BIPARTITE COMPETENCY SCHEMAS ON POLISH LABOUR MARKET

Lula, P., Kovaleva, A., Oczkowska, R., Tyrańska, M., Wiśniewska, S.

Paweł Lula, Renata Oczkowska, Małgorzata Tyrańska, Sylwia Wiśniewska / Cracow University of Economics, Poland. Email: pawel.lula@uek.krakow.pl, renata.oczkowska@uek.krakow.pl, malgorzata.tyranska@uek.krakow.pl, sylwia.wisniewska@uek.krakow.pl

Anna Kovaleva / Saint-Petersburg State University of Economics, Russia. Email: intermasterf@gmail.com

Abstract

The complexity and variability of the contemporary labour markets creates the need for continuous improvement of methods used for their description, analysis and forecasting. Looking for a tool that allows for the simultaneous analysis of various aspects of contemporary labour markets, the authors focused their attention on k-partite graph models (with particular emphasis on bipartite graphs). The assessment of the usefulness of models based on bipartite graphs for analysis of regularities occurring on the Polish labour market is the main aim of the paper. The authors studied the regional distribution of the demand for employee competencies and evaluated the specificity of localities and competencies. The concept of bipartite competency schemas is also introduced in the paper. These schemas can be used as models representing strongly related competencies and localities. The usefulness of bipartite competency schemas was confirmed by empirical research presented in the paper. The content of job offers published online formed the main source of data examined. All analyses were performed with the use of the R programming language.

Introduction

The currently occurring technological and social changes have created new challenges for the modern labour market. There is a demand for new competencies and for new arrangements of previously existing competencies. A high rate of changes creates the need for continuous monitoring of the labour market situation. To realize this requirement, the authors, in their earlier works, proposed a concept of competency schemas, which can be defined as models of complex arrangements of competences required or offered on the labour market. In this approach, a competency schema is represented by a weighted graph in which competencies are represented by nodes and connections between them by edges. Weights assigned to nodes and edges inform about the importance of each element.

Despite many advantages, the proposed concept also has its disadvantages. The main one is related to the lack of possibility of taking into account factors other than competencies. Ensuring the ability to conduct competence analysis including other factors relevant to the modern labour market was the main goal of the paper. The authors based the proposed method on bipartite graph models known in ecology. The suggested approach allows the analysis of competences taking into account such factors as information about the sector, region, job position or enterprise. The researcher can estimate the importance and specificity of every factor and links between them. These data can be useful for evaluating the current situation on the labour market and can be fruitful for decision-making processes.

The paper has the following structure. The theoretical background of the proposed method of analysis can be found in the first section. In the second section, the survey concerning the application of graph models in labour market analysis is presented. Next, in the third section, the bipartite models for the Polish labour market are presented.

1 Graph models in labour market analysis

Graph models related to labour markets and competencies are considered by various authors from different perspectives. Researchers are faced with questions about the models and the variables responsible for explaining the relations between them.

In general, a mathematical approach is applied to the analysis of the labour market from the structural point of view capturing the main principles of supply and demand of households and firms. Lloret-Climent, Nescolarde-Selva, Mora-Mora & Signes-Pont (2018) propose the input-output analysis of the labour market within the framework of network theory and provide a mathematical model of labour relations based on network theory and graphs. This model explains the functioning of the market and the relations between different sectors of the economy considering the flow of goods and services in an example of a Spanish region. The new concepts have been introduced to the graph and network theories in relevance to the systematic view of labour market as orbits of companies, coverage, invariant set and circular flow.

Along this, in the subject literature, the graph models are used to analyse various areas of the labour market. Guerrero & Axtell (2013) emphasize that it is conventional in labour economics to treat all workers who are seeking new jobs as belonging to a labour pool, and all firms that have job vacancies as an employer pool, and then match workers to jobs. They develop a new approach to study labour and firm dynamics. By combining the emerging science of networks with newly available, comprehensive employment micro data in the examples of Finland and Mexico, they are able to broadly characterize the process through which workers move between firms. Five types of small and medium enterprises are considered according to their growth such as positive growth firm, labour flow network firm, high-growth firm, gazelle firm and high-impact firm. Specifically, for each firm in an economy, represented as a node in a graph, they draw edges between firms if a worker has migrated between them, possibly with a spell of unemployment in between. An economy's overall graph of firm-worker interactions is an object they call the labour flow network (LFN). The construction of a LFN is as follows: for a selected period, they count the total flows of labour between two firms in both directions. Although this is a directed network, they found that the most interesting insights come from studying its structural properties as an

undirected graph. Therefore, their analysis uses algorithms for undirected networks (with exception of in-degree and out-degree centralities). Additionally, the agent-based model has been used in this process of labour flow network connection to analyse job-to-job dynamics.

The graph method is also used to explore different types of networks at the intercompany level. Antonelli, Bruno, Taurino & Villa (2015) consider various types of networks of small and medium enterprises and their key parameters to reveal network modifications. These ideas have been continued in the next works. Thus, another example of studying labour flows across firms has been conducted on Finnish data examining the role of the aggregate matching function in considering the economy as a set of submarkets. However, this approach has revealed the restrictions of capturing the general picture of the labour market and not taking into account the impact of individual firms in reallocating of labour (Guerrero & Lopez, 2015).

In turn, Razakanirina & Chopard (2012) use the Multilayer Cellular Automata on Graph (MCAG) to investigate the dynamics of some abstract markets based on three main wealth ingredients: goods, working hours and money. In doing so, they consider particularly the cases of the goods, labour and money markets. It was indicated (Razakanirina & Chopard, 2012) that economic systems are complex ones built upon the complicated interactions between heterogeneous agents. These systems are defined using a graph due to the irregularities of the interactions. MCAG is an effective framework to simulate such complex systems. The artificial markets simulated in the article are defined using MCAG formalism. On the one hand, it was showed that all the agents belonging to circuits of goods and labour markets remain available during the whole simulation. These particular topologies are stable. On the other hand, the money market is sensitive to the existence of circuits due to the fact that systems with circuits crash and become chaotic with short-term loan duration.

Then, direct and indirect influence between particular elements creating the structure of the competence model are analysed within the framework of the graph model, which allows for the evaluation of the integrity of the competence model, relations between competences that create such model (Kvyatkovskaya, Sibikina & Berezhnov, 2013). Descriptive statistics as the analytical framework is developed for the analysis of labour market taking into account the particular parameters. Moreover, the applied researches use bipartite graph to reveal the dependencies in professional paths.

One of the examples is provided by Uliana &de Castro (2018) with a model of a career progression as a graph, called Map of Careers (MCar), in which nodes represent occupations and edges correspond to the flow of professionals among occupations. Based on MCar, some key concepts to investigate career evolution, namely, career boundary, occupational poles, and occupational islands are introduced. Finally, the authors propose the use of community detection techniques on a real database with millions of professional backgrounds to objectively identify career boundaries in Brazil and to study the topologies of the resultant graph. The obtained results provide a quantitative basis for career models, showing, for instance, the presence of hubs suggesting that less regular careers are common.

Tassier & Menczer (2001) consider a labour market with an agent-based simulation according to which jobseekers combine direct job search and social networks (mainly friends, relatives and other social contacts) to increase the effectiveness of looking for a job. The proposed model is elaborated on at the macro level where jobs are considered as supply and open jobs as demand. The model has two main results: (1) networks with different local structures are like small-world networks but don't guarantee the efficient transfer of information between agents and (2) the competition between jobseekers is increasing so people must compare their individual skills and goals at a global level. A simulation algorithm is applied to the model to measure the initial level of search intensity with an initial number of friends. Afterwards, the diameter or path length as the average of the shortest path between all pairs of vertices in the graph is applied which measures the interconnection between the strength of connection between people, education and status of agents (the level of their job for example). As a result, the model reveals that unemployed people prefer to keep their contacts with friends who have jobs in order to increase their chances of obtaining a job.

