

Цифровые следы и психологическое благополучие учащихся

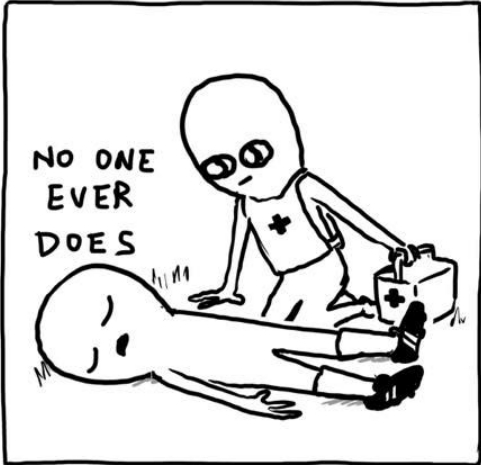
**Иван Смирнов
Юлия Дементьева
София Докука
ЦИСД**



Мотивация







“человеку в аспирантуре
должно быть плохо”

директор одного центра

**“современное поколение
ранимое и не готовое к
стрессу”**

проректор одного вуза о студентах

**“пускай готовится к
взрослой жизни”**

школьный учитель

Не правы морально и прагматические

“stress, depression or anxiety accounted for 40% of all work-related ill health cases and 49% of all working days lost due to ill health” // 2016/17, UK

Не правы морально и прагматические

“stress, depression or anxiety accounted for 40% of all work-related ill health cases and 49% of all working days lost due to ill health” // 2016/17, UK

“Болят голова/живот”

Не правы морально и прагматические

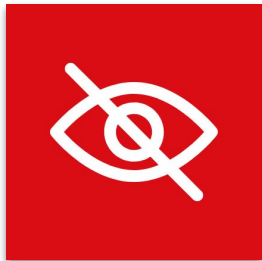
“stress, depression or anxiety accounted for 40% of all work-related ill health cases and 49% of all working days lost due to ill health” // 2016/17, UK

“Болят голова/живот”

Падение продуктивности

Цифровые следы

Ничего не знаем



Скрытые переменные

Измеряем то, что можем, а не то, что важно

паттерны сна



Скрытые переменные

Измеряем то, что можем, а не то, что важно

Паттерны сна



Гетерогенные и сложные паттерны

Предполагаем гомогенность, используем линейные модели

использование социальных медиа подростками и удовлетворенность жизнью



Скрытые переменные

Измеряем то, что можем, а не то, что важно

Паттерны сна



Гетерогенные и сложные паттерны

Предполагаем гомогенность, используем линейные модели

использование социальных медиа подростками и удовлетворенность жизнью



Неспособность делать предсказания

Недостаточно выявить некоторые закономерности

obesity system map



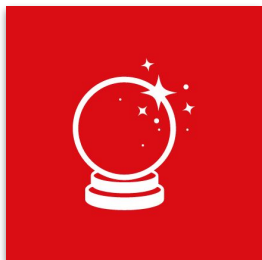
Скрытые переменные

Измеряем то, что можем, а не то, что важно
паттерны сна



Гетерогенные и сложные паттерны

Предполагаем гомогенность, используем линейные модели
использование социальных медиа подростками и удовлетворенность жизнью



Неспособность делать предсказания

Недостаточно выявить некоторые закономерности
obesity system map

Решение



Скрытые переменные

Новые источники данных



Скрытые переменные

НОВЫЕ ИСТОЧНИКИ ДАННЫХ



Гетерогенные и сложные паттерны

Big data и machine learning



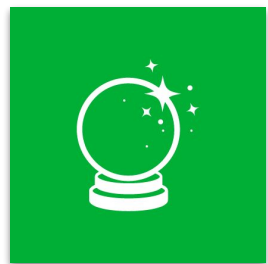
Скрытые переменные

Новые источники данных



Гетерогенные и сложные паттерны

Big data и machine learning



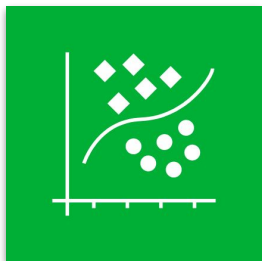
Неспособность делать предсказания

Моделирование



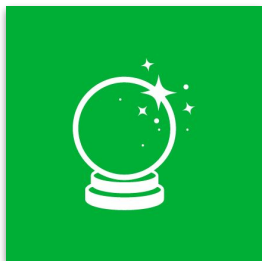
Скрытые переменные

Новые источники данных



Гетерогенные и сложные паттерны

Big data и machine learning



Неспособность делать предсказания

Моделирование

Благотворительные

Депрессия



Достаточно много знаем (контекст)



Достаточно много знаем (контекст)

Доказано связана с важными
жизненными исходами

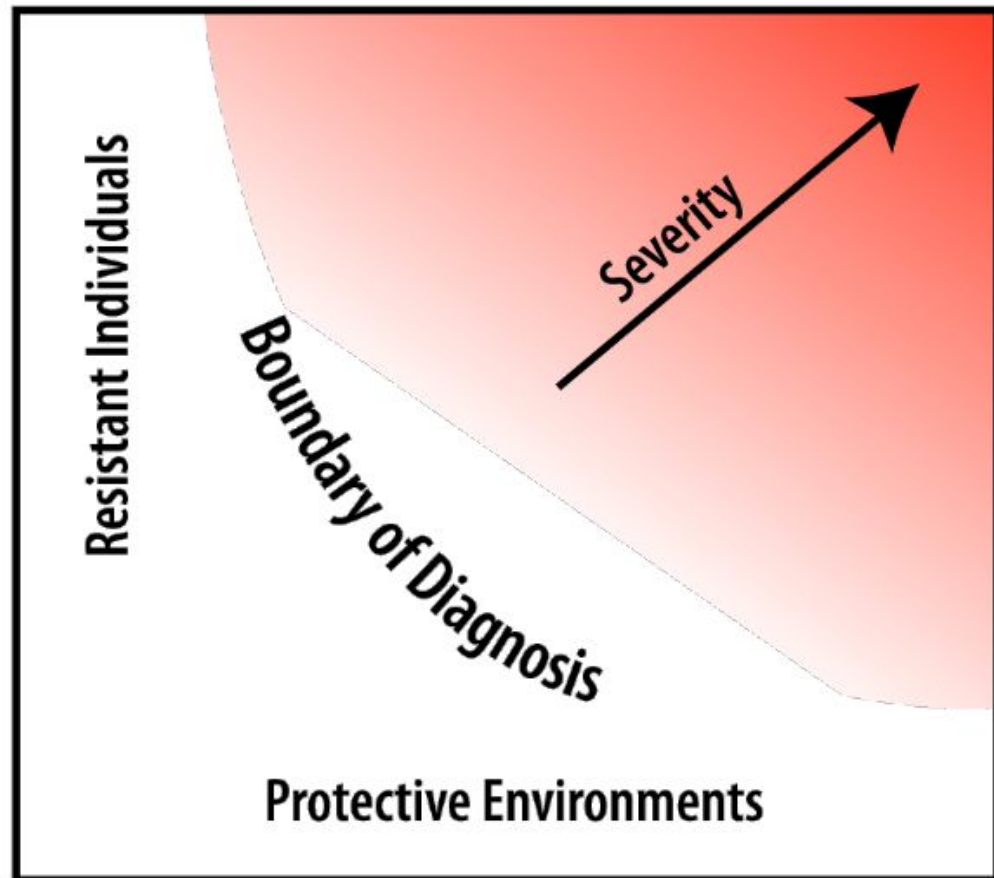


Достаточно много знаем (контекст)

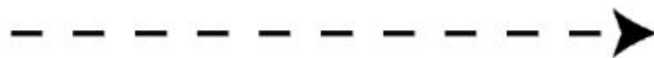
Доказано связана с важными
жизненными исходами

Социальный контекст

Low stress... High stress



Few genes... Many genes



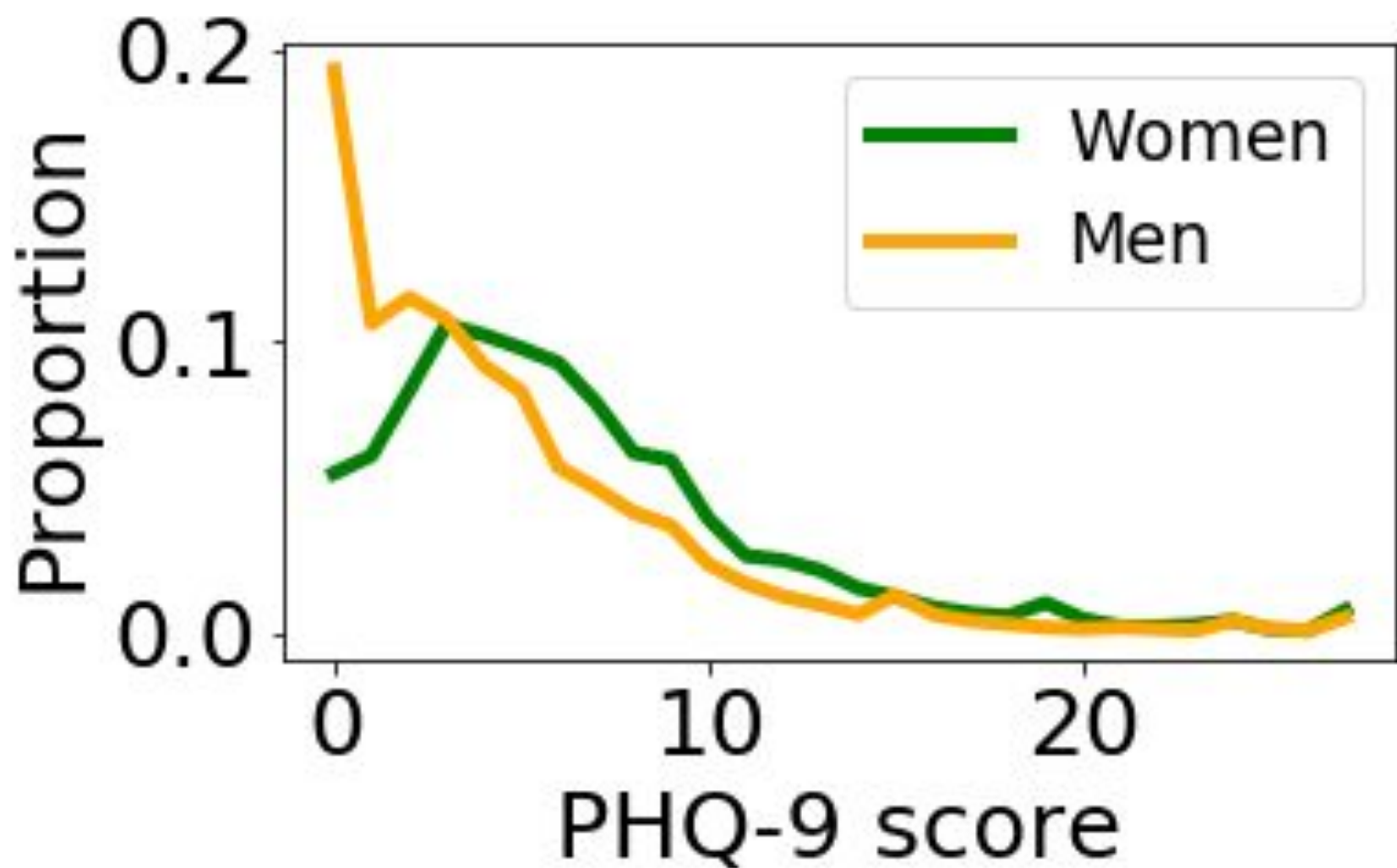
Траектории в образовании и карьере

репрезентативная панель с 2011

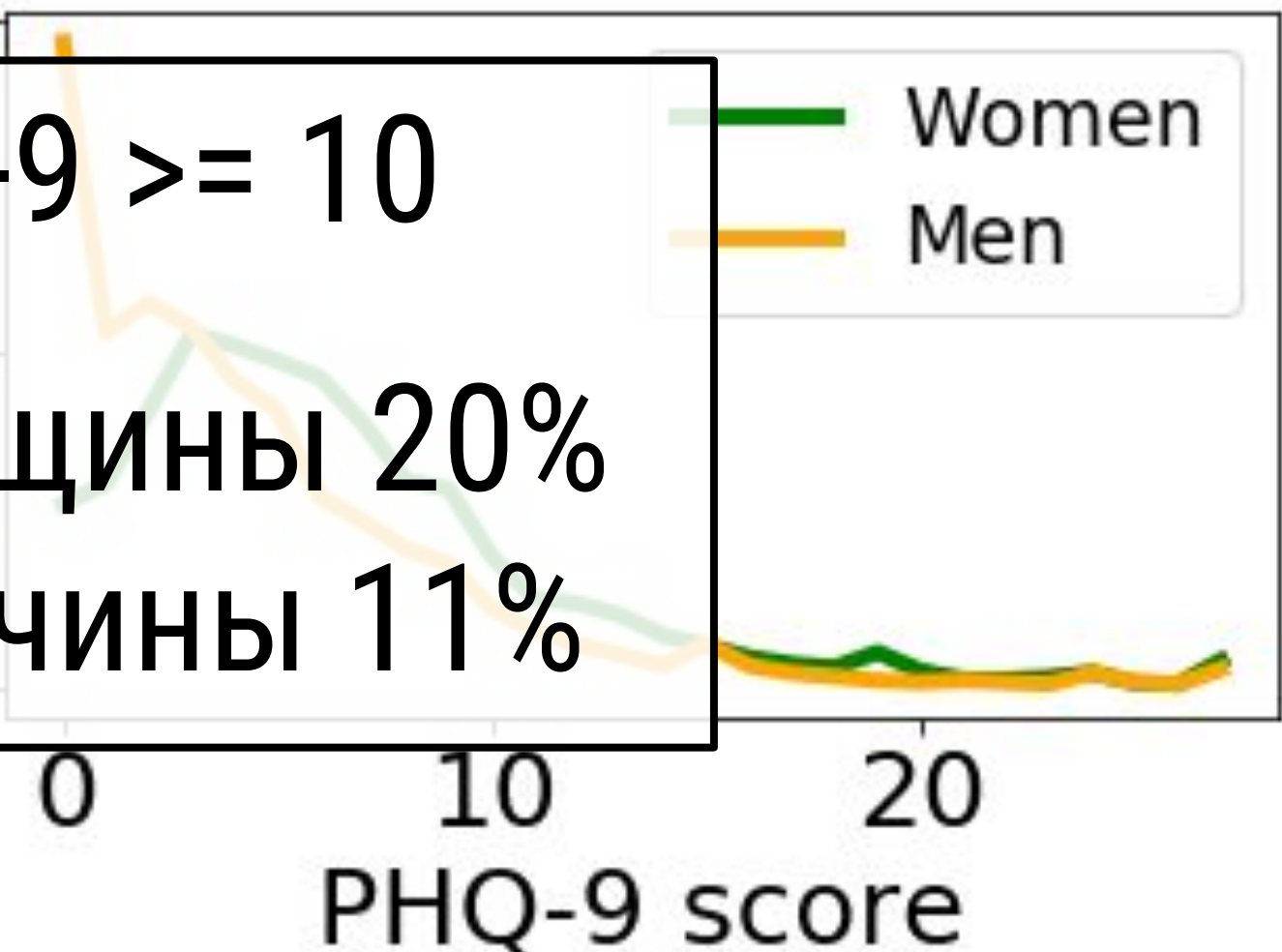
4000+ респондентов, 20-21 (в 2018)

