Fundamenta Informaticae XXI (2018) 1–21
DOI 10.3233/FI-2016-0000
IOS Press

Gender-Based Analysis of Intra-Institutional Research Productivity and Collaboration

Miloš Savić, Mirjana Ivanović, Miloš Radovanović
University of Novi Sad, Faculty of Sciences
Department of Mathematics and Informatics
Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia
{*svc, mira, radacha*}*@dmi.uns.ac.rs*

Bojana Dimić Surla
Union University, School of Computing
Knez Mihailova 6, 11000 Belgrade, Serbia
bdimicsurla@raf.rs

Abstract. Current Research Information Systems (CRISs) offer great opportunities for assessments of institutional research outputs and extraction of useful and actionable knowledge based on various data-analysis techniques. However, many of these opportunities have not been explored in depth, especially in culture-sensitive areas such as gender-based analysis of research productivity and collaboration. In this paper we present GERBER, a network-based methodology and accompanying tool for gender-based analysis of publication data stored in institutional CRISs. GERBER relies on statistically robust techniques applied on weighted co-authorship networks whose nodes are enriched with different types of researcher evaluation metrics. The functionality of GERBER is demonstrated on publication data stored in the institutional CRIS of the Faculty of Sciences, University of Novi Sad, Serbia. The obtained results show that GERBER enables institutional research managers and policy makers to detect gender inequalities and homophily in research productivity and collaboration. Finally, we discuss different possibilities to integrate GERBER with CRISs in order to facilitate continuous gender-based evaluation of researchers.

Address for correspondence: Miloš Savić, University of Novi Sad, Faculty of Sciences, Department of Mathematics and Informatics, Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia, svc@dmi.uns.ac.rs

Keywords: research information systems, research productivity, research collaboration, gender-based analysis, gender inequalities, institutional analysis

1. Introduction

Measuring gender differences in research and teaching productivity has been a topic of interest for researchers and decision makers for many decades [1]. The basic generalization found in the literature is that male faculty outperform female faculty [2, 3] (for a more comprehensive list, please see [1] and references therein). However, more recent studies have shown that this is by no means due to some inherent superiority of one gender over the other. Xie and Shauman [4] conducted four large, nationally representative, cross-sectional surveys spanning several decades, observing that differences in research productivity declined over the period 1969–1993. Similarly, Gander [1] concludes that, although analysis of data at first glance supports the generalization, further scrutiny reveals that patterns of employment and distribution of funding are actually the main causes of the observed differences in productivity. Once the analysis is adjusted, it is revealed that female faculty have significant research productivity [1]. On the other hand, gender differences in research collaboration have received considerably less research attention in spite of the fact that contemporary scientific research is increasingly conducted in collaboration [5].

Research productivity and collaboration can be studied at various levels: intra-institutional (within an institution), institutional (between institutions), national (within a country), international (between countries), disciplinary (within a scientific discipline), inter-disciplinary (between scientific disciplines), etc. Current Research Information Systems (CRISs) offer rarely exploited opportunities for scientometrics studies at the intra-institutional and national level. Exceptional examples are studies by Leeuwen et al. [6] who performed bibliometric analysis of research performance of a Dutch university using data stored in its institutional CRIS and Perc [7] and Kastrin et. al [8] who investigated research productivity and collaboration of Slovenian researchers relying on publication data recorded in the national CRIS system of Slovenia. In our previous work we proposed a methodology based on co-authorship networks extracted from CRIS databases to analyze intra-institutional research collaboration [9]. In spite of the fact that CRIS systems can be exploited to investigate gender differences in research productivity and collaboration at the intra-institutional level we are not aware of any such study.

In this paper we propose a methodology for gender-based analysis of intra-institutional research productivity and collaboration. The methodology is based on statistically robust analysis of enriched co-authorship networks whose nodes have gender labels. Co-authorship networks reflect the structure of research collaboration. Enriched co-authorship networks are co-authorship networks whose nodes are extended by a rich vector of metrics which reflect productivity, collaboration and institutional importance of researchers. The methodology is implemented within a tool called GERBER (**GE**nde**R** **B**ased **E**valuation of **R**esearchers). We also performed an XML-based integration of the tool with CRIS UNS. CRIS UNS is an information system for storing and managing data about scientific research activity at the University of Novi Sad (UNS), Serbia. It was developed by the recommendations of the non-profit organization euroCRIS [10]. CRIS UNS provides a comprehensive list of publications of researchers affiliated with UNS and enables automated evaluation of UNS researchers and institutions. Using GERBER we performed the gender-based analysis of productivity and collaboration of researchers employed at UNS-PMF – the Faculty of Sciences, University of Novi Sad, Serbia.

The rest of the paper is structured as follows. The overview of related research works is given in Section 2. The next section describes our methodology for gender-based analysis of intra-institutional research productivity and collaboration. The accompanying tool is presented in Section 4. The analysis conducted using the GERBER tool is presented in Section 5. Integration of analytic tools with CRIS systems is discussed in Section 6. In the last section we give conclusions and directions for future work.

2. Related Work

In this section we will review related studies of gender differences in scientific productivity and collaboration (Section 2.1) and give the background of the CRIS UNS information system for storing and managing data about scientific research (Section 2.2).

2.1. Scientometric Analysis of Gender in Research

The comprehensive multi-decade study by Xie and Shauman [4], besides observing the decline of gender differences in research productivity, also correlates gender productivity differences with gender differences in personal characteristics, structural positions, and marital status, implying that gender differences in research productivity stem from gender differences in structural locations, and as such respond to the secular improvement of women's position in science.

Contemporary data in the field of social science shows that not only did gender differences disappear in the younger generations of researchers, but that if some differences exist, it is the young female researchers that outperform their male peers [11]. In educational psychology, on the other hand, although females are gaining ground in terms of primary and secondary article authorship and journal editorial board membership, this increase does not keep pace with the male-female ratio in organizational memberships [12]. Also, in the industrial and organizational psychology there are significant gender differences with respect to publication output (fewer publications authored by female researchers) and career courses (male researchers have longer careers, but with longer interruptions) [13]. However, the projected future promises a much more balanced situation in this field [13].

In the domains of science, engineering and technology, the trends are also varying, with the overall impression that participation and performance of women improved in recent times. Within Spanish natural resources and chemistry scientists, no significant differences in productivity were found between genders within professional categories, but the outliers with the highest production were for the most part male [14]. In nano science and technology, female researchers are scarce in number, but perform equally in terms of scientific production and impact [15]. A comprehensive study involving the DBLP database of computer science publications ranging from 1936 to 2010 also indicates a low percentage of women in this field, albeit a steadily rising one [16]. The authors observed that men publish more than women, but attributed this to the fact that the average research life of men is longer. In software engineering journals, the percentage of women authors is roughly on par with the general trend in computer science (∼16–17%), comparable to the percentage of female editors (∼18%), but women editors-in-chief are under-represented (∼9.5%) [17].

