



Факультет компьютерных наук

Прикладная математика и
информатика

Москва 2023

Сетевые модели по публикациям в области научных исследований болезни Паркинсона

Network Analysis of Publications on Studies of Parkinson Disease

Работу выполнили:
Степочкина Анна, БПМИ194
Зиновьева Ксения, БПМИ194
Научный руководитель:
Алескеров Ф. Т.

Тематическая область

Болезнь Паркинсона (БП) - это нейродегенеративное, прогрессирующее заболевание, в основном характерное для людей старшего возраста.

В США в 2020 году 930 тысяч человек живут с БП, а к 2030 это число возрастет до 1.2 миллиона, каждый год болезнь диагностируется у 60 тысяч людей.

Прямые и косвенные затраты на БП в США составляют 52 миллиарда долларов ежегодно, на лекарства - \$2500, а операции достигают стоимости в 100 тысяч долларов на одного человека.



**Маскообразное
лицо
(гипомимия)**

Сутулая поза

Ригидность

**Тремор покоя
в руке**

**Согнутые
бедра
и колени**

**Шаркающая
походка
мелкими
шажками**



Цели

- Проанализировать основных участников в области исследований БП
- Провести апробацию новых методов сетевого анализа
- Разработать подход к анализу научной области с использованием сетевых моделей

Обзор литературы

Анализ публикаций по БП:

- количество цитирований [1],
- индекс Хирша [2],
- анализ международных коллабораций, моделирование с помощью экспоненциальной регрессии [3],
- анализ статей с числом цитирования большим 400 [4].

Индексы центральности SRIC и LRIC в работе 2018 года [5].

Используя сетевой анализ цитирований, можно выделить ключевые исследования и журналы, в которых они публикуются.

Этапы работы

1. Обзор литературы и источников
2. Сбор данных из базы Microsoft Academic
3. Обработка данных
4. Предварительный анализ
5. Построение сети цитирования для статей, журналов, авторов и аффилиаций
6. Расчет индексов центральности по сетям
7. Выделение топ-10 вершин по индексам, их сопоставление
8. Анализ динамических изменений в индексах по годам
9. Основные выводы и направление дальнейших исследований
10. Подготовка статьи по данной теме

Описание данных

Из базы научных публикаций Microsoft Academic были скачаны статьи 2015-2021 года со словами “parkinson” и “disease” в названии или аннотации.

Атрибуты статьи:

- Id – ID публикации
- W, AW – уникальные нормализованные слова (приведенные к единой форме) в заголовке и абстракте
- Y – год публикации
- RId – список ID статей, на которые ссылается публикация
- DOI – цифровой идентификатор объекта
- J.Id – ID журнала

