

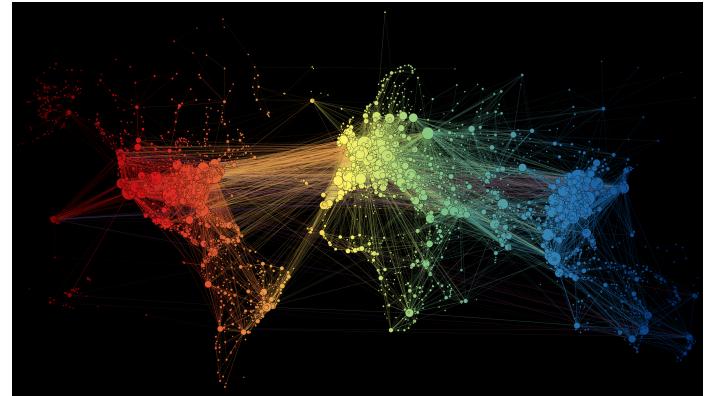


Анализ русской википедии

Меры центральности

Меры центральности определяют наиболее важные вершины графа. Приложения показателя применяются для выявления **наиболее влиятельного лица** в социальной сети, **ключевых узлов** инфраструктуры в интернете или городских сетей и разносчиков болезни. При анализе больших графов большое внимание уделяют различным центральностям.

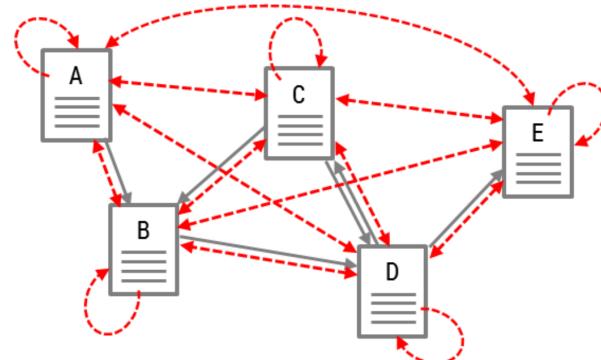
1. Degree centrality
2. Pagerank
3. Eigenvector centrality
4. Hubs and authorities



Сеть воздушного движения

Постановка задачи

- ◆ **Дано:** Набор из n страниц
- ◆ Каждая страница может содержать произвольное количество ссылок на другие страницы
- ◆ **Требуется:** Присвоить каждой странице score - «важность» (действительное неотрицательное число)



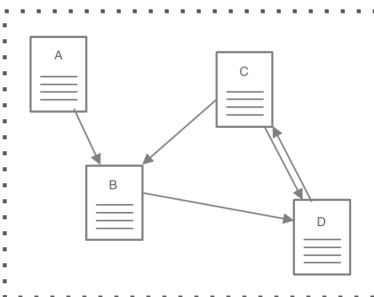
Degree centrality

Важность узла в графе можно анализировать разными способами. Самый простой — отсортировать участников **по количеству входящих ребер**. У кого больше — тот более важен.

Pagerank

В поиске по интернету одним из критериев для важности страниц является **PageRank**.

Он вычисляется при помощи случайного блуждания по графу, где в качестве узлов — страницы, а ребро между узлами — гиперссылка с одной страницы на другую. Случайный блуждатель двигается по графу и время от времени перемещается на случайный узел и начинает блуждание заново. PageRank равен доли пребывания на каком-то узле за все время блуждания. Чем он больше, тем узел важнее.



Sergey Brin and Larry Page, 1998

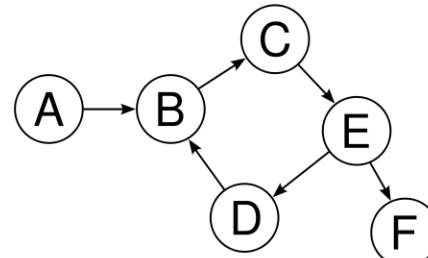
Формальное определение pagerank

Случайное блуждание по ориентированному графу:

$$p_i^{t+1} = \sum_{j \in N(i)} \frac{p_j^t}{d_j^{\text{out}}} = \sum_j \frac{A_{ji}}{d_j^{\text{out}}} p_j$$

$$D_{ii} = \text{diag}(d_i^{\text{out}})$$

$$p^{t+1} = (D^{-1} A)^T p^t$$



Степенной метод:

$$p^{t+1} \leftarrow (P = D^{-1} A) = P^T p^t$$

$$p^{t+1} \leftarrow \alpha P^T p^t + (1 - \alpha) \frac{e}{n}$$

α – коэффициент телепортации

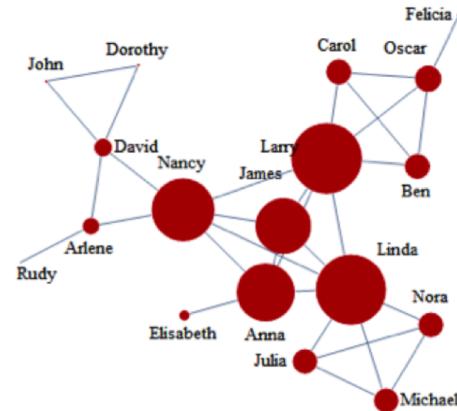
Eigenvector centrality (степень влиятельности)

Важность узла зависит от важности его соседей (рекурсивное определение)

$$v_i \leftarrow \sum_j A_{ij} v_j$$

$$v_i = \frac{1}{\lambda} \sum_j A_{ij} v_j$$

$$\mathbf{Av} = \lambda \mathbf{v}$$



В общем случае имеется много различных собственных значений, для которых существует ненулевой собственный вектор. Однако, из дополнительного требования, чтобы все элементы собственного вектора были неотрицательны, следует (по теореме Фробениуса — Перрона), что только **наибольшее собственное значение** приводит к желательной мере центральности

Теорема Перрона-Фробениуса

$$P = \{p_{ij}\}_{i,j}^t, p_{ij} \geq 0$$

- Существует λ_1 такое, что $|\lambda_j| < \lambda_1, j = 2, \dots, t$
- $\min_i \sum_{j=1}^t p_{ij} \leq \lambda_1 \leq \max_i \sum_{j=1}^t p_{ij}$
- Существует вектор $\pi = (\pi_1, \dots, \pi_t), \pi_i > 0$:

$$\pi P = \lambda_1 \pi$$

Hubs and authorities

Авторитетная страница - содержит полезную информацию.

