# The rich club phenomenon in the classroom

Luis M. Vaquero[1] & Manuel Cebrian[2,3]

[1]Hewlett-Packard Laboratories, Bristol BS34 8QZ, UK, [2]NICTA, Melbourne, Victoria 3010, Australia, [3]Department of Computer Science and Engineering, University of California at San Diego, La Jolla, CA 92093, USA.

We analyse the evolution of the online interactions held by college students and report on novel relationships between social structure and performance. Our results indicate that more frequent and intense social interactions generally imply better score for students engaging in them. We find that these interactions are hosted within a "rich-club", mediated by persistent interactions among high performing students, which is created during the first weeks of the course. Low performing students try to engage in the club after it has been initially formed, and fail to produce reciprocity in their interactions, displaying more transient interactions and higher social diversity. Furthermore, high performance students exchange information by means of complex information cascades, from which low performing students are selectively excluded. Failure to engage in the rich club eventually decreases these students' communication activity towards the end of the course.

More than 1.2 million students drop out of school every year in the U.S., one every 26 seconds[1]. Year 2007 dropouts will cost more than $300 billion in lost wages, taxes and productivity to the U.S. Dropouts contribute about $60,000 less in federal and state income taxes. Each cohort of dropouts costs the U.S. $192 billion in lost income and taxes[2]. A dropout student is more than 8 times as likely to be in jail or prison as a high school graduate and nearly 20 times as likely as a college graduate[3].

Early detection of poor performance will allow more time to take corrective actions and will likely help to reduce the number of dropouts. Therefore, it is of the utmost importance to be able to assess the performance of students in a continuous manner.

Computer science is not unaware of this need for close follow up of students. Computer Supported Collaborative Learning (CSCL) is a branch of computer science that intersects with pedagogy and social sciences. Indeed, one of the goals of CSCL is to explore appropriate methods/tools for evaluating collaboration so that more insight can be gained into the results of lecturing/teaching procedures[4].

However, systematic gathering and analysis of educational data *in-natura* has only recently started. So far this analysis has mainly tried to determine static structural features of the social learning network formed by the students. For instance, Nurmela *et al.* looked at the structure of the interactions trying to determine the central actors in a CSCL environment[5]. In this social structure, "key communicators" were assumed to be the most connected individuals in time-aggregated networks[6]. Similar analyses were carried out by Martínez *et al.*[7] and Chen and Watanabe, who focused on other structural parameters that are important for the final score: group structure, member's physical location distribution, and member's social position[8].

Beyond this merely static structural analysis, the literature also highlights the key role of student interaction for effective learning. At a societal scale, Granovetter's pioneering work[9] recognised the importance of interaction patterns and proposed his well-known "strength of weak ties" phenomenon, where he hypothesised that isolated social ties offer limited access to external prospects, while heterogeneous social ties diversify one's opportunities.

While the relevance of the social network structure and interactions has been widely recognised in the educational context[10], some other factors have recently been under the spotlight, e.g. social acceptance or willingness to communicate[11]. In general, it is not just about knowing "who" the students interact with, but "how" and "when" they do it and, importantly, what is the result of these interactions with regards to the educational outcome[12].

Preliminary answers to the "how" question come from different works. The effects of analysing the relationships between web forum users on the structure of the network (reconstructed from the messages sent) were studied in[13,14]. Also, the type of interaction or content being exchanged have been considered[6,16]. However, these previous analyses were based on a static snapshot of the structure and interactions of the network at some point in

time or included a reduced number of samples. For instance[7], analysed these macroscopic metrics in the four different assignments the course was structured in ($\simeq$ once a month).

Acquiring full knowledge on "how" students interact would be facilitated by having access to dynamic interactions and their changes with time. Timing is a determinant element to understand the correspondence between student behaviour and performance. Therefore, this paper tries to determine the individual and group-level behavioural patterns that lead to low scoring and possible dropout. Gaining insight into these data could help in identifying "groups at risk", enabling educators to act sooner and hopefully reduce dropout rates.

The rest of this paper is organised as follows. Next section presents the main results obtained from our analysis. This is followed by a broader discussion.

## Results

We analysed a record of college student interactions and compared social interaction data with the academic scores of the students (see

third paragraph of Course Details in Methods in the Supplementary Information (SI) for a concrete definition on what an interaction is in this context) and how this relationship evolves with time. To this end, we analysed records of 80, 000 interactions by 290 students - approximately 16 times more interactions with almost 3 times more students than previous studies on educational networks *in natura*[5–8,10,12,15]. Even so the data can still be considered to be sparse ($\simeq$ 4.6 interactions per person per day). This sparseness is partly due to the fact that our work does not include verbal in classroom interactions or other communication mechanisms, like discussion groups that are typical in most universities.

Figure 1A shows a snapshot of the social graph for one of the classes being analysed. Supplementary video S1 offers a complete weekly sequence of interactions between students in one of the courses we analysed.

