

Базовый анализ социальных графов организаций в социальных сервисах на примере МЧС России

Артюхин Валерий Викторович

к. э. н., доцент, ведущий научный сотрудник Центра анализа и управления рисками, Всероссийский научно-исследовательский институт по проблемам гражданской обороны и чрезвычайных ситуаций МЧС России,
ул. Давыдовская, 7, г. Москва, 121352, +7 (499) 233-25-40
ikshot@mail.ru

Аннотация

В данной работе автором ставится задача изучения реальных организаций на основе анализа социальных графов, построенных из связанных с организацией учетных записей в социальных сервисах. В качестве конкретного примера рассматриваются социальные графы, составленные из учетных записей *Facebook* и *Twitter*, предположительно связанных с Министерством Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий. Рассмотрены перспективы развития описанного подхода и практические задачи, решению которых он потенциально может служить.

In this article the author formulates a problem of studying of the real organizations on the basis of analysis of the social graphs made of social services accounts associated with the organization. As a concrete example it is considered the social graphs, made of the *Facebook* and *Twitter* accounts which are allegedly connected with The Ministry of the Russian Federation for Civil Defense, Emergencies and Elimination of Consequences of Natural Disasters. Prospects of development of the described approach and the practical tasks to which decision it can potentially serve are considered.

Ключевые слова

анализ социальных сетей, социальный граф, *Facebook*, *Twitter*, организация; social network analysis, social graph, *Facebook*, *Twitter*, organization.

Введение

Со словами «социальная сеть» у большинства людей ассоциируются такие громкие названия, как *Facebook*, *Twitter*, «Одноклассники», «В контакте» и другие. Однако ученые знают, что эти сервисы, на самом деле, являются всего лишь примерами воплощения мощной концепции «социальных сетей» (о чем создатели сервисов, вполне возможно, изначально не подозревали). Социальная сеть – это теоретическая конструкция, предназначенная для изучения взаимоотношений между индивидами, группами, организациями и т. д. При этом сами индивиды представлены узлами сети, а взаимоотношения между ними – ребрами или дугами между этими узлами. Предполагается, что характеристики узлов и связей между ними являются взаимозависимыми. Единицей анализа в данном случае является именно группа узлов вместе со связями между ними [1].

В данной работе мы будем употреблять термин «социальная сеть» для обозначения таких теоретических конструкций, в то время как *Facebook*, *Twitter* и тому подобные службы, позволяющие людям устанавливать и фиксировать различные виды взаимоотношений между ними (выстраивая таким образом «социальные сети»), мы будем собирательно называть «сетевыми социальными сервисами», «социальными сервисами» или просто «сервисами».

Анализ социальных сетей в качестве самостоятельной дисциплины сформировался относительно недавно и берет свое начало в исследованиях Эмиля Дюркгейма и Фердинанда Тённиса, проводимых ими во второй половине XIX века и посвященных социальным группам. В 30-х годах XX века психологи, антропологи и математики работали в данном направлении независимо. В частности собиралась и анализировалась информация о взаимоотношениях в малых группах, например учеников в классах и сотрудников в рабочих группах (Якоб Морено – психолог) и о взаимоотношениях внутри сообществ в Индии, южной Африке и Великобритании (Макс Глакман и другие антропологи), была разработана теория обмена (Джордж Хоманс, Питер Блау – социологи). В 1970-х годах теории и традиции, принятые в разных науках, стали сливаться, формируя новую дисциплину [2]. В начале 2000-х годов интерес к ней резко возрос, в частности в связи с распространением Интернета и сетевых социальных сервисов – накопился огромный объем информации в формате пригодном для машинной обработки, которую можно было анализировать и использовать в коммерции (в частности, речь идет о коллаборативной фильтрации, то есть о формировании для индивида рекомендаций на основании предпочтений связанных с ним индивидов или индивидов по тем или иным признакам схожих с ним). Однако анализ социальных сетей также применяется в исследованиях:

- специфики связей между членами сообществ, расположенных на определенных географических территориях нашей планеты;
- диффузии инноваций;
- скрытых сетей, в частности криминальных [3];
- экономической социологии;
- распространения эпидемий;
- эволюционной лингвистики (изменений в языковых системах, переносимых в другие языковые системы посредством сетей социальных взаимодействий);
- цитируемости научной литературы (какие работы цитируются чаще других, и с чем это связано);
- формальных и неформальных связей в организациях и во многих других областях.

Анализ социальных сетей может применяться и / или использовать достижения из многих областей: математики (сети обычно визуализируются в виде графов), имитационного моделирования, журналистики, маркетинга, связей с общественностью, социологии, теории коммуникации, поисковой оптимизации, психологии, коммуникации риска, системного анализа, защиты информации и т. д. – это действительно междисциплинарная теория.

В общем случае согласно следующему распространенному определению модели: «*M* моделирует объект *A*, если *M* отвечает на вопросы относительно *A*» [4] социальная сеть организации, построенная тем или иным способом (узлами могут быть сотрудники, отделы, территориально-распределенные подразделения и т. д., а дугами или ребрами – сообщения электронной почты, связи типа руководитель-подчиненный и пр.) представляет собой модель этой организации. В зависимости от способа, которым построена модель, вопросы относительно исходного объекта, на которые она может ответить, естественно, различаются. Разные социальные сети могут использоваться, чтобы выявить скрытые взаимодействия, оценить эффективность организационной структуры или какие-либо иные параметры,

выработать рекомендации для руководства. Во всех описанных случаях реальный объект является в определенном смысле фиксированным: мы изучаем реальную организацию и «списываем» с нее модель в виде социальной сети.

Одной из интересных задач, касающихся одновременно и социальных сетей, и современных сетевых социальных сервисов, представляется задача изучения реально существующего сообщества (сообщества физически существующих людей, как-либо связанных) или организации через представление таковых в социальном сервисе типа *Facebook* (хотя это может быть и любой другой подобный сервис). В этом случае мы предполагаем, что нам ничего или очень мало известно о реальном объекте. Мы фиксируем правила отбора интересующих нас узлов в сетевом сервисе (например, мы выбираем узлы, у которых в имени пользователя имеются определенные ключевые слова), собираем информацию об этих узлах, включая информацию о наличии или отсутствии связей между ними, строим социальный граф (графическое представление социальной сети) на основе этих данных и в дальнейшем пытаемся сделать определенные обоснованные заключения о реальном объекте, то есть об организации, посредством изучения этой сети. В таких обстоятельствах мы продолжаем рассматривать социальную сеть в качестве модели организации, но и организация становится, в определенном смысле, моделью сети – по уже имеющейся у нас фиксированной (если мы рассматриваем ее в статике) сети мы можем предположить (или восстановить) бесконечное количество описаний возможных организаций.

