Оглавление

5.1. Составление схемы отношений

Большинство из нас имеет множество кругов общения, включающих такие категории людей, как родственники, коллеги или одноклассники. Чтобы выяснить, как устроены отношения всех этих людей, определив, например, активных персон и то, как они влияют на групповую динамику, мы можем воспользоваться методом под названием *анализ социальных сетей* (Social Network Analysis). Этот метод можно применять в вирусном маркетинге, моделировании эпидемий и даже для стратегий в командных играх. Тем не менее он больше известен своим использованием для анализа отношений в социальных сетях, что и дало ему название. На рис. 1 пример того, как анализ социальных сетей показывает отношения.

Рисунок 1 показывает сеть из четырех индивидов, также известную как *граф*, в котором каждый из этих персон представлен *узлом* (node). Отношения между узлами представлены линиями, называемыми *ребрами* (edges).

Каждое ребро может иметь *вес*, показывающий силу отношений.

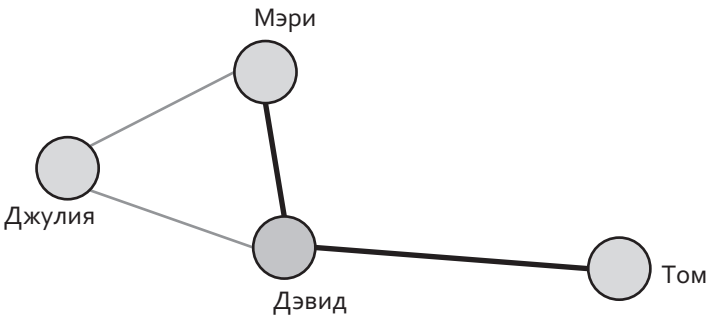

Рис. 1. Простая сеть друзей. Более близкие отношения показаны утолщенными линиями

Из рис. 1 мы можем заключить:

- Дэвид имеет больше всех связей, будучи знакомым с остальными тремя персонами;
- Том не знает никого, кроме Дэвида, с которым они близкие друзья;
- Джулия знает Мэри и Дэвида, но не близка с ними.

Кроме отношений анализ социальных сетей может строить схемы и для других сущностей, при условии, что между ними есть связи. В этой главе мы воспользуемся им для анализа международной сети торговли оружием, чтобы выявить доминирующие силы и их сферы влияния.

5.2. Пример: геополитика в торговле оружием

Мы получили данные о двусторонних трансферах основных видов обычных вооружений из *Стокгольмского международного института по исследованию проблем мира*. Военные поставки были выбраны в качестве косвенного показателя двусторонних отношений, поскольку должны свидетельствовать о тесной связи стран на международной арене.

В этом анализе мы стандартизировали стоимость оружия на уровне цен 1990 года в долларах США, после чего приняли в расчет только сделки, сумма которых превысила 100 млн долларов. Чтобы учесть флуктуации в торговле оружием, обусловленные производственными циклами новых технологий, мы рассмотрели 10-летний период, с 2006 по 2015 год, построив сеть из 91 узла и 295 ребер.

Для визуализации сети использовался *силовой алгоритм* (force-directed algorithm): узлы без связей отталкиваются друг от друга, а связанные узлы, наоборот, притягиваются с той степенью близости, которая отражает силу их связи (рис. 2). Например, максимальный объем торговли зафиксирован между Россией и Индией ($ 22,3 млрд), поэтому эти государства соединены толстой линией и близко расположены.

После анализа получившейся сети лувенским методом (Louvain Method, описан в следующем разделе) геополитические альянсы были сгруппированы в три кластера.

