



Clustering Multiplex Networks in Schools Through ERGMs and Tandem Analysis

Angela Pacca^{1,4}✉, Luka Kronegger², Giancarlo Ragozini³,
and Maria Prosperina Vitale⁴

¹ University of Florence, Florence 50134, Italy

² University of Ljubljana, Ljubljana 1000, Slovenia

³ University of Naples Federico II, Naples 80138, Italy

⁴ University of Salerno, Fisciano, (SA) 84084, Italy

angela.pacca@unifi.it

Abstract. The present contribution proposes an analytical strategy designed both to identify clusters of schools characterised by distinctive local network structural configurations as well as to enhance the interpretation of the obtained results by integrating supplementary variables capturing students' socioeconomic background and educational contexts. Using data collected from high schools in Southern Italy, the proposed approach combines results from network models and exploratory data analysis techniques. First, Exponential Random Graph Models are estimated separately for each school network, and then the resulting parameters are used to compare relational configurations. Given the presence of some outliers, in a second step, school networks are clustered by using the coordinates obtained via a robust Principal Component Analysis on the model parameters. Clusters are then interpreted in light of students' characteristics, the scholastic context, types of relationships among classmates, and educational choices after high school. The main findings reveal distinct groups, each characterised by relational patterns—from highly centralised and hierarchical to locally cohesive or expansive patterns—shaped by the interplay of peer network structures, individual characteristics, and scholastic contexts.

Keywords: Exponential Random Graph Models · robust PCA · Peer network · Higher education

1 Introduction

The school environment constitutes a crucial context of socialisation in which students form and develop different types of social relationships and support networks. A substantial body of research has reported that interactions among classmates are closely associated with access to information during key educational transitions. They play a significant role in shaping students' aspirations and choices, including the decision to pursue higher education after high school

[1–3]. These complex structures, in which relationships among students convey crucial information about social dynamics and the processes shaping the diffusion of behaviours, attitudes, and orientations, require dedicated methodological approaches [4, 5]. Within this context, the present study aims to propose an analytical strategy to find clusters within a collection of networks referring to a similar context. A two-step procedure that combines advanced network modelling results with dimensionality reduction and cluster analysis is proposed to capture the most relevant relational patterns. Specifically, in a first step, Exponential Random Graph Models (ERGMs) [6] are estimated for each network, and the resulting estimated parameters are used to map the networks in a common parametric space. Given the exploratory framework adopted, in the present study, the parameters' standard errors and significance are not included in the analysis.

In the second step, the so-called tandem analysis [7], i.e. a dimension reduction and a cluster analysis applied sequentially, is used to obtain clusters of networks based on their structural properties and model characteristics. The proposed procedure is applied to a sample of final-year classes of high schools in Southern Italy. A survey gathered information on different types of relationships (friendship and support ties) among classmates. The network data are related to four distinct layers within each classroom, resulting in several multiplex networks [8] at the school level. In this paper, each layer within a school is treated as a separate statistical unit. The obtained clusters are then interpreted in light of both average model parameters and students' characteristics related also to the study track, the type of support, and the educational choices after high school. This strategy seems to be a flexible and scalable methodology for the comparative analysis of multiplex social networks. Note that the networks under analysis are comparable, even if they are slightly different in size and density. These differences are captured and included in the analysis by considering specific ERGM parameters. In this way, such heterogeneity contributes to identifying groups of networks characterised by similar relational patterns, while also accounting for structural differences. In educational settings, it could provide new insights into the interplay between social relations and educational trajectories.

2 The Analytical Strategy at Work

The study is based on data¹ collected through a survey on a sample of 28 Italian high schools in the Campania region, encompassing 1,532 students across 91 final-year classes [9]. Information on personal and social factors affecting life trajectories in the transition from high school to university or the labour market was collected. Particular attention has been devoted to the measurement of the peer effect, intended as informal ties among classmates. To this end, friendship ties and three different types of support (emotional, scholastic, and appraisal)

¹ The data come from a survey financially supported by the PRIN 2022 project “From high school to university: Assessing peers’ influence in educational inequalities and performances.” For details see: <https://prin.unica.it/peer-up/>.

were measured to capture the multifaceted social interactions, resulting in multiplex networks at the school level. Although drawn from a geographically limited area, this comprehensive dataset is unique in its scope and detail, offering a rich foundation for analysing peer interactions within schools. It enables subsequent multivariate analyses that explore how the formation of networks interacts with both scholastic and individual characteristics to affect educational pathways and shape students' life trajectories.

