

МЕТОДЫ АНАЛИЗА КОМПЬЮТЕРНЫХ СОЦИАЛЬНЫХ СЕТЕЙ

Представлен обзор работ, посвященных проблеме анализа компьютерных социальных сетей. Существует четыре основных направления исследований в данной области: структурное, ресурсное, нормативное и динамическое. Для решения различных задач при анализе социальных сетей используются графовые и стохастические модели, модели эволюции сетей, методы с привлечением онтологий, структурные и реляционные модели, методы машинного обучения, методы визуализации графов и т. д. Приведено краткое описание наиболее популярных в настоящее время компьютерных социальных сетей и перечислены отдельные интересные программные приложения для их анализа. Намечены некоторые возможные пути дальнейших исследований, а именно: необходимость создания интегральной теории социальных сетей, более существенная адаптация методов обработки текстовой информации к сетевому контенту и др.

Ключевые слова: анализ социальных сетей, модель сети, граф сети, интеллектуальный анализ данных, центральность.

Введение

Социальные сети как общественное явление появились довольно давно. Анализ социальных сетей используется для исследования взаимодействий между участниками сети, прогнозирования их поведения, классификации, моделирования информационных потоков в сетях. В настоящее время с развитием компьютерных технологий у людей появилась возможность общаться виртуально при помощи компьютерных социальных сетей. Несомненно, анализировать информацию и поведение людей в процессе общения в таком формате значительно быстрее и удобнее. Поэтому анализ именно компьютерных социальных сетей вызывает большой интерес у современных исследователей.

В данной статье приведен обзор по анализу социальных сетей: перечислены основные направления исследований в этой области; описаны некоторые полезные при изучении социальных сетей характеристики; кратко изложены методы и алгоритмы, применяемые в различных моделях анализа сетей; приведено описание наиболее популярных в настоящий момент компьютерных социальных сетей; намечены некоторые возможные пути дальнейших исследований.

Основные направления исследования компьютерных социальных сетей

В настоящее время в анализе социальных сетей выделяют [1] четыре основных направления исследований: структурное, ресурсное, нормативное и динамическое.

В *структурном подходе* все участники сети рассматриваются как вершины графа, которые влияют на конфигурацию ребер и других участников сети. Основное внимание уделяется геометрической форме сети и интенсивности взаимодействий (весу ребер), поэтому исследуются такие характеристики, как взаимное расположение вершин, центральность, транзитивность взаимодействий. Для интерпретации результатов в данном направлении используются структурные теории и теории сетевого обмена.

Ресурсный подход рассматривает возможности участников по привлечению индивидуальных и сетевых ресурсов для достижения определенных целей и дифференцирует участников, находящихся в идентичных структурных позициях социальной сети, по их ресурсам. В качестве индивидуальных ресурсов могут выступать знания, престиж, богатство, раса, пол. Под сетевыми ресурсами понимаются влияние, статус, информация, капитал.

Нормативное направление изучает уровень доверия между участниками, а также нормы, правила и санкции, которые влияют на поведение участников в социальной сети и процессы их взаимодействий. В этом случае анализируются социальные роли, которые связаны с данным ребром сети, например, отношения руководителя и подчиненного, дружеские или родственные связи. Комбинация индивидуальных и сетевых ресурсов участника с нормами и правилами, действующими в данной социальной сети, образует его «сетевой капитал». В упрощенном виде «сетевой капитал» можно рассматривать как сумму некоторых преимуществ, которые участник может получить в произвольный момент времени для достижения некоторой цели.

Динамический подход – направление в изучении социальных сетей, в котором объектами исследований являются изменения в сетевой структуре с течением времени: по каким причинам исчезают и появляются ребра сети, как сеть изменяет свою структуру при внешних воздействиях, существуют ли какие-либо стационарные конфигурации социальной сети и др.

Рассмотрим немного подробнее перечисленные направления анализа социальных сетей в терминах решаемых задач [2].

Структурный анализ и анализ поведения связей в социальных сетях необходим для того, чтобы определить наиболее важные вершины, связи, сообщества и развивающиеся регионы сети. Такой анализ позволяет осуществлять обзор глобального эволюционного поведения сети. При структурном анализе и анализе поведения связей используются методы статистического анализа, методы определения сообществ, алгоритмы классификации.

Статистический анализ социальных сетей. В работе [3] приведен развернутый анализ структурных свойств сетей большого размера. Исследуются «типичные» социальные сети и изучается вопрос, как будет выглядеть сеть, если ее увеличить. Изучается взаимное поведение вершин сети исходя из предположения, что у большинства вершин имеется мало связей, возникают ли при этом «ядра» (скопления) или степени вершин распределяются более равномерно. Изучается поведение вершин при кластеризации. Другой из вопросов, изучаемых в этом разделе, – поведение типичных временных характеристик социальных сетей. Например, как меняется структура сети в процессе роста или как меняется поведение и распределение связанных компонентов графа. Со временем к сети добавляются новые сущности, но, несмотря на это, некоторые свойства графа могут сохраняться.

Определение сообществ в социальных сетях. Этот вопрос является наиболее важным в анализе социальных сетей, хотя довольно близок к задаче классификации. Цель – попытаться определить регионы сети, внутри которых происходит активное взаимодействие участников. Алгоритмически эту задачу можно отнести к задаче о разделении графов. Нужно разделить сеть на плотные регионы на основе поведения связей между вершинами. Компьютерные социальные сети динамические, что приводит к затруднениям с точки зрения выявления сообществ. В некоторых случаях удается интегрировать информационное содержимое сети в процесс определения сообществ. Тогда контент является вспомогательным средством для выявления групп участников с похожими интересами.

Анализ содержания социальных сетей. Можно выделить четыре вида анализа контента сети:

- анализ общей информации с произвольными типами данных;
- анализ текста;
- анализ мультимедиа;
- сенсорный и потоковый анализ.

Анализ медиаданных. Для обнаружения полезных бизнес-приложений можно анализировать социальные медиасети. Техники анализа данных предоставляют исследователям и специалистам инструменты для анализа больших, комбинированных, постоянно меняющихся медиаданных.

Анализ текстовой информации в социальных сетях. В вершинах социальной сети содержится много текстовой информации в различных формах, например, ссылки на посты (сообщения), блоги или статьи с новостями. Иногда пользователи могут отмечать друг друга, что тоже является формой текстовой информации в виде ссылок. Использование контента сети может сильно улучшить качество выводов при анализе социальных сетей, например, в задачах кластеризации и классификации.

Интеграция данных, поступающих с датчиков, и социальных сетей. Многие современные сотовые телефоны поддерживают возможность взаимодействия пользователей друг с другом динамически в режиме реального времени в зависимости от их местоположения и статуса. Подобные приложения также приводят к образованию потоков массивов в режиме реального времени. Их применяют для того, чтобы получить информацию о человеке или совокупности свойств объектов, которые отслеживаются. Поскольку информация о местоположении пользователя является личной, естественно возникает ряд проблем с точки зрения обработки, исследуются методы интеграции данных, поступающих с датчиков, и данных в социальных сетях.