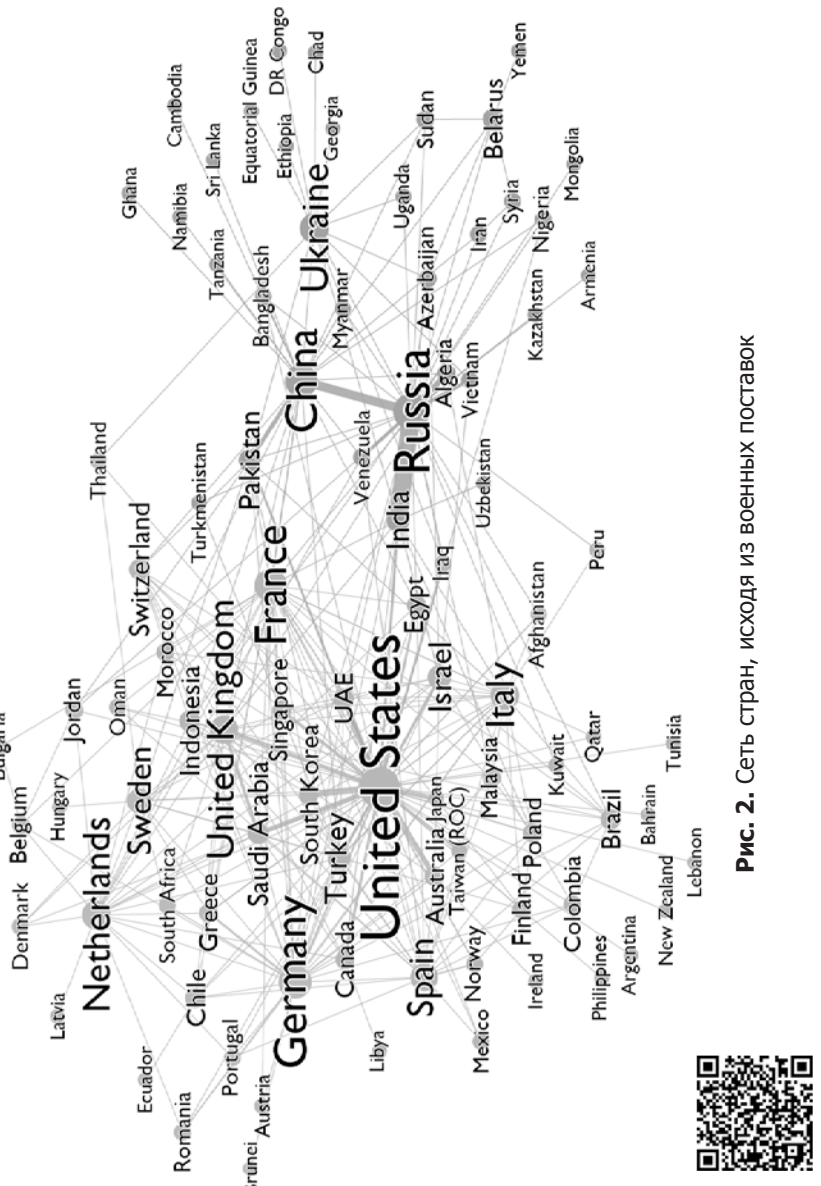

Рис. 2. Сеть стран, исходя из военных поставок

- **Светло-серый**: это крупнейший кластер, в котором доминируют США и который включает их союзников, таких как Великобритания и Израиль.
- **Светлый**: в нем лидирует Германия, и он включает в основном европейские страны, а также тесно связан со светло-серым кластером.
- **Темный**: в этом кластере доминируют Россия и Китай, он дистанцирован от двух других и включает в основном азиатские и африканские государства.

Кластеры отражают геополитические реалии XXI столетия, такие как долгосрочные альянсы между западными нациями, поляризацию между демократическими и коммунистическими странами и возрастающую роль противостояния между США и Китаем.

Кроме группировки в кластеры мы также проранжировали отдельные страны по уровню их влияния, воспользовавшись алгоритмом PageRank (описывается дальше). На рис. 3 представлены 15 самых влиятельных государств, которые также отмечены на рис. 2 более крупными узлами и подписями.

