

Послание Президента Федеральному Собранию

Выполнили:

Дасаев Гарри
Птичкина Анастасия
Файн Егор
Филатов Михаил

Актуальность

- Сетевой анализ является удобным инструментом для анализа человеческой речи
- Речи российских политиков редко становились объектом изучения с помощью сетевого анализа
- Анализ послания с помощью сетевых методов полезен как с точки зрения применения сетевых методов для анализа речи, так и с точки зрения изучения повестки российских политиков

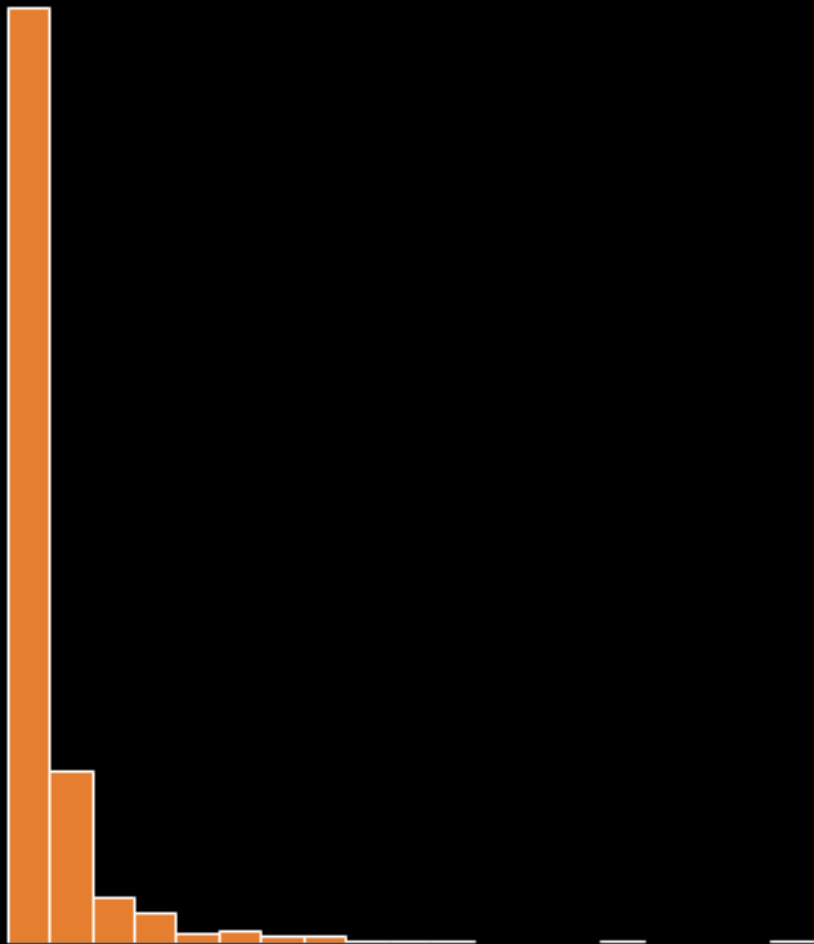
Обзор литературы

- Сетевой анализ биграмм абстрактов диссертаций: (Tunalı&Bilgin, 2014)
 - выявление трендов по темам диссертаций
 - выявление наиболее связанных друг с другом тем
- Анализ публичного дискурса Центрального Банка ЕС: (Nerghes et al., 2014)
 - основные акценты в публичном дискурсе ЕвроЦБ до, во время, после кризиса-2008 и периода восстановления
- Анализ тем дискуссий в Твиттере: (Nachař, Ogiela, 2018)
 - сетевой анализ биграмм определяет темы хуже чем хэштег-анализ
 - анализ биграмм может улучшить хэштег анализ

Сбор и подготовка данных

- Парсинга сайта Президента РФ
 - Использование библиотек `requests` и `bs4`
- Удаление стоп-слов и лемматизация
 - Модуль `nltk` и библиотека `py morphology2`
- Создание биграмм
 - Модуль `nltk`