Studies of gender differences in scientific productivity also produce varying results in different countries. In Croatia [18], within the studied young research population, females are somewhat less productive than males, with structural variables being the most powerful factor influencing this distinction, which is in line with the observations by Xie and Shauman [4] discussed earlier. In Italy [19], there is also evidence of higher overall male productivity, but with difference smaller than reported in a large part of the literature, confirming an ongoing tendency towards decline. Russia [20], on the other hand, still exhibits strong gender disparity, which can also be said for Turkish social sciences [21]. In Spain, there is no overall significant difference among young postdoctoral researchers in terms of productivity [22], with women leading the way in the number of citations, but also in the number of researchers with no output after obtaining their Ph.D.

The above constitutes only a small representative sample of studies dealing with gender differences in scientific production and teaching. A comprehensive meta-study of scientific literature on women in science and higher education, considering almost 1500 articles, shows continued growing interest in the topic, featuring more than 3000 authors, 67 countries, and 86 research areas [23].

Gender differences in research collaboration have received considerably less research attention compared to gender differences in research productivity (an overview of existing studies can be found in [5]). Additionally, the literature shows mixed conclusions about gender aspects of research collaboration. For example, the study by Bozeman and Gaughan [24] from 2011 indicated that U.S. female researchers tend to have more collaborators than U.S. male researchers, while the study by Zeng et al. [25] from 2016 indicated exactly the opposite. The most comprehensive analysis of gender differences in research collaboration at the national level was performed by Abramo et al. [5]. Their study involving the entire population of Italian researchers in the "hard" sciences and economics showed that gender differences in research collaboration do not exist at the intra-institutional and national level, but gender gap can be observed in the propensity to collaborate at the international level.

2.2. CRIS UNS

The starting point in developing CRIS UNS was creating a well-structured and comprehensive metadata set for describing scientific results, as well as researchers and institutions. Ivanović et al. [26] proposed the metadata model based on the MARC 21 library standard and compatible with CERIF (Common European Research Information Format). The CERIF data model provides a very rich and well-structured set of metadata. The core of CERIF are three basic entities Person, Project and OrganisationUnit, and three result entities ResultPatent, ResultPublication and ResultProduct [27]. The structure of CERIF enabled the development of the information system in which authors are uniquely identified and connected to their results, institutions and projects. The metadata set for some entities was further enriched by introducing the MARC 21 format of bibliographic data for presenting publications and the MARC 21 format of authority data for presenting authors [28]. The described model was the basis for developing the information system CRIS UNS for tracking research activity at the University of Novi Sad [29].

Speaking of research activity, one of the main purposes for developing the CRIS UNS system was providing automated evaluation of scientific results, researchers and institutions, which has become extremely important. Paper [30] proposed an extension of CERIF by data for evaluation of published scientific results. The extension is based on the CERIF semantic layer that enables classification of entities and their relationships by different classification schemas. The rules for evaluation proposed by the academic regulatory bodies were implemented within CRIS UNS and exposed as a service for evaluation of scientific results [31].

The architecture of CRIS UNS with its rich matadata model and module for evaluation of scientific results provided the environment for developing various techniques and tools for business analysis and gathering important information used by institutional management. These tools included a module for creating periodical reports on research activity and different analyses [32]. Architecture of the CRIS UNS and adoption of international standards for presenting research data provided appropriate environment for interoperability with other systems [33] including ontology-based integration [34, 35].

Each researcher employed at our institution is obligated to have his/her CRIS UNS profile and periodically update his/her bibliographic references. Additionally, each CRIS UNS profile contains all institutionally relevant information about the researcher (academic rank, research department within the institution, year of birth, etc.) including also his/her gender.

3. GERBER Methodology

The GERBER methodology for gender-based analysis of intra-institutional research productivity and collaboration is based on the notion of enriched institutional co-authorship networks [9]. An enriched co-authorship network associated to an institution I is an undirected, labeled and weighted graph $G = (V, E, W)$ with the following properties:

1. The set of nodes V corresponds to researchers employed at I and their collaborators not affiliated with I. Each node in V is labeled either as *local* or *external* – local nodes correspond to researchers from I, while external nodes represent researchers not affiliated with I.

2. The set of links E corresponds to research collaborations among researchers from V. Two researchers are connected by an undirected link if they co-authored at least one publication together (with or without other co-authors). Each link is classified either as *intra-institutional* (a link connecting two local nodes) or *inter-institutional* (a link connecting a local to an external node).

3. The function $W : E \to \mathbb{R}$ determines link weights which reflect the strength of research collaboration among connected researchers. The GERBER methodology uses the link weighting scheme proposed by Newman [36]. The Newman scheme takes into account the total number of authors of a paper when quantifying the strength of research collaboration between two authors. The weight of link connecting two researchers a and b is computed according to the following formula:
$$W(a \leftrightarrow b) = \sum_{k \in J} \frac{1}{n_k - 1},$$

where J denotes the set of joint publications of a and b and n_k is the total number of authors of publication k. The main characteristics of the scheme is that the sum of weights of all links incident to a node is equal to the total number of multi-authored publications of the corresponding researcher.

4. The network has an underlying institutional structure. This means that the set of local nodes can be partitioned into non-overlapping groups where each group corresponds to one organizational unit (research department) within the institution I.

5. A metric vector containing different types of researcher evaluation metrics (productivity, collaboration and metrics of institutional importance) is attached to each local node.

The GERBER methodology extends the previously described notion with one more node attribute representing gender of researchers. This means that each local node can be classified either as *male* or *female*. Consequently we can distinguish between three types of intra-institutional links:

1. links representing collaboration between two male researchers (MM links),
2. links representing collaboration between two female researchers (FF links), and
3. links representing collaboration between male and female researchers (MF links).

The GERBER methodology relies on both domain-independent and country-specific researcher productivity metrics. Three commonly used researcher productivity metrics known as normal counting, fractional (adjusted) counting, and straight counting scheme [37] are used to enrich nodes of the co-authorship network. The productivity of a researcher estimated by the normal counting scheme is equal to the total number of publications he/she (co-)authored. The straight counting scheme assigns the whole credit for a publication only to the first author, which means that the productivity of a researcher measured by this scheme is equal to the number of publications in which he/she is the first co-author (or the only author in the case of single-authored publications). The fractional counting scheme assigns credit equal to $1/n$ to each of n authors of a publication. Therefore, the productivity of a researcher a is equal to

$$P_{\text{fractional}}(a) = \sum_{k \in S} \frac{1}{n_k},$$

where S is the set of publications (co-)authored by a and n_k denotes the number of authors of publication k.

In our implementation of the methodology we also use a Serbian-specific researcher productivity metric known as *Serbian research competency index*. Namely, research publications (co-)authored by Serbian researchers are categorized according to the rule book prescribed by the Serbian Ministry of Education, Science and Technological Development. The rule book defines several categories of publication venues and each category correspond to a certain number of points. For example, papers published in the top 30% SCI ranked journals in appropriate scientific discipline are worth 8 points, papers presented at conferences are worth 1 point, etc. The competency index of a Serbian researcher is then defined as the sum of points of publications he/she (co)authored. This index is used as one of the criteria in the process of academic promotions at Serbian universities, as well as in researcher evaluation within Serbian national research projects.

