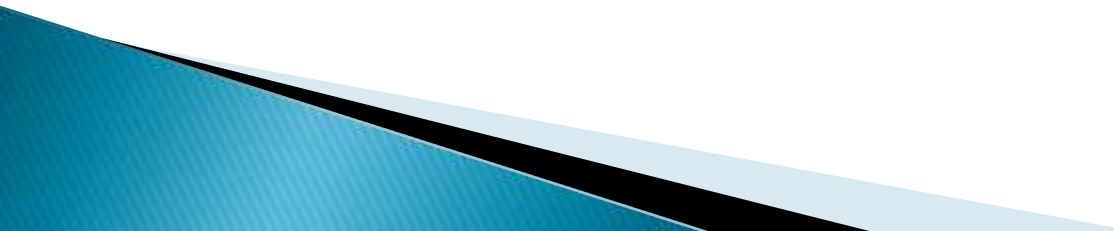


Методология и инструменты Big Data в социально-политических исследованиях.

**Мастер-класс доктора политических наук
Бродовской Елены Викторовны
30 марта 2018 года, СевГУ.**

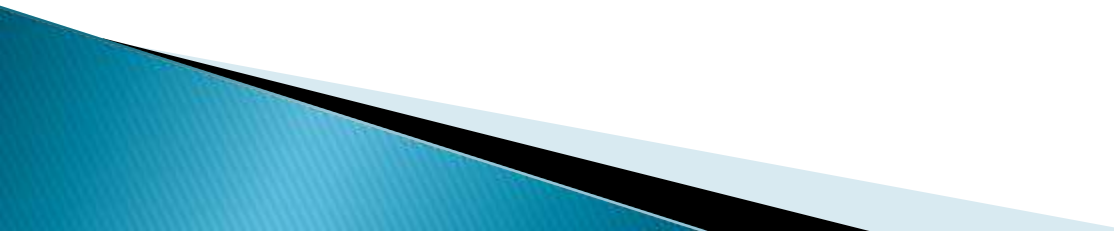


Predictor Mining

Predictor Mining - новое направление исследования политического поведения и политических процессов, отраженных в сети Интернет.

Predictor Mining возник на пересечении интеллектуального анализа данных и социального компьютеринга.

Посредством использования методов автоматического извлечения закономерностей из больших объемов данных для анализа и прогнозирования политических протестов и их динамики, Predictor Mining способствует развитию нового теоретического и прикладного инструментария исследования политических процессов по цифровым следам.



Теоретический фундамент исследования

- ▶ Концепция глобальной информационной сети (С. Колеман, Д. Карран, Д. Тапскотт).
- ▶ Теории коллективного поведения (Г. Лебон, Г. Тард, Г. Ортега-и-Гассет , С. Московичи, Г. Блумер).
- ▶ Концепции социального и политического конфликта (Р. Дарендорф, М. Дойч, Л. Козер).
- ▶ Теории относительной депривации (Н. Смелзер Т. Гарр, Р. Тед).
- ▶ Концепция мобилизации ресурсов (Дж. Маккарти, А. Обершол).
- ▶ Ценностно-ориентированный подход "новых общественных движений" (К.Оффе).
- ▶ Подход сетевого взаимодействия (М. Кастельс, А. Чевик, Х. Квак, С. Ли).
- ▶ Концепция "сетевого клуба" (Дж. Коэн, Э. Шмидт).
- ▶ Теория "умной толпы" (Х. Дженкинс, Г. Рейнгольд).
- ▶ Концепции политических мимов (Р. Докинз, С. Шомова, С. Грин).

Методология исследования

- ▶ *Новый институционализм* позволяет рассматривать в качестве факторов, влияющих на политическую систему, политические институты, акторов, их ресурсы и стратегии (Х. Линц, А. Степан, Г.О. Доннелл, Ф. Шмиттер и др.), а также анализировать рациональность выбора стратегии политического поведения (Э. Остром, К. Шепсл, Д. Норт).

Методология исследования

- ▶ *Сетевой подход* делает возможным исследовать взаимосвязь между мобилизацией политического действия и сетевой активностью Интернет-пользователей (Г. Градосельская, С. Грин, Л.Сморгунов, Ф. Ховард, А. Шерстобитов, А. Роменков, Д. Мартьянов, И. Мирошниченко и др.), сетевые интеракции как коллективное действие (Л. Карлссон, И. Кравченко), инструменты вовлечения в сетевой политический протест (З. Бауман, Э.Кац, П. Лазарсфельд, Р. Эпштейн, В. Костюшев, В. Сафронов и др.).

Методология исследования

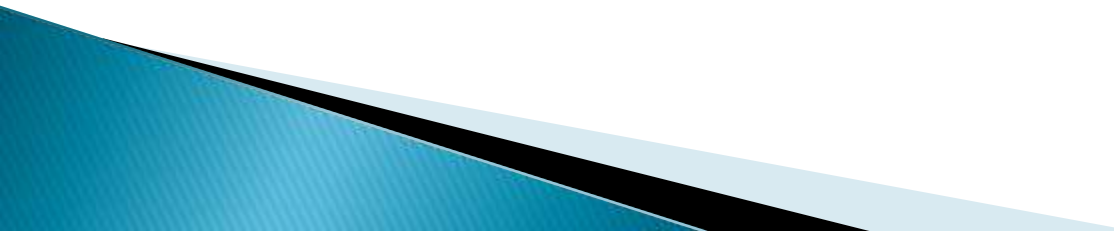
- ▶ Когнитивный подход акцентирован на анализе процессов формирования субъективного политического мира, субъективного восприятия и интерпретации сообщений, транслируемых в пространстве социальных медиа, выбора стратегии политического поведения (Р.Дж. Далтон, П. Данливи, Г. Пушкарева, Е. Шестопап и др.).

Значимыми для решения задач интеллектуального поиска цифровых маркеров политических процессов являются такие аспекты когнитивного подхода, как технологии информационного влияния, манипулятивное воздействие, убеждающая коммуникация и др. (Э. Аронсон, Р. Эйкерт, Э.Р. Пратканис, С. Кара-Мурза, Г. Почепцов, М. Грачев, М. Вершинин, А. Балашов, И. Быков, М. Щенников и др.).

Эмпирическая модель исследования

- ▶ киберметрический анализ: автоматизированный анализ информационных потоков социальных медиа с помощью онлайн-сервиса для мониторинга социальных медиа IQBuzz (выгрузка сообщений по всем исследуемым кейсам составила составила более 10 млн. документов, глубина выгрузки – 01.01.2006 г. по 08.04.2017 г., 30 социальных медиа и микроблогов);
- ▶ построение социальных графов, позволяющих определить плотность социальных связей между участниками онлайн сообществ, выделить лидеров общественного мнения;
- ▶ дискурс-анализ сообщений пользователей социальных медиа для определения ключевых тем, релевантных кейсам исследования.

Исследовательские кейсы:

- ▶ анализ активно-нигилистического и пассивно-нигилистического информационных потоков в дискурсе пользователей социальных медиа Украины (2014 г., автоматизированная классификация информационных потоков).
 - ▶ анализ облака тегов политического протеста в сообщениях пользователей социальных медиа России (2011—2017 гг., дискурс анализ);
 - ▶ технологии формирования установок и мобилизации протестной политической активности у пользователей социальных медиа России (2011—2017 гг., киберметрия);
 - ▶ выявление структуры ролей и плотности социальных связей в российских и украинских онлайн-сообществах массового политического протеста в социальной сети «ВКонтакте» (2017 гг., построение социальных графов);
 - ▶ социальный и политический протест коллективных акторов в современной России (Манежная площадь, Кондопога, Пугачев, Ростов-на-Дону, Бирюлево, Минеральные Воды (2006—2014 гг., триангуляция методов ивент-анализа, дискурс анализа, киберметрии).
- 

Эвристические возможности Predictor Mining

Содержательные:

- ▶ автоматизированное сегментирование социально-медийных информационных потоков;
- ▶ выявление облаков тегов сообщений социальных медиа.

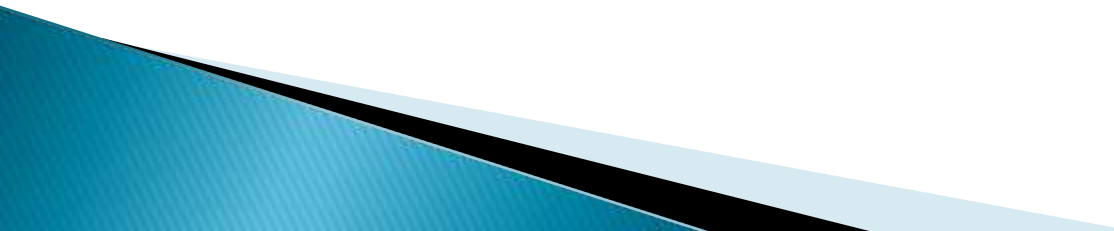
Эвристические возможности Predictor Mining

Социально-демографические:

- ▶ половозрастной анализ авторов социально-медийных документов;
- ▶ геолокационный анализ сообщений социальных медиа.