RHQ-9

данные ВКонтакте (3000+)



PHQ-9 ≥ 10
Женщины 20%
Мужчины 11%



Депрессия и академическая успеваемость

Депрессия и академическая успеваемость

Отрицательная связь =>

больше ресурсов и внимания наименее благополучным

Депрессия и академическая успеваемость

Отрицательная связь =>

больше ресурсов и внимания наименее благополучным

Положительная или отсутствие =>

важность включения этого измерения

Депрессия и академическая успеваемость

Отрицательная связь =>

больше ресурсов и внимания наименее благополучным

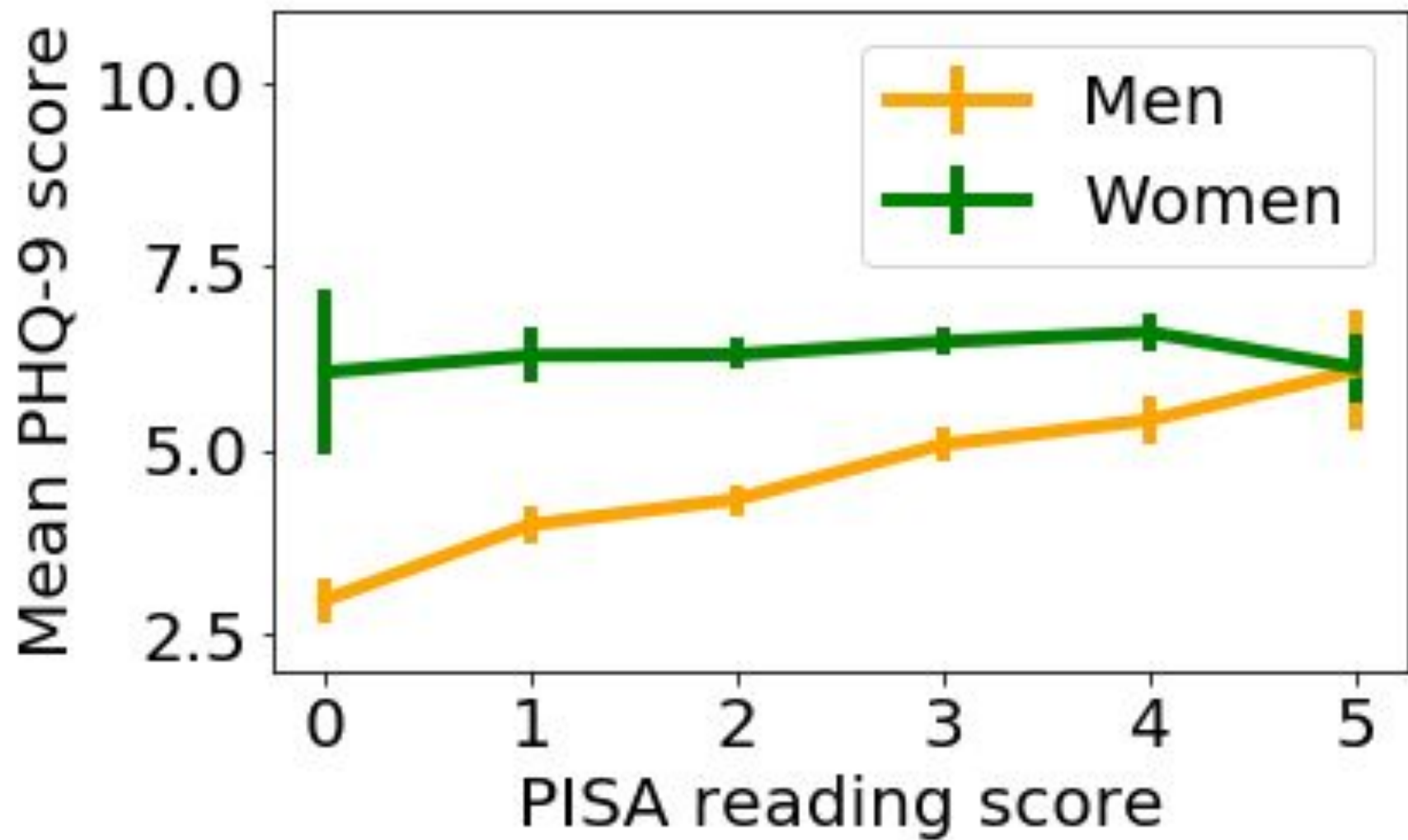
Положительная или отсутствие =>

важность включения этого измерения

Стандартизированные тесты (PISA, ЕГЭ)

Корреляция между степенью выраженности симптомов депрессии и академической успеваемостью

	Женщины	Мужчины
Чтение	0.02 (P = 0.3)	0.14 (P < 10 ⁻⁹)
Математика	0.02 (P = 0.3)	0.13 (P < 10 ⁻⁷)
Естественные науки	0.03 (P = 0.1)	0.13 (P < 10 ⁻⁷)
ЕГЭ	0.01 (P = 0.6)	0.10 (P < 10 ⁻³)



Сохраняется при контроле на
факт обучения в вузе и СЭС

Сохраняется при контроле на
факт обучения в вузе и СЭС

Относительный риск: 1.6

Цифровые следы

Предсказание депрессии по текстам постов на ВКонтакте

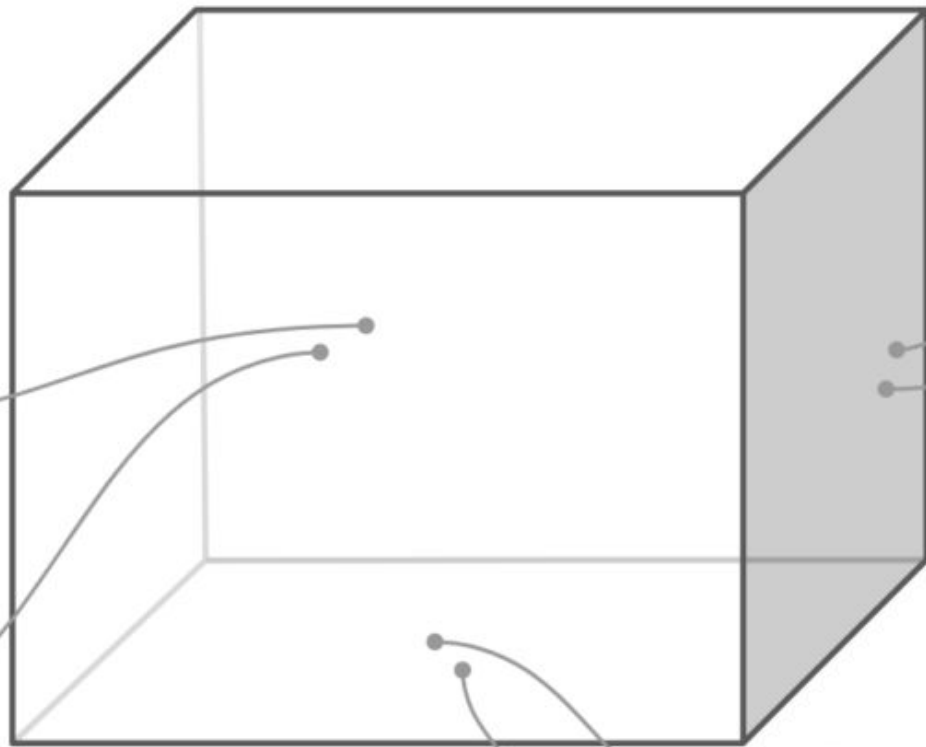
Предсказание депрессии по текстам постов на ВКонтакте

Не так быстро!

успеваемости

Предсказание ~~депрессии~~ по
текстам постов на ВКонтакте

Шаг 1. Векторные представления

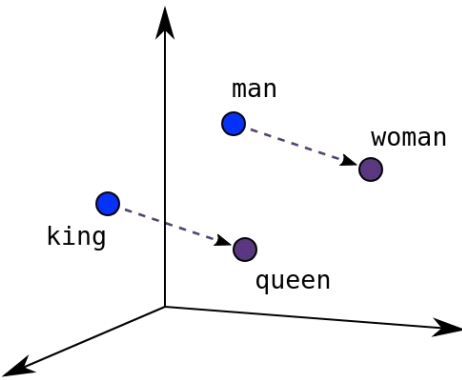


Audi

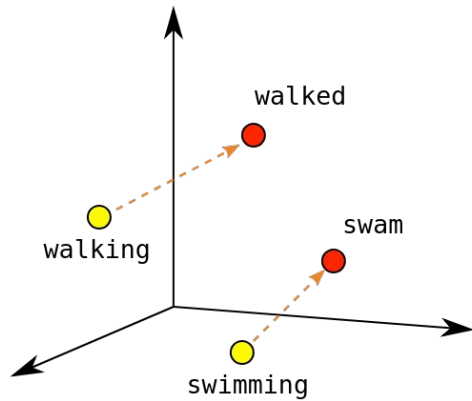


USC

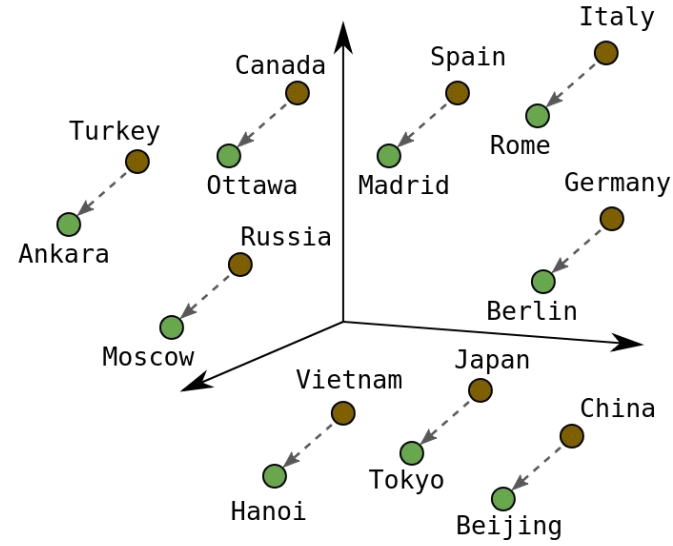




Male-Female



Verb Tense



Country-Capital

Низкая успеваемость

овен

гороскоп

стрелец

водолей

телец

Высокая успеваемость

дамблдор

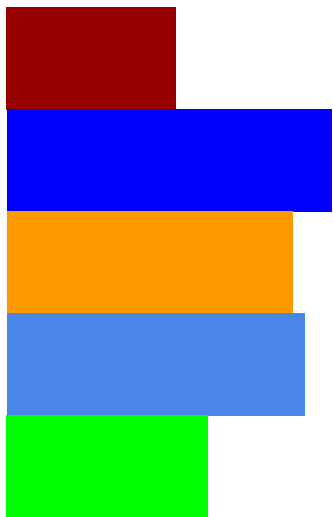
ньют

гермиона

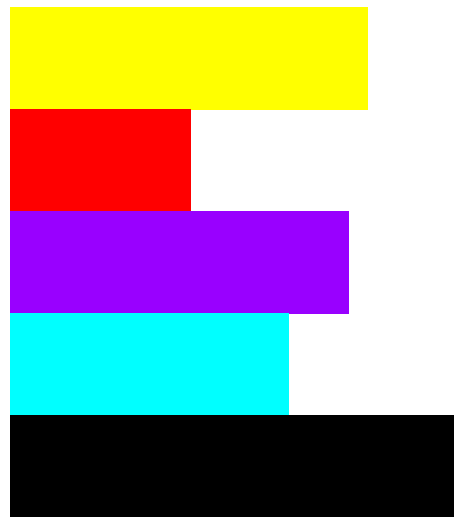
северус

макгонагалл

Низкая успеваемость



Высокая успеваемость



Низкая успеваемость

овен

гороскоп

стрелец

водолей

телец

Высокая успеваемость

дамблдор

ньют

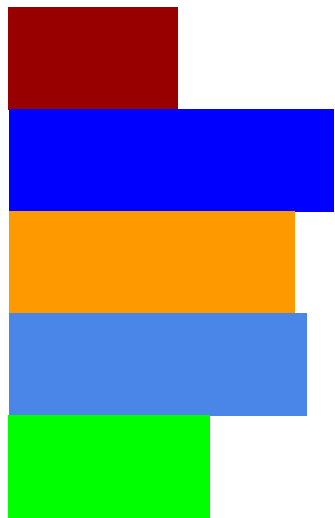
гермиона

северус

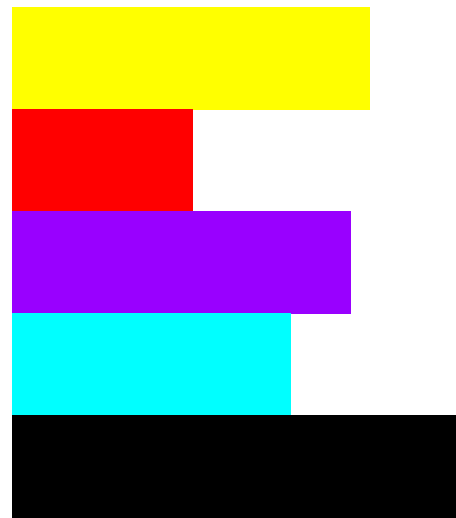
макгонагалл

зодиак ?

