Measuring Peer Effects in a Network Perspective. Survey Design and Privacy Issues

Angela Pacca[1,2(✉)], Giancarlo Ragozini[3], and Maria Prosperina Vitale[2]

[1] University of Florence, 50134 Florence, Italy
angela.pacca@unifi.it
[2] University of Salerno, 84084 Fisciano, SA, Italy
[3] University of Naples Federico II, 80138 Naples, Italy

Abstract. The present contribution discusses a strategy to address methodological and ethical challenges in network survey designs. The approach seeks to balance statistical rigor with the protection of participants' privacy. Its applicability is demonstrated through a case study involving high school students in Southern Italy aiming at exploring the factors related to their post-diploma educational choices. Recognising the role of peers' influence, the study adopts a statistical model that includes the students' relationships within a class, i.e. the class adjacency matrix, considering different kinds of interactions among classmates. All students in the classes selected through a quota sampling procedure are included in a whole network study, raising important ethical and privacy concerns. To address these issues, anonymisation protocols are implemented to ensure data protection and management.

Keywords: Peer effects · Network survey design · Research ethics · University choices

1 Introduction

Scholars have shown that peers can exert an influence on educational choices [1–4] as well as personal expectations closely related to social context [5]. Given the relevance of these dynamics in shaping students' trajectories, it is important to explore research perspectives capable of capturing the underlying relational dynamics.

To operationalise peer effects, a network analysis perspective is particularly useful for gathering information on interactions and connections in educational research [6], providing an in-depth understanding of social influence and selection mechanisms. Nevertheless, the adoption of such designs raises significant ethical and privacy concerns [7,8], making it essential to employ methodologically sound strategies that reconcile the protection of sensitive data with the validity of statistical analyses. This requires the implementation of rigorous anonymisation

E. di Bella et al. (Eds.): SIS 2025, ISSSAS, pp. 120–126, 2025.
https://doi.org/10.1007/978-3-031-96033-8_21

procedures and advanced data protection protocols to ensure participants' confidentiality throughout the entire research process, including the dissemination of findings [9].

Within this context, the present contribution discusses a strategy to address methodological and ethical challenges in network survey designs. The approach seeks to balance statistical rigor with the protection of participants' privacy. Its practical application is illustrated through a case study involving high school students in Southern Italy, focused on exploring the factors influencing their post-diploma educational choices. The role of peers' influence is included in statistical models that use the intra-class relationships, i.e. the class adjacency matrix, as explicative variable. All students in the classes selected through a quota sampling procedure are included in a whole network study, raising important ethical and privacy concerns. To address these issues, anonymisation protocols are implemented to ensure data protection and management.

2 Peer Effects and Network Research Ethics

The analysis of peer effects presents inherent complexities, particularly due to the simultaneous interactions among individuals, which can generate reciprocal relationships that are difficult to isolate. This perspective relates to the reflection problem, which explores whether and to what extent an individual's behaviour is influenced by their reference group. To identify mechanisms of social influence, it is necessary to adopt an approach that jointly studies the interactions between endogenous, exogenous, and correlated effects [10]. Some researchers suggest that the reflection problem can be addressed through methodologies that focus on interactions within social networks, enabling a more accurate distinction of peer influence mechanisms [11].

From this perspective, measuring peer effects presents a considerable challenge, as it requires the formulation of a well-designed research framework that can accurately capture and interpret relationships among people. This is compounded by the complexity of the interactions themselves, which involve a range of social, psychological, and environmental factors that have an impact on individual behaviour. To address this, a carefully structured research design is essential to facilitate the collection of network data but also to incorporate analytical techniques to interpret the subtle dynamics at play. Social Network Analysis examines this phenomenon in terms of an inter-subjective perspective, focusing on the entire system. This methodological approach is not limited to investigating personal connections but the overall architecture of the network, identifying structural patterns, substructures, and ways in which the social capital is distributed among actors.