Применяя такой подход и предполагая, что мы смогли собрать достаточно информации о реальной организации и сделать верные выводы (эта работа является установочной – возможности такого подхода, как и его обоснованность пока далеко неочевидны), мы можем, в частности:

- получить представление о масштабе организации, ее информационной политике, предпочитаемых каналах распространения информации, выявить изъяны информационной политики (и другие проблемы, которые в реальности вуалируются) и сформировать рекомендации по их устранению;
- разделить учетные записи на легитимные (представляющие официальную информацию об организации) и нелегитимные (например, в случае МЧС России это имеет огромное значение – информация, предоставляемая в качестве предупреждений, в ходе чрезвычайной ситуации или ее ликвидации, может носить критический характер для жизни и здоровья людей);
- определить наиболее эффективные каналы передачи информации (с точки зрения уменьшения ее искажений, численности получателей, скорости распространения или всего перечисленного одновременно);
- выделить «главные» узлы (трактовка понятия зависит от конкретной задачи);
- в качестве гипотезы мы предполагаем возможным восстановить структуру организации – иерархию управления или территориального распределения ее подразделений, хотя, как мы увидим позже, на пути решения этой задачи стоит множество проблем, часть из которых связана со спецификой функционирования организации значительно сильнее, чем со спецификой социальных сервисов, поэтому такое восстановление структуры организации возможно не для всех организаций, и не во всех случаях, но как минимум эти причины могут быть установлены.

Разумеется, степень приближения и характер сходства между нашей «восстановленной» таким образом организацией и реально существующей изучаемой будет зависеть от конкретной задачи, а также от самой организации: ее структуры, принятых бизнес-процессов, того, насколько системно организация обеспечивает свое присутствие в социальных сервисах, что видит в качестве цели этого присутствия и т. д.

Актуальность задачи связана с тенденцией к формированию государственными структурами и коммерческими компаниями своих «представительств» в социальных сервисах. Часто это организации со сложной структурой. Предлагаемый подход позволяет руководителям организации увидеть, как построенное «представительство» выглядит извне с точки зрения стороннего наблюдателя (в анализе не используются данные, полученные изнутри организации), и если оно не соответствует ожиданиям, то скорректировать свою политику в этой области, а при ее полном отсутствии – выработать.

Помимо указанных вариантов практического применения подобная задача может с успехом использоваться и в образовательных целях в качестве группового проекта, например, для студентов-программистов: группа делится на подгруппы по организациям (или брендам, или компаниям) для исследования, и каждому члену подгруппы в качестве задания поручается составить и проанализировать граф в том или ином социальном сервисе. Позже студенты могут сравнить свои результаты и выводы. Такая задача тесно связана с информационными технологиями, позволяет освоить сразу несколько инструментальных средств (например, указанных ниже), в зависимости от конкретной формулировки может развивать навыки работы с графами, метриками, текстом, аналитические способности и, не в последнюю очередь, она может быть весьма интересна студентам, поскольку в ней присутствует некий «авантюризм».

Нельзя сказать, что эта задача абсолютно нова – по сути, она включает элементы нескольких хорошо известных задач:

- анализ социальных сетей, полученных посредством изучения взаимодействия и структуры реальных организаций [5, 6, 7];
- идентификация пользователей – обнаружение профилей, принадлежащих одному человеку, в нескольких социальных сетях;
- социальный поиск – поиск социальных объектов (пользователей, их данных, их записей и т. д.), основанный на анализе набора связей, в которых находятся искомые объекты;
- генерация рекомендаций;
- выявление «настоящих» связей – применение подхода «разведки на основе открытых источников» для выявления истинных связей между пользователями, то есть настоящих друзей, родственников и т. п.;
- сбор информации – построение социального графа на основе данных, полученных в результате обработки информации социальных веб-сервисов [8].

Основных отличий рассматриваемой задачи от перечисленных два:

- мы пытаемся отразить собранную информацию в виде социального графа, полученного из сетевого сервиса, и результаты ее обработки обратно на реальный объект;
- мы изначально нацелены на изучение организаций, а не индивидов – в социальном сервисе одновременно могут присутствовать учетные записи, соответствующие нашим критериям отбора и ведущиеся:
 - сотрудниками организаций от ее имени или от имени ее отдельного подразделения – легитимные узлы;
 - сотрудниками организации, расценивающими учетную запись, как свою собственную (публикуя личные материалы, устанавливая личные связи);
 - индивидами, которые не имеют к организации никакого отношения;
 - индивидами, по какой либо причине нацеленными на причинение вреда организации или ее репутации;
 - сотрудниками конкурирующих организаций и т. д.

В ходе исследования автор построил социальные графы подразделений МЧС России на основе данных нескольких сетевых социальных сервисов. Основной вопрос, на который должно было дать ответ исследование: можно ли что-либо

определенно сказать о структуре МЧС России и его информационной политике, исследуя социальный граф, то есть имеет ли очерченная задача (получение существенной информации о конкретной организации путем анализа ее социальной сети в сетевом социальном сервисе) предпосылки к решению.

МЧС России выбрано в качестве организации для рассмотрения отчасти потому, что оно декларирует одним из приоритетов своей деятельности максимальную открытость, в том числе и в области связей с общественностью, которые сегодня немислимы без соответствующей работы в Интернете, отчасти поскольку автор руководил подготовкой ряда аналитических документов, разработанных в интересах МЧС России и касающихся эффективности освещения (как официальными, так и неофициальными источниками) в Интернете событий, связанных с наводнением в Крымске, всероссийской тренировкой по гражданской обороне, ураганом Сэнди в США и т. д.

В данной работе рассматриваются социальные графы МЧС России в сервисах *Facebook* и *Twitter*.

В ходе исследования применялось следующее программное обеспечение:

- открытая платформа визуализации графов *Gephi* [9];
- открытый язык статистических вычислений *R* [10];
- открытая среда разработки *R Studio* [11].

В качестве методов исследования применяется визуальный анализ социальных графов и диаграмм, полученных на основе исходных данных, а также расчет и анализ простых метрик социальных сетей.

Сбор данных

В данном исследовании мы принимаем сторону постороннего неквалифицированного в сетевых исследованиях наблюдателя. Чаще всего при широкомасштабных исследованиях сбор данных, будь это система ссылок между веб-сайтами или связей между пользователями в социальных сервисах, производится с использованием программ-«пауков» (англ. *spider, crawler*). Программа просматривает страницы в Интернете (в том числе это могут быть страницы пользователей сервиса) согласно заранее определенному списку, считывает ссылки на другие страницы (профили других пользователей сервиса) и повторяет этот процесс рекурсивно, следуя установленным правилам обхода, чтобы не «заблудиться» во всемирной паутине и избежать заикливания [12].

В данном же случае сбор информации производился ручным способом по следующим причинам:

1. Это эксперимент, и прежде чем переходить к анализу тысяч и тысяч профилей и ссылок, представлялось разумным «почувствовать», если можно так выразиться, процесс во всех его деталях.

