

ИНТЕР- И ИНТРАСЕТЕВЫЕ СВЯЗИ: ВИЗУАЛИЗАЦИЯ И АНАЛИЗ

Дудкин Е. В., студент кафедры информатики СПбГУ,
evgene.dudkin@gmail.com

Аннотация

В настоящей работе рассмотрена реализация программного обеспечения, способного автоматизировать сбор, обработку и визуализацию информации об онлайн-сообществах социальной сети «ВКонтакте».

Введение

В наши дни социальные сети стали неотъемлемой частью нашей жизни. Они являются источником данных и средством распространения информации. Поэтому в последнее время увеличилось количество исследований, проводимых в социальных сетях. При этом огромное внимание уделяется изучению онлайн-сообществ как особых социальных общностей [1].

Для проведения исследований онлайн-сообществ необходимы программы способные собирать необходимую информацию о группах. Но зачастую исследователям бывает недостаточно существующих инструментов для решения поставленных задач. Специалисты, составившие план изучения сообществ в социальных сетях, вынуждены тратить время и средства на сбор и анализ данных. Поэтому возникла необходимость в программном обеспечении, которое позволит собирать необходимые данные онлайн-сообществ, а также предоставит набор инструментов для их базового анализа.

Цель данной работы — автоматизировать сбор, обработку и визуализацию информации об онлайн-сообществах социальной сети «ВКонтакте» (в рамках поставленных ниже задач).

Постановка задачи

Для достижения цели были поставлены следующие задачи:

- обеспечить сбор данных о структурах онлайн-сообществ социальной сети «ВКонтакте», включая список контактов, город, ссылки, страну и информацию об участниках;

- визуализировать структуру сообщества в виде графа, вершинами которого являются участники этого сообщества, а ребрами — соответствующие связи между этими пользователями (дружба, обучение в одном вузе, город, родной город, обучение в одной школе, наличие общих сообществ);
- визуализировать структуру связей между несколькими сообществами в виде графа, вершинами которого являются эти сообщества, а ребрами — соответствующие связи между ними (город, ссылки на одни и те же ресурсы или на друг друга, страна, наличие общих пользователей в графе контакты);
- предоставить инструменты для базового анализа полученных графов (вычисление показателей центральности, среднего для показателей центральности, количества вершин и ребер, вычисление количества изолированных вершин, количества компонент связности, количества вершин максимальной компоненты связности и др.) ;
- предоставить возможность импортировать данные о структурах в виде CSV файлов.

Описание решения

Интерсетевые связи

Программное обеспечение, разработанное в рамках работы, позволяет анализировать интерсетевые связи (связи между участниками внутри сообщества). Для этого строится сетевая схема онлайн-сообщества. Сначала программа предоставляет пользователю различные типы связей. Связи нужны для того, чтобы формировать ребра между вершинами графа. Вершинами графа в социальной сетевой схеме являются участники сообщества. А ребро между вершинами будет существовать только тогда, когда между участниками в социальной сети есть соответствующая связь. Программа предоставляет следующие типы связей: город, родной город, страна, школа, вуз, дружба и наличие общих сообществ. Затем с помощью http-запроса получается список участников сообщества. Это и будут вершины нашего графа. Затем для каждой выбранной связи, тоже с помощью запросов, получают нужные данные об участниках. На основании этих данных по-

лучается список ребер, соответствующий этой связи. После этого ребра добавляются граф и помечаются своим цветом.

Интрасетевые связи

Программное обеспечение позволяет анализировать связи между сообществами (интрасетевые связи). После того, как составлен набор сообществ, который необходимо проанализировать, программа предоставляет пользователю различные типы связей. Связи здесь также нужны для того, чтобы формировать ребра между вершинами графа. Но в данном случае вершинами графа в социальной сетевой схеме являются сами сообщества. А ребро между вершинами будет существовать только тогда, когда между сообществами есть соответствующая связь. Программа предоставляет следующие типы связей: город, ссылки, страна, контакты и наличие общих участников. Вершинами графа будет набор сообществ. Для каждой выбранной связи, с помощью запросов к серверу «ВКонтакте», получаются необходимые данные о сообществах. На основании этих данных получается список ребер, соответствующий этой связи. После этого ребра добавляются граф и помечаются своим цветом.

