal{P}}, w_{\mathcal{P}\mathcal{R}}^A)$ is defined as:

$$V^A = \left(\bigcup_{R \in \mathcal{P}} \bigcup_{P \in R} auth(P) \right),$$

$$E_{\mathcal{P}}^A = \left(\bigcup_{R \in \mathcal{R}} \bigcup_{P \in R} \bigcup_{\substack{A_i, A_j \in auth(P) \\ A_i \neq A_j}} (A_i, A_j) \right), \quad E_{\mathcal{P}\mathcal{R}}^A = \left(\bigcup_{R_i, R_j \in \mathcal{R}} \bigcup_{\substack{P_i \in R_i \\ P_j \in R_j}} \bigcup_{\substack{A_i \in auth(P_i) \\ A_j \in auth(P_j) \\ A_i \neq A_j \\ pro(A_i, P_i) = pro(A_j, P_j)}} (A_i, A_j) \right),$$

$$w_{\mathcal{P}}^A(A_i, A_j) = |\{P : \exists_R P \in R \wedge A_i, A_j \in \text{auth}(P)\}|,$$

$$w_{\mathcal{P}\mathcal{R}}^A(A_i, A_j) = \left| \left\{ PR : \begin{array}{l} \exists_R \exists_P (P \in R \wedge A_i \in \text{auth}(P) \wedge \text{pro}(A_i, P) = PR) \wedge \\ \exists_R \exists_P (P \in R \wedge A_j \in \text{auth}(P) \wedge \text{pro}(A_j, P) = PR) \end{array} \right\} \right|.$$

The function $w_{\mathcal{P}}^A(A_i, A_j)$ simply tells us how many joint publications have in their records the two authors A_i, A_j . The more such publications they have, the higher is the wedge on the edge $e_{\mathcal{P}} = (A_i, A_j)$. This wedge is the cardinal representation of the density of their collaboration. The function $w_{\mathcal{P}\mathcal{R}}^A(A_i, A_j)$ also gives us information about the density of collaboration between two scientists. However, this measure interprets the notion of collaboration as participating in the same projects.

The second network being under consideration in this article is the network between scientific institutions: $N^I = (V^I, E_{\mathcal{P}}^I, E_{\mathcal{P}\mathcal{R}}^I, w^I, w_{\mathcal{P}\mathcal{R}}^I)$.

$$V^I = \left(\bigcup_{R \in \mathcal{P}} \bigcup_{P \in R} \text{inst}(P) \right),$$

$$E_{\mathcal{P}}^I = \left(\bigcup_{R \in \mathcal{R}} \bigcup_{P \in R} \bigcup_{\substack{I_i, I_j \in \text{auth}(P) \\ I_i \neq I_j}} (I_i, I_j) \right), \quad E_{\mathcal{P}\mathcal{R}}^I = \left(\bigcup_{R_i, R_k \in \mathcal{R}} \bigcup_{\substack{P_i \in R_i \\ P_k \in R_k}} \bigcup_{\substack{A \in \text{auth}(P_i) \\ A \in \text{auth}(P_k) \\ \text{inst}(A) \neq \text{inst}(A) \\ \text{pro}(A_i, P_i) = \text{pro}(A_j, P_j)}} (\text{inst}(A_i), \text{inst}(A_j)) \right),$$

$$w^I(I_1, I_2) = |\{P : \exists_R P \in R \wedge I_1, I_2 \in \text{inst}(P)\}|,$$

$$w_{\mathcal{P}\mathcal{R}}^I(I_i, I_j) = \left| \left\{ PR : \begin{array}{l} \exists_R \exists_P (P \in R \wedge A_i \in \text{auth}(P) \wedge \text{inst}(A_i) = I_i \wedge \text{pro}(A_i, P) = PR) \wedge \\ \exists_R \exists_P (P \in R \wedge A_j \in \text{auth}(P) \wedge \text{inst}(A_j) = I_j \wedge \text{pro}(A_j, P) = PR) \end{array} \right\} \right|.$$

The above two constructive definitions of collaboration networks show that based on the available stores we can build fully informative collaboration networks between scientists or scientific institutions and, what is of particular pragmatic value, we are able to quantitatively measure their features. Now, we shall present a set of network analysis methods that constitute a basic methodological toolkit of our model. It is important to emphasise that the methods described below are focused on the structure of the network only, without entering upon any content or meaning recognition/analysis of the network. Here are the methods we have chosen at the outset of building the toolkit.

