CCC Publications

Determining IT Student Profile Using Data Mining and Social Network Analysis

L. Stanca, R. Lacurezeanu, A. Tiron-Tudor, V.P. Bresfelean, I. Pandelica

Liana Stanca, Ramona Lacurezeanu*

Babes-Bolyai University of Cluj-Napoca
Faculty of Economics and Business Administration
Cluj-Napoca, Teodor Mihali 58-60, 400591,Romania
liana.stanca@econ.ubbcluj.ro, ramona.lacurezeanu@econ.ubbcluj.ro
*Corresponding author: ramona.lacurezeanu@econ.ubbcluj.ro

Adriana Tiron-Tudor, Vasile Paul Bresfelean

Babes-Bolyai University of Cluj-Napoca
Faculty of Economics and Business Administration
Cluj-Napoca, Teodor Mihali 58-60, 400591,Romania
adriana.tiron@econ.ubbcluj.ro, paul.bresfelean@econ.ubbcluj.ro

Ionut Pandelica

1. Academy of Economical Sciences in Bucharest
Bucharest, Piata Romana 6, 010374, Romania
ionut.pandelica@rei.ase.ro
2. Agora University of Oradea
410526 Oradea, P-ta Tineretului 8, Romania,

Abstract

To become higher competitive a university needs to develop a viable students' absorption strategy on the labor market. A key to the successful development of such a strategy rests to synchronize jobs descriptions with profiles and behavior of IT students. In order to generate this synchronization, it is essential to identify a way to improve university curricula, learning and teaching process based on the students' profile and on the labor market needs. In this manner, universities could offer IT companies information about their IT students' profile and behavior. Our paper proposes a data mining and social network analysis to examine IT students' skills and behavior in order to generate their actual profile. The results contribute to the development of knowledge concerning the IT graduates' profile and based on this, a solution that might match the university curricula with the labor market requirements. Finally, the results attempt to provide IT companies with information with the aim of better understanding the IT students' profile and to create a realistic description of the job in the recruitment software on the digital market.

Keywords: data mining, social network analysis, personas, IT graduates' skills, job profile.

1 Introduction

Contemporary higher education has known an unprecedented level of development at a European and worldwide level, under the influence of social and economic changes [37, 48]. It is in this context, terms such as big data, data mining, visual analytics data mining, design thinking, partner recognition, education 3.0, e-learning 3.0 appear and were introduced in the university's curricula [7]. All these concepts have a significant impact on the academic environment, while the business environment modifies its dimension and increases the accuracy of the results obtained based on the decision analysis of teaching & learning. Some researchers [43] state that e-learning shifts the focus from how to introduce technology in teaching to analytical learning, perceiving it as an essential tendency for the learning and teaching process based on technology [37].

The explosion of digital technologies has changed how talent is recruited. For the young generation Z (born after 1994), which now forms a significant part of the workforce, a new manner of recruiting talented employees is needed [48]. As compared to past generations of students, the new generation is much more oriented and set on obtaining the information that the youngsters consider essential for their development [59]. The educator, teacher or professor now has the rather difficult task of demonstrating and justifying the necessity and usefulness of all the information taught [7]. With this kind of direct learning demand and need, the providers of education (universities, colleges and even the providers of professional formation courses) ought to provide courses and a curriculum that is not only flexible and accessible anytime and anywhere, but most important of all, courses that are relevant and applicable to the demands of the labor market [48]. Otherwise, they would not be able to remain in demand, given the competitive environment of this field. Up until now the labor market has required a CV, representing a form of a genuine professional card, but from now on everything will have to do with the aptitudes and skills of the individuals as well as how they can present these abilities on various social platforms. Furthermore, they need to be able to prove their assumed abilities from the very moment when they step into a company [37].

In addition to professional and transversal competencies offered by the universities to their graduates, companies search for specific socio-psychological identities that create in a synergic manner the "ideal profile" of the IT student graduate in accordance with the companies' vision and organizational culture. Based on this "ideal profile" that combines mentality, hard and soft skills, psychometric information, personality dimension, companies create the "hunted" profiles for specific IT data mining on recruitment platforms. The knowledge of this profile can contribute to enhance the connection between the market needs with the universities "deliveries" of IT students "products" and eliminate the existing discrepancies between the demand and offer that characterize the contemporary labor market [4].

In this context, the paper's objective is to analyze whether the profile of the IT graduates based on the university curricula corresponds to the market's requirements, using a mix of empirical research methods in an innovative manner. The paper results contribute to the matching process between technical and transversal competencies and labor market needs, bringing insights specific to one of the most dynamic workforce domains, in particular, the IT domain. Moreover, the paper contributes to knowledge development in two applicative directions. First of all, the profiles identified will help universities improve curricula and teaching/learning processes based on the labor market needs. Secondly, by presenting information to IT companies, they will be able to understand the profile of the IT students and to create a realistic job description in the recruiter software on the digital market. The novelty of this paper results also in the combination of data mining and social network analysis methods that facilitate the attempt to carry out a sophisticated analysis of the problem.

Our study is organized as follows: section 2 presents the debates from the research literature concerning the job matching system, cooperative education and the psycho-social profile of the IT employees. Section 3 delivers the theoretical framework based on the psychological and social profiles, followed by the methodology from section 4. The fifth section describes our research results, in the sixth section we present the final discussions and the last section comprises main conclusions, limits and further possible developments of this study.

2   Nowadays directions of research

2.1   Job Matching System and Cooperative Education

The bidirectional relation between education and the business environment is recognized by the literature via the concept of Cooperative Education . The main factors involved in Cooperative Education are the teachers, students and the business environment [36, 38]. The development of a course in the university field, according to Cooperative Education theory, has the main advantage of optimizing the recruitment process in the IT domain. One of the arguments brought in favor of this statement is that the recruiters will get to know information about the personality, level of academic preparation and skills of the students who are involved in the sphere of business information systems education [4]. Therefore, this would allow them to present job demands anchored in the reality of the labor market. The academic environment would benefit from the fact that it would be able to adapt its curriculum, programs and basic courses according to the direct feedback provided by the business environment.

