

УДК 005

**Обход социального графа для определения людей в социальных сетях,
обладающих общим значением произвольного атрибута с заданной выборкой
пользователей**

И.Г. Шомполов, Р.А. Сидорец

Московский физико-технический институт (государственный университет)
Всероссийский научно-исследовательский институт по проблемам гражданской
обороны и чрезвычайных ситуаций МЧС России» (федеральный центр науки и высоких
технологий) (ФГБУ ВНИИ ГОЧС (ФЦ))

В работе описан метод поиска множества пользователей в социальной сети “ВКонтакте”, обладающих общим значением произвольного атрибута. Обладая информацией о заданной выборке пользователей заранее, посредством найденного объединения или же списка группы лиц, обладающих общим значением произвольного атрибута, можно найти большее (отличное от выбранного) множество с той же характеристикой. Данная задача имеет практическое бизнес-применение и изучено недостаточно подробно. Рассмотрена возможность применения метода в области работы с абитуриентами МФТИ или же поиск объединений студентов по тем или иным интересам.

В частности, в работе исследовались пользователи социальной сети ВКонтакте (<http://vk.com/>). В качестве связей, определяющих общий атрибут пользователей, была выбрана информация о принадлежности пользователя к тем или иным группам ВКонтакте, встречам (далее – группа).

1. Начальными (входными) данными служит произвольное подмножество пользователей социальной сети (начальное). В нашем случае пользователи (их уникальные идентификаторы) в сети ВКонтакте. Нам известно, что данные пользователи имеют схожий атрибут, природу которого, сам факт его наличия и детерминированность в данной работе не обсуждается.
2. Данный атрибут может быть интересом, увлечением, желанием купить что-то или же территориальной принадлежностью. Мы определяем это как некоторое общее значение атрибута для данной группы пользователей.
3. Мы предполагаем, что информация об атрибуте (о его значении/наличии) для всех пользователей социальной сети заложена в социальных связях принадлежности к группе. Как для начального подмножества, так и для искомого
4. Формализуя задачу, мы имеем ненаблюдаемый социальный граф пользователей и их связей с множеством групп. Однако, информация о связях является доступной и опирается лишь на производительность вычислительных машин и ограничения API
5. Задача ставится в нахождении подмножества пользователей (искомое) с тем же значением атрибута.

Целью данной работы является исследование и разработка метода поиска подмножества пользователей (далее – *искомое подмножество*) социальной сети, обладающих общим значением произвольного атрибута, природу которого, вообще говоря, не обсуждаем. Тестирование метода будет проводиться путем применения метода к заданной выборке пользователей (далее – *исходное множество*) с заведомо заданным атрибутом. Мерой точности метода будут выступать стандартные метрики Precision, Recall и F1-мера. Для достижения цели необходимо решить следующие задачи:

1. Исследовать предметную область, изучить существующие методы кластеризации пользователей и вычисления значения того или иного конкретного атрибута.

2. Разработать и реализовать по меньшей мере 2 метода поиска людей с заданным значением атрибутов, на основе связей типа «Друзья» и информации о принадлежности группам
3. Провести экспериментальное исследование и сравнение разработанных алгоритмов.

Основной задачей, поставленной в данной работе является исследование и разработка метода поиска *искомого подмножества* на основе информации о членстве пользователей в группах. Соответственно решение данной задачи включает в себя следующие пункты:

1. Реализация метода поиска пользователей, обладающих общим значением произвольного атрибута на основе информации о членстве пользователей в группах и *исходном множестве*
2. Подбор параметров фильтров для получения оптимального результата
3. Сравнение результатов с «наивным» методом и методом анализа дружественных связей

В данной работе исследовались и разрабатывались методы поиска подмножества пользователей социальной сети, обладающих общим значением произвольного атрибута с заданной выборкой пользователей путем обхода социального графа. Все поставленные задачи были выполнены, в частности:

1. Исследована предметная область, изучены существующие методы кластеризации пользователей и вычисления значения того или иного конкретного атрибута.
2. Разработаны и реализованы 2 метода поиска людей с заданным значением атрибутов, на основе связей типа «Друзья» и информации о принадлежности группам
3. Проведено экспериментальное исследование и сравнение разработанных алгоритмов

Литература

1. *D. M. Boyd, N.B. Ellison.* Social network sites: Definition, history, and scholarship // Journal of Computer-Mediated Communication, 2007, 13(1), article 11
2. *George Pallis, Demetrios Zeinalipour-Yazti, Marios D. Dikaiakos.* Online Social Networks: Status and Trends // New Directions in Web Data Management 1, Studies in Computational Intelligence Volume 331, 2011, pp 213-234
3. *Najork M., Wiener J. L.* Breadth-first crawling yields high-quality pages // Proceedings of the 10th international conference on World Wide Web. – ACM, 2001. – С. 114-118.
4. *Leskovec J., Faloutsos C.* Sampling from large graphs // Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. – ACM, 2006. – С. 631-636.
5. *Nazar Buzun, Anton Korshunov.* Innovative Methods and Measures in Overlapping Community Detection // Proceedings of the International Workshop on Experimental Economics and Machine Learning (EEML 2012), Brussel, Belgium
6. Facebook Open Graph. <https://developers.facebook.com/docs/opengraph/>
7. Social Network Data Analytics. Editors: Charu C. Aggarwal // Springer, 2011
8. *Шомполов И.Г.* Новые образовательные технологии научно-педагогической системы выявления, отбора и методического сопровождения одаренных школьников в рамках межвузовской системы образования в московском физико-техническом институте в 2014/2015 учебном году. // МФТИ. – 2015.
9. *Корицунов А.* Определение демографических атрибутов пользователей микроблогов // Труды Института системного программирования РАН Том 25. 2013 г. Стр. 179-194.
10. *Шомполов И.Г., Сидорец Р.А.* Обход социального графа для определения людей в социальных сетях, обладающих общим значением произвольного атрибута с заданной выборкой пользователей // «European research» 9 (20)
11. *Шомполов И.Г., Сидорец Р.А.* Обход социального графа для определения людей в социальных сетях, обладающих общим значением произвольного атрибута с заданной выборкой пользователей // «Проблемы современной науки и образования» 26 (68)