

Nonlinear Estimation of Similarity Among Scientists' Disciplinary Profiles. A Case Study

Veslava Osińska

ORCID 0000-0002-1306-7832

*Institute of Information and Communication Research,
Nicolaus Copernicus University in Toruń, Poland*

Oleksandr Sokolov

ORCID 0000-0002-6531-2203

*Faculty of Physics, Astronomy and Informatics,
Nicolaus Copernicus University in Toruń, Poland*

Aleksandra Mreła

ORCID 0000-0002-2059-864X

*Institute of Informatics,
Kazimierz Wielki University, Poland*

Abstract

Purpose/Thesis: Authors estimate the disciplinary similarity of researchers according to selected academic units with a different cross-section of specializations. The paper presents the model for studying disciplinary diversity of scientific units. The premise of the article is that knowledge of disciplinary profiles of researchers can be applied to create interdisciplinary teams, or one disciplinary team with a focused specializations.

Approach/Methods: The approach is based on the visualization and comparison of disciplinary space and space of co-authorship. Fuzzy logic and aggregation norm were used to calculate disciplinary weights of each journal listed in the database. For visualization, new, dimension reduction algorithm t-SNE was applied. Achieved results were verified by using the expert's knowledge.

Results and conclusions: In the evaluation of scientific collaboration, a co-authorship relationship can be complemented by researchers' disciplinary profiles represented by aggregation norm. Thanks to the continuity of researchers' publishing activity, the proposed measure based on the disciplinary profile is stable.

Research limitations: The sample of both selected teams and journals database is limited. The journals from WoS/Scopus list were considered because analyzed researchers publish articles there. Additionally, during linking these two databases, problems of matching journals titles appeared.

Practical implications: The authors proposed a model of evaluating scientists' disciplinary similarity and further, to estimate the potential of their collaboration.

Originality/Value: This approach applies fuzzy logic algorithms to quantifying scientific interests and is another rare instance of practical application of artificial intelligence algorithms (fuzzy logic) in scientometric studies.

Keywords

Aggregation norm. Authorship. Disciplinary profiles. Fuzzy logic. Scientometrics. t-SNE plot.

Received: 25 August 2019. Reviewed: 28 October 2019. Accepted: 30 November 2019.

1. Introduction

In scientific communication, the disciplinary orientation of an individual researcher carries essential information about them, their groups, cooperation areas, fields in which they conduct their research. One can say that it constitutes the principal identifier of individuals in terms of specialization, but it can also indicate a shared research framework within a particular scientific community. If fields of research were to be represented graphically, such a representation would show where they activities overlap, and where they lie far away from each other, which could be used for organizing and designing the multidisciplinary teams. This insight would be especially valuable in the context of contemporary science policy in Poland, where the Ministry of Science emphasizes conducting joint research in all scientific domains, team fusion and multi-authorship publishing.

From the researcher's disciplinary profile, their publications output narrowed to the journal data set, and their association to disciplines might be inferred. Scientific journals are usually assigned a specific database, such as Web of Science (WoS) or Scopus collections by the professionals such as editors, domain classifiers, librarians, information specialists, and so on (Leydesdorff et al., 2019). A table of journal titles, mapping their disciplinary allegiance according to last Science Classification (MNSW, 2018) changes, introduced by the Ministry of Science and Higher Education has been recently published online.

The classification of sciences in Poland consists of scientific disciplines (previously) and eight branches of science, resembling WoS research areas: medicine and health sciences, technical sciences and engineering, exact and natural sciences, humanities, social sciences, theological sciences, agricultural sciences and arts.

While creating scientific profiles of researchers, bibliometrics should take into consideration that some scholar databases are oriented towards a particular scientific domain (Kulczycki & Rozkosz, 2017). For example, WoS covers mostly natural sciences and engineering (Kokowski, 2015); Scopus also specializes in hard sciences, although to a lesser degree, and PubMed database is dedicated to medical sciences.

2. The objectives

By using mapping algorithms, a visual representation of researcher network can be constructed. Ties will represent collaboration between people in co-authorship relationships (Fields, 2015; Small & Garfield, 1986). Based on the citations database, it is possible to see which research fields are close, thanks to techniques such as direct citations, co-citations, or bibliographic coupling (see Boyack et al., 2005; Huang, 2015; Jarneving, 2005; Shibata et al., 2009). In the bibliometric study, it is a common practice to use the linear counting of records after aggregation processes according to selected units, for example, author, keyword, discipline, or topic. However, the calculation of impact weights can be based on fuzzy logic algorithms, rather than on direct operations. Scientometricians turn to fuzzy logic rarely, which is confirmed by only nine items retrieved from WoS (the result of a query “fuzzy logic” AND “scientometric*”). Most papers refer to bibliometric study on fuzzy sets and algorithms used in other scientific domains. The query also yielded analyses of citation networks presenting the ties between information sciences and fuzzy systems (Merigóa et al., 2108; Yu & Shi, 2015).

The main difficulty which the authors met in the context of citations study is that there is no unified citation database concerning Polish scientists; however the authors proposed to analyze the similarity between the researchers established by the consideration of the characteristics of journals where they publish. Thus, taking into account both co-authorships and disciplinary embedding should compose a comprehensive view of the similarity and collaboration between the members of the scientific community. Final evaluation of these characteristics was made using visual maps generated by t-SNE algorithm.

Two sets of scientists were taken into consideration, the first team from the Department of Informatics and the second one from the Interdisciplinary Centre for Modern Interdisciplinary Technologies at the Nicolaus Copernicus University in Torun (NCU), Poland. These teams were chosen because two of the authors know the staff of these groups of scientists and can interpret the achieved results. Thus, the authors can include their experts' assignment of the spatial arrangement of data on the visualization thanks to authors' acquaintance with the members. Moreover, the first team is rather homogeneous (it is called GROUP_H) because the members represent one knowledge domain – computer science with the relation to technology and engineering domain. The other team (GROUP_M) is more multidisciplinary, gathering researchers whose research encompasses natural sciences (chemistry, physics, biology), social sciences (psychology, cognitive sciences) and health sciences (neurology, physiotherapy).

