

DataFest⁶, 11 мая 2019

Визуализация и анализ комментариев на ютубе

Антон Костин
МФТИ

Слайд 1. Исследовательская установка

- **Проблема:** анализ ютуб-каналов затруднён для невладельцев
- **Решение:** анализировать комментарии
- **Результат:** имеем картину аудитории ютуб-каналов и их связи друг с другом

Слайд 2. Инструменты и метод

Алгоритм:

- 1) Привязать ID комментаторов к объектам (топикам)
- 2) Повторяющиеся в объектах комментаторы дают пересечение

Инструменты:

- 1) Gephi: <https://gephi.org>

Алгоритм укладки Force Atlas 2

Для отображения статистики: функции «Средняя (взвешенная) степень» и «Модулярность»

- 2) Скрепер (например, <http://ytcomments.klostermann.ca>)

	A	B
1	Source	Target
2	User1	1
3	User2	1
4	User3	1
5	User2	1
6	User3	1
7	User4	1
8	User4	2
9	User5	2
10	User6	2
11	User7	2
12	User2	2
13	User8	2
14	User9	2
15	User10	2

Слайд 3. Визуализация в Gephi

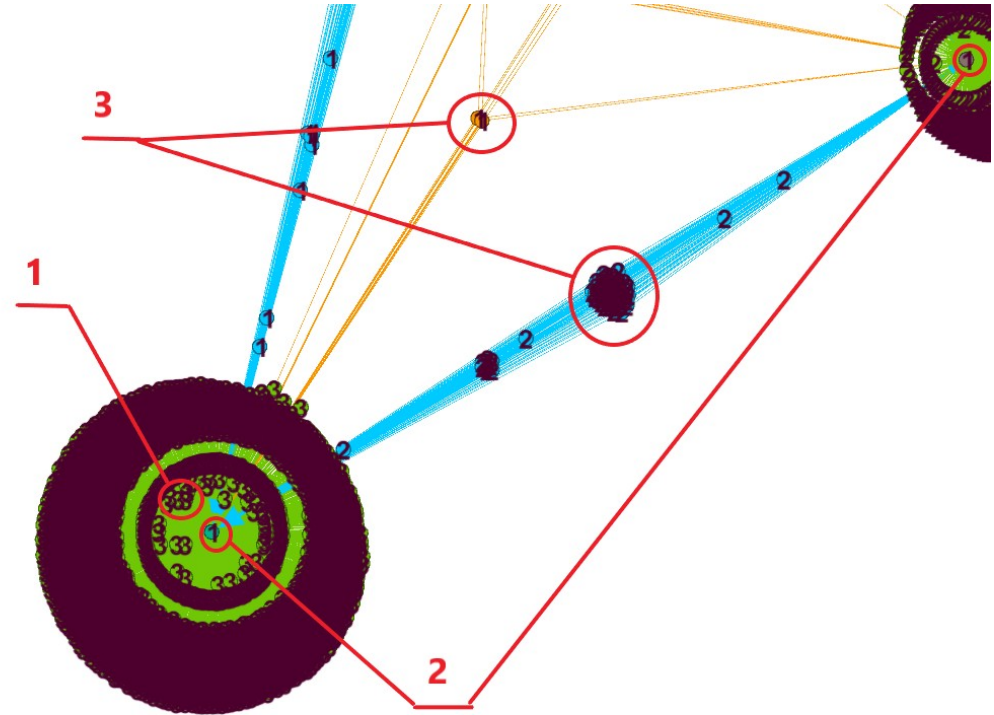
Двудольный ориентированный граф
 $G=(U,V,E)$

1 — вершины $U_1 \dots U_n$ (комментаторы)

2 — вершины $V_1, V_2 \dots V_m$ (топики)

3 — U_i (общие комментаторы)

соединены с V_j (топиками) рёбрами E
(комментариями)



Слайд 4. Источник данных: Youtube Data API

Постоянно используем:

AuthorChannelUrl —

уникальные
комментаторы

VideoId — уникальные
ролики

Иногда используем:

LikeCount — сортировка по
лайкам

PublishedAt — дата
комментария

ModerationStatus —
маркировка спамеров

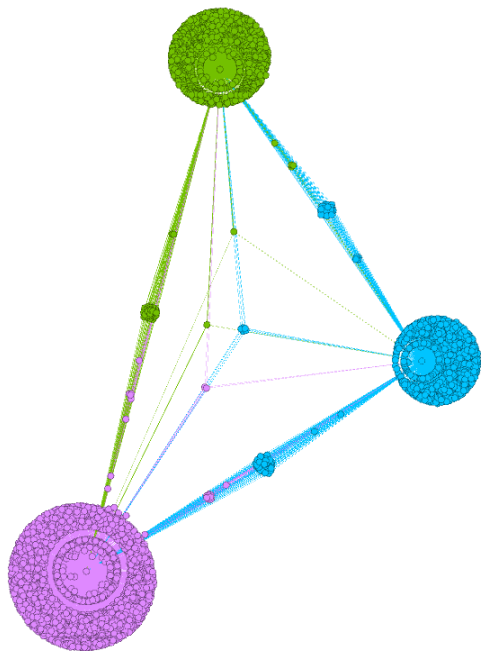
TextOriginal — контент-
анализ содержания
комментариев

ParentId — сортировка по
реплайам

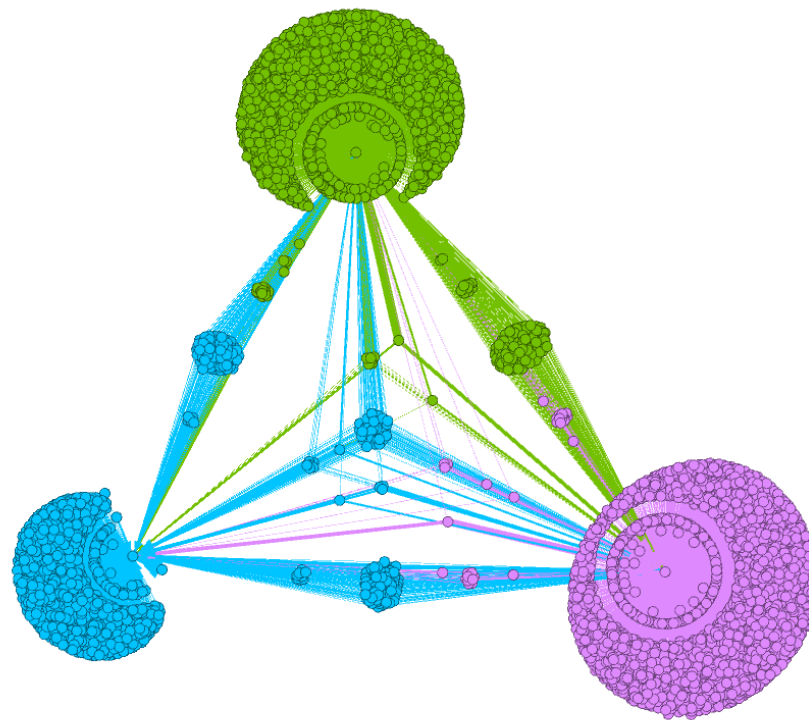
Пока не используем:

Id
AuthorDisplayName
AuthorProfileImageUrl
AuthorChannelId
ChannelId
TextDisplay
CanRate
ViewerRating
UpdatedAt