Хаб-страница - это документ, содержащий ссылки на авторитетные документы.

- на хорошие авторитетные страницы переходят через хорошие хаб-страницы

$$a_i \leftarrow \sum_j A_{ji} h_j$$

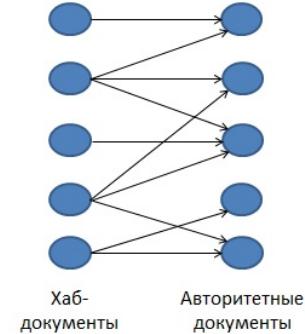
- хорошие хаб-страницы ссылаются на хорошие авторитетные страницы

$$h_i \leftarrow \sum_j A_{ij} a_j$$

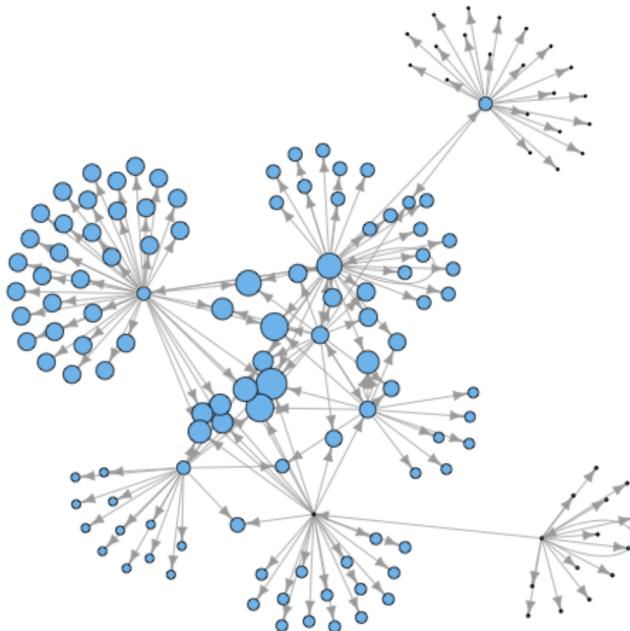
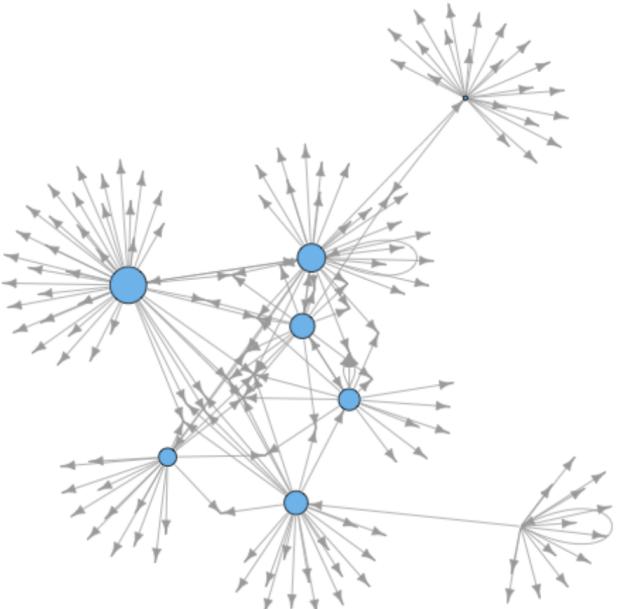
$$\begin{aligned}\mathbf{a} &= \alpha \mathbf{A}^T \mathbf{h} \\ \mathbf{h} &= \beta \mathbf{A} \mathbf{a}\end{aligned}$$

$$\begin{aligned}(\mathbf{A}^T \mathbf{A})\mathbf{a} &= \lambda \mathbf{a} \\ (\mathbf{A} \mathbf{A}^T)\mathbf{h} &= \lambda \mathbf{h}\end{aligned}$$