Mapping results are represented in t-SNE space and co-authorships graphs, which are compared and interpreted by experts.

Research questions:

- How do two approaches, co-authorship and disciplinary similarity, characterize researchers, and do they complement each other?
- Do co-authorship and disciplinary similarity characterize researchers, and do they complement each other?
- Does disciplinary similarity studies foster collaboration potential?
- Is fuzzy logic appropriate for the representation of relationships between disciplines and journals, and disciplines and authors, in the multidisciplinary world of science?
- Is fuzzy logic more fit to calculate the contribution to the researchers' disciplines (numbers from the interval) which constitutes their scientific profiles, than the linear sum of published papers?

Hypothesis: In the disciplinary representation of scientists, co-disciplinary relationships based on fuzzy logic should be compared with the classically calculated co-authorships. These combinations give the best results of a real disciplinary profile of the selected group; the expert's assignment can confirm that.

3. Nonlinear aggregation of scientific contribution based on fuzzy logic

The paper aims to find a method to calculate the scientists' contribution to discipline/science. The basis of calculating scientific achievements is the number of articles published by researchers considered. Next, aggregation functions are applied, of which one is the total number of papers. However, this method has weaknesses; the result belongs to the interval $[0, +\infty)$, so it is difficult to compare the contribution to science between two

different scientific units (with a varied number of their members). The authors propose the nonlinear function of aggregation, whose result belongs to the interval $[0,1]$, makes it much easier to compare the contribution to science of a few different scientific units.

Fuzzy logic is one of the most efficient tools for describing uncertainty characterizing artificial intelligence. The research area analysis, the characterization of the researchers in terms of their contribution to the discipline, level of collaboration, the similarity of discipline, etc. can be more efficiently described based on fuzzy logic. All of these concepts are fuzzy for two reasons.

The first reason is the lack of objective measuring instruments. It is difficult to find a unit of measurement of the scientist's contribution to the discipline. Of course, it may be measured by a number of articles in a journal; however, a linear dependence of the sum of papers may lead to the need for constant scaling during the comparison, for instance, of two scientists. The limitation of the linear calculation of contribution values became more significant in the case of more than two scientists with the high diversity of their output. If we have teams of researchers publishing very differently, then normalization needs to be updated with every new article. Linear normalization by the maximum number of articles leads to a huge divergence of contribution value, which does not coincide with reality. The authors propose a fuzzy value for describing a contribution unit for the scientist.

The second reason is the subjective nature of discipline/disciplines assigned to a scientific journal, which is established by experts/editorial boards. Instead of a binary relation, it is more natural to use fuzzy one, which describes the degree of such a relation more realistically or accurately. Moreover, in the current version of the scientific journals' classification, measures assigned to the journals, for example, Impact Factor, are defined. They also need to be taken into account. The following example will demonstrate this approach.

Let the fuzzy value A be the contribution of one scientist to one discipline after publishing one paper. Let D, J, S denote spaces of disciplines (as said above, according to the new classification of scientific domains and disciplines used by Polish Ministry of Science and Higher Education (MNSW, 2018), there currently are disciplines in Poland), journals, and scientists, respectively. Moreover, let

$$D = \{D_i, i = 1, 2, \dots, I\}, J = \{J_m, m = 1, 2, \dots, M\}, S = \{S_n, n = 1, 2, \dots, N\}.$$

Let $R \subseteq J \times D$ be a fuzzy relation between journals and disciplines for one scientist, where $J \times D$ denotes the Cartesian product of spaces J and D . Then $R(J_m, D_i)$ is equal to A if discipline D_i is assigned to journal J_m for each i and m and it is equal to 0 otherwise. Hence, if according to the Classification Table journal J_1 is assigned only to discipline D_3 , then $R(J_1, D_1) = R(J_1, D_2) = 0$ and $R(J_1, D_3) = A$. If the next paper is published in journal J_2 which is assigned to disciplines D_2 and D_3 , then $R(J_2, D_1) = 0$ and $R(J_2, D_2) = R(J_2, D_3) = A$. Table 1 presents the total achievement of this scientist when the both journals are taken into account.

Tab. 1. Values of relation for the discussed scientist

Journals \ Disciplines	D_1	D_2	D_3
J_1	0	0	A
J_2	0	A	A

To calculate the contribution $R(D_i)$, where $i = 1,2,3$, of this scientist to disciplines D_i , the following function is applied:

$$R(D_i) = R(D_i, J_1) + R(D_i, J_2) - R(D_i, J_1) \cdot R(D_i, J_2). \quad (1)$$

Hence,

$$R(D_1) = R(D_1, J_1) + R(D_1, J_2) - R(D_1, J_1) \cdot R(D_1, J_2) = 0 + 0 - 0 \cdot 0 = 0,$$

$$R(D_2) = R(D_2, J_1) + R(D_2, J_2) - R(D_2, J_1) \cdot R(D_2, J_2) = 0 + A - 0 \cdot A = A,$$

$$R(D_3) = R(D_3, J_1) + R(D_3, J_2) - R(D_3, J_1) \cdot R(D_3, J_2) = A + A - A \cdot A = 2A - A^2.$$

This formula is a special case of an optimistic fuzzy aggregation norm (Sokolov et al., 2018).

