Wifi Usage on Campus and Students Academic Performance

Pedro Ferreira pedrof@cmu.edu[1,2]

Rodrigo Belo rbelo@rsm.nl[3]

Yael Inbar yinbar@cmu.edu[1,2]

Ryan Turner rjturner@cmu.edu[2]

1 – Heinz College, Carnegie Mellon University

2 – Dept Engineering and Public Policy, Carnegie Mellon University

5000 Forbes Ave. Pittsburgh PA, 15203

3 – Rotterdam School of Management, Erasmus University

Postbus 1738, 3000 DR Rotterdam, Netherlands

Significance

We show that Wifi usage in a University campus is associated to better grades and to completing more courses per semester. This effect is driven by Wifi usage during the day and by more advanced students. We also find that student performance tends to align with that of "collocated peers" -- students that share the same space when using Wifi on campus at the same time. Furthermore, we show that low performing students benefit from being paired with high performing students, with the latter affected only slightly. This pairing policy is productive only for advanced students. Our findings help shape education policy and aid teachers and administrators when it comes to planning shared learning spaces and team work.

Abstract

The extensive use of Internet by students in higher education raises questions about the impact of campus Wifi networks on academic performance. In this study, we use a unique dataset covering 3,030 students in a European university, comprising data on all their Wifi sessions and academic performance over 5 consecutive semesters. Academic performance is measured by grades and by the number of courses completed per semester. Using Dynamic Propensity Score Matching, we find a significant positive association between Wifi usage and the performance of students. We also find that this association is driven by Wifi usage during the day and by more advanced students (juniors and seniors). In addition, we use randomization techniques (Shuffle Tests) to estimate the effect of peers who frequently share the same space on campus for Wifi connectivity on each other's grades. On average, our results show that the performance of a student tends to align with that of the peer that she spends most time with while using Wifi on campus. Still, results are heterogeneous across curricular years and across different levels of prior performance. We find empirical evidence that pairing high and low performance students is likely to be productive only for advanced students. Overall, our results show that more mature students are more likely to benefit from Wifi connectivity and from peer effects. Our findings inform students, teachers, managers and administrators in higher education about how best to *plan shared learning spaces on campus and to assemble students into teams.*

Paper's body:

The determinants of academic performance in higher education are extensively studied, and include personal, behavioral and environmental factors. For example, class attendance and 'orderness' (i.e., regularity of daily activities) were found to be positively correlated with grades (1–3). The use of Information and Communication Technologies (ICT) in education and, in particular, the use of Internet among students, has been a major focus of research in recent years given that digital technologies can significantly help learning (4) in and outside the classroom. However, there is still significant debate among scholars and educators as how to best take advantage of ICTs in educational settings. Doubts arise because the academic performance of students in higher education was found to be negatively correlated with Internet use (5–7), cell phone use (8), in-class laptop use (9) and social media use (10). Heavy use of the Internet was also found to be negatively correlated to students grades (6, 11), presumably because it increases distraction (12).

Traditionally, studies of the factors that affect the performance of students were limited to students' self-reports, such as surveys and questionnaires, which are considered prone to biases and errors (2). Nowadays, the ubiquity of mobile digital devices allows for collecting the "digital traces" of students. These traces – originated from mobile phone calls, Wifi usage and other digital sensors – offer a more complete mosaic of how students behave and use ICTs on campus potentially leading to a better understanding of the antecedents of good academic performance. Several recent studies have used data collected from students' digital traces: Wang et al. (3) used data from a smartphone sensing app on students cellphones to learn about their activities and their mobility around campus. Cao et al. (1) used data from students' activities as captured by their smart cards to learn about their daily activities on campus. Zhou et al. (13) used students' locations obtained from Wifi connectivity logs, along with data from smartphones apps to evaluate the students attendance and behavior during class time. Scanlon (14) utilized Wifi logs and students location data to infer their sociability throughout the semester, measured by the interactions with other students on campus. Our Wifi usage data include, for each Wifi session, the location, the duration and the amount of data transferred. This allows us to derive insights about how students use Wifi on campus, which is a meaningful activity that students engage with on a daily basis on campus.

Our study aligns with a recent broad strand of research looking at human behavior from digital footprints to learn about socially relevant outcomes. For example, large datasets of digital traces were used by scholars to study various social issues such as mobility (15), urban planning (16), poverty and wealth status (17), crime hotspots (18), health services allocation (19) and literacy in developing countries (20). A few studies have used data on Wifi usage in higher education to proxy student behavior and learn about mobility patterns (21, 22). Instead, we use this type of data to study the effect of Wifi usage and of collocated peers on academic performance.

We use a unique dataset of Wifi session logs, coupled with students' administrative data, including their grades and the number of courses completed per semester for 5 consecutive semesters. We create a quasi-experimental setting using Dynamic Propensity Score Matching and show that the duration and volume (number of bytes exchanged) of Wifi usage on campus has a significant positive association with both measures of academic performance. Additional tests show heterogeneous effects with regards to the students' maturity and to when Wifi is used – more advanced students (juniors and seniors) benefit from Wifi more, and usage during the day (8am-8pm) is more productive.

We also leverage our Wifi logs, and in particular their timestamp and location data, to learn about how the performance of "collocated peers" affects one's performance (23). Collocated peers are students that share the same space on campus for Internet use at the same time. We use the term "collocated peers" to encompass all relationships that might arise among students who frequently share learning spaces on campus. In the academic environment, these relationships are, most likely, friendships and co-studying. Two "collocated peers" share the same space on campus often and, consequently, their behavior may influence each other's performance. Collocation has been used before as a way to identify peers (25). In this respect our study is similar to prior works looking at the effect of peers on educational outcomes. Before, peers have been defined as college roommates (24) or as students that exchange text messages (2).

We use a randomization technique known as the Shuffle Test to account for the underlying baseline correlation across students' grades that, by construction, cannot come from peer influence. We find evidence that one's academic performance tends to align with that of the peer that they spend most time with while using Wifi on campus. This effect is again heterogeneous with respect to the maturity of students and, in this case, also with respect to the students' prior performance (measured by the students' application scores when entering the university): for freshmen and sophomores peer effects are positive and small for those with high prior performance and inexistent for those with low prior performance. However, for juniors and seniors, peer effects are always positive but significantly larger for those with low prior performance. Therefore, our empirical evidence shows that pairing low and high performance students is likely to be productive for more advanced students, but otherwise unproductive when involving early stage students.

Our findings offer new insights for students, teachers, managers as well as administrators at institutions of higher education, when it comes to planning shared spaces for learning and to allocating students into groups.

Data

We perform our analysis using an anonymized panel of data that combines (i) session-level Wifi usage data of students in a European university, including: access point, time and duration of connection, and number of megabytes transferred; and (ii) administrative data, at the student-semester level, including GPA and number of courses completed. We also have information on the student's application score when admitted to the university, year started, curricular year, and major. We analyze the activity of 3,030 students that use Wifi, spanning five consecutive semesters (Fall 2006 to Fall 2008), which results in a dataset with 6,425 student-semester observations. For details about the setting, the data, and descriptive statistics see the Empirical Context and the Data sections of the SI. The median Wifi usage of a student in a semester is 69 hours, with 6% of the students in our dataset spending more than 400 hours per semester on Wifi. The mean number of courses completed in a semester is 4, and the mean grade points earned in a semester is 53 (grades are between 0 and 20, students need at least 9.5 points to pass a course).

Wifi usage and academic performance

We use Dynamic Propensity Score Matching (DPSM) to estimate the effect of Wifi usage on student performance by comparing the performance of students that use a significant amount of Wifi (treated) and that of students that use only a trivial amount of Wifi (control). We compare only across students that are similar on several covariates such as curricular year, major, cohort and application score when applied to the university. Comparing only across these users increases significantly our ability to avoid selection bias arising from potential unobserved confounding factors. We follow the approaches in (26) and (27), and match observations separately per semester and then combine semesters to obtain again a full matched panel. In this way, treated students always match to control students in the same semester but we do not force one treated student to be matched to the same control student for the whole duration of our panel, which increases significantly our ability to perform matching. Our matching procedure significantly reduced the bias in pre-treatment covariates as shown in detail in the Propensity Score Matching and Balance Checks section of the SI.

