

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра социальной информатики

**ХАРАКТЕРИЗАЦИЯ СЛУЧАЙНЫХ ГРАФОВ НА ПРИМЕРЕ ГРУПП В
СОЦИАЛЬНЫХ СЕТЯХ**

(автореферат бакалаврской работы)

студента 5 курса 531 группы
направления 09.03.03 - Прикладная информатика
профиль Прикладная информатика в социологии
Социологического факультета
Киналиева Мербулата Константиновича

Научный руководитель

кандидат физико-математических наук, доцент _____ Л.Б. Тяпаев
подпись, дата

Зав. кафедрой

кандидат социологических наук, доцент _____ И.Г.Малинский
подпись, дата

Саратов 2020

ВВЕДЕНИЕ

Актуальность проблемы исследования. Термин «Социальная сеть» был введен задолго до появления Интернета и современных интернет-сетей. Еще в 1954 году социологом из «Манчестерской школы» Джеймсом Барнсом¹. Он развил изобретенный еще в 30-е годы подход к исследованию взаимосвязей между людьми с помощью социограмм- визуальных диаграмм, в которых отдельные люди представлены в виде точек, а связи между ними – в виде линий. На этом базируется социометрия – психологический метод для выяснения распределения в коллективе. К 70-м годам окончательно сформировался комплекс социологических и математических методов исследования, которые составляют научный фундамент современного анализа социальных сетей.

Визуально социальную сеть можно представить в виде социального графа, вершинами которого будут являться люди, а ребрами – связи между ними.

Социальные графы позволяют понять, как формируется, и развиваются социальные сети, прогнозировать распространение информации, выбор поведения людьми, поведение рынка и другое.

Не всегда есть возможность построить реальный граф, это связано с тем, что в большинстве случаев, часть данных скрыта от посторонних и их нельзя получить законным путем. К тому же реальный граф может достигать огромных размеров. Поэтому реальную картину иногда целесообразно заменить моделью.

Актуальность нашей работы заключается в том, что хоть по теории графов и написано большое количество работ, ее достаточно редко применяют в социологической науке для изучения интернет сообществ.

¹ История возникновения социальных сетей и профессиональных сообществ. URL: <https://sssvu.minobr63.ru/istoriya-vozniknoveniya-socialnyx-setej-i-professionalnyx-soobshhestv-y-seti/> (Дата обращения 2.04.2020) Загл. с экрана. Яз.рус.

Цель. Применить программные средства для построения графа пользователей социальной сети.

Задачи.

1. Рассмотреть модели случайных графов
2. Построить реальный граф пользователей социальной сети
3. Программно реализовать алгоритм построения графов
4. Рассмотреть и применить возможности ВКонтакте API и Graphi
5. Построить граф крупных новостных сообществ.

Выпускная квалификационная работа состоит из введения, 3 разделов, заключения, списка использованных источников и приложения.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В первом разделе «Теоретико-методологические обоснования теории графов» рассматривается появление понятия «Социальная сеть» в современном мире. Теория графов, сеть социальных взаимодействий. История развития данных понятий.

Современный человек реализует свои базовые потребности в коммуникации через социальные сети. Можно выделить две пересекающиеся трактовки социальной сети – как социальной структуры и ее специфической программной Интернет – реализацией.

Первыми компьютерными социальными сетями стали группы людей, использовавшие для создания и поддержания социальных связей средства компьютерного общения, которыми стала электронная почта. Случилось это 2 октября 1971 года – день первого сообщения, отправленного на удаленный компьютер.

Следующим шагом стало изобретение IRC (англ. Internet Relay Chat – ретранслируемый интернет-чат) – сервисной системы для общения в режиме реального времени. Чат был создан в 1988 году финским студентом Ярко

Ойкариненом. Это были уже более «продвинутые» социальные сети, однако еще далекие от современных².

В 1995 году появилась первая, приближенная к современным, социальная сеть «Classmates.com», которую создал Рэнди Конрад, владелец компании Classmates Online. Этот сайт помогал зарегистрированным посетителям находить и поддерживать отношения со знакомыми людьми. Сейчас в этой сети зарегистрировано более 40 миллионов людей, преимущественно из США и Канады. Концепция Classmates оказалась успешной, и с 2005 года она развивается уже не только в пределах этой сети, появились такие мировые гиганты как MySpace, Facebook, Bebo и LinkedIn или гиганты рунета Одноклассники.ру, Вконтакте, МойМир.

Что же такое «Социальная сеть»? Современное понятие в простом виде означает некий круг знакомых человека, где есть сам человек – центр социальной сети, его знакомые – ветки этой социальной сети и отношения между этими людьми – связи.

SNA (Social network analysis – анализ социальных сетей) – направление современной компьютерной социологии, которое занимается описанием и анализом возникающих в ходе социального взаимодействия и коммуникации связей (сетей) различной плотности и интенсивности³.

