Academic Performance and Behavioral Patterns

Valentin Kassarnig1*, Enys Mones2, Andreas Bjerre-Nielsen3,4, Piotr Sapiezynski2,5, David Dreyer Lassen3,4 and Sune Lehmann2,4,6

Abstract

Identifying the factors which influence academic performance is an essential part of educational research. Previous studies have documented the importance of personality traits, class attendance, and social network structure. However, because most of these analyses were based on a single behavioral aspect and/or small sample sizes, we lack a quantification of the interplay of these factors. Here, we infer the academic performance among a cohort of almost 1000 freshmen based on data collected through smartphones, which the students used as their primary phone for two years. The availability of multi-channel data from a single population allows us to directly compare the predictive power of individual and social characteristics. We find that a student’s performance is best inferred from their social ties. Network indicators out-perform models based on individual characteristics. We confirm earlier findings indicating that class attendance is the most important predictor among the individual characteristics. Finally, our results indicates potential presence of strong homophily and/or peer effects among university students.

Keywords: academic performance; data collection; homophily; peer effect

Introduction

Since research on academic achievement began to emerge as a field in the 1960s, it has guided educational policy makers on admissions and dropout prevention [1]. Although much of the literature has focused on higher education, the knowledge obtained on behavioral phenomena observed in colleges and universities can potentially guide research on student behavior in primary and secondary schools. A number of behavioral patterns have been linked to academic performance, such as time allocation [2], active social ties [3], sleep duration and sleep quality [4], or participation in sport activity [5]. Most of the existing studies, however, suffer from the typical biases and limitations associated with surveys and self-reports [6, 7]. Here we investigate student performance within a novel dataset collected as part of the Copenhagen Network Study (CNS), which involved approximately 1000 university students, with data collection ongoing for more than 2 years [8]. Due to the scale of the CNS, and the inclusion of observational data in place of self-reports we are able to mitigate some of the limitations encountered in existing ‘traditional’ studies. Using smartphones, we collected data from multiple channels: person-to-person proximity (using Bluetooth scans), calls and text messages, activity on online social networks (Facebook), and mobility traces.
Related Work

Cell phone based studies Through a variety of methods, a large number of studies have investigated the factors determining academic performance. Vandamme et al. [9] investigated three sets of characteristics: personal history, behavior, and perception. Similarly, the StudentLife study [10] used smartphones to collect data on student activity, social behavior, personality, and mental health. Both research groups observed correlations between performance and all feature categories, indicating that factors influencing academic performance are not limited to a single aspect of an individual's life. Nghe et al. [11] reframed the problem as a prediction task: using data to predict performance in a population of undergraduate and postgraduate students at two different institutions. Using a wide range of features, they predicted GPA after third year with high accuracy. One of the features included GPA after the second year, while in this work we show that it is possible to make accurate predictions without the knowledge of past performance.

Online Social Media Only a few prior studies have investigated the impact of social media activity on academic performance, despite the growing availability of such data and undisputed presence of these media in our daily lives. The majority of existing studies found a decrease in academic performance with increasing time spent on social media [12–19]. However, not all studies confirm this result. In some studies, time spent on social media was found to be unrelated to academic performance [20,21] or had even a positive effect on performance [22,23].

Online Social Interactions The relationship between social interactions and academic performance has gained growing importance [3, 24–42]. There are two dominant approaches found in the literature. The first approach focuses on the relation between own performance and that of peers [24–31], based on a hypothesis of similarity in achievements of peers. The similarity between pairs of individuals connected via social ties are attributed to various aspects: selection into friendships by similarity (i.e., homophily); influence by social peers (also know as peer effect); and correlated shocks (e.g., being exposed to the same teacher). As noted by [43] the issue of separating these effects are inherently difficult. The second approach emphasizes the positive influence having a central position in the contact network between students [35–40]. The results in the existing research, however, are based on self-reports and therefore subject to various biases that are eliminated by using smartphones to measure the social network [44].

Personality A large body of research at the intersection of psychology and education investigates the relationship between personality and performance, as pioneered by [45]. Finally, prior research has emphasized the positive influence of attending classes as shown in [46–49]. In particular, the study by Crede et al. [49] concludes that attendance is the most accurate known predictor of academic performance. See [50] for a more detailed analysis of the impact of class attendance on academic performance based on the CNS experiment.

Results

Our research strategy is as follows. We identified a number of individual and social indicators (features) of academic performance in the collected data. We then used
these features to train machine learning models designed to predict performance and help us understand the importance of particular features. Although all feature categories are correlated with academic performance, we find that features describing the social networks of students have the highest predictive power. Based on these results, we conduct an in-depth analysis of the most correlated impact factors of each category. In particular, we elaborate on the social behavioral features which have only been considered to a limited extent in previous studies.