2. Автор стремился избежать лишней массы ссылок и профилей, очевидно не имеющих отношения к системе МЧС России. Исследователь-человек при взгляде на страницу пользователя «Вася МЧС», у которого нет ни одного «друга», а публикации посвящены видеоклипам сомнительного содержания, может с достаточной долей уверенности сделать вывод о том, что данный пользователь к МЧС как к системе отношения не имеет, но программа-паук представления об этом не имеет. Возможно, подобные профили при исследовании организаций следует учитывать (и к этому мы позже вернемся), но не на данном экспериментальном этапе.

3. Наш метод поиска учетных записей во многом соответствует проводимому сторонним пользователем, желающим получать информацию или общаться с официальными представителями МЧС.

Поиск и отбор сайтов и учетных записей пользователей социальных сервисов производился с помощью поисковой системы *Google*, а также встроенных механизмов поиска социальных сервисов. В каждом случае отбирались страницы / профили пользователей, найденные по ключевым словам «МЧС» и «МЧС России» (содержащие эти слова в имени пользователя). В случае поиска сайтов и веб-страниц, не относящихся к социальным сервисам, было принято определенное ограничение: ссылки собирались только с первых 20 страниц поисковой выдачи *Google* по запросу «МЧС» и 30 страниц выдачи по запросу «МЧС России» (количество страниц определялось эмпирически на основе доли нерелевантных системе МЧС ссылок на странице: в случае запроса «МЧС России», который длиннее, чем просто «МЧС», общее количество релевантных организации ссылок, очевидно, больше – более конкретный запрос в большинстве случаев дает более релевантные результаты, отсюда и большее число обработанных страниц поисковой выдачи).

Результаты проведенного таким образом сбора информации отражены в Таблица. 1. Каждый обнаруженный профиль стал узлом в социальном графе. Следует отметить, что в ходе построения социальных графов число узлов увеличивалось. Например, добавлялись профили, на которые обнаруживались множественные ссылки из уже отобранных профилей.

Таблица. 1.
Количество и типы узлов, использованных при построении социальных графов МЧС России.

Тип узлов	Кол-во узлов
Сайты, обнаруженные посредством <i>Google</i>	136
Каналы <i>YouTube</i>	2
Профили пользователей сервиса <i>Facebook</i>	49
Профили пользователей сервиса <i>Twitter</i>	92
Профили пользователей сервиса <i>LiveJournal</i>	14
Профили пользователей сервиса «ВКонтакте»	90
Профили пользователей сервиса «Одноклассники»	25

Если мы принимаем гипотезу о том, что присутствие МЧС как организации с официальными учетными записями в социальных сервисах является не случайным и не стихийным, а системным, то информация в Таблица. 1 может быть использована в качестве предварительного рейтинга социальных сервисов по уровню фактического присутствия в них МЧС России: верхние строки занимают *Twitter* и «ВКонтакте», нижние строки – «Одноклассники» и *LiveJournal*. Мы называем рейтинг «условным», потому что в данном случае не учитывается, что некоторые из обнаруженных нами записей, официально могут не иметь к МЧС отношения (более точный рейтинг можно составить на основе количества узлов, входящих в гигантские компоненты социальных графов – см. след. раздел).

Социальные графы строились отдельно по узлам сайтов и каждого сервиса (исключение составляют узлы каналов *YouTube* – в силу их малого числа, они были включены в граф сайтов).

По каждому узлу (за исключением сайтов) были собраны следующие данные (для *Facebook* – по состоянию на 25 марта 2013 г., для *Twitter* – по состоянию на 3 апреля 2013 г.):

- наименование учетной записи;
- дата последнего обновления;
- общее количество публикаций;
- для *Facebook*:
 - общее количество «друзей»;
- для *Twitter*:

- общее количество подписчиков;
- общее количество учетных записей, на публикации которых подписана данная учетная запись;
- дата первой публикации.

Связи между узлами из списка устанавливались на основе открытых списков «друзей» / подписчиков из числа других отобранных узлов.

В качестве меток узлов были использованы коды шаблона $M[\text{номер узла}]$, чтобы до определенной степени обезличить данные для анализа. Автор понимает, что при ручном сборе исходных данных заявлять о полном обезличивании или «очистке данных» было бы нелепо, однако, учитывая изыскательский характер исследования, начальный анализ графов производился именно по кодированным названиям узлов, а выводы, полученные в ходе анализа сверялись с информацией, которую можно получить, зная, что представляют собой те или иные узлы на самом деле.

Граф Facebook

Социальный граф узлов – профилей пользователей *Facebook*, содержащих в названии учетной записи слова «МЧС» или «МЧС России» изображен на Рис. 1. Названия учетных записей (имена пользователей) заменены их кодами в соответствии со схемой, описанной выше.

Каждый социальный сервис имеет свои особенности и правила, свою систему терминов, и все это накладывает отпечаток на то, что мы можем извлечь из сервиса, какую информацию получить проще, а какую – сложнее. В частности, что касается *Facebook*, то количество «друзей» пользователя может быть обнаружено на странице его хроники, там же можно посмотреть и дату последней публикации, а вот общее количество публикаций обнаружить визуально нельзя (можно сделать это программно, но в данном случае мы опирались на «ручной» анализ), их приходится пересчитывать. Кроме того, отношения «дружбы» – основной тип перманентной связи пользователей в *Facebook* (существуют также связи типа «нравится» и «поделится» в отношении конкретных публикаций) – не имеют направления. Имеется в виду, что пользователь *A* посылает запрос на добавление в «друзья» пользователю *B*, в случае получения подтверждения от пользователя *B*, оба пользователя получают нового «друга» в соответствующих списках, и с этого момента понять, кто инициировал эту связь, становится очень невозможно, можно увидеть только дату, с которой «дружба» началась. Это означает, что наш социальный граф является неориентированным.

Граф на Рис. 1 уложен в соответствии с алгоритмом Фрухтермана-Рейнгольда [13]. Граф содержит 49 вершин и 386 ребер. Узлы показаны в виде кругов, большая насыщенность заполнения круга соответствует большему числу связей (ребер) узла с другими, размер круга пропорционален количеству «друзей» соответствующего пользователя, включая и тех, которые не включены в данный граф в виде узлов, то есть не соответствуют нашим критериям отбора, размер шрифта метки узла пропорционален количеству публикаций пользователя в его «хронике».

