

Институт проблем управления РАН



Губанов Д.А.
с.н.с. ИПУ РАН, ст. исследователь
"Лаборатории цифрового общества"

Анализ социальных медиа

Социальные медиа

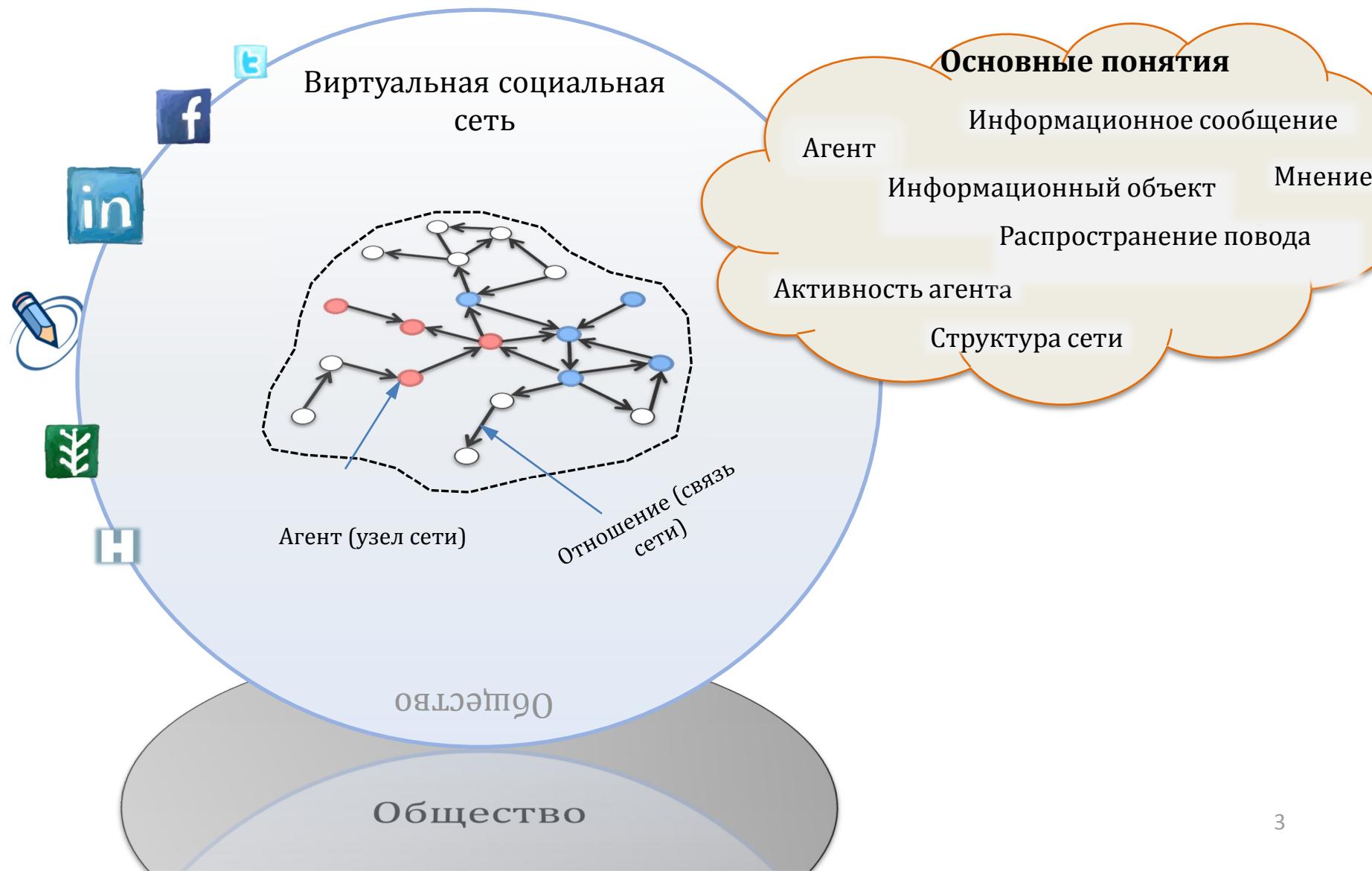
- **Социальные медиа.**
 - **Онлайновые сообщества:** Facebook (1 миллиард пользователей), LinkedIn (> 175 миллионов пользователей). **Коммуникация:** Windows Live Messenger. **Новостные и социальные медиа:** блоги (свои блоги ведут президенты США и России)..
- **Мотивация исследователей:**
 - Как возникают связи между пользователями социальной сети? ..
 - Как распространяется информация по сети? Как возникают информационные каскады? Какие пользователи являются их инициаторами?
 - Какие пользователи социальных сетей самые влиятельные? ...
 - **Мотивация практиков:** маркетинг, информационная безопасность, политика, бизнес, онлайневые игры, .. – там, где нужно учитывать взаимодействие между пользователями.

... социальные сети позволили заранее увидеть реальное число желающих участвовать в митинге и это число росло с каждым днем и часом. ... протест приобретает реальные очертания и массовость...

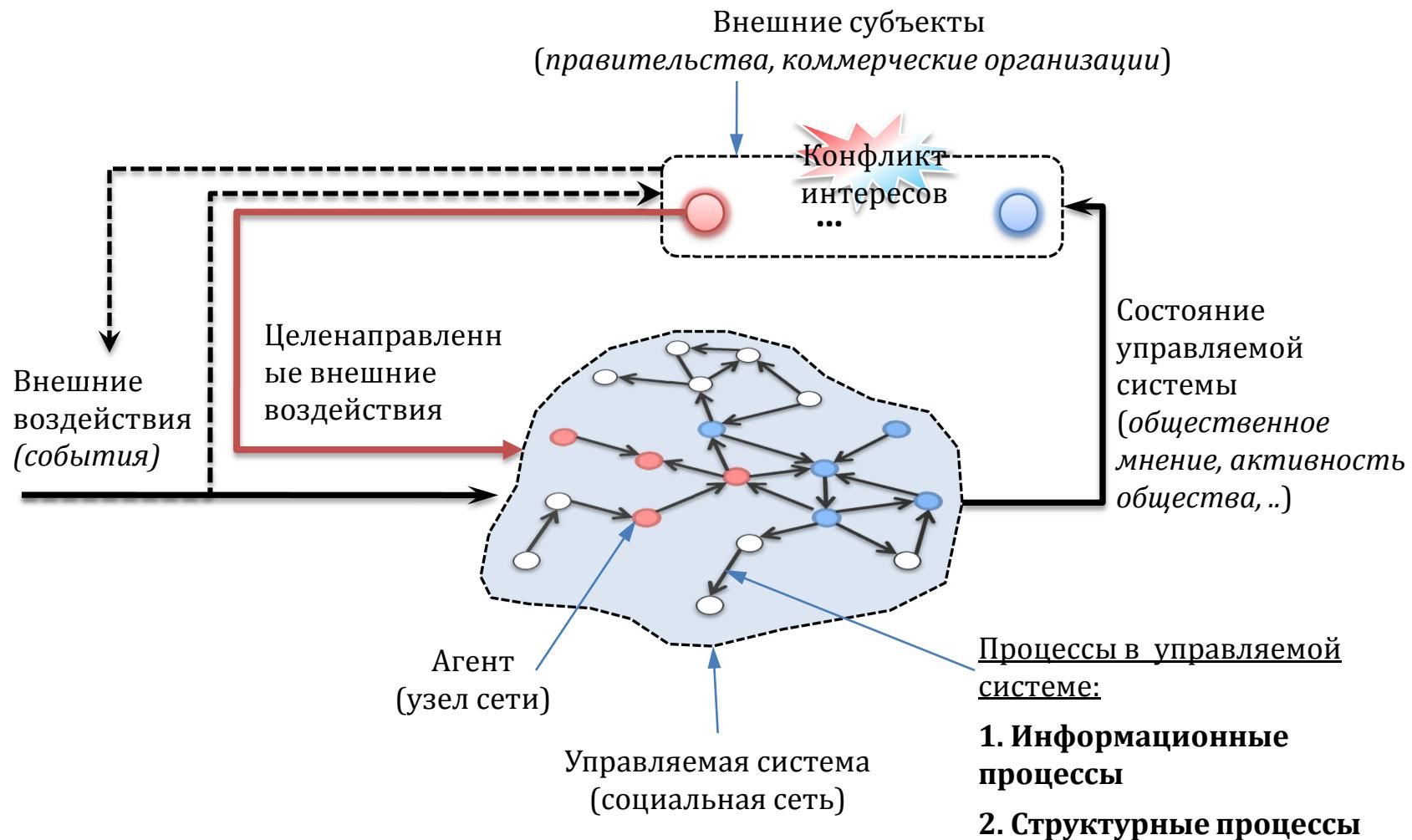
... впервые на массовый митинг люди пришли с лозунгами «из интернета». Именно в ЖЖ подробно освещались темы, которые мы увидели на плакатах митингующих на Болотной...

... влияние организующей силы сети Facebook в российских политических событиях по-прежнему велико...

Социальные сети



Управление и социальные сети



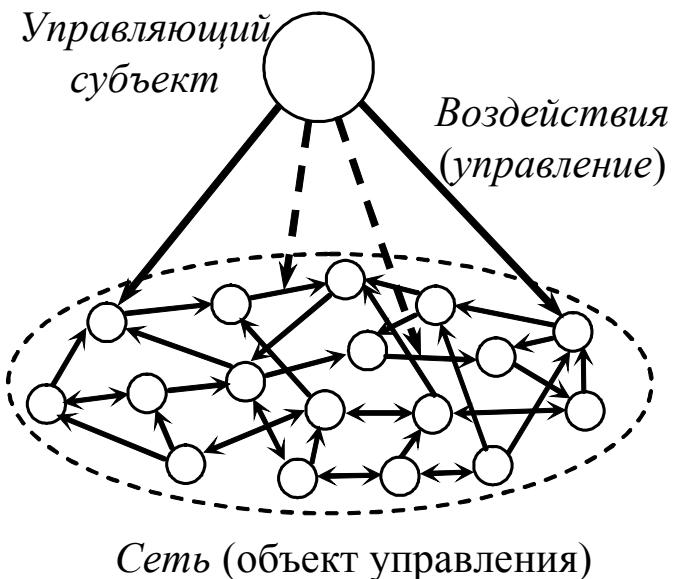
Виды управляющих воздействий

Объект управления описывается графом.
Управление может заключаться в
целенаправленном воздействии на
следующие компоненты объекта
управления:

- состав управляемой системы;
- структуру (связи между элементами)
управляемой системы;
- значения параметров, соответствующих
вершинам графа и его дуг.