Согласно нашему анализу, в пятерку самых влиятельных стран входят США, Россия, Германия, Франция и Китай. Эти результаты подтверждаются тем обстоятельством, что четыре из пяти этих государств имеют влияние еще и как члены Совета Безопасности ООН.

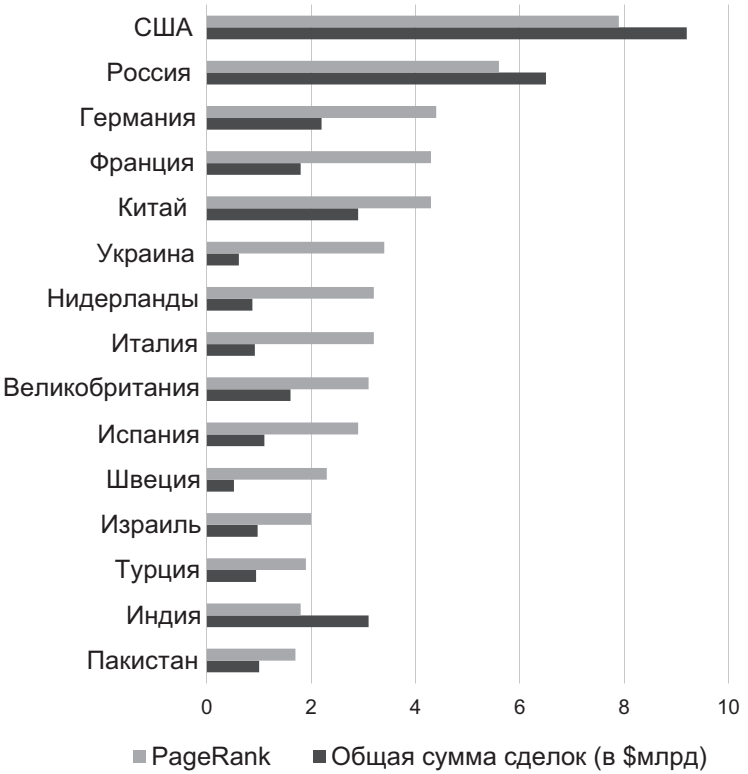

Рис. 3. Самые влиятельные страны в торговле оружием, согласно алгоритму PageRank. Значение PageRank для каждой страны показано светлым, а торговый объем — темным

В следующих разделах мы рассмотрим методы, использованные для выделения кластеров и ранжирования стран.

5.3. Лувенский метод

Как видно на рис. 2, можно найти кластеры сети путем группировки узлов. Изучение этих кластеров поможет понять, чем различаются части сети и как они соединены.

Лувенский метод — один из способов определения кластеров сети. Он подбирает различные кластерные конфигурации, чтобы: 1) максимизировать число и силу связей между узлами в одном кластере; 2) минимизировать при этом связи между узлами различных кластеров. Степень удовлетворения этим двум условиям известна как *модулярность* (modularity), и более высокая модулярность — признак более оптимальных кластеров.

Чтобы получить оптимальную конфигурацию кластеров, лувенский метод итеративно проходит следующие стадии.

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.

Таким образом, лувенский метод помогает нам выявить более значимые кластеры, начав с обнаружения малых из них, а затем при необходимости соединяя их.

Простота и эффективность делают лувенский метод популярным решением для кластеризации сети. Однако он имеет свои ограничения.

Важные, но малые кластеры могут быть поглощены. Итеративный процесс слияния кластеров может привести к тому, что значимые, но небольшие кластеры будут обойдены вниманием. Чтобы избежать этого, мы можем при необходимости проверять идентифицированные кластеры на промежуточных фазах итераций.

Множество возможных кластерных конфигураций. Для сетей, содержащих перекрывающиеся или вложенные кластеры, определить оптимальное кластерное решение может оказаться трудным. Тем не менее, когда имеются несколько решений с высокой модулярностью, мы можем сверить кластеры с другими информационными источниками, что мы и проделали на рис. 2, приняв во внимание географическое местоположение и политическую идеологию.