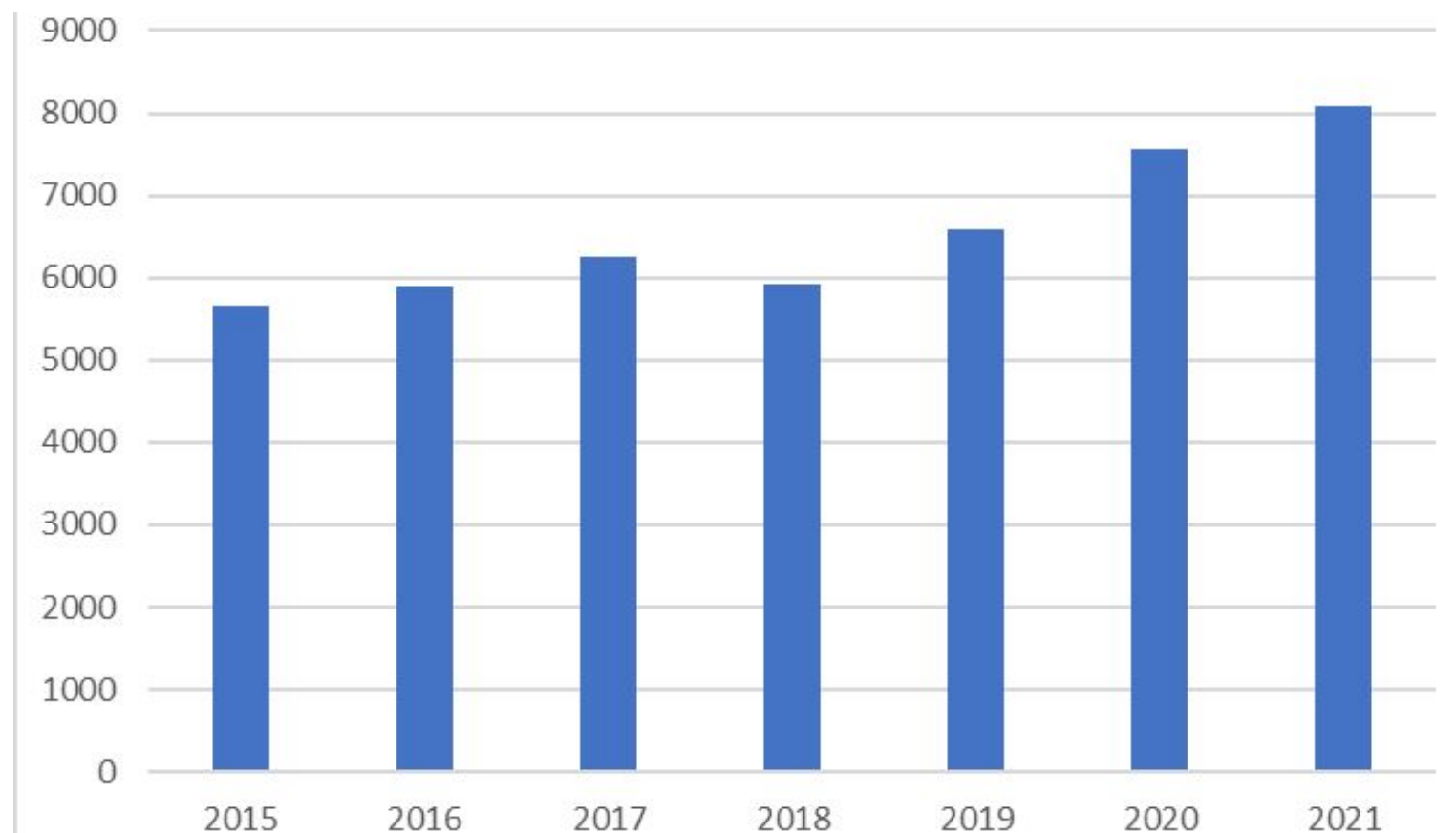
Всего статей скачано: 70119.

Из них

- **10681 без DOI**
- **7315 с DOI и без аннотации**
- **52123 с DOI и аннотацией**
- **45940 с DOI, аннотацией и журналом**



Распределение статей с DOI, аннотацией и журналом по годам



Сеть цитирований по статьям

Сеть цитирований – ориентированный граф, где статьи – вершины, а ребра – цитирования. Если статья ID1 ссылается на статью ID2, то появляется направленное ребро от ID1 к ID2. Y – год статьи ID1.

Количество ребер 2698197, вершин - 836533.