**Diversity and assortativity analysis.** Our first finding is that, in this environment, social diversity is negatively correlated with performance. This is explained by our second finding: high
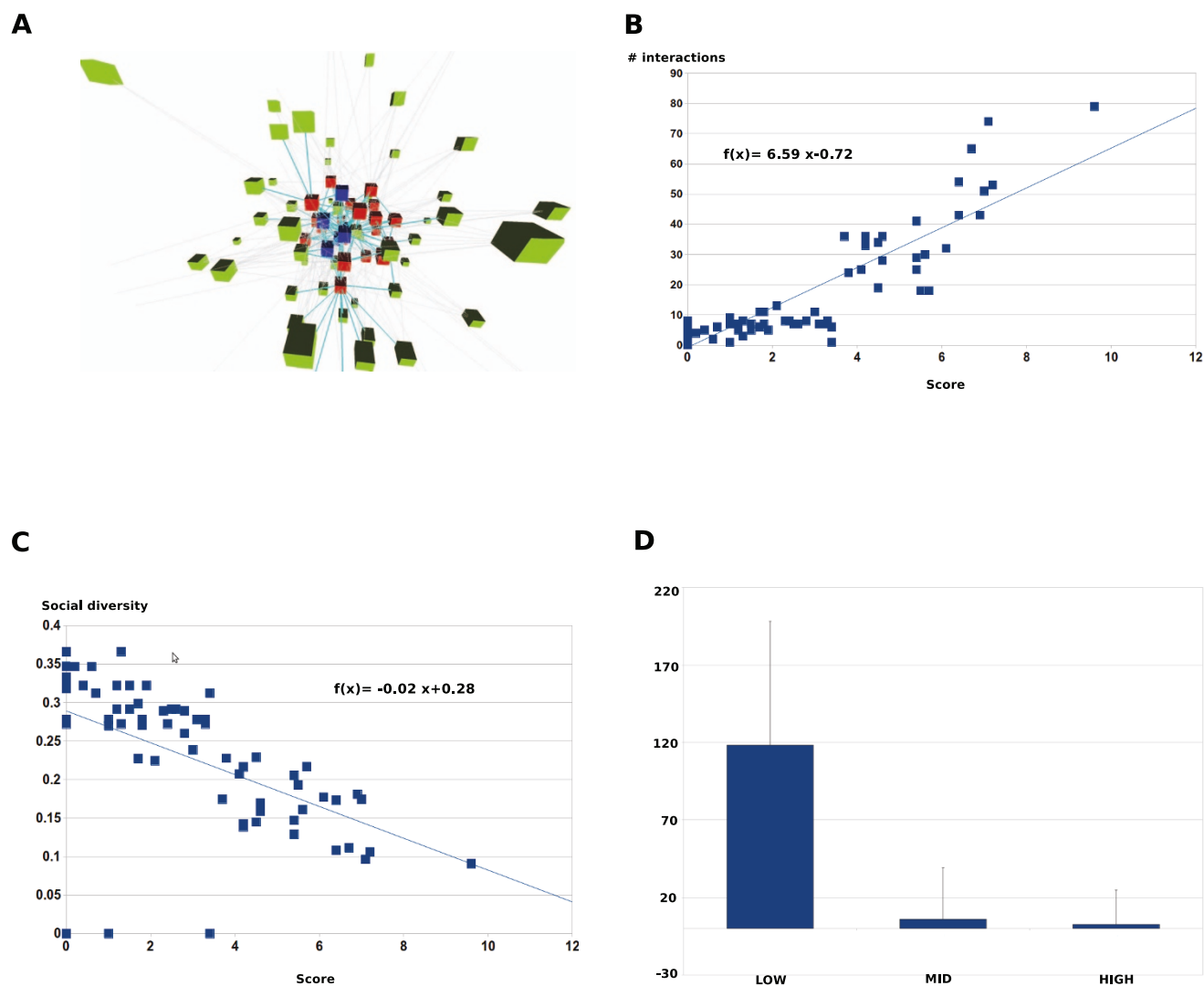