- (1) A pretty common information need among researchers is as follows: A researcher looks for new collaboration opportunities in her/his area of interest. Thus, how can s/he find other researchers who could be potentially interested in setting up a collaboration platform? This task can be boiled down to the topic known in networks analysis as a *link prediction* problem (Lü & Zhou, 2011): Given the links in a network at the point of time T or over a period P , one aims at predicting all the links that will be added to the network during the time interval of a determined length starting at the point of time T' , where $T' > T$. Incidentally, this is not a trivial task; among many algorithms to resolve the link prediction problem discussed in the paper (Szczepański

et al., 2015) we experiment with the algorithms proposed by P. Szczepański et al. in this paper because of their high precision and accuracy.

- (2) A well-known fact of academic and/or research ventures is that while collaborating researchers form informal groups that cannot be identified by affiliations or other metadata provided by their publications stored in digital libraries or repositories. Identifying such groups is a pretty common topic for various reasons, for instance, because one would like to identify existing informal groups with the intent to learn what research they carry out, or try to join them, or perhaps to avoid them. The process of groups' identification can be implemented by means of the so-called *community detection* approach (Girvan & Newman, 2002). The community detection process produces clusters of objects sharing a given characteristic; the organization of nodes (members of a community) in a cluster is such that it includes the nodes with many edges joining the nodes of the same cluster and comparatively few edges joining the nodes of different clusters. Such clusters can be considered as fairly independent networks.
- (3) As indicated a propos the community detection, it is characteristic that the communities within networks are groups of densely connected nodes, while the number of connections between different groups is low. Here, an interesting problem is to find out who is responsible for inter-community connections. Such agents may play a particularly important role as leaders or communicators in establishing cooperation patterns, disseminating information, organising awareness outreach activities, etc. Detecting these agents can be performed by means of algorithms intended to identify *gate-keepers* (also referred as *bridges*) in social networks (Girvan & Newman, 2002).
- (4) Interdisciplinarity is one of the most characteristic features of contemporary science. By identifying "compatible" research networks in different scientific disciplines and combining activities of two or more such networks into one project and thus crossing boundaries within science, one can achieve innovative and valuable outcomes. Towards this end, we developed a methodology dubbed *game-theoretic social analysis technique* (Szczepański et al., 2014), which applies methods of community detection and link prediction.
- (5) A particularly desired capability of the network analysis toolkit is the one that helps establish rankings given a predefined criterion(a). This could be useful, for instance, for identifying the most influential persons in a group, for setting up a list of the institutions best suited to collaboration, for ranking researchers to be hired by a department, etc. The procedures for constructing such rankings can be based on resolving the problem of *centrality* that, generally speaking, consists in quantifying, in one way or another, individual nodes (members of a network/community) in order to identify the most important nodes in a network. In our approach for resolving centrality-like problems we stick to Freeman's methodology (Freeman, 1979), which assumes that a centrality measure can be based on the number of connections (*degree-based centralities*), the distance to the others (*closeness-based centralities*), and/or the number of paths between the others (*betweenness-based centralities*).

4. Summary

Scientific digital libraries and repositories containing hundreds of thousands of publications and other digital artefacts are natural environments within which resides undiscovered/ unspecified information about existing and potential collaboration networks between scientists, scientific institutions, funding agencies, etc. Such networks are undoubtedly valuable sources of data and knowledge whose deep analysis can not only support on-going research processes, but also inspire and initiate new ones. This can be achieved by extracting data and information encoded in thousands of links between collaborating scientists, research institutions and academia by means of network analysis methodologies and tools, some of which we already apply in our work on enhancing the Ω^{PR} system towards having the built-in advanced functionality for discovering the networks of collaborating scientists and scientific institutions and for getting in-depth knowledge residing/hidden in these networks.

Acknowledgment

Piotr Szczepański's contribution to this paper was founded by the Polish National Science Centre through the grant DEC – 2013/09/N/ST6/04095.

