

Майнор

«Обработка и анализ данных» v3.0

http://electives.hse.ru/2016/minor_data_spb/
http://electives.hse.ru/minor_data_spb/

А.В. Сироткин, А.В. Суворова, И.Л. Мусабиров

Data Science

Наука о том, что делать с данными

- Где брать?
- Что брать?
- Как анализировать и делать выводы?
- Как проверять выводы?

- Кому все это надо?

Мы будем учиться:

- Собирать данные:
 - Работать с базами
 - Грабить интернет
- Анализировать данные
 - Ну тут очень много чего...
- Рассказывать о данных и наших выводах
 - Рисовать красивые картинки и графики
 - Объяснять, почему это круто

Курсы майнора

- Программирование для анализа данных и воспроизводимые исследования
- Анализ данных и технологии работы с данными
- Интеллектуальный анализ данных и основы машинного обучения
- Приложения и практика анализа данных

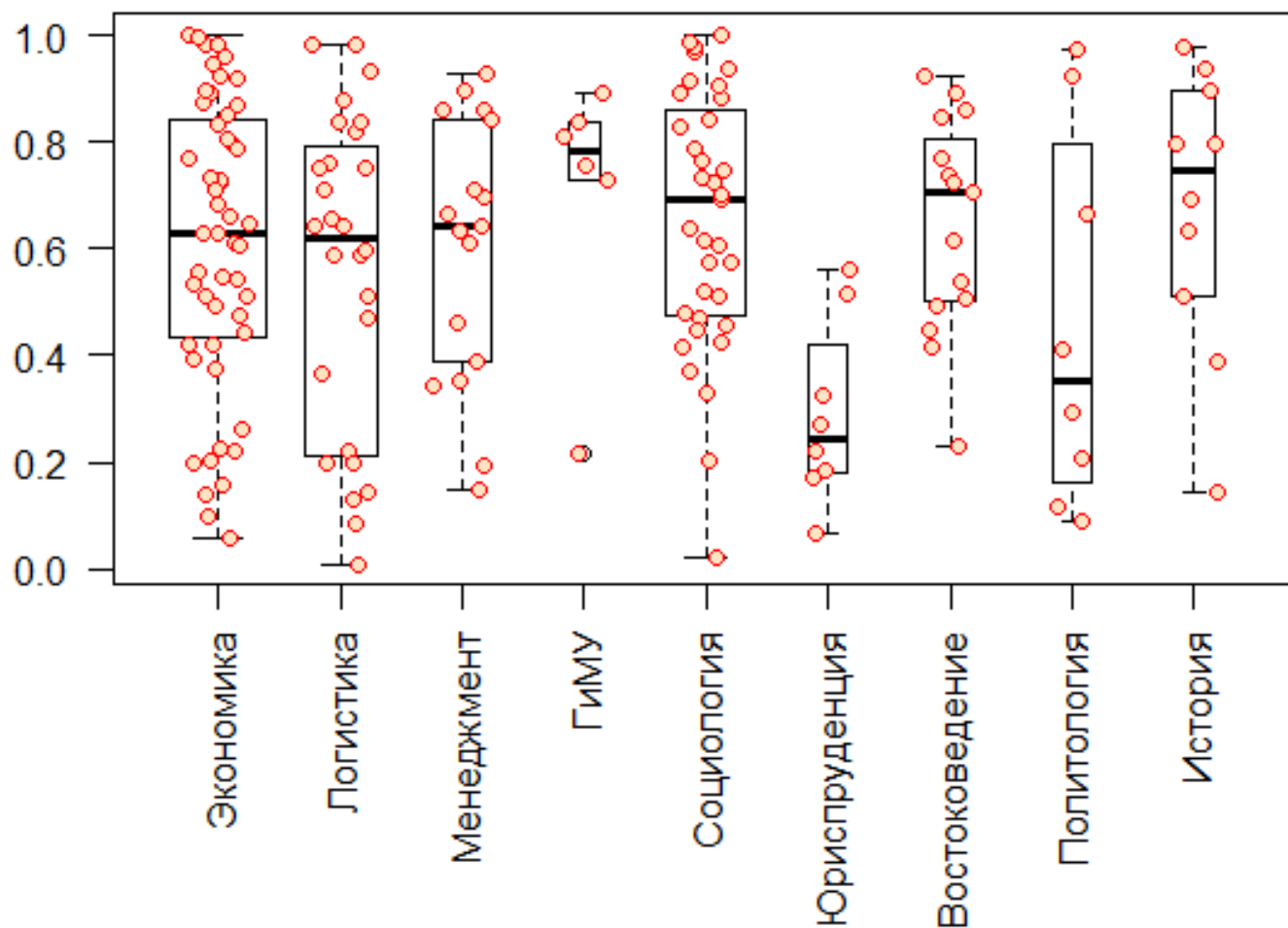
Проектная работа

- Мы будем делать проекты
- Уже делались или делаются:
 - Музыкальные предпочтения студентов Вышки и студентов ПТУ
 - Анализ фанфиков
 - Анализ поведения игроков в онлайн играх
 - Что обсуждают в кавказских группах Вконтакте
 - Предсказание победителя в Dota 2
 - и много чего еще...
- Можно придумать свой проект и сделать его.
- В последнем семестре никто не уйдет без проекта

Требования к студентам

- Хотеть изучать анализ данных
- Не испытывать непреодолимого страха перед компьютером
- Быть готовым поработать
- Немножко знать английский

Кто приходил к нам



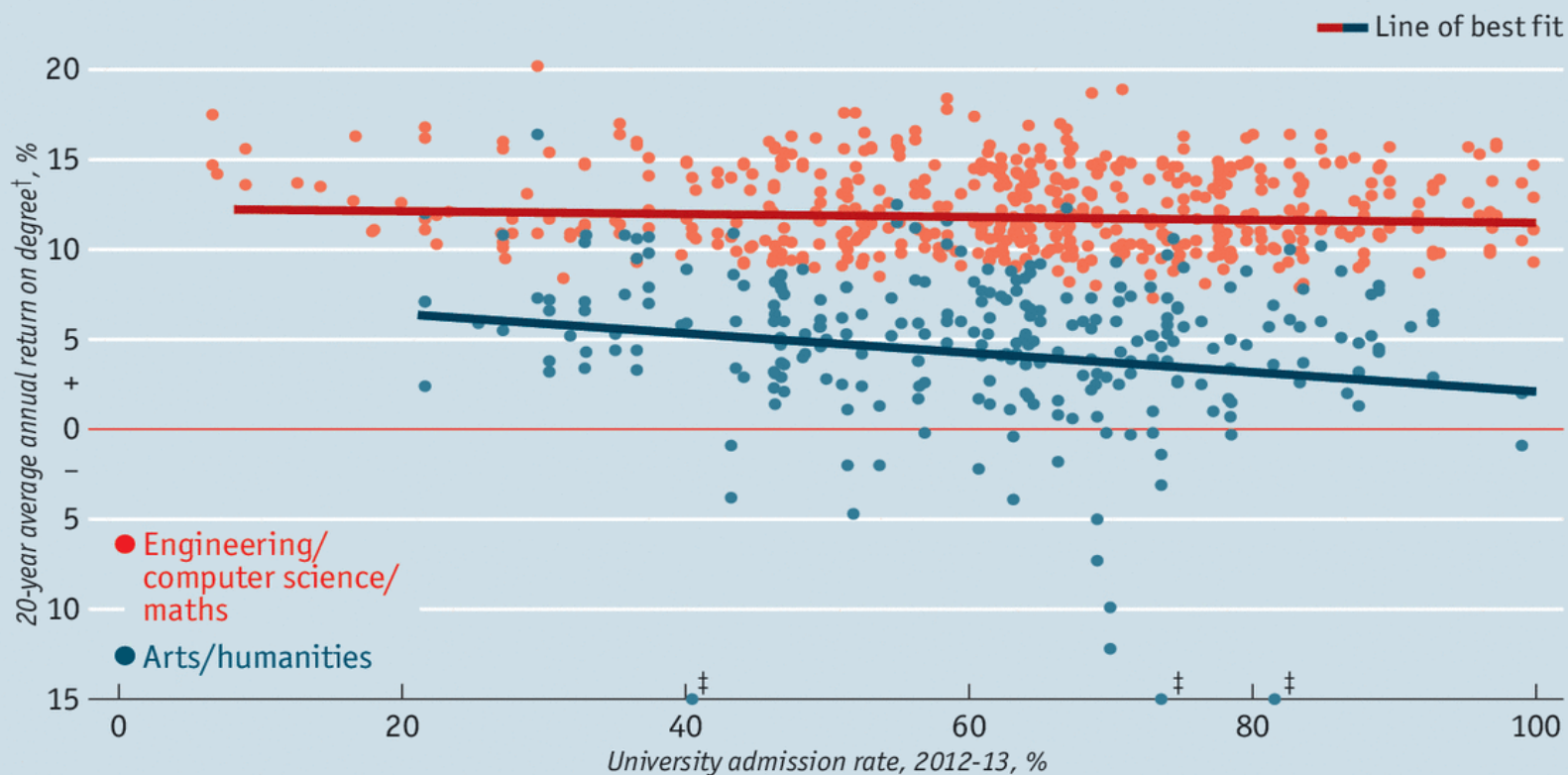
А теперь немного веселых историй про анализ данных