Слайд 5. Визуализация пересечения комментариев к роликам

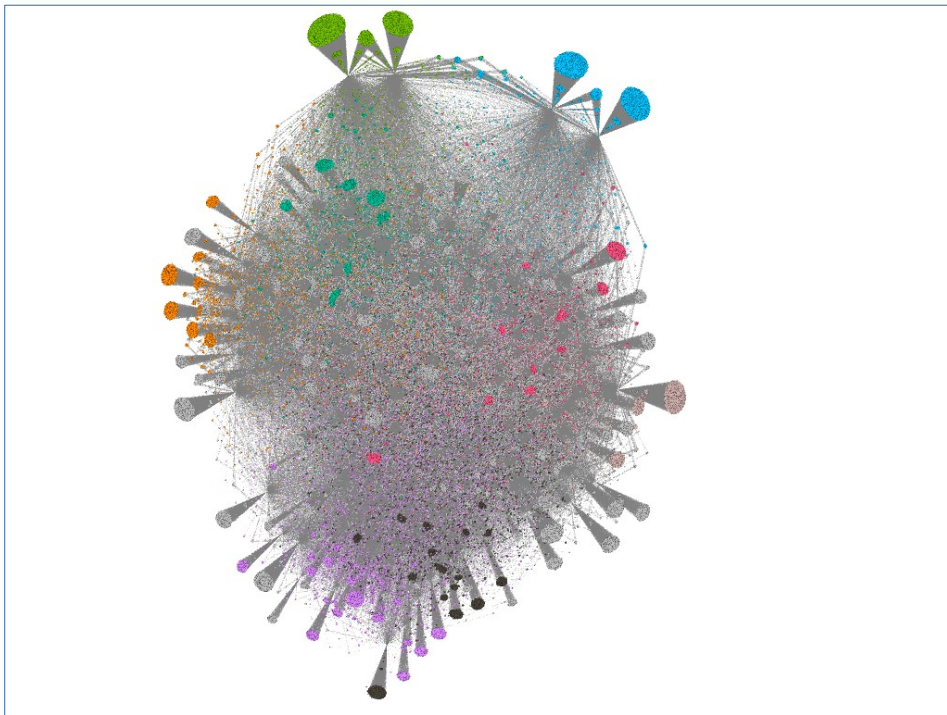


1) Ролики одного канала, не связанные популярным персонажем (10462 вершины, 10692 ребра, средняя степень: 1,022)

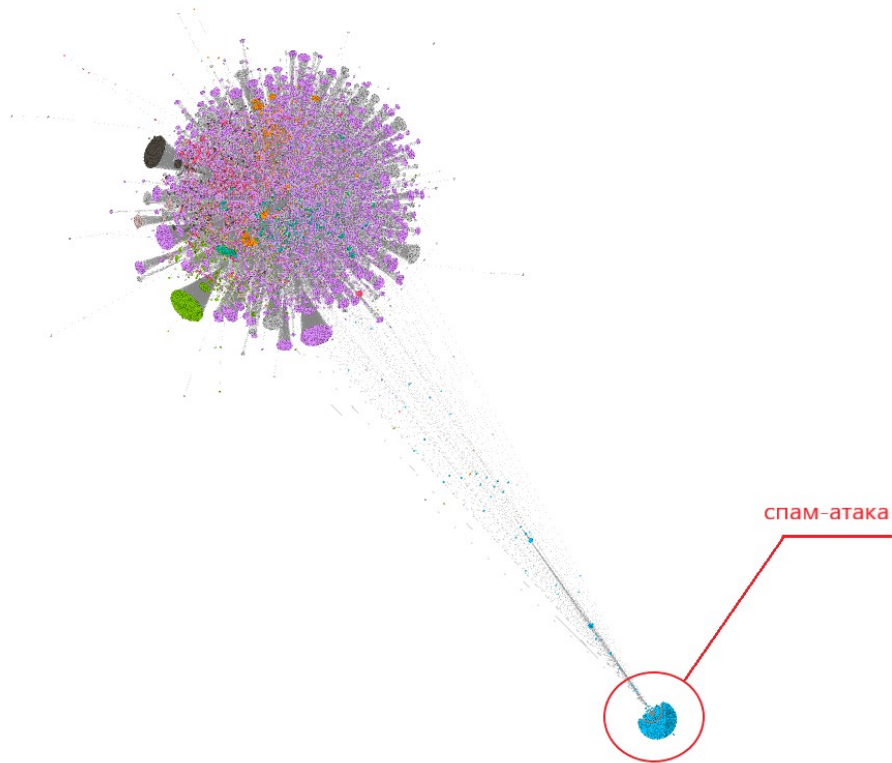


2) Ролики того же канала, связанные популярным персонажем (10108 вершин, 10857 рёбер, средняя степень: 1,074)

Слайд 6. Выделение регионов - новая аудитория



Слайд 7. Выделение регионов - спамеры



Inquire98 3 года назад
"Thank You", thank you very much...

