

# Formation of homophily in academic performance

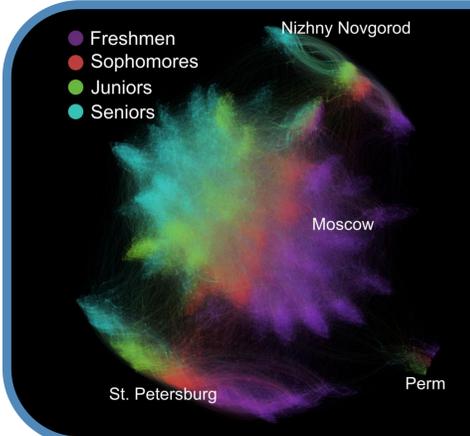
Ivan Smirnov<sup>1</sup> and Stefan Thurner<sup>2,3,4</sup>

ibsmirnov@hse.ru

<sup>1</sup> Institute of Education, National Research University Higher School of Economics

<sup>2</sup> Section for Science of Complex Systems, Medical University of Vienna

<sup>3</sup> Santa Fe Institute <sup>4</sup> IIASA



## Motivation

Information technologies may help to break the cycle of inequality reproduction because students are not limited by their immediate environment anymore

Information technologies may also reinforce segregation and prevent social mobility

**What is happening with segregation by academic performance in the digital world?**



Friendship network of university students on a popular social networking site (extended data set)

## Data

Academic performance records of 6,580 high school and university students

History of their interaction ("likes") on a popular web site across 42 months

## Methods

For each 3 months period we compute homophily index H:

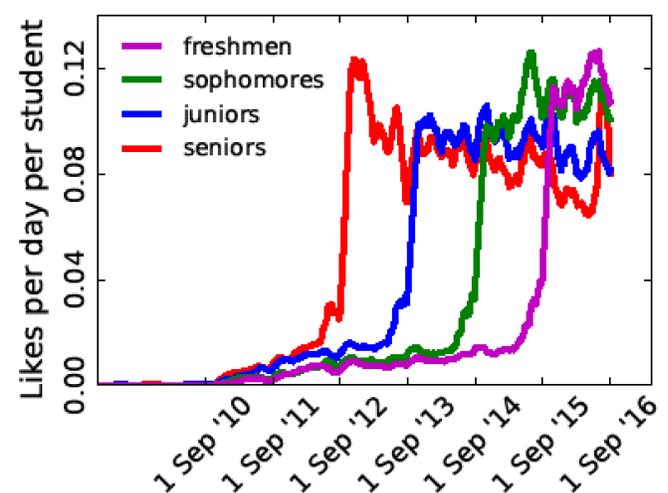
$$H(t) = \text{corr}(G_i(t), \langle G_j(t) \rangle_{j | A_{ij}(t)=1}), \text{ where}$$

$G_i(t)$  is GPA of student  $i$  at time  $t$ ,

$A_{ij}(t) = 1$  if student  $i$  places at least one "like" to student  $j$  from time  $t - 1$  to time  $t$ ,

$\langle \cdot \rangle_j$  means average over all  $j$  satisfying condition.

Changes in average number of student interactions with time (extended data set)



## Results

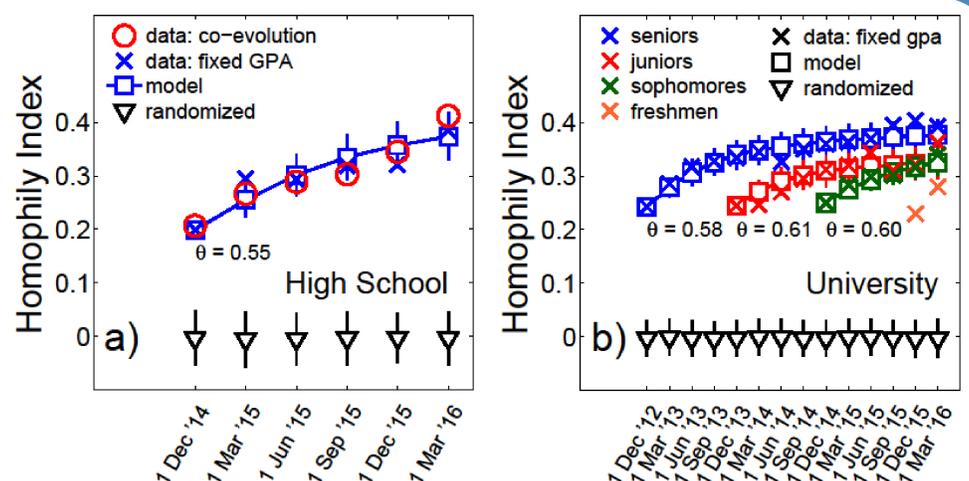
There is a **strong homophily** in academic performance for high school (a) and university (b) students

The significance of the observed effect is measured with a randomization test (triangles)

Homophily increases with time by almost a factor of 2 (circles)

By fixing grades (crosses) we show that it is **explained by social selection** rather than by peer influence

Results can be understood with a simple model (squares)



## Model

Initial network is equal to the observed network at time step 1

From time  $t$  to  $t + 1$  the model runs through following steps

For each **student**:

Pick random **friend** and random **not friend**

If  $|GPA_{\text{student}} - GPA_{\text{not friend}}| < |GPA_{\text{student}} - GPA_{\text{friend}}|$ :

Swap link from **friend** to **not friend**

Else:

Swap link from **friend** to **not friend** with probability  $\theta$

## Discussion

Physical mixing of students in the same educational institution does not lead to a homogeneous mixing of social ties

This mechanism is potentially reinforced by the modern technologies where maintaining links does not require physical presence anymore

It could have important implications for the understanding of segregation, inequality and social immobility in the digital age