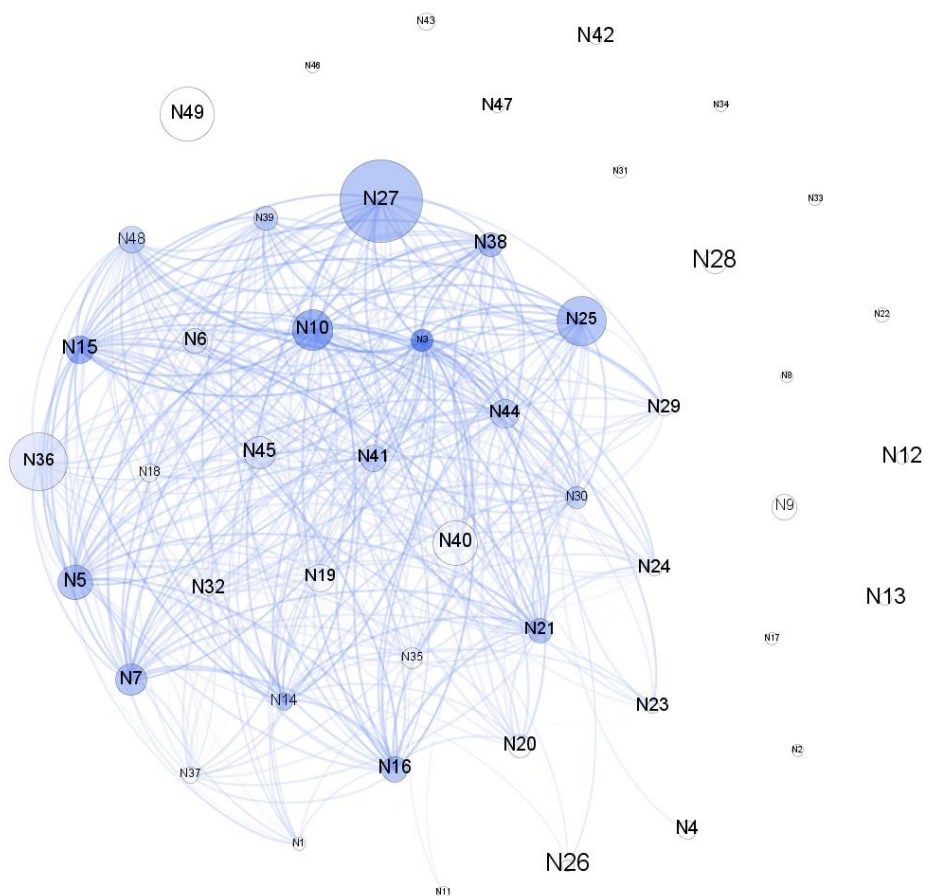


Рис. 1. Социальный граф МЧС России в социальном сетевом сервисе Facebook.

Первый очевидный результат чисто графического анализа (к рассмотрению связей между кодированными именами узлов и реальными наименованиями учетных записей мы перейдем несколько позже) – это возможность разделения графа на два подмножества: гигантский компонент (большая часть узлов графа, связанных между собой) и отдельно стоящие узлы. Что касается гигантского компонента, то если только в мире не существует другого тесно связанного сообщества людей или официальных представителей организаций, которые по непонятным причинам любят включать в названия своих учетных записей слова «МЧС» и «МЧС России», то мы имеем дело именно с учетными записями, соответствующими подразделениям рассматриваемого министерства. Наличие столь значительного количества узлов может свидетельствовать не только о развитости структуры организации, но о наличии некоторой установленной информационной политики, обязывающей или рекомендующей подразделениям обеспечивать свое присутствие в социальных сервисах. Это, вообще, важный аспект, появляющийся только при анализе социальных графов организаций и отсутствующий или могущий отсутствовать в случае анализа социальных графов индивидов: *в организациях есть правила*. Эти правила могут носить формальный или неформальный характер, характер требований или рекомендаций, они могут быть политиками, кодексами и устанавливаться в уставах, должностных инструкциях или приказах, но они имеют

место быть, и даже если мы ничего не знаем об их содержании, мы всегда с уверенностью может предположить их наличие, что может дать нам дополнительную информацию. В частности, *мы можем и должны учитывать возможность централизованного создания и управления учетными записями в социальном сервисе*. В случае с сообществами индивидов централизованное управление совершенно не имеет смысла, и никаких правил может и не быть.

Еще одной отличительной чертой организаций (наряду с обязательным наличием неких правил) является *обязательное наличие некоей структуры, которую в идеале можно выявить по социальному графу*. К сожалению, по данному графу какие-либо заключения о структуре МЧС сделать сложно: средняя степень узла (количество соединенных с ним ребер) внутри гигантского компонента составляет 22.06, что при наличии в графе всего 35 узлов, имеющих связи, означает, фактически, что мы имеем плотный клубок, где «каждый связан почти с каждым». Представляется, что *структуру организации значительно проще выявить на основе социального графа сайтов и ссылок* (причем дугам – это будет ориентированный граф, – соответствующим ссылкам, можно придать определенный вес на основе того, где эта ссылка размещена: на главной странице или в глубине сайта). Например, головной сайт организации чаще всего содержит ссылки на сайты или страницы региональных представительств, а те, в свою очередь, ссылки на центральный сайт, но необязательно, что сайты региональных представительств будут ссылаться друг на друга.

С отдельно стоящими узлами дело обстоит несколько сложнее. Учитывая все, сказанное выше, это могут быть узлы:

- находящиеся вне контроля системы МЧС, то есть не представляющие официальную позицию подразделений МЧС («Иван МЧС Рязань» и т. п.) – таким узлам следует уделить внимание, особенно, если они достаточно «общительны», то есть имеют большое число «друзей» и публикаций;
- соответствующие подразделениям МЧС, но в силу ряда причин трудно обнаруживаемые обычным поиском и системой рекомендаций Facebook – это сигнал к тому, что *нужно что-то менять в поведении людей, отвечающих за узел, либо корректировать информационную политику, если таких узлов много;*
- «умершие» или «мертворожденные» узлы, то есть узлы, соответствующие подразделениям МЧС, деятельность которых в сети характеризуется низкой активностью или отсутствием таковой (учетная запись была создана, но нет ни одной публикации) – это *может, хотя и не обязательно, свидетельствовать либо об уровне дисциплины сотрудников, ответственных за ведение учетной записи, либо об их квалификации;*
- конкретно в случае Facebook такие узлы могут соответствовать «страницам» подразделений МЧС – такие «страницы» могут быть заведены в сервисе для брендов, организаций, мест, музыкальных групп, общих идей и т. д., и они отличаются от главных страниц учетных записей, в частности, имеют другие адреса и не могут связываться «дружескими» отношениями.

Узлы нашего социального графа, соответствующие людям или организациям, не имеющим отношения к нашей организации с точки зрения высказывания официальной позиции *могут как представлять опасность* (сторонние пользователи в поисках информации от МЧС могут принять их за официальных представителей), *так и являться кандидатами для потенциального сотрудничества* (например, можно выделить блогеров, с которыми МЧС стоит установить взаимовыгодные отношения, хотя для этого необходимо также принять во внимание и содержание публикаций этих людей). *Если речь идет о большом числе узлов или учетных записей, принадлежащих сотрудникам организации, предпочитающим*

указывать место работы вместо своих реальных имени или фамилии, то это повод скорректировать условия договоров найма и правил внутреннего распорядка.

