The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/2514-9288.htm

Received 18 January 2021
Revised 5 August 2021
21 December 2021
Accepted 19 January 2022

Social recruiting: an application of social network analysis for preselection of candidates

Stevan Milovanović and Zorica Bogdanović
Faculty of Organisational Sciences, University of Belgrade, Belgrade, Serbia
Aleksandra Labus
Department for E-business, Faculty of Organisational Sciences, University of Belgrade, Belgrade, Serbia
Marijana Despotović-Zrakić
Faculty of Organisational Sciences, University of Belgrade, Belgrade, Serbia, and
Svetlana Mitrović
Project Management College, Belgrade, Serbia

Abstract

Purpose – The paper aims to studiy social recruiting for finding suitable candidates on social networks. The main goal is to develop a methodological approach that would enable preselection of candidates using social network analysis. The research focus is on the automated collection of data using the web scraping method. Based on the information collected from the users' profiles, three clusters of skills and interests are created: technical, empirical and education-based. The identified clusters enable the recruiter to effectively search for suitable candidates.

Design/methodology/approach – This paper proposes a new methodological approach for the preselection of candidates based on social network analysis (SNA). The defined methodological approach includes the following phases: Social network selection according to the defined preselection goals; Automatic data collection from the selected social network using the web scraping method; Filtering, processing and statistical analysis of data. Data analysis to identify relevant information for the preselection of candidates using attributes clustering and SNA. Preselection of candidates is based on the information obtained.

Findings – It is possible to contribute to candidate preselection in the recruiting process by identifying key categories of skills and interests of candidates. Using a defined methodological approach allows recruiters to identify candidates who possess the skills and interests defined by the search. A defined method automates the verification of the existence, or absence, of a particular category of skills or interests on the profiles of the potential candidates. The primary intention is reflected in the screening and filtering of the skills and interests of potential candidates, which contributes to a more effective preselection process.

Research limitations/implications – A small sample of the participants is present in the preliminary evaluation. A manual revision of the collected skills and interests is conducted. The recruiters should have basic knowledge of the SNA methodology in order to understand its application in the described method. The reliability of the collected data is assessed, because users provide data themselves when filling out their social network profiles.

Practical implications – The presented method could be applied on different social networks, such as GitHub or AngelList for clustering profile skills. For a different social network, only the web scraping instructions would change. This method is composed of mutually independent steps. This means that each step can be implemented differently, without changing the whole process. The results of a pilot project evaluation indicate that the HR experts are interested in the proposed method and that they would be willing to include it in their practice.

Social implications – The social implication should be the determination of relevant skills and interests during the preselection phase of candidates in the process of social recruitment.

Originality/value – In contrast to previous studies that were discussed in the paper, this paper defines a method for automatic data collection using the web scraper tool. The described method allows the collection of more data in a shorter period. Additionally, it reduces the cost of creating an initial data set by removing the

This research was funded by the Ministry of education, science and technological development, Republic of Serbia, the Grant number 11143.

Data Technologies and Applications
Vol. 56 No. 4, 2022
pp. 536-557

2514-9288
DOI 10.1108/DTA-01-2021-0021

cost of hiring interviewers, questioners and people who collect data from social networks. A completely automated process of data collection from a particular social network stands out from this model from currently available solutions. Considering the method of data collection implemented in this paper, the proposed method provides opportunities to extend the scope of collected data to implicit data, which is not possible using the tools presented in other papers.

Keywords E-recruitment, Social recruiting, Social network analysis, Web scraping

Paper type Research paper

1. Introduction

In an information-driven society, where knowledge and skills are recognized as an important resource for growth and competitiveness, human resource management (HRM) becomes a business process of strategic importance, both for the design and implementation of corporate strategy, motivation, recruitment and retention of highly qualified candidates (Toteva and Gourova, 2011). In such a defined business environment, social networks sites (SNS) enable connecting candidates and companies.

SNS significantly influence the candidate recruitment process. Due to their high attendance, SNS provide effectiveness to find and analyze potential candidates for the job (Hedenus *et al.*, 2019). Main strengths of SNS are finding candidates with specific knowledge, skills and recommendations. Social recruiting on SNS provides benefits that are reflected in cost reduction, search efficiency of candidates, opportunities to contact passive candidates who are not active in the search for a new job and company brand development (Okolie and Irabor, 2017) (Ramaabaanu and Saranya, 2014). Social network analysis (hereinafter: SNA) can be used for social recruiting to make the selection of potential candidates as efficient as possible (Golovko and Schumann, 2019).

Preselection of candidates in the social recruitment process requires advanced search with candidate filtering. The main disadvantages of this feature are the cost of using specialized tools, predefined and limited functionalities and the use of unknown algorithms during filtering. Previous research, such as (Baruffaldi *et al.*, 2017), used manually collected data sets or, as (Chiang and Suen, 2015), conducted surveys and interviews to collect data. Both time and budget resources are required to implement manual data collection methods. Literature research (Tifferet and Vilnai-Yavetz, 2018) has shown that automated data collection methods can be more efficient and as accurate as manual ones. However, the main disadvantage of automated tools is that they are designed only for a specific platform. No tool is generic enough that you can use it to collect data from a variety of sources.

This paper presents an alternative methodological approach for the preselection of candidates in the social recruiting process by using SNA. Using the SNA methodology, it is possible to analyze the structure of the social network through concepts from graph theory and network analysis, with the help of defined mathematical models and clustering algorithms (Toteva and Gourova, 2011). Based on the definition of SNA methodology (Milovanović *et al.*, 2019), it can be concluded that the greatest contribution of this methodology in the process of recruiting candidates is in the phase of preselection of qualified people for a certain job position.

The main goal of this research is to propose new methodological approach for the preselection of candidates based on SNA. The proposed methodological approach enables automated identification of users' social network profiles according to their main skills and interests. The advantages of this approach are the automation of the data collection process using web scraping method, and preselection of candidate suitable for the job position using SNA. Proposed method for the collecting users' data from SNSreduces the time and cost of the recruitment process and training recruiters to use specialized tools for different SNS. For this research LinkedIn was chosen because this is the most widely used SNS for social recruitment (Tifferet and Vilnai-Yavetz, 2018). The result of applying this approach should enable identification of clusters that would enable the recruiters to effectively search suitable candidates for the job and to serve as an integral part of a wider framework for e-recruitment.

The rest of the paper is organized as follows: section 2 provides a literature review related to SNS, e-recruiting, SNA and automated data collection. Section 3 discusses the methodological approach for preselection of candidates based on SNA. Pilot project is shown in section 4, and Evaluation and analysis of results are provided in section 5. Section 6 covers Discussion and section 7 shows summarized Conclusions.

2. Literature review

2.1 SNS as tools for e-recruiting

Recruitment is one of the HRM processes involved in attracting, selecting and deciding to hire qualified candidates for a particular job position (Jiang *et al.*, 2012). E-recruitment is the use of resources from the Internet to find, select and recruit new candidates for employment (Melanthiou *et al.*, 2015; Girard and Fallery, 2009). By using data from SNSfor candidates selection within the recruitment process, this task becomes more efficient and cost-effective (Han and Han, 2009). The competitiveness of e-recruitment platforms is determined primarily by the effectiveness to find, select and analyze potential candidates for the preselected skillset by using specialized tools and analysis methods.

