

Analysis of the social impact of research activities carried out in the area of selected social disciplines in Poland and in the UK

Paweł Lula

Department of Computational Systems, Krakow University of Economics, Poland

e-mail: lulap@uek.krakow.pl

ORCID: [0000-0003-2057-7299](https://orcid.org/0000-0003-2057-7299)

Magdalena Talaga

Krakow University of Economics, Poland

e-mail: talagam@uek.krakow.pl

ORCID: [0000-0003-4110-0449](https://orcid.org/0000-0003-4110-0449)

Agata Dobranowska

Krakow University of Economics, Poland

e-mail: dobranoa@uek.krakow.pl

ORCID: [0009-0009-3460-1764](https://orcid.org/0009-0009-3460-1764)

©2024 Paweł Lula, Magdalena Talaga, Agata Dobranowska

This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/4.0/>

Quote as: Lula, P., Talaga, M., & Dobranowska, A. (2024). Analysis of the social impact of research activities carried out in the area of selected social disciplines in Poland and in the UK. *Argumenta Oeconomica*, 2(53), 52-65.

DOI: [10.15611/aoe.2024.2.04](https://doi.org/10.15611/aoe.2024.2.04)

JEL: C63

Abstract: The analysis of the social impact of research activity carried out in the area of selected social disciplines in the United Kingdom and in Poland was the main goal of the paper. The research process was conducted with the use of documents prepared for the evaluation procedures taking place in the United Kingdom in 2014 and in Poland in 2022 and presenting cases of the impact of research activity on society and the economy. This analysis, performed with the use of exploratory text analysis methods (particularly with the latent Dirichlet allocation model), allowed to identify topics and evaluate their significance in documents prepared by individual universities or groups of universities. A comparative analysis of the contents of descriptions was also performed. All the analyses presented in the paper were conducted by the authors with the use of computer programs written in the R language.

Keywords: evaluation of the quality of research activity, social impact of research activity, topic modelling, latent Dirichlet allocation method, Research Excellence Framework

1. Social impact as the basic function of universities in light of legal regulations in the United Kingdom and Poland

The Law on Higher Education and Science, introduced in Poland in 2018, empowered the so-called third mission of the higher education and research system, understood as the participation of universities in the social development and the development of the economy based on innovation. In 2016, the Ministry of Science and Higher Education, in the White Paper of Innovation, included a provision on the introduction of the *social impact* – an element in the assessment of individual institutions following the example of the United Kingdom, i.e. the impact that universities have on the economy and society. Thus, during the evaluation of the quality of research activity carried out in 2022, Polish universities for the first time were obliged to prepare the so-called ‘impact descriptions’ and complementary evidence of impact.

The assessment of the social impact that universities have on the environment is one of the evaluation criteria adopted in several European countries, and also used outside Europe. The Polish solution is modelled on the Research Excellence Framework, a British system for assessing the quality of research carried out in the UK for the first time in 2014 (Wróblewska, 2017, 2021).

In the British assessment of research institutions, social impact was defined as change, benefit or in general the impact on the economy, society, culture, legislation and budgetary sphere, health, environment or quality of life, outside the world of science. Social impact refers to the impact on behaviour, practices, knowledge, awareness, and understanding; this may also include the limitation of damage and/or the reduction of risk.

The scientific research included in the assessment of the impact of British universities on the environment was carried out in the period 1993-2013, while the social impact itself had in 2008-2013.

The assessment for the British descriptions was based on the descriptions of case studies created according to a mandatory template consisting of the following elements:

1. summary of the described *social impact*,
2. description of the research activity which caused the studied effect,
3. publications presenting the results of the research,
4. description of *social impact*,
5. facts confirming the social impact – its existence and quality.

The main task faced by the authors of the case studies was to demonstrate cause-and-effect relationships between research and changes in the external world. The evaluation of this criterion was based on an expert assessment of the individual impact descriptions. The descriptions were assessed in terms of the extent and significance of the impact.

In the Polish system of assessing the quality of research activity, the legislator did not indicate a clear definition of social impact. In the prepared guidelines, however, it is pointed out that the social impact that is the impact of scientific activity on the functioning of society and the economy shows that research activity carried out in institutions of higher education affects not only the academic world, but also the socio-economic environment, thus helping society understand culture and history. This description was the starting point for the institutions assessed within the meaning of the third evaluation criterion.

Only the scientific research carried out from 1 January 1997 to 31 December 2021 was included in the assessment. The prior activities or those planned for realisation after this period could be taken into account only if they were performed during the period indicated by the evaluated institution. In turn, the social impact itself had to appear in the time line mentioned above. The impact occurring outside this period was not taken into consideration.

The basis of the evaluation within the third criterion was based on two types of documents:

1. Impact description form – prepared in accordance with the evaluator's guidelines, registered in the Integrated Information System on Science POL-on.
2. Evidence of influence – placed in the system of making available documentation of scientific and artistic activity STUDNIA.

The impact description form, valid for all the institutions assessed in the same form, consisted of the following information:

1. year of evaluation, field, discipline, type,
2. title,
3. abstract,
4. characteristics of the main conclusions of scientific research or development work,
5. role of the evaluated institution,
6. description of publications constituting the theoretical background of the social impact,
7. description of the social impact of scientific activity,
8. evidence of the social impact,
9. facts justifying the interdisciplinary character of the research activity.

An evaluation of the Polish social impact was performed, in line with the British assessment, in the form of an expert's impact based on the above-mentioned types of documents, within the scope and the significance of the impact. Such assessment could be extended thanks to the interdisciplinarity of the scientific research presented in the prepared impact description.

2. Research methodology

The analysis of the social impact of research activity carried out in selected areas of social sciences at British and Polish universities was the main goal of the research process. Documents prepared by higher education institutions during evaluation processes conducted in the UK in 2014 and in Poland in 2022 were used as the main source of information. This section presents consecutive steps of analysis required for the assessment of social impact of universities in the UK and in Poland and their comparison.

