

# Схема отношений и алгоритм PageRank

Лекция 7

# Составление схемы отношений

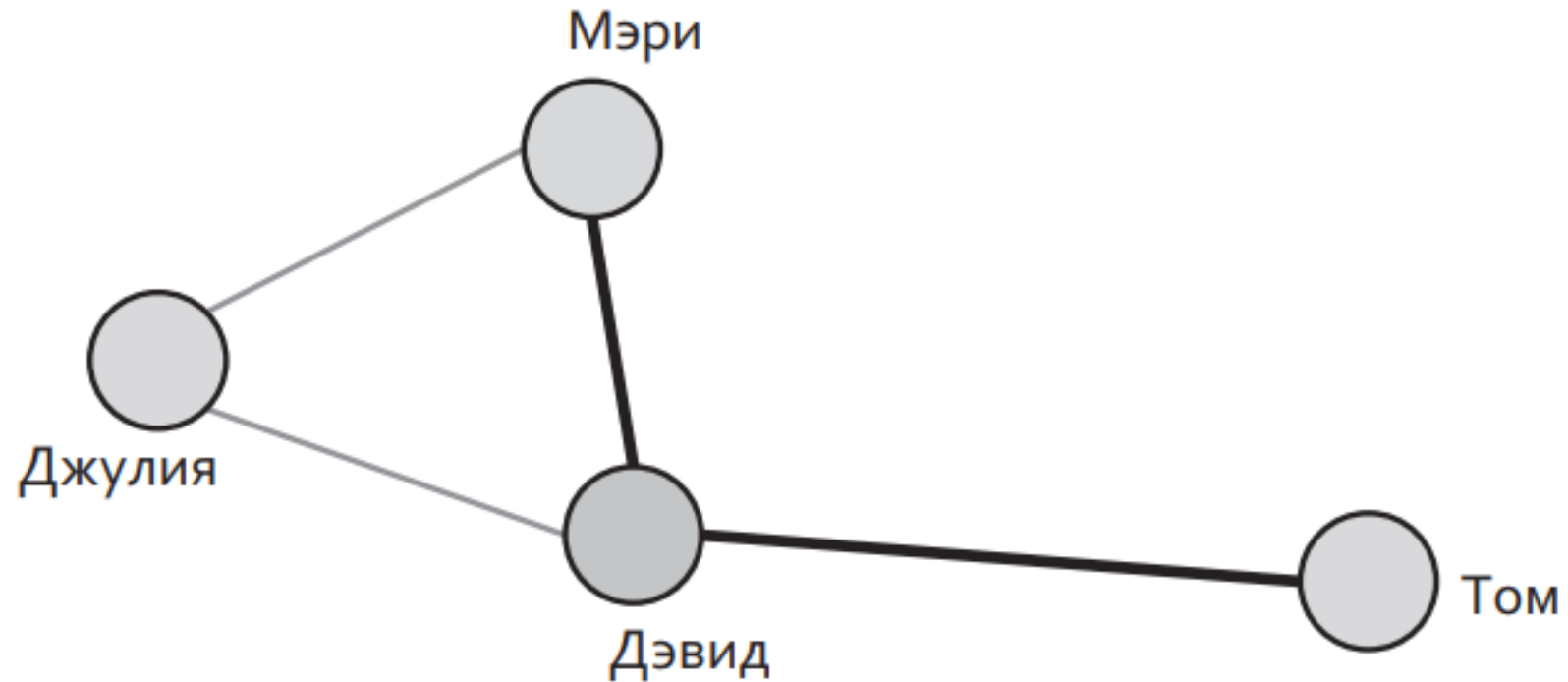
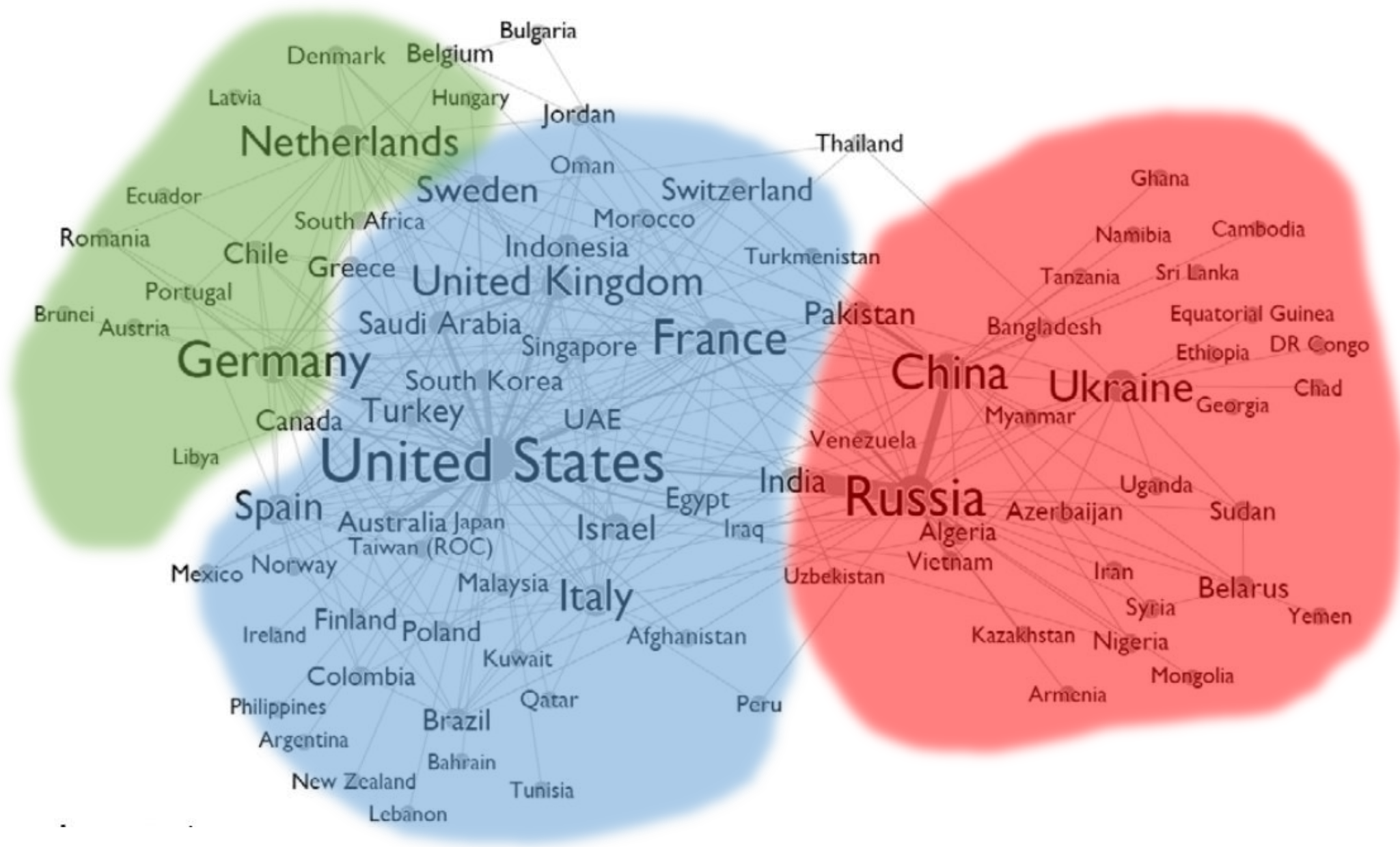