SNS represent virtual spaces where individuals can create public or semi-public profiles with a detailed description of their skills, interests, education and working experience, articulate a list of other users with whom they are connected, view connections made by others, achieve social interaction, maintaining different type of relationships and build their personal or professional identity (Boyd and Ellison, 2007). From a candidate's perspective, the first step in the process of looking for business opportunities is presenting themselves properly to potential employers via SNS by self-promotion (Tifferet and Vilnai-Yavetz, 2018). Candidates' profiles and connections contribute to their visibility by networking with colleagues, recruiters, companies, or communities with similar skills and interests (Ou *et al.*, 2013). From companies' perspective, SNS provide a suitable tool for e-recruiting in the sense of the following (Sharma, 2014):

(1) Cost-effectiveness. Companies can significantly reduce recruitment costs by using SNS as the platform to advertise open job positions, run low cost digital campaigns and automatically make a preselection of suitable candidates for the job.

(2) Time Efficiency. SNS save both recruiters' and candidates' time and reduces the time elapsed from initial contact with candidates to the final decision for the employment.

(3) Branding. By participating in discussions, advertising and marketing campaigns, companies gain the opportunity to become leaders on market or representatives of particular skills and interests.

(4) Larger candidate base. Companies can attract a large number of qualified candidates by advertising open positions on SNS.

(5) Better quality of candidates. On SNS, companies have access to the most competent experts, thus improving the recruitment process itself.

(6) Reduced administration. Using social media platforms and services reduces the need for unnecessary administrative processes, which further facilitates the recruitment process.

One of the most used SNS for creating users' business profiles, establishing business contacts, connecting with communities of the same skills and business interests, job searching and e-recruitment is LinkedIn (Tifferet and Vilnai-Yavetz, 2018; Castillo de Mesa and Gómez Jacinto, 2020; Van Dijck, 2013). LinkedIn is often chosen as an e-recruiting tool by companies (Hosain and Liu, 2020). According to research (Hosain and Liu, 2020), which included 260,000 recruiters, 97% of respondents said that their daily activities are largely tied to LinkedIn. Two-thirds of respondents using LinkedIn confirmed that campaigns created on this SNS were successful and achieved the created goals.

In the research (Baruffaldi *et al.*, 2017) LinkedIn profiles were analyzed and it was concluded that LinkedIn is a significant tool for e-recruiting candidates with high skill levels and without geographical boundaries.

Research paper (Chiang and Suen, 2015) investigated how user profiles influence the recruiter's recommendations on LinkedIn. Survey results showed that argumentative profile features, such as education and support skills, make the greatest contribution when selecting candidates for particular positions. All these results imply that LinkedIn serves as an efficient recruitment tool recognized by both employees and employers.

The development of e-recruiting via SNS is noticeable by analyzing statistical data (Parks-Yancy and Cooley, 2018):

(1) 94% of recruiters use social networks for recruiting candidates,

(2) Companies that use social networks for recruiting candidates have improved the quality of their candidates by 49%,

(3) Almost three-quarters of employees between the ages of 18 and 34 found their last job using SNS,

(4) 89% of recruiters reported that they hired at least one candidate using LinkedIn,

2.2 SNA

In context of SNS, SNA represents an interdisciplinary methodology for mapping and measuring relationships and flows between people, groups, organizations, computers, URLs and other connected information entities (Milovanović *et al.*, 2019; Wasserman and Faust, 1994). SNA is a special branch of data science that refers to the use of graph theory concepts to study and model social networks (Steketee *et al.*, 2015; Southekal, 2017; Provost and Fawcett, 2013).

Data science and social networks are connected in many different ways (Voytek, 2017). Analyzing data from social networks provides an understanding of the social, scientific and economic phenomena which generate the data. For example, collecting digital tags from social network users is a necessary step to understand, model and predict user behavior (Malhotra *et al.*, 2012). In economic terms, companies collect and analyze data on competitors' activities. This identifies market opportunities, anticipates competitors' moves and learns from other people's experiences. By analyzing data from social networks, companies can improve trend management, which is based on identifying customer interests (Ferrara *et al.*, 2014).

The authors of the paper (El Ouirdi *et al.*, 2016) researched the topic of social recruiting by interviewing recruiters. It was concluded that some of the benefits of social recruiting include a significant reduction in the time and cost of candidate selection, transparency of data on candidate profiles and more effective communication with potential candidates. The next paper (Melanthiou *et al.*, 2015) analyzed the role of SNS in the process of e-recruitment. The authors examined the extent to which the social recruitment process was accepted by companies, as well as whether there were opportunities for this type of recruitment to become the main strategy for recruiting candidates. This research concludes that a well-designed system and strategic use of available information about potential candidates can significantly help in recruiting employees with the most appropriate skills and competencies.

2.3 Automatic data collection for SNA

Data collection is the first step in the process of analyzing data from social networks. Further research flow largely depends on the extent and quality of the data collected, so the methods of data collection have to be clearly and precisely defined. The choice of data collection method depends primarily on the type of collected data. The data can be created for specific research or collected from a specific social network site. The most common method for creating research data is to conduct a survey. The questions in the survey are created based on the defined goals of the survey. On the other hand, the disadvantages of conducting the survey are the time it takes for the survey to be designed, created and conducted, and the problem to find a large number of respondents (Peleshchyshyn and Mastykash, 2017). Due to these drawbacks, the automatic collection of data from SNS is increasingly used. Since it is necessary to create software for extracting data from a social network site, it is crucial to analyze the structure of data on a particular social network site beforehand.

Data from SNS can be collected by the use of the publicly available Application Program Interface (hereinafter: API) of SNS. In this way, network requests that meet certain search criteria can be sent to the social network site. With the introduction of the General Data Protection Regulation (hereinafter: GDPR) (Voigt and Von dem Bussche, 2017) on data protection and privacy, as of May 2018, the rules for using these services have been significantly tightened. Since then, the amount of data that can be accessed through these services has been reduced to data for which users have given explicit consent to the data collector.

GDPR also affects the recruitment process. Sources from which data are collected and how long the data are stored must be clearly defined and transparent to the candidates. During data collection, the candidate must be informed of the purpose for which they will be used. Each recruitment company must have a clearly defined data privacy policy. The policy must specify how the candidate can request the deletion of their information from the company database. This rule forms the backbone of GDPR and is called the "right to be forgotten". Furthermore, GDPR also influences the development of tools for collecting information about potential candidates on recruiting platforms. According to this regulation, it is not allowed to collect information about candidates unless there is a legitimate interest of the company and the need for the collection of the information. Collecting data from social profiles is legal under the GDPR if those profiles are publicly available and it can reasonably be assumed that candidates expect to be contacted. For example, it can be assumed that a public profile on any social network indicates a legitimate expectation of contact (Calopereanu *et al.*, 2019). Due to this assumption, the processing of publicly available data from social network profiles is under GDPR. As mentioned in (Backman and Hedenus, 2019), in recruitment, the contemporary practice of cybervetting–the use of search engines and social media platforms to evaluate jobseekers–highlights how information previously considered to belong to the private realm is now also relevant in the public work realm. This paper also explains the meaning of the concept "relevant information as public information". Recruiters prefer to access online information as public. From their perspective, the relevant information is described as that which is valuable for the recruitment process and the evaluation of the candidate even if it is not easily accessible for the recruiters.

However, when social networks do not provide publicly available data collection services, it is necessary to find alternative ways of collecting data. The first and simplest method is observation (also called manual scraping) (Adamic and Adar, 2005). This method involves manually viewing material from social networks (user profiles, posts, comments, likes) and collecting this information manually. This method is time-consuming and error-prone.

Because of these shortcomings, a new, fully automated way of collecting data directly from SNS has been devised. This method is called web scraping and involves writing programs that automate all observation steps (Pereira and Vanitha, 2015). Once the execution with the given instructions is finished, the collected data can be downloaded in CSV (Comma-separated values) format (Nylen and Wallisch, 2017).

Using the API is more robust, efficient, accurate and informative than scraping the network but also more restrictive since it requires special permissions from the social network site and its users to collect data while specialized programs can collect all publicly available data.