ID1	ID2	weight	Y
2754967293	2339791932	1	2017
2112455323	1920030402	1	2015
2177834950	2123627348	1	2015
2558041282	1934236512	1	2016
2558041282	1947901277	1	2016
2558041282	1564387586	1	2016
2558041282	1903888485	1	2016
2558041282	2338408645	1	2016
2558041282	2508023300	1	2016
2558041282	2097335121	1	2016

Обработка данных

Удалены:

- 115 статей с синдром Вольфа-Паркинсона-Вайта
- Вершины не из первоначального множества (например, если есть ссылка на статью, которая не содержит слов “parkinson” и “disease”)
- Изолированные вершины
- Ребра, в которых год ссылающей публикации меньше чем год цитируемой (4625 ребер, 105 вершин)

	Количество вершин	Количество ребер
Исходная сеть	780796	2488343
После удаления ребер, где хотя бы одна вершина не входит в исходный набор статей	39825	312044
После удаления неправильных ребер	39811	310829

Сеть цитирований по статьям

Количество вершин	39811
Количество ребер	310829
Количество компонент связности	94
Размер наибольшей компоненты связности	39618
Плотность графа	0.0002
Минимальное количество цитирований	0
Максимальное количество цитирований	1563
Среднее количество цитирований	7.8

$$density = \frac{m}{n(n-1)}$$

m - количество ребер
n - количество вершин

Индексы центральности

Классические индексы

1. In-degree index – сумма весов входящих ребер

$$x_i = \sum_j w_{ij}$$

2. Eigenvector index – решение уравнения $Ax = \lambda_1 x$, где λ_1 максимальное собственное значение матрицы смежности A

$$x_i = \frac{1}{\lambda_1} \sum_j A_{ij} x_j$$

Классические индексы

3. PageRank centrality – это разновидность центральности по собственному вектору, которая учитывает исходящую степень

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

4. Betweenness centrality – показывает долю кратчайших путей между двумя вершинами, на которых лежит исследуемая вершина

$$x_i = \sum_{kj} \frac{n_{kj}^i}{g_{kj}}$$

Новые индексы

2. Pivotal index (PI) показывает влияние ключевых вершин.

Вершина j_p называется ключевой, если $\sum_{j \in S} w_{ji} \geq q_i$ и $\sum_{j \in S \setminus \{j_p\}} w_{ji} < q_i$.

Значение индекса для критического множества равно количеству ключевых вершин в нем. Для вершины: $PI(i) = \sum_S |S| \times PI_i(S)$

Общее влияние: $TI(i) = \frac{1}{3} \ln - degree(i) + \frac{1}{3} BI(i) + \frac{1}{3} PI(i)$

Новые индексы

S – критическое множество для вершины i , если $S \subseteq V \setminus \{i\}$,

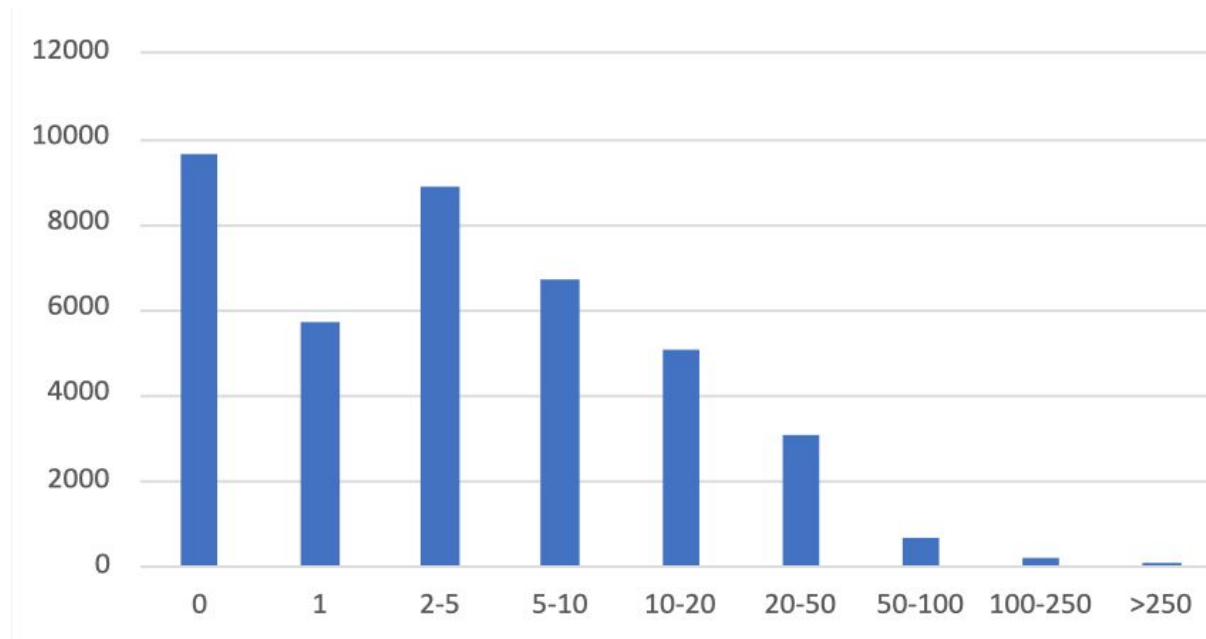
$|S| \leq k, \sum_{j \in S} w_{ji} \geq q_i$, где квота q_i - процент суммы весов входящих ребер, k – количество вершин, которые одновременно могут влиять на узел. В работе $k=3$. С увеличением квоты уменьшается количество критических множеств.

