

Urszula MARKOWSKA-PRZYBYŁA<sup>1</sup>

David M. RAMSEY<sup>2</sup>

## A GAME THEORETICAL STUDY OF GENERALIZED TRUST AND RECIPROCATION IN POLAND. II. A DESCRIPTION OF THE STUDY GROUP

The first article describing this project presented the three games that the participants played: the Ultimatum Game, the Trust Game and the Public Goods Game. This article describes the study group on the basis of a questionnaire regarding where they study and come from, their social contacts, interest in current issues, views on inequality and outlook on life. A description of the migratory decisions of students is given. In particular, two exploratory methods are used to investigate the data's structure: Bayesian networks and principal component analysis. Bayesian networks are used to illustrate the associations between categorical variables. Principal component analysis is designed to describe latent variables which reflect the associations between numerical variables. We present the results of this analysis and discuss the advantages and disadvantages of these two methods.

**Keywords:** *Bayesian networks, principal component analysis, social survey, experimental game theory*

### 1. Introduction

This article is the second in a series on a project studying the level of generalised trust and reciprocation among Polish students using experimental game theory. The first article [15] presented the games that the students played: the Ultimatum Game [9], the Trust Game [2], the Public Goods Game [10] and how these games were played in our study. This article describes the questionnaire used alongside the games as well as the

---

<sup>1</sup>Faculty of Economics, Management and Tourism, Wrocław University of Economics, ul. Kochnowskiego 8, 58-500 Jelenia Góra, e-mail: ump@ue.wroc.pl

<sup>2</sup>Department of Computer Science and Management, Wrocław University of Technology, ul. Łukasiewicza 5, 50-371 Wrocław, e-mail: david.ramsey@pwr.edu.pl

study group with the aid of both Bayesian networks and principal component analysis (PCA).



Fig. 1. Map of the regions of Poland; as in the text, the Polish names of the regions are used.  
Based on: [d-maps.com/carte.php?num\\_car=4309&lang=en](http://d-maps.com/carte.php?num_car=4309&lang=en) (accessed 6/11/2014)

The study was carried out at state universities in a capital of each of the sixteen Polish regions (Fig. 1). The 16 regional capitals used (in order of descending population, based on data from the Polish Statistical Office for the end of 2013 [22]) were: Warsaw (Mazowieckie), Cracow (Małopolskie), Łódź (Łódzkie), Wrocław (Dolnośląskie), Poznań (Wielkopolskie), Gdańsk (Pomorskie), Szczecin (Zachodnio-Pomorskie), Lublin (Lubelskie), Katowice (Śląskie), Białystok (Podlaskie), Toruń (Kujawsko-Pomorskie), Kielce (Świętokrzyskie), Rzeszów (Podkarpackie), Olsztyn (Warmińsko-Mazurskie), Opole (Opolskie), Zielona Góra (Lubuskie). Note that the Lubuskie and Kujawsko-Pomorskie have two capitals. How the study was conducted is described in Section 2.

Section 3 presents the answers given in the questionnaire, which covered:

- The subject studied: it has been noticed that, in particular, economics students show different behaviour in such settings [13], as they may have come into contact with the concepts of game theory. Also, it has been noted that younger students may show more pro-social behaviour in such a setting [8].

- Where the student comes from and their feeling of attachment to the region in which they live/study. Lewicka [14] notes that the way that people bond to the place in which they live is strongly related to their character and outlook on life.

- The expressed level of generalised trust, membership in social organisations and social contacts. Growiec [7] describes social capital in terms of the strength of links with family (bonding capital) and acquaintances (bridging capital). The relation between these forms of capital influences an individual's decision on whether to move, which in turn effects individual's social contacts. Putnam [17] understands generalised trust to be the degree to which one expects unknown people to exhibit positive reciprocation or behaviour that is beneficial to the group as a whole, even at the risk of a personal loss. He argues that membership in social organisations is positively associated with generalised trust and, to a large extent, understands social capital in terms of involvement in civic matters.

- Interest in current affairs at national, regional and local levels. For example, Norris [16] considers the relation between social capital and access to information via the media.

- General values (e.g. relative importance of ethical and legal norms, attitude towards inequality, expressed readiness to reciprocate and level of individualism). These questions are designed to elicit the attitudes of students, as well as their readiness to adopt these attitudes in practice. Kaiser et al. [11] note that individuals may express the same attitudes for different reasons and that the same observed behaviour can result from different reasons. Hence, attitudes and behaviour should be seen as formally, not causally, related.

Section 4 gives a brief description of the methods used to explain the relations between the variables observed in the questionnaire: Bayesian networks and PCA. Section 5 presents the results of the analysis using the theory of Bayesian networks and relates these to the results of standard tests of association.

Section 6 presents the results of PCA, along with an interpretation of the components obtained in terms of types of students. The final section sums up the advantages and disadvantages of the methods used and the results of this analysis.

## **2. How the study was conducted**

The study was carried out at the state university in each of the 16 cities listed between 16.04.2014 and 12.06.2014. At each site, 88–100 students participated in the study (Table 2). The studies were conducted by the EU-CONSULT Ltd. under the su-

pervision of Dr. Urszula Markowska-Przybyła and Ewa Starczewska, both from the Faculty of Economics, Management and Tourism of the Wrocław University of Economics, based in Jelenia Góra, where a pilot study was observed by the authors on 11.03.2014. At each site, the students were split into three or four sessions, run in immediate succession. Statistical analysis indicated that the decisions made in the games neither depended on the session number, nor the number of participants in a session. Each session lasted about one hour. On average, students obtained a payoff of around 45 PLN (approx. 11 €).

All decisions and questionnaires were written on forms which were coded to identify the player and their “opponents” in the relevant games. In each session, students were split randomly into two groups (without knowing which group any of the other players were in). They first made their decision in the Public Goods Game and the decision of Player 1 in the game appropriate to their group (the Trust Game or the Ultimatum Game). The players then obtained the appropriate instructions for Player 2 in the game they had not played yet, together with the decision of the student randomly assigned to them (who played the role of Player 1). These decisions were collected and the payoffs calculated, while the students filled in the questionnaire. Filling in the questionnaire took from 20 to 30 minutes and was required for the students to obtain their payoff. These decisions were then collected, the individual payoffs calculated and the students paid.