The increase in the number of users on the internet has contributed to the increase in the amount of data relevant to SNA (Liang and Zhu, 2017). The result of this phenomenon is the development of various tools for automatic data collection from the internet such as web scraping. This method of data collection is defined as the process of extracting data from the internet by any method or technique (Pereira and Vanitha, 2015).

The main feature of network scraping is the conversion of data gathered from a website into a set of structured data that can be further analyzed (Zhu *et al.*, 2019). The advantages of this type of data collection over the formal approach to using publicly available services are as follows (Pereira and Vanitha, 2015):

(1) Companies care more about the data they present to customers than the data they deliver through publicly available services. By collecting data directly from websites, it is possible to collect more relevant data.

(2) There is no limit to data collection–when using publicly available services, in most cases the maximum number of requests for a certain period is defined, while very few platforms have a defense system against automatic data collection.

(3) Authentication is not required–it is not possible to track the activities of the data collector during automatic data collection, whereas authentication requires the use of publicly available services.

(4) Data can be collected even though there are no publicly available services.

The authors of (Dewi and Chandra, 2019) state that shortcomings of the web scraping approach of data collecting from SNS are complexity, overload, redundancy and relevance of data.

3. Methodological approach for preselection of candidates based on SNA

The goal of this research is to develop a methodological approach that should improve preselection of candidates using SNA. The questions this research needs to answer are as follows:

(1) What are the candidates' skills and interests that affect the profile's visibility?

(2) How can candidates' skills and interests be classified?

(3) Is there a correlation between the candidate's skills and interests?

(4) Is it possible, by interpreting the candidates' skills and interests and their correlation, to select a candidate for a particular job position?

Developed methodological approach for preselection of candidates based on SNA includes the following phases (Figure 1):

(1) Social network selection. Depending on the job position, recruiters can define requirements for searching suitable candidates based on professional and soft skills. After defining these requirements, recruiters choose adequate social network site for searching users' profiles.

(2) Data collection. In this phase, recruiter should use the scraping method for automatic collection data from chosen social network. As stated in section 2.3, this method of processing publicly available data from SNS profiles must be under the GDPR (Calopereanu *et al.*, 2019). It is also possible to use other methods of collecting data, if a specific social network site offers such features. However, scraping has been proposed as a generic approach that can be applied to any social network website.

(3) Data processing. After collecting available information from the SNS users' profiles, the next step is filtering, processing and defining a descriptive data set.

(4) Data analysis. After processing the data, the next step is to determine the optimal number of clusters. The proposed method for this step is the Elbow method that aims to determine the optimal number of clusters of a given data set depending on the consistency of the clusters created using a particular clustering algorithm (Bholowalia and Kumar, 2014). Cluster consistency is calculated as a percentage of variance relative to the number of clusters (Syakur *et al.*, 2018). After clusters are defined, SNA can be applied using suitable SNA tools.

(5) Preselection of candidates. Analysis of the results should provide data of users' profiles that possess skills and interests which are defined in scope of the preselection requirements for the job position.

For using the proposed methodological approach, it is necessary for recruiters to have knowledge related to SNA methodology and statistical analysis, while knowledge of other HR analytic methods can be useful as well. Recruiters would need help from IT specialists in the data collection step, because the implementation of the scraping method requires programming skills.

4. Pilot project

4.1 Social network selection

The proposed methodological approach was applied within the pilot project of the preselection of candidates from the Information technologies sector in Serbia. The first step of the pilot project was to select the most appropriate social network site by the defined research goals for the candidates preselection.

The social network site LinkedIn was selected for the following reasons:

(1) At the time when the pilot project was conducted, LinkedIn Terms & Conditions were not against data scraping (This has been changed with the LinkedIn T&C update from April 8.2021).

(2) User profiles contain more relevant data that can be used for the project purposes, such as professional and soft skills, the number of people who have confirmed that a

Figure 1. Methodological approach for preselection of candidates based on SNA

user possesses particular skills, interests, education, current profession and company where the user is employed and many others.

(3) The data is strictly structured by sections and each profile contains the same sections.

(4) This social network site is intended for business users, which adds to the seriousness and responsibility of users when creating profiles. Therefore, the information on the profiles is easily verifiable and the data trustworthiness is higher than on other SNS (Gërxhani *et al.*, 2013).

(5) LinkedIn has over 660 million users from over 200 countries worldwide (LinkedIn Inc, 2020).

4.2 Data collection

With respect to GDPR, the proposed methodological approach can be used if the data is public (Calopereanu *et al.*, 2019). Therefore, it was decided to use a social network site that is suitable for automatic data collection using specialized programs, web scrapers.

The web scraper tool (WebScraper, 2020) was used as a method for automatic data collection. After managing initial configuration and defining the seed as a URL from where the crawling will start, navigation steps were defined (scrolling, clicking, selecting, etc.). Created selector tree was used to define how the program will be navigated. The last step was to define which data will be scraped and that was defined by graph leaves. After defining the selector graph, the resulting structure looks like in Figure 2.

In order to collect all the necessary data from the LinkedIn users' profiles, the next step is to run an automatic data collection program with a defined scraping plan. Execution of the program stops when all possible navigation steps are completed and all available data is collected from the searched pages. Once the data collection is complete, the data can be downloaded in CSV format. After the successful execution of the web scraper program, the collected data is organized into the following 9 columns:

(1) Name and surname – this column is used as the record identifier, based on the values in this column, the values of all other columns can be verified (this column is omitted to anonymize the data).

(2) Profession – defines the profession of the user (CEO, faculty assistant, manager, software engineer and more).

(3) The number of connections – defines how many connections a particular user has on a LinkedIn social network site.

(4) Current employment – defines where a particular user is currently employed (In the data set sample table, this column is named "Employment").

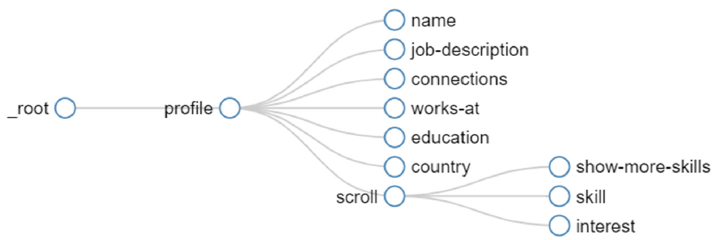

Figure 2.
Selector graph

(5) Education – defines users' education.

(6) Country – defines which country the particular user is from. (This column is omitted from the data set sample table).

(7) Interest – defines a particular user interest.

(8) Skill – defines a particular skill of the user.

(9) Endorsement – describes how many people have confirmed that a user has a particular skill.

The original data set consists of 5,641 records, each record containing the skill or interest of a particular user. For better readability, a sample of data without columns "number of connections" and "country" and merged records of columns "interest" and "skill" are shown in the table data set sample. Given that all the faculties listed in the "education" column are from the University of Belgrade, this information is omitted for readability purposes (see Table 1).

After data collection and observation of the resulting data set, the next step is to filter the initial data. For a record from a set of collected data to be considered valid, it must have a user name, so that the data collected can be verified, and skill or interest, to be useable in further research. Furthermore, users who are not from Serbia will not be considered as the survey is focused exclusively on users of this country. Additionally, this data set can be used with different research aims. Another possible application of the collected data set is classification of acquired occupations based on acquired education and previous experience.