1. Bundle index (BI) учитывает групповое влияние на вершину

$$BI_i(S) = 1, \text{ если } \sum_{j \in S} w_{ji} \geq q_i \quad BI_i(S) = 0, \text{ иначе}$$

$$BI(i) = \sum_S BI_i(S)$$

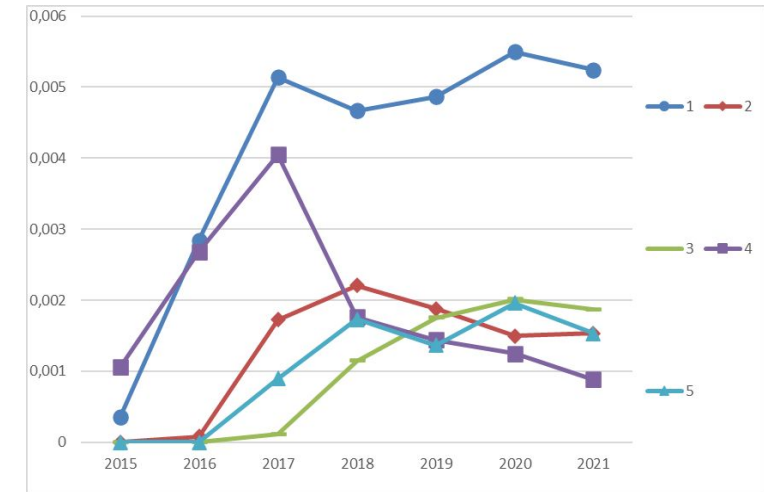
Результаты для статей



Распределение количества цитирований

Топ-5 статей по индексу In-degree

Название	DOI	In-degree	Ранг по Eigenvector	Ранг по PageRank
1. MDS clinical diagnostic criteria for Parkinson's disease	10.1002/MDS.26424	0.005	15	1
2. Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease	10.1016/J.CELL.2016.11.018	0.0016	62	8
3. Epidemiology of Parkinson's disease	10.1007/S00702-017-1686-Y	0.00156	1224	5
4. MDS research criteria for prodromal Parkinson's disease	10.1002/MDS.26431	0.00152	13	3
5. The epidemiology of Parkinson's disease: risk factors and prevention	10.1016/S1474-4422(16)30230-7	0.0015	705	4