It should be noted that many of the answers given by the students are subjective in nature. It is possible that these answers may have been influenced by the decisions he/she made in the game (e.g. in order to justify them). A more serious problem could be that the decisions made by other players might influence the answers given by a student. It should be noted that, at the time the students filled in the questionnaire, the only such relevant information regarded the game in which a student was assigned the role of Player 2 (i.e. either in the Ultimatum Game or in the Trust Game). Analysis of the associations between the actions of the “opponent” in these games and the subjective answers given by the students showed that there was only one borderline significant association at the 5% level. The amount transferred by Player 1 to Player 2 in the Trust Game is negatively associated with Player 2’s interest in local affairs (Spearman’s test of association,  $r = -0.072$ ,  $p = 0.05$ ). Since there is no clear interpretation of such an association and this was the only significant result obtained from nearly 50 tests, it is reasonable to assume that there is no association between the answers obtained in the questionnaire and the behaviour of “opponents”.

### **3. The variables observed in the questionnaire**

The variables observed in the study are listed in Table 1 and described in this section. Some studies have indicated that females act more pro-socially than males [5]. Also, students of economics or mathematical subjects are most likely to have had contact

Table 1. Variables observed in the study

Binary	Nominal	Ordinal/numerical	Ordinal/numerical
Sex	City	Year of Study	No. of organisations
Form of course	Subject studied	Size of home town	Family
Attachment A	Type of course	National	Friends
Preference	Home region	Regional	Acquaintances
Voluntary work	Attachment B	Local	Reciprocation (public)
Organisation membership	Reason A	Trust	Reciprocation (private)
	Reason M	Tax evasion (attention)	Intention
	Strategy	Tax evasion (distaste)	Norms
		Reaction to inequality	

Source: The authors' own work.

with the concepts of game theory [13]. There were 96 economics students (6.2% of the total sample) and 26 mathematics/physics students. All the studies were carried out at universities rather than more specialist institutions, in order to obtain a wide profile of students in each city. However, there were more than twice as many economics students in Białystok (24) than in any other city (10 or less). Due to space restrictions, the numbers studying each subject are not given here. The types of course were: bachelor – 983 students (63.8%), masters – 384 (24.9%), unified – 169 (11.0%) and doctoral – 3 (0.2%). The form of the course was either “day studies” or “non-day studies”. Day students are full time students who have classes during the week. Non-day students have classes at weekends. Mature students are much more common among the non-day students than among the day students. The study group was fairly uniform with respect to this variable (only 42 non-day students, 2.7%). Since students wrote the year of their course (rather than total years of study), in order to get a proxy for the year of study, when appropriate 3 years were added in the case of a masters student and 5 in the case of a doctoral student. According to this, the distribution of the year of study is as follows: year 1 – 397 (26.0%), year 2 – 359 (23.5%), year 3 – 335 (21.9%), year 4 – 257 (16.8%), year 5 – 173 (11.3%), Years 6 to 8 – 8 (0.6%).

### 3.1. Demographics of the study group

A summary of the demographic variables is given in Table 2. The home region is the region in which a student was registered as being resident immediately before starting their studies (one of the sixteen Polish regions corresponding to the cities in which the study was carried out). The place of study and home region were used together to define a 0–1 variable taking the value 1 if and only if the student moved to a different region to study (Move Away). The size of a student's home town was categorised as follows: 1 – up to 5000 inhabitants, 2 – between 5000 and 20 000, 3 – be-

tween 20 000 and 100 000 and 4 – above 100 000. The cities included in the study are all in category 4.

Table 2. Demographic statistics for the study group

City	Number of						Net immigration
	Students	Females	Coming to region	Registered in region	Females	Leaving region	
Białystok	96	72 (75%)	9 (9.4%)	113	81 (71.7%)	26 (23.0%)	-17
Gdańsk	100	77 (77%)	34 (34%)	83	62 (74.7%)	18 (21.7%)	17
Katowice	96	72 (75%)	10 (10.4%)	138	99 (71.7%)	52 (37.7%)	-42
Kielce	100	68 (68%)	16 (16%)	104	74 (71.2%)	20 (19.2%)	-4
Cracow	96	67 (69.8%)	53 (55.2%)	52	41 (78.8%)	12 (23.1%)	44
Lublin	88	68 (77.3%)	5 (5.7%)	108	83 (76.9%)	25 (23.1%)	-20
Łódź	100	75 (75%)	14 (14%)	114	88 (77.2%)	28 (24.6%)	-14
Olsztyn	96	56 (58.3%)	28 (29.2%)	94	60 (63.8%)	27 (28.7%)	2
Opole	96	74 (77.1%)	40 (41.7%)	69	52 (75.4%)	13 (18.8%)	27
Poznań	100	77 (77%)	34 (34%)	90	67 (74.4%)	24 (26.7%)	10
Rzeszów	100	79 (79%)	15 (15%)	108	80 (74.1%)	23 (21.3%)	-8
Szczecin	88	60 (68.2%)	22 (25%)	88	61 (69.3%)	22 (25%)	0
Toruń	100	80 (80%)	22 (22%)	100	78 (78%)	22 (22%)	0
Warsaw	88	70 (79.5%)	39 (44.3%)	79	62 (78.5%)	36 (45.6%)	9
Wrocław	96	73 (76%)	43 (44.8%)	77	58 (75.3%)	24 (31.2%)	19
Zielona Góra	100	51 (51%)	14 (14%)	112	63 (56.3%)	26 (23.2%)	-12
Total	1540	1119 (72.7%)	398 (25.8%)	1529	1109 (72.5%)	398 (26.0%)	

Source: The authors' survey.

Eleven students were not registered in any region (e.g. lived outside Poland). The proportion of females was well above 60% except in Olsztyn and Zielona Góra. This may be explained by the facts that according to the OECD [20], about 60% of young Polish university graduates are female and females are more likely to volunteer for surveys [19].

The migration rates quoted do not give a full reflection of the number of students travelling between regions to study. For example, the Opolskie is the smallest region (both in population and area) and the university is one of the smallest. A large majority of the students coming to Opole from other regions are from neighbouring regions. On the other hand, the University of Warsaw and the Jagiellonian University in Cracow are the largest universities in Poland and attract students from all over the country. Although the proportion of students from other regions at the University of Wrocław is very similar to the University of Warsaw, its smaller size and region of attraction means that it should be categorized with Adam Mickiewicz University in Poznań and the University of Gdańsk in terms of attractiveness. Net immigration also gives an indication of the attractiveness of a university and its region. Due to their

smaller size and the low proportion of students from other regions, the following universities may be classified as provincial: Maria Curie-Skłodowska University in Lublin, the University of Białystok, Rzeszów University, the Jan Kochanowski University in Kielce and the University of Zielona Góra. Apart from Zielona Góra, these universities are in the less developed, more rural east of the country. The University of Silesia in Katowice and the University of Łódź are both large universities in urban regions with a small proportion of students from other regions. This probably reflects the fact that both cities are commonly regarded as unattractive [6, 12], as these universities are placed in the top half of the rankings of Polish universities according to Perspektywy [21].