Рис. 1. Простая сеть друзей. Более близкие отношения показаны утолщенными линиями

# Пример: геополитика в торговле оружием

- стоимость оружия на уровне цен 1990 года в долларах США
- сделки, сумма которых превысила 100 млн долларов
- рассмотрели 10-летний период, с 2006 по 2015 год
- построили сеть из 91 узла и 295 ребер.



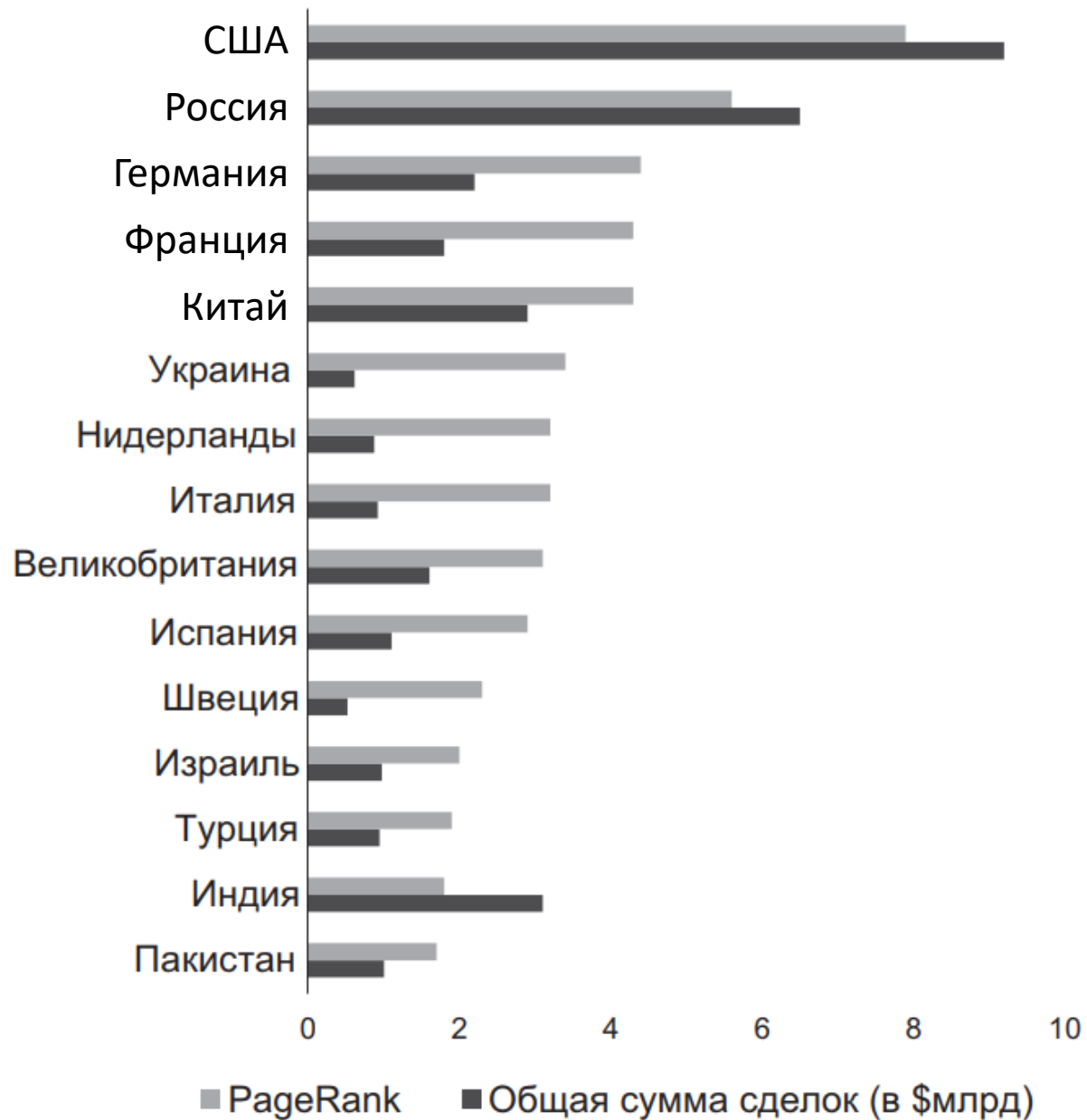
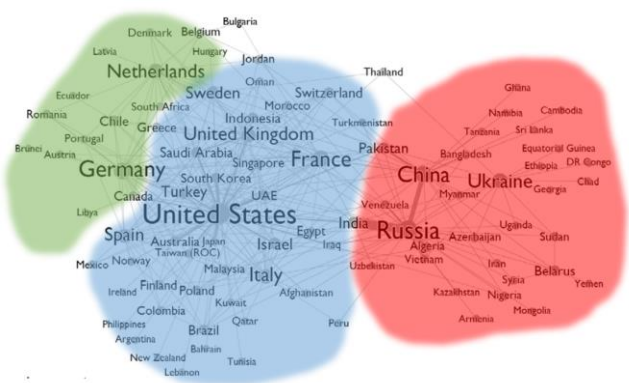


