

Коллин, Массимилиано, Айрис: спасибо.

– Филиппо Менцер

Моим родителям и брату.

– Санто Фортунато

Лиз, Джина, Мэри Джо и Джей: ваша любовь и поддержка значат для меня все.

– Клейтон Дэвис

Содержание

Предисловие	10
Признательности	16
Введение	17
0.1 Социальные сети.....	18
0.2 Коммуникационные сети	21
0.3 Всемирная паутина и «Википедия»	24
0.4 Интернет	26
0.5 Транспортные сети.....	27
0.6 Биологические сети.....	29
0.7 Резюме.....	30
0.8 Дальнейшее чтение.....	31
Упражнения.....	32
1 Сетевые элементы	34
1.1 Базовые определения	34
1.2 Манипулирование сетями в исходном коде	36
1.3 Плотность и разреженность	39
1.4 Подсети	42
1.5 Степень.....	43
1.6 Направленные сети	44
1.7 Взвешенные сети	45
1.8 Многослойные и темпоральные сети	46
1.9 Представления сетей	49
1.10 Рисование сетей.....	51
1.11 Резюме.....	52
1.12 Дальнейшее чтение.....	53
Упражнения.....	53
2 Малые миры	58
2.1 Рыбак рыбака видит издалека.....	58
2.2 Пути и расстояния	62
2.3 Соединенность и компоненты	67
2.4 Деревья.....	69

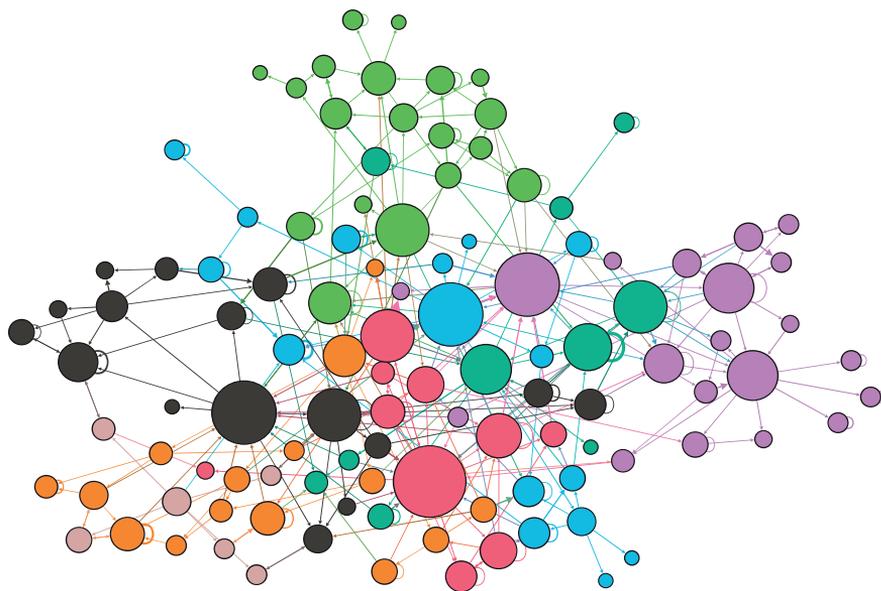
2.5	Отыскание кратчайших путей	71
2.6	Социальное расстояние.....	75
2.7	Шесть степеней сепарации.....	78
2.8	Друг моего друга	81
2.9	Резюме.....	84
2.10	Дальнейшее чтение.....	85
	Упражнения.....	86
3	Хабы.....	94
3.1	Меры центральности	95
3.1.1	Степень	95
3.1.2	Близость.....	95
3.1.3	Промежуточность.....	96
3.2	Распределения значений центральности.....	99
3.3	Парадокс дружбы	104
3.4	Ультрамалые миры.....	107
3.5	Устойчивость.....	108
3.6	Разложение ядра	110
3.7	Резюме.....	112
3.8	Дальнейшее чтение.....	113
	Упражнения.....	113
4	Направления и веса	119
4.1	Направленные сети	119
4.2	Всемирная паутина	120
4.2.1	Краткая история Всемирной паутины	121
4.2.2	Как работает Всемирная паутина.....	122
4.2.3	Обходчики Всемирной паутины.....	124
4.2.4	Структура Всемирной паутины	127
4.2.5	Тематическая локальность.....	129
4.3	Метрика PageRank	132
4.4	Взвешенные сети	137
4.5	Информация и дезинформация.....	138
4.6	Сети совместной встречаемости.....	143
4.7	Весовая гетерогенность.....	147
4.7.1	Трафик Всемирной паутины	147
4.7.2	Фильтрация связей	149
4.8	Резюме.....	151
4.9	Дальнейшее чтение.....	153
	Упражнения.....	155
5	Сетевые модели.....	162
5.1	Случайные сети	162
5.1.1	Плотность	165
5.1.2	Степенное распределение.....	166