**Figure 1 | Diversity and Assortativity Analysis.** (A) shows a graph of one of the analysed courses including 82 students at the end of the last week of the course. Continuous thick blue edges indicate persistent interactions while dotted thin grey edges indicate transient interactions. High performing students are shown in dark blue, mid performing ones in red and low performing ones in green. As can be observed, high performance students form a "core" where the highest density of persistent interactions can be observed. Low performance students remain in the periphery of the graph, mainly holding transient interactions. (B) Scatter plot and linear regression for one of the variables analysed (number of interactions) vs. scoring in one of the classes ($R^2 = 0.72$). (C) Scatter plot and linear regression for social diversity vs. scoring in one of the classes ($R^2 = 0.12$). (D): Ratio of transient to persistent interactions obtained for different groups of students with different levels of interaction (LOW, MID, HIGH).
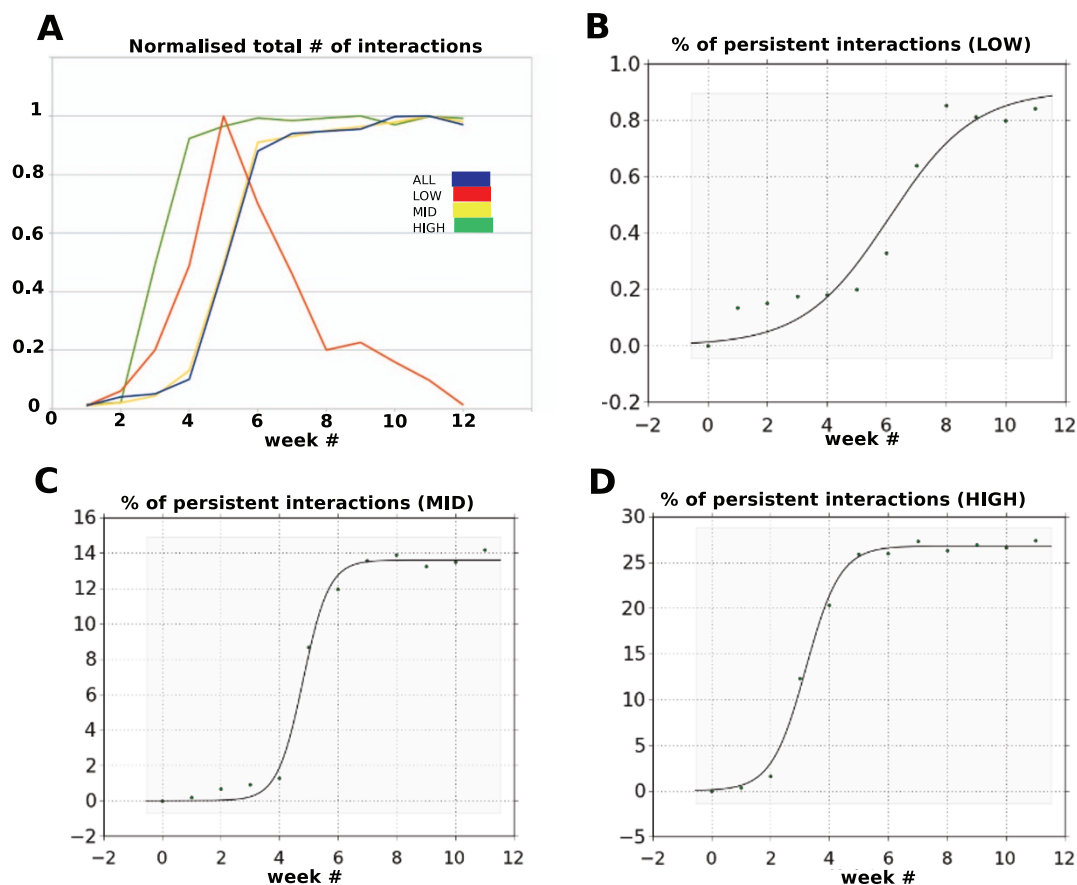
**Figure 2 | Persistent Interaction Analysis.** (A) Temporal Evolution of the total number of interactions in all groups. The y-axis indicates the number of interactions per group per week normalised to the value of the week when the maximum number of interactions was recorded for that group. This figure pools normalised data from all three courses available. High performing students start to interact before and keep interactions throughout the whole course. The same applies to mid performing students, although their interactions start a bit later in the course. Low performing students start interacting later than high performing ones and their interactions drop with time. The maximum values used for normalising these curves were 150, 36, 57 and 63 all, high, mid and low interactions, respectively. (B, C and D) Evolution of the % of persistent interactions (relative to the average total # of interactions of that group) per week and per student group (low, (B); mid, (C); and high, (D)) relative to the total number of interactions per group per week. Continuous lines represent the fit of a curve to the points as indicated in Methods. As can be observed, the % of persistent interaction increases as the course progresses for all groups of students. High performing students achieved a higher % of persistent interactions than mid and low performing ones.

performing students interact in groups of similarly performing peers. This effect is stronger the higher the performance of the student. Indeed, low performance students tend to initiate many transient interactions regardless of the performance of the students they interact with. These interactions held by low performance students start late in the course, allowing high performers to establish a closely knitted group. In the following, we give details of these findings.

We start by comparing the score of each student with diversity metrics associated with the interactions held by each member of the social network (as shown in the SI). We characterise the nature and diversity of interaction ties within an individual's social network. Specifically, social diversity is defined as Shannon's entropy associated with individual communication behaviour, normalised to the total number of interactions (see Methods in SI for more details). Since both Shannon's entropy and the total number of interactions depend on the degree (number of connections), this normalisation reduces the correlation between low degree and high social diversity (see Figure S1 in Supplementary material).

The number of connections (students that a student has interacted with) and number of interactions (times a student has contacted or been contacted with/by other students), (see Methods in SI) were all positively correlated with the final score of the student (Pearson's correlations of 0.81, 0.85, respectively; $p < 0.01$), as shown in Figure 1B. Principal component analysis of these metrics revealed that all of them were closely interrelated, resulting in a non-significant improvement when combined (see Methods in SI). However, social diversity negatively correlated with final scores (Pearson's

**Table 1 | Summary of the cascade analysis performed across the three groups of students ($p < 0.05$ between any two groups)**

|  | LOW | MID | HIGH |
|---|---|---|---|
| % of Cascades | 13.03 | 35.97 | 51 |
| % trivial | 96.36 | 65.13 | 34.1 |
| Average Cascade length | 1.05 | 1.95 | 2.9 |
| % stars | 97.8 | 46.18 | 23.71 |
| % chain | 2.2 | 53.82 | 76.29 |

