Formation of homophily in academic performance: Students change their friends rather than performance

Ivan Smirnov[1], Stefan Thurner[2,3,4,5] *

1 Institute of Education, National Research University Higher School of Economics, Moscow, Russia,
2 Section for Science of Complex Systems, Medical University of Vienna, Vienna, Austria, 3 Santa Fe
Institute, Santa Fe, NM, United States of America, 4 IIASA, Laxenburg, Austria, 5 Complexity Science Hub
Vienna, Josefstädterstr. 39, A-1080, Vienna, Austria

* stefan.thurner@meduniwien.ac.at

Abstract

Homophily, the tendency of individuals to associate with others who share similar traits, has been identified as a major driving force in the formation and evolution of social ties. In many cases, it is not clear if homophily is the result of a socialization process, where individuals change their traits according to the dominance of that trait in their local social networks, or if it results from a selection process, in which individuals reshape their social networks so that their traits match those in the new environment. Here we demonstrate the detailed temporal formation of strong homophily in academic achievements of high school and university students. We analyze a unique dataset that contains information about the detailed time evolution of a friendship network of 6,000 students across 42 months. Combining the evolving social network data with the time series of the academic performance (GPA) of individual students, we show that academic homophily is a result of selection: students prefer to gradually reorganize their social networks according to their performance levels, rather than adapting their performance to the level of their local group. We find no signs for a pull effect, where a social environment of good performers motivates bad students to improve their performance. We are able to understand the underlying dynamics of grades and networks with a simple model. The lack of a social pull effect in classical educational settings could have important implications for the understanding of the observed persistence of segregation, inequality and social immobility in societies.

Introduction

Homophily is the tendency of humans to associate with others who share similar traits. It has been observed for a multitude of different traits, including gender [1, 2], race [1–3], academic achievements [2, 4, 5], genotypes [6], aggression [7], obesity [8], happiness [9], divorce [10], smoking [11], or sexual orientation [12]. Homophily is found for different types of relationships such as between spouses [13], friends [14] and co-workers [15], and occurs in a wide range of environments including kindergarten [16], large human gatherings [17], Wall Street

[18], populations of hunter-gatherers [19], or virtual societies of online gamers [20, 21]. Homophily is considered as one of the fundamental organizational principles of human societies [22], and has a number of important social implications such as the origin of segregation [23] or the perpetuation of economic inequality and social immobility [24].

Even though there exists an extensive body of research on homophily, it remains a challenging question to understand its origins and how it forms and develops over time. For traits that can not be changed, such as race or gender, homophily arises through a re-structuring process of inter-human relationships, where on average links between people with similar traits are created, while links between dissimilar people are dissolved. If traits can be changed over time, the situation becomes more involved. In this case, there exist two mechanisms to explain the formation of homophily from an initially homogeneous population: *socialization* and *social selection* [25]. The mechanism of socialization means that people change their traits to increase similarity to those they are connected to in a static social environment (network). This is schematically shown in Fig 1(a). This mechanism is sometimes also referred to as 'social contagion' or 'peer influence'. Under the mechanism of social selection, individuals re-arrange their social ties so that they become linked to people that are similar in traits, see Fig 1(b). If both mechanisms are at work at the same time, traits and social networks are said to *co-evolve*. The literature on the aspect of co-evolution is limited because of the lack of the simultaneous availability of longitudinal social networks and traits data.

In this paper we focus on homophily in academic achievements. It might be expected that academic performance plays an important role in friendship sorting. Homophily in achievements may arise from several factors, including school organization [26] and school policies such as ability-grouping [27] that increases the probability to meet students with similar academic performance. However, such policies are not common in the Russian education system that is characterized by its egalitarian nature and its high level of standardization. There is no tracking or ability-grouping in the particular school we use in this study. Students from that school are from the same neighborhood, the majority of them lives within 10 minutes walk from the school (see S1 Text).

Adolescence is the period of major social changes in the lives of young people [28]. Many of these changes lie in the area of peer relationships [29]. Starting from early adolescence young people become much less involved with their parents and more with their peers [30]. In this period, adolescents are concerned about their popularity among friends and seek peer acceptance [31]. To no surprise, peer influence extends to many areas of their life, including academic [32]. It was previously shown that peers may influence school engagement [33], disruptive behaviour in the classroom [34], and academic performance [35].

However, similarities between friends are not simply explained by peer influence. Adolescents also choose friends with similar behaviours and attitudes [36]. One major feature of adolescence is formation of a personal identity [37], it means that young people begin to explore and examine psychological characteristics of the self in order to discover who they really are, and how they fit in the social world in which they live [38]. Peer-group membership is also a part of identity as the group of friends to which adolescents belong helps define who they are [39]. We might expect, for example, that high performing students seek friendship with other high performing students as part of their academic identity formation. There is evidence that the importance of pears peaks in the early adolescence and then gradually declines when young people develop a mature sense of autonomy [31].

Homophily in academic performance has been studied with network data collected with questionnaire-based surveys [4, 5]. The design of these studies makes it hard to follow the temporal evolution of social networks. The availability of new technologies and big datasets provides researchers with novel tools to observe the dynamics of social networks with high