5.1.3	Короткие пути	168
5.1.4	Коэффициент кластеризации.....	169
5.2	Малые миры.....	170
5.3	Конфигурационная модель	174
5.4	Преференциальное прикрепление	177
5.5	Другие преференциальные модели	182
5.5.1	Модель на основе привлекательности	184
5.5.2	Модель на основе приспособленности	185
5.5.3	Модель на основе случайного блуждания	187
5.5.4	Модель на основе копирования.....	190
5.5.5	Модель на основе ранга	191
5.6	Резюме.....	193
5.7	Дальнейшее чтение.....	194
	Упражнения.....	194
6	Сообщества	200
6.1	Базовые определения	203
6.1.1	Переменные сообщества	203
6.1.2	Определения сообщества	205
6.1.3	Разделы.....	207
6.2	Смежные проблемы	209
6.2.1	Деление сети на разделы.....	209
6.2.2	Кластеризация данных	212
6.3	Обнаружение сообществ.....	215
6.3.1	Устранение мостов.....	216
6.3.2	Оптимизация модулярности	218
6.3.3	Распространение меток.....	225
6.3.4	Стохастическое блочное моделирование	227
6.4	Оценивание методов	230
6.4.1	Искусственные эталоны.....	230
6.4.2	Реально существующие эталоны.....	233
6.4.3	Сходство между разделами.....	234
6.5	Резюме.....	236
6.6	Дальнейшее чтение.....	237
	Упражнения.....	238
7	Динамика.....	244
7.1	Идеи, информация, влияние	246
7.1.1	Пороговые модели	247
7.1.2	Независимо-каскадные модели	250
7.2	Распространение эпидемий.....	252
7.2.1	Модели SIS и SIR	254
7.2.2	Распространение слухов.....	259
7.3	Динамика мнений	261
7.3.1	Дискретные мнения	262
7.3.2	Непрерывные мнения	265

7.3.3	Козволюция сетей и динамика	267
7.4	Поиск	270
7.4.1	Локальный поиск.....	270
7.4.2	Доступность поиска	273
7.5	Резюме.....	278
7.6	Дальнейшее чтение.....	280
	Упражнения.....	281
Приложение А. Руководство по языку Python.....		288
A.1	Блокнот Jupyter.....	288
A.2	Условный блок	289
A.3	Списки.....	290
A.4	Циклы	292
A.5	Кортежи.....	295
A.6	Словари	297
A.7	Комбинирование типов данных	300
A.7.1	Список кортежей	300
A.7.2	Список словарей.....	301
A.7.3	Словарь словарей.....	302
A.7.4	Словарь с кортежными ключами.....	302
A.7.5	Еще один словарь словарей	303
Приложение В. Модели NetLogo.....		305
B.1	Модель PageRank.....	306
B.2	Гигантская компонента.....	307
B.3	Малые миры.....	308
B.4	Преференциальное прикрепление	309
B.5	Вирус в сети.....	310
B.6	Изменение языка.....	312
Справочные материалы		314
Предметный указатель.....		331

Предисловие

Сети присутствуют во всех аспектах нашей жизни: круг друзей, коммуникационные и транспортные сети, а также Веб как Всемирная паутина – все это примеры, которые мы воспринимаем внешне, тогда как нейроны в нашем мозге и белки в нашем теле образуют сети, которые определяют наш интеллект и выживание. Когда люди общаются в Facebook или Twitter, покупают что-то в Amazon, ищут в Google или покупают авиабилет, чтобы навестить семью, они используют сети, не осознавая того. Сегодня базовое понимание сетевых процессов требуется в различных сферах деятельности – от технологий до маркетинга, от менеджмента до дизайна, от биологии до искусства и гуманитарных наук. В этом учебнике проводится разведывательный анализ учения о сетях и то, как сети помогают нам понимать сложные шаблоны взаимоотношений, которые формируют наши жизни.



Эта книга тоже является сетью! На приведенном выше рисунке показаны взаимоотношения между главами, разделами и подразделами. Связи представляют и иерархическую структуру книги (как показано в Оглавлении), и перекрестные ссылки между главами, разделами,

рисунками, таблицами, уравнениями и вставками. Цвета узлов представляют главы, а размер узла пропорционален числу соседей.

Зачем нужен «Вводный курс» по науке о сетях?

Это не первая книга о науке о сетях – на самом деле есть несколько отличных книг на выбор, и мы перечислим некоторые из них в главе 1. Мы преподаем эти темы уже в течение нескольких лет в Университете Индианы для широкой аудитории студентов старших курсов по информатике, теории вычислительных машин, науке о данных, теории информации, бизнесу, естественным и социальным наукам. Этот опыт научил нас тому, что студенты стремятся «пачкать свои руки» выполнением черновой работы и писать исходный код, чтобы понимать и использовать сети в интересующих их областях применения, даже если они только учатся программировать и не имеют математического и компьютерного образования за пределами средней школы и курсов начального уровня колледжей. Поэтому мы разработали широкий круг учебно-практических занятий и задач, как теоретических, так и вычислительных, предоставив студентам обилие практических занятий по науке о сетях. Используя такой подход, книга знакомит с сетями широкую аудиторию студентов, не имеющих никаких технических предпосылок, кроме какого-то вводного программирования и готовности учиться на деле. Это делает наш учебник пригодным для «вводного курса» науки о сетях.

Синописис

После проведения обзора сетей, существующих во многих областях человеческих знаний и деятельности, мы поговорим о социальных сетях, которые знакомы студентам больше всего. Это позволяет ввести такие понятия, как маломировое свойство (короткие пути) и кластеризация (треугольники и транзитивность). Указанные темы объясняются с использованием увлекательных учебных занятий, таких как игра *«Шесть степеней Кевина Бейкона»*. Затем мы погрузимся в роль хабов, используя Парадокс дружбы, и обсудим тему устойчивости сетей. Далее мы вводим соответственно направленные и взвешенные сети. Всемирная паутина, «Википедия», цитирование, трафик и Twitter используются для иллюстрации роли направления и веса. Последние три главы охватывают более сложные темы, а именно модели возникновения сетей, методы обнаружения сообществ и динамические процессы, происходящие в сетях.