Profession	Employment	Education	Interest	Skill	Endorsement
Tech recruiter	Fenix human resources	University of Belgrade	Online Gambling and Gaming professionals	Human resources	301
Android developer	Just Raspberry		Connect IT	Business strategy	246
Marketing manager	Belgrade Banging		Romanian Startup Academy	Event management	235
Software developer	3AP AG	Faculty of organizational sciences	Endava	Communication	8
Program manager	Microsoft	Faculty of Electrical engineering	Bill Gates	Computer science	8
CTO	Between	Faculty of organizational sciences	Factors Chain International	Continuous Integration	8
CEO	Eric–Rotaract Europe	Faculty of Electrical engineering	James Altucher	Crisis management	8
Teaching assistant	Faculty of organizational sciences	Faculty of organizational sciences	Blockchain	Databases	8

Table 1.
Data set sample

4.3 Data processing

The collected data in CSV format was imported into the Microsoft Excel program and by using filtering functionality all records that did not meet the above conditions were removed. The condition that the record is from the user from Serbia meets 4,847 records. The condition that the record has either skill or interest meets 4,827 records. This confirms the quality of the web scraper tool and defined data collection plan. The analysis of the collected data is based on the users' skills and interests sections as the most appropriate data for clustering.

For initial clustering and statistical processing of gathered data, scripts written in Python programming language (Flowers, 2019) were used. Scripts include functionalities such as counting and grouping the skills and interests of the users, standardization, determination of the optimal number of clusters using the K-means clustering method and displaying with two-dimensional graphs. This phase could be manual as well but we wanted to automate it in order to further improve method efficiency. Yet, we are aware that for this kind of implementation, specific knowledge is required and that is why we mark automatization of this step as optional. Being manual, this phase of the process takes the most time but contributes to the accuracy of the results. Grouping was performed on attribute type level. Since the value distribution was not close to the Gaussian distribution, we empirically chose the threshold in a way that minimizes the misclassified attributes, by excluding 5% of skills and interests with the least number of occurrences. In our case, to analyze the obtained values, the pandas library function called *describe* was used. This function generates a statistical report that includes sample size, mean, standard deviation, quartiles and the minimum and maximum sample values (McKinney and Team, 2015). The statistical report of the newly created data set is shown in Table 2.

Statistical report (Table 2) indicates a significant difference between the values of the skills and values of interests. Statistical analysis of the data after standardization is shown in Table 3.

Distribution of values after standardization of data reflects the values in Table 3.

After processing the data, the next step is to determine the optimal number of clusters. The Elbow method is selected for this step using the K means algorithm. The K means clustering algorithm is considered as a type of unsupervised learning (Raval and Jani, 2016). As a result,

	Skills	Interests
Sample size	62.000	62.000
Mean	36.065	5.984
Standard deviation	36.678	4.194
min	10.000	3.0000
25%	14.250	3.000
50%	21.500	4.000
75%	39.750	6.750
max	181.000	22.000

Table 2. Statistical analysis of grouped skills and interests

	Skills	Interests
Sample size	62.000	62.000
Mean	0.000	0.000
Standard deviation	1.008	1.008
min	−0.725	−0.726
25%	−0.588	−0.726
50%	−0.397	−0.487
75%	0.096	0.232
max	3.957	3.823

Table 3. Statistical analysis of grouped skills and interests after standardization of data

the data are grouped based on the similarity of the given functions (Garbade, 2018). Using the K Means class *fit* function, from the *scikit-learn* library, which receives standardized data as an input parameter (Pedregosa *et al.*, 2011), a graph of the Elbow method was created for values of variable K from 1 to 10 (Figure 3).

Based on the graph shown, it can be concluded that the optimal number of clusters is three because this value creates a curvature of the line of the graph in the form of an elbow, which means that the ratio of the number of clusters and their variance is optimal.

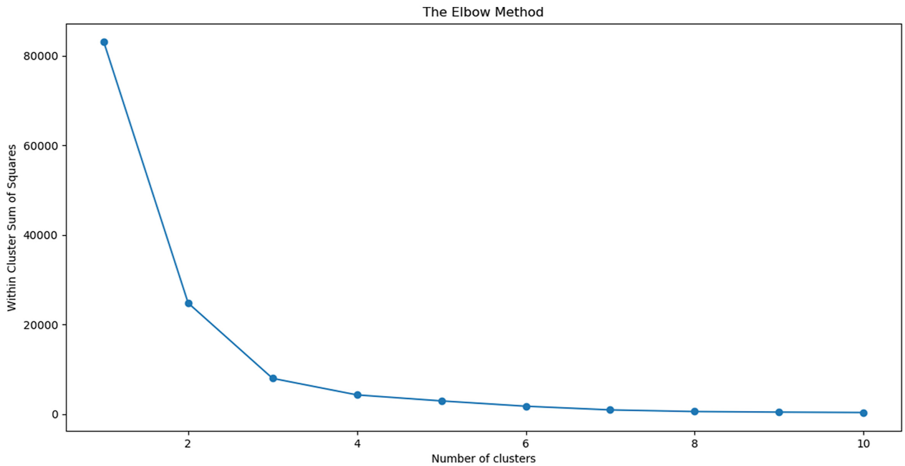

Figure 3.
Elbow method for determining the optimum number of clusters using the K means algorithm

4.4 Data analysis

After determining the optimal number of clusters, a network of the most frequently used terms was created. Since visual network representations produce superior problem-solving outcomes than text (Foucault Welles and Xu, 2018), data from the skill, interest and current employment columns were imported into the VOSviewer tool for SNA. VOSviewer is a free software tool that uses suitable mapping techniques that cluster outputs in large bibliometric networks (Veloutsou and Ruiz Mafe, 2020). For this step we used text mining and visualization feature (Van Eck and Waltman, 2011). In our case, we used it to cluster resulting terms. The resulting network of imported terms should define the relationships between the collected data. When defining the parameters for creating the network, data processing results were used. The created network is shown in Figure 4.

From a social network site perspective, small groups appear as network clusters—subsets of nodes in a social network that have a high density within (group cohesion) and are sparsely connected between (group boundaries) (Stadtfeld *et al.*, 2020). Three clusters can be observed on the shown network:

(1) The first cluster covers terms related to user education. Representatives of this group of terms are “organizational science”, “Belgrade”, “business”, “assistant”, “university”, “Faculty of Organizational Sciences” (abbreviation in Serbian is “FON”) and Student Informatics Association named “FONIS” (Figure 5). The terms presented here group candidates who are in some way affiliated with presented educational organizations. Connections between these terms indicate that for a significant number of candidates, a certain group of terms from a grouped set occurs together. Belgrade has the biggest and the oldest University in Serbia and that is why it is recognized as the most prestigious university center in our country. To further

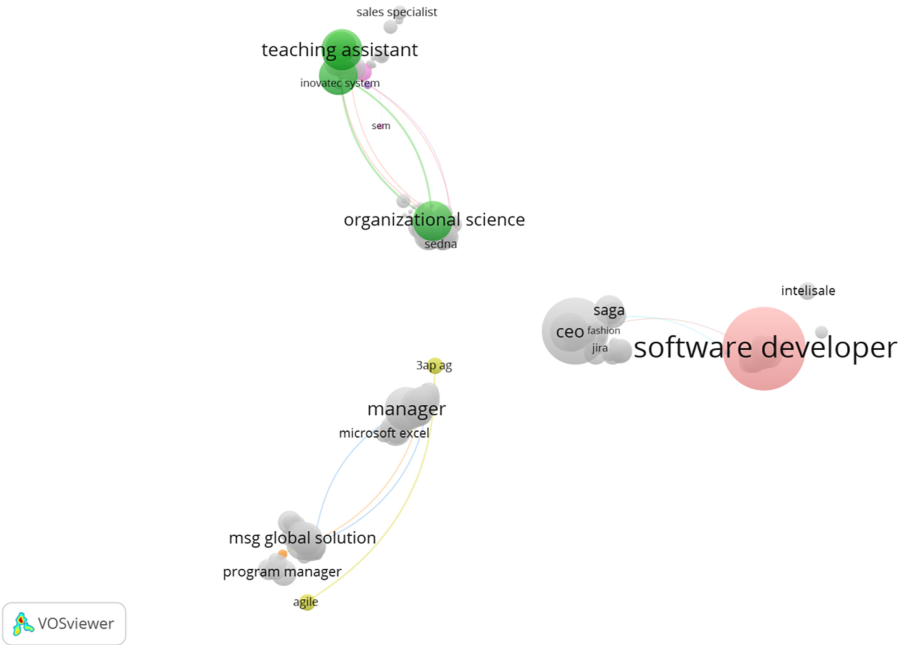

Figure 4.
Network of most common terms and their relations

validate this cluster we performed a manual check on a certain number of the scraped profiles and we discovered that most of the positions in the education are defined by mentioning university. For example, for the position of teaching assistant, the position on LinkedIn is stated as *"Teaching assistant at the University of Belgrade"*. For example, in the profile of users who have graduated from the Faculty of Organizational Sciences, the terms "Belgrade" or "university" appear in the education section, while the terms "organizational science" or "FONIS" (Student organization) appear in the interest section. This group of terms indicates the connection between education and current employment, as well as with the skills and interests of candidates (Figure 5).

