Bibliometric analysis by network models

Identifying trends in scientific literature

Fuad Aleskerov
HSE University, Institute of Control Sciences

Co-authors

Olga Khutorskaya, Institute of Control Sciences of Russian Academy of Sciences

Anna Stepochkina, HSE University

Ksenia Zinovyeva, HSE University

Vyacheslav Yakuba, Institute of Control Sciences, HSE University

Bibliometric analysis

One of the main tools that allows tracking global research activity is bibliometric analysis. Bibliometric analysis is a highly demanded and widely used method for measuring publication activity, monitoring the development of scientific directions, and the research activity of the scientific community. It allows to

- identify the most popular journals where studies on a particular topic are published,
- assess their impact and status within the scientific community,
- identify the most relevant research directions in a specific field,
- determine the countries and organizations leading in research on a particular topic,
- study cooperation between groups in the form of scientific publications, for example, a collaboration network between scientists and research institutions.

Citation network analysis

Network analysis is a powerful tool for analyzing complex interactions within research communities. We use a comprehensive approach to analyze publication activity based on network analysis methods. The objective of our work is to develop a methodology for studying a subject area using network models based on bibliometric data. The developed approach makes it possible to identify connections between research clusters, rank their significance, and track changes in the research directions of scientific groups.

Scientific publications network analysis methodology

To test the methodology we use scientific publications related to Parkinson’s disease (PD). The following tasks have been formulated

- analyze the dynamics of publications by year (for a given time period, namely 2015–2021) to determine how and in which directions research on Parkinson’s disease is evolving;

- investigate the most productive and influential authors in this field;

- examine the number of citations for each publication and identify the most influential and cited works in the field;

- identify the most popular journals where research on Parkinson’s disease is published;

- analyze the most frequently occurring keywords and topics in publications on Parkinson’s disease.

Parkinson’s disease

Parkinson's disease and Alzheimer's disease are two scourges of the modern world and the two most common neurodegenerative diseases. More than 10 million people worldwide suffer from Parkinson's disease. According to the Parkinson’s Foundation, about one million people in the United States are currently living with Parkinson's disease, and this number is expected to rise to 1.2 million by 2030. Over the past 25 years, the number of people with PD has doubled, and experts predict it will double again by 2040 due to population aging.

The economic consequences of the disease are catastrophic. The cost of the disease increases creating an economic burden for healthcare systems, society, and the patients themselves. Total cost estimates vary by country. The cost associated with Parkinson’s disease in Europe amounts to €13.9 billion per year. The highest direct costs are typically associated with inpatient treatment and nursing home care. Costs rise from €5,000 annually in the early stages of the disease to €17,000 or more in the late stages. Total costs in the United Kingdom are estimated at between £449 million and £3.3 billion annually, depending on the cost model used and the level of prevalence.

Parkinson’s disease

A study by the Lewin Group showed that the total cost of PD to individuals, families, and the U.S. government is $51.9 billion annually. Of this, $25.4 billion is attributed to direct medical costs and $26.5 billion to non-medical costs such as missed work, lost wages, early forced retirement, and family caregiver time. In 2017, Parkinson’s disease cost the United States approximately $52 billion per year, and it is projected that by 2037 this figure will increase to nearly $80 billion.

The significant socio-economic impact of Parkinson’s disease due to the rising global incidence has led to a sharp increase in scientific interest in the issue. The number of studies has grown, and consequently, the number of publications on the topic has also increased. A rapid and representative analysis of publications has become highly relevant for the scientific and medical community to obtain information about the most significant and influential works, research teams, and current trends in studies. Such analysis allows the scientific and medical community to track the emergence of new research directions, assess their potential, and determine the influence of an author or research team on the perceptions of colleagues.

Network analysis models

Various methods can be used to identify the most significant vertices in a network. The characteristic "centrality" allows determining the degree of importance of a graph node based on its position. Centrality in a graph is a vector that assigns a certain number (index) to each node of the graph, indicating its importance. By calculating centrality indices and ranking them, the most influential nodes in a graph can be identified and analyzed. There is a large number of different indices that can be calculated for a network depending on the task.

Data description

The Microsoft Academic scientific publication search system was used to access data – an open and publicly accessible search engine for scientific publications developed by Microsoft Research. The system indexes approximately 250 million publications, including scientific articles, books and individual chapters, patents, and conference proceedings.

To construct the citation network and analyze centrality, articles were selected that met the following query: the words "parkinson" and "disease" are present in the title/abstract for the years 2015–2021. This time frame was chosen to trace the main research interests and trends in their evolution over recent years.

All data were compiled into a single table. Abstracts and titles of articles were normalized, i.e., all words were reduced to their base form (the main immutable form of the word from which all its grammatical variants are derived, e.g., for verbs – this is the infinitive). Each article had more than 30 attributes.

Data description

The following 17 attributes were used in this study

- Id – unique publication identifier. Example: Id 2754967293
- Ti – normalized title. A title in which all words are written in lowercase letters. Example: "global regional and national burden of neurological disorders 1990–2016, a systematic analysis for the global burden of disease study 2016"
- W – unique normalized words in the title. Example: ['global', 'regional', 'national', 'burden', 'neurological', 'disorders', 'systematic', 'analysis']
- AW – unique normalized words in the abstract. Example: ['summary', 'background', 'comparable', 'data', 'global', 'country', 'specific']
- Y – year of publication. Example: 2017
- CC – citation count. Example: 1535.
- RId – list of IDs of the articles referenced by the publication. Example: [RId 2159011576, RId 2751884637, RId 2125065061]
- DN – title. Example: "Global, regional, and national burden of neurological disorders, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016"
- DOI – digital object identifier. Example: 10.1016/S1474-4422(18)30499-X

Data description

- IA – abstract in Inverted Index format. Example: {'summary': [1], 'background': [5], 'comparable': [10], 'data': [15, 20], 'global': [25], 'country': [30], 'specific': [35]}. In this example, the numbers indicate the positions of the respective words in the abstract, e.g., 'summary': [1] means that the word 'summary' occurs in the first position of the abstract, while 'data': [15, 20] means that the word 'data' appears in the 15th and 20th positions.
- AA.AfId – author's affiliation ID. Example: AA.AfId 39854758
- AA.AfN, AA.DAfN – name of the author's affiliation. Example: 'AfN': 'auckland university of technology', 'DAfN': 'Auckland University of Technology'
- AA.AuId – author ID. Example: AA.AuId 130753112
- AA.AuN, AA.DAuN – author's name. Example: 'AuN': 'valery l feigin', 'DAuN': 'Valery L Feigin'
- J.JId – journal ID. Example: J.JId 70053155
- J.JN – journal name. Example: “Lancet Neurology”
- D – publication date in the format YYYY-MM-DD. Example: 2017-11-01.

Data description

A total of 70,119 articles were downloaded. To improve the quality of network construction and subsequent analysis, additional processing was carried out. This stage included preliminary data analysis, processing of abstracts of the articles, removal of missing values and errors, graph preparation (converting the data into the "source vertex – target vertex" structure), and semantic analysis. Isolated vertices were not included in the network. For this stage, Python libraries were used: numpy, pandas, matplotlib. The next step involved graph construction, network analysis, and community detection.

For the correct construction of the citation graph of publications, the DOI of the works was used, since this field is present in the article descriptions and the identification numbers are indicated in the citation field. However, detailed data analysis showed that the DOI was not specified for all works – 10,681 publications lacked DOI information. Therefore, the number of publications used in graph construction decreased to 59,438. Of these, 7,315 had no abstracts, and 52,132 had them. Journal information was available for 45,940 articles.

Further analysis showed that a number of articles referenced publications that were released after the citing article, meaning an article from the past cited a future one. A total of 4,184 such edges were removed. For further analysis, only articles with DOI, annotation, and journal information were used to enable result interpretation.

Data description

After all filtering steps, the network consisted of

- 39,811 articles,
- 423 terms in the analysis of co-occurrence of terms,
- 27,551 authors in the author network,
- 3,029 affiliations,
- 3,292 journals in the journals citation network.

Formal model of network analysis

Let $G = (V, E)$ be a graph, where $V = (1, \ldots, n)$ is a non-empty set of vertices, $E \subseteq V \times V$ is a set of edges, $A = [a_{ij}]$ is the adjacency matrix ($a_{ij} = 1$ if there is an edge from vertex i to vertex j, otherwise $a_{ij} = 0$), $W = [w_{ij}]$ is the weight matrix (w_{ij} is the weight of the edge from vertex i to vertex j).

A graph is called directed if each edge has a direction, given by an ordered pair of vertices (i, j). The existence of an edge (i, j) indicates a connection from i to j.

A graph is called unweighted if $\forall i_1, i_2, j_1, j_2 \in V: (i_1, j_1), (i_2, j_2) \in E \Rightarrow w_{i_1 j_1} = w_{i_2 j_2}$ for all edges, i.e., all edge weights are equal. Otherwise, a graph is called weighted.

To calculate graph characteristics and centrality indices, it is necessary to represent the graph as an adjacency matrix $A = [a_{ij}]_{n \times n}$, where a_{ij} =1 if $(i, j) \in E$, and $a_{ij} = 0$ otherwise, and a weighted adjacency matrix $W = [w_{ij}]_{n \times n}$), where w_{ij} is the weight of the edge from i to j.

Formal model of network analysis

We used the following classical indices.

In the task of identifying the most influential nodes in a network, the in-degree centrality index is used, as it shows the number of links to a node.

Weighted in-degree centrality – the sum of the weights of incoming edges, i.e., the number of citations of the node if the graph is a citation network

$$In - degree(i) = \sum_{j=1}^{n} w_{ij}$$

where i – number of the considered node, w_{ij} – weight of edge (i, j), n – number of nodes in the graph.

Degree centrality does not take into account any other parameters of the node or the network as a whole. For example, a node with a small sum of incoming edges, but referenced by several popular nodes, should also have high centrality.

Formal model of network analysis

Eigenvector centrality can be evaluated by the following formula

$$x_i = \frac{1}{\lambda_1} \sum_{j}^{n} A_{ij}\, x_j$$

where λ_1 – the largest eigenvalue of the adjacency matrix A, x – eigenvector, i.e., in vector form, this formula is written as

$\lambda x = Ax$.

Eigenvector centrality determines the importance of a node by considering the importance of its neighbors. However, in some cases, it is also worth considering the number of outgoing edges of nodes. For example, if there is a node with high centrality that refers to many other nodes, their centrality should not increase as much as if the important node referred to only one node. In terms of a journal citation network, if there is a significant journal in which various articles are published that cite a large number of other journals, their centrality should not increase as much as if those articles cited only one journal. It is necessary to consider the fact that in a large and popular journal, there are always more references than in less significant journals.

Formal model of network analysis

PageRank centrality index, essentially, is a variation of eigenvector centrality, but the centrality value of the nodes neighboring the given node is divided by the number of their outgoing edges, i.e.

$$x_i = \alpha \sum_{j}^{n} A_{ij} \frac{x_j}{k_j^{out}} + \beta$$

Constants α and β are used to avoid zero index values.

Newly developed centrality indices

The described indices are classical and are used in most works on network analysis. A serious drawback of these centrality indices is the lack of consideration for node parameters and ways to account for group influence. To take into account these and other node parameters, new centrality indices were introduced. These are the BI (Bundle Index) – the group influence index, and the PI (Pivotal Index) – the pivotal node influence index.