Картинки из разных исследований с комментариями

Точные науки против гуманитарных

It's what you learn, not where

American universities*, selectivity and returns



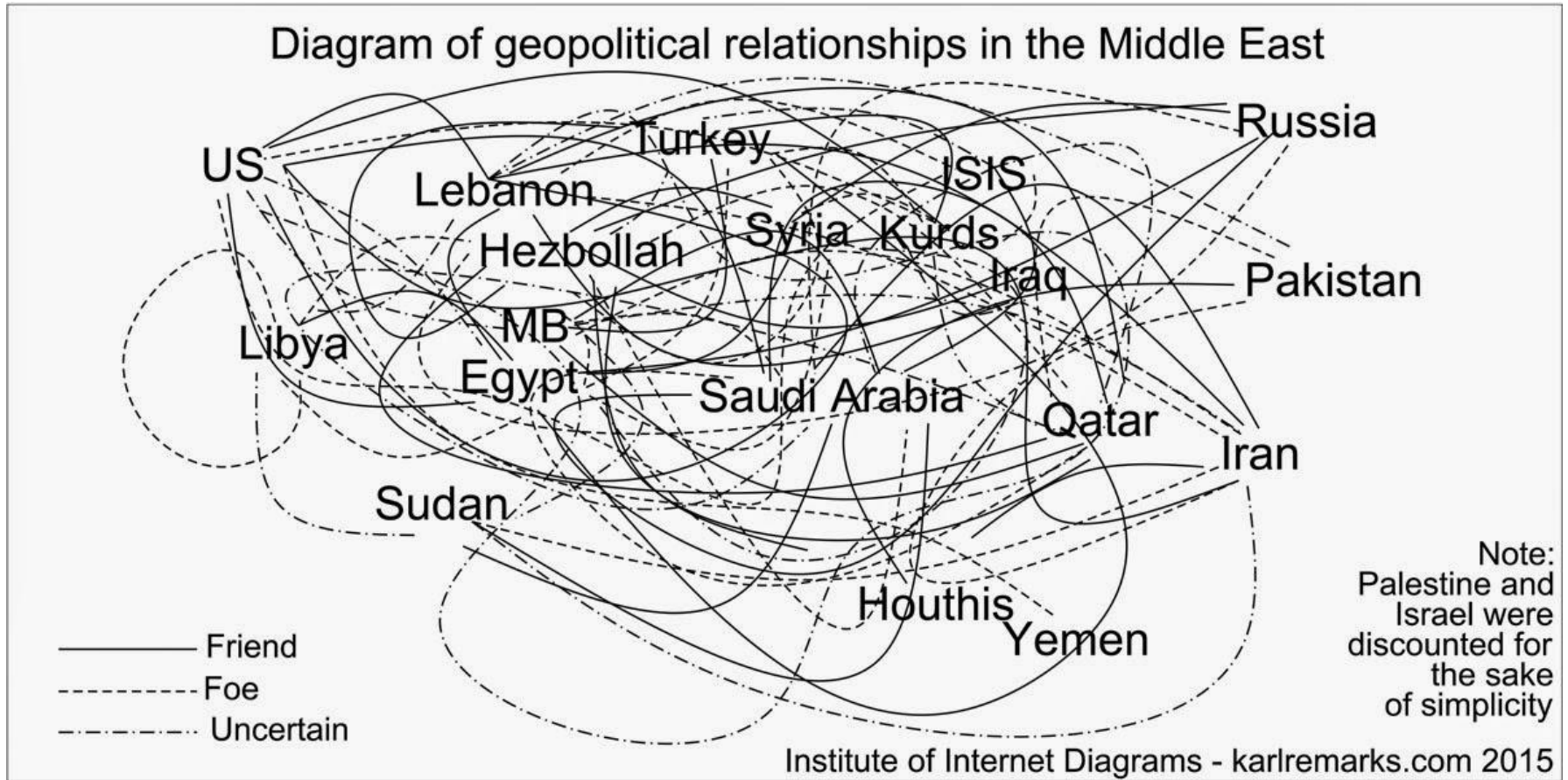
Sources: PayScale; National Centre for Education Statistics