2.1. Retrieving documents and corpus construction

The analysis of the social impact of research activity carried out in Poland and in the United Kingdom was performed based on cases of impact described for the evaluation purpose. All cases of impact prepared by British universities in 2014 are available at: <https://impact.ref.ac.uk/casestudies/Search1.aspx>. Similarly, Polish cases of impact prepared in 2021 are stored at: <https://radon.nauka.gov.pl/dane>. Data retrieved from both web sites were transformed into plain text format and saved as separate files creating the corpus D which formally can be defined as a set of M documents (Eq. 1):

$$D = \{D_1, D_2, \dots, D_M\}, \quad (1)$$

where every document D_j is a sequence of tokens having a form of words or punctuation marks.

All the words were transformed into their basic form during the lemmatisation process. Next, the following two groups of words were removed:

- those appearing very rarely in the corpus,
- those which are not nouns or adjectives.

Next, the process of identification of relevant multi-word sequences was carried out using the rapid automatic keyword extraction) (RAKE) algorithm (Rose et al., 2010), which belongs to machine learning

solutions based on unsupervised training. This means that the training sets with keywords or key-phrases identified manually are not necessary. Below are the following stages in the RAKE algorithm:

- splitting the text into *candidate key-phrases* in positions in which punctuation marks or irrelevant words (included into a stop-list) appear;
- generating a graph $G = \{V, E\}$ of word co-occurrence in which:
 - V is a set of lemmas of words included into candidate key-phrases which are used as vertices (Eq. 2):

$$V = \{l_1, l_2, \dots, l_C\}, \quad (2)$$

- edges are created between two vertices representing words appearing at least once in the same candidate key-phrase,
- weights assigned to edges inform how many times a given pair of words appeared jointly at the same candidate key-phrase,
- weights assigned to vertices inform about the number of occurrences of words represented by the given lemma;
- evaluation of the relevance of the words' lemmas performed by calculating one of three relevance measures:

$$relevance(l_i) = freq(l_i), \quad (3)$$

$$relevance(l_i) = freq(l_i) + deg(l_i), \quad (4)$$

$$relevance(l_i) = \frac{freq(l_i) + deg(l_i)}{freq(l_i)}, \quad (5)$$

where $freq(l_i)$ denotes the number of occurrences of the word represented by the l_i lemma and $deg(l_i)$ represents the degree of the vertex corresponding to the l_i lemma,

- calculation of the relevance of candidate key-phrases by adding measures of relevance for every word appearing in a given key-phrase,
- for sequences composed of two candidate key-phrases separated by a word from a stop-list which appeared at least twice, the operation of merging into one key-phrase is performed,
- sorting candidate key-phrases decreasing by their relevance and taking approximately 30% of items as key-phrases,
- adding identified key-phrases as tokens in place of tokens corresponding to words forming these phrases.

As a result of the above algorithm, every document can be defined as a sequence of tokens taken from the set:

$$W = \{w_1, w_2, \dots, w_N\}, \quad (6)$$

where w_i is a lemma of a single word or a sequence of lemmas corresponding to words forming a key-phrase.

Then each document D_j (where $j = 1, \dots, M$) can be treated as a sequence of tokens:

$$D_j = (w_{j,1}, \dots, w_{j,Q_j}), \quad (7)$$

where Q_j is a number of tokens in document D_j .

a. Identification of topics with the Latent Dirichlet Allocation algorithm

The main assumption of the topic modelling is that documents cover issues related to P topics: T_1, \dots, T_P (Blei et al., 2003).

Topics are represented by matrix Φ :

$$\Phi = \begin{bmatrix} \phi_{1,1} & \dots & \phi_{1,N} \\ \dots & \dots & \dots \\ \phi_{P,1} & \dots & \phi_{P,N} \end{bmatrix}, \quad (8)$$

The T_k ($k = 1, \dots, P$) topic is represented by k -th row of matrix Φ with N elements representing the probability of occurrence of w_i token in T_k topic.

Vector ϕ_k representing T_k topic meets the condition:

$$\sum_{i=1}^N \phi_{k,i} = 1. \quad (9)$$

The identified topics can be used for modelling the contents of documents from corpus D as every document D_j can be presented as a mixture of topics T_1, \dots, T_P . The contribution of topics in documents is described by matrix θ :

$$\theta = \begin{bmatrix} \theta_{1,1} & \dots & \theta_{1,P} \\ \dots & \dots & \dots \\ \theta_{M,1} & \dots & \theta_{M,P} \end{bmatrix}, \quad (10)$$

where $\theta_{j,k}$ describes the probability of occurrence of topic T_k in document D_j .

The description of document D_j meets the requirement:

$$\sum_{k=1}^P \theta_{j,k} = 1. \quad (11)$$

In the latent Dirichlet allocation model it is assumed that ϕ_k is a random variable from the Dirichlet distribution with parameter $\beta_k = [\beta_{k,1} \dots \beta_{k,N}]$. This means that the density function describing ϕ_k has the form:

$$p_k(\phi_{k,1}, \dots, \phi_{k,N} | \beta_{k,1}, \dots, \beta_{k,N}) = \frac{1}{B(\beta_{k,1}, \dots, \beta_{k,N})} \prod_{i=1}^N \phi_{k,i}^{\beta_{k,i}-1}, \quad (12)$$

where

$$B(\beta_{k,1}, \dots, \beta_{k,N}) = \frac{\prod_{i=1}^N \Gamma(\beta_{k,i})}{\Gamma(\sum_{i=1}^N \beta_{k,i})}. \quad (13)$$

Likewise, random variable θ_j describing the structure of document D_j also has the Dirichlet distribution with parameter $\alpha_j = [\alpha_{j,1} \dots \alpha_{j,P}]$. In this case, the density function has the form:

$$p_j(\theta_{j,1}, \dots, \theta_{j,P} | \alpha_{j,1}, \dots, \alpha_{j,P}) = \frac{1}{B(\alpha_{j,1}, \dots, \alpha_{j,P})} \prod_{k=1}^P \theta_{j,k}^{\alpha_{j,k}-1}, \quad (14)$$

where

$$B(\alpha_{j,1}, \dots, \alpha_{j,P}) = \frac{\prod_{k=1}^P \Gamma(\alpha_{j,k})}{\Gamma(\sum_{k=1}^P \alpha_{j,k})}. \quad (15)$$