👍 3 🗨️ ОТВЕТИТЬ

Показать 9 ответов ▾



Sinan Kasap 3 года назад
@sinan_alqassab the architect

👍 1 🗨️ ОТВЕТИТЬ



Tyesha7575 4 года назад
I just got myself another 10,000 [#followers](#). I [#use](#):
INSTACREASE.NET

👍 1 🗨️



Skylar Childers 2 года назад
Whee, I'm Really happy that i finally got Twitter Followers, such a big amount also :) try this TWITTFOLLOWERSNET

👍 🗨️ ОТВЕТИТЬ



Maeve Dye 2 года назад
Whee, I'm Really happy that i finally got Twitter Followers, such a big amount also :) try this TWITTFOLLOWERSNET

👍 🗨️ ОТВЕТИТЬ



duminda deshapriya 4 года назад
My business received over 15k new followers this month from using [TwitIncrease.COM](#)

👍 🗨️ ОТВЕТИТЬ



jeremy spangler 2 года назад
I'm on [#twitter](#) @jeremyspangler following back all

👍 🗨️ ОТВЕТИТЬ




roshan dhakal 4 года назад
Yeah, I started my [TwitIncrease.COM](#) lastnight. I have over 1,400 new followers so far.


Слайд 8. Рекомендации на основе выделения родственных групп каналов

убер


ФИЛЬТРЫ




Работа в Убер. Как заработать в такси от \$1000. Бизнес идеи
FranchTV • 75 тыс. просмотров • Год назад
УПАКОВКА ФРАНШИЗ + бесплатные материалы и шаблоны: https://franch.site/?utm_source=tv&utm_campaign=uber Как ...



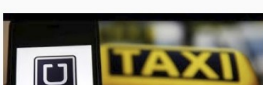
ВОДИТЕЛЬ УБЕР ТАКСИ НА ЛАМБОРГАНИ
AMERIKOZTV • 161 тыс. просмотров • 5 месяцев назад
Оригинал: <https://www.youtube.com/watch?v=KDglY7QyKM> Страница в вк: <https://vk.com/jerrytin>
Скинуть чирик на ...



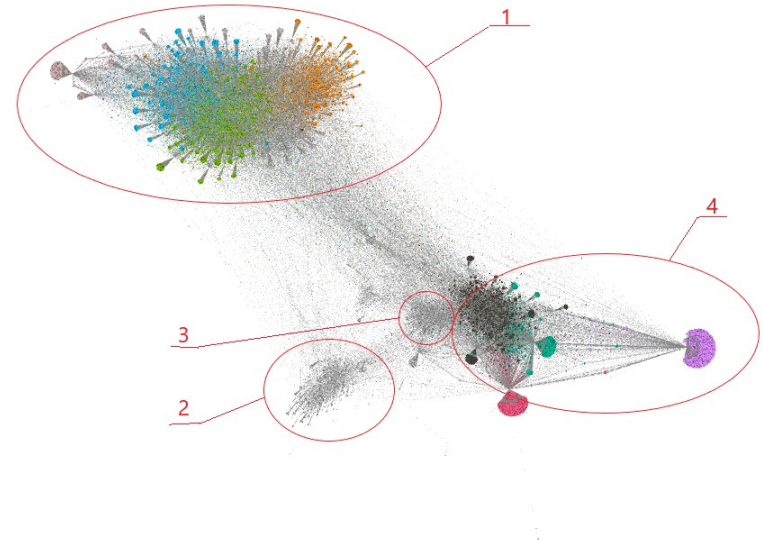
Типичный пассажир Убер: школьник в 60й раз едет из школы домой по промокоду
Такси Провинция • 280 тыс. просмотров • 2 года назад
Наконец-то найден типичный пассажир "Убер-такси"! Это ученик/ученица 7-9 класса, который (ая) каждый день ездит...



Последний заказ по Убер БТ#8
Рафис Камалетдинов • 363 тыс. просмотров • 8 месяцев назад
Телеграмм канал https://t.me/rafis_kamaletdinov Телеграмм чат <https://t.me/rafistaxi> Вк <https://vk.com/id4814476> Инстаграм ...