The degree of collaboration of a researcher with other researchers can be quantified by its *degree centrality* in the co-authorship network. The degree centrality of researcher a is equal to the number of links incident to a, or equivalently, to the total number of co-authors of a. Authors without collaborators (authors whose scientific output solely consists of solo-authored publications) have degree centrality equal to 0 and appear as isolated nodes in the co-authorship network. Since there are two types of nodes in the network (local and external) we can derive two additional degree centrality measures:

- *local degree centrality* – the number of local co-authors of a, and
- *external degree centrality* – the number of external co-authors of a.

Clearly, degree centrality is equal to the sum of local and external degree centralities.

Another local centrality measure used in our methodology is *shell index* associated to the k-core decomposition of networks [38]. The k-core of a graph is a maximal subgraph in which the degree centrality of each node is higher than or equal to k. The k-core can be obtained by recursively deleting all nodes whose degree centrality is smaller than k. A node has shell index k if it belongs to the k-core but not to $(k+1)-$core. Shell index enables us to distinguish between two types of researchers having a high degree centrality: those connected to a large number of other researchers with a low degree centrality have a low shell index, while those connected to a large number of other researchers with a high degree centrality have a high shell index.

Enriched co-authorship networks are weighted graphs. Thus, we can compute *weighted degree centrality* for nodes. This measure reflects the strength of research collaboration of an author with other authors. The weighted degree centrality of a node is equal to the sum of weights of all links incident to the node. Additionally, the GERBER methodology includes the following variants of the weighted degree centrality measure according to different types of links in the network:

- *local weighted degree centrality* – the sum of weights of local co-authorship links incident to a researcher reflecting the strength of its intra-institutional research collaboration,
- *external weighted degree centrality* – the sum of weights of external co-authorship links incident to a researcher reflecting the strength of its inter-institutional research collaboration,
- *intra-department weighted degree centrality* – the sum of weights of intra-department co-authorship links incident to a researcher reflecting the strength of its collaboration with other researchers from the same department, and
- *inter-department weighted degree centrality* – the sum of weights of inter-department co-authorship links incident to a researcher reflecting the strength of its collaboration with institutional colleagues from other departments.

Clearly, weighted degree centrality is equal to the sum of local and external weighted degree centralities, while local weighted degree centrality is equal to the sum of intra- and inter-department weighted degree centralities.

Local degree centrality can be viewed as the simplest local measure of institutional importance. A researcher having a large number of local collaborators can be considered institutionally important because he/she is in the position to act as a bridge between a large number of his/her co-authors. The GERBER methodology also relies on two global node centrality metrics to quantify institutional importance of researchers: *betweenness centrality* [39, 40] and *closeness centrality* [41].

The betweenness centrality of node (author) a is the extent to which a is located on the shortest paths connecting two arbitrary selected nodes in the network. More formally,

$$\text{Betweenness}\ (a) = \sum_{b,c \in V, a \neq b \neq c} \frac{\sigma(a, b, c)}{\sigma(b, c)},$$

where V denotes the set of nodes in the network, $\sigma(b, c)$ is the number of shortest paths connecting nodes (authors) b and c, and $\sigma(a, b, c)$ is the number of shortest paths connecting b and c that pass through a. If a large fraction of shortest paths contain a, then a can be viewed as an important node of the network in the sense that it has a vital role to the overall connectivity of the network. If the network has a clustered or community organization, then nodes with high betweenness centrality tend to be located at the intersections of communities, which means that they connect together different cohesive research groups. Betweenness centrality can be also viewed as a measure of the influence that a node has over the spread of information through the network, i.e. nodes having a high betweenness centrality are in the position to maintain and control the spread of information over the network.

The closeness centrality of node a is inversely proportional to the cumulative distance between a to other nodes in the network. More formally,

$$\text{Closeness}\ (a) = \frac{1}{\sum_{b \in V, b \neq a} d_{ab}},$$

where V is the set of nodes in the network, and d_{ab} is the length of the shortest path connecting a and b. Nodes with high closeness centrality can be considered as socially important since they are in proximity to a large number of other nodes. In other words, nodes with high closeness centrality tend to be located in the core of the network, while nodes having low closeness centrality are located on the periphery of the network.

The last category of researcher evaluation metrics present in the GERBER methodology are metrics related to characteristics of ego-networks. The ego network of node a in an undirected graph G, denoted by $Ego(a)$, is a sub-graph of G induced by a and its nearest neighbors. The cohesiveness of ego-networks can be quantified by *clustering coefficient* [42]. The clustering coefficient of node a, denoted by $CC(a)$, is the probability that two randomly selected neighbors of a are directly connected. If $CC(a) = 1$ then neighbors of a form the most cohesive ego-network – a clique. The lowest value of $CC(a)$ is equal to 0 and happens when co-authors of a have never collaborated among themselves. To quantify the gender structure of ego-networks we introduce a metric called *co-author gender disbalance* (CGD). Let $M(a)$ and $F(a)$ denote the fraction of male and female collaborators of author a. Then, the co-author gender imbalance of a is defined as $CGD(a) = |M(a) - F(a)|$. If $CGD(a) = 1$ then all collaborators of a have the same gender, while $CGD(a) = 0$ implies that a equally collaborates with male and female researchers. It is important to emphasize that global centrality metrics and ego-network metrics are computed on the reduced co-authorship network that encompasses only local researchers for two reasons:

1. The reduced co-authorship network provides an institutional boundary for the interpretation of obtained metric values, i.e. global centrality and ego-network metrics computed on the reduced co-authorship network reflect the importance of researchers within the institution and intra-institutional cohesiveness of research collaboration, respectively.

2. Co-authorship networks extracted from institutional bibliographic databases may not contain all links between external collaborators since institutional bibliographic databases do not contain full bibliographies of external collaborators, but only those publications made in cooperation with local researchers. Consequently, ego-network metrics computed on non-reduced institutional co-authorship networks may be biased towards local researchers without external collaborators. On the other hand, global centrality metrics computed on non-reduced institutional co-authorship networks cannot be used to assess the international importance of local researchers since institutional bibliographic databases do not provide a broad publication coverage in respective scientific disciplines.

The GERBER methodology relies on non-parametric statistical tests applied to the sets of metric values of independent groups of nodes/links in enriched co-authorship networks in order to detect gender inequalities regarding research productivity and collaboration. The used statistical tests are the Mann-Whitney U (MWU) test [43] and the two-sample Kolmogorov-Smirnov (KS) test [44].

