

бионических алгоритмов, предусматривающей их автоматическую настройку.

Подытоживая, отметим, что в данной работе предложен и апробирован подход к построению адаптивных бионических алгоритмов комбинаторной оптимизации, основанный на автоматической подстройке вероятностей выбора типов операторов или дискретных значений параметров. Показано, что полученные таким образом алгоритмы близки по эффективности к своим оптимально настроенным аналогам, хотя и не требуют усилий по настройке. Данный подход не может напрямую распространяться на алгоритм IDWs, обладающий большим количеством вещественных параметров, требующих тонкой подстройки. Для его применения в данном случае требуются отдельные исследования.

В дальнейшем предполагаются разработка адаптивных версий других бионических алгоритмов, расширение круга решаемых ими задач, а также их практическое применение.

Литература

1. Гуменикова А.В., Емельянова М.Н., Семенкин Е.С., Сопов Е.А. Об эволюционных алгоритмах решения сложных

задач оптимизации // Вестн. СибГАУ. 2003. № 4. С. 14–23.

2. Dorigo M., Gambardella L.M., Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem, *IEEE Transactions on Evolutionary Computation*, Vol. 1, Is. 1, 1997, pp. 53–66.

3. Shah-Hosseini H., Optimization with the Nature-Inspired Intelligent Water Drops Algorithm, *Evolutionary Computation*, Vienna, InTech, 2009, pp. 299–319.

4. Lin Sh., Kernighan B.W., An Effective Heuristic Algorithm for the Traveling-Salesman Problem, *Operations Research*, 1973, no. 21 (2), pp. 498–516.

5. Oliver I.M., Smith D.J., Holland J.R.C., A study of permutation crossover operators on the travelling salesman problem, *Proc. 2nd Intern. Conf. on Genetic Algorithms and their application*, 1987, pp. 224–230.

References

1. Gumennikova A.V., Emelyanova M.N., Semekin E.S., Sopov E.A., *Vestnik SibGAU* [Bulletin of the SibSAU], 2003, no. 4, pp. 14–23.

2. Dorigo M., Gambardella L.M., *IEEE Transactions on Evolutionary Computation*, Vol. 1, iss. 1, 1997, pp. 53–66.

3. Shah-Hosseini H., *Evolutionary Computation*, Vienna, InTech, 2009, pp. 299–319.

4. Lin Sh., Kernighan B.W., *Operations Research*, 1973, no. 21 (2), pp. 498–516.

5. Oliver I.M., Smith D.J., Holland J.R.C., *Proc. of the 2nd Int. Conf. on Genetic Algorithms and their application*, 1987, pp. 224–230.

УДК 519.68

МОДЕЛИ И МЕТОДЫ АНАЛИЗА КОМПЬЮТЕРНЫХ СОЦИАЛЬНЫХ СЕТЕЙ

Т.В. Батура, к.ф.-м.н.

(Институт систем информатики им. А.П. Ершова СО РАН,
просп. Лаврентьева, 6, г. Новосибирск, 630090, Россия, tbatura@ngs.ru)

Статья представляет собой обзор методов анализа компьютерных социальных сетей, который используется для исследования взаимодействий между участниками сети, прогнозирования их поведения, классификации участников, моделирования информационных потоков в сетях. Существуют четыре основных направления исследований при таком анализе: структурное, ресурсное, нормативное и динамическое. Для решения различных задач в данной области 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. В работе дается краткое описание популярных компьютерных социальных сетей, рассмотрены различные подходы и модели, применяемые при вычислении характеристик сети, а также перечислены программные приложения, помогающие автоматизировать процесс анализа. Намечены некоторые возможные направления дальнейших исследований в этой области, в частности, более существенная адаптация методов обработки текстовой информации к сетевому контенту. Чтобы обобщить поведение конкретных участников сети на всю сеть, необходимо создание единой теории социальных сетей, в рамках которой можно было бы изучать методы обнаружения и описания характеристик сетей, закономерности распространения этих характеристик, создание методов, позволяющих по структуре социальной сети определять причины взаимодействий участников.