A significant difference of the newly proposed centrality indices BI and PI lies in the fact that they take into account the parameters of the vertices, as well as the fact that the centrality of the vertices in the network is determined by the group connections of the given vertex with the others.

To describe these indices, certain definitions must be introduced.

Bundle centrality index (BI)

For each node, a quota q_i is defined. As a rule, the quota is calculated as a certain percentage of the sum of the weights of incoming edges. For each node, the value of the quota is defined as

$$q_i = q \cdot \sum_j w_{ji}, \text{ where } q \text{ is a share of the sum of weights}$$

i.e., the quota is a percentage of the sum of the weights entering the node. Each node has a different quota value, allowing its individual parameters to be taken into account.

A critical set S for the node i is defined as a set of nodes not exceeding size k entering node i such that

$$\sum_{j \epsilon S} w_{ji} \geq q_i.$$

The parameter k is the number of nodes that can simultaneously influence node i. In this work, k = 3 (this choice is due to computational complexity limitations).

Bundle centrality index (BI)

Thus, the Bundle Index (BI) can be defined as the number of critical sets for node i not exceeding size k.

Bundle index (BI) is a group influence index. For all critical sets $S, |S| \leq k$

$$BI_i(S) = \begin{cases} 1, if \sum_{j \in S} w_{ji} \geq q_i \\ 0, else \end{cases}$$

BI for a node equals the sum of index values over all its critical sets,

$$BI(i) = \sum_{S} BI_i(S)$$

i.e., $BI(i)$ equals the number of critical sets of node i of size no greater than k.

Pivotal centrality index (PI)

The second centrality index, PI (Pivotal Index), shows the number of pivotal nodes for each critical set. To introduce this index, it is necessary to define what constitutes a pivotal node. A node j_p is called a pivotal node for node i in a critical set S if

$$\sum_{j \in S} w_{ji} \geq q_i \text{ и } \sum_{j \in S \setminus \{j_p\}} w_{ji} < q_i \ ,$$

i.e. the total incoming weight is no less than the quota (S is critical), but upon removing the node j_p from the set, the sum falls below the quota (S ceases to be critical). For a specific critical set, the value of the $PI_i(S)$ index is equal to the number of pivotal nodes in it. The PI value for node i is the sum over all critical sets of the product of the size of the set $|S|$ and the $PI_i(S)$ value for this set.

$$PI(i) = \sum_{S} |S| \times PI_i(S)$$

Total influence index (TI)

The combined index, called the Total influence index (TI) is defined as follows

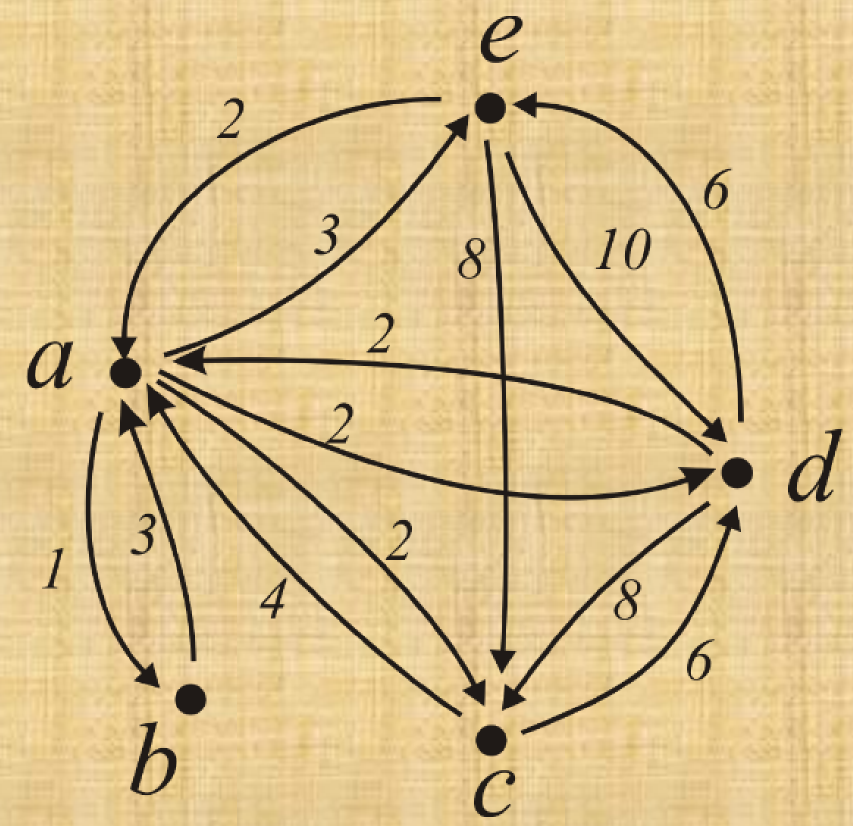
$$TI(i)= \alpha_1 IN(i)+\alpha_2 BI(i)+\alpha_3 PI(i)$$

Here α_i, i=1,2,3 are weighting coefficient, $\alpha_i \in [0;1]$, $\alpha_1+\alpha_2+\alpha_3=1$.

Normally, the equal weights are used, i.e. $\alpha_i=\frac{1}{3}$.

Example

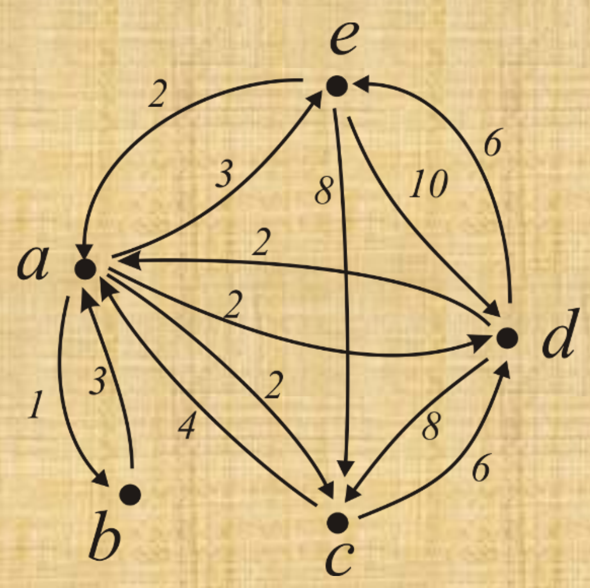
Let us consider the described indices using an example of a graph consisting of five nodes



Adjacency matrix for a graph consisting of five vertices

	a	b	c	d	e
a	0	1	2	2	3
b	3	0	0	0	0
c	4	0	0	6	0
d	2	0	8	0	6
e	2	0	8	10	0

Example



Let the quota be q = 1, k = 3. Then for node e we obtain the following critical sets: $\{a\}$, $\{d\}$, and $\{a, d\}$, with total weights 3, 6, and 9 respectively.

Now the BI index for these critical sets can be calculated. Since for all critical sets the sum of weights is greater than or equal to the quota q, BI for these sets equals 1, hence $BI(e)$ = 3. Similar calculations can be performed for the remaining nodes to obtain the index values.

To calculate the PI index, it is necessary to identify the pivotal nodes. For node e in the first two critical sets, nodes a и d are pivotal, since their exclusion causes the sum of incoming edge weights to fall below the quota. For the critical set $\{a, d\}$, there are no pivotal nodes, as removing any one node still leaves the sum above the quota. Thus, $PI(e)$ = 2.

The values of the In-degree, BI, PI, and TI indices for a graph

Vertices	In-degree	BI, q=1, k=3	PI, q=1, k=3	TI
a	0.19	0.44	0.31	0.27
b	0.02	0.03	0.08	0.06
c	0.32	0.22	0.23	0.26
d	0.32	0.22	0.23	0.26
e	0.16	0.09	0.15	0.15

Example

Let us consider the calculation of BI and PI indices for various quota values under the same restriction $k = 3$.

Quota $q = 1$

Vertices	BI, q=1	PI, q=1
a	0.44	0.31
b	0.03	0.08
c	0.22	0.23
d	0.22	0.23
e	0.09	0.15

Quota $q = 9$

Vertices	BI, q=9	PI, q=9
a	0.18	0.43
b	0	0
c	0.36	0.29
d	0.36	0.19
e	0.09	0.10

Quota $q = 10$

Vertices	BI, q=10	PI, q=10
a	0	0
b	0	0
c	0.5	0.6
d	0.5	0.4
e	0	0

For quota $q = 10$, according to the BI index, the most significant vertices are c and d, and the remaining vertices have the same rank.

When the quota is decreased to 9 ($q = 9$), vertices c and d have a high rank, but the significance of the other vertices has changed. The ranking now is as follows

$$c \backsim d \succ a \succ e \succ b,$$

When the quota is reduced to 1 (q =1), the ranking changes completely

$$a \succ c \sim d \succ e \succ b,$$

i.e., the most important vertex becomes a.

In terms of article citations at the journal level, this example can be interpreted as follows. For quota $q = 1$, i.e., if we simply consider the fact of citation of journals, then the most important journal turns out to be a, and the influence of a is twice as high as that of c and d. However when the quota is increasing, journals c and d, actively citing each other, turn out to be the most influential. In other words, the considered models allow us to identify groups of narrowly specialized journals that actively cite each other.

Network stability analysis

In networks considered dynamically, various changes may occur – new vertices may appear or old ones may disappear, and the interconnections between them may change. It is important to understand how much the network has changed over time to identify patterns and trends, and the stability of the scientific community. The simplest methods are the calculation of the correlation of two successive adjacency matrices or the ranks of vertices in such networks, which do not account for topological changes in the network. An approach is developed to measure graph stability, taking into account topological similarity and the similarity of the importance of elements. Two networks are considered similar if they have a similar structure (vertices exert the same influence on each other) and similar key elements.

