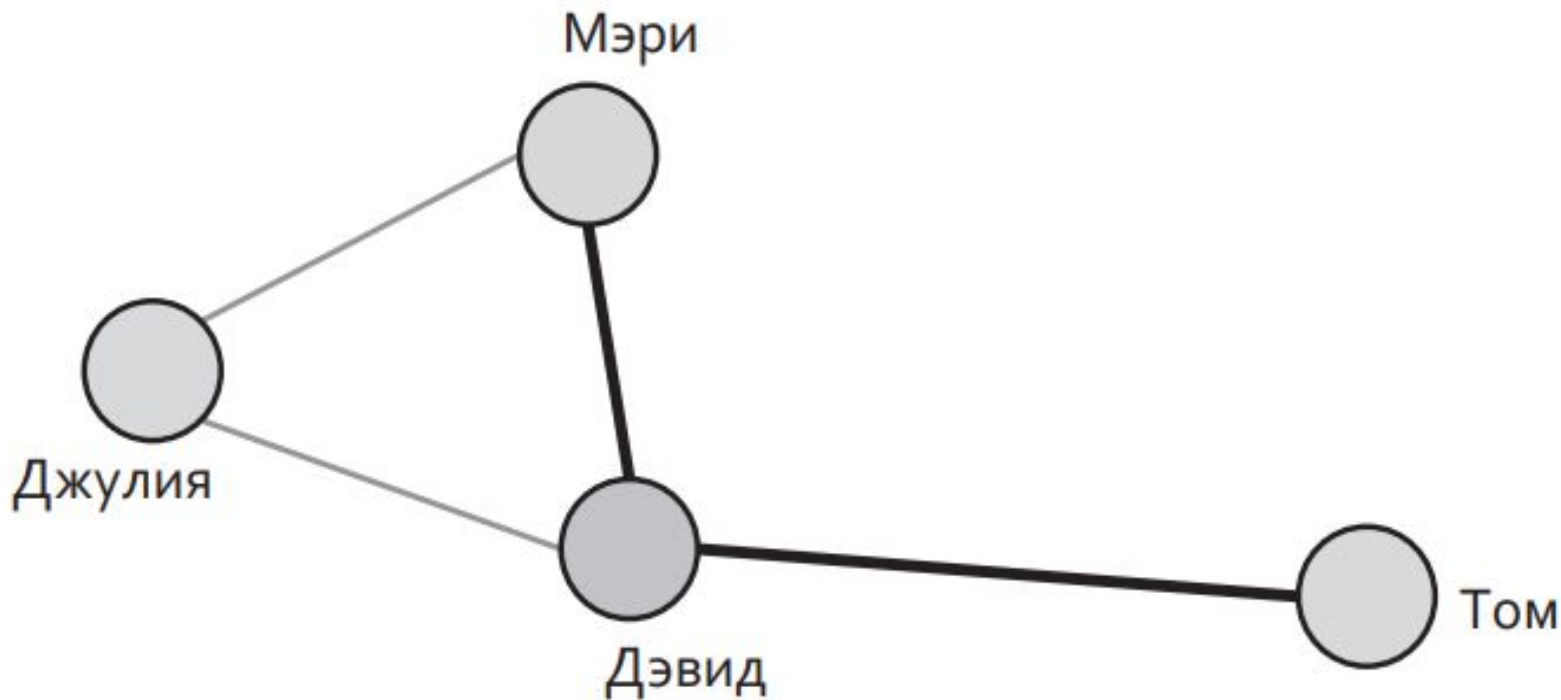


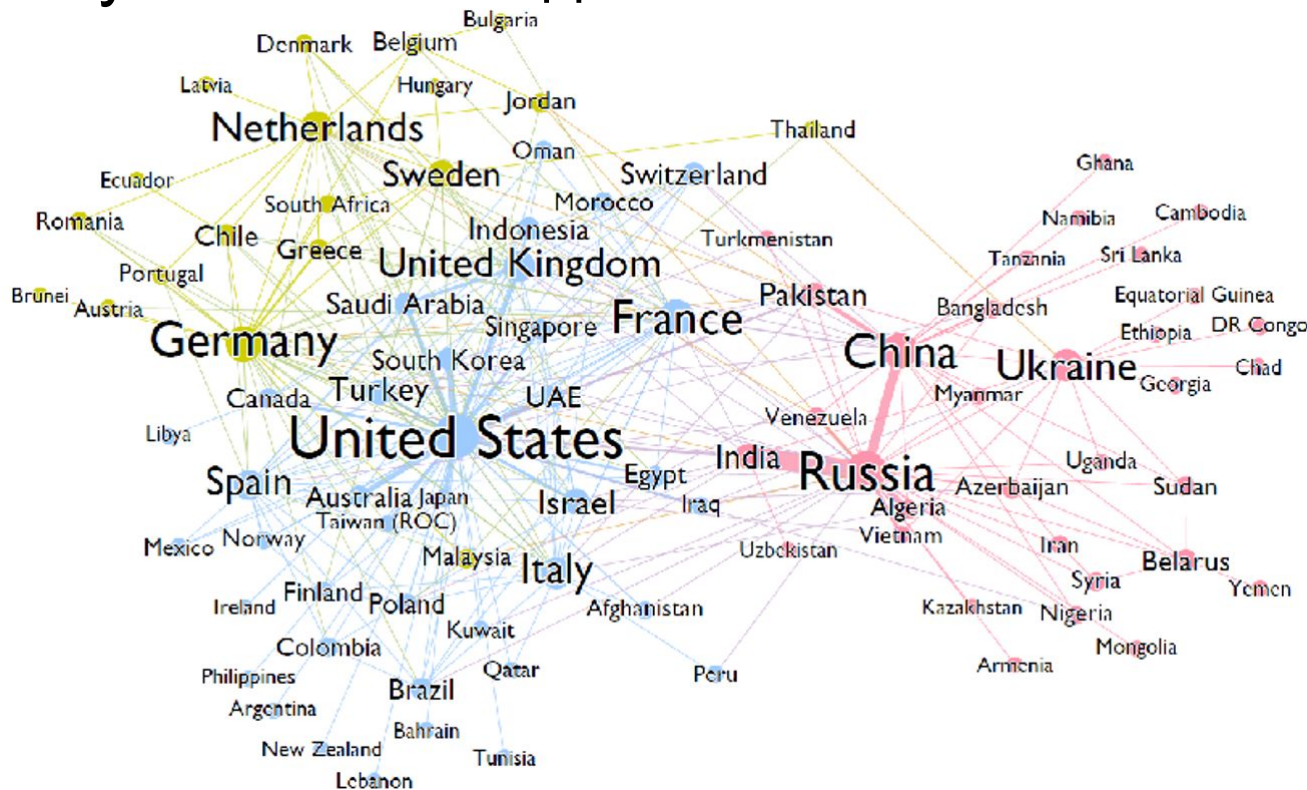
# Анализ социальных сетей

Гуляева Ольга

Анализ социальных сетей — это метод изучения взаимосвязей между объектами в сети, где **узлы** представляют людей или другие объекты, а **ребра** — связи между ними.



# Лувенский метод



Лувенский метод помогает находить **кластеры** в сети. Кластеры — это группы объектов, которые имеют больше связей между собой, чем с другими частями сети.

Военные поставки были выбраны в качестве косвенного показателя двусторонних отношений, поскольку должны свидетельствовать о тесной связи стран на международной арене.

## Формула для расчета модулярности

$$Q = \frac{1}{2m} \sum_{i,j} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

$m$  — общее количество весов всех ребер в графе (если граф не взвешенный, то это просто количество ребер).

$A_{ij}$  — вес ребра между узлами  $i$  и  $j$  (если граф не взвешенный, то 1, если есть ребро, и 0, если нет).

$k_i$  — степень узла  $i$  (общее количество соединений узла  $i$ ).

$\delta(c_i, c_j)$  — 1, если узлы  $i$  и  $j$  находятся в одном сообществе, и 0, если в разных.

# Основные этапы Лувенского метода

- **Инициализация**
  - Каждый узел сети рассматривается как отдельный кластер.
- **Оптимизация модулярности**
  - Узлы перемещаются между кластерами, чтобы максимизировать *модулярность* — показатель, который оценивает плотность связей внутри кластеров по сравнению с межкластерными связями.
- **Агрегация**
  - После оптимизации модулярности на локальном уровне, каждый кластер представляется как единый узел, и процесс повторяется для новой сети, состоящей из этих агрегированных узлов.
- **Повторение**
  - Процесс продолжается до тех пор, пока модулярность не перестанет улучшаться.