В каждой главе рассматриваются базовые концепции, необходимые для понимания фундаментального аспекта сетей; избегаются

сложные темы и формализм. Когда это полезно, мы включаем немного математики во вставки, обрамленные рамкой. В них находится чуть-чуть более технический материал, и его можно пропустить без потери базового понимания темы. Но студенты, которые будут следить за этими дополнительными примечаниями, смогут получить более глубокое понимание материала. Каждая глава включает в себя учебно-практические занятия по программированию и упражнения, позволяющие читателям применять и проверять свои знания с помощью практических занятий по строительству и анализу сетей. Указанные учебно-практические занятия основаны на примерах реально существующих сетей, которые используются для иллюстрации концепций на протяжении всей книги. И учебно-практический исходный код, и сетевые данные доступны в репозитории книги на GitHub¹.

Целевая аудитория

С ростом популярности и коммерческого успеха онлайн-социальных сетей многие студенты заинтересованы в том, чтобы узнать немного о том, что находится «под капотом» таких сетей. Данный учебник предназначен для всех этих студентов в основном на уровне бакалавриата, хотя книга, возможно, будет полезна и для вводных курсов аспирантуры в нетехнических областях. Студенты, обучающиеся по таким программам, как наука о данных, информатика, бизнес, теория вычислительных машин, машиностроение, теория информации, биология, физика, статистика и социальные науки, получают пользу от курсов, основанных на этом учебнике. Их интерес будет достаточно велик, чтобы изучить науку о сетях глубже, и, возможно, они выберут карьеру, которая поможет найти им работу в Google, Facebook, Twitter или организовать свой собственный сетевой стартап.

Педагогика

Настоящий курс не требует никакого технического образования в области математики или программирования, что делает эту книгу пригодной для вводных курсов любого уровня, включая курсы сетевой грамотности и грамотности в программировании. Подобного рода курсы могут пропускать математические вставки. Отрабатывая учебно-практические занятия по программированию в коллаборативной вычислительной лаборатории и назначая упражнения по программированию, преподаватели предоставят студентам возможность приобрести технические навыки, достаточные для выполнения задач

¹ См. github.com/CambridgeUniversityPress/FirstCourseNetworkScience.

анализа данных, связанных с сетями. Таков наш подход в Университете Индианы, где мы преподаем материал данной книги в течение двух курсов: первый вводный курс, предназначенный для студентов-второкурсников / младших курсов, которые прошли или проходят курсы конкурентного программирования на Python; и второй курс, предназначенный для студентов младших/старших курсов. Первый курс примерно охватывает материал глав с 0 по 4. Второй курс сосредоточен на главах 5–7 после расширенного обзора и нескольких более продвинутых учебно-практических занятий по предыдущему материалу.

Обширные учебно-практические занятия по программированию и упражнения позволяют преподавателям легко руководить учебным процессом и проводить практические мероприятия, а также позволяют студентам укрепить и проверить свое понимание сетевых концепций. Мероприятия включают учебно-практические занятия по *NetworkX*, широко распространенной библиотеке для сетевой аналитики; и по всем затронутым в книге темам, от базовых упражнений до передовых методов. Например, на одном учебно-практическом занятии студенты знакомятся с шагами извлечения данных социальных сетей из Всемирной паутины. Используя интерфейс прикладного программирования (API) Twitter, студенты смогут анализировать популярные темы, выявлять влиятельных пользователей и реконструировать сети диффузии информации, показывающие процесс онлайн-распространения хештегов. Студенты, которые проходят учебно-практические занятия и выполняют упражнения по программированию, наберутся опыта в строительстве, импортировании/экспортировании, анализировании, манипулировании и визуализировании сетей любого типа.

Учебно-практические занятия основаны на языке Python, т. е. самом популярном языке для написания скриптов/программ. Учебное руководство, в котором рассматриваются главные концепции программирования на Python, включен в приложение А книги. Все учебно-практические занятия доступны онлайн в виде блокнотов Jupyter/IPython. Со временем библиотека *NetworkX* (и даже язык Python), возможно, эволюционирует, и, возможно, потребуется обновить часть исходного кода книги. Мы будем отмечать такие обновления в репозитории книги на GitHub.

Разумеется, для программирования сетей существуют и другие библиотеки, например *igraph*, *SNAP* и *graph-tool*. Наш выбор библиотеки *NetworkX* основан на том факте, что она написана на чистом Python, что облегчает отладку для студентов, знакомых с Python. Многие альтернативы имеют интерфейсы Python, но написаны на языке C, что делает их эффективнее, но и сложнее в отладке.

Наконец, в некоторых главах используются интерактивные модели для демонстрации сетевых явлений, таких как гигантские компоненты, малые миры, алгоритм PageRank, предпочтительное прикрепле-

ние и эпидемии. Эти модели работают в NetLogo, популярной симуляционной платформе. Учебно-практический материал по NetLogo и несколько наиболее актуальных моделей представлены в приложении В книги.