References

- Albert, R.; Jeong, H.; Barabasi, A. (2000). Error and attack tolerance of complex networks. *Nature*, 406, 378–382.
- Bork, P.; Jensen, L. J.; Mering, C. von; Ramani, A. K.; Lee, I.; Marcotte, E. M. (2004). Protein interaction networks from yeast to human. *Current Opinion in Structural Biology*, 14 (3), 292–299.
- Freeman, L. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, 1 (3), 215–239.
- Freixas, X.; Parigi, B.; Rochet, J. C. (2000). Systemic risk, interbank relations and liquidity provision by the central bank. *Journal of Money Credit and Banking*, 32 (3), 611–638.
- Girvan, M.; Newman, M.E.J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99 (12), 7821–7826.
- Hockney, M. (2012). *The Last Man Who Knew Everything*. Hyperreality books.
- Leskovec J.; Kleinberg J.; Faloutsos C. (2007). Graph evolution: Densification and shrinking diameters. *ACM Transaction Knowledge Discovery*, vol. 1 (1), article no. 2.
- Lü, L.; Zhou, T. (2011). Link prediction in complex networks: A survey, *Physica A: Statistical Mechanics and its Applications*, 390 (6), 1150–1170.
- Muraszkiewicz, M. (2004). Mobile Network Society, *Dialog and Universalism*, 14 (1–2), 113–124
- Muraszkiewicz, M. (2013). W stronę społeczeństwa sieciowego i inteligentnych miast. Propozycja programu I'Miasto. *Przegląd Telekomunikacyjny*, 8–9, 609–613.
- Muraszkiewicz, M.; Szmidt, J.; Zaremba, K (2014). SYNAT i Ω^{PR} – ku ekosystemowi wsparcia informacyjnego nauki i uczelni polskich. *Zagadnienia Informatyki Naukowej – Studia Informacyjne*, 54 (2), 7–22.
- Newman, M.E.J. (2001): The structure of scientific collaboration networks [online]. In: *Proceedings of the National Academy of Sciences*, 98 (2), 404–409 [24.10.2015], <http://www.pnas.org/content/98/2/404.full>
- Radosevic, S.; Yoruk, E. (2014). Are there global shifts in the world science base? Analysing the catching up and falling behind of world regions. *Scientometrics*, 101 (3), 1897–1924.

- Szczepański, P.L.; Michalak, T.P.; Talal, R.; Barcz, A. (2015). The Game-Theoretic Interaction Index on Social Networks With Applications to Link Prediction and Community Detection. In: *Proceedings of the 24th International Joint Conference on Artificial Intelligence*. Palo Alto: AAAI Press / International Joint Conferences on Artificial Intelligence, 638–644.
- Szczepański, P.L.; Michalak, T.P.; Wooldridge, M. (2014). A Centrality Measure for Networks With Community Structure Based on a Generalization of the Owen Value. *Frontiers in Artificial Intelligence and Applications*, 263, 867–872.
- Wasserman, S.; Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, New York: Cambridge University Press.

Odkrywanie sieci współpracy badawczej z naukowych bibliotek cyfrowych i repozytoriów

Abstrakt

Cel/teza: Celem pracy jest naszkicowanie praktycznego modelu odkrywania sieci współpracy badawczej korzystając z zasobów naukowych bibliotek cyfrowych i repozytoriów. Odkryte zależności wiążące badaczy, projekty, instytucje naukowe i inne naukowe przedsięwzięcia i artefakty stanowią podstawę do wyodrębnienia sieci współpracy naukowców i instytucji naukowych zainteresowanych wspólną tematyką badawczą. Sieci takie mogą być następnie przedmiotem analizy w celu uzyskania pogłębionej wiedzy na ich temat.

Koncepcja/metody badań: Pracę oparto na metodzie, która ma dwa składniki, a mianowicie: (i) wykorzystano proste mechanizmy odkrywania sieci współpracy badawczej opracowane i zastosowane w ramach systemu bazy wiedzy akademickiej Ω^{VR} , który zrealizowano na Politechnice Warszawskiej oraz (ii) opracowano zarys formalnego modelu sieci współpracy naukowej, który bierze pod uwagę specyfikę naukowych bibliotek cyfrowych i repozytoriów oraz zawiera zbiór technik analizy sieciowej, pozwalających na odkrywanie wiedzy zawartej/ukrytej w sieciach współpracy naukowej.

Wyniki i wnioski: Przedstawiono zarys formalnego modelu sieci współpracy naukowej, który ma dwa komponenty, a mianowicie: (i) mechanizm odkrywania tematycznie skorelowanych badaczy, projektów, instytucji naukowych i innych podmiotów i artefaktów naukowych oraz (ii) zbiór metod analizy sieciowej, które umożliwiają wykrywanie wiedzy zawartej w sieciach współpracy naukowej. Zaproponowany model jest skalowalny zarówno w zakresie jego funkcjonalności, jak i technik analizy sieciowej. Został on oparty na sprawdzonych rozwiązaniach zrealizowanych w ramach systemu Ω^{VR} i jest obecnie wykorzystany w pracach nad rozszerzeniem tego systemu.