*Sample of 240 institutions

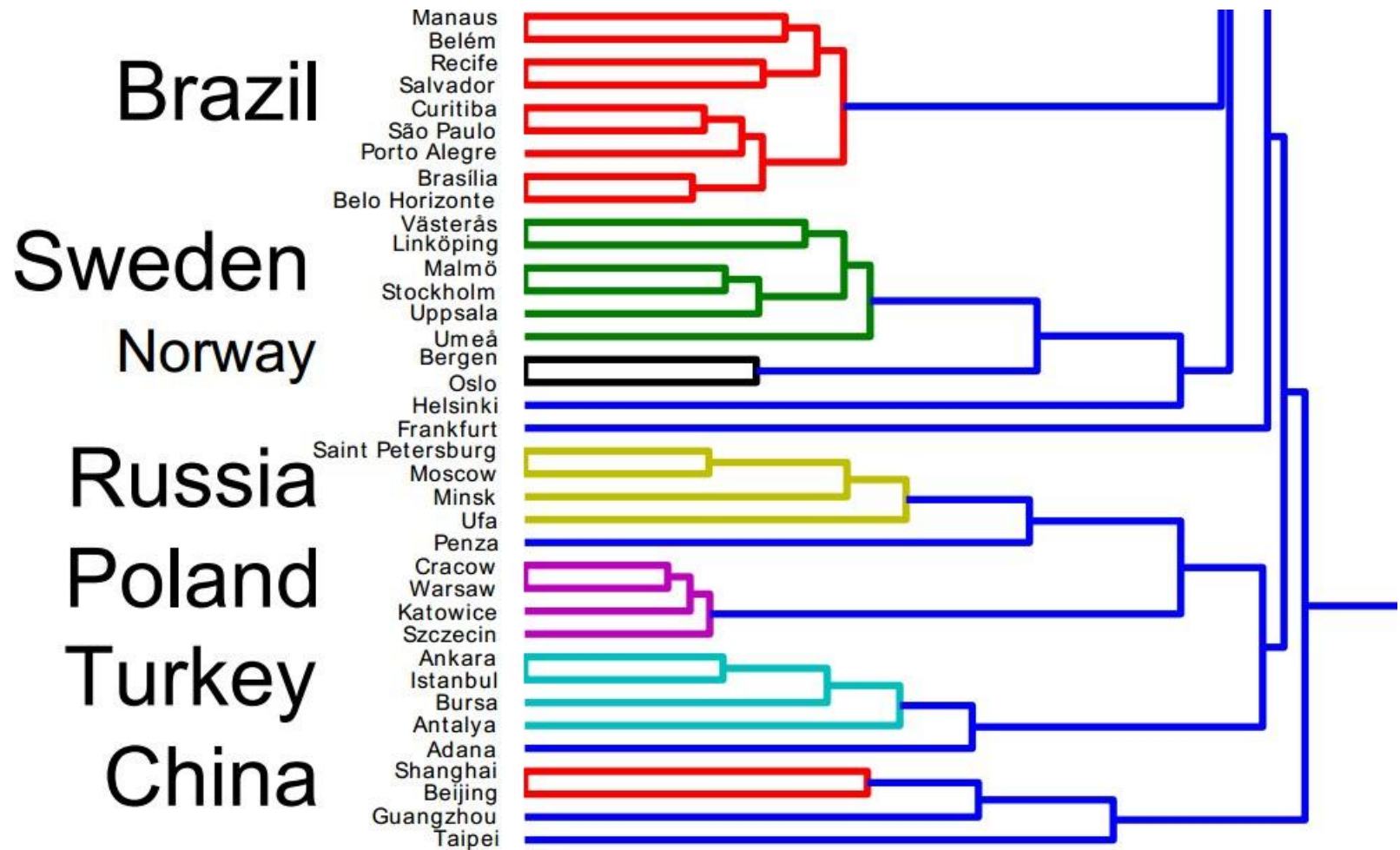
†Based on 2012-13