Помимо возможности создавать «страницы» вдобавок к учетным записям, дополнительные сложности в наш анализ вносит довольно навязчивая система рекомендаций «друзей», реализованная на *Facebook*. Отчасти ее наличием объясняется невозможность выделения из гигантского компонента какой-либо структуры или сообществ (то есть групп узлов, которые теснее связаны между собой, чем со всеми остальными узлами; выделение сообществ – это один из центральных моментов анализа социальных сетей). Мы, по сути, имеем на графе несколько пересекающихся систем сообществ:

- сообщество или сообщества учетных записей, рекомендованных друг другу в качестве «друзей» системой рекомендаций *Facebook*;
- сообщество учетных записей, которые должны быть связаны в силу информационной политики организации (если таковая существует и предусматривает это);
- сообщества учетных записей, связанных между собой в качестве отражения того факта, что соответствующие подразделения связаны организационно, территориально или функционально;
- сообщества учетных записей, объединенных на основе предпочтений администраторов этих записей (человеческий фактор).

На социальном графе мы *можем попытаться выделить «главный» узел*. Трактовка понятия «главный» может меняться от задачи к задаче и от организации к организации, но если принимать за основную цель присутствия организации в социальном сервисе распространение информации (и в данном случае мы ее видим именно так), то *рейтинг узла может быть определен, например, на основе некоторой комбинации показателей, соответствующих входящей степени узла, общему количеству связей в сервисе (то есть количеству «друзей»), а также количеству публикаций*. В данном случае выделить наиболее интересные узлы можно визуально, посмотрев на Рис. 1.

На приведенном графе в составе гигантского компонента визуально выделяются узлы *N10, N25, N27, N36, N40*. По совокупности признаков наиболее «общительным» узлом выглядит *N27*. Результаты визуального наблюдения подтверждаются результатами иерархической кластеризации (применялась кластеризация по методу полной связи [14], приведенными на Рис. 2. Мы можем предположить, что именно узел *N27* является главным в системе с точки зрения распространения информации и, возможно, контроля над другими узлами.

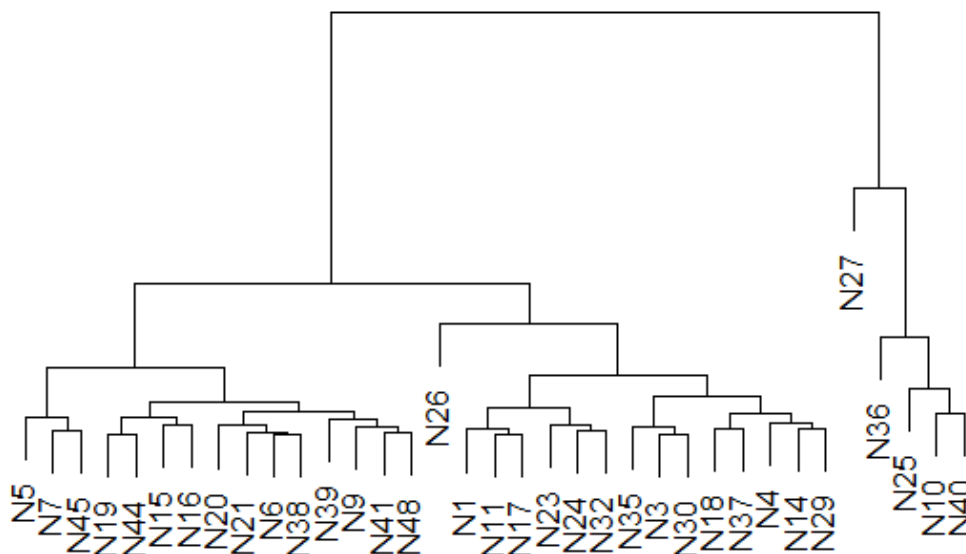


Рис. 2. Результаты кластеризации узлов гигантского компонента графа на основании степени, количества «друзей» и количества публикаций по методу полной связи.

В заключение нашего анализа декодируем имена узлов, сопоставив их с наименованиями учетных записей *Facebook*. Результаты отражены в Табл.2.

Таблица 2.

Соответствие между кодированными именами узлов и наименованиями учетных записей *Facebook*.

Код узла	Наименование учетной записи / страницы <i>Facebook</i>	URL
N1	MCHS SVAO	<i>facebook.com/profile.php?id=100005527804857</i>
N2	MCHS UR	<i>facebook.com/profile.php?id=100005274512054</i>
N3	MchsYanao GU Mchs Yanao	<i>facebook.com/MchsYanao</i>
N4	МЧС России Амурская область	<i>facebook.com/mchs.rossiamurskayaoblast</i>
N5	МЧС Астрахань	<i>facebook.com/MchsAstrakhan</i>
N6	МЧС Белгород	<i>facebook.com/vikamchs</i>
N7	МЧС Владимир	<i>facebook.com/grsmi.mchs</i>
N8	ГИМС МЧС	<i>facebook.com/gims.mchs</i>
N9	Главное управление МЧС Бурятии	<i>facebook.com/profile.php?id=100005347753748</i>
N10	ГУ МЧС Краснодарский Край	<i>facebook.com/mchs.press23</i>
N11	Главное управление МЧС НСО (МЧС по НСО)	<i>facebook.com/gumchs.novosibirsk</i>
N12	ГУ МЧС России по Камчатскому краю	<i>facebook.com/gumchskam</i>
N13	Главное управление МЧС России по Московской области (страница)	<i>facebook.com/pages/...МЧС-Росси-14332385332785</i>

Код узла	Наименование учетной записи / страницы Facebook	URL
N14	ГУ МЧС Тюменская область	facebook.com/profile.php?id=100004683186990
N15	МЧС Дагестана (Пресс-служба)	facebook.com/profile.php?id=100004389362928
N16	МЧС Ивановская область	facebook.com/press.sluzhba.mchs
N17	Казселезащита МЧС	facebook.com/kazselezashita
N18	МЧС Калининград	facebook.com/profile.php?id=100001727586514
N19	МЧС Калмыкии	facebook.com/profile.php?id=100004258923732
N20	МЧС Карачаево-Черкесия	facebook.com/mchs09kchr
N21	МЧС КБР	facebook.com/pressmchskbr
N22	КИИ МЧС	facebook.com/profile.php?id=100004499855434
N23	МЧС России Курган	facebook.com/pressa.mchs.kgn
N24	МЧС Липецкая область	facebook.com/profile.php?id=100004651293211
N25	Mchsmedia Ru	facebook.com/mchsmedia.ru
N26	МЧС Москвы (страница)	facebook.com/mosmchs
N27	МЧС России	facebook.com/mchs.press
N28	МЧС России (страница)	facebook.com/MchsRussia
N29	МЧС Рязань	facebook.com/mchs.ryazan
N30	МЧС Сибирь	facebook.com/src.mchsrif
N31	ОГПС МЧС	facebook.com/ogps.mchs
N32	МЧС Орёл	facebook.com/analitikmchs
N33	Пресс-служба МЧС РК	facebook.com/profile.php?id=100004868102862
N34	Пресс-служба МЧС РК	facebook.com/profile.php?id=100005487725369
N35	Пресс Якутия	facebook.com/emecom.press
N36	Пресс-служба МЧС Москвы	facebook.com/mchs.moskvy
N37	МЧС РО	facebook.com/mchsro
N38	МЧС Саратовской области	facebook.com/sarmchs
N39	МЧС Свердловская область	facebook.com/profile.php?id=100001661372955
N40	Сотрудник МЧС	facebook.com/sotrudnik.mchs
N41	МЧС Тульская область	facebook.com/pr.mchs.tula
N42	Уральский региональный центр МЧС России	facebook.com/uralmchs
N43	ЦГИМС МЧС по АК	facebook.com/tsgims.mchspoak
N44	Центр Лидер	facebook.com/csoor
N45	ЦРЦ МЧС России	facebook.com/pressa.crc
N46	ЦУКС Хабаровск МЧС	facebook.com/tsuxhabarovsk.mchs
N47	МЧС России по Челябинску	facebook.com/vg.centr