Networks have similar key elements if their vertices have the same ranks. Vertices can be ranked using any centrality indices suitable for the task. Let C_i be the rank of vertex i, n – the number of vertices, then the interval order matrix $R^t = [r_{ij}^t]$ is introduced at time t

$$r_{ij}^t = \begin{cases} 1, if \ |c_i^t - c_j^t| > \varepsilon \\ 0, else \end{cases}$$

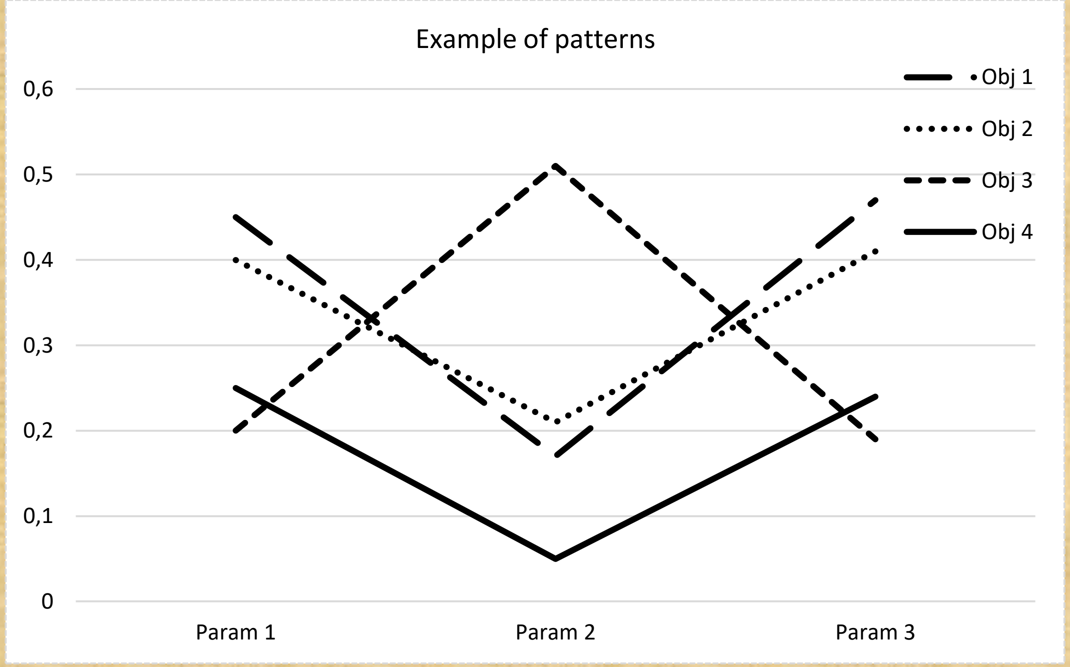
That is, its values are equal to 1 if the ranks of the vertices differ by more than ε. Then the distance between matrices at two successive points in time can be calculated (Wiener, 1914; Aleskerov et al., 2007)

$$d(R^t, R^{t+1}) = \frac{\sum_{i \neq j}^n |r_{ij}^t - r_{ij}^{t+1}|}{n \cdot (n-1)}$$

It is equal to zero in the case of matching ranks and greater than zero when the ranks differ.

Pattern analysis

Various clustering methods can be applied to bibliometric data to identify specific groups of objects. Pattern analysis is the analysis of objects grouped into clusters with the same pattern. A pattern is a sample or template; in this work, a pattern is understood as a set of objects characterized by a certain combination of indicator values of the objects. Groups of similar objects will be called clusters. Objects in one cluster are similar to each other and significantly differ from others. Visually, the pattern can be represented by depicting objects (vectors) in parallel coordinates: each vector value is located on its own vertical axis, and all values of one vector are connected by a line for four objects.



Pattern analysis

There are two ways to group into clusters: grouping identical objects with and without considering the scale. Identical objects are defined by falling into the same neighborhood, called the ε-tube. For the two methods, it is constructed differently:
1. The ε -tube is built on normalized data. Let $v \in \mathbb{R}^k$ be a data vector. In this case, objects that are close in values of vector elements fall into one pattern.
2. The ε -tube is built on the tangents of the angles of polyline slopes. Let $v \in \mathbb{R}^{k-1}$ be a vector of data slope angles. Clustering by tangents combines into one pattern objects whose graphs have a similar shape (points connected by segments in the parallel coordinates system can be represented as a graph of a piecewise linear function).

For each vector v, its ε-tube is calculated, for all v_i the maximum deviation $(d_{max})_i$ is calculated, which is of two types:

- Non-adaptive, does not depend on the value of the parameter $(d_{max})_i = 1$.
- Adaptive, calculated based on the parameter $(d_{max})_i = v_i$.

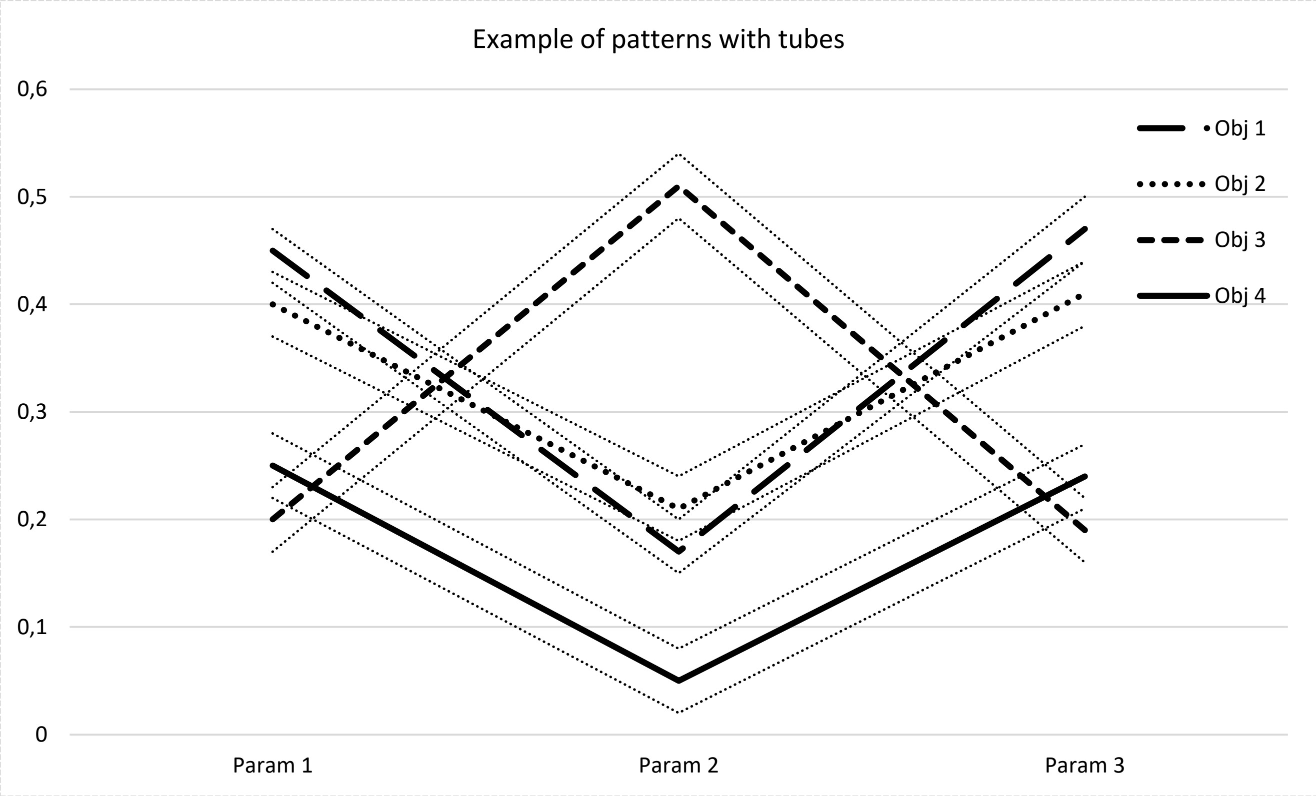
After that, all values are multiplied by ε: $d = \varepsilon \cdot d_{max}$. Then, if the values $N(v)$ are in the same cluster with vector v, then

$$u \in N(v) \Leftrightarrow \forall\, i: |u_i - v_i| < d_i.$$

That is, if for each coordinate with number i of vector u the condition is met that the value of this coordinate differs from the value of the corresponding coordinate of vector v by no more than d_i, then vector u falls into the cluster formed by vector v.

Pattern analysis

Thus, in the considered example, for each object we introduce an ε-tube, as shown in the figure.



As can be seen, for objects numbered 1 and 2 the ε -tubes intersect, i.e., these objects are placed in one cluster, while objects numbered 3 and 4 each form their own cluster. Thus, it is possible to group objects into various clusters by choosing the appropriate method for accounting for their structure and scale.

Publication citation analysis

Publication citation graph is constructed. The purpose of constructing this graph is to determine the influence of vertices. The vertices of the graph are publications, the edges represent citations. The type of graph is directed. We analyze a graph containing 39,811 vertices and 310,829 edges. Citation analysis allows the identification of key research directions and determination of trends in their development.

The citation graph of publications makes it possible to analyze the interaction and influence exerted by studies presented in scientific publications on each other. The nodes in this graph are articles, and the directed unweighted edges are citations (all weights are equal to 1). Thus, if an article with ID1 cites an article with ID2 in year Y, then a directed edge from node ID1 to node ID2 appears in the network. The 'weight' parameter contains the weight of the edge or the number of citations, i.e., in this network it is equal to 1, because each article can cite another no more than once.

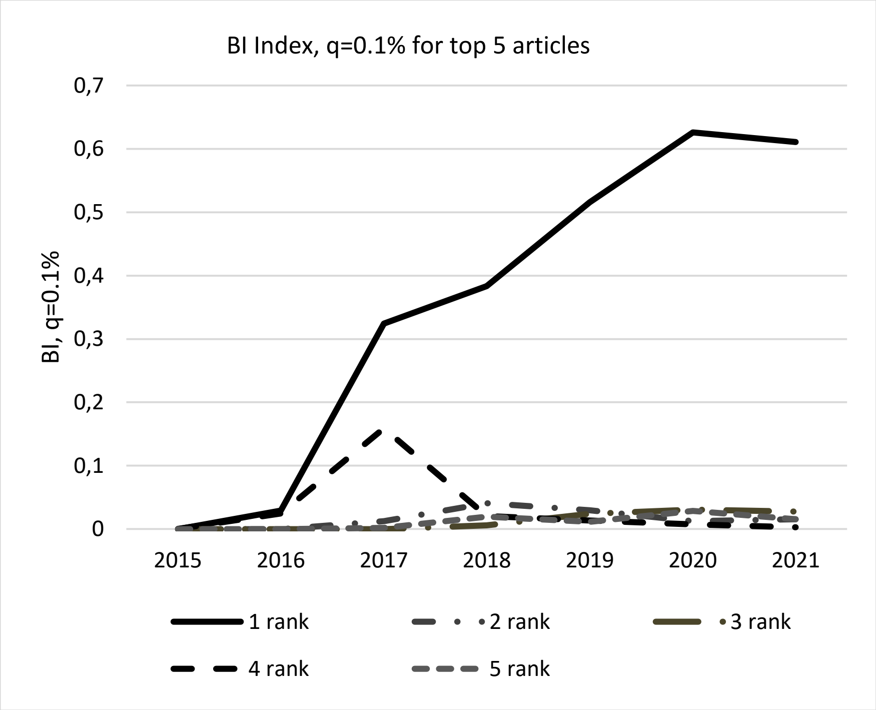
Publication citation analysis

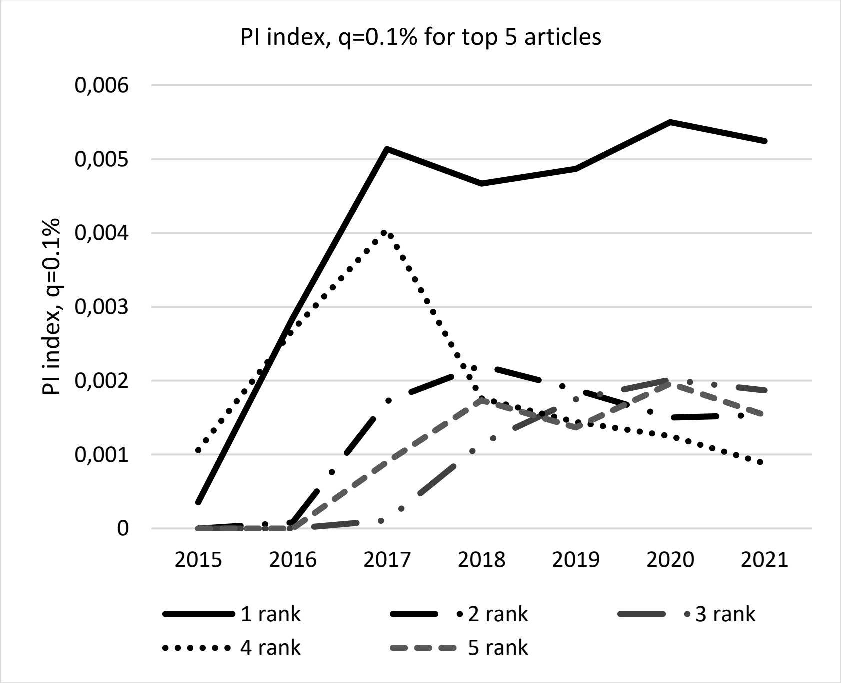
New centrality indices were applied to identify groups of specialized articles that actively cite each other. The maximum size of the critical set is 3, that is, 3 nodes can simultaneously influence another node. The following indices were calculated: PI, BI, In-degree, and the overall influence index TI. For the new indices, the following quota values were considered as percentages of the total number of citations per article, 0.1% of citations, 0.5% of citations, 1% of citations. The table presents the values of the PI, BI, TI indices for the 0.1% quota and the In-degree index for the Top-10 articles. The list of publications is ranked by TI.

