

УДК 519.87 + 004.94 + 004.89

ББК 22.18

ОБЗОР ИНФОРМАЦИОННЫХ СИСТЕМ АНАЛИЗА СОЦИАЛЬНЫХ СЕТЕЙ¹

Базенков Н. И.², Губанов Д. А.³

(ФГБУН Институт проблем управления РАН, Москва)

Статья содержит обзор современных систем анализа социальных сетей. Приводится классификация пользователей таких систем и задач, которые они решают. Рассматриваются системы, используемые обычными пользователями, коммерческими структурами, органами государственной власти и силовыми структурами, а также научными организациями. Большинство рассмотренных систем предназначено для мониторинга и анализа социальных сетей и не затрагивают вопросы прогноза и управления. В конце статьи приведен обзор прикладных исследований социальных сетей.

Ключевые слова: социальная сеть, интеллектуальный анализ данных, информационные технологии.

1. Введение

Основными задачами информационно-аналитической поддержки пассивной и активной работы с социальными сетями являются (рис. 1): мониторинг и анализ социальных сетей (для достижения понимания происходящих в социальных сетях процессов), прогнозирование и управление (для перевода социальной сети в требуемое состояние).

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

² Николай Ильич Базенков, аспирант (n.bazenzov@gmail.com)

³ Дмитрий Алексеевич Губанов, к.т.н., с.н.с. (dmitry.a.g@gmail.com)

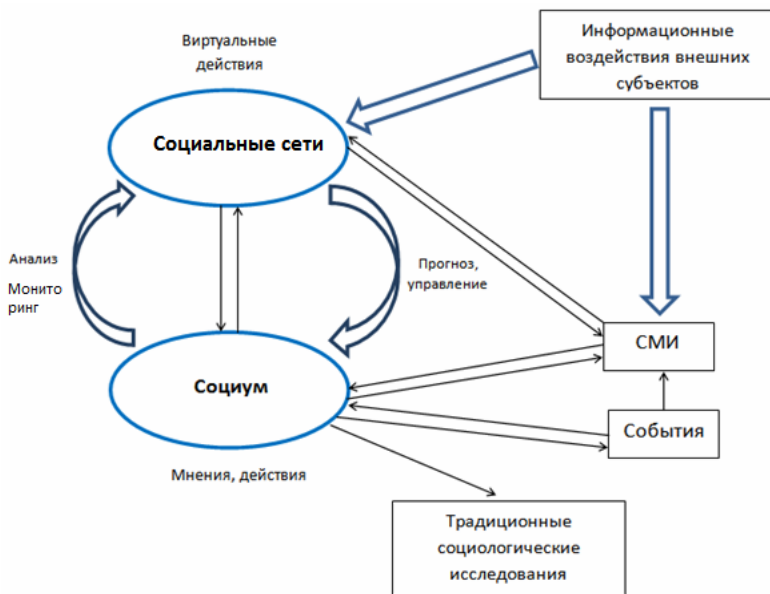


Рис. 1. Социальная сеть как объект анализа, прогнозирования и управления

Мониторинг включает получение и структурирование первичных данных. Собираются тексты сообщений, связи между пользователями, ссылки на внешние ресурсы. Возможности системы во многом определяются богатством используемых данных и режимом их обработки. Системы, поддерживающие мониторинг в режиме реального времени, сложнее в реализации, чем использующие ретроспективный сбор данных.

Анализ подразумевает несколько этапов обработки первичных данных. Во-первых, вычисление базовых показателей, которое позволяет отвечать на простые количественные вопросы типа «сколько сообщений написал пользователь А?». Далее выявление статистических и структурных закономерностей в данных дает понимание природы исследуемой сети. Например, типы распределений, к которым относятся обсуждения тех или иных тем. С точки зрения практических приложений наиболь-

ший интерес представляет выявление специфических закономерностей в узких предметных обсуждениях.

Прогноз возможен после идентификации математической модели информационного процесса. Могут использоваться статистические модели и модели динамических процессов на графах (распространение эпидемий, каскадное поведение).

Управление заключается в оказании целенаправленных воздействий на социальную сеть для перевода информационных процессов в желаемое состояние. Возможны «мягкие» качественные рекомендации пользователю и «жесткие» количественные оценки необходимых воздействий. Количественные рекомендации требуют формальных моделей управляемых процессов. Авторам не известны системы, предоставляющие такую возможность.

Задачи анализа, прогнозирования и управления могут быть разными, в первую очередь, в зависимости от того, кто ставит задачу, т.е. кто является конечным пользователем системы. Можно выделить следующие типы конечных пользователей, заинтересованных в анализе, прогнозировании и управлении социальными сетями (рис. 2):

1. Органы государственной власти и местного самоуправления.
2. Предприятия государственного и частного сектора экономики, в том числе
 - коммерческие организации (в первую очередь, «брендовые»);
 - исследовательские организации;
 - средства массовой информации;
 - силовые структуры.
3. Общество, в том числе
 - политические партии;
 - отдельные физические лица.

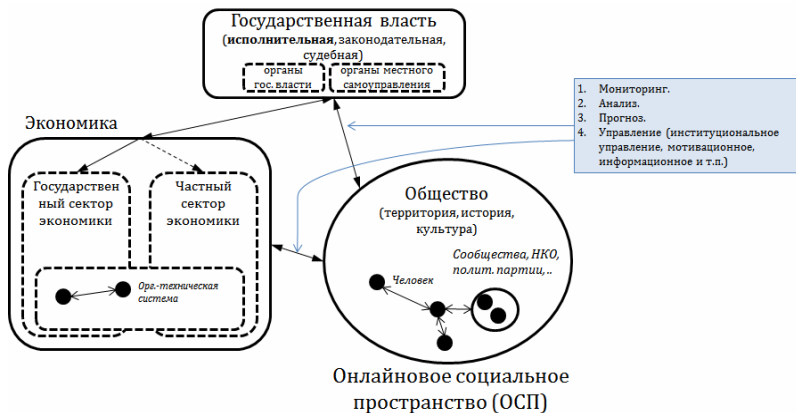


Рис. 2. Управляющие субъекты и взаимосвязи между ними

На данный момент, по нашим наблюдениям, наиболее развиты системы анализа социальных сетей для коммерческих организаций. Однако, независимо от конечных пользователей, системы анализа социальных сетей можно классифицировать по следующим основаниям.

1. Уровни анализа социальных сетей.

Системы могут осуществлять простой мониторинг социальных сетей, анализ социальных сетей, прогнозирование процессов в социальных сетях, управление социальными сетями.

2. Модели социальных сетей.

В системах могут быть реализованы те или иные модели социальных сетей: модели структуры сетей (модели случайных графов, модели безмасштабных сетей) [8, 4], модели распространения информации (марковские модели, конечные автоматы, модели диффузии инноваций, модели заражения) и др. [2]. Зачастую на практике системы либо не основываются на каких-либо математических моделях, либо не предоставляют информацию об их использовании.

3. Методы анализа данных.

Можно выделить следующие основные классы методов, используемых в системах анализа социальных сетей: статистические методы и методы анализа графов. Для классификации систем также удобно выделить отдельно методы семантического

анализа и анализа тональности текстов (необходимо обратить внимание на поддержку системой языка, на котором общаются пользователи анализируемой онлайн-социальной сети).

4. Объекты анализа социальных сетей.

Системы могут фокусироваться на анализе следующих объектов социальной сети: сети «в целом» (при помощи некоторых агрегированных глобальных показателей); подсетей и сообществ; отдельно взятых пользователей (при помощи актор-ориентированных локальных показателей); информационных сообщений (при помощи показателей упоминания некоторых информационных объектов); мнений (при помощи показателей тональности сообщения относительно некоторых информационных объектов); внешних узлов (или информационных ресурсов сети Интернет).

Примечание: информационным объектом может быть некоторая персона, событие, организация и т.п.

5. Режимы анализа данных.

Системы могут не предоставлять возможность анализа данных (отсутствует), или предоставлять возможность ретроспективного анализа данных и/или анализа данных в режиме реального времени.

6. Режимы сбора данных.