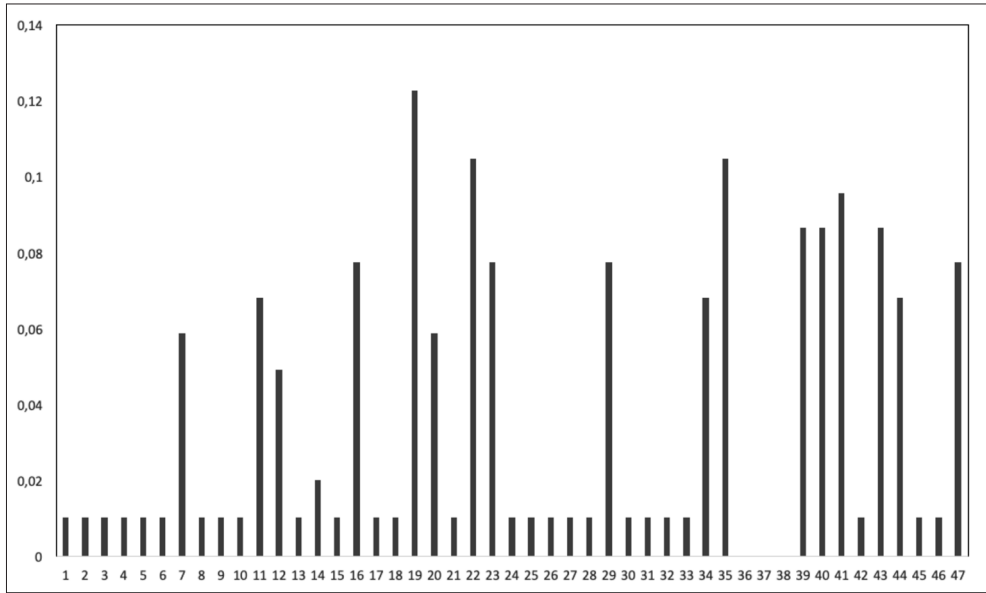


Fig. 1. The graphs of scientific contribution to disciplines of one scientist

Thus, the graph above (Fig. 1) shows the scientific contribution to disciplines of one exemplary scientist. The x-axis presents the numbered disciplines, and the y-axis shows the scientific contribution to these disciplines of this scientist calculated by formula (1). Moreover, by the application of the fuzzy optimistic aggregation norm, the total contribution to science by one scientist can be calculated, and for this scientist, it is equal to 0.818873.

4. Data and methods

4.1. Data

As authors found in previous research (Osińska et al., 2018), the bibliographic database Expertus (n.d.) is the most complementary and updated source of scientific output on the local level, i.e., in a selected university. This Web platform is included in the science-policy

of NCU, whose authorities force the staff to update their bibliographic records of published papers indicating that it is the primary condition to get promoted. Expertus platform has the same interface in most universities in Poland; it allows the user to filter records by researchers' names and organizational units. Bibliographical metadata in both basic and extended forms of selected records can be downloaded in text format. An article's metadata such as author(s), title, source, and publication year have been processed for further analysis.

Following the expectations of authorities to publish in selected domains, the members of these two organizational units (Department of Informatics and CMIT), publish journal articles rather than books and book chapters because the primary Web filter was predefined to set on formal publication type to be equal "002" (journal articles) and the authors assumed that Web of Science and Scopus databases fully cover downloaded data. The current conditions do not require to apply the set Polish journals, but the authors take into account the need for extending their approach to the broader scope of analysis units in the future. Then the database such as Arianta (<http://ariant.pl>) a continuously updated Polish journals reference list, will be useful (Kulczycki & Rozkosz, 2017).

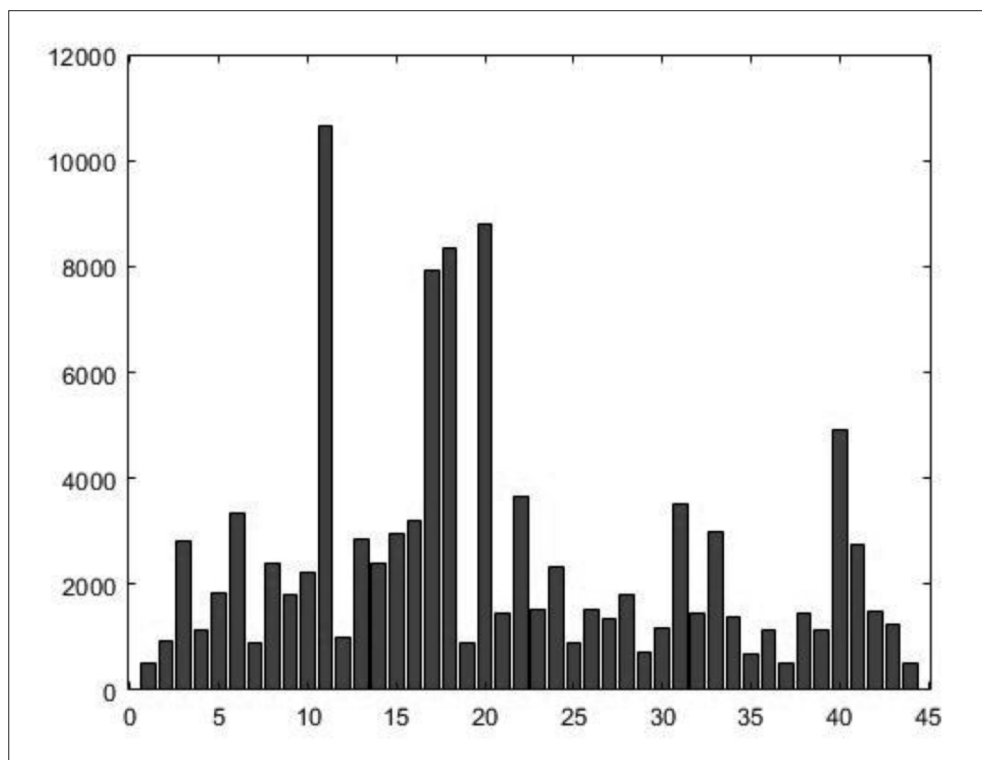


Fig. 2. The histogram of journal disciplinary distribution in the dataset of journals title list (N=29037) in both WoS and Scopus databases in 2019 (x-axis denotes the numbered disciplines and y-axis denotes the frequency of journals connected with these disciplines)

The ministerial table of disciplinary appurtenance of analyzed journals (Komunikat MNSW, 2019) was used to construct the mapping space which would account for all the researchers who belong to selected units. It consists of a list of journals ($N=29037$) from both WoS and Scopus databases as well as of the sequence of assigned disciplines, in some cases amounting to as many as 20. The disciplinary distribution of journals dataset is presented in Figure 2, where the x-axis denotes the numbered disciplines, and y-axis – the frequency of journals connected with these disciplines. As it can be seen, the majority of journals are described by one, two, or three categories, but often multidisciplinary titles can be found in the right tail of the distribution. Outstanding discipline in Figure 2 numbered as 11, describing more than 10,000 journals, is biomedical engineering.