Socialization/adaptation

Trait change

a) t   t+1

Social selection

Link change

b) t   t+1
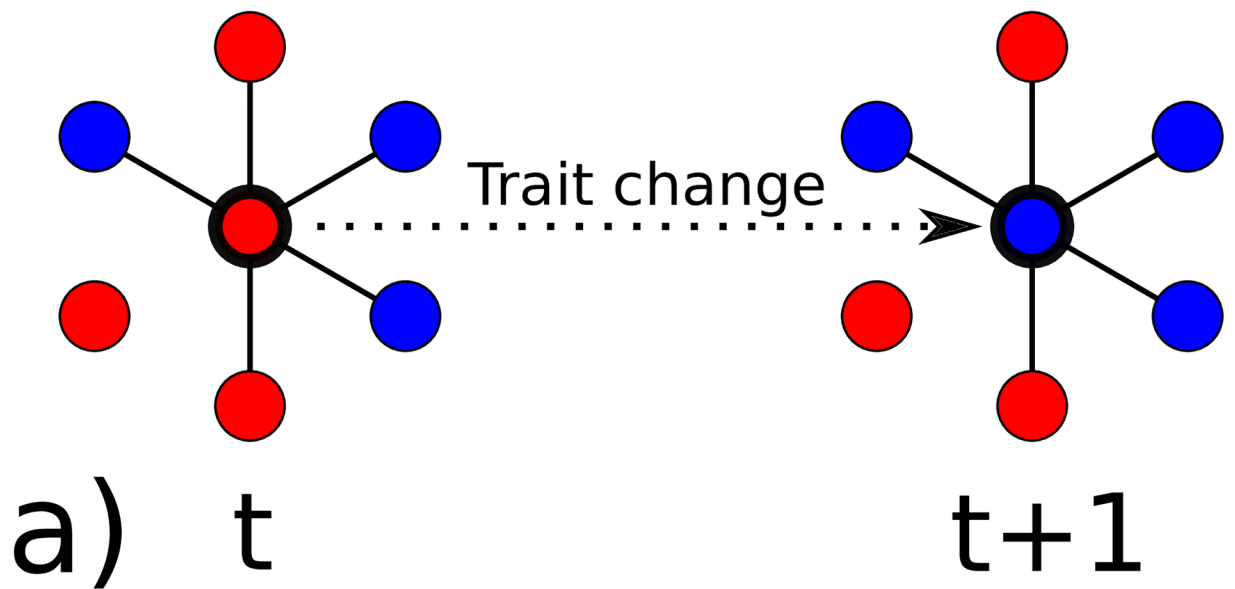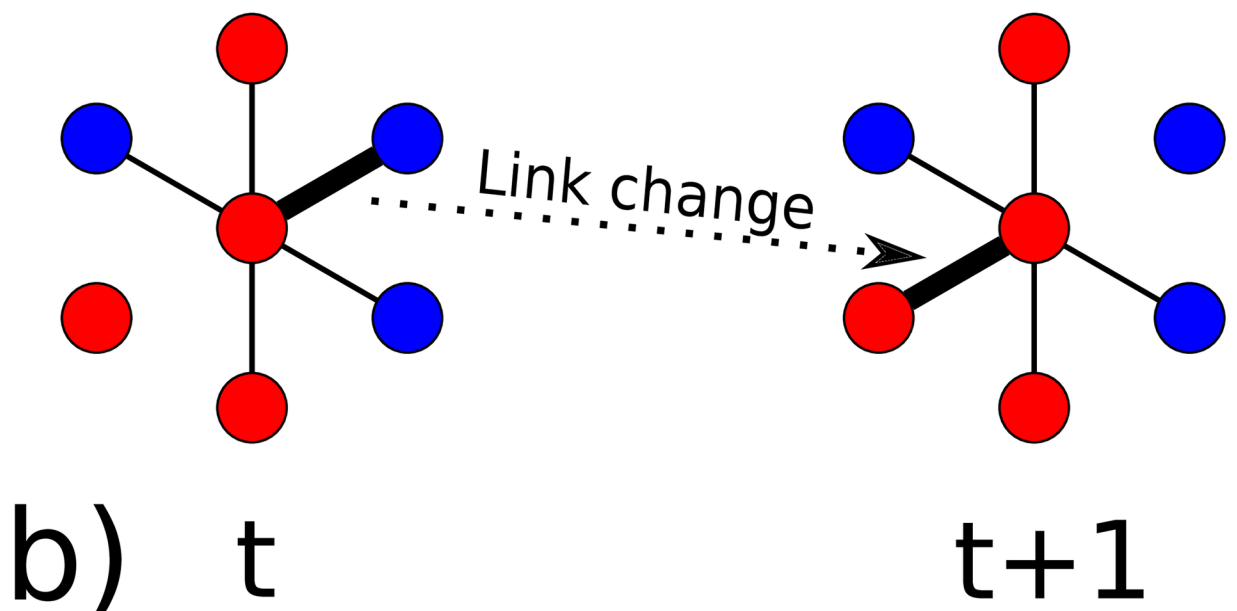
Fig 1. Two basic mechanisms to understand the origin of homophily. Nodes represent individuals, different colors indicate different traits. Links correspond to social ties, e.g. friendship. Increase of similarity between connected individuals may either arise from changes in traits (socialization process) (a), where individuals change their trait according to the dominance of that trait in their local social networks, or through re-wiring of their local social networks (social selection process) (b), where individuals re-shape their social contacts such that their trait matches those in the new environment better.

temporal precision. For example, the temporal structure of social networks has been reconstructed by email data [40], computer game logs [41], or interactions on learning management system platforms [42].

The quantification of social ties remains a challenging task [43]. Even traditional approaches that are based on self-reported friendship ties may contradict the common definition of friendship. For example, it was shown that only half of self-reported friendship links was reciprocal (despite the fact that almost all of them were perceived as reciprocal) [44]. Alternatively, friendship links may be inferred from digital records of human behaviour, allowing to track the detailed evolution of social ties [45], including those among high school students [46]. We approximate friendship links between students from their activities on the social networks site, in particular from the placement of "likes" on other students' pages. Some classical sociological theory suggests that social relationships are maintained through the symbolic exchanges [47], where exchange of "likes" may be considered as such a "symbolic ritual" [48]. Indeed, there is direct empirical evidence that "likes" correlate with real friendship ties [49].

In this paper, we use a unique anonymized dataset to observe the temporal formation of academic homophily based on social interactions between Russian students from a public high school and a university. The dataset contains information on the students' academic performance at several time points during their studies together with the detailed information about the evolution of their friendship networks (see S1 Text). These networks between students were obtained from the largest European social network site VK (http://vk.com), that provides a functionality similar to Facebook. VK users create their profiles with information about their identity, education, interests, etc. The use of the real name is required by VK. Users may indicate other users as their friends. VK friendship is mutual and requires confirmation. However, using VK friendship links is not the most efficient way to study the dynamical evolution of actual friendships, since only information about the current friendship is available, which makes it practically impossible to extract the dynamics of VK friendship links. It is also impossible to distinguish active friendship ties from obsolete ones since VK friendship links are rarely dissolved. We, therefore, approximate friendship links between students from the placement of "likes" on other students' pages. A link from one student to another is created if a "like" was placed at least once within a given period of time (see S1 Text). This approximation of actual friendships by social interaction strengths allows us to track the effective network evolution between students with much higher precision (see S1 Fig).

The network of university students (seniors) on March 2016 is shown in Fig 2. Previous studies on the Facebook (or its Russian analog VK) have focused on the (relatively static) friendship marking options that are provided by the sites [50–52]. This dataset not only allows us to quantify the extent of academic homophily among students but also to see its detailed evolution over time. In particular, we are able to clarify the mechanism behind the emergence of academic homophily from an initially homogeneous population across several years.

