

# Университетский Консорциум Исследователей Больших Данных: Примеры Проектов

Михаил Мягков, Цифровизация 2019

[Opendata.university](https://opendata.university)



Национальный  
исследовательский  
**Томский  
государственный  
университет**



**Университетский  
консорциум  
исследователей  
больших данных**



Национальный  
исследовательский  
**Томский  
государственный  
университет**



СЕВЕРО-ВОСТОЧНЫЙ ФЕДЕРАЛЬНЫЙ  
УНИВЕРСИТЕТ ИМЕНИ М.Е. АММОСОВА

# Университетский консорциум исследователей больших данных

Объединение образовательных организаций, реализующих фундаментальные и прикладные исследования в области сбора и анализа больших данных, а также ведущие разработку продуктов и инструментария для работы с большими данными.



Национальный  
исследовательский  
**Томский  
государственный  
университет**



СЕВЕРО-ВОСТОЧНЫЙ ФЕДЕРАЛЬНЫЙ  
УНИВЕРСИТЕТ ИМЕНИ М.Е. АММОСОВА



**Уральский  
федеральный  
университет**  
имени первого Президента  
России Б.Н. Ельцина



СЕВЕРНЫЙ  
АРКТИЧЕСКИЙ,  
ФЕДЕРАЛЬНЫЙ  
УНИВЕРСИТЕТ  
ИМЕНИ М.В. ЛОМОНОСОВА



Факультет  
государственного  
управления



[www.opendata.university](http://www.opendata.university)



Национальный  
исследовательский  
Томский  
государственный  
университет



СЕВЕРНО-ВОСТОЧНЫЙ ФЕДЕРАЛЬНЫЙ  
УНИВЕРСИТЕТ ИМЕНА М.В. АММОСОВА

# Университетский консорциум исследователей больших данных

Индустриальными партнерами Консорциума выступают ведущие компании в области сбора и анализа данных:





# Данные

- 89 млн. валидных аккаунтов социальных сетей (ВКонтакте, Одноклассники, Facebook и др.) с информацией о дружеских связях и подпискам на тематические сообщества
- Контент 142 млн. сообществ в социальной сети ВКонтакте
- Данные онлайн-СМИ, блогов, тематических и региональных форумов
- 13.4 млн контрактов госзакупок с 2014-ого года
- Текстовый контент интернет-энциклопедии Wikipedia на 18 языках
- Профдиагностика 3000 школьников Томской области
- Собственные данные университетов



# Приложения

## Данные + Инфраструктура

### Социально значимые

- Исследование проявлений благотворительности в онлайн-пространстве
- Цифровое качество жизни
- Предсказание политических предпочтений пользователей социальных сетей
- Гражданские инициативы
- Оценка коррупционных рисков на основе данных государственных закупок

### Образование

- Определение образовательных интересов и признаков одаренности у школьников
- Оценка влияния университетов на общество
- Анализ социально-психологического профиля личности

### Коммерческие проекты

- Мониторинг региональных мнений и настроений
- Расчет показателей для Московского международного рейтинга вузов «Три миссии университетов»
- Маркетинговые исследования (“Триколор ТВ”, “Castrol”)
- Аналитические отчеты (МЧС России)

### Безопасность

- Исследование экстремистских онлайн-сообществ в социальных медиа
- Анализ проявлений девиантного поведения среди школьников



Национальный  
исследовательский  
Томский  
государственный  
университет



СЕВЕР-ВОСТОЧНЫЙ ФЕДЕРАЛЬНЫЙ  
УНИВЕРСИТЕТ ИМЕНИ М.К. АММОСОВА

# Московский международный рейтинг «Три миссии университета»



The MosiUR advisory council



# Московский международный рейтинг «Три миссии университета»

Учредитель рейтинга – Российский союз ректоров.

Оператор – Ассоциация составителей рейтингов («Эксперт РА», ВЦИОМ, «Репутация» и др.).

Группа показателей «Университет и общество»:

- Количество онлайн-курсов вуза, размещенных на крупнейших глобальных онлайн-платформах
- Доля вуза в общем объеме публикаций университетов по стране
- Общее количество страниц веб-сайта университета, индексируемых крупнейшими поисковыми системами
- **Количество просмотров страницы вуза в Википедии**
- **Количество подписчиков аккаунта университета в Twitter**
- **Количество выпускников вуза, которым посвящена отдельная страница в Википедии**

\* Показатели, выделенные красным цветом, предложены и рассчитаны ТГУ



## Поиск выдающихся выпускников университета с использованием данных интернет-энциклопедии Wikipedia

- Источник данных — данные страниц интернет-энциклопедии на национальном для университета и английском языках
- Поиск основан на применении лингвистических маркеров и технологий обработки естественного языка
- Формируется список выпускников университета, у которых имеется страница в Wikipedia
- Учитывается статистика посещаемости страницы выпускника для выделения выдающихся и наиболее влиятельных выпускников



# MosiUR 2016

Идентифицировано 187 тысяч выпускников из 346 университетов

ФИО выпускника	Университет	Количество просмотров страницы
<b>Барак Обама</b>	Columbia University	<b>16 673 705</b>
<b>Владимир Путин</b>	Saint Petersburg State University	<b>12 139 843</b>
<b>Илон Маск</b>	Queen's University	<b>11 721 349</b>



# MosIUR 2017

Идентифицировано 265 тысяч выпускников из 473 университетов

ФИО выпускника	Университет	Количество просмотров страницы
<b>Меган, герцогиня Сассекская</b>	Northwestern University	<b>30 430 581</b>
<b>Стивен Хокинг</b>	University of Cambridge	<b>19 183 278</b>
<b>Илон Маск</b>	Queen's University	<b>15 791 090</b>
<b>Чарльз, принц Уэльский</b>	Queen's University	<b>12 944 420</b>
<b>Владимир Путин</b>	Saint Petersburg State University	<b>11 426 497</b>

Топ 5 выпускников по количеству просмотров за 2017 год



The ranking includes 336 international universities from 39 countries.

Country	Number of universities/ number of graduates*	Total number of graduates' page views in 2016
<b>Russia</b>	50 universities / 11,347 graduates	<b>187.8 million</b>
<b>USA</b>	41 universities / 52,460 graduates	<b>988.49 million</b>
<b>China</b>	21 universities / 3,604 graduates	<b>60.8 million</b>
<b>Great Britain</b>	18 universities / 13,533 graduates	<b>350.8 million</b>

\* Including historic figures who graduated from the university



# Поиск талантов в регионе. Кейс СФО

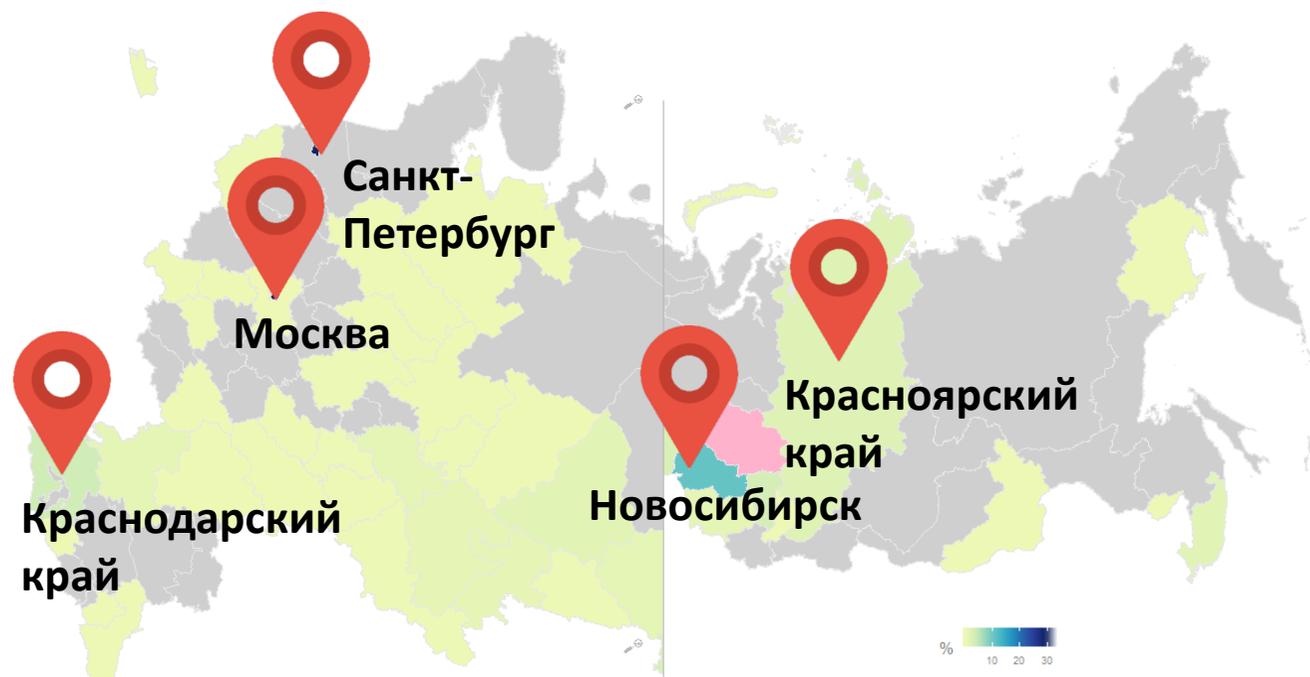
## Цель проекта:

- 1** Идентификация аккаунтов потенциальных абитуриентов в СФО.
- 2** Выявление абитуриентов с высоким образовательным потенциалом.
- 3** Приглашение в приёмную кампании университета через социальную сеть «ВКонтакте»



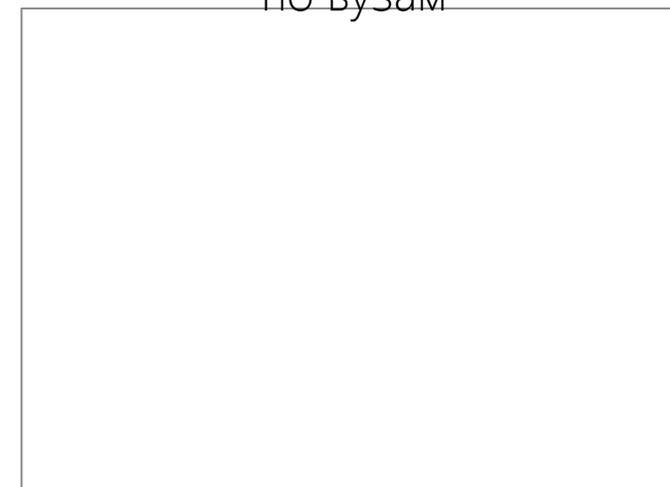
# Поиск талантов в регионе. Кейс СФО

## Центры академической миграции выпускников школ Томской области



22 000 аккаунтов выпускников школ Томской области за 2016 и 2017 годы.  
Данные о географическом положении пользователей и об учебном заведении пользователей

по вузам



по регионам





# Поиск талантов в регионе Кейс СФО

## Поиск и привлечение абитуриентов в социальных сетях

- Цифровые следы абитуриентов в социальных сетях
- Открытые пользовательские данные
- Big Data mining
- Анализ данных с помощью методов машинного обучения и психологии\*
- SMM+контент маркетинг

\* Большая пятёрка OCEAN: экстраверсия, доброжелательность, добросовестность, нейротизм и открытости опыту.

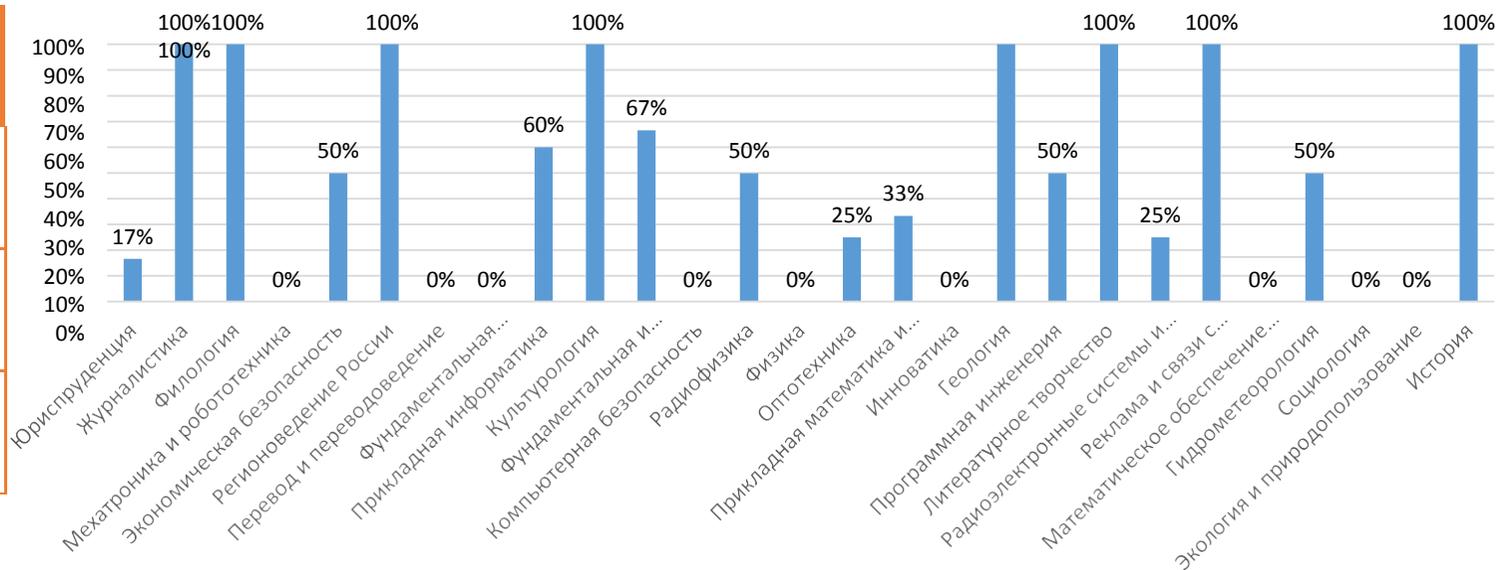


# Результаты рекрутинговой кампании ТГУ 2018

## Общий балл ЕГЭ

	Заявления	Поступления
Рекрутинг 1.0	215,1	219,3
Рекрутинг 2.0	219,8	223,8
<b>дельта</b>	<b>+4,7</b>	<b>+4,5</b>

## Поступаемость «одарённых»



**302 177**

Проанализировано  
профилей  
старшеклассников  
«Вконтакте»

**63 471**

Выявлено абитуриентов  
с высоким  
образовательным  
потенциалом

**25 000**

Приглашены в  
приёмную кампанию  
ТГУ через «Вконтакте»

**530**

Абитуриентов  
подали  
заявления

**228**

Абитуриентов поступило в  
ТГУ: 51 направление  
подготовки. Гуманитариев  
58%, технарей 23%,  
естественников 19%

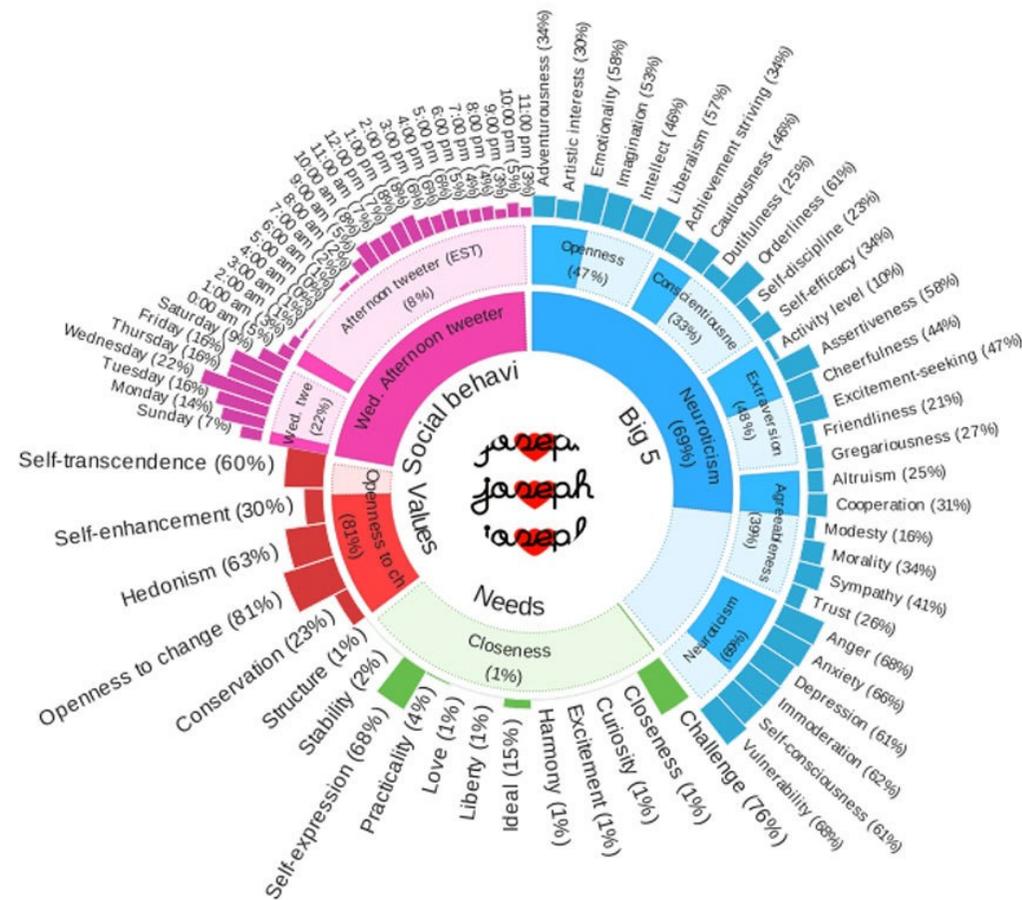
# Идентификация социально-психологического профиля личности

## Цель:

Определение социально-психологического профиля личности, отражающего её возможности обучения и профессиональной самореализации, политических предпочтений, экономическое поведение на основе сопоставимых офлайн и онлайн-данных.