†Returns of minus 15% or worse

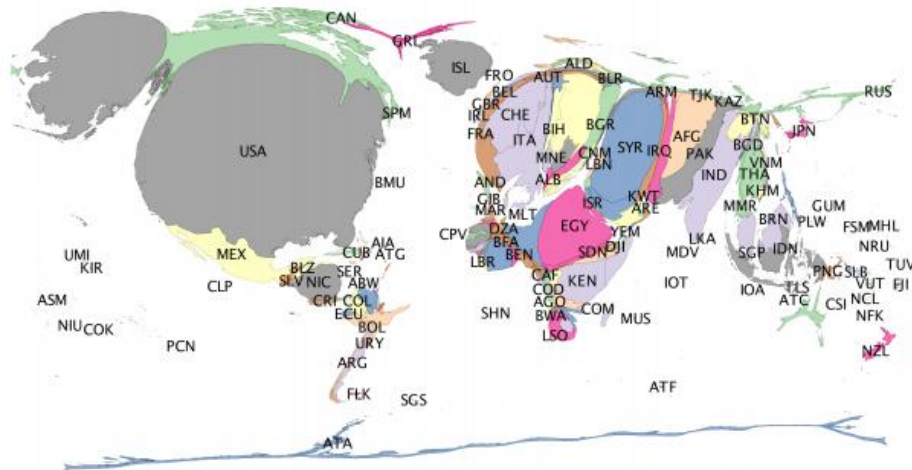
Все сложно?...