We use two datasets of academic performance records measured as grade point averages (GPA), one with 655 students from the 5th to 11th grades (age from 11 to 18) of a Russian public high school in Moscow (for reasons of anonymity we do not state the name of the school), the other with 5,925 bachelor students of the Higher School of Economics in Moscow. High school students receive their grades at the end of each trimester, their GPAs for the last 5 trimesters were available. Since the academic year of 2014/15 the Higher School of Economics started to publish a public ranking of its students. It contains information about their GPAs for the current semester along with the aggregated average GPA across the whole period of their studies. We collect the temporal GPA data into a vector $G_i^{\mathrm{HS/U}}(t)$ that represents the GPA
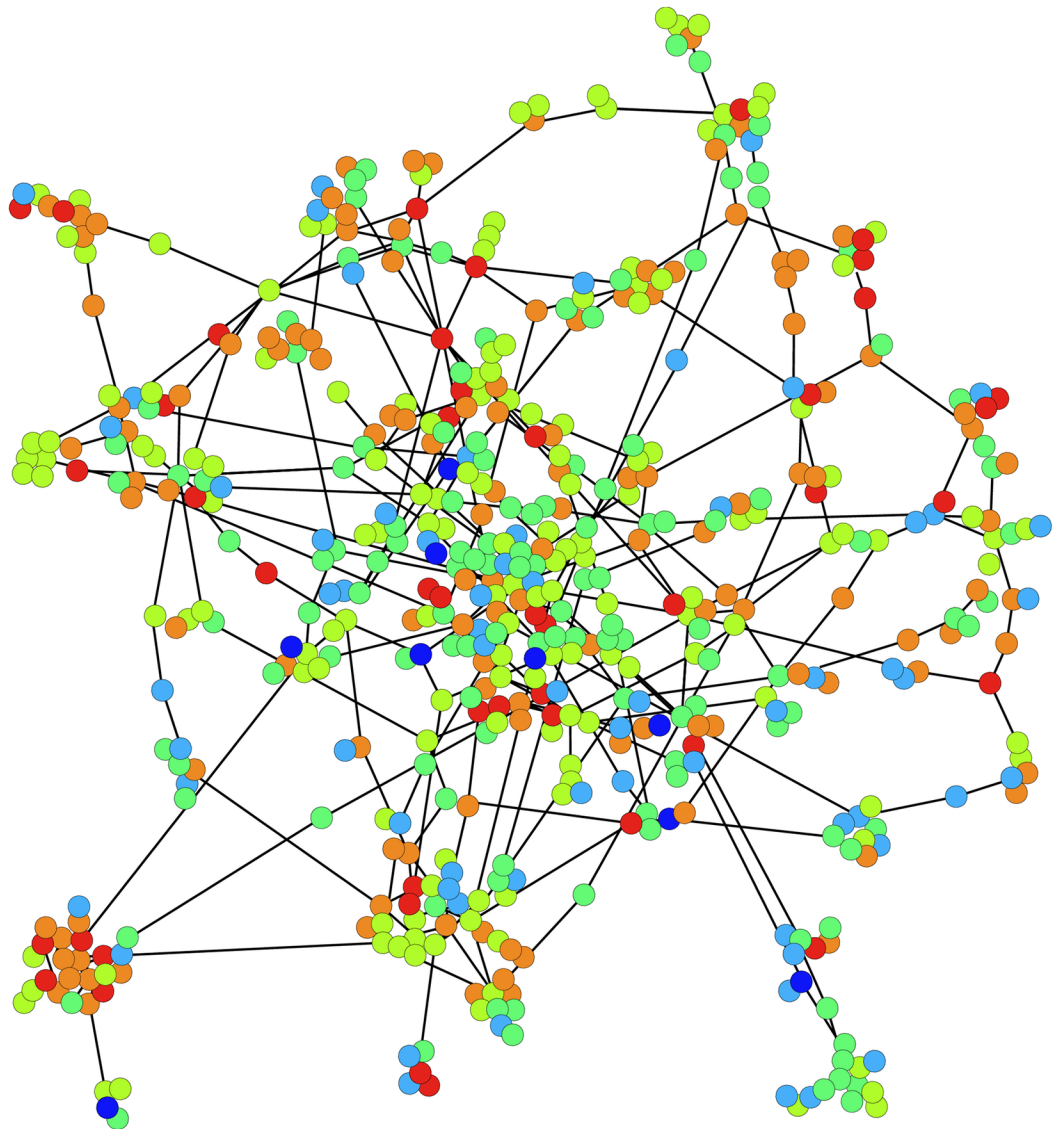
GPA

Fig 2. Snapshot of the friendship network of university students. The network is reconstructed from students' interactions on the social network site VK, the Russian variant of Facebook. Nodes represent students, links exist if one student gave a "like" to another at least once in March 2016. Color represent the performance (GPA) of students across the whole period of studies. There is visible clustering of students with similar GPA.

of student i at time t and corresponds to a student's performance within the time period from $t - 1$ to time t. $t = 1$ indicates the end of the first trimester/semester, $t = T$ is the end of the last trimester/semester for a given group of students. HS indicates "high school", U is "university". For detailed information about time points corresponding to GPAs collection see S6 Fig. Note that grades are different for high school and university. For high school grades range from 2 (worst) to 5 (best), for university from 4 (worst) to 10 (best). The average GPA of a student across the entire available time period we denote by $\bar{G}_i^{\text{HS/U}}$. For university students we follow 4 cohorts that are labeled by X in the following way: $G_i^{\text{U,X}}$, where $X = 1, 2, 3, 4$ stands for freshmen, sophomores, juniors, and seniors, respectively. The average GPA for high school students and the cohorts of university students are presented in S1 Table.

To generate a proxy for the temporal friendship interaction network between students we use the popular SNS VK, whose main component is a user-generated news feed. This feed contains all content that was generated (posted) by users and is generally visible to friends only. If users like the content that was posted by their friends they can indicate this by an instant feedback called a "like". "Likes" may mean different things to different people [53], however, "likes" can, in general, be seen as an indication of *active* friendship contacts between users.

VK provides an application programming interface (API) that allows to download information systematically in an open JSON format. In particular, it is possible to download user profiles from particular educational institutions and within selected age ranges. For each user, it is possible to obtain the list of their friends and the content that was published by them along with the VK identifiers of users that liked this content. Posting times are known with a time resolution of one second. "Likes" for specific content are almost always placed within 1-2 days after the content was posted. Using specially developed software the profiles of students of a given institution were downloaded and automatically matched by their first and last names with the available data on students' performance. 88% of all high school students and 95% of university students could be identified on VK (see S1 Text). The matching procedure was performed by authorized representatives of the high school and the Higher School of Economics, respectively. After the matching procedure, all names and VK identifiers were irrevocably deleted. The "likes" of all users were collected with corresponding timestamps, those from users outside the educational institutions were removed. "Likes" were then aggregated to intervals of 3 months periods. For each group of students, we obtain a $N \times N$ adjacency matrix $A(t)$, where $A_{ij}(t) = 1$ if student i places at least one "like" to student j from time $t - 1$ to time t. For detailed information about time periods corresponding to collected network data see S6 Fig. The subsequent deletion of all information on individual "likes" and respective timestamps prevents the possibility of any de-anonymization. The resulting datasets were transferred to the Institute of Education, which made it available for research in fully anonymized form.

Results

We first demonstrate the existence of academic homophily and then try to understand its origin. For all groups of students we find strong homophily. To make it comparable with other homophily studies such as in [14], we use a standard way of quantifying it by the conditional probability increase, I_X that a student belongs to top Xth percentile of performers, given that his/her friends also belong to the same percentile (see S1 Text). $I_X(t) = 0$ means that grades and friendship network are uncorrelated, $I_X(t) = 100\%$ means that the probability to be in the top Xth percentile is 2 times higher if the student's friends are also in the top Xth percentile, compared to the situation when they are not. $I_X(t)$ can not only be computed for friends (social distance 1) but also for friends of friends (social distance 2), and friends of friends of friends (social distance 3), etc. In Fig 3 we fix X to be the 50th (above average students) (a) and (b) and 80th (excellent students) percentile (c) and (d), respectively. For social distances up to 2 we observe significant homophily for all student groups at the last time point $T = 6$ for high school, and $T = 14$ for university. We find $I_{50\%}^{\text{HS}}(6) = 23\%$, $I_{80\%}^{\text{HS}}(6) = 57\%$, $I_{50\%}^{\text{U,4}}(14) = 30\%$, $I_{80\%}^{\text{U,4}}(14) = 49\%$, p-value $< 10^{-4}$. Significance was tested with a permutation test (10,000 permutations), see Methods. Note that the corresponding values at the first time point are smaller, $I_{50\%}^{\text{HS}}(1) = 22\%$, $I_{80\%}^{\text{HS}}(1) = 34\%$, $I_{50\%}^{\text{U,4}}(1) = 16\%$, $I_{80\%}^{\text{U,4}}(1) = 28\%$.