Системы могут не предоставлять возможность сбора данных (отсутствует) или предоставлять возможность ретроспективного сбора данных и/или сбора данных в режиме реального времени. Системы могут осуществлять сбор всего объема данных или осуществлять сбор данных по определенной тематике.

7. Охват источников данных.

Системы могут собирать и анализировать данные классических онлайн-социальных сетей (*Facebook, VKontakte*), блогов (*LiveJournal*), микроблогов (*Twitter*), сервисов обмена фотографиями и видео (*YouTube, Flickr*), форумов и т.п.

8. Объемы обрабатываемых данных.

Системы могут быть рассчитаны на модельные объемы данных или на промышленные объемы данных (программные продукты *BigData* [7]).

Далее будут рассмотрены системы анализа онлайн-социальных сетей для конкретных вышеупомянутых типов конеч-

ных пользователей (см. соответствующие описания соответствующих разделов ниже). Системы будут классифицированы согласно приведенным выше основаниям.

Структура настоящего обзора следующая: в разделе 2 рассматриваются массовые системы анализа социальных сетей; в разделе 3 рассматриваются системы для коммерческих организаций; в разделе 4 рассматриваются системы, предназначенные для использования органами государственной власти и силовыми структурами; в разделе 5 рассматриваются исследовательские разработки для анализа социальных сетей. В разделе 6 приведены примеры экспериментальных исследований социальных сетей.

2. Массовые системы анализа социальных сетей

Конечным пользователем системы является интернет-пользователь, использующий ее для получения представления об интересующем информационном объекте (сущности, событии, персоне и т.п.). Массовыми системами являются:

1. **Поисковые системы:** поиск в Твиттере – search.twitter.com; поиск в блогах – blogsearch.google.com; поиск людей в социальных сетях – people.yandex.ru.

2. **Уведомляющие системы,** использующие поисковые машины для поиска документов по запросу пользователя и регулярно отсылающие наиболее релевантные результаты поиска на электронную почту пользователя: *GoogleAlerts* (работает поверх поисковой машины *Google*), *Twilert* (работает поверх поисковой машины *Twitter*).

3. **Простые агрегирующие системы:** «Пульс блогосферы» Яндекса, *GoogleTrends*.

4. **Системы сбора информации** по заданному множеству источников (*RSS*-подписка): *GoogleReader*.

5. **Системы сбора и объединения информации** из различных источников («мэшапы» данных): *Yahoo! Pipes*.

Основные характеристики массовой системы приведены в таблице 1 на примере системы «Пульс блогосферы» Яндекса.

Таблица 1. Унифицированное (в рамках используемой системы классификаций) описание системы «Пульс блогосферы» Яндекса

Вендор	Компания Яндекс (Россия)
Сайт	blogs.yandex.ru/pulse
Пользователи	Интернет-пользователи
Уровень анализа данных	Мониторинг с элементами первичного анализа
Методы анализа	Базовые методы анализа текстов (поддерживается в том числе русский язык)
Объекты анализа социальных сетей	Сеть в целом, информационные сообщения
Режим анализа	Ретроспективный анализ
Объемы обрабатываемых данных	Большие (десятки и сотни миллионов информационных сообщений)
Сбор данных	Данные предоставляется другими сервисами Яндекса
Охват источников данных	Совокупность источников, индексируемых Яндексом: социальные сети (ВКонтакте), блоги (Живой Журнал, Твиттер), форумы и т.д.
Клиенты	Используется широкими кругами интернет-пользователей для получения первичного представления об упоминаемости информационных объектов за выбранный период времени

Достоинства массовых систем – их доступность, низкие требования к квалификации пользователя; недостатки массовых систем заключаются в слабых возможностях анализа. Такие системы, в сущности, предназначены для получения при помощи методов информационного поиска приближенного представления об информационных процессах, протекающих в социальных сетях.

3. Системы мониторинга и анализа социальных сетей для коммерческих организаций

Конечным пользователем рассматриваемого класса систем обычно является коммерческая организация (впрочем, в той или иной степени такими системами могут воспользоваться и другие организации, например, органы государственной власти).

Система используется для решения задач как внутри самой организации (интранет), так и за ее пределами (интернет).

Задачи внутри организации (социальная сеть организации либо уже явно определена, либо идентифицируется каким-то образом):

- комплексная оценка персонала;
- управление информационными потоками в социальной сети организации;
- сокращение вертикального разрыва между сотрудниками; формирование и сплочение команд;
- реорганизация бизнес-процессов.

За пределами организации:

- исследование рынка (анализ присутствия в социальных сетях брендов и их решений, анализ мнений пользователей о брендах и решениях, определение характеристик целевой аудитории, определение ключевых пользователей и площадок, выявление негативной волны обсуждений);
- поддержание связей с партнерами и клиентами (обратная связь);
- повышение информированности пользователей о компании (бренде) и предлагаемых ею решениях;
- повышение репутации бренда и продвижение решений (размещение информационных материалов в ключевых точках сети, постоянное участие);
- оценка результативности воздействий, оказанных на социальные сети (определение и интерпретация ключевых показателей эффективности);
- прогнозирование исходов событий (величины продаж решений).

В данном обзоре рассматриваются задачи, ориентированные за пределы организации. На данный момент на рынке наиболее развитыми являются системы, направленные прежде всего на управление взаимоотношениями с клиентами (можно считать такие системы социально-сетевыми *CRM*, на подходе системы социально-сетевой аналитики (наподобие *business intelligence*)). Обобщенная схема такой системы представлена на рис. 3.

В целом существующие на рынке системы могут предоставлять следующие возможности:

Мониторинг упоминания брендов. Лица, принимающие решения, могут использовать такие системы для отслеживания упоминаний брендов и продуктов в онлайн-социальных сетях; отслеживание ведется при помощи заданных пользователем ключевых слов. Соответствующей функцией обладают системы *Radian 6*, *VisibleTechnologies*, *Techrify/Alterian*, *Buzzmetrics*, *Symfony* и другие.

Определение рыночных рисков и возможностей (включают в себя функционал уведомляющих системы (см. раздел о массовых системах выше) и специализированных систем мониторинга). Лица, принимающие решения, могут использовать системы с данным функционалом для поиска важных обсуждений в социальных сетях, которые могут помочь определить перспективные направления развития организации (возможности) либо могут стать резонансными и тем самым повлиять на развитие организации (риски).

Веб-аналитика. Лица, принимающие решения, могут использовать такие системы для отслеживания и анализа поведения пользователей на собственных ресурсах. Примеры систем: *Omniure*, *CoreMetrics*, *WebTrends*, *GoogleAnalytics*, *YahooAnalytics*.



Рис. 3. Схема процесса управления взаимоотношениями с пользователями социальных сетей

Поддержка работы в онлайн-социальных сетях (социально-сетевые системы управления взаимоотношениями с клиентами, см. рис. 3). При помощи таких систем пользователи могут завести учетные записи в онлайн-социальных сетях и осуществлять диалог с клиентами в режиме реального времени. Наиболее развитые системы дают возможность вводить профили клиентов с социальной составляющей (поведение пользователей, местоположение пользователей, предпочтения пользователей и т.п.), используя партнерские соглашения с ведущими социальными сетями. Например, *Salesforce* и *SAP* имеют договоренности с блогерской площадкой Твиттер. Примеры систем, поддерживающих работу в онлайн-социальных сетях: *LITHIUM* (www.lithium.com), *SYSOMOS* (www.sysomos.com), *CYMFONY* (www.cymfony.com), *CRIMSON HEXAGON* (www.crimsonhexagon.com), *VISIBLE TECHNOLOGIES* (www.visibletechnologies.com), *ALTERIAN* (www.alterian.com),

RIGHTNOW TECHNOLOGIES (www.rightnow.com), *BRAND AURA* (www.brand-aura.com).

Прогнозирования и управления социальными сетями. Системы данного класса позволяют прогнозировать действия клиентов при помощи прогностических моделей и соответственно предпринимать упреждающие шаги.

Современные системы, предназначенные для коммерческих организаций, предоставляют возможности по п. 1–4; автоматизация п. 5 пока остается перспективным направлением их дальнейшего развития. Далее приведены краткие описания некоторых систем анализа социальных сетей.