Код узла	Наименование учетной записи / страницы Facebook	URL
N48	ЮРЦ МЧС России	facebook.com/yurts.mchsrossii
N49	МЧС России Ярославль	facebook.com/Mchs.yaroslavl76

Большая часть выбранных нами узлов и все узлы, составляющие гигантский компонент графа, действительно являются подразделениями МЧС России. В основном это региональные центры и главные управления по федеральным округам и субъектам РФ.

Количество соответствующих записей, а также имена учетных записей не позволяют сделать заключение о существовании определенной на высшем уровне политики организации в отношении использования *Facebook*, особенно, если сравнивать эти данные с аналогичными данными по сетевому сервису «ВКонтакте», где и территориальные органы представлены значительно более широко и системно, и схема именования соответствующих учетных записей отчетливо прослеживается (в данном случае идет речь об информации, полученной в ходе составления списков узлов для последующего построения социального графа: многие узлы имеют схожие по формату названия, например «Пресс-Служба-Гу-Мчс По-Воронежской-Области», «Пресс-Служба Гу-Мчс-России-По-Курганской-Обла», «Уральский региональный центр МЧС России», «Северо-Кавказский региональный центр МЧС России» и т. д.).

В целом верны наши заключения о невозможности различить структуру организации по данным учетных записей *Facebook* и о природе отдельно стоящих узлов.

В таблице отмечены узлы, выбранные нами в качестве наиболее влиятельных на основе результатов графического анализа, подтвержденного результатами иерархической кластеризации. Подразделения МЧС, соответствующие четырем из пяти отобранных узлов непосредственно связаны с распространением информации («Сотрудник МЧС»), на самом деле, также является частью системы МЧС, присутствует в большинстве рассмотренных социальных сервисов, судя по содержанию публикаций и реплик обладает внутренней информацией, кроме того, в качестве адреса домашнего сайта в *Twitter* у него указан <http://mchs.gov.ru> – это официальный сайт МЧС России). Страница Главного управления МЧС России по Краснодарскому краю попала в список влиятельных либо благодаря интенсивной работе отвечающих за нее сотрудников, либо из-за всплеска общественного интереса к региону в связи со стихийным бедствием в г. Крымск в 2012 году, либо в силу обеих причин. Также интерес представляет узел N44, соответствующий учетной записи ФГКУ «294 Центр по проведению спасательных операций особого риска» – хотя данная учетная запись не была обнаружена на этапе первоначального составления списков учетных записей (в ее имени «МЧС» не содержится), узел «всплыл» в графе в виде составляющей гигантского компонента, поскольку среди других составляющих оказалось значительное количество его «друзей» (и, действительно, «Центр Лидер» является организацией МЧС России, одним из спасательных воинских подразделений).

Социальный граф Twitter

В отличие от социального сервиса *Facebook* основной тип перманентной связи между учетными записями в *Twitter* – «подписка» – является направленным. Если пользователь *A* подписан на сообщения пользователя *B*, то (хотя взаимная подписка – явление довольно распространенное) это не обязательно означает, что и *B* получает сообщения *A*. Это означает, что в данном случае мы можем построить

ориентированный граф, где дугами будут связаны текущие узлы (в качестве источника) и все узлы, на которые он подписан (в качестве приемников).

Очевидное благо социальных сервисов – это система рекомендаций контактов: пользователю-индивиду она позволяет найти и связаться с людьми, которых он, возможно, знает в реальной жизни, или которые могут быть ему полезны, но о присутствии которых в сервисе он не предполагал. К сожалению эта же система создает значительные проблемы для анализа – невозможно достоверно определить, почему установлена связь одного узла с другим: потому что оператор учетной записи посчитал, что это имеет смысл в силу причин, связанных со структурой организации или функций соответствующего подразделения, или потому что указанный оператор просто согласился с предложениями системы рекомендаций. Первая причина имеет отношение к нашему анализу и связана с деятельностью организации, вторая – нет.

Мы не можем полностью обойти систему рекомендаций или полностью нивелировать ее влияние на наш граф, устранив нерелевантные связи, однако мы можем учесть их наличие, придав дугам веса, в зависимости от исходящей степени узла-источника. Назначим каждому узлу одинаковый исходный вес и будем считать, что он разделяется поровну между всеми исходящими дугами. Логика этого подхода такова: чем больше дуг соединяет узел с другими узлами внутри графа, тем больше вероятность того, что оператор соответствующей учетной записи подписался на сообщения от других записей, исходя не из неких конструктивных соображений, а согласившись с рекомендациями сервиса, чем меньше таких дуг, тем больше вероятность того, что учетные записи для подписки были выбраны вручную на основании некой логики.

Получившийся социальный граф изображен на Рис. 3, он состоит из 111 узлов и 1904 дуги. Большая насыщенность заполнения кругов, представляющих узлы на графе, в данном случае соответствует большей входящей степени узла (программа *Gephi*, с помощью которой было получено изображение автоматически масштабирует толщину дуги на графе в зависимости от ее веса, и некоторые дуги имеют столь малый вес, что они не видны, но об их наличии можно судить по интенсивности заполнения соответствующего круга).

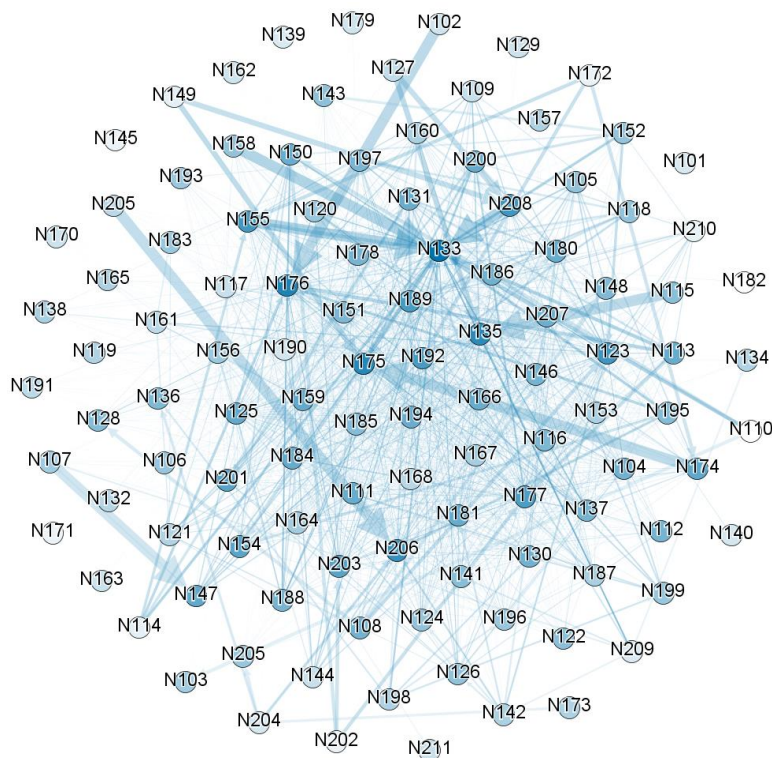


Рис. 3. Социальный граф МЧС России в социальном сетевом сервисе Twitter.

