



Научная статья

УДК 311.216+311.42

## Разработка алгоритма для сбора социологических данных из онлайн-сообществ университетов в социальной сети ВКонтакте



А. А. Фомин

Саратовский национальный исследовательский государственный университет имени Н. Г. Чернышевского, Россия, 410012, г. Саратов, ул. Астраханская, д. 83

Фомин Андрей Анатольевич, заведующий лабораторией социологических исследований и информатики кафедры социальной информатики, [Fominaa@sgu.ru](mailto:Fominaa@sgu.ru), <https://orcid.org/0000-0003-1075-2962>

**Аннотация.** В статье представлены данные исследования онлайн-сообществ Саратовского государственного университета, которое проводилось на базе лаборатории социологических исследований и информатики в период с января по июнь 2022 г. В ходе исследования разработан алгоритм на языке Python для поиска онлайн-сообществ по определенным критериям, построен социальный граф, который в дальнейшем был преобразован с помощью алгоритма кластеризации. Полученные результаты исследования могут быть использованы для повышения узнаваемости вуза в социальных сетях, что повлияет на увеличение престижа вуза и привлечение в него большего числа абитуриентов.

**Ключевые слова:** социальные сети, сбор данных, алгоритм, онлайн-сообщества, ВКонтакте, университет

**Для цитирования:** Фомин А. А. Разработка алгоритма для сбора социологических данных из онлайн-сообществ университетов в социальной сети ВКонтакте // Известия Саратовского университета. Новая серия. Серия: Социология. Политология. 2022. Т. 22, вып. 4. С. 413–417. <https://doi.org/10.18500/1818-9601-2022-22-4-413-417>, EDN: MIGCQE

Статья опубликована на условиях лицензии Creative Commons Attribution 4.0 International (CC-BY 4.0)

Article

### Development of the algorithm for collecting sociological data from universities' online communities in the social network VKontakte

A. A. Fomin

Saratov State University, 83 Astrakhanskaya St., Saratov 410012, Russia

Andrey A. Fomin, [Fominaa@sgu.ru](mailto:Fominaa@sgu.ru), <https://orcid.org/0000-0003-1075-2962>

**Abstract.** The article tells about the study of online communities at Saratov State University. In the course of the study, a Python algorithm was developed to search for online communities according to certain criteria and a social graph was built, which was further transformed using a clustering algorithm. The obtained results of the study can be used to increase the recognition of the university in social networks, which will raise the prestige of the university and attract more applicants.

**Keywords:** social networks, data collecting, algorithm, online communities, VKontakte, university

**For citation:** Fomin A. A. Development of the algorithm for collecting sociological data from universities' online communities in the social network VKontakte. *Izvestiya of Saratov University. Sociology. Politology*, 2022, vol. 22, iss. 4, pp. 413–417 (in Russian). <https://doi.org/10.18500/1818-9601-2022-22-4-413-417>, EDN: MIGCQE

This is an open access distributed under the terms of Creative Commons Attribution 4.0 International License (CC-BY 4.0)

Развитие социальных сетей в последнее время открыло многим исследователям доступ к новой информации об интересах людей, в связи с этим большую популярность набирает такое направление, как анализ социальных сетей.

Анализ социальных сетей является междисциплинарной проблемой, в него входят социология, компьютерные науки, психология, дискретная математика. На данный момент не

существует общепринятого математического определения онлайн-сообщества. Часто под ним понимают группу узлов или группу людей, которые связаны внутри группы теснее между собой, чем со всем остальным миром [1].

Также анализ социальных сетей становится актуальным в связи с увеличением рынка образовательных услуг. Для этого вузам необходимо совершенствовать свои маркетинговые



коммуникации с целью привлечения аудитории. В настоящее время неотъемлемой частью деятельности современных высших образовательных учреждений является разработка собственных программ продвижения с использованием разнообразных средств маркетинговых коммуникаций. Грамотно используя коммуникационные инструменты с применением современных технологий, учебное заведение может не только добиться привлечения к себе большего количества потенциальных абитуриентов, но и расширить границы своей деятельности, что позволит ему выгодно отличаться от своих конкурентов. На данном этапе большинство вузов используют традиционные способы продвижения, такие как реклама, связи с общественностью, участие в выставках, проведение дней открытых дверей и др. Однако сегодня одним из приоритетных направлений работы маркетинговой деятельности в вузе становится продвижение образовательных услуг через интернет-каналы.

Для анализа социальных данных требуется специализированное программное обеспечение. Для сбора информации из различных онлайн-источников используются специализированные алгоритмы под названием «Парсеры». Парсер – это программа для сбора и систематизации информации, размещенной на различных онлайн-ресурсах, источником данных могут выступать как текстовая информация, так и медиа-данные (фото, аудио и видео). Часто под каждую задачу разрабатывается уникальный

алгоритм, который нацелен на решение узкого спектра задач исследования. С примерами разработанных нами алгоритмов можно ознакомиться на странице лаборатории социологических исследований и информатики (<https://www.sgu.ru/structure/sociological/structuresociological>).