This result holds independent from the method used. Following an alternative approach for scalar variables we compute the *assortativity coefficient r* [54] and again find highly significant
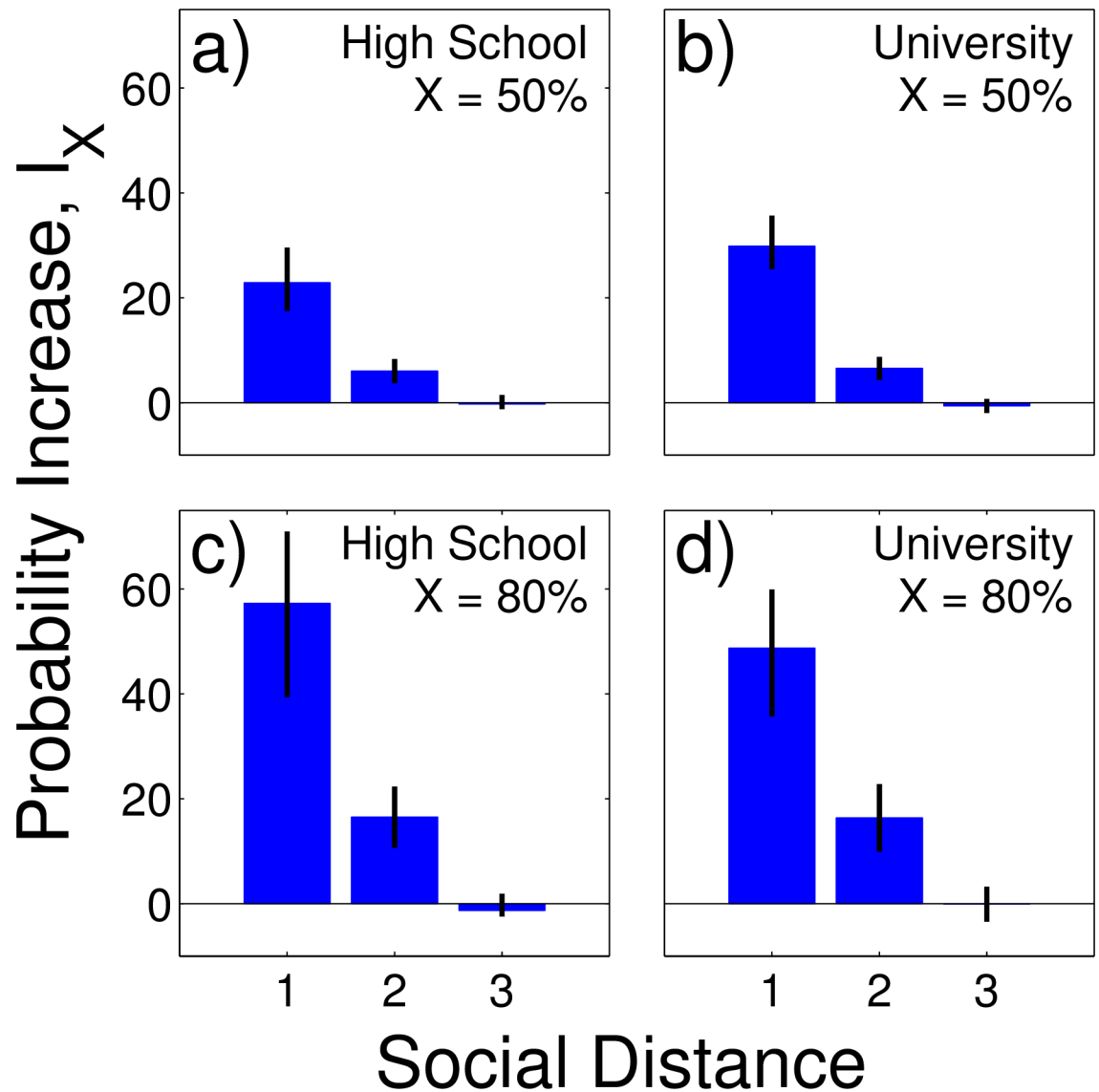
Fig 3. Homophily of students with good (a) and (b), and excellent grades (c) and (d), as a function of social distance. Observed increase in probability I_X that a student is in the top Xth percentile of students, given that their friends are also in the top Xth percentile. Results for the high school are shown in (a) and (c), for university in (b) and (d). Vertical lines indicate 95% confidence intervals computed with the permutation test. The social distance of 1 means friends, the social distance of 2 means friends of friends and the social distance of 3 means friends of friends of friends.

homophily at the last time point $r^{HS}(6) = 0.20$ (p-value $< 10^{-4}$) and $r^{U,4}(14) = 0.21$ (p-value $< 10^{-4}$). At the first time point homophily is much smaller, $r^{HS}(1) = 0.12$ and $r^{U,4}(1) = 0.12$.

In Fig 4 we show the time evolution of homophily over 1.5 years for high school students (a) and over 3.5 years for university students (b). We employ a transparent definition of a *Homophily Index*, H,(see Methods). We see a clear increase of H from the first to the last trimester from about $H = 0.20$ to $H = 0.41$ for the high school (a) (circles), and from $H = 0.24$ to $H = 0.40$ for university (b) (crosses). We next show in a series of three arguments that the increase in homophily over time can not be explained by the socialization/adaptation mechanism, i.e. by the changes in GPAs over time.
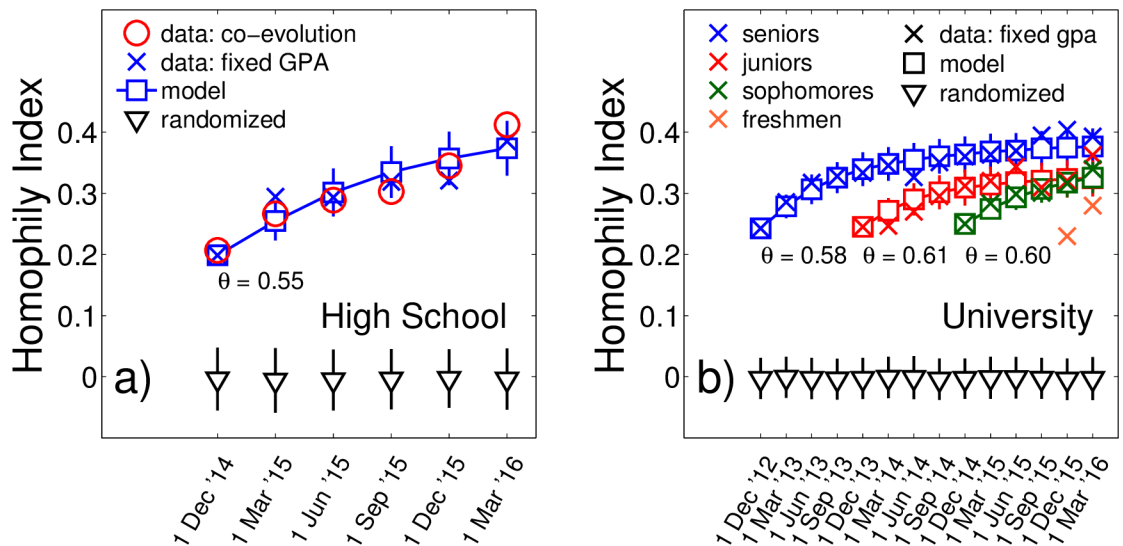
Fig 4. Evolution of homophily (Homophily index) in friendship networks of high school (a) and university students (b). Homophily increases with time by almost a factor of 2 (circles). The significance of the observed effect is measured with a randomization test (triangles), where grades were reshuffled randomly between the nodes in the network. It is amazing that when the GPAs of individual students are fixed to their temporal average (crosses), practically the same increase of homophily is observed, which signals the dominance of network restructuring. Results can be understood with a simple model (squares). Vertical bars are standard deviations.