Many researchers [32, 33, 34, 43] consider that Cooperative Education facilitates the cooperation between business and academic environments for courses creation as a context for the preparation of the student following the demands of the employer. At the same time, the employer learns how to define the available jobs that would be attractive for potential employees on the online platforms, while the educators could adapt their curricula to the demands of the labor market [4]. The courses developed on the philosophy of cooperation in education place emphasis on studying the psychological and social profiles of the students to generate descriptions of jobs so that these would be correctly and accurately understood by all those involved. Furthermore, according to cognitive philosophy, the psychological and social profiles which one analyses include the students' learning in connection to the mental models and possible wrong mental representations, intending to identify the typical mistakes made by the student in the process of learning, so as to avoid them [43]. As a consequence, the learning process depends on how the students represent and process knowledge and information, while treating the students as creators and masters of their learning process, of their own content of information, abilities and newly gathered knowledge. The course model thus created, will allow students to correctly and efficiently assimilate a programming language.

Following the constructivist theory, our paper intents to gather information related to the profile of the IT student, the future employee in the IT domain. Technology in the educational domain has changed what we provide, how we can evaluate and how we can connect the attendants to the instructors and the content. Furthermore, education 3.0 enables mass personalized education through mixed learning. More precisely, one talks about bringing together, of course, attendants from different domains and with different abilities that work together [41]. Moreover, education 3.0 incorporates e-Learning 3.0 and characterizes it as being both "collaborative" and "intelligent" [59]. Collaboration in education 3.0 appears as social learning and places the focus on social aspects of importance, such as participation, collaboration, building together with an improvement, all of which become vital in the learning and teaching process [4]. A possible solution for building a better learning space that offers through discovery not through memorization the skills of the 21st century include: ways/learning spaces that allow students to acquire creative thinking, flexible problem solving, collaboration and innovative skills needed to be successful at work and in life [5, 26, 63].

Moreover, universities' curricula influence the profile of graduate students which is the future worker profile [40, 72]. Our paper sustains the idea that the psychosocial profiles of graduates are based on the learning space the teachers created following the university curriculum and in line with the 21st century education theories.

2.2   Psycho-social profile of the IT employees

Personality refers to individual differences among people in behavior, cognition and emotion patterns [53].

Starting from the set of central features characterizing the individual, that define the basic personality: agreeableness, likability, being conscientious, extroverted, intellectual, but also exhibiting

neuroticism [25], the evolution of the psycho-social profiles of the IT employees are debated below. There is a significant amount of research interest related to the creation of the psycho-social profiles of IT employees, starting in the 80s that has increased constantly. Many attempts and experiments have the purpose of determining/ identifying techniques that would match the persons to the appropriate workplace. Some authors [44, 50, 55, 68] consider personality as being more important than intelligence in the case of the IT employee. This stereotype of the programmer wards off youngsters from the programmer profession in the USA [51]. Based on the job announcements published in the New York Times and USA Today, a study [51] identified that 20-40% of the programmers have a personality characterized by introversion, detection, thinking, judgment, while the others are described in terms of extroverted, sociable and agreeable personalities.

Moreover, both [62] following [51] researches, consider programmers are social autisms that cannot "think outside the box", statements that lead us to the assumption that programmers may have distinct personalities from other professionals. Other studies reveal that in programming, personality is more important than intelligence, and many programmers fail mostly because of their personality flaws, rather than from the lack of intelligence [68]. Furthermore, the programmer's work is situated at the crossroads between the act of creation and technique, and that employs adequate education, so the programmer's personality can be changed [51, 56]. Also, several studies have examined the role of personality in IT related courses [45] using Meyers-Briggs or similar measures of cognitive style to determine personality and its relationship to course performance or potential educational strategies [9, 70].

The research was carried out by observing students in a learning environment that comprised physical and virtual spaces. The virtual and physical space in a course had the aim of obtaining and gathering information about in what way students use, interact with and manage course content. The virtual environment allowed us to measure the real degree of involvement of each individual in a course, based on which one can identify the profiles and personalities of future IT employees. The course outline to create this experiment is intended to represent the prototype of a course created on the concepts of Cooperative Education [6, 16]. The experiment ends with the analysis of the choices of job announcements by each student. The announcements were collected from a dedicated IT job announcements platform. The purpose of this action is to verify the link between the students' personality features and the choice of a specific job announcement. The results of this approach are useful for both academic and business areas in the subsequent manner. The result of our research will generate an actual stereotype of the IT graduate students that would allow IT firms to have a clearer vision over the students' profile and most important to create more suitable job announcements. Furthermore, this work provides useful information to academic environment educators for more customized courses contents and styles of delivering, with regard to the students' profile and labor market's demands.

3   Theoretical framework

The psychological and social profile does not represent a method of investigation per se, but a way to represent the visible results of measures by using various pieces of evidence applied to a subject or a sample-group of subjects [18, 36, 54]. These profiles highlight the performances obtained by the individual in a set of tasks. The psycho-social profile offers the global but in the same time detailed image of the level of development characteristics for various psychological and social functions of the individual.

The individual's personality portrait has applications in many domains. For example, the personality model IPIP (International Personality Item Pool) was used to measure the degree of gratefulness and spiritual wellbeing [66], in examining the effects of the personality features on the processes and results of teamwork [21], in examining the manner of using medicine issued on prescription [62] and in exploring how the investor's disposition can affect and influence the concentration of the investment portfolio [15]. Close related to our objective, the managers are aware of the necessity to evaluate the personalities of their employees in a task distribution process as well as evaluating the performances at the workplace, being convinced by its efficiency [47].

The studies dedicated to personality identification use five principal dimensions of the personal-

ity, called Big Five Personality or OCEAN: openness to experiences, conscientiousness, extraversion, agreeableness, and neuroticism [61] (see Table 1).