The Expertus database is constructed with coding external and internal co-authors (i.e. University employees). This information allows filtering the data about collaboration within the University. Finally, by selecting certain units, the co-authorship records are obtained, and the symmetric matrixes are constructed. The rows and columns consist of all members of analyzed Departments, the cells – of the numbers of their joint publications. The Departments GROUP_H and GROUP_M numbered 18 and 16 members, respectively. This representation of data follows mapping by graph layouts algorithms.

4.2. Methods

Figure 3 presents the stages of the research. The data regarding the particular researcher's output was collected, taking into account journals where the articles were published. Journals' datasets were matched with the WoS/Scopus classification table.

In Table 2, the rows represent scientists and columns – their extracted disciplines, as gathered from the journals' list. The table was filled out with weights calculated by using the optimistic fuzzy aggregation norm. The data from the table was visualized in the t-SNE space representing the disciplinary similarity between scientists.

Tab. 2. A part of the fuzzy relation between scientists and disciplines

Disciplines Scientist	Arche- ology	Architecture and urban studies	Astro- nomy	Automatics, electronics and electrotechnics	Economy and Finances	Philoso- phy
S1	0	0	0	0	0.01	0
S2	0	0	0	0	0	0
S3	0	0	0	0	0	0
S4	0	0	0	0	0	0
S5	0	0.0199	0.01	0.05852	0	0
S6	0	0	0	0	0	0
S7	0.01	0	0.095618	0.05852	0.039404	0.029701

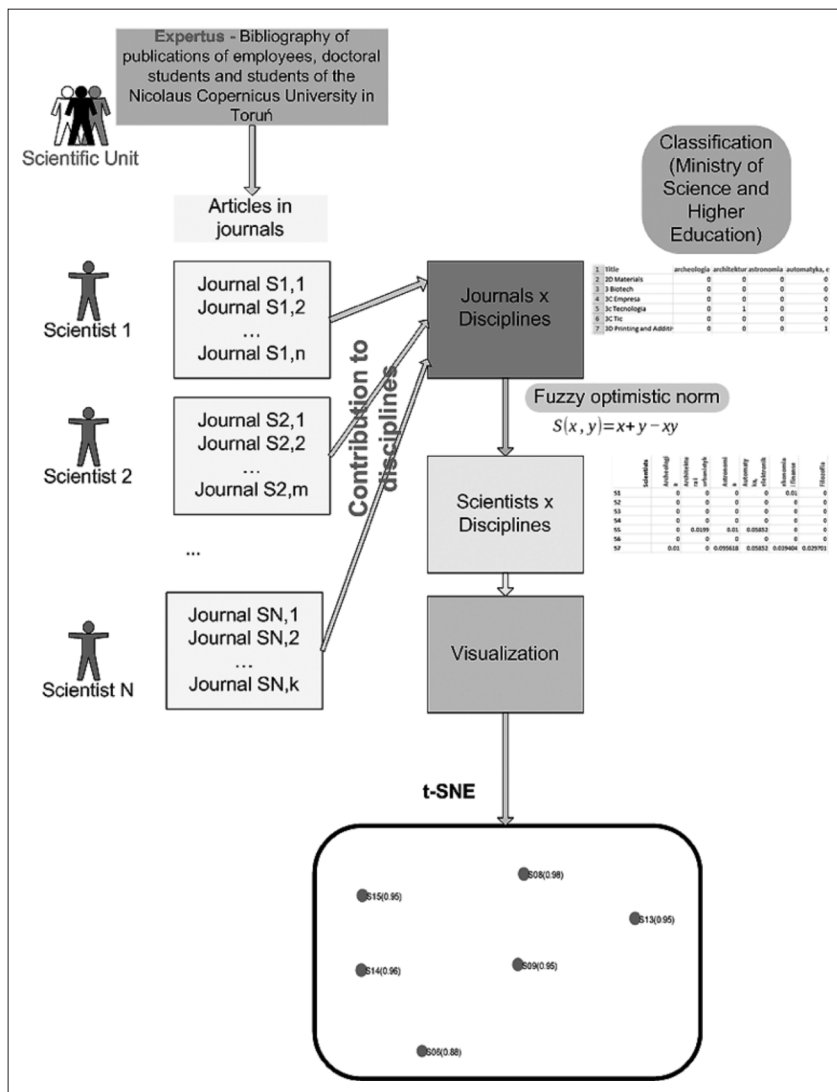


Fig. 3. The stages of the research

4.3. Optimistic fuzzy aggregation norm

Being a fuzzy variable, A must be between 0 and 1. The function S must increase with the scientist's contribution to the disciplines by adding the following article, so the authors used an optimistic fuzzy aggregation norm (Sokolov et al., 2018). One of examples of the function S is the following function (Fig. 4), for $x, y \in [0, 1]$.