Низкая успеваемость



Высокая успеваемость



северус

НЬЮТ

дамблдор

макгонагалл

гермиона

овен

гороскоп

водолей

стрелец

телец

зодиак ?

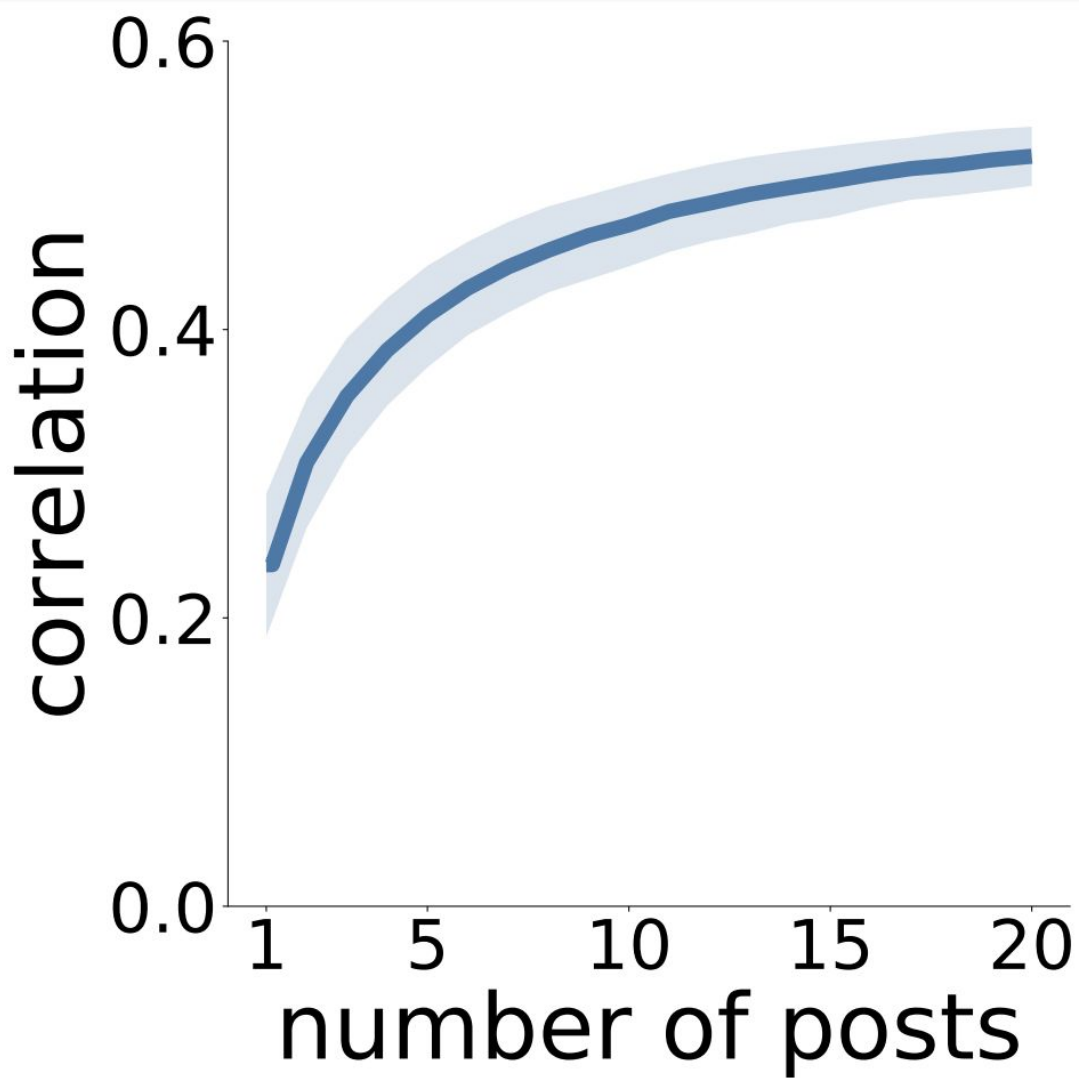


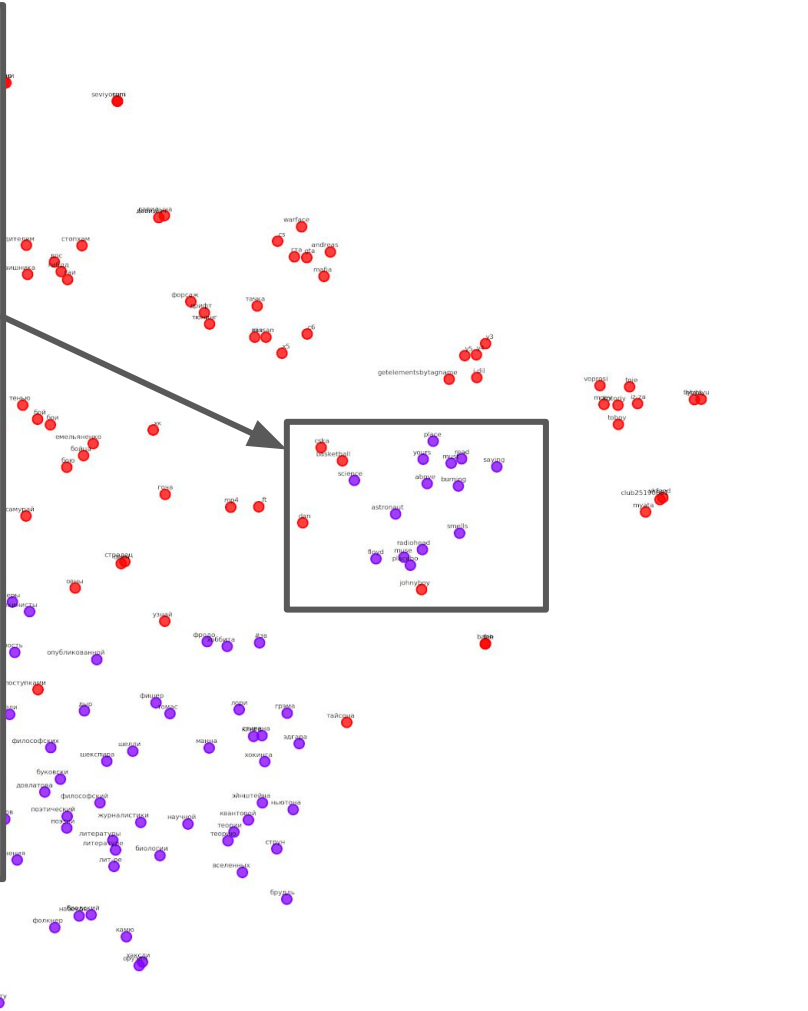
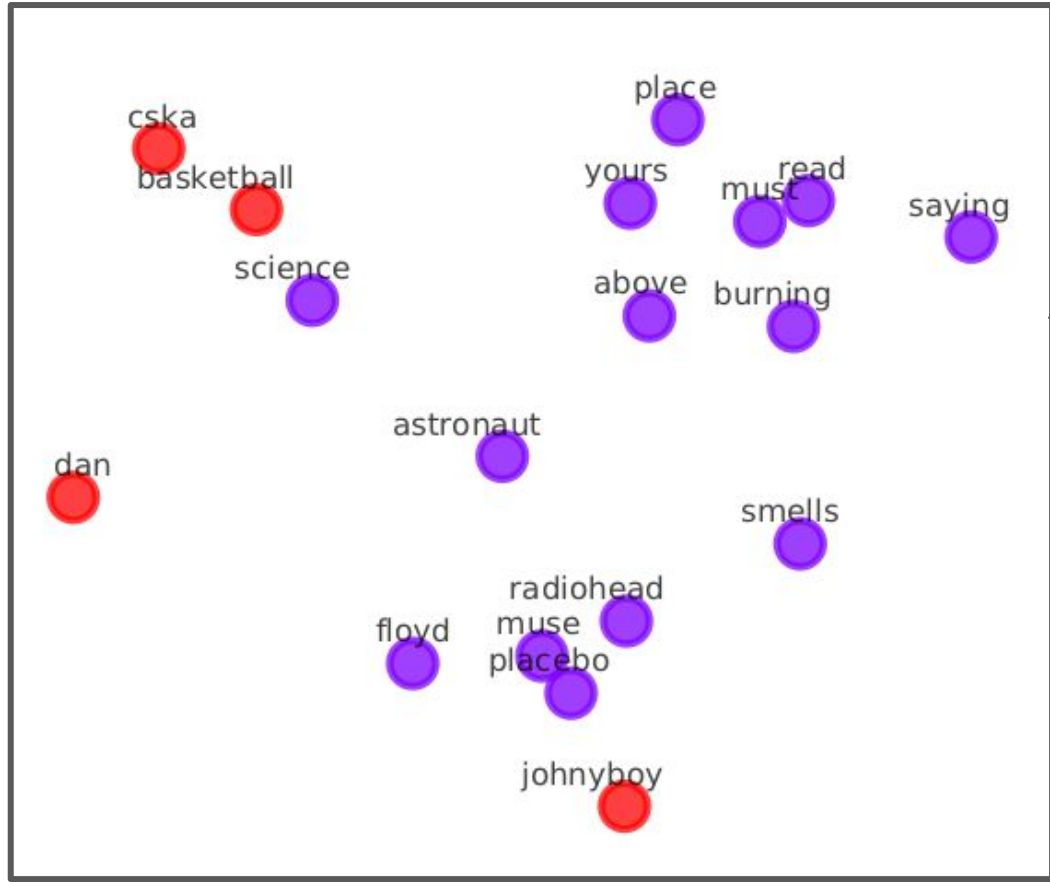
Обучаем модель на 100Гб+ постов ВКонтакте

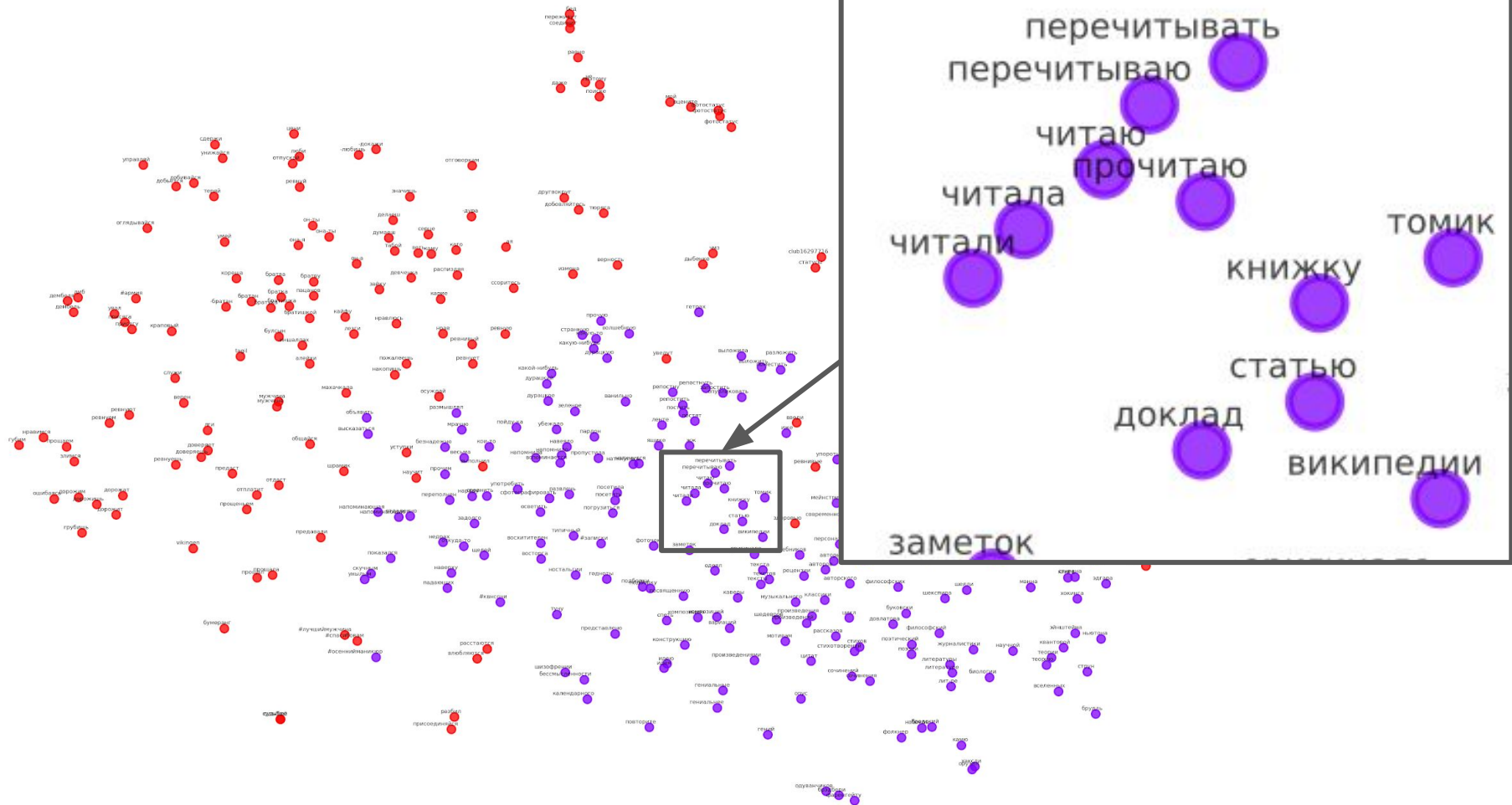
Переводим слова в вектора размерности 300

