

КОНЦЕПТУАЛЬНЫЙ ПОДХОД К АНАЛИЗУ ОНЛАЙНОВЫХ СОЦИАЛЬНЫХ СЕТЕЙ¹

Губанов Д. А.², Чхартишвили А. Г.³

(ФГБУН Институт проблем управления

им. В.А. Трапезникова РАН, Москва

Лаборатория Цифрового Общества (DSL), Москва)

В статье предложен конструктивный подход к анализу социальных сетей. Предложена концептуальная модель социальной сети, поставлены базовые задачи анализа и управления в социальных сетях, определены методы и алгоритмы анализа активности в социальных сетях, а также технологии мониторинга и анализа социальных сетей.

Ключевые слова: социальная сеть, концептуальная модель, информационное управление, информационные технологии.

1. Введение

В последнее десятилетие онлайн-социальные сети прочно вошли в нашу жизнь и стали ее неотъемлемой частью. Практически у каждого сейчас есть своя страница в тех или иных онлайн-социальных сетях; наиболее популярные сети (FaceBook – 955 миллионов пользователей, ВКонтакте –

¹ Работа выполнена при поддержке гранта РФФИ № 10-07-00104.

² Дмитрий Алексеевич Губанов, кандидат технических наук, старший научный сотрудник ИПУ РАН, старший исследователь Лаборатории Цифрового Общества (DSL) (dmitry.a.g@gmail.com)

³ Александр Гедеванович Чхартишвили, доктор физико-математических наук, главный научный сотрудник ИПУ РАН, научный директор Лаборатории Цифрового Общества (DSL) (sandro_ch@mail.ru)

140 миллионов пользователей¹) пытаются обеспечить пользователям возможность общаться, обсуждать новости, вступать в группы по интересам, слушать музыку и видео, обмениваться ссылками и т.п., не покидая их пределы. В результате российский интернет-пользователь проводит в среднем в месяц в социальных сетях 10,4 часа².

В целом складывающуюся в онлайн-социальных сетях (блоги, форумы, классические социальные сети) ситуацию можно охарактеризовать следующими тезисами: быстрыми темпами растет число и интенсивность взаимодействия пользователей социальных сетей; пользователи оставляют много информационных следов в интернете (эти следы могут быть неявными); наращивание активности обсуждения может происходить «стихийно» (латентный, фрактальный характер) или по определенному поводу; на конкурентном рынке совершенствуется методический и ИТ-инструментарий поддержки взаимодействия пользователей (краудсорсинг, экспертные сети, виртуальное сотрудничество, интеллектуальные технологии).

Бурное развитие технологий и сопутствующие ему изменения в экономической, социальной, политической жизни делает актуальным разработку концептуального подхода к анализу онлайн-социальных сетей, чему и посвящена данная работа. Необходимо отметить, что существуют монографии [см. например, 2, 8], посвященные анализу онлайн-социальных сетей, но в них отсутствуют сколько-нибудь развернутые содержательные модели онлайн-социальных сетей (тем более с точки зрения постановки задач управления).

2. Подход к анализу социальных сетей

Основными задачами информационно-аналитической поддержки пассивной и активной работы с социальными сетями являются: мониторинг и анализ социальных сетей (для дости-

¹ По состоянию на июнь 2012 года (<http://pdf.secdatabase.com/700/0001193125-12-316895.pdf>) и сентябрь 2012 года (<http://vk.com/help.php?page=about>)

² ComScoreWorldMetrics, декабрь 2011 г.

жения понимания происходящих в социальных сетях процессов), прогнозирование и управление (для приведения социальной сети в требуемое состояние).

Для выполнения этих задач предлагается конструктивный подход, согласно которому:

1) определяются содержательные вопросы, на которые предполагается получить ответы;

2) определяются показатели и их системы для ответа на содержательные вопросы (отметим, что для определения показателей требуются содержательные модели, для расчета показателей – математические модели и методы);

3) разрабатываются технологии мониторинга и анализа социальных сетей;

4) разрабатываются технологии генерации отчетов и выдачи рекомендаций в зависимости от предметной области и заинтересованных лиц.