Since the number of participants from each university is roughly the same, the emigration rates will tend to exaggerate the level of migration from larger regions (e.g. the two highest emigration rates correspond to the two most populated regions: Mazowieckie and Śląskie).

As described in the introduction, our survey also measures various components of social capital, namely: contact with family, friends and acquaintances, expressed level of generalised trust, membership in social organisations, interest in current affairs, norms of reciprocation, norms of redistribution, relation between legal and ethical norms, attitudes on the benefits of cooperation, feeling of belonging. We now describe these factors in turn.

### 3.2. Social contact and generalised trust

Students described their level of social contact with family, close friends<sup>3</sup> and acquaintances on a seven-point scale: 1 – never, 2 – less than once a month, 3 – once a month, 4 – two/three times a month, 5 – once a week, 6 – a few times a week, 7 – daily.

In our study, expressed generalised trust is measured by the answer, on a five-point scale, to the following question: can the majority of people be trusted? The possible answers were: 1 – no, 2 – generally no, 3 – I do not know, 4 – generally yes, 5 – yes. The distributions of the answers to these questions are given in Table 3.

Compared to the results from the European Social Survey [7], students are unsurprisingly more socially active than the population in general. The level of expressed generalised trust seems very similar to that in the general population (the mean level is slightly below the midpoint of the scales used).

---

<sup>3</sup>The concept of a close friend (*przyjaciół*) is quite specific in the Polish language and in sociological terms can be seen as lying between bonding and bridging capital.

Table 3. Distributions of social contact and expressed generalised trust

Contact with family		Contact with close friends		Contact with acquaintances		Generalised trust	
1	4 (0.3%)	1	9 (0.6%)	1	4 (0.3%)	1	192 (12.5%)
2	139 (9.0%)	2	49 (3.2%)	2	98 (6.4%)	2	644 (41.8%)
3	205 (13.3%)	3	52 (3.4%)	3	89 (5.8%)	3	92 (6.0%)
4	427 (27.8%)	4	157 (10.2%)	4	251 (16.3%)	4	558 (36.2%)
5	239 (15.5%)	5	226 (14.7%)	5	219 (14.3%)	5	54 (3.5%)
6	232 (15.1%)	6	688 (44.8%)	6	502 (32.7%)		
7	291 (18.9%)	7	355 (23.1%)	7	373 (24.3%)		
Missing	3		4		4		0

Source: The authors' survey.

### 3.3. Interest in current affairs

Interest in national, regional and local affairs is measured on a five-point scale: 1 – very uninterested, 2 – rather uninterested, 3 – neither interested nor uninterested, 4 – rather interested, 5 – very interested. The distributions of these variables are given in Table 4.

Table 4. Interest in current affairs

National matters		Regional matters		Local matters	
1	52 (3.4%)	1	76 (5.0%)	1	71 (4.6%)
2	146 (9.5%)	2	289 (18.8%)	2	175 (11.4%)
3	439 (28.6%)	3	529 (34.5%)	3	368 (24.0%)
4	557 (36.3%)	4	476 (31.0%)	4	492 (32.1%)
5	340 (22.2%)	5	164 (10.7%)	5	429 (27.9%)
Missing	6		6		5

Source: The authors' survey.

### 3.4. Membership in social organisations

555 of 1532 respondents (36.2%) stated that they belonged to some form of social organisation. 685 of 1531 respondents (44.7%) stated that they carried out voluntary work over the past year. Multiple memberships occurred: 340 individuals were members of one group, 144 were members of two groups, 55 were members of three groups and 18 were members of four or more groups. The questionnaire also asked about the types of organisation an individual was a member of, but this is not considered here.



### 3.5. Norms of reciprocation

Two questions were asked on negative reciprocation. The first asks whether students would call the relevant authority if they saw that public property was being damaged (the variable “public”), on a four-point scale: 1 – never, 2 – very rarely, 3 – sometimes, 4 – generally, yes. The second asked how students would react if someone acted unfairly towards them (the variable “private”), on a four-point scale: 1 – I would just carry on as normal, 2 – I react if it does not cost me anything, 3 – I react if the costs of reaction are small, 4 – I would react, even if that means changing plans and incurring significant costs. Hence, the first question refers to reciprocation in public matters, the second to reciprocation in private matters. The distributions of these expressed levels of reciprocation are illustrated in Table 5.

Table 5. Distribution of expressed negative reciprocation

Public matters		Private matters	
1	83 (5.4%)	1	450 (29.3%)
2	380 (24.7%)	2	633 (41.2%)
3	639 (41.6%)	3	301 (19.6%)
4	434 (28.2%)	4	151 (9.8%)
Missing	4		5

Source: The authors' survey.

### 3.6. Norms of redistribution

There were three questions regarding redistribution. The first asked how much attention participants paid to the possibility that people paid lower taxes than they should. It should be noted the answer may be highly subjective, since many students would not distinguish between tax avoidance (lowering tax bills via legitimate means) and tax evasion (not paying the taxes required by law). The answer was on a four-point scale: 1 – no attention at all, 2 – little attention, 3 – quite a lot of attention, 4 – a lot of attention. The second asked the participants whether they approved of the fact that some people paid lower taxes than they should. This answer was also on a five-point scale: 1 – like it a lot, 2 – rather like it, 3 – neither like nor dislike, 4 – rather dislike it, 5 – do not like it at all. This question asks the respondent to express their distaste of tax evasion (the higher the score, the greater the distaste). The third question asks the respondent what they thought of economic inequality. The answer was on a three-point scale: 1 – inequality results from the free market and thus is just, 2 – inequality is inevitable and, to some degree, even good, but should be controlled by the government, 3 – the government should act to minimize inequality. The higher the score on

this scale, the stronger the support for government intervention and the lower the support for the free market was. The distributions of these answers are given in Table 6.