3.1. СИСТЕМА АНАЛИЗА СОЦИАЛЬНЫХ СЕТЕЙ *RADIAN6*

Система *Radian 6* (www.radian6.com) предназначена для отслеживания в реальном времени упоминаний брендов с учетом тональности в социальных сетях (предоставляется панель управления мониторингом) и для участия в происходящих обсуждениях (предоставляется панель управления участием). Панель управления участием позволяет реагировать на активность в социальных сетях из одного места, используя имеющиеся учетные записи в блогах, площадках *Twitter* и *Facebook*.

Для ретроспективного анализа доступны данные, накопленные за последние 30 дней. Такое ограничение представляется существенным для анализа продолжительных кампаний в социальных сетях. Заметим, что система *Radian 6* в большей степени фокусируется на оперативном реагировании на происходящие события, нежели на бизнес-аналитике (стратегический уровень принятия решений), поэтому управляющие воздействия могут привести лишь к кратковременному всплеску продаж.

Основные характеристики системы *Radian 6* приведены в таблице 2.

Таблица 2. Унифицированное (в рамках используемой системы классификации) описание системы *Radian6*

Вендор	<i>SalesForce</i>
Сайт	www.radian6.com
Пользователи	Коммерческие организации

Уровень анализа данных	Мониторинг и анализ
Методы анализа	Базовые методы анализа и поиска текстов на уровне ключевых слов, анализ тональности текстов (поддерживается в том числе русский язык), визуальный анализ (инфографика)
Объекты анализа социальных сетей	Сеть в целом, пользователи (см. примечание ниже), информационные сообщения, мнения – анализ при помощи простых агрегированных показателей
Режим анализа	Анализ в режиме реального времени, ретроспективный анализ с ограничением в 30 дней
Объемы обрабатываемых данных	Отсутствует информация
Сбор данных	Сбор данных в режиме реального времени
Охват источников данных	Блоги, форумы, новостные медиа, сайты обмена изображениями и видео, социальные сети, такие как <i>YouTube, GoogleVideo, Flickr, Metacafe, Revver, Vech, DailyMotion, Bebo, Grouper, BrightCove, Twitter, Facebook, LinkedIn.</i>
Дополнительно	– Управление рабочим процессом (совместная классификация и категоризация постов, назначение приоритетов и выполнение работ по плану) – Возможность работы в соц.сетях: интеграция с <i>Twitter</i> и <i>Facebook</i> .
Клиенты	Более 50 компаний из <i>Fortune 100 (Pepsi, Dell, Kodak</i> и др. (см. полный список на www.radian6.com/about-us/customers/))

Примечание: пользователям системы предоставляется возможность настраивать (и сохранять) профили ранжирования по следующим показателям: по количеству постов заданной темы, по количеству комментариев заданной тематики, по количеству уникальных комментаторов, по количеству входящих ссылок,

по количеству голосов, по количеству ответов на тематических форумах.

3.2. СИСТЕМА АНАЛИЗА СОЦИАЛЬНЫХ СЕТЕЙ *ALTERIAN SM2*

Основное решение компании *SDL* в области анализа социальных медиа – система *Alterian SM2* в связке с дополнительными приложениями и сервисами. Система *Alterian SM2* – типичная для своего класса система, которая позволяет отслеживать упоминания брендов в социальных сетях с учетом тональности (определяется положительная, отрицательная и нейтральная тональность). Кроме того утверждается, что система позволяет локализовать места обсуждений и определять демографические характеристики пользователей социальных сетей.

Основные характеристики системы *Alterian SM2* приведены в таблице 3.

Таблица 3. Унифицированное (в рамках используемой системы классификаций) описание системы *AlterianSM2*

Вендор	<i>SDL</i> (www.sdl.com)
Сайт	www.alterian.com
Пользователи	Коммерческие организации
Уровень анализа данных	Мониторинг и анализ
Методы анализа	Базовые методы анализа и поиска текстов на уровне ключевых слов (поддерживается в том числе русский язык), анализ тональности текстов (русский язык не поддерживается), тематический анализ, визуальный анализ (инфографика)
Объекты анализа социальных сетей	Сеть в целом, пользователи, информационные сообщения, мнения – анализ при помощи простых агрегированных показателей
Режим анализа	Анализ в режиме реального времени, ретроспективный анализ (5 лет)
Объемы обрабатываемых данных	Более 60 миллионов постов, комментариев и твитов ежедневно
Сбор данных	Сбор данных в режиме реального времени

Охват источников данных	Блоговые площадки (Живой журнал, TypePad, Twitter, Plurk, Identi.ca), доски объявлений и форумы, вики сайты, сервисы обмена фотографиями и видео (YouTube, Flickr), социальные сети (Ning, Facebook, LinkedIn), сайты электронных объявлений (Craigslist), сайты обзоров потребителей (Epinions)
Клиенты	MD Anderson Cancer Center, Pursuit, YouCast, Red Bricks Media и др.

3.3. СИСТЕМА АНАЛИЗА СОЦИАЛЬНЫХ СЕТЕЙ BRANDSPOTTER

Позиционируется как система мониторинга и управления репутацией бренда в социальных сетях: отслеживаются упоминания бренда с учетом тональности; отслеживаются наиболее значимые пользователи социальных сетей по данной тематике (значимые с точки зрения количества упоминаний, тональности упоминаний, количеству последователей и друзей).

Таблица 4. Унифицированное (в рамках используемой системы классификаций) описание системы BrandSpotter

Вендор	ООО «Мониторинг социальных медиа»
Сайт	http://brandspotter.ru/
Пользователи	Коммерческие организации
Уровень анализа данных	Мониторинг и анализ
Методы анализа	Базовые методы анализа и поиска текстов на уровне ключевых слов, анализ тональности текстов (русский язык поддерживается), тематический анализ, визуальный анализ (инфографика)
Объекты анализа социальных сетей	Сеть в целом, пользователи, информационные сообщения, мнения – анализ при помощи простых агрегированных показателей
Режим анализа	Анализ в режиме реального времени
Объемы обрабатываемых данных	Отсутствует информация
Сбор данных	Сбор данных в режиме реального времени

Охват источников данных	ВКонтакте, Твиттер, ЖивойЖурнал, ..
Клиенты	Отсутствует информация

В категорию российских систем входят, помимо *BrandSpotter*, системы *IQBuzz* (iqbuzz.ru), *YouScan* (www.youscan.ru), *Monitorix* (monitorix.biz).

Выше представлены примеры типичных систем анализа социальных сетей для коммерческих организаций; более полный список таких систем можно получить по адресу <http://wiki.kenburbarry.com>.

Достоинства систем данного класса в том, что они, в отличие от массовых систем, позволяют проводить целенаправленную работу с социальными сетями. В частности, отслеживать упоминания брендов и продуктов по заданному набору ключевых слов, определять значимость обсуждений при помощи типичного набора показателей, поддерживать работу с пользователями социальных сетей при помощи унифицированного доступа к учетным записям организации, интегрировать данные социальных сетей с имеющимися данными организации (например, с данными продаж).

Ограниченность возможностей таких систем в том, что они ориентированы на оперативное реагирование, а не на стратегическую бизнес-аналитику, прогнозирование и управление. Кроме того, они, в отличие от массовых систем, требуют наличия дорогостоящей инфраструктуры и квалифицированных сотрудников.

4. Системы анализа социальных сетей для силовых структур и органов государственной власти

Конечными пользователями такой системы обычно являются некоторые силовые структуры (в том числе службы безопасности) или органы государственной власти.

Интерес к использованию систем анализа сетей государственными и силовыми структурами подтверждается новостями из СМИ. По информации «Коммерсанта», Служба внешней разведки РФ в январе-феврале 2012 года провела три закрытых

конкурса на исследование и разработку систем мониторинга блогосферы и формирования общественного мнения с помощью массового распространения информации в соцсетях [1].

