

Инструкция по работе в VOSviewer с данными из программы OpenAlex

Подготовлено: Полина Ермолаева, Ирина Павлова, Дарья Мальцева, МЛ прикладного сетевого анализа НИУ ВШЭ

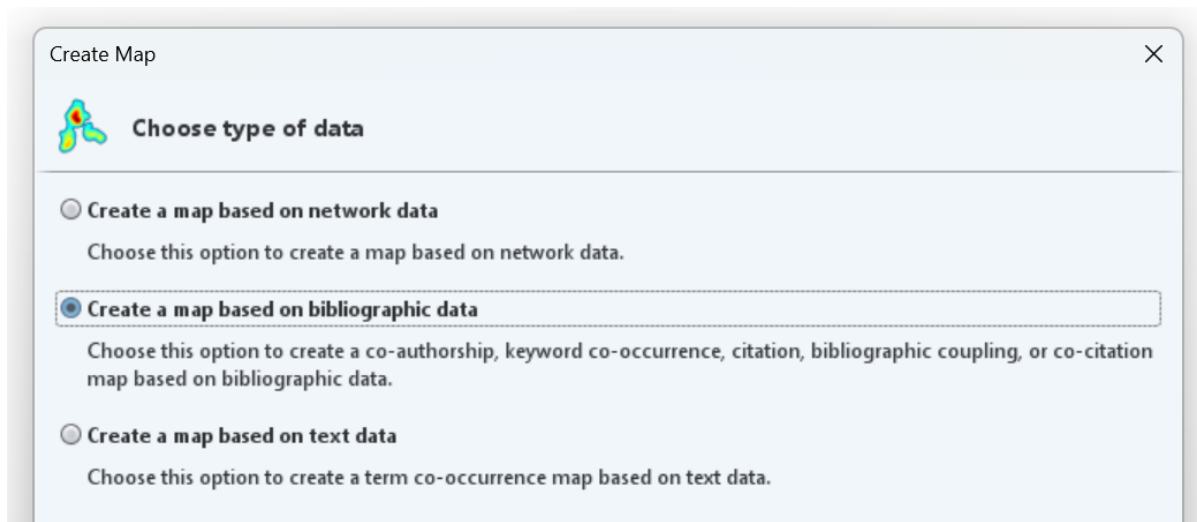
VOSviewer: Общая информация

VOSviewer — это программное обеспечение для визуализации библиометрических данных, позволяющее создавать карты научных областей, наглядно отображающие взаимосвязи между публикациями, авторами, ключевыми словами и журналами. Эти карты, называемые "карты соавторства" (co-authorship maps), "карты цитирования" (citation maps), или "карты ключевых слов" (keyword maps), помогают исследователям понять структуру и развитие научной области.

Типы анализа в VOSviewer:

- Соприсутствие ключевых слов (Keyword co-occurrence): Показывает связи между ключевыми словами, которые часто используются в одних и тех же публикациях.
- Цитирование (Citation): Показывает связи между публикациями на основе цитирования.
- Соавторство (Co-authorship): Показывает связи между авторами на основе совместных публикаций.
- Библиографическое сопряжение (Bibliographic coupling): Показывает связи между публикациями, которые цитируют одни и те же статьи.

Источниками информации для библиометрического анализа выступают библиографические базы данных (однако обратите внимание, что программа также позволяет создавать карты на основании стандартных сетевых данных и на основе текстовых источников).



Библиографические данные могут быть прочитаны разными способами. Их можно скачать через API, открыть из имеющегося файла или программы-менеджера по работе с литературой (например, Zotero).



Мы будем сети соприсутствия ключевых слов и цитирования публикаций на основе данных из базы OpenAlex, которые мы будем собирать через API.

Установить ПО можно здесь: <https://www.vosviewer.com/download>

Подробная инструкция по работе с VOSviewer:

https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.20.pdf

или: <https://www.vosviewer.com/getting-started>

OpenAlex как источник библиографической информации

Общая информация

[OpenAlex](#) — это открытый библиографический каталог, предоставляющий доступ к обширной базе данных научных публикаций, исследователей и институтов. Он представляет собой единую, связанную сеть информации, извлеченную из множества источников. Проект начал работу в 2022 году, но база данных постоянно обновляется, включая новые публикации и данные.

Основные преимущества базы:

- Открытый доступ: В отличие от коммерческих платформ, OpenAlex предоставляет бесплатный доступ к своим данным, что делает его ценным инструментом для исследователей с ограниченным бюджетом.
- Обширная база данных: OpenAlex охватывает огромное количество научных публикаций, исследователей и институтов, интегрируя данные из различных источников. Это обеспечивает более полное представление о научном ландшафте.

“OpenAlex, как и Google Scholar, не ведет экспертный отбор источников. Приоритет отдан не селективности, а охвату, причем препринты, не прошедшие рецензирование, индексируются вместе с остальными типами публикаций. Это следует учитывать при трактовке числа публикаций и цитирований по данной базе”.

- Связанные данные: OpenAlex не просто предоставляет списки публикаций; он связывает данные, создавая сеть взаимосвязей между исследователями, их публикациями, институтами и темами исследований. Это позволяет проводить более комплексный анализ. Например, можно отследить коллаборации между исследователями или развитие конкретных тем исследований.
- Подробные метаданные: OpenAlex предоставляет подробные метаданные о публикациях, включая информацию об авторах, аффилиациях, ключевых словах и тематических областях.

- API и инструменты: OpenAlex предлагает простой в использовании API, что позволяет легко интегрировать его данные в другие инструменты для анализа (например, VOSViewer, как в нашем случае).