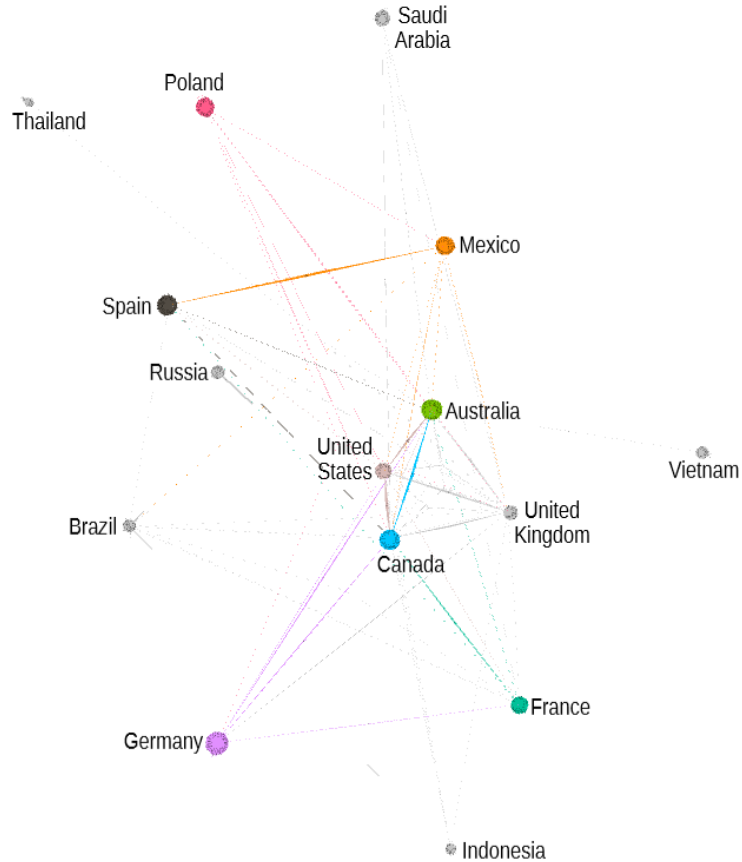


Вся правда про убер в Польше. Новое рабство для заробитчан или золотые горы?
Провинциальный Стиль • 2,4 тыс. просмотров • 1 месяц назад



1 — обзоры автомобилей
2,3,4 — видеоблоги таксистов

Слайд 9. Выделение групп комментаторов: инфлюенсеры



Всего: 150 каналов (группы по 15 странам, 10 каналов в каждой) с подписчиками от 1 до 2 млн., категория «Gaming»

Источник: stats.video

Критерий отбора комментариев:
100+ лайков

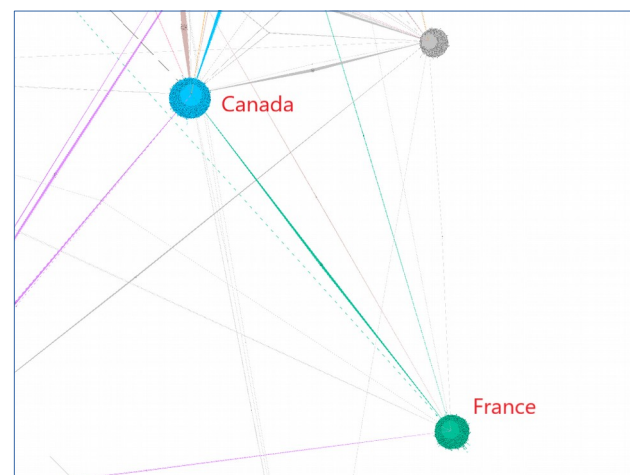
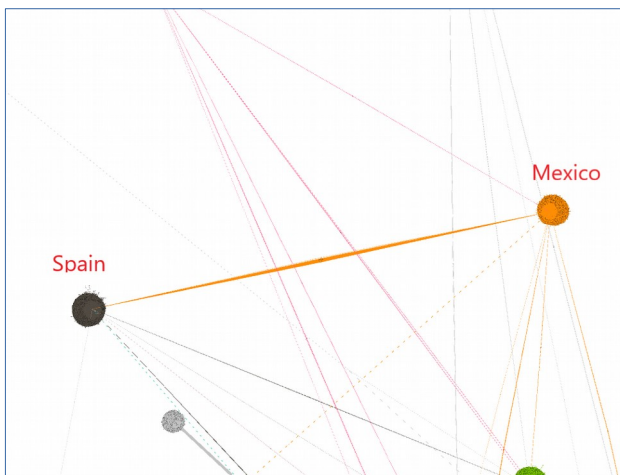
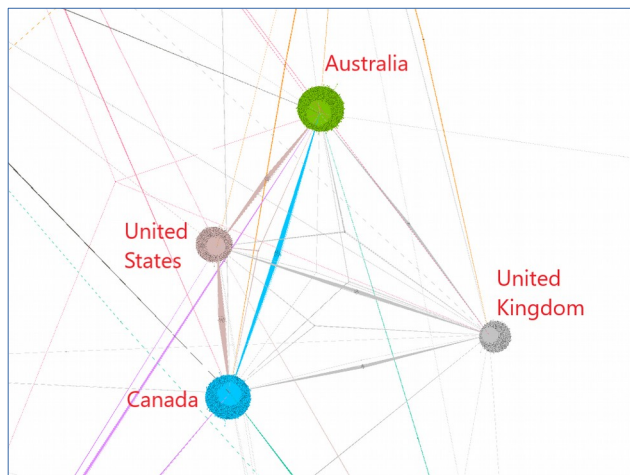
Статистика по графу:

186 946 вершин

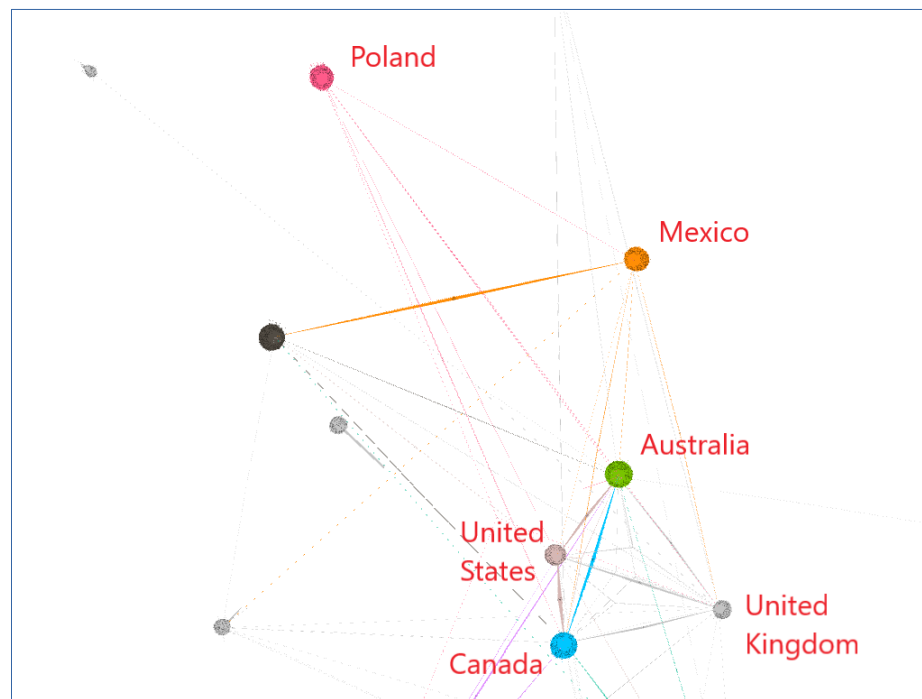
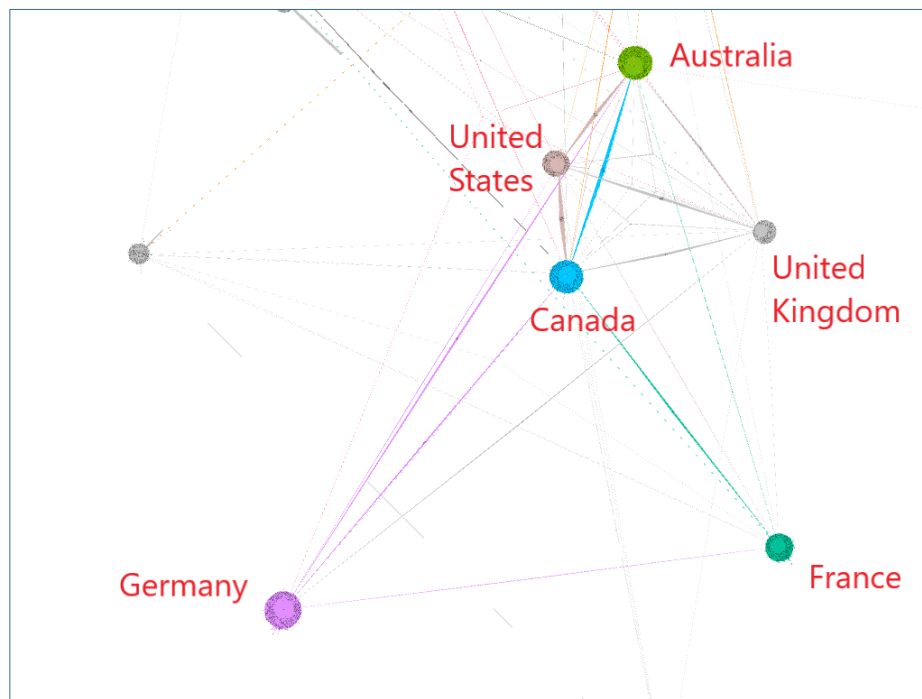
188 648 рёбер