Table 1: Personalities types, characteristics and indications

Personalities	Characteristics	Indicates how the person is
Openness (Intellect)	People who like to learn new things and enjoy new experiences usually score high in openness. Openness includes traits like being insightful and imaginative and having a wide variety of interests	Open-minded and authority-challenging
Conscientiousness	People are reliable and prompt. Traits include being organized, methodic, and thorough	Self-disciplined and organized
Extraversion	Extraverts get their energy from interacting with others, while introverts get their energy from within themselves. Extraversion includes the traits of energetic, talkative, and assertive	Out-going and social
Agreeableness	These individuals are friendly, cooperative, and compassionate. People with low agreeableness may be more distant. Traits include being kind, affectionate, and sympathetic.	Warm, friendly and tactful
Neuroticism (Emotional Stability)	This dimension relates to one's emotional stability and degree of negative emotions. People that score high on neuroticism often experience emotional instability and negative emotions. Traits include being moody and tense	Ability to remain stable and balanced

Programmers' attitude is strictly connected to openness because they need to poses the ability to conceive of unconventional methods in order to obtain new solutions to a problem [52]. As a consequence, their top characteristics should be imagination, creativity, and intellectual curiosity [15], but at the same time, they are somehow incompetent from the social point of view. Also, programmers' attitude is associated with being introverted, with lack of concord and neuroticism. Therefore, in our view, cognitive competences could model and refine the programmer's personality [47]. Built-up on this framework, our paper illustrates an up to date profile for the IT graduate student. In our experiment, the student was observed for 14 weeks in a web programming course. The teaching environment is a complex one, conceived and developed so that it allows the development of cognitive and digital competences. The evaluation of the student involved a practical test followed by a project, then written evaluation over the 14 weeks, focusing on how they interacted in teams and in which way they solved their tasks. At the end of the course, we created a questionnaire with 33 items following Cattell's instrument 16 PF (The Sixteen Personality Factor Questionnaire (16PF) [12]. The 16PF is one of the oldest, most researched, and most influential measures of normal personality [3, 11] using the Big-Five personality factors: Extraversion (E), Agreeability (A), Conscientiousness (C), Emotional Stability (ES) and Intellect (I). The answers to the 33 questions were presented on a scale of 5 Likert points, from (1) standing for "I do not agree" to (5) "I agree". The personality features are considered

by [12], as basic tendencies that refer to the individual's potential, while the attitudes, roles and relationships to the adapting objectives reflect the interaction between the elementary tendencies and the environment. Based on this instrument, the research intends to identify the features/facets of the personality that can represent predictors of the natural absorption of students in the IT domain and work environment.

Finally, we achieve the prototype of a profile that needs to be validated in time. The identified profile is useful for the academic environment and the labor market. In academia, the contribution of the prototype will allow a smooth adaptation of the curriculum and the teaching/learning activity to the demands of the labor market. For the latter, it will provide the IT companies with information that would enable them to understand the profile of IT students and will create a realistic description of the job advertised in the digital labor market, in order to facilitate the superposition of the offer and the request on the supposed market.

4 Material and research methods

In order to achieve this objective concerning the IT graduate student profile as a future employee in the 21st century, our paper follows an exploratory empirical research using the survey method and a questionnaire based on Cattell's personality test to collect the data. The analysis and appreciation of each personas item materialized itself in a grade with values ranging from 1 to 5.

We utilized several software packages with the purpose of testing the research hypothesis: SPSS 12, Statistica 7.0 and NodeXL. The statistical process performed is described in the next steps:

1. We applied statistical data analysis to illustrate the study participants' characteristics.

2. We tested whether the items that form the personality measurement instrument have a single dimension, a single latent factor. This basic rule is known as unidimensionality [20] and is linked to the second assumption, namely the local Independence of items [20]. The unidimensional does not refer strictly to the presence of single dimension, but to the existence of a dominant dimension (ability or coverage in the latent factor [20] that would influence the performance in the test [39]. In the case of a test that measures presence we cannot have the pretention that the items could measure pure presence. The performance of a subject in connection to a test item is foreseen and explained by the existence of a latent factor, the relationship between this performance and the latent factors that represent its basis is described by a monotonous and increasing function called the item response function or the particular curb of the item [39]. Such a reliability-validity analysis will estimate the viability of the instrument by measuring the internal consistency of the elements, to the extent that they correlate well. The viability of our instrument was calculated according to the theory developed by Spearmen [58]. Going further, a scale or a test presents a better fidelity when the fidelity coefficient is more significant than 0.70 [20]. At this stage, we decided whether the instrument employed has an average degree of trust or consistency [20], so that its result would be the same throughout time and so that it can be used for scenarios that are similar to the ones analyzed in this article.

3. The analysis of principal components (PCA) with the purpose to highlight pre-existent structures in a multitude of multivariable data [19]. These structures are generally expressed through the variation and co-variation of variables and of similarities, respectively differences among objects. In this first stage, we identify the values and own vectors of a square matrix obtained by multiplying the data matrix (objects) with its transposed function, which is a relatively standard procedure in the framework of matrix calculus. The role of the principal component analysis was to build decision models based on the answers to the questions from the Sixteen Personality Factor Questionnaire. These answers would help us formulate a pattern that would contain the restraining factors, which explains the variation of the studied variables. The variables with the non-representative correlations (<0.3 of the correlation matrix) were eliminated [19]. The next step of the analysis eliminates the variables based on the Kaiser-Meyer-Olkin measurements (the values of the co-variation indicators from the main diagonal of the anti-imagines matrix >0.5) [20].

4. In order to analyze the existence possibility of hidden aspects that have not been noticed or foreseen, aspects that could generate different and inexplicable results in the context of the problem, we utilize Social Network Analysis. We considered appropriate to create an analysis that enters the

sphere of Social Network Analysis, based on which we intend to determine the attributes present in job advertisements by the choices made by students following their personality features. Centrality is considered the most often used conceptual tool for exploring the actors' (locations) roles within social networks. Central locations have to be more active in the sense that they have the highest number of connections with other actors-locations in the network [67].