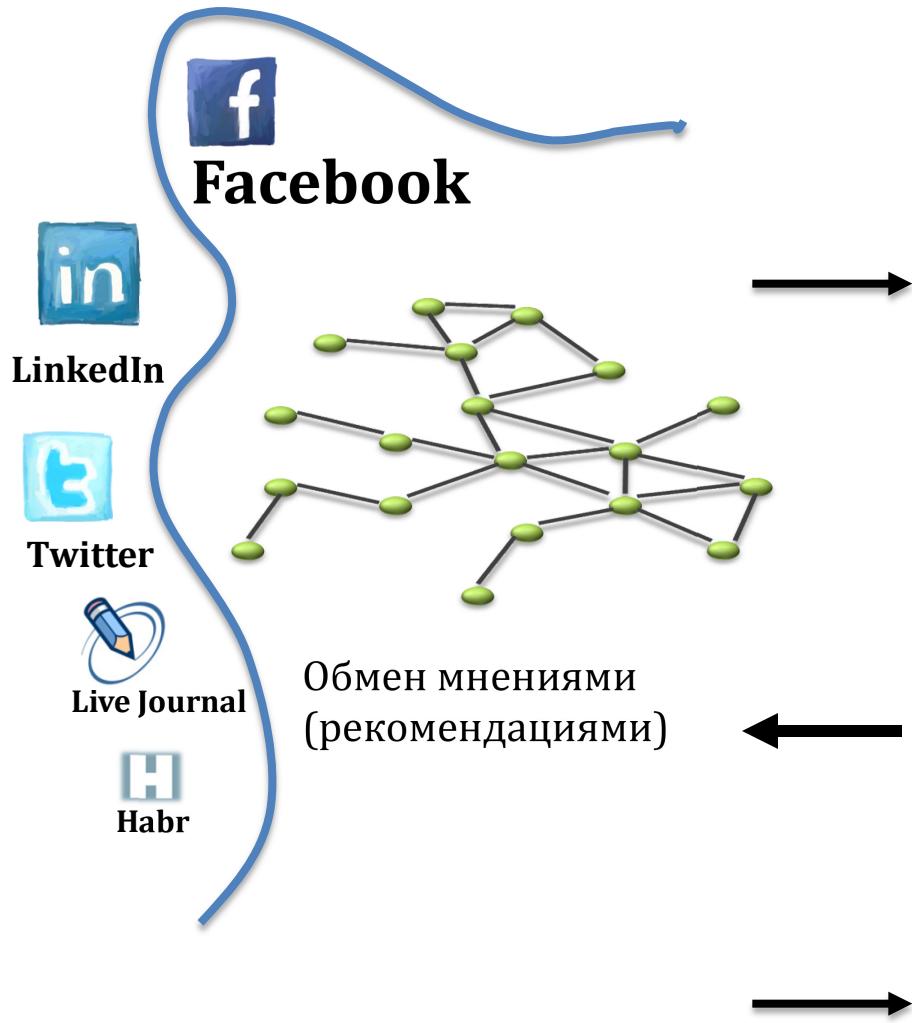


*Информационное управление,
Институциональное управление,
Мотивационное управление*



*Рис. Объект управления,
описываемый сетью*

Пример управления в социальных сетях

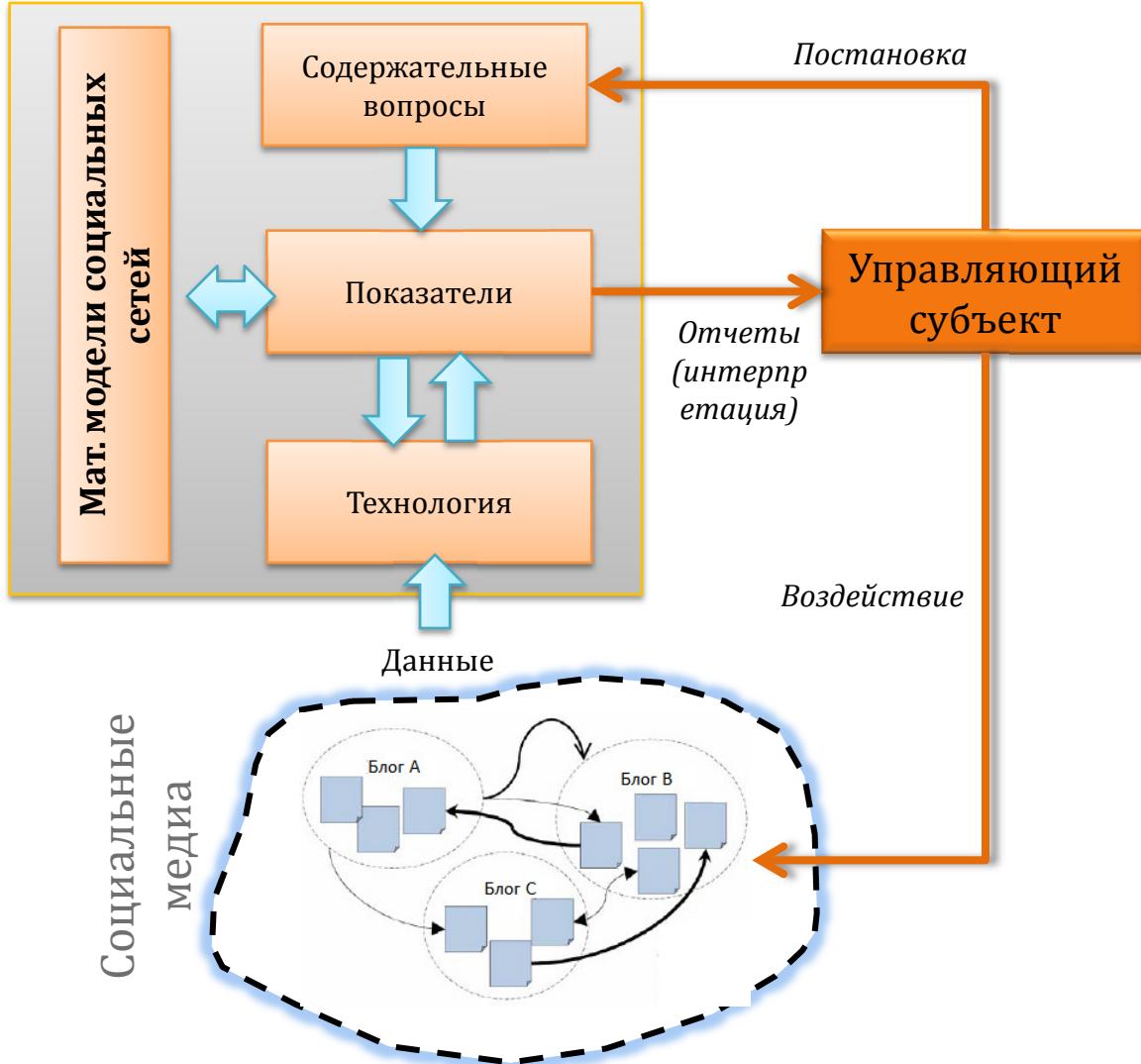


Субъект управления (цель – создать требуемое мнение в сети)

- Использует модели социальных сетей и алгоритмы для поиска влиятельных пользователей в сети или для увеличения влияния уже имеющихся агентов
- Оказывает воздействие на влиятельных пользователей
- Требуемое общественное мнение

Подход к анализу социальных медиа

Подход:



Уровни подхода:

- Определение списка содержательных вопросов.
- Разработка моделей и методов социальных сетей.
- Определение показателей и системы показателей для ответа на содержательные вопросы.
- Разработка технологии анализа социальных медиа.

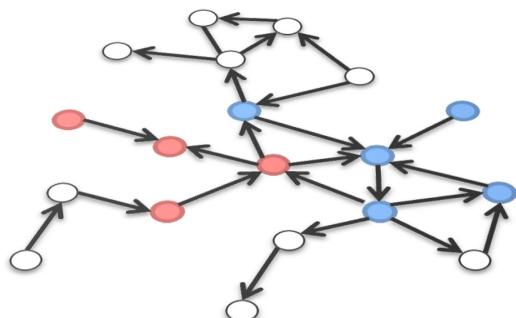
МОДЕЛИРОВАНИЕ СОЦИАЛЬНЫХ МЕДИА ПРИКЛАДНЫЕ ИССЛЕДОВАНИЯ БЛОГОСФЕРЫ

Основные подходы к моделированию социальных сетей

1. Моделирование структуры социальной сети

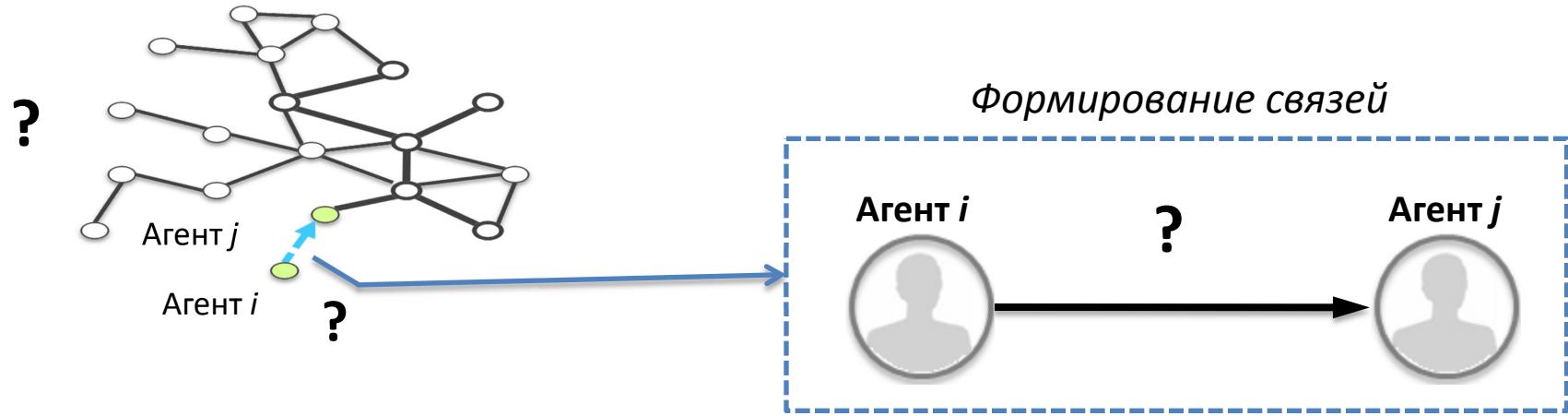
2. Моделирование распространения информации

3. Моделирование формирования мнений



Основные подходы к моделированию социальных сетей: Моделирование структуры социальной сети

Структура социальной сети – граф $\Gamma \langle V, E \rangle$



Ключевые характеристики реальных социальных сетей

Глобальные:

- Диаметр/среднее расстояние: малый/тесный мир (~ 6)
- Распределение степеней: степенное распределение $P_k = ck^{-\gamma}$, где $\gamma > 0$ ($2 \leq \gamma \leq 3$) и $c > 0$ – константы.

Локальные:

- Степень узла, центральность узла.
- Распределение узлов в эгоцентрической сети.

