

Asocial Networks

[Главная страница](#) [О нас](#)

вторник, 19 апреля 2016 г.

7 известных баз данных про социальные сети

Сетевые базы данных можно достаточно просто выстроить на основе известных фильмов, книг, исторических событий и даже на основе своих собственных дружеских связей. Это просто и увлекательно, а также намного интереснее изучать, как связаны люди, если мы этих людей знаем. Например, когда появился последний эпизод [Звездных войн](#) этой зимой, научные блоггеры тут же проанализировали сети в этом фильме. Например, [в этом посте](#) и [в этом посте](#) Эвелина Габасова показывает, как связаны друг с другом герои на протяжении всего фильма, выкладывая свой код в открытый доступ.

Помимо того, что работать с известными данными намного любопытнее, часто приходится учиться новому, а это также приятно делать на уже много раз изученных данных. В этом посте мы сделали подборку из наиболее часто используемых баз данных, с которых просто начинать изучение сетевого анализа и на них просто демонстрировать его базовые идеи.

Сеть флорентийских семей

(Florentine Families Network)

Пожалуй, это самая известная база данных, без которой не обходится ни один базовый учебник по сетевому анализу. Данные представляют собой 16 известных флорентийских семей XV века и брачные и деловые связи между ними, собранные из исторических источников. Брачная связь определялась как присутствующая, если члены одной семьи состояли были женаты или замужем за членами другой семьи. Деловая связь определялась как присутствующая, если члены одной семьи давали кредиты, займы или были в деловых партнерствах с членами другой семьи. Дополнительно есть данные о власти семьи: богатство семьи, количество мест в городском совете, общее количество деловых и брачных связей.

Популярные сообщения

[Что делать с графами в Python: пример в NetworkX](#)

С графами сегодня можно работать в большом количестве программ. Как правило, начинают с UCINET, Gephi, ORA. В них просто рассчи...

[С чего начать знакомство с социальными сетями: миллион вопросов в одном месте](#)

У нас часто спрашивают, с чего начать знакомство с социальными сетями. Какие книги и статьи прочитать, в каких программах работать с сет...

["Вне контакта". Вспомним, что мы могли изучать, когда данные "ВК" были открытыми.](#)

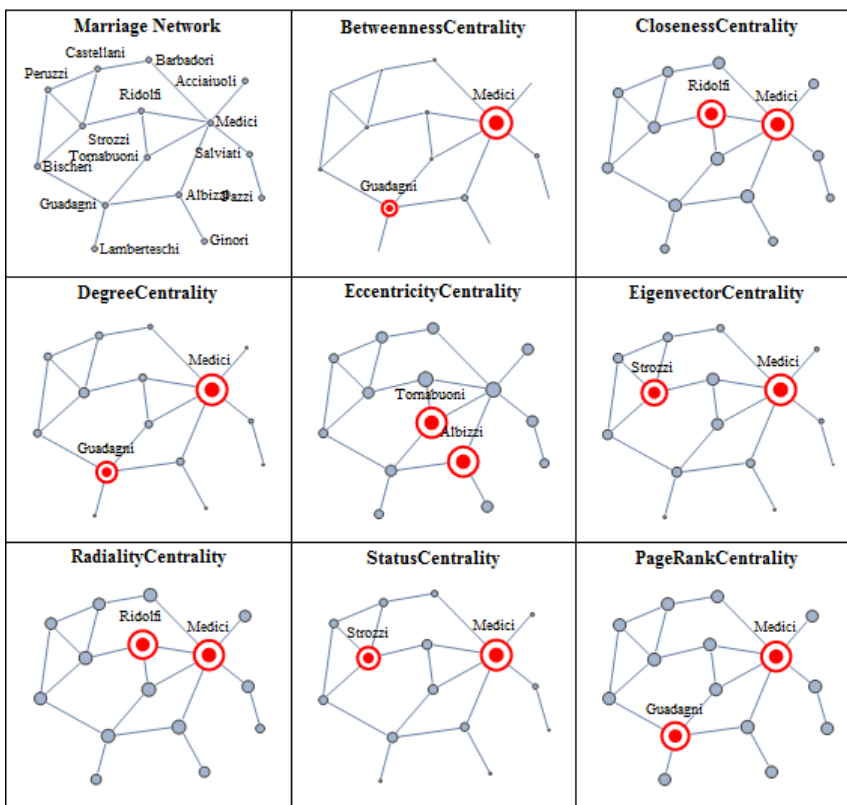
31 июля газета «Известия» написала о том, что Роскомнадзор (РКН) запретил сторонним компаниям собирать открытую персональную информац...

[Работа с графами в R](#)

Мы уже писали о том, как готовить описательную статистику для социальных сетей в Python, а в этом посте расскажем о том, как посчита...

[Загадка семи мостов, анархические социогаммы и кумовство норвежских рыбаков: к истории сетевых исследований](#)

Представляем гостевой пост от Егора Лавренчука - нашего коллеги и друга. Егор преподаватель, кандидат философских наук. Главный фил...



Сеть брачных связей флорентийских семей XV века. На рисунке проиллюстрированы различные меры сетевых центральностей. [Источник изображения.](#)

Эта база данных, как правило, используется при изучении центральностей. Из истории мы знаем, что самыми влиятельными были семьи Медичи и Строщи, но в ходе анализа мы узнаем, что ключевые акторы в сети — не всегда те, о которых мы думаем изначально.

Базу данных можно скачать [отсюда](#). Почитать подробнее: Padgett, J. F., & Ansell, C. K. (1993). *Robust Action and the Rise of the Medici, 1400-1434*. *American Journal of Sociology*, 1259-1319.

Клуб каратэ Захария (Zachary's karate club)

Это данные о 34 членах университетского клуба каратэ и дружеских связях между ними, собранные Уэйном Захарием в 1970-х. Дополнительно, есть данные о количестве ситуаций, когда между членами клуба происходили взаимодействия в клубе и вне его.

Данные часто используются для демонстрации идеи сетевых сообществ, так как первоначально перед У. Захарием стояла цель разрешить конфликтную ситуацию. В клубе было 2 человека (преподаватель и администратор), между которыми произошел конфликт, который разделил клуб на 2 сообщества. Половина участников сформировала новый клуб вокруг преподавателя, другая половина ушла в другой клуб или вовсе бросила заниматься этим видом спорта.

Tweets by @AsocialNets

Asocial Networks
@AsocialNets

Wow! Cool complexity science guide!
<https://twitter.com/DirkBrockman/status/1144730846605324294>

Jun 29, 2019

Asocial Networks Retweeted

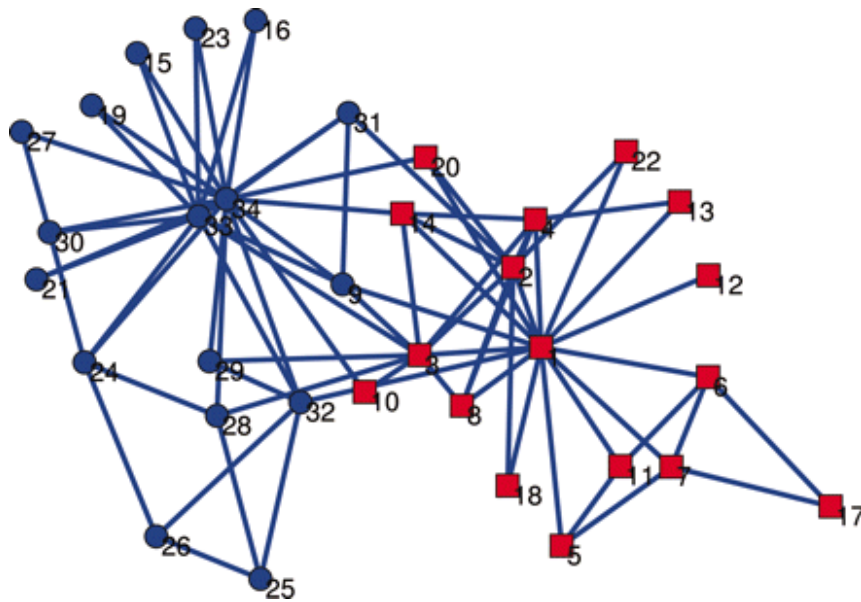
NICO
@NICOatNU

New research in @NatureComms finds that balanced professional #networks are more important than individual talent when it comes to high-risk decision making. @UzziLeadership @KelloggSchool news.northwestern.edu/stories/2019/0...

Embed View on Twitter

Ярлыки

агентные модели алкоголь
антропология базы данных
бимодальные сети биология блоги
большие данные брокер
визуализация ВИЧ Вконтакте
вычислительные социальные
науки генеалогическое древо
генетика гомофилия гостевой
пост динамические сети
друзья животные жизнь
журналы загадки сетей
звезда здоровье игры изолят
изоморфизм графов иллюзия
большинства искусство история
итоги месяца кино книги курсы
лекции лингвистика малый мир
машинное обучение медицина
мультисети нейронные сети
неравенство нетворкинг
образование одиночество
онлайн организации
открытые данные парадокс
дружбы перколяция популярность
породничество протесты
процессы в сетях
психология размер сети
регрессия родственные связи секс
селфи семантический анализ сети
скрипты советское социальная
селекция социальное
влияние степенной закон
текстовый анализ ученые
центральность школьники
эксперименты эпидемии языки
AddHealth complexity Coursera
CSS ERGM Facebook Gephi
graphtool igraph Instagram Nature
networkx ORA Pajek PNAS Python
R REM SAOM Science TED
Twitter



Сеть клуба каратэ Захария. Администратор клуба под номером 1, инструктор - под номером 33. Сеть разбита на 2 сетевых сообщества согласно кластеризации, предложенной в работе [Weinan & Vanden-Eijnden \(2008\)](#).

Сейчас есть [традиция вручать статуэтку клуба Захария](#) ученому, который на любой сетевой конференции представляет работу на основе этих данных.

Базу данных можно скачать [отсюда](#). Почитать подробнее: Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 452-473.

Сеть поддержки в «Гарри Поттере»

(*Harry Potter support network*)

Эти данные основаны на известной [серии детских книг о Гарри Поттере](#) — волшебнике, который учится в магической школе. Исследователи, на основе текстов всех 7 книг, сделали сети поддержки 64 учеников в школе (например, Гермиона помогает Гарри делать домашнее задание) и собрали информацию об атрибутах (пол, год обучения и факультет).

Особенность базы данных в том, что она лонгитюдная (всего 7 волн) и ее можно использовать при изучении лонгитюдных сетевых методов. Как это делать в R, мы уже писали [в одном из постов](#) как раз на примере данных из Гарри Поттера.

Базу данных можно скачать [отсюда](#). Почитать подробнее: Bossaert, G., & Meidert, N. (2013). "We are only as strong as we are united, as weak as we are divided" A dynamic analysis of the peer support networks in the Harry Potter books. *Open Journal of Applied Sciences*. 3(2), 174-185.

Стэнфордская коллекция больших сетевых данных

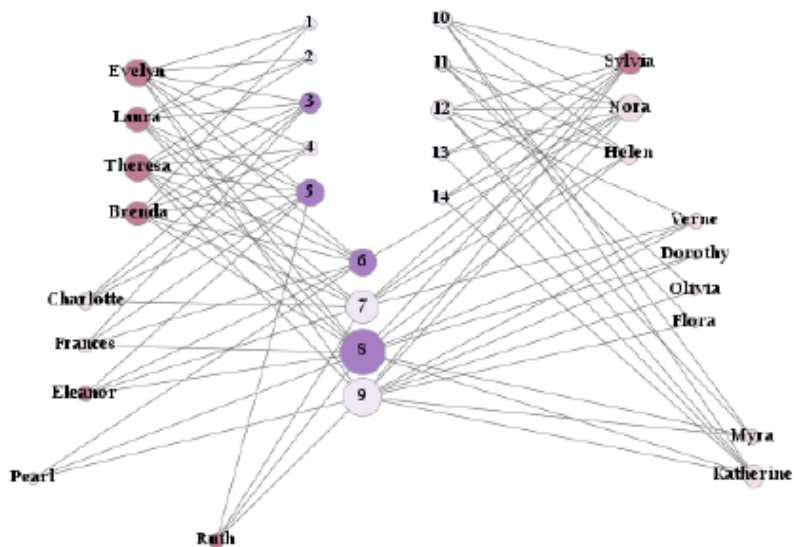
(*Stanford Large Network Dataset Collection*)

Здесь представлены различные данные по онлайн-социальным сетям, например, по Фейсбуку, Твиттеру. Особенность этих баз данных в том, что они представляют собой большие графы, что полезно при изучении онлайн-сетей, которые, конечно, состоят из сотен и тысяч узлов, а не из пары десятков. Часть из них выложена Ю. Лесковцом, о котором мы писали [в этом посте](#) и используется на Coursera-курсе "[Mining Massive Datasets](#)". Данные можно скачать [отсюда](#).