Задачи, решаемые службами информационной безопасности в поле онлайн-социальных сетей:

- обнаружение информационных атак (определение узлов, с которых производится атака; оптимальное размещение сигнальных элементов);
- предотвращение информационных атак (оценка затрат на нападение на объекты атаки и затрат на защиту);
- формирование и разрушение различных сетей (социальных/информационных);
- обнаружение сообществ злоумышленников (например, террористов), отслеживание активности злоумышленников.

Пример анализа такого рода – анализ и прогноз событий в ходе войны в Афганистане на основе открытых источников, в данном случае, дневников *WikiLeaks* [45].

Задачи, решаемые органами государственной власти в поле социальных сетей:

- оценка и объективное понимание состояния российского общества и происходящих в нем процессов;
- определение значимости (влиятельности) онлайн-социального пространства как политической площадки (также определение значимости отдельных социальных сетей и степени их перекрытия, значимости отдельных сообществ, отдельных пользователей);
- определение значимости политических тем (напр., «коррупция»), событий (например, «принятие резолюции по Ливии СБ ООН»), процессов (например, «ситуация в Сирии») и персон (например, «Навальный»), а также отношения к ним;
- поддержание взаимоотношений с гражданским обществом; обнаружение радикальных сообществ и отслеживание их активности и т.п.

Заметим, что, с одной стороны, для решения определенных выше задач могут быть использованы социально-сетевые CRM («С» можно расшифровывать в данном случае как *Citizen*), упомянутые выше, с другой стороны – требуются системы социально-сетевой аналитики (системы, так сказать, *social intelligence*),

а также системы, основанные на методологии *OSINT*. Методология *OSINT (Open Source Intelligence)* – разведка на основе открытых источников. Она включает в себя поиск, выбор и сбор информации, полученной из общедоступных источников, и её согласованный и перекрёстный анализ. В качестве основных поставщиков данных выступают: СМИ, социальные медиа, публичные отчеты органов государственной власти и частных компаний, материалы официальных пресс-конференций, различные публичные заявления, материалы различных конференций, семинаров, и т.п. Существует ряд курсов и тренингов по практическим аспектам применения этой методологии, см., например, onstrat.com.

Ниже приведены примеры систем данного класса.

4.1. СИСТЕМЫ АНАЛИЗА ОТ КОМПАНИИ RECORDEDFUTURE

ИТ-компания *RecordedFuture*, поддержанная *Google*, на своём web-сайте (www.recordedfuture.com) заявляет, что открывает новый этап в сборе и анализе разведывательной информации. Программные средства этой фирмы в реальном времени отслеживают десятки и сотни тысяч веб-сайтов, блогов и аккаунтов *Twitter*. Цель анализа – выявление паттернов событий: взаимосвязей между людьми, закономерностям по которым происходят эти события и предсказание хода аналогичных событий в дальнейшем.

4.2. СИСТЕМА АНАЛИЗА «ПРИЗМА»

Позиционируется как инструмент оперативного мониторинга и анализа политико-социальной активности населения в *интернет блогах и форумах*. Работает на базе информационно-аналитической системы «Медialogия».

Система предназначена для управления репутацией и рисками в социальных медиа.

Охват источников информации: сообщения более 40 млн. русскоязычных социальных медиа: блогов, микроблогов, форумов и социальных сетей.

Основные функции работы с информацией:

1. Анализ тональности высказываний по отношению к информационным объектам.

2. Кластеризация сообщений по сюжетам с возможностью найти каждое отдельное сообщение и его характеристики (в том числе тональность); ранжирование сюжетов по обсуждаемости.

3. Анализ динамики параметров обсуждений информационных объектов.

4. Определение интереса блогосферы к тем или иным информационным объектам, выявление аномального интереса.

5. Оценка реакции социальных медиа на события, связанные с определенными ведомствами, регионами, руководителями.

6. Предупреждение о возможных рисках – возможно, заключается в слежении за количественной и качественной динамикой обсуждения информационного объекта и предсказании дальнейшей динамики.

7. Определение ботов и слежение за их активностью.

Возможные слабости системы: «Там берется несколько сотен блогеров, и система анализирует то, что происходит внутри их замкнутого сообщества. Это слишком малая часть, она не дает представления об общей картине. Мне же нужно понимать, что в реальном времени происходит в сетях» (Константин Костин, бывший глава управления внутренней политики Администрации Президента РФ). В целом такая система похожа на системы, предназначенные для коммерческих организаций, поэтому достоинства и недостатки те же.

4.3. СИСТЕМА «PALANTIR» (КОРПОРАЦИЯ PALANTIR)

Система *Palantir* оценивает транзакции в интернете (финансовые транзакции через кредитные карты, звонки по сотовому телефону, записи об адресатах и темах писем электронной почты, покупки и использования авиабилетов, логи разыскиваемой человеком информации в интернете и т.д.) как часть «общего паттерна активности» пользователя. В такой системе выявляется информация о компьютере человека, организующего транзакцию, о других людях, с которыми он ведет дела, а также о том, как это все вписывается в общую историю транзакций [36].

Идея состоит в скрещивании алгоритмов искусственного интеллекта с опытом человека-аналитика, каждая информационное событие – часть «общего паттерна активности».

Основная цель создания системы – обнаружение и разоблачение злонамеренных организаций. Раскрытие связанных событий на основе огромного массива разрозненной информации.

Ниже перечислены функции системы:

1. Работа с произвольными типами неструктурированной и структурированной информации.

2. Человеко-читаемое представление данных (в основном графы связей и различные графики).

3. Способность выполнять статистическую, временную, геопространственную и реляционную обработку всех данных.

4. Способность строить социальные сети интересующих объектов/субъектов (т.е. отображать графовую структуру их взаимодействия/знакомств и т.п.).

5. Возможность «связывать» и оценивать влияние внешних событий на ключевые события/мероприятия.

Достоинства систем данного класса в расширенных аналитических возможностях и умении работать с различными открытыми источниками информации, а не только с социальными сетями (продукты *OSINT*). Однако такие системы являются закрытыми, их доступные описания неполны, туманны и мифологизированы.

4.4. СИСТЕМЫ АНАЛИЗА СОЦИАЛЬНЫХ СЕТЕЙ, ИСПОЛЬЗУЕМЫЕ В НАУЧНЫХ ИССЛЕДОВАНИЯХ

Системы данного рода реализуют некоторый набор математических методов, которые могут быть использованы для анализа социальных сетей. При этом могут использоваться методы многомерного статистического анализа и искусственного интеллекта (в том числе методы *data mining*, *text mining*, *image/video mining*), а также методы анализа сетевых структур. Системы, реализующие первый класс методов, широко известны (например, статистический пакет *IBM SPSS Statistics* или *Statsoft STATISTICA*), в отличие от систем, реализующих второй класс методов. Поэтому в данном разделе остановимся подробнее на системах анализа сетевых структур.

Существует множество систем, используемых аналитиками (в основном исследователями), как для визуализации сетевых структур, так и для проведения вычислений. Это могут быть готовые продукты с пользовательским интерфейсом и набором реализованным функций, а также библиотеки вычислительных методов. Ниже кратко описаны некоторые системы, разработанные для научных исследований. Все рассмотренные системы, кроме *Gephi*, не имеют пользовательского интерфейса и представляют собой просто библиотеки вычислительных функций для анализа и визуализации графов. В приложении 2 приведены некоторые примеры исследований с применением таких библиотек.

4.5. СИСТЕМА АНАЛИЗА И ВИЗУАЛИЗАЦИИ СЕТЕВЫХ СТРУКТУР GEPHI

Система предназначена для визуализации и базового анализа графов среднего размера (до миллиона узлов). Имеет многофункциональный пользовательский интерфейс и богатый набор способов визуализации сетей (включая визуализацию в реальном времени динамических графов).

Gephi развивает *Gephi Consortium*, включающий более десятка организаций из США и Европы. Среди спонсоров проекта – *Google* и *Oracle*. *Gephi* используется в научных исследованиях и образовательных программах (например, в онлайн курсе *Michigan University Social Network Analysis* (www.coursera.org/course/sna)).

Архитектура системы и открытый исходный код предусматривают добавление новых функций за счет разработки плагинов.

Основные характеристики системы *Gephi* приведены в таблице 5.