$$\lambda = (\alpha \beta)^{-1}$$



Hubs and authorities



Практическая часть

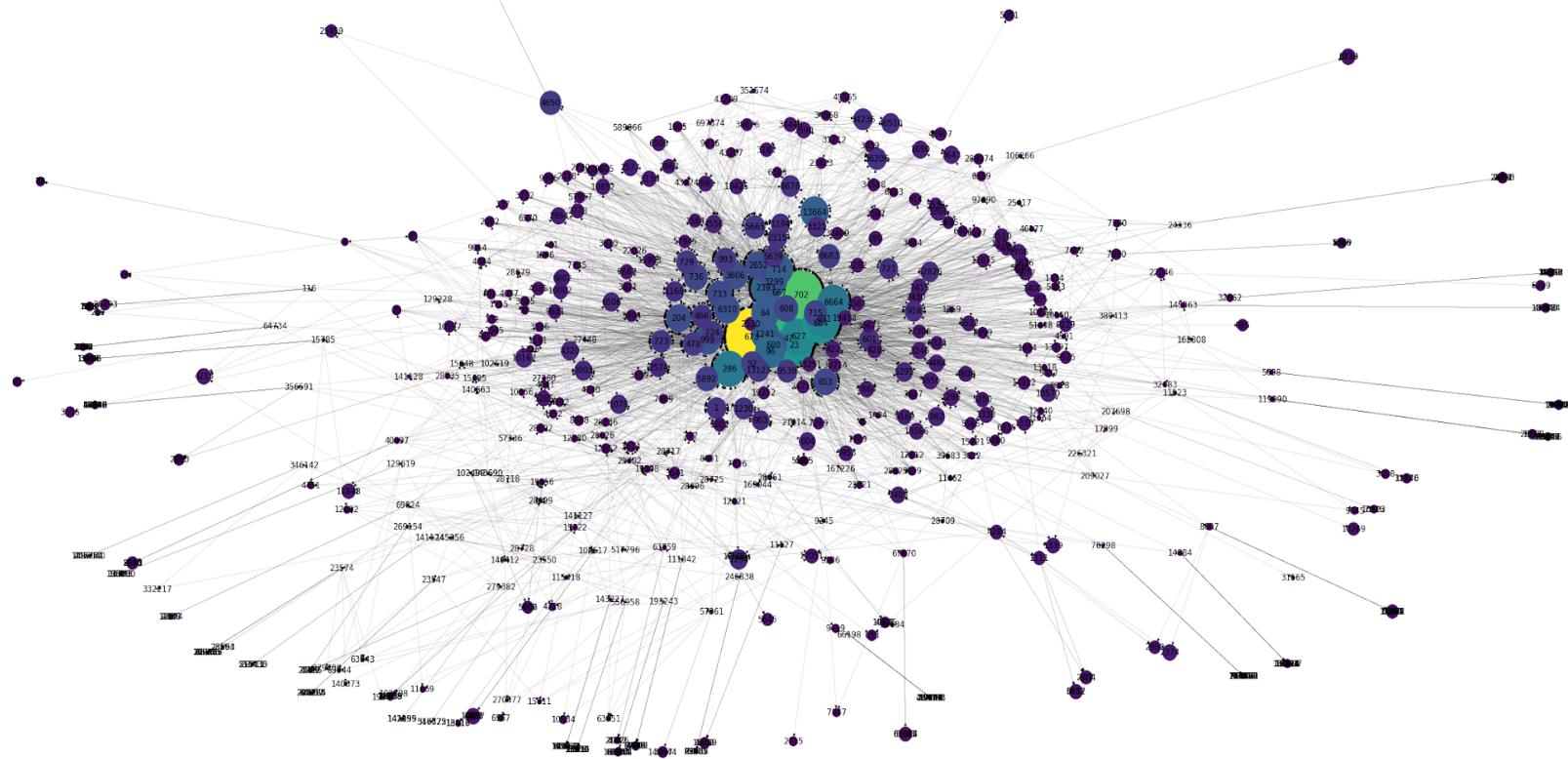
- Для анализа была выбрана Wikipedia на русском языке (<https://ru.wikipedia.org/wiki/>);
- Был реализован краулер страниц;
- Для каждой страницы запоминался адрес и набор ссылок на другие страницы;
- После этого строилась разреженная квадратная матрица (**1 900 000 x 1 900 000**)

Практическая часть

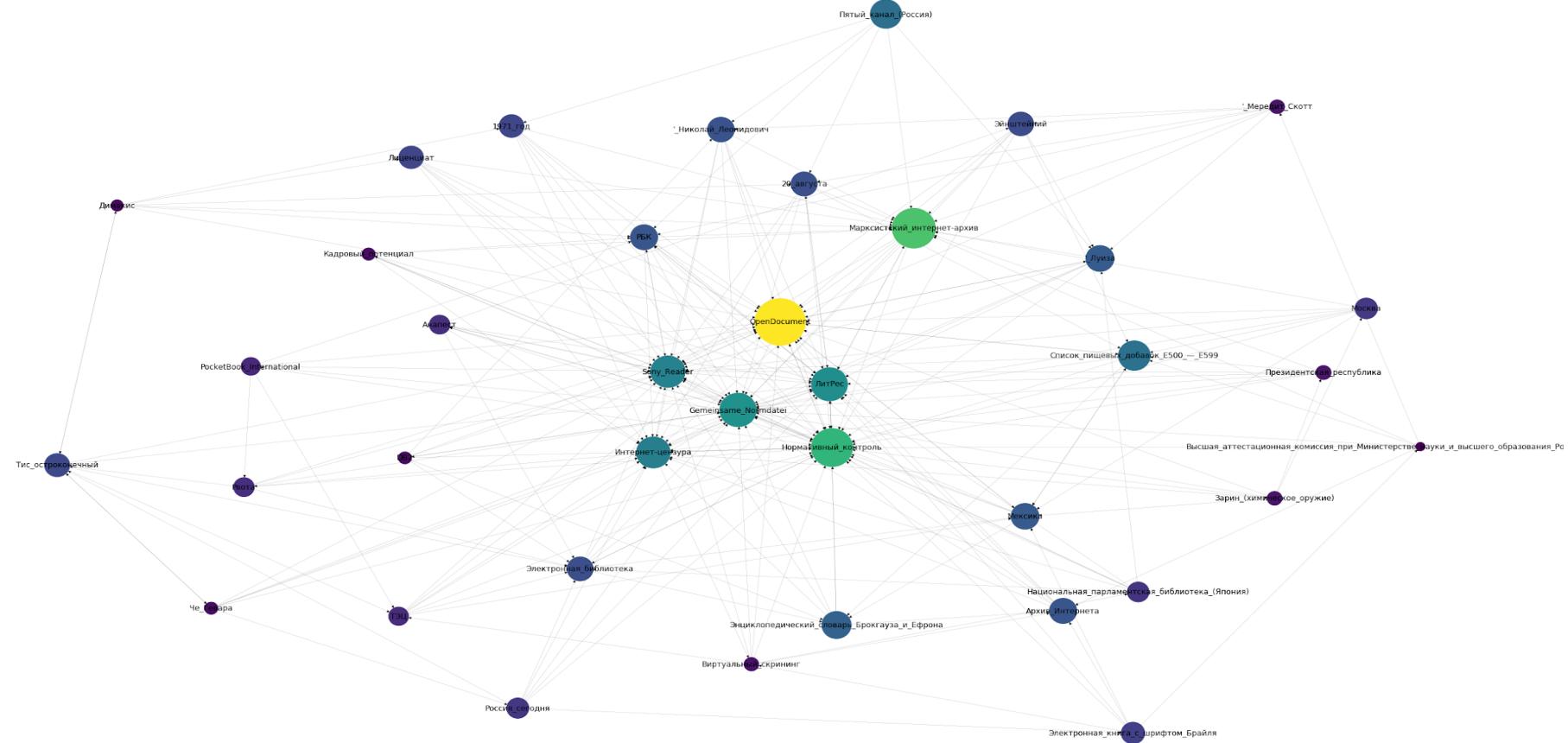
Пример краулинга:

'<https://ru.wikipedia.org/wiki/Озон>', 'https://ru.wikipedia.org/wiki/Древнегреческий_язык', '<https://ru.wikipedia.org/wiki/Аммиак>', 'https://ru.wikipedia.org/wiki/Теплота_испарения', '<https://ru.wikipedia.org/wiki/SMILES>', '<https://ru.wikipedia.org/wiki/Фреон>', 'https://ru.wikipedia.org/wiki/Экстракорпоральная_мембранные_оксигенации', 'https://ru.wikipedia.org/wiki/Химическая_энциклопедия', 'https://ru.wikipedia.org/wiki/Гидроксид_калия', '<https://ru.wikipedia.org/wiki/Дебай>', 'https://ru.wikipedia.org/wiki/1785_год', 'https://ru.wikipedia.org/wiki/Внутривенное_вливание', 'https://ru.wikipedia.org/wiki/Соляная_кислота', '<https://ru.wikipedia.org/wiki/Озон>', '<https://ru.wikipedia.org/wiki/Дезинфекция>', '<https://ru.wikipedia.org/wiki/Детонация>', 'https://ru.wikipedia.org/wiki/Температура_кипения', '<https://ru.wikipedia.org/wiki/Фосфор>', '<https://ru.wikipedia.org/wiki/Диоксин>', 'https://ru.wikipedia.org/wiki/Нитрат_аммония', 'https://ru.wikipedia.org/wiki/Молярная_теплоёмкость', '<https://ru.wikipedia.org/wiki/Азот>', '<https://ru.wikipedia.org/wiki/Хлорирование>', '<https://ru.wikipedia.org/wiki/PubChem>', 'https://ru.wikipedia.org/wiki/Gemeinsame_Normdatei'

PageRank



PageRank



Анализ PageRank



Википедия
Свободная энциклопедия

Заглавная страница
Рубрикация
Указатель А – Я
Избранные статьи
Случайная статья
Текущие события

Участие

Сообщить об ошибке
Сообщество
Форум
Свежие правки
Новые страницы
Справка
Пожертвовать

Инструменты

Ссылки сюда
Связанные правки
Служебные страницы
Постоянная ссылка
Сведения о странице

Печать/экспорт

Создать книгу 
Скачать как PDF
Версия для печати

В других проектах

Викисклад
MediaWiki
Мета-вики
Викивиды
Викиучебник
Викиданнане
Викимания
Викиновости
Викицитатник

Добро пожаловать в Википедию,
свободную энциклопедию, которую может редактировать каждый.

Сейчас в Википедии 1 585 245 статей на русском языке.

Порталы Справка Создать статью

Избранная статья

Клуб Винкс: Тайна затерянного королевства

«Клуб Винкс: Тайна затерянного королевства» (итал. *Winx Club: Il Segreto del Regno Perduto*) — итальянский 3D-анимированный полнометражный мультфильм производства Rainbow CGI совместно с Rai Fiction^{TV}, режиссёром которого выступил Идженнио Страffi. Фильм дебютировал на экранах кинотеатров 30 ноября 2007 года в Италии и стал первым опытом для Rainbow в создании мультфильмов с применением CGI-анимации. Мультфильм завершает сюжетную линию первых трёх сезонов анимационного сериала «Клуб Винкс» про команду девушки-волшебниц, сражающихся со злом в мире волшебства. При этом впервые в истории франшизы персонажи представлены в трёхмерной графике. Идея фильма состоит в том, что в каждом есть что-то волшебное и особенное, что делает его неповторимым, а волшебство позиционируется автором как энергия, формирующая личность человека. По сюжету главная героиня Блум однажды узнаёт, что она фея и к тому же принцесса заселкованного королевства. С тех пор она подружилась с пятью другими феями и стала защищать волшебный мир от зла, но не переставала искать информацию о себе и своих родителях. Мультфильм подводит черту в этой сюжетной линии и проливает свет на оставленные без ответа в сериале вопросы.

Читать

Все 1256 избранных статей

...

Хорошая статья

Административно-территориальное деление Башкортостана

Из новых материалов

Знаете ли вы?