(2) The second cluster covers terms related to users' technical skills. The most important terms of this cluster are "software engineer", "java", "web", "Zuhlke Group", "Saga" and "SQL" (Figure 6). This group of terms defines the candidate's technical skills such as programming languages and different technologies or frameworks. The connection between these terms indicates the most common combinations of technical skills in a given number of candidates or the frequent candidate skills of a particular company. Here we were also led by the construction of the position definition on the LinkedIn network. Current position on LinkedIn profiles is defined as *"name of the position" at "name of the company"*. We used that fact to come to the conclusion that, for example, this cluster was created from profiles as *"Java developer at Zuhlke Group"*. That is how we determined that the vast majority of candidates with the term "software engineer" as part of the name of their current job position and "Zuhlke Group" as their current employer cited the java and SQL programming languages as some of their skills. The interests of this group of users are aimed at software development companies, such as Zuhlke Group or Saga, or to specific tech stack such as for example Java.

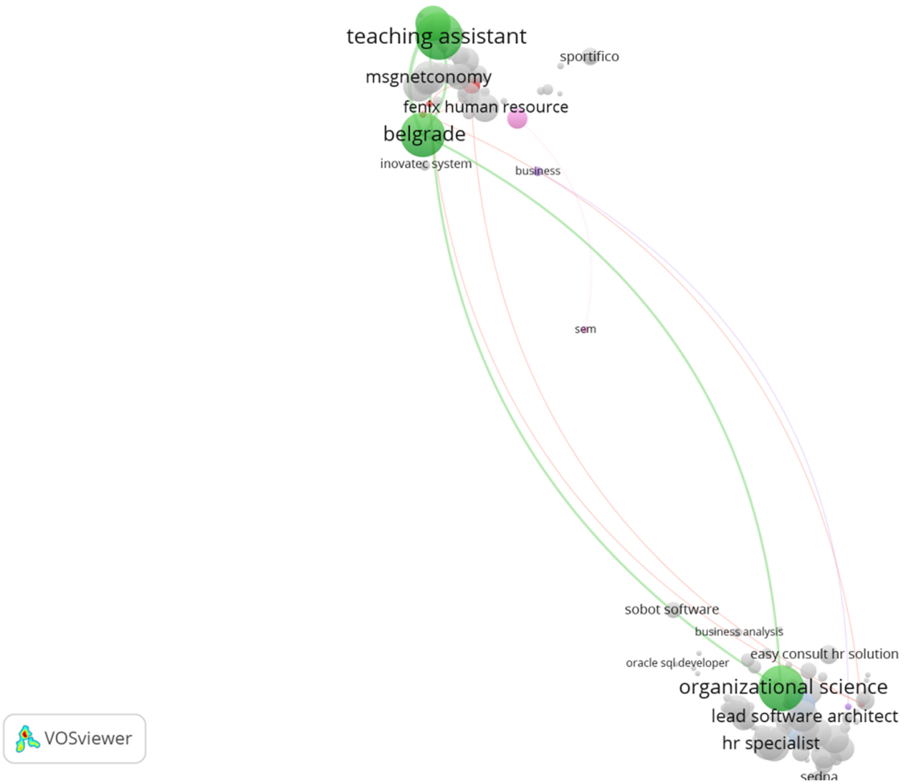

Figure 5.
A cluster of concepts related to user education

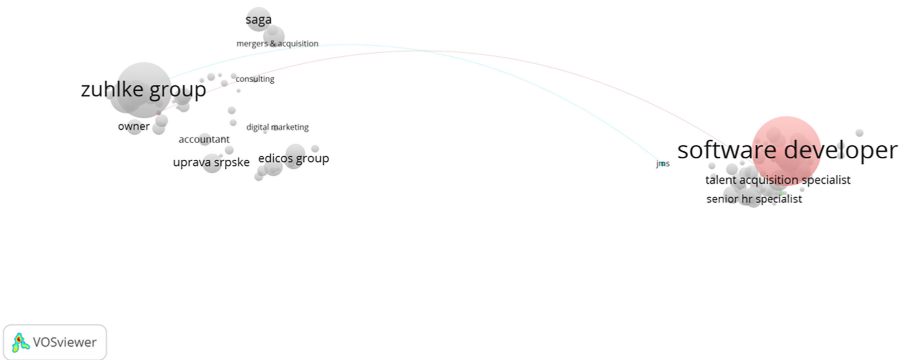

Figure 6.
A cluster of terms related to users' technical skills

(3) The third cluster covers terms related to users with more years of experience. The most common terms in this group are “manager”, “agile business”, “Microsoft”, “expert”, “UML”, “event management” and “employment” (Figure 7). For this group of users, the terms “manager”, “agile business”, “Microsoft” or “UML” are evident, especially in the profile of users whose skills are focused on project management. On the other hand, the combination of the terms “manager” and “hiring” indicates positions in the field of HRM. Most of the terms in this group reflect management or

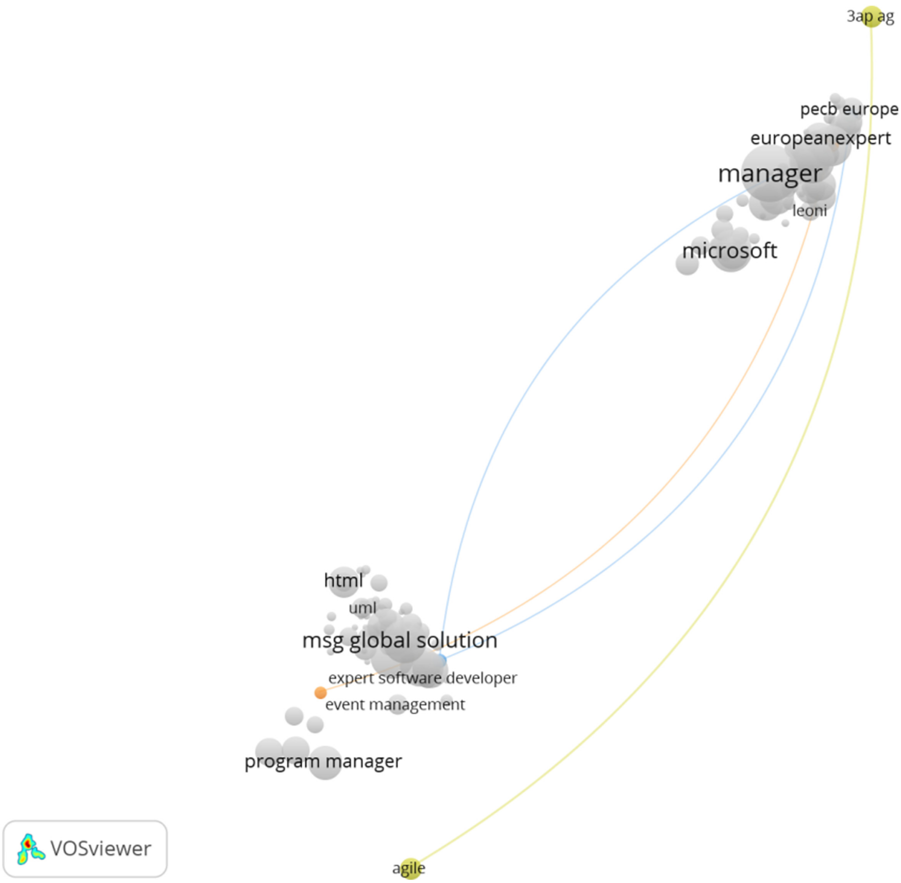

Figure 7.
A cluster of terms related to user experience

leadership skills. Based on this, it can be concluded that this group of terms defines the skills and interests relevant to candidates in senior positions where certain experience is required.