$$S(x, y) = x + y - xy \quad (2)$$

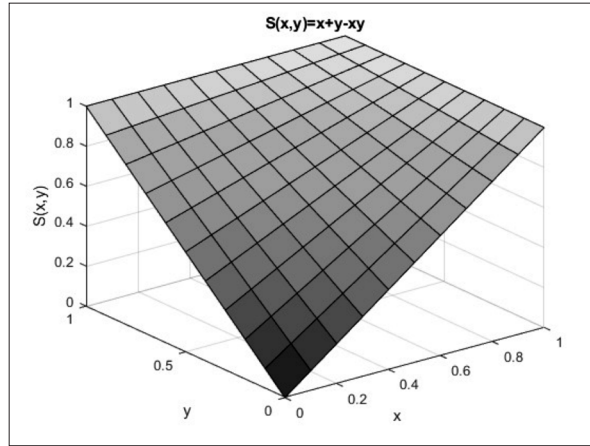


Fig. 4. 3D graph of function

To demonstrate the method, let the abovementioned value A be equal to $A = 0.01$. If we use the optimistic fuzzy aggregation norm as a function of scientific contribution, then $R(D_3) = R(D_3, J_1) + R(D_3, J_2) - R(D_3, J_1) \cdot R(D_3, J_2) = A + A - A \cdot A = S(A, A) = 0.0199$. In such a way, the relation between scientists and disciplines is established.

The value of the variable A is chosen so that the estimated scientist's contribution to the disciplines representative of the research team would be as diverse as possible. It is set experimentally. If it is observed that the total scientist's contribution to the disciplines is approximately the same and close to 1 (maximum value), the value of variable A should be reduced.

4.4. T-SNE algorithm

T-distributed Stochastic Neighbor Embedding (t-SNE) is a technique based on machine learning developed by Laurens van der Maaten and Geoffrey Hinton (2008). This technique is non-linear and well suited for embedding multidimensional data for visualization in a 2 or 3-dimensional space. This method uses a t-distribution rather than a Gaussian distribution to compute the similarity between two points in the low-dimensional space (Maaten & Hinton, 2008). Explicitly, it models each high-dimensional object with a two- or three-dimensional point in such a way that there is a high probability of similar objects being modeled by nearby points, and non-similar objects are modeled by distant points.

The advantage of t-SNE visualization relies on reducing the tendency to crowd points together in the center of a map. It better reveals structure on many different scales. This algorithm has a wide range of applications, including computer security research, music analysis, cancer research, bioinformatics, and biomedical signal processing.

4.5. Expert assessment

The two authors are the members of considered scientific units, and they are knowledgeable about these teams, which impacts the internal relations between researchers: for example, "who collaborates with whom?". In this case, they are the experts because they have more

knowledge than others about the research interests of a particular scientist. It is worth mentioning that the considered teams have got only a small number of members (18 and 16). Hence, the authors-experts can estimate the results of visualization taking into account the context of relationships between the scientists.

The number of co-authorship links is called *a degree* in network analysis (Barabasi, 2010). The larger the degree, the closer the connection (and/or the disciplinary similarity) between the authors is.

However if we focus on the disciplinary similarity, we will face a serious problem. Portals such as Research Gate or Academia.edu use specialized algorithms very effectively to search for researchers with similar interests. In the case explored in this article, the authors played the role of experts and evaluated the results of visual representation of contribution to science and co-authorship of these scientists.

4.6. Tools

For bibliographic data cleansing, processing and extraction R scripts were used. In the Matlab environment, the weights for the scientists-disciplines table were calculated. The t-SNE technique helps to visualize the relations between scientists in 2D space. Co-authorship graph layout was provided by Gephi platform (<https://gephi.org>) with embedded appearance tools. Data and results were communicated and discussed between the authors in the Microsoft Excel environment.

5. Visualization results

Below are the visualization layouts made by both algorithms: t-SNE and force-directed graph for two analyzed groups. The t-SNE maps illustrate disciplinary relationships, the graphs – collaboration in publishing. According to the mapping principle, the distances between data points reflect the similarity between the researchers in the terms of disciplinary activity. The links on the co-authorship graph show the ties between the co-authors. Accordingly, the node size depends on the number of the articles, so the larger the data point, the more significant scientific the output of the considered scientist is.

5.1. The co-authorship graph

Figure 5 presents co-authorship networks within both groups. The researchers' names are coded by letters: "S" – for the employees within the department, and "O" – for the external scholars. The links between nodes indicate the existence of joint publications. As the amount becomes higher, the edges thicken. The size of the node is proportional to the number of the papers written by the given researcher. Thanks to linked components, we can track the subgroups of collaborators and the people individually conducting research (separate, regularly arranged nodes in Fig. 5). As can be expected, there are more collaboration units in the multidisciplinary group GROUP_M.

By the application of fuzzy optimistic aggregation norm S, the total contribution to science of these scientists was calculated and is presented in the graphs (Fig. 6) by the size

of circles representing scientists. A large size of a circle (the high contribution to science) can be achieved in two cases:

- (1) The scientist is devoted to one discipline and achieved excellent results publishing in many journals assigned to this discipline.
- (2) The scientist's output is interdisciplinary and he/she has published papers in many journals assigned to many disciplines.

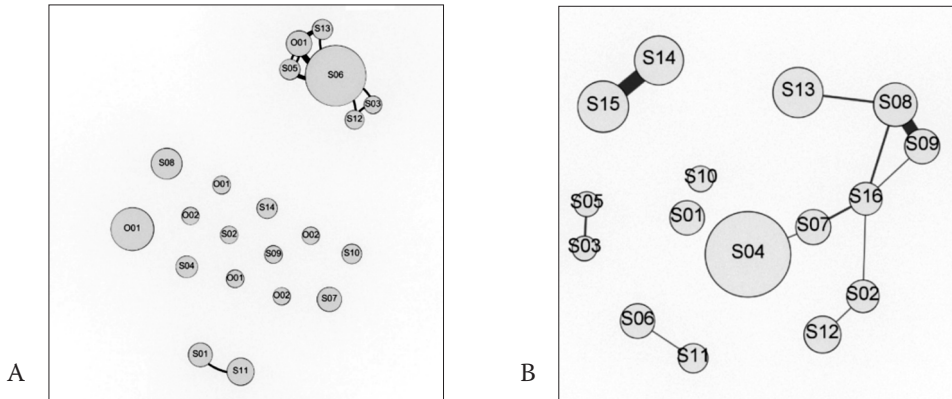


Fig. 5. Co-authorships graph layout for GROUP_H (A) and GROUP_M (B)

The more papers published, the higher the value of contribution to science. To reduce the 47-dimensional space to 2-dimensional one, a modern t-SNE technique was applied. Thanks to such visualization the relationships between points (scientists) can be easily considered and discussed.

