

# РАЗРАБОТКА МОДЕЛИ СЛУЧАЙНОГО ГРАФА, СООТВЕТСТВУЮЩЕЙ УСРЕДНЕННЫМ ХАРАКТЕРИСТИКАМ СОЦИАЛЬНЫХ ГРАФОВ, ПОЛУЧЕННЫМ ПО ДАННЫМ СОЦИАЛЬНОЙ СЕТИ TWITTER

М.Э. Грачева, В.В. Степаненко, А.А. Аманжолов, Лунева Е.Е.

Научный руководитель: Е.Е. Лунева  
Томский политехнический университет  
gramar\_98.98@mail.ru

## Введение

За относительно небольшой промежуток времени, после своего появления, Интернет стал основой современного информационного пространства и основным средством для передачи разного рода информации.

Одной из важнейших составляющих Интернета являются социальные сети, которые в настоящее время получили широкое распространение. У каждого человека появилась возможность практически моментально делиться своим мнением с огромными группами людей и получать от них ответные реакции. Социальные сети стали важным инструментом информационного влияния и, в связи с этим, огромной сферой для различных исследований, включая маркетинговые, политические, в том числе, формирование прогнозов, исследования, связанные с задачами антитеррористической направленности, социологические и некоторые другие.

Важное место при анализе данных из социальных сетей занимает проблема идентификации ключевых пользователей некоторой социальной сети в заданной предметной области [1]. При этом в качестве модели при решении данной проблемы используется ориентированный взвешенный социальный граф.

Развитие новых способов решения данной задачи, а также оценка эффективности существующих способов требует на этапе экспериментального анализа использования как социальных графов, построенных по реальным данным, так и моделей случайных графов, обладающих характеристиками реальных социальных графов.

Таким образом, целью данной работы является разработка модели случайного ориентированного графа, а также программного обеспечения для генерации моделей, предназначенных для тестирования способов идентификации пользователей-экспертов в заданной предметной области.

## Обзор существующих моделей

На этапе выбора генератора модели социального графа для проведения экспериментального анализа был проведен обзор существующих моделей случайных графов.

При рассмотрении различных моделей учитывались следующие характеристики: диаметр графа, среднее расстояние между вершинами, закон распределения степени вершин, коэффициент кластеризации.

Модель Эрдеша-Реньи. Каждое ребро возникает с вероятностью  $p$  независимо от других. Плюсом данной модели является малое среднее расстояние. Графы, построенные на основе этой модели, имеют маленький коэффициент кластеризации, ребра возникают согласно схеме Бернулли, а не по степенному закону [2].

Модель с предпочтительным распределением Барабаши-Альберт. Основные идеи модели: постепенный рост сети, новые вершины с большей вероятностью присоединяются к тем, у кого текущая степень выше. Данная модель примечательна тем, что имеет степенной закон распределения, маленький диаметр, небольшой коэффициент кластеризации [3].

Также была рассмотрена модель случайных графов с нелинейным правилом предпочтительного связывания В.Н. Задорожного, которая позволяет выращивать случайные графы с использованием механизма предпочтительного связывания, при этом вероятность связывания зависит не от степени связности анализируемого узла, а от функции предпочтения [4].

Большинство моделей направлено на работу с неориентированными графами, хотя следует отметить, что модель Задорожного В.Н. может быть использована для генерации ориентированного графа, но в рамках настоящего исследования ее использование осложнено отсутствием эталонной модели социального графа.

## Разработка модели

Был проведен анализ десяти реальных графов, составленных на основе данных выгруженных из социальной сети Twitter. Выгруженные данные представляют собой группу пользователей социальной сети, которые отреагировали на определенное событие. Узлы графов, созданных на основе этих данных, представляют собой пользователей, которые написали пост по определенной теме или каким-либо образом отреагировали на пост другого пользователя по этой теме. Ребро между вершинами возникает в следующих случаях:

1. Комментарий к посту другого пользователя
2. Репост записи другого пользователя
3. Упоминание другого пользователя в своем твите

Исследуемые графы являются направленными, ребра направлены в сторону вершины, представляющей пользователя, отреагировавшего на пост.