Таблица 5. Унифицированное (в рамках используемой системы классификаций) описание системы *Gephi*

Сайт	gephi.org
Пользователи	Научные, образовательные организации
Объемы данных	До 1 миллиона узлов и ребер
Сбор данных	Отсутствует
Источники данных	Отсутствуют

Режим анализа	Ретроспективный анализ
Методы	Визуальный анализ. Базовые статистические методы Базовые методы теории графов
Рассматриваемые объ- екты	Структура сети (узлы, направленные и ненаправленные связи)
Условия распространения	<i>OpenSource (CDDL 1.0, GPL 3.0)</i>
Языковая поддержка	<i>English</i>
Разработчик	<i>GephiConsortium</i> (более 10 организаций). США, Франция, Германия, др.
Клиенты	Используется в исследовательских проектах, для визуализации данных и в образовательных программах.

4.6. СИСТЕМА АНАЛИЗА СЕТЕВЫХ СТРУКТУР *IGRAPH*

В отличие от *Gephi*, *igraph* – это не готовая система, а библиотека вычислительных функций [15, 41]. *Igraph* реализован на языке *C*, но имеет программные оболочки для *Python*, *Ruby* и *R*. Библиотека содержит большое количество методов теории графов, как классических, так и появившихся недавно, например, методы ранжирования узлов, поиск сообществ и т.д. Предназначена библиотека *igraph* для анализа больших сетей в исследовательских и учебных проектах. Есть примеры использования *igraph* для масштабных распределенных вычислений [15].

Основные характеристики *igraph* приведены в таблице 6.

Таблица 6. Унифицированное (в рамках используемой системы классификаций) описание системы *igraph*

Сайт	igraph.sourceforge.net
Пользователи	Научные, образовательные организации
Объемы данных	До нескольких миллионов узлов и ребер
Сбор данных	Отсутствует
Источники данных	Отсутствуют
Режим анализа	Ретроспективный анализ
Методы	Широкий набор методов теории графов
Рассматриваемые объ-	Структура сети (узлы, направленные и нена-

екты	правленные связи)
Условия распространения	<i>OpenSource (GPL 2.0+)</i>
Языковая поддержка	<i>English</i>
Разработчик	Gábor Csárdi (<i>Harvard University</i> , США), Tamás Nepusz (<i>Eötvös University</i> , Венгрия)
Клиенты	Используется в исследовательских проектах

4.7. СИСТЕМА АНАЛИЗА СЕТЕВЫХ СТРУКТУР NETWORKX

Эта система напоминает *igraph*, но реализована на языке *Python* [23]. Поскольку этот язык программирования широко используется для научных расчетов, библиотека ориентирована на удобную интеграцию с *Python*-проектами. Основные характеристики системы *NetworkX* приведены в таблице 7.

Таблица 7. Унифицированное (в рамках используемой системы классификаций) описание системы *NetworkX*

Сайт	networkx.lanl.gov
Пользователи	Научные, образовательные организации
Объемы данных	До нескольких миллионов узлов и ребер
Сбор данных	Отсутствует
Источники данных	Отсутствуют
Режим анализа	Ретроспективный анализ
Методы	Широкий набор методов теории графов
Рассматриваемые объемы	Структура сети (узлы, направленные и ненаправленные связи)
Условия распространения	<i>OpenSource (BSD License)</i>
Языковая поддержка	<i>English</i>
Разработчик	Aric Hagberg, Dan Schult, Pieter Swart и другие
Клиенты	Научные организации

4.8. СИСТЕМА АНАЛИЗА СЕТЕВЫХ СТРУКТУР SNAP (STANFORD NETWORK ANALYSIS PROJECT)

Система представляет собой библиотеку вычислительных методов, разработанную для исследований Стэнфордского уни-

верситета. Система реализована на языке C++, относится к тому же классу продуктов, что *igraph* и *networkX*. Система *SNAP* ориентирована на обработку больших массивов данных и использовалась во многих исследованиях.

Помимо библиотеки, на сайте проекта *SNAP* доступны массивы данных о взаимодействии реальных пользователей *LiveJournal*, *Epinion*, *Amazon* и др.

Полный список публикаций проекта находится на сайте snap.stanford.edu/papers.html.

Основные характеристики системы *SNAP* приведены в таблице 8.

Таблица 8. Унифицированное (в рамках используемой системы классификаций) описание системы *SNAP*

Сайт	snap.stanford.edu
Пользователи	Научные, образовательные организации
Объемы данных	До нескольких миллионов узлов и ребер
Сбор данных	Отсутствует
Источники данных	Отсутствуют
Режим анализа	Ретроспективный анализ
Методы	Широкий набор методов теории графов
Рассматриваемые объемы	Структура сети (узлы, направленные и ненаправленные связи)
Условия распространения	<i>OpenSource (BSD License)</i>
Языковая поддержка	<i>English</i>
Разработчик	<i>Stanford University</i>
Клиенты	Используется в исследовательских проектах, в частности, Стэнфордским университетом.

Выше представлены типичные системы анализа сетевых структур, список таких систем можно найти по адресу en.wikipedia.org/wiki/Social_network_analysis_software.

Как можно видеть, представленные системы ориентированы прежде всего на ретроспективный анализ графов, состоящих из относительно небольшого числа вершин (до миллиона вершин). Также такие системы работают только с одним уров-

нем представления – графом, что требует большой предварительной подготовки и структурирования данных. Никак не затронут уровень анализа текстов и, как правило, весьма скуден арсенал методов *data mining*.

Для анализа реальных онлайн-социальных сетей (напомним, что сеть *Facebook* насчитывает миллиард пользователей) нужны новые системы анализа сетевых структур, основанные на технологиях *BigData* [12, 7].

Системы для визуального анализа, как *Gephi*, широко применяются в исследовательских проектах для визуализации результатов. Вычислительные библиотеки, как *igraph*, обладают богатыми возможностями интеллектуального анализа графов и применяются для научных расчетов. Их использование в качестве компонентов полноценной информационной системы возможно, но требует интеграции с технологиями анализа текстов и статистическими библиотеками.

5. Примеры экспериментальных исследований социальных сетей

В 1940-х годах ученые, занимающиеся исследованием человеческого мозга, выдвинули гипотезу, что его уникальные свойства обусловлены не характеристиками отдельных нервных клеток, а структурой связей между ними. К настоящему моменту исследования сетей самой разной природы – биологических, физических, социальных и экономических – получили общее название *network science*, или наука о сетях.

Допустим, мы представили группу людей как сеть, т.е. набор узлов, соединенных связями, или ребрами. Естественно ожидать, что влияние, авторитет, другие характеристики каждого агента в группе, можно определить на основе положения данного узла, соответствующего данному агенту в сети.

Политологи Паджет (Pagett) и Ансель (Ansell) в 1993 году исследовали структуру взаимоотношений между влиятельными родами Флоренции в начале 15 века [35]. Цель состояла в том, чтобы объяснить возвышение семьи Медичи в 1434 году и их могущество на протяжении следующих шестидесяти лет.

Гипотезы об экономическом превосходстве Медичи не подтвердились. Существовали куда более состоятельные семьи, которые, несмотря на это, не добились заметных политических успехов. Паджет и Ансель предположили, что род Медичи занимал особое положение в структуре социальных связей среди других 215 влиятельных родов.

Исследователи построили сеть взаимоотношений между флорентийскими родами. Узел в сети представлял одну семью, а ребро – наличие родственных, экономических или политических отношений. Учитывались такие факты, как брак между членами семей, деловое партнерство, наличие совместной собственности и другие. Рассматривались только долговременные устоявшиеся контакты, которые в терминологии Грановеттера (Granovetter) [22] называются сильными связями. Центральная часть этой сети показана на рис. 4 (визуализация из книги [25]).

