

Мастер-класс по работе с данными социальных медиа

Материалы подготовлены Анной Карташевой, Тамарой Щегловой, Ириной Павловой и Дарьей Мальцевой, МЛ прикладного сетевого анализа

В рамках этого семинара вы получите навыки по работе в программе Gephi с реальными данными:

- проанализируете представленность НИУ ВШЭ в медиа на основе дайджестов мониторинга СМИ;
- используете онлайн ресурс для автоматической подготовки текстовых документов в нужном формате для Gephi;
- поработаете в командах;
- проведете анализ наборов сетевых данных;
- презентуете результат своего анализа.

(1) Начало работы – pdf файлы с мониторингом СМИ

Набор данных: [ссылка](#)

(2) Перевод текстовых файлов в сетевой формат

На этом семинаре вы можете работать в командах. Возьмите pdf-файлы мониторинга пресс-службы НИУ ВШЭ «Информационная картина дня» и сделайте анализ о том, как именно упоминается в СМИ деятельность сотрудников Вышки.

- Альтернативный вариант № 1: можно раздать разные дни (или недели) разным студентам (или командам).
- Альтернативный вариант № 2: можно раздать одинаковый набор данных разным студентам (или командам) – и посмотреть лучший вариант анализа.


Откройте сайт **nocodefunctions.com** и загрузите pdf-файлы мониторинга СМИ ВШЭ:

https://nocodefunctions.com/cowo/semantic_networks_tool.html


русский (Россия)

Nocode functions


Menu

 **Бесплатный инструмент для преобразования ваших текстов в семантические графики**

Используйте его напрямую, выбрав данные для анализа:

Обычный текст (файлы txt или pdf). Поддерживает очень большое количество файлов. 

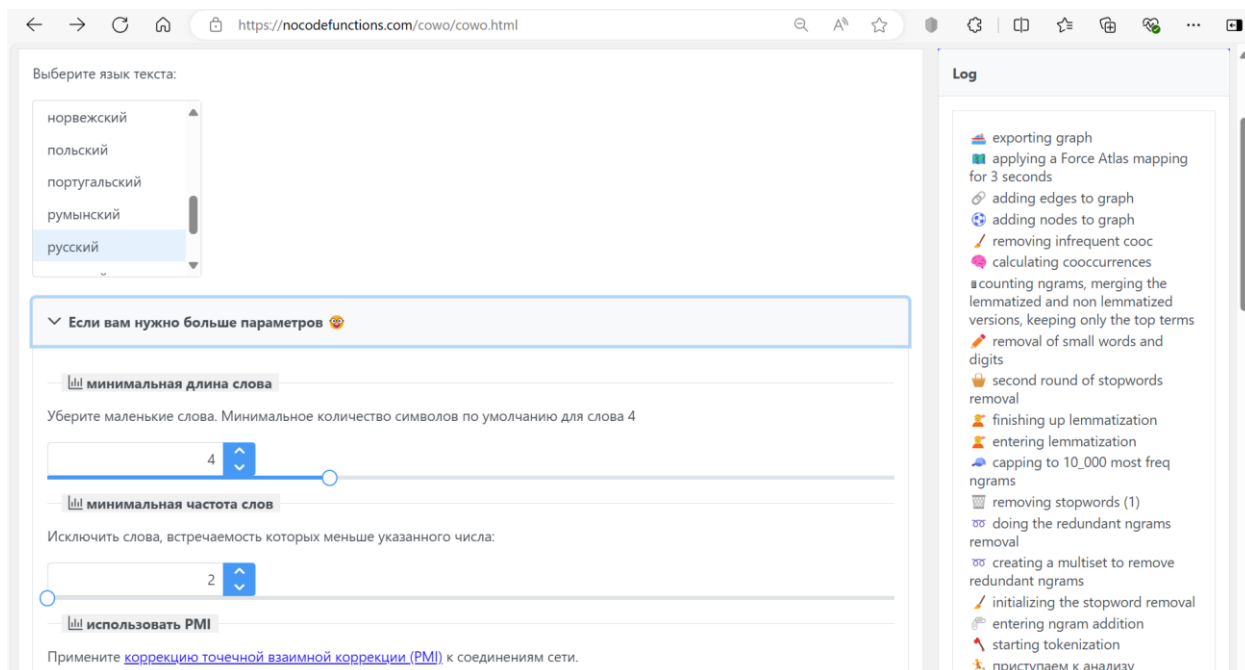
или

Структурированный текст в столбцах или таблицах (файлы csv или Excel). 