Let M be an arbitrary node/link metric (a metric of researcher productivity, collaboration or institutional importance in case of nodes and link weight in case of links). Let G_1 and G_2 denote two sets of M values for two independent groups of nodes and links. The MWU test is a rank-based test of stochastic superiority and it can be employed to test the null hypothesis that the values in G_1 do not tend to be systematically smaller or greater than the values in G_2. The test is based on the U statistic which is the number of times a value from G_2 precedes a value from G_1 in the sorted sequence of values from both groups. Under the null hypothesis U closely follows a normal distribution. The null hypothesis is rejected if the obtained p-value is smaller than 0.05, and in such cases we can conclude that there is a statistically significant difference between two groups of nodes/links regarding the aspect quantified by M. To quantify the effect size of the difference we use two probabilities of superiority [45]:

1. PS_1 – the probability that a randomly selected value from G_1 is strictly higher than a randomly selected value from G_2, and

2. PS_2 – which is the opposite probability of superiority, i.e. the probability that a randomly selected value from G_2 is strictly higher than a randomly selected value from G_1.

$PS_1 + PS_2$ is not necessarily equal to 1: $1 - (PS_1 + PS_2)$ is the probability that a randomly selected value from G_1 is equal to a randomly selected value from G_2.

The KS test checks the null hypothesis that cumulative distributions of G_1 and G_2 are not significantly different. The test relies on the maximal vertical distance between two empirically observed distributions (the D statistic). The null hypothesis is rejected if the obtained p-value is smaller than 0.05.

The MWU and KS tests are employed to:

1. Detect gender inequalities regarding researcher productivity, collaboration and institutional importance at the institutional level and the level of research departments within institution. In other words, for each of the previously mentioned researcher evaluation metrics the tests are applied on G_1 and G_2 where G_1 corresponds to all male researchers (resp., male researchers from a department D) and G_2 to all female researchers (resp., female researchers from D) at the institutional level (resp., the level of the department D).

2. Detect gender homophily (the preference for collaboration with researchers of the same gender) in research collaboration at the institutional level, the level of intra-department collaboration and the level of inter-department collaboration. This means that the tests are used to detect statistically significant differences in link weights for all, intra-department and inter-department MM, FF and FM links, respectively.

4. GERBER Tool

Generally speaking, there are two options for preparing data for gender-based analysis of research productivity and collaboration: the first one is to export all relevant data to a format suitable for analysis tools, while the second one is to provide an interface for retrieving relevant data directly from a CRIS database. In this paper we opted for the first option, and discuss the second option in Section 6.

GERBER is a standalone tool implemented in Java that performs gender-based analysis of data exported from the CRIS UNS system according to the methodology described in the previous section. The tool consists of three modules: Data Loader, Author Metric and Gender Analyzer. The architecture of GERBER is shown in Figure 1. We used the existing interoperability architecture of CRIS UNS to develop a module for exporting publications and authors metadata to XML documents.

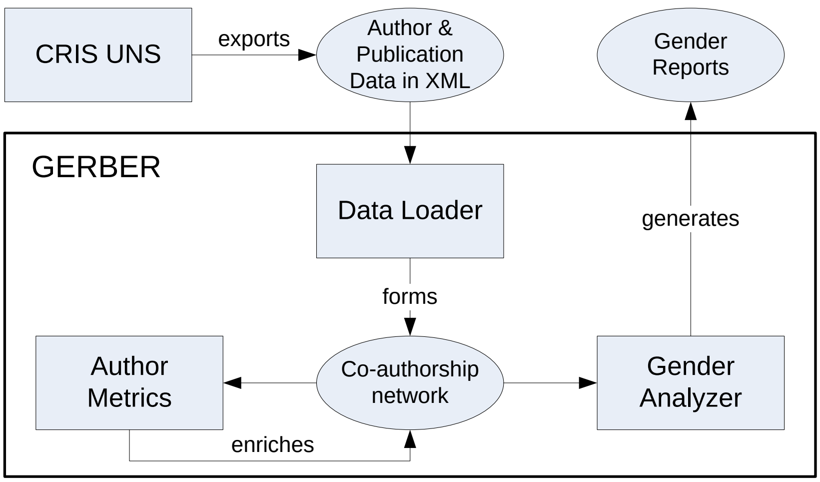

Figure 1. The architecture of GERBER.

The Data Loader (DL) module parses two XML files that contain the data exported from the CRIS UNS information system and forms the co-authorship network of researchers appearing in the data. The first XML file contains metadata about all UNS-PMF researchers and their direct external collaborators (researchers not affiliated with UNS-PMF). Each author is described by an XML element which includes unique author identifier, author name, date of birth, institution to which the author is affiliated, organizational unit within the institution, academic rank, and gender. The second XML file contains metadata about publications which are authored by UNS-PMF researchers. Each publication is described by an XML element which consists of the following information: unique publication identifier, the complete list of author identifiers, publication year, title, publication type (journal, conference, monograph, etc.), information about publication venue and the quantitative evaluation of the publication by the rule book prescribed by the Serbian ministry of science.

The extraction of co-authorship networks from CRIS-UNS data is a straightforward task since researchers present in the CRIS-UNS database are uniquely identified and consequently there are no name disambiguation problems. The DL module forms a co-authorship network in three phases. The set of nodes is formed in the first stage. The DL module iterates through the list of author XML elements and for each element creates one node in the co-authorship network. In the second phase links in the network are formed. A decentralized inverted index which maps authors to their publications is also formed in this phase. The inverted index is decentralized in the sense that a list of all publications of a researchers is directly attached to the corresponding node in the co-authorship network. The DL module iterates through the list of publication XML elements and for each publication p does the following:

- Connects each two authors of p by an undirected link.
- For each author a of publication p, adds p in the list of publications a authored.

In the last phase, weights of co-authorship links are determined according the Newman weighting scheme. The decentralized inverted index is used to compute the set of joint publications for two researchers directly connected in the network.

The Author Metrics (AM) module enriches nodes of the co-authorship network formed by the Data Loader module with research evaluation metrics that reflect author productivity, collaboration, institutional importance and characteristics of ego networks. Table 1 shows the complete list of metrics computed by the AM module. The productivity metrics are computed using the decentralized inverted index constructed by the DM module. Other researcher evaluation metrics are computed from the co-authorship network. Since the co-authorship network is extracted from the institutional bibliographic database it cannot be instrumented to estimate the international importance of UNS-PMF researchers.

The Gender Analyzer (GA) module performs statistical comparison of male and female researchers considering researcher evaluation metrics computed by the AM module and statistical comparison of weights of different types of links. This module implements two non-parametric statistical tests: the Mann-Whitney U test and the two-sample Kolmogorov-Smirnov test. The GA module makes three types of reports:

- basic gender statistics considering organizational units covered by the CRIS-UNS database (different departments at our faculty),

Table 1. The list of author metrics computed by the AM module.