In this research, we adopted [24] theory of nodes in a network, according to which group centrality shows the centrality degree of a group meaning the number of actors outside the group that are connected to the member of that group. Different ties to the same actors by different group members are only counted once. Closeness centrality measures the closeness of a node to the other. Consequently, closeness centrality can be used to determine the degree of influence a node has in the network. Group closeness centrality refers to the normalized inverse sum of distances from the group to all nodes outside the group. Group betweenness centrality reveals the "proportion of geodesics connecting pairs of non-group members that pass through the group" [24]. These three centrality indicators allowed us to measure the role of attribute played in job description in a network for personality attributes. The results of our research are in line with results from other studies [13].

5 Study findings

5.1 Sample description

The group of respondents is composed of 106 students, a representative number taking into consideration the next two arguments based on the research literature. The group is higher than 100, represents a homogeneous population and according to the literature is considered to be a representative number [17, 28, 64, 65]. Moreover, certain statisticians [2] agree that if the population of interest was homogenous, then the sample might be smaller than 100.

All students are enrolled at Babes-Bolyai University of Cluj-Napoca, in the web-programming course. Group demographic characteristics concerning gender and age are presented in Table 2.

Table 2: Demographic variables

Gender	Female	60.4 %
	Male	39.6 %
Age	18-25	82.1 %
	25-30	12.3 %
	over 30	5.7 %

The data was collected for the period of a 14 weeks course. During the course, the students took part in two types of face-to-face meetings: lectures (dedicated to theoretical teaching conception) and practical (laboratory sessions). Additionally, an e-learning platform (Moodle) was used for teaching/learning activities owing to the advantages it offered. We mention just two of them: quick access for students to the training materials and professors' constant support. The records from the Moodle platforms allowed us to determine the degree of involvement presented by the students in solving tasks involving the study of course materials and bibliography uploaded at the beginning of each week. All through the semester, students were evaluated during face-to-face meetings in the classroom as well as on the e-learning platform's virtual environment. The descriptive analysis allowed us to create a profile of the participants, presenting the following characteristics:

1. 82.1% come from urban environments and the rest of 17.9% from rural areas.

2. After at least three years of using the Moodle platform in IT learning processes, 88.7% state that they enjoy it and that it helps them learn, while 11.3% expressed total disagreement.

3. The grades they obtained during these 3 years of university studies are as follows: 3.8% obtained 10, 4.9% between 9-10; 31.6% got grades in the interval of 8-9; 8.8% got grades from 7-8; 20.9 % obtained grades from 6-7; 28.1% from 5-6; 1.9% got grades below 5. (Romania uses a 10-point

scale grading system: 1 being the lowest, 10 being the top grade, and 5 the minimum passing grade).

4. 65.1% from the study participants enjoy teamwork, and 34.9% prefer to work alone.

5. 39.6% from the participants enjoy dealing with assessments in good time, while 60.4% work against a time limit or cram for exams.

6. The preferred manner of examination is the practical exam for 34.9%, while 30.2% prefer oral project presentation and 34.9% written examines (multiple choice questions, open questions).

7. Concerning the learning style of the participants, 23.6% learn mostly visually; 0.9% mostly auditory; 22.7% predominantly by writing and reading; 51.9% rely on learning through demonstrations, simulations, case studies, practical tasks; the rest of 0.9% are mixed.

5.2 Instrument for the profile of the IT students

The next step of the study was to create an analysis of the reliability-validity type of the technique we propose in order to obtain an instrument that measures the profile of the IT students. The results of the analysis are:

- The test is sensitive to the measured characteristics: alpha Cronbach =0.767; Mean=110.77; Std.Dv.=9.261. The alpha value is significant and suggests that the instrument is adequate to the purpose for which it was built so one can state with precision and accuracy that the test is unidimensional so there is an agreement between the students' assessments.

- The students manifested high levels of agreement in appreciating the items that compose the measurement of their personality and course attendance (average Inter-Item Correlation: 0.764).

- The study continued with ANOVA to test if the responses were unilateral (without differences between students and the effect of interaction with the instrumental). The results of the test F=1364.385; p-value=0.000 confirm that there are differences between the answers provided by students according to their personality and transversal competences.

5.3 IT students' personality types

The analysis continued by applying the Principal Component Analysis (PCA), which allowed us to test whether our instrument contained latent factors or not. More exactly, we wanted to see if there were central items of the instrument and secondary ones, or whether they formed an entity that measured the same dimension. The PCA purpose is to obtain a small number of linear combinations (main components) from a set of variables that would retain as much information as possible from the initial variables. The Principal Component Analysis method provides a series of results such as descriptive statistics, correlation matrix, the Kaiser-Meyer-Olkin and Bartlett tests, communality, and the total explained variance, the Screen Plot image, the matrix of components and the matrix of components after the rotation. By applying these methods, we could identify the existence of a smaller number of variables from the mass of studied variables that express most of the total variance. To be more precise, we wanted to verify if the analyzed instrument could identify the 16 latent factors in Cattell's manner.

PCA with equamax rotation allowed us to verify the validity of constructing the scale utilized in measuring the behavior/ profile of the IT student. During the process, we tried to eliminate items that presented less than 0.3 correlations, but there were no such items. The results of the Kaiser-Meyer-Olkin test showed how adequate the sampling process was 0.736>=0.50 [30, 69]. This proved that the sample used was sufficient for the study and as a result, the variability of data was caused by the instrument created in Bartlett's Test of Sphericity assumption that all correlation coefficients are not quite far from zero. The result of Bartlett's Test of Sphericity, (chi-Square=1287.792; p=0.000) meant our variables were related and therefore suitable for structure detection, so the analysis of the factors was useful in the prevailing of data and the subscales of the scale were inter-independent. The

patterned relationship among the variables is seen from the correlation matrix. According to Kaiser's criteria (from commonalities the values need to be bigger than 0.4), the latent root is formed of items taken from the study. Therefore, there is a limited set of items that determine the validity of the tool, stating that there is a determined set of items which determines the validity of the instrument in a proportion of 79.641% according to which there are items that moderately correlate in the study.