Following the conducted research, three different groups of LinkedIn users' skills and interests can be identified based on the data collected:

(1) *Technical skills and interests* – skills are predominantly oriented toward programming languages, technologies, frameworks, platforms and services. The interests of this cluster, by technical skills, are directed towards leading companies in the field of software development in the territory of Serbia.

(2) *Empirical skills and interests* – skills that fall into this category are recruitment, event management and organization, HRM, project management and knowledge of specific tools and methodologies such as creating UML diagrams. With the skills and interests of this cluster, the positions of senior management or expert in a particular field are most commonly identified. The interests of this category relate to influential personalities from the business world, such as CEOs of leading global companies (e.g. Bill Gates, Satya Nadella and Jack Welch).

(3) *Education related skills and interests* – the skills of this group are directed towards the area of acquired knowledge and education. Interests in this group are oriented towards academic institutions and scientific and professional communities and organizations. Based on the correlation of terms from this group and current employment, it can be determined to what extent the education obtained corresponds to current employment.

5. Evaluation and analysis of results

Within the pilot project, the data were gathered from 476 users' profiles. There were no age or gender restrictions although we had geographical constraint – profiles of users living outside of Serbia were excluded from further research. This user base was used as our sample for conducting a proof of concept research with the method defined in this paper. Using the web scraper tool, nearly 6,000 skills and interests were collected. Extraction process was performed on November 28, 2019 and it took close to an hour. The constraint of the scraping tool we used is that starting URL address of the automatic search plan has to be defined. For this research it will be the URL of all available connections from the authors' network: https://www.linkedin.com/mynetwork/invite-connect/connections/. In this way, we scraped all 1st level connections of authors' LinkedIn network. We are aware that defining starting point manually will bias the results at some level, but from the other side, we expect that this method should be performed in the same way by the recruiters, for them to be able to search adequate profiles inside of their LinkedIn network. Nevertheless, this can clearly be marked as one of the most important limitations of our survey.

The preliminary evaluation of the pilot implementation of the proposed approach was conducted with 6 HR experts whose jobs include the candidate recruiting and preselection. Although the proposed system can be used by a wider population of HR experts, the preliminary evaluation was conducted on a small convenient sample, in order to get information about the usability of the proposed method. The HR experts came from different contexts: 2 were from IT companies, 1 from a public administration institution, 2 from an education institution and 1 from an HR company. During the short session, they were each presented the proposed method, and allowed to analyze the scraped data from LinkedIn through the VOSviewer tool.

The evaluation was done in the form of a survey, using the widely used questionnaire developed by authors Davis (1989) and Lund (2001). This questionnaire enables us to assess perceived usefulness, perceived ease of use and ease of learning of the proposed method. The formulation of a few questions was slightly adapted to fit the context of the research, and two questions were removed from the perceived ease of use section because they would require more experience in using the proposed approach. A Likert type five-point scale is provided for all the questions. In addition, the participants were given opportunity to comment each aspect. The results are given in Table 4.

The results of the preliminary evaluation indicate that the participants find the proposed approach generally useful, and as such they are willing to learn it. They recognized that they would need help related to the web scraping method, and proposed this step to be integrated as a feature in the data analysis software. When looking at individual responses, we saw that the participants from the IT companies and from the HR companies gave higher grades for the questions related to the ease of use and ease of learning. This may be the results of their higher proficiency in using various HR tools and techniques.

6. Discussion

It is possible to achieve effective candidate selection in the recruiting process by identifying key categories of skills and interests of LinkedIn users. For example, when looking for

Table 4. Evaluation results

	Avg	Std. Dev	Selected comments and suggestions
Perceived usefulness	4.00	0.63	- It would be useful to have a tool like this one, to analyze candidates‘ data from social networks
Using the proposed method in my job would enable me to accomplish tasks more quickly	4.17	0.75	- I find the proposed approach useful, but I would prefer to have single software where I can set all the scraping parameters, and then focus on data analysis - I would like to have candidates‘ data from other websites, such as Git, not just LinkedIn
Using the proposed method would improve my job performance	3.83	0.41	- I like that the method can be implemented without heavy investments. I could improve recruitment and preselection i my company using the open source tools
Using the proposed method in my job would increase my productivity	4.67	0.52	
Using the proposed method would enhance my effectiveness on the job	3.83	0.75	
Using the proposed method would make it easier to do my job	4.50	0.55	
I find the proposed method would be useful in my job			

(*continued*)

Table 4.

	Avg	Std. Dev	Selected comments and suggestions
Perceived ease of use			
Learning to use the proposed method would be easy for me	4.17	0.98	- I would need the help of an IT person for performing the scraping step, but the processing and analysis phase were easy for me, since I have experience in various HR analytics methods
I find it easy to get the system to do what I want it to do	4.17	0.98	
The interaction is clear and understandable	3.67	0.82	
I find the system to be flexible to interact with	4.00	0.63	
It is easy for me to become skillful at using this method	4.17	0.98	
It was easy to use	4.50	0.55	
It was effortless	4.17	0.98	
I can use it without written instructions	3.33	1.21	
I did not notice any inconsistencies	4.00	0.63	
Both occasional and regular users would like it	4.00	0.89	
Ease of learning			
I could learn to use it quickly	4.00	1.26	- I would learn it more quickly if the data collection step was included as a feature in the analysis software
I could easily remember how to use it	3.67	1.21	
It will be easy to learn to use it	3.67	1.51	- I would need some training for the tools, but I find the concept easy to include in my job
I quickly became skillful with it	4.33	0.82	

candidates for a particular position, using a defined methodological approach would allow candidates who possess the skills and interests defined by the search to be identified. If it is necessary to define a set of skills and interests, that is, a certain category, using a defined methodological approach it is possible to automate the verification of the existence, or absence, of a particular category of skills or interests on the profiles of potential candidates. The primary intention is reflected in the screening and filtering of the skills and interests of potential candidates, which contributes to more effective recruitment processes. Moreover, presented method could be applied on different social networks, such as GitHub or AngelList for clustering profile skills. For a different social network, only the web scraping instructions would change. For example, for GitHub, we would start scraping from the user profile and by collecting repositories and programming languages they are written in, we could already gain a list of skills for each potential candidate. Moreover, starred projects could give us a list of user interests, which could be used in further research.

In contrast to previous studies, such as Baruffaldi *et al.* (2017) and Chiang and Suen (2015), this paper defines a method for automatic data collection using the web scraper tool. The main advantages of the proposed method for the recruiters are saving the time required for searching and preselection of the candidates, automated collection of more data about the suitable candidates for the job positions in a shorter period, creating an initial data set with reduction of the costs of hiring interviewers, questioners and people who collect data from SNS. Additionally, this method is composed of mutually independent steps. This means that each step can be implemented differently, without changing the whole process. This flexibility distinguishes our method from already known methods, such as direct filtering.