Потенциальные недостатки:

- Неполный охват: Хотя OpenAlex стремится к всеобъемлющему охвату, он всё ещё не включает все научные публикации и исследователей в мире. Особенно это касается публикаций из менее распространенных языков или издательств, которые не так хорошо индексируются (например, многие публикации на русском языке не представлены в базе).
- Качество данных: Поскольку OpenAlex агрегирует данные из различных источников, качество данных может быть неоднородным. Могут присутствовать ошибки, неточности или неполные записи. Особенно заметны ошибки в информации, касающейся исследователей. Так, OpenAlex не всегда различает однофамильцев и тёзок, присваивая работы одних исследователей другим или объединяя их в одного человека. Поэтому для анализа сетей соавторства/социтирования по документам мы советуем проводить тщательную проверку и очистку данных или воспользоваться другими базами данных (например, Dimensions).

Более подробно с информацией о базе можно ознакомиться здесь:

<https://docs.openalex.org/>

и здесь: <https://sciguide.hse.ru/sources/openalex/>

Инструкция по работе с OpenAlex

OpenAlex предоставляет интуитивный интерфейс для поиска.

1. Основные сущности поиска:

Поиск в OpenAlex можно проводить по различным “сущностям”, основными из которых являются:

- Ключевые слова: Позволяют найти работы, содержащие заданные слова в названии, аннотации (abstract) или полном тексте. Поиск по умолчанию осуществляется по полному тексту, что может привести к получению нерелевантных результатов. Рекомендуется использовать фильтры для уточнения области поиска.
- Авторы: Поиск по авторам позволяет найти все публикации указанного исследователя. Для этого необходимо знать имя автора или их ORCID.

2. Формирование поискового запроса:

Поисковый запрос строится на основе логических операторов и ключевых слов.

Логические операторы (обязательно заглавными буквами):

- AND: Возвращает результаты, содержащие все указанные слова.
- OR: Возвращает результаты, содержащие хотя бы одно из указанных ключевых слов.
- NOT: Исключает результаты, содержащие указанное ключевое слово.
- (): Скобки используются для группировки ключевых слов и задания приоритета операторов.

Пример: Найдем работы об этнической сегрегации в школах:

- *"ethnic segregation" AND school*: Запрос найдет все работы, в тексте которых упоминается *этническая сегрегация* (указываем в кавычках, потому что нас интересует точное словосочетание) и *школы*
- *"ethnic segregation" AND (school OR adolescent)*: Аналогичен предыдущему запросу, но включает статьи, в которых упоминаются *этническая сегрегация* и либо *школы*, либо *подростки* (или и те, и другие)

- *"ethnic segregation" AND (school OR adolescent) NOT migrant*: Этот запрос, в дополнение к предыдущему исключает упоминания слова "migrant".
- *"ethnic segregation" AND (school OR adolescent) NOT (migrant OR migration)*: Этот запрос найдет работы, содержащие "ethnic segregation" и "school" или "adolescent", но не содержащие "migrant" или "migration"

3. Выбор области поиска:

По умолчанию поиск осуществляется по полному тексту работы. Для повышения релевантности результатов можно использовать фильтры для указания области поиска:

- title: Поиск только в заголовках работ (самый строгий критерий)
- title&abstract: Поиск в заголовках и аннотациях.

4. Использование фильтров:

Для уточнения результатов поиска OpenAlex предоставляет множество фильтров. Их можно добавить, нажав на кнопку "+" под полем поискового запроса. Основные фильтры:

- Год публикации (publication year): Позволяет ограничить поиск определенным годом или диапазоном лет
- Тип источника (source type): Позволяет ограничить поиск определенным типом источников (например, журналы, книги, препринты)
- Язык публикации (language): Ограничивает поиск работами на определенном языке
- Количество цитирований (citation count): Позволяет выбрать работы с определенным количеством цитирований (например, ">100").
- Область исследования (field): Для более точной фильтрации по тематике

Для того, чтобы получить ссылку с API-запросом нажимаем на значок в правом углу и выбираем опцию Copy link to share.

OpenAlex Works Search OpenAlex

Unsaved search

Show works where:

1 title & abstract includes ("ethnic segregation" OR "ethnic homophily" OR "race homophily" OR "race segregation") AND (school OR adole

Get QR code to share

Copy link to share

Works

Ethnic segregation of friendship networks in school: Testing a rational-choice argument of differences in ethnic homophily between classroom- and grade-level networks
2015 · Lars Leszczensky, Sebastian Pink · *Social Networks*
Cited by 114 PDF

Parallel Lives? Ethnic Segregation in Schools and Neighbourhoods
2005 · Simon Burgess, Deborah Wilson, et al. · *Urban Studies*
Cited by 212 PDF

Ethnic Segregation in Arizona Charter Schools
1999 · Casey D. Cobb, Gene V. Glass · *Education Policy Analysis Archives*
Cited by 185 PDF

School Choice and Ethnic Segregation
2003 · Sjoerd Karsten, Guuske Ledoux, et al. · *Educational Policy*