The orthogonal factorial model obtained by the equamax method is formed by 9 factors, not 16 as in Cartel's case. The result obtained after the omission of the saturations that are smaller than 0.4 [19] are presented in Table 3.

To summarize, the basis of the analysis of the main components allowed us to identify nine typologies of the IT students' personality according to Cattell's theory which was also validated in our study. The general profile of the IT students based on the analyzed sample can be described in the next pharagraphs.

Table 3: Personality types

Types	Description
A	start projects and conversations easily (0.690); comfortable around people (0.668); interested in people (0.665); talk to many people in a party context (0.664); do not mind being at the center of the party, the life and soul of the party (0.630); animate a party (0.547); have excellent ideas (0.550); are full of ideas (0.544) make people feel at ease (0.523); complete their tasks at once (0.456); share the others' emotions with empathy(0.401); make jokes all the time(-0.568).
B	orderly (0.667); are considered extroverted persons (0.525); create a schedule for themselves (0.463); want everything to be accurate and correct (0.340)
C	difficulty in understanding abstract ideas (0.698); they are not imaginative persons (0.684); solve their tasks at once (-0.400); introverted (0.525)
D	make jokes all the time (-0.568); impress people around them (0.520); like reading books (0.492); make sure they have time for the others (0.440); empathic (0.439)
E	like to listen to music (0.456); excellent ideas (0.417); paying close attention to the details (-0.479)
F	without sense of humour, the last to laugh (0.550); don't have a lot of imagination (0.417); demanding in their work (0.411)
G	think that people have good intentions (0.600); empathic towards other people's emotions (0.401); deficient in imagination (0.417).
H	use complicated words (0.508); understanding things fast (0.483)
I	always ready and well-prepared (0.432); pay much attention to details (0.472); ready for action (0.462); make people feel comfortable (0.482); do not mind being in the center of attention (-0.414)

IT students declare themselves to be extroverted (in the positive sense), stating that they are communicative, spontaneous, disinherited, talkative, sincere, humorous, and optimistic; in the negative sense: distant, quiet, reserved, shy, and inhibited.

Concerning the positive side of agreeableness, the IT students are open to cooperation, politeness, empathy, tolerance, flexibility. The main features of the Conscientiousness positive dimension are referring to organization, efficiency, trustworthiness, precision, perseverance, punctuality, logical thinking whereas the negative side refers at inconsistency, lack of attention, emotional instability or neuroticism. The positive dimension of Agreeability is a dimension characterized by calmness, independence. The Intellect or Level of Culture is characterized in the positive dimension by an intellectual mindset, profoundness, intuition, intelligence, creativity and curiosity, while the negative dimension is related to superficiality, lack of imagination, and the incapacity to understand abstract ideas.

According to the result of the study and [38, 46], we can observe that IT typology excludes the negative sides of the dimensions of being extroverted, to emotional stability or neuroticism and agreeability. However, they also have the power to admit that part of them is not empathetic, without humor, not sociable; do not understand jokes or abstract ideas.

5.4   About type of personality and job-opportunity

The experiment continued with the stage where the students were asked to extract the job announcements that appealed to them from an e-jobs platform. The purpose of this experiment was to verify whether the personality influenced the differentiated choice of jobs. To identify essential attributes that determine the choice of job-opportunity announcements characterizing the individual, we proceeded to determine the values of the knots in the graph, in order to create a classification for them. More precisely, we concentrated on the term of "importance" with the intention of determining the role played by the students' personality and the level of appreciation of attributes of the job description, to see how many these factors influenced the accumulation of information and knowledge related to the choices generated on the grounds of personality traits characterizing each participant. In order to carry out this stage, we used the classical topological measuring methods, more specifically we calculated the degree of centrality [8, 31], the closeness centrality [29] the betweenness centrality and the eigenvector centrality [8]. The theoretical background for the indicators is presented in detail in [10, 27, 29]. Based on these indicators, data was processed to obtain a topological network of importance. The findings indicate that the studied attributes of job description act as ties between personality descriptor and attributes of a job describing which influence the students in their choosing process.

Finally, according to the centrality measuring factor, we explored the place of each attribute that influenced the choice of the job announcement. As a result, we obtained three clusters:

1. One cluster comprises job announcements that have attributes related to the firm's culture, to the team organization, team building and necessary skills for the job.

2. The second cluster comprises the job announcements created from the majority of attributes related to technical skills, soft skills.

3. The third cluster comprises job advertisements mostly created from attributes related to technical skills and the second related to the culture and policy of the company; the free time.

6   Discussions

The personality characteristics defining the IT employee, as described in [51, 55] are much improved with the attributes that characterize people who are well-adapted to the social and cultural environment. The justification that we found is that the IT specialist is a person that earns a lot, having been raised in a society that is based on computer science [14]. At the same time, in an era relying on communication there is a high percentage of introverted students, that is to say, reserved persons with low levels of sociability, who tend to focus on their inner self rather than on the outside social environment. This is beneficial to the IT labor market for the reason that introverted persons

require fewer stimulants in order to reach an adequate level of excitement, a fact evidenced by the results of social network analysis (introverted persons choose job advertisements which place more emphasis on technical skills). As a consequence, they could be more suitable for software development tasks that require more complex and advanced analysis of the concepts and projection or development algorithms. Furthermore, the results of the study identify a considerable percentage of extroverted students. These persons could manifest levels of under-excitement during programming tasks [14]. This is a result of their stronger need for social interactions, which shows that they would be more suitable in positions that accompany programming with social interaction, a fact demonstrated by the result of social network analysis (extroverted persons choose job advertisements that are more concerned with the organization's culture, with team building and less with technical skills). Extroverted persons described by attitudes of conscientiousness and openness correspond to the new characteristics of education 3.0 as participation, co-creation, becoming, assuming the part of 21st century skills [38, 59]. On the contrary, the introverted ones will have adaptability issues regarding these ideas and principles.

