



“Livejournal Libra!” Влияние политической блогосферы на политическую мобилизацию в России в 2011-12 гг..

Щербак Андрей (ЛССИ, ВШЭ-СПб)

Кольцова Олеся (ЛИНИС, ВШЭ-СПб)

При поддержке РЭШ

Введение

- Избирательная кампания 2011-12 гг.:
снижение голосов за «партию власти»,
массовые протесты
- Роль Интернета? Особенно – роль блогов?
- Анализ постов топ-2000 блогеров
LiveJournal, за 3 периода
- Contribution: 1) блоги – оплот оппозиции, 2)
связь между политикой онлайн и политикой
оффлайн

Интернет и протесты

- Интернет и политика. В целом, существует точка зрения, что интернет создает новые возможности политической коммуникации и мобилизации
- «Арабская весна», протесты в Иране и Молдове
- Ютуб, Твиттер, социальные сети, блоги
- Часто преобладает описательность: изучение структуры, функционирования и использования интернета во время протестов
- Каузальные связи? Интернет и явка на выборах, доля голосов. Проникновение/ потребление интернета и протесты.
- Старые медиа и новые медиа; как отделить эффекты? Случай полуавторитарных стран

Выборы в России 2011-12: роль Интернета

- Распространение Интернета в России: 50 млн пользователей, 1 место в Европе по числу пользователей. Аудитория Яндекса больше аудитории 1 Канала (2012)
- Относительная автономия Интернета в России
- Рост доходов и смена ценностей (Инглхарт)
- Наталья Зубаревич: «4 России». «Первая Россия» - жители крупных городов, т.е. основная масса интернет-пользователей
- Протесты – из-за смены политической культуры (Макарин и Полищук). Доступ к новым технологиям – новые возможности мобилизации (Бодрунова и Литвиненко)
- Протесты = средний класс + доступ к Интернету (?)
- Недостаточное внимание к блогам
- Мы предсказываем преобладание протестных настроений в блогах

Данные и методы

- Выбор блогплатформы – Живой Журнал
- Отбор топ-2000 блогеров по рейтингу
- 3 периода:
 - *15 августа – 15 сентября 2011 года (т.н. «тихий» период).*
 - *27 ноября – 27 декабря 2011 года*
 - *1 февраля – 6 марта 2012 года*
- подход «больших данных»



Данные и методы

- Методика автоматизированного вычленения тем
- Каждый текст относится к каждой теме с определенной вероятностью, то есть осуществляет нечеткую кластеризацию.
- Алгоритм тематического моделирования, называемый латентным размещением Дирихле (LDA) с сэмплированием Гиббса, примененный в программном обеспечении Stanford Topic Modeling Toolbox
- За корпусом текстов стоят латентные переменные (темы). Происходит оценка вероятности отнесения поста к той или иной теме. Далее происходит кластеризация корпуса текстов на заданное количество тем (группы постов с наиболее похожими словами)
- Алгоритм позволяет относить к искомым темам релевантные тексты, даже если в них нет топовых слов по теме.
- Алгоритм позволяет выявлять темы, наличие которых исследователь не предполагал, либо лексический состав которых исследователю не ясен.