A two-step analytical procedure was proposed for clustering these four types of networks in order to identify schools that share homogeneous patterns of relational mechanisms and structural interaction dynamics. In detail, first, ERGMs were estimated separately for each school and network type to capture the local structural mechanisms driving the ties' formation, while accounting for endogenous dependencies within the networks. A total of 84 ERGM models were obtained across 21 schools and network types.² As the modelling phase aims to represent networks in a common parameter space, the same set of ERGM parameters was selected for all networks based on theoretical considerations on peer relational mechanisms. This approach enables each network to be represented by a comparable parameter vector, thereby facilitating a model-based comparison analogous to methodologies employed within the framework of time series analysis [10]. Thus, parameters obtained for all networks are: density that captures the baseline tendency for tie formation; reciprocity that accounts for mutual ties; some degree-related effects including out-degree and in-degree equal to 1 (controlling for minimum sending and receiving ties), and geometrically weighted out-degree (1.75) and in-degree (2), capturing heterogeneity and centralisation; geometrically weighted edge-wise shared partners and dyad-wise shared partners modelling higher-order dependencies and capturing triadic closure and shared partners outside closed triads; the outgoing two-path as indirect connections; and, finally, homophily effects capturing the tendency of actors to form ties with others sharing the same characteristics. All these parameters provide detailed relational patterns within each school, reflecting both cohesion and hierarchy in social interactions among classmates. In the second step, these multivariate parameter vectors were analysed using a tandem approach, in which dimensionality reduction and clustering were applied sequentially to identify homogeneous groups of schools. Dimensionality reduction was performed via PCA, which revealed three outliers—three networks exhibiting no heterogeneity in parents' educational level or gender composition, and displaying parameters related to homophily effects. For this reason, a robust version of PCA was employed, based on eigenvectors and eigenvalues of the robust covariance matrix estimated through the minimum covariance determinant [11]. Given the coordinates of the networks in the reduced space, i.e. on the k first relevant factors, schools were clustered through a hierarchical clustering procedure applying the Ward criterion [12], yielding distinct groups characterised by similar structural network profiles across multiple relational dimensions. The resulting clusters were then substantively interpreted taking into account the significant model parameters

² ERGM models did not converge for seven schools.

and the variables related to the school gender composition, the proportion of students intending to pursue higher education after high school; the type of study track (classical lyceum, scientific lyceum, technical institute, vocational institute, multitrack schools with different kind of lyceum or mixed tracks of different types); the province in which the school is located (Avellino, Benevento, Caserta, Naples, and Salerno); and the type of network.

3 First Clustering Results

Without going into details, robust PCA provides five relevant principal components accounting for approximately 81% of the total variance. Looking at the dendrogram of the clustering result (Fig. 1), three clustering solutions were examined, yielding 3-, 5-, and 7-cluster partitions. Based on the between-cluster to within-cluster variance ratio, as well as considerations of interpretability, the 7-cluster solution was selected and described below.

- The *first cluster* (19 schools) shows positive in-degree alongside negative out-degree and low density, indicating a hierarchical relational structure. A small number of central students receive the majority of ties, while the capacity to form outgoing connections is limited. Triadic closure is slightly positive, reflecting the presence of some locally cohesive subgroups, but overall, the network remains sparse, with some students playing central roles. The cluster includes schools from the province of Salerno, mainly multitrack institutes and scientific lyceums, with a higher proportion of students intending to pursue higher education. This suggests that the observed relational patterns are associated with academically oriented environments where peer interactions are centralised.
- A second *cluster* (14 schools) contains networks marked by very high out-degree and positive triadic closure, with negative dispersion. Ties are broadly distributed among students, resulting in low overall fragmentation. This group is mainly composed of vocational schools. The networks show a significant association between social class homophily and students who are undecided about university enrolment after high school. This suggests that ties are more likely to form among peers who share similar family backgrounds and experience uncertainty regarding their educational trajectories.
- The *third cluster* (15 schools) exhibits positive density but negative in-degree and geometrically weighted in-degree, indicating that ties are relatively evenly distributed across students with low network centralisation. Interaction patterns reveal local subgroups connected through homophily. Ties are more likely among students planning to continue their studies who share the same educational level of parents, as well as among undecided students with the same educational level of fathers. This group also reflects a scholastic support network, structured around locally cohesive groups defined by educational choices and family background.
- The *fourth cluster* (13 schools) exhibits high reciprocity alongside negative in-degree and triadic closure, reflecting selective and reciprocal ties. Students

tend to form mutual ties that provide scholastic support, particularly in relation to future educational choices, emphasising the connection between relational cohesion and orientation toward higher education. Negative coefficients for undecided students within the same social class further indicate that social class homophily shapes tie formation. This group is predominantly composed of classical lyceums, with several schools showing a strong orientation toward university enrolment. This configuration highlights environments where peer networks are reciprocal, selectively cohesive, and closely linked to students' educational trajectories.