Анализ мультимедийной информации сети. Существует много сайтов по обмену средствами массовой информации, например, Flickr и YouTube, которые обеспечивают возможность совместного использования этой информации. Такие средства массовой информации общего пользования часто сочетаются с взаимодействием пользователей – размещением тегов и комментариев в различных изображениях. Поэтому подобные сети могут служить источником для широкого спектра приложений в процессе извлечения и анализа данных.

Расстановка тегов. Большинство взаимодействий между пользователями происходит в форме тегирования (расстановки тегов, отметок), в которых пользователи прикрепляют описания различных объектов в социальной сети, такие как картинки, текст, видео или другая мультимедийная информация. В рамках данного подхода изучают свойства потоков тегов, моделей тегирования, семантику тегов, рекомендации по использованию тегов, визуализацию тегов, приложения для расстановки тегов, интеграцию различных систем тегирования и проблем, связанных с использованием тегов. Интересным вопросом является, например, почему люди расставляют теги, что влияет на выбор при тегировании, как промоделировать процесс подобной разметки, разновидности тегов, как создаются теги и как выбрать правильные теги для рекомендации.

Случайные блуждания и их применение в социальных сетях. Классификация – один из наиболее известных методов в веб-поиске. Например, можно упомянуть алгоритм ссылочно-ранжирования (PageRank) для приписывания веса веб-документам. Его основной принцип может применяться для поиска и классификации сущностей и участников в социальной сети. Этот алгоритм использует подход случайного блуждания для того, чтобы оценить вероятность посещения той или иной вершины. Естественно, что вершины, которые лучше расположены со структурной точки зрения, имеют более высокий вес, а значит, являются более важными. Методы случайного блуждания могут быть также полезны для объединения участников в группы относительно наиболее влиятельных участников.

Классификация вершин в социальных сетях. Некоторые вершины удобно снабжать пометками, чтобы их отличительные особенности и структурную информацию можно было распространить на всю сеть. Например, в маркетинговых исследованиях определенные вершины могут обозначать заинтересованность участников сети в конкретном продукте, и было бы желательно применить характерные особенности этих вершин для изучения других участников на предмет заинтересованности этим продуктом. Для этих целей, кроме того, можно использовать информацию о контенте и структуре социальной сети. Другой пример, когда об одной из двух связанных вершин получены некоторые сведения, для второй эти сведения с большой долей вероятности тоже будут верны. Вот почему структуру связей можно применять для распространения меток среди вершин. Содержимое сети и структурные особенности в дальнейшем могут пригодиться для подтверждения качества полученной классификации.

Анализ социального влияния. Так как в основе социальных сетей лежит взаимодействие между различными участниками, естественно предположить, что это взаимодействие оказывает влияние на участников в терминах их поведения. Классическим примером является

применение «вирусного маркетинга» для распространения сообщения между взаимосвязанными участниками через всю сеть. Вопросы этого направления:

- как моделировать влияние на основе информации об участниках;
- как моделировать распространение влияния;
- кто из участников является наиболее влиятельным в процессе распространения?

Конфиденциальность в социальных сетях. В социальных сетях содержится большое количество личной информации об участниках, например, интересы, информация о дружбе, демографическая информация и др. Это может привести к несанкционированному распространению личной информации в сетях. В решении такого типа задач полезно применять модели на основе механизмов конфиденциальности.

Обнаружение экспертов в сетях. Социальная сеть может являться инструментом для выявления экспертов в конкретной области. Часто в реальности эксперты образуют сеть, которая соответствует социальной сети или организационной структуре компании. Многие сложные задачи требуют коллективного решения нескольких экспертов. В подобных случаях получается, что более эффективно можно достигнуть общую цель, когда специалисты сотрудничают друг с другом. Важной задачей данного направления является обнаружение групп специалистов в определенных узких областях.

Эволюция в динамических социальных сетях. С течением времени в социальных сетях появляются новые участники, некоторые участники прекращают взаимодействие, возникают новые связи, некоторые связи устаревают, так как участники перестают взаимодействовать. Это приводит к изменениям в структуре социальных сетей в целом и в отдельных сообществах. При этом возникает два важных вопроса: 1) согласно каким законам происходят долгосрочные изменения между крупными сообществами в социальных сетях; 2) как развиваются сами сообщества во времени. Какие изменения могут происходить, как можно отследить и представить их?

Прогноз формирования связей в социальных сетях. Для извлечения интересующей информации из социальной сети полезны исследования, направленные на определение и предсказание возможных связей между вершинами в будущем.

В большинстве приложений для анализа социальных сетей связи считаются динамическими и могут сильно изменяться с течением времени. Например, отношение «дружбы» не меняется. В процесс прогнозирования связей может быть вовлечена как структура сети, так и информация об особенностях различных вершин. Для решения таких задач предлагается строить разнообразные структурные и реляционные модели.

Визуализация социальных сетей. Социальные сети становятся крупнее и имеют все более сложную структуру. Визуализация помогает естественным образом свести воедино информацию о сетях и сделать ее более доступной для понимания. Визуализация в сочетании с взаимодействием помогают аналитикам в описании социальных сетей. Целью этого направления является также поиск ответа на вопрос, как различные модели могут быть использованы для изучения различных аспектов сетей, таких как структура и семантика. Важным является создание алгоритмов, сочетающих методы анализа и визуализации, чтобы улучшить понимание структуры и динамики сети.

Как видно, при анализе социальных сетей решается довольно большой круг задач и применяются методы из различных областей знаний.

Некоторые наиболее известные социальные сети

К крупнейшим социальным сетям по числу пользователей относятся Facebook, LinkedIn, ВКонтакте, Twitter, Одноклассники.ru, YouTube и др.¹

Facebook (<http://www.facebook.com>). Сеть основана в 2004 г. Марком Цукербергом. По данным за апрель 2012 г.², аудитория Facebook составляет 901 млн пользователей. Каждый

¹ Социальная сеть. URL: http://ru.wikipedia.org/wiki/Социальная_сеть; List of social networking websites. URL: http://en.wikipedia.org/wiki/List_of_social_networking_websites

день в сети пользователи оставляют 3,2 млрд «лайков» и комментариев и публикуют 300 млн фотографий. Facebook позволяет создать профиль с фотографией и информацией о себе, приглашать друзей, обмениваться с ними сообщениями, изменять свой статус, оставлять сообщения на своей и чужой «стенах», загружать фотографии и видеозаписи, создавать группы (сообщества по интересам). Существует возможность создавать приложения для Facebook (игры, средства обмена музыкой, фотографиями и т. д.), что повышает посещаемость сайта.