The latent Dirichlet allocation model has a generative character. This means that it describes a process of document generation. Taking into account the fact that document D_j is a sequence of Q_j tokens $(w_{j,1}, \dots, w_{j,Q_j})$, the process of document creation can be presented as:

1. choose $\theta_j \sim \text{Dir}(\alpha)$, where $j = 1, \dots, M$ and $\text{Dir}(\alpha)$ is a symmetric Dirichlet distribution with parameter α ,
2. choose $\phi_k \sim \text{Dir}(\beta)$, where $k = 1, \dots, P$ and $\text{Dir}(\beta)$ is a symmetric Dirichlet distribution with parameter β ,

3. for each document D_j , where $j = 1, \dots, M$:
 - a. for each token $w_{j,l}$, where $l = 1, \dots, Q_j$:
 - i. choose a topic for the token in l -th position $z_{j,l} \sim \text{Multi}(\theta_j)$, where $\text{Multi}(\theta_j)$ is the multinomial distribution (generalised Bernoulli distribution),
 - ii. choose a token in l -th position $w_{j,l} \sim \text{Multi}(\phi_{z_{j,l}})$, where $\text{Multi}(\phi_{z_{j,l}})$ is the multinomial distribution (generalised Bernoulli distribution).

Interference of the latent Dirichlet allocation model was performed with the use of the collapsed Gibbs sampling (Yao et al., 2009).

2.2. Topics interpretation

The most popular strategy for assigning the meaning to the given topic is based on the analysis of its most important tokens. The implementation of this rule is possible by sorting the tokens in descending order according to the probability of their appearance in the given topic (that is according to $\phi_{k,i}$ values).

An alternative method of calculating the tokens' relevance is presented in (Sievert & Shirley, 2015). Let $r_{k,i}$ denote the relevance of w_i token for T_k topic which can be defined as:

$$r_{k,i} = \lambda \log(\phi_{k,i}) + (1 - \lambda) \log\left(\frac{\phi_{k,i}}{p_i}\right), \quad (16)$$

where p_i is the probability of appearance of token w_i in the whole corpus; expression $\frac{\phi_{k,i}}{p_i}$ determines the importance of w_i in topic T_k in relation to its importance in the whole corpus; λ is the weight belonging to the range $[0; 1]$. For $\lambda = 1$, values $r_{k,i}$ are equal to $\phi_{k,i}$, whereas for $\lambda = 0$, only the relation $\frac{\phi_{k,i}}{p_i}$ is taken into account.

Formula (16) takes into account not only the probability of the token's occurrence in the topic, but also the lift, defined as the relationship between the token's significance in the given topic and in the whole corpus.

2.3. Analysis of the achievements of groups of universities in the context of the identified topics

Let us assume that $v_G^* = [v_{G,1}^* \dots v_{G,P}^*]$ is a vector containing sums of $\theta_{j,k}$ where j indicates documents belonging to set G and $k = 1, \dots, P$ represents a topic:

$$v_{G,k}^* = \sum_{j \in G} \theta_{j,k}. \quad (17)$$

The contribution of topics in the achievements described by documents belonging to set G can be expressed by vector $v_G = [v_{G,1} \dots v_{G,P}]$ where:

$$v_{G,k} = \frac{v_{G,k}^*}{\sum_{k=1}^P v_{G,k}^*}. \quad (18)$$

Formulas (17) and (18) can be useful for the estimation of the topics' contribution to the social impact of an individual university or research discipline.

2.4. Comparative analysis of research activity’s impact of the two groups of institutions

The comparative analysis of documents belonging to two groups, G_I and G_{II} , can be based on some similarity measures between vectors v_{G_I} and $v_{G_{II}}$ describing the contribution of the topics within these two groups of documents.

Taking into account the properties of v_{G_I} and $v_{G_{II}}$, it seems justified to use the Kullback-Leibler divergence as a measure of similarity (Kullback & Leibler, 1951):

$$sim(v_{G_I}, v_{G_{II}}) = \frac{KL(v_{G_I}, v_{G_{II}}) + KL(v_{G_{II}}, v_{G_I})}{2}, \tag{19}$$

where $KL(v, u)$ is the Kullback-Leibler divergence between two distributions v and u and is calculated as:

$$KL(v, u) = \sum_{k=1}^P v_i \log\left(\frac{v_i}{u_i}\right). \tag{20}$$

The justification for formula (19) is the fact that the Kullback-Leibler divergence is not symmetrical. This measure reflects the difference between two distributions and, depending on the logarithm base, is expressed in bits, dits or nips.

3. Social impact of research activity of British universities

The analysis of British universities was performed based on 890 case studies prepared for the 2014 evaluation by 126 institutions within four research disciplines: *Politics and International Studies, Law, Economics and Econometrics and Business and Management Studies*.

Using the latent Dirichlet allocation method, six topics were found. For every topic, a list of the most relevant words was identified (Figure 1).