Об обложке

Сеть на обложке, сгенерированная Онуром Варолом (Onur Varol) (Феррара и соавт., 2016), изображает диффузию хештега #SB277 в Twitter. Этот хештег относится к калифорнийскому закону 2015 года об обязательствах в отношении вакцинации и освобождении от нее, и указанная сеть изображает обсуждение, которое проходило онлайн среди сторонников и противников указанного законопроекта. Узлы изображают пользователей Twitter, а связи показывают информацию, распространяемую среди пользователей через ретвиты. Размер узла отражает влияние учетной записи (сколько раз пользователь ретвитнул), а цвета узлов иллюстрируют баллы ботов: красные узлы, скорее всего, являются учетными записями ботов, синие узлы, скорее всего, являются людьми.

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге, – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Скачивание исходного кода примеров

Скачать файлы с дополнительной информацией для книг издательства «ДМК Пресс» можно на сайте www.dmkpress.com на странице с описанием соответствующей книги.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг, мы будем очень благодарны, если вы сообщите о ней главному редактору по адресу dmpress@gmail.com. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательства «ДМК Пресс» и Manning Publications очень серьезно относятся к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу электронной почты dmpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

Признательности

Первоначальная идея этой книги возникла из бесед с нашим бывшим коллегой Алексом Веспиньяни. На протяжении многих лет наши коллеги Сандро Фламмини, Йи Ан и Филиппо Радиччи давали ценные советы. Несколько студентов оказало помощь в преподавании наших курсов науки о сетях в Университете Индианы. Среди них мы хотим отметить Майка Коновера, который первым задумал некоторые упражнения, описанные на страницах данной книги. Мы также благодарны коллегам, которые предоставили отзывы о ранних набросках книги, в особенности Клаудио Кастеллано, Чато Кастильо и нескольким анонимным рецензентам.

Мы благодарим наших замечательных сотрудников, студентов, аспирантов и посетителей: Ану Магитман, Бена Маркинеса, Бруно Гонсалвеса, Ченчен Шао, Дип Тхи Хоанга, Димитара Николова, Эмилио Феррару, Джованни Луку Чампалью, Якоба Раткевича, Джаслин Каур, Хосе Рамаско, Кай-Чен Яна, Кариссу Маккелви, Кадзу Сасахару, Ле-Шин Ву, Лилиан Венг, Луку Марию Алелло, Марка Мейса, Маркуса Якобссона, Михая Аврама, Никола Перра, Онура Варола, Пик-Май Хуэй, Прашанта Ширалкара, Пшемека Грабович, Руж Акавипата, Сяодан Лу, Сяолин Сун, Зохера Качвалу и многих других среди прочих. Это удивительно яркая и веселая группа людей, которые внесли свой значительный вклад в идеи, наборы данных и приводимые к книге иллюстрации.

Наша работа была бы невозможна без поддержки многих преданных своему делу сотрудников Центра исследований сложных сетей и системных исследований, Школы информатики, вычислительной техники и машиностроения и Института науки о сетях Университета Индианы. Прежде всего мы с благодарностью должны отметить Тару Холбрук, Мишель Домпке, Роба Хендерсона, Дейва Кули, Пэтти Мэбри, Энн Маккрейни, Вэла Пентчева, Мэтью Хатчинсона, Чатури Пели Канканамалаге и Бена Серретта. Спасибо также Нику Гиббонсу из издательства Кембриджского университета за его поддержку и отзывы.

Мы благодарны Арику Хагбергу, Питеру Сварту и Дэну Шульцу, авторам NetworkX, а также Ури Виленски и Центру связного обучения и компьютерного моделирования Северо-Западного университета за разработку и поддержку NetLogo.

Наконец, мы в огромном долгу перед нашими семьями, которые любят, поддерживают и терпят нас, даже когда мы работаем больше, чем должны.

Введение

Сеть: взаимосвязанная или взаимодействующая цепочка, группа или система.

Вообразите мир, в котором у людей нет друзей. Где дороги никогда не пересекаются. Где компьютеры не связаны между собой. Этот мир без сетей был бы очень грустным и скучным местом, где ничего не происходит, – и даже если бы что-то случилось, никто бы об этом не узнал. Такой мир невообразим, потому что наша жизнь полностью определяется сетями: взаимоотношениями, взаимодействиями, каналами связи и Всемирной паутиной. Биологические сети, управляющие взаимодействиями между генами в наших клетках, определяют наше развитие, нейронные сети в мозге наделяют нас возможностью думать, информационные сети направляют наши знания и культуру, транспортные сети позволяют двигаться, а социальные сети подпирают нашу жизнь.

Сети – это общий, но мощный способ представления и изучения простых и сложных взаимодействий. В этой книге проводится разведывательный анализ учения о сетях и того, как они помогают нам понимать закономерности соединений и взаимоотношений, которые формируют нашу жизнь. По своей сути сеть – это простейшее описание множества взаимосвязанных сущностей, которые мы называем *узлами*, и их соединений, которые мы называем *связями*. Сетевое представление является столь общим и мощным, потому что оно устраняет многие детали конкретной системы и фокусируется на взаимодействиях между ее элементами. Отсюда сети используются для изучения самых разнообразных систем. Узлы могут представлять все виды сущностей: людей, города, компьютеры, веб-сайты, концепции, клетки, гены, виды животных и т. д. Связи представляют взаимоотношения или взаимодействия между этими сущностями: дружеские связи между людьми, рейсы между аэропортами, пакеты, которыми компьютеры обмениваются в интернете, связи между страницами Всемирной паутины, синапсы между нейронами и т. д.