Эвристические возможности Predictor Mining

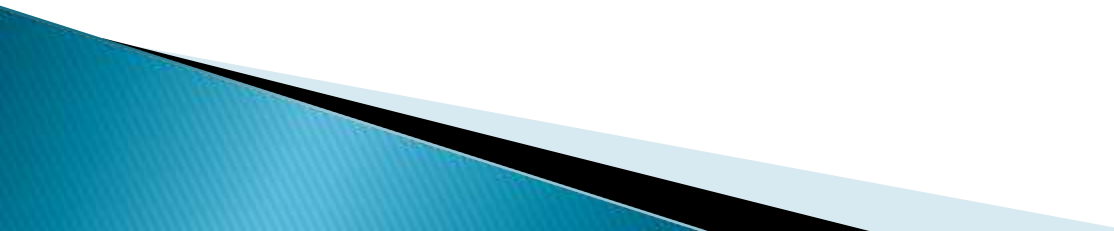
Технологические:

- ▶ исследование динамики информационных потоков;
 - ▶ изучение публикационной активности авторов информационных потоков и их аудитории;
 - ▶ построение социальных графов, отражающих структуру взаимодействия внутри и между онлайн-сообществами.
- 

Инструменты киберметрии

Сервисы для мониторинга социальных медиа:

Seesmic, Socialmention, Hootsuite, YouScan, BuzzLook, IQBuzz, Socialbakers, SocialSeek, Socialpointer, PerrIndex, PostRank, Topsy, HowSociable, Twitalyzer, WildFire, Kurrently, Trackur, Babkee, Buzzware, Wobot и ряд др. позволяют осуществить автоматическую выгрузку документов, релевантных предмету исследования, из социальных медиа.

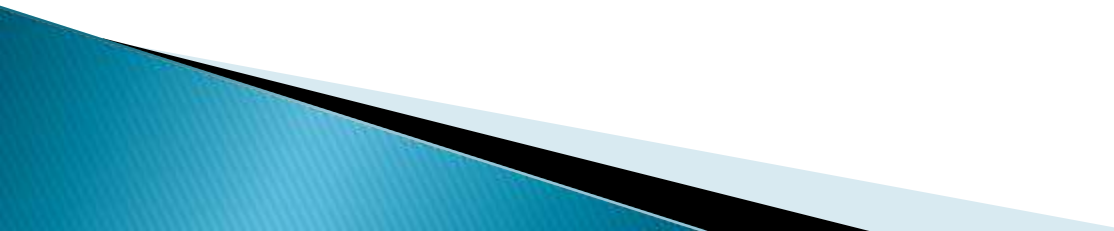


Достоинства и недостатки IQBuzz

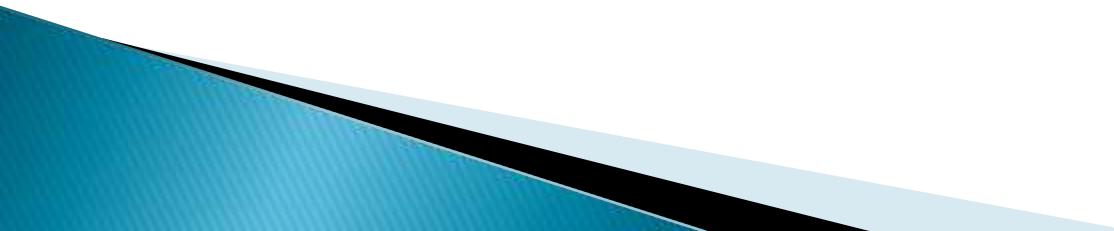
Преимущества:

- ▶ относительно большой объем выгружаемых сообщений за одну итерацию (до 100 000);
- ▶ высокий уровень идентификации социально-демографических характеристик авторов сообщений;
- ▶ высокие показатели объема и глубины аккумулируемых документов;
- ▶ использование большого объема социальных медиа и др. Интернет-ресурсов

Ограничения:

- ▶ Низкая доля идентификации модальности сообщений;
 - ▶ Отсутствие возможностей по распознаванию иронии в тексте.
- 

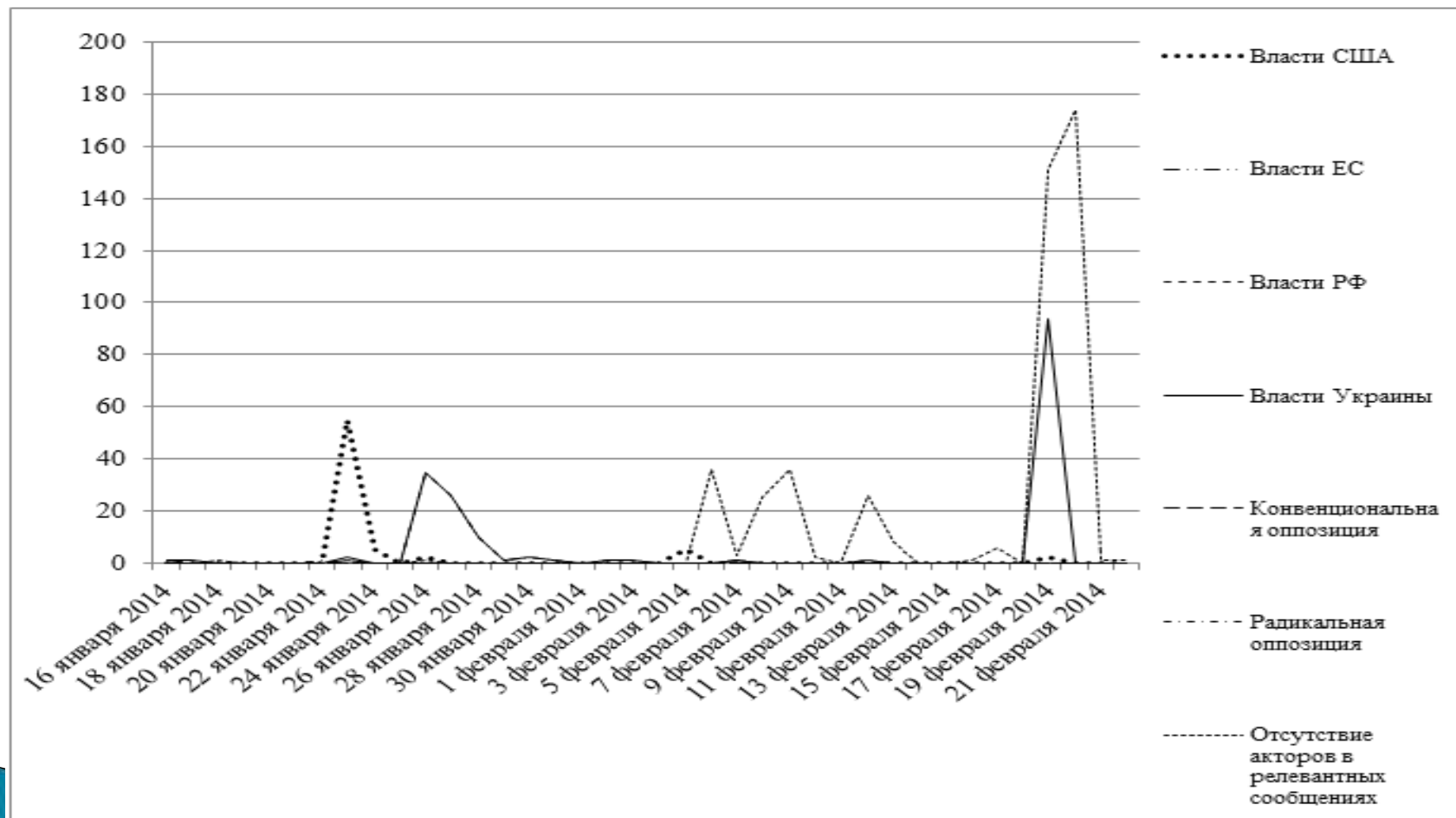
Автоматизированное сегментирование информационных потоков: поисковые запросы

- ▶ Поисковые запросы — это те слова (словосочетания), присутствие которых в документе указывает на то, что сообщение релевантно теме исследования.
 - ▶ Возможности онлайн сервисов позволяют аккумулировать сообщения, в которых содержатся слова, имеющие такой же корень, как и в поисковом запросе, поэтому в процессе ввода маркеров достаточно указать корни слов — поисковых критериев.
- 