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

SOCIAL NETWORKS ANALYSIS MODELS AND METHODS

Batura T.V., Ph.D. (A.P. Ershov Institute of Informatics Systems, Siberian Branch of RAS,
Acad. Lavrentjev Av., 6, Novosibirsk, 630090, Russia, tbatura@ngs.ru)

Abstract. The paper represents the survey of the social network analysis methods. That analysis is used to study the interactions between network members, prediction their behavior, classification of the members, modeling information flow in

networks. There are four main research areas: structural, resource, regulatory and dynamic. To solve the problems in social network analysis the 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. This paper provides a brief description of popular computer social networks, various approaches and models used in the network characteristics calculation, and lists the software applications that help automate the process of analysis. Several possible directions of further studies in this area are outlined: creation of social networks integrated theory, adaptation of natural language text processing methods to the online content, etc. In order to generalize the behavior of specific members of the network to the entire network, it is necessary to study methods for the detection and characterization of networks, patterns of distribution of these characteristics, the creation of methods to determine the cause of the participants' interaction.

Keywords: social network analysis, network mapping, dynamic network analysis, network models, graph of network, data mining, centrality, network representation, network characteristics.

Компьютерные социальные сети становятся все более популярными. Одним из результатов взаимодействия людей посредством таких сетей является получение огромного количества информации различных форматов: тексты, картинки, аудио, видео и пр. Для того чтобы делать определенные выводы о процессах, протекающих в обществе, прогнозировать поведение его участников, моделировать социальное взаимодействие, необходимо осуществлять анализ информации, существующей в социальных сетях. Поэтому создание новых и интеграция уже созданных методов и моделей анализа компьютерных социальных сетей представляет интерес для исследования.

Основные направления исследований компьютерных социальных сетей

В анализе социальных сетей можно выделить четыре основных направления исследований: структурное, ресурсное, нормативное и динамическое [1]. В каждом из них решается довольно большой круг задач и применяются методы из различных областей знаний.

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.

При структурном анализе и анализе поведения связей используются методы статистического анализа, определения сообществ, алгоритмы классификации. Изучается поведение вершин в процессе кластеризации и типичных временных характеристик социальных сетей. Например, как меняется структура сети в процессе роста или как меняются поведение и распределение связных компонентов графа.

Большое значение придается определению сообществ в социальных сетях. Цель – попытаться определить регионы сети, внутри которых происходит активное взаимодействие участников. Алгоритмически эту задачу можно отнести к задаче о разделении графов [2]. Необходимо разделить сеть на плотные регионы на основе поведения

связей между вершинами. Компьютерные социальные сети динамичны, что приводит к затруднениям с точки зрения выявления сообществ. В некоторых случаях удастся интегрировать информационное содержимое сети в процесс определения сообществ. Тогда контент является вспомогательным средством для выявления групп участников с похожими интересами.

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, объем и характер информации. Основным показателем, определяющим различия в ресурсах участников сети, является сила структурной позиции участника.

Важная задача данного направления – анализ содержания социальных сетей. Сетевой контент служит источником для широкого спектра приложений, ориентированных на извлечение и анализ данных. Использование содержания сети помогает значительно улучшить качество выводов при анализе социальных сетей, например, в задачах кластеризации и классификации. Можно выделить четыре вида анализа контента сети [2].

1. При *анализе общей информации с произвольными типами данных* применяются, например, методы случайных блужданий. Одним из наиболее известных алгоритмов, использующих подобные методы, является алгоритм ссылочного ранжирования (PageRank) для приписывания веса веб-документам: 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, то есть вершина, ссылающаяся на другую вершину с большим весом, сама получает больший вес. Этот алгоритм может также применяться для поиска и классификации сущностей и участников в социальной сети, для оценки вероятности посещения той или иной вершины. Естественно, что вершины, лучше расположенные со структурной точки зрения, имеют более высо-

кий вес, а значит, являются более важными. Методы случайного блуждания могут быть также полезны для объединения участников в группы относительно наиболее влиятельных участников.

2. Для *сенсорного и потокового анализа* используются методы интеграции данных, поступающих с датчиков, и данных, имеющихся в социальных сетях. Многие современные сотовые телефоны поддерживают возможность взаимодействия пользователей друг с другом динамически в режиме реального времени в зависимости от их местоположения и статуса. Их применяют для получения информации о человеке или совокупности свойств объектов, которые отслеживаются.

3. *Анализ мультимедиа*. Существует много сайтов (Flickr, YouTube и др.) по обмену и совместному использованию средствами массовой информации: фото, видео, аудио. При наличии тегов или комментариев анализ мультимедиа можно свести к анализу текстовой информации в сети.