Additionally, the model has been used to analyse mobility pathways to jobs with higher levels of wages, as in cases using data from a national grocery store chain (McDonald & Benton, 2017). Additional uses of the model have been evident in the studies concentrated on the matching of education background and job occupation, as in the study utilizing the databases of the National Tax Administration, the National Health Insurance Fund of Hungary and the data set of the Hungarian higher education. The combined database contains 70 variables about such parameters as personal, occupational, employer and educational statistics, and ownership in business. The bipartite network-type model has also been used to analyse the connection between the weights of the edges with the expected number of edges of a random graph with the same degree as the considered network. The graph has been clustered into sub graphs and then the weak connections have been removed in series based on a proposed multi-resolution type analysis of network connections measured by their strengths.

One more relevant research is depicted in a graph-based data structure to analyse the career path of Universities graduates. Furthermore, the gender pay gap and the spatial distribution of over-education have been studied (Gadar & Abonyi, 2018). In this research the authors introduce the bipartite competency schemas to overcome the limitations of applying bipartite models for competency analysis in explaining the connection between the labour market and competences of agents (for instance, students, job seekers etc.).

2 Research methodology

2.1 Competency schema concept

The competency schema concept was introduced in Lula, Oczkowska, Kovaleva & Wiśniewska (2019). It can be defined as a set of competencies and a set of relations between them, together with the information about the importance of every competency and the significance of every relationship between any two competencies. A competency schema can be used for describing the demand or supply side of the labour market and subsequently allows for the identification of competency gaps.

The construction of the competency co-occurrence graph is the first stage in the process of competency schemas identification. It has the form of a weighted graph with weights assigned to nodes (representing competencies' importance) and to edges (representing connections' importance). Competency co-occurrence graphs form a base for competency schemas' identification. This process is realized by the detection of strongly related sub graphs (called communities) in the competency co-occurrence graph.

It is worth highlighting that due to the relative difficulty of measuring the strength of relations between competences in individual cases, competency co-occurrence graphs and competency schemas are not appropriate for describing the requirements related to an individual job position or abilities of an individual employees or job candidates. In short, competency co-occurrence graphs and competency schemas should be used for representing only generalized information about the labour market.

The application of the competency schema concept allows one to:

- identify and evaluate competencies possessed by employees, job candidates or graduates,
- analyse competences expected by the labour market,
- describe and examine connections between competencies,
- compare the demand and supply side of labour market and evaluate the competency gap also in cross-country perspective.

Despite the advantages indicated above, the competency schema concept also has weaknesses. The most important is its lack of taking into account factors different than competencies already existing on the labour market. An attempt of removing this disadvantage was considered as the main methodological goal of this study. The work carried out was focused on the possibility of applying bipartite graph models.

2.2 Bipartite model for competency analysis

As stated above, competency schemas allow for the presentation of competencies and relations between them, but they do not take regard to the context of the analysis of embracive information related to the region, sector, position, company and other attributes crucial for the labour market. It may indicate that k-partite graphs instead of undirected graphs can be used for building models describing competencies in the context of other attributes characterizing a given labour market. A k-partite graph is a graph whose nodes belong to k disjoint sets and the edges always connect nodes belonging to two different sets. Edges between nodes from the same set are not allowed.

The considerations presented later in the paper concern competency schemas based on bipartite graphs. The authors propose to call them *bipartite competency schemas*.

Let's assume that the C is a set of competencies:

$$C = \{c_1, c_2, \dots, c_M\} \tag{1}$$

and that the context is described by a set V:

$$V = \{v_1, v_2, \ldots, v_N\} \tag{2}$$

Values belonging to the set V represent various possible states in which competencies should be analyzed.

The relations between elements of C and V sets are defined by an interaction matrix G:

$$G = \begin{bmatrix} g_{11} & \cdots & g_{1M} \\ \cdots & \cdots & \cdots \\ g_{N1} & \cdots & g_{NM} \end{bmatrix} \tag{3}$$

The G matrix columns represent competencies and rows – elements of the V set. Element g_{ij} gives the number of interactions between v_i and c_j values.

Models of this type are widely used in ecology for describing relations between two groups of species which are called "higher group of species" and "lower group of species". Species belonging to the "higher" group correspond always to columns, and species belonging to "lower" group always correspond to rows of the G matrix. Using this convention, competencies are equivalent to species belonging to the "higher" group, and values of the V set correspond to species from the "lower" group (Pavlopoulos et al., 2018).

There are many factors confirming the correspondence between interactions of species analysed in ecological models and relationships between objects existing on the labour market. In these two cases the important role is played by the analysis of:
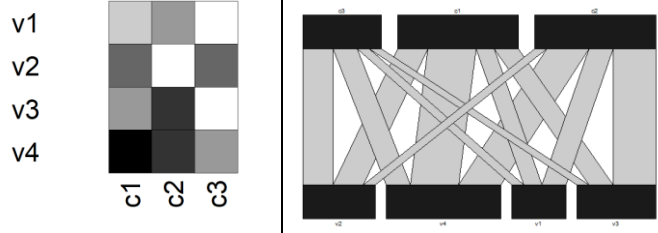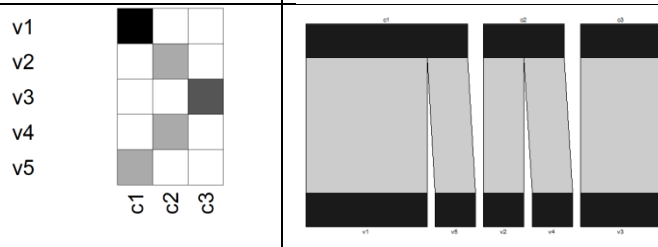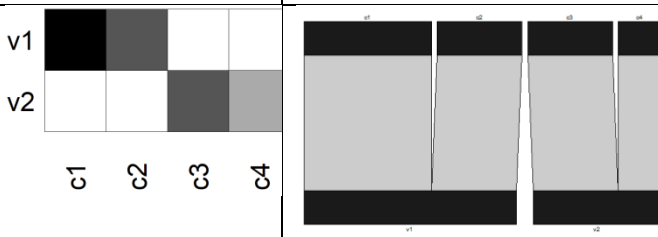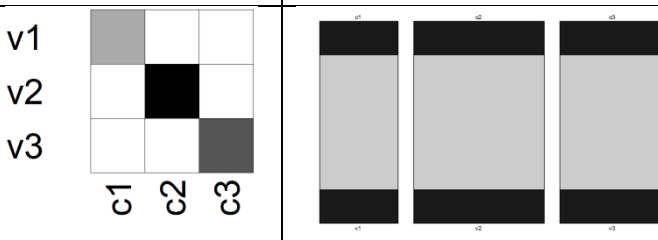
- connections between two types of objects – in this case of labour market analysis, the character and strength of relationships between competencies and other factors (job offers, position, companies or sectors),
- objects' importance – labour market analysis should include an assessment of importance of competencies in general case and for individual positions, companies or sectors,

These two groups of analysis allow to identify groups of interrelated competencies, groups of other factors of labour market and all connections existing in this complex system.

2.3 Exemplary networks

The analysis of bipartite networks can be performed with the use of different methods. To present their potentiality, they will be used for analysis of the four various bipartite networks presented in Table 1.