7   Conclusion

Education is and will remain a constant preoccupation for all the parts involved in a society's development. Educational theories have evolved, one might say, together with the evolution of digital technologies.

This article examined teachers' perspectives on recent graduates' skills and employment readiness to identify the mismatch between their skills in university programs and the skills needed in the job market. Our work is in line with the research trend on increasing the quality of higher education, improving the quality of teaching and, therefore, the quality of graduates with a specific focus on the IT area and the personality of future graduates. Personality research offers great potential for providing educators and researchers with more information on the students' style of learning, based on their personality profile [46].

Furthermore, recruitment methods in the labor market are closely connected to digital technologies. Nowadays we are dealing with real recruitment social platforms in which intelligent digital techniques manage to place and analyze large amounts of data in order to assist those who recruit and also those who look for a job. The study proved that many of the participating students are characterized by high levels of personality traits such as Extroversion, Agreeability, Conscientiousness and Openness and also experience low levels of Neuroticism. The results obtained contribute to the development of knowledge in two appropriate directions. Firstly, identified profiles will be of assistance for the universities to improve their curricula and the learning and teaching processes based on the students' needs. Secondly, it is useful in this attempt to provide IT companies with information in order to better understand the IT students' profile and to create a realistic description of the job in the recruitment software on the digital market.

The importance of knowing about the personality features of the IT specialists allow both employers and educators to form teams, to create various types of projects and to build learning environments suitable for every type of employee. They can also avoid the failure of a project caused by inappropriate assignment of the human resource, and this might also help to avoid conflicts and improve the collaboration among employees or students by the correct identification of the personalities working well with each other. Our research is also useful in identifying the techniques that are appropriate to explore the students' profile and proposes data mining techniques and social network analysis to examine IT students' skills and behavior so as to generate the actual IT students' profile.

The limitations of our study are due to the fact that our involvement model was determined by analyzing a set of participants from a specific socio-economic, cultural context.

The current study offers the potential for several follow-up studies. For example, we intent to compare the profile identified in this study with profiles of the students from other universities (such as MIT, EU universities etc.) to analyze whether these characteristics are nowadays common for all students. Further research may well expand the study of identifying personality typologies of IT students, by creating a prototype of the Motivational Map for them, based on the types and degree

of motivations. The purpose of this prototype is to check whether the IT profile is directly influenced by the student's types of motivation.

Furthermore we plan to adapt our results by including a new dimension for the study: modeling emotion with Markov chains developed from [57] in order to attain a visual analysis of IT students' profile. We also intend to modernize the programming teaching techniques by implementing pedagogical agents named HAPA based on the teaching scenario adapted from [42].

This development might help companies accurately describe their vacancies and job opportunities in IT, selecting the right person for that position and identifying the techniques of communication and motivation that are adequate to each typology, so that they obtain and maintain a high percentage of employees who remain loyal to their working environment. Therefore, in our future research, the goal is to continue this analysis with the determination of the groups of personas' behaviors by applying the k-means clustering algorithm developed by MacQueen [23] and by perfecting it based on the method of the dynamic clouds [22]. Another option of extending the research targets is to analyze if and how the determination of personality traits influences the manner of choosing and interpreting job ads from employment platforms.

In conclusion, there are many actors interested in the results of this research area. Knowing the student profile might contribute to a better understanding from the academic part, and allow them to adjust their curricula, learning/teaching methods, tools and interactions to students' characteristics. The primary benefit for universities which are aware of the students' profile is to enhance their competitiveness in the educational market, also measured by their graduates' absorption on the labor market. The research results propose a solution that might synchronize the university curricula to the market's requirements. The results attempt to provide companies with information in order to understand better graduate students' profile and to create a realistic description of the jobs that will match this profile.

References

[1] Adedoyin-Olowe, M.; Gaber, M.M.; Stahl, F. (2014). A Survey of Data Mining Techniques for SocialNetwork Analysis, *Journal of Data Mining and Digital Humanities*, 2014.

[2] Alreck, P.; Settle, R. (2004). *The Survey Research Handbook. 3rd ed.*, McGraw-Hill Education, 2004.

[3] Bahner C.A., Clark C.B. (2020). Sixteen Personality Factor Questionnaire (16PF), In *Zeigler-Hill V., Shackelford T.K. (eds.) Encyclopedia of Personality and Individual Differences*, Springer 2020.

[4] Bapna, R.; Langer, N.; Mehra, A.; Gopal, R.; Gupta, A. (2013). Human capital investments and employee performance: An analysis of IT services industry, *Management Science*, 59(3), 641–658, 2013.

[5] Benta, D.; Bologa, G.; Dzitac, I. (2014). E-learning Platforms in Higher Education. Case Study, *Procedia Computer Science*, 31, 1170–1176, 2014.

[6] Benta, D.; Bologa, G.; Dzitac, S.; Dzitac, I. (2015). University level learning and teaching via e-learning platforms, *Procedia Computer Science*, 55, 1366-1373, 2015.

[7] Borden, J.; Pearson, S. (2015). Education 3.0: Embracing Technology to 'Jump the Curve', *Wired Magazine*, 2015.

[8] Borgatti, S.P. (2005). Centrality and network flow, *Social Networks*, 27(1), 55–71, 2005.

[9] Capretz, L.F.(2002). Implications of MBTI in Software Engineering Education, *ACM SIGCSE Bulletin*, 34(4), 134–137, 2002.

[10] Carrington, P.J.; Scott, J.; Wasserman, S. (2005). *Models and Methods in Social Network Analysis*, Cambridge University Press, 2005.

[11] Cattell, H; Mead, A. D. (2008). The Sixteen Personality Factor Questionnaire (16PF), *The Sage handbook of personality theory and assessment*, SAGE Publications Ltd., 135–159, 2008.