## Задачи проекта

- Разработка модели цифровой коммуникативной личности.
- Сбор данных по методам психосемантического графа и моделирования коммуникативных миров.
- Компаративный анализ соответствия цифровой коммуникативной личности реальному индивиду.

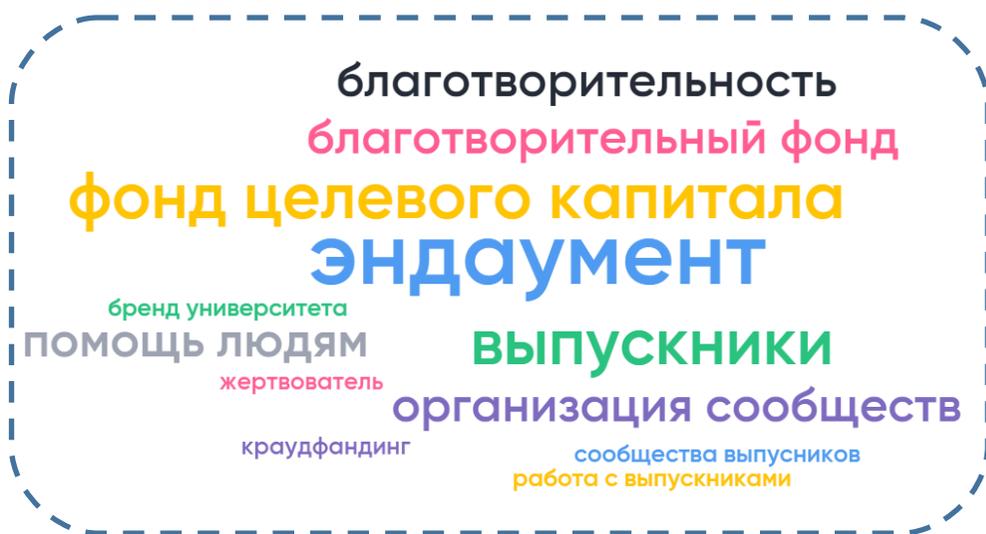




## Вызовы общества и инструменты социальных сетей

Успех Благотворительных фондов и Фондов целевого капитала во многом зависит от решения людей. В настоящее время социальные сети оказывают огромное влияние на принятие решений людьми.

**Вызовы**



**Социальные сети**

Применить инструменты социальных сетей для задач ФЦК и Университетов

Аудитория в РФ: 68.1 млн. 31.6 млн. 25.8 млн. 22.3 млн.



## Цели и задачи исследования

101 

фонд целевых капиталов:

72 

университеты

11 

социальное развитие и благотворительность

10 

культура и искусство

4 

здравоохранение

4 

школьные фонды

- Анализ **форм присутствия** ФЦК онлайн: сайты, соц. сети, площадки, ресурсы
- Оценка **эффективности онлайн-каналов взаимодействия** с жертвователями
- Определение **целевой аудитории** (потенциальной и реальной) ФЦК на онлайн-площадках
- Построение **карты взаимодействия пользователей** – участников сообществ ФЦК в социальных сетях

Выпускники **72**  университетов:

- Построение **социального портрета** выпускника через сбор данных социальных сетей
- Мониторинг **траектории развития выпускников** посредством сбора данных из социальных медиа
- Анализ **активности выпускников** в сообществах университета
- Создание **инструментов** для дополнения существующих методов работы с выпускниками университетов на основе цифровых следов



# Данные

Социальные  
сети



**1.37 млн.**  
профилей выпускников

**1.23 млн.**  
профилей подписчиков ФЦК

**173 млн.** просмотров

**8.9 млн.** лайков

**385 тысяч** комментариев

**437 тысяч** репостов

Местоположение

Должность

Жертвовател?ь?

Образование

Подписки

Тексты постов

Социальные связи

Интересы

Реакция на контент



ФИО  
Жертвователей  
Структура и контент страниц  
Ссылки на аккаунты в социальных  
сетях

ФЦК  
Сайты



**Цель:** построение модели идентификации девиантных сообществ и прогнозирования факторов риска безопасности в социальных сетях

**Задачи:**

создание методологии идентификации сообществ, аккаунтов пользователей, продуцирующих и распространяющих небезопасный контент;

анализ индивидуально-психологических характеристик пользователей во взаимосвязи с паттернами потребления и продуцирования контента социальных сетей;

разработка алгоритма поиска пользователей, являющихся группой риска для воздействия небезопасного контента;

определение групп риска среди пользователей социальных сетей (потребителей небезопасного контента), описание их характеристик.



### Ключевые личности:

- Эрик Харрис (Eric Harris) или Эрик Харис (Eric Haris) или Ерик Харрис или Ерик Харис, Дилан Клиболд (Dylan Klebold);
- Владислав Росляков (Влад Росляков).

### Ключевые даты:

- 9.04.1981 (д.р. Эрика), 11. 09.1981 (д.р. Дилана), 20.04.1999 (массовое убийство в школе «Колумбайн»);
- 17.10.2018 (массовое убийство в Керченском политехническом колледже), 2.05.2000 (д.р. Влада).

### Ключевые места:

- школа «Колумбайн», Литлтон (Littleton), штат Колорадо;
- Керчь, ул. Войкова, д. 1, Керченский политехнических колледж.

### Ключевые слова:

- колумбайн (columbine), колумбина, клмб (clmb), клмбс, клумба, сады, колумбайнеры, NBK – Natural Born Killers (прирожденные убийцы), естественный отбор (national selection), DOOM (любимая игра), золофт и лювокс (антидепрессанты Эрика), Hi-Point 995 и Savage-Springfield 67H и TEC-9 и Stevens 311D (оружие), дневники шутеров, KMFDM (музыкальная группа);
- ружьё «Сайга», НЕНАВИСТЬ (надпись на футболке);
- Яровая, Я\*ровая.



# Идентифицированные сообщества

Категория	Количество идентифицированных сообществ	Количество подписчиков
Пропаганда суицида	39	438850
Жесть (расчлененка, кровь и т.п.)	105	856460
Аниме	289	1756973
Порно, сексуальные	42	269622
True Crime (биографии маньяков, серийных убийц)	6	69499
Сатанисты	49	0 (сообщества заблокированы)
Скулшутинг	11	90983
Национализм	4	47250
Пропаганда депрессии	85	1316238
Алкоголь	15	455356
Оружие	15	168275
Смерть (эстетика смерти)	35	105476
Разжигание ненависти	30	229560
Прочие	13	246491



# Пример девиантного контента



**Columbine Reb&voDKa**

26 авг в 3:21

Ребята,максимальный репост записей ! Чтоб мы понимали,что мы не зря стараемся!



12



3



261

Сначала интересные ▾



**Clair-De-Lune Dans-La-Nuit**

Не стоит, такие группы банят. Чем позже о вас будут узнавать - тем больше инфы вы будете выкладывать сюда.

28 авг в 3:29 Ответить

4



**Игорь Сухарев**

Лол вы хотите, чтобы ко мне пришли фэсэбэ?

30 авг в 4:15 Ответить

1



**Dylan Klebold**

Уже больше сотки, ждём блокировку сегодня в 0:56 Ответить



Написать комментарий...



**Август Вишня**

вчера в 15:08

будьте внимательны садоводы снова в деле)



13



2



56

Сначала интересные ▾



**Софья Роммель**

Чееееее

вчера в 20:11 Ответить

1



**Анастасия Кащенко**



вчера в 15:44 Ответить

1



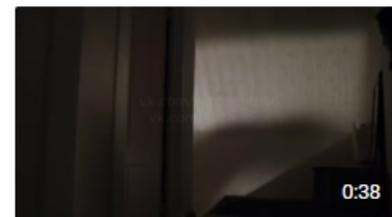
Написать комментарий...