4. *Анализ текстовой информации*. В социальной сети содержится много текстовой информации в различных формах, например, можно оставлять комментарии, ссылки на посты (сообщения), блоги или статьи с новостями. Иногда пользователи могут отмечать друг друга, что тоже является формой текстовой информации в виде ссылок. Особый интерес представляет расстановка тегов (меток или ключевых слов), описывающих различные объекты: картинки, текст, видео. В рамках данного подхода изучают свойства потоков тегов, моделей тегирования, семантику тегов, визуализацию тегов, приложения для их расстановки и т.д. Интересными являются, например, следующие вопросы: почему люди расставляют теги, что влияет на выбор людей при тегировании, как промоделировать процесс подобной разметки, какие разновидности тегов существуют, как создаются теги.

Нормативное направление изучает уровень доверия между участниками, а также нормы, правила и санкции, влияющие 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. Так как в основе социальных сетей лежит взаимодействие между различными участниками, естественно предположить, что это взаимодействие оказывает влияние на участников в терминах их поведения. Вопросы этого направления: как моделировать влияние на основе информации об участниках; как моделировать распространение влияния; кто из участников наиболее влиятельный в процессе распространения.

В социальных сетях содержится большое количество личной информации об участниках, например, интересы, друзья, демография и др. Это

может привести к несанкционированному распространению личной информации в сетях. В решении такого типа задач полезно применять модели на основе механизмов конфиденциальности.

Так как для эффективности и устойчивости социальной сети важны функциональные роли ее участников, социальная сеть может быть 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 достичь общей цели, когда специалисты сотрудничают друг с другом. Помимо экспертов, при анализе социальных сетей представляют интерес так называемые брокеры (лидеры). 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).

Динамический подход – направление в изучении социальных сетей, в котором объектами исследований являются изменения в сетевой структуре с течением времени: 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 возникают вопросы: согласно каким законам происходят долгосрочные изменения между крупными сообществами в социальных сетях, существуют ли какие-либо стационарные конфигурации социальной сети, как развиваются сами сообщества во времени, какие изменения могут происходить, как можно отследить и представить их.

Важной задачей является прогноз формирования связей в социальных сетях. В большинстве приложений для анализа социальных сетей связи считаются динамическими и могут изменяться с течением времени. В процесс прогнозирования связей могут быть вовлечены как структура сети, так и информация об особенностях различных вершин. Для решения таких задач предлагается строить разнообразные структурные и реляционные модели [2].

Визуализация помогает естественным образом свести воедино информацию о сетях и сделать ее более доступной для понимания. Важным является создание алгоритмов, сочетающих в себе методы анализа и методы визуализации, чтобы улучшить понимание структуры и динамики сети.

Некоторые наиболее известные социальные сети

К крупнейшим социальным сетям по числу пользователей относятся Facebook, LinkedIn, ВКонтакте, Twitter, Одноклассники.ru, YouTube и др. [3, 4].

Facebook. Сеть основана в 2004 году Марком Цукербергом. По данным за апрель 2012 года, аудитория Facebook составляет около 901 млн пользователей. Каждый 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.

YouTube. Сервис, предоставляющий услуги видеохостинга, основан в 2005 году. 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 года. Ежеминутно на YouTube загружают 60 часов видео. В январе 2012 года ежедневное количество просмотров видео на сайте достигло 4 млрд.

LinkedIn. Социальная сеть была основана Ридом Хоффманом в декабре 2002 года, запущена в мае 2003 года. В основном сеть используется для поиска и установления деловых контактов. По данным на февраль 2012 года, в LinkedIn зарегистрировано свыше 160 млн пользователей. Чуть меньше половины пользователей LinkedIn являются жителями США.

ВКонтакте. Сеть основана в 2006 году Павлом Дуровым. По данным на март 2012 года, 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 медиа-архивов Рунета. Facebook, Одноклассники.ru и другие социальные сети используют протокол обмена сообщениями XMPP (Extensible Messaging and Presence Protocol), ранее известный как Jabber.

Twitter. Создана в 2006 году Джеком Дорси. На начало 2012 года сервис насчитывал более 140 млн пользователей. 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.