Stats

313 results

year

open access 37,1% 116

topic

inequality in Education: Impact of Socioeconomic Status Effects of Residential

institution

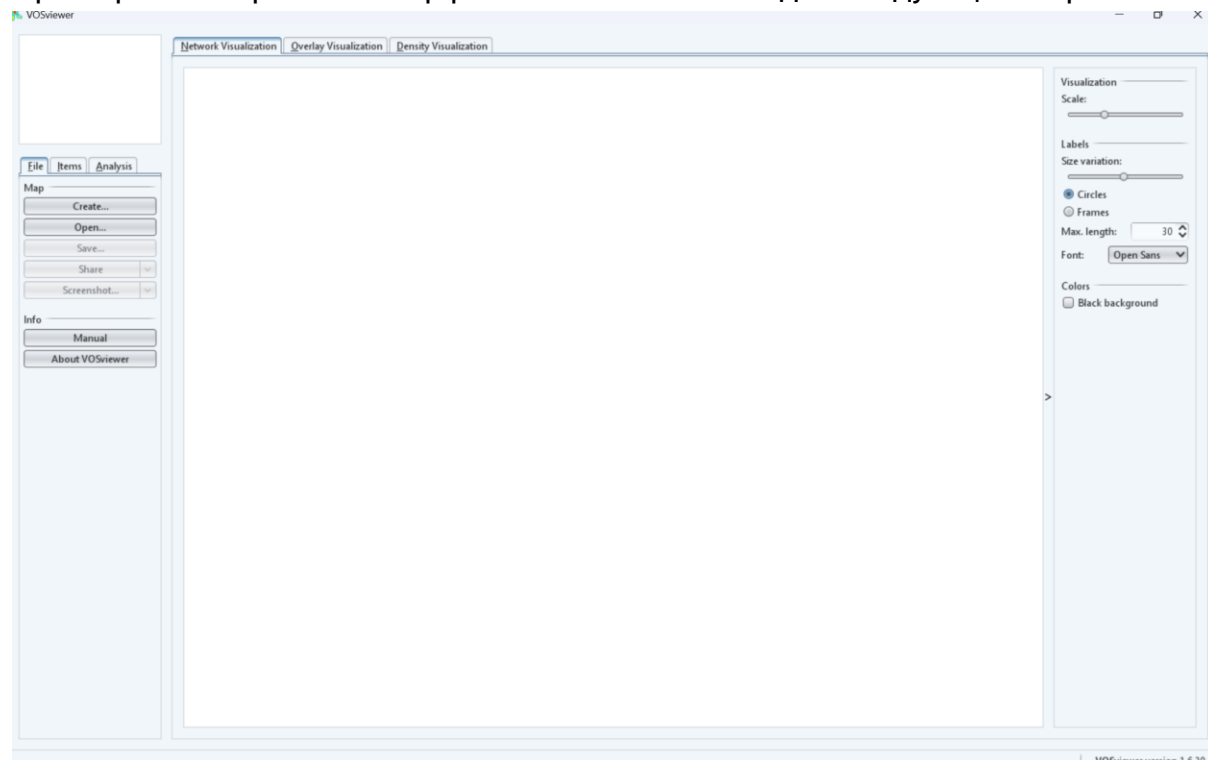
University of Bristol 15

University of London 7

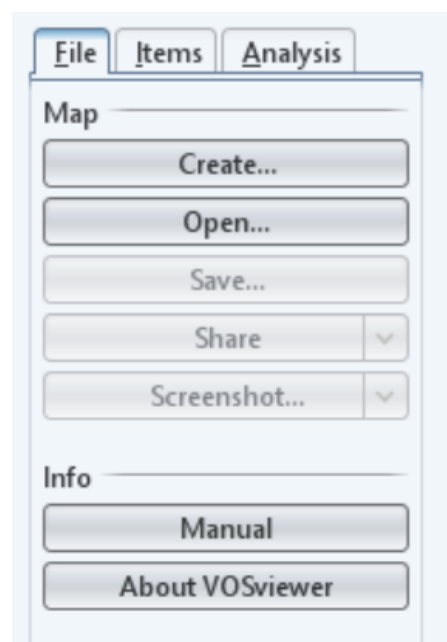
Подробнее про поиск в OpenAlex можно узнать здесь:
<https://www.youtube.com/watch?v=ao0jsdjLU2I>

Инструкция по работе с VOSviewer на данных OpenAlex

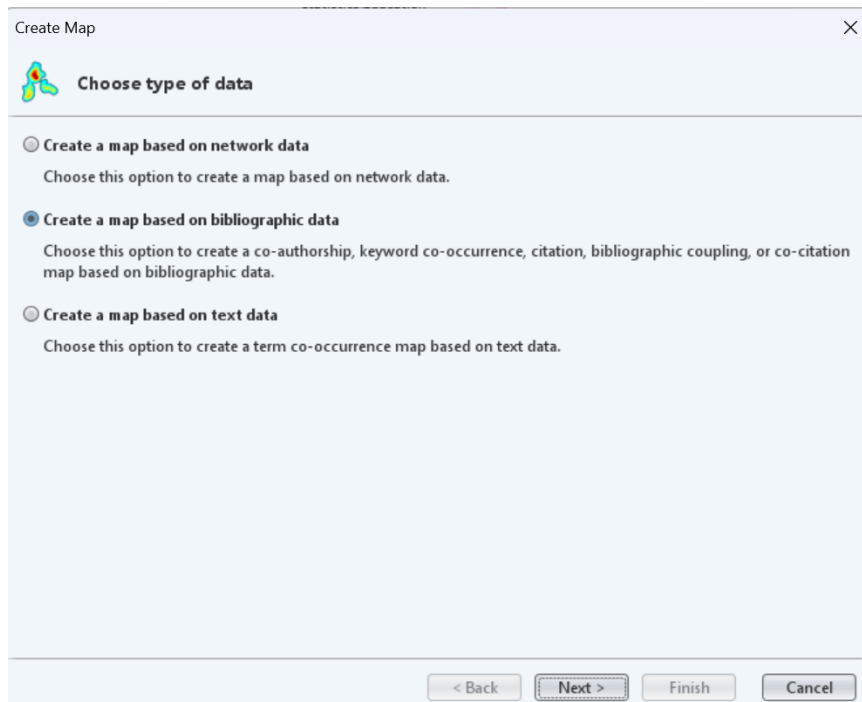
При первом открытии интерфейс VOSviewer выглядит следующим образом:



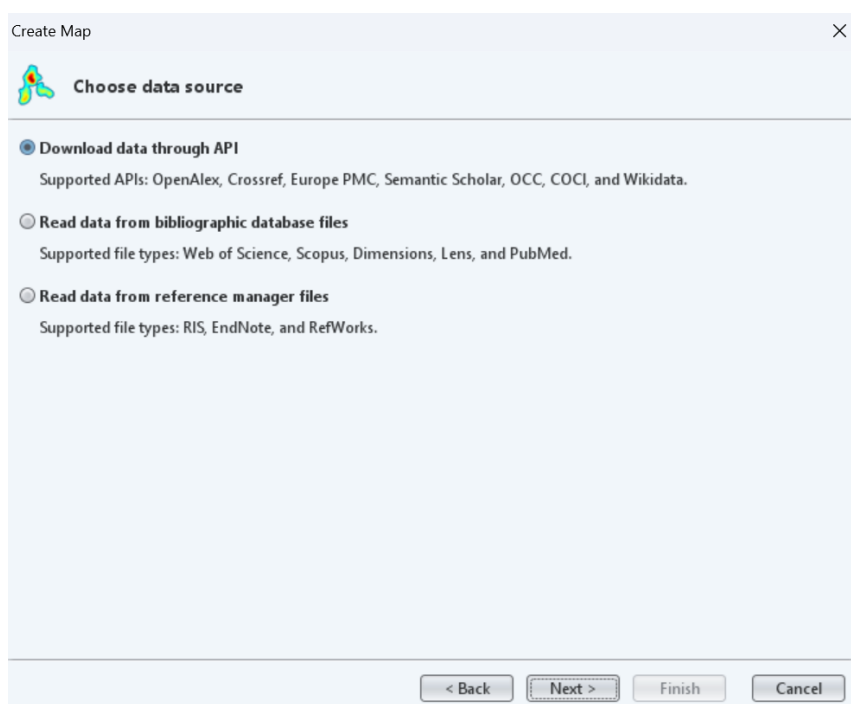
Панель, которая нас интересует, находится слева. Для того, чтобы создать новый проект нажимаем Create



Из предложенных опций выбираем Create a map based on bibliographic data, потому что именно с библиографическими данными мы будем работать.



Как уже было отмечено выше, OpenAlex предлагает простой в использовании API, что позволяет легко интегрировать его данные в другие инструменты для анализа. Так, возможность загрузить данные из OpenAlex была интегрирована в VOSviewer.



Ссылку с API-запросом, которую мы получили выше, необходимо вставить в соответствующее поле. Стоит отметить, что выгрузка данных с помощью API является наиболее удобной и простой в работе с OpenAlex в VOSviewer и мы советуем придерживаться ее.

Create Map

Specify API request or search query, or select files

API Request Search Query DOI JSON

API:
OpenAlex

Request URL (required):
https://openalex.org/works?page=1&filter=title_and_abstract.search%3A%28%22ethnic%20segregation%22%20OR%20%22ethnic%20homophily%22%20OR%20%22race%20homophily%22%20OR%20%22race%20segregation%22%20AND%20%28school%20OR%20adolescent%29%20

Example: <https://api.openalex.org/works?search=vosviewer&filter=type:article,language:en>
See the [OpenAlex API documentation](#) for more information.

JSON file (optional):

< Back Next > Finish Cancel

Есть, однако и другие. Например, поисковой запрос можно сформулировать с самом VOSviewer, но это занимает больше времени и ограничивает нас в количестве фильтров.

Create Map

Specify API request or search query, or select files

API Request Search Query DOI JSON

API:
OpenAlex

Author ID: ORCID: Affiliation: ROR ID: Source ID: ISSN: Text: "ethnic segregation" Year: to

☒ Title ☐ Abstract ☐ Full text

JSON file (optional):

< Back Next > Finish Cancel

Построение карты соприсутствия концептов (co-occurrence map)

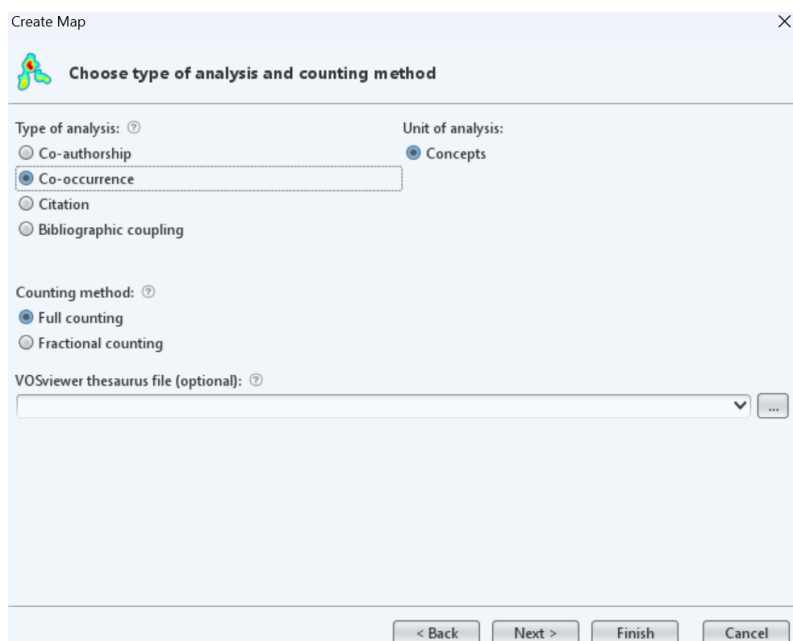
Поработаем с данными, полученными по запросу ("ethnic segregation" OR "ethnic homophily" OR "race homophily" OR "race segregation") AND (school OR adolescent)

API:

https://openalex.org/works?page=1&filter=title_and_abstract.search%3A%28%22ethnic%20segregation%22%20OR%20%22ethnic%20homophily%22%20OR%20%22race%20homophily%22%20OR%20%22race%20segregation%22%29%20AND%20%28school%20OR%20adolescent%29

Чтобы построить карту соприсутствия, выбираем соответствующий тип анализа.

В разделе counting method предпочтительнее выбирать вариант «fractional counting», чтобы избежать избыточной репрезентации работ, имеющих много ключевых слов (в данном случае «вклад» каждого слова равен не 1, а 1/количество слов, указанных для работы).