Table 6. Norms for redistribution

Attention to tax evasion		Distaste for tax evasion		Attitude to economic inequality	
1	126 (9.0%)	1	68 (4.9%)	1	296 (20.1%)
2	319 (22.9%)	2	64 (4.6%)	2	769 (52.3%)
3	642 (46.1%)	3	272 (19.4%)	3	405 (27.6%)
4	307 (22.0%)	4	573 (41.0%)		
		5	422 (30.2%)		
Missing	146		141		70

Source: The authors' survey.

### 3.7. Attitudes regarding cooperation and the relation between legal and ethical norms

Respondents were asked two closed questions on what type of behaviour they viewed as being most likely to bring success. The first of these questions was as follows: Which of the following types of strategy is most likely to bring success? (write × in the appropriate box below – only one strategy type should be indicated), see Table 7.

Table 7. Possible answers to the question:  
*Which of the following types of strategy is most likely to bring success?*

Acting	Primarily individual effort, a minimum of cooperation with others	Cooperation with others
In line with the law		
On the edge of the law or even beyond it		

Source: The authors.

The four possible types of strategy are referred to as: co-op/leg – in cooperation with others and in line with the law, ind/leg – individualistic and in line with the law, co-op/edge – in cooperation with others and on the edge of legality, ind/edge – individualistic and on the edge of legality. The first dimension indicates whether a student regards cooperation or individual effort as being a key to success. The second dimension indicates whether a student feels that it is easier to achieve success by acting on the edge of the law or not. The second question asked whether participants intended to follow such a strategy, measured on a five-point scale: 1 – no, 2 – probably not, 3 – do

not know, 4 – probably, 5 – yes. This is aimed at measuring the discomfort that students feel when their own strategy is not in line with what is seen to bring success in real life. For example, some might say that acting on the edge of the law is most likely to bring success, but do not intend to follow such a strategy. This indicates that they see people often achieve success by acting in such a way, but they themselves would not use such a strategy, as e.g. it is not in line with their ethics or they fear the consequences of adopting such a strategy. It should be noted that a number of students highlighted two types of strategy, although it was stressed that only one answer was required. In these cases, the answers to the two questions above were assumed to be missing completely at random.

The participants were also asked to state which of ethical and legal norms were the most important on the following three-point scale: 1 – legal norms, even when in conflict with ethical norms, 2 – ethical norms, as long as there is no major legal punishment, 3 – ethical norms. The distributions of the answers to these questions are presented in Table 8.

Table 8. Attitudes regarding cooperation and relation between legal and ethical norms

Optimal strategy		Intention to use strategy		Importance of ethical norms	
Co-op/leg	716 (55.2%)	1	77 (5.9%)	1	164 (10.7%)
Ind/leg	207 (16.0%)	2	116 (8.9%)	2	967 (62.9%)
Co-op/edge	256 (19.7%)	3	137 (10.6%)	3	406 (26.4%)
Ind/edge	118 (9.1%)	4	601 (46.3%)		
		5	366 (28.2%)		
Missing	243		243		3

Source: The authors' study.

### 3.8. Feeling of attachment to home region and choice of place of study

The students were asked if they had a preference for their own region in the sense that they would buy a product from their home region rather than another if there were no other factors (yes/no answer, variable preference). Of the 1537 students who answered, 716 said yes (46.6%). Students studying in their home region were asked if they felt attached to that region (yes/no answer – attachment A). Of these 1131 students, 1066 (94.3%) answered yes. Students who left their home region were asked if they felt more attached to their home region, the region in which they study or neither (attachment B). Of these 379 students, 184 (48.5%) felt more attached to their home region, 133 (35.1%) to the region in which they study, while 62 (16.4%) did not feel more attached to either region. This could indicate attachment to both regions or neither, which was investigated by the next question. All students were asked what their attachment resulted from (reason A). The possible answers were: 1 – I was born here,

2 – I have lived here for some time, 3 –I have lived most of my life here, 4 – I was not brought up here, but intend to live here, 5 – I am not attached to any region, 6 – other (please state). The first answer is an expression of the importance of bonding capital, while answers 2, 4 and 6 stress bridging capital. Those expressing no attachment are expected to have low levels of social capital [14]. Students who moved to a different region were asked why they moved (reason M). They could state that the courses and level available in the region were not appropriate, or give another reason. The answers to the last two questions above are summarised in Table 9. Those who study in another region for reasons of cost/convenience probably live close to the border between regions attend a university which is closer and/or easier to get to than the regional university.

Table 9. Reasons for attachment to a region and reasons for moving

Reason for attachment to given region		Reason for moving to another region	
Birth	1110 (72.8%)	Best course available	262 (64.7%)
Spent some time here	51 (3.3%)	Not accepted by first choice	12 (3.0%)
Spent most of my life here	125 (8.2%)	The city	17 (4.2%)
I want to spend my future here	105 (6.9%)	Friends/family/past	21 (5.2%)
No attachment	97 (6.4%)	Independence	34 (8.4%)
Attachment to two regions	12 (0.8%)	Cost/convenience	10 (2.5%)
Other	25 (1.6%)	Work/self-development	17 (4.2%)
		Change	24 (5.9%)
Missing	15 (1.0%)	Other/unknown	8 (2.0%)

Source: The authors' survey.

#### 4. The methods of analysis

Since we are dealing with a data set containing a relatively large number of mutually correlated variables, it is difficult, if not impossible, to describe the relationship between these variables using standard tests of association. For this reason, we decided to use two exploratory techniques designed to illustrate the associations between large numbers of variables: Bayesian networks [3] and PCA [4].

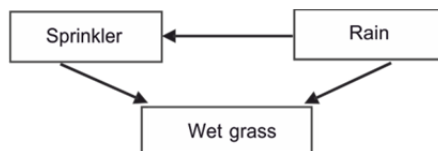


Fig. 2. Example of a Bayesian network.

Source: [en.wikipedia.org/wiki/Bayesian\\_network](http://en.wikipedia.org/wiki/Bayesian_network)

Bayesian networks are used to describe the relationship between categorical variables via a directed, acyclic graph, although discrete variables taking a small number of values are also suitable. If variable  $A$  influences variable  $B$ , then there is a directed edge between the two variables. For example, grass may be wet due to rain or a sprinkler. The sprinkler is only used when it is dry. Figure 2 illustrates a very simple Bayesian network.