Основные подходы к моделированию социальных сетей: Моделирование структуры социальной сети

Класс оптимизационных и имитационных моделей

Класс теоретико-игровых моделей

Статические модели

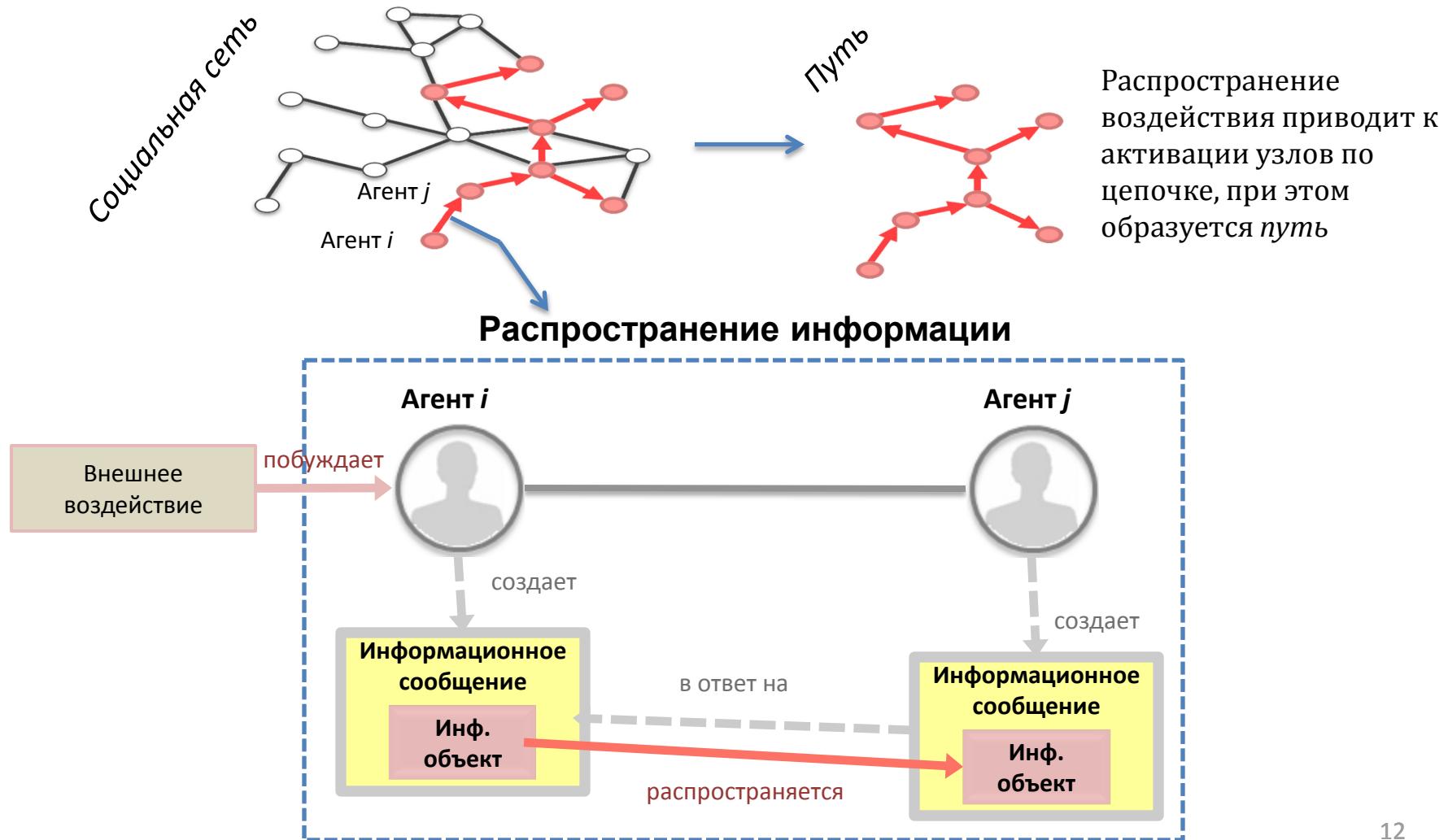


Динамические модели

Модели, рассматривающие предпочтительность для узлов тех или иных конфигураций связей

Основные подходы к моделированию социальных сетей: Моделирование распространения информации

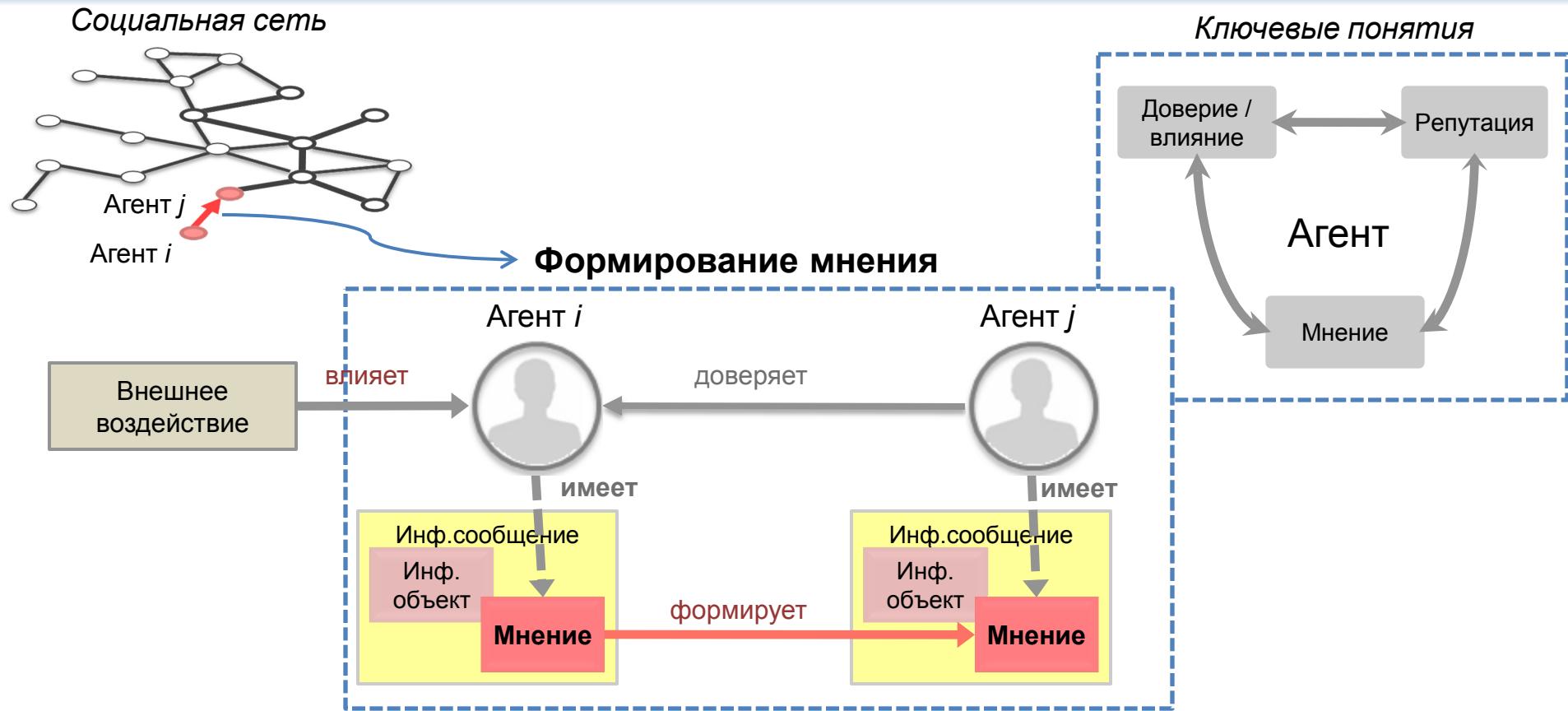
Распространение информации – процесс, посредством которого некоторый информационный объект распространяется по коммуникационным каналам во времени и в пространстве среди узлов сети



Основные подходы к моделированию социальных сетей: Моделирование распространения информации



Основные подходы к моделированию социальных сетей: Моделирование формирования мнений



Ключевые характеристики реальных социальных сетей

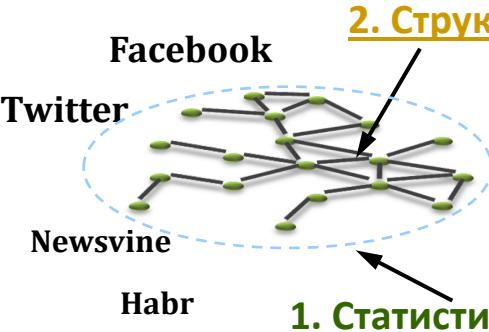
- Наличие собственных мнений агентов.
- Изменение мнений под влиянием других членов социальной сети.
- Различная значимость мнений (влияния, доверия) одних агентов для других агентов.
- Воздействие структурных свойств социальных сетей на динамику мнений

Основные подходы к моделированию социальных сетей: Моделирование формирования мнений

Множество
управляющих
субъектов M



Множество
управляемых
субъектов N



2. Структурный анализ

1. Статистический анализ

5. Модели информационного противоборства

Каждый игрок из множества M имеет возможность влиять на начальные мнения агентов u_{ij} и заинтересован в формировании итоговых мнений X_M .
Задача – найти равновесные действия игроков в игре

$$\Gamma = (M, \{U_j\}_{j \in M}, \{G_j(\cdot)\}_{j \in M}).$$

4. Модели информационного управления

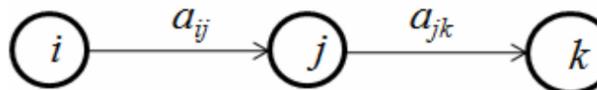
Задача – найти такой вектор управлений u , что:

$$\Phi(X, u) = H(X) - c(u) \rightarrow \max_{u \in U},$$

где $H(\cdot)$ – выигрыш, $c(\cdot)$ – затраты на управление.

3. Модели информационных взаимодействий

Агенты из N образуют социальную сеть $G = (N, E)$.
Вектор начальных мнений x , конечных X .



$a_{ij} \geq 0$ – степень доверия i -го агента j -му,
 k -й агент косвенно влияет на i -го.

$$x^{k+1} = A [x^k + B u^k].$$

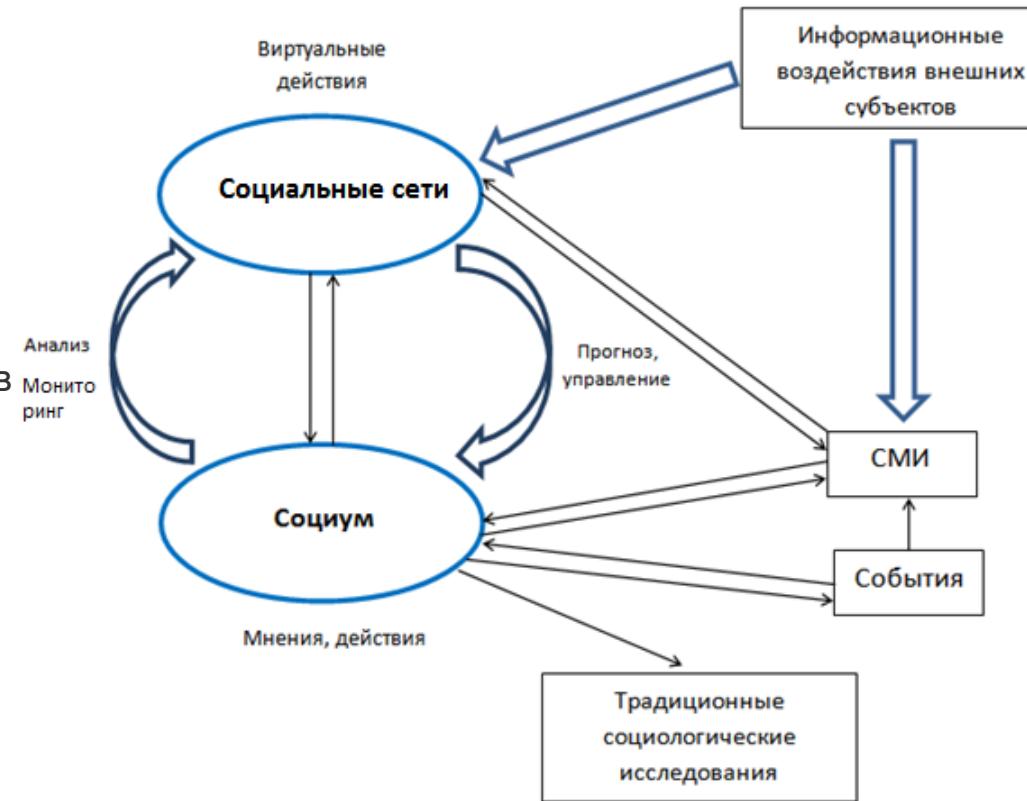
Задача – найти результирующее
влияние одних агентов на других;
найти агентов, формирующих
итоговое мнение в сети.