https://doi.org/10.1371/journal.pone.0183473.g004

Ruling out socialization/adaptation

1. The first argument why socialization/adaptation is not the relevant mechanism behind the observed homophily increase, is due to the fact that academic performance is known to be a relatively persistent feature of students. It was shown that school-entry academic skills have large predictive power for later academic performance [55], and that academic performance might be heritable [56]. We find the persistence of performance in our data. The average GPA over high school students (3.85 ± 0.55) and its variance do practically not change over time, see S2 Fig. The average absolute difference between two consecutive time points $\langle |G_i(t) - G_i(t-1)| \rangle_i$ is 0.130, which means that the variation between GPAs of the same student at different time points is much smaller than the variation across students. Similar results are observed for the university students, with an average GPA of 7.41 ± 1.03 and a mean absolute difference 0.40.

2. The second argument why the socialization/adaptation mechanism can be ruled out is due to the observation that if we fix the GPAs for the high school students and do not let evolve them over time (we use the average GPA over all trimesters $\bar{G}_i$), we observe practically the same homophily increase as for the co-evolving GPAs, Fig 4(a).

3. Finally, we use a regression model to explain the GPA of students $G_i(t)$ by the explanatory variables: GPA at the previous trimester/semester $G_i(t-1)$, by the influence of friends' GPAs, by gender and by age (see S1 Text). The results are presented in S2 Table. For high school and university alike we find that the coefficients for $G_i(t-1)$ (α_1) and gender (γ) are significant and the coefficient for friends' GPA (α_2) is not. Again, this suggests that GPAs are rather stable over time and are almost fully determined by the GPA at the previous time point. The regression shows no evidence for an adaptation effect.

Social selection and network re-organization

Due to the second argument above the explanation of the observed homophily increase can only come through changes in social networks over time, i.e. the social selection mechanism, where students preferentially select new friends that are similar in performance. A simple model allows us to understand the situation. It assumes that whenever students select new friends they prefer students who are more similar to them than their current friends. Every student i is endowed with a fixed GPA $\bar{G}_i$ (constant). There exists an initial friendship network that we initialize with the observed network at timestep 1, $A_{ij}^{\text{model}}(1) = A_{ij}(1)$. From time t to $t + 1$ the model runs through the following steps

- Pick a student i at random,

- Pick a random friend j of i, $(A_{ij}^{\text{model}}(t) = 1)$,

- Pick a random potential new friend k $(A_{ik}^{\text{model}}(t) = 0)$,

- If k is closer to i than j, i.e. if $|G_i - G_k| \leq |G_i - G_j|$, rewire the link from ij to ik. Otherwise, rewire the link from ij to ik anyhow, with probability θ,

- Repeat until all students are updated, then continue with next timestep until $t = T$.

Clearly, if $\theta = 0$, rewiring happens only if a potential friend is closer in GPA than a current one (strict homophily increase); if $\theta = 1$ we have pure random rewiring. For a fixed θ we compute the Homophily Index $H^{\text{model}}(t)$ based on model networks. θ is fitted from the data such that $\sum_{t=1}^{T} \left(H^{\text{model}}(t) - H(t) \right)^2$ is minimized.

   We find θ values within the range of 0.55 and 0.61 for all groups. This means that students choose new friends among those who are similar about 64%-81% more often than among those who are not similar. The results of the model are presented in Fig 4 (boxes). The experimental homophily increase is recovered. Remarkably, for all student groups, the model is able to reproduce even details in the empirical GPA distances between stable, discontinued, and new friendships, see S3 Table.

   We have to show that the homophily increase is not explained as a trivial consequence of network densification. In both datasets we observe that friendship networks are dynamically changing over time. In Fig 5 the relative change of the average degree and the clustering coefficient of the networks are shown in comparison with the relative change of homophily. To see that the observed homophily increase is not a trivial consequence of network densification, observe that while for the high school degree and clustering increase, for university (seniors) they decrease. In both cases homophily increases. This is a first indication that degree and clustering are not the drivers of homophily change. As a second indication we test if H and I_X are significant with respect to a permutation test that preserves network topology. This is indeed the case (see Methods). Thirdly, by re-defining time intervals in a way that for each time interval the average degree is approximately the same, we find the same homophily increase (see S3 Fig), indicating that the degree is not an explanatory variable.

   Finally, in S4 Fig we show that there exist slight gender differences in the homophily increase. While both genders show about the same increase over time, the homophily index H is slightly larger for females in the sophomore and senior groups, and larger for males for the high school students and juniors. However, the noise in our data is too large to confirm that homophily indeed peaks in early adolescence, as seen in [57].
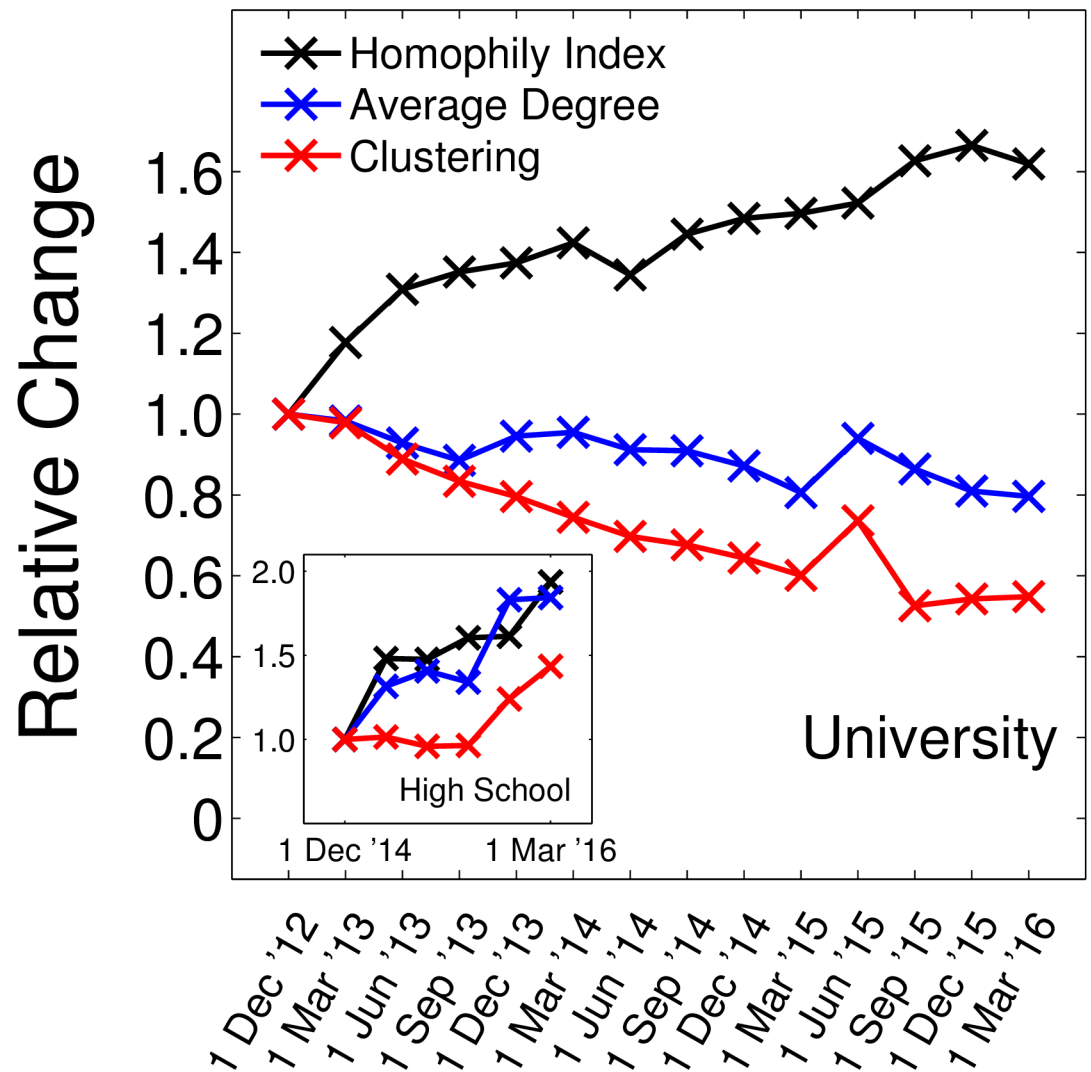
Fig 5. The network properties degree and clustering change over time (relative changes are shown, first time point is 1). While the network of seniors becomes sparser, there is a densification of the high school network (inset). Therefore degree and clustering coefficients can not be the drivers behind the observed homophily increase in both groups.