В данной работе рассматриваются первые три уровня предлагаемого нами подхода.

3. Основные элементы модели социальной сети

Основными элементами модели социальной сети предлагается считать следующие типы объектов:

- внешний узел (СМИ, информационный ресурс и т.п.; внешние узлы обычно существуют в том же виртуальном пространстве, что и онлайн-социальные сети (например, онлайн-ресурсы типа *lenta.ru*));

- внешний управляющий субъект, оказывающий те или иные воздействия на социальную сеть (политический деятель, партия, коммерческая организация и т.п.);

- социальная сеть в целом либо подсеть (множество узлов, выделенных по какому-либо признаку, и связей между ними);

- пользователь сети – член социума;

- узел сети (агент);

- информационное сообщение (пост, комментарий, сообщение);

- информационный объект – некоторая сущность, событие, персона и т.п. (информационные объекты могут быть сложными и состоять из других информационных объектов);
- информация – описание некоторых информационных объектов в информационном сообщении;
- мнение (суждение по какому-то поводу, точка зрения на какой-то объект, оценка).

Рассмотрим подробнее вышеперечисленные типы объектов.

Социальная сеть понимается нами как социальная структура, состоящая из множества агентов (пользователей, групп и сообществ) и определенного на нем множества отношений (информационного взаимодействия, знакомства, участия в группах и сообществах). В рамках этих отношений в сети происходят различные информационные процессы.

В данной модели достаточно определить явные и легко идентифицируемые отношения: отношения информационного взаимодействия (отношения post/reply, отношения цитирования, отношение подписки на сообщения); отношения знакомства (всевозможные виды контактов в социальных сетях); отношения участия в группах и сообществах. Эти виды отношений важны с точки зрения распространения информации и формирования мнений в социальных сетях. Все остальные виды отношений неявны (например, пользователи неявно связаны между собой, если они подписаны на одних и тех же пользователей, пишут на одни и те же темы и т.п.), эти виды отношений могут выявляться по мере необходимости для решения конкретных задач.

Социальная сеть не существует сама по себе, на процессы в ней могут повлиять какие-то внешние воздействия – либо события (например, выборы Президента РФ), либо какие-то целенаправленные воздействия со стороны внешних субъектов. *Внешние субъекты* (политические силы, коммерческие организации и пр.) могут быть заинтересованы в достижении сетью некоторого целевого состояния: требуемой степени связности (либо, наоборот, разобщенности) сети, формировании необходимой информированности в сети, в формировании требуемого общественного мнения в сети и т.п. Эта заинтересованность побуждает внешних субъектов осуществлять управление (рис. 1).