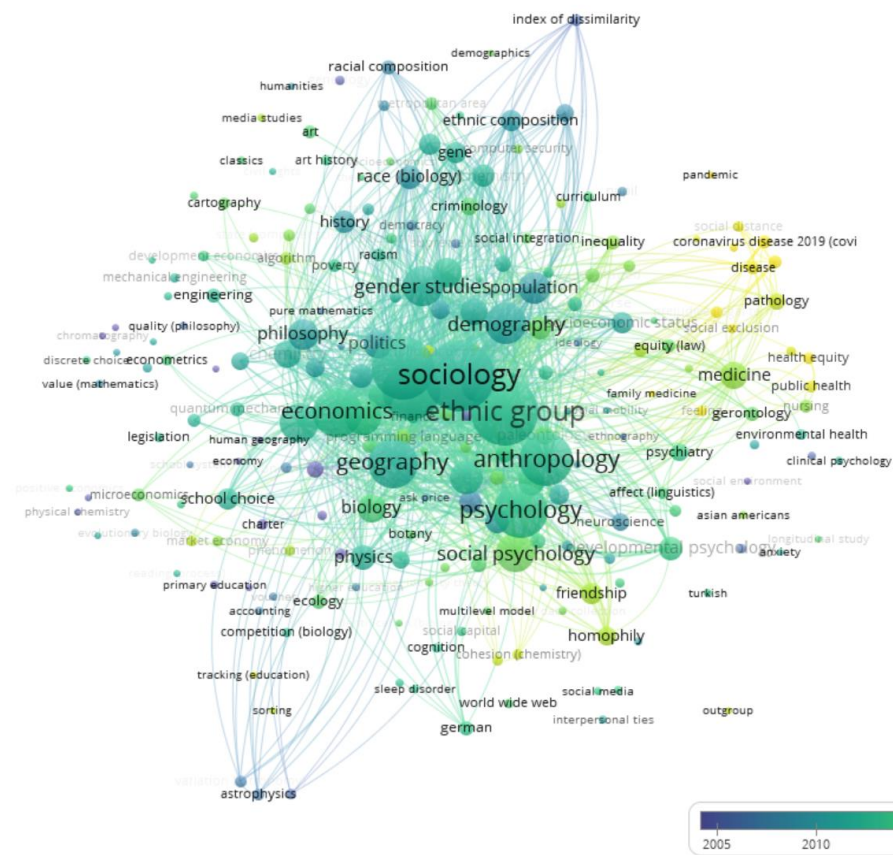
Определяем минимальное количество упоминаний для концепта (например, 3) и нажимаем Finish.

автоматически на основе принципа visualization of similarities (VOS clustering) и раскрасил разными цветами. На концепты, которые попали в тот или иной кластер, можно посмотреть во вкладке Items слева.

Например, социологические концепты попали в красный и фиолетовый кластеры, а синий левый кластер включает в себя медицинские исследования и исследования о неравенстве доступа к медицине различных этнических групп.

- Центральные узлы: Узлы с большим количеством связей, представляющие наиболее влиятельные ключевые слова в данной области. Например, ожидаемо, самыми популярными концептами являются Социология, демография, психология, антропология и пр. С другой стороны, мы наблюдаем, что медицинские исследования развиты хуже и находятся на периферии карты.
- Связи между узлами: Показывают взаимосвязи между разными темами исследования.

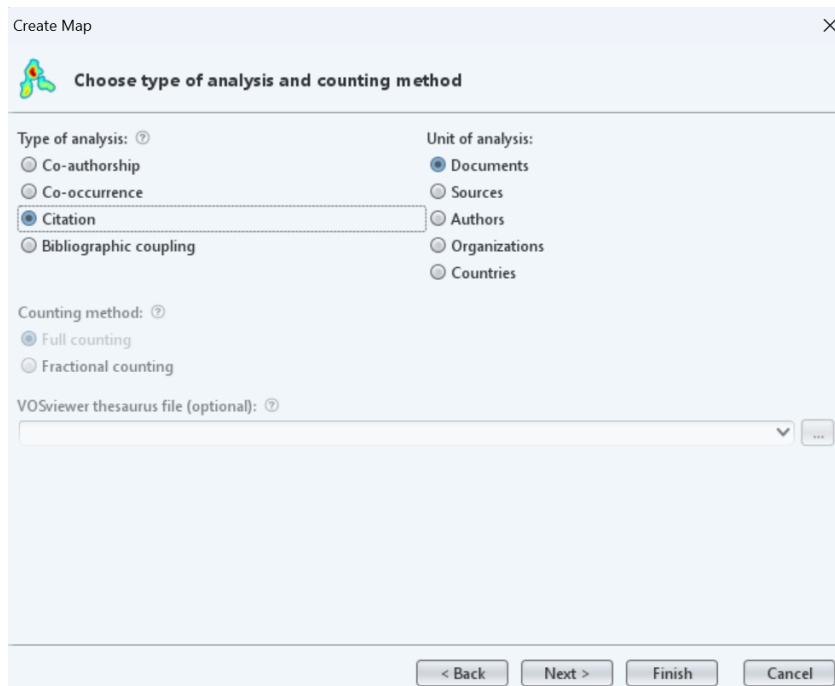
Дополнительно VOSviewer позволяет окрасить узлы на основе числовой переменной (например, год публикации). Это делается с помощью переключения на вкладку Overlay Visualization. В нашем случае для каждого концепта рассчитан средний год публикации. Мы видим, что концепты, связанные с медициной, здоровьем и доступом к медицинской помощи начали развиваться в связке с этнической сегрегацией относительно недавно. Возможно, причиной такого развития стала пандемия COVID-19.