Изменение индекса по годам

Топ-5 статей по индексу TI, $q = 0,1\%$

Название	DOI	In-degree	BI, $q=0,1\%$	PI, $q=0,1\%$	TI, $q=0,1\%$
1. MDS clinical diagnostic criteria for Parkinson's disease	10.1002/MDS.26424	0,005	0,6307	0,9404	0,5254
2. Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease	10.1016/J.CELL.2016.11.018	0,0016	0,0206	0,000096	0,0075
3. Epidemiology of Parkinson's disease.	10.1007/S00702-017-1686-Y	0,00156	0,0188	0,000093	0,0068
4. MDS research criteria for prodromal Parkinson's disease	10.1002/MDS.26431	0,00152	0,0175	0,000091	0,0064
5. The epidemiology of Parkinson's disease: risk factors and prevention	10.1016/S1474-4422(16)30230-7	0,0015	0,01704	0,00009	0,0062

Топ-5 статей по индексу TI, $q = 1\%$

Название	DOI	In-degree	BI, $q=1\%$	PI, $q=1\%$	TI, $q=1\%$
1. Vagotomy and subsequent risk of Parkinson's disease	10.1002/ANA.24448	0,0009	0,0204	0,068	0,0298
2. The clinical symptoms of Parkinson's disease	10.1111/JNC.13691	0,00089	0,0202	0,0673	0,0295
3. Oxidative stress and Parkinson's disease	10.3389/FNANA.2015.00091	0,00087	0,0183	0,061	0,0267
4. Neuroinflammation in Parkinson's disease and its potential as therapeutic target	10.1186/S40035-015-0042-0	0,00086	0,0177	0,0589	0,0258
5. Short chain fatty acids and gut microbiota differ between patients with Parkinson's disease and age-matched controls	10.1016/J.PARKRELDIS.2016.08.019	0,00082	0,0156	0,052	0,0228

Статьи с высоким значением индексов в 2021 году

Название	DOI	Год публикации	In-degree, ранг в 2015-2021	Индекс, ранг в 2021
1. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies	10.1016/S1474-4422(19)30320-5	2019	12	In-degree, 3
2. Diagnosis and Treatment of Parkinson Disease: A Review	10.1001/JAMA.2019.22360	2020	32	In-degree, 5
3. Extensive graft-derived dopaminergic innervation is maintained 24 years after transplantation in the degenerating parkinsonian brain	10.1073/PNAS.1605245113	2016	108	Eigenvector, 1
4. Personalized iPSC-Derived Dopamine Progenitor Cells for Parkinson's Disease	10.1056/NEJMOA1915872	2020	164	Eigenvector, 2
5. Pre-clinical study of induced pluripotent stem cell-derived dopaminergic progenitor cells for Parkinson's disease	10.1038/S41467-020-17165-W	2020	182	Eigenvector, 3
6. The future of stem cell therapies for Parkinson disease	10.1038/S41583-019-0257-7	2020	174	Eigenvector, 4



Сеть цитирований по журналам

Атрибуты сети:

- Jld1 – ID цитирующего журнала
- Jld2 – ID цитируемого журнала
- Y – год публикации журнала 1
- Weight – количество цитирований журнала 2 журналом 1 в год Y

Количество вершин в сети - 3292, количество ребер в сети 152203.

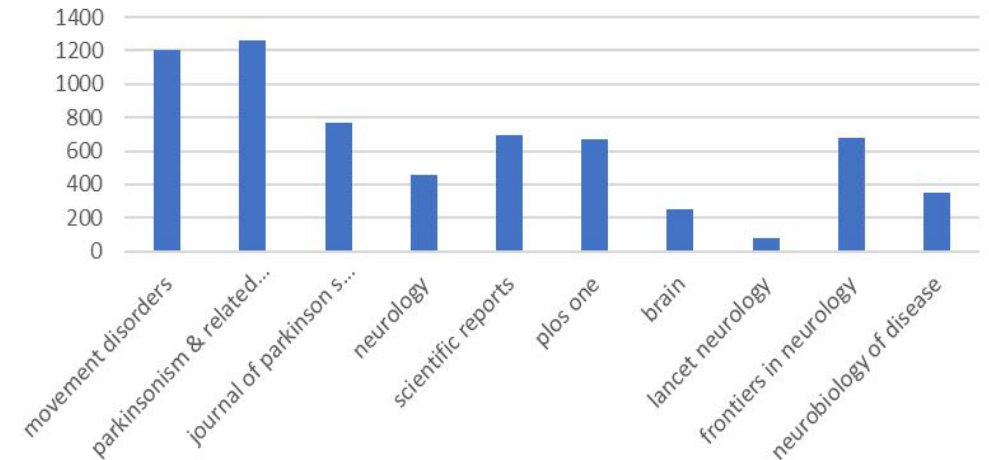
Jld1	Jld2	weight	Y
10623703	10623703	101	2021
10623703	163027424	150	2021
115201632	163027424	143	2020
115201632	163027424	134	2021
118428158	163027424	146	2020
118428158	163027424	127	2021
147691530	147691530	120	2017
147691530	147691530	247	2019
147691530	147691530	139	2020
147691530	147691530	195	2021
147691530	163027424	123	2016
147691530	163027424	115	2017
147691530	163027424	102	2018