Table 1 | Exemplary bipartite networks

Id	Network presentation
Network 1	$\begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \begin{array}{ccc} c_1 & c_2 & c_3 \\ \left[\begin{array}{ccc} 3 & 4 & 2 \\ 5 & 2 & 5 \\ 4 & 7 & 2 \\ 8 & 7 & 4 \end{array}\right] \end{array}$
Network 2	$\begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{array} \begin{array}{ccc} c_1 & c_2 & c_3 \\ \left[\begin{array}{ccc} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{array}\right] \end{array}$
Network 3	$\begin{array}{c} \\ v_1 \\ v_2 \end{array} \begin{array}{cccc} c_1 & c_2 & c_3 & c_4 \\ \left[\begin{array}{cccc} 3 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \end{array}\right] \end{array}$
Network 4	$\begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \end{array} \begin{array}{ccc} c_1 & c_2 & c_3 \\ \left[\begin{array}{ccc} 3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 4 \end{array}\right] \end{array}$

Source: own elaboration

2.4 Descriptive statistics for network models

Descriptive statistics show the main features of the networks (Dormann, Fründ, Blüthgen & Gruber, 2009; Dormann, Gruber & Fründ, 2008). These measures can be calculated of three levels:

- for individual nodes,
- for groups of nodes,
- for a network

and are discussed in sections 2.4.1 – 2.4.3.

2.4.1 Statistics for individual nodes

2.4.1.1 Degree

For every node, the degree value informs about the number of partners from the other group. For nodes belonging to the **V** set the degree is defined as:

$$N_D(v_i) = \sum_{j=1}^{M} \text{sgn}(g_{ij})$$ (4)

where $\text{sgn}(.)$ is the signum function.

For nodes belonging to the **C** set, the definition takes the form:

$$N_D(c_j) = \sum_{i=1}^{N} \text{sgn}(g_{ij})$$ (5)

The degree values for exemplary networks are presented in Table 2.

Table 2 | Degree value for individual nodes

	Network 1	Network 2	Network 3	Network 4
$N_D(\mathbf{C})$	$N_D(c_1) = 4$ $N_D(c_2) = 4$ $N_D(c_3) = 4$	$N_D(c_1) = 2$ $N_D(c_2) = 2$ $N_D(c_3) = 1$	$N_D(c_1) = 1$ $N_D(c_2) = 1$ $N_D(c_3) = 1$ $N_D(c_4) = 1$	$N_D(c_1) = 1$ $N_D(c_2) = 1$ $N_D(c_3) = 1$
$N_D(\mathbf{V})$	$N_D(v_1) = 3$ $N_D(v_2) = 3$ $N_D(v_3) = 3$ $N_D(v_4) = 3$	$N_D(v_1) = 1$ $N_D(v_2) = 1$ $N_D(v_3) = 1$ $N_D(v_4) = 1$ $N_D(v_5) = 1$	$N_D(v_1) = 2$ $N_D(v_2) = 2$	$N_D(v_1) = 1$ $N_D(v_2) = 1$ $N_D(v_3) = 1$

Source: own elaboration

2.4.1.2 Normalized degree

Normalized degree is calculated by dividing degree value by the number of possible partners for a node. For nodes existing in the model the appropriate formulas take the form:

$$N_{ND}(v_i) = \frac{N_D(v_i)}{M}$$ (6)

and:

$$N_{ND}(c_j) = \frac{N_D(c_j)}{N}$$ (7)

Table 3 | Normalized degree value for individual nodes

	Network 1	Network 2	Network 3	Network 4
$N_{ND}(C)$	$N_{ND}(c_1) = 1$	$N_{ND}(c_1) = 0.4$	$N_{ND}(c_1) = 0.5$	$N_{ND}(c_1) = 0.333$
	$N_{ND}(c_2) = 1$	$N_{ND}(c_2) = 0.4$	$N_{ND}(c_2) = 0.5$	$N_{ND}(c_2) = 0.333$
	$N_{ND}(c_3) = 1$	$N_{ND}(c_3) = 0.2$	$N_{ND}(c_3) = 0.5$	$N_{ND}(c_3) = 0.333$
			$N_{ND}(c_4) = 0.5$	
$N_{ND}(V)$	$N_{ND}(v_1) = 1$	$N_{ND}(v_1) = 0.333$	$N_{ND}(v_1) = 0.5$	$N_{ND}(v_1) = 0.333$
	$N_{ND}(v_2) = 1$	$N_{ND}(v_2) = 0.333$	$N_{ND}(v_2) = 0.5$	$N_{ND}(v_2) = 0.333$
	$N_{ND}(v_3) = 1$	$N_{ND}(v_3) = 0.333$		$N_{ND}(v_3) = 0.333$
	$N_{ND}(v_4) = 1$	$N_{ND}(v_4) = 0.333$		
		$N_{ND}(v_5) = 0.333$		

Source: own elaboration

2.4.1.3 Node specificity

Node specificity reflects the diversity of a node's interactions with partner nodes. Low diversity is equivalent to low specificity, and high diversity indicates high specificity.

The specificity of the j-th node can be expressed by the variance of interactions with all nodes from the partner group. Poisot, Canard, Mouquet & Hochberg (2012) measured node specificity, a standardized coefficient of variation:

$$N_{NS}(i) = \frac{\sqrt{\sum_{j=1}^{P}(g_{ij}-\mu_i)^2}}{\mu_i P \sqrt{\frac{P-1}{P}}} = \frac{\sqrt{\sum_{j=1}^{P}(g_{ij}-\mu_i)^2}}{\mu_i \sqrt{P}\sqrt{P-1}} \tag{8}$$

where P is the number of nodes in the partner group.

To calculate the $N_{NS}(i)$ value, the coefficient of variation is defined as:

$$CoV = \frac{\sqrt{\frac{\sum_{j=1}^{P}(g_{ij}-\mu_i)^2}{P}}}{\mu_i} \tag{9}$$

which should be divided by its maximum value $\sqrt{P-1}$.

The coefficient $N_{NS}(i)$ belongs to the $[0;1]$ range. Value 0 means low specificity, and value 1 indicates for the maximal specificity.

The values of specificity indexes for nodes belonging to four exemplary networks are presented in the Table 4.

Table 4 | Values of specificity index for nodes

	Network 1	Network 2	Network 3	Network 4
$N_{NS}(\boldsymbol{C})$	$N_{NS}(c_1) = 0.216$	$N_{NS}(c_1) = 0.729$	$N_{NS}(c_1) = 1$	$N_{NS}(c_1) = 1$
	$N_{NS}(c_2) = 0.245$	$N_{NS}(c_2) = 0.612$	$N_{NS}(c_2) = 1$	$N_{NS}(c_2) = 1$
	$N_{NS}(c_3) = 0.231$	$N_{NS}(c_3) = 1.000$	$N_{NS}(c_3) = 1$	$N_{NS}(c_3) = 1$
			$N_{NS}(c_4) = 1$	
$N_{NS}(\boldsymbol{V})$	$N_{NS}(v_1) = 0.192$	$N_{NS}(v_1) = 1$	$N_{NS}(v_1) = 0.600$	$N_{NS}(v_1) = 1$
	$N_{NS}(v_2) = 0.250$	$N_{NS}(v_2) = 1$	$N_{NS}(v_2) = 0.638$	$N_{NS}(v_2) = 1$
	$N_{NS}(v_3) = 0.335$	$N_{NS}(v_3) = 1$		$N_{NS}(v_3) = 1$
	$N_{NS}(v_4) = 0.190$	$N_{NS}(v_4) = 1$		
		$N_{NS}(v_5) = 1$		

Source: own elaboration

2.4.2   Group level

Group level statistics are calculated separately for competences (represented by the set $\boldsymbol{C}$) and for a set describing the analysis context (represented by the set $\boldsymbol{V}$).

2.4.2.1   Number of nodes

The number of nodes in a given group forms a basic measure for each of two groups of nodes occurring in the bipartite graph. For exemplary networks, the value of these statistics is presented in the Table 5.