YouTube (<http://www.youtube.com>). Сервис, предоставляющий услуги видеохостинга, основан в 2005 г. Пользователи могут добавлять, просматривать и комментировать те или иные видеозаписи, добавлять аннотации и титры к видео, а также выставлять рейтинг просмотренным видео, если такую возможность им предоставил автор. Благодаря простоте и удобству использования YouTube стал популярнейшим видеохостингом и третьим сайтом в мире по количеству посетителей на июнь 2012 г. Ежеминутно на YouTube загружают 60 часов видео³. В январе 2012 г. ежедневное количество просмотров видео на сайте достигло 4 млрд.

LinkedIn (<http://www.linkedin.com>). Социальная сеть была основана Ридом Хоффманом в декабре 2002 г., запущена в мае 2003 г. В основном сеть используется для поиска и установления деловых контактов. По данным на февраль 2012 г.⁴, в LinkedIn зарегистрировано свыше 160 млн пользователей. Чуть меньше половины пользователей LinkedIn являются жителями США.

ВКонтакте (<http://vk.com>). Сеть основана в 2006 г. Павлом Дуровым. По данным на март 2012 г.⁵, аудитория ВКонтакте составила около 150 млн человек, около 70 % из них проживают в России. Подобно Facebook, пользователи ВКонтакте могут обмениваться сообщениями приватно (через личные сообщения) и публично (с помощью записей на «стене», а также через механизм групп и встреч), отслеживать через ленту новостей активность друзей и сообществ. В сети есть возможность обмена и загрузки файлов довольно большого объема, так как используется технология распределенного распространения файлов BitTorrent, что делает ВКонтакте одним из крупнейших медиаархивов Рунета. Facebook, Одноклассники.ru и другие социальные сети используют протокол обмена сообщениями XMPP (Extensible Messaging and Presence Protocol), ранее известный как Jabber.

Twitter (<https://twitter.com>). Создана в 2006 г. Джеком Дорси. По состоянию на начало 2012 г. сервис насчитывает более 140 млн пользователей⁶. Ежедневно пользователи отправляют около 340 млн сообщений. Система позволяет отправлять короткие текстовые сообщения (до 140 символов), используя веб-интерфейс, SMS, средства мгновенного обмена сообщениями или сторонние программы-клиенты. Отличительной особенностью Twitter является публичная доступность размещенных сообщений, что позволяет называть его микроблогом. С 2011 г. Twitter перешел с MySQL на Lucene и с Ruby on Rails на Java и Scala для повышения производительности и масштабируемости.

Одноклассники.ru (<http://www.odnoklassniki.ru>). Проект запущен в 2006 г., его автором является российский веб-разработчик Альберт Попков. По состоянию на июнь 2011 г.⁷, зарегистрировано более 70 млн пользователей. Особенность этой сети в том, что каждый пользователь видит имена всех, кто заходил посмотреть на его анкету, все публичные действия пользователей (сообщения в форумах, добавление друзей, загрузка фотографий) отображаются в доступной другим пользователям «ленте активности». Является русскоязычным аналогом американской сети Classmates.com.

Flickr (www.flickr.com). Создана в 2004 г. По данным на июнь 2011 г.⁸, в сети зарегистрировано 51 млн пользователей. Сервис предназначен для хранения и дальнейшего использования пользователем цифровых фотографий и видеороликов. Является одним из первых Web

² U.S. Securities and Exchange Commission. URL: <http://www.sec.gov/Archives/edgar/data/1326801/000119312512175673/d287954ds1a.htm>

³ Alexa – The Web Information Company. URL: <http://www.alexa.com/siteinfo/youtube.com>

⁴ LinkedIn – Press Center. URL: <http://press.linkedin.com/about>

⁵ ВКонтакте. URL: <http://vk.com/catalog.php>

⁶ Twitter. URL: <http://blog.twitter.com/2012/03/twitter-turns-six.html>

⁷ Благовещенский А. Количество пользователей «Одноклассников» превысило 70 миллионов // Российская газета. URL: <http://www.rg.ru/2011/06/21/odnoklassniki-site-anons.html>

⁸ Yahoo! Advertising solutions. URL: <http://advertising.yahoo.com/article/flickr.html>

2.0 сервисов. Есть возможность к каждой фотографии добавить название, краткое описание и ключевые слова (тег) для дальнейшего поиска.

Модели анализа социальных сетей

Один из самых известных примеров анализа сетей был проведен в 1970-е гг. американским социологом Марком Грановеттером [4]. Он показал, что для многих социальных задач, таких как поиск работы, слабые связи оказываются намного эффективнее, чем сильные. Этот эффект он назвал «силой слабых связей».

Для эффективности и устойчивости социальной сети являются важными функциональные роли ее участников, такие как информационные брокеры, эксперты и др. Информационные брокеры – люди, которые играют роль посредника в социальной сети, связывая между собой группы людей, налаживая связи между специалистами и тем самым открывая им доступ к информации. Для идентификации экспертов в социальных сетях используется, например, оптимизационный подход муравьиной колонии (ACO = Ant Colony Optimization). Это эффективный полиномиальный алгоритм для нахождения приближенных решений оптимального пути в графе.

Оригинальная идея алгоритма исходит из наблюдения за муравьями в процессе поиска кратчайшего пути от колонии до источника питания. Биологи заметили, что среди экспериментов по выбору между двумя путями неравной длины, ведущих от колонии к источнику питания, муравьи, как правило, используют кратчайший маршрут. Модель такого поведения заключается в следующем. Муравей проходит случайным образом от колонии. Если он находит источник пищи, то возвращается в гнездо, оставляя за собой след из феромонов. Эти феромоны привлекают других муравьев находящихся вблизи, которые, вероятнее всего, пойдут по этому маршруту. Вернувшись в гнездо, они укрепят феромонную тропу. Если существует два маршрута, то по более короткому за то же время успеют пройти больше муравьев, чем по длинному. Поэтому короткий маршрут станет более привлекательным, а длинные пути, в конечном итоге, исчезнут из-за испарения феромонов.

Работа начинается с размещения муравьев в вершинах графа, затем начинается движение муравьев. Направление определяется вероятностным методом на основании формулы вида

$$P_i = \frac{l_i^q f_i^p}{\sum_{k=0}^N l_k^q f_k^p},$$

где

P_i – вероятность перехода по пути i ;

l_i – величина, обратная весу (длине) i -го перехода;

f_i – количество феромонов на i -м переходе;

q – величина, определяющая «жадность» алгоритма;

p – величина, определяющая «стадность» алгоритма и $q + p = 1$.

Решение не является точным и даже может быть одним из худших, однако в силу вероятности решения повторение алгоритма может выдавать (достаточно) точный результат.