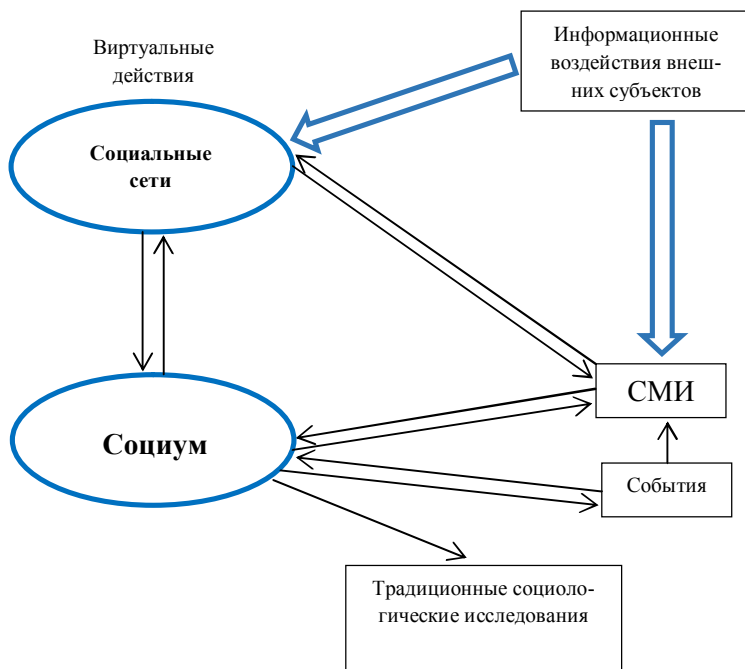


Рис. 1. Социальная сеть как открытая система

Внешнее воздействие (освещение некоторого события в новостях и др.) может повлиять на пользователя социальной сети, в результате чего пользователь может (пользуясь своим «отражением» в сети – агентом) публиковать *информационные сообщения* (посты, комментарии, записи и т.п.), которые содержат соответствующие *информационные объекты* (сущности, события, персоналии и т.п.). Публикация информационных сообщений приводит к постепенному формированию мнений и убеждений относительно информационных объектов (при этом пользователи формируют и меняют свои мнения по тем или иным вопросам под *влиянием* других членов сети, обладающих различной значимостью (авторитетностью)). Мнения и убеждение могут вызвать действия (например, резкое повышение спроса на

некий товар, стихийную массовую акцию протеста и пр.), которые, в свою очередь, могут получить отображение в СМИ.

Внешними воздействиями могут быть хакерские атаки, рекламные кампании, создание бот-сетей, воздействия на поведение пользователей социальных сетей и т.п. Внешнее воздействие на сеть можно выявить (с большой долей уверенности) по некоторым сетевым критериям «странности»/«неестественности». Например, можно предположить, что в некоторый момент на сеть оказывается внешнее воздействие посредством пользователя в каждом из таких случаев, когда пользователь: а) демонстрирует неестественную активность; б) резко меняет взгляды; в) действует согласованно с другими пользователями. Если при этом в данный момент наблюдается связанное с активностью пользователей внешнее событие и определяется круг заинтересованных внешних субъектов, то уверенность во внешнем воздействии возрастает.

Возникают две проблемы: как отличить внешнее воздействие от внутреннего (например, пользователь может поменять взгляды под влиянием своих внутренних причин) и как отличить целенаправленное воздействие (управление) от нецеленаправленного. Приведем некоторые соображения: (1) формальный анализ позволит выявить сам факт воздействия (для некоторых задач анализа этого будет достаточно); (2) формальный анализ, проводимый совместно с содержательным анализом экспертов в соответствующей предметной области (например, политологический анализ), позволит определить, является ли воздействие внешним и целенаправленным.

4. Основные подходы к моделированию социальных сетей

В данном пункте кратко описаны три взаимосвязанных подхода к моделированию социальных сетей: моделирование структуры сети, моделирование распространения информации, моделирование формирования мнений. Отметим, что если конечной целью управляющего субъекта является управление мнениями в социальной сети, то для ее достижения необходимо использовать все три подхода, поскольку структура сети во мно-

гом определяет распространение информации, а оно, в свою очередь, существенно влияет на формирование мнений.

Моделирование структуры социальной сети. Математической абстракцией социальной сети является граф. В социальных сетях вершинами графа являются социальные агенты (например, люди или организации), а ребрами – социальные взаимоотношения (например, отношения дружбы, соавторства). Реальные социальные сети являются большими и сложными (насчитывают десятки и сотни миллионов вершин), иными словами для реальных сетей нелегко ответить на вопросы относительно их структуры и, в конечном итоге, определить модели структуры сети. Поэтому необходимо определить множество ключевых признаков, характеризующих структуру больших и сложных сетей.