Table 5 | Number of nodes belonging to each of two groups

	Network 1	Network 2	Network 3	Network 4
$G_{NN}(\boldsymbol{C})$	3	3	4	3
$G_{NN}(\boldsymbol{V})$	4	5	2	3

Source: own elaboration

2.4.2.2   Mean number of partners

Nodes v_i and c_j can be considered as partners if an interaction between them exists, i.e. an element $g_{ij} > 0$.

The mean number of partners is calculated as a sum of partners of every node belonging to a given group divided by the number of nodes in this group. For sets C and V these measures are calculated as:

$$G_{MNP}(\mathbf{C}) = \frac{I}{M} \tag{10}$$

and:

$$G_{MNP}(\mathbf{V}) = \frac{I}{N} \tag{11}$$

where I is the number of non-zero elements in the matrix $\boldsymbol{G}$.

The values of these statistics calculated for exemplary networks are presented in the Table 6.

Table 6 | Mean number of partners

	Network 1	Network 2	Network 3	Network 4
$G_{MNP}(\mathbf{C})$	4.000	1.667	1.000	1.000
$G_{MNP}(\mathbf{V})$	3.000	1.000	2.000	1.000

Source: own elaboration

2.4.2.3 *Mean number of shared partners*

The number of shared partners can be calculated for two nodes belonging to the same group. It expresses the number of partners in the other group that both nodes interact with. The statistics presented here reflect the averaged value of shared partners calculated for a whole group (separately for C and for V set).

Table 7 | Mean number of shared partners

	Network 1	Network 2	Network 3	Network 4
$G_{MNSP}(\mathbf{C})$	4.000	0.000	0.333	0.000
$G_{MNSP}(\mathbf{V})$	3.000	0.200	0.000	0.000

Source: own elaboration

2.4.2.4 *Cluster coefficient*

The idea of the cluster coefficient was introduced in Watts & Strogatz (1998). For a given node u, a cluster coefficient calculates the probability that. selected at random, neighbours of u are connected by an edge (it means that they are neighbors to each other). In other words, it expresses the tendency to form a clique (a graph in which all nodes are connected directly by an edge) by neighbours of a given node u.

The cluster coefficient calculation is relatively simple for one-mode networks. Assume that two nodes are called neighbours if an edge between them exists. Let $N(u)$ be a set of neighbours of a node u. Then a possible number of edges between neighbors is calculated:

$$t_{N(u)} = \frac{|N(u)|(|N(u)|-1)}{2} \tag{12}$$

where $|N(u)|$ is the number of neighbours of a node u.

Let $o_{N(u)}$ be the observed numbers of edges between neighbors of the node u. The cluster coefficient for the u node can be defined as:

$$n_{CC}(u) = \frac{o_{N(u)}}{t_{N(u)}} \tag{13}$$

where $t_{N(u)} = \frac{|N(u)|(|N(u)|-1)}{2}$ is the number of links that could possibly exist between neighbours of the node u.

If the $n_{CC}(u)$ is equal to 1, then all neighbors of u are connected by a single edge (they create a clique).

The idea of a cluster coefficient for bipartite networks is similar, but second order neighbours (neighbours of neighbours) are considered. The calculation process for this

case is composed of several steps (Latapy, Magnien & Del Vecchio, 2011). First, a clustering coefficient for two nodes (we assume that two of them belong to the set C or V) is calculated:

$$cc(u, v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|} \tag{14}$$

After it, the clustering coefficient can be expressed for a given node u (as an averaged value of $cc(u, v)$):

$$cc(u) = \frac{\sum_{v \in N(N(u))} cc(u,v)}{|N(N(u))|} \tag{15}$$

The above value can be aggregated for:

- competencies (elements of the C set):

$$G_{CC}(C) = \frac{\sum_{u \in C} cc(u)}{M} \tag{16}$$

- elements of the V set:

$$G_{CC}(V) = \frac{\sum_{v \in V} cc(v)}{N} \tag{17}$$

- the whole network:

$$N_{CC} = \frac{\sum_{u \in C} cc(u) + \sum_{v \in V} cc(v)}{M + N} \tag{18}$$

The values of cluster coefficients for each of two groups existing in four exemplary networks are presented in the Table 8.

Table 8 | Cluster coefficients for groups

	Network 1	Network 2	Network 3	Network 4
$G_{CC}(C)$	1.000	0.350	0.500	0.333
$G_{CC}(V)$	1.000	0.333	0.500	0.333

Source: own elaboration

2.4.2.5 *Niche overlap*

In the study by Lomolino, Riddle & Brown (2009), an ecological niche is defined as the total requirements of a population or species for resources and physical conditions. The niche of a given species shows its interactions with the environment. In the context of bipartite graph models, a niche can be defined as a set of partners from the other group which are connected with a given node.

In Horn (1966) a niche overlap index was proposed:

$$R_{jk} = \frac{\sum_{i=1}^{P}(p_{ij}+p_{ik}) \ln(p_{ij}+p_{ik}) - \sum_{i=1}^{P} p_{ij} \ln p_{ij} - \sum_{i=1}^{P} p_{ik} \ln p_{ik}}{2 \ln 2} \tag{19}$$

where:

- R_{jk} – a niche overlap index defined for nodes j and k,
- p_{ij} – calculated for the j-th node, the relation calculated as: $p_{ij} = \frac{g_{ij}}{\sum_{i=1}^{P} g_{ij}}$, where P is a number of nodes in the partner group
- p_{ik} – calculated for the k-th node, the relation calculated as: $p_{ik} = \frac{g_{ik}}{\sum_{i=1}^{P} g_{ik}}$, where P is a number of nodes in the partner group

The niche overlap index belongs to the range $[0; 1]$ and 0 means that niches do not overlap, whereas 1 informs about the maximum overlapping.

The niche overlap index for a group is equal to an averaged value of indexes calculated for all pairs of nodes belonging to this group.

The values of niche overlap coefficients for exemplary networks are presented in the Table 9.

Table 9 | Niche overlap coefficients

	Network 1	Network 2	Network 3	Network 4
$G_{NO}(C)$	0.884	0.000	0.333	0.000
$G_{NO}(V)$	0.891	0.200	0.000	0.000

Source: own elaboration

2.4.3   Network level (web level)

The characteristics of network level statistics are presented in Dormann et al. (2009).

2.4.3.1   Connectance

The connectance index is defined as:

$$W_C = \frac{I}{M*N} \tag{20}$$

where I is the number of non-zero elements in the matrix G. The N_c index takes into account only the existence of relations between elements of C and V sets and ignores its strength.

Table 10 | Connectance index

	Network 1	Network 2	Network 3	Network 4
W_c	1.000	0.333	0.500	0.333

Source: own elaboration

2.4.3.2   Web asymmetry

The web asymmetry index informs about the balance in the bipartite model size and shows the difference between the number of elements in C and V set with respect to the sum of these two values. It is defined as:

$$W_{WA} = \frac{|C|-|V|}{|C|+|V|} \tag{21}$$

where $|C|$ and $|V|$ means the number of elements in sets C and V respectively.

Table 11 | Web asymmetry index

	Network 1	Network 2	Network 3	Network 4
W_{WA}	-0.143	-0.250	0.333	0.000

Source: own elaboration

2.4.3.3 Mean number of links per nodes

The mean number of links in the network is calculated as:

$$W_{MNP} = \frac{I}{|C|+|V|} \tag{22}$$

where I is the number of non-zero elements in the matrix G.

Table 12 | Mean number of links per node

	Network 1	Network 2	Network 3	Network 4
W_{MNP}	1.714	0.625	0.667	0.500

Source: own elaboration

2.4.3.4 Number of compartments

Compartments can be defined as connected components of a given network. The index W_{NC} expresses the number of compartments for a network.