+++

†‡‡ WITCH3RY †‡

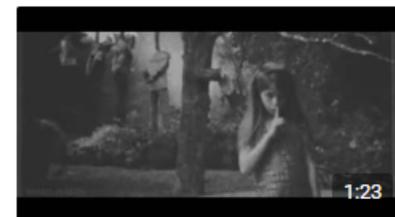
7 534 просмотра · четыре месяца назад



+++

†‡‡ WITCH3RY †‡

1 624 просмотра · пять месяцев назад



†

Александр Эйбрамсон

94 200 просмотров · 11 месяцев назад



# Распространение событий

- **Реальный инфоповод.** Контент (медийный повод) возникает сам. Если он соответствует ожиданиям пользователей – возникает вирусный эффект.
- **Присоединение сторон события к горячей теме:**
  - Участники события, пострадавшие, свидетели, журналисты
  - Эксперты, чиновники, следователи, медики, ответственные лица
  - **Присоединение грибков и паразитов.** На открытую рану садится вредная флора:
    - Политические игроки, ангажированные СМИ и блогеры
    - **Нормальный жизненный цикл:** без подогрева извне естественное событие сейчас живёт 3-5 дней, максимум неделю, с затуханием



# Схемы распространения событий

## 1. Естественное событие

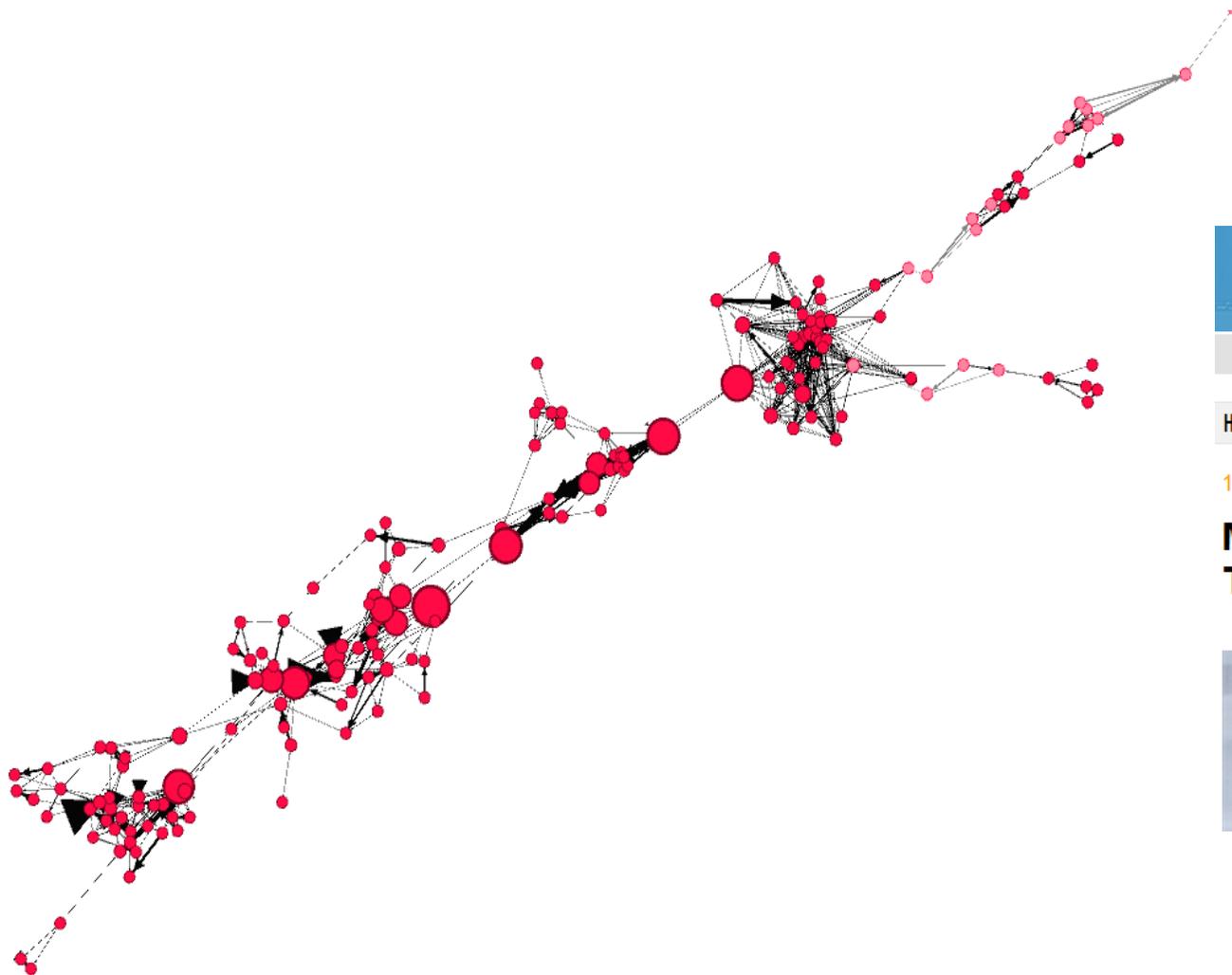
- Много свидетелей и участников - > много разных оригиналов новостей
- Одновременное возникновение новостей в разных местах (СМИ, аккаунты).
- Есть «центры» - много крупных блогеров и СМИ

## 2. Вброс:

- Единый источник контента - > мало оригиналов, много дублей
- Растянутая во времени «колбаса» перепоста и комментирования
- Нет крупных центров на первом этапе, участники – «спящие» аккаунты, мелкие блогеры, крупняк подтягивается позже



# Органичное распространение события



## Томская область

РЕГИОНАЛЬНОЕ ИНФОРМАЦИОННОЕ АГЕНТСТВО

10 декабря 2016 | 13:29 | -16 C

**РИАТОМСК**  
RIATOMSK.RU

Новости Аналитика Мнения Фото Видео Инфографика

НОВОСТИ

11:47 25 ноября 2018 г.

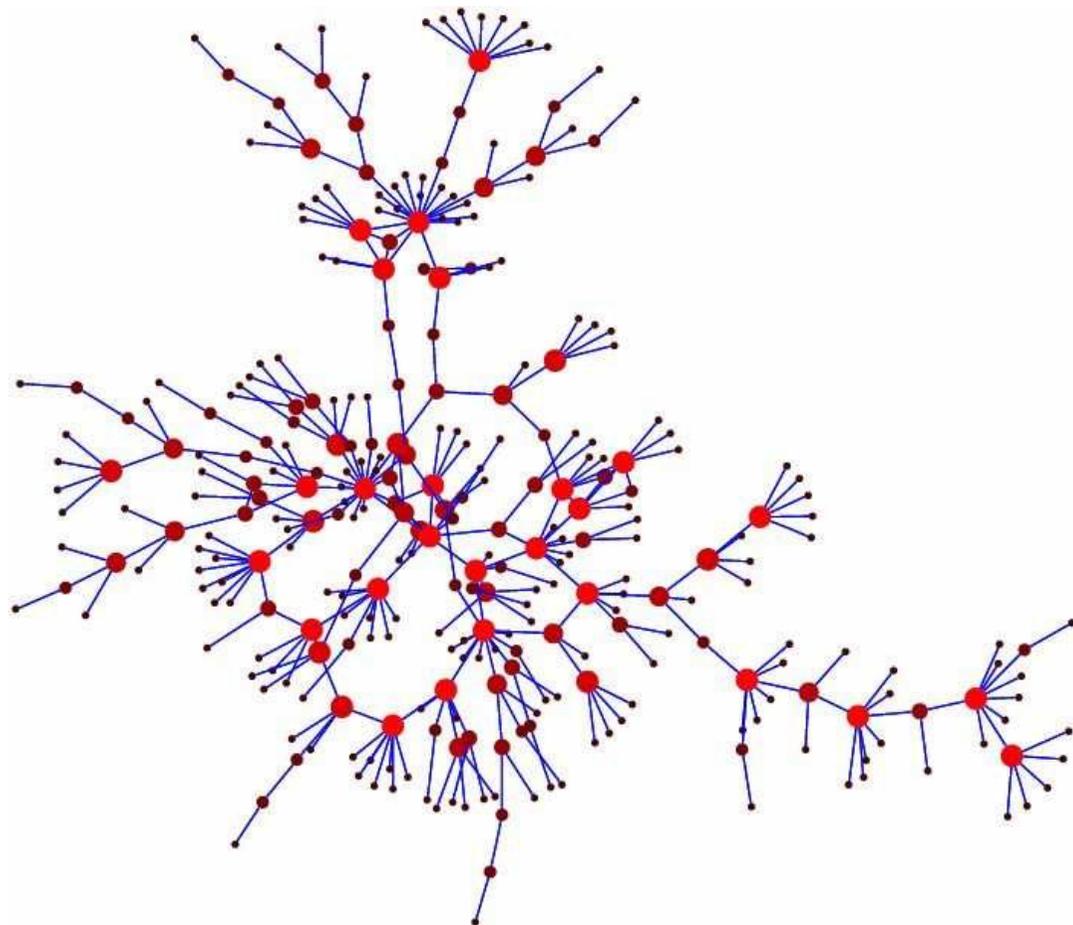
### МЧС объявило оперативное предупреждение в Томской области из-за метели