Данные и методы

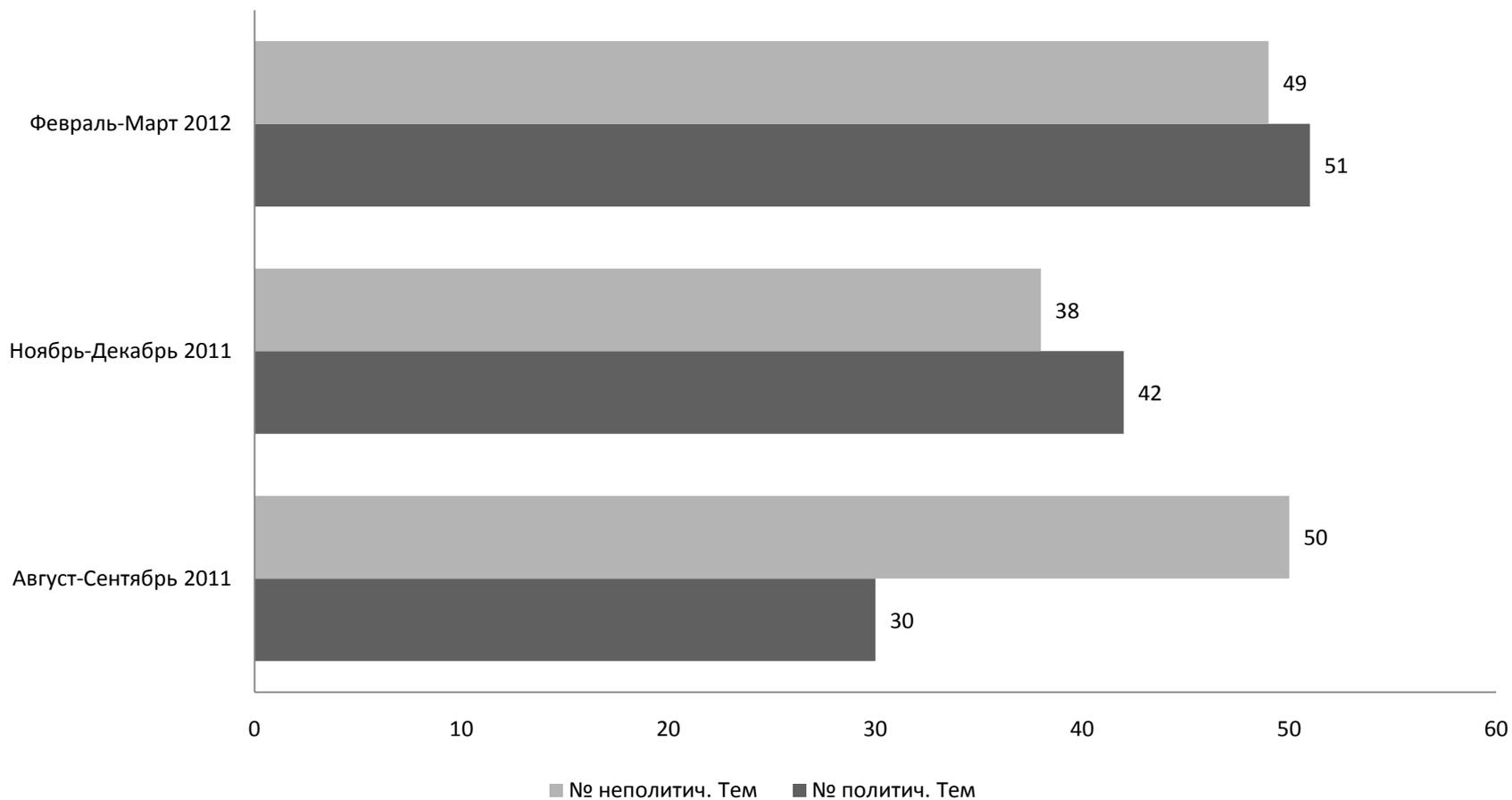
- 1 этап: выгружены три массива постов за каждый из трех периодов. Каждая выборка была кластеризована на 100 тем на основе алгоритма LDA. Два кодировщика вручную отобрали темы, связанные с политикой. Минимальным критерием отнесения темы к политической было наличие хотя бы 5 постов (из топовых 50) в теме. Политическими постами не считались посты исторического (советская история) или внешнеполитического (события в Египте или внутренняя украинская политика) характера, если автор не увязывал их содержание с текущей политической ситуацией в России.
- 55 политических тем: 11 за период август-сентябрь 2011 года, 24 – за период ноября-декабря 2011 года и 20 – за февраль-март 2012 года.
- 2 этап: все тексты, отнесенные к выбранным темам с вероятностью более 0,1, были объединены в один массив.
- 3 этап: новый массив был переформатирован. Во-первых, новая база была разбита по неделям (13 недель). Во-вторых, каждая недельная выборка была кластеризована на 20 тем. Эти темы были обработаны вручную также двумя кодировщиками, и из них были отобраны 123 темы (из каждой были взяты по 30 топ-постов), преимущественно связанных с внутренней российской политикой. **В итоге сформировалась база данных из 3690 постов**, с крайне высокой вероятностью отнесения их к политическим по содержанию.