https://doi.org/10.1371/journal.pone.0183473.g005

Discussion

We studied a unique dataset containing the academic performance of high school and university students together with detailed information about the evolution of their social ties. In accordance with previous research [2, 4, 5, 50] we found strong homophily in academic performance. The strength of academic homophily is found to be stronger than for homophily in sexual activity [58] or alcohol abuse among adolescents [59] but weaker than for homophily in smoking marijuana [59], or for age [54].

We are not only able to demonstrate the strong homophily in academic performance but also to monitor how it emerges from a homogenous population and how it solidifies over time. We show that the observed gradual homophily increase can be explained predominantly by the process of social selection, meaning that students re-arrange their local social networks

to form ties and clusters of individuals that have similar performance levels. We could exclude the alternative explanations of social adaptation and co-evolution of social ties and performance. With a series of tests we ruled out the possibility that the increase of homophily results from adapting their academic performance to the one by their close friends. As an important consequence, this means that there are no indications for a pull effect, where groups of friends with good grades stimulate poor performing friends to increase their performance. The opposite effect of a negative group influence on students is also not found. It can be concluded that academic homophily in the studied groups arises and strengthens almost entirely through network re-linking.

We are able to understand the social-selection based homophily increase with a simple dynamical one-parameter model. The estimate of the parameter from the data means that students choose a new friend among those who are similar to them 64%-81% more often than dissimilar ones.

Note that even though this model is much simpler than others previously used [60], remarkably it is able to recover the increase over the whole time period for all groups, and even allows to understand details of the dynamics. It would be interesting to see in further work if these findings hold more generally also true for other student groups with different social contexts and in different countries.

It is important to note that the observed changes in social ties might be driven or facilitated by various factors. In the absence of ability tracking, other institutional factors may play a role in the segregation by academic achievements. For example extracurricular activities may provide an additional opportunity for similar individuals to meet and to from friendship ties [61]. Future research is needed to clarify the role of such specific factors.

Our findings might shed light or even confirm that access does not necessarily lead to equity. We find indications that physical mixing of students in the same educational institution does not lead to a homogeneous mixing of social ties. Even if the initial distribution is rather homogenous, students constantly re-organize their social network during the studies, which eventually results in segregation by academic performance. It is possible to conjecture that this mechanism is potentially reinforced by the accessibility of modern information technologies where maintaining links does not require physical presence anymore.

Academic achievements are the result of various factors, ranging from innate abilities to teacher qualification and family background. Regardless of these factors, it is the achievements that have direct implications for the students' future success. For example, Russian universities select students solely on the basis of their final school examination. Thus, academic achievements determine which students are selected for elite universities and which are not. As social networks play a crucial role in social mobility [62, 63], a selective university may provide a unique opportunity to create ties that will benefit students in the future. However, if initially low-performing students from a disadvantaged background predominantly create ties with other lower-performing students it significantly reduces their upward social mobility and may explain the persistence of inequality in societies.

Methods

Homophily index

We introduce a Homophily Index, H, as the Pearson correlation coefficient between the vector of students' GPAs, $G_i(t)$, and the vector of the average of the GPAs of their direct friends,

$$H(t) = \mathrm{corr}\left(G_i(t), \frac{\sum_j A_{ij}(t)G_j(t)}{\sum_j A_{ij}(t)}\right).$$ (1)

If students' grades are independent from average grades of their friends than $H(t) = 0$. Positive $H(t)$ means that better average grades of friends lead to better average grades of students and negative $H(t)$ means that better average grades of friends lead to worse students' performance. $H(t) = 1$ means a linear relation between students' performance and average performance of their friends.

Randomization test

One of the challenges in understanding correlations of traits between connected individuals is to test if the observed homophily effect is significant or if it results trivially from the topology of the underlying network. To test for this we employ a typical permutation test, see e.g. [14], where we preserve the network topology and randomly reshuffle the assignment of the GPAs to the node. We repeat this procedure 10,000 times to obtain a distribution of the measures H and I_X. We can then test the null hypothesis that GPAs are independent of network topology, and to compute corresponding p-values.

Supporting information

S1 Text.
(TEX)

S1 Fig. The average number of interaction ("likes") per day between university students is presented for each cohort. The maximum observed value is 200 or 0.13 "likes" per day per student. The steep increase in September marks the beginning of studies. Some students knew each other before the matriculation.
(TIF)

S2 Fig. Average GPA for high school and university students (inset) over time. Results are shown as mean values ± standard deviations. Females have better grades on average. GPAs and their variance do practically not change with time.
(TIF)

S3 Fig. In the high school data the network is getting more connected over time. It is therefore possible to re-define new time intervals in such a way that for each time interval the average degree in the network is approximately the same. Clearly the homophily index H increases as before, indicating that the degree is not an explanatory variable. The same argument holds for the clustering coefficient.
(TIF)

S4 Fig. There are no consistent differences in gender. While both genders show about the same increase over time, it is larger for females in the sophomore and senior groups, and larger for males for the high school students and juniors.
(TIF)

S5 Fig. Homophily increase varies from subject to subject. Since there are only 4 possible values of grades (scores) possible for the individual subjects, we expect to observe less stable results than for the GPA. However, the general pattern of homophily increase over time holds, for mathematics it is not much pronounced.
(TIF)