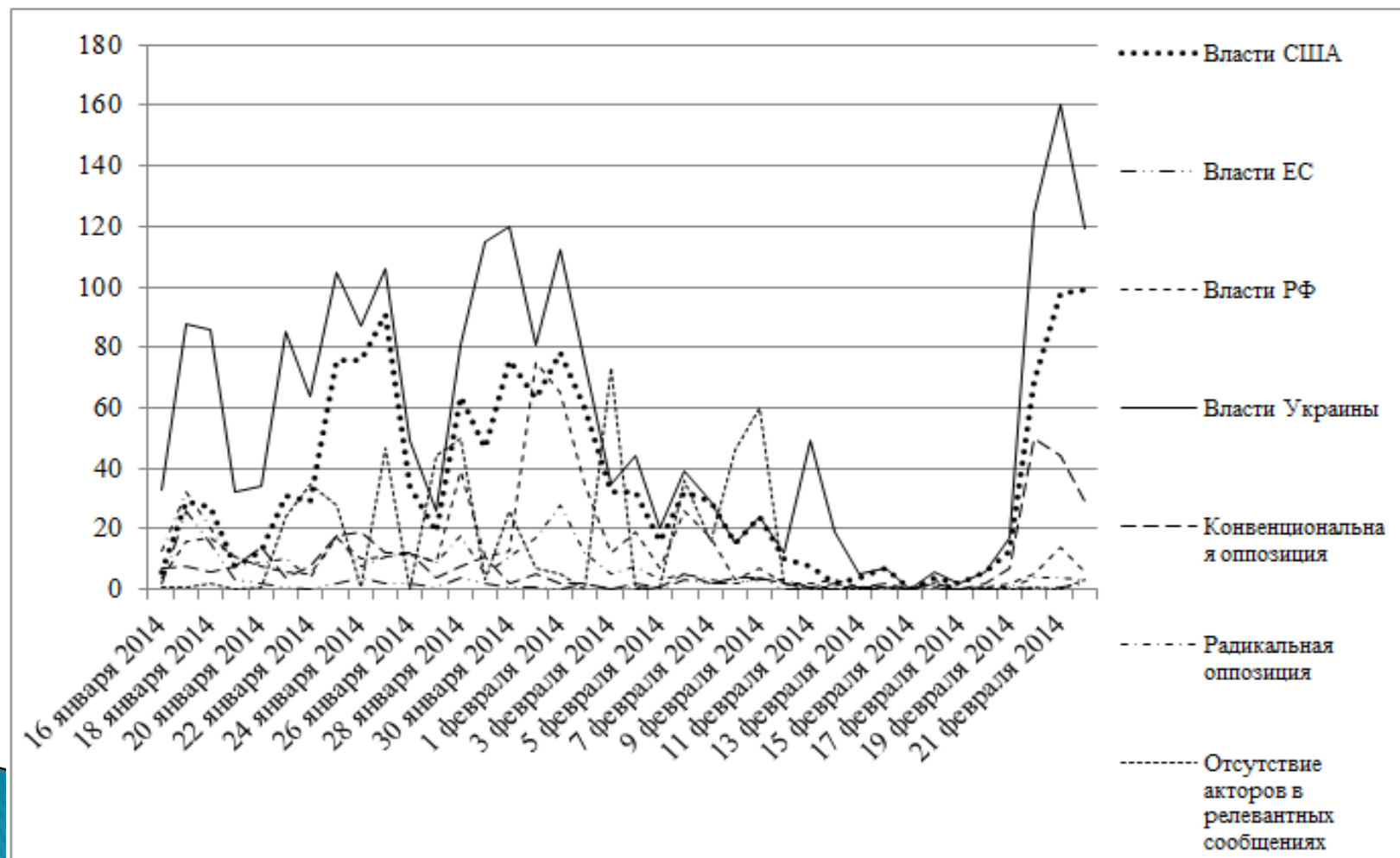
Автоматизированное сегментирование информационных потоков: словарь маркеров

- ▶ Пример словаря маркеров для темы формирования протестных установок в украинских социальных медиа: «Майдан|Евромайдан|євромайдан|Ukraine». Довольно эффективными поисковыми запросами были слова, обозначающие неприязненное отношение к ныне бывшему Президенту Украины («Предатель|Вор|Януч|Бандюкович» и т. д.).

Хронологическое распределение сообщений активно-нигилистического характера с учетом акторов (16 января - 22 февраля 2014 года)



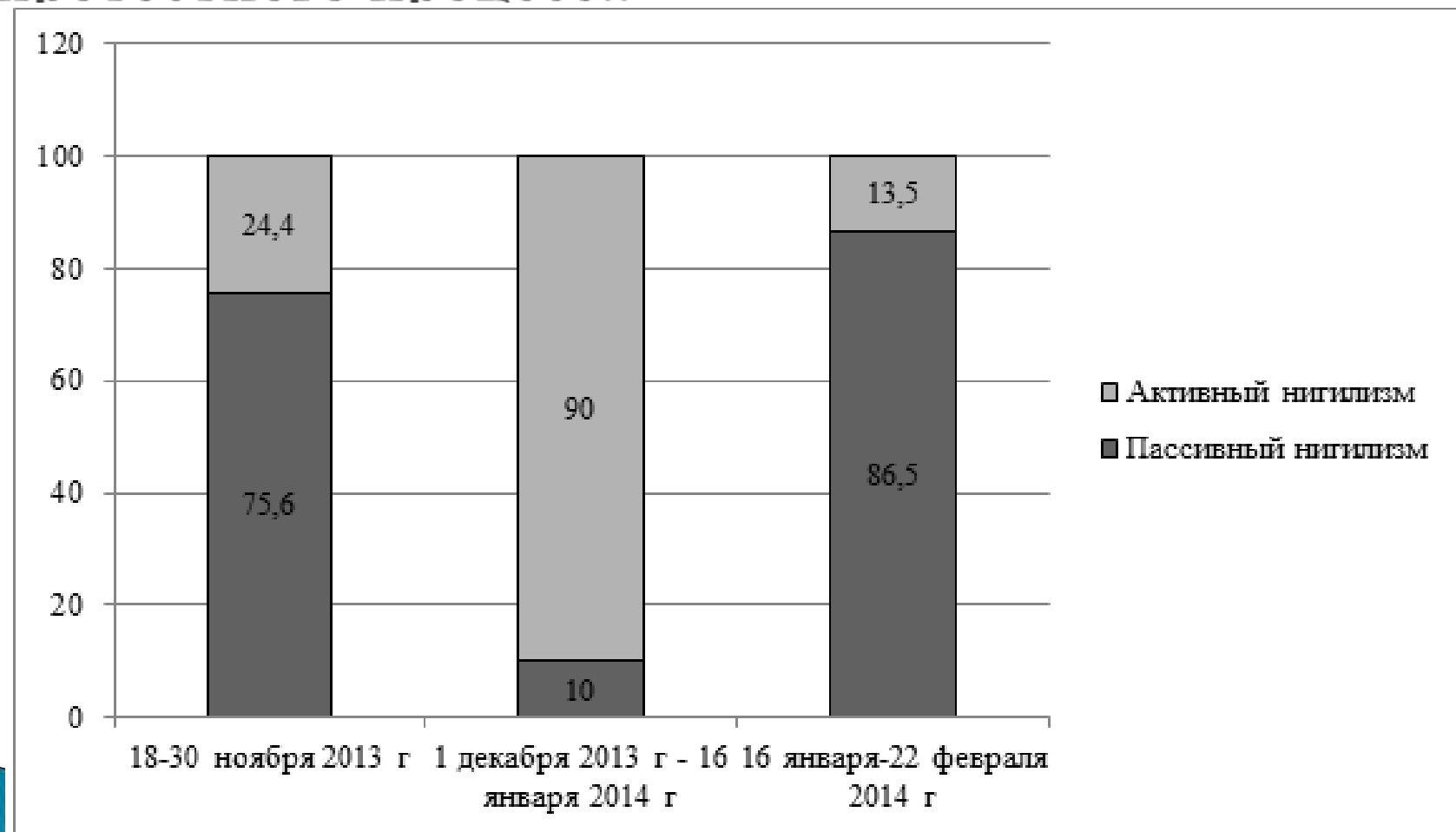
Хронологическое распределение сообщений пассивно-нигилистического характера с учетом акторов (16 января - 22 февраля 2014 года)