Сеть социальных взаимодействий – это сеть, состоящая из так называемых социальных акторов и наборов взаимосвязей между ними. Термин «Актор» используется в русскоязычной литературе и имеет более широкое значение, чем слово «лицо», то есть это может быть не только человек, но и совокупность организаций или целая страна. Под связями могут рассматриваться не только коммуникационные воздействия, но и связи по

² IRC. Национальная библиотека им. Н.Э. Баумана. URL: [https://ru.bmstu.wiki/IRC_\(Internet_Relay_Chat\)](https://ru.bmstu.wiki/IRC_(Internet_Relay_Chat)) (Дата обращения 3.04.2020) Загл. с экрана. Яз. рус.

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.

обмену различными ресурсами, взаимодействия, связанные с совместной деятельностью, включая конфликтные отношения.

В отличие от классических методов анализа, которые исследуют индивидуальные свойства объектов, основные цели анализа сетей – это исследование взаимодействия между социальными объектами (актерами) и выявление условий возникновения этого взаимодействия.

Отдельным направлением исследования является визуализация – графическое отображение социальных сетей (графов). Визуализация имеет большое значение, поскольку сама возможность увидеть сеть позволяет сделать важные выводы о характере взаимодействия акторов, не прибегая к другим методам анализа графа.

Термин «граф» впервые был введен в 1936 году венгерским математиком Денни Кёнигом.

В теории графов и информатике граф – это совокупность непустого множества вершин и множества пар вершин (связей между ними). Объекты представляются как вершины или узлы графа, а связи – как дуги или ребра⁴.

Граф, в котором ключевыми элементами являются люди и связи между ними, называется социальным графом. В абстракции социальный граф выглядит как набор узлов, которые обозначают людей, и линии между узлами которые обозначают существующие между людьми связи. Социальные графы идеально подходят для исследования крупных онлайн сообществ в социальных сетях. В современном понимании социальная сеть — это Интернет-сервис, позволяющий пользователям публиковать на своих страницах персональные и иные данные и служащий для упрощения коммуникации и обмена информацией между пользователями сети Интернет. Можно выделить следующие группы социальных сетей в зависимости от основного предназначения:

⁴ Пронькин Ю.С., Егоров Ю.А. Элементы теории графов и их техническое приложение. Тверь. 2007 г. С.4.

Можно выделить следующие группы социальных сетей в зависимости от основного предназначения:

1) Сети общего назначения (Facebook, ВКонтакте, Одноклассники) — предназначены для поддержания существующих контактов из реального мира, обсуждения повседневных событий и развлечений;

2) Нишевые сети (LinkedIn, Comon) — предназначены для поддержания существующих и установления новых профессиональных контактов, а также для дискуссий по общим интересам и решения профессиональных вопросов;

3) Контентные сети (Twitter, YouTube, Last.fm) — предназначены для обмена контентом, распространения новостей, создания и развития сообществ по интересам и развлечений;

4) Другие сети — геосоциальные сервисы (FourSquare), сервисы вопросов (StackOverflow, Quora), социальные Интернет-закладки (delicious), онлайн-игры (World of Warcraft) и т.д.

Кроме того, появление средств эффективного взаимодействия между пользователями во многом определило развитие других сервисов, таких как системы рекомендаций и онлайн-энциклопедии.

Помимо коммуникационной функции, сервисы социальных сетей играют роль баз пользовательских данных, в которых с каждым пользователем ассоциирован набор персональной информации, составляющий его "виртуальную личность". Данные всех пользователей одного сервиса образуют его социальный граф, — динамическую структуру, полностью описывающую состояние и поведение составляющих её пользователей, а также их отношения между собой и объектами внешнего мира в некоторый момент времени.

Во втором разделе «Модели случайных графов» рассматриваются примеры моделей случайных графов, предложенные различными авторами.

Модель Эрдеша-Реньи. На рубеже 50-х и 60-х годов XX века эту модель предложили классики современной комбинаторики и теории вероятностей П. Эрдеш и А. Реньи.

Ребра графа возникают в случайном графе независимо друг от друга с одной и той же вероятностью, а ребра, которых в графе нет, не возникают в случайном графе вовсе.

В начале 2000-х выяснилось, что модель Эрдеша-Реньи плохо описывает реальные графы, возникающие в различных областях. Это породило много новых исследований математических моделей случайных графов.

Модель Барабаши-Альберт представляет собой алгоритм генерации случайных графов с использованием принципа предпочтительного присоединения.

А. М. Райгородский в своей статье «Модели случайных графов и их применения», опубликованной в сборнике статей «Труды МФТИ» перечисляет основные моменты исследования Барабаши-Альберт, по существу этих моментов всего три⁵.

Во-первых, веб-граф (граф, вершинами которого являются сайты, а ребрами ссылки) – это весьма разряженный граф.

Во-вторых, диаметр (диаметром графа называется максимально возможное расстояние между двумя его вершинами) веб-графа исключительно скромен. В 1999 году он имел величину 5-7.

В-третьих, у веб-графа весьма характерное распределение степеней вершин. Эмпирическая вероятность того, что вершина веб-графа имеет степень d , оценивается как c/d^{λ} , где $\lambda = 2.1$, а c - нормирующий множитель, вычисляемый из условия «сумма вероятностей равна 1». Этот любопытный факт роднит Интернет с очень многими реальными сетями – биологическими, социальными, транспортными. Все они подчиняются степенному закону, только у каждой из них свой показатель λ .