Исследователями уже определены десятки признаков/показателей, характеризующих структуру социальных сетей [12]: диаметр и среднее расстояние (малый диаметр социальной сети обуславливает короткую цепочку распространения воздействия по сети), распределение степеней (степенное/безмасштабное распределение в социальной сети обуславливает наличие доминирующих узлов (хабов), связанных с сотнями, тысячами и даже миллионами других узлов, в большинстве своем имеющих всего по несколько связей), показатели связанности и другие показатели.

На основе определенного множества ключевых структурных признаков и характерных их значений можно разработать модели социальных сетей, обладающих структурным сходством с реальными сетями. Долгое время все сложные сети рассматривались как полностью случайные [7], и сети моделировались соединением случайных вершин случайным образом. Однако в 1969 году Милгрэм (Milgram) обнаружил, что любых двух людей «разделяет всего шесть шагов». Появилось множество статических моделей так называемого «малого мира» [11]. Однако, несмотря на то, что статические модели дают возможность учесть сложность сети, они не объясняют причину появления безмасштабных сетей, их природу. Наиболее распространенной динамической моделью появления таких структур является модель «богатый становится богаче» [3], в которой в растущем

графе вероятность появления новых связей у вершины пропорциональна ее степени, т.е. количеству уже имеющихся связей. Существуют и другие подходы к объяснению причин появления структуры социальной сети. Например, в теоретико-игровых моделях [5] стоимость добавления новых связей влияет на механизм преимущественного присоединения новых связей (т.е. для агентов предпочтительны те или иные конфигурации связей).

Моделирование распространения информации. Распространение информации (или, если смотреть шире, активности) – это процесс, посредством которого некоторый информационный объект (информация, информационный вирус, мнение) распространяется по коммуникационным каналам во времени и в пространстве среди узлов сети [12], образуя путь воздействия в сети.

В социальной сети пользователи создают информационные сообщения (посты, комментарии, записи), которые содержат устойчивые информационные объекты (пример информационного объекта – «взрыв в минском метро»).

Информационные объекты могут распространяться по социальной сети от пользователя к пользователю, активируя их и образуя траекторию распространения информационного объекта. Пользователь становится активным по отношению к данному информационному объекту с момента использования информационного объекта в своих информационных сообщениях (совокупность состояний активности всех пользователей в сети в определенный момент времени назовем состоянием сети).

Как уже было сказано выше, на активность пользователей может повлиять внешнее воздействие, которое оказывается определенным внешним субъектом, например, средством массовой информации, освещающим события в новостях, или хакером, осуществляющим дефейс значимого веб-ресурса. В конечном итоге некоторый пользователь становится источником воздействий, используя соответствующие информационные объекты в информационных сообщениях.

Определив основные элементы информационных процессов в социальных сетях, можно разработать модели динамики активности (распространения информации) в социальных сетях. Эти модели делятся на две категории: макроуровня и микро-

уровня. Модели макроуровня (например, [6]) рассматривают сеть в целом, без учета связей между узлами. Модели микроуровня, учитывающие эти связи, в свою очередь разбиваются на следующие категории моделей: модель с порогами [10]; модель независимых каскадов (в которой каждый узел получает на некотором шаге шанс активировать другие узлы) [9]; модели пространства, базирующиеся на аналогиях с физикой, медициной и другими разделами науки [4] и т.д.

Моделирование формирования мнений. Обмен информацией пользователями сети приводит к формированию мнений (убеждений). Мнение – это суждение по какому-то поводу, точка зрения на какой-то информационный объект, оценка информационного объекта. В общем случае оно выражается в тексте сообщения или в виде численной оценки (пример – «по моему мнению, курс доллара через месяц составит 30,5 рублей»). Мнения пользователей по тем или иным вопросам могут меняться под влиянием других членов сети, обладающих различной влиятельностью. Вопросам формирования мнения, а также моделям информационного влияния, управления и противоборства посвящена монография [1].