Performance Predictions

We report the results in two stages. First, we divide the student population in quintiles based on performance and report the correlations between individual characteristics and quintile assignment. Second, we divide the population in tertiles: low, moderate, and high performers. This division into tertiles allows us to compare our results to previous work on performance prediction while quintiles provide a more detailed insight into the behavioral differences of the individual groups. To account for the different predictive power of the individual and network aspects, we construct four feature sets, each representing a certain aspect of life and corresponding to a specific level of information: personality (personality traits only), individual (all individual features), network (network metrics) and combined (full feature set). See Table 2 for a complete list of features in each category. Figure 1 shows the feature importance for features achieving significance of $p < .001$. Note that in general, network properties dominate the predictions with more than half of the significant features corresponding to this category. The fraction of low performing friends as well as the mean GPA of peers contacted over text messages and calls are the most important features. Class attendance proves to be the most important individual feature. Centrality in the communication networks (call and text messages) are also found to be significant descriptors with moderate importance. Among personality traits, only self-esteem and conscientiousness have significant predictive power.

For the evaluation we investigate the performance of four different classifier, namely Logistic Regression, Linear Discriminant Analysis (LDA), Support Vector Classifier, and Random Forest Classifier. The achieved predictive accuracies of every model are visualized in Fig. 2. Each data point is the result of a 10-fold cross-validation and is based on 30 independent measurements (see Materials and Methods for more details on the experimental setup).

In the following, we report the in-class precision and recall values only for the classification problem with the LDA model and the combined feature set, which achieved the best results. Fig. 3 shows these results along with the corresponding $F_1$ values. As the results indicate, once the GPA class is provided, the model has high predictive power among the low and high performers (compared to that of the moderate performers) with $F_1$ values of .649 and .646, respectively. A corresponding accuracy of 58.7% is achieved (significantly higher than that of the baseline of 33.33%).

Identifying students in risk of failing or dropping out at an early stage, is of particular interest from a practical standpoint. In a practical scenario, these are the students that we might want to offer extra help. Modeling the question of identifying these students as a binary classification allows us to identify 77.4% of the at-risk
students (bottom third) with a precision of 58.3%. This corresponds to a $F_1$ score of .666. On the other hand, non-risk students were correctly classified as such with a precision of 86.5% ($F_1 = .788$). These results underline the predictive power of the observed variables, in particular in the application of ascertaining at-risk students.

**Individual behavior**

Among individual effects we considered, class attendance is found to be the strongest predictor for academic performance. This observation is robust regardless of the measure used to quantify academic performance: correlation coefficients of $r_S = .255$ for course grades, $r_S = .255$ for term GPAs, and $r_S = .274$ for cumulative GPAs were determined (all values are significant with $p < .001$). An in-depth analysis of the observed class attendance patterns along with a detailed description of measuring attendance applied to the CNS dataset is discussed in [50].

Assuming that any posting activity on Facebook is positively correlated with the total time spent on the social network site, we quantify the time devoted to the social networking site by the number of posts a student performed during the semester. Figure 4 illustrates the results after dividing the population into quintiles according to their average number of posts per week. As Fig. 4a shows, the distribution of posts among students is heavy-tailed and is described by the vast majority of the students having less than 3 posts in a typical week. Distribution of term GPA values in the different quintiles reveals that, on average, students with lower activity perform better (see Fig. 4b). The last two groups (individuals with the highest number of average posts per week) show high similarity, whereas the first quintile differs from the other to the largest extent.

**Social interactions**

We find that a student’s performance can be accurately predicted from the achievements of their peers. We have quantified this effect using three measures: the mean GPA of friends and the fraction of low performers (bottom tertile) and high performers (top tertile) among the friends. Results from this analysis are displayed in Fig. 5, where we divide the population into deciles based on term GPA and calculated mean GPA of their contacts in various communication and interaction channels. First, regardless of the channel considered, the curves show a strong increasing trend. The general relationship between peers from each channel and performance is quantified in Table 1. Each channel displays higher correlations than any of the personality traits. The most pronounced effect is obtained with calls and text messages, which are considered a proxy for the strongest social ties because they require more effort to initiate and maintain [51].

We further assess the validity of pairwise similarity in the network by focusing exclusively on social ties based on text messages. Figure 6 shows a scatter plot of the correlation between the own GPA and mean GPA of the texting peers for every student in the dataset. Once again, we observe a clear linear trend; the trend is especially strong in the region where the majority of the students is located (GPAs in the range between 2 and 3). In Fig. 7 we divide the population in quintiles based on term GPA and calculate the fraction of text messages exchanged with members of the different groups. Beyond the correlation, we can see that students’
communication in each group is dominated by members of the same or neighboring
groups.