S6 Fig. Time schedule of data collection for university (a) and high school (b) students. Network data is in the form of adjacency matrices $A_{ij}(t)$, where $A_{ij}(t) = 1$ means that student i gave at least one "like" to student j from time $t - 1$ to time t. The time period from $t - 1$ to t is

equal to 3 months. (a) For the university students (seniors, juniors, sophomores) the aggregated average GPA, $\bar{G}_i^{\mathrm{U}}$, from the beginning of their studies on the 1st of September (2012/2013/2014) until the 1st of March, 2016 is collected. This period is equal to 3.5 years for seniors, 2.5 years for juniors and 1.5 years for sophomores respectively. The temporal GPA data, $G_i^{\mathrm{U}}(t)$, was also collected for the last 3 semesters for all 3 cohorts (arrows). (b) For the high school students the temporal GPA data, $G_i^{HS}(t)$, is collected at the end of each trimester for the last 5 trimesters (arrows). As students do not study in summer, we assume the same performance at that period as at the last available time point i.e. spring, $G_i^{HS}(3) = G_i^{HS}(4)$. $\bar{G}_i^{HS}$ is computed as the average over the 5 trimesters.
(TIF)

S1 Table. Descriptive statistics of students' GPA scores across the whole period of their studies. $\langle . \rangle_i$ means average over all students in the group. Mean values and standard deviations (in brackets) are presented. Females have better grades than males on average.
(PDF)

S2 Table. Coefficients from the regression model. The GPA at the current time point is almost fully explained by the GPA at the previous time point. The influence of gender is also significant, males have lower grades also after controlling for their previous GPA. The average GPA of friends at the previous time point is not significant.
(PDF)

S3 Table. Re-organization of the students' network over time. The GPA distance for new friends is consistently and significantly smaller (tested with two-sample Students' test) than the GPA distance for discontinued friends in the observed data. Comparable results are obtained with the model.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Ivan Smirnov, Stefan Thurner.

Data curation: Ivan Smirnov.

Formal analysis: Ivan Smirnov, Stefan Thurner.

Investigation: Ivan Smirnov, Stefan Thurner.

Methodology: Ivan Smirnov, Stefan Thurner.

Writing – original draft: Ivan Smirnov, Stefan Thurner.

Writing – review & editing: Ivan Smirnov, Stefan Thurner.

References

1. Shrum W, Cheek NH, Saundra MH. Friendship in School: Gender and Racial Homophily. Sociology of Education. 1988; 61(4):227–239. https://doi.org/10.2307/2112441

2. Tuma NB, Hallinan MT. The effects of sex, race, and achievement on schoolchildren's friendships. Social Forces. 1979; 57(4):1265–1285. https://doi.org/10.1093/sf/57.4.1265

3. Currarini S, Jackson MO, Pin P. Identifying the roles of race-based choice and chance in high school friendship network formation. Proceedings of the National Academy of Sciences. 2010; 107(11): 4857–4861. https://doi.org/10.1073/pnas.0911793107

4. Flashman J. Academic achievement and its impact on friend dynamics. Sociology of Education. 2012; 85(1):61–80. https://doi.org/10.1177/0038040711417014 PMID: 25705057

5. Lomi A, Snijders TA, Steglich CE, Torló VJ. Why are some more peer than others? Evidence from a longitudinal study of social networks and individual academic performance. Social Science Research. 2011; 40(6):1506–1520. https://doi.org/10.1016/j.ssresearch.2011.06.010 PMID: 25641999

6. Fowler JH, Settle JE, Christakis NA. Correlated genotypes in friendship networks. Proceedings of the National Academy of Sciences. 2011; 108(5):1993–1997. https://doi.org/10.1073/pnas.1011687108

7. Espelage DL, Holt MK, Henkel RR. Examination of peer-group contextual effects on aggression during early adolescence. Child development. 2003; 74(1):205–220. https://doi.org/10.1111/1467-8624.00531 PMID: 12625446

8. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. New England journal of medicine. 2007; 357(4):370–379. https://doi.org/10.1056/NEJMsa066082 PMID: 17652652

9. Fowler JH, Christakis NA. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. BMJ. 2008; 337:2338. https://doi.org/10.1136/bmj.a2338

10. McDermott R, Fowler JH, Christakis NA. Breaking up is hard to do, unless everyone else is doing it too: Social network effects on divorce in a longitudinal sample. Social Forces. 2013; 92(2):491–519. https://doi.org/10.1093/sf/sot096

11. Ennett ST, Bauman KE. The contribution of influence and selection to adolescent peer group homogeneity: the case of adolescent cigarette smoking. Journal of personality and social psychology. 1994; 67(4):653–663. https://doi.org/10.1037/0022-3514.67.4.653 PMID: 7965611

12. Thelwall M. Homophily in myspace. Journal of the American Society for Information Science and Technology. 2009; 60(2):219–231. https://doi.org/10.1002/asi.20978

13. Blossfeld HP. Educational assortative marriage in comparative perspective. Annual review of sociology. 2009; 35:513–530. https://doi.org/10.1146/annurev-soc-070308-115913

14. Christakis NA, Fowler JH. Social contagion theory: examining dynamic social networks and human behavior. Statistics in medicine. 2013; 32(4):556–577. https://doi.org/10.1002/sim.5408 PMID: 22711416

15. Ibarra H. Homophily and differential returns: Sex differences in network structure and access in an advertising firm. Administrative science quarterly. 1992; 37(3):422–447. https://doi.org/10.2307/2393451

16. Hanish LD, Martin CL, Fabes RA, Leonard S, Herzog M. Exposure to externalizing peers in early childhood: Homophily and peer contagion processes. Journal of abnormal child psychology. 2005; 33(3): 267–281. https://doi.org/10.1007/s10802-005-3564-6 PMID: 15957556

17. Barnett I, Khanna T, Onnela JP. Social and Spatial Clustering of People at Humanity's Largest Gathering. PloS One. 2016; 11(6):1–12. https://doi.org/10.1371/journal.pone.0156794

18. Roth LM. The social psychology of tokenism: Status and homophily processes on Wall Street. Sociological Perspectives. 2004; 47(2):189–214. https://doi.org/10.1525/sop.2004.47.2.189

19. Apicella CL, Marlowe FW, Fowler JH, Christakis NA. Social networks and cooperation in hunter-gatherers. Nature. 2012; 481(7382):497–501. https://doi.org/10.1038/nature10736 PMID: 22281599

20. Szell M, Lambiotte R, Thurner S. Multirelational organization of large-scale social networks in an online world. Proceedings of the National Academy of Sciences. 2010; 107(31):13636–13641. https://doi.org/10.1073/pnas.1004008107

21. Szell M, Thurner S. How women organize social networks different from men. Scientific reports. 2013; 3:1214. https://doi.org/10.1038/srep01214 PMID: 23393616

22. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: Homophily in social networks. Annual review of sociology. 2001; 37:415–444. https://doi.org/10.1146/annurev.soc.27.1.415

23. Currarini S, Jackson MO, Pin P. An economic model of friendship: Homophily, minorities, and segregation. Econometrica. 2009; 77(4):1003–1045. https://doi.org/10.3982/ECTA7528

24. DiMaggio P, Garip F. Network effects and social inequality. Annual Review of Sociology. 2012; 38: 93–118. https://doi.org/10.1146/annurev.soc.012809.102545

25. Kandel DB. Homophily, selection, and socialization in adolescent friendships. American journal of Sociology. 1978; p. 427–436. https://doi.org/10.1086/226792

26. Karweit N, Hansell S. School organization and friendship selection. In: Friends in school: Patterns of selection and influence in secondary schools. Elsevier; 1983. p. 29–38.