5. Задачи анализа, прогнозирования и управления в социальных сетях

Лица, принимающие решения, заинтересованы в информационно-аналитической поддержке пассивной и активной работы с социальными сетями. В связи с этим основными задачами являются (см. рис. 2): мониторинг и анализ сети; прогнозирование состояния сети; информационное управление.

Перечислим некоторые задачи мониторинга и анализа социальных сетей: мониторинг развития существующих и появления новых актуальных обсуждаемых информационных объектов, определение характера обсуждения (критика, одобрение, призывы к действиям и т.п.), оценка естественности обсуждений, прогнозирование роста популярности информационного объекта, выявление активных групп и сообществ, социодемографический анализ сообществ, определение инициаторов

обсуждений, оценка влиятельности пользователей, деанонимизация пользователей.

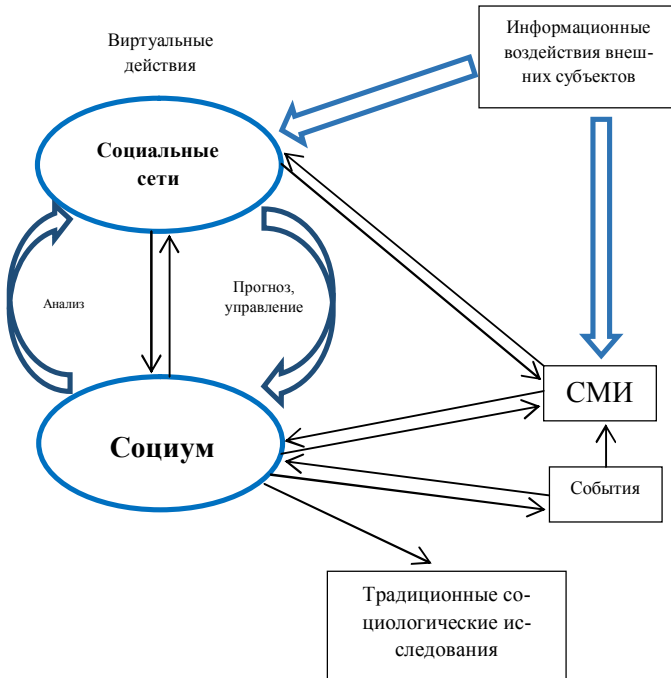


Рис. 2. Социальная сеть как объект анализа, прогнозирования и управления

Перечислим некоторые задачи прогнозирования и управления: прогнозирование развития информационных процессов, формирование общественного мнения, поддержка и повышение рейтинга виртуальных пользователей в заданных социальных группах.

6. Методы и алгоритмы анализа активности в социальной сети

В зависимости от поставленной задачи выбираются типы объектов (внешний субъект – внешний источник – сеть – агент – информационное сообщение – информационный объект – мне-

ние). Для каждого выбранного объекта составляется набор характеристик, отражающих те или иные его свойства, важные с точки зрения моделирования. Например, агенту (узлу сети) соответствует его важность с точки зрения распространения информации, сети – активность сети по отношению к информационному объекту и т.п. Показатели являются детализацией и (или) конкретизацией характеристик объекта. Например, активность сети по отношению к информационному объекту конкретизируется в виде таких показателей, как количество описывающих информационный объект информационных сообщений, суммарное количество упоминающих информационных объектов блогеров пр.

После того как определяются объекты и их характеристики, выбирается необходимый уровень анализа данных. Выделим три уровня анализа социальных сетей.

- Расчет базовых показателей, позволяющих ответить на вопросы «Как часто пишут пользователи социальных сетей?», «Сколько друзей у пользователя социальной сети?» и т.п.

- Выявление закономерностей и паттернов, позволяющих ответить на вопросы «В какой тональности пишут пользователи?», «Насколько естественно протекает информационный процесс?» и т.п.

- Нахождение конкретных ответов на специфические вопросы «Как распространяется та или иная информация?», «Как формируются то или иное мнения в сети?» и т.п.