The structure of the interaction networks provide further insight into how students
position in their social environment is correlated with performance. To investigate
the network effects, we consider two descriptors here: number of contacts (degree)
and eigenvector centrality. Eigenvector centrality is related to flows in the social
network and individuals of high centrality are regarded as influencers in case of a
spreading dynamics (outbreak, rumors, etc.) [52]. Here we consider networks based
on various channels (person-to-person, calls, text messages, and Facebook friendship
or direct interaction). The correlations between cumulative GPA and the network
descriptors are summarized in Fig. 8. We find weak to moderate positive correlations
in most of the networks, in agreement with the existing literature [35–40], however,
proximity interactions display stronger values relative to rest of the interactions.
It should be noted that the proximity networks include all connections, even those
formed on campus, and during lectures, which may account for some of the observed
correlations.

Discussion

For the participants of the CNS study, we found that the most efficient predictor
for academic performance for an individual is the academic performance of her
contacts. We observed this effect across different channels of social interactions with
calls and text messages showing the strongest correlations, further emphasizing the
phenomena. As mentioned in the literature review, this effect could be caused by
multiple factors: peer effects, homophily or the impact from taking the same classes.

These finding contrast with previous conclusions from a meta-study [49], which
found class attendance to be the strongest predictor of academic achievements. In
our study, however, when network effects (such as social ties) are disregarded and
only individual features are considered, class attendance is indeed the important
feature. We note that our results on social ties are limited due to the limited num-
ber of interactions recorded (only contacts between participants of the CNS are
considered). Because of this limitation, our results show the lower bound for the
importance of social ties in inferring academic performance. The availability of more
data about interactions would allow us to make even more precise inferences.

We found network centrality to have positive correlation with academic perfor-
manance, in agreement with the literature [35–40]. However, from among all types
of interaction networks, only proximity network exhibited strong effect. A possible lim-
itation in measuring centrality is that the mere physical proximity of two individuals
does not necessarily involve direct communication. Nevertheless, it is reasonable to
expect an increased level of information exchange in a group of individuals if they
are in close proximity, which was the case in our dataset.\[1\]

Consistent with findings in existing literature, we found that class attendance
showed the strongest correlation with academic performance when we consider only
individual effects [46, 48, 49, 54–58]. We also found that Facebook activity had a
negative relation to academic performance – also in agreement with the majority
\[1\]The Copenhagen Networks study uses Bluetooth visibility as an indicator of
person-to-person proximity.
of the studies that investigated Facebook and social media usage [12–19]. We note, however, that our the data is limited to Facebook activities such as posting a status update or uploading a picture etc, and that we have no information regarding ‘passive’ Facebook usage, such as scrolling and reading. It is a reasonable assumption, however, that active participation on Facebook has a positive correlation with the (passive) time spent on Facebook. Also, our data does not include direct messages which may constitute a relevant fraction of communications performed via the social network site.

The analysis of the different personality traits revealed that two aspects of personality, namely conscientiousness and self-esteem are useful for predicting academic performance. These two traits reached a correlation coefficient between 0.2 and 0.3 corresponding to the upper limit achievable for any correlation with a personality trait, according to Mischel [59]. These results also agree with existing literature [60–89].

In the prediction experiment we achieved a classification accuracy of around 25.4 percentage points above baseline, a result similar to that of Vandamme et al. [9], although they utilized nearly ten times as many features to build a model as we did. In addition the accuracy of Vandamme et al. [9] is driven by using prior achievement (grades), which is known to be a strong predictor of performance (e.g. due to persistence of skill and motivation). Our findings—together with the results in the literature—emphasize that there is a considerable dependence of academic performance on personality and social environment. However, there are also other factors that play a role which we were unable to capture. See Materials and Methods for a review of the major limitations of our approach.

Materials and Methods
Data collection
Results presented in this paper are based on the data collected in the Copenhagen Network Study (CNS) [8]. In the CNS experiment, various data types were recorded by dedicated smartphones: Bluetooth scans, call and text message meta data, Facebook activity logs, and mobility traces among around 1,000 students at the Technical University of Denmark. The study covered the academic years 2013/14 and 2014/15. Participating students responded to a survey on personality. Finally, course grades were provided by the Technical University of Denmark. Due to the possibility to exit the experiment at any given point, the number of participants varied over time.