- На Землю ежедневно падает от 60 до 100 тонн космической пыли (на илл.).
- Югославский ответ кока-коле последовал на 20 лет раньше советского.
- В путешествии на край света богу грома пришлось прятаться в рукавице великаны.
- Кусочки пласти массы добавляют в косметику и обнаруживают в пиве и питьевой воде.
- ООО — общество с ограниченной ответственностью, а ООО — псевдоним Гоголя.
- Самиздат в США выпускал комиксы.
- Железные дороги Непала (на илл.) имеют суммарную протяжённость 34 километра и состоят из двух веток.
- Руководителя «селекционеров» узников Освенцима судили, но отпустили.
- За три года до выхода в финал Южноамериканского кубка эквадорский вратарь подрабатывал водителем и продавал манго.
- На случай нехватки скафандром при спасении американцы придумали специальный шар.
- На знаменитой суриковской картине изображены неправильные кремлёвские башни.
- Центр железного дела северо-восточной Нигерии страдал сначала от оттока, а потом от притока населения.



Анализ PageRank

Создание книги



Создатель книг претерпевает изменения

Из-за серьёзных проблем с существующей системой «Создатель книг» пока не поддерживает сохранение книг в формате PDF. Мы прилагаем все усилия, чтобы заменить свою систему и снова позволить сохранять PDF-файлы в «Создателе книг».

[Узнать больше](#)



Загрузите отдельные страницы

Поскольку PDF-рендер с помощью Book Creator был заменён, вы можете попробовать функцию **Скачать как PDF**, которая находится на боковой панели в режиме чтения статей.

С помощью **книжного мастера** вы можете создать книгу, содержащую выбранные вами вики-страницы. Вы сможете экспорттировать книги в различные форматы (например, в PDF или ODF) или заказать печатный экземпляр.

[Запустить функцию книжного мастера](#)

[Отменить](#)

Дополнительную информацию см. в [справке о книгах](#).

Анализ PageRank

Википедия:Книги

Материал из Википедии — свободной энциклопедии



Книги

[Создать книгу](#) · [Справка для начинающих](#) · [Книжный поиск](#) · [Справка для эксперта](#)

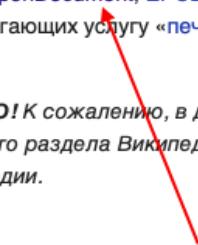
Краткая информация

Вики-книга — это самостоятельно сделанная подборка статей Википедии, которую легко сохранить «как есть» или же в форматах [PDF](#), [OpenDocument](#), [EPUB](#) или [ZIM](#). Сохранённую в совместимом формате вики-книгу можно заказать в печать в одной из компаний, предлагающих услугу «[печать по требованию](#)».

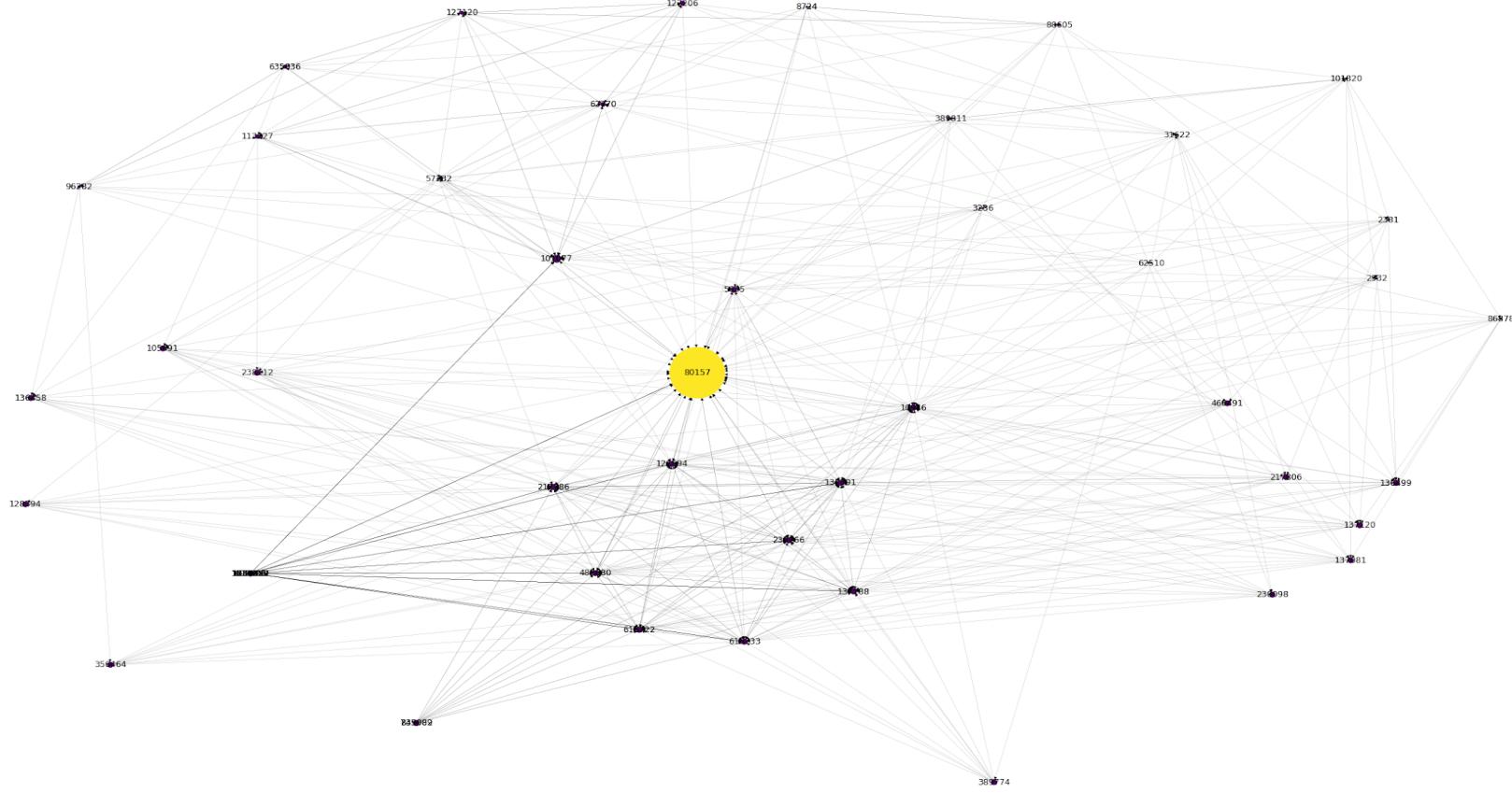
ВАЖНО! К сожалению, в данный момент функция сохранения книги как файла в электронном виде и функция заказа печати для Русского раздела Википедии [не работают](#)! Вы можете только создавать книги, как персональные сборники статей Российской Википедии.