Примеры кластеризации

масло	1494	участок	1712	русский	6480
добавлять	1126	избирательный	1661	советский	2165
вода	933	наблюдатель	1573	ссср	1513
мясо	810	голос	1491	язык	1016
ложка	788	голосование	1160	народ	1010
минута	777	результат	1071	сталин	912
сахар	740	кандидат	995	еврей	806
соль	727	бюллетень	981	латвия	666
2	725	избиратель	960	национальный	605
рецепт	708	комиссия	830	германий	584
вкусный	698	фальсификация	807	немец	537
г	674	4	735	немецкий	525
блюдо	664	март	714	союз	505
вкус	650	голосовать	702	история	495
мука	642	нарушение	574	еврейский	455
яйцо	639	опрос	556	государство	405
перец	601	проголосовать	555	война	403
сырой	599	уик	531	нация	389
сок	589	партия	450	население	376
молоко	588	протокол	445	гитлер	360



Данные и методы



Данные и методы

- 4 этап: посты кодировались вручную пятью кодировщиками (коэффициент Каппа не менее 0.73).
- Индикаторы для кодировки: «*политика*» (отнесение содержания поста к политике), «*власть*» (отношение к власти), «*оппозиция*» (отношение к оппозиции)
- Если пост признавался неполитическим, то его кодировка прекращалась; кодировщики выделили около 30% таких постов. Диапазон значений для каждой из содержательных переменных (т.е. всех кроме «Политика») ранжировался от «-1» (отрицательное отношение к категории) до «2» (крайне положительное отношение)