**Table 2 | Sigmoid Fitting Results. Constants obtained on fitting a sigmoid curve to the data**

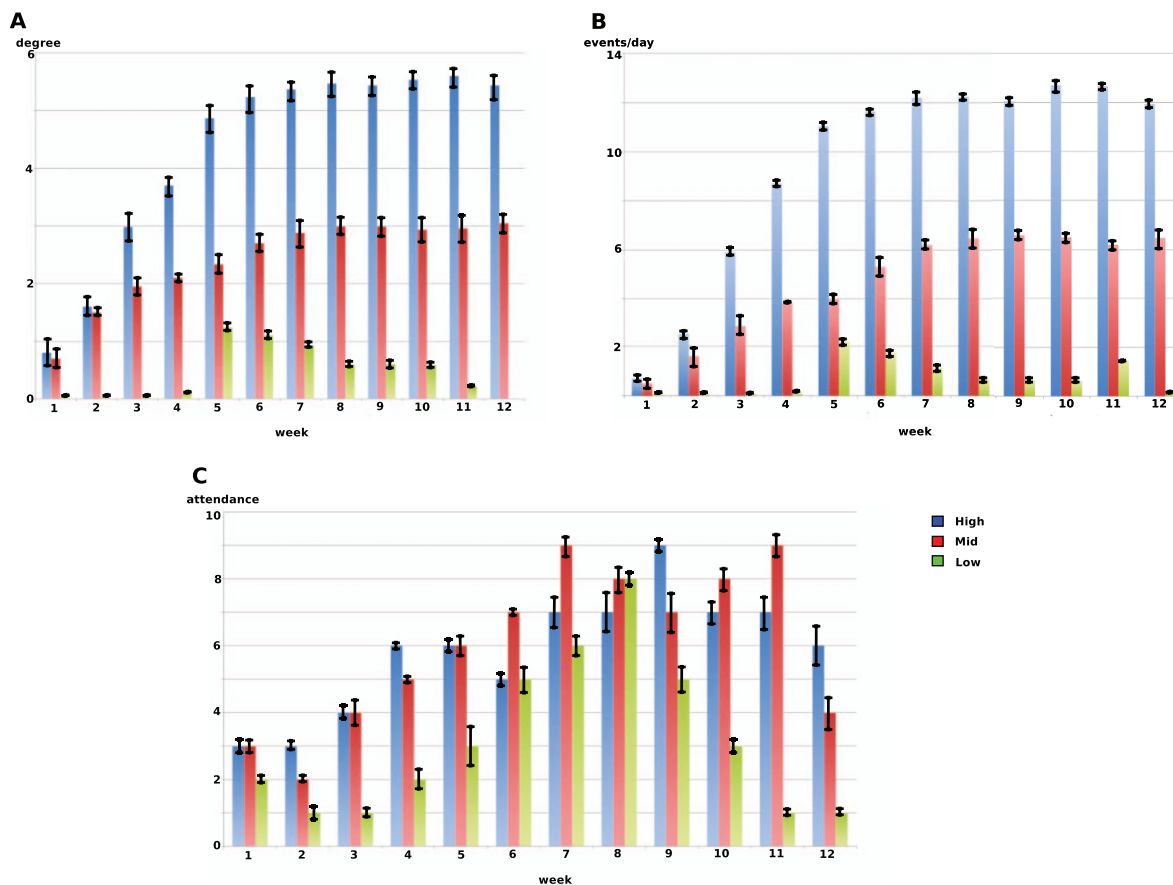|  | LOW | MID | HIGH | Total |
|---|---|---|---|---|
| $a$ | 0.9 | 13.6 | 26.79 | 0.973 |
| $b$ | 6.08 | 4.81 | 3.2 | 5.007 |
| $c$ | 1.43 | 0.44 | 0.58 | 0.4769 |

**Figure 3 | Course Data Details.** (A) Shows the evolution of the degree of the nodes in the graph per week per scoring group for all three courses. (B) Number of actual communications held per day on a given week grouped per scoring group. (C) An estimation of the attendance of the students to the course, based on the number of log-ons performed on any day in that week in any of the systems available for them to communicate. As can be observed, the degree remained almost constant for mid and high performing students, while it started to increase around week 4 and slowly declined later on for low performance students. This same pattern is observed for the number of interactions held by the students. These data are consistent with our estimation of "attendance", where log performing students have a significantly lower number of logins into the system. All panels show data from one of the courses under study only. The whiskers in the Figure show the estimated error in the mean.
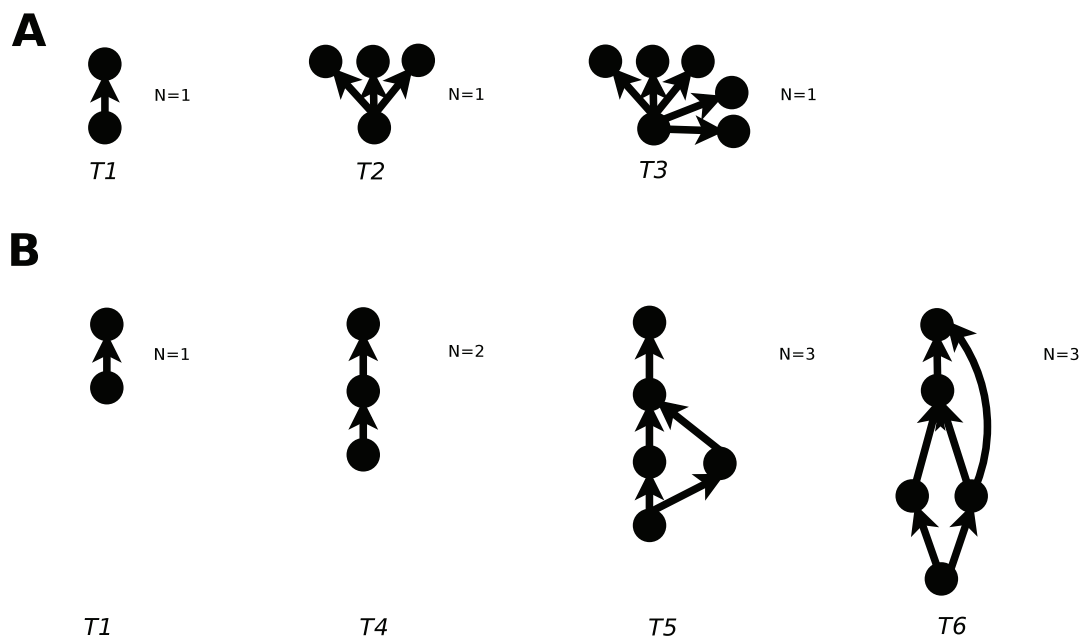


**Figure 4 | Information Cascades.** Most Frequent Cascades for Low Performing (A) and High Performing (B) students. Students initiating, relaying or receiving a document were supposed to be part of the cascade. As can be observed high performance students keep more complex information cascades in sharing documents in the systems available. Low performing students use a more straightforward "relay" strategy, forwarding documents to other students.