Table 13 | Number of compartments

	Network 1	Network 2	Network 3	Network 4
W_{NC}	1	3	2	3

Source: own elaboration

2.4.3.5 Cluster coefficient

The cluster coefficient for a given network is based on calculations presented in section 2.4.2.4 and then averaged with the use of the formula:

$$W_{CC} = \frac{\sum_{u \in C} cc(u) + \sum_{v \in V} cc(v)}{M+N} \tag{23}$$

Table 14 | Cluster coefficient for exemplary networks

	Network 1	Network 2	Network 3	Network 4
W_{CC}	1.000	0.400	0.500	0.333

Source: own elaboration

2.4.3.6 Specialization index H_2'

The specialization index H_2' is in a study by Blüthgen, Menzel & Blüthgen (2006).

First, probabilities associated with the G matrix should be defined. To perform this task, the sum of all elements ought to be calculated:

$$s = \sum_{i=1}^{|V|} \sum_{j=1}^{|C|} g_{ij} \tag{24}$$

Next, probabilities p_{ij} may be expressed as:

$$p_{ij} = \frac{g_{ij}}{s} \tag{25}$$

also marginal probabilities can be calculated:

$$p_{i*} = \frac{\sum_{j=1}^{M} g_{ij}}{s} \tag{26}$$

and:

$$p_{j*} = \frac{\sum_{i=1}^{N} g_{ij}}{s} \tag{27}$$

Having a probability matrix, the two-dimensional Shannon entropy can be expressed:

$$H_2 = -\sum_{i=1}^{N} \sum_{j=1}^{M} \left(p_{ij} \ln p_{ij} \right) \tag{28}$$

Lower values of H_2 indicates higher specialization and higher values of H_2 informs about higher generalization. Unfortunately, values of H_2 are not limited to the $[0; 1]$ range and therefore they should be normalized. For this operation, for a given structure (number of rows and number of columns) of the $\boldsymbol{G}$ matrix, a maximum (H_2^{max}) and minimum (H_2^{min}) value of the H_2 index should be calculated. Finally, an index of specialization can be formulated:

$$H_2' = \frac{H_2^{max} - H_2}{H_2^{max} - H_2^{min}} \tag{29}$$

The H_2' always belongs to the $[0; 1]$ range and for high generalisation (nodes interact with many partners from the other group) it is close to 0, and for high specialization (nodes interact with small number of partners from the other group) it is close to 1.

Table 15 | The H_2' coefficient for exemplary networks

	Network 1	Network 2	Network 3	Network 4
$W_{H_2'}$	0.041	1.000	1.000	1.000

Source: own elaboration

2.5 Competency schema identification

A bipartite competency schema can be defined as a strongly related sub graph of a bipartite graph describing competencies together with another factor existing on the labour market. Bipartite competency schemas can be formed by compartments (sub graphs completely separated) or communities (sub graphs which are connected with other parts of a graph, but their ties with other sub graphs are notably weaker than ties existing within them). As long as identification of compartments is relatively simple, then the communities' analysis can be a bigger challenge.

2.5.1 Compartment detection

The problem of compartment detection can be solved by the use of the depth-first search (Newman & Girvan, 2004). First, a starting point should be chosen. Next, all element reachable from the starting point using the DFS algorithm form the first compartment (connected sub graph). Performing the same procedure for unvisited nodes allows for the possibility to identify all compartments existing in the graph.

The most popular approach of community detection is based on the maximization of the modularity measure (Fortunato, 2010). Modularity can be treated as a measure of quality calculated for a given graph division and it reflects the number of edges within clusters compared with number of edges existing between different clusters. Having a graph divided into clusters, the intuitive measure of modularity can be defined:

$$m = f_{within} - f_{between} \tag{30}$$

where:

- f_{within} is a fraction of edges which join nodes belonging to the same cluster,
- $f_{between}$ is a fraction of edges which join nodes belonging to two different clusters.

The m measure belongs to the range $[-1, +1]$ and its positive value indicates that the density of inner connections is greater than the density of outer connections.

The most common measure of modularity was proposed by Newman & Girvan (2004). It compares the structure of connections in a given graph (which is divided into clusters) with connections existing in a *null model* which has the same number of nodes and random structure of edges (it guarantees lack of clusters in the graph). The general form of the modularity measure for undirected graphs can be defined as:

$$Q = \frac{1}{2|E|} \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} \left(A_{ij} - P_{ij} \right) \delta \left(C_i, C_j \right) \tag{31}$$

where:

- $|E|$ – number of edges,
- $|V|$ – number of nodes,
- A_{ij} – an element of the adjacency matrix,
- P_{ij} – an expected number of edges between i-th and j-th node in the null model,
- C_i – a cluster to which the i-th node belongs,
- $\delta \left(C_i, C_j \right)$ – a function which returns 1 if the i-th and j-th node belong to the same cluster (it means that $C_i = C_j$) and returns 0 otherwise.

The idea of modularity calculations proposed by Newman & Girvan was adapted to weighted bipartite networks by Barber (2007). Efficient algorithms for finding the optimal division of a given graph into clusters were proposed by Dormann & Strauss (2013) and Beckett (2016).

3 Bipartite competency schemas of the Polish labour market

The methodology presented in Section 3 was used for the analysis of the demand for employee competencies taking into account the location of enterprises seeking candidates for employment. The information about locality was represented by the V set.

3.1 Network building

During empirical research, a set of 6667 job offers retrieved from the https://www.pracuj.pl/ portal was analysed. Analyses of offers allowed for the identification of two crucial elements for every offer:

- list of required competences,
- list of localities in which a given position was offered (very often in one job offer many localities were mentioned).

Identification of competencies listed in offers was realized with the use of the ontology-based exploratory system for analysis of textual documents presented in a study by Lula, Oczkowska, Wiśniewska & Wójcik (2018). During the analysis, 61 competencies belonging to three groups (individual, social and professional) were found. The number of occurrences of particular competencies fluctuated from several (P_BEAUTY – 2, P_ENERGY – 3, O_DEAL_STRESS - 4) to above two thousand (P_GEN_LAN – 2061, P_SELL - 2517). The number of occurrences for competencies taken into account in the study is presented in Figure 1.

**Figure 1 | Number of occurrences for competencies**
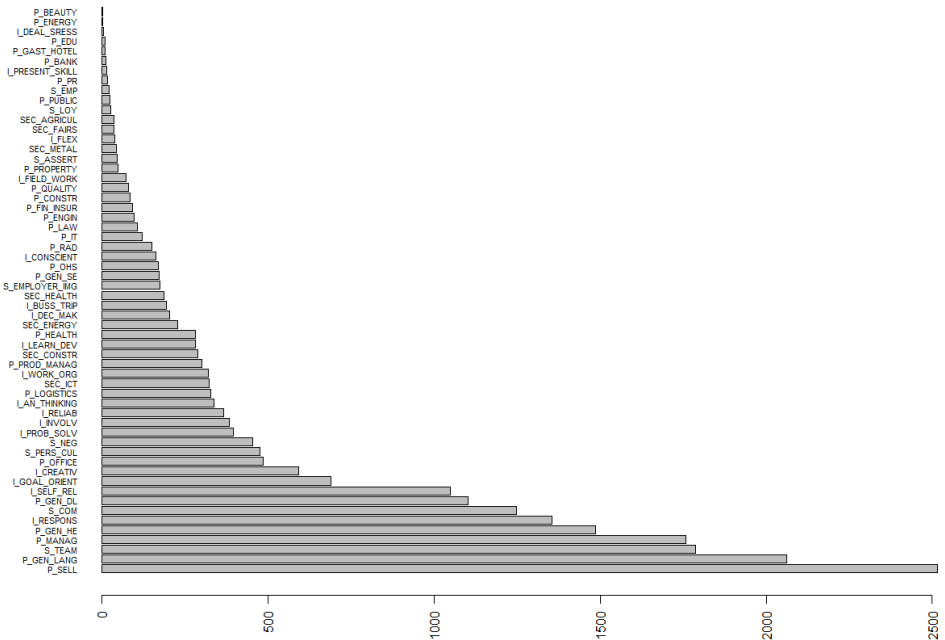
Source: own elaboration

Job offers studied during the research were located in 962 different localities but during further analysis only localities mentioned in at least 75 offers were taking into consideration

(this limit decreased the number of localities to 60). The number of occurrences for them is presented in Figure 2.