© РИА Томск. Павел Стеф



# Распространение «вброса»

## Астраханская область



**1** ПРЯМОЙ ЭФИР **НОВОСТИ** ТЕЛЕПРОЕКТЫ ФИЛЬМЫ И СЕРИАЛЫ ДОККИНО СПОР

Погода Юбилей программы "Время" Москва

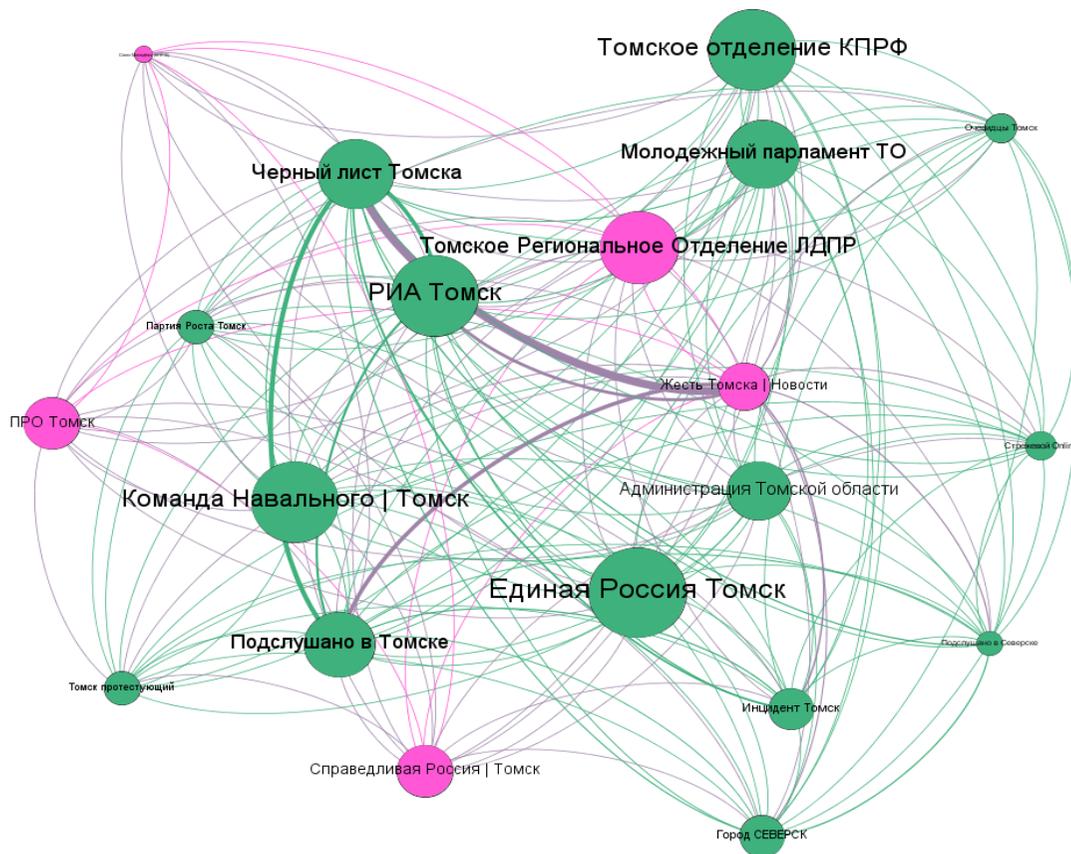
© 14 августа 2018, 15:16

## В «Пусть говорят» — история о чудовищных методах воспитания в одном из детских садов Астрахани



# Ключевые региональные сообщества

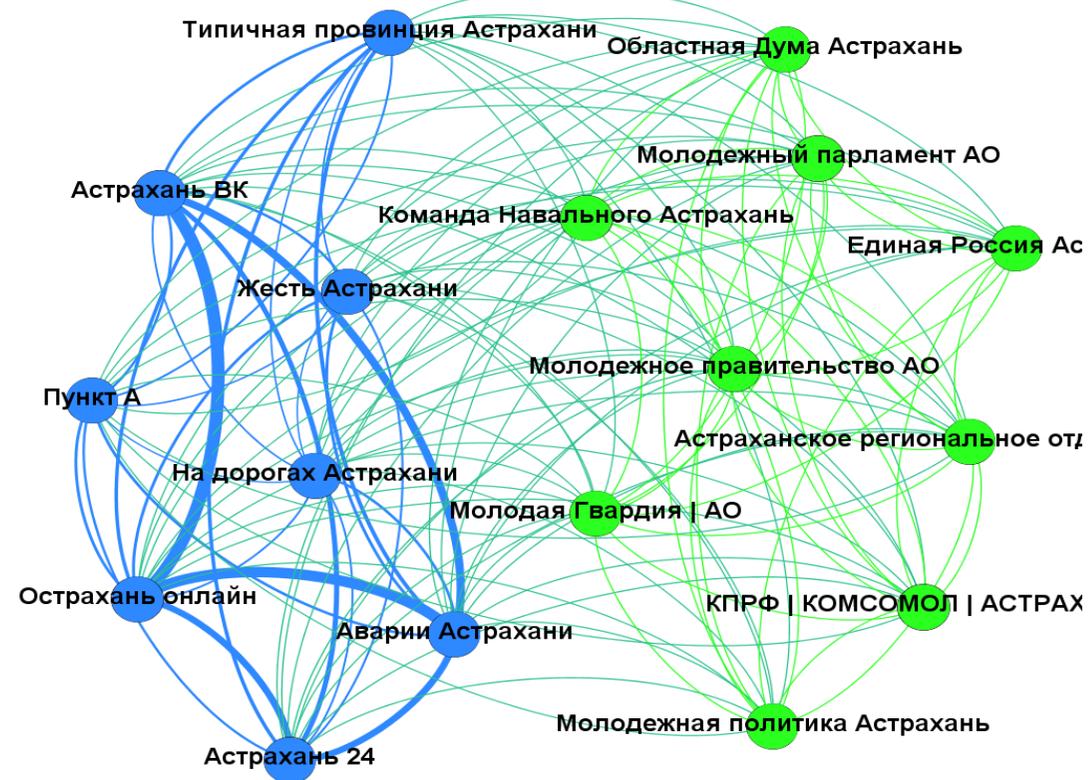
## Томская область



## Томская область

2 кластера, децентрализованная структура

## Астраханская область



## Астраханская область

2 кластера, централизованная структура

# Политические сообщества

## Томская область



Плотное взаимодействие между сообществами «Команда Навального», «Томск протестующий» и «Томское отделение КПРФ»

## Астраханская область



Плотное взаимодействие между молодежными движениями Астрахани



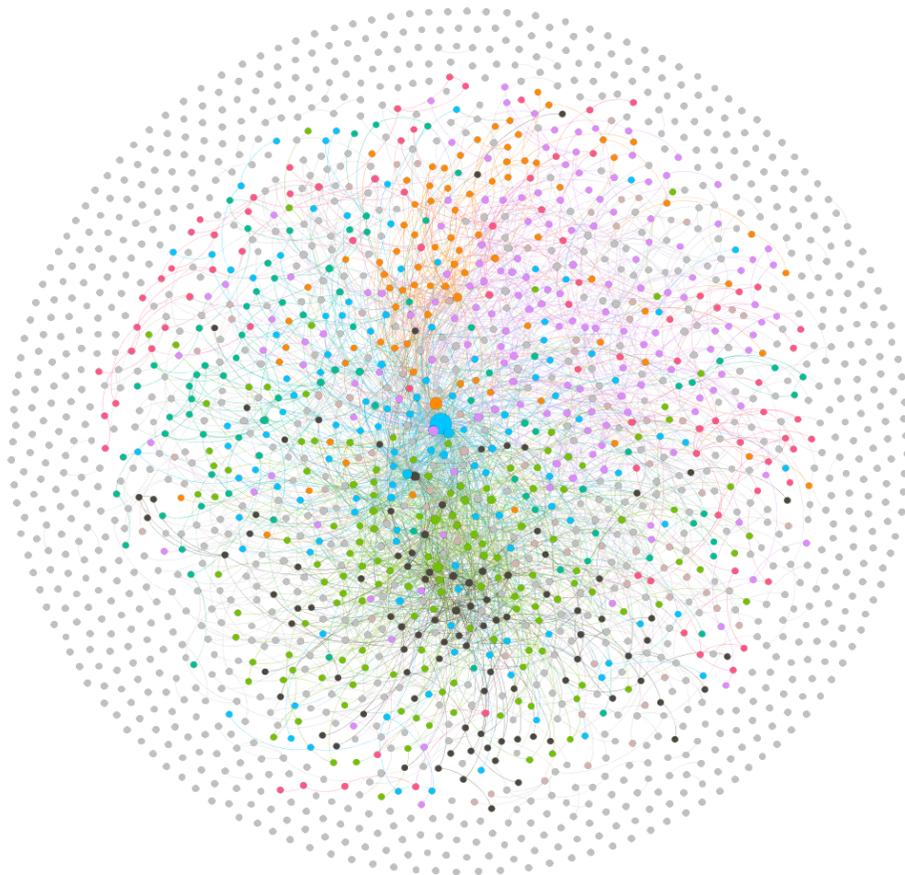
Национальный  
исследовательский  
Томский  
государственный  
университет