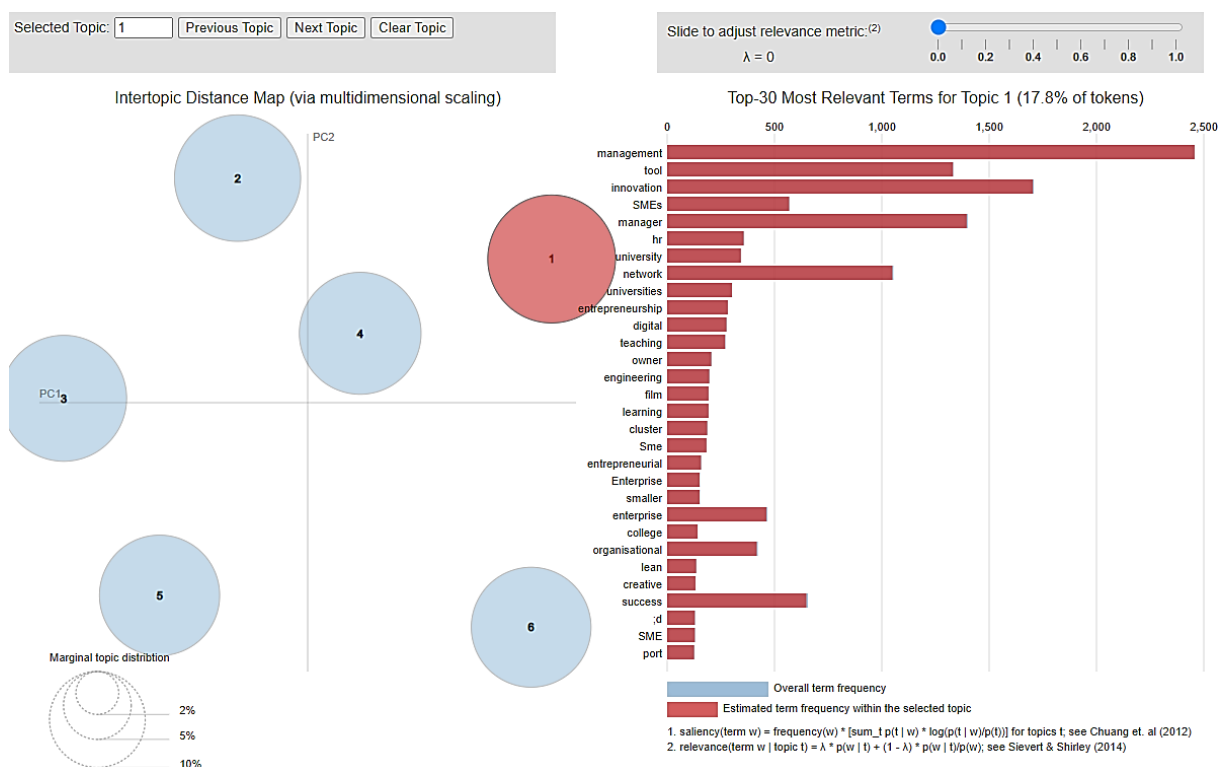


Fig. 1. Topics identified for British universities.

Source: own elaboration.

The list of words, shown on the right side of the figure, allowed to assign the label to the given topic. The characteristics of topics is presented in Table 1.

Table 1. Topics identified in the documents presenting the social impact of research activity of British universities in selected areas within social sciences.

Topic number	Topic name	The most relevant words
1	Small and medium enterprise. Innovation	<i>management, tool, innovation, SME, manager, HR, university, network, entrepreneurship, digital, teaching, owner, engineering, film, learning, cluster, enterprise, smaller, college, organizational, learn, creative, success, port</i>
2	Foreign policy	<i>security, conflict, military, actor, OECD, peace, land, intelligence, corruption, ethic, democracy, governmental, testimony, war, governance, foreign, aimed, club, post-conflict, mission, African, indicator, Asian, ecosystem, ministry, Arab, sport, gender equality</i>
3	Internal policy	<i>parliament, Scottish, election, inquiry, parliamentary, citizen, politician, newspaper, journalist, referendum, devolution, alcohol, radio, select, blog, Australian, prominent, candidate, democratic, magazine, final report, vote, daily, commentator, petition, quota, commons, distinct, minister, minority</i>
4	Employee matters	<i>employment, union, employer, health, equality, charity, arrangement, wage, career, healthcare, patient, poverty, minimum wage, research team, merger, hospital, labour, worker, public sector, adult, youth, evidence base, association, population, skills, social, scheme, NHS, unemployment</i>
5	Legal system	<i>law, legal, rights, court, legislation, justice, human rights, criminal, judge, guideline, lawyer, protection, victim, enforcement, parent, code, prison, act, judgement, offender, appeal, asylum, treatment, offence, terrorism, jurisdiction, police, prisoner</i>
6	National economy and public finance	<i>market, bank, regulatory, credit, competition, regulator, method, food, capital, software, revenue, banking, energy, investor, asset, modelling, cost, electricity, insurance, debt, economic, monetary policy, emission, pension, statistics, table, price, incentive, forecast</i>

Source: own elaboration.

Using the approach presented above, the contribution of topics was calculated for every discipline. The results are presented in the Figure 2.