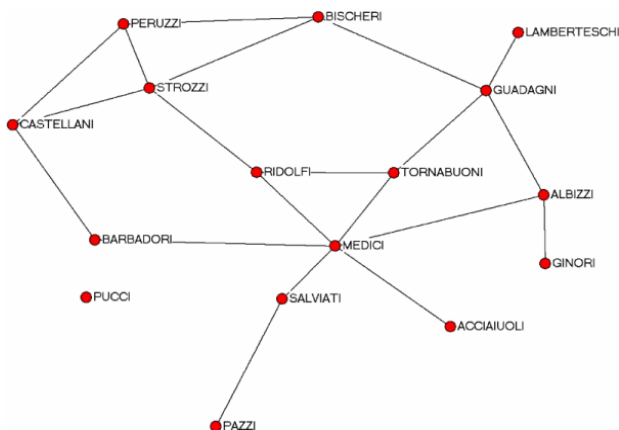


Рис. 4. Ядро сети связей между влиятельными родами во Флоренции начала 15 века.

Узел, представляющий Медичи (Medici), имеет шесть связей. Это больше, чем у любого другого узла, но существует несколько узлов с четырьмя связями. Простого различия в числе связей недостаточно, чтобы объяснить превосходство Медичи.

Особая роль Медичи становится понятна, если рассмотреть кратчайшие пути между всеми узлами сети. Например, между

узлами Барбадори (Barbadori) и Гуадани (Guadagni) есть два возможных кратчайших пути: Барбадори-Медичи-Торнабуони (Tornabuoni)-Гуадани и Барбадори-Медичи-Альбицци (Albizzi)-Гуадани. Оба пути проходят через Медичи. Таким образом, семья Медичи становится ключевой в установлении контактов между Барбадори и Гуадани.

Формально доля всех кратчайших путей, пролегающих через узел, называется промежуточностью (*betweenness*) узла. Для Медичи промежуточность равна 0,522. Это означает, что через семью Медичи проходит более половины всех кратчайших путей в сети. Следующая семья – Гуадани – имеет промежуточность 0,255.

Центральное положение Медичи сделало их ключевой фигурой в том, что касалось обмена информацией, заключения деловых и политических соглашений, и позволило сконцентрировать в своих руках огромную власть. Таким образом, ключом к пониманию причин возвышения Медичи является не их экономическая мощь, а наличие сети прочных взаимоотношений с другими участниками общественной жизни Флоренции.

Подобные исследования относительно небольших сообществ сейчас остаются актуальными. После терактов 11 сентября активно начали развиваться исследования социальных сетей террористических и криминальных группировок [27, 37, 16].

Для ученых, которые изучали структуру отношений в политической элите Флоренции, основной проблемой было получение первичных данных. Им приходилось лично изучать массив исторических документов, который наверняка не полностью охватывал всю структуру взаимоотношений между родами. С появлением интернета и особенно онлайн-социальных сетей огромное количество данных стало доступным для изучения. Размеры исследуемых сетей возросли на несколько порядков.

Сегодня количество открытых и закрытых исследований онлайн-социальных сетей настолько велико, что охватить все направления просто невозможно. Мы рассмотрим те работы, которые имеют прикладную направленность и, на наш взгляд, отражают тенденции в этой области.

5.1. СТРУКТУРНЫЕ СВОЙСТВА СОЦИАЛЬНЫХ СЕТЕЙ

В настоящее время статические структурные свойства крупномасштабных саморазвивающихся сетей, таких как Интернет, сети цитирования ученых, онлайн-социальные сети, хорошо изучены. Например, степенной закон распределения [8], «тесный мир» (*small world*) [43], т.е. малое среднее расстояние между узлами. Эти свойства наблюдались первоначально для социальных сетей до появления Интернета [38, 33, 22]. Затем было установлено, что структура Интернета обладает такими же характеристиками [10, 24, 20]. Аналогичные результаты подтвердились для сетей контактов *e-mail* [17], популярных социальных сервисов *Flickr*, *Livejournal*, *YouTube* [34]. В 2011 году исследование сети Facebook показало, что ее топология имеет сходные свойства, но, например, распределение числа друзей описывается законом, отличным от степенного [42].

С точки зрения прикладных исследований, интерес представляет изучение сообществ, существующих в социальных сетях, как отражение структуры общества. Центр исследований взаимосвязей общества и Интернета при Гарвардском университете провел серию исследований политически ориентированных сообществ в иранской [26], арабской [19] и русскоязычной [18] блогосферах.

Основным объектом анализа были внешние СМИ, которые цитируют блогеры в своих сообщениях. Из текстов постов автоматически извлекались ссылки на внешние сайты. Из всего множества полученных сайтов были выбраны те ресурсы, которые могут служить индикаторами политических взглядов блогеров. Это сайты онлайн-СМИ, политических партий и общественных организаций.

Блогеры, которые цитировали сходный набор сайтов, относились к одному кластеру. Также исследовались ссылки на других блогеров. Одна из целей исследования состояла в том, чтобы выяснить, насколько часто блогеры ссылаются на посты своих идеологических оппонентов, присутствует ли тенденция к образованию замкнутых сообществ.

Для американской блогосферы аналогичное исследование показало, что существует ярко выраженное разделение блогеров на либералов и консерваторов [8]. При этом из всех ссылок

только 15% относится к перекрестному цитированию между кластерами. Большинство предпочитает цитировать своих сторонников по идеологии.

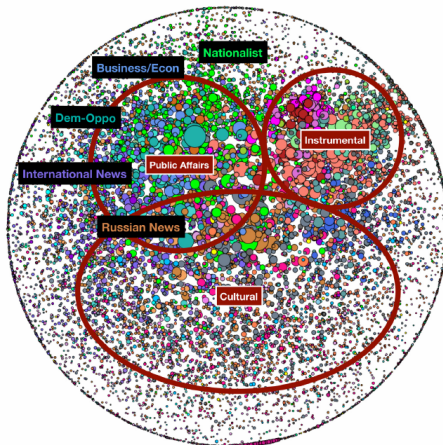


Рис. 5. Карта политических предпочтений в русскоязычной блогосфере. Точки обозначают блогеров. Размер точки показывает популярность блогера; цветом обозначен кластер, к которому относится блогер

Остановимся подробнее на результатах, полученных для российской блогосферы. На рис. 5 показана карта разбиения блогосферы на кластеры, группы блогеров, которые цитируют похожие ресурсы. Для визуализации использовался алгоритм *Fruchterman–Reingold* [21], имитирующий «физическое взаимодействие» блогеров. На всех блогеров действует одна общая сила, «сдувающая» их к краям карты. В то же время каждые два блога «притягиваются» друг к другу с силой, пропорциональной числу ссылок между ними. Получается, что ближе друг к другу расположены блогеры, активно цитирующие друг друга. Видно, что блогеры, принадлежащие к разным кластерам, часто располагаются рядом. Нет тенденции к образованию замкнутых идеологически однородных сообществ, как в американской блогосфере.

В исследовании центра *Berkman* кроме компьютерных методов анализа активное участие принимали русскоязычные эксперты. Они классифицировали 1200 случайно выбранных блогов по демографическим параметрам и обсуждаемым темам. Дискуссионное ядро разделилось на четыре зоны: 1) политические и общественные темы; 2) региональные блоги и экспатрианты; 3) культурные темы; 4) инструментальные блоги. Эти четыре зоны содержали 24 основных кластера.

Основное внимание было уделено шести кластерам в зоне политических и общественных дискуссий: 1) международные темы, 2) обсуждение российских новостей, 3) националисты, 4) новости экономики, бизнеса и финансов, 5) общественные и экологические активисты, 6) демократическая оппозиция.

Стоит отметить, что под националистами авторы, по-видимому, подразумевают всех патриотически настроенных блогеров, выступающих за сильное государство, возрождение империи и т.д.

Ученые подчеркивают, что российская блогосфера имеет ряд уникальных характеристик. С одной стороны, используется множество блог-платформ, в основном слабо связанных друг с другом. С другой стороны, существует дискуссионное ядро с высокой ссылочной активностью, расположенное преимущественно в *Livejournal*. Внутри центрального ядра не прослеживается тенденция к изоляции сообществ единомышленников, как это наблюдалось в американской и иранской блогосферах.

Несмотря на то, что два наибольших по размеру кластера, где обсуждаются международные и внутрироссийские события, явно не связаны с какими-либо организациями в оффлайне, среди блогеров здесь преобладает критика действий власти. Кластеры демократической оппозиции и националистов представляют два противоположных тренда в российской политике, хотя и не полностью аналогичны либералам и консерваторам в США.