A completely automated and a generally applicable process of data collection from a particular social network site stands out in this methodological approach from the currently available solutions. Considering the method of data collection implemented in this paper, the proposed method provides opportunities to extend the scope of collected data to implicit data, which is not possible using the tools presented in other papers. Furthermore, the presented methodological approach to data collection differs from the previously presented tools for automatic data collection (Tifferet and Vilnai-Yavetz, 2018) due to its applicability on different SNS.

The results of a pilot project evaluation indicate that the HR experts are interested in the proposed method, and that they would be willing to include it in their practice. Although the phase of data collection needs help from a programmer, this can be applicable in many scenarios. On the other hand, the future work should be directed to automation of the web scraping phase, where the users would be provided with an easy-to-use interface that could enable them to configure web scraping parameters without any programming knowledge. A larger sample of participant in the evaluation process would be needed to gain more precise results.

The main constraint of this process is manual revision of the collected skills and interests. In that way, this process still requires domain expert to perform this step which increases the cost of entire process. So, assumption of this process is that recruiters in most cases have at least basic knowledge of relevant skills and interests for the opened position so they should be able to facilitate that part of the process. Additionally, recruiters should have basic knowledge of the SNA methodology in order to understand its application in described method. We believe that this may not be a problem, having in mind the current trends in HR analytics.

The main drawback is the reliability of the collected data because this research was conducted only on explicit information, which includes data that users enter themselves when filling out their profiles. However, compared to other SNS, the reliability of the data on LinkedIn is at an adequate level.

7. Conclusions

The significance of this research is reflected in the definition of a methodological approach for preselection of candidates using social network analysis to improve the social recruiting process. The methodological approach is based on the automatic collection of data from the recruiting platform and is, therefore, platform-independent. As the unique method can be applied to different recruitment platforms, the cost of use is nearly fixed regardless of how many platforms are used. Since it has one manual phase of merging synonyms this step would linearly increase the time and costs of implementation. Additionally, the time needed to learn to use the methodology does not scale with the number of used recruiting platforms.

The main scientific contribution of this paper is the proposition of a methodological approach for an automated process for preselection of candidates from SNS using SNA. The candidate preselection process, presented by the proposed methodological approach, involves the collection, processing and analysis of data from SNS. The proposed method can be applied to similar projects for the preselection of candidates by adopting a list of users' skills and interests required for the analysis in the social recruiting process. Furthermore, depending on the data required for the preselection of candidates for a specific job position, the number of clusters can be extended.

The further direction of this research is to expand the spectrum of user profile analysis. To extend defined methodology process we can cross check given clusters with defined skills and interests for a certain job offer. In addition to the analysis of skills and interests, the analysis of acquired education and acquired vocation could indicate the professional development of users. Another direction of further research is the generalization of the defined methodological approach through implementation in the context of different candidate recruitment platforms. As proposed by the participants in the pilot evaluation, the authors will continue work on development of a software solution that enables HR experts to easily and without technical knowledge apply the proposed method.

References

Adamic, L. and Adar, E. (2005), "How to search a social network", *Social Networks*, Vol. 27 No. 3, pp. 187-203, doi: 10.1016/j.socnet.2005.01.007.

Backman, C. and Hedenus, A. (2019), "Online privacy in job recruitment processes? Boundary work among cybervetting recruiters", *New Technology Work Employment*, Vol. 34 No. 2, pp. 157-173.

Baruffaldi, S.H., Di Maio, G. and Landoni, P. (2017), "Determinants of PhD holders' use of social networking sites: an analysis based on LinkedIn", *Research Policy*, Vol. 46 No. 4, pp. 740-750.

Bholowalia, P. and Kumar, A. (2014), "EBK-means: a clustering technique based on elbow method and k-means in WSN", *International Journal Computer Applications*, Vol. 105 No. 9, pp. 17-24.

Boyd, D.M. and Ellison, N.B. (2007), "Social network sites: definition, history, and scholarship", *Journal Computer Communications*, Vol. 13 No. 1, pp. 210-230.

Calopereanu, M.F., Şuică, M. and Calopereanu, A.N. (2019), "Considerations regarding compliance with the regulation (EU) 2016/679 (GDPR) in human resources departments", *Proceedings of the International Conference on Business Excellence*, Vol. 13 No. 1, pp. 548-559.

Castillo-de Mesa, J. and Gómez-Jacinto, L. (2020), "Connectedness, engagement, and learning through social work communities on LinkedIn", *Psychosocial Interviews*, Vol. 29 No. 2, pp. 103-112.

Chiang, J.K.-H. and Suen, H.-Y. (2015), "Self-presentation and hiring recommendations in online communities: lessons from LinkedIn", *Computational Human Behaviour*, Vol. 48, pp. 516-524.

Davis, F.D. (1989), "Perceived usefulness, perceived ease of use, and user acceptance of information technology", *MIS Quarterly*, Vol. 13 No. 3, pp. 319-340.

Dewi, L.C. and Chandra, A. (2019), "Social media web scraping using social media developers API and Regex", *Procedia Computational Science*, Vol. 157, pp. 444-449.

El Ouirdi, M., Pais, I., Segers, J. and El Ouirdi, A. (2016), "The relationship between recruiter characteristics and applicant assessment on social media", *Computational Human Behavior*, Vol. 62, pp. 415-422.

Ferrara, E., De Meo, P., Fiumara, G. and Baumgartner, R. (2014), "Web data extraction, applications and techniques: a survey", *Knowledge-based System*, Vol. 70, pp. 301-323.

Flowers, A. (2019), "Indeed tech skills explorer: today's top tech skills", available at: https://www.hiringlab.org/2019/11/19/todays-top-tech-skills/ (accessed 18 April 2020).

Foucault Welles, B. and Xu, W. (2018), "Network visualization and problem-solving support: a cognitive fit study", *Social Networks*, Vol. 54, pp. 162-167, doi: 10.1016/j.socnet.2018.01.005.

Freepik (2020), "Candidate search", *Flat Icon*, available at: https://www.flaticon.com/free-icon/job_1759339 (accessed 15 April 2020).

Garbade, M.J. (2018), "Understanding K-means clustering in machine learning", available at: https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1 (accessed 19 April 2020).

Gërxhani, K., Brandts, J. and Schram, A. (2013), "The emergence of employer information networks in an experimental labor market", *Social Networks*, Vol. 35 No. 4, pp. 541-560, doi: 10.1016/j.socnet.2013.08.001.

Girard, A. and Fallery, B. (2009), "E-recruitment: new practices, new issues. An exploratory study", in Boudarouk, T. and Ruel, H. (Eds), *Human Resource Information System*, INSTICC Press, pp. 39-48, 978-989-8111-97-5.

Golovko, D. and Schumann, J.H. (2019), "Influence of company Facebook activities on recruitment success", *Journal Business Research*, Vol. 104, pp. 161-169, doi: 10.1016/j.jbusres.2019.06.029.

Han, J. and Han, J. (2009), "Network-based recruiting and applicant attraction in China: insights from both organizational and individual perspectives", *International Journal Human Resource Management*, Vol. 20 No. 11, pp. 2228-2249, doi: 10.1080/09585190903239633.

Hedenus, A., Backman, C. and Håkansson, P. (2019), "Whom do you know? Recruiters' motives for assessing jobseekers' online networks", *International Journal Human Resource Management*, March, pp. 1-24, doi: 10.1080/09585192.2019.1579245.

Hosain, S. and Liu, P. (2020), "Recruitment through LinkedIn: employers' perception regarding usability", *Asian Journal of Management*, Vol. 11 No. 1, pp. 54-60.