№	Title	In-degree	q=0.1%		
			BI	PI	TI
1	MDS clinical diagnostic criteria for Parkinson's disease	0.005	0.63	0.94	0.53
2	Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease	0.0016	0.02	0.000096	0.0075
3	Epidemiology of Parkinson's disease	0.00156	0.018	0.000093	0.0068
4	MDS research criteria for prodromal Parkinson's disease	0.00152	0.0175	0.000091	0.0064
5	The epidemiology of Parkinson's disease: risk factors and prevention	0.0015	0.0170	0.00009	0.0062
6	Non-motor features of Parkinson disease	0.00147	0.016	0.000088	0.0058
7	Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016	0.00147	0.016	0.000088	0.0058
8	The Roles of PINK1, Parkin and Mitochondrial Fidelity in Parkinson's Disease	0.00147	0.0159	0.0000882	0.0058
9	α-Synuclein strains cause distinct synucleinopathies after local and systemic administration	0.0014	0.014	0.000085	0.0052
10	Gut microbiota are related to Parkinson's disease and clinical phenotype	0.0013	0.012	0.00008	0.0044

Publication citation analysis

The figures show the dynamics of changes in the BI and PI indices by year for the Top-5 publications. The Bundle index for the first article was the highest after 2016, whereas for the other publications the index was significantly lower and fluctuated between 0 and 0.15, with a maximum value 1–3 years after publication. The dynamics of the Pivotal index by years fully coincides with the In-degree index. In this case, the small quota (q=0.1%) does not lead to a change in the number of pivotal nodes. The value of TI over the years behaves similarly to the BI index.





Journal citation analysis

Journal citation graph is constructed. Information for each publication contains the journal in which it was published, which allows the construction of a network of interactions among scientific journals on a specific topic. The graph in this case is constructed as follows: vertices represent journals, and the edge weight is the number of references from one journal to another over a certain period. The type of graph is weighted and directed (edges are assigned weights – the number of references). Unlike publications, journals may cite themselves. If journals cite themselves, loops are formed in the graph (a vertex is connected to itself). The constructed journal citation graph contained 3,292 vertices and 152,203 edges. Journal citation analysis makes it possible to identify the most influential journals, determine the journal's rank in the research field, its productivity, and scientific popularity.

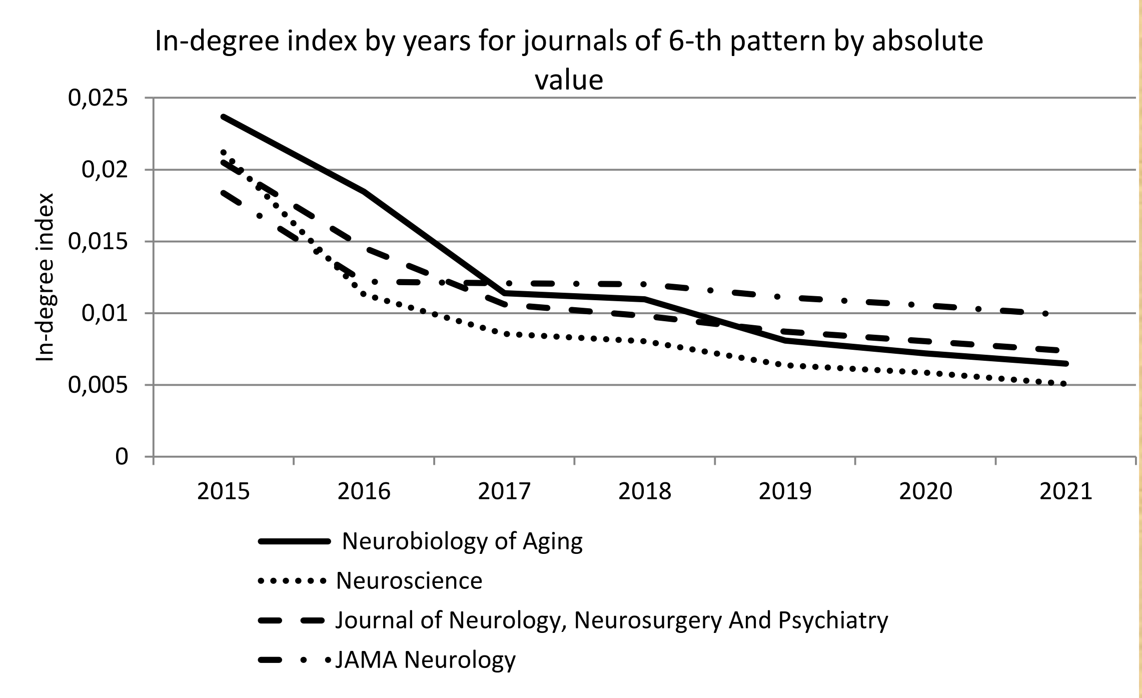
Journal citation analysis

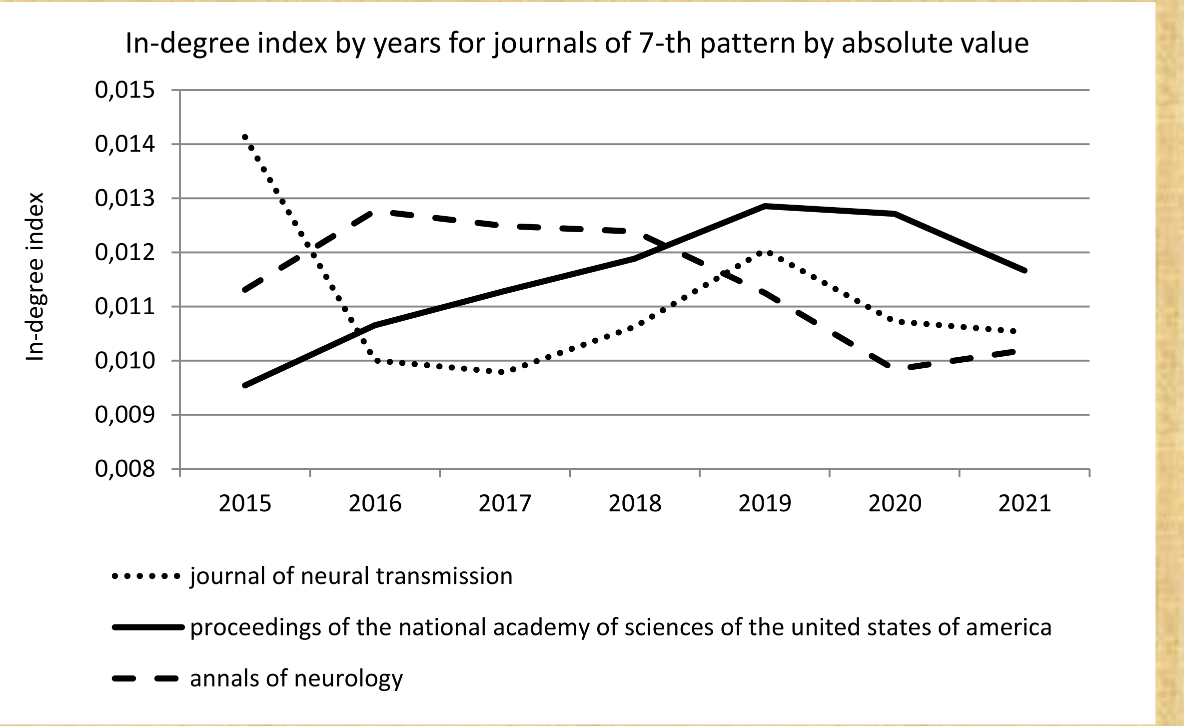
New centrality indices were calculated, as well as overall influence, by which journals were ranked. The calculation results showed that the Top-10 journal list for the quotas 0.5%, 1%, and 3% is practically identical to the list obtained for classical indices. As an example for the Top-10 journals, the value of the TI index and the ranks of the BI and PI indices for the quota q=0.5% are provided. The calculations were carried out taking into account journal self-citation. For comparison, the values of the journals' In-degree are given. The journals are ranked by the In-degree index.

№	Journal	In-degree values		Rank q=0.5%		
			TI	TI	BI	PI
1	Movement Disorders	0.08	0.06	1	1	1
2	Parkinsonism & Related Disorders	0.05	0.04	2	2	2
3	Neurology	0.02	0.02	6	7	10
4	Journal of Parkinson's Disease	0.02	0.027	4	4	4
5	Plos One	0.019	0.031	3	3	3
6	Brain	0.019	0.017	8	11	13
7	Scientific Reports	0.018	0.024	5	5	5
8	Lancet Neurology	0.016	0.020	7	6	6
9	Neurobiology of Disease	0.015	0.016	11	12	12
10	Proceedings of the National Academy of Sciences of the United States of America	0.012	0.012	20	22	24

Clustering of journals centrality patterns

Clustering of journal patterns by change in their citation by years. An analysis of patterns with various parameters was conducted for 3,292 journals. The main task of such analysis is clustering of journals and identification of differing patterns. This made it possible to determine the journals whose citation increased during the period from 2015 to 2021, and those journals whose citations declined. Pattern analysis was carried out based on changes in the In-degree index by year. For the analysis, an epsilon tube of $\varepsilon = 0.05$ was selected. Two variants of clustering were performed – clustering by absolute value and by slope angle.