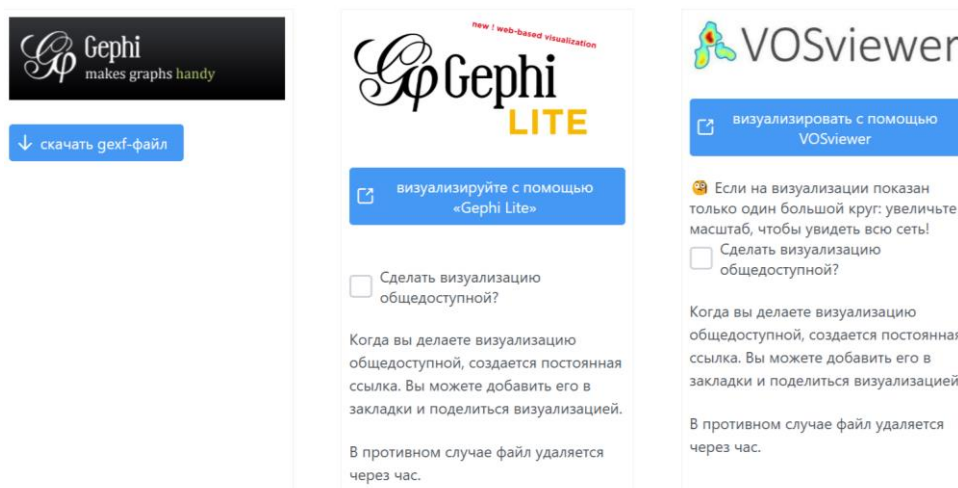
На этом сайте текстовые файлы автоматически конвертируются в сеть соприсутствия ключевых слов.

После загрузки файлов выберите язык текста и нужные вам параметры.

В окошке Log будет записано, какие шаги по анализу текстов делает программа.



Далее воспользуйтесь опциями по визуализации, доступными на сайте – через Gephi Lite или VOSviewer:



Скачайте файлы в сетевом формате с расширением gexf, который читается в программе Gephi. GEXF помечен значком программы Gephi:

Network data

Co-occurrence network:



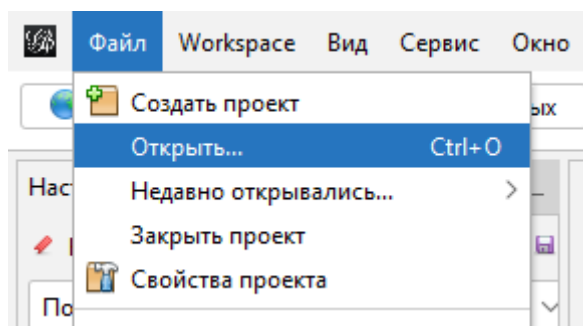
(3) Анализ полученных файлов в Gephi

Полученные данные откройте и проанализируйте в Gephi.

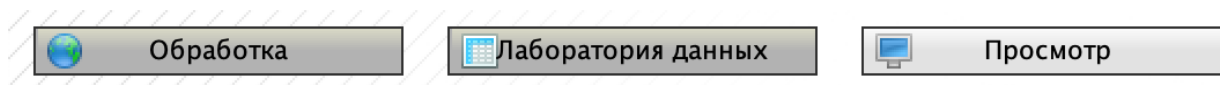
При открытии программы вас встречает приветственное окошко, через которое можно либо «Создать проект», либо «Открыть файл с графом».



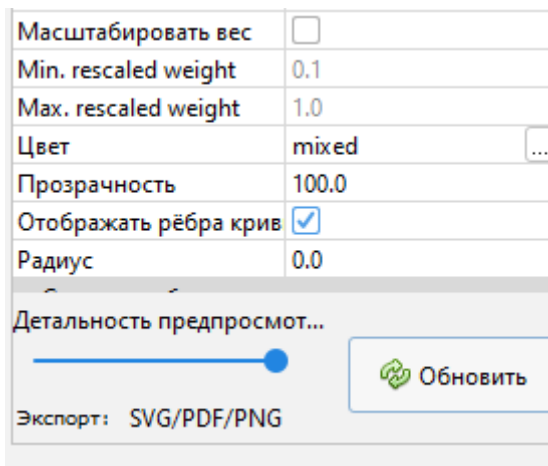
Или откройте файл через меню «Файл»



Для начала обратимся к вкладке Просмотр: все вкладки подписаны в самом верху.