**Table 3 | Percentage of Interactions per Communication Channel.** Average % of interactions taking place over the different communication channels employed in our study. No significant differences were found between different groups of students. Moodle interaction count was increased only if the post received an answer. The collaborative workspace let us include interactions from blog posts, document shares, reminders or messages in the collaborative space. Each chat and classroom IRC session (sequence of messages exchanged without stopping for more than 3 min) counted as a single interaction

| Moodle | Collaborative workspace | Chat | Classroom IRC |
|--------|-------------------------|------|---------------|
| 39.63% | 13.49% | 35.34% | 11.54% |

correlation of –0.34, $p < 0.01$) (Figures 1C). The reader is reminded that correlation does not imply causation and that diversity cannot be regarded as the cause of low score from these results.

To further analyse the effects on score, students were grouped into high ($> 6.5$), mid (between 6.5 and 3.5) and low ($< 3.5$) scoring (scores in Spain are typically given in a 0–10 scale, being 10 the top score). To verify the suggested existence of less effective interactions, we also classified the type of interactions in two types: 1) persistent, those sustained over time, and 2) transient, those not reciprocated within a week. We found that at the end of the course up to $28 \pm 12\%$ of the interactions held by high performing students were persistent, which is statistically different to those held by mid ($14 \pm 5\%$) or low ($1 \pm 0.5\%$) performance students ($n = 290$, $p < 0.05$).

We analysed the average ratio of transient to persistent interactions per neighbour: a higher number indicated less targeted interactions. This is illustrated in Figure 1D for one of the three classes under analysis (results were similar for the other two classes).

The presence of more focused and sustained interactions did not stop high scoring students from interacting with colleague students with mid or low scores in a transient manner (similar number of transient interactions regardless of the score). An assortativity analysis[17] on these persistent interactions with regards to score indicated the existence of preferential interaction initiation ($r = 0.5$, $p < 0.05$ by using the Jackknife method, see Methods in SI). In other words, similarly scoring students tended to keep persistent interactions only between themselves.

This assortative behaviour with regards to scoring is highly suggestive of a "rich club" phenomenon (see Methods in SI and[18,19]). A "rich club" is defined as a set of nodes with degree larger than $k$ that tend to be more densely connected among themselves than the nodes with degree smaller than $k$. When we performed this analysis taking all the types of interaction into account, we could observe no "rich club" effect ($\phi(r) \simeq 3$ for the students with more links, indicating they also interacted with students outside the "rich club"). However, when only persistent interactions were taken into account, we obtained $\phi(r) \simeq 1$, which is in line with the idea of high scoring students keeping persistent interactions between themselves as indicated by our assortativity analysis. The "rich club" phenomenon could not be observed during the first weeks, $\phi(r) \ll 1$, and it became apparent only after week 4–5 for the top performing students, remaining stable afterwards.

**Temporal analysis.** One interesting finding is that the total number of interactions per week (normalised to the maximum value in all weeks) for all groups increased over time and it saturated around week 6 for mid performing students and around week 4 for high performing students (Figure 2A). In both cases, the number of persistent and transient interactions increased until saturation as the weeks went by. However, the number of interactions for low scoring students behaved in a strikingly different manner. The number of total interaction increased until week 4, where it started to drop steadily until the end of the course (Figure 2A). We believe this may be due to a lack of incentives to interact as revealed by our reciprocity measurements (see two paragraphs below).

A closer look at the data revealed that the percentage of persistent interactions increased in all groups, but with different timing, as shown in the persistent interaction analysis (see Figure 2B, C, D). As indicated in Table 2, the midpoint for the sigmoid function was 6.08, 4.81 and 3.2 weeks for low, mid and high performing students ($p < 0.05$). This suggested that high performing students on average established persistent interactions before mid and low performance students did (1 and 2 weeks earlier, respectively). Also, mid performing students started to establish persistent interactions 1 week before low performance students did. If one takes the slope of the sigmoid as a reference, it can be observed that there was no significant difference in the rate of change from a "low interaction mode" to a "high interaction mode" between mid and high performing students (0.58 vs. 0.4769). These data are in line with those on the number of connections, interactions and attendance (Figure 3 A, B and C), which showed that low performance students tried to engage later in the course, while mid and high performing students started their interactions earlier. These data are aligned with the number of students that stopped delivering their assignments and therefore did not pass the course. The average percentage of students dropping the course was 24.5%, 31.5% and 0% for low, mid and high performance students, respectively. $\simeq 80\%$ of these dropouts occurred after the 9th week of course. The higher attendance level by high performing students may also be causing the higher number of persistent interactions, although our analysis does not let us conclude any causality relationship.