Южные женщины

(*Southern Women*)

Эта база данных состоит из 18 женщин с юга Америки, которые в 1930 участвовали в 14 социальных событиях. Она часто используется как пример бимодальной сети. Напомним, что бимодальная сеть — это сеть с двумя различными уровнями, которые состоят из акторов разного рода. Например, люди и организации, женщины и клубы. В обычных же сетях (одномодальных) у нас акторы принадлежат одному роду. Например, в клубе каратэ Захария у нас есть только люди, а в сети флорентийских семей — только семьи.



Сеть женщин, посещающих события из базы данных Southern Women. [Источник](#) изображения.

Данные можно скачать [отсюда](#) или [отсюда](#). Почитать подробнее: Davis, A., Gardner, B. B., & Gardner, M. R. (2009). *Deep South: A social anthropological study of caste and class*. Univ of South Carolina Press.

Сети соавторства физиков

(*Condensed matter collaborations*)

Это серия баз данных, собранная Марком Ньюманом с сайта препринтов [Condensed Matter E-Print Archive](#). База данных была впервые представлена за период с 1995 по 1999, затем были ее обновления включая 2003 и 2005 гг. Эта база данных используется для изучения сетей соавторства, также на основе ее были протестированы различные алгоритмы нахождения сообществ для больших сетей.

Данные можно скачать [отсюда](#) Почитать подробнее: 1) Newman, M. (2001). The structure of scientific collaboration networks. *Proc Natl Acad Sci.* 98, 404-409; 2) Duch, J., & Arenas, A. (2005). Community detection in complex networks using extremal optimization. *Physical review E*, 72(2), 027104.

Сеть серферов

(*Windsurfers on a beach*)

Это данные о группе из 43 серферов в Калифорнии, которые, по наблюдениям исследователей, были разделены на несколько сетевых сообществ. Линтон Фриман с коллегами сначала наблюдали за взаимоотношениями серферов и кодировали их. Затем они взяли интервью, где у каждого серфера спрашивали, как он или она считают, как связаны люди в этом клубе. База данных часто используется для работы с когнитивной социальной структурой (cognitive social structure) и для работы с пропущенными данными в сетевых исследованиях.



Серферы на пляже. [Источник изображения.](#)

Данные можно скачать [отсюда](#). Почитать подробнее: 1) Freeman, L. C., Freeman, S. C., & Michaelson, A. G. (1988). On human social intelligence. *Journal of Social and Biological Structures*, 11(4), 415-425; 2) Freeman, L. C., Freeman, S. C., & Michaelson, A. G. (1989). How humans see social groups: A test of the Sailer-Gaulin models. *Journal of Quantitative Anthropology*, 1(3), 229-238.

И это еще не все

Также большое количество открытых сетевых данных можно найти на следующих сайтах:

- [Базы данных](#) на странице Марка Ньюмана. Здесь можно найти большинство стандартных баз данных, а также данных, специфических для программного обеспечения (для R, UCINET);
- [Базы данных](#) на странице Линтона Фримана. Один из самых полных списков стандартных данных, на которых хорошо обучаться;
- [Базы данных](#) на странице Тома Снайдерса. Большое количество лонгитюдных данных о школьниках и их поведении;
- [Базы данных](#) на странице Альберта-Ласло Барабаши. Небольшой список больших сетевых данных.

Автор: [Diliara Valeeva](#) на 15:20



Ярлыки: [базы данных](#), [большие данные](#), [друзья](#), [процессы в сетях](#), [центральность](#), [Facebook](#), [Twitter](#)

Комментариев нет:

Отправка комментария

Подпись комментария: Аккаунт Google ▼

Опубликована

Предварительная

[Следующее](#)

[Главная страница](#)

[Предыдущее](#)

Подписаться на: [Комментарии к сообщению \(Atom\)](#)