Одноклассники.ru. Проект запущен в 2006 году, его автором является российский веб-разработчик Альберт Попков. На июнь 2011 года 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. Создана в 2004 году. По данным на июнь 2011 года, в сети зарегистрирован 51 млн пользователей. Сервис, предназначенный для хранения и дальнейшего использования пользователями цифровых фотографий и видеороликов, является одним из первых Web 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 – анализ, проведенный в 1970-е годы американским социологом Марком Грановеттером. Он показал, что для многих социальных задач, таких как поиск работы, слабые связи оказываются намного эффективнее сильных [5]. Слабые связи являются важными источниками информации, так как они помогают получить дополнительные сведения об участнике или сообществе, в котором он состоит, из других источников. Этот эффект Грановеттер назвал *силой слабых связей*. Сила связей между участниками определяется как линейная комбинация продолжительности, эмоциональной насыщенности, интимности или конфиденциальности и значимости взаимных услуг, которые характеризуют данное взаимодействие и соответствующее ему ребро графа.

Еще один широко известный пример анализа социальных сетей – эксперимент американского психолога Милгрэма, проведенный в 1969 году [6]. 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 каж-

дый 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.

Рассмотрим немного подробнее основные модели анализа компьютерных социальных сетей.

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.

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, рассматривают два основных типа отношений: дружба (люди знакомы друг с другом) и интересы (есть общие интересы, люди входят в одну группу по интересам).

Выделяют [1] три следующих вида графовых моделей.

- *Стохастические блочные 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.

Чтобы определить относительную важность (вес) вершин графа (то есть насколько влиятель-

ным в рамках конкретной сети является ее участник), вводят понятие центральности – меры близости к центру графа. Следует отметить, что речь идет не о геометрической центральности при визуализации графа отношений. Центральность можно определить разными способами, поэтому существуют различные меры центральности. Центральность по степени (Degree centrality) определяется как количество связей, инцидентных вершине, входящие связи характеризуют популярность человека, выходящие – его общительность. Центральность по близости (Closeness centrality) является показателем того, насколько быстро распространяется информация в сети от одного участника к остальным, то есть насколько близок рассматриваемый участник ко всем остальным участникам сети. Центральность по посредничеству (Betweenness centrality) – характеристика участника, показывающая его важность при распространении информации. Центральность по собственному вектору (Eigenvector centrality) демонстрирует зависимость между центральностью участника и центральностями его друзей.

Полезной характеристикой при анализе социальных сетей является уровень доверия. Алгоритм вычисления уровня доверия (TrustRank) 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.

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 максимально полный подграф данного графа. Если анализировать процессы распространения информации в графах, можно дать другое определение сообщества: множество участников, путь между двумя любыми

участниками которого 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. Более детальный обзор методов обнаружения сообществ можно найти например в [7].

Структурная эквивалентность участников сети. Этот подход является противоположным исследованию связанных групп. Участники эквивалентны, когда они занимают одинаковые позиции в социальной структуре сети, то есть 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 [8]. Наряду со структурной эквивалентностью используется регулярная эквивалентность участников. В этом случае участники эквивалентны, когда они одинаковым образом взаимодействуют с участниками одного типа.

Рольевые алгебры. Этот метод анализа социальных сетей направлен на выявление логики взаимодействий участников сети в блоковых моделях, что позволяет выявлять сходство 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. 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 ». Важной характеристикой, описывающей локальные связи участников и часто используемой при анализе диад и триад, является сбалансированность. Сбалансированность – это

отсутствие ситуаций типа «положительное 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 ». Предполагается, что сбалансированные сети психологически более комфортны для участников и более устойчивы по сравнению с несбалансированными.

Модели динамики сети. Для исследования динамики сети используются подходы, описанные в [9, 10].

Моделирование эволюции графа сети изучают различные стратегии формирования сети и показывают, что расположение ребер играет важную роль в эволюции сетей. Например, в [11] было обнаружено, что со временем плотность сети увеличивается. При этом количество ребер обычно увеличивается линейно с ростом количества вершин. В итоге 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.

Среди работ, представляющих алгоритмические инструменты для анализа эволюции сетей, можно выделить [12], 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. Для поиска закономерностей в данных и обнаружения сообществ в динамических сетях применяется принцип минимума длины описания MDL (Minimum Description Length).