Задачи прикладных исследований

Мониторинг, прогноз и анализ

- мониторинг развития существующих и появления новых актуальных обсуждаемых тем
- определения характера обсуждения
- выявление активных групп и сообществ
- социо-демографический анализ сообществ
- определение групп согласованно действующих пользователей
- определение инициатора обсуждения
- оценка влиятельности пользователей
- оценка естественности обсуждения темы
- прогнозирование роста популярности тем
- деанонимизация пользователей
- определение ботов...

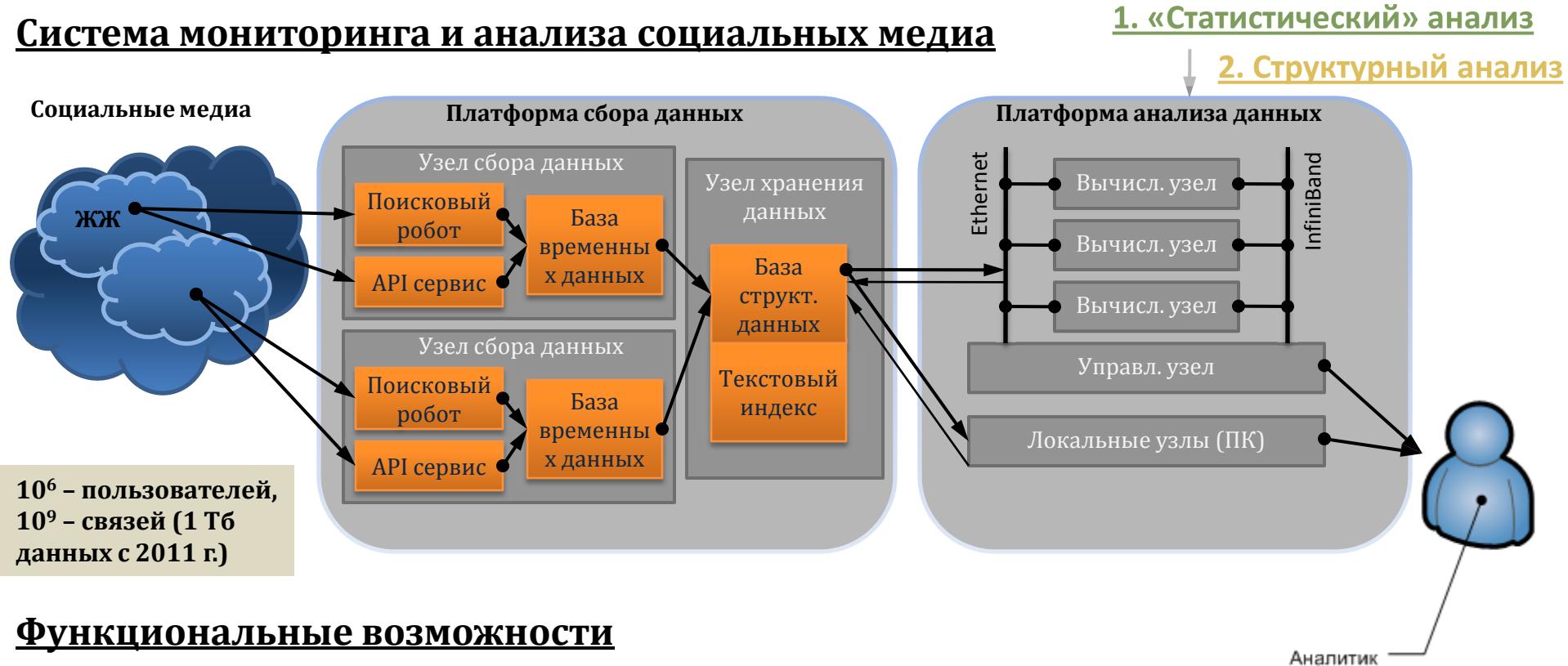


Управление и противоборство

- рекомендации по формированию общественного мнения
- создание виртуальных пользователей социальных сетей
- поддержка и поднятие рейтинга виртуальных пользователей в заданных социальных группах
- ...

Технологии мониторинга и анализа социальных медиа

Система мониторинга и анализа социальных медиа



1. «Статистический» анализ

2. Структурный анализ

Функциональные возможности

1. Анализ закономерностей информационных процессов

- ✓ Количество тематических авторов
- ✓ Память социальных медиа
- ✓ «Естественность» обсуждений
- ✓ Смена режимов написания сообщений

2. Анализ взаимовлияния социальных медиа и внешнего мира

- ✓ Связь онлайн-активности и реальности
- ✓ Рейтинг СМИ в соц. медиа
- ✓ Влияние СМИ на обсуждения в соц. медиа
- ✓ «Неестественность» присутствия СМИ в соц.медиа

3. Анализ закономерностей внутренней структуры

- ✓ Структура сети взаимодействий
- ✓ Устойчивость сети к внешним воздействиям
- ✓ Влиятельность пользователей
- ✓ Взаимовлияние пользователей
- ✓ И многое другое..

Перспективы

- Последовательное и системное развитие: *Математические модели социальных медиа, Математические методы анализа социальных медиа → Алгоритмы → Программы*
- Разработка комплексной модели динамики социальных медиа (привлечение социологов и психологов)
- Разработка методов анализа социальных медиа с учетом мнений агентов
- Разработка технологий для социальных медиа:
«Большие Данные», «Большие Вычисления», «Большая Визуализация»

Администрация Президента США обязала профинансировать NSF исследовательские проекты по Big Data (22 октября 2012 г).

"The awards were among eight Big Data fundamental research projects, totaling \$15 million, developed in response to the federal Big Data Initiative announced last spring."

МОДЕЛИРОВАНИЕ СОЦИАЛЬНЫХ МЕДИА

ПРИКЛАДНЫЕ ИССЛЕДОВАНИЯ БЛОГОСФЕРЫ

Прикладные исследования блогосферы

1. ЗАКОНОМЕРНОСТИ ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

- Характерные параметры обсуждений заданных тем (число блогеров, параметры активности и т.д.)
- Аномальные отклонения в обсуждениях
- «Память» блогосферы (автокорреляции, характерное время обсуждения события)

2. ВЛИЯНИЕ ВНЕШНЕГО МИРА НА БЛОГОСФЕРУ

- Источники информации в блогосфере (популярность, классификация)
- Реакция блогеров на события внешнего мира

3. СТРУКТУРА БЛОГОСФЕРЫ

- Составление «карты» блогосферы относительно заданных предметных областей
- Устойчивость к удалению вершин и связей

Исходные данные

Источник данных:

«живой журнал» (ЖЖ, <http://www.livejournal.com>).

Информационные объекты:

текстовый контент блогов, конкретно постов.

Период: 11 ноября 2011 г. по 29 июля 2012 г.

Выборка данных – четыре запроса:

Q1 – «Путин \ Медведев»

Q2 – «Медведев \ Путин»

Q3 – «Путин AND Медведев»

Q4 – «Путин OR Медведев»

Объём ЖЖ: ~ 2.5 млн. аккаунтов

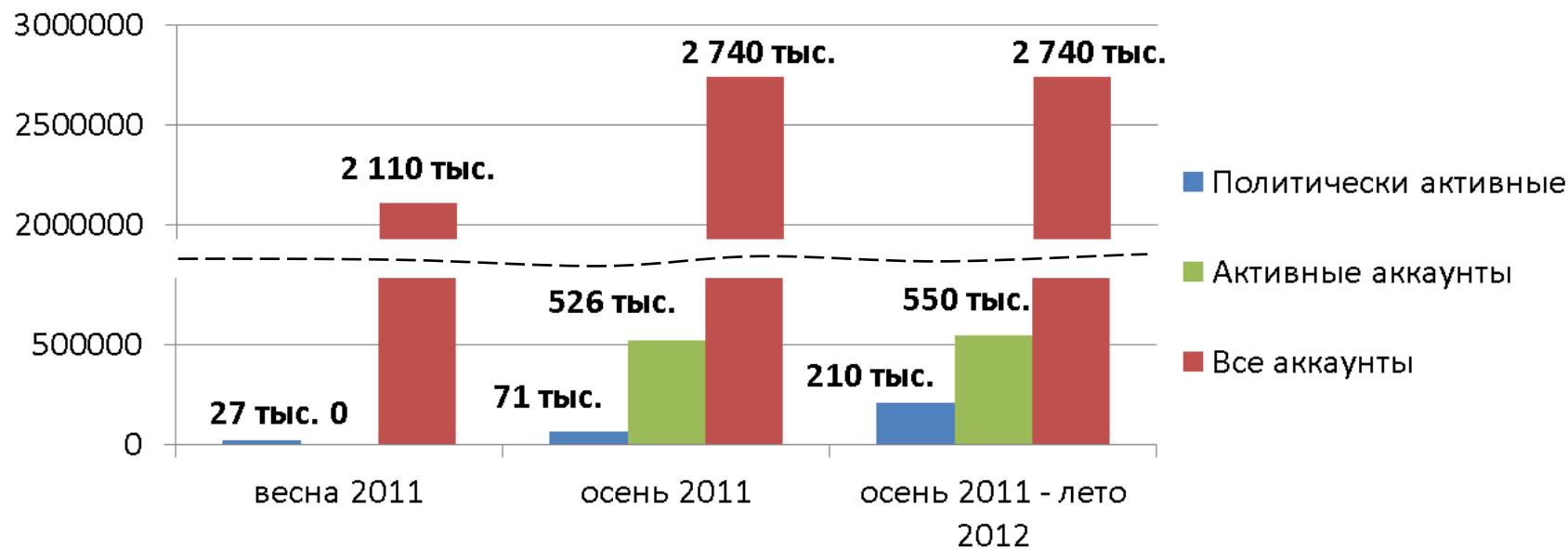
Объём активных аккаунтов: ~ 550 тыс.