Рис. 3. Самые влиятельные страны в торговле оружием, согласно алгоритму PageRank. Значение PageRank для каждой страны показано светлым, а торговый объем — темным



# Лувенский метод

- **Стадия 0:** рассматривает каждый узел в качестве кластера, то есть начинает с числа кластеров, равного числу узлов.
- **Стадия 1:** меняет кластерное членство узла, если это приводит к улучшению модулярности. Если модулярность больше нельзя улучшить, узел остается на месте. Это повторяется для каждого узла до тех пор, пока изменения кластерного членства не будут исчерпаны.
- **Стадия 2:** строит грубую версию сети, в которой каждый кластер, найденный на стадии 1, представлен отдельным узлом, и объединяет бывшие межкластерные соединения в утолщенные ребра этих новых узлов в соответствии с их весом.
- **Стадия 3:** повторяет стадии 1 и 2 до тех пор, пока не закончатся дальнейшие изменения членства и размера связей.

# Лувенский метод: ограничения

- Важные, но малые кластеры могут быть поглощены
- Множество возможных кластерных конфигураций

# Алгоритм PageRank

Значение PageRank для веб-сайта определяется тремя факторами.

- **Число ссылок.** Если на один веб-сайт ссылаются другие, то он, скорее всего, привлекает больше пользователей.
- **Сила ссылок.** Чем чаще переходят по этим ссылкам, тем больше трафик сайта.
- **Источник ссылок.** Ранг веб-сайта повышается и оттого, что на него ссылаются другие высокоранговые сайты.



