

**DataFest<sup>6</sup>**, 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

### 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>)

### Алгоритм:

1) Привязать **ID** комментаторов к **объектам** (топикам)

2) Повторяющиеся в **объектах** комментаторы дают пересечение

	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

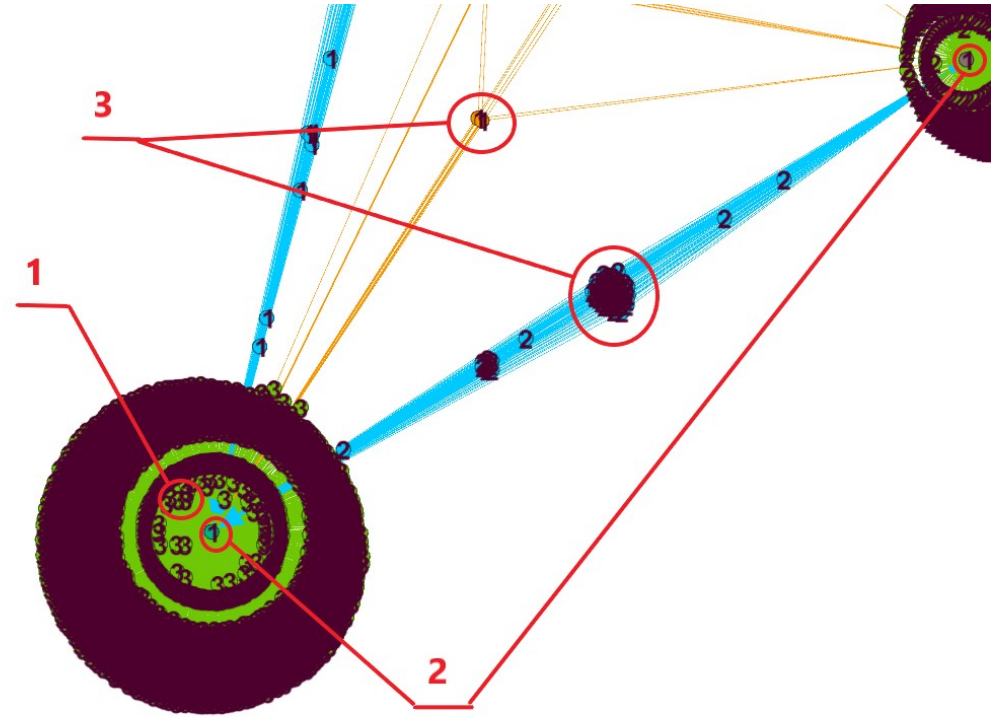
Двудольный ориентированный граф  
 $G=(U,V,E)$

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. Источники данных

Источник данных: Youtube Data API

Используемые категории:

**AuthorChannelUrl** — уникальные комментаторы

**VideoId** — уникальные ролики

**LikeCount** — сортировка по лайкам

**PublishedAt** — дата комментария

YouTube > Data API

Поиск

НАЧАЛО РУКОВОДСТВА СПРАВОЧНИКИ ПРИМЕРЫ ПОДДЕРЖКА

Overview

» Activities

» Captions

» ChannelBanners

» Channels

» ChannelSections

» Comments

» **Comments**

» list

» insert

» update

» markAsSpam

» setModerationStatus

» delete

» CommentThreads

» GuideCategories

» i18nLanguages

» i18nRegions

» PlaylistItems

» Playlists

» Search

» Subscriptions

» Thumbnails

» VideoAbuseReportReasons

» VideoCategories

» Videos

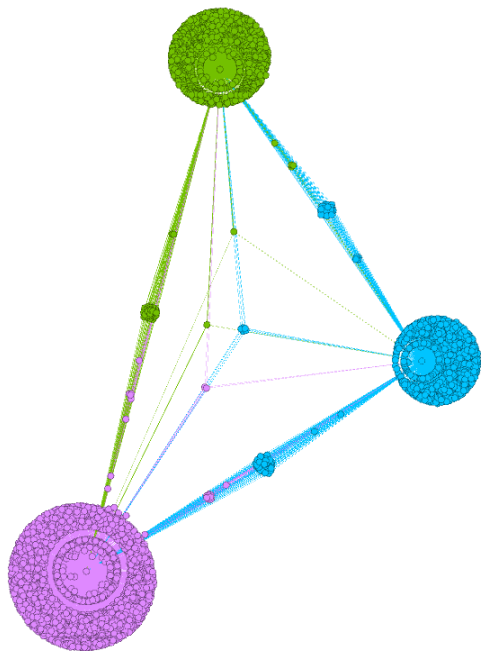
» Watermarks

» Standard Query Parameters

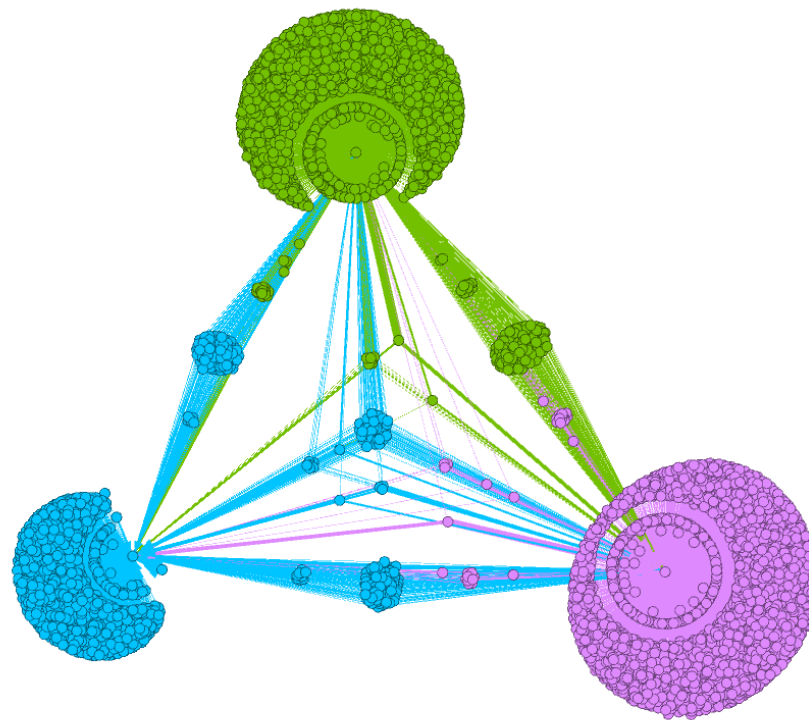
Properties

kind	string	Identifies the API resource's type. The value will be youtube#comment.
etag	etag	The Etag of this resource.
id	string	The ID that YouTube uses to uniquely identify the comment.
snippet	object	The snippet object contains basic details about the comment.
snippet.authorDisplayName	string	The display name of the user who posted the comment.
snippet.authorProfileImageUrl	string	The URL for the avatar of the user who posted the comment.
snippet.authorChannelUrl	string	The URL of the comment author's YouTube channel, if available.
snippet.authorChannelId	object	This object encapsulates information about the comment author's YouTube channel, if available.
snippet.authorChannelId.value	string	The ID of the comment author's YouTube channel, if available.
snippet.channelId	string	The ID of the YouTube channel associated with the comment. <ul style="list-style-type: none"><li>If the comment is a video comment, then this property identifies the video's channel, and the snippet.videoId property identifies the video.</li><li>If the comment is a channel comment, then this property identifies the channel that the comment is about.</li></ul>