⁵ Райгородский А.М. Модели случайных графов и их применение// Труды МФТИ. 2010. Том 2, №4. С.137.

Модель LCD. В статье «Случайные графы, модели и генераторы безмасштабных графов» М. Берновский и Н. Кузюрин рассматривают модель LCD⁶.

Модель линейной хордовой диаграммы (LCD) возникла в топологии и теории узлов, но ее структура оказалась напрямую связана с формированием веб-графа.

Модель Бакли-Остгуса. Обзор модели взят из статьи М. Берновского и Н. Кузюрина «Случайные графы, модели и генераторы безмасштабных графов»⁷.

Бакли и Остгус и другие исследователи предложили модификацию модели Барабаши-Альберт, в которой вершины обладают «изначальной привлекательностью»: вероятность того, что старая вершина будет выбрана соседом новой вершиной пропорциональна ее входящей степени плюс константе a , т.е. «изначальной привлекательности».

Модель копирования. Данный материал взят из статьи А. Райгородского «Модели случайных графов и их применения», опубликованной в сборнике статей «Труды МФТИ»⁸.

Модель копирования возникла одновременно с моделью Барабаши-Альберт. Она принадлежит Р. Кумару, П. Рагхавану, С. Раджагопалану, Д. Сивакумару, А. Томкинсу и Э. Упфалу.

В целом распределение степеней вершин в модели копирования очень похоже на распределение вершин в модели Боллобаша-Риордана. В остальном. Модели сильно разнятся. Например, в модели Боллобаша-Риордана практически отсутствуют плотные двудольные подграфы, в модели копирования таких подграфов много. Это особенно важно ввиду того, что

⁶ Берновски М.М., Кузюрин Н.Н. Случайные графы, модели и генераторы безмасштабных графов. Труды Института системного программирования РАН. 2012 С.425

⁷ Там же.

⁸ Райгородский А.М. Модели случайных графов и их применение// Труды МФТИ. 2010. Том 2, №4. С.136.

спамерские структуры, зачастую образуют именно двудольные графы с плотной перелинковкой.

В третьем разделе «Применение теории графов на примере социальной сети ВКонтакте» рассматривается применение теории случайных графов на примере сети «ВКонтакте». Для этого используется библиотека Python – networkx и программу Gephi.

«ВКонтакте» - крупнейшая в Рунете социальная сеть. Она представляет собой сетевой проект, который помогает людям высказываться и находить слушателей. Вы можете общаться с широким кругом интересных людей или поддерживаться связь с друзьями и близкими. Задача «ВКонтакте» - в каждый отдельно взятый момент может оставаться наиболее современным, быстрым и эстетичным способом общения в сети.

Для начала рассмотрим, что такое API (application programming interface). API – это посредник между разработчиком и приложений и какой-либо средой, с которой это приложение должно взаимодействовать. API упрощает создание кода, поскольку представляет набор готовых классов, функций или структур для работы с имеющимися данными. API ВКонтакте – это интерфейс, который позволяет получать информацию из баз данных vk.com с помощью http-запросов к специальному серверу.

В социальной сети «ВКонтакте» существует возможность скрывать личную информацию от пользователей не зарегистрированных или не авторизованных в этой социальной сети. По этой причине часть методов могут потребовать авторизации (передачи access токена). В таких случаях есть ограничение в виде лимитов на API. В документации к API ВКонтакте указано, что для клиентского приложения лимит – 3rps, а для серверного приложения прогрессивная шкала в зависимости от числа установок приложения. (rps/число установок): 5/<10000, 8/<100000, 20/<1000000, 35/>1000000.

Программа Gephi – это программа с открытым исходным кодом для анализа и визуализации графов, написанная на Java. Gephi имеет большую функциональность, которая расширяется с помощью большого количества

плагинов. Также с помощью данного инструментария можно реализовать задачу построить граф на основе заведомо известного списка групп из социальной сети ВКонтакте. В котором: Вершины – группы социальной сети; ребра – наличие общих подписчиков; чем больше у данной группы подписчиков, тем больше размер вершины; чем больше у групп общих подписчиков, тем ближе друг к другу располагаются вершины.

Представленные в данной главе результаты показывают простоту и доступность использования теории графов с социальными сетями. Использование описанных методов могут быть использованы для создания социального профиля пользователя социальной сети (возраст, интересы, круг общения), что в значительной степени может упростить работу социологов.

ЗАКЛЮЧЕНИЕ

Благодаря выводам, сделанным в рамках данной ВКР, мы проверили, подходят ли модели случайных графов для применения их на графах групп пользователей социальной сети, и можем сделать вывод, что использование графов является наиболее перспективным направлением в социологии.

Для достижения цели работы были построены две модели графов. Первый с помощью API Вконтакте, языка программирования Python и обработаны в специализированной программе Gephi на основе группы пользователей. Второй граф был построен на основе подписчиков популярных интернет пабликов в социальной сети ВКонтакте.