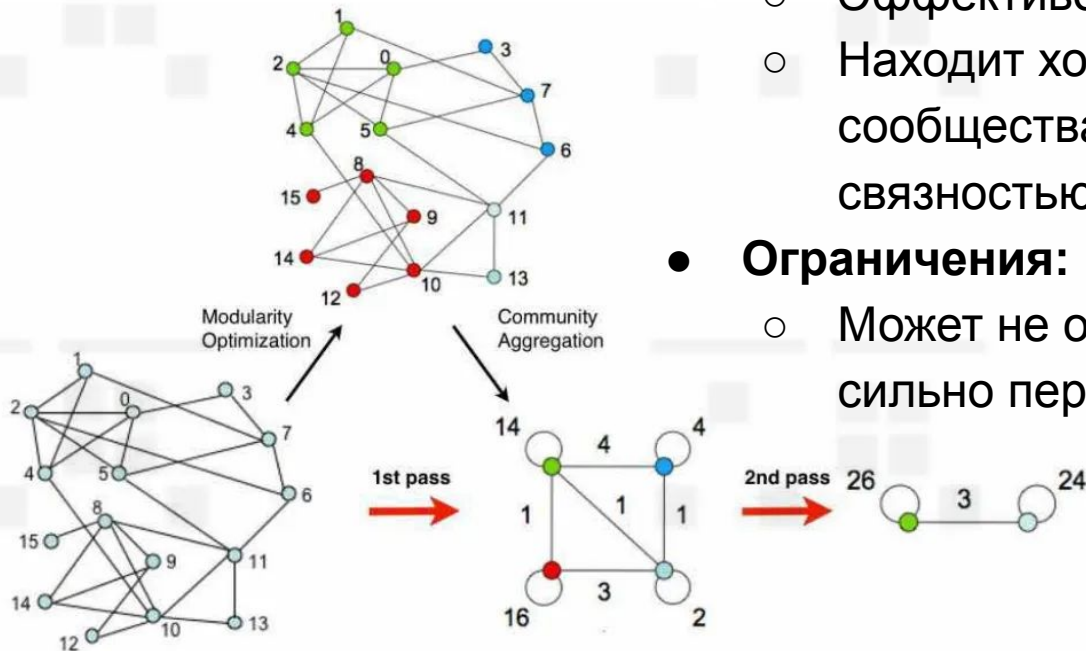
# Преимущества и ограничения

- **Преимущества:**

- Эффективен для больших графов.
- Находит хорошо определенные сообщества с высокой внутренней связностью.

- **Ограничения:**

- Может не обнаруживать мелкие или сильно перекрывающиеся сообщества.



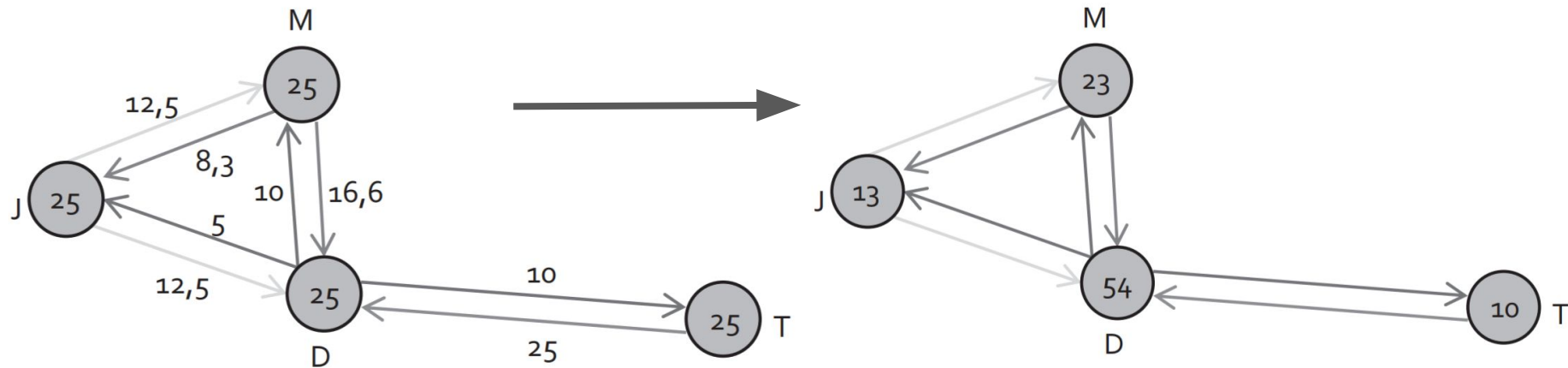
# PageRank

**PageRank** — это алгоритм, который используется для определения наиболее значимых узлов в сети на основе структуры связей.

Главные факторы важности узлов:

- **Число ссылок** — это количество ссылок, которое не учитывает качество ссылающихся сайтов (необъективность к новым ссылкам).
- **Качество ссылок** — это учет того, насколько авторитетны страницы, которые ссылаются на вашу страницу.
- **Переходы по ссылкам** — это реальные действия пользователей, которые показывают, насколько интересен сайт.

# Демонстрация силы ссылок



После многократных перераспределений число пользователей на каждом сайте стабилизируется, и их количество соответствует значению PageRank: чем больше пользователей он привлечет, тем выше его ранг.



# Интерпретация и визуализация

**ТОП-10 фильмов по кассовым  
сборам в 2015 г., млн.\$**



**ТОП-10 фильмов по кассовым  
сборам в 2015 г., млн.\$**