С декабря 2007 года [Фонд Викимедиа](#) имеет партнёрское соглашение по печати вики-книг с онлайновым сервисом [PediaPress](#), который перечисляет 10% прибыли от заказов на развитие проектов Фонда Викимедиа. Этот онлайновый сервис предлагается по умолчанию, если вы решите заказать печать книги. Вы можете выбрать и другие компании, осуществляющие «[печать по требованию](#)» (англ. *print on demand*).

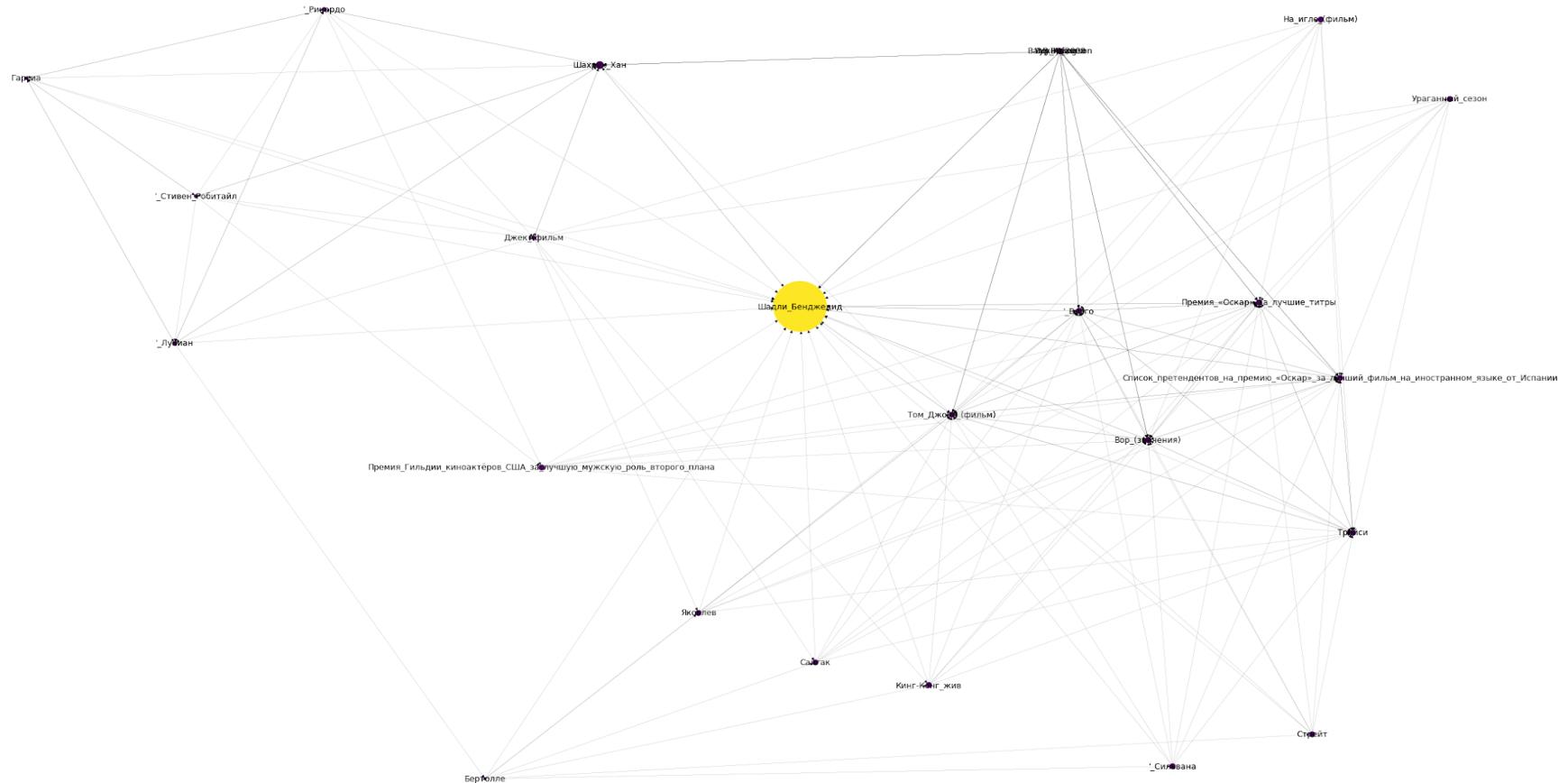
С конкретными вопросами по составлению и печати вики-книг вы можете обратиться к [справочной странице](#) или на [специальный форум](#). С более общими вопросами по использованию материалов Википедии можно обратиться на [форум «Вопросы»](#).