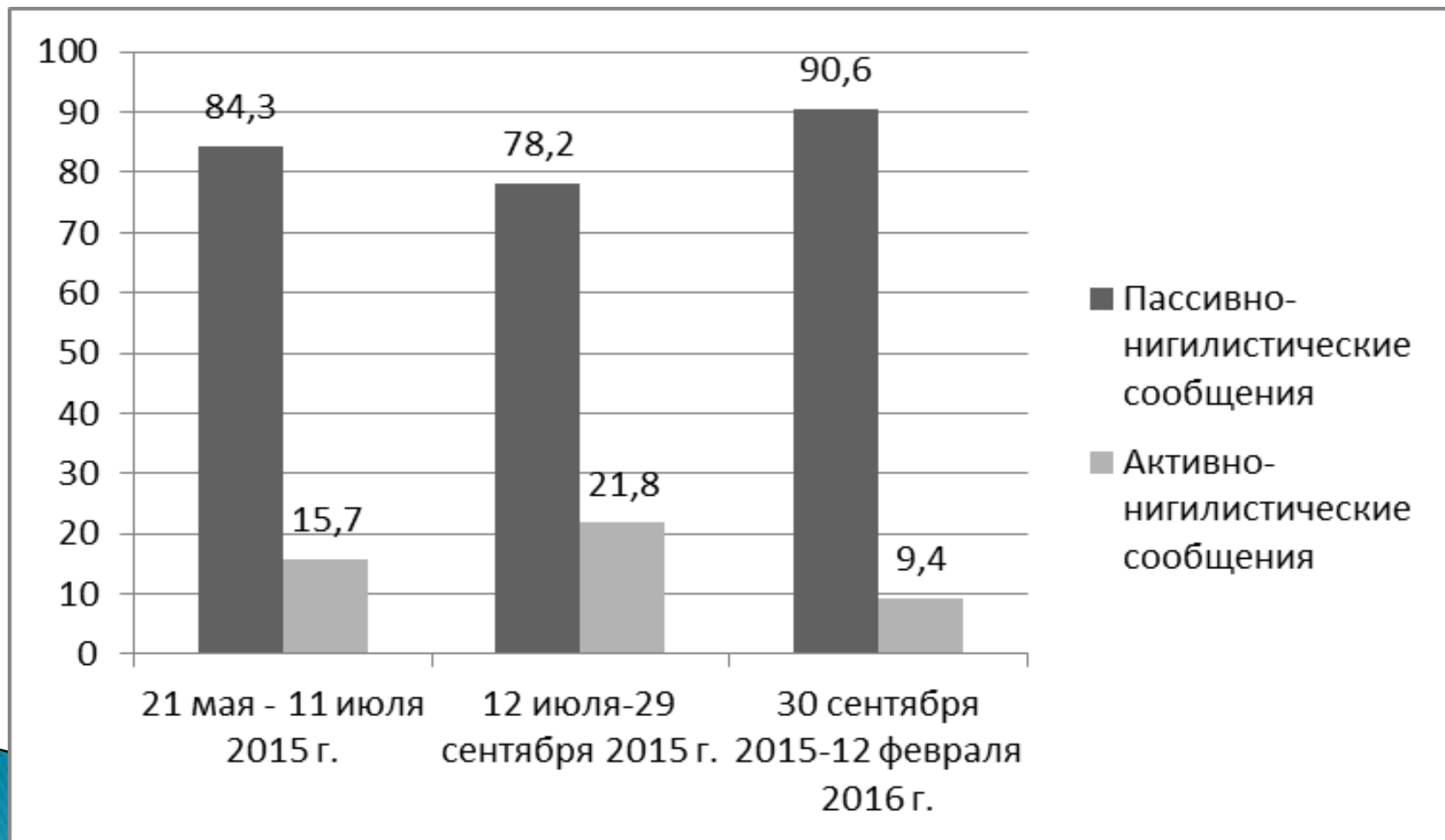
Браузерный классификатор

- ▶ Браузерный классификатор – это приложение к социальному медиа, способное сегментировать информационные потоки по заданным параметрам.
- ▶ Браузерный классификатор обладает большими преимуществами с точки зрения надежности классификации и получения дополнительных функций автоматизированного анализа.

Распределение долей сообщений с лексикой активного и пассивного нигилизма в украинском сегменте социальных медиа в разные периоды протестного процесса

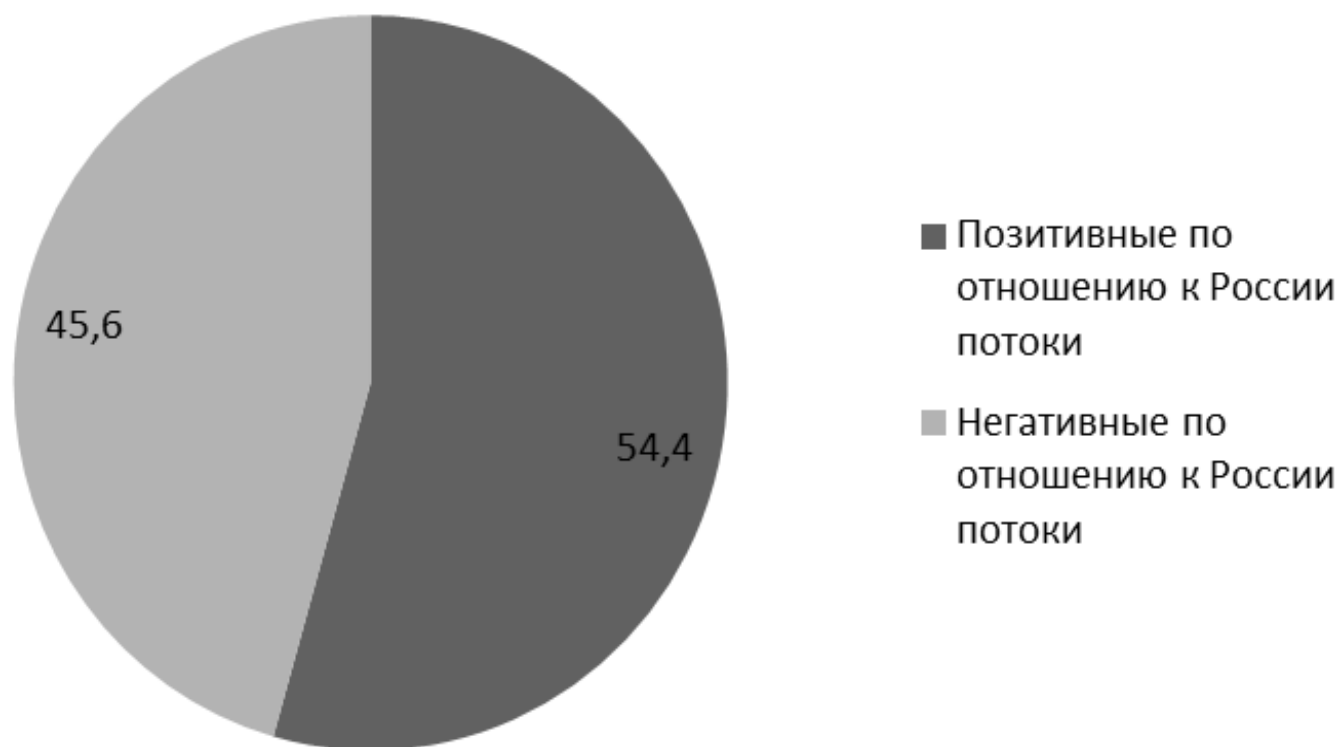


Распределение долей сообщений с лексикой активного и пассивного нигилизма в молдавском сегменте социальных медиа в разные периоды протестного процесса

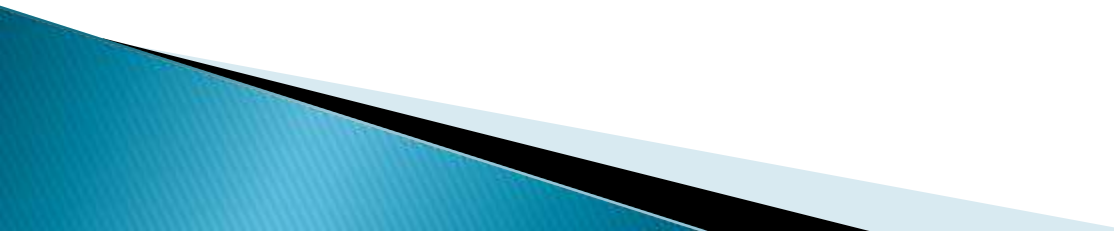


Определение модальности сообщений.

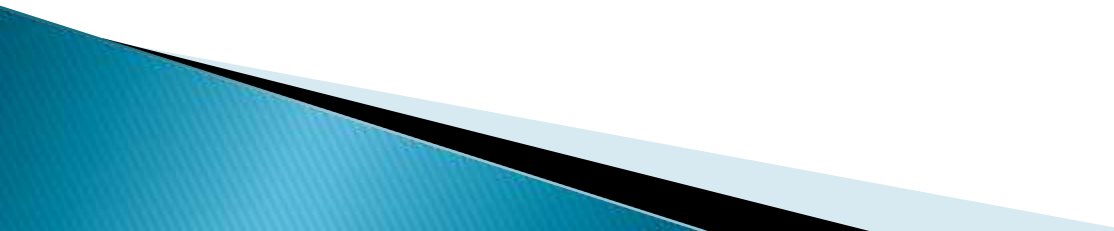
Кейс: информационные потоки в социальных медиа о воссоединении России и Крыма.