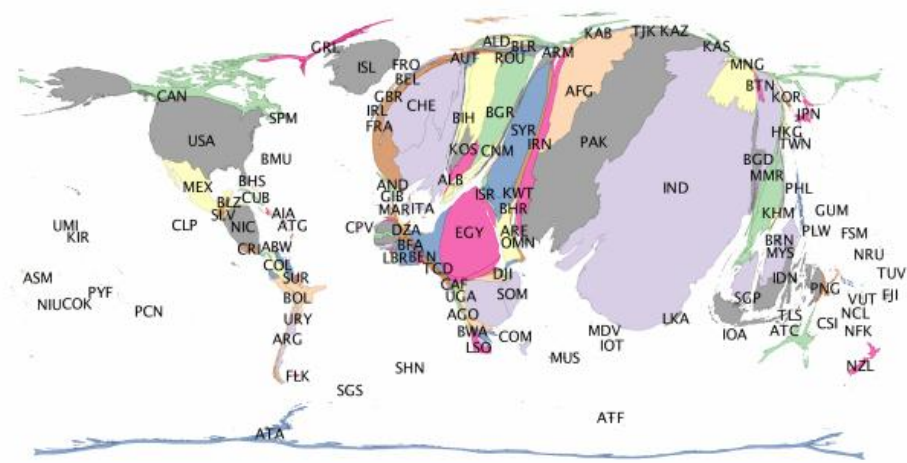
Схожесть музыкальных вкусов



Как видят мир в разных регионах



(a) North America

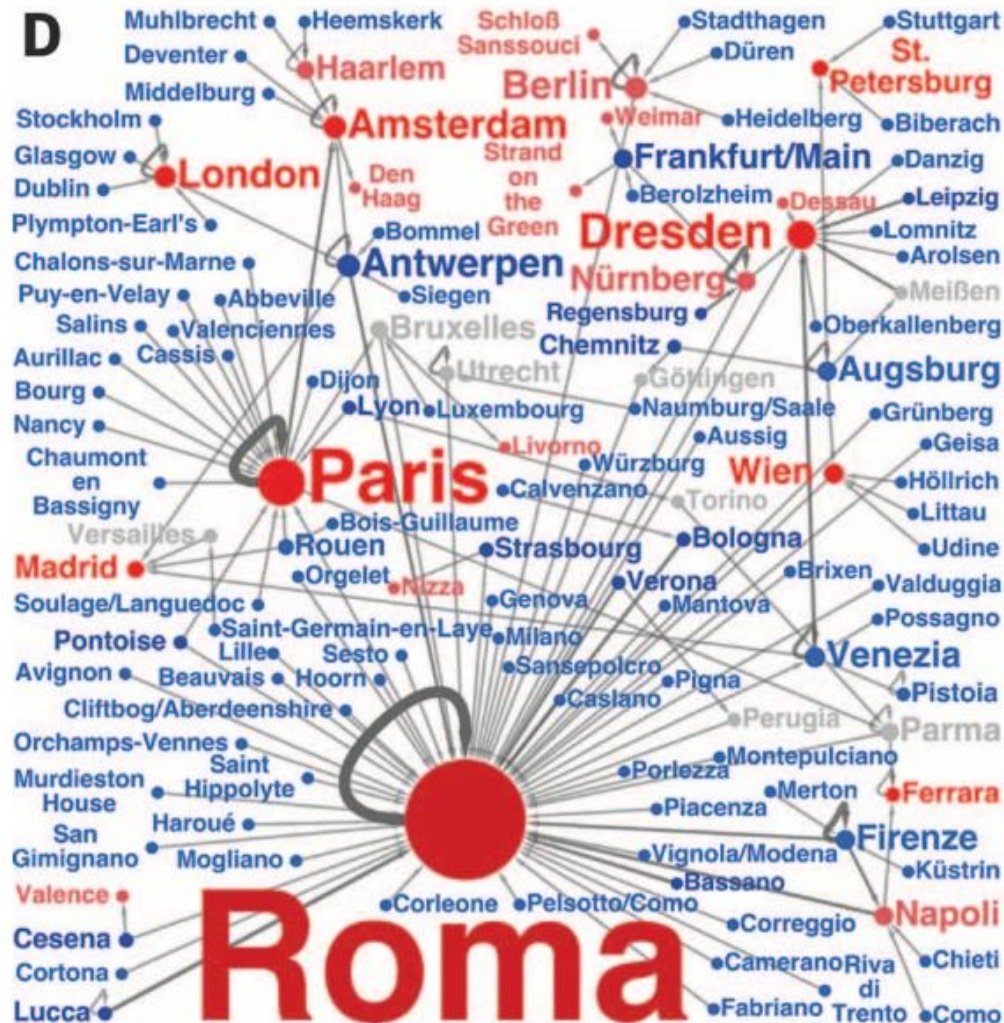


(b) South Asia

Fig. 1. News geography seen by each region

From A First Look at Global News Coverage of Disasters By Using the GDELT Dataset,
H. Kwak, J. An, SocInfo'14

Где умирали знаменитости в 18 веке



**From “A network
framework of cultural
history”**

Science 1 August 2014:
Vol. 345 no. 6196
pp. 558-562
DOI: 10.1126/science.1240064