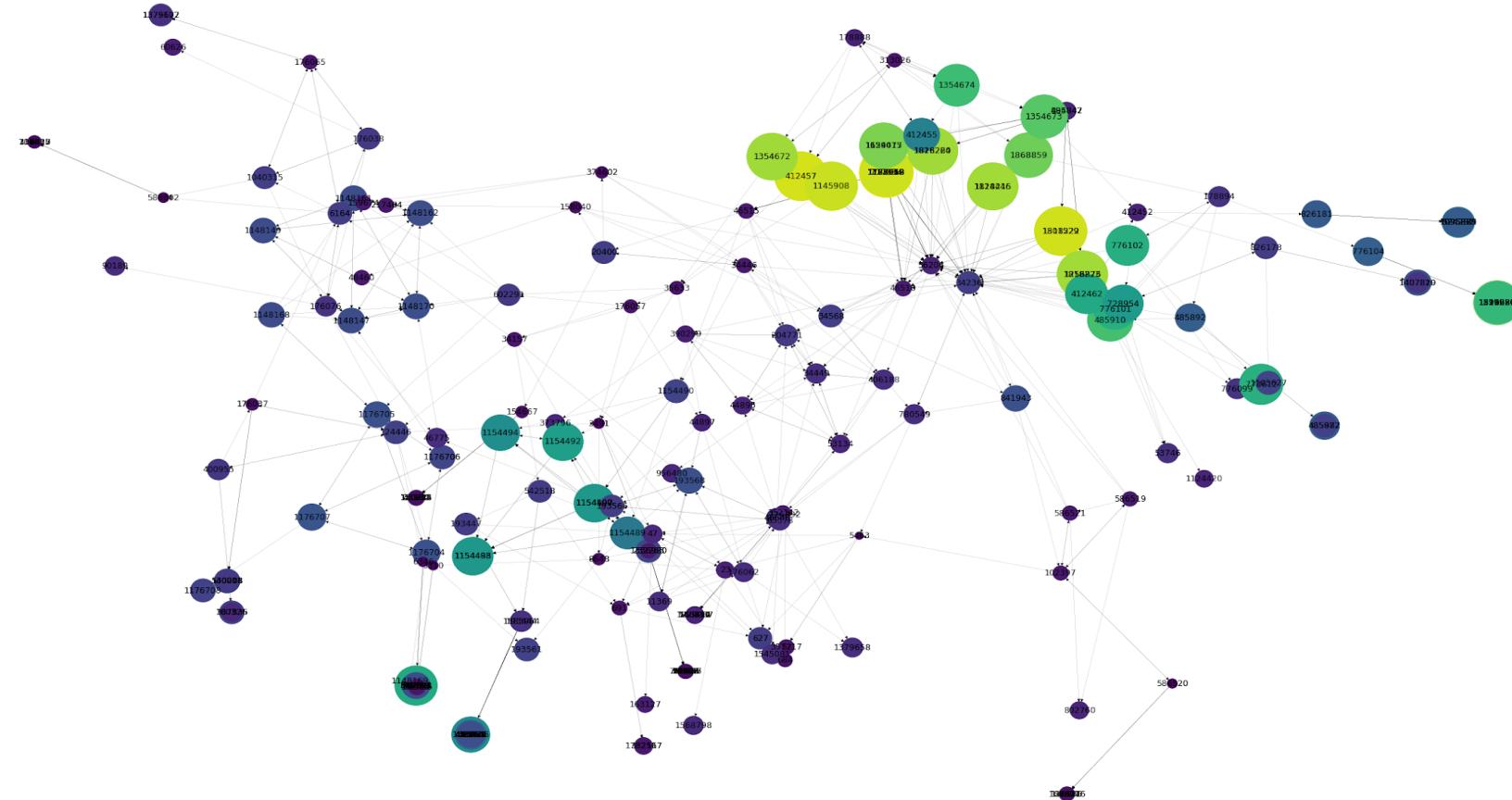
Eigenvector centrality



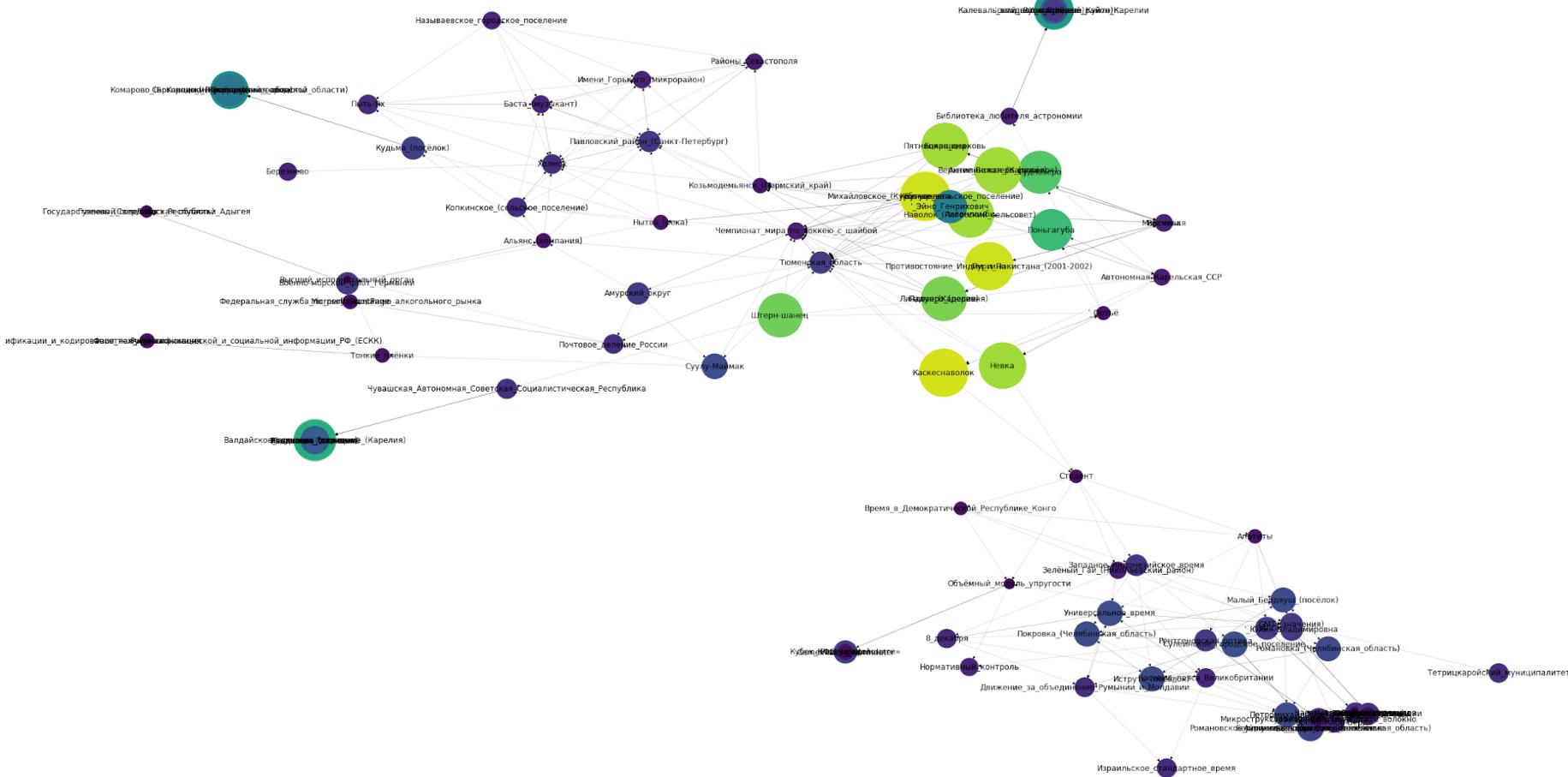