Ограничения автоматизированного сегментирования информационных потоков:

- ▶ объем выгружаемых сообщений единовременно;
 - ▶ глубина выгрузки, лимитирующая аккумулялирование сообщений определенной давности;
 - ▶ ирония и подтекстовые смыслы, обусловливающих попадание в выгрузку нерелевантных документов.
- 

Маркеры мобилизации протестных движений в социальных медиа:

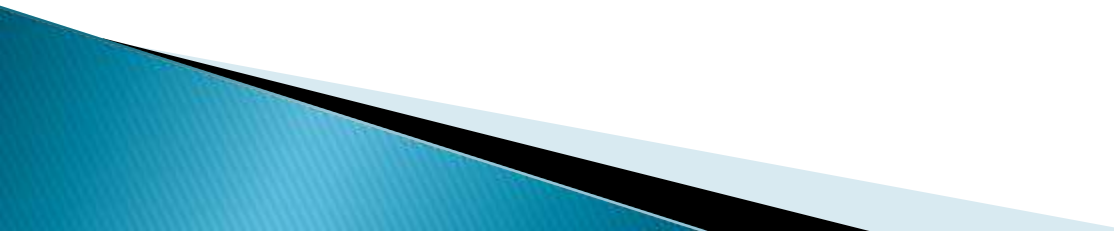
- ▶ маркеры интенсивности информационных потоков, отражающих формирование протестных установок пользователей в социальных медиа;
 - ▶ маркеры содержания сообщений информационных потоков, отражающих формирование протестных установок пользователей в социальных медиа.
- 

Анализ облака тегов.

Облако тегов сообщений о КПРФ и ЛДПР.

Ключевое слово	ЛДПР	КПРФ
Россия	28356	24955
Депутат	22551	24703
Человек	18847	17507
Госдума	18021	14940
Жириновский	14024	-

Маркеры облака тегов

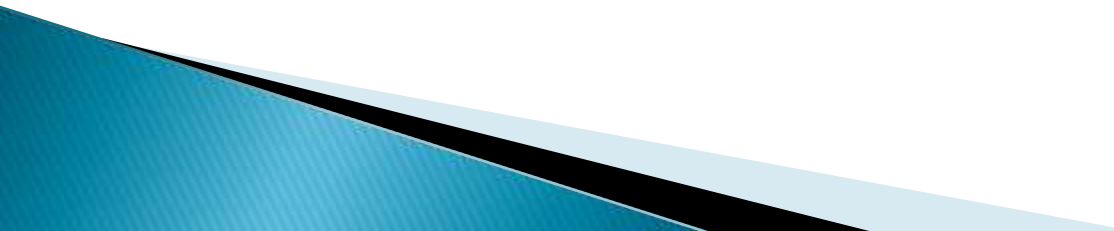
- ▶ частота упоминания слов негативной модальности;
 - ▶ частота упоминания слов позитивной модальности;
 - ▶ частота упоминания слов, имеющих отношения к контекстам поисковых запросов киберметрии.
- 

Геолокационный анализ сообщений социальных медиа.

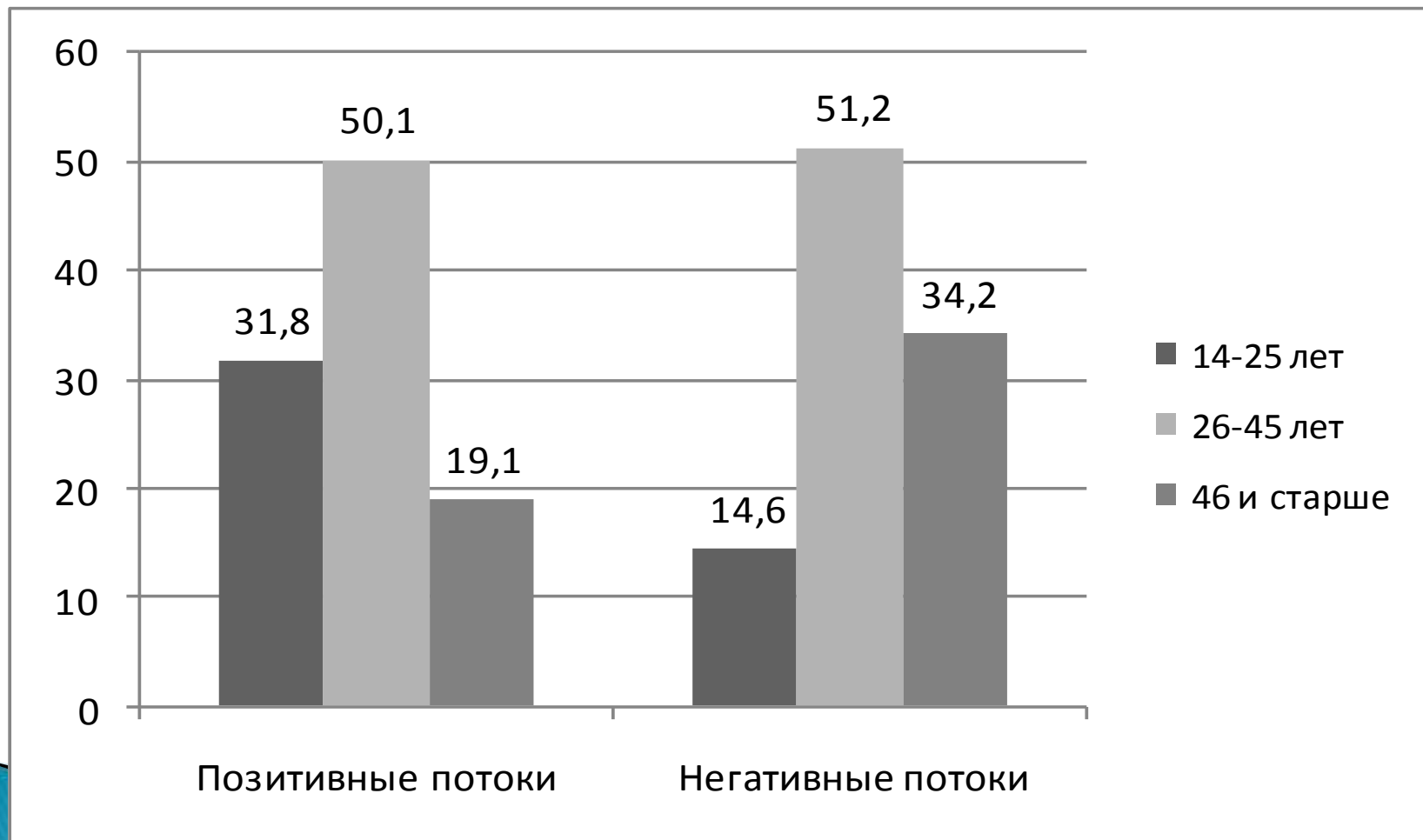
Кейс: антикоррупционная и антиолигархическая риторика в русскоязычном сегменте Интернета 2014-2017 гг.

Страна/Регион	Авторов	Документов
--Москва	1779	3070
--Санкт-Петербург	950	1525
--Московская область	200	318
--Свердловская область	191	227
--Ростовская область	188	252

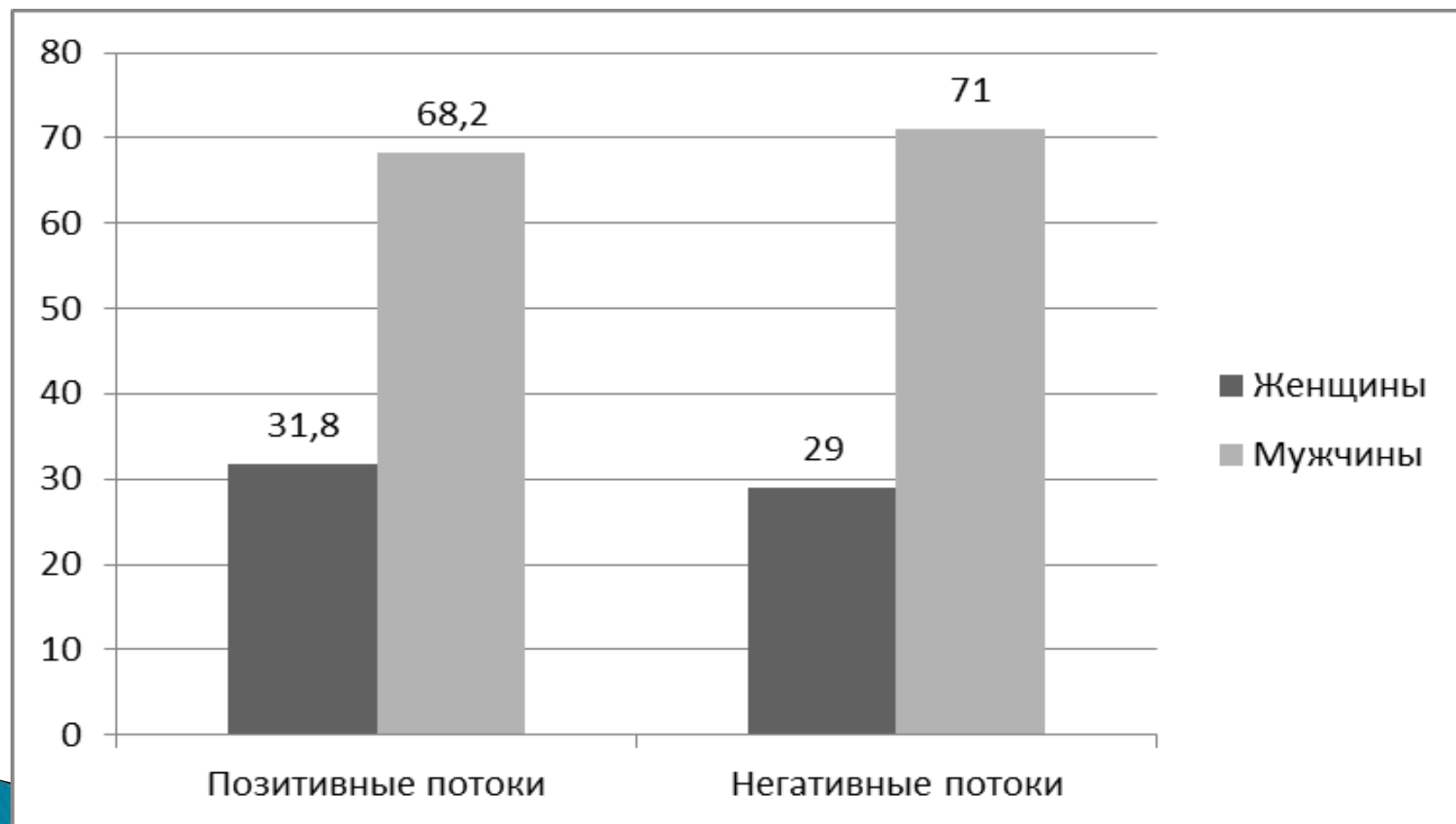
Маркеры геолокационного анализа сообщений