**Figure 2 | Number of occurrences for localities**
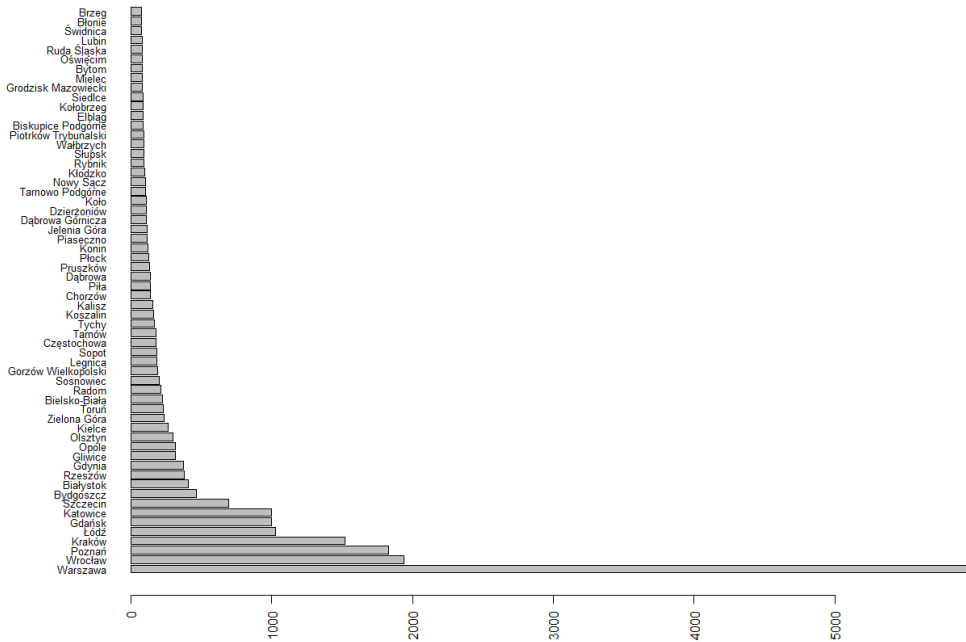
Source: own elaboration

Data described above were used for building a bipartite graph with nodes belonging to two groups: competencies and localities and the interaction matrix G having the following structure:

$$Competencies$$

$$Localities \quad \begin{bmatrix} g_{11} & \cdots & g_{1M} \\ \vdots & \ddots & \vdots \\ g_{N1} & \cdots & g_{NM} \end{bmatrix}$$

The element g_{ij} informs how many times the j-th competence was mentioned in the context of the i-th locality. The structure of the G network is presented in Figure 3.

Figure 3 | Bipartite interaction graph presenting relations between competencies and localities
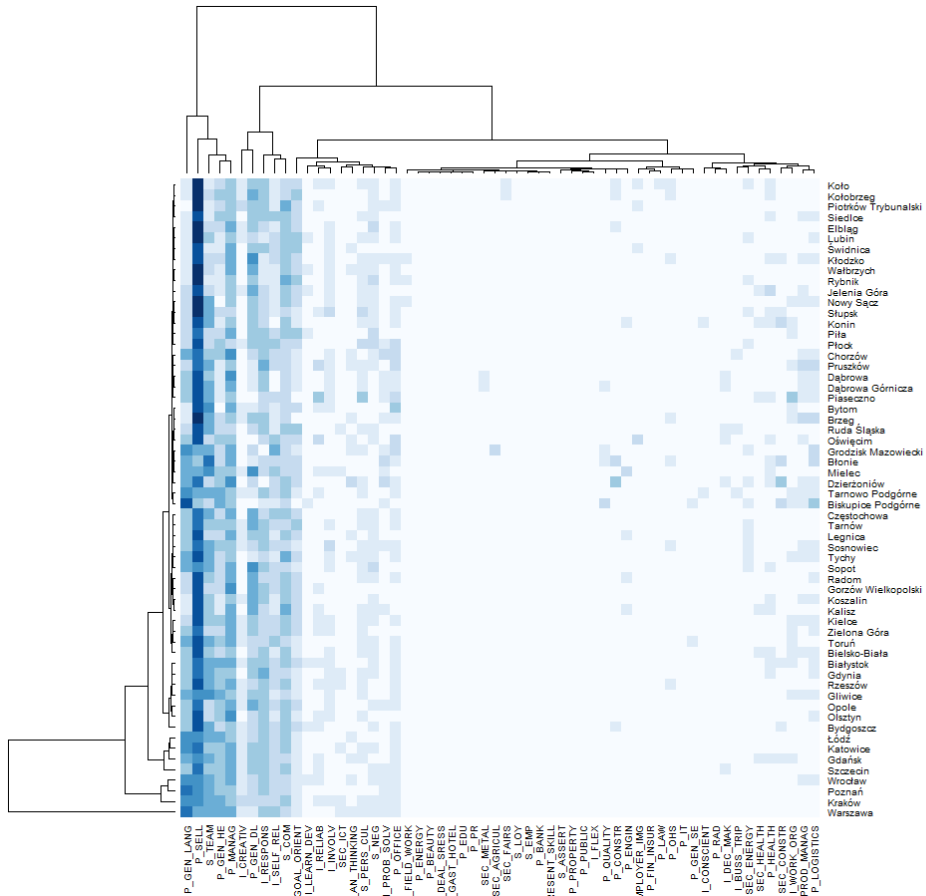
Source: own elaboration

3.2 Network analysis

The analysis of the bipartite graph G was realized on three levels: network, groups and nodes.

3.2.1 Network statistics

The main network statistics are presented in the Table 16.

Table 16 | Network statistics for the G graph.

Statistics	Value
Connectance	0.607
Web asymmetry	-0.017
Mean number of links per nodes	17.744
Number of compartments	1.000
Cluster coefficient	0.583
Specialization index H'_2	0.026

Source: own elaboration

The value of the connectance index shows that slightly above 60% of elements in the G matrix are different than zero. Web asymmetry confirms that the number of localities is a little greater than the number of competencies. The number of connections in the graph in relation to one node is equal to 17.74. The G graph is connected (number of compartments = 1). The relatively high value of the cluster coefficient (0.583) also indicates the high density of connections. Whereas the small value of H'_2 suggest that the specificity of competences and localities is low.

3.2.2   Group statistics

Group statistics calculated for the graph G are presented in Table 17.

Table 17 | Group statistics for the G graph

| Statistics	Value
Number of nodes (competencies)	58
Number of nodes (localities)	60
Mean number of partners (competencies)	54.809
Mean number of partners (localities)	47.223
Mean number of shared partners (competencies)	22.824
Mean number of shared partners (localities)	27.458
Cluster coefficient (competencies)	0.914
Cluster coefficient (localities)	0.814
Niche overlap (competencies)	0.764
Niche overlap (localities)	0.885

Source: own elaboration

All values presented in Table 17 confirm high density of connections and low specificity of competencies (mean number of shared partners = 54.81, cluster coefficient = 0.91, niche overlap index = 0.76) and localities (mean number of shared partners = 47.22, cluster coefficient = 0.81, niche overlap index = 0.88).

