



Исследование многоуровневых алгоритмов для визуализации графов большого объема

Мачульскис Сергей Викторович

гр. 9202 ФИТ НГУ

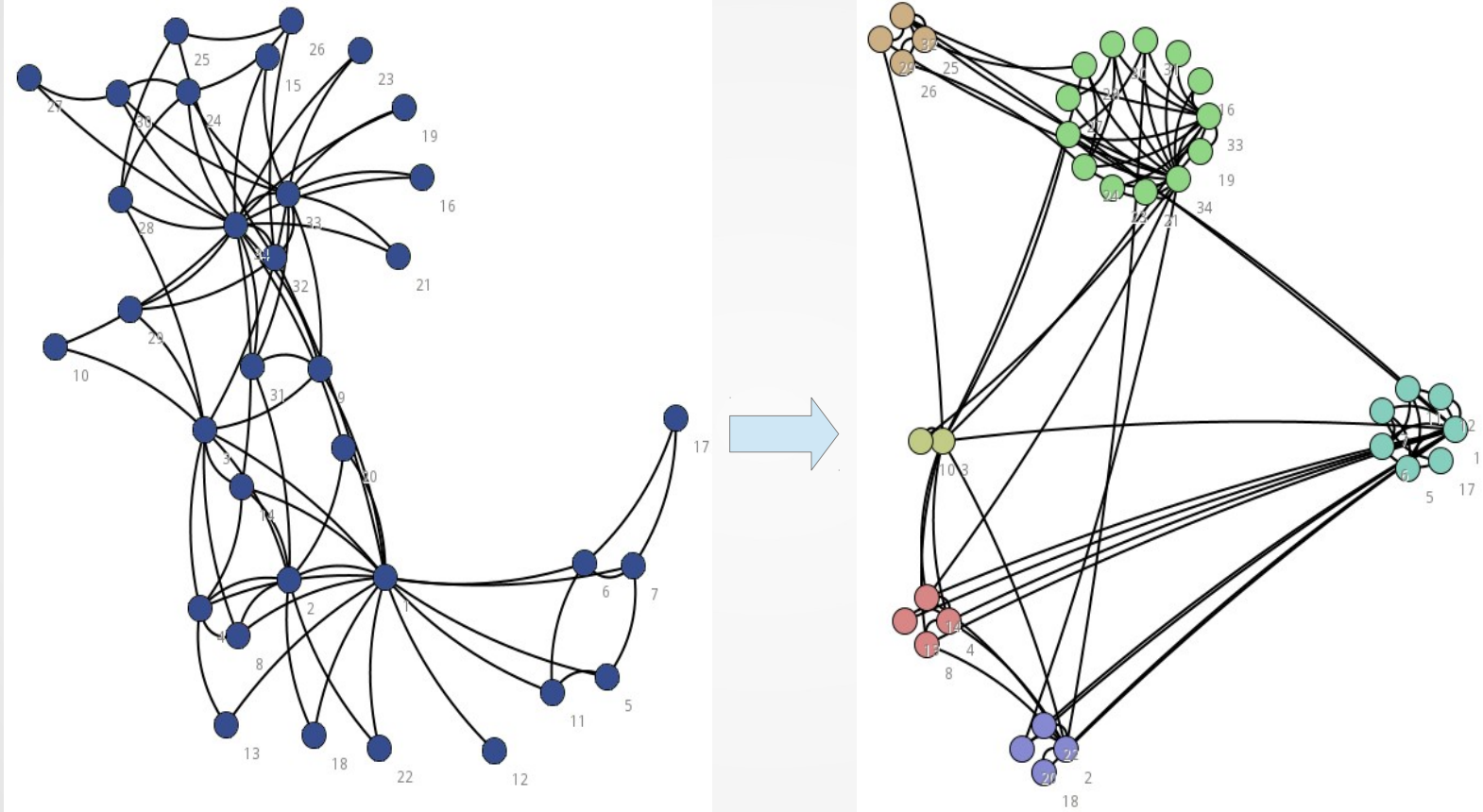
Научный руководитель:

к. ф.-м.н., с.н.с., Апанович Зинаида Владимировна

Введение

- 1) Задача кластеризации
- 2) Специфика: заранее неизвестно количество сообществ
- 3) Оценка качества на основе модулярности

Кластеризация

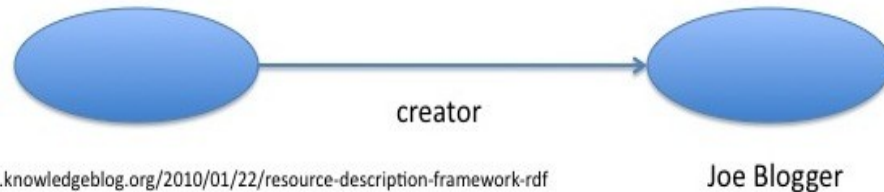
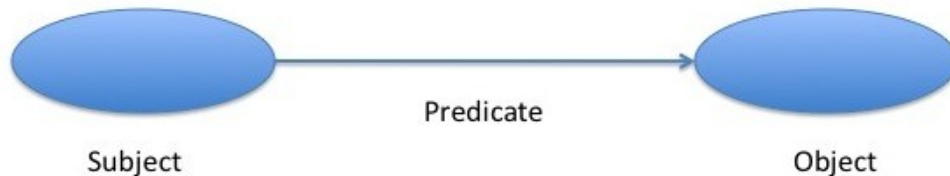


Кластеризация C_c : $\{C_1, \dots, C_n\}$

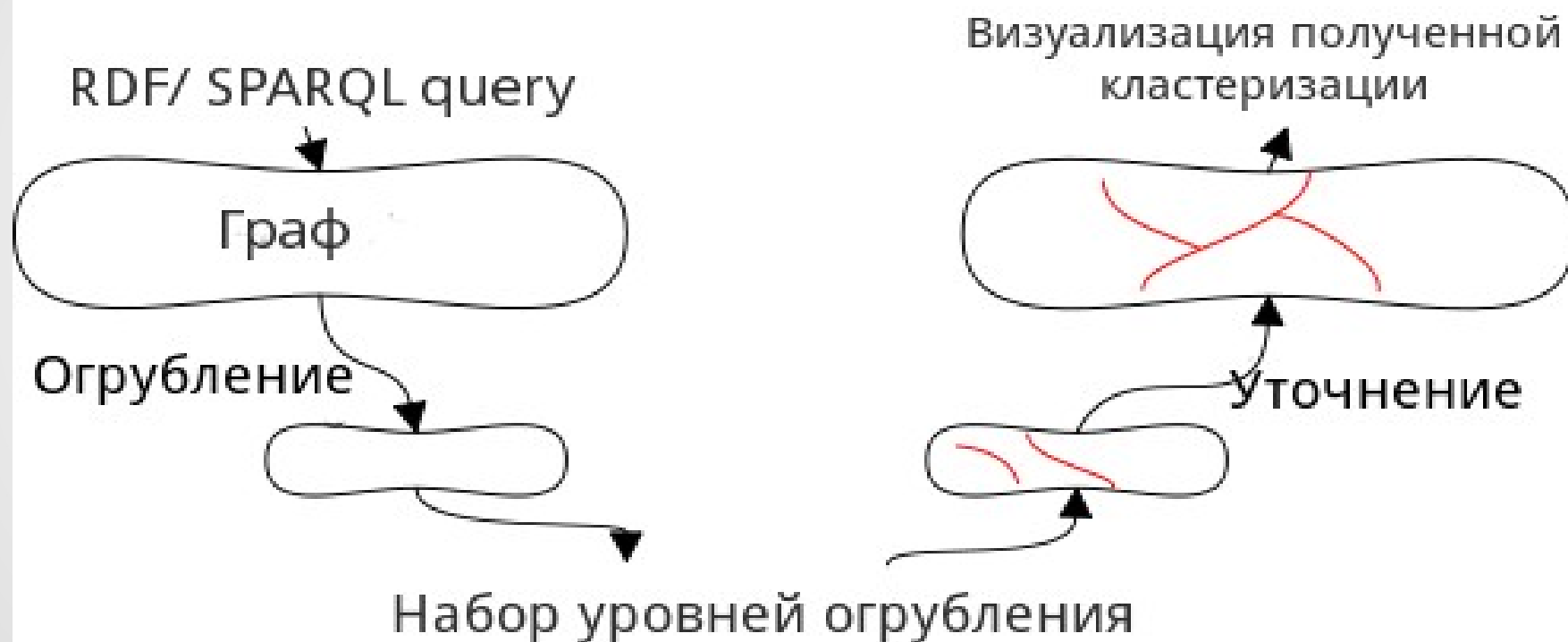
Модулярность кластеризации C_c :
$$Q(C_c) = \sum_{C \in C_c} \left(\frac{\deg(C, C)}{\deg(V, V)} - \frac{\deg(C)^2}{\deg(V)^2} \right)$$