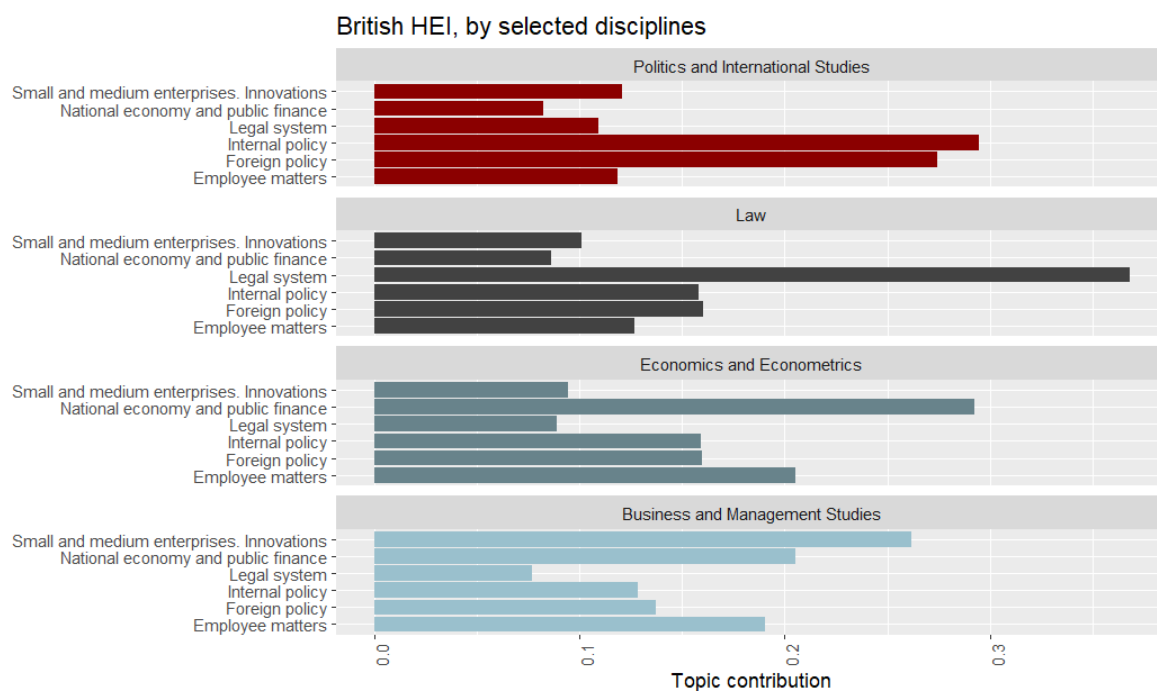


Fig. 2. The contribution of identified topics in research disciplines in the United Kingdom.

Source: own elaboration.

4. Social impact of research activity of Polish universities

For the evaluation process conducted in 2022, 94 Polish higher education institutions prepared 414 case studies in four research disciplines: *Law* (in Polish: *nauki prawne*), *Management and Quality Studies* (in Polish: *nauki o zarządzaniu i jakości*), *Politics and Administration* (in Polish: *nauki o polityce i administracji*) and *Economics and Finance* (in Polish: *ekonomia i finanse*). For evaluation purposes, the Polish universities prepared all the required documents in Polish and English. During the analysis, the documents prepared in English were used.

The main topics identified during the analysis are presented in Figure 3.

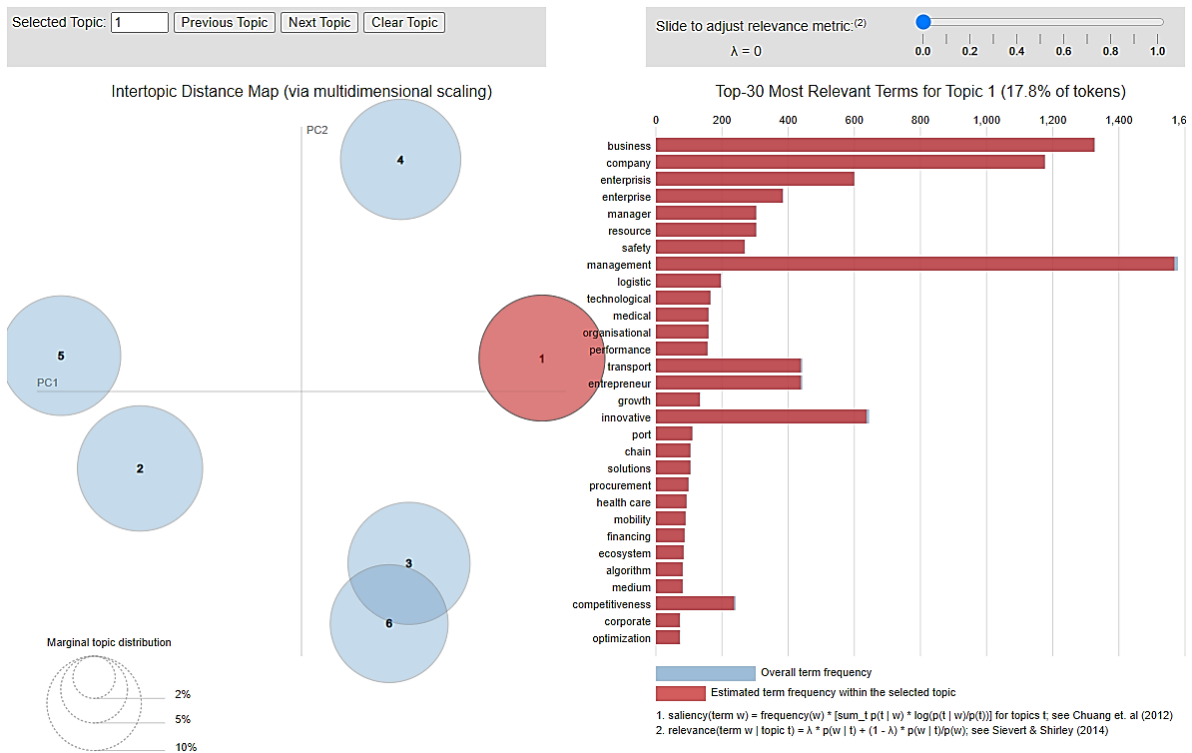


Fig. 3. Topics identified for Polish universities

Source: own elaboration.

Taking into account the list of words relevant for every topic, labels were assigned. The results are presented in Table 2.

Table 2. Topics identified in the documents presenting the social impact of research activity of Polish universities in selected areas within social sciences

Topic number	Topic name	The most relevant words
1	Enterprise management	<i>business, company, enterprise, manager, resource, safety, management, logistic, technological, medical, organizational, performance, transport, entrepreneur, growth, innovative, port, chain, solutions, procurement, health care, mobility, financing, ecosystem, algorithm, medium, competitiveness, corporate, optimization</i>
2	Foreign policy	<i>political, foreign, security, discussion, meeting, researcher, affair, debate, threat, culture, seminar, description, interested, series, armed, history, military, minority, heritage, recognition, politics, exhibition, third, foreign policy, identity, memory, migration, forum, eastern, recipients</i>