Анализ социальной сети также можно рассматривать под углом исследования связанных структур пользователей и отношений между ними. В представленном исследовании объектами выступили онлайн-сообщества Саратовского государственного университета в социальной сети «ВКонтакте». Собранные данные были использованы для расчета математических характеристик, после чего визуализированы при построении социальных графов.

Для анализа связей онлайн-сообществ Саратовского государственного университета мы разработали алгоритм на языке Python для автоматического поиска сообществ. Также мы добавили в алгоритм критерии для фильтрации. В ходе исследования нами рассматривались только те сообщества, в которых более 50 участников и последняя запись на стене сообщества датировалась текущем годом, что говорило об активности и ведении модерации сообщества.

В ходе первого этапа исследования мы получили список из 119 сообществ и связи между этими сообществами, на основании чего смогли построить социальный граф со всеми связями между сообществами университета (рис. 1).

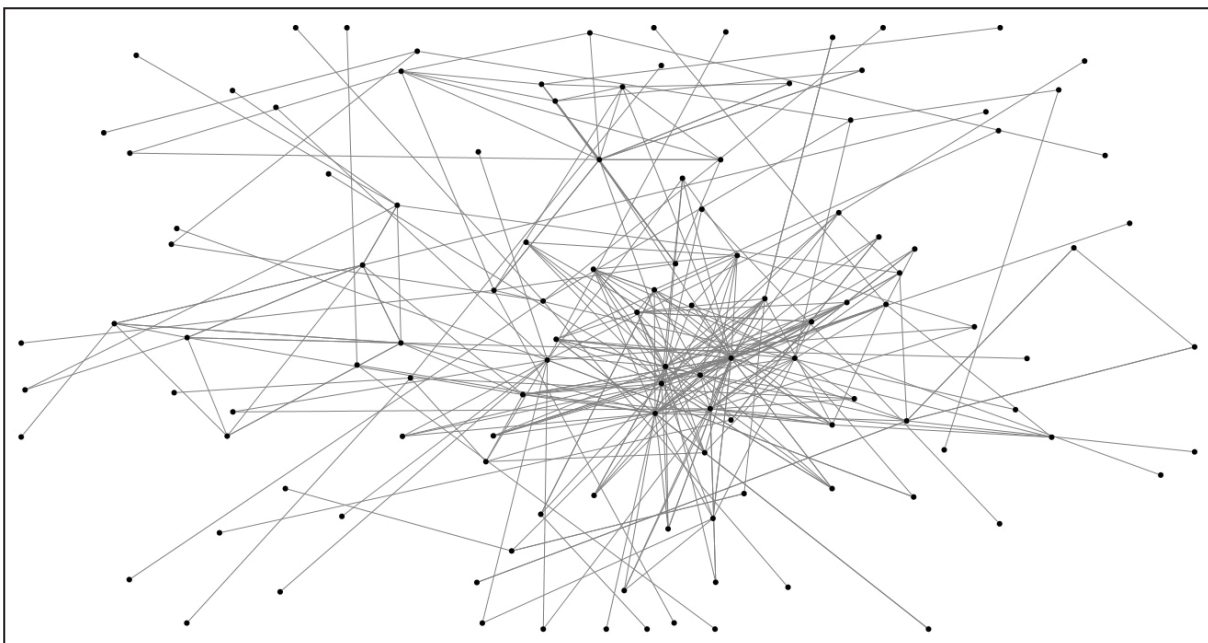


Рис. 1. Социальный граф онлайн-сообществ СГУ



Построив математическую модель представленного графа, можно сказать, что он имеет максимальный диаметр 7, что является наибольшим расстоянием между вершинами, однако среднее геодезическое равняется 3,6, что говорит о том, что внутри сети сообщества находятся друг от друга в среднем на расстоянии трех с половиной связей. Плотность представленного графа составляет всего 2%, что говорит о его слабой загруженности.