Еще один широко известный пример анализа социальных сетей – эксперимент американского психолога Милгрэма [3], проведенный в 1969 г. Этот эксперимент получил название «феномен малого мира» (Milgram's small world experiment), или «теория шести рукопожатий» (Six degrees of separation). Гипотеза заключается в том, что каждый человек знаком с любым другим жителем планеты через цепочку общих знакомых, в среднем состоящую из шести человек. Пока что это утверждение не было опровергнуто. Наоборот, в качестве доказательства правильности гипотезы выдвигается наблюдение, что диаметр большинства сетей относительно небольшой.

Среди методов анализа социальных сетей основными [1] являются: методы теории графов, в частности изучение ориентированных графов и представляющих их матриц, применяемых для изучения структурных взаимосвязей участника сети; методы нахождения локальных

свойств участников, например, центральности, влияния, положения, принадлежности к некоторым подгруппам; методы определения эквивалентности участников, включая их структурную эквивалентность; блоковые модели и ролевые алгебры; анализ диад и триад; вероятностные модели, включая модели марковских процессов.

Графовые модели. Любую социальную сеть можно математически представить в виде графа

$$G = (V, E),$$

где V – множество вершин графа; E – множество ребер графа; $|V| = N$ – количество вершин в графе.

В графе социальной сети вершинами являются участники, а ребра означают наличие отношений между ними. Отношения могут быть как направленными, так и ненаправленными. Как правило, выделяют два типа отношений: «дружба» (люди знакомы друг с другом) и «интересы» (есть общие интересы, люди входят в одну группу по интересам). Эти отношения используются, например, в FOAF (Friend of a friend) – онтологии описания людей, их активности и отношений к другим людям и объектам. В FOAF описание социальных связей между людьми основывается на транзитивности доверия. Описание алгоритма вычисления уровня доверия (TrustRank) приведено ниже.

Можно выделить три типа графовых моделей [1].

Стохастические блоковые модели задаются матрицей A размера $N \times N$, где N – число групп (блоков) участников. Элемент $a_{ij} \in [0, 1]$ показывает плотность связей между участниками сети, принадлежащими к группе v_i , и участниками, принадлежащими к группе v_j . При этом граф не содержит дополнительных ребер и вершин, соответствующих связям участников внутри одной группы.

Вероятностные графовые модели задаются матрицей A размера $N \times N$, где N – число участников сети. Элемент $a_{ij} \in [0, 1]$ показывает вероятность взаимодействия участника v_i и участника v_j в течение определенного периода времени.

Обычные графовые модели задаются матрицей связности A размера $N \times N$.

Для анализа графовых моделей социальных сетей иногда удобно использовать коэффициент плотности, определенный как отношение числа ребер в анализируемом графе к числу ребер в полном графе с тем же числом вершин (полный граф – это граф, в котором все вершины соединены между собой). Кроме этого, сеть могут характеризовать такие величины, как число путей заданной длины (путь – последовательность вершин, связанных между собой), минимальное число ребер, удаление которых разбивает граф на несколько частей.

Графовые модели социальных сетей используются для моделирования экономических и коммуникационных связей людей, анализа процессов распространения информации, нахождения сообществ и связанных подгрупп, на которые можно разбить всю социальную сеть.

Анализ центральности и других локальных свойств. Чтобы определить относительную важность (вес) вершин графа (т. е. насколько участник в рамках конкретной сети является влиятельным), вводят понятие центральности – меры близости к центру графа. Центральность можно определить разными способами, поэтому существуют различные меры центральности⁹. Следует отметить, что речь идет не о геометрической центральности при визуализации графа отношений.

Центральность по степени (Degree centrality) определяется как количество связей, инцидентных вершине:

$$C_D(v) = \deg(v).$$

Выделяют входящие и исходящие связи. Входящие связи характеризуют популярность человека, выходящие – его общительность. Полученную величину можно нормировать, разделив на общее число участников в сети.

⁹ *Hanneman R., Riddle M.* Introduction to social network methods. Riverside, CA: University of California, Riverside, 2005. URL: <http://faculty.ucr.edu/~hanneman/nettext>; Centrality. URL: <http://en.wikipedia.org/wiki/Centrality>

Другими словами, центральность по степени предполагает, что среди участников сети более влиятельным является тот, у кого больше друзей, либо тот, кто входит в большее количество сообществ. Тем не менее участник сети, имеющий большое количество друзей, может быть связан с остальным графом маленьким количеством ребер. Поэтому вводится следующее понятие.

Центральность по близости (Closeness centrality) является показателем, насколько быстро распространяется информация в сети от одного участника к остальным. В качестве меры расстояния между двумя участниками используется кратчайший путь по графу (геодезическое расстояние). Так, непосредственные друзья участника находятся на расстоянии 1, друзья друзей – на расстоянии 2, друзья друзей друзей – на расстоянии 3 и т. д. Далее берется сумма всех расстояний и нормируется. Полученная величина называется удаленностью вершины v от других вершин. Близость определяется как величина, обратная удаленности:

$$C_c(v) = \frac{N-1}{\sum_{t \in V \setminus v} d_G(v, t)},$$

где $d_G(v, t)$ – кратчайший путь от вершины v до вершины t .

Другими словами, центральность по близости позволяет понять, насколько близок рассматриваемый участник ко всем остальным участникам сети. Таким образом, важно не только наличие непосредственных друзей, но и чтобы у самих этих друзей тоже были друзья.

Центральность по посредничеству (Betweenness centrality). Еще одной характеристикой участника является его важность при распространении информации. Именно в этом контексте центральность по посредничеству оценивает участника. Она рассчитывается как число кратчайших путей между всеми парами участников, проходящих через рассматриваемого участника:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}},$$

где

σ_{st} – общее количество кратчайших путей из вершины s к вершине t ;

$\sigma_{st}(v)$ – количество кратчайших путей из вершины s к вершине t , проходящих через вершину v .

Для нормализации нужно разделить на количество пар вершин, за исключением самой вершины v , т. е. для ориентированного графа нужно разделить на $(N-1)(N-2)$, для неориентированного – на величину, равную $(N-1)(N-2)/2$. Недостатком центральности по посредничеству является ее вычислительная сложность.

Центральность по собственному вектору (Eigenvector centrality). Пусть центральность рассматриваемого участника – x_v , а центральность его непосредственных друзей (соседних вершин) x_j, x_k, x_l и т. д. Центральность по собственному вектору определяется как сумма центральностей соседних вершин, поделенных на константу λ , т. е. $x_v = (x_j + x_k + x_l) / \lambda$. Выписав аналогичные уравнения для всех друзей, получим вектор неизвестных $X = (x_1, \dots, x_v, \dots, x_n)$. Правила сложения определяются матрицей смежности $A = (a_{vt})$, т. е. $a_{vt} = 1$, если вершина v соединена с вершиной t , $a_{vt} = 0$ – иначе. Далее требуется решить уравнение $AX = \lambda X$, т. е. найти собственные значения и собственные векторы матрицы A . Полученную задачу можно переписать иначе:

$$C_E(v) = x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{vt} x_t,$$

где $M(v)$ – множество вершин, соседних с вершиной v ; λ – константа.