The raw data records are cleaned and transformed to meaningful information before the analysis. Bluetooth scans are used to estimate person-to-person interactions corresponding to a physical distance of up to 10 m between participants. After constructing proximity links, data is binned in 15 minutes time bins. Facebook data was obtained by the Facebook Graph API, and contains both static friendship connections as well as various interactions on the social network. All types of interactions are treated equally. Private messages, however, are unavailable at they cannot be obtained from Facebook using the official Graph API. Locations are estimated by selecting the ones with the highest accuracy in 15 minute bins. Only courses using the Danish 7-point grading scale are considered and all grades are converted to GPAs by linear scaling, with negative grades being converted to 0. We also calculate
cumulative and term average GPAs weighted by the course credit points, and only students attending at least three courses are considered. In total, 3494 term GPAs and 957 cumulative GPAs are computed.

In order to compute attendance we used the smartphone locations as well as person-to-person proximity obtained from Bluetooth scans. A detailed description of how we compute attendance can be found in a companion paper [50].

The different channels collected (proximity, Facebook, calls, and text messages) may indicate several types of social connections. Due to the nature and high usage frequency of these channels, in case of physical proximity and Facebook interactions links are weighted by the total number of encounters or actions. On the contrary, calls and text messages are reduced to unweighted relationships, as these interactions show insignificant variations over the pairs of students. To further decrease the impact of noise due to the wide distribution of weights in the proximity and Facebook interaction networks, a non-linear scaling function is applied to remove large deviations:

\[
    w'_{ij} = \lfloor \sqrt{w_{ij}} \rfloor,
\]

where \( w_{ij} \) denotes the original number of interactions between students \( i \) and \( j \). Qualitative results and the findings are not affected by the above scaling. Finally, all interactions are considered as undirected, that is, social ties are symmetric.

In order to increase the stability of the results we apply bootstrap resampling. That is, analyses are performed on 100 bootstrap samples, where each has the same size as the original sample. Results are given by the mean of the bootstrap analyses with approximated standard errors described by the \textit{Standard Error of the Mean}.

**Performance predictions**

During performance prediction, we define four different types of feature sets based on their origin and function: personality, individual, network and, combined (see Table 2 for the different feature sets along with the features included). The personality feature set is based on solely personality traits, while the individual set contains all individual related features that do not require knowledge on the structure of the social ties. On the other hand, the network feature set includes only features that are extracted from the social interactions. Finally, the combined set contains every feature introduced in this paper. Instances with missing attribute values were excluded. At total a sample of 540 students were considered for the prediction.

As a performance measure for prediction, we use classification accuracy. Model selection is carried out by a 10-fold cross-validation, and due to variations in the different realizations of predictions, we present the average performance of 30 independent predictions.

**Limitations**

Although we utilize wider and more insightful data than most other studies our approach has also some drawbacks which need to be taken into account. The first limitation comes from the fact that we only observed students from a single, technical, Danish university. For this reason, the findings may not be generalizable to
students at other institutions, of other academic disciplines or with other demographics. Furthermore, only a subset of all the students participated in our study – for first year students the rate was around 40%. There was some tendency of selection into the study as the average student tend to achieve higher grades [90]. Yet, within our dataset we observe a high degree of variation with respect to behavioral and network measures as well as academic performance. Thus, we argue that our results are valid and therefore appropriate precursors of the trends present in the larger student population.

Our measures of attendance and student social networks are also subject to some noise. Our method of inferring attendance and its drawbacks is discussed in [50] where we show it is quite similar to official records where available but with some error. Also our measures of social networks is limited to phone calling/texting, meetings and public Facebook messages. Although these are probably some of the most important means of communication, some students may communicate via other smartphone apps. Furthermore, attendance does not always imply neither in class participation nor attention to the taught material.

As a final remark we note that measuring academic success is not only exam performance but also includes long term learning, skills not measured in the tests etc.

Competing interests
The authors declare that they have no competing interests.

Funding
This work was supported by the Villum Foundation, the Danish Council for Independent Research, and University of Copenhagen (via the UCPH-2016 grant Social Fabric and The Center for Social Data Science).

Author’s contributions
All authors contributed equally to this work.

Acknowledgments
Due to privacy implications we cannot share data but researchers are welcome to visit and work under our supervision.

Author details
1 Institute of Software Technology, Graz University of Technology, Graz, Austria. 2 Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark. 3 Department of Economics, University of Copenhagen, Copenhagen, Denmark. 4 Center for Social Data Science, University of Copenhagen, Copenhagen, Denmark. 5 College of Information and Computer Science, Northeastern University, Boston, USA. 6 The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark.