3	Regional policy	<i>local, regional, government, public administration, self, region, tourism, socio-economic, unit, smart, city, planning, voivodeship, audit, commune, powiat, revitalization, urban, functional, subsidy, spatial, local government, crisis management, modernization, province, regeneration, complex, integration, municipality</i>
4	Industry production. Agriculture. Energy.	<i>product, production, cost, agricultural, farm, consumer, agriculture, income, accounting, animal, statement, amount, insurance, plant, value, bank, emission, energy, average, valuation, software, primary, progress, price, rate, climate, farms, lower, rural, investor</i>
5	Legal system	<i>legal, law, court, administrative, provision, body, rights, tax, interpretation, legislative, draft, amendment, proceedings, judgement, act, judge, resolution, crime, juridical, provincial, matter, criminal, freedom, commentary, procedural, administration, dispute</i>
6	Education and professional skills development	<i>educational, university, school, employee, skill, student, education, program, employer, competition, internet, higher education, attention, labour market, senior, civic, secondary, assistance, humanity, qualifications, migrants, anti-corruption, vocational education, commercialization, qualification, social entrepreneurship, understood, career</i>

Source: own elaboration.

The significance of the topics was different for the disciplines considered within the research topics. The contribution of the topics into each discipline is presented in Figure 4.

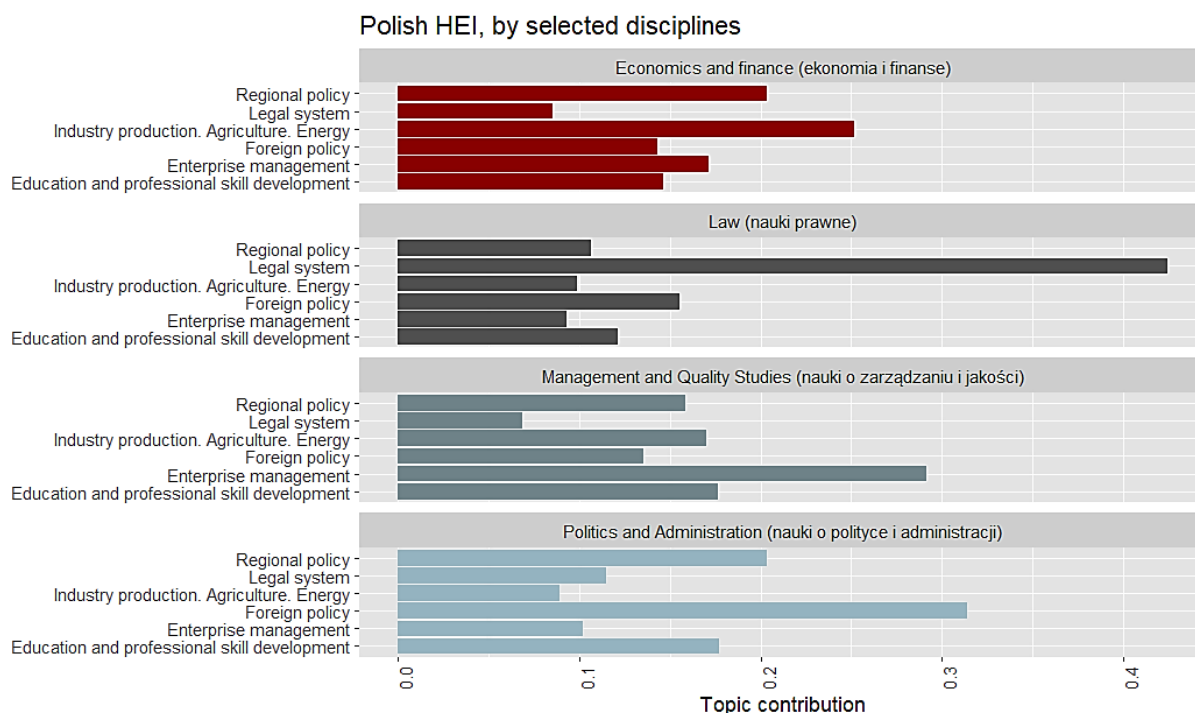


Fig. 4. The contribution of the identified topics in research disciplines

Source: own elaboration.

The analysis allowed also to build profiles showing the most important areas in which each university exerts an influence on society and the economy. The profiles for the leading Polish universities of economics are presented in Figure 5.

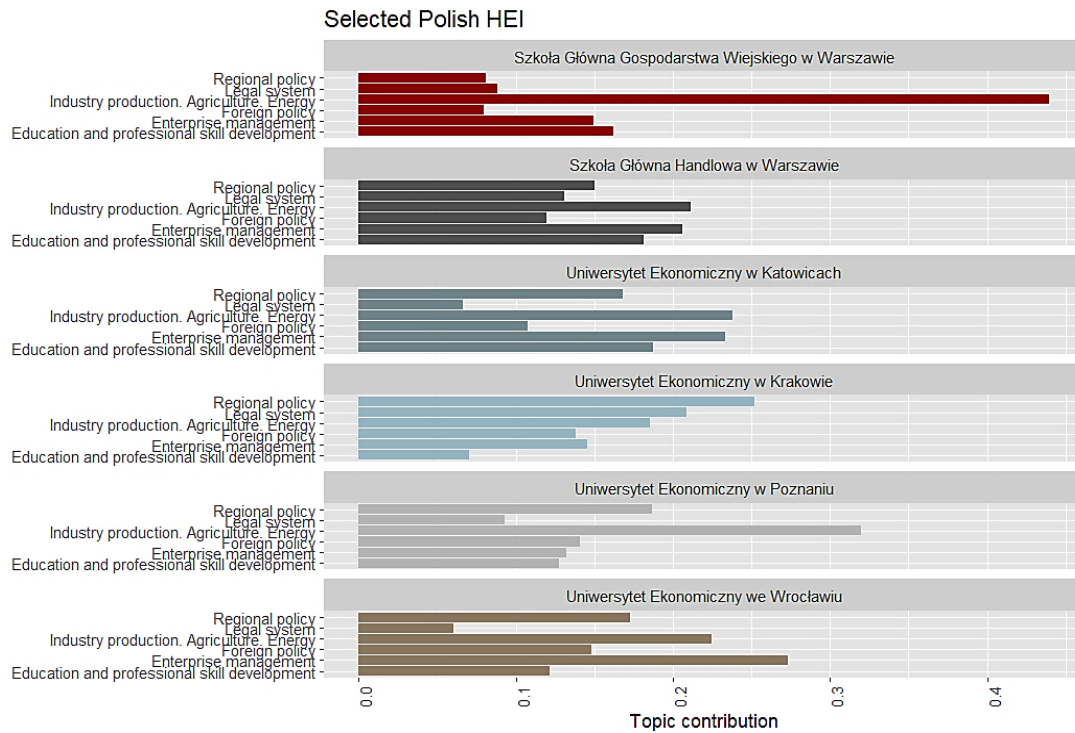


Fig. 5. The most important areas of the social impact of the leading universities of economics in Poland
Source: own elaboration.