Собственный вектор, соответствующий самому большому собственному значению, как раз образован центральностями соответствующих участников сети.

Таким образом, чем больше у участника друзей и чем они центральнее, тем больше его центральность. Верно и обратное: чем больше центральность участника, тем больше центральность его друзей. Недостатком центральности по собственному вектору также является вычислительная сложность.

Обобщением центральности по степени является *центральность Каца (Katz centrality)*. Отличие в том, что центральность по степени учитывает количество непосредственных соседей вершины, а центральность Каца учитывает количество всех вершин, которые могут быть соединены путем

$$C_{Katz}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^N \alpha^k (a^k)_{ji},$$

где $\alpha \in (0, 1)$ – доля участия удаленных вершин, называемая *коэффициентом затухания*.

Центральность Каца можно представить как разновидность центральности по собственному вектору:

$$C_{Katz}(i) = x_i = \alpha \sum_{j=1}^N a_{ij} (x_j + 1).$$

Центральность можно вычислить при помощи *алгоритма ссылочного ранжирования (PageRank)*, который используется в поисковой системе Google. В основу положен принцип «важности» веб-страницы: чем больше ссылок на страницу, тем она «важнее». Кроме того, вес самой страницы определяется весом ссылки передаваемой на нее страницы. Таким образом, PageRank – это метод вычисления веса страницы путем подсчета важности ссылок на нее, т. е. вершина, ссылающаяся на другую вершину с большим весом, сама получает большой вес:

$$C_{PageRank}(i) = x_i = \alpha \sum_{j=1}^N a_{ji} \frac{x_j}{L(j)} + \frac{1-\alpha}{N},$$

где $L(j) = \sum_j a_{ji}$ – количество вершин, соседних с вершиной j (или количество выходящих связей в ориентированном графе).

Отличием данного алгоритма от вычисления центральности по собственному вектору и центральности Каца является наличие коэффициента пересчета $L(j)$. Следует заметить также, что в алгоритме ссылочного ранжирования используется обратная индексация матрицы смежности a_{ji} в сравнении с вычислением центральности по собственному вектору. Предшественником алгоритма PageRank является алгоритм *HITS (Hyperlink-Induced Topic Search)*, предложенный Кляйнбергом [5].

Помимо перечисленных методов определения центральности, существует большое количество введенных неклассическим образом способов вычисления этой характеристики сети.

Важными характеристиками связей сети являются сбалансированность и транзитивность. Сбалансированность – это отсутствие ситуаций типа «позитивное взаимодействие (дружба, партнерство) между v_1 и v_2 , а также между v_1 и v_3 , но негативное взаимодействие (вражда, соперничество) между v_2 и v_3 ». Утверждается, что сбалансированные сети психологически более комфортабельны для участников и более устойчивы по сравнению с несбалансированными [6]. Транзитивность – это выполнение условий вида «если есть взаимодействие между v_1 и v_2 , а также между v_2 и v_3 , то имеет место взаимодействие между v_1 и v_3 ». Данные характеристики описывают локальные связи участников и часто используются при анализе диад и триад.

Полезной характеристикой при анализе социальных сетей является уровень доверия. *Алгоритм вычисления уровня доверия (TrustRank)* предложен в [7]. Изначально был создан для отделения информативных веб-страниц от спама. Если говорить об этом алгоритме в терминах сайтов, для контрольной выборки эксперты вручную оценивают степень доверия небольшого количества сайтов, которые можно считать надежными. Эти сайты принимаются за

эталон. Далее в основу алгоритма положено утверждение, что хорошие сайты редко ссылаются на плохие, а вот плохие очень часто ссылаются на хорошие. TrustRank – величина, которая дает оценку того, можно ли доверять конкретному сайту, считая, что он не содержит спама. Чем больше ссылок на сайте, тем меньше доверия «передается» по каждой такой ссылке. Степень доверия сайту (TrustRank) убывает с увеличением расстояния между ним и первоначальной выборкой.

Сила структурной позиции участника является основным показателем, определяющим различия в ресурсах участников сети. В теории сетевого обмена для измерения данной характеристики вводится [8] индекс GPI силы участника v_i (Genuine Progress Indicator):

$$GPI_i = \sum_{k=1}^{g-1} (-1)^{k-1} P[i]_k,$$

где $P[i]_k$ – число непересекающихся путей длины k , проходящих через вершину v_i . Сила участника v_i по сравнению с силой участника v_j вычисляется как $GPI_{ij} = GPI_i - GPI_j$.

Методы обнаружения сообществ и анализ связанных подгрупп. Связанные подгруппы (сообщества) в сети характеризуются наличием большого числа связей между входящими в них участниками и существенно меньшим числом связей с остальными участниками. Анализ сообществ позволяет изучать устойчивость социальных структур. Простейший случай связанной группы – это сообщество, где каждый участник связан с каждым, и в данную группу не могут быть включены другие участники сети, поскольку они не имеют связей со всеми членами сообщества (клики). Таким образом, клика – это максимальный полный подграф данного графа. Если анализировать процессы распространения информации в графах, то можно дать другое определение сообщества, как множества участников, где путь между двумя любыми участниками не содержит более одной промежуточной вершины. В результате информация от одного участника к другому в связанной группе передается с минимальными искажениями. Связанные группы также могут быть выделены с помощью многомерного шкалирования или факторного анализа матрицы связей графа [1].

Для анализа устойчивости групповой структуры во времени используется следующая техника. Вначале строится трехмерная матрица, в которой строки представляют оценки взаимодействий участника со всеми другими участниками, данные самими участниками; столбцы являются собственными оценками взаимодействий участника; на третьей оси расположены периоды времени. Далее может быть построен график, показывающий изменения структуры подгрупп с течением времени.

После этого применяются *методы уменьшения размерности данных* (например, метод главных компонент), т. е. рассматривается проекция вершин сети в евклидово пространство пониженной размерности для описания зависимостей между строками и столбцами данной матрицы. В результате можно визуализировать изменения статуса пользователя сети на фоне изменений статусов подгрупп [6].

Полученную проекцию можно кластеризовать при помощи стандартных *алгоритмов кластеризации* как статистических (например, k -средних) [9], так и иерархических. Преимуществом иерархических методов является возможность представления результата кластеризации в виде дендрограммы, т. е. на выходе будет получено не просто разбиение графа на группы, а иерархия групп и подгрупп в графе. Основная сложность подобных методов – подобрать подходящую меру расстояния (кратчайшего пути между вершинами) или меру сходства (similarity). Наиболее часто применяются меры сходства, использующие косинусный коэффициент (cosine similarity, также известен как коэффициент Охай) и коэффициент Жаккара (Jaccard coefficient). Кластеризацию можно проводить не снизу вверх, а *сверху вниз*, т. е. сначала вся сеть рассматривается как одна группа, а на каждой итерации происходит последовательное отделение по одной связи.

