



OSTIS-2016

(Open Semantic Technologies for Intelligent Systems)

УДК 004.853

СЕМАНТИЧЕСКИЙ АНАЛИЗ СОЦИАЛЬНЫХ СЕТЕЙ С ИСПОЛЬЗОВАНИЕМ СРЕДЫ WOLFRAM MATHEMATICA

Тякунов А.С., Славский В.В.

*Югорский государственный университет,
г. Ханты-Мансийск, Российская Федерация*

atyakunov@mail.ru

slavsky2004@mail.ru

В работе проводится обзор возможностей среды Wolfram Mathematica, используемых для семантического анализа открытых данных в социальных сетях. Рассмотрены варианты использования запросов Application Programming Interface (API) и их интерпретации. Подготовлена основа для создания учебно-методических пособий по использованию среды Wolfram Mathematica.

Ключевые слова: семантический анализ; социальные сети; API, теория графов.

Работа выполнена при поддержке Совета по грантам Президента РФ (грант НШ 2263.2014.1), Правительства РФ (госконтракт 14.B25.31.0029), РФФИ 15-41-00092 р-урал-а, 15-41-00063 р-урал-а, 15-01-06582 А.

Введение

Онлайновые социальные сети, помимо выполнения функций поддержки общения, обмена мнениями и получения информации их членами, в последнее время все чаще становятся объектами и средствами информационного управления и ареной информационного противоборства. В недалеком будущем они неизбежно станут существенным инструментом информационного влияния, в том числе, в целях манипулирования личностью, социальными группами и обществом в целом.

Таким образом, потребность в полном и качественном анализе социальных взаимодействий посредством социальных сетей является важной задачей современности, в том числе и с математической точки зрения. Для ее решения требуется наличие инструментов, позволяющих обрабатывать запросы к социальным сетям с целью получения понятных результатов анализа. Одним из таких инструментов является среда Wolfram Mathematica.

Математическое программное обеспечение Mathematica, известно как самое мощное в мире вычислительное приложение. Более того, Mathematica служит единой платформой для проведения исследований и вычислений, которые находят свое отражение в живых интерактивных документах издательского качества. Используя

мощный вычислительный аппарат и удобный интерфейс, пользователи Mathematica могут легко решить трудоемкие задачи и разобраться в сложных концепциях, используя весь имеющийся арсенал аппаратных средств.

Цель данной работы – продемонстрировать возможности среды Wolfram Mathematica 10 для масштабного анализа связей пользователей социальных сетей, графического представления и анализа данных, размещенных в открытом доступе.

1. Общие определения

Термин "сеть" имеет разные значения в разных дисциплинах. В социальных науках, сеть, как правило, определяется как набор субъектов сети (или агентов, или узлов, или точек, или вершин), которые могут иметь отношения (или ссылки, или ребра, или связи) друг относительно друга [Hanneman et al, 2005]. Сети могут иметь несколько или много узлов, и один или более видов отношений между парами узлов (Рисунок 1).

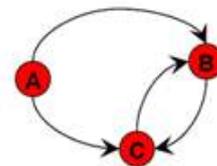


Рисунок 1 – Пример направленной сети

Социальная сеть – платформа, онлайн-сервис или веб-сайт, предназначенные для построения, отражения и организации социальных взаимоотношений в Интернете.

Получать данные из социальной сети удобнее всего посредством готовых API-функций. API (интерфейс программирования приложений) — набор готовых классов, процедур, функций, структур и констант, предоставляемых приложением (библиотекой, сервисом) для использования во внешних программных продуктах.

Стандартный формат ответа от API — это JSON (англ. JavaScript Object Notation) — текстовый формат обмена данными, основанный на JavaScript и обычно используемый именно с этим языком. Как и многие другие текстовые форматы, JSON легко читается людьми.

Это универсальные структуры данных: как правило, любой современный язык программирования поддерживает их в той или иной форме. Они легли в основу JSON, так как он используется для обмена данными между различными языками программирования. В частности, JSON легко конвертируется в списки и правила внутренними средствами Mathematica.

2. Запросы в социальных сетях

2.1. Использование API-функций социальной сети "ВКонтакте"

Социальная сеть "ВКонтакте" (<http://vk.com>) выбрана полем для исследования по следующим причинам:

- это одна из крупнейших мировых и крупнейшая русскоязычная социальная сеть;
- формат запроса достаточно прост и масштабируется, позволяя получать разнообразные данные о пользователе;
- для создания запросов на получение данных не обязательно (по крайней мере, на сегодняшний день) создавать standalone-приложение, передающее ключи доступа;
- учетная запись данной сети имеется у большинства пользователей в российском сегменте Интернет.

В общем случае, для того чтобы вызвать метод API сети "ВКонтакте", необходимо осуществить POST- или GET-запрос по протоколу HTTPS на URL <https://api.vk.com>. Для получения данных внешнего сайта или файла в Mathematica используется команда Import. Передавая в нее URL запроса, мы получим данные, которые размещаются в сети Интернет по соответствующему адресу. Используя необходимый формат (в данном случае - описанный выше JSON), можно легко получать необходимые для исследования данные.

2.2. Запрос данных о пользователе социальной сети

Для получения данных из сети "ВКонтакте" необходимо проанализировать структуру запроса. К примеру, мы желаем получить самую общую информацию о пользователе. Для этого будет использоваться метод `users.get`, имеющий параметр `user_id`, в котором размещается ID пользователя.

Перечисляя необходимые поля, которые необходимо вернуть, можно, к примеру, получить имя, фамилию, фотографию пользователя, его место жительства и пол. Mathematica позволяет совмещать типы данных внутри одного списка, что чрезвычайно удобно при обработке данных, полученных из социальной сети (Рисунок 2).

```
{Александр, Тякунов, Ханты-Мансийск, , Мужской}
```

Рисунок 2 – Вывод общей информации о пользователе

Получение информации, размещенной на стене и в ленте пользователя производится аналогичными средствами и позволяет исследователю, в том числе, анализировать и интерпретировать полученные результаты согласно выдвигаемым гипотезам. Так, например, аналитическим агентством "Скориста" (г.Москва) в 2014 г. был представлен алгоритм предсказания поведения кредитного заемщика на основе открытых данных. Проверяя гипотезы о связи открытости профиля предполагаемого заемщика, а также частоты и направленности публикаций с предполагаемой добросовестностью, был сделан вывод, что чем больше люди говорят в социальной сети, тем они хуже возвращают займы.

2.3. Анализ социальных связей пользователя

Особый интерес для исследователя представляет, как правило, изучение и анализ информации о тех пользователях, с которыми связан данный узел. Для получения информации о "друзьях" пользователя в социальной сети "ВКонтакте" используется метод `friends.get`, который формируется аналогично `users.get`. К примеру, достаточно простым, но весьма наглядным могут быть графические представления ответов на запросы о половом разделении списка "друзей" (Рисунок 3) или градации по месяцам рождения (Рисунок 4). В контексте изучения Wolfram Mathematica подобные примеры позволяют осваивать функционал, не теряя интереса обучающегося, поскольку основаны на данных его собственного профиля социальной сети.

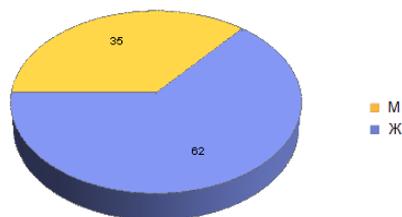


Рисунок 3 – Распределение "друзей" пользователя по полу

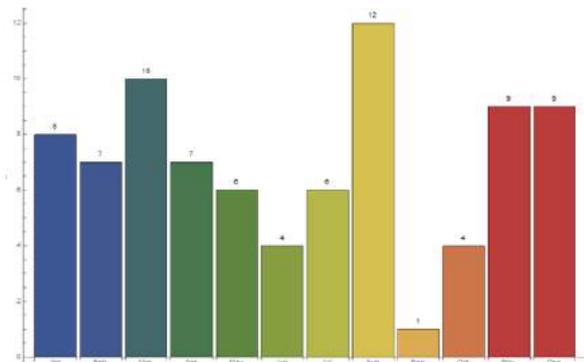


Рисунок 4 – Распределение дат рождения "друзей" пользователя по месяцам

С другой стороны, анализ данных о "друзьях" пользователя может быть использован и при проверке определенных гипотез [Watts, 1999]. В частности, практически подтвержденный факт того, что значительную (от 40% всех "друзей") часть связей в социальной сети составляют пользователи, проживающие в том же населенном пункте, что и сам пользователь, позволяет предположить, что подобное распределение характерно для всех подлинных пользователей сети. Это аналогично и для возраста, социальных убеждений, интересов и т.д. В то же время, как известно, многие социальные сети позволяют скрыть дату рождения, город проживания и другие поля. Таким образом, проверяя распределение "друзей" пользователя по параметру "город" с определенной долей уверенности мы можем сказать, что сам пользователь проживает в том городе, что и большинство его друзей (Рисунок 5). Подобное предположение позволяет, например, выявить недостоверные ("фейковые") учетные записи, созданные для рекламных или иных целей. Пользователи, у которых нет четко выделенной группы "друзей" (Рисунок 6), проживающих в одном населенном пункте (или регионе), безусловно, могут существовать, однако, это не является характерным для большинства. Аналогичная ситуация обстоит и с указываемым годом рождения. Такая проверка достоверности может быть использована, в том числе, при проверке добросовестности в кредитовании, наборе персонала, при пересылке информационных сообщений в социальных сетях и тому подобных операциях.

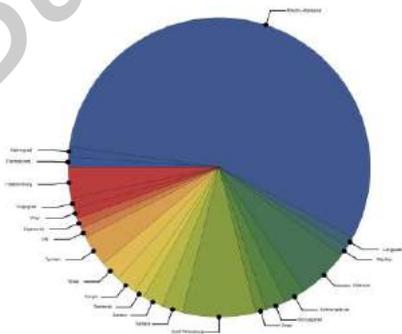


Рисунок 5 – Распределение "друзей" по городам

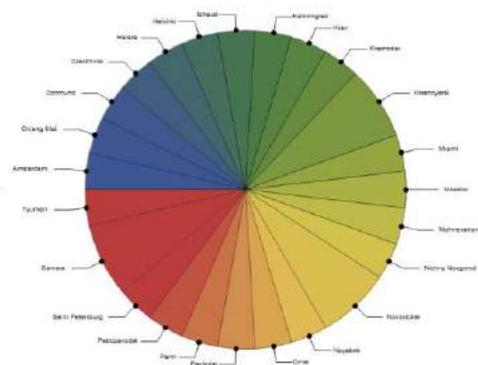


Рисунок 6 – Распределение "друзей" по городам в случае недостоверной учетной записи

3. Использование элементов теории графов в анализе социальных сетей

Одним из наиболее удобных и информативных методов представления информации о связях в социальной сети является изображение их в виде графа. В Wolfram Mathematica реализован значительный функционал по работе с графами, в том числе, масса встроенных функций, позволяющих анализировать структуру графов.

Так, например, чрезвычайно полезным как в образовательном, так и в исследовательском плане является получение графа связей "друзей" пользователя между собой. При этом средства Mathematica позволяют не только указать связи, но и продемонстрировать дополнительные свойства пользователей. Так на полученном графе (Рисунок 7) разным цветом обозначен пол пользователя, а диаметр круга зависит от количества друзей каждого из пользователей. Активность в сети можно использовать для понимания, какие пользователи из нашего окружения могут обладать наибольшим влиянием [Губанов и др., 2010].

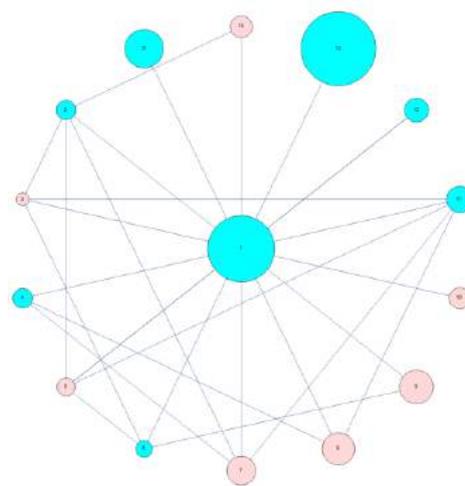


Рисунок 7 – Граф "друзей" пользователя

Используя встроенные функции для обработки графов можно представить полученный выше рисунок в виде "графа сообществ" (community graph), т.е. разделения всех "друзей" пользователя на

группы, имеющие связи между собой и, как следствие, максимальное влияние друг на друга (Рисунок 8).

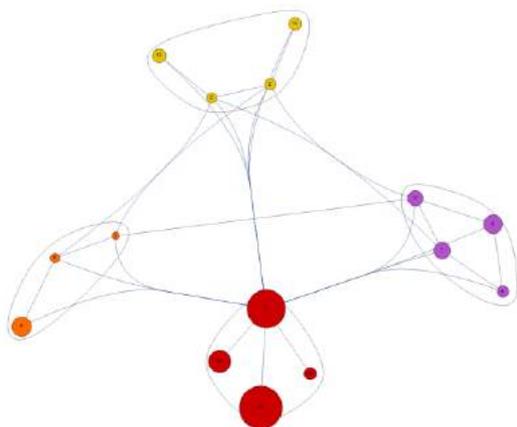


Рисунок 8 – Граф "друзей" пользователя

Заключение

Социальные сети на сегодняшний день прочно вошли в повседневную жизнь людей во всем мире. Изучение сетевых взаимодействий пользователей, а также той информации, которая передается между пользователями представляет собой важную задачу современных информационных технологий.

Wolfram Mathematica как наиболее перспективная и мощная среда для семантического анализа социальных данных позволяет производить разнообразные эксперименты по проверке социологических гипотез. Как следствие, можно утверждать, что изучение и развитие подобной среды является необходимостью для современного исследователя.

Библиографический список

[Губанов и др., 2010] Губанов Д.А., Новиков Д.А., Чхартишвили А.Г. Модели информационного влияния и информационного управления в социальных сетях / Губанов Д.А. [и др.]; - ФИЗМАТЛИТ, 2010

[Hanneman et al, 2005] Hanneman, R. A., M. Riddle Introduction to Social Network Methods/ R. A. Hanneman [et al]// University of California, Riverside. - 2005

[Watts, 1999] Watts, D.J. Small Worlds: The Dynamics of Networks between Order and Randomness. Princeton University Press, Princeton, NJ. - 1999

SEMANTIC ANALYSIS OF SOCIAL NETWORKS USING WOLFRAM MATHEMATICA

Tyakunov A.S., Slavsky V.V.

Ugra State University, Khanty-Mansiysk, Russia

atyakunov@mail.ru

slavsky2004@mail.ru

This article examines the use of Wolfram Mathematica for data analysis posted on social networks. First of all, interest is public data, on the basis of which the researcher can check various social hypothesis. The purpose of this work - to show the possibilities of Wolfram Mathematica for large-scale analysis of the

relations of social network users, graphical reporting and analysis of data placed in the public domain.

Introduction

Online social networks, in addition to support functions of communication, exchange of views and information of their members in recent years have increasingly become targets and means of information management and information warfare arena. In the near future, they will inevitably become an essential tool of information influence, including, for the purpose of manipulating the individual, social groups and society as a whole.

Thus, the need for full and qualitative analysis of social interactions through social networks is a major challenge of our time, including from a mathematical point of view. To solve it requires tools to handle requests to social networks in order to obtain clear results of the analysis. One such tool is an Wolfram Mathematica.

Main Part

The mathematical software Mathematica, is known as the most powerful in the world of computing application. Moreover, Mathematica provides a single platform for research and calculations, which are reflected in the live interactive publishing quality documents. Using a powerful computing unit and user-friendly interface, users can easily Mathematica to solve the time-consuming tasks and to understand the complex concepts using all available arsenal of hardware. Of particular interest to the researchers is, usually, the study and analysis of information about the users, which is associated with the node. Analysis of the data on the "Friends" of the user can be used for verification of various hypotheses. One of the most convenient and informative methods of presenting information about connections in the social network is an image of a graph. The Wolfram Mathematica implemented significant functionality for working with graphs, including plenty of built-in features to help you analyze the structure of graphs.

Conclusion

Social networks today have become part of the daily lives of people around the world. The study of network user interactions, as well as the information that is transferred between users is an important task of modern information technologies. Mathematica as the most promising and powerful environment for semantic analysis of social data allows for a variety of experiments to test sociological hypotheses. As a consequence, it can be argued that the study and the development of such an environment is a necessity for the researcher.