In this model, the rain variable is said to be a parent node, since no other variable is assumed to influence it. Let  $S$ ,  $R$ ,  $W$  denote the events that on a given day: a) the sprinkler has been used, b) it has rained, c) the grass is wet. The model above leads to the following model for the joint distribution of these three events:

$$P(S, R, W) = P(R)P(S|R)P(W|S, R)$$

This can be used to calculate the likelihood of the data under the assumed model.

In our analysis we used the *catnet* package implemented in the R program by Balov and Salzman [1]. They adopt the Bayesian information criterion (BIC) and the Akaike information criterion (AIC) to select a model. These criteria impose a penalty for the complexity of the model in terms of the number of parameters used. Salzman and Almudevar [18] propose estimates of this complexity for a given model. The BIC minimises the expression  $-2 \ln \hat{L} + k \ln(n)$ , where  $\hat{L}$  is the likelihood of the data under a given model,  $k$  is the number of parameters and  $n$  is the sample size. The AIC minimises the expression  $-2 \ln \hat{L} + k$ . Since the BIC places a higher penalty on complexity, it tends to suggest simpler models.

The *catnet* package allows the user to create a hierarchy of variables, so that a variable cannot be influenced by one lower on the list. For our data, sex and home region should come at the top of this list, since the other variables cannot influence these two. However, some pairs of variables can show feedback. For example, an individual's contact with their family influences a student's decision on whether to move to a different region which, in turn, affects contact with their family. For this reason, no hierarchy was defined.

The package also allows the user to define the maximum number of parent nodes in the graph and the complexity of the model. Since the problem of finding the optimal network is NP-hard, these parameters have a highly significant influence on the run time.

PCA is designed to explain a number  $m$  of numerical variables using a smaller number of components describing a large proportion of the variation. It is based on the eigenvalues of the correlation matrix  $\Sigma$ . Denote the  $i$ -th largest eigenvalue by  $\lambda_i$ . The proportion of the variance explained by the  $i$ -th principal component, which corresponds to the  $i$ -th eigenvector, is equal to  $\lambda_i/m$ . This follows as the sum of the eigenvalues must equal the trace of  $\Sigma$  (the sum of the elements along the leading diagonal). The principal components explaining the most variation can be used to define latent

variables, which often have intuitive interpretations. One advantage of these latent variables is their independence [4].

Hence, for these methods to be effective, variables must be correlated. However, they should not be correlated by definition (e.g. in our survey, year of study and type of course).

## 5. Analysis of the characteristics of the study group via Bayesian networks

The variables used in this analysis are those listed in Table 1 except for the following: a) No. of organisations and Type of Course (since they are correlated by definition with other variables included), b) attachment A, attachment B and reason for moving (since they were not asked to all participants) c) subject studied, form of course (too much and too little variation, respectively). The binary variable defining whether a student moved away was also used. Hence, in total 25 variables were used. Using BIC, allowing the number of parent nodes to be at most four and defining a maximum complexity large enough not to be a constraint, the Bayesian network illustrated in Fig. 3 was found to be optimal.

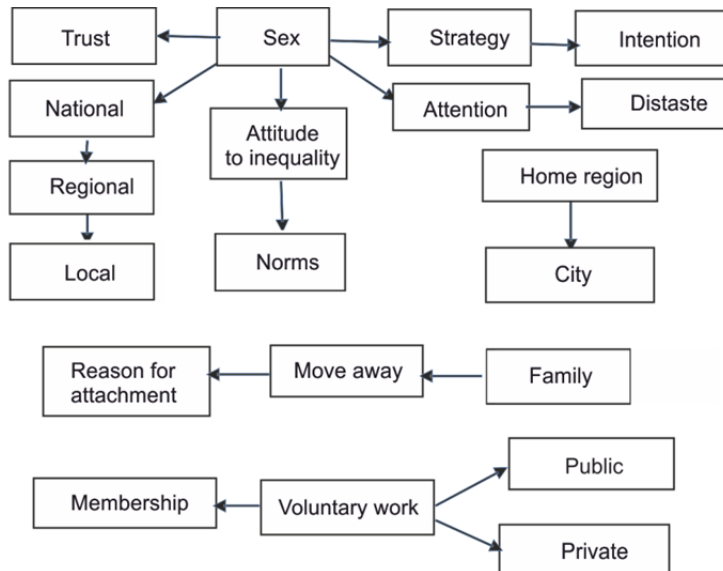


Fig. 3. Bayesian network based on the Bayes Information Criterion. Source: Authors' analysis using the *catnet* package in the R program

It should be noted that for our data, when the maximum number of parent nodes was set to be five, six and seven, the run times on a standard laptop were several a) seconds, b) minutes, c) hours, respectively. The direction of influence was defined by the algorithm, except in cases where it should obviously be in the opposite direction (e.g. home region influenced the city in which a student studies, not vice versa).

From the model presented in Fig. 3, there are four networks of association between the variables studied and 20 variables are included. In one network, the city in which a student studies and the region from they come are associated, but independent of the remaining variables. From the analysis in Section 3.1, the majority of students study in their home region. In the second network, moving away to study is associated with both a student's reason for attachment to a region and contact with their family. Moving to a different region has a large impact on a student's social network. Almost by definition, contact with one's family will decrease. It is possible that friends also move to study in the same city. However, students who move to a different region will often have to build up a new network of contacts with non-relations. This is reflected in the results from the study. Moving away to another region has a very significant negative association with contact with family members ( $p < 0.001$ , the Mann–Whitney test). It also has a negative association with contact with friends and a positive association with contact with acquaintances (although these associations are only significant at the 10% level). The reasons for attachment I was born there or I have lived there most of my life are associated with students who have not moved away to study ( $p < 0.001$ , likelihood test of association). Given these associations, it is unsurprising that those who gave place of birth as their reason for attachment had most contact with their family ( $p < 0.001$ , the Kruskal–Wallis test).

The third network of associations is between membership of an organisation, voluntary work and expressed reciprocation (both in public and private matters). Voluntary work is positively associated with both membership in an organisation ( $p < 0.001$ , likelihood test of association) and expressed reciprocation in public matters but negatively associated with expressed reciprocation in private matters (in both cases  $p < 0.001$ , the the Mann–Whitney test). Similarly, membership in an organisation is positively associated with expressed reciprocation in public matters but negatively associated with expressed reciprocation in private matters (in both cases  $p < 0.001$ , the Mann–Whitney test). There is no significant association between the expressed levels of reciprocation in public and private matters.