# Алгоритм PageRank

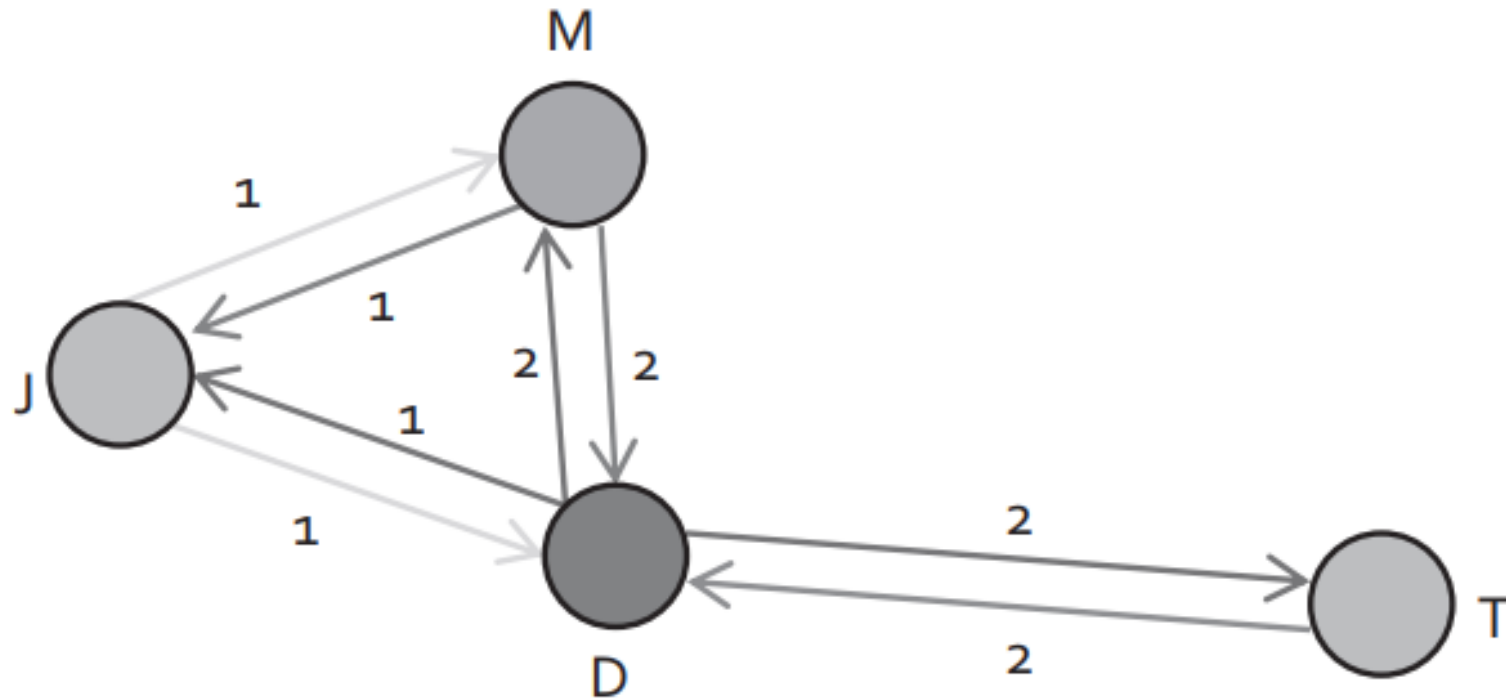


Рис. 4. Сеть, в которой узлы — это веб-сайты, а ребра — гиперссылки

# Алгоритм PageRank

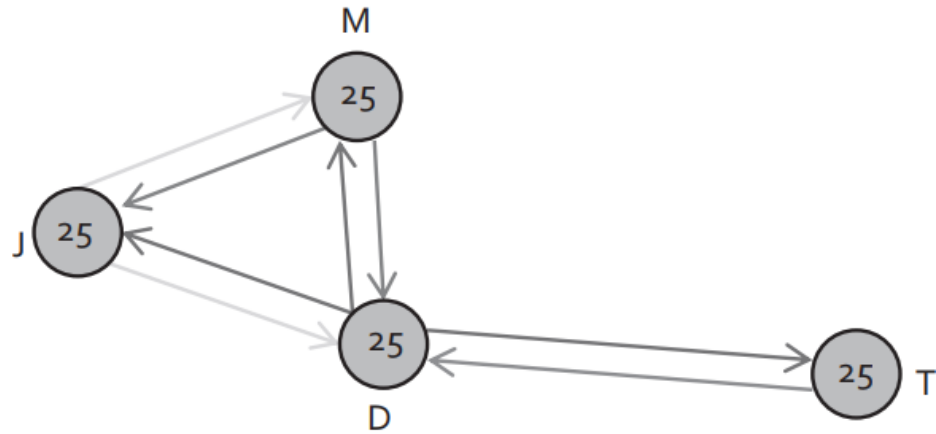


Рис. 5. Начальное положение, в котором 100 пользователей распределены по четырем веб-сайтам