На следующем этапе нашего исследования необходимо было разбить представленный выше граф на сообщества, что было проведено при помощи алгоритма Клозета – Ньюмана – Мура. В результате произошло разбиение графа на части на основе плотности связей. При помощи данного алгоритма вычисляются связи между узлами, принадлежащими разным кластерам, и узлами одного кластера. Сообщество можно охарактеризовать как часть графа, в которой средняя плотность между узлами превышает плотность связей между остальными подграфами. Алгоритм Клозета – Ньюмана – Мура вычисляется по формуле

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{d_i d_j}{2m}) \delta(C_i, C_j),$$

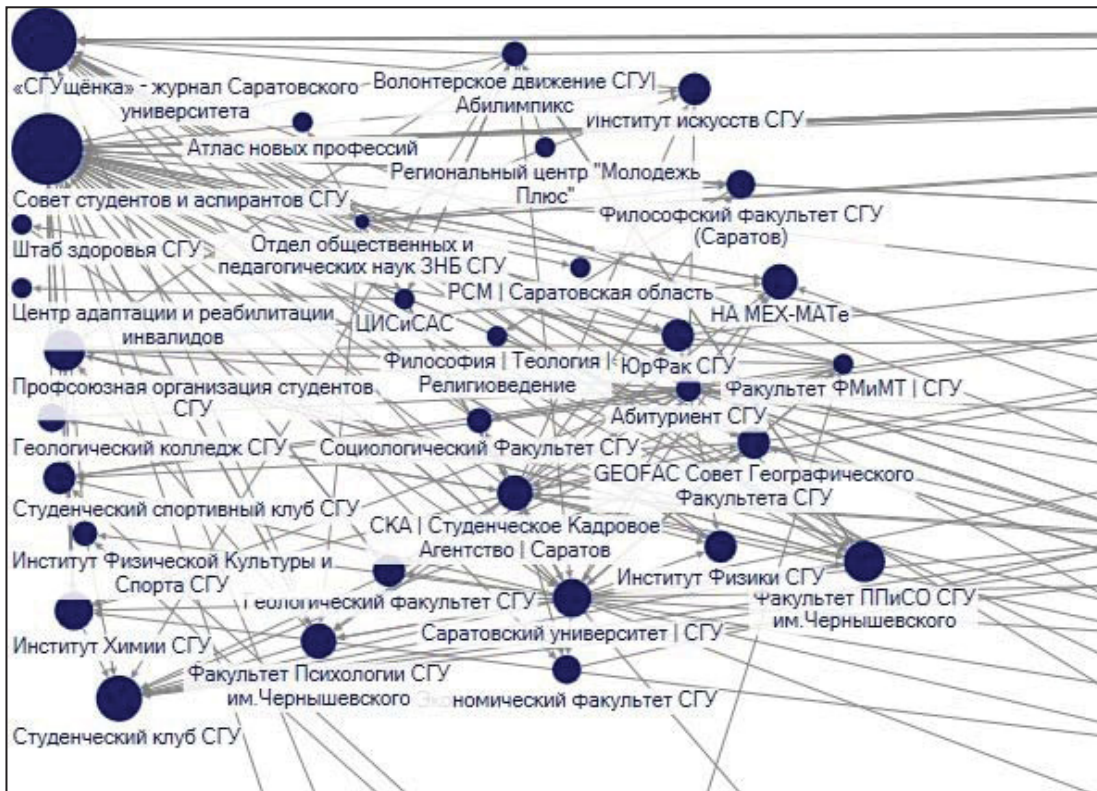
где  $A$  – матрица смежности графа,  $A_{ij}$  –  $(i, j)$  элемент матрицы,  $d_i$  – степень  $i$  вершины графа,

$C_i$  – метка вершины (номер сообщества, к которому относится вершина),  $m$  – общее количество ребер в графе.  $\delta(C_i, C_j)$  – дельта-функция: равна единице, если  $C_i = C_j$ , если иначе – нулю.

В результате нами были получены 11 кластеров:

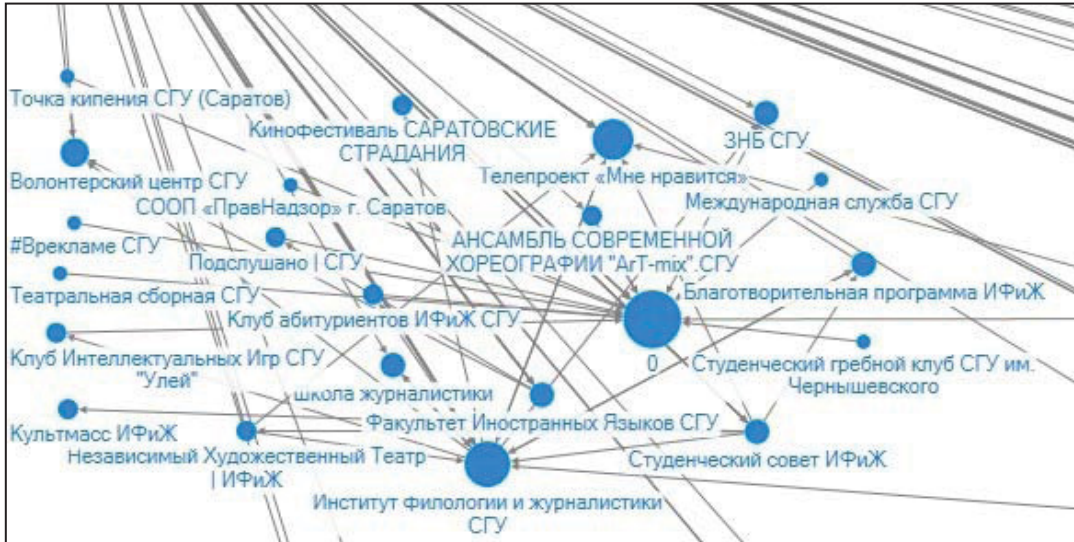
1) в первый кластер вошли самые крупные сообщества и официальные сообщества факультетов и институтов. Данный кластер состоит из 32 вершин и 117 уникальных связей. Хотя доля взаимных связей внутри кластера равняется 0,157, что является достаточно низким показателем (рис. 2, а);