Metric	Abbreviation	Metric category
Productivity, normal count	PRO-N	Productivity
Productivity, fractional count	PRO-F	Productivity
Productivity, straight count	PRO-S	Productivity
Serbian Research Competency Index	CI	Productivity
Degree centrality	DEG	Collaboration
Local degree centrality	LDEG	Collaboration
External degree centrality	EDEG	Collaboration
Shell index	SHI	Collaboration
Weighted degree centrality	WD	Collaboration
Weighted local degree centrality	WD-LOCAL	Collaboration
Weighted external degree centrality	WD-EXTERNAL	Collaboration
Weighted intra-department degree centrality	WD-INTRAD	Collaboration
Weighted inter-department degree centrality	WD-INTERD	Collaboration
Betweenness centrality	BET	Institutional importance
Closeness centrality	CLO	Institutional importance
Clustering coefficient	CC	Collaborators
Co-author gender disbalance	CGD	Collaborators

- the results of the non-parametric statistical tests for the whole institution and each department within the institution, and
- tables that contain values of the Spearman correlation coefficient between different researcher evaluation metrics considering male and female researchers separately.

5. Results and Discussion

Using GERBER we performed the gender-based analysis of research productivity and collaboration at the Faculty of Sciences, University of Novi Sad (UNS-PMF). The publication and author metadata exported from the CRIS-UNS database covers 423 researchers employed at UNS-PMF and their 15097 publications written in collaboration with 5267 researchers not affiliated with UNS-PMF. The co-authorship network extracted from the exported data contains 34111 links, where 2859 links (8.38%) represent local collaborations – collaborations between UNS-PMF researchers. Only 13 nodes in the network are isolated, while other nodes belong to a giant connected component. The existence of the giant connected component in the network indicates that UNS-PMF researchers overall form a cohesive, mature research community.

Table 2 shows basic gender statistics per organizational units (departments) of UNS-PMF. As can be observed, the majority of UNS-PMF researchers are female (60.76% of the total number). Female researchers are in a strong majority at the Department of Biology and Ecology and the Department of Chemistry. The smallest gender gap can be observed at the Department of Mathematics and Informatics where male and female researchers are almost equally represented.

Table 2. Basic gender statistics of UNS-PMF researchers. R denote the absolute number of researchers, while M and F are percentages of male and female researchers, respectively.

Department	R	M[%]	F[%]
Mathematics and Informatics	87	49.43	50.57
Geography	66	57.58	42.42
Biology and Ecology	118	25.42	74.58
Physics	57	56.14	43.86
Chemistry	95	24.21	75.79
Total	426	39.24	60.76

The results of statistical tests performed by the GA module of GERBER are summarized in Table 3. Although UNS-PMF male researchers on average have slightly higher values of all productivity metrics compared to UNS-PMF female researchers, the application of non-parametric statistical tests revealed that there are no statistically significant gender differences regarding scientific productivity. Also, we noticed that there are strong positive Spearman correlations between different productivity metrics for both genders – the lowest value of Spearman correlations for a randomly selected pair of productivity metrics is equal to 0.83.

Regarding research collaboration we can see that UNS-PMF male researchers do not tend to have more both local and external collaborators than UNS-PMF female researchers, and vice versa. Collaboration metrics also exhibit strong Spearman correlations to productivity metrics (see Table 4). Moreover, the external degree centrality stronger correlates to productivity metrics compared to the local degree centrality which means that external collaborations have stronger impact to productivity for both UNS-PMF male and female researchers compared to local collaborations.

From the data presented in Table 3 it can be observed that the null hypothesis of both non-parametric statistical tests were rejected for the betweenness centrality metric, but not for the closeness centrality metric. This means that there are statistically significant differences between UNS-PMF male and female researchers considering their institutional importance. Namely, UNS-PMF male researchers are not dominant in the core of the co-authorship network, but they more frequently appear as bridges that connect different, highly cohesive research groups. Statistically significant gender differences can be also observed for the clustering coefficient: ego-networks of UNS-PMF female researchers tend to be slightly more cohesive than ego-networks of UNS-PMF male researchers. This suggests that UNS-PMF female researchers tend to stimulate their unconnected collaborators to work together more often compared to UNS-PMF male researchers. Having in mind that UNS-PMF male researchers tend to have higher betweenness centrality we can conclude the following:

- UNS-PMF male researchers tend to be more important for the cohesion of the institution at the

Table 3. The results of statistical comparison of UNS-PMF male and female researchers. $\langle M \rangle$ and $\langle F \rangle$ denote the average values of corresponding metric for male and female researchers, respectively. U is the value of the Mann-Whitney test statistic, MWU-p denotes the p-value of the MWU test, PS_m and PS_f are male and female probabilities of superiority, respectively. D is the value of the Kolmogorov-Smirnov test statistic and KS-p denotes the p-value of the KS test. Bold p values indicate statistically significant differences.

Metric	$\langle M \rangle$	$\langle F \rangle$	U	MWU-p	PS_m	PS_f	D	KS-p
PRO-N	82.72	65.40	20741	0.63	0.51	0.48	0.07	0.71
PRO-F	25.52	17.85	19975.5	0.27	0.53	0.47	0.10	0.29
PRO-S	25.83	19.33	20754	0.64	0.50	0.47	0.09	0.40
CI	132.05	101.40	20056	0.30	0.53	0.47	0.10	0.23
DEG	51.24	44.98	20440	0.47	0.52	0.47	0.06	0.78
LDEG	13.91	13.26	21194	0.91	0.49	0.48	0.05	0.95
EDEG	37.33	31.72	20301.5	0.40	0.52	0.47	0.07	0.66
SHI	7.85	8.1	20275.5	0.39	0.44	0.49	0.11	0.14
WD	77.43	62.84	20805	0.67	0.51	0.49	0.06	0.88
WD-LOCAL	39.46	34.07	20886.5	0.72	0.5	0.48	0.05	0.86
WD-EXTERNAL	37.97	28.77	20336.5	0.41	0.52	0.47	0.09	0.33
WD-INTRAD	37.3	31.66	20668	0.59	0.51	0.48	0.06	0.87
WD-INTERD	2.15	2.41	19739	0.17	0.35	0.42	0.09	0.34
BET	579.38	389.05	18564.5	**0.02**	0.55	0.42	0.14	**0.04**
CLO	0.30	0.31	18666	0.23	0.46	0.53	0.12	0.10
CC	0.49	0.55	18545	**0.02**	0.42	0.55	0.13	**0.04**
CGD	0.39	0.41	19004	0.35	0.45	0.51	0.12	0.10

Table 4. The values of the Spearman correlation coefficients between productivity and collaboration metrics.

	Male researchers				Female researchers			
	PRO-N	PRO-F	PRO-S	CI	PRO-N	PRO-F	PRO-S	CI
DEG	0.86	0.75	0.68	0.79	0.87	0.74	0.74	0.82
LDEG	0.66	0.54	0.48	0.59	0.7	0.55	0.56	0.64
EDEG	0.88	0.78	0.72	0.82	0.88	0.77	0.76	0.82

macro scale – they more often connect different research groups, but

- UNS-PMF female researchers tend to be more important for the cohesion of the institution at the micro scale – they more often connect researchers from the same research group which previously have not collaborated.

UNS-PMF consists of 5 departments corresponding to different scientific disciplines (see Table 2). GERBER performs non-parametric statistical tests to detect significant differences between male and female researchers within departments. Table 5 shows UNS-PMF departments and researcher metrics for which the null hypothesis of at least one of implemented non-parametric statistical test is rejected. It can be noticed that there are no statistically significant gender differences at the Department of Mathematics and Informatics and the Department of Biology and Ecology for each of considered researcher evaluation metrics. Consequently, we can conclude that there are no significant gender gaps regarding research productivity and collaboration at those two departments.

Table 5. Statistically significant gender differences per UNS-PMF departments.

Department	Metric	$\langle M \rangle$	$\langle F \rangle$	U	MWU-p	PS_m	PS_f	D	KS-p
Geography	DEG	58.45	36.18	361	0.03	0.65	0.33	0.35	0.03
Geography	EDEG	38.39	19.96	369	0.03	0.64	0.34	0.28	0.11
Geography	WD-INTERD	0.71	0.48	379	0.03	0.49	0.21	0.31	0.06
Geography	BET	523.52	157.82	336	0.01	0.68	0.31	0.31	0.06
Geography	CLO	0.32	0.29	242.5	0.001	0.74	0.26	0.44	0.004
Physics	PRO-N	129.66	60.64	295	0.09	0.62	0.36	0.39	0.02
Physics	WD	127	58.04	286	0.06	0.64	0.355	0.39	0.01
Physics	WD-LOCAL	56.75	32.27	289	0.07	0.63	0.36	0.35	0.04
Physics	WD-EXTERNAL	70.25	25.77	282	0.05	0.645	0.35	0.37	0.03
Physics	CGD	0.42	0.28	240	0.02	0.65	0.3	0.31	0.11
Chemistry	CGD	0.33	0.46	555.5	0.02	0.32	0.65	0.32	0.04

Two UNS-PMF departments possessing a high level of gender imbalance are the Department of Geography and the Department of Physics. We can see that male researchers from the Department of Geography tend to have drastically higher number of external and total collaborators compared to female researchers. Additionally, male UNS-PMF geographers established significantly stronger intra-department research collaboration amongst themselves than females. This implies that males have a stronger impact to the cohesiveness of the department than females. Female UNS-PMF geographers exhibit drastically lower centrality in the co-authorship network implying that their male colleagues are more important for the cohesiveness of the department. From the results presented in Table 5 we can also observe that male researchers from the Department of Physics and female researchers from the Department of Chemistry have significantly higher co-author gender disbalance compared to their departmental colleagues of the opposite gender. Male researchers from the Department of Physics tend to have significantly higher research productivity than female researchers when research productivity is measured by the normal counting scheme. On the other hand, gender disparity in research productivity is absent when productivity is estimated by other three productivity metrics indicating that the normal counting scheme is a gender-biased research productivity metric. The normal counting scheme assigns equal credit to each author of a paper ignoring the total number of authors.

Therefore, we can conclude that male UNS-PMF physicists tend to produce papers which on average have a larger number of authors than papers produced by female UNS-PMF physicists. Additionally, male UNS-PMF physicists established significantly stronger intra-institutional and inter-institutional research collaboration implying that they are more willing to collaborate with researchers from other UNS-PMF departments and researchers from other institutions compared to their departmental female colleagues.

The UNS-PMF co-authorship network contains 548 links (19.17% of the total number of links) connecting male researchers (MM links), 1098 links (38.4%) connecting female researchers (FF links) and 1213 links (42.43%) representing collaborations between male and female researchers (MF links). The GERBER tool tests for statistically significant differences in the weight of MM, FF and MF links at different granularity levels. The obtained results for the UNS-PMF co-authorship network are shown in Table 6. It can be seen that statistically significant gender differences are not present at the level of the whole institution. Additionally, statistically significant differences in the weight of MM, FF and MF links are not present considering all inter-department and all inter-department links separately. However, statistically significant gender differences appear at the level of individual departments. The strength of research collaboration between male UNS-PMF physicists tends to be significantly higher than the strength of research collaboration between female UNS-PMF physicists and the strength of research collaboration between UNS-PMF physicists of opposite genders. In other UNS-PMF departments statistically significant differences in the weight of MM, FF and MF intra-department links are absent. Therefore, we can conclude that the Department of Physics is the only UNS-PMF department which exhibits gender homophily in research collaboration.

6. Integration of Analytic Tools with CRIS UNS

The main subject in the future work in this area is an integration of GERBER and similar analytic tools with CRIS UNS that facilitates continuous evaluation of researchers. This can be done by applying a service oriented architecture in which CRIS UNS exposes services to the analytics tools. These services will include operations for obtaining relevant data for analytic tools.

Instead of dividing the analysis in two independent stages, (1) export of all relevant data to XML documents and (2) loading data from these documents, the better solution is to load data directly from the CRIS UNS database through appropriate services. Some benefits of the proposed approach are:

- retrieving the updated states of the entities (as CRIS UNS is in constant use by authors who enter their publications, the latest state can be obtained only by real-time access),
- obtaining only data that are relevant for the analysis, and omitting redundant elements that will influence memory usage,
- exposing the complete CRIS UNS metadata set for analysis, including the attributes that are not recognized as relevant in time of export in the current solution.

As for the concrete technology and implementation of the services there are basically two solutions: implementing the web service with WSDL and SOAP and exchanging XML documents, or

Table 6. Statistical comparison of MM, FF and MF link weights in the UNS-PMF co-authorship network. G_1 denotes the first group, while G_2 denotes the second group of links. PS_1 is the probability of superiority of G_1 over G_2, while PS_2 denotes the opposite probability of superiority. Bold p values indicate statistically significant differences.

	G_1	G_2	$\langle G_1 \rangle$	$\langle G_2 \rangle$	U	MWU-p	PS_1	PS_2	D	KS-p
All links	MM	FF	2.87	2.44	291232.5	0.29	0.48	0.51	0.06	0.2
	MM	MF	2.87	2.8	330556	0.85	0.49	0.49	0.05	0.39
	FF	MF	2.44	2.8	642627	0.14	0.51	0.47	0.05	0.07
All	MM	FF	0.74	0.86	8860	0.29	0.44	0.52	0.1	0.48
inter-department links	MM	MF	0.74	0.85	11153	0.66	0.49	0.46	0.07	0.84
	FF	MF	0.86	0.85	27417	0.04	0.53	0.43	0.09	0.19
All	MM	FF	3.26	2.85	192337	0.15	0.47	0.52	0.08	0.36
intra-department links	MM	MF	3.26	3.36	211322.5	0.35	0.48	0.51	0.08	0.34
	FF	MF	2.85	3.36	405048.5	0.61	0.5	0.49	0.04	0.37
Mathematics	MM	FF	4.06	2.61	1200.5	0.81	0.46	0.49	0.09	0.96
intra-department links	MM	MF	4.06	5.75	2969	0.68	0.45	0.49	0.11	0.67
	FF	MF	2.61	5.75	1785	0.92	0.46	0.48	0.13	0.65
Geography	MM	FF	1.38	1.66	10354.5	0.28	0.44	0.52	0.11	0.33
intra-department links	MM	MF	1.38	1.48	25431.5	0.43	0.5	0.46	0.08	0.39
	FF	MF	1.66	1.48	10318	0.1	0.54	0.42	0.14	0.11
Biology	MM	FF	2.93	3.11	7894	0.2	0.43	0.54	0.12	0.47
intra-department links	MM	MF	2.93	3.41	6522.5	0.24	0.43	0.54	0.1	0.69
	FF	MF	3.11	3.41	45680.5	0.97	0.49	0.49	0.05	0.67
Physics	MM	FF	8.15	2.64	1613.5	**0.03**	0.6	0.39	0.23	**0.04**
intra-department links	MM	MF	8.15	4.21	3124	**0.04**	0.58	0.41	0.21	**0.04**
	FF	MF	2.64	4.21	2711.5	0.41	0.45	0.52	0.11	0.72
Chemistry	MM	FF	3.59	2.98	6537.5	0.26	0.43	0.54	0.17	0.18
intra-department links	MM	MF	3.59	3.82	4227	0.24	0.43	0.54	0.18	0.17
	FF	MF	2.98	3.82	38455	0.59	0.48	0.5	0.06	0.61

using the REST architectural style and exchanging either XML documents or JSON objects. Although there are some advantages of SOAP-based services in terms of tools support and type safety, REST services become very popular these days mostly due to ease of implementation based on the HTTP protocol.

The suggested architecture is shown in Figure 2. GERBER (or some other analytic tool) accesses CRIS UNS data through the REST API providing at least the following operations: (1) retrieve all researchers for the given institution, (2) retrieve all publications for the given researcher, and (3) retrieve all publications entered after the given date. The third operation will improve the efficiency of an analytic tool such as GERBER because instead of loading all publications, it can load only those entered after the date of last access.

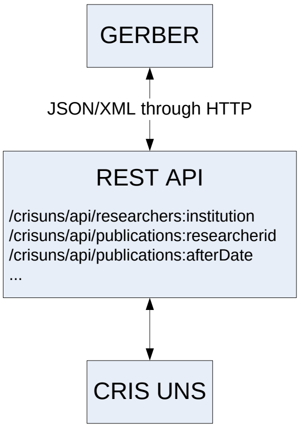

Figure 2. GEBER and CRIS UNS integration through the REST API.

The architecture presented in Figure 2 is applicable for any analytic tool and any research information system. In other words, if we define the complete set of operations for the REST API, and extend GERBER to load data through that API, any other research information system that implements the API can use GERBER for gender-based analysis.

7. Conclusions and Future Work

As the main contribution of the paper we described the methodology and accompanying tool (GERBER) for gender-based analysis of intra-institutional research productivity and collaboration. The methodology is based on the application of non-parametric statistical tests to weighted co-authorship networks with an underlying institutional structure whose nodes have gender labels and associated metric vectors containing metrics reflecting productivity, collaboration and institutional importance of researchers. Additionally, the tests are conducted on different levels of granularity in order to detect gender inequalities and homophily at the institutional level and the level of research departments within institution.

Using GERBER we performed gender-based analysis of our faculty (UNS-PMF) using data exported from the institutional CRIS. The obtained results showed that there are no significant gender differences at the institutional level considering productivity and collaboration of UNS-PMF researchers. On the other hand, gender differences can be observed with respect to the role researchers have in institutional cohesion – male UNS-PMF researchers tend to be more important for collaboration between different research groups, while female UNS-PMF researchers tend to be more important for collaboration within research groups. At the level of research departments situation is slightly different – there are two (out of five) UNS-PMF research departments exhibiting significant gender inequalities in research collaboration. Additionally, for one of those two departments we detected gender homophily in research collaboration.

In our future work we plan to improve GERBER by including other culture-sensitive attributes such as age and academic rank in gender-based analysis of research performance. We will also investigate possibilities to enrich nodes of co-authorship networks extracted from institutional bibliographic databases with metrics reflecting the importance of researchers at the international level. To achieve this goal, GERBER has to be able to extract, retrieve or fuse field co-authorship networks having a broad coverage of individual scientific disciplines from publicly available bibliographic databases and research networking platforms [46], locate local researchers within such networks and compute centrality metrics for corresponding nodes. Additionally, our aim is to integrate GERBER into our institutional CRIS system (CRIS UNS) as an analytic service. Such integration will enable continuous gender-based analysis of researchers employed at our university and the evaluation of policies and actions conducted to reduce gender gaps in research performance. On the basis of previously mentioned integration we will also be in position to propose general methodological and technical guidelines (APIs) for culture-sensitive extensions of CRIS systems.

Acknowledgments. The authors thank the Ministry of Education, Science and Technological Development of the Republic of Serbia for support through project no. OI174023, "Intelligent techniques and their integration into wide-spectrum decision support," and for additional support in cooperation with the Slovenian Research Agency through bilateral project no. 451-03-3095/2014-09/43, "Culture-sensitive aspects in data technologies."

References

[1] Gander JP. Faculty gender effects on academic research and teaching. *Research in Higher Education*, 1999. **40**(2):171–184.

[2] Bayer AE. Teaching Faculty in Academe: 1972–73. ACE Research Reports 8, American Council on Education, Washington, DC, USA, 1973.

[3] Cole JR. Fair Science: Women in the Scientific Community. The Free Press, New York, USA, 1979.

[4] Xie Y, Shauman KA. Sex differences in research productivity: New evidence about an old puzzle. *American Sociological Review*, 1998. pp. 847–870.

[5] Abramo G, D'Angelo CA, Murgia G. Gender differences in research collaboration. *Journal of Informetrics*, 2013. **7**(4):811–822. doi:10.1016/j.joi.2013.07.002.

[6] Van Leeuwen TN, Van Wijk E, Wouters PF. Bibliometric analysis of output and impact based on CRIS data: a case study on the registered output of a Dutch university. *Scientometrics*, 2016. **106**(1):1–16. doi:10.1007/s11192-015-1788-y.

[7] Perc M. Growth and structure of Slovenia's scientific collaboration network. *Journal of Informetrics*, 2010. **4**(4):475–482. doi:10.1016/j.joi.2010.04.003.

[8] Kastrin A, Klisara J, Lužar B, Povh J. Analysis of Slovenian research community through bibliographic networks. *Scientometrics*, 2017. **110**(2):791–813. doi:10.1007/s11192-016-2203-z.

[9] Savić M, Ivanović M, Dimić Surla B. Analysis of intra-institutional research collaboration: a case of a Serbian faculty of sciences. *Scientometrics*, 2017. **110**(1):195–216. doi:10.1007/s11192-016-2167-z.

[10] euroCRIS. euroCRIS Current Resesarch Information Systems. URL `http://www.eurocris.org/`.

[11] van Arensbergen P, van der Weijden I, van den Besselaar P. Gender differences in scientific productivity: A persisting phenomenon? *Scientometrics*, 2012. **93**(3):857–868.