Построение карты цитирования документов (citation)

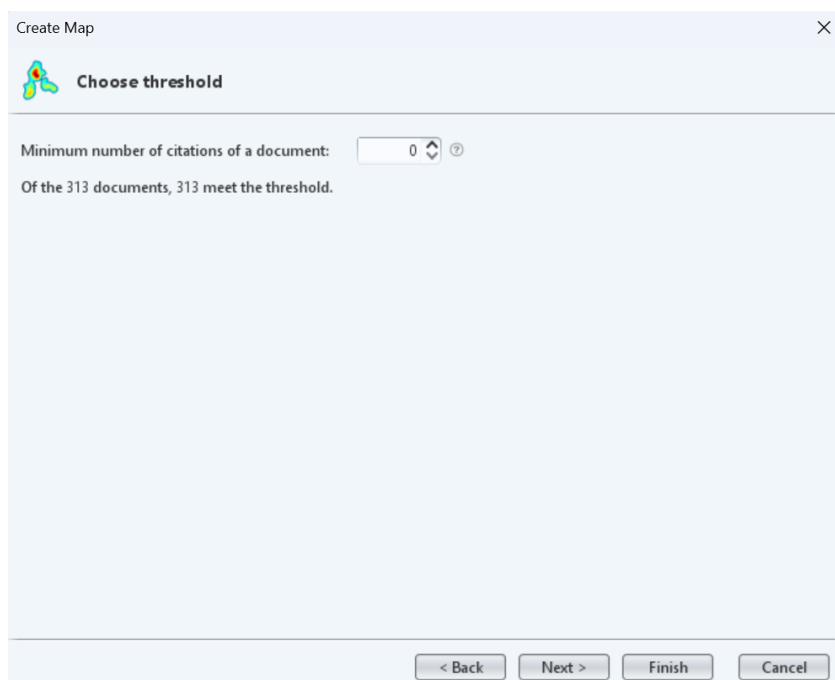
Продолжим работать с запросом (*"ethnic segregation" OR "ethnic homophily" OR "race homophily" OR "race segregation"*) AND (*school OR adolescent*).

Теперь построим карту цитирования документов. Выбираем Citation в типе анализа и Documents как единицы (units) анализа.



The screenshot shows the 'Create Map' dialog box with the title 'Choose type of analysis and counting method'. It contains two columns of radio button options. The left column, 'Type of analysis:', includes 'Co-authorship', 'Co-occurrence', 'Citation' (selected), and 'Bibliographic coupling'. The right column, 'Unit of analysis:', includes 'Documents' (selected), 'Sources', 'Authors', 'Organizations', and 'Countries'. Below these is a 'Counting method:' section with 'Full counting' (selected) and 'Fractional counting'. At the bottom is a 'VOSviewer thesaurus file (optional):' field with a dropdown arrow and a file selection button. Navigation buttons at the bottom are '< Back', 'Next >', 'Finish', and 'Cancel'.

Определим минимальное количество цитирований для работы. По умолчанию оно равно 0. Нажимаем Finish.

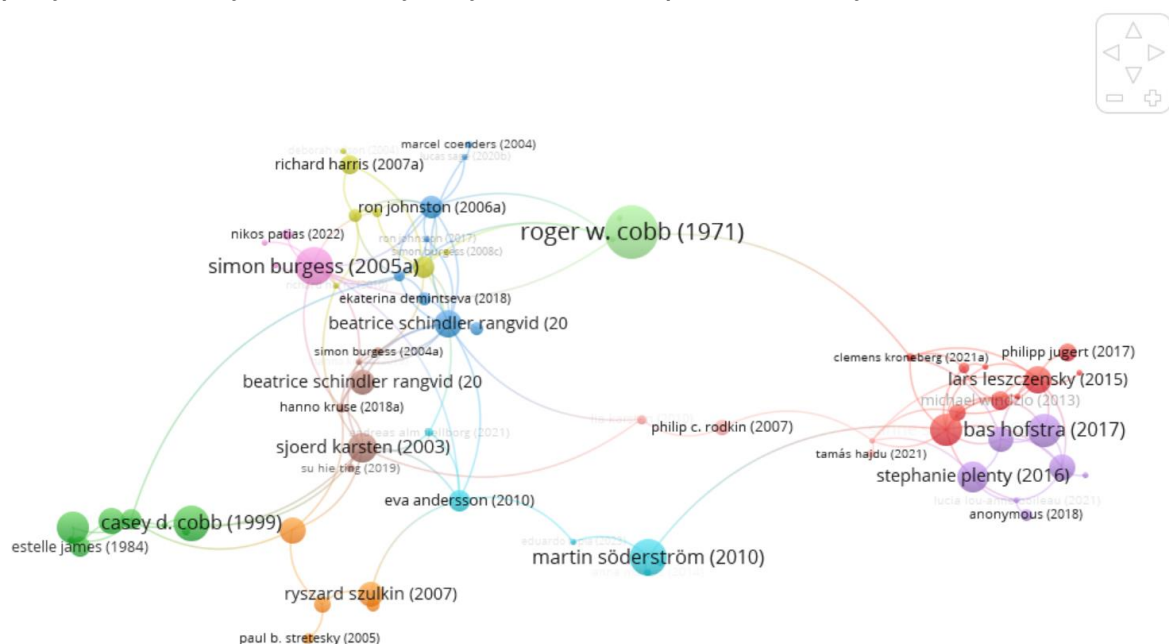


The screenshot shows the 'Create Map' dialog box with the title 'Choose threshold'. It features a label 'Minimum number of citations of a document:' followed by a text input field containing '0' and a help icon. Below this, it states 'Of the 313 documents, 313 meet the threshold.' Navigation buttons at the bottom are '< Back', 'Next >', 'Finish', and 'Cancel'.

Получаем следующее сообщение. Не все работы связаны друг с другом. Это логично. Чтобы не отвлекаться на работ, которые не цитируют друг друга, стоит нажать Yes. Вы, однако, можете попробовать поработать и с иным вариантом.