В данном случае лидерами по взвешенной входящей степени являются узлы *N133*, *N175* и *N176*. Узлы *N133* и *N176* соответствуют уже обнаруженным нами ранее подразделениям МЧС России непосредственно ответственным за распространение информации (это «Пресс-служба МЧС России» и «Объединенная редакция МЧС России»), а вот третий лидер в *Twitter* иной, нежели в *Facebook* – это некий «Пожарный спасатель» (https://twitter.com/Ilya_Malikov), который пользуется среди наших учетных записей значительной популярностью, и сам имеет множество исходящих соединений. Интересно, что это единственный случай из всех рассмотренных нами, когда соответствующая узлу учетная запись не принадлежит подразделению МЧС России, являясь, тем не менее, частью гигантского компонента графа. Иными словами, мы имеем определенную аномалию, намеренно порожденную неаффилированным субъектом, а именно, если верить информации «ВКонтакте», Ильёй Андреевичем Маликовым, предпринимателем, имеющим некое отношение к «САМОСПАСу» – негосударственному учреждению по делам эвакуации, спасения и пожарной безопасности (<http://www.samospas.ru/about/>) и, таким образом, по сути, ведущим (грамотную) маркетинговую кампанию.

В случае графа *Twitter* мы имеем плотный клубок из узлов и дуг, похожий на ранее рассмотренный граф *Facebook*, однако ориентированность первого и тот факт, что мы придали дугам вес, может помочь понять, что в этом клубке происходит. Действительно, если мы считаем дуги с малым весом с большей вероятностью порожденными исключительно на основе рекомендаций сервиса, а не на основе неких здравых рассуждений, то можем задать для веса пороговое значение и не учитывать на графе дуги с весом меньшим его. Если мы установим пороговое значение веса величиной в 1, то есть будем учитывать только те дуги, которые являются единственными для своих источников, мы получим:

- что «Главное управление (ГУ) МЧС России по Тюменской области» (в этом списке указываются реальные наименования подразделений МЧС

России) считает для себя авторитетным и необходимым для ознакомления источником «ГУ МЧС России по Ульяновской области» (виной этому название учетной записи последнего – «Пресс-служба МЧС»);

- такие же отношения связывают «ГУ МЧС России по Амурской области» и «ГУ МЧС России по Еврейской автономной области» (в этом есть логика – эти области граничат);

- в аналогичных отношениях находятся «ГУ МЧС России по Республике Калмыкия» и «Южный региональный центр МЧС России» (Республика Калмыкия действительно входит в состав Южного федерального округа);

- несколько региональных центров и главных управлений подписаны только на публикации «главных» узлов, выделенных нами ранее.

Снизив пороговое значение до 0.33, мы получим дополнительную информацию:

- «ГУ МЧС России по Саратовской области» подписано на публикации «ГУ МЧС России по Тверской области», «ГУ МЧС России по г. Москве» и «Приволжского регионального центра» (область входит в соответствующий федеральный округ);

- «ГУ МЧС России по Ленинградской области» подписано на публикации «Пресс-службы МЧС России», «ГУ МЧС России по Санкт-Петербургу» и «Северо-западного регионального центра» (область входит в соответствующий федеральный округ);

- «ГУ МЧС России по Кемеровской области» подписано на публикации «ГУ МЧС России по Республике Удмуртия», «МЧС Республики Беларусь» и «Приволжского регионального центра» (республика входит в соответствующий федеральный округ) и т. д.

Как мы видим, *действительно возможно получить из социального графа некоторую информацию о структуре организации*: 6 из 12 дуг со значением веса равным 1, 0 из 4 дуг со значением веса равным 0.5, 8 из 18 дуг со значением веса 0.33 и 9 из 16 со значением веса равным 0.25 отражают параллельные, подчинительные или территориальные связи, которые имеются между соответствующими подразделениями МЧС России в реальности. Существует обоснованное предположение, что по мере смягчения ограничения снизу на вес учитываемых дуг мы будем получать из графа все больше информации, которая одновременно будет все менее надежной с точки зрения отражения структуры организации, но это предмет отдельного исследования.

Анализ усложняет не только система рекомендации контактов, влияние которой мы до определенной степени снизили посредством введения весов, но и «человеческий фактор». Дополнительно на основе анализа графа *Twitter можно выделить три основных шаблона поведения операторов учетных записей*:

- расценивание учетной записи в качестве своей личной и официальной в равной степени – с точки зрения общего списка подписки характеризуется большим числом позиций, не относящихся к МЧС России (этот шаблон может комбинироваться как со двумя другими);

- выбор для подписки исключительно учетных записей, организационно или территориально связанных с данным подразделением МЧС России;

- подписка на все или многие учетные записи, найденные по ключевым словам «МЧС» или «МЧС России» – фактически, это метод работы, примененный нами для поиска кандидатов в узлы социальных графов.

В части информационной политики МЧС России в отношении социальных сетей также можно сделать ряд заключений:

- она подразумевает целенаправленное создание учетных записей подразделений в социальных сетях;

- она не предусматривает схемы именования таких учетных записей (как минимум это можно с уверенностью утверждать в отношении *Facebook* и *Twitter*);
- она не регламентирует содержание публикаций;
- она не регламентирует списки подписки для учетных записей подразделений и, вероятно, не содержит рекомендаций на этот счет.

Заключение

Данная работа является первой в серии работ автора, посвященных методологии анализа организаций посредством анализа социальных графов в социальных сервисах. В ней показано, что можно получить большой объем информации об организации, зная, как в данном случае, только ее название. Непосредственно в данной работе, анализируя социальные графы, построенные из учетных записей, предположительно имеющих отношение к МЧС России, в сервисах *Facebook* и *Twitter*:

- был получен подграф учетных записей подразделений МЧС России в виде гигантского компонента графа *Facebook*;
- кратко описаны особенности социальных сервисов, упрощающих и усложняющих анализ;
- выделены «главные» узлы графов *Twitter* и *Facebook* (разными способами – на основе визуального анализа, с помощью кластеризации и на основе упорядочения по взвешенной входящей степени узлов), выделенные узлы оказались информационными центрами МЧС России;
- выделена аномалия на графе *Twitter* в виде узла в гигантском компоненте, не соответствующего подразделению МЧС России;
- получена частичная информация о структуре МЧС России посредством анализа социального графа *Twitter*;
- выделены шаблоны поведения операторов учетных записей;
- сформулированы обоснованные (на основе анализа графов) предположения о характере и содержании информационной политики МЧС России в области социальных сервисов.