Прежде чем мы представим базовые понятия, определения и номенклатуру сетей, давайте рассмотрим несколько примеров социальных, инфраструктурных, информационных и биологических сетей. Данные для всех представленных здесь примеров доступны в репозитории книги на GitHub¹. Сети, на которых мы сосредоточимся

¹ См. github.com/CambridgeUniversityPress/FirstCourseNetworkScience.

в этой книге, как правило, являются крупными, хотя многому можно научиться, изучая и меньшие системы, такие как социальные сети, созданные на основе опросов или собеседований. В этих случаях имеет смысл детально проинспектировать отдельные узлы и соединения, тогда как анализ крупных сетей, как правило, фокусируется на макроскопических свойствах, классах узлов и связей, типичных проявлений поведения и аномалиях.

0.1. Социальные сети

Социальная сеть – это группа людей, связанных каким-либо типом взаимоотношения. Дружба, сотрудничество, романтика или простое знакомство – все это примеры социальных взаимоотношений, которые соединяют пары людей. Когда мы говорим о социальной сети, мы обычно думаем об определенном типе взаимоотношения. Человек представляется узлом в социальной сети, а взаимоотношение представляется связью между двумя людьми. Таким образом, сеть является представлением взаимоотношения. Это позволяет нам говорить о взаимоотношениях, описывать их и анализировать на уровне, выходящем за рамки пары людей.

Существует много разных типов социальных сетей, и их важно изучать. Медицинские работники анализируют сети сексуальных отношений, чтобы отыскивать способы борьбы с распространением заболеваний, передающихся половым путем. Экономисты изучают сети направления на работу для решения проблемы неравенства и сегрегации на рынках труда. А ученые инспектируют сети соавторства в научных публикациях, чтобы выявлять влиятельных мыслителей и идеи.

В наши дни мы используем веб-сайты онлайн-социальных сетей, чтобы отслеживать социальные связи. Такие платформы, как Facebook и Twitter, позволяют нам поддерживать связь со многими людьми – партнерами, друзьями, коллегами и знакомыми, иногда сотнями, – и комфортно с ними общаться независимо от расстояния. На рис. 0.1 показана сеть знакомств, часть социального графа Facebook. В этой сети узлами являются люди с учетной записью Facebook в университетах США, и соединения могут представлять различные типы взаимоотношения, от настоящей дружбы до простого знакомства. Просто взглянув на визуализацию сети, вы узнаете кое-что о лежащей в ее основе социальной структуре. У некоторых людей связей больше; мы представляем это, делая соответствующие узлы больше и темнее. Это могут быть популярные студенты, преподаватели или администраторы. Мы также замечаем, что сеть примерно поделена на две части. Данные анонимны, поэтому мы не можем сказать наверняка, но возможной интерпретацией будет то, что крупная подсеть

включает в себя в основном студентов старших курсов, а меньшая – в основном аспирантов. Между узлами в двух группах есть соединения, но их не так много, как между узлами внутри каждой группы. Другими словами, студенты старших курсов с большей вероятностью будут дружить с другими студентами, чем с аспирантами. Позже для всех этих наблюдений, которые типичны для большинства социальных сетей, мы введем формальные названия.

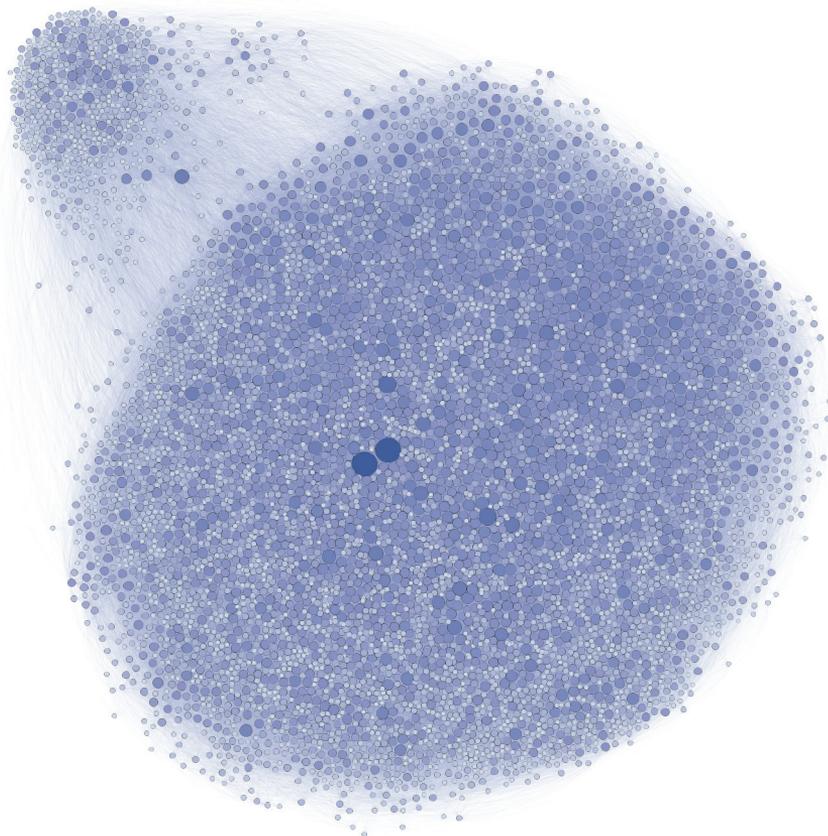


Рис. 0.1 Визуализация сети пользователей Facebook в Северо-Западном университете. Узлы обозначают людей, а связи обозначают соединения между друзьями в Facebook