Taking data on increasing percentage of persistent student interactions together with the assortativity analysis (students preferred to interact with those who have similar scores/performance), our results suggested that at some point reciprocity $R_{i,j}$ (measured as the fraction of times a student $i$ in any given group responds to a student $j$ outside her same group) should start to drop. However, reciprocity remained unchanged with time and was similar between groups ($\simeq 0.7$). By analysing the direction of the initiation of the interaction we could see that persistent interactions held between members of different groups are highly symmetric (having almost even initiations starting from both ends). On the contrary, transient interactions between members of different groups are almost always initiated by the student with lower performance (with 0.87 probability). In addition, the timing of responses was different. While persistent interactions are responded in $8.1 \pm 0.3$ hours on average, the response time for transient interactions is delayed $7.21 \pm 0.46$ days.

This could be indicating that low performance was due to either a lack of interest of the students or just that no valuable content was conveyed in these delayed interactions. Since the content of these interactions was not logged, we restricted ourselves to find whether there was any differences in the way content flowed between students and groups of students.

**Information cascades.** Information cascades reveal spread mechanisms in which an action or idea becomes adopted due to the influence of others, typically, neighbours in some network. A well-known example are cascades in the context of large product recommendation networks[21–24].

In order to detect the presence of information cascades and determine the actual value of the communication, we needed to gain insight on the content of the messages exchanged by students. Since this would be a clear violation of students' privacy, we decided to analyse another source of information: file exchange of students in their home directories and in their Moodle and collaborative workspace accounts (see "Information Cascades" in Methods in SI).
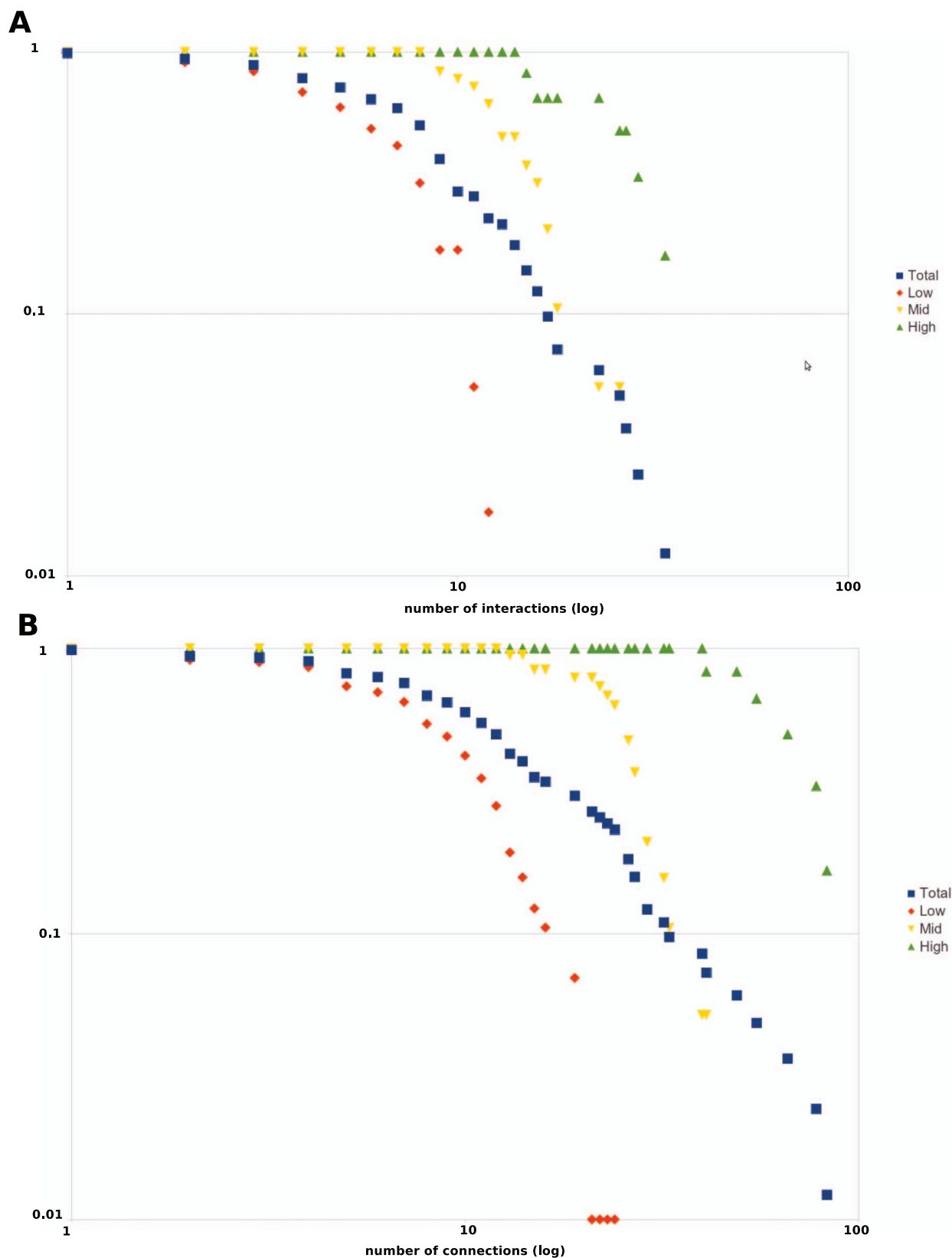
**Figure 5 | Quality of the Data.** Probability density distribution of the number of iterations (A) and connections (B) per group in one of the courses being analysed.

We defined as trivial cascades those implying a single transfer (a single originating source and a single destination) of information about the course, and non-trivial cascades, those with more complex patterns. We found a total of 845 cascades, and 53.37% of which were trivial cascades (*T1* in Figure 4), 25% were non-trivial cascades involving transfer from a single source to many destinations in the same time frame, and the remaining 11% of the cascades were topologically more complex.