Network studies are often difficult to perform, particularly outside of constrained settings, such as classrooms, schools, or organizational environments [12]. The primary advantage of a network research design lies in its capacity to facilitate the identification of ties between respondents, offering tools to assess actors' centrality within the broader network, their access to information and

resources across distinct network regions, and other critical measures. A classroom, for instance, represents a bounded context for the implementation of whole network research designs [13]. The widespread use of digital technologies and the proliferation of data collection methods have made the potential risks associated with the disclosure of personal information more apparent. Privacy-related issues have taken on a critical dimension, as social networks can lead to the inadvertent disclosure of sensitive data [14]. In network data collection, the information gathered is not limited to involved respondents. Participants may, in fact, nominate other individuals who are unaware of being mentioned, which significantly amplifies the privacy-related issues. In this regard, the non-participation of an individual in the study does not necessarily mean that the person is excluded from the data collection process, as they may have been named by other participants as part of their social network. The implication of such social dynamics underlines the need of implementing rigorous measures to protect privacy and ensure the security of the information processed [7,15]. To address these issues, network research design must integrate protection procedures with data quality and validity. The gathered information complies with privacy regulations, without compromising the reliability of data, and establishes the foundations for research that respects participants' rights while maintaining high standards of scientific rigour.

3 Methodological Challenges in Network Designs

Before presenting the survey conducted in high schools in Southern Italy (Campania region) and the proposed strategy at work, let us introduce the statistical model we will use to investigate the factors influencing students' educational choices at the end of high school. In our case, if we exclude interactions between students from different classes, each class can be considered a subnetwork, within the students' educational choices can be influenced by their classmates. Given intra-class relational dynamics, i.e. the adjacency matrix $\mathbf{G}$ coding the network ($\mathbf{G} = (g_{ij})$, with $g_{ij} = \frac{1}{d_i}$ if there is link between the student i and the student j, $g_{ij} = 0$ if the link does not exist, and d_i the degree of the student i), following [11] the model to explore the role of peer interactions in shaping educational choices can be described as follows:

$$\mathbf{y} = \boldsymbol{\alpha}_C + \boldsymbol{\gamma}\mathbf{X} + \boldsymbol{\delta}\mathbf{G}\mathbf{X} + \beta\mathbf{G}\mathbf{y} + \boldsymbol{\epsilon}.$$

where $\mathbf{y}$ represents the outcome variable, namely students' educational choices. The term $\boldsymbol{\alpha}_C$ captures class fixed effects, while $\boldsymbol{\gamma}\mathbf{X}$ represents the influence of students' own characteristics (such as age, sex, family background, and soft skills). The term $\boldsymbol{\delta}\mathbf{G}\mathbf{X}$ captures the effects of the characteristics of students' classmates, and $\beta\mathbf{G}\mathbf{y}$ captures the influence of classmates' educational choices. The error term $\boldsymbol{\epsilon}$ is assumed to satisfy $E(\boldsymbol{\epsilon} \mid \boldsymbol{\alpha}_C, \mathbf{X}, \mathbf{G}) = 0$, meaning that both students' characteristics $\mathbf{X}$ and the network structure $\mathbf{G}$ are exogenous with respect to the outcome variable $\mathbf{y}$.

It is evident that, in order to apply this kind of model to analyse peer effects, it is necessary to design a survey that, beyond the students' outcome and characteristics, thoroughly collect classes network structures $\mathbf{G}$, preserving the linkage among $\mathbf{G}$, $\mathbf{X}$ and $\mathbf{y}$.