Анализ графов развития сети иногда удобно проводить, опираясь на парадигму извлечения ассоциативных правил и анализа частотных моделей. Вводятся правила эволюции графа, новый тип частотных моделей и рассматривается проблема поиска типичных моделей структурных изменений в динамических сетях. Сначала вычисляются набор частотных моделей графа, который описывает характерные эволюционные механизмы, а затем находят правила эволюции графа, удовлетворяющие заданному ограничению минимальности доверия. Для анализа графов развития сети применяются также методы извлечения часто встречающихся подграфов, меняющихся со временем. Релевантным подграфом считается не наиболее частый, а наиболее значимый. Историю ребра в динамическом графе можно представлять в виде последовательности нулей и единиц, соответствующих наличию или отсутствию того или иного ребра. Затем для получения частотных моделей графа применяются традиционные методы полу-

чения графа из описания в виде последовательности нулей и единиц.

Задача прогнозирования формирования связей состоит в определении, будут ли две конкретные вершины соединены друг с другом через некоторый промежуток времени. Для ее решения 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 связей, основанные на машинном обучении и 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 [13] за основу предлагается брать сами свойства пользователей и, например, наличие большого количества связей (в блогосфере) может быть объяснено сопоставлением демографических групп, общих интересов или географической близостью.

Методы на основе онтологий. Оценить параметры социальных сетей (диаметр, количество участников, среднюю длину пути и др.) можно при помощи онтологий [14]. Сначала анализируются виды элементов сети: люди, объекты (музыка, фото, видео, сообщения) и взаимодействия (знает, сообщает, комментирует и т.д.). Затем применяется онтология FOAF (Friend of a friend) для определения участников социальной сети и контента, который они добавляют в сеть. Эта онтология описывает людей, их активность и отношения к другим людям и объектам. Описание социальных связей между людьми в FOAF основывается на транзитивности доверия. Для описания тегов можно использовать SCOT. В результате 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 сети удалось вычислить параметры подграфов по разным типам семантических связей (семья, мне нравится, дружба) 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 или Facebook. В них ис-

пользуются различные методы и алгоритмы, которые описаны выше.

К наиболее известным средствам автоматического анализа социальных взаимодействий относятся NetMiner, NetworkX, SNAP, UCInet, Pajek, ORA, Cytoscape 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», то есть выстраивают цепочку из связей (друзей) между двумя пользователями сети: для русскоязычной сети ВКонтакте и для англоязычных сетей. Эти цепочки, как правило, действительно получаются небольшой длины.

Более подробную информацию о существующих приложениях для анализа социальных сетей можно найти например в [15, 16].

В заключение следует отметить, что существование разных подходов к анализу компьютерных социальных сетей приводит к проблеме объединения результатов, полученных в ходе исследований. Поэтому к актуальным проблемам сетевого анализа можно отнести следующие: создание единой теории социальных сетей, создание универсального набора мер расстояний, позволяющих определять дистанции между элементами сети, а также проблему систематизации различных мер 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 [1] 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 [9] выдвигается предположение о необходимости определения аналитической структуры сети, в которой участники являются противниками, и отношения между ними не все-

гда 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 обобщения поведения 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.

Литература

1. Чураков А.Н. Анализ социальных сетей // Социологические исследования. 2001. № 1. С. 109–121.
2. Charu C., *Social network data analytics*, 2011, 520 p.
3. Социальная сеть. URL: http://ru.wikipedia.org/wiki/Социальная_сеть (дата обращения: 22.06.2012).
4. List of social networking websites, 2012. URL: http://en.wikipedia.org/wiki/List_of_social_networking_websites (дата обращения: 25.06.2012).
5. Granovetter M.S., The Strength of Weak Ties. *American Journ. of Sociology*, 1973, Vol. 78, no. 6, pp. 1360–1380.
6. Milgram S., The Small World Problem. *Psychology Today*, 1967, Vol. 2, pp. 60–67.
7. Fortunato S., Community detection in graphs. *Physics Reports*, 2010, Vol. 486, Iss. 3–5, pp. 75–174.
8. Wasserman S., Faust K., *Social Network Analysis: Methods and Applications*, Cambridge Univ. Press, 1994, 825 p.
9. Bonchi F., Castillo C., Gionis A., Jaimes A., *Social Network Analysis and Mining for Business Applications*, ACM TIST, 2011, Vol. 2, Iss. 3, pp. 22–58.
10. Hanneman R., *Computer-Assisted Theory Building: Modeling Dynamic Social Systems*, Riverside, Univ. of California, 1988. URL: <http://faculty.ucr.edu/~hanneman> (дата обращения: 05.07.2012).
11. Leskovec J., Kleinberg J., Faloutsos C., Graphs over time: Densification laws, shrinking diameters and possible explanations, *Proc. 11th ACM SIGKDD Intern. Conf. on Knowledge Discovery in Data Mining*, NY, 2005, pp. 177–187.
12. Tantipathananandh C., Berger-Wolf T., Kempe D., A framework for community identification in dynamic social networks. *Proc. 13th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining*, NY, 2007, pp. 717–726.