Описательные статистики



Количество вершин в графе: 2306

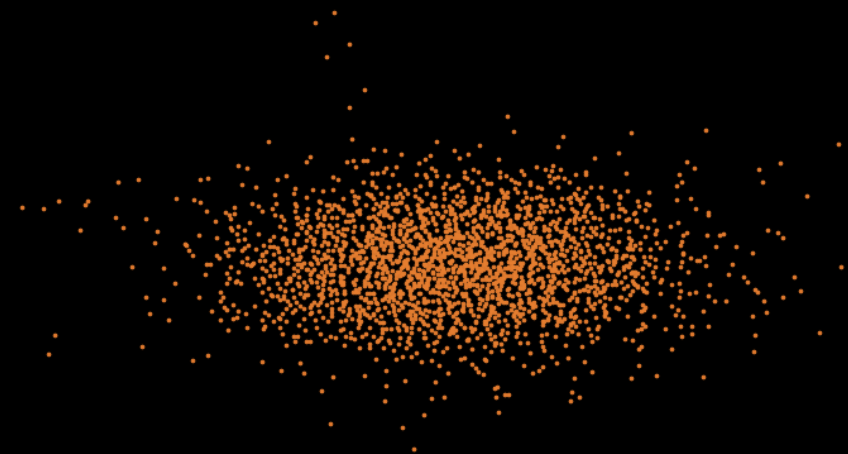
Количество ребер в графе: 6173

Диаметр графа: 10

Радиус графа: 6

Плотность графа: 0.001