Подобные исследования представляют интерес для социологов, политологов, хотя некоторые результаты быстро теряют актуальность и требуют возобновления.

5.2. ПРОЦЕССЫ РАСПРОСТРАНЕНИЯ ИНФОРМАЦИИ

Помимо структурных характеристик, большой объем работ посвящен вопросам распространения информации в социальных сетях. Разработано множество моделей динамических процессов на сетевых структурах. Подробный обзор доступен в [2]. Здесь мы рассмотрим исследования, которые имели прикладную направленность.

Группа исследователей Стэнфордского университета с начала 2000-х годов изучает сетевую структуру Интернета, активность пользователей в блогосфере и онлайн-СМИ. Они опубликовали несколько работ по распространению новостей во время президентских выборов в США в 2004 и в 2008 годах [8, 28, 29, 32].

Объектом изучения были популярные выражения, так называемые мемы, которые слабо видоизменяются при повторном цитировании. Оценивалась популярность разных мемов, отслеживались пути их распространения. Прикладным результатом исследований стала разработанная технология *Memetracker*. На рис. 6 приведен пример ее использования, взятый с сайта <http://memetracker.org>.

Основой технологии является метод классификации фраз, который позволяет определить происхождение той или иной цитаты. Визуально история изменений высказываний представляется в виде направленного ациклического графа. Пример такого графа показан на рис. 7. В результате несколько сотен и тысяч высказываний, происходящих из одного источника, можно рассматривать как одно высказывание. Это позволяет точнее оценить масштаб распространения и отследить динамику популярности, что и показано на рис. 6.

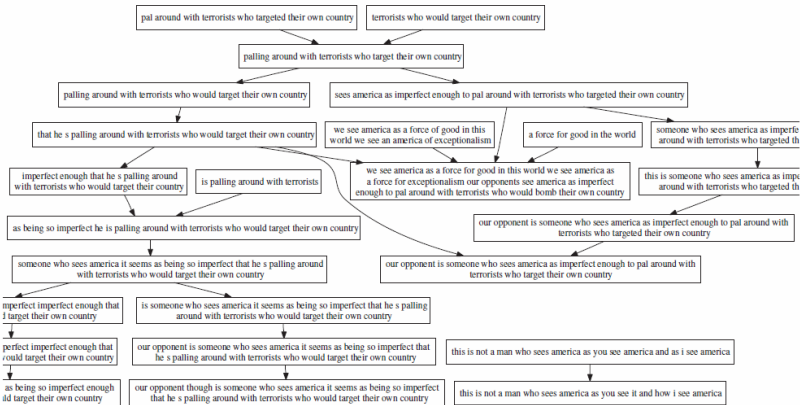


Рис. 7. Ациклический граф вариаций высказывания Сары Пэйлин «Our opponents is someone who sees America, it seems, as being so imperfect, imperfect enough that he's palling around with terrorists who would target their own country»

Помимо анализа временных закономерностей, группа из Стэнфордского университета занимается изучением каскадов распространения информации, в том числе при неполных наблюдениях [44, 39], структуры сообществ в социальных сетях [30, 31], моделированием информационных эпидемий [14].

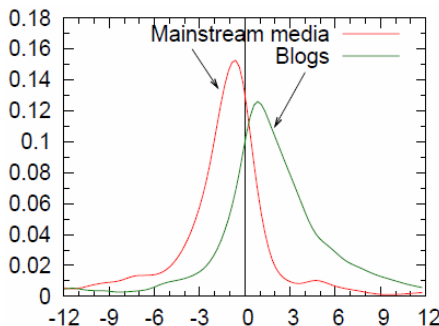


Рис. 8. Обсуждения в блогах и СМИ (горизонтальная ось – время в часах относительно суммарного максимума упоминаний, вертикальная ось – относительный объем упоминаний фразы, в процентах от общего числа упоминаний)

6. Заключение

В данной статье представлен краткий обзор информационных систем анализа онлайн-социальных медиа: (1) предназначенных для массового использования; (2) предназначенных для коммерческих организаций; (3) предназначенных для силовых структур и органов государственной власти; (4) открытых систем, предназначенных в основном для научных исследований.

Отметим, что слабость систем (1) и отчасти систем (2) состоит в бедности и поверхности методов анализа; системы (3) закрыты; системы (4) не в полной мере учитывают специфику социальных сетей, не обладают возможностью сбора данных и обрабатывают относительно малые их объемы.

Исходя из анализа достоинств и недостатков систем, в заключение можно сделать следующий вывод: идеальная система анализа социальных сетей должна:

- работать на всех уровнях анализа (от мониторинга социальных сетей до прогноза и управления) в различных режимах (в режиме реального времени и ретроспективном режиме),
- анализировать различные объекты социальной сети (от отдельно взятого информационного сообщения и отдельно взятого пользователя до социальной сети в целом и внешних по отношению к ней источников) и учитывать различные отношения между такими объектами (связи знакомств между пользователями, связи цитирования, связи комментирования и пр.),
- базироваться на математических моделях и методах интеллектуального анализа данных (статистических и графовых),
- интегрироваться с подсистемами сбора данных из различных открытых источников (социальных сетей, блоговых площадок, новостных ресурсов и т.п.),
- обрабатывать очень большие массивы данных (терабайты данных, миллионы узлов сети и сотни миллионов связей между ними).

Кроме того, важно, чтобы такая система была ориентированной на «обычного» аналитика в определенной предметной области. Представляется очевидным, что разработка серьезной

системы интеллектуального анализа социальных сетей для всех возможных пользователей весьма сложна и экономически нецелесообразна (сходство по функциональности современных систем разного предназначения объясняется тем, что они находятся только в начале своего развития). Поэтому можно предположить, что в ближайшие годы появятся системы для конкретных конечных пользователей, решающих частные задачи в тех или иных предметных областях.

Литература

1. БАРАБАНОВ И., САФРОНОВ И., ЧЕРНЕНКО Е. Разведка ботом // Газета «Коммерсантъ» [Электронный ресурс]. – 2012.– №158. – URL: <http://kommersant.ru/doc/2009256> (дата обращения 27.01.2013).
2. 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.
3. *Разведка на основе открытых источников* [Электронный ресурс]. – URL: <http://ru.wikipedia.org/wiki/OSINT> (дата обращения 27.01.2013).
4. РАЙГОРОДСКИЙ А.М. *Модели случайных графов и их применение* // Труды МФТИ. – 2010. –Т. 2, №4. – С. 130–140.
5. *Список систем анализа сетевых структур* [Электронный ресурс]. – URL: http://en.wikipedia.org/wiki/Social_network_analysis_software (дата обращения 27.01.2013).
6. *Список систем анализа социальных медиа* [Электронный ресурс]. – URL: <http://wiki.kenburbary.com> (дата обращения 27.01.2013).
7. *Технологии анализа больших данных* [Электронный ресурс]. –URL: http://en.wikipedia.org/wiki/Big_data (дата обращения 27.01.2013).