5. Comparison of topics mentioned in case studies prepared in the UK and in Poland

A comparative analysis of the cases of impact in the United Kingdom and in Poland was conducted using the corpus containing documents prepared during the evaluation process in the UK in 2014 and in Poland in 2022. It allowed to identify nine topics presented in Figure 6.

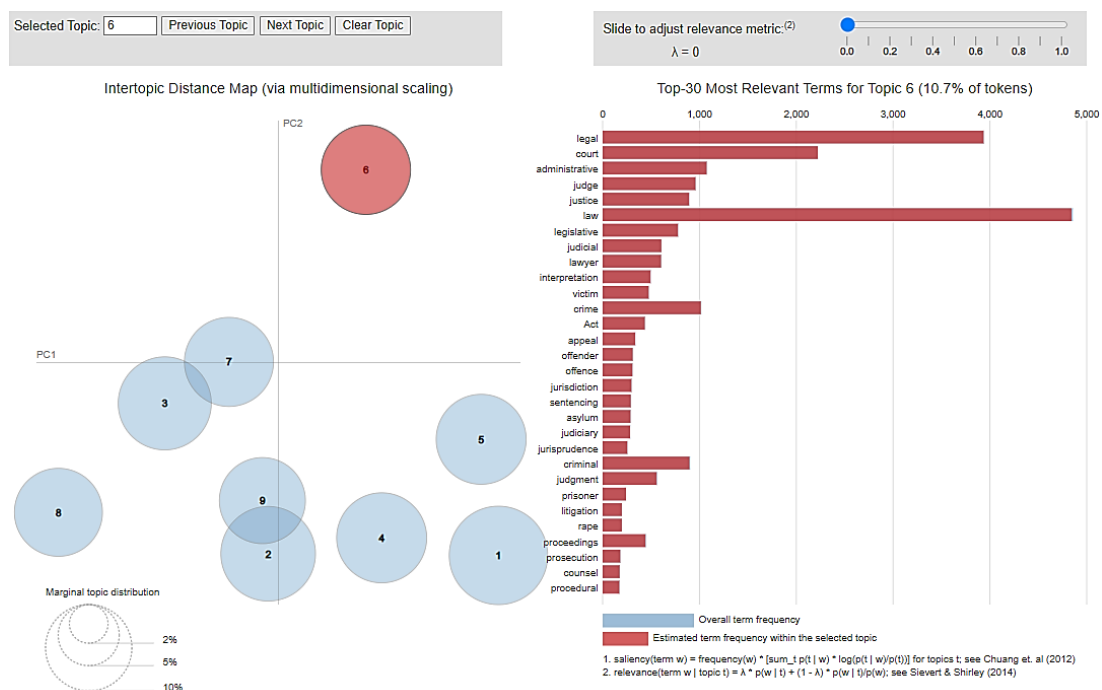


Fig. 6. Topics identified for British and Polish universities
Source: own elaboration.

The most relevant words for every topic are presented in Table 3.

Table 3. Topics identified in the documents presenting the social impact of research activity of British and Polish universities in the selected areas within social sciences

Topic number	Topic name	The most relevant words
1	Industrial policy. Energy. Production process and resources.	<i>source, chair, co-production, OECD, policy debate, minister, flagship, corporate governance, federal, taskforce, nanomaterials, renewable energy, treasury, impact, find, industrial policy, servant, undeclared work, budgeting, impact, board, written, briefs, capable, curfew, donation, fiscal policy</i>
2	Managerial issues.	<i>leadership, feedback, HR, collaborative, manager, firm, module, learning, college, teacher, lean, master, coaching, employee engagement, probation, capability, social enterprise, trainer, site, lecturer, desistance, co-operative, entrepreneurship education, mentor, sprint, HEI, improvements, personality, professionalization, volunteers</i>
3	Regional development. Knowledge transfer.	<i>scientific, Polish, achievements, cooperation, thanks, proof, entity, program, competence, functioning, description, scientific activity, tourism, nationwide, part, higher education, voivodeship, center, municipality, specialization, diagnosis, revitalization, regeneration, foreigner, border, object</i>
4	National security.	<i>conflict, peace, security, donor, war, military, ministry, delegation, intelligence, armed, foreign policy, DFID, NGO, state, foreign, headquarter, civilian, fora, corruption, arm, humanitarian, diplomat, fragile, intergovernmental, trafficking, COE, relations, western, atrocity</i>
5	Internal policy.	<i>parliament, election, parliamentary, public debate, representation, politician, newspaper, media, minority, devolution, candidate, vote, campaign, tank, petition, TV, ethic, distinct, public, inquiry, guardian, MPs, voting, Welsh, social media, privatization, conservative, voter, broadcast, poll</i>
6	Legal system.	<i>legal, court, administrative, judge, justice, law, legislative, judicial, lawyer, interpretation, victim, crime, act, appeal, offender, offence, jurisdiction, sentencing, asylum, judiciary, jurisprudence, criminal, judgement, prisoner, litigation, rape, proceedings, prosecution, counsel, procedural</i>
7	Finance and banking system	<i>cost, bank, accounting, tax, estimate, credit, estimate, pension, finance, monetary policy, price, taxation, financial, utility, fiscal, banking, sport, calculation, macroeconomic, auditor, trading, insurer, taxpayer, modelling, forecast, card, financial crisis, mortality, insurance</i>
8	Entrepreneurship and innovations	<i>innovation, production, enterprise, customer, product, SME, entrepreneurship, infrastructure, agricultural, farmer, procurement, marketing, cluster, carbon, transport, owner, smart, environmental, energy, agriculture, engineering, port, digital, entrepreneurial, plant, small, supplier, brand, chain</i>
9	Employee matters. Labour market	<i>employment, care, employer, woman, health, equality, charity, healthcare, worker, patient, age, wage, labour market, minimum wage, hospital, inequality, public sector, union, female, disability, old, medical, clinic, unemployment, wellbeing, health care, NHS, older people, male, employee ownership</i>

Source: own elaboration.

