

УДК 004.738: 519.7

DOI 10.21661/r-472064

*А.А. Печников, А.Н. Ермолаев*

## **ПРОГРАММА ДЛЯ ИССЛЕДОВАНИЯ МАЛЫХ СООБЩЕСТВ «ВКОНТАКТЕ» И НЕКОТОРЫЕ РЕЗУЛЬТАТЫ**

*Аннотация:* авторы сосредоточили внимание на исследовании малых сообществ социальной сети ВКонтакте. Для этих целей была разработана специализированная программа сбора открытых данных о таких сообществах, размещенная в публичном домене. Программа проста в использовании и позволяет как собирать некоторые данные об участниках, так и вычислять основные метрики малых сообществ. Формат выходных данных дает возможность использовать их без предварительной обработки в программе анализа и визуализации сетей Gephi, что существенно усиливает возможности исследователей. Приводятся некоторые результаты исследований, показывающие возможности программы.

*Ключевые слова:* социальная сеть, ВКонтакте, программа, граф, метрики, ассортативность, предпочтительное присоединение.

*А.А. Pechnikov, A.N. Ermolaev*

## **SOFTWARE FOR THE STUDY OF SMALL COMMUNITIES VKONTAKTE AND SOME OF THE RESULTS**

*Abstract:* authors focused on study of small communities of the social network V Kontakte. For this purpose, specialized program for collecting open data from such communities was developed, which is available in the public domain. The program is easy to use and allows you to collect data about the participants and calculate basic metrics of small communities. The output format is organized in a such way that it can be used without any additional processing in the program for analysis and visualization of networks – Gephi, which significantly enhances the capabilities of researchers. Numerical simulations were made to demonstrate features of the program.

**Keywords:** *social network, VKontakte, software, graph, metrics, assortativity preferential attachment emergence.*

### *Введение*

Стремительный рост пользователей социальных сетей привёл к тому, что внимание исследователей в области социальных графов сфокусировалось главным образом именно на лидерах рынка. По состоянию на май 2018 года самой популярной социальной сетью в России является ВКонтакте, имеющая около 17 млн активных участников [1]. Этим обуславливается актуальность данной работы. Мы сосредоточили внимание на так называемых «малых сообществах», к которым в этой работе отнесем: эго-сообщество, эго-эго-сообщество, тематическую группу.

Эго (или центральный индивид) идентифицируется по его уникальному идентификатору *id*. Эго-сообщество это сам эго, другие индивиды, с которыми он поддерживает отношения, связи эго с этими другими индивидами и связи между другими индивидами (участниками).

В свою очередь эго-эго-сообщество можно определить так: все индивиды, входящие в эго-сообщество центрального индивида (первичного), принимаемые за вторичных центральных индивидов; все индивиды, входящие в эго-сообщества вторичных центральных индивидов; все связи между всеми индивидами.

Тематическая группа – это индивиды, входящие в тематическое сообщество, идентифицируемое по его названию, и все связи между ними. К тематической группе с 1 млн участников термин «малое сообщество» плохо применим, поэтому в этом случае мы ограничим число участников тысячами, учитывая имеющиеся вычислительные ресурсы.

### *Программа SGMaker*

Для исследования малых сообществ была разработана программа SGMaker. При разработке использовались следующие программные средства:

– Python 3.6 – интерпретируемый язык программирования высокого уровня общего назначения;

- NetworkX – библиотека Python для создания, обработки и изучения структуры, динамики и функций сложных сетей;
- VK API ВКонтакте – интерфейс прикладного программирования (application programming interface – API), позволяющий получать необходимую информацию из базы данных vk.com с помощью http-запросов к специальному серверу;
- vk\_requests – библиотека Python для работы с VK API;
- PyQt – «привязка» кроссплатформенного фреймворка Qt для Python, реализованная в виде его модуля, позволяющая упростить создание графического интерфейса пользователя (graphical user interface – GUI);
- Gephi – интерактивная платформа визуализации и исследования для сетей и сложных систем, динамических и иерархических графов.

Далее «json-ответ» обозначает результат выполнения определённой функции в формате JSON (англ. JavaScript Object Notation), текстовом формате обмена данными, основанном на JavaScript, а id является идентификатором участника социальной сети или именем тематической группы.

Для обработки запросов к ВКонтакте через VK API используются функции, реализованные с помощью библиотеки vk\_requests и возвращающие json-ответ с некоторыми из полей: id, имя, фамилия, город проживания, дата рождения, список друзей запрашиваемого участника и список участников запрашиваемого сообщества. На основе полученных данных основная функция строит графы, используя структуру библиотеки networkx, после чего сохраняет готовые социальные графы в папку '/graphs'. Для оценки характеристик имеющихся графов реализованы функции, позволяющие вычислить возраст по указанной дате рождения (если указанных данных недостаточно, то результат 'n/a'), плотность и коэффициенты ассортативности по возрасту и городу для заданного графа.