Переменные

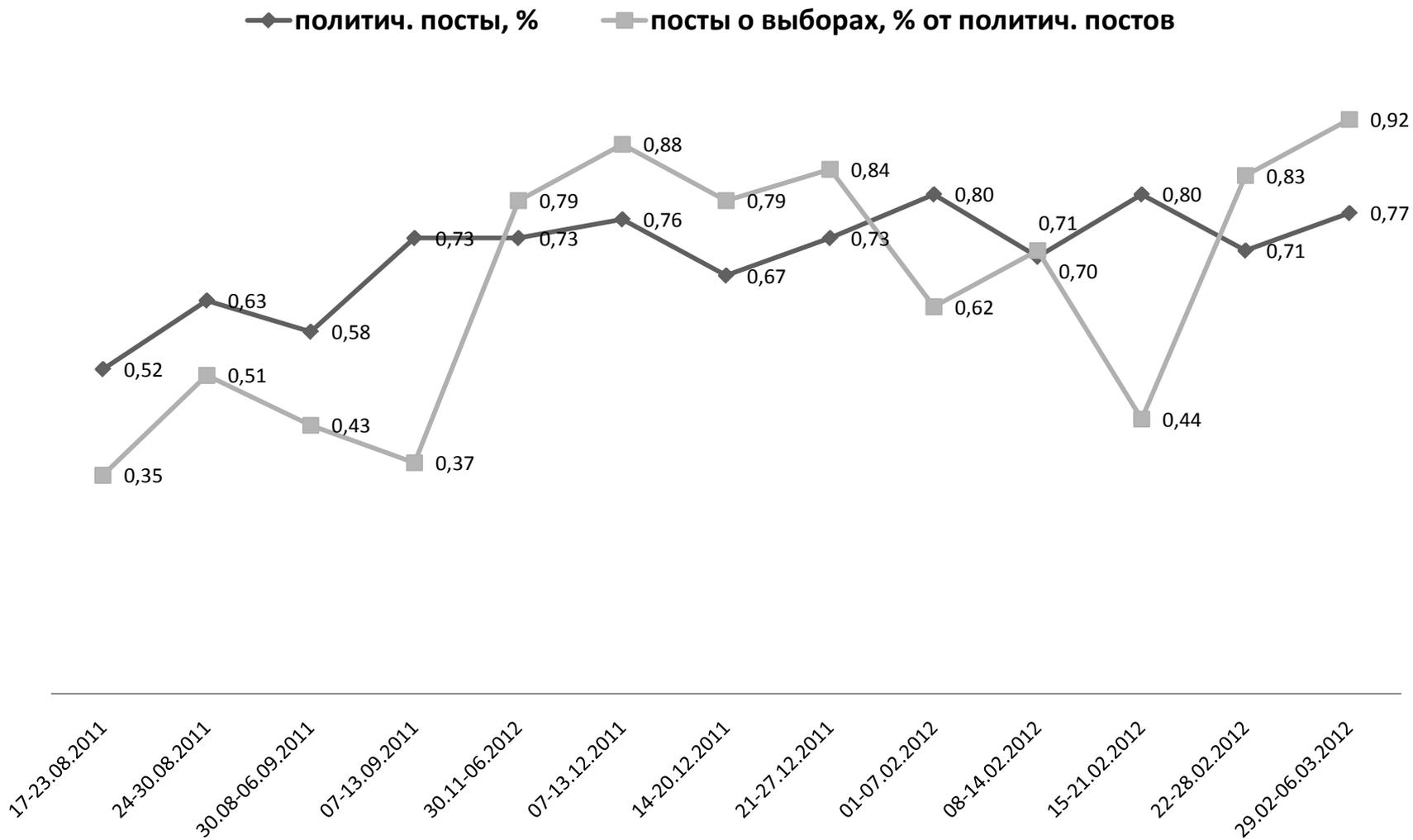
- **Доля постов о политике** (отношение числа постов о политике к общему числу постов в изучаемый период); изменяется от 0 до 100%;
- **Доля постов о выборах** (и протестах) среди всех политических постов; изменяется от 0 до 100%;
- **Доля позитивных оппозиционных постов** (отношение числа постов, имеющих значение 1 или 2 по переменной «оппозиция», к общему числу политических постов в изучаемый период); изменяется от 0 до 100%;
- **Доля позитивных «провластных» постов** (отношения числа постов, имеющих значение 1 и 2 по переменной «власть», к общему числу политических постов в изучаемый период); изменяется от 0 до 100%;
- **Отношение к оппозиции** (среднее по переменной «оппозиция» в изучаемый период); изменяется от -1 до 2;
- **Отношение к власти** (среднее по переменной «власть» в изучаемый период); изменяется от -1 до 2

Гипотезы

- Сопоставление еженедельных показателей политической активности блогеров с еженедельными предвыборными рейтингами (ФОМ) партий и кандидатов в президенты
- Метод корреляционного анализа.
- *H1. Доля позитивных оппозиционных постов значимо выше, чем доля позитивных провластных постов.*
- *H2. Отношение к оппозиции значимо позитивнее, чем отношение к власти.*
- *H3. Чем выше политическая активность блогеров (доля постов о политике, доля постов о выборах, доля позитивных оппозиционных постов, отношение к оппозиции, разница между отношением к оппозиции и власти), тем ниже рейтинги власти (электоральный рейтинг Единой России, В.Путина, рейтинг доверия Д.Медведева).*
- *H4. Чем выше политическая активность блогеров, тем выше рейтинги оппозиции (электоральный рейтинг КПРФ, СР, ЛДПР, Г.Зюганова и В.Жириновского).*