[12] Cattell, R. B.; Eber, H. W.; Tatsuoka, M.M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16PF)s*, Institute for Personality and Ability Testing, 1970.

[13] Chaoqun Ni; Sugimoto, C.R.; Jiang, J. (2011). Degree, Closeness, and Betweenness: Application of group centrality measurements to explore macro-disciplinary evolution diachronically, *Proceedings of ISSI*, 2011.

[14] Charlton, J.P.; Birkett, P.E. (1999), An integrative model of factors related to computing course performance, *Journal of Educational Computing Research*, 20(3), 237–257, 1999.

[15] Cheryan, S.; Plaut, V.C.; Handron, C.; Hudson, L. (2013), The stereotypical computer scientist: Gendered media representations as a barrier to inclusion for women, *Sex Roles*, 69, 58–71, 2013.

[16] Chopra, S.; Golab, L.; Pretti, T.J.; Toulis, A. (2018). Using data mining methods for research in co-operative education, *Asia-pacific journal of cooperative education*, 19(3), 297–310, 2018.

[17] Conroy, R. (2016). *The RCSI Sample size handbook. A rough guide*, 2016. Available: http://www.rcsi.ie/files/research/docs/20160811111051_Sample%20size%202016.pdf, Accesed on 10 May 2020.

[18] Cristea, D. (2001). *Tratat de psihologie sociala*, ProTransilvania Bucuresti, 2001.

[19] Culic, I.(2004). *Metode avansate in cercetarea sociala. Analiza multivariata de interdependenta*, Editura Polirom, 230–245, 2004.

[20] Dan C. (2013), Analiza Componentelor Principale pentru date Categoriale (CATPCA), *Psihologia Resurselor Umane Journal*, 103–117, 2013.

[21] DeYoung, C. G. (2011). Intelligence and personality, In *JR. J. Sternberg & S. B. Kaufman(Eds.), The Cambridge handbook of intelligence*, Cambridge University Press, 711–737, 2011.

[22] Dhillon I.; Modha D.(2001). Concept Decomposition for Large Sparse Text Data Using Clustering, *Machine Learning*, 42, 143–175, 2001.

[23] Enachescu, D. (2009). *Data mining, Metode si aplicatii*, Editura Academiei Romane, 2009.

[24] Everett M.; Borgatti, S.P. (2005). *Extending centrality, Models and Methods in Social Network Analysis*, Cambridge University Press, 57–76, 2005.

[25] Feldman, R.S.; Fienman, J.A. (1992). *Who You Are*, Venture Books New York, 1992.

[26] Fisher, D.; Frey, N. (2013). *Better Learning Through Structured Teaching: A Framework for the Gradual Release of Responsibility (2nd ed.)*, Association for Supervision and Curriculum Development, 2013.

[27] Fortunato, S. (2010), Community detection in graphs, *Physics Reports*, 486(3), 75–174, 2010.

[28] Fowler, S.B.; Lapp, V. (2019). Sample Size in Quantitative Research: Sample Size Will Affect the Significance of Your Research, *American Nurse Today*, 14(5), 2019.

[29] Freeman, L.C. (1979), Centrality in social networks conceptual clarification, *Social Networks*, 1(3), 215–239, 1979.

[30] Gie Yong A.; Pearce S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis, *Tutor Quant Methods Psychol*, 9, 79–94, 2013.

[31] Girvan, M.; Newman, M.N. (2002). Community structure in social and biological networks, *Proceedings of the National Academy of Sciences*, 99(12), 7821–7826, 2002.

[32] Goldberg L.R.; Johnson, J.A.; Eber, H.W.; Hogan, R.; Ashton M.C.; Cloninger, C.R.; Gough, H.G.(2006). The International Personality Item Pool and The Future of Public-Domain Personality Measures, *Journal of Research in Personality*, 40, 84–96, 2006.

[33] Goldberg L.R. (1990). An alternative "description of personality": The Big-Five factor structure, *Journal of Personality and Social Psychology*, 59, 1216–1229, 1990.

[34] Goldberg L.R.; Sweeney, D.; Merenda, P.F.; Hughes Jr, J.E. (1996). The Big Five Factor Structure as an Integrative Framework: An analysis of Clarke's AVA Model, *Journal of Personality Assessment*, 66(3), 441–71., 1996.

[35] Goldberg L.R.; Sweeney, D.; Merenda, P.F.; Hughes Jr, J.E. (1999). A Broad-Bandwidth, Public-Domain, Personality Inventory Measuring the Lower-Level Facets of Several Five-Factor Model, *Personality Psychology in Europe*, 7, 2–28, 1999.

[36] Goodwin, K. (2008). Perfecting your personas, *The Cooper Journal*, 2008.

[37] Halasz, G.; Michel, A. (2011). Key Competencies in Europe: interpretation, policy formulation and implementation, *European Journal of Education*, 46, 289–306, 2011.

[38] Halpern D.F. (2000). Creating cooperative learning environments, *APS Observer*, 2000.

[39] Hambleton, R.K.; Swaminathan, H.; Rogers, H.J. (1991). *Fundamentals of Item Response Theory*, Sage Press, 1991.

[40] Han-Haas, G.H.; Peng G. (2018). Big Data Analytics in Human Resource Management and Its Impact: Theoretical Development and Empirical Results, *Twenty-fourth Americas Conference on Information Systems*, New Orleans, 2018.

[41] Horn, M. (2014). KAIST Does not Wait For Change In Korea, Pioneers 'Education 3.0, *Forbes Magazine*, 2014.

[42] Ivanovic, M.; Mitrovic, D.; Budimac, Z.; Jerinic, L.; Badica, C. (2015). HAPA: Harvester and pedagogical agents in e-learning environments, *International Journal of Computers Communications & Control*, 10(2), 200–210, 2015.

[43] Johnson L.; Becker A. S.; Cummins, M.; Freeman, A.; Ifenthaler, D.; Vardaxis N. (2013). *FTechnology Outlook for Australian Tertiary Education 2013-2018: An NMC Horizon Project Regional Analysis*, New Media Consortium, 2013.