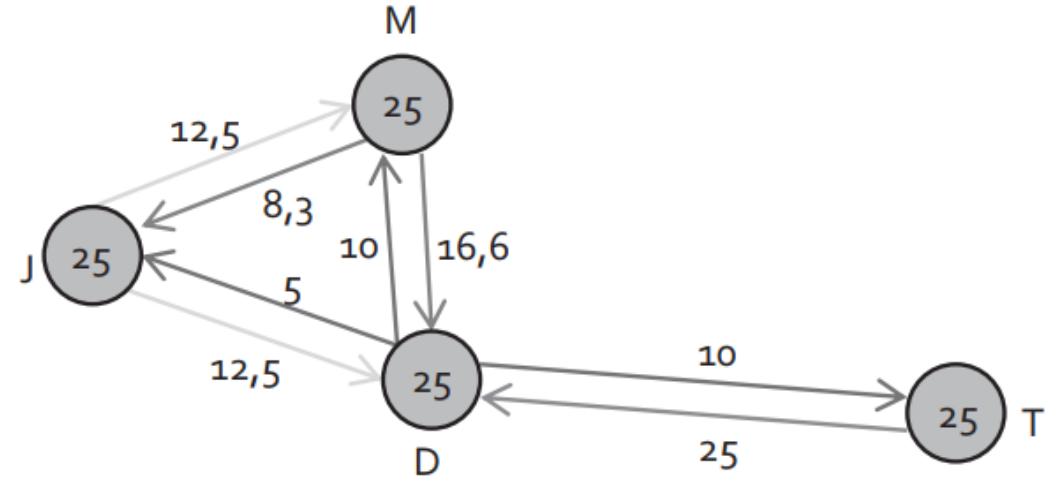


Рис. 6. Перераспределение пользователей, основанное на силе исходящих ссылок

# Алгоритм PageRank

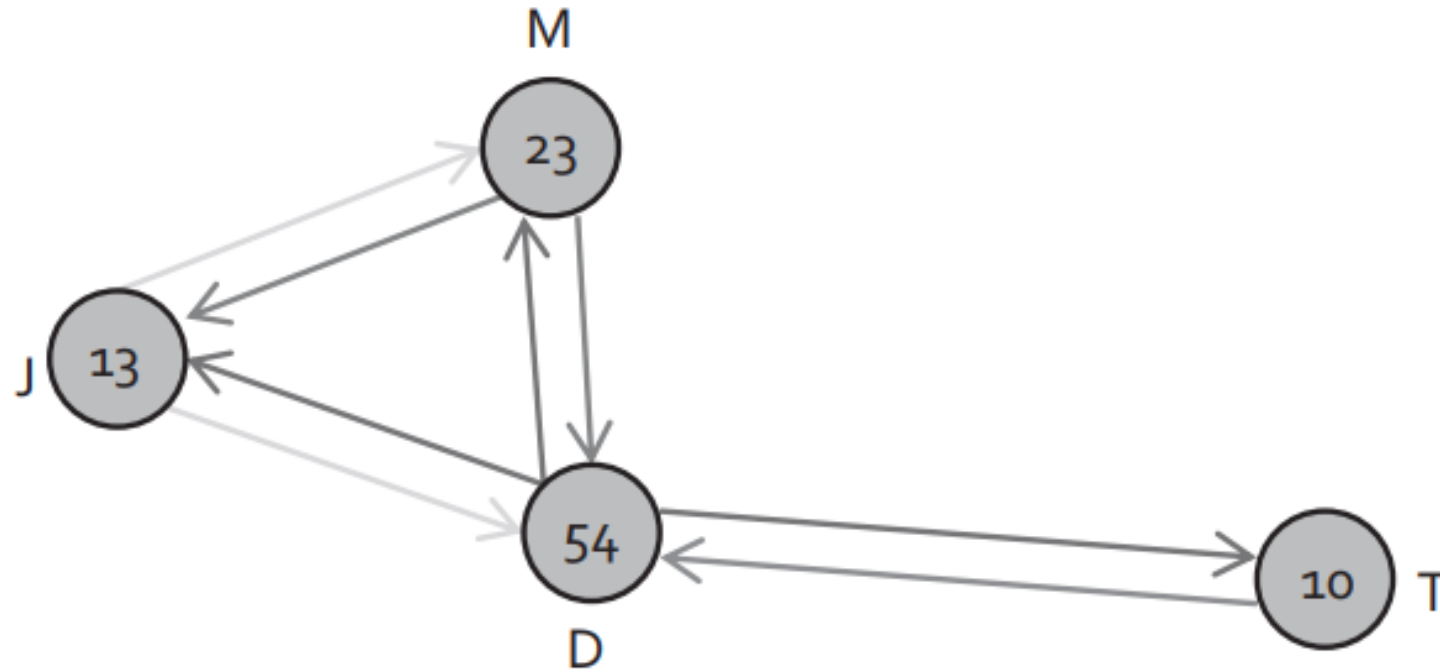


Рис. 7. Число пользователей на каждом веб-сайте после распределения

# Алгоритм PageRank: недостаток

- Необъективность в отношении старых узлов

# Ограничения

- Игнорирование дипломатических отношений при отсутствии покупок вооружения
- Игнорирование других соображений

# Краткие итоги

- Анализ социальных сетей — это метод, позволяющий строить схему и анализировать отношения между сущностями.
- Лувенский метод определяет кластеры внутри сети тем способом, который максимизирует взаимодействие внутри кластеров и минимизирует — между. Он лучше работает, когда кластеры имеют сходный размер и дискретны.
- Алгоритм PageRank ранжирует узлы в сети, исходя из числа ссылок, а также из их силы и источника