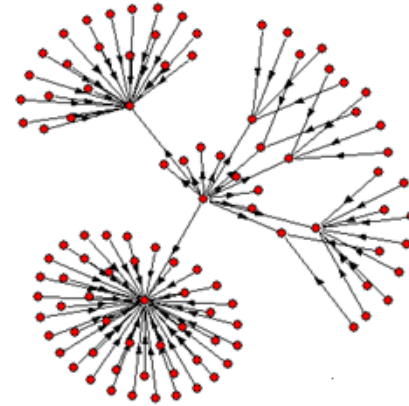




ИНСТИТУТ
ПРОБЛЕМ
УПРАВЛЕНИЯ
ИМ. В.А. ТРАПЕЗНИКОВА
РОССИЙСКОЙ АКАДЕМИИ НАУК



Прикладные задачи и методы анализа социальных сетей

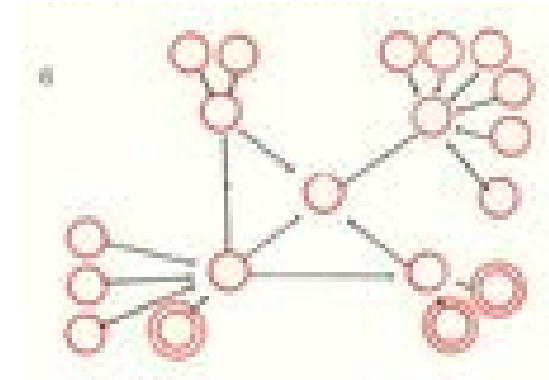
Николай Базенков
ИПУ РАН
n.bazenzkov@gmail.com

История исследований социальных сетей

1930-1950 Я.Л. Морено, Д. Барнс

Малые группы, замкнутые сообщества

Личные наблюдения



1960-1980 С. Милгрэм, М. Грановеттер

Универсальные свойства всего общества

Эксперименты, опросы

1990-2000 Watts, Strogatz, Barabasi, Albert, Jackson, Newman, et.al.

Автоматизированный сбор и обработка данных

Модели формирования сетей, распространения информации и др.



Задачи исследования

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

- Группы влиятельных блогеров
- Устойчивость к удалению вершин и связей

Исходные данные

Источник данных:

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

Информационный объект:

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

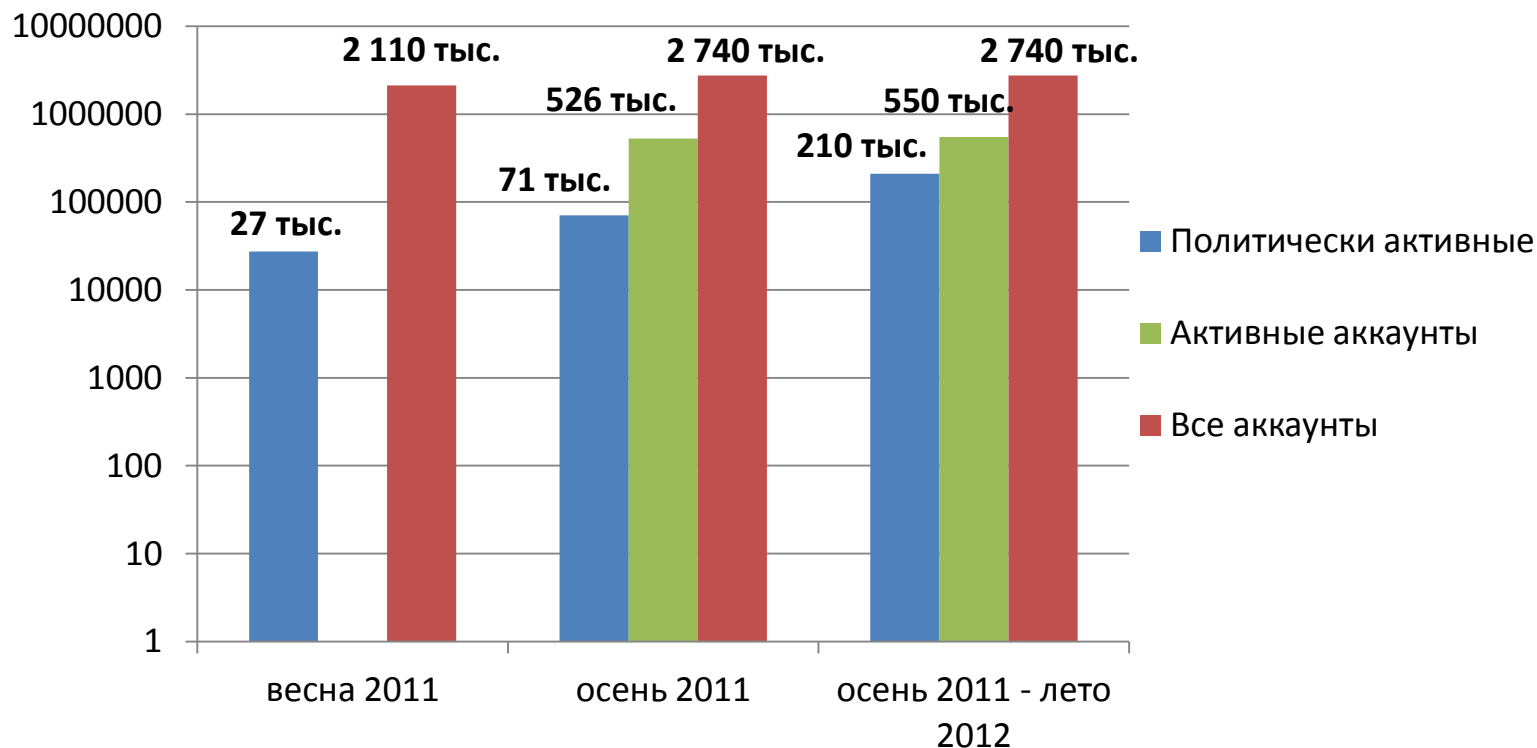
Объём активных аккаунтов: ~ 550 тыс.

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



Задачи исследования

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

- Группы влиятельных блогеров
- Устойчивость к удалению вершин и связей

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:

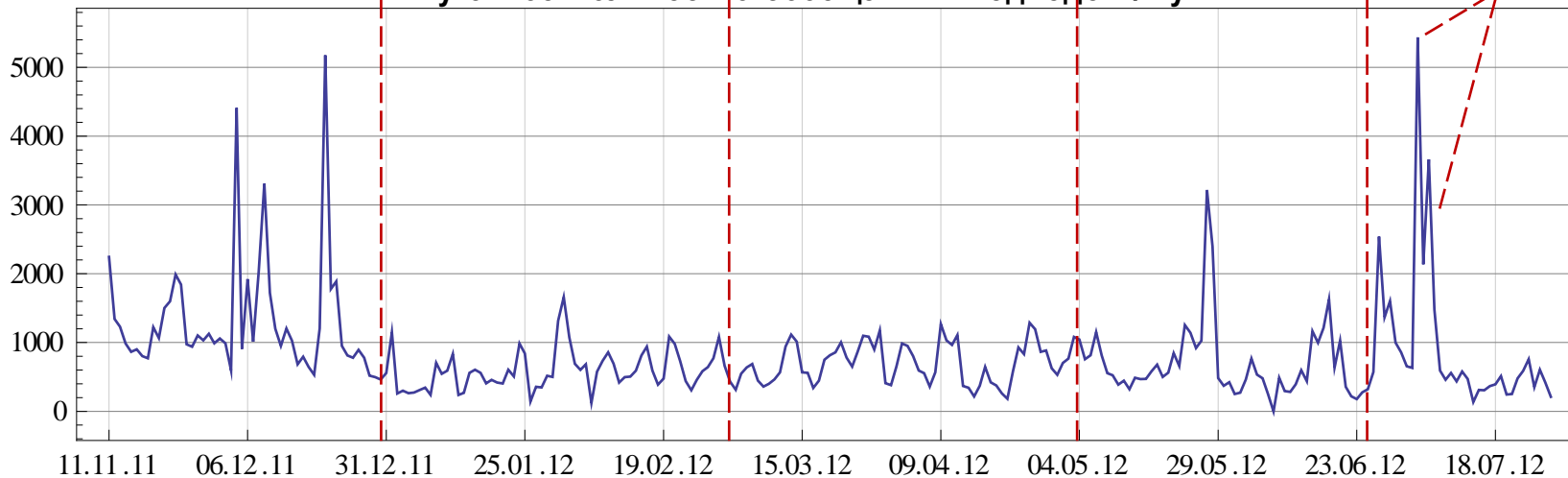
- Характер обсуждений Путина и Медведева за рассматриваемый период отличается количественно (авторов, пишущих исключительно о Путине, в 2.37 раз больше), так и качественно (разный тип распределений)
- В обсуждениях обоих субъектов значительная доля неестественной реакции
- Длительность «памяти»: 2 недели для Путина, 5 дней для Медведева
- Характерное время обсуждения 10-11 дней (осенью и зимой)

Динамика по сообщениям

Суточное количество сообщений : Путин \ Медведев

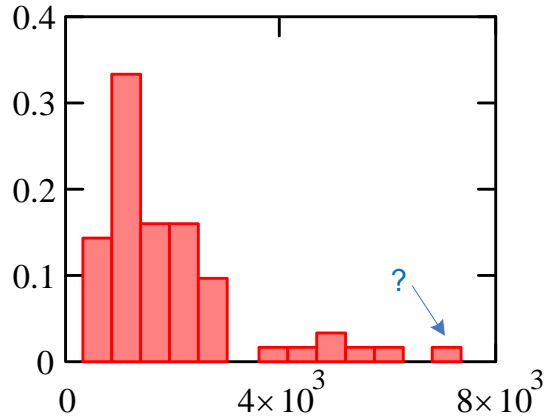


Суточное количество сообщений : Медведев \ Путин

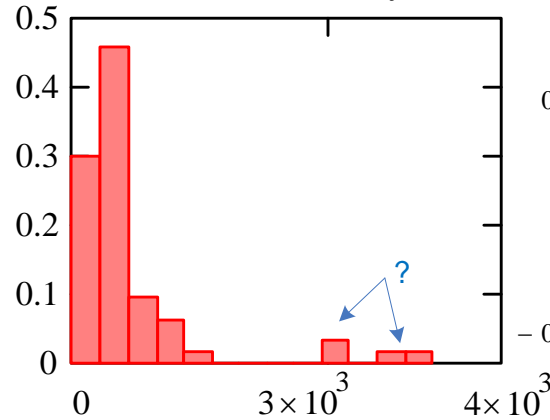


Характеристики обсуждений политических субъектов

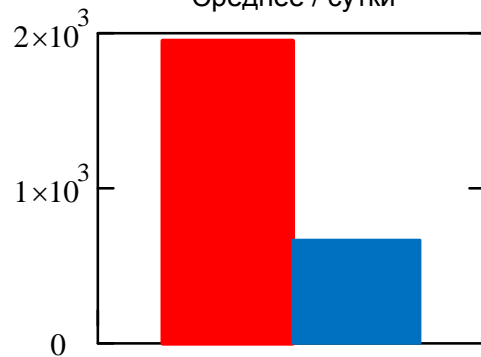
Q5 – Путин \ Медведев



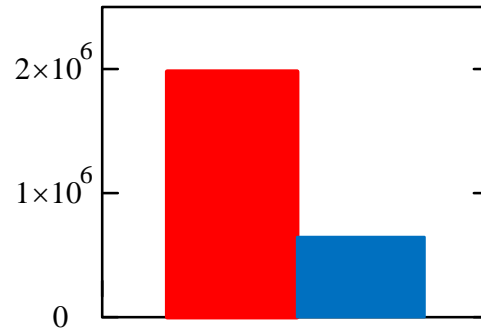
Q6 – Медведев \ Путин



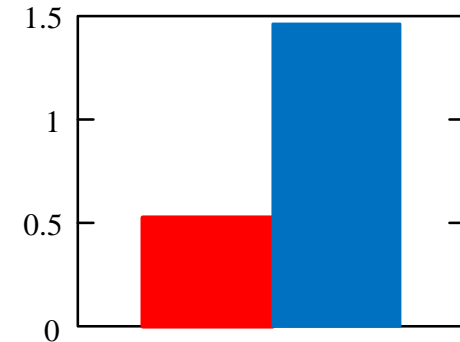
Среднее / сутки



Дисперсия / сутки



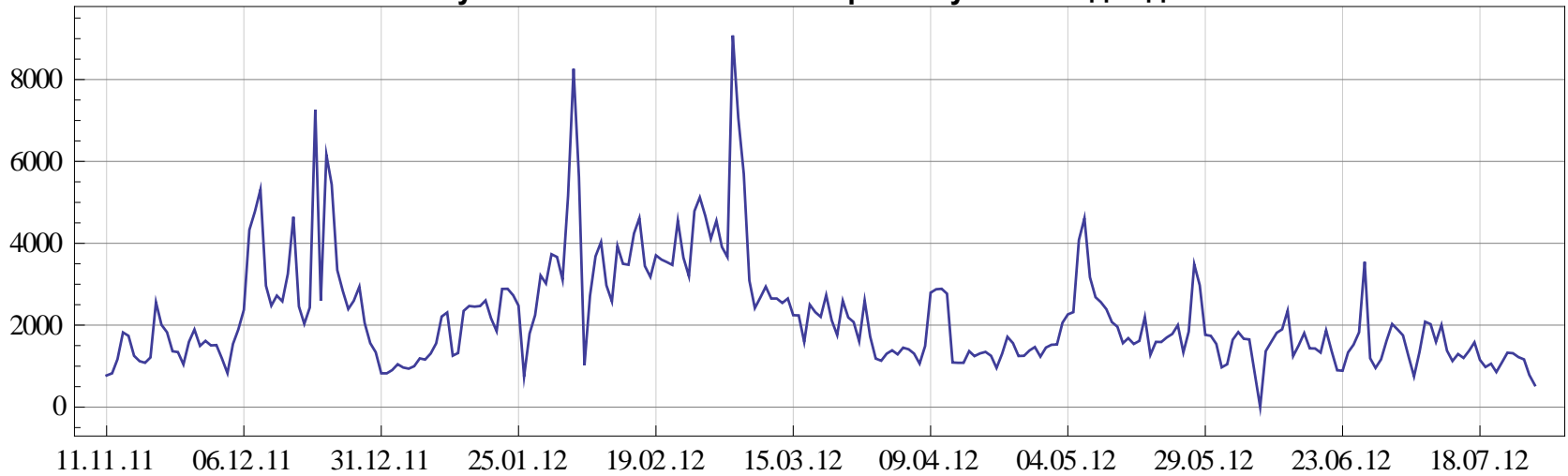
«бета» / сутки



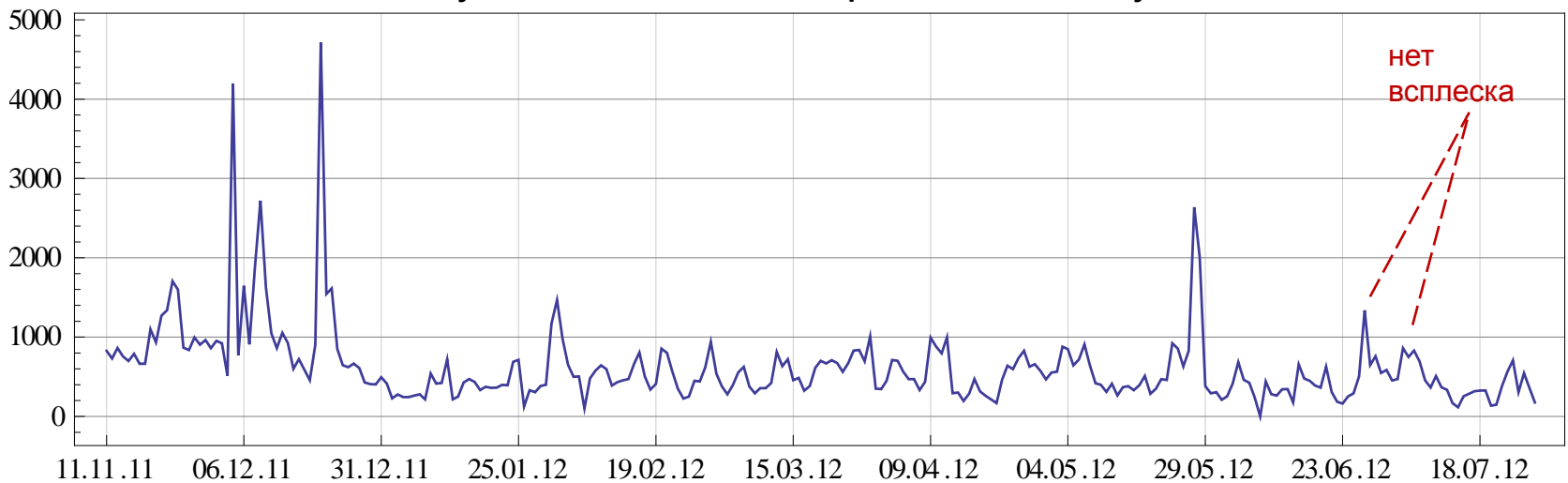
Период с 12.11.2011 г. по 13.01.2012 г.

Динамика по авторам

Суточное количество авторов : Путин \ Медведев



Суточное количество авторов : Медведев \ Путин



Производительность авторов постов

Отношение суточных количеств сообщений и авторов : Путин \ Медведев



Отношение суточных количеств сообщений и авторов : Медведев \ Путин



В среднем 3 сообщения (поста) за 2 дня

Задачи исследования

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

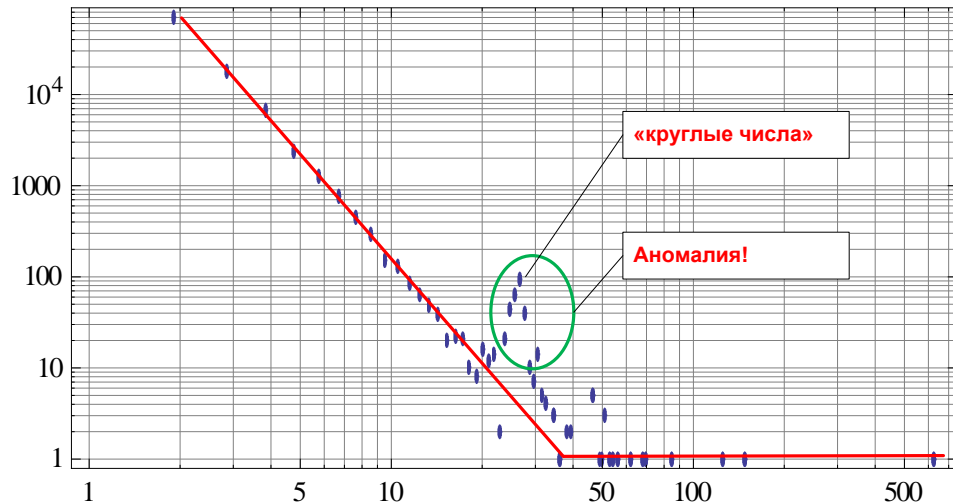
- Источники информации в блогосфере (популярность, классификация)
- Реакция блогеров на события внешнего мира

3. СТРУКТУРА БЛОГОСФЕРЫ

- Группы влиятельных блогеров
- Устойчивость к удалению вершин и связей

Активность политических авторов

Гистограмма активности уникальных авторов : Путин \ Медведев



Всплески у «круглых» значений:

Путин \ Медведев
100 блог., по 25 сообщ. каждый;

Медведев \ Путин
40 блог., 30 сообщ. каждый.

Гистограмма активности уникальных авторов : Медведев \ Путин



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

Задачи исследования

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

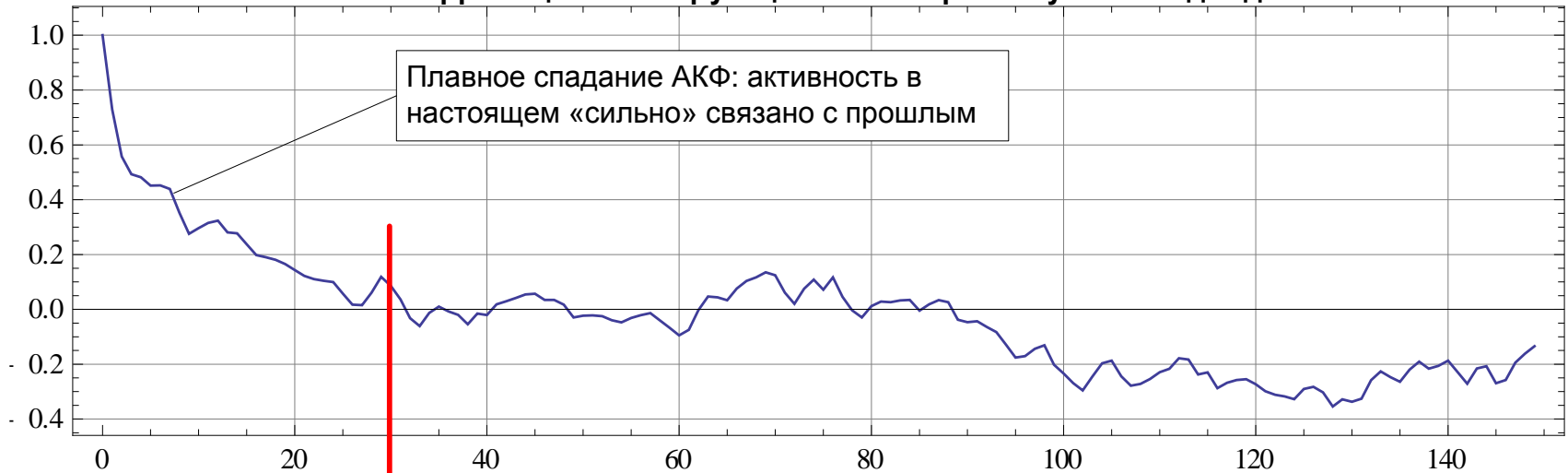
- Источники информации в блогосфере (популярность, классификация)
- Реакция блогеров на события внешнего мира

3. СТРУКТУРА БЛОГОСФЕРЫ

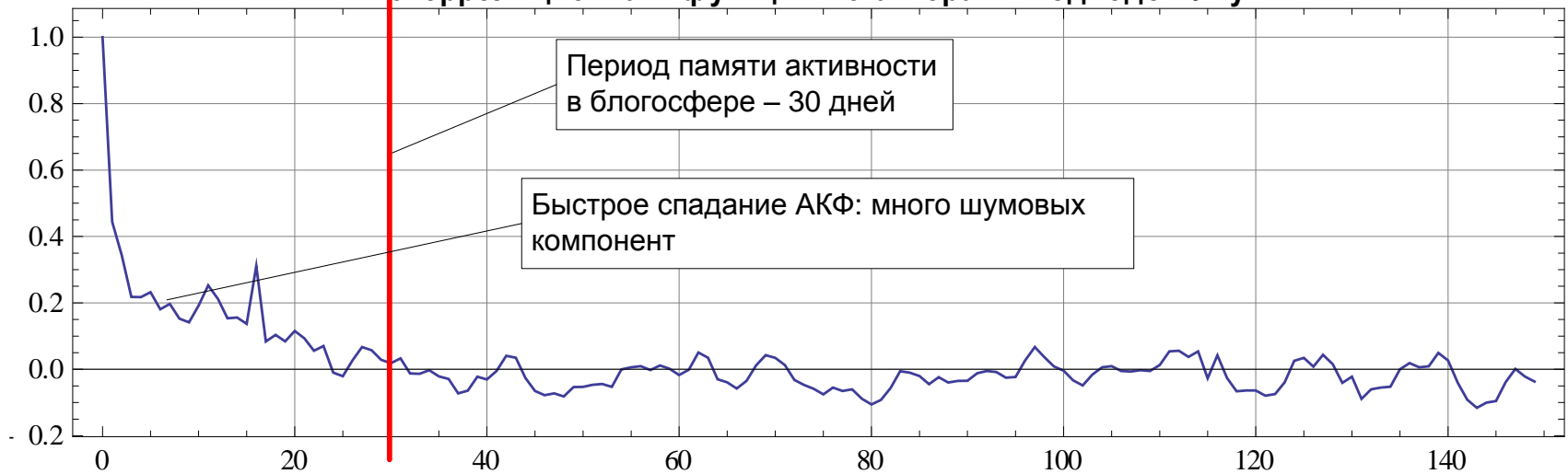
- Группы влиятельных блогеров
- Устойчивость к удалению вершин и связей

Память блогосферы

Автокорреляционная функция по авторам: Путин \ Медведев

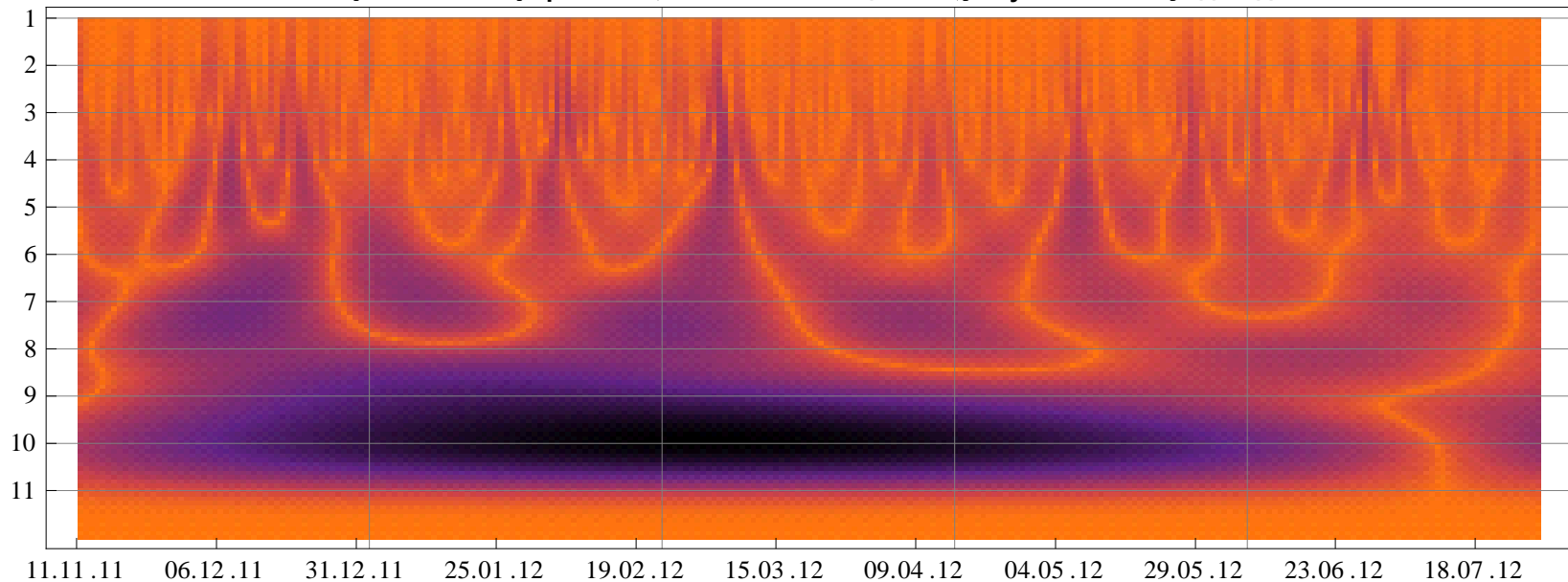


Автокорреляционная функция по авторам: Медведев \ Путин

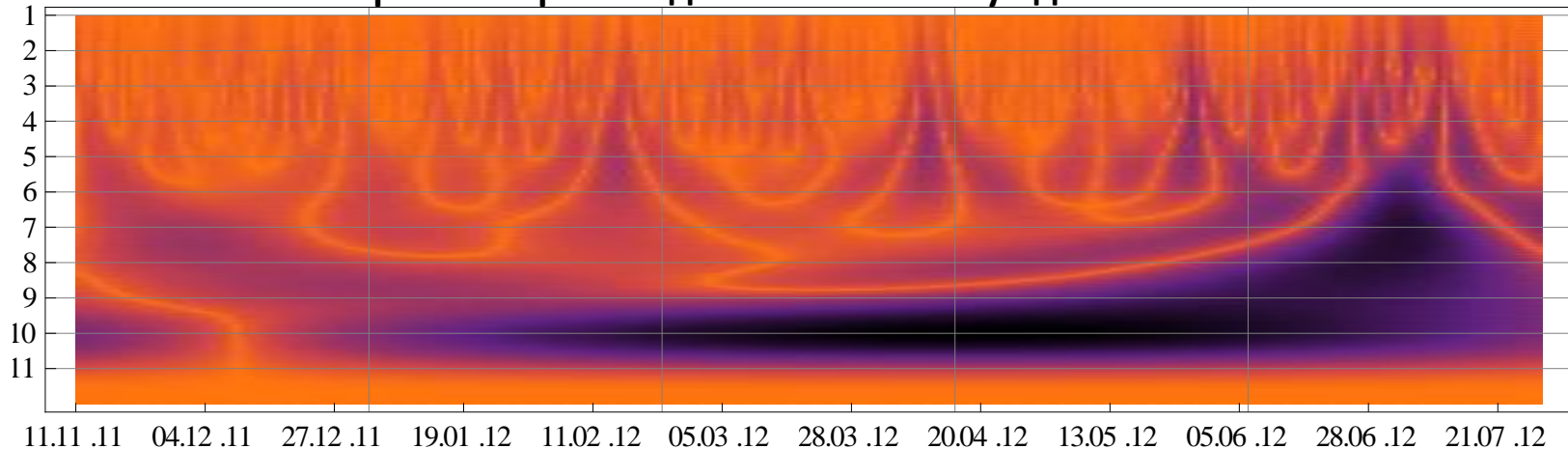


Длительность обсуждений инфоповодов

Вэйвлет скалограмма. Время – длительность обсуждений. Путин ИЛИ Медведев



Вэйвлет скалограмма. Время – длительность обсуждений. Все посты



Все инфоповоды обсуждаются не более 11 дней!

Задачи исследования

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

- Группы влиятельных блогеров
- Устойчивость к удалению вершин и связей

Взаимодействие блогосферы с внешним миром

Задачи

- Рейтинг источников, 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

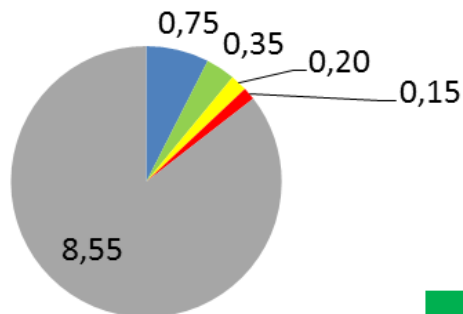
№	Название	Цитируемость блогеров/сут.	Оппозиционность	Консерватизм	Либерализм	Социализм	Радикализм	Нейтральность
1	lenta.ru	7,58	3	2	3	2	1	2
2	ria.ru	6,90	2	1	1	0	0	8
3	echo.msk.ru	6,78	4	0	5	2	3	0
4	kp.ru	6,24	1	3	1	3	1	2
5	gazeta.ru	4,90	4	1	3	1	2	3
6	kommersant.ru	4,39	3	1	3	1	1	4
7	novayagazeta.ru	3,86	4	0	5	1	4	0
8	newsru.com	3,84	3	0	3	0	2	5
9	vz.ru	3,00	0	8	1	1	0	0
			Медианная оценка для каждого идеологического показателя					
Путин/Медведев			3	1	3	1	1	3
Медведев/Путин			3	1	2	1	1	3
Путин AND Медведев			3	1	3	1	1	3
Путин OR Медведев			3	1	1,5	1	1	3

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

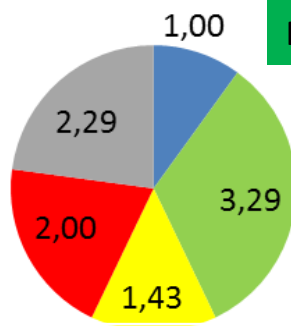
В обсуждениях, с упоминанием Путина цитируется больше источников, либеральной направленности, чем в обсуждениях Медведева

Кластеры источников

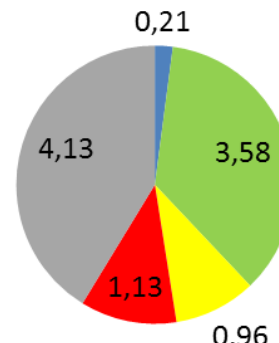
Нейтральные (20)



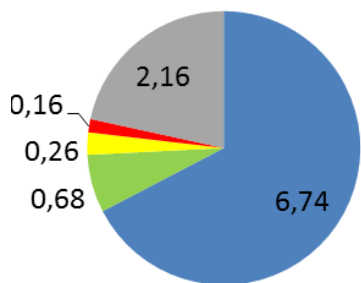
Популярные (9)



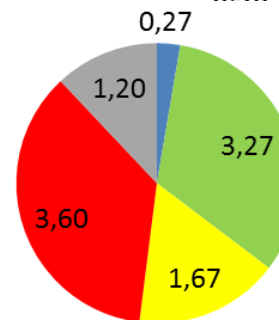
Либерально-нейтральные (24)



Консервативные (19)



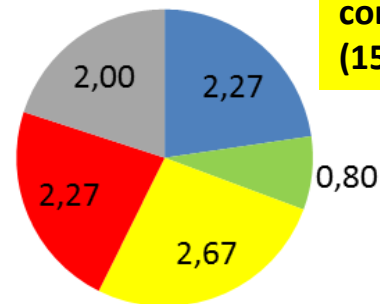
Либерально-радикальные (15)



№ Пример ресурса

1	lenta.ru
2	ria.ru
3	echo.msk.ru
9	vz.ru
10	vedomosti.ru
11	top.rbc.ru
12	vesti.ru
13	interfax.ru
14	lifenews.ru
17	svobodanews.ru

Смешанно-социалистические (15)

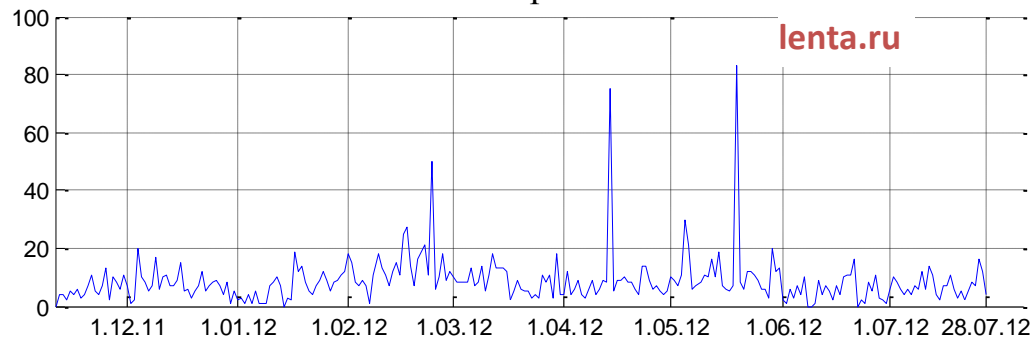


- Консерватизм
- Либерализм
- Социализм
- Радикализм
- Нейтральность

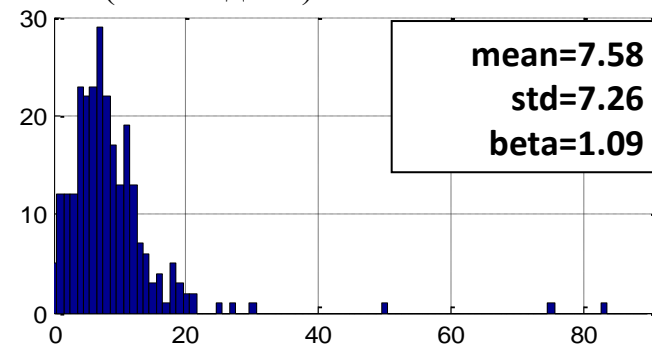
Признак «Либерализм» связан с признаком «Радикализм»

Динамика новостных источников

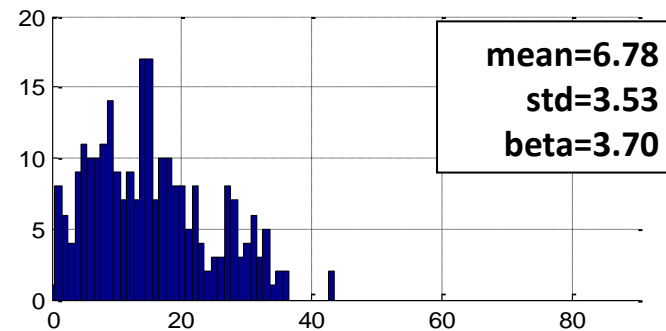
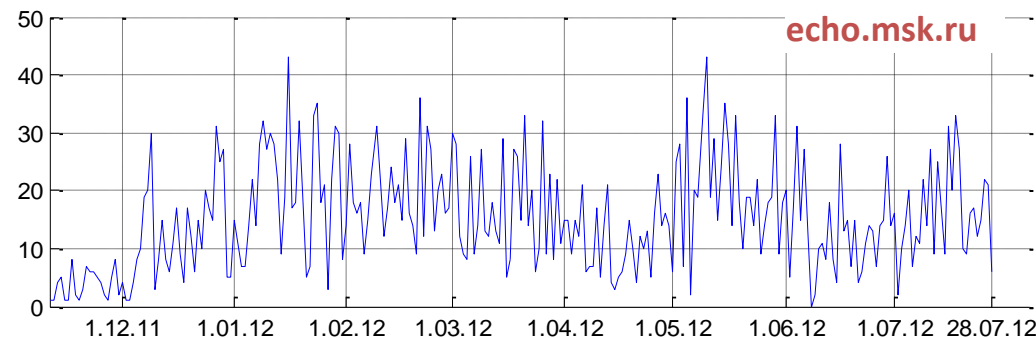
Количество блогеров



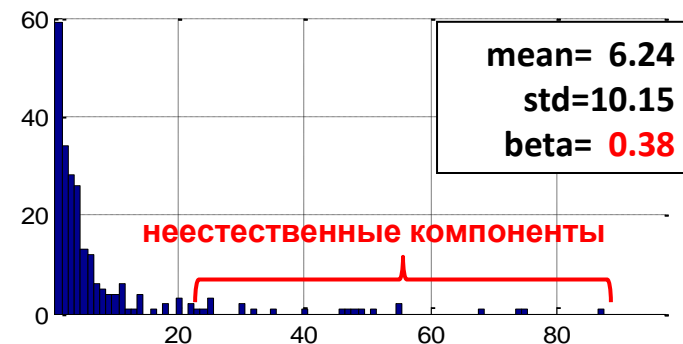
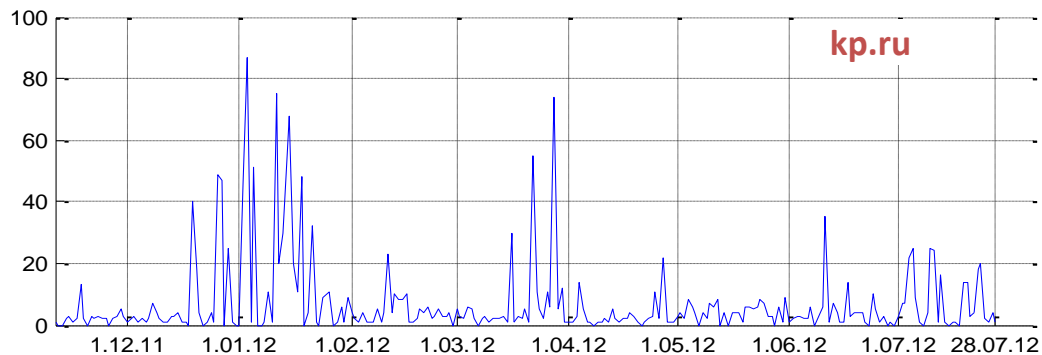
Частота (кол-во дней)



echo.msk.ru



kp.ru



Задачи исследования

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

- Группы влиятельных блогеров
- Устойчивость к удалению вершин и связей

Реакция на сообщения в СМИ

Корреляции с задержкой в 0 суток

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

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.

Задачи исследования

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

- **Группы влиятельных блогеров**
- Устойчивость к удалению вершин и связей

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 влияния разных групп блогеров. Наблюдается спад влияния радикальных групп блогеров (как оппозиционных, так и лояльных)
- Выявлены сообщества блогеров, устойчивые к удалению связей и 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 антинационалистические (-1), ярко выраженные антинационалистические (-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
aillarionov.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)

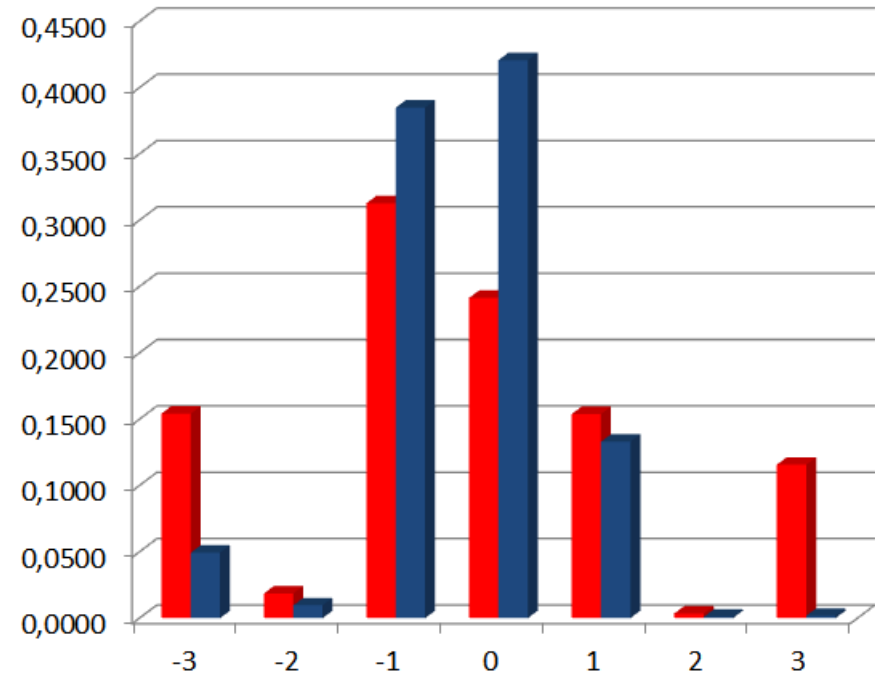
Период T1 (ноябрь-декабрь 2011)

	-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

Период T2 (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

Влиятельность



■ T1: Ноябрь-декабрь 2011 ■ T2: Июль-август 2012

	-3	-2	-1	0	1	2	3
0,1542	0,0186	0,3128	0,2414	0,1539	0,0035	0,1156	
0,0493	0,0096	0,3849	0,4206	0,1329	0,0009	0,0018	

T1

T2

Индекс радикальности: $(-3)+(3)-(-1)-(1)-(0)$

-0,4382 **-0,8874**



Индекс оппозиционности: $(-3)+(-2)+(-1)-(1)-(2)-(3)$

0,2126 **0,3082**



Задачи исследования

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)

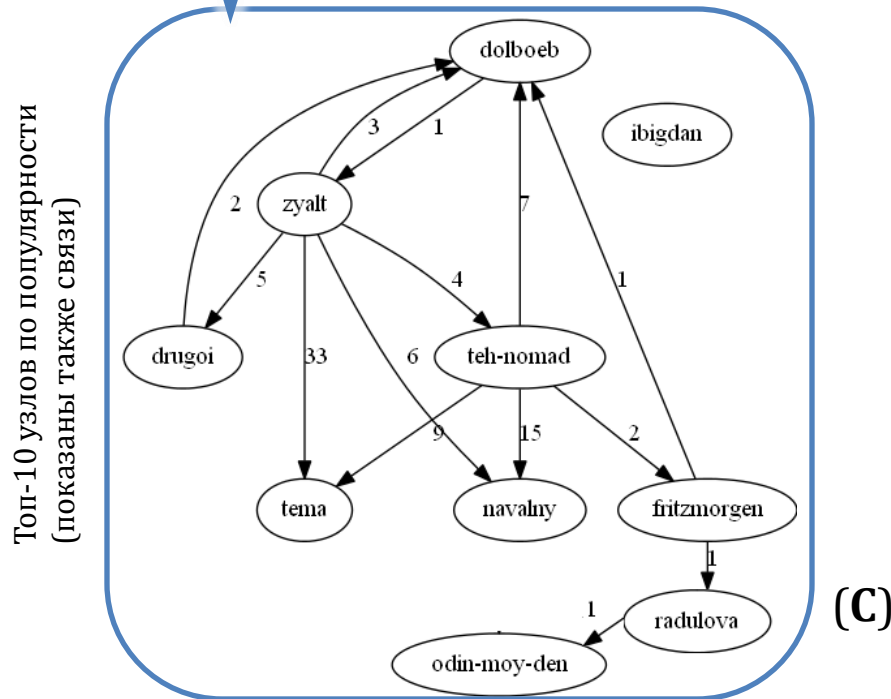
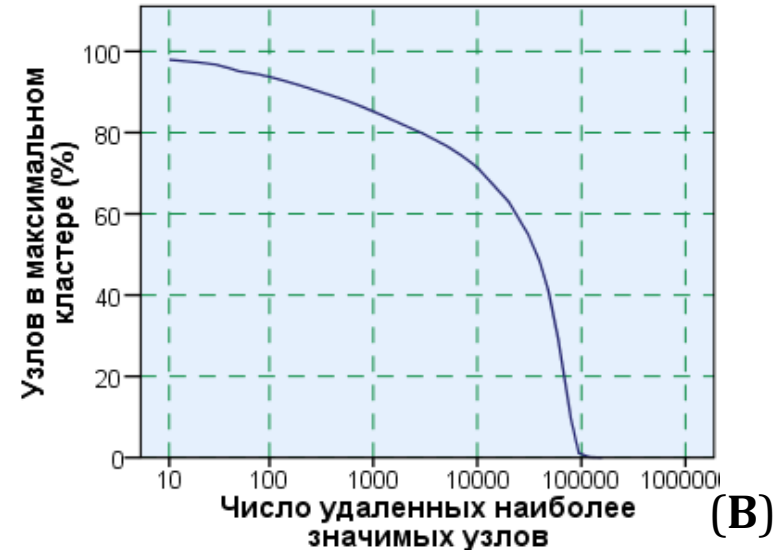
2. ВЛИЯНИЕ ВНЕШНЕГО МИРА НА БЛОГОСФЕРУ

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

3. СТРУКТУРА БЛОГОСФЕРЫ

- Группы влиятельных блогеров
- **Устойчивость к удалению вершин и связей**

Устойчивость к целенаправленному удалению популярных блогеров



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

Возможные задачи

1. Анализ

Избыток неструктурированных данных

Исследования «под микроскопом»?

Интеграция данных из разных источников?

2. Идентификация

Модели формирования сетей, распространения активности, коллективных действий и т.д.

Идентификация и использование – адекватность, вычислительная сложность?

3. Прогноз

Анализ временных рядов, прогноз на основе классификации сообщений

Прогноз на основе моделей?

4. Управление

Формальные модели

Верификация и использование?

Связь с внешним миром

«Если вы просите через Твиттер прислать машину, не ждите немедленного ответа, но будьте готовы, что через **четыре часа** вам пришлют сразу **три**» Мария Баронова, о сборе гуманитарной помощи после наводнения в Крымске, 2012

Самоорганизация в случае ЧС



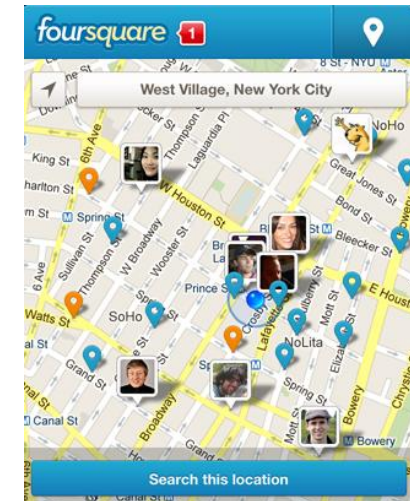
Доноры крови



Краудфандинг

Политическая деятельность
Журналистика
Благотворительность
Научные экспедиции

Сервисы check-in



Задачи

Количественные исследования
Модели коллективных действий
Предсказание эффективности кампании
Конкуренция ?