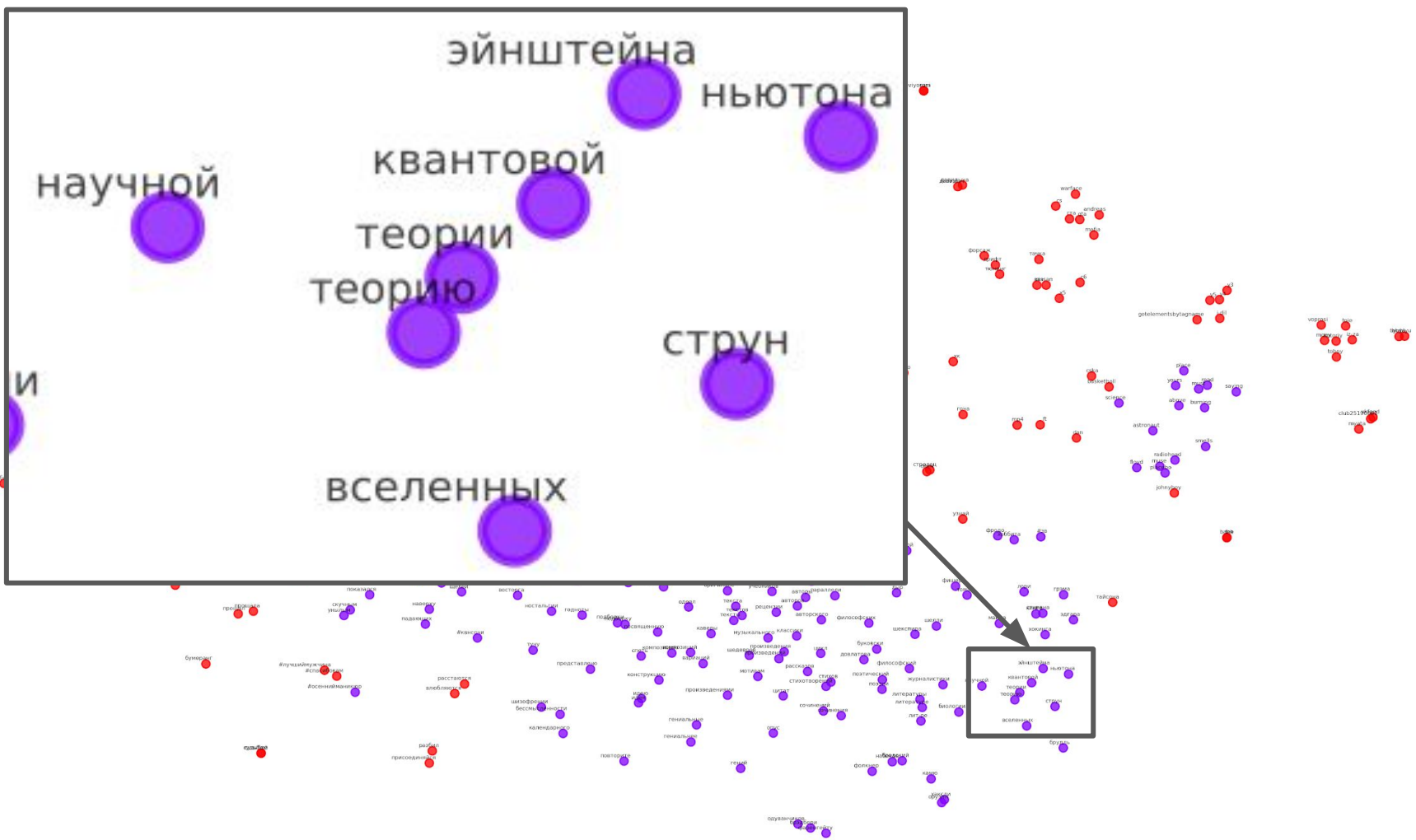
Пост = среднее всех слов

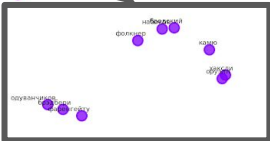
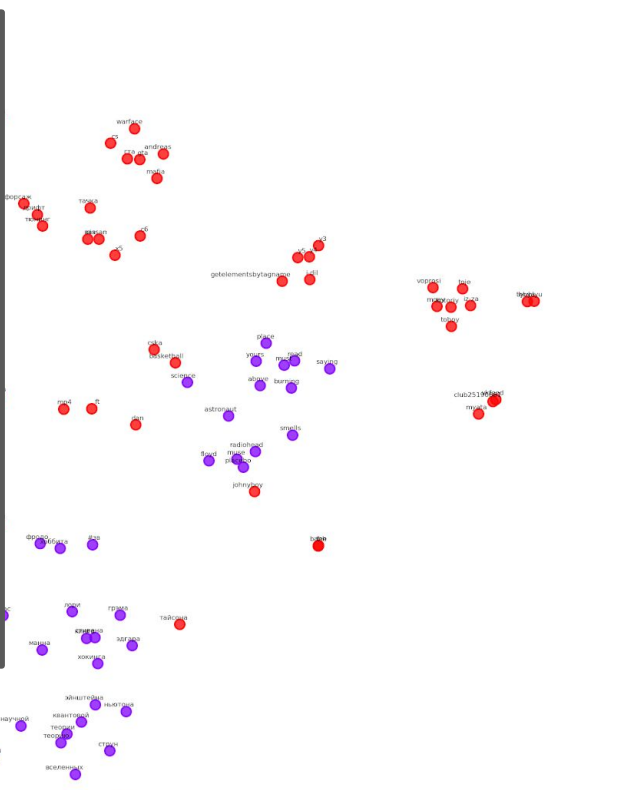
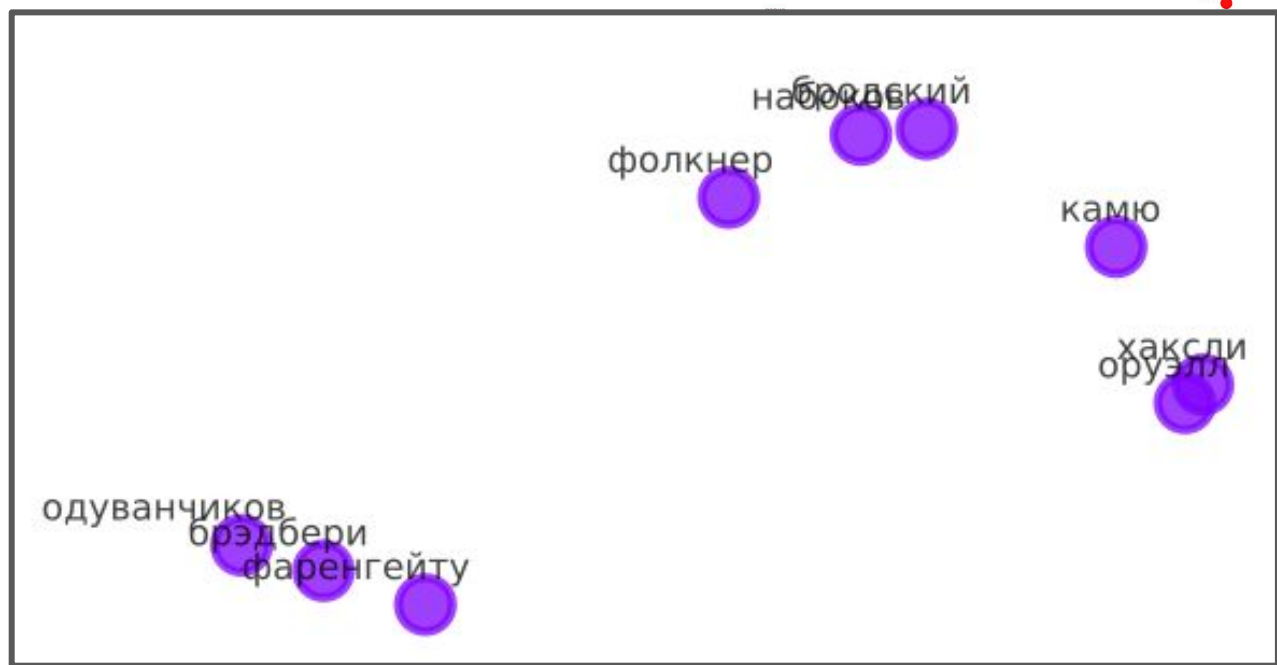
Линейная регрессия

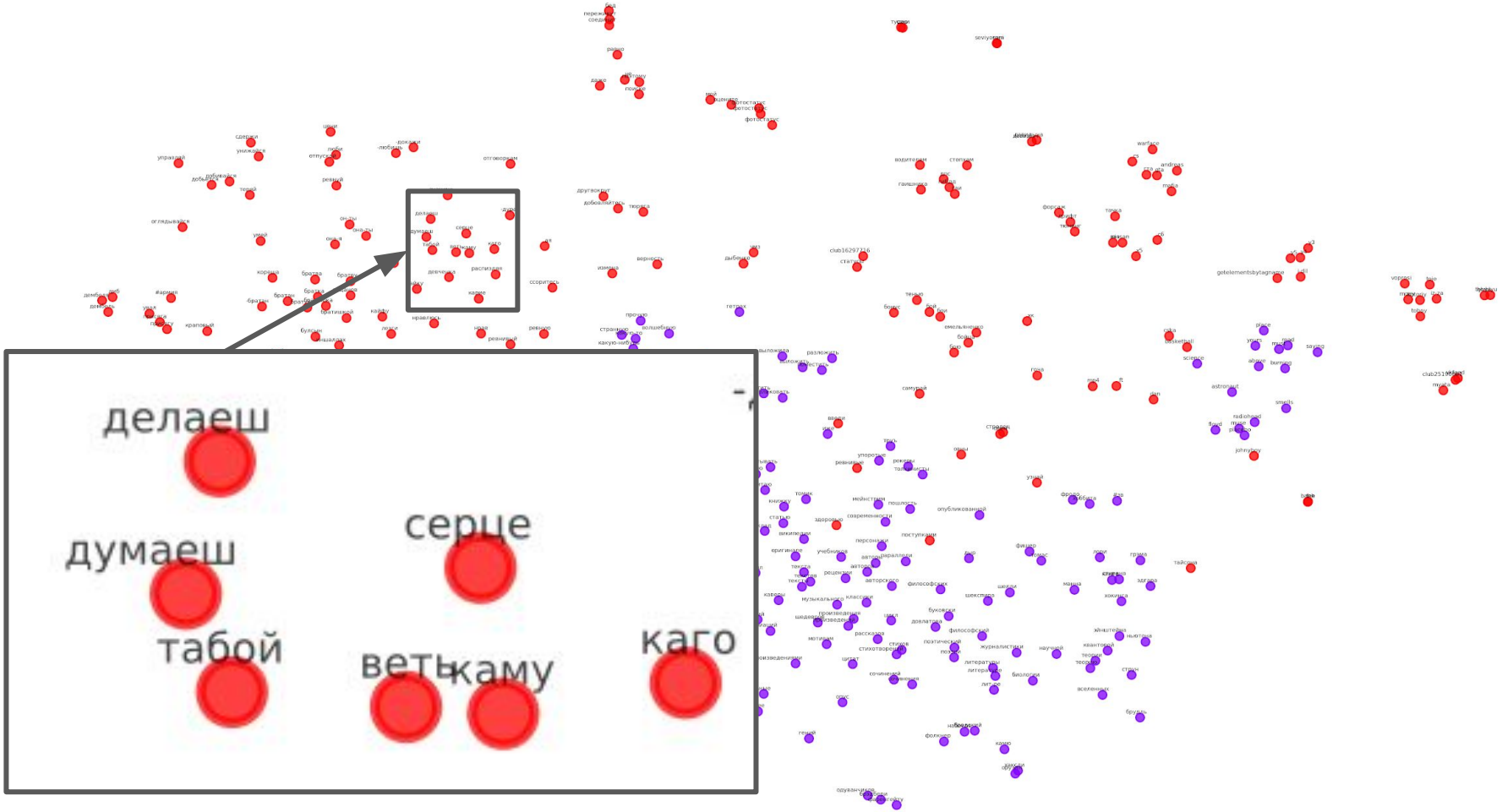












бурдые 0.01%

312204 out of 1000000 by frequency (31.22%) and 82 out of 1000000 (0.01%) by score

набоков 2.02%

30139 out of 1000000 by frequency (3.01%) and 20193 out of 1000000 (2.02%) by score

книга 29.45%

1626 out of 1000000 by frequency (0.16%) and 294503 out of 1000000 (29.45%) by score

университет 37.15%

5637 out of 1000000 by frequency (0.56%) and 371504 out of 1000000 (37.15%) by score

спо 75.27%

56590 out of 1000000 by frequency (5.66%) and 752737 out of 1000000 (75.27%) by score

армия 87.51%

4313 out of 1000000 by frequency (0.43%) and 875097 out of 1000000 (87.51%) by score

НЬЮТ 3.45%

325429 out of 1000000 by frequency (32.54%) and 34541 out of 1000000 (3.45%) by score

макгонагалл 3.3%

188638 out of 1000000 by frequency (18.86%) and 32983 out of 1000000 (3.3%) by score

гарри 8.6%

4677 out of 1000000 by frequency (0.47%) and 85971 out of 1000000 (8.6%) by score

дамблдор 13.5%

70450 out of 1000000 by frequency (7.05%) and 134985 out of 1000000 (13.5%) by score

весы 92.8%

3997 out of 1000000 by frequency (0.4%) and 928027 out of 1000000 (92.8%) by score

гороскоп 93.61%

5209 out of 1000000 by frequency (0.52%) and 936102 out of 1000000 (93.61%) by score

овен 98.78%

5657 out of 1000000 by frequency (0.57%) and 987802 out of 1000000 (98.78%) by score

ВИНО 51.74%

1836 out of 1000000 by frequency (0.18%) and 517380.5 out of 1000000 (51.74%) by score

ПИВО 61.19%

1510 out of 1000000 by frequency (0.15%) and 611897 out of 1000000 (61.19%) by score