В данном случае в качестве примера рассматривалось МЧС России – достаточно открытое государственное ведомство, но подобный анализ может проводиться и для других организаций и компаний, каким-то образом обеспечивающих свое присутствие в социальных сервисах. По сути, это вариант разведки с применением открытых источников, но со своей спецификой (сбор информации на основе минимальных знаний об организации, например, на основе ее названия, а также априорные предположения о наличии у организации структуры и некоторых политик, требований, правил). В данной области автором планируются значительные изыскания, включая:

- попытки применения известных метрик и алгоритмов и, возможно, разработки новых, для получения большего объема более конкретной информации о реальных организациях на основе анализа социальных графов;
- применение знаний, накопленных в других областях, таких как теория коммуникации, теория организаций, теория систем и др. [15];
- разработку собственных и настройку существующих открытых программных средств для автоматизации сбора информации из социальных сервисов и т. д.

В будущем предстоит формализовать все этапы подобных исследований от выбора метода поиска узлов и дуг/ребер для построения социального графа до интерпретации результатов графического и прочих компонент анализа. Предстоит

определимся с критериями качества исследования. Также в данной работе и в исследовании в целом не рассматривались такие потенциально важные аспекты, как содержание публикаций и изменения социального графа в динамике. Для выполнения таких более сложных исследований, необходимо получать из социального сервиса дополнительные данные, такие как даты регистрации учетных записей, даты и время публикаций, местоположение (оно может быть необходимо для корректировки времени), количество «друзей» в разные моменты времени, даты начала «дружеских» отношений и т. д. Безусловно, здесь открываются богатейшие возможности для изысканий.

Потенциально описанный в работе подход совместно с современными программными средствами может применяться:

- для получения представления о структуре организации;
- получения представления об информационной политике организации (как и по многим позициям ниже, конкретность этого представления и его полнота будут различаться при рассмотрении социальных графов, полученных из разных социальных сервисов, а также от организации к организации);
- построения модели роста данного социального графа при соблюдении организацией текущей информационной политики с учетом особенностей рассматриваемого социального сервиса [16, 17,18];
- обнаружения «главного» узла или совокупности таковых на основе определенной функции от имеющихся данных (например, в виде линейной комбинации количества сообщений, контактов внутри и вне графа), зависящей от конкретной задачи;
- обнаружения потенциально опасных для организации пользователей с точки зрения распространяемой информации и схожести по реквизитам учетных записей с подразделениями организации (проще говоря, это те пользователи, которых посторонний человек, пытающийся получить официальную информацию от нашей организации или связаться с ее представителями, может принять за этих официальных представителей, хотя они таковыми не являются);
- обнаружения кандидатов для сотрудничества (например, в части распространения информации это могут быть блогеры, с которыми стоит установить взаимовыгодные отношения);
- подтверждение или опровержение информации субъекта – частного лица или представителя организации о тесном сотрудничестве между эти частным лицом или организацией с другой организацией;
- повышения эффективности присутствия организации в сетевом сервисе (например, за счет выявления узлов, соответствующих подразделениям организации, но в силу тех или иных причине редко обнаруживаемых сторонними пользователями);
- выявления подразделений организации, уклоняющихся от следования общей информационной политике;
- выбора наиболее эффективных каналов распространения информации в виде социальных сервисов;
- обоснованной корректировки информационной политики.

Литература

1. Wasserman S. и Faust K. Social Network Analysis. Cambridge: Cambridge University Press, 1994. – 825 p.

2. Social Network / Wikipedia. 2013. URL: http://en.wikipedia.org/wiki/Social_network (дата обращения: 26.03.2013).
3. Baker Wayne E., Faulkner Robert R. The Social Organization of Conspiracy: Illegal Networks in the Heavy Electrical Equipment Industry // *American Sociological Review*. 1993. Vol 6, № 56. – P. 837-860.
4. Анфилатов В. С., Емельянов А. А., Кукушкин А. А. Системный анализ в управлении. – М.: Финансы и статистика, 2002. – С. 41-43.
5. Brzozowski Michael J. WaterCooler: Exploring an organization through enterprise social media / Hewlett Packard. 13.05.2009. URL: <http://www.hpl.hp.com/research/scl/papers/watercooler/group2009/> (дата обращения: 29.03.2013).
6. Sozen Cenk, Basim Nejat, Hazir Koksal. Social Network Analysis in Organizational Studies // *International Journal of Business and Management*. 2009. Vol. 1, № 1. – P. 21-35.
7. Zack Michael H. Researching Organizational Systems using Social Network Analysis // *Proceedings of the 33rd Hawaii International Conference on System Sciences*. 2000.
8. Социальный граф / Википедия. 17.03.2013. URL: http://ru.wikipedia.org/wiki/Социальный_граф (дата обращения: 26.03.2013).
9. Gephi / Gephi.org. 2012. URL: <https://gephi.org/> (дата обращения: 29.03.2013).
10. What is R? / r-project.org. 2013. URL: <http://www.r-project.org/> (дата обращения: 29.03.2013).
11. Welcome to RStudio / RStudio, Inc. 2013. URL: <http://www.rstudio.com/> (дата обращения: 29.03.2013).
12. Ye Shaozhi, Lang Juan, Wu Felix. Crawling Online Social Graphs // *12th International Asia-Pacific Web Conference*. Busan, Korea, 2010.
13. Fruchterman T. M. J., Reingold E. M. Graph Drawing by Force-Directed Placement // *Software: Practice and Experience*. 1993. Vol 21, № 11. – P. 1129-1164.
14. Сегаран Т. Программируем коллективный разум. – СПб.: Символ-Плюс, 2008. – С. 50-76.
15. Буханцева Н. В. Методология исследования информационного пространства // *Международный электронный журнал "Образовательные технологии и общество (Educational Technology & Society)"*. 2011. № 2. – С. 373-381. ISSN 1436-4522. URL: http://ifets.ieee.org/russian/depository/v14_i2/html/8r.htm
16. Артюхин В. В. CASE-технологии создания имитационных моделей в среде PiIgrim5 // *Прикладная информатика*. 2007. № 4. – С. 52-82.
17. Артюхин В. В. Постановка задач для имитационного моделирования. Ключевые аспекты // *Актуальные инновационные исследования: наука и практика*. 2011. № 3. – С. 93–99.
18. Патаракин Е. Д., Ярмахов Б. Б. Моделирование организационных отношений с использованием «связей» NetLogo // *Международный электронный журнал "Образовательные технологии и общество (Educational Technology & Society)"*. 2009. № 2. – С. 409–422. ISSN 1436-4522. URL: http://ifets.ieee.org/russian/depository/v12_i2/html/8.htm