Для использования программы можно скачать архив программы, доступный по ссылке <http://webometrics.krc.karelia.ru/section.php?id=23>, распаковать его и запустить исполняемый файл SGMaker.exe, размещенный в папке SGMaker. В Windows 8 и выше может потребоваться отключение функции SmartScreen,

защищающей от запуска не доверительных файлов, сомнительных сайтов или загрузки вирусных приложений.

Пример интерфейса SGMaker приведен на рис. 1. После ввода идентификатора id следует выбрать одну из опций Type (Friends – это-сообщество, Friend’s Friends – это-это-сообщество, Group – тематическая группа).

В поле Save as нужно ввести имя, под которым сохранятся два файла в директории приложения в папке '/graphs' в формате Gephi: в файле age.gexf сохранится структура графа с пометками узлов «возраст», а в файле city.gexf – с пометками «город». Далее следует нажать кнопку Download graph и дождаться загрузки графа. Процесс выполнения можно отслеживать на консоли (рис. 2), успешное завершение подтверждается Completed!, остальные варианты означают ошибку ввода id.

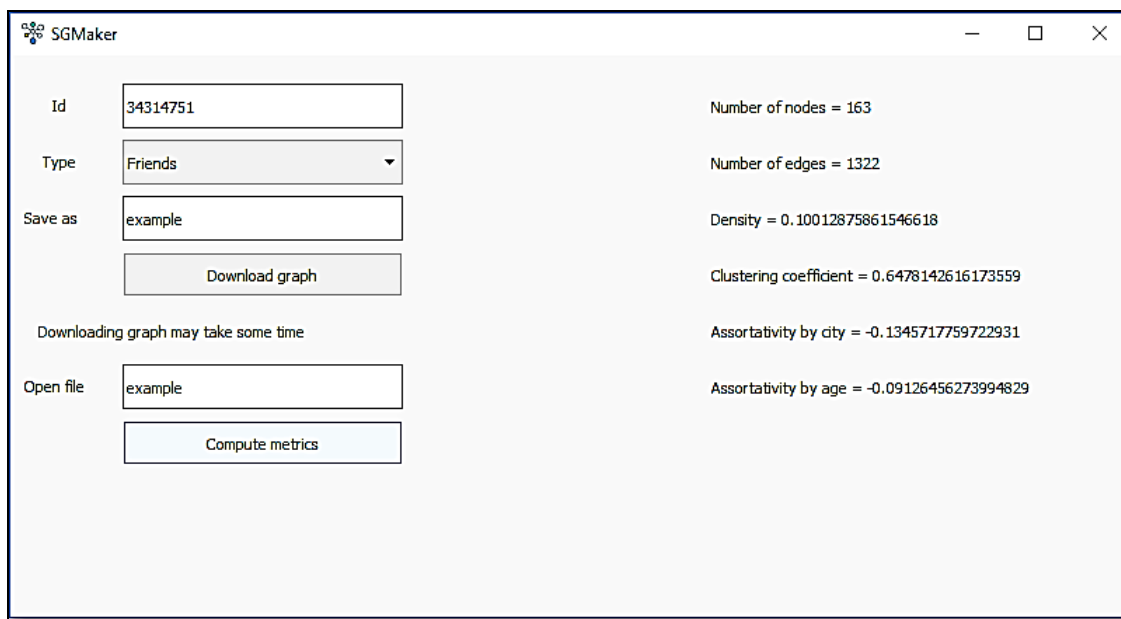


Рис. 1. Пример работы с интерфейсом в варианте Friends

Для только что загруженного файла достаточно нажать на Compute metrics и в правой части для этого социального графа появится информация о количестве узлов, рёбер, плотности, коэффициенте кластеризации, коэффициента ассортативности по возрасту и городу (рис. 1).

```

C:\Users\SGMaker.exe
87.73006134969326 %
88.34355828220859 %
88.95705521472392 %
89.57055214723927 %
90.1840490797546 %
90.79754601226993 %
91.41104294478528 %
92.02453987730061 %
92.63803680981594 %
93.25153374233128 %
93.86503067484662 %
94.47852760736197 %
95.0920245398773 %
95.70552147239265 %
96.31901840490798 %
96.93251533742331 %
97.54601226993866 %
98.15950920245399 %
98.77300613496932 %
99.38650306748467 %
100.0 %
Importing graf...
Completed!
--- 16.651442289352417 seconds ---

```

Рис. 2. Пример успешного завершения процесса

Для вычисления метрик ранее загруженного графа достаточно указать в поле Open file название файла, сохраненного в папке '/graphs', и нажать на Compute metrics. Эти же файлы могут быть использованы в качестве исходных данных для работы в Gephi.