СЕВЕР-ВОСТОЧНЫЙ ФЕДЕРАЛЬНЫЙ  
УНИВЕРСИТЕТ ИМЕНИ М.Е. АММОСОВА

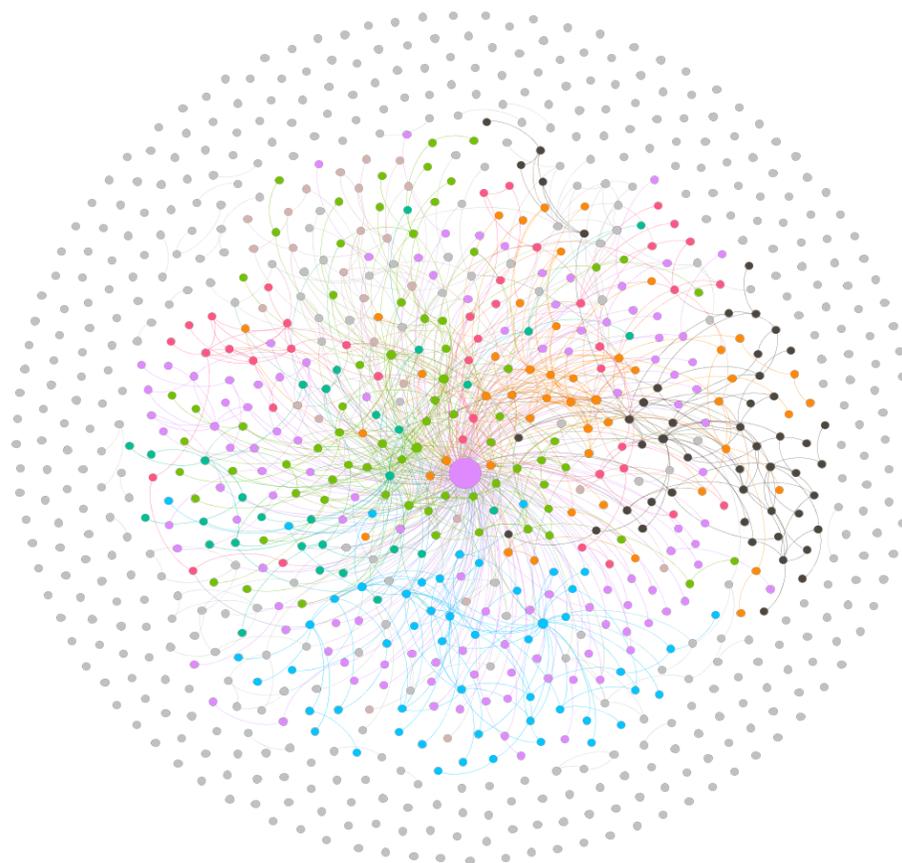
# Команда Навального

## Томская область



1 917 участников  
2 950 связей  
Граф децентрализован  
Низкая плотность взаимодействия

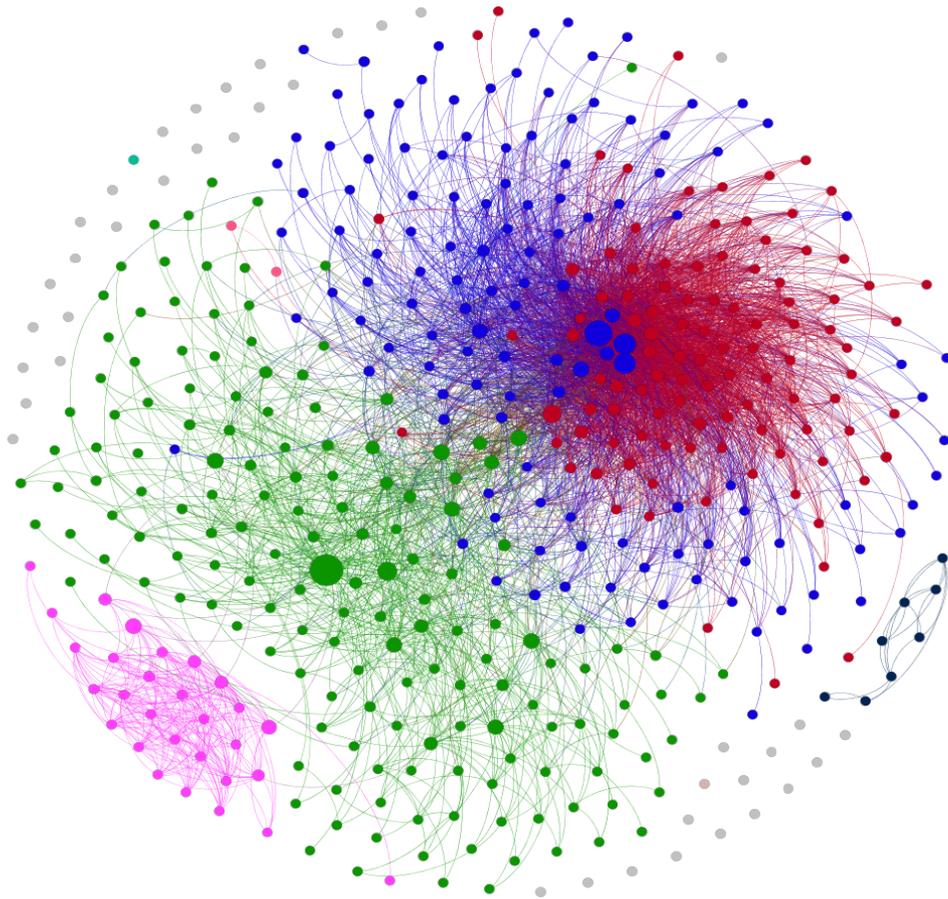
## Астраханская область



907 участников  
1074 связей  
Граф децентрализован  
Низкая плотность взаимодействия