3.2.3   Node statistics

At the node's level the number of connections to partner's nodes and node's specificity was analysed. The analysis was performed separately for nodes representing competencies and nodes representing localities.

Figure 4 shows the relation between nodes' degree and their specificity.

Figure 4 | The relation between the degree and specificity index for competencies
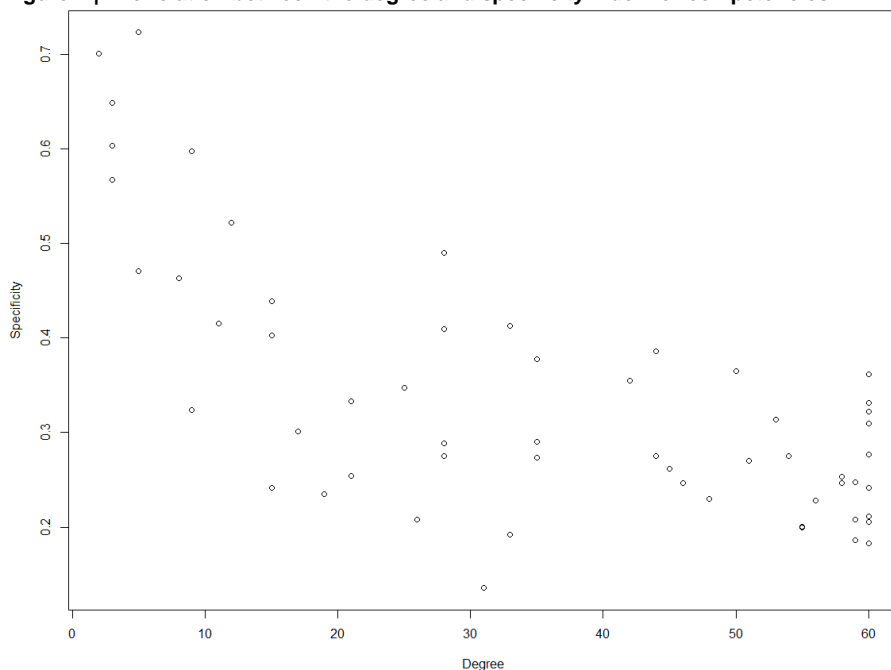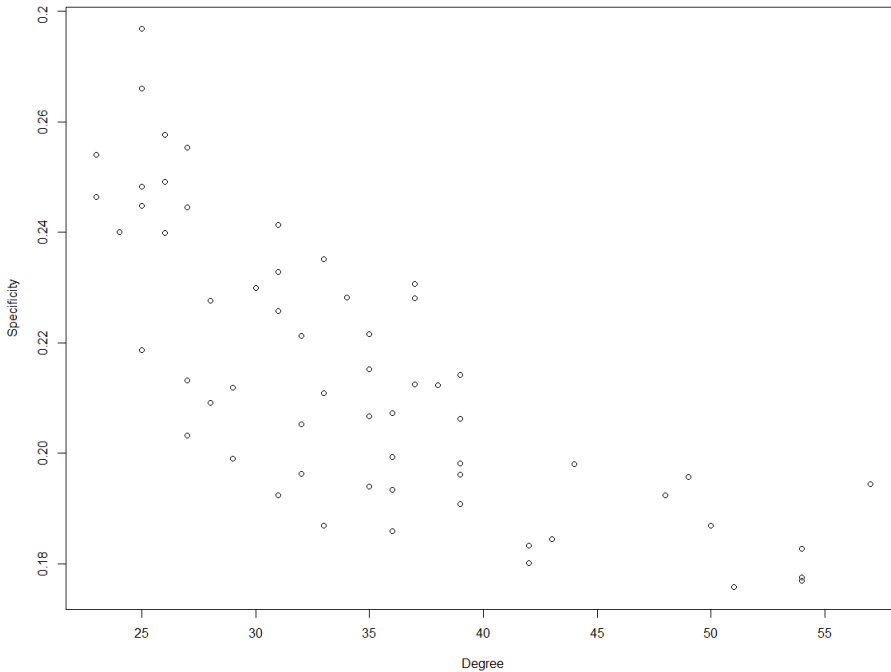

Source: own elaboration

Generally, the level of specificity is low (smaller than 0.5). On the left side of the picture, some competencies having higher value of specificity index can be observed. It seems, however, that their higher specificity was caused by their infrequent appearance in job offers. This can be observed for the following competencies:

- P_PUBLIC – specificity: 0.598, number of occurrences: 24,
- I_DEAL_STRESS – specificity: 0.604, number of occurrences: 4,
- P_BANK – specificity: 0.649, number of occurrences: 12
- P_BEAUTY – specificity: 0.701, number of occurrences: 2,
- I_PRESENT_SKILL – specificity: 0.723, number of occurrences: 14.

After excluding competencies appearing sporadically, the rest of them are distributed among many different localities. For localities, values of the specificity index are also low (Figure 5).

Figure 5 | The relation between node's degree and specificity index for localities
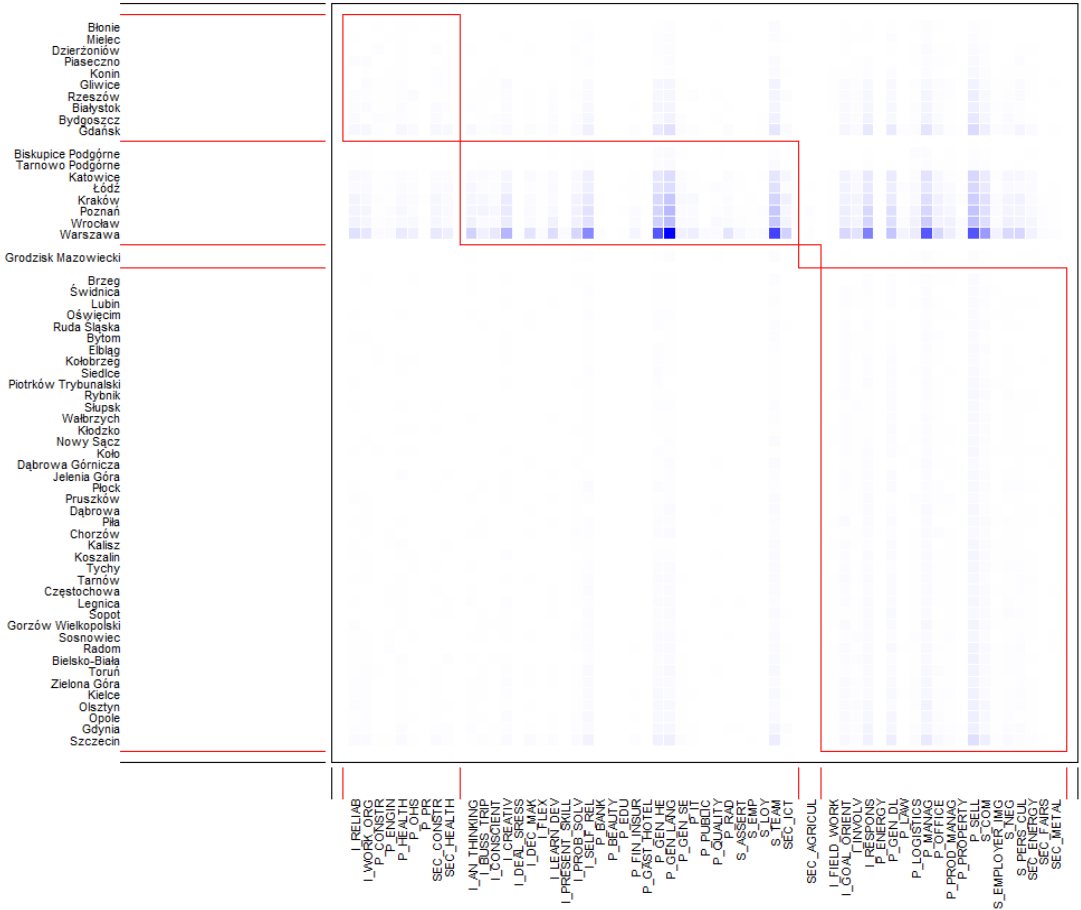

Source: own elaboration

The graph shows the explicit relationship between the value of degree and specificity.