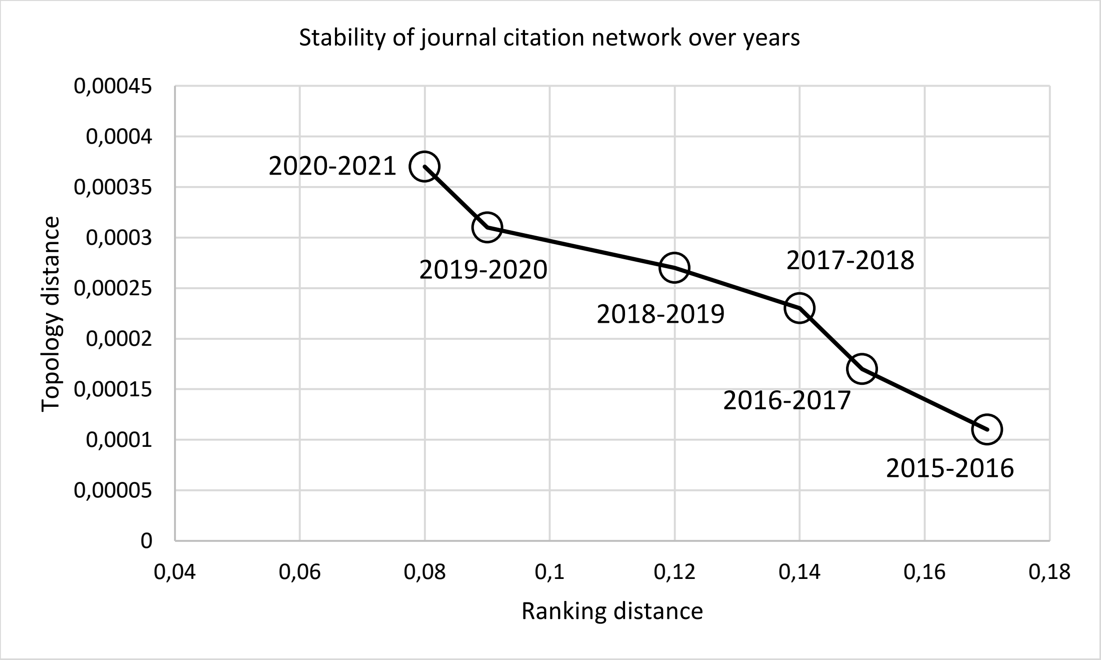




Journal citation network stability analysis

A stability analysis was conducted for the journal citation network consisting of 3,292 nodes. Six networks were constructed for each year from 2015 to 2021, evaluating the dynamic changes in the citation network. The in-degree index was chosen as the centrality indicator of a node for a specific year. The coefficient of influence of one node on another was calculated as follows:

$$\tilde{c}_{ij}^{t} = \begin{cases} \frac{w_{ij}}{\sum_k w_{kj}}, if\ an\ edge\ from\ i\ to\ j\ exists \\ 0\ , else \end{cases}$$



The results of the stability analysis are represented on a two-dimensional plane. The horizontal axis indicates the similarity of the networks in terms of centrality, and the vertical axis indicates similarity in terms of structure.

Author citation analysis

Author citation graph is constructed. Each publication also contains information about the authors. For each article, it was decided to select the first author from the list, as in medical articles the first author is usually the leading author or the project leader. Thus, the vertices of the graph are the first authors of the articles.

An edge corresponds to the citation of an article by author A of an article by author B. The edge weight is the number of citations for a given year. The constructed author citation graph contained 27,551 author vertices and 271,623 edges. The analysis of the author publication activity graph is mainly carried out to identify the most significant authors.

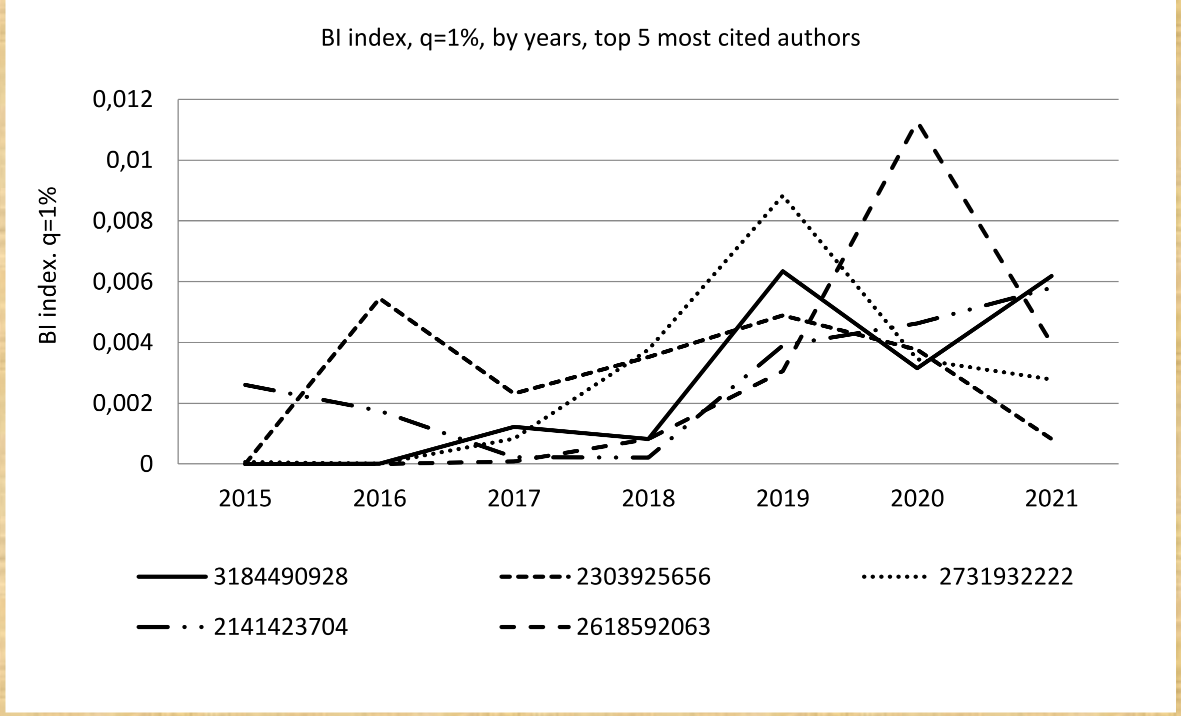
Author citation analysis

The table presents the Top-10 authors by BI, PI, and TI indices with a quota of 0.1%, ordered by the TI index. As in the case of journals, the results are almost identical to the Top-10 by the In-degree index or the list of most cited authors. It is interesting that the PI index of the first author (AuId 2047227543) is nearly equal to 1, while the other authors have values close to zero. This means that the first author has the highest number of pivotal nodes, an order of magnitude greater than the others.

№	Author Id	Affiliation	In-degree values	q = 0.1%		
				BI	PI	TI
1	2047227543	Montreal General Hospital	0.0071	0.577	0.99	0.528
2	2303827053	University of Rochester	0.0023	0.0251	0	0.0091
3	2064657699	University College London	0.0024	0.0245	0	0.009
4	2600996391	National Institute of Neurological Disorders and Stroke (NINDS)	0.0018	0.0142	0	0.0053
5	2195539447	Haukeland University Hospital	0.00164	0.0122	0	0.0046
6	2112352720	California Institute of Technology	0.00168	0.012	0	0.00457
7	78852038	University of Cincinnati	0.0018	0.011	0	0.0043
8	1761635969	National Institutes of Health	0.00197	0.0104	0	0.0041
9	1991986455	Harvard University	0.0016	0.0105	0	0.004
10	2100557140	Veterans Health Administration	0.0017	0.0096	0	0.0038

Author citation analysis

The dynamics of changes over the years in BI and PI indices for authors are presented using the example of quota q = 1%. The figure shows the changes in the BI index values from 2015 to 2021 for the Top-5 authors.



For example the author with AuId: 2731932222 had an index value close to 0 before 2017, which indicates low collective influence during that period. However, from 2017 to 2019 the index value increased significantly. This may be explained by the fact that during this period the author became part of more influential research groups or his works began to gain collective recognition from more significant authors. Since 2019, the index has begun to decline again, which may be due to a weakening of his role in collaborations or a decrease in citation of his works by other key authors.

Institutions citation analysis

Affiliation citation graph is constructed. At the data processing stage, the affiliations of researchers were identified for each article by processing the relevant fields using the Python programming language. As a result, a list of affiliations for the authors of each article was obtained. Two graphs were constructed – a network based on the first author's affiliation and based on all authors' affiliations. The network based on the first author's affiliation contained 3,029 organization vertices and 128,602 edges. The analysis of such a network determines the significance of scientific teams and research centers in a given research field. To analyze the organizational collaboration network, a graph was constructed in which the vertices represented the affiliations of all authors, and edges were drawn between them if authors from these affiliations had co-authored publications. The organizational collaboration network is an undirected graph with 4,697 vertices and 85,900 edges.

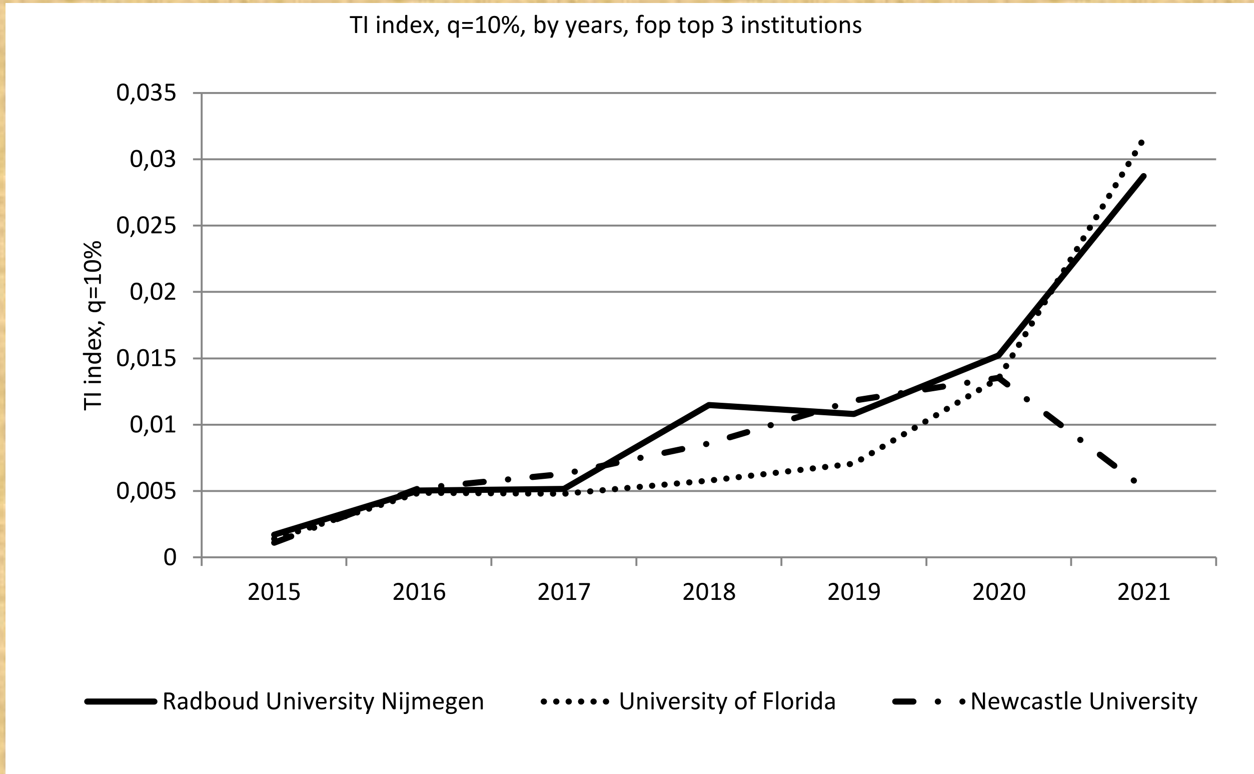
Institutions citation analysis

Since authors of publications can cite both their own works and those of colleagues working in the same organization, loops appear in the affiliation network. In this network, a node may belong to its own critical set.
The table presents the BI, PI, and TI index values with the smallest quota q=0.1% for the Top-10 affiliations. The list is ranked by TI value. For comparison, the table also includes In-degree index values for the affiliations.