References
1. Lavin, D.E.: The Prediction of Academic Performance. Russell Sage Foundation, New York City, USA (1965)
2. Macan, T.H., Shahani, C., Dipboye, R.L., Phillips, A.P.: College students’ time management: Correlations with academic performance and stress. Journal of educational psychology 82(4), 760 (1990)
3. Galević, D., Zouaq, A., Janzen, R.: “choose your classmates, your gpa is at stake!”: The association of cross-class social ties and academic performance. American Behavioral Scientist, 0002764213479362 (2013)
4. Curcio, G., Ferrara, M., De Gennaro, L.: Sleep loss, learning capacity and academic performance. Sleep medicine reviews 10(5), 323–337 (2006)
5. Singh, A., Uijtdewilligen, L., Twisk, J.W., Van Mechelen, W., Chinapaw, M.J.: Physical activity and performance at school: a systematic review of the literature including a methodological quality assessment. Archives of pediatrics & adolescent medicine 166(1), 49–55 (2012)
6. Van de Mortel, T.F., et al.: Faking it: social desirability response bias in self-report research. Australian Journal of Advanced Nursing, The 25(4), 40 (2008)
7. Junco, R.: Comparing actual and self-reported measures of facebook use. Computers in Human Behavior 29(3), 626–631 (2013)
8. Stopczynski, A., Sekara, V., Sapiezynski, P., Cutoffe, A., Madsen, M.M., Larsen, J.E., Lehmann, S.: Measuring large-scale social networks with high resolution. PloS one 9(4), 95978 (2014)
9. Vandamme, J.-P., Meskens, N., Superby, J.-F.: Predicting academic performance by data mining methods. Education Economics 15(4), 405–419 (2007)
10. Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., Campbell, A.T.: Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 3–14 (2014). ACM

11. Nghe, N.T., Janeczek, P., Haddawy, P.: A comparative analysis of techniques for predicting academic performance. In: Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE’07. 37th Annual, pp. 2–7 (2007). IEEE

12. Maqableh, M.M., Rajab, L., Quteshat, W., Moh’d Taisir Masa, R., Khatib, T., Karajeh, H., et al.: The impact of social media networks websites usage on students’ academic performance. Communications and Network \(7(04), 159\) (2015)

13. Al-Menayes, J.J.: Social media use, engagement and addiction as predictors of academic performance. International Journal of Psychological Studies \(7(4), 86\) (2015)

14. Al-Menayes, J.J.: The relationship between mobile social media use and academic performance in university students. New Media and Mass Communication \(25, 23–29\) (2014)

15. Karpinski, A.C., Kirschner, P.A., Ozer, I., Mellott, J.A., Ochwo, P.: An exploration of social networking site use, multitasking, and academic performance among united states and europe non-creative discipline. Higher Education, 1–16 (2015)

16. Paul, J.A., Baker, H.M., Cochran, J.D.: Effect of online social networking on student academic performance. Computers in Human Behavior \(28(6), 2117–2127\) (2012)

17. Junco, R.: The relationship between frequency of facebook use, participation in facebook activities, and student engagement. Computers & Education \(58(1), 162–171\) (2012)

18. Jacobsen, W.C., Forste, R.: The wired generation: Academic and social outcomes of electronic media use among university students. Cyberpsychology, Behavior, and Social Networking \(14(5), 275–280\) (2011)

19. Kirschner, P.A., Karpinski, A.C.: Facebook® and academic performance. Computers in human behavior \(26(6), 1237–1245\) (2010)

20. Pasek, J., Hargittai, E., et al.: Facebook and academic performance: Reconciling a media sensation with data. First Monday \(14(5)\) (2009)

21. Ainin, S., Naqshbandi, M.M., Moghavvemi, S., Jafar, N.I.: Facebook usage, socialization and academic performance. Computers & Education \(83, 64–73\) (2015)

22. Kolek, E.A., Saunders, D.: Online disclosure: An empirical examination of undergraduate facebook profiles. NASPA journal \(45(1), 1–25\) (2008)

23. Tayseer, M., Zoghieb, F., Alcheikh, I., Awadallah, M.N.: Social Network: Academic & Social Impact on College Students. Retrieved 20th November (2014)

24. Sacerdote, B.: Peer effects with random assignment: Results for dartmouth roommates. Quarterly Journal of Economics, 681–704 (2001)

25. Zimmerman, D.J.: Peer effects in academic outcomes: Evidence from a natural experiment. Review of Economics and statistics \(85(1), 9–23\) (2003)

26. Stonebrinckner, R., Stonebrinckner, T.R.: What can be learned about peer effects using college roommates? evidence from new survey data and students from disadvantaged backgrounds. Journal of public Economics \(90(8), 1435–1454\) (2006)

27. Carrell, S.E., Sacerdote, B.I., West, J.E.: From natural variation to optimal policy? the importance of endogenous peer group formation. Econometrica \(81(3), 855–882\) (2013)

28. Vitale, M.P., Porzio, G.C., Doreian, P.: Examining the effect of social influence on student performance through network autocorrelation models. Journal of Applied Statistics \(43(1), 115–127\) (2016)