The most extensive network of associations is centred around the sex of a student. Firstly, females express a higher level of generalised trust than males ( $p < 0.001$ , the Mann–Whitney test). Females are more likely to state that it is optimal to act in line with the law ( $p < 0.001$ , likelihood test of independence). In the choice of optimal strategy, cooperative behaviour is associated with behaviour in line with the law ( $p = 0.001$ , likelihood test of independence). However, there is no significant associa-

tion between sex and stating that it is optimal to be cooperative. In turn, the strategy assumed to be optimal is associated with a student's intention to follow such a strategy. Those stating behaviour on the edge of legality is optimal were less likely to follow such a strategy ( $p < 0.001$ , the Mann–Whitney test). A similar association was found for those who saw individualistic behaviour as being optimal ( $p = 0.001$ , the Mann–Whitney test). This may well reflect that cooperative behaviour in line with the law is favoured by legal and social norms and so those stating that such a strategy is optimal have no reason not to implement such a strategy. The association between sex and intention to follow the “optimal” strategy, which according to the model is indirect, is relatively weak. Females are more likely to intend following this strategy ( $p = 0.026$ , the Mann–Whitney test).

Sex is associated with attitudes towards tax evasion, in particular the attention paid to tax evasion, where males take more extreme positions (i.e. more often very concerned or not concerned at all,  $p < 0.001$ , likelihood test of association). Attention paid to tax evasion is positively associated with distaste for tax evasion ( $p < 0.001$ , Spearman's test of association). Also, females show a higher level of distaste for tax evasion ( $p = 0.032$ , the Mann–Whitney test).

Sex is related to interest in current affairs, particularly national affairs. Males are more interested in national affairs than females ( $p < 0.001$ , the Mann–Whitney test). However, females are more interested in local matters ( $p = 0.039$ , the Mann–Whitney test). There is no difference in the level of interest in regional matters according to sex. The levels of interest in national, regional and local matters are mutually positively associated ( $p < 0.001$ , Spearman's test of association). The strongest association is between interest in regional and interest in local matters.

Finally, sex is also associated with attitudes towards economic inequality and the emphasis placed on legal and ethical norms. Females express a higher level of support for government intervention to reduce economic inequality ( $p < 0.001$ ). Those who express support for the free market tend to have more extreme views in relation to the relation between legal and ethical rules (i.e. more often state that legal or ethical norms are categorically more important,  $p < 0.001$ , likelihood test of association).

Note that this network of associations assumes that variables which are not directly linked in Fig. 3 may be associated but are conditionally independent given the observations of the variables on the route between them, e.g. stress on ethical norms is associated with lower expressed reciprocation in personal matters and voluntary work ( $p = 0.003$ , Spearman's test of association,  $p = 0.016$ , the Mann–Whitney test).

A model was also constructed based on Akaike's information criterion. As this criterion is more liberal, this model contains more links and is difficult to present graphically. Hence, we only present the direct associations, which are not found to be direct or indirect associations according to Fig. 3. Table 10 summarises 19 associa-



tions found in this way and their level of significance. It should be noted that this model includes all 25 variables.

It is interesting to note that the levels of contact with family, friends and acquaintances are mutually positively correlated, although individuals must divide their time between various activities. The level of contact with friends is significantly correlated with both contact with family and contact with friends ( $p < 0.001$ , Spearman's test of correlation). Hence, individuals who score highly (lowly) on all three of these dimensions may be described as sociable (anti-social, respectively). Although these two associations are relatively strong (according to a test designed for ordinal variables), they are not detected by the more conservative model. This may well be due to the fact that this method treats the variables as being nominal. Hence, this approach will be relatively more powerful in finding associations between nominal variables rather than finding monotonic associations involving an ordinal (or discrete) variable.

It is unclear why an association is found between the level of expressed generalised trust and preference for a region in economic matters. However, it is possible that the complex network of associations between the complete set of 25 variables results in the lack of a visible association between these two variables when observed in isolation.

## **6. Analysis of the data using principal component analysis**

Unlike the Bayesian network approach, PCA assumes that the data come from a multivariate normal distribution. Obviously, this does not hold for our data but PCA can give us qualitative information regarding the mutual association between non-nominal variables. The variables considered in this analysis are the binary and ordinal/discrete variables described in Table 1, apart from attachment A (only asked to students who did not move), form of course (showed very little variance) and organisation membership (correlated by definition to the number of organisations a student is a member of). In addition, the three following 0–1 variables were used: move away, individual and legal. The second (third) variable is equal to one if and only if the student states that the optimal strategy involves being individualistic (acting in line with the law, respectively). Hence, 23 variables were used in the analysis.

We considered the components for which the corresponding eigenvalue was greater than one. Nine components were obtained that explained 57.8% of the variance in the data. An interpretation of the type of a student with a high score according to a given component is based on the four variables most strongly associated with that component (see Table 11).

Table 10. Associations according to the model based on the AIC criterion

Variable 1	Variable 2	Type of association	<i>p</i> -value	Test
Sex	city	more males in Zielona Góra, Olsztyn	<0.001	likelihood
Sex	membership	males are more often members	0.006	likelihood
Sex	acquaintances	males have more contact with acquaintances	<0.001	M–W
Sex	public	males react more often in public matters	0.004	M–W
Home	move away	more students move away from Mazowieckie and Śląskie	0.004	likelihood
Family	friends	positive association of contact with friends and contact with family	<0.001	Spearman
Friends	acquaintances	positive association of contact with friends and contact with acquaintances	<0.001	Spearman
City	size of home town	students at Rzeszów from rural areas, students at Warsaw, Łódź, Gdańsk from urban areas	<0.001	K–W
City	move away	more students move to Warsaw, Cracow, Opole	<0.001	likelihood
City	preference	Highest preference for region in Białystok, Rzeszów	<0.001	likelihood
Trust	preference	no significant association	0.196	M–W
Regional	voluntary	voluntary workers more interested in regional matters	<0.001	M–W
Voluntary	size of home town	volunteers are more likely to come from smaller towns	0.002	M–W
Strategy	voluntary	volunteers most likely to state that cooperation within the law is optimal	<0.001	likelihood
Trust	membership	members tend to be more trusting	<0.001	M–W
Public	attention	positive association	0.003	Spearman
City	year	youngest students in Katowice, Rzeszów, oldest in Szczecin, Białystok	<0.001	K–W
Reason A	strategy	attachment due to birth associated with stating that acting within the law is optimal	0.016	likelihood
Private	strategy	least reaction in those stating that cooperation within the law is optimal, greatest in those proposing individualism on the edge of the law	<0.001	likelihood

Likelihood: likelihood test of association, K–W – the Kruskal–Wallis test, M–W – the Mann–Whitney test, Spearman: Spearman’s test of association.