- ▶ доля авторов сообщений и количество документов из столичных регионов;
 - ▶ доля авторов сообщений и количество документов из крупных региональных центров;
 - ▶ доля авторов сообщений и количество документов из зарубежных стран.
- 

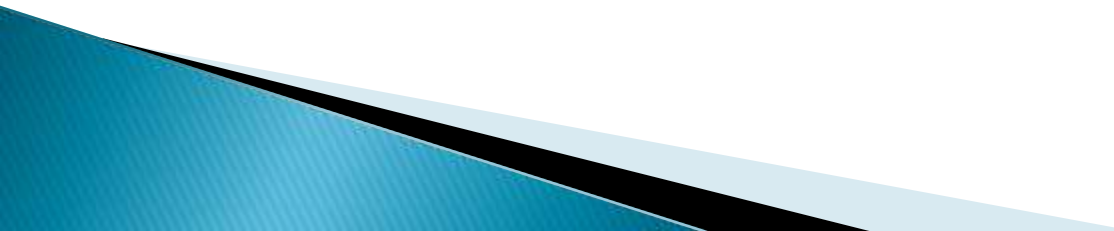
Соотношение пользователей Крыма разного возраста, выражающих позитивное или негативное отношение к воссоединению Крыма и России (03.04.2014-07.06.2016 годы).



Соотношение пользователей Крыма разного гендера, выражающих позитивное или негативное отношение к воссоединению Крыма и России (03.04.2014-07.06.2016 годы).

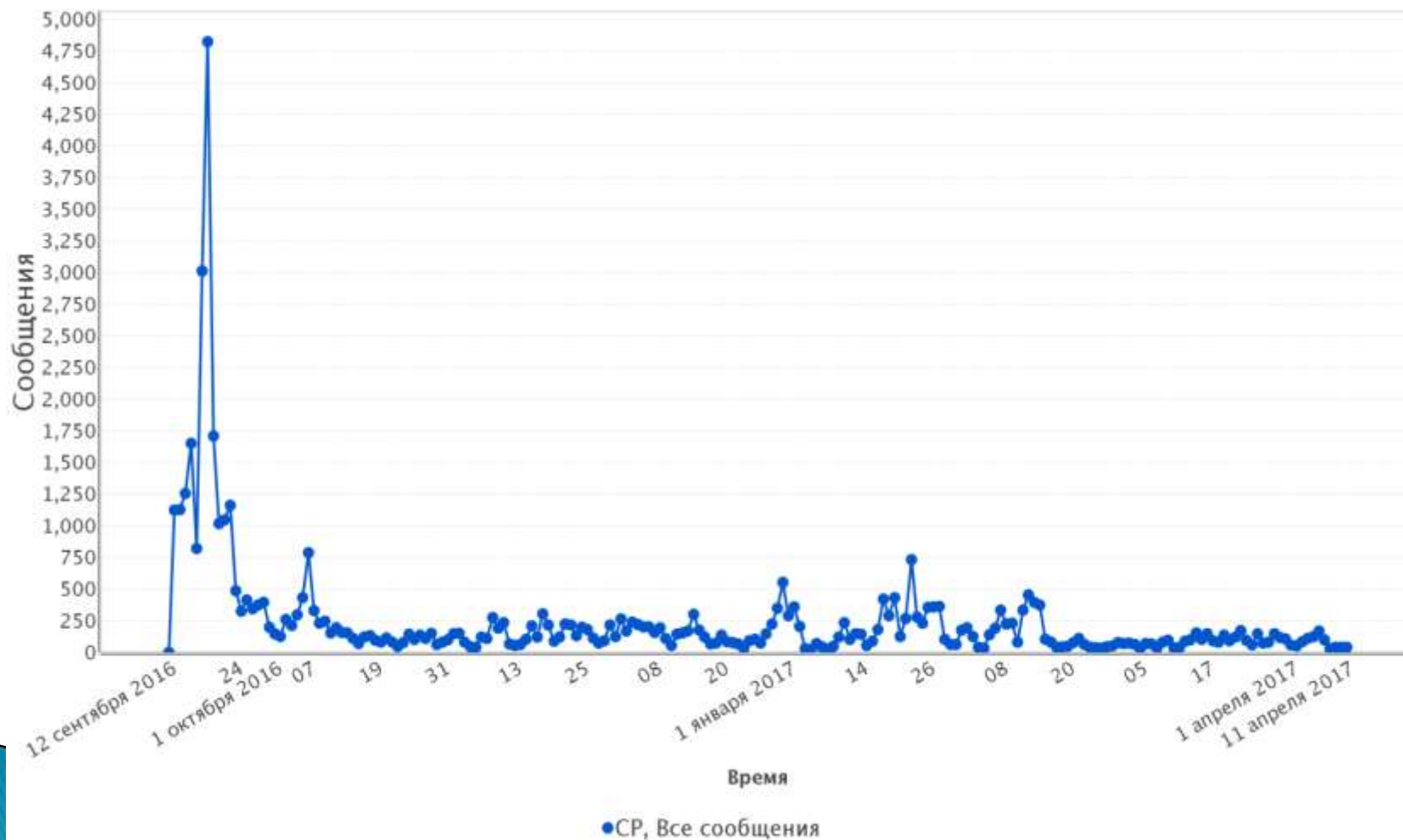


Маркеры половозрастного анализа сообщений:

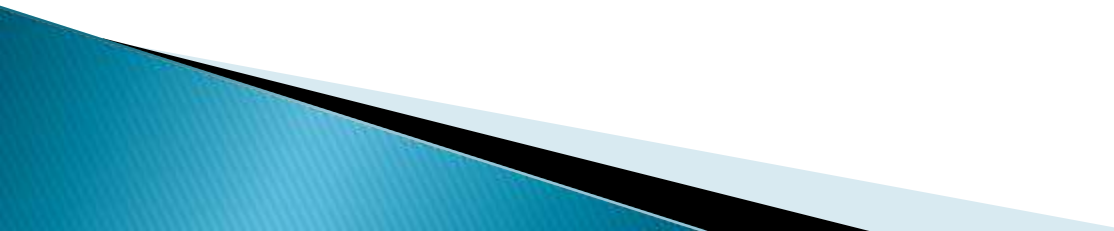
- ▶ доля молодых пользователей среди авторов сообщений;
 - ▶ доля пожилых пользователей среди авторов сообщений;
 - ▶ доля мужчин среди авторов сообщений;
 - ▶ доля женщин среди авторов сообщений.
- 

Анализ динамики информационных потоков.

Кейс: Потоки о «Справедливой России».