The total number of cascades was significantly different across all three groups 51%, 35.97% and 13.03% for high, mid and low performance students, respectively (see Table 1).

Our data revealed that the length of the cascade (number of synchronous transfers) gradually increased as the average score of the students involved in the cascade increased. This is also supported by the fact that among non trivial cascades, the most common pattern for low performance students was star-like (*T2 and T3* in Figure 4, 97.8%), while chained cascades (*T4, T5 and T6* in Figure 4) were more common for mid (53.82%) and high (76.29%) performing students.

## Discussion

Being limited to non-verbal interactions between students prevented us from capturing a wealth of valuable interactions and led to some sparseness in our data. We combined fine-grained educational data at unprecedented temporal resolution in educational settings ($\simeq 4.6$ events per student per day) and gained insight into the type of interaction patterns that are associated to lower performance.

The major finding is that a higher number of online interactions (independently of the number of distinct students involved) is usually an indicator of higher score.

Our data show that increased social diversity is negatively correlated with high scores; most diversity metrics are correlated with the degree of the vertices (e.g. Shannon's entropy or topological diversity as in[25]) and this may lead to think that social diversity is high in low performing students because their number of connections (degree) is low. We minimised this fact with the normalisation of Shannon's entropy to degree.

The results also show that the higher the score of the students, the higher the percentage of their interactions that were persistent. These results were independent of gender differences (correlation of gender to score was $-0.04$). As the score of the student increases, these persistent interactions are initiated with a reduced number of similarly performing colleagues (assortative interaction pattern). Low performance students have a larger number of transient interactions spread over a large number of neighbours.

The dynamics of these interactions reveal that once students start to establish persistent interactions they do it more and more until a maximum saturation point is reached. High performing students tend to initiate persistent interactions before low performance ones, suggesting more willingness to collaborate. A striking fact is that these high performance students still maintain more than $\simeq 70\%$ of transient interactions, mostly with mid performance students. Our reciprocity analysis shows that students try to contact high performance students and these respond although the latter do not usually initiate disassortative interactions with low performance students.

These early persistent interactions enable high performance students to build a "rich club", while low performance students barely interact. Low performance students start to interact later (around week 4–5), when their "attendance" also increased just to decrease again towards the end of the course. This delay may help to explain why low performance students initiated more interactions that decreased after they failed to engage in persistent interactions with high performing students, since the "rich-club" had already been formed.

We could not monitor the content of the private message of students and decided to perform an information diffusion analysis that could help us gain insight on the value of the content actually being exchanged. Our results revealed that low performance students generally exchange documents in a trivial manner (i.e. in a forwarding manner that spans a single hop). On the contrary, more complex and longer cascades occur in high performing groups. This indicates the existence of a highly organised network where similarly performing students exchange information in a well-structured fashion, following characteristic patterns that are different across groups. While high performing students mainly exchange documents in a chained manner, low performance students spread the information to many other students at the same time, without this document apparently being relayed to other students beyond the recipient. Indeed, low performance students were not typically included in the information chains developed by high performing students. By this we do not mean to imply a deliberate behaviour of students, but it most likely indicates the presence of a benefit maximisation process by which students focus their efforts on potentially more fruitful connections.

Low performance students drastically reduce the number of interactions after week 5, which may be indicating a lack of motivation that leads them to drop the course and focus on other tasks. This *per se* does not let us conclude a lack of skills or motivation by low performance students. For instance, external factors may cause both less interactions and dropping the course (e.g. too many extracurricular activities). The lack of data that could enable causality inference in our analysis precludes us from concluding whether inefficient interactions, external factors or both are the cause of the dropout/reduced performance.

Even when we cannot directly build a causality chain, our empirical data suggest that: 1) low performing students engage later in the course; 2) this late engagement is related with their exclusion from the highly-structured and persistent information exchanges held by high performing students; 3) low performing students try to compensate by initiating larger number of weak interactions; 4) since this attempt to catch up is not successful low performance students drastically reduce the number of interactions.

Our study did not allow us to distinguish the root cause (initial delay in interacting, low degree or a combination of both) of the increased social diversity found in low performing students.

As part of our future work, we aim to perform a detailed causality analysis to detect the root cause of the low performance. This may help to get low performing students involved in high performing chains and hopefully increase their final score and reduce dropout rates. On the other hand, this may have a negative effect on high scoring students who will get many more interactions. We also plan to expand this analysis to non university environments.

## Methods

The data consist of the interactions of 290 students at a Spanish university, during two consecutive years of a 12-week long course on Basic Computer Science Skills (in Linux such as OpenOffice, GIMP, or content licensing techniques such as Creative Commons) for freshmen students of journalism.

An interaction is defined as a communication attempt via the aforementioned systems. We logged the time and direction of the interaction in the Chat and the class IRC (see Table 3 for a detailed list of interactions and types). Confidentiality prevented us from performing an examination on the content of these interactions. Moodle and our collaborative workspace let us keep track of documents shared by students.

These interactions were used to build a graph with a fine grained temporal granularity (see Communication Channels in the SI). Diversity, grouping and connectivity metrics were calculated on the graph (see SI)[20]. These metrics were analysed and compared throughout the course. A snapshot of the quality of the data set can be observed in Figure 5.

Finally, we studied how files appeared and spread across the HOME directory students kept in the servers of the Lab (see SI).