- The *fifth cluster* (11 schools) is characterised by high out-degree but low overall density, reflecting expansive structures in which many outgoing ties are activated without generating broad cohesion. Triadic closure is positive but limited, indicating the presence of locally cohesive subgroups without overall network integration. The prominence of friendship networks highlights the role of informal ties among peers. It includes a multitrack lyceum with a relatively high proportion of female students.
- Finally, two small clusters of 6 schools appear. One group is marked by high geometrically weighted in-degree alongside negative out-degree, reciprocity, and triadic closure, indicating a strongly hierarchical, vertical structure in which some central students receive most incoming ties, while mutual and locally cohesive connections are limited. It includes a few schools from the province of Naples and institutes with different types of lyceum. This configuration reflects scholastic environments where interactions are highly centralised and hierarchical, with peer ties strongly shaped by both network position and similar socioeconomic background. The last group shows positive triadic closure, indicating locally cohesive subgroups, but negative in-degree, reciprocity, and low propensity for post-secondary enrolment. Ties remain concentrated within small groups, preventing the formation of an integrated network. Negative coefficients for students who intend to continue their studies within the same social class highlight homophily, indicating that such similarities may limit the formation of broader peer connections. This group reflects school environments, where interactions are highly localised and selective, shaped by both structural patterns and socioeconomic factors.

All these first findings indicate that the functioning of peer networks within schools arises from a complex interplay of structural, compositional, and contextual factors. Peer relational patterns emerge from the joint influence of the school environment, the student background, and the family context. Finally, the proposed approach presents some limitations that deserve further investigation. In particular, the use of ERGM parameter estimates—being subject to uncertainty—may affect the stability and the robustness of clustering results. Moreover, differences in network size, density, and composition may limit the comparability of parameters across networks, while variability in model fit across schools may further influence the results. Future research should address these issues by assessing the robustness of the clustering procedure and by considering the significance of the structural parameters and the overall goodness of fit mea-

tures. These developments will contribute to a more reliable and interpretable framework for network classification.

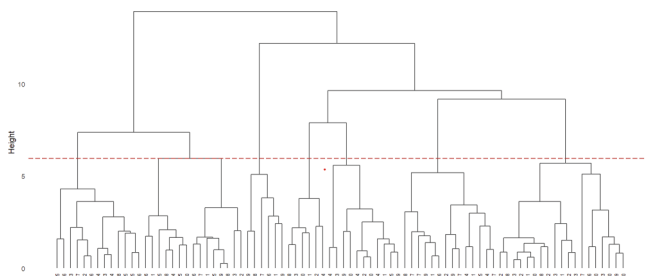


Fig. 1. Dendrogram with the 7-cluster solution at the school level

References

1. Raabe, I.J., Wolfer, R.: What is going on around you: peer milieus and educational aspirations. *Eur. Sociol. Rev.* **35**(1), 1–14 (2019)
2. Morgan, S.L.: Modeling preparatory commitment and non-repeatable decisions: information processing, preference formation, and educational attainment. *Ration. Soc.* **14**(1), 387–429 (2002)
3. Smith, E.: Peer preferences and educational decisions: heterogeneous associations across student socioeconomic status. *Brit. J. Sociol. Educ.* **44**(2), 374–393 (2023)
4. Bródka, P., Chmiel, A., Magnani, M., Ragozini, G.: Quantifying layer similarity in multiplex networks: a systematic study. *R. Soc. Open Sci.* **5**, 171747 (2018)
5. Giordano, G., Ragozini, G., Vitale, M.P.: Analyzing multiplex networks using factorial methods. *Soc. Networks* **59**, 154–170 (2019)
6. Lusher, D., Koskinen, J., Robins, G.: Exponential random graph models for social networks: Theory, methods and applications. Cambridge University Press, New York (2013)
7. Markos, A., D’Enza, A.I., van de Velden, M.: Beyond tandem analysis: joint dimension reduction and clustering in R. *J. Stat. Softw.* **91**(10), 1–24 (2019)
8. Dickison, M.E., Magnani, M., Rossi, L.: Multilayer social networks. Cambridge University Press, Cambridge (2016)
9. Sulis, I., Tocchioni, V., Vitale, M.P. (eds.): Dalla scuola superiore all’università: Valutare l’influenza dei pari sulle diseguaglianze educative e sulle performance. Franco Angeli, Milano (2026)
10. Corduas, M., Piccolo, D.: Time series clustering and classification by the autoregressive metric. *Comput. Stat. Data An.* **52**(4), 1860–1872 (2008)
11. Hubert, M., Rousseeuw, P.J., Vanden Branden, K.: ROBPCA: a new approach to robust principal component analysis. *Technometrics* **47**(1), 64–79 (2005)
12. Husson, F., Lê, S., Pagès, J.: Exploratory Multivariate Analysis by Example Using R. Chapman and Hall/CRC, London (2010)