Объём политически активных: ~ 210 тыс.

Рост числа политически активных блогеров

Критерий политической активности – упоминание Путина или Медведева хотя бы в одном сообщении

Число блогеров на конец рассматриваемого периода



К лету 2012 года политически активные блогеры составляют 38% всех активных аккаунтов (13% осенью)

Исследование динамики политических дискуссий

Задачи:

- определить характер политических обсуждений в блогосфере, определить моменты наибольшей активности
- выявить различия в обсуждениях Путина и Медведева
- проверить наличие неестественных компонент
- оценить «память» блогосферы – влияние прошлых значений на текущие обсуждения, характерное время обсуждения событий

Результаты:

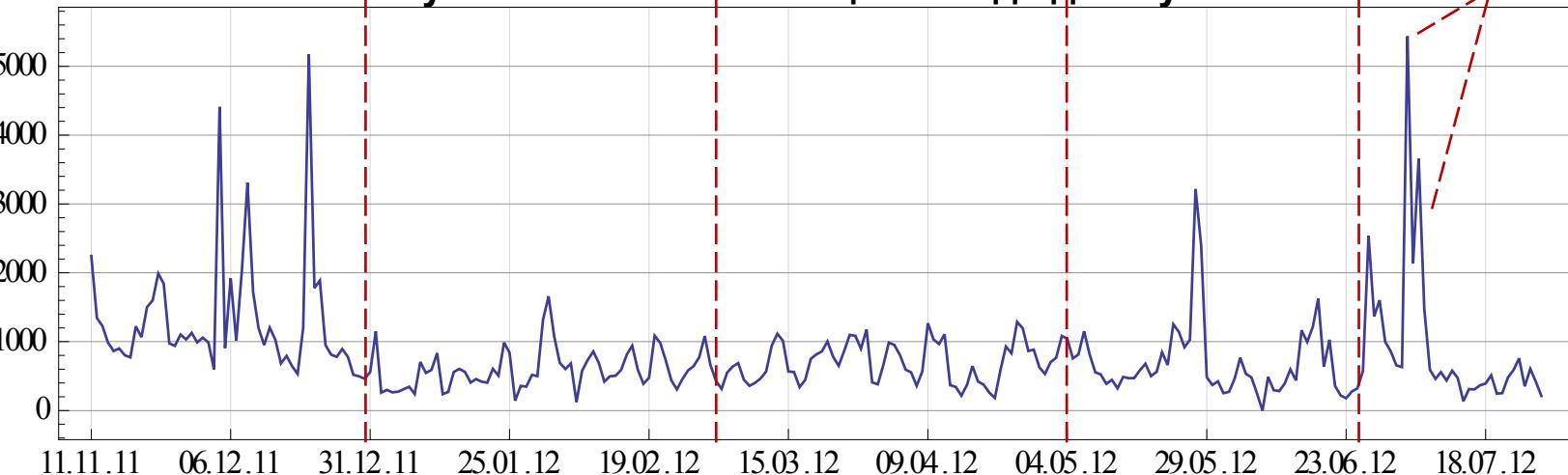
- Характер обсуждений Путина и Медведева за рассматриваемый период отличается количественно (авторов, пишущих исключительно о Путине, в 2.37 раз больше), так и качественно (разный тип распределений)
- В обсуждениях обоих субъектов значительная доля неестественной реакции
- Длительность «памяти»: 30 дней как для Путина, так и для Медведева
- Характерное время обсуждения 10-11 дней (осенью и зимой)

Динамика по сообщениям

Суточное количество сообщений : Путин \ Медведев

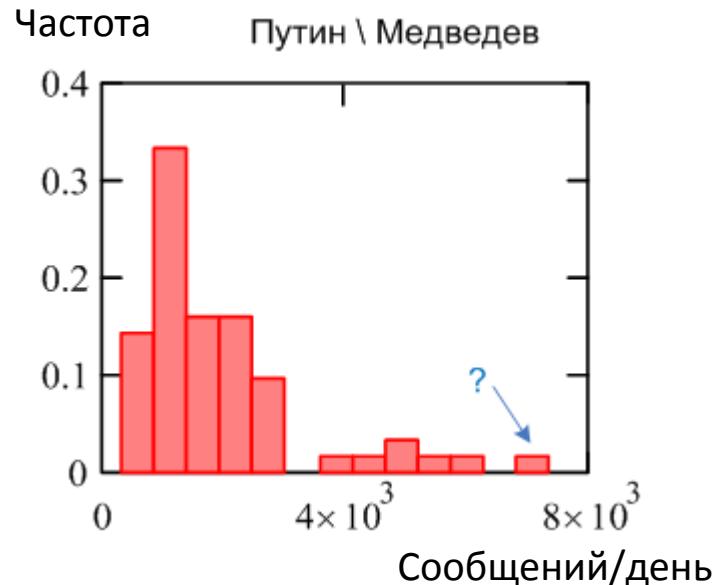


Суточное количество сообщений : Медведев\ Путин



Характеристики обсуждений политических субъектов

Гистограммы обсуждений за период с 12.11.2011 г. по 13.01.2012

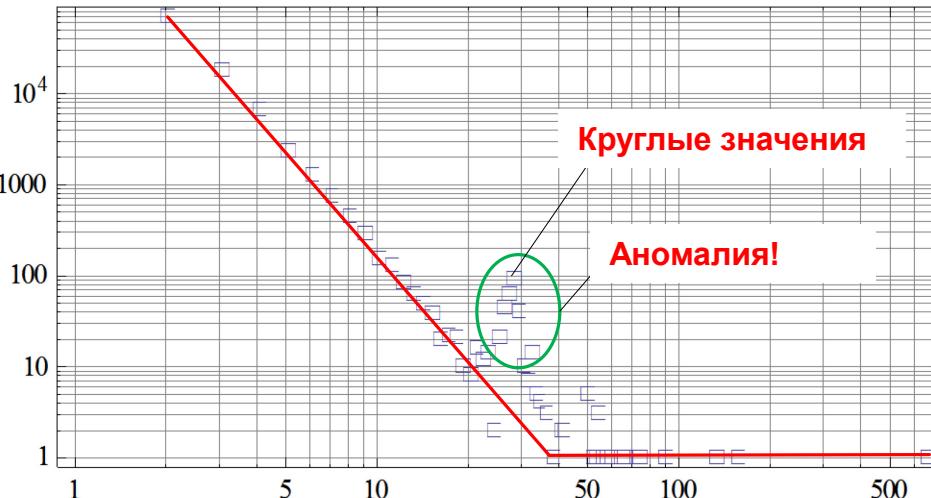


Производительность авторов постов

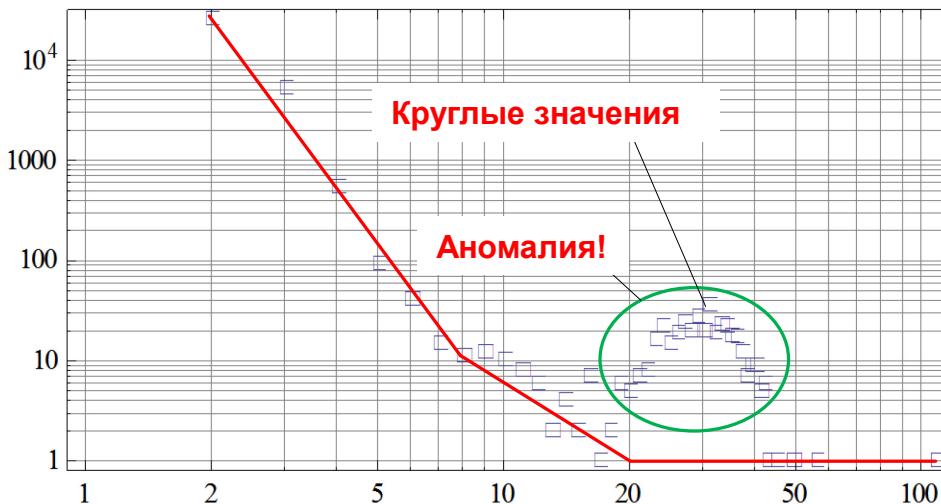


Активность политических авторов

Гистограмма активности уникальных авторов : Путин \ Медведев



Гистограмма активности уникальных авторов : Медведев \ Путин



По оси x - Всего написано сообщений

По оси y – количество авторов, написавших столько сообщений

Всплески у «круглых» значений:

Путин \ Медведев

100 блог., по 25 сообщ. каждый;

Медведев \ Путин

40 блог., 30 сообщ. каждый.

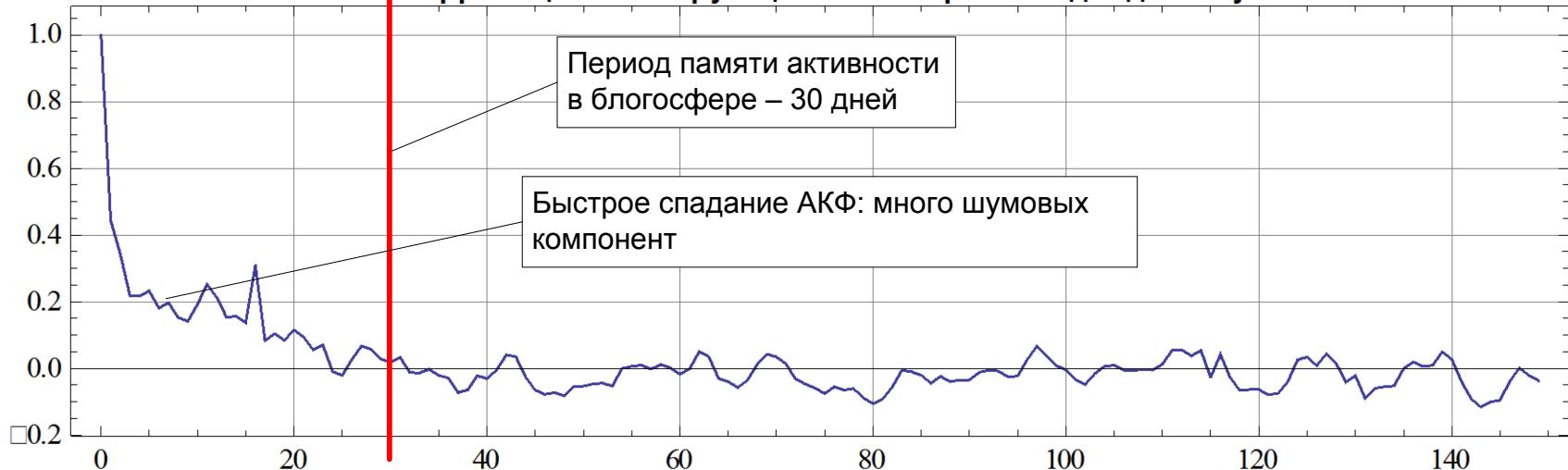
Объём предположительно «искусственной компании», проводимой в адрес тех или иных политиков.