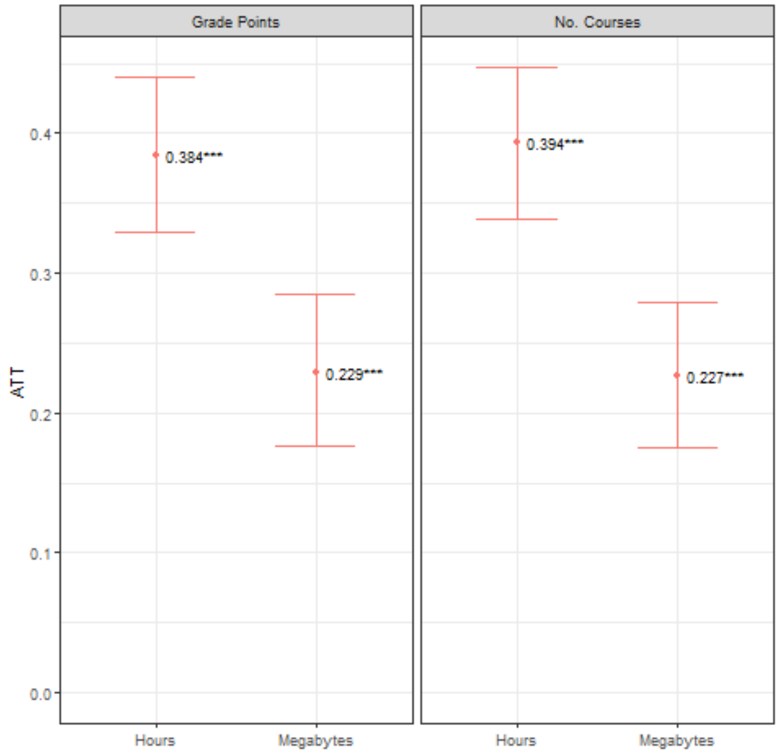

Figure 1. Average Treatment effect on the Treated (ATT) (in percentage terms) by performance measure (Grade Points and Number Courses Completed) and by Wifi usage measure: time connected (hours) and amount of information exchanged (megabytes).

We match students in the upper quintile of Wifi usage with students in the lower quintile of Wifi usage (the latter are therefore students that use only a trivial amount of Wifi). This way, we eliminate the potential impact of unobserved factors limiting access to Wifi, such as laptop ownership or lack of know-how for how to log into the Wifi network, and focus on the effect of Wifi usage *per se*. Wifi usage is defined by the amount of time connected to access points as well as by the amount of traffic exchanged with access points. Details of the models and of the DPSM design that we use are provided in the Propensity Score Matching section of the SI.

Fig. 1 shows the Average Treatment effect on the Treated (ATT). Students in the upper quintile of Wifi usage exhibit an increase of 38% in academic performance (both in terms of grade points and number of courses completed) compared to students in the lower quintile of Wifi usage, when Wifi usage is measured by time. Therefore, we find clear evidence that Wifi usage is positively associated with better academic performance. One may however argue that the duration of Wifi sessions does not reflect the intensity of Internet consumption. Therefore, we also test the effect of Wifi usage on academic performance measuring the former with the number of megabytes exchanged with access points. The results shown in Fig. 1 for this case come in line with the ones discussed above, with students in the upper quintile of Wifi usage exhibiting a 23% increase in performance compared to those in the lower quintile.

In another specification, we take advantage of the fact that the same student shows up in our panel in multiple semesters and use first-differences to provide additional results, in this case explicitly controlling for unobserved student fixed effects. We find again a positive association between Wifi usage and academic performance. Separating Wifi usage into daytime (8am-8pm) and nighttime usage, we find that this positive association is essentially driven by Wifi usage during the day. Interacting Wifi usage with curricular year, we find that the positive effect of Wifi usage is significantly stronger for more advanced students – juniors and seniors (curricular years 3 and above) – compared to that for freshman and sophomores (students in curricular years 1 and 2). This result hints at the idea that student maturity is likely to play a role on how productive Wifi usage is. In another specification, we show that our results remain unchanged when we try to control for the time that students spend on campus, given that more time on campus is likely associated to more Wifi usage but also to higher grades through, for example, more dedication. We proxy time on campus by looking at Wifi logs early and late in the same day. All models that we estimate using first-differences are detailed in the Empirical Strategy section and the Results Obtained using First-Differences section of the SI.

Collocated peers and academic performance

Students spend the majority of their time on campus in shared spaces, such as classrooms, libraries, study rooms, and cafes, which allows us to further explore the students' Wifi logs by considering what happens to their performance when they use Wifi at these shared spaces. A student that connects to an access point that serves one of these spaces shares the space with other students that connect to the same access point at the same time. We call “collocated peers” to students that spend a significant amount of time connected to the same access points at the same time. This setup allows us to estimate the effect of “collocated peers” on student performance. More specifically, we estimate the effect of the performance of one’s top collocated peer on her own performance. The “top collocated peer” of a student is the student that she spends most time with when using Wifi on campus. This exercise requires a careful definition of “collocated peers”. To this end, we draw a social network of “collocated peers” using the session-level Wifi data. This network is represented by a graph of students that use Wifi on campus. An edge between two students indicates that they have been at the same access point at the same time using Wifi. The weight on each edge represents the number of times that they shared access points across campus (for at least 5 consecutive minutes). We obtain a graph with 5 sub-networks (one for each semester), where each node represents a student in a given semester, and each edge represents a collocated relationship between two students. Our overall graph includes 12,486 vertices and 3,942,550 edges. In these graphs, the “top collocated peer” of a student is student connected to her through the edge with the highest weight.

We use a randomization technique known as the “Shuffle Test” (26, 28) to disentangle the effect of “collocated peers” on performance from other potential unobserved sources of correlation. For example, dedicated students may spend more time on campus, which increases the likelihood of sharing common spaces with other students. If being a dedicated student leads to higher grades, then students' grades can be correlated with the grades of their "collocated peers" simply because they spend time at the same shared spaces. We are interested in the effect of "collocated peers" on a student's own performance net of these spurious correlations. To measure this effect, we compare the results obtained using the real-world data with the results obtained using data from a pseudo-world obtained from shuffling (within the same building) the locations where students access Wifi. By shuffling students' Wifi session locations, we are also shuffling the "collocated peers" but keeping all other student-level variables untouched. This produces data describing a pseudo-world in which each student uses as much Wifi as in the original world, as well as at the same times of the day, from the same subset of access points, but her "collocated peers" are different. In fact, they become randomly defined. Therefore, any correlation between one's performance and the performance of her "collocated peers" in this pseudo-world cannot come from peer influence because in this world "collocated peers" are not the real-world peers. Rather, such correlation represents an underlying level of interdependence in the performance of "collocated peers" that arises from unobserved sources, which should be subtracted from the estimate of peer influence obtained using the original data. More precisely, we create 100 pseudo worlds using the technique described above and subtract from the estimate obtained using the original dataset the average of the correlations obtained from these simulated worlds. The SI provides additional details about this procedure.

We employ a regression model in which one's performance in a given semester is a function of the performance of her "top collocated peer" in the same semester. The results obtained show a significant positive effect of 14-15%, on both grade points and number of courses completed per semester. Therefore, and on average, one's performance in a given semester tends to align with the performance of her "top collocated peer". These results are discussed in more detail in the SI.

We also estimate a heterogeneous peer effects model to evaluate the effect of "collocated peers" per prior level of performance and per curricular year. We split students into below the median and above the median on prior performance, measured using the score with which they applied to the University (which was determined before the Wifi activity recorded in our panel of data thus avoiding simultaneity bias in our analysis).