Доступность данных из онлайн-социальных сетей очень увлекает ученых. Мы можем изучать человеческие взаимодействия в масштабе и в разрешающей способности, которые никогда не были возможны в прошлом: кто с кем дружит, кто на что обращает внимание, кому что нравится, что рекомендуется и как эта информация распространяется по сети. Эти данные предоставляют нам беспрецедентную возможность обнаруживать, отслеживать, добывать и моделировать

то, что делают люди. Подобно тому, как телескоп позволил нам впервые увидеть далекие планеты и звезды, а микроскоп – заглянуть в живые ткани и микроорганизмы, социальные сети позволяют изучать социальные системы и человеческую деятельность. Однако какими бы захватывающими ни были эти возможности для исследователей, они не обходятся без риска злоупотреблений. Онлайн-взаимодействия раскрывают нашу приватную информацию. Мы все слышали истории о том, как работодатели находили неловкие фотографии потенциальных сотрудников или о скандалах, имеющих отношение к хакерам и политическим организациям, собирающим данные о миллионах пользователей. Опасности бывают едва уловимыми. Обладание небольшим объемом информации о большом числе людей может раскрывать гораздо больше, чем предполагалось. Используя данные из Facebook, два студента Массачусетского технологического института обнаружили, что, просто взглянув на пол и сексуальность онлайн-друзей человека, они могут предсказывать, является этот человек геем или нет. Онлайн-социальные сети также облегчают выдачу себя за другого человека и затрудняют ее обнаружение. Выживание информации из социальных сетей (социальный фишинг) – это метод выставления себя за друга жертвы (логически выводимого из данных онлайн-социальной сети), чтобы побуждать жертву раскрывать конфиденциальную информацию. Два студента Университета Индианы продемонстрировали, что таким образом им удалось получить секретные пароли 72 % жертв.

Данные о социальной сети можно извлекать из многочисленных источников. Если мы хотим картировать шаблоны мобильности людей с целью улучшения городских транспортных сетей, то мы можем собирать данные о звонках с мобильных телефонов. Если хотим картировать соавторство среди ученых, то можем извлекать имена из базы данных научных публикаций; два соавтора одной и той же статьи будут связаны друг с другом. (Это не тривиальное упражнение, потому что у нескольких ученых могут быть общие имена.) Если мы хотим картировать сотрудничество между кинозвездами, можем извлекать данные о титрах кинофильмов из интернет-базы данных кинофильмов (Internet Movie Database, IMDb.com). На рис. 0.2 показаны две такие сети. В одном случае на самом деле существует два вида узлов: кинофильмы и актеры/актрисы. Мы проводим связь между актрисой и кинофильмом, в котором она снялась. В другом случае мы фокусируемся на связях между актерами/актрисами, которые снимались в фильмах вместе. Хотя изображенные сети улавливают лишь крошечные части базы данных кинофильмов, мы снова замечаем некоторые четкие закономерности. Более крупные узлы имеют больше соединений, представляющих звезд, которые снимались во многих кинофильмах. Мы также видим, что сети структурированы в несколько плотных групп, связанных с периодами, языками или жанрами кинофильмов: голливудские (синие), европейские (голубые), мексиканские (фиолетовые), китайские (желтые), филиппинские (оранжевые),

турецкие и восточноевропейские (зеленые), индийские (красные), греческие (белые) кинозвезды и кинозвезды фильмов для взрослых (розовые) на рис. 0.2(b). В главе 6 вы узнаете, как обнаружить эти группы и выяснить, чему они посвящены.