# Единая Россия

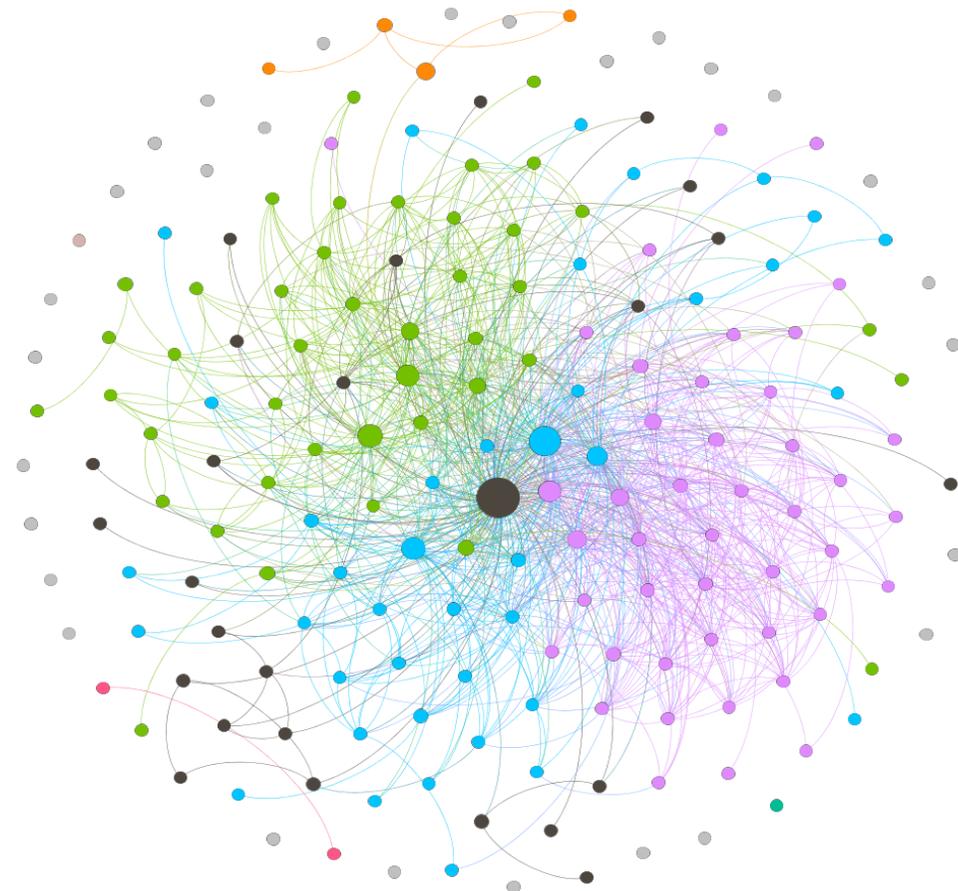
## Томская область



**470 участников**  
**4 606 связей**

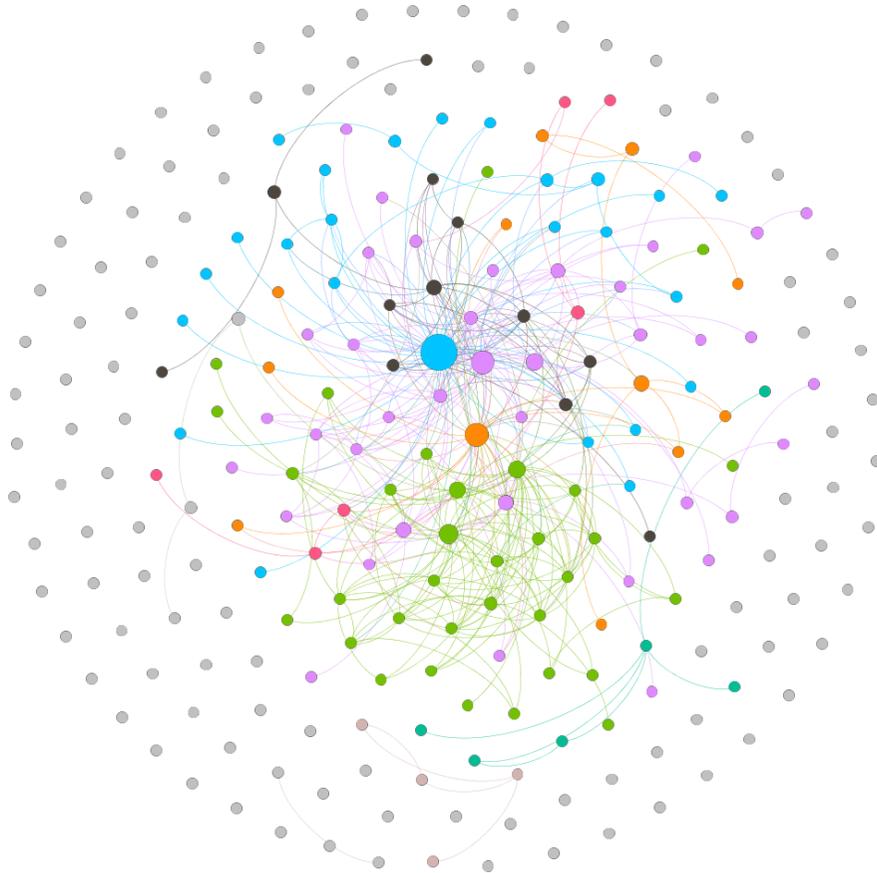
**5 плотных кластеров**  
3 – централизация  
2 - децентрализация

## Астраханская область



**189 участников**    **3 слабых кластера**  
**1 096 связей**

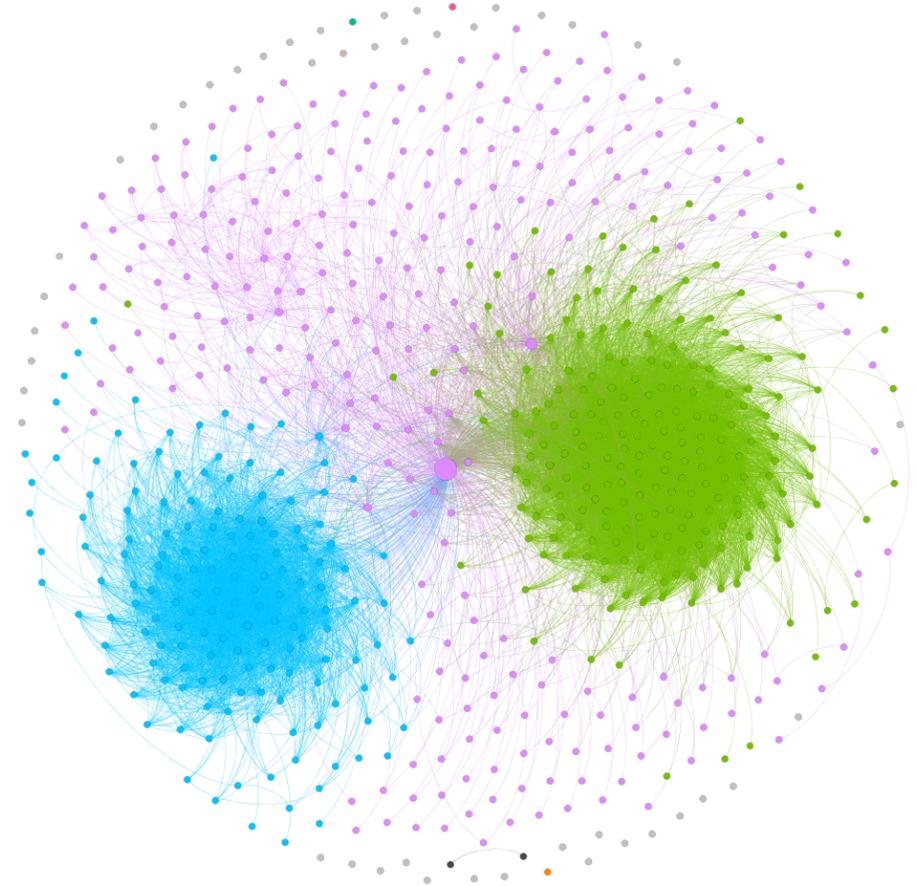
## Томская область



**245 участников**  
**385 связей**

**Разреженная сеть**  
**Парное взаимодействие**

## Астраханская область



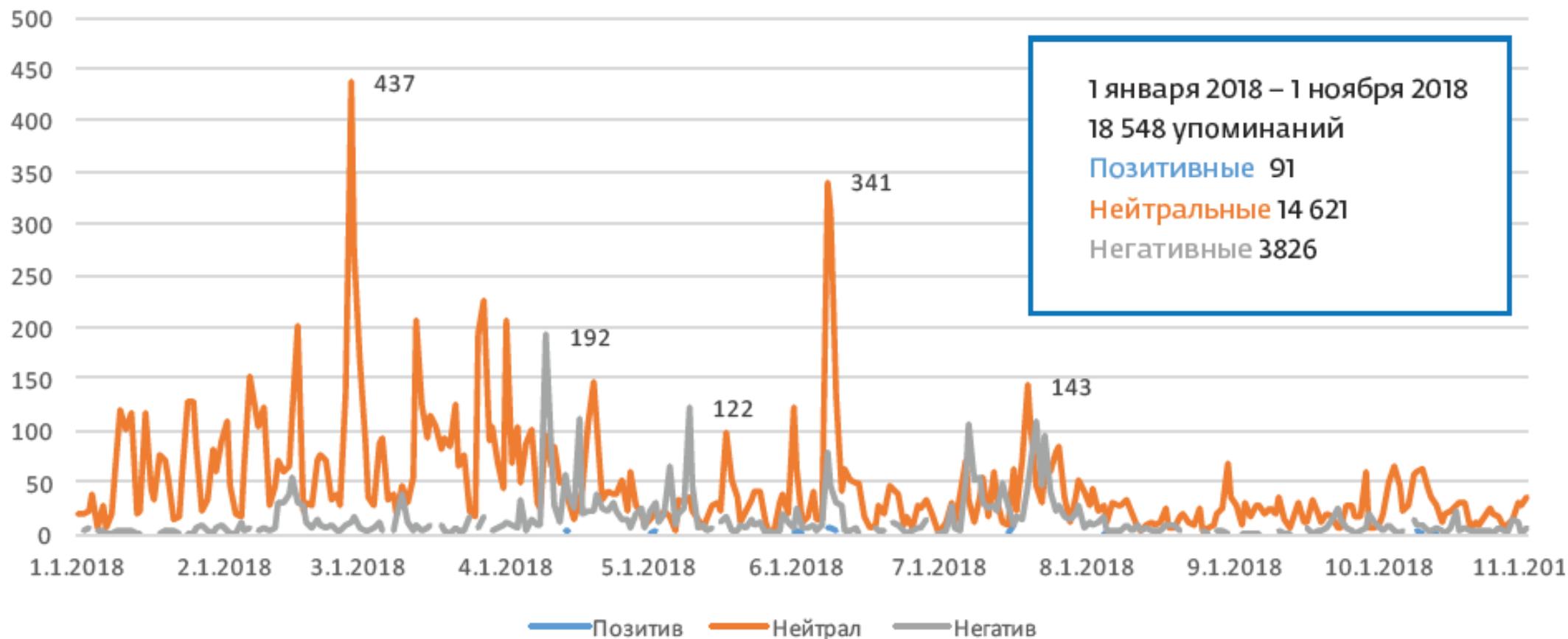
**684 участников** **2 плотных кластера**  
**13 689 связей** **взаимодействие через «мост»**