Jamoury (2020), "Data science icon", available at: https://www.cleanpng.com/png-data-analysis-analytics-management-big-data-data-p-2235063/ (accessed 15 April 2020).

Jiang, J., Wang, S. and Zhao, S. (2012), "Does HRM facilitate employee creativity and organizational innovation? A study of Chinese firms", *International Journal Human Resource Management*, Vol. 23 No. 19, pp. 4025-4047, doi: 10.1080/09585192.2012.690567.

Liang, H. and Zhu, J.J.H. (2017), "Big data, collection of (social media, harvesting)", *International Encyclopedia Communication Research Methods*, pp. 1-18.

LinkedIn Inc (2020), "About LinkedIn", available at: https://about.linkedin.com/ (accessed 15 April 2020).

Lund, A.M. (2001), "Measuring usability with the use questionnaire12", *Usability Interface*, Vol. 8 No. 2, pp. 3-6.

Malhotra, A., Totti, L., Meira, W. Jr, Kumaraguru, P. and Almeida, V. (2012), "Studying user footprints in different online social networks", *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 1065-1070.

McKinney, W. and Team, P.D. (2015), *Pandas: Powerful Python Data Analysis Toolkit*, Pandas—Powerful Python Data Anal. Toolkit, p. 1625, available at: https://pandas.pydata.org/pandas-docs/version/0.7.3/pandas.pdf.

Melanthiou, Y., Pavlou, F. and Constantinou, E. (2015), "The use of social network sites as an e-recruitment tool", *Journal Transnational Management*, Vol. 20 No. 1, pp. 31-49.

Milovanović, S., Bogdanović, Z., Labus, A., Barać, D. and Despotović-Zrakić, M. (2019), "An approach to identify user preferences based on social network analysis", *Future Generation Computer Systems*, Vol. 93, pp. 121-129, doi: 10.1016/j.future.2018.10.028.

Nylen, E. and Wallisch, P. (2017), *Web Scraping*, Neural Data Science, pp. 277-288.

Okolie, U.C. and Irabor, I.E. (2017), "E-recruitment: practices, opportunities and challenges", *European Journal Business Management*, Vol. 9 No. 11.

Ou, C.X., Sia, C.L. and Hui, C.K. (2013), "Computer-mediated communication and social networking tools at work", *Informational Technology People*, Vol. 26 No. 2, pp. 172-190.

Pace, R.D. (2016), "The best new social networks for marketers", available at: https://everything-pr.com/social-media-network-marketing/ (accessed 15 April 2020).

Parks-Yancy, R. and Cooley, D. (2018), "Who gets the job? First-generation college students' perceptions of employer screening methods", *Journal Education Business*, Vol. 93 No. 1, pp. 1-10.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J. (2011), "Scikit-learn: machine learning in Python", *Journal Machine Learning Research*, Vol. 12, October, pp. 2825-2830.

Peleshchyshyn, A. and Mastykash, O. (2017), "Analysis of the methods of data collection on social networks", *2017 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, Vol. 1, pp. 175-178.

Pereira, R.C. and Vanitha, T. (2015), "Web scraping of social networks", *International Journal Innovation Research Computer Communication Engineering*, Vol. 3 No. 7, pp. 237-240.

Provost, F. and Fawcett, T. (2013), *Data Science for Business: what You Need to Know about Data Mining and Data-Analytic Thinking*, O'Reilly Media.

Ramaabaanu, R. and Saranya, M. (2014), "Importance and problems of E-recruitment", *International Journal Research*, Vol. 1 No. 9, pp. 445-450.

Raval, U.R. and Jani, C. (2016), "Implementing & improvisation of K-means clustering algorithm", *International Journal Computer Science Mobile Computational*, Vol. 5 No. 5, pp. 191-203.

Rocchini, C. (2020), "Betweenness graph", *Wikipedia*, available at: https://en.wikipedia.org/wiki/Social_network_analysis#/media/File:Graph_betweenness.svg (accessed 15 April 2020).

Sharma, N. (2014), "Recruitment strategies: a power of E-recruiting and social media", *International Journal Core Engineering Management*, Vol. 1 No. 5, pp. 15-35.

Southekal, P.H. (2017), *Data for Business Performance: the Goal-Question-Metric (GQM) Model to Transform Business Data into an Enterprise Asset*, Technics Publications.

Stadtfeld, C., Takács, K. and Vörös, A. (2020), "The emergence and stability of groups in social networks", *Social Networks*, Vol. 60, pp. 129-145, doi: 10.1016/j.socnet.2019.10.008.

Steketee, M., Miyaoka, A. and Spiegelman, M. (2015), "Social network analysis", *International Encyclopedia Social Behaviour Science*, January, pp. 461-467, doi: 10.1016/B978-0-08-097086-8.10563-X.

Syakur, M.A., Khotimah, B.K., Rochman, E.M.S. and Satoto, B.D. (2018), "Integration K-means clustering method and elbow method for identification of the best customer profile cluster", *IOP Conference Series: Materials Science and Engineering*, Vol. 336 No. 1, 12017.

Tifferet, S. and Vilnai-Yavetz, I. (2018), "Self-presentation in LinkedIn portraits: common features, gender, and occupational differences", *Computational Human Behavior*, Vol. 80, pp. 33-48.

Toteva, K. and Gourova, E. (2011), "Social network analysis in professional e-recruitment", *Third International Conference on Software, Services and Semantic Technologies S3T 2011*, pp. 75-80.

Van Dijck, J. (2013), "'You have one identity': performing the self on Facebook and LinkedIn", *Media, Culture Society*, Vol. 35 No. 2, pp. 199-215.

Van Eck, N.J. and Waltman, L. (2011), "Text mining and visualization using VOSviewer", arXiv Prepr. arXiv1109.2058.

Veloutsou, C. and Ruiz Mafe, C. (2020), "Brands as relationship builders in the virtual world: a bibliometric analysis", *Electronic Commerce Research Application*, Vol. 39, 100901, doi: 10.1016/j.elerap.2019.100901.

Voigt, P. and Von dem Bussche, A. (2017), "The EU general data protection regulation (GDPR)", in *A Practice Guideliness*, 1st ed., Cham Springer International Publishing.

Voytek, B. (2017), "Social media, open science, and data science are inextricably linked", *Neuron*, Vol. 96 No. 6, pp. 1219-1222.

Wasserman, S. and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press.

WebScraper (2020), "WebScraper - making web data extraction easy and accessible for everyone", available at: https://webscraper.io/ (accessed 15 April 2020).

Zhu, Z., Zhou, Y., Deng, X. and Wang, X. (2019), "A graph-oriented model for hierarchical user interest in precision social marketing", *Electronics Commerce Research Application*, Vol. 35, 100845, doi: 10.1016/j.elerap.2019.100845.

About the authors

Stevan Milovanović is PhD student at Faculty of organizational sciences, University of Belgrade. His areas of interest are software engineering, mobile application development, social network analysis and web solutions.

Zorica Bogdanović is an associate professor at Faculty of organizational sciences, University of Belgrade. Her areas of interest are: e-business, data science, big data, business intelligence, digital marketing, social media, Internet of things, blockchain, mobile business.

Aleksandra Labus is an associate professor at Faculty of organizational sciences, University of Belgrade. Her areas of interest are: e-business, social media, data science, digital marketing, social network analysis, Internet of things, blockchain, e-education. Aleksandra Labus is the corresponding author and can be contacted at: aleksandra@elab.rs

Marijana Despotović-Zrakić is full professor at Faculty of organizational sciences, University of Belgrade. Her areas of interest are: computer simulation, e-business, web programming, mobile business, digital marketing, social media, Internet of things, risk management, e-education.

Svetlana Mitrović is an assistant professor at Project Management College. Her areas of interest are: e-business, e-HRM, e-recruitment, project management, social networks, e-government.