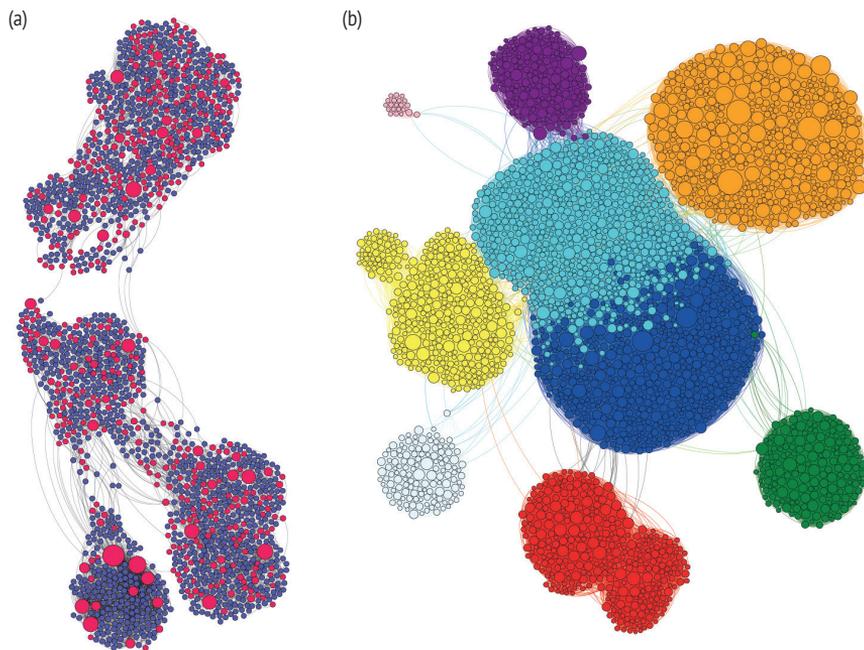


Рис. 0.2 (a) Сеть кинозвезд, основанная на небольшой выборке кинофильмов, актеров и актрис из интернет-базы данных кинофильмов. Узлы представляют кинофильмы (синие) или актеров/актрис (красные). Связь соединяет актера или актрису с фильмом, в котором они снимались. (b) Сеть кинозвезд, основанная на небольшой выборке актеров и актрис из интернет-базы данных кинофильмов, снимавшихся вместе с другой кинозвездой. Связь соединяет двух людей, которые снялись вместе по меньшей мере в одном фильме. Цвета представляют жанры фильмов или языки/страны

0.2. Коммуникационные сети

В сетях Facebook и кинофильмов связи взаимны: вы не можете подружиться с кем-либо на Facebook, если он не согласен, и вы не можете быть снятым в фильме, не будучи упомянутым в титрах. Однако не все социальные сети имеют взаимные связи. Например, Twitter представляет собой популярную социальную сеть со связями, которые не обязательно являются взаимными: Алиса может следить за Бобом без того, чтобы Боб обязательно следил за Алисой. Как следствие, запечатленные в сети Twitter отношения не являются дружбой; вы подписываетесь на кого-то, чтобы увидеть, что он публикует. Когда вы ретвитите сообщение, его видят ваши подписчики. Это хороший

способ широко обмениваться информацией, поэтому Twitter является социальной сетью, в основном направленной на распространение информации, т. е. коммуникационной сетью. Ретвитная сеть на рис. 0.3 иллюстрирует распространение политических сообщений во время выборов в США. Более крупные узлы являются узлами с большим числом исходящих связей, потому что число ретвитов пользователями другими пользователями является способом измерить их влияние. Вы, вероятно, сразу заметили более поразительную закономерность: консервативные пользователи (красные узлы) в основном ретвитят сообщения от других консерваторов, в то время как прогрессивные пользователи (синие узлы) аналогичным образом делятся прогрессивным контентом. На самом деле такие предпочтительные регулярности социальных связей позволяют нам с высокой точностью угадывать политические склонности человека. Это свойство, именуемое *гомофилией*, будет обсуждаться в главе 2; алгоритм определения политических предпочтений по структуре сети будет представлен в главе 6.

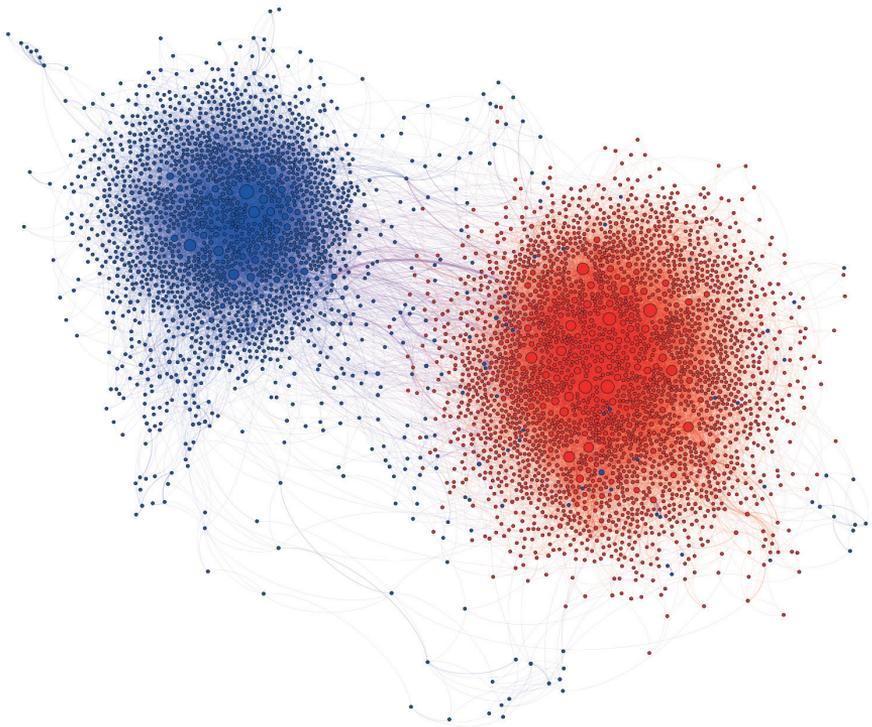


Рис. 0.3 Ретвитная сеть в Twitter среди людей, делящихся постами о политике США. Связи представляют собой ретвиты сообщений, в которых использовались хештеги, такие как #tscot и #p2, связанные соответственно с консервативными (красными) и прогрессивными (синими) сообщениями во время промежуточных выборов в США 2010 года. Когда Боб ретвитит Алису, мы рисуем направленную связь от Алисы к Бобу, чтобы обозначить, что сообщение перешло от нее к нему. Направление связей не показано