Визуализация

После обработки данных и представления их в виде графа программа предоставляет пользователю набор инструментов для работы с ним. Одной из возможностей является его отображение на экране. Программное обеспечение дает возможность визуализировать граф как статическое изображение с параметрами (размер и цвет вершин, размер и цвет названия вершины, цвет ребер, алгоритм отрисовки, наличие названия вершин) по умолчанию. Также программа дает возможность пользователю настроить параметры самостоятельно. Кроме того, программное обеспечение позволяет визуализировать граф на интерактивной форме, где можно настраивать параметры и перемещать вершины графа. На форме можно вручную менять цвета и размеры вершин, цвета и размеры ребер, алгоритмы отрисовки, наличие названия вершин.

Для больших графов предусмотрен специальный режим, который позволяет наглядно представить их структуру. На рисунке 1 представлен граф, состоящий из 550 вершин. При большом количестве изолированных вершин, можно воспользоваться отображением сетевой схемы

без изолянтов.

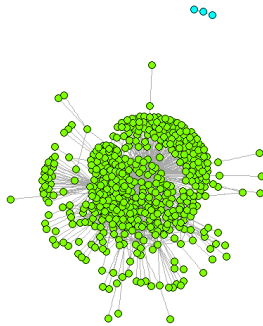


Рис. 1: Пример визуализации графа

Создание CSV файла

У пользователя должна быть возможность воспользоваться другими средствами визуализации или анализа. Такие программы принимают на вход файлы формата csv. Поэтому возникла необходимость представления графа в виде csv файла.

Граф онлайн-сообщества представляется как таблица из трех столбцов: Source, Target, Type. С помощью этой таблицы описываются все ребра графа.

Анализ графа

Программное обеспечение, рассматриваемое в данной работе, позволяет пользователю считать некоторые базовые метрики графа: количество вершин, число ребер, плотность, количество изолированных вершин (их доля в процентах относительно общего количества вершин), количество компонент связности, количество вершин в каждой компоненте связности, доля вершин максимальной компоненты связности (в %), средняя геодезическая дистанция, центральность по посредничеству каждого узла, средняя центральность по посредничеству (без

изолянтов), центральность по близости каждого узла, средняя центральность по близости (без изолянтов), центральность собственного вектора каждого узла, средняя центральность собственного вектора (без изолянтов), центральность по степени каждого узла, средняя центральность по степени (без изолянтов), взаимность, транзитивность.

Заключение

Таким образом были решены все поставленные задачи. Был обеспечен сбор данных о структурах онлайн-сообществ социальной сети «ВКонтакте», включая список контактов, город, ссылки, страна и информацию об участниках. Была обеспечена визуализация структуры сообщества в виде графа, вершинами которого являются участники этого сообщества, а ребрами — соответствующие связи между этими пользователями (дружба, обучение в одном вузе, город, родной город, обучение в одной школе, наличие общих сообществ). Была обеспечена визуализация структуры связей между несколькими сообществами в виде графа, вершинами которого являются эти сообщества, а ребрами — соответствующие связи между ними (город, ссылки на одни и те же ресурсы или на друг друга, страна, наличие общих пользователей в графе контакты). Были предоставлены инструменты для базового анализа полученных графов (вычисление показателей центральности, среднего для показателей центральности, количества вершин и ребер, вычисление количества изолированных вершин, количества компонент связности, количества вершин максимальной компоненты связности и др.). Была предоставлена возможность импортировать данные о структурах в виде CSV файлов. Таким образом, цель — автоматизировать сбор, обработку и визуализацию информации об онлайн-сообществах в социальной сети «ВКонтакте», можно считать достигнутой.

Литература

- [1] Martinez M. G., Walton B. The wisdom of crowds: The potential of online communities as a tool for data analysis // Technovation. — 2014. — Vol. 34, no. 4. — P. 203–214.