Information, advice, friendship, notes and trust network: evidence on learning from classmate [1]

Emma Zavarrone

Iulm University, Milan, Italy, emma.zavarrone.iulm.it

Agnese Vitali

Bocconi University and Dondena Centre, Milan, Italy; agnese.vitali@unibocconi.it

Abstract: This paper contribute to the literature on the influence of network structure and performance of university students over time. We move from the assumption that students' school performance is influenced by: friendship, exchange of general information about the course, contents, lecture notes and trust networks. Social influence has been modeled through SARAR model and several spatial weight matrixes W and M have been compared.

Keywords: SARAR model, social influence, social network analysis

1. Introduction

A number of studies in social network analysis, economic and sociology have recently focused on the association between friend networks (or peer effect) and school performance.

In this paper we aim at contributing to the literature on the influence of network structure and performance of university students. We hypothesize that class mates can develop four different types of relationships. Existing literature usually focuses on two main types of such relationships, namely friendship and study networks. In this paper we explore also the role of two other relationships which might be associated with students' performance: the exchange of general information and the exchange of lecture notes. We measure school performance by the means of students' University Human Capital (UHC).

We test whether the ego's UHC is influenced by the UHC of the subgroup of class mates with which he/she has a relationship of one of the four kinds and which of these relationships has the higher marginal effect on UHC (i.e., we investigate whether UHC is influenced by the UHC of study-network members, as well as by the UHC of friendship-, information- and lecture notes-network members) (*hypothesis 1*). Also, we inspect if students' UHC is influenced by unobserved characteristics common to the ego's networks' structure (*hypothesis 2*). We will test the hypothesis that high-performance students tend to relate themselves with other high-performance students, to isolate low-performance students and to have less free time to spend hanging out with friends (*hypothesis* 3).

2. Materials and Methods

We developed an ad-hoc survey in which students in a given class are asked to detail the structure of four different networks to which they belong. Respondents are master students in the age range 22-23, attending the Statistics course during their first year of a two-year master degree at Iulm University in Milan, Italy.

Students are asked whether, in order to prepare the Statistics exam, they studied on their own. If they did not, they are asked to identify the class mates with whom they studied. We consider these peers as members of the ego's study network. Students are also asked whether they have class mates with which they get together outside the university environment, and if they do, we ask to identify them and we consider them as members of the ego's friendship network. In order to identify the information network and the lecture notes network, we look at class mates whom the ego considers a reliable source of information for what concerns the Stats course and with which he/she exchange/compare his/her notes, respectively. The structure of all the four networks we consider are such that relationships do not necessarily need to be reciprocal. Students are surveyed twice: before the mid-term exam and before the final exam at the end of the course. Student UHC is then measured using the difference between the student's grade obtained at time 1 and the grade he/she obtained at time 0. Both grades are expressed in thirtieths (minimum for sufficiency is 18), a UHC equal to 0 is interpreted as no change in performance between time 1 and time 0, while a positive (negative) UHC is interpreted as increased (decreased) performance between time 1 and time 0.

In addition to information relating to the structure of the four networks discussed above, the survey also collects information on the students' field of education during their bachelor studies, whether their university career took place in the same University in which they are surveyed, and if this is not the case, in which university they took their bachelor. Further, the students are asked whether, during their university studies, they took a Stats class, and if this was the case, they are asked to specify which class it was. Finally, a question is asked to identify who the ego subjectively perceives as the central subject among his/her class mates ("You are the person located in the bottom part of this picture. Could you specify, among your class mates, the initials of the person in the upper left of the picture?"). This question does not refer to one particular network, rather it aims at catching the ego's perception about the central subject among his school mates, in general.

In order to test our three assumptions, we employ respectively the models:

1) the spatial lag model, $UHC = \rho_1 WUHC + X'\beta + \varepsilon; \ \varepsilon \sim N(0,\sigma_\varepsilon^2 I)$
2) the spatial error model, $UHC = \mathbf{X}\beta + \varepsilon; \ \varepsilon = \rho \boldsymbol{W} \varepsilon + v; \ v \sim N(0,\sigma_v^2 I)$
3) the spatial auto-regressive auto-regressive model (SARAR):