[12] Fong CJ, Yoo JH, Jones SJ, Torres LG, Decker ML. Trends in female authorships, editorial board memberships, and editorships in educational psychology journals from 2003 to 2008. *Educational Psychology Review*, 2009. **21**(3):267–277.

[13] König CJ, Fell CB, Kellnhofer L, Schui G. Are there gender differences among researchers from industrial/organizational psychology? *Scientometrics*, 2015. **105**(3):1931–1952.

[14] Bordons M, Morillo F, Fernández MT, Gómez I. One step further in the production of bibliometric indicators at the micro level: Differences by gender and professional category of scientists. *Scientometrics*, 2003. **57**(2):159–173.

[15] Sotudeh H, Khoshian N. Gender differences in science: The case of scientific productivity in Nano Science & Technology during 2005–2007. *Scientometrics*, 2014. **98**(1):457–472.

[16] Cavero JM, Vela B, Cáceres P, Cuesta C, Sierra-Alonso A. The evolution of female authorship in computing research. *Scientometrics*, 2015. **103**(1):85–100.

[17] Vela B, Cáceres P, Cavero JM. Participation of women in software engineering publications. *Scientometrics*, 2012. **93**(3):661–679.

[18] Prpić K. Gender and productivity differentials in science. *Scientometrics*, 2002. **55**(1):27–58.

[19] Abramo G, D'Angelo CA, Caprasecca A. Gender differences in research productivity: A bibliometric analysis of the Italian academic system. *Scientometrics*, 2009. **79**(3):517–539.

[20] Paul-Hus A, Bouvier RL, Ni C, Sugimoto CR, Pislyakov V, Larivière V. Forty years of gender disparities in Russian science: A historical bibliometric analysis. *Scientometrics*, 2015. **102**(2):1541–1553.

[21] Ozel B, Kretschmer H, Kretschmer T. Co-authorship pair distribution patterns by gender. *Scientometrics*, 2014. **98**(1):703–723.

[22] Borrego A, Barrios M, Villarroya A, Ollé C. Scientific output and impact of postdoctoral scientists: A gender perspective. *Scientometrics*, 2010. **83**(1):93–101.

[23] Dehdarirad T, Villarroya A, Barrios M. Research on women in science and higher education: A bibliometric analysis. *Scientometrics*, 2015. **103**(3):795–812.

[24] Bozeman B, Gaughan M. How do men and women differ in research collaborations? An analysis of the collaborative motives and strategies of academic researchers. *Research Policy*, 2011. **40**(10):1393–1402. doi:10.1016/j.respol.2011.07.002.

[25] Zeng XHT, Duch J, Sales-Pardo M, Moreira JAG, Radicchi F, Ribeiro HV, Woodruff TK, Amaral LAN. Differences in Collaboration Patterns across Discipline, Career Stage, and Gender. *PLOS Biology*, 2016. **14**(11):e1002573. doi:10.1371/journal.pbio.1002573.

[26] Ivanovic D, Surla D, Konjovic Z. CERIF compatible data model based on MARC 21 format. *The Electronic Library*, 2011. **29**(1):52–70.

[27] euroCRIS. CERIF 1.3 Full Data Model (FDM) Introduction and Specification, 2012. URL `http://eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_FDM.pdf`.

[28] The Library of Congress. MARC Standards. URL `http://www.loc.gov/marc/`.

[29] Milosavljevic G, Ivanovic D, Surla D, Milosavljevic B. Automated construction of the user interface for a CERIF-compliant research management system. *The Electronic Library*, 2011. **29**(5):565–588.

[30] Ivanović D, Surla D, Racković M. Journal evaluation based on bibliometric indicators and the CERIF data model. *Computer Science and Information Systems*, 2012. **9**(2):791–811.

[31] Nikolić S, Penca V, Ivanovića D, Surla D, Konjovića Z. CRIS service for journals and journal articles evaluation. *11th International Conference on Current Research Information Systems, Prague, Czech Republic*, 2012. pp. 323–332.

[32] Dimić Surla B, Ivanović D. Using templates for presenting publication references in CRIS. In: Proceedings of the CRIS 2012 Conference. euroCRIS, 2012 pp. 61–66.

[33] Ivanović D, Ivanović L, Dimić Surla B. Multi-interoperable CRIS Repository. *Procedia Computer Science*, 2014. **33**:86–91.

[34] Dimić Surla B, Segedinac M, Ivanović D. A BIBO ontology extension for evaluation of scientific research results. In: Proceedings of the Fifth Balkan Conference in Informatics. ACM, 2012 pp. 275–278.

[35] Ivanović L, Dimić Surla B, Segedinac M, Ivanović D. CRISUNS ontology for theses and dissertations. *Proceedings of the ICIST 2012 Conference (CD), Kopaonik, February 29*, 2012. pp. 164–169.

[36] Newman MEJ. Who Is the Best Connected Scientist? A Study of Scientific Coauthorship Networks. In: Ben-Naim E, Frauenfelder H, Toroczkai Z (eds.), Complex Networks, volume 650 of *Lecture Notes in Physics*, pp. 337–370. Springer Berlin Heidelberg. ISBN 978-3-540-22354-2, 2004. doi:10.1007/978-3-540-44485-5_16.

[37] Lindsey D. Production and Citation Measures in the Sociology of Science: The Problem of Multiple Authorship. *Social Studies of Science*, 1980. **10**(2):145–162.

[38] Seidman SB. Network structure and minimum degree. *Social Networks*, 1983. **5**(3):269–287. doi: http://dx.doi.org/10.1016/0378-8733(83)90028-X. URL `http://www.sciencedirect.com/science/article/pii/037887338390028X`.

[39] Freeman LC. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 1977. **40**:35–41.

[40] Kósa B, Balassi M, Englert P, Kiss A. Betweenness versus Linerank. *Computer Science and Information Systems*, 2015. **12**(1):33–48. doi:10.2298/CSIS141101092K.

[41] Bavelas A. Communication patterns in Task-Oriented Groups. *The Journal of the Acoustical Society of America*, 1950. **22**(6):725–730. doi:10.1121/1.1906679.

[42] Watts DJ, Strogatz SH. Collective dynamics of "small-world" networks. *Nature*, 1998. **393**:440–442.

[43] Mann HB, Whitney DR. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 1947. **18**(1):50–60. doi:10.2307/2236101.

[44] Feller W. On the Kolmogorov-Smirnov limit theorems for empirical distributions. *The Annals of Mathematical Statistics*, 1948. **19**(2):177–189.

[45] Erceg-Hurn DM, Mirosevich VM. Modern robust statistical methods: an easy way to maximize the accuracy and power of your research. *The American Psychologist*, 2008. **63**(7):591–601. doi:10.1037/0003-066X.63.7.591.

[46] De Vocht L, Softic S, Verborgh R, Mannens E, Ebner M. ResXplorer: Revealing Relations between Resources for Researchers in the Web of Data. *Computer Science and Information Systems*, 2017. **14**(1):25–50. doi:10.2298/CSIS151028031D.