Детальный обзор методов обнаружения сообществ можно найти, например, в [10].

Структурная эквивалентность участников сети. Этот подход является противоположностью исследованию связанных групп. Участники эквивалентны, когда они занимают одинаковые позиции в социальной структуре сети, т. е. когда эквивалентны структура и тип

взаимодействий этих участников с другими, при этом эквивалентные участники сети не должны взаимодействовать друг с другом.

В качестве меры эквивалентности может выступать плотность связей со структурными подгруппами участников сети [11]. Наряду со структурной эквивалентностью используется регулярная эквивалентность участников. В этом случае участники эквивалентны, когда они одинаковым образом взаимодействуют с участниками одного типа.

Для определения структурной эквивалентности двух участников необходимо сравнить структуру их взаимодействий с другими участниками, т. е. нужно сравнить соответствующие столбцы в матрице связей графа. Это может быть осуществлено с помощью вычисления расстояния между этими векторами (например, по метрике Евклида или Чебышева) или коэффициентов связи (например, корреляции Пирсона). Для направленных графов необходимо учитывать входящие и выходящие ребра, с этой целью одновременно рассматриваются две соответствующие матрицы.

На следующем этапе в матрицах для каждого типа связей переставляются столбцы таким образом, чтобы сгруппировать те из них, которые соответствуют структурно эквивалентным участникам. В результате матрица разбивается на структурные блоки, в каждом из которых вычисляется плотность. Далее строится новая матрица связей между найденными структурными блоками, например, по следующему правилу: если плотность связей между двумя блоками выше, чем средняя плотность связей в первоначальной матрице, то соответствующий элемент новой матрицы равен 1, в противном случае он равен 0. Такие матрицы называются блоковыми моделями и являются средством построения ролевых алгебр [1].

Ролевые алгебры. Это направление анализа социальных сетей фокусируется на выявлении логики взаимодействий участников сети в блоковых моделях, что позволяет выявлять сходства принципов взаимоотношений участников в различных социальных сетях. Определим, например, матрицы симпатии и антипатии следующим образом:

$$LIKE = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad DISLIKE = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Теперь можем анализировать комбинации взаимодействий участников сети, перемножая соответствующие матрицы.

Анализ диад и триад. Диады – это набор из двух участников сети (вершин) и всех взаимодействий (ребер) между ними. Диада для каждого типа взаимодействий может находиться в одном из четырех состояний: нет связи между участниками, связь направлена от первого участника ко второму, связь направлена от второго участника к первому, взаимные связи участников. Анализ диад помогает установить вероятность наличия ребра между ними, степень зависимости от свойств участников, определить условия и направления передачи информации и т. д. Для триад (три взаимодействующих участника) дополнительно исследуются вопросы транзитивности взаимодействий.

Важным показателем является сила связей между участниками, которая определяется как линейная комбинация продолжительности, эмоциональной насыщенности, интимности или конфиденциальности и значимости взаимных услуг, которые характеризуют данное взаимодействие и соответствующее ему ребро графа. Слабые связи являются важными источниками информации [4], так как они помогают получить дополнительные сведения об участнике или сообществе, в котором он состоит, из других источников.

Стохастические модели. Основная идея вероятностных моделей ориентированных графов состоит в том, что каждая социальная сеть может быть рассмотрена как реализация случайного двумерного бинарного массива. Так как элементы этого массива являются зависимыми случайными величинами, то можно анализировать структуру зависимостей между соответствующими участниками социальной сети, находить вероятности существования определенных связей и получать оценки различных параметров сети.

Применение статистических моделей в анализе социальных сетей приведено, например, в [12]. Предлагается также применять методы машинного обучения и анализа данных для вычисления относительной автокорреляции, плотности связей и некоторых других характеристик сети. Более подробно про стохастические модели можно посмотреть в [1].

Модели динамики сети. Для определения динамики сети используются [13; 14] следующие модели.

Модели эволюции графа. Согласно этой модели, когда новая вершина добавляется к сети, происходит выбор вершин, к которым можно осуществить присоединение при помощи присоединяющего правила предпочтений. Также выбор вершины может осуществляться случайным образом или «копированием» некоторых ее внешних ссылок.

В [15] было эмпирически обнаружено, что сети со временем становятся плотнее (плотность сети увеличивается), т. е. количество ребер увеличивается линейно с ростом количества вершин. Более того, плотность сети меняется по степенному закону. В этой же работе описано еще одно наблюдение: диаметр сети часто уменьшается с течением времени, что противоречит общепринятому мнению о том, что меры расстояний должны медленно увеличиваться в зависимости от количества вершин.

Модель «закрытого треугольника» (Triangle-Closing Model) [15] утверждает, что новые вершины, добавленные к сети, имеют тенденцию к закрытию треугольника. Если считать, что связи, возникающие между участниками, образуют треугольник, то «открытый» треугольник возникает, когда два участника могут быть связаны друг с другом только посредством третьего, т. е. одна из трех связей пропущена. Когда добавляется третья связь, получается «закрытый» треугольник.

Модель «лесных пожаров» [16] является в некотором смысле обобщением модели закрытого треугольника. Новая вершина присоединяется к существующей путем выбора подграфа, содержащего эту вершину и связывающего ее со всеми вершинами этого подграфа. Процесс начинается в выбранной вершине и напоминает распространение пожара через все вершины сети.

Несмотря на то, что существует довольно много работ, посвященных анализу глобальных свойств эволюции социальных сетей, вопросу эволюции графов на микроскопическом уровне посвящено совсем небольшое количество исследований. К этому направлению можно отнести, например, работы [14; 16], изучающие различные стратегии формирования сети и показывающие, что расположение ребер играет важную роль в эволюции сетей.

Среди работ, представляющих алгоритмические инструменты для анализа эволюции сетей, можно выделить [17], в которой предложены алгоритмы оценки принадлежности пользователя сообществу и ее изменения с течением времени. Алгоритмы базируются на динамическом программировании, полном переборе, максимальном соответствии и «жадных» эвристиках. Основное внимание уделяется определению приблизительных кластеров пользователей и их временным изменениям. В [18] предложено применить принцип минимума длины описания MDL (Minimum Description Length) для поиска закономерностей в данных и обнаружения сообществ в динамических сетях, который создает структуру, «свободную от параметров». В [19] предлагается использовать принцип MDL для мониторинга эволюции сети.