# Мониторинг социальных сетей

## Губернатор Томской области Сергей Анатольевич Жвачкин

vk.com, facebook.com, twitter.com, riatomsk.ru, odnoklassniki.ru, instagram.com, tvtomsk.ru...



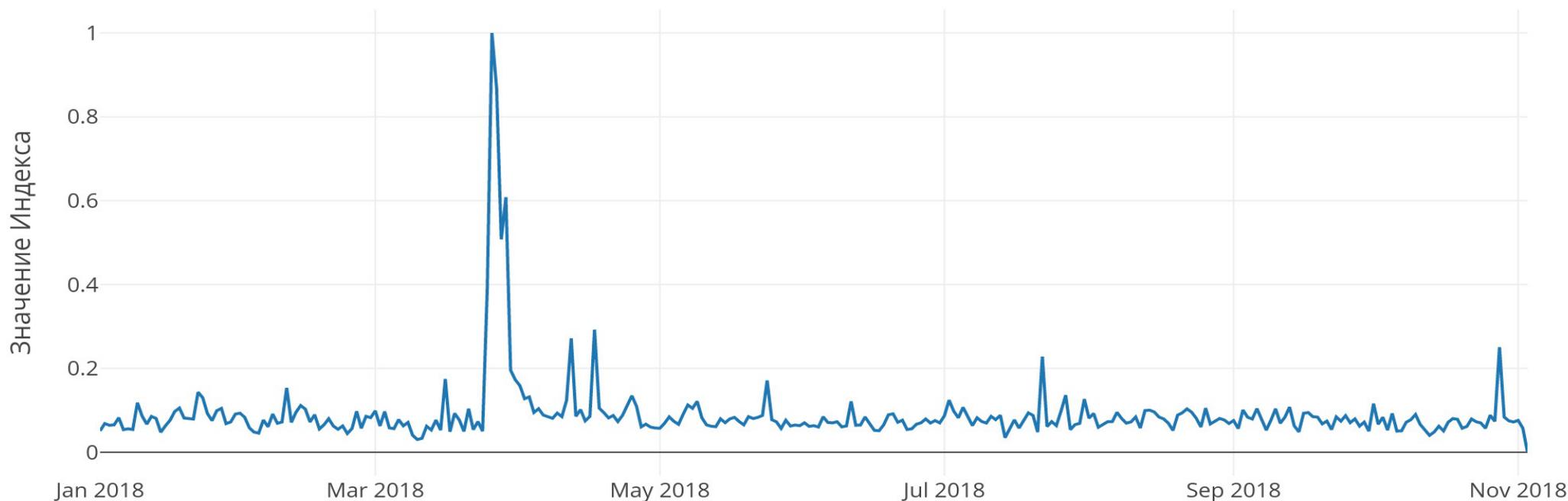


# Индекс информационной активности

## Кемеровская область

1 января 2018 — 1 ноября 2018

(Типичный Кемерово, Регион 42 Кемерово, Новокузнецк ОНЛАЙН, Говорящий Междуреченск, VSE42.RU Новости Кемерово Новокузнецк Кузбасс 42, А42.RU Новости Кемерово и Кузбасса и др.)



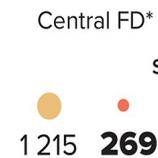
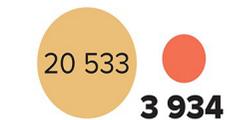
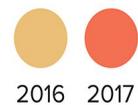
$$I = \frac{\textit{likes} + 5 * \textit{reposts} + \textit{comments}}{\text{количество людей}}$$

$$I^* = \frac{I_1 + I_2 + \dots + I_n}{\text{количество сообществ}}$$



# Мониторинг радикальных настроений

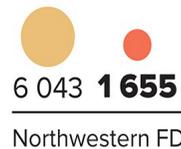
Total number  
of members



Crimean FD +  
Sevastopol



North-  
Caucasus FD



Northwestern FD

St. Petersburg



Moscow



Central FD\*

Sevastopol



Volga  
(Privolzhsky) FD

1844

Ural FD

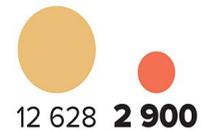


Siberian FD

\* FD =  
Federal  
District

2 342 685

Far East FD



Southern  
FD

CIS  
countries 549 151

USA 1 765 614

Germany 2 158 713

Other  
countries 5 855 3 116

Ukraine 21 498 8 496

Belarus 3 628 948

Kazakhstan 1 239 412





# ТОП 5 медицинских учреждений города Томск



## Черный список Томска

24 дек 2016 в 15:04

✓ Вы подписаны

ЦСМ на Киевская, 15

Ужасное обслуживание, при обращении к аллергологу даже кровь на анализы не взяли, а просто со слов назначили лечение, прописали лекарства, ночью стало ещё хуже, персонал относился ужасно, все процедуры проходила платно

Инесса Грачева-Джаббарова



## Белый список Томска

15 янв 2016 в 13:22

✓ Вы подписаны

Ветеринарная клиника "ТомВет" пер. Совпаптшкольный 10а

Если Вашему питомцу нужна помощь СПЕЦИАЛИСТА то Вам сюда! )

Обратились именно в эту клинику, поскольку в нашей ситуации необходима была срочная помощь именно специалистов да и еще чтоб клиника была технически оснащена (не просто стол и холодильник с лекарствами). В голову пришел ТомВет, и мы на всех парах полетели туда. Увидя состояние животного, администратор немедленно определила нас к врачу менуя очередь (да простят меня ожидающие). Пока питомец проходил все процедуры (порядка шести часов) мне и чаю налили (дабы нервишки пришли в себя) и про питомца моего не забывали и добрым словом поддерживали и если я на время отходила от своей животинки "подменяли" меня, так сказать, держали друга за лапу в место меня. ))) Теперь мы регулярно наблюдаемся в ТомВете. Что мне нравится?

## Топ-5 плохих клиник

1. Центр семейной медицины
2. Поликлиника СибГМУ
3. МСЧ №2
4. Тимирязевская больница
5. Городская больница №3

## Топ-5 хороших клиник

1. Стоматологическая клиника «Зубок»
2. Стоматологическая клиника «Колибри»
3. Ветеринарная клиника ТомВет
4. Центр семейной медицины
5. Лечебно-диагностический центр



Университетский  
консорциум  
информационной  
безопасности



National Research  
Tomsk  
State  
University



Факультет  
государственного  
управления

# Школа прикладного анализа данных

Совместная школа Томского государственного университета и Факультета государственного управления МГУ им. Ломоносова при поддержке университетов Консорциума



г. Москва



10 - 13 декабря 2019г.

Школа ориентирована на исследователей из разных областей наук социологов, политологов, лингвистов, экономистов, технических специалистов

Индустриальные партнеры Школы:



ИНСТРУМЕНТЫ ПАРТНЕРОВ

Обучаем навыкам **прикладного анализа данных** с применением инструментов партнеров. Решаем реальные кейсы в социально-значимых областях. Разрабатываем **стратегии создания** ваших исследовательских **центров** и **лабораторий**. Предоставляем площадку для **нетворкинга** и формирования **межуниверситетского** исследовательского сообщества.

ЗНАНИЯ И НАВЫКИ

ТРЕНДЫ И ПЕРСПЕКТИВЫ

РЕАЛЬНЫЕ КЕЙСЫ

СВОЙ ПРОЕКТ

**ЦЕНТР** или **ЛАБОРАТОРИЯ**  
в вашем университете

[www.opendata.university/school](http://www.opendata.university/school)