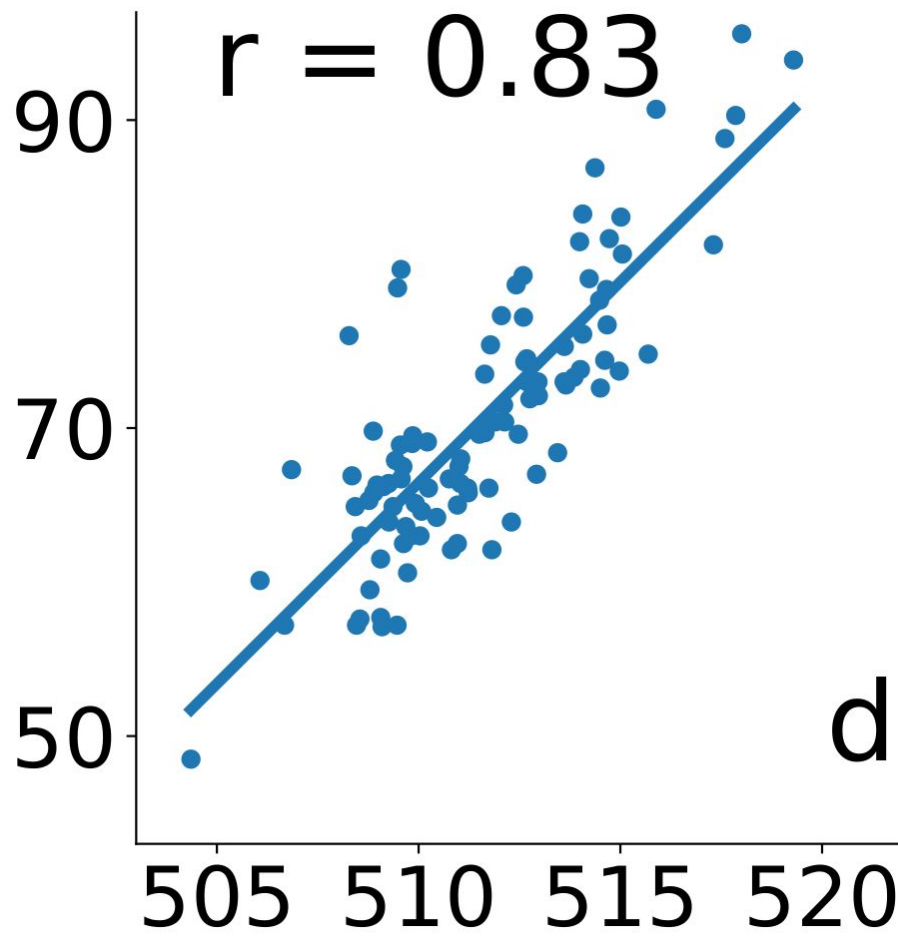
ВОДКА 70.26%

3925 out of 1000000 by frequency (0.39%) and 702605 out of 1000000 (70.26%) by score

Пользователи ВКонтакте, указавшие, что в
одном из 100 крупнейших вузов

Предсказываем их успеваемость

Средний балл ЕГЭ
абитуриентов



Предсказанный балл PISA



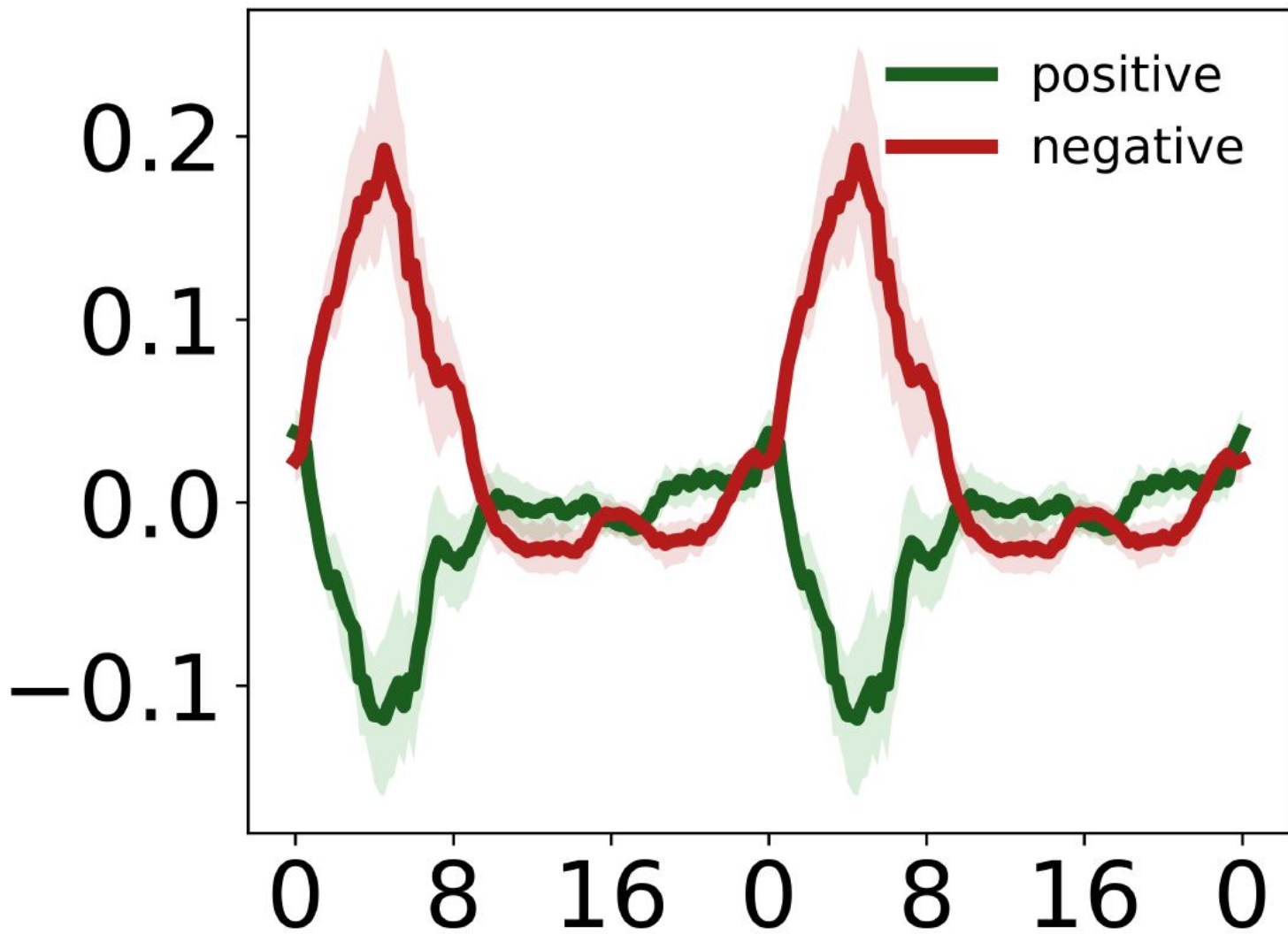
Но не теперь...

Тональность постов

Модель обученная на размеченных твитах

Корреляция с депрессией до 0.7 (зависит от количества постов и времени написания)