В зависимости от поставленных вопросов необходимо осуществить анализ на том или ином уровне. Для этого используются следующие классы методов:

- методы многомерного статистического анализа (рассматривающие на макроуровне сеть и информационные процессы, протекающие в ней);

- методы анализа сетей (учитывающие структуру связей между узлами сети и, соответственно, протекающие в рамках этих связей информационные процессы).

Следует отметить, что в некоторых случаях полезно совместно использовать как методы многомерного статистического анализа, так и методы анализа сетей.

7. Технологии мониторинга и анализа социальных сетей

Для информационно-аналитической поддержки пассивной и активной работы с онлайн-социальными сетями необходима разработка автоматизированной системы мониторинга и анализа онлайн-социальных сетей.

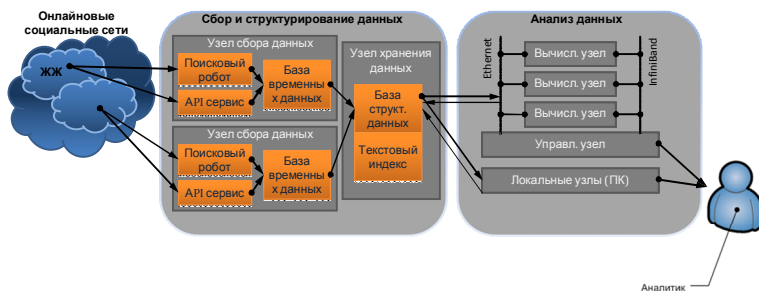


Рис. 3. Архитектура системы

Система состоит из двух больших аппаратно-программных комплексов (см. рис. 3).

Аппаратно-программный комплекс сбора и структурирования данных отвечает за сбор, структурирование и хранение первичных разнородных данных, обеспечивает эффективный доступ к ним. Комплекс состоит из узлов двух типов: узла сбора данных и узла хранения данных. Узел сбора данных осуществляет сбор данных (как при помощи поискового робота, так и при помощи программного интерфейса онлайн-социальной сети), а также их первичную структуризацию и временное хранение. Далее данные перемещаются в единое хранилище – узел хранения данных. Узел хранения данных производит их вторичную структуризацию, осуществляет резервное копирование и организует доступ к данным со стороны аналитического комплекса.

Аппаратно-программный комплекс анализа данных отвечает за анализ процессов, протекающих в социальных сетях. Комплекс ориентирован на высокопроизводительные вычисления с большими объемами входных данных. Поэтому комплекс может быть ре-

ализован на базе вычислительного кластера, узлы которого работают в режиме общей памяти. Обмен данными между узлами может осуществляться в режиме распределённой памяти.

К настоящему моменту разработан базовый вариант комплекса, осуществляющего сбор данных из русскоязычного сегмента блогерской площадки LiveJournal (<http://www.livejournal.com>). Опишем кратко системы, разработанные в рамках этого комплекса.

Система сбора данных. Сбор проводится по блогерской площадке LiveJournal, в которой пользователи публикуют посты, а также комментируют посты и комментарии других пользователей. Сбор постов осуществляется в реальном времени, сбор комментариев осуществляется регулярно (по расписанию). При этом первичные параметры постов и комментариев сохраняются в промежуточной базе метаданных, а их тексты – в файловой системе.

Система структурирования данных. По мере накопления необработанных данных на сервере их нужно актуализировать, т.е. надлежащим образом преобразовать, затем поместить в индекс и в конечную нормализованную базу данных. Такая процедура производится (при помощи специально разработанного программного обеспечения) в два этапа: актуализация данных в индексе подразумевает под собой создание текстовых индексов; актуализация постов в базе данных заключается в том, чтобы заполнить конечную базу данных, используя промежуточную базу метаданных и индекс. Индексы нужны для осуществления поиска по текстам постов и комментариев, а конечная база данных – при дальнейшей работе, для анализа онлайн-социальной сети.