8. ADAMIC L., GLANCE N. *The Political Blogosphere and the 2004 U.S. Election: Divided They Blog* // Proc. of the 3rd ACM international workshop on Link discovery, 2005. – P. 36–43
9. ALBERT R., BARABASI A.L. *Statistical Mechanics of Complex Networks* // Review of modern physics. – 2002. – Vol. 74, №1. – P. 47–97.
10. ALBERT R., JEONG H., BARABASI A.-L. *Diameter of the World-Wide Web* // Nature. – 1999. – №401. – P. 130–131.
11. BASTIAN M., HEYMANN S., JACOMY M. *Gephi: An Open Source Software for Exploring and Manipulating Networks* // Proc. of the 3rd International ICWSM conference, in American Journal of Sociology. – 2009. – P. 361–362
12. *Big Data Jobs will be Plentiful*. – URL: <http://www.computer.org/portal/web/buildyourcareer/Big-Data-Jobs-will-be-Plentiful> (дата обращения 27.01.2013)
13. BRUNS A. *How long is a tweet? Mapping dynamic conversation networks on Twitter using Gawk and Gephi* // Information, Communication & Society. – 2012. – Vol. 15, №9. – P. 1323–1351.
14. CHAKRABARTI D., WANG Y., WANG C., LESKOVEC J., FALOUTSOS C. *Epidemic Thresholds in Real Networks* // ACM Transactions on Information and System Security (ACM TISSEC). – 2007. – Vol. 10, №4. – P. 1–26.
15. CSÁRDI G., NEPUSZ T. *The igraph software package for complex network research* // Inter Journal Complex Systems. – 2006. [Электронный ресурс]. – URL: <http://www.necsi.edu/events/iccs6/papers/c1602a3c126ba822d0bc4293371c.pdf> (дата обращения 27.01.2013).
16. DUVAL R.D., CHRISTENSEN K., SPAHIU A. *Bootstrapping a Terrorist Network* // Open SIUC Working Papers. – 2010. – Paper 20. – URL: http://opensiuc.lib.siu.edu/pnconfs_2010/20 (дата обращения 27.01.2013).
17. EBEL H., MIELSCH L.-I. BORNHOLDT S. *Scale-free topology of e-mail networks* // Physical Review E. – 2002. – Vol. 66 (2A), №3. – 035103.
18. ETLING B., ALEXANYAN K., KELLY J., FARIS R., PALFREY J., GASSER U. *Public Discourse in the Russian Blogosphere: Mapping RuNet Politics and Mobilization* // Berkman

- Center Research Publication. – №2010-11. [Электронный ресурс] – URL: http://cyber.law.harvard.edu/publications/2010/Public_Discourse_Russian_Blogosphere (дата обращения 27.01.2013.)
19. ETLING B., KELLY J., FARIS R., PALFREY J. *Mapping the Arabic blogosphere: politics and dissent online* // *New Media & Society*. – 2010. – Vol. 12, №8. – P. 1225–1243.
 20. FALOUTSOS M., FALOUTSOS P., FALOUTSOS C. *On Power-Law Relationships of the Internet Topology* // Proc. of the conference on Applications, technologies, architectures, and protocols for computer communication (ACM SIGCOMM'99). – 1999. – Vol. 29., №4. – P. 251–262.
 21. FRUCHTERMAN M.J., REINGOLD E.M. *Graph drawing by force-directed placement* // *Software: Practice and Experience*. – 1991. – Vol. 21., №11. – P.1129–1164.
 22. GRANOVETTER M. *The Strength of Weak Ties* // *American Journal of Sociology*. – 1973. – Vol. 78. – P. 1360–1380.
 23. HAGBERG A., SWART P., CHULT D.S. *Exploring network structure, dynamics, and function using NetworkX* // Proceedings of the 7th Python in Science conference, 2008. – P. 11–15.
 24. HUBERMAN B.A., ADAMIC L. *Growth dynamics of the World-Wide Web* // *Nature*. – 1999. – Vol. 401. – P. 131–131.
 25. JACKSON M.O. *Social and Economic Networks*. – Princeton University Press, 2008. – 520 p.
 26. KELLY J., ETLING B. *Mapping Iran's Online Public: Politics and Culture in the Persian Blogosphere* // Berkman Center Research Publication. – 2008. – №2008-01 [Электронный ресурс] URL: http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/Kelly&Etling_Mapping_Irans_Online_Public_2008.pdf (дата обращения 27.01.2013.)
 27. KREBS V. *Mapping Networks of Terrorist Cells* // *Connections*. – 2002. –Vol. 24, №3. – P. 43–52.
 28. LESKOVEC J., ADAMIC L., HUBERMAN B. *The Dynamics of Viral Marketing* // Proc. of the 7th ACM Conference on Electronic Commerce (EC'06). – 2006. –P. 228–237.
 29. LESKOVEC J., BACKSTROM L., KLEINBERG J. *Meme-tracking and the Dynamics of the News* // Proc. of the 15th ACM

- SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09). – 2009. – P. 497–506.
30. LESKOVEC J., LANG K., DASGUPTA A., MAHONEY M. *Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters* // *Internet Mathematics*. – 2009. – Vol. 6, №1. – P. 29–123.
 31. LESKOVEC J., LANG K., MAHONEY M. *Empirical Comparison of Algorithms for Network Community Detection* // *Proc. of the 19th ACM WWW International conference on World Wide Web (WWW'10)*. – 2010. – P. 631–640
 32. LESKOVEC J., YANG J. *Patterns of Temporal Variation in Online Media* // *Proc. of 4th ACM International Conference on Web Search and Data Mining (WSDM'11)*. – 2011. P. 177–186.
 33. MILGRAM S. *The Small-World Problem* // *Psychology Today*. – 1967. – Vol. 1, №1. – P. 61–67.
 34. MISLOVE A., MARCON M., GUMMADI K.P., DRUSCHEL P., BHATTACHARJEE B. *Measurement and Analysis of Online Social Networks* // *Proc. of 7th ACM SIGCOMM Internet Measurement Conference*. - 2007. – P. 29–42.
 35. PADGETT J.F., ANSELL C.K. *Robust Action and the Rise of the Medici, 1400-1434* // *American Journal of Sociology*. – 1993. – Vol. 98. – P. 1259–1319.
 36. *Palantir, или Говорящие камни на службе ЦРУ*. – URL: <http://www.3dnews.ru/offsyanka/621533>. (дата обращения: 27.01.2013).
 37. PERLIGER A., PEDAHZUR A. *Social Network Analysis in the Study of Terrorism and Political Violence* // *Open SIUC Working Papers*. – 2010. – Paper 48. — URL: http://opensiuc.lib.siu.edu/pn_wp/48/ (дата обращения: 27.01.2013).
 38. PRICE D.S. *A General Theory of Bibliometric and Other Cumulative Advantage Processes* // *Journal of the American Society for Information Science*. – 1976. – Vol. 27. – №5. – P. 292–306.
 39. SADIKOV E., MEDINA M., LESKOVEC J., GARCIA-MOLINA H. *Correcting for Missing Data in Information Cascades* // *Proc. of 4th ACM International Conference on Web Search and Data Mining (WSDM)*. – 2011. – P. 55–64.

40. *Spinn3r*. – URL: <http://spinn3r.com>. (дата обращения: 27.01.2013).
41. SU G., KUCHINSKY A., MORRIS J.H., STATES D.J., MENG F. *GLay: community structure analysis of biological networks* // *Bioinformatics*. – 2010. – Vol. 26, №24. – P. 3135–3137.
42. UGANDER J., KARRER B., BACKSTROM L., MARLOW C. *The Anatomy of the Facebook Social Graph* // *ArXiv:1111.4503 [cs.SI]*. 2011. – URL: <http://arxiv.org/abs/1111.4503> (дата обращения: 27.01.2013).
43. WATTS D.J., STROGATZ S.H. *Collective dynamics of 'small world' networks* // *Nature*. – 1998. – Vol. 393. – P. 440–442.
44. YANG J., LESKOVEC J. *Modeling Information Diffusion in Implicit Networks* // *Proc. of the 2010 IEEE International Conference On Data Mining (ICDM'10)*. – 2010. – P. 599–608.
45. ZAMMIT-MANGION A., DEWAR M., KADIRKAMANTHAN V., SANGUINETTI G. *Point process modelling of the Afghan War Diary* // *PNAS*. – 2012. – Vol. 109, №31. – P. 12414–12419.

INFORMATION SYSTEMS FOR SOCIAL NETWORKS ANALYSIS: A SURVEY

Nikolay Bazenkov, Institute of Control Sciences of RAS, Moscow, software engineer (n.bazenkov@gmail.com)

Dmitry Gubanov, Institute of Control Sciences of RAS, Moscow, Ph.D., senior staff scientist (dmitry.a.g@gmail.com)

Abstract: We survey contemporary information systems for social network analysis and give classifications of users and typical tasks. We describe systems, which are used by ordinary people, business and government structures, security services and scientific organizations. Most considered systems are used for monitoring and analysis of social networks and implement no forecast or control methods. In the end of the article we make a short review of applied researches of social networks.

Keywords: social network, data mining, information technologies.

Статья представлена к публикации членом редакционной коллегии В. В. Мазаловым