И на этой вкладке мы, вероятно, ничего не увидим. Нужно нажать «Обновить» в Настройке предпросмотра.



Чтобы сделать наш граф наглядным и понятным, нужно будет настроить цвет и размер узлов. Для этого заходим во вкладку Обработка.

Чтобы раскрасить граф: Appearance - Nodes - значок палитры.

По умолчанию все узлы и ребра раскрашены одним цветом (Unique).

Первый вариант раскраски:

- разными цветами по одному из признаков (Partition).

Второй вариант раскраски:

- одним цветом, но будет отличаться интенсивность и яркость: по признакам «количество произнесенных слов» и «степень» (количество связей этого узла) (Ranking > Degree), чем больше значение признака, тем ярче цвет.

После каждого вносимого изменения нужно нажать «Применить».

Над словом «Применить» есть надпись Palette (возможность выбора цвета). По умолчанию в палитре 8 цветов, для генерации палитры нужно снять галочку с «Limit number of colours», а затем нажать Generate и ОК.

Экспериментируйте, но не забывайте главное правило работы с Gephi.

Главное правило при работе с Gephi: здесь нет кнопки «Назад», так что будьте осторожны с изменениями, чтобы потом не пришлось все переделывать с начала.

Чтобы включить подписи узлов, нажимаем на букву T на нижней панели под графом:



Укладка

Изначально, граф уложен случайным образом, все элементы внутри расположены хаотично. Используйте встроенные способы укладки в Gephi.

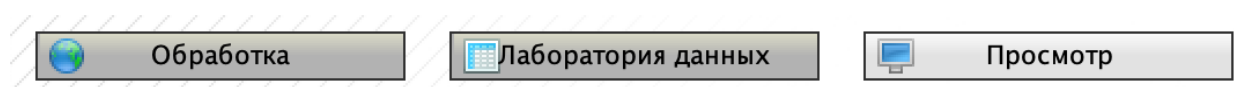
Сначала нажимайте «Пуск», а потом «Стоп», когда граф примет нужный вид.

В этом же меню можно выбрать «Укладка меток», чтобы названия узлов (метки) не наслаивались друг на друга.

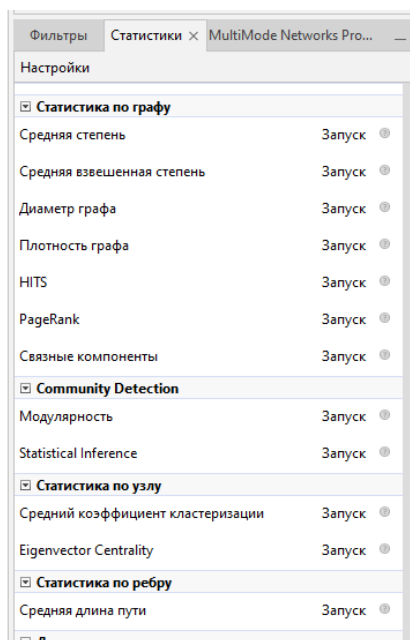
Размер узлов можно настроить во вкладке Appearance — Nodes, под значком совмещенных кругов. Можем задать единый размер для всех или проранжировать по какому-то признаку.

- Посчитайте метрики

Во вкладке «Лаборатория данных» вы можете посмотреть на все наши данные в табличном виде. А во вкладке «Обработка» вы можете посчитать статистики.



Для анализа статистик используется вкладка «Статистики»



Чтобы посмотреть уровень центральности нажимаем во вкладке «Appearance» - «Узлы» – «Ranking» – и выбираем разные типы центральности, например, «Betweenness Centrality»

Чтобы отразить среднюю степень на графе (показать, у кого больше число взаимодействий), перейдем во вкладку «Appearance» - «Ranking» и выбираем ранжирование по критерию «Degree»: После установки параметров нажимаем «Применить».