Анализ графов развития сети. В работе [20] представлены различные подходы к анализу эволюции сети, основанные на парадигме извлечения ассоциативного правила (association-rule mining) и анализа частотной модели (frequent-pattern mining). Вводятся правила эволюции графа, новый тип частотных моделей, и рассматривается проблема поиска типичных моделей структурных изменений в динамических сетях. Сначала вычисляется набор частотных моделей графа, который описывает характерные эволюционные механизмы, а затем находят правила эволюции графа, которые удовлетворяют заданному ограничению минимальности доверия.

Проблема получения временно развивающихся веб-графов рассмотрена в [21]. Определены три уровня анализа графов: графы с единственной вершиной, подграфы и анализ графа в целом – для каждого из них используются свои техники. Изучены изменения свойств на каждом из трех уровней анализа. Для представления изменений в подграфах динамического графа в [22] предложен быстрый метод извлечения часто встречающихся «подпоследовательностей» из графа. Однако в описании модели не оговаривается время, в течение которого наблюдались изменения графа во времени. Другой способ определения подграфов, меняющихся со временем, описан в [23]. Он учитывает оценку важности вершин (vertex importance

scores) и изменения близости вершин (vertex closeness changes). Релевантным подграфом считается не наиболее частый, а наиболее значимый.

Историю ребра в динамическом графе предложено в [24] представить в виде последовательности нулей и единиц, соответствующих наличию или отсутствию того или иного ребра. Затем для получения частотных моделей графа применяются традиционные методы получения графа. Разработанный алгоритм GREW использует эвристики и, вообще говоря, не извлекает все частые модели.

Прогнозирование формирования связей. Модели эволюции графа обычно создаются для оценки общестатистических свойств существующих графов. Можно пытаться также вычислить, будут ли две конкретные вершины соединены друг с другом через некоторый промежуток времени. Это вычислительная задача, в основе которой лежит анализ эволюции социальной сети во времени, и называется проблемой прогнозирования связей.

Пусть дана краткая характеристика социальной сети в момент времени t и задано будущее время t_0 . Задача состоит в том, чтобы предсказать новые связи, которые, скорее всего, появятся в сети за промежуток времени $[t_0, t_1]$. Для ее решения в [25] применяется автоматическое моделирование процесса развития социальной сети с привлечением некоторых характеристик сети, таких как количество общих соседей, геодезическое расстояние (кратчайший путь), влияние вершины, момент первого попадания в социальную сеть.

Есть модели прогнозирования возникновения связей¹⁰, основанные на машинном обучении, использующие личную информацию о пользователях сети для повышения точности предсказания. Иногда применяют иерархические, вероятностные (марковские) и реляционные модели для обнаружения связей между пользователями.

В других моделях [26] за основу предлагается брать сами свойства пользователей, и, например, наличие большого количества связей (в блогосфере) может быть объяснено путем сопоставления демографических групп, общих интересов или географической близостью.

Методы на основе онтологий. Исследования [27] показали, что оценить параметры социальных сетей (диаметр, количество участников, среднюю длину пути и др.) можно при помощи онтологий. Сначала производится анализ видов элементов сети: люди, объекты (музыка, фото, видео, сообщение), взаимодействия (знает, сообщает, комментирует и т. д.). Затем авторы использовали существующие ресурсы онтологий и добавили варианты всевозможных связей, включая «папа», «мама», «друг», применили онтологию FOAF для определения участников социальной сети и контента, который они добавляют в сеть. Для описания тегов использовали новую версию SCOT.

Была создана онтология SemSNI (Semantic Social Network Interactions) взаимодействий в социальной сети (посещений страниц, комментариев, личных сообщений) и онтология для анализа социальных сетей SemSNA. При помощи этих онтологий в рамках семантического анализа социальной сети удалось вычислить параметры подграфов социальной сети по разным типам семантических связей («семья»/«family», «мне нравится» / «favorite», «дружба»/«isFriendOf») и типам взаимодействий («комментирует», «создает сообщение» и др.).

Программные приложения для анализа социальных сетей

Для анализа социальных сетей существует множество приложений для моделирования взаимодействий и процессов в сети, для вычисления определенных параметров сети и для визуализации графа сети. Например, приложения по визуализации сети ВКонтакте (см. <http://www.yasiv.com/vk>) или Facebook (<http://www.touchgraph.com/facebook>). В них используются различные методы и алгоритмы, которые описаны ранее в данной работе.

К наиболее известным средствам автоматического анализа социальных взаимодействий относятся: NetMiner (<http://www.netminer.com/index.php>), NetworkX (<http://networkx.lanl.gov>),

¹⁰ Taskar B., Wong M., Abbeel P., Koller D. Label and Link Prediction in Relational Data. URL: http://kdl.cs.umass.edu/srl2003_upload/files/taskar-paper.pdf

SNAP (<http://snap.stanford.edu>), UCINET (<http://www.analytictech.com/ucinet>), Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek>), ORA (см. <http://www.casos.cs.cmu.edu/projects/ora>), Cytoscape (<http://www.cytoscape.org>) и др. Для подобных приложений важным требованием является возможность обрабатывать очень большое количество данных. В связи с этим процесс обработки часто распараллеливают.

Существуют приложения, которые моделируют «теорию шести рукопожатий», которые выстраивают цепочку из связей (друзей) между двумя пользователями сети: для русскоязычной сети ВКонтакте (<http://ienot.ru/hand>), для англоязычных сетей (<http://www.sixdegrees.org>, <http://sixdegrees.com>). Эти цепочки, как правило, действительно получаются небольшой длины.

Более подробную информацию о существующих приложениях для анализа социальных сетей можно найти, например, в [28; 29]¹¹.

Заключение

Среди наиболее актуальных проблем анализа социальных сетей можно назвать создание интегральной теории сетей, так как разнообразие подходов к их анализу приводит к проблеме объединения результатов, полученных с помощью каждого из подходов, в рамках единого исследования. Необходимо подобрать универсальный набор мер расстояний, позволяющих определять дистанции между элементами сети, а также необходимо систематизировать различные меры полноты сетей.

В исследовании атрибутов участников сети, связей между ними, выявлении закономерностей построения сетей между участниками могут быть полезны новые методы статистического анализа, комбинации их с алгоритмами из теории графов. Иногда отношения между участниками сети удобно рассматривать как вероятностные (стохастические) характеристики для описания процесса эволюции сетей.

Задачу поиска человека в социальной сети можно в некотором смысле сравнить с задачей поиска релевантного документа в коллекции документов, снабженных ссылками. Поэтому многие методы обработки текстовой информации можно адаптировать для анализа сетей.