## Слайд 5. Визуализация пересечения комментариев к роликам

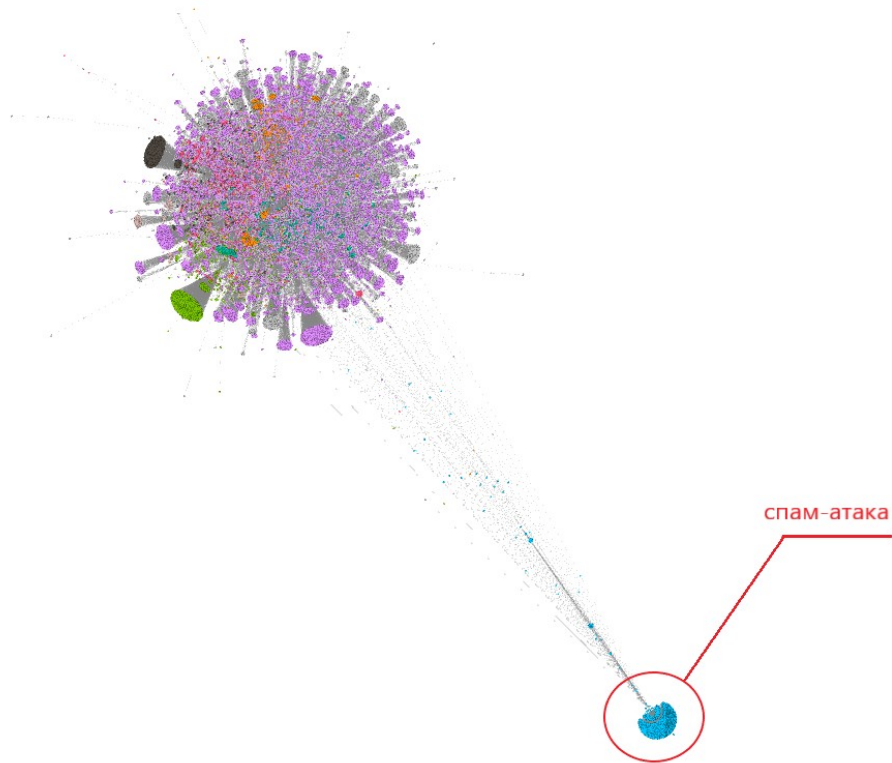


1) Ролики одного канала, не связанные популярным персонажем (10462 вершины, 10692 ребра, средняя степень: 1,022)



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

## Слайд 6. Выделение регионов (паттернов)



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 7. Рекомендации на основе выделения родственных групп каналов

убер

ФИЛЬТРЫ

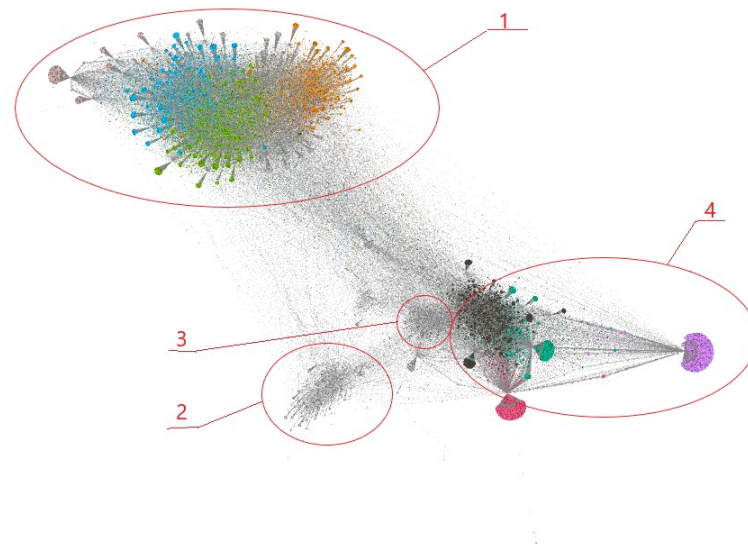
**Работа в Убер. Как заработать в такси от \$1000. Бизнес идеи**  
FranchTV • 75 тыс. просмотров • Год назад  
УПАКОВКА ФРАНШИЗ + бесплатные материалы и шаблоны: [https://franch.site/?utm\\_source=tv&utm\\_campaign=uber](https://franch.site/?utm_source=tv&utm_campaign=uber) Как ...