Входные данные

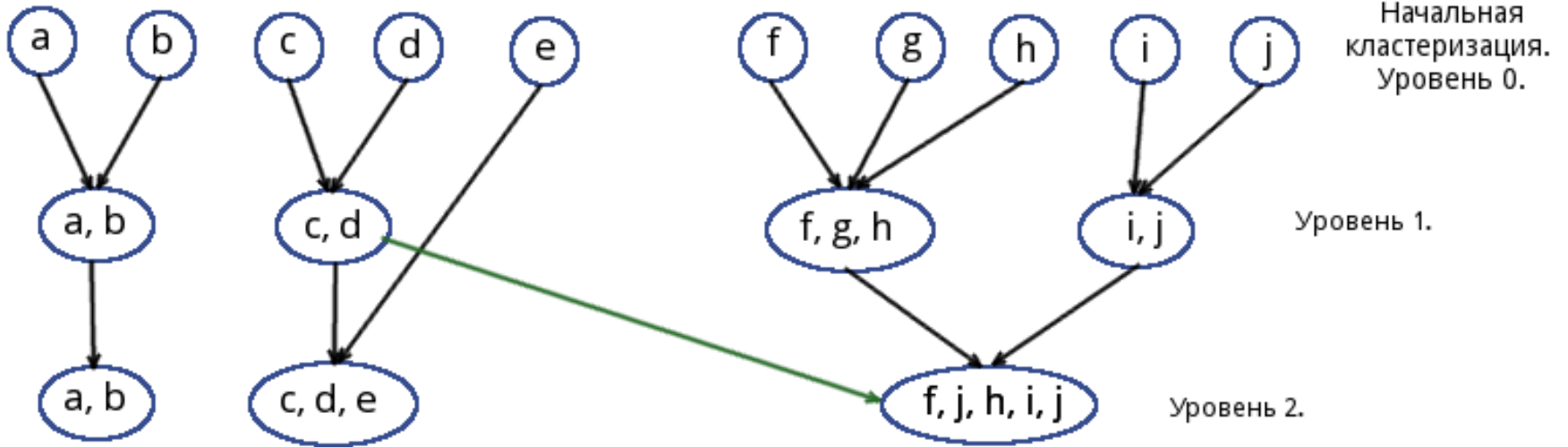
- Открытый архив СО РАН. 600К троек.
- Сети цитирования (CiteSeer). 8.2М троек.
- Сети соавторства (DBPL). 22М троек.