### *Некоторые результаты*

Был построен эго-граф одного из авторов статьи (id=34314751), содержащий 159 вершин и 1322 рёбер. Для каждой из вершин этого графа были построены свои эго-графы, имеющие от 12 до 2548 вершин и от 13 до 152695 ребер. В среднем построенные эго-графы имеют низкую плотность  $\sim 0.08$ , и при этом высокий коэффициент кластеризации  $\sim 0.6$ . То есть пользователи склонны объединяться в небольшие, но тесно связанные группы. Размерность наибольшей максимальной клики [4] в среднем близка к 6.5, из чего следует, что пользователи предпочитают «тесно дружить» по 6–7 человек.

Характерным представителем небольших тематических групп является Creedence Clearwater Revival (имя – c\_c\_r), содержащая 526 участников и 84 ребра. Граф чрезвычайно разреженный, 434 участника не имеют друзей в этой группе, выделяется компонента связности, содержащая 55 участников.

Эго-эго граф, для которого первичным индивидом был взят тот же участник с id=34314751, содержит 35138 вершин и 677460 ребер. По своим свойствам он

очень близок к известному графу Барабаши-Альберт [5]. Во-первых, это весьма разреженный граф с плотностью 0.001. Во-вторых, диаметр графа очень мал и равен 4. В-третьих, у графа весьма характерное распределение степеней вершин. Эмпирическая вероятность того, что вершина графа имеет степень  $d$ , оценивается как  $c/d^\lambda$ , где  $\lambda=1.797$ , что весьма близко к теоретическому значению  $\lambda \approx 2.1$  ( $c$  – нормирующий множитель, вычисляемый из условия «сумма вероятностей равна 1»). Похоже, что здесь работает принцип предпочтительного присоединения, когда новый участник преимущественно хочет дружить с теми участниками, у которых много друзей.

Все три приведенных результата и разработанная программа открывают широкий спектр задач для исследователей.

### ***Список литературы***

1. Статистика социальных сетей [Электронный ресурс]. – Режим доступа: <https://br-analytics.ru/statistics/author>

2. Newman M.E.G. Assortative mixing in networks // *Physical Review Letters*. – 2002. – Vol. 89. – P. 1–5.

3. Definition of dense graph, possibly with links to more information and implementations [Электронный ресурс]. – Режим доступа: <https://xlinux.nist.gov/dads/HTML/densegraph.html>

4. Moon J.W. On cliques in graphs / J.W. Moon, L.Moser // *Israel Journal of Mathematics*. – 1965. – Vol. 3. – Is. 1. – P. 23–28.

5. Barabasi L.-A. Scale-free characteristics of random net-works: the topology of the world-wide web / L.-A. Barabasi, R. Albert, H. Jeong // *Physica*. – 2000. – Vol. A281. – P. 69–77.

### ***References***

1. Statistika social'nyh setei [Elektronnyi resurs]. – Rejim dostupa: <https://br-analytics.ru/statistics/author>

2. Newman M.E.G. Assortative mixing in networks // *Physical Review Letters*. – 2002. – Vol. 89. – P. 1–5.

3. Definition of dense graph, possibly with links to more information and implementations [Elektronnyi resurs]. – Rejim dostupa: <https://xlinux.nist.gov/dads/HTML/densegraph.html>

4. Moon J.W. On cliques in graphs / J.W. Moon, L.Moser // Israel Journal of Mathematics. – 1965. – Vol. 3. – Is. 1. – P. 23–28.

5. Barabasi L.-A. Scale-free characteristics of random net-works: the topology of the world-wide web / L.-A. Barabasi, R. Albert, H. Jeong // Physica. – 2000. – Vol. A281. – P. 69–77.

---

**Печников Андрей Анатольевич** – д-р техн. наук, канд. физ.-мат. наук, главный научный сотрудник Института прикладных математических исследований ОП ФГБУН Федерального исследовательского центра «Карельский научный центр РАН», Россия, Петрозаводск.

**Pechnikov Andrei Anatolevich** – doctor of technical sciences, candidate of physical and mathematical sciences, chief research assistant at the Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences, Russia, Petrozavodsk.

**Ермолаев Алексей Николаевич** – студент ФГБОУ ВО «Санкт-Петербургский государственный университет», Россия, Санкт-Петербург

**Ermolaev Aleksei Nikolaevich** – student at the Saint-Petersburg State University, Russia, Saint-Petersburg.

---