Маркеры динамики информационных потоков:

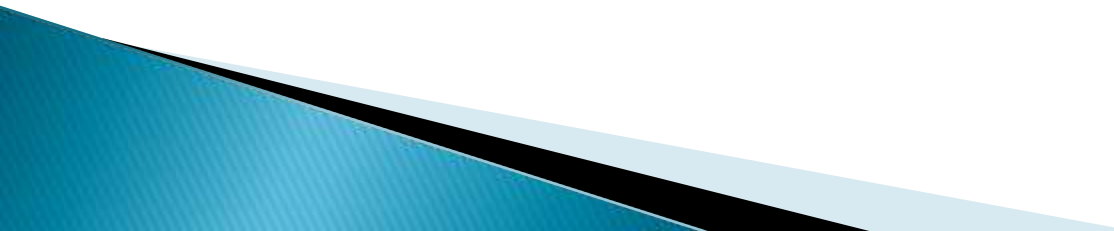
- ▶ интенсивность информационного потока;
 - ▶ вектор информационного потока;
 - ▶ взаимосвязь траектории информационного потока и событиями в оффлайне.
- 

Анализ публикационной активности лидеров общественного мнения.

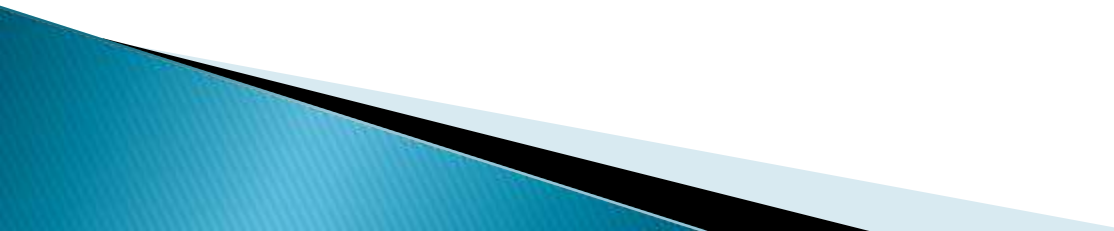
Кейс: лидеры общественного мнения «Евромайдана» и «АнтиЕвромайдана»

Название блога	Блогохостинг	Аудитория	Документов
<u>Типичный Донецк</u>	ВКонтакте	448351	325
<u>РУССКАЯ ВЕСНА</u> (rusvesna.su)	ВКонтакте	63417	207
<u>Новости RT на русском</u>	ВКонтакте	966513	112
<u>Русские Онлайн</u>	ВКонтакте	277329	107
<u>Інформаційний опір</u>	ВКонтакте	90626	87
<u>Сводки ДНР ЛНР Новороссия</u>	ВКонтакте	67254	83
<u>АНГРИФ</u>	ВКонтакте	61228	78
<u>РИА Новости</u>	ВКонтакте	2113225	78
<u>0629 Мариуполь</u>	ВКонтакте	96696	71
<u>Типова Україна</u>	ВКонтакте	344838	64
<u>Типичный Киев</u>	ВКонтакте	693555	63
<u>Коломойский вещает</u>	ВКонтакте	97063	61

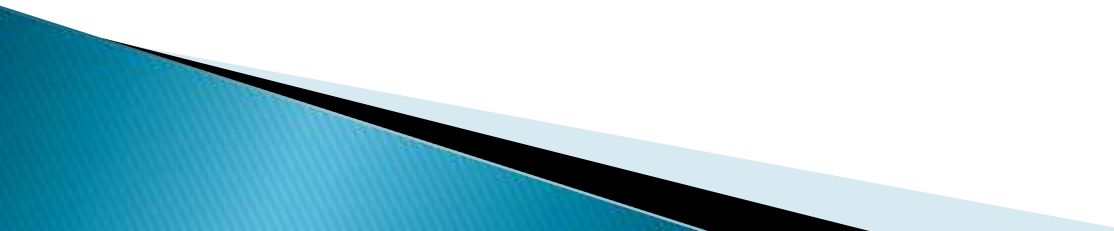
Маркеры лидеров общественного мнения:

- количество подписчиков на странице;
 - быстрая накрутка количества лайков и репостов публикуемых сообщений;
 - перекрёстные ссылки между лидерами общественного мнения;
 - сообщения содержат манипулятивные воздействия.
- 

Построение социальных графов. Программное обеспечение.

- ▶ Авторское программное обеспечение «Социальный граф» позволяет установить структуру взаимосвязей внутри и между онлайн сообществами.
 - ▶ Возможности автоматизированного построения социальных графов обеспечивают понимание схемы связей внутри интересующих исследователя онлайн групп и того, каков характер связи между изучаемыми онлайн сообществами.
- 

Построение социальных графов. Три типа структуры взаимодействия:

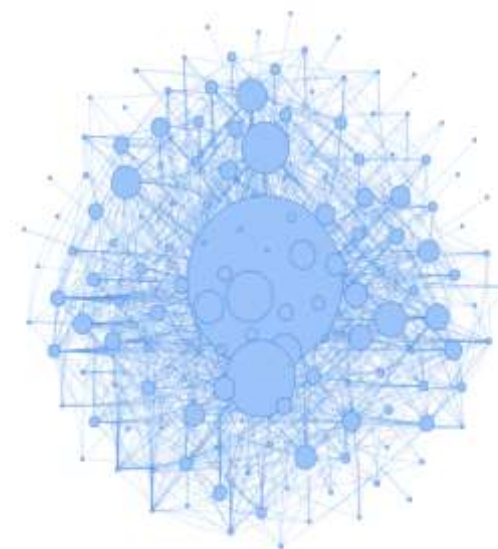
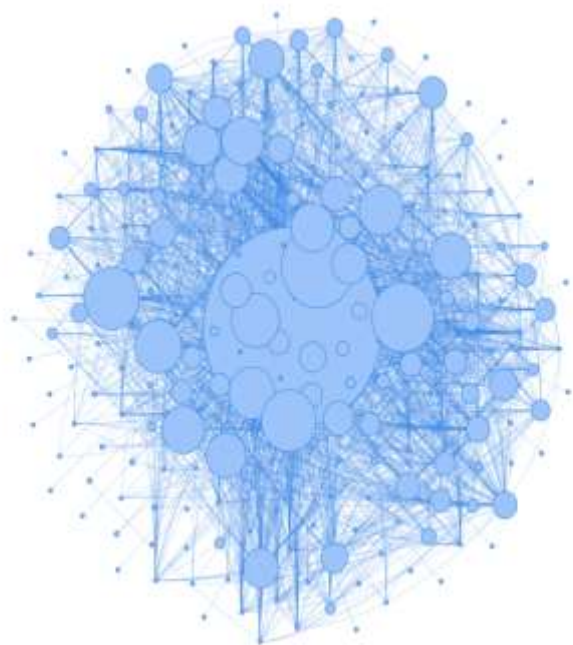
- ▶ группы с большим количеством микросообществ, объединенных общими темами, внутри которых установлены горизонтальные взаимосвязи;
 - ▶ группы с большим количеством лидеров общественного мнения, имеющих относительно небольшую пользовательскую аудиторию;
 - ▶ группы с явным доминированием узкого круга лидеров, формирующих протестные установки широкого круга пользователей.
- 

Первый тип. Структура взаимосвязей внутри группы ВКонтакте «Марш миллиона» (слева) и club 31182341 (справа).

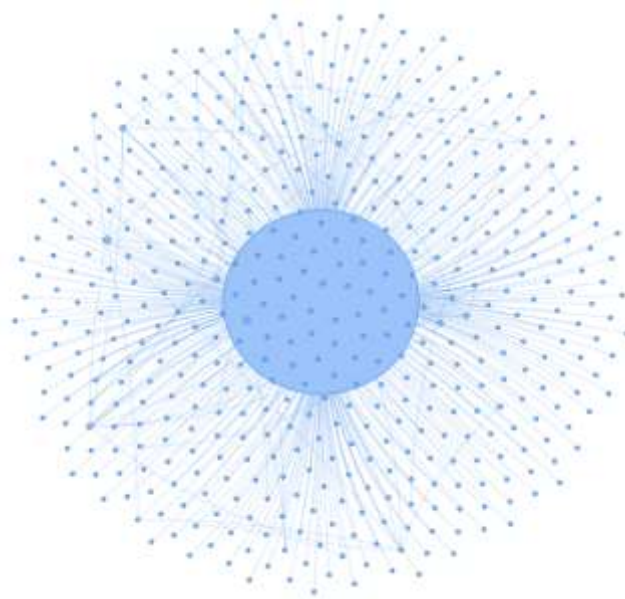
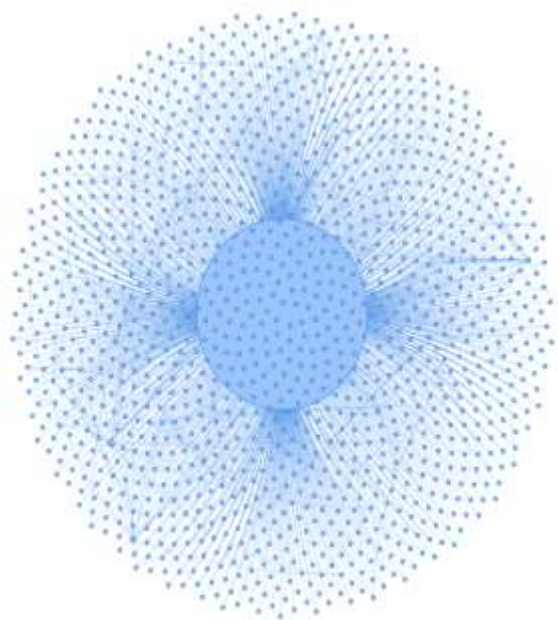
•



Второй тип. Структура взаимосвязей внутри группы Вконтакте «Club 43639012» (слева) и «Манежка» (справа).