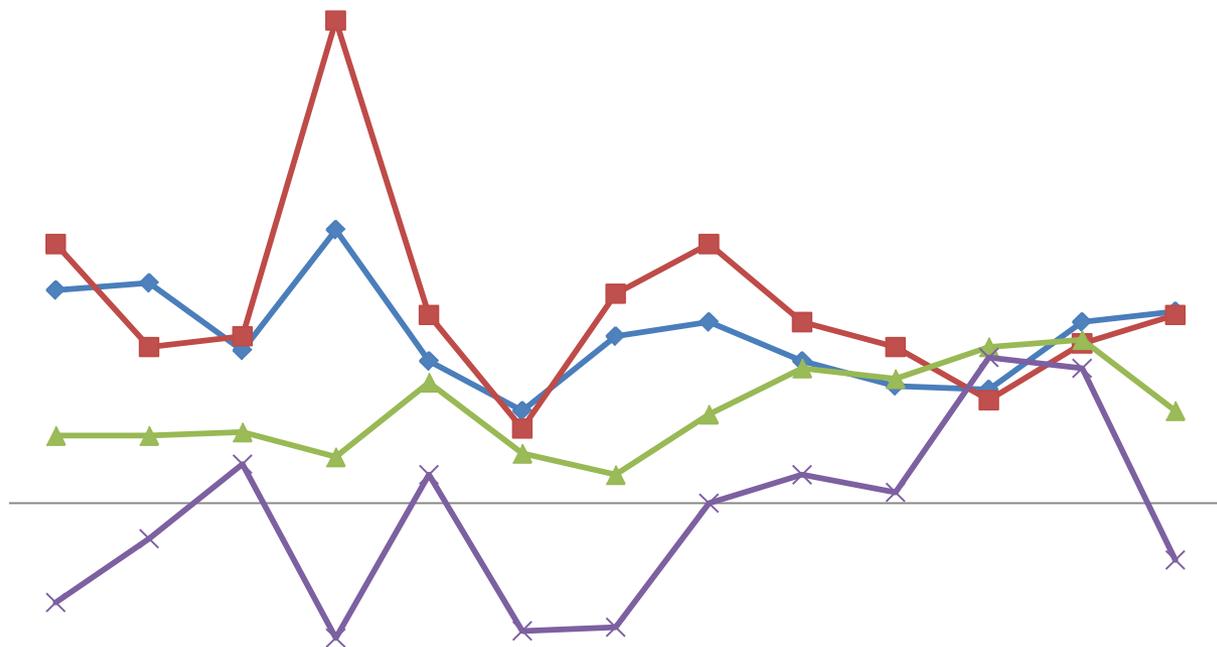
-

Политическая активность блогеров



Политическая активность блогеров

- ◆ про-опозиционные посты, % от политич. постов
- Отношение к оппозиции, среднее
- ▲ провластные посты, % от политич. Постов
- × Отношение к власти, среднее



	17-23. 08.11	24- 30.08. 11	30.08- 06.09. 11	07- 13.09. 11	30.11- 06.12. 11	07-13. 12.11	14- 20.12. 11	21- 27.12. 11	01- 07.02. 12	08- 14.02. 12	15- 21.02. 12	22-28. 02.12	29.02- 06.03. 12
про-опозиционные посты, % от политич. постов	0,60	0,62	0,43	0,77	0,40	0,26	0,47	0,51	0,40	0,33	0,32	0,51	0,54
Отношение к оппозиции, среднее	0,73	0,44	0,47	1,36	0,53	0,21	0,59	0,73	0,51	0,44	0,29	0,45	0,53
провластные посты, % от политич. Постов	0,19	0,19	0,20	0,13	0,34	0,14	0,08	0,25	0,38	0,35	0,44	0,46	0,26
Отношение к власти, среднее	-0,28	-0,10	0,11	-0,38	0,08	-0,36	-0,35	0,00	0,08	0,03	0,41	0,38	-0,16