3.1 Survey Design at Work

The network research design was implemented through a structured set of steps, summarised as follows. First, a quota sample was drawn from high schools in the Campania region. A total of 28 schools were selected from a population of 324, using stratification based on provinces and school types. Specifically, the sample was divided into 20 strata, combining the five provinces (Avellino, Benevento, Caserta, Naples, and Salerno) with four school types (lyceum, vocational, technical, and multi-track institutions), as described in Table 1. The province of Naples, the largest stratum, includes six lyceums, three technical schools, one vocational school, and two multi-track schools. Salerno is represented by eight schools in total, while the provinces of Avellino, Benevento, and Caserta each contribute with two or three schools. Within the selected schools, a random sample of 91 classes was drawn. In total, 1,551 students participated in the study. In line with the quotas and population strata, lyceums and schools from the provinces of Salerno and Benevento were oversampled to ensure adequate presence of scholastic oriented tracks and different territorial contexts. Conversely, multi-track schools and schools in Naples and Caserta were undersampled. Despite the inherent limitations and potential biases associated with the selected sample and the complexity of conducting research in school settings, the resulting data can be considered robust and suitable for further analysis.

An online self-administrated questionnaire was presented by the research staff in February 2025 to students attending the final year of the sample of high schools selected. The instrument was implemented using the SurveyMonkey platform and was structured in three sections. In the first section students' sociodemographic characteristics, including age, gender, parents' education level and employment status, and individual soft skills ($\mathbf{X}$) were gathered. The second section focused on post-graduation choices, collecting information on the intention to pursue higher education, the field of study, the preference to attend a university within or outside the region of residence ($\mathbf{y}$). The final section was devoted to describe interactions among classmates. The network data ($\mathbf{G}$) are collected asking students to indicate the classmates with whom they have the closest ties in terms of friendship, advice in school-related activities, emotional support for personal matters, and appraisal support related to educational and professional choices after high school. For each question, the survey participants might have indicated up to five students in order to obtain a list of around 20 different classmates, reconstructing a whole network structure within each classroom able to explore social influence and social selection mechanisms among peers during educational transitions.

To ensure participants' anonymity during the network data collection phase, a protocol was implemented that avoided directly revealing students' names.

Each school provided class rosters in which student names were paired with randomized codes. These lists were accessible only to the teachers; the research staff had no access to this information and was solely responsible for facilitating the survey by providing technical instructions. Students completed the questionnaire by selecting the code corresponding to their own name and responding to network-related questions using the codes associated with their classmates. This method allowed for the collection of data on various types of peer relationships without revealing the identities of the respondents. At the end of the survey, teachers also provided the codes of students who did not participate, enabling researchers to account for missing nodes and ties during the data cleaning process. Although separate data collectors were created for each school and classroom, the responses were later integrated into a unified database. This procedure preserved student privacy, ethical standards and principles of transparency in the protection of personal data, while successfully reconstructing the social network structure within each class at school.

During the data cleaning process, the proposed strategy highlighted the need for procedures to manage missing data, in order to preserve the integrity of the information and to reconstruct network data that are both complete and representative. This issue has been extensively studied [16,17], and a wide range of solutions for imputing missing links has been proposed within the field of network analysis. In this context, advanced imputation methods will be employed, taking into consideration the specific structural properties of the network and the mechanisms underlying link formation. Finally, additional analyses will pursue two objectives. First, we will investigate the heterogeneity of peer effects on educational choices, examining how these effects vary according to students' backgrounds, school types, and geographic contexts. Second, we will aim to identify particularly vulnerable or influential subgroups within educational settings.

Table 1. Distribution of high schools in the population and sample (in brackets) by Campania provinces and School Types (20 strata).

School types	Avellino	Benevento	Caserta	Naples	Salerno	Total
Lyceum	8 (2)	4 (2)	22 (2)	67 (6)	26 (3)	**127** (15)
Vocational	2 (0)	1 (0)	1 (1)	11 (1)	8 (1)	**23** (3)
Technical	2 (0)	1 (0)	7 (0)	16 (3)	5 (2)	**31** (5)
Multi-track	13 (0)	12 (1)	21 (2)	69 (2)	28 (2)	**143** (5)
Total	**25** (**2**)	**18** (**3**)	**51** (**3**)	**163** (**12**)	**67** (**8**)	**324** (**28**)

Acknowledgement. We acknowledge financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 104 published on 2.2.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union – NextGenerationEU– Project Title "From high school to university: Assessing peers' influence in educational inequalities and

performances" –CUP F53D23006150006- Grant Assignment Decree No. 1060 adopted on 07/17/2023 by the Italian Ministry of University and Research (MUR).