Результаты для журналов

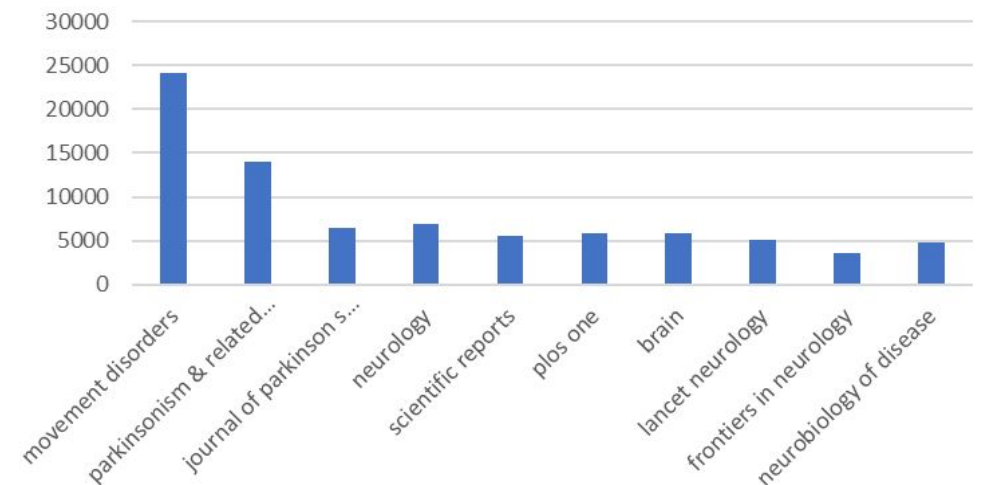
Name of journal	Number of citations
1. Movement Disorders	24122
2. Parkinsonism & Related Disorders	14035
3. Journal of Parkinson's Disease	6524
4. Neurology	6851
5. Scientific Reports	5613
6. PLOS One	5911
7. Brain	5909
8. Lancet Neurology	5085
9. Frontiers in Neurology	3641
10. Neurobiology of Disease	4757

Самые цитируемые журналы

Number of publications in top 10 cited journals



Citation count



Топ-10 журналов по индексу TI, $q = 0,5\%$

Название	In-degree	BI, $q=0,5\%$	PI, $q=0,5\%$	TI, $q=0,5\%$
1. Movement Disorders	0,0472	0,0276	0,0247	0,0332
2. Parkinsonism & Related Disorders	0,029	0,0248	0,0248	0,0262
3. Journal of Parkinson's Disease	0,0165	0,0168	0,0156	0,0163
4. PLOS one	0,014	0,0167	0,016	0,0156
5. Lancet Neurology	0,0137	0,0144	0,0145	0,0142
6. Scientific Reports	0,014	0,0129	0,0131	0,0134
7. Neurology	0,015	0,0105	0,0101	0,0119
8. Frontiers in Neuroscience	0,009	0,0114	0,014	0,0115
9. Frontiers in Neurology	0,011	0,0117	0,0113	0,01145
10. npj Parkinson's Disease	0,0087	0,011	0,0127	0,0108