29. Sminov, I., Thurner, S.: Formation of homophily in academic performance: students prefer to change their friends rather than performance. arXiv preprint arXiv:1606.09082 (2016)

30. Poldin, O., Valeeva, D., Yudkevich, M.: How social ties affect peer group effects: Case of university students. Higher School of Economics Research Paper No. WP BPR 15 (2013)

31. Mayer, A., Puller, S.L.: The old boy (and girl) network: Social network formation on university campuses. Journal of public economics \(92(1), 329–347\) (2008)

32. Yuan, Y.C., Gay, G., Hembrooke, H.: Focused activities and the development of social capital in a distributed learning “community”. The Information Society \(22(1), 25–39\) (2006)

33. Rizzato, T.E., LeDoux, J., Hatala, J.P.: It’s not just what you know, it’s who you know: Testing a model of the relative importance of social networks to academic performance. Social Psychology of Education \(12(2), 175–189\) (2009)

34. Tomás-Miquel, J.-V., Expósito-Langa, M., Nicolau-Julía, D.: The influence of relationship networks on academic performance in higher education: a comparative study between students of a creative and a non-creative discipline. Higher Education, 1–16 (2015)

35. Sparrowe, R.T., Liden, R.C., Wayne, S.J., Kraimer, M.L.: Social networks and the performance of individuals and groups. Academy of management journal \(44(2), 316–325\) (2001)

36. Smith, R.A., Peterson, B.L.: “past… what do you think?” the relationship between advice prestige, type of advice, and academic performance. Communication Education \(56(3), 278–291\) (2007)

37. Hommes, J., Rienties, B., De Grave, W., Bos, G., Schuwirth, L., Scherpbier, A.: Visualising the invisible: a network approach to reveal the informal social side of student learning. Advances in Health Sciences Education \(17(5), 743–757\) (2012)

38. Baldwin, T.T., Bedell, M.D., Johnson, J.L.: The social fabric of a team-based mba program: Network effects on student satisfaction and performance. Academy of management journal \(40(6), 1369–1397\) (1997)

39. Yang, H., Tang, J.: Effects of Social Network On Students, Performance: A web-based Forum Study in Taiwan. JALN, 7 (3), 93. Retrieved 11 26, 2011 (2003)

40. Cho, H., Gay, G., Davidson, B., Ingraffia, A.: Social networks, communication styles, and learning performance in a cisl community. Computers & Education \(49(2), 309–329\) (2007)