№	Institutions	In-degree	q=0.1%.		
			BI	PI	TI
1	UCL Institute of Neurology	0.013	0.044	0.21	0.089
2	National Institutes of Health	0.016	0.037	0.16	0.07
3	University of Cambridge	0.013	0.032	0.139	0.061
4	University College London	0.012	0.037	0.106	0.051
5	Harvard University	0.010	0.033	0.109	0.05
6	University of Oxford	0.012	0.031	0.103	0.049
7	Northwestern University	0.011	0.031	0.097	0.046
8	University of Pennsylvania	0.0085	0.019	0.069	0.032
9	Capital Medical University	0.0078	0.023	0.0004	0.01
10	Karolinska Institutet	0.0079	0.0198	0.0003	0.0093

Institutions citation analysis

The dynamics of changes in the TI index with a quota of q = 10% from 2015 to 2021 are shown for three affiliations: Radboud University Nijmegen, University of Florida, and Newcastle University



Term co-occurrence analysis

Term co-occurrence graph is constructed. The research topic can often be determined by the terminology used in publications. Therefore, term analysis provides a broad understanding of the development of various research directions and helps to identify future prospects. From the abstracts of the downloaded publications, we identified six sets of terms that define different research directions on Parkinson’s disease. These sets include from 20 to 147 terms. The following six research directions were identified

- Biochemistry, cell, genetics (BCG)
- Medicines
- Symptoms
- Central nervous system (CNS)
- Diagnostics
- Research methods

Term co-occurrence analysis

The analysis of the number of publications by year showed that there were no sharp changes in the proportion of articles mentioning terms from different fields. The presents the number of publications in each year (from 2015 to 2021) in which at least one term from the considered research area appeared. The table also shows the relative frequency of terms from the given group for each considered year. Note, that an article can contain several terms.

		2015	2016	2017	2018	2019	2020	2021
Number of publications	Total 39811	4613	4908	5396	5139	5876	6816	7063
BCG	articles	3166	3318	3618	3538	4016	4680	4736
	% to total	68.6	67.6	67.0	68.8	68.3	68.7	67.1
Medications	articles	1136	1212	1254	1187	1323	1490	1580
	% to total	24.6	24.7	23.2	23.1	22.5	21.9	22.4
Symptoms	articles	3129	3287	3598	3493	3912	4604	4779
	% to total	67.8	67.0	66.7	68.0	66.6	67.5	67.7
CNS	articles	2114	2203	2432	2294	2645	2918	3057
	% to total	45.8	44.9	45.1	44.6	45.0	42.8	43.3
Diagnosis	articles	1204	1275	1428	1363	1662	1899	2008
	% to total	26.1	25.9	26.4	26.5	28.2	27.8	28.4
Research methods	articles	1141	1189	1300	1260	1417	1601	1672
	% to total	24.7	24.2	24.1	24.5	24.1	23.5	23.7

Network of terms

For each of the six research directions, a co-occurrence network of terms in articles was constructed taking into account the centrality of the articles. The network is built as follows.

In the abstracts of articles, at least one pair of terms from the same topic is identified. There may be several such pairs. These terms are the nodes of the network. There is an edge between terms if a pair of terms simultaneously appears (as a pair) in article abstracts. The weight of the undirected edge between the terms (network nodes) is calculated as the sum of the centralities of all articles whose abstract simultaneously includes this pair of terms. Then, for each network, the centrality of the terms is calculated. For the term network built based on article centrality using the In-degree Index, the centrality of terms is calculated using the In-degree index. For the network built based on the Bundle Index (BI), the centrality of terms is calculated using BI. For the term network built based on the Pivotal Index (PI), the centrality of terms is calculated using the PI index. For BI and PI indices, quotas of 1%, 5%, and 10% were considered. Group size restrictions k=5 were applied.

Thus, the network allows for identifying terms that not only occur frequently but also appear in the most cited publications. At the same time, the new centrality indices allow for highlighting terms that most frequently appear in narrowly specialized articles actively citing each other. We carried out the calculation of the considered In-degree index, Bundle Influence index (BI), and Pivotal Influence index (PI) for the compiled list of terms for each research direction of PD – BCG (biochemistry, cell, genetics), Medications, Symptom, CNS (central nervous system), Diagnosis, Research methods.

Centrality in the network of terms

The Top-10 terms by the BI index for the quota q=1% are presented in the table. For comparison, the table shows the ranks of these terms by the In-degree and PI indices. The Top-10 terms by the Pivotal Influence Index PI (q=1%) are presented, which also provides the ranks of these terms by the In-degree and BI indices for the Top-10 terms of the BCG group.

BI index, BCG research direction						PI index, BCG research direction				
№	Terms	BI q=1%	Rank			№	Terms	PI q=1%	Rank	
			In-degree	PI					In-degree	BI
1	dopamine	0.031	3	1		1	dopamine	0.038	3	1
2	synuclein	0.0298	2	2		2	synuclein	0.036	2	2
3	protein	0.0297	1	3		3	protein	0.033	1	3
4	gene	0.0289	4	6		4	mrna	0.032	32	15
5	genetic	0.026	5	5		5	genetic	0.029	5	5
6	cellular	0.025	7	9		6	gene	0.0288	4	4
7	oxidation	0.0244	8	10		7	molecule	0.0275	6	8
8	molecule	0.0243	6	7		8	membrane	0.0268	14	12
9	enzyme	0.021	19	12		9	cellular	0.0266	7	6
10	neurotoxin	0.01	25	13		10	oxidation	0.0266	8	7

Centrality in the network of terms

The table presents the Top-10 CNS terms most frequently encountered in the total number of publications. For comparison, it presents the Top-10 terms ranked by the value of the In-degree index. This table also provides the term ranks according to the BI and PI indices (q=1%). It is noted that the two terms "nigra" and "striatum" are leaders in both lists, i.e., they most frequently occur both in the total number of publications and in the most cited publications. The two terms "subthalamic" and "microglia", which entered the Top-10 by total number of publications, dropped out of the Top-10 most cited list and were replaced by "putamen" and "thalamus".

BI index values for the Top-10 terms of the CNS group						PI index values for the Top-10 terms of the CNS group				
№	Terms	BI q=1%	Rank			№	Terms	PI q=1%	Rank	
			In-degree	PI q=1%					In-degree	BI q=1%
1	basal	0.064	4	3		1	subcortical	0.079	14	9
2	ganglia	0.062	5	2		2	ganglia	0.072	5	2
3	cortex	0.057	3	5		3	basal	0.070	4	1
4	thalamus	0.057	10	18		4	hippocampus	0.068	15	13
5	nigra	0.057	1	9		5	cortex	0.0599	3	3
6	nigrostriatal	0.056	6	6		6	nigrostriatal	0.057	6	6
7	striatum	0.05	2	11		7	putamen	0.054	9	12
8	cerebral	0.049	13	8		8	cerebral	0.053	13	8
9	subcortical	0.043	14	1		9	nigra	0.052	1	5
10	cerebellar	0.042	17	28		10	neurogenesis	0.048	26	19

Conclusion

Historically, the impact of any article, author or journal had been evaluated by a scientific community, sometimes too late. Today, due to enormous increase of the number of publications there is a need in high quality quantitative models analyzing the impact of a journal or an institution or an author to the corresponding field of research. Responding to this demand many colleagues suggested some models based on very different tools – from aggregation models to semantic analysis, from statistics to optimization techniques. We presented an attempt to construct such models based on a new technique proposed for social network analysis. This very technique include the new centrality measures, models of pattern analysis and stability.

Thank you

References

Abdrakhmanova, G., Arundel, A., De Fuentes, C., de Jong, P. J., De Marchi, V., Gokhberg, L., Roud, V., Streltsova, Ye., & Vishnevskiy, K. (2023). Handbook of Innovation Indicators and Measurement, 2nd edition. Ed. by F. Gault, A. Arundel, E. Kraemer-Mbula. Edward Elgar Publishing.

Abeynayake, I., & Tanner, C. M. (2020). The economic impact of OFF periods in Parkinson disease. J Manag Care, 26, 265–269. doi:10.37765/ajmc.2020.88518

Akintunde, T. Y., Musa, T. H., Musa, H. H., Musa, I. H., Chen, S., Ibrahim, E., Tassang, A., & Helmy, M. (2021). Bibliometric analysis of global scientific literature on effects of COVID-19 pandemic on mental health. Asian Journal of Psychiatry, 63, 102753. doi: 10.1016/j.ajp.2021.102753

Aleskerov, F. T., Belousova, V. Yu., Egorova, L. G., & Mirkin, B. G. (2013). Analysis of patterns in statics and dynamics, part 2: Examples of application to the analysis of socio-economic processes. Business informatics, 4 (26), 3–20. (in Russian)

Aleskerov, F. T., Kazachinskaya, A. I., Karabekyan, D. S., Semina, A. S., & Yakuba, V. I. (2021). Economic journals of Russia, their characteristics and network analysis. Journal of the New Economic Association, 50(2), 170–182. doi: 10.31737/2221-2264-2021-50-2-9. (in Russian)

Aleskerov, F., & Alper, C. (2000). A clustering approach to some monetary facts: a long-run analysis of cross-country data. The Japanese Economic Review, 51(4), 555–567. doi: 10.1111/1468-5876.00170

Aleskerov, F., & Shvydun, S. (2017). A mathematical approach to conflict resolution in the arctic region. In : 3rd IEEE International Conference on Cybernetics (CYBCON). Exeter: IEEE, 145–150.

Aleskerov, F., & Yakuba, V. (2020). Matrix-vector approach to construct generalized centrality indices in networks. SSRN 3597948. ssrn.com/abstract=3597948

Aleskerov, F., Andrievskaya, I., & Permjakova, E. (2014). Key borrowers detected by the intensities of their short-range interactions. HSE University Basic Research Program Working Paper, Series Financial Economics, WR BRP, 33/FE/2014, ssrn.com/abstract=2479272

Aleskerov, F., Bouyssou, D., & Monjardet, B. (2007). Utility Maximization, Choice and Preference. Springer Berlin, Heidelberg, pp. XII, 283. ISBN 978-3-540-34182-6. doi: 10.1007/978-3-540-34183-3

Aleskerov, F., Buldyaev, A., Khutorskaya, O., & Yamilov, A. (2017). Parkinson’s disease: network analysis of patents and publications’ activity. Proceedings of the 11th IEEE International Conference on Application of Information and Communication Technologies (AICT2017, Moscow) 430–432. ISBN/ISSN: 978-1-5386-0500-4. doi:10.1109/ICAICT.2017.8686912

Aleskerov, F., Khutorskaya, O., Buldyaev, A., & Yamilov, A. (2018). Parkinson's disease: Network analysis of publications' impact. In 7th International Conference on Computers Communications and Control (ICCCC 2018, Oradea, Romania), 82–85. IEEE. ISBN/ISSN: 978-1-5386-1934-6

Aleskerov, F., Khutorskaya, O., Stepochkina, A., Yakuba, V., & Zinoveva, K. (2023). Network analysis of publications on studies of Parkinson Disease. Procedia Computer Science, 219, 1380–1387. doi:10.1016/j.procs.2023.01.426 www.sciencedirect.com/science/article/pii/S1877050923004350.