Архитектура приложения



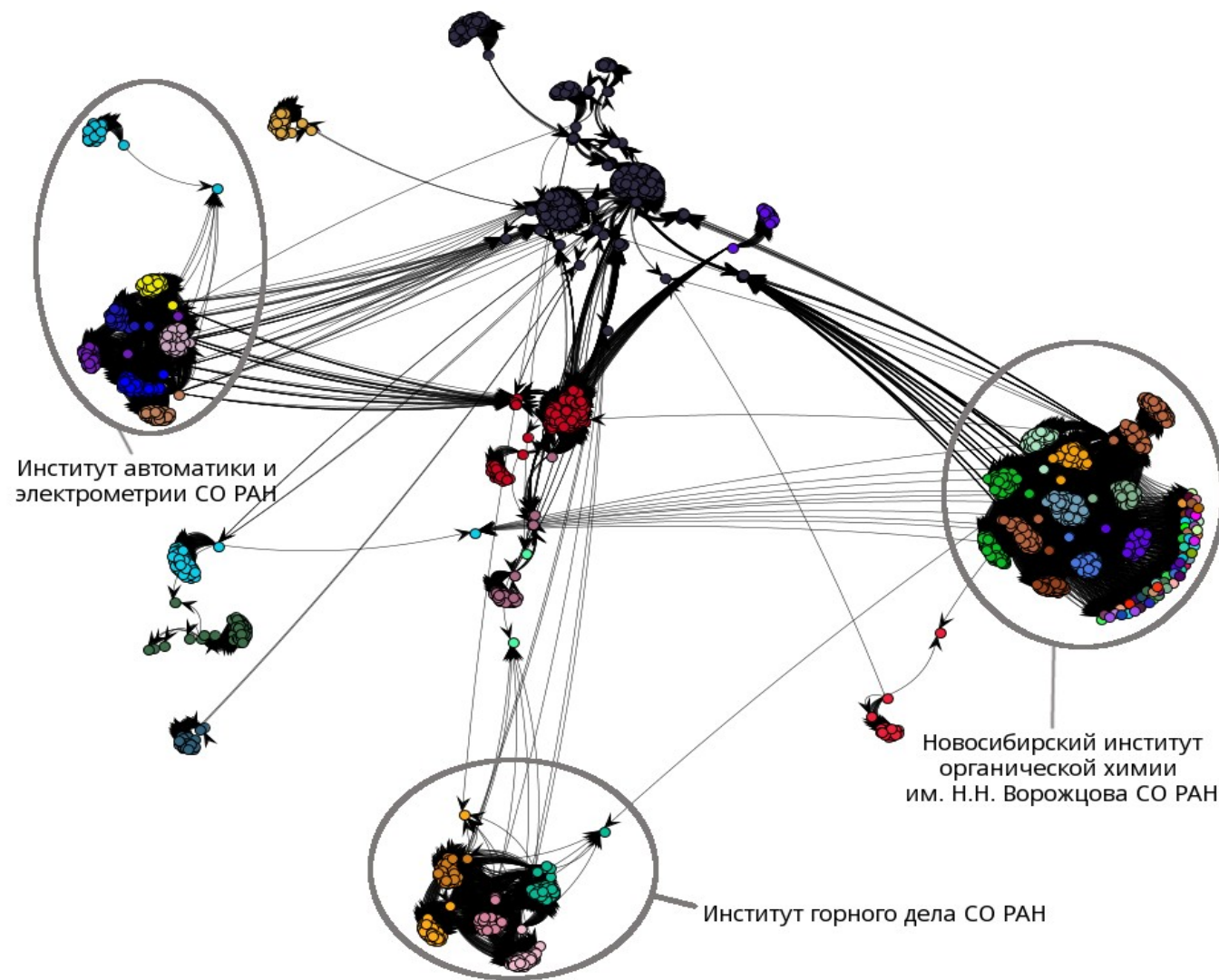
Уровни огрубления



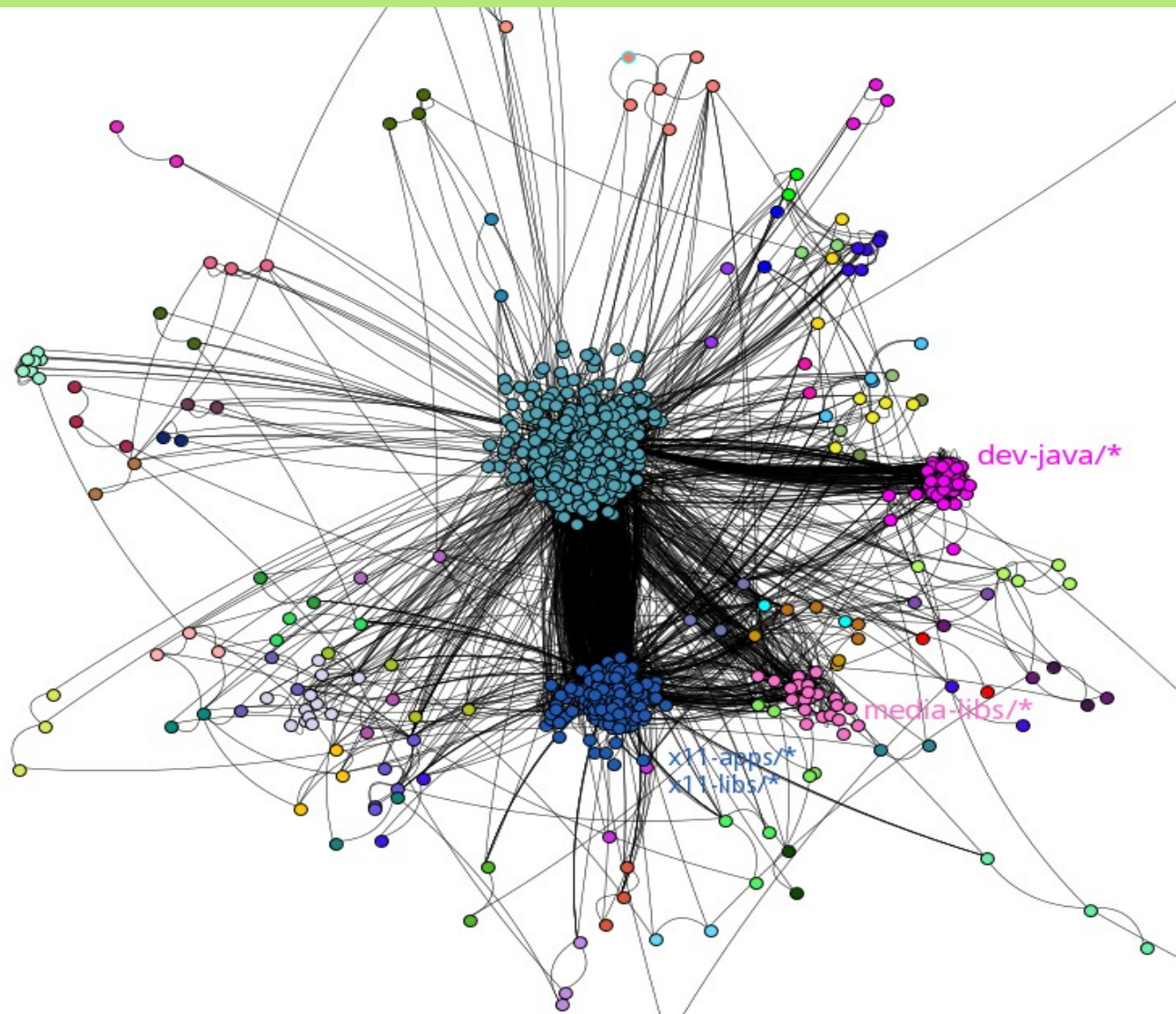
Оптимизация алгоритмов.

- Хранение уровней огрубления в виде сбалансированного дерева.
- Создание очереди с приоритетом позволяет гораздо реже пересчитывать модулярность при слияниях кластеров
- Использование силовых алгоритмов для визуализации.

Граф сотрудничества СО РАН. 20 тысяч ребер.



Граф зависимостей приложений и библиотек ОС GNU/Linux.



Результаты

- Реализован многоуровневый алгоритм кластеризации.
- Реализован модуль обработки SPARQL-запросов.
- Проведены оптимизации, реализовано несколько представлений промежуточных результатов алгоритма и несколько способов их получения. Выбран оптимальный вариант.
- Реализована интерактивная визуализация полученных результатов, размещение полученных кластеров, вершин кластеров.
- Анализ полученных кластеров, нахождение причин образования кластеров в графах.

Возможные направления исследования.

- 1) Обобщение понятия модулярности на случай ориентированного графа.
- 2) Работа с покрытиями.

**Исследование многоуровневых алгоритмов кластеризации
для визуализации графов большого объема**

Спасибо за внимание.