41. Thomas, S.L.: Ties that bind: A social network approach to understanding student integration and persistence.
42. Johnson, D.W., Johnson, R.T.: Structuring groups for cooperative learning. Journal of Management Education 9(4), 8–17 (1984)
43. Manski, C.F.: Identification of endogenous social effects: The reflection problem. The review of economic studies 60(3), 531–542 (1993)
44. Eagle, N., Pentland, A.S., Lazer, D.: Mobile phone data for inferring social network structure. In: Social Computing, Behavioral Modeling, and Prediction, pp. 79–88. Springer, ??? (2008)
45. Prociuk, T.J., Breen, L.J.: Locus of control, study habits and attitudes, and college academic performance. The Journal of Psychology 88(1), 91–95 (1974)
46. Buckalew, L., Daly, J.D., Coffield, K.: Relationship of initial class attendance and seating location to academic performance in psychology classes. Bulletin of the Psychonomic Society 24(1), 63–64 (1986)
47. Marburger, D.R.: Does mandatory attendance improve student performance? The Journal of Economic Education 37(2), 148–155 (2006)
48. Chen, J., Lin, T.-F.: Class attendance and exam performance: A randomized experiment. The Journal of Economic Education 39(3), 213–227 (2008)
49. Credé, M., Roch, S.G., Kiesczynka, U.M.: Class attendance in college a meta-analytic review of the relationship of class attendance with grades and student characteristics. Review of Educational Research 80(2), 272–295 (2010)
50. Kassarnig, V., Bjerre-Nielsen, A., Mones, E., Lehmann, S., Dreyer Lassen, D.: Class attendance, peer similarity, and academic performance in a large field study. PloS one, (2017)
51. Van Cleemput, K.: “I’ll see you on im, text, or call you”: A social network approach of adolescents’ use of communication media. Bulletin of Science, Technology & Society 30(2), 75–85 (2010)
52. Fowler, J.H., Christakis, N.A.: Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. Bmj 337, 2338 (2008)
53. Strickland, B.R., Haley, W.E.: Sex differences on the rotter i-e scale. Journal of Personality and Social Psychology 39(5), 930 (1980)
54. Stanca, L.: The effects of attendance on academic performance: Panel data evidence for introductory microeconomics. The Journal of Economic Education 37(3), 251–266 (2006)
55. Van Blerkom, M.L.: Class attendance in undergraduate courses. The Journal of psychology 126(5), 487–494 (1992)
56. Brocato, J.: How much does coming to class matter? some evidence of class attendance and grade performance. Educational Research Quarterly (1989)
57. Gump, S.E.: The cost of cutting class: Attendance as a predictor of success. College Teaching 53(1), 21–26 (2005)
58. Lin, T.-F., Chen, J.: Cumulative class attendance and exam performance. Applied Economics Letters 13(14), 937–942 (2006)
59. Mischel, W.: Personality and Assessment. Lawrence Erlbaum Associates, New Jersey, USA (2013)
60. Dollinger, S.J., Orf, L.A.: Personality and performance in "personality": Conscientiousness and openness. Journal of Research in Personality 25(3), 276–284 (1991)
61. Goff, M., Ackerman, P.L.: Personality-intelligence relations: Assessment of typical intellectual engagement. Journal of Educational Psychology 84(4), 537 (1992)
62. Rothstein, M.G., Paunonen, S.V., Rush, J.C., King, G.A.: Personality and cognitive ability predictors of performance in graduate business school. Journal of educational psychology 86(4), 516 (1994)
63. Wolfe, R.N., Johnson, S.D.: Personality as a predictor of college performance. Educational and psychological measurement 55(2), 177–185 (1995)
64. De Fruyt, F., Mervielde, I.: Personality and interests as predictors of educational streaming and achievement. European journal of personality 10(5), 405–425 (1996)
65. Paunonen, S.V.: Hierarchical organization of personality and prediction of behavior. Journal of Personality and Social Psychology 74(2), 538 (1998)
66. Busato, V.V., Prins, F.J., Elshout, J.J., Hamaker, C.: Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. Personality and Individual Differences 29(6), 1057–1068 (2000)
67. Paunonen, S.V., Ashton, M.C.: Big five predictors of academic achievement. Journal of Research in Personality 35(1), 78–90 (2001)
68. Gray, E.K., Watson, D.: General and specific traits of personality and their relation to sleep and academic performance. Journal of personality 70(2), 177–206 (2002)
69. Lievens, F., Coetzier, P., De Fruyt, F., De Maeseneer, J.: Medical students’ personality characteristics and academic performance: A five-factor model perspective. Medical education 36(11), 1050–1056 (2002)
70. Bauer, K.W., Liang, Q.: The effect of personality and precollege characteristics on first-year activities and academic performance. Journal of College Student Development (2003)
71. Chamorro-Premuzic, T., Furnham, A.: Personality traits and academic examination performance. European Journal of Personality 17(3), 237–250 (2003)
72. Chamorro-Premuzic, T., Furnham, A.: Personality predicts academic performance: Evidence from two longitudinal university samples. Journal of Research in Personality 37(4), 319–338 (2003)
73. Diseth, A.: Personality and approaches to learning as predictors of academic achievement. European Journal of Personality 17(2), 143–155 (2003)
74. Farsides, T., Woodfield, R.: Individual differences and undergraduate academic success: The roles of personality, intelligence, and application. Personality and Individual differences 34(7), 1225–1243 (2003)
75. Furnham, A., Chamorro-Premuzic, T., McDougall, F.: Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. Learning and Individual Differences 14(1), 47–64 (2002)
76. Lounsbury, J.W., Sundstrom, E., Loveland, J.M., Gibson, L.W.: Intelligence, “big five” personality traits, and work drive as predictors of course grade. Personality and Individual Differences 35(6), 1231–1239 (2003)
77. Phillips, P., Abraham, C., Bond, R.: Personality, cognition, and university students’ examination performance. European Journal of Personality 17(6), 435–448 (2003)
78. Duff, A., Boyle, E., Dunleavy, K., Ferguson, J.: The relationship between personality, approach to learning and academic performance. Personality and individual differences 36(8), 1907–1920 (2004)
79. Furnham, A., Chamorro-Premuzic, T.: Personality and intelligence as predictors of statistics examination grades. Personality and Individual Differences 37(5), 943–955 (2004)
80. Hair, P., Hampson, S.E.: The role of impulsivity in predicting maladaptive behaviour among female students. Personality and Individual Differences 40(5), 943–952 (2006)
81. Conard, M.A.: Aptitude is not enough: How personality and behavior predict academic performance. Journal of Research in Personality 40(3), 339–346 (2006)
82. Barchard, K.A.: Does emotional intelligence assist in the prediction of academic success? Educational and psychological measurement 63(5), 840–858 (2003)
83. Langford, P.H.: A one-minute measure of the big five? evaluating and abridging shafe’s (1999a) big five markers. Personality and Individual Differences 35(5), 1127–1140 (2003)
84. Oswald, F.L., Schmitt, N., Kim, B.H., Ramsay, L.J., Gillespie, M.A.: Developing a biodata measure and situational judgment inventory as predictors of college student performance. Journal of applied psychology 89(2), 187 (2004)
85. Leong, F.T., Gibson, L.W., Lounsbury, J.W., Huffstetler, B.C.: Sense of identity and collegiate academic achievement. Journal of College Student Development 46(5), 501–514 (2005)
86. Ridgell, S.D., Lounsbury, J.W.: Predicting academic success: General intelligence,” big five” personality traits, and work drive. College Student Journal 38(4), 607 (2004)
87. Komarraju, M., Karau, S.J., Schmeck, R.R.: Role of the big five personality traits in predicting college students’ academic motivation and achievement. Learning and individual differences 19(1), 47–52 (2009)
88. Noftle, E.E., Robins, R.W.: Personality predictors of academic outcomes: big five correlates of gpa and sat scores. Journal of personality and social psychology 93(1), 116 (2007)
89. Lane, J., Lane, A.M., Kyprianou, A.: Self-efficacy, self-esteem and their impact on academic performance. Social Behavior and Personality: an international journal 32(3), 247–256 (2004)
90. Bjerre-Nielsen, A., Dreyer Lassen, D.: Opportunity and similarity in dynamic friendships. Technical report (2017)

Figures

| Feature                          | ANOVA F-value |
|---------------------------------|---------------|
| low performer friends texts     | 68.02         |
| mean GPA of texts               | 53.7          |
| mean GPA of calls               | 46.06         |
| low performer friends calls     | 42.9          |
| high performer friends texts    | 30.35         |
| mean GPA of p2p contacts        | 30.33         |
| mean GPA of FB interactions     | 28.41         |
| high performer friends calls    | 24.39         |
| class attendance                | 24.8          |
| mean GPA of FB friends          | 19.79         |
| p2p centrality                  | 19.66         |
| conscientiousness               | 12.58         |
| self esteem                     | 9.16          |
| call centrality                 | 7.08          |

Figure 1 Feature importance ranking. Results from ANOVA F-test for 3-class classification task. Features which did not achieve sufficient significance ($p \geq 0.001$) are omitted.

Tables
Figure 2 Performance of the models based on the different feature sets. Bars show the classification accuracy of the models. Grey color marks the model and feature set with the best performance.

Figure 3 Precision-recall plot of the most efficient model and feature set. Dots represent the model performance in the low (red), moderate (green) and high (blue) performer classes. Dashed lines mark the profile of constant $F_1$ corresponding to the measured values for the specific class.

| Channel                      | $r_s$ |
|------------------------------|-------|
| Texts                        | .432  |
| Calls                        | .415  |
| Facebook interactions        | .323  |
| Facebook friendships         | .300  |
| Person-to-person interactions | .284  |

Table 1 Correlation between the cumulative GPA of the students and the mean cumulative GPA of their contacts based on different communication channels. Corresponding $p$-values are below 0.001.
Figure 4 Facebook usage and performance in the quintiles. (a) Division of students into five groups of equal size according to their active Facebook updates. Each bar represents a single quintile, width corresponds to the span of Facebook activity in the specific group and height shows the mean term GPA. (b) Grade distribution inside each Facebook activity class.

Figure 5 Similarity in academic performance for social ties. Students are divided in 10 deciles and curves show the mean GPA of the contacts in the respective performance group for each communication channel.
Figure 6 Correlation between performance of strong contacts. For each student, we show their cumulative GPA versus the mean GPA of their contacts obtained by their text messages. Color denotes density of points in arbitrary units.

Figure 7 Own academic performance and peers’ academic performance. Each histogram displays how students distribute their text messages exchanged with others over the various performance groups. Groups are defined by quintiles based on their term GPA.

| Feature set  | Features                                                                                                                                                                                                 |
|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| personality  | BFI: neuroticism, openness, conscientiousness, extraversion, agreeableness; satisfaction with life; locus of control; PANAS: positive and negative; self-esteem; loneliness; stress; depression; narcissism: rivalry, admiration, overall. |
| individual   | Facebook activity; class attendance, gender, and all personality traits.                                                                                                                                     |
| network      | Number of contacts, mean GPA of friends, and fraction of low/high performing friends in networks based on calls, text messages, proximity, Facebook interactions and Facebook friends.                        |
| combined     | All individual features and all network features together.                                                                                                                                                  |

Table 2 Feature sets for performance predictions.
Figure 8 Correlation between network measures and cumulative GPA. Bars denote the correlation measured between network descriptors and cumulative GPA: degree (gray) and eigenvector centrality (red).