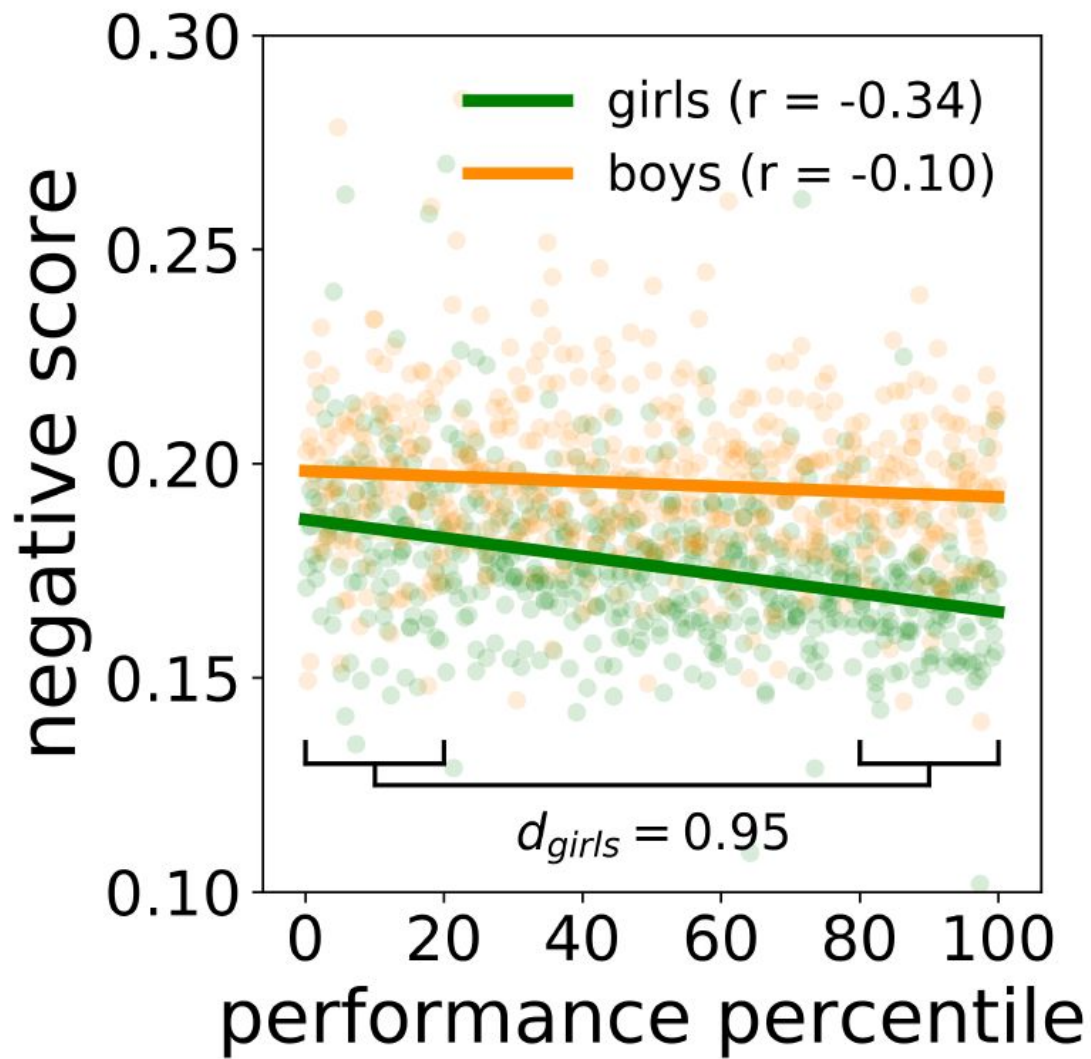
sentiment



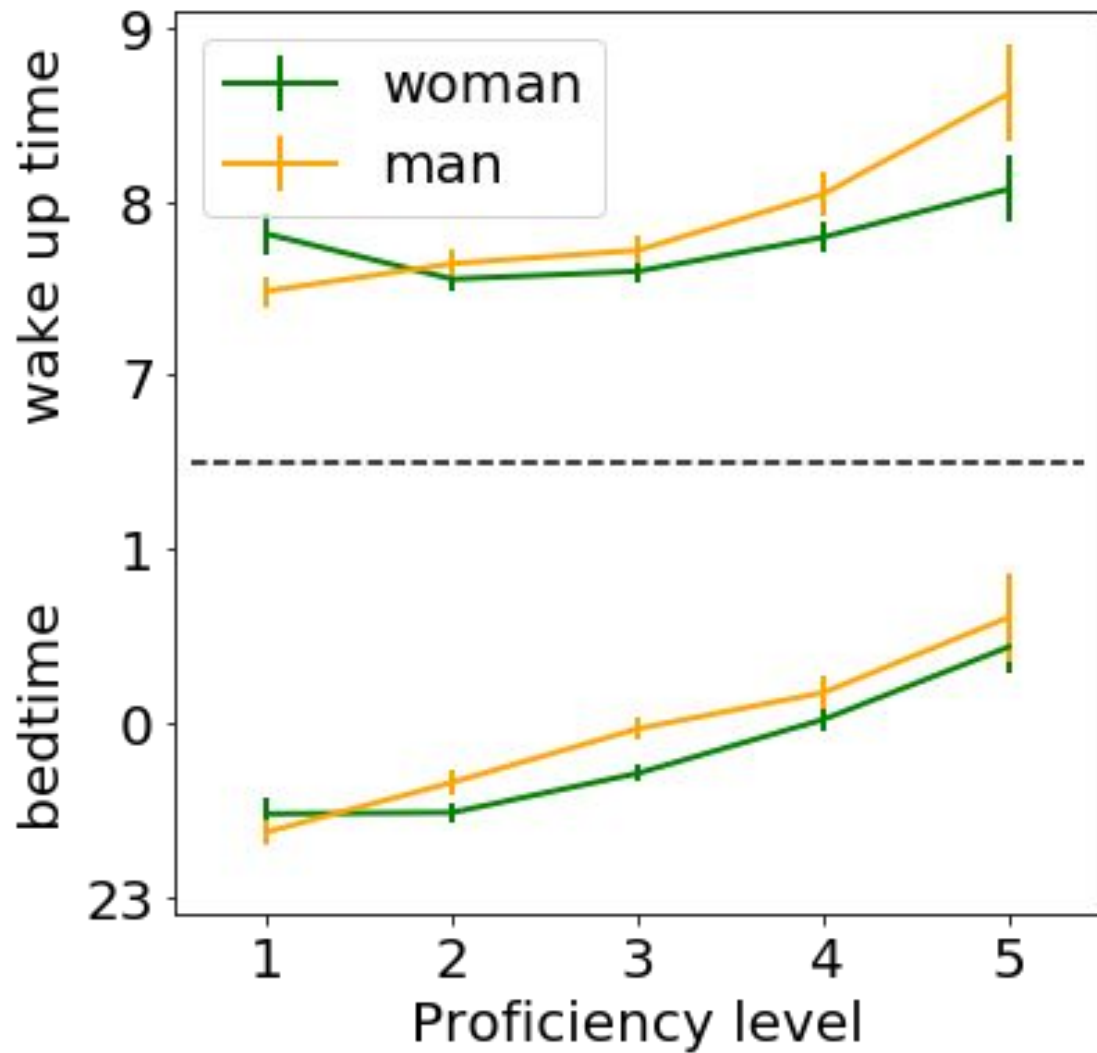
Тональность постов

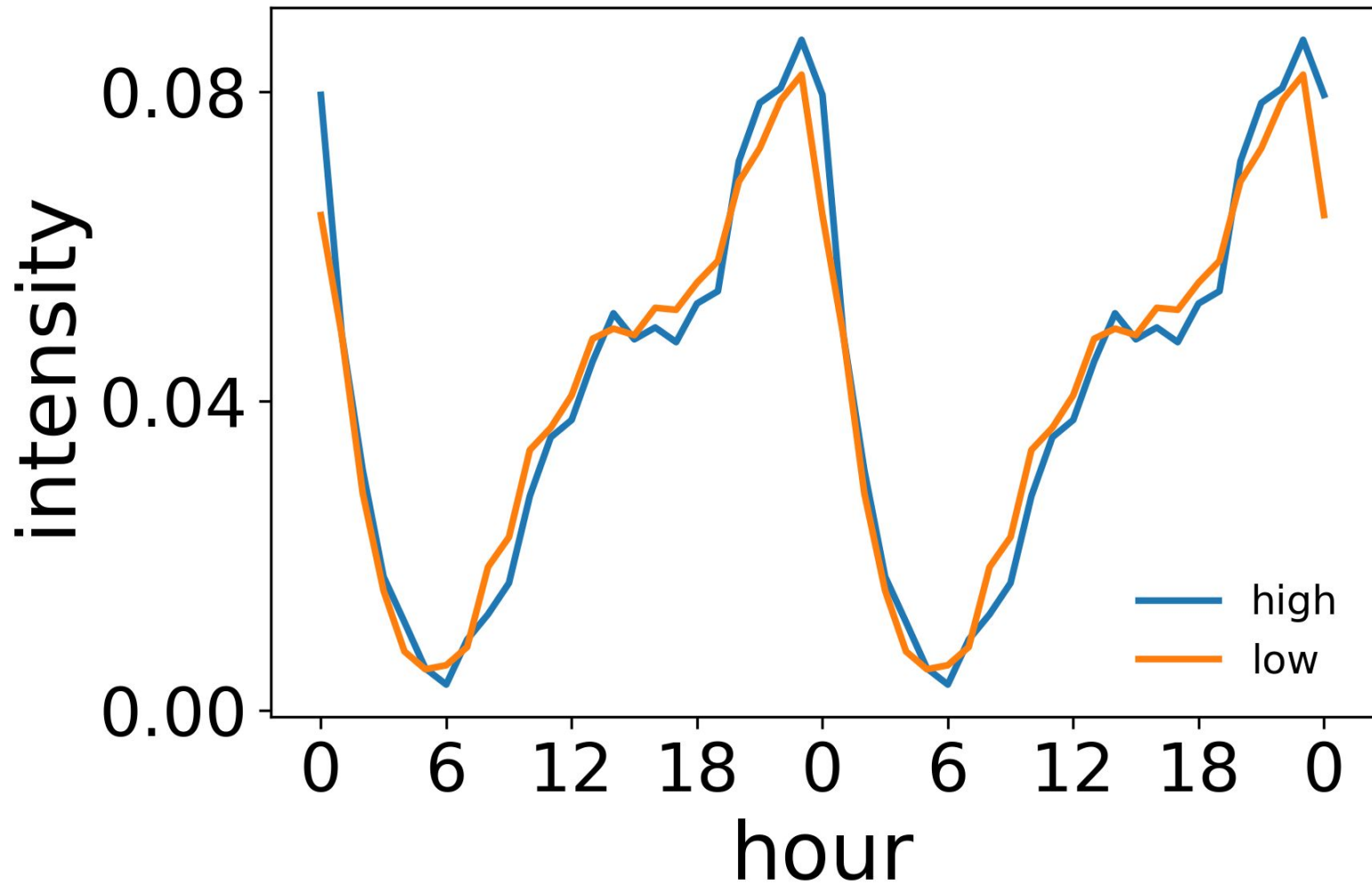
500+ школ Петербурга

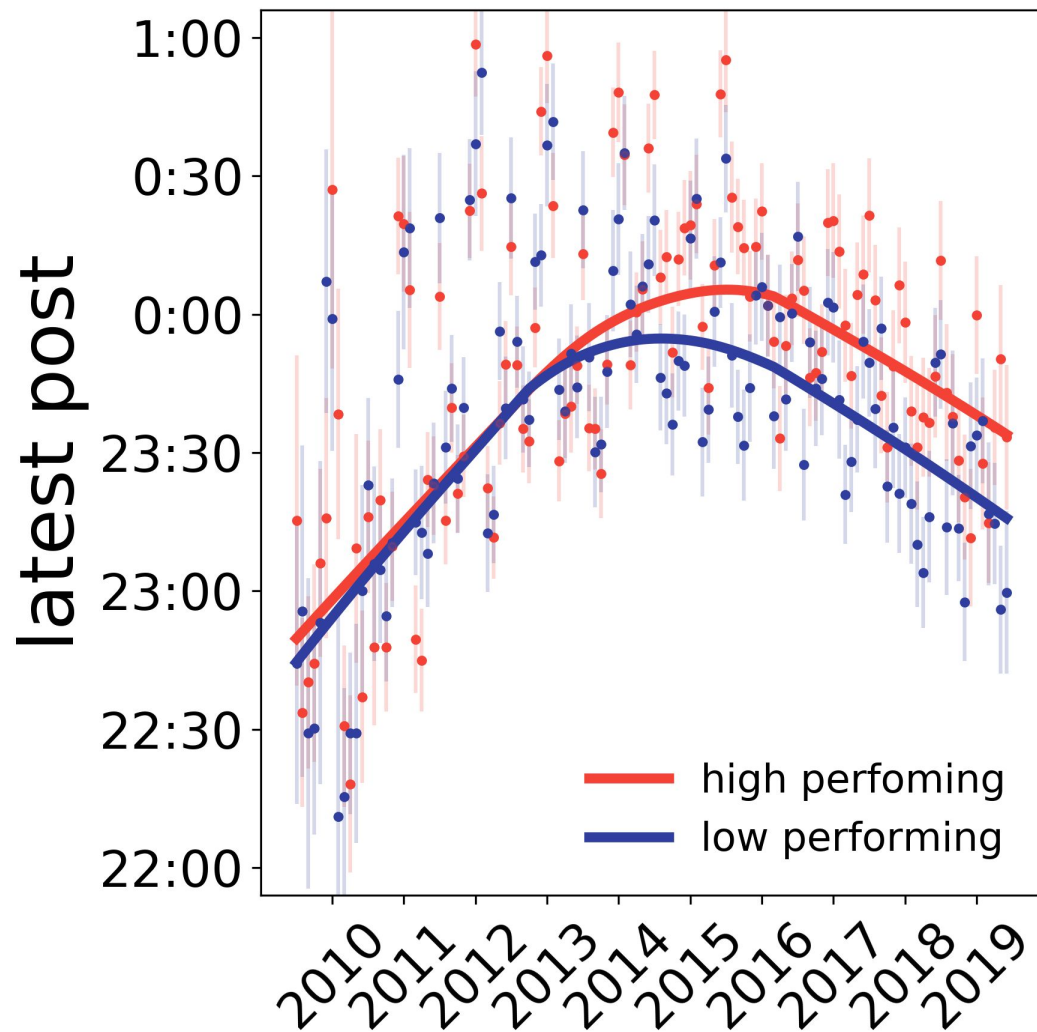
известен ЕГЭ выпускников



СОН

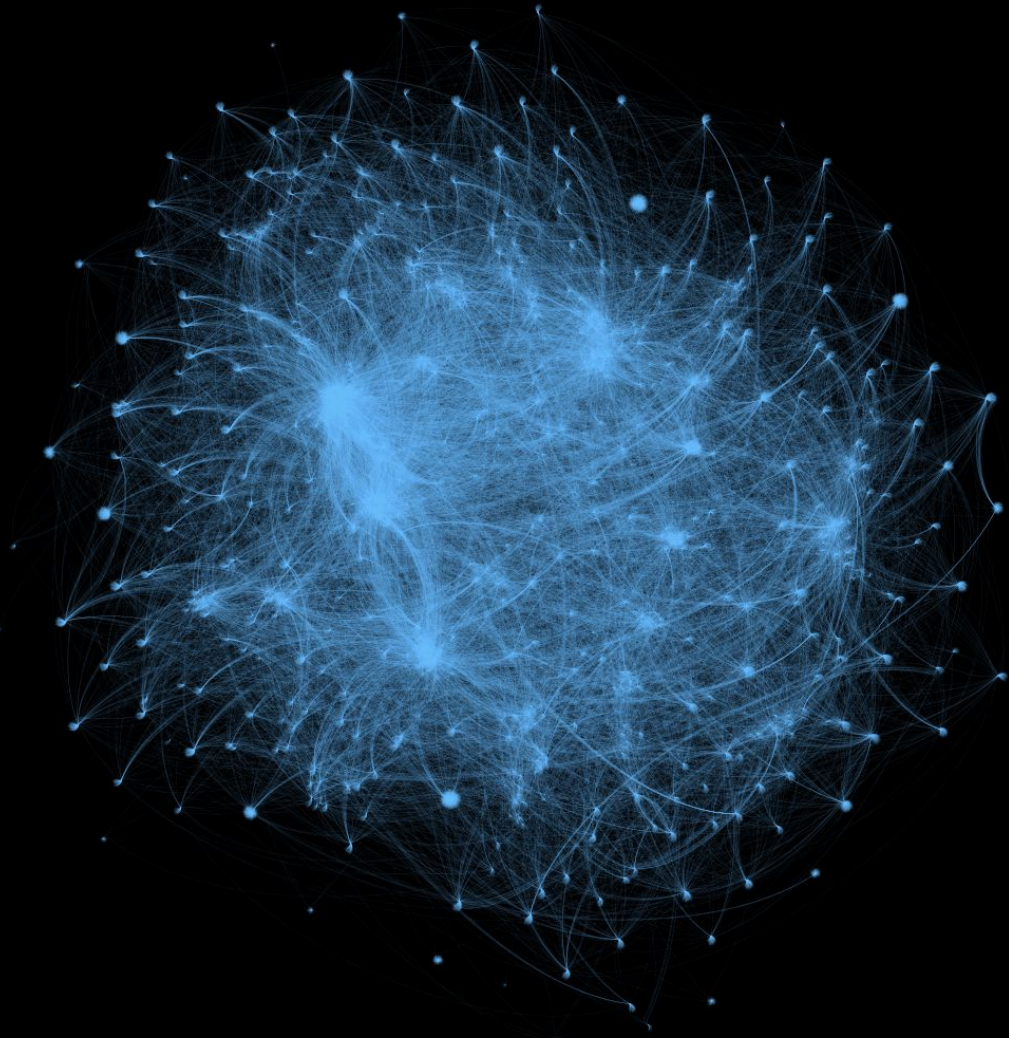


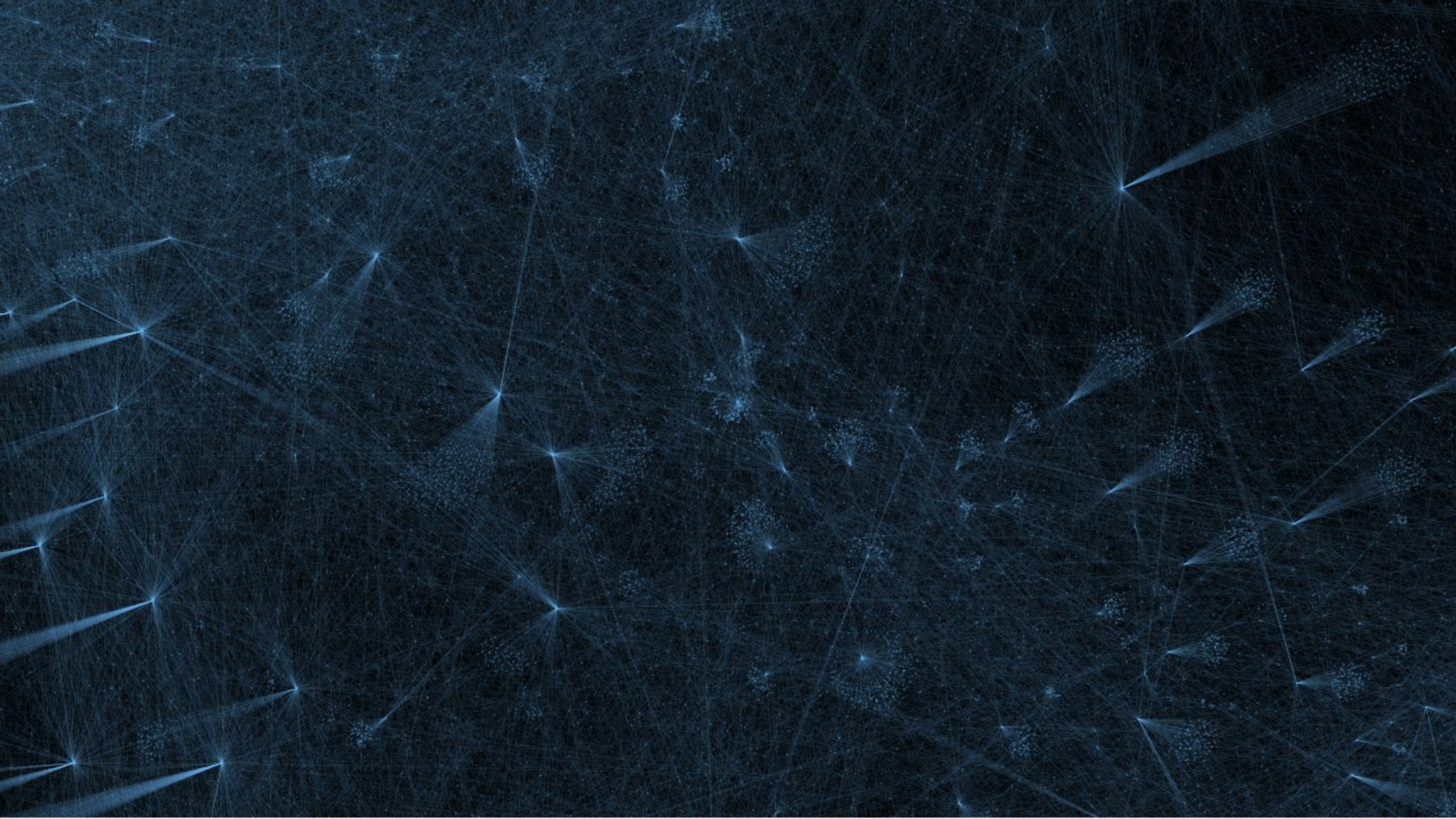