Топ-10 журналов по индексу TI, $q = 5\%$

Название	In-degree	BI, $q=5\%$	PI, $q=5\%$	TI, $q=5\%$
1. Movement Disorders	0,0472	0	0	0,0157
2. Parkinsonism & Related Disorders	0,029	0	0	0,0097
3. Age and Ageing	0,00033	0,008	0,0134	0,0072
4. British Medical Bulletin	0,0003	0,0067	0,0117	0,0062
5. Inflammopharmacology	0,00033	0,0071	0,011	0,0061
6. Expert Opinion on Drug Metabolism & Toxicology	0,0003	0,0063	0,0105	0,0057
7. Journal of Parkinson's Disease	0,0165	0	0	0,0055
8. Cells	0,0027	0,0087	0,005	0,00547
9. Genes & Development	0,0003	0,0059	0,0095	0,0052
10. Neurology	0,015	0	0	0,005

Топ-10 журналов по индексу TI, $q = 10\%$

Название	In-degree	BI, $q=10\%$	PI, $q=10\%$	TI, $q=10\%$
1. Movement Disorders	0,0472	0	0	0,0157
2. Parkinsonism & Related Disorders	0,029	0	0	0,0097
3. Environmental Toxicology and Pharmacology	0,00017	0,0062	0,0103	0,0056
4. Springer plus	0,00017	0,0062	0,0103	0,0056
5. Journal of Parkinson's Disease	0,0165	0	0	0,0055
6. Neuroimmunomodulation	0,00018	0,00626	0,0096	0,0053
7. Gastroenterology Research and Practice	0,00018	0,00626	0,0096	0,0053
8. Neurology	0,015	0	0	0,005
9. Amino Acids	0,00017	0,0056	0,0086	0,0048
10. Scientific Reports	0,014	0	0	0,0047

Сеть цитирований по аффилиациям

Строится по аффилиации первого автора

Количество вершин	3029
Количество ребер	128602
Количество компонент связности	5
Размер наибольшей компоненты связности	3025
Плотность графа	0.0244
Минимальное количество цитирований	0
Максимальное количество цитирований	3757
Среднее количество цитирований	79.6

Топ-10 аффилиаций по In-degree

Название	In-degree	Betweenness ранг	Eigenvector ранг	Pagerank ранг
1. National Institutes of Health	0,0156	5	1	1
2. UCL Institute of Neurology	0,0134	3	4	2
3. University of Cambridge	0,0131	1	3	4
4. University of Oxford	0,0122	2	2	3
5. University College London	0,0120	6	5	5
6. Northwestern University	0,0114	14	6	6
7. Harvard University	0,0100	8	8	7
8. University of Pennsylvania	0,0085	24	7	8
9. Katholieke Universiteit Leuven	0,0080	26	10	12
10. Karolinska Institutet	0,0079	10	9	9

Сеть цитирований по авторам

Количество вершин	27551
Количество ребер	271623
Количество компонент связности	82
Размер наибольшей компоненты связности	27402
Плотность графа	0.0004
Минимальное количество цитирований	0
Максимальное количество цитирований	2416
Среднее количество цитирований	14.75

Топ-5 авторов по In-degree

Имя	Аффилиация	In-degree	Betweenness rank	Eigenvector rank	Pagerank rank
1. Ronald B. Postuma	Montreal General Hospital	0,0071	2	1	1
2. Anthony H.V. Schapira	University College London	0,0024	73	17	5
3. E. Ray Dorsey	University of Rochester	0,0023	19	62	3
4. Mike A Nalls	National Institutes of Health	0,002	207	10	10
5. Alberto J. Espay	University of Cincinnati	0,00185	6	12	24