Source: Authors’ analysis using the *catnet* package in the R program.

Table 11. Description and interpretation of the principal components

Component	Percent of variance explained	Traits associated with a high score with respect to this component	Type
1	10.8	interested in regional and local matters, intends to follow “optimal” strategy which is in line with the law	grounded locally
2	7.9	male, favours the free market, accepting of tax evasion, states that acting on the edge of the law is optimal	male individualist
3	7.2	volunteer, member of social organisation, distaste for tax evasion	social activist
4	6.3	from large town and studies in home region, high level of contact with family and little interest in tax evasion	still lives in family home
5	6.0	high level of contact with friends and acquaintances, does not study in home region, intends to follow strategy assumed to be optimal	open and confident
6	5.4	comes from a small town, preference for home region, individualistic, interested in local issues	closed
7	5.0	advanced in their studies, low level of contact with friends and acquaintances, interested in national matters	career/studies orientated
8	4.7	comes from small town, no preference for region, stresses ethical norms, does not react to personal injury	small town “hippy”
9	4.4	low level of trust, young, low level of contact with acquaintances, individualistic	ungrounded socially

Source: Authors’ analysis using the IBM SPSS package.

## 7. Comments and conclusion

In the article, the theory of Bayesian networks and PCA have been used to describe the social relations and views of 1540 Polish students. When dealing with such data sets, the use of standard tests of association alone does not enable us to infer the underlying structure in the data.

Bayesian networks are very useful when investigating how a set of categorical variables interact. Out of the 35 associations found using this approach, 34 were clearly significant, based on traditional tests of association. Since this approach treats variables as being nominal, it is relatively weak at discovering monotonic relationships between ordinal variables, e.g. the associations between different forms of social contact are seen to be of lower importance using the Bayesian network approach rather than PCA. However, it gives a much better understanding of the underlying structure of the data than traditional tests of association.

PCA is designed to elucidate how a set of numerical variables are related and define latent variables which often have a physical interpretation. However, as the results from this study show, this approach can be successfully used on a set of ordinal and

binary variables. Based on the variables most strongly associated with each of the components corresponding to an eigenvalue of greater than one, we interpret latent variables describing various types of student. Of course, these categories are not distinct; a student may show traits corresponding to two or more types.

The results from the Bayesian network analysis give support to Putnam's [17] theory that membership in social organisations is associated with an increased level of generalised trust. The results also suggest that such membership is more strongly associated with both a higher level of condemnation of anti-social behaviour and a lower level of reaction to personal injury. However, it is unclear whether individuals with such an outlook are more likely to join a social group or whether membership in a social group actively promotes such an outlook. It is likely that feedback occurs here. The Bayesian network developed using BIC also clearly shows how the sex of an individual shapes their level of generalised trust and attitude to inequality.

Some of the latent types that result from PCA seem to coincide with Lewicka's [14] interpretation of types based on social interaction and feeling of attachment to a region. The analysis carried out here will be used in future work to describe the behaviour of students in the three experimental games played in the study.

### Acknowledgements

The authors thank the National Science Centre of Poland for funding on the basis of decision No. DEC-2012/07/B/HS5/03954.

### Appendix

#### **Spearman correlation coefficients between discrete and ordinal variables**

The variables considered are:

1. Year of study
2. Size of home town: scale 1–4, increasing in the size of a student's home town.
3. Interest in national affairs: scale 1–5, increasing in level of interest.
4. Interest in regional affairs: scale 1–5, increasing in level of interest.
5. Interest in local affairs: scale 1–5, increasing in level of interest.
6. Level of generalised trust: scale 1–5, increasing in level of trust.
7. Attention paid to tax evasion: scale 1–4, increasing in level of attention.
8. Distaste for tax evasion: scale 1–4, increasing in level of distaste.
9. Reaction to inequality: scale 1–3, increasing in support for government intervention.
10. Number of organisations a student is involved in.
11. Contact with family: scale 1–7, increasing in level of contact.
12. Contact with friends: scale 1–7, increasing in level of contact.

Table A1. Spearman's correlation coefficients between the ordinal and discrete variables considered. The first line in each cell gives the correlation coefficient, the second gives the  $p$ -value for the corresponding test of association. \*  $0.01 < p < 0.05$ , \*\*  $0.001 < p < 0.01$ , \*\*\*  $p < 0.001$