В результате получаем следующую сеть цитирования документов:

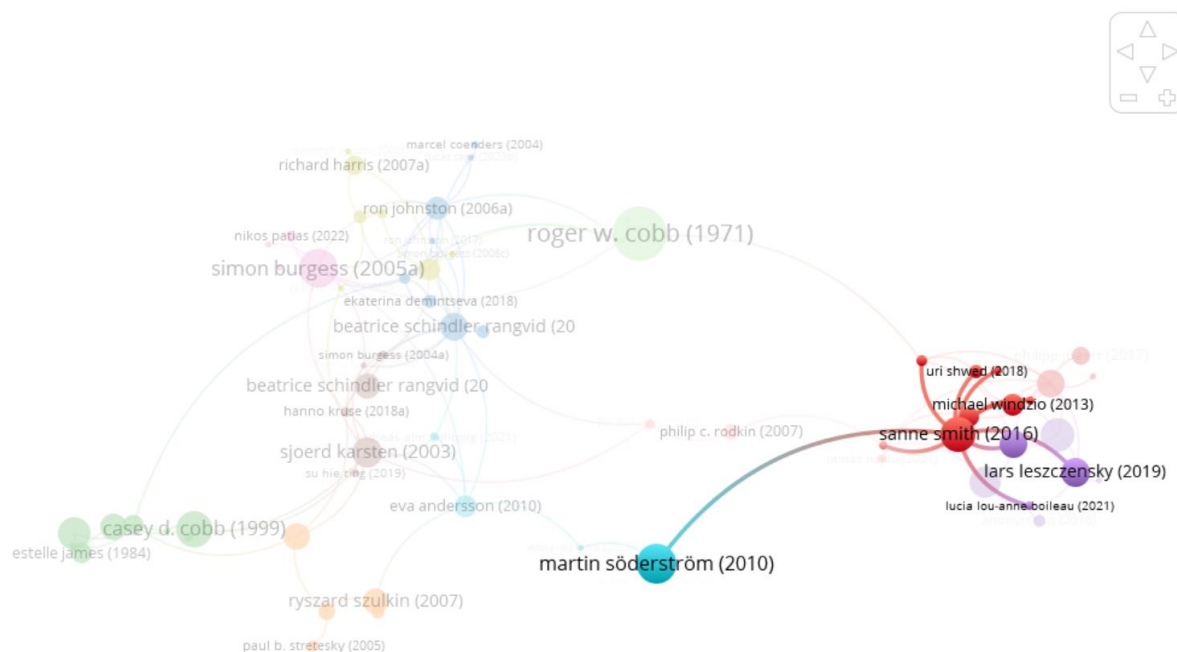


На карте можно выделить:

- Узлы: Узлами на карте являются работы. Размер узла задается количеством цитирований работы. Чем больше узел - тем более цитируемой является работа.
- Связи между узлами: Цитирование. Например, видим, что работа авторства Eva Andersson 2010 года цитирует работу Ryszard Szulkin 2007 года.

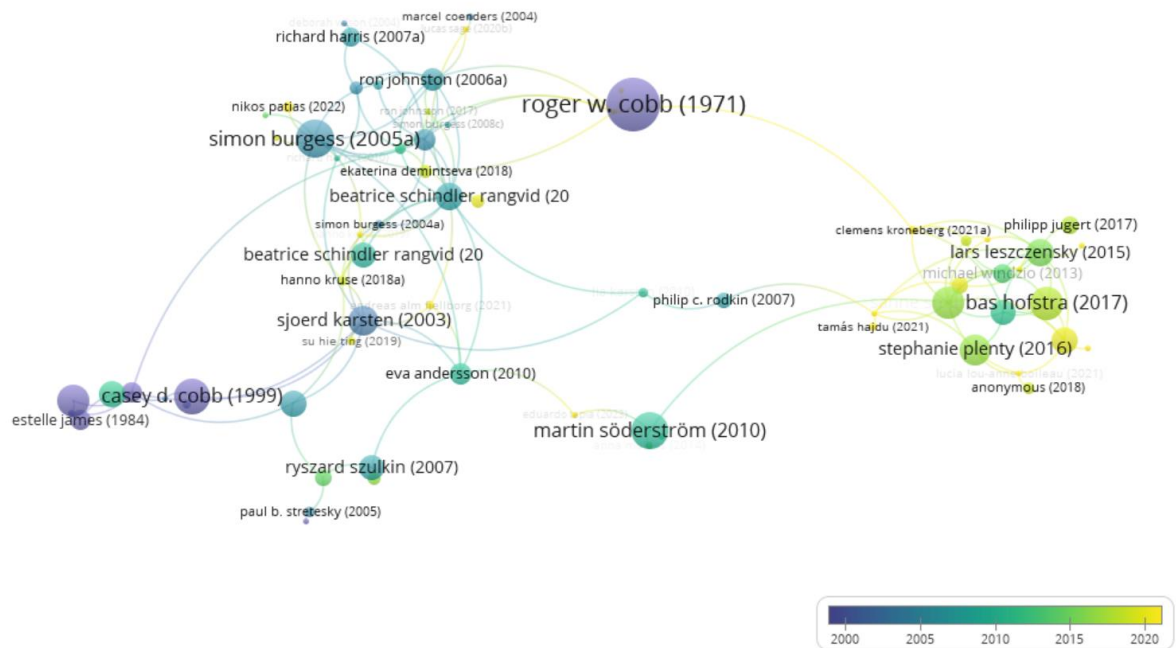
- Кластеры: Группы тесно связанных узлов, представляющие отдельные направления исследований. VOSviewer определил их автоматически на основе принципа visualization of similarities (VOS clustering) и раскрасил разными цветами. Зачастую работы, которые цитируют друг друга, посвящены схожей тематике и для ее определения можно обратиться к названию работы. Для этого нужно навести курсор на узел (выходные данные о публикации появятся внизу страницы).

Например, многие работы в правом кластере посвящены влиянию контекстуальных факторов (например, этнического состава или плотности класса) на этническую сегрегацию.



Authors:	sanne smith; daniel a. mcfarland; frank van tubergen; ineke maas
Title:	ethnic composition and friendship segregation: differential effects for adolescent natives and immigrants
Source:	american journal of sociology, 121(4), 1223-1272
Year:	2016

Мы также опять можем использовать Overlay Visualization. Теперь цветом отмечен год публикации работы и мы видим, что еще одной характеристикой правого кластера является относительная “свежесть” этих работ.