The comparison of the topics' significance in cases of the social impact of British and Polish universities is presented in Figure 7.

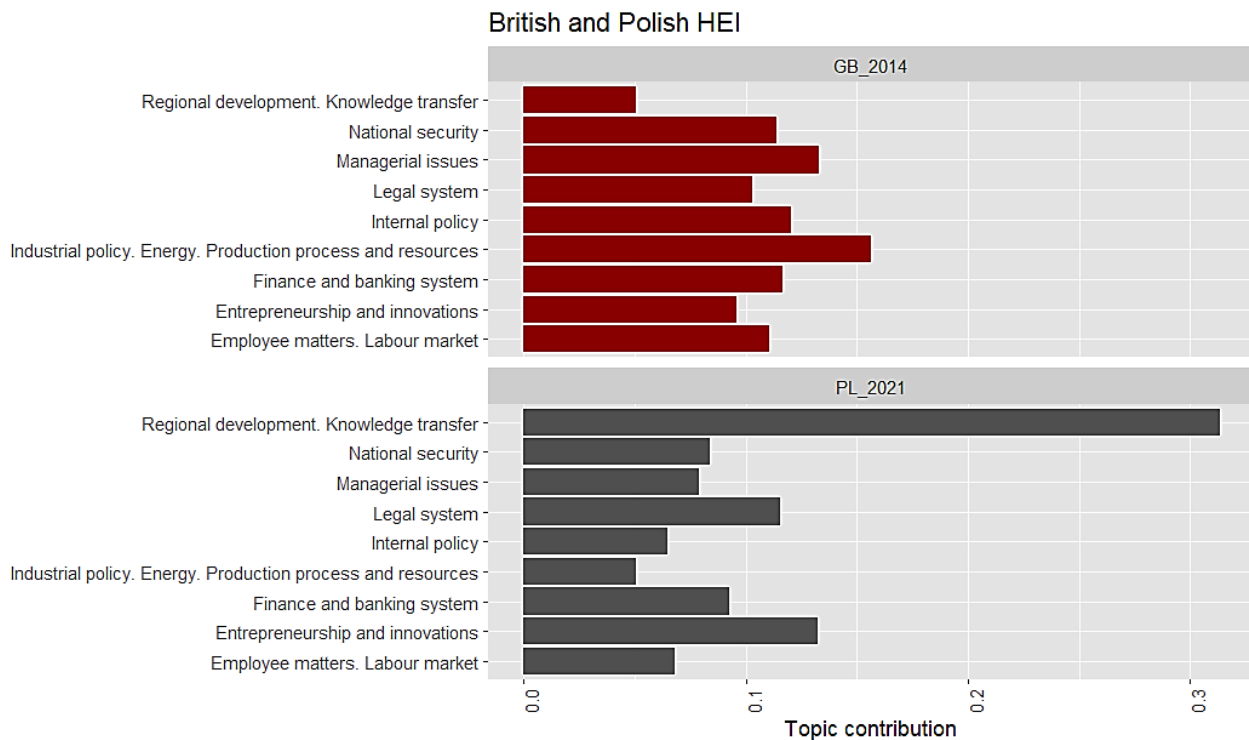


Fig. 7. Comparative analysis of the main areas of social impact in the UK and in Poland

Source: own elaboration.

For British universities the impact on industrial policy, energy and production process was the most important. In Poland issues related to regional development, knowledge transfer, entrepreneurship and innovations played a crucial role.

6. Conclusions

The research process allowed to formulate the following conclusions:

- 1) the scope of impact reported by British universities in 2014 within the disciplines *Politics and International Studies*, *Law*, *Economics and Econometrics*, and *Business and Management Studies* was broader than the one reported by Polish universities in 2022 within the disciplines: *Economics and Finance*, *Management and Quality Studies*, *Politics and Administration* and *Law*;
- 2) the topics of major importance in the United Kingdom include: national and foreign politics, the economy and security;
- 3) the most important issues in Poland are: regional economy, regional tourism, legal issues and innovations;
- 4) it seems that the latent Dirichlet allocation method is the appropriate algorithm for the identification of topics in the descriptions of social impact and in calculating their significance.

References

- Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1). <https://doi.org/10.1214/aoms/1177729694>
- Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic keyword extraction from individual documents. In *Text Mining: Theory and Applications*. John Wiley & Sons.

- Sievert, C., & Shirley, K. (2015). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70. Association for Computational Linguistics. <https://doi.org/10.3115/v1/w14-3110>
- Wróblewska, M. N. (2017). Ewaluacja „wpływu społecznego” nauki. Przykład REF 2014 a kontekst polski. *Nauka i Szkolnictwo Wyższe*, 1(49). <https://doi.org/10.14746/nisw.2017.1.5>
- Wróblewska, M. N. (2021). Research impact evaluation and academic discourse. *Humanities and Social Sciences Communications*, 8(1). <https://doi.org/10.1057/s41599-021-00727-8>
- Yao, L., Mimno, D., & McCallum, A. (2009). Efficient methods for topic model inference on streaming document collections. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. <https://doi.org/10.1145/1557019.1557121>

Received: December 2022