Aleskerov, F., Khutorskaya, O., Stepochkina, A., Yakuba, V., & Zinoveva, K. (2024). Affiliations based bibliometric analysis of publications on Parkinson’s disease. J Computational Management Science, 21(13), 1–14. doi:10.1007/s10287-023-00495-7

Aleskerov, F., Pislyakov, V., & Subochev, A. (2016). Ranking journals using social choice theory methods: a novel approach in bibliometrics. in : Proceedings of the 21 International Conference on Science and Technology Indicators. Ed. by I. Ràfols, J. Molas-Gallart, E. Castro-Martínez, R. Woolley. Valencia: Editorial Universitat Politècnica de València, 16, 1266–1273. doi: 10.4995/STI2016.2016.4543

Aleskerov, F., Shvydun, S., & Meshcheryakova, N. (2022). New centrality measures in networks: how to take into account the parameters of the nodes and group influence of nodes to nodes. CRC Press. doi:10.1201/9781003203421

Alvarez-Peregrina, C., Martinez-Perez, C., Villa-Collar, C., & Sánchez-Tena, M. (2021). A bibliometric and citation network analysis of myopia genetics. Genes, 12(3), 447. doi: 10.3390/genes12030447

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics, 11(4), 959–975. doi: 10.1016/j.joi.2017.08.007

Aroeira, R., & Castanho, M. (2020). Can citation metrics predict the true impact of scientific papers? The FEBS Journal, 287(2), 2440–2448. doi: 10.1111/febs.15255

Ashoka, R., & Harinarayana, N. S. (2022). Keyword Analysis of Research Trends in ‘Scientometrics’ Journal from 2014 to 2018. Journal of Information and Knowledge, 59(3), 163–169. doi: 10.17821/srels/2022/v59i3/165826

Beauchamp, L. C., Finkelstein, D. I., Bush, A. I., Evans, A. F., & Barnham, K. J. (2020). Parkinsonism as a third wave of the COVID-19 pandemic? Journal of Parkinson’s Disease, 10(4), 1343–1353. doi: 10.3233/JPD-202211

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvreet, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 10, P10008. doi: 10.1088/17425468/2008/10/P10008

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique Identification. Journal of mathematical sociology, 2(1), 113–120.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems, 30 (1–7), 107–117.

Brundin, P., Nath, A., & Beckman, J. D. (2020). Is COVID-19 a perfect storm for Parkinson’s disease? J Trends in Neurosciences, 43(12), 931–933. doi:10.1016/j.tins.2020.10.009

Cabanillas-Lazo, M., Quispe-Vicuna, C., Barja-Ore, J., Fernandez-Giusti, A., Munive-Degregori, A., Retamozo-Siancas, Ye., Guerrero, M. E., & Mayta-Tovalino, F. (2022). A 10-year bibliometric analysis of global research on gut microbiota and Parkinson’s disease: characteristics, impact, and trends. Hindawi BioMed Research International, 4144781, ISSN: 2314-6133. doi:10.1155/2022/4144781

Chahrour, M., Assi, S., Bejjani, M., Nasrallah, A. A., Salhab, H., Fares, M., & Khachfe, H. H. (2020). A bibliometric analysis of COVID-19 research activity: a call for increased output. Cureus, 12(3), e7357. doi: 10.7759/cureus.7357

References

Chen, L., Chen, J., Wu, M., Yan, P., & Zhou, X. (2024a). Analyzing the bibliometrics of brain-gut axis and Parkinson's disease. Front. Neurol. 15, 1343303. doi: 10.3389/fneur.2024.1343303

Chen, Q., Sun, T., & Wang, T. (2023). Network centrality, support organizations, exploratory innovation: Empirical analysis of China's integrated circuit industry. J Heliyon, 9(7), e17709. doi:10.1016/j.heliyon.2023.e17709

Chen, Y., Koch, T., Zakiyeva, N., Liu, K., Xu, Z., Chen, Ch-h., Nakano, J., & Honda, K. (2022). Article's Scientific Prestige: measuring the impact of individual articles in the Web of Science. arXiv:2202.08695 [cs.DL]. doi: 10.48550/arXiv.2202.08695

Chen, Y.-J., Xie, M.-R., Zhou, S.-Q., & Liu, F. (2024b). Mapping the research of mitochondria and Parkinson's disease: a bibliometric analysis. J Frontiers in Neurology, 15, 1413762. doi: 10.3389/fneur.2024.1413762

Cinpolat, H. Y., & Akar, A. (2023). A visual and bibliometric analysis of Parkinson's disease biomarker research. Ann Med Res, 30(10), 1359–64. doi: 10.5455/annalsmedres.2023.09.253

Cordasco, G., & Gargano, L. (2010). Community detection via semi-synchronous label propagation algorithms. IEEE international workshop on: business applications of social network analysis (BASNA), IEEE. 1–8. doi: 10.1109/BASNA.2010.5730298

Cowhitt, T., & Cutts, A. (2020). Using network analysis to compare bibliographic database journal coverage. Journal of Electronic Resources Librarianship, 32(3), 195–210. doi: 10.1080/1941126X.2020.1790952

Delardas, O., & Giannos, P. (2022). How COVID-19 Affected the journal impact factor of high impact medical journals: bibliometric analysis. J Med Internet Res. 24(12), e43089. doi: 10.2196/43089

Ding, J., Liu, C., & Kandonga, G. A. (2020). Exploring the limitations of the h-index and h-type indexes in measuring the research performance of authors. Scientometrics, 122, 1303–1322. doi: 10.1007/s11192-020-03364-1

Duggan, E. W., Atwood, G. S., Sanford, J. A., Tsai, M. H., Egbaria, J. K., Carmichael-Tanaka, N., & Outland, N. B. (2024). Using bibliometric data to define and understand publishing network equity in anesthesiology. Anesthesia & Analgesia, 139(5), 944–954. doi: 10.1213/ANE.0000000000006877

Fan, F. (2024). The new science of COVID-19: A bibliographic and network analysis arXiv:2407.15867 [cs.DL]. doi: 10.48550/arXiv.2407.15867

Fearon, C., & Fasano, A. (2021). Parkinson's disease and the COVID-19 pandemic. Journal of Parkinson's Disease, 11(2), 431–444. doi: 10.3233/JPD-202320

Findley, L. J. (2007). The economic impact of Parkinson's disease. J Parkinsonism Relat Disord, 13, 8–12. doi: 10.1016/j.parkreldis.2007.06.003

Fonseca, B. D., Sampaio, R. B., Fonseca, M. Vd., & Zicker, F. (2016). Co-authorship network analysis in health research: method and potential use. Health Research Policy and Systems. 14, 34. doi: 10.1186/s12961-016-0104-5.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. Sociometry, 40(1), 35–41.

Funada, S., Yoshioka, T., Luo, Y., Iwama, T., Mori, V., Yamada, N., Yoshida, H., Katanoda, K., & Furukawa, T. A. (2023). Global trends in highly cited studies in COVID-19 research. JAMA Netw Open, 6(9), e2332802. doi: 10.1001/jamanetworkopen.2023.32802

Furstenau, L. B., Rabaioli, B., Sott, M. K., Cossul, D., Bender, M. S., De Mattos, E. M. J., Barcellos, F. N., Severo, P. P., Dohan, M. S., & Bragazzi, N. L. (2021). A bibliometric network analysis of coronavirus during the first eight months of COVID-19 in 2020. Int. J Environ. Res. Public Health, 18(3), 952. doi: 10.3390/ijerph18030952

Gandasari, D., Tjahjana, D., Dwidienawati, D., & Sugiarto, M. (2024). Bibliometric and visualized analysis of social network analysis research on Scopus databases and VOSviewer. J Cogent Business & Management, 11(1), 2376899. doi: 10.1080/23311975.2024.2376899

Ganti, L., Persaud, N. A., & Stead, T. S. (2025). Bibliometric analysis methods for the medical literature. Academic Medicine & Surgery. 129134. doi: 10.62186/001c.129134 ISBN: 2996-2617

Gershman, M., Gokhberg, L., & Kuznetsova, T. (2025). Science, technology, and innovation in Russia: development challenges and opportunities. In: Zhao X, Sokolov A, Cassiolato J E (eds) The Innovation Competitiveness of BRICS Countries. Science, Technology and Innovation Studies. Springer, Singapore, 207–225. ISSN2570-1509. doi: 10.1007/978-981-97-8001-3_10

Gokhberg, L., Abdrakhmanova, G., Streltsova, E., & Vishnevskiy, K. (2023a). Measuring the digital transformation. in : Handbook of Innovation Indicators and Measurement, 2nd edition. Ed. by F. Gault, A. Arundel, E. Kraemer-Mbula. Edward Elgar Publishing, 12, 221–239.

Gokhberg, L., Kuznetsova, T., & Kotsemir, M. N. (2023b). From the Soviet Union to the Russian Federation: publication activity dynamics along the evolution of national science policies. Scientometrics, 128(11), 6195–6246. doi: 10.1007/s11192-023-04838-8

HabibAgahi, M. R., Kermani, M. A., & Maghsoudi, M. (2022). On the Co-authorship network analysis in the process mining research community: a social network analysis perspective. Expert Systems with Applications, 206, 117853. doi: 10.1016/j.eswa.2022.117853

Hakkaraki, V. P. (2024). Exploring the evolution of bibliometric analysis: A comprehensive study of scientific publications from 1974 to 2024 using the Dimensions AI database. Asian Journal of Information Science and Technology, 14(1), 24–31. doi: 10.70112/ajist-2024.14.1.3878

Ho, Y. S., & Shekofteh, M. (2021). Performance of highly cited multiple sclerosis publications in the Science Citation Index expanded: A scientometric analysis. Multiple Sclerosis and related disorders, 54, 103112. doi: 10.1016/j.msard.2021.103112

Ioannidis, J. (2006). Concentration of the most-cited papers in the scientific literature: analysis of journal ecosystems. PLoS ONE, 1(1), e5. doi: 10.1371/journal.pone.0000005

Koseoglu, M. A. (2016). Growth and structure of authorship and co-authorship network in the strategic management realm: Evidence from the Strategic Management Journal. BRQ Business Research Quarterly, 19(3), 153–170. doi: 10.1016/j.brq.2016.02.001

Li, X. L., Gao, R. X., Zhang, Q., Li, A., Cai, L.-N., Zhao, W.-W., Gao, S.-L., Wang, Y., & Yue, J. (2022). A bibliometric analysis of neuroimaging biomarkers in Parkinson disease based on Web of Science. Medicine (Baltimore), 101(33), e30079. doi: 10.1097/MD.0000000000030079

References

Liao, X. Y., Gao, Y. X., Qian, T. T., Zhou, L.-H., Li, L.-Q., Gong, Y., & Ye, T. F. (2024). Bibliometric analysis of electroencephalogram research in Parkinson’s disease from 2004 to 2023. Front. Neurosci., Sec. Neurodegeneration, 18, 1433583. doi: 10.3389/fnins.2024.1433583