5.2. *T-SNE space*

The disciplinary similarity of the researchers within the groups are presented by t-SNE space in Figure 6. The smaller the distance between the nodes, the more significant the disciplinary similarity between the respective scientists. For each scientist belonging to one of these two teams, with the application of the S norm, contributions to each discipline and finally to science were calculated, and the result is additionally presented in brackets beside the node.

6. Discussion

The classical approach used in bibliometrics is based on the co-authorships analysis. If we refer to the t-SNE maps, the name-coding reveals regularities in both arrangements. Close relationships on the co-authorship map (Fig. 5) mean that there are publications available by the pair of researchers, identified by the linkage. What we can observe in Figure 6 is that the scientists represented by the large circles play a central role in co-authorship and are attractors (see Buskell, 2017) to other scientists. So, they attract other researchers to collaborate with them. On the other hand, the close research interests do not always cause collaboration in

publishing. The co-authorship graph shows no permanent collaboration between researchers, but only intermittent connections, which assist with selected study themes and directions. Nor does it manifest the scientific interests of researchers and their relationships.

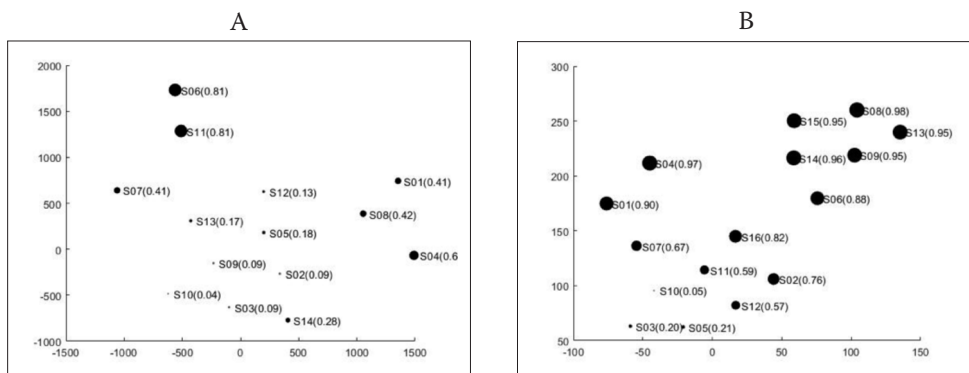


Fig. 6. The disciplinary similarity of scientists in t-SNE space for GROUP_H (A) and GROUP_M (B)

Quantitative comparison of two spatial arrangements of the sets of scientists is presented in Table 3 and Table 4. For $n=18$ nodes, it is theoretically possible to extract $n(n-1)/2 = 153$ links. A close relationship here means direct neighboring between the nodes, a distant relationship – all other pairs. Table 3 shows an extreme case: researchers who publish joint articles but are affiliated with different disciplinary and vice versa – represented by the most significant numbers (6, 6), which are in diagonal cells. In the case of GROUP_H, which is a homogenous computer science team, all scientists work in the same discipline but in different specializations, and they do not cooperate.

Tab. 3. Proximity matrix of two visual representations for GROUP_H

	Close relationships on t-SNE map	Far relationships on t-SNE map
Linkage on co-authorship map	3	6
No linkage on co-authorship map	6	4

Tab. 4. Proximity matrix of two visual representations for GROUP_M

	Close relationships on t-SNE map	Far relationships on t-SNE map
Linkage on co-authorship map	7	3
No linkage on co-authorship map	2	11

It is noticeable that they do not create teams which involve members of different specializations. In the case of GROUP_M, two approaches became complementary because they show the similar relationships: a linkage reflects the similarity in disciplines and vice versa, what can be observed in Table 4 when right diagonal numbers dominate (7, 11).

The co-authorship data reflects the real state in particular conditions, depending on the collaboration period. For example, the data can indicate a common study within the grants, at a particular stage of the researcher's career, affected by individual predispositions and other factors. The disciplinary profile of a scientist can give valuable information about the researcher's interests, their history, and development, that qualifies as a stable and independent parameter describing their scientific output.

The authors learned about the t-SNE algorithms and experienced within interpreting the visualizations. Relationships between data can be found on the periphery of the visual layout; non-clustered data remains in the center in the form of a regular grid. They can be interpreted as data with no or loose relationships.

The applied optimistic fuzzy aggregation norm, defined by the formula (2), is not linear, and the S norm does not imply linearity. Moreover, to make direct conclusions about complementing both approaches in a study of collective profiles, we need to operate on a big data set. In the future, research will be extended to include the data from several departments.

7. Conclusion

Authors proposed a combined approach to evaluate the researchers' disciplinary profiles and represent their similarity according to their interests. This approach is a combination of two unique relationships disciplinary profiles and co-authorships of the researchers within one team.

The disciplinary similarity of researchers allows predicting their future cooperation, even in the case of scientists who have not written a joint article yet. Thus, the chance for their collaboration can be estimated. If the disciplinary similarity is high, it can be expected that the scientists will write a joint article in the future. Moreover, it encourages them to do research together.

The cooperation in an interdisciplinary group is distributed according to the disciplinary interests of members; however, in a homogeneous group, there are no observed direct dependencies in collaborators selection. This conclusion from the pilot study is essential for future planning research.