Транзитивность: 0.029



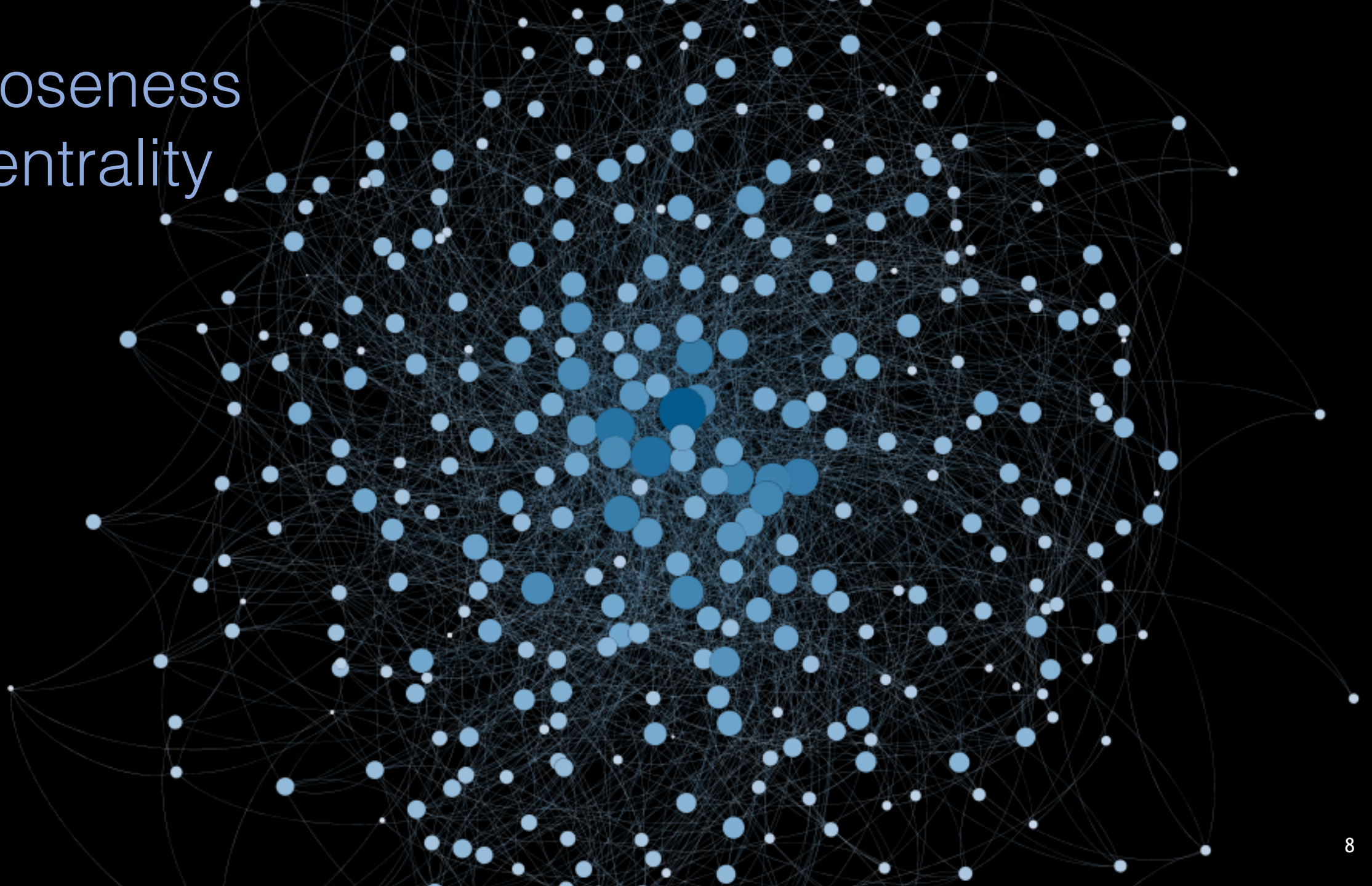
Degree Centrality



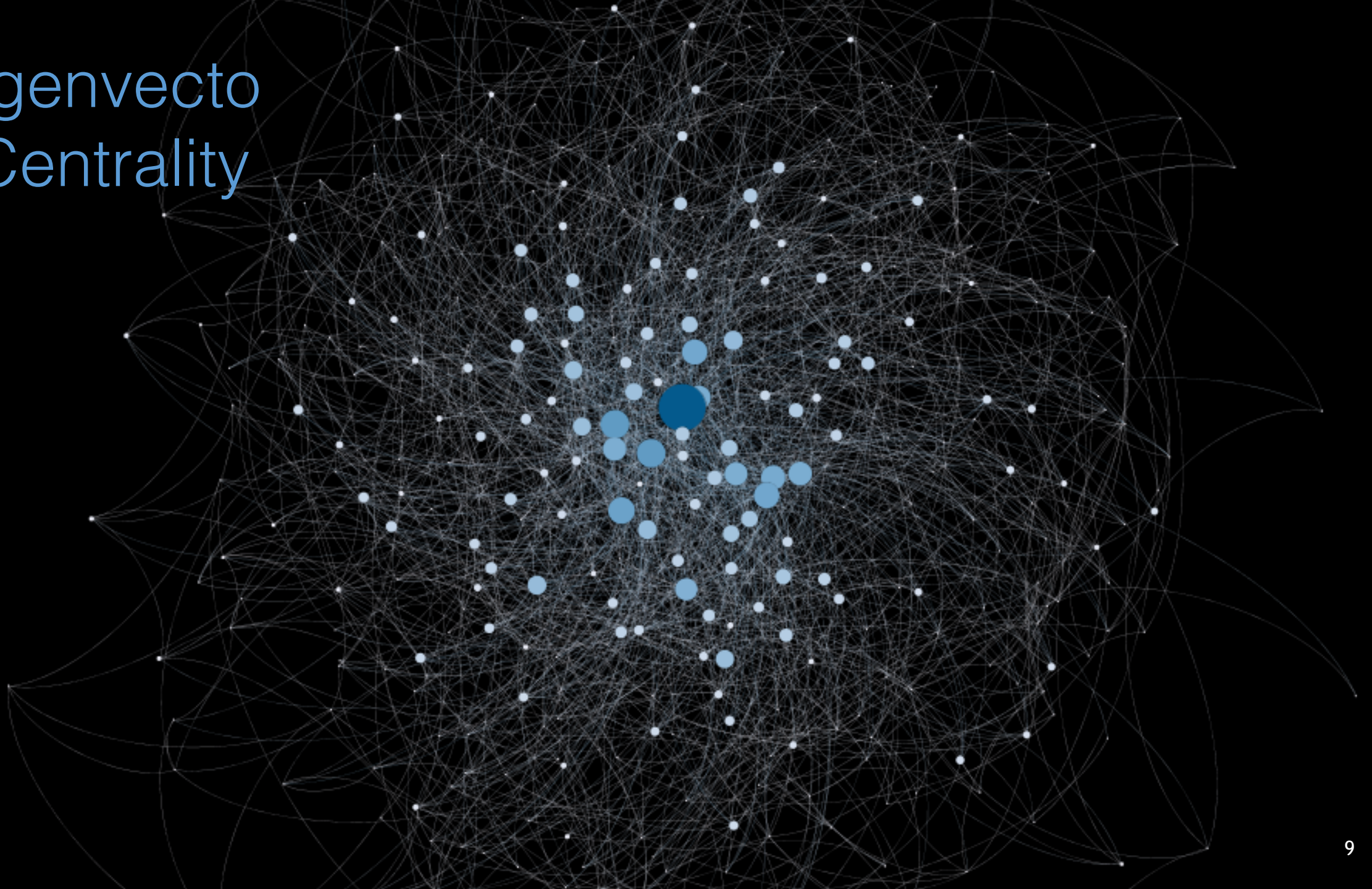
Betweenness Centrality



Closeness Centrality



Eigenvector Centrality