Память блогосферы

Автокорреляционная функция по авторам: Путин \ Медведев



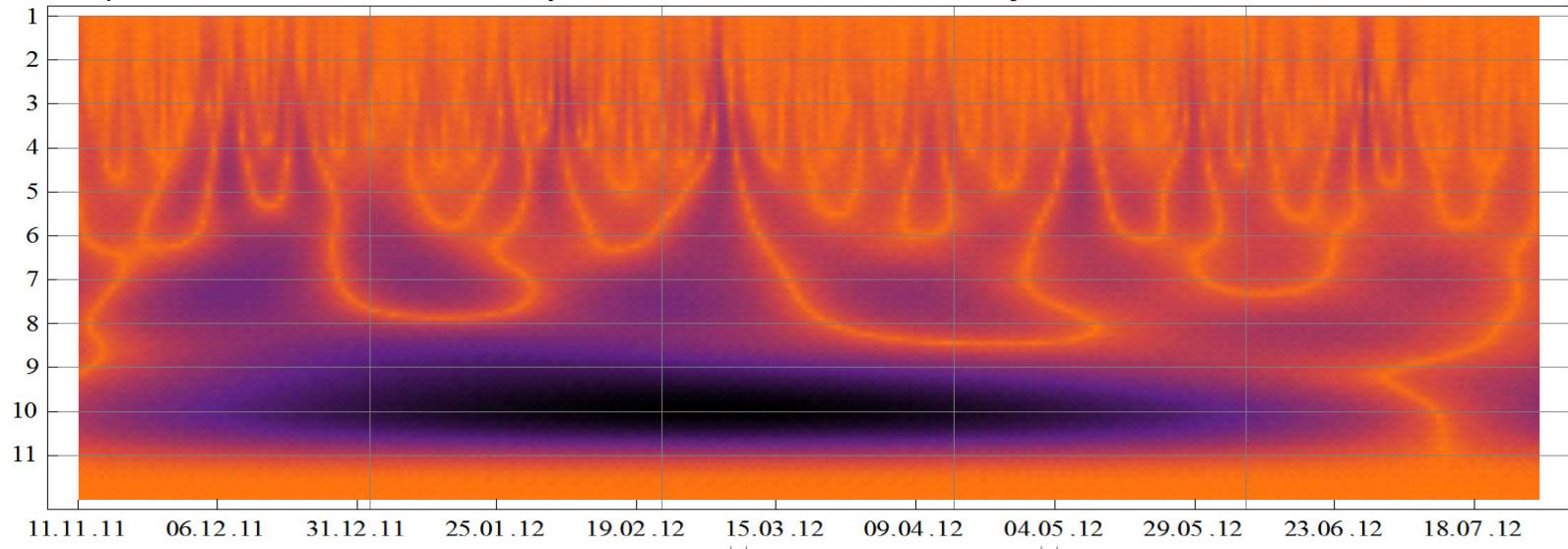
Автокорреляционная функция по авторам: Медведев \ Путин



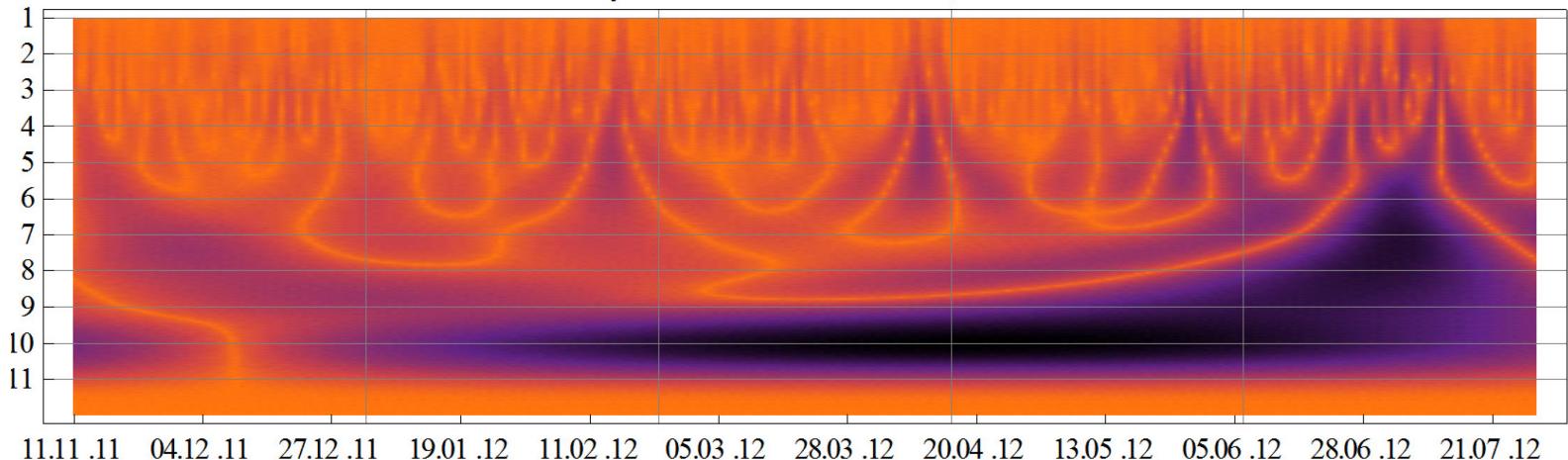
Длительность обсуждений инфоповодов

Длительность
обсуждения, сутки

Вейвлет скалограмма | Объём сообщений | Путин или Медведев



Вейвлет скалограмма | Объём сообщений | Все посты



Все инфоповоды обсуждаются не более 11 дней!

Взаимодействие блогосферы с внешним миром

Задачи

- Рейтинг источников информации, популярных в течение всей избирательной кампании.
- Классификация ресурсов по оппозиционности и идеологической направленности. Группировка в кластеры по общим признакам.
- Реакция блогеров на события внешнего мира и сообщения информационных источников

Результаты

- Популярность сайтов подчиняется степенному закону. Есть небольшая доля наиболее популярных сайтов
- Большинство популярных сайтов (55%) – оппозиционные СМИ
- Есть признаки неестественной активности аудитории некоторых сайтов
- Есть 11 ресурсов, слабо влияющих на общий уровень активности (корреляция >0.2), из них 3 наиболее влиятельны (корреляция >0.38)
- Есть слабая связь (корреляция 0.3) активности блогеров с интенсивностью событий в официальных источниках

Популярные источники

Путин / Медведев

№	Название	Блогеров / сутки	Общее кол-во сообщений
1	kp.ru	3,46	4893
2	lenta.ru	2,72	3752
3	ria.ru	2,35	4083
4	echo.msk.ru	2,23	3026
5	gazeta.ru	1,90	2650
6	kommersant.ru	1,59	2358
7	novayagazeta.ru	1,52	2253
8	inosmi.ru	1,46	1799
9	newsru.com	1,41	3495
10	vedomosti.ru	1,35	1001
11	odnako.org	1,34	2907
12	nstarikov.ru	1,04	1257
	...		
55	expert.ru	0,30	84
	...		
87	rusrep.ru	0,24	75

Медведев / Путин

№	Название	Блогеров / сутки	Общее кол-во сообщений
1	ria.ru	3,16	1005
2	lenta.ru	3,02	883
3	echo.msk.ru	2,60	2153
4	gazeta.ru	1,99	571
5	kommersant.ru	1,94	551
6	kp.ru	1,86	548
7	novayagazeta.ru	1,62	716
8	newsru.com	1,57	451
9	vz.ru	1,51	635
10	rbcdaily.ru	1,09	293
11	top.rbc.ru	1,09	348
12	podkontrol.ru	1,07	289
	...		
86	forbes.ru	0,25	64
	...		
101	expert.ru	0,20	52

Первые 7 источников, встречающиеся в обсуждениях обоих факторов, совпадают.

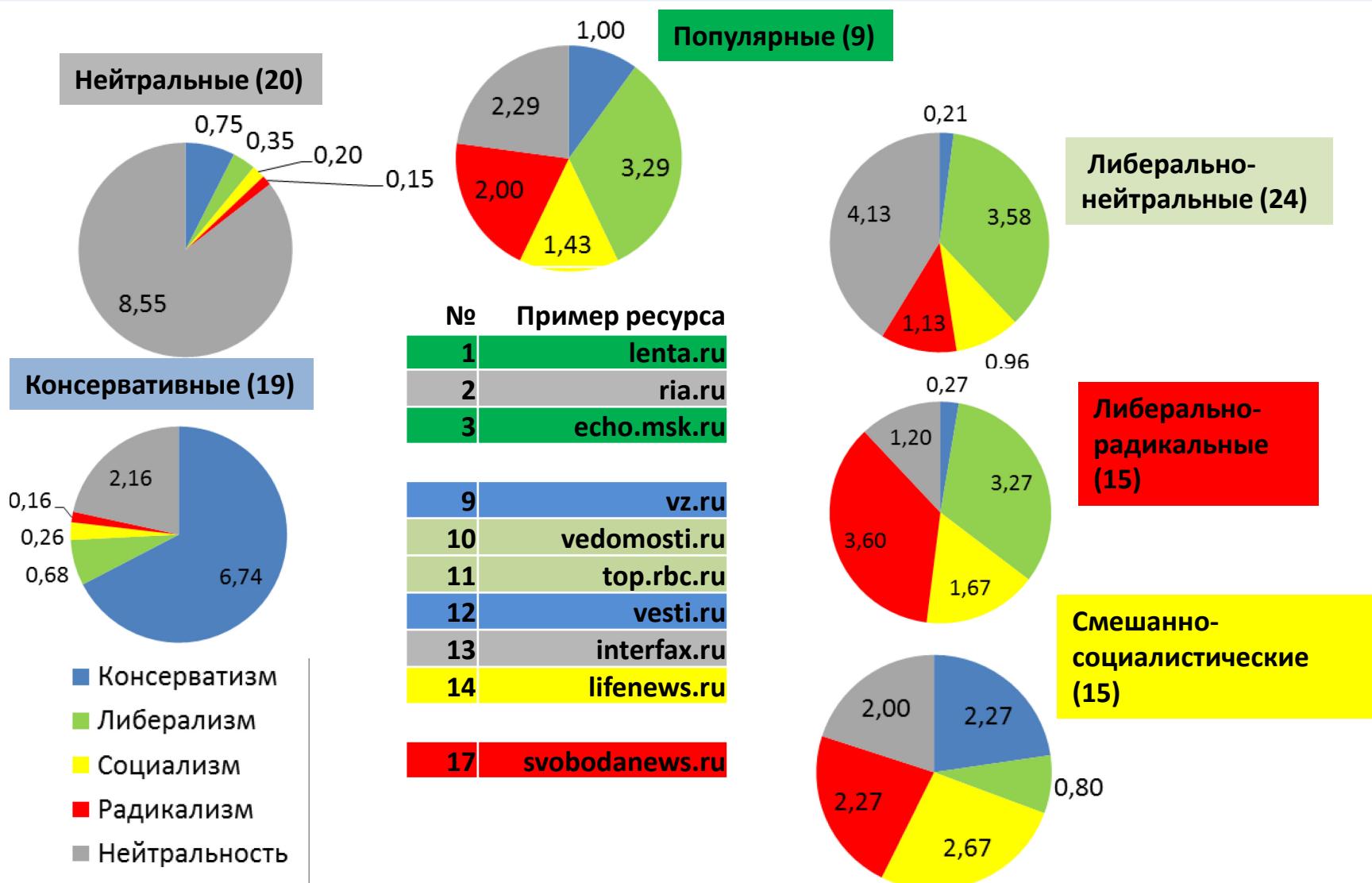
Основные отличия – в наборе лояльных власти сайтов, которые «специализируются» на разных политических субъектах

Классификация популярных источников

Наборы источников, в обсуждениях всех четырех факторов отличаются только по показателю «Либерализм».

В обсуждениях, с упоминанием Путина цитируется больше источников, либеральной направленности, чем в обсуждениях Медведева

Кластеры источников



Реакция на сообщения в СМИ

Корреляции с задержкой в 0 суток

№	Коэффициент корреляции Пирсона	URL СМИ
1	0.4086	gazeta.ru
2	0.4024	kommersant.ru
3	0.3807	ria.ru
4	0.2839	ridus.ru
5	0.2660	nstarikov.ru
6	0.2638	svobodanews.ru
7	0.2361	newsland.ru
8	0.2305	echo.msk.ru
9	0.2198	lifenews.ru
10	0.2024	rg.ru
11	0.2022	izvestia.ru

Задержки, для которых достигается максимум корреляции

№	Коэффициент корреляции Пирсона	Задержка, сутки	URL СМИ
1	0.4086	0	gazeta.ru
2	0.4024	0	kommersant.ru
3	0.3807	0	ria.ru
4	0.3470	14	bbc.co.uk
5	0.3390	2	ridus.ru
6	0.2802	24	kp.ru
7	0.2745	21	newsland.ru
8	0.2638	0	svobodanews.ru
9	0.2604	12	lifenews.ru

Есть 11 ресурсов, цитируемость которых имеет корреляцию с общим объемом сообщений выше 0.20.

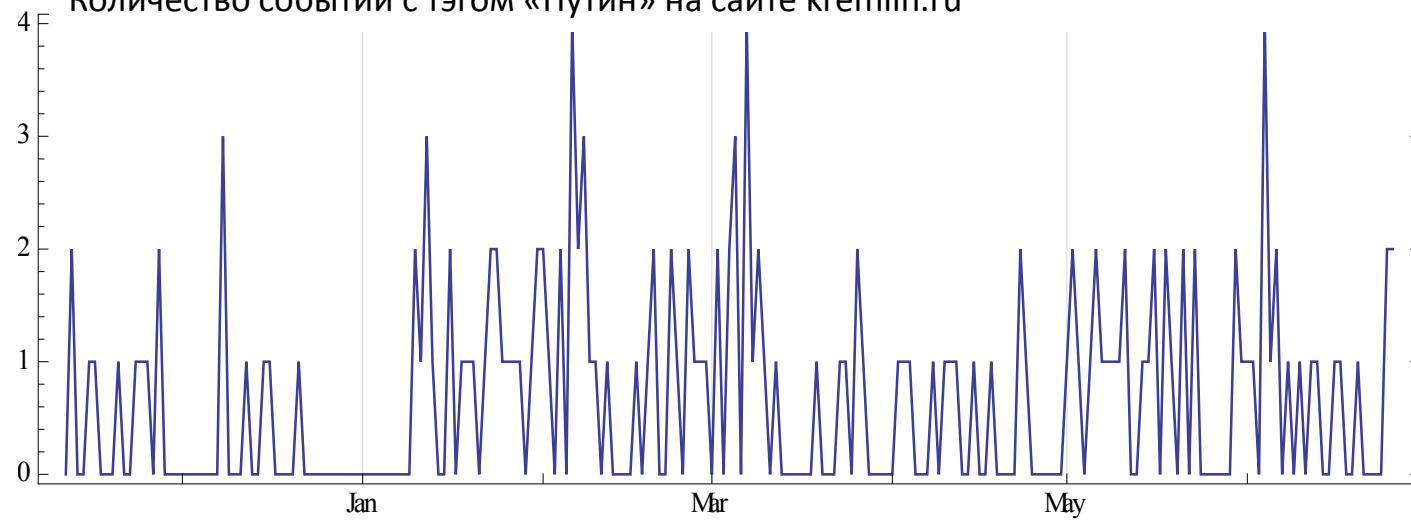
3 ресурса имеют корреляцию выше 0.35, причем с задержкой в 0 суток

Эти ресурсы поставляют наиболее актуальную или интересную информацию о событиях, либо имеют существенное влияние на политически активные группы блогеров.

Реакция на события внешнего мира



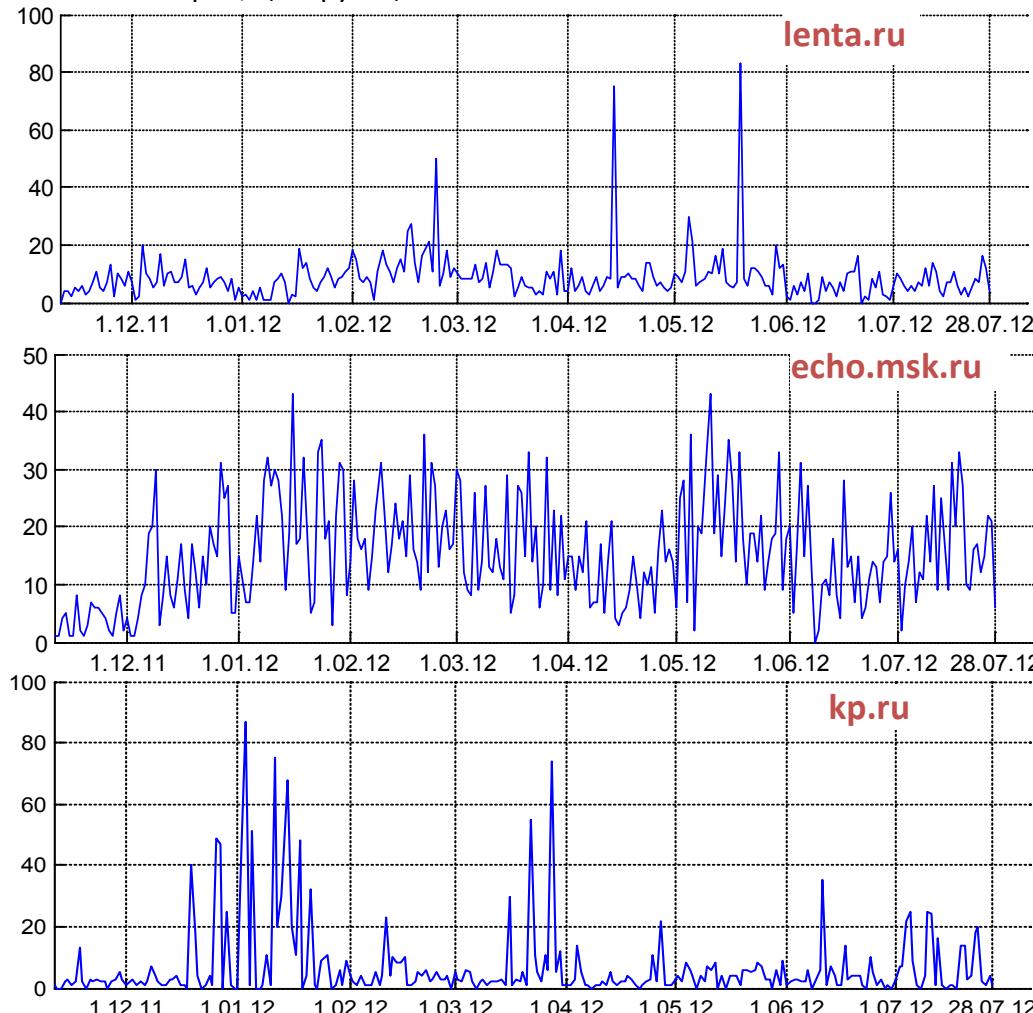
Количество событий с тэгом «Путин» на сайте kremlin.ru



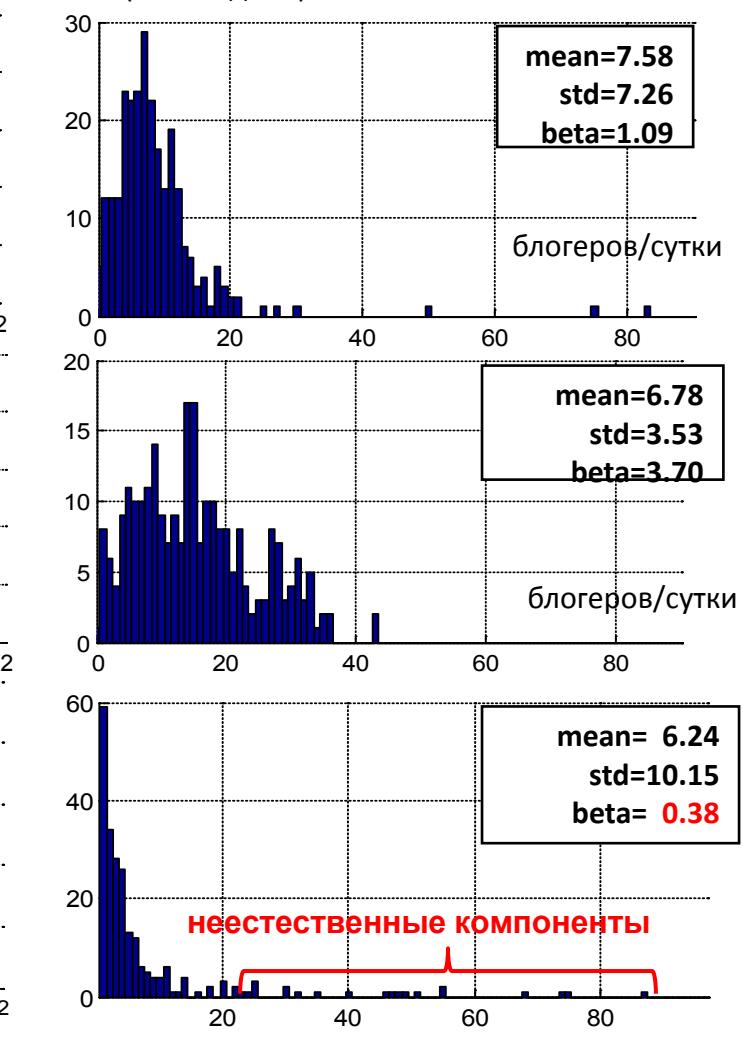
Корреляция 0.3026 – слабая корреляция с официальными источниками

Динамика новостных источников

Количество блогеров, цитирующих источник



Частота (кол-во дней)



Среди 10 популярных источников только аудитория Комсомольской правды имеет неестественное поведение

Исследование структуры блогосферы по отношению к политическим обсуждениям

Задачи

- Классификация популярных блогеров по политическим взглядам
- Исследование структуры связей блогеров из разных групп
- Поиск сильно связанных сообществ (тематических)

Результаты

- Первые 300 популярных блогеров классифицированы по трем основаниям. Выявлено три характерных кластера блогеров, отсутствует кластер явных сторонников власти
- Построена сеть связей комментирования между группами блогеров разной степени оппозиционности
- Исследована динамика влиятельности разных групп блогеров. Наблюдается спад влияния радикальных групп блогеров (как оппозиционных, так и лояльных)
- Выявлена устойчивость главной компоненты сети к удалению узлов

Классификация 300 популярных блогеров (рейтинг ЖЖ + Яндекс)

Основания классификации

Отношение к власти. Системные провластные (3), лояльные организованные (2), провластно настроенные (1), нейтральные (0), оппозиционно настроенные обыватели (-1), системная оппозиция (-2), внесистемная оппозиция (-3).

