

О ПОДТВЕРЖДЕНИИ ТЕОРИИ ШЕСТИ РУКОПОЖАТИЙ

Ефимов А.А.; Гришаков А.О.

Кафедра информационных технологий автоматизированных систем
Научный руководитель: Севернёв А.М., доцент каф. ИТАС, к.т.н., доцент
e-mail: soks.cokc@gmail.com

Аннотация– Теория шести рукопожатий – теория, согласно которой любые два человека на Земле разделены в среднем лишь пятью уровнями общих знакомых и, соответственно, шестью уровнями связей.

Ключевые слова: социальная сеть, выборка, алгоритм Дейкстры

Историческая справка

Теория была выдвинута в 1969-м году американскими психологами Стэнли Милгрэмом и Джеффри Трэверсом. Предложенная ими гипотеза заключалась в том, что каждый человек опосредованно знаком с любым другим жителем планеты через цепочку общих знакомых, в среднем состоящую из пяти человек.

Милгрэм опирался на данные эксперимента, проведенного в двух американских городах – Нью-Йорке и Бостоне. Жителям одного города было роздано 300 конвертов, которые надо было передать определенному человеку, живущему в другом городе. Конверты можно было передавать только через своих знакомых и родственников. До адресата дошло 60 конвертов. Произведя подсчёты, Милгрэм определил, что в среднем каждый конверт прошёл через цепочку из пяти человек. Так родилась теория «шести рукопожатий».

Проверка теории

Для проверки теории шести рукопожатий необходима репрезентативная выборка данных о пользователях сети Интернет. Для её получения мы решили воспользоваться популярной социальной сетью *ВКонтакте*. Но предварительно надо было решить следующие задачи:

- на каких данных выборки рассчитывать взаимосвязи людей;
- как эти данные сохранять;
- каким алгоритмом воспользоваться для обработки полученных данных.

Для решения первой задачи использовали стандартный интерфейс API сайта *ВКонтакте*, возвращающий данные о пользователе и его друзьях в json-формате; это позволяет минимизировать вычислительные мощности для вычисления числа связей.

Для сохранения данных о пользователях предложены на выбор две стратегии:

– запись сразу в БД – сайт *ВКонтакте* на запрос возвращает 100 пользователей в секунду, у каждого в среднем 130 друзей, итого 13000 вставок в БД в секунду;

– запись данных в текстовый файл на диск, а потом запись его в базу данных. В этом случае БД будет «весить» примерно 166 Гб [4 байта (размер поля user_id) + 4 байта (размер поля friend_id) + 8 байтов на индексы]*80 млн. пользователей сайта *ВКонтакте* * 130 друзей].

Для реализации была выбрана вторая стратегия.

Далее необходимо решить, с помощью какого алгоритма осуществлять поиск дистанции в

«количестве рукопожатий» между людьми. Рассматривались следующие алгоритмы:

– алгоритм Флойда-Уоршелла – позволяет рассчитать дистанции от всех пользователей ко всем. Недостатком этого алгоритма является требование памяти – необходимо хранить квадратную матрицу user_id/user_id, которая бы занимала 1 байт * 80 млн. пользователей * 80 млн пользователей = 6400 Тб;

– алгоритм Дейкстры – позволяет найти дистанции от одного пользователя до всех остальных;

– двунаправленный поиск – позволяет найти кратчайшую дистанцию между любыми двумя пользователями.

Выбран алгоритм Дейкстры, как наиболее подходящий для нашей цели.

После получения выборки пользователей *ВКонтакте* нами был произведён процесс отсеивания так называемых «ботов» путём исключения пользователей с нулевыми связями, а также пользователей, включённых в изолированные графы. Таким образом, было проанализировано 67730 пар пользователей, и результат статистической обработки представлен на рисунке 1.

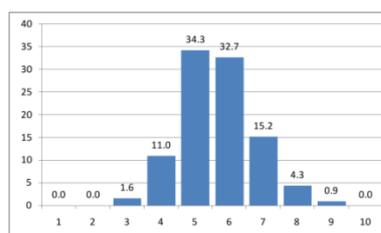


Рис. 1. Статистический ряд распределения числа связей

Выводы

Таким образом, в среднем между двумя случайными пользователями сайта *ВКонтакте* есть цепочка 5.65 друзей (т.е. 6.65 рукопожатий). Эта цифра вполне согласуется с проверяемой теорией, к тому же довольно точно совпадает с результатом, полученным фирмой Microsoft (6.6 рукопожатий). Так что полученный нами результат можно считать одним из подтверждений теории шести рукопожатий.

[1] Душенко, К. Мир тесен, или Шесть рукопожатий: Читаем Вместе, 2011. – 4 с.

[2] E.W. Dijkstra. A note on two problems in connexion with graphs. // Numerische Mathematic. V.1 (1959), P. 269-271.