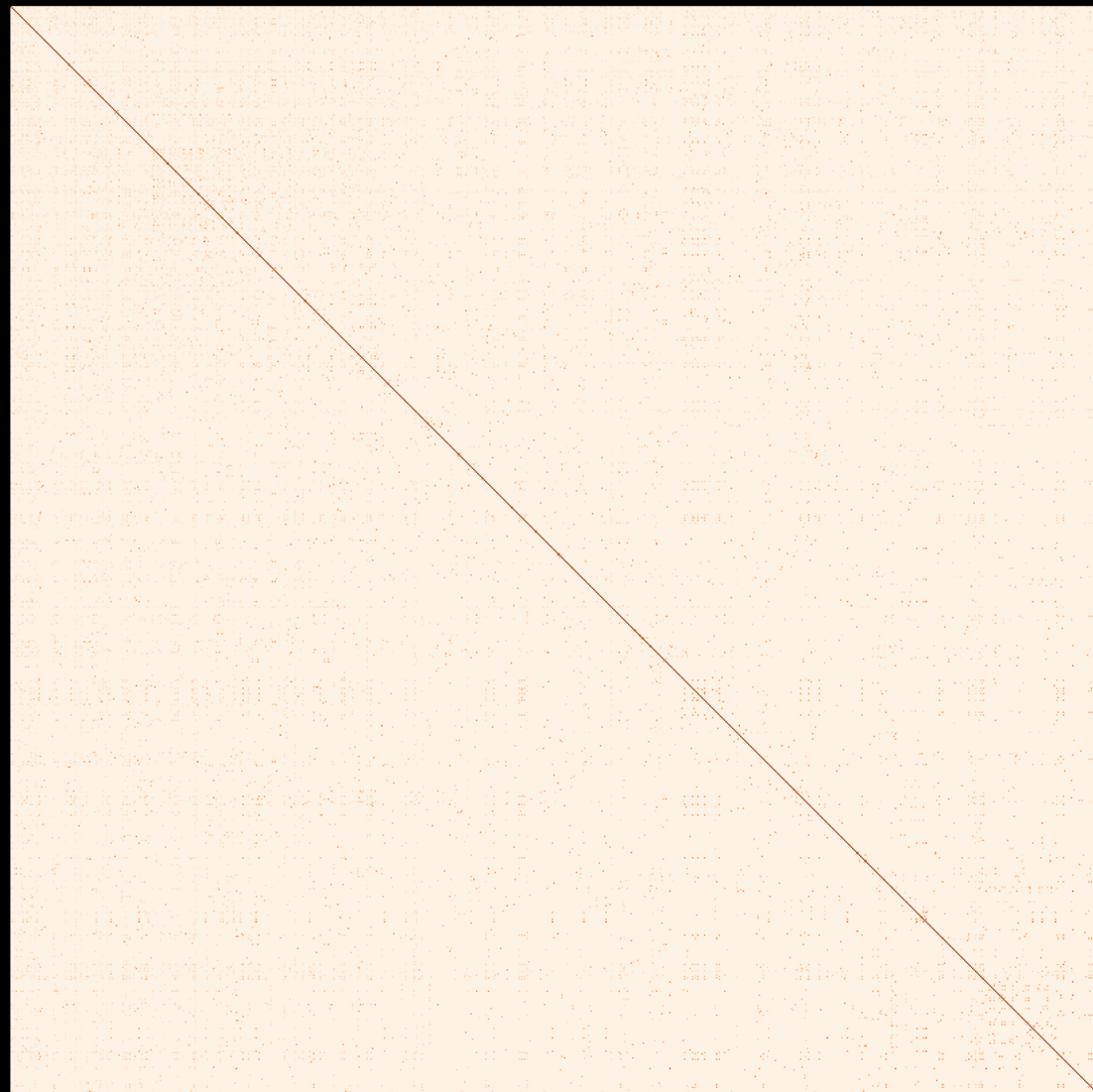


Структурная эквивалентность

Вершины графа, построенного на биграммах, очень слабо коррелируют между собой.

В ходе анализа были выделены следующие пары слов, коэффициент корреляции для которых равен единице

помогать	точно
критически	крайне
критически	чрезвычайно
избавляться	жёстко
заострить	обратить
челябинский	пермский
взаимовыгодный	дружеский
осуществляться	нарушение