Oryginalność/wartość poznawcza: Opracowany i realizowany w ramach prac nad rozszerzeniem systemu Ω^{VR} własny model pozwala odkrywać i w pogłębiony sposób analizować naukowe sieci współpracy badawczej, co zwiększa zakres, wartość i wpływ przedsięwzięć naukowych na rozwój nauki i społeczeństwa.

Słowa kluczowe

Sieci współpracy badawczej. Model sieci współpracy badawczej. Analiza sieci. Odkrywanie wiedzy. System Ω^{VR} .

Prof. MIECZYSLAW MURASZEWICZ is a professor of computer science at Warsaw University of Technology interested in knowledge representation and the relationships between technology and culture. His recent publications are: B. Jacobfeuerborn, M. Muraszewicz: Media, Information Overload, and Information Science. In: R. Bemberik et. al. (eds.) Intelligent Tools for Building Scientific Information Platform. Advanced Architectures and Solutions Berlin, Heidelberg, Springer 2013, 3–13 [Studies in Computational Intelligence

No. 467]; Muraszkwicz, M.: *Information Science Needs Cognitivism. An Essay. Miscellanea Informatologica Varsoviensia*, SBP, 2012; Muraszkwicz, M., Szmiedt, J., Zaremba, K (2014). SYNAT i Ω^{WR} – ku ekosystemowi wsparcia informacyjnego nauki i uczelni polskich. *Issues on Information Science (ZIN)*, 54(2), 7–22.

Contact to the Author:

m.muraszkiewicz@ii.pw.edu.pl

Institute of Computer Science

Faculty of Electronics and Information Technology

Warsaw University of Technology

Nowowiejska 15/19

00–665 Warszawa

Prof. HENRYK RYBIŃSKI is a professor of computer science at Warsaw University of Technology. His research areas cover knowledge representation, data and text mining, and information retrieval. His recent publications are: J. Koperwas, Ł. Skonieczny, M. Kozłowski, H. Rybiński, W. Struk: *University Knowledge Base: Two Years of Experience*, In: *Intelligent Tools for Building a Scientific Information Platform: From Research to Implementation / Bembenik Robert [et al.] (eds.), Studies in Computational Intelligence 2014*, vol. 541, Springer; R. Krajewski, M. Kozłowski, H. Rybiński A novel method for dictionary translation, *Journal of Intelligent Information Systems*, 2015, 1–24, J. Koperwas, Ł. Skonieczny, M. Kozłowski, P. Andrzejewicz, H. Rybiński, W. Struk: *AI Platform for Building University Research Knowledge Base*, In: *Foundations of Intelligent Systems. Proceedings / A. Troels [et.al.] (eds), LNAI*, vol. 8502, 2014, Springer.

Contact to the Author:

h.rybinski@ii.pw.edu.pl

Institute of Computer Science

Faculty of Electronics and Information Technology

Warsaw University of Technology

Nowowiejska 15/19

00–665 Warszawa

PIOTR SZCZEPAŃSKI, MSc Eng., is a Ph.D. student of computer science at Warsaw University of Technology. His research topics mostly cover the intersection of Social Networks and Game-Theory including knowledge representation and information mining. His recent publications are: T. Michalak, K. Aaditha, P. Szczepański, B. Ravindran, N.Jennings: *Efficient computation of the Shapley value for game-theoretic network centrality. Journal of AI Research 2013*, vol. 46, 607–650; P. Szczepański, T. Michalak, M. Wooldridge: *A Centrality Measure for Networks with Community Structure Based on a Generalization of the Owen Value. Frontiers in Artificial Intelligence and Applications*, 2014, vol. 263, 867–872; P. Szczepański, M. Tarkowski, T. Michalak, P. Harrenstein, M. Wooldridge: *Efficient Computation of Semivalues for Game-Theoretic Network Centrality. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015, vol. 1, 461–469.

Contact to the Author:

p.szczepanski@ii.pw.edu.pl

Institute of Computer Science

Faculty of Electronics and Information Technology

Warsaw University of Technology

Nowowiejska 15/19

00–665 Warszawa