**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](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

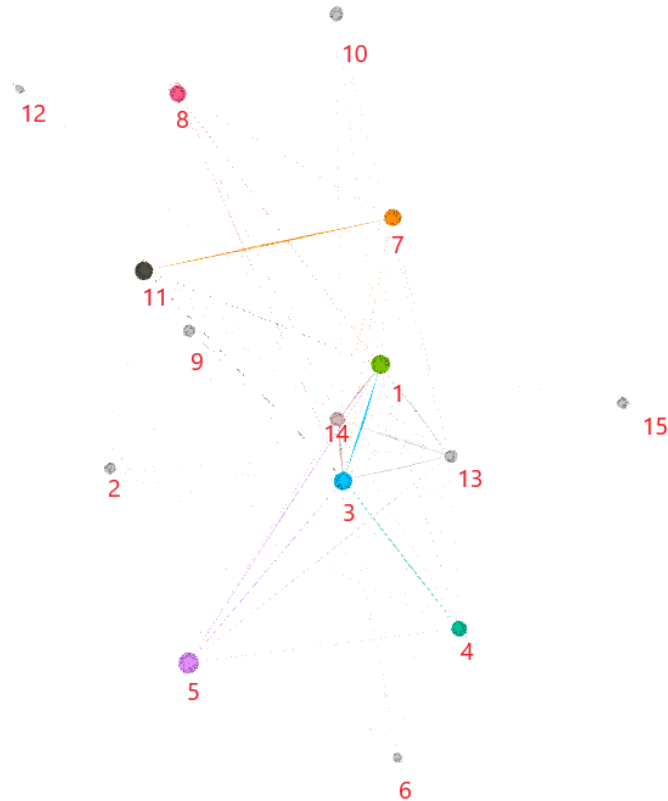


1 — обзоры автомобилей  
2,3,4 — видеоблоги таксистов



## Слайд 8. Выделение групп комментаторов: инфлюенсеры

Australia	1
Brazil	2
Canada	3
France	4
Germany	5
Indonesia	6
Mexico	7
Poland	8
Russia	9
Saudi Arabia	10
Spain	11
Thailand	12
United Kingdom	13
United States	14
Vietnam	15

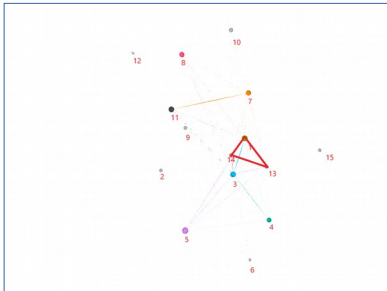


Всего: 150 каналов  
(15 языковых групп по  
10 каналов в каждой)  
с подписчиками от 1  
до 2 млн., категория  
«Gaming»

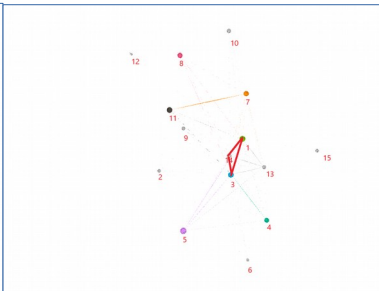
Критерий отбора  
комментариев: 100+  
лайков

Источник: [stats.video](https://stats.video)

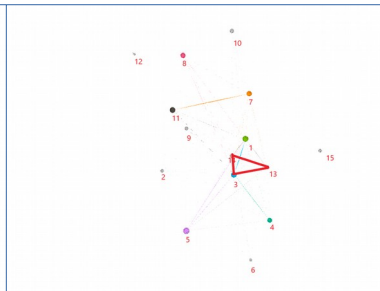
## Слайд 9. Комментаторы-инфлюенсеры: поиск маршрутов



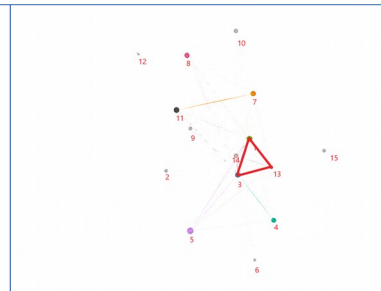
1, 13, 14 (7)



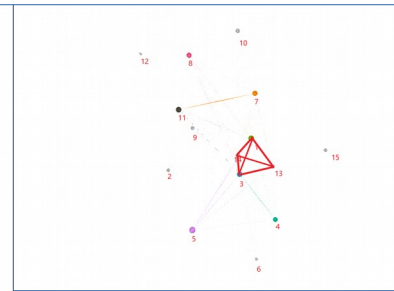
1, 3, 14 (7)



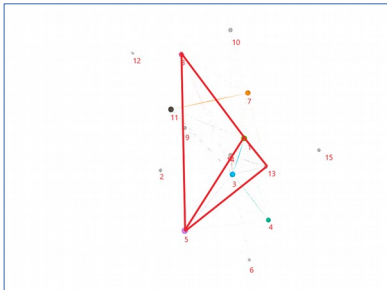
3, 13, 14 (5)