Возможности анализа данных в VOSviewer

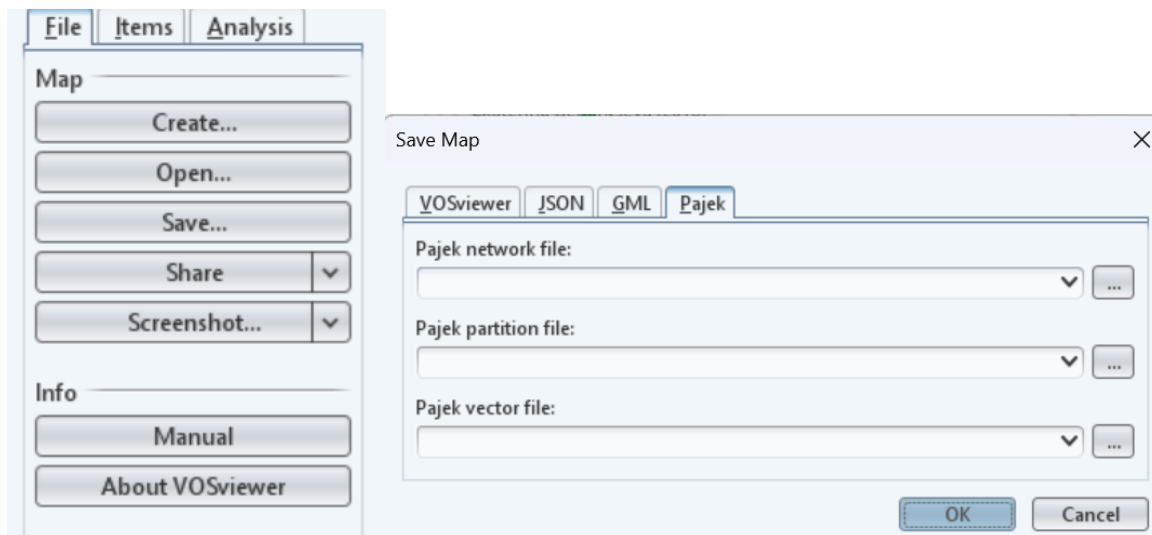
Вкладку Анализ можно использовать для обновления отображения и кластеризации текущей активной карты. Это делается с помощью техник VOS layout technique и VOS clustering.

- Normalization: Используйте раскрывающийся список методов нормализации, чтобы определить, как нормализуется сила связей между узлами.
- Layout: Attraction и Repulsion. Эти параметры влияют на способ расположения узлов на карте с помощью метода VOS layout technique. Параметр Attraction должен иметь целое значение от -9 до +10. Параметр Repulsion должен иметь целое значение от -10 до +9. Значение параметра Repulsion должно быть ниже значения параметра Attraction. Можно также использовать предзаданные параметры.
- Update Layout: При внесении изменений обновите визуализацию.
- Clustering: В этом разделе можно поменять параметры для выделения кластеров методом VOS clustering – изменить параметр resolution (определяющий уровень детализации кластеризации), определить минимальный размер кластеров (Min. cluster size) и изменить ряд продвинутых параметров (для понимания вам нужно подробнее познакомиться с методами выделения сообществ – community detection – и, в частности, методом VOS clustering).
- Update Clustering: При внесении изменений обновите кластеризацию.
- Rotate/flip: возможность отобразить визуализацию по вертикали или горизонтали.

Обратите также внимание на возможности во вкладке Visualization.

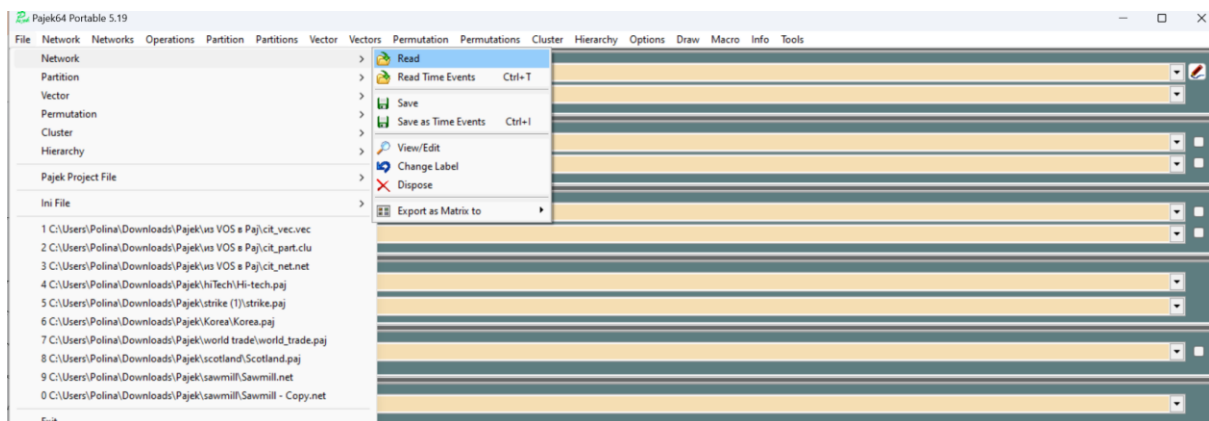
Сохранение файлов и открытие в Pajek

VOSviewer позволяет быстро и удобно сохранять файлы в других форматах. Для этого выбираем опцию Save на панели слева. Форматы, которые доступны для сохранения, перечислены наверху всплывающего окна. Нас интересует Pajek, выбираем его.



На компьютере выбираем папку, в которой сохраним три файла для загрузки в Pajek - network, partition, vector и нажимаем OK.

Переходим в Pajek и открываем файл там. Делаем это с помощью File -> Network/Partition/Vector -> Read.



После этого можем работать с файлами как обычно. Построить визуализацию, например (Draw -> Network+First Partition; Layout -> Energy -> Kamada-Kawai -> Separate Components).