Такие сети, как Twitter, позволяют нам отслеживать диффузию хештегов и новостей, наблюдая за тем, как идеи и культурные концепции распространяются от человека к человеку. Но социальные сети также используются для распространения дезинформации, которая неосознанно передается доверчивыми пользователями. Используя поддельные новостные веб-сайты и автоматизированные или полуавтоматические учетные записи, именуемые «социальными ботами», вредоносная организация может дешево и эффективно генерировать и усиливать кампанию по дезинформации в политических целях либо для монетизации трафика с помощью рекламы. В последние годы мы наблюдаем резкий рост подобных видов манипуляций социальными сетями в глобальном масштабе. Если кто-то может контролировать информацию, которую люди видят онлайн, то он может манипулировать их мнением. Во многих странах это явление представляет угрозу демократии, потому что без хорошо информированных избирателей невозможно проводить свободные выборы. Академические исследователи и промышленные инженеры усердно работают над разработкой контрмер. Понимание структуры и динамики сетей, обеспечивающих распространение информации, является важнейшим компонентом этих усилий.

Социальные связи в Twitter создаются до того, как пользователь создает сообщение, которое обычно транслируется всем подписчикам пользователя. В электронной почте, как и в социальных сетях, узлы являются людьми. Однако каждое сообщение предназначено для одного или нескольких конкретных получателей. Связи основаны на обмениваемых сообщениях. Электронная почта не зависит от конкретной платформы; протокол открыт и распространяется, вследствие чего ни одна организация не контролирует весь трафик. Как следствие, электронная почта по-прежнему остается одной из наиболее широко используемых коммуникационных сетей. На рис. 0.4 показан пример сети электронной почты. Опять же, связи направляются от отправителя к получателю электронного письма обозначенными стрелками. Размер и цвет узла представляют два разных признака: число соответственно входящих и исходящих связей. Более крупный узел получает электронные письма от большего числа отправителей, а более темный узел отправляет электронные письма большему числу получателей. Тот факт, что более крупные узлы, как правило, темнее и наоборот, говорит нам о том, что между отправкой и получением электронных писем существует корреляция.

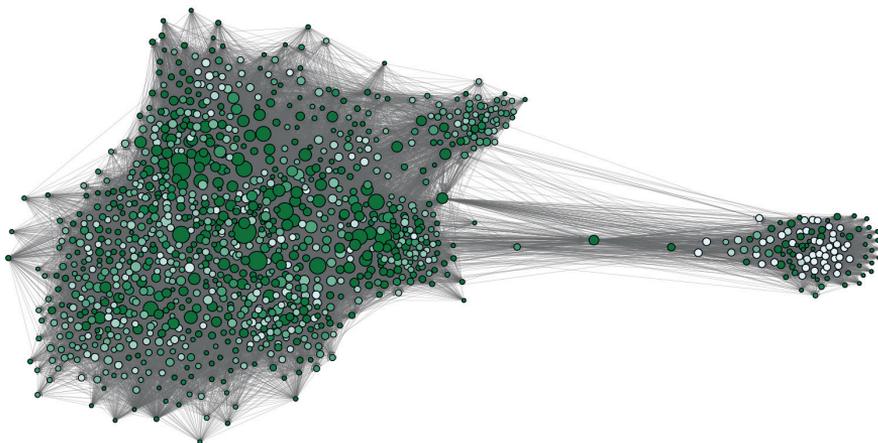


Рис. 0.4 Сеть, опирающаяся на базу данных электронных писем, сгенерированных сотрудниками энергетической компании Enron. Эти данные были получены Федеральной комиссией по регулированию энергетики США в ходе расследования, проведенного после краха компании в 2001 году. По завершении расследования электронные письма были признаны как находящиеся в публичном пространстве и сделаны общедоступными для исторических исследований и академических целей. Показана только небольшая часть центрального ядра сети. Направление связей показано стрелками

0.3. Всемирная паутина и «Википедия»

Всемирная паутина (Веб) – это крупнейшая информационная сеть. Хотя сейчас она используется для предоставления всех видов услуг, изначально это была просто сеть документов (страниц), соединенных «гиперсвязями», или кликабельными связями. В начале 1990-х годов Тим Бернерс-Ли захотел упростить доступ ученых к информации об экспериментах по физике высоких энергий в Европейской организации ядерных исследований (CERN) недалеко от Женевы. Он выдвинул три ключевые идеи: (1) систему именования страниц, Единый локатор ресурсов (Uniform Resource Locator, URL); (2) простой язык для написания документов, именуемый языком разметки гипертекста (HyperText Markup Language, HTML), включая гиперсвязи из одной страницы на другую; и (3) простой протокол, именуемый протоколом передачи гипертекста (HyperText Transfer Protocol, HTTP), для программ-клиентов (браузеров), чтобы общаться с серверами. Благодаря этим трем компонентам родилась Всемирная паутина. Бернерс-Ли даже имплементировал первый веб-сервер и программно-информационное обеспечение для браузера, чтобы скачивать страницы и мультимедиа с серверов, нажимая на связи. На самом деле мы можем видеть здесь участие двух сетей: статический «граф связей», состоящий из моментального снимка веб-страниц и связей в данный момент времени,

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru