

АНАЛИЗ ПОЛЬЗОВАТЕЛЬСКИХ ПРОФИЛЕЙ С ЦЕЛЬЮ ОПРЕДЕЛЕНИЯ НАИБОЛЕЕ ЗНАЧИМЫХ ОБЪЕКТОВ ПОДМНОЖЕСТВА СОЦИАЛЬНОЙ СЕТИ «ВКОНТАКТЕ»

Алымов А.С.¹, Бабийчук Г.А.² Email: Alymov1140@scientifictext.ru

¹Алымов Алексей Сергеевич – бакалавр, студент;

²Бабийчук Георгий Александрович – бакалавр, студент,
кафедра инструментального и прикладного программного обеспечения,
Институт информационных технологий
Московский технологический университет,
г. Москва

Аннотация: в данной статье рассматриваются основные вопросы проектирования и разработки макета системы определения наиболее значимых объектов подмножества в социальной сети «ВКонтакте». Основное внимание уделяется представлению описания социального влияния пользователей друг на друга, а также рассматривается процесс распространения информации в социальных сетях. Для определения влияния конкретного пользователя на других используется линейная пороговая модель, а для непосредственного определения наиболее влиятельного объекта социальной сети рассматривается реализация «жадного» алгоритма.

Ключевые слова: социальная сеть, распространение информации, наиболее значимый объект подмножества, интеллектуальный анализ, ВКонтакте.

ANALYSIS OF USER PROFILES TO DETERMINE THE MOST SIGNIFICANT OBJECTS OF A SUBSET IN SOCIAL NETWORK «VKONTAKTE»

Alymov A.S.¹, Babiychuk G.A.²

¹Alymov Alexey Sergeevich – Bachelor, Student,

²Babiychuk Georgiy Alexandrovich – Bachelor, Student,
DEPARTMENT OF INSTRUMENTAL AND APPLIED SOFTWARE,
INSTITUTE OF INFORMATION TECHNOLOGIES
MOSCOW TECHNOLOGICAL UNIVERSITY,
MOSCOW

Abstract: this article discusses the main issues of design and development of the layout of the system for determining the most significant objects of a subset in social network Vkontakte. The focus is on presenting a description of the social impact of users on each other, as well as the process of disseminating information in social networks. To determine the influence of a particular user on others, a linear threshold model is used, and for the direct determination of the most influential object of a social network we used the implementation of the "greedy" algorithm.

Keywords: social network, information dissemination, the most significant object of a subset, intellectual analysis, VKontakte.

УДК 004.021

Со времен введения Веб 2.0 социальные взаимоотношения людей существенно изменились, появились информационные порталы, способные отслеживать и отображать все активности человека в обществе, а также накапливать информацию об этих активностях. Одним из таких информационных порталов является социальная сеть «ВКонтакте» [1].

Ядром социальной сети «ВКонтакте» является набор пользовательских профилей, где зарегистрированные участники могут размещать информацию, которой они хотят поделиться с другими. По большей части пользователи участвуют в двух видах деятельности на сайте: либо создают новый контент (например, добавление фотографий, загрузка музыки, написание блогов и сообщений), или они потребляют контент, который создают другие пользователи (например, просмотр фотографий, прослушивание музыки, чтение блогов и сообщений). Также пользователи могут добавлять других пользователей в «друзья». Обычно один пользователь инициирует приглашение, а другой пользователь принимает или отклоняет его.

На страницах социальных сетей контент почти полностью создан пользователем. Пользователи сильно отличаются друг от друга с точки зрения частоты, объема, типа и качества цифрового сгенерированного и потребляемого контента. С управленческой точки зрения, очень важно понимание того, кто является наиболее влиятельным объектом социальной сети, в частности, определение пользователей, которые влияют на деятельность других.

На рисунке 2 Группа А, друзья (люди, которым Алексей отправил приглашения) - Александр, Георгий, Дарья, Анастасия, Денис, Иван.

The graph illustrates a social network with 25 individuals. The central nodes, which are dark grey, represent highly connected individuals: Александр, Алексей (top center), Иван (middle left), Максим (middle left), Денис (bottom center), and Анастасия (bottom right). These central nodes are connected to numerous peripheral nodes (light grey circles) representing less connected individuals. The network is highly interconnected, with many overlapping connections between the central and peripheral nodes, suggesting a complex social structure.

Рис. 1. Сеть друзей Алексея

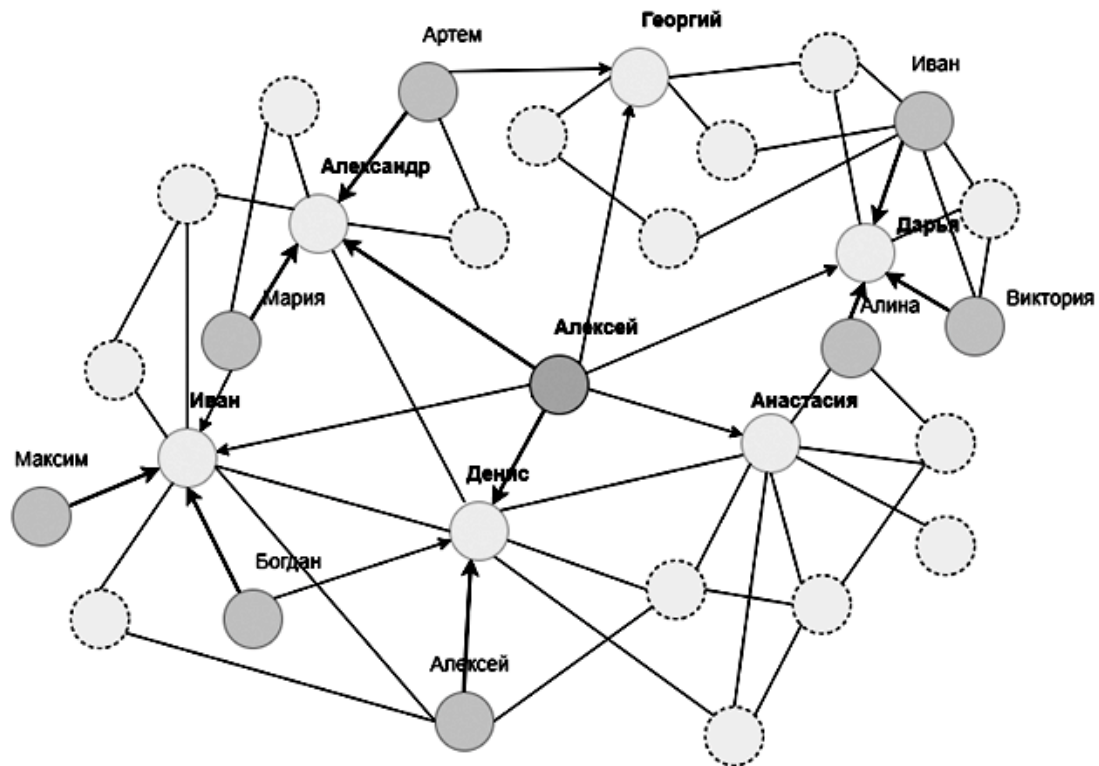


Рис. 2. Сеть «важных» друзей Алексея

В свою очередь, с точки зрения некоторых из его друзей, Алексей также может быть важен. Возможно, что некоторые из друзей Алексея регулярно проверяют содержание, которое он производит, или мотивированы на то, чтобы внести новый контент на страницы социальных сетей в надежде, что Алексей его увидит. В этом смысле онлайн-активность Алексея влияет на их поведение. Целью данного исследования является разработка метода оценки степени и направления влияния, связанного с каждым ребром в графе. Изучение отношений между взаимодействующими единицами является краеугольным камнем анализа социальной сети (SNA) - совокупностью теорий и методов, которые позволяют анализировать социальные структуры; эти методы специально ориентированы на исследование реляционных аспектов этих структур. Обычно важность отдельного актера (в данном случае члена сообщества) может быть выведена из его или ее местоположения в сети. В большинстве социальных сетей пользовательские взаимоотношения основаны на дружбе или ссылках, созданных посредством обмена электронными приглашениями. Поскольку эти ссылки легко можно отследить, может возникнуть соблазн применения SNA для определения важности человека. Вероятно, это подразумевает, что человек, имеющий более связанные профили, более важен, чем человек с меньшим количеством ссылок.

Для определения наиболее важного объекта подмножества социальной сети нам необходимо выбрать объект с которого будет начинаться сбор данных (центр выборки), получить начальный набор данных о всех связях пользователя, затем построить сеть взаимного влияния пользователей друг на друга.

Для получения наиболее корректной информации о месте объекта в сети необходимо проанализировать взаимодействия всех объектов, не разбивая их на группы и не исключая из рассмотрения ни одного из них, даже если на первый взгляд он кажется абсолютно не значимым.

Используя исходные данные о фактах связей, построим матрицу смежности графа. Матрицей смежности $A=[a_{ij}]$ графа G называется квадратная матрица с размерностью $n \times n$ (где n – кол-во вершин), а элемент a_{ij} определяется по следующему правилу:

$$A[i, j] = \begin{cases} 1, & \text{если объект } i \text{ и объект } j \text{ являются друзьями} \\ 0, & \text{иначе} \end{cases}$$

Введем понятие итерированной силы объекта i порядка k .

$$p^i(0) = 1 \quad i = 1..n \quad (1)$$

$$p^i(k) = \sum_{j=1}^n A[i, j] \times p^j(k-1), \quad i = 1..n, \quad (2)$$

$$k \geq 1$$

Итерированная сила 1-го порядка объекта i – это количество связей рассматриваемого объекта с остальными. На данный момент она не учитывает влияние других объектов подмножества. Начиная со 2-го порядка в данную сумму включается влияние остальных объектов подмножества [3].

Для проведения анализа нам необходимо будет собрать информацию из социальной сети «ВКонтакте». Реализуемый модуль сбора информации будет загружать наборы друзей для 2-го и 3-го уровня (т.е. два набора друзей), таким образом мы получаем выборку с максимальным удалением от центра посредством 3-х связей.

Помимо друзей, модуль сбора информации будет загружать и другие данные, необходимые для анализа:

- список фото, на которые пользователь поставил лайк/репост;
- список видео, на которые пользователь поставил лайк/репост;
- список записей, на которые пользователь поставил лайк/репост.

Анализируем активность пользователя в форме лайков, сколько и кому ставил, ищем максимальное количество лайков в нашей выборке для задания максимального значения.

Далее обозначаем коэффициент влияния пользователей в промежутке от 0 до 1, основываясь на максимальном значении и далее по убыванию.

В результате этого получаем влияние всех пользователей выборки друг на друга. Используя линейную пороговую модель, присваиваем каждому пользователю порог, т.е. вероятность им принимать новую информацию из сети.

После определения всех коэффициентов влияния пользователей друг на друга, необходимо переходить к реализации алгоритма выявления наиболее влиятельного объекта в рассматриваемом подмножестве. Необходимо будет реализовать «жадный» алгоритм, который на каждом шаге будет выбирать пользователя, который мог бы повлиять (сделать активным) на максимальное количество пользователей. Активным пользователь становится только в случае, когда влияние выбранного алгоритмом пользователя будет выше коэффициента влияния рассматриваемого пользователя.

Информация распространяется только если влияние активного пользователя больше чем порог неактивного. Таким образом, после завершения работы алгоритма мы получим 3 группы пользователей:

- пользователи, на которых уже повлияли другие пользователи сети;
- пользователи, которые повлияли на первую группу;
- пользователи, на которых остальные пользователи в сети не имеют влияния.

На современном этапе развития информационных технологий разработка подобной системы является актуальной задачей, позволяющей решить ряд экономических, социальных и бытовых проблем.

Важно отметить, что область применения реализуемой системы для поиска наиболее значимых объектов социальной сети широка и не ограничена, начиная с исследований в различных областях и заканчивая разнообразным коммерческим использованием.

Список литературы / References

1. *Лыфенко М.Д.* Виртуальные пользователи в социальных сетях: Мифы и реальность. [Текст] / Лыфенко М.Д. // Вопросы кибербезопасности, 2014. № 4. С. 1-4.
2. *Градосельская Г.В.* Сетевые измерения в социологии: Учебное пособие / Под ред. Г.С. Батыгина. М.: Издательский дом «Новый учебник», 2004. С. 3.
3. *Habrahabr.ru* Поиск наиболее влиятельных объектов подмножества социальной сети. [Электронный ресурс]. Режим доступа: <https://habrahabr.ru/post/183548/> (дата обращения: 21.11.2017).