По результатам анализа, реальные социальные графы обладают следующими характеристиками:

коэффициент кластеризации в диапазоне 0 – 0.104, диаметр сети в диапазоне 2 – 5.

Таким образом, оценивая характеристики социальных графов, построенных по выборкам данных из социальной сети Twitter, было принято решение использовать модель предпочтительного связывания, с ее адаптацией под ориентированный социальный граф.

Случайный ориентированный граф формируется из небольшого графа «затравки», с количеством вершин от 3 до 6. На рисунке 1 представлен пример графа затравки.

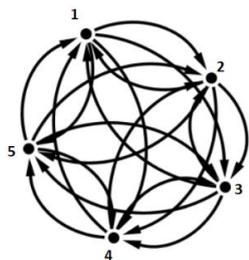


Рис. 1. – Пример графа затравки, состоящего из 5 узлов

Далее добавляется новая вершина, которая связывается направленной входной дугой с другими вершинами в соответствии с моделью Барабаша-Альберта. Таким образом, вероятность связывания новой вершины с существующими определяется по формуле:

$$P(k_i) = \frac{k_i}{\sum_j k_j}, \quad (1)$$

Где  $i$  – текущая вершина, с которой происходит связывание новой вершины,  $k_i$  – полустепень исхода,  $j=1, \dots, N$ ,  $N$  – текущее число вершин в графе.

Из формулы (1) видно, что чем больше полустепень исхода связности у вершины, тем выше вероятность её выбора. Такой способ построения исходящих связей имитирует большую значимость тех пользователей, чьи сообщения находят более активный отклик среди читателей в виде репостов, комментариев и т.п.

Входные дуги с новой добавляемой вершиной формируются с постоянной вероятностью независимо для каждого узла в соответствии с моделью Эрдеша-Реньи. Предлагается использовать небольшие значения данной вероятности 0,001-0,02. Такой способ построения дуг, позволяет построить

«горизонтальные», связи социального графа, т.е. имитировать заинтересованность любого пользователя в социальном графе другим пользователем вне зависимости от его значимости.

На основании разработанной модели случайного ориентированного графа была создана программа для генерации графов.

Были сгенерированы и проанализированы графы с числом вершин от 300 до 500, средние значения их характеристик составили: коэффициент кластеризации – 0.127, диаметр сети – 7.

### Заключение

В результате проведенной работы был выполнен анализ существующих моделей социальных графов. На основании этого анализа и на основании изучения характеристик реальных графов была разработана модель случайного ориентированного графа.

Для того чтобы изучить характеристики графов построенных, на основании разработанной модели, была создана программа для генерации графов.

Полученная модель обеспечивает генерацию графа с характеристиками схожими с характеристиками реальных графов, созданных на основе данных выгруженных из Twitter.

В дальнейшем разработанная модель может быть усовершенствована путем добавления весов ребрам, в зависимости от типа активности одного пользователя по отношению к другому.

Работа выполнена при финансовой поддержке РФФИ (проект №17-07-00034 А).

### Список использованных источников

1. Ефремов А. А. Использование процедуры ранжирования Кендалла–Уэя для идентификации ключевых игроков социального графа / А. А. Ефремов [и др.] // Доклады ТУСУР. – 2018. – Т. 21, № 1. – С. 80–85. DOI: 10.21293/1818-0442-2018-21-1-80-85.
2. Erdős P., Rényi A. On random graphs I. // Publ. Math. Debrecen. – 1959. – Vol. 6. – С. 290-297.
3. Barabási, Albert-László and Albert, Réka. Emergence of scaling in random networks // Science, 286:509-512, October 15, 1999.
4. Задорожный В.Н. Случайные графы с нелинейным правилом предпочтительного связывания // Проблемы управления. – 2010. – №6. – С. 2-11.