2) второй по величине кластер с 22 вершинами относится больше к внеучебной деятельности (рис. 2, б), хотя в него вошли официальные страницы Института филологии и журналистики и официальная страница факультета компьютерных наук и информационных технологий, что может быть связано с недавними изменениями адреса или названий сообществ. Также на рис. 2, б заметно крупное сообщество «0», которым мы обозначили отсутствие ссылок из сообществ. Размер данного «пустого» сообщества говорит о том, что среди сообществ СГУ много таких, которые получают только внешние ссылки на себя, но сами ни на кого не ссылаются;



а

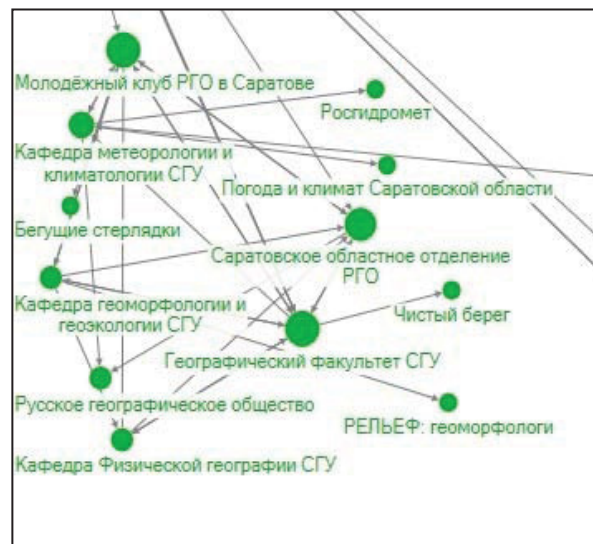




а



б



в

Рис. 2. Самые крупные кластеры онлайн-сообществ СГУ: а – официальные сообщества, б – внеучебная деятельность, в – студенческие отряды, г – географические сообщества (цвет online)

3) третий кластер представляет собой сбор сообществ, посвященных студенческим отрядам. Несмотря на наличие внутри данного кластера двух крупных сообществ: «Студенческие отряды ПФО» и «Штаб студенческих отрядов СГУ», доля взаимных связей внутри кластера достаточно низкая и равняется 0.188 (рис. 2, в);

4) четвертый кластер состоит из географических сообществ во главе с официальной страницей географического факультета. В данном кластере самый высокий уровень взаимных связей, равный 0,368 (рис. 2, г);

5) остальные кластеры имеют достаточно низкое количество вершин (в среднем 3) и в большинстве своем посвящены конкретному внеучебному направлению, например КВН, кино клуб и т. д.

В заключение можно подвести итоги. В ходе исследования нами был разработан алгоритм по поиску онлайн-сообществ в социальной сети «ВКонтакте», с помощью которого обнаружены 119 сообществ, удовлетворяющих нашим критериям. Диаметр полученной сети – 6 связей. Это означает, что кратчайший путь



между двумя наиболее удаленными сообществами состоит лишь из 6 связей. Плотность всей сети равна всего 2.2%. Данный критерий свидетельствует о том, что данная сеть является незагруженной, сообщества внутри сети крайне редко ссылаются друг на друга. По количеству входящих связей заметно выделяется сообщество «Совет студентов и аспирантов СГУ», также часто цитируются такие сообщества, как «СГУщётка – журнал Саратовского университета», «Институт филологии и журналистики СГУ» и «Студенческий клуб СГУ». По количеству исходящих связей лидирует официальное

сообщество вуза – «Саратовский университет | СГУ». Основываясь на полученных результатах данного исследования, можно разработать ряд рекомендаций для высших учебных заведений по самопрезентации вуза в социальных сетях и увеличить продуктивность ведения онлайн-сообществ.

#### Список литературы

1. Гусарова Н. Ф. Анализ социальных сетей. Основные понятия и метрики. СПб. : Университет ИТМО, 2016. 67 с.

Поступила в редакцию 08.08.2022; одобрена после рецензирования 19.08.2022; принята к публикации 29.08.2022  
The article was submitted 08.08.2022; approved after reviewing 19.08.2022; accepted for publication 29.08.2022