1. Diplomas Count 2007: Ready for What? Preparing Students for College Careers and Life after High School. *Education Week* **26** (2007).
2. Rouse, C. The Labor Market Consequences of an Inadequate Education. Princeton University and NBER. In *Equity Symposium on The Social Costs of Inadequate Education at Teachers College, Columbia University*, edited by Clive Belfield and Henry M. Levin (Washington: Brookings Institution Press,2007). Available: http://devweb.tc.columbia.edu/manager/symposium/Files/77_Rouse_paper.pdf Last visited: 4-1-2013
3. Harlow, C. Education and Correctional Populations. In *U.S. Department of Justice, Bureau of Justice*, (Washington DC, 2003). Available: www.ojp.usdoj.gov/bjs/pub/pdf/ecp.pdf Last visited: 4-1-2013
4. Neale, D. C. & Carroll, J. M. Multi-faceted evaluation for complex, distributed activities. In *Proceedings of the 1999 Conference on Computer Support for Collaborative Learning* Article 53 (International Society of the Learning Sciences,1999).

5.  Nurmela, K., Lehtinen, E. & Palonen, T. Evaluating CSCL log files by social network analysis. In *Proceedings of the 1999 conference on Computer support for collaborative learning* Article 54 (International Society of the Learning Sciences,1999).

6.  Cho, H., Gay, G., Davidson, B. & Ingraffea, A. Social networks, communication styles, and learning performance in a CSCL community. *Computers & Education* **49**, 309–329 (2007).

7.  Martinez, A., Dimitriadis, Y., Rubia, B., Gomez, E. & De La Fuente P: Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education* **41**, 353–368 (2003).

8.  Chen, Z. & Watanabe, S. A Case Study of Applying SNA to Analyze CSCL Social Network. In *ICALT 2007. Seventh IEEE International Conference on Advanced Learning Technologies* 18–20 (2007).

9.  Granovetter, M. The strength of weak ties. *The American Journal of Sociology* **78**, 1360–1380 (1973).

10. Sundararajan, B. Impact of communication patterns, network positions and social dynamics factors on learning among students in a cscl environment. *PhD thesis*, (Troy, NY, USA, 2007).

11. Yu, AY., Tian, SW., Vogel, D. & Chi-Wai Kwok R: Can learning be virtually boosted? An investigation of online social networking impacts. *Computers Education* **55**, 1494–1503 (2010).

12. Ullrich, C., Borau, K. & Stepanyan, K. Who students interact with? a social network analysis perspective on the use of twitter in language learning. In *Proceedings of the 5th European Conference on Technology Enhanced Learning Conference on Sustaining TEL: from innovation to learning and practice* 432–437 (Berlin, Heidelberg: Springer-Verlag, 2010).

13. Yeung, Y. Y. Macroscopic study of the social networks formed in web-based discussion forums. In *Proceedings of the Conference on Computer Support for Collaborative Learning: the next 10 years!* 727–731 (International Society of the Learning Sciences, 2005).

14. Kepp, S. J. & Schorr, H. Analyzing collaborative learning activities in wikis using social network analysis. In *Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems* 4201–4206 (New York, 2009).

15. Cho, H., Gay, G., Davidson, B. & Ingraffea, A. Social networks, communication styles, and learning performance in a CSCL community. *Computers and Education* **49**, 309–329 (2007).

16. Erlin, B., Yusof, N. & Rahman, A. Integrating Content Analysis and Social Network Analysis for analyzing Asynchronous Discussion Forum. In *ITSim 2008. International Symposium on Information Technology* **3**, 1–8 (2008).

17. Newman, M. Mixing patterns in networks. *Physical Review E* **67**, 026126 (2003).

18. Zhou, S. & Mondragon, R. J. The rich-club phenomenon in the Internet topology. *IEEE Communications Letters* **8**, 180–182 (2004).

19. Colizza, V., Flammini, A., Serrano, M. A. & Vespignani, A. Detecting rich-club ordering in complex networks. *Nature Physics* **2**, 110–115 (2006).

20. Wang, P., Hunter, T., Bayen, A. M., Schechtner, K., Gonzlez, M. C. Understanding Road Usage Patterns in Urban Areas.*Scientific Reports* **2**, 1001 (2012).

21. Leskovec, J., Singh, A. & Kleinberg, J. Patterns of influence in a recommendation network. In *Proceedings of the 10th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining* 380–389 (Springer-Verlag, 2006).

22. Leskovec, J., Adamic, L. A. & Huberman, B. A. The dynamics of viral marketing. In *Proceedings of the 7th ACM conference on Electronic commerce* 228–237 (New York, 2006).

23. Yang, J. & Leskovec, J. Temporal Variation in Online Media. In *Proceeding of the ACM International Conference on Web Search and Data Mining* 177–186 (New York, 2011).

24. Leskovec, J., Adamic, L. A. & Huberman, B. A. The dynamics of viral marketing. *ACM Transactions on the Web* **1**, 5 (2007).

25. Eagle, N., Macy, M. & Claxton, R. Network diversity and economic development. *Science* **328**, 1029 (2010).

## Acknowledgements

## Author contributions

Conceived, designed and performed the experiments: L.M.V. Analysed the data: L.M.V., M.C. Wrote the paper: L.M.V., M.C.

## Additional information

**How to cite this article:** Vaquero, L.M. & Cebrian, M. The rich club phenomenon in the classroom. *Sci. Rep.* **3**, 1174; DOI:10.1038/srep01174 (2013).