To sum up the authors conclude that co-authorship and disciplinary similarity characterize researchers in two different ways. The disciplinary similarity profile can show the collaboration potential because scientists with a high level of disciplinary similarity attract young scientist or are a center of a group of researchers. Fuzzy logic is developed to deal with different kinds of uncertainty and estimation in social sciences, so fuzzy relations, such as relations between disciplines and journals, or between disciplines and authors, are the best way to represent such subjective coupling in the multidisciplinary world of science. The application of fuzzy logic to calculate the contribution to discipline/science or discipline similarity is possible because all discussed values belong to the interval $[0,1]$. Hence, they are easy to compare and there is no need for continuous scaling. In this context, the assumed hypothesis, i.e., that two combinations of relationships give the best results of a real disciplinary profile of the selected group, was confirmed.

Fuzzy logic used for the design of intelligent systems, which is an area entering to information sciences (Babik, 2016), can be a natural way to find levels of connection between two

spaces in scientometrics such as disciplines, journals and authors. In this paper, not all problems and research questions are fully developed, mainly because of the limited data sample, but this way of solving these problems seem to be appropriate, and the authors are going to perform a series of further research activities on the disciplinary similarity of researchers.

The proposed approach favors scientists having significant output. For young researchers, the results can be misleading because they have published few articles so far, so perhaps they will derive a high disciplinary profile of co-authorship. However, thanks to the presented data, it is possible to predict their interests in the future.

- (1) In the case of poor data about co-authorships, the mapping does not reflect the real state of scientific interest. In this case, the authors recommend using the disciplinary profile.
- (2) Robust disciplinary dissemination (Ministerial WoS categorization) makes it difficult to describe the primary domain, and therefore, the authors' analysis can be based only on disciplines.
- (3) One can be tempted to say that Ministerial WoS/Scopus categorization of journals is not appropriate for study researchers disciplinary profiles.

References

- Babik, W. ed. (2016). *Nauka o informacji*. Warszawa: Wydaw. SBP.
- Barabasi, A. (2010). Network Science [online] [10.08.2019], <http://networksciencebook.com/>
- Boyack, K.W., Klavans, R., Börner, K. (2005). Mapping the Backbone of Science. *Scientometrics*, 64(3), 351–374, <http://dx.doi.org/10.1007/s11192-005-0255-6>
- Buskell, A. (2017). Cultural Attractor Theory and Explanation. *Philosophy, Theory and Practice in Biology*, 9 (13), <http://dx.doi.org/10.3998/ptb.6959004.0009.013>
- Expertus (n.d.). Expertus – Bibliografia publikacji pracowników, doktorantów i studentów Uniwersytetu Mikołaja Kopernika w Toruniu [online]. Biblioteka Uniwersytecka w Toruniu [12.08.2019], <http://bg.cm.umk.pl/splendor/umk/>
- Fields, C. (2015). Co-authorship Proximity of A. M. Turing Awarded John von Neumann Medal Winners to the Disciplinary Boundaries of Computer Science. *Scientometrics*, 104(3), 809–825, <https://doi.org/10.1007/s11192-015-1575-9>
- Huang, M-H.; Chang, C-P. (2015). A Comparative Study on Detecting Research Fronts in the Organic Light-Emitting Diode (OLED) Field Using Bibliographic Coupling and Co-Citation. *Scientometrics*, 102(3), 2041–2057, <https://doi.org/10.1007/s11192-014-1494-1>
- Jarnevig, B. (2005), A Comparison of Two Bibliometric Methods for Mapping of the Research Front. *Scientometrics*, vol. 65(2), 245–263, <https://doi.org/10.1007/s11192-005-0270-7>
- Kokowski, M. (2015). Science of Science (naukoznawstwo) in Poland. The Changing Theoretical Perspectives and Political Contexts – A Historical Sketch from the 1910s to 1993. *Organon* [online], (47), 147–237 [23.10.2019], <http://www.cyfronet.krakow.pl/~n1kokows/Kokowski-Organon-2015.pdf>
- Komunikat MNSW (2019). *Komunikat Ministra Nauki i Szkolnictwa Wyższego z dnia 31 lipca 2019 r. w sprawie wykazu czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych wraz z przypisaną liczbą punktów* [online]. Ministerstwo Nauki i Szkolnictwa Wyższego [23.10.2019], <http://www.bip.nauka.gov.pl/akty-prawne-mnisw/komunikat-ministra-nauki-i-szkolnictwa-wyzszego-z-dnia-31-lipca-2019-r-w-sprawie-wykazu-czasopism-naukowych-i-recenzowanych-materialow-z-konferencji-miedzynarodowych-wraz-z-przypisana-liczba-punktow.html>
- Kulczycki, E., Rozkosz, E. (2017). Publication Patterns in the Social Sciences and Humanities in Flanders and Poland. In: *Proceedings of ISSI 2017 Wuhan: 16th International Society of Scientometrics and Informetrics Conference*. Wuhan, China, 16–20 October 2017, 95–104.