Кто с кем общается

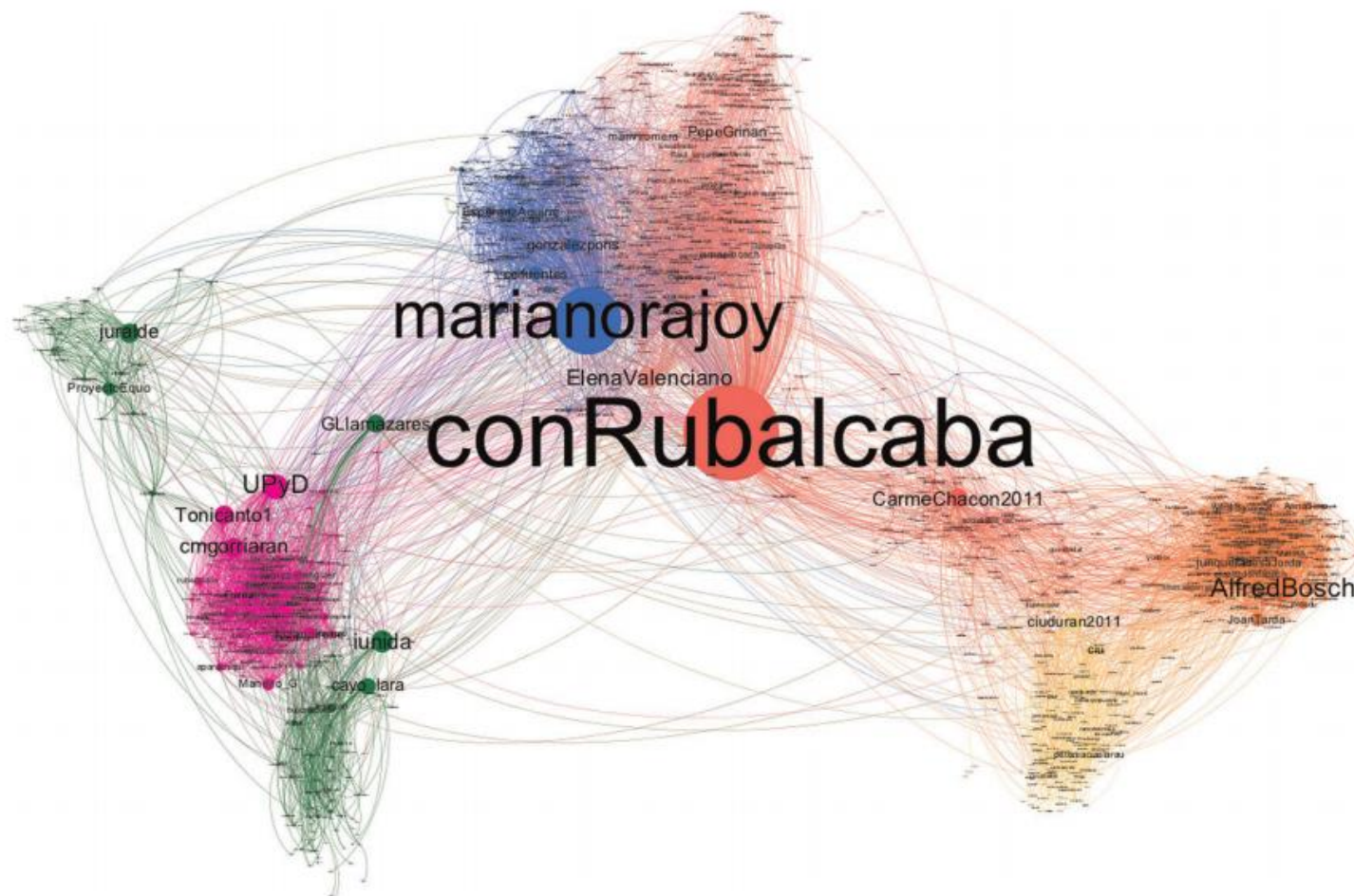
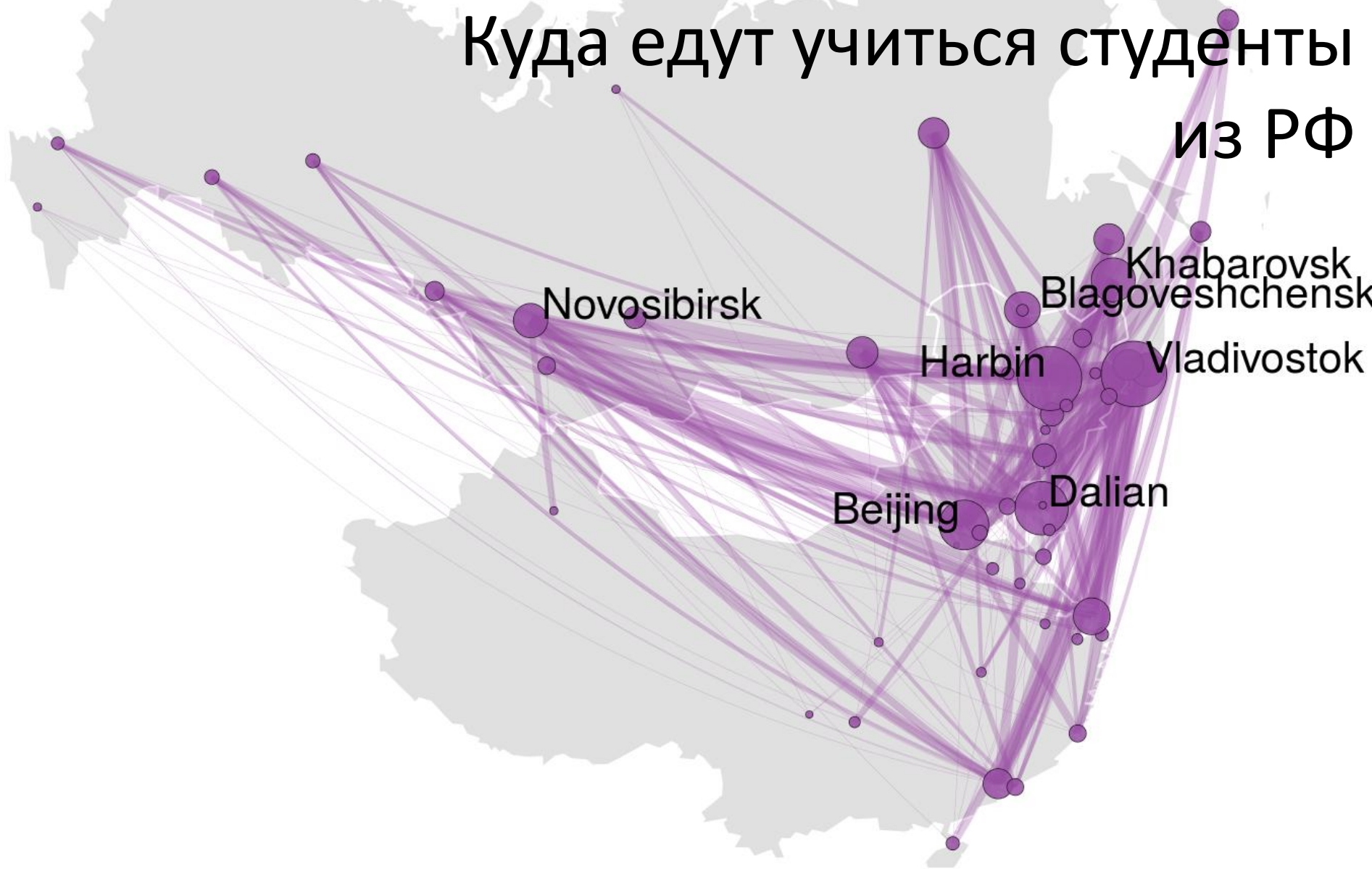


Figure 7. Reply Graph.

Notes: The size of each node corresponds to its in-degree and the color represents the cluster it belongs to. PSOE, red; PP, blue; IU-EQUO, green; UPyD, pink; CiU, yellow; ERC, orange.



Куда едут учиться студенты из РФ



Спасибо за внимание!

- Свежая информация про майнор:
http://electives.hse.ru/2016/minor_data_spb/
http://electives.hse.ru/minor_data_spb/
- НУГ «Машинное обучение и социальный компьютеринг»
<http://mlsoc.io>
- Истории социологов про курсы и проекты:
<http://nosoc.io>

Ждем вас!