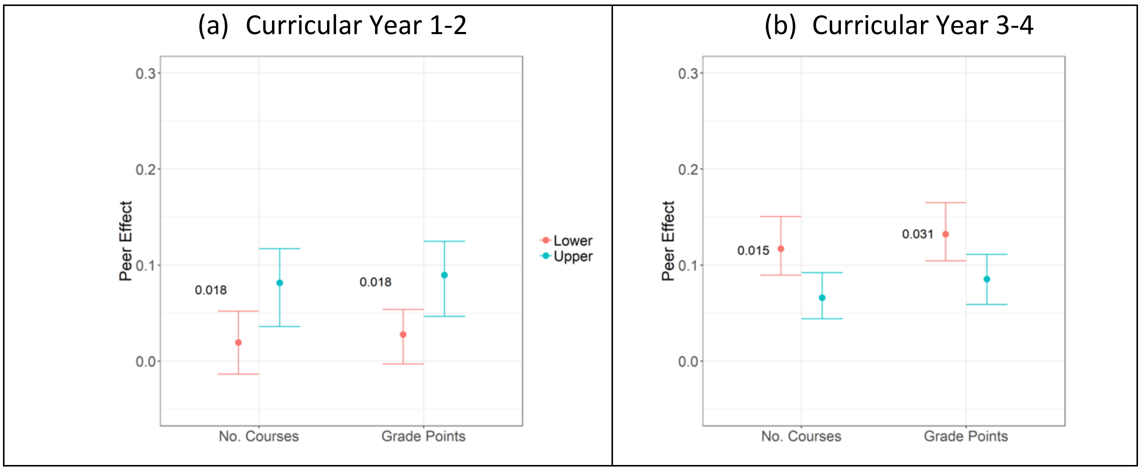

Fig. 2. Peer effect by students' prior performance (measured by the score with which they applied to University). Effect on number of courses completed per semester and grade points for curricular years 1 and 2 (a) and for curricular years 3 and 4 (b).

Figure notes: p-values shown in black inside figures are for t-tests comparing high and low performance students (high performance students have an application score above the median score, low performance students have an application score below the median).

Fig. 2 shows clear heterogeneity in the effect of the performance of the "top collocated peer" on one's performance, measured either by grade points or by number of courses completed. This effect is not statistically different from zero for freshmen and sophomores (students in curricular years 1 and 2) with prior performance below the median, and positive and significant for those with prior performance above the median. However, the effect of the "top collocated peer" is positive but small for juniors and seniors (students in curricular years 3 and above) with prior performance above the median and significantly higher for those with prior performance below the median. These results show that pairing high and low prior performance students is likely unproductive across freshmen and sophomores but may be interesting across juniors and seniors. In particular, pairing across the former students is likely to hurt high performing students without benefiting significantly the low performing ones. However, pairing across the more advanced students is likely to improve the performance of low performing students without hurting too much the performance of high performing students. Incentivizing pairing students across juniors and seniors can be accomplished by promoting collocated and collaborative learning, for example, in teams or by frequently using shared spaces.

Still, we acknowledge that the heterogeneous effects described above may arise because freshmen, sophomores, juniors and seniors perform different tasks. However, we also note that our findings on peer effects are consistent with the idea that pairing students is more productive when involving more mature students, who may be better equipped to help their lower performance classmen, and thus come in line with our prior findings that using Wifi is also more productive for juniors and seniors.

Finally, we estimate several specifications as robustness tests, including different definitions for high and low performance students, models with lagged grades and estimations without data from access points serving classrooms because using these data may potentially lead us to classify students that take the same classes as "collocated peers", which might not capture well the idea of relationships that arise organically across students on campus outside class time. A detailed description of all these models is provided in the Peers Effect section of the SI. All models provide similar results and support the conclusions described above.

Discussion

We use a granular dataset with Wifi logs from students in a European university, coupled with their corresponding performance measures (grade points and number of courses completed per semester) to learn about how Wifi usage and peers affect academic performance. We use Dynamic PSM and first differences to estimate the effect of Wifi usage on performance, thus comparing only similar students, which reduces our concerns with selection bias. Our results show that, on average, Wifi usage on campus is positively associated with academic performance. Still, this result is essentially driven by Wifi usage during the day and by junior and senior students.

We also investigate the effect of "collocated peers" on academic performance using the Wifi logs to learn when students share learning spaces on campus (outside classrooms). Using randomization techniques, we show that, on average, the performance of students tends to align with that of the peer they spend most time with while using Wifi on campus. Still, effects are heterogeneous and, in particular, we find that pairing low and high performance classmen is likely unproductive for freshmen and sophomores but may be appropriate for juniors and seniors. This finding has significant implications for students, teachers, managers as well as administrators at institutions of higher education when it comes to planning shared learning spaces, and potentially team work too, outside classrooms.

Methodologically, our work exemplifies how using the digital footprints of students on the campus in the era of big data analytics can help unveil new findings about factors that affect student performance. Using these logs, coupled with the appropriate analytical methods to examine the behavior of students on campus, is likely to provide more accurate, and thus more meaningful and reliable information for policy making, compared to the traditional self-reports that became popular over the years in both research and practice. Our work shows how these data can be used to inform planning decisions and our approach can be generalized to other domains that study big datasets of digital traces to better understand the dynamics of other social outcomes.

Materials and Methods

See SI Appendix for a detailed description of all materials and methods used in this study, including the models estimated and additional robustness checks performed.

References

1. Cao Y, et al. (2018) Orderliness predicts academic performance: behavioural analysis on campus lifestyle. *J R Soc Interface*. Available at: https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2018.0210 [Accessed April 15, 2019].

2. Kassarnig V, Bjerre-Nielsen A, Mones E, Lehmann S, Lassen DD (2017) Class attendance, peer similarity, and academic performance in a large field study. *PLOS ONE* 12(11):e0187078.

3. Wang R, et al. (2014) StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp ’14. (ACM, New York, NY, USA), pp 3–14.

4. Lundin J, Nuldén U, Svensson L (2008) A new agenda for research on higher education and information technology. *31th Information Systems Research Seminar in Scandinavia, IRIS 31*.

5. Belo R, Ferreira P, Telang R (2014) Broadband in School: Impact on Student Performance. *Manag Sci* 60(2):265–282.

6. Chen Y-F, Peng SS (2008) University Students’ Internet Use and Its Relationships with Academic Performance, Interpersonal Relationships, Psychosocial Adjustment, and Self-Evaluation. *Cyberpsychol Behav* 11(4):467–469.

7. Englander F, Terregrossa RA, Wang Z (2010) Internet use among college students: tool or toy? *Educ Rev* 62(1):85–96.

8. Lepp A, Barkley JE, Karpinski AC (2015) The Relationship Between Cell Phone Use and Academic Performance in a Sample of U.S. College Students. *SAGE Open* 5(1):2158244015573169.

9. Fried CB (2008) In-class laptop use and its effects on student learning. *Comput Educ* 50(3):906–914.

10. Junco R (2015) Student class standing, Facebook use, and academic performance. *J Appl Dev Psychol* 36:18–29.

11. Akhter N (2013) Relationship between internet addiction and academic performance among university undergraduates. *Educ Res Rev* 8(19):1793–1796.

12. Wentworth DK, Middleton JH (2014) Technology use and academic performance. *Comput Educ* 78:306–311.

13. Zhou M, et al. (2016) EDUM: Classroom Education Measurements via Large-scale WiFi Networks. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16. (ACM, New York, NY, USA), pp 316–327.

14. Scanlon P (2016) Identifying the impact of friends on their peers academic performance. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp 1368–1369.

15. González MC, Hidalgo CA, Barabási A-L (2008) Understanding individual human mobility patterns. *Nature* 453(7196):779–782.

16. Zheng Y, Capra L, Wolfson O, Yang H (2014) Urban Computing: Concepts, Methodologies, and Applications. *ACM Trans Intell Syst Technol* 5(3):38:1–38:55.