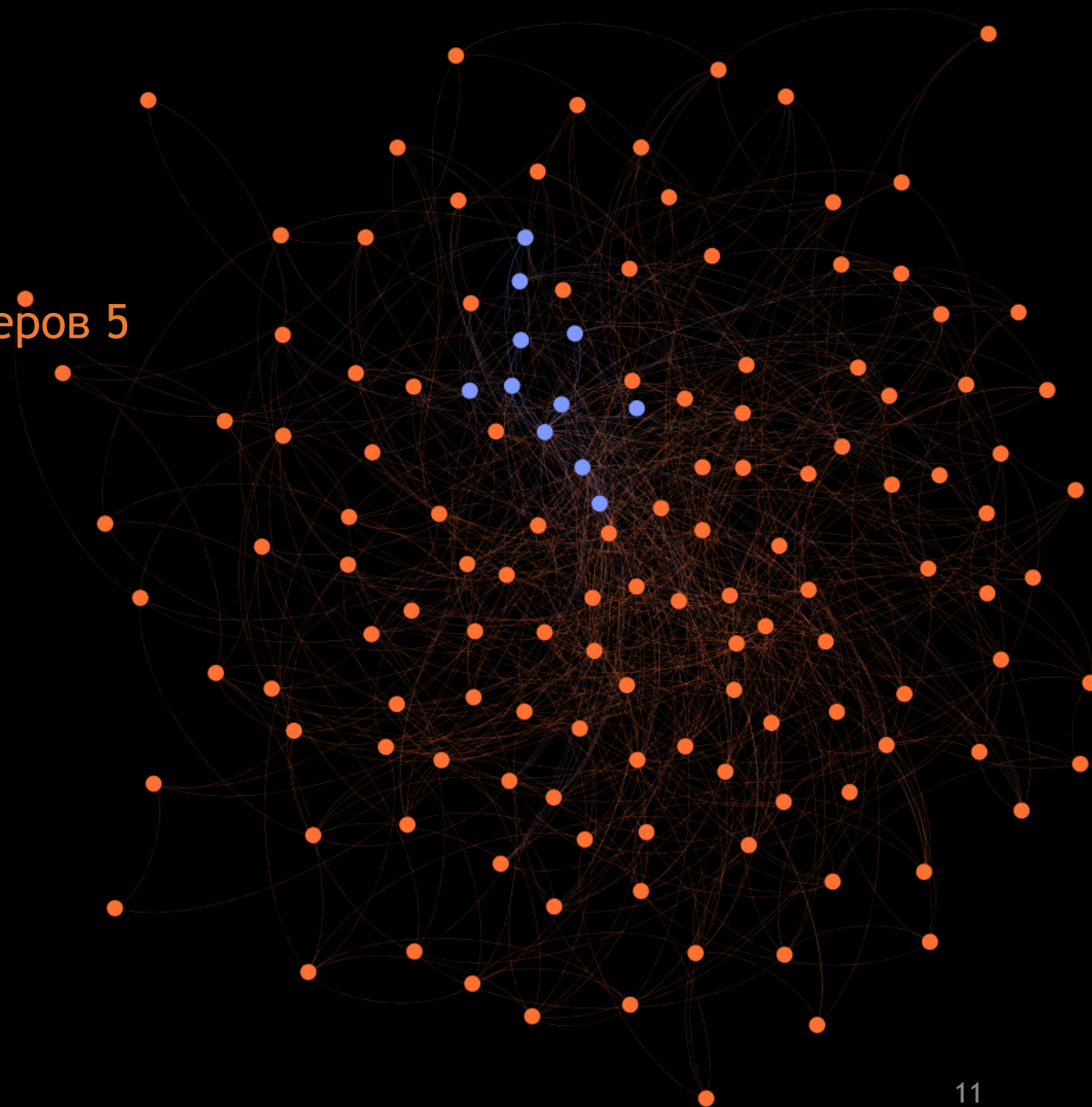
Поиск клик

Было найдено 3 максимальные клики размеров 5

это, вообще, её, хотеть, сделать

это, сегодня, прямо, сказать, хотеть

это, понимать, человек, делать, нужно



Выделение тематических блоков в послании

- Для кластеризации текста послания использовалось два метода выделения сообществ в сети - The asynchronous fluid communities algorithm и Clauset-Newman-Moore greedy modularity maximization algorithm
- V-мера схожести меток - 0.759
- Было выявлено 34 кластера. Определенную смысловую нагрузку в них несут лишь первые 4. Остальные кластеры можно выделить как подблоки основных
- Наиболее предпочтительным для анализа оказался второй метод

Основные тематические блоки

1. Социально-экономическая политика

- бедность, безопасность, гуманитарный, доступность, инфраструктура, наука, образование поликлиника

2. Экономический рост и крупные федеральные проекты

- администрация, дальневосточный, действовать, инвестиция, правительство, рубль, федеральный, экономить