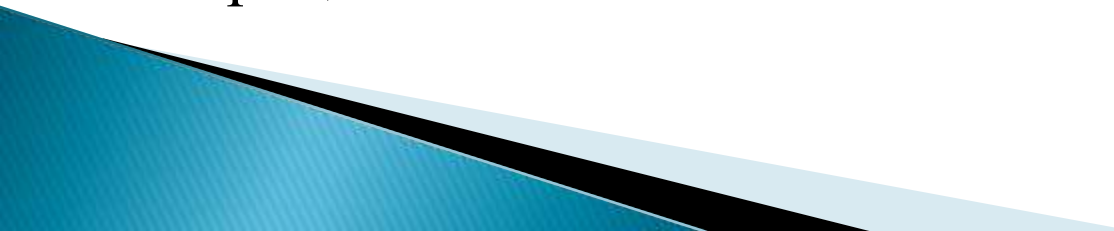


Третий тип. Структура взаимосвязей внутри группы ВКонтakte«ВКонтakte» «Анти_Димон» (слева) и «Манежка_2» (справа)



Выводы

1. Интеллектуальный поиск цифровых маркеров политических процессов основан на высоком измерительном потенциале метода Predictor Mining, позволяющем по цифровым следам исследовать массовое сознание и поведение пользователей, устанавливать тональность общественного мнения, выявлять приемы и субъектов формирования протестных установок в социальных медиа, оценивать степень радикализованности информационных потоков, охарактеризовывать состояние социальной напряженности в обществе в целом и отдельных регионах, анализировать протестный потенциал пользователей, прогнозировать развитие конфликтных процессов.



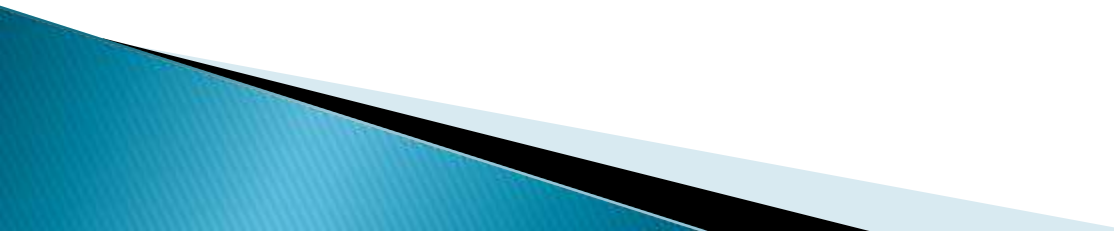
Выводы

2. Эвристический потенциал интеллектуального поиска цифровых маркеров политических процессов основан на комплексе преимуществ автоматического и автоматизированного анализа информационных потоков социальных медиа: неспровоцированный характер анализируемых документов (создаваемых пользователями добровольно, нет жестких рамок предлагаемого социологического инструментария, что повышает степень объективности и достоверности данных); возможность оперирования большими данными, освобождающими от необходимости построения репрезентативной выборки по значимым социальным признакам; использование ряда автоматизированных аналитических опций: геолокационный, социально-демографический, семантический, динамический и др. виды анализа массива данных, выгруженных онлайн сервисом для мониторинга социальных медиа, позволяющих анализировать структуру и факторы политических процессов по цифровым маркерам.

Выводы

3. Построение социальных графов как опция, связанная с анализом взаимодействий внутри и между онлайн-сообществами обладает особой ценностью для исследования политических процессов, т.к. обеспечивает понимание механизма влияния лидеров, формирующих политические установки, на участников групп, силу этого воздействия и схему интеракций внутри и между онлайн-сообществами. Данный вид анализа может служить основанием для разработки альтернативного информационного воздействия, нивелирующего негативные эффекты влияния, направленного на мобилизацию неконвенциональной активности в социальных медиа.

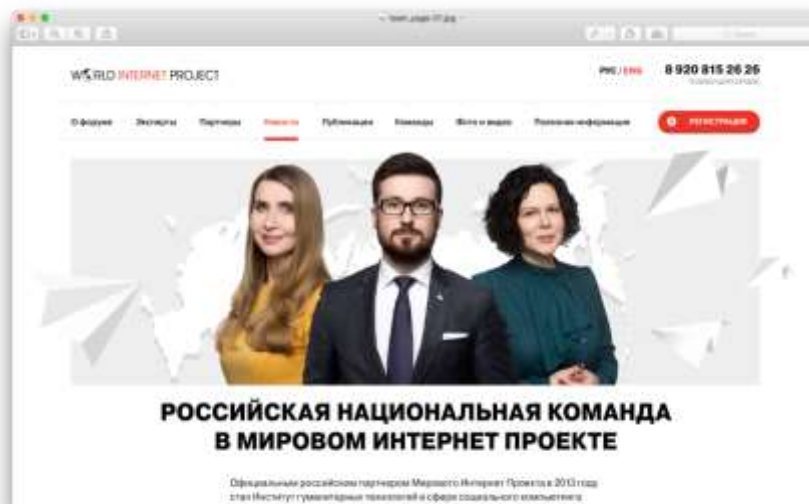
Перспективы исследования.

- ▶ Создание базы цифровых маркеров социальной напряженности на основных языках народов России и постсоветского пространства.
 - ▶ Создание базы цифровых маркеров межнациональных и межрелигиозных конфликтов на основных языках народов России и постсоветского пространства.
 - ▶ Создание базы цифровых маркеров неконвенциональных массовых политических протестов на основных языках народов России и постсоветского пространства.
 - ▶ Создание базы цифровых маркеров распространения экстремизма и терроризма на основных языках народов России и постсоветского пространства.
 - ▶ Создание базы цифровых маркеров фальсификации отечественной и мировой истории на основных языках народов России и постсоветского пространства.
 - ▶ Создание базы цифровых маркеров фейковой новостной информации на основных языках народов России и постсоветского пространства.
 - ▶ Создание базы цифровых маркеров сецессии субъектов Федерации /дезинтеграции общества на основных языках народов России.
 - ▶ Создание базы цифровых маркеров электорального поведения на основных языках народов России.
- 

Основные публикации.

- ▶ Brodovskaya E.V., Dombrovskaya A., Karzubov D. Online mobilization of mass protests in Ukraine, Moldova, Armenia, and Kazakhstan (2013 -- 2016.): the results of comprehensive comparative empirical study // Proceedings of the International Conference on Electronic Governance and Open Society: Challenges in Eurasia St. Petersburg, Russia — September 04 - 06, 2017. ACM New York, NY, USA., 2017 - Pages: 32-36.
- ▶ Бродовская Е.В., Домбровская А.Ю., Карзубов Д.Н. Развитие методологии и методики интеллектуального поиска цифровых маркеров политических процессов в социальных медиа // Мониторинг общественного мнения : Экономические и социальные перемены. 2017. № 5. С. 79—104.
- ▶ Бродовская Е.В., Домбровская А.Ю., Синяков А.В. Стратегии использования социальных сетей в современной России: результаты многомерного шкалирования // Мониторинг общественного мнения: Экономические и социальные перемены. 2016. № 1. С. 283—296.
- ▶ Бродовская Е.В., Домбровская А.Ю., Иванов И.С. Изменение стратегий онлайн-поведения российской интернет-аудитории: результаты сравнительного кластерного анализа (2012–2014 гг.) // Мониторинг общественного мнения: экономические и социальные перемены. - 2016. - 3. С. 173 – 188.
- ▶ Азаров А.А., Бродовская Е.В., Дмитриева О.В., Домбровская А.Ю., Фильченков А.А. Стратегии формирования установок протестного поведения в сети интернет: опыт применения киберметрического анализа (на примере евромайдана, ноябрь 2013 г.) Часть I. // Мониторинг общественного мнения: экономические и социальные перемены. 2014. № 2 (120). С. 63-78.
- ▶ Азаров А.А., Бродовская Е.В., Дмитриева О.В., Домбровская А.Ю., Фильченков А.А. Стратегии формирования установок протестного поведения в сети интернет: опыт применения киберметрического анализа (на примере Евромайдана, ноябрь 2013 г.) Часть II // Мониторинг общественного мнения: экономические и социальные перемены. 2014. № 3 (121).

Контактная информация



Бродовская Елена Викторовна, e-mail:
brodovskaya@inbox.ru

Домбровская Анна Юрьевна, e-mail: an-doc@yandex.ru

Карзубов Дмитрий Николаевич, e-mail:
karzubovdn@gmail.com