Средняя степень: 1,009

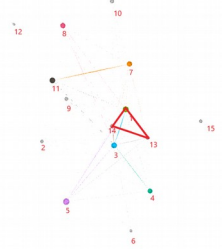
Слайд 10. Моноязыковые паттерны



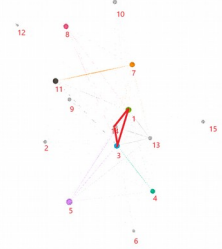
Слайд 11. Смешанные языковые паттерны



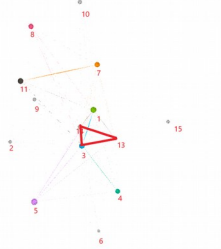
Слайд 12. Комментаторы-инфлюенсеры: поиск самых длинных маршрутов



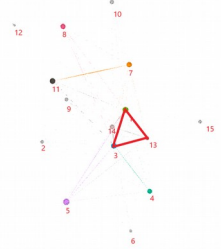
Австралия,
Великобритания, США (7)



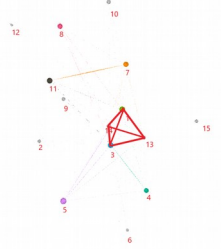
Австралия, Канада,
США (7)



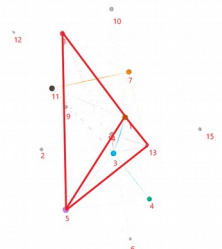
Канада, Великобритания,
США (5)



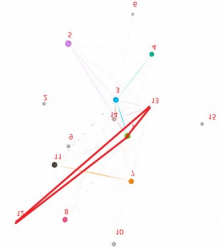
Австралия, Канада,
Великобритания (4)



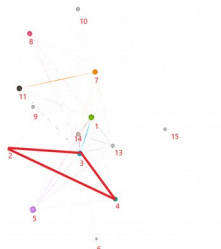
Австралия, Канада,
Великобритания, США (2)



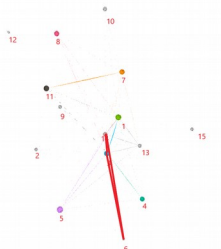
Австралия, Германия,
Польша,
Великобритания (1)



Австралия, Тайланд,
Великобритания (1)



Бразилия, Канада,
Франция (1)



Канада, Индонезия,
США (1)