Lin, T. Z., Tang, K. U., Lin, S. S., Changlai, M.-L., & Hsu, Y.-S. (2022). Co-word Analysis of Selected Science Education Literature: Identifying Research Trends of Scaffolding in Two Decades (2000–2019). Front. Psychol., 13, 844425. doi: 10.3389/fpsyg.2022.844425

Lindgren, P., von Campenhausen, S., Spottke, E., Siebert, U., & Dodel, R. (2005). Cost of Parkinson's disease in Europe. Eur J Neurol, 12(1), 68–73. doi:10.1111/j.1468-1331.2005.01197.x

Liu, Z., Yin, Y., Liu, W., & Dunford, M. (2015). Visualizing the intellectual structure and evolution of innovation systems research: a bibliometric analysis. Scientometrics, 103, 135–158. doi: 10.1007/s11192-014-1517-y

Lozano, S., Calzada-Infante, L., Adenso-Díaz, B., & García, S. (2019). Complex network analysis of keywords co-occurrence in the recent efficiency analysis literature. Scientometrics, 120, 609–629. doi: 10.1007/s11192-019-03132-w

Maltseva, D., & Batagelj, V. (2024). Citation and bibliographic coupling between authors in the field of social network analysis. Journal of Data and Information Science, 9(4), 110–154. doi:10.2478/jdis-2024-0028

McBenedict, B., Hauwanga, W. N., Ienaco, G., Dulci, P., Syeda, S. K., Jonatha, M. L., Barakat, B. O., Asaju, F., Sujood, A., Shivadeva, S., Phoh, W. C., & Bruno, L. P. (2024). Parkinson's disease treatment: A bibliometric analysis. Cureus, 16(9), e69613. doi:10.7759/cureus.69613

Mejia, C., Wu, M., Zhang, Y., & Kajikawa, Y. (2021). Exploring topics in bibliometric research through citation networks and semantic analysis. Front. Res. Metr. Anal., 6, 742311. doi: 10.3389/frma.2021.742311

Merigó, J., & Yang, J. (2017). Accounting research: a bibliometric analysis. Australian Accounting Review, 27, 71–100. doi: 10.1111/AUAR.12109

Minati, G. (2024). Linked Links – A Research Project: The Multiple Superimposed Soft Networks as Network Profiles. Systems, 12(8), 303. doi: 10.3390/systems12080303

Molontay, R., & Nagy, M. (2020). Twenty Years of Network Science: A Bibliographic and Co-Authorship Network Analysis. arXiv:2001.09006 [physics.soc-ph], doi: 10.48550/arXiv.2001.09006

Newman, M. E. (2006). Modularity and community structure in networks. Proceedings of the national academy of sciences, 103(23), 8577-8582. doi: 10.1073/pnas.0601602103

Newman, M. E. (2010). Networks: an introduction. New York, NY: Oxford University Press

Newman, M. E. (2018). Networks: Second Edition. Oxford: Oxford University Press.

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. Physical review E, 69(2), 026113. doi: 10.1103/PhysRevE.69.026113

Olesen, J., Gustavsson, A., Svensson, M., Wittchen, H.-U., & Jönsson, B. (2012). The economic cost of brain disorders in Europe. Eur J Neurol, 19(1), 155–162. doi: 10.1111/j.1468-1331.2011.03590.x

Pagani, R., Kovaleski, J., & Resende, L. (2015). Methodi Ordinatio: a proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citation, and year of publication. Scientometrics, 105, 2109–2135.

Pan, X., Liang, B., & Cao, T. (2024). A bibliometric analysis of speech and language impairments in Parkinson’s disease based on Web of Science. Front. Psychol., Sec. Psychology of Aging, 15, 1374924. doi: 10.3389/fpsyg.2024.1374924

Phan, T. L. (2021). Mapping the social entrepreneurship research: Bibliographic coupling, co-citation and co-word analyses. Cogent Business & Management, 8(1), 1896885. doi: 10.1080/23311975.2021.1896885

Prichard, A. (1969). Statistical bibliography or bibliometrics? Journal of Documentation, 25(4), 348–349.

Rashid, F., & Dudley, J. (2024). Bibliometric analysis of the 100 most-cited articles on the methods of shade-matching in dentistry. Clinical and Experimental Dental Research, 10(6), e70037. doi: 10.1002/cre2.70037

Raut, S., Hossain, N., & Fazio, S. (2025). Healthcare resilience: bibliometric and network analysis. Enterprise Information Systems, Art. No. 2464752. doi: 10.1080/17517575.2025.2464752

Rawat, K. S., & Sood, S. K. (2021). Emerging trends and global scope of big data analytics: a scientometric analysis. Qual Quant, 55, 1371–1396. doi: org/10.1007/s11135-020-01061-y

Ruiz-Roman, R., Martinez-Perez, C., Gil Prados, I., Cristóbal, I., & Sánchez-Tena, M. Á. (2022). COVID-19 and pregnancy: citation network analysis and evidence synthesis. JMIR Pediatr Parent, 5(1), e29189. doi:10.2196/29189

Saritas, O., Veselitskaya, N., & Sokolov, A. (2024). Foresight methodology and commonly used methods in the new generation foresight. In: 21st Century Foresight: Shaping the Future for Sustainable Social, Economic and Environmental Development in South Africa. Ed. by Saritas O, Sokolov A, Cele M. Cham: Springer. 4, 35–53. doi: 10.1007/978-3-031-69557-5_4

Segarra, S., & Ribeiro, A. (2016). Stability and Continuity of Centrality Measures in Weighted Graphs. IEEE Transactions on Signal Processing, 64(3), 543–555.

Sheng, Y., Huiting, Z., Qiang, Z., & Chenhui, L. (2024). A corpus-based bibliometric study of highly cited papers in sport sciences. Sage Open, 14(1). doi: 10.1177/21582440231225856

Shi, L., & Yih, B. (2024). Knowledge mapping and research trends of accidental falls in patients with Parkinson’s disease from 2003 to 2023: a bibliometric analysis. Frontiers in Neurology, 15, 1443799. doi: 10.3389/fneur.2024.1443799

Shishehgarkhaneh, M. B., Keivani, A., Moehler, R. C., Jelodari, N., & Laleh, S. R. (2022). Internet of Things (IoT), Building Information Modeling (BIM), and Digital Twin (DT) in construction industry: a review, bibliometric, and network analysis. Buildings, 12(10), 1503. doi: 10.3390/buildings12101503

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J., & Wang, K. (2015). An overview of Microsoft Academic Service (MAS) and applications. Proceedings of the 24th international World Wide Web Conferences, 243–246. doi: 10.1145/2740908.2742839

References

Sokolov, A., Shashnov, S., & Kotsemir, M. (2025). Intra-BRICS research collaboration: status and prospects. In: Zhao X, Sokolov A, Cassiolato J.E. (eds) The Innovation Competitiveness of BRICS Countries. Science, Technology and Innovation Studies, Springer, Singapore 3, 41–65. ISBN 978-981-97-8000-6. doi: 10.1007/978-981-97-8001-3_2

Stachura, A., & Banaszek. Ł., & Włodarski, P. K. (2023). Should experts guide research merit assessment? Discrepancies between Journal Impact Factor and national expert-based journal ranking list for evaluation of scientific achievement in Poland; a bibliometric analysis. medRxiv, doi: 10.1101/2023.10.26.23297621

Su, W. S., & Chang, C. Y. (2024). Root caries in older adults: A co-citation network analysis. Gerodontology, 41(4), 445-450. doi: 10.1111/ger.12745

Suresh, V., Dave, T., Ghosh, S., Jena, R., & Sanker, V. (2024). Deep brain stimulation in Parkinson's disease: A scientometric and bibliometric analysis, trends, and research hotspots. Medicine, 103(20), e38152. doi:10.1097/MD.0000000000038152.

Swanson, D., & Santamaria, L. (2021). Pandemic supply chain research: a structured literature review and bibliometric network analysis. Logistics, 5(1), 7. doi: 10.3390/logistics5010007

The economic burden of Parkinson's disease (2019). The Lewin group report, date posted : July 5, 2019 www.parkinson.org/sites/default/files/documents/2019-parkinsons-economic-burden-study.pdf

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Scientific reports, 9(1), 5233. doi: 10.1038/s41598-019-41695-z

Tsai, S. T., Lu, M. K., San, S., & Tsai, C. H. (2020). The neurologic manifestations of coronavirus disease 2019 pandemic: a systemic review. Frontiers in Neurology, 11, 498. doi:10.3389/fneur.2020.00498

Utegenova, A. B., Yermagambetova, A. P., Kabdrakhmanova, G. B., Khamidulla, A. A., Urasheva, Z. U., Kenzhina, N. K., Zhussupova, Z., & Nurlanova, G. N. (2025). Bibliometric Analysis of Alpha-Synuclein Determination by Biopsy in Peripheral Tissues of Patients With Parkinson's Disease. International Journal of Telemedicine and Applications, 2025:6469893. doi: 10.1155/ijta/6469893

Waltman, L., & van Eck, N. J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. Eur. Phys. J., 86, 471. doi: 10.1140/epjb/e2013-40829-0

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Eide, D., Dong, Y., Qian, J., Kanakia, A., Chen, A., & Rogahn, R. (2019). A review of Microsoft Academic Services for science of science studies. Frontiers in Big Data, 2, 45. doi: 10.3389/fdata.2019.00045

Wiener, N. (1914). A simplification in the logic of relations. Proceedings of the Cambridge Philosophical Society, 13, 387–390.

Yazdanjue, N., Yazdanjouei, H., Gharoun, H., Khorshidi, M. S., Rakhshaninejad, M., Amiri, B., & Gandomi, A. H. (2025). A comprehensive bibliometric analysis on social network anonymization: current approaches and future directions. Knowledge and Information Systems 67, 29–108 doi: 10.1007/s10115-024-02289-y

Yeung, A. W. K., Litvinova, O., Bragazzi, N. L., et al. (2024). Digital health and mobile health: a bibliometric analysis of the 100 most cited papers and their contributing authors. Explor Digit Health Technol, 2, 86–100. doi: 10.37349/edht.2024.00013

Zeng, Z., Jia, C., Zhang, W., Zhuang, X., & Li, X. (2024). Assessing journals through a three-dimensional framework based on article citation, author and institution influence. Journal of Information Science, 50, 786–807. doi: 10.1177/01655515241238401

Zhang, T., Yang, R., Pan, J., & Huang, S. (2023). Parkinson's disease related depression and anxiety: a 22-year bibliometric analysis (2000-2022). Neuropsychiatric disease and treatment, 19, 1477–1489. doi: 10.2147/NDT.S403002

Zhou, H., Zhu, X., Dai, J., & Wu, W. (2023). Innovation evolution of industry-university-research cooperation under low-carbon development background: In case of 2 carbon neutrality technologies. Sustainable Energy Technologies and Assessments 55, 102976. doi: 10.1016/j.seta.2022.102976

Zhu, L., Zhu, P., Wang, J., Yan, K., Zhao, S., Jiang, Y., & Zhang, H. (2024). A bibliometric and visual analysis of Parkinson's disease sleep disorders: articles from 2008 to 2023. Frontiers in Psychiatry, 15, 1468568. doi: 10.3389/fpsyt.2024.1468568