17. Blumenstock J, Cadamuro G, On R (2015) Predicting poverty and wealth from mobile phone metadata. *Science* 350(6264):1073–1076.

18. Bogomolov A, et al. (2014) Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data. *Proceedings of the 16th International Conference on Multimodal Interaction*, ICMI '14. (ACM, New York, NY, USA), pp 427–434.

19. Ruiz-Ruiz AJ, Blunck H, Prentow TS, Stisen A, Kjærgaard MB (2014) Analysis methods for extracting knowledge from large-scale WiFi monitoring to inform building facility planning. *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp 130–138.

20. Sundsøy P (2016) Can mobile usage predict illiteracy in a developing country? Available at: https://arxiv.org/abs/1607.01337 [Accessed October 2, 2018].

21. Redondi AE, Cesana M, Weibel DM, Fitzgerald E (2016) Understanding the WiFi usage of university students. *7th IEEE International Workshop on TRaffic Analysis and Characterization (TRAC)*, pp 44–49.

22. Sevtsuk A (2009) Mapping the MIT Campus in Real Time Using WiFi. *Handb Res Urban Inform Pract Promise Real-Time City*:326–338.

23. Huang H, Gartner G, Krisp JM, Raubal M, Weghe NV de (2018) Location based services: ongoing evolution and research agenda. *J Locat Based Serv* 0(0):1–31.

24. Sacerdote B (2001) Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Q J Econ* 116(2):681–704.

25. Scanlon P, Smeaton AF (2017) Using WiFi Technology to Identify Student Activities Within a Bounded Environment. *Data Driven Approaches in Digital Education*, Lecture Notes in Computer Science., eds Lavoué É, Drachsler H, Verbert K, Broisin J, Pérez-Sanagustín M (Springer International Publishing), pp 473–478.

26. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci U S A* 106(51):21544–21549.

27. Young JK (2013) Repression, Dissent, and the Onset of Civil War. *Polit Res Q* 66(3):516–532.

28. Anagnostopoulos A, Kumar R, Mahdian M (2008) Influence and Correlation in Social Networks. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’08. (ACM, New York, NY, USA), pp 7–15.

Supplemental Material

The Empirical Context

This study uses data from a top engineering school in Europe. Students at this institution are awarded a standard undergraduate degree after three years of study. Most students continue an additional 1-2 years to complete a Masters degree. Wifi access in this school began in 2001 with a pilot project that deployed only a few access points across campus. Later in the Fall 2006 the Wifi network evolved into a campus-wide network with 205 access points, located in classroom buildings, student study rooms, cafes, and other research and study areas on campus. Students could access the Internet using their laptops from any hotspot on campus. The Wifi network is part of the Eduroam network, which provides roaming Internet access at participating institutions across Europe, and increasingly around the world. Eduroam relies on a distributed authorization protocol called Remote Authentication Dial-In User Service (RADIUS), which also provides the network accounting data used in this study.

Data and Descriptive Statistics

The data used in this paper come from two sources: i) students administrative records including grades and courses completed per semester; ii) students Wifi usage data. The RADIUS protocol for Wifi service provides logs with the duration of each session (time wise) and the number of bytes sent and received. Using these data, we created an anonymized panel on the performance of students and their Wifi usage on campus on a semester basis. This panel covers the activity of 3,030 Wifi-using students, spanning 5 consecutive semesters (Fall 2006 to Fall 2008), resulting in 6,425 student-semester records. Our key variables (Wifi usage and grades) are standardized to facilitate interpreting results. Table S1 shows all relevant variable definitions.

Our outcome of interest is the students' academic performance. We have two measures for it: i) the number of courses completed per semester, that is, a student that attempts 6 courses in a semester but obtains a passing grade only for 4 registers 4 for this outcome variable; ii) the cumulative grade points obtained in a semester across successfully completed courses, that is, a student that completes 4 courses in a semester with grades 11, 13, 12 and 15 registers 51 points for this outcome variable. The first outcome measure proxies the amount of work successfully completed, while the second measure captures the quality of that work (grades can go up to 20 points per course). On average, students complete 4 courses and obtain 53 grade points per semester.

A measure of pre-university performance is given by the Application Score. This is the weighted average of the grades obtained in 12^{th} grade and of the scores obtained in national field exams, such as Physics and Math. This score is used to determine admission to the University and is comparable to the Scholastic Assessment Test (SAT). Wifi usage is measured per semester by adding all Wifi sessions of the same student in that semester. On average, students connect to Wifi for 125 minutes per semester, transferring about 11 GB of data. Additional summary statistics are provided in Table S2.

Empirical Strategy

Propensity Score Matching and Balance Checks

We analyze only students that use Wifi to capture the effect of Wifi usage *per se*. This reduces our concerns with unobserved effects related to both the student's ability to connect to the Wifi network and their performance, such as owning a laptop or lack of knowledge on how to connect and use Wifi on campus. Furthermore, we use Dynamic Propensity Score Matching (DPSM) to identify the effect of Wifi usage on the performance of students. Matching reduces the pre-treatment differences in observed covariates across students with low and high Wifi usage, thus lowering the likelihood that they drive the observed results. Matching also reduces model dependence, which is helpful when one cannot control for all desired covariates (1). DPSM is an extension of PSM to panel data (2). In our case, it is performed by matching control and treated students separately for each semester and then combining semesters to assemble a matched panel. In this way, treated students always match to control students in the same semester, effectively controlling for semester, but we do not require the same treated student to be always matched to the same control student, which increases our ability to match and thus reduces bias.

The treatment and control groups are defined as following. A student belongs to the *treatment group* in a specific semester if her Wifi usage is in the upper quintile of Wifi usage in this semester. A student belongs to the *control group* in a given semester if instead her Wifi usage is in the lower quintile of Wifi usage of that semester. Therefore, students in the control group are able to connect to Wifi but use only a trivial amount of it.

We estimate propensity scores using a probit model, controlling for the student's Application Score, her major, cohort and curricular year.

Fig. S1 shows the reduction in bias in percentage terms by covariate and per semester for a model where we assess performance using Grade Points and Wifi usage by the number of hours connected (other models exhibit similar reduction in bias). The reduction in bias is positive when matching reduces the bias relative to the original world and therefore Fig. S1 shows that matching works very well in our setting. Furthermore, Table S3 shows the total number of observations and the number of observations used for the analysis after matching. Our matching procedure drops only a few of the treated observations (<2%), allowing us to still claim that our results are internally valid.

First Differences and Additional Specifications

The same student appears in our dataset in multiple consecutive semesters. Therefore, we can use first-differences to estimate the effect of Wifi usage on campus on the performance of students controlling for time-invariant unobserved student characteristics. Our initial model is defined as follows:

$$(1)\ \ P_{i,t} = \beta_0 + W_{i,t}\beta_W + Z'_{i,t}\beta_Z + U_i\beta_U + \epsilon_{i,t}$$

where $P_{i,t}$ represents student i's performance in semester t. $W_{i,t}$ represents student i's Wifi usage in semester t, and $Z_{i,t}$ is a vector of student, semester, and student-semester covariates. The term U_i represents student-specific time-constant effects. Table S4 provides additional information on how P, W and Z are operationalized in practice.

We use *Application Score* to control for prior student performance, and thus proxy some unobserved attributes of the students, such as aptitude and socioeconomic status. We note that this covariate drops from a first-differences specification given that it does not change over time for each student. However, we forced it into our regressions to still control for potential (non-linear) effects associated with these unobserved attributes. Hence, in this case, the coefficient on *Application Score* is interpreted as the average effect of this covariate on the rate of change in performance across two consecutive semesters.

We use *Cohort* to control for differences in the overall performance of students across cohorts. We add dummy variables for *Curricular Year*, *Major*, and *Semester* to control, respectively, for variations in grades across different curricular years, differences in major difficulty or grading policies, and other unknown differences that arise over time that apply similarly to all students on campus.

Taking first-differences eliminates student-specific time-constant effects, Ui, and yields the model below, where $\Delta P_{i,t} = P_{i,t} - P_{i,t-1}$, and so forth. The effect of Wifi usage on student performance is given by β_W in this model:

$$(2)\ \ \Delta P_{i,t} = \Delta W_{i,t}\beta_W + Z'_{i,t}\beta_Z + \Delta\epsilon_{i,t}$$

We use several additional models based on this first-differences specification, as detailed below.

<u>The Effect of Day and Night Usage:</u> Let $D_{i,t}$ be the total Wifi usage of student i in semester t that occurred between 8 a.m. to 8 p.m. (daytime). Let $N_{i,t}$ be the total Wifi usage for student i during semester t that occurred between 8 p.m. to 8 a.m. (nighttime). Thus, $D_{i,t} + N_{i,t} = W_{i,t}$. We test whether Wifi usage during the day and during the night have different effects on the performance of students using the following model:

$$(3)\ \ \Delta P_{i,t} = \beta_0 + \Delta D_{i,t}\beta_d + \Delta N_{i,t}\beta_n + Z'_{i,t}\beta_Z + \Delta\epsilon_{i,t}$$

<u>The Effect of Advanced Curricular Year</u>: In this model, we interact advanced curricular years (year 3 and up) with Wifi usage to control for potential differences in the effect of Wifi for

different stages of the students' studies. For instance, more mature students, such as juniors and seniors, may be more productive than freshman and sophomores in how they use Wifi, due to, for example, the experience that they accumulate during their studies. We interact Wifi usage with the dummy variable *Advanced*, which takes the value 0 when the student is in curricular years 1 and 2, and 1 otherwise.

Controlling for Time on Campus: In one specification, we try to control for the time that students spend on campus, which might be correlated to both Wifi usage and performance through unobserved effects (e.g., dedication). For each student in each day we compute the time elapsed between the first and the last Wifi session in that day, thus proxying the amount of time spent on campus. We then average these times across the semester for each student.

Results Obtained Using First Differences

All coefficients in our first-differences results are normalized by the standard deviation over the set of observations included in each regression, and thus may be interpreted in terms of percent changes of standard deviations. For example, a coefficient of 0.123 on independent variable x is interpreted as an increase of one standard deviation in x is associated with a 12.3% increase of a standard deviation in the dependent variable.

Basic Model Results

Our first-differences results show a positive statistically significant correlation between Wifi usage and students' performance. Table S5 reports the regressions results for both Number of Courses and Grade Points measuring Wifi usage both in terms of time (Hours) and volume (Megabytes). Performance increases 15%-16.2% of a standard deviation, for an increase of 1 standard deviation in Wifi usage, when the latter is measured in time. This statistic becomes 2.6%-3.4% when Wifi usage is measured in volume.

Note that there is a negative association between *Application Score* and performance in these models. This is expected because while prior performance is positively correlated with current performance, it is negatively usually correlated with the student's marginal ability to change her performance (i.e., the higher a score the harder to improve it).

The Effect of Day and Night Wifi Usage

We examine the effect of day and night when we split Wifi usage into day and night usage. Table S6 shows that the average effect of Wifi usage is similar to that obtained from usage during the day, i.e., day time Wifi usage is more productive for performance than night time usage. These results seem to imply that students that connect to the Internet on campus during the day appear to be more productive than their "night owl" colleagues.

The Effect of Advanced Curricular Years

Using our original first-differences model, we regress our performance measures on the interaction of $W_{i,t}$ and $Advanced_{i,t}$ These models estimate the impact of $W_{i,t}$ while students progress from an early stage in their studies (curricular years 1 and 2) to more advanced phases (curricular year 3 and above). Table S7 shows the results of these regressions, which exhibit a significantly higher effect of Wifi on the performance of advanced students, when usage is measured by the duration of connectivity. This difference is significant for both Number of Courses and Grade points. The interaction between advanced curricular years and the volume of Wifi usage is also positive but our specification is underpowered in this case to obtain statistical significance.

The reported increase in the effect of Wifi usage from freshmen and sophomores to juniors and seniors suggests that Wifi may be increasingly useful as students gain more maturity in their academic program.

Controlling for Time on Campus

Table S8 shows the results obtained controlling for our proxy for time on campus. As expected, the coefficients on Wifi usage reduce in magnitude compared to those obtained with our original specification, and time on campus is positively associated to performance, meaning that including this covariate is likely controlling for additional unobserved effects that are simultaneously positively correlated with Wifi usage and performance. We still observe the positive significant relationship between Wifi usage in terms of duration and performance, although our specification is again underpowered when we measure Wifi usage in terms of volume.

Network of Collocated Peers

In our network of collocated peers, there is an edge between two students if they share a 5-minutes session at the same access point. The weight on each edge represents the number of shared sessions of 5-minutes (for example, there is an edge with weight 3 between two students that share between 15 and 20 consecutive minutes at the same access point and shared sessions that are less than 5 minutes long do not affect these weights). Thus, the edge weight indicates the strength of the collocation relationship across peers. The resulting network includes 5 sub-networks, one per semester. Our overall network includes 3,942,550 edges between 12,486 students that use Wifi on campus.

We focus our analysis on the "top collocated peer" – the peer with which a student spends the most time with while using Wifi on campus. More precisely, the "top collocated peer" of student i is student j such that w(i,j)=max_k{w(i,k)}, where k loops over all students connected to student i and w(m,n) is the weight on the edge between students m and n. By definition, the "top collocated peer" is the student with which one spends most time with while using Wifi on campus and thus this student may have the highest potential to affect grades.

Shuffle Test

In order to identify the effect of peers on performance and to disentangle it from other sources of social correlation, we utilize a randomization technique known as the Shuffle Test.

Randomization has been shown to be effective at identifying peer influence in the presence of homophily and confounding factors (3). The basic idea behind this method is to shuffle the social network in a way that is orthogonal to the attributes under investigation. For example, if student performance and peer influence are independent from access point location, we can shuffle sessions among access points to break social ties without introducing misleading bias in the analysis. Since social links are defined by contemporaneous usage (same time and place), randomizing sessions by access point location effectively breaks the social connections among peer students, while retaining the behavior explained by temporal usage patterns.

We effectively argue that randomization produces an alternate world, called a pseudosample, in which students exhibit similar Wifi usage patterns (due to the homophily that drives when and how they choose to connect to the Wifi network) but zero peer influence, by construction (4). Any correlation between one's performance and that of her peers in the pseudosample cannot arise from peer influence; and, therefore, the effect of peer influence can be estimated as the difference between the correlation in performance in the real world and that obtained using randomization. Simulation over many pseudosamples yields a distribution for the underlying correlation in performance across neighbors, and the causal peer effect may thus be estimated as the difference between the peer effect obtained with the original data and the mean of the peer effects obtained from such (randomized) simulations.

The correlation between own Grades $P_{i,t}$ and Neighbor Grades $N_{i,t}$ is denoted by β_N in model

$$P_{i,t} = \beta_0 + N_{i,t}\beta_N + Z'_{i,t}\beta_Z + \epsilon_{i,t}$$

Let $E[\beta'_N]$ represent the mean of the distribution of pseudosample coefficients, i.e., each β'_N is obtained from running the model above in each pseudosample. Since $E[\beta'_N]$ estimates the correlation between one's performance and that of her peers in networks without peer influence, then $\beta_N - E[\beta'_N]$ measures the effect of peer influence in performance in the original network. In this case, we use a liner model without normalization. Therefore, coefficients are interpreted in percentage terms, i.e., a coefficient of 0.14 indicates, for example, that when the average Grade Points of the top collocated peer increases by 1 point the average Grade Points of the focal student increases by 0.14 points.

Shuffling has some potentially undesirable consequences, such as placing pseudosampled students in locations they would never visit in real life. We control for this by restricting shuffles to Wifi sessions of the same student, during the same semester, and in the same building. This way, each student has pseudosampled sessions only in places where she has been in real life in that semester.

Peers Effect Results and Additional Robustness Tests

Baseline Peer Effect Model

We first examine the average effect of the "top collocated peer" over all students in our sample. Fig. S2 shows the results obtained. This effect is positive, between 14%-15%, which gives us confidence that a measurable, non-trivial peer effect exists in our setting.

Heterogeneous Peer Effects

In the paper, we observe that low performance students exhibit greater peer effects in later curricular years, which leads to significant policy implications. Therefore, we would like to investigate the robustness of this result. In the paper, low performance and high performance students have an Application Score below and above the median, respectively. For robustness sake, we define low performance students as those with an Application Score in the first tercile of the distribution of Application Scores and high performance students as those with an Application Score in the upper two terciles of this distribution. Fig. S3 shows that results do not change when we do so. Note that in the lower curricular years, low and high performance students show statistically similar peer effects, although the former show a non-zero peer effect. In later curricular years, we observe a statistically different peer effect for low and high performance students, with the former exhibiting a large, positive, statistically significant effect. Fig. S4 shows that our results also remain unchanged when we split students into the lower quartile and the upper three quartiles of the distribution of Application Scores.

Results without Access Points Serving Classrooms

A key concern in our work so far is that we may be detecting effects over “uninteresting” collocated relationships, namely those that may arise from Wifi usage during classes. In fact, all students in the same classroom may use the same Wifi access point(s) for the duration of their (joint) classes and thus will all be named “collocated peers”. However, these collocated relationships may not really capture the gist of those that organically arise across students due to working in teams or sharing common learning spaces besides classrooms. The collocated relationships that arise from simultaneous use of Wifi at these access points are instead likely to simply capture class attendance. To address this concern we remove Wifi sessions at access points that serve classrooms (we remove these sessions from the original data and therefore our pseudosamples do not include them either). Fig. S5 shows the effect of the performance of the “top collocated peer” for all students in all curricular years irrespective of prior performance. Although smaller in magnitude than the original results, we still observe a positive effect on both Number of Courses and Grade Points. Without surprise, removing Wifi sessions at these access points from the data may also reduce the strength of the true relationships among students, which would in turn affect our statistical power to identify peer effects.

For robustness sake, we also check results splitting students according to prior performance and curricular year (still without Wifi sessions at access points serving classrooms). Fig. S6 shows the results obtained. Low performance students exhibit no peer effect in early years and a large positive peer effect for more advanced students. The latter is statistically larger than the peer effect observed for high-performing students when performance is measured by Grade Points. Therefore, these results strengthen our original findings. Even when we remove in-class Wifi usage from the analysis, positive peer effects still arise and pairing advanced students results only in minimal harm for the high-performing ones.

Results with Lagged Performance Models

Another concern in our empirical analysis is with simultaneity in the performance of students in the same semester. The idea behind this concern is that the grades of students connected in our network may be jointly determined by some unobserved. The Shuffle Tests used before already address this concern. In any case, we now explicitly regress one student’s grades on the grades that her “top collocated peer” obtained in the prior semester. Fig. S7 shows that also in this case there is a significant positive peer effect across all measures of performance, with and without classrooms. Therefore, these results are consistent with our original results, increasing our confidence in our findings.

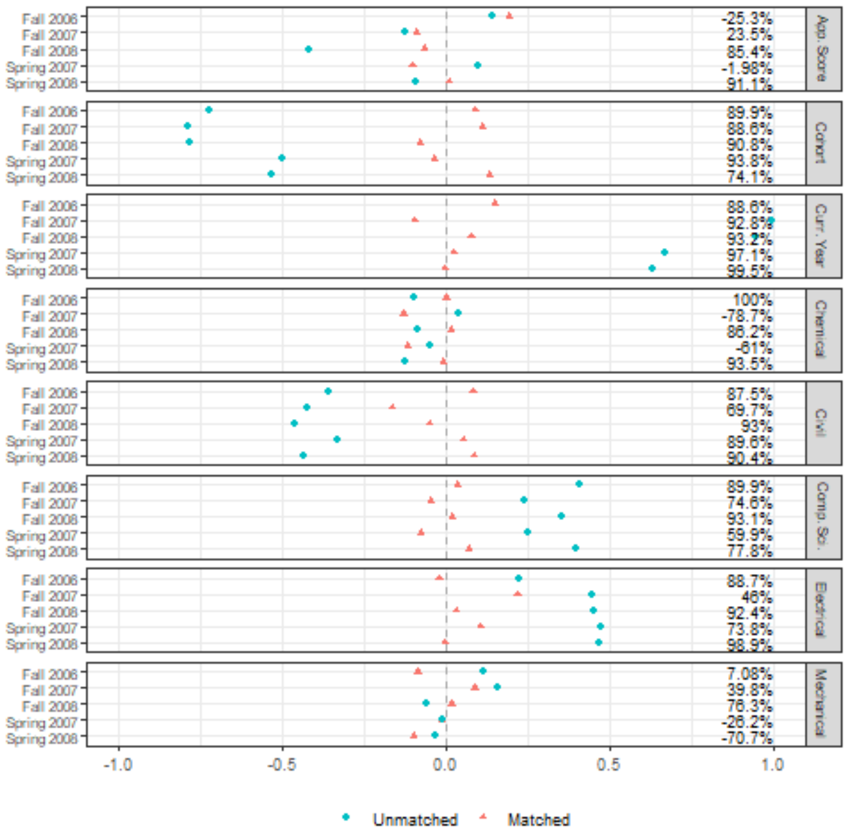

Fig. S1. Percent bias reduction (relative to largest absolute bias) obtained using Dynamic PSM for the Grade Points-Hours model (similar reductions in bias obtained for other models in the paper).

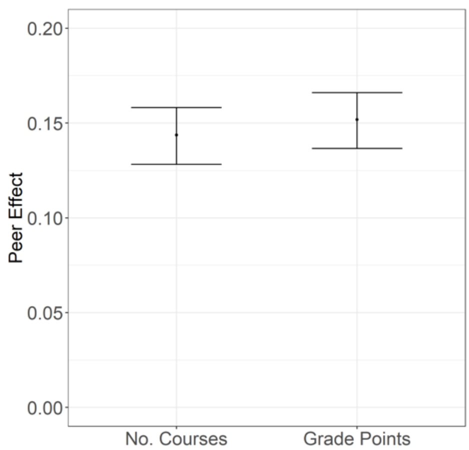

Fig. S2. The effect of the performance of the "Top Collocated Peer" on student's performance (the top peer of a student is the student with which she spends the most 5-minute shared sessions while using Wifi on campus).

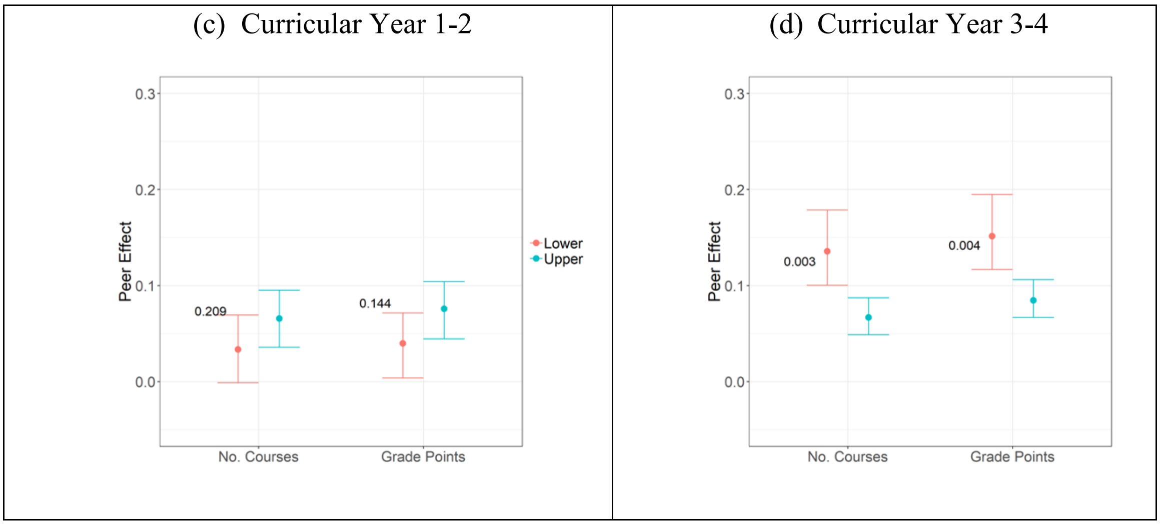

Fig. S3. Heterogeneous effect of the "Top Collocated Peer" per prior performance. Upper is the subset of students in the upper tercile of application average grades. Lower is the subset of students in the lower two terciles of application average grades.

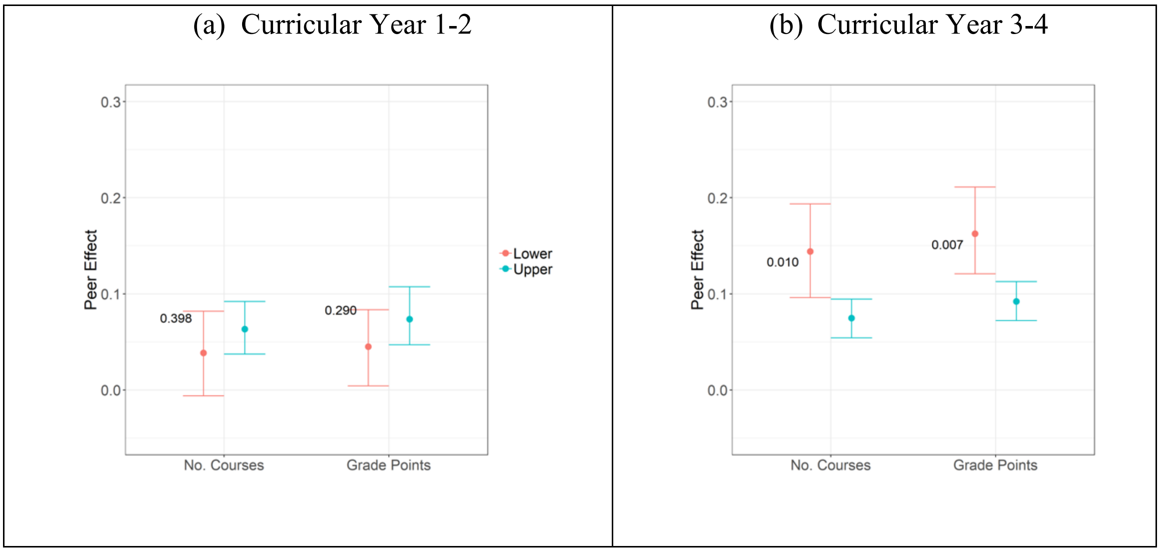

Fig. S4. Heterogeneous effect of the "Top Collocated Peer" per prior performance. Upper is the subset of students in the upper quartile of application average grades. Lower is the subset of students in the lower three quartiles of application average grades.

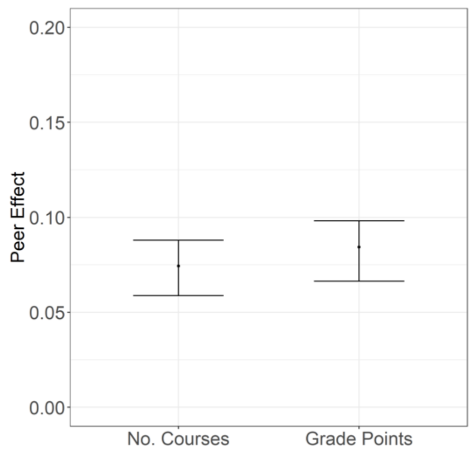

Fig. S5. The effect of the performance of the "Top Collocated Peer" on student's performance ignoring Wifi sessions at access points that serve classrooms.

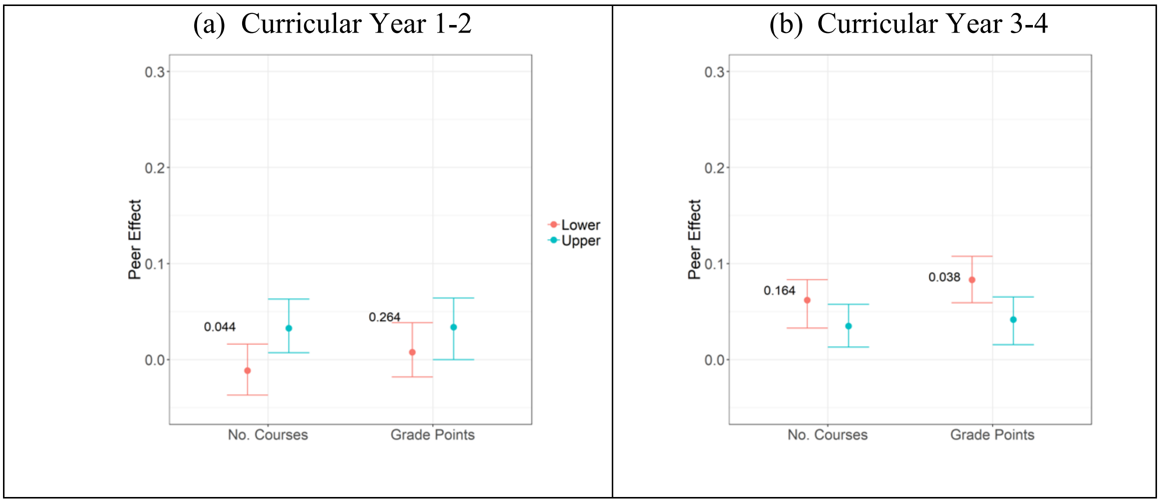

Fig. S6. Heterogeneous effect of the "Top Collocated Peer" per prior performance ignoring Wifi sessions at access points that serve classrooms. Upper is the subset of students in the upper halve of application average grades. Lower is the subset of students in the lower halve of application average grades.

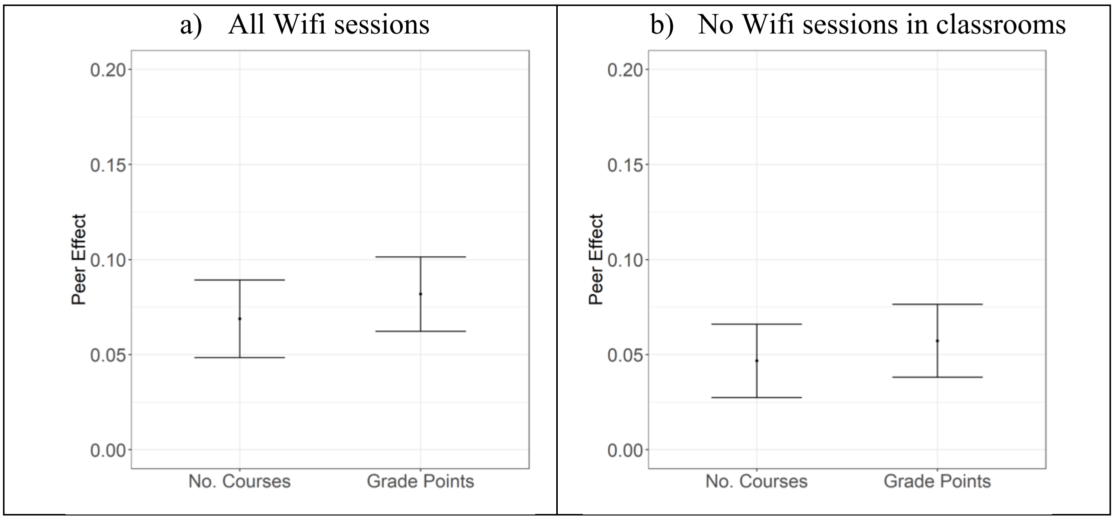

Fig. S7. The effect of the performance of the "Top Collocated Peer" using lagged performance for this peer. Effects using data from all Wifi sessions (a) and only from Wifi sessions at access points that do not serve classrooms (b).

Table S1. Description of covariates used in this study.

Name	Description
Student ID	Anonymized student identifier
Semester	Index for semester
Grade Points	Cumulative grade points earned per semester
No. Courses	Number of courses completed per semester
Total Hours	Total hours spent online per semester
Total Megabytes	Total megabytes transferred per semester
Hours (Day)	Daytime (8a-8p) hours online per semester
Megabytes (Day)	Daytime megabytes transferred per semester
Hours (Night)	Night time (8p-8a) hours online per semester
Megabytes (Night)	Night time megabytes transferred per semester
Application Score	Score for admission to the University
Cohort	Year that the student entered the university
Major Dummies	Indicator for engineering major
Advanced	Indicator for curricular years 3 and above

Table S2. Summary statistics of student-semester observations.

	Mean	SD	Median	Min	Max
Number Courses	3.980	1.833	4	1	22
Grade Points	52.825	26.033	54	10	332
Wifi Usage - Hours	125.518	149.212	69.187	0.009	1644.909
Wifi Usage - Hours (Day)	106.212	125.354	58.434	0	1007.670
Wifi Usage - Hours (Night)	19.306	42.494	3.435	0	683.096
Wifi Usage - Megabytes	10838.175	26454.104	2526.549	0.006	422173.53
Wifi Usage - Megabytes (Day)	9277.57	23347.86	2122.833	0	359161.64
Wifi Usage - Megabytes (Night)	1560.605	5847.301	62.393	0	157376.29
Application Score	149.088	17.037	148.8	97.5	199
Curricular Year	3.282	1.284	3	1	5
Cohort	2003.878	1.984	2004	1999	2007

Table S3. Number of observations selected by the matching algorithm.

	Number Courses		Grade Points	
	Hours	Megabytes	Hours	Megabytes
Treated - All	2059	2104	2059	2104
Treated - Matched	2020	2086	2020	2086
Control - All	1898	1871	1898	1871
Control - Matched	832	917	817	927

Table S4. Covariates used in our models.

Model Variable		Description
$P_{i,t}$	$GP_{i,t}$	Grade Points
	$NC_{i,t}$	Number Courses
$W_{i,t}$	$HR_{i,t}$	Hours Online
	$MB_{i,t}$	Megabytes Transferred
$Z_{i,t}$	AS_i	Application Score
	CH_i	Cohort Dummies
	$CY_{i,t}$	Curricular Year Dummies
	MJ_i	Major Dummies
	SM_i	Semester Dummies
	U_i	Unobserved Effects

Table S5. Results obtained using our baseline first-difference models.

	(1)	*(2)*	*(3)*	*(4)*
	ΔNo. Courses	*ΔNo. Courses*	*ΔGrade Points*	*ΔGrade Points*
ΔHours	0.150^{***} (0.016)		0.162^{***} (0.017)	
ΔMegabytes		0.034^{**} (0.015)		0.026^{***} (0.016)
App Score	-0.030^{***} (0.011)	-0.032^{***} (0.011)	-0.045^{***} (0.012)	-0.047^{***} (0.011)
Constant	-0.087 (0.139)	-0.087 (0.137)	-0.132 (0.136)	-0.129 (0.134)
Observations	5,337	5,337	5,337	5,337
Adjusted R^2	0.15	0.13	0.17	0.14

Notes:

1. Dummy variables for Cohort, Curricular Year, Major and Semester included.
2. Clustered robust standard errors shown in parentheses (clustered on Student ID).
3. **** $p<0.01$; ** $p<0.05$; * $p<0.1$*

Table S6. Results obtained partitioning Wifi usage by day and night.

	(1)	*(2)*	*(3)*	*(4)*
	ΔNo. Courses	*ΔNo. Courses*	*ΔGrade Points*	*ΔGrade Points*
ΔHours (Day)	0.137^{***} (0.017)		0.151^{***} (0.018)	
ΔHours (Night)	0.027 (0.017)		0.023 (0.017)	
ΔMegabytes (Day)		0.032^{**} (0.017)		0.023 (0.017)
ΔMegabytes (Night)		0.005 (0.015)		0.006 (0.015)
App Score	-0.030^{***} (0.011)	-0.032^{***} (0.011)	-0.046^{***} (0.012)	-0.047^{***} (0.011)
Constant	-0.087 (0.139)	-0.088 (0.137)	-0.131 (0.136)	-0.129 (0.134)
Observations	5,337	5,337	5,337	5,337
Adjusted R^2	0.15	0.13	0.17	0.17

Notes:

1. Dummy variables for Cohort, Curricular Year, Major and Semester included.
2. Clustered robust standard errors shown in parentheses (clustered on Student ID).
3. **** $p<0.01$; ** $p<0.05$; * $p<0.1$*

Table S7. Results obtained interacting Advanced Curricular Years with Wifi usage.

	(1) *ΔNo. Courses*	*(2)* *ΔNo. Courses*	*(3)* *ΔGrade Points*	*(4)* *ΔGrade Points*
ΔHours	0.085^{**} (0.036)		0.081^{**} (0.036)	
ΔMegabytes		-0.001 (0.032)		-0.013 (0.029)
Advanced	-0.176^{***} (0.018)	-0.155^{***} (0.018)	-0.162^{***} (0.017)	-0.138^{***} (0.017)
*ΔHours*Advanced*	0.068^{*} (0.037)		0.088^{**} (0.037)	
*ΔMegabytes*Advanced*		0.042 (0.033)		0.048 (0.031)
App Score	-0.060^{***} (0.012)	-0.056^{***} (0.011)	-0.067^{***} (0.012)	-0.069^{***} (0.012)
Constant	0.473^{**} (0.143)	0.469^{**} (0.141)	0.502^{***} (0.140)	0.499^{***} (0.137)
Observations	5,337	5,337	5,337	5,337
Adjusted R^2	0.09	0.07	0.11	0.08

Notes:

1. Dummy variables for Cohort, Major and Semester included.
2. Clustered robust standard errors shown in parentheses (clustered on Student ID).
3. **** p<0.01; ** p<0.05; * p<0.1*

Table S8. Results obtained controlling for time on campus.

	(1) *ΔNo. Courses*	*(2)* *ΔNo. Courses*	*(3)* *ΔGrade Points*	*(4)* *ΔGrade Points*
ΔHours	0.131^{***} (0.017)		0.140^{***} (0.018)	
ΔMegabytes		0.015 (0.015)		0.006 (0.016)
ΔCampus Hours	0.044^{***} (0.016)	0.100^{***} (0.016)	0.049^{***} (0.016)	0.111^{***} (0.015)
App Score	-0.033^{***} (0.011)	-0.035^{***} (0.011)	-0.048^{***} (0.012)	-0.051^{***} (0.011)
Constant	-0.052 (0.139)	-0.034 (0.137)	-0.099 (0.136)	-0.076 (0.134)
Observations	5,313	5,313	5,313	5,313
Adjusted R^2	0.15	0.14	0.17	0.15

Notes:

1. Dummy variables for Cohort, Curricular Year, Major and Semester included.
2. Clustered robust standard errors shown in parentheses (clustered on Student ID).
3. **** p<0.01; ** p<0.05; * p<0.1*

References

1. Rubin DB (1973) Matching to Remove Bias in Observational Studies. *Biometrics* 29(1):159–183.

2. Young JK (2013) Repression, Dissent, and the Onset of Civil War. *Polit Res Q* 66(3):516–532.

3. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci U S A* 106(51):21544–21549.

4. La Fond T, Neville J (2010) Randomization Tests for Distinguishing Social Influence and Homophily Effects. *Proceedings of the 19th International Conference on World Wide Web*, WWW ’10. (ACM, New York, NY, USA), pp 601–610.