[44] Karimi, Z.; Baraani-Dastjerdi, A.; Ghasem-Aghaee, N.; Wagner, S. (2016). Links between the personalities, styles and performance in computer programming, *Journal of Systems and Software*, 111, 228–24, 2013.

[45] Kim, E B.; Schniederjans, M. J.(2004). The Role of Personality in Web-Based Distance Education Courses, *Communications of the ACM*, 47(3), 95–98, 2004.

[46] Krause, L. (2000). *How We Learn and Why We Don't*, Thomson Learning, 2000.

[47] Kuncel, N.R.; Hezlett, S.A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions, *Current Directions in Psychological Science*, 19, 339–345, 2010.

[48] Kurekova, L.M.; Beblavy, M.; Haita C.; Thum A.E. (2016). Employers' skill preferences across Europe: between cognitive and non-cognitive skills, *Journal of Education and Work*, 29(6), 662–687, 2016.

[49] Ledward, B. C.; Hirata D. (2011). *An Overview of 21st Century Skills. Summary of 21st Century Skills for Students and Teachers*, Pacific Policy Research Center, Honolulu: Kamehameha Schools-Research & Evaluation, 2011.

[50] Lindsey, W. H. (2011). *The Relationship Between Personality Type and Software Usability Using the Myers-Briggs Type Indicator(MBTI®) and the Software Usability Measurement Inventory (SUMI), Doctoral dissertation*, Nova Southeastern University, 2011.

[51] McConnell, S. (1999). *After the Gold Rush: Creating a True Profession of Software Engineering*, Microsoft Press, 1999.

[52] McCrae, R. R.; Greenberg, D. M. (2014). Openness to experience, In *D. K. Simonton (Ed.), The Wiley Handbook of Genius Chisester*, Wiley, 222–243, 2014.

[53] Michel W.; Shoda Y.; Smith R. E. (2004). *Introduction to Personality: Towards an Integration*, John Wiley, 2004.

[54] Moisil, I.; Pitic, A.; Dzitac, S.; Popper, L. (2010). Adaptive Web Applications for Citizens' Education. Case Study: Teaching Children the Value of Electrical Energy, *International Journal of Computers Communications & Control*, 5(5), 819–824, 2010.

[55] Myers, I. B.; McCaulley M. H. (1992). *Manual: A guide to the development and use of the Myers-Briggs type indicator*, Consulting Psychologists Press, 1992.

[56] Oram, A., Wilson, G. (2010). *Making software: What really works, and why we believe it*, O'Reilly Media, Inc., 2010.

[57] Pitic, A. E.; Moisil, I.; Dzitac, S. (2013). Raising energy saving awareness through educational software, *International Journal of Computers Communications & Control*, 8(2), 255–262, 2013.

[58] Popa, M. (2008). *Statistica pentru psihologi. Teorie si aplicatii SPSS*, Colectia Collegium, Psihologie, 2008.

[59] Rubens, N.; Kaplan, D.; Okamoto, T. (2012). E-Learning 3.0: anyone, anywhere, anytime, and AI, In *Chiu D.K.W., Wang M., Popescu E., Li Q., Lau R. (Eds.) New Horizons in Web Based Learning. ICWL 2012. LNCS*, Springer, 171–180, 2012.

[60] Shaffer, D.W. (2004). Epistemic frames and islands of expertise: Learning from infusion experiences, In *Y. Kafai, W.A. Sandoval, N. Enyedy, A.S. Nixon, F. Herrera (Eds.) Proceedings of the Sixth International Conference of the Learning Sciences*, 473–480, 2004.

[61] Soto, C.J.; Kronauer, A.; Liang, J.K. (2015). Five-Factor Model of Personality, In *S.K. Whitbourne (Ed.) The Encyclopedia of Adulthood and Aging*, 2015.

[62] Tieger, P. D.; Barron-Tieger B. (2001). *Do What You Are*, Little, Brown and Company, 2001.

[63] Trilling, B.; Fadel, C. (2009). *21st Century Skills: Learning for Life in Our Times*, John Wiley & Sons, 2009.

[64] Verma, V. (1991). *Sampling Methods. Training Handbook*, Statistical Institute for Asia and the Pacific, World Bank, 1991.

[65] Waksberg, J. (1978). Sampling Methods for Random Digit Dialing, *The Journal of the American Statistical Association*, 73, 40–46, 1978.

[66] Wall, L. (2000). *DProgramming Perl, 3rd Edition*, O'Reilly, 2000.

[67] Wasserman, S.; Faust, K. (1994). *Social network analysis: Methods and applications*, Cambridge University Press, 1994.

[68] Weinberg, G.M. (1998). *The psychology of computer programming. 2nd ed.*, Van Nostrand Reinhold. xv. 288, 1998.

[69] Williams, B.; Onsman, A.; Brown, T. (1996). Exploratory factor analysis: A five-step guide for novices, *J Emerg Prim Health Care*, 19, 42–50, 1996.

[70] Woszczynski; A. B., Guthrie; T. C.; Shade, S. (2005). Personality and programming, *Journal of Information Systems Education*, 16(3), 293–299, 2005.

[71] World Bank, (1999). *Core Welfare Indicators Questionnaire (CWIQ) Handbook, Chapter 4: Preparing the CWIQ Sample Design*, World Bank, 1999.

[72] Yorke, M; Knight, P.T. (2004). Embedding employability into the curriculum, *Learning & Employability*, 3, 1–28, 2004.

C | **O** | **P** | **E**

Member since 2012
JM08090

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Stanca, L.; Lacurezeanu, R.; Tiron-Tudor, A.; Bresfelean, V.P.; Pandelica, I. (2020). Determining IT Student Profile Using Data Mining and Social Network Analysis, *International Journal of Computers Communications & Control*, 15(5), 3897, 2020. https://doi.org/10.15837/ijccc.2020.5.3897