Variable	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	-0.034 0.184	-0.001 0.972	0.008 0.763	0.040 0.116	0.066 0.010*	-0.018 0.514	0.006 0.819	-0.009 0.721	-0.002 0.933	-0.076 0.003**	-0.085 0.001**	-0.073 0.004**	0.008 0.765	-0.034 0.187	0.045 0.077	-0.002 0.953
2	-	0.050 0.049*	-0.059 0.021*	-0.069 0.007**	-0.001 0.978	-0.043 0.113	-0.003 0.904	-0.146 0.00***	0.019 0.462	0.135 0.00***	-0.029 0.255	-0.064 0.013*	-0.004 0.884	0.012 0.634	-0.063 0.014*	0.047 0.065
3	-	-	0.294 0.00***	0.176 0.00***	-0.013 0.599	0.138 0.00***	0.063 0.019*	-0.112 0.00***	0.063 0.014*	0.070 0.006*	0.033 0.195	0.080 0.002**	0.149 0.00***	-0.003 0.902	0.076 0.003*	-0.021 0.414
4	-	-	-	0.546 0.00***	0.007 0.787	0.096 0.00***	0.053 0.046*	0.042 0.109	0.090 0.00***	0.091 0.00***	0.089 0.00***	0.084 0.001**	0.144 0.00***	-0.043 0.092	0.115 0.00***	-0.021 0.415
5	-	-	-	-	0.045 0.077	0.078 0.004**	0.029 0.287	0.045 0.085	0.124 0.00***	0.115 0.00***	0.077 0.002**	0.029 0.248	0.156 0.00***	-0.024 0.341	0.119 0.00***	0.034 0.179
6	-	-	-	-	-	-0.025 0.357	-0.050 0.063	-0.039 0.135	0.137 0.00***	0.018 0.483	-0.015 0.547	-0.015 0.564	0.053 0.039*	-0.095 0.00***	0.058 0.022*	0.032 0.212
7	-	-	-	-	-	-	0.500 0.00***	0.094 0.001**	0.030 0.270	0.073 0.007**	-0.034 0.210	-0.015 0.563	0.079 0.003**	0.000 1.000	0.051 0.057	-0.054 0.045*
8	-	-	-	-	-	-	-	0.127 0.00***	-0.024 0.366	0.039 0.140	-0.038 0.158	-0.041 0.129	0.002 0.929	0.031 0.250	0.035 0.197	-0.070 0.009**
9	-	-	-	-	-	-	-	-	-0.054 0.039*	0.070 0.008**	-0.011 0.665	0.009 0.730	-0.025 0.341	-0.012 0.652	-0.006 0.821	-0.027 0.295
10	-	-	-	-	-	-	-	-	-	-0.050 0.048*	-0.039 0.126	-0.006 0.825	0.144 0.00***	-0.075 0.004**	0.066 0.010	0.048 0.058
11	-	-	-	-	-	-	-	-	-	-	0.144 0.00***	0.028 0.272	0.055 0.032*	-0.033 0.202	0.063 0.013*	0.005 0.852
12	-	-	-	-	-	-	-	-	-	-	-	0.213 0.00***	0.092 0.00***	0.060 0.019*	0.051 0.046*	0.002 0.942
13	-	-	-	-	-	-	-	-	-	-	-	-	0.029 0.257	0.014 0.578	-0.028 0.276	0.011 0.668
14	-	-	-	-	-	-	-	-	-	-	-	-	-	0.019 0.446	0.187 0.00***	-0.010 0.698
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-0.109 0.00***	-0.075 0.003**
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-0.024 0.352

Source: Author's analysis using the SPSS package.

13. Contact with acquaintances: scale 1–7, increasing in level of contact.
14. Level of reciprocation in public matters: scale 1–4, increasing in level.
15. Level of reciprocation in private matters: scale 1–4, increasing in level.
16. Level of intention to follow strategy: scale 1–5, increasing in level.
17. Importance of ethical norms: scale 1–3, increasing in importance of ethical norms compared to legal norms.

## References

- [1] BALOV N., SALZMAN P., *Catnet: categorical Bayesian network inference. R package version, 2012, 1.4.*
- [2] BERG J., DICKHAUT J., MCCABE K., *Trust, reciprocity and social history*, Games and Economic Behavior, 1995, 10, 122–142.
- [3] BORGELT C., STEINBRECHER M., KRUSE R.R., *Graphical models: representations for learning reasoning and data mining*, Wiley, Weinheim, Germany, 2009.
- [4] BRYMAN A., CRAMER D., *Quantitative data analysis with IBM SPSS 17, 18 & 19: a guide for social scientists*, Routledge Press, Abingdon, UK, 2011.
- [5] CADSBY C.D., MAYNES E., *Gender and free riding in a threshold public goods game. Experimental evidence*, Journal of Economic Behavior and Organization, 1998, 34 (4), 603–620.
- [6] CUDNY W., *Socio-economic changes in Lodz – the results of twenty years of system transformation*, Geographical Journal, 2012, 64, 3–27.
- [7] GROWIEC K., *Social capital. Origins and social consequences*, SWPS Academic Publishing, Warsaw 2011 (in Polish).
- [8] GUILLEN P., JI D., *Trust discrimination and acculturation. Experimental evidence on Asian international and Australian domestic university students*, Journal of Socio-Economics, 2001, 40, 594–608.
- [9] GÜTH W., SCHMITTBERGER R., SCHWARZE B., *An experimental analysis of ultimatum bargaining*, Journal of Economic Behavior and Organization, 1982, 3, 367–388.
- [10] ISAAC R., WALKER J., *Group size effects in public goods provision. The voluntary contribution mechanism*, Quarterly Journal of Economics, 1988, 103, 179–200.
- [11] KAISER F.G., BYRKA K., HARTIG T., *Reviving Campbell's paradigm for attitude research*, Personality and Social Psychology Review, 2010, 14 (4), 351–367.
- [12] KOMORNICKI T., *Transformation of the migrational processes in Poland (1988–1994)*, Discussion Papers, 1999, 49–58.
- [13] KOPCZEWSKI T., *Does a degree in economics reduce social capital? Experimental approach of measurement standards and social preferences*, Ekonomia, 2003, 11, 173 (in Polish).
- [14] LEWICKA M., *Localism and activity as two dimensions of people-place bonding. The role of cultural capital*, Journal of Environmental Psychology, 2013, 36, 43–53.
- [15] MARKOWSKA-PRZYBYŁA U., RAMSEY D., *A game theoretical study of generalised trust and reciprocation in Poland. I. Theory and experimental design*, Operations Research and Decisions, 2014 (3), 59–76.
- [16] NORRIS P., *Does television erode social capital? A reply to Putnam*, Political Science and Politics, 1996, 29 (3), 474–480.
- [17] PUTNAM R., *Making democracy work*, Princeton University Press, Princeton, NJ, 1993.
- [18] SALZMAN P., ALMUDEVAR A., *Using complexity for the estimation of Bayesian networks*, Statistical Applications in Genetics and Molecular Biology, 2006, 5.1.

- [19] SAX J.R., GILMARTIN S.K., LEE J.J., HAGEDORN L.S., *Using web surveys to reach community college students: an analysis of response rate and response bias*, Community College Journal of Research and Practice, 2008, 32, 712–729.
- [20] [www.oecd.org/gender/Closing%20the%20Gender%20Gap%20-%20Poland%20Final.pdf](http://www.oecd.org/gender/Closing%20the%20Gender%20Gap%20-%20Poland%20Final.pdf) accessed 20.11.2014).
- [21] [http://www.perspektywy.pl/portal/index.php?option=com\\_content&view=article&id=1593:rsw2014&catid=139&Itemid=287&idRanking=55](http://www.perspektywy.pl/portal/index.php?option=com_content&view=article&id=1593:rsw2014&catid=139&Itemid=287&idRanking=55) (accessed 24/11/2014).
- [22] <http://stat.gov.pl/statystyka-regionalna/rankingi-statystyczne/miasta-o-najwiekszej-liczbie-ludnosci-3018/>(accessed 19/11/2014).

*Received 28 November 2014*

*Accepted 15 April 2015*