- Kulczycki, E. Stan praktyk publikacyjnych polskich czasopism naukowych w 2017 roku. [online] Report 2018 [10.08.2019], <http://excellence-project.zut.edu.pl/>
- Leydesdorff, L., Wagner, C., Bornmann, L. (2019). Diversity Measurement: Steps towards the Measurement of Interdisciplinarity? *Journal of Informetrics*, 13(3), 904–905, <https://doi.org/10.1016/j.joi.2019.03.016>
- Maaten L. van der, Hinton G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9, pp. 2579–2605.
- Merigó, J.M., Pedrycz, W., Weber, R., la Sotta, C. (2018). Fifty Years of Information Sciences: A Bibliometric Overview. *Information Sciences*, 432, 245–268, <https://doi.org/10.1016/j.ins.2017.11.054>
- MNSW (2018). Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 20 września 2018 r. w sprawie dziedzin nauki i dyscyplin naukowych oraz dyscyplin artystycznych [online]. DzU, poz. 1818, <http://prawo.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20180001818>
- Osińska, V., Kowalska, M., Osiński, Z. (2018). The Role of Visualization in the Shaping and Exploration of the Individual Information Space: Part 1. *Knowledge Organization*, 45 (7), 547–558.
- Shibata, N., Kajikawa, Y., Takeda, Y., Matsushimaet, K.(2009), Comparative Study on Methods of Detecting Research Fronts. *Journal of the American Society for Information Science and Technology*, 60(3), 571–580, <https://doi.org/10.1002/asi.20994>
- Small, H., Garfield, E. (1986) The geography of science: disciplinary and national mappings. *Journal of Information Science*, 11(4), 147–159, <https://doi.org/10.1177/016555158501100402>
- Sokolov, O., Osińska, W., Mreła, A., Duch, W. (2018). Modeling of Scientific Publications Disciplinary Collocation Based on Optimistic Fuzzy Aggregation Norms. In: *Advances in Intelligent Systems and Computing*. Vol. 853. *Information Systems Architecture and Technology: Proceedings of 39th International Conference on Information Systems Architecture and Technology – ISAT 2018*, 145–156 [online]. Springer Link [10.08.2019], https://link.springer.com/chapter/10.1007/978-3-319-99996-8_14
- Yu, D., Shi, Sh. (2015). Researching the Development of Atanassov Intuitionistic Fuzzy Set: Using a Citation Network Analysis. *Applied Soft Computing* 32, 189–198, <https://doi.org/10.1016/j.asoc.2015.03.027>

Nieliniowa ocena podobieństwa profili dyscyplinarnych naukowców. Studium przypadku

Abstrakt

Cel/Teza: Autorzy oceniają podobieństwo dyscyplinarne badaczy wybranych jednostek akademickich o różnej naukowej specjalizacji i przedstawiają model badania różnorodności dyscyplinarnej jednostek naukowych. Założeniem artykułu jest to, że wiedza o profilach dyscyplinarnych badaczy może być wykorzystana do stworzenia zespołów interdyscyplinarnych lub jednego zespołu dyscyplinarnego o ukierunkowanych specjalizacjach.

Koncepcja/Metody badań: Metoda oparta jest na wizualizacji oraz porównaniu przestrzeni dyscyplinarnej i współautorstwa. Do obliczenia wag dyscyplinarnych każdego rozważanego czasopisma w bazie danych zastosowano logikę rozmytą i optymistyczną rozmytą normę agregacji. Do wizualizacji zastosowano nowy algorytm redukcji wymiarów t-SNE. Osiągnięte wyniki zostały zweryfikowane przy użyciu wiedzy ekspertów.

Wyniki i wnioski: W ocenie współpracy naukowej relację współautorstwa można uzupełnić profilami dyscyplinarnymi badaczy wyznaczonymi za pomocą rozmytej normy agregacji. Dzięki ciągłości działalności wydawniczej badaczy proponowana miara oparta na profilu dyscyplinarnym jest stabilna.

Ograniczenia badań: Próba badawcza jest ograniczona ponieważ autorzy rozważają dwa zespoły naukowców i ich publikacje z dwóch baz danych. Wybrano czasopisma z listy WoS/Scopus, ponieważ analizowani badacze tam publikują artykuły. Ponadto, podczas łączenia tych dwóch baz danych pojawiły się problemy z dopasowaniem tytułów czasopism.

Zastosowanie praktyczne: Autorzy zaproponowali model oceny podobieństwa dyscyplinarnego naukowców, a następnie oszacowania potencjału ich współpracy.

Oryginalność/Wartość poznawcza: Podejście to stosuje algorytmy rozmytej logiki do kwantyfikacji zainteresowań naukowych i jest kolejnym rzadkim przypadkiem praktycznego zastosowania algorytmów sztucznej inteligencji (logika rozmyta) w badaniach informatologicznych w szerokim kontekście.

Słowa kluczowe

Logika rozmyta. Mapowanie t-SNE. Naukometria. Profil interdyscyplinarny. Współautorstwo.

VESLAVA OSIŃSKA is an Associate Professor at the Nicolaus Copernicus University in Torun. She has a physics background and holds a PhD in information sciences. She is specializing in the mapping of information and data from both professional databases and the Internet. She is lecturing information visualization and graphic design, information architecture, and database management. She is certificated by ECDL training. Veslava Osińska is collaborating with Places @ Spaces International Project and is its ambassador in Poland. She is a local coordinator of the Polish Chapter of ISKO and the member of Polish Information Processing Society. More at her blog: www.wizualizacjainformacji.pl.

Contact to the Author:

wieo@umk.pl

Institute of Information and Communication Research

Nicolaus Copernicus University in Torun

Władysława Bojarskiego 1

87-100 Torun, Poland

OLEKSANDR SOKOLOV, Professor, Dr of Science, PhD. He took his degree in Applied Mathematics at Kharkiv Aviation Institute, and earned his PhD in Control Theory at Institute of Machinery Building of Academy of Science of Ukraine, and earned his Dr of Science in Control Theory at National Aerospace University, Kharkiv. Currently he is a professor in the Nicolaus Copernicus University in Torun in the Department of Informatics, Faculty of Physics, Astronomy and Informatics. His research interests are: Fuzzy Logic, Decision Support Systems, Intelligent Control Systems, Test Theory, Multiagent systems, Simulation of biological systems.

Contact to the Author:

osokolov@fizyka.umk.pl

Faculty of Physics, Astronomy and Informatics

Nicolaus Copernicus University in Torun

Grudziadzka 5

87-100 Torun, Poland

ALEKSANDRA MREŁA, PhD, took MSc degree in Mathematics at the Higher Pedagogical School in Bydgoszcz (Poland), and earned PhD in Mathematics at the University of Łódź, Poland. Currently she is an Assistant Professor in the Institute of Informatics, Kazimierz Wielki University in Bydgoszcz. Her research interests are: Fuzzy Logic, Decision Support Systems, Test Theory, Mathematics.

Contact to the Author:

Institute of Informatics

Kazimierz Wielki University

Kopernika 1

85-074 Bydgoszcz, Poland