Содержание идеологии. Крайне левые (-2), левые умеренные (-1), центристы и консерваторы (0), социал-либералы (1), радикальные демократы и либералы-рыночники (2).

Националистический настрой. Ярко выраженные националистические (2), слабо выраженные националистические (1), нейтральные или невыраженные (0), слабо выраженные антнационалистические (-1), ярко выраженные антнационалистические (-2).

Блогер	Отношение к власти	Идеологическая направленность	Национализм	Радикальность идей	Идентификация блогера/лейтмотив блога
13vainamoinen.livejournal.com	1	1	1	0	Алексей Востряков, Карельский иллюстрированный дневник
19viv69.livejournal.com	-1	1	0	0	Игорь; Разные вести - какие мысли?
abstract2001.livejournal.com	-3	-2	1	2	Марина Литвинович, российский политический деятель, политтехнолог, политолог, журналист, правозащитник, член исполкома коалиции «Другая Россия».
afranius.livejournal.com	-1	1	1	1	Кирилл Еськов, российский палеонтолог и писатель-фантаст
ailiarionov.livejournal.com	-3	2	-2	1	Андрей Илларионов, российский государственный деятель и экономист. Бывший советник президента В. Путина, ныне оппозиционер власти.
...

Результаты кластеризации популярных блогеров

I кластер (158 блогеров) - "политическая норма"

нейтральны
по отношению
к власти

консервативно
настроены

баланс
"идеологических
крайностей"

II кластер (99 блогеров) - "умеренно оппозиционный"

нет настроенных
провластно
блогеров

захвачен
весь спектр
идеологий

отсутствуют
проявления
радикальности

III кластер (69 блогеров) - "резко оппозиционный"

35 представителей
внесистемной
оппозиции
(navalny.livejournal.com,
aillarionov.livejournal.com)

58 сторонников
либеральной
идеологии

31 блогер
настроен
радикально

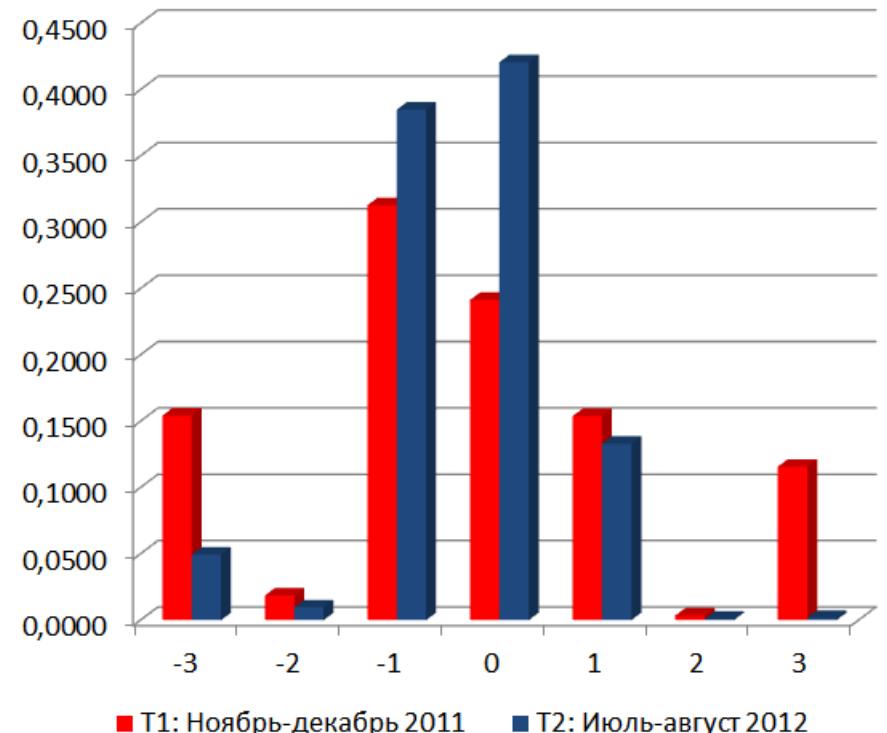
Анализ взаимовлияния 300 популярных блогеров

(«-» – оппозиционные, «+» – провластные)

Период Т1 (ноябрь-декабрь 2011)

	-3	-2	-1	0	1	2	3
-3	53	1	136	40	4	0	0
-2	2	1	27	10	1	1	1
-1	243	15	258	143	131	5	246
0	14	5	187	371	168	3	13
1	175	27	251	183	182	2	87
2	1	9	5	0	0	0	0
3	0	1	126	17	1	0	19

Влиятельность

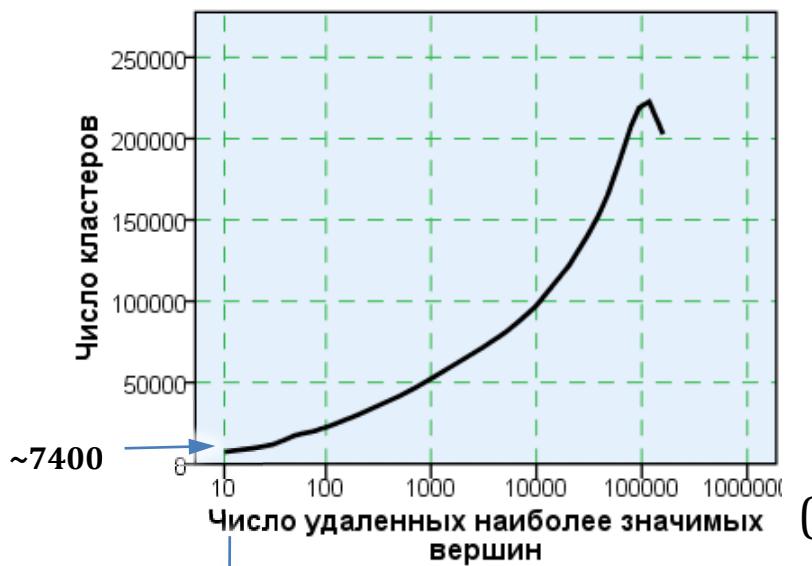


Период Т2 (июль-август 2012)

	-3	-2	-1	0	1	2	3
-3	36	3	143	67	16	0	1
-2	9	2	19	12	0	0	0
-1	172	18	642	802	202	1	5
0	34	16	989	944	340	0	3
1	25	9	326	530	186	4	1
2	0	6	12	0	0	0	0
3	1	0	32	9	3	0	0

Наблюдается спад влияния радикальных групп блогеров (как оппозиционных, так и лояльных) и усиление влиятельности умеренных

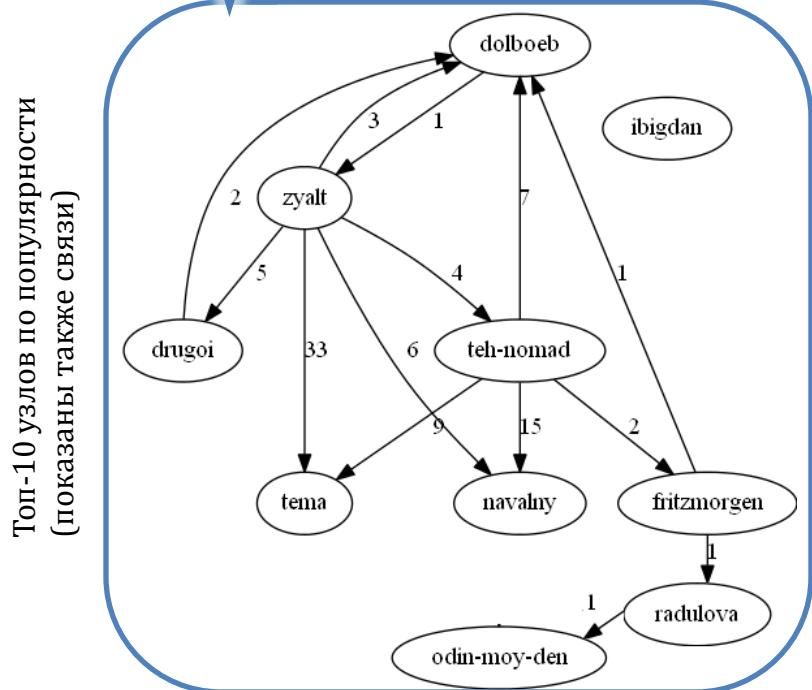
Устойчивость к целенаправленному удалению популярных блогеров



(A)



(B)



(C)

Удаление небольшого числа наиболее значимых узлов приводит к образованию большого числа несвязанных групп (A).

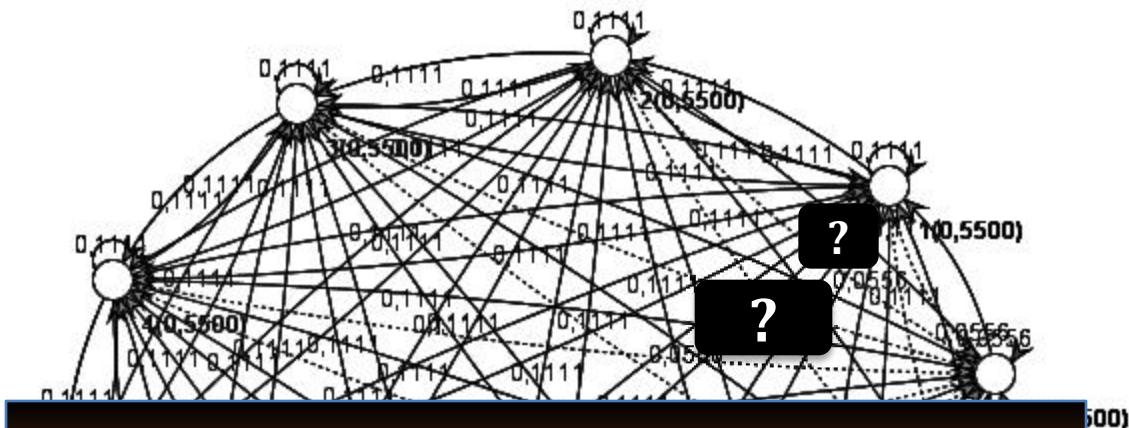
Однако эти группы по своему размеру малы, наибольшая группа остается связанной (B)

Блогосфера устойчива к последовательному и целенаправленному удалению наиболее популярных блогеров

Устойчивость к целенаправленному удалению блогеров (2)

Рассматриваются максимальные подсети из узлов с популярностью H и выше.





Спасибо за внимание!