Система доступа к данным. Подсистема принимает пользовательские запросы на специальном языке запросов (в частности, допускаются wildcard-символы, булевы операторы И/ИЛИ/НЕ и скобки произвольной вложенности), далее подсистема разбирает запрос и просматривает все индексы на наличие релевантных документов. В конечном итоге результаты (идентификаторы найденных документов и некоторая дополнительная информация) сохраняются в файл в определенном формате, пригодном для дальнейшего анализа.

8. Заключение

В настоящей работе определен конструктивный подход к анализу социальных сетей, рассмотрена концептуальная модель социальной сети, рассмотрены базовые задачи анализа и управления в социальных сетях, рассмотрены методы и алгоритмы анализа активности в социальных сетях, а также технологии мониторинга и анализа социальных сетей.

Перспективным направлением дальнейших теоретических и прикладных исследований является планомерное развитие и углубление методов анализа на основе сформулированной модели, а также применение этих методов для исследования реальных онлайн-социальных сетей.

Литература

1. ГУБАНОВ Д.А., НОВИКОВ Д.А., ЧХАРТИШВИЛИ А.Г. *Социальные сети: модели информационного влияния, управления и противоборства*. – М.: Изд-во физико-математической литературы, 2010. – 228 с.
2. AGGARWAL C.C. *Social Network Data Analytics*. – Kluwer Academic Publishers, 2011. – 502 p.
3. ALBERT R., BARABÁSI A.-L. *Statistical mechanics of complex networks* // *Reviews of Modern Physics*. – 2002. – Vol. 74. – P. 47–97
4. BAILEY N. *The Mathematical Theory of Infectious Diseases and Its Applications*. – New York: Hafner Press, 1975. – 413 p.
5. BALA V., GOYAL S.A. *Noncooperative Model of Network Formation* // *Econometrica*, *Econometric Society* – 2000. – Vol. 68(5). – P. 1181–1230.
6. BASS F. *A new product growth model for consumer durables* // *Management Science* – 1969. – № 15 (5). – P. 215–227.
7. ERDOS P., RENYI A. *On random graphs* // *Publicationes Mathematicae* – 1959. – №6 – P. 290–297.
8. EVERTON S.F. *Disrupting Dark Networks (Structural Analysis in the Social Sciences)*. – Cambridge University Press, 2012. – 482 p.

9. GOLDENBERG J., LIBAI B., MULLER E. *Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth* // Marketing Letters. – 2001. – №2. – P. 11–34.
10. GRANOVETTER M. *Threshold Models of Collective Behavior* // American Journal of Sociology. — 1978. – Vol. 83, №6. – P. 1420–1443.
11. NEWMAN M.E.J. *The structure and function of complex networks* // SIAM Review. – 2003. – P. 167–256. [Электронный ресурс]. – URL: <http://arxiv.org/pdf/condmat/0303516.pdf> (дата обращения 28.08.2013).
12. ROGERS E.M. *Diffusion of Innovations*. – New York, London: Free Press, 1983. – 453 p.

CONCEPTUAL APPROACH TO ONLINE SOCIAL NETWORKS ANALYSIS

Alexander Chkhartishvili, Institute of Control Sciences of RAS, Moscow, Digital Society Laboratory (DSL), Moscow, Doctor of science in mathematics (sandro_ch@mail.ru).

Dmitry Gubanov, Institute of Control Sciences of RAS, Moscow, Digital Society Laboratory (DSL), Moscow, Ph.D. (dmitry.a.g@gmail.com).

Abstract: We introduce a constructive approach to online social networks analysis. We suggest a conceptual model of a social network, describe basic problems of analysis and control in social networks, describe methods and algorithms for activity analysis in social networks, and also technologies for social networks monitoring and analysis.

Keywords: social network, conceptual model, informational control, information technologies.

*Статья представлена к публикации
членом редакционной коллегии М.В. Губко*

*Поступила в редакцию 18.12.2012.
Опубликована 30.09.2013.*