27. Hallinan MT, Sørensen AB. Ability grouping and student friendships. American Educational Research Journal. 1985; 22(4):485–499. https://doi.org/10.3102/00028312022004485

28. Blakemore SJ, Mills KL. Is adolescence a sensitive period for sociocultural processing?. Annual review of psychology. 2014; 65:187–207. https://doi.org/10.1146/annurev-psych-010213-115202 PMID: 24016274

29. Brown BB, Larson J. Peer relationships in adolescence. In: Handbook of adolescent psychology. John Wiley & Sons; 2009. p. 74–103.

30. Larson R, Richards MH. Daily companionship in late childhood and early adolescence: changing developmental contexts. Child Development. 1992; 62(2):284–300 https://doi.org/10.2307/1131003

31. Fuligni AJ, Eccles JS, Barber BL, Clements P. Early adolescent peer orientation and adjustment during high school. Developmental psychology. 2001: 37(1):28–36. https://doi.org/10.1037/0012-1649.37.1.28 PMID: 11206430

32. Brechwald WA, Prinstein MJ. Beyond homophily: A decade of advances in understanding peer influence processes. Journal of Research on Adolescence. 2011; 21(1):166–179. https://doi.org/10.1111/j.1532-7795.2010.00721.x PMID: 23730122

33. Simons-Morton B, Chen R. Peer and parent influences on school engagement among early adolescents. Youth & society. 2009: 41(1):3–25. https://doi.org/10.1177/0044118X09334861

34. Berndt TJ, Keefe K. Friends' influence on adolescents' adjustment to school. Child development. 1995; 66(5):1312–1329. https://doi.org/10.2307/1131649 PMID: 7555218

35. Burke MA, Sass TR. Classroom peer effects and student achievement. Journal of Labor Economics. 2013; 31(1):51–82. https://doi.org/10.1086/666653

36. Akers JF, Jones RM, Coyl DD. Adolescent friendship pairs: Similarities in identity status development, behaviors, attitudes, and intentions. Journal of Adolescent Research. 1998: 13(2):178–201. https://doi.org/10.1177/0743554898132005

37. Erikson E. Identity, Youth, and Crisis. New York: Norton; 1968

38. Steinberg L, Silverberg SB. The vicissitudes of autonomy in early adolescence. Child development. 1986; 57(4):841–851. https://doi.org/10.2307/1130361 PMID: 3757604

39. Sroufe LA, Cooper RG, DeHart GB, Marshall ME, Bronfenbrenner UE. Child development: Its nature and course. Mcgraw-Hill Book Company; 1992

40. Kossinets G, Watts DJ. Empirical analysis of an evolving social network. Science. 2006; 311(5757): 88–90. https://doi.org/10.1126/science.1116869 PMID: 16400149

41. Szell M, Thurner S. Measuring social dynamics in a massive multiplayer online game. Social networks. 2010; 32(4):313–329. https://doi.org/10.1016/j.socnet.2010.06.001

42. Vaquero LM, Cebrian M. The rich club phenomenon in the classroom. Scientific reports. 2013; 3:1174. https://doi.org/10.1038/srep01174 PMID: 23378908

43. Wuchty S. What is a social tie? Proceedings of the National Academy of Sciences. 2009; 106(36): 15099–15100. https://doi.org/10.1073/pnas.0907905106

44. Almaatouq A, Radaelli L, Pentland A, Shmueli E. Are you your friends' friend? Poor perception of friendship ties limits the ability to promote behavioral change. PloS One. 2016; 11(3):e0151588. https://doi.org/10.1371/journal.pone.0151588 PMID: 27002530

45. Eagle N, Pentland AS, Lazer D. Inferring friendship network structure by using mobile phone data. Proceedings of the National Academy of Sciences. 2009; 106(36): 15274–15278. https://doi.org/10.1073/pnas.0900282106

46. Mastrandrea R, Fournet J, Barrat A. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. PloS One. 2015; 10(9): e0136497. https://doi.org/10.1371/journal.pone.0136497 PMID: 26325289

47. Bourdieu P. The forms of capital. In Cultural theory: An anthology; 2011. p. 81–93.

48. Gamage KJ. Like it? Ritual Symbolic Exchange Using Facebook's Like Tool. 2012

49. Jones JJ, Settle JE, Bond RM, Fariss CJ, Marlow C, Fowler JH. Inferring tie strength from online directed behavior. PloS One. 2013; 8(1):e52168. https://doi.org/10.1371/journal.pone.0052168 PMID: 23300964

50. Mayer A, Puller SL. The old boy (and girl) network: Social network formation on university campuses. Journal of public economics. 2008; 92(1):329–347. https://doi.org/10.1016/j.jpubeco.2007.09.001

51. Lewis K, Gonzalez M, Kaufman J. Social selection and peer influence in an online social network. Proceedings of the National Academy of Sciences. 2012; 109(1):68–72. https://doi.org/10.1073/pnas.1109739109

52. Dokuka S, Valeeva D, Yudkevich M. Formation and Evolution Mechanisms in Online Network of Students: The Vkontakte Case. In: International Conference on Analysis of Images, Social Networks and Texts. Springer; 2015. p. 263–274.

53. Scissors L, Burke M, Wengrovitz S. What's in a Like?: Attitudes and behaviors around receiving Likes on Facebook. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. ACM; 2016. p. 1501–1510.

54. Newman M. Networks: an introduction. Oxford university press; 2010.

55. Duncan GJ, Dowsett CJ, Claessens A, Magnuson K, Huston AC, Klebanov P, et al. School readiness and later achievement. Developmental psychology. 2007; 43(6):1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428 PMID: 18020822

56. Krapohl E, Rimfeld K, Shakeshaft NG, Trzaskowski M, McMillan A, Pingault JB, et al. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. Proceedings of the National Academy of Sciences. 2014; 111(42):15273–15278. https://doi.org/10.1073/pnas.1408777111

57. Berndt TJ. The features and effects of friendship in early adolescence. Child development; 1982. p. 1447–1460. https://doi.org/10.2307/1130071

58. Brakefield TA, Mednick SC, Wilson HW, De Neve JE, Christakis NA, Fowler JH. Same-sex sexual attraction does not spread in adolescent social networks. Archives of sexual behavior. 2014; 43(2):335–344. https://doi.org/10.1007/s10508-013-0142-9 PMID: 23842784

59. Shakya HB, Christakis NA, Fowler JH. Parental influence on substance use in adolescent social networks. Archives of pediatrics & adolescent medicine. 2012; 166(12):1132–1139. https://doi.org/10.1001/archpediatrics.2012.1372

60. Snijders TA, Van de Bunt GG, Steglich CE. Introduction to stochastic actor-based models for network dynamics. Social networks. 2010; 32(1):44–60. https://doi.org/10.1016/j.socnet.2009.02.004

61. Karweit N. Extracurricular activities and friendship selection. In: Friends in school: Patterns of selection and influence in secondary schools. Elsevier; 1983. p. 131–139.

62. Granovetter MS. Getting a job: A study of contacts and careers. University of Chicago Press; 1995.

63. Podolny JM, Baron JN. Resources and relationships: Social networks and mobility in the workplace. American sociological review. 1997; 62(5):673–693. https://doi.org/10.2307/2657354