- Найдите сообщества

Разбиение по кластерам для наглядности можно дополнить цветовым делением:

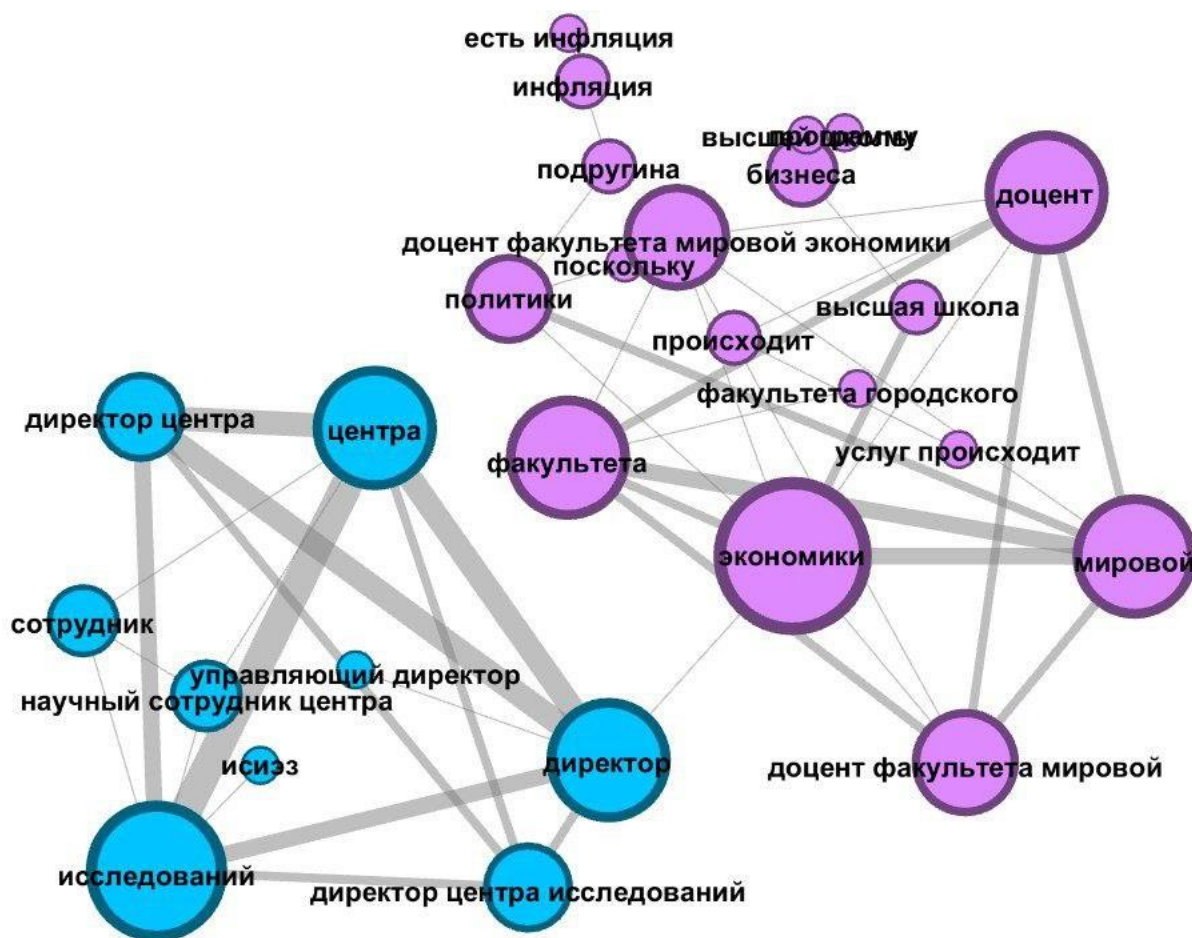
«Appearance» - выбираем значок «палитры» - «Partition» - «Modularity class» -

Применить

- Прокомментируйте проанализированные данные

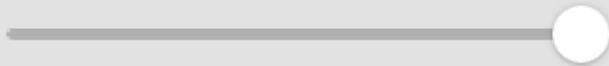
По большей части вы можете получить несвязанные группы, что тоже является результатом. Попробуйте найти в данных закономерности, интересные наблюдения.

На рисунке представлен пример расчёта метрик центральности и модулярности на основе файла мониторинга СМИ ВШЭ от 03.12.24.



Экспортировать наш граф мы можем в формате png, pdf или векторном svg. Экспорт графа осуществляется через самое нижнее окно во вкладке «Просмотр». Либо нажмите на кнопку «Экспорт: SVG/PDF/PNG» под кнопкой «Обновить», либо сохраните файл через «Файл» > Экспорт.

Детальность предпросмотра:



Обновить

Экспорт:

SVG/PDF/PNG

(4) Представление результатов

Составьте 2 слайда с полученными результатами

Прокомментируйте результаты анализа для всей аудитории.