Результаты

- Блогосфера принадлежит преимущественно оппозиционно настроенным блогерам.
- Среднее значение *«Доля позитивных оппозиционных постов»* за все недели – 0,47
- *«Доля проправительственных постов»* - только 0,26.
- *«Отношение к оппозиции»* - 0,56, а аналогичный показатель для *«Отношения к власти»* -0,05.
- Разница между этими показателями, т.е. между отношением к власти и к оппозиции, варьировалась по некоторым неделям от 1,74 до -0,12, при среднем значении 0,60 (по шкале от -1 до 2).
- Все это в целом согласуется с тезисом о доминировании в блогах *«недовольного среднего класса»*, с нереализованным запросом на ценности самовыражения и политические свободы.
- Корреляции с предвыборными рейтингами

Политическая активность и рейтинги

	ПОСТЫ О ВЫБОРАХ (%)	ПОСТЫ О ПОЛИТИКЕ (%)
Рейтинг КПРФ	0.872** (p=0.000)	0.571* (p=0.041)
Рейтинг Единой России	-0.172 (p=0.574)	0.091 (p=0.768)
Рейтинг Справедливой России	0.551 (p=0.051)	0.321 (p=0.285)
Рейтинг ЛДПР	0.220 (p=0.471)	0.027 (p=0.931)
Рейтинг В.Путина	-0.254 (p=0.402)	-0.054 (0.860)
Рейтинг Г.Зюганова	0.680* (p=0.011)	0.462 (p=0.112)
Рейтинг Д.Медведева	-0.278 (p=0.358)	-0.108 (p=0.724)
Рейтинг В.Жириновского	0.106 (p=0.730)	-0.155 (p=0.613)

Политическая активность и рейтинги

- КПРФ – главный выгодополучатель от политической активности блогеров?
- Эффект «варианта Навального»? Голосовать за любую другую партию кроме ЕР. От этой стратегии могли получить выгоду КПРФ, СР и ЛДПР.

Политическая активность и рейтинги

	Доля оппозиционных постов	Отношение к оппозиции, среднее	Доля провластных постов	Отношение к власти, среднее
Рейтинг КПРФ	0.310 (p=0.302)	0.026 (p=0.933)	0.089 (p=0.774)	-0.140 (p=0.649)
Рейтинг Единой России	-0.042 (p=0.892)	-0.333 (0.266)	0.733** (p=0.004)	0.636* (p=0.019)
Рейтинг Справедливой России	-0.696** (p=0.008)	-0.123 (p=0.688)	0.106 (p=0.730)	0.095 (p=0.758)
Рейтинг ЛДПР	-0.085 (p=0.783)	0.251 (p=0.408)	-0.462 (p=0.112)	-0.389 (p=0.188)
Рейтинг В.Путина	0.401 (p=0.174)	0.043 (p=0.889)	0.196 (p=0.521)	0.152 (p=0.620)
Рейтинг Г.Зюганова	-0.378 (p=0.202)	0.132 (p=0.667)	-0.053 (p=0.864)	-0.171 (p=0.576)
Рейтинг Д.Медведева	0.449 (p=0.124)	0.004 (p=0.989)	0.158 (p=0.607)	0.094 (p=0.760)
Рейтинг В.Жириновского	-0.044 (p=0.887)	0.215 (p=0.480)	-0.582* (p=0.037)	-0.397 (p=0.179)

Результаты

- Апелляция к общественному мнению в онлайне способна приносить определенные выгоды политикам в оффлайне
- «Рецепт успеха» для оппозиции оказался крайне прост: главное даже не «о чем» писать, а «сколько» писать. Любая информация о ходе российских выборов играла на руку оппозиции. Больше постов о выборах - выше предвыборные рейтинги оппозиционных партий.

Заключение

- Связь политики онлайн и политики оффлайн
- Роль Интернета в гибридных режимах. Не стоит игнорировать блоги!
- Методология LDA: подход «больших данных». Именно с ее помощью мы смогли получить ответы как о происходящем внутри блогосферы, так и за ее пределами