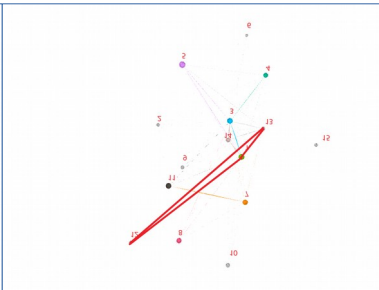
1, 3, 13 (4)



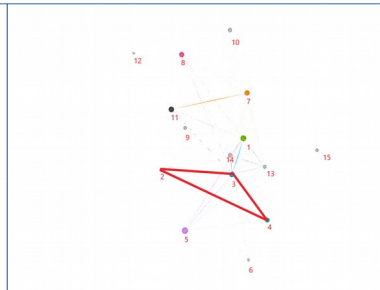
1, 3, 13, 14 (2)



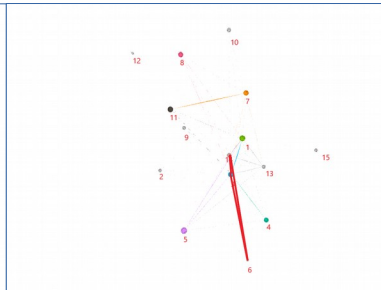
1, 5, 8, 13 (1)



1, 12, 13 (1)



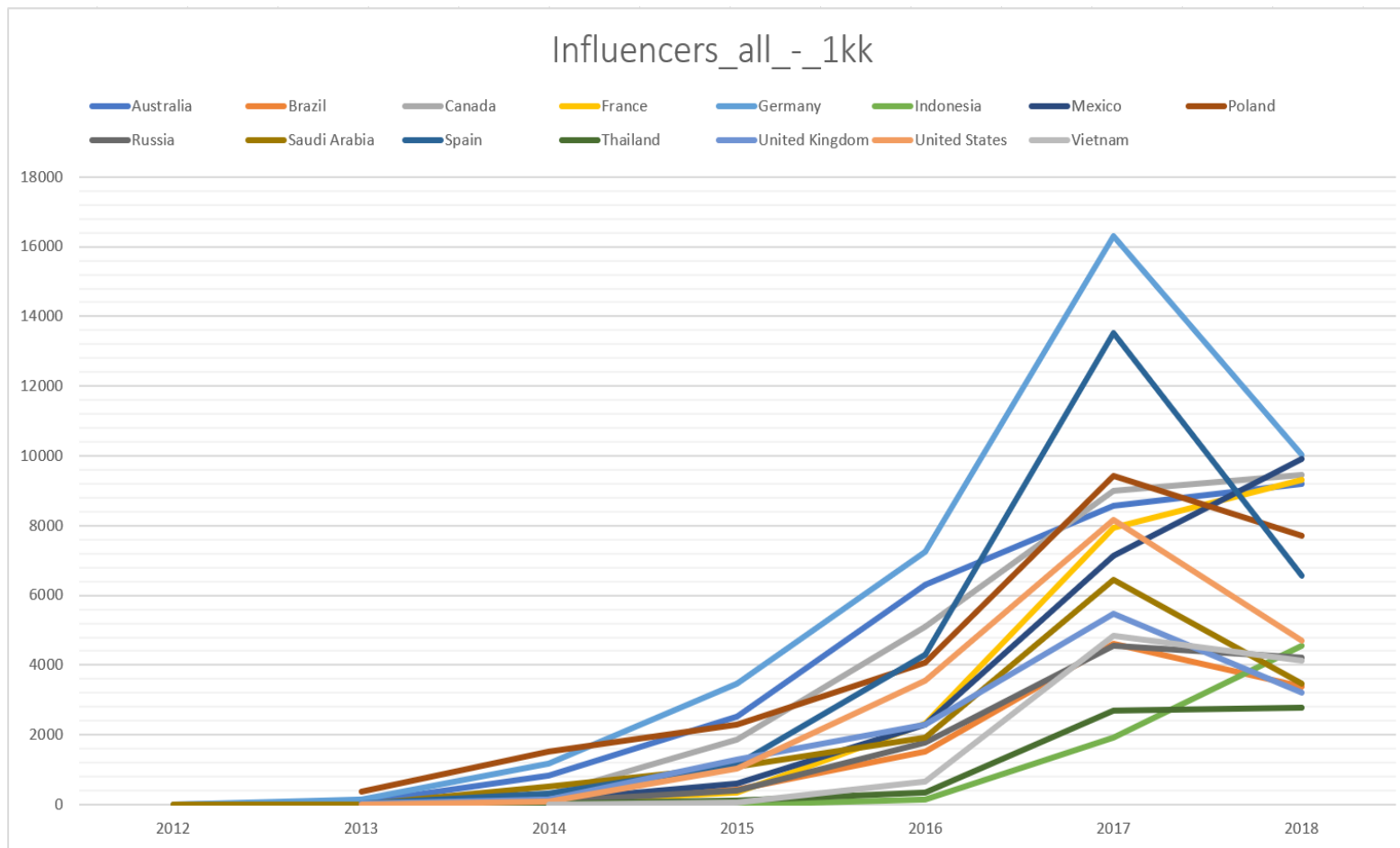
2, 3, 4 (1)