3.3 Bipartite competency schemas identification

Network statistics confirmed that in the graph G only one compartment exists. It indicates the lack of possibility of unambiguous identification of connected components which can be treated as competency schemas. Therefore, cluster analysis of the bipartite graph was performed to identify sub graphs containing strongly related nodes representing competencies and localities. The results are presented in Figure 6.

Figure 6 | The result of cluster analysis of the graph G

Source: own elaboration

The analysis allowed for the identification of four clusters which can form bipartite competency schemas. Schemas identified here are very indistinct and ambiguous. But this result was previously indicated by statistics confirming very low level of competencies' and localities' specificity.

Conclusions

The research confirmed the usefulness of bipartite graph models in labour market analysis. However, it seems that bipartite competency schemas do not replace competency schemas proposed by the authors in their previous works, but they complement their capabilities and remove some of their disadvantages. The enrichment of the competency analysis by including information about the analysis context is the main advantage of the proposed approach. Furthermore, it allows for the:

- estimation of the importance and the specificity of competencies and context factors,

- analysis of the relationships between them,
- identification of bipartite competency schemas in the form of strongly related nodes existing in the bipartite models of the labour market.

All analysis presented in the paper was performed with the use of publicly available data and R programming language.

It seems that the methods used here can be useful in various research projects focused on the analysis of ties between two (or more) groups of objects. They allow for the identification of significant interactions, evaluation of their specificity and finding of groups of objects closely connected.

In further research, the authors are going to extend the analysis to other countries and estimate the competency gap for various labour markets.

Acknowledgement

References

Antonelli, D., Bruno, G., Taurino, T., & Villa, A. (2015). Graph-based Models to Classify Effective Collaboration in SME Networks. *International Journal of Production Research*, 53(20), 6198-6209.

Barber, M. J. (2007). Modularity and Community Detection in Bipartite Networks. *Physical Review E*, 76(6), 066102.

Becket, S. J. (2016). Improved Community Detection in Weighted Bipartite Networks. *Royal Society Open Science*, 3(1), 1-18.

Blüthgen, N., Menzel, F., & Blüthgen, N. (2006). Measuring Specialization in Species Interaction Networks. *BMC Ecology*, 6(1), 1-12.

Dormann, C., Fründ J., Blüthgen N., & Gruber B. (2009). Indices, Graphs and Null Models: Analyzing Bipartite Ecological Networks. *The Open Ecology Journal*, 2, 7-24.

Dormann, C.F., Gruber B., & Fründ, J. (2008). Introducing the Bipartite Package: Analysing Ecological Networks. *R news*, 8(2), 8-11.

Dormann, C. F., & Strauss, R. (2013). *Detecting Modules in Quantitative Bipartite Networks: the QuaBiMo Algorithm*. Retrieved May 10, 2019, from https://arxiv.org/pdf/1304.3218.pdf.

Fortunato, S. (2010). *Community Detection in Graphs*. Retrieved May 10, 2019, from https://arxiv.org/pdf/0906.0612.pdf.

Gadar, L., & Abonyi, J. (2018). Graph Configuration Model Based Evaluation of the Education-occupation Match. *PLoS ONE*, 13(3). Retrieved May 10, 2019, from https://doi.org/10.1371/journal.pone.0192427.

Guerrero, O.A., & Axtell R.L. (2013). Employment Growth through Labour Flow Networks. *PLoS ONE*, 8(5), pp.1-12. Retrieved May 10, 2019, from https://doi:10.1371/journal.pone.0060808.

Guerrero, O. A., & Lopez E. (2015). Labor Flows and the Aggregate Matching Function: A Network-Based Test Using Employer-Employee Matched Records. *SSRN Electronic Journal*. Retrieved May 10, 2019, from https://doi: 10.2139/ssrn.2631045.

Horn, H.S. (1966). Measurement of "Overlap" in Comparative Ecological Studies. *The American Naturalist*, 100(914), 419-424.

Kvyatkovskaya, I., Sibikina, I., & Berezhnov, G. (2013). Procedure of the System Characteristics Competence Graph Model Calculation. *World Applied Sciences Journal. Information Technologies in Modern Industry, Education and Society*, 24, 111-116.

Latapy, M., Magnien, C., & Del Vecchio, N. (2011), *Basic Notions for the Analysis of Large Two-mode Networks*. Retrieved May 11, 2019, from http://www.complexnetworks.fr/wp-content/uploads/2011/01/socnet07.pdf.

Lloret-Climent, M., Nescolarde-Selva, J., Mora-Mora, H., & Signes-Pont, M. T. (2018). A New Network Perspective in the Study of Labour Markets. *Mathematical Methods in the Applied Sciences*, 41(6), 2261-2268.

Lomolino, M. V., Riddle, B. R. & Brown, J. H. (2009). *Biogeography*. Sunderland: Sinauer Associates, Inc.

Lula, P., Oczkowska, R., Wiśniewska, S., & Wójcik, K. (2018). Ontology-Based System for Automatic Analysis of Job Offers. In J. Ministr & M. Tvrdíková (Eds.), *Information Technology for Practice 2018* (pp. 205-212). Ostrava: VSB-Technical University.

Lula, P., Oczkowska, R., Kovaleva, A., & Wiśniewska, S. (2019). Methods of Competency Schema Identification and Evaluation. *Proceedings on Engineering Sciences*, 1(2), 11-18.

McDonald, S., & Benton, R. (2017). The Structure of Internal Job Mobility and Organizational Wage Inequality. *Research in Social Stratification and Mobility*, 47, 21-31.

Newman, M. E. J., & Girvan, M. (2004). Finding and Evaluating Community Structure in Networks. *Physical Review E*, 69, 026113, 1-15.

Pavlopoulos, G. A., Kontou, P. I., Pavlopoulou, A., Bouyioukos C., Markou, E., & Bagos, P. G. (2018). Bipartite Graphs in Systems Biology and Medicine: A Survey of Methods and Applications. *GigaScience*, 7(4). Retrieved May 10, 2019, from https://doi.org/10.1093/gigascience/giy014.

Poisot, T., Canard, E., Mouquet, N., & Hochberg, M. E. (2012). A Comparative Study of Ecological Specialization Estimators. *Methods in Ecology and Evolution*, 3, 537-544. https://doi: 10.1111/j.2041-210X.2011.00174.x.

Razakanirina, R. M. and Chopard, B. (2012). Dynamics of Artificial Markets on Irregular Topologies. In T. Gilbert, M. Kirkilionis, G Nicolis (Eds.), *Proceedings of the European Conference on Complex Systems 2012* (pp. 1019-1031). Springer Proceedings in Complexity. https://doi:10.1007/978-3-319-00395-5_123.

Tassier, T., & Menczer, F. (2001). Emerging Small-world Referral Networks in Evolutionary Labor Markets. *IEEE Transactions on Evolutionary Computation*, 5(5), 482-492.

Uliana, R.M., & de Castro, L.N. (2018), Identifying Career Boundaries Using Minimum Description Length on a Graph. *IEEE Access*, 6, 42407-42421.

Watts, D. J., & Strogatz, S. (1998). Collective Dynamics of 'Small-world' Networks. *Nature*, 393, 440-442.