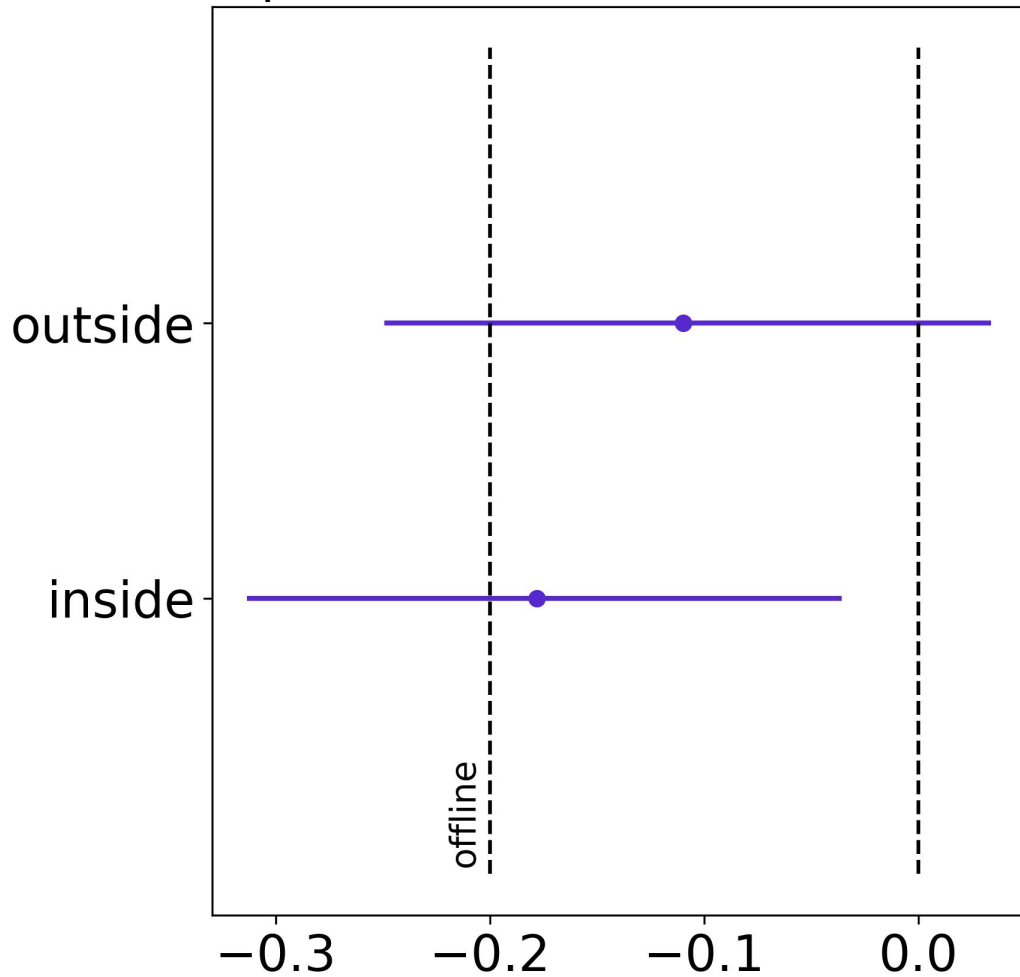
Социальные связи

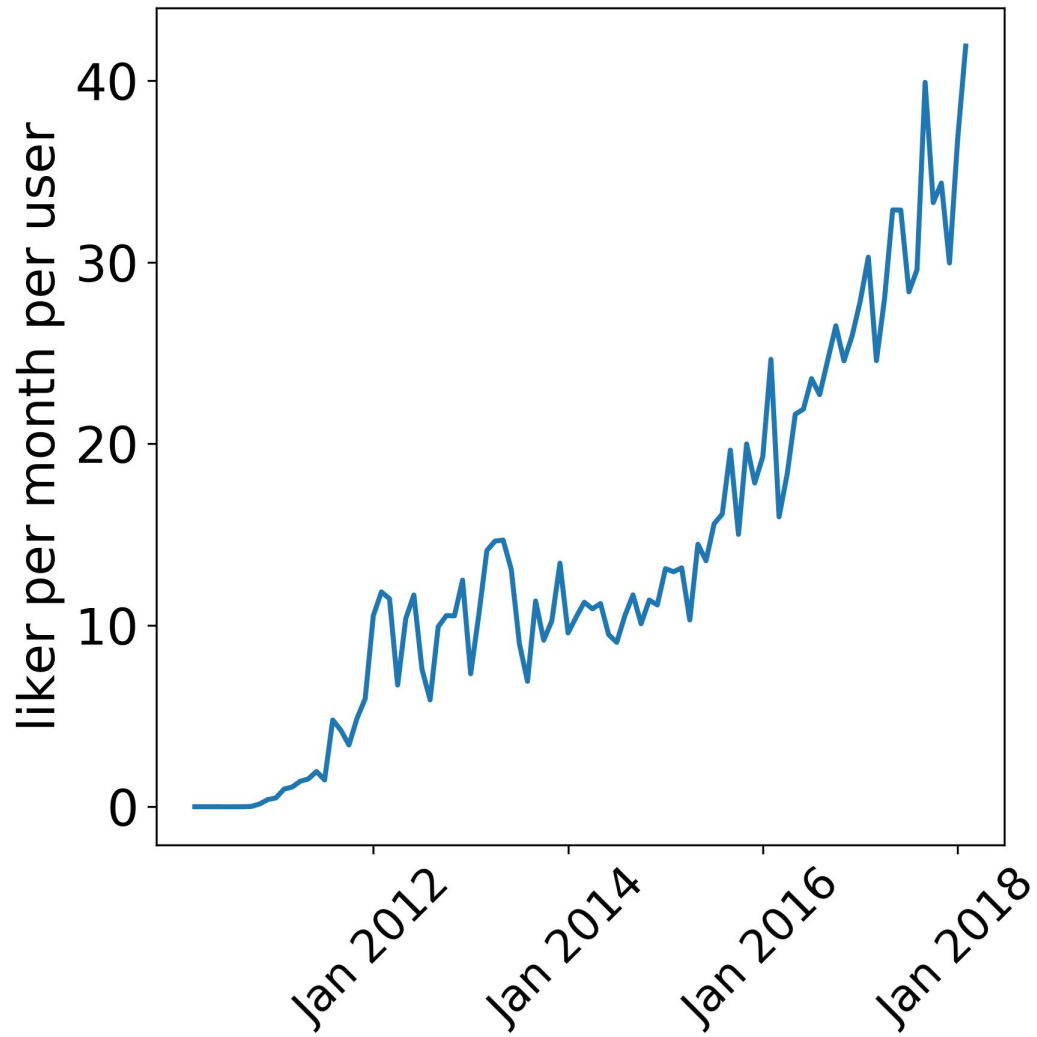
Учащиеся одной школы и
их друзья на ВКонтакте





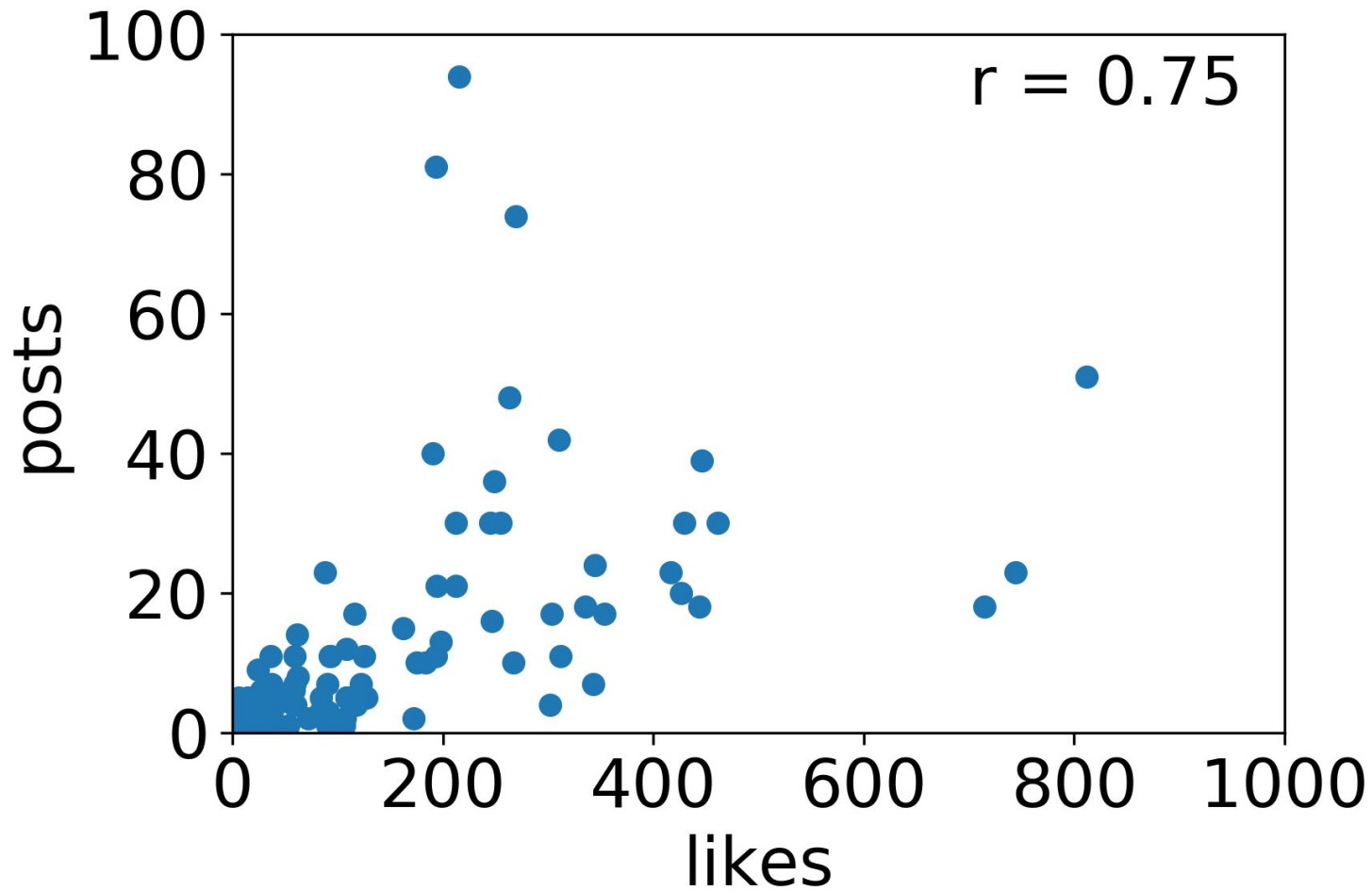
Depression and number of friends

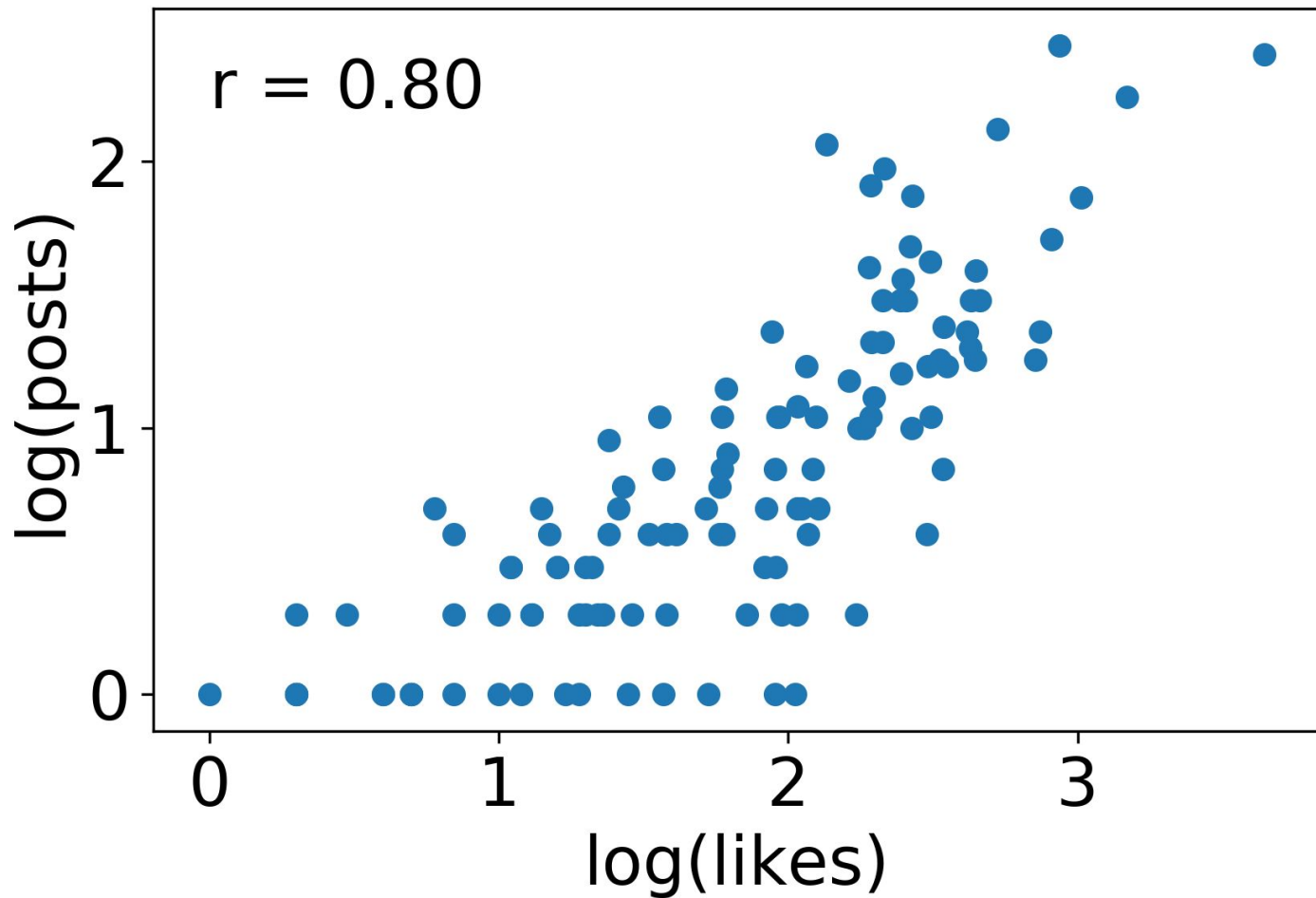




Корреляция 0.03 [-0.11; 0.17]