   $UHC = \rho_1 WUHC + X'\beta + u; \ u = \rho_2 Mu + \varepsilon$

The independent variables (gender and mark of previous statistics ability) are the same for all models; the coefficient ρ_1 measures the spatial autocorrelation in the dependent variable i.e. *a spatial lag* (Cliff et al., 1973; Leenders, 2002); if this coefficient is positively significant, there is evidence of spatial autocorrelation in UHC or, in other words, that students belonging to the same network tend to have similar grade differentials over the two time periods. The coefficient ρ_2 measures instead the spatial autocorrelation in the error term; if this coefficient is positively significant we interpret that there are common unobserved factors influencing all members of the same network (i.e., unobservable factors will have an effect on the network member's UHC to which they are related, but also on the UHC of his/her peers). The spatial weight matrixes W and M, which need not be equal, are non-stochastic spatial weight matrixes which take into account the neighbouring structure of the students, such that their entries are non-null (i.e., two students are neighbours) if the students belong to the same network.

For each weight matrix, we also define a different set of weights in order to assess the robustness of the results found, on the basis of different weight structures. To this aim, in the first place weights will be defined in such a way to assign the same weight to all members of a given network (thus weights will be proportional to the number of people belonging to the specified network). In the second place, weights will be defined to assign more weight to the peer who is central in the network.

Other model assumptions require that the spatial autoregressive ρ_1 and ρ_2 coefficients are bounded in absolute value (i.e. $|\rho_1|<1$ and $|\rho_2|<1$), ε_i is independently and identically normally distributed with zero mean and variance to be estimates. The model can be estimated via Maximum Likelihood or following a GMM procedure. Due to the narrowness of our sample size (n=41), we rely on the second approach. We test the significance of the two spatial autocorrelation coefficients using Lagrange Multiplier Tests.

3. Results

The final sample is constituted of 41 students Table 1 shows some socio-demographic information of the our sample: it is not surprising that male students represent only a minority (34%) at Iulm university (it is well known that a gender difference exists when the field of study is concerned; in particular, women are more often found in humanistic subjects).

Variable	Mean/Pro	Std. Dev.
Final grade in Stats	22	5.2
Sex (prop. of men)	34%	-
BSc in different	48%	-
Ever studied Stats	56%	-

Table 1: Descriptive statistics

For the lag spatial model we find a network effect on performance among class mates. The trust network exerts the most powerful effect on students' performance, followed by the friend and study networks. The Figure 1 summarize these results by the mean of a graphical representation. On the y-axis there is the magnitude of the spatial coefficient. Then we group models by weight matrix, so in the first case, using the friend network, the spatial weight matrix can be defined on the basis of 7 different criteria. And the same goes for each of the other 4 types of networks.
For the other two models we don't find a statistically significant effects and we are testing misspecification procedure on ρ_1 *and* ρ_2.

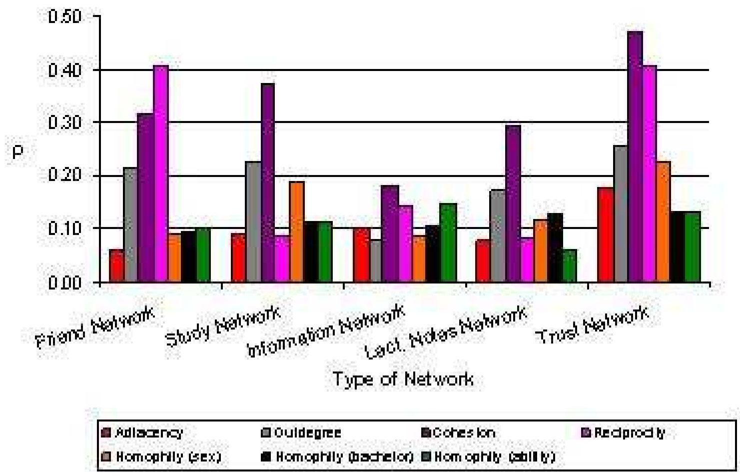

Figure 1: Comparison of ρ coefficient (*p-value* $\leq$*0.05*) in five networks

4. Concluding remarks

Using a set of different, equally theoretically-grounded weight matrices we show that:

i. in some cases results are robust to different specifications of *W*,
ii. however, in some other cases parameter estimates –hence conclusions– based on autocorrelation models can change according to the chosen specification of *W*,
iii. the network structure need to be translated into a meaningful and theory-guided choice of weight matrix (Leenders, 2002).

References

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*. Boston, Kluwer Academic.

Cliff, A.D. and Ord, J.K. (1973 *Spatial Autocorrelation*. London: Pion Limited.

Leenders, R. Th. A. J. (2002) Modeling social influence through network autocorrelation: Constructing the weight matrix. *Social Networks*, 24, 21-47.