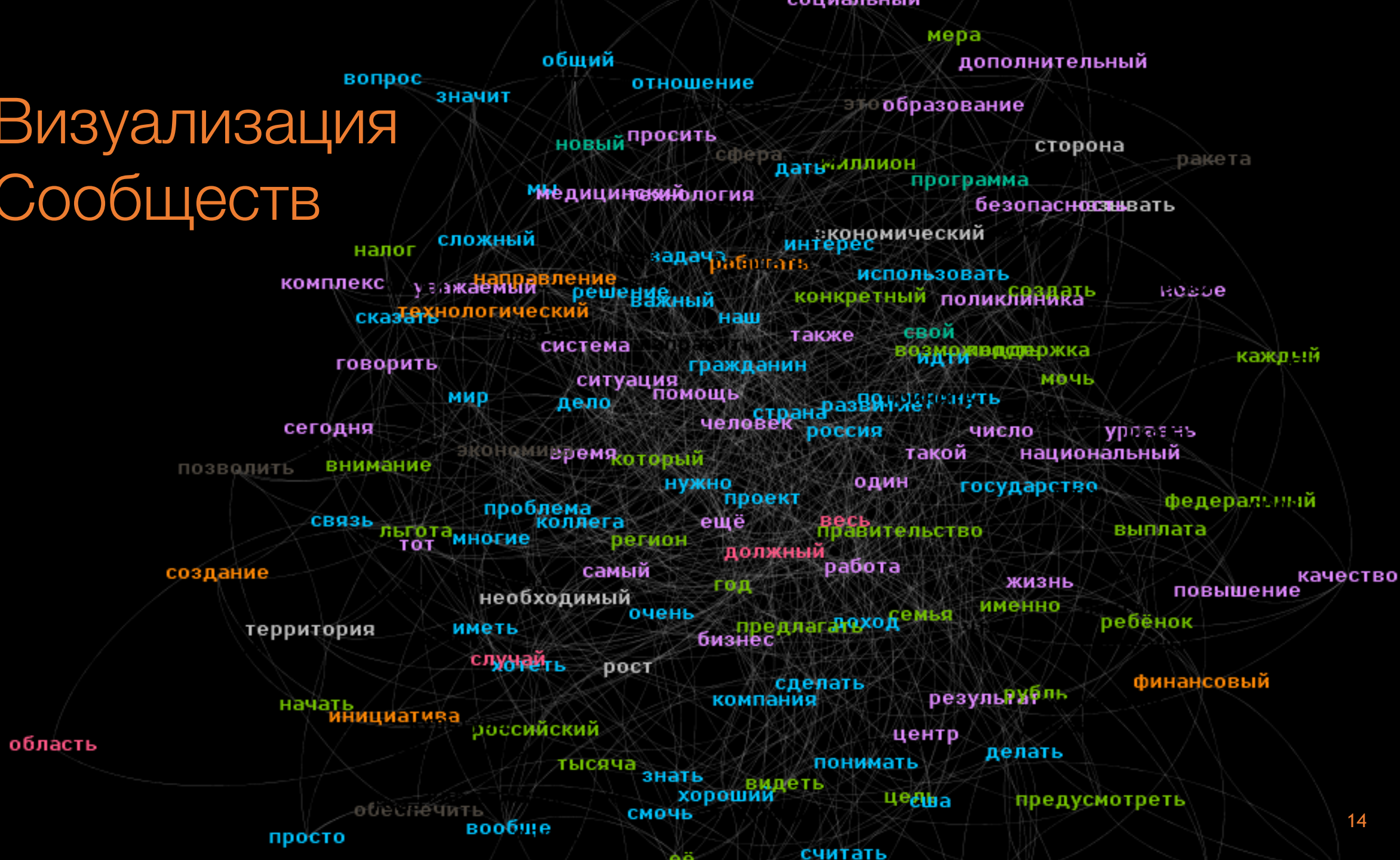
3. Внешняя политика

- США, Китай, Европа, восток, запад, команда, рынок, Россия, член

4. Военные достижения

- Буревестник, внешнеполитический, оборонный, сармат, ядерный

Визуализация Сообществ



Выводы

- Даже при низкой транзитивности графа, который построен на текстовых данных, можно выделять сообщества слов, но при этом появляется множество кластеров-спойлеров
- Невозможно выделить эмоциональную окраску каждой из частей
- Возможность «наложения» одного кластера на другой
- Существует проблема стоп-слов специфичных для текста

Дальнейшие шаги

- Составление сетей путем объединения корпусов текстов (выступления активистов на митингах, анализ текстов в СМИ, комментарии пользователей соцсетей)
- Выработка методологии по работе с текстами в политической науке
- Поиск наиболее подходящих алгоритмов для кластеризации сетей с низкой плотностью и транзитивностью
- Более качественная очистка текста

Спасибо за внимание!