13. Kumar R., Novak J., Raghavan P., Tomkins A., Structure and evolution of blogspace. *Communication of the ACM*, 2004, Vol. 47, no. 12, pp. 35–39.

14. Értéő G., Gandon F., Buffa M., Corby O., *Semantic Social Network Analysis*. Proc. 8th Intern. Semantic Web Conf., 2009, pp. 180–195.

15. Прохоров А., Ларичев Н. Компьютерная визуализация социальных сетей // КомпьютерПресс. 2006. № 9. С. 156–160.

16. Social network analysis software, 2012. URL: http://en.wikipedia.org/wiki/Social_network_analysis_software (дата обращения: 05.07.2012).

References

1. Churakov A.N., *Sotsiologicheskie issledovaniya* [Opinion surveys], Moscow, 2001, no. 1, pp. 109–121.
2. Charu C., *Social network data analytics*, 2011, 520 p.
3. *Sotsialnaya set* [Social network], 2012, available at: http://ru.wikipedia.org/wiki/Социальная_сеть (accessed 22 June 2012).
4. *List of social networking websites*, 2012, available at: http://en.wikipedia.org/wiki/List_of_social_networking_websites (accessed 25 June 2012).
5. Granovetter M.S., *American Journ. of Sociology*, 1973, Vol. 78, no. 6, pp. 1360–1380.
6. Milgram S., *Psychology Today*, 1967, Vol. 2, pp. 60–67.
7. Fortunato S., *Physics Reports*, 2010, Vol. 486, iss. 3–5, pp. 75–174.
8. Wasserman S., Faust K., *Social Network Analysis: Methods And Applications*, Cambridge Univ. Press, 1994, 825 p.
9. Bonchi F., Castillo C., Gionis A., Jaimes A., *ACM TIST*, 2011, Vol. 2, iss. 3, pp. 22–58.
10. Hanneman R., *Computer-Assisted Theory Building: Modeling Dynamic Social Systems*, Riverside, Univ. of California, 1988, available at: <http://faculty.ucr.edu/~hanneman> (accessed 5 July 2012).
11. Leskovec J., Kleinberg J., Faloutsos C., *Proc. of the 11th ACM SIGKDD Int. Conf. on Knowledge Discovery in Data Mining*, NY, 2005, pp. 177–187.
12. Tantipathananandh C., Berger-Wolf T., Kempe D., *Proc. of the 13th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, New York, 2007, pp. 717–726.
13. Kumar R., Novak J., Raghavan P., Tomkins A., *Communication of the ACM*, 2004, Vol. 47, no. 12, pp. 35–39.
14. Értéő G., Gandon F., Buffa M., Corby O., *Proc. of the 8th Int. Semantic Web Conf.*, 2009, pp. 180–195.
15. Prokhorov A., Larichev N., *ComputerPress*, 2006, no. 9, pp. 156–160.
16. *Social network analysis software*, 2012, available at: http://en.wikipedia.org/wiki/Social_network_analysis_software (accessed 7 July 2012).

УДК 007:519.816

ПРИМЕНЕНИЕ СОВРЕМЕННЫХ ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ ДЛЯ МОДЕРНИЗАЦИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

(Работа выполнена при финансовой поддержке РФФИ, проект № 12-07-31239)

И.Е. Куриленко, к.т.н., доцент; Ян Пайнг Хейн, аспирант

(Национальный исследовательский университет «Московский энергетический институт»,
ул. Красноказарменная, 14, г. Москва, 111250, Россия, ivan@appmat.ru, yanprainghein@gmail.com)

В статье рассматриваются возможности современных информационных технологий, полезные при модернизации образовательного процесса. Описывается пример построения современной образовательной среды (в плане создания