References

1. Palardy, G.J.: High school socioeconomic composition and college choice: multilevel mediation via organizational habitus, school practices, peer and staff attitudes. School Effect. School Improv. **26**(3), 329–353 (2015). https://doi.org/10.1080/09243453.2014.965182
2. Rosenqvist, E.: Two functions of peer influence on upper-secondary education application behavior. Sociol. Educ. **91**(1), 72–89 (2018). https://doi.org/10.1177/0038040717746113
3. Andersen, S.C., Hjortskov, M.: The unnoticed influence of peers on educational preferences. Behav. Public Policy **6**(4), 530–553 (2022). https://doi.org/10.1017/bpp.2019.14
4. Smith, E.: Peer preferences and educational decisions: heterogeneous associations across student socioeconomic status. Br. J. Sociol. Educ. **44**(2), 374–393 (2023)
5. Sewell, W.H., Haller, A.O., Portes, A.: The educational and early occupational attainment process. Am. Sociol. Rev. **34**(1), 82–92 (1969). https://doi.org/10.2307/2092789
6. Froehlich, D.E., Van Waes, S., Schäfer, H.: Linking quantitative and qualitative network approaches: a review of mixed methods social network analysis in education research. Rev. Res. Educ. **44**(1), 244–268 (2020). https://doi.org/10.3102/0091732X20903311
7. Kadushin, C.: Who benefits from network analysis: ethics of social network research. Soc. Netw. **27**(2), 139–153 (2005). https://doi.org/10.1016/j.socnet.2005.01.005
8. D'Angelo, A., Ryan, L.: The presentation of the networked self: ethics and epistemology in social network analysis. Soc. Netw. **67**, 20–28 (2021). https://doi.org/10.1016/j.socnet.2019.06.002
9. Tubaro, P.: Whose results are these anyway? Reciprocity and the ethics of "giving back" after social network research. Soc. Netw. **67**, 65–73 (2021). https://doi.org/10.1016/j.socnet.2019.10.003
10. Manski, C.: Identification of endogenous social effects: the reflection problem. Rev. Econ. Stud. **60**(3), 531–542 (1993). https://doi.org/10.2307/2298123
11. Bramoullé, Y., Djebbari, H., Fortin, B.: Peer effects in networks: a survey. Annu. Rev. Econ. **12**(1), 603–629 (2020). https://doi.org/10.1146/annurev-economics-020320-033926
12. Cornwell, B., Hoagland, E.: Survey methods for social network research. In: Handbook of Health Survey Methods, pp. 275–313 (2014). https://doi.org/10.1002/9781118594629.ch11
13. Marsden, P.V.: Network data and measurement. Annu. Rev. Sociol. **16**, 435–463 (1990)
14. Task, C., Clifton, C.: A guide to differential privacy theory in social network analysis. In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Istanbul, Turkey, pp. 411–417 (2012). https://doi.org/10.1109/ASONAM.2012.73

15. Borgatti, S.P., Molina, J.L.: Ethical and strategic issues in organizational social network analysis. J. Appl. Behav. Sci. **39**(3), 337–349 (2003). https://doi.org/10.1177/0021886303258111
16. Huisman, M., Krause, R.W.: Imputation of missing network data. In: Alhajj, R., Rokne, J. (eds.) Encyclopedia of Social Network Analysis and Mining. Springer, New York (2017). https://doi.org/10.1007/978-1-4614-7163-9_394-1
17. Krause, R.W., Huisman, M., Steglich, C., Snijders, T.A.B.: Missing network data: a comparison of different imputation methods. In: Proceedings of 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, Spain, pp. 159–163 (2018). https://doi.org/10.1109/ASONAM.2018.8508716