Eigenvector centrality



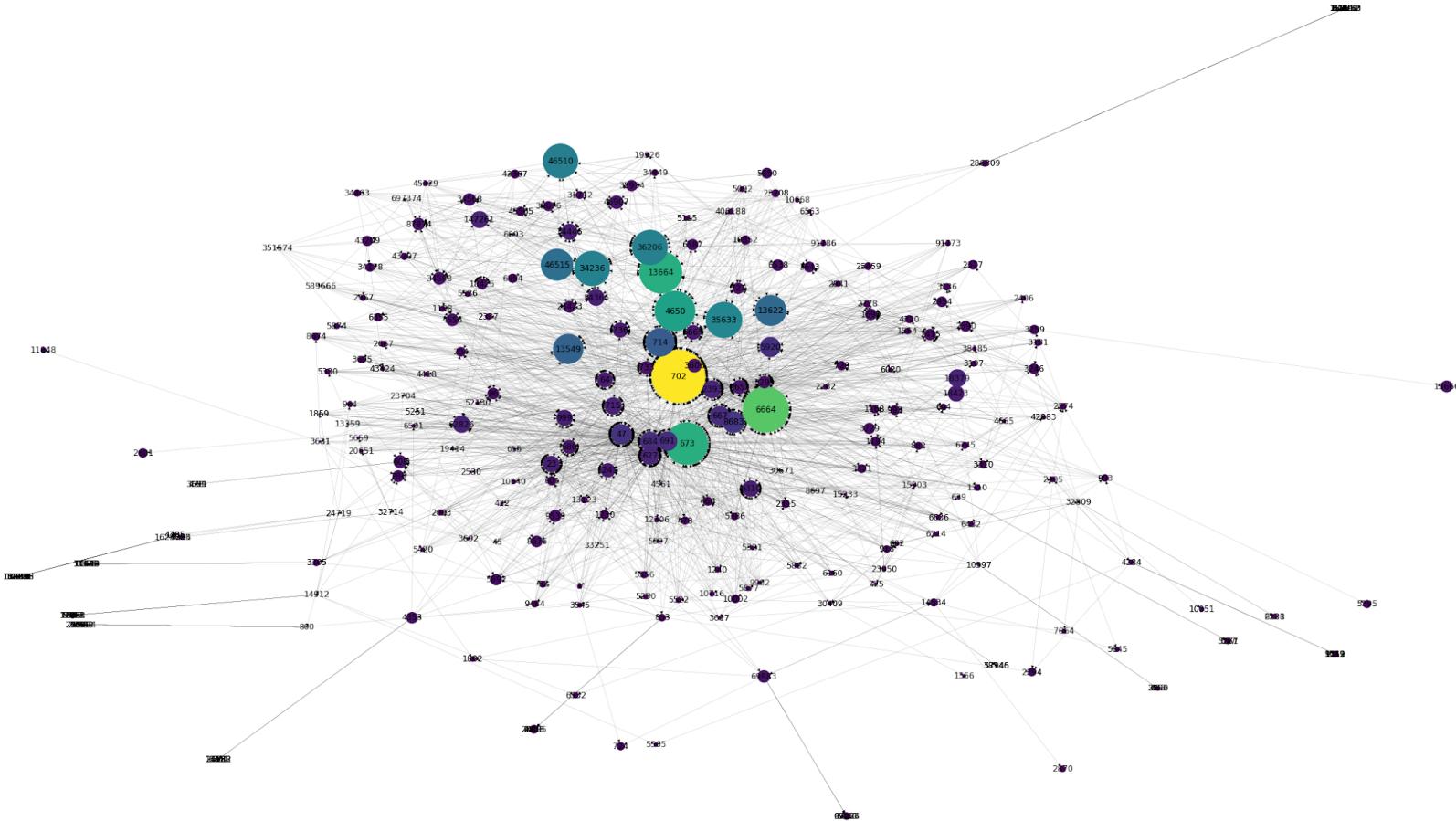
Hubs



Hubs



Authorities



Authorities