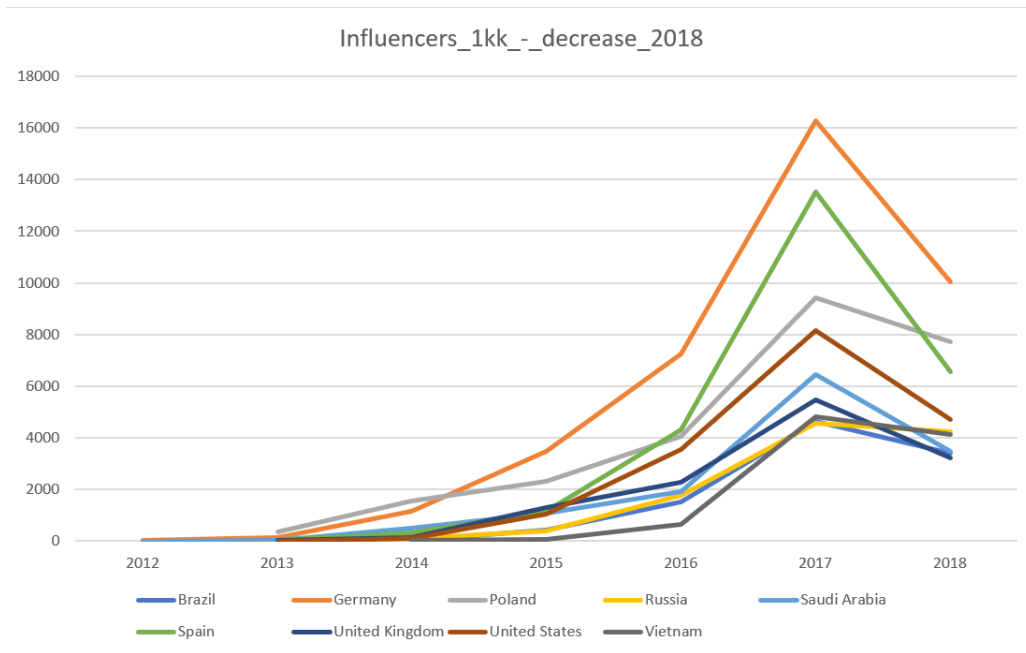
3, 6, 14 (1)

Australia	1
Brazil	2
Canada	3
France	4
Germany	5
Indonesia	6
Mexico	7
Poland	8
Russia	9
Saudi Arabia	10
Spain	11
Thailand	12
United Kingdom	13
United States	14
Vietnam	15

## Слайд 10. Комментаторы-инфлюенсеры: динамика изменения



## Слайд 11. Комментаторы-инфлюенсеры: спад



## Слайд 12. Области применения

- В определении ассоциаций каналов по **связности** и **родственным сообществам**
- В оптимизации использования каналов в маркетинговых целях на основе анализа **пересечения** аудиторий
- В оценке качества канала на основе анализа аудитории
- В исследовании **нечётких сообществ** (**fuzzy communities**)

Вы можете связаться со мной по адресу [aa.kostin@phystech.edu](mailto:aa.kostin@phystech.edu)

Спасибо за внимание!