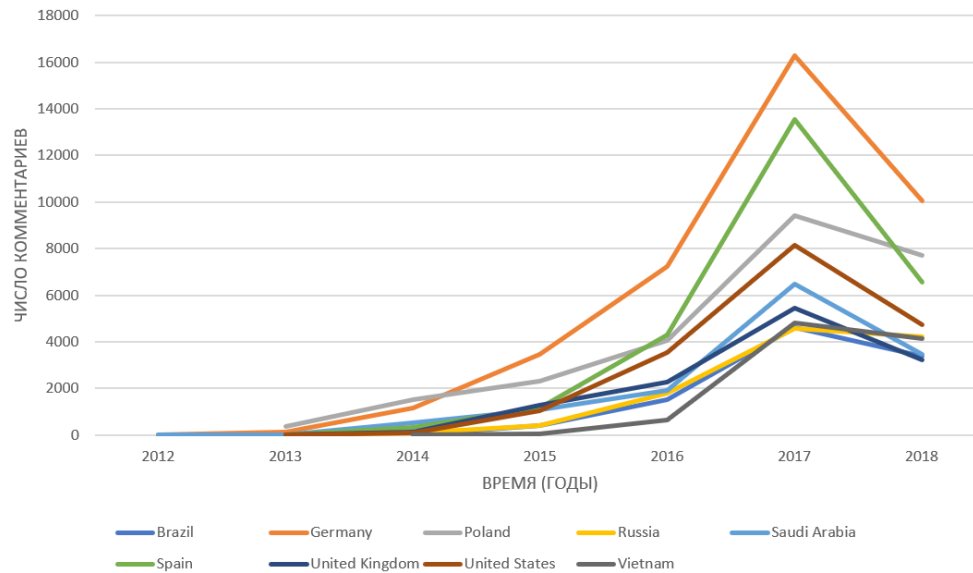
Динамика изменения числа комментариев

The chart displays the growth of online commentary across various nations over a seven-year period. Germany leads with the highest peak at approximately 16,300 comments in 2017. Other notable peaks occur in Mexico (approx. 13,500), Poland (approx. 9,400), and the United Kingdom (approx. 8,800). By 2018, several countries like Russia, Saudi Arabia, and Thailand have stabilized their comment counts, while others like Mexico and Poland show a downward trend.

Страна	2012	2013	2014	2015	2016	2017	2018
Australia	0	0	0	0	0	0	0
Brazil	0	0	0	0	0	0	0
Canada	0	0	0	0	0	0	0
France	0	0	0	0	0	0	0
Germany	0	0	0	0	0	0	0
Indonesia	0	0	0	0	0	0	0
Mexico	0	0	0	0	0	0	0
Poland	0	0	0	0	0	0	0
Russia	0	0	0	0	0	0	0
Saudi Arabia	0	0	0	0	0	0	0
Spain	0	0	0	0	0	0	0
Thailand	0	0	0	0	0	0	0
United Kingdom	0	0	0	0	0	0	0
United States	0	0	0	0	0	0	0
Vietnam	0	0	0	0	0	0	0

Слайд 14. Комментаторы-инфлюенсеры: спад

Динамика изменения числа комментариев - спад в 2018 г.



twitch.tv

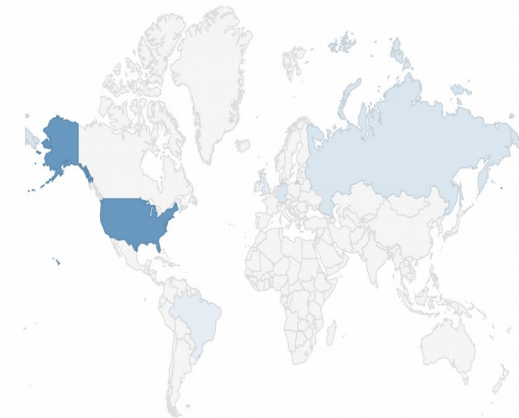
+ COMPARE

Feb 2019

DOWNLOAD

Traffic by countries ①

On desktop



United States	20.28%	▼ 8.58%
Russia	6.49%	▼ 14.01%
Germany	6.23%	▼ 7.90%
Brazil	5.12%	▼ 5.39%
United Kingdom	4.98%	▼ 5.21%

See 248 more countries

Input: определение ассоциаций каналов по связности и родственным сообществам

Output: рекомендации «похожих» каналов с сильным пересечением комментаторов

Input: оценка качества ютуб-канала на основе анализа аудитории

Output: сервис ранжирования ютуб-каналов

Input: оптимизация использования каналов на основе анализа **пересечения** аудиторий

Output: сервис размещения контента с максимальным охватом аудитории

Вы можете связаться со мной по адресу aa.kostin@phystech.edu

Спасибо за внимание!