К менее распространенным методам можно отнести, например, применение инструментария из топологии. В частности, в работе [20] предлагается рассмотреть вопрос об исследовании социальных сетей с помощью теоремы Кенига, утверждающей, что любой граф может быть уложен без самопересечений на некоторую компактную ориентируемую топологическую поверхность рода S . Это дает возможность с новых позиций рассмотреть вопрос о геометрии социального пространства. В исследовании [13] выдвигается предположение о необходимости определения аналитической структуры сети, в которой участники являются противниками, и отношения между ними не всегда в полной мере известны. Такой вид сетей является гораздо более сложным для исследований, так как связи не могут быть установлены априори. Подобные виды отношений было бы интересно исследовать в аналитических целях.

Для того чтобы обобщить поведение конкретных участников сети на всю сеть, необходимо изучение методов обнаружения и описания характеристик сетей, закономерностей распространения этих характеристик, создание методов, позволяющих определять по структуре социальной сети причины взаимодействий участников. Особенно важными эти процессы являются при анализе современных социальных сетей большого размера.

Список литературы

1. Чураков А. Н. Анализ социальных сетей // СоцИс. 2001. № 1. С. 109–121.
2. Charu C. Aggarwal. Social Network Data Analytics. 2011. 520 p.
3. Milgram S. The Small World Problem // Psychology Today. 1967. Vol. 2. P. 60–67.

¹¹ См. также: Social network analysis software. URL: http://en.wikipedia.org/wiki/Social_network_analysis_software

4. *Granovetter M. S.* The Strength of Weak Ties // *American Journal of Sociology*. 1973. Vol. 78. No. 6. P. 1360–1380.
5. *Kleinberg J. M.* Authoritative Sources in a Hyperlinked Environment // *J. ACM*. 1999. Vol. 46. No. 5. P. 604–632.
6. *Johnson J., Ironsmith M.* Assessing Children's Sociometric Status: Issues and the Application of Social Network Analysis // *Journal of Group Psychotherapy, Psychodrama & Sociometry*. 1994. Vol. 47. Is. 1. P. 36–49.
7. *Gyöngyi Z., Garcia-Molina H., Pedersen J.* Combating Web Spam with TrustRank // *Proceedings of the International Conference on Very Large Data Bases*. 2004. Vol. 30. P. 576.
8. *Davern M.* Social Networks and Economic Sociology: A Proposed Research Agenda for a More Complete Social Science // *American Journal of Economics & Sociology*. 1997. Vol. 56. Is. 3. P. 287–302.
9. *Koren Y.* On Spectral Graph Drawing // *Proceedings of the 9th International Computing and Combinatorics Conference*. Springer, 2003. P. 496–508.
10. *Fortunato S.* Community Detection in Graphs // *Phys. Rep.* 2010. Vol.486. No. 3–5. P. 75–174.
11. *Wasserman S., Faust K.* *Social Network Analysis: Methods And Applications*. N. Y.: Cambridge University Press, 1994. 825 p.
12. *Jensen D., Neville J.* Data Mining in Social Networks // *Proceedings of the National Academy of Sciences Symposium on Dynamic Social Network Analysis*. 2002. P. 289–302.
13. *Bonchi F., Castillo C., Gionis A., Jaimes A.* Social Network Analysis and Mining for Business Applications // *ACM TIST*. 2011. Vol. 2 (3). P. 22–58.
14. *Hanneman R.* *Computer-Assisted Theory Building: Modeling Dynamic Social Systems*. Riverside, CA: University of California, Riverside, 1988.
15. *Leskovec J., Kleinberg J., Faloutsos C.* Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations // *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (KDD)*. N. Y., 2005. P. 177–187.
16. *Leskovec J., Backstrom L., Kumar R., Tomkins A.* Microscopic Evolution of Social Networks // *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. N. Y., 2008. P. 462–470.
17. *Tantipathananandh C., Berger-Wolf T., Kempe D.* A Framework for Community Identification in Dynamic Social Networks // *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. N. Y.: ACM Press, 2007. P. 717–726.
18. *Sun J., Faloutsos C., Papadimitriou S., Yu P.* Graphscope: Parameter-Free Mining of Large Time-Evolving Graphs // *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. N. Y., 2007. P. 687–696.
19. *Ferlez J., Faloutsos C., Leskovec J., Mladenic D., Grobelnik M.* Monitoring Network Evolution Using MDL // *Proceedings of the International Conference on Data Engineering*. 2008. P. 1328–1330.
20. *Berlingerio M., Bonchi F., Bringmann B., Gionis A.* Mining Graph Evolution Rules // *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases. Lecture Notes in Computer Science*. Springer, 2009. Vol. 5781. P. 115–130.
21. *Desikan P., Srivastava J.* Mining Temporally Changing Web Usage Graphs // *Proceedings of the International Workshop on Mining Web Data for Discovering Usage Patterns and Profiles*. 2004. P. 1–17.
22. *Inokuchi A., Washio T.* A Fast Method to Mine Frequent Subsequences from Graph Sequence Data // *Proceedings of the IEEE International Conference on Data Mining*. 2008. P. 303–312.
23. *Liu Z., Yu J., Ke Y., Lin X., Chen L.* Spotting Significant Changing Subgraphs in Evolving Graphs // *Proceedings of the 8th International Conference on Data Mining*. 2008. P. 917–922.
24. *Borgwardt K. M., Kriegel H.-P., Wackersreuther P.* Pattern Mining in Frequent Dynamic Subgraphs // *Proceedings of the IEEE International Conference on Data Mining*. 2006. P. 818–822.
25. *Liben-Nowell D., Kleinberg J.* The Link Prediction Problem for Social Networks // *Proceedings of the 12th International Conference on Information and Knowledge Management*. N. Y.: ACM Press, 2003. P. 556–559.

26. Kumar R., Novak J., Raghavan P., Tomkins A. Structure and Evolution of Blogspace // Commun. ACM. 2004. Vol. 47. No. 12. P. 35–39.
27. Éréteo G., Gandon F., Buffa M., Corby O. Semantic Social Network Analysis // Proceedings of the 8th International Semantic Web Conference. 2009. P. 180–195.
28. Прохоров А., Ларичев Н. Компьютерная визуализация социальных сетей // КомпьютерПресс. 2006. № 9. С. 156–160.
29. Huisman M., Marijtje A. J. van Duijn. A Reader's Guide to SNA Software // The SAGE Handbook of Social Network Analysis. SAGE. 2011. P. 578–600.

Материал поступил в редколлегию 04.07.2012

T. V. Batura

METHODS OF SOCIAL NETWORKS ANALYSIS

This work is dedicated to social network analysis. There are four main research areas: structural, resource, regulatory, and dynamic. For the solving of the problems in social network analysis following methods are used: graph and stochastic models, models of network evolution, methods involving ontologies, structural and relational models, machine learning methods, network visualization techniques, etc. The article also describes the most popular computer social networks and some software applications to analyze them. It is identified some possible paths of research: the creation of an integrated theory of social networks, adaptation of methods of natural language text processing to the online content, etc.

Keywords: Social networks analysis, network model, graph of network, data mining, centrality.