Топ-5 авторов по TI $q = 0.1\%$

Имя	Аффилиация	In-degree	BI, $q = 0.1\%$	PI, $q = 0.1\%$	TI, $q = 0.1\%$
1.Ronald B. Postuma	Montreal General Hospital	0.0071	0,577	0,9999	0,528
2.E. Ray Dorsey	University of Rochester	0,0023	0,0251	0	0,0091
3.Anthony H.V. Schapira	University College London	0,0024	0,0245	0	0,009
4.Alicia M. Pickrell	National Institute of Neurological Disorders and Stroke (NINDS)	0,0018	0,0142	0	0,0053
5.Ole-Bjørn Tysnes	Haukeland University Hospital	0,00164	0,0122	0	0,0046



Топ-5 авторов по TI $q = 10\%$

Имя	Аффилиация	In-degree	BI, $q = 0.1\%$	PI, $q = 0.1\%$	TI, $q = 0.1\%$
1.Ronald B. Postuma	Montreal General Hospital	0,0071	0	0	0,0024
2.Thomas J. Hirschauer	Ohio State University	0,0001	0,0009	0,0015	0,0009
3.Yimeng Chen	Chinese Academy of Sciences	0,0001	0,0009	0,0015	0,0009
4.Tatsuya Sasaki	Okayama University	0,0001	0,0009	0,0015	0,0009
5.Michael Khalil	Medical University of Graz	0,0001	0,0009	0,0015	0,0009

Заключение

- Собрано почти 40 тысяч публикаций из более чем 3 тысяч журналов
- Подсчитаны классические индексы и недавно разработанные Bundle index и Pivotal index
- Работа представлена на конференции "The 12th International Conference on Network Analysis" в мае 2022 года
- Работа представлена на конференции "HCist 2022 - International Conference on Health and Social Care Information Systems and Technologies" в ноябре 2022 года

Результаты могут быть использованы исследователями БП для понимания ключевых областей исследований, их влияния на сообщество и для выявления новых направлений, которые раньше оставались без внимания.

Разработанная методика может применяться для выявления областей исследования для выгодных инвестиций.

Планы

1. Провести паттерн-анализ
2. Провести анализ аннотаций (семантическая близость публикаций)
3. Разработать модели по соавторам

Термины

Тема	Количество терминов	Количество статей
Биохимия, генетика, клетка	75	25037
Лекарства	175	9600
Симптом	56	25139
Диагноз	27	10975
Методы исследования	19	9505
ЦНС	40	21600

ИСТОЧНИКИ

1. Ruiz M. L., Benito-Le´on J. The top 50 most-cited articles in orthostatic tremor: A bibliometric review //Tremor and Other Hyperkinetic Movements. – 2019. – V. 9
2. Sorensen A. A., Weedon D. Productivity and impact of the top 100 cited Parkinson’s disease investigators since 1985 //Journal of Parkinson’s disease. – 2011. – V. 1. – No. 1. – C. 3-13
3. Li, T., Ho, Y. S., & Li, C. Y. (2008). Bibliometric analysis on global Parkinson's disease research trends during 1991–2006. *Neuroscience letters*, 441(3), 248-2q
4. Ponce, F. A., Lozano, A. M. (2011). The most cited works in Parkinson's disease. *Movement Disorders*, 26(3), 380-390.
5. Aleskerov F., Khutorskaya O., Buldyaev A., Yamilov A. Parkinson's disease: Network analysis of publications' impact //2018 7th International Conference on Computers Communications and Control (ICCCC). – IEEE, 2018. – C. 82-85.
6. Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of mathematical sociology*, 2(1), 113-120.
7. Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35-41.
8. 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.
9. Aleskerov F., Yakuba V. (2020). Matrix-vector approach to construct generalized centrality indices in networks. // SSRN 3597948.
Available at: <https://ssrn.com/abstract=3597948>



Спасибо за внимание!