количество лайков и депрессия

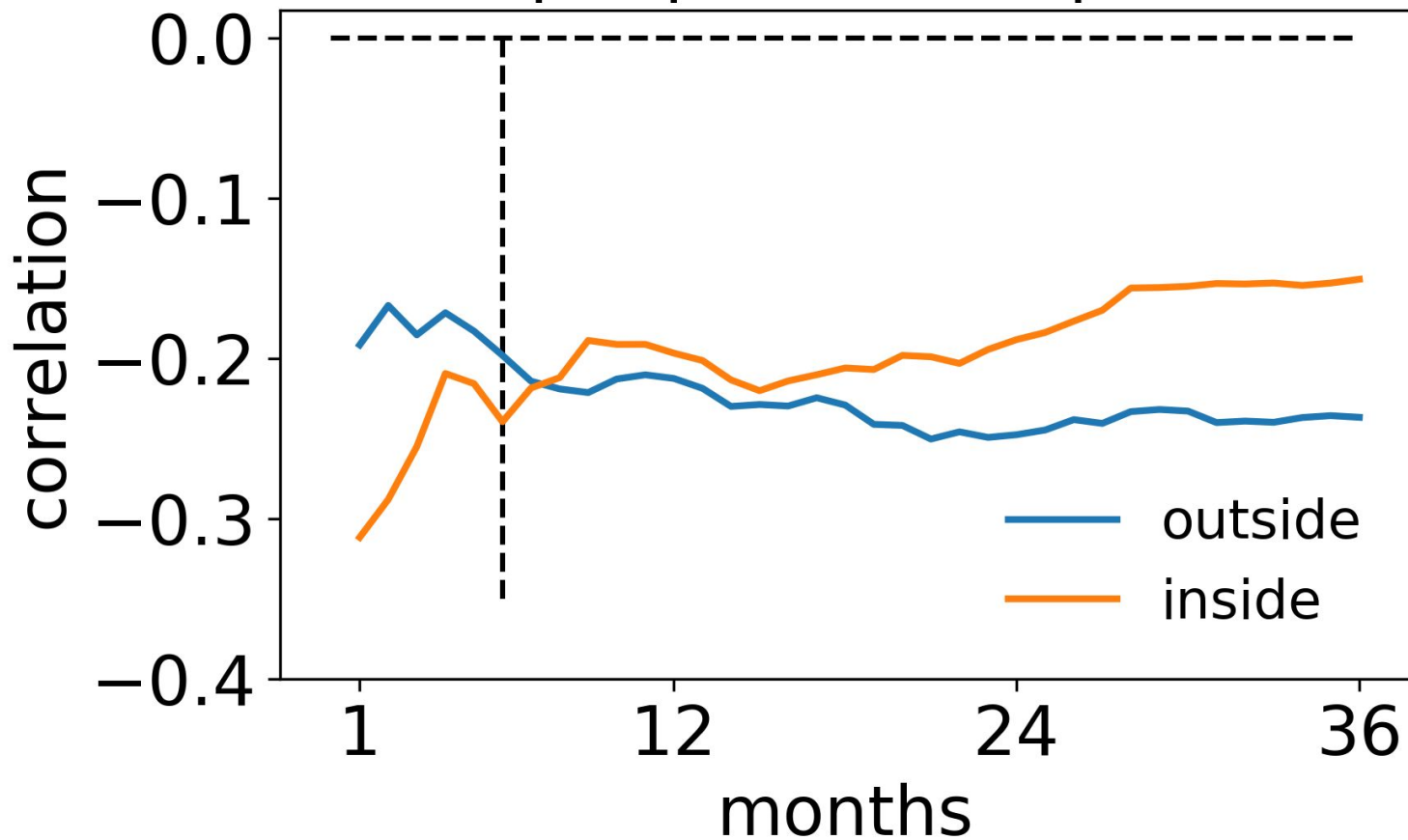




Корреляция 0.21 [0.05; 0.36]

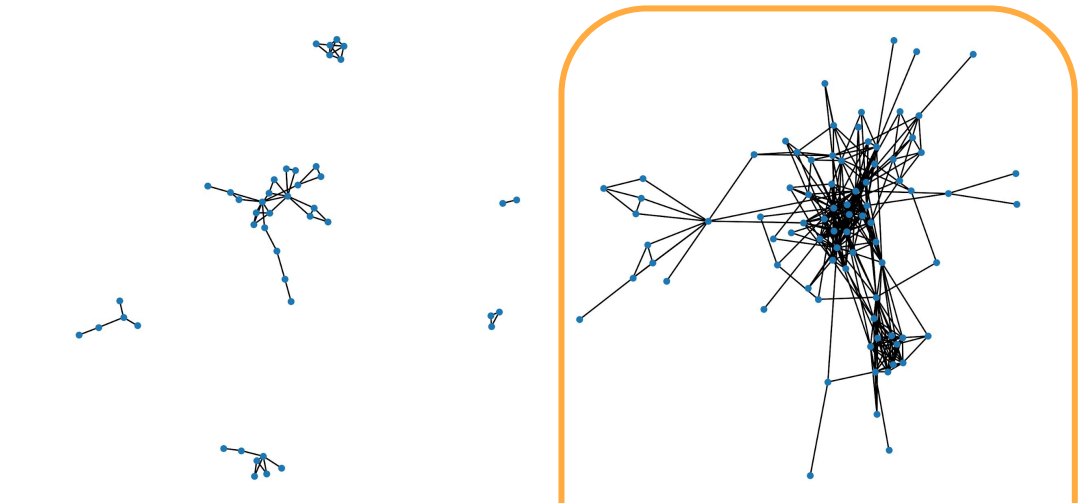
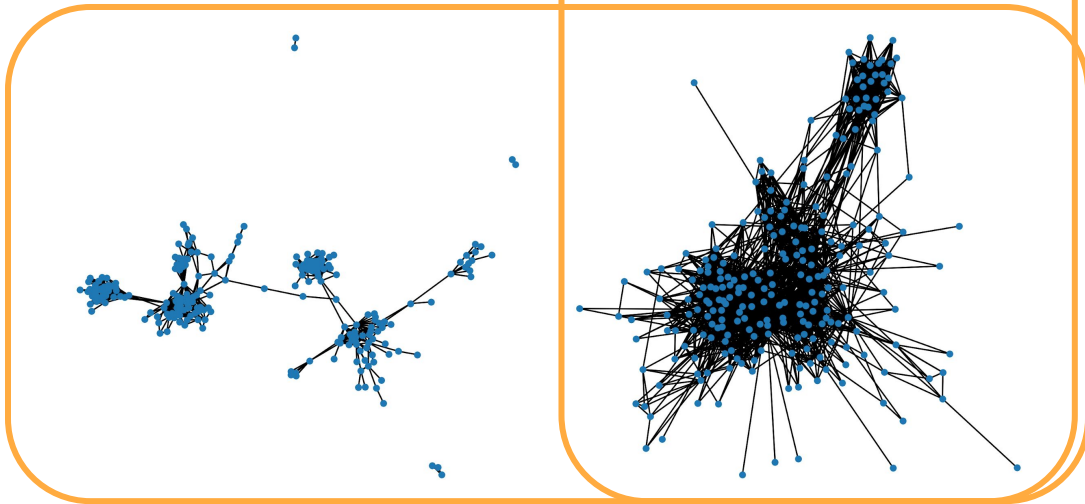
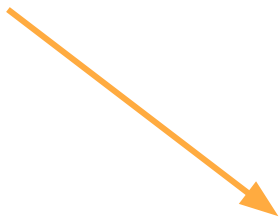
количество постов и депрессия

Likes per post and depression



3000 эго-сетей

Много друзей



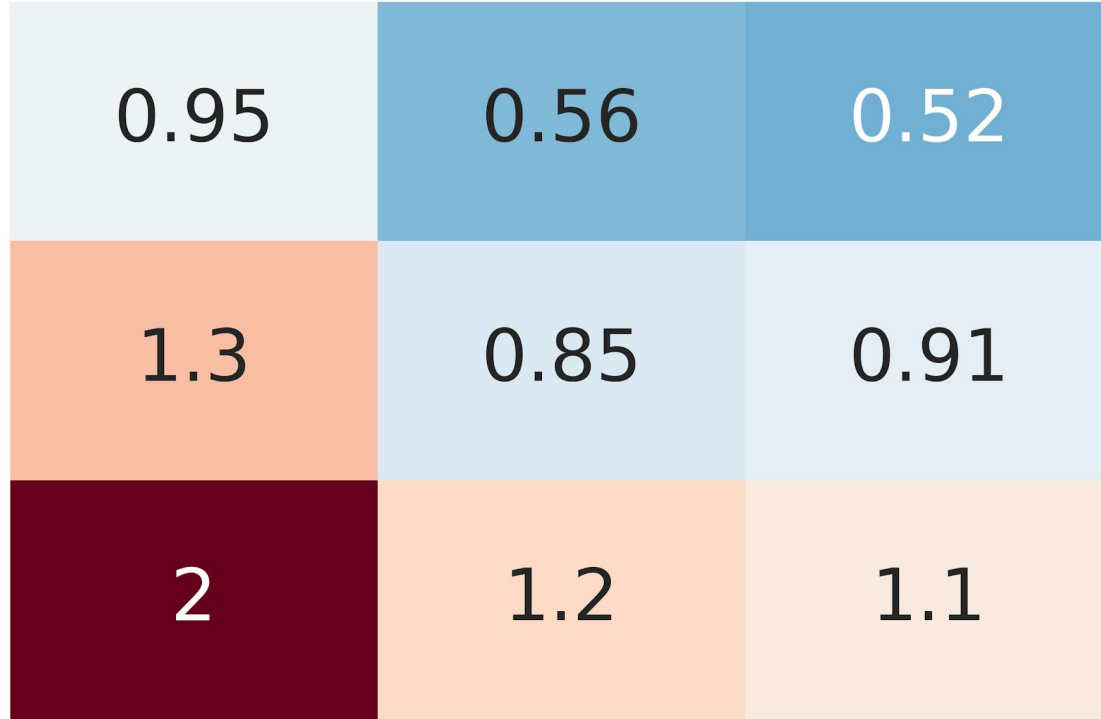
Связаны друг с другом



Relative risk of depression

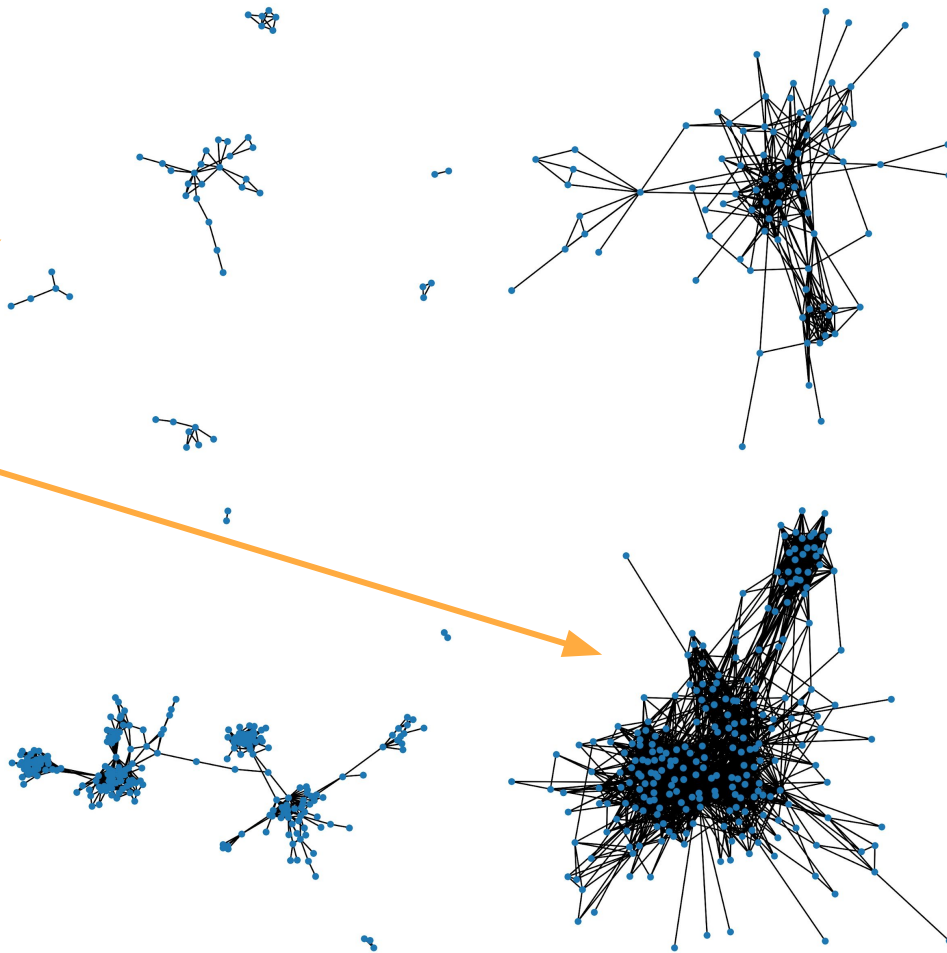
degree centrality

high
medium
low



low medium high
clustering coefficient

4x риск



0 BR CONTINUED

Ivan Smirnov



ibsmirnov@hse.ru

#education #computational social science
#inequality #social networks #well-being
#data science #institute of education #HSE